**Proposal**

**for a**

**Bachelor of Science in Data Science**

*SPONSORED BY:*

COMPUTER SYSTEMS TECHNOLOGY DEPARTMENT

SCHOOL OF TECHNOLOGY AND DESIGN

NEW YORK CITY COLLEGE OF TECHNOLOGY

THE CITY UNIVERSITY OF NEW YORK

Anticipated Initiation: Fall 2019

Governance Approval Date: TBD

**PROGRAM IDENTIFICATION**

COLLEGE New York City College of Technology

The City University of New York

PROGRAM TITLE Data Science

DEGREE Bachelor of Science

CONTACT PEOPLE Dr. Bonne August

Provost and Vice President of Academic Affairs

New York City College of Technology

BAugust@citytech.cuny.edu

Dr. Pamela Brown

Associate Provost

New York City College of Technology

PBrown@citytech.cuny.edu

Kevin Hom

Dean of the School of Technology and Design

New York City College of Technology

[KHom@citytech.cuny.edu](mailto:KHom@citytech.cuny.edu)



**TABLE OF CONTENTS**

Curriculum Modification Proposal Form 5

Section 1: Program Introduction 7

1.1 Introduction 7

1.2 Purpose and Goals 8

Section 2: Need and Justification 10

2.1 Employment Opportunities After Graduation 10

2.1.1 Overview of job market 10

2.1.2 Job market growth 11

2.1.3 Compensation Potential 11

2.2 Related Undergraduate and Graduate Programs in the New York area (as of November 5, 2018). 12

2.2.1 Related Programs in CUNY 12

2.2.2 Related Programs in SUNY 13

Section 3: Student Interest and Anticipated Enrollment 14

Section 4: Curriculum 17

4.1 Overview of the Courses in the Curriculum 17

4.2 Anticipated Learning Outcomes 18

4.3 Courses Required to Complete the Program 19

4.4 Example of a Four Years Course Sequence 21

4.5 Catalog Description of the Eight New Courses 24

4.5.1. CST 2312: Information and Data Management I 24

4.5.2. CST 2402: Introduction to Data Science 24

4.5.3. CST 3502: Data Mining 24

4.5.4. CST 3512: Information and Data Management II 25

4.5.5. CST 3602: Data Visualization 25

4.5.6. CST 4702: Machine Learning Fundamentals 25

4.5.7. CST 4802: Information Retrieval 26

4.5.8. CST 4812: Natural Language Processing 26

4.6 Mapping Anticipated Learning Outcomes to the Courses 27

Section 5: Faculty 28

Section 6: Cost Assessment 30

Section 7: Acknowledgements 31

References 32

New Course Proposal #1: CST2312 Information and Data Management I 33

New Course Proposal #2: CST2402 Introduction to Data Science 49

New Course Proposal #3: CST3502 Data Mining 64

New Course Proposal #4: CST3512 Information and Data Management II 82

New Course Proposal #5: CST3602 Data Visualization 99

New Course Proposal #6: CST4702 Machine Learning Fundamentals 115

New Course Proposal #7: CST4802 Information Retrieval 131

New Course Proposal #8: CST4812 Natural Language Processing 147

Proposal to Change Prerequisites for: CST 4714: Database Administration 164

Appendix A: Letters of Support 168

Appendix B: Sample Job Postings 173

Appendix C: Draft Articulation Agreements 188

Appendix D: Colleges Offering Degrees in Data Science 196

Appendix E: CST Industry Advisory Board Meeting 199

Appendix F: Evidence of Consultation with Other Departments 201

**New York City College of Technology, CUNY**

# Curriculum Modification Proposal Form

This form is used for all curriculum modification proposals. See the [Proposal Classification Chart](http://openlab.citytech.cuny.edu/collegecouncil/files/2014/08/2013-10-09-Proposal_Classification_Chart.pdf) for information about what types of modifications are major or minor. Completed proposals should be emailed to the Curriculum Committee chair.

|  |  |
| --- | --- |
| **Title of Proposal** | Bachelor of Science in Data Science |
| **Date** | 1/20/2018 |
| **Major or Minor** | Major |
| **Proposer’s Name** | Dr. Candido Cabo  Dr. Elena Filatova  Dr. Hong Li  Dr. Elizabeth Milonas  Dr. Ashwin Satyanarayana |
| **Department** | Computer Systems Technology |
| **Date of Departmental Meeting in which proposal was approved** | 11/17/2017 and 12/08/2017 |
| **Department Chair Name** | Dr. Hong Li |
| **Department Chair Signature and Date** | 1/15/2018 |
| **Academic Dean Name** | Kevin Hom |
| **Academic Dean Signature and Date** | 2/22/18 |
| **Brief Description of Proposal**  (Describe the modifications contained within this proposal in a succinct summary. More detailed content will be provided in the proposal body. | The Department of Computer Systems Technology of the School of Technology and Design at New York City College of Technology proposes a Bachelor of Science(BS) degree program in Data Science (DS). DS is about scientific methods, algorithm development and technology to extract knowledge or insights from data in various forms in order to solve analytically complex problems. The proposal includes design of the BS in Data Science degree program and eight new courses required in the program |
| **Brief Rationale for Proposal**  (Provide a concise summary of why this proposed change is important to the department. More detailed content will be provided in the proposal body). | The DS program prepares graduates to   * Identify and use appropriate data mining and analysis techniques to solve specific problems * Design and develop algorithms to identify reasonable trends in large data sets * Analyze big data for optimal data knowledge and business problem solving, for example, product recommendations or understanding consumer behavior   The DS program equips students with the technical skills necessary to enter careers in the design, application development, data management and data analytics of various data from different fields such as healthcare, business, biology, etc. According to various studies, jobs in the data science field are in high demand. Only two colleges within CUNY expose undergraduate students to the area of Data Analytics. Baruch College, offers a Data Analytics track as part of their CIS Major (Bachelor of Business Administration Degree). The College of Staten Island offers a certificate program (consisting of five courses) in Business Analytics of Large-Scale Data. |
| **Proposal History**  (Please provide history of this proposal: is this a resubmission? An updated version? This may most easily be expressed as a list). | 1/22/2018: 1st submission of this proposal  3/3/2018: 2nd revision  3/13/2018: 3rd revision  10/26/2018:4th revision  11/6/2018: 5th revision  11/7/2018: 6th revision  11/12/2018:7th revision  11/26/2018: 8th revision  12/02/2018: 9th revision |

# Section 1: Program Introduction

## Introduction

Data Science is “the study of data” combining statistics, informatics and computing in an effort to understand all aspects of data including environment, organization and the social and ethical implications. The results of such a data driven understanding is knowledge discovery leading to intelligent decision making (Cao, 2017). Students enrolled in the Bachelor of Science (BS) in Data Science program will gain a quantitative understanding of logical, statistical and business analysis as well as computing principles. They will be versed in various computing methods and tools used to analyze and understand massive amounts of data. Graduates of the program can benefit from the abundance of employment opportunities available both in public and private sectors within the tristate area and beyond. In addition, graduates of the program can pursue graduate studies at several graduate schools including the CUNY Graduate Center. The BS in Data Science is consistent with the mission of New York City College of Technology in that it will provide students with the educational foundation as well as the command of the technical skills necessary to succeed in the domains where Data Science is applied. The degree will offer a balance of technical and liberal arts courses in an effort to foster intellectual curiosity, an appreciation for the aesthetic dimension of life and work and a respect for cultural diversity. The BS in Data Science is also consistent with the mission of the Computer Systems Technology Department (CST) as courses offered will emphasize both the theoretical and practical foundation in the Data Science domain and will emphasize a “hands-on” approach for maximum learning and retention of concepts and practices.

## Purpose and Goals

The purpose and goal of the BS in Data Science program is to prepare graduates with the technical skills necessary to enter careers in the Data Science field, which is one of the fastest-growing fields today. According to Forbes magazine and industry experts (Kobielus, 2016) (Sentz, 2016) data scientists are in high demand and will continue to be in high demand as the big data market continues to gain momentum. Experts state (Kobielus, 2016) that high priority enterprise application projects require skilled data scientists who have knowledge in machine learning and cognitive computing. These key topics in addition to other integral topics in the field of data science are studied as part of the BS in Data Science program. Graduates of the BS in Data Science will acquire the conceptual and computational understanding as well as the technical expertise needed to excel in this field. In addition, they will also gain the real-world experience necessary to meet the industry need for skilled Data Scientists.

The BS in Data Science curriculum focuses on the knowledge and skills required to meet industry’s data challenges. Students will also complete the CUNY Pathway general education requirements which provide students with a solid liberal arts education. Students will also be required to complete several courses in mathematics including probability, discrete structures, linear algebra and calculus. Students will take two foundational courses in an area of their choosing such as business, economics, marketing, biomedical, and healthcare providing them with an understanding of the role data science plays within their chosen field. The courses in the major will consist of key computational topics such as natural language processing, machine learning, data mining, data retrieval and visualization as well as mathematical and analytical topics which will prepare students to meet the data challenges of the field. A combination of lecture, hands-on labs, group work and an internship will provide students with the opportunity to learn and become skilled data scientists who can then apply those skills on the job. In the first two years of the program students become well versed in the foundational principles of computer systems and data science topics. In the last two years, courses are designed to cover the breadth and depth of the data science field.

Consistent with City Tech’s educational goals, the BS in Data Science is designed to provide a well-rounded interdisciplinary education for the new generation of Data Scientists who will be prepared to meet the challenges of the domain. Graduates of the program should be able to secure employment as data scientists in various domains including financial, business, marketing and medical as well as others. In addition, graduates should be able to make positive contributions to the field.

# Section 2: Need and Justification

The BS in Data Science program has been designed so graduates of the program: (1) can benefit from the excellent employment opportunities in the public and private sectors in the tristate area and beyond; (2) can pursue graduate studies. This new program will significantly increase the opportunities available to City Tech and CUNY students transferring to City Tech for employment or graduate education and who are interested in the application of data science techniques and methods to a variety of domains, such as, finance, biology, medicine, business, etc.

## 2.1 Employment Opportunities After Graduation

### 2.1.1 Overview of job market

According to data produced by Emsi (http://www.economicmodeling.com) and CareerBuilder (https://hiring.careerbuilder.com), there were, on average, 2,900 unique job postings active per month for data scientists over the past nine months. (NOTE: The total number of job postings is, of course, much higher because employers advertise on many different job sites. Emsi has de-duplicated those postings to the real number of locations / businesses posting ads for those jobs.) The top states for data science job postings are California, Washington, New York, Virginia, and Massachusetts, while the top metros are San Jose, Seattle, New York, Washington D.C., Chicago, and San Francisco. 59% of all Data Science and Analytics job demand is in Finance and Insurance, Professional Services, and IT.

Companies looking for this talent represent virtually every industry. Oracle, Microsoft, Amazon, Apple, Booz Allen Hamilton, GE, State Farm, Walmart, Facebook, United Health Care, Aetna, AT&T, Intel, IBM, Nielsen, KPMG, eBay and many more all show up prominently in job postings for data scientists. Related job titles include:

* Data analyst
* Data scientist
* Data mining engineer
* Machine learning engineer
* Data architect
* Hadoop engineer
* Data warehouse architect
* Commercial intelligence manager
* Competitive intelligence analyst
* Consultant, strategic business and technology intelligence

### 2.1.2 Job market growth

Forbes predicts that annual demand for the fast-growing new roles of data scientist, data developers, and data engineers will reach nearly 700,000 openings by 2020. By 2020, the number of jobs for all US data professionals will increase by 364,000 openings to 2,720,000 according to IBM. Data Science and Analytics jobs remain open an average of 45 days, five days longer than the market average indicating difficulty in finding qualified talent (Career, 2017).

### 2.1.3 Compensation Potential

According to Forbes (Forbes, 2016) (Forbes, 2017), jobs requiring machine learning skills are paying an average of $114,000. Advertised data scientist jobs pay an average of $105,000 and advertised data engineering jobs pay an average of $117,000. According to Glassdoor, the annual median salary for a data scientist is $121,353.

In his article, [Big Data Salaries: An Inside Look](https://datajobs.com/big-data-salary), DataJobs.com founder Frank Lo conducted extensive research to identify the following job titles and salary ranges (Lo, 2017):

* Data analyst—entry level | Annual salary: $50,000–$75,000
* Data analyst—experienced | Annual salary: $65,000–$110,000
* Data science/analytics manager—1 to 3 direct reports | Annual salary: $90,000–$140,000
* Big data engineer—junior/generalist | Annual salary: $70,000–$115,000
* Big data engineer—domain expert | Annual salary: $100,000–$165,000

According to O’Reilly’s [2016 Data Science Salary Survey](http://www.oreilly.com/data/free/files/2016-data-science-salary-survey.pdf) (O'Reilly, 2016), experience is one of the most important factors in a data scientist’s salary. For every year of experience, data science professionals make an average of $2,000 to $2,500 more.

A 2017 [Burtch Works study](http://www.burtchworks.com/files/2017/05/DS-2017-Salary-Growth.pdf) (Burtch, 2017) of data science salaries reported the latest salary trends based on experience:

* *Entry-level data scientist salary.* Despite a recent influx of early-career professionals, the median starting salary for a data scientist remains high at $95,000.
* *Mid-level data scientist salary.* The median salary for a mid-level data scientist is $126,000. If this data scientist is also in a managerial role, the median salary rises to $190,000.
* Experienced data scientist salary. *The median salary for experienced data science professionals is $157,000—while the median salary for experienced manager-level professionals is considerably higher at $250,000.*

## 2.2 Related Undergraduate and Graduate Programs in the New York area (as of November 5, 2018).

### 2.2.1 Related Programs in CUNY

Three colleges within CUNY offer exposure to undergraduate students in the area of Data Analytics and Data Science.

* **Baruch College**, offers a Data Analytics track as part of their Computing in Information Systems (CIS) Major - Bachelor of Business Administration (BBA) Degree; as well as Data Analytics Minor (as one of IS Business Minors) for the Baruch Weissman School or Arts and Sciences or in the School of Public Affairs.
* **John Jay College of Criminal Justice** offers a BS degree in Applied Mathematics with concentration in Data Science.

The Baruch programs are designed to provide the background on “[…] how to aggregate, analyze, and monetize the growing surge of available data. The new track – Data Science and Analytics – would provide a strong foundation in technology, statistics, and quantitative modeling that is needed to develop business intelligence and drive organizational decision-making.”

The CSI programs focus on Data Science methods and techniques application to the fields of Business Management and Intelligence, Econometrics, and Marketing.

The John Jay Data Science concentration “presents the principles of data representation, big data management, and statistical modeling.”

At the Master’s level, as of November 5, 2018, four schools offer related degrees:

* **City College of New York** offers an MS program in Data Science and Engineering.
* **CUNY Graduate Center** offers an MS program in Data Science with topics including artificial intelligence, algorithms for big data, data visualization, information retrieval, data mining, and machine learning.
* **CUNY School of professional studies** offers an online degree program which provides foundational knowledge and hands-on programming competencies, resulting in project-based work samples.
* **Queens College**, a joint program between the Computer Science department and the department of Sociology offers and MS in Data Science.

### 2.2.2 Related Programs in SUNY

Two universities within SUNY offer master’s degrees in the area of data analytics. SUNY Albany offers a new program in data science in their department of Mathematics and Statistics. At SUNY Buffalo, the MS in Engineering Sciences degree on Data Sciences program provides students with a core foundation in big data and analysis by obtaining knowledge, expertise, and training in data collection and management, data analytics, scalable data-driven discovery, and fundamental concepts.

**For a more detailed listing of colleges (beyond CUNY and SUNY) offering Data Science degrees, see: Appendix D: Colleges Offering Degrees in Data Science**

# Section 3: Student Interest and Anticipated Enrollment

**Student Interest:**

According to the eFinancialCareers.com (a targeted career resource and employment website from the Wall Street Journal) “Most industries are struggling to find data science expertise, but Wall Street especially has particularly keen to hire in this area.” City Tech students, being New York City locals, are aware of this situation and are interested in the Data Science careers not only in the Finance industry, but in other industries including Medical, Biology, Hospitality, etc. In addition, many City Tech students are interested in continuing their education in the field of Data Science. Several universities in the New York City and its vicinities offer Master of Science degrees in Data Science. However, only a few universities offer Bachelor level degrees in Data Science related areas. The BS in Data Science degree proposed here will increase the opportunity for students in and around New York City.

A survey of 198 City Tech CST students (both AAS and BTech) was conducted in the Spring of 2017 after offering the students a brief description of a tentative Data Science program at City Tech. The following results were obtained:

1. How familiar are you with the terms big data, data analysis, data science?

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| This is the first time I see these terms. | I have heard these terms but do not have a good grasp of what they mean. | I am familiar with these terms and have a general understanding of what they mean. | I have a good understanding of these terms | Total |
| **11(5.56%)** | **77(38.89%)** | **85(42.93%)** | **25(12.63)** | **198 (100%)** |

1. Based on the description of the Bachelor of Science Degree in Data Science, would you be interested in pursuing this Bachelor of Science degree at City Tech?

|  |  |
| --- | --- |
| Answer Choice | Responses |
| Yes | 139 (70.20%) |
| No | 59 (29.80%) |
| Total | 198 (100%) |

Our preliminary conclusions are: 1) many City Tech CST students have an understanding of the various aspects of the Data Science field and, 2) students are interested in pursuing a BS degree in Data Science at City Tech.

**Enrollment Outlook**:

We can estimate enrollment outlook for the next five years based on the rate of growth of the CST student population for the last five years (2012-2016). This estimate is partly based on data acquired from Enrollment Trends found in the Assessment and Institutional Research (AIR) website (<http://air.citytech.cuny.edu/data-dashboard/enrollment-trends-fall>), and calculations of the average rate of growth over the five year period 2012-2016. These calculations show a 19% average growth for new incoming CST freshmen students and a 7% average growth for transfer students coming into the CST department. Based on these average growth rates and the survey results indicating a 75% interest in the BS in Data Science, we can estimate the following growth over the next five-year period 2017- 2021:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Year | No. of  CST Students | Estimated no. of students in BS/DS | No. of Transfer Students | Estimated no. of transfer students in BS/DS |
| 2019 | 409 | 76 | 219 | 41 |
| 2020 | 437 | 82 | 234 | 43 |
| 2021 | 467 | 87 | 250 | 46 |
| 2022 | 499 | 93 | 267 | 50 |
| 2023 | 499 | 93 | 267 | 50 |

For incoming new freshmen CST students, the rate of growth for this population in the last five years was 19%. For transfer students coming into the CST department, the rate of growth is 7%.

**Potential Students:**

City Tech students in the Bachelor of Technology (BTech) curriculum with an interest in Data Science and who meet the requirements of the new program will be potential candidates for the Bachelor of Science in Data Science. In addition, students with Associate Degrees in Computer Science, Computer Information Science and Mathematics from any of the CUNY community colleges including Borough of Manhattan, La Guardia, Kingsborough, Bronx Community College, Lehman College, York College, Medgar Evers College and the College of Staten Island would also be potential candidates for this program.

# Section 4: Curriculum

## Overview of the Courses in the Curriculum

The Bachelor of Science in Data Science curriculum includes:

1. The CUNY Pathways general education requirements (required core, flexible core, college option) which will help students develop a solid liberal arts education (**42 credits**).
2. Required general education courses in mathematics which will help students develop a strong foundation in the mathematics and statistics necessary to succeed in subsequent Data Science courses (**18 credits**).
3. Computer systems core courses taken mostly during the first two years of the program (**24 credits**) will introduce students to the fundamental concepts of Computer Systems (Databases, Problem Solving, Programming Fundamentals, Introduction to Security and Web Programming) and Data Science (Information and Data Management I, and Introduction to Data Science).
4. Data Science core with advanced topics are taken during the last two years of the program (**27 credits**). The program designed to balance breadth and depth in Data Science, offers different topics in the field of Data Science including: Data Mining, Information and Data Management II, Information Retrieval, Natural Language Processing, Machine Learning Fundamentals and Data Visualization.
5. Data Science Major Electives (**6 credits**). Those courses are designed so that students can take two major electives in a practical field of data science application (e.g. Finance, Healthcare, Economics and Biomedical Engineering, Internet of Things and Artificial Intelligence).
6. Internship (**3 credits**). Students are exposed to the practice of the profession in research or industry.

## Anticipated Learning Outcomes

The curriculum has been designed to meet the following learning outcomes:

**Student Learning Outcomes: General**

Students demonstrate:

1. An ability to apply the knowledge, techniques, skills, and modern tools of the discipline to Data Science activities.
2. An ability to apply a knowledge of mathematics, science, engineering, and technology to Data Science problems that require application of principles and practical knowledge.
3. An ability to conduct standard tests and measurements, and to conduct, analyze, and interpret experiments.
4. An ability to function effectively as a member of a technical team.
5. An ability to apply written, oral, and graphical communication in both technical and non-technical environments; and an ability to identify and use appropriate technical literature.
6. Demonstrate an understanding of the need for and an ability to engage in self-directed continuing professional development.
7. Demonstrate an understanding of and a commitment to address professional and ethical responsibilities, including a respect for diversity.
8. A commitment to quality, timeliness, and continuous improvement in professional practice.

**Student Learning Outcomes: Discipline-Specific**

Data Science students demonstrate knowledge and hands on competence in:

1. Analyzing, designing, and implementing data science algorithms along while performing analytics.
2. Demonstrating a deep knowledge of: Data Mining, Data Management, Data Analytics, Information Retrieval; enabling students to gain employment in the data science field.
3. Demonstrate an understanding of how management uses data science systems to operate business enterprises.
4. Demonstrate a deep understanding of techniques for visualizing multivariate, temporal, text-based, geospatial, hierarchical, and network/graph-based data.
5. Demonstrate an understanding of the ethics and privacy issues that arise from the process of managing and mining user data.

## 4.3 Courses Required to Complete the Program

The proposed curriculum for the Bachelor of Science in Data Science is detailed below. The curriculum reflects City Tech’s General Education requirements:

|  |  |
| --- | --- |
| **General Education Common Core** | **42 credits** |
| **Required General Education** | **18 credits** |
| **Computer Systems Fundamentals** | **24 credits** |
| **Data Science Core** | **27 credits** |
| **Data Science Major Electives** | **6 credits** |
| **Internship** | **3 credits** |
| **TOTAL** | **120 credits** |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **General Education Core:** |  | | |  |
| **Required Core:** |  | | |  |
| English Comp1 | ENG1101 | | | 3 |
| English Comp 2 | ENG1121 | | | 3 |
| Quantitative Reasoning | Any | | | 3 |
| Life & Physical Science | Any | | | 3 |
|  |  | | |  |
| **Flexible Core:** |  | | |  |
| World Culture and Global Issues | Any | | | 3 |
| US Experience and Diversity | Any | | | 3 |
| Creative Expression | Any | | | 3 |
| Individual and Society | Any | | | 3 |
| Scientific World | Any | | | 3 |
| Additional 6th course | Any | | | 3 |
|  |  | | |  |
| **College Option:** |  | | |  |
| Speech/Oral Communication | Any | | | 3 |
| Interdisciplinary Course | Any | | | 3 |
| Additional LA I | Any | | | 3 |
| Additional LA II | Any | | | 3 |
|  |  | | |  |
| **Sub-Total** |  | | | **42** |
| **Program-Specific Degree Requirements:** | | | | |
| Calculus I | | | MAT1475 | 4 |
| Calculus II | | | MAT1575 | 4 |
| Probability and Statistics I | | | MAT2572 | 4 |
| Discrete Structures and Algorithms I | | | MAT2440 | 3 |
| Introduction to Linear Algebra | | | MAT2580 | 3 |
|  | |  | |  |
| **Sub-Total** | |  | | **18** |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | | | | | | |
|  |  | | | |  | |
| **Computer Systems Fundamentals:** | | | | | | |
| Introduction to Computer Systems | CST1100 | | | | 3 | |
| Problem Solving with Computer Programming | CST1101 | | | | 3 | |
| Programming Fundamentals | CST1201 | | | | 3 | |
| Database Fundamentals | CST1204 | | | | 3 | |
| Information and Data Management I | CST2312 | | | | 3 | |
| Web Programming I | CST2309 | | | | 3 | |
| Introduction to Data Science | CST2402 | | | | 3 | |
| Introduction to Security | CST2410 | | | | 3 | |
|  |  | | | |  | |
| **Sub-Total** |  | | | | **24** | |
|  | |  | | |  | |
|  | |  | |  | | |
|  | |  | |  | | |
| **Data Science Core:** | |  | | |  |
| Information and Data Management II | | CST3512 | | | 3 |
| Data Mining | | CST3502 | | | 3 |
| Object Oriented Programming | | CST3513 | | | 3 |
| Data Visualization | | CST3602 | | | 3 |
| Data Structures | | CST3650 | | | 3 |
| Machine Learning Fundamentals | | CST4702 | | | 3 |
| Database Administration | | CST4714 | | | 3 |
| Information Retrieval | | CST4802 | | | 3 |
| Natural Language Processing | | CST4812 | | | 3 |
|  | |  | | |  |
| **Sub-Total** | |  | | | **27** |
|  | |  | | |  |
| Internship | | CST4900 | | | 3 |
|  | |  | | |  |
| **Sub-Total** | |  | | | **3** |
|  | |  | | |  |
| **Data Science major electives ( take any 2 courses)** | | | | | |
|  | | | | | |
|  | | | | | |
| Financial management | | BUS2339 | | | 3 |
| Financial Forecasting | | BUS2341 | | | 3 |
|  | |  | | |  |
| Macroeconomics | | ECON1101 | | | 3 |
| Money and Banking | | ECON2301 | | | 3 |
|  | |  | | |  |
| Medical Informatics Fundamentals | | MED2400 | | | 3 |
| Healthcare Databases | | MED4229 | | | 3 |
|  | |  | | |  |
| Fundamental Healthcare Data Analytics | | BMET4741 | | | 3 |
| Advanced Healthcare Data Analytics | | BMET4842 | | | 3 |
|  | |  | | |  |
| Internet of Things | | CET4925 | | | 3 |
| Introduction to Artificial Intelligence | | CET4973 | | | 3 |
|  | |  | | |  |
| Probability and Statistics II | | MAT3672 | | | 3 |
| Computational Statistics with Applications | | MAT4672 | | | 4 |
| \*Machine Learning for Physics and Astronomy | | PHYS3600 | | | 3 |
|  | |  | | |  |
| **Sub-Total** | |  | | | **6/7** |
|  | |  | | |  |
| **Total Credits** | |  | **120-121** | | |

\*Specific course indicates double duty course, i.e. program degree requirement that also meet general education requirements. Choose another elective to complete 120 credits if choosing to take double duty.

## 4.4 Example of a Four Years Course Sequence

|  |  |  |
| --- | --- | --- |
| **FIRST YEAR** | | |
| ***First Semester*** | | |
| CST 1100 | Introduction to Computer Systems | 3 |
| CST 1101 | Problem Solving | 3 |
| Required Core/Quantitative Res. | Any | 3 |
| Required Core/ENG 1101 | English Composition I | 3 |
| Flexible Core/World Cultures | Any | 3 |
| **TOTAL** | | **15** |
|  |  |  |
| ***Second Semester*** | | |
| CST 1201 | Programming Fundamentals | 3 |
| CST 1204 | Database Fundamentals | 3 |
| MAT 1475 | Calculus I | 4 |
| Required Core/Life &Physical Sc. | Any | 3 |
| Required Core/ENG 1121 | English Composition II | 3 |
| **TOTAL** | | **16** |
|  |  |  |
|  |  |  |
| **SECOND YEAR** | | |
| ***First Semester*** | | |
| CST 2312 | Information and Data Management I | 3 |
| CST 2309 | Web Programming | 3 |
| MAT 1575 | Calculus II | 4 |
| MAT 2440 | Discrete Structures and Algorithms I | 3 |
| Flexible Core/Any | Any | 3 |
| **TOTAL** | | **16** |
|  | | |
| ***Second Semester*** | | |
| CST 2402 | Intro to Data Science | 3 |
| CST 2410 | Intro to Computer Security | 3 |
| MAT 2572 | Probability and Statistics I | 4 |
| Flexible Core/US Experience | Any | 3 |
| **TOTAL** | | **13** |
|  | | |

|  |  |  |
| --- | --- | --- |
| **THIRD YEAR** | | |
| ***First Semester*** | | |
| CST 3512 | Information and Data Management II | 3 |
| CST 3513 | Object Oriented Programming | 3 |
| CST 3502 | Data Mining | 3 |
| Flexible Core/Scientific World | Any | 3 |
| College Option/LLAA I | Any | 3 |
| **TOTAL** | | **15** |
|  | | |
| ***Second Semester*** | | |
| CST 3602 | Data Visualization | 3 |
| CST 3650 | Data Structures | 3 |
| MAT 2580 | Intro to Linear Algebra | 3 |
| Flexible Core/Creative Expression | Any | 3 |
| College Option/LLAA II | Any | 3 |
| **TOTAL** | | **15** |
|  | | |
| **FOURTH YEAR** | | |
| ***First Semester*** | | |
| Data Science Major Elective I | Any | 3 |
| CST 4702 | Machine Learning Fundamentals | 3 |
| CST 4714 | Database Administration | 3 |
| College Option/Speech | Any | 3 |
| Flexible Core/Individual & Society | Any | 3 |
| **TOTAL** | | **15** |
|  | | |
| ***Second Semester*** | | |
| Data Science Major Elective II | Any | 3 |
| CST 4802 | Information Retrieval | 3 |
| CST 4812 | Natural Language Processing | 3 |
| CST 4900 | Internship | 3 |
| College Option/Interdisciplinary | Any | 3 |
| **TOTAL** | | **15** |

|  |
| --- |
|  |
|  |

## 4.5 Catalog Description of the Eight New Courses

### 4.5.1. CST 2312: Information and Data Management I

This course provides students with the introduction to the necessary informatics and intellectual tools to become efficient and effective information users. The course covers topics related to the digital infrastructure, acquisition, organization, management and curation of data. The course is structured around the Python tools for regular expression analysis, accessing data sources (crawling, Web APIs), analysis of structured data. At the end of the class the students complete a project to demonstrate the mastery of the technical topics discussed in class with an application to their domain of interest.

### 4.5.2. CST 2402: Introduction to Data Science

This is a fundamental course in the basic concepts, principles and ethical issues of the Data Science domain including: definition, framework, techniques, issues, business uses and ethics. Topics include data collection, processing and management, exploratory data analysis, data visualization and presentation and ethical issues. Projects involve the use of data science techniques and tools to solve business problems and improve business decision making.

### 4.5.3. CST 3502: Data Mining

Theory and practice of data mining methods for extracting knowledge from data using state-of-the-art software (WEKA) for solving scientific and business problems will be covered. Topics discussed in the class include: data preprocessing, classification and prediction, association mining, and cluster analysis. Projects either involve research in data mining or address a practical problem using the methods and tools introduced in the class.

### 4.5.4. CST 3512: Information and Data Management II

This course continues to familiarize students with the informatics and intellectual tools necessary for students to become efficient and effective information users. The course covers topics related to the digital infrastructure, management and curation of data both structured (record-based) and unstructured (such as text). For structured data (data series, data frames), the course introduces Time Series data analysis and the basics of data visualization; for unstructured data (text), the course introduces text mining techniques. The course is project-based and case studies of using Information and Data Management are discussed. Based on these case studies the students develop their own projects where they choose a data set, use the appropriate data processing techniques for data analysis, and present their finding and conclusions reached using the information and data management and analysis techniques discussed in class.

### 4.5.5. CST 3602: Data Visualization

This course introduces information visualization basics and provides knowledge about design and information literacy perspective, including what makes a good and bad visualization. The course covers basic visualization design techniques and evaluation principles, techniques for visualizing multivariate, temporal, text-based, geospatial, hierarchical, and network/graph-based data. The final project for the course is to acquire, parse, analyze a large dataset, and create meaningful, good visualization for this dataset.

### 4.5.6. CST 4702: Machine Learning Fundamentals

This course introduces fundamental machine learning algorithms and techniques, and their implementations applied to solving real world problems. Topics covered include Supervised learning (parametric/non-parametric algorithms, support vector machines, kernels, neural networks), unsupervised learning (clustering, dimensionality reduction, recommender systems, deep learning) and best practices in machine learning (bias/variance theory; innovation process in machine learning). The theory of machine learning and the practical know-how lead to numerous case studies with applications in text understanding (web search, anti-spam), medical informatics, audio, database mining, and other areas.

### 4.5.7. CST 4802: Information Retrieval

The course details basic and advanced concepts and principles of information retrieval including: definition, framework, uses and ethical issues. Issues related to interface, relevance, language processing, and indexing are discussed. Information retrieval techniques covering both effectiveness and run-time performance are covered as well as heuristics and algorithms used in information retrieval systems. Information retrieval mechanism used in both Websites and Web search engines are also discussed. Ethical issues including personal privacy, bias, informed consent, monitoring and surveillance are also examined and discussed.

### 4.5.8. CST 4812: Natural Language Processing

This course introduces a broad range of techniques that aim to read, understand, and extract information from natural language text resources. The course covers both knowledge-based and statistical approaches for syntax, semantics, and discourse. In addition to introducing theoretical aspects and methods of natural language processing (NLP), the course covers multiple applications such as Machine Translation, Summarization, Interactive Dialog Systems, etc.

## 4.6 Mapping Anticipated Learning Outcomes to the Courses

The following table maps CST courses to student outcomes a-k (see above: *section 4.2 Anticipated Student Outcomes*):

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Courses** | **a** | **b** | **c** | **d** | **e** | **f** | **g** | **h** | **i** | **j** | **k** |
| CST 1100: Introduction to Computer Systems | X | X |  |  | X | X | X | X | X | X | X |
| CST 1101: Problem Solving | X | X | X |  | X |  | X |  | X |  | X |
| CST 1201: Programming Fundamentals | X | X | X |  | X |  | X |  | X |  | X |
| CST 1204: Database Fundamentals | X | X | X |  | X |  | X |  | X |  | X |
| CST 2312: Information and Data Management I | X | X | X |  | X |  | X |  | X |  | X |
| CST 2309: Web Programming I |  | X | X |  | X |  | X |  |  |  |  |
| CST 2402: Introduction to Data Science | X | X | X |  | X |  | X |  |  |  |  |
| CST 2410: Introduction to Security | X | X | X | X | X | X | X |  |  |  |  |
| CST 3502: Data Mining | X | X | X | X | X | X | X | X | X | X | X |
| CST 3512: Information and Data Management II | X |  | X |  | X | X |  |  |  | X |  |
| CST 3513: Object Oriented Programming | X | X | X |  | X |  | X |  |  |  |  |
| CST 3602: Data Visualization |  |  |  |  |  |  | X | X | X | X |  |
| CST 3650: Data Structures | X | X | X | X | X | X | X |  |  |  | X |
| CST 4702: Machine Learning Fundamentals | X | X | X | X | X | X | X |  | X |  |  |
| CST 4714: Database Administration | X | X |  | X |  | X | X |  |  | X |  |
| CST 4802: Information Retrieval | X | X |  | X |  | X | X |  | X |  |  |
| CST 4812: Natural Language Processing | X | X | X | X | X | X | X | X | X | X | X |
| CST 4900: Internship |  |  |  |  |  |  |  |  | X | X | X |

# Section 5: Faculty

The Computer Systems Technology departments has 20 full time faculty and 90 adjunct professors and the most are qualified to teach all or a portion of the coursework related to the proposed B.S. degree program. We have full time faculty members who have done well published research in the related field and have developed the new courses proposed within the curriculum requirement.

There are faculty members from other departments such as Mathematics, Biology, who also conduct research in data analytics in various domain of interest.

Below is a list of core faculty members involved with the program:

1. Candido Cabo:
   1. Education: Ph.D. in Biomedical Engineering, Duke University.
   2. Areas of Interest: Computational Biology and High-Performance Computing.
2. Yu-Wen Chen:
   1. Education: Ph.D. in Computer Engineering, Iowa State University.
   2. Area of Interest: Big Data Optimization and Cloud computing.
3. Ossama Elhadary:
   1. Education: D.B.A. Maastricht School of Management (The Netherlands)
   2. Areas of Interest: Biomedical Information Technology and Biomedical Analytics.
4. Elena Filatova:
   1. Education: Ph.D. in Computer Science, Columbia University.
   2. Area of Interest: Natural Language Processing.
5. Hong Li:
   1. Education: Ph.D. in Mathematics, University of Oklahoma
   2. Area of Interest: Mathematical modeling, Artificial Neural Networks
6. Elizabeth Milonas:
   1. Education: Ph.D. in Computer Information Science, Long Island City University.
   2. Areas of Interest: Information Retrieval and Knowledge Organization.
7. Ashwin Satyanarayana:
   1. Education: Ph.D. in Computer Science, SUNY - Albany.
   2. Areas of Interest: Machine Learning and Data Mining

In addition, there are several faculty members in other Departments that can provide support and bring their expertise to the proposed curriculum, in particular, in the Mathematics, and Biology departments.

# Section 6: Cost Assessment

Current classroom equipment is sufficient for instruction delivery of all courses in the proposal. Other than our regular maintenance and replacement of current existing equipment, no additional equipment will be required.

However, the CST department currently has 20 full time faculty members and 90 adjunct faculty members providing academic support of more than two thousand students. We anticipate the total number of students will increase with the new proposed degree program. The department currently is recruiting for two positions of tenured track Assistant Professor.

# Section 7: Acknowledgements

The proposers would like to acknowledge the following for their support of the program:

* Kevin Hom, Dean of the School of Technology and Design
* Alexis Chaconis, Director of Admission Services
* Laina Karthikeyan, Chairperson of Biology Department
* Sandie Han, Chairperson of Mathematics Department
* Huseyin Yuce, Mathematics Department
* Jonathan Natov, Mathematics Department
* Boyan Kostanidov, Mathematics Department
* Urmi Ghosh-Dastidar, Mathematics Department
* Johann A Thiel, Mathematics Department
* Jean Hillstrom, Social Science Department
* Peter Parides, Social Science Department
* Evgenia Giannopoulou, Biology Department
* Joanne Weinreb, Biology Department
* Angel Rodriques, Technology Analyst at Microsoft
* Arup Das, JP Morgan Chase
* Eugene Kogan, Chief Technology Officer at Rainforest Alliance

# References

Burtch. (2017, May ). The Burtch Works Study, Salaries of Data Scientists . *https://www.burtchworks.com/files/2017/05/DS-2017-Salary-Growth.pdf*.

Cao, L. (2017). Data Science: Comprehensive Overview. *ACM Computing Surveys 50, no.3, article 43*, 1-35.

Career. (2017). Data Science Careers Outlook, University of Wisconsin. *https://datasciencedegree.wisconsin.edu/data-science/data-science-careers/*.

Forbes. (2016). Want To Become A Data Scientist? Where The Jobs Are And What Employers Are Looking For. *Forbes*, https://www.forbes.com/sites/emsi/2016/11/16/want-to-become-a-data-scientist-where-the-jobs-are-and-what-employers-are-looking-for/#753963535760.

Forbes. (2017). IBM Predicts Demand For Data Scientists Will Soar 28% By 2020. *Forbes*, https://www.forbes.com/sites/louiscolumbus/2017/05/13/ibm-predicts-demand-for-data-scientists-will-soar-28-by-2020/#40d109f7e3bd.

Kobielus, J. (2016). Big data analytics trends in 2017: James Kobielus’s predictions. *IBM Big Data & Analytics Hub December 5, 2016*, http://www.ibmbigdatahub.com/blog/big-data-and-analytics-trends-2017-james-kobielus-s-predictions.

Lo, F. (2017). Big Data Salaries: An Inside Look, Data Scientist Salaries, Data Analyst Salaries, DBA Salaries, etc. *https://datajobs.com/big-data-salary*.

O'Reilly. (2016). 2016 Data Science Salary Survey, Tools, Trends, What Pays (and What Doesn’t) for Data Professionals. *http://www.oreilly.com/data/free/files/2016-data-science-salary-survey.pdf*.

Sentz, R. (2016). Want to Become a Data Scientist? Where the Jobs are and What Employers are Looking for. *Forbes Magazine November 16, 2016*, https://www.forbes.com/sites/emsi/2016/11/16/want-to-become-a-data-scientist-where-the-jobs-are-and-what-employers-are.

# New Course Proposal #1: CST2312 Information and Data Management I

New York City College of Technology, CUNY

CURRICULUM MODIFICATION PROPOSAL FORM

This form is used for all curriculum modification proposals. See the [Proposal Classification Chart](http://openlab.citytech.cuny.edu/collegecouncil/files/2014/08/2013-10-09-Proposal_Classification_Chart.pdf) for information about what types of modifications are major or minor. Completed proposals should be emailed to the Curriculum Committee chair.

|  |  |
| --- | --- |
| **Title of Proposal** | **CST2312 Information and Data Management I** |
| **Date** | **Jan 4th, 2018** |
| **Major or Minor** | **Major** |
| **Proposer’s Name** | **Dr. Elena Filatova** |
| **Department** | **Computer Systems Technology** |
| **Date of Departmental Meeting in which proposal was approved** | **11/17/2017** |
| **Department Chair Name** | **Dr. Hong Li** |
| **Department Chair Signature and Date** | **1/16/2018** |
| **Academic Dean Name** | **Dean Hom** |
| **Academic Dean Signature and Date** | 2/22/18 |
| **Brief Description of Proposal**  (Describe the modifications contained within this proposal in a succinct summary. More detailed content will be provided in the proposal body. | Title: Information and Data Management I  This proposal explains the rationale for developing Information and Data Management I as a required course in the Data Science curriculum of the Computer Systems Technology Department. The course outline and sample syllabus are attached to the proposal. |
| **Brief Rationale for Proposal**  (Provide a concise summary of why this proposed change is important to the department. More detailed content will be provided in the proposal body). | Information and Data Management I is the first course in the two-course series. This two-course series provides a comprehensive introduction to programming with Python that goes beyond the basic concepts and focuses around data processing and analysis. In the modern world, data is the lifeblood of any organization and the driving force for development of new successful models for decision making. Competency in programming is an essential skill for successfully extracting information and knowledge form data. Also, this course is integral to the proposed Bachelors in Data Science curriculum program of the Computer Systems Technology department. |
| **Proposal History**  (Please provide history of this proposal: is this a resubmission? An updated version? This may most easily be expressed as a list). | This is a first submission. |
| **CUNY – Course Equivalencies**  Provide information about equivalent courses within CUNY, if any. | None |
| **Intent to Submit as Common Core**  If this course is intended to fulfill one of the requirements in the common core, then indicate which area. | N/A |

Please include all appropriate documentation as indicated in the Curriculum Modification Checklist.

For each new course, please also complete the New Course Proposal and submit in this document.

Please submit this document as a single .doc or .rtf format. If some documents are unable to be converted to .doc, then please provide all documents archived into a single .zip file.

**ALL PROPOSAL CHECK LIST**

|  |  |
| --- | --- |
| Completed CURRICULUM MODIFICATION FORM including: |  |
| * Brief description of proposal | X |
| * Rationale for proposal | X |
| * Date of department meeting approving the modification | X |
| * Chair’s Signature | X |
| * Dean’s Signature |  |
| Evidence of consultation with affected departments  List of the programs that use this course as required or elective, and courses that use this as a prerequisite. | X |
| Documentation of Advisory Commission views (if applicable). | N/A |
| Completed [Chancellor’s Report Form](http://openlab.citytech.cuny.edu/collegecouncil/files/2014/08/2013-10-09-Chancellor_Report_Quick_Reference_Guide1.doc). | X |

**EXISTING PROGRAM MODIFICATION PROPOSALS**

|  |  |
| --- | --- |
| Documentation indicating core curriculum requirements have been met for new programs/options or program changes. | X |
| Detailed rationale for each modification (this includes minor modifications) | N/A |

New York City College of Technology, CUNY

NEW COURSE PROPOSAL FORM

This form is used for all new course proposals. Attach this to the [Curriculum Modification Proposal Form](http://openlab.citytech.cuny.edu/collegecouncil/files/2014/08/2013-10-10-Curriculum_Modification_Proposal_Form.docx) and submit as one package as per instructions. Use one New Course Proposal Form for each new course.

|  |  |
| --- | --- |
| **Course Title** | Information and Data Management I |
| **Proposal Date** | Jan 4th, 2018 |
| **Proposer’s Name** | Dr. Elena Filatova |
| **Course Number** | CST 2312 |
| **Course Credits, Hours** | 3 credits, 2 lecture hours and 2 lab hours |
| **Course Pre / Co-Requisites** | Pre-req: CST 1101 (Problem Solving Using Computers) |
| **Catalog Course Description** | An introduction to informatics and intellectual tools covers topics related to the digital infrastructure, acquisition, organization, management and curation of data. The course is structured around Python tools for regular expression analysis, accessing data sources (crawling, Web APIs), and analysis of structured data. The final project includes an application of the technical topics discussed in class. |
| **Brief Rationale**  Provide a concise summary of why this course is important to the department, school or college. | Information and Data Management I is the first course in the two-course series. This two-course series provides a comprehensive introduction to programming with Python that goes beyond the basic concepts and focuses around data processing and analysis. In the modern world, data is the lifeblood of any organization and the driving force for development of new successful models for decision making. Competency in programming is an essential skill for successfully extracting information and knowledge form data. Also, this course is integral to the proposed Bachelors in Data Science curriculum program of the Computer Systems Technology department. |
| **CUNY – Course Equivalencies**  Provide information about equivalent courses within CUNY, if any. | N/A |
| **Intent to Submit as Common Core**  If this course is intended to fulfill one of the requirements in the common core, then indicate which area. | No. |
| **Intent to Submit as An Interdisciplinary Course** | No. |
| **Intent to Submit as a Writing Intensive Course** | No. |

Please include all appropriate documentation as indicated in the NEW COURSE PROPOSAL Combine all information into a single document that is included in the Curriculum Modification Form.

**NEW COURSE PROPOSAL CHECK LIST**

Use this checklist to ensure that all required documentation has been included. You may wish to use this checklist as a table of contents within the new course proposal.

|  |  |
| --- | --- |
| **Completed NEW COURSE PROPOSAL FORM** |  |
| * Title, Number, Credits, Hours, Catalog course description | X |
| * Brief Rationale | X |
| * CUNY – Course Equivalencies | X |
| Completed [Library Resources and Information Literacy Form](http://openlab.citytech.cuny.edu/collegecouncil/files/2014/08/curriculum_modification_library_form.doc) | X |
| **Course Outline**  Include within the outline the following. | X |
| Hours and Credits for Lecture and Labs  If hours exceed mandated Carnegie Hours, then rationale for this | X |
| Prerequisites/Co- requisites | X |
| Detailed Course Description | X |
| Course Specific Learning Outcome and Assessment Tables   * Discipline Specific * General Education Specific Learning Outcome and Assessment Tables | X |
| Example Weekly Course outline | X |
| Grade Policy and Procedure | X |
| Recommended Instructional Materials (Textbooks, lab supplies, etc) | X |
| Library resources and bibliography | X |
| **Course Need Assessment.**  Describe the need for this course. Include in your statement the following information. | X |
| Target Students who will take this course. Which programs or departments, and how many anticipated?  Documentation of student views (if applicable, e.g. non-required elective). | X |
| Projected headcounts (fall/spring and day/evening) for each new or modified course. | X |
| If additional physical resources are required (new space, modifications, equipment), description of these requirements. If applicable, Memo or email from the VP for Finance and Administration with written comments regarding additional and/or new facilities, renovations or construction. | X |
| Where does this course overlap with other courses, both within and outside of the department? | X |
| Does the Department currently have full time faculty qualified to teach this course? If not, then what plans are there to cover this? | X |
| If needs assessment states that this course is required by an accrediting body, then provide documentation indicating that need. | N/A |
| **Course Design**  Describe how this course is designed. | X |
| Course Context (e.g. required, elective, capstone) | X |
| Course Structure: how the course will be offered (e.g. lecture, seminar, tutorial, fieldtrip)? | X |
| Anticipated pedagogical strategies and instructional design (e.g. Group Work, Case Study, Team Project, Lecture) | X |
| How does this course support Programmatic Learning Outcomes? | X |
| Is this course designed to be partially or fully online? If so, describe how this benefits students and/or program. | X |
| **Additional Forms for Specific Course Categories** |  |
| [Interdisciplinary Form](http://openlab.citytech.cuny.edu/collegecouncil/files/2014/08/Application-for-Interdisciplinary-Course-Designation.docx) (if applicable) | N/A |
| Interdisciplinary Committee Recommendation (if applicable and if received)\*  \*Recommendation must be received before consideration by full Curriculum Committee | N/A |
| [Common Core (Liberal Arts) Intent to Submit](http://openlab.citytech.cuny.edu/collegecouncil/files/2014/08/CommonCoreCourseSubmissionForm_4.2.12.doc) (if applicable) | N/A |
| Writing Intensive Form if course is intended to be a WIC (under development) | N/A |
| If course originated as an experimental course, then results of evaluation plan as developed with director of assessment. | N/A |
| **(Additional materials for** [**Curricular Experiments**](http://www.300jaystreet.com/college-council/curriculum_proposals/curricular-experiments)**)** |  |
| Plan and process for evaluation developed in consultation with the director of assessment. (Contact Director of Assessment for more information). |  |
| Established Timeline for Curricular Experiment |  |

**Proposed Course Name: Information and Data Management I**

**Course Overview & Rationale**

Information and Data Management I is the first course in the two-course series. This two-course series provides a comprehensive introduction to programming with Python that goes beyond the basic concepts and focuses around data processing and analysis. In the modern world, data is the lifeblood of any organization and the driving force for development of new successful models for decision making. Competency in programming is an essential skill for successfully extracting information and knowledge form data. Also, this course is integral to the proposed Bachelors in Data Science curriculum program of the Computer Systems Technology department.

This course is for the students who are already familiar with the basics of programming in Python, and want to extend and apply this knowledge to a topic of their interest. This course introduces students to regular expressions and basic data structures that are essential for the data management process from data acquisition to analysis of data big data and small data (tuples, dictionaries). Furthermore, this course provides students with the introduction to the necessary informatics and intellectual tools to become efficient and effective information users. The course covers topics related to the digital infrastructure, acquisition, organization, management and curation of data. The course is structured around the Python tools for regular expression analysis, accessing data sources (crawling, Web APIs), analysis of structured data. At the end of the class the students complete a project to demonstrate the mastery of the technical topics discussed in class with an application to their domain of interest. The development of this course lays the foundation for the data acquisition and analysis step that is the first and necessary step in the decision making based on data pipeline. This course is designed to be a stepping stone for more advanced courses within the Data Science major, a new, fast-growing, and promising major in the Computer Systems Technology department at New York City College of Technology.

**Course Need**

**Students who would take this class:** students who intend to major in Data Science

**Department**: Computer Systems Technology

**Program**: Bachelors in Data Science

**The number of section (s) anticipated:** one section for the first year

**Projected headcount: 24** students

**Physical Resources required:** Basicsmartroomset**-**up**:** a screen, and an overhead projector/a TV set that is run by and connected to a computer

**Course overlap:** None

**Faculty** **qualified** **for** **teaching** **this** **course**: Yes, there are faculty members who have doctoral degrees in Computer Science with the concentration in Data Processing for various domains.

**Course design**

**Course context:** This course will be required of Data Science major students. Students are required to develop an independent project at the end of the semester.

**Course** **structure**: This course will be offered in a lecture style/format.

**Anticipated** **Pedagogical** **Strategies** **and** **Instructional** **Design**: This class will be run in a lecture-activity style/format. The class will start with a lecture, and then move on to creative in-class activities, such as using Python INotebook for Python-related exercises, collecting and analyzing datasets, using web APIs.

**Providing Support to Programmatic Learning Outcomes:** This course requires satisfactory completion of individual assignments, two major exams and a final term project.

**NEW YORK CITY COLLEGE OF TECHNOLOGY/CUNY**

**Computer Systems Technology Department**

**Course Outline**

**CST 2312 – Information and Data Management I**

**(2 class hours; 2 lab hours; 3 credits)**

**Course Description:**

This course provides students with the introduction to the necessary informatics and intellectual tools to become efficient and effective information users. The course covers topics related to the digital infrastructure, acquisition, organization, management and curation of data. The course is structured around the Python tools for regular expression analysis, accessing data sources (crawling, Web APIs), analysis of structured data. At the end of the class the students complete a project to demonstrate the mastery of the technical topics discussed in class with an application to their domain of interest.



**Course Objectives:**

Upon successful completion of this course, students should be able to:

* Demonstrate knowledge of regular expressions
* Demonstrate knowledge of web APIs
* Demonstrate the skills of processing information downloaded from Internet
* Demonstrate knowledge of web crawling

**Software**: Python and Python Libraries

**General Education Outcomes:**

* **SKILLS/Inquiry/Analysis:** Students will employ scientific reasoning and logical thinking.
* **SKILLS/Communication:** Students will communicate in diverse settings and groups, using written (both reading and writing), oral (both speaking and listening), and visual means
* **VALUES, ETHICS, RELATIONSHIPS / Professional/Personal Development:** Students will work with teams, including those of diverse composition. Build consensus. Respect and use creativity.

**Pre-requisites:**

CST 1101 Problem Solving with Computer Programming

**Required Text**:

* Python for Everybody: Exploring Information by Charles R. Severance

Free on-line version: http://www.pythonlearn.com/book.php#python-for-informatics

* Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython by W McKinney

**Additional Reading Materials:**

The instructor will identify several additional information resources during the semester including Internet resources, print material (handouts) and reference books.

**Tools and On-Line Documentation:**

* Python: https://www.python.org/
* Python Data Analysis Library: http://pandas.pydata.org/
* Python package for scientific computing: http://www.numpy.org/
* IPython command shell for interactive computing: https://ipython.org/
* Anaconda Data Science ecosystem: https://www.continuum.io/

**Grading Procedure:**

Midterm Exam 20%

Final Exam 30%

Lab Final Project 25%

Homework assignments 20%

In-class labs/quizzes 5%

=====

TOTAL 100%

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Letter Grade** | A | A- | B+ | B | B- | C+ | C | D | F |
| **Numerical Grade** | 93-100 | 90-92.9 | 87-89.9 | 83-86.9 | 80-82.9 | 77-79.9 | 70-76.9 | 60-69.9 | <=59.9 |

The grade distribution follows the information in the NYCCT Student Handbook (p.43).

**Academic Integrity**: During the course of the class you are required to follow the NYCCT Academic Integrity Standards described in the Student Handbook (pp.95 – 99)

**NYCCT Student Handbook** can be downloaded here: <http://www.citytech.cuny.edu/current-student/docs/StudentHandbook.pdf>.

**Course Outline:**

|  |  |  |
| --- | --- | --- |
| **Week** | **Topics** | **Assignments and Readings** |
| 1 | Review of Python Basics:  variables, basic data types (String, int, float, Boolean, None), lists, selection, loops, functions | <http://www.pythonlearn.com/html-270/book003.html> |
| 2 | Introduction to IPython and IPython NoteBooks | Lecture notes  Assignment 1 |
| 3 | Introduction to Python tuples, and dictionaries  Lab exercises: tuples, dictionaries | <http://www.pythonlearn.com/html-270/book011.html> |
| 4,5 | Introduction to regular expressions  Lab exercises: regular expressions | <http://www.pythonlearn.com/html-270/book012.html>  Assignment 2 |
| 6 | Working with files; Using Libraries  Lab exercises: importing libraries, using functions from imported libraries | <http://www.pythonlearn.com/html-270/book008.html>  Assignment 3 |
| 7 | **Midterm Exam:**  Python tuples, dictionaries, regular expressions | <http://www.pythonlearn.com/html-270/book010.html>  <http://www.pythonlearn.com/html-270/book012.html> |
| 8 | Introduction to Web APIs  Web APIs using Python  Lab exercises: using web APIs | Assignment 4 |
| 9 | Regular Expressions and Crawling |  |
| 10 | Reading CSV files, Selecting rows/columns | Assignment 5 |
| 11, 12 | Pandas and NumPy:  Introduction to data series, data frames  Lab exercises: numpy arrays, pandas dataframes |  |
| 13, 14 | Lab: Working on a project |  |
| 15 | **Review, project presentation, final exam** |  |

**ASSESSMENT CRITERIA:** For successful completion of this course the student should be able to:

|  |  |
| --- | --- |
| **For the successful completion of this course a student should be able to:** | **Evaluation methods and criteria** |
| Demonstrate knowledge of regular expressions | Students will create Python scripts (and run Python commands in the Shell mode) that use regular expressions to extracts and/or modify information in the source file / string. |
| Demonstrate knowledge of web APIs | Students will create Python scripts (and run Python commands in the Shell mode) that use web APIs to upload locally wen pages. |
| Demonstrate the skills of processing information downloaded from Internet | Students will create Python scripts (and run Python commands in the Shell mode) that read the source text of a downloaded web page and extract the desired information. |
| Demonstrate knowledge of  web crawling | Students will create Python scripts that can access the information form the web pages from the sites that do not have a custom web API. |
| Work effectively in a team | The final project is a team project. The students will group into teams and create a project that demonstrates their knowledge of the programming tools and techniques learned in class. Will use the Internet and other resources to complete the project. Also, there will be an oral presentation made to the class. It will include their learning experience in working in a group. |

**GENERAL EDUCATION LEARNING OUTCOMES/ASSESSMENT METHODS**



|  |  |
| --- | --- |
| **LEARNING OUTCOMES** | **ASSESSMENT METHODS** |
| **1.** Demonstrate the ability to work collaboratively and independently on assignments in and outside a classroom setting. | **1.** Classroom discussions, group assignments and individual oral presentations. |
| **2.** Understand and employ both quantitative and qualitative analysis to solve problems**.** | **2.** Classroom discussion, group activities, group presentations, quizzes, tests, final exam. |
| **3.** Develop reading, writing competencies, and listening skills. | **3.** Biweekly reading and writing assignments, individual and group presentation, classroom discussion. Each homework assignment requires writing. |
| **4.** Work with teams. Build consensus. Use creativity. | **4.** Group projects and presentations. |

**Bibliography:**

* Charles R. Severance (2013). *Python for Informatics: Exploring Information*. Publisher: CreateSpace Independent Publishing Platform; 1 edition,

ISBN-10: 1492339245

Free on-line version: <http://www.pythonlearn.com/book.php#python-for-informatics>

* Wes McKinney (2017) *Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython.*

Publisher: O'Reilly Media; 2 edition (October 20, 2017)

ISBN-10: 1491957662

* Armando Fandango (2017). *Python Data Analysi*s

Publisher: Packt Publishing - ebooks Account; 2nd Revised edition edition (March 27, 2017)

ISBN-10: 1787127486

* Jake VaderPlas (2016). *Python Data Science Handbook: Essential Tools for Working with Data.*

Publisher: O'Reilly Media; 1 edition (December 10, 2016)

ISBN-10: 1491912057

**CHANCELLOR’S REPORT FORM**

Department: Computer Systems Technology

Date: January 4th, 2018

**NEW COURSE PROPOSAL**

**Section AIV: New Course**

|  |  |
| --- | --- |
| **Department(s)** | **Computer Systems Technology** |
| **Academic Level** | **[X] Regular [ ] Compensatory [  ] Developmental [  ] Remedial** |
| **Subject Area** | Data Science |
| **Course Prefix** | **CST** |
| **Course Number** | **CST 2312** |
| **Course Title** | **Information and Data Management I** |
| **Catalog Description** | An introduction to informatics and intellectual tools covers topics related to the digital infrastructure, acquisition, organization, management and curation of data. The course is structured around Python tools for regular expression analysis, accessing data sources (crawling, Web APIs), and analysis of structured data. The final project includes an application of the technical topics discussed in class. |
| **Prerequisite** | **CST 1101** |
| **Corequisite** | **None** |
| **Credits** | **3** |
| **Contact Hours** | **2 Class Hours, 2 Lab Hours** |
| **Liberal Arts** | **[ ] Yes [ X  ] No** |
| **Course Attribute (e.g. Writing Intensive, etc)** | **It is not a writing intensive course** |
| **Course Applicability** | **[X] Major**  **[ ] Gen Ed Required [ ] Gen Ed - Flexible [ ] Gen Ed - College Option**  **[ ] English Composition [ ] World Cultures [ ] Speech**  **[ ] Mathematics [ ] US Experience in its Diversity**  **[ ] Interdisciplinary [ ] Science [ ] Creative Expression**  **[ ] Advanced Liberal Arts**  **[ ] Individual and Society**  **[ ] Scientific World** |
| **Effective Term** | **Fall 2019** |

**Rationale:** Information and Data Management I is the first course in the two-course series. This two-course series provides a comprehensive introduction to programming with Python that goes beyond the basic concepts and focuses around data processing and analysis. In the modern world, data is the lifeblood of any organization and the driving force for development of new successful models for decision making. Competency in programming is an essential skill for successfully extracting information and knowledge form data. Also, this course is integral to the proposed Bachelors in Data Science curriculum program of the Computer Systems Technology department.

**LIBRARY RESOURCES & INFORMATION LITERACY: MAJOR CURRICULUM MODIFICATION**

Please complete for **all** major curriculum modifications. This information will assist the library in planning for new courses/programs.

Consult with your library faculty subject specialist (<http://cityte.ch/dir>) **3 weeks before the proposal deadline**.

**Course proposer:** please complete boxes 1-4. **Library faculty subject specialist:** please complete box 5.

|  |  |  |
| --- | --- | --- |
| **1** | **Title of proposal**  CST 2312: Information and Data Management I | **Department/Program**  Computer and Systems Technology |
|  | **Proposed by** (include email & phone)  Elena Filatova  efilatova@citytech.cuny.edu | **Expected date course(s) will be offered**  Fall 2019  **# of students** 24 |

|  |  |
| --- | --- |
| **2** | **The library cannot purchase reserve textbooks for every course at the college, nor copies for all students. Consult our website (**[**http://cityte.ch/curriculum**](http://cityte.ch/curriculum)**) for articles and ebooks for your courses, or our open educational resources (OER) guide (**[**http://cityte.ch/oer**](http://cityte.ch/oer)**). Have you considered using a freely-available OER or an open textbook in this course?**  One of the books is a free on-line book, the other book is not free:  Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython by Wes McKinney, 2nd ed.  ISBN-10: 1449319793 |

|  |  |
| --- | --- |
| **3** | **Beyond the required course materials, are City Tech library resources sufficient for course assignments? If additional resources are needed, please provide format details (e.g. ebook, journal, DVD, etc.), full citation (author, title, publisher, edition, date), price, and product link.**  Tools and On-Line Documentation (free, on-line)  Python: https://www.python.org/  Python Data Analysis Library: http://pandas.pydata.org/  Python package for scientific computing: http://www.numpy.org/  IPython command shell for interactive computing: https://ipython.org/  Anaconda Data Science ecosystem: https://www.continuum.io/ |

|  |  |
| --- | --- |
| **4** | **Library faculty focus on strengthening students' information literacy skills in finding, critically evaluating, and ethically using information. We collaborate on developing assignments and customized instruction and research guides. When this course is offered, how do you plan to consult with the library faculty subject specialist for your area? Please elaborate.**  The project for the course concerns an on-line data set. There will be a discussion in class, what data can be downloaded and used for the project, whether there are any ethical consequences for using this data set and making conclusions about this data set. |

|  |  |
| --- | --- |
| **5** | **Library Faculty Subject Specialist Prof. Junior Tidal**  **Comments and Recommendations**  I believe that the library’s collection can adequately support this course.  **Date 12.13.17** |

# New Course Proposal #2: CST2402 Introduction to Data Science

New York City College of Technology, CUNY

CURRICULUM MODIFICATION PROPOSAL FORM

This form is used for all curriculum modification proposals. See the [Proposal Classification Chart](http://openlab.citytech.cuny.edu/collegecouncil/files/2014/08/2013-10-09-Proposal_Classification_Chart.pdf) for information about what types of modifications are major or minor. Completed proposals should be emailed to the Curriculum Committee chair.

|  |  |
| --- | --- |
| **Title of Proposal** | CST2402 Introduction to Data Science |
| **Date** | December 5, 2017 |
| **Major or Minor** | Major |
| **Proposer’s Name** | Dr. Elizabeth Milonas |
| **Department** | Computer Systems Technology |
| **Date of Departmental Meeting in which proposal was approved** | December 8, 2017 |
| **Department Chair Name** | **Dr. Hong Li** |
| **Department Chair Signature and Date** | **1/16/2018** |
| **Academic Dean Name** | Dean Hom |
| **Academic Dean Signature and Date** | 2/22/18 |
| **Brief Description of Proposal**  (Describe the modifications contained within this proposal in a succinct summary. More detailed content will be provided in the proposal body. | Title: Introduction to Data Science  This proposal explains the rationale for developing the Introduction to Data Science as a required course in the Data Science curriculum of the Computer Systems Technology Department. A course outline and sample syllabus are attached to the proposal. |
| **Brief Rationale for Proposal**  (Provide a concise summary of why this proposed change is important to the department. More detailed content will be provided in the proposal body). | The CST2402 Introduction to Data Science course introduces students to the fundamental concepts, principles and ethical issues of the Data Science domain and provides students with hands-on experience in the use of Data Science tools for business problem solving and decision making. Concepts discussed in this course are the underpinnings of topics covered in detail in upper level courses in the BS in Data Science. |
| **Proposal History**  (Please provide history of this proposal: is this a resubmission? An updated version? This may most easily be expressed as a list). | This is a first submission |
| **CUNY – Course Equivalencies**  Provide information about equivalent courses within CUNY, if any. | None |
| **Intent to Submit as Common Core**  If this course is intended to fulfill one of the requirements in the common core, then indicate which area. | N/A |

Please include all appropriate documentation as indicated in the NEW COURSE PROPOSAL Combine all information into a single document that is included in the Curriculum Modification Form.

Please submit this document as a single .doc or .rtf format. If some documents are unable to be converted to .doc then please provide all documents archived into a single .zip file.

**ALL PROPOSAL CHECK LIST**

|  |  |
| --- | --- |
| Completed CURRICULUM MODIFICATION FORM including: |  |
| * Brief description of proposal | X |
| * Rationale for proposal | X |
| * Date of department meeting approving the modification | X |
| * Chair’s Signature | X |
| * Dean’s Signature |  |
| Evidence of consultation with affected departments  List of the programs that use this course as required or elective, and courses that use this as a prerequisite. | X |
| Documentation of Advisory Commission views (if applicable). | N/A |
| Completed [Chancellor’s Report Form](http://openlab.citytech.cuny.edu/collegecouncil/files/2014/08/2013-10-09-Chancellor_Report_Quick_Reference_Guide1.doc). | X |

**EXISTING PROGRAM MODIFICATION PROPOSALS**

|  |  |
| --- | --- |
| Documentation indicating core curriculum requirements have been met for new programs/options or program changes. | X |
| Detailed rationale for each modification (this includes minor modifications) | N/A |

New York City College of Technology, CUNY

NEW COURSE PROPOSAL FORM

This form is used for all new course proposals. Attach this to the [Curriculum Modification Proposal Form](http://openlab.citytech.cuny.edu/collegecouncil/files/2014/08/2013-10-10-Curriculum_Modification_Proposal_Form.docx) and submit as one package as per instructions. Use one New Course Proposal Form for each new course.

|  |  |
| --- | --- |
| **Course Title** | Introduction to Data Science |
| **Proposal Date** | December 5, 2017 |
| **Proposer’s Name** | Dr. Elizabeth Milonas |
| **Course Number** | CST2402 |
| **Course Credits, Hours** | 3 credits, 2 lecture hours and 2 lab hours |
| **Course Pre / Co-Requisites** | Pre-reqs:CST1204 Database Fundamentals,  CST1201 Programming Fundamentals,  Pre or Co Req: MAT1475 |
| **Catalog Course Description** | This is a fundamental course in the basic concepts, principles and ethical issues of the Data Science domain including: definition, framework, techniques, issues, business uses and ethics. Topics include data collection, processing and management, exploratory data analysis, data visualization and presentation and ethical issues. Projects involve the use of data science techniques and tools to solve business problems and improve business decision making. |
| **Brief Rationale**  Provide a concise summary of why this course is important to the department, school or college. | The CST2402 Introduction to Data Science course introduces students to the fundamental concepts, principles and ethical issues of the Data Science domain and provides students with hands-on experience in the use of Data Science tools for business problem solving and decision making. Concepts discussed in this course are the underpinnings of topics covered in detail in upper level courses in the B.S. in Data Science. |
| **CUNY – Course Equivalencies**  Provide information about equivalent courses within CUNY, if any. | N/A |
| **Intent to Submit as Common Core**  If this course is intended to fulfill one of the requirements in the common core, then indicate which area. | No |
| **For Interdisciplinary Courses:**   * Date submitted to ID Committee for review * Date ID recommendation received   - Will all sections be offered as ID? Y/N | No |
|  |
|  |
| **Intent to Submit as a Writing Intensive Course** | No |

Please include all appropriate documentation as indicated in the NEW COURSE PROPOSAL Combine all information into a single document that is included in the Curriculum Modification Form.

**NEW COURSE PROPOSAL CHECK LIST**

Use this checklist to ensure that all required documentation has been included. You may wish to use this checklist as a table of contents within the new course proposal.

|  |  |
| --- | --- |
| **Completed NEW COURSE PROPOSAL FORM** |  |
| * Title, Number, Credits, Hours, Catalog course description | X |
| * Brief Rationale | X |
| * CUNY – Course Equivalencies | X |
| Completed [Library Resources and Information Literacy Form](http://openlab.citytech.cuny.edu/collegecouncil/files/2014/08/curriculum_modification_library_form.doc) | X |
| **Course Outline**  Include within the outline the following. | **X** |
| Hours and Credits for Lecture and Labs  If hours exceed mandated Carnegie Hours, then rationale for this | X |
| Prerequisites/Co- requisites | X |
| Detailed Course Description | X |
| Course Specific Learning Outcome and Assessment Tables   * Discipline Specific * General Education Specific Learning Outcome and Assessment Tables | X |
| Example Weekly Course outline | X |
| Grade Policy and Procedure | X |
| Recommended Instructional Materials (Textbooks, lab supplies, etc) | X |
| Library resources and bibliography | X |
| **Course Need Assessment.**  Describe the need for this course. Include in your statement the following information. | X |
| Target Students who will take this course. Which programs or departments, and how many anticipated?  Documentation of student views (if applicable, e.g. non-required elective). | X |
| Projected headcounts (fall/spring and day/evening) for each new or modified course. | X |
| If additional physical resources are required (new space, modifications, equipment), description of these requirements. If applicable, Memo or email from the VP for Finance and Administration with written comments regarding additional and/or new facilities, renovations or construction. | X |
| Where does this course overlap with other courses, both within and outside of the department? | X |
| Does the Department currently have full time faculty qualified to teach this course? If not, then what plans are there to cover this? | X |
| If needs assessment states that this course is required by an accrediting body, then provide documentation indicating that need. | N/A |
| **Course Design**  Describe how this course is designed. | X |
| Course Context (e.g. required, elective, capstone) | X |
| Course Structure: how the course will be offered (e.g. lecture, seminar, tutorial, fieldtrip)? | X |
| Anticipated pedagogical strategies and instructional design (e.g. Group Work, Case Study, Team Project, Lecture) | X |
| How does this course support Programmatic Learning Outcomes? | X |
| Is this course designed to be partially or fully online? If so, describe how this benefits students and/or program. | X |
| **Additional Forms for Specific Course Categories** | N/A |
| [Interdisciplinary Form](http://openlab.citytech.cuny.edu/collegecouncil/files/2014/08/Application-for-Interdisciplinary-Course-Designation.docx) (if applicable) | N/A |
| Interdisciplinary Committee Recommendation (if applicable and if received)\*  \*Recommendation must be received before consideration by full Curriculum Committee | N/A |
| [Common Core (Liberal Arts) Intent to Submit](http://openlab.citytech.cuny.edu/collegecouncil/files/2014/08/CommonCoreCourseSubmissionForm_4.2.12.doc) (if applicable) | N/A |
| Writing Intensive Form if course is intended to be a WIC (under development) | N/A |
| If course originated as an experimental course, then results of evaluation plan as developed with director of assessment. | N/A |
| **(Additional materials for** [**Curricular Experiments**](http://www.300jaystreet.com/college-council/curriculum_proposals/curricular-experiments)**)** |  |
| Plan and process for evaluation developed in consultation with the director of assessment. (Contact Director of Assessment for more information). |  |
| Established Timeline for Curricular Experiment |  |

**Proposal Course Name: Introduction to Data Science**

**Course Overview and Rationale**

In today’s society, data is being generated at an overwhelmingly rapid pace creating an unprecedented global challenge to analyze and understand this data and to use it to solve complex data-driven problems. This challenge has created an equally rapidly growing need for highly trained and skilled IT professionals who can take on the challenge by combining analytical, technical, mathematical, and statistical knowledge to solve data-driven problems. The Data Science program prepares students to meet the data challenge. Through lecture, hand-on, group and individual work as well as internship, the student in the Data Science major will be well versed in the tools and techniques required for meeting the big data challenges in many industries including finance, health, business, marketing and many more.

The Introduction to Data Science is one of the many courses in the Data Science program that prepares students to meet the data challenge. This fundamental course introduces students to the field of Data Science. Topics integral to the Data Science domain such as data mining, natural language processing, machine learning, visualization and ethics are introduced through lecture and real-life case studies. Students experience first-hand, the tools and techniques used in the Data Science field. Students also identify and examine the ethical issues that surround the field of Data Science. All CST students, transfer and non-transfer, will be required to complete this course when they have successfully completed the prerequisite courses: CST1204 Database Fundamentals and CST1201 Programming Fundamentals with a grade of C or better. This course does not overlap with other courses both within and outside of the department. The CST department currently has full time faculty qualified to teach this course.

**Course Need**

**Students who would take this class**: intend to major in Data Science

**Department**: Computer Systems Technology

**Program**: Bachelor of Science in Data Science

**The number of sections (s) anticipated**: one section for the first year

**Projected headcount**: 24 students

**Physical resources required**: Basic smart room set-up: a screen, and an overhead projector that is run by and connected to a computer

**Course overlap**: None

**Faculty qualified for teaching this course**: There are faculty members who have doctoral degrees in big data and related big data topics.

**Course Design**

**Course context**: This course is required for the Data Science major.

**Course structure**: This course is offered in a lecture format.

**Anticipated Pedagogical Strategies and Instructional Design**: This course is offered using a combination of lecture, hands-on, in class assignments, home assignments and group work.

**Providing Support to Programmatic Learning Outcomes**: This course requires satisfactory completion of in class assignment, home assignments, group project, three tests, final exam and online discussions.

**NEW YORK CITY COLLEGE OF TECHNOLOGY**

**COMPUTER SYSTEMS TECHNOLOGY DEPARTMENT**

**CST2402 – Introduction to Data Science**

**(2 lecture hours, 2 lab hours, 3 credits)**

**COURSE DESCRIPTION:** This is a fundamental course in the basic concepts, principles and ethical issues of the Data Science domain including: definition, framework, techniques, issues, business uses and ethics. Topics include data collection, processing and management, exploratory data analysis, data visualization and presentation and ethical issues. Projects involve the use of data science techniques and tools to solve business problems and improve business decision making.

**LEARNING OUTCOMES:** Upon successful completion of this course, the students acquire the following knowledge and skills:

1. Define and analyze the issues and challenges related to the data science domain including issues related to machine learning, data mining and data visualization.
2. Identify, analyze and use data science strategies and tools to solve business problems and improve business decision making.
3. Identify and examine ethical issues related to the data science domain, particularly issues related to privacy, data sharing and algorithmic decision-making.

**GENERAL EDUCATION OUTCOMES:**

* **SKILLS/Inquiry/Analysis:** Students use scientific reasoning and logical thinking.
* **SKILLS/Communication:** Students use written (both reading and writing), oral (both speaking and listening), and visual means to communicate.
* **VALUES, ETHICS, RELATIONSHIPS / Professional/Personal Development:** Students work in diverse teams utilizing key traits including respect, cooperation and creativity.

**PREREQUISITES**

The successful completion of the following courses with a grade of C or higher:

CST1201, CST1204

Pre or Co requisites: MAT1475

**REQUIRED TEXTBOOK**

Cielen, Davy, Arno D.B. Meysman, and Mohamed Ali. *Introducing Data Science, Big Data, Machine Learning and More, Using Python Tools.* Shelter Island: Manning, 2016.

ISBN-13: 9781633430037

**ADDITIONAL MATERIAL**

The instructor will identify several additional information resources during the semester including Internet resources, print material (handouts), journal articles and reference books.

**Academic Integrity –** Students and all others who work with information, ideas, texts, images, music, inventions, and other intellectual property owe their audience and sources accuracy and honesty in using, crediting, and citing sources. As a community of intellectual and professional workers, the College recognizes its responsibility for providing instruction in information literacy and academic integrity, offering models of good practice, and responding vigilantly and appropriately to infractions of academic integrity. Accordingly, academic dishonesty is prohibited in The City University of New York and at New York City College of Technology and is punishable by penalties, including failing grades, suspension, and expulsion. The complete text of the College policy on Academic Integrity may be found in the catalog.

**GRADING PROCEDURE:**

Tests (3) 30% Final Exam 15%

Labs 15%

Assignments 35%

Participation 5%

=====

TOTAL 100%

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Letter Grade** | A | A- | B+ | B | B- | C+ | C | D | F |
| **Numerical Grade** | 93-100 | 90-92.9 | 87-89.9 | 83-86.9 | 80-82.9 | 77-79.9 | 70-76.9 | 60-69.9 | <=59.9 |

**COURSE OUTLINE:**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Week** | **Topics** | **Reading** | **Labs/ Assignments** | **Tests** |
| 1 | Introduction to Data Science   * Benefits and uses of data science and Big Data * Types of Data * Data gathering, exploration, modeling and presentation * Ethical issues | Ch. 1 | Lab 1: Data Science in business |  |
| 2 | Data Science Process   * Define goals and create plan * Retrieve data; clean, integrate and transform data * Exploratory data analysis * Build models and present findings | Ch. 2 | Lab 2: The DS Process |  |
| 3 | Big Data   * Challenges of Big Data * Algorithms and data structure * Selecting tools * Case Study | Ch. 4 | Lab 3: Big Data |  |
| 4 | Big Data Techniques and Tools   * Distributing data storage and processing with frameworks (Hadoop, HDFS, and Map-Reduce) * Case Study | Ch. 5 | Lab 4: Tools and Techniques  Assignment 1 | **Test 1** |
| 5 | NoSQL   * Introduction to NoSQL * Case Study | Ch. 6 | Lab 5: NoSQL categories |  |
| 6 | Data on the Web and Linked Data   * Introduction to connected data and graph database * Introduction to RDF/RDFS and OWL * Case Study | Ch. 7 | Lab 6: Linked Data  Assignment 2 |  |
| 7 | Data Mining   * Introduction to Data Mining * Case study | Ch. 8 | Lab 7: Practical uses of data mining |  |
| 8 | Data Mining Techniques and Tools   * Bag of words * Stemming and lemmatization * Decision tree classifier | Ch. 8 | Lab 8: Data Mining  Assignment 3 | **Test 2** |
| 9 | Machine Learning   * Introduction to machine learning * The modeling process | Ch. 3 | Lab 9: Alexa and machine learning |  |
| 10 | Machine Learning (continued)   * Supervised and unsupervised learning * Semi-supervised learning | Ch. 3 | Lab 10: Machine Learning  Assignment 4 |  |
| 11  &  12 | Data Visualization, Techniques and Tools   * Introduction to Data Visualization * Introduction to Matplotlib * Plots, histograms, bar charts, scatterplots using Matplotlib | Ch. 9 | Lab 11&12: Decision making with visualization | **Test 3** |
| 13 | Ethics of Data Science   * Data and discrimination * Algorithms and discrimination * Data collection, storage, analysis and application ethics * Preserving Privacy | Handouts/  Internet resources/  journal articles | Lab 13: The Good, the bad and the big data |  |
| 14 | Advanced Topics in Data Science   * Social Networks * Data Journalism | Handouts/  Internet resources/  journal articles | Lab 14: Social Networks and Data Journalism |  |
| 15 | Review |  |  | Final Exam |

**COURSE INTENDED LEARNING OUTCOMES/ASSESSMENT METHOD**

|  |  |
| --- | --- |
| **LEARNING OUTCOMES** | **ASSESSMENT METHOD** |
| 1. Demonstrate an understanding of the technical issues and challenges related to the Data Science domain. | 1. Group activities, written assignment, in class and interactive online discussions |
| 1. Identify, examine and use the Data Science strategies and tools (such as Hadoop, MapReduce, NoSQL, RDF, Matplotlib) to retrieve, manage and explore data. | 1. Classroom exercises and programming assignments where students use non-relational databases such as MongoDB and JSON as well as code RDF files and SPARQL |
| 1. Identify the ethical concerns and examine the challenges and issues related to privacy, data sharing and algorithmic decision-making | 1. Group activities, Written assignment, in class and interactive online discussion |

**GENERAL EDUCATION LEARNING OUTCOMES/ASSESSMENT METHODS**

|  |  |
| --- | --- |
| **LEARNING OUTCOMES** | **ASSESSMENT METHODS** |
| **1.** Demonstrate the ability to work collaboratively and independently on assignments in and outside a classroom setting. | **1.** Classroom discussions, group assignments and individual oral presentations. |
| **2.** Understand and employ both quantitative and qualitative analysis to solve problems**.** | **2.** Classroom discussion, group activities, group presentations, quizzes, tests, final exam. |
| **3.** Develop reading, writing competencies, and listening skills. | **3.** Biweekly reading and writing assignments, individual and group presentation, classroom discussion. Each homework assignment requires writing. |
| **4.** Work with teams. Build consensus. Use creativity. | **4.** Group projects and presentations. |

**Bibliography**

Cady, Field. *The Data Science Handbook.* Hoboken: John Wiley & Sons Inc., 2017. ISBN-13: 978-1119092940

Grus, Joel. *Data Science from Scratch, First Principles with Python.* Sebastopol: O’Reilly Media, Inc., 2015. ISBN-13:978-1491901427

O’Neil, Cathy and Rachel Schutt. *Doing Data Science, Straight Talk from the Frontline.* Sebastopol: O’Reilly Media, Inc., 2014. ISBN-13: 978-1449358655

Provost, Foster and Tom Fawcett. *Data Science for Business.* Sebastopol: O’Reilly Media, Inc., 2013. ISBN-13: 978-1449361327

Stanton, Jeffrey M. and Jeffrey M. Saltz. *An Introduction to Data Science.*  New York: Sage Publishing Inc., 2017. ISBN-13: 978-1506377537

**CHANCELLOR’S REPORT FORM**

Department: Computer Systems Technology

Date: January 1st, 2018

**NEW COURSE PROPOSAL**

**Section AIV: New Course**

|  |  |
| --- | --- |
| **Department(s)** | **Computer Systems Technology** |
| **Academic Level** | **[X] Regular [  ] Compensatory [  ] Developmental [  ] Remedial** |
| **Subject Area** | Data Science |
| **Course Prefix** | **CST** |
| **Course Number** | **2402** |
| **Course Title** | **Introduction to Data Science** |
| **Catalog Description** | This is a fundamental course in the basic concepts, principles and ethical issues of the Data Science domain including: definition, framework, techniques, issues, business uses and ethics. Topics include data collection, processing and management, exploratory data analysis, data visualization and presentation and ethical issues. Projects involve the use of data science techniques and tools to solve business problems and improve business decision making. |
| **Prerequisite** | **CST 1204, CST1201** |
| **Pre-or Corequisite** | **MAT1475** |
| **Credits** | **3** |
| **Contact Hours** | **2 Lecture Hours, 2 Lab Hours** |
| **Liberal Arts** | **[ ] Yes [ X  ] No** |
| **Course Attribute (e.g. Writing Intensive, etc)** | **It is not a writing intensive course** |
| **Course Applicability** | **[X] Major**  **[ ] Gen Ed Required [ ] Gen Ed - Flexible [ ] Gen Ed - College Option**  **[ ] English Composition [ ] World Cultures [ ] Speech**  **[ ] Mathematics [ ] US Experience in its Diversity**  **[ ] Interdisciplinary [ ] Science [ ] Creative Expression**  **[ ] Advanced Liberal Arts**  **[ ] Individual and Society**  **[ ] Scientific World** |
| **Effective Term** | **Fall 2019** |

**LIBRARY RESOURCES & INFORMATION LITERACY: MAJOR CURRICULUM MODIFICATION**

Please complete for **all** major curriculum modifications. This information will assist the library in planning for new courses/programs.

Consult with your library faculty subject specialist (<http://cityte.ch/dir>) **3 weeks before the proposal deadline**.

**Course proposer:** please complete boxes 1-4. **Library faculty subject specialist:** please complete box 5.

|  |  |  |
| --- | --- | --- |
| **1** | **Title of proposal**  CST2402 - Introduction to Data Science | **Department/Program**  Computer Systems Technology (CST) Department |
|  | **Proposed by** (include email & phone)  Elizabeth Milonas (718)260-5170  emilonas@citytech.cuny.edu | **Expected date course(s) will be offered**  Fall 2019  **# of students 24** |

|  |  |
| --- | --- |
| **2** | **The library cannot purchase reserve textbooks for every course at the college, nor copies for all students. Consult our website (**[**http://cityte.ch/curriculum**](http://cityte.ch/curriculum)**) for articles and ebooks for your courses, or our open educational resources (OER) guide (**[**http://cityte.ch/oer**](http://cityte.ch/oer)**). Have you considered using a freely-available OER or an open textbook in this course?**  Yes I have however, none are suitable. |



|  |  |
| --- | --- |
| **3** | **Beyond the required course materials, are City Tech library resources sufficient for course assignments? If additional resources are needed, please provide format details (e.g. ebook, journal, DVD, etc.), full citation (author, title, publisher, edition, date), price, and product link.**  Yes the current library resources are sufficient for this course. |

|  |  |
| --- | --- |
| **4** | **Library faculty focus on strengthening students' information literacy skills in finding, critically evaluating, and ethically using information. We collaborate on developing assignments and customized instruction and research guides. When this course is offered, how do you plan to consult with the library faculty subject specialist for your area? Please elaborate.**  I will reach out to the library subject specialist via email to arrange an information session in which the library subject specialist can present to the students of this course, the use of library databases, citation convention and discuss copyright issues. |

|  |  |
| --- | --- |
| **5** | **Library Faculty Subject Specialist Prof. Junior Tidal**  **Comments and Recommendations**  After surveying the library’s collection, I believe that the library can adequately support this course. Upon course approval, besides the recommended texts, the library should add additional materials pertaining to NoSQL to augment the current collection.  **Date 12.13.17** |

# New Course Proposal #3: CST3502 Data Mining

New York City College of Technology, CUNY

CURRICULUM MODIFICATION PROPOSAL FORM

This form is used for all curriculum modification proposals. See the [Proposal Classification Chart](http://openlab.citytech.cuny.edu/collegecouncil/files/2014/08/2013-10-09-Proposal_Classification_Chart.pdf) for information about what types of modifications are major or minor. Completed proposals should be emailed to the Curriculum Committee chair.

|  |  |
| --- | --- |
| **Title of Proposal** | **CST 3502 Data Mining** |
| **Date** | **Dec 5th, 2017** |
| **Major or Minor** | **Major** |
| **Proposer’s Name** | **Dr. Ashwin Satyanarayana** |
| **Department** | **Computer Systems Technology** |
| **Date of Departmental Meeting in which proposal was approved** | **12/8/2017** |
| **Department Chair Name** | **Dr. Hong Li** |
| **Department Chair Signature and Date** | **1/16/2018** |
| **Academic Dean Name** | **Dean Hom** |
| **Academic Dean Signature and Date** | 2/22/18 |
| **Brief Description of Proposal**  (Describe the modifications contained within this proposal in a succinct summary. More detailed content will be provided in the proposal body. | Title: Data Mining  This proposal explains the rationale for developing Data Mining as a required course in the Data Science curriculum of the Computer Systems Technology Department. A course outline and sample syllabus are attached to the proposal. |
| **Brief Rationale for Proposal**  (Provide a concise summary of why this proposed change is important to the department. More detailed content will be provided in the proposal body). | Data Mining or knowledge discovery, is the computer-assisted process of digging through and analyzing enormous sets of data and then extracting the meaning of the data. Data mining models are at the heart of successful information and product search, automated merchandising, smart personalization, dynamic pricing, social network analysis, genetics, proteomics, and many other technology-based solutions to important problems in business. The ethical issues that arise from the process of mining customer data and thereby compromising the privacy of those customers will be discussed. |
| **Proposal History**  (Please provide history of this proposal: is this a resubmission? An updated version? This may most easily be expressed as a list). | **This is a first submission.** |
| **CUNY – Course Equivalencies**  Provide information about equivalent courses within CUNY, if any. | None |
| **Intent to Submit as Common Core**  If this course is intended to fulfill one of the requirements in the common core, then indicate which area. | N/A |

Please include all appropriate documentation as indicated in the Curriculum Modification Checklist.

For each new course, please also complete the New Course Proposal and submit in this document.

Please submit this document as a single .doc or .rtf format. If some documents are unable to be converted to .doc, then please provide all documents archived into a single .zip file.

**ALL PROPOSAL CHECK LIST**

|  |  |
| --- | --- |
| Completed CURRICULUM MODIFICATION FORM including: |  |
| * Brief description of proposal | X |
| * Rationale for proposal | X |
| * Date of department meeting approving the modification | X |
| * Chair’s Signature |  |
| * Dean’s Signature |  |
| Evidence of consultation with affected departments  List of the programs that use this course as required or elective, and courses that use this as a prerequisite. | X |
| Documentation of Advisory Commission views (if applicable). | N/A |
| Completed [Chancellor’s Report Form](http://openlab.citytech.cuny.edu/collegecouncil/files/2014/08/2013-10-09-Chancellor_Report_Quick_Reference_Guide1.doc). | X |

**EXISTING PROGRAM MODIFICATION PROPOSALS**

|  |  |
| --- | --- |
| Documentation indicating core curriculum requirements have been met for new programs/options or program changes. | X |
| Detailed rationale for each modification (this includes minor modifications) | N/A |

New York City College of Technology, CUNY

NEW COURSE PROPOSAL FORM

This form is used for all new course proposals. Attach this to the [Curriculum Modification Proposal Form](http://openlab.citytech.cuny.edu/collegecouncil/files/2014/08/2013-10-10-Curriculum_Modification_Proposal_Form.docx) and submit as one package as per instructions. Use one New Course Proposal Form for each new course.

|  |  |
| --- | --- |
| **Course Title** | Data Mining |
| **Proposal Date** | Dec 5th, 2017 |
| **Proposer’s Name** | Ashwin Satyanarayana |
| **Course Number** | CST 3502 |
| **Course Credits, Hours** | 3 credits, ( 2 lecture and 2 lab hours) |
| **Course Pre / Co-Requisites** | Pre-reqs: CST 2312 (Information and Data Management I), CST 2402 (Introduction to Data Science) |
| **Catalog Course Description** | Theory and practice of data mining methods for extracting knowledge from data using state-of-the-art software (WEKA) for solving scientific and business problems will be covered. Topics discussed in the class include: data preprocessing, classification and prediction, association mining, cluster analysis and ethical data mining. Projects either involve research in data mining or address a practical problem using the methods and tools introduced in the class. |
| **Brief Rationale**  Provide a concise summary of why this course is important to the department, school or college. | Data Mining or knowledge discovery, is the computer-assisted process of digging through and analyzing enormous sets of data and then extracting the meaning of the data. Data mining models are at the heart of successful information and product search, automated merchandising, smart personalization, dynamic pricing, social network analysis, genetics, proteomics, and many other technology-based solutions to important problems in businesses. The ethical issues that arise from the process of mining customer data and thereby compromising the privacy of those customers will be discussed. |
| **CUNY – Course Equivalencies**  Provide information about equivalent courses within CUNY, if any. | N/A |
| **Intent to Submit as Common Core**  If this course is intended to fulfill one of the requirements in the common core, then indicate which area. | No. |
| **Intent to Submit as An Interdisciplinary Course** | No. |
| **Intent to Submit as a Writing Intensive Course** | No. |

**NEW COURSE PROPOSAL CHECK LIST**

Use this checklist to ensure that all required documentation has been included. You may wish to use this checklist as a table of contents within the new course proposal.

|  |  |
| --- | --- |
| **Completed NEW COURSE PROPOSAL FORM** |  |
| * Title, Number, Credits, Hours, Catalog course description | X |
| * Brief Rationale | X |
| * CUNY – Course Equivalencies | X |
| Completed [Library Resources and Information Literacy Form](http://openlab.citytech.cuny.edu/collegecouncil/files/2014/08/curriculum_modification_library_form.doc) | X |
| **Course Outline**  Include within the outline the following. | X |
| Hours and Credits for Lecture and Labs  If hours exceed mandated Carnegie Hours, then rationale for this | X |
| Prerequisites/Co- requisites | X |
| Detailed Course Description | X |
| Course Specific Learning Outcome and Assessment Tables   * Discipline Specific * General Education Specific Learning Outcome and Assessment Tables | X |
| Example Weekly Course outline | X |
| Grade Policy and Procedure | X |
| Recommended Instructional Materials (Textbooks, lab supplies, etc) | X |
| Library resources and bibliography | X |
| **Course Need Assessment.**  Describe the need for this course. Include in your statement the following information. | X |
| Target Students who will take this course. Which programs or departments, and how many anticipated?  Documentation of student views (if applicable, e.g. non-required elective). | X |
| Projected headcounts (fall/spring and day/evening) for each new or modified course. | X |
| If additional physical resources are required (new space, modifications, equipment), description of these requirements. If applicable, Memo or email from the VP for Finance and Administration with written comments regarding additional and/or new facilities, renovations or construction. | X |
| Where does this course overlap with other courses, both within and outside of the department? | X |
| Does the Department currently have full time faculty qualified to teach this course? If not, then what plans are there to cover this? | X |
| If needs assessment states that this course is required by an accrediting body, then provide documentation indicating that need. | N/A |
| **Course Design**  Describe how this course is designed. | X |
| Course Context (e.g. required, elective, capstone) | X |
| Course Structure: how the course will be offered (e.g. lecture, seminar, tutorial, fieldtrip)? | X |
| Anticipated pedagogical strategies and instructional design (e.g. Group Work, Case Study, Team Project, Lecture) | X |
| How does this course support Programmatic Learning Outcomes? | X |
| Is this course designed to be partially or fully online? If so, describe how this benefits students and/or program. | X |
| **Additional Forms for Specific Course Categories** |  |
| [Interdisciplinary Form](http://openlab.citytech.cuny.edu/collegecouncil/files/2014/08/Application-for-Interdisciplinary-Course-Designation.docx) (if applicable) | N/A |
| Interdisciplinary Committee Recommendation (if applicable and if received)\*  \*Recommendation must be received before consideration by full Curriculum Committee | N/A |
| [Common Core (Liberal Arts) Intent to Submit](http://openlab.citytech.cuny.edu/collegecouncil/files/2014/08/CommonCoreCourseSubmissionForm_4.2.12.doc) (if applicable) | N/A |
| Writing Intensive Form if course is intended to be a WIC (under development) | N/A |
| If course originated as an experimental course, then results of evaluation plan as developed with director of assessment. | N/A |
| **(Additional materials for** [**Curricular Experiments**](http://www.300jaystreet.com/college-council/curriculum_proposals/curricular-experiments)**)** |  |
| Plan and process for evaluation developed in consultation with the director of assessment. (Contact Director of Assessment for more information). |  |
| Established Timeline for Curricular Experiment |  |

**Proposed Course Name: Data Mining**

**Course Overview & Rationale**

Data Mining or knowledge discovery, is the computer-assisted process of digging through and analyzing enormous sets of data and then extracting the meaning of the data. Data mining models are at the heart of successful information and product search, automated merchandising, smart personalization, dynamic pricing, social network analysis, genetics, proteomics, and many other technology-based solutions to important problems in business. Also, this course is integral to the proposed Bachelors in Data Science curriculum program of the Computer Systems Technology department.

This course will be an introduction to data mining. Topics will range from statistics to machine learning, with a focus on analysis of large data sets. This course covers popular data mining methods for extracting knowledge from data. It will balance theory and practice - the principles of data mining methods will be discussed, but students will also acquire hands-on experience using state-of-the-art software (WEKA) to develop data mining solutions to scientific and business problems. Topics and related methods discussed in the class include: data preprocessing, classification and prediction, association mining, and cluster analysis. This course also includes a project, which will either involve research in data mining or instead address a practical problem using the methods and tools introduced in the course. The ethical issues that arise from the process of mining customer data and thereby compromising the privacy of those customers will be discussed. The development of this course lays the foundation and adds tremendous value to the creation and further progression of Data Science as a new, fast-growing, and promising major in the Computer Systems Technology department at New York City College of Technology.

**Course Need**

**Students who would take this class:** students who intend to major in Data Science

**Department**: Computer Systems Technology

**Program**: Bachelor of Science in Data Science

**The number of section (s) anticipated:** one section for the first year

**Projected headcount: 24** students

**Physical Resources required:** Basicsmartroomset**-**up**:** a screen, and an overhead projector/a TV set that is run by and connected to a computer

**Course overlap:** None

**Faculty** **qualified** **for** **teaching** **this** **course**: Yes, there are faculty members who have doctoral degrees in Data Mining and have taught this class before in other institutions.

**Course design**

**Course context:** This course will be mandatory for Data Science major students. Students are required to develop an independent project at the end of the semester.

**Course** **structure**: This course will be offered in a lecture style/format.

**Anticipated** **Pedagogical** **Strategies** **and** **Instructional** **Design**: This course will be run in a lecture-activity style/format. The course will start with a lecture, and then move on to creative in-class activities, such as using existing Data Mining tools, analyzing large datasets, and applying Clustering/Classification algorithms on the data.

**Providing Support to Programmatic Learning Outcomes:** This course requires satisfactory completion of individual assignments, two major exams and a final term project.

**NEW YORK CITY COLLEGE OF TECHNOLOGY/CUNY**

**Computer Systems Technology Department**

**Course Outline**

**CST 3502 – Data Mining**

**(2 class hours; 2 lab hours; 3 credits)**

**Course Description:**

Theory and practice of data mining methods for extracting knowledge from data using state-of-the-art software (WEKA) for solving scientific and business problems will be covered. Topics discussed in the class include: data preprocessing, classification and prediction, association mining, and cluster analysis. Projects either involve research in data mining or address a practical problem using the methods and tools introduced in the class

**Learning Outcomes:**

Upon successful completion of this course, students should be able to:

* Develop familiarity with data mining techniques, and demonstrate the ability to apply these techniques to real world problems
* Demonstrate an understanding of the role and impact of data mining in our society.
* Demonstrate an understanding of the ethical and privacy issues that arise from the process of mining user data.

**Software**: WEKA (<http://www.cs.waikato.ac.nz/ml/weka/>)

**General Education Outcomes:**

* **SKILLS/Inquiry/Analysis:** Students will employ scientific reasoning and logical thinking.
* **SKILLS/Communication:** Students will communicate in diverse settings and groups, using written (both reading and writing), oral (both speaking and listening), and visual means
* **VALUES, ETHICS, RELATIONSHIPS / Professional/Personal Development:** Students will work with teams, including those of diverse composition. Build consensus. Respect and use creativity.

**Pre-requisites:**

CST 2402 – Introduction to Data Science and CST 2312 – Information and Data Management I

**Required Text**: "Introduction to Data Mining", P. Tan, M. Steinbach & V. Kumar. Addison

Wesley. 2005.

**Additional Text**: “Data Science for Business”, F. Provost, T. Fawcett. 2013

**Academic Integrity –** Students and all others who work with information, ideas, texts, images, music, inventions, and other intellectual property owe their audience and sources accuracy and honesty in using, crediting, and citing sources. As a community of intellectual and professional workers, the College recognizes its responsibility for providing instruction in information literacy and academic integrity, offering models of good practice, and responding vigilantly and appropriately to infractions of academic integrity. Accordingly, academic dishonesty is prohibited in The City University of New York and at New York City College of Technology and is punishable by penalties, including failing grades, suspension, and expulsion. The complete text of the College policy on Academic Integrity may be found in the catalog.

**Grading Procedure:**

Test1 15%

Test2 15% Final Exam 20%

Midterm Exam 20%

Labs (in class) 10%

Lab Project 10%

Homework Assignments 10%

=====

TOTAL 100%

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Letter Grade** | A | A- | B+ | B | B- | C+ | C | D | F |
| **Numerical Grade** | 93-100 | 90-92.9 | 87-89.9 | 83-86.9 | 80-82.9 | 77-79.9 | 70-76.9 | 60-69.9 | <=59.9 |

**Course Outline:**

|  |  |  |
| --- | --- | --- |
| **Week** | **Topic Covered** | **Lab Assignments and Readings** |
| Week 1 | Introduction: what is data mining, supervised  vs unsupervised data mining | Ch1, pp. 1 – 11 |
| Data: types, concepts, instances/examples, features/attributes, target/class, outliers, etc. *Other issues related to data for data mining tasks will be discussed during the course of the class.* | (Ch2, stop middle 76) |
| Week 2 | Data (contd.)  Introduction to classification supervised learning, features, target: linear classifier  Assignment 1: Classification. | Ch4, pp. 145 –  149 (4.1, 4.2), lecture slides |
| Lab 1: Linear Classifier |
| Week 3 | HW1 discussion  Inductive learning: Decision trees, Shannon’s information gain, brief math tutorial on probabilities and entropy | Ch4, pp. 150 –  172 (4.3), lecture notes  Due: Assignment 1 |
| Week 4 | Overfitting, evaluation (accuracy, precision, recall, cross-validation, leave-one-out cross- validation, brief introduction to ROC curves) | Ch4, pp. 4.4, lecture notes |
| Brief intro to WEKA | Lab 2: WEKA Basics |
| Week 5 | Evaluation (contd.) |  |
| HW2 discussion  Math tutorial: Bayesian theorem  Assignment 2: Decision Trees and Shannon’s information. | Ch.5 pp. 227 –  229  Lab 3: Decision Trees |
| Week 6 | HW2 discussion  Math tutorial: Bayesian theorem | Ch.5 pp. 227 –  229  Lecture notes |
| Midterm review  Rule-based classifier | Ch5, 5.1  Due: Assignment 2 |
| Week 7 | Midterm  Midterm Revision | Ch.5 pp. 229 –  238 |
| Week 8 | Naïve Bayesian classifier  Nearest neighbor | Ch5, 5.2  Lab 4: Naïve Bayes |
| Week 9 | HW3 discussion  Nearest neighbor (contd.)  Unsupervised learning, similarity, example- based learning, clustering  Assignment 3: Clustering (using WEKA). | Ch8 8.1-8.4 (skip 8.2.6,  8.3.3, 8.3.4) |
| Week 10 | Clustering (contd.)  Association rule mining | Browse pp.  325 – 341 |
| Week 11 | Brief introduction to Deep Learning  HW4 discussion  Ethical Data Mining: Privacy sensitivity mining | Lecture notes  Due: Assignment 3 |
| Week 12 | Natural Language learning (contd.) | Lecture notes |
| Project presentations |  |
| Week 13 | Logistic Regression | Lab 5: Logistic Regression |
| Week 14  Week 15 | Term Project Presentations  Finals |  |

**COURSE INTENDED LEARNING OUTCOMES/ASSESSMENT METHODS**

|  |  |
| --- | --- |
| **LEARNING OUTCOMES** | **ASSESSMENT METHODS** |
| 1. Understand of data mining concepts, supervised vs unsupervised data mining | **1.** Group activities, written report. |
| 1. Implement a classification technique such as Decision trees, SVMs. | **2.** Individual oral presentations, in‐class group activities. |
| 1. Understand Overfitting, Evaluation (accuracy, precision, recall, cross-validation, leave-one-out cross- validation, brief introduction to ROC curves) | **3.** Classroom discussion, in‐class estimation group assignments. |
| 1. Demonstrate an understanding on how data mining is applied on Clustering and association mining applications | **4.** Classroom discussion, in‐class group assignments learning logs. |
| 1. Demonstrate an understanding of the role and impact of data mining in our society. | **5.** Individual short essay related to functions (e.g., population growth, economics, climate change). |
| 1. Demonstrate an understanding of the ethical and privacy issues that arise from the process of mining user data. | **6.** Classroom discussion on the ethical issues of compromising the privacy of user data based on real world case studies. |
| 1. Demonstrate an understanding of basic introductory concepts on deep learning and natural language processing. | **7.** Written report and group presentation (e.g., an analysis of the garbage patch in the Pacific Ocean), learning logs. |

**GENERAL EDUCATION LEARNING OUTCOMES/ASSESSMENT METHODS**

|  |  |
| --- | --- |
| **LEARNING OUTCOMES** | **ASSESSMENT METHODS** |
| **1.** Demonstrate the ability to work collaboratively and independently on assignments in and outside a classroom setting. | **1.** Classroom discussions, group assignments and individual oral presentations. |
| **2.** Understand and employ both quantitative and qualitative analysis to solve problems**.** | **2.** Classroom discussion, group activities, group presentations, quizzes, tests, final exam. |
| **3.** Develop reading, writing competencies, and listening skills. | **3.** Biweekly reading and writing assignments, individual and group presentation, classroom discussion. Each homework assignment requires writing. |
| **4.** Work with teams. Build consensus. Use creativity. | **4.** Group projects and presentations. |

**Bibliography:**

Bezdek, J. C., & Pal, S. K. (1992). *Fuzzy models for pattern recognition: Methods that search for structures in data*. New York: IEEE Press

Fayyad, U. M., Piatetsky-Shapiro, G., Smyth, P., & Uthurusamy, R. (Eds.). (1996). *Advances in knowledge discovery and data mining.* AAAI/MIT Press.

Han, J., & Kamber, M. (2000). *Data mining: Concepts and techniques*: Morgan Kaufmann.

Hastie, T., Tibshirani, R., & Friedman, J. H. (2001). *The elements of statistical learning: Data mining, inference, and prediction:*New York: Springer.

Jain, A. K., & Dubes, R. C. (1988). *Algorithms for clustering data*. New Jersey: Prentice Hall.

Jensen, F. V. (1996). *An introduction to bayesian networks*.

London: University College London Press.

Kaufman, L., & Rousseeuw, P. J. (1990). *Finding groups in data: An introduction to cluster analysis*. New York: John Wiley.

Michie, D., Spiegelhalter, D. J., & Taylor, C. C. (1994). *Machine learning, neural and statistical classification*: Ellis Horwood.

**CHANCELLOR’S REPORT FORM**

Department: Computer Systems Technology

Date: January 1st, 2018

**NEW COURSE PROPOSAL**

**Section AIV: New Course**

|  |  |
| --- | --- |
| **Department(s)** | **Computer Systems Technology** |
| **Academic Level** | **[X] Regular [  ] Compensatory [  ] Developmental [  ] Remedial** |
| **Subject Area** | Data Science |
| **Course Prefix** | **CST** |
| **Course Number** | **CST 3502** |
| **Course Title** | **Data Mining** |
| **Catalog Description** | Theory and practice of data mining methods for extracting knowledge from data using state-of-the-art software (WEKA) for solving scientific and business problems will be covered. Topics discussed in the class include: data preprocessing, classification and prediction, association mining, cluster analysis and ethical data mining. Projects either involve research in data mining or address a practical problem using the methods and tools introduced in the class. |
| **Prerequisite** | CST 2312 (Information and Data Management I) and CST 2402 (Introduction to Data Science) |
| **Corequisite** | **None** |
| **Credits** | **3** |
| **Contact Hours** | **2 Class Hours, 2 Lab Hours** |
| **Liberal Arts** | **[ ] Yes [ X  ] No** |
| **Course Attribute (e.g. Writing Intensive, etc)** | **It is not a writing intensive course** |
| **Course Applicability** | **[X] Major**  **[ ] Gen Ed Required [ ] Gen Ed - Flexible [ ] Gen Ed - College Option**  **[ ] English Composition [ ] World Cultures [ ] Speech**  **[ ] Mathematics [ ] US Experience in its Diversity**  **[ ] Interdisciplinary [ ] Science [ ] Creative Expression**  **[ ] Advanced Liberal Arts**  **[ ] Individual and Society**  **[ ] Scientific World** |
| **Effective Term** | **Fall 2019** |

**Rationale:** Data Mining or knowledge discovery, is the computer-assisted process of digging through and analyzing enormous sets of data and then extracting the meaning of the data. Data mining models are at the heart of successful information and product search, automated merchandizing, smart personalization, dynamic pricing, social network analysis, genetics, proteomics, and many other technology-based solutions to important problems in businesses. Also, this course is integral to the proposed Bachelors in Data Science curriculum program of the Computer Systems Technology department.

**LIBRARY RESOURCES & INFORMATION LITERACY: MAJOR CURRICULUM MODIFICATION**

Please complete for **all** major curriculum modifications. This information will assist the library in planning for new courses/programs.

Consult with your library faculty subject specialist (<http://cityte.ch/dir>) **3 weeks before the proposal deadline**.

**Course proposer:** please complete boxes 1-4. **Library faculty subject specialist:** please complete box 5.

|  |  |  |
| --- | --- | --- |
| **1** | **Title of proposal**  CST 3502: Data Mining | **Department/Program**  Computer Systems Technology / BS in Data Science |
|  | **Proposed by** (include email & phone)  Dr. Ashwin Satyanarayana  [asatyanarayana@citytech.cuny.edu](mailto:asatyanarayana@citytech.cuny.edu)  518-330-7907 | **Expected date course(s) will be offered**  Fall 2019  **# of students:** 24 |

|  |  |
| --- | --- |
| **2** | **The library cannot purchase reserve textbooks for every course at the college, nor copies for all students. Consult our website (**[**http://cityte.ch/curriculum**](http://cityte.ch/curriculum)**) for articles and ebooks for your courses, or our open educational resources (OER) guide (**[**http://cityte.ch/oer**](http://cityte.ch/oer)**). Have you considered using a freely-available OER or an open textbook in this course?**  No, I have not considered freely available OER or an open textbook for this course. |

|  |  |
| --- | --- |
| **3** | **Beyond the required course materials, are City Tech library resources sufficient for course assignments? If additional resources are needed, please provide format details (e.g. ebook, journal, DVD, etc.), full citation (author, title, publisher, edition, date), price, and product link.**  Yes, City Tech Library resources are sufficient for the proposed course assignments because the main readings for the course are a required textbook and journal articles that will be assigned by instructor. Students should be able to locate the selected journal articles in library. |

|  |  |
| --- | --- |
| **4** | **Library faculty focus on strengthening students' information literacy skills in finding, critically evaluating, and ethically using information. We collaborate on developing assignments and customized instruction and research guides. When this course is offered, how do you plan to consult with the library faculty subject specialist for your area? Please elaborate** Yes, a library faculty will be invited for a guest lecture on how to find, and use information in conducting original research, which is a major portion of the proposed course assignments. |

|  |  |
| --- | --- |
| **5** | **Library Faculty Subject Specialist Prof. Junior Tidal**  **Comments and Recommendations**  After surveying the library’s collection, I believe that it may be necessary to add more materials related to data mining to support this course. We have very few print copies of books related to data mining, yet they are a few years old. I suggest adding more books in both the general and reference collections.  **Date 12.13.17** |

# New Course Proposal #4: CST3512 Information and Data Management II

New York City College of Technology, CUNY

CURRICULUM MODIFICATION PROPOSAL FORM

This form is used for all curriculum modification proposals. See the [Proposal Classification Chart](http://openlab.citytech.cuny.edu/collegecouncil/files/2014/08/2013-10-09-Proposal_Classification_Chart.pdf) for information about what types of modifications are major or minor. Completed proposals should be emailed to the Curriculum Committee chair.

|  |  |
| --- | --- |
| **Title of Proposal** | **CST 3512 Information and Data Management II** |
| **Date** | **Jan 4th, 2018** |
| **Major or Minor** | **Major** |
| **Proposer’s Name** | **Dr. Elena Filatova** |
| **Department** | **Computer Systems Technology** |
| **Date of Departmental Meeting in which proposal was approved** | **11/17/2017** |
| **Department Chair Name** | **Dr. Hong Li** |
| **Department Chair Signature and Date** |  |
| **Academic Dean Name** | **Dean Hom** |
| **Academic Dean Signature and Date** | 2/22/18 |
| **Brief Description of Proposal**  (Describe the modifications contained within this proposal in a succinct summary. More detailed content will be provided in the proposal body. | Title: Information and Data Management II  This proposal explains the rationale for developing Information and Data Management I as a required course in the Data Science curriculum of the Computer Systems Technology Department. The course outline and sample syllabus are attached to the proposal. |
| **Brief Rationale for Proposal**  (Provide a concise summary of why this proposed change is important to the department. More detailed content will be provided in the proposal body). | Information and Data Management II is the second course in the two-course series. This two-course series provides a comprehensive introduction to programming with Python that goes beyond the basic concepts and focuses around data processing and analysis. In the modern world, data is the lifeblood of any organization and the driving force for development of new successful models for decision making. Competency in programming is an essential skill for successfully extracting information and knowledge form data. Also, this course is integral to the proposed Bachelors in Data Science curriculum program of the Computer Systems Technology department. |
| **Proposal History**  (Please provide history of this proposal: is this a resubmission? An updated version? This may most easily be expressed as a list). | **This is a first submission.** |
| **CUNY – Course Equivalencies**  Provide information about equivalent courses within CUNY, if any. | None |
| **Intent to Submit as Common Core**  If this course is intended to fulfill one of the requirements in the common core, then indicate which area. | N/A |

Please include all appropriate documentation as indicated in the Curriculum Modification Checklist.

For each new course, please also complete the New Course Proposal and submit in this document.

Please submit this document as a single .doc or .rtf format. If some documents are unable to be converted to .doc, then please provide all documents archived into a single .zip file.

**ALL PROPOSAL CHECK LIST**

|  |  |
| --- | --- |
| Completed CURRICULUM MODIFICATION FORM including: |  |
| * Brief description of proposal | X |
| * Rationale for proposal | X |
| * Date of department meeting approving the modification | X |
| * Chair’s Signature | X |
| * Dean’s Signature |  |
| Evidence of consultation with affected departments  List of the programs that use this course as required or elective, and courses that use this as a prerequisite. | X |
| Documentation of Advisory Commission views (if applicable). | N/A |
| Completed [Chancellor’s Report Form](http://openlab.citytech.cuny.edu/collegecouncil/files/2014/08/2013-10-09-Chancellor_Report_Quick_Reference_Guide1.doc). | X |

**EXISTING PROGRAM MODIFICATION PROPOSALS**

|  |  |
| --- | --- |
| Documentation indicating core curriculum requirements have been met for new programs/options or program changes. | X |
| Detailed rationale for each modification (this includes minor modifications) | N/A |

New York City College of Technology, CUNY

NEW COURSE PROPOSAL FORM

This form is used for all new course proposals. Attach this to the [Curriculum Modification Proposal Form](http://openlab.citytech.cuny.edu/collegecouncil/files/2014/08/2013-10-10-Curriculum_Modification_Proposal_Form.docx) and submit as one package as per instructions. Use one New Course Proposal Form for each new course.

|  |  |
| --- | --- |
| **Course Title** | Information and Data Management II |
| **Proposal Date** | Jan 4th, 2018 |
| **Proposer’s Name** | Dr. Elena Filatova |
| **Course Number** | CST 3512 |
| **Course Credits, Hours** | 3 credits, 2 lecture hours and 2 lab hours |
| **Course Pre / Co-Requisites** | Pre-req: CST 2312 (Information and Data Management I) |
| **Catalog Course Description** | Topics related to the digital infrastructure, management and curation of data both structured (record-based) and unstructured (such as text). The course concentrates on the Python tools for accessing data sources (crawling, Web APIs), data cleaning, text data analysis, data visualization and applications to different domain of interest. |
| **Brief Rationale**  Provide a concise summary of why this course is important to the department, school or college. | Information and Data Management II is the second course in the two-course series. This two-course series provides a comprehensive introduction to programming with Python that goes beyond the basic concepts and focuses around data processing and analysis. In the modern world, data is the lifeblood of any organization and the driving force for development of new successful models for decision making. Competency in programming is an essential skill for successfully extracting information and knowledge form data. Also, this course is integral to the proposed Bachelors in Data Science curriculum program of the Computer Systems Technology department. |
| **CUNY – Course Equivalencies**  Provide information about equivalent courses within CUNY, if any. | N/A |
| **Intent to Submit as Common Core**  If this course is intended to fulfill one of the requirements in the common core, then indicate which area. | No. |
| **Intent to Submit as An Interdisciplinary Course** | No. |
| **Intent to Submit as a Writing Intensive Course** | No. |

Please include all appropriate documentation as indicated in the NEW COURSE PROPOSAL Combine all information into a single document that is included in the Curriculum Modification Form.

**NEW COURSE PROPOSAL CHECK LIST**

Use this checklist to ensure that all required documentation has been included. You may wish to use this checklist as a table of contents within the new course proposal.

|  |  |
| --- | --- |
| **Completed NEW COURSE PROPOSAL FORM** |  |
| * Title, Number, Credits, Hours, Catalog course description | X |
| * Brief Rationale | X |
| * CUNY – Course Equivalencies | X |
| Completed [Library Resources and Information Literacy Form](http://openlab.citytech.cuny.edu/collegecouncil/files/2014/08/curriculum_modification_library_form.doc) | X |
| **Course Outline**  Include within the outline the following. | X |
| Hours and Credits for Lecture and Labs  If hours exceed mandated Carnegie Hours, then rationale for this | X |
| Prerequisites/Co- requisites | X |
| Detailed Course Description | X |
| Course Specific Learning Outcome and Assessment Tables   * Discipline Specific * General Education Specific Learning Outcome and Assessment Tables | X |
| Example Weekly Course outline | X |
| Grade Policy and Procedure | X |
| Recommended Instructional Materials (Textbooks, lab supplies, etc) | X |
| Library resources and bibliography | X |
| **Course Need Assessment.**  Describe the need for this course. Include in your statement the following information. | X |
| Target Students who will take this course. Which programs or departments, and how many anticipated?  Documentation of student views (if applicable, e.g. non-required elective). | X |
| Projected headcounts (fall/spring and day/evening) for each new or modified course. | X |
| If additional physical resources are required (new space, modifications, equipment), description of these requirements. If applicable, Memo or email from the VP for Finance and Administration with written comments regarding additional and/or new facilities, renovations or construction. | X |
| Where does this course overlap with other courses, both within and outside of the department? | X |
| Does the Department currently have full time faculty qualified to teach this course? If not, then what plans are there to cover this? | X |
| If needs assessment states that this course is required by an accrediting body, then provide documentation indicating that need. | N/A |
| **Course Design**  Describe how this course is designed. | X |
| Course Context (e.g. required, elective, capstone) | X |
| Course Structure: how the course will be offered (e.g. lecture, seminar, tutorial, fieldtrip)? | X |
| Anticipated pedagogical strategies and instructional design (e.g. Group Work, Case Study, Team Project, Lecture) | X |
| How does this course support Programmatic Learning Outcomes? | X |
| Is this course designed to be partially or fully online? If so, describe how this benefits students and/or program. | X |
| **Additional Forms for Specific Course Categories** |  |
| [Interdisciplinary Form](http://openlab.citytech.cuny.edu/collegecouncil/files/2014/08/Application-for-Interdisciplinary-Course-Designation.docx) (if applicable) | N/A |
| Interdisciplinary Committee Recommendation (if applicable and if received)\*  \*Recommendation must be received before consideration by full Curriculum Committee | N/A |
| [Common Core (Liberal Arts) Intent to Submit](http://openlab.citytech.cuny.edu/collegecouncil/files/2014/08/CommonCoreCourseSubmissionForm_4.2.12.doc) (if applicable) | N/A |
| Writing Intensive Form if course is intended to be a WIC (under development) | N/A |
| If course originated as an experimental course, then results of evaluation plan as developed with director of assessment. | N/A |
| **(Additional materials for** [**Curricular Experiments**](http://www.300jaystreet.com/college-council/curriculum_proposals/curricular-experiments)**)** |  |
| Plan and process for evaluation developed in consultation with the director of assessment. (Contact Director of Assessment for more information). |  |
| Established Timeline for Curricular Experiment |  |

**Proposed Course Name: Information and Data Management II**

**Course Overview & Rationale**

Information and Data Management II is the second course in the two-course series. This two-course series provides a comprehensive introduction to programming with Python that goes beyond the basic concepts and focuses around data processing and analysis. In the modern world, data is the lifeblood of any organization and the driving force for development of new successful models for decision making. Competency in programming is an essential skill for successfully extracting information and knowledge form data. Also, this course is integral to the proposed Bachelors in Data Science curriculum program of the Computer Systems Technology department.

This course continues to familiarize students with the informatics and intellectual tools necessary for students to become efficient and effective information users. The course covers topics related to the digital infrastructure, management and curation of data both structured (record-based) and unstructured (such as text). For structured data (data series, data frames), the course introduces Time Series data analysis and the basics of data visualization; for unstructured data (text), the course introduces text mining techniques. The course is project-based. During the course several case studies of using Information and Data Management are discussed. Based on these case studies the students develop their own projects where they choose a data set, use the appropriate data processing techniques for data analysis, and present their finding and conclusions reached using the information and data management and analysis techniques discussed in class. The development of this course lays the foundation for the data acquisition and analysis step that is the first and necessary step in the decision making based on data pipeline. This course is designed to be a stepping stone for more advanced courses within the Data Science major, a new, fast-growing, and promising major in the Computer Systems Technology department at New York City College of Technology.

**Course Need**

**Students who would take this class:** students who intend to major in Data Science

**Department**: Computer Systems Technology

**Program**: Bachelors in Data Science

**The number of section (s) anticipated:** one section for the first year

**Projected headcount: 24** students

**Physical Resources required:** Basicsmartroomset**-**up**:** a screen, and an overhead projector/a TV set that is run by and connected to a computer

**Course overlap:** None

**Faculty** **qualified** **for** **teaching** **this** **course**: Yes, there are faculty members who have doctoral degrees in Computer Science with the concentration in Data Processing for various domains.

**Course design**

**Course context:** This course will be required of Data Science major students. Students are required to develop an independent project at the end of the semester.

**Course** **structure**: This course will be offered in a lecture style/format.

**Anticipated** **Pedagogical** **Strategies** **and** **Instructional** **Design**: This class will be run in a lecture-activity style/format. The class will start with a lecture, and then move on to creative in-class activities, such as using Python INotebook for Python-related exercises, collecting and analyzing datasets, using web APIs, using time series data analysis for the data store over time.

**Providing Support to Programmatic Learning Outcomes:** This course requires satisfactory completion of individual assignments, two major exams and a final term project.

**NEW YORK CITY COLLEGE OF TECHNOLOGY/CUNY**

**Computer Systems Technology Department**

**Course Outline**

**CST 3512 – Information and Data Management II**

**(2 class hours; 2 lab hours; 3 credits)**

**Course Description:**

This course continues to familiarize students with the informatics and intellectual tools necessary for students to become efficient and effective information users. The course covers topics related to the digital infrastructure, management and curation of data both structured (record-based) and unstructured (such as text). For structured data (data series, data frames), the course introduces Time Series data analysis and the basics of data visualization; for unstructured data (text), the course introduces text mining techniques. The course is project-based. During the course several case studies of using Information and Data Management are discussed. Based on these case studies the students develop their own projects where they choose a data set, use the appropriate data processing techniques for data analysis, and present their finding and conclusions reached using the information and data management and analysis techniques discussed in class.

**Final Project:**

The final project is a team project with an oral presentation. The teams choose an on-line data collection, obtain this data collection, analyze it, and create a visualization supporting interesting facts found in this data collection.

**Course Objectives:**

Upon successful completion of this course, students should be able to:

* Demonstrate knowledge of web APIs
* Demonstrate the skills of processing information downloaded from Internet
* Demonstrate knowledge of web crawling
* Demonstrate knowledge of basic text analysis techniques
* Demonstrate knowledge of basic visualization techniques

**Software**: Python and Python Libraries

**General Education Outcomes:**

* **SKILLS/Inquiry/Analysis:** Students will employ scientific reasoning and logical thinking.
* **SKILLS/Communication:** Students will communicate in diverse settings and groups, using written (both reading and writing), oral (both speaking and listening), and visual means
* **VALUES, ETHICS, RELATIONSHIPS / Professional/Personal Development:** Students will work with teams, including those of diverse composition. Build consensus. Respect and use creativity.

**Pre-requisites:**

CST 2312 Information and Data Management I

**Required Text**:

* Charles R. Severance. Python for Everybody: Exploring Information.

ISBN-13: 978-1492339243

Free on-line version: http://www.pythonlearn.com/book.php#python-for-informatics

* Wes McKinney. Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython, 2nd edition

ISBN-13: 978-1491957660

**Additional Reading Materials:**

The instructor will identify several additional information resources during the semester including Internet resources, print material (handouts) and reference books.

**Tools and On-Line Documentation:**

* Python: https://www.python.org/
* Python Data Analysis Library: http://pandas.pydata.org/
* Python package for scientific computing: http://www.numpy.org/
* IPython command shell for interactive computing: https://ipython.org/
* Anaconda Data Science ecosystem: https://www.continuum.io/

**Grading Procedure:**

Midterm Exam 20%

Final Exam 25%

Lab Project 20%

Homework assignments 25%

In-class labs/quizzes 10%

=====

TOTAL 100%

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Letter Grade** | A | A- | B+ | B | B- | C+ | C | D | F |
| **Numerical Grade** | 93-100 | 90-92.9 | 87-89.9 | 83-86.9 | 80-82.9 | 77-79.9 | 70-76.9 | 60-69.9 | <=59.9 |

The grade distribution follows the information in the NYCCT Student Handbook (p.43).

**Academic Integrity**: During the course of the class you are required to follow the NYCCT Academic Integrity Standards described in the Student Handbook (pp.95 – 99)

**NYCCT Student Handbook** can be downloaded here: <http://www.citytech.cuny.edu/current-student/docs/StudentHandbook.pdf>.

**Course Outline:**

|  |  |  |
| --- | --- | --- |
| **Week** | **Topics** | **Assignments and Readings** |
| 1, 2 | Review of Python Pandas basics:  data series and data frames (CST 2302) using IPython Notes | Lab 1: IPython – Data Series and Data Frames |
| 3, 4 | More on data series and data frames  Lab: data series, pandas dat frames | Assignment 1 |
| 5 | Basics of visualization using MatPlotLib  Lab: exercises using matplotlib library | Severance: Ch. 15  Lab 2: MatPlotLib for Visualization |
| 6,7 | Introduction to text analysis | McKinney: Ch. 6, Ch. 7.3  Lab 3: Text analysis using search engine data  Severance: Ch. 11 |
| 8 | **Midterm Exam**;  Introduction to Time Series data analysis, storing data over time | Assignment 2  Lab 4: Time series data analysis |
| 9 | Introduction to Time Series data analysis, storing data over time  Lab: time series analysis | McKinney: Ch. 11 |
| 10 | Lab: Practice with a sample API (e.g., Citibike API), part I;  Team project: proposals | <https://www.citibikenyc.com/system-data> |
| 11 | Practice with a sample API (e.g., Citibike API), part II;  Lab team project: data collection from on-line resources | <https://www.citibikenyc.com/system-data>  Assignment 3  Lab 5: Citibike API practice |
| 12 | Practice with a sample API (suggested in class);  Lab team project: data collection from on-line resources | Lab 6: Data Collection from online resources |
| 13 | Team project: data analysis and visualization | Lab 6: Data Analysis |
| 14 | Final exam review;  **Final exam (the final exam is cumulative)** |  |
| 15 | **Final Project Presentations / Discussion** |  |

**ASSESSMENT CRITERIA:** For successful completion of this course the student should be able to:

|  |  |
| --- | --- |
| **For the successful completion of this course a student should be able to:** | **Evaluation methods and criteria** |
| Demonstrate knowledge of web APIs | Students will create Python scripts (and run Python commands in the Shell mode) that use web APIs to upload locally wen pages. |
| Demonstrate the skills of processing information downloaded from Internet | Students will create Python scripts (and run Python commands in the Shell mode) that read the source text of a downloaded web page and extract the desired information. |
| Demonstrate knowledge of  web crawling | Students will create Python scripts that can access the information form the web pages from the sites that do not have a custom web API. |
| Demonstrate knowledge of basic text analysis techniques | Students will create Python scripts that analyze text information in the source files (e.g., find most frequent words, find most important entities mentioned in text, etc.) |
| Demonstrate knowledge of basic visualization techniques | Students will create Python scripts that visualize the input data. |
| Work effectively in a team | The final project is a team project. The students will group into teams and create a project that demonstrates their knowledge of the programming tools and techniques learned in class. Will use the Internet and other resources to complete the project. Also, there will be an oral presentation made to the class. It will include their learning experience in working in a group. |

**GENERAL EDUCATION LEARNING OUTCOMES/ASSESSMENT METHODS**

|  |  |
| --- | --- |
| **LEARNING OUTCOMES** | **ASSESSMENT METHODS** |
| **1.** Demonstrate the ability to work collaboratively and independently on assignments in and outside a classroom setting. | **1.** Classroom discussions, group assignments and individual oral presentations. |
| **2.** Understand and employ both quantitative and qualitative analysis to solve problems**.** | **2.** Classroom discussion, group activities, group presentations, quizzes, tests, final exam. |
| **3.** Develop reading, writing competencies, and listening skills. | **3.** Biweekly reading and writing assignments, individual and group presentation, classroom discussion. Each homework assignment requires writing. |
| **4.** Work with teams. Build consensus. Use creativity. | **4.** Group projects and presentations. |

**Bibliography:**

* Charles R. Severance (2013). *Python for Informatics: Exploring Information*. Publisher: CreateSpace Independent Publishing Platform; 1 edition,

ISBN-10: 1492339245

Free on-line version: <http://www.pythonlearn.com/book.php#python-for-informatics>

* Wes McKinney (2017) *Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython.*

Publisher: O'Reilly Media; 2 edition (October 20, 2017)

ISBN-10: 1491957662

* Armando Fandango (2017). *Python Data Analysi*s

Publisher: Packt Publishing - ebooks Account; 2nd Revised edition edition (March 27, 2017)

ISBN-10: 1787127486

* Jake VaderPlas (2016). *Python Data Science Handbook: Essential Tools for Working with Data.*

Publisher: O'Reilly Media; 1 edition (December 10, 2016)

ISBN-10: 1491912057

**CHANCELLOR’S REPORT FORM**

Department: Computer Systems Technology

Date: January 4th, 2018

**NEW COURSE PROPOSAL**

**Section AIV: New Course**

|  |  |
| --- | --- |
| **Department(s)** | **Computer Systems Technology** |
| **Academic Level** | **[X] Regular [  ] Compensatory [  ] Developmental [  ] Remedial** |
| **Subject Area** | Data Science |
| **Course Prefix** | **CST** |
| **Course Number** | **CST 3512** |
| **Course Title** | **Information and Data Management II** |
| **Catalog Description** | Topics related to the digital infrastructure, management and curation of data both structured (record-based) and unstructured (such as text). The course concentrates on the Python tools for accessing data sources (crawling, Web APIs), data cleaning, text data analysis, data visualization and applications to different domain of interest. |
| **Prerequisite** | **CST 2312** |
| **Corequisite** | **None** |
| **Credits** | **3** |
| **Contact Hours** | **2 Class Hours, 2 Lab Hours** |
| **Liberal Arts** | **[ ] Yes [ X  ] No** |
| **Course Attribute (e.g. Writing Intensive, etc)** | **It is not a writing intensive course** |
| **Course Applicability** | **[X] Major**  **[ ] Gen Ed Required [ ] Gen Ed - Flexible [ ] Gen Ed - College Option**  **[ ] English Composition [ ] World Cultures [ ] Speech**  **[ ] Mathematics [ ] US Experience in its Diversity**  **[ ] Interdisciplinary [ ] Science [ ] Creative Expression**  **[ ] Advanced Liberal Arts**  **[ ] Individual and Society**  **[ ] Scientific World** |
| **Effective Term** | **Fall 2019** |

**Rationale:** Information and Data Management II is the second course in the two-course series. This two-course series provides a comprehensive introduction to programming with Python that goes beyond the basic concepts and focuses around data processing and analysis. In the modern world, data is the lifeblood of any organization and the driving force for development of new successful models for decision making. Competency in programming is an essential skill for successfully extracting information and knowledge form data. Also, this course is integral to the proposed Bachelors in Data Science curriculum program of the Computer Systems Technology department.

**LIBRARY RESOURCES & INFORMATION LITERACY: MAJOR CURRICULUM MODIFICATION**

Please complete for **all** major curriculum modifications. This information will assist the library in planning for new courses/programs.

Consult with your library faculty subject specialist (<http://cityte.ch/dir>) **3 weeks before the proposal deadline**.

**Course proposer:** please complete boxes 1-4. **Library faculty subject specialist:** please complete box 5.

|  |  |  |
| --- | --- | --- |
| **1** | **Title of proposal**  CST 3512: Information and Data Management II | **Department/Program**  Computer and Systems Technology |
|  | **Proposed by** (include email & phone)  Elena Filatova  efilatova@citytech.cuny.edu | **Expected date course(s) will be offered**  Fall 2019  **# of students** 24 |

|  |  |
| --- | --- |
| **2** | **The library cannot purchase reserve textbooks for every course at the college, nor copies for all students. Consult our website (**[**http://cityte.ch/curriculum**](http://cityte.ch/curriculum)**) for articles and ebooks for your courses, or our open educational resources (OER) guide (**[**http://cityte.ch/oer**](http://cityte.ch/oer)**). Have you considered using a freely-available OER or an open textbook in this course?**  One of the books is a free on-line book, the other book is not free:  Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython by Wes McKinney, 2nd ed.  ISBN-10: 1449319793 |

|  |  |
| --- | --- |
| **3** | **Beyond the required course materials, are City Tech library resources sufficient for course assignments? If additional resources are needed, please provide format details (e.g. ebook, journal, DVD, etc.), full citation (author, title, publisher, edition, date), price, and product link.**  Tools and On-Line Documentation (free, on-line)  Python: https://www.python.org/  Python Data Analysis Library: http://pandas.pydata.org/  Python package for scientific computing: http://www.numpy.org/  IPython command shell for interactive computing: https://ipython.org/  Anaconda Data Science ecosystem: https://www.continuum.io/ |

|  |  |
| --- | --- |
| **4** | **Library faculty focus on strengthening students' information literacy skills in finding, critically evaluating, and ethically using information. We collaborate on developing assignments and customized instruction and research guides. When this course is offered, how do you plan to consult with the library faculty subject specialist for your area? Please elaborate.**  The project for the course concerns an on-line data set. There will be a discussion in class, what data can be downloaded and used for the project, whether there are any ethical consequences for using this data set and making conclusions about this data set. |

|  |  |
| --- | --- |
| **5** | **Library Faculty Subject Specialist Prof. Junior Tidal**  **Comments and Recommendations**  After surveying the collection, I believe that the library can adequately support this course. It may be necessary to add books on the Weka and Java to augment the current collection.  **Date 12.13.17** |

# New Course Proposal #5: CST3602 Data Visualization

New York City College of Technology, CUNY

CURRICULUM MODIFICATION PROPOSAL FORM

This form is used for all curriculum modification proposals. See the [Proposal Classification Chart](http://openlab.citytech.cuny.edu/collegecouncil/files/2014/08/2013-10-09-Proposal_Classification_Chart.pdf) for information about what types of modifications are major or minor. Completed proposals should be emailed to the Curriculum Committee chair.

|  |  |
| --- | --- |
| **Title of Proposal** | **CST3602 Data Visualization** |
| **Date** | **Jan 4th, 2018** |
| **Major or Minor** | **Major** |
| **Proposer’s Name** | **Dr. Elena Filatova** |
| **Department** | **Computer Systems Technology** |
| **Date of Departmental Meeting in which proposal was approved** | **12/8/2017** |
| **Department Chair Name** | **Dr. Hong Li** |
| **Department Chair Signature and Date** |  |
| **Academic Dean Name** | **Dean Hom** |
| **Academic Dean Signature and Date** | 2/22/18 |
| **Brief Description of Proposal**  (Describe the modifications contained within this proposal in a succinct summary. More detailed content will be provided in the proposal body. | Title: Data Visualization  This proposal explains the rationale for developing the Data Visualization course as a required course in the Data Science curriculum of the Computer Systems Technology Department. The course outline and sample syllabus are attached to the proposal. |
| **Brief Rationale for Proposal**  (Provide a concise summary of why this proposed change is important to the department. More detailed content will be provided in the proposal body). | Data Visualization allows to combat the information overload in the modern world. A well-designed visual data rendering can improve the data comprehension, information interpretation, and decision-making process. Also, this course is integral to the proposed Bachelors in Data Science curriculum program of the Computer Systems Technology department. |
| **Proposal History**  (Please provide history of this proposal: is this a resubmission? An updated version? This may most easily be expressed as a list). | This is a first submission. |
| **CUNY – Course Equivalencies**  Provide information about equivalent courses within CUNY, if any. | None |
| **Intent to Submit as Common Core**  If this course is intended to fulfill one of the requirements in the common core, then indicate which area. | N/A |

Please include all appropriate documentation as indicated in the Curriculum Modification Checklist.

For each new course, please also complete the New Course Proposal and submit in this document.

Please submit this document as a single .doc or .rtf format. If some documents are unable to be converted to .doc, then please provide all documents archived into a single .zip file.

**ALL PROPOSAL CHECK LIST**

|  |  |
| --- | --- |
| Completed CURRICULUM MODIFICATION FORM including: |  |
| * Brief description of proposal | X |
| * Rationale for proposal | X |
| * Date of department meeting approving the modification | X |
| * Chair’s Signature | X |
| * Dean’s Signature |  |
| Evidence of consultation with affected departments  List of the programs that use this course as required or elective, and courses that use this as a prerequisite. | X |
| Documentation of Advisory Commission views (if applicable). | N/A |
| Completed [Chancellor’s Report Form](http://openlab.citytech.cuny.edu/collegecouncil/files/2014/08/2013-10-09-Chancellor_Report_Quick_Reference_Guide1.doc). | X |

**EXISTING PROGRAM MODIFICATION PROPOSALS**

|  |  |
| --- | --- |
| Documentation indicating core curriculum requirements have been met for new programs/options or program changes. | X |
| Detailed rationale for each modification (this includes minor modifications) | N/A |

New York City College of Technology, CUNY

NEW COURSE PROPOSAL FORM

This form is used for all new course proposals. Attach this to the [Curriculum Modification Proposal Form](http://openlab.citytech.cuny.edu/collegecouncil/files/2014/08/2013-10-10-Curriculum_Modification_Proposal_Form.docx) and submit as one package as per instructions. Use one New Course Proposal Form for each new course.

|  |  |
| --- | --- |
| **Course Title** | Data Visualization |
| **Proposal Date** | Jan 4th, 2018 |
| **Proposer’s Name** | Dr. Elena Filatova |
| **Course Number** | CST 3602 |
| **Course Credits, Hours** | 3 credits, 2 lecture hours and 2 lab hours |
| **Course Pre / Co-Requisites** | Pre-req: CST 3512 (Information and Data Management II) and MAT 2572 (Probability and Mathematical Statistics I) |
| **Catalog Course Description** | Basic visualization design techniques and evaluation principles, techniques for visualizing multivariate, temporal, text-based, geospatial, hierarchical, and network/graph-based data. The final project for the course is to acquire, parse, analyze a large dataset, and create meaningful, good visualization for this dataset. |
| **Brief Rationale**  Provide a concise summary of why this course is important to the department, school or college. | Data Visualization allows to combat the information overload in the modern world. A well-designed visual data rendering can improve the data comprehension, information interpretation, and decision-making process. Also, this course is integral to the proposed Bachelors in Data Science curriculum program of the Computer Systems Technology department. |
| **CUNY – Course Equivalencies**  Provide information about equivalent courses within CUNY, if any. | N/A |
| **Intent to Submit as Common Core**  If this course is intended to fulfill one of the requirements in the common core, then indicate which area. | No. |
| **Intent to Submit as An Interdisciplinary Course** | No. |
| **Intent to Submit as a Writing Intensive Course** | No. |

**NEW COURSE PROPOSAL CHECK LIST**

Use this checklist to ensure that all required documentation has been included. You may wish to use this checklist as a table of contents within the new course proposal.

|  |  |
| --- | --- |
| **Completed NEW COURSE PROPOSAL FORM** |  |
| * Title, Number, Credits, Hours, Catalog course description | X |
| * Brief Rationale | X |
| * CUNY – Course Equivalencies | X |
| Completed [Library Resources and Information Literacy Form](http://openlab.citytech.cuny.edu/collegecouncil/files/2014/08/curriculum_modification_library_form.doc) | X |
| **Course Outline**  Include within the outline the following. | X |
| Hours and Credits for Lecture and Labs  If hours exceed mandated Carnegie Hours, then rationale for this | X |
| Prerequisites/Co- requisites | X |
| Detailed Course Description | X |
| Course Specific Learning Outcome and Assessment Tables   * Discipline Specific * General Education Specific Learning Outcome and Assessment Tables | X |
| Example Weekly Course outline | X |
| Grade Policy and Procedure | X |
| Recommended Instructional Materials (Textbooks, lab supplies, etc) | X |
| Library resources and bibliography | X |
| **Course Need Assessment.**  Describe the need for this course. Include in your statement the following information. | X |
| Target Students who will take this course. Which programs or departments, and how many anticipated?  Documentation of student views (if applicable, e.g. non-required elective). | X |
| Projected headcounts (fall/spring and day/evening) for each new or modified course. | X |
| If additional physical resources are required (new space, modifications, equipment), description of these requirements. If applicable, Memo or email from the VP for Finance and Administration with written comments regarding additional and/or new facilities, renovations or construction. | X |
| Where does this course overlap with other courses, both within and outside of the department? | X |
| Does the Department currently have full time faculty qualified to teach this course? If not, then what plans are there to cover this? | X |
| If needs assessment states that this course is required by an accrediting body, then provide documentation indicating that need. | N/A |
| **Course Design**  Describe how this course is designed. | X |
| Course Context (e.g. required, elective, capstone) | X |
| Course Structure: how the course will be offered (e.g. lecture, seminar, tutorial, fieldtrip)? | X |
| Anticipated pedagogical strategies and instructional design (e.g. Group Work, Case Study, Team Project, Lecture) | X |
| How does this course support Programmatic Learning Outcomes? | X |
| Is this course designed to be partially or fully online? If so, describe how this benefits students and/or program. | X |
| **Additional Forms for Specific Course Categories** |  |
| [Interdisciplinary Form](http://openlab.citytech.cuny.edu/collegecouncil/files/2014/08/Application-for-Interdisciplinary-Course-Designation.docx) (if applicable) | N/A |
| Interdisciplinary Committee Recommendation (if applicable and if received)\*  \*Recommendation must be received before consideration by full Curriculum Committee | N/A |
| [Common Core (Liberal Arts) Intent to Submit](http://openlab.citytech.cuny.edu/collegecouncil/files/2014/08/CommonCoreCourseSubmissionForm_4.2.12.doc) (if applicable) | N/A |
| Writing Intensive Form if course is intended to be a WIC (under development) | N/A |
| If course originated as an experimental course, then results of evaluation plan as developed with director of assessment. | N/A |
| **(Additional materials for** [**Curricular Experiments**](http://www.300jaystreet.com/college-council/curriculum_proposals/curricular-experiments)**)** |  |
| Plan and process for evaluation developed in consultation with the director of assessment. (Contact Director of Assessment for more information). |  |
| Established Timeline for Curricular Experiment |  |

**Proposed Course Name: Data Visualization**

**Course Overview & Rationale**

Data Visualization allows to combat the information overload in the modern world. A well-designed visual data rendering can improve the data comprehension, information interpretation, and decision-making process. Also, this course is integral to the proposed Bachelors in Data Science curriculum program of the Computer Systems Technology department.

This course introduces students to information visualization basics and provides students with the knowledge about design and information literacy perspective, touching on what makes good and bad visualization. The course covers basic visualization design techniques and evaluation principles. Students learn techniques for visualizing multivariate, temporal, text-based, geospatial, hierarchical, and network/graph-based data. The final project for the course is to acquire, parse, analyze a large dataset, and create meaningful, good visualization for this dataset.

**Course Need**

**Students who would take this class:** students who intend to major in Data Science

**Department**: Computer Systems Technology

**Program**: Bachelors of Science in Data Science

**The number of section (s) anticipated:** one section for the first year

**Projected headcount: 24** students

**Physical Resources required:** Basicsmartroomset**-**up**:** a screen, and an overhead projector/a TV set that is run by and connected to a computer

**Course overlap:** None

**Faculty** **qualified** **for** **teaching** **this** **course**: Yes, there are faculty members who have doctoral degrees in Computer Science with the concentration in Data Processing for various domains.

**Course design**

**Course context:** This course will be required of Data Science major students. Students are required to develop an independent project at the end of the semester.

**Course** **structure**: This course will be offered in a lecture style/format.

**Anticipated** **Pedagogical** **Strategies** **and** **Instructional** **Design**: This class will be run in a lecture-activity style/format. The class will start with a lecture, and then move on to creative in-class activities, such as using Python INotebook for creating visualizations using Python libraries.

**Providing Support to Programmatic Learning Outcomes:** This course requires satisfactory completion of individual assignments, two major exams and a final term project.

**NEW YORK CITY COLLEGE OF TECHNOLOGY/CUNY**

**Computer Systems Technology Department**

**Course Outline**

**CST 3602 – Data Visualization**

**(2 class hours; 2 lab hours; 3 credits)**

**Course Description:**

This course introduces information visualization basics and provides knowledge about design and information literacy perspective, including what makes good and bad visualization. The course covers basic visualization design techniques and evaluation principles, techniques for visualizing multivariate, temporal, text-based, geospatial, hierarchical, and network/graph-based data. The final project for the course is to acquire, parse, analyze a large dataset, and create meaningful, good visualization for this dataset.

**Final Project:**

The final project will be a team project with an oral presentation. The teams choose an on-line data collection, obtain this data collection, analyze it, and create a visualization supporting interesting facts found in this data collection.

**Course Objectives:**

Upon successful completion of this course, students should be able to:

* Demonstrate knowledge of the issues and challenges (including, ethics issues) of data / information visualization.
* Demonstrate knowledge of the fundamentals of good visualization.
* Analyze data visualization examples and reason about the quality of these visualization examples.
* Create data visualizations following the principles of good visualization.

**Software**: Python and Python Libraries

**General Education Outcomes:**

* **SKILLS/Inquiry/Analysis:** Students will employ scientific reasoning and logical thinking.
* **SKILLS/Communication:** Students will communicate in diverse settings and groups, using written (both reading and writing), oral (both speaking and listening), and visual means
* **VALUES, ETHICS, RELATIONSHIPS / Professional/Personal Development:** Students will work with teams, including those of diverse composition. Build consensus. Respect and use creativity.

**Pre-requisites:**

CST 3512 (Information and Data Management II) and MAT 2572 (Probability and Mathematical Statistics I)

**Required Text**:

Alberto Cairo. The Truthful Art: Data, Charts, and Maps for Communication

ISBN-13: 0321934075

**Additional Reading Materials:**

* Edward Tufte. The Visual Display of Quantitative Information

ISBN-13: 0961392142

* Cole Nussbaumer Knaflic. Storytelling with Data: A Data Visualization Guide for Business Professionals

ISBN-13: 978-1119002253

The instructor will identify several additional information resources during the semester including Internet resources, print material (handouts) and reference books.

**Tools and On-Line Documentation:**

* Python: https://www.python.org/
* Python Data Analysis Library: http://pandas.pydata.org/
* Python package for scientific computing: http://www.numpy.org/
* IPython command shell for interactive computing: https://ipython.org/
* Anaconda Data Science ecosystem: https://www.continuum.io/

**Grading Procedure:**

Midterm Exam 20%

Final Exam 30%

Lab Final Project 25%

Homework assignments 20%

In-class labs/quizzes 5%

=====

TOTAL 100%

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Letter Grade** | A | A- | B+ | B | B- | C+ | C | D | F |
| **Numerical Grade** | 93-100 | 90-92.9 | 87-89.9 | 83-86.9 | 80-82.9 | 77-79.9 | 70-76.9 | 60-69.9 | <=59.9 |

The grade distribution follows the information in the NYCCT Student Handbook (p.43).

**Academic Integrity**: During the course of the class you are required to follow the NYCCT Academic Integrity Standards described in the Student Handbook (pp.95 – 99)

**NYCCT Student Handbook** can be downloaded here: <http://www.citytech.cuny.edu/current-student/docs/StudentHandbook.pdf>.

**Course Outline:**

|  |  |  |
| --- | --- | --- |
| **Week** | **Topics** | **Assignments and Readings** |
| 1 | Introduction  Importance of Ethics in Visualization. Implications of (Good / Bad) Data Visualization. Examples and discussion of bad data visualization. | Alberto Cairo |
| 2, 3 | Review of Matplotlib basics (CST3502 Information and Data Management II)  Lab: Plotting, axes, legend placement | <http://pandas.pydata.org/pandas-docs/stable/whatsnew.html#restore-matplotlib-datetime-converter-registration> |
| 4 | Axes customization  Lab: axes customization for various plots and diagrams | Assignment 1 |
| 5 | Histograms (for different types of distributions)  Lab: creating histograms | <http://pandas.pydata.org/pandas-docs/stable/10min.html#histogramming> |
| 6 | Legends, annotation, and styles  Labs: customizing legends, annotations, style | Assignment 2 |
| 7,8 | Bar Charts: choosing the plotted data intervals, adding confidence intervals  Lab: creating bar charts  **Midterm** | <http://pandas.pydata.org/pandas-docs/stable/style.html#Bar-charts> |
| 9, 10 | Analyzing and plotting time series data:  line plots, dot plots, density plots, whisker plots by interval  Lab: creating charts and graphs for times series analysis | Assignment 3 |
| 11 | Heatmaps  Lab: creating heatmaps |  |
| 12 | Time series with moving images | <http://pandas.pydata.org/pandas-docs/stable/timeseries.html> |
| 13 | Lab: Case study |  |
| 14 | Lab: Working on a team project |  |
| 15 | **Review, project presentation, final exam** |  |

**ASSESSMENT CRITERIA:** For successful completion of this course the student should be able to:

|  |  |
| --- | --- |
| **For the successful completion of this course a student should be able to:** | **Evaluation methods and criteria** |
| 1. Demonstrate knowledge of the issues and challenges of data / information visualization. 2. Demonstrate knowledge of the fundamentals of good visualization. 3. Analyze data visualization examples and reason about the quality of these visualization examples. 4. Create data visualizations following the principles of good visualization. | Exams, homework assignments, in-class lab exercises and project |

**GENERAL EDUCATION LEARNING OUTCOMES/ASSESSMENT METHODS**

|  |  |
| --- | --- |
| **LEARNING OUTCOMES** | **ASSESSMENT METHODS** |
| **1.** Demonstrate the ability to work collaboratively and independently on assignments in and outside a classroom setting. | **1.** Classroom discussions, group assignments and individual oral presentations. |
| **2.** Understand and employ both quantitative and qualitative analysis to solve problems**.** | **2.** Classroom discussion, group activities, group presentations, quizzes, tests, final exam. |
| **3.** Develop reading, writing competencies, and listening skills. | **3.** Biweekly reading and writing assignments, individual and group presentation, classroom discussion. Each homework assignment requires writing. |
| **4.** Work with teams. Build consensus. Use creativity. | **4.** Group projects and presentations. |

**Bibliography:**

* Alberto Cairo. *The Truthful Art: Data, Charts, and Maps for Communication*

Publisher: New Riders; 1 edition (February 28, 2016)

ISBN-13: 0321934075

* Edward Tufte (2001). *The Visual Display of Quantitative Information*

Publisher: Graphics Press; 2nd edition edition (May 1, 2001)

ISBN-13: 0961392142

* Cole Nussbaumer Knaflic (2015). *Storytelling with Data: A Data Visualization Guide for Business Professionals*

Publisher: Wiley

ISBN-13: 978-1119002253

**CHANCELLOR’S REPORT FORM**

Department: Computer Systems Technology

Date: January 4th, 2018

**NEW COURSE PROPOSAL**

**Section AIV: New Course**

|  |  |
| --- | --- |
| **Department(s)** | **Computer Systems Technology** |
| **Academic Level** | **[X] Regular [  ] Compensatory [  ] Developmental [  ] Remedial** |
| **Subject Area** | Data Science |
| **Course Prefix** | **CST** |
| **Course Number** | **CST 3602** |
| **Course Title** | **Data Visualization** |
| **Catalog Description** | Basic visualization design techniques and evaluation principles, techniques for visualizing multivariate, temporal, text-based, geospatial, hierarchical, and network/graph-based data. The final project for the course is to acquire, parse, analyze a large dataset, and create meaningful, good visualization for this dataset. |
| **Prerequisite** | **CST 3512 and MAT 2572** |
| **Corequisite** | **None** |
| **Credits** | **3** |
| **Contact Hours** | **2 Class Hours, 2 Lab Hours** |
| **Liberal Arts** | **[ ] Yes [ X  ] No** |
| **Course Attribute (e.g. Writing Intensive, etc)** | **It is not a writing intensive course** |
| **Course Applicability** | **[X] Major**  **[ ] Gen Ed Required [ ] Gen Ed - Flexible [ ] Gen Ed - College Option**  **[ ] English Composition [ ] World Cultures [ ] Speech**  **[ ] Mathematics [ ] US Experience in its Diversity**  **[ ] Interdisciplinary [ ] Science [ ] Creative Expression**  **[ ] Advanced Liberal Arts**  **[ ] Individual and Society**  **[ ] Scientific World** |
| **Effective Term** | **Fall 2019** |

**Rationale:** Data Visualization allows to combat the information overload in the modern world. A well-designed visual data rendering can improve the data comprehension, information interpretation, and decision-making process. Also, this course is integral to the proposed Bachelors in Data Science curriculum program of the Computer Systems Technology department.

**LIBRARY RESOURCES & INFORMATION LITERACY: MAJOR CURRICULUM MODIFICATION**

Please complete for **all** major curriculum modifications. This information will assist the library in planning for new courses/programs.

Consult with your library faculty subject specialist (<http://cityte.ch/dir>) **3 weeks before the proposal deadline**.

**Course proposer:** please complete boxes 1-4. **Library faculty subject specialist:** please complete box 5.

|  |  |  |
| --- | --- | --- |
| **1** | **Title of proposal**  CST 3602: Data Visualization | **Department/Program**  Computer and Systems Technology |
|  | **Proposed by** (include email & phone)  Elena Filatova  efilatova@citytech.cuny.edu | **Expected date course(s) will be offered**  Fall 2019  **# of students** 24 |

|  |  |
| --- | --- |
| **2** | **The library cannot purchase reserve textbooks for every course at the college, nor copies for all students. Consult our website (**[**http://cityte.ch/curriculum**](http://cityte.ch/curriculum)**) for articles and ebooks for your courses, or our open educational resources (OER) guide (**[**http://cityte.ch/oer**](http://cityte.ch/oer)**). Have you considered using a freely-available OER or an open textbook in this course?**  One of the books is a free on-line book, the other book is not free:  Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython by Wes McKinney, 2nd ed.  ISBN-10: 1449319793 |

|  |  |
| --- | --- |
| **3** | **Beyond the required course materials, are City Tech library resources sufficient for course assignments? If additional resources are needed, please provide format details (e.g. ebook, journal, DVD, etc.), full citation (author, title, publisher, edition, date), price, and product link.**  Tools and On-Line Documentation (free, on-line)  Python: https://www.python.org/  Python Data Analysis Library: http://pandas.pydata.org/  Python package for scientific computing: http://www.numpy.org/  IPython command shell for interactive computing: https://ipython.org/  Anaconda Data Science ecosystem: https://www.continuum.io/ |

|  |  |
| --- | --- |
| **4** | **Library faculty focus on strengthening students' information literacy skills in finding, critically evaluating, and ethically using information. We collaborate on developing assignments and customized instruction and research guides. When this course is offered, how do you plan to consult with the library faculty subject specialist for your area? Please elaborate.**  The project for the course concerns an on-line data set. There will be a discussion in class, what data can be downloaded and used for the project, whether there are any ethical consequences for using this data set and making conclusions about this data set. |

|  |  |
| --- | --- |
| **5** | **Library Faculty Subject Specialist Prof. Junior Tidal**  **Comments and Recommendations**  I believe that the library can support this course. We have several monographs related to Python. It may be necessary to look into non-print materials, such as software, to further augment the course.  **Date 12.13.17** |

# New Course Proposal #6: CST4702 Machine Learning Fundamentals

New York City College of Technology, CUNY

CURRICULUM MODIFICATION PROPOSAL FORM

This form is used for all curriculum modification proposals. See the [Proposal Classification Chart](http://openlab.citytech.cuny.edu/collegecouncil/files/2014/08/2013-10-09-Proposal_Classification_Chart.pdf) for information about what types of modifications are major or minor. Completed proposals should be emailed to the Curriculum Committee chair.

|  |  |
| --- | --- |
| **Title of Proposal** | **CST4702 Machine Learning Fundamentals** |
| **Date** | **Dec 5th, 2017** |
| **Major or Minor** | **Major** |
| **Proposer’s Name** | **Dr. Ashwin Satyanarayana** |
| **Department** | **Computer Systems Technology** |
| **Date of Departmental Meeting in which proposal was approved** | **12/8/2017** |
| **Department Chair Name** | **Dr. Hong Li** |
| **Department Chair Signature and Date** |  |
| **Academic Dean Name** | **Dean Hom** |
| **Academic Dean Signature and Date** | 2/22/18 |
| **Brief Description of Proposal**  (Describe the modifications contained within this proposal in a succinct summary. More detailed content will be provided in the proposal body. | Title: Machine Learning Fundamentals  This proposal explains the rationale for developing Machine Learning Fundamentals as a required course in the Data Science curriculum of the Computer Systems Technology Department. A course outline and sample syllabus are attached to the proposal. |
| **Brief Rationale for Proposal**  (Provide a concise summary of why this proposed change is important to the department. More detailed content will be provided in the proposal body). | Machine Learning is concerned with computer programs that automatically improve their performance through experience (e.g., programs that learn to recognize human faces, recommend music and movies, and drive autonomous robots). This course covers the theory and practical algorithms for machine learning from a variety of perspectives. Also, this course is integral to the proposed Bachelors in Data Science curriculum program of the Computer Systems Technology department. |
| **Proposal History**  (Please provide history of this proposal: is this a resubmission? An updated version? This may most easily be expressed as a list). | This is a first submission. |
| **CUNY – Course Equivalencies**  Provide information about equivalent courses within CUNY, if any. | None |
| **Intent to Submit as Common Core**  If this course is intended to fulfill one of the requirements in the common core, then indicate which area. | N/A |

Please include all appropriate documentation as indicated in the Curriculum Modification Checklist.

For each new course, please also complete the New Course Proposal and submit in this document.

Please submit this document as a single .doc or .rtf format. If some documents are unable to be converted to .doc, then please provide all documents archived into a single .zip file.

**ALL PROPOSAL CHECK LIST**

|  |  |
| --- | --- |
| Completed CURRICULUM MODIFICATION FORM including: |  |
| * Brief description of proposal | X |
| * Rationale for proposal | X |
| * Date of department meeting approving the modification | X |
| * Chair’s Signature |  |
| * Dean’s Signature |  |
| Evidence of consultation with affected departments  List of the programs that use this course as required or elective, and courses that use this as a prerequisite. | X |
| Documentation of Advisory Commission views (if applicable). | N/A |
| Completed [Chancellor’s Report Form](http://openlab.citytech.cuny.edu/collegecouncil/files/2014/08/2013-10-09-Chancellor_Report_Quick_Reference_Guide1.doc). | X |

**EXISTING PROGRAM MODIFICATION PROPOSALS**

|  |  |
| --- | --- |
| Documentation indicating core curriculum requirements have been met for new programs/options or program changes. | X |
| Detailed rationale for each modification (this includes minor modifications) | N/A |

New York City College of Technology, CUNY

NEW COURSE PROPOSAL FORM

This form is used for all new course proposals. Attach this to the [Curriculum Modification Proposal Form](http://openlab.citytech.cuny.edu/collegecouncil/files/2014/08/2013-10-10-Curriculum_Modification_Proposal_Form.docx) and submit as one package as per instructions. Use one New Course Proposal Form for each new course.

|  |  |
| --- | --- |
| **Course Title** | Machine Learning Fundamentals |
| **Proposal Date** | Dec 5th, 2017 |
| **Proposer’s Name** | Ashwin Satyanarayana |
| **Course Number** | CST 4702 |
| **Course Credits, Hours** | 3 credits, 2 lecture hours and 2 lab hours |
| **Course Pre / Co-Requisites** | CST3650: Data Structures, CST3502 Data Mining and MAT2572: Statistics and Probability |
| **Catalog Course Description** | Introduces fundamental machine learning algorithms and techniques, and their implementations applied to solving real world problems. Topics include supervised learning (parametric/non-parametric algorithms, support vector machines, kernels, neural networks), unsupervised learning (clustering, dimensionality reduction, recommender systems, deep learning) and best practices in machine learning (bias/variance theory; innovation process in machine learning). The theory of machine learning and the practical know-how lead to numerous case studies with applications in text understanding (web search, anti-spam), medical informatics, audio, database mining, and other areas |
| **Brief Rationale**  Provide a concise summary of why this course is important to the department, school or college. | Machine Learning is concerned with computer programs that automatically improve their performance through experience (e.g., programs that learn to recognize human faces, drive autonomous robots, etc). This course covers the theory and practical algorithms for machine learning from a variety of perspectives, which is a core fundamental tool for data scientists, researchers and engineers of today’s world. |
| **CUNY – Course Equivalencies**  Provide information about equivalent courses within CUNY, if any. | N/A |
| **Intent to Submit as Common Core**  If this course is intended to fulfill one of the requirements in the common core, then indicate which area. | No. |
| **Intent to Submit as An Interdisciplinary Course** | No. |
| **Intent to Submit as a Writing Intensive Course** | **Yes** |

Please include all appropriate documentation as indicated in the NEW COURSE PROPOSAL Combine all information into a single document that is included in the Curriculum Modification Form.

**NEW COURSE PROPOSAL CHECK LIST**

Use this checklist to ensure that all required documentation has been included. You may wish to use this checklist as a table of contents within the new course proposal.

|  |  |
| --- | --- |
| **Completed NEW COURSE PROPOSAL FORM** |  |
| * Title, Number, Credits, Hours, Catalog course description | X |
| * Brief Rationale | X |
| * CUNY – Course Equivalencies | X |
| Completed [Library Resources and Information Literacy Form](http://openlab.citytech.cuny.edu/collegecouncil/files/2014/08/curriculum_modification_library_form.doc) | X |
| **Course Outline**  Include within the outline the following. | X |
| Hours and Credits for Lecture and Labs  If hours exceed mandated Carnegie Hours, then rationale for this | X |
| Prerequisites/Co- requisites | X |
| Detailed Course Description | X |
| Course Specific Learning Outcome and Assessment Tables   * Discipline Specific * General Education Specific Learning Outcome and Assessment Tables | X |
| Example Weekly Course outline | X |
| Grade Policy and Procedure | X |
| Recommended Instructional Materials (Textbooks, lab supplies, etc) | X |
| Library resources and bibliography | X |
| **Course Need Assessment.**  Describe the need for this course. Include in your statement the following information. | X |
| Target Students who will take this course. Which programs or departments, and how many anticipated?  Documentation of student views (if applicable, e.g. non-required elective). | X |
| Projected headcounts (fall/spring and day/evening) for each new or modified course. | X |
| If additional physical resources are required (new space, modifications, equipment), description of these requirements. If applicable, Memo or email from the VP for Finance and Administration with written comments regarding additional and/or new facilities, renovations or construction. | X |
| Where does this course overlap with other courses, both within and outside of the department? | X |
| Does the Department currently have full time faculty qualified to teach this course? If not, then what plans are there to cover this? | X |
| If needs assessment states that this course is required by an accrediting body, then provide documentation indicating that need. | N/A |
| **Course Design**  Describe how this course is designed. | X |
| Course Context (e.g. required, elective, capstone) | X |
| Course Structure: how the course will be offered (e.g. lecture, seminar, tutorial, fieldtrip)? | X |
| Anticipated pedagogical strategies and instructional design (e.g. Group Work, Case Study, Team Project, Lecture) | X |
| How does this course support Programmatic Learning Outcomes? | X |
| Is this course designed to be partially or fully online? If so, describe how this benefits students and/or program. | X |
| **Additional Forms for Specific Course Categories** |  |
| [Interdisciplinary Form](http://openlab.citytech.cuny.edu/collegecouncil/files/2014/08/Application-for-Interdisciplinary-Course-Designation.docx) (if applicable) | N/A |
| Interdisciplinary Committee Recommendation (if applicable and if received)\*  \*Recommendation must be received before consideration by full Curriculum Committee | N/A |
| [Common Core (Liberal Arts) Intent to Submit](http://openlab.citytech.cuny.edu/collegecouncil/files/2014/08/CommonCoreCourseSubmissionForm_4.2.12.doc) (if applicable) | N/A |
| Writing Intensive Form if course is intended to be a WIC (under development) | N/A |
| If course originated as an experimental course, then results of evaluation plan as developed with director of assessment. | N/A |
| **(Additional materials for** [**Curricular Experiments**](http://www.300jaystreet.com/college-council/curriculum_proposals/curricular-experiments)**)** |  |
| Plan and process for evaluation developed in consultation with the director of assessment. (Contact Director of Assessment for more information). |  |
| Established Timeline for Curricular Experiment |  |

**Proposed Course Name: Machine Learning Fundamentals**

**Course Overview & Rationale**

Machine Learning is concerned with computer programs that automatically improve their performance through experience (e.g., programs that learn to recognize human faces, recommend music and movies, and drive autonomous robots). This course covers the theory and practical algorithms for machine learning from a variety of perspectives. Also, this course is integral to the proposed Bachelors in Data Science curriculum program of the Computer Systems Technology department.

This course covers fundamental machine learning algorithms and techniques, and their implementations applied to solving real world problems. Topics include Supervised learning (parametric/non-parametric algorithms, support vector machines, kernels, neural networks), unsupervised learning (clustering, dimensionality reduction, recommender systems, deep learning) and best practices in machine learning (bias/variance theory; innovation process in machine learning). The theory of machine learning and the practical know-how needed to quickly and powerfully apply these techniques will be presented. The course will draw from numerous case studies and applications with applications in text understanding (web search, anti-spam), medical informatics, audio, database mining, and other areas. This course also includes a project, which may either involve research in machine learning or may address a practical problem using the methods and tools introduced in the class. It is vital to emphasize the teaching and learning of these crucial concepts and incorporate them into the practical training of Machine Learning algorithms in and out of the classroom. The development of this course lays the foundation and adds tremendous value to the creation and further progression of Data Science as a new, fast-growing, and promising major in the Computer Systems Technology department at New York City College of Technology.

**Course Need**

**Students who would take this class:** students who intend to major in Data Science

**Department**: Computer Systems Technology

**Program**: Bachelors in Data Science

**The number of section (s) anticipated:** one section for the first year

**Projected headcount: 24** students

**Physical Resources required:** Basicsmartroomset**-**up**:** a screen, and an overhead projector/a TV set that is run by and connected to a computer

**Course overlap:** None

**Faculty** **qualified** **for** **teaching** **this** **course**: Yes, there are faculty members who have doctoral degrees in Data Mining and have taught this class before in other institutions.

**Course design**

**Course context:** This course will be required of Data Science major students. Students are required to develop an independent project at the end of the semester.

**Course** **structure**: This course will be offered in a lecture style/format.

**Anticipated** **Pedagogical** **Strategies** **and** **Instructional** **Design**: This class will be run in a lecture-activity style/format. The class will start with a lecture, and then move on to creative in-class activities, such as using existing Machine Learning tools, analyzing large datasets, and applying algorithms on the data.

**Providing Support to Programmatic Learning Outcomes:** This course requires satisfactory completion of individual assignments, two major exams and a final term project.

**NEW YORK CITY COLLEGE OF TECHNOLOGY/CUNY**

**Computer Systems Technology Department**

**CST 4702 – Machine Learning Fundamentals**

**(2 class hours; 2 lab hours; 3 credits)**

**Course Description:**

This course introduces fundamental machine learning algorithms and techniques, and their implementations applied to solving real world problems. Topics covered include Supervised learning (parametric/non-parametric algorithms, support vector machines, kernels, neural networks), unsupervised learning (clustering, dimensionality reduction, recommender systems, deep learning) and best practices in machine learning (bias/variance theory; innovation process in machine learning). The theory of machine learning and the practical know-how lead to numerous case studies with applications in text understanding (web search, anti-spam), medical informatics, audio, database mining, and other areas.

**Learning Outcomes:**

Upon successful completion of this course, students should be able to:

* Demonstrate an understanding of univariate and multivariate linear regression and apply it to machine learning.
* Use tools which are important to implement machine learning solutions like Octave and/or Matlab
* Demonstrate an understanding of logistic regression and apply it to machine learning.
* Demonstrate an understanding of the differences between supervised and unsupervised learning.
* Demonstrate an understanding of the structure of neural networks and their use to solve problems
* Demonstrate an understanding of dimensionality reduction using principal components analysis
* Demonstrate an understanding of anomaly detection techniques.
* Demonstrate an understanding if clustering analysis techniques.

**General Education Outcomes:**

* **SKILLS/Inquiry/Analysis:** Students will employ scientific reasoning and logical thinking.
* **SKILLS/Communication:** Students will communicate in diverse settings and groups, using written (both reading and writing), oral (both speaking and listening), and visual means
* **VALUES, ETHICS, RELATIONSHIPS / Professional/Personal Development:** Students will work with teams, including those of diverse composition. Build consensus. Respect and use creativity.

**Pre-requisites:**

CST3650: Data Structures, CST3502 Data Mining and MAT2572: Statistics and Probability I

**Required Text**: Ethem Alpaydin, *Introduction to Machine Learning, Third Edition*  **ISBN-10:** 8120350782, **ISBN-13:** 978-8120350786

**Recommended** **Texts**:

* Stephen Marsland, *Machine Learning: An Algorithmic Perspective*
* Christopher M. Bishop, *Pattern Recognition and Machine Learning*
* Tom Mitchell, *Machine Learning*

**Academic Integrity –** Students and all others who work with information, ideas, texts, images, music, inventions, and other intellectual property owe their audience and sources accuracy and honesty in using, crediting, and citing sources. As a community of intellectual and professional workers, the College recognizes its responsibility for providing instruction in information literacy and academic integrity, offering models of good practice, and responding vigilantly and appropriately to infractions of academic integrity. Accordingly, academic dishonesty is prohibited in The City University of New York and at New York City College of Technology and is punishable by penalties, including failing grades, suspension, and expulsion. The complete text of the College policy on Academic Integrity may be found in the catalog.

**Grading Procedure:**

Test1 15%

Test2 15%

Final Exam 25%

Midterm 20%

Projects 15%

Homeworks 10%

=====

TOTAL 100%

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Letter Grade** | A | A- | B+ | B | B- | C+ | C | D | F |
| **Numerical Grade** | 93-100 | 90-92.9 | 87-89.9 | 83-86.9 | 80-82.9 | 77-79.9 | 70-76.9 | 60-69.9 | <=59.9 |

**Course Outline:**

|  |  |  |
| --- | --- | --- |
| **Week** | **Topics** | **Assignments and Readings** |
| Week 1 | Introduction | Reading: CH 1  HW: CH 1.1,6,7,8,9 |
| Supervised Learning | Reading: CH 2  HW: CH 2.1, 2, 4, 5, 7, 8 |
| Week 2 | Bayesian Decision Theory  Nonparametric Methods | Reading: CH 3.1-3.5, 8.1-8.5  HW: CH 3.1, 2, 5; 8.2, 3, 4 |
| Decision Trees | Reading: CH 9, 19.1, 19.5-19.7  HW: CH 9.1, 2, 4  Assignment: Project1 |
| Week 3 | Linear Discrimination  Multilayer Perceptrons | Reading: CH 10, 11.1-11.8.2  HW: 10.1,7-9; 11.1-3 |
| Week 4 | Exam 1 | Reading: CH 1-4, 8-11  Due: Project1 |
| Exam 1 and Project 1 Review |  |
| Week 5 | Parametric & Multivariate | Reading: CH 4, 5 |
| Dimensionality Reduction | Reading: CH 6, PCA Primer |
| Week 6  Week 7 | Clustering | Reading: CH 7 |
| Kernel Machines | Reading: CH 13  Assignment: Project 2 |
| Week 8 | Combining Learners | Reading: CH 17 |
|  |  |
| Week 9 | Exam 2 Review Session |  |
| Exam 2 | Reading: CH 4-7, 12-13, 17 |
| Week 10 | Reinforcement Learning | Reading: CH 18.1-18.4  Due: Project 2 |
| Week 11 | Reinforcement Learning | Reading: CH 18 all  Assignment: Project 3 |
| Week 12  Week 13 | Hidden Markov Models | Reading: CH 15 |
| Course Review | Due: Project3 |
| Week 14  Week 15 | Research Paper Presentations  Finals | Reading: CH 15-16, 18 |

**ASSESSMENT CRITERIA:**

**COURSE INTENDED LEARNING OUTCOMES/ASSESSMENT METHODS**

|  |  |
| --- | --- |
| **LEARNING OUTCOMES** | **ASSESSMENT METHODS** |
| 1. Understand Bayesian Decision Theory and Nonparametric Methods | **1.** Group activities, written report. |
| 1. Implement a supervised learning technique such as Decision trees on real world dataset | **2.** Individual oral presentations, in‐class group activities. |
| 1. Understand Linear Discrimination and Multilayer Perceptrons | **3.** Classroom discussion, in‐class estimation group assignments. |
| 1. Understand Reinforcement Learning and Hidden Markov Models | **4.** Classroom discussion, in‐class group assignments, learning logs. |
| 1. Demonstrate an understanding of dimensionality reduction using principal components analysis | **5.** Individual short essay related to functions. |
| 1. Demonstrate an understanding of the structure of neural networks and their use to solve problems | **6.** Written report and group presentation, learning logs. |
| 1. Demonstrate an ability to write research articles which includes sections the following sections: Introduction, Related Work, Methodology, Empirical evidence and Conclusions | 1. Written research paper |

**GENERAL EDUCATION LEARNING OUTCOMES/ASSESSMENT METHODS**

|  |  |
| --- | --- |
| **LEARNING OUTCOMES** | **ASSESSMENT METHODS** |
| **1.** Demonstrate the ability to work collaboratively and independently on assignments in and outside a classroom setting. | **1.** Classroom discussions, group assignments and individual oral presentations. |
| **2.** Understand and employ both quantitative and qualitative analysis to solve problems**.** | **2.** Classroom discussion, group activities, group presentations, quizzes, tests, final exam. |
| **3.** Develop reading, writing competencies, and listening skills. | **3.** Biweekly reading and writing assignments, individual and group presentation, classroom discussion. Each homework assignment requires written |
| **4.** Work with teams. Build consensus. Use creativity. | **4.** Group projects and presentations. |

**Bibliography:**

* Hastie, Friedman, and Tibshirani, *The Elements of Statistical Learning*, 2001
* Bishop, *Pattern Recognition and Machine Learning*, 2006
* Ripley, *Pattern Recognition and Neural Networks*, 1996
* Duda, Hart, and Stork, *Pattern Classification*, 2nd Ed., 2002
* Tan, Steinbach, and Kumar, [Introduction to Data Mining](http://www-users.cs.umn.edu/~kumar/dmbook/index.php), Addison-Wesley, 2005.
* Scholkopf and Smola, *Learning with Kernels*, 2002
* Mardia, Kent, and Bibby, *Multivariate Analysis*, 1979
* [Computational Statistics](http://www.quantlet.org/mdstat/scripts/csa/html/csahtml.html) (online book)
* Sutton and Barto, *Reinforcement Learning: An Introduction*, MIT Press, 1998.
* Bertsekas and Tsitsiklis, *Neuro-Dynamic Programming*, Athena Scientific, 1996.

**CHANCELLOR’S REPORT FORM**

Department: Computer Systems Technology

Date: January 1st, 2018

**NEW COURSE PROPOSAL**

**Section AIV: New Course**

|  |  |
| --- | --- |
| **Department(s)** | **Computer Systems Technology** |
| **Academic Level** | **[X] Regular [  ] Compensatory [  ] Developmental [  ] Remedial** |
| **Subject Area** | Data Science |
| **Course Prefix** | **CST** |
| **Course Number** | **CST 4702** |
| **Course Title** | **Machine Learning Fundamentals** |
| **Catalog Description** | Introduces fundamental machine learning algorithms and techniques, and their implementations applied to solving real world problems. Topics include supervised learning (parametric/non-parametric algorithms, support vector machines, kernels, neural networks), unsupervised learning (clustering, dimensionality reduction, recommender systems, deep learning) and best practices in machine learning (bias/variance theory; innovation process in machine learning). The theory of machine learning and the practical know-how lead to numerous case studies with applications in text understanding (web search, anti-spam), medical informatics, audio, database mining, and other areas. |
| **Prerequisite** | **CST3650: Data Structures, CST3502 Data Mining and MAT2572: Statistics and Probability I** |
| **Corequisite** | **None** |
| **Credits** | **3** |
| **Contact Hours** | **2 Class Hours, 2 Lab Hours** |
| **Liberal Arts** | **[ ] Yes [ X  ] No** |
| **Course Attribute (e.g. Writing Intensive, etc)** | **It is a writing intensive course** |
| **Course Applicability** | **[X] Major**  **[ ] Gen Ed Required [ ] Gen Ed - Flexible [ ] Gen Ed - College Option**  **[ ] English Composition [ ] World Cultures [ ] Speech**  **[ ] Mathematics [ ] US Experience in its Diversity**  **[ ] Interdisciplinary [ ] Science [ ] Creative Expression**  **[ ] Advanced Liberal Arts**  **[ ] Individual and Society**  **[ ] Scientific World** |
| **Effective Term** | **Fall 2019** |

**Rationale:** Machine Learning is concerned with computer programs that automatically improve their performance through experience (e.g., programs that learn to recognize human faces, recommend music and movies, and drive autonomous robots). This course covers the theory and practical algorithms for machine learning from a variety of perspectives. Also, this course is integral to the proposed Bachelors in Data Science curriculum program of the Computer Systems Technology department.

**LIBRARY RESOURCES & INFORMATION LITERACY: MAJOR CURRICULUM MODIFICATION**

Please complete for **all** major curriculum modifications. This information will assist the library in planning for new courses/programs.

Consult with your library faculty subject specialist (<http://cityte.ch/dir>) **3 weeks before the proposal deadline**.

**Course proposer:** please complete boxes 1-4. **Library faculty subject specialist:** please complete box 5.

|  |  |  |
| --- | --- | --- |
| **1** | **Title of proposal**  CST 4702: Machine Learning Fundamentals | **Department/Program**  Computer Systems Technology / BS in Data Science |
|  | **Proposed by** (include email & phone)  Dr. Ashwin Satyanarayana  [asatyanarayana@citytech.cuny.edu](mailto:asatyanarayana@citytech.cuny.edu)  518-330-7907 | **Expected date course(s) will be offered**  Fall 2019  **# of students:** 24 |

|  |  |
| --- | --- |
| **2** | **The library cannot purchase reserve textbooks for every course at the college, nor copies for all students. Consult our website (**[**http://cityte.ch/curriculum**](http://cityte.ch/curriculum)**) for articles and ebooks for your courses, or our open educational resources (OER) guide (**[**http://cityte.ch/oer**](http://cityte.ch/oer)**). Have you considered using a freely-available OER or an open textbook in this course?**  No, I have not considered freely available OER or an open textbook for this course. |

|  |  |
| --- | --- |
| **3** | **Beyond the required course materials, are City Tech library resources sufficient for course assignments? If additional resources are needed, please provide format details (e.g. ebook, journal, DVD, etc.), full citation (author, title, publisher, edition, date), price, and product link.**  Yes, City Tech Library resources are sufficient for the proposed course assignments because the main readings for the course are a required textbook and journal articles that will be assigned by instructor. Students should be able to locate the selected journal articles in library. |

|  |  |
| --- | --- |
| **4** | **Library faculty focus on strengthening students' information literacy skills in finding, critically evaluating, and ethically using information. We collaborate on developing assignments and customized instruction and research guides. When this course is offered, how do you plan to consult with the library faculty subject specialist for your area? Please elaborate.**  Yes, a library faculty will be invited for a guest lecture on how to find, and use information in conducting original research, which is a major portion of the proposed course assignments. |

|  |  |
| --- | --- |
| **5** | **Library Faculty Subject Specialist Prof. Junior Tidal**  **Comments and Recommendations**  After surveying the collection, I highly recommend that the library acquire more books on machine learning upon approval of this course. The collection has a few books on machine learning, but they are several years old. I also recommend additional books on statistics to properly support this course.  **Date 12.13.17** |

# New Course Proposal #7: CST4802 Information Retrieval

New York City College of Technology, CUNY

CURRICULUM MODIFICATION PROPOSAL FORM

This form is used for all curriculum modification proposals. See the [Proposal Classification Chart](http://openlab.citytech.cuny.edu/collegecouncil/files/2014/08/2013-10-09-Proposal_Classification_Chart.pdf) for information about what types of modifications are major or minor. Completed proposals should be emailed to the Curriculum Committee chair.

|  |  |
| --- | --- |
| **Title of Proposal** | CST4802 Information Retrieval |
| **Date** | December 5, 2017 |
| **Major or Minor** | Major |
| **Proposer’s Name** | Dr. Elizabeth Milonas |
| **Department** | Computer Systems Technology |
| **Date of Departmental Meeting in which proposal was approved** | December 8, 2017 |
| **Department Chair Name** | **Dr. Hong Li** |
| **Department Chair Signature and Date** |  |
| **Academic Dean Name** | Dean Kevin Hom |
| **Academic Dean Signature and Date** | 2/22/18 |
| **Brief Description of Proposal**  (Describe the modifications contained within this proposal in a succinct summary. More detailed content will be provided in the proposal body. | Title: Information Retrieval  This proposal explains the rationale for developing the Information Retrieval course as a required course in the Data Science curriculum of the Computer Systems Technology Department. A course outline and sample syllabus are attached to the proposal. |
| **Brief Rationale for Proposal**  (Provide a concise summary of why this proposed change is important to the department. More detailed content will be provided in the proposal body). | The CST4802 Information Retrieval course provides students with hands-on experience in the design, implementation and use of techniques and tools related to information retrieval which is an integral component of the Data Science domain. In addition, ethical issues including personal privacy, bias, informed consent, monitoring and surveillance are examined. Technical and ethical concepts, techniques and tools covered in this course are integral in the process of data acquisition and presentation. |
| **Proposal History**  (Please provide history of this proposal: is this a resubmission? An updated version? This may most easily be expressed as a list). | This is a first submission |
| **CUNY – Course Equivalencies**  Provide information about equivalent courses within CUNY, if any. | None |
| **Intent to Submit as Common Core**  If this course is intended to fulfill one of the requirements in the common core, then indicate which area. | N/A |

Please include all appropriate documentation as indicated in the NEW COURSE PROPOSAL Combine all information into a single document that is included in the Curriculum Modification Form.

Please submit this document as a single .doc or .rtf format. If some documents are unable to be converted to .doc then please provide all documents archived into a single .zip file.

**ALL PROPOSAL CHECKLIST**

|  |  |
| --- | --- |
| Completed CURRICULUM MODIFICATION FORM including: |  |
| * Brief description of proposal | X |
| * Rationale for proposal | X |
| * Date of department meeting approving the modification | X |
| * Chair’s Signature |  |
| * Dean’s Signature |  |
| Evidence of consultation with affected departments  List of the programs that use this course as required or elective, and courses that use this as a prerequisite. | X |
| Documentation of Advisory Commission views (if applicable). | N/A |
| Completed [Chancellor’s Report Form](http://openlab.citytech.cuny.edu/collegecouncil/files/2014/08/2013-10-09-Chancellor_Report_Quick_Reference_Guide1.doc). | X |

**EXISTING PROGRAM MODIFICATION PROPOSALS**

|  |  |
| --- | --- |
| Documentation indicating core curriculum requirements have been met for new programs/options or program changes. | X |
| Detailed rationale for each modification (this includes minor modifications) | N/A |

New York City College of Technology, CUNY

NEW COURSE PROPOSAL FORM

This form is used for all new course proposals. Attach this to the [Curriculum Modification Proposal Form](http://openlab.citytech.cuny.edu/collegecouncil/files/2014/08/2013-10-10-Curriculum_Modification_Proposal_Form.docx) and submit as one package as per instructions. Use one New Course Proposal Form for each new course.

|  |  |
| --- | --- |
| **Course Title** | Information Retrieval |
| **Proposal Date** | December 5, 2017 |
| **Proposer’s Name** | Dr. Elizabeth Milonas |
| **Course Number** | CST4802 |
| **Course Credits, Hours** | 3 credits, ( 2 lecture hours and 2 lab hours ) |
| **Course Pre / Co-Requisites** | CST4702 Machine Learning Fundamentals |
| **Catalog Course Description** | The course details basic and advanced concepts and principles of information retrieval including: definition, framework, uses and ethical issues. Issues related to interface, relevance, language processing, and indexing are discussed. Information retrieval techniques covering both effectiveness and run-time performance are covered as well as heuristics and algorithms used in information retrieval systems. Information retrieval mechanism used in both Websites and Web search engines are also discussed. Ethical issues including personal privacy, bias, informed consent, monitoring and surveillance are also examined and discussed. |
| **Brief Rationale**  Provide a concise summary of why this course is important to the department, school or college. | The CST4802 Information Retrieval course provides students with hands-on experience in the design, implementation and use of techniques and tools related to information retrieval which is an integral component of the Data Science domain. In addition, ethical issues including personal privacy, bias, informed consent, monitoring and surveillance are examined. Technical and ethical concepts, techniques and tools covered in this course are integral in the process of data acquisition and presentation. |
| **CUNY – Course Equivalencies**  Provide information about equivalent courses within CUNY, if any. | N/A |
| **Intent to Submit as Common Core**  If this course is intended to fulfill one of the requirements in the common core, then indicate which area. | No |
| **For Interdisciplinary Courses:**   * Date submitted to ID Committee for review * Date ID recommendation received   - Will all sections be offered as ID? Y/N | No |
|  |
|  |
| **Intent to Submit as a Writing Intensive Course** | No |

Please include all appropriate documentation as indicated in the NEW COURSE PROPOSAL Combine all information into a single document that is included in the Curriculum Modification Form.

**NEW COURSE PROPOSAL CHECK LIST**

Use this checklist to ensure that all required documentation has been included. You may wish to use this checklist as a table of contents within the new course proposal.

|  |  |
| --- | --- |
| **Completed NEW COURSE PROPOSAL FORM** |  |
| * Title, Number, Credits, Hours, Catalog course description | X |
| * Brief Rationale | X |
| * CUNY – Course Equivalencies | X |
| Completed [Library Resources and Information Literacy Form](http://openlab.citytech.cuny.edu/collegecouncil/files/2014/08/curriculum_modification_library_form.doc) | X |
| **Course Outline**  Include within the outline the following. | **X** |
| Hours and Credits for Lecture and Labs  If hours exceed mandated Carnegie Hours, then rationale for this | X |
| Prerequisites/Co- requisites | X |
| Detailed Course Description | X |
| Course Specific Learning Outcome and Assessment Tables   * Discipline Specific * General Education Specific Learning Outcome and Assessment Tables | X |
| Example Weekly Course outline | X |
| Grade Policy and Procedure | X |
| Recommended Instructional Materials (Textbooks, lab supplies, etc) | X |
| Library resources and bibliography | X |
| **Course Need Assessment.**  Describe the need for this course. Include in your statement the following information. | X |
| Target Students who will take this course. Which programs or departments, and how many anticipated?  Documentation of student views (if applicable, e.g. non-required elective). | X |
| Projected headcounts (fall/spring and day/evening) for each new or modified course. | X |
| If additional physical resources are required (new space, modifications, equipment), description of these requirements. If applicable, Memo or email from the VP for Finance and Administration with written comments regarding additional and/or new facilities, renovations or construction. | X |
| Where does this course overlap with other courses, both within and outside of the department? | X |
| Does the Department currently have full time faculty qualified to teach this course? If not, then what plans are there to cover this? | X |
| If needs assessment states that this course is required by an accrediting body, then provide documentation indicating that need. | N/A |
| **Course Design**  Describe how this course is designed. | X |
| Course Context (e.g. required, elective, capstone) | X |
| Course Structure: how the course will be offered (e.g. lecture, seminar, tutorial, fieldtrip)? | X |
| Anticipated pedagogical strategies and instructional design (e.g. Group Work, Case Study, Team Project, Lecture) | X |
| How does this course support Programmatic Learning Outcomes? | X |
| Is this course designed to be partially or fully online? If so, describe how this benefits students and/or program. | X |
| **Additional Forms for Specific Course Categories** | N/A |
| [Interdisciplinary Form](http://openlab.citytech.cuny.edu/collegecouncil/files/2014/08/Application-for-Interdisciplinary-Course-Designation.docx) (if applicable) | N/A |
| Interdisciplinary Committee Recommendation (if applicable and if received)\*  \*Recommendation must be received before consideration by full Curriculum Committee | N/A |
| [Common Core (Liberal Arts) Intent to Submit](http://openlab.citytech.cuny.edu/collegecouncil/files/2014/08/CommonCoreCourseSubmissionForm_4.2.12.doc) (if applicable) | N/A |
| Writing Intensive Form if course is intended to be a WIC (under development) | N/A |
| If course originated as an experimental course, then results of evaluation plan as developed with director of assessment. | N/A |
| **(Additional materials for** [**Curricular Experiments**](http://www.300jaystreet.com/college-council/curriculum_proposals/curricular-experiments)**)** |  |
| Plan and process for evaluation developed in consultation with the director of assessment. (Contact Director of Assessment for more information). |  |
| Established Timeline for Curricular Experiment |  |

**Proposal Course Name: Information Retrieval**

**Course Overview and Rationale**

In today’s society, data is being generated at an overwhelmingly rapid pace creating an unprecedented global challenge to analyze and understand this data and to use it to solve complex data-driven problems. This challenge has created an equally rapidly growing need for highly trained and skilled IT professionals who can take on the challenge by combining analytical, technical, mathematical, and statistical knowledge to solve data-driven problems. The Data Science program prepares students to meet the data challenge. Through lecture, hand-on, group and individual work as well as internship, the student in the Data Science major will be well versed in the tools and techniques required for meeting the big data challenges in many industries including finance, health, business, marketing and many more.

This upper level course prepares students for the challenging task of data retrieval. Integral topics related to data retrieval such as language processing, indexing, relevance and feedback are discussed. Students use data retrieval tools and techniques to understand the challenges and methods of retrieving complex data. Through hands-on exercises, classroom and home assignments and group projects, students gain experience in the use of data retrieval tools. Students will also identify and examine the key ethical issues involving the process of information retrieval including personal privacy, bias, informed consent, monitoring and surveillance. All CST students majoring in Data Science will be required to complete this course after successfully completing the prerequisite CST4702 Machine Learning Fundamental course with a grade of C or better. This course does not overlap with other courses both within and outside of the department. The CST department currently has full time faculty qualified to teach this course.

**Course Need**

**Students who would take this class**: intend to major in Data Science

**Department**: Computer Systems Technology

**Program**: Bachelors of Science in Data Science

**The number of sections (s) anticipated**: one section for the first year

**Projected headcount**: 24 students

**Physical resources required**: Basic smart room set-up: a screen, and an overhead projector that is run by and connected to a computer

**Course overlap**: None

**Faculty qualified for teaching this course**: There are faculty members who have doctoral degrees in big data and related big data topics.

**Course Design**

**Course context**: This course is required for the Data Science major.

**Course structure**: This course is offered in a lecture format.

**Anticipated Pedagogical Strategies and Instructional Design**: This course is offered using a combination of lecture, hands-on, in class assignments, home assignments and group work.

**Providing Support to Programmatic Learning Outcomes**: This course requires satisfactory completion of in class assignment, home assignments, group project, three tests, final exam and online discussions.

**NEW YORK CITY COLLEGE OF TECHNOLOGY**

**COMPUTER SYSTEMS TECHNOLOGY DEPARTMENT**

**CST4802 – Information Retrieval**

**(2 class hours, 2 lab hours, 3 credits)**

**COURSE DESCRIPTION:** The course details basic and advanced concepts and principles of information retrieval including: definition, framework, uses and ethical issues. Issues related to interface, relevance, language processing, and indexing are discussed. Information retrieval techniques covering both effectiveness and run-time performance are covered as well as heuristics and algorithms used in information retrieval systems. Information retrieval mechanism used in both Websites and Web search engines are also discussed. Ethical issues including personal privacy, bias, informed consent, monitoring and surveillance are also examined and discussed.

**LEARNING OUTCOMES:** Upon successful completion of this course, the students have acquired knowledge and skills to:

1. Define and analyze tasks associated with information retrieval, web search and clustering
2. Define and analyze information retrieval strategies and models, and demonstrate an understanding of the issues and challenges related to information retrieval.
3. Develop strategies suited for specific retrieval, clustering and classification scenarios and recognize the limits of these strategies
4. Design and implement efficient and effective queries within Information Retrieval systems.
5. Analyze and utilize techniques used for relevance feedback.
6. Analyze and design an interface for information retrieval systems.
7. Develop and utilize knowledge organization methods including clusters, indexes and facets. Analyze these knowledge organization methods for their impact on retrieval.
8. Analyze and understand the issues and challenges related to non-text retrieval.
9. Analyze and understand the ethical issues related to the process of information retrieval.

**GENERAL EDUCATION OUTCOMES:**

* **SKILLS/Inquiry/Analysis:** Students use scientific reasoning and logical thinking.
* **SKILLS/Communication:** Students use written (both reading and writing), oral (both speaking and listening), and visual means to communicate.
* **VALUES, ETHICS, RELATIONSHIPS / Professional/Personal Development:** Students work in diverse teams utilizing key traits including respect, cooperation and creativity.

**PREREQUISITES**

Student must have successfully completed CST4702 with a grade of C or higher.

**REQUIRED TEXTBOOKS**

Baeza-Yates, Ricardo and Berthier Ribeiro-Neto. *Modern Information Retrieval the Concepts and Technology Behind Search* 2nd Ed. New York: Pearson Education, 2011.

ISBN-13: 978-0321416919

Manning, Christopher D., Prabhakar Raghavan, & Hinrich Schütze. *An Introduction to Information Retrieval*. London: Cambridge University Press, 2009. Online: <https://nlp.stanford.edu/IR-book/pdf/irbookonlinereading.pdf>

**ADDITIONAL MATERIAL**

The instructor will identify several additional information resources during the semester including Internet resources, print material (handouts) and reference books.

**Academic Integrity –** Students and all others who work with information, ideas, texts, images, music, inventions, and other intellectual property owe their audience and sources accuracy and honesty in using, crediting, and citing sources. As a community of intellectual and professional workers, the College recognizes its responsibility for providing instruction in information literacy and academic integrity, offering models of good practice, and responding vigilantly and appropriately to infractions of academic integrity. Accordingly, academic dishonesty is prohibited in The City University of New York and at New York City College of Technology and is punishable by penalties, including failing grades, suspension, and expulsion. The complete text of the College policy on Academic Integrity may be found in the catalog.

**GRADING PROCEDURE:**

Tests (3) 30% Final Exam 20%

Labs 15%

Assignments 30%

Participation 5%

=====

TOTAL 100%

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Letter Grade** | A | A- | B+ | B | B- | C+ | C | D | F |
| **Numerical Grade** | 93-100 | 90-92.9 | 87-89.9 | 83-86.9 | 80-82.9 | 77-79.9 | 70-76.9 | 60-69.9 | <=59.9 |

**COURSE OUTLINE:**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Week** | **Topics** | **Reading** | **Assignments** | **Tests** |
| 1 | Introduction to Information Retrieval (IR) domain and to ethical issues related to this domain. | Baeza-Yates (Ch. 1) | Lab 1: Getting started with IR |  |
| 2 | Information Retrieval Models: Classic, Set Theoretic, Algebraic, Probabilistic | Baeza-Yates (Ch. 3)  Manning (Ch. 1) | Lab 2:IR models |  |
| 3 | Retrieval Evaluation Methods and Experiments: Cranfield Paradigm, retrieval metrics, user-base evaluation, reference collections | Baeza-Yates (Ch. 4)  Manning (Ch. 8) | Lab 3: Indexing and the Cranfield experiments |  |
| 4 | Relevance Feedback – Relevance, precision, feedback (Implicit/Explicit) | Baeza-Yates (Ch. 5)  Manning (Ch. 9) | Lab 4:Relevance | **Test 1** |
| 5 | Documents - Languages and Properties: Metadata, document formats, markup languages, text properties, document processing, organizing documents, text compression | Baeza-Yates (Ch. 6)  Manning (Ch. 2)  Manning (Ch. 12) | Lab 5: Data about data |  |
| 6 | Queries - Languages and Properties: Query languages, query properties. | Baeza-Yates (Ch. 7)  Manning (Ch. 12) | Lab 6:Queries  **Assignment 1** |  |
| 7 | Text Classification: Supervised/Unsupervised algorithms, evaluation metrics (Recall/Precision), organizing classes | Baeza-Yates (Ch. 8)  Manning (Ch. 13) | Lab 7: Google vs. Bing is there a difference in terms of precision/recall? | **Test 2** |
| 8 | Indexing and Searching: Inverted indexes, sequential searching, multi-dimensional indexing | Baeza-Yates (Ch. 9)  Manning (Ch. 4)  Manning (Ch. 5) | Lab 8:Indexing  **Assignment 2** |  |
| 9 | Parallel and Distributed IR: Data partitioning, parallel IR, cluster-based IR, distributed IR, federated search, peer-to-peer retrieval networks | Baeza-Yates (Ch. 10) | Lab 9:Parallel & Distributed |  |
| 10 | User Interfaces For IR: How people search, search interfaces, visualization in search interfaces | Baeza-Yates (Ch. 2) | Lab 10:User Interface  **Assignment 3** |  |
| 11 | Web Retrieval: Search engine architectures, search engine ranking (HITS/PageRank), managing Web data, search engine user interaction, browsing, bias and privacy | Baeza-Yates (Ch. 11)  Manning (Ch. 19)  Manning (Ch. 21) | Lab 11:SE  **Assignment 4** |  |
| 12 | Web Crawling: Architecture and implementation, scheduling algorithms, evaluation, bias, informed consent, surveillance and monitoring | Baeza-Yates (Ch. 12)  Manning (Ch. 20) | Lab 12: Dissecting a spider: how do web crawlers do it? | **Test 3** |
| 13 | Structured Text Retrieval: Explicit/Implicit, static/dynamic, XML retrieval, query languages | Chapter 13  Manning (Ch. 10) | Lab 13: Relational vs. NoSQL query languages |  |
| 14 | Multimedia IR: Content-based image retrieval, audio/music, video, segmentation, compression, MPEG | Baeza-Yates (Ch. 14) | Lab 14: Multimedia IR |  |
| 15 | Ethics of Information Retrieval (bias, privacy, informed consent, monitoring, surveillance)    Review |  |  | **Final Exam** |

**COURSE INTENDED LEARNING OUTCOMES/ASSESSMENT METHODS**

|  |  |
| --- | --- |
| **LEARNING OUTCOMES** | **ASSESSMENT METHODS** |
| 1. Demonstrate an understanding of the technical and ethical issues and challenges related to information retrieval | 1. Classroom lecture, in class and online interactive group discussions, written assignments |
| 1. Examine and analyze various types of retrieval systems | 1. Classroom lecture, in class and online interactive group discussions, written assignments |
| 1. Build and use categories, clusters and other organization techniques to gain first-hand experience with issues concerning the organization of information in relation to information retrieval | 1. Utilize various industry tools and techniques to code organized information elements |
| 1. Analyze and use mathematical methods and computer code related to relevance and feedback | 1. Utilize various techniques and information retrieval models to code retrieval algorithms |
| 1. Analyze, design and use interface design techniques | 1. Utilize design models to generate varying efficient user interfaces |
| 1. Analyze and understand issues and challenges related to non-text retrieval | 1. Classroom lecture, in class and online interactive group discussions, written assignments |

**GENERAL EDUCATION LEARNING OUTCOMES/ASSESSMENT METHODS**

|  |  |
| --- | --- |
| **LEARNING OUTCOMES** | **ASSESSMENT METHODS** |
| **1.** Demonstrate the ability to work collaboratively and independently on assignments in and outside a classroom setting. | **1.** Classroom discussions, group assignments and individual oral presentations. |
| **2.** Understand and employ both quantitative and qualitative analysis to solve problems**.** | **2.** Classroom discussion, group activities, group presentations, quizzes, tests, final exam. |
| **3.** Develop reading, writing competencies, and listening skills. | **3.** Biweekly reading and writing assignments, individual and group presentation, classroom discussion. Each homework assignment requires writing. |
| **4.** Work with teams. Build consensus. Use creativity. | **4.** Group projects and presentations. |

**Bibliography**

Chowdhury, G. G. *Introduction to Modern Information Retrieval* 3rd edition. London: Facet Publishing, 2011. ISBN-13: 978-1856046947

Chu, Heting. *Information Retrieval for a Digital Age* 2nd edition. New York: ASIS&T, 2010. ISBN-13:978-1573873932

Hearst, Marti. *Search User Interfaces.* New York: Cambridge University Press, 2009. ISBN-13: 978-0521113793

Manning, Christopher D., Hinrich Schutze, and Prabhakar Raghavan. *Introduction to Information Retrieval.* New York: Cambridge University Press, 2008. ISBN-13: 978-0521865715

Ouzzani, Mourad, Athaman Bouquettaya. *Semantic Web services for Web Database*. New York: Springer, 2011. ISBN-13: 978-1461416432

**CHANCELLOR’S REPORT FORM**

Department: Computer Systems Technology

Date: January 1st, 2018

**NEW COURSE PROPOSAL**

**Section AIV: New Course**

|  |  |
| --- | --- |
| **Department(s)** | **Computer Systems Technology** |
| **Academic Level** | **[X] Regular [  ] Compensatory [  ] Developmental [  ] Remedial** |
| **Subject Area** | Data Science |
| **Course Prefix** | **CST** |
| **Course Number** | **4802** |
| **Course Title** | **Information Retrieval** |
| **Catalog Description** | The course details basic and advanced concepts and principles of information retrieval including: definition, framework, uses and ethical issues. Issues related to interface, relevance, language processing, and indexing are discussed. Information retrieval techniques covering both effectiveness and run-time performance are covered as well as heuristics and algorithms used in information retrieval systems. Information retrieval mechanism used in both Websites and Web search engines are also discussed. Ethical issues including personal privacy, bias, informed consent, monitoring and surveillance are also examined and discussed. |
| **Prerequisite** | **CST4702 Machine Learning Fundamentals** |
| **Corequisite** | **None** |
| **Credits** | **3** |
| **Contact Hours** | **2 Class Hours, 2 Lab Hours** |
| **Liberal Arts** | **[ ] Yes [ X  ] No** |
| **Course Attribute (e.g. Writing Intensive, etc)** | **It is not a writing intensive course** |
| **Course Applicability** | **[ X ] Major**  **[ ] Gen Ed Required [ ] Gen Ed - Flexible [ ] Gen Ed - College Option**  **[ ] English Composition [ ] World Cultures [ ] Speech**  **[ ] Mathematics [ ] US Experience in its Diversity**  **[ ] Interdisciplinary [ ] Creative Expression [ ] Advanced Liberal Arts**  **[ ] Science [ ] Individual and Society**  **[ ] Scientific World** |
| **Effective Term** | **Fall 2019** |

**LIBRARY RESOURCES & INFORMATION LITERACY: MAJOR CURRICULUM MODIFICATION**

Please complete for **all** major curriculum modifications. This information will assist the library in planning for new courses/programs.

Consult with your library faculty subject specialist (<http://cityte.ch/dir>) **3 weeks before the proposal deadline**.

**Course proposer:** please complete boxes 1-4. **Library faculty subject specialist:** please complete box 5.

|  |  |  |
| --- | --- | --- |
| **1** | **Title of proposal**  CST4802 – Information Retrieval | **Department/Program**  Computer Systems Technology (CST) Department |
|  | **Proposed by** (include email & phone)  Elizabeth Milonas (718)260-5170  emilonas@citytech.cuny.edu | **Expected date course(s) will be offered**  Fall 2019  **# of students 24** |

|  |  |
| --- | --- |
| **2** | **The library cannot purchase reserve textbooks for every course at the college, nor copies for all students. Consult our website (**[**http://cityte.ch/curriculum**](http://cityte.ch/curriculum)**) for articles and ebooks for your courses, or our open educational resources (OER) guide (**[**http://cityte.ch/oer**](http://cityte.ch/oer)**). Have you considered using a freely-available OER or an open textbook in this course?**  Yes I have however, none are suitable. |

|  |  |
| --- | --- |
| **3** | **Beyond the required course materials, are City Tech library resources sufficient for course assignments? If additional resources are needed, please provide format details (e.g. ebook, journal, DVD, etc.), full citation (author, title, publisher, edition, date), price, and product link.**  Yes the current library resources are sufficient for this course. |

|  |  |
| --- | --- |
| **4** | **Library faculty focus on strengthening students' information literacy skills in finding, critically evaluating, and ethically using information. We collaborate on developing assignments and customized instruction and research guides. When this course is offered, how do you plan to consult with the library faculty subject specialist for your area? Please elaborate.**  I will reach out to the library subject specialist via email to arrange an information session in which the library subject specialist can present to the students of this course, the use of library databases, citation convention and discuss copyright issues. |

|  |  |
| --- | --- |
| **5** | **Library Faculty Subject Specialist Prof. Junior Tidal**  **Comments and Recommendations**  After surveying the collection, I believe that the library will need additional materials to properly support this course. I believe that monographs related to search algorithms, web crawling, search engines, user interfaces, and other related subjects will need to be acquired.    **Date 12.13.17** |

# New Course Proposal #8: CST4812 Natural Language Processing

New York City College of Technology, CUNY

CURRICULUM MODIFICATION PROPOSAL FORM

This form is used for all curriculum modification proposals. See the [Proposal Classification Chart](http://openlab.citytech.cuny.edu/collegecouncil/files/2014/08/2013-10-09-Proposal_Classification_Chart.pdf) for information about what types of modifications are major or minor. Completed proposals should be emailed to the Curriculum Committee chair.

|  |  |
| --- | --- |
| **Title of Proposal** | **CST 4812 Natural Language Processing** |
| **Date** | **Jan 4th, 2018** |
| **Major or Minor** | **Major** |
| **Proposer’s Name** | **Dr. Elena Filatova** |
| **Department** | **Computer Systems Technology** |
| **Date of Departmental Meeting in which proposal was approved** | **12/8/2017** |
| **Department Chair Name** | **Dr. Hong Li** |
| **Department Chair Signature and Date** |  |
| **Academic Dean Name** | **Dean Kevin Hom** |
| **Academic Dean Signature and Date** | 2/22/18 |
| **Brief Description of Proposal**  (Describe the modifications contained within this proposal in a succinct summary. More detailed content will be provided in the proposal body. | Title: Natural Language Processing  This proposal explains the rationale for developing the Natural Language Processing course as a required course in the Data Science curriculum of the Computer Systems Technology Department. The course outline and sample syllabus are attached to the proposal. |
| **Brief Rationale for Proposal**  (Provide a concise summary of why this proposed change is important to the department. More detailed content will be provided in the proposal body). | Vast amounts of new information and data that are generated every day are presented in a text form (e.g., new articles, product reviews, scientific articles, etc.). This text data needs to be accessed, analyzed, annotated, processed. Natural Language Processing studies methods, techniques, and tools that are designed for automatics analysis and processing of text data. Also, this course is integral to the proposed Bachelors in Data Science curriculum program of the Computer Systems Technology department. |
| **Proposal History**  (Please provide history of this proposal: is this a resubmission? An updated version? This may most easily be expressed as a list). | This is a first submission. |
| **CUNY – Course Equivalencies**  Provide information about equivalent courses within CUNY, if any. | None |
| **Intent to Submit as Common Core**  If this course is intended to fulfill one of the requirements in the common core, then indicate which area. | N/A |

Please include all appropriate documentation as indicated in the Curriculum Modification Checklist.

For each new course, please also complete the New Course Proposal and submit in this document.

Please submit this document as a single .doc or .rtf format. If some documents are unable to be converted to .doc, then please provide all documents archived into a single .zip file.

**ALL PROPOSAL CHECK LIST**

|  |  |
| --- | --- |
| Completed CURRICULUM MODIFICATION FORM including: |  |
| * Brief description of proposal | X |
| * Rationale for proposal | X |
| * Date of department meeting approving the modification | X |
| * Chair’s Signature | X |
| * Dean’s Signature |  |
| Evidence of consultation with affected departments  List of the programs that use this course as required or elective, and courses that use this as a prerequisite. | X |
| Documentation of Advisory Commission views (if applicable). | N/A |
| Completed [Chancellor’s Report Form](http://openlab.citytech.cuny.edu/collegecouncil/files/2014/08/2013-10-09-Chancellor_Report_Quick_Reference_Guide1.doc). | X |

**EXISTING PROGRAM MODIFICATION PROPOSALS**

|  |  |
| --- | --- |
| Documentation indicating core curriculum requirements have been met for new programs/options or program changes. | X |
| Detailed rationale for each modification (this includes minor modifications) | N/A |

New York City College of Technology, CUNY

NEW COURSE PROPOSAL FORM

This form is used for all new course proposals. Attach this to the [Curriculum Modification Proposal Form](http://openlab.citytech.cuny.edu/collegecouncil/files/2014/08/2013-10-10-Curriculum_Modification_Proposal_Form.docx) and submit as one package as per instructions. Use one New Course Proposal Form for each new course.

|  |  |
| --- | --- |
| **Course Title** | Natural Language Processing |
| **Proposal Date** | Jan 4th, 2018 |
| **Proposer’s Name** | Dr. Elena Filatova |
| **Course Number** | CST 4812 |
| **Course Credits, Hours** | 3 credits, 2 lecture hours and 2 lab hours |
| **Course Pre / Co-Requisites** | Pre-req: CST 3512 (Information and Data Management II) and MAT 2572 (Probability and Mathematical Statistics I) |
| **Catalog Course Description** | Introduces a broad range of techniques that aim to read, understand, and extract information from natural language text resources using both knowledge-based and statistical approaches for syntax, semantics, and discourse. In addition to introducing theoretical aspects and methods of natural language processing (NLP), the course covers multiple applications such as Machine Translation, Summarization, Interactive Dialog Systems, etc. |
| **Brief Rationale**  Provide a concise summary of why this course is important to the department, school or college. | Vast amounts of new information and data that are generated every day are presented in a text form (e.g., new articles, product reviews, scientific articles, etc.). This text data needs to be accessed, analyzed, annotated, processed. Natural Language Processing studies methods, techniques, and tools that are designed for automatics analysis and processing of text data. Also, this course is integral to the proposed Bachelors in Data Science curriculum program of the Computer Systems Technology department. |
| **CUNY – Course Equivalencies**  Provide information about equivalent courses within CUNY, if any. | N/A |
| **Intent to Submit as Common Core**  If this course is intended to fulfill one of the requirements in the common core, then indicate which area. | No. |
| **Intent to Submit as An Interdisciplinary Course** | No. |
| **Intent to Submit as a Writing Intensive Course** | No. |

Please include all appropriate documentation as indicated in the NEW COURSE PROPOSAL Combine all information into a single document that is included in the Curriculum Modification Form.

**NEW COURSE PROPOSAL CHECK LIST**

Use this checklist to ensure that all required documentation has been included. You may wish to use this checklist as a table of contents within the new course proposal.

|  |  |
| --- | --- |
| **Completed NEW COURSE PROPOSAL FORM** |  |
| * Title, Number, Credits, Hours, Catalog course description | X |
| * Brief Rationale | X |
| * CUNY – Course Equivalencies | X |
| Completed [Library Resources and Information Literacy Form](http://openlab.citytech.cuny.edu/collegecouncil/files/2014/08/curriculum_modification_library_form.doc) | X |
| **Course Outline**  Include within the outline the following. | **X** |
| Hours and Credits for Lecture and Labs  If hours exceed mandated Carnegie Hours, then rationale for this | X |
| Prerequisites/Co- requisites | X |
| Detailed Course Description | X |
| Course Specific Learning Outcome and Assessment Tables   * Discipline Specific * General Education Specific Learning Outcome and Assessment Tables | X |
| Example Weekly Course outline | X |
| Grade Policy and Procedure | X |
| Recommended Instructional Materials (Textbooks, lab supplies, etc) | X |
| Library resources and bibliography | X |
| **Course Need Assessment.**  Describe the need for this course. Include in your statement the following information. | X |
| Target Students who will take this course. Which programs or departments, and how many anticipated?  Documentation of student views (if applicable, e.g. non-required elective). | X |
| Projected headcounts (fall/spring and day/evening) for each new or modified course. | X |
| If additional physical resources are required (new space, modifications, equipment), description of these requirements. If applicable, Memo or email from the VP for Finance and Administration with written comments regarding additional and/or new facilities, renovations or construction. | X |
| Where does this course overlap with other courses, both within and outside of the department? | X |
| Does the Department currently have full time faculty qualified to teach this course? If not, then what plans are there to cover this? | X |
| If needs assessment states that this course is required by an accrediting body, then provide documentation indicating that need. | N/A |
| **Course Design**  Describe how this course is designed. | X |
| Course Context (e.g. required, elective, capstone) | X |
| Course Structure: how the course will be offered (e.g. lecture, seminar, tutorial, fieldtrip)? | X |
| Anticipated pedagogical strategies and instructional design (e.g. Group Work, Case Study, Team Project, Lecture) | X |
| How does this course support Programmatic Learning Outcomes? | X |
| Is this course designed to be partially or fully online? If so, describe how this benefits students and/or program. | X |
| **Additional Forms for Specific Course Categories** |  |
| [Interdisciplinary Form](http://openlab.citytech.cuny.edu/collegecouncil/files/2014/08/Application-for-Interdisciplinary-Course-Designation.docx) (if applicable) | N/A |
| Interdisciplinary Committee Recommendation (if applicable and if received)\*  \*Recommendation must be received before consideration by full Curriculum Committee | N/A |
| [Common Core (Liberal Arts) Intent to Submit](http://openlab.citytech.cuny.edu/collegecouncil/files/2014/08/CommonCoreCourseSubmissionForm_4.2.12.doc) (if applicable) | N/A |
| Writing Intensive Form if course is intended to be a WIC (under development) | N/A |
| If course originated as an experimental course, then results of evaluation plan as developed with director of assessment. | N/A |
| **(Additional materials for** [**Curricular Experiments**](http://www.300jaystreet.com/college-council/curriculum_proposals/curricular-experiments)**)** |  |
| Plan and process for evaluation developed in consultation with the director of assessment. (Contact Director of Assessment for more information). |  |
| Established Timeline for Curricular Experiment |  |

**Proposed Course Name: Data Visualization**

**Course Overview & Rationale**

Vast amounts of new information and data that are generated every day are presented in a text form (e.g., new articles, product reviews, scientific articles, etc.). This text data needs to be accessed, analyzed, annotated, processed. Natural Language Processing studies methods, techniques, and tools that are designed for automatics analysis and processing of text data. Also, this course is integral to the proposed Bachelors in Data Science curriculum program of the Computer Systems Technology department.

This course introduces a broad range of techniques that aim to read, understand, and extract information from natural language text resources. The course covers both knowledge-based and statistical approaches for syntax, semantics, and discourse. In addition to introducing theoretical aspects and methods of natural language processing (NLP), the course covers multiple applications such as Machine Translation, Summarization, Interactive Dialog Systems, etc.

**Course Need**

**Students who would take this class:** students who intend to major in Data Science

**Department**: Computer Systems Technology

**Program**: Bachelors in Data Science

**The number of section (s) anticipated:** one section for the first year

**Projected headcount: 24** students

**Physical Resources required:** Basicsmartroomset**-**up**:** a screen, and an overhead projector/a TV set that is run by and connected to a computer

**Course overlap:** None

**Faculty** **qualified** **for** **teaching** **this** **course**: Yes, there are faculty members who have doctoral degrees in Computer Science with the concentration in Data Processing for various domains.

**Course design**

**Course context:** This course will be required of Data Science major students. Students are required to develop an independent project at the end of the semester.

**Course** **structure**: This course will be offered in a lecture style/format.

**Anticipated** **Pedagogical** **Strategies** **and** **Instructional** **Design**: This class will be run in a lecture-activity style/format. The class will start with a lecture, and then move on to creative in-class activities, such as using Python-based NLTK software for text analysis and processing activities.

**Providing Support to Programmatic Learning Outcomes:** This course requires satisfactory completion of individual assignments, two major exams and a final term project.

**NEW YORK CITY COLLEGE OF TECHNOLOGY**

**COMPUTER SYSTEMS TECHNOLOGY DEPARTMENT**

**CST4812 – Natural Language Processing**

**(2 class hours, 2 lab hours, 3 credits)**

**Course Description:** This course introduces a broad range of techniques that aim to read, understand, and extract information from natural language text resources. The course covers both knowledge-based and statistical approaches for syntax, semantics, and discourse. In addition to introducing theoretical aspects and methods of natural language processing (NLP), the course covers multiple applications such as Machine Translation, Summarization, Interactive Dialog Systems, etc.

**Pre-requisites:** CST3512 (Information and Data Management II) and MAT2572 (Probability and Mathematical Statistics I)

**Learning Outcomes:**

1. Demonstrate knowledge of the issues and challenges NLP.
2. Demonstrate knowledge of the elements of NLP techniques and applications.
3. Process text through the language levels using Python-based libraries and software tools.
4. Work with NLP tools to create NLP applications.
5. Describe how NLP is used in many types of real world applications.

**Required Textbook:**

S. Bird, E. Klein, E. Loper. Analyzing Text with Natural Language Toolkit

Free on-line version: [*http://www.nltk.org/book/*](http://www.nltk.org/book/)

**Required Textbooks:**

D. Jurafsky & James H. Martin. *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics and Speech Recognition*, Second Edition, 2009.

ISBN-13: 978-0131873216

**Additional Reading Materials:**

Foundations of Statistical Natural Language Processing, by Chritopher Manning and Hinrich Schuetze

The instructor will identify several additional information resources during the semester including Internet resources, print material (handouts) and reference books.

**Grading:**

Homework assignments 25% Mid-term exam 20%

Project 30%

Final exam 25%

=====

TOTAL 100%

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Letter Grade** | A | A- | B+ | B | B- | C+ | C | D | F |
| **Numerical Grade** | 93-100 | 90-92.9 | 87-89.9 | 83-86.9 | 80-82.9 | 77-79.9 | 70-76.9 | 60-69.9 | <=59.9 |

The grade distribution follows the information in the NYCCT Student Handbook (p.43).

During the course of the class you are required to follow the NYCCT Academic Integrity Standards described in the Student Handbook (pp.95 – 99)

NYCCT Student Handbook can be downloaded here: <http://www.citytech.cuny.edu/current-student/docs/StudentHandbook.pdf>.

**COURSE OUTLINE (tentative):**

|  |  |  |
| --- | --- | --- |
| **Week** | **Topics** | **Assignments and Readings** |
| 1 | Introduction | Ch. 0 |
| 2 | Regular expressions;  Computing with language: text and words;  Searching text  Lab: regular expressions | Ch. 1, Ch. 2, Ch. 3, Ch. 4  Assignment 1 |
| 3 | Language modelling, n-grams, tf, tf\*idf counts  Lab: computing text frequencies | Ch. 5 |
| 4 | Text corpora;  Accessing / generating text corpora | Ch. 3, Ch. 4 |
| 5 | Text classification using n-grams with tf\*idf counts as features  Lab: examples of text classification | Ch. 6  Assignment 2 |
| 6 | Part-Of-Speech tagging;  Lab: Using POS taggers | Ch. 2 |
| 7 | Using Part-of-Speech tags for classification | Ch. 2, Ch. 6 |
| 8 | **Mid-term**;  Analyzing sentence structure  Lab: the basics of syntactic analysis | **Ch. 8** |
| 9 | Grammars; Context Free Grammars (CFG)  Syntactic parsing; statistical parsing | Ch. 9 |
| 10 | Dependencies and Dependency Grammar | Ch. 9 |
| 11 | Building Linguistic features  Semantic Analysis;  The representation of meaning | Assignment 4 |
| 12 | Word Sense Disambiguation (WSD) | Ch. 10 |
| 13 | Applications review: Information extraction, question-answering, summarization, machine translation, dialogue systems | Ch. 11 |
| 14 | **Review/Final** |  |
| 15 | **Project discussion and presentation** |  |

**ASSESSMENT CRITERIA:** For successful completion of this course the student should be able to:

|  |  |
| --- | --- |
| **For the successful completion of this course a student should be able to:** | **Evaluation methods and criteria** |
| 1. Demonstrate knowledge of the issues and challenges NLP. 2. Demonstrate knowledge of the elements of NLP techniques and applications. 3. Process text through the language levels using Python-based libraries and software tools. 4. Work with NLP tools to create NLP applications. 5. Describe how NLP is used in many types of real world applications. | Exams, homework assignments, in-class lab exercises and project |

**GENERAL EDUCATION LEARNING OUTCOMES/ASSESSMENT METHODS**

|  |  |
| --- | --- |
| **LEARNING OUTCOMES** | **ASSESSMENT METHODS** |
| **1.** Demonstrate the ability to work collaboratively and independently on assignments in and outside a classroom setting. | **1.** Classroom discussions, group assignments and individual oral presentations. |
| **2.** Understand and employ both quantitative and qualitative analysis to solve problems**.** | **2.** Classroom discussion, group activities, group presentations, quizzes, tests, final exam. |
| **3.** Develop reading, writing competencies, and listening skills. | **3.** Biweekly reading and writing assignments, individual and group presentation, classroom discussion. Each homework assignment requires writing. |
| **4.** Work with teams. Build consensus. Use creativity. | **4.** Group projects and presentations. |

**Bibliography:**

* S. Bird, E. Klein, E. Loper. *Analyzing Text with Natural Language Toolkit*

Free on-line version: [*http://www.nltk.org/book/*](http://www.nltk.org/book/)

* Dan Jurafsky & James H. Martin (2009). *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics and Speech Recognition*, Second Edition, 2009.

Publisher: Prentice Hall; 2nd edition

ISBN-13: 978-0131873216

* Chritopher Manning and Hinrich Schuetze (1999). *Foundations of Statistical Natural Language Processing*

Publisher: The MIT Press; 1 edition (June 18, 1999)

ISBN-13: 978-0262133609

**CHANCELLOR’S REPORT FORM**

Department: Computer Systems Technology

Date: January 4th, 2018

**NEW COURSE PROPOSAL**

**Section AIV: New Course**

|  |  |
| --- | --- |
| **Department(s)** | **Computer Systems Technology** |
| **Academic Level** | **[X] Regular [  ] Compensatory [  ] Developmental [  ] Remedial** |
| **Subject Area** | Data Science |
| **Course Prefix** | **CST** |
| **Course Number** | **CST 4812** |
| **Course Title** | **Natural Language Processing** |
| **Catalog Description** | Introduces a broad range of techniques that aim to read, understand, and extract information from natural language text resources using both knowledge-based and statistical approaches for syntax, semantics, and discourse. In addition to introducing theoretical aspects and methods of natural language processing (NLP), the course covers multiple applications such as Machine Translation, Summarization, Interactive Dialog Systems, etc. |
| **Prerequisite** | **CST 3512 and MAT2572** |
| **Corequisite** | **None** |
| **Credits** | **3** |
| **Contact Hours** | **2 Class Hours, 2 Lab Hours** |
| **Liberal Arts** | **[ ] Yes [ X  ] No** |
| **Course Attribute (e.g. Writing Intensive, etc)** | **It is not a writing intensive course** |
| **Course Applicability** | **[X] Major**  **[ ] Gen Ed Required [ ] Gen Ed - Flexible [ ] Gen Ed - College Option**  **[ ] English Composition [ ] World Cultures [ ] Speech**  **[ ] Mathematics [ ] US Experience in its Diversity**  **[ ] Interdisciplinary [ ] Science [ ] Creative Expression**  **[ ] Advanced Liberal Arts**  **[ ] Individual and Society**  **[ ] Scientific World** |
| **Effective Term** | **Fall 2019** |

**Rationale:** Vast amounts of new information and data that are generated every day are presented in a text form (e.g., new articles, product reviews, scientific articles, etc.). This text data needs to be accessed, analyzed, annotated, processed. Natural Language Processing studies methods, techniques, and tools that are designed for automatics analysis and processing of text data. Also, this course is integral to the proposed Bachelors in Data Science curriculum program of the Computer Systems Technology department.

**LIBRARY RESOURCES & INFORMATION LITERACY: MAJOR CURRICULUM MODIFICATION**

Please complete for **all** major curriculum modifications. This information will assist the library in planning for new courses/programs.

Consult with your library faculty subject specialist (<http://cityte.ch/dir>) **3 weeks before the proposal deadline**.

**Course proposer:** please complete boxes 1-4. **Library faculty subject specialist:** please complete box 5.

|  |  |  |
| --- | --- | --- |
| **1** | **Title of proposal**  CST 4812: Natural Language Processing | **Department/Program**  Computer and Systems Technology |
|  | **Proposed by** (include email & phone)  Elena Filatova  efilatova@citytech.cuny.edu | **Expected date course(s) will be offered**  Fall 2019  **# of students** 24 |

|  |  |
| --- | --- |
| **2** | **The library cannot purchase reserve textbooks for every course at the college, nor copies for all students. Consult our website (**[**http://cityte.ch/curriculum**](http://cityte.ch/curriculum)**) for articles and ebooks for your courses, or our open educational resources (OER) guide (**[**http://cityte.ch/oer**](http://cityte.ch/oer)**). Have you considered using a freely-available OER or an open textbook in this course?**  All the materials for the class are on-line, free (http://www.nltk.org/book/). |

|  |  |
| --- | --- |
| **3** | **Beyond the required course materials, are City Tech library resources sufficient for course assignments? If additional resources are needed, please provide format details (e.g. ebook, journal, DVD, etc.), full citation (author, title, publisher, edition, date), price, and product link.**  Tools and On-Line Documentation (free, on-line)  Python: https://www.python.org/  http://www.nltk.org/ |

|  |  |
| --- | --- |
| **4** | **Library faculty focus on strengthening students' information literacy skills in finding, critically evaluating, and ethically using information. We collaborate on developing assignments and customized instruction and research guides. When this course is offered, how do you plan to consult with the library faculty subject specialist for your area? Please elaborate.**  The project for the course concerns an on-line data set. There will be a discussion in class, what data can be downloaded and used for the project, whether there are any ethical consequences for using this data set and making conclusions about this data set. |

|  |  |
| --- | --- |
| **5** | **Library Faculty Subject Specialist Prof. Junior Tidal**  **Comments and Recommendations**  I believe that the library can adequately support this course. I recommend that in addition to acquiring the recommended texts, pending course approval, to also acquire books on computational linguistics to the library’s collection.  **Date 12.13.17** |

# Proposal to Change Prerequisites for: CST 4714: Database Administration

New York City College of Technology, CUNY

CURRICULUM MODIFICATION PROPOSAL FORM

This form is used for all curriculum modification proposals. See the [Proposal Classification Chart](http://openlab.citytech.cuny.edu/collegecouncil/files/2014/08/2013-10-09-Proposal_Classification_Chart.pdf) for information about what types of modifications are major or minor. Completed proposals should be emailed to the Curriculum Committee chair.

|  |  |
| --- | --- |
| **Title of Proposal** | **Change prerequisites of CST 4714: Database Administration** |
| **Date** | **February 28, 2018** |
| **Major or Minor** | **Minor** |
| **Proposer’s Name** | **Ashwin Satyanarayana** |
| **Department** | **Computer Systems Technology** |
| **Date of Departmental Meeting in which proposal was approved** | **2/28/2018** |
| **Department Chair Name** | **Hong Li** |
| **Department Chair Signature and Date** | 3/3/2018 |
| **Academic Dean Name** | **Kevin Hom** |
| **Academic Dean Signature and Date** |  |
| **Brief Description of Proposal**  (Describe the modifications contained within this proposal in a succinct summary. More detailed content will be provided in the proposal body. | **The proposal is to change the prerequisites of CST4714**  **from: CST 2405 or CST 2415 or CST 3604**  **to: CST 2405 or CST 2415 or CST 3604 or CST 2402** |
| **Brief Rationale for Proposal**  (Provide a concise summary of why this proposed change is important to the department. More detailed content will be provided in the proposal body). | **This change is necessary to accommodate students in the proposed Data Science program to take CST4714 (Database Administration). The newly added pre-requisite CST 2402 is part of the Data Science curriculum. The other pre-requisites (CST 2405 or CST 2415 or CST 3604) for this course are not part of the proposed program.**  **In addition, CST 2402 (Introduction to Data Science) provides intellectual foundation of prior knowledge within the discipline of Data Science and will enhance the quality of the learning experience in CST4714.** |
| **Proposal History** | **This is a new submission.** |

Please include all appropriate documentation as indicated in the Curriculum Modification Checklist.

**ALL PROPOSAL CHECK LIST**

|  |  |
| --- | --- |
| Completed CURRICULUM MODIFICATION FORM including: |  |
| * Brief description of proposal | X |
| * Rationale for proposal | X |
| * Date of department meeting approving the modification | X |
| * Chair’s Signature | X |
| * Dean’s Signature | X |
| Evidence of consultation with affected departments  List of the programs that use this course as required or elective, and courses that use this as a prerequisite. | N/A |
| Documentation of Advisory Commission views (if applicable). | N/A |
| Completed [Chancellor’s Report Form](http://openlab.citytech.cuny.edu/collegecouncil/files/2014/08/2013-10-09-Chancellor_Report_Quick_Reference_Guide1.doc). | X |

**EXISTING PROGRAM MODIFICATION PROPOSALS**

|  |  |
| --- | --- |
| Documentation indicating core curriculum requirements have been met for new programs/options or program changes. | N/A |
| Detailed rationale for each modification (this includes minor modifications) | X |

**2,3. Chancellor’s Report Template for Modification of Existing Courses (also used for minor changes)**

**Section AV: Changes in Existing Courses**

**AV.1. CST4714 Database Administration**

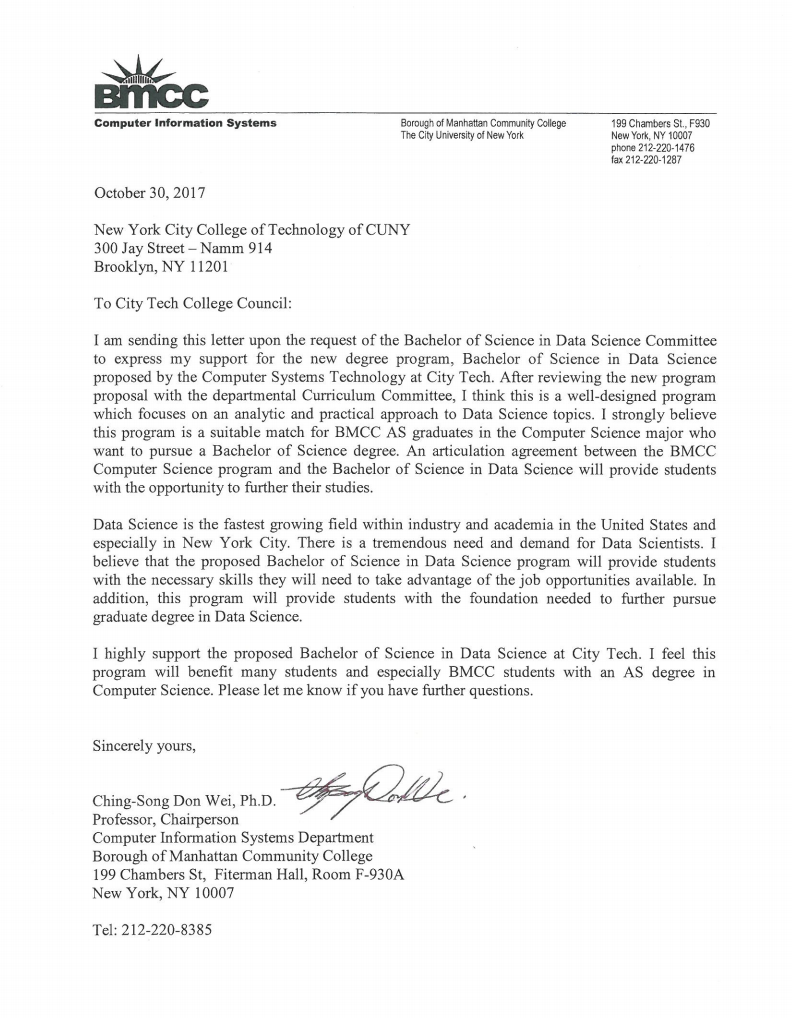
|  |  |  |  |
| --- | --- | --- | --- |
| **CUNYFirst Course ID** | 037917 |  |  |
| **FROM:** |  | **TO:** |  |
| **Department(s)** |  | **Department(s)** |  |
| **Course** |  | **Course** |  |
| **Prerequisite** | ~~CST 2405 or CST 2415 or CST 3604~~ | **Prerequisite** | CST 2405 or CST 2415 or CST 3604 or CST 2402 |
| **Corequisite** |  | **Corequisite** |  |
| **Pre- or corequisite** |  | **Pre- or corequisite** |  |
| **Hours** |  | **Hours** |  |
| **Credits** |  | **Credits** |  |
| **Description** |  | **Description** |  |
| **Requirement Designation** |  | **Requirement Designation** |  |
| **Liberal Arts** | [ ] Yes [ ] No | **Liberal Arts** | [ ] Yes [ ] No |
| **Course Attribute (e.g. Writing Intensive, Honors, etc** |  | **Course Attribute (e.g. Writing Intensive, Honors, etc** |  |
| **Course Applicability** | |  | | --- | | [ ] Major | | [ ] Gen Ed Required | | [ ] English Composition | | [ ] Mathematics | | [ ] Science | | [ ] Gen Ed - Flexible | | [ ] World Cultures | | [ ] US Experience in its Diversity | | [ ] Creative Expression | | [ ] Individual and Society | | [ ] Scientific World | | [ ] Gen Ed - College Option | | [ ] Speech | | [ ] Interdisciplinary | | [ ] Advanced Liberal Arts | | **Course Applicability** | |  | | --- | | [ ] Major | | [ ] Gen Ed Required | | [ ] English Composition | | [ ] Mathematics | | [ ] Science | | [ ] Gen Ed - Flexible | | [ ] World Cultures | | [ ] US Experience in its Diversity | | [ ] Creative Expression | | [ ] Individual and Society | | [ ] Scientific World | | [ ] Gen Ed - College Option | | [ ] Speech | | [ ] Interdisciplinary | | [ ] Advanced Liberal Arts | |
| **Effective Term** | Fall 2019 | | |

**Rationale:** This change is necessary to accommodate students in the proposed Data Science program to take CST4714 (Database Administration). The newly added pre-requisite CST 2402 is part of the Data Science curriculum. The other pre-requisites (CST 2405 or CST 2415 or CST 3604) for this course are not part of the new proposed program.

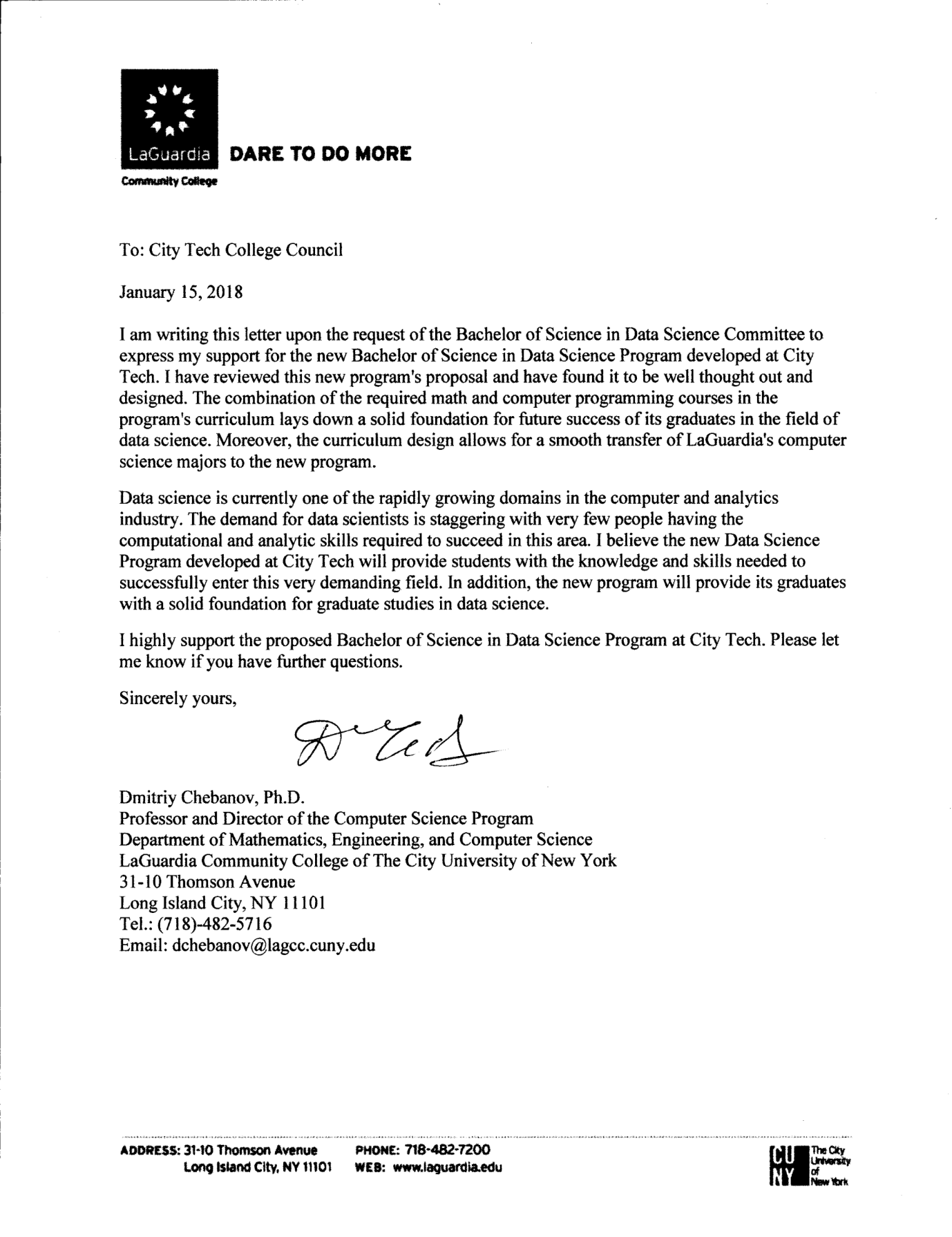
In addition, CST 2402 (Introduction to Data Science) provides intellectual foundation of prior knowledge within the discipline of Data Science and will enhance the quality of the learning experience in CST4714.

# Appendix A: Letters of Support

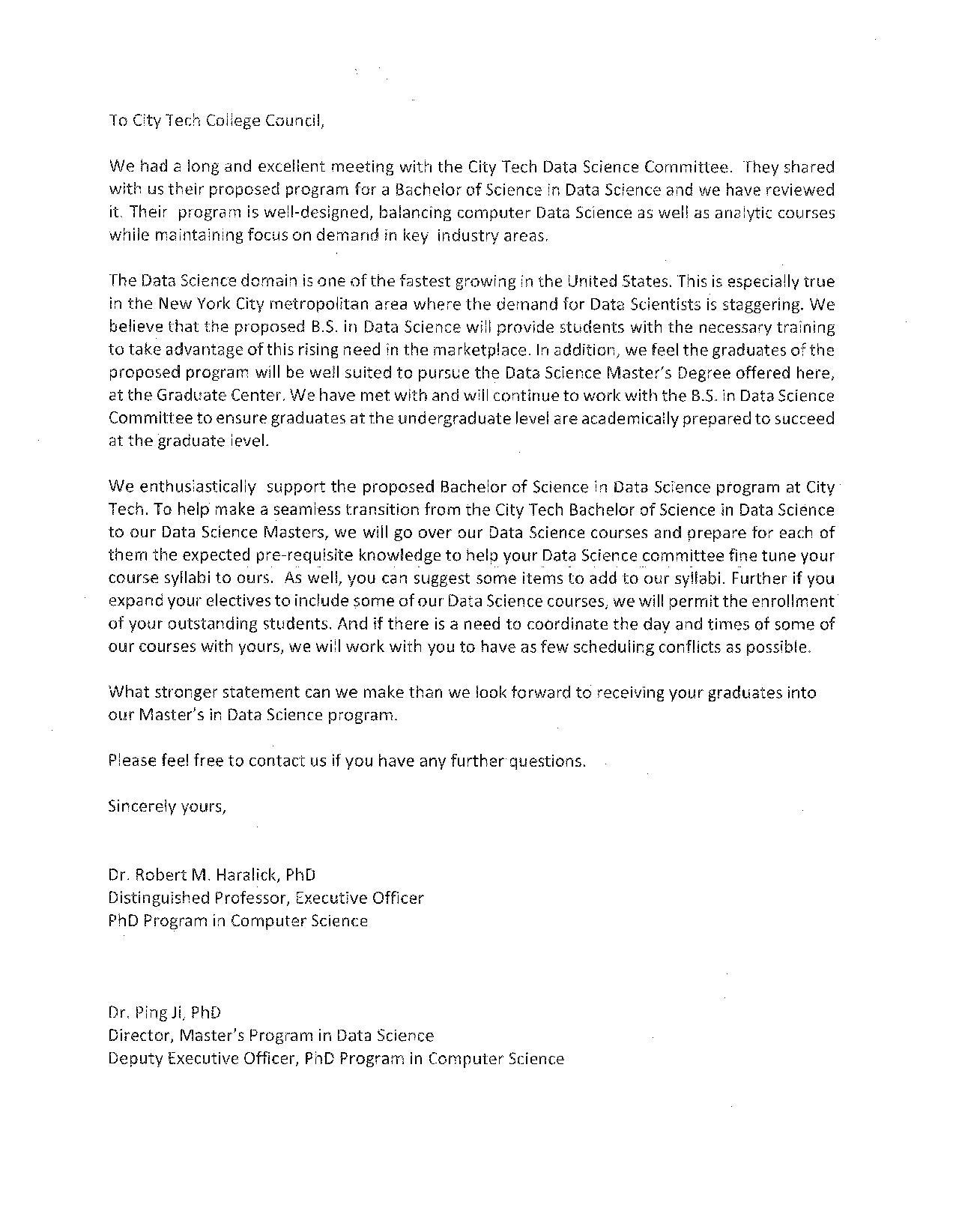
**1. Borough of Manhattan Community College**



**2. LaGuardia Community College**

****

**3. CUNY Graduate Center (Computer Science Department)**

****

1. **CitiGroup**



1. **Ernst & Young**



# Appendix B: Sample Job Postings

**1. Data Engineer** (<https://www.glassdoor.com/job-listing/data-engineer-mount-sinai-health-system-JV_IC1132348_KO0,13_KE14,39.htm?jl=2546746975>)

Mount Sinai Health System - New York, NY 10019

**Job Summary**

**The Mount Sinai Health System**

**Do you have what it takes to wear the badge?**

The Mount Sinai Health System’s commitment to excellence extends beyond delivering world-class health care. The System’s ongoing success is dependent upon our highly motivated, nonclinical professionals working to improve business operations. Our leadership team is driven to provide exceptional service by cultivating a workforce that is dedicated to upholding Mount Sinai’s mission of delivering innovative, breakthrough medicine with compassion and integrity.

Are you ready to discover the world of limitless possibilities that comes with wearing the badge? Explore more about this opportunity and how you can help us write a new chapter in our story of unrivaled patient care!

**What You’ll Do:**

The Data Engineer II focuses on data collection, movement, storage, transformation processing, and storage of Big Data. This individual works with both current ETL/Data Warehousing and future Big Data/Streaming/Pipeline architectures. The focus is on choosing optimal solutions to use for these purposes, then implementing, maintaining, and monitoring them, always being mindful of the overarching goal of accelerating translational research and improving clinical care.

* Facilitates data collection from a variety of sources, getting it in the right formats, assuring that it adheres to data quality standards, and assuring that downstream users can get that data quickly and with a common standard interface.
* Ensures that data streams/pipelines are scalable, repeatable, and secure, and can serve multiple users within the Institute.
* Develops as a core member of an Agile team, using Agile tools and methodology. Work closely with other team members including Application Developers, Database Developers, and Data Scientists.
* Responsible for creating the infrastructure that provides insight from raw data and handles diverse sources of data seamlessly.
* Enables big data and batch/real-time analytical solutions that leverage emerging technologies.
* Additional responsibilities include developing prototypes and proof of concepts for the selected solutions, and implementing complex big data projects with a focus on collecting, parsing, and managing large sets of data using multiple platforms to allow for Research and Data Science initiatives.
* Translates business requirements into modern data pipeline solutions. Create centralized documents and diagrams of all solutions.
* Creates a data catalog store of all metadata.
* Designs and implements monitoring, backup, and disaster recovery of data systems.
* Approaches all relationships with a world-class customer service approach. Maintains a customer-focused approach with users to provide solutions that are science/research-driven.
* Responsible for the integrity and security of data in all forms of storage throughout the Data Architecture.
* Works with other IT professionals through Mount Sinai effectively. Comply with the Institutional Review Board and HIPAA to follow all applicable policies and procedures.
* Assists in the development of standards and procedures affecting data management, design and maintenance. Documents all standards and procedures.
* Provides presentations and training to other team members in the above.
* Possesses an extremely flexible attitude. Willing to work with multiple types of technologies and languages with an open mind and without technology bias. Continuous interest in updating skill sets and knowledge of trends in the Big Data Technology space.
* Other duties as assigned.

***What You’ll Bring:***

Bachelor’s degree in Computer Science or a related discipline; Advanced degree preferred

* 4+ years relevant professional development experience, preferably in a LINUX environment.
* Strong SQL and NoSQL Database Knowledge: Oracle, PostgreSQL/MYSQL, and Mongo DB (or similar).
* Proficiency with at least 2 programming languages among Scala/Python/ Java. Must be flexible and fast to pick up new languages.
* Proficiency in Restful service development, preferably with Node JS, Django and PHP.
* Experience with micro-services and SOA.
* Strong SQL and NoSQL Database Knowledge: Oracle, PostgreSQL/MYSQL, and Mongo DB (or similar).
* Proficiency on installation and configuration of big data software and technology
* Knowledge of Hadoop, Spark, Kafka and other big data technology stacks and streaming tools.
* Familiarity with and the ability to leverage a wide variety of open source technologies and tools.
* Working knowledge of cloud architecture and implementation on Azure or AWS, is a big plus. Experience with server-less computing, creating VMs, cloud security, and other cloud services is also a big plus.
* Experience working in an Agile methodology. Experience working with JIRA is a plus.

***Who We Are:***

Over 35,000 employees strong, the mission of the Mount Sinai Health System is to provide compassionate patient care with seamless coordination and to advance medicine through unrivaled education, research, and outreach in the many diverse communities we serve.

Formed in September 2013, The Mount Sinai Health System combines the excellence of the Icahn School of Medicine at Mount Sinai with seven premier hospital campuses, including Mount Sinai Beth Israel, Mount Sinai Beth Israel Brooklyn, The Mount Sinai Hospital, Mount Sinai Queens, Mount Sinai Roosevelt, Mount Sinai St. Luke’s, and New York Eye and Ear Infirmary of Mount Sinai.

The Mount Sinai Health System is committed to the tenets of diversity and workforce that are strengthened by the inclusion of and respect for our differences. We offer our employees a highly competitive compensation and benefits package, a 403(b) savings plan, and much more.

The Mount Sinai Health System is an equal opportunity employer. We promote recognition and respect for individual and cultural differences, and we work to make our employees feel valued and appreciated, whatever their race, gender, background, or sexual orientation.

EOE Minorities/Women/Disabled/Veterans

**2. Big Data Consultant** (<https://www.amazon.jobs/en/jobs/554290/big-data-consultant>)

**Job ID**

554290

**Location**

US-NY-New York

**Posted Date**

7/5/2017

**Company**

Amazon Web Services, Inc.

**Job Description**

At Amazon Web Services (AWS), we’re hiring highly technical cloud computing architects to help our partners develop technical expertise and capacity, while also collaborating with our customers and partners on key engagements. Our consultants will deliver proof-of-concept projects, topical workshops, and lead implementation projects. These professional services engagements will focus on key customer solutions such as, web applications, enterprise applications, HPC, batch processing and big data, archiving and disaster recovery, education and government.

Professional Services engage in a wide variety of projects for customers and partners, providing collective experience from across the AWS customer base and are obsessed about strong success for the Customer. Our team collaborates across the entire AWS organization to bring access to product and service teams, to get the right solution delivered and drive feature innovation based upon customer needs.

Responsibilities include:

* **Expertise** - Collaborate with AWS field sales, pre-sales, training and support teams to help partners and customers learn and use [AWS services](http://aws.amazon.com/products/) such as Amazon Elastic Compute Cloud (EC2), Amazon Data Pipeline, S3, DynamoDB NoSQL, Relational Database Service (RDS), Elastic Map Reduce (EMR) and Amazon Redshift.
* **Solutions** - Deliver on-site technical engagements with partners and customers. This includes participating in pre-sales on-site visits, understanding customer requirements, creating consulting proposals and creating packaged Big Data service offerings.
* **Delivery** - Engagements include short on-site projects proving the use of AWS services to support new distributed computing solutions that often span private cloud and public cloud services. Engagements will include migration of existing applications and development of new applications using AWS cloud services.

**Basic Qualifications**

* BA/BS degree or equivalent experience; Computer Science or Math background preferred
* 5+ years’ experience of IT platform implementation in a highly technical and analytical role.
* Customer facing skills to represent AWS well within the customer’s environment and drive discussions with senior personnel regarding trade-offs, best practices, project management and risk mitigation
* Demonstrated ability to think strategically about business, product, and technical challenges in an enterprise environment.
* Deep understanding of database and analytical technologies in the industry including MPP databases, noSQL storage, Data Warehouse design, BI reporting and Dashboard development.
* Implementation and tuning experience for Apache Hadoop + tools such as Pig and Hive.
* Current hands-on implementation experience required; individual contributors only need apply.
* Strong verbal and written communications skills and ability to lead effectively across organizations.
* Ability to travel to client locations when needed.

**Preferred Qualifications**

* Hands on experience leading large-scale global data warehousing and analytics projects.
* Demonstrated industry leadership in the fields of database, data warehousing or data sciences.
* Implementation and tuning experience specifically using Amazon Elastic Map Reduce + tools such as Pig and Hive.
* Track record of implementing AWS services in a variety of distributed computing, enterprise environments.

Amazon is an Equal Opportunity-Affirmative Action Employer-inority/Female/Disability/Vet

**3. Database Analyst** (<https://www.nespressojobs.com/job/new-york/database-analyst/5736/5761202>)

**Job location:** New York, New York **Position type:** Professional **Req No:** 17006809

Our story began over 25 years ago with a simple but revolutionary idea – to create the perfect cup of coffee. An industry pioneer, Nestlé Nespresso has become an international reference for the highest quality coffee and an iconic symbol of refined elegance. We are now in more than 50 countries and our team has grown well beyond 7,000 employees. We are part of Nestlé SA in Switzerland, the world's largest food company. Nespresso USA continues to drive momentum and innovation in our market segment, and in North America, we're just getting started.

Nestle Nespresso USA is searching for a driven, detail oriented, Database Analyst to lead deep dive analyses and advanced analytics projects to support database/CRM activation. This role will work closely with the CRM Insights Manager to develop and execute analytics plans including: promotional dashboards, big picture business analyses, and other vendor driven projects.

**Responsibilities**

* Act as primary point of contact for database related questions for internal and external stakeholders
* Drive database and technological advancements to improve level and efficiency of analyses as well as database hygiene
* Generate queries and in general leverage data to address key business questions
* Mine data for insights and key trends that address business questions and inform strategic decisions
* Deepen understanding of the Nespresso consumer through customer level analysis
* Drive omni channel strategy through holistic, cross-channel analyses
* Communicate key insights and recommendations to stakeholders
* Determine where predictive and attribution modeling will best enhance the businesses core competencies
* Work with CRM Insights Manager to develop a deployment plan and implementation of the model
* Review business objectives and key trends with CRM Insights Manager and other CRM counterparts to develop predictive model-based segmentation strategies
* Create personalized communication strategy that increases consumers’ engagement with the brand and lifetime value

**Requirements**

* Bachelor’s Degree required, preferably in business, marketing, statistics, or computer science
* 4+ years of experience in an analytics focused role
* 2+ years of experience in direct to consumer marketing analytics
* Advanced SQL querying skills required
* Prior experience using data analysis tools required (SPSS, SAS, etc.)
* Prior experience using business intelligence tools strongly preferred (Cognos, SAP Business Objects, etc.)
* Prior experience using data visualization tools preferred (Tableau, Power BI, etc.)
* Strong analytical skills
* Ability to clearly deliver insights and findings
* Ability to integrate multiple data sources into a single narrative
* Strong strategic thinking skills
* Strong presentation skills

**The Nestle companies are equal opportunity and affirmative action employers and are looking for diversity in candidates for employment: Minority/Female/Disabled/Protected Veteran**

**4. Data Scientist** (<https://www.glassdoor.com/job-listing/data-scientist-bankers-healthcare-group-JV_IC1132348_KO0,14_KE15,39.htm?jl=2625375731>)

Bankers Healthcare Group  [20 reviews](https://www.indeed.com/cmp/Bankers-Healthcare-Group/reviews) - New York, NY

|  |
| --- |
| **Job Summary**  We are seeking a talented data scientist to work on revenue predictive analytics team at BHG.  You will provide advanced analytics and predictive modeling skills to drive business decision making. Utilize data mining techniques to understand revenue life circle of BHG customers. Automate complicated works and create efficiency.  You will be deeply involved in evaluation of new data source, and extracting, cleaning, modeling of the data and experimental design.  **Requirement**:   * 2-5 years of professional experience working as a Data Scientist * Master's degree, or PhD in Computer Science, Statistics, Mathematics, Engineering, Econometrics, or related fields from an accredited college or university * Proficient in statistics and machine learning, including exploratory data analysis, regression, classification, experimental design and clustering. * Proficient in extracting, cleaning, and modeling of the data. * 2-5 years of professional Python or R experience * 2-5 years of professional SQL experience * Detailed oriented and work well under pressure * Good communication skills and team player * Ideally 2 plus years marketing data science experience   **Job Type**: Full-time  **Required education**:   * Bachelor's   **Required experience**:   * Machine Learning: 2 years * R: 1 year * SQL: 2 years * Predictive Modeling: 2 years |

**5. Data Analyst** (<https://www.glassdoor.com/job-listing/data-analyst-popular-community-bank-JV_IC1132348_KO0,12_KE13,35.htm?jl=2626947239>)

Apply Now

Date: Jan 3, 2018

Location: New York, NY, US, 10004

**Join the Popular Team!**

Founded in 1893, Popular Inc. has been built upon strong institutional values while providing broad financial services within the United States, Puerto Rico, and the Caribbean.

We strive to create an extraordinary legacy with a passion for customer service, tremendous dedication to our employees and strong partnerships in the communities where we reside. Come explore Popular Community Bank and see why we are “The Human Side of Banking.”

**Strategic Planning**

We are seeking a **Data Analyst** who can combine strong analytic capabilities, database skills, and communication to help shape our reporting and analytical needs. The Analyst will need to be able to drive forward internal projects with a combination of business stakeholders, technical stakeholders, and support teams by being the bridge between business and systems. The Analyst will play an integral part in creating and enhancing business intelligence systems and frameworks. Please note that this position will be more technical and would be ideal to applicants who wish to begin or further their experience with coding, practice, and development of BI systems with financial institutions.

**In this position, you will:**

* Develop, analyze, and evaluate data to create and maintain business intelligence frameworks that supports the Bank’s strategic objectives;
* Demonstrate a sufficient understanding of BI systems and engage and communicate effectively both to non-technical personnel, as well as internal developers regarding regular business and QA requirements;
* Work with internal and external IT partners to ensure data feeds are accurate and delivered in a timely manner;
* Analyze and assess business problems and processes by applying conceptual knowledge and discretion to independently define functional and non-functional requirements for technology solutions, consult and make recommendations accordingly
* Maintain/update/create documentation of knowledge assets, e.g. business rules, technical specifications (data mapping, data flows, dashboard content, data dictionaries, relational diagrams, QA plans, etc.)
* Administer security for various applications
* Responsible for the timely resolution of various Compliance and Audit items.

**To be considered, you will need:**

* Bachelor’s degree required, preferred in Information Systems, Computer Science, or other technical background
* Minimum of 2-3 years in a similar position, or experience in data mining, date a management, financial analysis or other related fields in the financial services industry
* Knowledge of business intelligence applications and/or statistical analysis software, e.g., Microsoft Office (Excel VBA), SAS, Tableau, SQL, Power Pivot, Business Objects Desktop/Web Intelligence
* Familiarity with the IBM Cognos TM1 platform is preferred, but not necessarily required
* Understand concepts of database structures, data querying, ETL, data mining, data auditing and technical documentation is preferred
* The ability to deal with and resolve conflicting and competing demands, prioritize and organize work to meet deadlines and collaborate with a variety of stakeholders with little supervision
* Flexibility with general ambiguity as it refers to stakeholder requirements, needs, and daily work as this position interfaces with many departments and stakeholders

**6. Data Scientist, Analytics** (New York, NY)

Facebook's mission is to give people the power to build community and bring the world closer together. Through our family of apps and services, we're building a different kind of company that connects billions of people around the world, gives them ways to share what matters most to them, and helps bring people closer together. Whether we're creating new products or helping a small business expand its reach, people at Facebook are builders at heart. Our global teams are constantly iterating, solving problems, and working together to empower people around the world to build community and connect in meaningful ways. Together, we can help people build stronger communities — we're just getting started.

We’re looking for Data Scientists to work on our core and business products (Instagram, Ads, Messaging, Identity, Growth & Engagement, Mobile, Search, Privacy, Payments) with a passion for Internet technology to help drive informed business decisions for Facebook. You will enjoy working with one of the richest data sets in the world, cutting edge technology, and the ability to see your insights turned into real products on a regular basis. The perfect candidate will have a background in a quantitative or technical field, will have experience working with large data sets, and will have some experience in data-driven decision making. You are scrappy, focused on results, a self-starter, and have demonstrated success in using analytics to drive the understanding, growth, and success of a product. This position is located in our New York City office.

**Responsibilities**

* Apply your expertise in quantitative analysis, data mining, and the presentation of data to see beyond the numbers and understand how our users interact with both our consumer and business products
* Partner with Product and Engineering teams to solve problems and identify trends and opportunities
* Inform, influence, support, and execute our product decisions and product launches
* The Data Scientist Analytics role has work across the following four areas:
* Product Operations
* Forecasting and setting product team goals
* Designing and evaluating experiments
* Monitoring key product metrics, understanding root causes of changes in metrics
* Building and analyzing dashboards and reports
* Building key data sets to empower operational and exploratory analysis
* Evaluating and defining metrics
* Exploratory Analysis
* Proposing what to build in the next roadmap
* Understanding ecosystems, user behaviors, and long-term trends
* Identifying new levers to help move key metrics
* Building models of user behaviors for analysis or to power production systems
* Product Leadership
* Influencing product teams through presentation of data-based recommendations
* Communicating state of business, experiment results, etc. to product teams
* Spreading best practices to analytics and product teams
* Data Infrastructure
* Working in Hadoop and Hive primarily, sometimes MySQL, Oracle, and Vertica
* Automating analyses and authoring pipelines via SQL and Python based ETL framework

**Minimum Qualifications**

* 5+ years of experience doing quantitative analysis.
* BA/BS in Computer Science, Math, Physics, Engineering, Statistics or other technical field. Advanced degrees.
* Experience in SQL or other programming languages.
* Development experience in any scripting language (PHP, Python, Perl, etc.)
* Ability to communicate the results of analyses.
* Understanding of statistics (e.g., hypothesis testing, regressions).
* Experience manipulating data sets through statistical software (ex. R, SAS) or other methods.

**Preferred Qualifications**

* Experience with distributed computing (Hive/Hadoop)

**7. Data Science Analyst I - (Mount Sinai Hospital) Day Shift** (<https://www.glassdoor.com/job-listing/data-science-analyst-i-mount-sinai-hospital-day-shift-mount-sinai-queens-JV_IC1132348_KO0,53_KE54,72.htm?jl=2610877506>)

The Mount Sinai Hospital

**The Mount Sinai Health System**

Do you have what it takes to wear the badge?

The Mount Sinai Health System's commitment to excellence extends beyond delivering world-class health care. The System's ongoing success is dependent upon our highly motivated, nonclinical professionals working to improve business operations. Our leadership team is driven to provide exceptional service by cultivating a workforce that is dedicated to upholding Mount Sinai's mission of delivering innovative, breakthrough medicine with compassion and integrity.

Are you ready to discover the world of limitless possibilities that comes with wearing the badge? Explore more about this opportunity and how you can help us write a new chapter in our story of unrivaled patient care!

***What You'll Do:***

The Data Science Analyst for Access/Transfer Center collaborates with stakeholders from across the organization to develop sophisticated analytics to provide information, insights and BI (Business Intelligence) solutions that contribute to sound strategic planning, decision-making, goal setting, and effective performance measurement. This individual demonstrates sound understanding of the healthcare domain, technical data manipulation and analytic development skills and impact the patient community of the Mount Sinai Health System. This analyst will be working at Mount Sinai Corporate Office near Grand Central.

**Duties and Responsibilities**:

Analyzes data requests for Access/Transfer Center using information technology, enrollment, claims, clinical, contract, medical management, financial, administrative and other corporate data from both modeled and disparate internal and external sources.

1. Works with departmental staff to identify requirements for reporting and / or business intelligence tools.
2. Identifies necessary data, data sources and methodologies.
3. Collects, organizes, integrates, analyzes and interprets data.
4. Leverages advanced statistical analysis methods to create insightful recommendations and conclusions that may be communicated to the stakeholder.
5. Identifies and addresses expected and unforeseen data complexities to mitigate their impact on the analytic outcome and associated business decisions. Works to improve data quality where possible within created analytical models. Feeds data quality issues back to IT or identified data stewards to facilitate creation of high quality metrics.
6. Develops and may present reports, analyses and findings to senior management and others as scheduled or requested.

**Responsible for one or more of the following stakeholder groups:**

1. Contracting and Commercialization – May assist in the modeling and forecasting contract scenarios, measuring ongoing performance and identify trends in performance to inform our clinical or contracting staff to improve contract outcomes.
2. Care Management – Helps to identify, understand and prioritize at-risk members in need of care management. Helps stratify our membership to optimally use resources to focus on the patients most in need, currently or in the future.
3. Medical Directors – Helps to identify utilization trends and variations across the different categories of health care services to assist the Medical Directors to focus their efforts to maximize contract performance and clinical effectiveness.
4. Quality and Documentation – Helps to link payer quality and documentation opportunities into operational analytic processes to maximize our quality scores, top line revenue and optimize the use of resources in concert with MS Health System contracts.
5. I.T. / High Performance Computing in any ongoing projects.
6. Acts as liaison/analyst for internal stakeholders, understands their needs and translates them into reporting and analytic solutions.
7. Effectively communicates with stakeholders and customers and ensures all requests are properly triaged, recorded and tracked.
8. Adheres to corporate standards for performance metrics, data collection, data integrity, query design, and reporting format to ensure high quality, meaningful analytic output.
9. Helps identify and understand data from internal and external sources for competitive, scenario and performance analyses, and financial modeling to gain member/provider insight into new and existing processes and business opportunities.
10. Works closely with IT on the ongoing improvement of Mount Sinai's integrated data warehouse, driven by strategic and business needs, and designed to ensure data and reporting consistency throughout the organization.

Other duties assigned as per Director.

**Education**:

BA or BS degree minimum, in a relevant field of study; Masters degree preferred.

**Experience**:

* years minimum in analytics development expertise, preferably in health care, or for a health provider, health plan or accountable care organization, including either:
* Working knowledge of a health care EMR such as Epic/Clarity, aCW, etc.; a payor claims system such as Facets, Amisys, etc.; or a hospital/provider system such as IDX, Soarian, etc.
* Knowledge of the New York State Medicaid and CMS Medicare regulations and related reporting requirements, such as STARS,QARR, MMCOR, MEDS, RAPS and HEDIS is a strong plus.
* Experience working in a health plan or consulting actuarial, financial reporting or medical economics departments highly valuable.
* Experience working in healthcare provider analytics related to revenue modeling, managed care contracting, population management, case management, clinical or financial decision report

# Appendix C: Draft Articulation Agreements

**1. Draft Articulation with Borough of Manhattan Community College (BMCC)**

BOROUGH OF MANHATTAN COMMUNITY COLLEGE

City University of New York

and

New York City College of Technology

City University of New York

Borough of Manhattan Community College, hereafter known as Borough of Manhattan, and New York City College of Technology, hereafter known as City Tech, enter into the following articulation agreement for the transfer of credit and courses, effective \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_. This agreement will involve the Department of Computer Systems Technology, City Tech and the Department of Computer Information Systems, Borough of Manhattan. It is restricted to students enrolled in the Computer Science programs offered at Borough of Manhattans which leads to an Associate in Sciences Degree.

Borough of Manhattan students with the BS degree in Computer Science who meet City Tech’s specified criteria, including the passage of the CUNY Basic Skills Assessment Tests in Mathematics, Reading, and Writing, shall be admitted as candidates for the flowing upper level baccalaureate degree:

Data Science Bachelor of Science (B.S.) 60 Credits

Acceptance into this program will be under the requirements in effect at the time of admission and may be subject to such changes as shall be determined by City Tech’s academic policies and curricula. To earn a baccalaureate degree, admitted students must earn a minimum of 60 credits which 34 must be taken in residence. Certification of graduation is assured upon completion of a minimum of 60 credits with a cumulative index of 2.0. This 2.0 index is also required in the major, the minor and/or interdisciplinary program.

City Tech agrees to accept the following Borough of Manhattan courses as the equivalent to City Tech courses offered in the Bachelor of Science in Data Science:

|  |  |
| --- | --- |
| City Tech Courses:  B.S. in Data Science | Borough of Manhattan Course:  Computer Science |
| CST1100 | CSC101 |
| CST1101 | CSC215 |
| CST1201 | CSC111 |
| CST3513 | CSC211 |
| CST3650 | CSC331 |
| CST2301 | CSC350 |
| MAT2440 | CSC231 |

To satisfy the Bachelor of Science degree requirements in Data Science at City Tech 42 credits in major are required as follow:

|  |  |  |
| --- | --- | --- |
| CST1204 | 3 | Database Fundamentals |
| CST2312 | 3 | Information and Data Management I |
| CST2309 | 3 | Web Programming I |
| CST2402 | 3 | Introduction to Data Science |
| CST3512 | 3 | Information and Data Management II |
| CST3502 | 3 | Data Mining |
| CST3602 | 3 | Data Visualization |
| CST4702 | 3 | Machine Learning Fundamentals |
| CST4802 | 3 | Information Retrieval |
| CST4812 | 3 | Natural Language Processing |
| Two Electives from the following: | | |
| BUS2339 | 3 | Financial Management |
| BUS2341 | 3 | Financial Forecasting |
| MED2400 | 3 | Medical Informatics Fundamentals |
| MED4229 | 3 | Healthcare Databases |
| BMET4741 | 3 | Fundamental Healthcare Data Analytics |
| BMET4842 | 3 | Advanced Healthcare Data Analytics |
| ECON1101 | 3 | Macroeconomics |
| ECON2301 | 3 | Money and Banking |

In addition to satisfy the Bachelor of Science in Data Science degree requirements in Computer Systems at City Tech the 42-credit distribution in the Liberal Arts & Sciences Baccalaureate Core must be fulfilled. Up to 30 credits of these courses satisfying the Liberal Arts & Sciences Baccalaureate Core may be included in the courses earned at Borough of Manhattan and transferred to City Tech. Borough of Manhattan students should take the following Liberal Arts courses. The following courses must be taken at Borough of Manhattan. If not taken at Borough of Manhattan, these courses would then be taken at City Tech.

General Education Core

|  |  |  |  |
| --- | --- | --- | --- |
| Required Common Core | English Comp 1 | ENG 1101 | 3 |
|  | English Comp 2 | ENG 1121 | 3 |
|  | Quantitative reasoning | Any (MAT1475 - Strongly Recommended) | 3 |
|  | Life & Physical Sciences | Any | 3 |
| Flexible Common Core | World Cultures | Any | 3 |
|  | US Experience | Any | 3 |
|  | Creative Expression | Any | 3 |
|  | Individual and Society | Any | 3 |
|  | Scientific World | Any | 3 |
|  | Additional 6th course | Any | 3 |
| College Option | Speech/Oral Communication | Any | 3 |
|  | Interdisciplinary Course | Any | 3 |
|  | Additional LA I | Any (MAT2440 - Strongly Recommended) | 3 |
|  | Additional LA II | Any (MAT2540 - Strongly Recommended) | 3 |
|  | SUB-TOTAL |  | 42 |
| Program GenEd | Calculus 2 | MAT1575 S2 | 4 |
| Required Courses | Calculus 3 | MAT2675 S3 | 4 |
|  | Statistics and Probability | MAT2572 S5 | 4 |
|  | Probability and Statistics II | MAT3672 | 4 |
|  | Linear Algebra | MAT2580 S6 | 3 |
|  | SUB-TOTAL |  | 19 |
| Gen Ed Electives  to complete 120 credits | Differential Equations | MAT2680 | 3 |
| Numerical Methods | MAT2630 | 3 |
|  | Computational Statistics with Application | MAT4672 | 3 |

In order to facilitate the efficient transition between our institutions, interested Borough of Manhattan students are invited to utilize the pre-transfer advisement services of City Tech. Such services may be performed at City Tech, or, by pre-arrangement, on-site at Borough of Manhattan. Successful graduates are also assured of availability to all ancillary services at City Tech.

Both parties agree to notify each other of any changes in their respective programs. Such changes will commence discussions to as to how best to amend this agreement.

All the preceding agreements and commitments are herein certified as binding upon both Borough of Manhattan and City Tech until declared terminated by mutual consent.

FOR Borough of Manhattan FOR New York City College of Technology

Community College

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Sadie Bragg Bonne August

Vice President – Academic Affairs Provost and Vice President of Academic Affairs

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Ching-Song Wei Hong Li

Chairperson, Computer Information Chairperson, Computer Systems

Systems Department Technology Department

**2. Draft Articulation with La Guardia Community College**

ARTICULATION AGREEMENT

(DRAFT)

between

La Guardia Community College

City University of New York

and

New York City College of Technology

City University of New York

La Guardia Community College, hereafter known as LAGCC, and New York City College of Technology, hereafter known as City Tech, enter into the following articulation agreement for the transfer of credit and courses, effective \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_. This agreement will involve the Department of Computer Systems Technology, City Tech and the Department of Math, Engineering and Computer Science, La Guardia Community College. It is restricted to students enrolled in the Computer Science programs offered at LAGCC which leads to an Associate in Sciences Degree.

LAGCC students with the AS degree in Computer Science who meet City Tech’s specified criteria, including the passage of the CUNY Basic Skills Assessment Tests in Mathematics, Reading, and Writing, shall be admitted as candidates for the flowing upper level baccalaureate degree:

Data Science Bachelor of Science (B.S.) 60 Credits

Acceptance into this program will be under the requirements in effect at the time of admission and may be subject to such changes as shall be determined by City Tech’s academic policies and curricula. To earn a baccalaureate degree, admitted students must earn a minimum of 60 credits which 34 must be taken in residence. Certification of graduation is assured upon completion of a minimum of 60 credits with a cumulative index of 2.0. This 2.0 index is also required in the major, the minor and/or interdisciplinary program.

City Tech agrees to accept the following LAGCC courses as the equivalent to City Tech courses offered in the Bachelor of Science in Data Science:

|  |  |
| --- | --- |
| **City Tech Courses: B.S. in Data Science** | **LAGCC Computer Science** |
| CST1100 | MAC283 |
| CST1101 | MAC101 |
| CST1201 | MAC125, MAC108 |
| CST3513 | MAC190 |
| CST3650 | MAC286 |
| CST2301 | CSC350 |
| MAT2440 | MAT231 |
| MAT2540 | MAC281 |
| MAT1575 | MAT202 |
| MAT2580 | MAT210 |

To satisfy the Bachelor of Science degree requirements in Data Science at City Tech 42 credits in major are required as follow:

|  |  |  |
| --- | --- | --- |
| CST1204 | 3 | Database Fundamentals |
| CST2312 | 3 | Information and Data Management I |
| CST2309 | 3 | Web Programming I |
| CST2402 | 3 | Introduction to Data Science |
| CST3512 | 3 | Information and Data Management II |
| CST3502 | 3 | Data Mining |
| CST3602 | 3 | Data Visualization |
| CST4702 | 3 | Machine Learning Fundamentals |
| CST4802 | 3 | Information Retrieval |
| CST4812 | 3 | Natural Language Processing |
| Two Electives from the following: | | |
| BUS2339 | 3 | Financial Management |
| BUS2341 | 3 | Financial Forecasting |
| MED2400 | 3 | Medical Informatics Fundamentals |
| MED4229 | 3 | Healthcare Databases |
| BMET4741 | 3 | Fundamental Healthcare Data Analytics |
| BMET4842 | 3 | Advanced Healthcare Data Analytics |
| ECON1101 | 3 | Macroeconomics |
| ECON2301 | 3 | Money and Banking |

In addition to satisfy the Bachelor of Science in Data Science degree requirements in Computer Systems at City Tech the 42-credit distribution in the Liberal Arts & Sciences Baccalaureate Core must be fulfilled. Up to 30 credits of these courses satisfying the Liberal Arts & Sciences Baccalaureate Core may be included in the courses earned at LAGCC and transferred to City Tech. LAGCC students should take the following Liberal Arts courses. The following courses must be taken at LAGCC. If not taken at LAGCC, these courses would then be taken at City Tech.

General Education Core

|  |  |  |  |
| --- | --- | --- | --- |
| Required Common Core | English Comp 1 | ENG 1101 | 3 |
|  | English Comp 2 | ENG 1121 | 3 |
|  | Quantitative reasoning | Any (MAT1475 - Strongly Recommended) | 3 |
|  | Life & Physical Sciences | Any | 3 |
| Flexible Common Core | World Cultures | Any | 3 |
|  | US Experience | Any | 3 |
|  | Creative Expression | Any | 3 |
|  | Individual and Society | Any | 3 |
|  | Scientific World | Any | 3 |
|  | Additional 6th course | Any | 3 |
| College Option | Speech/Oral Communication | Any | 3 |
|  | Interdisciplinary Course | Any | 3 |
|  | Additional LA I | Any (MAT2440 - Strongly Recommended) | 3 |
|  | Additional LA II | Any (MAT2540 - Strongly Recommended) | 3 |
|  | SUB-TOTAL |  | **42** |
| Program GenEd | Calculus 2 | MAT1575 S2 | 4 |
| Required Courses | Calculus 3 | MAT2675 S3 | 4 |
|  | Statistics and Probability | MAT2572 S5 | 4 |
|  | Probability and Statistics II | MAT3672 | 4 |
|  | Linear Algebra | MAT2580 S6 | 3 |
|  | SUB-TOTAL |  | **19** |
| Gen Ed Electives  to complete 120 credits | Differential Equations | MAT2680 | 3 |
| Numerical Methods | MAT2630 | 3 |
| Computational Statistics with Application | MAT4672 | 3 |

In order to facilitate the efficient transition between our institutions, interested LAGCC students are invited to utilize the pre-transfer advisement services of City Tech. Such services may be performed at City Tech, or, by pre-arrangement, on-site at LAGCC. Successful graduates are also assured of availability to all ancillary services at City Tech.

Both parties agree to notify each other of any changes in their respective programs. Such changes will commence discussions to as to how best to amend this agreement.

All the preceding agreements and commitments are herein certified as binding upon both LAGCC and City Tech until declared terminated by mutual consent.

For La Guardia Community College For City Tech

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Paul Arcario, Provost and Senior Bonne August

Vice President – Academic Affairs Provost and Vice President of Academic Affairs

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Abderrazak Belkharraz, Hong Li

Chairperson, Department of Mathematics, Chairperson, Computer Systems

Engineering and Computer Science Technology Department

# Appendix D: Colleges Offering Degrees in Data Science

1. **Sample Bachelor of Science in Data Science Programs**

* NYU School of Professional Studies
  + <http://www.sps.nyu.edu/academics/departments/mcghee/undergraduate/bachelors/bs-applied-data-analytics-and-visualization.html>
* Columbia University
  + <https://mice.cs.columbia.edu/c/d.php?d=245>
* Penn State University
  + <https://datasciences.psu.edu/>
* Drexel University
  + <http://catalog.drexel.edu/undergraduate/collegeofcomputingandinformatics/datascience/>
* George Mason University
  + <https://cos.gmu.edu/cds/bs-in-computational-and-data-sciences/>
* Purdue University
  + <http://catalog.purdue.edu/preview_program.php?catoid=8&poid=10127>
* University of Massachusetts - Dartmouth
  + <http://www.umassd.edu/programs/data-science/>
* University of Michigan
  + <https://www.eecs.umich.edu/eecs/undergraduate/data-science/>
* University of Rochester
  + <http://www.sas.rochester.edu/dsc/undergraduate/major.html>
* University of San Francisco, College of Arts and Science <https://www.usfca.edu/arts-sciences/undergraduate-programs/data-science>
* University of California Irvine (UCI)
  + <http://datascience.uci.edu/data-science-degree/>
* University of Wisconsin River Falls <https://www.uwrf.edu/Academics/Undergraduate/Data-Science-Degree.cfm>
* Luther College
  + <http://www.luther.edu/computer-science/data-science-major/>
* Valparaiso University
  + <https://www.valpo.edu/mathematics-statistics/academics/degree-programs/b-s-in-data-science/>
* Becker College
  + <https://www.becker.edu/academics/undergrad/school-of-design-technology/data-science>
* Smith College
  + <https://www.smith.edu/sds/major.php>
* Winona State University
  + <https://www.winona.edu/math-stat/data-science.asp>
* University of Nebraska
  + <https://www.unomaha.edu/college-of-arts-and-sciences/mathematics/_files/datascience.pdf>
* College of Charleston
  + <http://datascience.cofc.edu/program-information/index.php>
* Marist College
  + <http://www.marist.edu/compscimath/undergraduate/data-science-and-analytics.html>

**2. Sample Masters of Science in Data Science Programs**

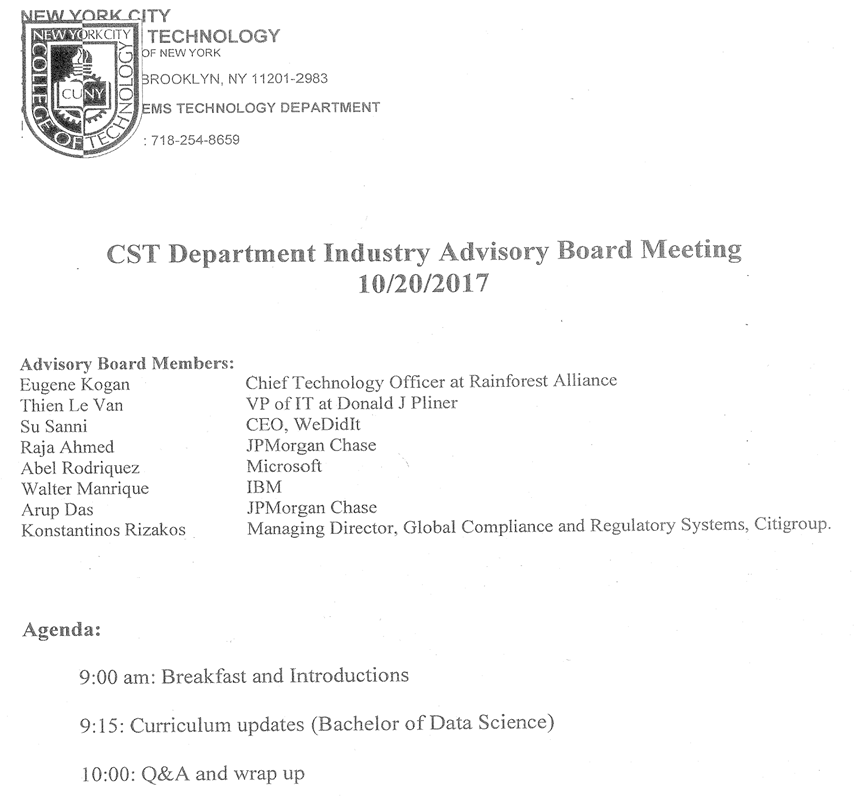
* CUNY - Graduate Center
  + <https://www.gc.cuny.edu/Page-Elements/Academics-Research-Centers-Initiatives/Masters-Programs/Data-Science>
* Columbia University
  + <http://datascience.columbia.edu/master-of-science-in-data-science>
* New York University
  + <https://cds.nyu.edu/academics/ms-in-data-science/>
* University of Rochester
  + <http://www.sas.rochester.edu/dsc/graduate/ms.html>
* Cornell
  + <http://stat.cornell.edu/academics/mps>
* Fordham University
  + <https://www.fordham.edu/info/25661/master_of_science_in_data_analytics/9052/courses_and_degree_requirements>
* Rutgers
  + <https://mbs.rutgers.edu/program/analytics-discovery-informatics-data-sciences>
* Stanford
  + <https://statistics.stanford.edu/academics/ms-statistics-data-science>
* University of California - San Diego
  + <http://jacobsschool.ucsd.edu/mas/dse/>
* University of Illinois at Urbana-Champaign
  + <https://online.illinois.edu/mcs-ds>
* University of San Francisco, College of Arts and Science
  + <https://www.usfca.edu/arts-sciences/graduate-programs/data-science>
* University of Southern California
  + <https://viterbigradadmission.usc.edu/programs/masters/msprograms/computer-science/ms-cs-data-science/>
* UMass Amherst
  + <https://www.cics.umass.edu/grads/data-science-concentration-elective-requirements>

**COURSES IN DATA SCIENCE PROGRAMS OF 12 COLLEGES/UNIVERSITIES**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Courses** | **Rochester** | **Penn State** | **Univ. of SF** | **UC**  **Berkeley** | **UC Irvine** | **Univ. of Wisconsin River Falls** | **Luther College** | **Smith College** | **Univ. of Michigan** | **Winona State Univ..** | **College of Charleston** | **Univ. of Nebraska** | **Count** |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| **Statistics** | **X** | **X** | **X** | **X** | **X** |  | **X** | **X** | **X** | **X** | **X** | **X** | **14** |
| **Programming/**  **Algorithms** | **X** | **X** | **X** |  | **X** | **X** | **X** | **X** | **X** | **X** | **X** |  | **13** |
| **Data Visualization** |  | **X** | **X** | **X** | **X** | **X** | **X** |  |  | **X** |  | **X** | **11** |
| **Artificial Intelligence and machine learning** | **X** | **X** |  | **X** | **X** | **X** | **X** | **X** | **X** |  | **X** |  | **11** |
| **Calculus** | **X** | **X** | **X** |  |  | **X** |  |  | **X** | **X** | **X** |  | **10** |
| **Data Structures** | **X** |  | **X** | **X** | **X** | **X** |  | **X** | **X** | **X** | **X** |  | **9** |
| **Data Mining** | **X** | **X** | **X** | **X** |  |  |  | **X** | **X** |  | **X** |  | **9** |
| **Linear algebra** | **X** |  | **X** |  |  | **X** |  | **X** |  |  | **X** | **X** | **9** |
| **Database Management Systems/Data Warehousing** | **X** |  |  | **X** |  | **X** | **X** |  | **X** | **X** |  |  | **9** |
| **Probability** | **X** | **X** | **X** | **X** | **X** |  |  |  | **X** |  |  | **X** | **8** |
| **Discrete Math** |  | **X** | **X** |  |  | **X** |  |  | **X** |  | **X** |  | **7** |
| **Applied or Data Analytics** | **X** | **X** |  |  | **X** | **X** | **X** |  |  |  |  |  | **7** |
| **Business Intelligence and Data Management** |  | **X** |  |  | **X** |  |  |  |  | **X** |  |  | **4** |
| **Visual Analytics** |  |  |  |  |  |  |  | **X** |  |  |  |  | **2** |
| **Knowledge Organization** |  | **X** |  |  |  |  |  |  |  |  | **X** |  | **2** |
| **Privacy/ Security/ Ethics** |  | **X** |  | **X** |  |  |  |  |  |  |  |  | **2** |
| **Mathematical Modeling** |  |  | **X** |  |  |  |  |  |  |  |  |  | **1** |
| **Microeconomics** |  |  | **X** |  |  |  |  |  |  |  |  |  | **1** |
| **Designing Data – Infographics** |  |  |  |  |  |  |  |  |  |  |  |  | **1** |
| **HCI** |  |  |  |  | **X** |  |  |  |  |  |  |  | **1** |
| **Information Retrieval** |  |  |  |  | **X** |  |  |  |  |  |  |  | **1** |
| **Networking** |  |  |  |  |  |  |  |  |  |  |  |  | **1** |
| **Systems Analysis** |  |  |  |  |  |  |  |  |  |  |  |  | **1** |

|  |  |
| --- | --- |
| **COMPUTER COURSES:**   * Programming/Algorithms * Data Visualization * Artificial Intelligence/Machine Learning * Data Mining * Data Base Management Systems/Data Warehousing * Business Intelligence/Data Management | **MATH COURSES:**   * Statistics * Calculus * Data Structures * Linear Algebra * Probability * Discrete Math * Applied Analytics |

# Appendix E: CST Industry Advisory Board Meeting



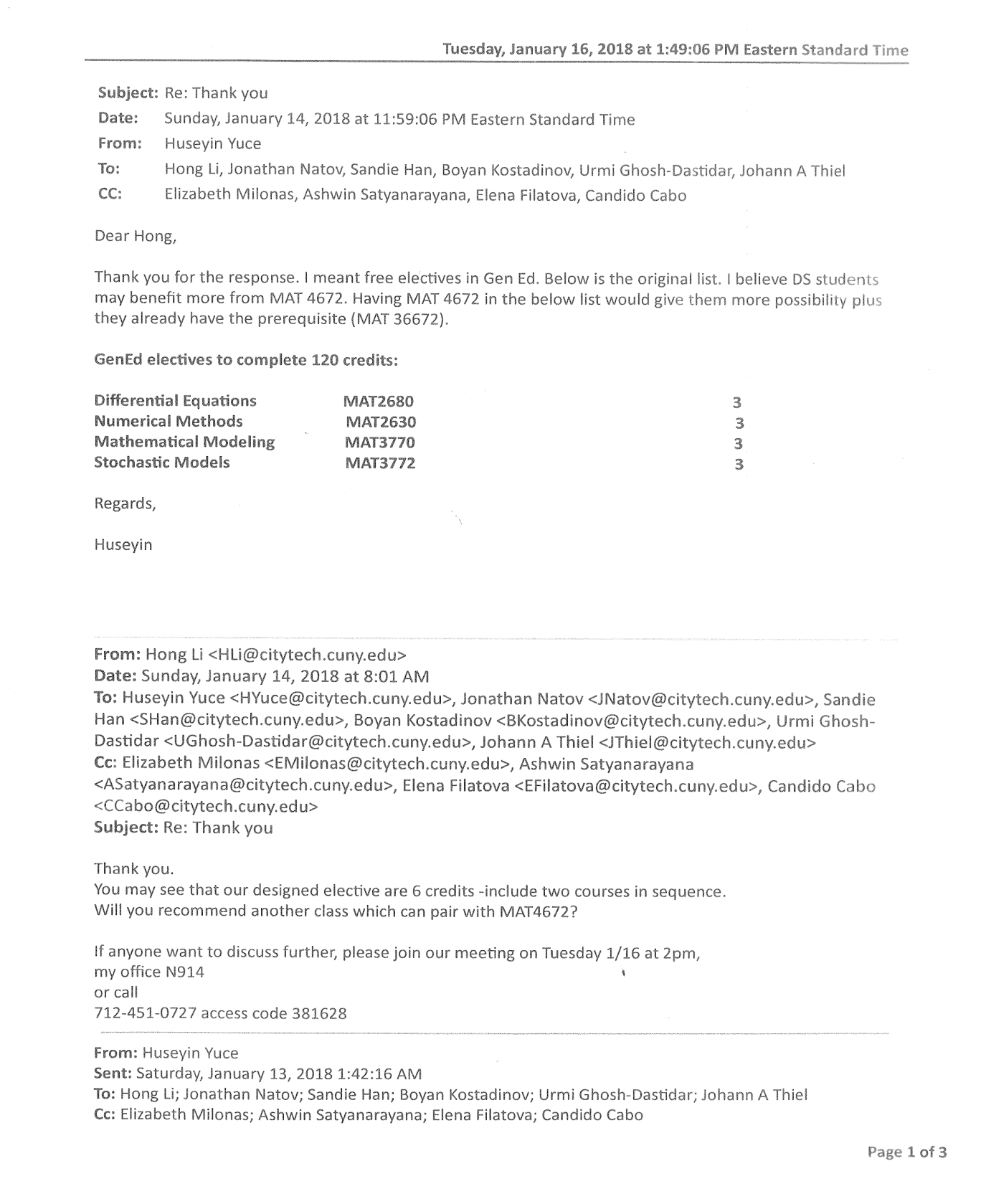
**Minutes**

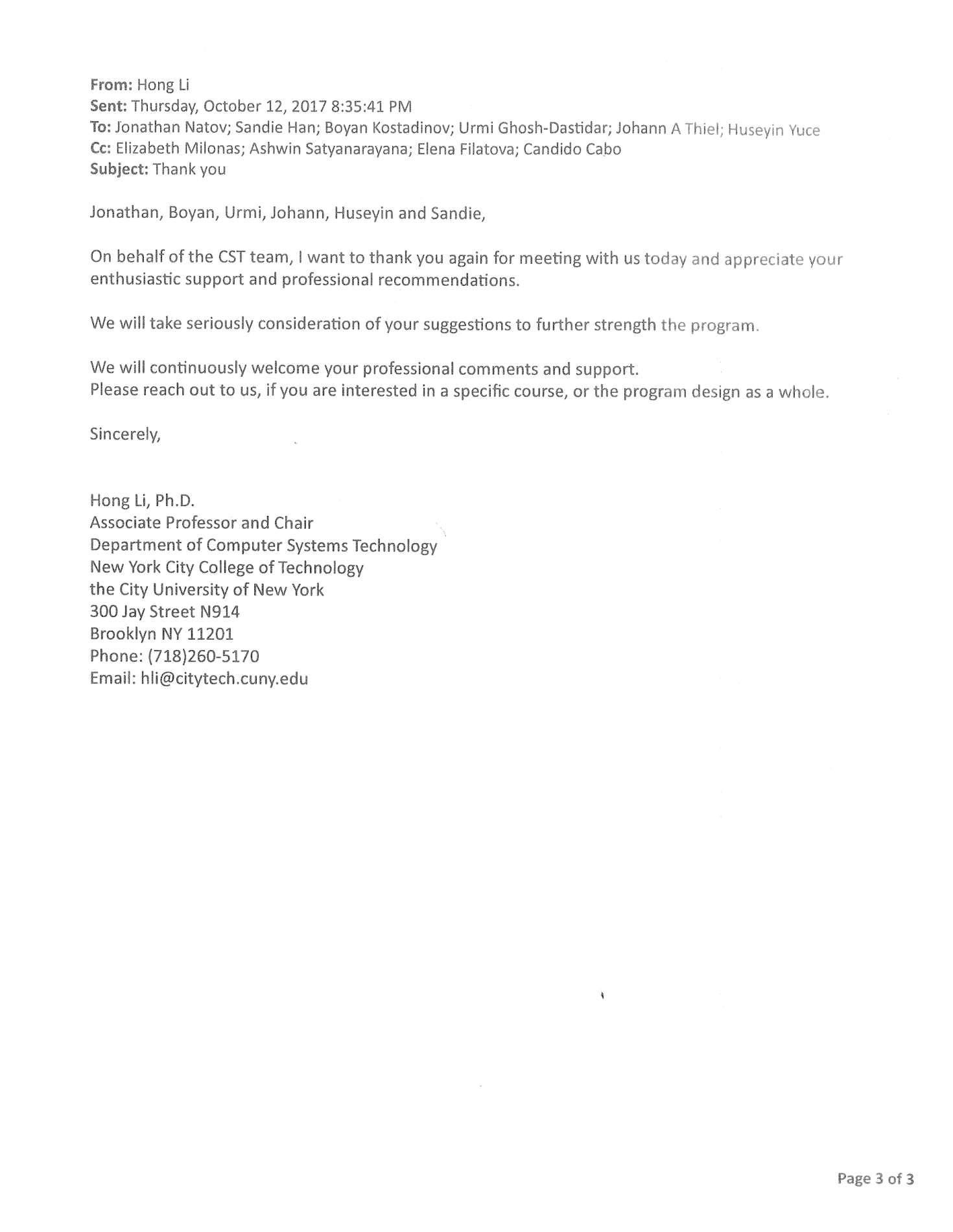
**In attendance:** Jean Boulet, Yu-Wen Chen, Arup Das (via Skype), Ossama Elhadary, Elena Filatova, Tony Holley, Eugene Kogan, Hong Li, Elizabeth Milonas, Angel Rodriques, Ashwin Satyanarayana, Fangyang Shen

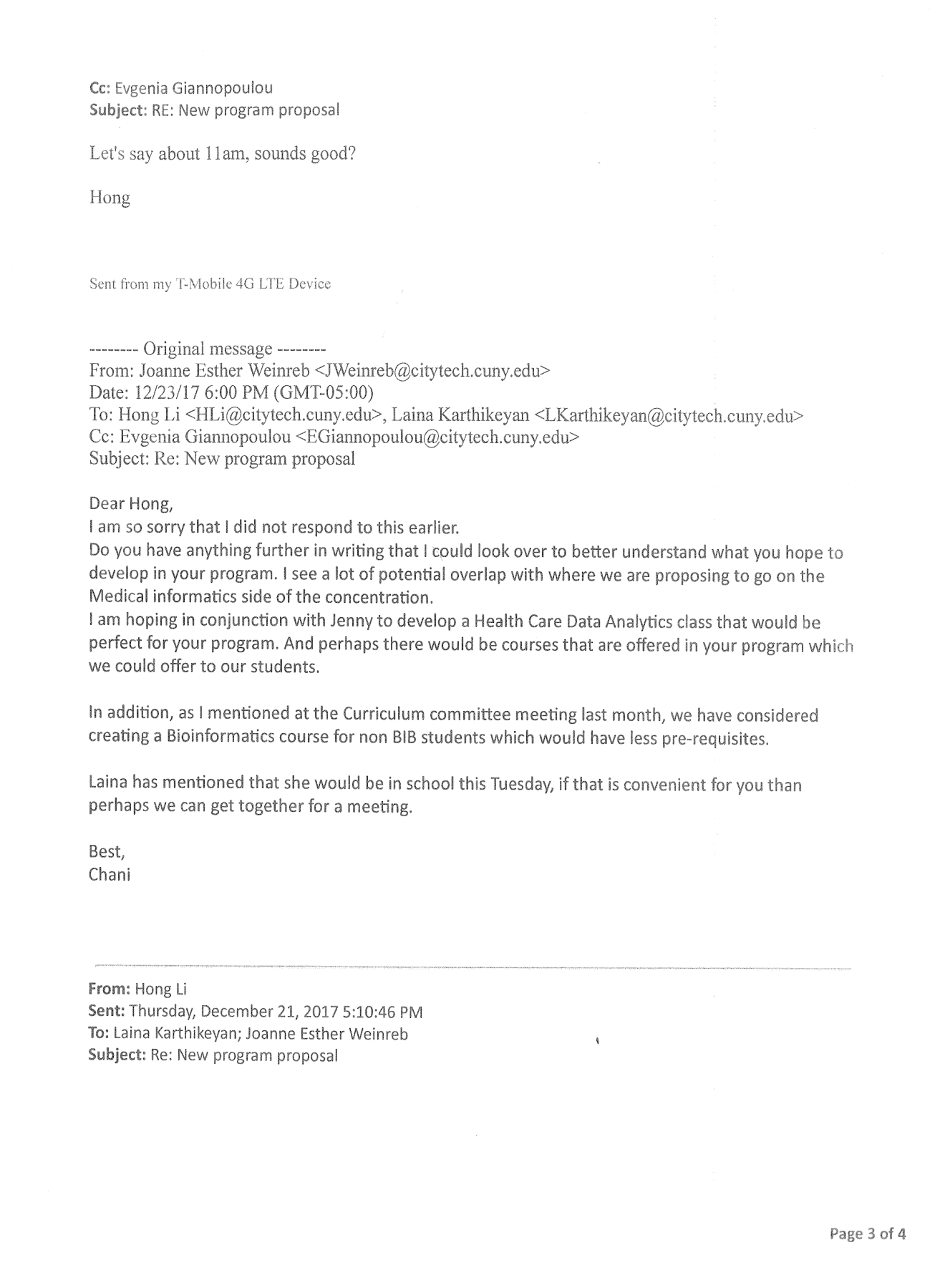
* Meeting was called to order at 9:00am
* A PowerPoint presentation of the Bachelor of Science in Data Science was given
* Handouts were distributed which included:
  + Bachelor of Science in Data Science curriculum spreadsheet identifying General Education courses, elective courses and required core courses
  + Brief descriptions of new Data Science courses
* Mr. Rodriquez comments:
  + Curriculum should include more machine learning topics/courses
  + Priority should be given on the data science tools/software/languages which are currently being used and are popular in industry
  + Recommended the addition of courses or topics in the following areas
    - Artificial Intelligence (Bots)
    - Decision engines
    - Recommendation engines
  + Curriculum has too many Calculus courses – remove Calculus 3
  + Add additional statistics course
  + Students should be familiar with financial industry data science needs and applications
* Mr. Das comments:
  + Curriculum should include
    - a focus on the use of data analytics and autonomous programming
    - logical thinking
    - broad understanding of popular industry tools available and the difference among these tools
  + Create partnerships with industry in order to place graduates once they have successfully completed the program
  + Will send information regarding JP Morgan’s Hackathon program to Hong
* Mr. Kogan comments:
  + Machine learning course (CST4702) should include tools such as Einstein and Watson
  + Students should be familiar with R
  + Entrance requirements should be carefully considered. Warned that students who may have the requirements for the math and science may not have the requirements for the general education courses such as English
* Ossama comments:
  + Create an Enterprise Application course which will include the various tools used in the data science domain
  + Evaluate the economics electives to include econometrics and courses more in line to analysis
* Meeting adjourned 10:00am

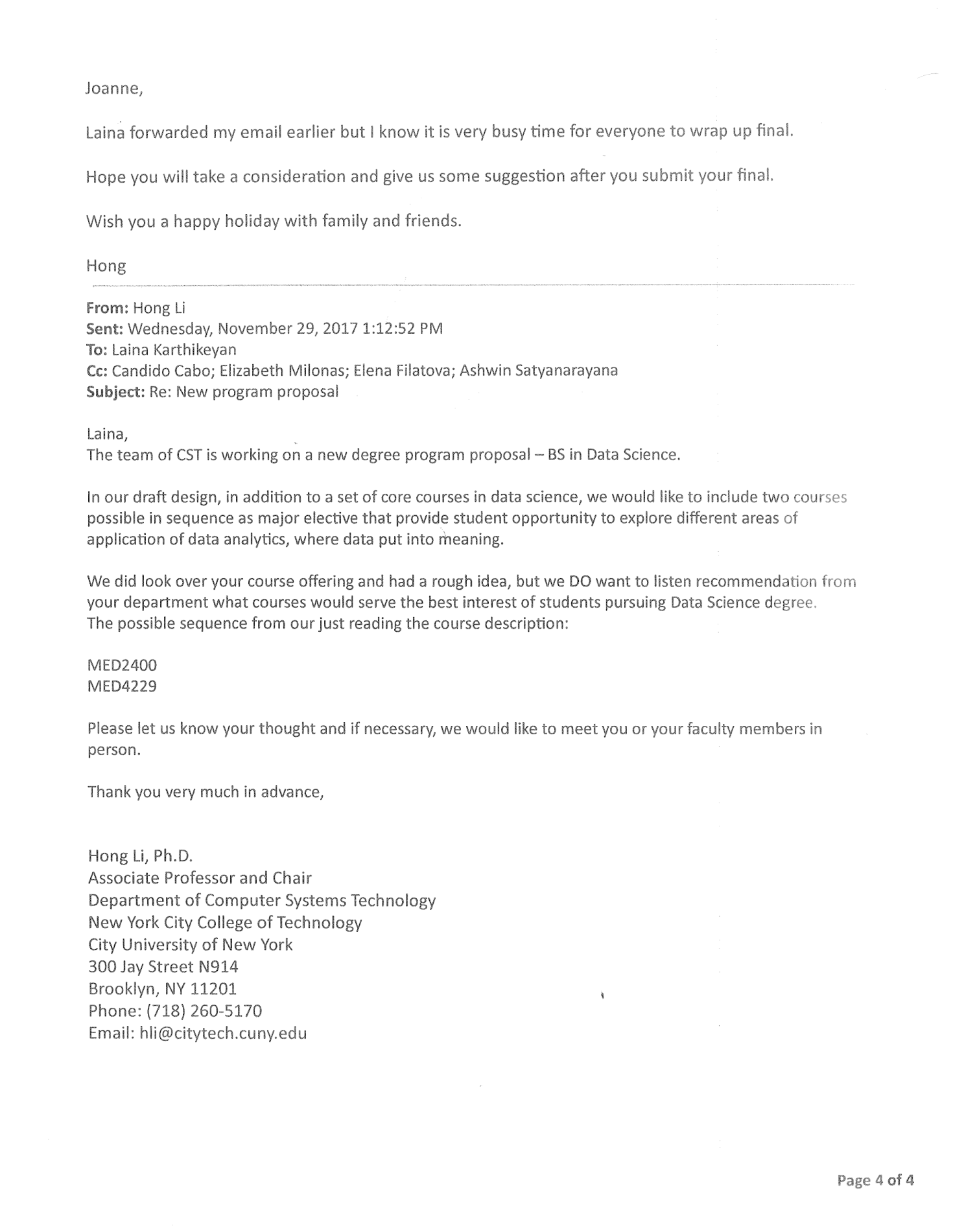
# Appendix F: Evidence of Consultation with Other Departments

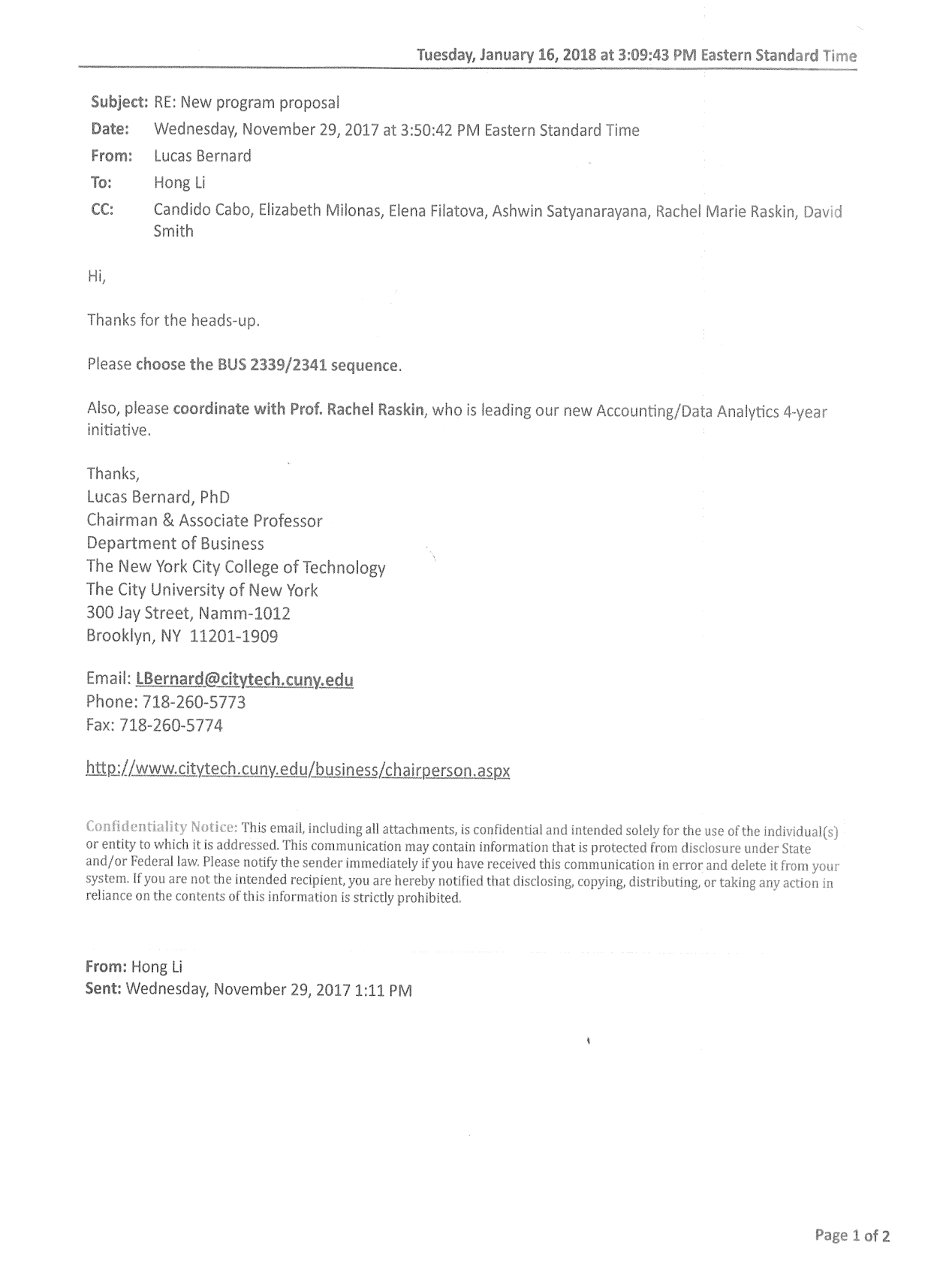
**1. Mathematics department**



****

**2. Biology department**

****

**3. Business department **