Bilkent University
Faculty of Engineering

Data Science and Engineering Certificate

1 INTRODUCTION

Data science is an inter-disciplinary field that is using scientific methods and algorithms, and systems to analyze and extract knowledge and insight from complex data [12] for use in a broad range of applications [15]. Data is usually big and complex. It can be structured or unstructured. The goal is to create data centric products, applications, solutions that address scientific, social or business questions [14]. Complex and massive data from various sources need to be analyzed. Data science spans activities that involve applying principles for data collection, storage, integration, analysis, inference, communication, and ethics.

The improvement of algorithms, tools, and platforms for dealing with data is the focus of data science research. The development of next-generation data scientists as well as data-literate workforce that can utilize data-driven techniques to increase innovation and success is the focus of data science education and training. Adequate platforms and infrastructure are needed for data science research and education [19].

In Bilkent University Engineering Faculty, to equip our graduates with data science knowledge and skills and to make them ready to take data science related jobs, we consider a certificate program that will include a set of related courses. Students taking the required number of courses from the related courses will be eligible to get a certificate, which will explicitly indicate the know-how of our students in the field.

Our goal with the selection of courses to be included in the certificate program is to instill some basic understanding of data science to our students and to develop their data science skills to some degree. We can start with what we have already as courses, and later we can add other courses related to data science.

The certificate shall be as inclusive as possible, and should provide pathways to acquire data science knowledge and skills for all the students of Engineering Faculty.

We could have a major, minor, track, focus, or certificate, or just some courses, as the means to offer data science focus. Hence, a range of educational programs are possible to consider for our undergraduate students. We selected certificate approach that we believe is currently the most appropriate for our Faculty and University. It seems to be a viable approach, since the program is to be offered to engineering students and the

current definition of minor program in Bilkent University is more suitable for being inter-faculty or inter-departmental. Furthermore, a minor program needs to satisfy Bilkent University regulations that are somewhat restrictive.

A key goal of the certificate program is to enable students to make good judgments and decisions in problems involving large sets of data and to be able to use appropriate tools (in broader sense) effectively. We need to impart, to some degree, the related knowledge and skills to our students, so that they can be successful in data science and related jobs.

A data scientist can be doing or engaged with one or more of the following:

- Set up and operate and manage the systems underlying the big data, over which analysis can be done.
- Data storage and access management, databases.
- Collecting and preparing data.
- Doing and coordinating the data analysis and data analytics. Statistical modeling, computational modeling, and machine learning.
- Solving data-driven problems with suitable algorithmic approach and software.
- Visualization.
- Supporting data driven decision making, uncovering the stories buried in data.

Therefore, there are various data science related roles and works to do for our graduates.

The following are key concept areas for data science: mathematical foundations, computational foundations, statistical foundations, data management and curation, data description and visualization, data modeling and assessment, workflow and reproducibility, communication and teamwork, domain specific considerations, and ethical problem solving [15].

a) Key mathematical skills involve: set theory and basic logic, multivariate thinking via functions and graphical displays, basic probability theory and randomness, matrices and basic linear algebra, networks and graph theory, optimization. Additionally, knowledge on partial derivatives, advanced linear algebra (eigenvalues and decompositions), Big-O and analysis of algorithms, and numerical methods (interpolation and approximation) are desirable [15].

b) Key computational foundations include: basic abstractions, algorithmic thinking, programming concepts, data structures, simulations [15].

c) Important statistical foundations include: variability, uncertainty, sampling error, inference; multivariate thinking, non-sampling error, design, experiments, tests, biases, causal inference; exploratory data analysis; statistical modeling and model assessment; simulations and experiments [15].

d) Key data management and curation skills include: data preparation, data transformation, data management, dealing with missing and conflicting data, modern databases [15].

e) Data description and visualization concepts/skills include: data consistency checking, exploratory data analysis, grammar of graphics, attractive and sound and dynamic visualizations, dashboards [15].

f) Data modeling and assessment related skills/concepts include: machine learning, multivariate modeling, supervised learning, dimensionality reduction, unsupervised learning, deep learning, model assessment and sensitivity analysis, model interpretation [15].

g) Workflows and reproducibility skills include: ability to understand client needs, combining simpler tools to solve more complex problems, communication and teamwork, technical writing, presentation [15].

In Bilkent, we can touch to a, b, c, d, and f.

2 ALL RELATED COURSES IN BILKENT UNIVERSITY NOW

Below are the courses that we have in Bilkent University, most of them in Faculty of Engineering, that are related either directly or indirectly with data science. Some courses in other faculties may also be related (some courses in Economics and Management departments). They are not investigated yet.

Foundational College Mathematics Courses

MATH101/102 - Calculus I and II
MATH230, MATH250/MATH260, MATH255 - Probability and Statistics for Engineers
MATH225, MATH220/MATH240, MATH241/MATH242 - Linear Algebra and Differential Equations

Programming Foundation Courses

CS101, CS115

Math – Computing and Data Theory /Foundations Related Courses

CS471 Numerical Methods
CS473 Algorithms I
CS478 Computational Geometry
IE411 Introduction to Nonlinear Optimization
ME361 Numerical Methods for Engineers

MATH132 Discrete Mathematics
MATH260 Introduction to Statistics
MATH465 Mathematical Foundations of Data Science

Computing – Programming and Data Systems Related Courses

CS102 Algorithms and Programming II
CS201 Fundamental Structures of Computer Science I
CS202 Fundamental Structures of Computer Science II
CS353 Database Systems
CS281 Computer and Data Organization
CS425 Algorithms for Web scale data
CS426 Parallel Computing
CS442 Distributed Systems
CS443 Cloud Computing and Mobile Applications
IE324 Simulation

Techniques for/and Applications of Data Science Related Courses

GE461 Introduction to Data Science
CS461 Artificial Intelligence
CS464 Introduction to Machine Learning
CS481 Bioinformatics Algorithms
CS483 Natural Language Processing
CS484 Introduction to Computer Vision
IE425 Forecasting Methods and Applications
IE451 Applied Data Analysis
IE452 Algebraic and Geometric Methods in Data Analysis
IE469 Industrial Applications of Operations Research
EEE443 Neural Networks
EEE482 Computational Neuroscience
EEE485 Statistical Learning And Data Analytics
EEE486 Statistical Foundations of Natural Language Processing

Graduate Level Courses (not double coded) related to Data Science Techniques and Applications

CS550 Machine Learning
CS551 Pattern Recognition
CS553 Intelligent Data Analysis
CS554 Computer Vision
CS557 Computational Systems Biology
CS558 Data Mining
CS559 Deep Learning
CS579 Biometrics
IE553 Applied Statistical Modeling and Data Analysis

3 APPLICATION FOR CERTIFICATE

Since this will be a certificate program, not a minor, there will be no application. Any student who thinks he/she is satisfying the requirements will apply and will get their certificate after checks are done. Therefore, it may be better not to set entrance requirements.

4 CERTIFICATE COURSES

We consider three sets of courses. A student should have taken some number of courses from each set to be eligible for the certificate. We may or may not name the sets of courses. We can just say Set 1, etc., and have a rough idea about what each set should have. We can add a course to more than one set if that seems to be better.

Set 1 includes courses that provide *general* foundations for data science: mathematical foundations related to data science, programming and data systems foundations for data science, or computational thinking foundations for data science.

Set 1
General Foundations for Data Science
1-2 courses from this set
CS281 Computer and Data Organization CS353 Database Systems CS426 Parallel Computing CS471 Numerical Methods CS473 Algorithms I IE411 Introduction to Nonlinear Optimization ME361 Numerical Methods for Engineers MATH260 Introduction to Statistics

Set 2 includes courses that are about foundations of data analysis and analytics, and about models, tools, and techniques for data science (mathematical, statistical, and computational/algorithmic).

Set 2
(Mathematical, Statistical and Computational/Algorithmic) Foundations, Models, Tools and Techniques of Data Analysis, Analytics, and Science.
1-3 courses from this set
GE461 Introduction to Data Science CS461 Artificial Intelligence CS464 Introduction to Machine Learning CS478 Computational Geometry

EEE443 Neural Networks
EEE485 Statistical Learning and Data Analytics
IE451 Applied Data Analysis
IE452 Algebraic and Geometric Methods in Data Analysis
IE553 Applied Statistical Modeling and Data Analysis
MATH465 Mathematical Foundations of Data Science

Set 3 incudes courses that are about applications of data science, different domains of data science, and more advances topics related with data science.

Set 3
Applications and Advanced Topics in Data Science
1-3 courses from this set
CS425 Algorithms for Web-scale Data
CS443 Cloud Computing and Mobile Applications
CS481 Bioinformatics Algorithms
CS483 Natural Language Processing
CS484 Introduction to Computer Vision
EEE482 Computational Neuroscience
EEE486 Statistical Foundations of Natural Language Processing
IE469 Industrial Applications of Operations Research
CS550 Machine Learning
CS551 Pattern Recognition
CS553 Intelligent Data Analysis
CS554 Computer Vision
CS558 Data Mining
CS559 Deep Learning
CS579 Biometrics

5 RULES FOR COMPLETION OF THE PROGRAM

A student must complete a total of 6 courses listed above with the indicated number of courses from each pool, with a grade of B or above from all 6 courses.

The courses that are in the three sets for the certificate program can be changed in the future by the Faculty of Engineering.

The certificate program can be terminated by Engineering Faculty when it seems to be necessary to do so.

6 EXAMPLES FROM OTHER UNIVERSITIES

Reference [2] states that knowledge in statistics, computer science and linear algebra and calculus is important. SQL, Python, R and Hadoop skills are needed. Sample course overview includes courses in foundations on data analysis, data management, data analytics and visualization, big data analytics and management, machine learning. Same reference says that data science skill set includes experience in many of the following programs: SAS, Matlab, R, Python, Java, C, C++, Hadoop, SQL/NoSQL databases. Technical skills include Math (linear algebra, calculus, probability), statistics, machine learning tools and techniques, data mining, data cleaning and wrangling, data visualization and reporting, unstructured data techniques.

Universities differ in the courses offered, number of credits, and total number of courses required. But there are some common aspects. [3] is listing a summary of various certificate programs in USA. Columbia university [11], for example, offers an online certificate program that is 12 credits, and courses cover algorithms for data science, probability and statistics, machine learning, and exploratory data analysis and visualization.

When we look to various programs and the courses included, we see that universities are quite often including the courses that they already offer in various departments.

The number of sets, from which students select courses, is also changing from one program to another. MIT data science minor [4] has 4 sets of courses, for example. Set 1 is Foundations, and includes courses on linear algebra, introduction to computer science, programming in python, differential equations (1 course will be selected); Set 2 is Statistics I, and includes courses on probability, statistics, random variables (1 course will be selected); Set 3 is Statistics II, and includes courses on econometrics and statistics (1 course will be selected); and Set 4 is Computer and Data Analysis, and incudes courses on engineering computation and data science, numerical computation for mechanical engineers, inference, machine learning, computational systems biology, vision, econometrics, optimization, matrix methods in data analysis, business and finance (2 courses will be selected). There is also a capstone project.

7 NAME OF THE PROGRAM

The name of the program can be:

- Data Science and Engineering (currently most preferred) (DSE)
- Data Science (DS)
- Data Analytics (DA)
- Data Science and Analytics (DSA)

8 EVOLUTION

Depending on the needs of the industry, government, academia and society for data scientist qualifications, and the advancements in the data science field, the program needs to evolve to adapt to the changing needs and focus. We also need to learn from our experience with the program. Hence, the program should be designed in such a way to include the flexibility for evolution. We keep the certificate program open for changes and improvements.

9 NOTES

- The following courses are at design level. If they are offered in the future, they can be included in the course sets as well.
    - EEE3XX Matrix Methods for Machine Learning.
    - IE362 Optimization Methods for Data Science.
- We need to evaluate the program and the outcomes periodically.
- We may include an ethics related course or we may embed ethical issues to some courses.

10 CERTIFICATE DOCUMENT CONTENT

We need to decide what the certificate document will contain. The average gpa from courses, signatures of Dean, Chair, list of courses, are some options to consider.

11 REFERENCES

[1]. Data Science Report. Bilkent University, 2019.
[2]. Data Science Certification Programs in Universities.
https://www.discoverdatascience.org/programs/data-science-certification/
[3] Data Science Degree Programs.
https://www.datasciencedegreeprograms.net/rankings/certificate/. Shows the courses (topics) included.
[4]. MIT Minor in Statistics and Data Science. https://stat.mit.edu/academics/minor-in-statistics/.
http://catalog.mit.edu/interdisciplinary/undergraduate-programs/minors/statistics-data-science/statistics-data-science.pdf
[5]. Data Science, CMU. https://www.cmu.edu/graduate/data-science/data-science-programs-graphic-fall-2014.pdf.
[6]. NYU Center for Data Science.
https://cas.nyu.edu/academic-programs/bulletin/departments-and-programs/data-science/program-of-study-cas-bulletin.html.

[7]. Berkeley Institute for Data Science. https://bids.berkeley.edu/about.
[8]. A Very Short History of Data Science. https://www.forbes.com/sites/gilpress/2013/05/28/a-very-short-history-of-data-science/#11a99d1855cf.
[9]. Data Science, Cornell. https://www.engineering.cornell.edu/data-science.
[10] University of Washington, Data Science. https://escience.washington.edu/data-science-courses-at-the-university-of-washington/.
[11]. Columbia, Data Science Certification. https://datascience.columbia.edu/certification.
[12]. Data Science Definition – Wikipedia. https://en.wikipedia.org/wiki/Data_science.
[13] Data Science Career Path after College. https://www.datasciencesociety.net/data-science-career-path-after-college/.
[14] Data Science, UMD. https://www.cs.umd.edu/data-science.
[15]. Data Science for Undergraduates: Opportunities and Options. National Academies of Science, Engineering and Medicine. 2018.
[16]. College and University Data Science Degrees. http://datascience.community/colleges.
[17]. Curriculum Guidelines for Undergraduate Programs in Data Science. Park City Math Institute (PCMI) Undergraduate Faculty Group. 2016.
[18]. Data Science at NSF. https://www.nsf.gov/attachments/130849/public/Stodden-StatsNSF.pdf.
[19]. Realizing the Potential of Data Science. NSF. https://www.nsf.gov/cise/ac-data-science-report/CISEACDataScienceReport1.19.17.pdf.
[20]. Data 8: The Foundation of Data Science. http://data8.org/.
[21]. DS100. Principles and Techniques of Data Science. http://www.ds100.org/.