

Proposal of Certificate in Data Science

Summary: The proposed Certificate in Data Science (hereinafter CDS) will provide a basic set of skills for data scientists. It will provide students a broad overview of the field with practical applications of data curation and management, data preprocessing and wrangling, data processing and integration, programming, data mining and statistical learning, technical reporting and communication.

The primary goal of the CDS is to improve the quantitative and analytical learning outcomes for students of all degree programs at Florida Atlantic University. The proposed curriculum of the CDS exposes students to the nature of data science and builds a basic set of skills in data analytics. The proposed CDS enables students to complement any major with a focus in data analytics obtained from the foundational courses offered through the CDS program. The proposed CDS incorporates RI components in three courses to enhance the initiative of OURI at FAU.

Introduction and Rationale

Today's demand for expertise in data mining, predictive analytics and machine learning far exceeds the current supply. Data science is a rapidly accelerating interdisciplinary field with an increasing demand in industry, research, and government. Data science addresses knowledge extraction via data curation and management, data mining, predictive analytics, and machine learning in structured or unstructured large-scale and complex data. Data science combines mathematics, statistics, operations research, information science, and computer science.

Data Science graduates are highly employable in business, industry, and professional fields such as medicine, law, and government. Graduates wishing to pursue higher degrees in machine learning or more domain specific fields such as bioinformatics will be well prepared for graduate study by completing the proposed CDS. A widely cited McKinsey Global Institute study (McKinsey Global Inst. 2011) forecasts a need for hundreds of thousands of data science jobs in the next decade. In a recent "Peer-Research Big Data Analytics Survey" conducted by Intel, seventy four percent of the respondents have agreed that data science is adding value to their organization and provides vital information for making timely and effective business decisions of great importance. In a recent MIT Sloan Management Review survey, four in ten (43%) companies report their lack of staff with appropriate analytical skills as a key challenge.

Our society generates 2.5 quintillion bytes of data per day, and roughly ninety percent of the entire data were generated in the past two years. The growth rate keeps increasing every year. Finding meaningful information from those huge data in a timely manner is a challenging task which requires specialized data science and analytics skills. Meanwhile, the rise of big data has been accompanied by rapid growth in the demand for data analysts and scientists. To embrace such a great opportunity and job market need, we propose to create CDS at the undergraduate level for Florida Atlantic University. The CDS can be completed in conjunction with any major from Florida Atlantic University.

While data science is best taught at the master's level, a few bachelor's level programs do exist. A selection of undergraduate programs in data science and related fields can be found at <https://discoverdatascience.com/data-science-bachelors-programs/>. The CDS program is designed to allow students in a variety of majors to add on a certificate of specialization in Data Science. The authors of this proposal have seriously referenced similar programs from several countries, and the proposal follows the curriculum guidelines for undergraduate programs in Data Science (<http://www.amstat.org/asa/files/pdfs/EDU-DataScienceGuidelines.pdf>). Data science is an interdisciplinary discipline about methods and systems to extract knowledge or insights from large quantities of data delivered in various forms. Data Science employs techniques and theories drawn from many fields within the broad areas of mathematics, statistics, computer science and information science and applies them to a wide range of data-rich domains such as biomedical sciences, physical sciences, geoscience, social science, engineering, business, and education. A CDS program at the undergraduate level provides a synergistic approach to problem solving, which leverages the content in all three disciplines. It offers the opportunity to integrate and use both computational and statistical thinking to solve problems rather than emphasizing one over the other. A graduate with an undergraduate data science certificate should be prepared to interact with data at all stages of an investigation and will be equipped to work within a team environment.

The goal of this proposal is to create a CDS. The proposed CDS will provide a basic set of skills for data scientists. It will prepare students a broad overview of the field with practical applications of data curation and management, data preprocessing and wrangling, data processing and integration, programming, data mining and statistical learning, technical reporting and communication. The proposed CDS requires 15 credits from Mathematics, Statistics, Computer Science, and Operations Management.

The College of Science is positioned to offer a distinctive certificate program in Data Science. A Data Science Certificate will be issued to a student if s/he successfully completes **five 3-credit courses** from the Curriculum.

Curriculum

Students enrolled to the CDS program are required to complete 15 credits including three core courses and two elective courses. The two elective courses can be chosen from the list of elective courses.

Three core courses:

- MAD 2502 Introduction to Computational Mathematics (or COP 2220 Introduction to Programming in C)
- CAP 3786 RI: Introduction to Data Science

- STA 4442 Probability and Statistics 1 (or STA 4032 Probability and Statistics for Engineers)

Elective courses:

- STA 4102 RI: Computational Statistics
- STA 4222 Statistical Designs
- STA 4234 Applied Statistics 1/STA 4202L Applied Statistics 1 Lab (or ECO 4421 Introduction to Econometric Methods)
- STA 4241 RI: Statistical Learning (or CAP 4770 Introduction to Data Mining and Machine Intelligence)
- STA 4853 Applied Time Series and Forecasting

Admission and Completion

The certificate program is for students who satisfy the prerequisites required for each course in the program at Florida Atlantic University. All five courses in the program must be completed with a grade C or better.