Data Analysis Foundations for Mathematics

Data Analysis Foundations for Mathematics	
Offered by	Department of Mathematics and Computer Science
Language	English
Primarily interesting for	All students, but most relevant for students with background in Applied Mathematics and other BSc students interested in a Master program in Data Science (e.g., Data Science in Engineering)
Prerequisites	Statistics, linear algebra, propositional logic, predicate logic, set theory, imperative or object-oriented programming
Contact person	Dr. Dirk Fahland (d.fahland@tue.nl)

Content and composition

This package provides courses to students in the Applied Mathematics major, who wish to advance their knowledge in the data analysis foundations of Data Science: data visualization and data mining and machine learning. Data Mining and Machine Learning introduces the theoretical foundations of data mining and machine learning. The course Visualization teaches techniques to enable the explorative analytics of unknown data through interactive visualization.

This elective package is one of two packages intended for students in the bachelor program Applied Mathematics and other programs (except the bachelor program Computer Science), who are interested in a technical master program on Data Science (e.g., Data Science in Engineering).

- Elective Package: Data and Algorithmic Foundations for Mathematics
- Elective Package: Data Analysis Foundations for Mathematics

Following both packages gives Mathematics students the necessary prior knowledge to enroll in a technical master program on Data Science (e.g., Data Science in Engineering).

Course code	Course name	Level classification
2IRR50	Statistics and Machine learning	3
JBI100	Visualization	2

The courses can be taken in any order or in parallel.

Course description

2IRR50, Statistics and Machine learning

The main focus of this course is on the theoretical and mathematical foundations of Data Mining and Machine Learning. A secondary focus is on low-level practical aspects (e.g. vanilla implementations of various models and algorithms). After the course the students will be able to:
* distinguish the two main Machine Learning paradigms of supervised and unsupervised learning.
* assess the suitability of common machine learning methods for a given learning problem

• compare weaknesses and strengths of common machine learning methods based on theoretical properties and practical considerations
• implement, and evaluate some of the most widely used methods
• reflect upon the choice of model, its theoretical properties and resulting influences on the result.

JBI100 Visualization

This is a course about visualization concepts in which the students should learn about the difficulties that visualization researchers face today. To reach this goal we focus on one specific scenario about designing and implementing interactive dynamic graph visualizations with a focus on exploring time-varying trends in network data. In addition, the development of professional skills is an important aspect of this project. Each group of at most 5 participants specifies, designs, and actually builds a visualization tool. Each group must document the implemented code and also report on the activities that are typically motivated by weekly assignments. Each project is concluded with a presentation at the end of the quartile and a written final report.