Data Modeling Foundations for Data Science

Data Modeling Foundations for Data Science	
Offered by	Department of Mathematics and Computer Science
Language	English
Primarily interesting for	Data Science (BSc) and other BSc students interested in the MSc Data Science in Artificial Intelligence
Prerequisites	Logic and set theory, Programming, Calculus B (or equivalent), Probability and Statistics
Contact person	Dr. Dirk Fahland (d.fahland@tue.nl)

Content and composition

Data analysis is at the core of most professional, engineering, and scientific disciplines and is fundamental to Data Science. Yet, analyzing data requires storing and accessing it in ways that are most suitable for the analysis at hand. The courses in this package provide the necessary foundational knowledge and skills to choose adequate data models for a task at hand, pick the right system to store data, and access, analyze, and get insights into very large data sets through data querying languages, data mining and machine learning, and linear algebra techniques.

This elective package is for students in the bachelor program Data Science and other programs (except the bachelor program Computer Science), who are interested in the master program Data Science and Artificial Intelligence (DS&AI). The courses in this package offer required and highly recommended prior knowledge for enrollment in this master program. Bachelor Data Science students acquire the necessary knowledge to be admissible to Master program DS&AI if they have successfully completed the following packages

- Data Modeling Foundations for Data Science
- Computer Science for Data Science.

Note: admission to DS&AI requires more prior knowledge than listed in these packages, see https://educationguide.tue.nl/programs/graduate-school/masters-programs/master-data-science-and-ai/?L=2 .

Course code	Course name	Level classification
2DBI00*	Linear algebra and applications	1
JBI050**	Data management for data analytics	2
2IIG0***	Data mining and machine learning (optional)	3

* 2DBI00 may be replaced by JBM075 Linear Algebra for Data Science or 2WF20 Linear Algebra 1
** JBI050 may be replaced by 2ID50 Data modeling and databases
*** 2IIG0 has Linear Algebra as prerequisite knowledge, the contents of 2IIG0 overlaps with JBI030 since AY 2019-2020, students who took JBI030 in AY 2019-2020 are admissible to DS&AI without following 2IIG0, however following 2IIG0 will aid students in getting more proficient in the theoretical foundations of data mining and machine learning

Course description

2DBI00, Linear algebra and applications

This course offers a wide range of interesting linear algebra techniques and very nice applications, including Big Data applications. These techniques are almost indispensable for the current Big Data era. You will learn how to solve linear systems, arguably the most common and important scientific problem. Least squares methods can be used to determine an approximate line (or polynomial, or spline) through a number of points (used in Computer Graphics). Rotations and reflections are useful for Computer Graphics and Robotics. Angles between

vectors can be used to compare tastes of movies, music, or books, and to predict such tastes for the future. You will learn the idea behind Google PageRank and its connection with eigenvalues, how it is computed, and you can also apply similar techniques to finally determine which of Ajax, Feyenoord, or PSV is the best team. We can use text mining methods to determine the most important tweets or keywords from a set of tweets; useful for effective communication. The course is designed to be interesting for many students, including students from Computer Science, Data Science, and Robotics. The subjects are fascinating on their own, but also form an ideal preparation for subsequent courses such as Data Mining and Machine Learning and Computer Graphics. The course offers a nice combination of theoretical (math) and practical (implementing) work, partly done in groups. You can use a language you already know (for instance Python or Matlab) or learn a new language.

JBI050, Data management for data analytics

As we enter an era of big data and data science, core knowledge and skills in data modeling and data management are now recognized as essential in many disciplines. The primary goal of JBI050 is to master the core best-practices of data management systems, applied towards using contemporary tools to support effective data analytics. In particular, this course focuses on preparing students to meet contemporary data modeling and data management challenges which arise in applications in their own fields of study. The focus of the course is on practical problem-solving in an application domain. Students will gain practical experience developing the ability to design effective databases based on a solid understanding of the underlying principles. By design, hands-on practical assignment(s) using contemporary frameworks and technologies are a central component of the course. The following topics will be covered: Data modeling: conceptual modeling in the ER model and UML; logical data modeling in the relational database model; optimization of logical models, basics of normalization; Querying databases: SQL basic queries, aggregation; Datalog basic queries, recursion.

2IIG0, Data mining and machine learning

The main focus of this course is on the theoretical 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: define Data Mining; define Machine Learning; define all Machine Learning paradigms: supervised, unsupervised, and reinforcement learning; work with data; preprocess data; identify which Machine Learning methods are the most suitable for a specific learning paradigm; derive, implement, and evaluate some of the most widely used methods (listed in the course content) for a specific learning paradigm; derive, implement, and evaluate some of the most used deep learning models (listed in the course content) and their learning algorithms; solve a Data Mining problem using a Machine Learning model.