Data Science for CSE elective package

Data Science for CSE	
Offered by	Department of Mathematics and Computer Science
Language	English
Primarily interesting for	All students, but primarily for students with BSc in Computer Science and Engineering (CSE) who want to pursue the Data Science & Artificial Intelligence master
Prerequisites	Required courses: Recommended courses: Students are assumed to have basic skills in logic, set theory, calculus, discrete mathematics, databases, algorithms and programming.
Contact person	Dr. Dirk Fahland (d.fahland@tue.nl)

Content and composition

Analysis of information, data, and knowledge is increasingly important, with broad application across science, engineering, society, and industry. To tackle these challenges, knowledge and skills in the management, mining, and analysis of (big) data collections is necessary. This elective package provides deeper study of the foundations and applications of analysis of data and information systems.

Students should select a coherent package of courses from the following list. The courses can be followed in any order, but need to take the specific prerequisites for each course into account when scheduling this package. For example, 2IX30 has as prerequisite 2IIG0 given in Q2/C and 2DI90 assumes the successful completion of a calculus course, particularly integral calculus.

Course code	Course name	Level classification
2DI90	Probability and statistics	2.
JBI100	Visualization	2.
2IX30	Responsible data science	3.

Course description

Probability and statistics (2DI90)
This course gives an introduction to probability and statistics. In many situations it is impossible to predict how a complex system will behave. Models capturing this uncertainty are essential to study such systems. Probability theory gives us the tools to do this in a powerful manner. Statistics, on the other hand, concerns what can be learned from data. In this course you will learn how to build probabilistic models, perform statistical analyses of real data, and the theory behind all this.

Visualization (JBI100)
In the visualization course you will learn about the challenges of visually representing data that comes in a variety of forms. Starting from simple primitive data types like categorical, ordinal, or quantitative data, we will have a look into more complex dataset scenarios including relational data like graphs/networks or hierarchies, multivariate data, text data, or trajectory data that contains an inherent spatio-temporal aspect.

Data Science for CSE elective package

In this course you will learn about the data processing, data transformation, data visualization, and finally, the interaction with the visual output. To make a visualization interpretable, readable, and intuitive, we will also have a look at perceptual issues like pre-attentive processing, the visual memory, or Gestalt principles. Moreover, a number of laws or no-goes will be discussed to make the diagrams or visualization techniques more perceptually effective.

Responsible data science (2IX30)

The course is focused on studying the problems of fairness, accountability, confidentiality, and transparency (FACT) in data science, and data mining and machine learning in particular. One important challenge to face is that machine learnt models typically are not 100% accurate, i.e. in some ways these models are wrong. Thus, it is important to study how we can make a good use of models that are not perfect, how we can understand the strengths and weaknesses of these models, how we can help a decision maker to trust (or not trust) the model or its particular prediction, and how we can get insights into impact of input features and some inner logic of a predictive model. We need techniques not just to explain the decision of a model, but also to uncover and characterize undesired or even unlawful biases in its performance. Hence, the other important challenge to study is how to formally define such biases, how to uncover and quantify them and how to design machine learning solutions that would enable the so-called fair algorithmic decision making by design. On the other side of the spectrum, there are challenges of privacy and confidentiality. We will study the main principles and techniques that have been researched and employed in data mining for privacy preserving and secure computation to inducemodels from data and to apply them in real-life scenarios.