Data Science for IE&IS students

Offered by: Department of Industrial Engineering & Innovation Sciences
Language: English
Primarily interesting for: All majors of the Industrial Engineering and Innovation Sciences fac.
Contact person: dr. Ksenia Podoynitsyna (k.s.podoynitsyna@tue.nl)

Content and composition
This coherent package provide students with the basic knowledge on data science including programming, databases and data mining techniques. This coherent package also clears up the requirement on 15ECTS data science courses necessary for the direct admission of students majoring Industrial Engineering or Innovation Sciences to the master “Data Science and Entrepreneurship” in JADS, Den Bosch. JADS is the Joint Graduate School of Tilburg University and Eindhoven University of Technology (see www.jads.nl for more information).

<table>
<thead>
<tr>
<th>Course code</th>
<th>Course name*</th>
<th>Level classification</th>
</tr>
</thead>
<tbody>
<tr>
<td>JBI010**</td>
<td>Programming</td>
<td>1. Basic</td>
</tr>
<tr>
<td>JBI030</td>
<td>Data Mining</td>
<td>2. Intermediate</td>
</tr>
<tr>
<td>2ID50*** / JBI050</td>
<td>Data modelling &amp; Databases/ Data management for data analytics</td>
<td>2. Intermediate</td>
</tr>
</tbody>
</table>

* - JBI020 “Foundations of computing” is a highly recommended supplement to this coherent package.
** - JBI010 is not allowed for the BPT students. This course has a strong overlap with the obligatory course 0HV120 1e jaars vak BPT studenten. 0HV120 serves as the replacement for JBI010 in this case.
*** - Prior knowledge requirement for the 2ID50 course is 2IT60 “Logic and set theory”, which contents are partially covered in the obligatory courses in the major TBK and P&T. 2ID50 has an entrance test which will validate whether this prior knowledge is sufficient, allowing students to qualify for this course.

Please note that students have to be directly admissible to the master “Data Science and Entrepreneurship” as a whole in order to follow one or more courses from this master, including an exchange semester. This means that students following master programs of the IE&IS faculty are only directly admissible in case they have followed this coherent package as part of their bachelor.

BP&T students generations (2016/2017 onwards) have likely fulfilled most of the requirements as part of their major (10 ECTS math, 10 ECTS statistics, 15 ECTS databases, data mining and programming) and would only need to finish 1 extra course on data mining to be able to join.

Since the current 15 ECTS coherent package is the bare minimum, we highly recommend students interested in joining the master “Data Science and Entrepreneurship” to also follow the Additional Recommended Courses listed below next to the current coherent package.
Relevant courses descriptions

2ID50 Datamodelling & Databases
Our lives are awash in data (e.g., social, business, and web) which only continues to grow in both quantity and variety. Database management systems are the key technologies which facilitate our practical use of these massive data sets. In this course, we study fundamental concepts, such as data model design and formulation of queries against databases, which underpin the effective practical use of industrial strength data management systems.

JBI050 Data management for data analytics
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.

JBI010 Programming
This course introduces imperative and object-oriented programming using Java and Python. Topics: basic imperative programming (assignment, choice, repetition, input/output, functions), typing, recursion, objects (both data objects, or records, and domain objects), a few collection classes, inheritance, interfaces, specification of methods, coding style practice, API use, basic handling of large data sets.

JBI030 Data Mining
The course fits with the educational philosophy of the program by emphasizing the interdisciplinary perspective of data science and introducing students to research in the field of data science. During this course, the students will learn the foundations of data mining and gain hands-on experience in applying data mining in practice.

Additional recommended courses:

JBI020 Foundations of computing
- Upon completion of this course students:
  understand basic proving techniques and can apply the right technique(s) to prove formal statements
- understand computability and models of computations
- understand what algorithms are and what they are used for, as well as some principles of data representation
- are able to apply techniques from computer science to understand and solve problems
**2IOI0 DBL Process mining**

The goal of this DBL is for students to get a first glimpse on the difficulties faced when introducing the process perspective in a data analytics context. At the end of this course students:

- Understand the CRISP-DM framework for data mining and are able to apply it in a project setting,
- Are able to hypothesise on data analytics questions related to process aspects,
- Understand the complexity of process aspects when doing predictions,
- Are able to build prediction models for process questions,
- Can communicate their findings in a concise and visually appealing manner,
- Can explain the validity of the results by referring back to the assumptions made on the data.

**JBI040 Algorithmic Aspects of Data Analysis**

After this course students

- understand how certain data-analysis tasks, and in particular clustering, can be formulated as well-defined optimization or decision problems;
- can reason about algorithmic speedups for data-analysis tasks, and understand that choosing the right algorithmic techniques can make a significant performance difference;
- can describe and compare different algorithms to detect clusters in various types of data;
- can implement various algorithms for clustering in an effective manner.