Introduction to the course

Introduction to the course

INTRODUCTION

Welcome to the Data Analytics course! Before you start studying the course we would like to inform you about our intentions concerning the course, the structure and content of the course and the mode of working for this course. Therefore, this introduction does not contain any study material. But is essential for a good understanding of the setup of the course.

1   **General course information**

In the Data Analytics course you will learn about the main types of data analytics techniques and how to apply them in a business context. You will learn to recognize data analytics problems, to propose analytics techniques to address the problem and to interpret the outcomes of data different analytics techniques.

The Data Analytics course is developed within the faculty of Science, Management & Technology (MST) of Open University, The Netherlands. It concerns a scientific course on master level in the Data Science Management variant of the Business Process Management & IT (BPMIT) master program. The study load of the course is approximately 210 hours (7,5 EC).

The course material consists of digital and analogue parts:

- *Workbook (digital document)*

The workbook (the document you are now reading) guides you through the course and contains instructions regarding the use of the other study materials and assignments for this course. Furthermore, the workbook contains theory that is not addressed by the textbook.

- *Textbook (analogue document)*

The textbook used for this course is: Data Science for Business by Provost & Fawcet and published by O'REILLY (2013, first edition). It contains the theory that will be used in the Data Analytics course.

- *Assignment bundle (digital document)*

The assignment bundle contains the exercises that belong to the different learning units in the course. Additionally, it describes the formal tasks that will be graded and need to be completed in order to pass the course.

- *RapidMiner (software)*

In the course we will use software to apply data analytics technique to data sets, which will be provided during the course on the YouLearn website. The software is free to use for educational purposes and can be downloaded from the RapidMiner website.

The YouLearn website forms an integral part of the course. Here you will find all the digital materials for the course as well as current information about meetings, support and exams. Specifically, the students will have access to study resources such as:

- Presentations in the form of Powerpoint or PDF slides;
- Datasets for exercises;
- Form 1&2 necessary for the examination;
- Videos of the presentations in Collaborate.

Furthermore, it provides a discussion forum to ask questions to the lecturers and fellow students.

All course material and lectures will be in English. A general understanding of English (orally and written) is therefore required for this course.

2   **Applying data analytics in a business context**

Data analytics has received much attention lately due to the growing amount of data that is at our disposal today. All this data is considered a valuable asset that needs to be explored and analyzed to turn it into information and actionable knowledge. Turning data into actionable knowledge requires techniques that are generally referred to as data analytics or data mining. Due to the fact that there are so many different analytical techniques, a good understanding of their strengths and limitations is essential for their application in a business context and is the main focus of this course.

In a broader context, the course also addresses the data safari process, which is the process of exploring and analyzing the data. This process is essential to turn data into actionable knowledge because it is more than just applying the right analytical technique according to the problem at hand. For this purpose, the students will learn to distinguish between supervised and unsupervised learning and to apply algorithms for classification, clustering, regression and associative rule mining.

The knowledge gained about data analytics in this course is either intended for people from the business that want to understand the potential of data analytics for their organization. Or for people that will act as an intermediary between the business on one end and data scientists at the other hand. People from the business often lack a general understanding of data analytics techniques to understand how it can help them to solve business problems. Students that followed this course should be able to help them to formulate their business problems in such a way that they can be analyzed by data scientists.

3   **Course learning objectives**

The learning objectives presented here describe the knowledge, insight and skills you should acquire at the end of this course. After studying this course, we expect that you are able to:

- To understand the key elements of machine learning, such as dataset, classification algorithm, label, mean classification accuracy and standard deviation.

- Understand the main concepts of data analytics and its role in organizational decision making and innovation.

- Understand the difference between: predictive tools, prescriptive tools and descriptive tools.

- Derive a researchable data mining question from business needs.

- Apply the following analytical techniques multiple regression analysis, clustering, classification trees, naive Bayes, Bayesian networks, support vector machines, principal component analysis and linear discriminant analysis.

- Using a tool like RapidMiner.

4   **Structure and contents of the course**

The contents of the course consists of 7 themes called blocks:

Data and charting (block 1)
The first block deals with a gentle introduction to the data analytics concepts and with charting techniques that are often used when performing a data mining task.

The data mining process (block 2)
The second block introduces the student to the concept of data set, data set description and data cleansing.

Machine learning (block 3)
The third block introduces the student to supervised machine learning algorithms and metrics for classification tasks.

Regression (block 4)
The fourth block introduces the student to supervised machine learning algorithms and metrics for regression tasks.

Clustering (block 5)
The fifth block introduces the student to unsupervised machine learning algorithms and metrics.

Rule mining (block 6)

The sixth block introduces the students to the issue of association rule mining.

Text mining (block 7)
The seventh block introduces the student to text mining concepts and algorithms.

The blocks are spread out over the 11 week duration of the course resulting in an average study load of max 19 hours per week. Each block consists of studying the theory from the textbook/workbook and making exercises from the assignment bundle.

Examination of the course is based on two individual assignments, made during the course, and a webinar at the end of the course.

5   **Study instructions**

5.1   MANNER OF STUDY

Learning unit

The course covers a number of themes (called blocks) and each block consists of one or more learning units. Learning units are the elementary building blocks of the course and they contain:
-   <u>Introduction</u>: explains what the learning unit is about and what you will learn in the specific learning unit.
-   <u>Study core</u>: contains the study material including exercises and tasks.

Example of a study assignment

You will find the learning units in the workbook (this document) and is therefore the starting point for all your study activities. At the start of a learning unit you will begin with reading the introduction and after that you will study the material in the study core and will make the exercises and/or tasks. In the workbook, the term *'Reading assignment'* in the margin indicates that the adjacent text in the workbook provides instructions on what part of the textbook you should study at that point. Additionally, the workbook will make references to other sources where to find the study materials, such as assignment bundle and YouLearn website.

Online tutoring

Besides self-study, this course also consist of five *online tutoring* sessions during each cycle. One can find more information about these meetings on the course's yOULearn website. It is strongly recommended to attend these tutoring sessions in order to successfully pass the course.

Exercises

The exercises in this course are meant for self-study and getting a better comprehension of the theory. Hence, they are not graded and do not count towards the grade of the course.

Additionally, there are three tasks that together form the basis for examination in this course. The tasks test the student's knowledge acquired during the readings. The first two tasks are the execution of a machine learning task, which each student performs on an individual basis. At the end of the course, the students communicate their test results as a whole in a final *Webinar (task 3)*, which is the core of the third and last task. The teacher assesses these assignments and Webinar. Assessment of task 1 and 2 help determine the final mark, task 3 only confirms the vote.

Task one:
Machine learning
with Rapid Miner

The *first tasks* concerns the application of machine learning algorithms in a dataset. The students will be provided with a set of datasets to analyze, the students will have to specify what sort of analysis is most appropriate in the dataset and provide an explanation of their reasoning, comparing the results of multiple algorithms.

Task two:
Text Mining or
Association Rule
Mining

The *second task* will concern an exercise with association rule mining or with text mining. The student will be asked to classify or cluster a set of documents or to specify which association rules hold in a dataset, providing an explanation of the rules found with the existing association rule mining algorithms.

Task three:
Webinar
Presentation

The final task, the presentation, will involve a Webinar meeting with the examiners to confirm the grade obtained in the first two tasks. More detailed information concerning the assignments can be found in the assignment bundle.

5.2   STUDY LOAD

The course has a total workload of about 177.5 hours net. The beginning of each unit has an indication of that learning unit's workload. In several places, you will find optional parts, such as reading literature as additional background information or perform optional tasks. We would like to stress that these parts do not fall within the specified study. Each stated workload is an indication of how long the average student needs to perform the task. In each study unit, we indicate in detail which parts you should study or read.

6   **Prerequisite knowledge**

This course requires no previous knowledge beyond that needed to start any course in the Masters of Business Process Management and IT or Computer Science programs.

7   **Form of assessment**

Participation in the course Data Analytics is assessed through the assignments, which consists of two tasks, and a final presentation in the form of a Webinar, during which the teacher asks questions. The complete instructions for the assignments and for preparing for the final presentation are in the assignment bundle.

After submission of task 1, you will receive some feedback and a mark. After submission of task 2 you will receive a second mark. The final mark

is calculated as the average of task 1 and 2. This is then confirmed during task 3. The grade is assigned on the basis of task 1&2 considered as a whole, task 3 only confirms the mark. To pass the course the student must score an average of at least 55/100 and get more than 40/100 in both task 1 and task 2.