



Business Intelligence & Data Management

Syllabus – Business Intelligence & Data Management (320092):
2019-2020

Master: Information Management

Course teachers: Dr. E. Caron & Dr. E. Ioannou

Date: January 2020

Table of Contents

Master Information Management	3
Course teachers	3
Contents	3
Course aims	3
Course organization	4
Grading	4
Reading	5
Software	5
Database prior knowledge	5
Python prior knowledge	6
Course outline	6

A banner image for the course 'Business Intelligence & Data Management'. It features a collage of business-related graphics including bar charts, line graphs, and data tables in various colors (blue, yellow, red, green).

Business Intelligence & Data Management

Master Information Management

The course *Business Intelligence & Data management* (BI & DM) is part of the *Master Information Management* (MIM). The course (BI & DM) is a course in the heart of Information Management, with many applications in all business functions, like producing analytics based on data for: sales, marketing, accounting, human resource management, supply chain management, etc.

Next to the Master Information Management, the course is open to students from the masters: Research Master in Business, Master Marketing Analytics, and the Master Data Science and Society.

Notice that the course is tailor-made for the Master IM, therefore the course might not be perfectly aligned with the contents of the other masters mentioned.

Course teachers

The following people are involved in the course:

- Dr. E. (Emiel) Caron: lecturer & lab teacher (eamcaron@uvt.nl);
- Dr. E. (Ekaterini) Ioannou: lecturer & lab teacher (ekaterini.ioannou@uvt.nl);
- Dr. P. (Poonacha) Medappa: lab teacher (P.K.Medappa@uvt.nl).

Contents

The main issues in this course concern the identification and extraction of new and useful knowledge from company databases. We will start with the fundamentals behind these databases, and introduce you to database management and database querying with SQL. The company databases are the sources for the development of the data-warehouse, which is a dedicated database for managerial decision-making.

In addition, different types of knowledge can be derived from data-warehouses. OLAP data cubes for descriptive business analytics, rules characterizing potential customer classes, knowledge classifying groups with larger risks, and so on. Quite often causal relations are hidden in company databases and the goal of the data mining process is to induce these from the data and to represent them in meaningful ways to improve business processes. The emphasis will be on the methodological and practical aspects of data mining.

Course aims

After completing the '*database management part*' of the course, students will be able to:

- apply the concepts of *database management* and use the *Structured Query Language* (SQL) to query and analyze business databases;
- comment on the development and business implications of *data transformation* and *data warehouses*;
- design and analyze an *OLAP data cube* for *descriptive business analytics*.

Business Intelligence & Data Management

After completing the '*business intelligence/analytics part*' of the course, students will be able to:

- comment on the fundamental issues of *knowledge discovery in databases*, i.e. (the) data mining (process), such as learning algorithms for classification and prediction. Discuss the business relevance of data mining models.
- interpret the substance of *regression*, fitting and supervised learning, including: interpreting parameters, designing the model, making choices.
- comment on the *Nearest Neighbor algorithm*. Perform classification on data tables with Naïve Bayes.
- design a *decision tree* on data.
- analyze data sets using *frequent item sets / association rules*.
- design *clusters* on data sets with different clustering techniques.
- evaluate performance issues, complexity issues, business relevance and implementation;
- work with (basic) *data mining software*.

Course organization

Course organization entails:

1. Lectures (with video registration);
 - a. Regular lectures on Data Management by Dr. E. Caron (3x).
 - b. Regular lectures on Business Intelligence by Dr. E. Ioannou (6-8x).
2. Weekly computer labs & Online tests;
 - a. Students can only take the lab for which they are registered.
3. Self-study.

Grading

The grading is 100% exam-based, with a bonus arrangement. The mark for the exam should at least be 5 out of 10, to obtain the 0.5 bonus point. *For the retake the bonus point cannot be used*, so the retake has no bonus arrangement. Partial results (exam, bonus) are not valid in the next year.

The following rules are in place to obtain the 0.5 bonus point:

1. The student finishes *all tests* in Testvision within the time limit provided.
2. For each test at least 67% ($>2/3$) of the questions need to be correctly answered.
3. Partial results for the bonus are not given.
4. Tests are made by students individually.
5. Per test a maximum of 3 attempts is allowed.
6. It is not possible to ask questions regarding the tests via e-mail, due to the number of participants in this course. Therefore, we encourage you to *ask questions about the test during the labs* that the test is opened. Only ask questions about the current lab.
7. The test opens on the day of the lab itself and is available for several indicated days. The closing day of the lab is typically before the next lab.
8. There is no retake for a test.

Business Intelligence & Data Management

Reading

The compulsory reading for the course is given by:

- Book:
 - Galit Shmueli, Peter C. Bruce, Nitin R. Patel: '*Data Mining for Business Analytics: Concepts, Techniques, and Applications with XLMiner*', 2016, 3rd Edition, ISBN: 978-1-118-72924. *Or this new edition:*
 - Galit Shmueli, Peter C. Bruce, Peter Gedeck, Nitin R. Patel: '*Data Mining for Business Analytics, Concepts, Techniques and Applications in Python*', 2019, 1st Edition, ISBN: 978-1-119-54986-4. (New book, not widely available yet).
- Scientific articles, made available on Canvas:
 - '*Database systems: design, implementation, and management*' (8 edition), P. Rob and C. Coronell.
 - '*Data Warehouse Design: Modern Principles and Methodologies*' / Golfarelli & Rizzi / Chapter 1.
 - Pedersen, Torben Bach, and Christian S. Jensen. "*Multidimensional database technology*." IEEE Computer, 2001, p.40-46.
 - Some pages on Performance Dashboards, taken from the book *Business Intelligence & Analytics* (Sharda, Delen & Turban).
 - For the labs related to databases we use the online book '*Microsoft SQL Server T-SQL*'
 - All slides from the lectures.
 - Other material made available by the teachers.

Recommended Reading

- Turban et al., Business Intelligence, A Managerial Approach, Pearson, 2011, ISBN 0-13-247882-X.
- Han, J. & M. Kamber, Data Mining, Morgan Kaufmann Publishers, 2011, ISBN 1-55860-901-6.

Software

In the lab you will work with various software environments:

- In labs 1-3, you will work with *MS SQL Server Management Studio*.
 - Download from: <https://docs.microsoft.com/en-us/sql/ssms/download-sql-server-management-studio-ssms?view=sql-server-ver15>
- In lab 3, you work with *MS Excel*, *MS Visio/Draw.IO*, and *MS PowerBI*.
 - Download from <https://powerbi.microsoft.com/en-us/downloads/>
- In labs 4-7, you will work with *Python*, specially *Anaconda* (Spyder, Jupyter, or MS Visual code)
 - Download from: <https://www.anaconda.com/distribution/>

Database prior knowledge

We assume that you have a basic understanding of databases and SQL. You understand tables and relationships between tables and know about retrieving, sorting, and filtering data with SQL. If you do have this prior knowledge of SQL or if you want to refresh the SQL basics a special lab test will be offered in Testvision.

Business Intelligence & Data Management

Python prior knowledge

We assume that you have a basic understanding of the programming concepts within Python. Students should already be able to develop a small Python programs, explain the advantages of functions and modules, and can learn new Python packages on their own. More precisely, students should understand the following programming fundamentals in Python:

- Data types;
- Decisions;
- Iterations;
- Functions;
- Arrays and lists.

If you do not have any programming background – complete “Units 1-5, and units 7 and 8” from –<https://www.codecademy.com/learn/learn-python>. You don’t have to pay for a PRO version of Codecademy. Just create a free account and do what is available in the free version.

In addition, a free book ‘How to Think Like a Computer Scientist: Learning with Python 3 Documentation’ to study the main programming concepts within Python can be found here: <https://buildmedia.readthedocs.org/media/pdf/howtothink/latest/howtothink.pdf>.

Course outline

The course setup might change slightly due to unknown circumstances. Please inspect the table below.

Nr	Date	Contents	Lecturer	Lecture	Video	Lit.	Num./Chap.	Lab date	Lab topics	Test	Lab teachers
									SQL-0 (intro)	0	
1	27-1	Intro. to BI+ Data Management	Caron	yes	yes	Article	I	30-01	SQL-1	1	Caron & Medappa
2	28-1	Data warehousing	Caron	yes	yes	Article	II	6-2	SQL-2	2	Caron & Medappa
3	03-2	OLAP business databases & dashboard	Caron	yes	yes	Article	II, III, IV	13-02	SQL-3 & OLAP	3a & 3b	Caron & Medappa
4	10-2	Data mining introduction	Ioannou	yes	yes	Book	1, 2, 3				
5	10-2	Regression models	Ioannou	yes	yes	Book	2, 6				
6	17-2	Naive Bayes	Ioannou	yes	yes	Book	8	20-02	Bayes & neighbors	4	Ioannou & Medappa
7	17-2	k nearest neighbors	Ioannou	yes	yes	Book	7				
8	27-2	Performance measures	Ioannou	yes	yes	Book	5				
9	2-3	Decision trees	Ioannou	yes	yes	Book	9	5-03	Dec. trees	5	Ioannou & Medappa
10	9-3	Association rules	Ioannou	yes	yes	Book	14	11,12&13-03	Ass. Rules	6	Ioannou & Medappa
11	16-3	Clustering (+20 mins exam preparation)	Ioannou	yes	yes	Book	15	19-03	Clustering	7	Ioannou & Medappa
		Exam 24-3									

Lecturers will give regular lectures and video registrations of lectures will be provided. There are computer labs each week to apply the concepts and techniques introduced in the lectures. In total there are 7 computer labs. Each lab is associated with one or more tests in TestVision. Labs visits are highly recommended, but are non-mandatory. *You can only visit the lab session for which you have registered.*

If you want to participate in the bonus arrangement, *you have to do all tests in TestVision*. We assume that prior to a certain lab session:

- you have visited the lecture, or have watched the video material;
- you have studied the relevant article or book chapter(s).

So, for example, to prepare for the lab on Decision Trees, you have to study the material introduced in the lecture on this topic and you have studied Chapter 9 from the book.