# Business analytics and data mining

## Prof. Marco L. Della Vedova; Prof. Walter Ballardin

### **COURSE AIMS AND INTENDED LEARNING OUTCOMES**

Data is more and more a competitive asset of any organization. Competency in programming is an essential skill for successfully extracting information and knowledge from data. The goal of this course is to build this skill with the help of the Python programming language, which is one of the leading language for data science because it is easy to write, easy to read, and performs well. Moreover, it combines the flexibility of a general purpose programming language with an ecosystem of well-documented libraries specifically designed for processing, manipulating and analyzing data.

The second module aims to gain an understanding of how managers use business analytics to solve business problems and support decision making. Several use cases selected by industries and business processes will be analyzed to give a practical overview of how advanced analytics are applied in organizations. After the course, learners are equipped with proficiency in data management for developing business applications of modern data mining.

***COURSE CONTENT***

Module I - *Principles of programming for data mining* (30h) (Prof. Della Vedova)

1. Introduction to algorithms.

– The role of the algorithms in computer science and in data science.

– Elementary data structures and graphs.

– Introduction to computational complexity theory.

– Programming techniques.

2. Python3 advanced programming.

3. Environments for programming in Python.

– Integrated Development Environments.

– Interactive programming with IPython/Jupyter.

4. Creating and manipulating numerical data with Numpy and Scipy.

5. Data visualization.

6. Connecting databases with SQLAlchemy.

7. Data analysis and data mining with Pandas.

– Import, build, and manipulate DataFrames.

– Exploratory data analysis.

Module 2 - *Applications of data mining in business* (30h) (Prof. Walter Ballardin)

1. Business case overview.
2. Business context exploration and feature analysis.
3. Feature engineering and input features analysis.
4. Machine Learning models.
5. Analysis and presentation of model results.

***READING LIST[[1]](#footnote-1)***

Textbooks

J. VanderPlas, *Python Data Science Handbook,* O'Reilly Media, 2016 – available online at *https://jakevdp.github.io/PythonDataScienceHandbook/*.

Further readings

Scipy Lecture Notes, *http://www.scipy-lectures.org/index.html.*

***TEACHING METHOD***

Lectures, lab projects under professor’s guidance, revision of individual and group assignments. It is strongly advised to attend to lectures in order to face step by step the complex subject of enterprise data management, for working on case studies and examples, and for revising materials.

***ASSESSMENT METHOD AND CRITERIA***

The assessment method is composed by:

– individual written text (during the mid-term week);

– group project work of a practical application.

The final grade is composed equally by both evaluation.

***NOTES AND PREREQUISITES***

The student should have basic knowledge of statistics and python.

*In case the current Covid-19 health emergency does not allow frontal teaching, remote teaching will be carried out following procedures that will be promptly notified to students.*

1. I testi indicati nella bibliografia sono acquistabili presso le librerie di Ateneo; è possibile acquistarli anche presso altri rivenditori. [↑](#footnote-ref-1)