**Introduction to Data Science**

Prof. Eugenio Tacchini

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

The course provides a basic introduction to modern data science using SQL and the Python programming language. It also covers some fundamental concepts of Information Systems in a business environment.

At the end of the course, students will be able:

* to understand the importance and the impact of information in modern organizations and society;
* to identify the various components of an Information System;
* to understand and support the process of software procurement
* to design a database according to the information needs of an organization
* to perform simple SQL queries on relational databases.

Furthermore, student will be introduced to modern data science with the Python programming language: they will be able to do basic exploratory data analysis, to graphically present the results of an analysis with simple graphs using the Matplotlib library and to apply some basic machine learning methods to real data using the Scikit-learn library.

***COURSE CONTENT***

1.Information Systems and Databases

a. Introduction to information systems.

b. Databases: the concept of data, information, database, DBMS.

c. The relational model: the concept of relation, integrity constraints.

d. Designing databases: conceptual and logical design

e. SQL: performing simple SELECT queries on a database

f. Brief overview of: Transaction Processing Systems, Functional Areas Information Systems, Enterprise Resource Planning Systems, Customer Relationship Management Systems, Supply Chain Management Systems.

g. Software Procurement: how organizations procure software

2. Introduction to Data Science

a. What is data science and why it is important;

b. Use case: Recommender Systems;

c. Introduction to computing in Python (Data structures, Flow control structures, Functions);

d. Basic Exploratory Data Analysis;

f. Introduction to data visualization with Matplotlib

g. Introduction to Machine Learning with Python Scikit-learn (possible topics: unsupervised clustering, regressions analysis)

***READING LIST***

Copies of the slides used in class and class notes. The books to be used as references will be indicated during the course.

***TEACHING METHOD***

Lectures, computer laboratory lessons.

***ASSESSMENT METHOD AND CRITERIA***

All students will be graded based on a written final exam. The exam consists of a database design exercise, an SQL exercise, a simple Python exercise and some open questions.

***NOTES AND PREREQUISITES***

The course does not need prerequisites.

***OFFICE HOURS***

Information on office hours available on the teacher's personal page at <http://docenti.unicatt.it/>.