



UNIVERSITÀ
DEGLI STUDI DI MILANO-BICOCCA

SYLLABUS DEL CORSO

Data Science

2425-2-E1803M119

Learning objectives

The objective of this course is to introduce students to programming and provide them with an overview of the data science and artificial intelligence techniques employed in a business environment. It covers the fundamentals of programming (Python), databases (SQL), and machine learning. **All classes and course materials will be in English.**

The course can be taken as an alternative to 'Informatica Generale e Laboratorio Informatico', which focuses instead on providing a general understanding of the issues related to technological innovation in financial companies and society, data and information management, digital business models, business information systems and their organisational and management implications.

At the end of the course, the student will be able to:

- Understand the fundamental concepts of imperative programming;
- Implement simple programs in Python;
- Query a database using the SQL language;
- Understand and apply artificial intelligence and data mining techniques to prepare, analyze, and process data;
- Manage change in a business environment by proposing modern data management and analysis solutions and collaborating in cross-functional teams;
- Understand the trends of the digital world and their impact on businesses, especially in the financial sector and society;
- Acquire a technical language that enables effective communication with both highly skilled technical-informatics professionals and end users;
- Develop strong learning skills that allow for greater autonomy in pursuing advanced studies.

Contents

The course "Data Science" aims to acquaint students with programming and give them a comprehensive

understanding of the data science and artificial intelligence methods utilized within a business setting. The program includes a brief introduction to business information systems, an introductory overview of programming using the Python language, learning the basic concepts of databases using SQL, and gaining practical skills in machine learning techniques. **All classes and course materials will be in English.**

Detailed program

Introduction to Business Information Systems: architectures and applications supporting the operational, tactical, and strategic needs of enterprises.

Digital trends: digital transformation processes in companies and technological evolution in the information society.

Programming Basics: basic concepts of imperative programming and development of simple programs in Python.

Methods for Data Management and Analysis: relational databases, SQL, data warehousing, knowledge graphs, business analytics.

Machine Learning and Data Mining: supervised and unsupervised solutions, with a specific focus on neural networks.

Natural Language Processing: methods for text processing and analysis.

Foundations of Security: risk management and prevention in the business environment.

Foundations of Computer Ethics: an overview of fundamental ethical topics, particularly regarding the use of artificial intelligence.

Computer Lab: the lab component aims to provide students with intermediate-level proficiency in using Microsoft Excel to solve typical business problems.

Prerequisites

Mathematical and logical knowledge acquired in high school. Proficiency in the English language.

Teaching methods

Data Science - 42 hours

- 5 lessons of 2 hours each conducted in an in-person lecture format;
- 16 lessons of 2 hours each, initially delivered in a in-person lecture format, followed by an interactive section (primarily focusing on Python and SQL).

Laboratorio Informatico (Excel) - 12 hours

- 6 laboratory activities of 2 hours each conducted in an interactive in-person format.

Assessment methods

The assessment of learning includes a written test and, optionally, an additional oral examination. The written test, which includes multiple-choice questions and open-ended questions, will take place in the teaching labs (laboratori didattici) in order to evaluate the student's skills in using the Python language and the Microsoft Excel application.

The final exam consists of two sections.

Data Science Section:

- 16 multiple-choice questions (1 correct answer out of 4), each worth 1 point (0 points for incorrect or missing answers). These questions will be randomly selected from a predefined list. Specifically, there will be three types of questions: 1) Data Science and Computer Science theory, 2) Python theory, 3) Python code interpretation (where a small code snippet is provided, and the output is requested).
- 1 Python exercise worth a maximum of 8 points (0 points for a substantially or mostly incomplete response). The exercise requires writing a code snippet to perform a specific task (e.g., iterating over a list and removing all even numbers, finding the minimum in a numeric list).
- 1 SQL query worth a maximum of 8 points (0 points for a substantially or mostly incomplete response; please note that the query must be executable). The exercise (similar to the one presented in the "SQL Exercises - Example SQL question in the final exam" section) will present a database consisting of several tables and will require formulating a specific query (usually involving multiple tables).

This section awards the maximum achievable score in the exam.

Computer Lab Section (Excel):

- 3 additional multiple-choice questions (1 correct answer out of 4) on the contents of the Computer Lab (answering these questions requires the use of Excel). The questions will cover both theory and practice.

Errors or omissions in the Computer Lab section result in the following penalties applied to the score obtained in the General Computer Science section:

- 3 correct answers: no penalty
- 1 incorrect or missing answer: -1 point
- 2 incorrect or missing answers: -3 points
- 3 incorrect or missing answers: -5 points

The additional 3 questions do not need to be answered by: (1) students who have already completed the "Abilità Informatiche Preparazione Tesi" module, and (2) previous academic year students who have already passed the Computer Lab module.

The final grade is determined by the Computer Science section score minus any penalties from the Computer Lab section (e.g., 27 points in the General Computer Science section, -3 points in the Computer Lab section - i.e., 24 points - resulting in a total score of 24/30).

Textbooks and Reading Materials

- The slides used in the classroom and other study materials available on e-Learning platform.
- W3C Python tutorial (<https://www.w3schools.com/python/>)
- Downey, A. B. (2015). Think Python: How to Think Like a Computer Scientist, Version 2.4. Green Tea

Press. (Freely available online with Creative Commons licence,
<https://greenteapress.com/thinkpython2/thinkpython2.pdf>)

Optional textbook

- Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow, 2nd Edition. Aurélien Géron. O'Reilly, 2019.

Semester

First semester.

Teaching language

English.

Sustainable Development Goals

INDUSTRY, INNOVATION AND INFRASTRUCTURE
