Big Data and Predictive Analytics

Lecturer: Maria Chiara Debernardi

Language

English

Course description and objectives

In today's digital environment, organizations generate and store vast amounts of data through transactions, digital platforms, sensors, and online interactions. This Big Data context has created new opportunities for analysis, but it has also increased the need for tools and methods that can transform large, complex datasets into meaningful insights for decision-making.
This course examines Big Data primarily as the context in which predictive analytics operates. The emphasis is not on the technological infrastructure behind Big Data, but on the practical use of data to identify patterns, support forecasts, and inform business decisions. Through hands-on activities, students will learn how data can be prepared, explored, and modeled to extract value from real-world information.
To support this learning process, the course uses KNIME Analytics Platform, an open-source, graphical, no-code environment for data analysis and machine learning. Using KNIME workflows, participants will move from raw data to interpretable results through data preparation, visualization, predictive modeling, and model evaluation.

Upon successful completion of the course, students will be able to:

- explain the role of Big Data as a context for modern predictive analytics
- describe the main phases of a predictive analytics project
- prepare and explore datasets using KNIME Analytics Platform
- build basic predictive models for regression and classification tasks
- evaluate model performance using appropriate metrics and visual tools
- interpret analytical results and discuss their relevance for decision-making
- recognize the opportunities and limitations of predictive analytics in practical settings

Audience

The course is reserved for students enrolled in the Master of Science programs at Bocconi University and is offered as part of the *Enhancing Experience – Curricular Integrative Activities*.

It is designed for students with little or no prior experience in predictive analytics, who want to understand what Big Data means in practice and develop hands-on analytical skills using a no-code analytics platform.
Upon successful completion of the course - requiring **attendance of at least 75%** of scheduled lessons and **passing the final exam** - students will be awarded **2 credits** and an Open Badge, which may be shared online, for example on LinkedIn, or included in their CV.

Prerequisites

No prior coding experience or familiarity with KNIME Analytics Platform is required. It is helpful having a good understanding of:

- descriptive and inferential statistical methodologies, equivalent to a foundational university-level Statistics exam
- basic mathematical concepts with specific emphasis on function optimization techniques (e.g., finding the minimum of a loss function), even if core concepts will be introduced intuitively during the course

Duration

16 academic hours

Teaching mode

Distance learning. The course will be delivered **online** in **synchronous mode**.

NB: please note that the last lecture, including the **final test,** will take place **in person only**. No online alternative will be available.

Calendar

Lecture	Date	Time	Room
1	Wed 03/06/2026	16.30 - 18.00	Virtual room
2	Thu 04/06/2026	16.30 - 18.00	Virtual room
3	Tue 09/06/2026	16.30 - 18.00	Virtual room
4	Wed 10/06/2026	16.30 - 18.00	Virtual room
5	Thu 11/06/2026	16.30 - 18.00	Virtual room

6	Wed 17/06/2026	16.30 - 18.00	Virtual room
7	Thu 18/06/2026	16.30 - 18.00	Virtual room
8	Tue 23/06/2026	14.45 - 16.15	InfoAS04/05

Syllabus of the course

Lecture	Topics	Supporting material
1	**Introduction** - Big Data: definitions and taxonomy - Predictive analytics - The KNIME environment - Building a KNIME workflow - Basic visual exploration ***Exercises***	**Slides** **Parr. 1.1, 1.3.3, 12.1***
2	**Business understanding and data preparation** - CRISP-DM: how to efficiently create a predictive analytics model - Data preparation: the ETL step - Exploring the dataset - Data types, missing values, and encoding ***Exercises***	**Slides** **Parr. 3.3, 4.2, 4.3, 2.2, 2.3, 6.1, 9.2***
3	**Predictive analytics techniques** - Predictive analytics algorithms: characteristics and taxonomy - When to use which model - Sampling: train, test, and cross-validation - Quantitative prediction: regression - Regression metrics ***Exercises***	**Slides** **Parr. 11.5, 1.3, 1.4, 7.1, 7.2, 7.3, 9.1***
4	**The classification problem** - Data preparation for classification tasks - Setting up a classification model - Model performance and evaluation - Confusion matrix - Accuracy, precision, recall, F1-score ***Exercises***	**Slides** **Parr. 9.6, 9.7, 7.4, 7.5***

Lecture	Topics	Supporting material
5	**Classification algorithms** - Naïve Bayes - k-NN - Decision tree - Support vector machine (SVM) - Comparing different models: ROC curve - Hyperparameter tuning ***Exercises***	**Slides** **Parr. 11.1, 10.2,11.2, 11.4, 7.4***
6	**Model ensembles: introductory overview** - Bagging - Boosting - Random forest - Stacking ***Exercises***	**Slides** **Parr. 11.3, 11.2.3***
7	**Intro to Neural Networks** - From linear regression to neural network models - From machine learning to deep learning (*hints only*) ***Exercises***	**Slides** **Parr. 11.6, 9.4***
8	**Q&A and final test – *in person only*** - Guided recap exercise - Last doubts and clarifications - **Exam**	

Software used

KNIME Analytics Platform (knime.com): latest standard release available (5.11.0 or higher)
Download it here
Please note that registration is not required for download, but users must accept the *terms and conditions* of the open-source license.

Suggested bibliography

For the final exam, only the slides and commented exercises provided by the Lecturer are required.
Optional additional reading cited in Supporting material (*):

- Skiena S. S., *The Data Science Design Manual*, Springer, 2017

Available seats

This activity is limited to **110** participants and reserved for **students enrolled in the Master of Science programs**.
Registration is not possible once capacity has been reached or after the registration period has ended.