Prototyping of Data Mining Workflows

(John Puentes)

This course focuses on the practical application of data mining concepts, emphasizing the development and evaluation of prototypes for real-world use cases. Students learn to design, implement, and validate data mining workflows using RapidMiner software, following the CRISP-DM methodology.

Key concepts covered:

- CRISP-DM methodology for data mining projects
- Data handling, visualization, and exploratory analysis
- Classification models (decision trees, rule induction, K-nearest neighbors, Naive Bayes, support vector machines, ensemble modeling)
- Feature selection and parameter optimization
- Clustering techniques
- Contextual performance evaluation of data mining workflows

By the end of this course, students will be able to:

- [BC-03] Evaluate and critically analyze the advantages and limitations of developed data mining workflows
- [BC-04] Analyze and compare prototypes' configurations for improved model performance
- [BC-07] Design and implement data mining workflows using RapidMiner for real-world scenarios

Prerequisites:

- Full understanding of algorithms design principles
- Basis of data mining, performance metrics, supervised and non-supervised learning