COURSE SYLLABUS
Tools for Analytic Lab – SPSS Modeler

Department of Economics / CEU Business School

Term: Winter 2015/16
Credits: 2 (4 ECTS)
Instructor: Clementine Consulting

Course description

IBM SPSS Modeler is an extensive predictive analytics platform with an intuitive graphical interface. It provides a wide range of advanced statistical and data mining algorithms, which allows users to perform complex data analysis and to extract business value from data.

This introductory course aims to give an overview of the fundamentals of using IBM SPSS Modeler through real business cases. The course structure follows the stages of a typical data mining project, from collecting data, to data exploration, data transformation, and modeling to effective interpretation of the results. The last part of this course will be dedicated to additional features and extensions in Modeler, allowing students to add more advanced functionalities. These techniques include text analytics, entity analytics, social network analysis, Big Data and open source integration such as R and Python.

Assessment

50% final exam
50% assignments

Course schedule

1. Working with Modeler: describing the Modeler user-interface, working with nodes and running streams, the basic framework of data analysis in Modeler
2. Importing, exporting, integrating and understanding your data: importing and exporting data from/to databases and from/to various formats, appending records and merging fields from multiple datasets, sampling records, data audit, handling missing data
3. Data transformations: aggregating records, field options: deriving and filling fields, using the reclassify, binning and restructure nodes, string functions
4. Data visualization: charts, diagrams, map visualization
5. Modeling and evaluation: regression, generalized linear models, goodness of fit, overfitting, confounders
6. Modeling: PCA, association rules

7. Clustering and Classification: supervised and unsupervised methods, hierarchical and k-means cluster analysis, logistic regression
8. Machine learning: decision trees, random forest
9. Text Mining: handling and reading unstructured data sources, building patterns and libraries and using the text mining results in Modeler
10. Entity Analytics and Social Network Analysis: improving the coherence and consistency of data by resolving identity conflicts, transforming information about relationships to show the social behavior of individuals and groups
11. Integration of open source tools: R and Python
12. Big Data with Modeler: Analytic Server, Hadoop, Spark