Syllabus

[Tools for Analytics Lab – SPSS Modeler]

- **Instructor:** György Körmendi, Judit Pancza ([KormendiGy@ceu.edu, PanczaJ@ceu.edu])
- **Credits:** 1 (2 ECTS)
- **Term:** Winter 2017-2018
- **Course level:** [PhD or MA/MSc]
- **Prerequisites:** -

Course availability

Cap: 30. Students from the MS in Business Analytics and MS in Finance programs will have direct entry (priority 1) upon registration.

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, Big Data and open source integration such as R and Python.

Learning outcomes

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

- Understand the basic concept of data mining in IBM SPSS Modeler
- Define when and how various data mining techniques should be applied
- Handle unstructured data in order to use in predictive modeling
- Execute basic data mining projects using IBM SPSS Modeler and interpret the result

Assessment

50% final exam

50% assignments

Course schedule and materials for each session

1. Working with Modeler: describing the Modeler user-interface, working with nodes and running streams, the basic framework of data analysis in Modeler. 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.

2. Data transformations: aggregating records, field options: deriving and filling fields, using the reclassify, binning and restructure nodes, string functions. Sampling records, data audit, handling missing data

3. Modeling and evaluation I.: regression, generalized linear models, goodness of fit, overfitting, confounders, clustering methods

4. Modeling and evaluation II.: Machine learning: decision trees, random forest

5. Text Mining: handling and reading unstructured data sources, building patterns and libraries and using the text mining results in Modeler

6. Data Science deployment: Automation, integration of open source tools: R and Python, Big Data with Modeler: Analytic Server, Hadoop, Spark