

## Advanced Predictive Analytics (F000882)

Course size (nominal values; actual values may depend on programme)  
 Credits 8.0 Study time 240 h Contact hrs 80.0 h

### Course offerings and teaching methods in academic year 2018-2019

Offering	Language	Teaching Method	Hours
A (semester 2)	English	lecture	7.5 h
		self-reliant study activities	5.0 h
		seminar: coached exercises	7.5 h
		group work	5.0 h
		seminar: practical PC room classes	55.0 h

### Lecturers in academic year 2018-2019

Van den Poel, Dirk EB23 lecturer-in-charge

### Offered in the following programmes in 2018-2019

Programme	crdts	offering
<a href="#">Master of Science in Marketing Analysis</a>	8	A

### Teaching languages

English

### Keywords

Advanced predictive models, data mining, machine learning, artificial neural networks, mixture models for classification, ensembles, computational statistics, R

### Position of the course

In addition to the required conceptual skills, marketing managers must increasingly master the use of fact-based analysis and decision-models to plan the marketing strategy of their business unit or company. In Advanced Predictive Analytics we will explore the computer-assisted interactive application of decision models.

The course Analytical CRM is a necessary prerequisite to this course as well as Pricing and Revenue Management.

The program aims at providing a thorough scientific training. The overall objective is to train and educate marketing graduates specialised in marketing analysis, who can support the strategy and action plans of a company. Emphasis is placed on the practical use of these analysis techniques within a company setting.

### Contents

Applying advanced statistical methods and machine learning algorithms (including both supervised and unsupervised learning techniques) to solve business and marketing related problems. The focus lies in the understanding and implementation of several supervised and unsupervised learning techniques. For every modeling technique, we will focus on the trade-off between model interpretability and accuracy and their impact on the bias-variance problem. The following supervised learning techniques are discussed:

- (Non-)Linear regression
- Logistic regression
- Decision Trees
- Neural Networks
- Bagging
- Adaboosting
- Random forest
- Rotation forest
- Hybrid ensembles

Next to supervised learning, several unsupervised learning algorithms will be taught:

- Hierarchical clustering
- K-means clustering
- Spectral clustering
- Principal component analysis
- Self-organizing maps (Kohonen maps)

The evaluation of all these techniques is crucial. Thereby, we rely on several performance measures such as AUC, accuracy, F1 and lift for classification and  $R^2$ , RMSEP for regression.

Moreover, we show data visualization, which the students will need to apply to a business case.

#### Initial competences

This course builds on the final competences of the courses "Analytical Customer Relationship Management", "Pricing and Revenue Management" and "Big Data" (SQL), which in turn implies knowledge about intermediate statistics and econometrics. Basic knowledge of R.

#### Final competences

- 1 Mastering of advanced data mining techniques.
- 2 Apply advanced knowledge R to effectively use in real-life projects
- 3 Using the appropriate techniques for model building and developing creative approaches to solving real life problems, based on large data sets.
- 4 Participate in a competition to solve a real life business problem in the most effective way, including exploration of advanced methodology ( R- packages).
- 5 Taking appropriate business decisions (taking into account the different views of all stakeholders) based on the outcomes of analytical/decision models and communicating results en conclusions towards professionals and laymen.
- 6 Exploring oportunities offert by new data sources.

#### Conditions for credit contract

Access to this course unit via a credit contract is determined after successful competences assessment

#### Conditions for exam contract

This course unit cannot be taken via an exam contract

#### Teaching methods

Group work, lecture, self-reliant study activities, seminar: coached exercises, seminar: practical PC room classes

#### Learning materials and price

- Own syllabus
- Scientific papers
- Books:
  - James G., Witten D., Hastie T., Tibshirani R. (2015), An Introduction to Statistical Learning: with Applications in R, Springer.
  - Ballings M. & Van den Poel D. (2016), Analytical Customer Relationship Management Using R.

#### References

Academic articles:

- Ballings, M., Van den Poel, D., 2015. CRM in Social Media: Predicting Increases in Facebook Usage Frequency. *European Journal of Operational Research*.
- Prinzie, A., Van den Poel, D., 2008. Random Forests for multiclass classification: Random MultiNomial Logit. *Expert Systems with Applications* 34, 1721-1732.
- Fernández-Delgado, M., Cernadas, E., Barro, S., Amorim, D., 2014. Do we Need Hundreds of Classifiers to Solve Real World Classification Problems? *Journal of Machine Learning Research* 15, 3133-3181.
- Breiman, L., 2001. Random Forests. *Machine Learning* 45, 5-32.
- Breiman, L., 1996. Bagging predictors. *Mach Learn* 24, 123-140.
- Sokolova, M., Lapalme, G., 2009. A systematic analysis of performance measures for classification tasks. *Information Processing & Management* 45, 427-437.
- Rodriguez, J.J., Kuncheva, L.I., Alonso, C.J., 2006. Rotation Forest: A New Classifier Ensemble Method. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 28, 1619-1630.
- Dietterich, T.G., 2000. Ensemble Methods in Machine Learning, in: *Multiple Classifier Systems*, Lecture Notes in Computer Science. Springer Berlin Heidelberg, pp. 1-15.
- Ngai, E.W.T., Xiu, L., Chau, D.C.K., 2009. Application of data mining techniques in customer relationship management: A literature review and classification. *Expert Systems with Applications* 36, 2592-2602.

### Course content-related study coaching

Numerous exercises are being solved during sessions. In addition, assignments (to be solved in teams) are handed out. Students receive coaching in the process of solving the assignments and feedback afterwards (collectively, by team and individually). After numerous tests about the programming language R (including packages), students will receive individual feedback & coaching.

### Evaluation methods

continuous assessment

Examination methods in case of periodic evaluation during the first examination period

Examination methods in case of periodic evaluation during the second examination period

Examination methods in case of permanent evaluation

Written examination, oral examination, assignment, skills test, peer assessment

Possibilities of retake in case of permanent evaluation

examination during the second examination period is possible in modified form

Extra information on the examination methods

Presentation of a mini-project on predictive analytics

Written exam to determine to what extent the student mastered:

- the higher principles of predictive analytics,
- the higher programming language R,
- the principles of advanced analysis techniques,
- the use of R to solve non-trivial business problems by means of predictive models,
- taking appropriate business decisions (under constraints and integrating the views of the different stakeholders)

Calculation of the examination mark

Written part: 80%

Oral part: 20%

potentially adjusted by peer assessment.