Course ADVANCED METHODS OF DATA ANALYSIS

Basic Information

This is a course, which contributes to MSc award in Mathematics

Duration	**180 hours**
Starting date	September 1st
Study credits	5 ECTS credits
Language of instruction	English level B1(European Framework of Reference of Communicative Skills)
Academic requirements	– BSc degree in Mathematics, Physics, Computer Science, Engineering or equivalent (a copy of your diplomas from previous university studies and transcripts of completes courses and grades) – Skype interview

Course Description

"Advanced methods of Data Analysis" is the second course on data analysis, which broadens students' toolkit of data analysis methods and develops skills of combining technics for building complex models.

Students will deal both with problems of supervised and unsupervised learning, using such advanced instruments as tree-based methods and various type of nets, nonlinear and additive models, time series and a wide range of clustering methods.

Special Features of the Course

The course provides a brief overview of major advanced analytical tools and develops practical skills of implementing these tools in projects.

While programming algorithms in Python – the most popular programming language among data scientists – student will learn its key packages and become independent users of Python.

Course Aim

To introduce advanced methods of data analysis;

- To develop skills of choosing appropriate methods and combining different approaches for achieving better model quality.

Course Objectives

- To broaden the range of familiar analytical tools with logical, probabilistic and nonlinear geometrical methods;

- To introduce time series and observe applied aspects of statistical time series modelling;
- To give students a notion of machine learning ensemble meta-algorithm, such as boosting, bagging and random forests models;
- To observe methods of unsupervised learning for predictive and descriptive analysis;
- To develop skills of thoughtful choosing and programming algorithms.

Learning Outcomes of the Course

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

- formulate the problem of analysis for given datasets and goals of research;
- choose appropriate analytical methods and apply meta-algorithms to produce more powerful models;
- use appropriate Python packages and tools for programming machine-learning algorithms;
- interpret the results of model implementation.

Course (module) Structure

Learning Activities	Hours
Lectures	36
Practice sessions / Seminars,	36
Self-study Assignments	72
Final Exam (including preparation)	36
Total study hours	**180**

Course Outline

Week	Lectures	Practice session / Assignments	Hours[1]
1-4	Time series analysis	– Autocorrelation and Periodic Movements – Stationarity and Nonstationarity – Trends. Volatility – Transforming Time Series – ARMA Models for Stationary Time Series – Home assignment No 1 – Home assignment No 2	18
5-6	Distance-based models	– k nearest neighbors – Parzen-window estimation in classification problems – Home assignment No 3	20
7-8	Logical models	– Tree-based methods – CART modelling – Tree-pruning algorithm – Home assignment No 4	28

[1] Hours designed for Classroom sessions, Web-sessions, Home Assignments etc.

		– Home assignment No 5	
9-11	Ensemble algorithms	– Bagging – Boosting – Random forest – Home assignment No 6 – Home assignment No 7	24
12-15	Deep learning	– Neural networks – Convolutional Neural Networks and Recurrent neural network – Overlearning of neural networks – Neural network ensembles – Home assignment No 8 – Home assignment No 9	20
16-18	Unsupervised learning	– Association Rules – Cluster Analysis – Home assignment No 10 – Home assignment No 11	34
19-20	Final exam		36

Assessment

Grade policy for both practical home assignments and the final exam is:

- A (excellent work) 91–100 points
- B (above average work) 81–90 points
- C (average work) 71–80 points
- D (below average work) 50–70 points
- F (failed work) < 50 points

The exam is taken orally. Each exam ticket consists of 2 theoretical questions from the list of exam questions.

The final grade includes grade for homework and exam in a ratio of 50:50.

Attendance Policy

Students are expected to attend classes regularly. However, occasional skipping classes is permissible if a student does all necessary in-class work at home.

Lecturer(s) and Tutors, Contact Information

Nikolai Kuzenkov
Associate Professor of Applied Mathematics and Computer Safety Department
School of Space and Information Technologies, Siberian Federal University
(room 311) 26-1, Kirensky st, Krasnoyarsk, Russia
Tel: +7 (391) 291-27-90,
nkuzenkov@sfu-kras.ru

Core Reading

All the books are available on-line in Elsevier at SFU library.

1. **Sergios Theodoridis,** Machine Learning, 1st Edition, A Bayesian and Optimization Perspective, Hardcover ISBN: 9780128015223, eBook ISBN: 9780128017227, Imprint: Academic Press, Published Date: 27th March 2015, Page Count: 1062
2. **Ryszard Michalski Jaime Carbonell Tom Mitchell,** Machine Learning, 1st Edition, Hardcover ISBN: 9780934613095, Paperback ISBN: 9781493303489, eBook ISBN: 9780080510545, Imprint: Morgan Kaufmann, Published Date: 28th June 2014, Page Count: 572
3. **Xian Liu,** Methods and Applications of Longitudinal Data Analysis, 1st Edition, Hardcover ISBN: 9780128013427, eBook ISBN: 9780128014820, Imprint: Academic Press, Published Date: 1st September 2015, Page Count: 530