PREDICTIVE ANALYSIS

Basic Information

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

Course period	From September 1st till February 1st, 1 semester
Study credits	5 ECTS credits
Duration	180 hours
Language of instruction	English
Academic requirements	- BSc degree in Biology, Chemistry, Physics or Environmental Sciences or equivalent (transcript of records), - Good command of English (certificate or other official document)

Course Description

The course «Predictive analysis» includes the information from all previous courses and integrates them in one big task of predicting the medical result with a given input.

The course is divided into three steps – the uncertainties estimation, the prediction performance and the estimation of the final prediction. It requires the strong background in mathematics and statistics, the advanced level is mandatory for the successful graduation.

The major part of the course is dedicated to the machine-learning predictions. The students need to apply all his or her knowledge from the courses “Machine Learning in Biomedical Data” and “Classification of Biomedical Data”, as well as other courses.

The final stage of the course is designed to give the students the ability to estimate their own results critically and to choose the best solution from the set of existing ones.

The course is designed to cover a wide range of possible prediction techniques and algorithms, which are special for a given type of the medical dataset.

Course Aims

- To help students study new discipline or to ameliorate the existing knowledge in prediction analysis.
- To assist students to estimate their own results of medical data prediction algorithms.
- To give students the understanding of how to estimate the errors of the prediction.
- To introduce the basic principles of dealing with the uncertainties in the medical data.

Course Objectives

- To ensure that students are familiar with a basic prediction strategies.
- To ensure that students know the verification and validation techniques, which are mandatory for the prediction analysis.
- To give students the prediction and optimal control algorithms for the medical data analysis.
- To give students an appreciation of the uncertainty propagation in the prediction routines.
- To provide students with the principles of the prediction strength estimation.

Learning Outcomes of the Course

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

- range the features of the given phenomenon and choose the predictor for the analysis,
- construct the machine-learning algorithm for the prediction,
- estimate the prediction result.

Course Structure

Learning Activities	Hours
Lectures	18
Practice sessions / Seminars	36
Self-study Assignments	90
Final Exam (including preparation)	36
Total study hours	**180**

Course Outline

Week	Lectures	Practice session / Assignments	Hours
1-6	Introduction to the Predictive Analysis	• Uncertainty quantification • Stochastic processes • Input parameter distributions • Home assignment No 1	(6/12/30)
7-12	Statistical and Machine-Learning Predictions	• Forecasting with regression models • Black box machine-learning methods • Simulation and Surrogate usage for prediction • Local sensitivity analysis • Home assignment No 2	(6/12/30)
13-18	Predictive and Optimal Control	• Feature selection • Uncertainty propagation • Nonlinear model predictive and optimal control • The predictor importance • Home assignment No 3	(6/12/30)
19	Final Exam		36

Course Instructors and Tutors, Contact Information

Instructor	Contact Information
Andrey SHUVAEV Ph.D. in Biophysics, Associate professor, School of Biology and Biotechnology Siberian Federal University, Krasnoyarsk	Svobodny, 79 Room 32-13 Tel +7 391 206 2165 Email: AShuvaev@sfu-kras.ru, andrey.n.shuvaev@gmail.com

Assessment

The final exam is an individual problem in the form of a simulated experimental dataset. The student must demonstrate the abilities in:

• calculation of QoI and uncertainties propagation in the prediction analysis (20 points maximum),

• construct the machine-learning algorithm for the prediction analysis (40 points maximum),

• estimate the final prediction and the predictor strength (40 point maximum).

Grade policy:

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

Core reading

1. Dinov, I. (2018) *Data Science and Predictive Analytics*. Springer, Cham.
2. Grune, L. and Pannek, J. (2011) *Nonlinear Model Predictive Control*. Springer, Cham.
3. Grancharova, A. and Johansen, T.A. (2012) *Explicit Nonlinear Model Predictive Control*. Springer, Berlin, Heidelberg.
4. McClaren, R.G. (2018) *Uncertainty Quantification and Predictive Computational Science*. Springer, Cham.
5. Cinar, A. and Turksoy, K. (2018) *Advances in Artificial Pancreas Systems*. Springer, Cham.

6. Berliner, L. and Lemke, H. U. (ed.) (2015) *An Information Technology Framework for Predictive, Preventive and Personalised Medicine*. Springer, Cham.
7. Kuhn, M., and Johnson, K. (2013) *Applied Predictive Modeling*. Springer, New York, NY.