Z1-PU7	WYDANIE N1

(*faculty stamp*)

COURSE DESCRIPTION

1. Course title: MACHINE LEARNING, Statistical learning	**2. Course code ML_SL**
3. Validity of course description: 2018/2019	
4. Level of studies: MSc programme	
5. Mode of studies: intramural studies	
6. Field of study: CONTROL, ELECTRONIC AND INFORMATION ENGINEERING (MACRO)	(FACULTY SYMBOL) RAU-2
7. Profile of studies: ACADEMIC	
8. Programme: DATA SCIENCE	
9. Semester: 2	
10. Faculty teaching the course: Faculty of Automatic Control, Electronics and Computer Science	
11. Course instructor: Prof. dr hab. inż. Joanna Polańska	
12. Course classification: common courses	
13. Course status: compulsory ~~/elective~~	
14. Language of instruction: English	
15. Pre-requisite qualifications: Algebra and analytic geometry, Calculus and differential equations, Physics, Computer programming, Optimization methods, Numerical methods, Statistics and probability theory, Algorithms and data structures. Classifiers.	
16. Course objectives: The aim of the course is making students familiar with statistical problems related to machine learning, feature engineering for classification, model rank estimation and model selection in the aspect of machine learning, regression models for machine learning, model integration and analyses of significantly correlated datasets.	

17. Description of learning outcomes:

Nr	Learning outcomes description	Method of assessment	Teaching methods	Learning outcomes reference code
1.	Student understands statistical character of problems arising in classification of data.	Exam	Lecture	K2A_W02, K2A_W04
2.	Student understands the notion and character of high throughput data and the importance of adjusting suitable methods for their analysis.	Exam	Lecture	K2A_W06, K2A_W07, K2A_W08
3.	Student has knowledge on methods of estimating reliability of classification models.	Exam	Lecture	K2A_W07, K2A_W08
4.	Student has knowledge on principles of integration of machine learning methods.	Exam	Lecture	K2A_W10, K2A_W26
5.	Student can use different algorithms for feature selection and signature construction.	Laboratory tasks	Laboratory	K2A_U14, K2A_U15, K2A_K01, K2A_K02
6.	Student can use algorithms of integration of statistical models.	Laboratory tasks	Laboratory	K2A_U14, K2A_U15
7.	Student understands the influence of correlations on statistical inference.	Exam	Lecture	K2A_W01, K2A_W02

8.				
9.				

18. Teaching modes and hours

Lecture 30 / ~~BA /MA Seminar / Class / Project~~ / Laboratory 30

19. Syllabus description:

Lecture:

1. Introductory information. Topics and importance of statistical learning in data analysis. Reliability of inference based on classifiers. High throughput data.
2. Feature engineering for classification. Overview and comparisons of methods and algorithms for feature selection. Assessments of the influence of data sets signatures on the quality of classification. Tradeoffs between precision and generalizability.
3. Model rank estimation and model selection. Information criteria. Stepwise algorithms for constructions of statistical models and classifiers. Model quality assessment and model validation.
4. Regression models as backgrounds for classification methods. Linear and generalized linear regression, logistic regression, Poisson regression, additive models and pursuit regression.
5. Data compression and data representation by unsupervised models. Data mining by unsupervised models.
6. Integration of statistical models. Integration vectors of p-values. Integration of supervised and unsupervised classifiers.
7. Supervised and unsupervised classification methods for significantly correlated datasets.

Laboratory:

1. Algorithms for computing gene signatures for DNA expression data.
2. Analysis of the MILE gene expression dataset.
3. Feature engineering.
4. Application of "Spectre" algorithm and program for Maldi ToF proteomic datasets.
8. Problems in analysis of "radiomics" data. Multistep algorithms for radiomics data analysis.
9. Model rank estimation

20. Examination: semester: 2

21. Primary sources:

Shai Shalev-Shwartz, Shai Ben-David, (2014), Understanding Machine Learning:
From Theory to Algorithms, Cambridge University Press.

22. Secondary sources:

T. Hastie, R. Tibshirani, J. Friedman, (2008), The elements of statistical learning, Springer

23. Total workload required to achieve learning outcomes

Lp.	Teaching mode :	Contact hours / Student workload hours
1	Lecture	30/30
2	Classes	/
3	Laboratory	30/30
4	Project	/
5	BA/ MA Seminar	/
6	Other	/
	Total number of hours	60/60

24. Total hours: 120

25. Number of ECTS credits: 4

26. Number of ECTS credits allocated for contact hours: 2

27. Number of ECTS credits allocated for in-practice hours (laboratory classes, projects): 2

26. Comments:

Approved:

…… ……

(date, Instructor's signature) (*date , the Director of the Faculty Unit signature)*