Zał. nr 4 do ZW 64/2012

FACULTY OF PURE AND APPLIED MATHEMATICS
SUBJECT CARD

Name in Polish ANALIZA NIEUPORZĄDKOWANYCH ZBIORÓW DANYCH
Name in English ANALYSIS OF UNSTRUCTURED DATA
Main field of study (if applicable): APPLIED MATHEMATICS
Specialization (if applicable): COMPUTATIONAL MATHEMATICS
Level and form of studies: ~~1st~~/ 2nd* level, full-time / ~~part-time*~~
Kind of subject: ~~obligatory~~ / optional ~~/ university-wide*~~
Subject code MAT001578
Group of courses YES / ~~NO~~*

	Lecture	Classes	Laboratory	Project	Seminar
Number of hours of organized classes in University (ZZU)	30			30	
Number of hours of total student workload (CNPS)	150				
Form of crediting	crediting with grade				
For group of courses mark (X) final course	X				
Number of ECTS points	5			3	
including number of ECTS points for practical (P) classes	2			2	
including number of ECTS points for direct teacher-student contact (BK) classes	1,5			1,5	

*delete as applicable

PREREQUISITES RELATING TO KNOWLEDGE, SKILLS AND OTHER COMPETENCES

1. Student has basic programming skills.

\

SUBJECT OBJECTIVES

C1 Searching, extracting, storing ond computer-aided analysis of unstructered data (texts, blogs, web sites, social media posts etc.)

SUBJECT EDUCATIONAL EFFECTS

relating to knowledge:
PEK_W12 knows how to use Python and its scientific modules for data analysis

relating to skills:
PEK_U12 can perform an analysis of unstructured data by making use of Python and its modules

relating to social competences:
PEK_K06 can, without assistance, search for necessary information in the literature, also in foreign languages

PEK_K02 can accurately formulate questions for deeper understanding of a given topic

PROGRAMME CONTENT

Form of classes - lecture		Number of hours
Lec 1	Data analysis in Python – PANDAS primer	8
Lec 2	Retrieving and storing data	6
Lec 3	Data visualisation	2
Lec 4	Data wrangling	2
Lec 5	Natural language processing with NLTK	4
Lec 6	Sentiment analysis	2
Lec 7	Document classification	4
Lec 8	Handling big data	2
	Total hours	**30**

Form of classes - project		Number of hours
Pr1	Practical Preparation and presentations of projects illustrating methods given in the lectures.	30
	Total hours	**30**

TEACHING TOOLS USED

N1. Lecture – traditional method and presentations
N2. Student partial project presentation and final presentation
N3. Consultations
N4. Student's self work – work related to the project development

EVALUATION OF SUBJECT EDUCATIONAL EFFECTS ACHIEVEMENT

Evaluation (F – forming (during semester), P – concluding (at semester end)	Educational effect number	Way of evaluating educational effect achievement
F1	PEK_W12 PEK_U12	mid-term exams
F2	PEK_U12 PEK_K06 PEK_K02	Oral presentations
C P==0.5*F1+0.5*F2		

PRIMARY AND SECONDARY LITERATURE
PRIMARY LITERATURE: [1] S. Bird, E. Klein i E. Loper, „Natural Language Processing with Python” [2] I. H. Witten & E. Frank, „Data Mining. Practical Machine Learning Tools and Techniques” **[3]** W. McKinney, „Python for Data Analysis” **SECONDARY LITERATURE:** [1] P. Giudici, „Applied Data Mining” [2] T. Segaran, „Programming Collective Intelligence” [3] I. Idris, „Python Data Analysis”
SUBJECT SUPERVISOR (NAME AND SURNAME, E-MAIL ADDRESS)

MATRIX OF CORRELATION BETWEEN EDUCATIONAL EFFECTS FOR SUBJECT
ANALYSIS OF UNSTRUCTURED DATA MAT001578
AND EDUCATIONAL EFFECTS FOR MAIN FIELD OF STUDY **APPLIED MATHEMATICS**
AND SPECIALIZATION COMPUTATIONAL MATHEMATICS

Subject educational effect	Correlation between subject educational effect and educational effects defined for main field of study and specialization (if applicable)**	Subject objectives***	Programme content***	Teaching tool number***
PEK_W12	K2MST_W12 K2MST_cm_W01 K2MST_cm _W02 K2MST_cm _W03	C1	Lec1-Lec8	1,3
PEK_U12 (skills)	K2MST_U21, K2MST_U20 K2MST_U24 K2MST_U25 K2MST_cm _U01 K2MST_cm_U02 K2MST_cm_U03	C1	Pr1	2,3,4
PEK_K02 **PEK_K06** **(competences)**	K2MST_K02, K2MST_K06 K2MST_cm_K01 K2MST_cm_K02	C1	Lec1-Le8, Pr1	1,2,3,4

** - enter symbols for main-field-of-study/specialization educational effects
*** - from table above