FACULTY OF PURE AND APPLIED MATHEMATICS
SUBJECT CARD

Name of subject in Polish: ANALIZA NIEUPORZĄDKOWANYCH ZBIORÓW DANYCH
Name of subject in English: ANALYSIS OF UNSTRUCTURED DATA
Main field of study (if applicable): APPLIED MATHEMATICS
Specialization (if applicable): COMPUTATIONAL MATHEMATICS
Profile: academic / practical*
Level and form of studies: ~~1st~~/ 2nd* level, full-time / ~~part-time*~~
Kind of subject: ~~obligatory~~ / optional ~~/ university-wide*~~
Subject code:
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)	90			60	
Form of crediting	crediting with grade				
For group of courses mark (X) final course	X				
Number of ECTS points	3			2	
including number of ECTS points for practical classes (P)	2			2	
including number of ECTS points corresponding to classes that require direct participation of lecturers and other academics (BU)	1,5			1,5	

*delete as not necessary

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:
PEU_W12 knows how to use Python and its scientific modules for data analysis
Relating to skills:
PEU_U12 can perform an analysis of unstructured data by making use of Python and its modules
Relating to social competences:
PEU_K06 can, without assistance, search for necessary information in the literature, also in foreign languages

PEU_K02 can accurately formulate questions for deeper understanding of a given topic

PROGRAMME CONTENT		
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**

Project		**Number of hours**
Pr1	Practical Preparation and presentations of projects illustrating methods given in the lectures.	30
	Total hours	30

TEACHING TOOLS USED
1. Lecture – traditional method and presentations 2. Student partial project presentation and final presentation 3. Consultations 4. Student's self work – work related to the project development

EVALUATION OF SUBJECT LEARNING OUTCOMES ACHIEVEMENT

Evaluation (F – forming (during semester), P – concluding (at semester end)	Educational effect number	Way of evaluating educational effect achievement
F1	PEU_W12 PEU_U12	mid-term exams
F2	PEU_U12 PEU_K06 PEU_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)

JANUSZ SZWABIŃSKI