

INTRODUCTION TO DATA SCIENCE WITH PYTHON

University of Ljubljana

FACULTY OF
ECONOMICS

Doctoral seminar (6 ECTS)

Faculty of Economics, University of Ljubljana

September 2015

SEMINAR SUMMARY

The Introduction to Data Science with Python seminar will consist of two parts. In the first part, we will cover the basics of Python programming language. In the second part we will survey some of the foundational topics in Data Science, such as Data Analysis, Data Visualization, Web Intelligence, and working with Big Data. The use of several practical tools and libraries in Python will be introduced through various case studies and practical examples from different fields of Economics.

SEMINAR OBJECTIVES

The students will first learn the basics of Python programming language, and revise the basic programming concepts such as variables, functions, loops, and data structures like lists, sets, or dictionaries. The students will then learn how to use various state-of-the-art Python tools for data scientists, and how to apply them on real-world problems in Economics.

SCHEDULE

Day 1-2: Introduction to programming in Python

installation, basic overview, control flow statements, functions, data structures, file manipulations, using modules

Day 3-4: Data Analysis, Data Visualization, Web Intelligence

Python libraries: (1) SciPy library for scientific computing, (2) NumPy for working with N-dimensional array packages, (3) Matplotlib for comprehensive data visualization; web scraping, string manipulations, regular expressions, text mining, basic data mining

Day 5: Big Data

working with a lot of data: practical case studies from marketing and finance

The seminar will be conducted in five (5) consecutive weekdays in March or April 2016. The exact dates can be determined with respect to preferences of the participants.

GRADING

The students will use Python to analyze a problem or some data related to a problem, and then describe their experiment in a research paper (8 – 12 pages). Students can either use their own research data (e.g., data from their Ph.D. research), or data provided by the lecturers. The students will have enough time (after the seminar) to carry out the research and describe their work. Every paper will be carefully examined. Grading will be based on the research paper written by the student. Precise instructions will be given during the course.

REQUIREMENTS AND LITERATURE REVIEW

The ideal students for this class are prepared individuals who have strong interest in data science, some background in intro level statistics, and some basic programming skills.

The recommended literature:

- Magnus Lie Hetland. *Beginning Python: From Novice to Professional*. Apress, 2009.
- Y. Daniel Liang. *Introduction to Programming Using Python*. Pearson, 2013.
- Wes Mckinney. *Python for Data Analysis*. O'Reilly Media, 2013.
- Yuxing Yan. *Python for Finance*. Packt Publishing Ltd., 2014.

SEMINAR INSTRUCTORS

Matej Guid, Faculty of Computer and Information Science, University of Ljubljana, is assistant professor of Computer Science and Chief Data Scientist in a multinational corporation, with more than ten years of research experience from the field of Artificial Intelligence.

Martin Možina, Faculty of Computer and Information Science, University of Ljubljana, is a senior researcher and an experienced lecturer, also with more than ten years of research experience from the field of Artificial Intelligence.