

**Федеральное государственное автономное образовательное учреждение
высшего образования
"Национальный исследовательский университет
"Высшая школа экономики"**

**Программа учебной дисциплины
«Введение в машинное обучение в финансах»
(Introduction to Machine Learning for Finance)**

Утверждена

Академический совет основных образовательных программ по направлениям подготовки
38.03.01 Экономика, 38.04.08 Финансы и кредит

Протокол № от _____ 2019

Академический руководитель ОП

_____ Белых С.А.

Подпись ФИО

Разработчик	Найденова Юлия Николаевна, доцент, департамент экономики и финансов
Число кредитов	4
Контактная работа (час.)	2
Самостоятельная работа (час.)	150
Образовательная программа, курс	Экономика, направление подготовки 38.03.01 Экономика, уровень бакалавриат, 4 курс
Формат изучения дисциплины	На английском языке, с использованием онлайн курса

Syllabus
Machine Learning in Finance
(4 ECTS)

Iuliia Naidenova, lecturer (e-mail: naidenovayn@gmail.com, web-page: <https://www.hse.ru/en/org/persons/14538835>)

School of Economics and Finance

Meeting Minute # ___ dated _____ 20__

1. Course Description

a) Pre-requisites

- English
- Microeconomics
- Corporate Finance
- Statistics
- Econometrics
- Familiarity with programming language such as R or Python is advantageous.

b) Abstract

The aim of the course is to apply main financial concepts and make students acquainted with machine learning techniques relevant for finance.

Python is a general-purpose programming language that is becoming ever more popular for data science. The course focuses on Python specifically for data science. The course is about ways to import, store and manipulate data, and helpful data science tools to conducting data analyses. The course is intended for students with basic background in finance, statistical methods. Experience in programming is not required, but advantageous. The learning process is facilitated with DataCamp platform.

2. Learning Objectives

At the end of the course, students should be able to write short scripts to import, prepare and analyze financial data for making decisions.

3. Learning Outcomes

By the end of the course, a student develops the following competencies, where

GK – general competence

GPK – general professional competence

Code	Competencies
GK-3	Is able to solve problems in professional activities based on analysis and synthesis
GK-5	Is able to work with information: find, evaluate and use information from various sources, necessary to solve scientific and professional problems (including on the basis of a systematic approach)

GPK-3	Is able to analyze socially significant problems and processes and predict their possible development in the future
GPK-11	Is able to collect, analyze and process statistical data, information, scientific and analytical materials needed to solve economic problems
GPK-12	Is able to choose tools for processing economic data in accordance with the task, analyze the results of calculations and substantiate the findings
GPK-17	Is able to use modern hardware/software and information technologies for solving analytical and research problems.

4. Course Plan

№	Topics	All hours	Lecture hours	Seminar hours	Self-study hours
Topic 1. Introduction to Python for Finance					
1	Lists and Arrays	20	-	-	30
2	DataFrames	20	-	-	20
3	Importing, cleaning and merging data	20	-	-	20
Topic 2. Statistical Methods in Python					
4	Statistical Thinking in Python	20	-	-	20
5	Introduction to Time Series Analysis in Python	20			20
Topic 3. Machine learning in Python					
6	Supervised Learning with scikit-learn	20	-	-	20
7	Machine Learning for Finance in Python	20	-	-	20
Topic 4. Environment for scientific programming in Python					
8	Jupyter Notebook	12	2	-	10
Sum		152	2	0	150

Topic 1. Introduction to Python for Finance

1. Lists and Arrays.

Introduction to basics in Python, including how to name variables and various data types in Python. NumPy and Matplotlib packages.

<https://www.datacamp.com/courses/intro-to-python-for-finance>

2. DataFrames.

Using of pandas to import and inspect a variety of datasets. Building DataFrames and the intrinsic data visualization capabilities of pandas.

<https://www.datacamp.com/courses/pandas-foundations>

3. Importing, cleaning and merging data

Importing, cleaning and combining data from Excel workbook sheets into a pandas DataFrame. Grouping data, summarizing information for categories, and visualizing the result using subplots and heatmaps.

<https://www.datacamp.com/courses/importing-managing-financial-data-in-python>

Topic 2. Statistical Methods in Python

4. Statistical Thinking in Python

The principles of statistical inference. Graphical exploratory data analysis, quantitative exploratory data analysis, statistical inference for discrete and continuous variables.

<https://www.datacamp.com/courses/statistical-thinking-in-python-part-1>

5. Introduction to Time Series Analysis in Python

Correlation and autocorrelation, autoregressive (AR) models, moving average (MA) and ARMA models in Python.

<https://www.datacamp.com/courses/introduction-to-time-series-analysis-in-python>

Topic 3. Financial data analysis in Python

6. Supervised Learning with scikit-learn

Building predictive models, tuning their parameters, and determining how well they will perform with unseen data. Scikit-learn library for machine learning in Python.

<https://www.datacamp.com/courses/supervised-learning-with-scikit-learn>

7. Machine Learning for Finance in Python

Calculation of technical indicators from historical stock data, the historical stock data analysis. Linear models, xgboost models, and neural network models. Decision trees, random forests, and neural networks to predict the future price of stocks in the US markets.

<https://www.datacamp.com/courses/machine-learning-for-finance-in-python>

Topic 4. Environment for scientific programming in Python

8. Jupiter Notebook as an environment for scientific programming in Python, its structure and features.

5. Reading List

a) Required

Vanderplas JT. Python Data Science Handbook : Essential Tools for Working with Data. First edition. Sebastopol, CA: O'Reilly Media; 2016. Available from: search.ebscohost.com/login.aspx?direct=true&db=nlebk&AN=1425081

b) Optional

Thomas, Seemon. Basic Statistics, Alpha Science Internation, 2014. ProQuest Ebook Central. Available from: <https://ebookcentral.proquest.com/lib/hselibrary-ebooks/detail.action?docID=5190782>

Hastie, T., Tibshirani, R., Friedman, J., & Franklin, J. (2005). The elements of statistical learning: data mining, inference and prediction. The Mathematical Intelligencer, 27(2), 83-85. Available from: <https://link.springer.com/book/10.1007%2F978-0-387-84858-7>

6. Grading System

The teacher assesses a self-study students' work as fulfilment of tasks in DataCamp. Each fulfillment task before deadline gives 2 point, each fulfillment task after deadline gives 1 point. The whole grade (O_{DC}) is calculated as a percentage of maximum possible grade and then converted to 10 point scale.

And **the resulting course grade** is calculated as follows:

$$O_{result} = 0,5 * O_{DC} + 0,5 * O_{exam}$$

with O_{exam} – a grade for the exam on a 10 point scale, O_{DC} – a grade for fulfilment of tasks in DataCamp (10 point scale).

7. Examination Type

Final student assessment (exam) is a project that is performed in a team of no more than 2 people. Each team uses provided dataset and apply one or a combination of the learnt methods of data analysis in Python. As a result of the project each team write down the report and prepare working file. The grade for the exam includes the grade for the report, grade for the working file and the grade for answering questions.

Example of exam tasks:

1. Import the file “data_for_exam.xlsx”
2. Describe the variable X1 – its type, number of observations, descriptive statistics.
3. Create a dataframe named df1 from the dataset
4. Predict future price changes of stocks using a linear model.
5. Predict future price changes of stocks using machine learning tree methods.
6. Predict future price changes of stocks using neural networks.
7. Compare the predictions and actual future returns. Which prediction is better? Why? Plot the results.

8. Methods of Instruction

At the start of the course, students receive the schedule for DataCamp courses fulfillment. At the end of the course, the lecture that combines the material of the course and a discussion of particular aspects take place.

9. Special Equipment and Software Support (if required)

Software that is essential for the course is:

- Access to DataCamp
- Anaconda
- Microsoft office
- Chrome or any other web browser