

**Санкт-Петербургский филиал федерального государственного
автономного образовательного учреждения высшего образования
«Национальный исследовательский университет
"Высшая школа экономики"»**

Факультет Санкт-Петербургская школа экономики и менеджмента
Департамент менеджмента

Рабочая программа дисциплины

Машинное обучение и его применение для решения финансовых задач /
Machine Learning and its Application for Finance

для образовательной программы «Финансы»
направления подготовки 38.04.08 «Финансы и кредит»
уровень магистратура

Разработчик: Терников Андрей Александрович, aternikov@hse.ru

Согласована начальником ОСУП магистратуре
Неклюдова М.А.

«_____» _____ 2019 г.

Утверждена Академическим руководителем образовательной программы
магистратура «Финансы»

Санкт-Петербург, 2019

*Настоящая программа не может быть использована другими подразделениями
университета и другими вузами без разрешения подразделения-разработчика программы.*

Аннотация

Название дисциплины	Машинное обучение и его применение для решения финансовых задач		
Образовательная программа	«Финансы» 38.04.08 «Финансы и кредит»		
Тип дисциплины	Обязательный		
Требования к уровню знаний студентов, необходимых для освоения дисциплины (пререквизиты)	Базовые знания статистики и навыки работы в среде R		
Объем з.е.	4		
Объем в часах	Аудиторная работа	Самостоятельная работа	Всего
	40	112	152
Краткое описание курса	Практико-ориентированный курс, посвященный анализу данных и методам их обработки при помощи алгоритмов машинного обучения в специализированных программных средах таких как Python. Основным принципом является изучение существующих закономерностей в данных, построение моделей прогнозирования для решения сложных задач в сфере финансов.		
Образовательные результаты по дисциплине	<ul style="list-style-type: none"> Способен к самостоятельному освоению новых методов исследований, изменению научного и производственного профиля своей деятельности Способен принимать управленческие решения и готов нести за них ответственность Способен выявлять данные, необходимые для решения поставленных исследовательских задач в сфере управления; осуществлять сбор данных, как в полевых условиях, так и из основных источников социально-экономической информации: отчетности организаций различных форм собственности, ведомств и т.д., баз данных, журналов, и др., анализ и обработку этих данных, информацию отечественной и зарубежной статистики о социально-экономических процессах и явлениях Способен выбирать инструментальные средства, современные технические средства и информационные технологии для обработки информации в соответствии с поставленной научной задачей в сфере управления 		
Краткое содержание дисциплины	<ol style="list-style-type: none"> Анализ данных в MS Excel Введение в язык программирования Python Обработка данных в Python Визуализация данных в Python Сбор данных из открытых Интернет-источников в Python Алгоритмы машинного обучения в сфере финансов 		
Образовательные технологии	<ul style="list-style-type: none"> Каждую неделю дается 2-часовой практический семинар для изучения методов обработки данных и алгоритмов машинного обучения. Развитие навыков программирования в Python. Интерактивные дискуссии и презентации, направленные на то, чтобы сделать обучение более интересным и мотивировать студентов на просмотр материалов лекций и семинаров для получения бонусных баллов. 		
Формы контроля	<p>Формула оценки: $O_{\text{итоговая}} = 0.25 * (O_1 + O_2 + O_3 + O_4)$</p> <p> O_1 – лабораторная работа в MS Excel O_2 – домашнее задание в Python O_3 – лабораторная работа в Python O_4 – проект в Python </p> <p> O_1, O_2, O_3, O_4 – целые числа от 0 до 10 включительно. $O_{\text{итоговая}}$ – итоговая оценка, которая округляется по математическим правилам </p>		
Литература	<p><u>Основная</u></p> <p>Muller, A. C., & Guido, S. (2017). Introduction to machine learning with Python: a guide for data scientists. O'Reilly Media. (HSE access: http://ebookcentral.proquest.com/lib/hselibrary-ebooks/detail.action?docID=4698164)</p>		

	<p><u>Дополнительная</u></p> <ol style="list-style-type: none"> 1. Cirillo A. R Data Mining. Birmingham, UK: Packt Publishing; 2017. https://proxylibrary.hse.ru:2119/login.aspx?direct=true&db=nlebk&AN=1643003. 2. Vanderplas, J.T. (2016). Python data science handbook: Essential tools for working with data. Sebastopol, CA: O'Reilly Media, Inc. https://proxylibrary.hse.ru:2119/login.aspx?direct=true&db=nlebk&AN=1425081.
Преподаватель	Терников Андрей Александрович, преподаватель, деп.менеджмента

Title of the course	Machine Learning and its Application for Finance				
Title of the Academic Programme	38.04.08. Finance Master program “Finance”				
Type of the course	Compulsory				
Prerequisites	Basic knowledge of statistics and R programming language.				
ECTS workload	3				
Total indicative study hours	Directed Study	Self-directed study	Total		
	40	112	152		
Course Overview	During this practically oriented data science module students will learn how machine learning uses computers to run predictive models. The main principal is to explore existing data to build new knowledge, forecast future behaviour, anticipate outcomes and trends. Explore theory and practice, and work with tools like Python to solve advanced data science problems.				
Intended Learning Outcomes (ILO)	Upon completing this course, students should be able to do the following: <ul style="list-style-type: none">• Collect, store, process and analyze data automatically with the use of scripting languages.• Develop and apply new research methods of basic machine learning algorithms and ways to collect information using data mining techniques.• Solve economic, financial and managerial problems using best practices of data analysis using modern computational tools.• Can identify the data needed for addressing the financial and business objectives.				
Teaching and Learning Methods	<ul style="list-style-type: none">• Every week a 2-hour tutorial is given to practice real-world data mining and machine learning skills.• Developing programming skills in Python in desktop and web-based interfaces.• Interactive discussions and presentations aimed to make learning more engaging and motivate students to review the lecture and seminar materials to earn bonus points.				
Content and Structure of the Course (see Annex 1)					
№	Topic / Course Chapter	Total	Directed Study	Self-directed Study	
			Lectures	Tutorials	
1	Data Analysis in MS Excel	12	6	6	27
2	Introduction to Python	4	2	2	15
3	Managing Datasets in Python	4	2	2	15
4	Data Visualisation	4	2	2	15
5	Getting Data from Web	4	2	2	15
6	Machine Learning Algorithms in Finance	8	4	4	25
Total study hours		40	20	20	112
Indicative Assessment Methods and Strategy		In-class Activity <ul style="list-style-type: none">• Regular weekly in-class tasks on which students are required to do some Python coding and interpretation applied to business and scientific tasks.• Presentations and discussion of topics are related to data science application in finance.			

	<u>Assessment (see Annex 2)</u> Assessment formula: $O_{\text{final}} = 0.25 * (O_1 + O_2 + O_3 + O_4)$ O ₁ – Lab in MS Excel O ₂ – Hometask in Python O ₃ – Lab in Python O ₄ – Project in Python O ₁ , O ₂ , O ₃ , O ₄ – integer numbers from 0 to 10 including. O _{final} – final grade rounded with mathematical rules.		
Readings / Indicative Learning Resources	<u>Mandatory</u> Muller, A. C., & Guido, S. (2017). Introduction to machine learning with Python: a guide for data scientists. O'Reilly Media. (HSE access: http://ebookcentral.proquest.com/lib/hselibrary-ebooks/detail.action?docID=4698164) <u>Optional</u> 1. Cirillo A. R Data Mining. Birmingham, UK: Packt Publishing; 2017. https://proxylibrary.hse.ru:2119/login.aspx?direct=true&db=nlebk&AN=1643003 . 2. Vanderplas, J.T. (2016). Python data science handbook: Essential tools for working with data. Sebastopol, CA: O'Reilly Media, Inc. https://proxylibrary.hse.ru:2119/login.aspx?direct=true&db=nlebk&AN=1425081 .		
Indicative Self- Study Strategies	Type	+/-	Hours
	Reading for seminars / tutorials (lecture materials, mandatory and optional resources)	+	52
	Assignments for seminars / tutorials / labs	+	20
	E-learning / distance learning (MOOC / LMS)	+	10
	Project work	+	30
Academic Support for the Course	Academic support for the course is provided via LMS and Slack conference, where students can find: guidelines and recommendations for doing the course; guidelines and recommendations for self-study; samples of assessment materials. Several practical tasks are provided in online platform Stepik. Recommended external support source is stackoverflow.com.		
Facilities, Equipment and Software	Microsoft Excel 2013/2016/365 Python: Python 3 with Jupyter Notebook (installed with Anaconda)		
Recommendations for students on the organization of self-study	Self-study work can be considered as an organizational form of training - a system of pedagogical conditions that ensure the management of educational activities for the development of knowledge and skills in the field of educational activities without assistance. The student needs to clearly understand that independent work is not just a prerequisite but a necessary condition for obtaining knowledge of the discipline and the development of competencies required in future professional activities. In the educational process there are two types of independent work: <ul style="list-style-type: none"> • in-class; • out-of-class. In-class work on the discipline is carried out in the classroom under the direct supervision of the teacher and his task. Out-of-class work - the planned educational work of students, performed in out-of-class time on the instructions and methodological guidance of the teacher but without his direct participation.		

	<p>Types of tasks for out-of-class independent work, their content and nature can be variable and differentiated, take into account the individual characteristics of the student.</p> <p>Independent work can be carried out individually or by groups of students online and in the classroom, depending on the purpose, volume, specific topics of independent work, the level of complexity.</p> <p>Control of results of out-of-class independent work is carried out within the time allotted for compulsory training sessions on discipline at seminars or control classes.</p>
Special conditions for organization of learning process for students with special needs	<p>The following types of comprehension of learning information (including e-learning and distance learning) can be offered to students with disabilities (by their written request) in accordance with their individual psychophysical characteristics:</p> <ol style="list-style-type: none"> 1) <i>for persons with vision disorders</i>: a printed text in enlarged font; an electronic document; audios (transferring of learning materials into the audio); an individual advising with an assistance of a sign language interpreter; individual assignments and advising. 2) <i>for persons with hearing disorders</i>: a printed text; an electronic document; video materials with subtitles; an individual advising with an assistance of a sign language interpreter; individual assignments and advising. 3) <i>for persons with muscle-skeleton disorders</i>: a printed text; an electronic document; audios; individual assignments and advising.
Course Instructor	Andrei A. Ternikov, Lecturer, Department of Management

Programme ILO(s)	Course ILO(s)	Teaching and Learning Methods for delivering ILO(s)	Indicative Assessment Methods of Delivered ILO(s)
LO1 Understand the challenges of uncertain economic environment, assess them and take appropriate financial and investment decisions	ILO1 Choose methods adequately corresponding to the objectives of a research project	Individual problem solving Discussion of real-time issues using terms and concepts studied in class. Case studies	In-class discussions Reports
LO3 Use strong analytical skills and apply them to solve practical problems	ILO3 Collect, store, process and analyze data automatically with the use of scripting languages; develop and apply new research methods of basic machine learning algorithms and ways to collect information using data mining techniques	In-class/home assignments to solve computer exercises in data analysis software: Excel, Python.	In-class quizzes Home assignments
LO4 Examine and critically appraise research methods and tools relevant for research in finance	ILO4 Able to choose tools, modern technical means and information technologies for processing information in accordance with the assigned scientific task in the field of finance	Analysis of publications in leading academic journals	In-class discussions Reports
LO9 Demonstrate a range of generic skills including information and time management, team and project work, computing and autonomous learning, digital skills	ILO9 Students should know how to: use ICT solutions in solving real-life problems, work together with other team members, develop personal knowledge and skills.	Individual/team projects Research projects MOOCs	Individual projects Group projects
LO10 Demonstrate an innovative, open and ethical mindset	ILO10 Planning and beginning to perform a research project requires an open and innovative mindset.	Problem-solving Discussions	Reports Participation in group discussions Participation in group exercises

Course Content

- – obligatory content
- – additional (recommended) content

1. Data Analysis in MS Excel

- 1.1. Manipulating with Data in Excel (Import, Formats, VLOOKUPs)
- 1.2. Text & Financial functions + PivotTables in Excel
- 1.3. Financial Models in Excel (OLS + Forecasting)

2. Data Collection and Processing (in Python)**2.1. Introduction to Python (basic syntax)**

- Scripting languages itself and Graphical User Interface (GUI)
- Reading developers' documentation (packages, libraries, forums)
- Code iterations (loops)
- Writing function

2.2. Managing Datasets (Dictionaries, Numpy, Pandas)**2.2.1. Data Sources**

- Minalable Data examples (text, data tables, time-series, images, etc)
- *.csv-format: separators (delimiters) and encoding

2.2.2. Data Structures

- Data formats (types) in Python
- Data arrangement (matrices, lists, data frames)

2.2.3. Data Processing

- Cleaning noisy data
- Merging and reorganizing data
- Concatenating strings
- Date formats
- Regular expressions & Encoding issues

2.3. Data Visualisation

- Types of graphics
- Exploratory data analysis

2.4. Getting Data from Web

- Reading, uploading and saving data
- Code debugging
- Basic HTML syntax
- Special formats of data *.xml and *.json
- Working with Application Programming Interfaces (APIs)

3. Machine Learning Algorithms in Finance (in Python)**3.1. Supervised Learning**

- 3.1.1. Regression Algorithms
- 3.1.2. Classification Algorithms

3.2. Unsupervised Learning

Assessment Methods and Criteria

Assessment Methods

Types of Assessment	Forms of Assessment	Modules			
		1	2	3	4
Summative assessment	Lab in MS Excel	*			
	Hometask in Python	*			
	Lab in Python		*		
	Project in Python		*		

Assessment Criteria

Written Assignments (labs, hometask, project)

Grades	Assessment Criteria
«Excellent» (8-10)	Has a clear argument, which addresses the topic and responds effectively to all aspects of the task. Fully satisfies all the requirements of the task; rare minor errors occur.
«Good» (6-7)	Responds to most aspects of the topic with a clear, explicit argument. Covers the requirements of the task; may produce occasional errors.
«Satisfactory» (4-5)	Generally addresses the task; the format may be inappropriate in places; display little evidence of (depending on the assignment): independent thought and critical judgement include a partial superficial coverage of the key issues, lack critical analysis, may make frequent errors.
«Fail» (0-2)	Fails to demonstrate any appropriate knowledge.