FEDERAL STATE BUDGET EDUCATIONAL INSTITUTION OF HIGHER EDUCATION “LOMONOSOV MOSCOW STATE UNIVERSITY”

FACULTY OF ECONOMICS

«APPROVED»

Dean of the Faculty of Economics, MSU

professor _____________ A.A.Auzan

«___» ____________2024

COURSE SYLLABUS

Course title:

Machine Learning and Data Analysis – 2

Level of higher education:

MASTER STUDIES

Field of study:

38.04.01. ECONOMICS

Mode of study:

FULL-TIME

Course syllabus is considered and approved by
the Educational and Methodological Council of the Faculty of Economics
(minutes №__________, date)

Moscow 2024

The course syllabus is developed in accordance with the self-established MSU educational standard (ES MSU) for implemented main professional educational programs of higher education for Master's degree in the field of study 38.04.01. Economics

ES MSU is approved by the decision of MSU Academic Council dated December 28, 2020, minutes №7

Year (years) of enrollment: 2024 and forthcoming

1. **Place and status of the course in the structure of the Master program**

Course status: *elective (by students selection)*

Trimester: 3

2. **Course Prerequisites**

This discipline is based on the knowledge and skills acquired as a result of studying following courses:

— Probability theory
— Mathematical statistics
— Python
— Machine Learning and Data Analysis – 1

3. **Intended learning outcomes (ILO) of the course associated to the required competencies of the graduates**

Competencies of graduates (codes)	Indicators of achievement of competencies	Intended learning outcomes of the course (module) associated to the required competencies of the graduates

4. **Workload of the course by types of activity**

The workload of the discipline is 3 ECTS: 108 academic hours, including 48 academic hours of contact work with a professor, 56 academic hours of self-directed studies, 4 academic hours of exam.

5. **Learning format:** full-time, with the use of educational platform On.Econ (use of distant learning technologies is allowed if necessary)

6. **Content of the course structured by topics (sections) indicating the number of academic hours allocated to them and types of training**

Title and brief content of sections	**Total**, hours	**Including**			
		Contact work (work in contact with a professor)			**Student self-directed studies,** hours
		Lections, hours	Seminars, hours	Group consultations, hours	
1. Introduction to deep learning	16	2	2	4	8
2: Architectures	18	2	2	4	10
3.Computer vision	16	2	2	4	8
4. Audio Processing and Speech Technologies	18	2	2	4	10
5. Natural Language Processing	18	2	2	4	10
Тема 6. Graph Neural Network	18	2	2	4	10
Midterm assessment:	4	4			
Total	**108**	**16**	**12**	**24**	**56**

Brief content of the course topics

Topic 1. Introduction to deep learning

The first lesson covers the history of DL and its current trends. It gives an introduction to the basics of DL (forward/backward passes, gradient descent and its types, single and multilayer neural network architectures, basic layers: activation, batch normalization etc) and to the PyTorch framework.
The practical part of the lesson explains how to build a simple neural network from scratch and compares the hand-made network with PyTorch's proper implementation.

Topic 2 Architectures
The second lesson covers more advanced architectures. It briefly touches upon computer vision problems and gives an understanding of Convolutional Layers. In addition to that the NLP problems are also being discussed and Recurrent Layers are explained as well. Also the lesson contains information about useful practical DL tools, such as MLFlow, TensorBoard and PyTorchLightning.
The homework asks students to build more layers by hand and add them to their mini neural network library

Topic 3 Computer vision
This lecture introduces tools for the main tasks of computer vision: classification, object detection and image segmentation. Students will gain a deep understanding of convolutional neural networks. Classic architectures such as AlexNet, VGGNet, ResNet will be covered. RNNs and Transformers for CV tasks are also covered in the lectures. Building an end-to-end object detection system is proposed as the main practice for this part of the course.

Topic 4 Audio Processing and Speech Technologies
This lec provides an in-depth exploration of audio processing techniques and applications, with a particular focus on speech technologies such as ASR, TTS, and KWS. Through theoretical concepts and hands-on projects, students will learn how to preprocess audio, create and interpret spectrograms, and build robust audio models using deep learning techniques. The lec places a strong emphasis on ASR applications in news, but also covers a wide range of audio processing tasks relevant to modern speech technology.

Topic 5 Natural Language Processing (NLP)
This NLP lec provides a structured path from foundational concepts to advanced applications. Starting with basic text classification techniques, the lec progresses through essential NLP models, representation techniques, language modeling, modern architectures like Transformers, and culminates in working with large language models (LLMs). By the end of the lec, students will gain a deep understanding of NLP's theoretical foundations and practical skills, making them proficient in developing and fine-tuning state-of-the-art NLP systems.

Topic 6 Graphs
The lecture on GNN introduces one of the modern topics of DL application to complex data structures. Students will get acquainted with graph data structures and classical approaches for graph visualization and analysis. The lecture covers classical ML tasks such as regression, classification and clustering at different levels of graph analysis: node-level, edge-level and graph-level tasks. Additionally, GNN applications for CV and NLP tasks will be considered.

7. Assessment tools to assess the course learning outcomes

7.1. Sample assessment tools:

Learning outcomes of the course	Types of assessment tools
М.УК-1.Ум.1	Home assignment, project assignment, exam
М.ОПК-5.Ум.1	Home assignment, project assignment, exam
М.ПК-3.Зн.1	Home assignment, project assignment, exam
М.ПК-3.Ум.1	Home assignment, project assignment, exam
М.ПК-4.Ум.1	Home assignment, project assignment, exam
М.ПК-4.Ум.2	Home assignment, project assignment, exam
М.ПК-9.Ум.1	Home assignment, project assignment, exam
М.ПК-9.Ум.2	Home assignment, project assignment, exam
М.УК-1.Ум.1	Home assignment, project assignment, exam
М.ОПК-5.Ум.1	Home assignment, project assignment, exam
М.ПК-3.Зн.1	Home assignment, project assignment, exam
М.СПК-4.Ум.1	Home assignment, project assignment, exam
М.СПК-4.Ум.2	Home assignment, project assignment, exam

7.2. Course assessment criteria (scores):

Types of assessment tools	Score
Home assignment	70
Project assignment	50
Exam	30
Total	**150**

7.3. Grade for the course is determined based on the following criteria:

Grade	Minimum score	Maximum score

Excellent	127,5	150,0
Good	97,5	127,0
Satisfactory	60,0	97,0
Failed	0,0	59,5

Note: in case a student's score obtained during the trimester is less than 20% of the maximum score of the discipline, the following rule of passing the course should be applied at the midterm assessment (and further re-examination): 'a student can obtain only a satisfactory mark and only in case she/he receives for the midterm assessment, including all the course material, no less than 85% of the score allocated to this assessment'.

7.4. Typical tasks and other materials necessary to assess the learning outcomes:

Example of home assignment
The homework is to expand upon the practical part and upgrade the network, adding bias, new activation layer and to make further comparison with PyTorch's implementation.

Example of project assignment.
Put into practice one (or several) machine learning models discussed in the course. Training data should be taken from open sources. The task is performed in groups or individually.
Submit this work in the Jupiter Notebook.

7.5. Methodological guidelines and assignment requirements:

Home assignments *are practical tasks aimed at consolidating the skills of machine learning and data analysis in Python within various topics. The task is submitted in the Jupyter-notebook .ipynb format with the code, comments and answers to the questions of the task. The comments should describe the methods and data used to solve the problem, justify the choice of the algorithm in detail, and present the results of calculations (using tables and figures if necessary). Homework allows you to score up to 200 points.*

Project assignments *are practical tasks that require the student to apply the acquired machine learning skills to solve problems. The task is given in the format of an oral presentation, as well as the source code in Python in a convenient format (for example, Jupyter laptop, code on the Github repository, zip archive). The project can be presented both individually and in a team. During the presentaion of the project, it is required to briefly describe the problem being solved, describe the methods of machine learning and data analysis used, formulate and explain the results obtained, answer the questions of the host. Project work allows you to score up to 100 points.*

8. Resources

8.1. List of main and additional literature

Main literature:

- Christopher M. Bishop. Pattern Recognition and MachineLearning
- Kevin P. Murphy. Machine Learning: A Probablistic Perspective
- Ian Goodfellow, Yoshua Bengio, Aaron Courville. Deep Learning
- Richard S. Sutton, Andrew G. Barto. Reinforcement Learning: An Introduction
- Aurélien Géron. Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow

8.2. List of licensed software

- Anaconda (Python distribution)
- Data-science packages for Python programming language

8.3. List of professional databases and information referral systems

- Yandex School of data analysis.
- MADE Academy.
- Coursera.

8.4. List of Internet resources (if necessary)

- course CS229 Stanford.
- course CS231N Stanford.
- course CS234 Stanford.
- Towards Data Science.
- Kaggle platform.
- А.Г. Дьяконова blog.

- course Open Data Science.

8.5. Description of material and technical support

To organize classes in the discipline the following technical training tools are needed: a computer class with a projector and a blackboard.

9. **Language of instruction:** *English. Additional educational materials, including literature sources, can be in English*

10. **Professor (professors):** S. Ilishaev (Илишаев Семён Исрольевич), Head of Department, Risks Unit, Sberbank, V. Kirpa (Кирпа Вадим Дмитриевич), Head of Data Research at the Risks Unit of Sberbank, M. Nikonov (Никонов Максим Викторович), Head of Analytics and Big Data, VK

11. **Syllabus authors:** S. Ilishaev (Илишаев Семён Исрольевич), Head of Department, Risks Unit, Sberbank, V. Kirpa (Кирпа Вадим Дмитриевич), Head of Data Research at the Risks Unit of Sberbank, M. Nikonov (Никонов Максим Викторович), Head of Analytics and Big Data, VK