

**ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ ОБРАЗОВАТЕЛЬНОЕ
УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ
«МОСКОВСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ ИМЕНИ
М.В.ЛОМОНОСОВА»**

ЭКОНОМИЧЕСКИЙ ФАКУЛЬТЕТ

«УТВЕРЖДАЮ»

Декан экономического факультета МГУ
профессор. _____ А.А.Аузан

«» 2020 год

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ

Машинное обучение и анализ данных-2 (на англ.языке)

Уровень высшего образования
Магистратура

Направление подготовки (специальность)
38.04.01 Экономика

Направленность (профиль) ОПОП
Анализ данных в экономике

Форма обучения
очная

Рабочая программа рассмотрена и одобрена
Учебно-методической комиссией экономического факультета
(протокол № _____, дата)

Москва 2024

На оборотной стороне титульного листа указывается:

Рабочая программа дисциплины разработана в соответствии с самостоятельно установленным МГУ образовательным стандартом (ОС МГУ) для реализуемых основных профессиональных образовательных программ высшего образования по направлению подготовки «_____» магистратуры

ОС МГУ утвержден решением Ученого совета МГУ имени М.В.Ломоносова от _____ 20____ года (протокол № ____).

Год (годы) приема на обучение: 2024 и последующие

1 Место и статус дисциплины в структуре основной профессиональной образовательной программы подготовки магистра

Статус дисциплины: *вариативная (по выбору)*

Триместр: 3

2 Входные требования для освоения дисциплины

Для успешного освоения данного курса требуются знания и навыки, полученные в следующих дисциплинах:

- основы программирования python;
- Машинное обучение и анализ данных-1

3 Планируемые результаты обучения по дисциплине, соотнесенные с требуемыми компетенциями выпускников

Формируемые компетенции	Планируемые результаты обучения по дисциплине, соотнесенные с требуемыми компетенциями
Способность формулировать научно обоснованные гипотезы, создавать теоретические модели явлений и процессов, применять методологию научного познания в профессиональной деятельности (М.УК-1)	УМЕТЬ выдвигать научно обоснованные гипотезы, поддающиеся операционализации, моделировать явления и процессы на основе системного видения различных отраслей знаний М.УК-1.Ум.1
Способность применять продвинутые инструментальные методы экономического анализа в прикладных и/или фундаментальных исследованиях (М.ОПК-5)	УМЕТЬ обрабатывать информацию при помощи методов машинного обучения М.ОПК-5.Ум.1
Способность проводить самостоятельные исследования в соответствии с разработанной программой (М.ПК-3)	ЗНАТЬ современные научные методы машинного обучения и анализа данных М.ПК-3.Зн.1
	УМЕТЬ применять современные научные методы машинного обучения и анализа данных в экономических исследованиях М.ПК-3.Ум.1
Способность представлять результаты проведенного исследования научному сообществу в виде статьи или доклада (М.ПК-4)	УМЕТЬ представлять результаты научного исследования в систематизированном виде в письменной форме М.ПК-4.Ум.1
	УМЕТЬ создавать презентации по итогам исследований и делать устные научные доклады М.ПК-4.Ум.2

Способность анализировать и использовать различные источники информации для проведения экономических расчетов (М.ПК-9)	УМЕТЬ оценивать качество источников экономической информации М.ПК-9.Ум.1
	УМЕТЬ применять качественные и количественные методы для проведения прикладных экономических исследований М.ПК-9.Ум.2
Способность разрабатывать эконометрические модели и модели машинного обучения исследуемых экономических процессов и явлений, интерпретировать полученные результаты (М.СПК-1)	ЗНАТЬ современные инструментальные методы, применяемые в экономических исследованиях М.СПК-1.Зн.1
	УМЕТЬ применять современные инструментальные методы к релевантным данным для решения заданного или самостоятельно сформулированного исследовательского вопроса М.СПК-1.Ум.1
Способность видеть логические связи в системе собранной, обработанной и проанализированной информации, и на основании этого разрабатывать рекомендации для лиц, принимающих решения на микро- и макроуровне, или бизнес-решения (М-СПК-4)	УМЕТЬ при помощи методов анализа данных сравнивать альтернативные решения и находить оптимальные по заданным метрикам качества М.СПК-4.Ум.1
	УМЕТЬ на основе сделанных выводов об оптимальности решения разрабатывать рекомендации для лиц, принимающих решения М.СПК-4.Ум.2

4 Объем дисциплины по видам занятий

Объем дисциплины составляет 6 зачетных единицы: 108 академических часов, из которых 56 академических часов составляет аудиторная нагрузка, из них 24 академических часа — лекции и 24 академических часа — семинары, 48 академических часов — групповая контактная работа, 0 академических часов — индивидуальная контактная работа, 4 часа составляет аттестация (экзамен), 56 академических часов составляет самостоятельная работа магистранта.

5 Формат обучения: используется электронная информационная среды экономического факультета МГУ имени М.В.Ломоносова «ON.ECON».

6 Содержание дисциплины, структурированное по темам (разделам) с указанием отведенного на них количества академических часов и виды учебных занятий

Название раздела/темы	Всего, часы	В том числе			
		Контактная работа с преподавателем			Самостоятельная работа магистранта, часы
		Лекции, часы	Семинары, часы	Групповая, часы	
Тема 1. Введение в глубокое обучение	16	2	2	4	8
Тема 2: Архитектуры	18	2	2	4	10
Тема 3. Компьютерное зрение	16	2	2	4	8
Тема 4. Технологии обработки звука и речи	18	2	2	4	10
Тема 5. Обработка естественных языков	18	2	2	4	10
Тема 6. Графовые нейронные сети	18	2	2	4	10
Экзамен	4	4			
Всего	108	16	12	24	56

Краткое содержание тем дисциплины

Тема 1. Введение в глубокое обучение

История глубокого обучения и его современным тенденциям. Он знакомит с основами глубокого обучения (прямой/обратный проход, градиентный спуск и его типы, однослойные и многослойные архитектуры нейронных сетей, базовые слои: активация, пакетная нормализация и т.д.), а также с фреймворком PyTorch. В практической части урока объясняется, как создать простую нейронную сеть с нуля, и сравнивается созданная вручную сеть с её реализацией в PyTorch.

Тема 2. Архитектуры

Вторая тема посвящена более сложным архитектурам. Он кратко затрагивает проблемы компьютерного зрения и дает представление о сверточных слоях. Кроме того, обсуждаются задачи обработки естественного языка и объясняются рекуррентные слои. Урок также содержит информацию о полезных практических инструментах глубокого обучения, таких как MLFlow, TensorBoard и PyTorchLightning. Домашнее задание: предлагается вручную создать больше слоев и добавить их в свою мини-библиотеку нейронных сетей.

Тема 3. Компьютерное зрение

Инструменты для решения основных задач компьютерного зрения: классификации, обнаружения объектов и сегментации изображений. Студенты получают глубокое понимание сверточных нейронных сетей. Будут рассмотрены классические архитектуры, такие как AlexNet, VGGNet, ResNet. Также на занятии будут рассмотрены рекуррентные нейронные сети (RNN) и преобразователи для задач

компьютерного зрения (CV). В качестве основной практики в этой части курса предлагается создание сквозной системы обнаружения объектов.

Тема 4. Технологии обработки звука и речи

Углубленное изучение методов и приложений обработки звука, уделяя особое внимание таким речевым технологиям, как ASR, TTS и KWS. Благодаря теоретическим концепциям и практическим проектам студенты научатся выполнять предварительную обработку звука, создавать и интерпретировать спектрограммы, а также создавать надежные аудиомодели с использованием методов глубокого обучения. Курс уделяет особое внимание применению ASR в новостях, но также охватывает широкий спектр задач обработки звука, актуальных для современных речевых технологий.

Тема 5. Обработка естественных языков (NLP)

Структурированный путь от фундаментальных концепций к продвинутым приложениям. Начиная с базовых методов классификации текста, переход к основным моделям обработки естественного языка (NLP), методам представления, языковому моделированию и современным архитектурам, таким как трансформеры, и завершается работой с большими языковыми моделями (LLM). К концу курса студенты получают глубокое понимание теоретических основ обработки естественного языка (NLP) и приобретут практические навыки, что позволит им эффективно разрабатывать и настраивать современные системы обработки естественного языка (NLP).

Тема 6. Графовые нейронные сети

GNN знакомит с одной из современных тем применения глубокого машинного обучения к сложным структурам данных. Студенты познакомятся с графовыми структурами данных и классическими подходами к визуализации и анализу графов. Лекция охватывает классические задачи машинного обучения, такие как регрессия, классификация и кластеризация, на разных уровнях анализа графов: на уровне узлов, ребер и графов. Кроме того, будут рассмотрены применения GNN для задач компьютерного зрения и обработки естественного языка.

Фонд оценочных средств для оценивания результатов обучения по дисциплине

Шкала оценивания результатов (баллы) по дисциплине:

Результаты обучения по дисциплине	Виды оценочных средств
УМЕТЬ выдвигать научно обоснованные гипотезы, поддающиеся операционализации, моделировать явления и процессы на основе системного видения различных отраслей знаний М.УК-1.Ум.1	Домашние работы Выполнение проектных заданий. Экзамен
УМЕТЬ обрабатывать информацию при помощи методов машинного обучения М.ОПК-5.Ум.1	Домашние работы Выполнение проектных заданий
ЗНАТЬ современные научные методы машинного обучения и анализа данных М.ПК-3.Зн.1	Домашние работы Выполнение проектных заданий
УМЕТЬ применять современные научные методы машинного обучения и анализа данных в экономических исследованиях М.ПК-3.Ум.1	Домашние работы Выполнение проектных заданий

УМЕТЬ представлять результаты научного исследования в систематизированном виде в письменной форме М.ПК-4.Ум.1	Домашние работы Выполнение проектных заданий
УМЕТЬ создавать презентации по итогам исследований и делать устные научные доклады М.ПК-4.Ум.2	Выполнение проектных заданий
УМЕТЬ оценивать качество источников экономической информации М.ПК-9.Ум.1	Домашние работы Выполнение проектных заданий
УМЕТЬ применять качественные и количественные методы для проведения прикладных экономических исследований М.ПК-9.Ум.2	Домашние работы Выполнение проектных заданий
УМЕТЬ выдвигать научно обоснованные гипотезы, поддающиеся операционализации, моделировать явления и процессы на основе системного видения различных отраслей знаний М.УК-1.Ум.1	Домашние работы Выполнение проектных заданий
УМЕТЬ обрабатывать информацию при помощи методов машинного обучения М.ОПК-5.Ум.1	Домашние работы Выполнение проектных заданий
ЗНАТЬ современные научные методы машинного обучения и анализа данных М.ПК-3.Зн.1	Домашние работы Выполнение проектных заданий. Экзамен
УМЕТЬ при помощи методов анализа данных сравнивать альтернативные решения и находить оптимальные по заданным метрикам качества М.СПК-4.Ум.1	Домашние работы Выполнение проектных заданий
УМЕТЬ на основе сделанных выводов об оптимальности решения разрабатывать рекомендации для лиц, принимающих решения М.СПК-4.Ум.2	Домашние работы Выполнение проектных заданий. Экзамен

Виды оценочных средств	Баллы
Домашние работы	70
Проектные задания	50

Экзамен	30
Всего	150

Оценка по дисциплине выставляется, исходя из следующих критериев:

Оценка	Минимальное количество баллов	Максимальное количество баллов
<i>Отлично</i>	127,5	150,0
<i>Хорошо</i>	97,5	127,0
<i>Удовлетворительно</i>	60,0	97,0
<i>Неудовлетворительно</i>	0,0	59,5

Примечание: в случае, если магистрант за триместр набирает менее 20% баллов от максимального количества по дисциплине, то уже на промежуточном контроле (и далее на пересдачах) действует следующее правило сдачи: «магистрант может получить только оценку «Удовлетворительно», и только если получит за промежуточный контроль, включающий весь материал дисциплины, не менее, чем 85% от баллов за промежуточный контроль».

Типовые задания, методические рекомендации по их подготовке и требования к их выполнению:

Домашние работы представляют собой практические задания, ориентированные на закрепление навыков машинного обучения и анализа данных на языке Python в рамках различных тем. Задание сдается в формате Jupyter-ноутбука .ipynb с кодом, комментариями и ответами на вопросы задания. В комментариях следует описать используемые для решения задачи методы и данные, подробно обосновать выбор алгоритма, представить результаты расчётов (при необходимости используя таблицы и рисунки).

Проектные работы представляют собой практические задания, требующие от студента применения полученных навыков машинного обучения для решения проблем. Задание сдается в формате устной презентации, а также исходного кода на языке Python в удобном формате (например, Jupyter-ноутбук, код на репозитории Github, zip-архив). Проект может быть представлен как индивидуально, так и в команде. В ходе защиты проекта требуется кратко описать решаемую задачу, описать применяемые методы машинного обучения и анализа данных, сформулировать и объяснить полученные результаты, ответить на вопросы принимающих.

Ниже приведены примеры домашних заданий, которые будут задаваться студентам в процессе обучения.

Пример домашнего задания расширить практическую часть и усовершенствовать сеть, добавив смещение, новый активационный слой и проведя дальнейшее сравнение с реализацией в PyTorch.

Пример проектного задания. Применить на практике одну (или несколько) моделей машинного обучения, рассмотренных в курсе. Данные для обучения следует брать из открытых источников. Задание выполняется в группах или индивидуально.

Результат оформить в среде Jupyter Notebook.

8 Ресурсное обеспечение

8.1 Перечень основной и дополнительной литературы

Основная литература:

- Christopher M. Bishop. Pattern Recognition and Machine Learning
- Kevin P. Murphy. Machine Learning: A Probabilistic 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 Перечень лицензионного программного обеспечения

- Среда разработки Anaconda для языка Python.
- Библиотеки анализа данных для языка Python.

8.3 Перечень профессиональных баз данных и информационных справочных систем

- Школа Анализа Данных Яндекса.
- Академия MADE.
- Coursera.

8.4 Перечень ресурсов информационно-телекоммуникационной сети «Интернет»

- Курс CS229 Stanford_
- Курс CS231N Stanford_
- Курс CS234 Stanford_
- Towards Data Science_
- Платформа Kaggle_
- Блог А.Г. Дьяконова_
- Курс Open Data Science_

8.5. Описание материально-технической базы

Для организации занятий по дисциплине необходимы следующие технические средства обучения: компьютерный класс с проектором и доской.

1 Язык преподавания: английский

- 2 Преподаватель (преподаватели):** Илишаев Семён Исрольевич, руководитель направления, Блок Риски, ПАО Сбербанк (внешний преподаватель), Кирпа Вадим Дмитриевич, руководитель направления по исследованию данных в Блоке Риски ПАО Сбербанк (внешний преподаватель), Никонов Максим Викторович, руководитель аналитики и больших данных, ВК (внешний преподаватель)

- 3 Автор (авторы) программы:** Илишаев Семён Исрольевич, руководитель направления, Блок Риски, ПАО Сбербанк (внешний преподаватель), Кирпа Вадим Дмитриевич, руководитель направления по исследованию данных в Блоке Риски ПАО Сбербанк (внешний преподаватель), Никонов Максим Викторович, руководитель аналитики и больших данных, ВК (внешний преподаватель)