

Федеральное государственное бюджетное образовательное учреждение высшего профессионального образования

Московский государственный университет имени М.В. Ломоносова

Механико-математический факультет

Кафедра инженерной механики и прикладной математики

УТВЕРЖДАЮ
Заведующий кафедрой

_____/Ганиев Р.Ф./

«__» _____ 20__ г.

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ (МОДУЛЯ)

Наименование дисциплины (модуля):

Анализ данных в Python и введение в машинное обучение

наименование дисциплины (модуля)

Уровень высшего образования:

Специалитет, подготовка кадров в аспирантуре

Направление подготовки (специальность):

Фундаментальные математика и механика

(код и название направления/специальности)

Направленность (профиль) ОПОП:

Механика жидкости, газа и плазмы

(если дисциплина (модуль) относится к вариативной части программы)

Форма обучения:

очная

(очная, очно-заочная)

Рабочая программа рассмотрена и одобрена
на заседании кафедры инженерной механики и прикладной математики

(протокол № _____, «__» _____ 20__ года)

Москва 2022

Рабочая программа дисциплины (модуля) разработана в соответствии с самостоятельно установленным МГУ образовательным стандартом (ОС МГУ) для реализуемых основных профессиональных образовательных программ высшего образования по направлению подготовки / специальности «Фундаментальные математика и механика» (программы бакалавриата, магистратуры, реализуемых последовательно по схеме интегрированной подготовки; программы специалитета; программы магистратуры) в редакции приказа МГУ от 30 декабря 2016 г.

Год (годы) приема на обучение _____

1. Место дисциплины (модуля) в структуре ОПОП ВО: *относится к вариативной части ОПОП ВО.*

2. Входные требования для освоения дисциплины (модуля), предварительные условия (если есть):

Знание основ теории вероятности/статистики, линейной алгебры, математического анализа и программирования.

3. Результаты обучения по дисциплине (модулю), соотнесенные с требуемыми компетенциями выпускников.

Компетенции выпускников (коды)	Планируемые результаты обучения по дисциплине (модулю), соотнесенные с компетенциями
ОПК-1 Готовность использовать фундаментальные знания в области механики сплошной среды в будущей профессиональной деятельности	Знать: <i>Способы использования на практике математических алгоритмов, в том числе с применением современных программных систем</i> Уметь: <i>Решать стандартные задачи профессиональной деятельности на основе информационной и библиографической культуры с применением информационно-коммуникационных технологий</i>
ПК-2 способность к самостоятельному анализу физических аспектов в классических постановках задач механики	Знать: <i>Знать python для его применения в задачах анализа данных и для реализации основных методов машинного обучения</i> Уметь <i>Правильно выбирать метод решения заданной задачи анализа данных</i>

4. Формат обучения **очный**

5. Объем дисциплины (модуля) составляет 72 академических часов, в том числе 33 академических часа, отведенных на контактную работу обучающихся с преподавателем, 33 академических часа на самостоятельную работу обучающихся, 2 академических часа на коллоквиум, 4 академических часов на экзамен и защиту результатов самостоятельной работы.

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

Наименование и краткое содержание разделов и тем дисциплины (модуля), Форма промежуточной аттестации по дисциплине (модулю)	Всего (часы)	В том числе			
		Контактная работа (работа во взаимодействии с преподавателем) Виды контактной работы, часы			Самостоятельная работа обучающегося, часы (виды самостоятельной работы – эссе, реферат, контрольная работа и пр. – указываются при необходимости)
		Занятия лекционного типа*	Занятия семинарского типа*	Всего	
Тема 1. Основы Python Установка и настройка окружения, импорт библиотек. Знакомство с основными структурами данных: список, множество, кортеж, словарь, базовые методы работы с ними. Основы работы со строками. Библиотеки NumPy и SciPy для работы с векторами и матрицами.	8	4	0	4	4
Тема 2. Работа с табличными данными Преобразование табличных данных, селекция таблиц, библиотека Pandas, объекты Pandas.Series и Pandas.DataFrame. Анализ и исследования данных одной таблицы (работа с единым датафреймом).	12	6		6	6

Работа с несколькими таблицами, операции merge, join, различные типы join'ов. Основы sql-синтаксиса, библиотека pandasql.					
Тема 3. Визуализация данных Исследовательский анализ данных (EDA) и визуализация при помощи Pandas. Работа с библиотеками для визуализации данных matplotlib и seaborn. Интерактивная визуализация с помощью plotly.	14	7	0	7	7
Коллоквиум по темам 1–3	2				2
Тема 4. Элементы статистики, библиотека SciPy Основные вероятностные распределения. Нормальное распределение. Центральная предельная теорема и визуализация ее утверждения методами Python. Доверительные интервалы. Сравнение двух средних, t-критерий Стьюдента. Проверка распределения на нормальность.	8	4		4	4
Тема 5. Градиентные методы Методы градиентного спуска и стохастического градиентного спуска в задачах машинного обучения. Реализация градиентного спуска для задачи бинарной классификации. Визуализация метода для двумерного случая.	6	3	0	3	3
Тема 6. Методы обучения с учителем Задачи линейной, логистической регрессии, базовые модели sklearn. Функции потерь и метрики качества. Решающие деревья. Случайный лес. Градиентный бустинг. Современные реализации градиентного бустинга: XGBoost, LightGBM, Catboost.	10	5	0	5	5

Тема 7. Методы обучения без учителя Методы кластеризации: метод К-средних, иерархическая кластеризация, DBSCAN. Метрики качества кластеризации. Методы понижения размерности: Метод Главных Компонент (Principal Component Analysis), tSNE, UMAP	8	4	0	4	4
Промежуточная аттестация: экзамен и защита результатов самостоятельной работы.	4				4
Итого	72				36

**Внимание! В таблице должно быть зафиксировано проведение текущего контроля успеваемости, который может быть реализован, например, в рамках занятий семинарского типа.*

*** Часы, отводимые на проведение промежуточной аттестации, выделяются из часов самостоятельной работы обучающегося*

7. Фонд оценочных средств (ФОС) для оценивания результатов обучения по дисциплине (модулю)

7.1. Типовые контрольные задания или иные материалы для проведения текущего контроля успеваемости.

1. Коллоквиум по темам 1–3:

1.1 Опишите основные структуры данных в Python и операции с ними

1.2 Применение пакетов pandas и plotly на примере разбора задачи анализа данных (данные соревнований платформы Kaggle)

7.2. Типовые контрольные задания или иные материалы для проведения промежуточной аттестации.

1. Список билетов к экзамену:

- 1) Основные структуры данных и операции с ними: список, множество, словарь. Работа с векторами и матрицами в библиотеке NumPy.
- 2) Работа с табличными данными в при помощи pandas. Вычисление основных статистик, преобразование и группировка данных, визуализация в библиотеке pandas.
- 3) Работа с несколькими таблицами, операции merge, join, различные типы join'ов.
- 4) Визуализация данных с помощью matplotlib, seaborn. Интерактивная визуализация с plotly.
- 5) Основные вероятностные распределения. Нормальное распределение. Центральная предельная теорема и визуализация ее утверждения методами Python. Доверительные интервалы. Сравнение двух средних, t-критерий Стьюдента.
- 6) Градиентные методы в задачах машинного обучения. Реализация градиентного спуска.
- 7) Линейная и логистическая регрессии, функции потерь, метрики качества.
- 8) Решающие деревья. Случайный лес.
- 9) Градиентный бустинг. Современные реализации.
- 10) Задача кластеризации: метод K-средних, иерархическая кластеризация, DBSCAN. Метрики качества кластеризации
- 11) Методы понижения размерности: Метод Главных Компонент (Principal Component Analysis), tSNE, UMAP

2. Защита результатов самостоятельной работы.

ШКАЛА И КРИТЕРИИ ОЦЕНИВАНИЯ результатов обучения (РО) по дисциплине (модулю)				
Оценка	2	3	4	5

РО и соответствующие виды оценочных средств				
Знания <i>(виды оценочных средств: устные и письменные опросы и контрольные работы, тесты, и т.п.)</i>	Отсутствие знаний	Фрагментарные знания	Общие, но не структурированные знания	Сформированные систематические знания
Умения <i>(виды оценочных средств: практические контрольные задания, написание и защита рефератов на заданную тему и т.п.)</i>	Отсутствие умений	В целом успешное, но не систематическое умение	В целом успешное, но содержащее отдельные пробелы умение (допускает неточности непринципиального характера)	Успешное и систематическое умение
Навыки (владения, опыт деятельности) <i>(виды оценочных средств: выполнение и защита курсовой работы, отчет по практике, отчет по НИР и т.п.)</i>	Отсутствие навыков (владений, опыта)	Наличие отдельных навыков (наличие фрагментарного опыта)	В целом, сформированные навыки (владения), но используемые не в активной форме	Сформированные навыки (владения), применяемые при решении задач

8. Ресурсное обеспечение:

- Перечень основной и дополнительной литературы:
 - Конспект лекций
 - Материалы курса с примерами реализации методов
- Перечень лицензионного программного обеспечения (при необходимости):
Не требуется.

- Перечень профессиональных баз данных и информационных справочных систем:
Не требуется.
- Перечень ресурсов информационно-телекоммуникационной сети «Интернет» (при необходимости):
Не требуется.
- Описание материально-технического обеспечения:
 - Аудитория
 - Доска
 - Мел
 - Ноутбук или ПК
 - Проекционное оборудование (экран, проектор)

9. Язык преподавания:

Русский

10. Преподаватель (преподаватели):

К.ф.-м.н. Русинова Анна Михайловна

11. Автор (авторы) программы:

К.ф.-м.н. Русинова Анна Михайловна