

МИНИСТЕРСТВО ТРАНСПОРТА РОССИЙСКОЙ ФЕДЕРАЦИИ
ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ ОБРАЗОВАТЕЛЬНОЕ
УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ
«РОССИЙСКИЙ УНИВЕРСИТЕТ ТРАНСПОРТА»
(РУТ (МИИТ))



Рабочая программа дисциплины (модуля),
как компонент образовательной программы
высшего образования - программы бакалавриата
по направлению подготовки
09.03.01 Информатика и вычислительная техника,
утвержденной первым проректором РУТ (МИИТ)
Тимониным В.С.

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ (МОДУЛЯ)

Библиотеки Python для анализа данных

Направление подготовки: 09.03.01 Информатика и вычислительная техника

Направленность (профиль): IT-сервисы и технологии обработки данных на транспорте (российско-китайская программа)

Форма обучения: Очная

Рабочая программа дисциплины (модуля) в виде
электронного документа выгружена из единой
корпоративной информационной системы управления
университетом и соответствует оригиналу

Простая электронная подпись, выданная РУТ (МИИТ)
ID подписи: 937226
Подписал: руководитель образовательной программы
Проневич Ольга Борисовна
Дата: 26.05.2025

1. Общие сведения о дисциплине (модуле).

Целями освоения факультативной учебной дисциплины являются формирование у обучающихся профессиональных знаний и навыков в области использования библиотек Python для анализа данных.

Задачами дисциплины являются:

- Освоение основных концепций и методов анализа данных, включая обработку, визуализацию и моделирование данных.
- Развитие навыков работы с популярными библиотеками Python для анализа данных, такими как Pandas, NumPy, Matplotlib, Seaborn, Scikit-learn и другие.
- Обучение применению методов машинного обучения для решения задач классификации, регрессии и кластеризации данных.
- Приобретение опыта работы с большими данными и оптимизации вычислительных процессов.
- Освоение основ построения интерактивных отчетов и дашбордов для представления результатов анализа.
- Развитие умений интерпретации полученных результатов и формирования рекомендаций на основе данных.
- Изучение принципов работы с внешними источниками данных (API, базы данных) и их интеграция в аналитические процессы.

2. Планируемые результаты обучения по дисциплине (модулю).

Перечень формируемых результатов освоения образовательной программы (компетенций) в результате обучения по дисциплине (модулю):

ОПК-8 - Способен разрабатывать алгоритмы и программы, пригодные для практического применения;

ПК-2 - Способен осуществлять концептуальное, функциональное и логическое проектирование систем среднего и крупного масштаба и сложности .

Обучение по дисциплине (модулю) предполагает, что по его результатам обучающийся будет:

Знать:

- различные методы чтения, обработки и анализа данных с использованием Python, включая фильтрацию, сортировку, группировку и агрегацию данных

- основные принципы библиотек для работы с данными, основанными на Python, и уметь выбирать подходящий инструмент для конкретной задачи

Уметь:

- проводить визуальный анализ данных с помощью построения графиков и диаграмм, что позволит лучше понять структуру и закономерности данных

- разрабатывать и применять простые модели машинного обучения для решения задач классификации, регрессии и кластеризации

Владеть:

- навыками разделения данных на обучающую и тестовую выборки, оценки качества моделей, настройки параметров и интерпретации результатов

- навыками применения методов визуализации данных для выявления скрытых закономерностей и тенденций в данных.

3. Объем дисциплины (модуля).

3.1. Общая трудоемкость дисциплины (модуля).

Общая трудоемкость дисциплины (модуля) составляет 2 з.е. (72 академических часа(ов)).

3.2. Объем дисциплины (модуля) в форме контактной работы обучающихся с педагогическими работниками и (или) лицами, привлекаемыми к реализации образовательной программы на иных условиях, при проведении учебных занятий:

Тип учебных занятий	Количество часов	
	Всего	Семестр №2
Контактная работа при проведении учебных занятий (всего):	8	8
В том числе:		
Занятия семинарского типа	8	8

3.3. Объем дисциплины (модуля) в форме самостоятельной работы обучающихся, а также в форме контактной работы обучающихся с педагогическими работниками и (или) лицами, привлекаемыми к реализации образовательной программы на иных условиях, при проведении промежуточной аттестации составляет 64 академических часа (ов).

3.4. При обучении по индивидуальному учебному плану, в том числе при ускоренном обучении, объем дисциплины (модуля) может быть реализован

полностью в форме самостоятельной работы обучающихся, а также в форме контактной работы обучающихся с педагогическими работниками и (или) лицами, привлекаемыми к реализации образовательной программы на иных условиях, при проведении промежуточной аттестации.

4. Содержание дисциплины (модуля).

4.1. Занятия лекционного типа.

Не предусмотрено учебным планом

4.2. Занятия семинарского типа.

Практические занятия

№ п/п	Тематика практических занятий/краткое содержание
1	Тема 1. NumPy для работы с многомерными массивами Рассматриваемые вопросы: - Создание массивов (np.array, np.zeros, np.ones). - Индексация, срезы и изменение формы массивов. - Векторизованные операции: преимущества перед циклами. - Базовые математические функции (sum, mean, std). - Работа с NaN-значениями: обнаружение и обработка.
2	Тема 2. Pandas для работы с табличными данными Рассматриваемые вопросы: - Загрузка данных из CSV, Excel и словарей в DataFrame. - Фильтрация строк и столбцов, сортировка по условиям. - Группировка данных (groupby) и агрегация (agg). - Обработка пропусков: удаление, заполнение, интерполяция. - Объединение таблиц: merge, concat, join.
3	Тема 3. Matplotlib и Seaborn для визуализации данных Рассматриваемые вопросы: - Построение графиков: линейные, столбчатые, круговые. - Настройка осей, заголовков, легенд и цветов. - Визуализация распределений: гистограммы, boxplot, kde-plot. - Тепловые карты (heatmap) и парные графики (pairplot). - Экспорт графиков в PNG, PDF и интерактивные форматы.
4	Тема 4. Scikit-learn для предварительной обработки и ML Рассматриваемые вопросы: - Нормализация данных (StandardScaler, MinMaxScaler). - Кодирование категориальных признаков (OneHotEncoder). - Разделение данных на обучающую и тестовую выборки. - Обучение модели линейной регрессии и оценка качества. - Визуализация результатов: матрица ошибок, ROC-кривая

4.3. Самостоятельная работа обучающихся.

№ п/п	Вид самостоятельной работы
1	Изучение литературы
2	Подготовка к практическим занятиям
3	Подготовка к промежуточной аттестации.

5. Перечень изданий, которые рекомендуется использовать при освоении дисциплины (модуля).

№ п/ п	Библиографическое описание	Место доступа
1	Программирование в MS Office Excel на Visual Basic Т.Н. Глебова, Н.А. Зайцева; МИИТ. Каф. "Путевые, строительные машины и робототехнические комплексы" Однотомное издание МИИТ , 2007	НТБ (ЭЭ); НТБ (уч.6)
2	Бонцанини, М. Анализ социальных медиа на Python. Извлекайте и анализируйте данные из всех уголков социальной паутины на Python / М. Бонцанини ; перевод с английского А. В. Логунова. — Москва : ДМК Пресс, 2018. —	https://e.lanbook.com/book/108129

	288 с. — ISBN 978-5-97060-574-5	
3	Жуков, Р. А. Язык программирования Python. Практикум : учебное пособие / Р.А. Жуков. — Москва : ИНФРА-М, 2024. — 216 с. + Доп. материалы [Электронный ресурс]. — (Среднее профессиональное образование). - ISBN 978-5-16-015638-5	https://znanium.ru/catalog/document?id=439174&ysclid=lx1ogullfu764081385

6. Перечень современных профессиональных баз данных и информационных справочных систем, которые могут использоваться при освоении дисциплины (модуля).

Документация библиотеки pandas. <https://pandas.pydata.org/docs/>

Документация библиотеки matplotlib. <https://matplotlib.org/stable/users/index>

Документация библиотеки seaborn. <https://seaborn.pydata.org/tutorial.html>

Документация библиотеки pandas_profiling. <https://pandas-profiling.github.io/pandas-profiling/docs/master/index.html#pandas-profiling>

Документация scikit-learn https://scikit-learn.org/stable/user_guide.html

Документация библиотеки lime <https://github.com/marcotcr/lime>

7. Перечень лицензионного и свободно распространяемого программного обеспечения, в том числе отечественного производства, необходимого для освоения дисциплины (модуля).

PyCharm — интегрированная среда разработки для языка программирования Python, community, версия не ниже 2021.2 <https://www.jetbrains.com/ru-ru/pycharm/>

8. Описание материально-технической базы, необходимой для осуществления образовательного процесса по дисциплине (модулю).

Компьютер преподавателя

Компьютеры студентов

экран для проектора, маркерная доска,

Проектор

9. Форма промежуточной аттестации:

Зачет во 2 семестре.

10. Оценочные материалы.

Оценочные материалы, применяемые при проведении промежуточной аттестации, разрабатываются в соответствии с локальным нормативным актом РУТ (МИИТ).

Авторы:

руководитель образовательной
программы

О.Б. Проневич

Согласовано:

Директор

Б.В. Игольников

Руководитель образовательной
программы

О.Б. Проневич

Председатель учебно-методической
комиссии

Д.В. Паринов