

Документ подписан простой электронной подписью
Информация о владельце:
ФИО: Ястребов Олег Александрович
Должность: Ректор
Дата подписания: 15.05.2026 09:41:00
Уникальный программный ключ:
ca953a01204891083f939673078ef1a989dae18a

**Федеральное государственное автономное образовательное учреждение высшего образования
«Российский университет дружбы народов имени Патриса Лумумбы»
Филологический факультет**

(наименование основного учебного подразделения (ОУП)-разработчика ОП ВО)

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ

PYTHON ДЛЯ АНАЛИЗА ДАННЫХ

(наименование дисциплины/модуля)

Рекомендована МССН для направления подготовки/специальности:

45.03.01 ФИЛОЛОГИЯ

(код и наименование направления подготовки/специальности)

Освоение дисциплины ведется в рамках реализации основной профессиональной образовательной программы высшего образования (ОП ВО):

ФИЛОЛОГИЯ

(наименование (профиль/специализация) ОП ВО)

2026 г.

1. ЦЕЛЬ ОСВОЕНИЯ ДИСЦИПЛИНЫ

Дисциплина «Python для анализа данных» входит в программу бакалавриата «Филология» по направлению 45.03.01 «Филология» и изучается в 3 семестре 2 курса. Дисциплину реализует Вечерне-заочное отделение филологического факультета. Дисциплина состоит из 3 разделов и 9 тем и направлена на изучение является продолжением курса «Основы программирования на Python» и направлена на углублённое изучение методов анализа данных для решения прикладных задач в гуманитарных и социальных науках, включая: продвинутую обработку табличных данных и объединение разнородных источников; построение многоуровневых визуализаций для научных публикаций и управленческих отчётов; автоматизированный сбор данных из веб-источников и API; применение базовых методов машинного обучения для классификации и прогнозирования; текстовую аналитику и тематическое моделирование; создание интерактивных аналитических приложений

Целью освоения дисциплины является формирование у студентов продвинутых компетенций в области обработки структурированных данных (Pandas, NumPy) для подготовки управленческих решений, визуализации результатов анализа (Matplotlib, Seaborn, Plotly) для отчётов и презентаций, автоматизации сбора данных (API, веб-скрейпинг) из различных источников, применения базовых методов ML (scikit-learn) для прогнозной аналитики.

2. ТРЕБОВАНИЯ К РЕЗУЛЬТАТАМ ОСВОЕНИЯ ДИСЦИПЛИНЫ

Освоение дисциплины «Python для анализа данных» направлено на формирование у обучающихся следующих компетенций (части компетенций):

Таблица 2.1. Перечень компетенций, формируемых у обучающихся при освоении дисциплины (результаты освоения дисциплины)

| Шифр | Компетенция | Индикаторы достижения компетенции (в рамках данной дисциплины) |
|-------|--|---|
| УК-12 | Способен: искать нужные источники информации и данные, воспринимать, анализировать, запоминать и передавать информацию с использованием цифровых средств, а также с помощью алгоритмов при работе с полученными из различных источников данными с целью эффективного использования полученной информации для решения задач; проводить оценку информации, ее достоверность, строить логические умозаключения на основании поступающих информации и данных. | УК-12.1 Эффективно ищет и использует информацию, применяя цифровые средства и алгоритмы работы с данными из различных источников; УК-12.2 Использует предварительно проверенные на достоверность данные и информацию для построения умозаключений; |

3. МЕСТО ДИСЦИПЛИНЫ В СТРУКТУРЕ ОП ВО

Дисциплина «Python для анализа данных» относится к части, формируемой участниками образовательных отношений блока 1 «Дисциплины (модули)» образовательной программы высшего образования.

В рамках образовательной программы высшего образования обучающиеся также осваивают другие дисциплины и/или практики, способствующие достижению запланированных результатов освоения дисциплины «Python для анализа данных».

Таблица 3.1. Перечень компонентов ОП ВО, способствующих достижению запланированных результатов освоения дисциплины

| Шифр | Наименование компетенции | Предшествующие дисциплины/модули, практики* | Последующие дисциплины/модули, практики* |
|-------|--|---|--|
| УК-12 | Способен: искать нужные источники информации и данные, воспринимать, анализировать, запоминать и передавать информацию с использованием цифровых средств, а также с помощью алгоритмов при работе с полученными из различных источников данными с целью эффективного использования полученной информации для решения задач; проводить оценку информации, ее достоверность, строить логические умозаключения на основании поступающих информации и данных. | Цифровая грамотность; | Основы экономики и менеджмента; <i>Искусственный интеллект в литературоведении**;</i> <i>Искусственный интеллект в языкознании**;</i> <i>Искусственный интеллект в преподавании русского языка**;</i> |

* - заполняется в соответствии с матрицей компетенций и СУП ОП ВО

** - элективные дисциплины /практики

4. ОБЪЕМ ДИСЦИПЛИНЫ И ВИДЫ УЧЕБНОЙ РАБОТЫ

Общая трудоемкость дисциплины «Python для анализа данных» составляет «2» зачетные единицы.

Таблица 4.1. Виды учебной работы по периодам освоения образовательной программы высшего образования для очной формы обучения.

| Вид учебной работы | ВСЕГО, ак.ч. | | Семестр(-ы) |
|---|----------------|-----------|-------------|
| | | | 3 |
| Контактная работа, ак.ч. | 17 | | 17 |
| Лекции (ЛК) | 0 | | 0 |
| Лабораторные работы (ЛР) | 17 | | 17 |
| Практические/семинарские занятия (СЗ) | 0 | | 0 |
| Самостоятельная работа обучающихся, ак.ч. | 46 | | 46 |
| Контроль (экзамен/зачет с оценкой), ак.ч. | 9 | | 9 |
| Общая трудоемкость дисциплины | ак.ч. | 72 | 72 |
| | зач.ед. | 2 | 2 |

Общая трудоемкость дисциплины «Python для анализа данных» составляет «2» зачетные единицы.

Таблица 4.2. Виды учебной работы по периодам освоения образовательной программы высшего образования для очно-заочной формы обучения.

| Вид учебной работы | ВСЕГО, ак.ч. | | Семестр(-ы) |
|---|----------------|-----------|-------------|
| | | | 4 |
| Контактная работа, ак.ч. | 17 | | 17 |
| Лекции (ЛК) | 0 | | 0 |
| Лабораторные работы (ЛР) | 0 | | 0 |
| Практические/семинарские занятия (СЗ) | 17 | | 17 |
| Самостоятельная работа обучающихся, ак.ч. | 46 | | 46 |
| Контроль (экзамен/зачет с оценкой), ак.ч. | 9 | | 9 |
| Общая трудоемкость дисциплины | ак.ч. | 72 | 72 |
| | зач.ед. | 2 | 2 |

Общая трудоемкость дисциплины «Python для анализа данных» составляет «2» зачетные единицы.

Таблица 4.3. Виды учебной работы по периодам освоения образовательной программы высшего образования для заочной формы обучения.

| Вид учебной работы | ВСЕГО, ак.ч. | | Семестр(-ы) |
|--|----------------|-----------|-------------|
| | | | 5 |
| <i>Контактная работа, ак.ч.</i> | 8 | | 8 |
| Лекции (ЛК) | 4 | | 4 |
| Лабораторные работы (ЛР) | 0 | | 0 |
| Практические/семинарские занятия (СЗ) | 4 | | 4 |
| <i>Самостоятельная работа обучающихся, ак.ч.</i> | 60 | | 60 |
| <i>Контроль (экзамен/зачет с оценкой), ак.ч.</i> | 4 | | 4 |
| Общая трудоемкость дисциплины | ак.ч. | 72 | 72 |
| | зач.ед. | 2 | 2 |

5. СОДЕРЖАНИЕ ДИСЦИПЛИНЫ

Таблица 5.1. Содержание дисциплины (модуля) по видам учебной работы

| Номер раздела | Наименование раздела дисциплины | Наименование темы | | Содержание темы | Вид учебной работы* |
|---------------|---------------------------------|-------------------|---------------------------------------|--|---------------------|
| Раздел 1 | Обработка и анализ данных | 1.1 | Продвинутая работа с Pandas | Повторение основ: DataFrame, загрузка, фильтрация. Продвинутые операции: множественная фильтрация (query), работа с MultiIndex, метод apply для построчных преобразований, замена значений (map, replace). Цепочки методов (method chaining). Практика: загрузка датасета международных индексов (Democracy Index / Rule of Law Index) за несколько лет; фильтрация по регионам и диапазону значений; создание нового столбца-категории через apply; ранжирование стран внутри регионов | ЛР |
| | | 1.2 | Объединение и реструктуризация данных | Объединение таблиц: merge (inner, left, outer), concat. Сводные таблицы: pivot_table, crosstab. Преобразование формата: melt (из «широкого» в «длинный»), pivot. Обработка дубликатов (duplicated, drop_duplicates). Практика: объединение таблицы социально-экономических показателей стран (World Bank) с таблицей политических индексов по коду страны; построение сводной таблицы «регион × год → среднее значение индекса»; преобразование данных из широкого формата в длинный для последующей визуализации. | ЛР |
| | | 1.3 | NumPy и вычисления над массивами | Массивы ndarray: создание, индексация, срезы. Векторизованные операции (поэлементное сложение, умножение, сравнение). Математические функции (np.mean, np.std, np.corrcoef). Булева индексация. Генерация случайных чисел (np.random). Практика: расчёт нормализованных показателей (z-score) для сравнения стран по нескольким индексам; вычисление матрицы корреляций вручную через NumPy и сравнение с Pandas; генерация случайной выборки для моделирования социологического опроса. | ЛР |
| Раздел 2 | Автоматизация аналитики | 2.1 | Продвинутая работа с API | Повторение: REST API, GET-запрос, JSON. Новое: параметризованные запросы, пагинация (автоматический перебор страниц), обработка ошибок и тайм-аутов (try/except, status_code). Библиотека httpx (асинхронные запросы — обзорно). Практика: автоматическая загрузка данных по 50 странам из API Всемирного банка с пагинацией; обработка | ЛР |

| Номер раздела | Наименование раздела дисциплины | Наименование темы | | Содержание темы | Вид учебной работы* |
|---------------|---------------------------------|-------------------|---|--|---------------------|
| | | | | ошибок при недоступности сервера; формирование итогового DataFrame и сохранение в CSV для дальнейшего анализа. | |
| | | 2.2 | Продвинутый веб-скрейпинг | Повторение: BeautifulSoup, find, find_all. Новое: обход нескольких страниц в цикле; извлечение таблиц (pd.read_html); работа с динамическими страницами (Playwright — обзорно). Задержки между запросами (time.sleep). Правовые ограничения. Практика: сбор таблицы данных со страницы Transparency International (CPI) или международного рейтинга; обход нескольких страниц каталога с автоматическим сбором записей; объединение результатов в единый DataFrame. | ЛР |
| | | 2.3 | Очистка и подготовка данных | Типичные проблемы «грязных» данных: пропуски, дубликаты, несогласованные форматы, выбросы. Стратегии обработки пропусков (удаление, заполнение средним / медианой / интерполяция). Обнаружение выбросов (IQR, z-score). Приведение типов (astype, pd.to_numeric, pd.to_datetime). Нормализация текстовых полей (lower, strip, replace). Практика: «грязный» датасет с реальными проблемами (дубликаты строк, пропуски в числовых столбцах, даты в разных форматах, выбросы); пошаговая очистка и подготовка до состояния, пригодного для анализа; документирование каждого шага. | ЛР |
| Раздел 3 | Прикладная аналитика | 3.1 | Введение в машинное обучение: классификация | Что такое ML: обучение с учителем, задача классификации. Подготовка данных: разделение выборки (train/test), нормализация. Логистическая регрессия (sklearn LogisticRegression). Оценка качества: accuracy, матрица ошибок (confusion_matrix), classification_report. Практика: классификация стран по типу политического режима (демократия / автократия) на основе социально-экономических показателей; разделение данных; обучение модели; оценка точности; интерпретация — какие признаки наиболее значимы. | ЛР |
| | | 3.2 | Текстовая аналитика: продвинутый уровень | Повторение: spaCy, NER, облако слов. Новое: TF-IDF (sklearn TfidfVectorizer) — числовое представление текстов. Анализ тональности (sentiment analysis — базовый уровень). Сравнение текстов по TF-IDF (косинусное сходство). Практика: загрузка корпуса текстов (политические выступления / аннотации статей / резолюции); вычисление TF-IDF; определение ключевых терминов для каждого документа; попарное сравнение текстов | ЛР |

| Номер раздела | Наименование раздела дисциплины | Наименование темы | | Содержание темы | Вид учебной работы* |
|---------------|---------------------------------|-------------------|----------------------------------|---|---------------------|
| | | | | по сходству; визуализация результатов в виде тепловой карты сходства. ЛР | |
| | | 3.3 | Комплексный аналитический проект | Объединение всех навыков курса в сквозной исследовательский проект. Этапы: формулирование вопроса → сбор данных (API + скрейпинг + файлы) → очистка → анализ (статистика + ML/NLP) → визуализация (графики + карты) → представление результатов (Streamlit-дашборд или Word-отчёт + презентация). Практика: мини-проект на выбор: (а) прогнозирование типа политического режима по социально-экономическим показателям (классификация + дашборд); (б) кластеризация регионов / стран с интерактивной картой и профилями кластеров; (в) текстовый анализ корпуса документов (TF-IDF + сходство + визуализация); (г) комплексный страновой / региональный аналитический продукт — от сбора данных до готового дашборда. Защита проекта. | ЛР |

* - заполняется только по **ОЧНОЙ** форме обучения: ЛК – лекции; ЛР – лабораторные работы; СЗ – практические/семинарские занятия.

6. МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

Таблица 6.1. Материально-техническое обеспечение дисциплины

| Тип аудитории | Оснащение аудитории | Специализированное учебное/лабораторное оборудование, ПО и материалы для освоения дисциплины (при необходимости) |
|----------------------------|---|--|
| Компьютерный класс | Компьютерный класс для проведения занятий, групповых и индивидуальных консультаций, текущего контроля и промежуточной аттестации, оснащенная персональными компьютерами (в количестве 3 шт.), доской (экраном) и техническими средствами мультимедиа презентаций. | |
| Для самостоятельной работы | Аудитория для самостоятельной работы обучающихся (может использоваться для проведения семинарских занятий и консультаций), оснащенная комплектом специализированной мебели и компьютерами с доступом в ЭИОС. | |

* - аудитория для самостоятельной работы обучающихся указывается **ОБЯЗАТЕЛЬНО!**

7. УЧЕБНО-МЕТОДИЧЕСКОЕ И ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

Основная литература:

1. Федоров, Д. Ю. Программирование на языке высокого уровня Python : учебное пособие для вузов / Д. Ю. Федоров. — 5-е изд., перераб. и доп. — Москва : Издательство Юрайт, 2025. — 227 с. — (Высшее образование). — ISBN 978-5-534-18715-1. — URL: <https://urait.ru/bcode/562700>

2. Чернышев, С. А. Основы программирования на Python : учебник / С. А. Чернышев. — 2-е изд., перераб. и доп. — Москва : Издательство Юрайт, 2025. — 349 с. — (Профессиональное образование). — ISBN 978-5-534-17056-6. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: <https://urait.ru/bcode/567823>

Дополнительная литература:

1. Лучано Рамальо. Свободный Python. Чистое, лаконичное и эффективное программирование. — 2-е изд. — СПб.: Питер, 2024. — 896 с. — ISBN 978-5-4461-1950-7

2. Шаблон научных вычислений на Python: VanderPlas, J. Python Data Science Handbook. — 2nd ed. — O'Reilly Media, 2023. — 583 p. — ISBN 978-1-098-12122-8. — URL: <https://jakevdp.github.io/PythonDataScienceHandbook/> (открытый доступ)

Ресурсы информационно-телекоммуникационной сети «Интернет»:

1. ЭБС РУДН и сторонние ЭБС, к которым студенты университета имеют доступ на основании заключенных договоров

- Электронно-библиотечная система РУДН – ЭБС РУДН

<https://mega.rudn.ru/MegaPro/Web>

- ЭБС «Университетская библиотека онлайн» <http://www.biblioclub.ru>

- ЭБС «Юрайт» <http://www.biblio-online.ru>

- ЭБС «Консультант студента» www.studentlibrary.ru

- ЭБС «Знаниум» <https://znanium.ru/>
- 2. Базы данных и поисковые системы
 - Sage <https://journals.sagepub.com/>
 - Springer Nature Link <https://link.springer.com/>
 - Wiley Journal Database <https://onlinelibrary.wiley.com/>
 - Наукометрическая база данных Lens.org <https://www.lens.org>

Учебно-методические материалы для самостоятельной работы обучающихся при освоении дисциплины/модуля:*

1. Курс лекций по дисциплине «Python для анализа данных».

* - все учебно-методические материалы для самостоятельной работы обучающихся размещаются в соответствии с действующим порядком на странице дисциплины **в ТУИС!**

РАЗРАБОТЧИК:

| | | |
|-----------------------|----------------|---|
| <hr/> | <hr/> | <hr/> |
| <i>Должность, БУП</i> | <i>Подпись</i> | Широкова Евгения Павловна <i>Фамилия И.О.</i> |

РУКОВОДИТЕЛЬ БУП:

| | | |
|--------------------------------|----------------|---|
| <hr/> | <hr/> | <hr/> |
| Доцент <i>Должность БУП</i> | <i>Подпись</i> | Подолько Павел Михайлович <i>Фамилия И.О.</i> |

РУКОВОДИТЕЛЬ ОП ВО:

| | | |
|---------------------------------|----------------|--|
| <hr/> | <hr/> | <hr/> |
| Доцент <i>Должность, БУП</i> | <i>Подпись</i> | Рыбаков Михаил Анатольевич <i>Фамилия И.О.</i> |