Документ подписан простой электронной подписью
Информация о владельце:
ФИО: Ястребов Олег Александрович
Должность: Ректор
Дата подписания: 03.11.2026 12:34:54
Уникальный программный ключ:
ca953a0120d891083f939673078ef1a989dae18a

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

Факультет гуманитарных и социальных наук

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

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ

PYTHON ДЛЯ АНАЛИЗА ДАННЫХ

(наименование дисциплины/модуля)

Рекомендована МССН для направления подготовки/специальности:

41.03.04 ПОЛИТОЛОГИЯ

(код и наименование направления подготовки/специальности)

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

ПОЛИТОЛОГИЯ

(наименование (профиль/специализация) ОП ВО)

2026 г.

1. ЦЕЛЬ ОСВОЕНИЯ ДИСЦИПЛИНЫ

Дисциплина «Python для анализа данных» входит в программу бакалавриата «Политология» по направлению 41.03.04 «Политология» и изучается в 4 семестре 2 курса. Дисциплину реализует Кафедра прикладного искусственного интеллекта. Дисциплина состоит из 3 разделов и 18 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 Осуществляет поиск нужных источников информации и данных, воспринимает, анализирует, запоминает и передает 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;; УК-12.2 Проводит оценку информации, ее достоверность, строит логические умозаключения на основании поступающих информации и данных;;

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	Продвинутый Excel**; Основы программирования на	Преддипломная практика; Искусственный интеллект в

Шифр	Наименование компетенции	Предшествующие дисциплины/модули, практики*	Последующие дисциплины/модули, практики*
	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	Python**; Инфографика и технология презентаций**; Цифровая трансформация политических отношений;	профессиональной деятельности; Идентичность в политике; Политическая стабильность**;

* - 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)**
			4
Контактная работа, ак.ч	*34*		*34*
Лекции (**ЛК**)	0		0
Лабораторные работы (**ЛР**)	0		0
Практические/семинарские занятия (**СЗ**)	34		34
Самостоятельная работа обучающихся, ак.ч.	*29*		*29*
Контроль (экзамен/зачет с оценкой), ак.ч.	*9*		*9*
Общая трудоемкость дисциплины ак.ч.	**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	*34*		*34*
Лекции (**ЛК**)	0		0
Лабораторные работы (**ЛР**)	0		0
Практические/семинарские занятия (**СЗ**)	34		34
Самостоятельная работа обучающихся, ак.ч.	*29*		*29*
Контроль (экзамен/зачет с оценкой), ак.ч.	*9*		*9*
Общая трудоемкость дисциплины ак.ч.	**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.	СЗ
		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.	СЗ
		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.	СЗ
		1.4	Продвинутая визуализация: Seaborn	Библиотека Seaborn: философия и отличие от Matplotlib. Графики распределений (histplot, kdeplot). Категориальные графики (boxplot, violinplot, swarmplot). Тепловые карты (heatmap) для корреляционных матриц. Парные графики (pairplot). Настройка стилей и палитр. Практика: построение корреляционной тепловой карты социально-экономических показателей; boxplot сравнения индекса демократии по типам политических режимов; pairplot для выявления зависимостей между ВВП, образованием и свободой прессы.	СЗ
		1.5	Интерактивная визуализация: Plotly	Plotly Express: line, bar, scatter, choropleth, sunburst, treemap. Интерактивные элементы: hover-подписи, зум, фильтры. Plotly Graph Objects: настройка layout, annotations, subplots. Экспорт в HTML. Практика: интерактивная хороплетная карта мира по индексу верховенства права с hover-информацией; анимированный scatter по годам (bubble chart: ВВП vs. ожидаемая продолжительность жизни, размер — население); sunburst-диаграмма структуры расходов бюджета.	СЗ
		1.6	Статистический анализ и проверка гипотез	Повторение: корреляция, линейная регрессия. Новое: множественная регрессия (statsmodels OLS), интерпретация коэффициентов, R^2, p-value. Проверка гипотез: t-тест, хи-квадрат, ANOVA (scipy.stats). Доверительные интервалы. Практика: построение множественной регрессии «индекс демократии ~ ВВП + образование + урбанизация»; интерпретация результатов (какой фактор значим); проверка гипотезы о различии среднего уровня коррупции между группами стран; оформление результатов в виде таблицы для публикации.	СЗ

Номер раздела	Наименование раздела дисциплины	Наименование темы		Содержание темы	Вид учебной работы*
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 ошибок при недоступности сервера; формирование итогового DataFrame и сохранение в CSV для дальнейшего анализа.	СЗ
		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.	СЗ
		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.	СЗ
		2.4	Временные ряды: продвинутый уровень	Повторение: datetime, rolling. Новое: ресемплирование (resample: дни → месяцы → кварталы), декомпозиция (statsmodels seasonal_decompose — тренд, сезонность, остатки). Расчёт темпов роста (year-over-year). Сравнение нескольких рядов на одном графике. Практика: загрузка ежемесячных данных (число обращений граждан / заявлений / публикаций) за 5 лет; декомпозиция ряда; выявление сезонности и тренда; расчёт year-over-year; визуализация каждого компонента на отдельном subplot	СЗ
		2.5	Автоматизация отчётов: продвинутый уровень	Повторение: python-docx. Новое: шаблонизация с Jinja2 (подстановка данных в текстовый шаблон); генерация Excel-отчётов с форматированием (openpyxl: стили, условное форматирование, формулы); пакетная генерация документов в цикле. Практика: создание шаблона аналитической записки (Jinja2 + python-docx); загрузка данных по 5 странам; автоматическая генерация 5 отдельных Word-документов (по одному на страну) с индивидуальными таблицами и графиками; параллельно — Excel-файл со сводной таблицей и условным форматированием.	СЗ
		2.6	Data Pipeline: сбор → обработка → визуализация	Понятие аналитического пайплайна. Проектирование цепочки: источник данных → загрузка → очистка → анализ → визуализация → сохранение результатов. Организация кода: функции для каждого этапа. Воспроизводимость: фиксация шагов в Jupyter-ноутбуке. Практика: построение полного пайплайна — запрос данных из API Всемирного банка → очистка (обработка пропусков, приведение типов) → расчёт показателей → построение 3 графиков (линейный, столбчатый, карта) → сохранение графиков и итоговой таблицы в файлы. Весь код организован в виде функций.	СЗ
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;	СЗ

Номер раздела	Наименование раздела дисциплины	Наименование темы		Содержание темы	Вид учебной работы*
				разделение данных; обучение модели; оценка точности; интерпретация — какие признаки наиболее значимы.	
		3.2	Машинное обучение: кластеризация	Обучение без учителя: задача кластеризации. Алгоритм k-средних (KMeans). Выбор числа кластеров (метод локтя). Визуализация кластеров (scatter plot с цветовой кодировкой). Интерпретация результатов. Практика: кластеризация стран по набору показателей (ВВП, образование, индекс свободы, урбанизация); подбор числа кластеров; визуализация на плоскости (2 главных признака); описание профиля каждого кластера — какие страны попали в какую группу и почему.	СЗ
		3.3	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 по сходству; визуализация результатов в виде тепловой карты сходства.	СЗ
		3.4	Геоаналитика: продвинутый уровень	Повторение: Folium, маркеры, хороплет. Новое: GeoPandas (GeoDataFrame, пространственные операции — обзорно). Несколько слоёв на одной карте (LayerControl). Динамические хороплеты (анимация по годам — TimestampedGeoJson, обзорно). Plotly choropleth с анимацией (animation_frame). Практика: построение анимированной хороплетной карты изменения индекса демократии по странам за 10 лет (Plotly animation_frame); карта Folium с двумя переключаемыми слоями (например, ВВП и индекс свободы прессы).	СЗ
		3.5	Интерактивные дашборды: продвинутый уровень Интерактивные дашборды: продвинутый уровень	Повторение: Streamlit, базовые виджеты. Новое: многостраничное приложение (st.navigation / pages). Динамическая фильтрация: связанные виджеты (выбор региона → обновление списка стран). Кэширование (@st.cache_data). Отображение карт (streamlit-folium). Вкладки (st.tabs). Метрики (st.metric с delta). Практика: создание многостраничного дашборда: страница 1 — обзорная таблица и карта по выбранному индексу; страница 2 — сравнение двух стран (графики, метрики с delta); страница 3 — временная динамика с выбором периода. Все виджеты связаны.	СЗ
		3.6	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.	СЗ

* - 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.	
Семинарская	Аудитория для проведения занятий семинарского типа, групповых и индивидуальных консультаций, текущего контроля и промежуточной аттестации, оснащенная комплектом специализированной мебели 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**!

РАЗРАБОТЧИКИ:

Ассистент		Широкова Е.П.
Должность	*Подпись*	*Фамилия И.О*

РУКОВОДИТЕЛЬ ОП ВО:

Профессор		Почта Ю.М.
Должность	*Подпись*	*Фамилия И.О*

РУКОВОДИТЕЛЬ БУП:

Заведующий кафедрой		Подолько П.М.
Должность	*Подпись*	*Фамилия И.О*