

МИНОБРНАУКИ РОССИИ



Федеральное государственное автономное образовательное учреждение
высшего образования
«Российский государственный гуманитарный университет»
ФГАОУ ВО «РГГУ»

Кафедра фундаментальной и прикладной математики

**ПРОГРАММИРОВАНИЕ И ОБРАБОТКА ДАННЫХ
В СОЦИОЛОГИЧЕСКИХ ИССЛЕДОВАНИЯХ**

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ

Социология (магистратура) 39.04.01

Цифровая социология и социальная аналитика

Уровень квалификации выпускника (магистр)

Форма обучения (очная)

РПД адаптирована для лиц
с ограниченными возможностями
здоровья и инвалидов

Москва 2026

«Программирование и обработка данных в социологических исследованиях»

Рабочая программа дисциплины

Составитель:

Кандидат экономических наук, доцент *А.Ю. Журавлев*

УТВЕРЖДЕНО

Протокол заседания кафедры фундаментальной и прикладной математики

№ 5 от 19.12. 2025

ОГЛАВЛЕНИЕ

1. Пояснительная записка	4
1.1 Цель и задачи дисциплины.....	4
1.2 Перечень планируемых результатов обучения по дисциплине (модулю), соотнесенных с индикаторами достижения компетенций.....	6
2. Структура дисциплины	9
3. Содержание дисциплины	9
4. Образовательные технологии	15
5. Оценка планируемых результатов обучения	18
5.1 Система оценивания.....	18
5.2 Критерии выставления оценки по дисциплине.....	19
5.3 Оценочные средства (материалы) для текущего контроля успеваемости, промежуточной аттестации обучающихся по дисциплине.....	20
6. Учебно-методическое и информационное обеспечение дисциплины	37
6.1 Список источников и литературы.....	37
Основная литература.....	37
Дополнительная литература.....	37
6.2. Перечень ресурсов информационно-телекоммуникационной сети «Интернет».....	39
7. Материально-техническое обеспечение дисциплины	40
8. Обеспечение образовательного процесса для лиц	41
с ограниченными возможностями здоровья и инвалидов.....	41
9. Методические материалы	43
9.1 Планы практических занятий.....	43
9.2. Методические указания.....	48
Приложение 1. Аннотация рабочей программы дисциплины	50

1. Пояснительная записка

1.1 Цель и задачи дисциплины

Цель дисциплины: формирование у студентов целостного системного представления о возможностях языка программирования Python и современных технологий искусственного интеллекта (включая нейросетевые модели) как инструментальной базы для решения широкого круга профессиональных задач цифрового социолога, а также развитие устойчивых практических навыков автоматизированного сбора, обработки, анализа и визуализации социальных данных, включая применение методов машинного обучения для выявления скрытых закономерностей, моделирования социальных процессов и интерпретации результатов с учетом этических и методологических требований современной социологической науки.

Задачи дисциплины:

1. **Сформировать методологическую основу** для понимания роли программирования и искусственного интеллекта в трансформации современной социологии: раскрыть концепцию AI Society, объяснить переход от традиционных методов анализа к нейросетевому моделированию, обосновать необходимость синтеза технологических инноваций с социально-гуманитарным знанием для обеспечения валидности и надежности результатов.
2. **Систематизировать и углубить знания языка Python** применительно к задачам обработки социологических данных: освоить продвинутые возможности библиотек Pandas, NumPy, SciPy для работы с большими массивами, включая методы очистки, трансформации, агрегации и реструктуризации данных различной природы (опросные данные, цифровые следы, открытые источники).
3. **Обучить студентов методам автоматизированного сбора данных** из современных цифровых источников: освоить техники веб-скрапинга динамических страниц (Selenium, Playwright), работу с API социальных сетей и открытых платформ (ВКонтакте, Telegram, YouTube), извлечение данных из PDF-отчетов и других неструктурированных форматов.
4. **Развить навыки визуально-аналитического исследования данных** с использованием современных библиотек визуализации (Matplotlib, Seaborn, Plotly) для создания интерактивных дашбордов и подготовки публикационно-качественной графики, отражающей многомерную структуру социальных явлений.
5. **Сформировать компетенции в области применения методов машинного обучения** для решения социологических задач: обучить студентов использованию

библиотеки Scikit-learn для построения классификационных и кластерных моделей (сегментация аудиторий, типологизация респондентов, выявление предикторов социального поведения), а также методам оценки качества и интерпретации полученных моделей.

6. **Дать системное представление о нейросетевых технологиях** и их прикладном потенциале в социологии: объяснить архитектурные особенности различных типов нейронных сетей (полносвязные, сверточные для анализа изображений, рекуррентные и трансформеры для текстов), обучить использованию библиотек TensorFlow/Keras и PyTorch для построения прогностических моделей социальных процессов.
7. **Обучить студентов применению больших языковых моделей (LLM)** в социологической практике: освоить методы промпт-инжиниринга для анализа тональности, тематического моделирования, извлечения сущностей и суммаризации текстов (открытые вопросы, комментарии, посты), а также критической оценки результатов генерации и выявления алгоритмических смещений.
8. **Сформировать компетенции в области моделирования социальных процессов** с использованием агент-ориентированных моделей (Mesa) и графовых нейронных сетей для анализа распространения информации, влияния в социальных сетях и прогнозирования коллективных действий.
9. **Воспитать культуру ответственного применения технологий ИИ** в социальных исследованиях: сформировать понимание этических проблем (алгоритмическая предвзятость, проблема «черного ящика», цифровое неравенство, информированное согласие), обучить методам обеспечения прозрачности (Explainable AI) и интерпретируемости моделей, привить навыки документирования аналитических пайплайнов для обеспечения воспроизводимости результатов.
10. **Подготовить студентов к самостоятельной исследовательской деятельности**, требующей разработки комплексных аналитических решений на стыке социологии и Data Science, включая постановку задач, обоснованный выбор инструментария, реализацию полного цикла анализа, критическую интерпретацию результатов и формулировку научно обоснованных выводов и практических рекомендаций.

1.2 Перечень планируемых результатов обучения по дисциплине (модулю), соотнесенных с индикаторами достижения компетенций

Компетенция (код и наименование)	Индикаторы компетенций (код и наименование)	Результаты обучения
ОПК-1. Способен обоснованно отбирать и использовать современные информационно-коммуникационные технологии для решения профессиональных задач	ОПК-1.1. Обосновывает выбор информационно-коммуникационных технологий для постановки и решения задач социологического исследования	<p>Знать:</p> <ul style="list-style-type: none"> классификацию современных инструментов программирования и библиотек Python для анализа данных (Pandas, NumPy, Scikit-learn, TensorFlow) и их функциональные возможности; методологические критерии выбора технологий в зависимости от типа исследовательской задачи (описательный анализ, прогнозирование, кластеризация, моделирование), характера данных (структурированные/неструктурированные, статические/поточные, малые/большие данные) и требований к интерпретируемости результатов; границы применимости методов машинного обучения и нейросетевого моделирования в социологическом контексте и риски, связанные с некорректным выбором инструментария. <p>Уметь:</p> <ul style="list-style-type: none"> проводить критический анализ альтернативных технологических решений для конкретной социологической задачи; аргументированно обосновывать выбор языка программирования, библиотек и фреймворков в научных отчетах, проектных заявках и технических заданиях; прогнозировать потенциальные ограничения и этические последствия применения выбранных технологий и предлагать способы их минимизации. <p>Владеть:</p> <ul style="list-style-type: none"> методологией технико-методологического обоснования выбора ИКТ в исследовательских проектах; навыками презентации и защиты принятых технологических решений перед научным сообществом и заказчиками; способностью разрабатывать критерии оценки эффективности применения

		различных технологий в социологических исследованиях.
ОПК-1. Способен обоснованно отбирать и использовать современные информационно-коммуникационные технологии для решения профессиональных задач	ОПК-1.2. Предлагает постановку задач по поиску, обобщению социологических данных.	<p>Знать:</p> <p>типологию источников социологических данных в цифровой среде (опросные массивы, открытые государственные данные, API социальных сетей, веб-архивы, цифровые следы); современные методы и технологии автоматизированного сбора данных (парсинг, скрапинг, работа с API), включая их технические и правовые ограничения; понимать, как исследовательская задача транслируется в формальные критерии отбора и структуру собираемых данных, и как качество исходных данных влияет на валидность последующих выводов.</p> <p>Уметь:</p> <p>формулировать задачи сбора данных в терминах, допускающих программную реализацию (определять источники, поля, периоды сбора, критерии фильтрации); разрабатывать технические спецификации для сбора данных из различных цифровых источников; проектировать структуры хранения собранных данных (схемы баз данных, форматы файлов), обеспечивающие их целостность и удобство последующего анализа.</p> <p>Владеть:</p> <p>навыками разработки комплексных стратегий сбора данных для многокомпонентных исследовательских проектов; методами оценки полноты, репрезентативности и качества собранных данных; способностью документировать процессы сбора данных для обеспечения прозрачности и воспроизводимости исследования.</p>
ПК-1. Способен к разработке предложений по совершенствованию методов проведения социологических исследований	ПК-1.2. Разрабатывает предложения по совершенствованию технологических процессов, методов сбора и анализа информации в	<p>Знать:</p> <p>современные тенденции развития инструментальных средств анализа данных (автоматизация, интеграция, использование нейросетей) и их потенциал для повышения эффективности исследовательских процессов; понимать принципы построения</p>

	социологическом и маркетинговом исследовании.	<p>аналитических пайплайнов (конвейеров обработки данных), обеспечивающих автоматизацию рутинных операций, масштабируемость и воспроизводимость результатов;</p> <p>знать методы оценки эффективности внедряемых технологических инноваций (сокращение времени, снижение затрат, повышение качества анализа).</p> <p>Уметь:</p> <p>выявлять узкие места и неэффективные операции в существующих процессах сбора и анализа данных;</p> <p>разрабатывать и внедрять усовершенствованные процедуры на основе Python-автоматизации: создавать скрипты для пакетной обработки, проектировать ETL-процессы, настраивать регулярно обновляемые дашборды, интегрировать нейросетевые модули для автоматической обработки текстов или изображений;</p> <p>документировать предложенные улучшения в форме методических рекомендаций и регламентов.</p> <p>Владеть:</p> <p>навыками реинжиниринга исследовательских процессов с использованием современных технологий;</p> <p>методами количественной и качественной оценки эффективности внедренных улучшений;</p> <p>способностью разрабатывать программы обучения и повышения квалификации для сотрудников по работе с новыми инструментами и методами;</p> <p>навыками презентации предложений по совершенствованию руководству и заказчикам.</p>
--	---	--

1.3 Место дисциплины в структуре образовательной программы

Дисциплина «Программирование и обработка данных в социологических исследованиях» относится к вариативной части, формируемой участниками образовательных отношений по направлению «Социология», базового блока дисциплин учебного плана.

Для освоения дисциплины необходимы знания, умения и владения, сформированные в ходе изучения дисциплин «Инструменты аналитической обработки данных в табличных редакторах» и «Базы данных».

В результате освоения дисциплины формируются знания, умения и владения, необходимые для прохождения практик и выполнения выпускной квалификационной работы.

2. Структура дисциплины

Структура дисциплины для очной формы обучения

Общая трудоёмкость дисциплины составляет 7 з.е., 252 академических часа. Объем дисциплины в форме контактной работы обучающихся с педагогическими работниками и (или) лицами, привлекаемыми к реализации образовательной программы на иных условиях, при проведении учебных занятий:

Семестр	Тип учебных занятий	Количество часов
2	Лекции	12
2	Практические занятия	48
3	Лекции	20
3	Практические занятия	60
Всего:		140

Объем дисциплины в форме самостоятельной работы обучающихся составляет 94 академических часа.

3. Содержание дисциплины

Семестр 1. Основы программирования и аналитики данных на Python

Тема 1. Python в арсенале современного социолога: от цифровых следов к AI Society.

Роль программирования в трансформации социологической методологии. Концепция AI Society: искусственный интеллект как актор социальных изменений. Экосистема Python для анализа данных: обзор ключевых библиотек (Pandas, NumPy, Scikit-learn, TensorFlow). Настройка профессиональной среды разработки (PyCharm, Jupyter Lab, Google Colab). Понятие воспроизводимости исследований (reproducible research) и роль кода в современной науке.

Тема 2. Продвинутая работа с типами данных и структурами в Python.

Углубленное изучение встроенных типов: изменяемые и неизменяемые объекты, хэшируемость. Списки, кортежи, словари, множества: методы оптимизации памяти и

скорости доступа. Генераторы списков и словарей (comprehensions) для эффективной обработки данных. Итераторы и генераторы: работа с потоками данных, функция yield. Применение для обработки больших массивов социологической информации.

Тема 3. Функциональное программирование в анализе данных.

Функции высшего порядка: map, filter, reduce. Лямбда-функции для краткой записи преобразований. Декораторы для расширения функциональности (логирование, тайминг, кеширование). Модуль itertools для эффективной работы с итераторами. Применение функционального подхода для трансформации социологических данных.

Тема 4. Объектно-ориентированное программирование для структурирования кода.

Основы ООП: классы, объекты, атрибуты, методы. Инкапсуляция и сокрытие данных. Наследование и полиморфизм для создания иерархий объектов (например, классы для разных типов респондентов или источников данных). Магические методы (str, repr, call). Создание собственных классов для хранения и обработки социологических данных.

Тема 5. Работа с файловыми системами и форматами данных.

Модули os, shutil, pathlib для управления файлами и директориями. Чтение и запись данных в различных форматах: CSV (углубленно: обработка сложных разделителей, кодировок), JSON (работа с вложенными структурами), XML (ElementTree), Excel (openpyxl). Пакетная обработка множества файлов. Создание структурированных архивов данных.

Тема 6. Библиотека NumPy: эффективные вычисления с многомерными массивами.

Массивы ndarray: преимущества перед списками. Создание, индексация, срезы, булева индексация. Универсальные функции (ufunc) для векторизованных вычислений. Линейная алгебра (dot, linalg). Статистические функции (mean, std, var, percentile). Применение для обработки числовых социологических матриц (например, батарей вопросов).

Тема 7. Библиотека Pandas: структуры Series и DataFrame.

Объекты Series и DataFrame: создание, индексация, основные атрибуты. Чтение данных из различных источников (read_csv, read_excel, read_json, read_sql). Методы info(), describe(), head(), tail() для первичного знакомства с данными. Работа с индексами (set_index, reset_index, reindex). Применение для загрузки и первичного анализа реальных социологических датасетов.

Тема 8. Очистка и предобработка данных в Pandas.

Проблема «грязных данных» в социологии. Работа с пропущенными значениями (`isna`, `dropna`, `fillna`, `interpolate`). Поиск и удаление дубликатов. Преобразование типов данных (`astype`, `to_numeric`, `to_datetime`). Работа с категориальными переменными (`CategoricalDtype`, `cat`-методы). Применение строковых методов (`str`-аксессор) для очистки текстовых полей.

Тема 9. Фильтрация, выборка и трансформация данных в Pandas.

Булева индексация для сложных условий отбора. Методы `loc` и `iloc` для гибкой выборки. Применение функций к данным: `apply`, `map`, `applymap`. Замена значений (`replace`). Условная замена с помощью `np.where`. Создание новых столбцов на основе существующих. Практикум: подготовка «чистого» датасета для анализа.

Тема 10. Агрегация и группировка данных.

Группировка данных с `groupby`. Агрегация с использованием встроенных функций и пользовательских функций (`agg`). Многомерная группировка. Трансформация данных внутри групп (`transform`). Фильтрация групп (`filter`). Анализ различий между социально-демографическими группами с помощью группировки.

Тема 11. Анализ частот и таблицы сопряженности.

Метод `value_counts()` для анализа распределений. Нормированные частоты. Таблицы сопряженности: `pd.crosstab()`, добавление маргинальных итогов (`margins`). Расчет долей по строкам и столбцам. Визуализация таблиц сопряженности. Выявление взаимосвязей между категориальными переменными.

Тема 12. Работа с временными рядами в Pandas.

Тип данных `datetime`. Создание временных индексов (`DatetimeIndex`). Ресемплинг (`resample`) для изменения частоты данных. Сдвиги (`shift`, `tshift`) для анализа лагов. Скользящие окна (`rolling`) для сглаживания и расчета скользящих средних. Анализ сезонности и трендов в социальных показателях.

Тема 13. Объединение и интеграция данных из разных источников.

Конкатенация `DataFrame` (`pd.concat`). Соединение таблиц по ключу: `pd.merge()` и его типы (`inner`, `outer`, `left`, `right`). Обработка дублирующихся ключей. Соединение по индексу (`join`).

Практикум: обогащение данных опроса внешней статистической информацией (например, данными Росстата по регионам).

Тема 14. Визуализация данных с Matplotlib.

Философия Matplotlib: figure и axes. Построение базовых графиков: линейные, столбчатые, гистограммы, ящики с усами, точечные диаграммы. Настройка параметров графиков: заголовки, подписи осей, легенда, цветовые схемы, размеры. Создание многопанельных композиций (subplots). Сохранение графиков в высоком разрешении для публикаций.

Тема 15. Статистическая визуализация с Seaborn.

Seaborn как надстройка над Matplotlib для статистической визуализации. Построение сложных графиков: violinplot, boxenplot, stripplot, swarmplot. Визуализация распределений (histplot, kdeplot, displot). Визуализация взаимосвязей (lmpplot, regplot, pairplot, jointplot). Тепловые карты (heatmap, clustermap) для матриц корреляций. Настройка стилей и палитр.

Тема 16. Исследовательский анализ данных (EDA) на Python.

Понятие EDA и его роль в социологическом исследовании. Комплексный подход к анализу: описательная статистика, визуализация распределений, выявление выбросов, анализ пропусков, поиск взаимосвязей. Разведочный анализ реального социологического датасета с формулировкой предварительных гипотез. Оформление результатов EDA в формате Jupyter Notebook.

Семестр 2. Машинное обучение, нейросети и моделирование социальных процессов

Тема 17. Введение в машинное обучение для социологов.

Основные парадигмы ML: обучение с учителем, без учителя, с подкреплением. Типы задач: классификация, регрессия, кластеризация, снижение размерности. Понятие признаков (features) и целевой переменной (target). Обучающая и тестовая выборки. Проблема переобучения и недообучения. Библиотека Scikit-learn: структура и основные компоненты.

Тема 18. Предобработка данных для машинного обучения.

Масштабирование признаков: стандартизация (StandardScaler), нормализация (MinMaxScaler). Кодирование категориальных переменных: One-Hot Encoding, Ordinal Encoding. Обработка выбросов. Разделение данных на обучающую, валидационную и

тестовую выборку (`train_test_split`). Создание конвейеров обработки (Pipeline) для воспроизводимости.

Тема 19. Линейные модели для классификации и регрессии.

Линейная регрессия для прогнозирования количественных показателей. Логистическая регрессия для бинарной классификации. Интерпретация коэффициентов моделей в социологическом контексте. Метрики качества: для регрессии (MSE, MAE, R^2), для классификации (accuracy, precision, recall, F1-score, ROC-AUC). Матрица ошибок (confusion matrix).

Тема 20. Деревья решений и ансамблевые методы.

Деревья решений: принцип работы, интерпретируемость, проблема переобучения. Случайный лес (Random Forest) как ансамбль деревьев. Градиентный бустинг (XGBoost, LightGBM, CatBoost). Важность признаков (feature importance) в ансамблевых моделях. Применение для выявления ключевых факторов социального поведения.

Тема 21. Кластеризация и сегментация социальных групп.

Задачи кластеризации в социологии: типологизация респондентов, выделение аудиторных сегментов. Метод k-средних (k-means): выбор числа кластеров (метод локтя, силуэтный анализ). Иерархическая кластеризация (dendrogram). Алгоритм DBSCAN для выделения кластеров произвольной формы. Интерпретация полученных кластеров.

Тема 22. Снижение размерности и визуализация многомерных данных.

Проблема многомерности в социологических данных. Метод главных компонент (PCA): математические основы, интерпретация компонент, доля объясненной дисперсии. t-SNE и UMAP для визуализации многомерных структур. Применение для визуализации социальных пространств и дистанций между группами.

Тема 23. Обработка естественного языка (NLP) для анализа текстов.

Особенности текстовых данных в социологии (открытые вопросы, посты, комментарии). Предобработка текста: токенизация, лемматизация, стемминг, удаление стоп-слов. Векторизация текстов: мешок слов (CountVectorizer), TF-IDF (TfidfVectorizer). Визуализация текстовых данных: облака слов.

Тема 24. Тематическое моделирование и анализ тональности.

Тематическое моделирование (Latent Dirichlet Allocation, LDA) для выявления скрытых тем в массиве текстов. Выбор оптимального числа тем. Интерпретация тематических структур. Анализ тональности (sentiment analysis) с использованием правил и машинного обучения. Применение для анализа общественных настроений.

Тема 25. Введение в нейронные сети и глубокое обучение.

Биологическая и искусственная нейронная сеть. Архитектура полносвязной сети (MLP): входной, скрытые, выходной слои. Функции активации (ReLU, sigmoid, tanh). Функции потерь и оптимизаторы. Библиотеки TensorFlow и Keras: создание простых нейросетевых моделей. Применение для прогнозирования социальных явлений.

Тема 26. Продвинутое моделирование архитектур нейронных сетей.

Сверточные нейронные сети (CNN): принципы свертки и пулинга, применение для анализа изображений (анализ визуального контента в соцсетях). Рекуррентные нейронные сети (RNN, LSTM, GRU) для анализа последовательностей (тексты, временные ряды). Механизм внимания (attention) и трансформеры (transformers). Обзор современных архитектур.

Тема 27. Большие языковые модели (LLM) в социологических исследованиях.

Архитектура трансформера и принципы работы больших языковых моделей (GPT, LLaMA, DeepSeek). Использование API языковых моделей. Промпт-инжиниринг для решения социологических задач: анализ тональности, извлечение сущностей, суммаризация, категоризация открытых вопросов. Создание систем для автоматической обработки текстовых массивов.

Тема 28. Агент-ориентированное моделирование (ABM) социальных процессов.

Принципы агент-ориентированного моделирования: агенты, среда, правила взаимодействия. Библиотека Mesa для создания ABM на Python. Проектирование простых моделей: распространение мнений, сегрегация (модель Шеллинга), диффузия инноваций. Визуализация и анализ результатов моделирования. Сравнение ABM с нейросетевым моделированием.

Тема 29. Анализ социальных сетей (SNA) с Python.

Основные понятия SNA: графы, узлы, связи, направленные/ненаправленные графы. Библиотека NetworkX для создания и анализа графов. Метрики центральности (степень, близость, посредничество, собственный вектор). Выявление сообществ (community detection). Визуализация графов. Применение для анализа структуры социальных медиа.

Тема 30. Графовые нейронные сети (GNN) для анализа социальных взаимодействий.

Ограничения классического SNA. Принципы работы графовых нейронных сетей: агрегация информации от соседей. Библиотеки PyTorch Geometric и DGL. Применение GNN для прогнозирования влияния (social influence prediction), рекомендации связей, классификации узлов. Социологически-информированное представление графов (SIGRL).

Тема 31. Этика и ответственность в применении ИИ в социальных исследованиях.

Проблема алгоритмической предвзятости (bias) и способы ее выявления. Прозрачность моделей (Explainable AI, XAI): методы LIME, SHAP для интерпретации "черных ящиков". Приватность данных и федеративное обучение. Информированное согласие при сборе цифровых следов. Этические протоколы и международные стандарты. Проблема «черного ящика» и доверия к алгоритмам.

Тема 32. Итоговый проект: комплексное исследование с применением Python и нейросетей.

Обобщение и систематизация знаний, полученных в курсе. Полный цикл исследовательского проекта: постановка социологической задачи, выбор источников данных, сбор данных (API/парсинг/готовые датасеты), предобработка и EDA, выбор и применение методов анализа (статистика, ML, нейросети), визуализация результатов, интерпретация и формулировка выводов. Оформление проекта в формате Jupyter Notebook с полным кодом и выводами. Подготовка презентации и защита проекта.

4. Образовательные технологии

№	Тема	Образовательные технологии
1	Python в арсенале современного социолога	Лекция-визуализация с элементами дискуссии «AI Society: вызовы и возможности», установка ПО (тренинг), самостоятельная работа
2	Продвинутая работа с типами данных	Проблемная лекция (как эффективно хранить большие данные?), лабораторное занятие (оптимизация кода), тренинг
3	Функциональное	Лекция с демонстрацией, лабораторное занятие

№	Тема	Образовательные технологии
	программирование	(применение map/filter/reduce к социологическим данным), парное программирование
4	Объектно-ориентированное программирование	Лекция-визуализация, лабораторное занятие (создание классов для опросных данных), мини-проект
5	Работа с файловыми системами	Лекция с разбором кейсов (пакетная обработка 1000 файлов), лабораторное занятие, тренинг
6	Библиотека NumPy	Лекция-визуализация, лабораторное занятие (матричные операции с социологическими данными), самостоятельная работа
7	Библиотека Pandas: Series и DataFrame	Лекция с демонстрацией, лабораторное занятие (загрузка реального датасета), работа в группах
8	Очистка и предобработка данных	Проблемная лекция (проблема «грязных данных»), лабораторное занятие (кейс по очистке), тренинг «Найди ошибку»
9	Фильтрация, выборка и трансформация	Лекция-визуализация, лабораторное занятие (подготовка данных для анализа), консультация
10	Агрегация и группировка данных	Лекция с разбором примеров, лабораторное занятие (анализ различий между группами), мини-исследование
11	Анализ частот и таблицы сопряженности	Лекция-визуализация, лабораторное занятие (анализ взаимосвязей), дискуссия об интерпретации
12	Работа с временными рядами	Проблемная лекция (анализ динамики социальных показателей), лабораторное занятие, кейс-стади
13	Объединение данных из разных источников	Лекция с демонстрацией, лабораторное занятие (обогащение данных опроса статистикой), практикум
14	Визуализация с Matplotlib	Лекция-визуализация с элементами дискуссии «Как не обмануть графиком», лабораторное занятие
15	Статистическая визуализация с Seaborn	Лекция-визуализация, лабораторное занятие (конкурс на лучшую визуализацию), тренинг
16	Исследовательский анализ данных (EDA)	Проектная деятельность (индивидуально), консультации, мини-конференция по результатам EDA
17	Введение в машинное обучение	Лекция-визуализация, дискуссия о применимости ML в социологии, самостоятельная работа с литературой
18	Предобработка данных для ML	Лекция, лабораторное занятие (подготовка данных для моделей), тренинг
19	Линейные модели	Лекция с разбором примеров, лабораторное занятие (прогнозирование и классификация), интерпретация результатов
20	Деревья решений и ансамбли	Лекция-визуализация, лабораторное занятие (выявление ключевых факторов), групповая работа
21	Кластеризация и сегментация	Проблемная лекция (как выделить целевые группы?), лабораторное занятие, кейс по маркетинговым исследованиям
22	Снижение размерности	Лекция с демонстрацией, лабораторное занятие (визуализация социальных пространств), дискуссия
23	Обработка естественного языка (NLP)	Лекция-визуализация, лабораторное занятие (предобработка открытых вопросов), тренинг
24	Тематическое	Лекция, лабораторное занятие (анализ комментариев)

№	Тема	Образовательные технологии
	моделирование и тональность	в соцсетях), мини-исследование
25	Введение в нейронные сети	Лекция-визуализация, лабораторное занятие (первая нейросеть на Keras), консультация
26	Продвинутое архитектуры	Лекция с обзором современных исследований, лабораторное занятие (эксперименты с архитектурами)
27	Большие языковые модели (LLM)	Лекция-визуализация, мастер-класс по промпт-инжинирингу, лабораторное занятие с API LLM
28	Агент-ориентированное моделирование	Проблемная лекция, лабораторное занятие (модель Шеллинга), групповая проектная работа
29	Анализ социальных сетей (SNA)	Лекция-визуализация, лабораторное занятие (NetworkX), анализ реальной сети
30	Графовые нейронные сети (GNN)	Лекция с разбором статей (SIGRL), лабораторное занятие (PyTorch Geometric), дискуссия
31	Этика и ответственность в ИИ	Лекция с элементами дискуссии, разбор кейсов (алгоритмическая предвзятость), круглый стол
32	Итоговый проект	Проектная деятельность (индивидуально/в группах), консультации, защита проектов с презентацией

При реализации программы дисциплины «Программирование и обработка данных в социологических исследованиях» используются различные образовательные технологии: во время аудиторных занятий занятия проводятся в виде лекционных и семинарских занятий.

Лекции проводятся в виде: лекции-визуализации, проблемные лекции, лекции с разбором конкретных ситуаций с использованием ПК и компьютерного проектора.

На семинарских занятиях, проводимых по типу развернутой беседы с обсуждением докладов, предусмотрены выступления студентов по определенной тематике с групповым обсуждением этих сообщений. Формой текущего контроля по таким темам служит качество подготовленного доклада (критерии: содержательность, актуальность, мастерство изложения) и оценка участия в обсуждении доклада.

Самостоятельная работа студентов подразумевает изучение источников и литературы, предусмотренной программой дисциплины, написание реферата, рецензии, перевода иностранной статьи (см. список примерных тем, список источников и литературы).

В период временного приостановления посещения обучающимися помещений и территории РГГУ для организации учебного процесса с применением электронного обучения и дистанционных образовательных технологий могут быть использованы следующие образовательные технологии:

~ видеолекции;

- ~ онлайн-лекции в режиме реального времени;
- ~ электронные учебники, учебные пособия, научные издания в электронном виде и доступ к иным электронным образовательным ресурсам;
- ~ системы для электронного тестирования;
- ~ консультации с использованием телекоммуникационных средств.

5. Оценка планируемых результатов обучения

5.1 Система оценивания

Оценка знаний студентов складывается из *промежуточных* аттестаций (60%), а также *итоговой* аттестации в письменной форме (40%).

Форма контроля	Макс. количество баллов	
	За одну работу	Всего
Текущий контроль:		
~ экспресс-тест;	1 балл	15 баллов
~ участие в дискуссии на семинаре;	1 балл	13 баллов
~ подготовка доклада;	5 баллов	5 баллов
~ промежуточная контрольная работа;	10 баллов	10 баллов
~ реферат.	17 баллов	17 баллов
<i>Промежуточная аттестация (итоговая практическая работа)</i>		40 баллов
Итого за семестр экзамен		100 баллов

Полученный совокупный результат конвертируется в традиционную шкалу оценок и в шкалу оценок Европейской системы переноса и накопления кредитов (European Credit Transfer System; далее – ECTS) в соответствии с таблицей:

100-балльная шкала	Традиционная шкала		Шкала ECTS
95 – 100	отлично	зачтено	A
83 – 94			B
68 – 82	хорошо		C
56 – 67	удовлетворительно		D
50 – 55			E
20 – 49	неудовлетворительно	не зачтено	FX
0 – 19			F

5.2 Критерии выставления оценки по дисциплине

Баллы/ Шкала ECTS	Оценка по дисциплине	Критерии оценки результатов обучения по дисциплине
100-83/ A,B	«отлично»/ «зачтено (отлично)»/ «зачтено»	<p>Выставляется обучающемуся, если он глубоко и прочно усвоил теоретический и практический материал, может продемонстрировать это на занятиях и в ходе промежуточной аттестации.</p> <p>Обучающийся исчерпывающе и логически стройно излагает учебный материал, умеет увязывать теорию с практикой, справляется с решением задач профессиональной направленности высокого уровня сложности, правильно обосновывает принятые решения.</p> <p>Свободно ориентируется в учебной и профессиональной литературе.</p> <p>Оценка по дисциплине выставляется обучающемуся с учётом результатов текущей и промежуточной аттестации.</p> <p>Компетенции, закреплённые за дисциплиной, сформированы на уровне – «высокий».</p>
82-68/ C	«хорошо»/ «зачтено (хорошо)»/ «зачтено»	<p>Выставляется обучающемуся, если он знает теоретический и практический материал, грамотно и по существу излагает его на занятиях и в ходе промежуточной аттестации, не допуская существенных неточностей.</p> <p>Обучающийся правильно применяет теоретические положения при решении практических задач профессиональной направленности разного уровня сложности, владеет необходимыми для этого навыками и приёмами.</p> <p>Достаточно хорошо ориентируется в учебной и профессиональной литературе.</p> <p>Оценка по дисциплине выставляется обучающемуся с учётом результатов текущей и промежуточной аттестации.</p> <p>Компетенции, закреплённые за дисциплиной, сформированы на уровне – «хороший».</p>
67-50/ D,E	«удовлетвори- тельно»/ «зачтено (удовлетвори- тельно)»/ «зачтено»	<p>Выставляется обучающемуся, если он знает на базовом уровне теоретический и практический материал, допускает отдельные ошибки при его изложении на занятиях и в ходе промежуточной аттестации.</p> <p>Обучающийся испытывает определённые затруднения в применении теоретических положений при решении практических задач профессиональной направленности стандартного уровня сложности, владеет необходимыми для этого базовыми навыками и приёмами.</p> <p>Демонстрирует достаточный уровень знания учебной литературы по дисциплине.</p> <p>Оценка по дисциплине выставляется обучающемуся с учётом результатов текущей и промежуточной аттестации.</p>

Баллы/ Шкала ECTS	Оценка по дисциплине	Критерии оценки результатов обучения по дисциплине
		Компетенции, закреплённые за дисциплиной, сформированы на уровне – «достаточный».
49-0/ F,FX	«неудовлетворительно»/ не зачтено	<p>Выставляется обучающемуся, если он не знает на базовом уровне теоретический и практический материал, допускает грубые ошибки при его изложении на занятиях и в ходе промежуточной аттестации.</p> <p>Обучающийся испытывает серьёзные затруднения в применении теоретических положений при решении практических задач профессиональной направленности стандартного уровня сложности, не владеет необходимыми для этого навыками и приёмами. Демонстрирует фрагментарные знания учебной литературы по дисциплине.</p> <p>Оценка по дисциплине выставляется обучающемуся с учётом результатов текущей и промежуточной аттестации.</p> <p>Компетенции на уровне «достаточный», закреплённые за дисциплиной, не сформированы.</p>

5.3 Оценочные средства (материалы) для текущего контроля успеваемости, промежуточной аттестации обучающихся по дисциплине

Оценка знаний, умений, опыта деятельности студента осуществляется преподавателем в ходе реализации каждой из предусмотренных рабочей программой форм занятий. Основным способом оценки выступает контроль за посещением занятий, уровнем подготовленности к ним, качеством предъявляемых устных и письменных знаний, умений, навыков, формирующих профессиональный опыт студентов-социологов. Предусмотрены следующие виды контроля:

Экспресс-опрос

Формируемые компетенции и индикаторы их достижения: ОПК-1 (ОПК-1.1, ОПК-1.2) и ПК-1 (ПК-1.2)

В конце лекции преподаватель задает один вопрос по теме лекции. В течение максимум 5 минут, студент обязан написать краткий (максимум 10 строк) ответ на поставленный вопрос. Экспресс-тест оценивается в 1 балл.

При экспресс-опросе по каждой из прочитанных лекций выявляется формирование компетенции по критическому восприятию, обобщению, анализу профессиональной информации.

Доклад

Формируемые компетенции и индикаторы их достижения: ОПК-1 (ОПК-1.1, ОПК-1.2) и ПК-1 (ПК-1.2)

При оценивании устного доклада, учитываются следующие характеристики:

- ~ полнота подготовки задания, информативность;
- ~ глубина проработки изучаемых вопросов (проработка материала должна отражать владение основными понятиями и концепциями);
- ~ системность, наличие выводов.
- ~ максимально за один доклад можно получить 5 баллов.

Участие в дискуссии на семинарах

Формируемые компетенции и индикаторы их достижения: ОПК-1 (ОПК-1.1, ОПК-1.2) и ПК-1 (ПК-1.2)

Выступления студентов на семинарских занятиях позволяют контролировать формирование у них способности использовать полученные знания в преподавании социологии, формировать практические навыки и умения для участия в научных и научно-прикладных исследованиях, аналитической и консалтинговой деятельности. Максимально за одно выступление можно получить 1 балл.

Подготовка реферата

Формируемые компетенции и индикаторы их достижения: ОПК-1 (ОПК-1.1, ОПК-1.2) и ПК-1 (ПК-1.2)

Требования к содержанию реферата. Реферат должен отражать основные подходы к изучаемому объекту, а также содержать авторскую оценку описываемого объекта. Объем работы: 10 стр. печатного текста.

Оформление работы

Шрифт TimesNewRoman, кегль 14, поля 3 см слева, 2 сверху и внизу, 1,5 справа. Красная строка (абзацный отступ) 1,25 см.

Фамилия автора – распределение справа, название статьи и фамилия автора статьи по центру, основной текст – по ширине. Реферат оценивается максимум в 17 баллов.

Тематика рефератов:

1. Эволюция методов анализа данных в социологии: от статистических пакетов к нейросетевым моделям.
2. Концепция AI Society: теоретические основания и эмпирические исследования.
3. Сравнительный анализ библиотек Python для обработки и анализа социологических данных (Pandas, Polars, Dask).
4. Методы визуализации многомерных социологических данных: современные подходы и инструменты.
5. Применение методов машинного обучения для прогнозирования электорального поведения.
6. Кластерный анализ в маркетинговых исследованиях: сегментация потребительских аудиторий.
7. Тематическое моделирование в анализе социальных медиа: выявление повесток дня и общественных настроений.
8. Анализ тональности комментариев в социальных сетях как инструмент мониторинга общественного мнения.
9. Нейросетевые методы анализа визуального контента в социальных исследованиях (анализ фотографий, мемов, рекламы).
10. Применение рекуррентных нейронных сетей (LSTM) для прогнозирования временных рядов социально-экономических показателей.
11. Большие языковые модели (LLM) в социологии: возможности, ограничения, этические аспекты.
12. Промпт-инжиниринг как методологическая проблема: как корректно «общаться» с ИИ для получения валидных данных.
13. Анализ социальных сетей (SNA) в изучении распространения информации и влияния.
14. Графовые нейронные сети (GNN) для прогнозирования социального влияния: обзор современных архитектур.
15. Агент-ориентированное моделирование социальных процессов: от теории к практике на Python.
16. Моделирование диффузии инноваций с помощью агент-ориентированного подхода.
17. Веб-скрапинг как метод сбора социологических данных: правовые и этические аспекты.

18. Использование API социальных сетей (ВКонтакте, Telegram) для сбора данных в исследовательских целях.
19. Проблема алгоритмической предвзятости в системах искусственного интеллекта и социальные последствия.
20. Методы объяснимого ИИ (XAI) и их роль в обеспечении доверия к алгоритмическим решениям.
21. Приватность данных в цифровую эпоху: федеративное обучение и дифференциальная приватность.
22. Цифровое неравенство как социальная проблема: методы измерения и анализа.
23. Использование методов машинного обучения для выявления деструктивного контента в социальных медиа.
24. Анализ цифровых следов для изучения социального капитала и сетевых связей.
25. Прогнозирование протестной активности на основе анализа данных социальных сетей.
26. Моделирование динамики общественного мнения с использованием нейросетевых подходов.
27. Сравнительный анализ инструментов для работы с большими данными в Python (Dask, Vaex, Modin).
28. Интеграция качественных и количественных методов в цифровой социологии (mixed-methods с поддержкой Python).
29. Разработка интерактивных дашбордов для мониторинга социальных показателей (Plotly Dash, Streamlit).
30. Будущее социологии в эпоху искусственного интеллекта: новые исследовательские парадигмы и вызовы.

Вопросы для самопроверки:

1. Какой метод Pandas используется для чтения данных из CSV-файла?
 - A) read_excel()
 - B) read_csv() (*верно*)
 - C) load_csv()
 - D) open_csv()
2. Какая библиотека Python предназначена для численных расчетов с многомерными массивами?
 - A) Pandas
 - B) Matplotlib

- C) NumPy (*верно*)
 - D) Scikit-learn
3. Какой метод Pandas удаляет строки с пропущенными значениями?
- A) fillna()
 - B) dropna() (*верно*)
 - C) remove_na()
 - D) clean()
4. Какая функция в Pandas позволяет применить произвольную функцию к каждому элементу Series или DataFrame?
- A) map() / applymap() (*верно*)
 - B) filter()
 - C) transform()
 - D) agg()
5. Какой метод используется для группировки данных в Pandas?
- A) groupby() (*верно*)
 - B) pivot_table()
 - C) aggregate()
 - D) stack()
6. Какая метрика качества классификации рассчитывается как $(TP+TN)/(TP+TN+FP+FN)$?
- A) Precision
 - B) Recall
 - C) Accuracy (*верно*)
 - D) F1-score
7. Какой метод машинного обучения относится к обучению без учителя?
- A) Логистическая регрессия
 - B) Случайный лес
 - C) К-средних (k-means) (*верно*)
 - D) Градиентный бустинг
8. Какая библиотека используется для построения нейросетевых моделей?
- A) Scikit-learn
 - B) TensorFlow/Keras (*верно*)
 - C) Statsmodels
 - D) NetworkX

9. Что такое переобучение (overfitting) в машинном обучении?
- A) Модель слишком простая и не улавливает закономерности
 - B) Модель слишком сложная и запоминает шум в данных (*верно*)
 - C) Модель обучалась слишком мало времени
 - D) Модель показывает одинаковые результаты на train и test
10. Какой метод используется для снижения размерности данных?
- A) PCA (метод главных компонент) (*верно*)
 - B) KNN (метод k-ближайших соседей)
 - C) SVM (метод опорных векторов)
 - D) K-means
11. Какая функция активации чаще всего используется в скрытых слоях нейронных сетей?
- A) Сигмоида
 - B) Тангенс гиперболический
 - C) ReLU (*верно*)
 - D) Линейная
12. Что такое TF-IDF в анализе текстов?
- A) Мера важности слова в документе относительно коллекции (*верно*)
 - B) Метод кластеризации текстов
 - C) Библиотека для обработки естественного языка
 - D) Формат хранения текстовых данных
13. Какая библиотека используется для анализа графов и социальных сетей?
- A) NetworkX (*верно*)
 - B) Seaborn
 - C) Plotly
 - D) BeautifulSoup
14. Какой тип нейронной сети лучше всего подходит для анализа последовательностей (например, текста)?
- A) Сверточная (CNN)
 - B) Полносвязная (DNN)
 - C) Рекуррентная (RNN/LSTM) (*верно*)
 - D) Однослойный перцептрон
15. Что такое промпт-инжиниринг?
- A) Разработка архитектуры нейронной сети
 - B) Искусство составления запросов к языковым моделям для получения нужных

- результатов (*верно*)
- C) Метод оптимизации гиперпараметров
 - D) Способ визуализации данных
16. Какая метрика позволяет оценить, какая доля предсказанных положительных классов действительно является положительной?
- A) Accuracy
 - B) Recall
 - C) Precision (*верно*)
 - D) ROC-AUC
17. Какой метод используется для выявления скрытых тем в коллекции текстовых документов?
- A) Логистическая регрессия
 - B) Латентное размещение Дирихле (LDA) (*верно*)
 - C) Метод главных компонент (PCA)
 - D) Кластеризация k-means
18. Что такое алгоритмическая предвзятость (bias) в контексте ИИ?
- A) Смещение в оценке модели, когда она систематически ошибается в определенную сторону (*верно*)
 - B) Техническая ошибка в коде
 - C) Слишком медленная работа алгоритма
 - D) Недостаточная глубина нейронной сети
19. Какая библиотека используется для агент-ориентированного моделирования в Python?
- A) Mesa (*верно*)
 - B) Pygame
 - C) Django
 - D) Flask
20. Что означает аббревиатура XAI?
- A) Extreme AI
 - B) Explainable AI (объяснимый искусственный интеллект) (*верно*)
 - C) XML Artificial Intelligence
 - D) X-ray Artificial Intelligence
21. Какой метод используется для оценки важности признаков в моделях типа Random Forest?
- A) feature_importances_ (*верно*)

- B) coefficients_
 - C) loadings_
 - D) components_
22. Как в Pandas объединить два DataFrame по строкам (вертикально)?
- A) pd.merge()
 - B) pd.concat(axis=0) (*верно*)
 - C) pd.join()
 - D) pd.append() (устаревший)
23. Какой метод визуализации лучше всего подходит для отображения корреляционной матрицы?
- A) Гистограмма
 - B) Ящик с усами
 - C) Тепловая карта (heatmap) (*верно*)
 - D) Точечная диаграмма
24. Что такое графовая нейронная сеть (GNN)?
- A) Сеть, которая рисует графы
 - B) Нейронная сеть, работающая с данными, представленными в виде графов (*верно*)
 - C) Сеть для визуализации графов функций
 - D) Тип рекуррентной сети
25. Какая функция в Python позволяет создать генератор?
- A) Использование ключевого слова yield (*верно*)
 - B) Использование return в цикле
 - C) Использование global
 - D) Использование nonlocal
26. Какой метод используется для преобразования категориальных переменных в бинарные признаки (0/1)?
- A) Label Encoding
 - B) One-Hot Encoding (*верно*)
 - C) Ordinal Encoding
 - D) Binary Encoding
27. Что такое валидационная выборка в машинном обучении?
- A) Данные, на которых модель обучается
 - B) Данные, на которых оценивается финальная модель

- C) Данные, используемые для настройки гиперпараметров (*верно*)
- D) Все собранные данные
28. Какая функция в NumPy создает массив из последовательности чисел с заданным шагом?
- A) argray()
- B) linspace()
- C) arange() (*верно*)
- D) zeros()
29. Какой метод визуализации в Seaborn сочетает ящик с усами и оценку плотности распределения?
- A) boxplot
- B) violinplot (*верно*)
- C) swarmplot
- D) stripplot
30. Что такое федеративное обучение (federated learning)?
- A) Обучение модели на данных из разных стран
- B) Подход, при котором модель обучается на устройствах пользователей без передачи данных на центральный сервер (*верно*)
- C) Обучение с участием нескольких университетов
- D) Метод ансамблирования моделей
31. Какая метрика используется для оценки качества кластеризации?
- A) Silhouette score (*верно*)
- B) R^2
- C) MAE
- D) Log-loss
32. Что делает метод apply() в Pandas?
- A) Применяет функцию ко всем элементам DataFrame (*верно*)
- B) Фильтрует данные
- C) Группирует данные
- D) Объединяет таблицы
33. Какая архитектура нейронной сети лежит в основе больших языковых моделей?
- A) Сверточная (CNN)
- B) Трансформер (Transformer) (*верно*)
- C) Рекуррентная (RNN)
- D) Автокодировщик

34. Какой метод используется для анализа тональности текста?
- A) Sentiment analysis (*верно*)
 - B) Topic modeling
 - C) Named Entity Recognition
 - D) Summarization
35. Что такое LIME в контексте ХАИ?
- A) Библиотека для визуализации
 - B) Метод объяснения предсказаний модели путем аппроксимации локально-интерпретируемой моделью (*верно*)
 - C) Тип нейронной сети
 - D) Метод оптимизации
36. Какой тип графа в NetworkX является ориентированным?
- A) Graph
 - B) DiGraph (*верно*)
 - C) MultiGraph
 - D) WeightedGraph
37. Какой метод используется для масштабирования признаков к диапазону [0, 1]?
- A) StandardScaler
 - B) MinMaxScaler (*верно*)
 - C) RobustScaler
 - D) Normalizer
38. Что такое ресемплинг (resample) в Pandas?
- A) Изменение частоты временного ряда (*верно*)
 - B) Повторная выборка из данных
 - C) Изменение размера DataFrame
 - D) Замена пропущенных значений
39. Какая библиотека используется для работы с API социальных сетей ВКонтакте?
- A) tweepy
 - B) vk_api (*верно*)
 - C) facebook-sdk
 - D) telegram-bot-api
40. Что такое кросс-валидация (cross-validation)?
- A) Метод оценки модели, при котором данные многократно разбиваются на обучающие и тестовые части (*верно*)
 - B) Проверка модели на кросс-платформенную совместимость

- C) Валидация данных из разных источников
 - D) Метод объединения моделей
41. Какой метод используется для извлечения данных из динамических веб-страниц, требующих выполнения JavaScript?
- A) BeautifulSoup
 - B) Requests
 - C) Selenium / Playwright (*верно*)
 - D) Scrapy
42. Что такое SHAP в контексте интерпретации моделей?
- A) Метод, основанный на теории игр для объяснения вклада каждого признака в предсказание (*верно*)
 - B) Библиотека для визуализации графов
 - C) Метод кластеризации
 - D) Архитектура нейронной сети
43. Какая функция в Python используется для создания декоратора?
- A) @decorator
 - B) Функция, принимающая другую функцию и возвращающая функцию (*верно*)
 - C) lambda
 - D) yield
44. Какой метод Pandas создает сводную таблицу, аналогичную Excel PivotTable?
- A) groupby()
 - B) pivot_table() (*верно*)
 - C) crosstab()
 - D) stack()
45. Что такое модель «черного ящика» (black box model)?
- A) Модель, которую нельзя запустить
 - B) Модель, внутренняя логика которой непрозрачна и трудно интерпретируема (*верно*)
 - C) Модель, защищенная паролем
 - D) Модель с открытым исходным кодом
46. Какая библиотека используется для построения агент-ориентированных моделей?
- A) Mesa (*верно*)
 - B) AgentPy
 - C) NetLogo (не Python)
 - D) AIMA

47. Какой метод используется для поиска сообществ в графах?
- A) community detection (например, алгоритм Лувена) (*верно*)
 - B) shortest path
 - C) PageRank
 - D) betweenness centrality
48. Что такое градиентный бустинг?
- A) Метод последовательного построения ансамбля моделей, где каждая следующая модель исправляет ошибки предыдущей (*верно*)
 - B) Метод параллельного построения ансамбля моделей
 - C) Метод оптимизации нейросетей
 - D) Метод визуализации
49. Какой тип данных в Pandas оптимизирован для хранения категориальных переменных?
- A) object
 - B) category (*верно*)
 - C) string
 - D) categorical (устаревший)
50. Что такое цифровое неравенство в контексте AI Society?
- A) Неравномерный доступ к цифровым технологиям и данным между разными группами населения (*верно*)
 - B) Разница в скорости интернета
 - C) Неравенство в зарплатах IT-специалистов
 - D) Отсутствие цифровых навыков у старшего поколения

Открытые вопросы (на размышление):

1. Предложите аналитический пайплайн для исследования связи между активностью пользователей в социальных сетях и их реальным социальным самочувствием. Какие методы сбора и анализа данных вы бы использовали?
2. Как можно использовать большие языковые модели (LLM) для обработки открытых вопросов в массовых опросах? Какие риски и ограничения необходимо учитывать?
3. Опишите, как методы машинного обучения могут помочь в выявлении групп риска (например, склонных к депрессии или экстремизму) по их цифровым следам. Какие этические проблемы возникают при таком анализе?

4. Сравните метод главных компонент (PCA) и t-SNE для визуализации многомерных социологических данных. В каких случаях предпочтительнее каждый из них?
5. Предложите систему метрик для оценки качества работы чат-бота, созданного для консультирования граждан по социальным вопросам.
6. Как можно использовать анализ социальных сетей (SNA) и графовые нейронные сети (GNN) для изучения распространения дезинформации?
7. Что такое алгоритмическая предвзятость и как она может проявляться при использовании нейросетей в социальных исследованиях? Приведите примеры и предложите способы минимизации.
8. Разработайте план агент-ориентированной модели для изучения динамики общественного мнения в условиях информационного воздействия.
9. Какие методы интерпретации моделей (XAI) вы примените, чтобы объяснить заказчику (неспециалисту), почему модель относит определенных респондентов к группе риска?
10. Опишите потенциальные социальные последствия широкого внедрения систем распознавания эмоций в публичных пространствах. Какие исследования нужны для оценки этих последствий?

Промежуточная контрольная работа

Формируемые компетенции и индикаторы их достижения: ОПК-1 (ОПК-1.1), ПК-1 (ПК-1.2) и ПК-2 (ПК-2.1)

Промежуточная контрольная работа проводится в середине семестра с целью сделать «срез» знаний по классической части курса «Инструменты аналитической обработки данных в табличных редакторах».

Тематика вопросов для промежуточной контрольной:

1. Классификация типов данных в Excel. Почему некорректное определение типа данных может исказить результаты социологического исследования?
2. Правила организации "чистых данных" (Tidy Data) для последующего анализа в табличных редакторах.
3. Сравнительный анализ относительной и абсолютной адресации. Приведите пример из практики социолога, где использование каждого типа критически важно.
4. Методы импорта данных в Excel из внешних источников (CSV, TXT, веб-страницы).

5. Статистические функции Excel для описания центральной тенденции (СРЗНАЧ, МЕДИАНА, МОДА). Когда применение медианы предпочтительнее среднего арифметического?
6. Функции для подсчета частот (СЧЁТЕСЛИ, СЧЁТЕСЛИМН) в анализе результатов закрытых вопросов анкеты.
7. Использование логических функций (ЕСЛИ, И, ИЛИ) для сегментации респондентов на социально-демографические группы.
8. Применение функций для работы с текстом (ЛЕВСИМВ, ПСТР, НАЙТИ) при обработке открытых вопросов и "грязных" данных.
9. Вычисление возраста респондента и длительности событий с помощью функций дат и времени (РАЗДАТ, ГОД, МЕСЯЦ).
10. Алгоритм поиска и устранения дубликатов записей в базе данных респондентов.
11. Простой и расширенный фильтр: сравнительный анализ возможностей для отбора подвыборок по сложным критериям.
12. Инструмент "Текст по столбцам" как способ парсинга структурированных текстовых данных.
13. Сортировка данных: простая, многоуровневая, пользовательская. Значение сортировки для подготовки отчетов.
14. "Умные таблицы" в Excel: преимущества перед обычными диапазонами при добавлении новых данных и написании формул.
15. Использование инструмента "Промежуточные итоги" для автоматического расчета показателей по группам (например, средний балл по факультетам).
16. Логика построения сводной таблицы: строки, столбцы, значения, фильтры.
17. Настройка типов вычислений в сводных таблицах: сумма, среднее, количество, доля от общей суммы.
18. Группировка данных в сводных таблицах (например, объединение возраста в когорты).
19. Создание срезов и шкал времени для интерактивной фильтрации данных в сводных таблицах.
20. Принципы выбора типа диаграммы для визуализации социологических данных.
21. Построение и настройка гистограммы для сравнения категориальных переменных.
22. Использование круговых диаграмм: правила применения и ограничения (когда круговая диаграмма неинформативна).
23. Построение графиков для отображения динамики социальных процессов во времени.

24. Точечная диаграмма (Scatter plot) как инструмент первичного выявления корреляционной связи.
25. Добавление и форматирование элементов диаграммы: легенда, подписи данных, оси.
26. Построение комбинированных диаграмм (например, столбцы + график) для отображения разнородных данных.
27. Создание и настройка сводных диаграмм на основе сводных таблиц.
28. Методы проверки данных на ошибки ввода и выбросы.
29. Использование условного форматирования для быстрого визуального анализа аномалий.
30. Подготовка данных для построения частотных распределений и их визуализация с помощью гистограмм.

Промежуточная аттестация по курсу проводится в форме итоговой контрольной работы.

Тематика вопросов для промежуточной контрольной работы:

Формируемые компетенции и индикаторы их достижения: ОПК-1 (ОПК-1.1, ОПК-1.2) и ПК-1 (ПК-1.2)

1. Роль Python в экосистеме цифрового социолога.
2. Продвинутое структурирование данных в Python: генераторы, итераторы, comprehensions.
3. Функциональное программирование: map, filter, reduce, lambda-функции.
4. Основы объектно-ориентированного программирования в Python.
5. Работа с файлами различных форматов (CSV, JSON, XML, Excel) в Python.
6. Библиотека NumPy: создание и индексирование массивов.
7. Универсальные и агрегирующие функции NumPy.
8. Библиотека Pandas: структуры Series и DataFrame.
9. Загрузка данных в Pandas из различных источников.
10. Первичный анализ данных в Pandas: info, describe, head, tail.
11. Очистка данных: работа с пропущенными значениями.
12. Очистка данных: поиск и удаление дубликатов, преобразование типов.
13. Фильтрация и выборка данных в Pandas (булева индексация, loc, iloc).
14. Применение функций к данным: apply, map, applymap.
15. Группировка данных с groupby и агрегация (agg).
16. Анализ частот с value_counts и таблицы сопряженности crosstab.
17. Работа с категориальными переменными в Pandas.

18. Работа с временными рядами в Pandas: DatetimeIndex, resample.
19. Скользящие окна и сглаживание временных рядов (rolling).
20. Объединение данных: merge, concat, join.
21. Визуализация с Matplotlib: основные типы графиков и настройка параметров.
22. Статистическая визуализация с Seaborn: violinplot, boxplot, lmpplot.
23. Тепловые карты (heatmap) для визуализации корреляций.
24. Понятие исследовательского анализа данных (EDA) и его этапы.
25. Выявление выбросов и их обработка.
26. Анализ распределений и проверка на нормальность.
27. Корреляционный анализ и его интерпретация.
28. Создание многопанельных графиков (subplots).
29. Экспорт результатов анализа и графиков для отчетов.
30. Оформление EDA-отчета в Jupyter Notebook.

Тематика вопросов для итоговой контрольной работы:

Формируемые компетенции и индикаторы их достижения: ОПК-1 (ОПК-1.1, ОПК-1.2) и ПК-1 (ПК-1.2).

1. Основные парадигмы машинного обучения: обучение с учителем, без учителя, с подкреплением.
2. Этапы построения модели машинного обучения.
3. Предобработка данных для ML: масштабирование, кодирование категориальных признаков.
4. Проблема переобучения и способы борьбы с ней. Кросс-валидация.
5. Линейная регрессия: принцип работы, интерпретация, метрики качества.
6. Логистическая регрессия для бинарной классификации. Метрики классификации.
7. Деревья решений: принцип работы, визуализация, интерпретация.
8. Случайный лес: ансамблевый метод, оценка важности признаков.
9. Градиентный бустинг (XGBoost, LightGBM, CatBoost): принципы и преимущества.
10. Методы кластеризации: k-means, выбор числа кластеров.
11. Иерархическая кластеризация и дендрограммы.
12. Метод главных компонент (PCA): математические основы и интерпретация.
13. t-SNE и UMAP для визуализации многомерных данных.
14. Предобработка текстов для NLP: токенизация, лемматизация, стемминг.
15. Векторизация текстов: мешок слов, TF-IDF.
16. Тематическое моделирование (LDA): принципы и применение.

17. Анализ тональности текстов: подходы и инструменты.
18. Основы нейронных сетей: перцептрон, функции активации, обратное распространение ошибки.
19. Библиотеки TensorFlow и Keras для построения нейросетей.
20. Сверточные нейронные сети (CNN) и их применение.
21. Рекуррентные нейронные сети (RNN, LSTM) для анализа последовательностей.
22. Архитектура трансформера и большие языковые модели (LLM).
23. Промпт-инжиниринг для социологических задач.
24. Анализ социальных сетей (SNA): основные понятия и метрики (NetworkX).
25. Графовые нейронные сети (GNN): принципы и применение в социологии.
26. Агент-ориентированное моделирование (Mesa): концепции и примеры.
27. Веб-скрапинг динамических страниц (Selenium, Playwright).
28. Работа с API социальных сетей и открытых платформ.
29. Этические проблемы применения ИИ: алгоритмическая предвзятость, приватность, дискриминация.
30. Методы объяснимого ИИ (XAI): LIME, SHAP.

Контрольные вопросы для экзамена

1. Роль и место программирования на Python в современной цифровой социологии. Концепция AI Society.
2. Библиотека Pandas: структуры данных, загрузка, очистка и трансформация данных. Применение в социологических исследованиях.
3. Разведочный анализ данных (EDA): цели, методы, инструменты визуализации (Matplotlib, Seaborn).
4. Группировка и агрегация данных в Pandas. Анализ различий между социальными группами.
5. Введение в машинное обучение: основные парадигмы, типы задач, этапы построения моделей.
6. Методы классификации в социологических исследованиях: логистическая регрессия, деревья решений, случайный лес. Метрики качества.
7. Методы кластеризации для сегментации социальных групп: k-means, иерархическая кластеризация, DBSCAN.
8. Снижение размерности данных: PCA, t-SNE, UMAP. Визуализация социальных пространств.

9. Обработка естественного языка (NLP) в социологии: предобработка текстов, векторизация, тематическое моделирование (LDA).
10. Анализ тональности текстов и его применение для изучения общественных настроений.
11. Введение в нейронные сети: архитектура, функции активации, обучение. Библиотеки TensorFlow/Keras.
12. Продвинутое архитектуры нейронных сетей: CNN, RNN/LSTM, трансформеры. Области применения в социальных исследованиях.
13. Большие языковые модели (LLM) и промпт-инжиниринг в задачах анализа социальных данных.
14. Анализ социальных сетей (SNA) с Python: основные метрики, выявление сообществ, визуализация.
15. Графовые нейронные сети (GNN) для моделирования социальных взаимодействий и прогнозирования влияния.
16. Агент-ориентированное моделирование (ABM) социальных процессов на Python (Mesa).
17. Сбор данных из цифровых источников: веб-скрапинг (BeautifulSoup, Selenium) и работа с API.
18. Этические проблемы применения ИИ в социальных исследованиях: алгоритмическая предвзятость, проблема «черного ящика», приватность данных.
19. Методы объяснимого ИИ (XAI): LIME, SHAP для интерпретации моделей.
20. Полный цикл анализа данных в социологическом проекте: от постановки задачи и сбора данных до презентации результатов.

6. Учебно-методическое и информационное обеспечение дисциплины

6.1 Список источников и литературы

Основная литература

1. **Маккинни, У.** Python и анализ данных. — 3-е изд. — М.: ДМК Пресс, 2023. — 536 с.
2. **Вандер Плас, Дж.** Python для сложных задач: наука о данных и машинное обучение. — СПб.: Питер, 2021. — 576 с.
3. **Жерон, О.** Прикладное машинное обучение с помощью Scikit-Learn, Keras и TensorFlow. — 3-е изд. — М.: Вильямс, 2023. — 848 с.
4. **Васильев, Ю.** Python для data science. — СПб.: Питер, 2023. — 416 с.
5. **Любанович, Б.** Простой Python. Современный стиль программирования. — 3-е изд. — СПб.: Питер, 2023. — 592 с.

Дополнительная литература

1. **Zhang, W., Zinoviev, D.** Introduction to Quantitative Social Science with Python. — CRC Press, 2025. — 352 p. (Специализированное издание для социальных наук)
2. **Chen, D. Y.** Pandas for Everyone: Python Data Analysis. — 2nd ed. — Addison-Wesley, 2023. — 400 p.
3. **Титов, А. Н., Тазиева, Р. Ф.** Обработка данных в Python. Основы работы с библиотекой Pandas. — Казань: КНИТУ, 2022. — 116 с.
4. **VanderPlas, J.** Python Data Science Handbook. — 2nd ed. — O'Reilly Media, 2023. — 600 p.
5. **Géron, A.** Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow. — 3rd ed. — O'Reilly Media, 2022. — 850 p.
6. **Müller, A. C., Guido, S.** Introduction to Machine Learning with Python. — 2nd ed. — O'Reilly Media, 2021. — 400 p.
7. **Chollet, F.** Deep Learning with Python. — 2nd ed. — Manning, 2021. — 504 p.
8. **Bird, S., Klein, E., Loper, E.** Natural Language Processing with Python. — Updated ed. — O'Reilly Media, 2024. — 650 p. (Актуальное издание по NLTK)
9. **Jurafsky, D., Martin, J. H.** Speech and Language Processing. — 3rd ed. — Pearson, 2024. — 1088 p. (Фундаментальный учебник по NLP)
10. **Leitgöb, H., Prandner, D., Wolbring, T.** Big data and machine learning in sociology // *Frontiers in Sociology*. — 2023. — Т. 8. — С. 1173155.
11. **Burrell, J., Fourcade, M.** The society of algorithms // *Annual review of sociology*. — 2021. — Т. 47. — P. 213–237.
12. **Чернавский, А. С.** От Big Data к AI Society: новая парадигма социального прогнозирования в эпоху нейросетевого моделирования // *Петербургская социология сегодня*. — 2025. — № 28. — С. 14-37.
13. **Xu, H., Gao, C., Li, X., Wang, Z.** SIGRL: Sociologically-Informed Graph Representation Learning for Social Influence Prediction // *IEEE Transactions on Network Science and Engineering*. — 2025. — Vol. 12. — P. 4164-4181.
14. **Okawa, M., Iwata, T.** Predicting Opinion Dynamics via Sociologically-Informed Neural Networks // *Proceedings of the ACM Web Conference*. — 2022.
15. **Metzler, H., Garcia, D.** Social drivers and algorithmic mechanisms on digital media // *Perspectives on Psychological Science*. — 2024. — Т. 19. — №. 5. — P. 735–748.
16. **Liu, Z., Song, T.** Big Data Analysis and User Behavior Prediction of Social Networks Based on Artificial Neural Network // *Journal of computing and information technology*. — 2023. — Т. 31. — №. 3. — P. 185–201.
17. **Airlangga, G., Liu, A.** A Hybrid Gradient Boosting and Neural Network Model for Predicting Urban Happiness // *Machine Learning and Knowledge Extraction*. — 2025. — Т. 7. — №. 1. — С. 4.
18. **Brevini, B., Pasquale, F.** Revisiting the Black Box Society by rethinking the political economy of big data // *Big Data & Society*. — 2020. — Т. 7. — №. 2.
19. **Contini, F., Minissale, A., Bergman Blix, S.** Artificial intelligence and real decisions: predictive systems and generative AI vs. emotive-cognitive legal deliberations // *Frontiers in Sociology*. — 2024. — Т. 9. — С. 1417766.
20. **Kawamleh, S.** Algorithmic evidence in US criminal sentencing // *AI and Ethics*. — 2024. — С. 1-14.
21. **Жуков, Р. А.** Язык программирования Python. Практикум. — М.: НИЦ ИНФРА-М, 2023. — 216 с.
22. **Иванцова, О. Н., Мирошниченко, И. Д.** Практикум по дисциплине «Языки программирования (Python)». Часть 1. — СПб.: СПбГУ, 2024. — 120 с.

23. **Larooij, M., Törnberg, P.** Do Large Language Models Solve the Problems of Agent-Based Modeling? A Critical Review of Generative Social Simulations // arXiv preprint arXiv:2504.03274. – 2025.
24. **Chung, M., Wihbey, J.** The algorithmic knowledge gap within and between countries: Implications for combatting misinformation // Harvard Kennedy School Misinformation Review. – 2024.
25. **Dwivedi, Y. K. et al.** Evolution of artificial intelligence research in Technological Forecasting and Social Change // Technological Forecasting and Social Change. – 2023. – Т. 192. – С. 122579.
26. **Li, Q., King, B. G., Uzzi, B.** Quantifying social media predictors of violence during the 6 January US Capitol insurrection using Granger causality // Journal of the Royal Society Interface. – 2024. – Т. 21. – №. 220.
27. **Bradshaw, S., Bailey, H., Howard, P. N.** Industrialized disinformation: 2020 global inventory of organized social media manipulation. – Oxford Internet Institute, 2021.
28. **Мозалевская, Д. А.** Акторно-сетевая теория как методология современных социальных исследований // Журнал Белорусского государственного университета. Философия. Психология. – 2022. – №. 3. – С. 67–75.
29. **Малахова, Е. В.** Проблема аутопойезиса техногенной цивилизации и формирование ценностных основ применения цифровых технологий // Философские науки. – 2022. – Т. 65. – №. 1. – С. 109-123.
30. **Klein, B.** Python for Social Scientists: A Practical Guide. – University of Edinburgh, 2025.

При изучении дисциплины «Программирование и обработка данных в социологических исследованиях» студенту следует использовать монографии, учебники, учебные пособия, справочники (в том числе энциклопедические), а также *новейшие* научные публикации в *научных изданиях* (материалах научных и практических конференций, тематических сборниках и т.д.), в том числе периодических (журналах). Для поиска необходимых материалов помимо фондов научной библиотеки РГГУ и других библиотек целесообразно обращаться к ресурсам интернета.

6.2. Перечень ресурсов информационно-телекоммуникационной сети «Интернет»

Профессиональные базы данных и информационно-справочные системы

- ~ Национальная электронная библиотека (НЭБ) www.rusneb.ru
- ~ ELibrary.ru Научная электронная библиотека www.elibrary.ru
- ~ Электронная библиотека Grebennikon.ru www.grebennikon.ru
- ~ Cambridge University Press
- ~ ProQuest Dissertation & Theses Global
- ~ SAGE Journals
- ~ Taylor and Francis
- ~ JSTOR

Иные ресурсы, необходимые для усвоения дисциплины

1. Официальная документация Python: <https://docs.python.org/3/>
2. Официальная документация Pandas: <https://pandas.pydata.org/docs/>
3. Официальная документация Scikit-learn: <https://scikit-learn.org/stable/documentation.html>
4. Официальная документация TensorFlow: https://www.tensorflow.org/api_docs
5. Официальная документация PyTorch: <https://pytorch.org/docs/>
6. Официальная документация NetworkX: <https://networkx.org/documentation/>
7. Официальная документация Mesa (ABM): <https://mesa.readthedocs.io/>
8. Stepik. Курс «Анализ данных на Python»: <https://stepik.org/course/724/promo>
9. Stepik. Курс «Машинное обучение»: <https://stepik.org/course/8057/promo>
10. Kaggle Learn (обучающие курсы и соревнования): <https://www.kaggle.com/learn>
11. Единый архив экономических и социологических данных (ЕАЭСД) НИУ ВШЭ: <http://sophist.hse.ru/>
12. ВЦИОМ. Данные опросов: <https://wciom.ru/database/>
13. ФОМ. База данных опросов: <https://fom.ru/>
14. Росстат. Базы данных: <https://rosstat.gov.ru/>
15. GitHub (репозитории с кодом для анализа данных): <https://github.com/>

Профессиональные базы данных и информационно-справочные системы

Доступ к профессиональным базам данных: <https://liber.rsuh.ru/ru/bases>

Информационные справочные системы:

1. Консультант Плюс
2. Гарант

7. Материально-техническое обеспечение дисциплины

Для обеспечения дисциплины используется материально-техническая база образовательного учреждения: учебные аудитории, оснащённые компьютером и проектором для демонстрации учебных материалов.

Состав программного обеспечения:

1. Windows
2. Microsoft Office

Профессиональные полнотекстовые базы данных:

1. Национальная электронная библиотека (НЭБ) www.rusneb.ru
2. ELibrary.ru Научная электронная библиотека www.elibrary.ru
3. Электронная библиотека Grebennikon.ru www.grebennikon.ru
4. Cambridge University Press
5. ProQuest Dissertation & Theses Global
6. SAGE Journals
7. Taylor and Francis
8. JSTOR

8. Обеспечение образовательного процесса для лиц с ограниченными возможностями здоровья и инвалидов

В ходе реализации дисциплины используются следующие дополнительные методы обучения, текущего контроля успеваемости и промежуточной аттестации обучающихся в зависимости от их индивидуальных особенностей:

1. для слепых и слабовидящих:
 - ~ лекции оформляются в виде электронного документа, доступного с помощью компьютера со специализированным программным обеспечением;
 - ~ письменные задания выполняются на компьютере со специализированным программным обеспечением, или могут быть заменены устным ответом;
 - ~ обеспечивается индивидуальное равномерное освещение не менее 300 люкс;
 - ~ для выполнения задания при необходимости предоставляется увеличивающее устройство; возможно также использование собственных увеличивающих устройств;
 - ~ письменные задания оформляются увеличенным шрифтом;
 - ~ экзамен и зачёт проводятся в устной форме или выполняются в письменной форме на компьютере.
2. для глухих и слабослышащих:
 - ~ лекции оформляются в виде электронного документа, либо предоставляется звукоусиливающая аппаратура индивидуального пользования;
 - ~ письменные задания выполняются на компьютере в письменной форме;
 - ~ экзамен и зачёт проводятся в письменной форме на компьютере; возможно проведение в форме тестирования.
3. для лиц с нарушениями опорно-двигательного аппарата:
 - ~ лекции оформляются в виде электронного документа, доступного с помощью компьютера со специализированным программным обеспечением;
 - ~ письменные задания выполняются на компьютере со специализированным программным обеспечением;
 - ~ экзамен и зачёт проводятся в устной форме или выполняются в письменной форме на компьютере.

При необходимости предусматривается увеличение времени для подготовки ответа.

Процедура проведения промежуточной аттестации для обучающихся устанавливается с учётом их индивидуальных психофизических особенностей. Промежуточная аттестация может проводиться в несколько этапов.

При проведении процедуры оценивания результатов обучения предусматривается использование технических средств, необходимых в связи с индивидуальными особенностями обучающихся. Эти средства могут быть предоставлены университетом, или могут использоваться собственные технические средства.

Проведение процедуры оценивания результатов обучения допускается с использованием дистанционных образовательных технологий.

Обеспечивается доступ к информационным и библиографическим ресурсам в сети Интернет для каждого обучающегося в формах, адаптированных к ограничениям их здоровья и восприятия информации:

1. для слепых и слабовидящих:
 - ~ в печатной форме увеличенным шрифтом;
 - ~ в форме электронного документа;
 - ~ в форме аудиофайла.
2. для глухих и слабослышащих:
 - ~ в печатной форме;
 - ~ в форме электронного документа.
3. для обучающихся с нарушениями опорно-двигательного аппарата:
 - ~ в печатной форме;
 - ~ в форме электронного документа;
 - ~ в форме аудиофайла.

Учебные аудитории для всех видов контактной и самостоятельной работы, научная библиотека и иные помещения для обучения оснащены специальным оборудованием и учебными местами с техническими средствами обучения:

1. для слепых и слабовидящих:
 - ~ устройством для сканирования и чтения с камерой SARA CE;
 - ~ дисплеем Брайля PAC Mate 20;
 - ~ принтером Брайля EmBrailleViewPlus;
2. для глухих и слабослышащих:
 - ~ автоматизированным рабочим местом для людей с нарушением слуха и слабослышащих;
 - ~ акустический усилитель и колонки;
3. для обучающихся с нарушениями опорно-двигательного аппарата:
 - ~ передвижными, регулируемые эргономическими партами СИ-1;
 - ~ компьютерной техникой со специальным программным обеспечением

9. Методические материалы

9.1 Планы практических занятий

Семестр 1. Основы программирования и аналитики данных на Python

Тема 1. Настройка среды разработки и первый анализ.

Вопросы для обсуждения и задания: Установка Python, Jupyter Lab, PyCharm. Знакомство с интерфейсом. Создание первого ноутбука. Загрузка демо-датасета (например, опрос ВЦИОМ). Простейшие операции: выгрузка статистик, построение графиков.

Литература: [1, осн.], [5, осн.], [22, доп.], Интернет-ресурс [8].

Тема 2. Продвинутая работа со структурами данных.

Вопросы для обсуждения и задания: Генераторы списков для быстрой обработки. Использование множеств для поиска уникальных значений. Словари для хранения вложенных структур. Решение задач на оптимизацию кода.

Литература: [1, осн.], [5, осн.], [22, доп.].

Тема 3. Функциональное программирование на практике.

Вопросы для обсуждения и задания: Применение map и filter для трансформации данных. Использование lambda для краткой записи функций. Практикум: подсчет частот с использованием функционального подхода.

Литература: [1, осн.], [5, осн.], [22, доп.].

Тема 4. ООП для социолога.

Вопросы для обсуждения и задания: Создание класса Respondent с атрибутами и методами. Создание класса Survey для управления массивом респондентов. Добавление методов для фильтрации и агрегации.

Литература: [1, осн.], [5, осн.], [22, доп.].

Тема 5. Пакетная обработка файлов.

Вопросы для обсуждения и задания: Написание скрипта для обработки всех CSV-файлов в папке. Чтение, объединение, вычисление общих статистик. Сохранение объединенного файла.

Литература: [1, осн.], [4, доп.], [22, доп.].

Тема 6. NumPy: матричные операции.

Вопросы для обсуждения и задания: Создание матрицы "респонденты x вопросы". Вычисление средних по столбцам (вопросам) и по строкам (респондентам). Нормализация данных.

Литература: [2, осн.], [4, доп.], [1, доп.].

Тема 7. Pandas: загрузка и первичный анализ.

Вопросы для обсуждения и задания: Загрузка реального датасета (например, European Social Survey). Использование info, describe, head. Анализ типов данных. Поиск проблемных столбцов.

Литература: [1, осн.], [2, доп.], [3, доп.], Интернет-ресурс [11].

Тема 8. Очистка данных: работа с пропусками.

Вопросы для обсуждения и задания: Анализ пропусков в датасете. Принятие решений о стратегии обработки: удаление или заполнение. Заполнение пропусков средним, медианой, константой. Сравнение результатов.

Литература: [1, осн.], [2, доп.], [3, доп.].

Тема 9. Фильтрация и трансформация данных.

Вопросы для обсуждения и задания: Сложная фильтрация по нескольким условиям. Создание новых признаков (например, возрастные когорты, индексы). Применение функций для перекодировки ответов.

Литература: [1, осн.], [2, доп.], [3, доп.].

Тема 10. Группировка и агрегация.

Вопросы для обсуждения и задания: Анализ различий в доходах, образовании, удовлетворенности жизнью по полу, возрасту, региону. Построение сравнительных таблиц. Визуализация результатов группировки.

Литература: [1, осн.], [2, доп.], [3, доп.].

Тема 11. Анализ взаимосвязей категориальных переменных.

Вопросы для обсуждения и задания: Построение таблиц сопряженности для различных пар признаков. Расчет долей. Проверка гипотез о наличии связи. Визуализация с помощью тепловых карт.

Литература: [1, осн.], [2, доп.], [3, доп.].

Тема 12. Анализ временных рядов.

Вопросы для обсуждения и задания: Загрузка временного ряда (например, данные Росстата по безработице). Преобразование в формат дат. Построение графика. Выявление тренда и сезонности. Расчет скользящего среднего.

Литература: [2, осн.], [1, доп.], Интернет-ресурс [14].

Тема 13. Объединение данных из разных источников.

Вопросы для обсуждения и задания: Загрузка данных опроса и данных Росстата по регионам. Объединение по коду региона. Анализ зависимости ответов от региональных характеристик.

Литература: [1, осн.], [2, доп.], [3, доп.].

Тема 14. Визуализация с Matplotlib.

Вопросы для обсуждения и задания: Построение гистограмм распределения ключевых переменных. Ящики с усами для сравнения групп. Точечные диаграммы для выявления корреляций. Настройка графиков для публикации.

Литература: [2, осн.], [4, доп.], [1, доп.].

Тема 15. Визуализация с Seaborn.

Вопросы для обсуждения и задания: Построение violinplot для сравнения распределений. Матрица корреляций и тепловая карта. Парные графики (pairplot) для многомерного анализа. Экспорт графиков.

Литература: [2, осн.], [4, доп.], [1, доп.].

Тема 16. Защита EDA-проекта.

Вопросы для обсуждения и задания: Презентация результатов исследовательского анализа данных по выбранному датасету. Демонстрация кода, визуализаций, ключевых выводов. Ответы на вопросы.

Литература: Все источники семестра.

Семестр 2. Машинное обучение, нейросети и моделирование социальных процессов

Тема 17. Введение в Scikit-learn.

Вопросы для обсуждения и задания: Загрузка датасета. Разделение на обучающую и тестовую выборки. Построение простой модели (логистическая регрессия). Оценка ассигасы.

Литература: [3, осн.], [6, доп.], [7, доп.].

Тема 18. Предобработка данных для ML.

Вопросы для обсуждения и задания: Масштабирование признаков. One-Hot Encoding категориальных переменных. Создание пайплайна предобработки. Сравнение результатов до и после обработки.

Литература: [3, осн.], [6, доп.], [7, доп.].

Тема 19. Линейные модели.

Вопросы для обсуждения и задания: Построение модели линейной регрессии для прогнозирования (например, дохода). Интерпретация коэффициентов. Построение логистической регрессии для классификации. Анализ матрицы ошибок.

Литература: [3, осн.], [6, доп.], [7, доп.].

Тема 20. Деревья решений и случайный лес.

Вопросы для обсуждения и задания: Построение дерева решений, его визуализация. Анализ важности признаков. Построение случайного леса. Сравнение качества с логистической регрессией.

Литература: [3, осн.], [6, доп.], [7, доп.].

Тема 21. Кластеризация k-means.

Вопросы для обсуждения и задания: Подготовка данных для кластеризации. Определение оптимального числа кластеров (метод локтя). Интерпретация полученных кластеров. Визуализация кластеров в 2D.

Литература: [3, осн.], [6, доп.], [7, доп.].

Тема 22. Снижение размерности с PCA.

Вопросы для обсуждения и задания: Применение PCA к многомерному датасету. Анализ доли объясненной дисперсии. Визуализация данных в пространстве первых двух главных компонент. Интерпретация компонент.

Литература: [2, осн.], [3, осн.], [6, доп.].

Тема 23. Предобработка текстов для NLP.

Вопросы для обсуждения и задания: Загрузка корпуса текстов (например, открытые вопросы). Токенизация, удаление стоп-слов. Лемматизация/стемминг. Создание мешка слов и TF-IDF матрицы.

Литература: [8, доп.], [9, доп.], Интернет-ресурс [10].

Тема 24. Тематическое моделирование и тональность.

Вопросы для обсуждения и задания: Построение LDA-модели. Визуализация тем. Применение предобученной модели для анализа тональности. Интерпретация результатов.

Литература: [8, доп.], [9, доп.], Интернет-ресурс [10].

Тема 25. Нейронные сети на Keras.

Вопросы для обсуждения и задания: Создание простой полносвязной сети для классификации. Компиляция и обучение модели. Оценка качества на тестовой выборке. Визуализация процесса обучения.

Литература: [3, осн.], [7, доп.], [8, доп.].

Тема 26. Продвинутое моделирование нейросетей.

Вопросы для обсуждения и задания: Знакомство с CNN для обработки изображений (анализ фотографий). Знакомство с LSTM для анализа текстов (генерация/классификация). Эксперименты с готовыми архитектурами.

Литература: [3, осн.], [7, доп.], [8, доп.].

Тема 27. Промпт-инжиниринг с LLM.

Вопросы для обсуждения и задания: Работа с API большой языковой модели. Разработка промптов для классификации тональности, извлечения ключевых тем, суммаризации текстов. Сравнение результатов с классическими методами.

Литература: [13, доп.], [24, доп.], Интернет-ресурсы.

Тема 28. Агент-ориентированное моделирование (Mesa).

Вопросы для обсуждения и задания: Установка Mesa. Реализация модели Шеллинга (сегрегация). Запуск симуляций, визуализация, анализ результатов. Модификация параметров модели.

Литература: [7, доп.], Интернет-ресурс [7].

Тема 29. Анализ социальных сетей (NetworkX).

Вопросы для обсуждения и задания: Загрузка или создание графа социальных связей. Расчет метрик центральности. Выявление сообществ. Визуализация графа.

Литература: [13, доп.], [14, доп.], Интернет-ресурс [6].

Тема 30. Графовые нейронные сети (PyTorch Geometric).

Вопросы для обсуждения и задания: Установка PyTorch Geometric. Загрузка графового датасета. Создание простой GNN модели. Обучение и оценка. Анализ результатов.

Литература: [13, доп.], [14, доп.], Интернет-ресурс [5].

Тема 31. Этические аспекты: анализ предвзятости моделей.

Вопросы для обсуждения и задания: Анализ обученной модели на предмет смещений (bias). Использование LIME/SHAP для интерпретации предсказаний. Обсуждение этических дилемм в кейсах.

Литература: [10, доп.], [11, доп.], [18, доп.], [19, доп.].

Тема 32. Защита итогового проекта.

Вопросы для обсуждения и задания: Презентация комплексного исследовательского проекта, включающего сбор данных, предобработку, анализ (ML/нейросети) и интерпретацию результатов. Демонстрация кода, выводов, рекомендаций.

Литература: Все источники.

9.2. Методические указания

1. Освоение дисциплины строится по принципу «от данных — к знанию»: последовательно пройдите все этапы — от ввода и очистки информации до её многомерного анализа и визуализации.
2. На лекционных занятиях фиксируйте не только теоретические положения, но и конкретные названия инструментов, чтобы сразу применять их на практике.
3. Выполнение лабораторных работ обязательно должно сопровождаться самостоятельным экспериментированием: не бойтесь изменять параметры функций, чтобы понять логику их работы.
4. При подготовке к практическим занятиям заранее знакомьтесь с массивом данных, который предстоит обработать, — это позволит сосредоточиться на инструментари, а не на структуре программы.

5. Ведите личный глоссарий терминов и «шпаргалку» по синтаксису функций: Python насчитывает сотни функций, и мгновенное запоминание всех невозможно.
6. Для закрепления навыков автоматизации обязательно повторяйте запись макросов для однотипных операций, даже если задание этого прямо не требует.
7. В процессе выполнения контрольных работ особое внимание уделяйте интерпретации полученных цифр: умение объяснить результат ценится выше, чем просто правильный расчёт.
8. Итоговый проект должен демонстрировать владение полным циклом обработки данных — от импорта «сырого» файла до формулировки аналитического вывода, подкреплённого визуализацией.

АННОТАЦИЯ РАБОЧЕЙ ПРОГРАММЫ ДИСЦИПЛИНЫ

Дисциплина «Программирование и обработка данных в социологических исследованиях» реализуется на социологическом факультете кафедрой фундаментальной и прикладной математики.

Цель дисциплины: формирование у студентов целостного системного представления о возможностях языка программирования Python и современных технологий искусственного интеллекта (включая нейросетевые модели) как инструментальной базы для решения широкого круга профессиональных задач цифрового социолога, а также развитие устойчивых практических навыков автоматизированного сбора, обработки, анализа и визуализации социальных данных с применением методов машинного обучения.

1. Сформировать методологическую основу для понимания роли программирования и ИИ в трансформации современной социологии (концепция AI Society).
2. Систематизировать и углубить знания языка Python и библиотек (Pandas, NumPy) для обработки социологических данных.
3. Обучить методам автоматизированного сбора данных из цифровых источников (веб-скрапинг, API).
4. Развить навыки визуально-аналитического исследования данных и создания интерактивных дашбордов.
5. Сформировать компетенции в области применения методов машинного обучения (Scikit-learn) и нейросетевого моделирования (TensorFlow, Keras) для решения социологических задач.
6. Обучить применению больших языковых моделей (LLM) и промпт-инжиниринга в анализе текстов.
7. Сформировать компетенции в области моделирования социальных процессов (агент-ориентированные модели, графовые нейронные сети).
8. Воспитать культуру ответственного применения ИИ, понимание этических проблем (алгоритмическая предвзятость, прозрачность моделей).

Дисциплина направлена на формирование следующих компетенций студентов:

~ **ОПК-1.1** определяет релевантные для решения поставленной задачи источники информации, включая национальные и международные базы

данных, электронные библиотечные системы, специализированные пакеты прикладных программ;

~ **ОПК-1.2** предлагает постановку задач по поиску, обобщению социологических данных;

~ **ПК-1.2** разрабатывает предложения по совершенствованию технологических процессов, методов сбора и анализа информации в социологическом и маркетинговом исследовании.

В результате освоения дисциплины обучающийся должен:

Знать: современные методы сбора, обработки и анализа социальных данных с использованием Python; архитектуру и функциональные возможности библиотек для анализа данных и машинного обучения; принципы работы нейросетевых моделей и больших языковых моделей; этические и правовые аспекты применения ИИ в социальных исследованиях.

Уметь: разрабатывать на Python скрипты для автоматизированного сбора, очистки и анализа данных; применять методы машинного обучения для классификации, кластеризации и прогнозирования; использовать нейросетевые модели и LLM для обработки текстов; интерпретировать результаты анализа и представлять их в наглядной форме.

Владеть: навыками работы в среде Jupyter Notebook; методами построения аналитических пайплайнов; техниками промпт-инжиниринга; инструментами визуализации и создания дашбордов; навыками документирования и воспроизводимости исследований.

По дисциплине предусмотрена промежуточная аттестация в форме экзамена. Общая трудоемкость освоения дисциплины составляет 7 зачетных единиц.