

УТВЕРЖДАЮ
Первый проректор, проректор
по УР

_____ А.Е. Рудин

Рабочая программа дисциплины

Б1.В.03

Модели и методы обработки данных

Учебный план: 2026-2027 09.03.04 ИИТА Программ инженерии ОО №1-1-195.plx

Кафедра: **33** Цифровых и аддитивных технологий

Направление подготовки:
(специальность) 09.03.04 Программная инженерия

Профиль подготовки:
(специализация) Программная инженерия

Уровень образования: бакалавриат

Форма обучения: очная

План учебного процесса

Семестр (курс для ЗАО)		Контактная работа обучающихся		Сам. работа	Контроль, час.	Трудоёмкость, ЗЕТ	Форма промежуточной аттестации
		Лекции	Практ. занятия				
4	УП	17	34	102	27	5	Экзамен
	РПД	17	34	102	27	5	
Итого	УП	17	34	102	27	5	
	РПД	17	34	102	27	5	

Рабочая программа дисциплины составлена в соответствии с федеральным государственным образовательным стандартом высшего образования по направлению подготовки 09.03.04 Программная инженерия, утверждённым приказом Минобрнауки России от 19.09.2017 г. № 920

Составитель (и):

старший преподаватель

Кокорин Евгений Сергеевич

От кафедры составителя:

Заведующий кафедрой цифровых и аддитивных технологий

Сошников Антон
Владимирович

От выпускающей кафедры:

Заведующий кафедрой

Сошников Антон
Владимирович

Методический отдел:

1 ВВЕДЕНИЕ К РАБОЧЕЙ ПРОГРАММЕ ДИСЦИПЛИНЫ

1.1 Цель дисциплины: Сформировать у обучающегося базовые и прикладные компетенции в области обработки, анализа и представления данных, включая навыки работы с современными инструментами и технологиями подготовки и хранения данных, основами генерации, визуализации и распределённой обработки больших объёмов информации.

1.2 Задачи дисциплины:

Изучить этапы обработки данных: получение, очистку, подготовку и первичный анализ;
 Ознакомиться с основными типами данных и форматами их хранения (CSV, Excel, JSON и др.);
 Научиться использовать современные библиотеки Python (NumPy, Pandas) для обработки и анализа данных;
 Освоить приёмы визуализации данных с помощью специализированных инструментов (matplotlib, pandas);
 Сформировать навыки генерации и обогащения данных, в том числе для работы с временными рядами;
 Освоить основы оценки качества данных и применения основных метрик для анализа полноты и корректности информации;
 Получить базовые знания о технологиях распределённой обработки данных, включая MapReduce и Hadoop;
 Развить умения самостоятельной работы с инструментами для анализа и визуализации данных в средах Jupyter Notebook и Google Colab;
 Выработать навыки командной работы и презентации результатов анализа данных.

1.3 Требования к предварительной подготовке обучающегося:

Предварительная подготовка предполагает создание основы для формирования компетенций, указанных в п. 2, при изучении дисциплин:
 Алгоритмизация и программирование
 Учебная практика (научно-исследовательская работа (получение первичных навыков научно-исследовательской работы))
 Архитектура бизнес-системы
 Прикладная статистика

2 КОМПЕТЕНЦИИ ОБУЧАЮЩЕГОСЯ, ФОРМИРУЕМЫЕ В РЕЗУЛЬТАТЕ ОСВОЕНИЯ ДИСЦИПЛИНЫ

ПК-2: Способен на основе применения информационных технологий управлять информацией из различных источников	
Знать: Методологические подходы к разработке программного обеспечения для обработки данных; подходы к моделированию потоков и хранилищ информации.	
Уметь: Вырабатывать варианты реализации требований к компьютерному программному обеспечению в зависимости от типа данных и поставленных целей; строить информационные модели, отражающие сущности, связи и правила предметной области; анализировать полноту и непротиворечивость данных в контексте системных требований; проектировать структуры данных, поддерживающие функциональные и нефункциональные цели программного продукта.	
Владеть: Навыками проведения сравнительного анализа вариантов реализации требований к компьютерному программному обеспечению; инструментами концептуального моделирования данных; навыками трансформации бизнес-семантики в формализованные информационные структуры.	
ПК-6: Способен подготавливать данные для проведения аналитической работы с использованием программного продукта	
Знать: Теоретические основы и алгоритмические подходы к обработке структурированных и неструктурированных данных, включая методы фильтрации, нормализации, кластеризации и снижения размерности.	
Уметь: Проектировать и реализовывать конвейеры обработки данных с использованием современных библиотек и фреймворков для аналитических задач.	
Владеть: Навыками построения эффективных моделей трансформации данных, обеспечивающих их пригодность для последующего анализа и визуализации	

3 РЕЗУЛЬТАТЫ ОБУЧЕНИЯ ПО ДИСЦИПЛИНЕ

Наименование и содержание разделов, тем и учебных занятий	Семестр (курс для ЗАО)	Контактная работа		СР (часы)	Инновац. формы занятий	Форма текущего контроля
		Лек. (часы)	Пр. (часы)			
Раздел 1. Введение в обработку данных	4					О
Тема 1. Источники и категории данных. Понятие пайплайна обработки данных. Получение, очистка и подготовка данных. Введение в NumPy и Pandas. Практические занятия: Анализ источников, схема пайплайна. Загрузка, очистка, преобразование данных.		2	4	14		

Тема 2. Основные типы данных и форматы хранения. Первичный анализ данных и выявление проблем. Практические занятия: разбор CSV, Excel, JSON. Поиск пропусков, базовые статистики.		2	4	14		
Тема 3. Введение в инструменты Jupyter Notebook и Google Colab. Практические: Работа с блокнотами, импорт данных.		2	4	14		
Раздел 2. Основы работы с данными в Python						
Тема 4. Массивы и таблицы в NumPy и Pandas. Практические занятия: базовые операции с массивами, фильтрация		2	4	15		Пр
Тема 5. Агрегация и группировка данных. Практические занятия: группировка, сводные таблицы		2	6	15		
Раздел 3. Генерация и обогащение данных						
Тема 6. Генерация синтетических данных и аугментация. Обогащение данных: новые признаки, объединение таблиц. Практические занятия: Генерация выборок, добавление шума. Merge, join, feature engineering.		3	6	15		О,Пр
Тема 7. Основы работы с временными рядами. Визуализация временных рядов. Практические занятия: Разметка временных данных, выделение трендов. Построение графиков временных рядов		4	6	15		
Итого в семестре (на курсе для ЗАО)		17	34	102		
Консультации и промежуточная аттестация (Экзамен)		2,5		24,5		
Всего контактная работа и СР по дисциплине		53,5		126,5		

4 КУРСОВОЕ ПРОЕКТИРОВАНИЕ

Курсовое проектирование учебным планом не предусмотрено

5. ФОНД ОЦЕНОЧНЫХ СРЕДСТВ ДЛЯ ПРОВЕДЕНИЯ ПРОМЕЖУТОЧНОЙ АТТЕСТАЦИИ

5.1 Описание показателей, критериев и системы оценивания результатов обучения

5.1.1 Показатели оценивания

Код компетенции	Показатели оценивания результатов обучения	Наименование оценочного средства
ПК-2	Раскрывает список методологий для обработки данных. Формулирует требования к программному обеспечению для обработки данных для достижения поставленных целей. Представляет аналитическое заключение по анализу требований к программному обеспечению.	
ПК-6	Формулирует понятие большие данные и раскрывает классификацию типов больших данных. Раскрывает основные этапы работы с данными из широкого круга источников. Демонстрирует результат построения структуры для хранения данных.	

5.1.2 Система и критерии оценивания

Шкала оценивания	Критерии оценивания сформированности компетенций	
	Устное собеседование	Письменная работа

5 (отлично)	Полный, исчерпывающий ответ, явно демонстрирующий глубокое понимание предмета и широкую эрудицию в оцениваемой области. Критический, оригинальный подход к материалу. Учитываются баллы, накопленные в течение семестра.	Качество исполнения всех элементов задания полностью соответствует всем требованиям. Учитываются баллы, накопленные в течение семестра.
4 (хорошо)	Ответ стандартный, в целом качественный, основан на всех обязательных источниках информации. Присутствуют небольшие пробелы в знаниях или несущественные ошибки. Учитываются баллы, накопленные в течение семестра.	Работа выполнена в соответствии с заданием. Имеются отдельные несущественные ошибки или отступления от правил оформления работы. Учитываются баллы, накопленные в течение семестра.
3 (удовлетворительно)	Ответ неполный, основанный только на лекционных материалах. При понимании сущности предмета в целом – существенные ошибки или пробелы в знаниях сразу по нескольким темам, незнание (путаница) важных терминов. Учитываются баллы, накопленные в течение семестра.	Задание выполнено полностью, но с многочисленными существенными ошибками. При этом нарушены правила оформления или сроки представления работы. Учитываются баллы, накопленные в течение семестра.
2 (неудовлетворительно)	Неспособность ответить на вопрос без помощи экзаменатора. Незнание значительной части принципиально важных элементов дисциплины. Многочисленные грубые ошибки. Не учитываются баллы, накопленные в течение семестра.	Отсутствие одного или нескольких обязательных элементов задания, либо многочисленные грубые ошибки в работе, либо грубое нарушение правил оформления или сроков представления работы. Не учитываются баллы, накопленные в течение семестра.

5.2 Типовые контрольные задания или иные материалы, необходимые для оценки знаний, умений, навыков и (или) опыта деятельности

5.2.1 Перечень контрольных вопросов

№ п/п	Формулировки вопросов
Семестр 4	
1	Какие этапы включает в себя процесс очистки данных?
2	Перечислите основные источники и категории данных.
3	Что такое пайплайн обработки данных? Приведите пример.
4	Назовите распространённые форматы хранения данных.
5	Как загрузить данные из CSV-файла в Pandas DataFrame?
6	В чём особенности формата JSON по сравнению с Excel?
7	Назовите распространённые форматы хранения данных.
8	Как загрузить данные из CSV-файла в Pandas DataFrame?
9	В чём особенности формата JSON по сравнению с Excel?
10	Как осуществляется фильтрация данных в Pandas?
11	Какие функции Pandas позволяют группировать данные?
12	Как построить простой график по данным с помощью matplotlib?
13	В чём разница между Series и DataFrame в Pandas?
14	Как сохранить обработанные данные обратно в файл?
15	Какую функцию в Pandas использовать для объединения таблиц?
16	Какие инструменты используются для работы с ноутбуками?
17	Какие шаги включает первичный анализ данных?
18	Для чего используется Jupyter Notebook?
19	Как импортировать библиотеку Pandas в Python?
20	Каковы основные преимущества использования Google Colab?
21	Для чего используется генерация синтетических данных?
22	В каких случаях применяется аугментация данных?
23	Как создать новый признак (feature) в Pandas DataFrame?
24	В чём заключается объединение данных с помощью merge?

25	Какие задачи решаются при анализе временных рядов?
26	Как разметить временные данные для анализа?
27	Какие методы позволяют выявить тренды во временных рядах?
28	Как построить график временного ряда?
29	Почему важно проверять корректность и полноту данных?
30	Какие бывают метрики качества данных?
31	В чём отличие полноты и точности как метрик?
32	Как определить наличие дубликатов в наборе данных?
33	Как проверить распределение данных?
34	Какие графики используются для визуального контроля качества?
35	Чем полезна гистограмма для анализа данных?
36	Какие шаги включает в себя процесс валидации данных?
37	Что такое feature engineering?
38	Как осуществляется объединение нескольких DataFrame?
39	Почему важно анализировать пропуски во временных рядах?
40	Какие ошибки могут возникать при работе с временными рядами?
41	Объясните принцип работы MapReduce.
42	В чём преимущества распределённой обработки данных?
43	Какова основная задача этапа Map в MapReduce?
44	Для чего используется этап Reduce?
45	Что такое кластер Hadoop?
46	Как устроена распределённая файловая система HDFS?
47	Какие задачи решаются с помощью Hadoop?
48	Какие форматы данных поддерживаются в Hadoop?
49	Как смоделировать работу MapReduce на простом примере?
50	Какие существуют типовые ошибки при работе с большими данными?
51	Как организовать хранение больших массивов данных?
52	Для чего нужны распределённые вычисления?
53	Какие существуют типовые ошибки при работе с большими данными?
54	Как организовать хранение больших массивов данных?
55	Для чего нужны распределённые вычисления?
56	Как анализировать результат выполнения MapReduce?
57	Какие существуют способы оптимизации производительности при работе с большими данными?
58	В чём разница между централизованной и распределённой обработкой?
59	Какова роль мастер-узла в кластере Hadoop?
60	Какие программные инструменты используются для распределённого анализа данных?
61	Как осуществляется мониторинг и отладка процессов в распределённой системе?

5.2.2 Типовые тестовые задания

Не предусмотрено

5.2.3 Типовые практико-ориентированные задания (задачи, кейсы)

Собрать и загрузить набор данных из открытого источника, выполнить его очистку и первичный анализ (выявить пропуски, удалить дубликаты).

Провести агрегацию и группировку данных с помощью библиотеки Pandas, представить результаты в виде таблицы и диаграммы.

Выполнить визуализацию набора данных с помощью matplotlib и Pandas: построить графики, гистограммы и круговые диаграммы.

Сгенерировать синтетические данные для учебной задачи, обогатить их новыми признаками и сохранить результат в файл.

Провести анализ и визуализацию временного ряда: выделить тренды, построить соответствующие графики.

Разработать пайплайн обработки данных в Jupyter Notebook или Google Colab для решения прикладной задачи (например, анализ результатов опроса).

Оценить качество данных, рассчитать основные метрики (полнота, точность, количество дубликатов), оформить результаты анализа в виде отчёта.

Смоделировать работу алгоритма MapReduce для обработки большого объёма данных на примере учебного датасета.

Выполнить интеграцию и объединение данных из двух разных источников, выявить и устранить несоответствия.

Подготовить презентацию (или мини-отчёт) с визуализацией результатов анализа данных и выводами.

5.3 Методические материалы, определяющие процедуры оценивания знаний, умений, владений (навыков и (или) практического опыта деятельности)

5.3.1 Условия допуска обучающегося к промежуточной аттестации и порядок ликвидации академической задолженности

Проведение промежуточной аттестации регламентировано локальным нормативным актом СПбГУПТД «Положение о проведении текущего контроля успеваемости и промежуточной аттестации обучающихся»

5.3.2 Форма проведения промежуточной аттестации по дисциплине

Устная ☒ Письменная ☐ Компьютерное тестирование ☐ Иная ☐

5.3.3 Особенности проведения промежуточной аттестации по дисциплине

экзамена

- время на подготовку к устному собеседованию составляет 30 минут;
- выполнение кейс-задания осуществляется на компьютере за 60 минут.

6. УЧЕБНО-МЕТОДИЧЕСКОЕ И ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

6.1 Учебная литература

Автор	Заглавие	Издательство	Год издания	Ссылка
6.1.1 Основная учебная литература				
Бобонова, Е. Н.	Методы математической обработки данных	Москва, Вологда: Инфра-Инженерия	2024	https://www.iprbookshop.ru/143529.html
Глотова, М. Ю., Самохвалова, Е. А.	ИКТ и математические методы обработки данных	Москва: Московский педагогический государственный университет	2024	https://www.iprbookshop.ru/145465.html
Курносков, М. Г.	Введение в методы машинной обработки данных	Новосибирск: Автограф	2020	http://www.iprbookshop.ru/102117.html
Родионова, Т. Е.	Информационные технологии обработки данных	Ульяновск: Ульяновский государственный технический университет	2020	http://www.iprbookshop.ru/106094.html
6.1.2 Дополнительная учебная литература				
Самуйлов, С. В.	Алгоритмы и структуры обработки данных	Саратов: Вузовское образование	2016	https://www.iprbookshop.ru/47275.html
Волков, М. М.	Практикум по дисциплине Структуры и алгоритмы обработки данных	Москва: Московский технический университет связи и информатики	2016	https://www.iprbookshop.ru/61551.html
Медведев, Д. М.	Структуры и алгоритмы обработки данных в системах автоматизации и управления	Саратов: Ай Пи Эр Медиа	2018	https://www.iprbookshop.ru/71591.html

6.2 Перечень профессиональных баз данных и информационно-справочных систем

Информационная система «Единое окно доступа к образовательным ресурсам. Раздел. Информатика и информационные технологии» [Электронный ресурс]. URL: http://window.edu.ru/catalog/?p_rubr=2.2.75.6

Электронно-библиотечная система IPRbooks [Электронный ресурс]. URL: <http://www.iprbookshop.ru/>

6.3 Перечень лицензионного и свободно распространяемого программного обеспечения

MicrosoftOfficeProfessional

Microsoft Windows

Adobe Audition CC ALL Multiple Platforms Multi European Languages Team LicSub Level 4 (100+) Education

Device license

Figma

6.4 Описание материально-технической базы, необходимой для осуществления образовательного процесса по дисциплине

Аудитория	Оснащение
Компьютерный класс	Мультимедийное оборудование, компьютерная техника с возможностью подключения к сети «Интернет» и обеспечением доступа в электронную информационно-образовательную среду
Лекционная аудитория	Мультимедийное оборудование, специализированная мебель, доска