

Федеральное государственное бюджетное образовательное учреждение высшего образования
«Алтайский государственный технический университет им. И.И. Ползунова»

СОГЛАСОВАНО

Директор УТК И.А.
Бахтина

Рабочая программа дисциплины

Код и наименование дисциплины: **ПМ.2.МДК.1 «Инжиниринг данных»**

Код и наименование направления подготовки (специальности): **09.02.12**

Техническая эксплуатация и сопровождение информационных систем

Квалификация: **Специалист по технической эксплуатации и сопровождению информационных систем**

Статус дисциплины: **обязательная**

Форма обучения: **очная**

Статус	Должность	И.О. Фамилия
Разработал	профессор преподаватель преподаватель почасовик	Н.Н. Барышева О.А. Лыскова Я.Ю. Музоватова
Согласовал	Зав. кафедрой «ИСЭ»	А.С. Авдеев
	руководитель образовательной программы	Н.Н. Барышева

г. Барнаул

1. Перечень планируемых результатов обучения по дисциплине, соотнесенных с планируемыми результатами освоения образовательной программы

Код компетенции из УП	Содержание компетенции	В результате изучения дисциплины обучающиеся должны:		
		знать	уметь	иметь практический опыт
ПК 2.1	Выполнять подготовку данных для проведения аналитических работ	Предметную область анализа Теоретические и прикладные основы анализа больших данных	Использовать инструментальные средства для извлечения, преобразования, хранения и обработки данных из разнородных источников, в том числе в режиме реального времени	Получения и фильтрации больших объемов данных из гетерогенных источников; Извлечения, проверки и очистки больших объемов данных из гетерогенных источников
ПК 2.2	Строить статистические и математические модели для систем анализа данных	Основы теории вероятностей и ее применения в статистике. Матрицы, векторы и операции с ними.	Обрабатывать пропущенные данные и аномалии. Представлять данные и результаты анализа с помощью графиков и диаграмм. Самостоятельно изучать новые инструменты и технологии.	Понимания основных статистических понятий. Умения проводить описательную и инференциальную статистику. Опыта работы с языками программирования. Знания библиотек для анализа данных.

2. Место дисциплины в структуре образовательной программы

Дисциплины (практики), предшествующие изучению дисциплины, результаты освоения которых необходимы для освоения данной дисциплины.	Основы алгоритмизации и программирования
Дисциплины (практики), для которых результаты освоения данной дисциплины будут необходимы, как входные знания, умения и владения для их изучения.	Искусственный интеллект для бизнеса, Конфигурирование аналитических решений, Прикладные решения на базе 1С

3. Объем дисциплины в акад. часах

Общий объем дисциплины в час: 52

Форма промежуточной аттестации: Зачет

Форма обучения	Виды занятий, их трудоемкость (час.)								
	Лекции	Лабораторные работы	Практические занятия	Уроки	В т.ч. в форме практической подготовки	Консультации	Семинары	Курсовое проектирование	Самостоятельная работа
очная	16	32	0	0	32	2	0	0	2

4. Содержание дисциплины, структурированное по темам (разделам) с указанием отведенного на них количества академических часов и видов учебных занятий

Форма обучения: очная

Семестр: 4

Лекционные занятия (16ч.)

1. Знакомство с Jupyter Notebook и Google Colab. Основы Python. {лекция с разбором конкретных ситуаций} (2ч.)[3,8] Обзор процесса анализа данных. Введение в концепции программирования на Python. Установка и настройка Python и необходимых библиотек. Переменные и типы данных

Язык Python. Переменные. Вывод объектов и данных на экран. Обработка ошибок, оператор try-except. Типы данных. Преобразования типов данных.

Индексы в строках. Срезы строк. Операции над строками. Методы строк. Форматирование строк, метод format(), f-строки.

Индексы в списках. Срезы списков. Добавление и удаление элементов. Сложение и умножение, сортировка списков. Поиск элементов в списке. Разделение строки в список строк, соединение списка строк.

Цикл For. Перебор элементов. Перебор индексов элементов. Обработка элементов списков с помощью циклов: нахождение суммы и произведения элементов.

Вложенные списки. Циклы по вложенным спискам с подсчётом значений.

Добавление элементов во вложенные списки. Сортировка вложенных списков. Условный оператор. Цикл While. Логический тип данных. Булевы значения. Логические и составные логические выражения. Условный оператор if-elif-else. Ветвления. Фильтрация списков с использованием условного оператора.

Функции. Назначение функций. Параметры и аргументы. Параметры со значениями по умолчанию. Позиционные и именованные аргументы. Возвращение результата из функции.

Словари. Ключи и значения. Поиск значения по ключу. Добавление элементов в словарь. Список словарей.

2. Библиотека Pandas. Библиотека NumPy. {лекция с разбором конкретных ситуаций} (2ч.)[1,3,5,8] Библиотека Pandas. Чтение CSV-файлов. Датафрейм. Конструктор датафрейма. Вывод первых и последних строк датафрейма.

Индексация в датафреймах. Индексация в столбцах Series.

Библиотека NumPy. Работа с векторами, матрицами и другими многомерными массивами в Python. Функции и методы для анализа числовых данных высокой размерности.

3. Импорт и экспорт данных {лекция с разбором конкретных ситуаций} (2ч.)[1,2,3,4,5,6,7,8] Импорт и экспорт данных на Python. Обработка недостающих данных. Методы очистки и предварительной обработки данных. Данные различных открытых форматов: XML, JSON, CSV.

Чтение текстовых файлов. Индексы, мультииндекс. Операции с колонками и строками

Навигация по таблице. Работа с несколькими таблицами. Чтение-запись таблиц в файл, формат хранения данных .XML, JSON, CSV.

4. Статистический анализ. Описательная статистика. Распределения. Проверка гипотез. Теория вероятностей. {лекция с разбором конкретных ситуаций} (4ч.)[1,2,3,4,5,6,7,8] 1. Описательная статистика. Категориальные и количественные переменные. Мода и медиана. Среднее значение. Дисперсия. Стандартное отклонение. Квартили и процентили. Диаграмма размаха. Столбчатая диаграмма. Плотность частоты. Гистограмма.

2. Случайные величины. Случайная и дискретная случайная величина. Распределение вероятностей для дискретной случайной величины. Кумулятивная функция, математическое ожидание и дисперсия дискретной случайной величины.

3. Распределения. Эксперимент Бернулли. Биномиальный эксперимент. Распределения: непрерывное равно-мерное, нормальное и стандартное нормальное. CDF, PPF для нормального распределения. Распределение Пуассона. Апроксимация одного распределения другим.

4. Проверка гипотез

Генеральная совокупность. Выборка. Выборочное распределение. Центральная предельная теорема. Односторонние и двусторонние гипотезы. P-Value. Проверка гипотезы о равенстве средних двух генеральных совокупностей.

5. Статистический анализ данных. Теория вероятностей. Эксперимент. Вероятностное пространство. События. Вероятность. Пересекающиеся и взаимоисключающие события. Диаграмма Эйлера–Венна. Закон больших чисел.

5. Обработка и визуализация данных. Основы библиотеки Pandas. {лекция с разбором конкретных ситуаций} (4ч.)[1,2,3,4,5,6,7,8] Основы библиотеки Pandas, обзор и типы данных, работа с датой и временем, индексация в датафреймах, фильтрация данных, работа с пропущенными значениями и дубликатами, категоризация данных

Группировка данных. Сортировка данных. Основы описательной статистики. Оформление результатов.

Первые графики и выводы. Применение сводных таблиц. Гистограмма. Распределения.

Диаграмма размаха.

Изучение срезов данных. Метод query(). Работа с датой и временем. Построение

графиков методом `plot()`. Бритва Оккама.

- Работа с несколькими источниками данных

Срез данных на основе внешних объектов. Добавление новых столбцов в датафрейм. Добавление данных из других датафреймов. Переименование столбцов. Объединение таблиц.

Взаимосвязь данных. Диаграмма рассеяния. Корреляция переменных. Матрица диаграмм рассеяния.

Валидация результатов. Укрупнение групп. Разбиение данных по группам.

Линейные графики, точечные графики и столбчатые диаграммы. Настройка визуализаций с помощью `Matplotlib`. Введение в библиотеку `Plotly`. Создание интерактивных графиков и информационных панелей. Встраивание визуальных элементов в записные книжки `Jupyter`. Библиотека `Seaborn` для статистической визуализации. Построение графиков категориальных данных.

6. Визуализация данных с помощью DataLens, создание дашбордов. {лекция с разбором конкретных ситуаций} (2ч.)[1,2,3,4,5,6,7,8] Визуализация данных с помощью `DataLens`, создание дашбордов.

Интерфейс, подключения, датасеты, основы визуализации, чарты в `DataLens`, вычисляемые поля, создание дашбордов, параметры.

Консультации (2ч.)

1. Консультация(2ч.)[1,2,3,4,5,6,7,8]

Лабораторные работы (32ч.)

1. Лабораторная работа на тему: "Основы Python".(практическая подготовка

- 6ч.)[1,2,3,4,5,6,7,8] Знакомство с `Jupyter Notebook` и `Google Colab`. Основы Python. Решение задач на темы:

1. Переменные и типы данных Язык Python. Переменные. Вывод объектов и данных на экран. Обработка ошибок, оператор `try-except`. Типы данных. Преобразования типов данных.

2. Индексы в строках. Срезы строк. Операции над строками. Методы строк. Форматирование строк, метод `format()`, f-строки.

3. Индексы в списках. Срезы списков. Добавление и удаление элементов. Сложение и умножение, сортировка списков. Поиск элементов в списке. Разделение строки в список строк, соединение списка строк.

4. Цикл `For`. Перебор элементов. Перебор индексов элементов. Обработка элементов списков с помощью циклов: нахождение суммы и произведения элементов. Вложенные списки. Циклы по вложенным спискам с подсчётом значений. Добавление элементов во вложенные списки. Сортировка вложенных списков.

5. Условный оператор. Цикл `While`. Логический тип данных. Булевы значения. Логические и составные логические выражения. Условный оператор `if-elif-else`. Ветвления. Фильтрация списков с использованием условного оператора.

6. Функции. Назначение функций. Параметры и аргументы. Параметры со значениями по умолчанию. Позиционные и именованные аргументы. Возвращение результата из функции. Словари. Ключи и значения. Поиск значения по ключу. Добавление элементов в словарь. Список словарей.

2. Лабораторная работа на тему: Библиотека Pandas. Библиотека NumPy.(практическая подготовка - 6ч.)[1,2,3,4,5,6,7,8] Решение практических задач с использованием библиотеки Pandas и NumPy. Импорт и экспорт данных на Python. Работа с таблицами.

3. Лабораторная работа на тему "Статистический анализ. Теория вероятностей. Описательная статистика. Распределения. Проверка гипотез".(практическая подготовка - 8ч.)[1,2,3,4,5,6,7,8]

5. Обработка и визуализация данных. Основы библиотеки Pandas.(практическая подготовка - 6ч.)[1,2,3,4,5,6,7,8] Обработка и визуализация данных. Основы библиотеки Pandas. Решение практических задач.

6. Лабораторная работа на тему "Визуализация данных с помощью DataLens, создание дашбордов."(практическая подготовка - 6ч.)[2,4]

Самостоятельная работа (2ч.)

1. Подготовка к защите лабораторных работ и промежуточной аттестации(1ч.)[1,2,3,4,5,6,7,8]

2. Подготовка к экзамену.(1ч.)[1,2,3,4,5,6,7,8]

5. Перечень учебно-методического обеспечения самостоятельной работы обучающихся по дисциплине

Для каждого обучающегося обеспечен индивидуальный неограниченный доступ к электронной информационно-образовательной среде АлтГТУ:

1. Фоминых, Е. И. Инструментальное программное обеспечение : учебное пособие / Е. И. Фоминых, Т. Е. Фоминых. – Минск : РИПО, 2022. – 413 с. : ил., табл., схем. – Режим доступа: по подписке. – URL: <https://biblioclub.ru/index.php?page=book&id=711499> (дата обращения: 06.02.2025). – Библиогр. в кн. – ISBN 978-985-895-023-1. – Текст : электронный.

6. Перечень учебной литературы

6.1. Основная литература

2. Сотник, С. Л. Проектирование систем искусственного интеллекта : учебное пособие для СПО / С. Л. Сотник. — 2-е изд. — Саратов : Профобразование, 2024. — 228 с. — ISBN 978-5-4488-1009-1. — Текст : электронный // Цифровой образовательный ресурс IPR SMART : [сайт]. — URL: <https://www.iprbookshop.ru/139762.html> (дата обращения: 05.02.2025). — Режим

доступа: для авторизир. пользователей

3. Абдрахманов, М. И. Основы языка программирования Python : учебное пособие для СПО / М. И. Абдрахманов. — Москва : Ай Пи Ар Медиа, 2023. — 142 с. — ISBN 978-5-4497-2310-9. — Текст : электронный // Цифровой образовательный ресурс IPR SMART : [сайт]. — URL: <https://www.iprbookshop.ru/132567.html> (дата обращения: 05.02.2025). — Режим доступа: для авторизир. пользователей

6.2. Дополнительная литература

4. Абрамов, Г. В. Проектирование и разработка информационных систем : учебное пособие для СПО / Г. В. Абрамов, И. Е. Медведкова, Л. А. Коробова. — 2-е изд. — Саратов : Профобразование, 2024. — 169 с. — ISBN 978-5-4488-2259-9. — Текст : электронный // Цифровой образовательный ресурс IPR SMART : [сайт]. — URL: <https://www.iprbookshop.ru/143685.html> (дата обращения: 06.02.2025). — Режим доступа: для авторизир. пользователей

7. Перечень ресурсов информационно-телекоммуникационной сети «Интернет», необходимых для освоения дисциплины

5. Документация по библиотеке Pandas. Ссылка на ресурс: https://pandas.pydata.org/docs/user_guide/

6. Руководство пользователя scikit-learn. Ссылка на ресурс: https://scikit-learn.org/stable/user_guide.html

7. Документация NumPy. Ссылка на ресурс: <https://numpy.org/>

8. Руководство пользователя Python. Ссылка на ресурс - <https://python.land/python-tutorial>

8. Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине, включая перечень программного обеспечения и информационных справочных систем

Для успешного освоения дисциплины используются ресурсы электронной информационно-образовательной среды, образовательные интернет-порталы, глобальная компьютерная сеть Интернет. В процессе изучения дисциплины происходит интерактивное взаимодействие обучающегося с преподавателем через личный кабинет студента. Для изучения данной дисциплины профессиональные базы данных и информационно-справочные системы не требуются.

№пп	Используемое программное обеспечение
1	Python
2	Яндекс.Браузер

№пп	Используемые профессиональные базы данных и информационные справочные системы
1	IEEE Xplore - Интернет библиотека с доступом к реферативным и полнотекстовым

№пп	Используемые профессиональные базы данных и информационные справочные системы
	статьям и материалам конференций. Бессрочно без подписки (https://ieeexplore.ieee.org/Xplore/home.jsp)
2	Springer - Издательство с доступом к реферативным и полнотекстовым материалам журналов и книг (https://www.springer.com/gp https://link.springer.com/)
3	Wiley - Издательство с доступом к реферативным и полнотекстовым материалам журналов и книг. Содержит большой раздел Computer Science & Information Technology, содержащий pdf-файлы с полными текстами журналов и книг издательства. Фиксируется пользователь информации на уровне вуза (Access by Polzunov Altai State Technical University) (https://www.wiley.com/en-ru https://www.onlinelibrary.wiley.com/)

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

Наименование специальных помещений и помещений для самостоятельной работы
учебные аудитории для проведения занятий лекционного типа
учебные аудитории для проведения групповых и индивидуальных консультаций
учебные аудитории для проведения текущего контроля и промежуточной аттестации
помещения для воспитательной, самостоятельной работы
лаборатории
учебные аудитории для проведения лабораторных занятий
мастерские

Материально-техническое обеспечение и организация образовательного процесса по дисциплине для инвалидов и лиц с ограниченными возможностями здоровья осуществляется в соответствии с «Положением об обучении инвалидов и лиц с ограниченными возможностями здоровья».

10. Контроль и оценка результатов освоения учебной дисциплины

Код компетенции из УП	Содержание компетенции	Формы и методы оценки
ПК 2.1	Выполнять подготовку данных для проведения аналитических работ	Защита лабораторных работ, зачет
ПК 2.2	Строить статистические и математические модели для систем анализа данных	Защита лабораторных работ, зачет

ПРИЛОЖЕНИЕ А **МЕТОДИЧЕСКИЕ РЕКОМЕНДАЦИИ И УКАЗАНИЯ**

Лекции (в том числе уроки, проводимые в виде лекций) составляют основу теоретического обучения студентов. Они позволяют систематизировать знания студентам о наиболее сложных и актуальных проблемах изучаемой дисциплины. На лекциях особое внимание уделяется не только усвоению студентами изучаемых проблем, но и стимулированию их активной познавательной деятельности, творческого мышления, развитию профессионально-значимых свойств и качеств. Для лучшего освоения учебной дисциплины перед каждой лекцией студент повторяет предыдущий лекционный материал и прорабатывает рассмотренные ранее вопросы с использованием рекомендованной преподавателем основной и дополнительной литературы.

Студенты должны аккуратно вести конспект. В случае недопонимания какой-либо части предмета следует задать вопрос в установленном порядке преподавателю. В процессе работы на лекции необходимо так же выполнять в конспектах модели изучаемого предмета (рисунки, схемы, чертежи и т.д.), которые использует преподаватель.

При подготовке к лабораторным работам студенту, кроме повтора лекционного материала по теме занятия, необходимо также изучить методические рекомендации, выданные преподавателем.

Выполнение всех видов работы в соответствующие сроки позволит студентам в течение семестра вести подготовку к промежуточной аттестации. Промежуточная аттестация проводится в письменном виде в конце семестра.

Методические указания студентам по подготовке к лабораторным работам

Лабораторные работы необходимы для усвоения теоретического материала и формирование учебных и профессиональных практических навыков.

Выполнение лабораторных работ направлено на обобщение, систематизацию, углубление, закрепление теоретических знаний и приобретения практического опыта по конкретным темам дисциплин.

Содержание лабораторных работ представлено в настоящей программе.

При подготовке к лабораторным работам студенту, кроме повтора лекционного материала по теме занятия, необходимо также изучить методические рекомендации к лабораторной работе.

Подготовка к промежуточной аттестации

Промежуточная аттестация является приемом проверки степени усвоения учебного материала и лекционных занятий, качества усвоения обучающимися отдельных разделов, сформированных умений и навыков.

При подготовке к промежуточной аттестации целесообразно:

- внимательно изучить перечень вопросов и определить, в каких источниках находятся сведения, необходимые для ответа на них;
- внимательно прочитать рекомендованную литературу, изучить конспекты по занятиям;
- составить краткие конспекты ответов (планы ответов).