



**МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ
РОССИЙСКОЙ ФЕДЕРАЦИИ**

федеральное государственное бюджетное образовательное учреждение
высшего образования

**«ИРКУТСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ»
(ФГБОУ ВО «ИГУ»)**

Институт математики и информационных технологий
Кафедра информационных технологий



Рабочая программа дисциплины (модуля)

Б1.В.ДВ.01.01 Программное обеспечение для анализа данных

Направление подготовки 02.03.02 Фундаментальная информатика и
информационные технологии

Направленность (профиль) подготовки Фундаментальная информатика и
программная инженерия

Квалификация выпускника бакалавр

Форма обучения очная

Иркутск 2025 г.

1. ЦЕЛИ И ЗАДАЧИ ДИСЦИПЛИНЫ

Цель дисциплины: Формирование у студентов бакалавриата комплекса знаний, умений и навыков в области применения современного программного обеспечения для решения практических задач сбора, обработки, анализа и визуализации данных.

Задачи дисциплины:

- Изучить архитектуру и функциональные возможности современного программного обеспечения для анализа данных.
- Освоить основные библиотеки и инструменты экосистемы Python для анализа данных.
- Сформировать навыки работы с реляционными базами данных с использованием SQL.
- Научить методам и приемам визуализации данных для анализа и представления результатов.
- Привить навыки проведения первичного анализа данных и решения базовых задач машинного обучения." Освоить принципы воспроизводимости исследований и организации проектов по анализу данных.

2. МЕСТО ДИСЦИПЛИНЫ В СТРУКТУРЕ ОПОП ВО

Учебная дисциплина Б1.В.ДВ.01.01 Программное обеспечение для анализа данных относится к части Блока 1 образовательной программы, формируемой участниками образовательных отношений.

3. ТРЕБОВАНИЯ К РЕЗУЛЬТАТАМ ОСВОЕНИЯ ДИСЦИПЛИНЫ

Процесс освоения дисциплины направлен на формирование следующих компетенций в соответствии с ФГОС ВО и ОП ВО по направлению подготовки 02.03.02

Фундаментальная информатика и информационные технологии:

ПК-3 Способен планировать и организовывать аналитические работы с использованием технологий больших данных;

ПК-4 Способен подготавливать и обрабатывать данные для аналитических исследований.

4. СОДЕРЖАНИЕ И СТРУКТУРА ДИСЦИПЛИНЫ

Объем дисциплины составляет 2 зачетных ед., 72 час.

Форма промежуточной аттестации: зачет.

4.1. Содержание дисциплины, структурированное по темам, с указанием видов учебных занятий и отведенного на них количества академических часов

Раздел дисциплины / тема	Виды учебной работы			Самост. работа	Формы текущего контроля; Формы промежут. аттестации
	Контактная работа преподавателя с обучающимися				
	Лекции	Лаб. занятия	Практ. занятия		
Тема 1. Введение в ПО для анализа данных Экосистема инструментов анализа данных. Настройка рабочего окружения: Python, Jupyter Notebook, VS Code. Основы Python для анализа данных		5		5	
Тема 2. Библиотеки Python для анализа данных NumPy: массивы, операции, broadcasting. Pandas: Series, DataFrame, базовые операции, работа с индексами		5		5	
Тема 3. Обработка и очистка данных Загрузка данных из различных источников. Обработка пропущенных значений. Фильтрация и преобразование данных. Группировка и агрегация		5		5	
Тема 4. Визуализация данных Принципы эффективной визуализации. Библиотека Matplotlib. Библиотека Seaborn для статистической визуализации		5		5	
Тема 5. Работа с базами данных Основы SQL для анализа данных. Подключение к БД из Python. Выполнение запросов и анализ результатов		5		5	
Тема 6. Введение в машинное обучение Задачи классификации, регрессии, кластеризации. Библиотека Scikit-learn. Предобработка данных для ML		9		5	
Итого (3 семестр):		34		30	зач.

4.2. Содержание учебного материала

Тема 1. Введение в ПО для анализа данных Экосистема инструментов анализа данных. Настройка рабочего окружения: Python, Jupyter Notebook, VS Code. Основы Python для анализа данных

Тема 2. Библиотеки Python для анализа данных NumPy: массивы, операции, broadcasting. Pandas: Series, DataFrame, базовые операции, работа с индексами

Тема 3. Обработка и очистка данных Загрузка данных из различных источников. Обработка пропущенных значений. Фильтрация и преобразование данных. Группировка и агрегация

Тема 4. Визуализация данных Принципы эффективной визуализации. Библиотека Matplotlib. Библиотека Seaborn для статистической визуализации

Тема 5. Работа с базами данных Основы SQL для анализа данных. Подключение к БД из Python. Выполнение запросов и анализ результатов
Тема 6. Введение в машинное обучение Задачи классификации, регрессии, кластеризации. Библиотека Scikit-learn. Предобработка данных для ML

4.3. Методические указания по организации самостоятельной работы студентов

Самостоятельная работа студентов всех форм и видов обучения является одним из обязательных видов образовательной деятельности, обеспечивающей реализацию требований Федеральных государственных стандартов высшего образования. Согласно требованиям нормативных документов самостоятельная работа студентов является обязательным компонентом образовательного процесса, так как она обеспечивает закрепление получаемых на лекционных занятиях знаний путем приобретения навыков осмысления и расширения их содержания, навыков решения актуальных проблем формирования общекультурных и профессиональных компетенций, научно-исследовательской деятельности, подготовки к семинарам, лабораторным работам, сдаче зачетов и экзаменов. Самостоятельная работа студентов представляет собой совокупность аудиторных и внеаудиторных занятий и работ. Самостоятельная работа в рамках образовательного процесса в вузе решает следующие задачи:

- закрепление и расширение знаний, умений, полученных студентами во время аудиторных и внеаудиторных занятий, превращение их в стереотипы умственной и физической деятельности;
- приобретение дополнительных знаний и навыков по дисциплинам учебного плана;
- формирование и развитие знаний и навыков, связанных с научно-исследовательской деятельностью;
- развитие ориентации и установки на качественное освоение образовательной программы;
- развитие навыков самоорганизации;
- формирование самостоятельности мышления, способности к саморазвитию, самосовершенствованию и самореализации;
- выработка навыков эффективной самостоятельной профессиональной теоретической, практической и учебно-исследовательской деятельности.

Подготовка к лекции. Качество освоения содержания конкретной дисциплины прямо зависит от того, насколько студент сам, без внешнего принуждения формирует у себя установку на получение на лекциях новых знаний, дополняющих уже имеющиеся по данной дисциплине. Время на подготовку студентов к двухчасовой лекции по нормативам составляет не менее 0,2 часа.

Подготовка к практическому занятию. Подготовка к практическому занятию включает следующие элементы самостоятельной деятельности: четкое представление цели и задач его проведения; выделение навыков умственной, аналитической, научной деятельности, которые станут результатом предстоящей работы. Выработка навыков осуществляется с помощью получения новой информации об изучаемых процессах и с помощью знания о том, в какой степени в данное время студент владеет методами исследовательской деятельности, которыми он станет пользоваться на практическом занятии. Подготовка к практическому занятию нередко требует подбора материала, данных и специальных источников, с которыми предстоит учебная работа. Студенты должны дома подготовить к занятию 3–4 примера формулировки темы исследования, представленного в монографиях, научных статьях, отчетах. Затем они самостоятельно осуществляют поиск соответствующих источников, определяют актуальность

конкретного исследования процессов и явлений, выделяют основные способы доказательства авторами научных работ ценности того, чем они занимаются. В ходе самого практического занятия студенты сначала представляют найденные ими варианты формулировки актуальности исследования, обсуждают их и обосновывают свое мнение о наилучшем варианте. Время на подготовку к практическому занятию по нормативам составляет не менее 0,2 часа.

Подготовка к семинарскому занятию. Самостоятельная подготовка к семинару направлена: на развитие способности к чтению научной и иной литературы; на поиск дополнительной информации, позволяющей глубже разобраться в некоторых вопросах; на выделение при работе с разными источниками необходимой информации, которая требуется для полного ответа на вопросы плана семинарского занятия; на выработку умения правильно выписывать высказывания авторов из имеющихся источников информации, оформлять их по библиографическим нормам; на развитие умения осуществлять анализ выбранных источников информации; на подготовку собственного выступления по обсуждаемым вопросам; на формирование навыка оперативного реагирования на разные мнения, которые могут возникать при обсуждении тех или иных научных проблем. Время на подготовку к семинару по нормативам составляет не менее 0,2 часа.

Подготовка к коллоквиуму. Коллоквиум представляет собой коллективное обсуждение раздела дисциплины на основе самостоятельного изучения этого раздела студентами. Подготовка к данному виду учебных занятий осуществляется в следующем порядке. Преподаватель дает список вопросов, ответы на которые следует получить при изучении определенного перечня научных источников. Студентам во внеаудиторное время необходимо прочитать специальную литературу, выписать из нее ответы на вопросы, которые будут обсуждаться на коллоквиуме, мысленно сформулировать свое мнение по каждому из вопросов, которое они выскажут на занятии. Время на подготовку к коллоквиуму по нормативам составляет не менее 0,2 часа.

Подготовка к контрольной работе. Контрольная работа назначается после изучения определенного раздела (разделов) дисциплины и представляет собой совокупность развернутых письменных ответов студентов на вопросы, которые они заранее получают от преподавателя. Самостоятельная подготовка к контрольной работе включает в себя: — изучение конспектов лекций, раскрывающих материал, знание которого проверяется контрольной работой; повторение учебного материала, полученного при подготовке к семинарским, практическим занятиям и во время их проведения; изучение дополнительной литературы, в которой конкретизируется содержание проверяемых знаний; составление в мысленной форме ответов на поставленные в контрольной работе вопросы; формирование психологической установки на успешное выполнение всех заданий. Время на подготовку к контрольной работе по нормативам составляет 2 часа.

Подготовка к зачету. Самостоятельная подготовка к зачету должна осуществляться в течение всего семестра. Подготовка включает следующие действия: перечитать все лекции, а также материалы, которые готовились к семинарским и практическим занятиям в течение семестра, соотнести эту информацию с вопросами, которые даны к зачету, если информации недостаточно, ответы находят в предложенной преподавателем литературе. Рекомендуются делать краткие записи. Время на подготовку к зачету по нормативам составляет не менее 4 часов.

Подготовка к экзамену. Самостоятельная подготовка к экзамену схожа с подготовкой к зачету, особенно если он дифференцированный. Но объем учебного материала, который нужно восстановить в памяти к экзамену, вновь осмыслить и понять, значительно больше, поэтому требуется больше времени и умственных усилий. Важно сформировать целостное представление о содержании ответа на каждый вопрос, что предполагает знание разных научных трактовок сущности того или иного явления,

процесса, умение раскрывать факторы, определяющие их противоречивость, знание имен ученых, изучавших обсуждаемую проблему. Необходимо также привести информацию о материалах эмпирических исследований, что указывает на всестороннюю подготовку студента к экзамену. Время на подготовку к экзамену по нормативам составляет 36 часов для бакалавров.

В ФБГОУ ВО «ИГУ» организация самостоятельной работы студентов регламентируется Положением о самостоятельной работе студентов, принятым Ученым советом ИГУ 22 июня 2012 г.

5. УЧЕБНО-МЕТОДИЧЕСКОЕ И ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

5.1. Литература, базы данных, информационно-справочные и поисковые системы

1. 1. Маккинни У. Python и анализ данных. - СПб.: Питер, 2022. - 482 с.
2. 2. Дж. Вандер Плас. Python для сложных задач: наука о данных и машинное обучение. - СПб.: Питер, 2022. - 576 с.
3. 3. Грубер М. Изучаем SQL. - СПб.: Питер, 2021. - 448 с.

6. МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

6.1. Учебная аудитория для проведения:

- занятий лекционного типа,
- занятий семинарского (практического) типа,
- групповых и индивидуальных консультаций,
- текущего контроля и промежуточной аттестации.

Оснащение:

Учебная аудитория, укомплектованная специализированной мебелью и техническими средствами обучения, служащими для представления учебной информации большой аудитории, для проведения занятий лекционного типа, практических занятий (семинарского типа), курсового проектирования (выполнения курсовых работ), групповых и индивидуальных консультаций, текущего контроля и промежуточной аттестации. Для проведения занятий лекционного типа обучающимся предлагаются наборы демонстрационного оборудования и учебно-наглядные пособия, обеспечивающие тематические иллюстрации.

6.2. Помещения для самостоятельной работы обучающихся.

Оснащение:

Помещения для самостоятельной работы обучающихся, оснащенные учебной мебелью. Рабочие места обучающихся оборудованы компьютерной техникой и подключены в локальную вычислительную сеть, в т.ч. с использованием беспроводного Wi-Fi подключения, с возможностью выхода в глобальную сеть Интернет и с доступом в электронную информационно-образовательную среду.

6.3. Программное обеспечение

Приложение для чтения PDF-файлов, браузер для просмотра интернет контента, приложение для создания PDF-файлов.

7. ОЦЕНОЧНЫЕ МАТЕРИАЛЫ ДЛЯ ТЕКУЩЕГО КОНТРОЛЯ И ПРОМЕЖУТОЧНОЙ АТТЕСТАЦИИ

7.1. Оценочные средства для промежуточной аттестации

Список вопросов для промежуточной аттестации:

1. Классификация программного обеспечения для анализа данных. Преимущества и недостатки различных подходов.
2. Библиотека NumPy: основные структуры данных и операции. Концепция broadcasting.
3. Библиотека Pandas: структуры Series и DataFrame. Основные операции и методы.
4. Методы загрузки данных из различных источников в Pandas. Форматы файлов и их особенности.
5. Обработка пропущенных значений в анализе данных. Стратегии и методы.
6. Группировка и агрегация данных в Pandas. Методы groupby и агрегирующие функции.
7. Принципы эффективной визуализации данных. Критерии выбора типа диаграммы.
8. Библиотека Matplotlib: архитектура и основные компоненты. Типы графиков и их построение.
9. Библиотека Seaborn: особенности и преимущества. Основные типы статистических визуализаций.
10. Язык SQL в контексте анализа данных. Основные конструкции для аналитических запросов.
11. Соединения таблиц в SQL. Виды JOIN и их применение в анализе данных.
12. Оконные функции в SQL. Примеры использования для аналитических задач.
13. Интеграция Python и СУБД. Библиотеки и методы подключения к базам данных.
14. Задачи машинного обучения: классификация, регрессия, кластеризация. Примеры и особенности.
15. Библиотека Scikit-learn: архитектура и основные компоненты. Процесс построения модели.
16. Предобработка данных для машинного обучения. Кодирование категориальных признаков.
17. Масштабирование признаков в машинном обучении. Методы и необходимость применения.
18. Разделение данных на обучающую и тестовую выборки. Методы и их обоснование.
19. Метрики качества для задач классификации: accuracy, precision, recall, F1-score.
20. Метрики качества для задач регрессии: MAE, MSE, R². Интерпретация результатов.
21. Принципы воспроизводимости в анализе данных. Системы контроля версий и виртуальные окружения.
22. Jupyter Notebook как инструмент для анализа данных. Преимущества и лучшие практики.
23. Методы обработки текстовых данных. Векторизация текста и мешок слов.
24. Анализ временных рядов: основные понятия и методы визуализации.
25. Этика в анализе данных. Проблемы смещенных данных и конфиденциальности.