



**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



Рабочая программа дисциплины (модуля)

**ФТД.02 Анализ и обработка данных при разработке цифровых
продуктов**

Направление подготовки	01.03.02 Прикладная математика и информатика
Направленность (профиль) подготовки	Системы искусственного интеллекта
Квалификация выпускника	бакалавр
Форма обучения	очная

Иркутск 2023 г.

1. ЦЕЛИ И ЗАДАЧИ ДИСЦИПЛИНЫ

Цели: сформировать у студентов компетентностный подход к анализу, обработке и представлению больших данных, а также умение применять полученные знания на практике.

Задачи:

освоить основные тенденций развития теории и практики анализа и интерпретации данных;

приобрести практические навыки работы с современными пакетами прикладных программ для решения задач анализа и интерпретации данных;

сформировать целостную систему знаний и методов анализа данных.

2. МЕСТО ДИСЦИПЛИНЫ В СТРУКТУРЕ ОПОП ВО

Учебная дисциплина ФТД.02 Анализ и обработка данных при разработке цифровых продуктов относится к факультативным дисциплинам.

Для изучения данной учебной дисциплины необходимы знания, умения и навыки, формируемые предшествующими дисциплинами: Базы данных и системы управления базами данных, Структуры и алгоритмы компьютерной обработки данных, Теория вероятностей и математическая статистика, Эконометрика.

Перечень последующих учебных дисциплин, для которых необходимы знания, умения и навыки, формируемые данной учебной дисциплиной: Обработка больших объемов данных.

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 01.03.02 Прикладная математика и информатика:

ПК-3 Способен создавать, модифицировать и сопровождать информационные системы, автоматизирующие задачи организационного управления и процессов функционирования производственных организаций, социальных институтов и структур.

В результате освоения дисциплины обучающийся должен

знать: современные методы анализа и обработки данных;

уметь: решать основные задачи, возникающие при анализе данных;

владеть: навыками решения задач, связанными с анализом и визуализацией данных различной природы.

4. СОДЕРЖАНИЕ И СТРУКТУРА ДИСЦИПЛИНЫ

Объем дисциплины составляет 1 зачетных ед., 36 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	Сем.	Виды учебной работы	Формы
--------------------------	------	---------------------	-------

		Контактная работа преподавателя с обучающимися			Самост. работа	текущего контроля; Формы промежут. аттестации
		Лекции	Лаб. занятия	Практ. занятия		
Тема 1 Основы обработки и хранения больших массивов данных			10			вопросы к зачету
Тема 2 Анализ и визуализация данных с помощью BI-инструментов			6			вопросы к зачету
Итого (8 семестр):		16	16		2	зач.

4.2. План внеаудиторной самостоятельной работы обучающихся по дисциплине

Раздел дисциплины / тема	Самостоятельная работа обучающихся			Оценочное средство	Учебно-методическое обеспечение самост. работы
	Вид самост. работы	Сроки выполнения	Затраты времени		
Тема 1 Анализ поведения пользователей по графовым структурам.	подготовка конспекта	в течение семестра	1	вопросы к зачёту	материалы лекций и презентаций
Тема 2 Клиент-ориентированные и бизнес-ориентированные системы: архитектура и применение в продуктовых приложениях	подготовка конспекта	в течение	1	вопросы к зачёту	материалы лекций и презентаций
Общая трудоемкость самостоятельной работы (час.)			2		
Из них с использованием электронного обучения и дистанционных образовательных технологий (час.)			1		

4.3. Содержание учебного материала

Тема1. Основы обработки и хранения больших массивов данных

Архитектура систем обработки данных. Консолидация данных. Сравнение организаций СУБД для аналитических целей. Системы аналитической обработки данных: назначение и применение. Обработка данных по запросу (пассивный сбор данных). Организация потоковой обработки (real-time clickstream). Парадигмы построения систем аналитической обработки данных. Методы аналитической обработки данных. Задачи из теории измерений. Возможностями предобработки данных с использованием Python. Подготовка данных к анализу. Пайплайны обработки данных. Изучение структуры данных, описательные статистики. Выдвижение и проверка гипотез. Сбор данных для A/B тестирования. Представление поведения пользователей в форме конечного автомата. Поведенческий анализ. Клиент-ориентированные и бизнес-ориентированные системы: архитектура и применение в продуктовых приложениях.

Тема2. Анализ и визуализация данных с помощью BI-инструментов

Понятие BI-инструментов и их место в анализе данных. Предварительная подготовка данных. Подключение источников данных. Интерактивные отчеты. Вычисляемые поля. Табличные вычисления. Визуализация географических данных

Основы дизайна отчетов и цветового кодирования. Создание дашбордов. Основные типы визуализаций. Разведочный и описательный анализ данных с помощью визуализации. Практические аспекты визуализации данных с помощью программы Tableau Public.

4.3.1. Перечень лабораторных работ

Тема занятия	Всего часов	Оценочные средства	Формируемые компетенции
Консолидация данных. Выбор СУБД для использования в системах аналитической обработки данных	4	домашнее задание презентация	ПК-3
Назначение и классификация систем аналитической обработки данных Парадигмы построения систем аналитической обработки данных	2	домашнее задание презентация	ПК-3
Подходы к построению архитектуры систем аналитической обработки данных	2	домашнее задание	ПК-3
Описательные статистики. Возможности выявления структуры данных с использованием Python.	2	домашнее задание презентация	ПК-3
Понятие BI-инструмента	2	конспект	ПК-3
Tableau Public – знакомство, интерфейс.	2	домашнее задание	ПК-3
Создание дашбордов	2	домашнее задание	ПК-3

4.3.2. Перечень тем (вопросов), выносимых на самостоятельное изучение студентами в рамках самостоятельной работы

Тема	Задание	Формируемые компетенции
Тема 1 Визуализация географических данных	Подготовка презентации	ПК-3
Тема 2 Разведочный и описательный анализ данных с помощью визуализации	Подготовка презентации	ПК-3
Тема 3 Знакомство с системами Web-аналитики (Google Analytics, Яндекс.Метрика)	Подготовка презентации	ПК-3
Тема 4 Маркетинговая аналитика	Подготовка презентации	ПК-3

4.4. 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 лабораторному 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 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. Рекомендуется делать краткие записи. Время на подготовку к зачету по нормативам составляет не менее 4 часов.

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

а) основная литература:

1. Бабенко М. А., Левин М. В. Введение в теорию алгоритмов и структур данных. — М.: МЦНМО. 2020. 144 с+

2. **Макшанов, А. В.** Технологии интеллектуального анализа данных [Электронный ресурс] : учебное пособие / А. В. Макшанов, А. Е. Журавлев. - 2-е изд., стер. - Электрон.текстовые дан. - Санкт-Петербург : Лань, 2019. - 212 с. - **Режим доступа:** <https://e.lanbook.com/book/120063>, <https://e.lanbook.com/img/cover/book/120063.jpg>. - ЭБС "Лань". - Неогранич. доступ. - ISBN 978-5-8114-4493-9

3. **Алексеев, Д. С.** Технологии интеллектуального анализа данных [Электронный ресурс] / Д. С. Алексеев, О. В. Щекочихин. - Электрон.текстовые дан. - Санкт-Петербург : Лань, 2022. - 176 с. - **Режим доступа:** <https://e.lanbook.com/book/187559>, <https://e.lanbook.com/img/cover/book/187559.jpg>. - ЭБС "Лань". - Неогранич. доступ. - ISBN 978-5-8114-8299-

4. **Мхитарян, Владимир Сергеевич.** Анализ данных [Электронный ресурс] : Учебник для вузов / В. С. Мхитарян, М. Ю. Архипова [и др.]. - Электрон.текстовые дан. - Москва : Юрайт, 2021. - 490 с. - (Высшее образование). - **Режим доступа:** <https://urait.ru/bcode/469022>. - ЭБС "Юрайт". - неогранич. доступ. - ISBN 978-5-534-00616-2 : 1299.00 р. URL: <https://urait.ru/bcode/469022> (дата обращения: 28.06.2021).

5. Грабер Мартин SQL. – М.: Лори, 2007. – 644 с.+

6. Грасс Дж. DataScience. Наука о данных с нуля. СПб.: – БХВ-Петербург, 2017. +

в) базы данных, информационно-справочные и поисковые системы:

1. Научная электронная библиотека «ELIBRARY.RU» [Электронный ресурс] : сайт. – Режим доступа: <http://elibrary.ru/defaultx.asp>.

2. Открытая электронная база ресурсов и исследований «Университетская информационная система РОССИЯ» [Электронный ресурс] : сайт. – Режим доступа: <http://uisrussia.msu.ru>

3. Государственная информационная система «Национальная электронная библиотека» [Электронный ресурс] : сайт. – Режим доступа: <http://нэб.рф>.

6. МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

6.1. Учебно-лабораторное оборудование

ЭТОТ РАЗДЕЛ НЕ ЗАПОЛНЯТЬ

6.2. Программное обеспечение

ПЕРЕЧИСЛИТЬ ПРОГРАММНОЕ ОБЕСПЕЧЕНИЕ ТРЕБУЕМОЕ ДЛЯ ДИСЦИПЛИНЫ

7. ОЦЕНОЧНЫЕ МАТЕРИАЛЫ ДЛЯ ТЕКУЩЕГО КОНТРОЛЯ И ПРОМЕЖУТОЧНОЙ АТТЕСТАЦИИ

7.1. Оценочные средства текущего контроля

Вид контроля	Контролируемые темы	Контролируемые компетенции
Домашнее задание	1,2	ПК-3
Вопросы к экзамену	1,2	ПК-3
Устный опрос	1,2	ПК-3

Примеры оценочных средств текущего контроля

Темы презентаций

1. Визуализация географических данных
2. Разведочный и описательный анализ данных с помощью визуализации
3. Методы аналитической обработки данных
4. Архитектура Data Lake
5. Архитектура Data Mesh
6. Пайплайны обработки данных. Подходы к построению системы предобработки данных.
7. Продуктовая аналитика с использованием графовых представлений.
8. Основы дизайна отчетов
9. Основы продуктовой аналитики

7.2. Оценочные средства для промежуточной аттестации

Список вопросов для промежуточной аттестации (зачета)

1. Консолидация данных.
2. Сравнение организаций СУБД для аналитических целей: Реляционные СУБД vs NoSQL СУБД vs Колоночные СУБД.
3. Системы аналитической обработки данных: назначение и применение.
4. Обработка данных по запросу (пассивный сбор данных).
5. Организация потоковой обработки (real-time clickstream).
6. Парадигмы построения систем аналитической обработки данных.
7. Архитектура ETL-DWH.
8. Архитектура Data Lake.
9. Архитектура Data-mesh (микросервисный подход к обработке данных).
10. Тестирование vs реактивный мониторинг в системах аналитической обработки данных.
11. Задачи из теории измерений. Возможностями предобработки данных с использованием Python.
12. Подготовка данных к анализу. Пайплайны обработки данных.
13. Изучение структуры данных, описательные статистики.
14. Выдвижение и проверка гипотез. Сбор данных для A/B тестирования.
15. Представление поведения пользователей в форме конечного автомата. Поведенческий анализ.
16. Клиент-ориентированные и бизнес-ориентированные системы: архитектура и применение в продуктовых приложениях.
17. Предварительная подготовка данных.
18. Подключение источников данных.
19. Основные типы визуализаций.
20. Интерактивные отчеты.
21. Вычисляемые поля.
22. Табличные вычисления.
23. Визуализация географических данных.
24. Основы дизайна отчетов и цветового кодирования.
25. Создание дашбордов.
26. Разведочный и описательный анализ данных с помощью визуализации.

Разработчик: Захарова И. В., к.ф.-м.н., доцент кафедры математического анализа и дифференциальных уравнений