

Министерство науки и высшего образования Российской Федерации
Федеральное государственное бюджетное образовательное учреждение высшего
образования
«ИРКУТСКИЙ НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ ТЕХНИЧЕСКИЙ
УНИВЕРСИТЕТ»

Структурное подразделение «Институт информационных технологий и анализа данных»

УТВЕРЖДЕНА:

на заседании Совета института ИТиАД им. Е.И.Попова

Протокол №8 от 24 февраля 2025 г.

Рабочая программа дисциплины

«МЕТОДЫ АНАЛИЗА ДАННЫХ»

Направление: 09.03.02 Информационные системы и технологии

Информационные системы и технологии в административном управлении

Квалификация: Бакалавр

Форма обучения: очная

Документ подписан простой электронной подписью Составитель программы: Бучнев Олег Сергеевич Дата подписания: 05.06.2025

Документ подписан простой электронной подписью Утвердил: Говорков Алексей Сергеевич Дата подписания: 11.06.2025

Документ подписан простой электронной подписью Согласовал: Аршинский Вадим Леонидович Дата подписания: 06.06.2025

Год набора – 2025

Иркутск, 2025 г.

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

1.1 Дисциплина «Методы анализа данных» обеспечивает формирование следующих компетенций с учётом индикаторов их достижения

Код, наименование компетенции	Код индикатора компетенции
ПКС-1 Способен проводить научные исследования при разработке, внедрении и сопровождении информационных технологий и систем на всех этапах жизненного цикла	ПКС-1.4

1.2 В результате освоения дисциплины у обучающихся должны быть сформированы

Код индикатора	Содержание индикатора	Результат обучения
ПКС-1.4	Способен выбирать и применять различные методы анализа данных для решения профессиональных задач	Знать основные типы задач в современном анализе данных и методы их решения; возможности и ограничения применяемых методов и моделей для анализа. Уметь осуществлять постановку задачи и выбор методов для решения задачи с учетом имеющихся ограничений, осуществлять предобработку данных, осуществлять интерпретацию результатов анализа. Владеть современными средствами и инструментами для реализации этапов и методов анализа данных.

2 Место дисциплины в структуре ООП

Изучение дисциплины «Методы анализа данных» базируется на результатах освоения следующих дисциплин/практик: «Математика», «Теория вероятностей и математическая статистика», «Программирование на языке высокого уровня», «Базы данных»

Дисциплина является предшествующей для дисциплин/практик: «Интеллектуальные системы и технологии», «Нейросетевые технологии», «Основы СППР»

3 Объем дисциплины

Объем дисциплины составляет – 4 ЗЕТ

Вид учебной работы	Трудоемкость в академических часах (Один академический час соответствует 45 минутам астрономического часа)	
	Всего	Семестр № 4
Общая трудоемкость дисциплины	144	144

Аудиторные занятия, в том числе:	64	64
лекции	32	32
лабораторные работы	32	32
практические/семинарские занятия	0	0
Контактная работа, в том числе	0	0
в форме работы в электронной информационной образовательной среде	0	0
Самостоятельная работа (в т.ч. курсовое проектирование)	44	44
Трудоемкость промежуточной аттестации	36	36
Вид промежуточной аттестации (итогового контроля по дисциплине)	Экзамен	Экзамен

4 Структура и содержание дисциплины

4.1 Сводные данные по содержанию дисциплины

Семестр № 4

№ п/п	Наименование раздела и темы дисциплины	Виды контактной работы						СРС		Форма текущего контроля
		Лекции		ЛР		ПЗ(СЕМ)				
		№	Кол. Час.	№	Кол. Час.	№	Кол. Час.	№	Кол. Час.	
1	2	3	4	5	6	7	8	9	10	11
1	Сбор и подготовка данных; Разведочный анализ	1, 2	4	1	8			1	10	Отчет по лаборатор ной работе
2	Статистический анализ	3	4							Тест
3	Исследование зависимостей	4, 5	6	2	8			1	10	Отчет по лаборатор ной работе
4	Задачи кластеризации и классификации	6, 7	4	3	6			1	8	Отчет по лаборатор ной работе
5	Методы понижения размерности и факторный анализ	8, 9	6	4	4			1	8	Отчет по лаборатор ной работе
6	Логистическая регрессия	10	2							Тест
7	Метод опорных векторов	11	2							Тест
8	Деревья решений для классификации	12	4	5	6			1	8	Отчет по лаборатор ной работе
	Промежуточная аттестация								36	Экзамен
	Всего		32		32				80	

4.2 Краткое содержание разделов и тем занятий

Семестр № 4

№	Тема	Краткое содержание
1	Сбор и подготовка данных; Разведочный анализ	Сбор данных из различных источников. Типы шкал. Описательная статистика. Выявление выбросов
2	Статистический анализ	Точечные оценки, метод максимального правдоподобия. Статистические критерии, ошибки первого и второго рода. Мощность критерия.
3	Исследование зависимостей	Корреляционный и регрессионный анализ. Частная, множественная корреляция, индекс корреляции. Нелинейная и множественная регрессия. Парадоксы регрессии.
4	Задачи кластеризации и классификации	Обучение с учителем и без учителя на примере кластерного и дискриминантного анализа. Иерархический кластерный анализ и метод k-средних. Дискриминантный анализ, квадратичный дискриминантный анализ. Дискриминантный анализ для многоклассовой классификации. Классификация методом kNN.
5	Методы понижения размерности и факторный анализ	Главные компоненты, главные факторы. Нахождение факторных нагрузок. Вращение факторов.
6	Логистическая регрессия	Метод логистической регрессии в задаче классификации. Оценка чувствительности и специфичности модели. ROC-кривая, AUC-оценка.
7	Метод опорных векторов	Метод опорных векторов. Теорема Куна-Таккера. Задача квадратичного программирования, ядра классификации.
8	Деревья решений для классификации	Деревья решений CART и CTree. Критерии разбиения. Оценка качества классификации. Переобученные модели, способы борьбы с переобучением.

4.3 Перечень лабораторных работ

Семестр № 4

№	Наименование лабораторной работы	Кол-во академических часов
1	Сбор и подготовка данных для анализа	8
2	Корреляционный и регрессионный анализ	8
3	Решение задачи кластеризации и классификации	6
4	Метод главных компонент и факторный анализ	4
5	Построение деревьев решений для решения задачи классификации	6

4.4 Перечень практических занятий

Практических занятий не предусмотрено

4.5 Самостоятельная работа

Семестр № 4

№	Вид СРС	Кол-во академических часов
1	Оформление отчетов по лабораторным и практическим работам	44

В ходе проведения занятий по дисциплине используются следующие интерактивные методы обучения: публичная презентация, кейс-технология

5 Перечень учебно-методического обеспечения дисциплины

5.1 Методические указания для обучающихся по освоению дисциплины

5.1.1 Методические указания для обучающихся по лабораторным работам:

Бучнев О. С. Методы анализа данных : лабораторный практикум / О. С. Бучнев, 2022. - 114 с.

5.1.2 Методические указания для обучающихся по самостоятельной работе:

Бучнев О. С. Методы анализа данных: электронный курс / О. С. Бучнев
<https://el.istu.edu/course/view.php?id=957> (дата обращения 03.06.25) Режим доступа: для зарегистрир. пользователей.

6 Фонд оценочных средств для контроля текущей успеваемости и проведения промежуточной аттестации по дисциплине

6.1 Оценочные средства для проведения текущего контроля

6.1.1 семестр 4 | Отчет по лабораторной работе

Описание процедуры.

Выполнение каждой лабораторной работы состоит из двух частей: практической части в ходе которой необходимо выполнить индивидуальное задание и подготовка и защита отчета по лабораторной работе. Реализация решения индивидуального задания осуществляется средствами языка программирования Python. В ходе выполнения лабораторной работы студент должен применить для выполнения задания теоретические знания, полученные в ходе изучения курса, а также приобрести практические навыки использования процедур и функций языка Python. Подготовка отчета по лабораторной работе необходима для закрепления и переосмысления полученных знаний и навыков. Для успешной сдачи лабораторной работы необходимо продемонстрировать работу программы, реализующую индивидуальное задание, устранить выявленные преподавателем недостатки, подготовить и защитить отчет. Требования к оформлению программного кода, а также содержанию отчета указаны в соответствующем разделе методических указаний для каждой лабораторной работы.

Критерии оценивания.

В процессе защиты отчета по лабораторной работе, студенту задаются контрольные вопросы теоретического и практического характера, соответствующие теме работы. Для успешной защиты отчета студенту необходимо дать краткое изложение основных результатов полученных в ходе выполнения лабораторной работы, устно ответить на теоретические вопросы по теме лабораторной работы, а также продемонстрировать

умение ориентироваться в написанном программном коде. Успешная защита отчета является необходимым условием для выставления оценки «зачтено» по соответствующей лабораторной работе.

6.1.2 семестр 4 | Тест

Описание процедуры.

В конце лекции за 20 минут до окончания лекции в системе дистанционного обучения открывается доступ к тесту, каждый студент самостоятельно отвечает на пять вопросов.

Критерии оценивания.

Для успешного прохождения необходимо ответить на три вопроса из пяти.

6.2 Оценочные средства для проведения промежуточной аттестации

6.2.1 Критерии и средства (методы) оценивания индикаторов достижения компетенции в рамках промежуточной аттестации

Индикатор достижения компетенции	Критерии оценивания	Средства (методы) оценивания промежуточной аттестации
ПКС-1.4	Приводит примеры моделей, условий применимости моделей. Владеет инструментами и технологиями реализации методов и моделей для анализа данных. Работает со специализированными программными средствами. Способен осуществлять выбор методов и инструментов для анализа, осуществлять предобработку данных, интерпретировать полученные результаты.	Устное собеседование по теоретическим вопросам и защита лабораторных работ.

6.2.2 Типовые оценочные средства промежуточной аттестации

6.2.2.1 Семестр 4, Типовые оценочные средства для проведения экзамена по дисциплине

6.2.2.1.1 Описание процедуры

Экзаменационный билет содержит два теоретических вопроса и одно практическое задание, для допуска до экзамена необходимо сдать все лабораторные работы и успешно пройти тестирование по трем темам, не охваченным лабораторными работами. Оформляются 15 билетов, в каждом билете 2 теоретических вопроса и один практический. Оценка выставляется по результатам собеседования по теоретическим вопросам, и по результатам выполнения практического задания.

Пример задания:

1. Меры зависимости. Корреляционный и регрессионный анализ. Необходимость оптимизации или улучшения моделей регрессии.
2. Факторный анализ. Отличие от главных компонент. Оценка факторов. Вращение факторов.
3. Выполнить стандартизацию данных.

Список вопросов к экзамену:

1. Задачи и методы анализа данных.
2. Процесс анализа данных. Входные данные. Результат анализа данных.
3. Разведочный анализ данных. Назначение. Этапы разведочного анализа.
4. Типы шкал, их атрибуты, примеры. Меры центральной тенденции.
5. Пропуски в данных. Механизмы формирования пропусков. Подходы и методы заполнения.
6. Статические критерии, типы. Гипотезы, ошибки I и II рода, уровень значимости. Критическая область, p -value, доверительный интервал. Примеры статистических критериев.
7. Меры зависимости. Корреляционный и регрессионный анализ.
8. Виды корреляции и регрессии. Ранговая корреляция. МНК. Примеры.
9. Сложности при оценке зависимостей. Примеры.
10. Надежность коэффициентов регрессии. Коэффициент детерминации.
11. Уравнение множественной регрессии. Назначение, вычисления, применение.
12. Критерии и оценки качества регрессионной модели.
13. Необходимость оптимизации или улучшения моделей регрессии.
14. Проблема переобучения.
15. Ридж-регрессия, Лассо-регрессия. Назначение, вычисления, применение.
16. Машинное обучение. Задачи кластеризации и классификации (обучение с учителем и без учителя). Понятие обучающей выборки.
17. Задачи кластеризации. Этапы кластерного анализа. Примеры.
18. Нормировка и стандартизация. Меры близости.
19. Иерархическая кластеризация, k -средних. Примеры итераций.
20. Подходы к определению хорошего числа кластеров.
21. Понятие о Байесовской классификации.
22. Алгоритм kNN . «Недообучение» и переобучение в этом алгоритме.
23. Задачи дискриминантного анализа.
24. Требования к данным при проведении дискриминантного анализа.
25. Алгоритм дискриминантного анализа.
26. Назначение методов снижения размерности.
27. Метод главных компонент. Алгоритм метода. Собственные векторы и собственные числа.
28. Факторный анализ. Отличие от главных компонент. Оценка факторов.
29. Вращение факторов.
30. Метод деревьев решений. Задачи, решаемые методом деревьев решений.
31. Обобщенный алгоритм деревьев решений.
32. Алгоритм CART.
33. Переобученное и недообученное дерево. Способы борьбы с переобучением.
34. Уравнение логистической кривой. Коэффициенты кривой.
35. Классификация с помощью модели логистической регрессии.
36. ROC-анализ.
37. Выбор порога отсечения в модели логистической регрессии.
38. Метод опорных векторов. Назначение и алгоритм. Случай линейной неразделимости.

Практические задания

1. Построить гистограммы и диаграммы размаха.
2. Выполнить стандартизацию данных.
3. Вывести тепловую карту.
4. Построить уравнение регрессии и вывести его график.
5. Построить уравнение множественной регрессии и оценить качество модели.
6. Вывести дерево иерархической кластеризации.
7. Выполнить кластеризацию методом k-средних.
8. Построить модель дискриминантного анализа и оценить качество модели.
9. Выполнить снижение размерности.
10. Выполнить факторный анализ с вращением и без. Интерпретировать результаты.
11. Построить дерево решений. Оценить качество.
12. Построить переобученное и недообученное дерево решений.

-

6.2.2.1.2 Критерии оценивания

Отлично	Хорошо	Удовлетворительно	Неудовлетворительно
Демонстрирует полученные знания. Умеет осуществлять постановку задачи, выбирать методы анализа. Работает со специализированными программными средствами и библиотеками. Знает ограничения применяемых методов. Владеет способами верификации моделей. Уверенно интерпретирует полученные результаты.	Демонстрирует полученные знания. Умеет осуществлять постановку задачи, колеблется при выборе метода анализа и обосновании выбора. Неуверенно работает со специализированными программными средствами и библиотеками. Знает ограничения применяемых методов. Владеет способами верификации моделей. Может интерпретировать полученные результаты.	Плохо демонстрирует полученные знания. На вопросы отвечает плохо, не приводит примеры. Работает со специализированными программными средствами. Не знает ограничения применяемых методов. Не владеет способами верификации моделей. Не умеет интерпретировать полученные результаты.	Непонимание назначения методов анализа данных и математических моделей, которые лежат в их основе. Не умеет осуществить постановку задачи.

7 Основная учебная литература

1. Татарникова, Т. М. Интеллектуальный анализ данных : учебное пособие / Т. М. Татарникова. — Вологда : Инфра-Инженерия, 2024. — 172 с. — ISBN 978-5-9729-1772-3. — Текст : электронный // Лань : электронно-библиотечная система. — URL:

<https://e.lanbook.com/book/427826> (дата обращения: 03.06.2025). — Режим доступа: для авториз. пользователей.

2. Золкин, А. Л. Математическое моделирование и анализ данных : учебное пособие для вузов / А. Л. Золкин, М. В. Сартаков. — Санкт-Петербург : Лань, 2025. — 128 с. — ISBN 978-5-507-51616-2. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <https://e.lanbook.com/book/455660> (дата обращения: 03.06.2025). — Режим доступа: для авториз. пользователей.

3. Маккинни, У. Python и анализ данных. Первичная обработка данных с применением pandas, NumPy и Jupiter : справочник / У. Маккинни ; перевод с английского А. А. Слинкина. — 3-е изд. — Москва : ДМК Пресс, 2023. — 536 с. — ISBN 978-5-93700-174-0. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <https://e.lanbook.com/book/348086> (дата обращения: 03.06.2025). — Режим доступа: для авториз. пользователей.

8 Дополнительная учебная литература и справочная

1. Миркин Б. Г. Введение в анализ данных : учебник и практикум для вузов / Б. Г. Миркин, 2022. - 174.

2. Брантон С. Л. Анализ данных в науке и технике. Машинное обучение, динамические системы и управление / С. Л. Брантон, Дж. Н. Куц, 2021. - 542.

3. Сапрыкин О. Н. Интеллектуальный анализ данных : учебное пособие / О. Н. Сапрыкин, 2020. - 80.

4. Криволапов С. Я. Введение в анализ данных. Поиск структуры данных с применением языка Python : учебное пособие / С. Я. Криволапов, 2024. - 177.

9 Ресурсы сети Интернет

1. <http://library.istu.edu/>
2. <https://e.lanbook.com/>

10 Профессиональные базы данных

1. <http://new.fips.ru/>
2. <http://www1.fips.ru/>

11 Перечень информационных технологий, лицензионных и свободно распространяемых специализированных программных средств, информационных справочных систем

1. Свободно распространяемое программное обеспечение Google Colab

12 Материально-техническое обеспечение дисциплины

1. Проектор Epson EB-W04LCD.WXGA 1280*800.3000:1.2800 ANSI Lumens
2. Персональный компьютер