

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ
Федеральное государственное бюджетное образовательное учреждение
высшего образования
«КУБАНСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ»
Факультет компьютерных технологий и прикладной математики

УТВЕРЖДАЮ:

Проректор по учебной работе,
качеству образования – первый
проректор

Хагуров Т.А.

подпись

«02» июня 2026 г

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ

Б1.О.27 Основы прикладной статистики

Направление подготовки 02.03.03 Математическое обеспечение и администрирование информационных систем

Направленность Искусственный интеллект и аналитика данных

Форма обучения _____ очная

Квалификация выпускника _____ бакалавр

Краснодар 2026

Рабочая программа дисциплины «Основы прикладной статистики» составлена в соответствии с федеральным государственным образовательным стандартом высшего образования (ФГОС ВО) по направлению 09.03.03 Прикладная информатика

Программу составил:

Халафян А.А., доцент, д-р техн. наук

И.О. Фамилия, должность, ученая степень, ученое звание

подпись

Рабочая программа дисциплины утверждена на заседании центра искусственного интеллекта

протокол № 04 «14» мая 2026 г.

Руководитель центра ИИ Коваленко А.В.

подпись

Утверждена на заседании учебно-методической комиссии факультета компьютерных технологий и прикладной математики
протокол № 03 «15» мая 2026 г.

Председатель УМК факультета Добровольская Н. Ю.

подпись

Рецензенты:

Мостовой Евгений Викторович, генеральный директор ООО «Портал-Юг»,
e-mail: mostovoy@portal-yug.ru

Луценко Евгений Вениаминович, доктор экономических наук, кандидат технических наук, профессор кафедры компьютерных технологий и систем Федерального государственного бюджетного образовательного учреждения высшего образования «Кубанский государственный аграрный университет имени И.Т. Трубилина», e-mail: prof.lutsenko@gmail.com

1. Цели и задачи учебной дисциплины

1.1 Цель освоения дисциплины.

Цели определены государственным образовательным стандартом высшего образования и соотнесены с общими целями ООП ВО по направлению подготовки «02.03.03 Математическое обеспечение и администрирование информационных систем», в рамках которой преподается дисциплина.

Целью освоения учебной дисциплины «Основы прикладной статистики» является изучение основных методов многомерного статистического анализа данных с точки зрения их практического применения; формирование навыков работы с соответствующими разделами ППП STATISTICA.

1.2 Задачи дисциплины:

- помочь студентам понять и, освоить методологию методов прикладной статистики;
- привить теоретические и практические знания в области прикладного многомерного анализа данных;
- познакомить студентов и обучить максимально широкому инструментарию статистического анализа данных в среде ППП STATISTICA;
- выработать в процессе обучения у студентов навыки грамотного использования аппарата вероятностно-статистического моделирования, многомерного анализа посредством применения передовых информационных технологий.

1.3 Место дисциплины (модуля) в структуре образовательной программы

Дисциплина «Основы прикладной статистики» относится к дисциплинам обязательной части Блока 1 Дисциплины (модули).

Данная дисциплина тесно связана с дисциплинами: «Математический анализ», «Векторная алгебра», «Математические методы и модели исследования операций».

Материал курса предназначен для использования в дисциплинах, связанных с количественным анализом реальных экономических явлений, таких как, например, «Математические модели анализа экономических субъектов», «Теория риска и моделирование рискованных ситуаций» и др.

В совокупности изучение этой дисциплины готовит студентов, как к различным видам практической деятельности, так и к научно-теоретической, исследовательской работе.

Особо следует подчеркнуть, что многомерный статистический анализ лежит в основе методов ИИ. Ниже перечислены наиболее часто используемые разделы прикладной статистики в методах ИИ.

РСА (метод главных компонент) – уменьшение размерности.

Кластерный анализ (k-means, hierarchical clustering) – неконтролируемое обучение (без учителя).

Факторный анализ – сокращение размерности данных и выявление скрытых переменных.

Дискриминантный анализ, общие модели дискриминантного анализа, деревья классификации, логистическая регрессия – в решении задач классификации (контролируемое обучение) методами машинного обучения.

Регрессионный анализ, общие линейные модели в решении задач регрессии методами машинного обучения.

Дискриминантный, кластерный, корреляционный анализ – в выявлении признаков построения моделей машинного обучения с высокими прогностическими свойствами.

1.4 Перечень планируемых результатов обучения по дисциплине (модулю), соотнесенных с планируемыми результатами освоения образовательной программы

Роль 1: Data Analyst (Аналитик данных)

Задачи:

1. Статистический анализ, визуализация данных, предварительная обработка.
2. Создание прогнозных моделей
3. Построение аналитических моделей для поддержки бизнес-решений.

Роль 2: MLOps (Специалист по эксплуатации ИИ)

Задачи:

1. DevOps для ML.
2. Автоматизация, мониторинг ML-систем.
3. Операционное управление жизненным циклом ML-моделей.

Роль 3: AI PM (Менеджер проектов ИИ)

Задачи:

1. Управление ИИ-проектами от идеи до внедрения
2. Анализ бизнес-требований и постановка задач
3. Оценка эффективности и ROI ИИ-решений

Изучение данной учебной дисциплины направлено на формирование у обучающихся следующих компетенций:

Код и наименование индикатора*	Результаты обучения по дисциплине
ОПК-2 Способен применять современный математический аппарат, связанный с проектированием, разработкой, реализацией и оценкой качества программных продуктов и программных комплексов в различных областях человеческой деятельности	
ОПК-2.2 Применяет современный математический аппарат при построении моделей в различных областях человеческой деятельности	Знать: Регрессионный, дисперсионный, кластерный анализ, критерии качества моделей (R^2 , точность). Уметь: Строить модели под задачу (регрессию, логистическую регрессию, кластеризацию), оценивать параметры, интерпретировать результаты. Владеть: Построением и оценкой статистических моделей в Statistica, на Python (statsmodels, sklearn) для разных предметных областей
MF-1 Способен применять современную теоретическую математику для разработки новых алгоритмов и формулирования перспективных задач искусственного интеллекта	
MF-1.2 Применяет аппарат теории вероятностей, матстатистики и теории информации для формулирования и анализа задач искусственного интеллекта	Знать: Основы теории вероятностей, матстатистики и теории информации для анализа данных ИИ Уметь: Оценивать распределения, выявлять зависимости, применять энтропию и взаимную информацию Владеть: Выбором и применением статистико-информационного аппарата для задач ИИ
MF-2 Способен применять байесовский подход для построения вероятностных моделей анализа неопределенности и создания адаптивных систем ИИ	
MF-2.1 Использует теорему Байеса и её следствия, понимает отличия байесовской статистики от частотного подхода	Знать: Теорему Байеса, априорное/апостериорное распределения, отличие от частотного подхода Уметь: Вычислять апостериорное распределение и объяснять различия подходов на примерах Владеть: Байесовским выводом и сравнением с частотными методами для задач ИИ
MF-2.2 Применяет байесовские методы оценивания и байесовские интервалы для решения задач статистики и построения моделей.	Знать: Определения байесовских точечных оценок (апостериорное среднее, апостериорная мода), байесовских интервалов (кредибельные интервалы, HPD-интервалы — наибольшей апостериорной плотности), а также условия их применения для построения вероятностных

Код и наименование индикатора*	Результаты обучения по дисциплине
	моделей в задачах статистики и ИИ Уметь: Вычислять байесовские оценки параметров (апостериорное среднее, медиану, моду) и строить байесовские интервалы для одно- и многопараметрических моделей, интерпретировать их как вероятностные утверждения о параметре (в отличие от доверительных интервалов в частотной статистике) Владеть: Навыками применения байесовских методов оценивания для построения вероятностных моделей в задачах ИИ (включая биномиальные, нормальные, иерархические модели), обоснованного выбора типа байесовского интервала (равнохвостый или HPD) и сравнения полученных результатов с классическими статистическими подходами
BD-1 Способен осуществлять поиск сбор очистку и предварительный анализ данных	
BD-1.3 Применяет методы понижения размерности для первичной интерпретации и визуализации многомерных данных	Знать: Цели понижения размерности, методы (PCA, t-SNE, UMAP), критерии выбора числа компонент Уметь: Применять PCA, t-SNE, UMAP в Python (sklearn), интерпретировать нагрузки и визуализировать Владеть: Выбором подходящего метода (PCA или t-SNE или UMAP) для визуализации и интерпретации многомерных данных
BD-2 Способен определять требования к наборам данных для решения задач машинного обучения проводить разметку и анализ наборов данных оценивать качество данных обеспечивать непрерывную интеграцию данных	
BD-2.1 Определяет требования к наборам и качеству данных для решения задач машинного обучения	Знать: Критерии качества данных (точность, полнота, согласованность) и зависимость требований от типа задачи ML Уметь: Формулировать требования к признакам, объёму, разметке и допустимым дефектам данных под конкретную задачу Владеть: Составлением спецификации требований к данным (Data Requirements) и оценкой пригодности существующих наборов

Результаты обучения достигаются в рамках осуществления всех видов контактной и самостоятельной работы обучающихся в соответствии с утвержденным учебным планом.

Индикаторы достижения компетенций считаются сформированными при достижении соответствующих им результатов обучения.

2. Структура и содержание дисциплины

2.1 Распределение трудоемкости дисциплины по видам работ

Общая трудоёмкость дисциплины составляет 4 зач.ед. (144 часов), их распределение по видам работ представлено в таблице
(для студентов ОФО).

Вид учебной работы	Всего часов	Семестр (часы)
		5
Контактная работа, в том числе:	77,3	77,3
Аудиторные занятия (всего):	68	68
Занятия лекционного типа	34	34
Занятия семинарского типа (семинары, практические занятия)	—	—
Лабораторные занятия	34	34
Иная контактная работа:	9,3	9,3

Вид учебной работы	Всего часов	Семестр (часы)
		5
Контроль самостоятельной работы (КСР)	9	9
Промежуточная аттестация (ИКР)	0,3	0,3
Самостоятельная работа, в том числе	31	31
Курсовая работа	10	10
Проработка учебного (теоретического) материала	10	8
Подготовка к текущему контролю	11	11
Контроль:		
Подготовка к экзамену	35,7	35,7
Общая трудоемкость	час.	144
	в том числе контактная работа	77,3
	зач. ед	4

2.2 Структура учебной дисциплины

Распределение видов учебной работы и их трудоемкости по разделам дисциплины.

Разделы дисциплины, изучаемые в 5 семестре:

№ п/п	Наименование раздела, темы	Количество часов			
		Всего	Аудиторная работа		Внеаудиторная работа
			Л	ЛР	
1.	Элементарные понятия статистики. Измерительные шкалы	6	4	2	–
2.	Работа с данными. Основные операции над случаями и переменными	7	2	2	3
3.	Основные статистики, корреляционный анализ, графика	4	2	2	–
4.	Подгонка закона распределения	4	2	2	–
5.	Сравнение средних величин критерием Стьюдента	4	2	2	–
6.	Группировка и однофакторная ANOVA	7	2	2	3
7.	Дисперсионный анализ	4	2	2	–
8.	Непараметрическая статистика. Сравнение средних величин методами непараметрической статистики	7	2	2	3
9.	Таблицы частот, сопряженности, флагов и заголовков	8	2	2	4
10.	Линейное и нелинейное моделирование взаимосвязей	4	2	2	–
11.	Кластерный анализ	8	2	2	4
12.	Дискриминантный анализ	8	2	2	4
13.	Деревья классификации	7	2	2	3
14.	Факторный анализ	9	2	2	5
15.	Канонический анализ. Многомерное шкалирование	6	2	2	2
16.	Обзор пройденного материала.	6	2	4	–
Контроль самостоятельной работы (КСР)		9			
Промежуточная аттестация (ИКР)		0,3			
Подготовка к текущему контролю		35,7			

№ п/ п	Наименование раздела, темы	Количество часов			
		Всего	Аудиторная работа		Внеаудиторная работа
			Л	ЛР	
Общая трудоемкость по дисциплине:		144	34	34	31

Примечание: Л – лекции, ЛР – лабораторные занятия, СРС – самостоятельная работа студента

2.3 Содержание разделов (тем) дисциплины

2.3.1 Занятия лекционного типа

№	Наименование раздела	Содержание раздела	Форма текущего контроля
1	Основные приемы работы с ППП STATISTICA	Тема 1. Элементарные понятия статистики. Измерительные шкалы. Тема 2. Работа с данными. Основные операции над случаями и переменными	Вопросы к экзамену 1-4
2	Основные статистики и методы сравнения средних величин	Тема 3. Основные статистики, корреляционный анализ, подгонка закона распределения Тема 4. Графика Тема 5. Сравнение средних величин критерием Стьюдента Тема 6. Группировка и однофакторная ANOVA Тема 7. Дисперсионный анализ Тема 8. Непараметрическая статистика. Сравнение средних величин методами непараметрической статистики	Вопросы к экзамену 5-12
3	Методы многомерного анализа	Тема 9. Таблицы частот, сопряженности, флагов и заголовков. Тема 10. Линейное и нелинейное моделирование взаимосвязей Тема 11. Дискриминантный анализ Тема 12. Кластерный анализ Тема 13. Деревья классификации Тема 14. Факторный анализ. Тема 15. Многомерное шкалирование, канонический анализ Тема 16. Обзор пройденного материала.	Вопросы к экзамену 13-27

2.3.2 Занятия семинарского типа/Лабораторные занятия

№	Наименование раздела	Содержание раздела	Форма текущего контроля
---	----------------------	--------------------	-------------------------

1	Основные приемы работы с ППП STATISTICA	Тема 1. Элементарные понятия статистики. Измерительные шкалы. Тема 2. Работа с данными. Основные операции над случаями и переменными	Проверка домашнего задания Контрольная работа
2	Основные статистики и методы сравнения средних величин	Тема 3. Основные статистики, корреляционный анализ, подгонка закона распределения Тема 4. Графика Тема 5. Сравнение средних величин критерием Стьюдента Тема 6. Группировка и однофакторная ANOVA Тема 7. Дисперсионный анализ Тема 8. Непараметрическая статистика. Сравнение средних величин методами непараметрической статистики	Проверка домашнего задания Контрольная работа
3	Методы многомерного анализа	Тема 9. Таблицы частот, сопряженности, флагов и заголовков. Тема 10. Линейное и нелинейное моделирование взаимосвязей Тема 11. Дискриминантный анализ Тема 12. Кластерный анализ Тема 13. Деревья классификации Тема 14. Факторный анализ. Тема 15. Многомерное шкалирование, канонический анализ Тема 16. Обзор пройденного материала.	Проверка домашнего задания Тестирование

2.3.3 Примерная тематика курсовых проектов

1. Статистический анализ вторичного рынка отечественных автомобилей с факторизацией пространства количественных признаков.
2. Многомерный анализ вторичного рынка автомобилей иностранного производства с кластеризацией пространства количественных признаков
3. Прогнозирование заработной платы сотрудников компании методами регрессионного анализа
4. Общие линейные модели как обобщение регрессионного анализа
5. Общие модели дискриминантного анализа как метод дискриминации с категориальными переменными
6. Позиционный анализ как метод оценки согласованности экспертов
7. Анализ экономической деятельности банков методами многомерного анализа
8. Анализ эффективности вузов методами многомерного анализа
9. Метод главных компонент в средах STATISTICA, R, преимущества и недостатки
10. Факторный анализ в средах STATISTICA, Python, R, преимущества и недостатки

11. Кластерный анализ в средах STATISTICA, Python, R, преимущества и недостатки
12. Деревья классификации в средах STATISTICA, Python, R преимущества и недостатки

Рекомендации

- Регулярно консультироваться с научным руководителем.
- Использовать современные источники литературы и открытые данные.
- Проверять работу на уникальность (оригинальность текста должна быть не менее 70%).

Методические указания помогут студентам успешно выполнить курсовой проект и продемонстрировать свои навыки в области многомерного статистического анализа.

2.4 Перечень учебно-методического обеспечения для самостоятельной работы обучающихся по дисциплине (модулю)

№	Вид СРС	Перечень учебно-методического обеспечения дисциплины по выполнению самостоятельной работы
1	2	3
1	Проработка и повторение лекционного материала, материала учебной и научной литературы, подготовка к семинарским занятиям	Методические указания для подготовки к лекционным и семинарским занятиям, утвержденные на заседании кафедры анализа данных и искусственного интеллекта факультета компьютерных технологий и прикладной математики ФГБОУ ВО «КубГУ», протокол №7 от 22.03.2023 г. Методические указания по выполнению самостоятельной работы, утвержденные на заседании кафедры анализа данных и искусственного интеллекта факультета компьютерных технологий и прикладной математики ФГБОУ ВО «КубГУ», протокол №7 от 22.03.2023 г.
2	Подготовка к лабораторным занятиям	Методические указания по выполнению лабораторных работ, утвержденные на заседании кафедры анализа данных и искусственного интеллекта факультета компьютерных технологий и прикладной математики ФГБОУ ВО «КубГУ», протокол №7 от 22.03.2023 г.
3	Подготовка к решению задач и тестов	Методические указания по выполнению самостоятельной работы, утвержденные на заседании кафедры анализа данных и искусственного интеллекта факультета компьютерных технологий и прикладной математики ФГБОУ ВО «КубГУ», протокол №7 от 22.03.2023 г.
4	Подготовка к текущему контролю	Методические указания по выполнению самостоятельной работы, утвержденные на заседании кафедры анализа данных и искусственного интеллекта факультета компьютерных технологий и прикладной математики ФГБОУ ВО «КубГУ», протокол №7 от 22.03.2023 г.
5	Подготовка докладов	Методические указания для подготовки эссе, рефератов, курсовых работ, утвержденные на заседании кафедры анализа данных и искусственного интеллекта факультета компьютерных технологий и прикладной математики ФГБОУ ВО «КубГУ», протокол №7 от 22.03.2023 г.

Учебно-методические материалы для самостоятельной работы обучающихся из числа инвалидов и лиц с ограниченными возможностями здоровья (ОВЗ) предоставляются в формах, адаптированных к ограничениям их здоровья и восприятия информации:

Для лиц с нарушениями зрения:

- в печатной форме увеличенным шрифтом,
- в форме электронного документа,
- в форме аудиофайла,
- в печатной форме на языке Брайля.

Для лиц с нарушениями слуха:

- в печатной форме,
- в форме электронного документа.

Для лиц с нарушениями опорно-двигательного аппарата:

- в печатной форме,
- в форме электронного документа,
- в форме аудиофайла.

Данный перечень может быть конкретизирован в зависимости от контингента обучающихся.

3. Образовательные технологии

С точки зрения применяемых методов используются как традиционные информационно-объяснительные лекции, так и интерактивная подача материала с мультимедийной системой. Компьютерные технологии в данном случае обеспечивают возможность разнопланового отображения алгоритмов и демонстрационного материала. Такое сочетание позволяет оптимально использовать отведенное время и раскрывать логику и содержание дисциплины.

Лекции представляют собой систематические обзоры задач прикладной статистики с подачей материала в виде презентаций.

Лабораторное занятие позволяет научить студента применять теоретические знания при решении и исследовании конкретных задач. Лабораторные занятия проводятся в компьютерных классах, при этом практикуется работа в группах. Подход разбора конкретных ситуаций широко используется как преподавателем, так и студентами при проведении анализа результатов самостоятельной работы. Это обусловлено тем, что в процессе исследования часто встречаются задачи, для которых единых подходов не существует. Каждая конкретная задача при своем исследовании имеет множество подходов, а это требует разбора и оценки целой совокупности конкретных ситуаций.

Для лиц с ограниченными возможностями здоровья предусмотрена организация консультаций с использованием электронной почты.

4. Оценочные и методические материалы

4.1 Оценочные средства для текущего контроля успеваемости и промежуточной аттестации

Оценочные средства предназначены для контроля и оценки образовательных достижений обучающихся, освоивших программу учебной дисциплины «многомерный статистический анализ».

Оценочные средства включает контрольные материалы для проведения текущего контроля в форме домашних заданий и выполнения самостоятельной работы и промежуточной аттестации в виду зачета.

Оценочные средства для инвалидов и лиц с ограниченными возможностями здоровья выбираются с учетом их индивидуальных психофизических особенностей.

– при необходимости инвалидам и лицам с ограниченными возможностями здоровья предоставляется дополнительное время для подготовки ответа на экзамене;

– при проведении процедуры оценивания результатов обучения инвалидов и лиц с ограниченными возможностями здоровья предусматривается использование технических средств, необходимых им в связи с их индивидуальными особенностями;

– при необходимости для обучающихся с ограниченными возможностями здоровья и инвалидов процедура оценивания результатов обучения по дисциплине может проводиться в несколько этапов.

Процедура оценивания результатов обучения инвалидов и лиц с ограниченными возможностями здоровья по дисциплине (модулю) предусматривает предоставление информации в формах, адаптированных к ограничениям их здоровья и восприятия информации:

Для лиц с нарушениями зрения:

- в печатной форме увеличенным шрифтом,
- в форме электронного документа.

Для лиц с нарушениями слуха:

- в печатной форме,
- в форме электронного документа.

Для лиц с нарушениями опорно-двигательного аппарата:

- в печатной форме,
- в форме электронного документа.

Данный перечень может быть конкретизирован в зависимости от контингента обучающихся.

Структура оценочных средств для текущей и промежуточной аттестации

№ п/п	Контролируемые разделы (темы) дисциплины	Код контролируемой компетенции (или ее части)	Наименование оценочного средства	
			Текущий контроль	Промежуточная аттестация
1	Элементарные понятия статистики. Измерительные шкалы.	ОПК-2.2, MF-1.2, BD-1.3	УО, ЛР	ЭКВ (1)
2	Работа с данными. Основные операции над случаями и переменными	MF-1.2, MF-2.2, BD-1.3, BD-2.1	УО, ПДР, ЛР	ЭКВ (1-3)
3	Описательные статистики. Корреляционная матрица	ОПК-2.2, MF-1.2, MF-2.2, BD-1.3	УО, ЛР	ЭКВ (6-7)
4	Подгонка закона распределения	ОПК-2.2, MF-1.2, MF-2.2, BD-2.1	УО, ПДР, ЛР	ЭКВ (8)
5	Сравнение двух средних t-критерием для независимых и зависимых выборок	MF-2.2, BD-1.3	УО, ПДР, ЛР	ЭКВ (9)
6	Группировка и однофакторная ANOVA	ОПК-2.2, MF-1.2,	УО, ПДР, ЛР	ЭКВ (10)
7	Дисперсионный анализ	ОПК-2.2, MF-1.2, MF-2.2, BD-1.3	УО, ПДР, ЛР	ЭКВ (10)
8	Таблицы частот, сопряженности, флагов и заголовков	ОПК-2.2, MF-2.2, BD-1.3	УО, ПДР, ЛР	ЭКВ (13)
9	Непараметрическая статистика	ОПК-2.2, MF-1.2, BD-2.1	УО, ПДР	ЭКВ (12)
10	Линейное и нелинейное моделирование взаимосвязей	ОПК-2.2, MF-1.2, MF-2.2	УО, ПДР	ЭКВ (15-21)
11	Дискриминантный анализ	ОПК-2.2, MF-1.2, MF-2.2, BD-2.1	УО, ПДР, ЛР	ЭКВ(23)
12	Кластерный анализ	MF-1.2, MF-2.2,	УО, ПДР, ЛР	ЭКВ(22)
13	Деревья классификации	ОПК-2.2, MF-1.2,	УО, ПДР, ЛР	ЭКВ(25)
14	Факторный анализ.	ОПК-2.2, MF-1.2, MF-2.2, BD-1.3	УО, ПДР, ЛР	ЭКВ(24)
15	Многомерное шкалирование, канонический анализ	MF-1.2, BD-1.3, BD-2.1	УО, ПДР, ЛР	ЭКВ(27)

Сокращения: УО – устный опрос, ПДР – проверка самостоятельной, КР – контрольная работа, ЛР – лабораторная работа, ЭКВ – экзаменационные вопросы

Показатели, критерии и шкала оценки сформированных компетенций

Код и наименование индикатора компетенции	Соответствие уровней освоения компетенции планируемым результатам обучения и критериям их оценивания		
	пороговый	базовый	продвинутый
	Оценка		
	Удовлетворительно / зачтено	Хорошо / зачтено	Отлично / зачтено
ОПК-2.2 Применяет современный математический аппарат при построении моделей в различных областях человеческой деятельности	Воспроизводит основные понятия и формулы: линейная регрессия, коэффициент детерминации R^2 , метод наименьших квадратов. Строит простейшие модели по образцу (парная линейная регрессия, базовая кластеризация k-means). Интерпретирует результаты на уровне «коэффициент положительный/отрицательный», но испытывает затруднения при оценке качества модели и проверке условий применимости.	Самостоятельно строит статистические модели (множественная регрессия, логистическая регрессия, кластеризация, ANOVA) в зависимости от типа задачи и структуры данных. Оценивает качество моделей (R^2 , точность, полнота, F1-мера), проверяет условия применимости (нормальность остатков, мультиколлинеарность). Интерпретирует коэффициенты и статистическую значимость факторов.	Свободно применяет современный математический аппарат для построения сложных статистических моделей в нестандартных или междисциплинарных задачах. Критически сравнивает альтернативные модели с использованием критериев. Обосновывает выбор модели с учётом особенностей предметной области, ограничений данных и целей исследования.
МФ-1.2 Применяет современный математический аппарат теории вероятностей для исследования методов и моделей машинного обучения	Знает математические основы методов (теория вероятностей и статистики) для анализа данных. Может воспроизвести и применить ключевые формулы.	Умеет применять математический аппарат (теория вероятностей, мат. статистика) для исследования и сравнения моделей машинного обучения (оценка параметров, проверка гипотез). Может модифицировать известные алгоритмы.	Владеет навыками самостоятельного выбора и применения современного математического аппарата для исследования и модификации алгоритмов МО в рамках перспективных задач ИИ.

Код и наименование индикатора компетенции	Соответствие уровней освоения компетенции планируемым результатам обучения и критериям их оценивания		
	пороговый	базовый	продвинутый
	Оценка		
	Удовлетворительно / зачтено	Хорошо / зачтено	Отлично / зачтено
MF-2.2 Применяет байесовские методы оценивания и байесовские интервалы для решения задач статистики и построения моделей	Знает основные принципы и формулы байесовского вывода (априорное/апостериорное распределение). Может применить их для простых случаев (оценка доли).	Умеет применять байесовские методы (оценивание параметров, построение доверительных интервалов) для решения стандартных задач анализа данных и машинного обучения.	Владеет навыками построения сложных байесовских моделей (иерархические, с использованием метода Монте-Карло). Способен применять их для создания адаптивных систем и решения сложных статистических задач.
BD-1.3 Применяет методы понижения размерности для первичной интерпретации и визуализации многомерных данных	Знает основные методы понижения размерности (Факторный, Канонический анализ). Может применить один метод по инструкции для визуализации данных.	Умеет применять методы понижения размерности (Факторный и Кластерный анализ) для первичной интерпретации и визуализации данных. Может оценить и сравнить качество снижения размерности разными методами.	Владеет навыками сравнительного анализа и выбора оптимального метода понижения размерности для конкретных данных и задач. Способен проводить углубленную интерпретацию результатов и интегрировать их в процесс моделирования сложных взаимосвязей
BD-2.1 Определяет требования к наборам и качеству данных для решения задач машинного обучения	Знает основные критерии качества данных (полнота, точность, отсутствие дубликатов). Может перечислить базовые требования к данным для типовых задач ML (классификация, регрессия) по шаблону, но испытывает затруднения при самостоятельной	Самостоятельно определяет требования к набору данных на основе постановки задачи ML: обосновывает необходимые признаки (целевые, входные), их типы (числовые, категориальные, текстовые), минимальный объём выборки (с учётом числа признаков и сложности модели).	Проводит глубокий анализ требований к данным для сложных и нестандартных задач ML (обучение с подкреплением, временные ряды, обработка естественного языка). Учитывает компромиссы между качеством данных и затратами на их сбор/разметку. Формулирует

Код и наименование индикатора компетенции	Соответствие уровней освоения компетенции планируемым результатам обучения и критериям их оценивания		
	пороговый	базовый	продвинутый
	Оценка		
	Удовлетворительно / зачтено	Хорошо / зачтено	Отлично / зачтено
	формулировке требований для новой задачи.	Формулирует критерии качества данных и предлагает способы их проверки (доли пропусков, допустимые выбросы, форматы данных).	количественные критерии качества с обоснованием (например, «доля пропусков не более 5%», «точность разметки не менее 95%»).

Соответствие курсовых работ индикаторам компетенций

Тема	Индикатор
Статистический анализ вторичного рынка отечественных автомобилей с факторизацией пространства количественных признаков.	ОПК-2.2
Многомерный анализ вторичного рынка автомобилей иностранного производства с кластеризацией пространства количественных признаков	MF-1.2
Прогнозирование заработной платы сотрудников компании методами регрессионного анализа	BD-1.3
Общие линейные модели как обобщение регрессионного анализа	ОПК-2.2
Общие модели дискриминантного анализа как метод дискриминации с категориальными переменными	MF-2.2
Позиционный анализ как метод оценки согласованности экспертов	BD-2.1
Анализ экономической деятельности банков методами многомерного анализа	MF-2.2
Анализ эффективности вузов методами многомерного анализа	ОПК-2.2
Метод главных компонент в средах STATISTICA, R, преимущества и недостатки	MF-1.2
Факторный анализ в средах STATISTICA, Python, R, преимущества и недостатки	BD-1.3
Кластерный анализ в средах STATISTICA, Python, R, преимущества и недостатки	MF-2.2

Деревья классификации в средах STATISTICA, Python, R преимущества и недостатки	MF-1.2
--	--------

В результате студент получит:

- Навыки работы с математическим софтом (STATISTICA, Python, R)
- Понимание, как методы многомерного анализа применяются в решении прикладных задач
- Понимание, как методы классификации, применяемые в ИИ использовались в многомерном анализе

Зачетно-экзаменационные материалы для промежуточной аттестации (экзамен/зачет)

Примерный перечень вопросов к экзамену (перечень компетенции)

1. Инструменты для работы с данными (ОПК-2.2, MF-1.2)
2. Структура электронной таблицы (MF-2.2, MF-1.2)
3. Основные операции над переменными и случаями (MF-2.2, MF-1.2)
4. Основные операции с таблицами данных (ОПК-2.2, MF-1.2)
5. Двухмерная и трехмерная графика (MF-2.2, BD-2.1)
6. Описательные статистики (MF-2.2, BD-1.3)
7. Корреляционная матрица (MF-2.2, BD-1.3)
8. Подгонка закона распределения (MF-2.2, MF-1.2)
9. t-критерий сравнения средних (ОПК-2.2, MF-1.2)
10. Группировка и однофакторная ANOVA, дисперсионный анализ (MF-2.2, BD-1.3)
11. Непараметрическая корреляция (MF-2.2, MF-1.2)
12. Непараметрические методы сравнения (MF-2.2, MF-1.2)
13. Таблицы частот, кросстабуляции, флагов и заголовков (ОПК-6.2, MF-1.2)
14. Линейная регрессионная модель (MF-2.2, BD-1.3)
15. Модуль множественная регрессия (ОПК-2.2, MF-2.2)
16. Линеаризующие преобразования (ОПК-2.2, MF-2.2)
17. Модели бинарных откликов (ОПК-2.2, MF-1.2)
18. Описание модуля Нелинейное оценивание (ОПК-2.2, MF-2.2)
19. Экспоненциальная регрессия (MF-2.2, MF-1.2)
20. Кусочно-линейная регрессия (MF-2.2, BD-2.1)
21. Определенная пользователем регрессия (MF-2.2, MF-1.2)
22. Кластерный анализ. Описание модуля Кластерный анализ (ОПК-2.2, MF-1.2).
23. Дискриминантный анализ. Описание модуля Дискриминантный анализ (MF-2.2, BD-1.3)
24. Задача факторного анализа. Описание модуля Факторный анализ (ОПК-2.2, MF-2.2)
25. Деревья классификации (MF-2.2, MF-1.2)
26. Задача канонического анализа. Описание метода. Модуль Канонический анализ (MF-2.2, BD-1.3)
27. Многомерное шкалирование, канонический анализ (ОПК-2.2, MF-2.2)

4.2. Типовые контрольные задания или иные материалы, необходимые для оценки знаний, умений, навыков и (или) опыта деятельности, характеризующих этапы формирования компетенций в процессе освоения образовательной программы

По теме: Элементарные понятия статистики

Контрольные вопросы (ОПК-2.2, MF-2.2):

- 1). Дайте определение математического ожидания, дисперсии, медианы, моды и т.д.

- 2). Какие измерительные шкалы вы знаете?
- 3). Чем отличается шкала наименований от порядковой шкалы?

По теме: Работа с данными.

Контрольные вопросы (MF-1.2, BD-1.3):

- 1). Какие способы существуют объединения таблиц в STATISTICA?
- 2). Как осуществить перекодировку переменных, зачем она нужна?
- 3). Какие существуют способы ранжирования значений переменной?
- 4). Как реализована сортировка переменных в STATISTICA?

По теме: Основные статистики, корреляционный анализ, подгонка закона распределения (ОПК-2.2, BD-1.3):

- 1). Какие методы подгонки распределений реализованы в STATISTICA?
- 2). Опишите последовательность действий для оценки взаимосвязей и сформулируйте нулевую гипотезу.

По теме: t- критерий сравнения средних

Контрольные вопросы (MF-1.2, MF-2.2):

- 1). В каких случаях рекомендовано применение t-критерия?
- 2). Какие модификации t-критерия вы знаете?
- 3). Почему метод называется параметрическим?

По теме: Однофакторный дисперсионный анализ.

Контрольные вопросы (ОПК-2.2, BD-1.3):

- 1). В каких случаях рекомендовано применение t-критерия?
- 2). Какие модификации t-критерия вы знаете?
- 3). Почему метод называется параметрическим?

По теме: Дисперсионный анализ.

Контрольные вопросы (MF-1.2, BD-2.1):

- 1). Какую основную статистическую задачу решает дисперсионный анализ.
- 2). Какие преимущества дисперсионного анализа перед t-критерием.
- 3). Существуют ли ограничения в применении дисперсионного анализа, если да, то перечислите их.

По теме: Непараметрическая статистика.

Контрольные вопросы (MF-2.2, BD-2.1):

- 1). Перечислите коэффициенты непараметрической корреляции.
- 2). Когда применяется критерий Манна-Уитни?
- 3). Когда применяется критерий Краскела-Уоллиса

По теме: Таблицы частот, сопряженности, флагов и заголовков.

Контрольные вопросы (ОПК-2.2, MF-1.2, BD-1.3):

- 1). В каких случаях целесообразно применение таблиц частот.
- 2). Что такое таблица сопряженности?
- 3). Как при помощи таблиц сопряженности исследовать зависимость между категориальными переменными?

По теме: Анализ взаимосвязи между группами переменных.

Контрольные вопросы (ОПК-2.2, MF-1.2, BD-1.3):

- 1). Что такое каноническая корреляция?
- 2). Что такое канонический корень?

3). Каким образом определяется значимость канонической корреляции?

По теме: Линейное и нелинейное моделирование взаимосвязей.

Контрольные вопросы (ОПК-2.2, MF-2.2, BD-2.1):

- 1). Что называется откликом, предиктором?
- 2). Как построить систему нормальных уравнений?
- 3). Как оценить направление и вклад предикторов в значение отклика?
- 4). Перечислить линеаризующие преобразования. Для чего они нужны?

По теме: Методы классификации с обучением.

Контрольные вопросы (MF-1.2, MF-2.2, BD-1.3):

- 1). Почему дискриминантный анализ называется классификация с обучением?
- 2). Постановка задачи дискриминантного анализа?
- 3). Какие ограничения существуют на переменные?
- 4). Для чего нужна статистика лямбда Уилкса.

По теме: Кластерный анализ

Контрольные вопросы (ОПК-2.2, MF-1.2, MF-2.2):

- 1). Чем отличается дискриминантный анализ от кластерного?
- 2). Постановка задачи кластерного анализа?
- 3). Какие ограничения существуют на переменные?
- 4). Как оценивается качество кластеризации.

По теме: Деревья классификации.

Контрольные вопросы (MF-1.2, MF-2.2, BD-2.1):

- 1). Отличие метода от других методов классификации
- 2). Как строится бинарное дерево
- 3). Что такое цена правильной классификации

По теме: Факторный анализ.

Контрольные вопросы (MF-2.2, BD-1.3):

- 1). Постановка задачи факторного анализа?
- 2). Какие ограничения существуют на переменные?
- 3). Для чего предназначен метод каменистой осыпи, критерий Кайзера?
- 4). Каково предназначение факторных нагрузок?

По теме: Многомерное шкалирование, канонический анализ

Контрольные вопросы (MF-1.2, BD-2.1):

- 1). Цель многомерного шкалирования?
- 2). чем отличается канонический анализ от множественного корреляционного анализа?
- 3). Основные предположение многомерного шкалирования?

Критерии выставления оценок на экзамене

Оценка «отлично»:

- систематизированные, глубокие и полные знания по всем разделам дисциплины, а также по основным вопросам, выходящим за пределы учебной программы;
- точное использование научной терминологии систематически грамотное и логически правильное изложение ответа на вопросы;
- безупречное владение инструментарием учебной дисциплины, умение его эффективно использовать в постановке научных и практических задач;
- выраженная способность самостоятельно и творчески решать сложные проблемы и нестандартные ситуации;

- полное и глубокое усвоение основной и дополнительной литературы, рекомендованной учебной программой по дисциплине;
- умение ориентироваться в теориях, концепциях и направлениях дисциплины и давать им критическую оценку, используя научные достижения других дисциплин;
- творческая самостоятельная работа на практических/семинарских/лабораторных занятиях, активное участие в групповых обсуждениях, высокий уровень культуры исполнения заданий;
- высокий уровень сформированности заявленных в рабочей программе компетенций.

Оценка «хорошо»:

- достаточно полные и систематизированные знания по дисциплине;
- умение ориентироваться в основном теориях, концепциях и направлениях дисциплины и давать им критическую оценку;
- использование научной терминологии, лингвистически и логически правильное изложение ответа на вопросы, умение делать обоснованные выводы;
- владение инструментарием по дисциплине, умение его использовать в постановке и решении научных и профессиональных задач;
- усвоение основной и дополнительной литературы, рекомендованной учебной программой по дисциплине;
- самостоятельная работа на практических занятиях, участие в групповых обсуждениях, высокий уровень культуры исполнения заданий;
- средний уровень сформированности заявленных в рабочей программе компетенций.

Оценка «удовлетворительно»:

- достаточный минимальный объем знаний по дисциплине;
- усвоение основной литературы, рекомендованной учебной программой;
- умение ориентироваться в основных теориях, концепциях и направлениях по дисциплине и давать им оценку;
- использование научной терминологии, стилистическое и логическое изложение ответа на вопросы, умение делать выводы без существенных ошибок;
- владение инструментарием учебной дисциплины, умение его использовать в решении типовых задач;
- умение под руководством преподавателя решать стандартные задачи;
- работа под руководством преподавателя на практических занятиях, допустимый уровень культуры исполнения заданий;
- достаточный минимальный уровень сформированности заявленных в рабочей программе компетенций.

Оценка «неудовлетворительно»:

- фрагментарные знания по дисциплине;
- отказ от ответа (выполнения письменной работы);
- знание отдельных источников, рекомендованных учебной программой по дисциплине;
- неумение использовать научную терминологию;
- наличие грубых ошибок;
- низкий уровень культуры исполнения заданий;
- низкий уровень сформированности заявленных в рабочей программе компетенций.

Формируемые компетенции:

- ОПК-2 Способен применять современный математический аппарат, связанный с проектированием, разработкой, реализацией и оценкой качества программных продуктов и программных комплексов в различных областях человеческой деятельности

- MF-1 Способен применять современную теоретическую математику для разработки новых алгоритмов и формулирования перспективных задач ИИ
- MF-2 Способен применять байесовский подход для построения вероятностных моделей анализа неопределенности и создания адаптивных систем ИИ
- BD-1 Способен осуществлять поиск сбор очистку и предварительный анализ данных
- BD-2 Способен определять требования к наборам данных для решения задач машинного обучения проводить разметку и анализ наборов данных оценивать качество данных обеспечивать непрерывную интеграцию данных

Оценочные средства для инвалидов и лиц с ограниченными возможностями здоровья выбираются с учетом их индивидуальных психофизических особенностей.

– при необходимости инвалидам и лицам с ограниченными возможностями здоровья предоставляется дополнительное время для подготовки ответа на экзамене;

– при проведении процедуры оценивания результатов обучения инвалидов и лиц с ограниченными возможностями здоровья предусматривается использование технических средств, необходимых им в связи с их индивидуальными особенностями;

– при необходимости для обучающихся с ограниченными возможностями здоровья и инвалидов процедура оценивания результатов обучения по дисциплине может проводиться в несколько этапов.

Процедура оценивания результатов обучения инвалидов и лиц с ограниченными возможностями здоровья по дисциплине предусматривает предоставление информации в формах, адаптированных к ограничениям их здоровья и восприятия информации:

Для лиц с нарушениями зрения:

- в печатной форме увеличенным шрифтом,
- в форме электронного документа.

Для лиц с нарушениями слуха:

- в печатной форме,
- в форме электронного документа.

Для лиц с нарушениями опорно-двигательного аппарата:

- в печатной форме,
- в форме электронного документа.

Данный перечень может быть конкретизирован в зависимости от контингента обучающихся студентов.

4.3. Методические указания по организации лабораторных работ по дисциплине. Примерные задания лабораторных работ

4.3.1. Общие сведения

Образовательная программа: Искусственный интеллект и аналитика данных

Дисциплина: «Многомерный статистический анализ».

Вид обеспечения: Проведение лабораторных работ.

Условия применения:

- Курс рассчитан на студентов 3-го года обучения.
- Техническое обеспечение: пакет STATISTICA, Python, R
- Использование открытых датасетов и библиотек.

4.3.2. Цели, задачи и ожидаемые

Цели:

- Закрепление теоретических знаний на практике.
- Развитие навыков проведения анализа данных в средах пакет STATISTICA, Python, R

- Подготовка к решению задач анализа данных в различных областях, включая индустрию, экономику, медицину и др.

Задачи:

1. Обеспечить студентов структурированными лабораторными работами.
2. Предоставить доступ к необходимому программному обеспечению.
3. Показать пример выполнения лабораторной работы
4. Организовать проверку и обратную связь по выполненным работам.

4.3.3. Ожидаемые результаты:

- Умение применять методы многомерного анализа на практике.
- Навыки работы с пакетом STATISTICA, языками программирования Python, R
- Опыт решения практических задач многомерного анализа в различных отраслях.

4.3.4. Примерные лабораторные задания. Соответствие заданий и формируемых компетенций

Задание 1. Произвести основные операции над переменными и случаями по представленным файлам данных

Задание 2. Построить двухмерные графики по представленным файлам данных

Задание 3. Вычислить описательные статистики, просчитать параметрические и непараметрические коэффициенты корреляции по представленным файлам данных

Задание 4. Произвести подгонку законов распределения по представленным файлам данных, генерацию случайных чисел

Задание 5. Произвести сравнение средних посредством t-критерия по представленным файлам данных

Задание 6. Реализовать метод Группировка и однофакторная ANOVA на представленных файлах данных

Задание 7. Реализовать метод Дисперсионный анализ на представленных файлах данных

Задание 8. Реализовать метод Непараметрическая статистика на представленных файлах данных

Задание 9. Реализовать метод частотный анализ на представленных файлах данных

Задание 10. Реализовать метод Таблицы кросстабуляции и таблицы флагов и заголовков на представленных файлах данных

Задание 11. Реализовать методы Множественная регрессия, фиксированная нелинейная регрессия на представленных файлах данных

Задание 12. Реализовать методы Нелинейная регрессия, модели бинарных откликов на представленных файлах данных

Задание 13. Реализовать метод Дискриминантный анализ на представленных файлах данных

Задание 14. Реализовать метод Кластерный анализ на представленных файлах данных

Задание 15. Реализовать метод Деревья классификации на представленных файлах данных

Задание 16. Реализовать метод Факторный анализ на представленных файлах данных

Задание 17. Реализовать метод Канонический анализ на представленных файлах данных

Задание 18. Реализовать метод Многомерное шкалирование на представленных файлах данных

Для успешного выполнения лабораторной работы обучающемуся следует ознакомиться с теоретической частью дисциплины по теме лабораторной работы,

изложенной в лекциях. С целью более полного и углубленного понимания теоретического материала могут быть использованы источники, указанные в списке литературы [1-5], а также Перечень ресурсов информационно-телекоммуникационной сети «Интернет», необходимых для освоения дисциплины, указанный в п.6.

Критерием должной подготовки студентов к выполнению лабораторных работ являются приобретенные знания, позволяющие дать безошибочные пояснения к теоретической части заданий по каждой теме лабораторных работ. Для приобретения должных навыков к решению задач предполагается проведение статистического анализа на лабораторных занятиях в компьютерных классах под руководством преподавателя. Закрепление приобретенных навыков осуществляется внеаудиторным самостоятельным выполнением статистического анализа по индивидуальным таблицам данных.

Соответствие лабораторных работ и индикаторов компетенций

В таблице ниже представлено, как каждая практическая работа способствует формированию заявленных компетенций.

Название индикатора	Тема/темы лабораторных работ	Обоснование соответствия
ОПК-2.2 Применяет современный математический аппарат при построении моделей в различных областях человеческой деятельности	Задания 5-12, 13, 15 (Сравнение средних, ANOVA, дисперсионный анализ, регрессионные модели, дискриминантный анализ, деревья классификации)	Эти работы требуют проверки адекватности моделей, сравнения с эталонными значениями, анализа ошибок и формулировки выводов
МФ-1.2 Применяет современный математический аппарат теории вероятностей для исследования методов и моделей машинного обучения	Задания 3-4, 11-18 (Описательные статистики, распределения, регрессия, дискриминантный, кластерный, факторный анализ)	Работы охватывают ключевые вероятностно-статистические методы, составляющие математическую основу машинного обучения
МФ-2.1 Использует теорему Байеса и её следствия, понимает отличия байесовской статистики от частотного подхода	Задания 9-10 (Частотный анализ на представленных файлах данных, таблицы кросстабуляции и таблицы флагов и заголовков на представленных файлах данных)	Одномерный и многомерный частотный анализ имеют прямое отношение к теореме Байеса
МФ-2.2 Применяет байесовские методы оценивания и байесовские интервалы для решения задач статистики и построения моделей	Задания 4, 9, 12 (Подгонка распределений, частотный анализ, модели бинарных откликов)	Эти методы являются основой для байесовского подхода - работа с распределениями, вероятностями и бинарными моделями
ВД-1.3 Применяет методы понижения размерности для первичной интерпретации	Задания 2, 16-18 (Визуализация, факторный анализ, канонический анализ, многомерное шкалирование)	Указанные методы непосредственно направлены на снижение размерности и визуальное

и визуализации многомерных данных		представление сложных данных
BD-2.1 Определяет требования к наборам и качеству данных для решения задач машинного обучения	Задание 3, 11-14 (Описательные статистики, коэффициенты корреляции, регрессия, модели бинарных откликов, дискриминантный, кластерный анализ)	Эти методы имеют непосредственное отношение к качеству данных

Детализация по направлениям:

Верификация и валидация моделей:

- Задания 5-8 - статистическая проверка гипотез и сравнение моделей
- Задания 11-12 - оценка качества регрессионных моделей
- Задания 13, 15 - валидация моделей классификации

Вероятностно-статистические методы:

- Задания 3-4 - базовые вероятностные распределения и статистики
- Задания 9-10 - анализ распределений и частот
- Задания 16-18 - многомерные вероятностные методы

Снижение размерности и визуализация:

- Задание 2 - базовые методы визуализации многомерных данных
- Задания 16-18 - продвинутое методы снижения размерности

Каждая лабораторная работа способствует формированию нескольких компетенций, обеспечивая комплексный подход к освоению методов анализа данных и машинного обучения.

4.3.5. Порядок реализации

Шаг №1. Обеспечить студентов структурированными лабораторными работами

Определение тем:

- Элементарные понятия математической статистики. Измерительные шкалы.
- Работа с данными. Основные операции над случаями и переменными
- Основные показатели статистики, корреляционный анализ
- Сравнение средних величин критерием Стьюдента
- Непараметрическая статистика. Сравнение величин в непараметрической статистике
- Группировка и однофакторная ANOVA
- Дисперсионный анализ
- Таблицы частот, сопряженности, флагов и заголовков.
- Канонический анализ
- Линейное и нелинейное моделирование взаимосвязей
- Дискриминантный анализ
- Кластерный анализ
- Деревья классификации
- Факторный анализ.
- Многомерное шкалирование

Шаг №2. Предоставить доступ к необходимому программному обеспечению

Доступ к программному обеспечению осуществляется посредством терминалов (рабочих станций), установленных в компьютерных классах и взаимодействующих с

университетским сервером с установленным программным обеспечением: пакет STATISTICA, языки программирования Python, R.

Шаг №3. Показать пример выполнения лабораторной работы

В качестве примера лабораторной работы выбрана тема «Дискриминантный анализ».

Цель работы: для «Ирисов Фишера» (Irisdat) провести дискриминантный анализ с предварительным анализом данных и последующей интерпретацией результатов.

Среды реализации (пакет STATISTICA). Реализацию в Python и R следует студентам выполнить как самостоятельную работу

Этапы выполнения лабораторной работы:

Этап 1. Применяя частотный анализ исследовать количественный состав групп (сортов ириса)

Этап 2. Исследовать закон распределения независимых переменных при помощи критерия Хи-квадрат Пирсона. Самостоятельно исследовать закон распределения независимых переменных по сортам ириса.

Вопросы по этапу 2:

- распределение каких переменных для всех групп и отдельно в каждой группе соответствует, или не соответствует нормальному закону?, обосновать ответ;
- каким образом несоответствие распределений нормальному закону может повлиять на результаты дискриминации?

Этап 3. Вычислить и сделать сравнительный анализ их описательных статистик по сортам ирисов

Вопросы по этапу 3:

- дать определения всех описательных статистик;
- в каких случаях целесообразно применение среднего арифметического, стандартного отклонения, дисперсии, доверительного интервала; а в каких медианы, нижней и верхней квартили, минимального и максимального значения переменных;
- какими преимуществами обладает коэффициент вариации перед стандартным отклонением и дисперсией

Этап 4. Построить итоговую таблицу дискриминантного анализа

Вопросы по этапу 4:

- в каких пределах изменяется Лямбда Уилкса, каким образом вычисляется и, что можно сказать о построенной модели дискриминации?
- интерпретировать результаты каждого столбца таблицы
- перечислить статистически значимые предикторы модели дискриминации

Этап 5. Исследовать расстояния между сортами ирисов

Вопросы по этапу 5:

- каким образом вычисляется расстояние Махаланобиса?
- какие расстояния статистически значимы и что означает статистическая значимость применительно к расстояниям?
- что характеризуют расстояния между сортами ирисов?

Этап 6. Построить таблицу со значениями критерия Хи-квадрат канонических корней

Вопросы по этапу 6:

- какое отношение имеют канонические корни к дискриминантным функциям?;
- как определяется количество канонических корней?;
- сколько канонических корней имеют данные, какие из них статистически значимы?

Этап 7. Построить таблицы с коэффициентами дискриминантных функций

Вопросы по этапу 6:

- чем отличаются исходные коэффициенты от стандартизованных?;
- как вычисляются собственные значения, и что характеризует кумулятивная доля?;
- написать уравнения дискриминантных функций

Этап 8. Построить таблицу со значениями дискриминантных функций для всех наблюдений (приведены первые 10 наблюдений).

Вопросы по этапу 8:

- каким образом вычисляются значения таблицы, и что они собою представляют.

Этап 9. Построить диаграмму рассеяния канонических значений и интерпретировать ее относительно расстояний между сортами, сходства и различия между ними

Этап 10. Построить таблицу с коэффициентами функций классификации

Вопросы по этапу 10:

- как выглядят функции классификации? Написать из уравнения;
- каким образом при помощи функций классификации можно предсказать принадлежность ирисов к тому, или иному сорту?

Этап 11. Построить матрицу классификации и дать ее интерпретацию

Вопрос по этапу 9:

- почему сорт SETOSA не содержит ошибочных классификаций, а два других сорта имеют ошибочные классификации

Этап 12. Построить таблицы с результатами классификации наблюдений, квадратов расстояний Махаланобиса, апостериорных вероятностей. Интерпретировать их и ответить на вопрос, как они могут быть использованы для классификации ирисов, не содержащихся в таблице (приведены таблицы с первыми 10 наблюдениями)

Шаг 4. Организовать проверку и обратную связь по выполненным работам.

4.3.6. Критерии оценки лабораторных заданий:

Отлично: Полное выполнение всех заданий, правильные ответы на все вопросы.

Хорошо: Незначительные недочеты в ответах

Удовлетворительно: не все этапы выполнены верно, есть ошибки в ответах

Неудовлетворительно: не все этапы выполнены, много ошибок при интерпретации результатов и ответах на вопросы

Формируемые компетенции:

– ОПК-6 Способен анализировать и разрабатывать организационно-технические и экономические процессы с применением методов системного анализа и математического моделирования

- MF-1 Способен применять современную теоретическую математику для разработки новых алгоритмов и формулирования перспективных задач ИИ
- MF-2 Способен применять байесовский подход для построения вероятностных моделей анализа неопределенности и создания адаптивных систем ИИ
- BD-1 Способен осуществлять поиск сбор очистку и предварительный анализ данных

- BD-2 Способен определять требования к наборам данных для решения задач машинного обучения проводить разметку и анализ наборов данных оценивать качество данных обеспечивать непрерывную интеграцию данных

4.4 Методические указания по организации курсовых проектов

4.4.1. Цель курсового проекта

Целью курсового проекта является закрепление теоретических знаний и практических навыков студентов в области многомерного статистического анализа. Проект направлен на развитие умений применять современные методы анализа данных для решения реальных задач, а также на формирование компетенций, связанных с обработкой, интерпретацией и визуализацией данных.

4.4.2. Задачи курсового проекта

- Освоить методы многомерного статистического анализа, такие как регрессионный анализ, кластерный анализ, факторный анализ, дискриминантный анализ и др.
- Научиться работать с программным обеспечением (STATISTICA, Python, R) для обработки и анализа данных.
- Развить навыки интерпретации результатов и их представления в виде отчетов или презентаций.
- Применить полученные знания для решения практических задач, связанных с анализом данных.

4.4.3. Ожидаемые результаты

- Умение применять методы многомерного статистического анализа на практике.
- Навыки работы с программными инструментами для анализа данных.
- Опыт решения реальных задач, включая подготовку данных, построение моделей и интерпретацию результатов.
- Способность оформлять результаты исследования в соответствии с требованиями.

4.4.4. Порядок реализации курсового проекта

1. Выбор темы:

Студент выбирает тему из предложенного списка или согласовывает собственную с научным руководителем.

Примерные темы:

- Прогнозирование заработной платы сотрудников методами регрессионного анализа.
- Кластерный анализ вторичного рынка автомобилей.
- Факторный анализ эффективности вузов.
- Сравнение методов многомерного анализа в средах STATISTICA, Python и R.

2. Составление плана:

План проекта согласовывается с научным руководителем и включает:

- Постановку задачи.
- Описание методов анализа.
- Этапы работы (сбор данных, предобработка, анализ, интерпретация).
- Сроки выполнения.

3. Сбор и подготовка данных:

- Использование открытых датасетов или данных, предоставленных организацией.
 - Очистка данных, обработка пропусков, преобразование переменных.
4. **Проведение анализа:**
 - Применение выбранных методов многомерного анализа.
 - Использование программного обеспечения для реализации моделей.
 5. **Оформление результатов:**
 - Подготовка отчета с описанием всех этапов работы.
 - Визуализация данных и результатов анализа.
 - Формулировка выводов и рекомендаций.
 6. **Защита проекта:**
 - Презентация результатов перед комиссией.
 - Ответы на вопросы по содержанию работы.

4.4.5. Структура курсовой работы

1. **Титульный лист** (оформляется по стандарту вуза).
2. **Оглавление.**
3. **Введение:**
 - Актуальность темы.
 - Цель и задачи работы.
 - Объект и предмет исследования.
4. **Теоретическая часть:**
 - Обзор методов многомерного анализа, используемых в работе.
 - Описание программного обеспечения.
5. **Практическая часть:**
 - Описание данных и их предобработка.
 - Применение методов анализа.
 - Интерпретация результатов.
6. **Заключение:**
 - Выводы по работе.
 - Практическая значимость результатов.
7. **Список литературы** (не менее 10 источников).
8. **Приложения** (если необходимы):
 - Исходные данные.
 - Код программы (для Python/R).
 - Дополнительные графики и таблицы.

Примерная тематика курсовых проектов

1. Статистический анализ вторичного рынка отечественных автомобилей с факторизацией пространства количественных признаков.
2. Многомерный анализ вторичного рынка автомобилей иностранного производства с кластеризацией пространства количественных признаков
3. Прогнозирование заработной платы сотрудников компании методами регрессионного анализа
4. Общие линейные модели как обобщение регрессионного анализа
5. Общие модели дискриминантного анализа как метод дискриминации с категориальными переменными
6. Позиционный анализ как метод оценки согласованности экспертов
7. Анализ экономической деятельности банков методами многомерного анализа
8. Анализ эффективности вузов методами многомерного анализа

9. Метод главных компонент в средах STATISTICA, R, преимущества и недостатки
10. Факторный анализ в средах STATISTICA, Python, R, преимущества и недостатки
11. Кластерный анализ в средах STATISTICA, Python, R, преимущества и недостатки
12. Деревья классификации в средах STATISTICA, Python, R, преимущества и недостатки

Рекомендации

- Регулярно консультироваться с научным руководителем.
- Использовать современные источники литературы и открытые данные.
- Проверять работу на уникальность (оригинальность текста должна быть не менее 70%).

Методические указания помогут студентам успешно выполнить курсовой проект и продемонстрировать свои навыки в области многомерного статистического анализа.

Критерии оценивания

Теоретическая часть (математическое обоснование):

Пороговый – Приведены базовые формулы и определения без глубокого анализа

Базовый – Дано развернутое объяснение методов, их плюсы/минусы, ссылки на источники.

Продвинутый – Собственный анализ методов, критика существующих подходов, предложения улучшений.

Практическая реализация

Пороговый – Работающий код на Python/MATLAB с минимальными комментариями.

Базовый – Чистый, модульный код с пояснениями, тестами и визуализацией.

Продвинутый – Оптимизированный код, сравнение нескольких методов, обработка edge-cases.

Эксперименты и анализ результатов

Пороговый – Простые тесты на синтетических данных без глубокого анализа.

Базовый – Сравнение с baseline-методами, интерпретация метрик (MSE, Accuracy).

Продвинутый – Анализ устойчивости, влияние гиперпараметров, статистическая значимость.

Оформление и структура

Пороговый – наличие введения, основной части и выводов.

Базовый – Четкая структура (оглавление, список литературы, графики).

Продвинутый – Публикация кода на GitHub, оформление в LaTeX, видео-презентация.

Балл	Уровень выполнения	Соответствие критериям
5 (Отлично)	Продвинутый	Все критерии выполнены на высоком уровне + инновации.
4 (Хорошо)	Базовый	Полное соответствие без серьезных недочетов.
3 (Удовл.)	Пороговый	Минимальные требования выполнены, но есть недочеты.
2 (Неуд.)	Ниже порога	Критические ошибки в теории/практике.

- **Отлично:**
 - Полное соответствие структуре и требованиям.
 - Глубокий анализ данных, корректное применение методов.
 - Четкие выводы и рекомендации.
 - Грамотное оформление и презентация.
- **Хорошо:**
 - Незначительные недочеты в анализе или оформлении.
 - Удовлетворительная интерпретация результатов.
- **Удовлетворительно:**
 - Нарушения в структуре или методах анализа.
 - Поверхностные выводы.
- **Неудовлетворительно:**
 - Несоответствие темы и содержания.
 - Грубые ошибки в анализе.
 - Нарушение академических норм (плагиат).

Формируемые компетенции:

- ОПК-6 Способен анализировать и разрабатывать организационно-технические и экономические процессы с применением методов системного анализа и математического моделирования
- MF-1 Способен применять современную теоретическую математику для разработки новых алгоритмов и формулирования перспективных задач ИИ
- MF-2 Способен применять байесовский подход для построения вероятностных моделей анализа неопределенности и создания адаптивных систем ИИ
- BD-1 Способен осуществлять поиск сбор очистку и предварительный анализ данных
- BD-2 Способен определять требования к наборам данных для решения задач машинного обучения проводить разметку и анализ наборов данных оценивать качество данных обеспечивать непрерывную интеграцию данных

5. Перечень учебной литературы, необходимой для освоения дисциплины

5.1. Литература

1. Халафян, А.А. Методы машинного обучения в Data Mining пакета STATISTICA / А. А. Халафян. - М.: [Горячая линия-Телеком], 2022. - 260 с.
2. Халафян, А.А. Методы искусственного интеллекта в медицинских задачах классификации и регрессии / А. А. Халафян. - М.: [Горячая линия-Телеком], 2023. - 352 с.
3. Чураков, Е. П. Введение в многомерные статистические методы / Е. П. Чураков. — 3-е изд., стер. — Санкт-Петербург: Лань, 2023. — 148 с. — ISBN 978-5-507-47141-6. — Текст: электронный // Лань: электронно-библиотечная система. — URL: <https://e.lanbook.com/book/330530> (дата обращения: 13.06.2023). — Режим доступа: для авториз. пользователей.
4. Ганичева, А. В. Прикладная статистика / А. В. Ганичева. — 4-е изд., стер. — Санкт-Петербург: Лань, 2023. — 172 с. — ISBN 978-5-507-47980-1. — Текст : электронный // Лань: электронно-библиотечная система. — URL: <https://e.lanbook.com/book/336800> (дата обращения: 13.06.2023). — Режим доступа: для авториз. пользователей.
5. Халафян, А.А. Статистический анализ данных. STATISTICA 6 / А. А. Халафян. - М.: [БИНОМ-Пресс], 2010. - 522 с.
6. Халафян, А.А. Статистический анализ данных. STATISTICA 6 / А. А. Халафян. - М.: [Бином-Пресс], 2009. - 522 с.

7. Халафян, А.А. Математическая статистика с элементами теории вероятностей. STATISTIKA 6 / А. А. Халафян. - М. : БИНОМ, 2010. - 491 с.
8. Туганбаев, А.А. Теория вероятностей и математическая статистика / А.А. Туганбаев, В.Г. Крупин. – Санкт-Петербург: Лань, 2011. – 320 с. [Электронный ресурс]: – Режим доступа: <https://e.lanbook.com/book/652>
9. Боровиков, В. П. Популярное введение в современный анализ данных в системе STATISTICA. Методология и технология современного анализа данных : учебное пособие / В. П. Боровиков. — Москва: Горячая линия-Телеком, 2018. — 288 с. — ISBN 978-5-9912-0326-5. — Текст: электронный // Лань : электронно-библиотечная система. — URL: <https://e.lanbook.com/book/111023> (дата обращения: 13.06.2023). — Режим доступа: для авториз. Пользователей.
10. Сенько, А Работа с BigData в облаках. Обработка и хранение данных с примерами из Microsoft Azure / А Сенько. - СПб.: Питер, 2019. - 448 с. ЭБС Университетская библиотека ONLINE. – URL: <https://biblioclub.ru/index.php?page=book&id=598404> (дата обращения: 07.12.2023). – Текст : электронный.
11. Sun, X., Li, J., Kovalenko, A.V., Feng, W., Ou, Y. Integrating Reinforcement Learning and Learning From Demonstrations to Learn Nonprehensile Manipulation //IEEE Transactions on Automation Science and Engineering, 2023, 20(3), 1735–1744, DOI: 10.1109/TASE.2022.3185071, Q1
12. Petukhova, A.V.; Kovalenko, A.V.; Ovsyannikova, A.V. Algorithm for Optimization of Inverse Problem Modeling in Fuzzy Cognitive Maps. Mathematics 2022, 10, 3452. DOI: 10.3390/math10193452, Q1
13. Kirillova, E.; Kovalenko, A.; Urtenov, M. Study of the Current–Voltage Characteristics of Membrane Systems Using Neural Networks. AppliedMath 2025, 5, 10. <https://doi.org/10.3390/appliedmath5010010>
14. Kadurin, Artur, et al. "The cornucopia of meaningful leads: Applying deep adversarial autoencoders for new molecule development in oncology." Oncotarget 8.7 (2016): 10883.
15. Kadurin, Artur, et al. "druGAN: an advanced generative adversarial autoencoder model for de novo generation of new molecules with desired molecular properties in silico." Molecular pharmaceutics 14.9 (2017): 3098-3104.
16. Polykovskiy, Daniil, et al. "Molecular sets (MOSES): a benchmarking platform for molecular generation models." Frontiers in pharmacology 11 (2020): 565644.
17. Khrabrov, Kuzma, et al. " ∇^2 DFT: A Universal Quantum Chemistry Dataset of Drug-Like Molecules and a Benchmark for Neural Network Potentials." Advances in Neural Information Processing Systems 37 (2024): 36869-36889..
18. Polykovskiy, Daniil, et al. "Entangled conditional adversarial autoencoder for de novo drug discovery." Molecular pharmaceutics 15.10 (2018): 4398-4405.
19. Николенко, Сергей, Кадури, Артур и Архангельская Екатерина. Глубокое обучение. Издательский дом "Питер", 2017..

5.2. Периодические издания и конференции (А*):

1. IEEE Transactions on Big Data – научные статьи по обработке больших данных.
2. Journal of Big Data (SpringerOpen) – открытый журнал с исследованиями в области Big Data.
3. Big Data Research (Elsevier) – публикации по анализу, управлению и визуализации данных.
4. Data Science Journal (CODATA) – междисциплинарные исследования данных.
5. ACM Transactions on Knowledge Discovery from Data (TKDD) – методы извлечения знаний из больших данных.
6. <https://openreview.net/forum?id=FMMF1a9ifL>
7. <https://openreview.net/forum?id=ElUrNM9U8c#discussion>
8. <https://openreview.net/forum?id=JoO6mtCLHD>
9. <https://aclanthology.org/2024.findings-emnlp.760/>

10. <https://aclanthology.org/2020.coling-main.588/>
11. https://link.springer.com/chapter/10.1007/978-3-030-72113-8_30
12. https://link.springer.com/chapter/10.1007/978-3-031-42448-9_10
13. <https://aclanthology.org/2024.findings-naacl.288/>

5.3. Интернет-ресурсы, в том числе современные профессиональные базы данных и информационные справочные системы

Электронно-библиотечные системы (ЭБС):

1. ЭБС «ЮРАЙТ» <https://urait.ru/>
2. ЭБС «УНИВЕРСИТЕТСКАЯ БИБЛИОТЕКА ОНЛАЙН» <http://www.biblioclub.ru/>
3. ЭБС «BOOK.ru» <https://www.book.ru>
4. ЭБС «ZNANIUM.COM» www.znanium.com
5. ЭБС «ЛАНЬ» <https://e.lanbook.com>

Профессиональные базы данных

1. Scopus <http://www.scopus.com/>
2. ScienceDirect <https://www.sciencedirect.com/>
3. Научная электронная библиотека (НЭБ) <http://www.elibrary.ru/>
4. Полнотекстовые архивы ведущих западных научных журналов на Российской платформе научных журналов НЭИКОН <http://archive.neicon.ru>
5. Национальная электронная библиотека (доступ к Электронной библиотеке диссертаций Российской государственной библиотеки (РГБ) <https://rusneb.ru/>
6. Президентская библиотека им. Б.Н. Ельцина <https://www.prilib.ru/>
7. Springer Journals: <https://link.springer.com/>
8. Springer Journals Archive: <https://link.springer.com/>
9. Nature Journals: <https://www.nature.com/>
10. Springer Nature Protocols and Methods: <https://experiments.springernature.com/sources/springer-protocols>
11. Springer Materials: <http://materials.springer.com/>
12. Nano Database: <https://nano.nature.com/>
13. Springer eBooks (i.e. 2020 eBook collections): <https://link.springer.com/>
14. "Лекториум ТВ" <http://www.lektorium.tv/>
15. Университетская информационная система РОССИЯ <http://uisrussia.msu.ru>

Информационные справочные системы

1. Консультант Плюс - справочная правовая система (доступ по локальной сети с компьютеров библиотеки)

Ресурсы свободного доступа

1. КиберЛенинка <http://cyberleninka.ru/>;
2. Министерство науки и высшего образования Российской Федерации <https://www.minobrnauki.gov.ru/>;
3. Федеральный портал "Российское образование" <http://www.edu.ru/>;
4. Информационная система "Единое окно доступа к образовательным ресурсам" <http://window.edu.ru/>;
5. Единая коллекция цифровых образовательных ресурсов <http://school-collection.edu.ru/> .
6. Проект Государственного института русского языка имени А.С. Пушкина "Образование на русском" <https://pushkininstitute.ru/>;
7. Справочно-информационный портал "Русский язык" <http://gramota.ru/>;
8. Служба тематических толковых словарей <http://www.glossary.ru/>;
9. Словари и энциклопедии <http://dic.academic.ru/>;
10. Образовательный портал "Учеба" <http://www.ucheba.com/>;

11. Законопроект "Об образовании в Российской Федерации". Вопросы и ответы http://xn--273--84d1f.xn--p1ai/voprosy_i_otvety

Собственные электронные образовательные и информационные ресурсы КубГУ

1. Электронный каталог Научной библиотеки КубГУ <http://megapro.kubsu.ru/MegaPro/Web>
2. Электронная библиотека трудов ученых КубГУ <http://megapro.kubsu.ru/MegaPro/UserEntry?Action=ToDb&idb=6>
3. Среда модульного динамического обучения <http://moodle.kubsu.ru>
4. База учебных планов, учебно-методических комплексов, публикаций и конференций <http://infoneeds.kubsu.ru/>
5. Библиотека информационных ресурсов кафедры информационных образовательных технологий <http://mschool.kubsu.ru;>
6. Электронный архив документов КубГУ <http://docspace.kubsu.ru/>
7. Электронные образовательные ресурсы кафедры информационных систем и технологий в образовании КубГУ и научно-методического журнала "ШКОЛЬНЫЕ ГОДЫ" <http://icdau.kubsu.ru/>

Для освоения дисциплины инвалидами и лицами с ограниченными возможностями здоровья имеются издания в электронном виде в электронно-библиотечной системе «Лань».

6. Методические указания для обучающихся по освоению дисциплины

По курсу предусмотрено проведение лекционных занятий, на которых дается основной теоретический материал, лабораторных занятий, позволяющих студентам в полной мере ознакомиться с понятиями дисциплины и освоиться в решении практических задач. Важнейшим этапом курса является самостоятельная работа. Целью самостоятельной работы бакалавра является углубление знаний, полученных в результате аудиторных занятий. Вырабатываются навыки самостоятельной работы. Закрепляются опыт и знания, полученные во время лабораторных занятий. Самостоятельная работа студентов в ходе изучения дисциплины состоит в выполнении индивидуальных заданий (отчет в электронной форме), подготовки теоретического материала к лабораторным занятиям, на основе конспектов лекций и учебной литературы, согласно календарному плану и подготовки теоретического материала к тестовому опросу, зачету и экзамену, согласно вопросам к экзамену.

Указания по оформлению работ:

- работа на лабораторных занятиях выполняется за компьютером в компьютерном классе;
- оформление индивидуальных заданий (отчетов) желательно в виде файлов в формате word.

Итогом самостоятельной работы студента является отчет, в котором на оригинальной таблице исходных данных студент самостоятельно проводит анализ данных всеми изученными в рамках курса методами и, представляет его на проверку в электронном виде.

Проверка индивидуальных заданий по темам, разобранным на лабораторных занятиях, осуществляется через неделю на текущем лабораторном занятии, либо в течение недели после этого занятия на консультации.

Для разъяснения непонятных вопросов лектором и ассистентом еженедельно проводятся консультации, о времени которых группы извещаются заранее.

В освоении дисциплины инвалидами и лицами с ограниченными возможностями здоровья большое значение имеет индивидуальная учебная работа (консультации) – дополнительное разъяснение учебного материала.

Индивидуальные консультации по предмету являются важным фактором, способствующим индивидуализации обучения и установлению воспитательного контакта

между преподавателем и обучающимся инвалидом или лицом с ограниченными возможностями здоровья.

Подход, определяющий установление соответствия кейсов ИП и УГТ (5-7), позволяет четко соотносить этапы развития технологии с вовлеченностью партнера и снижать риски при переходе от лабораторных испытаний к промышленному внедрению.

Ключевые аспекты взаимодействия с промышленными партнерами:

- Для УГТ 5 – ИП помогает определить реалистичные условия тестирования, но не рискует своей инфраструктурой.
- Для УГТ 6 – ИП предоставляет "песочницу" или изолированную среду, где можно выявить скрытые проблемы.
- Для УГТ 7 – ИП становится соразработчиком, так как технология адаптируется под его конкретные процессы.

Кейсы от ПАО «Сбербанк»

Кейс 1: Прогнозирование оттока клиентов

Описание:

Анализ поведения клиентов Сбера для выявления признаков возможного ухода к конкурентам.

Цель:

Снижение оттока клиентов за счёт своевременного предложения персонализированных условий.

Технологии:

- Логистическая регрессия / градиентный бустинг
- Кластеризация (k-means)
- Анализ выживаемости (Survival Analysis)

Реализация:

1. Сбор данных о транзакциях, взаимодействиях с колл-центром, использованием мобильного приложения.
2. Построение модели для оценки вероятности ухода клиента.
3. Сегментация клиентов и выявление ключевых факторов оттока.

Результат:

Снижение оттока на 15%, повышение лояльности за счёт таргетированных предложений.

Кейс 2: Оптимизация кредитного скоринга

Описание:

Улучшение скоринговой модели для более точной оценки кредитоспособности клиентов.

Цель:

Снижение уровня дефолтов при увеличении одобренных кредитов.

Технологии:

- Random Forest (случайный лес) / CatBoost (алгоритм градиентного бустинга, адаптированный к категориальным переменным)
- Методы feature engineering (генерация новых признаков) (WOE, IV)
- SHAP-анализ (это метод интерпретации моделей машинного обучения, который предоставляет информацию о вкладе каждого признака в предсказание конкретного наблюдения) для интерпретации модели

Реализация:

1. Анализ исторических данных по выданным кредитам.
2. Добавление альтернативных данных (например, цифровой след клиента).
3. Тестирование модели на новых данных.

Результат:

Уменьшение дефолтов на 10% при росте одобрений на 5%.

Кейс 3: Сегментация клиентов для кросс-продаж**Описание:**

Группировка клиентов для персонализированного предложения дополнительных продуктов (страховки, инвестиции).

Цель:

Увеличение кросс-продаж и среднего чека клиента.

Технологии:

- Метод главных компонент (PCA)
- Иерархическая кластеризация
- A/B-тестирование

A/B-тестирование – это метод исследования, при котором сравнивают эффективность двух вариантов какого-то объекта, например страницы сайта. Эти варианты показывают аудитории и оценивают, на какой из них люди реагируют лучше

Реализация:

1. Анализ транзакций и продуктовой корзины клиентов.
2. Выделение сегментов (например, "инвесторы", "заёмщики", "платёжные").
3. Разработка персонализированных предложений.

Результат:

Рост кросс-продаж на 20% за квартал.

Кейс 4: Обнаружение мошеннических операций**Описание:**

Разработка системы для выявления аномальных транзакций в режиме реального времени.

Цель:

Снижение финансовых потерь от мошенничества.

Технологии:

- Алгоритмы anomaly detection (алгоритмы детектирования аномалий: Isolation Forest, Autoencoders)
Обнаружение аномалий - это метод, используемый в машинном обучении (ML) для выявления редких элементов, событий или наблюдений, которые вызывают подозрения, поскольку значительно отличаются от большинства данных.
- Анализ временных рядов
- Graph-based анализ связей (используют представление данных при помощи графов и разбивают его на подграфы, соответствующие кластерам)

Реализация:

1. Обучение модели на исторических данных о мошеннических операциях.
2. Интеграция в процессинг транзакций.
3. Автоматическое блокирование подозрительных операций.

Результат:

Снижение ущерба от мошенничества на 30%.

Кейс 5: Оптимизация работы отделений банка**Описание:**

Анализ загруженности отделений для улучшения распределения ресурсов.

Цель:

Сокращение издержек и повышение удовлетворённости клиентов.

Технологии:

- Регрессионный анализ
- Прогнозирование временных рядов (ARIMA, Prophet – библиотека прогнозирования временных рядов)
- Оптимизация на основе линейного программирования

Реализация:

1. Сбор данных о посещаемости, времени обслуживания.
2. Прогнозирование пиковых нагрузок.
3. Оптимизация графика работы сотрудников.

Результат:

Снижение очередей на 25%, экономия на персонале до 10%.

Кейсы от «АВАЛАБ» (лаборатория искусственного интеллекта Сбербанка)

Кейс 1: Кластеризация клиентов банка на основе их финансового поведения

Описание:

Банк хочет сегментировать своих клиентов по их транзакциям, кредитной истории и другим финансовым показателям, чтобы предложить персонализированные услуги.

Цель:

Выделить однородные группы клиентов с помощью методов кластеризации для улучшения маркетинговых стратегий.

Технологии:

- Метод главных компонент (PCA) для снижения размерности
- K-means и иерархическая кластеризация
- Визуализация (дендрограммы, scatter plot)

Реализация:

1. Загрузка и предобработка данных (нормализация, удаление выбросов).
2. Применение PCA для сокращения числа переменных.
3. Кластеризация клиентов с помощью K-means и иерархических методов.
4. Интерпретация кластеров (средние значения признаков в группах).

Результат:

Выделено 5 кластеров клиентов с разными финансовыми профилями. На основе этого банк может предлагать таргетированные продукты.

Кейс 2: Прогнозирование оттока клиентов телеком-компаний

Описание:

Телекоммуникационная компания хочет предсказать, какие клиенты с высокой вероятностью уйдут к конкурентам.

Цель:

Построить модель классификации для выявления клиентов с риском оттока.

Технологии:

- Логистическая регрессия
- Random Forest / XGBoost
- Метрики: ROC-AUC, Precision-Recall (точность-полнота)

Реализация:

1. Разведочный анализ (EDA) и обработка пропусков.
2. Feature engineering (создание новых признаков, например, среднее количество звонков в месяц).
3. Обучение и валидация моделей.

4. Интерпретация важности признаков (SHAP, feature_importances_).

Результат:

Модель с $AUC = 0.89$ позволяет выявлять 80% уходящих клиентов, что дает возможность удерживать их специальными предложениями.

Кейс 3: Анализ эффективности рекламных кампаний с помощью MANOVA

Описание:

Маркетинговая команда запустила несколько рекламных кампаний и хочет оценить их влияние на разные группы потребителей.

Цель:

Определить, есть ли статистически значимые различия в реакции потребителей на разные рекламные стратегии.

Технологии:

- Многомерный дисперсионный анализ (MANOVA)
- Тесты на нормальность и гомогенность дисперсий (Box's M, Levene) для соблюдения условий применимости MANOVA
- Пост-хок тесты (Tukey, HSD)

Реализация:

1. Формирование групп по демографическим признакам.
2. Проверка предпосылок MANOVA.
3. Применение MANOVA для оценки влияния кампаний на несколько метрик (например, количество покупок, средний чек).

Результат:

Обнаружено, что кампания А значимо эффективнее для молодежи, а кампания В — для людей старше 40 лет.

Кейс 4: Факторный анализ для оценки удовлетворенности сотрудников

Описание:

Компания провела опрос сотрудников по 20 параметрам удовлетворенности работой и хочет сократить размерность данных.

Цель:

Выявить скрытые факторы, влияющие на удовлетворенность персонала.

Технологии:

- Факторный анализ (PCA, Varimax-вращение)
- Определение числа факторов (Kaiser criterion, scree plot)

Реализация:

1. Проверка данных на пригодность для факторного анализа (корреляционный анализ, Bartlett's test).
2. Выделение главных компонент.
3. Интерпретация факторов (например, "Удовлетворенность зарплатой", "Карьерные возможности").

Результат:

Выделено 4 ключевых фактора, объясняющих 75% дисперсии. HR-отдел может фокусироваться на улучшении этих аспектов.

Кейс 5: Построение рекомендательной системы для интернет-магазина

Описание:

Интернет-магазин хочет увеличить средний чек за счет персонализированных рекомендаций товаров.

Цель:

Разработать систему рекомендаций на основе поведения пользователей.

Технологии:

- Коллаборативная фильтрация (SVD – метод коллаборативной фильтрации, использующий сингулярное разложение, ALS – метод коллаборативной фильтрации, использующий метод наименьших квадратов)
- Метрики: Precision@k, RMSE
RMSE (Root Mean Squared Error), или среднеквадратическая ошибка (СКО), является метрикой оценки качества моделей регрессии;
Precision@k – метрика precision (точность) измеряет вероятность того, что модель верно спрогнозировала, что значение является истинным.

Реализация:

1. Построение user-item матрицы (покупки/просмотры).
2. Обучение модели SVD на исторических данных.
3. Генерация рекомендаций для каждого пользователя.

Результат:

Внедрение системы увеличило конверсию в покупку на 15%.

7. Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине

7.1 Перечень информационно-коммуникационных технологий

1. Электронная почта mail.ru, yandex.ru
2. Yandex Browser
3. Система управления обучением Moodle – сдача работ

7.2 Перечень лицензионного и свободно распространяемого программного обеспечения

1. OpenOffice
2. GIT
3. Yandex Browser
4. Mozilla Firefox
5. Google Chrome
6. Python + Jupyter + Google Colab
7. SymPy/SageMath
8. Octave (аналог MATLAB)

7.2 Перечень лицензионного и свободно распространяемого программного обеспечения

1. OpenOffice
2. GIT
3. Yandex Browser
4. Mozilla Firefox
5. Google Chrome
6. Python + Jupyter + Google Colab
7. SymPy/SageMath
8. Octave (аналог MATLAB)

8. Материально-техническое обеспечение по дисциплине

В качестве материально-технического обеспечения дисциплины используются - проекционное оборудование (цифровой проектор, экран, ноутбук, интерактивная доска).

Для проведения занятий используются аудитории с учебной мебелью (столы, стулья), соответствующей количеству студентов и позволяющей осуществлять упражнения по моделированию компьютерные классы. Студенты и преподаватели вуза имеют

постоянный доступ к электронному каталогу учебной, методической, научной литературе, периодическим изданиям и архиву статей.

№	Вид работ	Наименование учебной аудитории, ее оснащенность оборудованием и техническими средствами обучения
	Лекционные занятия	Аудитория, укомплектованная специализированной мебелью и техническими средствами обучения
	Лабораторные занятия	Аудитория, укомплектованная специализированной мебелью и техническими средствами обучения, компьютерами, проектором, программным обеспечением
	Групповые (индивидуальные) консультации	Аудитория, укомплектованная специализированной мебелью и техническими средствами обучения, компьютерами, программным обеспечением
	Текущий контроль, промежуточная аттестация	Аудитория, укомплектованная специализированной мебелью и техническими средствами обучения, компьютерами, программным обеспечением
	Самостоятельная работа	Кабинет для самостоятельной работы, оснащенный компьютерной техникой с возможностью подключения к сети «Интернет», программой экранного увеличения и обеспеченный доступом в электронную информационно-образовательную среду университета.

Виртуальные машины, кластер Managed Kubernetes и ресурсы GPU в облаке предоставляется индустриальным партнером ПАО «Сбербанк»:

№	Продукт	Параметры продукта	Кол-во	Кол-во конфигураций	Ед. изм.
1	Виртуальная машина	Виртуальная машина 10% vCPU 2 vCPU 4 RAM	1	60	Шт
		ОС Ubuntu 22.04	1		Шт
		Системный диск SSD	1		Шт
			10		Гб
		Аренда публичного IP	1		Шт
2	Виртуальная машина с GPU	Виртуальная машина с GPU NVIDIA® Tesla® V100 2 GPU 8 vCPU 128 Гб RAM	1	1	Шт
		ОС Ubuntu_24.04	1		Шт
		Системный диск SSD	1		Шт
			2000		Гб
		Диск SSD	1		Шт
			4096		Гб
		Диск SSD	1		Шт
			4096		Гб
3	K8S	Аренда публичного IP	1		Шт
		Master node 8 vCPU 16 RAM	1	1	Шт
		Worker node 10% доля 4 vCPU 32 RAM	5		Шт
		Worker node SSD-NVME	64		Гб

4	ML Inference Instance Type GPU	Аренда публичного IP	1		Шт
		Время работы в месяц	40	1	Ч
		Инстанс 8 x NVIDIA® H100 NVLink PCIe 160 vCPU 1520 GB RAM	1		Шт
		Количество запросов к ML- моделям	1		Млн. Шт
5	LLM	Кэш ML-моделей	160		Гб
		Токены GigaChat 2 Max	50		Млн. Шт
		Токены Embeddings	400		Млн. Шт

Дополнительные облачные ресурсы предоставляются технологическим партнером Yandex Cloud.