Документ подписан простой электронной подписью
Информация о владельце:
ФИО: Гриб Владислав Валерьевич
Должность: Ректор
Дата подписания: 23.09.2026 18:46:59
Уникальный программный ключ:
637517d24e103c3db032acf37e[illegible]

Образовательное частное учреждение высшего образования
«МОСКОВСКИЙ УНИВЕРСИТЕТ ИМЕНИ А.С. ГРИБОЕДОВА»
(ИМПЭ им. А.С. Грибоедова)

МЕЖДУНАРОДНЫЙ ИНСТИТУТ ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ И БИЗНЕС-ИНФОРМАТИКИ

УТВЕРЖДАЮ
Директор Международного
института информационных
технологий и бизнес-информатики
_____ Т.В. Новикова
«17» декабря 2025г.

Рабочая программа дисциплины

СТАТИСТИЧЕСКИЕ МЕТОДЫ АНАЛИЗА ДАННЫХ

Направление подготовки
09.03.03 Прикладная информатика
(уровень бакалавриат)

Направленность (профиль):
«Анализ данных»

Форма обучения: **очная, заочная**

Москва

Рабочая программа дисциплины «Статистические методы анализа данных». Направление подготовки 09.03.03 Прикладная информатика, направленность (профиль): «Анализ данных» / Т.В. Новикова – М.: ИМПЭ им. А.С. Грибоедова.

Рабочая программа дисциплины составлена на основании федерального государственного образовательного стандарта высшего образования – бакалавриат по направлению подготовки 09.03.03 Прикладная информатика, утвержденного приказом Министерства образования и науки Российской Федерации от «19» сентября 2017 № 922 (с изменениями и дополнениями) и Профессиональным стандартом «Программист», Утверждённым приказом Министерства труда и социальной защиты Российской Федерации от 20 июля 2022 № 424н (регистрационный номер 4).

Разработчики:	Т.В. Новикова, к.э.н., доцент
Ответственный рецензент:	Е.В. Михалёва, к. ф.-м. н. исполнительный директор института информационных систем и инженерно- компьютерных технологий

Рабочая программа дисциплины рассмотрена и одобрена на заседании кафедры информационных технологий и прикладной информатики 17.12.2025г., протокол №6

Заведующий кафедрой _____/Н.Н. Загускин, к.ю.н, доцент

Согласовано от библиотеки _____/ О. Е. Степкина

Раздел 1. Цели и задачи освоения дисциплины

Целью освоения дисциплины «Статистические методы анализа данных» является изучение современной методологии анализа данных и формирование навыков статистического анализа данных с целью разработки алгоритмов математического обеспечения информационных систем.

Задачи, решаемые в ходе освоения программы дисциплины:

– ознакомить обучающихся с методами обработки данных, со сферами применения статистических методов анализа данных в информационных системах;

– сформировать навыки решения профессионально-ориентированных задач с применением статистических методов обработки и анализа данных.

Раздел 2. Планирование результатов обучения по дисциплине, соотнесенные с планируемыми результатами освоения образовательной программы

Код компетенции	Формулировка компетенции	Индикаторы достижения компетенции
ПК-5	Способен осуществлять проектирование структур данных	**ИПК-5.1 Знать:** Методы и средства проектирования компьютерного программного обеспечения **ИПК-5.2 Уметь:** Применять методы и средства проектирования компьютерного программного обеспечения, структур данных, баз данных, программных интерфейсов
ОПК-6	Способен анализировать и разрабатывать организационно-технические и экономические процессы с применением методов системного анализа и математического моделирования	**ИОПК-6.1.** Знать основы теории систем и системного анализа, дискретной математики, теории вероятностей и математической статистики, методов оптимизации и исследования операций, нечётких вычислений, математического и имитационного моделирования на базовом уровне **ИОПК-6.2.** Уметь применять методы теории систем и системного анализа, математического, статистического и имитационного моделирования для автоматизации задач принятия решений, анализа информационных потоков, расчёта экономической эффективности и надёжности информационных систем технологий на базовом уровне

Раздел 3. Место дисциплины в структуре образовательной программы

Дисциплина «Статистические методы анализа данных» изучается в 6 семестре очной и в 7 семестре заочной форм обучения, относится к Блоку 1 «Дисциплины (модули)», «Часть, формируемая участниками образовательных отношений», «Дисциплины (модули) по выбору 1 (ДВ.1), образовательной программы по направлению подготовки 09.03.03 Прикладная информатика (уровень бакалавриат), направленность (профиль): «Анализ данных».

Раздел 4. Объем (трудоемкость) дисциплины (общая, по видам учебной работы, видам промежуточной аттестации)

Трудоемкость дисциплины и виды учебной нагрузки

на очной форме обучения

з.е.	Итого	Лекции	Практические занятия	Курсовое проектирование	Самостоятельная работа	Текущий контроль	Контроль, промежуточная аттестация
6 семестр							

4	144	26	26		56		36 Экзамен

на заочной форме обучения

з.е.	Итого	Лекции	Практические занятия	Курсовое проектирование	Самостоятельная работа	Текущий контроль	Контроль, промежуточная аттестация
7 семестр							
4	144	4	8		96		36 Экзамен

Тематический план дисциплины

Очная форма обучения

Разделы / Темы	Лекции	Практические занятия	Самостоятельная работа	Текущий контроль	Контроль, промежуточная аттестация	Всего часов
6 семестр						
Раздел №1. «Введение в статистические методы анализа данных».						
Тема 1.1. Введение в статистические методы анализа данных.			2			2
Тема 1.2. Методы первичной обработки и представления данных.	2	2	2			6
Тема 1.3. Методы выявления и устранения аномальных наблюдений.		2	2			4
Раздел №2. «Множественный корреляционно-регрессионный анализ. Классическая модель множественной регрессии»						
Тема 2.1. Корреляционно-регрессионный анализ. Модель множественной регрессии.		2	2			4
Тема 2.2. Фиктивные переменные. Критерий Чоу проверки	2		2			4

однородности.						
Тема 2.3. Множественные коэффициенты корреляции.		2	2			4
Тема 2.4. Статистический анализ адекватности модели множественной регрессии.	2		2			4
Тема 2.5. Мультиколлинеарность факторов.			2			2
Раздел №3. «Обобщенная линейная модель множественной регрессии. Регрессионные модели с автокоррелированными остатками»						
Тема 3.1. Обобщенная линейная модель множественной регрессии. Обобщенный МНК	2		2			4
Тема 3.2. Гетероскедастичность. Методы выявления гетероскедастичности.	2	2	2			6
Тема 3.3. Взвешенный МНК. Устранение гетероскедастичности.	2	2	2			6
Тема 3.4. Регрессионные модели с автокоррелированными остатками.			2			2
Тема 3.5. Устранение автокорреляции		2	2			4
Раздел №4 «Кластерный анализ»						
Тема 4.1. Кластерный анализ. Основные этапы кластерного анализа.	2		2			4
Тема 4.2.		2	2			4

Метрики расстояний. Свойства кластеров. Методы кластерного анализа.						
Тема 4.3. Расстояния между кластерами. Методы группировки кластеров.			2			2
Тема 4.4. Иерархические методы кластеризации. Агломеративные методы.	2		2			4
Тема 4.5. Иерархические дивизимные методы кластеризации.	2		2			4
Тема 4.6. Примеры реализации иерархических методов кластеризации.		2	2			4
Тема 4.7. Итеративные методы. Метод к-средних.		2	2			4
Тема 4.8. Метод к-средних. Критерии качества кластеризации.			2			2
Тема 4.9. Метод нечетких к-средних		2	2			4
Раздел №5 «Методы классификации»						
Тема 5.1. Понятие методов классификации.	2		2			4
Тема 5.2. Логистическая регрессия.	2		2			4
Тема 5.3. Метрики			2			2

качества классификации.						
Тема 5.4. Метрики качества классификации. Регуляризация в задаче регрессии.	2	2	2			6
Тема 5.5. Методы деревьев решений.		2	2			4
Тема 5.6. Ансамбли деревьев решений: Случайный лес.			1			1
Тема 5.7. Ансамбли деревьев решений: Градиентный бустинг.	2		1			3
Экзамен					36	36
Итого по дисциплине	26	26	56		36	144

Заочная форма обучения

Разделы / Темы	Лекции	Практические занятия	Самостоятельная работа	Текущий контроль	Контроль, промежуточная аттестация	Всего часов
8 семестр						
Раздел №1. «Введение в статистические методы анализа данных».						
Тема 1.1. Введение в статистические методы анализа данных.		1	3			4
Тема 1.2. Методы первичной обработки и представления данных.			3			3
Тема 1.3. Методы выявления и устранения аномальных наблюдений.		1	3			4
Раздел №2. «Множественный корреляционно-регрессионный анализ. Классическая модель множественной регрессии»						

Тема 2.1. Корреляционно-регрессионный анализ. Модель множественной регрессии.			3			3
Тема 2.2. Фиктивные переменные. Критерий Чоу проверки однородности.		1	3			4
Тема 2.3. Множественные коэффициенты корреляции.			3			3
Тема 2.4. Статистический анализ адекватности модели множественной регрессии.			3			3
Тема 2.5. Мультиколлинеарность факторов.			3			3
Раздел №3. «Обобщенная линейная модель множественной регрессии. Регрессионные модели с автокоррелированными остатками»						
Тема 3.1. Обобщенная линейная модель множественной регрессии. Обобщенный МНК			3			3
Тема 3.2. Гетероскедастичность. Методы выявления гетероскедастичности.		1	3			4
Тема 3.3. Взвешенный МНК. Устранение гетероскедастичности.			3			3
Тема 3.4. Регрессионные модели с автокоррелированными	1		3			4

остатками.						
Тема 3.5. Устранение автокорреляции			3			3
Раздел №4 «Кластерный анализ»						
Тема 4.1. Кластерный анализ. Основные этапы кластерного анализа.			3			3
Тема 4.2. Метрики расстояний. Свойства кластеров. Методы кластерного анализа.		1	3			4
Тема 4.3. Расстояния между кластерами. Методы группировки кластеров.			3			3
Тема 4.4. Иерархические методы кластеризации. Агломеративные методы.			3			3
Тема 4.5. Иерархические дивизимные методы кластеризации.	1	1	3			5
Тема 4.6. Примеры реализации иерархических методов кластеризации.			3			3
Тема 4.7. Итеративные методы. Метод к-средних.			3			3
Тема 4.8. Метод к-средних. Критерии качества кластеризации.			4			4

Тема 4.9. Метод нечетких к-средних		1	4			5
Раздел №5 «Методы классификации»						
Тема 5.1. Понятие методов классификации.			4			4
Тема 5.2. Логистическая регрессия.			4			4
Тема 5.3. Метрики качества классификации.	1		4			5
Тема 5.4. Метрики качества классификации. Регуляризация в задаче регрессии.			4			4
Тема 5.5. Методы деревьев решений.	1		4			5
Тема 5.6. Ансамбли деревьев решений: Случайный лес.		1	4			5
Тема 5.7. Ансамбли деревьев решений: Градиентный бустинг.			4			4
Экзамен					36	36
Итого по дисциплине	4	8	96		36	144

Структура и содержание дисциплины

Наименование разделов и тем	**Содержание тем**
Раздел №1. «Введение в статистические методы анализа данных».	
Тема 1.1. Введение в статистические методы анализа данных.	Предмет дисциплины «Статистические методы анализа данных». Основные типы данных. Основные этапы эмпирического исследования. Информационный подход к анализу данных.
Тема 1.2. Методы первичной обработки и представления данных.	Методы предобработки данных. Группировка данных. Методы нормализации или стандартизации данных. Приложение методов обработки данных к решению практических задач.

Наименование разделов и тем	Содержание тем
Тема 1.3. Методы выявления и устранения аномальных наблюдений.	Аномальные наблюдения или «выбросы». Классификация аномальных наблюдений. Критерий Смирнова-Граббса. Критерий Титьена-Мура. Критерий Ирвина. Графические методы выявления аномальных наблюдений. Методы устранения аномальных наблюдений
Раздел №2. «Множественный корреляционно-регрессионный анализ. Классическая модель множественной регрессии»	
Тема 2.1. Корреляционно-регрессионный анализ. Модель множественной регрессии.	Общий вид модели множественной регрессии. Классификация регрессии и корреляции. Модель множественной линейной регрессии. Матричная форма. МНК-оценки теоретических коэффициентов. Этапы построения модели множественной регрессии.
Тема 2.2. Фиктивные переменные. Критерий Чоу проверки однородности.	Фиктивные переменные во множественной регрессии. Критерий Чоу анализа структурных изменений. Методы формирования выборки наблюдений фиктивных переменных для построения регрессии.
Тема 2.3. Множественные коэффициенты корреляции.	Частный коэффициент корреляции, его свойства. Множественный коэффициент корреляции. Свойства. Частный коэффициент детерминации. Коэффициент детерминации. Свойства.
Тема 2.4. Статистический анализ адекватности модели множественной регрессии.	Теорема Гаусса-Маркова. Статистические свойства МНК-оценок. Проверка гипотез о статистической значимости теоретических коэффициентов регрессии. Проверка гипотезы о статистической значимости уравнения регрессии. Дисперсионный анализ. Дисперсионное отношение Фишера. Доверительные интервалы для теоретических коэффициентов регрессии. Практическая реализация проверки адекватности модели множественной регрессии
Тема 2.5. Мультиколлинеарность факторов.	Мультиколлинеарность факторов. Признаки мультиколлинеарности. Методы выявления мультиколлинеарности. Методы устранения мультиколлинеарности. Последствия мультиколлинеарности.
Раздел №3. «Обобщенная линейная модель множественной регрессии. Регрессионные модели с автокоррелированными остатками»	
Тема 3.1. Обобщенная линейная модель множественной регрессии. Обобщенный МНК	Обобщенная линейная модель множественной регрессии. Обобщенный МНК. Теорема Айткена. Взвешенный МНК.
Тема 3.2. Гетероскедастичность. Методы выявления гетероскедастичности.	Гетероскедастичность. Графический анализ остатков на гетероскедастичность. Тест Гольфельда-Квандта. Тест Уайта. Тест Глейзера.
Тема 3.3. Взвешенный МНК. Устранение гетероскедастичности.	Двухуровневая дисперсия. Взвешенный МНК. Пропорциональность дисперсии случайных ошибок значениям объясняющей переменной. Зависимость дисперсии случайных ошибок от значений объясняющей переменной. Примеры устранения гетероскедастичности.

Наименование разделов и тем	Содержание тем
Тема 3.4. Регрессионные модели с автокоррелированными остатками.	Понятие автокорреляции остатков. Причины автокорреляции остатков. Графический анализ остатков. Критерий Дарбина-Уотсона. Тест Бреуша-Годфри. Примеры практической реализации критериев проверки автокорреляции остатков.
Тема 3.5. Устранение автокорреляции	Метод взвешенного МНК для случая модели с автокоррелированными остатками. Методы построения оценки коэффициента автокорреляции первого порядка. Методы построения оценки коэффициента автокорреляции первого порядка.
Раздел №4 «Кластерный анализ»	
Тема 4.1. Кластерный анализ. Основные этапы кластерного анализа.	Понятие кластерного анализа. Основные этапы кластерного анализа. Представление данных в задаче кластеризации. Основные задачи и приложения кластерного анализа.
Тема 4.2. Метрики расстояний. Свойства кластеров. Методы кластерного анализа.	Метрики расстояний. Свойства метрик. Свойства кластеров. Классификация методов кластерного анализа.
Тема 4.3. Расстояния между кластерами. Методы группировки кластеров.	Расстояния между кластерами. Методы группировки кластеров. Дисперсионные методы группировки кластеров.
Тема 4.4. Иерархические методы кластеризации. Агломеративные методы.	Иерархические методы кластеризации. Алгоритм реализации агломеративных методов. Иерархические агломеративные методы кластеризации. Практическое приложение.
Тема 4.5. Иерархические дивизимные методы кластеризации.	Иерархические дивизимные методы. Алгоритм реализации дивизимных методов кластеризации. Иерархические дивизимные методы кластеризации. Практическое приложение.
Тема 4.6. Примеры реализации иерархических методов кластеризации.	Пример реализации агломеративного алгоритма. Пример реализации дивизимного алогритма. Дендограмма.
Тема 4.7. Итеративные методы. Метод к-средних.	Итеративные методы. Метод k-средних. Пример реализации метода к-средних.
Тема 4.8. Метод к-средних. Критерии качества кластеризации.	Функционалы качества разбиения. Критерии качества кластеризации. Статистические критерии качества разбиения.
Тема 4.9. Метод нечетких к-средних	Метод нечетких k-средних. Функционалы качества разбиения. Пример реализации метода нечетких к-средних.
Раздел №5 «Методы классификации»	
Тема 5.1. Понятие методов классификации.	Постановка задачи классификации. Основные методы классификации. Примеры задач классификации.
Тема 5.2. Логистическая регрессия.	Общий вид логистической регрессии. Оценка логистической регрессии. Понятие сбалансированной выборки. Методы оценки параметров логистической регрессии.
Тема 5.3. Метрики качества классификации.	Метрики качества классификации. Матрица ошибок. Доля правильных ответов, точность, полнота. Примеры расчета метрик качества классификации.
Тема 5.4. Метрики качества классификации. Регуляризация в задаче	ROC-кривая, площадь под кривой. L1-регуляризация. L2-регуляризация. Построение ROC-кривой.

Наименование разделов и тем	Содержание тем
регрессии.	
Тема 5.5. Методы деревьев решений.	Деревья решений. Основные алгоритмы построения деревьев решений. Основные алгоритмы построения деревьев решений.
Тема 5.6. Ансамбли деревьев решений: Случайный лес.	Понятие ансамблей решений. Случайный лес. Пример практической реализации алгоритма случайного леса.
Тема 5.7. Ансамбли деревьев решений: Градиентный бустинг.	Градиентный бустинг. Реализация ансамблей решений. Пример практической реализации алгоритма градиентного бустинга.

Занятия семинарского типа
(Практические занятия)

Общие рекомендации по подготовке к практическим занятиям. При подготовке к работе во время проведения занятий практического типа следует обратить внимание на следующие моменты: на процесс предварительной подготовки, на работу во время занятия, обработку полученных результатов, исправление полученных замечаний. Предварительная подготовка к учебному занятию практического типа заключается в изучении теоретического материала в отведенное для самостоятельной работы время, ознакомление с инструктивными материалами с целью осознания задач занятия.

Работа во время проведения занятия практического типа включает несколько моментов: а) консультирование обучающихся преподавателями с целью предоставления исчерпывающей информации, необходимой для самостоятельного выполнения предложенных преподавателем задач, б) самостоятельное выполнение заданий согласно обозначенной учебной программой тематики.

Раздел №1. «Введение в статистические методы анализа данных».

Тема 1.1. Введение в статистические методы анализа данных.

1. Классификация задач анализа данных: описательный, диагностический, предсказательный, предписывающий анализ — примеры из практики.
2. Обзор основных статистических понятий: генеральная совокупность, выборка, признаки, шкалы измерения (номинальная, порядковая, интервальная, отношений).
3. Знакомство с инструментами анализа: Python (pandas, scipy, statsmodels), R, Excel — выполнение простого расчёта среднего и дисперсии.

Тема 1.2. Методы первичной обработки и представления данных.

1. Построение частотных таблиц, гистограмм, box-plot’ов для количественных и категориальных переменных.
2. Расчёт основных описательных статистик: среднее, медиана, мода, дисперсия, стандартное отклонение, коэффициент вариации.
3. Визуализация распределений и выявление смещений (skewness) и эксцесса (kurtosis) с помощью Python или Excel.

Тема 1.3. Методы выявления и устранения аномальных наблюдений.

1. Применение правила 3σ и межквартильного размаха (IQR) для выявления выбросов.
2. Графическое обнаружение аномалий: scatter plot, box plot, Z-score heatmap.
3. Сравнение стратегий обработки выбросов: удаление, замена (медианой/средним), трансформация (логарифмирование), сохранение с пометкой.

Раздел №2. «Множественный корреляционно-регрессионный анализ. Классическая модель множественной регрессии»

Тема 2.1. Корреляционно-регрессионный анализ. Модель множественной регрессии.

1. Построение модели множественной линейной регрессии в Python (statsmodels) или Excel на основе реальных данных (например, цены на жильё).
2. Интерпретация коэффициентов регрессии и свободного члена.
3. Расчёт и анализ матрицы парных корреляций между признаками.

Тема 2.2. Фиктивные переменные. Критерий Чоу проверки однородности.

1. Кодирование категориальных признаков с помощью dummy-переменных и включение их в регрессионную модель.
2. Проверка гипотезы о структурных сдвигах: применение критерия Чоу на разбиении выборки (например, до/после пандемии).
3. Интерпретация результатов модели с фиктивными переменными (эффект «группы»).

Тема 2.3. Множественные коэффициенты корреляции.

1. Расчёт коэффициента множественной корреляции R и коэффициента детерминации R^2.
2. Сравнение R^2 и скорректированного R^2_{adj} при добавлении новых признаков.
3. Анализ вклада каждого фактора в объяснение дисперсии зависимой переменной.

Тема 2.4. Статистический анализ адекватности модели множественной регрессии.

1. Проверка значимости модели в целом (F-тест) и отдельных коэффициентов (t-тест).
2. Анализ остатков: проверка нормальности (график Q-Q), гомоскедастичности, автокорреляции.
3. Интерпретация p-value и доверительных интервалов коэффициентов.

Тема 2.5. Мультиколлинеарность факторов.

1. Вычисление VIF (Variance Inflation Factor) для оценки мультиколлинеарности.
2. Устранение мультиколлинеарности: исключение признаков, объединение, PCA.
3. Анализ последствий мультиколлинеарности: неустойчивость коэффициентов, завышенные стандартные ошибки.

Раздел №3. «Обобщенная линейная модель множественной регрессии. Регрессионные модели с автокоррелированными остатками»

Тема 3.1. Обобщенная линейная модель множественной регрессии. Обобщенный МНК

1. Постановка задачи ОМНК: когда обычный МНК неприменим.
2. Реализация взвешенного МНК в Python (statsmodels.WLS) на примере гетероскедастичных данных.
3. Сравнение оценок параметров по МНК и ОМНК.

Тема 3.2. Гетероскедастичность. Методы выявления гетероскедастичности.

1. Графический анализ: график остатков vs предсказанных значений.
2. Применение тестов: Бреуша–Пагана, Уайта, Голдфелда–Квандта.
3. Интерпретация результатов тестов и вывод о наличии/отсутствии гетероскедастичности.

Тема 3.3. Взвешенный МНК. Устранение гетероскедастичности.

1. Оценка весов на основе дисперсии остатков.
2. Построение модели с использованием WLS и сравнение её качества с OLS.
3. Проверка остатков после WLS на гомоскедастичность.

Тема 3.4. Регрессионные модели с автокоррелированными остатками.

1. Выявление автокорреляции: график остатков во времени, коэффициент Дарбина–Уотсона.
2. Построение модели временного ряда (например, продажи по месяцам) и анализ автокорреляции остатков.
3. Интерпретация последствий автокорреляции: смещение стандартных ошибок, неверные выводы о значимости.

Тема 3.5. Устранение автокорреляции

1. Применение метода Кохрана–Оркатта или Прайса–Уинстена для коррекции модели.
2. Использование авторегрессионной модели (AR(1)) для моделирования остатков.
3. Сравнение качества модели до и после устранения автокорреляции.

Раздел №4 «Кластерный анализ»

Тема 4.1. Кластерный анализ. Основные этапы кластерного анализа.

1. Формулировка цели кластеризации: сегментация клиентов, классификация товаров и т.п.
2. Подготовка данных: масштабирование (StandardScaler, MinMaxScaler), кодирование.
3. Выбор числа кластеров: метод локтя, silhouette score.

Тема 4.2. Метрики расстояний. Свойства кластеров. Методы кластерного анализа.

1. Сравнение метрик: Евклидово, Манхэттенское, косинусное расстояние — на одном наборе данных.
2. Анализ компактности и разделимости кластеров.
3. Классификация методов: иерархические vs итеративные.

Тема 4.3. Расстояния между кластерами. Методы группировки кластеров.

1. Реализация разных связок: одноэлементная, полная, средняя, центроидная.
2. Влияние типа связи на форму и количество кластеров.
3. Визуализация дендрограммы с разными методами связи.

Тема 4.4. Иерархические методы кластеризации. Агломеративные методы.

1. Построение агломеративной кластеризации вручную (на малом наборе точек).
2. Реализация с помощью scipy.cluster.hierarchy в Python.
3. Интерпретация дендрограммы и выбор порога отсечения.

Тема 4.5. Иерархические дивизимные методы кластеризации.

1. Сравнение агломеративного и дивизимного подходов.
2. Пример применения дивизимного метода (DIANA) на синтетических данных.
3. Обсуждение вычислительной сложности и применимости.

Тема 4.6. Примеры реализации иерархических методов кластеризации.

1. Кластеризация клиентов по поведенческим данным (RFM-анализ) с использованием иерархического метода.
2. Визуализация кластеров на 2D-графике после понижения размерности (PCA/t-SNE).
3. Оценка качества кластеров: внутрикластерная дисперсия, silhouette score.

Тема 4.7. Итеративные методы. Метод к-средних.

1. Реализация алгоритма k-means с нуля (на Python) и сравнение с sklearn.
2. Анализ чувствительности к начальным центрам (random_state).
3. Применение k-means к задаче сегментации студентов по успеваемости.

Тема 4.8. Метод к-средних. Критерии качества кластеризации.

1. Расчёт суммы квадратов расстояний до центроидов (inertia).
2. Построение графика "локтя" и silhouette plot для выбора оптимального k.
3. Сравнение устойчивости кластеров при многократном запуске.

Тема 4.9. Метод нечетких к-средних

1. Реализация Fuzzy C-Means (например, с помощью библиотеки scikit-fuzzy).
2. Интерпретация степеней принадлежности объектов к кластерам.
3. Сравнение чёткой и нечёткой кластеризации на данных с перекрытием групп.

Раздел №5 «Методы классификации»

Тема 5.1. Понятие методов классификации.

1. Классификация задач: бинарная vs многоклассовая, сбалансированные vs несбалансированные выборки.
2. Примеры прикладных задач: кредитный скоринг, диагностика, спам-фильтрация.
3. Разделение данных на train/test, стратификация.

Тема 5.2. Логистическая регрессия.

1. Построение модели логистической регрессии в Python (sklearn.linear_model.LogisticRegression).
2. Интерпретация коэффициентов и вероятностей.
3. Построение ROC-кривой и расчёт AUC.

Тема 5.3. Метрики качества классификации.

1. Построение confusion matrix и расчёт accuracy, precision, recall, F1-score.
2. Анализ метрик в зависимости от порога классификации.
3. Применение метрик к несбалансированным данным (например, обнаружение мошенничества).

Тема 5.4. Метрики качества классификации. Регуляризация в задаче регрессии.

1. Сравнение L1 (Lasso) и L2 (Ridge) регуляризации в логистической регрессии.
2. Подбор гиперпараметра регуляризации (C) с помощью кросс-валидации.
3. Анализ влияния регуляризации на переобучение и интерпретируемость модели.

Тема 5.5. Методы деревьев решений.

1. Построение дерева решений (sklearn.tree.DecisionTreeClassifier) и визуализация.
2. Анализ важности признаков по дереву.
3. Настройка гиперпараметров: глубина, min_samples_split, критерий (gini/entropy).

Тема 5.6. Ансамбли деревьев решений: Случайный лес.

1. Обучение модели Random Forest и сравнение с одиночным деревом.
2. Оценка важности признаков по случайным лесам.
3. Подбор числа деревьев и других гиперпараметров.

Тема 5.7. Ансамбли деревьев решений: Градиентный бустинг.

1. Обучение модели Gradient Boosting (sklearn.ensemble.GradientBoostingClassifier).
2. Сравнение производительности и скорости обучения с Random Forest.
3. Использование XGBoost/LightGBM для повышения качества предсказаний.

Раздел 5. Учебно-методическое обеспечение самостоятельной работы обучающихся по дисциплине

Наряду с чтением лекций и проведением практических занятий неотъемлемым элементом учебного процесса является *самостоятельная работа.* При самостоятельной работе достигается конкретное усвоение учебного материала, развиваются теоретические способности, столь важные для успешной подготовки и защиты выпускной работы бакалавра. Формы самостоятельной работы, обучаемых могут быть разнообразными. Самостоятельная работа включает: изучение литературы, веб-ресурсов, оценку, обсуждение и рецензирование публикуемых статей; ответы на контрольные вопросы; решение задач; самотестирование. Выполнение всех видов самостоятельной работы увязывается с изучением конкретных тем.

Типовые задания для самостоятельной, представлены в фонде оценочных средств по дисциплине.

5.1. Индивидуальные задания для курсовых работ

Курсовая работа позволяет выявить степень владения базовыми знаниями, умениями необходимыми для обучения, и определить уровень владения новым материалом. Рекомендации по подготовке, оформлению и защите курсовой работы (курсового проекта), предусмотренной учебным планом, изложены в Методических указаниях по соответствующей дисциплине.

5.1.1. Примерная тематика курсовых работ (проектов)

Курсовая работа не предусмотрена учебным планом.

Раздел 6. Оценочные и методические материалы по образовательной программе (фонд оценочных средств) для проведения текущего контроля успеваемости и промежуточной аттестации

В процессе освоения учебной дисциплины для оценивания сформированности требуемых компетенций используются оценочные материалы (фонды оценочных средств).
Типовые тестовые задания, типовые практические задания, типовые задания для контрольных работ и материалы для диагностической работы представлены в фонде оценочных средств по дисциплине. Промежуточная аттестация проводится в форме зачета/ зачета с оценкой/экзамена.

6.1. Материалы для оценки результатов промежуточной аттестации

Перечень вопросов для подготовки к промежуточной аттестации

1. Основные этапы эмпирического исследования.
2. Дискретные, непрерывные данные. Пространственная выборка, временной ряд, панельные данные.
3. Аномальные наблюдения. Критерии проверки аномальности наблюдений.
4. Общий вид модели множественной регрессии. Линейная модель множественной регрессии. Матричная форма.
5. Основные предпосылки регрессионного анализа. Метод наименьших квадратов. МНК-оценки параметров множественной линейной регрессии. Теорема Гаусса-Маркова.
6. Статистические свойства МНК оценок. Проверка гипотезы о статистической значимости теоретических коэффициентов регрессии. Доверительные интервалы для теоретических коэффициентов.
7. Частный коэффициент корреляции, множественный коэффициент корреляции. Проверка гипотезы о статистической значимости частного коэффициента корреляции.
8. Коэффициент детерминации. Свойства. Дисперсионное отношение Фишера. Проверка гипотезы о статистической значимости уравнения регрессии.
9. Обобщенная линейная модель множественной регрессии. Обобщенный МНК.
10. Гомоскедастичность, гетероскедастичность. Критерии проверки остатков на гетероскедастичность. Методы устранения гетероскедастичности.
11. Автокорреляция остатков. Критерии проверки автокорреляции. Устранение автокорреляции.
12. Понятие кластерного анализа. Основные этапы кластерного анализа.
13. Основные меры сходства объектов. Расстояние между объектами: основные виды, формулы расчета.
14. Свойства кластера: плотность, дисперсия, размеры, форма, отделимость. Расстояние между кластерами: основные виды, формулы расчета.
15. Иерархические агломеративные методы: понятие, алгоритм реализации.
16. Иерархические дивизимные методы: понятие, алгоритм реализации.
17. Метод k-средних: понятие, алгоритм реализации. Функционалы качества разбиения объектов методом k-средних.
18. Метод нечетких k-средних: понятие, алгоритм реализации. Функционалы качества разбиения объектов методом k-средних.
19. Постановка задачи классификации. Основные методы классификации.
20. Общий вид логистической регрессии. Оценка логистической регрессии. Понятие сбалансированной выборки.

21. Метрики качества классификации.
22. Методы деревьев решений.
23. Понятие ансамблей решений. Случайный лес. Градиентный бустинг.

6.2. Оценивание ответов на вопросы и выполнения заданий промежуточной аттестации

При оценке знаний учитывается уровень сформированности компетенций:

1. Уровень усвоения теоретических положений дисциплины, правильность формулировки основных понятий и закономерностей.
2. Уровень знания фактического материала в объеме программы.
3. Логика, структура и грамотность изложения вопроса.
4. Умение связать теорию с практикой.
5. Умение делать обобщения, выводы.

Шкала оценивания на экзамене, зачете с оценкой

Оценка	Критерии выставления оценки
Отлично	Обучающийся должен: - продемонстрировать глубокое и прочное усвоение знаний программного материала; - исчерпывающе, последовательно, грамотно и логически стройно изложить теоретический материал; - правильно формулировать определения; - продемонстрировать умения самостоятельной работы с литературой; - уметь сделать выводы по излагаемому материалу.
Хорошо	Обучающийся должен: - продемонстрировать достаточно полное знание программного материала; - продемонстрировать знание основных теоретических понятий; - достаточно последовательно, грамотно и логически стройно излагать материал; - продемонстрировать умение ориентироваться в литературе; - уметь сделать достаточно обоснованные выводы по излагаемому материалу.
Удовлетворительно	Обучающийся должен: - продемонстрировать общее знание изучаемого материала; - показать общее владение понятийным аппаратом дисциплины; - уметь строить ответ в соответствии со структурой излагаемого вопроса; - знать основную рекомендуемую программой учебную литературу.
Неудовлетворительно	Обучающийся демонстрирует: - незнание значительной части программного материала; - не владение понятийным аппаратом дисциплины; - существенные ошибки при изложении учебного материала; - неумение строить ответ в соответствии со структурой излагаемого вопроса; - неумение делать выводы по излагаемому материалу.

Шкала оценивания на зачете

Оценка	Критерии выставления оценки
«Зачтено»	Обучающийся должен: уметь строить ответ в соответствии со структурой излагаемого вопроса; продемонстрировать прочное,

	достаточно полное усвоение знаний программного материала; продемонстрировать знание основных теоретических понятий; правильно формулировать определения; последовательно, грамотно и логически стройно изложить теоретический материал; продемонстрировать умения самостоятельной работы с литературой; уметь сделать достаточно обоснованные выводы по излагаемому материалу.
«Не зачтено»	Обучающийся демонстрирует: незнание значительной части программного материала; не владение понятийным аппаратом дисциплины; существенные ошибки при изложении учебного материала; неумение строить ответ в соответствии со структурой излагаемого вопроса; неумение делать выводы по излагаемому материалу.

Методические материалы, определяющие процедуру оценивания сформированных компетенций в соответствии с ООП

Качество знаний характеризуется способностью обучающегося точно, структурированно и уместно воспроизводить информацию, полученную в процессе освоения дисциплины, в том виде, в котором она была изложена в учебном издании или преподавателем.

Умения, как правило, формируются на занятиях семинарского типа. Задания, направленные на оценку умений, в значительной степени требуют от обучающегося проявления стереотипности мышления, т.е. способности выполнить работу по образцам, с которыми он работал в процессе обучения. Преподаватель же оценивает своевременность и правильность выполнения задания.

Навыки можно трактовать как автоматизированные умения, развитые и закрепленные осознанным самостоятельным трудом. Навыки формируются при самостоятельном выполнении обучающимися практико ориентированных заданий, моделирующих решение им производственных и социокультурных задач в соответствующей области профессиональной деятельности, как правило, при выполнении домашних заданий, курсовых проектов (работ), научно-исследовательских работ, прохождении практик, при работе индивидуально или в составе группы и т.д.

Устный опрос – это процедура, организованная как специальная беседа преподавателя с группой обучающихся (фронтальный опрос) или с отдельными обучающимися (индивидуальный опрос) с целью оценки сформированности у них основных понятий и усвоения учебного материала. Устный опрос может использоваться как вид контроля и метод оценивания формируемых компетенций (как и качества их формирования) в рамках самых разных форм контроля, таких как: собеседование, коллоквиум, зачет, экзамен по дисциплине. Устный опрос (УО) позволяет оценить знания и кругозор обучающегося, умение логически построить ответ, владение монологической речью и иные коммуникативные навыки. УО обладает большими возможностями воспитательного воздействия преподавателя. Воспитательная функция УО имеет ряд важных аспектов: профессионально-этический и нравственный аспекты, дидактический (систематизация материала при ответе, лучшее запоминание материала при интеллектуальной концентрации), эмоциональный (радость от успешного прохождения собеседования) и др. Обучающая функция УО состоит в выявлении деталей, которые по каким-то причинам оказались недостаточно осмысленными в ходе учебных занятий и при подготовке к зачёту или экзамену. УО обладает также мотивирующей функцией: правильно организованные собеседование, коллоквиум, зачёт и экзамен могут стимулировать учебную деятельность студента, его участие в научной работе.

Тесты являются простейшей формой контроля, направленной на проверку владения терминологическим аппаратом, современными информационными технологиями и конкретными знаниями в области фундаментальных и прикладных дисциплин. Тест может предоставлять возможность выбора из перечня ответов (один или несколько правильных ответов).

Семинарские занятия. Основное назначение семинарских занятий по дисциплине – обеспечить глубокое усвоение обучающимися материалов лекций, прививать навыки

самостоятельной работы с литературой, воспитывать умение находить оптимальные решения в условиях изменяющихся отношений, формировать современное профессиональное мышление обучающихся. На семинарских занятиях преподаватель проверяет выполнение самостоятельных заданий и качество усвоения знаний, умений, определяет уровень сформированности компетенций.

Раздел 7. Методические указания для обучающихся по основанию дисциплины

Освоение обучающимся учебной дисциплины предполагает изучение материалов дисциплины на аудиторных занятиях и в ходе самостоятельной работы. Аудиторные занятия проходят в форме лекций, семинаров и практических занятий. Самостоятельная работа включает разнообразный комплекс видов и форм работы обучающихся.

Для успешного освоения учебной дисциплины и достижения поставленных целей необходимо внимательно ознакомиться с настоящей рабочей программы учебной дисциплины. Следует обратить внимание на список основной и дополнительной литературы, которая имеется в электронной библиотечной системе Университета. Эта информация необходима для самостоятельной работы обучающегося.

При подготовке к аудиторным занятиям необходимо помнить особенности каждой формы его проведения.

Подготовка к учебному занятию лекционного типа. С целью обеспечения успешного обучения обучающийся должен готовиться к лекции, поскольку она является важнейшей формой организации учебного процесса, поскольку: знакомит с новым учебным материалом; разъясняет учебные элементы, трудные для понимания; систематизирует учебный материал; ориентирует в учебном процессе.

С этой целью: внимательно прочитайте материал предыдущей лекции; ознакомьтесь с учебным материалом по учебнику и учебным пособиям с темой прочитанной лекции; внесите дополнения к полученным ранее знаниям по теме лекции на полях лекционной тетради; запишите возможные вопросы, которые вы зададите лектору на лекции по материалу изученной лекции; постарайтесь уяснить место изучаемой темы в своей подготовке; узнайте тему предстоящей лекции (по тематическому плану, по информации лектора) и запишите информацию, которой вы владеете по данному вопросу

Предварительная подготовка к учебному занятию семинарского типа заключается в изучении теоретического материала в отведенное для самостоятельной работы время, ознакомление с инструктивными материалами с целью осознания задач занятия.

Самостоятельная работа. Для более углубленного изучения темы задания для самостоятельной работы рекомендуется выполнять параллельно с изучением данной темы. При выполнении заданий по возможности используйте наглядное представление материала.

Подготовка к зачету, экзамену. К зачету, экзамену необходимо готовится целенаправленно, регулярно, систематически и с первых дней обучения по данной дисциплине. Попытки освоить учебную дисциплину в период зачетно-экзаменационной сессии, как правило, приносят не слишком удовлетворительные результаты. При подготовке к зачету обратите внимание на защиту практических заданий на основе теоретического материала. При подготовке к экзамену по теоретической части выделите в вопросе главное, существенное (понятия, признаки, классификации и пр.), приведите примеры, иллюстрирующие теоретические положения.

Раздел 8. Учебно-методическое и информационное обеспечение дисциплины

Основная литература

1. Вероятностно-статистический анализ данных с использованием MS Excel. В 2 частях. Ч.2. Математико-статистические методы анализа данных: учебник и практикум / В. С. Мхитарян, В. Ф. Шишов, Д. В. Искоркин, А. Ю. Козлов. — Москва: КУРС, 2024. — 415 с. — ISBN 978-5-907535-89-3. — Текст: электронный // Цифровой образовательный ресурс IPR SMART: [сайт]. — URL: https://www.iprbookshop.ru/144845.html

2. Дятлов, А. В. Методы многомерного статистического анализа данных в социологии: учебник / А. В. Дятлов, П. Н. Лукичев. — Ростов-на-Дону, Таганрог: Издательство Южного федерального университета, 2023. — 236 с. — ISBN 978-5-9275-4265-9. — Текст: электронный // Цифровой образовательный ресурс IPR SMART: [сайт]. — URL: https://www.iprbookshop.ru/135652.html

3. Дронов, С. В. Анализ многомерных статистических данных: монография / С. В. Дронов. — Москва, Вологда: Инфра-Инженерия, 2025. — 308 с. — ISBN 978-5-9729-2524-7. — Текст: электронный // Цифровой образовательный ресурс IPR SMART: [сайт]. — URL: https://www.iprbookshop.ru/155978.html

Дополнительная литература

4. Романко, В. К. Статистический анализ данных в психологии: учебное пособие / В. К. Романко. — 5-е изд. — Москва: Лаборатория знаний, 2025. — 313 с. — ISBN 978-5-93208-845-6. — Текст: электронный // Цифровой образовательный ресурс IPR SMART: [сайт]. — URL: https://www.iprbookshop.ru/149631.html

5. Орлов, А. И. Искусственный интеллект: статистические методы анализа данных: учебник / А. И. Орлов. — Москва: Ай Пи Ар Медиа, 2022. — 843 с. — ISBN 978-5-4497-1470-1. — Текст : электронный // Цифровой образовательный ресурс IPR SMART : [сайт]. — URL: https://www.iprbookshop.ru/117029.html

6. Вероятностно-статистический анализ данных с использованием MS Excel. В 2 частях. Ч.1. Вероятностные методы анализа данных: учебник и практикум / В. С. Мхитарян, В. Ф. Шишов, Д. В. Искоркин, А. Ю. Козлов. — Москва: КУРС, 2024. — 360 с. — ISBN 978-5-907535-92-3. — Текст: электронный // Цифровой образовательный ресурс IPR SMART: [сайт]. — URL: https://www.iprbookshop.ru/144901.html

8.1 Требования к материально-техническому и учебно-методическому обеспечению

8.1.1. Университет располагает материально-технической базой, соответствующей действующим противопожарным правилам и нормам и обеспечивающей проведение всех видов дисциплинарной и междисциплинарной подготовки, практической и научно-исследовательской работ обучающихся, предусмотренных учебным планом.

Помещения представляют собой учебные аудитории для проведения учебных занятий, предусмотренных программой магистратуры, оснащенные оборудованием и техническими средствами обучения, состав которых определяется в рабочих программах дисциплин (модулей).

В Университете имеются специализированные аудитории для проведения занятий по информационным технологиям.

Помещения для самостоятельной работы обучающихся оснащены компьютерной техникой с возможностью подключения к сети "Интернет" и обеспечением доступа к электронной информационно-образовательной среде Университета.

Электронная информационно-образовательная среда Университета включает:

1. Официальный сайт Университета (https://www.iile.ru/)
2. Электронная информационно-образовательная среда «1С: Университет» договор от 10.09.2018 г. №ПРКТ-18281 (бессрочно)
3. Программы для ЭВМ. Система дистанционного обучения «Mirapolis» - Лицензионный договор №107/06/24-к от 27.06.2024 (Спецификация к Лицензионному договору №107/06/24-к от 27.06.2024, срок действия с 01.07.2026 по 30.06.2027 г.) https://impe.lms.mirapolis.ru/mira/
4. Программа для ЭВМ. Виртуальная комната «Mirapolis» - Лицензионный договор №107/06/24-к от 27.06.2024 (Спецификация к Лицензионному договору №107/06/24-к от 27.06.2024, срок действия с 01.07.2026 по 30.06.2027 г.) https://impe.lms.mirapolis.ru/mira/
5. Система тестирования INDIGO лицензионное соглашение (Договор от 07.11.2018 г. №Д-54792, бессрочно) http://212.48.35.211:85/

8.1.2. Университет обеспечен необходимым комплектом лицензионного и свободно распространяемого программного обеспечения, в том числе отечественного производства (состав определяется в рабочих программах дисциплин (модулей) и подлежит обновлению при необходимости).

Перечень лицензионного программного обеспечения, в том числе отечественного производства:

1. Операционная система «Атлант» - Atlant Academ от 24.01.2024 г. (бессрочно)
2. Антивирусное программное обеспечение Kaspersky Endpoint Security для бизнеса – Расширенный Russian Edition договор-оферта № Tr000941765 от 16.10.2025 г.

8.1.3. Обучающимся обеспечен доступ (удаленный доступ) к современным профессиональным базам данных и информационным справочным системам, состав которых определяется в рабочих программах дисциплин (модулей) и обновляется при необходимости, но не реже одного раз в год.

Перечень современных профессиональных баз данных и информационных справочных систем:

1. Информационно-поисковая система «Консультант Плюс» - Договор №МИ-ВИП-79717-56/2022 (бессрочно)
2. Электронно-библиотечная система IPRsmart лицензионный договор от 01.09.2024 г. №11652/24С (срок действия до 31.08.2027 г.) https://www.iprbookshop.ru/
3. Научная электронная библиотека eLIBRARY лицензионный договор SCIENC INDEX № SIO - 3079/2026 от 30.01.2026 г. (срок действия до 29.01.2027г.) https://elibrary.ru

8.1.4. Обучающиеся из числа инвалидов и лиц с ОВЗ обеспечены электронными образовательными ресурсами в формах, адаптированных к ограничениям их здоровья.

Раздел 9. Материально-техническое обеспечение образовательного процесса

Учебная аудитория для проведения занятий лекционного типа, занятий семинарского типа, групповых и индивидуальных консультаций, текущего контроля и промежуточной аттестации	Оборудование: специализированная мебель (мебель аудиторная (столы, стулья, доска аудиторная навесная), стол преподавателя, стул преподавателя. Технические средства обучения: персональный компьютер; мультимедийное оборудование (проектор, экран).
Помещение для самостоятельной работы	Специализированная мебель (столы, стулья), персональные компьютеры с возможностью подключения к сети «Интернет» и обеспечением доступа в электронную информационно-образовательную среду Университета