



МИНОБРНАУКИ РОССИИ
федеральное государственное бюджетное образовательное учреждение
высшего образования
«ИРКУТСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ»
ФГБОУ ВО «ИГУ»

Кафедра социально-экономических и математических дисциплин



Директор МИЭЛ  О. В. Архипкин
«22» апреля 2026 г.

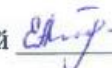
Рабочая программа дисциплины

| | |
|---------------------------|---|
| Наименование дисциплины | Б.В.05 Технологии интеллектуального анализа данных / Data Mining Technologies |
| Направление подготовки | 38.04.01 Экономика |
| Направленность подготовки | Экономика производства продукции и услуг/ Economics of Goods and Services Production |
| Квалификация выпускника | магистр |
| Форма обучения | очная с применением электронного обучения, дистанционных образовательных технологий |

Согласовано с УМК МИЭЛ ИГУ:
Протокол № 3 от «23» марта 2026 г.

Рекомендовано кафедрой:
Протокол № 7 от «05» марта 2026 г.

Председатель  Е. В. Крайнова

Зав. кафедрой  Е. В. Антонова

Иркутск – 2026 г.

СОДЕРЖАНИЕ

| | стр. |
|--|------|
| I. Цели и задачи дисциплины (модуля) | 3 |
| II. Место дисциплины (модуля) в структуре ОПОП. | 3 |
| III. Требования к результатам освоения дисциплины (модуля) | 4 |
| IV. Содержание и структура дисциплины (модуля) | 5 |
| 4.1 Содержание дисциплины, структурированное по темам, с указанием видов учебных занятий и отведенного на них количества академических часов | 5 |
| 4.2 План внеаудиторной самостоятельной работы обучающихся по дисциплине | 7 |
| 4.3 Содержание учебного материала | 9 |
| 4.3.1 Перечень семинарских, практических занятий и лабораторных работ | 10 |
| 4.3.2. Перечень тем (вопросов), выносимых на самостоятельное изучение в рамках самостоятельной работы студентов | 11 |
| 4.4. Методические указания по организации самостоятельной работы студентов | 11 |
| V. Учебно-методическое и информационное обеспечение дисциплины (модуля) | 17 |
| а) перечень литературы | 17 |
| б) периодические издания | |
| в) список авторских методических разработок г) базы данных, поисково-справочные и информационные системы | |
| VI. Материально-техническое обеспечение дисциплины (модуля) | 18 |
| 6.1. Учебно-лабораторное оборудование: | 18 |
| 6.2. Программное обеспечение: | 20 |
| 6.3. Технические и электронные средства обучения: | 20 |
| VII. Образовательные технологии | 20 |
| VIII. Оценочные материалы для текущего контроля и промежуточной аттестации | 21 |

I. ЦЕЛИ И ЗАДАЧИ ДИСЦИПЛИНЫ

Цели:

- формирование у студентов комплекса теоретических знаний в области проведения анализа данных с использованием интеллектуальных систем обработки информации, а также практических навыков, необходимых для внедрения и практического использования таких систем.

Задачи:

- получить представление о современных подходах и процедурах, выполняемых с использованием интеллектуальных систем;
- познакомиться с основными методами интеллектуального анализа данных;
- изучить основные технологии построения алгоритмов анализа данных с использованием информационных систем на основе технологий прикладного искусственного интеллекта;
- выработать навыки использования современных информационных технологий и программных средств, реализующих алгоритмы интеллектуального анализа данных.

II. МЕСТО ДИСЦИПЛИНЫ В СТРУКТУРЕ ОПОП ВО

Дисциплина «Технологии анализа данных» относится к части программы, формируемой участниками образовательных отношений. Изучается на 2 курсе в первом семестре.

Дисциплина имеет содержательно-логические связи с дисциплинами «Методология экономической науки», «Микроэкономика», «Макроэкономика», «Бизнес-проектирование», «Стратегический менеджмент», «Стратегический маркетинг».

Знания, умения и навыки, формируемые данной учебной дисциплиной, необходимы для подготовки к процедуре защиты и защиты выпускной квалификационной работы.

III. ТРЕБОВАНИЯ К РЕЗУЛЬТАТАМ ОСВОЕНИЯ ДИСЦИПЛИНЫ

Процесс освоения дисциплины направлен на формирование следующих компетенций в соответствии с ФГОС ВО и ОПОП ВО «Экономика производства продукции и услуг/ Economics of Goods and Services Production» по направлению подготовки 38.04.01 Экономика:

ПК-1 – Способен осуществлять экономический анализ деятельности организации.

Перечень планируемых результатов обучения по дисциплине, соотнесенных с индикаторами достижения компетенций

| Компетенция | Индикаторы компетенций | Результаты обучения |
|---|--|---|
| ПК-1 Способен осуществлять экономический анализ деятельности организации | ИДК _{ПК1.1} Осуществляет сбор, мониторинг и обработку данных для проведения расчетов экономических показателей организации | Знать: основные понятия, термины интеллектуального анализа данных (ИАД) Уметь: проводить моделирование и анализ данных с применением методов и программного инструментария интеллектуального анализа данных Владеть навыками выбора оптимальных способов решения стандартных экономических задач с применением методов интеллектуального анализа данных |
| | ИДК _{ПК1.2} Рассчитывает и анализирует экономические | Знать: основные алгоритмы методов и моделей ИАД и подходы к их применению Уметь: применять методы ИАД в среде табличного процессора "MsExcel" для решения экономических, |

| | | |
|--|---|--|
| | показатели результатов деятельности организации | управленческих, научно-исследовательских задач Владеть: навыками использования методов и программного инструментария интеллектуального анализа данных |
|--|---|--|

IV. СОДЕРЖАНИЕ И СТРУКТУРА ДИСЦИПЛИНЫ

Объем дисциплины составляет 3 зачетные единицы, 108 часов, в том числе 0,2 зачетных единиц, 8 часов - на зачет.

Из них реализуется с применением электронного обучения, дистанционных образовательных технологий 18 часов.

Форма промежуточной аттестации: зачет.

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

| № п/н | Тема | Семестр | Всего часов | Из них практическая подготовка обучающихся | Виды учебной работы, включая самостоятельную работу обучающихся, практическую подготовку и трудоемкость (в часах) | | | | Форма текущего контроля успеваемости/ Форма промежуточной аттестации |
|----------|--|----------|-------------|--|---|--|--------------|--|---|
| | | | | | Контактная работа преподавателя с обучающимися | | | Самостоятельная работа, в том числе, внеаудиторная самостоятельная | |
| | | | | | Лекция | Семинар/ Практическое, лабораторное занятие/ | Консультация | | |
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| 1 | Введение в интеллектуальный анализ данных | 3 | 8 | | 2 | 2 | | 4 | |
| 1.1. | Основные понятия интеллектуального анализа данных. Постановка задачи Подготовка данных. Просмотр данных. Построение моделей. Исследование и проверка моделей. Развертывание и обновление моделей. Интерпретация и прогнозирование. | | | | | | | | Тест |
| 2 | Технология и алгоритмы интеллектуального анализа | 3 | 8 | | 2 | 2 | | 4 | |

| | | | | | | | | | |
|------|---|---|--|------------|-----------|-----------|--|-----------|------------------------|
| | данных | | | | | | | | |
| 2.1. | Средства для прогноза, анализа и представления выводов. | | | | | | | | Собеседование |
| 3 | Статистические методы | 3 | | | | | | | |
| 3.1. | Методы предварительного анализа данных. Описательная статистика. Характеристики центральной тенденции. Характеристики вариации. Очистка данных. Нормировка и стандартизация данных. | | | 8 | 2 | 2 | | 4 | Индивидуальное задание |
| 3.2. | Анализ выбросов и аномальных значений. Анализ пропущенных значений. | | | 8 | 2 | 2 | | 4 | Индивидуальное задание |
| 3.3. | Корреляционный анализ. Регрессионный анализ. Дисперсионный анализ. | | | 17 | 3 | 2 | | 12 | Индивидуальное задание |
| 4 | Визуализация как этап анализа данных | 3 | | 20 | 2 | 2 | | 16 | |
| 4.1. | Визуализация и предварительное исследование данных. Графики и диаграммы. Инфографика и схемы. Визуализация данных в программе MsExcel. | | | | | | | | Индивидуальное задание |
| 5 | Бизнес аналитика и информационные панели (дашборды) | 3 | | 13 | 2 | 3 | | 8 | Индивидуальное задание |
| 6 | Метод анализа иерархий | 3 | | 6 | 1 | 1 | | 4 | Собеседование |
| 7 | Модели классификации | 3 | | 6 | 1 | 1 | | 4 | Собеседование |
| 8 | Модели кластеризации | 3 | | 6 | 1 | 1 | | 4 | Собеседование |
| | Итого | | | 100 | 18 | 18 | | 64 | Зачет |

4.2. План внеаудиторной самостоятельной работы обучающихся по дисциплине

| Семестр | Название раздела, темы | Самостоятельная работа обучающихся | | | Оценочное средство | Учебно-методическое обеспечение самостоятельной работы |
|---------|---|---|---------------------|---------------------|------------------------|--|
| | | Вид самостоятельной работы | Сроки выполнения | Трудоемкость (час.) | | |
| 3 | Введение в интеллектуальный анализ данных | Изучение материала по данной теме для выполнения заданий по текущему контролю, подготовки к практическому занятию | 1-2 недели семестра | 4 | Тест | |
| 3 | Технология и алгоритмы интеллектуального анализа данных | Изучение материала по данной теме для выполнения заданий по текущему контролю, подготовки к практическому занятию | 3-4 недели | 4 | Собеседование | |
| 3 | Статистические методы | Изучение материала по данной теме для выполнения заданий по текущему контролю, подготовки к практическому занятию | 5-6 недели | 20 | Индивидуальное задание | |
| 3 | Визуализация как этап анализа данных | Изучение материала по данной теме для выполнения заданий по текущему контролю, подготовки к практическому занятию | 7-8 недели | 16 | Индивидуальное задание | |
| 3 | Бизнес-аналитика и информационные панели (дашборды) | Изучение материала по данной теме для выполнения заданий по текущему контролю, подготовки к практическому занятию | 9-10 недели | 8 | Индивидуальное задание | |
| 3 | Метод анализа иерархий | Изучение материала по данной теме для выполнения заданий по текущему контролю, подготовки к практическому занятию | 11-12 недели | 4 | Собеседование | |
| 3 | Модели классификации | Изучение материала по данной теме для выполнения заданий по текущему контролю, подготовки к | 13-15 недели | 4 | Собеседование | |

| | | | | | | |
|--|----------------------|---|-------------|-----------|---------------|--|
| | | практическому занятию | | | | |
| 3 | Модели кластеризации | Изучение материала по данной теме для выполнения заданий по текущему контролю, подготовки к практическому занятию | 17-я неделя | 4 | Собеседование | |
| Общий объем самостоятельной работы по дисциплине (час) | | | | 64 | | |
| Из них объем самостоятельной работы с применением электронного обучения, дистанционных образовательных технологий (час) | | | | 18 | | |

4.3.Содержание учебного материала

| № п/п | Название темы | Содержание |
|-------|---|--|
| 1 | Введение в интеллектуальный анализ данных | Основные понятия интеллектуального анализа данных. Интеллектуальный анализ данных (ИАД). Постановка и классификация задач ИАД. Данные. Понятие данных. Типы шкал наборов данных. Типы наборов данных. Задача классификации и регрессии. Задача поиска ассоциативных правил. Задача кластеризации. Примеры практического применения дисциплины ИАД: медицина, биоинформатика, геология, банковское дело, торговля, другие области применения. |
| 2 | Процесс интеллектуального анализа данных | Основные этапы процесса интеллектуального анализа данных. Анализ предметной области. Постановка задачи. Сбор и подготовка данных. Этапы процесса интеллектуального анализа данных, связанные с построением, проверкой, оценкой, выбором и коррекцией моделей. Анализ данных. Интерпретация и прогнозирование. |
| 3 | Статистические методы | Методы предварительного анализа данных. Описательная статистика. Характеристики центральной тенденции. Характеристики вариации. Очистка данных. Нормировка и стандартизация данных. Анализ выбросов и аномальных значений. Обработка пропущенных значений. Корреляционный анализ. Коэффициент корреляции Пирсона. Ранговая корреляция. Критерий Спирмена. Множественная и частная корреляции. Регрессионный анализ. Дисперсионный анализ. |
| 4 | Визуализация как этап анализа данных | Понятие визуализации данных, базовые принципы. Исследовательская и презентационная задачи визуализации. Графики и диаграммы. Инфографика и схемы. Визуализация данных в программе MsExcel. |
| 5 | Бизнес аналитика и информационные панели (дашборды) | Бизнес аналитика и информационные панели (дашборды). Построение сводных таблиц и диаграмм для вычисления, сведения и анализа данных в программе MsExcel. Построение дашборда в программе MsExcel. |
| 6 | Метод анализа иерархий | Классификация задач. Участники процесса принятия решений. Альтернативы. Критерии. Типы задач принятия решений. Предварительная структуризация задач принятия решений. Преимущества и недостатки метода. Иерархическое представление. Шкала отношений. Матрицы парных сравнений. Собственные векторы и значения матриц. |
| 7 | Методы классификации | Деревья решений. Анализ данных с использованием деревьев решений. Анализ признаков объектов. Критерии и функции качества разбиения. Индекс Джини. Энтропия. Ошибка классификации. Остановка построения дерева. Сокращение дерева. Извлечение правил из деревьев. Алгоритмы построения деревьев решений. |

| | | |
|---|-------------------|--|
| 8 | Кластерный анализ | Основы кластерного анализа. Математические характеристики кластера. Критерии качества кластеризации. Выделение значимых кластеров. Иерархические агломеративные и дивизимные методы иерархического кластерного анализа. Неиерархические итерационные методы кластерного анализа. Алгоритм k-средних. |
|---|-------------------|--|

4.3.1. Перечень семинарских, практических занятий и лабораторных работ

| № темы | Наименование семинаров, практических и лабораторных работ | Трудоемкость (час.) | | Оценочные средства | Формируемые компетенции (индикаторы)* |
|--------|--|---------------------|----------------------------|------------------------|--|
| | | Всего часов | Из них практич. подготовка | | |
| 1 | 2 | 3 | 4 | 5 | 6 |
| 1 | Сущность интеллектуального анализа данных и машинного обучения | 2 | | Собеседование | ИДК _{ПК1.1} ИДК _{ПК1.2} |
| 2 | Основы интеллектуального анализа данных | 2 | | Собеседование | ИДК _{ПК1.1} ИДК _{ПК1.2} |
| 3 | Статистические методы | 6 | | Индивидуальное задание | ИДК _{ПК1.1} ИДК _{ПК1.2} |
| 4 | Визуализация как этап анализа данных | 2 | | Индивидуальное задание | ИДК _{ПК1.1} ИДК _{ПК1.2} |
| 5 | Бизнес аналитика и информационные панели (дашборды) | 3 | | Индивидуальное задание | ИДК _{ПК1.1} ИДК _{ПК1.2} |
| 6 | Метод анализа иерархий | 1 | | Собеседование | ИДК _{ПК1.1} ИДК _{ПК1.2} |
| 7 | Методы классификации | 1 | | Собеседование | ИДК _{ПК1.1} ИДК _{ПК1.2} |
| 8 | Кластерный анализ | 1 | | Собеседование | ИДК _{ПК1.1} ИДК _{ПК1.2} |
| | Итого часов | 18 | | | |

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

| № п/п | Тема | Задание | Формируемая компетенция | ИДК |
|-------|---|------------------------|-------------------------|--|
| 1 | 2 | 3 | 4 | 5 |
| 1 | Сущность интеллектуального анализа данных. | Доклад | ПК-1 | ИДК _{ПК1.1} ИДК _{ПК1.2} |
| 2 | Основы интеллектуального анализа данных | Доклад | ПК-1 | ИДК _{ПК1.1} ИДК _{ПК1.2} |
| 3 | Статистические методы | Индивидуальное задание | ПК-1 | ИДК _{ПК1.1} ИДК _{ПК1.2} |
| 4 | Визуализация как этап анализа данных | Индивидуальное задание | ПК-1 | ИДК _{ПК1.1} ИДК _{ПК1.2} |
| 5 | Бизнес аналитика и информационные панели (дашборды) | Индивидуальное задание | ПК-1 | ИДК _{ПК1.1} ИДК _{ПК1.2} |
| 6 | Метод анализа иерархий | Доклад | ПК-1 | ИДК _{ПК1.1} ИДК _{ПК1.2} |
| 7 | Методы классификации | Доклад | ПК-1 | ИДК _{ПК1.1} ИДК _{ПК1.2} |
| 8 | Кластерный анализ | Доклад | ПК-1 | ИДК _{ПК1.1} ИДК _{ПК1.2} |

4.4. Методические указания по организации самостоятельной работы студентов

В рамках теоретической подготовки к практическим занятиям студент должен самостоятельно подготовить ответы на следующие вопросы по темам дисциплины:

Тема 3. Статистические методы

1. Что такое методы предварительного анализа данных и почему они важны при работе с данными?
2. Какие основные методы описательной статистики используются для анализа данных?
3. Какие характеристики центральной тенденции помогают определить типичное значение в наборе данных?
4. Какие характеристики вариации используются для измерения разброса значений в данных?
5. Какие методы можно применить для очистки данных от выбросов и ошибок?
6. Что такое нормировка данных и какие преимущества она может иметь при анализе данных?
7. Какие методы можно использовать для обработки пропущенных значений в данных?
8. Какие статистические показатели и графики могут помочь визуализировать распределение данных?

9. Какие методы можно применить для исследования связи между переменными в наборе данных?
10. Методы отбора признаков в задачах линейной регрессии
11. Значимость коэффициентов линейной регрессии
12. Методы борьбы с переобучением в различных методах линейной регрессии
13. Вероятностная постановка задач регрессии
14. Взаимосвязь между сложностью регрессионной модели и размером набора данных
15. Вычисление оптимальных параметров логистической регрессии
16. Проблема определения оптимальных параметров линейной регрессии
17. Метрики качества регрессии
18. Что такое корреляционный анализ и какая его основная цель?
19. Что представляет собой коэффициент корреляции Пирсона и как он интерпретируется?
20. Какая роль ранговой корреляции в анализе данных и в чем ее отличие от коэффициента корреляции Пирсона?
21. Что такое критерий Спирмена и как он применяется в корреляционном анализе?
22. В чем заключается множественная корреляция и какие преимущества она имеет перед простой корреляцией?
23. Как можно провести частную корреляцию и какие выводы о взаимосвязи переменных можно сделать на основе результатов?

Тема 7. Методы классификации

1. Проблемы практического применения классификационных моделей
2. Сравнение методов классификации
3. Метрики качества разбиения в задачах классификации и регрессии.
4. Композиции алгоритмов, применяемые в задачах классификации
5. Вычисление оптимальных параметров моделей классификации
6. Методы борьбы с переобучением в различных методах классификации
7. Проблема определения оптимальных параметров классификации
8. Метрики качества классификации: доля правильных ответов, матрица ошибок, точность и полнота, AUC, индекс Джини, чувствительность к размеру классов.
9. Чувствительность и специфичность метода классификации
10. Что такое дерево решений и какой основной принцип лежит в его основе?
11. Какие виды задач можно решать с помощью деревьев решений в анализе данных?
12. Каким образом происходит анализ признаков объектов при построении дерева решений?
13. Какие критерии и функции качества разбиения используются для определения наилучшего разделения в дереве решений?
14. Что такое индекс Джини и как он используется для измерения неопределенности в дереве решений?
15. Как энтропия применяется в контексте деревьев решений и какую роль она играет в принятии решений?
16. Что представляет собой ошибка классификации и как она учитывается при построении дерева решений?
17. Какие методы и критерии можно использовать для остановки построения дерева решений и избежания переобучения?
18. Что такое сокращение дерева?

19. Как можно извлечь правила из построенного дерева решений и как они могут быть интерпретированы в контексте анализа данных?

Тема 8. Модели кластеризации

1. Проблемы практического применения моделей кластеризации
2. Сравнение методов кластеризации.
3. Подходы к классификации методов кластеризации
4. Методы итеративной кластеризации: преимущества и недостатки, область применения
5. Методы иерархической кластеризации: преимущества и недостатки, область применения
6. Методы плотностной кластеризации: преимущества и недостатки, область применения
7. Методы борьбы с переобучением в различных методах кластеризации
8. Проблема определения оптимальных параметров кластеризации
9. Метрики качества классификации

Материалы для самостоятельной работы обучающихся и методические рекомендации по её выполнению расположены на Образовательном портале ИГУ в личном кабинете магистранта <http://educa.isu.ru/>

Индивидуальные задания

1. Сформулируйте задачи, возникающие в заданном по вариантам виде деятельности, которые можно было бы решить с использованием машинного обучения (необходимо выделить как минимум по одной задаче регрессии, классификации и кластеризации).
2. Опишите каждую задачу по следующей схеме: сущность задачи, класс задачи, состав признаков (наименование, тип данных, ограничения на значения), состав меток (наименование, тип данных, ограничения на значения). Варианты видов деятельности: 1) банковская деятельность; 2) электронная коммерция; 3) риэлтерская деятельность; 4) информационная безопасность; 5) розничная торговля; 6) сельское хозяйство; 7) транспортные услуги; 8) туристические услуги; 9) услуги страхования
3. На основании вариантов видов деятельности, найдите во внешних открытых репозиториях несколько наборов данных, им соответствующих, и сделайте их описание по следующей схеме: а) наименование репозитория с указанием его интернет-адреса, б) краткое и полное наименование набора данных, в) идентификатор набора данных, г) краткая характеристика (решаемые задачи, первоисточник данных и т.д.), д) количественные параметры набора данных (размер выборки), е) описание признаков и меток.
4. Проведите анализ данных в наборе на доступность, точность, взаимосвязанность, полноту, непротиворечивость, однозначность, релевантность, надежность и своевременность.
5. Варианты видов деятельности: 1) банковская деятельность; 2) электронная коммерция; 3) риэлтерская деятельность; 4) информационная безопасность; 5) розничная торговля; 6) сельское хозяйство; 7) транспортные услуги; 8) туристические услуги; 9) услуги страхования.
6. На основании набора данных о покупках в розничном магазине, занимающемся продажей подарков и сувениров через интернет-сайт, осуществите кластеризацию заказов и покупателей. Источник данных: <https://archive.ics.uti.edu/ml/datasets/online+retail>

7. На основании данных о транзакциях по кредитным картам необходимо выявить те из них, которые являются мошенническими. Источник данных: <https://www.kaggle.com/mlgulb/creditcardfraud>
3. На основании данных о производстве конфет в США с января 1972 по настоящий момент необходимо спрогнозировать индустриальный продуктовый индекс (IPG3113N - универсальный индекс уровня производства, который измеряется как % от уровня производства 2012 года). Горизонт прогнозирования: временной интервал в 24 месяца. Источник данных: <https://fred.stlouisfed.org/series/IPG3113N>
8. Подготовьте Развернутый ответ на следующие вопросы: 1. Обзор задач, решаемых алгоритмами машинного обучения. 2. Классификация алгоритмов машинного обучения 3. Программная реализация алгоритма линейной регрессии 4. Программная реализация алгоритма логистической регрессии. 5. Программная реализация дерева решений. 6. Программная реализация алгоритма AdaBoost. 7. Обзор существующих алгоритмов классификации. Алгоритм k-means

Подготовка докладов по следующим темам:

- Тема 1.** Сущность интеллектуального анализа данных и машинного обучения. Сущность интеллектуального анализа данных. Интеллектуальный анализ данных в бизнесе (описание стандартных задач). Технологии интеллектуального анализа
- Тема 2.** Основы интеллектуального анализа данных. Обзор программных решений для интеллектуального анализа данных.
- Тема 3.** Контейнер для хранения больших наборов однородных данных. Обзор операций с наборами однородных данных
- Тема 4.** Обработка и анализ наборов разнородных данных Функциональные возможности python-библиотеки Pandas. Контейнер для хранения больших наборов разнородных данных. Обзор операций с наборами разнородных данных.
- Тема 5.** Визуализация и предварительное исследование данных Функциональные возможности python-библиотеки Matplotlib. Построение двумерных графиков и диаграмм. Построение трехмерных графиков и диаграмм
- Тема 6.** Python-библиотека Scikit-learn как инструмент интеллектуального анализа данных Общие сведения о библиотеке Scikit-learn. Обзор основных способов загрузки данных. Реализация подготовки данных к анализу. Моделирование данных и анализ качества моделей
- Тема 7.** Модели регрессии Сущность регрессионного анализа. Линейные модели регрессии. Нелинейные модели регрессии
- Тема 8.** Модели классификации Классификация как задача машинного обучения. Линейные модели классификации. Модели классификации с нелинейными разделяющими поверхностями
- Тема 9.** Модели кластеризации Кластерный анализ как задача машинного обучения. Модели итеративной кластеризации. Модели иерархической кластеризации. Модели плотностной кластеризации

Требования по подготовке доклада

Доклад — вид самостоятельной работы, представляющий собой краткое информативное сообщение по конкретному вопросу (проблеме). В докладе приводятся различные точки зрения на предмет исследования, а также высказывается собственная позиция в рамках тематической проблематики.

Доклад должен содержать:

- введение, содержащее постановку проблемы;
- основную часть, содержащую логически выдержанное изложение темы (предпосылок и путей решения поставленной проблемы);
- краткие выводы, обобщающие позицию автора по проблеме;
- список использованной литературы (указывается только та литература, которой фактически пользовался автор; все случаи использования источников - цитаты, сведения, оценки и т.д. - отмечаются ссылками в виде сносок или примечаний с указанием страниц источника).

Объем доклада должен составлять 7-10 страниц (до 4 тыс. слов) печатного текста (шрифт Times, размер 12, полуторный интервал). Включение в доклад материалов, не имеющих прямого отношения к теме, а также источников, не указанных в базовом списке литературы (в частности, текстов из Интернета), служит основанием для признания работы не соответствующей требованиям или существенного снижения общей оценки.

Доклад оценивается по следующим критериям:

- самостоятельность выполнения работы, способность аргументировано защищать основные положения и выводы;
- соответствие формальным требованиям: структура, наличие списка литературы, сносок, грамотность изложения;
- способность сформулировать проблему;
- уровень освоения темы и изложения материала: обоснованность отбора материала, использование первичных источников, способность самостоятельно осмысливать выявленные факты, логика изложения;
- четкость и содержательность выводов.

Требования по подготовке презентации

Общие требования к презентации:

- Презентация не должна быть меньше 10 слайдов.
- Первый лист - это титульный лист, на котором обязательно должны быть представлены: название проекта; название выпускающей организации; фамилия, имя, отчество автора; вуз, где учится автор проекта и его группа.
- Следующим слайдом должно быть содержание, где представлены основные части (моменты) презентации. Желательно, чтобы из содержания по гиперссылке можно перейти на необходимую страницу и вернуться вновь на содержание.
- Дизайн-эргономические требования: сочетаемость цветов, ограниченное количество объектов на слайде, цвет текста.
- Презентация не может состоять из сплошного не структурированного текста.
- Последними слайдами урока-презентации должны быть глоссарий и список литературы.

Создание презентации состоит из трех этапов:

- I. Планирование презентации - это многошаговая процедура, включающая определение целей, формирование структуры и логики подачи материала. Планирование презентации включает в себя:
 1. Определение целей;
 2. Определение основной идеи презентации;
 3. Подбор дополнительной информации;
 4. Планирование выступления;
 5. Создание структуры презентации;
 6. Проверка логики подачи материала;
 7. Подготовка заключения.

- II. Разработка презентации - методологические особенности подготовки слайдов презентации, включая вертикальную и горизонтальную логику, содержание и соотношение текстовой и графической информации.
- III. Репетиция презентации - это проверка и отладка созданной презентации. В оформлении презентаций выделяют два блока: оформление слайдов и представление информации на них. Для создания качественной презентации необходимо соблюдать ряд требований, предъявляемых к оформлению данных блоков.

Оформление слайдов:

| | |
|-----------------------------|--|
| Стиль | Соблюдайте единый стиль оформления. Избегайте стилей, которые будут отвлекать от самой презентации. Вспомогательная информация (управляющие кнопки) не должны преобладать над основной информацией (текстом, иллюстрациями). |
| Фон | Для фона предпочтительны холодные тона |
| Использование цвета | На одном, один для текста. Для фона и текста используйте контрастные цвета. слайде рекомендуется использовать не более трех цветов: один для фона, один для заголовка. Обратите внимание на цвет гиперссылок (до и после использования). |
| Анимационные эффекты | Используйте возможности компьютерной анимации для представления информации на слайде. Не стоит злоупотреблять различными анимационными эффектами, они не должны отвлекать внимание от содержания информации на слайде. |

Представление информации:

| | |
|--|---|
| Содержание информации | Используйте короткие слова и предложения. Минимизируйте количество предлогов, наречий, прилагательных. Заголовки должны привлекать внимание аудитории. |
| Расположение информации на странице | Предпочтительно горизонтальное расположение информации. Наиболее важная информация должна располагаться в центре экрана. Если на слайде располагается картинка, надпись должна располагаться под ней. |
| Шрифты | Для заголовков - не менее 24. Для информации не менее 18. Шрифты без засечек легче читать с большого расстояния. Нельзя смешивать разные типы шрифтов в одной презентации. Для выделения информации следует использовать жирный шрифт, курсив или подчеркивание. Нельзя злоупотреблять прописными буквами (они читаются хуже строчных). |
| Способы выделения информации | Следует использовать: а) рамки; границы, заливку; б) штриховку, стрелки; в) рисунки, диаграммы, схемы для иллюстрации наиболее важных фактов. |
| Объем информации | Не стоит заполнять один слайд слишком большим объемом информации: люди могут одновременно запомнить не более трех фактов, выводов, определений. |
| Виды слайдов | Для обеспечения разнообразия следует использовать разные виды слайдов: а) с текстом; б) с таблицами; в) с диаграммами. |

V. УЧЕБНО-МЕТОДИЧЕСКОЕ И ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

а) перечень литературы

9. Основная литература:

1. Алексеев Д.С., Щекочихин О.В. Технологии интеллектуального анализа данных: учебник для вузов. - СПб.: Лань, 2022. - 176 с.
2. Берман К. Основы Python для Data Science: пер. с англ. – СПб.: Питер, 2023. – 272 с.
3. Бессмертный, И. А. Интеллектуальные системы : учебник и практикум для вузов / И. А. Бессмертный, А. Б. Нугуманова, А. В. Платонов. — Москва : Издательство Юрайт, 2023. — 243 с. — (Высшее образование).
4. Петрунин Ю.Ю. Информационные технологии анализа данных. Data analysis: учебное пособие. - 2-е изд. - М.: КДУ, 2010. - 292 с.
5. Игнатов, Гейб Текст майнинг. Интеллектуальный анализ текста : дизайн исследований, сбор данных и методы анализа / Гейб Игнатов, Рада Михальча ; пер. с англ. А. М. Морозовой Харьков : Гуманитарный Центр, 2021 341 с. : ил. ISBN 978-617-7528-54-7 ISBN 978-1506337005 ((англ.))
6. Пальмов С.В. Интеллектуальный анализ данных : учебное пособие / Пальмов С.В.. — Самара: Поволжский государственный университет телекоммуникаций и информатики, 2017 — 127 с. — ISBN 2227-8397.

Дополнительная литература:

1. Тирни, Б Наука о данных. Базовый курс / Брендан Тирни – Альпина Паблишер., 2021. – 222 с.
2. Дж., Морроу, Как вытащить из данных максимум. Навыки аналитики для неспециалистов / Морроу Джордан – Альпина Паблишер, 2021. – 256 с.
3. Дюк, В. Data Mining: учебный курс / В. Дюк, А. Самойленко. – СПб., 2001. 368 с.
4. Нейтан Яу. Искусство визуализации в бизнесе. – М.: Манн, Иванов и Фербер, 2013. – 352 с.
5. Билл Фрэнкс Революция в аналитике: Как в эпоху Big Data улучшить ваш бизнес с помощью операционной аналитики / Билл Фрэнкс. — М.: Альпина Паблишер, 2020. — 320 с. — ISBN 978-5-9614-5302-7.
6. Желязны, Д. Говори на языке диаграмм: пособие по визуальным коммуникациям / Джин Желязны; пер. с англ. [А. Мучника и Ю.Корнилович] – 5-е изд. – М.: Манн, Иванов и Фербер, 2012. — 304 с.
7. Боровиков В.П. Популярное введение в современных анализ данных и машинное обучение на STATISTICA. – М.: Горячая линия – Телеком, 2021. – 354 с.
8. Груздев А.В. Прогнозное моделирование в IBM SPSS Statistics и R: Метод деревьев решений. – М.: ДМК Пресс, 2016. – 278 с.
9. Жерон О. Прикладное машинное обучение с помощью Scikit-Learn, Keras и TensorFlow: концепции, инструменты и техники для создания интеллектуальных систем: пер. с англ. – 2-е изд. – СПб.: ООО «Диалектика», 2020. – 1040 с.

б) периодические издания

1. Журнал «Инновации».
2. Журнал «Наука, инновации, технологии».
3. Журнал «КомпьютерПресс».
4. Журнал «PC Week / RE (Компьютерная неделя)».
5. Журнал «Информационное общество».
6. Журнал «CRN / RE (ИТ-бизнес)».

7. Издание о высоких технологиях.
8. Вычислительная математика и математическая физика
9. The Journal of Machine Learning Research (JMLR)
10. Machine Learning
11. Data Mining And Knowledge Discovery
12. Intelligent Data Analysis
13. Pattern Recognition And Image Analysis: Advances In Mathematical Theory And Applications

Интернет-ресурсы

1. <http://www.fasie.ru>
2. <http://www.government.gov.ru>
3. <http://www.innovbusiness.ru>
4. <http://www.rsci.ru>
5. <http://www.rvca.ru>
6. <http://www.technopark.al.ru>
7. <http://economics.edu.ru>
8. <http://e.lib.vlsu.ru>

VI. МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

6.1. Учебно-лабораторное оборудование:

Для реализации данной дисциплины используются специальные помещения:

- учебные аудитории для лекционных и семинарских занятий, групповых и индивидуальных консультаций, текущего контроля и промежуточной аттестации:

Аудитория на 100 посадочных мест, укомплектованная специализированной учебной мебелью и техническими средствами обучения с возможностью подключения к сети «Интернет» и обеспечением доступа в электронную информационно-образовательную среду организации:

- Wi-Fi;
- активный микшерный пульт Yamaha + микрофон;
- колонки – 2*200Вт;
- экран настенный – ScreenMedia 180x180;
- проектор – BenQ MX661;
- компьютер преподавателя (AMD ATHLON II x3);
- ПО – Microsoft Windows, Microsoft Office Professional Plus 2010, Kaspersky Endpoint Security;
- наборы демонстрационного оборудования и учебно-наглядных пособий, обеспечивающие тематические иллюстрации, соответствующие рабочей программе дисциплины Б.В.06 «Технологии интеллектуального анализа данных».

- аудитории для организации самостоятельной работы:

Аудитория на 40 посадочных мест, укомплектованная специализированной учебной мебелью и техническими средствами обучения с возможностью подключения к сети «Интернет» и обеспечением доступа в электронную информационно-образовательную среду организации:

- интерактивная доска – Panasonic UB-T880-G;
- проектор – AcerS1212;
- ноутбук – Lenovo (Intel Core2Duo);
- маркерная доска.
- ПО – Microsoft Office Professional Plus 2010; Архиватор WinRAR; Wi-Fi.

Аудитория на 15 посадочных мест, укомплектованная специализированной

учебной мебелью и техническими средствами обучения с возможностью подключения к сети «Интернет» и обеспечением доступа в электронную информационно-образовательную среду организации:

- 15 компьютеров (AMD Athlon64) с доступом к сети интернет;
- 1 компьютер оператора (AMD Athlon64);
- ПО – Microsoft Windows 7; Microsoft Office Professional Plus 2010; 1С:Предприятие 8. Комплект для обучения в высших и средних учебных заведениях; Архиватор RAR WinRAR 5; Far Manager v3; Консультант Плюс: Версия Проф; Kaspersky Endpoint Security.

Аудитория на 15 посадочных мест, укомплектованная специализированной учебной мебелью и техническими средствами обучения с возможностью подключения к сети «Интернет» и обеспечением доступа в электронную информационно-образовательную среду организации:

- 15 компьютеров (AMD Athlon64) с доступом к сети интернет;
- 1 компьютер оператора (AMD Athlon64);
- ПО – Microsoft Windows 7; Microsoft Office Professional Plus 2010; 1С:Предприятие 8. Комплект для обучения в высших и средних учебных заведениях; Архиватор RAR WinRAR 5; Far Manager v3; Консультант Плюс: Версия Проф; Kaspersky Endpoint Security.

- помещение для хранения и профилактического обслуживания учебного оборудования:

- шкафы, расходные материалы,
- 6 ноутбуков,
- 4 переносных комплекта:
- Экран на штативе ScreenMedia Apollo, проектор переносной – Epson EB-X24,
- ноутбук HP 255 G7 (Intel Core i5),
- ПО – Microsoft Windows 10 OEM, Microsoft Office Professional Plus 2010,
- Kaspersky Endpoint Security

6.2. Программное обеспечение

Операционные системы:

- Microsoft Windows (версии Vista, 7, 10)
- ALT Linux

Пакеты офисных приложений:

- Microsoft Office (ред. Professional Plus, Standard; вер. 2007, 2010, 2013)
- LibreOffice, OpenOffice

Интернет-браузеры:

- YandexBrowser
- Атом
- Mozilla Firefox
- Google Chrome

Прикладное ПО для работы с документами:

- Архиватор RAR WinRAR (5.x Версия Академическая)
- Far Manager
- Adobe Reader

Средства антивирусной защиты:

- Kaspersky Endpoint Security

Онлайн сервисы:

- Видео конференц система bbb.isu.ru (система BigBlueButton)
- Образовательный портал educa.isu.ru (система LMS Moodle)
- Видеохостинг cloud.isu.ru (система NextCloud)

Прикладное ПО для работы с графическими и аудио/видео файлами:

— VLC Player

6.3. Технические и электронные средства:

Компьютеры, проекторы, позволяющие проводить на лекциях и семинарах презентации, разработанные с помощью пакета прикладных программ MS Power Point, использовать наглядные, иллюстрированные материалы.

На образовательном портале Иркутского государственного университета <https://educa.isu.ru/> размещены презентации лекций и тесты для контроля знаний по всем темам дисциплины Б.В.06 «Технологии интеллектуального анализа данных».

VII. ОБРАЗОВАТЕЛЬНЫЕ ТЕХНОЛОГИИ

Наименование тем занятий с указанием форм/ методов/ технологий обучения

| № п/п | Тема занятия | Вид занятия | Форма / Методы//технологии дистанционного, интерактивного обучения | Количество часов |
|--------------|---|------------------------------|---|-------------------------|
| 1 | 2 | 3 | 4 | 5 |
| 1 | Введение в интеллектуальный анализ данных | Лекция, практическое занятие | Групповые проекты, дискуссии, кейс-методы | 4 |
| 2 | Технология и алгоритмы интеллектуального анализа данных | Лекция, практическое занятие | Групповые проекты, дискуссии, кейс-методы | 8 |
| 3 | Статистические методы | Лекция, практическое занятие | Групповые проекты, дискуссии, кейс-методы | 18 |
| 4 | Визуализация как этап анализа данных | Лекция, практическое занятие | Групповые проекты, дискуссии, кейс-методы | 8 |
| 5 | Бизнес аналитика и информационные панели (дашборды) | Лекция, практическое занятие | Групповые проекты, дискуссии, кейс-методы | 10 |
| 6 | Метод анализа иерархий | Лекция, практическое занятие | Групповые проекты, дискуссии, кейс-методы | 4 |
| 7 | Методы классификации | Лекция, практическое занятие | Групповые проекты, дискуссии, кейс-методы | 4 |
| 8 | Кластерный анализ | Лекция, практическое занятие | Групповые проекты, дискуссии, кейс-методы | 4 |
| | Итого | | | 68 |

VIII. ОЦЕНОЧНЫЕ МАТЕРИАЛЫ ДЛЯ ТЕКУЩЕГО КОНТРОЛЯ И ПРОМЕЖУТОЧНОЙ АТТЕСТАЦИИ

8.1. Оценочные материалы для проведения текущего контроля

Демонстрационный вариант теста

1. Что включает в себя ИНТЕЛЛЕКТУАЛЬНЫЙ АНАЛИЗ ДАННЫХ?
 - а) Статистический анализ данных
 - б) Машинное обучение
 - в) Обработку естественного языка
 - г) Все вышеперечисленное

2. Какие методы входят в ИНТЕЛЛЕКТУАЛЬНЫЙ АНАЛИЗ ДАННЫХ?
 - а) Кластеризация
 - б) Классификация
 - в) Регрессия
 - г) Все вышеперечисленное

3. Что такое данные?
 - а) Факты, цифры, статистика, тексты и т.д.
 - б) Только числовые значения
 - в) Только текстовые значения

4. Какие типы шкал наборов данных не существуют?
 - а) Номинальная
 - б) Порядковая
 - в) Интервальная
 - г) Ординальная
 - д) Кумулятивная

5. Какую задачу решает классификация?
 - а) Распределение объектов по заранее заданным классам
 - б) Поиск взаимосвязей между признаками
 - в) Поиск аномалий в данных

6. Какую задачу решает регрессия?
 - а) Распределение объектов по заранее заданным классам
 - б) Поиск взаимосвязей между признаками
 - в) Предсказание числовых значений на основе имеющихся данных

7. Какую задачу решает поиск ассоциативных правил?
 - а) Распределение объектов по заранее заданным классам
 - б) Поиск взаимосвязей между признаками
 - в) Поиск аномалий в данных

8. Какую задачу решает кластеризация?
 - а) Распределение объектов по заранее заданным классам
 - б) Поиск взаимосвязей между признаками
 - в) Группировка объектов на основе их сходства

9. Какой коэффициент используется для измерения силы линейной связи между двумя переменными?
- Коэффициент корреляции Пирсона
 - Ранговая корреляция
 - Критерий Спирмена
10. Какой тип графика наиболее удобен для сравнения значений различных категорий?
- Столбчатая диаграмма
 - Круговая диаграмма
 - Линейный график
 - Точечная диаграмма
11. Какой инструмент программы MsExcel используется для создания дашборда?
- Формулы и функции
 - Панель инструментов
 - Параметры страницы
 - Сводные таблицы
12. Что такое дашборд в программе MsExcel?
- Крупномасштабная диаграмма
 - Интерактивная панель управления для отображения данных
 - Сложный алгоритм обработки данных
 - Режим просмотра данных без возможности редактирования

8.2. Оценочные материалы для промежуточной аттестации

Примерный перечень вопросов к зачету

- Основные понятия интеллектуального анализа данных. Понятие данных. Типы шкал наборов данных. Типы наборов данных.
- Задача классификации и регрессии. Задача поиска ассоциативных правил. Задача кластеризации.
- Примеры практического применения дисциплины **ИНТЕЛЛЕКТУАЛЬНЫЙ АНАЛИЗ ДАННЫХ**: медицина, биоинформатика, геология, банковское дело, торговля, другие области применения.
- Основные этапы процесса интеллектуального анализа данных.
- Методы предварительного анализа данных. Описательная статистика.
- Характеристики центральной тенденции. Характеристики вариации.
- Очистка данных. Нормировка и стандартизация данных.
- Анализ выбросов и аномальных значений.
- Обработка пропущенных значений.
- Корреляционный анализ. Коэффициент корреляции Пирсона. Ранговая корреляция. Критерий Спирмена. Множественная и частная корреляции.
- Регрессионный анализ.
- Основные методы трансформации данных, включая фильтрацию, сортировку, объединение таблиц и другие операции в MsExcel. Создание запросов. Возможности Power Query для очистки и преобразования данных.
- Понятие визуализации данных, базовые принципы. Исследовательская и презентационная задачи визуализации.
- Графики и диаграммы. Инфографика и схемы. Визуализация данных в программе MsExcel.

15. Бизнес аналитика и информационные панели (дашборды). Построение сводных таблиц и диаграмм для вычисления, сведения и анализа данных в программе MsExcel.
16. Построение дашборда в программе MsExcel.
17. Метода анализа иерархий.
18. Методы классификации. Деревья решений. Анализ данных с использованием деревьев решений. Анализ признаков объектов. Критерии и функции качества разбиения.
19. Индекс Джини. Энтропия. Ошибка классификации. Остановка построения дерева. Сокращение дерева. Извлечение правил из деревьев. Алгоритмы построения деревьев решений.
20. Основы кластерного анализа. Математические характеристики кластера. Критерии качества кластеризации. Выделение значимых кластеров. Иерархические агломеративные и дивизимные методы иерархического кластерного анализа. Неиерархические итерационные методы кластерного анализа.
21. Алгоритм k-средних.

КРИТЕРИИ ОЦЕНКИ ПО ТЕКУЩЕМУ КОНТРОЛЮ

Критерии оценки результатов выполнения тестовых заданий

Основным критерием оценки результатов выполнения тестовых заданий является коэффициент усвоения учебного материала, который определяется как отношение правильных ответов к общему количеству заданий.

Критерии оценки результатов тестирования, 10-балльная шкала

10 баллов = 100% (правильных ответов)

96% ≤ 9 баллов ≤ 99%

91% ≤ 8 баллов ≤ 95%

86% ≤ 7 баллов ≤ 90%

80% ≤ 6 баллов ≤ 85%

70% ≤ 5 баллов ≤ 79%

60% ≤ 4 балла ≤ 69%

50% ≤ 3 балла ≤ 59%

40% ≤ 2 балла ≤ 49%

1 балл ≤ 40 %

КРИТЕРИИ ОЦЕНКИ ПО ПРОМЕЖУТОЧНОЙ АТТЕСТАЦИИ

Критерии оценки знаний студента на зачете:


Оценка «зачтено» ставится, если студент:

- обнаружил всестороннее, систематическое и глубокое знание учебного материала, успешно выполнил предусмотренные программой задания, усвоил рекомендованную литературу;
- обнаружил полное знание учебного материала, успешно выполнил предусмотренные программой задания, усвоил рекомендованную литературу, продемонстрировал систематический характер знаний по дисциплине и способен к их самостоятельному пополнению и обновлению в ходе дальнейшей учебы и профессиональной деятельности;
- показал знание основного учебного материала в объеме, необходимом для дальнейшей учебы и предстоящей работы по профессии; справлялся с выполнением заданий, предусмотренных программой, но допустил погрешности в ответе на зачете и при выполнении заданий, которые не носят принципиального характера (студент способен исправить ошибки под руководством преподавателя).

Оценка «не зачтено» выставляется, если студент:

- обнаружил пробелы в знаниях основного учебного материала, допустил принципиальные ошибки в выполнении предусмотренных программой заданий. Ответы студента носит несистематизированный, отрывочный, поверхностный характер. Студент не понимает существа излагаемых им вопросов, что свидетельствует о его неспособности продолжать обучение или приступать к профессиональной деятельности без дополнительных занятий по дисциплине.

Разработчик:



(подпись)

доцент
(занимаемая должность)

Т. Д. Ахмеджанова
(Ф.И.О.)

Программа составлена в соответствии с требованиями ФГОС ВО по направлению и направленности подготовки 38.04.01 Экономика «Экономика производства продукции и услуг / Economics of Goods and Services Production».

Программа рассмотрена на заседании кафедры социально-экономических и математических дисциплин «05» марта 2026 г. Протокол № 7

Зав. кафедрой социально-
экономических и математических
дисциплин, доцент



Е. В. Антонова

Настоящая программа, не может быть воспроизведена ни в какой форме без предварительного письменного разрешения кафедры-разработчика программы.