

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ
федеральное государственное бюджетное образовательное учреждение
высшего образования
«Тольяттинский государственный университет»

Б1.В.ДВ.01.01
(индекс дисциплины)

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ

Технологии анализа данных

(наименование дисциплины)

по направлению подготовки
09.04.03 Прикладная информатика

направленность (профиль)
Прикладной анализ данных

Форма обучения: очная

Год набора: 2023

Общая трудоемкость: 3 ЗЕ

Распределение часов дисциплины по семестрам

| Семестр | 3 | Итого |
|--|------------|------------|
| Форма контроля | Зачет | |
| Вид занятий | | |
| Лекции | 18 | 18 |
| Лабораторные | | |
| Практические | 18 | 18 |
| Руководство: курсовые работы (проекты) / РГР | | |
| Промежуточная аттестация | 0,25 | 0,25 |
| Контактная работа | 36,25 | 36,25 |
| Самостоятельная работа | 71,75 | 71,75 |
| Контроль | | |
| Итого | 108 | 108 |

Рабочую программу составил(и):

Доцент, к.т.н., Климов В.С.

(должность, ученое звание, степень, Фамилия И.О.)

Рецензирование рабочей программы дисциплины:



Отсутствует



Рецензент

(должность, ученое звание, степень, Фамилия И.О.)

Рабочая программа дисциплины составлена на основании ФГОС ВО и учебного плана направления подготовки (специальности)

09.04.03 Прикладная информатика

Срок действия рабочей программы дисциплины до «31» августа 2025г.

УТВЕРЖДЕНО

На заседании кафедры «Прикладная математика и информатика»

(протокол заседания №1 от «30» августа 2022 г.).

1. Цель освоения дисциплины

Цель освоения дисциплины – формирование представления о современных методах и технологиях интеллектуального анализа на основе хранилищ данных, которые применяются на практике в различных областях науки и техники.

2. Место дисциплины в структуре ОПОП ВО

Данная дисциплина (учебный курс) относится к дисциплинам по выбору вариативной части блока Б1 дисциплины (модули).

Дисциплины, учебные курсы, для которых необходимы знания, умения, навыки, приобретаемые в результате изучения данной дисциплины (учебного курса) – научно-исследовательская работа (подготовка к магистерской диссертации).

3. Планируемые результаты обучения

| Формируемые и контролируемые компетенции (код и наименование) | Индикаторы достижения компетенций (код и наименование) | Планируемые результаты обучения |
|--|---|--|
| - Способен управлять разработкой продуктов, услуг и решений на основе больших данных (ПК-3) | - ПК-3.1. Знает способы управления разработкой продуктов, услуг и решений на основе больших данных | Знать: способы решения прикладных задач с помощью методов интеллектуального анализа данных |
| | - ПК-3.2. Умеет управлять разработкой продуктов, услуг и решений на основе больших данных | Уметь: применять методы интеллектуального анализа данных при решении прикладных задач |
| | - ПК-3.3. Владеет навыками управления разработкой продуктов, услуг и решений на основе больших данных | Владеть: навыками программной реализации методов интеллектуального анализа данных для решения прикладных задач |
| - Способен разрабатывать и внедрять новые методы и технологии исследования больших данных (ПК-4) | - ПК-4.1. Знает способы разработки и внедрения новых методов и технологий исследования больших данных. | Знать: способы применения методов интеллектуального анализа данных при проведении научных исследований |
| | - ПК-4.2. Умеет разрабатывать и внедрять новые методы и технологии исследования больших данных. | Уметь: планировать научные эксперименты с использованием методов интеллектуального анализа данных |
| | - ПК-4.3. Владеет навыками разработки и внедрения новых методов и технологий исследования больших данных. | Владеть: навыками оценки результатов, получаемых в результате интеллектуального анализа данных |

4. Структура и содержание дисциплины

| Модуль (раздел) | Вид учебной работы | Наименование тем занятий (учебной работы) | Семестр | Объем, ч. | Баллы | Интерактив, ч. | Формы текущего контроля (наименование оценочного средства) |
|---|--------------------|--|---------|-----------|-------|----------------|--|
| Модуль 1. Хранилища данных, OLAP-системы | Лек. | Тема 1. Основные понятия принципы работы хранилищ данных | 1 | 2 | - | | Собеседование |
| | Лек. | Тема 2. Основные понятия принципы работы OLAP-систем | 1 | 2 | - | | Собеседование |
| | Лек. | Тема 3. Классификация при интеллектуальном анализе данных | 1 | 2 | - | | Собеседование |
| | Пр. | Кластеризация при интеллектуальном анализе данных | 1 | 4 | - | | Отчет по практической работе №1 |
| | СР | Работа с лекционным материалом, подготовка задания по практической работе к защите | 1 | 20 | | | Собеседование |
| Модуль 2. Алгоритмы интеллектуального анализа данных | Лек. | Тема 4. Классификация на основе алгоритмов ID3, C4.5 | 1 | 2 | - | | Собеседование |
| | Пр. | Классификация на основе алгоритмов ID3, C4.5 | 1 | 2 | - | | Отчет по практической работе №2 |
| | Лек. | Тема 5. Классификация на основе алгоритма CART | 1 | 2 | - | | Собеседование |
| | Пр. | Классификация на основе алгоритма CART | 1 | 4 | - | | Отчет по практической работе №3 |

| Модуль (раздел) | Вид учебной работы | Наименование тем занятий (учебной работы) | Семестр | Объем, ч. | Баллы | Интерактив, ч. | Формы текущего контроля (наименование оценочного средства) |
|-----------------|--------------------|--|---------|-----------|-------|----------------|--|
| | СР | Работа с лекционным материалом, подготовка задания по практической работе к защите | 1 | 20 | | | Собеседование |
| | Лек. | Тема 6. Регрессионный анализ при интеллектуальном анализе данных | 1 | 2 | - | | Собеседование |
| | Лек. | Тема 7. Регрессионный анализ на основе нейронных сетей | 1 | 2 | - | | Собеседование |
| | Пр. | Регрессионный анализ на основе нейронных сетей | 1 | 2 | - | | Отчет по практической работе №4 |
| | СР | Работа с лекционным материалом, подготовка задания по практической работе к защите | 1 | 20 | | | Собеседование |
| | Лек. | Тема 8. Регрессионный анализ на основе алгоритма CART | 1 | 2 | - | | Собеседование |
| | Пр. | Регрессионный анализ на основе алгоритма CART | 1 | 4 | - | | Отчет по практической работе №5 |
| | СР | Работа с лекционным материалом, подготовка задания по практической работе к защите | 1 | 11,75 | | | Собеседование |
| | Лек. | Тема 9. Аффинитивный анализ на основе алгоритма Apriori | 1 | 2 | - | | Собеседование |

| Модуль (раздел) | Вид учебной работы | Наименование тем занятий (учебной работы) | Семестр | Объем, ч. | Баллы | Интерактив, ч. | Формы текущего контроля (наименование оценочного средства) |
|-----------------|--------------------|---|---------|------------|----------|----------------|--|
| | Пр. | Аффинитивный анализ на основе алгоритма Apriori | 1 | 2 | - | | Отчет по практической работе №6 |
| | ПА, зачет | Зачет | 1 | 0,25 | - | | Зачет |
| Итого: | | | | 108 | - | | |

5. Образовательные технологии

В рамках изучения дисциплины «Анализ данных» предусмотрено использование следующих образовательных технологий:

- технология традиционного обучения: лекции и практические работы;
- интерактивные технологии: проблемная лекция.

Тема 1. Основные понятия принципы работы хранилищ данных

Учебные вопросы:

1. Концепция хранилища данных
2. Организация хранилища данных
3. Очистка данных
4. Анализ данных из хранилища данных
5. Формулирование выводов

Изучив данную тему, студент должен:

иметь представление о назначении хранилищ данных и способах представления в них информации;

знать:

- понятие хранилище данных;
- понятие витрины данных;
- способы организации хранилища данных;
- способы очистки данных;

уметь:

- на основе базы данных формировать хранилище данных;
- владеть навыками:
- пользования учебной и монографической литературой;
- формирования SQL-запросов для формирования хранилища данных.

Тема 2. Основные понятия принципы работы OLAP-систем

Учебные вопросы:

1. Многомерная модель данных
2. Определение OLAP-систем
3. Концептуальное многомерное представление
4. Архитектура OLAP-систем.

Изучив данную тему, студент должен:

иметь представление о назначении, структуре и основных выполняемых задачах OLAP- систем;

знать:

- назначение OLAP систем
 - структуру OLAP систем
 - задачи, решаемые с помощью OLAP систем
- уметь:
- выделять задачи, присущие OLAP системам
 - владеть навыками:
 - постановки требований к OLAP задачам.

Тема 3. Кластеризация при интеллектуальном анализе

Учебные вопросы:

1. понятие кластеризации данных
2. общие сведения об алгоритме k-means
3. математический аппарат алгоритма k-means
4. метрики, применяемые в алгоритме k-means
5. вопрос сходимости алгоритма к локальному решению.

Изучив данную тему, студент должен:

иметь представление о назначении кластеризации данных и математическом аппарате алгоритма k-means;

знать:

- понятие кластеризации данных
- математический аппарат алгоритма k-means
- проблему сходимости алгоритма к локальному решению

уметь:

- анализировать решаемую задачу для оценки возможности ее решения с использованием алгоритма k-means

владеть навыками:

- применения алгоритма k-means для решения прикладных задач.

Тема 4. Классификация при интеллектуальном анализе данных

Учебные вопросы:

1. понятие классификации данных
2. задачи классификации
3. проблемы классификации данных
4. обзор интеллектуальных алгоритмов для классификации данных

Изучив данную тему, студент должен:

иметь представление о назначении и задачах классификации данных;

знать:

- назначение классификации данных
- задачи, решаемые с помощью классификации данных

уметь:

- выделять задачи, которые можно решить с помощью классификации данных

владеть навыками:

- анализа задач по классификации данных.

Тема 5. Классификация на основе алгоритмов ID3, C4.5

Учебные вопросы:

1. общие сведения об алгоритме ID3
2. математический аппарат алгоритма ID3
3. пример использования алгоритма ID3
4. общие сведения об алгоритме C4.5
5. математический аппарат алгоритма C4.5
6. пример использования алгоритма C4.5

Изучив данную тему, студент должен:

- иметь представление о математическом аппарате алгоритмов ID3, C4.5 и их назначении;

знать:

- назначение алгоритмов ID3 и C4.5
 - различия между алгоритмами ID3 и C4.5
- уметь:
- применять алгоритмы ID3 и C4.5 для решения практических задач владеть навыками
 - математического моделирования деревьев принятия решений с помощью алгоритмов ID3, C4.5.

Тема 6. Классификация на основе алгоритма CART

Учебные вопросы:

1. общие сведения об алгоритме CART
2. математический аппарат алгоритма CART
3. пример использования алгоритма CART

Изучив данную тему, студент должен:

иметь представление о математическом аппарате алгоритма CART и его назначении;
знать:

- назначение алгоритма CART.
- уметь:
- применять алгоритм CART для решения практических задач владеть навыками
 - математического моделирования деревьев принятия решений с помощью алгоритма CART.

Тема 7. Регрессионный анализ при интеллектуальном анализе данных

Учебные вопросы:

1. понятие регрессионного анализа данных
2. задачи регрессионного анализа
3. проблемы регрессионного анализа данных
4. обзор интеллектуальных алгоритмов для регрессионного анализа данных

Изучив данную тему, студент должен:

иметь представление о назначении и задачах регрессионного анализа данных;
знать:

- назначение регрессионного анализа данных
 - задачи, решаемые с помощью регрессионного анализа данных
- уметь:
- выделять задачи, которые можно решить с помощью регрессионного анализа данных владеть навыками
 - анализа задач по регрессионному анализу данных

Тема 8. Регрессионный анализ на основе нейронных сетей

Учебные вопросы:

1. общие сведения об нейронных сетях
2. математический аппарат нейронных сетей
3. пример использования нейронных сетей

Изучив данную тему, студент должен:

иметь представление о проведении регрессионного анализа данных с помощью нейронных сетей;

знать:

- назначение нейронных сетей

уметь:

- использовать нейронные сети для решения задач регрессионного анализа
- владеть навыками математического моделирования искусственных нейронных сетей.

Тема 9. Регрессионный анализ на основе алгоритма CART

Учебные вопросы:

1. математический аппарат алгоритма CART
2. пример использования алгоритма CART для решения задач регрессионного анализа данных

Изучив данную тему, студент должен:

3. иметь представление о возможности применения алгоритма CART для решения задач регрессионного анализа данных

знать:

- назначение алгоритма CART.

уметь:

- применять алгоритм CART для решения практических задач
- владеть навыками математического моделирования деревьев принятия решений с помощью алгоритма CART

Тема 10. Аффинитивный анализ на основе алгоритма Apriori

Учебные вопросы:

1. общие сведения об алгоритме Apriori
2. математический аппарат алгоритма Apriori
3. пример использования алгоритма Apriori

Изучив данную тему, студент должен:

иметь представление о математическом аппарате алгоритма Apriori и его назначении;
знать:

- назначение алгоритма Apriori.

уметь:

- применять алгоритм Apriori для решения практических задач
- владеть навыками математического моделирования алгоритма Apriori

6. Методические указания по освоению дисциплины

Тема 1. Основные понятия принципы работы хранилищ данных

При освоении темы необходимо:

- изучить учебный материал;
- акцентировать внимание на понятии хранилище данных;
- понять способы формирования хранилищ данных;

- понять причины, по которым анализ данных удобнее производить с использованием хранилища данных

Тема 2. Основные понятия принципы работы OLAP-систем

При освоении темы необходимо:

- изучить учебный материал темы;
- акцентировать внимание на задачах и структуре OLAP систем;
- необходимо понять 12 правил Кодда и обратить внимание о наличии дополнительных правил Кодда;
- понять различия OLAP-систем: MOLAP, ROLAP и HOLAP.

Тема 3. Кластеризация при интеллектуальном анализе

При освоении темы необходимо:

- изучить учебный материал темы;
- акцентировать внимание на понятии кластеризации данных;
- акцентировать внимание на математическом аппарате алгоритма k-means;
- понять, что получаемая с помощью k-means кластерная структура не всегда является оптимальной.

Тема 4. Классификация при интеллектуальном анализе данных

При освоении темы необходимо:

- изучить учебный материал темы;
- акцентировать внимание на задачах, решаемых с помощью классификации данных;
- акцентировать внимание на различия между классификацией и кластеризацией данных.

Тема 5. Классификация на основе алгоритмов ID3, C4.5

При освоении темы необходимо:

- изучить учебный материал темы;
- акцентировать внимание на различиях между алгоритмами ID3 и C4.5
- акцентировать внимание на проблеме построения оптимальных классификационных моделей с помощью алгоритмов ID3 и C4.5.

Тема 6. Классификация на основе алгоритма CART

При освоении темы необходимо:

- изучить учебный материал темы;
- акцентировать внимание на различиях между алгоритмами CART, ID3 и C4.5
- акцентировать внимание на критерий оптимальности разбиения множества решений в каждом из узлов классификационной модели.

Тема 7. Регрессионный анализ при интеллектуальном анализе данных

При освоении темы необходимо:

- изучить учебный материал темы;
- акцентировать внимание на задачах, решаемых с помощью регрессионного анализа данных;

- акцентировать внимание на различия между регрессионным анализом, классификацией и кластеризацией данных.

Тема 8. Регрессионный анализ на основе нейронных сетей

При освоении темы необходимо:

- изучить учебный материал темы;
- акцентировать внимание на способы обучения нейронных сетей;
- акцентировать внимание на требованиях к обучающей выборке данных, применяемой для настройки нейронной сети.

Тема 9. Регрессионный анализ на основе алгоритма CART

При освоении темы необходимо:

- изучить учебный материал темы;
- акцентировать внимание на требованиях к обучающей выборке данных, применяемой для построения дерева принятия решения с использованием алгоритма CART.

Тема 10. Аффинитивный анализ на основе алгоритма Apriori

При освоении темы необходимо:

- изучить учебный материал темы;
- акцентировать внимание на объективных и субъективных метриках, применяемых в алгоритме Apriori

7. Оценочные средства

7.1. Паспорт оценочных средств

| Семестр | Код контролируемой компетенции (или ее части) | Наименование оценочного средства |
|---------|--|---|
| 1 | ПК-3, ПК-4 | Собеседование Практические задания №1-6 Экзамен (письменно) |

7.2. Типовые задания или иные материалы, необходимые для текущего контроля

7.2.1. Практическая работа (наименование оценочного средства)

Типовой(ые) пример(ы) задания(ий)

Практическая работа № 1. Кластеризация при интеллектуальном анализе данных

В соответствии с номером варианта сгенерировать данные, на которых будет тестироваться работа алгоритма k-means.

Варианты заданий:

1 Вариант: У записей должно быть 3 параметра. Первый изменяется в диапазоне [0,1; 5], второй параметр – в диапазоне [0,1; 3], третий параметр может принимать значения – 10%, 20%, 80%, 90%.

2 Вариант: У записей должно быть 3 параметра. Первый изменяется в диапазоне [0,01; 1], второй параметр – в диапазоне [1; 300], третий параметр может принимать значения – Самара, Тольятти, Чапаевск.

3 Вариант: Первый изменяется в диапазоне [0; 1], второй параметр – в диапазоне [-2; 2], третий параметр может принимать значения – да или нет.

4 Вариант: Первый изменяется в диапазоне [-10; 1], второй параметр – в диапазоне [1; 2], третий параметр может принимать значения – отрицательное значение, положительное значение.

3. На любом известном объектно-ориентированном языке программирования разработать программу, на практике реализующую выполнение кластеризации сгенерированных данных по алгоритму k-means с применением не менее двух метрик (например, Евклидова расстояние и расстояние Манхэттена). В программе должна быть предусмотрена возможность фиксирования после каждой итерации алгоритма суммы квадратов ошибок и координат центроидов.

4. Ответить на контрольные вопросы

1. Охарактеризуйте этапы выполнения алгоритма k-means.
2. Какими данными необходимо обладать для запуска алгоритма k-means?
3. Какие метрики определения расстояний применяются в алгоритме k-means?
4. В чем заключается рекурсивность алгоритма k-means.
5. По каким параметрам производится остановка выполнения алгоритма k-means?

Практическая работа № 2. Классификация на основе алгоритмов ID3, C4.5

В соответствии с номером варианта сгенерировать данные, на которых будет тестироваться работа алгоритма ID3.

Варианты заданий:

1 Вариант: Первый атрибут может принимать значения: А, В, С, D, Е. Второй атрибут может принимать значения: да, нет. Третий атрибут может принимать значения из диапазона [0, 600]. Четвертый атрибут может принимать значения из диапазона [0, 0.9]. Пятый атрибут может принимать значения: 6, 7, 10. Выходной атрибут может принимать значения С1, С2, С3, С4, С5. Набор данных должен содержать не менее 40 записей.

2 Вариант: Первый атрибут может принимать значения: блондинка, брюнетка, шатенка, рыжая. Второй атрибут может принимать значения: да, нет. Третий атрибут может принимать значения из диапазона [160, 185]. Четвертый атрибут может принимать значения 0, 1, 2, 3. Пятый атрибут может принимать значения: 80, 90, 100. Выходной атрибут может принимать значения С1, С2, С3, С4. Набор данных должен содержать не менее 40 записей.

3 Вариант: Первый атрибут может принимать значения: стрелок1, стрелок2, стрелок3, стрелок4, стрелок5. Второй атрибут может принимать значения: нарезной, гладкий. Третий атрибут может принимать значения из диапазона [50, 95]. Четвертый атрибут может принимать значения 100, 150, 200. Пятый атрибут может принимать значения: [500, 800]. Выходной атрибут может принимать значения С1, С2, С3, С4. Набор данных должен содержать не менее 40 записей.

4 Вариант: Первый атрибут может принимать значения: среднее, средне-специальное, высшее не оконченное, высшее, два высших. Второй атрибут может принимать значения из диапазона [18, 50]. Третий атрибут может принимать значения да, нет. Четвертый атрибут может принимать значения 0, 1, 2, 3. Четвертый атрибут может принимать значения из диапазона [50, 100]. Выходной атрибут может принимать значения С1, С2, С3, С5. Набор данных должен содержать не менее 40 записей.

2. На любом известном объектно-ориентированном языке программирования разработать программу, на практике реализующую выполнение алгоритма ID3 с использованием сгенерированных данных.

3. Ответить на вопросы для контроля

1. Охарактеризуйте этапы выполнения алгоритма ID3.

2. Какими данными необходимо обладать для запуска алгоритма ID3?
3. По какому критерию определяется условие разбиения в текущем узле дерева решений построенного с помощью алгоритма ID3?
4. В чем заключается рекурсивность алгоритма ID3.
5. По каким параметрам производится остановка выполнения алгоритма ID3?
4. На любом известном объектно-ориентированном языке программирования разработать программу на практике реализующую выполнение алгоритма C4.5 с использованием сгенерированных данных.
5. Сравнить построенные деревья принятия решений по алгоритму ID3 и C4.5
6. Ответить на вопросы для контроля.
 1. Охарактеризуйте этапы выполнения алгоритма C4.5.
 2. Какими данными необходимо обладать для запуска алгоритма C4.5?
 3. По какому критерию определяется условие разбиения в текущем узле дерева решений построенного с помощью алгоритма C4.5?
 4. В чем заключается рекурсивность алгоритма C4.5.
 5. По каким параметрам производится остановка выполнения алгоритма C4.5?
 6. В чем отличия алгоритмов ID3 и C4.5?

Практическая работа № 3. Классификация на основе алгоритма CART

1. Данные, на которых будет проводиться тестирование алгоритма CART необходимо взять из предыдущей практической работы.
2. На любом известном объектно-ориентированном языке программирования разработать программу на практике реализующую выполнение алгоритма CART с использованием сгенерированных данных.
3. Сравнить построенные деревья принятия решений по алгоритму ID3, C4.5, CART
4. Ответить на вопросы для контроля.
 1. Охарактеризуйте этапы выполнения алгоритма CART.
 2. Какими данными необходимо обладать для запуска алгоритма CART?
 3. По какому критерию определяется условие разбиения в текущем узле дерева решений построенного с помощью алгоритма CART?
 4. В чем заключается рекурсивность алгоритма CART.
 5. По каким параметрам производится остановка выполнения алгоритма CART?
 6. В чем отличия алгоритмов CART и C4.5?

Практическая работа № 4. Регрессионный анализ на основе нейронных сетей

1. Данные, на которых будет проводиться тестирование нейронных сетей необходимо сгенерировать следующим образом. Выбрать функцию с двумя независимыми переменными $y = f(x_1, x_2)$. При разных значениях x_1, x_2 рассчитать значения y . Принять значения x_1, x_2 – входными сигналами нейронной сети, а y – выходным сигналом.
2. На любом известном объектно-ориентированном языке программирования разработать программу на практике реализующую выполнение простейшей нейронной сети с использованием сгенерированных данных.
3. Построить контурную диаграмму исходной функции $y = f(x_1, x_2)$ (например в среде Mathcad) . Построить аналогичную контурную диаграмму по данным генерируемых нейронной сетью. Сравнить обе контурных диаграммы.
4. Ответить на вопросы для контроля.
 1. Охарактеризуйте понятие регрессионный анализ.
 2. Опишите общий алгоритм работы с нейронными сетями.
 3. Какие существуют типы нейронных сетей?
 4. Назовите рекомендации по выбору структуры нейронных сетей в зависимости от решаемой задачи.

Практическая работа № 5. Регрессионный анализ на основе алгоритма CART

1. Данные, на которых будет проводиться тестирование алгоритма CART необходимо сгенерировать следующим образом. Выбрать функцию с двумя независимыми переменными $y = f(x_1, x_2)$. При разных значениях x_1, x_2 рассчитать значения y . Принять значения x_1 – значениями первого атрибута, значения x_2 – значениями второго атрибута, а y – метками класса.

2. На любом известном объектно-ориентированном языке программирования разработать программу на практике реализующую выполнение алгоритма CART с использованием сгенерированных данных.

3. Построить контурную диаграмму исходной функции $y = f(x_1, x_2)$ (например в среде Mathcad) . Построить аналогичную контурную диаграмму по правилам выделенным из построенного дерева решений. Сравнить обе контурных диаграммы.

4. Ответить на вопросы для контроля.

1. Охарактеризуйте понятия регрессионного анализа данных.
2. Охарактеризуйте этапы выполнения алгоритма CART.
3. Назовите причины, по которым алгоритмы ID3 и C4.5 не могут применяться для регрессионного анализа данных?
4. В чем заключается рекурсивность алгоритма CART.
5. По каким параметрам производится остановка выполнения алгоритма CART?

Практическая работа № 6. Аффинитивный анализ на основе алгоритма Apriori

1. Выбрать предметную область, в которой возможно применение аффинитивного анализа и сгенерировать не менее 100 транзакций, содержащих не менее 25 различных предметов.

2. На любом известном объектно-ориентированном языке программирования разработать программу, на практике реализующую выполнение алгоритма Apriori с использованием сгенерированных данных.

3. Ответить на вопросы для контроля.

1. Охарактеризуйте понятие аффинитивного анализа
2. Охарактеризуйте алгоритм Apriori.
3. Какие параметры может задавать аналитик для управления процессом выполнения алгоритма Apriori.
4. Какие объективные и субъективные метрики применяются для оценки ассоциативных правил?
5. Что является результатом выполнения алгоритма Apriori?

Требования к оформлению

Отчет должен содержать подробное описание (включая иллюстративный материал) последовательности действий, проделанных студентом для выполнения заданий.

Процедура оценивания

Оценка выполненной работы проводится по критериям:

1. Наличие всей существенной информации по работе
2. Точность и полнота предоставляемых сведений
3. Непротиворечивость приводимой информации
4. Правильность интерпретаций и выводов, которые сделаны по результатам работы
5. Степень достижения студентом поставленной цели
6. Обоснованность применяемого решения
7. Грамотность (содержательная) используемых формулировок

Критерии оценки за отчеты по практическим работам:

- оценка «отлично» выставляется студенту, если он выполнил работу в полном объеме

с соблюдением необходимой последовательности действий; аккуратно, четко и без ошибок выполнил отчет, вывод исчерпывающий и доказательный. При защите отчета ответил на все вопросы по теме; хорошо ориентируется в материале, умеет определить взаимосвязь факторов и их влияние на конечную цель, умеет графически отобразить важнейшие функциональные зависимости;

- оценка «хорошо» выставляется студенту, если он выполнил работу в полном объеме с соблюдением необходимой последовательности действий; без ошибок выполнил отчет, вывод исчерпывающий. При защите отчета хорошо разбирается в материале, но не уверен и неполно отвечает на вопросы. Способность к обобщению причинно-следственных связей важнейших факторов выражена недостаточно;
- оценка «удовлетворительно» выставляется студенту, если он выполнил работу не полностью, но объем выполненной части таков, что позволяет получить правильные результаты и выводы; выполнен с несущественными замечаниями. Вывод по работе не раскрывает сути работы. Владение понятийным аппаратом темы недостаточны;
- оценка «неудовлетворительно» выставляется студенту, если он выполнил работу не полностью или объем выполненной части работы не позволяет сделать правильных выводов. В ответах на вопросы есть грубые ошибки. Нет знания принципиальных теоретических положений темы.

7.3. Оценочные средства для промежуточной аттестации по итогам освоения дисциплины

7.3.1. Вопросы к промежуточной аттестации

Семестр 3

| № п/п | Вопросы к зачету |
|-------|---|
| 1 | Концепция и организация хранилища данных. |
| 2 | Многомерная модель данных. |
| 3 | Определение и архитектура OLAP-систем. |
| 4 | Задачи интеллектуального анализа данных. |
| 5 | Практическое применение интеллектуального анализа данных. |
| 6 | Модели интеллектуального анализа данных. |
| 7 | Методы Data Mining. |
| 8 | Метода построения деревьев принятия решений. |
| 9 | Алгоритмы поиска ассоциативных правил. |
| 10 | Неиерархические алгоритмы кластеризации. |
| 11 | Нечеткая логика в интеллектуальном анализе данных. |
| 12 | Генетические алгоритмы в интеллектуальном анализе данных. |
| 13 | Нейронные сети в интеллектуальном анализе данных. |
| 14 | Алгоритм k-means в интеллектуальном анализе данных. |
| 15 | Алгоритм C-means в интеллектуальном анализе данных. |
| 16 | Алгоритм Apriori в интеллектуальном анализе данных. |
| 17 | Неэффективность использования OLTP-систем анализа данных. |
| 18 | Процесс обнаружения новых знаний в данных. |
| 19 | Алгоритм ID3 в интеллектуальном анализе данных. |
| 20 | Алгоритм C4.5 в интеллектуальном анализе данных. |
| 21 | Алгоритм CART в интеллектуальном анализе данных. |
| 22 | Генетический алгоритмы в интеллектуальном анализе данных. |
| 23 | Нейронные сети в интеллектуальном анализе данных. |

| № п/п | Вопросы к зачету |
|----------|--|
| 24 | Основные понятия и назначения алгоритма k-means. |
| 25 | Математический аппарат алгоритма k-means. |
| 26 | Начальные данные необходимые для использования алгоритма k-means. |
| 27 | Метрики применяемые в алгоритме k-means. |
| 28 | Критерии остановки выполнения алгоритма k-means. |
| 29 | Вопрос о необходимости нормировки данных, ее назначение в алгоритме k-means. |
| 30 | Алгоритм ID3: критерий выбора условия для разбиения исходного множества. |
| 31 | Особенности алгоритма ID3 по сравнению с алгоритмами построения ДПР. |
| 32 | Алгоритм ID3 и проблема переобучения. |
| 33 | Математический аппарат алгоритма ID3. |
| 34 | Алгоритм C4.5: критерий выбора условия для разбиения исходного множества. |
| 35 | Особенности алгоритма C4.5 по сравнению с алгоритмами построения ДПР. |
| 36 | Алгоритм C4.5 и проблема переобучения. |
| 37 | Математический аппарат алгоритма C4.5. |
| 38 | Основные понятия искусственных нейронных сетей (ИНС) |
| 39 | Биологическая модель нейрона. |
| 40 | Преобразование сигналов искусственным нейроном |
| 41 | Цель обучения нейронной сети |
| 42 | Обучение нейронной сети методом дельта-правила |
| 43 | Обучение многослойной нейронной сети как задача многопараметрической оптимизации |
| 44 | Обучение нейронных сетей методом обратного распространения ошибки |
| 45 | Проблема выбора количества слоев и нейронов в ИНС |
| 46 | Виды активационных функций, их графики |
| 47 | Взаимосвязь между используемой в нейронах активационной функцией и ограничениями по диапазону изменения входных сигналов |
| 48 | Способы улучшения результатов обучения многослойных нейронных сетей |
| 49 | Персептрон и проблема «исключающего ИЛИ» |
| 50 | Задачи, решаемые с помощью различных типов нейронных сетей |
| 51 | Проблема переобучения искусственных нейронных сетей и способы их преодоления |
| 52 | Оценка точности работы нейронных сетей. |
| 53 | Требования, предъявляемые к обучающей выборке. |
| 54 | Самоорганизующаяся карта Кохонена: назначение, особенности. |
| 55 | Назначение и особенности рекуррентных нейронных сетей. |
| 56 | Нейронные сети Хопфилда и Хемминга, их особенности и отличия. |
| 57 | Способы кодирования сигналов в биполярный вид для нейронных сетей Хопфилда и Хемминга. |
| 58 | Влияние скорости обучения нейронной сети на точность работы нейронной сети |
| 59 | Адаптивный выбор скорости обучения нейронных сетей |
| 60 | Коррекция весовых коэффициентов в процессе обучения нейронных сетей |

Процедура оценивания

Ответы на вопросы обеспечивают возможность адекватной оценки знаний и профессиональной подготовки магистров. Важным фактором при этом является умение экзаменуемого оперировать в своем ответе ссылками на соответствующие положения учебной и научной литературы. По результатам выполнения практического задания

определяется уровень сформированности профессиональных компетенций студента по использованию современных технологий решения прикладных задач предметной области.

Критерии оценки:

- оценка «зачтено» выставляется студенту, если он сдал все практические работы и исчерпывающе и грамотно дал ответы на вопросы;
- оценка «не зачтено» выставляется студенту, если он не дал ответ на вопросы или в ответе содержались фундаментальные ошибки, а также если им не были сданы все практические работы.

8. Учебно-методическое и информационное обеспечение дисциплины

8.1. Обязательная литература

| № п/п | Авторы, составители | Заглавие (заголовок) | Тип (учебник, учебное пособие, учебно-методическое пособие, практикум, др.) | Год издания | Количество в научной библиотеке / Наименование ЭБС |
|-------|-----------------------------|---|---|-------------|---|
| 1. | Нестеров С.А. | Интеллектуальный анализ данных средствами MS SQL Server 2008 | Учебное пособие | 2016 | ЭБС «IPRbooks» http://www.iprbookshop.ru/62813.html |
| 2. | Пальмов С.В. | Интеллектуальный анализ данных | Учебное пособие | 2017 | ЭБС «IPRbooks» http://www.iprbookshop.ru/75376.html |
| 3. | Воронова Л.И., Воронов В.И. | Machine Learning: регрессионные методы интеллектуального анализа данных | Учебное пособие | 2018 | ЭБС «IPRbooks» http://www.iprbookshop.ru/81325.html |

8.2. Дополнительная литература

| № п/п | Авторы, составители | Заглавие (заголовок) | Тип (учебник, учебное пособие, учебно-методическое пособие, практикум, др.) | Год издания | Количество в научной библиотеке / Наименование ЭБС |
|-------|------------------------|---|---|-------------|---|
| 1. | Федин Ф.О., Федин Ф.Ф. | Анализ данных. Часть 1. Подготовка данных к анализу | Учебное пособие | 2012 | ЭБС «IPRbooks» http://www.iprbookshop.ru/26444.html |
| 2. | Федин Ф.О., Федин Ф.Ф. | Анализ данных. Часть 2. Инструменты | Учебное пособие | 2012 | ЭБС «IPRbooks» |

| № п/п | Авторы, составители | Заглавие (заголовок) | Тип (учебник, учебное пособие, учебно- методическое пособие, практикум, др.) | Год издания | Количество в научной библиотеке / Наименование ЭБС |
|------------------|----------------------------|---|---|--------------------|---|
| | | Data Mining | | | http://www.iprbookshop.ru/26445.html |
| 3. | Мельниченко А.С. | Математическая статистика и анализ данных | Учебное пособие | 2018 | ЭБС «IPRbooks» http://www.iprbookshop.ru/78563.html |
| 4. | Истомина А.П. | Анализ данных качественных исследований | Практикум | 2016 | ЭБС «IPRbooks» http://www.iprbookshop.ru/66014.html |
| 5. | Жуковский О.И. | Информационные технологии и анализ данных | Учебное пособие | 2014 | ЭБС «IPRbooks» http://www.iprbookshop.ru/72106.html |

8.3. Перечень профессиональных баз данных и информационных справочных систем

Springer International Publishing, Part of Springer Science+Business Media [Электронный ресурс] – Springer International Publishing AG, 2020. Режим доступа к журн.: <http://link.springer.com> . – Загл. с экрана

8.4. Перечень программного обеспечения

| № п/п | Наименование ПО | Реквизиты договора (дата, номер, срок действия) |
|-------|-------------------------|--|
| 1. | Microsoft Access | До 01.07.2020 |
| 2. | Microsoft Visual Studio | До 01.07.2020 |
| 3. | MathCAD | MCD-7514-P/MCD-7503CP от 21.07.2009 |
| 4. | PyCharm | - |

8.5. Описание материально-технической базы, необходимой для осуществления образовательного процесса по дисциплине

| № п/п | Наименование оборудованных учебных кабинетов, лабораторий, мастерских и др. объектов для проведения практических и лабораторных занятий, помещений для самостоятельной работы обучающихся (номер аудитории) | Перечень основного оборудования |
|-------|---|---|
| 1 | Компьютерный класс УЛК-401 | Компьютер (монитор 19", системный блок Pentium (R) Dual-Core E5500 2,8 GHz / 4 Gb / 500 Gb) - 12 шт, стол ученический - 7 шт., стол компьютерный -14шт., стол преподавательский -1 шт., стулья -35шт. Доска аудиторная(меловая) - 1 шт. |
| 2 | Компьютерный класс УЛК-407 | Компьютер (монитор Samsung Sync Master 943n 19", системный блок Intel (R) Core 2 Quad 2,40 GHz 1 Gb) - 12 шт., стол лабораторный -10шт., стул -25 шт., доска 3-х секционная(меловая) -1 шт., стол преподавательский-1 шт. |
| 3 | Компьютерный класс УЛК-408 | Компьютер (монитор 17", системный блок Intel (R) Celeron (R) 2,66 GHz / 1 Gb / 80 Gb) - 18шт., маршрутизатор 2801 Router-6 шт., коммутатор Catalyst-6 шт., экран/интерактивная доска Smart Board TB-1 шт., проектор Acer P1303W., стол преподавательский-1шт., стол ученический-13шт., стол компьютерный-18 шт., стул- 50 шт., доска аудиторная (маркерная)-1 шт. |