0Министерство науки и высшего образования РФ
ФГБОУ ВО «Ульяновский государственный университет»
Факультет математики, информационных и авиационных технологий

Кафедра телекоммуникационных технологий и сетей

Липатова Светлана Валерьевна

МЕТОДИЧЕСКИЕ РЕКОМЕНДАЦИИ
для семинарских (практических) занятий, лабораторного практикума
и самостоятельной работы
по дисциплине

«Системы принятия решений»

для студентов факультета математики, информационных и авиационных технологий

Ульяновск
2022

Методические рекомендации для семинарских (практических) занятий, лабораторного практикума и самостоятельной работы по дисциплине «Системы принятия решений» / составитель: С.В. Липатова - Ульяновск: УлГУ, 2022 – 75 с.

Настоящие методические рекомендации предназначены для студентов факультета математики, информационных и авиационных технологий. В работе приведены литература по дисциплине, темы дисциплины и вопросы в рамках каждой темы, рекомендации по изучению теоретического материала, контрольные вопросы для самоконтроля, задания для самостоятельной работы, задачи и упражнения для самостоятельной подготовки к семинарам или полностью самостоятельного освоения практических навыков, задания для лабораторного практикума и рекомендации по их выполнению.

Студентам всех форм обучения следует использовать данные методические рекомендации при подготовке к семинарам, самостоятельной подготовке, а также промежуточной аттестации по дисциплине «Системы принятия решений».

Рекомендованы к введению в образовательный процесс

Учёным советом факультета математики, информационных и авиационных технологий УлГУ

протокол № 3/**22** от «19» апреля 2022 г.

СОДЕРЖАНИЕ

ОБЩИЕ ВОПРОСЫ

В результате изучения дисциплины «Системы принятия решений» студенты должны:

1) знать:

– основные идеи и алгоритмы оптимизации;
– теоретические основы математического и компьютерного моделирования
– основные понятия теории моделирования, основные требования, предъявляемые к разработке математических моделей;
– различные классы моделей,
– уметь применять их для решения практических задач, иметь навыки работы в средах моделирования.

2) уметь:

– планировать проведение экспериментов и обрабатывать их результаты;
– обосновывать выбор методов для поддержки принятия решений в конкретных ситуациях;
– разрабатывать наборы критериев для задач принятия решений;
– применять методы поддержки принятия решений;
– разрабатывать системы поддержки принятия решений; владеть:

3) владеть:

– терминологией, применяемой в теории принятия решений;
– методами поддержки принятия решений,
– информационными средствами поддержки принятия решений,
– навыками практической работы по решению оптимизационных задач;
– навыками применения алгоритмов и методов оптимизации, основных классов моделей и методов моделирования, принципов построения моделей информационных процессов, методов формализации, алгоритмизации и реализации моделей с помощью современных компьютерных средств; использования инструментальных средств моделирования систем.

Методические рекомендации для семинарских (практических) занятий, лабораторного практикума и самостоятельной работы по дисциплине «Системы принятия решений» направлены на повышение эффективности освоения знаний, умений, навыков и компетенций, связанных с:

– оптимизационными задачами,

– информационными системами и технологиями поддержки принятия решений,
– машинном обучении и т.д..

Методические рекомендации предлагают указания по всем темам дисциплины «Системы принятия решений». Методические рекомендации разбиты по темам и содержат набор вопросов для систематизации теоретического материала, полученного на лекционных занятиях, и самостоятельного изучения теории, вопросы (тесты) для текущего контроля на практических занятиях (семинарах), задачи для усвоения практических навыков. Для лабораторного практикума приведены задания, варианты и рекомендации по выполнению лабораторных работ.

Список литературы и информационного обеспечения, приведённый в конце методических указаний, может служить основой для изучения всех рассматриваемых тем. Дополнительная и учебно-методическая литература могут быть использованы обучающимися для закрепления изучаемого материала.

РЕКОМЕНДАЦИИ ПО ОТДЕЛЬНЫМ ТЕМАМ ДИСЦИПЛИНЫ

Тема 1. Процесс принятия решений.

Основные вопросы темы

1. Модель задачи принятия решений, методы и их классификация.
2. Основные этапы процесса принятия решений. Условия принятия решений. Методы описания процессов.

Рекомендации по изучению темы

Вопрос 1 изложен в учебнике [6] на с. 5-18 и [1] разделах 1,3,4.

Вопрос 2 изложен в учебнике [6] на с. 18-23.

Вопросы для самоподготовки

Рекомендуется после изучения материалов лекций и специальной литературы подготовить ответы на вопросы:

1. Что необходимо учитывать при принятии решений?
2. Какие этапы включает в себя процесс принятия решений?
3. Какие методы принятия решений существуют?
4. Какие условия принятия решения выделяют?
5. Какие существуют методы описания процессов и какие из них можно использовать для описания процесса принятия решений?

Контрольные тесты

1. Нормативный подход

Выберите один ответ:

a. нет правильного ответа.

b. предписывает, как должен поступать человек с нормальным интеллектом, желающий напряженно и систематизированно обдумывать все аспекты своей задачи;

c. описывает процесс выбора решений человеком в целях определения рационального зерна, характерного для всякого разумного выбора;

d. описывает методы получения решения сверхрациональным человеком;

e. все ответы выше верны;

2. На чем основывается принятие решения:

Выберите один ответ:

- a. все ответы выше верны;
- b. суждение;
- c. нет правильного ответа
- d. рациональность;
- e. интуиция;
- f. профессионализм;

3. Что является неопределенностью при принятии решения:

Выберите один ответ:

- a. нет правильного ответа.
- b. непонимание физической природы процессов, имеющих отношение к задаче;
- c. орфографические ошибки в исходных данных;
- d. не полное знание цели;
- e. все ответы выше верны;

4. Теория строгих решений включает:

Выберите один ответ:

- a. дедуктивные решения;
- b. индуктивные решения;
- c. все ответы выше верны;
- d. нет правильного ответа.

5. Если существует группа лиц, имеющих противоположные ЛПР цели, то это задача принятия решений в условиях:

Выберите один ответ:

- a. неопределенности;
- b. определенности;
- c. нет правильного ответа.
- d. все ответы выше верны;
- e. риска;

6. Задача принятия решений заключается в:

Выберите один ответ:

- a. классификации
- b. нет правильного ответа.
- c. выборе лучшей;
- d. ранжирование альтернатив;
- e. все ответы выше верны;

7. Если существует статистическая база данных исходов ситуации, то это задача принятия решений в условиях:

Выберите один ответ:

- a. конфликта;
- b. риска;
- c. неопределенности;
- d. определенности;
- e. все ответы верны.

8. Что является этапом принятия решения?

Выберите один ответ:

- a. все ответы выше верны;
- b. выдвижение альтернатив;
- c. нет правильного ответа.
- d. реализация решения;

9. Решение – это:

Выберите один ответ:

- a. все ответы выше верны;
- b. нет правильного ответа.
- c. выбор наилучшей альтернативы;
- d. определение критериев выбора
- e. ранжирование альтернатив по одному критерию;

10. ЛПР покупает телевизор выбирая его в магазине (альтернативы). Критерий «внешний вид» является:

Выберите один ответ:

a. Качественным

b. Неопределенным

c. Количественным

11. Стратегические цели являются:

Выберите один ответ:

a. долгосрочными;

b. краткосрочными;

c. среднесрочными;

d. все ответы выше верны;

e. нет правильного ответа.

12. Какие решения являются только вероятными:

Выберите один ответ:

a. индуктивные;

b. все перечисленные выше;

c. абдуктивные;

d. нет правильного ответа.

e. дедуктивные;

13. Лицо, принимающее решение несет ответственность за:

Выберите один ответ:

a. «непродуманные» решения;

b. за все принимаемые им решения

c. «моральные» решения;

d. решения, принятые в условиях неопределенности и риска;

e. нет правильного ответа.

14. За кем остается последнее слово при принятии решений?

Выберите один ответ:

a. описывает методы получения решения сверхрациональным человеком;

b. За ЛПР

c. За инициативной группой

d. За владельцем проблемы

e. нет правильного ответа.

Тема 2. Системы поддержки принятия решений.

Основные вопросы темы

1. Схема формальной системы поддержки принятия решений. Структура, подсистемы, функции.
2. Основные виды архитектур и примеры систем поддержки принятия решений

Рекомендации по изучению темы

Вопрос 1 изложен в учебнике [6] на с. 50-54.

Вопрос 2 изложен в учебнике [6] на с. 55-58.

Дополнительные материалы по теме изложены в [2] в главе 2 и 6.

Вопросы для самоподготовки

Рекомендуется после изучения материалов лекций и специальной литературы подготовить ответы на вопросы:

1. Какие подсистемы входят в СППР?
2. Какие существуют архитектуры построения СППР?
3. Какие классы СППР выделяют?
4. Какие методы используют при построении СППР?
5. Какие средства разработки СППР существуют?

Контрольные тесты

1. **Какие функциональные подсистемы может содержать СППР?**

Выберите один или несколько ответов:

a. все ответы выше верны;

b. поисковая машина;

c. подсистема генерации альтернатив;

d. база данных;

2. **Оптимизационная СППР используют**

Выберите один ответ:

- a. методы математического программирования;
- b. методы извлечения знаний;
- c. нет правильного ответа.
- d. все ответы верны;
- e. эвристические методы;

3. Какие утверждения верны для СППР?

Выберите один ответ:

- a. может быть адаптирована для группового и индивидуального использования;
- b. оперирует со слабоструктурированными решениями;
- c. все ответы выше верны;
- d. реализует разделение данных и моделей;
- e. нет правильного ответа.

Тема 3. Хранилища данных.

Основные вопросы темы

1. Определение и свойства хранилищ данных, виды данных, хранящихся в хранилищах. Многомерная модель представления данных. OLAP. Виды реализации многомерной модели данных. СУБД, обеспечивающие поддержку OLAP.
2. Технологии ETL.

Рекомендации по изучению темы

Вопрос 1 изложен в учебнике [6] на с. 58-73, [4] в главе 2.

Вопрос 2 изложен в учебнике [6] на с. 74-79, [4] в параграфах 2.9-2.11.

Вопросы для самоподготовки

Рекомендуется после изучения материалов лекций и специальной литературы подготовить ответы на вопросы:

1. На основе какой технологии строятся оперативные базы данных?
2. На основе какой технологии строится хранилище данных?
3. Для чего используют технологию ETL?

4. В чем отличия OLTP и OLAP?
5. Чем отличаются ROLAP, MOLAP и HOALP?
6. Как реализуют многомерное представление модели звезда и снежинка?

Контрольные тесты

1. Витрина данных - это:

Выберите один ответ:

- a. все ответы выше верны;
- b. реляционная база данных по выбранной тематике
- c. «срез» хранилища данных по выбранному измерению;
- d. ER-модель хранилища данных;
- e. нет правильного ответа.

2. Какие задачи решает технология ETL?

Выберите один ответ:

- a. нет правильного ответа.
- b. нахождение скрытых закономерностей в данных;
- c. извлечение, преобразование и загрузка;
- d. диагностика и анализ;
- e. все ответы выше верны;

3. Технология OLAP использует модель данных:

Выберите один ответ:

- a. реляционную;
- b. нет правильного ответа.
- c. иерархическую;
- d. все ответы верны;
- e. многомерную;

4. Подход по Кимбаллу предполагает

Выберите один ответ:

- a. все ответы выше верны;
- b. объединение всех витрин данных предприятия;

- c. нет правильного ответа.
- d. построение хранилища данных «с нуля»;
- e. построение виртуального хранилища данных на основе оперативных баз данных;

5. Какая из технологий опирается на понятие транзакции?

Выберите один ответ:

- a. Data Mining;
- b. нет правильного ответа.
- c. OLTP;
- d. все ответы выше верны;
- e. OLAP;

6. Хранилище данных:

Выберите один ответ:

- a. все ответы выше возможны для разных хранилищ;
- b. не учитывает хронологию данных;
- c. учитывает хронологию данных и хранит все версии;
- d. учитывает хронологию данных и стирает старые версии данных

7. На каком этапе принятия решений можно использовать технологию OLTP?

Выберите один ответ:

- a. анализ последствий;
- b. все ответы выше верны;
- c. нет правильного ответа.
- d. выдвижение или генерация альтернатив;
- e. диагностика и анализ ситуации

Тема 4. Средства СУБД для аналитической обработки данных.

Основные вопросы темы

1. На примере PostgreSQL рассматриваются средства системы для аналитической обработки данных: понятия окна (over), секционирование (partitioning), упорядочивание (order by), кадрирование (с использованием rows и range),

аналитических функций сведения, функций нумерации, получения значения строк, статистические.

Рекомендации по изучению темы

Справочные материалы изложены в разделе «Лабораторный практикум».

Вопросы для самоподготовки

Рекомендуется после изучения материалов лекций и специальной литературы подготовить ответы на вопросы:

1. Каково применение окна для получения аналитических данных?
2. В чем заключается секционирование данных?
3. Каково расширение оператора упорядочивание?
4. Приведите примеры использования функций сведения.
5. Как можно использовать статистические функции и что они позволяют делать?

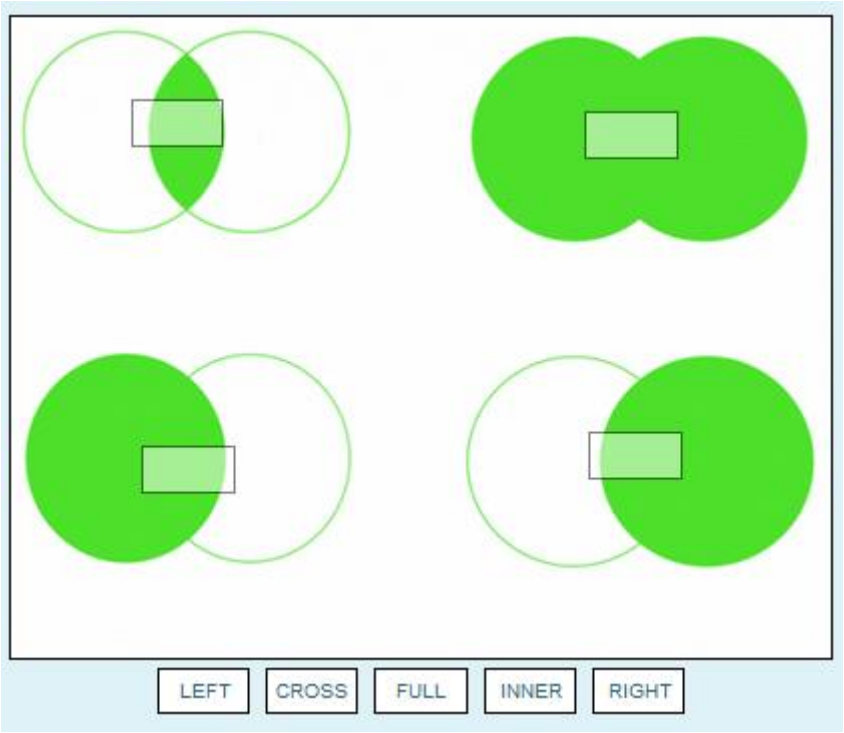
Контрольные тесты

1. **Общие табличные выражения являются отдельно применяемым оператором и сохраняют данные во временную таблицу**

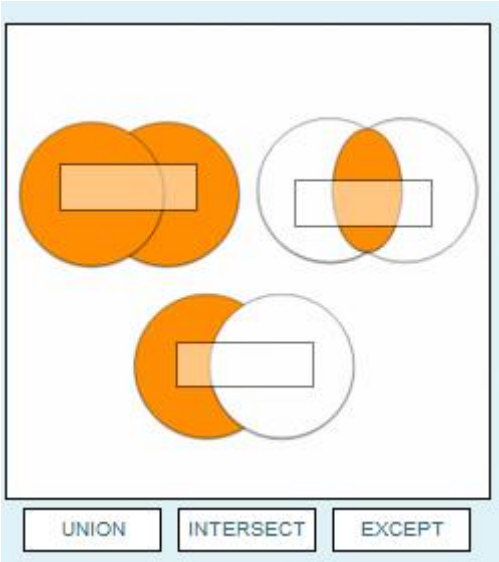
Выберите один ответ:

- Верно
- Неверно

2. **Сопоставьте соответствующий вид соединения таблиц (JOIN) на соответствующее отображение операции над множествами.**



3. **Сопоставьте операторы объединения результатов запроса схемам**



4. **Что выполняют оконные функции с их значениями (применяемые с OVER)**

- last_value,
- row_number
- lead
- first_value
- Lag

Тема 5. Методы работы с экспертами.

Основные вопросы темы

1. Задачи экспертов в процессе принятия решений.
2. Классификация методов работы с экспертами. Методы оценивания экспертов.

Рекомендации по изучению темы

Вопрос 1 изложен в учебнике [6] на с. 26-29.

Вопрос 2 изложен в учебнике [6] на с. 30-32.

Дополнительные материалы по теме изложены в [3] в разделе 4.

Вопросы для самоподготовки

Рекомендуется после изучения материалов лекций и специальной литературы подготовить ответы на вопросы:

1. В чем заключается метод «мозговой штурм»?
2. В чем заключается метод «круглый стол»?
3. В чем заключается метод «Дельфи»?
4. В чем заключается метод анализа иерархий?

Контрольные тесты

1. **Какой из методов генерирования альтернативных вариантов решений основан на использовании опыта решения предшествующих аналогичных проблем?**

Выберите один ответ:

- a. все ответы выше верны;
- b. нет правильного ответа.
- c. метод ассоциаций и аналогов;
- d.

метод «мозговой штурм»;

- e. тестирование;

2. **Метаданные бывают:**

Выберите один или несколько ответов:

- a. преобразования данных;
- b. все ответы выше верны;

- [ ] c. исходной системы;
- [ ] d. нет правильного ответа.
- [ ] e. извлечения скрытых закономерностей;

3. Для какой шкалы не требуется устанавливать степень обладания определенным свойством?

Выберите один ответ:

- a. номинальная;
- b. интервальная;
- c. нет правильного ответа.
- d. порядковая;
- e. ранговая;
- f. все ответы выше верные;

4. Какие методы являются групповыми?

Выберите один ответ:

- a. нет правильного ответа.
- b. метод рационального решения проблемы;
- c. интервью;
- d. метод комиссии;
- e. все ответы выше верны;

Тема 6. Методы выбора решений (рациональные).

Основные вопросы темы

1. Задача оптимизации. Классификация методов оптимизации.
2. Математическое программирование.
3. Методы минимизации функции одной переменной (попарного сравнения, дихотомии, золотого сечения), методы многомерной оптимизации (нулевого порядка: метод Хука-Дживса, метод Нелдера-Мида; первого: градиентного спуска с постоянным шагом, наискорейшего спуска; второго: Ньютона), линейное программирование и т.д.

Рекомендации по изучению темы

Вопрос 1 изложен в учебнике [6] на с. 32-36.

Вопрос 1 изложен в учебнике [6] раздел 2.

Вопрос 3 изложен в учебнике [7] и [8].

Дополнительные материалы по теме изложены в [3] в главе 3.

Вопросы для самоподготовки

Рекомендуется после изучения материалов лекций и специальной литературы подготовить ответы на вопросы:

1. Описание задачи оптимизации?
2. Классификация задач оптимизации?
3. Классификация методов решения задач оптимизации?
4. Математическое программирование?
5. Линейное программирование?
6. Динамическое программирование?
7. В чем заключается метод золотого сечения?
8. В чем заключается метод дихотомии?
9. В чем заключается метод попарного деления?
10. В чем заключается метод Хука-Дживса?
11. В чем заключается метод Нелера-Мида?
12. В чем заключается метод градиентного спуска с постоянным путем?

Контрольные тесты

1. **Алгоритм последовательного улучшения плана, позволяющий осуществлять переход от одного допустимого базисного решения к другому таким образом, что значение целевой функции непрерывно возрастают и за конечное число шагов находится оптимальное решение называется**

Выберите один ответ:

- a. Алгоритм двойственного симплекс-метода
- b. Алгоритм метода ветвей и границ
- c. Алгоритм симплекс-метода
- d. Алгоритм метода Гомори

2. **Вершина выпуклого многогранника это**

Выберите один ответ:

a. любая точка выпуклого многогранника, которая является концомотрезка целиком принадлежащего этому многограннику

b. любая точка выпуклого многогранника, которая является внутренней отрезка целиком принадлежащего этому многограннику

c. любая точка выпуклого многогранника, которая не является внутренней никакого отрезка целиком принадлежащего этому многограннику

d. любая точка выпуклого многогранника, которая является серединойотрезка целиком принадлежащего этому многограннику

3. **В каком направлении сдвигают линию уровня целевой функции при решении задачи линейного программирования на максимум?**

Выберите один ответ:

a. в направлении антиградиента

b. вверх

c. в направлении градиента

4. **В каком случае точка на отрезке между оптимальными планами задачи линейного программирования тоже будет оптимальным планом (задача не целочисленная)?**

Выберите один ответ:

a. если задача на максимум

b. всегда

c. никогда

d. если задача на минимум

5. **Выберите оптимальную альтернативу по следующим данным:**

Альтернатива\Критерий	К1	К2	К3
А1 .	8	4	7
А2 .	. 2	4	5
А3 .	. 4	7	3
Вес .	3	4	5

Выберите один ответ:

a. A3

b. A2

c. A1

6. **Графический способ решения задачи линейного программирования – это построение прямых, уравнения которых получаются в результате замены в ограничениях знаков неравенств на знаки точных равенств и включает:**

Выберите один ответ:

a. все перечисленные ответы

b. нахождение полуплоскости, определяемой каждым из ограничений задачи

c. построение прямой F = h = const >= 0, проходящей через многоугольник решений

d. нахождение многоугольника допустимых решений

e. построение вектора C, перпендикулярного прямой F = h = const

f. определение координат точки максимума функции и вычисление значения целевой функции в этой точке

g. передвижение прямой F = h = const в направлении вектора C (в сторону увеличения h), в результате чего находят либо точку (точки), в которой целевая функция принимает максимальное значение, либо устанавливают неограниченность сверху функции на множестве допустимых решений

7. **Если в прямой задаче, какое либо ограничение является неравенством, то в двойственной задаче соответствующая переменная**

Выберите один ответ:

a. неотрицательна

b. свободна от ограничений

c. положительна

d. отрицательная

8. **Если в транспортной задаче объем запасов превышает объем потребностей, в рассмотрение вводят**

Выберите один ответ:

a. фиктивный пункт производства

b. изменения структуры не требуются

c. фиктивный пункт потребления

9. Если в транспортной задаче объем спроса равен объему предложения, то такая задача называется

Выберите один ответ:

a. открытой

b. сбалансированной

c. замкнутой

d. незамкнутой

e. закрытой

10. Если задача линейного программирования имеет оптимальное решение, то целевая функция достигает нужного экстремального значения в одной из

Выберите один ответ:

a. внутренних точек многоугольника (многогранника) допустимых решений

b. вершин многоугольника (многогранника) допустимых решений

c. вне точек многоугольника (многогранника) допустимых решений

11. Если при попытке решить задачу линейного программирования симплекс-методом не обнаружено необходимого числа базисных переменных, ...

Выберите один ответ:

a. для решения задачи симплексметодом необходимо ввести искусственный базис

b. задача неразрешима

c. задачу можно решить только графически

12. Если целевая функция имеет линейный вид, а ограничения нелинейный, к какому классу следует отнести задачу?

Выберите один ответ:

a. нет правильного ответа

b. безусловная оптимизация

c. линейного программирования

d. нелинейного программирования

Задания для практических занятий и самостоятельной работы

Задание 1:

1) Привести задачу ко всем формам ЗЛП
2) Найти двойственную задачу
3) Решить графическим методом задачу линейного программирования.

Варианты задания

Вариант 1

Max f (x) = $3X_1 + 2X_2$

$X_1 + 2X_2 \leq 11$

$2X_1 - X_2 \geq 5$

$X_1 + 3X_2 \geq 14$

$X_1 , X_2 \geq 0$

Вариант 2

Max f (x) = 3X1 + 2X2

X1 + 2X2 ≤ 12

2X1 - X2 ≥ 7

X1 + 3X2 ≥ 14

X1 , X2 ≥ 0

Вариант 3

Max f (x) = 3X1 + 2X2

X1 + 2X2 ≥ 10

2X1 - X2 ≤ 18

X1 + 3X2 ≤ 13

X1 , X2 ≥ 0

Вариант 4

Min f (x) = 3X1 + 2X2

X1 + 2X2 ≥10

2X1 - X2 ≥ 10

X1 + 3X2 ≤ 13

X1 , X2 ≥ 0

Вариант 5

Max f (x) = 4x1+ 3x2

x1 + 2x2 ≤ 10

x1 + 2x2 ≥ 2

2x1 + x2 ≤ 10

x1 ≥ 0, x2 ≥ 0

Вариант 6

Min f (x) = 3X1 + 2X2

X1 + 2X2 ≥12

2X1 - X2 ≥ 12

X1 + 3X2 ≤ 14

X1 , X2 ≥ 0

Вариант 7

Max f (x) = 3x1+ 5x2

x1 + x2 ≤ 5

3x1 + 2 x2 ≤ 8

x1 ≥ 0, x2 ≥ 0

Вариант 8

Min f (x) = 3X1 + 2X2

X1 + 2X2 ≤ 11

2X1 - X2 ≥ 5

X1 + 3X2 ≥ 14

X1 , X2 ≥ 0

Вариант 9

Max f (x) = 3x1+ x2

2x1 + 3x2 ≥ 12

-x1 + x2 ≤ 2

2x1 - x2 ≤ 2

x1 ≥ 0, x2 ≥ 0

Вариант 10

Max f (x) = 3x1+ x2

x1 + x2 ≤ 5

$0.5x1 + x2 \geq 3$ $x1 - x2 \geq 1$

Задание 2:

1) решить задачу, согласно варианту (записать формулировку ЗЛП, получить решение).
2) Привести одну итерацию цикла решения задачи симплекс-методом

Варианты задания

1)Для изготовления четырех видов продукции используют три вида сырья. Запасы сырья, нормы его расхода и прибыль от реализации каждого продукта приведены в таблице.

Тип сырья	Нормы расхода сырья на одно изделие				Запасы сырья
	А	Б	В	Г	
I	1	2	1	0	18
II	1	1	2	1	30
III	1	3	3	2	40
Цена изделия	12	7	18	10	

1. Определить, как изменятся общая стоимость продукции и план ее выпуска при увеличении запасов сырья I и II вида на 4 и 3 единицы соответственно и уменьшении на 3 единицы сырья III вида.
2. Определить целесообразность включения в план изделия «Д» ценой 10 ед., на изготовление, которого расходуется по две единицы каждого вида сырья ед.

2)Для изготовления четырех видов продукции используют три вида сырья. Запасы сырья, нормы его расхода и прибыль от реализации каждого продукта приведены в таблице.

Тип сырья	Нормы расхода сырья на одно изделие				Запасы сырья
	А	Б	В	Г	
I	1	0	2	1	180
II	0	1	3	2	210
III	4	2	0	4	800
Цена изделия	9	6	4	7	

1. Определить, как изменятся общая стоимость продукции и план выпуска при увеличении запасов сырья II и III вида на 120 и 160 ед. соответственно и одновременном уменьшении на 60 ед. запасов сырья I вида;
2. Определить целесообразность включения в план изделия «Д» ценой 12 ед., на изготовление которого расходуется по две единицы каждого вида сырья.

3) Для изготовления трех видов продукции используют три вида сырья. Запасы сырья, нормы его расхода и прибыль от реализации каждого продукта приведены в таблице.

Тип Сырья	Нормы расхода сырья на одно изделие			Запасы сырья
	А	Б	В	
I	4	2	1	180
II	3	1	3	210
III	1	2	5	244
Цена	10	14	12	

1. Определить, как изменится общая прибыль продукции и план выпуска при увеличении запасов сырья I и III вида на 4 ед. каждого;
2. Определить целесообразность включения в план изделия «Г», на изготовление которого расходуется соответственно 1, 3 и 2 ед. каждого вида сырья ценой 13 ед. и изделия «Д» на изготовление которого расходуется по две единицы каждого вида сырья ценой 12 ед.

4) Для изготовления четырех видов продукции используют три вида сырья. Запасы сырья, нормы его расхода и прибыль от реализации каждого продукта приведены в таблице.

Тип Сырья	Нормы расхода сырья на одно изделие				Запасы сырья
	А	Б	В	Г	
I	2	1	3	2	200
II	1	2	4	8	160
III	2	4	1	1	170
Цена изделия	5	7	3	8	

1. Определить, как изменится общая стоимость продукции и план выпуска при увеличении запасов сырья I и II вида на 8 и 10 ед. соответственно и одновременном уменьшении на 5 ед. запасов сырья III вида;
2. Определить целесообразность включения в план изделия «Д» на изготовление которого расходуется по две единицы каждого вида сырья и ожидается прибыль 10 ед.

5) На основании информации приведенной в таблице была решена задача оптимального использования ресурсов на максимум общей стоимости.

Ресурсы	Нормы затрат ресурсов на единицу продукции			Запасы
	I вид	**II** вид	**III** вид	
Труд	1	4	3	200
Сырье	1	1	2	80
Оборудование	1	1	2	140

Цена	40	60	80	

1. Определить, как изменится общая стоимость продукции и план выпуска при увеличении запасов сырья на 18 единиц;.
2. Определить целесообразность включения в план изделия четвертого вида на изготовление которого расходуется по две единицы каждого вида ресурсов ценой 70 ед.

6) На предприятии выпускается три вида изделий, используется при этом три вида сырья:

Сырье	Нормы затрат ресурсов на единицу продукции			Запасы
	А	**Б**	**В**	сырья
I	18	15	12	360
II	6	4	8	192
III	5	3	3	180
Цена	9	10	16	

1. Как изменится общая стоимость выпускаемой продукции и план ее выпуска, если запас сырья I вида увеличить на 45 кг., а II - уменьшить на 9кг.?
2. Целесообразно ли выпускать изделие Г ценой 11 единиц, если нормы затрат сырья 9, 4 и 6 кг.?

7) Для изготовления трех видов продукции используют четыре вида ресурсов. Запасы ресурсов, нормы расхода и цена каждого продукта приведены в таблице.

Ресурсы	Нормы затрат ресурсов на единицу продукции			Запасы
	I вид	**II** вид	**III** вид	
Труд	3	6	4	2000
Сырье 1	20	15	20	15000
Сырье 2	10	15	20	7400
Оборудование	0	3	5	1500
Цена	6	10	9	

1. Как изменится общая стоимость выпускаемой продукции и план ее выпуска, если запас сырья I вида увеличить на 24?
2. Целесообразно ли выпускать изделие четвертого вида ценой 11 единиц, если нормы затрат ресурсов 8, 4, 20 и 6 единиц.?

8) Предприятие выпускает 4 вида продукции и использует 3 типа основного оборудования: токарное, фрезерное, шлифовальное. Затраты на изготовление единицы продукции приведены в таблице; там же указан общий фонд рабочего времени, а также цена изделия каждого вида.

Тип	Нормы расхода сырья на одно изделие	Общий фонд

оборудования	А	Б	В	Г	раб. времени
Токарное	2	1	1	3	300
Фрезерное	1	0	2	1	70
Шлифовальное	1	2	1	0	340
Цена изделия	8	3	2	1	

1. Как изменится общая стоимость выпускаемой продукции и план ее выпуска, если фонд времени шлифовального оборудования увеличить на 24 часа ?
2. Целесообразно ли выпускать изделие «Д» ценой 11 единиц, если нормы затрат оборудования 8, 2 и 2 ед.?

9) На предприятии выпускается три вида изделий, используется при этом три вида сырья:

Сырье	Нормы затрат ресурсов на единицу продукции			Запасы
	А	**Б**	**В**	сырья
I	1	2	1	430 кг
II	3	0	2	460 кг
III	1	4	0	420 кг
Цена	3	2	5	

1. Как изменится общая стоимость выпускаемой продукции и план ее выпуска, если запас сырья I вида увеличить на 80 кг., а II - уменьшить на 10кг.?
2. Целесообразно ли выпускать изделие Г ценой 7 единиц, если нормы затрат сырья 2, 4 и 3 кг.?

10) Для изготовления четырех видов продукции используют три вида сырья. Запасы сырья, нормы его расхода и прибыль от реализации каждого продукта приведены в таблице.

Тип	Нормы расхода сырья на одно изделие				Запасы
Сырья	А	Б	В	Г	сырья
I	2	1	0,5	4	2400
II	1	5	3	0	1200
III	3	0	6	1	3000
Цена изделия	7,5	3	6	12	

1. Как изменится общая стоимость выпускаемой продукции и план ее выпуска, если запас сырья I вида увеличить на 100 кг, а II - уменьшить на 150кг.?
2. Целесообразно ли выпускать изделие «Д» ценой 10 единиц, если нормы затрат сырья 2, 4 и 3 кг?

Задание 3:

1) Привести поиск допустимого решения с помощью метода с северо-западным углом
2) Привести поиск допустимого решения с помощью метода наименьшей стоимости
3) Найти оптимальное решение транспортной задачи

В каждом из заданий приведена таблица, в клетках которой представлены элементы матрицы **C = {C_{ij}},** справа от таблицы – значения величин **a_i (i = 1,4),** внизу – значения величин **b_j (j = 1,5)** транспортной задачи.

A –поставщики, B – потребители

Решить соответствующую транспортную задачу.

1.

14	5	27	29	23	18
17	7	16	19	2	14
20	12	15	29	5	16
14	24	18	7	14	22
8	11	11	9	21	

2.

6	17	26	14	8	24
18	14	27	6	20	8
8	24	11	17	26	12
4	18	21	16	12	14
11	11	11	12	16	

3.

17	10	7	5	13	34
12	28	25	9	10	18
14	15	18	9	28	6
25	16	21	12	8	10
10	10	10	10	30	

4.

19	9	14	17	9	17
4	21	2	8	29	17
22	30	4	1	24	14
16	22	8	5	27	10
9	9	9	9	24	

5.

25	16	26	43	23	38
30	23	28	48	27	12
37	23	25	49	28	8
22	1	4	25	10	20
13	13	13	13	28	

6.

12	21	19	29	4	23
27	13	22	19	4	23
20	27	18	2	23	20
30	12	3	20	24	23
22	22	22	22	4	

7.

10	15	14	28	1	14
16	7	30	8	29	14
1	21	22	19	12	10
8	25	28	5	19	16
11	11	11	8	15	

8.

17	16	15	29	9	25
6	27	20	25	20	25
6	15	12	8	14	10
10	24	23	5	22	15
16	16	16	16	16	

9.

24	19	5	10	23	30
15	16	3	13	6	30
7	5	24	11	23	33
4	28	29	21	20	33
25	25	25	25	30	

10.

24	23	6	29	3	30
20	8	13	2	27	35
30	17	10	23	28	20
4	7	23	27	26	10
20	20	15	15	30	

11.

3	25	11	22	12	20
9	15	4	26	12	25
13	22	15	12	27	10
6	19	8	11	8	30
18	18	18	18	18	

12.

23	2	1	10	3	35
20	19	47	16	14	10
7	3	121	21	10	10
9	9	29	8	18	25
17	17	17	17	17	

15.

33	22	14	34	19	15
26	16	7	29	16	17
28	18	17	23	30	20
35	25	11	22	9	16
14	14	14	18	10	

16.

25	18	14	3	16	24
29	15	27	16	17	6
21	2	29	2	22	15
5	13	1	5	17	13
11	16	11	11	11	

17.

8	28	17	19	11	20
27	5	10	6	19	24
29	11	3	7	8	21
25	16	19	24	13	15
19	16	16	16	16	

18.

27	6	8	12	23	28
1	25	19	11	12	15
28	19	15	17	29	15
16	22	18	5	13	14
14	15	15	15	15	

19.

13	7	19	18	27	15
1	21	8	20	12	18
50	17	14	23	21	15
7	14	29	18	22	10
12	18	10	10	10	

20.

39	28	37	27	46	30
21	4	20	3	14	15
25	27	25	24	29	15
12	26	10	5	22	15
13	13	13	21	20	

21.

16	26	12	24	3	12
5	2	19	27	2	14
29	23	25	16	8	14
22	25	14	15	21	14
13	5	13	12	13	

22.

29	4	8	11	5	15
10	19	26	1	27	15
16	7	4	29	23	15
9	10	24	25	17	15
11	12	13	14	12	

23.

14	27	6	16	8	20
2	4	19	4	27	22
26	23	1	20	3	20
24	5	12	30	5	20
18	20	19	19	9	

24.

23	2	1	4	12	30
24	17	27	3	5	30
26	2	19	22	11	35
7	1	2	14	9	35
22	43	20	17	35	

25.

14	27	5	18	19	24
17	20	1	24	3	20
11	7	28	23	9	20
8	26	19	2	24	24
19	25	20	13	13	

26.

14	6	1	12	19	30
28	13	22	18	4	15
21	27	30	10	14	20
2	5	6	25	7	15
15	15	18	17	16	

27.

6	30	25	7	15	35
5	29	21	4	13	40
18	22	5	28	1	25
19	23	8	2	14	15
24	25	30	20	21	

28.

22	23	16	12	14	18
17	30	1	8	25	18
27	15	13	23	22	18
3	12	21	26	7	18
17	17	17	17	18	

29.

9	21	22	14	10	18
30	34	42	23	26	12
8	17	30	27	9	20
11	20	24	7	25	18
14	11	17	15	14	

30.

12	15	9	19	22	40
20	15	11	2	19	30
21	26	23	7	16	25
11	24	8	3	29	15
34	39	24	8	8	

Задание 4:

Решить задачу о назначениях венгерским методом.

Вариант 1			
9	3	2	4
3	1	9	4
4	1	3	8
3	8	9	2

Вариант 4			
5	1	1	2
5	8	3	7
4	9	8	4
3	8	7	3

Вариант 7			
7	5	4	8
7	7	1	3
3	2	6	3
2	5	8	8

Вариант 9			
2	9	10	2
4	4	4	1
8	5	4	7
2	2	2	3

Вариант 2			
1	5	5	3
8	2	7	6
3	9	6	3
4	9	5	2

Вариант 5			
4	8	3	1
6	9	2	2
2	8	3	1
8	8	1	4

Вариант 8			
3	7	6	4
7	7	10	7
4	8	8	7
6	1	8	6

Вариант 10			
7	8	8	4
4	8	5	5
6	6	5	4
1	8	7	3

Вариант 3			
3	9	7	7
2	2	6	2
1	9	8	2
7	7	4	1

Вариант6			
1	5	9	3
4	5	8	8
2	2	7	9
5	9	2	4

Задание 5:

1) С помощью эвристического алгоритма Свенна найти интервал неопределенности.
2) С помощью указанного в варианте метода сделать одну итерацию цикла поиска оптимума.

Вариант	**Метод**	**Функция**
1	Метод дихотомии	$f(x) = 2x^2 + \frac{16}{x}$

2	Метод деления интервала пополам	$f(x)=\frac{127}{4}x^2-\frac{61}{4}x+2$
3	Метод золотого сечения	$f(x)=2x^2-12x$
4	Метод Фибоначчи	$f(x)=(500-2x)^2$
5	Метод дихотомии	$f(x)=2x^2-12x$
6	Метод золотого сечения	$f(x)=\frac{127}{4}x^2-\frac{61}{4}\lfloor x-0{,}3\rfloor+2$
7	Метод Фибоначчи	$f(x)=2\lfloor x-0{,}7\rfloor^2+\frac{16}{x}$
8	Метод дихотомии	$f(x)=(50-13x)^2$
9	Метод деления интервала пополам	$f(x)=\frac{25}{8}x^2-\frac{1}{3}x+4\cdot\lvert x-110\rvert$
10	Метод дихотомии	$f(x)=(50+x^2-2x)^2$
11	Метод золотого сечения	$f(x)=2\lceil x-0{,}3\rceil^2+\frac{16}{x}$
12	Метод Фибоначчи	$f(x)=(500-2x)^2$

Тема 7. Методы выбора решения (эвристические).

Основные вопросы темы

1. Понятие эвристики. Эволюционное моделирование (генетические алгоритмы и метод группового учёта аргумента), эвристическое программирование.

Рекомендации по изучению темы

Вопрос 1 изложен в учебнике [7] на с. 32-36.

Вопросы для самоподготовки

Рекомендуется после изучения материалов лекций и специальной литературы подготовить ответы на вопросы:

1. Какие недостатки у эвристического подхода?
2. Что такое эволюционное моделирование?
3. Что такое эвристическое программирование?
4. Что такое эвристика?
5. Что такое кроссовер?

6. Какие методы мутации существуют?
7. Какие бывают методы отбора?
8. Как осуществляется кодирование признаков?

Контрольные тесты

1. Выбор решений при неопределенности это

Выберите один ответ:

a. Игры, где одним из определяющих факторов является внешняя среда или природа, которая может находится в одном из состояний, которые известны лицу, принимающему решение

b. Правильного ответа нет

c. Игры, где всефакторыизвестны

d. Игры, где одним из определяющих факторов является внешняя среда или природа, которая может находится в одном из состояний, которые неизвестны лицу, принимающему решение

2. Выбор, сделанный только на основе ощущения того, что он правильный – это:

Выберите один ответ:

a. все ответы верны.

b. рациональное решение;

c. профессиональное решение;

d. интуитивное решение;

e. решение, основанное на суждении;

Тема 8. Методы извлечения знаний.

Основные вопросы темы

1. Технология Data Mining (определение, задачи, модели, методы, этапы).
2. Методы классификации и регрессии: построения правил классификации, деревьев решений, математических функций; поиска ассоциативных правил (алгоритм Aproiri), кластеризации (базовые и адаптивные методы).
3. Visual- и Text- Mining. Стандарты технологии.

Рекомендации по изучению темы

Вопрос 1,3 изложен в учебнике [6] на с. 79-89.

Вопрос 2 изложен в учебнике [5] в главах 1-5.

Вопросы для самоподготовки

Рекомендуется после изучения материалов лекций и специальной литературы подготовить ответы на вопросы:

1. Какие задачи DM выделяют?
2. Что такое DM?
3. Какие существуют модели DM?
4. Классификация методов DM?
5. Этапы DM?
6. В чем заключается метод одного правила?
7. В чем заключается наивный байевский метод?
8. В чем заключается метод «разделяй и властвуй»?

Контрольные тесты

1. Совокупность всех имеющихся описаний прецедентов называется:

Выберите один ответ:

- a. базой данных
- b. тестирующей выборкой
- c. данными
- d. обучающей выборкой

2. Как соотносятся Data Mining и Machine Learning?

Выберите один ответ:

- a. Data Mining и Machine Learning не пересекаются, это разные направления
- b. Data Mining входит в Machine Learning
- c. Machine Learning входит в Data Mining
- d. Data Mining и Machine Learning пересекаются

3. Как соотносятся Machine Learning и Artificial Intelligence?

Выберите один ответ:

- a. Machine Learning и Artificial Intelligence пересекаются
- b. Artificial Intelligence входит в Machine Learning
- c. Machine Learning и Artificial Intelligence не пересекаются

d. Machine Learning входит в Artificial Intelligence

4. При обучении без учителя результатом может быть

Выберите один или несколько ответов:

a. *тактика*

b. *вектор*

c. *кластеры*

d. *деревья решений*

e. *индикатор класса*

f. *число*

g. *стратегия*

h. *правила*

5. Явление, при котором ошибка обученной модели оказывается слишком большой, называется переобучением.

Выберите один ответ:

Верно

Неверно

6. Что из ниже приведенного является определением машинного обучения?

Выберите один или несколько ответов:

a. математические и кибернетические методы искусственного интеллекта, используемые для задач обучения людей

b. компьютерная программа обучается на основе опыта E по отношению к некоторому классу задач T и меры качества P, если качество решения задач из T, измеренное на основе P, улучшается с приобретением опыта E.

c. подраздел искусственного интеллекта, изучающий методы построения алгоритмов, способных обучаться

d. процесс, в результате которого машина (компьютер) способна показывать поведение, которое в неё не было явно заложено (запрограммировано)

7. При обучении с учителем результатом может быть

Выберите один или несколько ответов:

a. число/вектор

- [ ] b. деревья решений
- [ ] c. индикатор класса
- [ ] d. правила
- [ ] e. кластеры

8. Трансдуктивное обучение можно отнести к

Выберите один ответ:

- a. обучению с подкреплением
- b. обучению с учителем
- c. обучению без учителя
- d. частичному обучению

Задания для практических занятий и самостоятельной работы

Задание:

С помощью метода 1 Rule найти правила. Зависимые и независимые переменные определить самому (выбор обосновать).

Вариант 1

	A	B	C	D
1	Фамилия	Продукт	Дата	Сумма
2	Иванов	Книжка	01.01.2011	200
3	Петров	Ручка	16.01.2011	300
4	Сидоров	Линейка	28.01.2011	100
5	Иванов	Книжка	05.02.2011	250
6	Петров	Книжка	16.02.2011	150
7	Сидоров	Ручка	21.02.2011	50
8	Иванов	Линейка	28.02.2011	300
9	Петров	Линейка	04.03.2011	250
10	Сидоров	Книжка	09.03.2011	300
11	Иванов	Ручка	12.03.2011	100
12	Петров	Линейка	21.03.2011	150
13	Сидоров	Линейка	29.03.2011	300

Вариант 2

№ п.п.	Фамилия	Имя	Ср. балл	Дата рождения	Пол	Возраст	Отличница
1	Иванов	Сергей	3	12.01.1993	м	15	0
2	Петрова	Елена	3,7	14.05.1992	ж	16	0
3	Сидорова	Елизавета	4,4	30.03.1993	ж	15	0
4	Семенов	Роман	4,2	04.01.1993	м	15	0
5	Аникина	Инга	3,9	20.11.1992	ж	15	0
6	Сидоренко	Петр	4	06.06.1992	м	16	0
7	Прокошева	Оксана	4,9	22.05.1993	ж	15	0
8	Ошуркова	Ирина	4,3	21.04.1993	ж	15	0
9	Золотых	Игорь	5	05.07.1992	м	16	0
10	Дорошенко	Денис	3,6	04.08.1992	м	16	0
11	Светлаков	Михаил	3,1	01.03.1993	м	15	0
12	Серова	Наталья	5	15.02.1993	ж	15	1

Вариант 3

	A	B	C	D	E
1	Клиент	Счет	Тип счета	Отделение	Дата
2	Старый	23230	Срочный	Восточное	12.12.2004
3	Старый	23424	Текущий	Восточное	12.12.2004
4	Старый	543300	Депозит	Восточное	12.12.2004
5	Старый	34230	Депозит	Восточное	12.12.2004
6	Старый	3453	Текущий	Западное	12.12.2004
7	Новый	65400	Текущий	Западное	13.12.2004
8	Новый	23300	Текущий	Западное	13.12.2004
9	Новый	34200	Депозит	Южное	13.12.2004
10	Новый	45345	Депозит	Южное	13.12.2004
11	Старый	2344	Срочный	Южное	14.12.2004
12	Старый	34500	Срочный	Восточное	14.12.2004
13	Старый	345400	Срочный	Восточное	14.12.2004
14	Новый	34550	Депозит	Западное	14.12.2004
15	Новый	65004	Депозит	Западное	14.12.2004

Вариант 4

	A	B	C	D	E	F	G	H	I
1	Наименование	Категория	Квартал	Месяц	День	Город	Сумма	Менеджер	Заказчик
134	Ананас	Фрукты	I	Январь	18	Питер	$45 424	Иванов	Рамстор
235	Ананас	Фрукты	I	Февраль	2	Питер	$118 278	Дубинин	Пятерочка
237	Ананас	Фрукты	I	Февраль	2	Питер	$176 992	Дубинин	Рамстор
259	Ананас	Фрукты	I	Февраль	5	Питер	$121 452	Волина	Ашан
360	Ананас	Фрукты	I	Февраль	18	Питер	$188 710	Волина	Шангри-Ла
404	Ананас	Фрукты	I	Февраль	24	Питер	$181 246	Иванов	Пятерочка
420	Ананас	Фрукты	I	Февраль	26	Питер	$197 222	Михайлов	Ашан
484	Ананас	Фрукты	I	Март	9	Питер	$477 672	Дубинин	Рамстор
655	Ананас	Фрукты	II	Апрель	12	Питер	$211 662	Михайлов	Пятерочка
737	Ананас	Фрукты	II	Апрель	22	Питер	$13 308	Михайлов	Перекресток
778	Ананас	Фрукты	II	Апрель	28	Питер	$234 639	Петров	Перекресток
793	Ананас	Фрукты	II	Май	4	Питер	$55 311	Иванов	Пятерочка

Вариант 5

	A	B	C	D	E	F
1	Наименование	Выручка	Менеджер	Заказчик	Дата	
2	Персик	68 959	Петров	Рамстор	21.10.2013	
3	Лук	69 758	Тарасов	Пятерочка	26.02.2013	
4	Нектарин	88 432	Иванов	Перекресток	11.07.2013	
5	Картофель	11 634	Дубинин	Ашан	29.04.2013	
6	Грейпфрут	80 039	Петров	Перекресток	21.08.2013	
7	Грейпфрут	92 830	Михайлов	Рамстор	05.04.2013	
8	Морковь	13 634	Иванов	Шангри-Ла	24.08.2013	
9	Баклажан	63 729	Иванов	Шангри-Ла	01.12.2013	
10	Салат	49 137	Булкин	Ашан	04.04.2013	
11	Салат	34 911	Михайлов	Тандем	21.06.2013	
12	Киви	27 284	Иванов	Тандем	30.01.2013	
13	Капуста	98 018	Иванов	Пятерочка	06.09.2013	

Вариант 6

НОМЕР	АВТОР	НАЗВАНИЕ	ГОД	ПОЛКА
0001	Беляев А. Р.	Человек-амфибия	1987	5
0002	Кервуд Д.	Бродяги севера	1991	7
0003	Тургенев И. С.	Повести и рассказы	1982	1
0004	Олеша Ю. К.	Избранное	1987	5
0005	Беляев А. Р.	Звезда КЭЦ	1990	5
0006	Тынянов Ю. Н.	Кюхля	1979	1
0007	Толстой Л. Н.	Повести и рассказы	1986	1
0008	Беляев А. Р.	Избранное	1994	7

Вариант 7

Кольца : таблица

Металл	Проба	Камень	Цена (евро)
Белое золото	750	бриллиант	1 020,00€
Желтое золото	585	бриллиант	675,00€
Белое золото	585	бриллиант	340,00€
Серебро	925	гранат	210,00€
Красное золото	750	сапфир	420,00€
Желтое золото	585	сапфир	570,00€
Серебро	585	оникс	300,00€
Красное золото	585	бриллиант	743,00€
Белое золото	750	бриллиант	560,00€
Красное золото	585	янтарь	156,00€
Серебро	925	гранат	234,00€
Серебро	925	гранат	192,00€
Красное золото	925	бриллиант	2 175,00€

Вариант 8

Контракты : таблица

	Фамилия	Имя	Амплуа	Срок контракта	Зарплата
+	Бак	Брюс	Руководство	5	475 000€
+	Баллак	Михаэль	Полузащитни	3	102 000€
+	Бридж	Уэйн	Защитник	3	43 000€
+	Булахруз	Халид	Защитник	3	32 000€
+	Бэримор	Глен	Медик	3	50 000€
+	Джонсон	Скотт	Медик	2	23 000€
+	Диарра	Лассана	Полузащитни	4	81 000€
+	Дрогба	Дидье	Нападающий	3	105 000€
+	Жереми	Йон	Защитник	1	66 000€
+	Калу	Саломон	Нападающий	4	79 000€
+	Карвалью	Рикарду	Защитник	3	55 000€
+	Коул	Эшли	Защитник	4	73 000€
+	Коул Джо	Джо	Полузащитни	3	65 000€
+	Кудичини	Карло	Вратарь	1	45 000€

Запись: 1 из 32

Вариант 9

	A	B	C	D	E	F
1	**День**	**Наименование**	**Фирма**	**Количество**	**Цена**	**Общая сумма**
2	1	Пылесос	Bosch	3	1 200,00 ₽	3 600,00 ₽
3	2	Телевизор	LG	5	1 400,00 ₽	7 000,00 ₽
4	3	Ноутбук	Samsung	12	5 400,00 ₽	64 800,00 ₽
5	4	Телевизор	LG	7	1 400,00 ₽	9 800,00 ₽
6	5	Ноутбук	Bosch	1	4 700,00 ₽	4 700,00 ₽
7	6	Пылесос	LG	41	900,00 ₽	36 900,00 ₽
8	7	Ноутбук	Samsung	23	5 600,00 ₽	128 800,00 ₽
9	8	Ноутбук	LG	4	6 000,00 ₽	24 000,00 ₽
10	9	Пылесос	Bosch	27	1 200,00 ₽	32 400,00 ₽
11	10	Телевизор	Samsung	7	1 700,00 ₽	11 900,00 ₽
12	11	Телевизор	Bosch	9	2 100,00 ₽	18 900,00 ₽
13	12	Пылесос	LG	16	900,00 ₽	14 400,00 ₽
14	13	Ноутбук	LG	22	6 000,00 ₽	132 000,00 ₽
15	14	Телевизор	Samsung	14	1 700,00 ₽	23 800,00 ₽
16	15	Пылесос	Bosch	1	1 200,00 ₽	1 200,00 ₽

Вариант 10

Остатки товаров

Товар	Склад	Стоимость кон.	Количество к...	СтоимостьВа...	РезервТовара ...
		798957.00	13416.000		
Двери	Основной склад	10000.00	10.000		
Доски	Основной склад	13378.00	246.000		
ДСП ламинат	Склад столярного цеха	10000.00	1000.000		
ДСП пластикуб	Склад столярного цеха	48906.00	1650.000		
Кабель питания	Партионный склад	24599.00	42.000		
Кирпич	Основной склад	600000.00	10000.000		
Короба	Основной склад	8734.00	200.000		
Масло моторное "Лукойл-Стандар	Партионный склад	3340.00	48.000		
Рамы	Основной склад	40000.00	100.000		
Трубы	Основной склад	40000.00	120.000		

ЛАБОРАТОРНЫЙ ПРАКТИКУМ

Порядок выполнения лабораторных работ может быть произвольным и определяется уровнем освоения компетенций обучающегося.

Тема 4. Средства СУБД для аналитической обработки данных.

Задание лабораторной работы

Цель работы: получение практических навыков проектирования, разработки и использования хранилищ данных.

Задание: спроектируйте БД в многомерной модели представления данных использую модель звезды или снежинки (в реляционной базе) согласно полученному варианту (используя программу Open System Architect или аналогичное CASE-средство, модель должна включать не менее 5 сущностей), реализуйте спроектированную базу в СУБД PostgreSQL.

Внесите в базу тестовые данные (не менее 10 строк в каждую таблицу).

Реализуйте аналитические запросы к базе, используя следующие конструкции секционирование (partitioning), упорядочивание (order by), кадрирование (с использованием rows и range), аналитических функций сведения (crosstab), ранжирования функций (row_number, rank, dense_rank), получения значения строк (first_value, last_value, lead, lag), статистические (var, varp, stdevp, stdev).

Для справки по синтаксису используйте ресурсы:
https://postgrespro.ru/docs/postgrespro/9.5/tablefunc,
https://postgrespro.ru/docs/postgresql/9.5/tutorial-window,
http://www.sql-tutorial.ru/ru/book_crosstab.html,
https://postgrespro.ru/docs/postgrespro/9.5/functions-aggregate.

Отчет по лабораторной работе должен содержать:

1. Фамилию и номер группы учащегося, задание
2. Описание многомерной модели (схема)
3. Физическую модель БД (sql-код)
4. Перечень тестовых данных (в виде таблиц)
5. Код запросов, задача (вопрос) для решения которых можно использовать полученные наборы данных (для каждого запроса), и результаты их выполнения (принскрин с базы).

ROLAP (Relational OLAP)

ROLAP-системы позволяют представлять данные, хранимые в классической реляционной базе, в многомерной форме или в плоских локальных таблицах на файл-сервере, обеспечивая преобразование информации в многомерную модель через промежуточный слой метаданных. Агрегаты хранятся в той же БД в специально созданных служебных таблицах. В этом случае гиперкуб эмулируется СУБД на логическом уровне.

Преимущества ROLAP.

- Реляционные СУБД имеют реальный опыт работы с очень большими БД и развитые средства администрирования. При использовании ROLAP размер хранилища не является таким критичным параметром, как в случае MOLAP.
- При оперативной аналитической обработке содержимого хранилища данных инструменты ROLAP позволяют производить анализ непосредственно над хранилищем (потому что в подавляющем большинстве случаев корпоративные хранилища данных реализуются средствами реляционных СУБД).
- В случае переменной размерности задачи, когда изменения в структуру измерений приходится вносить достаточно часто, ROLAP системы с динамическим представлением размерности являются оптимальным решением, так как в них такие модификации не требуют физической реорганизации БД, как в случае MOLAP.
- Системы ROLAP могут функционировать на гораздо менее мощных клиентских станциях, чем системы MOLAP, поскольку основная вычислительная нагрузка в них ложится на сервер, где выполняются сложные аналитические SQL-запросы, формируемые системой.
- Реляционные СУБД обеспечивают значительно более высокий уровень защиты данных и хорошие возможности разграничения прав доступа.

Недостатки ROLAP.

- Ограниченные возможности с точки зрения расчета значений функционального типа.
- Меньшая производительность, чем у MOLAP. Для обеспечения сравнимой с MOLAP производительности реляционные системы требуют тщательной проработки схемы БД и специальной настройки индексов. Но в результате этих операций производительность хорошо настроенных реляционных систем при

использовании схемы "звезда" сравнима с производительностью систем на основе многомерных БД.

Схема звезда. Преимущества и недостатки

Схема типа звезды (Star Schema) - схема реляционной базы данных, служащая для поддержки многомерного представления содержащихся в ней данных.

Особенности ROLAP-схемы типа "звезда"

1. Одна таблица фактов (fact table), которая сильно денормализована. Является центральной в схеме, может состоять из миллионов строк и содержит суммируемые или фактические данные, с помощью которых можно ответить на различные вопросы.
2. Несколько денормализованных таблиц измерений (dimensional table). Имеют меньшее количество строк, чем таблицы фактов, и содержат описательную информацию. Эти таблицы позволяют пользователю быстро переходить от таблицы фактов к дополнительной информации.
3. Таблица фактов и таблицы размерности связаны идентифицирующими связями, при этом первичные ключи таблицы размерности мигрируют в таблицу фактов в качестве внешних ключей. Первичный ключ таблицы факта целиком состоит из первичных ключей всех таблиц размерности.
4. Агрегированные данные хранятся совместно с исходными.

Преимущества

Благодаря денормализации таблиц измерений упрощается восприятие структуры данных пользователем и формулировка запросов, уменьшается количество операций соединения таблиц при обработке запросов. Некоторые промышленные СУБД и инструменты класса OLAP / Reporting умеют использовать преимущества схемы "звезда" для сокращения времени выполнения запросов.

Недостатки

Денормализация таблиц измерений вносит избыточность данных, возрастает требуемый для их хранения объем памяти. Если агрегаты хранятся совместно с исходными данными, то в измерениях необходимо использовать дополнительный параметр - уровень иерархии.

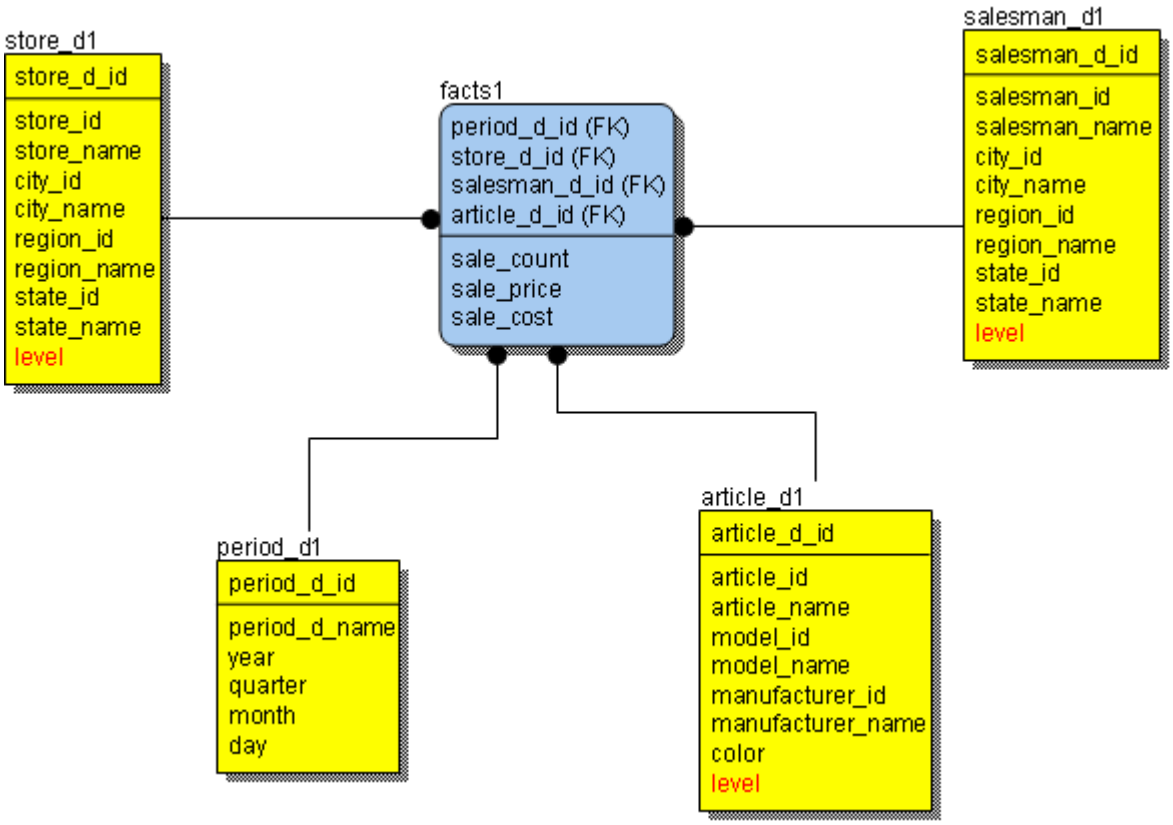


Схема снежинка. Преимущества и недостатки

Схема типа снежинки (Snowflake Schema) - схема реляционной базы данных, служащая для поддержки многомерного представления содержащихся в ней данных, является разновидностью схемы типа "звезда" (Star Schema).

Особенности ROLAP-схемы типа "снежинка"

1. Одна таблица фактов (fact table), которая сильно денормализована. Является центральной в схеме, может состоять из миллионов строк и содержать суммируемые или фактические данные, с помощью которых можно ответить на различные вопросы.
2. Несколько таблиц измерений (dimensional table), которые нормализованы в отличие от схемы "звезда". Имеют меньшее количество строк, чем таблицы фактов, и содержат описательную информацию. Эти таблицы позволяют пользователю быстро переходить от таблицы фактов к дополнительной информации. Первичные ключи в них состоят из единственного атрибута (соответствуют единственному элементу измерения).
3. Таблица фактов и таблицы размерности связаны идентифицирующими связями, при этом первичные ключи таблицы размерности мигрируют в таблицу фактов в качестве внешних ключей. Первичный ключ таблицы факта целиком состоит из первичных ключей всех таблиц размерности.

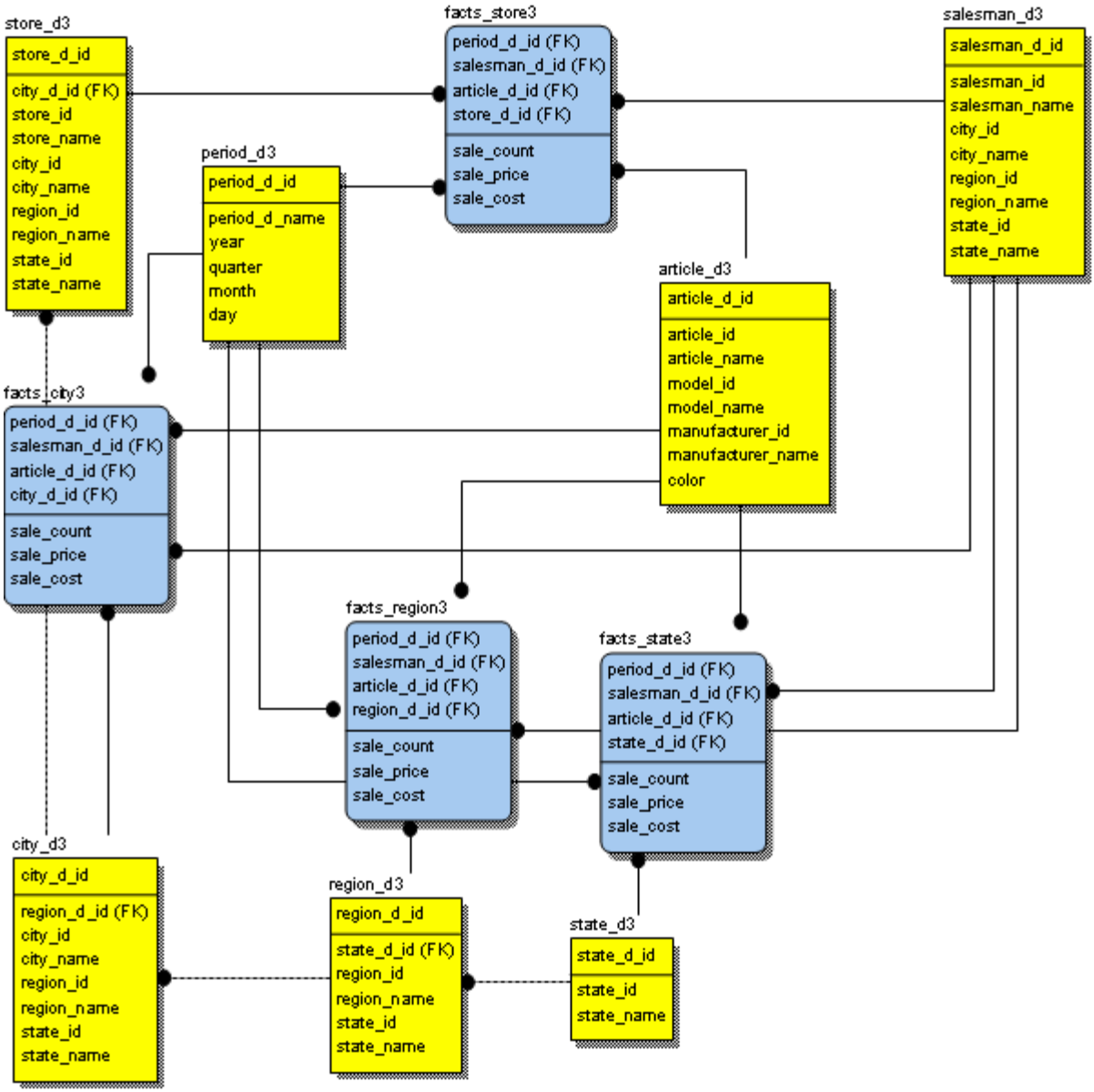
4. В схеме "снежинка" агрегированные данные могут храниться отдельно от исходных.

Преимущества

Нормализация таблиц измерений в отличие от схемы "звезда" позволяет минимизировать избыточность данных и более эффективно выполнять запросы, связанные со структурой значений измерений.

Недостатки

За нормализацию таблиц измерений иногда приходится платить временем выполнения запросов.



Варианты для выполнения лабораторной работы

1) Погодные условия в регионе
2) Продажа комплектующих изделий
3) Демографическая ситуация в регионе
4) Продажа земельных участков
5) Рынок труда
6) Больница
7) Железнодорожный транспорт
8) Авиационные перевозки
9) Олимпиада
10) Футбол
11) Туристический бизнес
12) Социальные сети
13) Интернет-провайдер
14) Здравоохранение
15) Автострахование
16) Кредитование
17) Экология
18) Правонарушения
19) Литература
20) Компьютеры

Тема 6. Методы выбора решений (рациональные).

Задание лабораторной работы

Цель работы: получение практических навыков оптимизации.

Задание: постройте согласно варианту программу, которая позволяет решать задачу одномерной оптимизации (для написания программы можно использовать любой язык программирования высокого уровня). Интервал неопределенности найти с помощью эвристического алгоритма Свенна (на этом интервале функция должна быть унимодальной). Программа должна выводить график функции на выбранном интервале.

Отчет по лабораторной работе должен содержать:

6. Фамилию и номер группы учащегося, задание
7. Описание поиска интервала
8. Результат выполнения программы.
9. Код программы.

Методические указания по выполнению лабораторной работы

Оптимизация функции одной переменной - наиболее простой тип оптимизационных задач. В методах одномерной оптимизации вместо X=R рассматривается отрезок X=[a,b], содержащей искомое решение Xmin. Такой отрезок называется отрезком неопределенности, или отрезком локализации. Целевая функция f(x) является унимодальной.

Функция f(x) называется унимодальной на X=[a,b], если существует единственная точка Xmin, в которой функция достигает минимума, причем слева от точки функция строго убывает, а справа строго возрастает.

Метод установления границ начального отрезка локализации минимума (алгоритм Свенна).

Шаг 1. Выбрать произвольную начальную точку $x^{(0)}$ и Δ – начальный положительный шаг.

Шаг 2. Вычислить $f(x^{(0)}),\ f(x^{(0)}+\Delta)$

Шаг 3. Сравнить $f(x^{(0)}),\ f(x^{(0)}+\Delta)$:

а) если $f(x^{(0)}) > f(x^{(0)}+\Delta)$ то, согласно предположению об унимодальности функции, точка минимума должна лежать правее, чем точка $x^{(0)}$. Положить $a = x^{(0)}$, $x^{(1)} = x^{(0)} + \Delta$, $f(x^{(1)}) = f(x^{(0)} + \Delta)$, k=2 и перейти на шаг 5.

б) если $f(x^{(0)}) \le f(x^{(0)} + \Delta)$, то вычислить $f(x^{(0)} - \Delta)$.

Шаг 4. Сравнить $f(x^{(0)} - \Delta),\ f(x^{(0)})$:

а) если $f(x^{(0)} - \Delta) \ge f(x^{(0)})$, то точка минимума лежит между точками $x^{(0)} - \Delta$ и $x^{(0)} + \Delta$, которые и образуют границы начального отрезка локализации минимума. Положить $a = x^{(0)} - \Delta,\ b = x^{(0)} + \Delta$ и завершить поиск.

б) если $f(x^{(0)} - \Delta) < f(x^{(0)})$ то, согласно предположению об унимодальности функции, точка минимума должна лежать левее, чем точка $x^{(0)}$. Положить $b = x^{(0)},\ x^{(1)} = x^{(0)} - \Delta,\ f(x^{(1)}) = f(x^{(0)} - \Delta),\ \Delta = -\Delta$, k=2 и перейти на шаг 5.

Шаг 5. Вычислить $x^{(k)} = x^{(0)} + 2^{k-1} \cdot \Delta,\ f(x^{(k)})$.

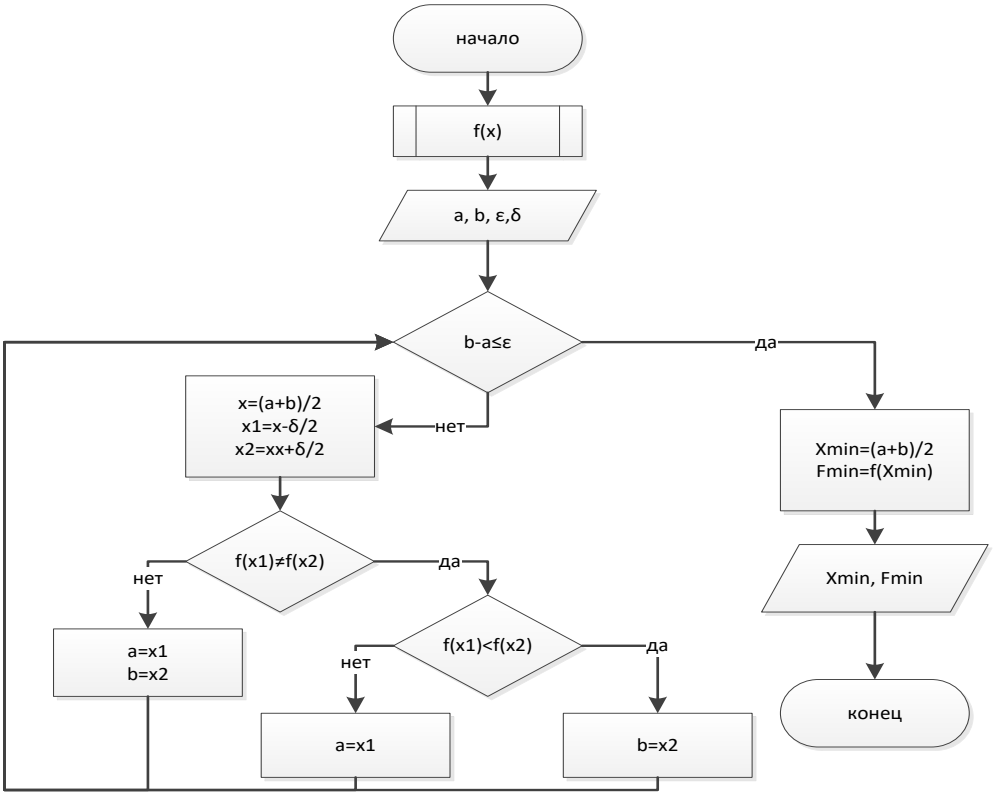
Шаг 6. Сравнить $f(x^{(k)}),\ f(x^{(k-1)})$:

а) если $f(x^{(k-1)}) \le f(x^{(k)})$, то при $\Delta > 0$ положить $b = x^{(k)}$ при $\Delta < 0$ положить $a = x^{(k)}$ и завершить поиск.

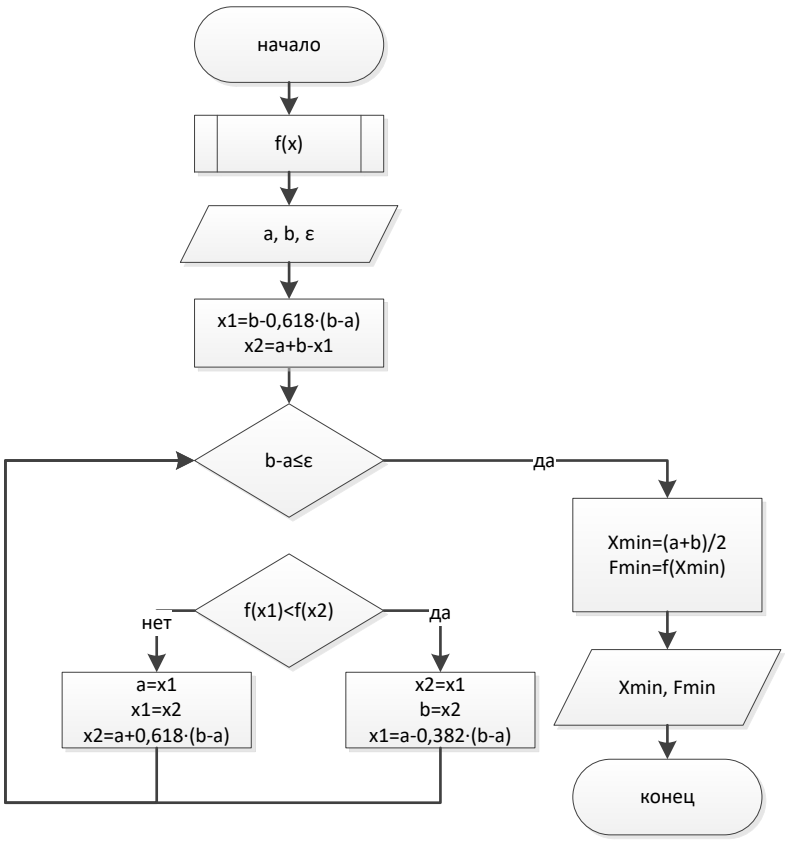
б) если $f(x^{(k-1)}) > f(x^{(k)})$, то при $\Delta > 0$ положить $a = x^{(k-1)}$ при $\Delta < 0$ положить $b = x^{(k-1)}$ и положить k=k+1 и перейти на шаг 5.

Для поиска точки Xmin с заданной точностью ε используют различные методы:

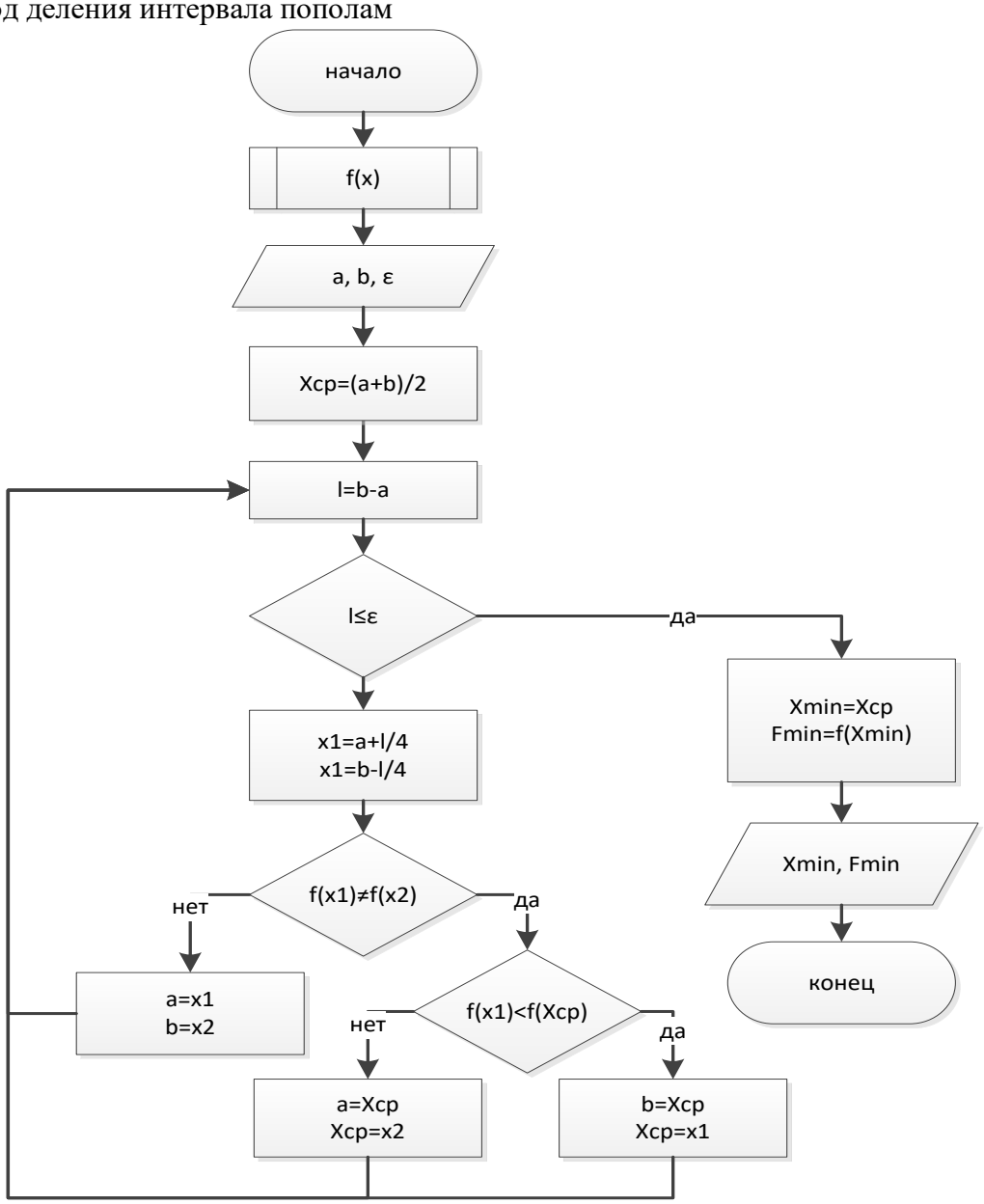
1) метод дихотомии



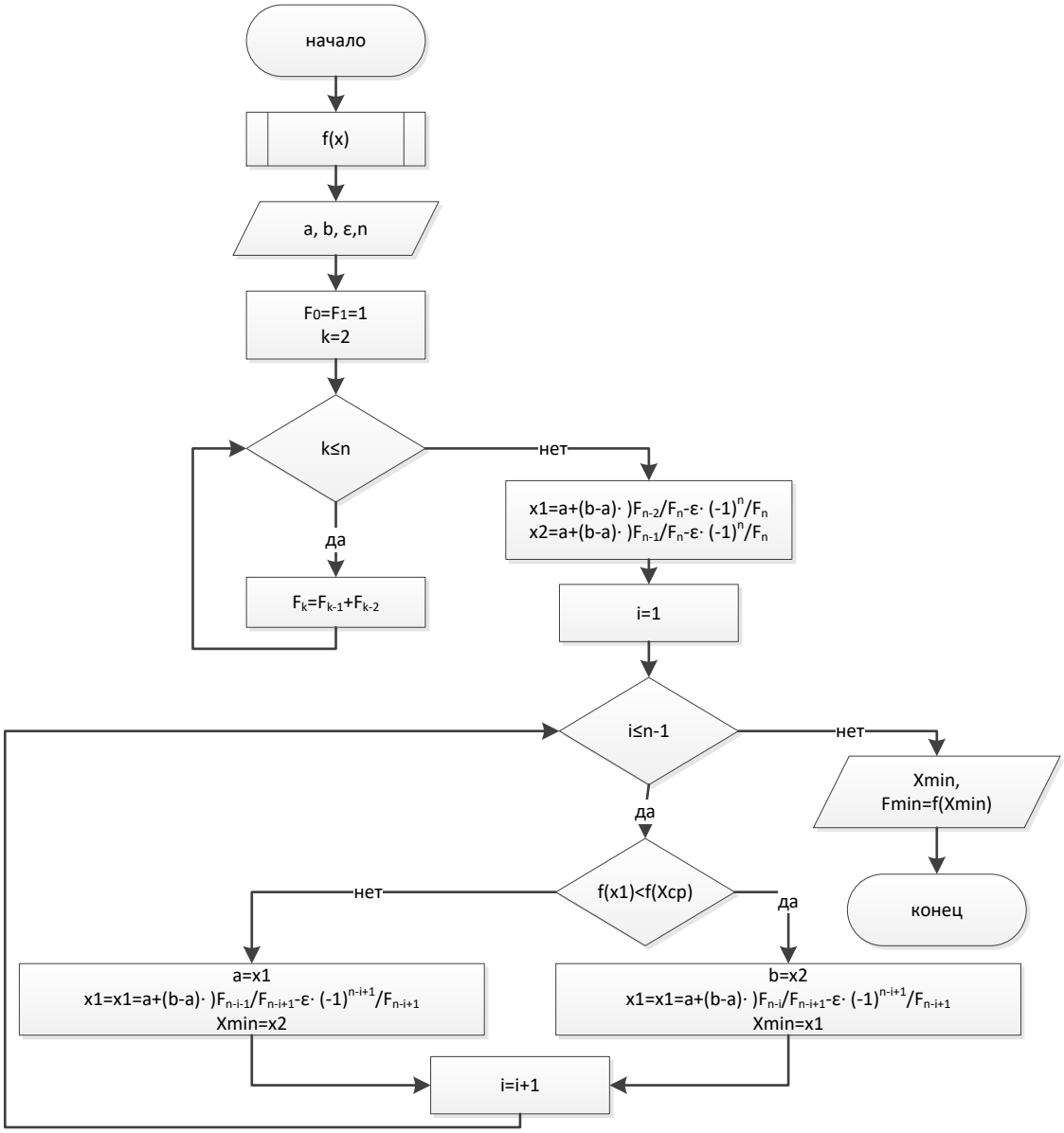
2) метод золотого сечения



3) метод деления интервала пополам



4) метод Фибоначчи



Варианты для выполнения лабораторной работы

Вариант	Метод	Функция
1	Метод дихотомии	$f(x) = 2x^2 + \frac{16}{x}$
2	Метод деления интервала пополам	$f(x) = \frac{127}{4}x^2 - \frac{61}{4}x + 2$
3	Метод золотого сечения	$f(x) = 2x^2 - 12x$
4	Метод Фибоначчи	$f(x) = (500 - 2x)^2$
5	Метод дихотомии	$f(x) = 2x^3 - (x\text{-}2) \cdot x$

6	Метод деления интервала пополам	$f(x)=\frac{215}{4}x^3-\frac{1}{2}x+4\cdot\left\|3+x-2^x\right\|$
7	Метод золотого сечения	$f(x) = 2x^3 - (x\text{-}2)\cdot x$
8	Метод Фибоначчи	$f(x)=\frac{127}{4}\left\|x-1\right\|^2-\frac{34}{4}\lceil x-0{,}4\rceil+2$
9	Метод дихотомии	$f(x) = 2x^2 - 12x$
10	Метод деления интервала пополам	$f(x) = 67x^3 - (x\text{-}4)\cdot x$
11	Метод золотого сечения	$f(x)=\frac{127}{4}x^2-\frac{61}{4}\lfloor x-0{,}3\rfloor+2$
12	Метод Фибоначчи	$f(x)=2\lfloor x-0{,}7\rfloor^2+\frac{16}{x}$
13	Метод дихотомии	$f(x) = (50 - 13x)^2$
14	Метод деления интервала пополам	$f(x)=\frac{25}{8}x^2-\frac{1}{3}x+4\cdot\left\|x-110\right\|$
15	Метод золотого сечения	$f(x) = 6x^3 - (x\text{-}2)\cdot x$
16	Метод Фибоначчи	$f(x) = 45x^3 - (x\text{-}2)\cdot x$
17	Метод дихотомии	$f(x)=\frac{215}{4}x^3-\frac{1}{2}x+4\cdot\left\|x\right\|$
18	Метод деления интервала пополам	$f(x) = (50 + x^2 - 2x)^2$
19	Метод золотого сечения	$f(x)=2\lceil x-0{,}3\rceil^2+\frac{16}{x}$
20	Метод Фибоначчи	$f(x) = (500 - 2x)^2$

Тема 8. Методы извлечения знаний.

Задание лабораторной работы

Цель работы: Получение практических навыков анализа данных.

Задание: Используя программное средство Weka, выполните анализ данных согласно полученному варианту. Работа состоит из нескольких этапов:

1) Подготовка данных для анализа в полученной согласно варианту предметной области (атрибутов должно быть не менее 10, строк с данными не менее 100, строки должны быть уникальными)
2) Загрузка данных в систему, рассмотрение описания данных (максимальных, минимальных значений и т.д.)

3) Построение моделей различными методами:

– Регрессионной,

– Классификации

– Кластеризации

– Ассоциативной

4) Исследование моделей, их интерпретация и выводы о возможности их применения

Отчет по лабораторной работе должен содержать:

10. Фамилию и номер группы, задание
11. Описание данных
12. Описание процесса построения моделей
13. Описание полученного результата (с визуализацией)
14. Интерпретация полученных результатов и выводы
15. Листинги данных и моделей.

Методические указания по выполнению лабораторной работы

Weka (Waikato Environment for Knowledge Analysis) — свободное программное обеспечение для анализа данных, написанное на Java в университете Уайкато (Новая Зеландия), распространяющееся по лицензии GNU GPL. Система представляет собой систему библиотек функции обработки данных, плюс несколько графических интерфейсов к этим библиотекам. Основной интерфейс системы - Explorer. Он позволяет выполнять практически все действия, которые предусмотрены в системе.

Также в системе Weka предусмотрены другие интерфейсы - Knowledge Flow для работы с большими массивами данных (Explorer загружает все данные в память сразу, и потому работа с большими массивами затруднена) и Experimenter для экспериментального подбора наилучшего метода анализа данных.

Weka предоставляет доступ к SQL-базам через Java Database Connectivity (JDBC) и в качестве исходных данных может принимать результат SQL-запроса. Возможность обработки множества связанных таблиц не поддерживается, но существуют утилиты для преобразования таких данных в одну таблицу, которую можно загрузить в *Weka.*

Explorer имеет несколько панелей.

1) Панель предобработки Preprocess panel позволяет импортировать данные из базы, CSV файла и т. д., и применять к ним алгоритмы фильтрации, например, переводить количественные признаки в дискретные, удалять объекты и признаки по заданному критерию.

2) Панель классификации Classify panel позволяет применять алгоритмы классификации и регрессии (в Weka они не различаются и называются classifiers) к выборке данных, оценивать предсказательную способность алгоритмов, визуализировать ошибочные предсказания, ROC-кривые, и сам алгоритм, если это возможно (в частности, решающие деревья).
3) Панель поиска ассоциативных правил Associate panel решает задачу выявления всех значимых взаимосвязей между признаками.
4) Панель кластеризации Cluster panel даёт доступ к алгоритму k-средних, EM-алгоритму для смеси гауссианов и другим.
5) Панель отбора признаков Select attributes panel даёт доступ к методам отбора признаков.
6) Панель визуализации Visualize строит матрицу графиков разброса (scatter plot matrix), позволяет выбирать и увеличивать графики, и т. д.

WEKA использует Java, так что если на вашем компьютере нет JRE, выберите для установки версию WEKA, включающую в себя JRE.

Рисунок 1 - Стартовое окно WEKA

При запуске WEKA, пакет предлагает вам на выбор 4 графических интерфейса для работы с WEKA и вашими данными. Выберете Explorer.

Анализ данных подразумевает наличие самих данных в системе. Для того чтобы загрузить данные в WEKA, их следует преобразовать в формат, понятный этому программному пакету. Наиболее подходящим форматом для загрузки данных в WEKA является формат Attribute-Relation File Format (ARFF), который сначала определяет тип загружаемых данных, а потом указывает собственно данные.

В файле формата ARFF вы указываете название и тип данных для каждого столбца таблицы, а затем собственно данные по строкам.

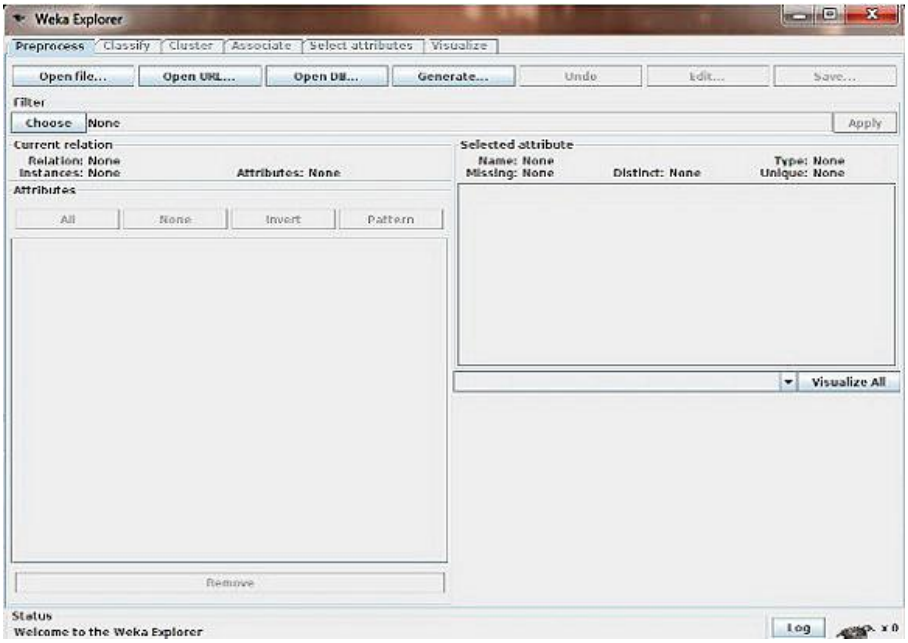


Рисунок 2 - Окно WEKA Explorer

Подготовка данных состоит в добавлении полей метаданных в начало файла. На отдельных строчках добавляется следующая информация:

-названия зависимости @relation имя,

-описания атрибутов @attribute: имя, тип

-@data перед началом самих данных.

Различают следующие типы данных:

-численные (numeric, real, integer),

-перечислимые(nominal) (задаются перечислением вида {i1, ..., in}),

-строковые (string),

-дата (date [date format]).

- составной тип (relational),

Регрессия

В моделях регрессионного анализа используются всего два типа данных: NUMERIC и DATE.

Метод регрессионного анализа является самым простым и, пожалуй, наименее эффективным методом интеллектуального анализа данных (удивительно, как часто эти качества сопутствуют друг другу). Самая простая модель анализа использует один входной (независимый) параметр и один результирующий (зависимый) параметр, модель можно усложнить, добавив несколько десятков входных параметров, но в любом случае общий подход будет один и тот же: на основании нескольких независимых переменных определяется один зависимый результат. Таким образом, модель регрессионного анализа

используется для прогнозирования значения одной зависимой переменной, исходя из известных значений нескольких независимых параметров.

Наверняка, каждый из нас хотя бы раз сталкивался с регрессионной моделью, а может быть, и проводил в уме самостоятельный регрессионный анализ. Наиболее очевидный пример – определение стоимости дома. Цена на дом (зависимая переменная) определяется несколькими независимыми параметрами: какова площадь дома и размер участка, используется ли в оформлении кухни гранитные плиты, каково качество и срок службы сантехники и так далее. Так что, если вам случалось когда-либо продавать или покупать дом, то, скорее всего, вы использовали регрессионный анализ для определения его цены. Вы оценивали параметры похожих домов в этом же районе и цену, по которой эти дома были проданы (т.е. создавали модель), а затем подставляли параметры вашего дома в полученную зависимость и рассчитывали предполагаемую стоимость вашего дома.

Давайте воспользуемся моделью регрессионного анализа для определения цены дома и разберем конкретный пример. В таблице внизу указаны фактические параметры домов, выставленных на продажу в моем районе. На основании этих данных я попробую оценить стоимость моего дома (и воспользуюсь этими результатами, чтобы опротестовать предъявленную мне сумму налога на недвижимость).

Таблица 1. Регрессионная модель оценки стоимости дома

Площадь дома (кв.футы)	**Размер участка**	**Количество спален**	**Гранитная отделка на кухне**	**Современное сантехническое оборудование?**	**Продажная цена**
3529	9191	6	0	0	$205,000
3247	10061	5	1	1	$224,900
4032	10150	5	0	1	$197,900
2397	14156	4	1	0	$189,900
2200	9600	4	0	1`	$195,000
3536	19994	6	1	1	$325,000
2983	9365	5	0	1	$230,000
3198	9669	5	1	1	????

Файл с данными (отношение \ таблица Дом, атрибуты размер дома, ванная и т.д. и собственно данные через запятую):

@RELATION house

@ATTRIBUTE houseSize NUMERIC

@ATTRIBUTE lotSize NUMERIC

@ATTRIBUTE bedrooms NUMERIC

@ATTRIBUTE granite NUMERIC

@ATTRIBUTE bathroom NUMERIC

@ATTRIBUTE sellingPrice NUMERIC

@DATA

3529,9191,6,0,0,205000

3247,10061,5,1,1,224900

4032,10150,5,0,1,197900

2397,14156,4,1,0,189900

2200,9600,4,0,1,195000

3536,19994,6,1,1,325000

2983,9365,5,0,1,230000

Запустите WEKA и выберите опцию Explorer. В результате откроется закладка Preprocess окна Explorer. Щелкните на кнопке Open File и выберите созданный вами ARFF-файл. Окно WEKA Explorer с загруженными данными о домах показано на рисунке 3.

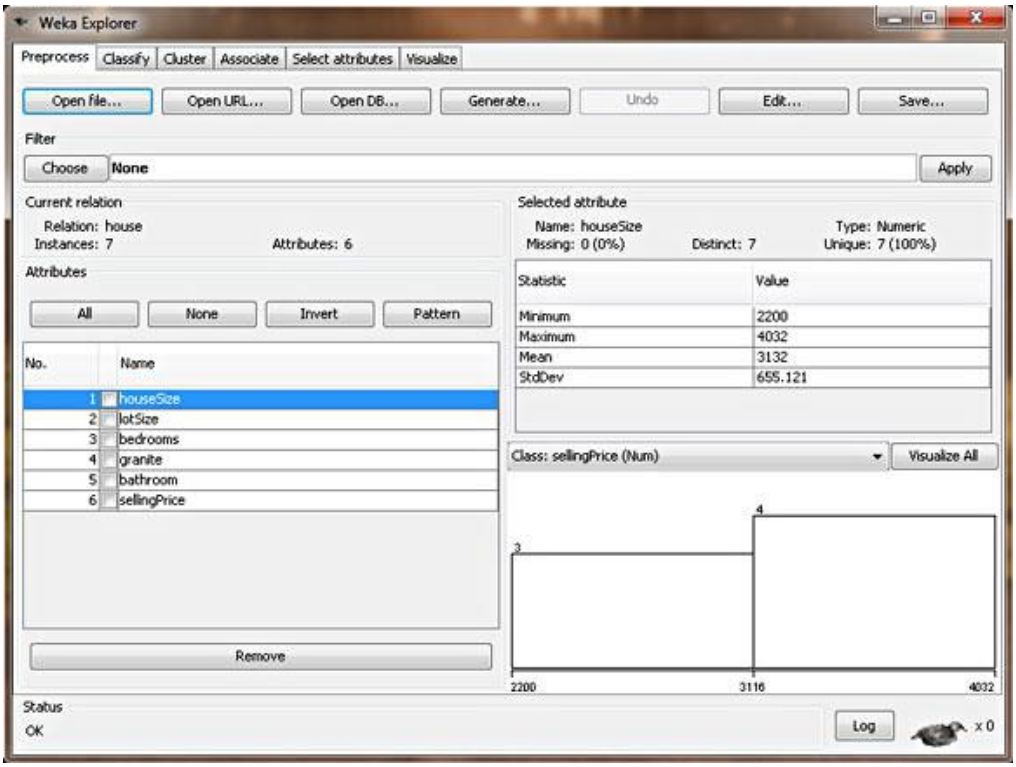


Рисунок 3- Окно WEKA Explorer с загруженными данными о домах

В этом окне вы можете проверить данные, на основании которых вы собираетесь строить модель. В левой части окна Explorer показаны параметры объектов (Attributes), которые соответствуют заголовкам столбцов нашей исходной таблицы, а также указано количество объектов (Instances), т.е. строк таблицы. Если вы щелкните мышкой на одном из заголовков столбцов, то в правой панели будет выведена полная информация о наборе данных в данном столбце. Например, если мы выберем столбец houseSize в левой панели (он выбран по умолчанию), то в правой панели отобразится дополнительная

статистическая информация по этому столбцу. Будет показано максимальное значение в столбце (4032 кв.фута) и минимальное значение (2200 кв.футов). Кроме того, будет подсчитано среднее значение (3131 кв.фут) и стандартное отклонение (655 кв.футов) (стандартное отклонение – статистический показатель рассеивания значений случайной величины). Наконец, здесь же вам предлагается возможность визуального анализа данных (кнопка Visualize All). Поскольку в нашей таблице данных не так много, то их визуальное отображение не дает такой наглядной аналитической картины, как в случае использования сотен или тысяч показателей.

Для того чтобы создать модель, откройте закладку **Classify**. В качестве первого шага, нам надо выбрать тип модели для анализа, чтобы указать WEKA, каким образом мы хотим анализировать наши данные, и какую модель построить:

1) Щелкните на копке **Choose** и разверните меню **functions**.
2) Выберите опцию **LinearRegression**.

Таким образом, мы указали WEKA, что мы хотим создать модель регрессионного анализа. Как вы заметили, меню включает целое множество моделей. Множество! Это еще раз подчеркивает факт нашего весьма поверхностного знакомства с областью интеллектуального анализа данных. Обратите внимание: в меню включена опция **SimpleLinearRegression**, однако мы не используем ее, поскольку этот тип модели определяет значение зависимой переменной по значениям одного независимого параметра, а у нас их целых шесть. Если вы выбрали правильную модель, то окно WEKA Explorer должно выглядеть так, как показано на рисунке 4.

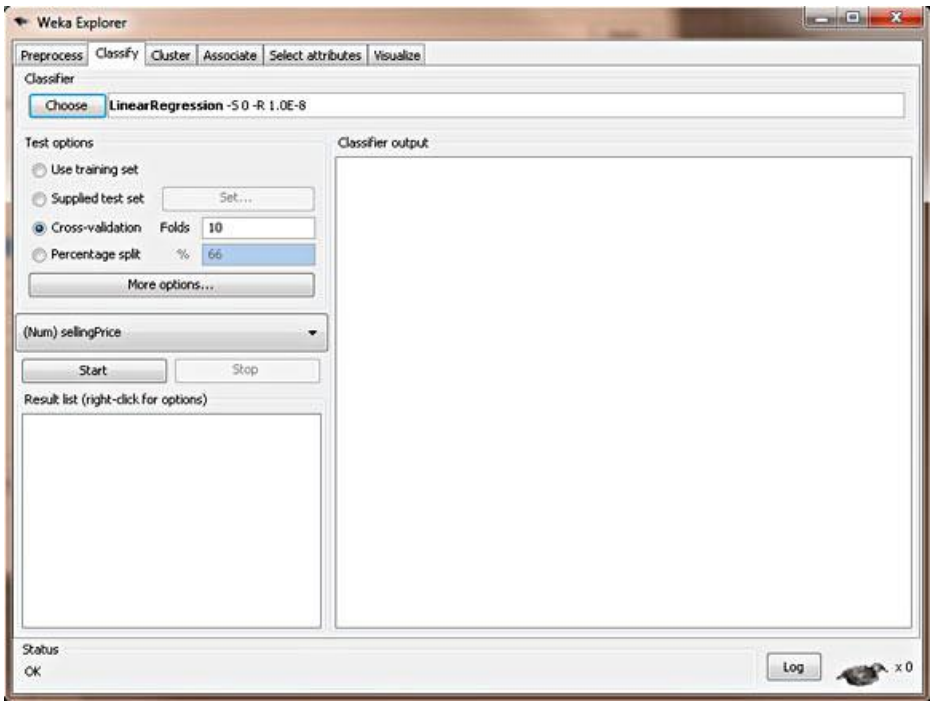


Рисунок 4 - Модель линейного регрессионного анализа WEKA

После того, как мы выбрали тип модели, нужно указать WEKA, какие данные должны использоваться для ее создания. Несмотря на то, что ответ на этот вопрос для нас вполне очевиден – нужно взять данные из созданного нами ARFF-файла – существует несколько других, более сложных, возможностей предоставления данных для анализа. Опция **Supplied test set** позволяет указать дополнительный набор тестовых данных для модели, опция **Cross-validation** использует несколько наборов данных, усредняет их и строит модель на основе средних значений, а опция**Percentage split** использует в качестве базы для модели процентили набора данных. Эти способы применяются для создания аналитических моделей, которые мы рассмотрим в следующих статьях этой серии. В случае регрессионного анализа нам нужна опция **Use training set**. В этом случае WEKA создаст модель на базе данных из загруженного ARFF-файла.

Завершающий этап создания модели – выбор зависимой переменной (столбца, в котором находится неизвестное нам значение, которое требуется рассчитать). В нашем примере – это цена дома, так как именно это значение мы и хотим узнать. Сразу после секции Test options находится раскрывающийся список, в котором вам нужно выбрать зависимый параметр. По умолчанию должен быть выбран атрибут **sellingPrice**. Если это не так, выберите сами этот параметр.

Мы определили все параметры и можем приступить к созданию модели. Нажмите кнопку **Start**. В результате окно WEKA должно выглядеть так, как показано на рисунке 5.

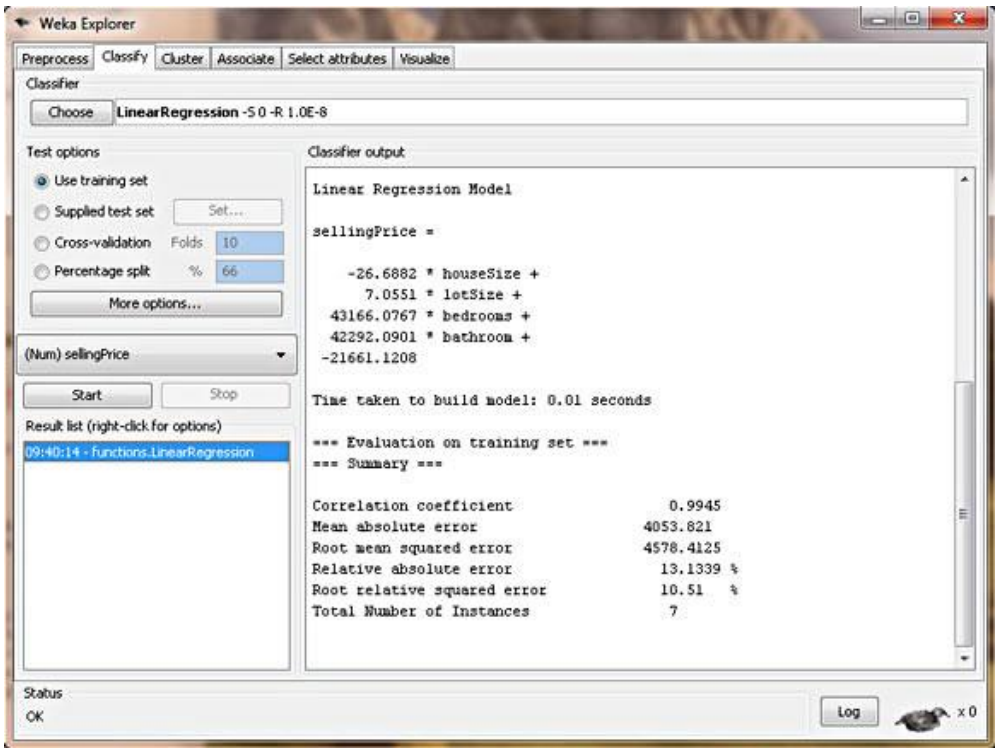


Рисунок 5 - Регрессионная модель WEKA для расчета стоимости дома

Готовая модель регрессионного анализа

```
sellingPrice = (-26.6882   * houseSize) +
               (7.0551     * lotSize) +
               (43166.0767 * bedrooms) +
               (42292.0901 * bathroom)
             - 21661.1208
```

Рассчитаем цену конкретного дома:

```
sellingPrice = (-26.6882   * 3198) +
               (7.0551     * 9669) +
               (43166.0767 * 5) +
               (42292.0901 * 1)
             - 21661.1208
```

```
sellingPrice = 219,328
```

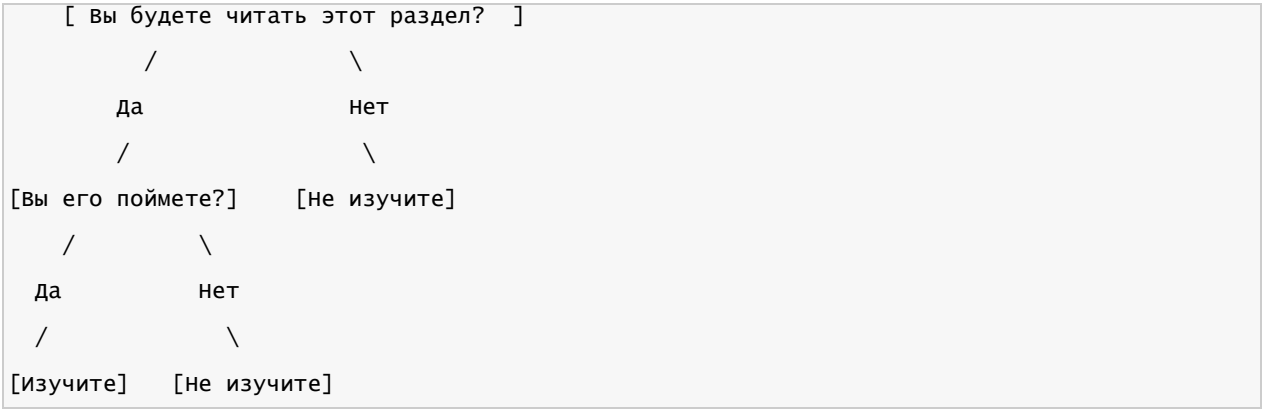
Рассмотрим зависимости между данными нашей модели и постараемся сделать определенные выводы относительно правил формирования цен на недвижимость.

- Гранитные элементы в оформлении кухни не влияют на цену дома — WEKA использует только те данные, которые, согласно статистике, влияют на точность модели (влияние каждого независимого параметра на зависимую переменную определяется с помощью коэффициента детерминации R-квадрат, обсуждение которого выходит за рамки этой статьи). Таким образом, параметры, не имеющие достаточного влияния на зависимую переменную, в модели не учитываются. Наша регрессионная модель свидетельствует о том, что использование гранита на кухне не влияет на цену дома.
- Состояние ванных комнат и сантехники влияет на цену дома — поскольку мы используем значения 0 или 1 в качестве показателя модернизации ванных комнат, то соответствующий коэффициент регрессионной модели демонстрирует нам, как современное сантехническое оборудование влияет на цену дома, а именно добавляет 42292$ к его цене.
- Большая площадь дома снижает его цену — Согласно модели WEKA, по мере роста площади домов, цена снижается. Это следует из того, что модель включает переменную houseSize с отрицательным коэффициентом. Что же получается? Увеличение площади дома на 1 кв.фут снижает его стоимость на 26$? Подобное утверждение кажется очевидной бессмыслицей. Мы же рассматриваем дома в Америке: чем больше, тем лучше, особенно в Техасе, где я живу. Как же это понимать? Это классический пример случая «каков

вопрос, таков и ответ». На самом деле, размер дома не является независимой величиной. Этот параметр связан, например, с количеством спален - очевидно, что в больших домах и количество спален больше. Так что наша модель, увы, не идеальна, но мы можем ее поправить. Запомните: закладка Preprocess позволяет удалить столбцы из набора данных. В качестве самостоятельного упражнения, удалите столбец houseSize и создайте новую модель. Проверьте, как изменение набора данных отразится на цене дома, и какая из двух моделей больше соответствует реальности (уточненная цена моего дома $217,894).

Классификация

Метод классификации (также известный как метод классификационных деревьев или деревьев принятия решений) - это алгоритм анализа данных, который определяет пошаговый способ принятия решения в зависимости от значений конкретных параметров. Дерево этого метода имеет следующий вид: каждый узел представляет собой точку принятия решения на основании входных параметров. В зависимости от конкретного значения параметра вы переходите к следующему узлу, от него – к следующему узлу, и так далее, пока не дойдете до листа, который и дает вам окончательное решение. Звучит довольно запутанно, но на самом деле метод достаточно прямолинеен. Давайте обратимся к конкретному примеру.



Это простое классификационное дерево определяет ответ на вопрос «Изучите ли вы принцип построения классификационного дерева?» В каждом узле вы отвечаете на соответствующий вопрос и переходите по соответствующей ветке к следующему узлу, до тех пор, пока не дойдете до листа с ответом «да» или «нет». Эта модель применима к любым сущностям, и вы сможете ответить, в состоянии ли эти сущности изучить классификационные деревья, с помощью двух простых вопросов. В этом и состоит основное преимущество классификационных деревьев – они не требует чрезмерного

количества информации для создания достаточно точного и информативного дерева решений.

J4.8 (модификация C4.5)

Набор данных, который мы будем использовать для примера классификационного анализа, содержит информацию, собранную дилерским центром BMW. Центр начинает рекламную компанию, предлагая расширенную двухгодичную гарантию своим постоянным клиентам. Подобные компании уже проводились, так что дилерский центр располагает 4500 показателями относительно предыдущих продаж с расширенной гарантией. Этот набор данных обладает следующими атрибутами:

- Распределение по доходам [0=$0-$30k, 1=$31k-$40k, 2=$41k-$60k, 3=$61k-$75k, 4=$76k-$100k, 5=$101k-$150k, 6=$151k-$500k, 7=$501k+]
- Год/месяц покупки первого автомобиля BMW
- Год/месяц покупки последнего автомобиля BMW
- Воспользовался ли клиент расширенной гарантией

Загрузите файл bmw-training.arff (см.раздел Загрузка) в программный пакет WEKA, используя те же шаги, которые мы проделали для загрузки данных в случае регрессионного анализа. Замечание: в предлагаемом файле содержатся 3000 из имеющихся 4500 записей. Мы разделили набор данных так, чтобы часть их использовалась для создания модели, а часть – для проверки ее точности, чтобы убедиться, что модель не является подогнанной под конкретный набор данных.

Откройте закладку Classify, выберите опцию trees, а затем опцию J48.Убедитесь, что выбрана опция Use training set, чтобы пакет WEKA при создании модели использовал именно те данные, которые мы только что загрузили в виде файла. Нажмите кнопку Start и предоставьте WEKA возможность поработать с нашими данными. Результирующая модель должна выглядеть так:

```
Number of Leaves  :        28

Size of the tree :         43


Time taken to build model: 0.18 seconds

=== Evaluation on training set ===
=== Summary ===

Correctly Classified Instances        1774               59.1333 %
```

```
Incorrectly Classified Instances      1226               40.8667 %
Kappa statistic                          0.1807
Mean absolute error                      0.4773
Root mean squared error                  0.4885
Relative absolute error                 95.4768 %
Root relative squared error             97.7122 %
Total Number of Instances             3000

=== Detailed Accuracy By Class ===

               TP Rate   FP Rate   Precision   Recall  F-Measure   ROC Area  Class
                 0.662     0.481      0.587     0.662     0.622      0.616    1
                 0.519     0.338      0.597     0.519     0.555      0.616    0
Weighted Avg.    0.591     0.411      0.592     0.591     0.589      0.616

=== Confusion Matrix ===

   a    b   <-- classified as
 1009  516 |    a = 1
 710 65 |    b = 0
```

Наиболее существенные данные – это показатели классификации "Correctly Classified Instances" (59.1%) и "Incorrectly Classified Instances" (40.9%). Кроме того, следует обратить внимание на число в первой строке столбца ROC Area (0.616). Чуть позже мы подробно обсудим эти значения, пока же просто запомните их. Наконец, таблица Confusion Matrix показывает количество ложноположительных (516) и ложноотрицательных (710) распознаваний. Поскольку показатель точности нашей модели – 59,1%, то в первоначальном рассмотрении ее нельзя назвать достаточно хорошей.

Вы сможете увидеть дерево, если щелкнете правой кнопкой мышки в панели результирующей модели. В контекстном меню выберите опцию Visualize tree. На экране отобразится визуальное представление классификационного дерева нашей модели (рисунок 3), однако в данном случае картинка мало чем нам поможет. Еще один способ увидеть дерево модели – прокрутить вверх вывод в окне Classifier Output, там вы найдете текстовое описание дерева с узлами и листьями.

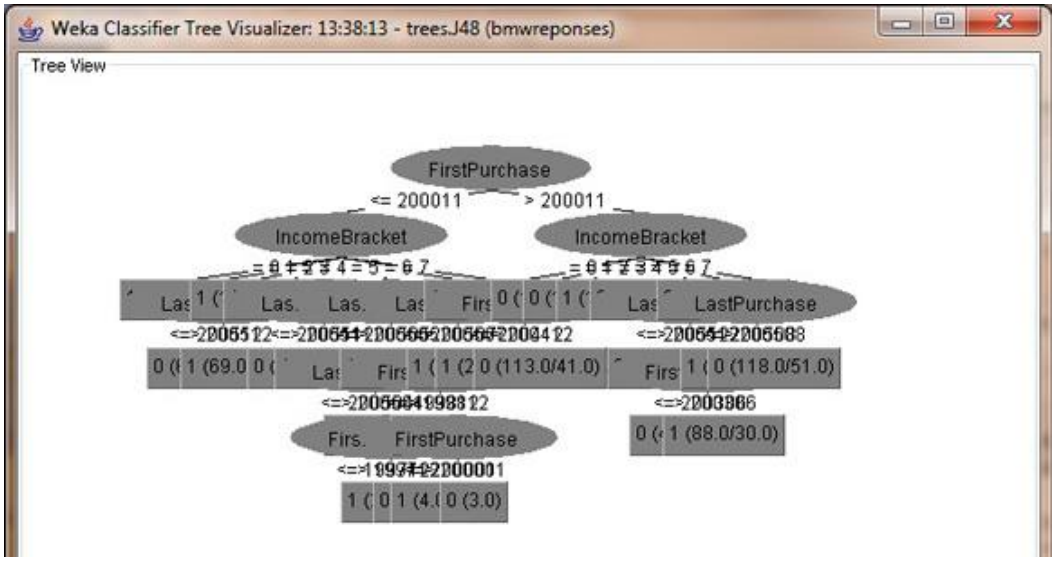


Рисунок 6 -Визуальное представление дерева классификации

Остался последний этап проверки классификационного дерева: нам надо пропустить оставшийся набор данных через полученную модель и проверить, насколько результаты классификации будут отличаться от реальных данных. Для этого в секции Test options выберите опцию Supplied test set и нажмите на кнопку Set. Укажите файл bmw-test.arff, содержащий оставшиеся 1500 данных, которые не были включены в обучающий набор. При нажатии на кнопку Start WEKA пропустит тестовые данные через модель и покажет результат работы модели. Давайте нажмем на Start и проверим, что у нас получилось.

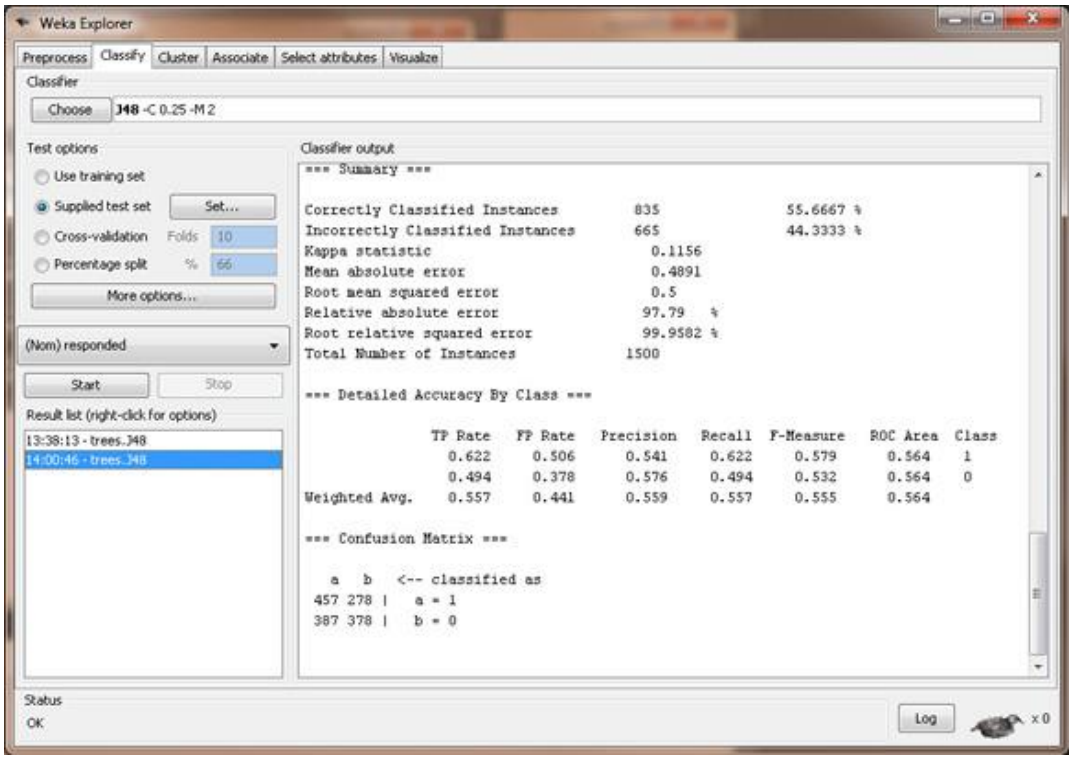


Рисунок 7 - Проверка классификационного дерева

Сравнивая показатель Correctly Classified Instances для тестового набора (55,7%) с этим же показателем для обучающего набора (59,1%), мы видим, что точность модели для двух разных наборов данных примерно одинакова. Это значит, что новые данные, которые будут использоваться в этой модели в будущем, не снизят точность ее работы.

Однако, поскольку собственно точность модели довольно низка (всего лишь 60% данных классифицировано верно), мы имеем полное право остановиться и сказать: «Она работает с точностью чуть выше 50%, с таким же успехом мы можем просто пытаться угадать значение случайным образом». Существуют случаи, когда использование алгоритмов интеллектуального анализа данных приводит к созданию неудачной аналитической модели.

Классификационная модель не подходит для анализа имеющихся у нас данных.

Метод ближайших соседей

Алгоритм метода ближайших соседей во многом схож с алгоритмом, используемым в методе кластеризации. Метод определяет расстояние между неизвестной точкой и всеми известными точками данных. Самый простой и наиболее распространенный способ определения расстояния – это нормализованное эвклидово расстояние.

Загрузим файл bmw-training.arff в WEKA, выполнив в закладке Preprocess.

Точно так же, как мы проделали это для методов регрессионного анализа и классификации в предыдущих статьях, мы должны открыть закладку Classify. В панели Classify нужно выбрать опцию lazy, а затем Ibk (здесь IB означает Instance-Based – обучение на примерах, а k указывает на количество соседей, поведение которых мы хотим исследовать). Убедитесь, что вы выбрали опцию Use training set, чтобы использовать набор данных, который мы только что загрузили в WEKA. Нажмите кнопку Start.

Результат вычислений IBk

```
=== Evaluation on training set ===
=== Summary ===

Correctly Classified Instances        2663               88.7667 %
Incorrectly Classified Instances       337               11.2333 %
Kappa statistic                          0.7748
Mean absolute error                      0.1326
Root mean squared error                  0.2573
Relative absolute error                 26.522  %
Root relative squared error             51.462  %
Total Number of Instances             3000
```

```
=== Detailed Accuracy By Class ===

               TP Rate   FP Rate   Precision   Recall  F-Measure   ROC Area  Class
                 0.95      0.177      0.847     0.95      0.896      0.972    1
                 0.823     0.05       0.941     0.823     0.878      0.972    0
Weighted Avg.    0.888     0.114      0.893     0.888     0.887      0.972

=== Confusion Matrix ===

    a    b   <-- classified as
 1449   76 |    a = 1
 261 1214 |    b = 0
```

У модели, использующей метод ближайших соседей, показатель точности равен 89% - совсем неплохо для начала, учитывая то, что точность предыдущей модели составляла всего 59%. Практически 90% точности – это вполне приемлемый уровень. Давайте рассмотрим результаты работы метода в терминах ложных определений, чтобы вы смогли на конкретном примере увидеть, как именно WEKA может использоваться для решения реальных вопросов бизнеса.

Результаты использования модели на нашем наборе данных показывают, что у нас есть 76 ложноположительных распознаваний (2.5%) и 261 ложноотрицательных распознаваний (8.7%). В нашем случае ложноположительное распознавание означает, что модель считает, что данный покупатель приобретет расширенную гарантию, хотя на самом деле он отказался от покупки. Ложноотрицательное распознавание, в свою очередь, означает, что согласно результатам анализа данный покупатель откажется от расширенной гарантии, а на самом деле он ее купил. Предположим, что стоимость каждой рекламной листовки, рассылаемой дилером, составляет $3, а покупка одной расширенной гарантии приносит ему 400$ дохода. Таким образом, ошибки ложного распознавание в терминах расходов и доходов нашего дилера будут выглядеть следующим образом: 400$ - (2.5% * $3) - (8.7% * 400) = $365. Следовательно, ложное распознавание ошибается в пользу дилера. Сравним этот показатель с данными модели классификации: $400 - (17.2% * $3) - (23.7% * $400) = $304. Как вы видите, использование более точной модели повышает потенциальный доход дилера на 20%.

Naive Bayes (наивный байесовский метод)

"Наивная" классификация - достаточно прозрачный и понятный метод классификации. "Наивной" она называется потому, что исходит из предположения о взаимной независимости признаков.

Свойства наивной классификации:

1. Использование всех переменных и определение всех зависимостей между ними.
2. Наличие двух предположений относительно переменных:
   - все переменные являются одинаково важными;
   - все переменные являются статистически независимыми, т.е. значение одной переменной ничего не говорит о значении другой.

Для использования этого метода в системе Weka панели Classify нужно выбрать опцию NativeBayes.

Результаты интерпретируются также как и при использовании других методов классификации.

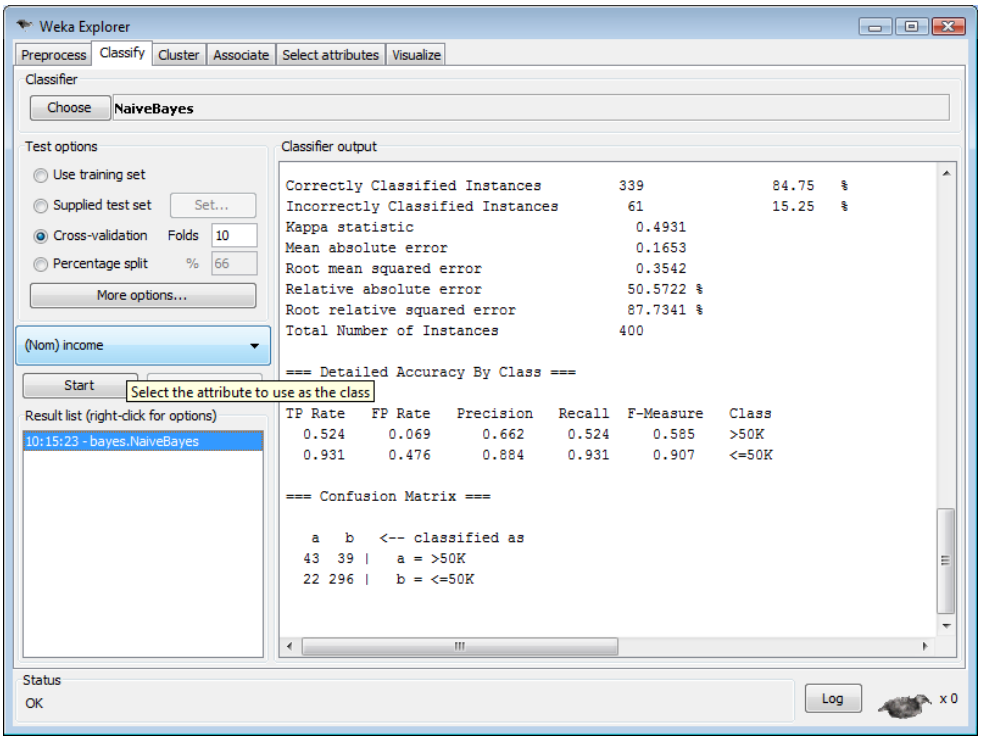


Рисунок 8 – Окно выбора метода

1R

Метод классификации 1R – один из самых простых и понятных методов классификации. Применяется как к числовым данным, которые разбиваются на промежутки, так и к данным типа nominal.

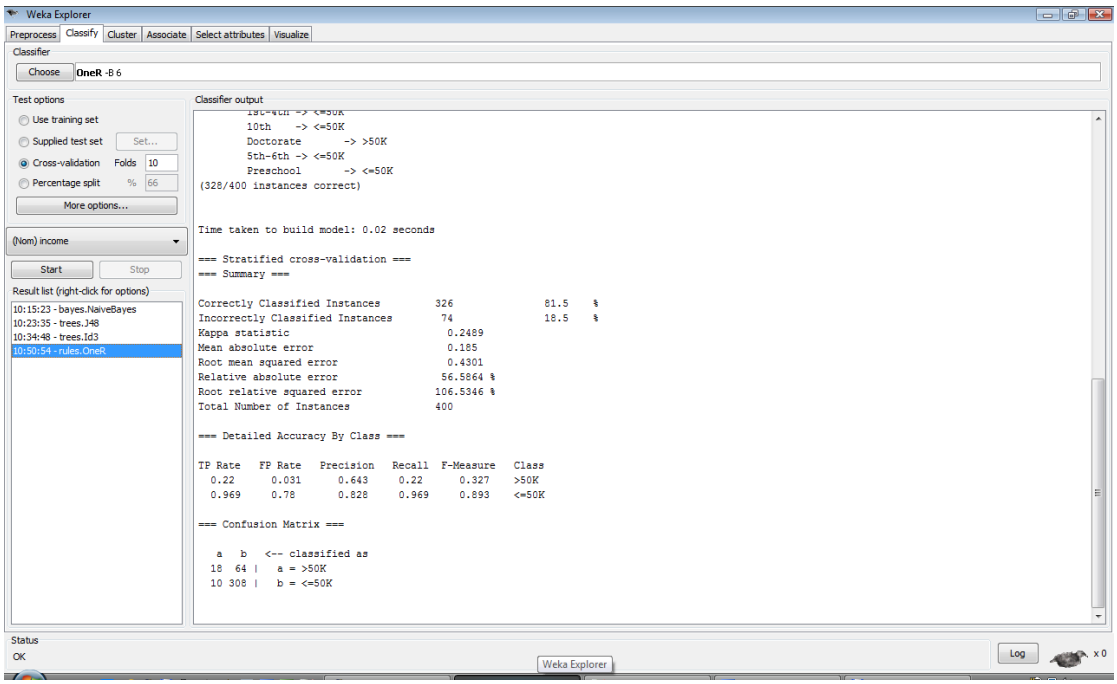


Рисунок 9

Для использования этого метода в системе Weka панели Classify нужно выбрать опцию OneR.

Результаты интерпретируются также как и при использовании других методов классификации.

SVM

Для этого метода не требуется каких-либо преобразований исходной выборки.

Данный метод является алгоритмом классификации с использованием математических функций. Метод использует нелинейные математические функции. Номинальные данные преобразуются в числовые. Основная идея метода опорных векторов – перевод исходных векторов в пространство более высокой размерности и поиск максимальной разделяющей гиперплоскости в этом пространстве.

Для использования этого метода в системе Weka панели Classify нужно выбрать опцию SMO.

Результаты интерпретируются также как и при использовании других методов классификации.

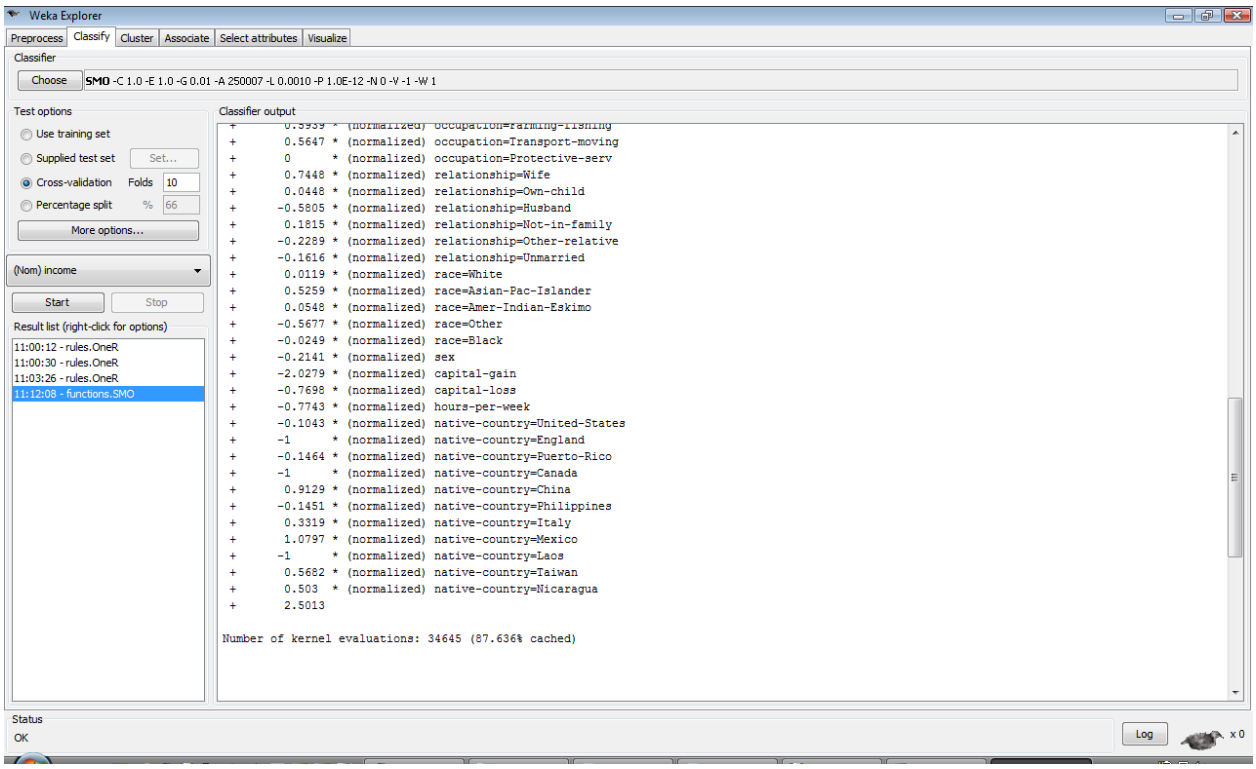


Рисунок 10

Кластеризация

Кластеризация позволяет разбить данные на группы, каждая из которых имеет определенные признаки. Метод кластерного анализа используется в тех случаях, когда необходимо выделить некоторые правила, взаимосвязи или тенденции в больших наборах данных. В зависимости от потребностей бизнеса, вы можете выделить несколько различных групп данных. Одно из явных преимуществ кластеризации по сравнению с классификацией состоит в том, что для разбиения множества на группы может использоваться любой атрибут (если вы помните, метод классификации использует только определенное подмножество атрибутов). В качестве основного недостатка метода кластеризации следует упомянуть тот факт, что составитель модели должен заранее решить, на сколько групп следует разбить данные. Для человека, который не имеет никакого представления о конкретном наборе данных, принять такое решение достаточно затруднительно. Следует ли нам создать три группы или пять групп? А может, нам нужно определить десять групп? Может потребоваться несколько итераций проб и ошибок, для того чтобы определить оптимальное количество кластеров.

Тем не менее, для среднестатистического пользователя кластеризация может оказаться наиболее полезным методом интеллектуального анализа данных. Этот метод позволит вам быстро разбить ваши данные на отдельные группы и сделать конкретные выводы и предположения относительно каждой группы. Математические методы,

реализующие кластерный анализ, довольно сложны и запутаны, так что в случае кластеризации мы будем целиком полагаться на вычислительные возможности WEKA.

Загрузите файл bmw-browsers.arff в WEKA, выполнив те же шаги, которые мы проделали ранее для открытия данных в закладке **Preprocess**.

Поскольку мы хотим разбить имеющиеся у нас данные на кластеры, вместо закладки **Classify** нам потребуется закладка Cluster. Нажмите на кнопку **Choose** и в предлагаемом меню выберите опцию SimpleKMeans.

Щелкните на опции SimpleKMeans (дизайн пользовательского интерфейса оставляет желать лучшего, но работать с ним можно). Единственный атрибут алгоритма, который нас интересует – это поле numClusters, которое указывает на количество кластеров для разбиения (напоминаем, что это значение вам нужно выбрать еще до создания модели). Изменим значение по умолчанию (2) на 5.

Результаты кластеризации

```
                           Cluster#
Attribute            Full Data      0          1          2          3          4
                       (100)       (26)       (27)        (5)       (14)       (28)
=============================================================================
Dealership              0.6      0.9615     0.6667          1     0.8571          0
Showroom               0.72      0.6923     0.6667          0     0.5714          1
ComputerSearch         0.43      0.6538          0          1     0.8571     0.3214
M5                     0.53      0.4615      0.963          1     0.7143          0
3Series                0.55      0.3846     0.4444        0.8     0.0714          1
Z4                     0.45      0.5385          0        0.8     0.5714     0.6786
Financing              0.61      0.4615     0.6296        0.8          1        0.5
Purchase               0.39           0     0.5185        0.4          1     0.3214



Clustered Instances

0       26 ( 26%)
1       27 ( 27%)
2        5 (  5%)
3       14 ( 14%)
4       28 ( 28%)
```

Данные кластеризации показывают, каким образом сформирован каждый кластер: значение «1» означает, что у всех данных в этом кластере соответствующий атрибут равен 1, а значение «0» означает, что у всех данных в этом кластере соответствующий атрибут равен 0. Данные соответствуют среднему значению атрибута на кластере. Каждый кластер

характеризует определенный тип поведения клиентов, таким образом, на основании нашего разбиения мы можем сделать некоторые полезные выводы:

- Кластер 0— эту группу посетителей можно было бы назвать «мечтатели». Они бродят вокруг дилерского центра, рассматривая машины, выставленные на внешней парковке, но никогда не заходят внутрь, и, хуже того, никогда ничего не покупают.
- Кластер 1— эту группу следовало бы назвать «поклонники M5», поскольку они сразу же подходят к выставленным автомобилям этой модели, полностью игнорируя BMW серии 3 или Z4. Тем не менее, эта группа не отличается высокими показателями покупки машин – всего 52%. Это потенциально может свидетельствовать о недостаточно продуманной стратегии продаж и о необходимости улучшить работу дилерского центра, например, за счет увеличения количества продавцов в секции M5.
- Кластер 2— эта группа настолько мала, что мы могли бы назвать ее выбраковкой. Дело в том, что данные этой группы статистически довольно разбросаны, и мы не можем сделать каких-либо определенных заключений относительно поведения посетителей, попавших в этот кластер (подобная ситуация может указывать на то, что вам следует сократить количество кластеров в модели)
- Кластер 3— эту группу следовало бы назвать «любимцы BMW», потому что посетители, попавшие в это кластер, всегда покупают машину и получают необходимое финансирование. Обратите внимание, данные этого кластера демонстрируют интересную модель поведения этих покупателей: сначала они осматривают выставленные на парковке машины, а затем обращаются к поисковой системе дилерского центра. Как правило, они покупают модели M5 или Z4, но никогда не берут модели третьей серии. Данные этого кластера указывают на то, что дилерскому центру следует активнее привлекать внимание к поисковым компьютерам (может быть, вынести их на внешнюю парковку), и кроме того, следует найти какой-нибудь способ выделить модели M5 и Z4 в результатах поиска, чтобы гарантированно обратить на них внимание посетителей. После того, как посетитель, попавший в этот кластер, выбрал определенную модель автомобиля, он гарантированно получает необходимый кредит и совершает покупку.
- Кластер 4— эту группу можно назвать «начинающие владельцы BMW», поскольку они всегда ищут модели 3 серии и никогда не интересуются более дорогими M5. Они сразу же проходят в демонстрационный зал, не тратя время на осмотр машин на внешней стоянке. Кроме того, они не пользуются поисковой системой центра. Примерно 50% этой группы получают одобрение по кредиту, тем не менее, покупку

совершают всего 32% участников. Анализируя данные этого кластера, можно сделать следующий вывод: посетители этой группы хотели бы купить свой первый BMW и точно знают, какая машина им нужна (модель 3 серии с минимальной конфигурацией). Однако, для того чтобы купить машину, им нужно получить положительное решение по кредиту. Чтобы повысить уровень продаж среди посетителей 4 кластера, дилерскому центру следовало бы понизить уровень требований для получения кредита или снизить цены на модели 3 серии.

Еще один интересный способ изучения результатов кластеризации – это визуальное представление данных. Щелкните правой кнопкой мышки в секции Result List закладки Cluster (и вновь элементы пользовательского интерфейса оставляют желать лучшего). В контекстном меню выберите опцию Visualize Cluster Assignments. В результате откроется окно с графическим представлением результатов кластеризации, настройки которого вы можете выбрать наиболее удобным для вас образом. Для нашего примера, измените настройку оси X так, чтобы она соответствовала количеству автомобилей M5 (M5 (Num)), а настройку оси Y – так, чтобы она показывала количество купленных автомобилей (Purchase (Num)), и укажите выделение каждого кластера отдельным цветом (для этого установите значение поля Color в Cluster (Nom)). Такие настройки помогут нам оценить распределение по кластерам в зависимости от того, сколько человек интересовалось BMW M5, и сколько человек купило эту модель. Кроме того, сдвиньте указатель Jitter примерно на три четверти в сторону максимума, это искусственным образом увеличит разброс между группами точек, чтобы вам было удобнее их просматривать.

Соответствует ли визуальное отображение кластеризации тем заключениям, которые мы сделали на основании данных в листинге 5? Как мы видим, в окрестности точки X=1, Y=1 (посетители, которые интересовались автомобилями модели M5 и купили их) расположены только два кластера: 1 и 3. Аналогично, в окрестности точки X=0, Y=0 расположены только два кластера: 4 и 0. Соответствует ли это нашим выводам? Да, соответствует. Кластеры 1 и 3 покупают BMW M5, в то время как кластер 0 не покупает ничего, а кластер 4 ищет BMW серии 3. На рисунке 8 показано визуальное отображение кластеров нашей модели. Мы предлагаем вам самостоятельно попрактиковаться в обнаружении других трендов и течений, меняя настройки осей X и Y.

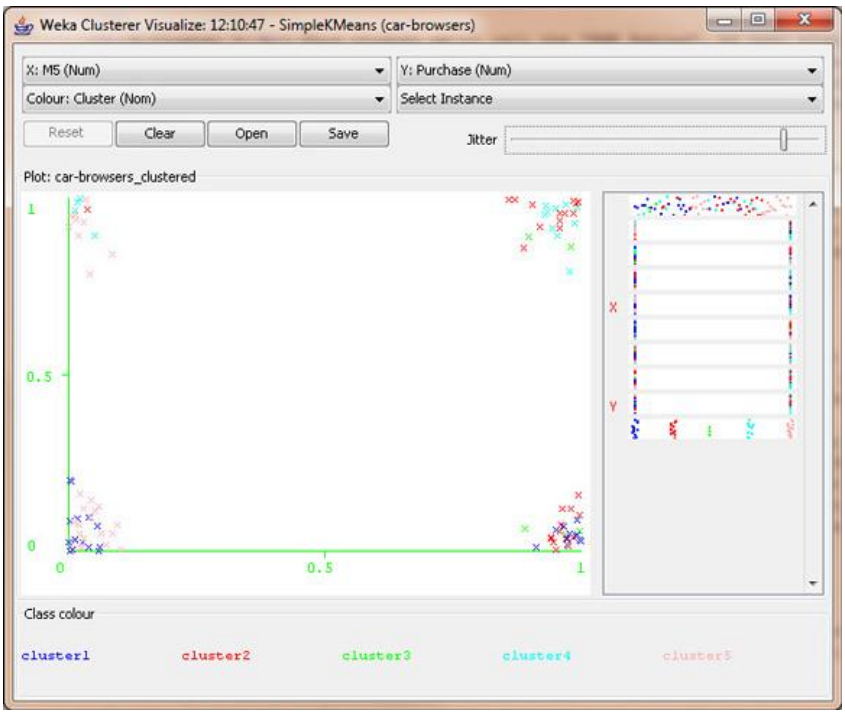


Рисунок 11 - Визуальное отображение кластеризации

Построение ассоциативных правил

Нахождение ассоциативных правил происходит почти так же, как и классификация. На вкладке Associate выбирается метод нахождения, для него выставляются параметры кликом на его названии, после чего нажимается кнопка Start и анализируется вывод (перед началом использования метода Априори необходимо применить фильтр RemoveType и удалить numeric-атрибуты). В нашем случае ассоциативные правила строятся по методу Априори.

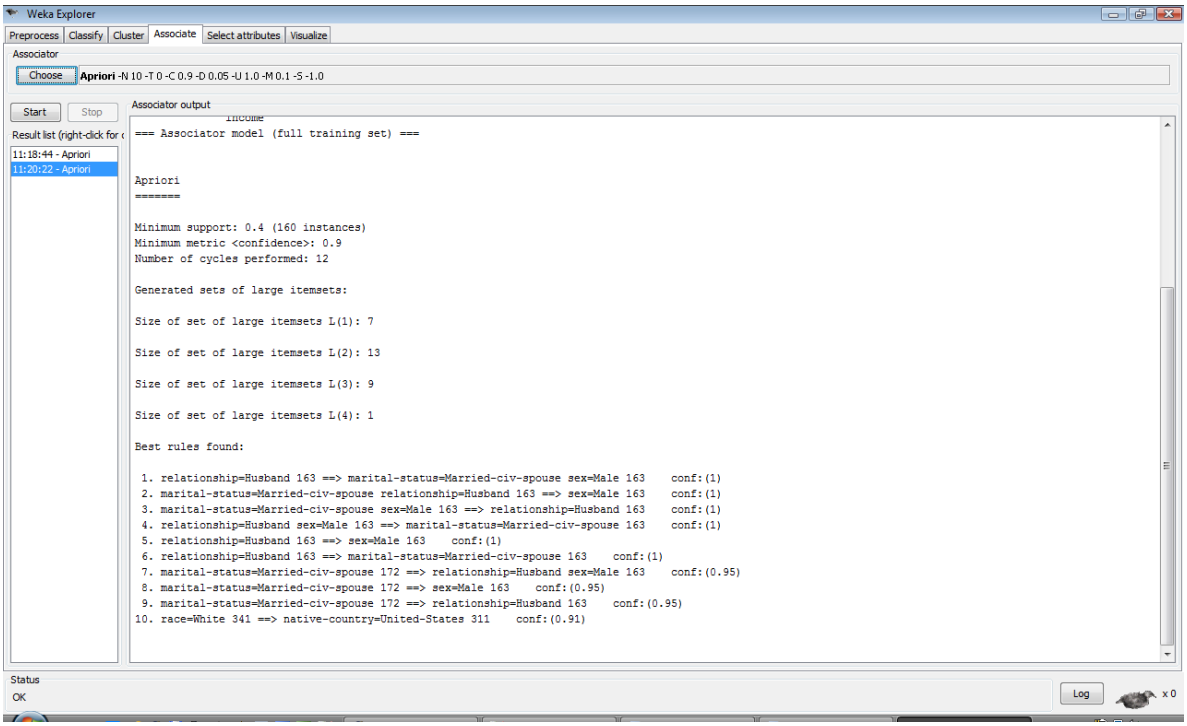


Рисунок 12

Варианты для выполнения лабораторной работы

Наборы данных можно брать из открытых источников (наборы не должны повторяться) или генерировать самостоятельно по вариантам:

1) Погодные условия в регионе
2) Продажа комплектующих изделий
3) Демографическая ситуация в регионе
4) Продажа земельных участков
5) Рынок труда
6) Больница
7) Железнодорожный транспорт
8) Авиационные перевозки
9) Олимпиада
10) Футбол
11) Туристический бизнес
12) Социальные сети
13) Интернет-провайдер
14) Здравоохранение
15) Автострахование
16) Кредитование
17) Экология
18) Правонарушения
19) Литература
20) Компьютеры

РЕКОМЕНДУЕМАЯ ЛИТЕРАТУРА И ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ

Список рекомендуемой литературы

основная

1) Федунец Н.И., Теория принятия решений [Электронный ресурс] : Учебное пособие для вузов / Федунец Н.И., Куприянов В.В. - М: Издательство Московского государственного горного университета, 2005. - 218 с. - ISBN 5-7418-0397-0 - Режим доступа: http://www.studentlibrary.ru/book/ISBN5741803970.html
2) Доррер Г.А., Методы и системы принятия решений [Электронный ресурс]: учеб. пособие / Доррер Г.А. - Красноярск : СФУ, 2016. - 210 с. - ISBN 978-5-7638-3489-5 - Режим доступа: http://www.studentlibrary.ru/book/ISBN9785763834895.html

дополнительная

3) Системы поддержки принятия решений : учебник и практикум для бакалавриата и магистратуры / В. Г. Халин [и др.] ; под редакцией В. Г. Халина, Г. В. Черновой. — Москва : Издательство Юрайт, 2019. — 494 с. — (Высшее образование). — ISBN 978-5-534-01419-8. — Текст : электронный // ЭБС Юрайт [сайт]. — URL: https://www.biblio-online.ru/bcode/432974
4) Федин, Ф. О. Анализ данных. Часть 1. Подготовка данных к анализу : учебное пособие / Ф. О. Федин, Ф. Ф. Федин. — М. : Московский городской педагогический университет, 2012. — 204 с. — ISBN 2227-8397. — Текст : электронный // Электронно-библиотечная система IPR BOOKS : [сайт]. — URL: http://www.iprbookshop.ru/26444.html
5) Федин, Ф. О. Анализ данных. Часть 2. Инструменты Data Mining : учебное пособие / Ф. О. Федин, Ф. Ф. Федин. — М. : Московский городской педагогический университет, 2012. — 308 с. — ISBN 2227-8397. — Текст : электронный // Электронно-библиотечная система IPR BOOKS : [сайт]. — URL: http://www.iprbookshop.ru/26445.html

учебно-методическая

6) Липатова Светлана Валерьевна. Системы принятия решений : учеб.-метод. пособие / Липатова Светлана Валерьевна; УлГУ, ФМИиАТ. - Ульяновск : УлГУ, 2016. – URL: ftp://10.2.96.134/Text/Lipatova2016.pdf

7) Воденин Дмитрий Ростиславович. Линейное программирование : учеб.-метод. пособие / Воденин Дмитрий Ростиславович; Ульяновск. гос. ун-т, Ин-т математики, физики и информ. технологий, Каф. прикл. математики. - Ульяновск : УлГУ, 2006. – URL: ftp://10.2.96.134/Text/vodenin1.pdf
8) Воденин Д. Р. Численные методы оптимизации : учеб.-метод. пособие / Д. Р. Воденин; УлГУ, ФМИиАТ. - Ульяновск : УлГУ, 2016. – URL: ftp://10.2.96.134/Text/Vodenin-2016.pdf

Профессиональные базы данных, информационно-справочные системы:

1. Электронно-библиотечные системы:

1.1. **IPRbooks** [Электронный ресурс]: электронно-библиотечная система / группа компаний Ай Пи Эр Медиа . - Электрон. дан. - Саратов, [2019]. - Режим доступа: http://www.iprbookshop.ru.

1.2. **ЮРАЙТ** [Электронный ресурс]: электронно-библиотечная система / ООО Электронное издательство ЮРАЙТ. - Электрон. дан. – Москва, [2019]. - Режим доступа: https://www.biblio-online.ru.

1.3. **Консультант студента** [Электронный ресурс]: электронно-библиотечная система / ООО Политехресурс. - Электрон. дан. – Москва, [2019]. - Режим доступа: http://www.studentlibrary.ru/pages/catalogue.html.

1.4. **Лань** [Электронный ресурс]: электронно-библиотечная система / ООО ЭБС Лань. - Электрон. дан. – С.-Петербург, [2019]. - Режим доступа: https://e.lanbook.com.

1.5. **Znanium.com** [Электронный ресурс]: электронно-библиотечная система / ООО Знаниум. - Электрон. дан. – Москва, [2019]. - Режим доступа: http://znanium.com.

2. **КонсультантПлюс** [Электронный ресурс]: справочная правовая система. /Компания «Консультант Плюс» - Электрон. дан. - Москва : КонсультантПлюс, [2019].

3**. База данных периодических изданий** [Электронный ресурс] : электронные журналы / ООО ИВИС. - Электрон. дан. - Москва, [2019]. - Режим доступа: https://dlib.eastview.com/browse/udb/12.

4. **Национальная электронная библиотека** [Электронный ресурс]: электронная библиотека. - Электрон. дан. – Москва, [2019]. - Режим доступа: https://нэб.рф.

5. **Электронная библиотека диссертаций РГБ** [Электронный ресурс]: электронная библиотека / ФГБУ РГБ. - Электрон. дан. – Москва, [2019]. - Режим доступа: https://dvs.rsl.ru.

6. Федеральные информационно-образовательные порталы:

6.1. Информационная система Единое окно доступа к образовательным ресурсам. Режим доступа: http://window.edu.ru

6.2. Федеральный портал Российское образование. Режим доступа: http://www.edu.ru

7. Образовательные ресурсы УлГУ:

7.1. Электронная библиотека УлГУ. Режим доступа : http://lib.ulsu.ru/MegaPro/Web

7.2. Образовательный портал УлГУ. Режим доступа : http://edu.ulsu.ru

Программное обеспечение

1. Редактор таблиц MS Excel.
2. Weka.
3. СУБД PostgreSQL (open source),
4. pgAdmin4 (open source).