МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ
Федеральное государственное автономное образовательное учреждение высшего образования
«СЕВЕРО-КАВКАЗСКИЙ ФЕДЕРАЛЬНЫЙ УНИВЕРСИТЕТ»
Институт сервиса, туризма и дизайна (филиал) СКФУ в г. Пятигорске

Методические рекомендации для студентов по организации самостоятельной работы по дисциплине «Технологии обработки информации»

Направление подготовки 09.03.02 Информационныесистемы и технологии
Направленность (профиль) Информационныесистемы и технологии
Квалификация (степень) выпускника
бакалавр

Пятигорск 2020

ВВЕДЕНИЕ

Целью изучения дисциплины «Технологии обработки информации» является формирование набора общепрофессиональных и профессиональных компетенций будущего бакалавра по направлению подготовки 09.03.02 «Информационные системы и технологии».

Задачи дисциплины: ознакомить обучающихся с основными видами и процедурами обработки информации, моделями и методами решения задач обработки (обработка числовых массивов данных, обработка экономической информации, обработки аудио, видеоинформации), обучить методам и средствам информационных технологий обработки числовых массивов данных, обработки экономической информации, сформировать умения и практические навыки эффективного использования программных средств обработки информации в профессиональной деятельности.

Предназначено для студентов вузов, обучающихся по направлению подготовки 09.03.02«Информационные системы и технологии».

Содержание данной учебной дисциплины опирается на знание дисциплин: Технологии программирования, Методы и средства проектирования информационных систем и технологий.

Знания, полученные при изучении данной дисциплины, необходимы для успешного освоения такой дисциплины, как Инструментальные средства информационных систем, Базы данных в распределенных системах обработки информации.

1. Организационно-методические рекомендации по освоению дисциплины

Самостоятельная работа студентов являетсяважнейшим условием формирования научного способа познания. Она проводится накануне каждого семинарского (практического, лабораторного) занятия и включает подготовку к выполнению лабораторной работы или подготовку к выступлению на практическом занятии.

Самостоятельные занятия (СЗ) являются одной из активных форм обучения.

Самостоятельные занятия по дисциплине «Технологии обработки информации» имеют целью:

- закрепить и углубить знания, полученные студентами на лекциях и в процессе лабораторных занятий;
- привить практические навыки при обработке различных видов информации.

Самостоятельные занятия проводятся в специализированных аудиториях, оборудованных СВТ и возможностью пользования Интернет-ресурсами, в библиотеке.

Предлагаемые методические рекомендации содержат информацию для студентов, необходимую при подготовке к проведению лабораторных занятий и практических занятий по дисциплине «Технологии обработки информации».

2. Технологическая карта самостоятельной работы обучающегося

Код реализуемой компетенции	Вид деятельности студентов	Итоговый продукт самостоятельной работы	Средства и технологии оценки*	Обьем часов		
				СРС	Контактная работа с преподавателем	Всего
ОПК-1, ОПК-6, ПК-22	Самостоятельное изучение литературы	Конспект	Собеседование	14,58	1,52	16,5

ОПК-1, ОПК-6, ПК-22	Проработка лекционного материала	Конспект	Собеседование	2,43	0,27	2,7
ОПК-1, ОПК-6, ПК-22	Подготовка к лабораторным занятиям	индивидуальное задание	отчет письменный	7,29	0,91	8,1
			Итого	24,3	2,7	27

3. Организация контроля знаний студентов

Формы контроля знаний студентов Контроль и оценка знаний, умений и навыков студентов осуществляется на лабораторных работах, консультациях, в ходе сдачи экзамена. В ходе контроля знаний преподаватель оценивает понимание студентом содержания дисциплины «Технологии обработки информации», его способность понимать фундаментальные технологий обработки информации и освоить практические навыки различных способов обработки информации.

Контроль знаний студентов может осуществляться в следующих формах:

- текущий контроль знаний;
- итоговый контроль знаний.

Текущий контроль знаний студентов имеет целью:

- дать оценку работы каждого студента по усвоению им учебного материала, выявить недостатки в его подготовке и оказать практическую помощь в их устранении;

Основными формами текущего контроля знаний студентов являются:

–устный контрольный опрос;

– письменный контрольный блиц-опрос; – защита лабораторной работы;

– выступление с докладом на практических занятиях;

–проверка конспектов лекций

Устный контрольный опрос студентов проводится на лекциях (и лабораторных занятиях). По его результатам преподаватель оценивает качество подготовки студента к занятию.

Письменный контрольный блиц-опрос студентов проводится в течение пяти минут на практических занятиях путем письменного ответа их на пять вопросов, заданных преподавателем. Результаты его проведения отмечаются в журнале. На лабораторных занятиях знания и практические навыки студентов оцениваются по 4-балльной системе. Полученные оценки выставляется в журнале.

При проверке конспектов лекций дается анализ качества их ведения. Отмечаются допущенные ошибки, врецензии преподавателя оценивается качество конспектирования учебного материала, даются рекомендации по улучшению качества конспектирования лекционного материала.

Итоговый контроль знаний осуществляется в форме экзамена.

Контрольные вопросы к экзамену:

Знать

1. Данные, информация, знания.
2. Различные подходы к определению понятия «информация».
3. Свойства и виды информации. Что такое информационные ресурсы?
4. Основные процессы сбора, накопления и преобразования информации.
5. Методы представления информации.

6. Единицы измерения информации в компьютерных системах: двоичная система исчисления, биты и байты. Методы представления информации.
7. Архивирование и сжатие информации. Виды архиваторов.
8. Понятие информационных технологий. Применение информационных технологий на практике.
9. Понятие информационного общества. Основные признаки и тенденции развития.
10. Технологии подготовки текстовых документов в текстовом процессоре MS Word. Функциональные возможности MS Word.
11. Технологии обработки информации в электронных таблицах MS Excel. Функциональные возможности MS Excel.
12. Технологии работы с мультимедийными данными. Разработка презентаций в MS Power Point.
13. Базы данных. Технология работы с базами данных.
14. Разработка баз данных. Модель ANSI/SPARC.
15. Индексирование; связывание таблиц. Потенциальные ключи. Внешние ключи.
16. Распределенные базы данных.
17. Язык SQL-3. Транзакции, триггеры и встроенные функции.
18. Графические редакторы. Разновидности, сферы использования.
19. Файловые системы. Системы FAT32 и NTFS и их особенности
20. Что такое архитектура и структура компьютера. Опишите принципы фон Неймана и «открытой архитектуры».

Уметь,

Владеть

1. Функциональная схема компьютера. Основные устройства компьютера, их назначение и взаимосвязь.
2. Виды и назначение устройств ввода и вывода информации.
3. Память компьютера – типы, виды, назначение. Внешняя память компьютера. Различные виды носителей информации, их характеристики (информационная емкость, быстродействие и т.д.).
4. Что такое BIOS и какова его роль в первоначальной загрузке компьютера? Каково назначение контроллера и адаптера.
5. Приведите основные описательные характеристики компьютера (характеристика процессора, объем оперативной и внешней памяти, мультимедийные и сетевые возможности, периферийные и другие составляющие).
6. Аппаратно-программное обеспечение компьютерной сети: основные устройства.
7. Методы интеллектуального анализа данных. Data Mining и KDD (Knowledge Discovery in Databases).
8. Система Deductor, ее предназначение, достоинства и ограничения.
9. Возможности и основная схема работы Deductor.
10. Понятие сценария и сценарии в Deductor.
11. Задачи решаемы в Deductor: очистка данных; корреляция и регрессия; кластеризация и классификация; прогнозирование; визуализация и др.
12. Опишите технологию «клиент-сервер». Приведите принципы многопользовательской работы с программным обеспечением.
13. Программное обеспечение компьютера, его классификация и назначение.
14. Что такое файловая система? Папки и файлы. Основные операции с файлами в операционной системе. Файловые системы NTFS и FAT – отличия в обеспечении надежности работы системы и безопасного хранения информации.
15. Понятие компьютерной сети. Виды компьютерных сетей.

16. Возможности глобальной сети Интернет. Базовые информационные ресурсы и ресурсы Интернета. Применение компьютерных сетей для обмена данными.
17. Топология и разновидности компьютерных сетей. Локальные и глобальные сети.
18. Сервисы и ресурсы Internet.
19. Что такое World Wide Web (WWW). Понятие гипертекста. Документы Internet.
20. Методы и средства поиска информации в компьютерной сети.

4. Рекомендации по работе с литературой и источниками

4.1. Рекомендации студентам по организации работы с литературой и источниками

Изучение литературы и источников необходимо начинать с прочтения соответствующих глав учебных изданий, учебных пособий или литературы, рекомендованной в качестве основной или дополнительной по дисциплине «Физические основы записи и хранения информации», которые прямо или косвенно относятся к изучаемой теме.

Изучая литературу и источники, студенту рекомендуется вести краткий конспект. Однако не следует переписывать все содержание изучаемой темы, нужно выписывать лишь основные идеи и главные на ваш взгляд мысли. В отдельных случаях, когда встречаются важные определения, понятия, необходимый фактический материал и примеры, статистическая информация, имеющие отношение к изучаемой теме, студенту следует выписать их в виде цитат с полным указанием библиографических источников.

Конспектирование рекомендуемой литературы и источников необходимо вести с распределением собранных материалов по отдельным главам и параграфам согласно учебно-тематическому плану. Необходимо выписывать все выходные данные по используемой литературе и источникам.

Важным этапом при работе с рекомендуемойлитературой и источниками является изучение законодательных и нормативных актов федерального, регионального, местного и ведомственного уровней. При изучении Указов Президента РФ, Законов и Кодексов РФ, постановлений, положений, рекомендаций и т.д., студент должен выяснить все изменения и дополнения, которые могли быть внесены после их выхода в свет.

Основой технологии интенсификации обучения на платформе цифровых образовательных технологий являются учебно-иллюстрационные материалы (опорный конспект) по дисциплине «Технологи обработки информации».

Работа с учебно-иллюстрационными материалами имеет следующие этапы.

1. Изучение теоретических основ учебного материала в аудитории: изложение преподавателем изучаемого материала студентам с объяснением по опорному конспекту;
2. Самостоятельная работа: индивидуальная работа студентов по опорному конспекту; фронтальное закрепление по блокам опорного конспекта.
3. Первое повторение - воспроизведение содержания заданной темы опорного конспекта по памяти.
4. Устное проговаривание материала опорногоконспекта - необходимый этап внешнеречевой деятельности при усвоении учебного материала.
5. Второе повторение - взаимоопрос и взаимопомощь студентов друг другу.

Применение учебно-иллюстрационных материалов позволяет обобщить сложный по содержанию материал, активизировать мыслительную деятельность студентов.

Необходимо помнить, что главное для студента в самостоятельной работе с рекомендуемой литературой и источниками - это формирование своего индивидуального стиля, который может стать основой в будущей профессиональной деятельности.

4.2. Рекомендации студентам по подготовке к выполнению лабораторных работ

1. Цель работы:

Освоение методов интеллектуального анализа данных, в частности: классификации объектов БД методом построения дерева решений; выявление связей между объектами БД методом ассоциаций.

2. Задачи:

2.1.На основании модели предметной области, разработанной в курсовом проекте по курсу ТПР, из перечня объектов – классов выбрать 1-2 наиболее важных, для которых могут существовать варианты (экземпляры классов). Определить для этих объектов – классов не менее 3-х свойств, на основе которых экземпляры классов могут быть разбиты на подклассы. Например, по содержанию предметной области, необходимо арендовать помещение (под офис, под магазин и т.п.). Выбор осуществляется на основе следующих свойств: цена за кв.м.; площадь, качество отделки помещения (высокое, среднее, низкое); расстояние от метро (0 мин., до 10 минут, до 20 минут) и т.п. На основе этих параметров сформировать прототип реляционной Базы Данных (РБД) из 15 записей с описанием конкретных помещений. Задача состоит в том, чтобы разбить имеющиеся варианты на 3 класса (например: евро-класс (1), бизнес – класс (2), эконом – класс (3)).

2.2.На основе той же модели предметной области сформировать прототип РБД из 15-ти транзакций, т.е. последовательности записей типа: (письменный стол, кресло, компьютер, офис), (обеденный стол, меню, кафе), (кресло, стол, настольная лампа, кабинет) и т.п.

2.3. Решить задачу классификации записей БД (п.2.1.) методом построения дерева решений. Обосновать порядок применения свойств для классификации объектов, выявить взаимосвязь данных, разработать правила классификации.

2.4. Выявить наиболее сильные взаимосвязи между элементами транзакций (гипотезами правил типа «если, то….».)на основе вычисления значений достоверности, поддержки, лифта, левереджа, улучшения для разработанных гипотез.

3. Методы решения.

3.1. Для решения задачи 2.3 применить метод построения дерева решений, а для определения наиболее информативного порядка применения свойств для классификации применить метод прироста информации (метод энтропии).

3.2. Для решения задачи 2.4 применить метод АССОЦИАЦИИ.

4.Программные средства.

Решение перечисленных задач выполнить на основе средств аналитической платформы DeductorStudio.

5 Порядок решения задачи:

5.1. Создать два прототипа РБД (15 записей).

5.2. Наполнить Хранилище Данных первичной информацией из прототипов РБД.

5.3. Провести анализ данных и их классификацию на основе метода дерева решений. Результат представить в виде дерева.

5.4. Сформировать множество гипотез правил на основании метода ассоциаций, найти наиболее сильное и наиболее слабое правила и доказать это расчетами. Расчеты лифта, левереджа осуществить вручную. Формирование заключения о достоверности и поддержке гипотезы осуществить средствами DeductorStudio.

6. Представление результатов:

В отчет о результатах лабораторной работы входят:

6.1. По дереву решений:

- Прототип РБД;
- Имена классов для разбиения объектов РБД на классы.
- Расчеты по выбору порядка применения свойств для классификации объектов (видеоформы).
- Дерево решений, классы объектов, классификационные правила (видеоформы и вручную).

6.2. По методу ассоциации:

- Прототип РБД;
- наиболее вероятное ассоциативное правило и расчеты его достоверности, поддержки, лифта, левереджа и улучшения (видеоформы и вручную);
- Результаты сравнения расчетов по ассоциативному правилу и обоснованный вывод о наиболее значимом ассоциативном правиле.

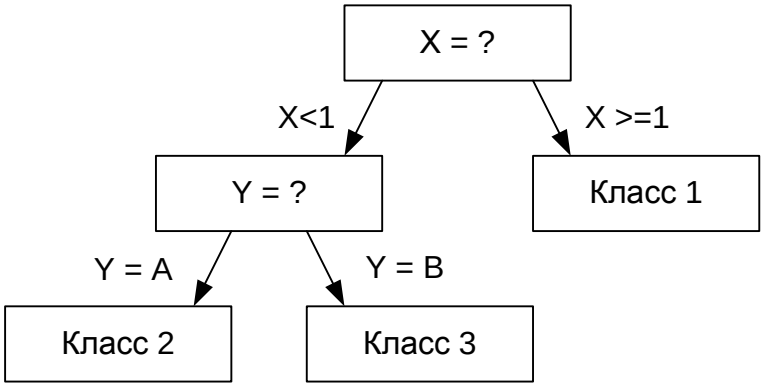
7. Приложение к методике выполнения лабораторной работы.

Классификацию данных можно рассматривать как процесс, состоящий из двух этапов. На первом этапе строится модель, описывающая предварительно определенный набор классов или категорий. Модель строится на основе анализа данных, содержащих признаки(атрибуты) объектов и соответствующую им метку класса. Такой набор называется обучающей выборкой. В контексте классификации записи могут упоминаться как наблюдения, примеры, прецеденты или объекты.

Поскольку метка класса каждого примера предварительно задана, построение классификационной модели часто называют обучением с учителем. В процессе обучения формируются правила, по которым производится оттеснение объектов к одному из классов.

На втором этапе модель применяется для классификации новых, ранее не известных объектов и наблюдений. Перед этим оценивается точность построенной классификационной модели.

Деревья решений (деревья классификаций) – классификационная техника, в ходе которой решающие правила извлекаются непосредственно из исходных данных в процессе обучения. Дерево решений – это иерархическая модель, где в каждом узле производится проверка определенного атрибута(признака) с помощью правила. Каждая выходящая из узла ветвь есть результат проверки, она содержит объекты, для которых значения данного атрибута удовлетворяют правилу в узле. Каждый конечный узел дерева(лист) содержит объекты, относящиеся к одному классу.



Пример дерева решений

Классический алгоритм построения деревьев решений использует стратегию «разделяй и властвуй». Начиная с корневого узла, где присутствуют все обучающие примеры, происходит их разделение на два подмножества или более на основе значений атрибута, выбранных в соответствии с критерием (правилом) разделения. Для каждого подмножества создается дочерний узел, с которым оно ассоциируется. Затем процесс ветвления повторяется для каждого дочернего узла до тех пор, пока не будет выполнено одно из условий остановки алгоритма, что служит упрощению дерева. Упрощение дерева заключается в том, что после его построения удаляется те ветви, правила в которых имеют низкую ценность, поскольку относятся к небольшому числу примеров.

Мерой оценки возможного разбиения является так называемая *чистота*, под которой понимается отсутствие *примесей*. Низкая чистота означает, что в подмножестве представлены объекты, относящиеся к различным классам. Высокая чистота свидетельствует о том, что члены отдельного класса доминируют. Наилучшим разбиением можно назвать то, которое дает наибольшее увеличение чистоты дочерних узлов относительно родительского. Кроме того, хорошее разбиении должно создавать узлы примерно одинакового размера или как минимум не создавать узлы, содержащие всего несколько записей.

Рассмотрим пример Дерева решений для следующей модели базы данных различных помещений:

Помещение	Площадь	Качество отделки	Близость к метро	Метка класса
1	Большая	Высокое	Близко	1(ЭЛИТ)
2	Большая	Среднее	Близко	2(БИЗНЕС)
3	Малая	Низкое	Далеко	3(ЭКОНОМ)
4	Малая	Высокое	Близко	1(ЭЛИТ)
5	Средняя	Среднее	Очень Далеко	3(ЭКОНОМ)
6	Большая	Высокое	Близко	1(ЭЛИТ)
7	Малая	Низкое	Очень Далеко	3(ЭКОНОМ)
8	Средняя	Среднее	Близко	2(БИЗНЕС)
9	Малая	Низкое	Очень Далеко	3(ЭКОНОМ)
10	Средняя	Высокое	Близко	1(ЭЛИТ)

1. Выбираем критерий наилучшего ветвления, то есть атрибут который даст наиболее чистое разбиение на классы. Для этого просчитаем возможные варианты разбиения на классы:

При разбиении по атрибуту ***Площадь*** мы будем иметь:

Большая площадь: 2 объекта 1го класса и 1 объект 2го класса.

Средняя площадь: 1 объект 1го класса, 1 объект 2го класса и 1 объект 3го класса.

Малая площадь: 3 объекта 3го класса и 1 объект 1го класса.

Как можно видеть, чистых классов при данном разбиении нет.

При разбиении по атрибуту ***Качество отделки*** мы будем иметь:

Высокое: 4 объекта 1го класса

Среднее: 2 объекта 2го класса и 1 объект 3го класса

Низкое: 3 объекта 3го класса

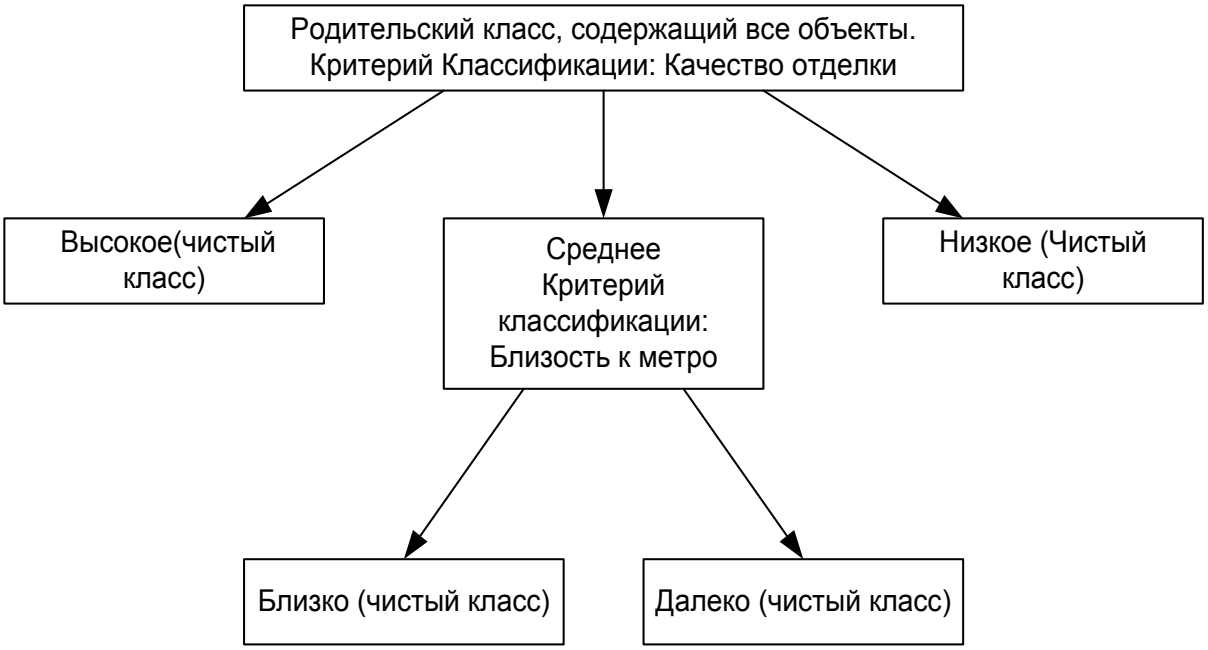
При разбиении по атрибуту Качество отделки мы можем видеть появление двух чистых классов(Высокое и низкое).

Аналогичным образом проверяются третий и любой последующий параметры, если они существуют.

В данном случае очевидно, что параметр Качество отделки дает наиболее чистое разбиение.

2. В случае если первое разбиение дает хотя бы один не чистый класс, для этого класса производится дополнительное разбиение по одному из оставшихся атрибутов, выбор которых производится аналогично. В нашем случае этим параметром будет ***Близость к метро.***

Таким образом итоговое дерево будет выглядеть следующим образом:



Работа с DeductorStudio.

1. Импорт данных для анализа.

Для анализа данных необходимо импортировать данные из текстового файла, содержащего набор необходимых транзакций. Для каждого задания по лабораторной работе данный файл создается исполнителем, в соответствии с выбранной темой. Подключение к файлу происходит путем следования указаниям диалоговых окон программы (на сайте разработчика можно ознакомиться с видеоматериалами подробнейшим образом описывающими необходимый набор действий). В ходе подключения к файлу указывается форматы анализируемых данных, и символы их разделения. В качестве примера используется файл pomesheniya.txt, в котором содержится таблица следующего вида:

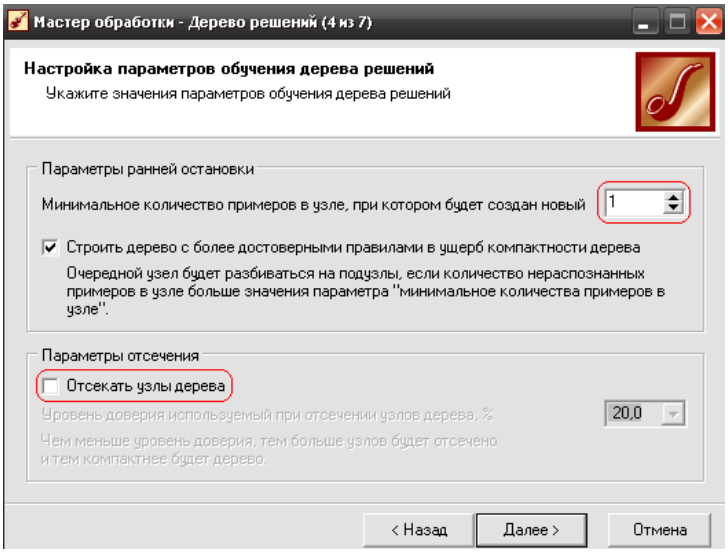
```
Помещение       Площадь КачествоОтделки БлизостьКМетро  МеткаКласса
1       большая высокое близко  ЭЛИТ
2       большая среднее близко  БИЗНЕС
3       малая   низкое  далеко  ЭКОНОМ
4       малая   высокое близко  ЭЛИТ
5       средняя среднее оченьдалеко     ЭКОНОМ
6       большая высокое близко  ЭЛИТ
7       малая   низкое  оченьдалеко     ЭКОНОМ
8       средняя среднее близко  БИЗНЕС
9       малая   низкое  оченьдалеко     ЭКОНОМ
10      средняя высокое близко  ЭЛИТ
```

В данной таблице «Помещение» является номером объектом классификации, а «Площадь», «Качество отделки» и «близость к метро» критериями

классификации(атрибутами). Как объект классификации, так и критерии классификации являются дискретными данными. При описании столбцов в ходе импорта из файла необходимо указывать соответствующие параметры для данных.

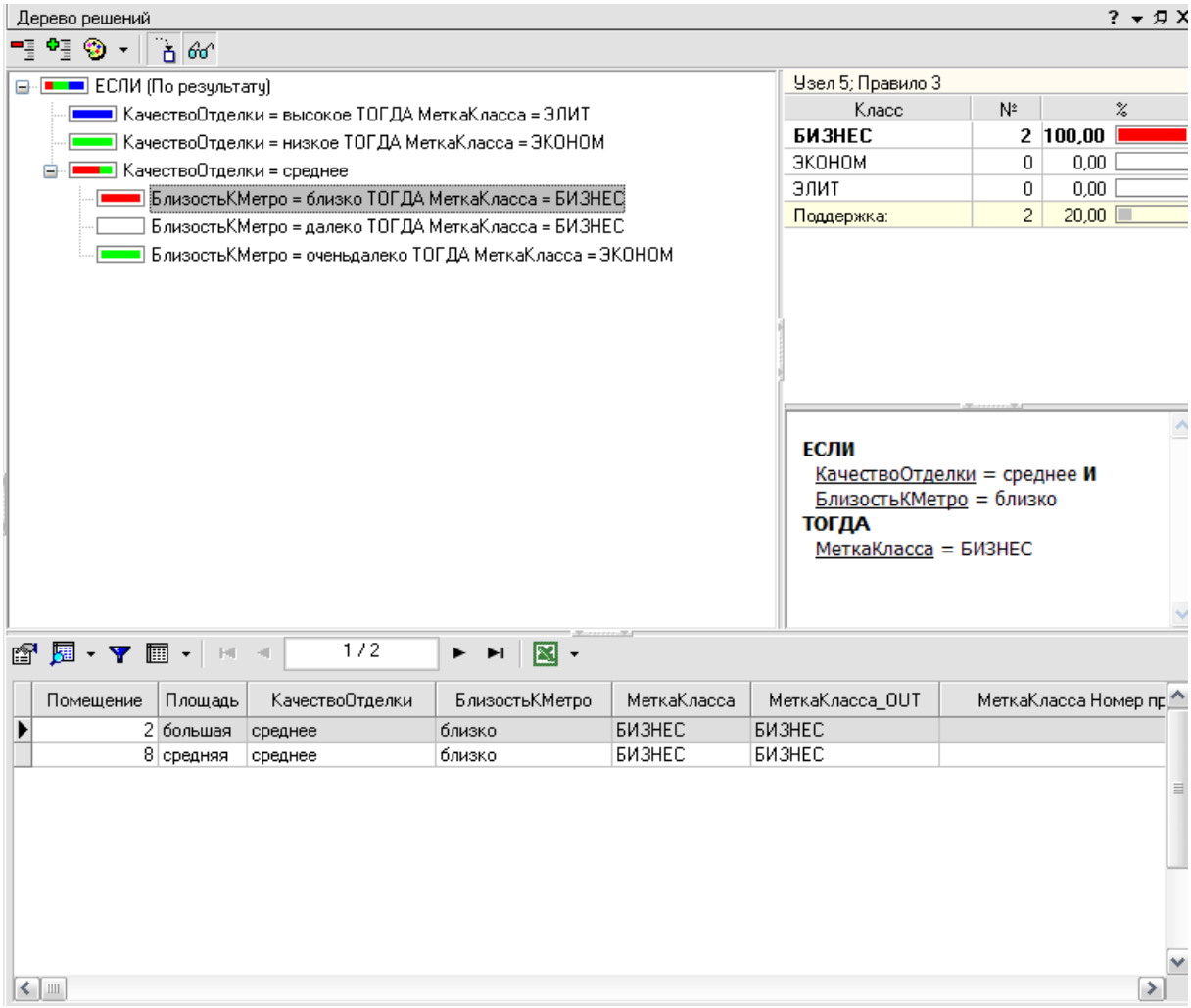
2. Анализ данных

Нажатием клавиши F7 вызывается Мастер обработки, который позволит проанализировать данные и для классификации. При настройке параметров построения **дерева решений** необходимо установить параметр *«Минимальное количество примеров в узле»* равным 1, и отменить выполнение параметра «отсекать узлы дерева». Это позволит добиваться максимальной чистоты получаемых классов. Ввиду крайне малого объема составляемой модели базы данных для исследования в ходе лабораторной работы новый класс должен создаваться до тех пор, пока не будет достигнута идеальная чистота класса.



3. Подготовка отчета по работе с DeductorStudio

Версия программы является обучающей, что обуславливает ряд функциональных ограничений по ее использованию. В связи с этим результаты исследования в DeductorStudio, необходимо сохранить в формате рисунка, нажатием клавиши **«*PrtSc*».**



Ассоциативные правила описывают связь между наборами предметов, соответствующими условию и следствию. Эта связь характеризуется двумя показателями— поддержкой (support) и достоверностью (confidence).

Обозначим базу данных транзакций как **D** а число транзакций в этой базе как **N**. Каждая транзакция **D** представляет собой некоторый набор предметов. Обозначим через 5 поддержку, через **C** — достоверность.

Поддержка ассоциативного правила — это число транзакций, которые содержат как условие, так и следствие. Например, для ассоциации **A->B**можно записать:

$$S(A \to B) = \frac{\text{Кол-во транзакций, содержащих A и B}}{\text{общее кол-во транзакций N}}$$

Достоверность ассоциативного правила **A->B**представляет собой меру точности правила и определяется как отношение количества транзакций, содержащих и условие, и следствие, к количеству транзакций, содержащих только условие:

C(A->B)=<u>количество транзакций, содержащих A и B</u>
количество транзакций, содержащих только A

Если поддержка и достоверность достаточно высоки, можно с большой вероятностью утверждать, что любая будущая транзакция, которая включает условие, будет также содержать и следствие.

Возьмем ассоциацию салат ->помидоры. Пусть количество транзакций, содержащих как салат, так и помидоры, равно 40, а общее число транзакций — 100, тогда поддержка данной ассоциации будет:

S(салат ->помидоры) = 4 / 10 = 0,4.

Пусть количество транзакций, содержащих только салат (условие), равно 40, то достоверность данной ассоциации будет:

C(салат ->помидоры) = 4 0 / 4 0 = 1.

Иными словами, все наблюдения, содержащие салат, также содержат и помидоры, из чего делаем вывод о том, что данная ассоциация может рассматриваться как правило.

Методики поиска ассоциативных правил обнаруживают все ассоциации, которые удовлетворяют ограничениям на поддержку и достоверность, наложенным пользователем. Это приводит к необходимости рассматривать десятки и сотни тысяч ассоциаций, что делает невозможным обработку такого количества данных вручную. Число правил желательно уменьшить таким образом, чтобы проанализировать только наиболее значимые из них.

Однако если условие и следствие независимы, то правило вряд ли представляет интерес независимо от того, насколько высоки его поддержка и достоверность. Например, если статистика дорожно-транспортных происшествий показывает, что из 100 аварий в 80 участвуют автомобили марки ВАЗ, то на первый взгляд это выглядит как правило «если *авария,* то *ВАЗ».* Но если учесть, что парк автомобилей ВАЗ составляет, 80 % от общего числа легковых автомобилей, то такое правило вряд ли можно назвать значимым.

По этой причине при поиске ассоциативных правил используются дополнительные показатели, позволяющие оценить значимость правила. Можно выделить объективные и субъективные меры значимости правил. Объективными являются такие меры, как поддержка й достоверность, которые могут применяться независимо от конкретного приложения. Субъективные меры связаны со специальной информацией, определяемой пользователем в контексте решаемой задачи. Такими субъективными мерами являются лифт (lift) и левередж (от англ. leverage — «плечо», «рычаг»).

Лифт (оригинальное название — интерес) вычисляется следующим образом:

L(A-+B) = C(A-+B) /S(B).

Лифт — это отношение частоты появления условия в транзакциях, которые также содержат и следствие, к частоте появления следствия в целом. Значения лифта большие, чем 1, показывают, что условие чаще появляется в транзакциях, содержащих следствие, чем в остальных. Можно сказать, что лифт является обобщенной мерой связи двух предметных наборов: при значениях лифта больше 1 связь положительная, при 1 она отсутствует, а при значениях меньше 1 — отрицательная.

Хотя лифт используется широко, он не всегда оказывается удачной мерой значимости правила. Правило с меньшей поддержкой и большим лифтом может быть менее значимым, чем альтернативное правило с большей поддержкой и меньшим лифтом, потому что последнее применяется для большего числа покупателей. Значит, увеличение числа покупателей приводит к возрастанию связи между условием и следствием.

Другой мерой значимости правила является левередж:

T(A->B) = S(A ->B) - S(A)*S(B).

Левередж— это разность между наблюдаемой частотой, с которой условие и следствие появляются совместно (то есть поддержкой ассоциации), и произведением частот появления (поддержек) условия и следствия по отдельности.

Улучшение(improvement) и вычисляется подобно левередж, только берется не разность, а отношение наблюдаемой частоты и частот появления по отдельности:

I(A->B)= *S(A->B) / S(A)*S(B)*

Улучшение показывает, полезнее ли правило случайного угадывания. Если *I(A* -> Л) > 1, это значит, что вероятнее предсказать наличие набора *B* с помощью правила, чем угадать случайно.

При практической реализации систем поиска ассоциативных правил используют различные методы, которые позволяют снизить пространство поиска до размеров, обеспечивающих приемлемые вычислительные и временные затраты, например *алгоритм apriori*

В основе алгоритма apriori лежит понятие *популярных наборов* (frequentitemset), которые также можно назвать частыми предметными наборами, часто встречающимися множествами (соответственно, они связаны с понятием частоты). Под частотойпонимается простое количество транзакций, в которых содержится данный предметный набор. Тогда популярными наборами будут те из них, которые встречаются чаще, чем в заданном числе транзакций.

Популярный предметный набор — предметный набор с поддержкой больше заданного порога либо равной ему. Этот порог называется минимальной поддержкой.

Методика поиска ассоциативных правил с использованием популярных наборов состоит из двух шагов.

1. Нахождение популярных наборов.

2. На их основе необходимо сгенерировать ассоциативные правила, удовлетворяющие условиям минимальной поддержки и достоверности.

Чтобы сократить пространство поиска ассоциативных правил, алгоритм apriorіиспользует свойство антимонотонности. Свойство утверждает, что если предметный набор **Z** не является частым, то добавление некоторого нового предмета **A**к набору **Z**не делает его более частым. Другими словами, если Zне является популярным набором, то и набор **Z**U**A**также не будет являться таковым. Данное полезное свойство позволяет значительно уменьшить пространство поиска ассоциативных правил.

Работа с DeductorStudio.

1.Импорт данных для анализа.

Для анализа данных необходимо импортировать данные из текстового файла, содержащего набор необходимых транзакций. Для каждого задания по лабораторной работе данный файл создается исполнителем, в соответствии с выбранной темой. Подключение к файлу происходит путем следования указаниям диалоговых окон программы (на сайте разработчика можно ознакомиться с видеоматериалами подробнейшим образом описывающими необходимый набор действий). В ходе подключения к файлу указывается форматы анализируемых данных, и символы их разделения. В качестве примера используется файл Supermarket.txt, с котором содержится таблица следующего вида:

Номер чека	Товар
160698	КЕТЧУПЫ, СОУСЫ, АДЖИКА
160698	МАКАРОННЫЕ ИЗДЕЛИЯ

160698 ЧАЙ
160747 МАКАРОННЫЕ ИЗДЕЛИЯ
160747 МЕД
160747 ЧАЙ
161217 КЕТЧУПЫ, СОУСЫ, АДЖИКА
161217 МАКАРОННЫЕ ИЗДЕЛИЯ
161217 СЫРЫ
161243 КЕТЧУПЫ, СОУСЫ, АДЖИКА
161243 МАКАРОННЫЕ ИЗДЕЛИЯ

В данной таблице «Номер чека» является номером транзакции, а «Товар» - элементом транзакции. Как номер транзакции, так и ее элементы являются дискретными данными. При описании столбцов в ходе импорта из файла необходимо указывать соответствующие параметры для данных.

2. Анализ данных

Нажатием клавиши F7 вызывается Мастер обработки, который позволит проанализировать данные и выявить в них необходимые ассоциативные правила. При настройке параметров построения ассоциативных правил используют минимальный и максимальный процент поддержки для часто встречающихся множеств и достоверности для ассоциативных правил. Это служит для исключения правил и наборов не представляющих интереса для исследования в виду своей очевидности, либо слишком малой значимости. Ввиду крайне малого объема составляемой модели базы данных для исследования в ходе лабораторной работы пороги процента поддержки часто встречающихся множеств и достоверности ассоциативных правил необходимо выставить от 0 до 100%.

3. Подготовка отчета по работе с DeductorStudio

Версия программы является обучающей, что обуславливает ряд функциональных ограничений по ее использованию. В связи с этим результаты исследования в DeductorStudio, необходимо сохранить в формате рисунка, нажатием клавиши **«*PrtSc*».**

5. Учебно-методическое и информационное обеспечение дисциплины

5.1. Перечень основной и дополнительной литературы, необходимой для освоения дисциплины

5.1.1. Перечень основной литературы:

1. Богданова, С.В. Информационные технологии : учебное пособие для студентов высших учебных заведений / С.В. Богданова, А.Н. Ермакова ; ФГБОУ ВПО Ставропольский государственный аграрный университет, Министерство сельского хозяйства РФ. - Ставрополь : Сервисшкола, 2014. - 211 с. : ил. - Библиогр. в кн. ; То же [Электронный ресурс]. - URL: http://biblioclub.ru/index.php?page=book&id=277476
2. Борисова И.В. Цифровые методы обработки информации [Электронный ресурс]: учебное пособие/ Борисова И.В.— Электрон. текстовые данные.— Новосибирск: Новосибирский государственный технический университет, 2014.— 139 с.— Режим доступа: http://www.iprbookshop.ru/45061.— ЭБС «IPRbooks», по паролю
3. Бельчик, Т.А. Основы математической обработки информации с помощью SPSS : учебное пособие / Т.А. Бельчик. - Кемерово : Кемеровский государственный

университет, 2013. - 232 с. - ISBN 978-5-8353-1265-8 ; То же [Электронный ресурс]. - URL: http://biblioclub.ru/index.php?page=book&id=232214

5.1.2. Перечень дополнительной литературы:

1. Исаев, Г. Н. Информационные технологии : учеб. пособие / Г.Н. Исаев. - М. : Омега-Л, 2012. - 464 с. : ил. - (Высшее техническое образование). - Библиогр.: с. 462-463. - ISBN 978-5-370-02165-7
2. Кузнецов, С.М. Информационные технологии : учебное пособие / С.М. Кузнецов. - Новосибирск : НГТУ, 2011. - 144 с. - ISBN 978-5-7782-1685-3 ; То же [Электронный ресурс]. - URL: http://biblioclub.ru/index.php?page=book&id=228789
3. Коноплева, И.А. Информационные технологии : учебное пособие / И.А. Коноплева, О.А. Хохлова, А.В. Денисов ; под ред. И.А. Коноплева. - 2-е изд., перераб. и доп. - М. : Проспект, 2014. - 328 с. - Библиогр. в кн. - ISBN 978-5-392-12385-8 ; То же [Электронный ресурс]. - URL: http://biblioclub.ru/index.php?page=book&id=251652

4.2. Перечень учебно-методического обеспечения самостоятельной работы обучающихся по дисциплине:

1. Методические указания по выполнению лабораторных работ по дисциплине «Технология обработки информации».
2. Методические рекомендации для студентов по организации самостоятельной работы по дисциплине «Технология обработки информации».

4.3. Перечень ресурсов информационно-телекоммуникационной сети «Интернет», необходимых для освоения дисциплины:

1. http://www.intuit.ru – сайт дистанционного образования в области информационных технологий
2. http://e.lanbook.com – ЭБС издательства «Лань».
3. http://www.biblioclub.ru – университетская библиотека онлайн.
4. http://window.edu.ru – образовательные ресурсы ведущих вузов
5. http://trpo.is-isea.ru/
6. http://www.iprbookshop.ru
7. Интернет-ресурсы: Все для учебы StudFiles - - http://www.studfiles.ru
8. Интернет-университет информационных технологий. - - www.intuit.ru
9. Научно-информационный портал - - http://sci-lib.com