



МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ

Федеральное государственное бюджетное образовательное учреждение
высшего образования

«ЮЖНО-УРАЛЬСКИЙ ГОСУДАРСТВЕННЫЙ ГУМАНИТАРНО-
ПЕДАГОГИЧЕСКИЙ УНИВЕРСИТЕТ»

(ФГБОУ ВО «ЮУрГГПУ»)

ФИЗИКО-МАТЕМАТИЧЕСКИЙ ФАКУЛЬТЕТ

КАФЕДРА ИНФОРМАТИКИ, ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ И
МЕТОДИКИ ОБУЧЕНИЯ ИНФОРМАТИКЕ

Методика изучения алгоритмов интеллектуального анализа данных на
факультативных занятиях

Выпускная квалификационная работа
по направлению 44.03.05 Педагогическое образование (два профиля подготовки)
Направленность программы бакалавриата

«Информатика. Математика»

Выполнил:

Студент группы ЗФ-513-111-5-1

Юрпалов Дмитрий Леонидович

Проверка на объем заимствований:

60,92 % авторского текста

Работа рекомендована к защите
рекомендована/не рекомендована

« 02 » июня 2017 г.

и.о. зав. кафедрой ИИТиМОИ

 к.п.н. Рузаков А.А.

Научный руководитель:

к.п.н., доцент кафедры ИИТиМОИ

Давыдова
Давыдова Надежда Алексеевна

« 22 » июня 2017 г.

Челябинск
2017



МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ

**Федеральное государственное бюджетное образовательное учреждение
высшего образования**

**«ЮЖНО-УРАЛЬСКИЙ ГОСУДАРСТВЕННЫЙ ГУМАНИТАРНО-
ПЕДАГОГИЧЕСКИЙ УНИВЕРСИТЕТ»
(ФГБОУ ВО «ЮУрГГПУ»)**

ФИЗИКО-МАТЕМАТИЧЕСКИЙ ФАКУЛЬТЕТ

**КАФЕДРА ИНФОРМАТИКИ, ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ И
МЕТОДИКИ ОБУЧЕНИЯ ИНФОРМАТИКЕ**

**Методика изучения алгоритмов интеллектуального анализа данных на
факультативных занятиях**

**Выпускная квалификационная работа
по направлению 44.03.05 Педагогическое образование (два профиля подготовки)
Направленность программы бакалавриата**

«Информатика. Математика»

Выполнил:
Студент группы ЗФ-513-111-5-1
Юрпалов Дмитрий Леонидович

Проверка на объем заимствований:
_____ % авторского текста

Работа _____ к защите
_____ рекомендована/не рекомендована

« ____ » _____ 2017 г.
и.о. зав. кафедрой ИИТиМОИ

_____ к.п.н. Рузаков А.А.

Научный руководитель:
к.п.н., доцент кафедры ИИТиМОИ

_____ Давыдова Надежда Алексеевна
« ____ » _____ 2017 г.

Работа защищена с оценкой

_____ « ____ » _____ 2017 г.

Челябинск
2017

ОГЛАВЛЕНИЕ

ВВЕДЕНИЕ.....	3
ГЛАВА 1. ИЗУЧЕНИЕ МЕТОДОВ ИНТЕЛЛЕКТУАЛЬНОГО АНАЛИЗА ДАННЫХ	6
1.1 Общие понятия интеллектуального анализа данных.....	6
1.2 Основные задачи интеллектуального анализа данных.....	9
1.3 Исследование алгоритмов интеллектуального анализа данных	19
Выводы по главе 1.....	24
ГЛАВА 2. РАЗРАБОТКА МЕТОДИКИ ИЗУЧЕНИЯ АЛГОРИТМОВ ИНТЕЛЛЕКТУАЛЬНОГО АНАЛИЗА ДАННЫХ НА ФАКУЛЬТАТИВНЫХ ЗАНЯТИЯХ.....	25
2.1 Организация внеурочной деятельности в условиях ФГОС.....	25
2.2 Программа факультативных занятий.....	31
2.3 Описание ресурса для изучения методов интеллектуального анализа данных	45
2.4 Результаты апробации факультативного курса в школе	56
Выводы по главе 2.....	57
ЗАКЛЮЧЕНИЕ	58
СПИСОК ЛИТЕРАТУРЫ.....	60
Приложения	62

ВВЕДЕНИЕ

Современный компьютерный термин Data Mining переведен как «извлечение информации» или «интеллектуальный анализ данных». Часто наряду с интеллектуальным анализом данных встречаются термины «Обнаружение знаний» и «Хранилище данных». Появление этих терминов, которые являются неотъемлемой частью Data Mining, связано с новым витком в разработке инструментов и методов обработки и хранения данных. Таким образом, цель Data Mining - выявить скрытые правила и шаблоны в больших (очень больших) объемах данных.

Дело в том, что сам человеческий разум не приспособлен к восприятию огромного количества разнородной информации. В среднем человек, за исключением некоторых лиц, не может выделить более двух-трех отношений даже в небольших выборках. Но традиционная статистика, которая долгое время считалась основным инструментом анализа данных, также часто не может решить проблемы из реальной жизни. Он работает с усредненными характеристиками выборки, которые часто являются фиктивными значениями (средняя платежеспособность клиента, когда в зависимости от функции риска или функции потерь нужно уметь прогнозировать последовательность и намерения клиента, средняя интенсивность сигнала, тогда как вас интересуют характерные особенности и фон пиков сигнала и т.д.).

Поэтому методы математической статистики полезны главным образом для проверки заранее сформулированных гипотез, тогда как определение гипотезы иногда является довольно сложной и трудоемкой задачей. Современные технологии интеллектуального анализа данных обрабатывают информацию для автоматического поиска паттернов (шаблонов), характерных для любых фрагментов разнородных многомерных данных. В отличие от оперативной аналитической обработки данных (OLAP) в Data Mining, формулировка гипотез и выявление необычных (неожиданных)

шаблонов переносится от человека к компьютеру. Data Mining - это не один, а комбинация большого числа различных методов обнаружения знаний. Выбор метода часто зависит от типа доступных данных и от того, какую информацию мы пытаемся получить.

Рассмотрим свойства обнаруженного знания, данные в определении, более подробно [6].

Знание должно быть новым, ранее неизвестным. Усилия, потраченные на обнаружение знаний, уже известных пользователю, не окупаются. Следовательно, значение представлено новым, ранее неизвестным знанием.

Знание должно быть нетривиальным. Результаты анализа должны отражать неочевидные, неожиданные закономерности в данных, которые составляют так называемое скрытое знание. Результаты, которые можно было получить более простым способом (например, визуальный просмотр), не оправдывают использование мощных методов интеллектуального анализа данных.

Знания должны быть практически полезными. Найденные знания должны быть применимы, в том числе к новым данным, с достаточно высокой степенью надежности. Полезность состоит в том, что эти знания могут принести определенную пользу в их применении.

Знание должно быть доступно понимающему человеку. Найденные образцы должны быть логически объяснены, в противном случае существует вероятность того, что они являются случайными. Кроме того, полученные знания должны быть представлены понятным человеку образом.

Объект исследования: методы интеллектуального анализа данных.

Предмет исследования: факультативный курс по интеллектуальному анализу данных.

Целью данной работы является разработка методики изучения алгоритмов Data Mining на факультативных занятиях в школе.

Гипотеза исследования: изучение методов интеллектуального анализа данных будет способствовать формированию представлений у учащихся о рациональных подходах к принятию решений.

Задачи исследования:

- рассмотрение общих понятий интеллектуального анализа данных;
- рассмотрение основных задач интеллектуального анализа данных;
- рассмотрение алгоритмов интеллектуального анализа данных;
- изучение особенностей организации внеурочной деятельности в условиях ФГОС;
- составление тематического плана факультативных занятий;
- разработка электронного ресурса для изучения алгоритмов интеллектуального анализа данных;
- разработка приложения для закрепления полученных знаний.

ГЛАВА 1. ИЗУЧЕНИЕ МЕТОДОВ ИНТЕЛЛЕКТУАЛЬНОГО АНАЛИЗА ДАННЫХ

1.1 Общие понятия интеллектуального анализа данных

В настоящее время ведется активная разработка технологии интеллектуального анализа данных (ИАД), появление которой связано в первую очередь с необходимостью аналитической обработки чрезвычайно больших объемов информации, накопленной в современных хранилищах данных. Возможность использования известных методов математической статистики и машинного обучения для решения подобных задач открыла новые возможности для аналитиков, исследователей, а также для лиц, принимающих решения, - менеджеров и руководителей компаний [2].

Сложность и разнообразие методов ИАД требуют создания специализированных инструментов конечного пользователя для решения типичных проблем анализа информации в конкретных областях. Поскольку эти инструменты используются в сложных многофункциональных системах поддержки принятия решений, они должны быть легко интегрированы в такие системы. Одной из наиболее важных и перспективных областей применения ИАД являются бизнес-приложения.

Системы ИАД используются в исследованиях и образовании, правоохранительной деятельности, производстве, здравоохранении и во многих других областях.

Data Mining - это процесс обнаружения в исходных данных (ранних данных) неизвестной, нетривиальной, практически полезной, доступной интерпретации знаний (шаблонов), необходимых для принятия решений в различных сферах человеческой деятельности.

Появление технологий DM&KDD (Data Mining and Knowledge Discovery in Databases (KDD)) связано с накоплением огромных объемов информации в компьютерных базах данных, которые нерентабельно хранить и которые стали трудно использовать традиционными методами. Последнее

обстоятельство связано с быстрым развитием компьютерных технологий и программных средств для представления и обработки данных. Большие объемы накопленных данных постоянно нуждаются в модификации из-за быстрой смены аппаратных и программных баз данных, в то время как потери и искажение информации неизбежны. Одним из способов преодоления таких трудностей является создание хранилищ данных, доступ к которым не будет сильно зависеть от изменения данных во времени и от используемого программного обеспечения. Другой подход сфокусирован на сжатии больших объемов данных путем поиска некоторых общих закономерностей (знаний) в накопленной информации. Оба направления актуальны с практической точки зрения. Второй подход более интересен специалистам в области ИИ, поскольку он связан с решением проблемы приобретения новых знаний. Следует отметить, что наиболее плодотворным является сочетание обоих направлений.

Наличие хранилища данных является необходимым условием успеха всего процесса KDD. Хранилище данных называется предметно-ориентированным, интегрированным, ограниченным по времени, неизменным набором данных, который используется для поддержки процесса принятия управленческих решений. Ориентация объекта означает, что данные сгруппированы по категориям и сохранены в соответствии с областями, которые они описывают, а не в соответствии с приложениями, которые их используют. Этот принцип хранения гарантирует, что отчеты, созданные разными аналитиками, будут опираться на один и тот же набор данных. Прикрепление ко времени означает, что хранилище можно рассматривать как набор исторических данных; специфические значения данных однозначно связаны с определенными моментами времени. Атрибут времени всегда явно присутствует в структурах хранилища данных. Данные, хранящиеся в репозитории, больше не изменяются, в отличие от операционных систем, в которых присутствуют только последние, постоянно меняющиеся версии данных. Хранилища данных характеризуются

операциями добавления, а не изменения данных. Современные средства администрирования хранилища данных обеспечивают эффективное взаимодействие с программными инструментами DM&KDD. В качестве примера можно привести развитие компании SAS Institute: SAS Warehouse Administrator и SAS Enterprise Miner [5].

Цель интеллектуального анализа данных (англ. Data mining, другие варианты перевода - «добыча данных», «раскопка данных») - обнаружение неявных закономерностей в наборах данных. В качестве научного направления он начал активно развиваться в 90-е годы XX века, что было вызвано широким распространением технологий автоматизированной обработки информации и накоплением больших объемов данных в компьютерных системах. И хотя существующие технологии позволили, например, быстро найти необходимую информацию в базе данных, этого во многих случаях было недостаточно. Необходимо было искать связи между отдельными событиями между большими объемами данных, которые требовали методов математической статистики, теории баз данных, теории искусственного интеллекта и ряда других областей [3].

Классическим является определение, данное одним из основателей направления Григорием Пятецким-Шапиро [6]: DataMining - исследование и обнаружение в сырых данных скрытых знаний, ранее не известных, нетривиальных, практически полезных, доступных для интерпретации.

Учитывая разнообразие форм представления данных, используемых алгоритмов и приложений, интеллектуальный анализ данных может быть выполнен с использованием программных продуктов следующих классов:

- специализированные «коробочные» программные продукты для интеллектуального анализа;
- математические пакеты;
- электронные таблицы (и надстройки над ними);

– средства, интегрированные в системы управления базами данных (СУБД).

Задача интеллектуального анализа данных состоит в том, чтобы провести исследование целого ряда объектов (или вариантов) во время интеллектуального анализа данных.

В большинстве случаев его можно представить в виде таблицы, каждая строка которой соответствует одной из опций, а столбцы содержат значения параметров, которые ее характеризуют. Зависимая переменная - это параметр, значение которого рассматривается как зависящее от других параметров (независимых переменных). Фактически эта зависимость также должна определяться с использованием методов интеллектуального анализа данных.

1.2 Основные задачи интеллектуального анализа данных

1.2.1 Задача классификации и регрессии

При анализе часто бывает необходимо определить, к какому известному классу относятся исследуемые объекты, т.е. классифицировать их. Например, когда человек обращается в банк за предоставлением ему кредита, сотрудник банка должен решить, является ли потенциальный клиент кредитоспособным или нет. Очевидно, такое решение принимается на основе данных об исследуемом объекте (в данном случае, лице): месте его работы, размере заработной платы, возрасте, составе семьи и т.д. В результате анализируя эту информацию, сотрудник банка должен отнести лицо к одному из двух известных классов «кредитоспособные» и «некредитоспособные» [4].

Другим примером проблемы классификации является фильтрация электронной почты. В этом случае программа фильтрации должна классифицировать входящее сообщение как спам (нежелательная электронная почта) или как письмо. Это решение принимается исходя из частоты появления в сообщении определенных слов (например, имя

получателя, безличное обращение, слова и фразы: приобрести, «зарабатывать», «выгодное предложение» и т.д.).

В общем случае число классов в задачах классификации может быть больше двух. Например, в задаче распознавания образа фигур таких классов может быть 10 (по числу цифр в десятичной системе). В такой задаче объект классификации представляет собой матрицу пикселей, представляющую изображение распознанной цифры. В этом случае цвет каждого пикселя является характеристикой анализируемого объекта.

В Data Mining задача классификации рассматривается как задача определения значения одного из параметров анализируемого объекта на основе значений других параметров. Параметр часто называют зависимой переменной, а параметры, участвующие в ее определении, являются независимыми переменными. В рассмотренных примерах независимыми переменными были:

- заработная плата, возраст, количество детей и т.д.;
- частота определенных слов;
- значения цвета матричных пикселей.

Зависимыми переменными в тех же примерах были:

- кредитоспособность клиента (возможные значения для этой переменной - «да» и «нет»);
- тип сообщения (возможные значения для этой переменной - «спам» и «почта»);
- цифра изображения (возможные значения этой переменной: 0,1,..., 9).

Следует отметить, что во всех рассмотренных примерах независимая переменная принимала значение из конечного набора значений: {да, нет}, {spam, mail}, {0, 1, ..., 9}. Если значения независимых и зависимых переменных являются вещественными числами, то проблема называется

проблемой регрессии. Примером регрессионной проблемы может быть задача определения суммы ссуды, которую банк может выдать клиенту.

Проблема классификации и регрессии решается в два этапа. На первом выбирается учебный образец. Он включает объекты, для которых известны значения как независимых, так и зависимых переменных. В примерах, описанных ранее, такими учебными образцами могут быть:

- информация о клиентах, которым ранее предоставлялись ссуды на различные суммы, и информация об их возврате;
- сообщения, классифицированные вручную как спам или как письмо;
- ранее распознанная матрица цифровых изображений.

На основе обучающей выборки строится модель для определения значения зависимой переменной. Его часто называют классификацией или функцией регрессии. Для получения наиболее точной функции на обучающую выборку накладываются следующие основные требования:

- количество объектов, включенных в выборку, должно быть достаточно большим. Чем больше объектов, тем точнее будет встроенная функция классификации или регрессии;
- образец должен включать объекты, которые представляют все возможные классы в случае проблемы классификации или весь диапазон значений в случае проблемы регрессии;
- для каждого класса проблемы классификации или каждого интервала диапазона значений в проблеме регрессии выборка должна содержать достаточное количество объектов.

На втором этапе построенная модель применяется к анализируемым объектам (объектам с неопределенным значением зависимой переменной).

Проблема классификации и регрессии имеет геометрическую интерпретацию. Рассмотрим его на примере с двумя независимыми

переменными, что позволит нам представить его в двумерном пространстве (Рисунок 1).

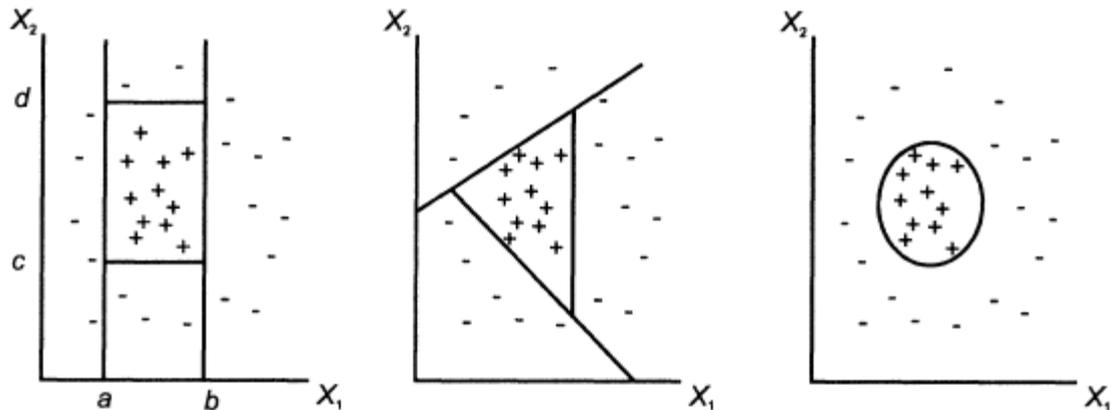


Рисунок 1 Классификация в двумерном пространстве

Каждому объекту назначается точка на плоскости. Символы «+» и «-» обозначают принадлежность объекта к одному из двух классов. Очевидно, данные имеют четко определенную структуру: все точки класса «+» сосредоточены в центральной области. Построение функции классификации сводится к построению поверхности, которая окружает центральную область. Он определяется как функция, которая имеет значения «+» внутри закрытой области и «-» - снаружи.

Как видно из рисунка, существует несколько возможностей построения контурной области. Форма функции зависит от используемого алгоритма.

Основными проблемами, возникающими при решении проблем классификации и регрессии, являются неудовлетворительное качество исходных данных, в которых есть как ошибочные данные, так и отсутствующие значения, различные типы атрибутов - числовые и категориальные, различное значение атрибутов, а также так называемые проблемы *overfitting* и *underfilling*. Суть первого из них заключается в том, что функция классификации при построении «слишком хорошо» адаптируется к данным, а ошибки и аномальные значения, встречающиеся в них, пытаются интерпретировать как часть внутренней структуры данных. Очевидно, что в будущем такая модель не будет работать корректно с другими данными, где характер ошибок будет несколько иным. Термин

underfitting обозначает ситуацию, когда количество ошибок при проверке классификатора на обучающем наборе слишком велико. Это означает, что специальные шаблоны в данных не были найдены, либо они совсем отсутствуют, либо необходимо выбрать другой метод их обнаружения.

1.2.2 Задача прогнозирования

Проблема прогнозирования новых значений выделяется отдельно на основе доступных значений числовой последовательности (или нескольких последовательностей между значениями, в которых наблюдается корреляция). Это может учитывать существующие тренды (тенденции), сезонность и другие факторы. Классическим примером является прогнозирование курсов акций на фондовой бирже [6].

Здесь требуется сделать небольшое отступление. Способ решения проблемы интеллектуального анализа можно разделить на два класса: обучение с преподавателем (с английского надзорного обучения) и обучение без учителя (с английского неконтролируемого обучения). В первом случае требуется набор обучающих данных, на которых создается и обучается модель интеллектуального анализа данных. Готовая модель тестируется и впоследствии используется для прогнозирования значений в новых наборах данных. Иногда в одном и том же случае они говорят об управляемых алгоритмах интеллектуального анализа. К этому типу относятся задачи классификации и регрессии.

Во втором случае цель состоит в том, чтобы определить закономерности существующего набора данных. В этом случае обучающий образец не требуется. В качестве примера можно привести задачу анализа потребительской корзины, когда в процессе исследования товаров, которые чаще всего покупаются вместе, идентифицируются. Проблема кластеризации принадлежит этому классу.

Также можно говорить о классификации задач интеллектуального анализа данных по целям, в соответствии с которыми они делятся на описательные и прогностические. Целью решения описательных задач

является лучшее понимание изучаемых данных, выявление закономерностей в них, даже если они не встречаются в других наборах данных. Для прогнозирующих проблем характерно, что в процессе их решения, на основе набора данных с известными результатами, строится модель для прогнозирования новых значений.

1.2.3 Задача кластеризации

Задача кластеризации состоит в разделении исследуемого набора объектов на группы «похожих» объектов, называемых кластерами. Часто решение проблемы разбиения набора элементов на кластеры называется кластерным анализом.

Кластеризация может быть применена практически в любой области, где необходимо изучить экспериментальные или статистические данные. Рассмотрим пример из области маркетинга, в котором эта задача называется сегментацией [5].

Концептуально сегментирование основано на предпосылке, что все потребители различны. У них разные потребности, разные требования к товарам, они ведут себя по-разному: в процессе выбора товаров, в процессе приобретения товара, в процессе использования товара, в процессе формирования реакции на товар. В этой связи необходимо по-разному подходить к работе с потребителями: предлагать им разные товары в соответствии с их характеристиками, продвигать и продавать товары по-разному. Чтобы определить, как потребители отличаются друг от друга и как эти различия отражаются в требованиях к продукту, проводится сегментация потребителей.

В маркетинге критериями (характеристиками) сегментации являются: географическое положение, социально-демографические характеристики, мотивы для покупки и т.д.

Основываясь на результатах сегментирования, маркетолог может определить, например, такие характеристики сегментов рынка, как реальная и потенциальная емкость сегмента, группы потребителей, потребности

которых не полностью удовлетворяются любым производителем, работающим в этом сегменте рынка, и т.д. На основании этих параметров маркетолог может сделать вывод о привлекательности фирмы в каждом из выбранных сегментов рынка.

Для научных исследований изучение результатов кластеризации, а именно выяснение причин группировки объектов, может открыть новые перспективные направления. Традиционный пример, который обычно дается для этого случая, - это периодическая таблица элементов. В 1869 году Дмитрий Менделеев разделил 60 известных в то время элементов на кластеры или периоды. Элементы, которые попадают в одну группу, имеют сходные характеристики. Изучение причин, по которым эти элементы были разбиты на явные кластеры, во многом определило приоритеты научных исследований на долгие годы. Но только через 50 лет квантовая физика дала убедительные объяснения периодической системы.

Кластеризация отличается от классификации тем, что для анализа не требуется выделенная зависимая переменная. С этой точки зрения это относится к классу неконтролируемого обучения. Эта проблема решается на начальных этапах исследования, когда мало известно о данных. Его решение помогает лучше понять данные, и с этой точки зрения задача кластеризации - описательная задача.

Проблема кластеризации характеризуется отсутствием каких-либо различий между обеими переменными и между объектами. Наоборот, ищутся группы ближайших, подобных объектов. Методы автоматической кластеризации редко используются сами по себе, просто для получения групп похожих объектов. После определения кластеров используются другие методы интеллектуального анализа данных, чтобы попытаться установить принцип секционирования.

Кластерный анализ позволяет рассматривать довольно большой объем информации и резко сокращать, сжимать большие объемы информации, делать их компактными и интуитивно понятными.

Отметим ряд особенностей, присущих проблеме кластеризации.

Во-первых, решение сильно зависит от природы объектов данных (и их атрибутов). Таким образом, с одной стороны, это могут быть однозначно определенные, четко определенные объекты, а с другой - объекты, которые имеют вероятностное или нечеткое описание.

Во-вторых, решение также зависит от представления кластеров и ожидаемых отношений объектов данных и кластеров. Таким образом, необходимо учитывать такие свойства, как возможность/невозможность объектов принадлежать нескольким кластерам. Необходимо определить само понятие принадлежности к кластеру: однозначное (принадлежит/не принадлежит), вероятностное (вероятность принадлежности), нечеткое (степень принадлежности).

1.2.4 Задача поиска ассоциативных правил

Поиск ассоциативных правил - одно из самых популярных приложений интеллектуального анализа данных. Суть проблемы заключается в определении часто встречающихся множеств объектов в большом наборе таких множеств. Эта задача является частным случаем проблемы классификации. Изначально она применялась при анализе тенденций поведения клиентов в супермаркетах. Анализу были подвергнуты данные о сделанных покупках, которые покупатели положили в тележку (корзину). Это стало причиной второго общепринятого названия - анализа рыночных корзин (Basket Analysis). Анализ этих данных позволяет судить о том, какие товары покупаются вместе, в каком порядке, какими категориями потребителей, какие продукты предпочтительней, в какие периоды времени и т.д. Такая информация позволяет более эффективно планировать закупку товаров, проведение рекламной кампании и т.д.

Например, из набора покупок, сделанных в магазине, вы можете определить следующие наборы продуктов, которые покупаются вместе [15]:

- {макароны, кетчуп};
- {сосиски, картофель}.

Поэтому можно заключить, что если куплены макароны или картофель, то, как правило, покупают кетчуп или сосиски соответственно. Обладая такими знаниями, можно разместить эти товары рядом, объединить их в один пакет со скидкой или предпринять другие действия, которые стимулируют покупателя к покупке товара.

Задача поиска ассоциативных правил актуальна не только в сфере торговли. Например, в сфере услуг интерес представляет то, какие услуги клиенты предпочитают использовать в совокупности. Чтобы получить эту информацию, проблема решается со ссылкой на данные об услугах, которые один клиент использует в течение определенного времени (месяц, год). Это помогает определить, например, как лучше всего составить пакеты услуг, предлагаемых клиенту.

В медицине симптомы и заболевания, наблюдаемые у пациентов, могут быть проанализированы. В этом случае знание комбинаций заболеваний и симптомов наиболее часто помогают правильно диагностировать в будущем.

В анализе часто встречается последовательность событий, которые происходят. Если обнаруживаются закономерности в таких последовательностях, можно с определенной вероятностью предсказать возникновение событий в будущем, что позволяет принимать более правильные решения. Такая задача является своего рода проблемой поиска ассоциативных правил и называется последовательным анализом.

Основное различие между задачей последовательного анализа и поиском ассоциативных правил - установление порядка отношений между изученными множествами. Это соотношение можно определить по-разному. При анализе последовательности событий, происходящих во времени, объекты таких множеств являются событиями, а порядок отношений соответствует хронологии их возникновения.

Последовательный анализ широко используется, например, в телекоммуникационных компаниях, для анализа данных об авариях в

различных узлах сети. Информация о последовательности аварий может помочь в выявлении проблем и предотвращении новых аварий.

В Таблица 1 приведены примеры задач интеллектуального анализа данных из различных областей.

Таблица 1

Примеры применения интеллектуального анализа данных

	Информационные технологии	Торговля	Финансовая сфера
Классификация			Оценка кредитоспособности
Регрессия			Оценка допустимого кредитного лимита
Прогнозирование		Прогнозирование продаж	Прогнозирование цен акции
Кластеризации		Сегментация клиентов	Сегментация клиентов
Определения взаимосвязей		Анализ потребительской корзины	
Анализ последовательно страниц	Анализ переходов по страницам веб-сайта		
Анализ отклонений	Обнаружение вторжений в информационные системы		Выявление мошенничества с банковскими картами

1.3 Исследование алгоритмов интеллектуального анализа данных

Алгоритм интеллектуального анализа данных представляет собой набор эвристик и вычислений, который создает модель для добычи данных из данных. Чтобы создать модель, алгоритм сначала анализирует предоставленные данные, ищет определенные шаблоны и тенденции. Алгоритм использует результаты этого анализа для выбора оптимальных параметров для создания модели интеллектуального анализа данных. Эти параметры затем применяются ко всему набору данных для идентификации используемых шаблонов и получения подробной статистики [1].

Модель интеллектуального анализа данных, созданная алгоритмом из предоставленных данных, может принимать различные формы, включая следующие:

- набор кластеров, описывающих взаимосвязи параметров в наборе данных;
- дерево решений, которое прогнозирует результат и описывает, какое влияние на этот результат оказывают разные критерии;
- математическая модель, которая прогнозирует те или иные события;
- набор правил, описывающих группировку объектов в транзакции, а также вероятность группировки данных объектов.

Эти алгоритмы являются реализациями некоторых из самых популярных методов, используемых в интеллектуальном анализе данных. Все алгоритмы интеллектуального анализа данных Microsoft настроены, они полностью программируются через API-интерфейсы служб SQL Server или компоненты интеллектуального анализа данных.

Кроме того, поддерживается использование сторонних алгоритмов, соответствующих спецификации OLE DB для интеллектуального анализа данных. Также возможно разработать пользовательские алгоритмы, которые

можно зарегистрировать как службы, а затем использовать их на платформе SQL Server для интеллектуального анализа данных.

Выбор правильного алгоритма для использования в конкретной аналитической задаче может быть довольно сложным. Хотя для выполнения одной и той же задачи могут использоваться разные алгоритмы, каждый алгоритм дает другой результат, а некоторые алгоритмы могут давать результаты более одного типа. Например, можно использовать алгоритм дерева решений (Microsoft) не только для прогнозирования, но и как способ уменьшения количества столбцов в наборе данных, потому что дерево решений может определять столбцы, которые не влияют на окончательную модель интеллектуального анализа данных.

Analysis Services включает следующие типы алгоритмов.

Алгоритмы классификации предсказывают одну или несколько дискретных переменных, основанных на других атрибутах набора данных.

Алгоритмы регрессии предсказывают одну или несколько непрерывных переменных, например, прибыль или убыток, на основе других атрибутов набора данных.

Алгоритмы сегментирования делят данные на группы или кластеры элементов, которые имеют схожие свойства.

Алгоритмы взаимосвязи ищут корреляцию между различными атрибутами в наборе данных. Наиболее частое применение этого типа алгоритма - создание правил взаимосвязи, которые могут быть использованы для анализа набора данных.

Алгоритмы анализа последовательности суммируют часто встречающиеся последовательности в данных, например, поток данных в Интернете.

Однако ничто не заставляет пользователя ограничивать себя одним алгоритмом в своих решениях. Опытные аналитики часто используют один алгоритм для определения наиболее эффективных входных данных (т.е. переменных), после чего применяют другой алгоритм для прогнозирования

определенного результата на основе этих данных. Интеллектуальный анализ данных SQL Server позволяет строить множество моделей на основе одной структуры интеллектуального анализа таким образом, что в рамках одного решения для интеллектуального анализа данных можно использовать алгоритм кластеризации, модель дерева решений и упрощенный алгоритм Байеса для получения различных данных представления. Несколько алгоритмов в одном решении также могут использоваться для выполнения отдельных задач. Например, используя регрессию, можно получить финансовые прогнозы и использовать алгоритм нейронной сети для анализа факторов, влияющих на продажи.

Чтобы облегчить выбор алгоритмов для решения определённой задачи, в таблице приведены типы задач, для решения которых обычно используется каждый алгоритм (Таблица 2).

Таблица 2

Примеры задач и возможные алгоритмы их решения

Примеры задач	Подходящие алгоритмы
Прогнозирование дискретного атрибута: <ul style="list-style-type: none"> – Пометка клиентов из списка потенциальных покупателей как хороших и плохих кандидатов. – Вычисление вероятности отказа сервера в течение следующих шести месяцев. – Классификация вариантов развития болезней пациентов и исследование связанных факторов. 	Алгоритм дерева принятия решений Упрощенный алгоритм Байеса Алгоритм кластеризации Алгоритм нейронной сети
Прогнозирование непрерывного атрибута:	Алгоритм дерева принятия решений Алгоритм временных рядов

Примеры задач	Подходящие алгоритмы
<ul style="list-style-type: none"> – Прогноз продаж на следующий год. – Прогноз количества посетителей сайта с учётом прошлых лет и сезонных тенденций. – Формирование оценки риска с учётом демографии. 	Алгоритм линейной регрессии
<p>Прогнозирование последовательности:</p> <ul style="list-style-type: none"> – Анализ маршрута перемещения по вебсайту компании. – Анализ факторов, ведущих к отказу сервера. – Отслеживание и анализ последовательностей действий во время посещения поликлиники с целью формулирования рекомендаций по общим действиям. 	Алгоритм кластеризации последовательностей
<p>Нахождение групп общих элементов в транзакциях:</p> <ul style="list-style-type: none"> – Использование анализа потребительской корзины для определения мест размещения продуктов. – Выявление дополнительных продуктов, которые можно предложить купить клиенту. – Анализ данных опроса, проведённого среди посетителей 	Алгоритм взаимосвязей Алгоритм дерева принятия решений

Примеры задач	Подходящие алгоритмы
<p>события, с целью выявления того, какие действия и стенды были связаны, чтобы планировать будущие действия.</p>	
<p>Нахождение групп схожих элементов:</p> <ul style="list-style-type: none"> – Создание профилей рисков для пациентов на основе таких атрибутов, как демография и поведение. – Анализ пользователей по шаблонам просмотра и покупки. – Определение серверов, которые имеют аналогичные характеристики использования. 	<p>Алгоритм кластеризации Алгоритм кластеризации последовательностей</p>

Выводы по главе 1

По результатам данной главы можно сделать вывод, что изучение алгоритмов интеллектуального анализа данных является важным в современных условиях.

Изучение методов интеллектуального анализа данных позволит достичь следующих результатов:

- развить логическое мышление;
- получить навыки обработки больших объемов информации;
- привить умение выделять значащие качества рассматриваемого или исследуемого объекта;
- принимать решения согласно взвешенным и обоснованным фактам;
- прогнозировать определенные события;
- привить умение разработки алгоритмов при изучении каких-либо событий.

Так как данные методы не включены в основную программу курса «Информатика», необходимо разработать план факультативных занятий, охватывающих данную тему.

Методика проведения факультативных занятий должна отвечать следующим требованиям:

- быть доступной для понимания учащимися 10-11 классов;
- быть в достаточной мере проиллюстрированной;
- содержать научные термины в доступной для понимания учащихся форме;
- использовать для пояснения повседневные примеры.

ГЛАВА 2. РАЗРАБОТКА МЕТОДИКИ ИЗУЧЕНИЯ АЛГОРИТМОВ ИНТЕЛЛЕКТУАЛЬНОГО АНАЛИЗА ДАННЫХ НА ФАКУЛЬТАТИВНЫХ ЗАНЯТИЯХ.

2.1 Организация внеурочной деятельности в условиях ФГОС

Федеральный государственный образовательный стандарт (ФГОС) утвержден Приказ Министерства образования и науки РФ от 17 мая 2012 г. N 413 "Об утверждении федерального государственного образовательного стандарта среднего общего образования" (ред. от 29.12.2014).

ФГОС содержит требования к структуре образовательных программ, формулирует требования к результатам освоения дисциплин.

Согласно ФГОС изучение предметной области "Информатика" должно обеспечить [9]:

- сформированность представлений о социальных, культурных и исторических факторах становления информатики;
- сформированность основ логического, алгоритмического и математического мышления;
- сформированность умений применять полученные знания при решении различных задач;
- сформированность представлений о роли информатики и ИКТ в современном обществе, понимание основ правовых аспектов использования компьютерных программ и работы в Интернете;
- сформированность представлений о влиянии информационных технологий на жизнь человека в обществе; понимание социального, экономического, политического, культурного, юридического, природного, эргономического, медицинского и физиологического контекстов информационных технологий;

– принятие этических аспектов информационных технологий; осознание ответственности людей, вовлеченных в создание и использование информационных систем, распространение информации.

Предметные результаты изучения предметной области "Информатика" должны отражать [9]:

1) сформированность представлений о роли информации и связанных с ней процессов в окружающем мире;

2) владение навыками алгоритмического мышления и понимание необходимости формального описания алгоритмов;

3) владение умением понимать программы, написанные на выбранном для изучения универсальном алгоритмическом языке высокого уровня; знанием основных конструкций программирования; умением анализировать алгоритмы с использованием таблиц;

4) владение стандартными приемами написания на алгоритмическом языке программы для решения стандартной задачи с использованием основных конструкций программирования и отладки таких программ; использование готовых прикладных компьютерных программ по выбранной специализации;

5) сформированность представлений о компьютерно-математических моделях и необходимости анализа соответствия модели и моделируемого объекта (процесса); о способах хранения и простейшей обработке данных; понятия о базах данных и средствах доступа к ним, умений работать с ними;

6) владение компьютерными средствами представления и анализа данных;

7) сформированность базовых навыков и умений по соблюдению требований техники безопасности, гигиены и ресурсосбережения при работе со средствами информатизации; понимания основ правовых аспектов использования компьютерных программ и работы в Интернете.

Требования к предметным результатам освоения углубленного курса информатики должны включать требования к результатам освоения базового курса и дополнительно отражать [9]:

1) владение системой базовых знаний, отражающих вклад информатики в формирование современной научной картины мира;

2) овладение понятием сложности алгоритма, знание основных алгоритмов обработки числовой и текстовой информации, алгоритмов поиска и сортировки;

3) владение универсальным языком программирования высокого уровня (по выбору), представлениями о базовых типах данных и структурах данных; умением использовать основные управляющие конструкции;

4) владение навыками и опытом разработки программ в выбранной среде программирования, включая тестирование и отладку программ; владение элементарными навыками формализации прикладной задачи и документирования программ;

5) сформированность представлений о важнейших видах дискретных объектов и об их простейших свойствах, алгоритмах анализа этих объектов, о кодировании и декодировании данных и причинах искажения данных при передаче; систематизацию знаний, относящихся к математическим объектам информатики; умение строить математические объекты информатики, в том числе логические формулы;

6) сформированность представлений об устройстве современных компьютеров, о тенденциях развития компьютерных технологий; о понятии "операционная система" и основных функциях операционных систем; об общих принципах разработки и функционирования интернет-приложений;

7) сформированность представлений о компьютерных сетях и их роли в современном мире; знаний базовых принципов организации и функционирования компьютерных сетей, норм информационной этики и права, принципов обеспечения информационной безопасности, способов и средств обеспечения надежного функционирования средств ИКТ;

8) владение основными сведениями о базах данных, их структуре, средствах создания и работы с ними;

9) владение опытом построения и использования компьютерно-математических моделей, проведения экспериментов и статистической обработки данных с помощью компьютера, интерпретации результатов, получаемых в ходе моделирования реальных процессов; умение оценивать числовые параметры моделируемых объектов и процессов, пользоваться базами данных и справочными системами;

10) сформированность умения работать с библиотеками программ; наличие опыта использования компьютерных средств представления и анализа данных.

В Стандарте основного общего образования предлагается новая структура учебного плана, в состав которого в качестве компонента включена внеурочная деятельность, которая, являясь неразрывной частью образовательного процесса, направлена на становление личностных характеристик выпускника, достижение личностных и метапредметных результатов освоения обучающимися основной образовательной программы основного общего образования. Внеурочная деятельность более эффективно, чем урочная, позволяет удовлетворить индивидуальные познавательные потребности обучающихся, организовать деятельность, направленную на развитие конкретного ученика, достижение личностных результатов освоения основной образовательной программы.

Цели внеурочной деятельности [12]:

- создать условия для развития личностных характеристик каждого ученика в процессе творческой коллективной деятельности;
- углубить и расширить знания обучающихся (за счет включения внепрограммного и расширения программного материала);
- способствовать формированию метапредметных универсальных способов учебной деятельности (путем выполнения творческих, исследовательских, проектных работ, не укладывающихся в рамки уроков);

- создать условия для сохранения и укрепления здоровья обучающихся;
- содействовать социализации, профориентации школьников.

Согласно Стандарту, «внеурочная деятельность организуется по направлениям развития личности (духовно-нравственное, физкультурно-спортивное и оздоровительное, социальное, общеинтеллектуальное, общекультурное). Вне зависимости от реализуемого педагогом направления, внеурочная деятельность отличается разнообразием организационных форм. Выбор содержания и организационных форм работы, как правило, определяется педагогом исходя из его компетентности, круга увлечений. При этом педагог обязан учитывать как интересы и возрастные особенности обучающихся, так и ресурсы образовательной среды, особенности и традиции конкретного образовательного учреждения [14].

Рекомендованная литература по предметной области «Информатика» [11] для учащихся 10-11 классов представлена в Таблица 3.

Таблица 3

Рекомендованная литература по предметной области «Информатика»

Авторы	Наименование	Класс	Издательство
Информатика (базовый уровень)			
Гейн А.Г., Ливчак А.Б., Сенокосов А.И. и др.	Информатика (базовый и углубленный уровень)	10	Издательство «Просвещение»
Гейн А.Г., Сенокосов АИ.	Информатика (базовый и углубленный уровень)	11	Издательство «Просвещение»
Семакин И.Г., Хеннер Е.К., Шеина Т.Ю.	Информатика. Базовый уровень: учебник для 10 класса	10	БИНОМ. Лаборатория знаний
Семакин И.Г., Хеннер Е.К.,	Информатика. Базовый уровень: учебник для 11	11	БИНОМ. Лаборатория

Авторы	Наименование	Класс	Издательство
Шейна Т.Ю.	класса		знаний
Информатика (углубленный уровень)			
Калинин И.А., Самылкина Н.Н.	Информатика. Углубленный уровень: учебник для 10 класса	10	БИНОМ. Лаборатория знаний
Калинин И.А., Самылкина Н.Н.	Информатика. Углубленный уровень: учебник для 11 класса	11	БИНОМ. Лаборатория знаний
Поляков К.Ю., Еремин Е.А.	Информатика. Углубленный уровень: учебник для 10 класса: в 2 ч.	10	БИНОМ. Лаборатория знаний
Поляков К.Ю., Еремин Е.А.	Информатика. Углубленный уровень: учебник для 11 класса: в 2 ч.	11	БИНОМ. Лаборатория знаний
Семакин И.Г., Шейна Т.Ю., Шестакова Л.В.	Информатика. Углубленный уровень: учебник для 10 класса: в 2 ч.	10	БИНОМ. Лаборатория знаний
Семакин И.Г., Хеннер Е.К., Шестакова Л.В.	Информатика. Углубленный уровень: учебник для 11 класса: в 2 ч.	11	БИНОМ. Лаборатория знаний
Фиошин М.Е., Рессин А.А, Юнусов СМ. / Под ред. Кузнецова	Информатика. Углублённый уровень	10	ДРОФА

Авторы	Наименование	Класс	Издательство
А.А.			
Фиошин М.Е., Рессин А.А, Юнусов СМ. / Под ред. Кузнецова А.А.	Информатика. Углублённый уровень	11	ДРОФА

Обзор заданных литературных источников [7, 8, 13] показал, что учебники по углубленному изучению информатики за 11 класс содержат краткий обзор алгоритмов интеллектуального анализа данных в контексте их использования для решения задач искусственного интеллекта, но не содержат их примеров, методики составления, обзора результатов применения.

Факультативный курс по предметной области информатика поможет расширить знания учащихся в области интеллектуального анализа данных и статистического анализа данных.

2.2 Программа факультативных занятий

2.2.1 Требования к оформлению и составлению программ факультативных курсов

Программа педагога (далее - Программа) - нормативно-правовой документ школы, учитывающий возможности методического, информационного и технического обеспечения учебного процесса и уровня подготовки обучающихся, определяющий объем, порядок, содержание изучения и преподавания факультативных курсов, курсов предпрофильной подготовки, элективных курсов, учебных предметов (не входящих в перечень учебных предметов федерального базисного учебного плана (в ред. Приказа Минобрнауки РФ от 20.08.2008 N 241)), составляющийся с учетом особенностей средней образовательной школы [10].

Порядок разработки и утверждения программы [9].

Программа составляется учителем-предметником на учебный год или ступень обучения.

Программа разрабатывается учителями-предметниками и проходит экспертизу на заседании научно-методического совета на предмет соответствия данным Требованиям, учебному плану, годовому календарному графику средней образовательной школы, целям и задачам образовательного учреждения. При соответствии программы установленным требованиям она визируется заместителем директора по учебно-воспитательной работе и утверждается приказом директора средней образовательной школы. При этом на титульном листе рабочей программы ставятся соответствующие грифы о согласовании с заместителем директора по учебно-воспитательной работе и утверждении программы директором средней образовательной школы. При несоответствии программы установленным требованиям зам. директора по учебно-воспитательной работе накладывает резолюцию о необходимости доработки с указанием конкретного срока.

Программа на следующий учебный год должна быть представлена на согласование заместителю директора по учебно-воспитательной работе до 30 мая и утверждена директором средней образовательной школы до 1 сентября.

Программа хранится у педагогического работника ведущего образовательную деятельность по этой программе и зам. директора по учебно-воспитательной работе.

В течение учебного года допускается корректировка программы, которая может быть отражена как в самой программе, так и в отдельном документе с указанием причин корректировки.

Программа имеет следующую структуру:

- титульный лист;
- пояснительная записка;
- учебно-методический комплекс;

- содержание, обозначенное в разделах и темах;
- календарно-тематическое планирование;
- требования к уровню подготовки учащихся, обучающихся по данной программе;
- список литературы (основной и дополнительной).

2.2.2 Программа факультативных занятий по информатике

Пояснительная записка

Основное содержание учебного предмета «Информатика» сосредоточено на формировании у учащихся знаний и умений, необходимых для использования персонального компьютера. Предлагаемая программа факультативных занятий предназначена для содействия изучению информатики в общеобразовательных учреждениях.

Программа факультативных занятий предназначена для учащихся 10-11 классов, которые проявляют интерес и склонность к изучению информатики. Это сделано с учетом содержания учебной программы по информатике.

Целью факультативных занятий является формирование познавательной деятельности и интерес к информатике, развитие способности учащихся эффективно использовать компьютер для решения практических задач.

Задачи факультативных занятий:

- расширять и углублять знания учащихся в соответствии с содержанием основного курса информатики;
- подготовить учащихся к систематическому изучению основ алгоритмизации и программирования;
- развить познавательную и творческую деятельность учащихся путем выполнения развлекательных заданий практического характера;
- содействовать формированию навыков учащихся в общеобразовательных умениях.

Этот факультативный курс составляет 17 часов. Курс состоит из восьми тем, организованных таким образом, чтобы сопровождать изучение предмета «Информатика» в 10-11 классах, без дублирования, а с расширением и углублением его содержания.

Расширение и углубление содержания основного курса информатики достигается при изучении нового материала и последующей работе при выполнении контрольных заданий, например, при прохождении тестовых заданий.

Развитию логического мышления способствует решение проблем, связанных с поиском алгоритма, ведущего к заданной цели. Это реализуется с помощью таких педагогических инструментов, как компьютерные образовательные среды при изучении и усвоении пройденного материала.

Регулярные задачи по работе с выделением основной идеи, отбором аргументов, составлением планов, будут полезны для формирования общеобразовательных навыков.

В структуре большинства классов есть теоретическая и практическая часть. Практическая часть может быть реализована как с компьютером, так и без него. Работа на компьютере должна быть организована с учетом возрастных особенностей учащихся, санитарных и гигиенических требований.

Для проведения занятий рекомендуются следующие формы и методы: презентация, демонстрация видеоматериалов, наглядных пособий, использование проектного метода, проблемный разговор, обсуждение, мозговой штурм, деловая игра и другие. Рекомендуется уделять первоочередное внимание активным формам и методам, привлекать учащихся к иллюстрации нового материала своими примерами, анализу способов работы и выбору оптимальных методов. Такие формы и методы позволят учителю осуществлять лично ориентированное, развивающее, гуманистически направленное обучение.

К данной программе разработаны и изданы дидактические материалы для учащихся и методические рекомендации для педагогов.

ЦЕЛИ И ЗАДАЧИ КУРСА

Интеллектуальный анализ данных широко применяется в мире при проведении исследований во многих областях, например, в биоинформатике, генетике, медицине, астрономии, образовании, экономике, информатике и других. Чтобы проводить эффективные эксперименты, получая доказательные и конкурентоспособные результаты, специалисты должны обладать знаниями и навыками в области интеллектуального анализа данных.

10-11 классы (17 часов)

Тематическое планирование занятий приведено в Таблица 4.

Таблица 4

Тематическое планирование занятий

№	Тема занятий	Количество часов			Форма проведения	Оборудование
		всего	теория	практика		
1	Общие понятия интеллектуального анализа данных: определение; области применения; примеры использования.	3	2	1	Лекция	Опорный конспект, электронный ресурс
2	Задача классификации: определение; рассмотрение наглядного примера; шкалирование.	2	1	1	Лекция, практикум	Опорный конспект, электронный ресурс, система контроля знаний
3	Задача регрессии: определение; рассмотрение наглядного примера; отличие от классификации.	2	1	1	Лекция, практикум.	Опорный конспект, электронный ресурс

№	Тема занятий	Количество часов			Форма проведения	Оборудование
		всего	теория	практика		
4	Задача прогнозирования: определение; методы; прогнозирование в MS Excel	3	1	2	Лекция, практикум.	Опорный конспект, электронный ресурс, решенные задания
5	Задача кластеризации: определение; сравнение с изученными методами; примеры.	2	1	1	Лекция, практикум.	Опорный конспект, электронный ресурс
6	Задача поиска ассоциативных правил: определение; области применения; исходные данные.	3	2	1	Лекция, практикум.	Опорный конспект, электронный ресурс, система контроля знаний
7	Алгоритмы интеллектуального анализа данных: основные модели; классификация моделей;	2	1	1	Лекция	Опорный конспект, электронный ресурс

№	Тема занятий	Количество часов			Форма проведения	Оборудование
		всего	теория	практика		
	применение.					
ИТОГО:		17	9	8		

СОДЕРЖАНИЕ

Тема 1. Общие понятия интеллектуального анализа данных (3 ч)

Получение представления о том, что такое интеллектуальный анализ данных, для чего он используется, как реализуется, какие методы и задачи в себе содержит.

Виды уроков: Комбинированные.

Цели урока:

Знать

- основные понятия интеллектуального анализа данных;
- области применения интеллектуального анализа данных;
- основные средства интеллектуального анализа данных;
- **Основные понятия:**
- Интеллектуальный анализ данных, Задачи интеллектуального анализа данных, Алгоритмы интеллектуального анализа данных, СУБД.
- **Методические рекомендации:**
- Учитель повторяет с учениками понятие СУБД, ИС, анализ данных.
- **Задания:**
- Изучить информационный ресурс Data-Mining.
- **Вопросы:**
- Что такое Data-Mining?
- Кто впервые ввел термин «интеллектуальный анализ данных»?
- Цели и задачи интеллектуального анализа данных?
- Какие аналитические платформы вы знаете?

Тема 2. Задача классификации (2 ч)

Знакомство с понятием классификации.

Использование алгоритмов решения данной задачи на конкретных примерах.

Рассмотрение частного случая классификации – шкалирования.

Поиск основных отличий между классификацией и регрессией.

Прохождение тестовых заданий для закрепления полученных знаний.

Вид урока: Комбинированный.

Цели урока:

Знать

- Понятие классификации.
- Основные отличия между классификацией и регрессией.

Основные понятия:

- Классификация, шкалирование, СУБД.

Методические рекомендации:

- Учитель повторяет с учениками пройденный материал.
- Показывает пример решения задачи.

Задания:

- Рассмотрение частного случая классификации – шкалирования.
- Использование алгоритмов решения данной задачи на конкретных примерах.
- Изучить информационный ресурс [Data-Mining](#).

Вопросы:

- Что такое классификация?
- Отличие классификации от регрессии?
- Приведите примеры классификации?

Тема 3. Задача регрессии (2 ч)

Знакомство с понятием регрессии.

Использование алгоритмов решения данной задачи на конкретных примерах.

Поиск основных отличий между классификацией и регрессией.

Прохождение тестовых заданий для закрепления полученных знаний.

Вид урока: Комбинированный.

Цели урока:

Знать

- Понятие Регрессии.
- Виды регрессии.
- Основные отличия между регрессией и классификацией.

Основные понятия:

- Регрессия, СУБД, ИАД.

Методические рекомендации:

- Учитель повторяет с учениками пройденный материал.
- Показывает пример решения задачи.

Задания:

- Изучить информационный ресурс Data-Mining.

Вопросы:

- Что такое регрессия?
- Виды регрессии?
- Отличие регрессии от классификации?
- Приведите примеры регрессии?

Тема 4. Задача прогнозирования (3 ч)

Знакомство с понятием прогнозирования.

Изучение методов прогнозирования.

Знакомство с программными возможностями прогнозирования.

Вид урока: Комбинированный.

Цели урока:

Знать

- Задача прогнозирования.
- Методы прогнозирования.

Основные понятия:

- Прогнозирование, СУБД, ИАД.

Методические рекомендации:

- Учитель повторяет с учениками пройденный материал.
- Показывает пример прогнозирования в MS Excel.

Задания:

- Изучить информационный ресурс Data-Mining.
- Прогнозирование в среде MS Excel

Вопросы:

- Что такое прогнозирование?
- Примеры прогнозирования?

Тема 5. Задача кластеризации (2 ч)

Знакомство с понятием кластеризации.

Установление отличий между кластеризацией, классификаций, регрессией и прогнозированием.

Изучение общеизвестных примеров кластеризации.

Прохождение тестовых заданий для закрепления полученных знаний.

Вид урока: Комбинированный.

Цели урока:**Знать**

- Что такое кластеризация.
- Цели кластеризации.

Основные понятия:

- Кластеризация или кластерный анализ, СУБД, ИАД.

Методические рекомендации:

- Учитель повторяет с учениками пройденный материал.
- Показывает примеры кластеризации в различных областях.

Задания:

- Изучить информационный ресурс Data-Mining.
- Прохождение тестовых заданий

Вопросы:

- Что такое кластеризация?
- Общеизвестные примеры кластеризации?

Тема 6. Задача поиска ассоциативных правил (3 ч)

Знакомство с понятием ассоциации.

Рассмотрение областей применения поиска ассоциативных правил.

Анализ корректности исходных данных для решения задачи поиска ассоциативных правил.

Прохождение тестовых заданий для закрепления полученных знаний.

Вид урока: Комбинированный.

Цели урока:**Знать**

- понятие ассоциации.
- области применения поиска ассоциативных правил.

Основные понятия:

- ассоциативные правила, СУБД, ИАД.

Методические рекомендации:

- учитель повторяет с учениками пройденный материал.
- показывает «жизненные» примеры ассоциативных правил.

Задания:

- изучить информационный ресурс Data-Mining.
- прохождение тестовых заданий

Вопросы:

- Что такое поиск ассоциативных правил?
- Приведите примеры ассоциативных правил?

Тема 7. Исследование алгоритмов интеллектуального анализа данных

(2 ч)

Изучение основных моделей интеллектуального анализа данных:

- набора кластеров, описывающих взаимосвязи параметров в наборе данных;
- дерева решений, которое прогнозирует результат и описывает, какое влияние на этот результат оказывают разные критерии;
- математической модели, которая прогнозирует те или иные события;
- набора правил, описывающих группировку объектов в транзакции, а также вероятность группировки данных объектов.

Классификация данных моделей.

Использование изученных алгоритмов для решения практических задач.

ОЖИДАЕМЫЕ РЕЗУЛЬТАТЫ

В результате изучения факультативных курсов «Интеллектуальный анализ данных» у учащихся:

- увеличивается познавательная активность, формируется когнитивный интерес, развивается интеллектуальный и творческий потенциал;
- знания, связанные с содержанием базового курса информатики, расширяются и углубляются;
- ведется подготовка к систематическому изучению основ алгоритмизации и программирования;
- формируются практические навыки работы с прикладными программами;

- развивает алгоритмическое и логическое мышление;
- формируются навыки самостоятельной и творческой работы.

РЕКОМЕНДУЕМАЯ ЛИТЕРАТУРА

1. Агеева, И. Д. Занимательные материалы по информатике и математике : метод. пособие / И. Д. Агеева. — М. : «ТЦ Сфера», 2006.
2. Златопольский, Д. М. Сборник заданий для внеклассной работы по информатике / Д. М. Златопольский. — М. : Чистые пруды, 2006. — 30 с. — (Б-чка «Первого сентября». Информатика. — Вып. 1 (7)/2006).
3. Зубрилин, А. А. Игровой компонент в обучении информатике : практикум. — Ч. 2 / А. А. Зубрилин. — Саранск : Морд. гос. пед. ин-т, 2006. — 180 с.
4. Кузнецов, А. Т. Практические задания по информатике: 7—10 классы : пособие для учителей и учащихся / А. Т. Кузнецов, О. Н. Масленникова, М. С. Цветкова. — Минск : Юнипресс, 2005. — 96 с.
5. Павловский, А. И. Виды и формы организации самостоятельной работы на уроке информатики / А. И. Павловский, А. Е. Пупцев // Информатизация адукацыі. — 1998. — № 1. — С. 100—108.
6. Дюк В., Самойленко А., «Data Mining. Учебный курс» Спб:Питер, 2001. - 386с.
7. Паклин Н.Б., Орешков В.И. Бизнес-аналитика: от данных к знаниям / 2-е изд., испр. - СПб: Питер, 2013.
8. Чубукова И.А. Основы информационных технологий. Data Mining. - М.: БИНОМ - Лаборатория знаний, 2008.

2.3 Описание ресурса для изучения методов интеллектуального анализа данных

Электронный ресурс содержит все необходимые данные для изучения факультативного курса «Интеллектуальный анализ данных» согласно приведенной тематике занятий.

Электронный ресурс представляет собой HTML приложение, которое легко может быть запущено в любом интернет-браузере.

Главная страница ресурса представляет собой введение в интеллектуальный анализ данных – здесь можно ознакомиться с определением понятия Data Mining, предпосылки для развития интеллектуального анализа данных, общей информацией по теме (Рисунок 2).

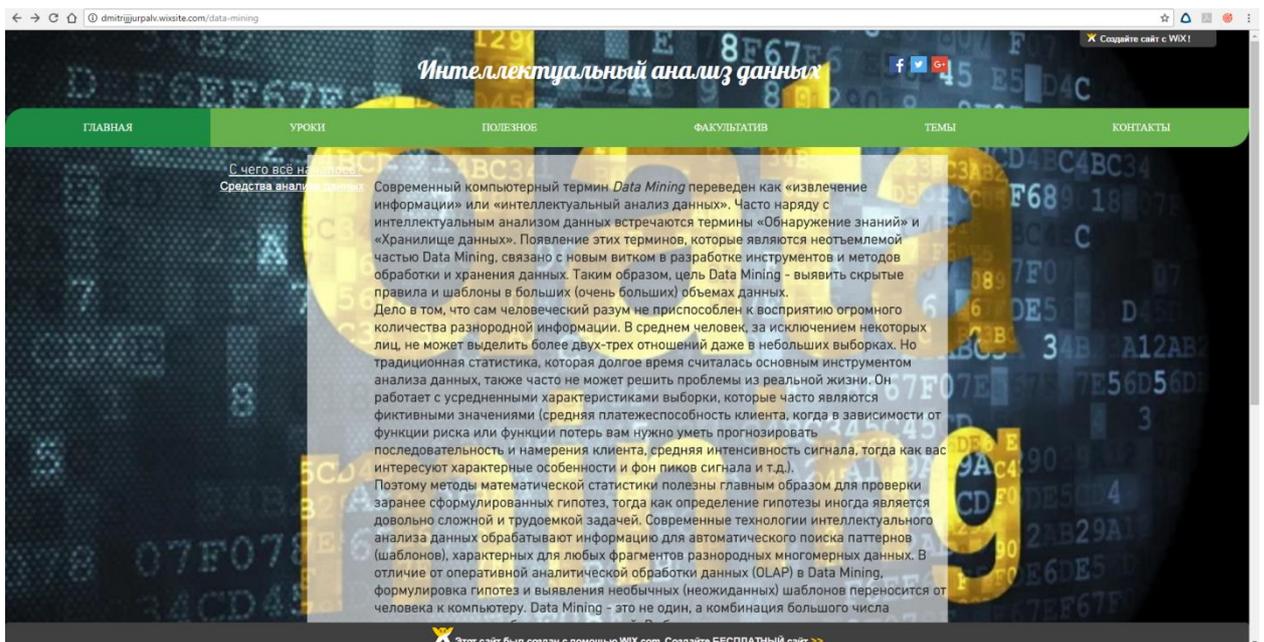


Рисунок 2 Главная страница электронного ресурса

Данная страница также содержит ссылку «С чего все начиналось» на дополнительную информацию по теме. Данная ссылка ведет на соответствующую страницу в Википедии, которая является проверенным ресурсом, следовательно, переход по ссылке безопасен для компьютера.

Ссылка «Средства анализа данных» обеспечивает переход в раздел «Полезное» электронного ресурса. Данный раздел содержит ссылки на страницы с описанием программных пакетов для интеллектуального анализа данных (рисунок 3).

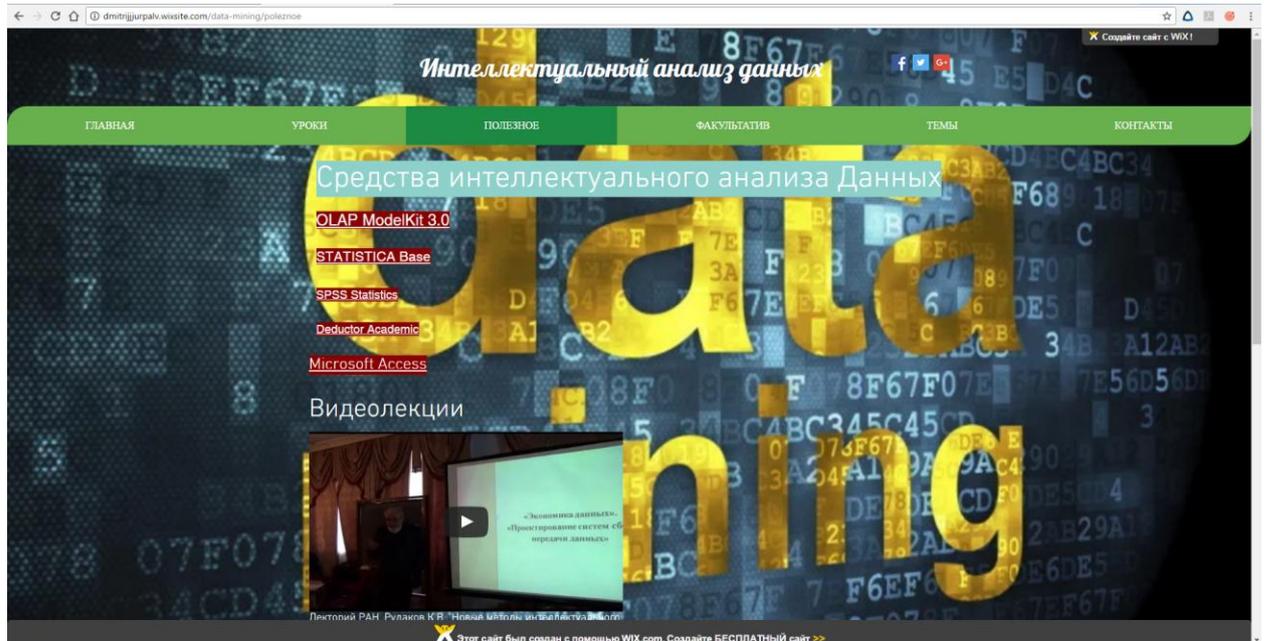


Рисунок 3 Электронный ресурс. Раздел «Полезное»

Данный раздел также содержит три видеолекции по теме интеллектуального анализа данных. Источником видеолекций является сайт YouTube, следовательно, данный материал также безопасен для просмотра.

Раздел «Уроки» электронного ресурса представляет собой перечень документов в формате MS Word, которые содержат подробные планы-конспекты факультативных занятий (приложение А).

Уроки можно скачать как по одному – щелчком на выбранном уроке, так и весь курс сразу – нажатием на кнопку «Скачать» (рисунок 4).

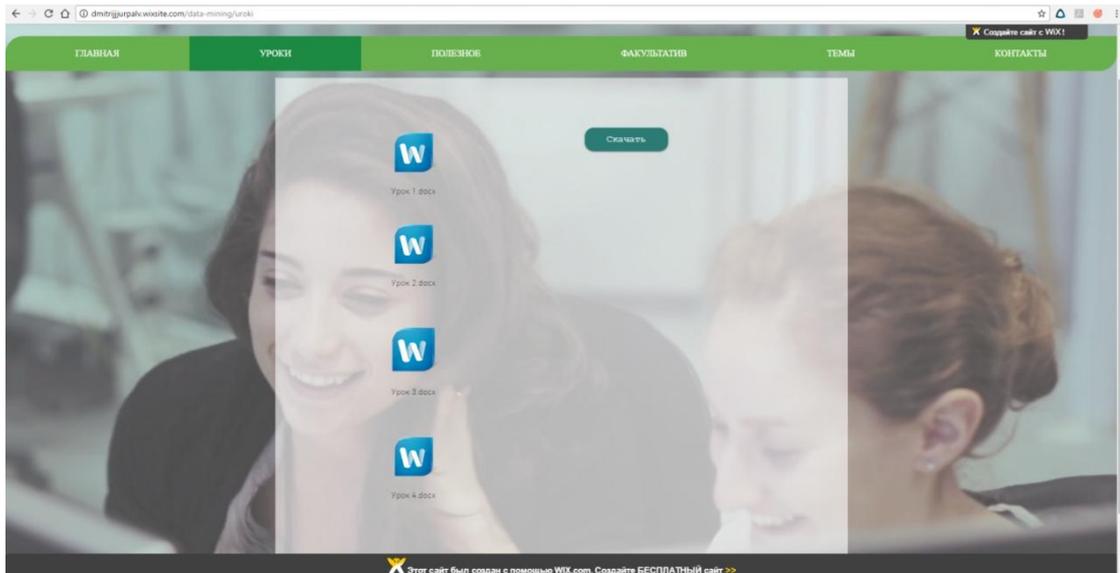


Рисунок 4 Электронный ресурс. Раздел «Уроки»

Раздел «Факультатив» электронного ресурса содержит таблицу с примерами применения интеллектуального анализа данных и таблицу с примерами задач, решаемых с помощью алгоритмов интеллектуального анализа данных (рисунок 5).

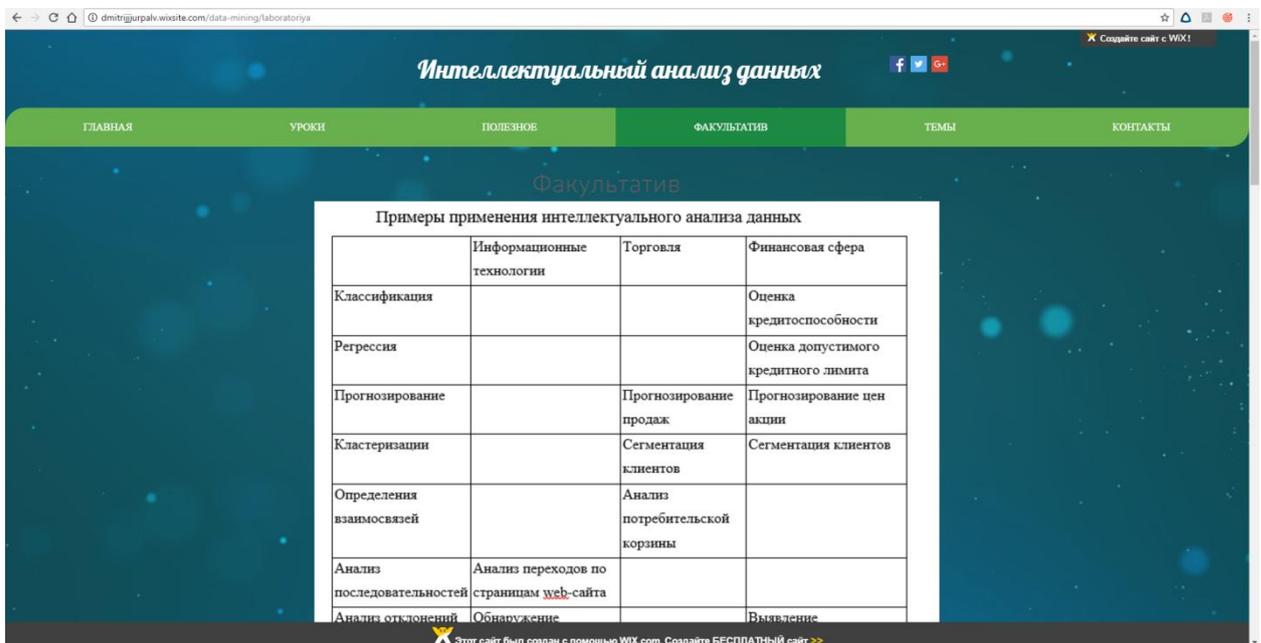


Рисунок 5 – Электронный ресурс. Раздел «Факультатив»

Данный раздел поможет учащимся определиться с выбором факультативного курса путем ознакомления с задачами, которые они смогут решать после его прохождения.

Раздел «Темы» содержит подразделы согласно темам занятий (рРисунок 6).

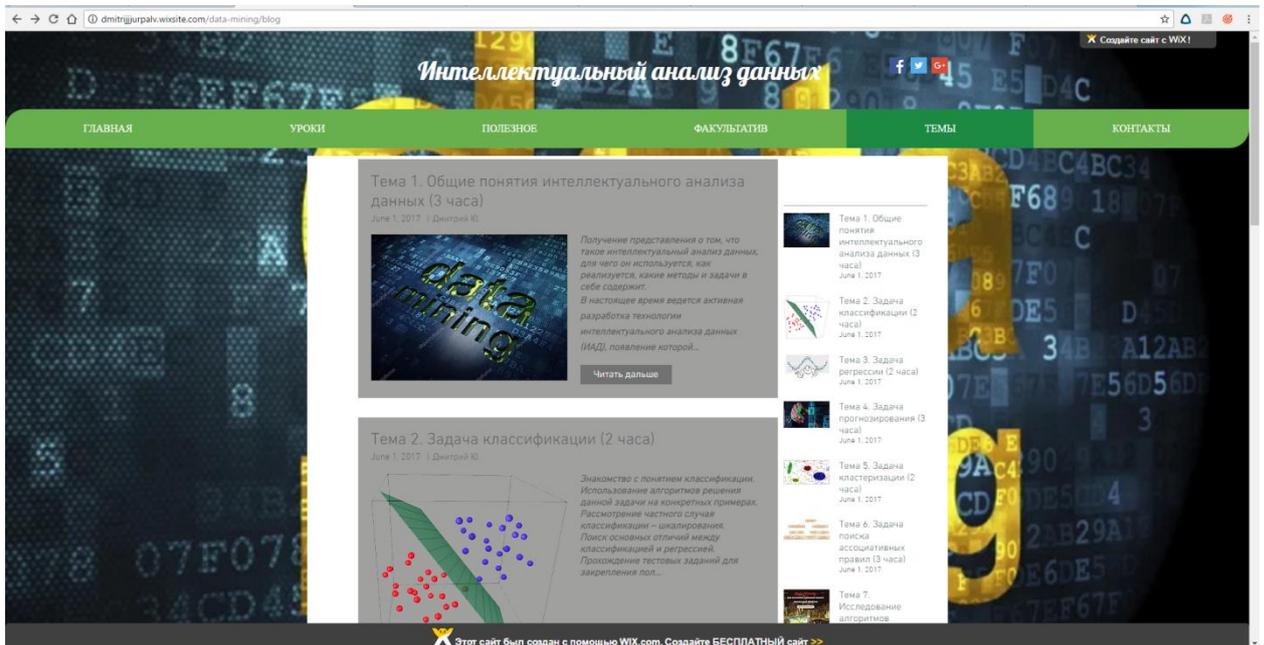


Рисунок 6 Электронный ресурс. Раздел «Темы»

Каждый из подразделов содержит следующую информацию:

- цели и задачи занятия;
- теоретические материалы по теме занятия;
- ссылки на дополнительные источники информации (тематические страницы Википедии).

Для перехода к подробному описанию темы необходимо нажать на кнопку «Читать дальше» в выбранной теме.

Навигация по темам факультативных занятий также может осуществляться с помощью панели навигации, расположенной справа.

Данная панель содержит иконку, соответствующую теме занятия, и название темы. Переход по ссылке откроет страницу ресурса с соответствующей темой (Рисунок 7 Электронный ресурс. Тема 1. Общие понятия интеллектуального анализа данных).



Рисунок 7 Электронный ресурс. Тема 1. Общие понятия интеллектуального анализа данных

С помощью панели навигации здесь также можно перейти к любой другой теме.

Раздел «Контакты» содержит форму обратной связи (Рисунок 8 Электронный ресурс. Раздел «Контакты»).

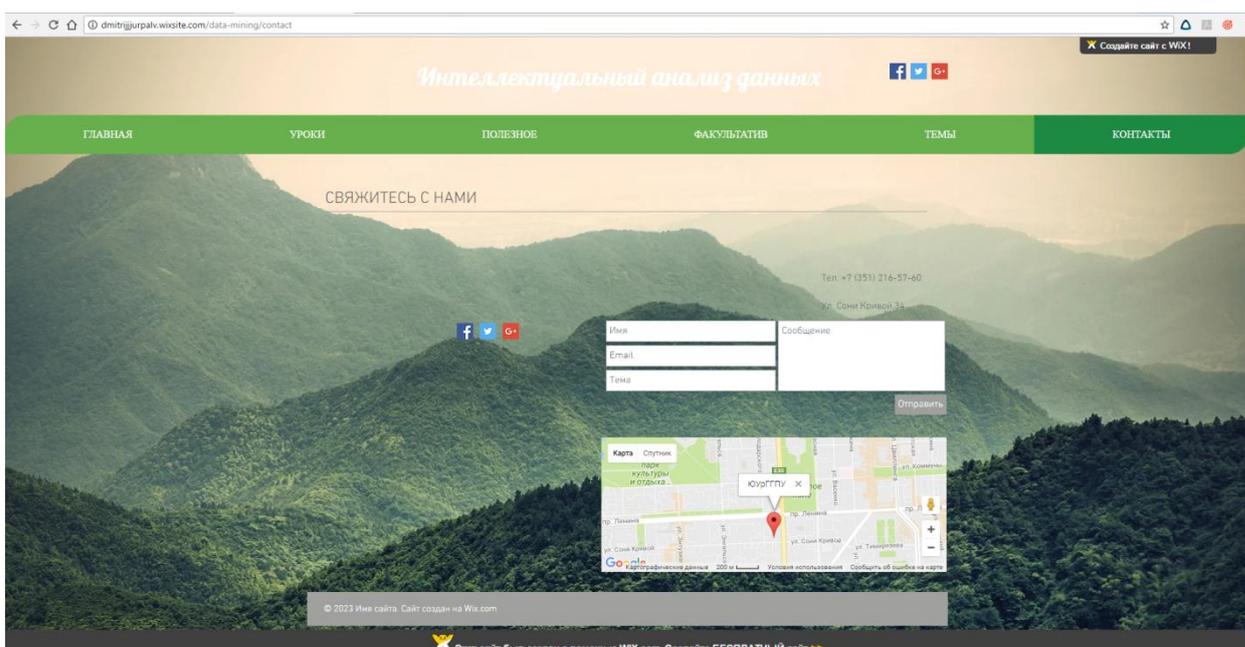


Рисунок 8 Электронный ресурс. Раздел «Контакты»
Данный ресурс предназначен как для учащихся, так и для учителя.

В дополнение к образовательному ресурсу разработан ресурс для закрепления знаний учащихся.

Для разработки ресурса использовалась СУБД Access.

Разработанный ресурс включает в себя три практических задания:

- два задания по теме «Задача классификации и регрессии»;
- одно задание по теме «Задача поиска ассоциативных правил».

Ресурс имеет главную форму с кнопками. Кнопки сгруппированы по занятиям для интуитивного понимания интерфейса программы (Рисунок).

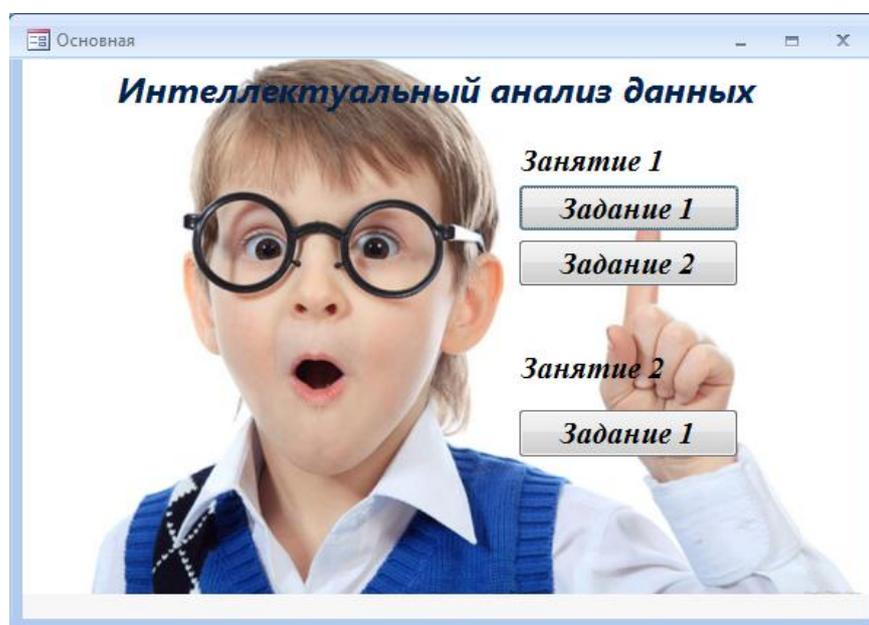


Рисунок 9 Главная форма ресурса

Первое занятие рассчитано на получение учащимся навыков в определении основных качеств для решения задачи классификации. Учащимся предлагается выбрать из списка качества, которые они считают важными при приеме на работу.

В результате выполнения программы учащийся получает список выбранных качеств с подробным пояснением, когда и как стоит выбирать то или иное качество, какие качества являются взаимоисключающими, какие качества должны всегда выбираться вместе.

Интерфейс окна для выбора качеств представлен на Рисунок .

Задание 1 - Учимся выбирать качества

Качество 1

Качество 2

Качество 3

Качество 4

Качество 5

Качество 6

Качество 7

Качество 8

Качество 9

Качество 10

Качество 11

Качество 12

Качество 13

Ознакомиться с результатом

Рисунок 10 Интерфейс окна для выбора качеств соискателя

Отчет, полученный в результате выполнения программы помогает учащемуся понять, какие из качеств он выбрал правильно, а где ошибся. Данное задание является познавательным.

Форма отчета представлена на Рисунок 11.

Наименование	Рекомендации
Есть высшее образование	Приоритет.
Нет проблем со здоровьем	Хорошо. Взаимоисключает качество "Есть проблемы со здоровьем".
Опыт работы более 3-х лет	Приоритет
Пол женский	Важным являются возраст и количество детей: если женщина старше 35 лет и без детей - приоритет; если моложе 35 лет и без детей - возможен декрет; если с детьми - возможны больничные. Использовать с качеством "Пол мужской"
Пол мужской	Обратите внимание на возраст. Использовать с качеством "Пол женский".

Рисунок 11 Форма отчета о выбранных качествах соискателя

Второе занятие направлено на развитие у учащегося понятия важности тех или иных качеств при принятии определенного решения.

Учащемуся необходимо расставить приоритеты возле каждого выбранного качества. Окно для простановки приоритетов представлено на Рисунок .

Качества

Задание 2 - Расставляем приоритеты

Наименование	Приоритет
Пол женский	4
Пол мужской	5
Есть высшее образование	4
Есть опыт работы	5
Опыт работы более 3-х лет	5
Есть второе высшее образование	3
Есть спортивный разряд	5
Замужем (женат)	5
Возраст более 35 лет	2
Возраст менее 35 лет	5
Есть дети	3
Нет проблем со здоровьем	5
Есть проблемы со здоровьем	5
Есть инвалидность	3
Нет инвалидности	5

Ознакомьтесь с результатами



Рисунок 12 Окно для выбора приоритетов определенных качеств при приеме на работу

В результате учащийся получает отчет, который сводно показывает, какие качества он считает важными, а какие – нет. Форма отчета представлена на Рисунок .

The screenshot shows a window titled "Приоритеты" (Priorities) with a main heading "Ознакомьтесь с результатами задания" (Get acquainted with the results of the task). Below the heading is a table with two columns: "Степень важности" (Degree of importance) and "Наименование" (Name). The table lists various qualities of an applicant, grouped into four categories of importance.

Степень важности	Наименование
Важно!	Пол мужской
	Опыт работы более 3-х лет
	Нет проблем со здоровьем
	Нет инвалидности
	Замужем (женат)
	Есть спортивный разряд
	Есть проблемы со здоровьем
	Есть опыт работы
	Возраст менее 35 лет
Важно, но не очень!	Пол женский
	Есть высшее образование
Можно пережить!	Есть инвалидность
	Есть дети
	Есть второе высшее образование
Совсем вообще не важно!	Возраст более 35 лет

Рисунок 13 Форма отчета о приоритетах качеств соискателя

Третье занятие предназначено для развития у учащегося ассоциативного мышления.

Учащемуся предлагается наполнить три потребительские корзины различными товарами (Рисунок).

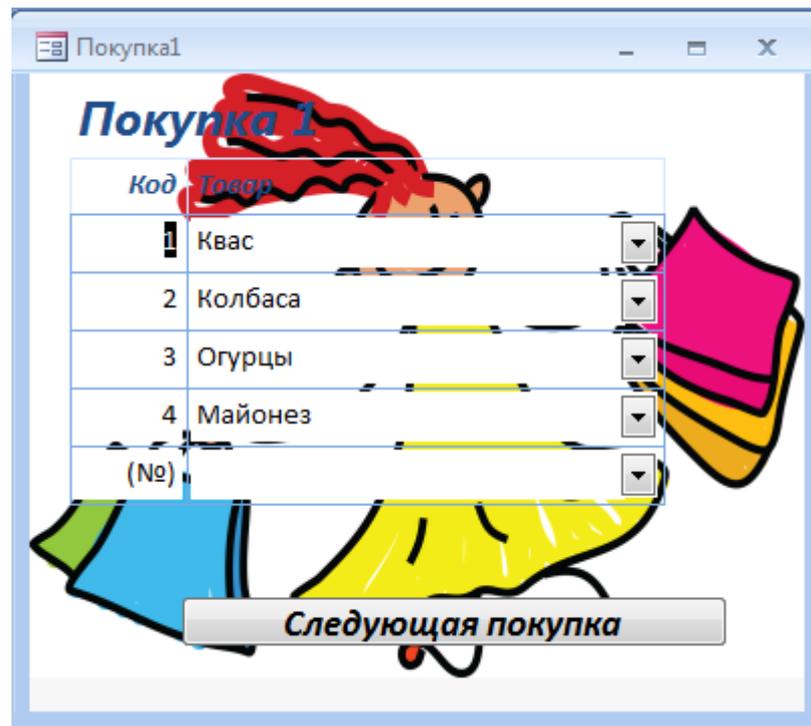


Рисунок 14 Форма для наполнения потребительской корзины товарами
 В результате учащийся получает отчет, который показывает ему, какие из выбранных товаров являются наиболее популярными, а какие из них чаще всего покупают вместе (Рисунок).

The screenshot shows a window titled "Покупка1 Запрос_перекрестный" with a table. The table has the following columns: "Наименование", "Покупка 1", "Покупка 2", "Покупка 3", and "Рейтинг товаров".

Наименование	Покупка 1	Покупка 2	Покупка 3	Рейтинг товаров
Квас	1	1		2
Колбаса	1	1		2
Майонез	1		1	2
Огурцы	1		1	2
Хлеб			1	1
Яйца		1		1
Лук		1		1
Помидоры			1	1

Рисунок 15 Форма отчета о потребительской корзине покупателей

По данному отчету учащемуся необходимо определить, какие товары наиболее часто покупаются вместе.

Разработанный ресурс интуитивно понятен, имеет красочный интерфейс, все задания, заключения и рекомендации в нем изложены на доступном языке.

2.4 Результаты апробации факультативного курса в школе

Апробация данного курса в школе показала, что данный курс интересен учащимся 10-11 классов, для которых он предназначен.

Материал курса изложен на доступном языке, иллюстрации и таблицы помогают визуализировать процесс изучения материала, что способствует его успешному освоению.

В материале курса научные определения и термины подкрепляются примерами из повседневной жизни учащихся, что делает данный курс доступным для понимания.

Выводы по главе 2

Анализ учебников и учебных пособий по информатике показал, что на сегодняшний день нет достаточной информации по разделу информатики «Интеллектуальный анализ данных». Вследствие этого был разработан тематический факультативный курс.

Данный курс содержит все необходимые разделы по изучению указанной темы. Для данного курса разработано тематическое планирование с указанием количества теоретических и практических занятий.

Для эффективного изучения курса был разработан электронный ресурс, содержащий необходимые учебные материалы. Ресурс имеет интуитивно понятный интерфейс и может быть доступен как для применения на факультативных занятиях, так и для самостоятельного изучения курса.

ЗАКЛЮЧЕНИЕ

Данная работа была посвящена разработке методики изучения алгоритмов интеллектуального анализа данных.

Данные алгоритмы в настоящее время широко используются при принятии решений и прогнозировании тех или иных событий на основе накопленного опыта.

В рамках интеллектуального анализа данных решаются следующие задачи:

- задача классификации и регрессии;
- задача прогнозирования;
- задача кластеризации;
- задача поиска ассоциативных правил.

Исходными данными для решения этих задач являются массивы информации, анализ которых помогает выявлять определенные зависимости, анализировать данные, принимать определенные решения на основании полученных результатов.

Изучение алгоритмов интеллектуального анализа данных является достаточно важным в современных условиях. Т.к. данные методы не включены в основную программу курса «Информатика», необходимо разработать план индивидуальных занятий, охватывающих данную тему.

В рамках разработки факультативных занятий по дисциплине "Информатика" с уклоном на изучение алгоритмов интеллектуального анализа данных были рассмотрены принципы организации внеурочной деятельности учащихся в условиях ФГОС.

По результатам была составлена программа факультативных занятий, которая рассчитана на 17 часов и предназначена для учащихся 10-11 классов. В работе представлен тематический план курса, сформулированы цели и

задачи, обозначены ожидаемые результаты, приведен список рекомендованной литературы.

Для изучения алгоритмов интеллектуального анализа данных разработан электронный ресурс, содержащий подробную информацию по курсу согласно представленной программе занятий. Для практической части курса разработан ресурс с заданиями по таким темам как задача классификации и задача поиска ассоциативных правил.

Данный ресурс помогает учащимся в освоении дисциплины и дать возможность самостоятельного изучения информации по курсу «Интеллектуальный анализ данных».

Апробация данного курса в школе показала, что он интересен учащимся 10-11 классов, для которых предназначен.

В приложении к работе представлены планы-конспекты двух факультативных занятий.

СПИСОК ЛИТЕРАТУРЫ

1. Абдикеев Н.М. Данько Т.П. Ильдеменов С.В. Киселев А.Д. «Реинжиниринг бизнес-процессов. Курс МВА», М.: Изд-во Эксмо, 2005. - 592 с.
2. Абдикеев Н.М., Киселев А.Д. «Управление знаниями в корпорации и реинжиниринг бизнеса» - М.:Инфра-М, 2011.- 382 с.
3. Барсегян А.А., Куприянов М.С., Степаненко В.В., Холод И.И. «Методы и модели анализа данных: OLAP и Data Mining», Спб:БХВ-Петербург, 2004. - 336с.
4. Дюк В.,Самойленко А., «Data Mining. Учебный курс» Спб:Питер, 2001. - 386с.
5. Зайцев К.С. Применение методов Data Mining для поддержки процессов управления ИТ-услугами: учебное пособие. - М.: МИФИ, 2009.
6. Паклин Н.Б., Орешков В.И. Бизнес-аналитика: от данных к знаниям / 2-е изд., испр. - СПб: Питер, 2013.
7. Поляков К.Ю., Еремин Е.А. Информатика. 11 класс. Углубленный уровень. В 2 ч. / Ч.1. - М.: БИНОМ. Лаборатория знаний, 2013. - 240 с.
8. Поляков К.Ю., Еремин Е.А. Информатика. 11 класс. Углубленный уровень. В 2 ч. / Ч.2. - М.: БИНОМ. Лаборатория знаний, 2013. - 304 с.
9. Приказ Министерства образования и науки РФ от 17 мая 2012 г. N 413 (ред. от 29.12.2014) "Об утверждении федерального государственного образовательного стандарта среднего общего образования"
10. Приказ Минобразования РФ от 09.03.2004 N 1312 (ред. от 20.08.2008) "Об утверждении федерального базисного учебного плана и примерных учебных планов для образовательных учреждений Российской федерации, реализующих программы общего образования"

11. Приказ Минобрнауки России от 31 марта 2014 г. № 253 «Об утверждении федерального перечня учебников, рекомендуемых к использованию при реализации имеющих государственную аккредитацию образовательных программ начального общего, основного общего, среднего общего образования» (с изменениями на 26 января 2016 года).

12. Ривкин Е.Ю. Внеурочная деятельность в условиях внедрения ФГОС ООО: целевые установки и организационные формы //Администратор образования. — 2012. - № 20. - С. 75 — 84.

13. Самылкина Н. Н., Калинин И. А. Информатика. 10–11 классы. Углубленный уровень: методическое пособие / И.А. Калинин, Н.Н. Самылкина - М.: БИНОМ. Лаборатория знаний, 2015.

14. Хуторской, А.В. Ключевые компетенции как компонент личностно-ориентированной парадигмы образования // Народное образование. - 2003. - № 2. - С. 58-64.

15. Чубукова И.А. Основы информационных технологий. Data Mining. - М.: БИНОМ - Лаборатория знаний, 2008.

Приложения

План-конспект факультативного занятия по методам интеллектуального анализа данных (занятие 1)

Тема: Задача классификации.

Цель: расширить и закрепить знания учащихся о таком методе интеллектуального анализа данных, как классификация.

Задачи:

1. Сформировать знания учащихся о классификации и ее частном случае - шкалировании.
2. Продолжить формирование умения переходить от общего к частному и наоборот.
3. Расширить кругозор учащихся.

Оборудование: иллюстрации, электронный ресурс для самопроверки.

Ход урока

Организационный момент.

Здравствуйте! Сегодняшнее занятие будет посвящено изучению такой задачи интеллектуального анализа данных, как классификация. Умение решать данную задачу поможет вам в будущем при принятии решений.

Формирование знаний учащихся

Для начала необходимо познакомиться с понятием классификации.

Классификация — это осмысленный порядок вещей, явлений, разделение их на разновидности согласно каким-либо важным признакам.

Рассмотрим данные понятия на определенных примерах.

Допустим, в нашу школу приехала комиссия по отбору участников для передачи «Умники и умницы».

В школе имеется два достойных кандидата, но для участия в передаче нужно выбрать только одного.

Рассмотрим пример выбора участника с применением классификации.

Вот наши кандидаты!

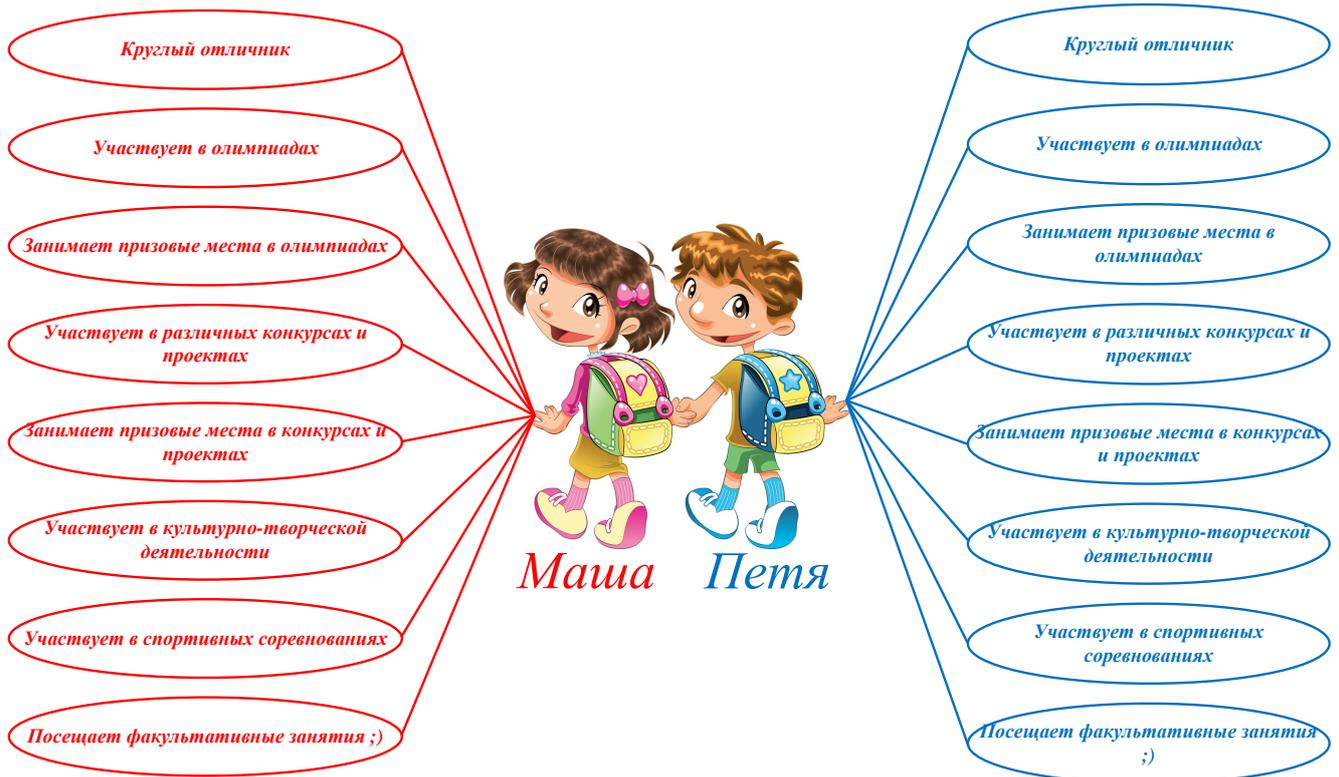


Маша

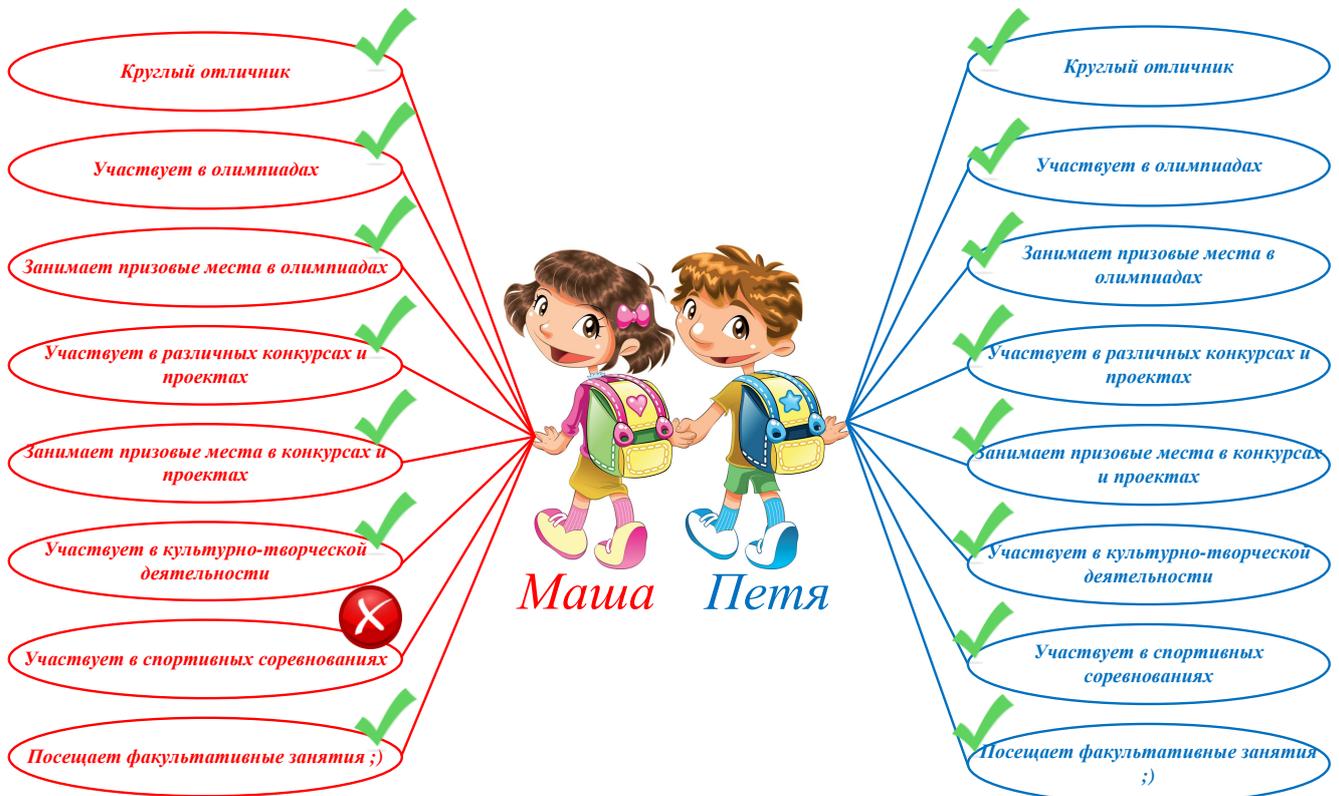
Петя

Каждый из них, несомненно, имеет определенные качества, оценив которые можно решить, кто из них поедет участвовать в передаче.

Обозначим эти качества.



Теперь отметим те качества, которые присущи каждому из счастливых претендентов, т.е просто ответим на вопрос «Да» или «Нет».



В итоге счастливым претендентом становится Петя, т.к. не имеет ни одного отрицательного качества.

Но правы ли мы? Так ли важно в данном случае, участвует кто-то из ребят в спортивных соревнованиях или нет? Передача-то интеллектуальная...

Здесь в определение наиболее достойного кандидата вступает ШКАЛИРОВАНИЕ.

Шкала порядка (порядковая, ранговая, ординальная) предназначена для измерения (обозначения) степени различия какого-либо признака или свойства у разных объектов. Самым ярким примером порядковой шкалы является пятибалльная система оценки знаний учащихся. Интервалы между отдельными баллами не отражают разрыва между реальными результатами. Мы знаем только, что ученик, получивший оценку "5" по какому-то предмету, знает этот предмет лучше того, кто получил "4". Но нельзя утверждать, что различие в знаниях этих учащихся такое же, как и в знаниях тех, кто получил "4" и "3".

Каждому из качеств присвоим весовой коэффициент по мере значимости данного качества для участия в интеллектуальной передаче. Возьмем привычную шкалу от 2 до 5 – самые важные качества получают «пятерку», самые неважные – «двойку».

Расставим оценки:

Круглый отличник – «важно» - 5;

Участвует в олимпиадах – «важно» - 5;

Занимает призовые места в олимпиадах – «важно» - 5;

Участвует в различных конкурсах и проектах – «важно» - 5;

Занимает призовые места в конкурсах и проектах – «важно» - 5;

Участвует в культурно-творческой деятельности – «не важно» - 2;

Участвует в спортивных соревнованиях – «не важно» - 2;

Посещает факультативные занятия – «важно, но не очень» - 4.

Теперь посчитаем количественные показатели для каждого из качеств наших претендентов, т.е. сколько раз они участвовали в олимпиадах, сколько заняли призовых мест и т.д. Результаты сведем в таблицу:

	Оценка	Маша	Петя
Круглый отличник	5	1	1
Участвует в олимпиадах	5	4	2
Занимает призовые места в олимпиадах	5	1	2
Участвует в различных конкурсах и проектах	5	5	4
Занимает призовые места в конкурсах и проектах	5	3	2
Участвует в культурно-творческой деятельности	2	3	7
Участвует в спортивных соревнованиях	2	0	3
Посещает факультативные занятия	4	4	6

По данным таблицы путем нехитрых вычислений, а именно умножением оценки качества на его количество, получаем результаты:

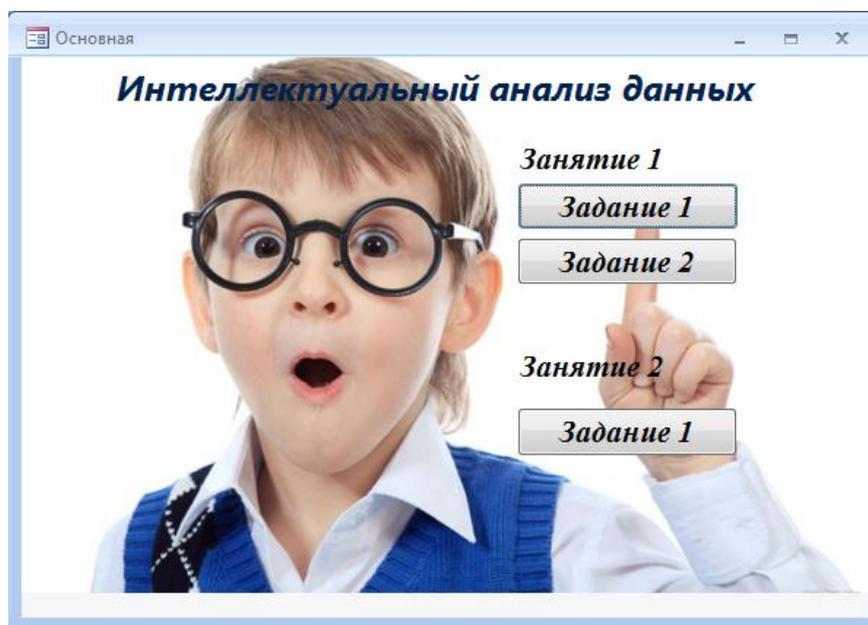
	Оценка	Маша	Петя
Круглый отличник	5	5	5
Участвует в олимпиадах	5	20	10
Занимает призовые места в олимпиадах	5	5	5
Участвует в различных конкурсах и проектах	5	25	20
Занимает призовые места в конкурсах и проектах	5	15	5
Участвует в культурно-творческой деятельности	2	6	14
Участвует в спортивных соревнованиях	2	0	6
Посещает факультативные занятия	4	16	24
ИТОГО:		92	89

По итогам данной таблицы на съемки передачи едет Маша, т.к. имеет больше научных достижений, чем Петя (хоть он и спортсмен, и активист).

Задания

Задание 1.

Теперь проверим, как усвоен материал. Вам предлагается пройти тестовое задание. Запускаем файл «Факультатив» на рабочем столе и в группе «Занятие 1» выбираем «Задание 1». Здесь вам необходимо выбрать качества, которые вы считаете важными при приеме на работу.



Будьте внимательны, т.к. некоторые из предложенных качеств взаимоисключаемы, а некоторые – всегда должны использоваться вместе.

Задание 1

Задание 1 - Учимся выбирать качества

Качество 1

Качество 2

Качество 3

Качество 4

Качество 5

Качество 6

Качество 7

Качество 8

Качество 9

Качество 10

Качество 11

Качество 12

Качество 13

Ознакомиться с результатом

После того, как вы выберете качества вашего нового сотрудника, нужно нажать кнопку «Ознакомиться с рекомендациями» и оценить результаты своих трудов.

Наименование	Рекомендации
Есть высшее образование	Приоритет.
Нет проблем со здоровьем	Хорошо. Взаимоисключает качество "Есть проблемы со здоровьем".
Опыт работы более 3-х лет	Приоритет
Пол женский	Важным являются возраст и количество детей: если женщина старше 35 лет и без детей - приоритет; если моложе 35 лет и без детей - возможен декрет; если с детьми - возможны больничные. Использовать с качеством "Пол мужской"
Пол мужской	Обратите внимание на возраст. Использовать с качеством "Пол женский".

Задание 2.

А теперь попробуем расставить приоритеты на качества нашего нового сотрудника.

В группе «Занятие 1» выбираем «Задание 2». Здесь вам необходимо оценить качества, которыми вы располагаете при приеме сотрудника на работу.

Качества

Задание 2 - Расставляем приоритеты

Наименование	Приоритет
Пол женский	4
Пол мужской	5
Есть высшее образование	4
Есть опыт работы	5
Опыт работы более 3-х лет	5
Есть второе высшее образование	3
Есть спортивный разряд	5
Замужем (женат)	5
Возраст более 35 лет	2
Возраст менее 35 лет	5
Есть дети	3
Нет проблем со здоровьем	5
Есть проблемы со здоровьем	5
Есть инвалидность	3
Нет инвалидности	5

Ознакомьтесь с результатами



После того, как вы оцените качества вашего нового сотрудника, нужно нажать кнопку «Ознакомьтесь с результатами» и оценить результаты своих трудов.

Приоритеты	
Ознакомьтесь с результатами задания	
Степень важности	Наименование
Важно!	<ul style="list-style-type: none"> Пол мужской Опыт работы более 3-х лет Нет проблем со здоровьем Нет инвалидности Замужем (женат) Есть спортивный разряд Есть проблемы со здоровьем Есть опыт работы Возраст менее 35 лет
Важно, но не очень!	<ul style="list-style-type: none"> Пол женский Есть высшее образование
Можно пережить!	<ul style="list-style-type: none"> Есть инвалидность Есть дети Есть второе высшее образование
Совсем вообще не важно!	<ul style="list-style-type: none"> Возраст более 35 лет

Подведение итогов урока.

Подведем итоги нашего занятия.

Мы ознакомились с двумя важными понятиями, используемыми при интеллектуальном анализе данных – классификация и ее составляющая - шкалирование.

Надеюсь, что приведенный в начале занятия пример наглядно показал вам, каким образом использовать данные методы.

Считаю, что полученные на занятии знания помогут вам при принятии решений в реальной жизни.

План-конспект факультативного занятия по методам интеллектуального анализа данных (занятие 2)

Тема: Задача поиска ассоциативных правил.

Цель: расширить и закрепить знания учащихся о таком методе интеллектуального анализа данных, как поиск ассоциативных правил.

Задачи:

1. Сформировать знания учащихся о поиске ассоциативных правил.
2. Продолжить формирование умения переходить от общего к частному и наоборот.
3. Расширить кругозор учащихся.

Оборудование: иллюстрации, электронный ресурс для самопроверки.

Ход урока

Организационный момент.

Здравствуйте! Сегодняшнее занятие будет посвящено изучению такой задачи, как поиск ассоциативных правил. Умение решать данную задачу поможет вам в будущем при прогнозировании тех или иных процессов.

Формирование знаний учащихся

Для начала необходимо познакомиться с понятием поиска ассоциативных правил.

Ассоциация - одна из задач Data Mining. Целью поиска ассоциативных правил является нахождение закономерностей между связанными событиями в базах данных.

Очень часто покупатели приобретают не один товар, а несколько. В большинстве случаев между этими товарами существует взаимосвязь. Так, например, покупатель, приобретающий макаронные изделия, скорее всего, захочет приобрести также кетчуп. Эта информация может быть использована для размещения товара на прилавках.

Часто встречающиеся приложения с применением ассоциативных правил:

– розничная торговля: определение товаров, которые стоит продвигать совместно; выбор местоположения товара в магазине; анализ потребительской корзины; прогнозирование спроса;

– перекрестные продажи: если есть информация о том, что клиенты приобрели продукты А, Б и В, то какие из них вероятнее всего купят продукт Г?

– маркетинг: поиск рыночных сегментов, тенденций покупательского поведения;

– оформление каталогов, анализ сбытовых кампаний фирмы, определение последовательностей покупок клиентов (какая покупка последует за покупкой товара А).

Приведем простой пример ассоциативного правила: покупатель, приобретающий банку краски, приобретет кисточку для краски с вероятностью 50%.

Рассмотрим данное понятие на определенных примерах.

Допустим, что мы – маркетологи. Наша задача изучить спрос покупателей и расставить товары на полках так, чтобы увеличить продажи.

Мы берем несколько чеков из мусорной корзины около кассы и составляем сводную таблицу о покупках по данным чекам:

	Чек 1	Чек 2	Чек 3
Кефир	v		v
Мука	v	v	
Яйца	v		v
Молоко		v	
Сахар		v	
Клубничный джем		v	
Французский батон			v

Анализируя данную таблицу, можно сделать вывод, что муку покупают чаще всего с яйцами и молоком или кефиром. Почему?

Потому что муку не намажешь на французский батон, но французский батон из муки вполне можно испечь.

Следовательно, муку и яйца нужно разместить недалеко друг от друга на прилавках нашего магазина.

Подойдем к проблеме с другой сторон. Какова вероятность того, что при покупке яиц покупатель купит муку? Все верно - никакой! Скорее он купит бекон или помидоры.

Следовательно, приведенного примера нам недостаточно, чтобы реально оценить механизм совместных покупок.

Здесь же стоит отметить, что поиск ассоциативных правил наиболее эффективен при больших количествах исходных данных, где зависимости между совместными покупками намного нагляднее.

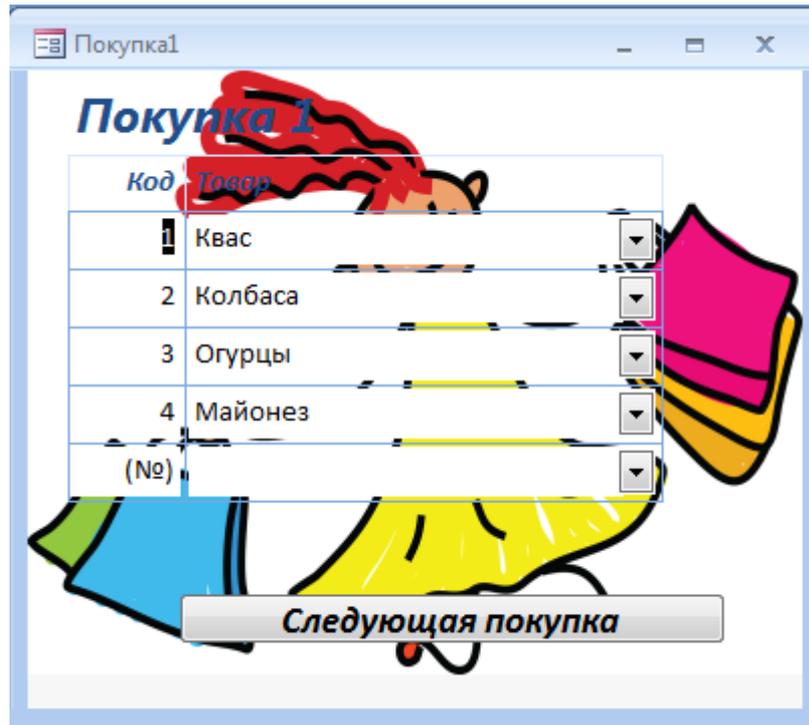
Чем меньше данных мы анализируем, тем более расплывчатым получается результат проведенного анализа.

Задания

Задание 1.

Однако представим себя владельцами небольшого продуктового магазина, например, «В шаге от дома». Наш магазин имеет небольшой ассортимент, но нам хотелось бы его расширить. Как же нам понять, что стоит привезти в наш магазин, а с чем не стоит даже связываться?

Вам предлагается пройти тестовое задание. Запускаем файл «Факультатив» на рабочем столе и в группе «Занятие 2» выбираем «Задание 1». Здесь вам необходимо наполнить покупательскую корзину предложенными в нашем магазине товарами.



Сделать это предлагается трижды. В результате вами будет получен отчет о рейтинге товаров нашего магазина и таблица совместных покупок.

Покупка1 Запрос_перекрестный

Данная таблица поможет Вам составить ассоциативные правила для сделанных покупок

Наименование	Покупка 1	Покупка 2	Покупка 3	Рейтинг товаров
Квас	1	1		2
Колбаса	1	1		2
Майонез	1		1	2
Огурцы	1		1	2
Хлеб			1	1
Яйца		1		1
Лук		1		1
Помидоры			1	1

По данным полученной таблицы предлагаю наиболее часто покупаемые продукты размещать ближе друг к другу на полках, а также

закупать их в достаточном количестве, для того чтобы предупредить их дефицит и, как следствие, потерю прогнозируемой прибыли.

Подведение итогов урока.

Подведем итоги нашего занятия.

Мы познакомились с новым понятием в составе интеллектуального анализа данных – поиск ассоциативных правил.

Надеюсь, что рассмотренная тема вызвала у вас интерес и желание продолжить исследования путем составления ассоциаций.

Считаю, что полученные на занятии знания вы уже в ближайшем будущем сможете применить на реальных примерах.

Спасибо за внимание!