

Министерство образования и науки Российской Федерации
Федеральное государственное автономное образовательное учреждение
высшего образования
«Уральский федеральный университет
имени первого Президента России Б.Н. Ельцина»

Институт Высшая школа экономики и менеджмента
Кафедра Анализа систем и принятия решений

ДОПУСТИТЬ К ЗАЩИТЕ ПЕРЕД ГЭК

И. о. зав. кафедрой д.ф.-м.н. Сеницын Е.В.

«6» июня 2018 г.

МАГИСТЕРСКАЯ ДИССЕРТАЦИЯ

Применение искусственного интеллекта при обработке анкетных данных

Руководитель:

к. т. н., доцент

Лапшина С. Н.

Консультант:

Старший преподаватель

Толмачев А.В.

Нормоконтролер:

к.ф.-м.н., доцент

Медведева М. А.

Студент группы ЭММ-261602

Рытова Т.А.

Екатеринбург
2018

Министерство образования и науки Российской Федерации
Федеральное государственное автономное образовательное учреждение
высшего образования
«Уральский федеральный университет имени первого Президента России Б.Н.Ельцина»

Институт Высшая школа экономики и менеджмента
Кафедра Анализа систем и принятия решений
Направление подготовки 09.04.03 Прикладная информатика
Образовательная программа ИТ-инновации в бизнесе

УТВЕРЖДАЮ
И. о. зав. кафедрой АСИПР
Синицын Е. В.
«20» марта 2018 г.

ЗАДАНИЕ
на выполнение магистерской диссертации

студента Рытовой Тамары Андреевны группы ЭММ-261602
(фамилия, имя, отчество)

1 Тема магистерской диссертации Применение искусственного интеллекта при обработке анкетных данных

Утверждена распоряжением по институту от «11» декабря 2017 г. № 33.01-05/344

2 Руководитель Лапшина Светлана Николаевна, доцент, кандидат технических наук
(Ф.И.О., должность, ученое звание, ученая степень)

3 Исходные данные к работе общедоступные литературные источники, техническая документация, нормативные документы предприятия, электронные ресурсы

4 Содержание пояснительной записки (перечень подлежащих разработке вопросов)

Введение

Глава 1. Обработка данных с использованием систем искусственного интеллекта

Глава 2. Разработка методики использования систем искусственного интеллекта при обработке анкетных данных

Глава 3. Применение систем искусственного интеллекта при сборе и обработке анкетных данных

Заключение

Список использованных источников

5 Перечень демонстрационных материалов Электронная презентация, выполненная в MS PowerPoint, раздаточный материал (печатный)

6 Консультанты по проекту (работе) с указанием относящихся к ним разделов проекта*

Раздел	Консультант	Подпись, дата	
		задание выдал	задание принял
Глава 3, подраздел 4	Толмачев А.В.	08.03.18	01.06.18

7 Календарный план

Наименование этапов выполнения работы	Срок выполнения этапов работы	Отметка о выполнении
1 глава	до 9 апреля 2018 г.	выполнено
2 глава	до 7 мая 2018 г.	выполнено
3 глава	до 21 мая 2018	выполнено
Магистерская диссертация в целом	до 1 июня 2018 г.	выполнено

Руководитель _____
подпись

Лапшина С. Н.

Задание принял к исполнению _____
подпись

Рытова Т.А.

8 Магистерская диссертация закончена «1» июня 2018 г.

Пояснительная записка и все материалы просмотрены

Оценка консультантов: * а) Толмачев А.В. _____ б) _____
в) _____ г) _____

Считаю возможным допустить Рытову Тamarу Андреевну к защите его магистерской диссертации в экзаменационной комиссии.

Руководитель _____
подпись

9 Допустить Рытову Тamarу Андреевну к защите магистерской диссертации в экзаменационной комиссии (протокол заседания кафедры № 6 от «6» июня 2018 г.)

И.о. зав. кафедрой _____
подпись

Синицын Е. В.

* - при наличии разделов, требующие привлечение консультантов

РЕФЕРАТ

Тема магистерской диссертации: Применение искусственного интеллекта при обработке анкетных данных.

Магистерская диссертация выполнена на 98 страницах, содержит 13 таблиц, 30 рисунков, 62 использованных источника.

Актуальность темы обусловлена большими трудозатратами и нерелевантными результатами обработки анкетных данных.

Целью работы является автоматизация процесса отбора анкетных данных в дистрибутиве Python Anaconda с использованием алгоритмов машинного обучения.

Задачи работы:

- изучить системы искусственного интеллекта;
- рассмотреть программное обеспечение для систем искусственного интеллекта;
- создать и обучить классификатор для сортировки анкетных данных;
- оценить экономическую эффективность создания проекта.

Объект исследования – система сбора и обработки анкетных данных отдела диспетчеризации ВШЭМ УрФУ.

Предмет исследования – автоматизация процесса ранжирования анкетных данных по релевантности.

В первой главе рассматривается обработка данных с использованием систем искусственного интеллекта.

Вторая глава посвящена разработке методики использования систем искусственного интеллекта при обработке анкетных данных.

В третьей главе представлены системы искусственного интеллекта при сборе и обработке анкетных данных

Результаты работы: практическим результатом работы стал разработанный классификатор, который определяет для заполненной анкеты: будет ли она учтена для анализа эффективности учебного процесса.

СОДЕРЖАНИЕ

ВВЕДЕНИЕ.....	6
1 ОБРАБОТКА ДАННЫХ С ИСПОЛЬЗОВАНИЕМ СИСТЕМ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА.....	9
1.1 СИСТЕМЫ СБОРА И ОБРАБОТКИ АНКЕТНЫХ ДАННЫХ.....	9
1.2 СИСТЕМЫ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА.....	27
1.3 ПРОГРАММНОЕ ОБЕСПЕЧЕНИЕ ДЛЯ СИСТЕМ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА.....	40
1.4 ВЫВОДЫ ПО РАЗДЕЛУ.....	46
2 РАЗРАБОТКА МЕТОДИКИ ИСПОЛЬЗОВАНИЯ СИСТЕМ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА ПРИ ОБРАБОТКЕ АНКЕТНЫХ ДАННЫХ.....	47
2.1 ПРИМЕНЕНИЕ СИСТЕМ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА ПРИ СБОРЕ И ОБРАБОТКЕ ДАННЫХ.....	47
2.2 СИСТЕМЫ СБОРА И ОБРАБОТКИ ДАННЫХ.....	49
2.3 ВЫВОДЫ ПО РАЗДЕЛУ.....	55
3 ПРИМЕНЕНИЕ СИСТЕМ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА ПРИ СБОРЕ И ОБРАБОТКЕ АНКЕТНЫХ ДАННЫХ.....	57
3.1 ОБЩАЯ ХАРАКТЕРИСТИКА ОРГАНИЗАЦИИ.....	57
3.2 СИСТЕМА СБОРА И ОБРАБОТКИ ДАННЫХ ОПРОСА СТУДЕНТОВ «AS-IS».....	76
3.3 СИСТЕМА СБОРА И ОБРАБОТКИ ДАННЫХ ОПРОСА СТУДЕНТОВ «TO-BE».....	81
3.4 ЭКОНОМИЧЕСКО ОБОСНОВАНИЕ.....	83
3.5 ВЫВОДЫ ПО РАЗДЕЛУ.....	94
ЗАКЛЮЧЕНИЕ.....	95
СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ.....	97

ВВЕДЕНИЕ

Обработка анкетных данных – трудоёмкий процесс, особенно в учебных заведениях, ведь необходимо собрать данные с многочисленного числа студентов. В результате сбора данных интервьюер имеет большое количество анкет. Ответы на опрос не подлежат отбору, то есть какие из них следует учитывать, а какие бесполезны, так потребуются дополнительные трудовые ресурсы, что означает увеличение затрат.

Выбор темы диссертации обуславливается сложившейся проблемой, описанной выше. Для автоматизации процесса отбора анкетных данных были разработаны и обучены алгоритмы машинного обучения, классификаторы, которые определяют для заполненной анкеты: будет ли она учтена для анализа эффективности учебного процесса.

Объект исследования – система сбора и обработки анкетных данных отдела диспетчеризации ВШЭМ УрФУ.

Предмет исследования – автоматизация процесса ранжирования анкетных данных по релевантности.

Целью данной работы является автоматизация процесса отбора анкетных данных в дистрибутиве Python Anaconda с использованием алгоритмов машинного обучения.

Для достижения цели поставленной в данной работе были определены следующие задачи:

- рассмотреть теоретические основы систем сбора и обработки данных;
- изучить системы искусственного интеллекта и программное обеспечение систем ИИ;
- проанализировать деятельность службы диспетчеризации ВШЭМ УрФУ и предоставить решение для более эффективной работы;
- автоматизировать процесс отбора анкетных данных в дистрибутиве Python Anaconda с использованием алгоритмов машинного обучения;

- определить эффективность автоматизации отбора анкетных данных для службы диспетчеризации ВШЭМ УрФУ.

Структура магистерской диссертации представлена введением, тремя разделами, заключением и списком использованной литературы.

Используемые методы исследования

- анализ литературы и нормативно-правовых документов по теме исследования;
- изучение и обобщение сведений;
- сравнение явлений и качеств;
- моделирование (получение информации о предмете через созданную модель;
- измерение (получение количественных данных).

Степень разработанности темы

Проблема автоматизации обработки данных является довольно разработанной и освещалась ранее в научных работах многих известных специалистов. Но данные работы не могут полностью быть применены для решения частного случая, а именно – для автоматизации отбора данных для последующей обработки. Это связано с тем, что нет единого подхода для автоматизации обработки данных. Данная магистерская работа направлена на решение конкретных задач, описанных выше.

Основные пункты научной/методологической новизны диссертации

В ходе написания магистерской диссертации был проведен анализ литературы и нормативно-справочной документации, по результатам исследования был написан классификатор, который определяет для заполненной анкеты: будет ли она учтена для анализа эффективности учебного процесса.

Практическая значимость исследования

В ходе создания данного проекта был автоматизирован процесс отбора анкетных данных. Разработка классификатора, безусловно, полезна документоведам и всему преподавательскому составу ВШЭМ УрФУ. Так как

сократятся трудозатраты при дальнейшей обработке данных, оценки, полученные по результатам обработки анкет будут отражать действительное качество обучения.

Эмпирическая база

При написании магистерской диссертации были учтены результаты собственных исследований, исследований других авторов, нормативные документы и иные источники.

Структура диссертации

Диссертация состоит из введения, трех разделов, заключения, списка литературы.

Первый раздел является теоретической базой магистерской работы, в нем рассмотрены существующие системы сбора и обработки данных. Изучены системы искусственного интеллекта и программное обеспечение для данных систем.

Во втором разделе раскрыт вопрос применения систем искусственного интеллекта при сборе и обработке данных, представлены разработанные классификаторы.

Третий раздел включает в себя общую характеристику ВШЭМ УрФУ, полную модель отдела диспетчеризации ВШЭМ, модели «as-is» и «to-be». Также в данной части работы представлено экономическое обоснование проекта: затраты на проект, эффективность от внедрения.

В качестве теоретической базы были использованы статьи и учебные пособия отечественных и зарубежных авторов, программа развития Института «Высшая школа экономики и менеджмента Уральского федерального университета им. первого Президента России Б.Н. Ельцина».

В процессе выполнения работы были использованы следующие инструменты: BPWin, MS Excel, MS Word.

1 ОБРАБОТКА ДАННЫХ С ИСПОЛЬЗОВАНИЕМ СИСТЕМ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА

1.1 СИСТЕМЫ СБОРА И ОБРАБОТКИ АНКЕТНЫХ ДАННЫХ

Анкета – основной инструмент количественных исследований.

Сбор данных осуществляется с помощью метода анкетирования, который используется в рамках определенного социального исследования и предполагает, что целевые группы респондентов самостоятельно заполняют анкеты и возвратят их интервьюеру [7].

Цель анкетирования заключается в выявлении различных фактов и тенденций.

В результате процесса анкетирования собирают большой объем данных, что обуславливает необходимость использования соответствующих программных средств для обработки этих данных.

В настоящее время набирает популярность статистический программный комплекс SPSS [8].

Общая информация о пакете SPSS и его структуре

Пакет SPSS (Statistical Package for the Social Sciences – статистический пакет для социальных наук) изначально¹ был разработан как компьютерная программа для статистической обработки данных, предназначенная для проведения прикладных исследований в социальных науках.

В 2009 г. пакет стал называться PASW Statistics (Predictive Analytics SoftWare – интеллектуальное аналитическое программное обеспечение) [11].

С июля 2009 г. пакет сопровождается фирмой IBM (International Business Machines) под именем IBM SPSS Statistics.

В 2013 г. планируется выход очередной версии пакета - IBM SPSS Statistics 22, работающей под управлением различных операционных систем - Windows, MacOSX, Linux [14].

Имея модульную структуру, пакет SPSS обеспечивает комплексную статистическую обработку – от планирования до управления данными, выполнения анализа и представления результатов. Мощные средства анализа и обработки данных с развитым графическим интерфейсом, удобные меню и простые диалоговые окна существенно упрощают работу пользователя [10].

Основные блоки SPSS:

- редактор данных – гибкая система, внешне похожая на электронную таблицу, служит для определения, ввода, редактирования и просмотра данных;
- визуализатор – средство просмотра, упрощающее отображение результатов, позволяющее показывать и скрывать отдельные элементы вывода, изменять порядок вывода результатов, перемещать готовые к презентации таблицы и диаграммы в другие приложения и получать их из других программных приложений;
- многомерные мобильные таблицы – служат для отображения результатов анализа, позволяя исследовать таблицы, перемещая строки, столбцы и слои и, таким образом, выявлять важные моменты, которые могут потеряться в стандартных отчетах. Обеспечивают также сравнение групп, расщепляя таблицы таким образом, чтобы каждый раз на экран выводилась только одна группа;
- высококачественная графика – средство формирования полноцветных диаграмм с высоким разрешением: круговых и линейчатых диаграмм, гистограмм, диаграмм рассеяния, объемных диаграмм и многих-многих других [40];
- доступ к базам данных – конструктор чтения баз данных, позволяющий легко, несколькими нажатиями кнопки мыши, загружать данные из любых источников;
- преобразование данных – средство преобразования данных, помогает готовить данные к анализу: можно с легкостью выделять подмножества в данных, объединять категории, добавлять, агрегировать, сливать,

- расщеплять и транспонировать файлы, а также проводить другие преобразования;
- справочная система – комплекс средств поддержки начинающего пользователя, включающий [9]:
 - 1) электронный учебник, содержащий детальный обзор средств и возможностей пакета;
 - 2) контекстную справку, помогающую разобраться в конкретных задачах при работе с диалоговыми окнами;
 - 3) всплывающие определения, объясняющие статистические термины в мобильных таблицах;
 - 4) репетитор по статистике, помогающий в поиске необходимой процедуры обработки;
 - 5) примеры анализа, облегчающие интерпретировать результаты в аналогичных типовых задачах;
 - командный язык, обеспечивающий доступ к дополнительным функциональным возможностям пакета, недоступным через меню и диалоговые окна, и позволяющий сохранять и автоматизировать различные повторяющиеся процессы и задачи. Полная документация по командному языку интегрирована в справочную систему и доступна в виде отдельного PDF-документа «Руководство по синтаксису», который можно вызвать в меню команды «Справка» [12].
 - Предусмотрены различные варианты поставки продуктов IBM SPSS Statistics:
 - IBM SPSS Statistics Standart – включает основные аналитические возможности для решения широкого спектра хозяйственных и исследовательских задач;
 - IBM SPSS Statistics Professional – дополнительно содержит средства, связанные с обеспечением качества данных и их полноты, а также автоматизации функций статистики и прогнозирования;

- IBM SPSS Statistics Premium – включает полный набор аналитических методик, систему моделирования на основе структурных уравнений (SEM), средства подробной оценки и проверки выборочных данных, процедуры прямого маркетинга.
- IBM SPSS Statistics for Education (только на английском языке) – включает основные модули, популярные в учебном процессе (для студентов, сотрудников и преподавателей) [13].

В настоящее время пакет представляет собой полнофункциональную исследовательскую систему для статистической обработки и анализа разноплановой информации – бизнес-структурами, представителями государственных органов, научными подразделениями, академическими организациями и вузами. Результаты анализа можно наглядно представлять в виде таблиц и диаграмм различных типов, распространять сетевым пользователям, внедрять полученные результаты в другие программные системы.

Последние версии пакета обладают мощными аналитическими инструментами и средствами ускорения работы, позволяют проще и быстрее выявлять ценную информацию в данных, делать более точные прогнозы и обеспечивать большую отдачу от аналитики при решении прикладных задач. Они содержат все современные средства импорта-экспорта данных, предобработки, реструктуризации и визуальной категоризации данных. Обеспечиваются широкие возможности имитационного моделирования данных с заданными характеристиками методом Монте-Карло (в том числе, с характеристиками фактических данных, формой распределения и связями между переменными). Тем самым предоставляется возможность специалистам, связанным с планированием и моделированием затрат и рисков, надежнее оценивать диапазон возможных значений прогнозируемого показателя и оценивать его отклик при воздействии на управляемые переменные [41].

Задачи и ключевые возможности поставок пакета SPSS

Профессиональная поставка пакета SPSS Statistics Professional обеспечивает решение всех задач полного жизненного цикла углубленного, нестандартного анализа – качества данных, их сложности, автоматизации и прогнозирования.

Ключевые возможности профессиональной поставки реализуются множеством инструментальных средств и процедур.

- Линейные модели включают разнообразные процедуры регрессионного и расширенного статистического анализа, спроектированные с учетом присущих характеристик составных отношений описания данных. Они предполагают применение:
 - 1) обобщенных линейных смешанных моделей (GLMM) для работы с иерархическими данными;
 - 2) процедур общих линейных моделей (GLM) и смешанных моделей;
 - 3) обобщенных линейных моделей (GENLIN): - линейной регрессии для ответов с нормальным распределением; - логарифмических моделей для двоичных данных; - логарифмически-линейных моделей для данных подсчета; - реализации других полезных статистических моделей с использованием самых общих формулировок;
 - 4) процедур обобщенных оценочных уравнений (GEE), расширяющих обобщенные линейные модели для возможности обработки связанных осевых и сгруппированных данных [60].
- Нелинейные модели предоставляют возможность применять к данным более сложные модели. Они обеспечивают использование:
 - 1) полиномиальной логарифмической регрессии (MLR) для прогнозирования категориальных исходов с числом категорий больше двух;
 - 2) двоичной логарифмической регрессии для разбивки данных на две группы;

- 3) нелинейной регрессии (NLR) и ограниченной нелинейной регрессии (CNLR) для оценки параметров нелинейных моделей;
 - 4) пробит-анализа для оценки значения стимулов.
- Настраиваемые таблицы позволяют пользователям легко определять суть данных и быстро выводить обзор результатов в разных стилях для разных потребителей:
- 1) при наличии дедуктивной статистики можно сравнивать средние значения и соотношения для демографических групп, сегментов потребителей или других категориальных переменных;
 - 2) обеспечивается формирование итоговой статистики (от простых подсчетов категориальных переменных до величин дисперсии) и сортировка категорий по любому из используемых итогов статистики;
 - 3) предусмотрено использование трех тестов значимости:
 - а) проверка независимости по критерию хи-квадрат;
 - б) сравнение средних значений столбцов (проверка по критерию Стьюдента);
 - в) сравнение соотношений столбцов (проверка по z-критерию);
 - 4) интерактивный редактор таблиц позволяет:
 - а) создавать сводные таблицы с помощью мыши;
 - б) исключать определенные категории;
 - в) показывать ячейки с отсутствующими значениями;
 - г) добавлять в таблицы промежуточные итоги;
 - 5) таблицы можно просматривать в реальном режиме времени и изменять по мере создания;
 - 6) таблицы можно экспортировать в документы Microsoft Word, Excel, PowerPoint или HTML для использования в отчетах [62].
- Подготовка данных рационализирует этап подготовки данных аналитического процесса:

- 1) позволяет определить подозрительные или недопустимые варианты, переменные и значения данных;
 - 2) позволяет просматривать шаблоны отсутствующих данных и создавать обзоры распределений переменных;
 - 3) функция Optimal Binning позволяет найти наилучший исход из возможных для алгоритмов, предназначенных для номинальных атрибутов;
 - 4) инструмент Автоматической подготовки данных (ADP) выявляет и устраняет ошибки качества и условно подставляет отсутствующие значения за один эффективно выполняемый этап;
 - 5) рекомендации и визуализации помогают определить, какие данные следует использовать.
- Значения, показывающие правильность или отсутствие данных повышают вероятность получения статистически значимых результатов:
- 1) данные проверяются с нескольких разных точек зрения с помощью одного из шести отчетов диагностики, затем оценивается итоговая статистика и подставляются отсутствующие значения;
 - 2) выполняется диагностика серьезных неполадок подстановки отсутствующих данных;
 - 3) заменяются отсутствующие значения приблизительными;
 - 4) отображается моментальная копия для каждого типа отсутствующего значения и крайние значения для каждого варианта;
 - 5) скрытое отклонение убирается путем замены отсутствующих значений приблизительными; это делается для того, чтобы включить все группы, даже с низким уровнем реагирования.
- Деревья решений упрощают определение групп, выявление отношений между группами и прогнозирование будущих событий:

- 1) пакет визуально определяет, как выполняется поток модели, чтобы можно было найти отдельные подгруппы и отношения;
 - 2) обеспечивается создание деревьев классификации непосредственно в пакете, чтобы можно было использовать результаты для сегментации и группировки вариантов непосредственно в данных;
 - 3) правила выбора или классификации/прогнозирования создаются в синтаксисе IBM SPSS Statistics, в операторах SQL или простом тексте (прямой синтаксис).
 - 4) Обеспечивается выбор одного из четырех алгоритмов формирования дерева:
 - а) CHAID – быстрого статистического алгоритма, который просматривает данные быстро и эффективно, создавая сегменты и профайлы с учетом предпочитаемого исхода;
 - б) исчерпывающего CHAID – модификации CHAID, проверяющей все возможные расщепления для каждой независимой переменной;
 - в) деревьев классификации и регрессии (C&RT) – полноценного двоичного алгоритма деревьев, разделяющего данные и создающего правильные однородные подмножества;
 - г) QUEST – статистического алгоритма, позволяющего выбрать переменные без отклонений и быстро и эффективно построить правильные двоичные деревья.
- Функции прогнозирования позволяют быстрее анализировать данные хронологии и выявлять тенденции:
- 1) позволяют доставлять информацию способом, более подходящим для ответственных руководителей организации, работающих с этой информацией;
 - 2) автоматически определяют наиболее подходящий алгоритм (модель) экспоненциального сглаживания для анализа данных хронологии;

- 3) вместо создания переменных по одной позволяют одновременно моделировать сотни рядов с различным временем;
- 4) модели сохраняются в центральном файле: при изменении данных прогнозы обновляются без необходимости повторного задания параметров или повторной оценки моделей;
- 5) можно создавать сценарии, позволяющие автоматически обновлять модели путем добавления новых данных.

Стандартная поставка пакета SPSS Statistics Standard содержит основные аналитические возможности для широкого спектра хозяйственных и исследовательских задач и позволяет просматривать данные, формулировать гипотезы для дополнительного тестирования и выполнять процедуры для выявления отношений между переменными, создания кластеров, определения тенденций и составления прогнозов.

Ключевые возможности стандартной поставки включают часть инструментальных средств и процедур профессиональной поставки.

- Линейные модели.
- Нелинейные модели.
- Настраиваемые таблицы.

Расширенная поставка пакета SPSS Statistics Premium позволяет использовать дополнительные средства, будучи ориентированной на аналитиков, специалистов по планированию, прогнозированию, анализу обзоров, оценке программ, маркетингу баз данных. Включает большое число полностью интегрируемых функций статистики и связанных продуктов для специализированных аналитических задач, выполняемых разными подразделениями предприятия. Позволяет значительно повысить производительность и добиться высоких результатов в ходе реализации конкретных проектов и бизнес-целей.

Ключевые возможности расширенной поставки включают часть инструментальных средств и процедур профессиональной поставки и новые возможности.

- Линейные модели.
- Нелинейные модели.
- Настраиваемые таблицы.
- Подготовка данных.
- Достоверность данных и отсутствующие значения.
- Категориальные и числовые данные:
 - 1) выявляются базовые отношения с помощью карт позиционирования, точечных диаграмм биплот (biplot) и трilinearных точечных диаграмм;
 - 2) используя общепринятые процедуры обработки, облегчается работа с номинальными (напр., зарплата) или порядковыми (напр., уровень образования) данными, делая их более понятными;
 - 3) визуально интерпретируются наборы данных указанием на взаимосвязь между строками и столбцами в больших таблицах баллов, подсчетов, рейтингов, приоритетов или подобий;
 - 4) обрабатываются отличные от нормальных остатки в числовых данных или нелинейные отношения между независимыми переменными (напр., атрибуты потребитель или продукт) и переменной исхода (напр., покупка/без покупки);
 - 5) реализуются приемы гребневой регрессии, лассо, эластической сети, выбора переменной и выбора модели как для числовых, так и для категориальных данных.
- Древовидные схемы принятия решения.
- Функции прогнозирования.
- Моделирование на основе структурных уравнений:
 - 1) проверяются гипотезы и подтверждаются отношения между наблюдаемыми или ненаблюдаемыми переменными, не останавливаясь на регрессии при выявлении дополнительной статистики;

- 2) обеспечивается создание моделей, более реалистично отражающих отношения, поскольку числовые переменные, как наблюдаемые (данные обследования, полученные не экспериментальным путем), так и ненаблюдаемые (удовлетворение и приверженность), могут использоваться для прогнозирования любых других числовых переменных;
 - 3) поддерживается визуальное сравнение, подтверждение и уточнение модели;
 - 4) многомерный анализ реализуется на расширении стандартных методов исследования - регрессии, факторного анализа, корреляции и дисперсионного анализа;
 - 5) используется три метода вычисления данных для подстановки: регрессия, стохастическая регрессия и байесовский подход.
- Применение начальной загрузки:
- 1) определяется оценка выборочного распределения оценочной функции путем повторного взятия выборки с заменой из исходной выборки;
 - 2) вычисляется оценка стандартных ошибок и доверительных интервалов параметра генеральной совокупности (таких как среднее, медиана, соотношение, отношение шансов, коэффициент корреляции, коэффициент регрессии и др.);
 - 3) создаются многовариантные наборы данных для более точного анализа;
 - 4) упрощается тестирование стабильности и надежности моделей.
- Дополнительная оценка выборки и тестирование:
- 1) предусмотрены специализированные инструменты планирования, статистики, необходимые для работы с планами сложной выборки (при многослойной, кластерной или многоэтапной выборке);
 - 2) обеспечивается получение улучшенных результатов за счет входа в состав обзорного анализа выборочного плана;

- 3) реализована более точная работа с числовыми и категориальными исходными данными в планах сложной выборки с использованием алгоритмов анализа и прогнозирования, в т.ч. времени события;
 - 4) предусмотрено использование Мастеров при создании планов, анализе данных и интерпретации результатов.
- Инструменты прямого маркетинга и принятия решений по продукту:
- 1) сегментация потребителей или контактов путем создания групп;
 - 2) профилирование потребителей или контактов с общими характеристиками (для улучшения выбора целевой аудитории при маркетинговых исследованиях);
 - 3) выявление уровней предрасположенности (при определении склонных к покупке потребителей);
 - 4) определение ответов по почтовому индексу;
 - 5) интеграция ответов потребителей.
- Высококачественные диаграммы и графики:
- 1) десятки встроенных шаблонов визуализации для отображения результатов анализа;
 - 2) создание графиков не требует навыков программирования;
 - 3) настройка таблиц стилей и шаблонов графиков;
 - 4) отображение графиков в различных ОС;
 - 5) поддержка разнообразных источников данных (специальных и общих).

Как видно, поставка Premium имеет существенно более широкие аналитические возможности по сравнению с профессиональной версией пакета.

Учебная поставка пакета SPSS Statistics for Education - специальная версия для учебных заведений. Поставляется в ВУЗы, прежде всего США, при заключении соответствующих договоров со значительной скидкой. Версия выходит только на английском языке. Возможности версии ограничены.

Схема обработки информации в SPSS

Использование пакета SPSS предусматривает выполнение следующих функциональных задач:

- оперативное получение аналитической информации – ключевых факторов, взаимосвязей и тенденций в исследуемых данных;
- наглядное представление результатов в виде таблиц и диаграмм;
- распространение полученных результатов сетевым пользователям – заказчикам проведенных исследований;
- поддержка процессов оперативного формирования научно-обоснованных оптимальных управленческих решений.

При этом пакет SPSS реализует семь этапов аналитического процесса:

- Планирование.
- Сбор данных.
- Обеспечение доступа к данным.
- Подготовка данных к анализу.
- Выполнение анализа.
- Формирование отчетов.
- Представление и распространение результатов.

При необходимости некоторые этапы можно пропускать (если они ранее были выполнены) или повторять несколько раз (для получения дополнительных данных) [43].

Назначение и общие возможности модулей IBM SPSS

Семейство программных продуктов IBM SPSS – это многомодульный, полностью интегрированный комплекс программ, охватывающий все этапы аналитического процесса – от планирования и сбора данных до анализа, создания отчетов и распространения результатов.

Все программы комплекса имеют интуитивно понятный интерфейс и включают в себя многочисленные функции и процедуры управления данными, реализующие современные статистические методы и алгоритмы, а также

средства создания отчетов, что обеспечивает проведение анализа любой степени сложности.

Использование комплекса обеспечивает решение бизнес-проблем и исследовательских задач при помощи статистического программного обеспечения.

Достигается возможность сбора и оперативной обработки массивов информации и получение аналитических сведений, в том числе с представлением результатов в виде наглядных таблиц и диаграмм.

Обеспечивается выявление ключевых факторов, взаимосвязей, тенденций, скрытых закономерностей в исходных данных.

Как результат: руководителям предоставляется научно-обоснованная информация, что позволяет своевременно принимать оптимальные управляющие воздействия.

Семейство SPSS представляет собой комплекс четырех взаимосвязанных инструментальных блоков (модулей, программных комплексов, систем, платформ) для прогностической аналитики и анализа данных, использование которых, например, позволяет:

- Собрать всю необходимую информацию о вкусах и мнениях потребителей.
- Спрогнозировать реакцию клиентов на специальные предложения до того, как они осуществляются.
- Влиять на результаты деятельности компании за счет внедрения прогностической аналитики в бизнес-процессы.

IBM SPSS Data Collection – сбор данных является программной платформой, позволяющей проводить опросы и маркетинговые исследования наиболее эффективными способами:

- по телефону,
- с помощью портативных компьютеров,
- через сеть Интернет [24].

IBM SPSS Statistics – статистический анализ – полнофункциональная система, предназначенная для решения бизнес- и исследовательских задач при помощи анализа данных, позволяющая наглядно представлять результаты в виде таблиц и диаграмм, а также распространять и внедрять полученные результаты.

IBM SPSS Modeler – моделирование – программный комплекс, позволяющий принимать обоснованные бизнес-решения, опираясь на надежные модели. Является признанным инструментом Data Mining – извлечения и добычи знаний, позволяет обнаруживать скрытые закономерности, предсказывать и оценивать возможные результаты альтернативных вариантов действий [22].

IBM SPSS Collaboration and Deployment Services – интеграция и внедрение в бизнес-процессы – семейство продуктов, предназначенное для быстрого и надежного внедрения прогностической аналитики во все процессы принятия решений, позволяющее соединить прогностическую аналитику и бизнес [19].

Основные возможности комплекса программ:

- Моделирование структурными уравнениями для поиска скрытых взаимосвязей в данных.
- Определение корректного объема выборки на этапе планирования исследования (размера выборки, соответствующего бюджету и решаемым задачам).
- Распространение отчетов, созданных в SPSS для Windows.
- Распространение в Интернете интерактивных отчетов, созданных в SPSS для Windows.
- Анализ сложных взаимосвязей в данных (методы одномерного и многомерного анализа).
- SPSS Statistics Base – доступ к данным, управление данными, подготовка данных к анализу, анализ данных и создание отчетов;

дополнительные модули встраиваются в SPSS Statistics Base и позволяют расширять аналитические возможности ПО.

- Прогнозирование категориальных данных (откликов) и их исследование при помощи карт восприятия (анализ громоздких двух- и многовходовых таблиц, визуализация и исследование категориальных данных).
- Построение деревьев классификаций и решений, идентифицирование групп, нахождение взаимосвязи в данных, предсказывание будущих событий.
- Разработка сложных выборок и учет сложности выборки в процессе анализа (детализация приводится далее).
- Определение предпочтений клиентов.
- Инструменты для организации ввода и чистки данных.
- Централизация сбора и ввода данных.
- Рабочее место оператора ввода и чистки данных.
- Точный анализ малых выборок и разрезов.
- Упрощение процесса проверки данных, устранение интенсивной ручной проверки, получение более точных результатов.
- Корректная импутация пропущенных значений (заполнение пропущенных значений для повышения информативности данных и построения адекватных моделей).
- Улучшение качества предсказаний при помощи мощных регрессионных процедур (моделей прогнозирования).
- Анализ централизованно хранящихся данных большого объема.
- Эффективное представление результатов.
- Прогнозирование временных рядов – мощного инструмента для анализа хронологической информации, построения моделей и выявления тенденций.

- Быстрый и полный анализ ответов на открытые вопросы опросов и обследований.

Работа в среде пакета SPSS основывается на применении конструкторов [32].

Предусмотрены следующие виды планов отбора:

- Расслоенный отбор – повышает точность результатов, полученных по выборочным данным, гарантирует репрезентативность выборки по ключевым группам за счет выборки из подгрупп обследуемой совокупности. Подгруппами, например, могут быть мужчины и женщины или люди определенных профессий, возрастных групп и т.п.
- Кластерный отбор – извлекаются кластеры, представляющие собой группы отбираемых для обследования единиц. Кластерами могут быть школы (вузы), больницы или географические районы, при этом отбираемыми единицами являются студенты, пациенты или проживающие на данной территории граждане. Кластерный отбор часто помогает снизить стоимость проведения обследования.
- Многоэтапный отбор – вначале, на первом этапе, извлекается выборка групп элементов генеральной совокупности; затем формируется выборка второго этапа путем извлечения подвыборки в каждой единице, отобранной на первом этапе.

Используя такую возможность многократно, можно сформировать выборку с большим числом этапов отбора. Например, в обследованиях населения можно формировать выборку членов домашних хозяйств, которые в свою очередь отбираются в районах города [18].

Направления практического использования SPSS

Анализ информации о клиентах (CRM, управление взаимоотношениями с клиентами):

- Обеспечение лучшего понимания поведения клиентов при помощи построения прогностических профилей.
- Прогнозирование различных вариантов поведения клиентов.

- Максимизация отдачи от клиентов за счет прогнозирования их интересов при помощи прогностической аналитики. Например, какие товары и услуги будут необходимы клиентам и, таким образом, максимизировать отдачу от клиентов.

Проведение опросов и обследований дает оптимальные результаты на каждом из семи этапов проведения обследования и опросов:

- Возможность принятия более нацеленных решений, основанных на надежной и достоверной информации, полученной благодаря опросам и обследованиям.
- Маркетинговые исследования.
- Анализ данных маркетинговых исследований.

Data mining (добыча знаний)

- Решение ключевых бизнес-проблем при помощи современных аналитических технологий.
- Обеспечение максимальной отдачи от полученных аналитических результатов за счет их внедрения.

Data mining (добыча знаний) – это процесс выявления скрытых фактов и взаимосвязей в больших массивах данных.

Data mining разрешает парадокс: чем больше накапливается данных о клиентах, тем сложнее и дольше эффективно анализировать эти данные и получать значимые выводы. То, что, по идее, должно быть золотоносной жилкой, обычно остается неисследованным из-за отсутствия подходящего персонала, недостатка времени и навыков.

Использование пакета:

- в государственном управлении:
 - 1) Аналитическая поддержка принятия управленческих решений.
 - 2) Повышение эффективности здравоохранения и инфраструктуры.
 - 3) Обнаружение и предотвращение случаев мошенничества, перерасхода и нецелевого использования средств.
 - 4) Проведение опросов общественного мнения.

- в здравоохранении:
 - 1) Анализ данных о пациентах.
 - 2) Поддержка планирования курсов лечения.
 - 3) Проведение медицинских обследований.
 - 4) Оптимизация работы медицинских учреждений.
- в образовательных учреждениях:
 - 5) Обучение студентов прикладной статистике и практическим навыкам анализа данных.
 - 6) Управление процессом приема/отбора студентов.
 - 7) Тестирование студентов.
- в финансовой сфере:
 - 8) Анализ финансовых рисков.
 - 9) Выявление и предотвращение случаев мошенничества в финансовой сфере.
 - 10) Прогнозирование финансовых показателей [21].

1.2 СИСТЕМЫ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА

Искусственный интеллект – это свойство автоматических вычислительных систем исполнять отдельные функции интеллекта человека, например, выбирать и принимать оптимальные и эффективные решения на основе полученного опыта и рационального анализа внешних воздействий [16].

Под мыслительным процессом или интеллектуальной деятельностью понимаем деятельность мозга, которая направлена на решение задач, требующих умственных усилий. Интеллект или мышление напрямую связаны с решением таких задач, как логический анализ, распознавание ситуаций, доказательство теорем, планирование поведения, а также управление в условиях неопределенности. Характерными чертами интеллекта, которые проявляются в процессе решения задач, являются следующие: способность к обобщению, и аккумуляции опыта (знаний и навыков), обучению и адаптации к изменяющимся условиям в процессе решения задач. Мозг может

решать разнообразные задачи благодаря вышеперечисленным качествам, а также быть мобильным, легко перестраиваться с решения одной задачи на другую. Следовательно, мозг, который наделен интеллектом, является универсальным средством решения широкого спектра задач, в том числе неформализованных, для которых нет заранее известных методов решения [17].

Безусловно, можно исключить задачи, которые не решаются стандартными методами, из класса интеллектуальных. В качестве примеров таких задач могут выступать вычислительные задачи: решение системы линейных алгебраических уравнений, численное интегрирование дифференциальных уравнений и прочее. Решением подобных задач выступают алгоритмы, которые представляют собой определенную последовательность элементарных операций; данная последовательность может быть легко реализована в виде компьютерной программ. Хотя для широкого класса интеллектуальных задач, таких, игра в шахматы, как распознавание образов, доказательство теорем и т.п., напротив, это формальное разбиение процесса поиска решения на отдельные элементарные шаги часто оказывается весьма затруднительным, даже если само их решение является поверхностным [15].

Существуют две научные школы с различными подходами к проблеме искусственного интеллекта (ИИ): вычислительный ИИ и конвенционный ИИ. Вычислительный ИИ под собой подразумевает итеративную разработку и обучение. Обучение основано на эмпирических данных и ассоциируется с не-символьным ИИ и нечеткими системами. В конвенционном ИИ главным образом используются методы машинного самообучения, основанные на статистическом анализе и формализме. Методы конвенционного ИИ реализуются в системах и подходах указанных ниже:

- Экспертные системы: программы, действующие по определенным правилам и обрабатывающие большое количество информации. В результате своей работы выдают заключение или рекомендацию.
- Рассуждение по аналогии.

- Байесовские сети доверия представляют собой вероятностные модели - систему из множества переменных и их вероятностных зависимостей.
- Поведенческий подход: модульный метод построения систем ИИ, система разбивается на несколько сравнительно независимых программ поведения, последние иницируются в зависимости от изменений внешней среды [25].

Основные методы вычислительного ИИ:

- Нечеткие системы: методики для рассуждения в условиях неопределенности.
- Нейронные сети: симуляция нервной системы, демонстрирующие, в частности, высокие способности к распознаванию образов.
- Эволюционные вычисления: модели, которые используют понятие естественного отбора, обеспечивающего отсеивание наименее подходящих согласно заданному критерию решений. В этой группе методов выделяют генетические алгоритмы и муравьиный алгоритм [26].

Экспертные системы

Экспертная система (ЭС) представляет собой компьютерную программу, которая способна выступать в роли специалиста-эксперта в решении проблемной ситуации. Эти системы начали разрабатываться исследователями ИИ в 1970-х годах, а в 1980-х получили коммерческое распространение. В общем случае структура экспертной системы может быть выражена следующей схемой (рисунок 1.1).

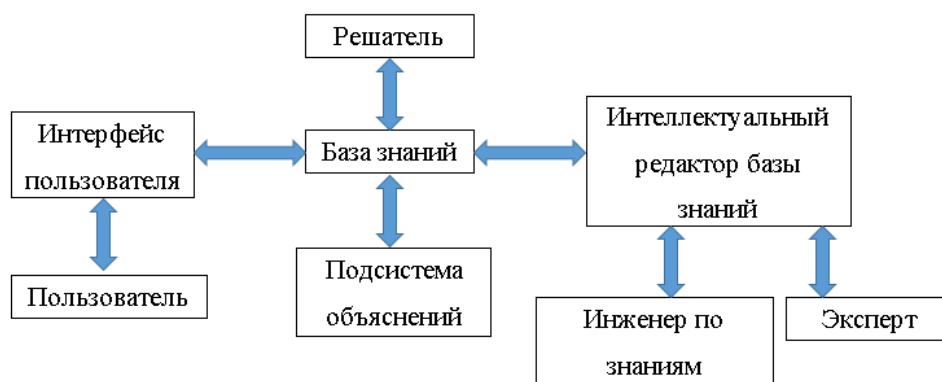


Рисунок 1.1 – Обобщенная структура экспертной системы¹

База знаний (БЗ) – главный элемент экспертной системы. Она состоит из правил анализа данных от пользователя по конкретной проблеме. Блок логического вывода, представляет собой программу, моделирующую ход рассуждений эксперта на основании знаний, содержащихся в БЗ.

Подсистема объяснений – это программа, с помощью которой пользователь может получить ответы на вопросы: «Почему система приняла то или иное решение?» и «Как была получена та или иная рекомендация?». Ответ на вопрос «как» – это трассировка всего процесса получения решения с указанием использованных фрагментов БЗ, т.е. всех шагов цепи умозаключений. Ответ на вопрос «почему» – это ссылка на умозаключение, которое предшествует полученному решению, т.е. отход на один шаг назад.

ЭС создается при помощи инженеров по знаниям (аналитиков), которые разрабатывают ядро ЭС и, зная организацию базы знаний, заполняют ее при помощи эксперта по специальности. Интеллектуальный редактор БЗ – программа, предоставляющая инженеру по знаниям возможность создавать БЗ в диалоговом режиме.

Интерфейсом пользователя является комплекс программных средств, реализующих диалог пользователя с ЭС как для ввода информации, так и для получения результатов работы ЭС [38].

Задачи, решаемые при помощи экспертных систем, чаще всего относятся к одной из следующих областей:

¹ Составлено автором по: [6]

- Интерпретация данных – это одна из традиционных задач для экспертных систем. Под интерпретацией понимается определение смысла данных, результаты которого должны быть согласованными и корректными. Примеры существующих ЭС: SIAP (обнаружение и идентификация различных типов океанских судов), АВТАНТЕСТ, МИКРОЛЮШЕР (определение основных свойств личности по результатам психодиагностического тестирования).
- Диагностика – это обнаружение неисправности в некоторой системе. Трактовка неисправности как отклонения от нормы позволяет с единых теоретических позиций рассматривать и неисправность оборудования в технических системах, и заболевания живых организмов, и всевозможные природные аномалии. Примеры существующих ЭС: ANGY (диагностика и терапия сужения коронарных сосудов), CRIB (диагностика ошибок в аппаратуре и математическом обеспечении компьютера).
- Мониторинг – это непрерывная интерпретация данных в реальном масштабе времени и сигнализация о выходе тех или иных параметров за допустимые пределы. Главные проблемы – «пропуск тревожной ситуации» и инверсная задача «ложного» срабатывания. Сложность этих проблем состоит в размытости симптомов тревожных ситуаций и необходимости учета временного контекста. Примеры существующих ЭС: СПРИНТ (контроль за работой электростанций), REACTOR (помощь диспетчерам атомного реактора), FALCON (контроль аварийных датчиков на химическом заводе).
- Проектирование состоит в подготовке спецификаций на создание «объектов» с заранее определенными свойствами. Под спецификацией понимается весь набор необходимых документов – чертеж, пояснительная записка и т.д. Примеры существующих ЭС: XCON (проектирование конфигураций ЭВМ), SYN (синтез электрических цепей).

- Прогнозирование – это логический вывод вероятных следствий из заданных ситуаций. В прогнозирующей системе обычно используется параметрическая динамическая модель, в которой значения параметров «подгоняются» под заданную ситуацию. Выводимые из этой модели следствия составляют основу для прогнозов с вероятностными оценками [5]. Примеры существующих ЭС: WILLARD (предсказание погоды), PLANT (оценки будущего урожая), ECON (экономические прогнозы).
- Планирование – нахождение планов действий, относящихся к объектам, способным выполнять некоторые функции. В таких ЭС используются модели поведения реальных объектов с тем, чтобы логически вывести последствия планируемой деятельности. Примеры существующих ЭС: STRIPS (планирование поведения робота), ISIS (планирование промышленных заказов), MOLGFN (планирование эксперимента).
- Обучение – процесс диагностирования ошибки при изучении какой-либо дисциплины с помощью компьютера и подсказывают правильные решения. Они аккумулируют знания о гипотетическом «ученике» и его характерных ошибках, а затем в ходе работы способны диагностировать слабости в знаниях обучаемых и находить соответствующие средства их ликвидации. Кроме того, они планируют процесс общения с учеником в зависимости от успехов ученика с целью передачи знаний. Примеры существующих ЭС: PROUST (обучение языку программирования Pascal) [39].

Рассуждение по аналогии

СBR-системы являются реализацией методологии ИИ, которую применяют при построении компьютеризированных консультационных систем, базирующихся на накопленном опыте. В отличие от классических экспертных систем, действующих на основе логических правил, СBR-системы хранят успешные решения ряда реальных проблем, такие решения называют

примерами или прецедентами и при появлении новой проблемы находят по определенному алгоритму (зачастую при помощи машины логического вывода, с количественной оценкой) наиболее подходящие (похожие) прецеденты, после чего предлагает соответственно измененную комбинацию их решений. Если новую проблему удастся решить таким способом, то это решение заносится в базу прецедентов для повышения эффективности работы системы в будущем. Главный недостаток CBR-систем состоит в том, что они не создают моделей или правил, которые обобщают накопленный опыт [47].

Байесовские сети доверия

Как было сказано выше, байесовская сеть – это вероятностная модель, представляющая собой множество переменных и их вероятностных зависимостей. Например, байесовская сеть может быть использована для вычисления вероятности того, чем болен пациент по наличию или отсутствию ряда симптомов, основываясь на данных о зависимости между симптомами и болезнями. Существуют эффективные методы, которые используются для вычислений и обучения байесовских сетей. Байесовские сети используются для моделирования в биоинформатике (генетические сети, структура белков), медицине, классификации документов, обработке изображений, обработке данных и системах принятия решений [27].

Нейронные сети

Нейронная сеть (НС) – это распределенный параллельный процессор, состоящий из элементарных единиц обработки информации, которые накапливают экспериментальные знания и предоставляют их для последующей обработки [28]. Она представляет собой действующую модель нервной системы и сходна с мозгом с двух точек зрения:

Знания поступают в нейронную сеть из окружающей среды и используются в процессе обучения.

Для накопления знаний применяются связи между нейронами, называемые синаптическими весами.

Структурная схема единичного нейрона представлена на рисунке 1.2.

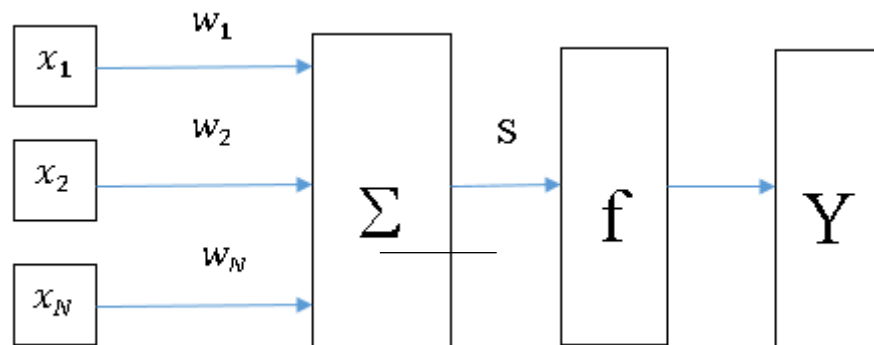


Рисунок 1.2 – Структурная схема нейрона²

Сигналы X_i , которые поступают на вход нейрона, умножаются на соответствующие весовые коэффициенты w_i , а затем суммируются. Результат суммирования поступает на нелинейный преобразователь, который реализует некоторую нелинейную функцию. Последняя называется функцией активации или передаточной функцией нейрона: результат ее действия поступает на выход нейрона. Используются множество способов построения нейронных сетей из отдельных нейронов. Слоистая архитектура НС является наиболее распространенной, ее обобщенная схема представлена на рисунке 1.3. Сигналы входного слоя поступают на входы нейронов первого слоя, после прохождения которого распространяются дальше, пока не достигнут выходов нейронной сети. Для обучения нейронных сетей такого типа используется, как правило, так называемый алгоритм обратного распространения ошибки, с помощью которого рассчитывают изменения весовых коэффициентов, необходимых для согласования выходных значений НС с выборкой образцов [29]. Использование нейронных сетей обеспечивает следующие полезные свойства систем:

- Нелинейность. Это качество нейронной сети особенно важно в том случае, если сам физический механизм, который отвечает за формирование входного сигнала, является нелинейным (например, человеческая речь).
- Адаптивность. Нейронные сети обладают способностью адаптировать свои синаптические веса к изменениям окружающей среды. Для работы

² Составлено автором по: [6]

в нестационарной среде создают нейронные сети, которые изменяют синаптические веса в реальном времени.

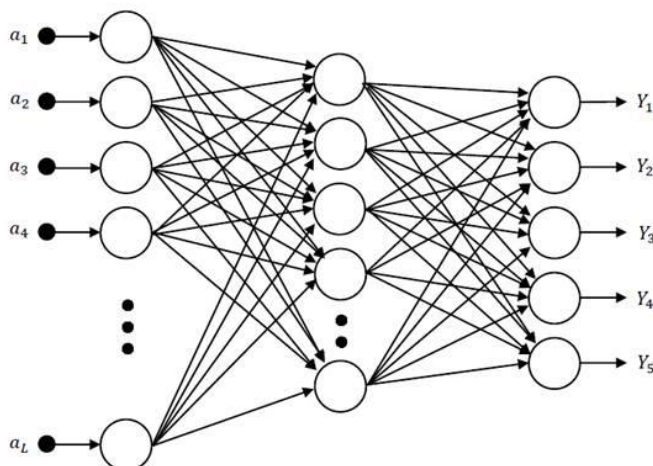


Рисунок 1.3 – Структура слоистой нейронной сети³

- Контекстная информация. Знания представляются в самой структуре нейронной сети. Каждый нейрон сети потенциально может быть подвержен влиянию всех остальных ее нейронов.
- Отказоустойчивость. Сети, которые реализованы аппаратно, потенциально отказоустойчивы. Это означает, что при неблагоприятных условиях их производительность уменьшается незначительно. Например, если поврежден какой-то нейрон или его связи, извлечение запомненной информации затрудняется. Однако, если брать в расчет распределенный характер хранения информации в нейронной сети, то можно утверждать, что только серьезные повреждения структуры нейронной сети существенно могут повлиять на ее работоспособность. Ниже представлены некоторые проблемы, которые решаются применением нейронных сетей:
 - а) Классификация образов. Задача состоит в том, чтоб указать принадлежности входного образа, который представлен набором признаков, одному или нескольким предварительно определенным классам. К известным приложениям относятся распознавание букв,

³ Составлено автором по: [6]

распознавание речи, классификация сигнала электрокардиограммы и прочее.

- б) Кластеризация/категоризация. Кластеризация основана на подобии образов: близкие образы размещены нейронной сетью в один кластер. Известны случаи применения кластеризации для извлечения знаний, сжатия данных и исследования свойств данных.
- в) Аппроксимация функций. Задача аппроксимации состоит в нахождении оценки некоторой искаженной шумом функции, которая генерирует обучающую выборку.
- г) Предсказание/прогноз. Задача прогнозирования состоит в предсказании некоторого значения для заданного момента времени на основании ряда значений, которые соответствуют другим моментам времени.
- д) Оптимизация. Задачей оптимизации является нахождение решения, удовлетворяющего системе ограничений и максимизирующего или минимизирующего целевую функцию.
- е) Ассоциативная память. Данная сфера применения НС состоит в организации памяти, адресуемой по содержанию, которое позволяет извлекать содержимое по искаженному или частичному образцу.

Дальнейшее повышение производительности компьютеров все в большей степени связывают с НС, в частности, с нейрокомпьютерами, в основе которых расположены аппаратно реализованные нейронные сети [51].

Нечеткие системы

Принципы нечеткой логики и теории нечетких множеств – то, на чем базируется направление. Данные принципы – раздел математики, который является обобщением классической логики и теории множеств. Данные понятия впервые были предложены американский ученым Лотфи Заде в 1965 г. Наличие нечетких и приближенных рассуждений при описании человеком процессов, систем, объектов – основная причина появления новой теории. В общем случае механизм логического вывода в рамках нечеткой логики

включает в себя четыре этапа: введение нечеткости (фазификация), нечеткий вывод, композиция и приведение к четкости или дефазификация (рисунок 1.4). Алгоритмы нечеткого вывода различаются главным образом видом используемых правил, логических операций и разновидностью метода дефазификации [42].



Рисунок 1.4 – Механизм нечеткого логического вывода⁴

Нечеткая логика оказала сильное влияние на другие парадигмы искусственного интеллекта. Объединение ее принципов с методами иных направлений породило следующие новые направления:

Нечеткие нейронные сети

- Нечеткие нейронные сети.
- Адаптивные нечеткие системы.
- Нечеткие запросы.
- Нечеткие ассоциативные правила.
- Нечеткие когнитивные карты.
- Нечеткая кластеризация.

Альтернативные методы искусственного интеллекта дополняют методологию нечеткой логики и используются в различных комбинациях для создания гибридных интеллектуальных систем [52].

Эволюционные вычисления

Генетический алгоритм, составляющий основу эволюционных вычислений, – это эвристический алгоритм поиска, который используют для решения задач оптимизации и моделирования путем последовательного подбора, комбинирования и вариации искомым параметров с использованием

⁴ Составлено автором по: [6]

механизмов, которые напоминают биологическую эволюцию. Сущность алгоритма состоит в следующем. Задача кодируется таким образом, чтобы ее решение могло быть представлено в виде вектора, хромосомы. Случайным образом создается некоторое количество начальных векторов (начальная популяция). Они оцениваются с использованием так называемой функции приспособленности, в результате чего каждому вектору присваивается определенное значение (приспособленность), определяющее вероятность выживания организма, которое представлено данным вектором. После этого с использованием полученных значений приспособленности выбираются вектора (селекция), допущенные к скрещиванию. К этим векторам применяются так называемые генетические операторы (в большинстве случаев – скрещивание и мутация), создавая таким образом следующее поколение. Особи следующего поколения также оцениваются, затем производится селекция, применяются генетические операторы и так далее.

Так происходит моделирование эволюционного процесса, продолжающегося несколько жизненных циклов, которые называют поколениями, пока не будет выполнен критерий останова алгоритма [33].

Главным образом, генетические алгоритмы применимы для поиска решений в очень больших, сложных пространствах поиска, и используются для решения следующих задач:

- оптимизация функций;
- разнообразные задачи на графах (задача коммивояжера, раскраска, нахождение паросочетаний);
- настройка и обучение нейронной сети;
- задачи компоновки;
- составление расписаний;
- игровые стратегии;
- аппроксимация функций;
- искусственная жизнь;
- биоинформатика [44].

Условия достижения интеллектуальности

Несмотря на существование множества методов ИИ, часть которых была рассмотрена выше, теоретики явно не определили, что именно считать необходимыми и достаточными условиями достижения интеллектуальности. С этим вопросом связан ряд гипотез, среди которых можно выделить следующие:

- гипотеза Ньюэлла-Саймона со следующей формулировкой: физическая символическая система имеет необходимые и достаточные средства для того, чтобы производить осмысленные действия. Другими словами, без символических вычислений невозможно выполнять осмысленные действия, а способность выполнять символические вычисления вполне достаточна для того, чтобы быть способным выполнять осмысленные действия. Независимо от того, справедлива ли эта гипотеза, символические вычисления стали реальностью, и весьма сложно отрицать полезность этой парадигмы для программирования.
- тест Тьюринга – мысленный эксперимент, который был предложен в качестве критерия и конструктивного определения интеллектуальности. Тест должен проводиться следующим образом. Судья (человек) переписывается на естественном языке с двумя собеседниками, один из которых – компьютер, а другой – человек. Если судья не может надежно определить, кто кем является, считается, что компьютер прошел тест. Предполагается, что каждый из собеседников стремится, чтобы человеком признали его [35].

Сверхзадачей искусственного интеллекта называют построение компьютерной интеллектуальной системы, обладающей уровнем эффективности решений неформализованных задач, сравнимым с человеческим либо превосходящим его [36].

1.3 ПРОГРАММНОЕ ОБЕСПЕЧЕНИЕ ДЛЯ СИСТЕМ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА

В рамках данного направления разрабатываются специальные языки для решения интеллектуальных задач, в которых упор делается на преобладание логической и символьной обработки над вычислительными процедурами. Языки ориентированы на символьную обработку информации: PROLOG, РЕФАЛ, LISP, CLIPS, PYTHON и прочие. LISP, один из старейших языков более высокого уровня, который обрабатывает списки. Prolog основан на логике. Сегодня популярны C ++ и Python. Также существует специальное программное обеспечение для разработки экспертных систем. Помимо этого, создаются пакеты прикладных программ, которые ориентированы на промышленную разработку интеллектуальных систем, или программные инструментарии искусственного интеллекта [2].

Несколько крупных пользователей ИИ предоставляют платформы для разработки, в том числе Amazon, Baidu, Google, IBM и Microsoft. Данными компаниями предложены заранее обученные системы как стартовая точка для некоторых распространенных приложений, например, распознавание голоса. Поставщики процессоров, такие как Nvidia и AMD, тоже предлагают определенную поддержку [3].

Структура и свойства программного обеспечения

Основными составными частями программного обеспечения (Про) систем искусственного интеллекта (СИИ) являются:

- программно-аппаратные средства СИИ;
- программные средства представления знаний в СИИ;
- языки программирования и среды функционирования СИИ;
- инструментальные программные средства создания СИИ и др [20].

Основными особенностями Про, которые существенно отличают их от Про традиционных систем управления и обработки данных, являются свойства интеллектуальности Про, характерные для СИИ в целом. Основным

назначением ПрО СИИ является ориентация на символьную обработку информации и решение трудно формализуемых задач, которые не могут быть описаны в математической форме и не имеют алгоритмического решения.

Отличительной чертой ПрО СИИ является также программная поддержка и обеспечение общительности СИИ, то есть способности вести интенсивный диалог с пользователем на языке, близком к естественному, с широким использованием компьютерной графики, видеотехники и других современных средств интеллектуального интерфейса. По известным данным до 50% программного обеспечения СИИ приходится на реализацию диалога с пользователем [30].

Для настоящего этапа исследований в области создания ПрО СИИ характерна концентрация усилий в следующих направлениях:

- создание современных средств автоматизации программирования в русле осуществления восходящей стратегии развития ПрО СИИ: автокоды – языки высокого уровня – языки сверхвысокого уровня – языки спецификаций;
- создание новых языков представления знаний, которые нацелены на программную поддержку решения трудно формализуемых задач;
- разработка новых программных средств представления знаний путем использования новых языков обработки символьной информации;
- расширение базовых языков программирования до программных систем СИИ с помощью создания пакетов программ и специализированных библиотек представления знаний и прочее [23].

Принципы и методы создания программного обеспечения

К ПрО СИИ так же, как и к традиционному ПрО, предъявляются следующие основные требования:

- модифицируемость, которая позволяет вводить необходимые изменения в ПрО или исправлять ошибки, которые внесены в процессе разработки и функционирования системы;

- эффективность, которая подразумевает оптимальное использование времени работы и памяти системы;
- устойчивость к ошибкам, которые возникают при проектировании, реализации и функционировании системы;
- прозрачность ПрО для заказчика, разработчика и пользователя, которая позволяет «в моменте» корректировать ПрО при изменении предметной области [31].

Основными принципами построения ПрО, обеспечивающие выполнение предъявляемых к ПрО требований, являются следующие:

- абстрагирование компонентов ПрО с выделением существенных свойств и игнорированием несущественных деталей;
- сокрытие (упрятывание) информации с целью сделать недоступными детали ПрО, которые могут непосредственно или косвенно повлиять на процесс функционирования СИИ (например, при ошибочных действиях пользователя);
- модульность предполагает построение ПрО в виде функциональных (процедурно-ориентированных) или декларативных (объектно-ориентированных) модулей, которые позволяют быстро менять их при поломке или моральном износе;
- локализация компонентов ПрО предполагает создание слабо связанных между собой программных модулей, имеющих сильные внутренние связи;
- под однородностью ПрО понимают использование однотипных программных модулей;
- полнота ПрО заключается в использовании компонентов, которые обеспечивают функциональность СИИ;
- подтверждаемость ПрО означает получение одних и тех же результатов при одинаковых исходных данных каждой решаемой в СИИ задачи [34].

Абстракция и сокрытие информации способствуют модифицируемости и прозрачности ПрО. На каждом уровне абстракции используется лишь определенный набор операций и не допускаются операции, которые нарушают логические функции данного уровня, что позволяет существенно повысить надежность работы СИИ [4].

В настоящее время существуют основные методы разбиения (декомпозиции) ПрО на модули, описанные ниже:

- нисходящее структурное проектирование предполагает создание функциональных программных модулей путем оформления каждого шага процесса создания ПрО в виде модуля;
- структурирование ПрО по данным является альтернативой нисходящему проектированию, когда сначала определяются структуры данных, а затем на их основе определяется структура программных модулей;
- объектно-ориентированное проектирование позволяет создавать ПрО в прямом соответствии с решаемой проблемой и объектами предметной области [45].

Методы нисходящего проектирования концентрируют внимание на операциях, мало заботясь о проектировании структур данных. Методы структурирования ПрО по данным оперируют с объектами, а операции трактуют в обобщенном виде. Объектно-ориентированные методы объединяют в себе достоинства указанных методов, так как рассматривают объекты ПрО в качестве активных элементов, каждый из элементов наделен своим собственным набором операций [46].

Стадии разработки программного обеспечения

В соответствии с этапами разработки СИИ (в частности – экспертных систем) процесс создания ПрО СИИ состоит из следующих стадий: проектирования, реализации, сопровождения и совершенствования. Жизненный цикл ПрО называют упорядоченной последовательностью этих стадий.

Ключевой стадией процесса разработки ПрО СИИ является проектирование, основными задачами которого являются исследование выполнимости требований заказчика и разработка общей концепции построения ПрО. Стадия проектирования ПрО соответствует этапам идентификации, концептуализации и формализации процесса создания СИИ.

Основная цель стадии проектирования ПрО заключается в соблюдении требований заказчика. Однако заказчик редко обладает полными и последовательными определениями требований. Обычно и заказчик, и разработчик не в полной мере понимают проблему, поэтому требования совершенствуются в процессе разработки СИИ [48].

Правильный выбор концепции построения ПрО является самой критической частью процесса создания СИИ. Ошибки и просчеты, которые допустили на этапе проектирования, считаются самыми "дорогими". Поэтому основные усилия в области технологии программирования направлены, в основном, на автоматизацию проектирования ПрО.

На этапе проектирования инженер по знаниям вместе с программистом составляют формальные или формализованные модели (спецификации) представления знаний в СИИ. Также они определяют, подходят ли используемые языки программирования и инструментальные средства для решения имеющейся проблемы либо необходимы новые разработки.

На этапе проектирования устанавливаются также:

- основные характеристики ПрО; структура и состав ПрО;
- входные и выходные данные;
- тестовые примеры для проверки ПрО и прочее [50].

На этапе реализации обычно уже создаются один или несколько прототипов ПрО, последовательно приближающихся к требуемому уровню развития ПрО. По мере увеличения знаний прототип может достигнуть такого состояния, когда он успешно решает все задачи СИИ. Преобразование прототипа в конечный программный продукт обычно приводит к перепрограммированию ПрО на языках более низкого уровня, которые

обеспечивают как увеличение быстродействия ЭС, так и уменьшение требуемого объема базы знаний. Прототипы работают на проверку правильности идей, методов и способов, которые выбраны для построения СИИ [58].

Этап сопровождения ПрО СИИ заключается в тестировании, внедрении и модификации разработанного ПрО, которые могут вернуть весь процесс к стадии проектирования (перепроектирования). Существуют следующие методы тестирования ПрО:

- тестирование исходных данных, покрывающих всю область возможных ситуаций, которые могут возникнуть в процессе работы системы;
- тестирование ПрО базы знаний для того, чтобы обнаружить логические ошибки, к числу которых относятся:
 - а) избыточные, циклические и конфликтные правила;
 - б) пропущенные и пересекающиеся правила;
 - в) несогласуемые условия и др.;
- концептуальное тестирование проводится с целью проверки общей структуры ПрО и учета в нем всех аспектов решения задач системы/

На этапе сопровождения проверяется также пригодность ПрО для конечного пользователя путем решения всех возможных задач на рабочих местах пользователей заказчика. Пригодность ПрО определяется его полезностью и удобством работы с ним. Полезность ПрО заключается в способности решать поставленные задачи, удовлетворять потребности пользователя, а также выявлять и устранять (в ходе диалога) причины сбоев и неудач в работе СИИ.

Удобство работы с ПрО проявляется следующим образом:

- естественность взаимодействия с ПрО, то есть общение в привычном, удобном пользователю диалоге с СИИ;

- гибкость ПрО, то есть возможность работы с различными пользователями, которые имеют различную степень подготовленности к работе с СИИ;
- устойчивость ПрО к ошибкам, то есть его способность не выходить из строя при ошибочных действиях пользователя.

Крайний этап совершенствования ПрО заключается в использовании новых прогрессивных технологий создания и функционирования ПрО путем замены морально устаревших программных модулей современными и перспективными программными средствами [49].

1.4 ВЫВОДЫ ПО РАЗДЕЛУ

На основании рассмотренного теоретического материала по обработке данных с использованием систем искусственного интеллекта было установлено следующее.

Наиболее значительными работами в области искусственного интеллекта являются разработки мощных компьютерных систем или экспертных систем, то есть систем, основанных на знаниях. Такие программы решения задач с представлением и применением фактических и эвристических знаний, совместной работой экспертов и инженеров по знаниям, разработчиков систем и логическим выводом позволяют переходить к новым информационным технологиям, к новой технологии программирования.

2 РАЗРАБОТКА МЕТОДИКИ ИСПОЛЬЗОВАНИЯ СИСТЕМ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА ПРИ ОБРАБОТКЕ АНКЕТНЫХ ДАННЫХ

2.1 ПРИМЕНЕНИЕ СИСТЕМ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА ПРИ СБОРЕ И ОБРАБОТКЕ ДАННЫХ

Машинное обучение (англ. machine learning) – обширный подраздел искусственного интеллекта, математическая дисциплина, использующая разделы математической статистики, численных методов оптимизации, теории вероятностей, дискретного анализа и извлекающая закономерности из данных. Машинное обучение уже нашло применение в следующих областях: в биоинформатике, в медицине (медицинская диагностика), в геологии и геофизике, в социологии, в экономике (кредитный скоринг, предсказание оттока клиентов, обнаружение мошенничества, биржевой технический анализ, биржевой надзор), в технике (техническая диагностика, робототехника, компьютерное зрение, распознавание речи), в офисной автоматизации (распознавание текста, обнаружение спама, категоризация документов, распознавание рукописного ввода). Сфера применений машинного обучения постоянно расширяется. Повсеместная информатизация приводит к накоплению огромных объемов данных в науке, производстве, бизнесе, транспорте, здравоохранении. Возникающие при этом задачи прогнозирования, управления и принятия решений часто сводятся к обучению по прецедентам. Раньше, когда таких данных не было, эти задачи либо вообще не ставились, либо решались совершенно другими методами.

Системы с машинным обучением и технологиями искусственного интеллекта предполагают, что программа проводит обработку данных не по заданным правилам, а создает эти правила сами. Такие системы помогают лучше распознавать тексты, изображения и голосовые команды. Поэтому в

сфере рекрутинга они позволяют, например, лучше сканировать резюме и описания вакансий, а также создавать программы-ассистенты, которые будут отвечать на распространенные вопросы и давать первоначальную информацию.

Задача классификации – это определение принадлежности объекта наблюдения к определенному классу. При наличии множества объектов, у которых уже известна принадлежность к классам, можно построить алгоритм, который сможет определять, к какому из классов будет принадлежать новый объект. В терминах машинного обучения объекты с известной принадлежностью к классам называются обучающей выборкой, а задача определения принадлежности нового объекта к классу – классификацией. Алгоритм, с помощью которого определяется принадлежность к классу, – классификатор [37].

Таким образом, применение машинного обучения для отбора ответных анкет студентов заключается в следующем: по ранее собранным ответам на вопросы использован алгоритм, который определяет для вновь поступившей анкеты: будет ли она учтена для анализа эффективности учебного процесса. Рассмотрим возможность построения такого алгоритма.

2.2 СИСТЕМЫ СБОРА И ООБРАБОТКИ ДАННЫХ

Анкета представляет собой набор из 5 вопросов, на каждый из которых можно ответить по шкале от 1 до 10. Опрос студенты проходят в google forms. Собрана обширная информационная база данных с ответами на вопросы.

Для характеристики анкеты были выделены следующие признаки:

- наличие пропущенных ответов;
- наличие ответов по порядку (нисходящий/восходящий порядок);
- анкета содержит одинаковые ответы;
- наличие граничных ответов в анкете.

Для построения моделей и анализа данных используется Python – высокоуровневый язык программирования общего назначения, ориентированный на повышение производительности разработчика и читаемости кода. Синтаксис ядра Python минималистичен. В то же время набор стандартных библиотек включает в себя большой объем полезных функций.

На первом этапе импортируем данные из google forms в Excel. Далее необходимо создать обучающую выборку, которая составлена мной как экспертом, то есть я разделила все анкеты на 2 группы: полезные и бесполезные. Под бесполезными анкетами понимаем анкеты, признаки которых принимают следующие значения:

- количество отвеченных вопросов меньше пяти;
- ответы на все вопросы одинаковые;
- ответы (числа) идут в порядке возрастания/убывания с шагом 1;
- в анкете присутствуют максимальные и минимальные числа (ответы).

В машинном обучении признаки описывают объект в доступной и понятной для компьютера форме.

Следующим этапом импортируем библиотеки pandas для удобной работы с данными и numpy для работы с числами (рисунок 3.1).

```
In [5]: # импортируем библиотеку pandas для удобной работы с данными
import pandas as pd
# импортируем библиотеку numpy для работы с числами
import numpy as np
```

Рисунок 3.1 – Импорт библиотек⁵

Далее считываем полезные (рисунок 3.2) и бесполезные анкеты (рисунок 3.3) при помощи библиотеки pandas.

```
In [15]: # считываем полезные анкеты при помощи pandas
good = pd.read_csv('/Users/Tamara/Downloads/good.csv', delimiter=';', header=None, dtype='int32')
# выводим первые 5 строк
good.head(5)
```

```
Out[15]:
```

	0	1	2	3	4
0	9	10	10	10	10
1	10	7	10	10	7
2	8	8	9	9	8
3	9	9	10	10	9
4	10	9	10	9	10

Рисунок 3.2 – Считывание полезных анкет⁶

```
In [12]: # считываем бесполезные анкеты при помощи pandas
bad = pd.read_csv('/Users/Tamara/Downloads/bad.csv', delimiter=';', header=None)
# заменяем отсутствующие значения на 0
bad = bad.fillna(0)
# делаем все значения целыми числами
bad = bad.astype('int32')
# выводим первые 5 строк
bad.head(5)
```

```
Out[12]:
```

	0	1	2	3	4
0	10	9	10	10	1
1	1	1	1	1	10
2	1	1	1	1	1
3	1	1	1	1	1
4	1	1	1	1	1

Рисунок 3.3 – Считывание бесполезных анкет⁷

Четвертым этапом объявляем функцию `get_features`, которая извлекает из анкет признаки для классификаторов. Все ответы принимают значения 0 или 1.

⁵ Составлено автором

⁶ Составлено автором

⁷ Составлено автором

Анкета принимает значение 1, если она удовлетворяет одному из четырех признаков бесполезности (рисунок 3.4).

```
In [17]: # функция, которая извлекает из анкет признаки для классификаторов
# ответы принимают значения 0 или 1. 1 значит, что в анкете наблюдается какое-либо свойство
def get_features(feedback):
    feedback = list(feedback)
    # будет 4 признака
    features = [0] * 4
    # первый признак - если больше одного 0 (ранее ноль был пустым значением)
    if feedback.count(0) > 0:
        features[0] = 1
    # второй признак - количество разных чисел = 1
    if len(set(feedback)) == 1:
        features[1] = 1
    # третий признак - в анкете были и 1, и 10
    if feedback.count(1) > 0 and feedback.count(10) > 0:
        features[2] = 1
    # четвертый признак - анкета представляет собой возрастающий или убывающий ряд
    ascending = True
    descending = True
    for i in range(1, len(feedback)):
        if ascending and feedback[i] - 1 < feedback[i-1]:
            ascending = False
        if descending and feedback[i] > feedback[i-1] - 1:
            descending = False
    if ascending or descending:
        features[3] = 1
    return features
```

Рисунок 3.4 – Задание признаков бесполезности⁸

Далее необходимо посчитать признаки и заполнить ответы для полезных и бесполезных анкет, то есть определяем, обладает ли анкета признаками бесполезности (рисунок 3.5).

```
In [19]: # X - признаки
X = []
# y - ответы
y = []
# вызовем функцию, которая формирует признаки сначала для полезных, потом для бесполезных анкет
for index, row in good.iterrows():
    X.append(get_features(row.tolist()))
    y.append(0)
for index, row in bad.iterrows():
    X.append(get_features(row.tolist()))
    y.append(1)
```

Рисунок 3.5 – Подсчет признаков и заполнение ответов⁹

После этого перемешиваем массивы с полученными ответами и признаками, чтобы нарушить последовательность (рисунок 3.6).

⁸ Составлено автором

⁹ Составлено автором

```
In [20]: # перемешаем массивы с признаками и ответами
from sklearn.utils import shuffle
X, y = shuffle(X, y)
```

Рисунок 3.6 – Изменение последовательности¹⁰

Когда данные для построения моделей классификации подготовлены, остается определить, каким образом необходимо оценивать точность работы моделей классификации – классификаторов. Для этого подготовленные данные разбиваются на две выборки – тренировочную и тестовую. После обучения классификатора на тренировочной выборке его оценивают на тестовой выборке. Разница в предсказаниях и будет являться критерием качества работы классификатора. Однако, при разбиении данных есть вероятность получить несбалансированные выборки (выборки, в которых объектов одного класса может быть значительно больше, чем другого). Поэтому для разбиения будем использовать кросс-валидацию. Кросс-валидация – это техника многостадийного разбиения данных. При каждом шаге данные разбиваются на части, затем происходит обучение модели на тренировочной выборке и валидация модели на тестовой выборке. Для уменьшения ошибки классификатора на каждом шаге происходит разбиение на разные части исходного набора данных (рисунок 3.7).

¹⁰ Составлено автором

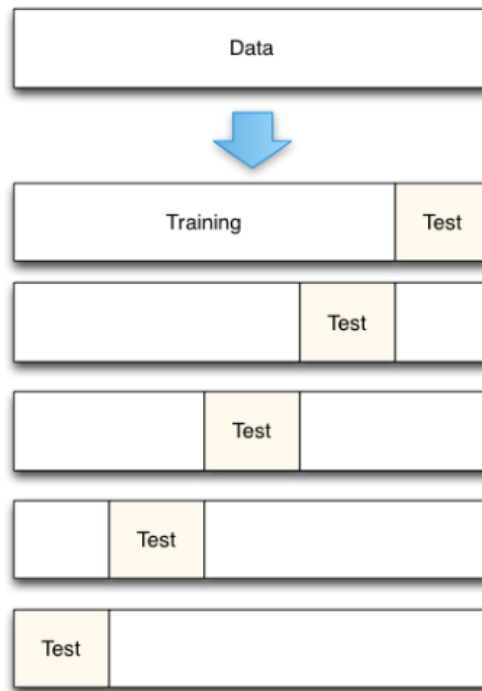


Рисунок 3.7 - Графическое пояснение метода кросс-валидации данных¹¹

Существует множество классификаторов, использующих разные алгоритмы классификации. У каждого классификатора есть свои преимущества и недостатки, потому главным критерием оценки классификатора является его точность. Построим ряд классификаторов, оценим точность их работы, а также важность каждого признака в конкретном алгоритме.

Логистическая регрессия – метод построения линейного классификатора, позволяющий оценивать апостериорные вероятности принадлежности объектов классам [56] (рисунок 3.8).

```
In [21]: # обучим логистическую регрессию
# точность будем мерять по кросс-валидации
from sklearn.linear_model import LogisticRegression
from sklearn.model_selection import cross_val_score
clf = LogisticRegression()
print(np.mean(cross_val_score(clf, X, y, scoring='accuracy', cv=5)))
# Посмотрим теперь еще важности признаков для данной модели
clf.fit(X, y)
print(clf.coef_[0])

0.9740986098609861
[3.08930446 5.91137186 0.          1.56719808]
```

Рисунок 3.8 – Логистическая регрессия¹²

¹¹ Составлено автором
¹² Составлено автором

Решающее дерево – алгоритм, воспроизводящий логические схемы, позволяющие получить окончательное решение о классификации объекта с помощью ответов на иерархически организованную систему вопросов [53]. Причем вопрос, задаваемый на последующем иерархическом уровне, зависит от ответа, полученного на предыдущем уровне (рисунок 3.9).

```
In [23]: # обучим решающее дерево
# точность будем мерять по кросс-валидации
from sklearn.tree import DecisionTreeClassifier
clf = DecisionTreeClassifier()
print(np.mean(cross_val_score(clf, X, y, scoring='accuracy', cv=5)))
clf.fit(X, y)
print(clf.feature_importances_)

0.9880194019401939
[0.06241381 0.91927884 0.          0.01830735]
```

Рисунок 3.9 – Решающее дерево¹³

Случайный лес – алгоритм машинного обучения, заключающийся в использовании комитета (ансамбля) решающих деревьев [61] (рисунок 3.10).

```
In [25]: # обучим случайный лес
# точность будем мерять по кросс-валидации
from sklearn.ensemble import RandomForestClassifier
clf = RandomForestClassifier(n_estimators=100)
print(np.mean(cross_val_score(clf, X, y, scoring='accuracy', cv=5)))
# Посмотрим теперь еще важности признаков для данной модели
clf.fit(X, y)
print(clf.feature_importances_)

0.9880194019401939
[0.04992395 0.93380547 0.          0.01627058]
```

Рисунок 3.10 – Случайный лес¹⁴

Стохастический градиентный спуск – оптимизационный алгоритм, при котором для корректировки параметров модели используется градиент [55] (рисунок 3.11).

¹³ Составлено автором

¹⁴ Составлено автором

```
In [26]: # обучим стохастический градиентный спуск
# точность будем мерять по кросс-валидации
from sklearn.linear_model import SGDClassifier
clf = SGDClassifier(max_iter=1000)
print(np.mean(cross_val_score(clf, X, y, scoring='accuracy', cv=5)))
# Посмотрим теперь еще важности признаков для данной модели
clf.fit(X, y)
print(clf.coef_)

0.9880194019401939
[[2.03187656 2.0119562 0.          2.0119562 ]]
```

Рисунок 3.11 - Стохастический градиентный спуск¹⁵

Метод К ближайших соседей – простейший метрический классификатор, основанный на оценивании сходства объектов. Классифицируемый объект относится к тому классу, которому принадлежат ближайшие к нему К объектов обучающей выборки [59] рисунок 3.12).

```
In [27]: # обучим метод ближайших соседей
# точность будем мерять по кросс-валидации
from sklearn.neighbors import KNeighborsClassifier
clf = KNeighborsClassifier(n_neighbors=5)
print(np.mean(cross_val_score(clf, X, y, scoring='accuracy', cv=5)))
# В этой модели важности признаков посмотреть нельзя

0.984039203920392
```

Рисунок 3.12 – Метод ближайших соседей¹⁶

2.3 ВЫВОДЫ ПО РАЗДЕЛУ

В ходе исследования была выявлена проблема обработки анкетных данных ввиду высоких трудозатрат и нерелевантных полученных результатов.

На основе проведенных исследований в качестве метода для автоматизации процесса отбора анкетных данных рекомендуется использовать классификаторы «Случайный лес», «Стохастический градиентный спуск» и «Решающее дерево». Также можно сказать, что параметр «Наличие граничных ответов в анкете» абсолютно не важен для каждого из алгоритмов.

¹⁵ Составлено автором

¹⁶ Составлено автором

Предложенные алгоритмы машинного обучения решают описанную выше проблему: после внедрения классификатора трудозатраты снизятся, так как уменьшится количество анкет, которые необходимо обрабатывать, а полученные результаты будут иметь значимость.

3 ПРИМЕНЕНИЕ СИСТЕМ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА ПРИ СБОРЕ И ОБРАБОТКЕ АНКЕТНЫХ ДАННЫХ

3.1 ОБЩАЯ ХАРАКТЕРИСТИКА ОРГАНИЗАЦИИ

3.1.1 ОПИСАНИЕ ИНСТИТУТА ВШЭМ УРФУ

Объектом рассмотрения данной магистерской диссертации является институт «Высшая школа экономики и менеджмента Уральского федерального университета им. первого Президента России Б.Н. Ельцина». ВШЭМ создана как полнопрофильное структурное подразделение УрФУ, где сосредоточена подготовка по всем основным (бакалавриат, магистратура) и дополнительным направлениям (программы МВА, программы профессиональной переподготовки и повышения квалификации) высшего профессионального образования, подготовка кадров высшей квалификации (аспирантура, докторантура) по укрупненной группе «Экономика и управление», а также фундаментальные и прикладные научные исследования и консалтинг в области экономики, менеджмента и промышленного бизнеса. ВШЭМ УрФУ выполняет функции бизнес-школы УрФУ.

ВШЭМ УрФУ создана на базе профильных подразделений Уральского федерального университета (факультет экономики и управления, факультет информационно-математических технологий и экономического моделирования, бизнес-школа), Уральского государственного университета им. А.М. Горького (экономический факультет) с включением кафедр непрофильных подразделений университетов:

- факультет экономики и управления УрФУ (ФЭУ УрФУ, 15 кафедр):
 - 1) кафедра экономики и управления на предприятиях металлургии;
 - 2) кафедра экономики и организации химической промышленности;
 - 3) кафедра экономики и организации предприятий машиностроения;

- 4) кафедра систем управления энергетикой и промышленными предприятиями;
 - 5) кафедра экономики энергетики и маркетинга;
 - 6) кафедра экономики и управления строительством и рынком недвижимости;
 - 7) кафедра управления внешнеэкономической деятельностью предприятий;
 - 8) кафедра экономики и управления качеством продукции;
 - 9) кафедра экономической безопасности;
 - 10) кафедра общей экономической теории;
 - 11) кафедра правового регулирования экономической деятельности;
 - 12) кафедра государственных и муниципальных финансов;
 - 13) кафедра банковского дела;
 - 14) кафедра финансового и налогового менеджмента;
 - 15) кафедра учета, анализа и экономики труда;
- факультет информационно-математических технологий и экономического моделирования УрФУ (ФИМТиЭМ УрФУ, 3 кафедры):
- 1) кафедра анализа систем и принятия решений;
 - 2) кафедра моделирования управляемых систем;
 - 3) кафедра мультимедиа технологий;
- бизнес-школа УрФУ (БШ УрФУ, 2 кафедры):
- 1) кафедра предпринимательства и инноваций;
 - 2) кафедра стратегического менеджмента;
- экономический факультет УрГУ (ЭФ УрГУ, 6 кафедр):
- 1) кафедра экономической теории;
 - 2) кафедра мировой экономики;
 - 3) кафедра теории и практики менеджмента;
 - 4) кафедра эконометрики и статистики;
 - 5) кафедра финансов, денежного обращения и кредита;

- б) кафедра финансового менеджмента и права;
- 7) кафедры других подразделений университетов (1 кафедра):
- 8) кафедра менеджмента факультета гуманитарного образования (ФГО) УрФУ [1].

3.1.2 СТРАТЕГИЧЕСКИЕ ЦЕЛИ И ЗАДАЧИ ВШЭМ УРФУ

Миссия ВШЭМ УрФУ – содействие инновационному росту и глобальной конкурентоспособности экономики Большого Урала и России путем опережающей подготовки интеллектуальной элиты, создания и распространения знаний мирового уровня и формирования соответствующих компетенций в сфере экономики, менеджмента и инновационного развития.

Стратегической целью ВШЭМ УрФУ является формирование в Уральском федеральном округе российской региональной бизнес-школы и школы подготовки экономистов с мировым признанием, ведущего российского центра исследований в области экономики, менеджмента и промышленного бизнеса, а также ведущего регионального центра партнерства власти, бизнеса, науки и образования в области экономики и менеджмента, что однозначно обеспечит лидерство университета в области экономики и менеджмента, его вхождение в число ведущих мировых образовательных и интеллектуальных центров в этой области. Достижение указанной цели возможно в результате решения следующих задач:

- модернизация образовательного процесса в области экономики, менеджмента и промышленного бизнеса;
- модернизация научно-исследовательского процесса и инновационной деятельности в области экономики, менеджмента и промышленного бизнеса;
- развитие кадрового потенциала ВШЭМ УрФУ и формирование качественного контингента обучающихся;
- модернизация инфраструктуры ВШЭМ УрФУ;

- совершенствование организационной структуры управления на уровне ВШЭМ УрФУ и повышение эффективности управления;
- формирование на базе ВШЭМ первой российской региональной бизнес-школы мирового уровня, что предполагает её системную интернационализацию:
 - 1) создание и успешную реализацию инновационных образовательных программ подготовки менеджеров мирового уровня, включая систему управления качеством образования в соответствии с международными стандартами;
 - 2) создание международного профессорско-преподавательского состава высшей квалификации, признание ВШЭМ как одного из ведущих центров исследований вопросов экономического развития и бизнеса на развивающихся рынках;
 - 3) получение наиболее престижных международных аккредитаций и построение в увязке с ними сильного бренда ВШЭМ.

Стратегическая цель и задачи создания и развития ВШЭМ УрФУ полностью соответствуют программе развития УрФУ. ВШЭМ УрФУ рассматривается как структурное подразделение университета, отвечающее за реализацию одного из приоритетных направлений его научно-образовательной и инновационной деятельности – направление «Экономика и управление».

На рисунке 1 представлены основные стратегические цели ВШЭМ УрФУ.

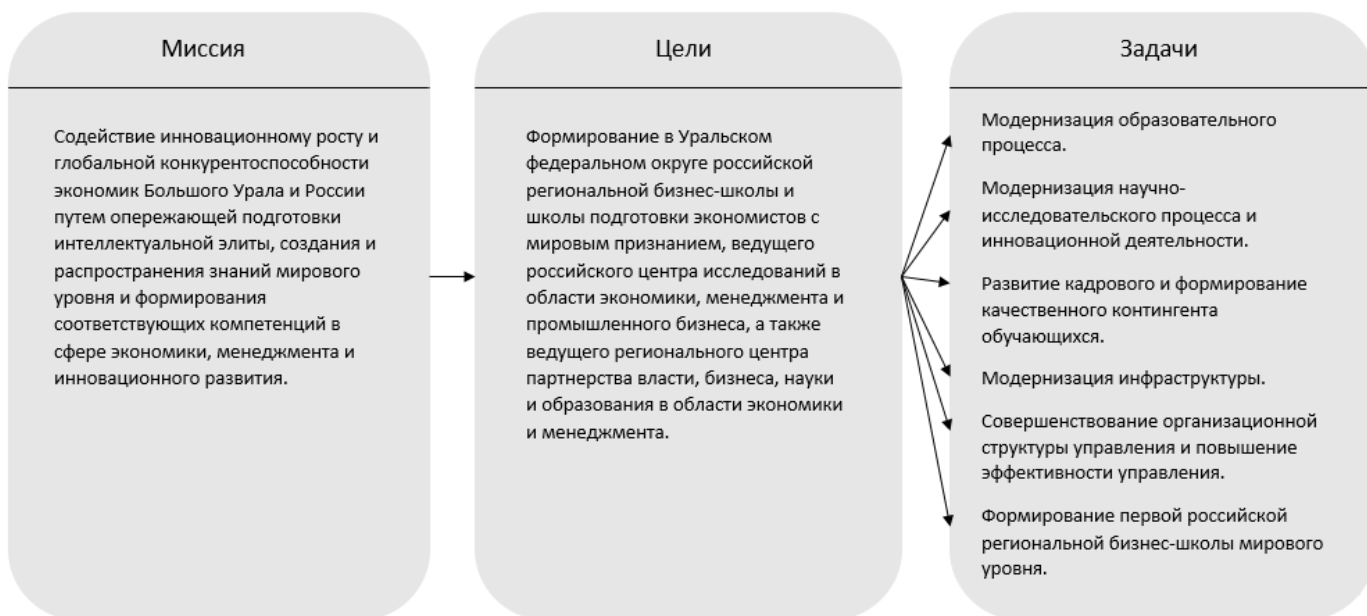


Рисунок 2.1 – Основные стратегические цели ВШЭМ УрФУ¹⁷

На рисунке 2 представлена бизнес-стратегия компании.

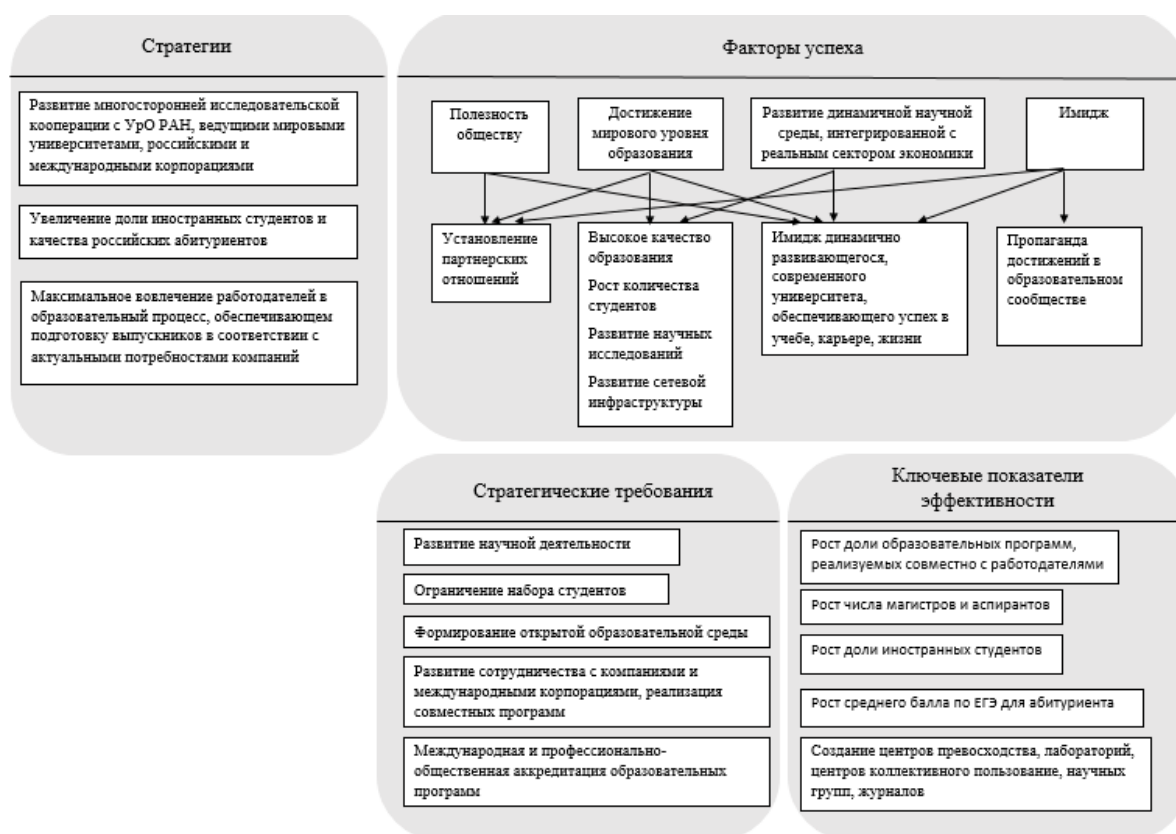


Рисунок 2.2 – Бизнес-стратегия УрФУ¹⁸

¹⁷ Составлено автором по: [1]

¹⁸ Составлено автором по: [1]

3.1.3 ОСНОВНЫЕ БИЗНЕС-ПРОЦЕССЫ СЛУЖБЫ

ДИСПЕТЧЕРИЗАЦИИ ВШЭМ УРФУ

Бизнес-процессы являются фундаментальными элементами, которые оказывают влияние на успешную деятельность организации. В настоящее время структура бизнес-процессов представляет собой совокупность трёх составляющих:

- основные бизнес-процессы;
- бизнес-процессы управления;
- вспомогательные бизнес-процессы.

Основными бизнес-процессами являются процессы, которые ориентированы на производство товара либо оказание услуги, являющиеся главными объектами создания предприятия и обеспечивающие получение прибыли.

Бизнес-процессы управления – это процессы, охватывающие весь комплекс функций управления на уровне каждого БП и бизнес-системы в целом. Это процессы стратегического, оперативного и текущего планирования, формирования и осуществления управленческих воздействий.

Вспомогательные бизнес-процессы – процессы, предназначенные для обеспечения выполнения основных БП и поддержания их специфических черт.

В службе диспетчеризации ВШЭМ УрФУ существует 4 основных бизнес-процесса:

- составление расписаний учебных занятий согласно рабочим планам по направлениям подготовки;
- составление расписаний зачётно-экзаменационных сессий согласно рабочим планам по направлениям подготовки;
- диспетчеризация аудиторного фонда учебного процесса ВШЭМ;
- анкетирование студентов на предмет удовлетворенности учебным процессом.

К бизнес-процессам управления можно отнести следующие процессы:

- контроль за формированием в электронной базе Университета учебных и рабочих планов кафедрами ВШЭМ;
- контроль планирования и организации мероприятий учебно-методического характера;
- контроль выполнения учебной нагрузки кафедр и реализации утвержденного расписания учебных занятий;
- решение текущих координационных задач, связанных с учебным процессом.

К вспомогательным бизнес-процессам относятся:

- оформление выписок из расписаний учебных занятий для преподавателей;
- подготовка консолидированных отчетов по направлениям деятельности;
- согласование расписаний с профессорско-преподавательским составом;
- обеспечение доступа к аудиториям;
- проверка целостности состояния оборудования в аудиториях;
- обеспечение непрерывности учебного процесса в течение дня и равномерности распределения учебной работы в течение учебной недели в расписании (рисунок 3).

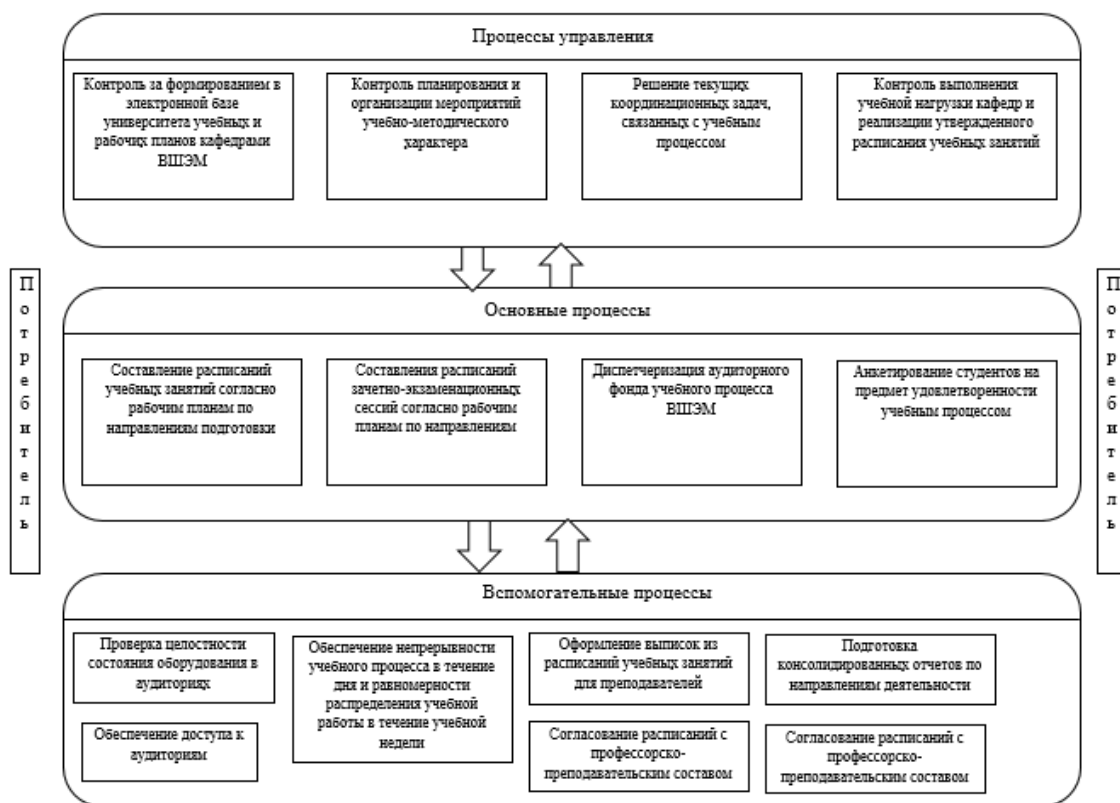


Рисунок 2.3 – Бизнес-процессы диспетчерской службы¹⁹

Исходя из описанных выше бизнес-процессов, можно выделить основные бизнес-функции компании (рисунок 4).

¹⁹ Составлено автором по: [1]

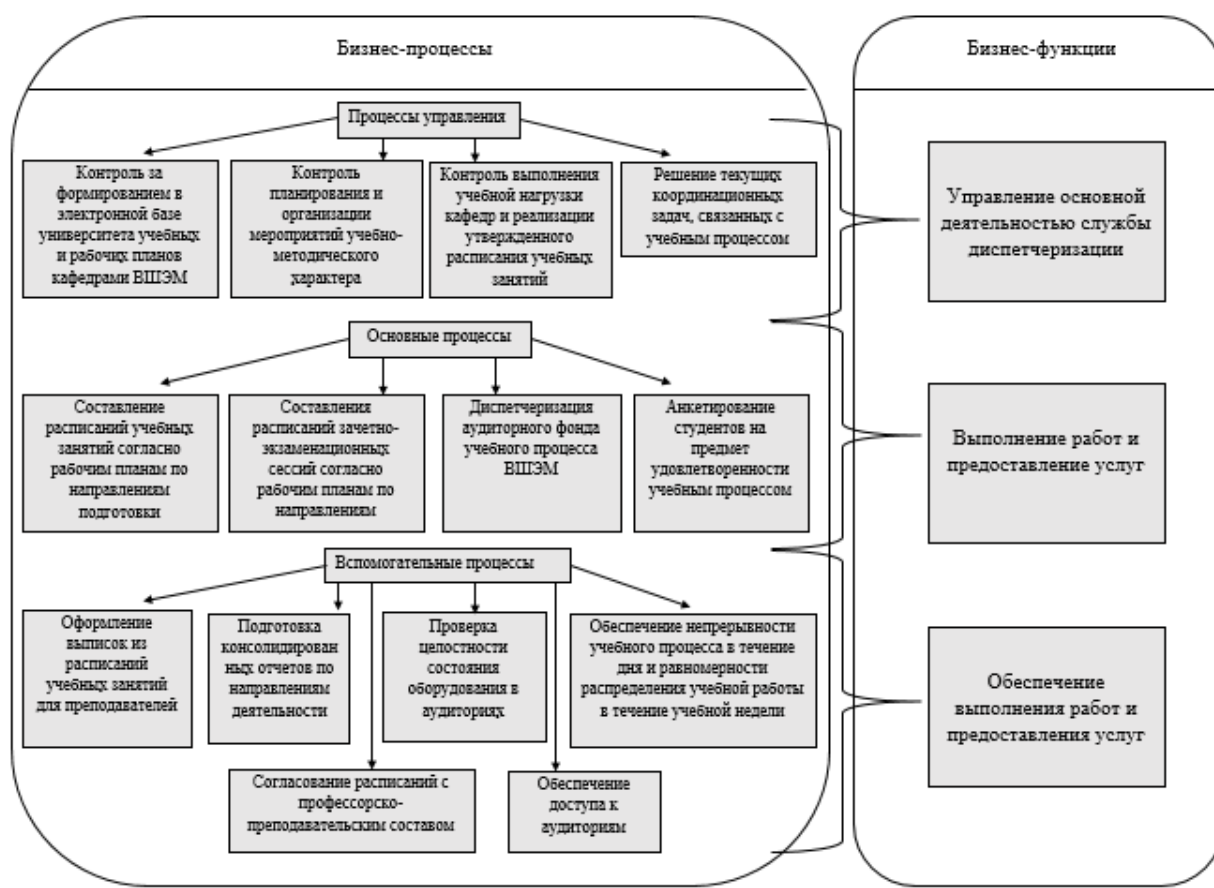


Рисунок 2.4 – Связь бизнес-процессов и бизнес-функций²⁰

3.1.4 ОРГАНИЗАЦИОННАЯ СТРУКТУРА СЛУЖБЫ ДИСПЕТЧЕРИЗАЦИИ ВШЭМ УРФУ

Организационная структура службы диспетчеризации ВШЭМ УрФУ является линейной, возглавляет службу начальник отдела, в подчинении которого заместитель начальника отдела и работники диспетчерской службы: ведущие документоведы и диспетчер. Организационная структура представлена на рисунке 5.

²⁰ Составлено автором по: [1]



Рисунок 2.5 – Модель организационной структуры службы диспетчеризации ВШЭМ УрФУ²¹

3.1.5 ПРОДУКТЫ И УСЛУГИ

Служба диспетчеризации предоставляет следующие услуги (рисунок 6):

- Составление расписаний учебных занятий: загрузка на сайт УрФУ расписаний учебных занятий, предоставление расписаний учебных занятий в деканаты, согласование расписаний с профессорско-преподавательским составом, назначение аудиторий.
- Составление расписаний экзаменационных сессий: загрузка на сайт УрФУ расписаний экзаменационных сессий, предоставление расписаний экзаменационных сессий в деканаты, согласование расписаний с профессорско-преподавательским составом, назначение аудиторий.
- Диспетчеризация аудиторного фонда учебного процесса ВШЭМ: назначение аудиторий, обеспечение доступа к аудиториям, настройка мультимедийного оборудования.

²¹ Составлено автором по: [1]

- Анкетирование студентов на предмет удовлетворенности учебным процессом: составление анкет, предоставление студентам анкет, предоставление кафедрам результатов анкетирования.
- Составление электронной отчетности.

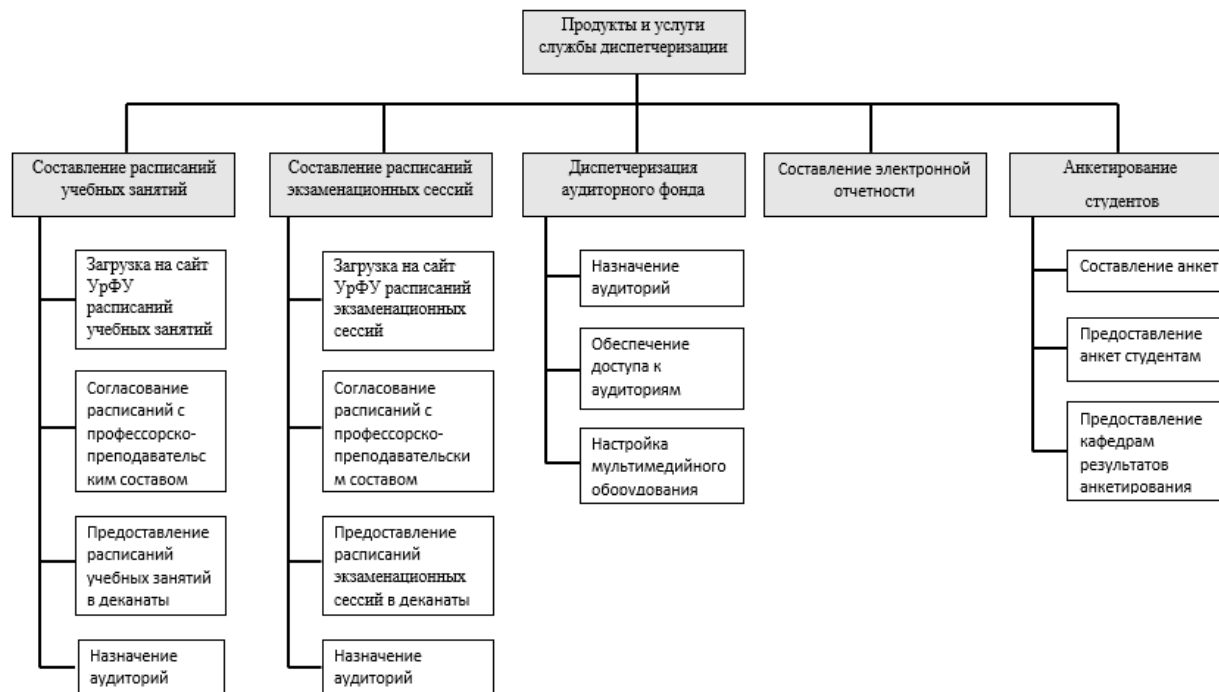


Рисунок 2.6 – Модель продуктов и услуг²²

Служба диспетчеризации ВШЭМ оказывает услуги для поддержки образовательного процесса, основными являются составление расписаний и диспетчеризация аудиторного фонда, без которых образовательный процесс невозможен. Составление электронной отчетности, анкетирование студентов – второстепенные услуги.

Расписание учебных занятий ВУЗа (в первую очередь для очного отделения) представляет собой результат распределения учебных занятий по учебным неделям с указанием преподавателя, времени и места проведения занятий. Должны быть устранены все накладки, неизбежно возникающие в процессе составления расписаний, занятия должны быть распределены на соответствующие смены и в подходящие для них по типу и объему аудитории. Кроме того, должны быть по возможности учтены пожелания отдельных

²² Составлено автором по: [1]

преподавателей относительно их занятий. Расписание составляется без "окон" для учебных групп с учетом особенностей учебного заведения, учебно-методических и нормативных документов учебно-методического управления, норм СанПиН. Таким образом, при составлении расписания необходимо учитывать большое число факторов, среди которых:

- временные факторы;
- временной фонд, заданный в учебных планах;
- допустимый диапазон пар для каждой группы;
- временной режим каждого преподавателя;
- факторы, связанные с взаимным расположением занятий;
- максимальное число занятий для каждой группы в каждый день недели;
- необходимое время для перехода из корпуса в корпус;
- группировка по 2 и более занятий непрерывно;
- имеющийся аудиторный фонд;
- список допустимых аудиторий для каждого предмета;
- список допустимых типов аудиторий для каждого предмета;
- список допустимых корпусов для каждого предмета.

Оперативное управление аудиторным фондом является одной из важнейших составляющих оперативного решения задачи оптимизационного составления расписания учебных занятий. К задачам управления аудиторным фондом ВУЗа можно отнести:

- ведение аудиторного фонда (назначение, вместимость, удаленность);
- закрепление аудиторий (за кафедрами, преподавателями, дисциплинами);
- оперативное распределение аудиторного фонда во времени;
- долговременное распределение аудиторного фонда;
- проведение занятий вне учебных корпусов;

- обработка статистики об эффективности использования аудиторного фонда для учебного процесса.

Составление отчетности – завершающий этап учетного процесса, поэтому она состоит из обобщающих итоговых показателей, которые получают в конце отчетного периода путем соответствующей обработки данных текущего учета. Специалисты службы диспетчеризации предоставляют отчетность начальнику отдела организации учебной работы о численности штата профессорско-преподавательского состава, его загруженности, количестве аудиторий и их оснащённости необходимым оборудованием. Отчётность, предоставляемая в бухгалтерию – информация о нагрузке преподавателей.

Анкетирование студентов проводится на предмет удовлетворенности образовательным процессом. Информация об анкетировании предоставляется кафедрам, где решаются выявленные проблемы для того, чтобы образовательный процесс стал более эффективным.

3.1.6 МОДЕЛЬ ИНФОРМАЦИОННЫХ СИСТЕМ

Модель информационных систем представлена на рисунке 7.

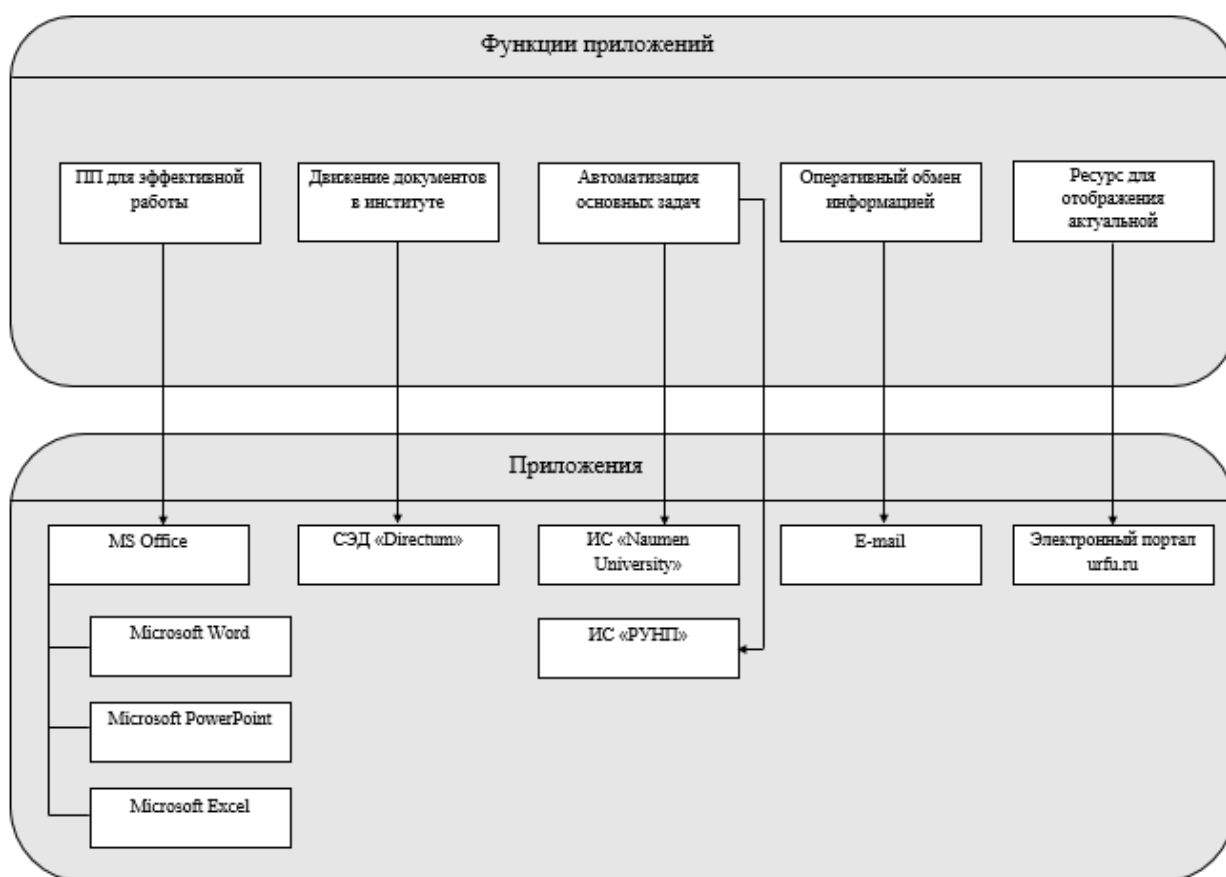


Рисунок 2.7 – Модель информационных систем²³

Microsoft Word – программа для работы с текстовыми документами. Этот текстовый процессор уже долгое время является самым популярным из всех используемых в данный момент. Именно поэтому, формат документов .doc стал стандартом для современного документооборота, а многие конкурирующие программы имеют поддержку совместимости с данным форматом. Документ, созданный в данном приложении, может включать в себя графические изображения, таблицы, звуковое сопровождение, фрагменты видеофильмов. Текстовый процессор ускоряет создание документа за счет сокращения количества действий, выполняемых пользователем. Word автоматизирует набор текста, оформление оглавления и предметного указателя, проверку правописания.

Microsoft PowerPoint – программа для создания красочных, наглядно оформленных мультимедийных презентаций. Материалы, подготовленные с

²³ Составлено автором по: [1]

помощью PowerPoint предназначены для отображения на большом экране – через проектор, либо телевизионный экран большого размера. Каждая презентация состоит из слайдов, которые сопровождаются различными эффектами. PowerPoint, как Word, позволяет пользователю менять шрифт, стиль текста, вставлять графические объекты и многое другое.

Microsoft Excel – программа, позволяющая с помощью электронных таблиц анализировать данные (таблицы, диаграммы, графики), выполнять вычисления и работать со списками. Оно упрощает доступ и анализ бизнес-информации, хранящейся на персональном компьютере, в сети и на веб-страницах. Прямоугольные таблицы широко используются для упорядоченного хранения данных и наглядного представления чисел или текстовой информации. В таблице может быть отображена как исходная информация, так и результаты выполнения арифметических, логических или иных операций над исходными данными. В службе диспетчеризации Excel применяется для статистической обработки информации, для анализа данных.

Электронная почта – технология и сервис по пересылке и получению электронных сообщений между пользователями компьютерной сети. Посредством электронной почты работники диспетчерской службы ведут деловую переписку, а также осуществляют рассылку анкет для заполнения студентами, производят их сбор.

Информационная система Naumen University – информационно-аналитическая система для организации управления процессом обучения в ВУЗах заведениях. Внедрение Naumen University позволяет комплексно подойти к решению задач, стоящих перед университетом. Основные задачи, которые решаются с помощью данной ИС (применительно к службе диспетчеризации ВШЭМ УрФУ):

- учет и ведение личных дел студентов, сотрудников, абитуриентов, аспирантов, совместителей;
- формирование и утверждение учебных и рабочих планов, справочник ГОСов;

- доступ к личному кабинету для сотрудника, студента, абитуриента, аспиранта, совместителя;
- расписание учебных мероприятий, аттестационных мероприятий в период сессии;
- управление аудиторным фондом;
- расчет нагрузок на ППС и читающие кафедры;
- возможность удаленного доступа к единому банку данных и получения актуальной информации.

Информационная система «РУНП», разработанная УрФУ, посредством которой производится расчет учебной нагрузки преподавателей.

Электронный портал ufu.ru используется службой диспетчеризации для загрузки расписаний учебных занятий, экзаменационных сессий как для студентов, так и для преподавателей.

Система электронного документооборота «Directum» и управления взаимодействием, нацеленная на рост эффективности работы всех сотрудников университета. Система «Directum» поддерживает все этапы производственного цикла управления документами, при этом традиционное «бумажное» делопроизводство имеет место быть и используется параллельно с электронным документооборотом. Данная система обеспечивает эффективную организацию и контроль деловых процессов: согласование документов, обработка сложных заказов, подготовка и проведение совещаний и других процессов взаимодействия.

3.1.7 ИТ-ИНФРАСТРУКТУРА СЛУЖБЫ ДИСПЕТЧЕРИЗАЦИИ ВШЭМ

УрФУ

На рисунке 8 представлена текущая ИТ-инфраструктура службы диспетчеризации ВШЭМ.

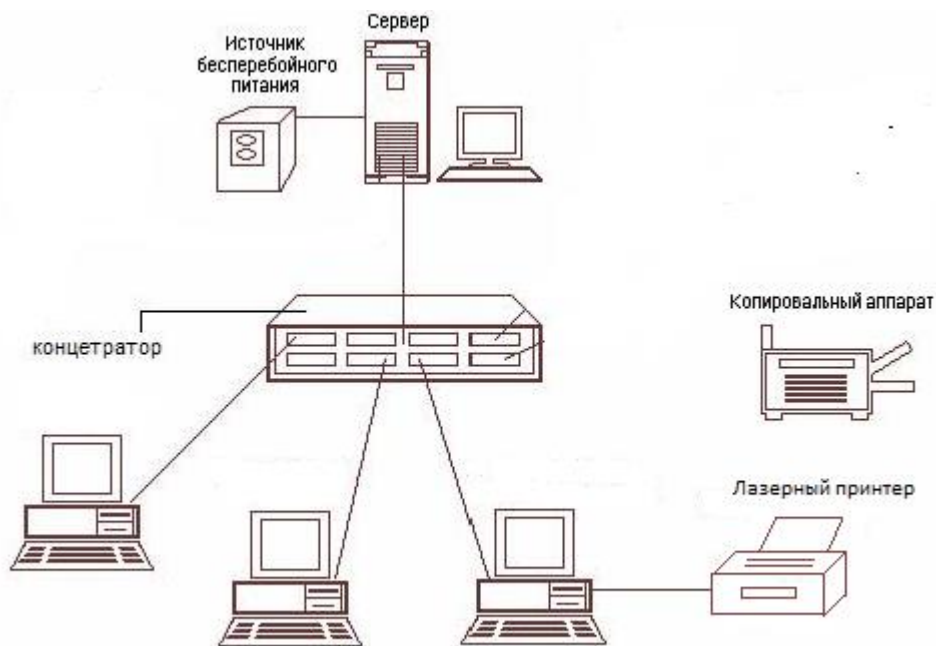


Рисунок 2.8 – Текущая ИТ-инфраструктура²⁴

3.1.8 МОДЕЛИРОВАНИЕ АРХИТЕКТУРЫ СЛУЖБЫ ДИСПЕТЧЕРИЗАЦИИ ВШЭМ УРФУ

Стратегические цели и задачи компании непосредственно связаны с бизнес-процессами. На рисунке 9 представлены связи между стратегиями развития университета и существующими бизнес-процессами.

²⁴ Составлено автором

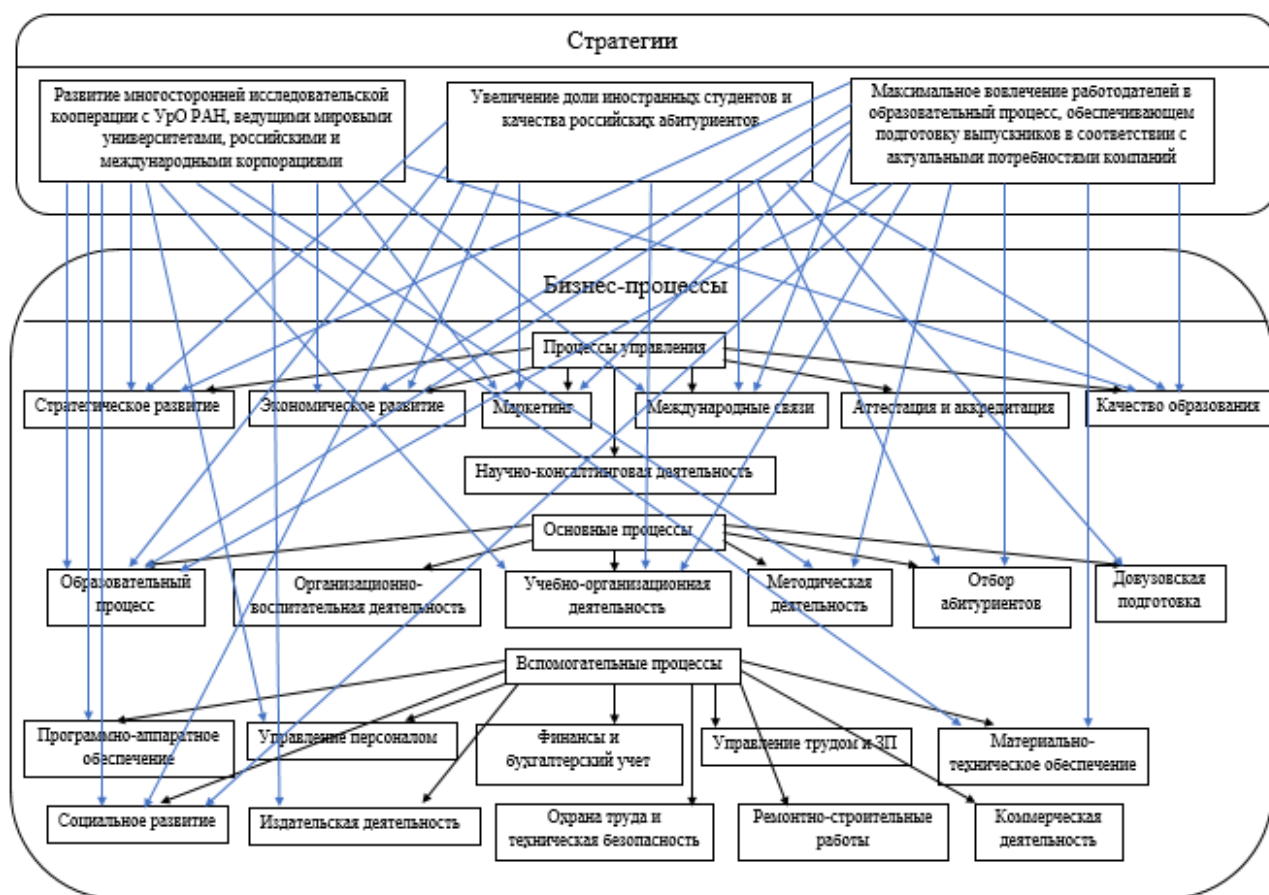


Рисунок 2.9 – Модель, связывающая стратегические цели и бизнес-процессы²⁵

Каждое приложение используется в компании для выполнения определенных бизнес-процессов. На рисунке 10 представлена схема связи бизнес-процессов и приложений.

²⁵ Составлено автором по: [1]

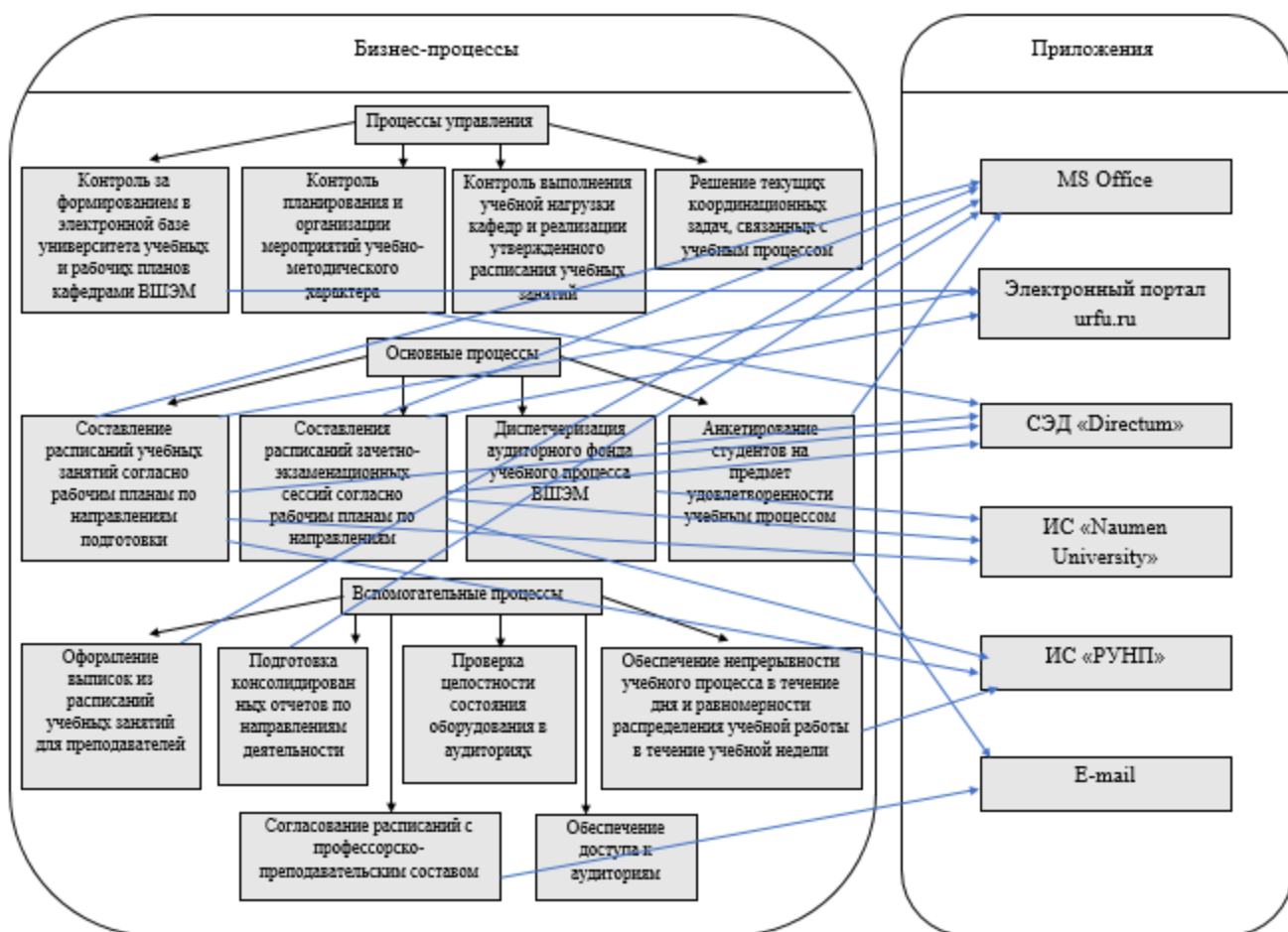


Рисунок 2.10 – Схема связи процессов и приложений²⁶

Общая модель архитектуры службы диспетчеризации ВШЭМ УрФУ представлена на рисунке 11. На модели представлены все объекты, которые можно использовать при построении архитектуры.

²⁶ Составлено автором по: [1]

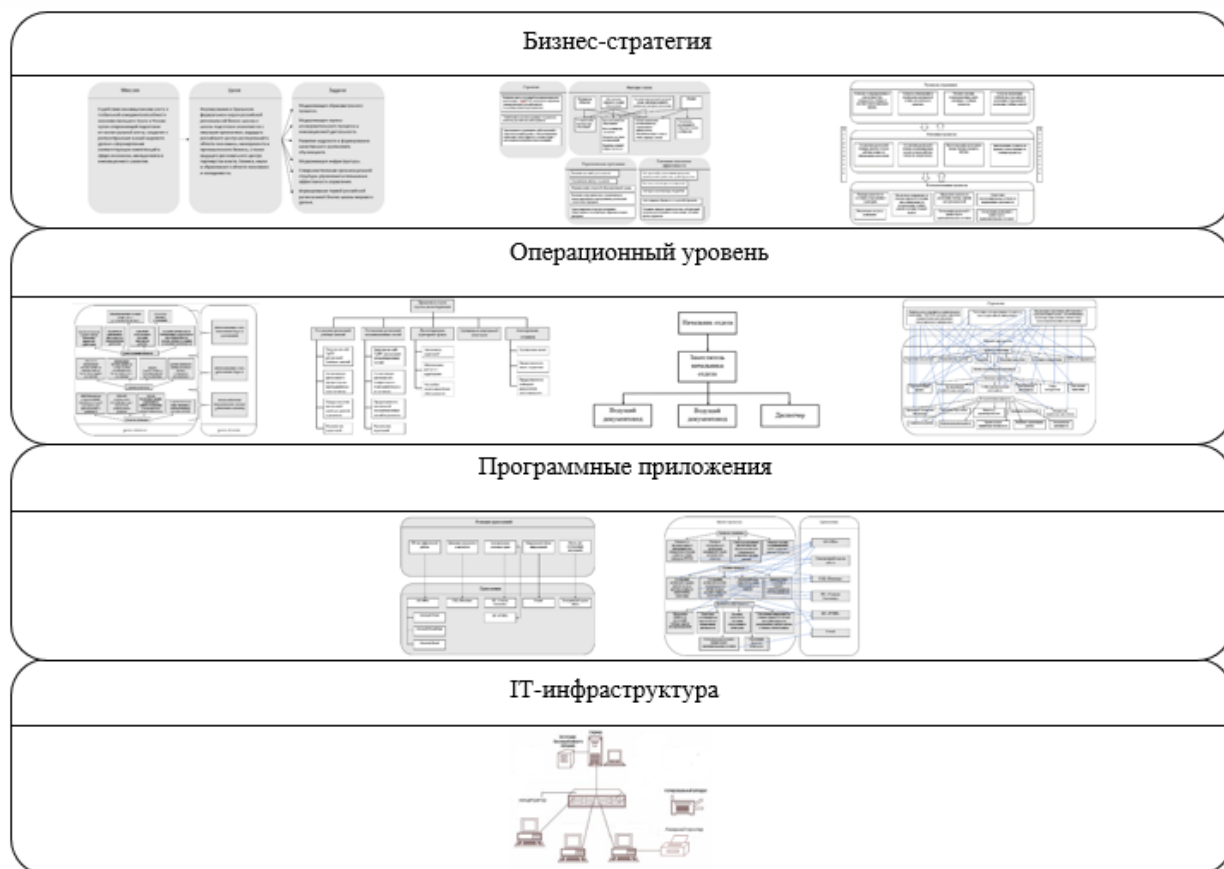


Рисунок 2.11 – Общая модель архитектуры службы диспетчеризации²⁷

3.2 СИСТЕМА СБОРА И ОБРАБОТКИ ДАННЫХ ОПРОСА СТУДЕНТОВ «AS-IS»

Современные тенденции развития экономики и сферы высшего образования в России заставляют вузы искать и внедрять новые методы и технологии, обеспечивающие конкурентные преимущества на рынке образовательных услуг. Одним из действенных рычагов, позволяющих вселить потенциальным потребителям уверенность в качестве предоставляемых вузом образовательных услуг, является проектирование и внедрение эффективных систем управления качеством образования. Для эффективного функционирования системы менеджмента качества в вузе необходимо постоянное совершенствование деятельности организации, которое невозможно без периодического анализа фактического состояния работы по

²⁷ Составлено автором по: [1]

качеству и ее результатов. Такой всесторонний анализ в мировой практике получил название самооценки. Самооценка может дать общее представление о деятельности организации и степени развития системы менеджмента качества. Она может также помочь организации определить области, нуждающиеся в улучшении, и приоритеты развития. Поэтому данная процедура не только предваряет процесс сертификации системы менеджмента качества, но также должна осуществляться периодически на всех этапах деятельности вуза. Одним из эффективных инструментов осуществления самооценки вуза может стать опрос студентов, поскольку последние являются не только основными «потребителями» образовательной деятельности вуза, но и активными участниками этого процесса. Выявление показателей, которые, по мнению студентов, определяют качество образовательных услуг, предоставляемых данным учебным заведением, позволяет осуществить ранжирование многочисленных факторов по степени их влияния на качество образовательного процесса, определить наиболее болезненные точки и направить основные усилия и средства вуза (материальные, технические, методические, информационные и др.) на совершенствование своей деятельности.

Одной из задач службы диспетчеризации ВШЭМ УрФУ являются анкетирование студентов по вопросам качества учебного процесса и последующая обработка ответов в MS Excel с целью предоставления консолидированных отчётов кафедрам ВШЭМ.

На данный момент документовед обрабатывает все собранные анкетные данные, не фильтруя их. Процесс занимает достаточно большое количество времени. На входе – анкетные данные, на выходе – консолидированный отчёт. Участники процесса – документовед и программное обеспечение MS Excel. Управление представляется в виде рабочих планов групп, содержащих фамилию преподавателя, предмет и организационную форму обучения (рисунок 2.12).

При обработке ответов выполняются 3 основных бизнес-процесса (рисунок 2.13):

- 1) сбор данных;
- 2) обработка данных;
- 3) создание консолидированного отчёта.

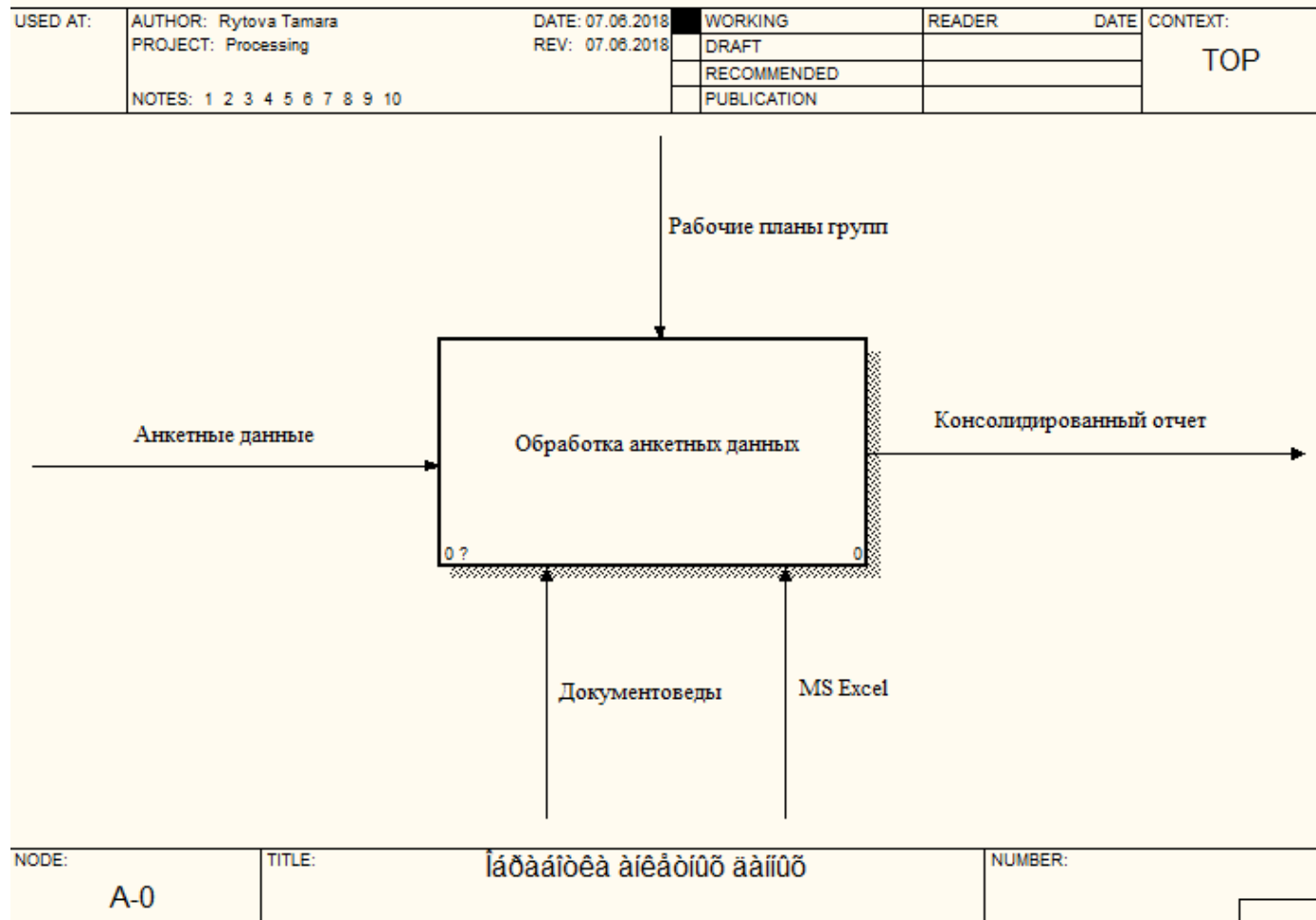


Рисунок 2.12 - Контекстная диаграмма «Обработка анкетных данных»²⁸

²⁸ Составлено автором

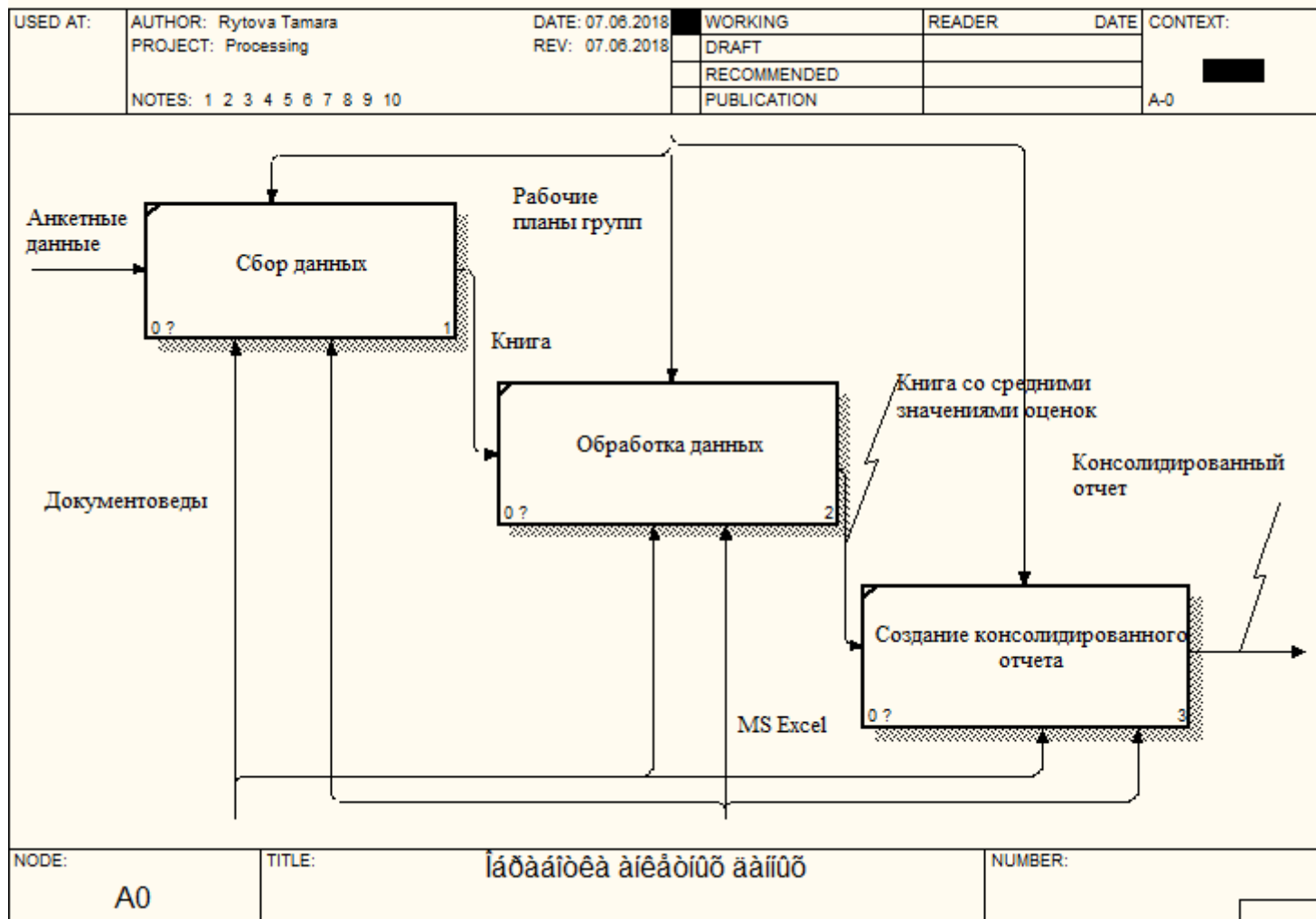


Рисунок 2.13 - Рисунок 13 – Диаграмма декомпозиции «Обработка анкетных данных»²⁹

²⁹ Составлено автором

3.3 СИСТЕМА СБОРА И ОБРАБОТКИ ДАННЫХ ОПРОСА СТУДЕНТОВ «ТО-ВЕ»

По окончании пройденных курсов студент заполняет анкету о качестве обучения, оставляет свои комментарии. Зачастую качество результатов, полученных в ходе анкетирования, делает нецелесообразным дальнейшую их обработку. Большинство комментариев «бесполезны», их трудно понять и расшифровать. Многие респонденты для сокращения времени заполнения анкеты указывают одинаковые ответы, чередуют их, либо проставляют оценки по порядку, например, на вопрос 9 отвечают – 1, на вопрос 10 – 2, на вопрос 11 – 3 и т. д. Разумеется, такие заполненные анкеты не отражают действительного мнения респондента, бесполезны и не должны учитываться при обработке данных.

На данный момент документовед обрабатывает анкетные данные «вручную». И, скорее всего, анкеты с неверными ответами будут учтены в анализе. Даже если сотрудники установят правила исключать такие анкеты, то всегда имеет место быть человеческий фактор, и диспетчер включит такие данные опроса к обработке.

Можно сделать вывод о том, что для получения релевантных результатов нужно исключать заведомо неверные ответы на опросы путём применения методов интеллектуальной обработки данных.

На рисунке 2.14 представлена диаграмма декомпозиции «Обработка анкетных данных» с использованием машинного обучения.

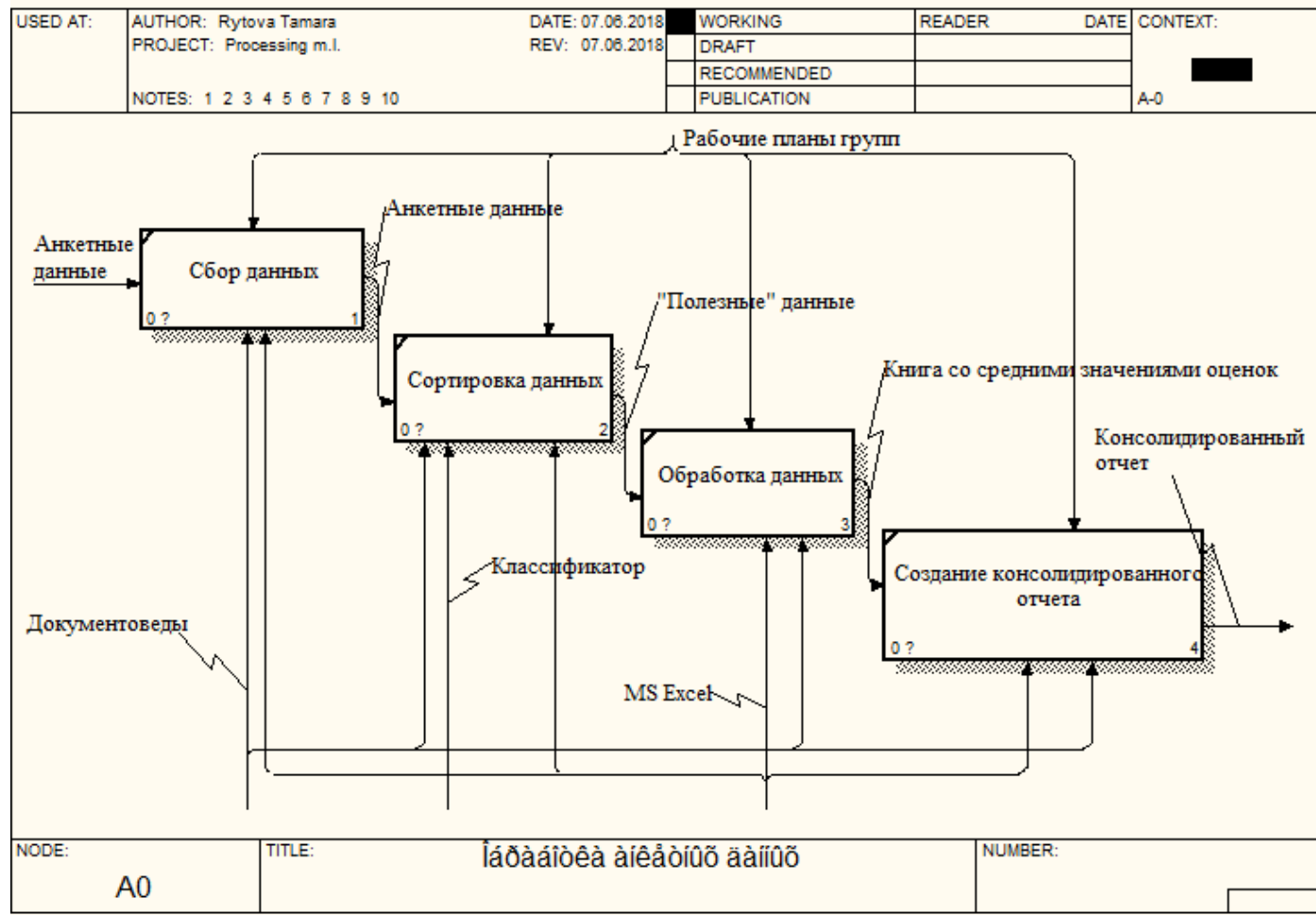


Рисунок 2.14 - Диаграмма декомпозиции «Обработка анкетных данных»³⁰

³⁰ Составлено автором

Как видно из диаграммы, количество бизнес-процессов увеличилось, добавился процесс «Сортировка данных». Выполнение данного процесса предполагает дальнейшее сокращение трудозатрат, так как уменьшится количество анкет, подлежащих обработке. Анкеты будут проходить через классификатор и принимать значения 0 либо 1, то есть учитывать ту или иную анкету в обработке либо нет.

3.4 ЭКОНОМИЧЕСКО ОБОСНОВАНИЕ

Характеристика выпускной работы

В экономической части представлено экономическое обоснование магистерской диссертации «Применение искусственного интеллекта при обработке анкетных данных».

Целью данной части выпускной работы является экономическое обоснование внедрения системы искусственного интеллекта, доказывающее экономическую целесообразность затрат на его внедрение.

В экономической части будут посчитаны затраты на заработную плату, налоги, материальные и нематериальные расходы на этапах реализации и эксплуатации проекта, рассчитаны чистый приведенный доход, внутренняя норма доходности. А также в заключении будут представлены эффект в денежном эквиваленте от внедрения проекта и суммарная стоимость проекта.

Описание статей расходов

Разобьем процесс магистерской диссертации на этапы по каждому виду работ и определим ожидаемую трудоёмкость выполнения каждого вида работ.

Исполнители выпускной работы:

- аналитик;
- python-разработчик.

Во-первых, рассчитаем стоимость часа работы специалистов, задействованных в проекте. Вспомогательные величины для расчетов представлены в таблице 1.

Таблица 1 – Вспомогательные величины³¹

Ставка НДС, % от ФОТ	13%
Страховые взносы, % от ФОТ, в том числе	30,2%
Пенсионное страхование	22,0%
Медицинское страхование	5,1%
Социальное страхование	2,9%
Взносы на травматизм	0,2%
Кол-во рабочих часов в месяце	168
Кол-во рабочих дней в месяце	21
Кол-во рабочих часов в день	8

С помощью формул (1), (2), (3), (4) находим значения для таблицы 2.

$$\text{НДС} = \text{ЗП в мес.} / (100\% - 13\%) * \text{ЗП в мес.}, \quad (1)$$

где НДС – налог на добавленную стоимость;

ЗП в мес. – заработная плата за месяц.

$$\text{Страховые взносы} = (\text{НДС} + \text{ЗП в мес.}) * 30,2\%, \quad (2)$$

где НДС – налог на добавленную стоимость;

ЗП в мес. – заработная плата за месяц.

$$\text{Затраты на труд 1} = \text{НДС} + \text{ЗП в мес.} + \text{Страховые Взносы}, \quad (3)$$

где Затраты на труд 1 – затраты на оплату труда в месяц;

НДС – налог на добавленную стоимость;

ЗП в мес. – заработная плата за месяц;

³¹ Составлено автором

Страховые взносы – страховые взносы.

$$\text{Затраты на труд 2} = \text{Затраты на труд 1} / 168, \quad (4)$$

где Затраты на труд – затраты на оплату труда в час;

Затраты на труд 1 – затраты на оплату труда в месяц.

В таблице 2 представлена стоимость часа работы специалистов.

Таблица 2 – Стоимость часа работы специалистов³²

№	Должность специалиста	Зарплата "на руки" в месяц, руб.	НДФЛ, руб.	Страховые взносы, руб.	Затраты на оплату труда, руб./мес.	Затраты на оплату труда, руб./ч
1	Аналитик	50 000	7 471,26	17 356,32	74 827,59	445
2	Python-разработчик	70 000	10 459,77	24 298,85	104 758,62	624

Следующим шагом рассчитаем трудозатраты и затраты на оплату труда.

Посчитаем показатели для двух этапов: этап реализации проекта (таблица 3) и этап эксплуатации внедренного решения (таблица 4).

Здесь разбивка процесса реализации проекта по этапам и задачам. Назначение на каждую задачу нужных специалистов и оценка продолжительности их работы в часах. Расчет затрат, исходя из ставки сотрудника и отработанных часов.

Используя формулу (5), рассчитаем затраты на оплату труда.

$$\text{Затраты на опл. Труда} = \text{Труд. Затрат.} * \text{Ставка}, \quad (5)$$

где Затраты на опл. Труда – затраты на оплату труда;

Труд.Затрат. – трудовые затраты специалиста;

³² Составлено автором

Ставка – стоимость часа работы специалиста.

Таблица 3 – Трудозатраты и затраты на оплату труда на этапе реализации проекта

№	Этап реализации и проекта	Задача/вид работы	Задействованный специалист	Трудозатраты специалиста, час.	Ставка, руб./час.	Затраты на опл. труда, руб.
1	Анализ	Исследование процесса "как есть"	Аналитик	6	445	2672
2	Анализ	Сбор данных	Аналитик	5	445	2227
3	Разработка	Подготовка данных для алгоритмов	Python-разработчик	3	624	1871
4	Разработка	Обучение классификаторов	Python-разработчик	3	624	1871
					ИТОГО:	8 641

Здесь оценка затрат на оплату труда ежемесячно в режиме эксплуатации внедрения. Учитывается количество часов работы в месяц.

Таблица 4 – Трудозатраты и затраты на оплату труда на этапе эксплуатации³³

№	Этап проекта	Задача/вид работы	Задействованный специалист	Трудозатраты специалиста, час.	Ставка, руб./час.	Затраты на опл. труда, руб.
1	Анализ	Оценка эффективности и алгоритмов	Аналитик	2	445	890
2	Эксплуатация	Регулярное обучение алгоритма	Python-разработчик	3	624	1 872
					ИТОГО:	1 872

Сводные данные по трудозатратам и затратам на оплату труда на этапе реализации проекта представлены в таблице 5.

Таблица 5 – Сводная таблица по трудозатратам и затратам на оплату труда из таблицы 3³⁴

³³ Составлено автором

Названия строк	Сумма по полю Трудозатраты специалиста, час.	Сумма по полю Затраты на опл. труда, руб.
Анализ	12	4899,42
Аналитик	12	4899,42
Разработка	6	3741,38
Python- разработчик	6	3741,38
Общий итог	18	8640,8

Сводные данные по трудозатратам и затратам на оплату труда на этапе эксплуатации проекта представлены в таблице 6.

Таблица 6 – Сводная таблица по трудозатратам и затратам на оплату труда из таблицы 4³⁵

Названия строк	Сумма по полю Трудозатраты специалиста, час.	Сумма по полю Затраты на опл. труда, руб.
Анализ	2	890
Аналитик	2	890
Эксплуатация	3	1872
Python- разработчик	3	1872
Общий итог	5	2762

Далее оцениваем материальные и нематериальные вложения на этапе реализации (таблица 7).

Таблица 7 – Материальные и нематериальные вложения на этапе реализации проекта³⁶

№	Категории и статьи вложений	Кол-во	Цена, руб.	Стоимость, руб.
I	Материальные вложения			0
1	Ноутбук	2	0	0
2	Офисная мебель			0
II	Нематериальные вложения			0

³⁴ Составлено автором

³⁵ Составлено автором

³⁶ Составлено атором

Продолжение таблицы 7

1	Дистрибутив Anaconda - свободно распространяемое ПО (opensource)	1	0	0
			ИТОГО:	0

Ноутбуки и офисная мебель уже имеются в собственности, поэтому расходы равны нулю.

Следующим действием оцениваем материальные и нематериальные вложения, а также прочие текущие вложения на этапе эксплуатации (таблица 8). Общая стоимость каждого вложения вычисляется по формуле (6).

$$C = N * P, \quad (6)$$

где С – стоимость;

N – количество;

P – цена.

Таблица 8 – Материальные и нематериальные вложения на этапе эксплуатации внедренного решения³⁷

№	Категории и статьи вложений	Кол-во	Цена, руб.	Стоимость, руб.
I	Материальные вложения			0
1	ПК	3	0	0
2	Офисная мебель		0	0
1	Нематериальные вложения			2000
A	Услуги по обучению персонала			2000
1	Обучение сотрудников			
2	Лицензия MS Office	1	0	0
			Итого:	2000

Персональные компьютеры, офисная мебель, лицензия MS Office – всё это имеется в отделе диспетчеризации.

Далее представлены расходы, несвязанные непосредственно с реализацией проекта.

³⁷ Составлено автором

Электроэнергия:

Тариф на электроэнергию - 5 руб. за 1 кВт.ч.

Потребление электричества ноутбуком за месяц: $900 / 1000 * 8 * 21 = 151$ кВт.

Расходы на электроэнергию за месяц: $5 * 151 = 755$ руб.

Амортизация ноутбука: $5000/12=417$ руб.

Стоимость расходов на интернет: 600 руб.

Сумма расходов составляет 1 772 руб.

В таблице 9 указан состав накладных расходов на этапе реализации проекта.

Таблица 9 – Состав накладных расходов на этапе реализации проекта³⁸

№	Статьи накладных расходов	Содержание статей накладных расходов
1	Электроэнергия	Стоимость услуги по тарифу за 1 месяц
2	Амортизация ноутбука	Расчет амортизации при условии, что остаточная стоимость ноутбука 5000 руб., остаточный срок службы - 1 год.
3	Интернет	Стоимость услуги по тарифу за 1 месяц
	Метод расчета накладных расходов	[% от трудозатрат в денежных единицах]
[A]	Сумма трудозатрат в денежных единицах (руб.)	8 641
[B]	Принятая доля (%) накладных расходов от [A]	20%
[C]	Оценка накладных расходов в денежных единицах (руб.)	1728

Электроэнергия:

Тариф на электроэнергию - 5 руб. за 1 кВт.ч.

³⁸ Составлено автором

Потребление электричества ноутбуком за месяц: $750 / 1000 * 6 * 5 = 23$ кВт

Расходы на электроэнергию за месяц: $5 * 23 = 115$ руб.

Амортизация ноутбука: $15000 / 60 = 250$ руб., $250 * 3 = 750$ руб.

Сумма расходов составляет 865 руб.

В таблице 10 указан состав накладных расходов на этапе эксплуатации.

Таблица 10 – Состав накладных расходов на этапе эксплуатации³⁹

№	Статьи накладных расходов	Содержание статей накладных расходов
1	Электронергия	Потребление электроэнергии тремя ПК
2	Амортизация ПК	Расчет амортизации при условии, что остаточная стоимость ПК 15000 руб., остаточный срок службы - 5 лет.
	Метод расчета накладных расходов	[% от трудозатрат в денежных единицах]
[A]	Сумма трудозатрат в денежных единицах (руб.)	2762
[B]	Принятая доля (%) накладных расходов от [A]	25%
[C]	Оценка накладных расходов в денежных единицах (руб.)	691

Расчет общих экономических показателей

NPV – это чистая приведенная (к сегодняшнему дню) стоимость. Метод оценки инвестиционных проектов, основанный на методологии дисконтирования денежных потоков [57]. Данный показатель вычисляется по формуле (7):

$$NPV = -I_0 + \sum_{t=1}^n \frac{CF_t}{(1+r)^t}, \quad (7)$$

где I_0 – первоначальные инвестиции;

³⁹ Составлено автором

n – кол-во периодов;

CF_t – денежный поток в t -й период;

r – ставка дисконтирования.

В таблице 11 представлен расчет чистой дисконтированной стоимости проекта, как мы видим проект окупается в период между 2-ым и 3-им годами. Столь длительный срок окупаемости обусловлен тем, что анкетирование проводится лишь раз в год в течение недели.

Для целей расчета NPV мной была использована ставка дисконтирования 10,25% или ставка рефинансирования ЦБ РФ в совокупности с годовым уровнем инфляции.

Таблица 11 – Чистая дисконтированная стоимость проекта⁴⁰

	Investments							
Период	0	1	2	3	4	5	6	7
Чистый приведенный доход (NPV)	-10 369	-8 432	-6 199	-3 698	-957	2 003	5 159	8 492

Выручку оцениваю как: сокращенная часть трудозатрат на обработку результатов анкетирования (в часах) * ставку в час.

Итого: учитывается то, что ежегодно затрачиваемое время на обработку анкет приблизительно составляет: 5дней * 6часов * 3сотрудника = 90 часов, и сэкономленное нами время согласно экспертной оценке сотрудников отдела диспетчеризации составляет до 80% от общих трудозатрат – эффект от автоматизации отбора анкетных данных: 90 * 50% = 45 ч.

Если учитывать заработную плату сотрудника отдела диспетчеризации равной 30 тыс. руб. в месяц, то в денежном эквиваленте эффект будет: = 90000 / (21*8) * 72 = 24107 руб.

⁴⁰ Составлено автором

С учетом общих затрат по заработной плате, материальных и нематериальных затрат, налогов стоимость внедрения проекта представлена в таблице 12.

Таблица 12 – Отток денежных средств⁴¹

Отток ДС	10 369	21 971	23 804	25 819	28 036	30 474	33 157	36 108
Себестоимость производства продукта/выполнения заказа	0	16 875	18 563	20 419	22 461	24 707	27 177	29 895
Инвестиционные и текущие вложения	10 369	4 563	4 563	4 563	4 563	4 563	4 563	4 563
Налоги (все учтено без НДС, налог на прибыль 20%)	0	534	679	838	1 013	1 205	1 417	1 650

IRR или внутренняя норма доходности – это ставка процента, при которой приведенная стоимость всех денежных потоков инвестиционного проекта (т.е. NPV) равна нулю [54]. Данную величину можно рассчитать по формуле (8):

$$NPV = \sum_{t=0}^n \frac{CF_t}{(1+r)^t} = 0, \quad (8)$$

где NPV – чистая приведенная стоимость;

CF – денежные потоки;

r – % ставка;

t – количество периодов.

Рассчитываем показатель IRR в Excel с помощью встроенной функции ВСД. При расчете используем значения, представленные в таблице 13.

⁴¹ Составлено автором

Таблица 13 – Денежные потоки⁴²

Период	0	1	2	3	4	5	6	7
Денежный поток	-10 369	2 136	2 714	3 351	4 051	4 821	5 668	6 600

Внутренняя норма доходности (IRR) составляет 28%, что говорит о целесообразности проекта.

Разность показателей IRR и ставки дисконтирования равна 17,75%, так как ставка дисконтирования равна 10,25%. Можно сделать вывод, что у организации есть резерв для страхового риска и премии за риск, это значит, что организация может корректировать эти величины до суммарного показателя IRR, не рискуя оказаться в убытке или увеличить срок окупаемости внедрения проекта. В нашем случае величина внутренней нормы доходности равна 2,7.

Индекс доходности равен 1,3, что больше единицы, следовательно, проект окупается.

Таким образом, данный проект по обработке анкетных данных с использованием машинного обучения имеет место быть и действительно позволит организации сэкономить денежные средства, которые организация может направить на увеличение заработной платы, автоматизацию процессов.

Оценка масштабируемости, выводы

Чтобы оценить масштабируемость, следует взять во внимание те технологии, которые используются в проекте. В данном проекте используются технологии машинного обучения.

С каждым днем популярность и применимость машинного обучения возрастает. Как использовать искусственный интеллект далее в этом проекте?

С каждым годом анкет для обработки становится больше, следовательно, растет и количество условий сортировки. В дальнейшем можно продолжить обучать представленные классификаторы на вновь появившейся базе ответов.

⁴² Составлено автором

Таким образом, можно сделать вывод, что проект является масштабируемым, и подлежит дальнейшему развитию.

3.5 ВЫВОДЫ ПО РАЗДЕЛУ

В третьей главе была проанализирована деятельность отдела диспетчеризации ВШЭМ УрФУ, рассмотрены бизнес-процессы обработки данных и проиллюстрирована модель процессов после внедрения проекта. Учитывая это, были разработаны мероприятия по повышению показателей эффективности. Были рассчитаны такие показатели как чистая приведенная стоимость проекта и внутренняя норма доходности. На основании внедрения предложенных мероприятий эффективность в денежном эквиваленте составит 24107 рублей, а экономия времени – 45 часов.

ЗАКЛЮЧЕНИЕ

Каждая современная организация оперирует большим объемом данных, который нужно систематически обрабатывать. Для этой цели предприятия, как правило, используют информационные технологии, позволяющие автоматизировать некоторые рутинные, постоянно повторяющиеся операции.

Преимуществами автоматизации работы является следующее:

- существенное сокращение числа ошибок в работе;
- снижение затрат времени и сил на выполнение привычных операций;
- существенное увеличение эффективности работы.

В ходе работы были выполнены следующие задачи:

- рассмотрены теоретические основы систем сбора и обработки данных;
- изучены системы искусственного интеллекта и программное обеспечение систем ИИ;
- проанализирована деятельность службы диспетчеризации ВШЭМ УрФУ и предоставлено решение для более эффективной работы;
- автоматизирован процесс отбора анкетных данных в дистрибутиве Python Anaconda с использованием алгоритмов машинного обучения;
- определена эффективность автоматизации отбора анкетных данных для службы диспетчеризации ВШЭМ УрФУ.

В результате проделанной работы были написаны и обучены классификаторы, которые позволяют сортировать отвеченные анкеты на «полезные» и «бесполезные» для более эффективной обработки анкетных данных. На основе проведенных исследований в качестве метода для автоматизации процесса отбора анкетных данных рекомендованы к использованию 3 классификатора. После внедрения алгоритма существенно сократятся трудозатраты, уменьшится количество допускаемых ошибок при вычислении.

В качестве исходных данных были использованы ответы анкетирования студентов и на основе данной базы определены экспертом признаки бесполезности.

Хочется отметить, что проект подлежит масштабированию, так как ежегодно база ответов пополняется, появляются новые признаки, следовательно, классификаторы можно обучать на новых выборках.

Также для имплементации выбранного классификатора в дальнейшем может быть разработан web-сервис, с которым будет взаимодействовать модуль системы сбора и обработки и учета анкетных данных посредством передачи JSON данных.

По итогам проделанной работы можно сделать вывод, что разработка проекта проведена успешно.

Обученный алгоритм будет использован при обработке анкетных данных документоведом службы диспетчеризации ВШЭМ.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. Программа развития Института «Высшая школа экономики и менеджмента Уральского федерального университета им. первого Президента России Б.Н. Ельцина» (ВШЭМ УрФУ) в 2011–2016 гг. и на период до 2020 г. [Текст] – Екатеринбург: Изд-во УрФУ, 2015.
2. Базы знаний интеллектуальных систем: Учебное пособие для студ. вузов, обуч. по направлениям "Прикладная математика и информатика", "Информатика и вычислительная техника" / Под ред. Т.А. Гавриловой, В.Ф. Хорошевского. СПб.: Питер, 2018 . С.382.
3. Введение в искусственный интеллект: Учебное пособие для студ. высш. учеб. заведений / Под ред. Л.Н. Ясницкого. М.: Издательский центр «Академия», 2015. С. 176.
4. Интеллектуальные информационные системы: Учебник для вузов / Под ред. А.В. Андрейчикова. М.: Финансы и статистика, 2014. С. 424.
5. Интеллектуальные информационные системы: Учебник для вузов / Под ред. Д. В. Гаскарова. М.: Высшая школа, 2015. С. 431.
6. Интеллектуальный анализ данных: Учебное пособие / Под ред. Г.Ю. Чернышевой, Л.В. Аброськиной. Саратов: СГСЭУ, 2014.
7. Компьютерный анализ данных: Учебное пособие / Под ред. О.Г. Берестневой, Е.А. Муратовой, А.М. Уразаевой. Томск: Изд-во ТПУ, 2013. С. 204
8. Маркетинговые исследования: Учебно-методическое пособие / Под ред. В.А. Бороденя. М.: БГЭУ, 2018. С. 156.
9. Математические методы психологического исследования: Анализ и интерпретация данных: Учебное пособие / Под ред. А.Д. Наследова. СПб.: Речь, 2014. С. 392.
10. Обработка и анализ социологических данных с помощью пакета SPSS: Учебно-методическое пособие / Под ред. Е.В. Дорогонько. Сургут: Издательский центр СурГУ, 2010. С. 67.
11. Основы маркетинга: Учебное пособие / Под ред. Л.В. Архиповой, П.В. Средник. М.: МФИ, 2015. С.72.

12. Основы статистического анализа. Практикум по статистическим методам и исследованию операций с использованием пакетов SPSS и EXCEL: Учебное пособие / Под ред. Э.А. Вуколова. М.: ФОРУМ: ИНФРА-М, 2014. С. 464.
13. Практикум по статистике с пакетом SPSS: Учебное пособие по курсу «Математическая статистика» / Под ред. Ю.А. Горицкого. Томск: Изд-во ТПУ, 2014. С. 116.
14. Системный анализ и синтез стратегических решений в инноватике. Математические, эвристические и интеллектуальные методы системного анализа и синтеза инноваций: Учебное пособие / Под ред. А.В. Андрейчикова, О.Н. Андрейчиковой. М.: Ленанд, 2015. С. 306.
15. Системы искусственного интеллекта: модели и технологии, основанные на знаниях: Учебник / Под ред. Л.С. Болотовой. М.: ФиС, 2014. С. 664.
16. Системы искусственного интеллекта. Практический курс: Учебное пособие / Под ред. И. Астаховой. М.: Бином. Лаборатория знаний, 2017. С. 292.
17. Системы искусственного интеллекта: Учебное пособие / Под ред. И.Г. Сидоркиной. М.: КноРус, 2015. С. 249.
18. SPSS для социологов: Учебное пособие / Под ред. В.В. Пациорковского, В.В. Пациорковской. М.: ИСЭПН РАН, 2015. С. 403.
19. Алексеева А.Ю., Ечевская О.Г., Ковалева Г.Д., Ростовцев П.С. Анализ социологических данных с применением пакета SPSS. Новосибирск: Редакционно-издательский центр НГУ, 2003. С. 105
20. Аляутдинов М. А., Галушкин А. И., Казанцев П. А., Остапенко Г. П. Нейрокомпьютеры. От программной к аппаратной реализации. М.: Горячая линия «Телеком», 2016. С. 152
21. Ахим Бююль, Петр Цёфель. SPSS: Искусство обработки информации. Анализ статистических данных и восстановление скрытых закономерностей: Пер. с нем. Спб.: «ДиаСофтЮП», 2015. С. 608.
22. Барсегян А.А. Методы и модели анализа данных: OLAP и Data Mining. СПб.: БХВ-Петербург, 2014. С. 336.

23. Беллман Р., Дрейфус С. Прикладные задачи динамического программирования. М.: Наука, 2016. С. 458.
24. Боровиков В.П. SPSS. Искусство анализа данных на компьютере: для профессионалов. СПб.: Питер, 2013. С. 656.
25. Веприцкая Д.А., Саламахина Ю.В. Применение систем искусственного интеллекта при решении экономических задач // Проблемы современной экономики (Новосибирск). 2015. № 27.
26. Гаврилова А.Н. Системы искусственного интеллекта. М.: КноРус, 2013. С. 248.
27. Девятков В. В. Системы искусственного интеллекта. М.: Изд-во МГТУ им. Н. Э. Баумана, 2013. С. 352.
28. Джарратано, Д. Экспертные системы: принципы разработки и программирование. 4-е изд. М.: ООО «И.Д. Вильямс», 2017. С. 1152.
29. Джордж Ф. Люгер. Искусственный интеллект: стратегии и методы решения. 4-е изд. М.: Издательский дом «Вильямс», 2014. С. 864.
30. Здор С.Е. Кодированная информация. От первых природных кодов до искусственного интеллекта. М.: Наука, 2014. С. 168.
31. Илющенко В.В., Балканов В.А., Сеньков И.В. Анализ инструментальных средств разработки экспертных систем // Повышение эффективности производства и использования электроэнергии в условиях Сибири: материалы Всероссийской научнопрактической конференции с международным участием. 2015. №1.
32. Кабанов С.В. Использование пакета SPSS для статистической обработки опытных данных: методические указания. Саратов: Изд-во Саратов. гос. агр. ун-та, 2015. С. 90.
33. Кошелев Д.А. Искусственный интеллект в информационных технологиях // Символ науки. 2016. № 1-2 (13).
34. Ладов В.А. Язык в системе искусственного интеллекта: синтаксис и семантика // Гуманитарная информатика. 2014. № 1 (1).
35. Лорьер Ж.-Л. Системы искусственного интеллекта. М.: Мир, 2016. С. 568.
36. Любарский Ю.Я. Интеллектуальные информационные системы. М.: Наука, 2015. С. 227.

37. Малышев В.В., Сливкин С.С., Рукавишников В.С., Базаркин Е.В. Применение методов машинного обучения для построения рекомендательной системы отбора анкет абитуриентов // Научный вестник НГТУ. 2017. №2.
38. Мелихов А.Н., Мелихова З.А., Мелихова О.А. Использование нечеткой математики при проектировании системы искусственного интеллекта // Известия ТРТУ. 2018. № 3 (9).
39. Мелихова О.А. Методы построения интеллектуальных систем на основе нечеткой логики. Таганрог: Изд-во ТРТУ, 2017. С. 92.
40. Наследов А.Д. Профессиональный статистический анализ данных СПб: Питер, 2013. С. 416.
41. Наследов А.Д. SPSS: Компьютерный анализ данных в психологии и социальных науках. 2-е изд. СПб.: Питер, 2006. С. 318.
42. Нильсон Н. Искусственный интеллект. М.: Мир, 2013. С. 373.
43. Ноткин Л.И. Искусственный интеллект и проблемы обучения. М.: КомКнига, 2015. С. 11.
44. Рассел Стюарт, Питер Норвиг. Искусственный интеллект: современный подход (AIMA)». 2-е изд. М.: Издательский дом «Вильямс», 2016. С. 1424.
45. Ричард В., Коэльо Л.П. Построение систем машинного обучения на языке Python. М.: ДМК Пресс, 2016. С. 97.
46. Романов П.С., Суворин Е.И. Искусственный интеллект: применение в технических системах. СПб.: МАА, 2016. С. 118.
47. Ручкин В.Н. Универсальный искусственный интеллект и экспертные системы. СПб.: БХВ-Петербург, 2016. С. 240.
48. Рытова Т.А. Как получить более эффективную обратную связь от студентов. 2017.
49. Рытова Т.А. Роль математики в экономике. 2017. № 7. С. 781-785.
50. Сукина С.А. Искусственный интеллект // Новая наука: Теоретический и практический взгляд. 2015. № 5-2.
51. Тарасов В.Б. Интеллектуальные системы в проектировании // Новости искусственного интеллекта. 2016. № 4

52. Теннант-Смит Дж. Бейсик для статистиков: Пер. с англ. М.: Мир, 2014. С. 208.
53. Ханмагомедов А.Х., Хасцаев Б.Д. Применение элементов искусственного интеллекта в системах сбора и передачи информации. Научно-технические ведомости Санкт-Петербургского государственного политехнического университета. Информатика. Телекоммуникации. Управление. 2017. № 76.
54. Шалютин С.М. Искусственный интеллект. М.: Мысль, 2015. С. 27.
55. Алгоритмы функционирования мультиагентной системы определения пространственно-временных состояний объекта. URL: <https://cyberleninka.ru> (дата обращения: 15.05.2018).
56. Внутренняя норма доходности. Формула расчета IRR инвестиционного проекта. URL: <http://msfo-dipifr.ru> (дата обращения: 27.05.2018).
57. Исследование применения методов машинного обучения для повышения надежности распознавания объектов в изображениях. URL: <http://elib.spbstu.ru> (дата обращения: 15.05.2018).
58. Логистическая регрессия. URL: <http://www.machinelearning.ru> (дата обращения: 15.05.2018).
59. Расчет NPV. URL: <http://forpm.ru> (дата обращения: 27.05.2018).
60. Электронный учебник по статистике. М.: StatSoft, Inc., 2012. URL: <http://www.statsoft.ru/home/textbook/default.html> (дата обращения: 11.04.2018).
61. Random forest. URL: <http://ru.cybernetics.wikia.com> (дата обращения: 15.05.2018).
62. SPSS – программа для статистической обработки данных. URL: <https://vuzlit.ru> (дата обращения: 03.04.2018).