

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

Институт математики, физики и информационных технологий
(наименование института полностью)

Кафедра Прикладная математика и информатика
(наименование)

09.04.03 Прикладная информатика
(код и наименование направления подготовки)

Управление корпоративными информационными процессами
(направленность (профиль))

ВЫПУСКНАЯ КВАЛИФИКАЦИОННАЯ РАБОТА (МАГИСТЕРСКАЯ ДИССЕРТАЦИЯ)

на тему «Модели и алгоритмы автоматизированной системы отбора студентов для
кадрового резерва вуза»

Обучающийся

И.Р. Хорошев

(Инициалы Фамилия)

(личная подпись)

Научный
руководитель

д.т.н., доцент, С.В. Мкртычев

(ученая степень (при наличии), ученое звание (при наличии), Инициалы Фамилия)

Тольятти 2023

Оглавление

Введение.....	3
Глава 1 Анализ современного состояния исследований по отбору персонала для кадрового резерва вуза.....	7
1.1 Анализ методов отбора персонала для кадрового резерва вуза.....	7
1.2 Анализ современных подходов к построению автоматизированных систем отбора персонала.....	18
Глава 2 Анализ методологий построения эффективных автоматизированных систем отбора персонала для кадрового резерва вуза.....	24
2.1 Методология объектно-ориентированного моделирования систем поддержки принятия решения.....	24
2.2 Методология построения систем поддержки принятия решения на основе интеллектуального анализа данных.....	27
Глава 3 Разработка моделей и алгоритмов автоматизированной системы отбора студентов для кадрового резерва вуза.....	31
3.1 Моделирование поведения студента на основе анализа его цифрового следа.....	31
3.2 Алгоритмы кластеризации данных цифрового следа студента.....	37
3.3 Функциональная модель автоматизированной системы отбора студентов для кадрового резерва вуза.....	45
Глава 4 Апробация проектных решений и оценка их эффективности.....	51
4.1 Апробация проектных решений.....	51
4.2 Оценка экономической эффективности проектных решений.....	62
Заключение.....	65
Список используемой литературы и используемых источников.....	68

Введение

«Одной из ключевых задач кадровой политики современного российского вуза является формирование кадрового резерва.

Кадровый резерв – это группа сотрудников (специалистов, руководителей), которые потенциально способны к руководящей деятельности, отвечают требованиям, предъявляемым должностью, прошли отбор и квалификационную подготовку, но еще не назначены на должность.

Создание кадрового резерва является инструментом эффективной управленческой политики предприятия» [5].

В настоящее время перед вузами поставлена задача создания внутреннего кадрового резерва в том числе из обучающихся в вузе студентов, используя эффективную методику отбора последних.

Создание кадрового резерва относится к категории задач отбора персонала предприятия или компании и играет важную роль в управлении вузом.

Неправильные решения при выборе студента для кадрового резерва вуза могут привести к значительным экономическим потерям, а, главное, могут негативно повлиять на будущую карьеру молодого специалиста, что соответствующим образом отразится на имидже вуза.

Для решения данной проблемы на предприятиях используются автоматизированные системы отбора персонала.

Совершенно очевидно, что в основу такой системы должны быть положены модели и алгоритмы, обеспечивающие эффективность управления процессом отбора студентов.

Таким образом, актуальность темы исследования обусловлена необходимостью разработки моделей и алгоритмов автоматизированной системы, обеспечивающей эффективность управления отбором студентов для кадрового резерва вуза.

Объектом настоящего исследования является автоматизированная система отбора студентов для кадрового резерва вуза.

Предметом исследования являются модели и алгоритмы автоматизированной системы отбора студентов для кадрового резерва вуза.

Целью работы является исследование и разработка моделей и алгоритмов эффективной автоматизированной системы отбора студентов для кадрового резерва вуза.

Для достижения поставленной цели необходимо решать следующие задачи:

- провести анализ современного состояния исследований в области проектирования систем отбора персонала для кадрового резерва вуза;
- провести анализ методологических подходов к построению систем отбора персонала для кадрового резерва вуза;
- разработать модели и алгоритмы автоматизированной системы отбора студентов для кадрового резерва вуза;
- выполнить апробацию предлагаемых проектных решений и оценить их эффективность.

Гипотеза исследования: использование разработанных в рамках диссертационного исследования моделей и алгоритмов для реализации автоматизированной системы отбора студентов для кадрового резерва вуза, обеспечит повышение эффективности данной системы.

Методы исследования. В процессе исследования будут использованы следующие положения и методы: системный анализ, методы отбора персонала, методы и технологии проектирования автоматизированных систем управления, интеллектуальный анализ данных.

Новизна исследования заключается в разработке моделей и алгоритмов, которые обеспечат повышение эффективности автоматизированной системы отбора студентов для кадрового резерва вуза.

Практическая значимость исследования заключается в возможности применения предлагаемых моделей и алгоритмов при проектировании

эффективной автоматизированной системы отбора студентов для кадрового резерва вуза.

Теоретической основой диссертационного исследования являются научные труды российских и зарубежных ученых, занимающихся проблемами управления персоналом учебных организаций.

Основные этапы исследования: исследование проводилось с 2020 по 2023 год в несколько этапов.

На первом (констатирующем) этапе формулировалась тема исследования, выполнялся сбор информации по теме исследования из различных источников, проводилась формулировка гипотезы, определялись постановка цели, задач, предмета исследования, объекта исследования и выполнялось определение проблематики данного исследования.

Второй этап – поисковый. В ходе проведения данного этапа осуществлялся анализ методов отбора персонала, разработаны модели и алгоритмы эффективной автоматизированной системы отбора студентов (АСОС) для кадрового резерва вуза, опубликованы статьи по теме исследования в научных сборниках.

На третьем этапе осуществлялась апробация предлагаемых проектных решений, произведена оценка их эффективности, сформулированы выводы о полученных результатах по проведенному исследованию.

На защиту выносятся:

- модели и алгоритмы эффективной автоматизированной системы отбора студентов для кадрового резерва вуза;
- результаты апробации и оценки эффективности предлагаемых проектных решений.

По теме исследования опубликовано 2 статьи:

Хорошев И.Р. Система поддержки принятия решения для отбора студентов в кадровый резерв вуза // Вестник научных конференций. 2022. N 9-1 (85). С. 115-117.

Шипилова А.М., Мкртычев С.В., Хорошев И.Р. Система отбора

студентов для кадрового резерва вуза // Тезисы докладов Международной научно-технической конференции «Материалы, оборудование и ресурсосберегающие технологии», 20–21 апреля 2023 г. в Белорусско-Российском университете.

Диссертация состоит из введения, четырех глав, заключения и списка литературы.

Во введении обоснована актуальность темы исследования, представлены объект, предмет, цели, задачи и положения, выносимые на защиту диссертации.

В первой главе дан анализ современного состояния исследований в области проектирования систем отбора персонала для кадрового резерва вуза.

Во второй главе дан анализ методологических подходов к построению автоматизированных систем отбора студентов для кадрового резерва вуза.

Третья глава посвящена разработке моделей и алгоритмов эффективной автоматизированной системы отбора студентов для кадрового резерва вуза.

В четвертой главе выполнены апробация предлагаемых проектных решений и оценка их эффективности.

В заключении приводятся результаты исследования.

Работа изложена на 72 страницах и включает 24 рисунка, 9 таблиц, 43 источника.

Глава 1 Анализ современного состояния исследований по отбору персонала для кадрового резерва вуза

1.1 Анализ методов отбора персонала для кадрового резерва вуза

Отбор персонала – это технология выбора кандидата, соответствующего требованиям к должности, заработной платы, условиям труда, профессиональному и карьерному росту [4].

Проблематике отбора персонала посвятили свои работы Н.Ш. Никитина, Н.Н. Скитер, Y. Filiberto, F. Färber и другие.

Как показал анализ, наиболее используемой для отбора кандидатов в кадровый резерв организации является модель компетенций.

Модель компетенций – это полный набор компетенций и индикаторов поведения, необходимых для успешного выполнения сотрудником его трудовых функций [9].

К основным методам отбора персонала по модели компетенций относятся [11]:

- сбор данных о кандидатах;
- тестирование;
- рекомендации с прошлых мест работы;
- традиционное интервью (собеседование);
- интервью по компетенциям;
- ассесмент-центр;
- brainteaser-интервью;
- стресс-интервью;
- настольные игры;
- графология.

В исследовании [13] представлена математическая модель, позволяющая получить на основе результатов психологического тестирования объективное мнение относительно того, в какой мере личностные качества

кандидата соответствуют требованиям вакантной должности.

В процессе формирования математической модели принятия решения о выборе кандидата на вакансию были предприняты шаги (этапы), представленные на рисунке 1.



Рисунок 1 – Этапы построения математической модели принятия решения о выборе кандидата на вакансию

В исследовании [17] рассматриваются основные проблемы, приводящие к негативным последствиям на предприятии, которые связаны с самим процессом отбора кадров (рисунок 2).

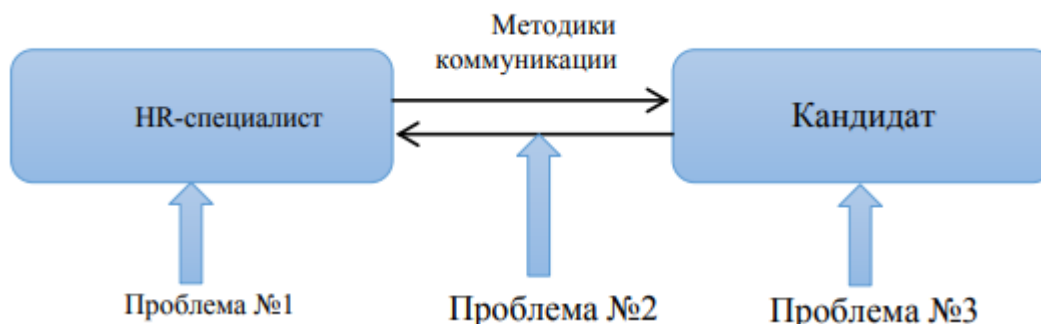


Рисунок 2 – Проблемы отбора и подбора кадров

«Первая проблема связана с вопросом, кто должен принимать участие в процессе отбора кадров, только HR-специалист или руководители отделов, в которые будет осуществляться прием, а возможно и сам руководитель?

Вторая проблема связана с выбором методик отбора персонала и формированием критериев отбора.

Третьей проблемой является психическое состояние кандидата на вакантное место, которое очень часто не позволяет человеку раскрыть весь свой потенциал, свои способности и уровень развития» [17].

В исследовании [28] предлагается подход к проблеме агрегирования рейтингов для получения общего рейтинга. Это также называется агрегированным рейтингом в задаче подбора персонала.

Подход основан на измерении расстояния между индивидуальным рейтингом и общим рейтингом и ищет решение, которое минимизирует расхождение между исходными рейтингами и результирующим агрегированием.

Этот метод использует подход обучения с подкреплением для построения агрегирования и его производительности, и по сравнению с другими подходами показывает многообещающие результаты.

В исследовании [27] описывается метод сопоставления с использованием вероятностного автоматизированного рекомендательного подхода в системе электронного найма и представляются первые весьма многообещающие результаты применения алгоритма к синтетическим данным.

На рисунке 3 показана гибридная модель оценки кандидата.

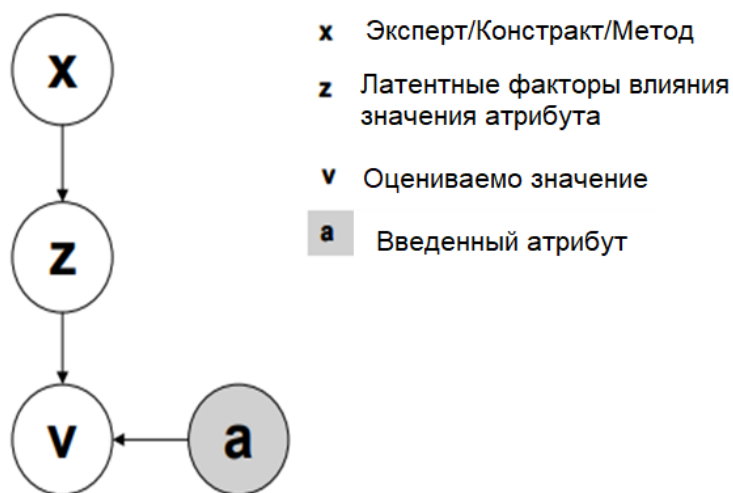


Рисунок 3 – Гибридная модель оценки кандидата

На рисунке 4 показаны три основные модели проекта оценивания характеристик и компетенций студентов [30].

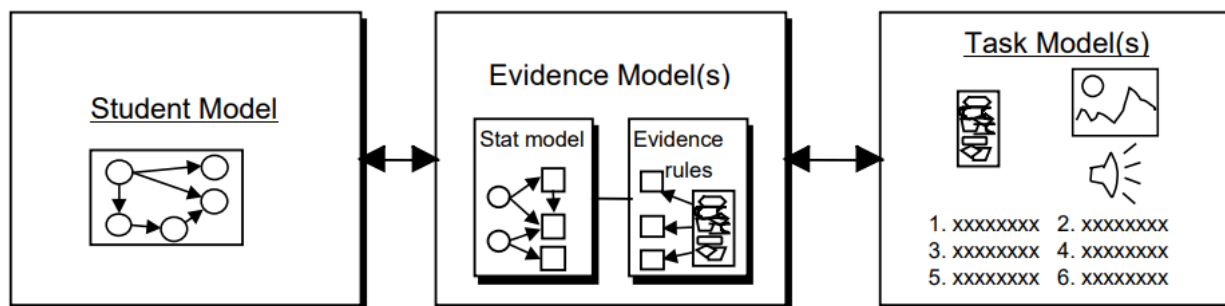


Рисунок 4 – Модели проекта оценивания характеристик и компетенций студентов

Рассмотрим свойства представленных моделей.

Модель студента (Student Model) содержит переменные, представляющие аспекты профессионального мастерства, которые являются объектами вывода при оценке, и именно здесь мы управляем нашими неопределенными знаниями об этих переменных.

Таким образом, переменные модели студента касаются характеристик студентов.

Переменные модели студента — это термины, в которых мы хотим обсуждать студентов. Это уровень, на котором строится история студента, чтобы определять оценки, принимать решения или планировать обучение.

Модель студента на рисунке 1 изображает переменные модели студента в виде кругов. Стрелки, соединяющие их, представляют собой важные эмпирические или теоретические ассоциации. Эти переменные и ассоциации неявно присутствуют в неформальных применениях рассуждений при оценивании, таких как дискуссия между студентом и преподавателем.

В более формальных приложениях используется вероятностная модель для управления знаниями о значениях данных переменных конкретного студента в любой заданный момент времени. В частности, модель студента принимает форму фрагмента байесовской логико-вероятностной сети или байесовской сети доверия [19].

Рассмотрим примеры модели студента.

На рисунке 5 представлена модель показана модель студента, которая лежит в основе наиболее известных оценок: одна переменная, обычно обозначаемая как θ , которая представляет уровень владения определенной областью задач.

Небольшая таблица в квадрате перед этой переменной студенческой модели представляет собой распределение вероятностей, которое выражает текущее убеждение.

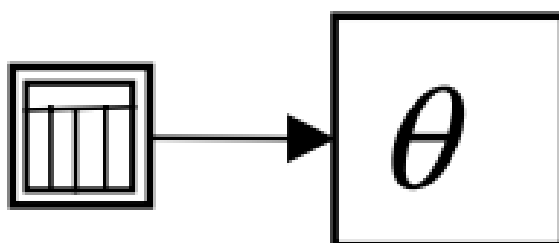


Рисунок 5 – Модель студента для оценки GRE-CAT

В качестве данных использована компьютерно-адаптивная версия (CAT) экзамена Graduate Record Examination (GRE), которые включают домены заданий для вербального, количественного и аналитического мышления [21].

Формально модель можно описать следующим образом:

$$M_{GRE} = (V, Q, A), \quad (1)$$

где V, Q и A - навыки вербального, количественного и аналитического мышления, соответственно.

Модель доказательства (Evidence Model) описывает, как извлечь ключевые элементы доказательства (значения наблюдаемых переменных) из того, что студент говорит или делает в контексте задания (результат работы), и моделирует взаимосвязь этих наблюдаемых переменных с переменными модели студента.

Наблюдаемые переменные касаются характеристик производительности.

На рисунке 4 показаны две части модели доказательства: оценочная субмодель и статистическая субмодель.

Оценочная подмодель извлекает характерные черты «продукта работы», то есть всего, что студент говорит, делает или создает в ситуации задания.

Статистическая подмодель обновляет модель студента в соответствии со значениями этих характеристик, эффективно синтезируя доказательную ценность результатов выполнения заданий.

На рисунке 6 показана статическая субмодель модели доказательств в GRE-CAT.

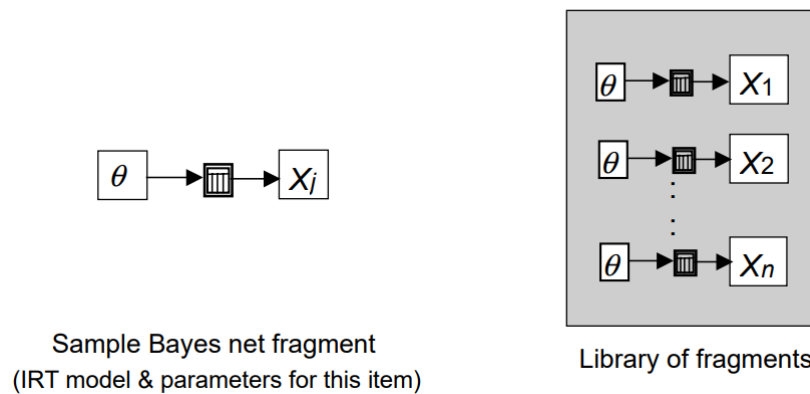


Рисунок 6 – Статистическая субмодель модели доказательств в GRE-CAT

В левой части показан фрагмент байесовской сети логического вывода для обновления распределения вероятностей параметра квалификации студента с учетом ответа на конкретный вопрос j .

Распределение между переменной θ модели студента и ответом на задание X_j представляет собой условное распределение вероятностей X_j при заданном θ .

В правой части рисунка находится библиотека всех элементов, которые могут быть предоставлены, а также структуры, необходимые для стыковки любого из них с моделью ученика, чтобы включить доказательства, которые вносит его ответ.

Информация, хранящаяся вместе с этими фрагментами, также предоставляет данные, необходимые для выбора следующего элемента для управления, чтобы оптимизировать информацию о θ испытуемого.

Сборка статистических моделей по мере необходимости из набора готовых фрагментов является примером того, что в литературе по экспертным системам называется «построением модели на основе знаний».

Модель задач (Task model) описывает функции задачи, которые необходимо указать при создании теста и задания.

Присвоение определенных значений переменным модели задачи и предоставление материалов, соответствующих заданным спецификациям,

приводит к конкретной задаче.

Таким образом, задание описывает конкретные обстоятельства, предназначенные для предоставления испытуемому возможности действовать таким образом, чтобы получить информацию о том, что он знает или может сделать в более общем плане.

Само задание не описывает, на что мы должны обратить внимание в конечном результате или как мы должны оценить то, что мы видим. Это указано в модели доказательств.

К одному и тому же рабочему продукту данной задачи могут быть применены различные, возможно, совершенно разные правила доказательства. Отдельные и, возможно, совершенно разные студенческие модели, подходящие для разных целей или полученные из разных концепций мастерства, могут быть основаны на данных одной и той же задачи [37].

Модель задач в GRE описывает класс тестовых заданий. Существует некоторое соответствие между моделями задач и «типами элементов» GRE (например, завершение предложения, понимание отрывка, количественное сравнение).

Для разных типов заданий обычно требуются разные модели задач, потому что для описания различных типов стимулирующих материалов и форматов представления необходимы разные наборы переменных, а разные функции могут быть важны при моделировании параметров заданий или управлении выбором заданий.

Для использования в SAT одного и того же элемента с точки зрения содержания потребуются разные модели задач, поскольку спецификации для представления элемента и управления им в этих двух режимах полностью различаются.

На рисунке 7 показан пример модели задач.

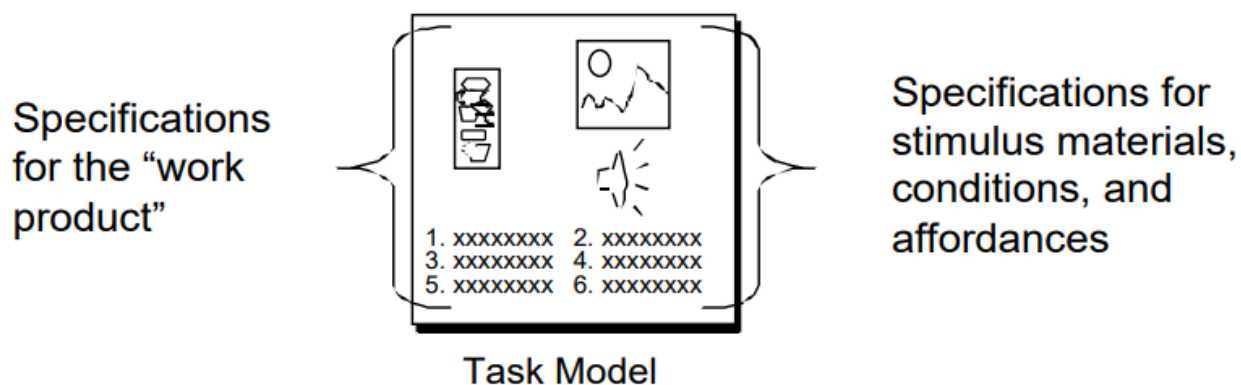


Рисунок 7 – Элементы модели задач в GRE-CAT

На входе модели – спецификации на «рабочий продукт».

На выходе модели – спецификации материального стимулирования, условий и возможностей.

Представляет интерес успешный опыт формирования кадрового резерва ТПУ, включающий студентов, аспирантов, и молодых сотрудников, перспективных для научно-педагогической работы.

«Отмечается, что претенденты на включение в кадровый резерв вуза должны иметь потенциал для успешной реализации управленческих задач в условиях изменяющейся среды. Соблюдение при отборе данных требований позволяет гарантировать высокий уровень профессионализма сотрудников, включенных в кадровый резерв, что формирует положительную репутацию и привлекательный имидж университета в научной среде. Оценка кандидатов в кадровый резерв базируется на моделях компетенций ТПУ. В вузе разработаны модели управленческих, корпоративных, а также модель функциональных (базовых) компетенций научно-педагогического работника и карьерная карта (рисунок 8)» [6].

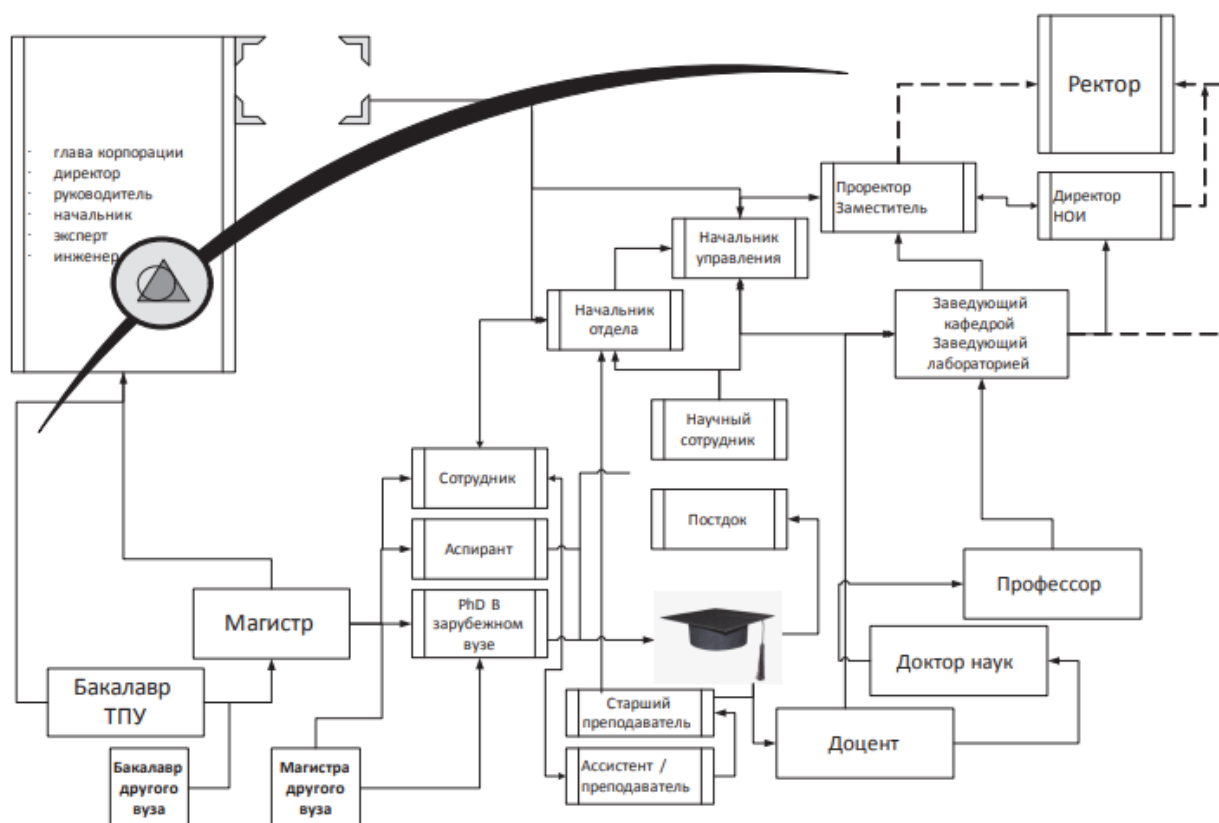


Рисунок 8 – Карьерная карта ТПУ

Альтернативными методами отбора кандидатов в кадровый резерв являются методы, основанные на анализе поведенческой модели кандидатов.

Модель поведения – это целый ценностно-нормативный комплекс характерных черт человека, которые являются образцами эмоций, действий, точек зрения, поступков и принципиальных установок индивидуума [16].

В [25] описан метод отбора кандидатов Success Insights DISC, основанный на модели оценки DISC.

Модель оценки DISC базируется на основных поведенческих факторах личности, которые классифицированы следующим образом:

- Dominance (D) — доминирование;
- Influence (I) — влияние;
- Steadiness (S) — стабильность;
- Compliance (C) — исполнительность, добросовестность.

«Различные комбинации данных факторов позволяют с высокой долей достоверности определить и предсказать основные поведенческие характеристики личности.

Система DISC представляет собой инструмент объективной оценки поведенческих характеристик и профессиональных компетенций сотрудников.

Технологии, основанные на модели DISC, позволяют существенно повысить производительность каждого сотрудника и персонала в целом и делают процесс управления понятным на каждом этапе. Они могут также применяться для оценки поведения студентов.

Оценки DISC тщательно исследованы и проверены временем. Издатель оценок DISC, Wiley, является одним из старейших и наиболее уважаемых в мире издателей научных и технических справочников.

Модель оценки DISC используется сегодня в крупнейших мировых компаниях для отбора, оценки, обучения и развития персонала» [34].

Студенты должны освоить навыки, чтобы лучше общаться и взаимодействовать друг с другом.

Для них важно добиться успеха в вузе, на работе и за ее пределами [36].

Оценки DISC могут помочь им добиться успеха. DISC может помочь студентам стать лучшими коммуникаторами, улучшая взаимодействие со всеми стилями DISC.

Следует отметить, что в методе DISC используются результаты тестирования сотрудника или кандидата на должность в компании. Для студентов вуза могут использоваться результаты их тестирования на платформах дистанционного образования.

Анализ когнитивных задач DISC позволит выявить «характеристики производительности», которые характеризуют модели поведения и различают уровни знаний.

Для анализа представленных моделей используются методы, которые относятся к области Educational Data mining – интеллектуальный анализ

образовательных данных [26].

Вместе с тем, необходимо констатировать недостаточность исследований, посвященных методологическим подходам к отбору студентов для кадрового резерва вуза.

Для выбора метода отбора студентов для кадрового состава вуза используем таблицу сравнения (таблица 1).

Таблица 1 – Сравнение методов отбора студентов для кадрового состава вуза

Метод	Преимущества	Недостатки
На основе модели компетенций	Наличие четких критериев отбора (компетенций) и возможность точной оценки по этим критериям	Длительный и сложный процесс разработки модели компетенций и проведение на ее основе оценки персонала
На основе модели DISC	Широкие возможности для применения, точность и валидность методики	Рассматривает только поведенческие особенности, но не оценивает такие важные параметры как мотиваторы и ценности

С учетом вышеизложенного выбираем для отбора студентов для кадрового состава вуза метод на основе модели DISC.

1.2 Анализ современных подходов к построению автоматизированных систем отбора персонала

По своим функциональным и архитектурным особенностям автоматизированные системы отбора персонала для кадрового резерва относятся к категории систем поддержки принятия решения (СППР).

Рассмотрим работы, в которых представлены подходы и решения для построения автоматизированных систем отбора персонала.

В исследовании [20] предложена модель отбора персонала,

использующая принципы ассоциативности и мажоритарности принятия решений на основе сигнатурного метода.

«Применение принципа ассоциативности в модели позволяет при вводе в систему отбора любой характеристики кандидата, например, образования или возраста, выбирать из базы вакансий параллельно все вакансии, соответствующие (ассоциирующие) значению введенной характеристики образа-кандидата.

На рисунке 9 представлен алгоритм программы мандатно-верификационного подхода, обеспечивающий целостность данных и организацию надежного доступа в систему отбора персонала» [20].



Рисунок 9 – Алгоритм программы мандатно-верификационного подхода

Рассмотренный подход позволяет проводить параллельно проверку соответствия каждого образа-кандидата по всем имеющимся вакансиям.

Решение о соответствии образа-кандидата вакантной должности

принимается по мажоритарному правилу или по правилу большинства, что повышает достоверность принятия решения. В процессе отбора формируется ранжированный ряд соответствий образа-кандидата по всем вакансиям базы вакансий.

В исследовании [31] рассмотрены наиболее эффективные методы машинного обучения для поддержки набора сотрудников и оценки их положения.

С этой целью была проведена серия тестов на основе алгоритмов классификации в отношении сотрудников государственного сектора с целью предсказать наилучшую пригодность на рабочих местах и распределение сотрудников.

На основе результатов проведенных тестов была построено дерево классификации на основе алгоритма J48, помогающая системе поддержки принятия решений по набору и оценке сотрудников (рисунок 10).

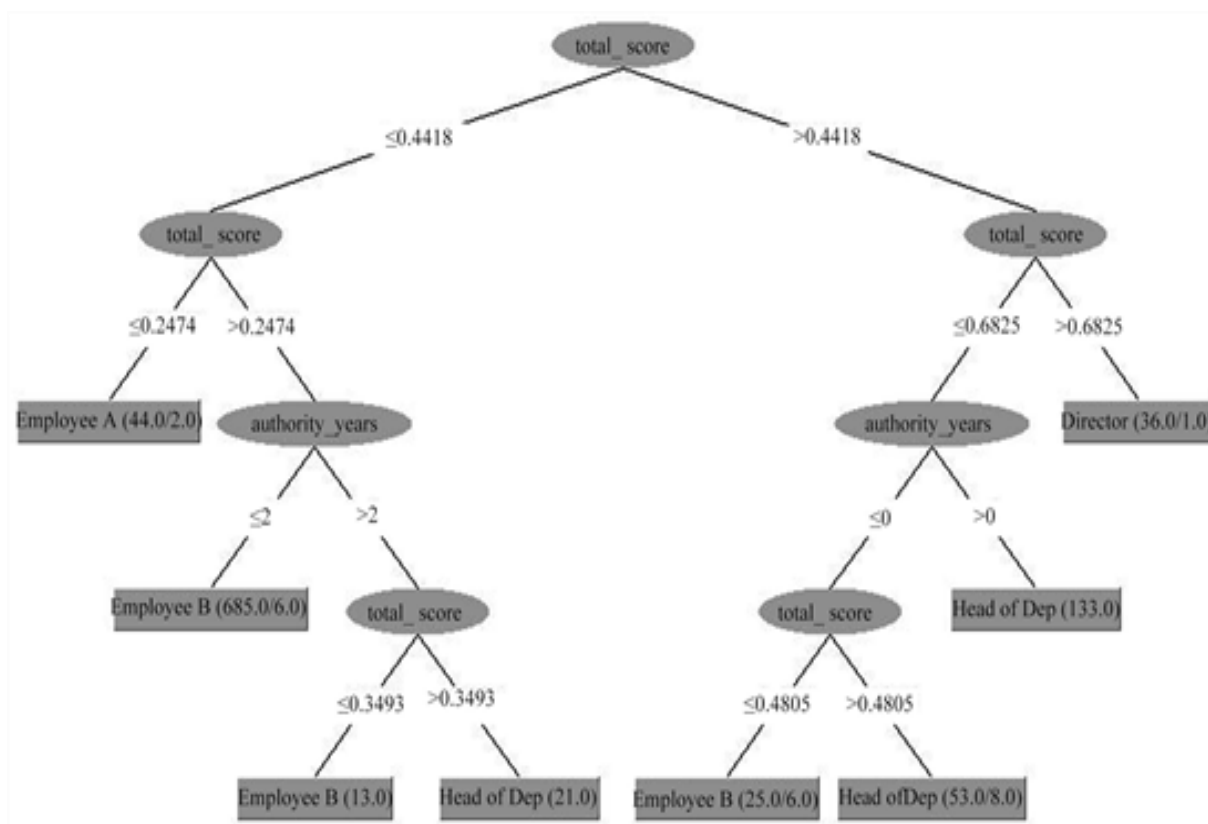


Рисунок 10 – Модель дерева, созданного с помощью алгоритма J48

В исследовании [23] представлена модель с акцентом на наилучшее возможное дерево или выбор деревьев, соответствующих банку данных.

В качестве метода отбора персонала предлагается использовать вступительный экзамен.

Статус повышения квалификации сотрудников был признан с использованием методов интеллектуального анализа данных, правил и взаимосвязи между оценками вступительных экзаменов и ситуацией сотрудников с производительностью труда.

Концептуальная модель открытия знаний из базы данных вступительных экзаменов представлена на рисунке 11.

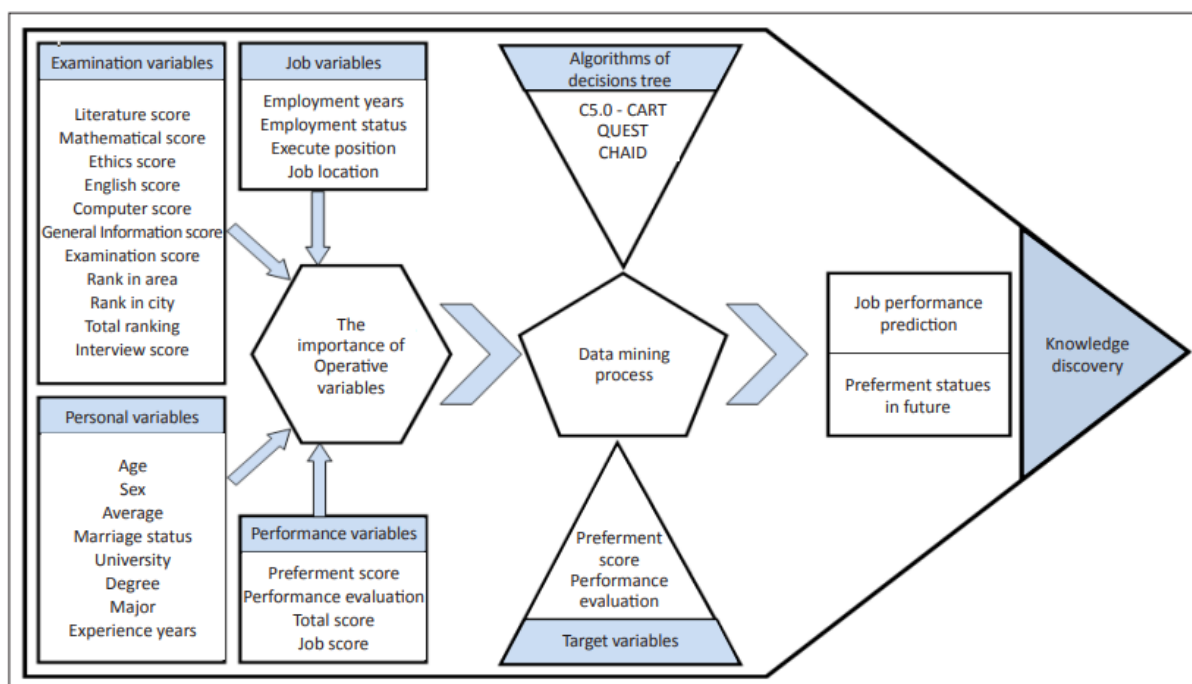


Рисунок 11 – Концептуальная модель открытия знаний из базы данных вступительных экзаменов

В исследовании [43] представлена модель, которая объединяет машинное обучение и методы принятия решения на основе мультиатрибутного подхода (Multiple-attribute decision-making, MADM), чтобы помочь менеджерам более объективно отбирать персонал и

способствовать повышению их компетенций.

На рисунке 12 представлен процесс моделирования интегрированной модели MADM, управляемой данными.

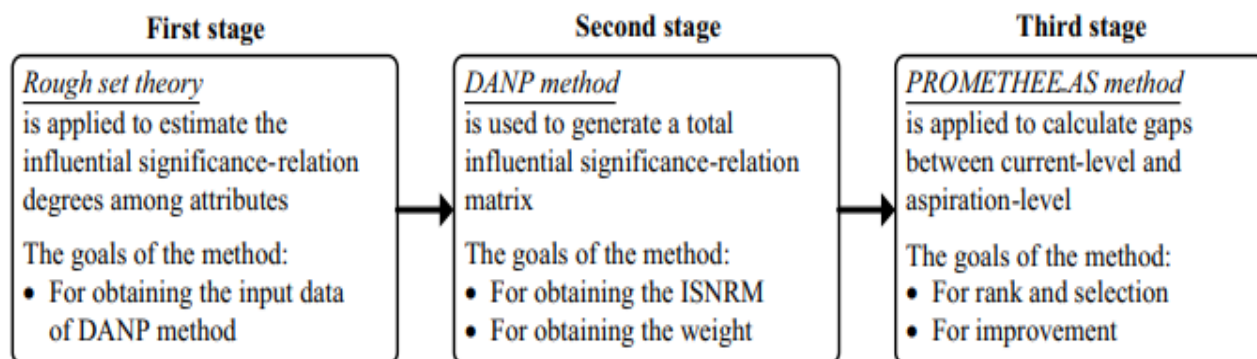


Рисунок 12 – Процесс моделирования интегрированной модели MADM, управляемой данными

Основанная на данных концепция предложенной модели MADM может быть применена для решения проблем принятия решений в различных областях, и модель также может быть применена для других компаний или в других отраслях.

Каждая компания должна устанавливать атрибуты компетенций на основе их собственной конкретной среды и характеристик должности; поэтому выводы, сделанные на основании исследований других компаний, могут быть другими.

Таким образом, анализ литературы и источников подтвердил интерес ученых и специалистов к проблеме отбора персонала для различных социальных систем.

Вместе с тем, необходимо констатировать отсутствие работ по проектированию автоматизированных систем отбора студентов для кадрового резерва вуза, что подтверждает актуальность темы исследования.

Выводы по главе 1

В результате проделанной работы были сделаны следующие выводы:

- для отбора кандидатов в кадровый резерв организации используются методы, основанные на модели компетенций и анализе поведенческой модели кандидатов;
- широкие возможности для применения, точность и валидность методики DISC обусловили ее выбор для отбора студентов для кадрового резерва вуза;
- по своим функциональным и архитектурным особенностям автоматизированные системы отбора персонала для кадрового резерва относятся к категории СППР;
- анализ литературы и источников подтвердил интерес ученых и специалистов к проблеме отбора персонала для различных социальных систем.

Вместе с тем, необходимо констатировать отсутствие работ по проектированию автоматизированных систем отбора студентов для кадрового резерва вуза, что подтверждает актуальность темы исследования.

Глава 2 Анализ методологий построения эффективных автоматизированных систем отбора персонала для кадрового резерва вуза

СППР должны объединять автоматизированную информацию и человеческий вердикт. Они должны поддерживать специальные стили принятия решений, а также процессы, как для нескольких, так и для отдельных лиц, принимающих решения. Эффективные СППР должны иметь возможность использоваться в качестве саморегулирующегося инструмента в дополнение к тому, чтобы помогать лицу, принимающему решение (ЛПР), принимать своевременные решения [22].

ЛПР должно быть в состоянии очень внимательно рассмотреть проблему.

Подходы к моделированию и разработке эффективных СППР привлекают большое внимание, поскольку технологические системы становятся все более сложными, чтобы отвечать постоянно растущим требованиям современных организационных потребностей и поддерживать ЛПР и менеджеров.

Рассмотрим наиболее востребованные методологии построения СППР на предмет применения для разработки моделей и алгоритмов АСОС для кадрового резерва вуза.

2.1 Методология объектно-ориентированного моделирования систем поддержки принятия решения

Методология объектно-ориентированного моделирования является основой для моделирования и проектирования СППР на основе унифицированного языка моделирования UML и моделирования дискретных событий.

UML имеет большое значение для разработки программного обеспечения. Как и каждый язык, он имеет стандартные обозначения и

синтаксис. Он состоит из тринадцати основных типов диаграмм, каждая из которых служит разным целям и описывает систему с разных точек зрения. В частности, можно определить два основных описания:

- структурное описание: в целом, система состоит из набора частей, часто называемых объектами. В среде UML структура системы описывается диаграммами, иллюстрирующими различные типы объектов, из которых состоит система, и их отношения;
- поведенческое описание: поведенческое представление очень важно, поскольку основная деятельность системы заключается в обеспечении соблюдения некоторых правил. Важно, чтобы эти правила были документированы в полной и правильной форме, чтобы избежать недопонимания как со стороны пользователя, так и со стороны разработчика.

Таким образом, моделирование СППР с использованием UML позволяет графически представить систему с различных аспектов, таких как поведение системы или ее структура.

Следует учесть, что СППР могут стать очень сложными. Поэтому их моделирование должно быть точным и структурированным [38].

Первым шагом для создания СППР является определение и идентификация проблемы. Это будет включать среду, в которой он будет применяться, параметры проблемы и ожидаемые результаты. Затем будет выбрана наиболее подходящая модель для конкретной задачи.

Далее мы опишем использование наиболее подходящих диаграмм UML для наших целей.

Диаграммы вариантов использования определяет акторов (действующих лиц системы), а также то, как система используется и взаимодействует с ними. Также показаны границы между СППР и «внешним миром», будь то другая система или пользователь СППР.

С помощью диаграмм классов определяются структура и элементы СППР.

Каждый элемент представляется как отдельный класс. Описываются атрибуты и операции каждого класса, а также типы атрибутов, входы и выходы операций. Кроме того, ограничения, касающиеся этих элементов, могут быть показаны явно.

Очень важным аспектом является использование интерфейсов. С их помощью можно четко описать взаимодействие между различными элементами СППР. С помощью интерфейсов можно показать взаимодействие СППР с остальной частью системы, что позволяет реализовать СППР в уже существующих системах, например, ERP и CRM.

Диаграммы деятельности и последовательности позволяют описать поведение СППР и всех процессов, составляющих систему.

Диаграммы деятельности показывают рабочий процесс действий, которые происходят в СППР, а при использовании дорожек показывают действующих лиц, ответственных за каждое действие.

Диаграммы последовательности обеспечивают динамическое описание системы, показывая сообщения, которые передаются от одного компонента к другому или к самим себе. Динамический аспект добавляется отображением течения времени по вертикальной оси.

Диаграммы последовательности позволяют описать поток информации с большой детализацией. В целом можно сказать, что диаграммы деятельности описывают систему с точки зрения «выбора», а диаграммы последовательности — с точки зрения «управления временем».

Важным аспектом описанного подхода является то, что он позволяет интегрировать СППР в уже существующие системы для поддержки принятия решений.

Таким образом, методология объектно-ориентированного моделирования может быть применена к конкретным проблемам, что позволяет описать и разработать СППР для сложных задач реального мира.

2.2 Методология построения систем поддержки принятия решения на основе интеллектуального анализа данных

Интеллектуальный анализ данных (Data mining) — это нетривиальный процесс получения от людей неизвестной, новой, эффективной, потенциально полезной и в конечном счете понятной информации и шаблонов знаний из большого количества неполных, зашумленных и нечетких случайных данных.

Его основными задачами являются анализ ассоциаций, кластерный анализ, классификация, прогнозирование, структура временных рядов и анализ отклонений. Другими словами, Data mining заключается в извлечении или «добыче» знаний из большого объема данных, но не все задачи по обнаружению информации являются интеллектуальным анализом данных.

Методы интеллектуального анализа данных применяются в различных областях управления организационными системами для улучшения процесса принятия решений [15].

На рисунке 13 представлена модель процесса использования СППР на основе интеллектуального анализа данных [33].

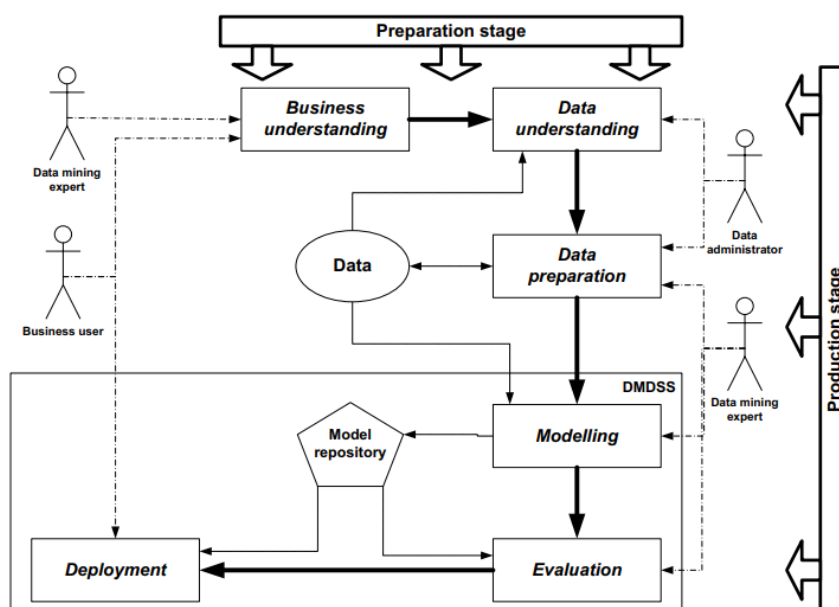


Рисунок 13 – Модель процесса использования СППР на основе интеллектуального анализа данных

Модель процесса использования СППР рассматривается как двухэтапная модель: этап подготовки и этап производства.

Разделение на два этапа основано на следующих двух требованиях, которые являются следствием используемого системного подхода к интеллектуальному анализу данных. Во-первых, СППР должна обеспечивать возможность многократного создания моделей интеллектуального анализа данных на основе актуального набора данных для каждой области анализа. Во-вторых, бизнес-пользователи должны использовать его только на этапе развертывания, обладая лишь базовым уровнем понимания концепции интеллектуального анализа данных.

Этап подготовки представляет собой модель процесса использования СППР в целях подготовки области анализа к производственному использованию.

Этап производства представляет собой производственное использование СППР для области анализа. На этапе производства акцент делается на стадиях моделирования, оценки и развертывания, что не означает, что другие этапы не включены в этап производства. Подготовка данных, например, выполняется автоматически на основе процедур, разработанных на этапе подготовки. Моделирование и оценка выполняются экспертом по интеллектуальному анализу данных, а этап развертывания выполняется бизнес-пользователями.

В последние годы, с быстрым развитием технологий интеллектуального анализа данных, все больше и больше предприятий начинают обращать внимание на завершение различных действий по принятию решений в управлении персоналом с помощью компьютерных информационных систем и интеллектуального анализа данных.

СППР на основе интеллектуального анализа данных используется с момента отбора сотрудников и проводит систематическое и всестороннее расследование, а также анализ предварительной подготовки сотрудников, навыков, заработной платы и других вопросов, которые могут помочь

компании лучше управлять сотрудниками. Кроме того, СППР позволяет ускорить переход управления персоналом к стандартизации, оцифровке и объединению в сеть, а также улучшить общий имидж компании.

Данные, генерируемые базовым управлением персоналом организации, такие как информация об организационной структуре предприятия, основная информация о персонале, информация об организации, информация о должности, информация об образовании и информация об изменениях персонала являются важной частью данных о человеческих ресурсах предприятия. Для получения этих данных необходима качественная обучающая выборка, обработка и стандартизированное управление.

Архитектура СППР для управления персоналом с использованием технологии Data mining, показана на рисунке 14 [29].

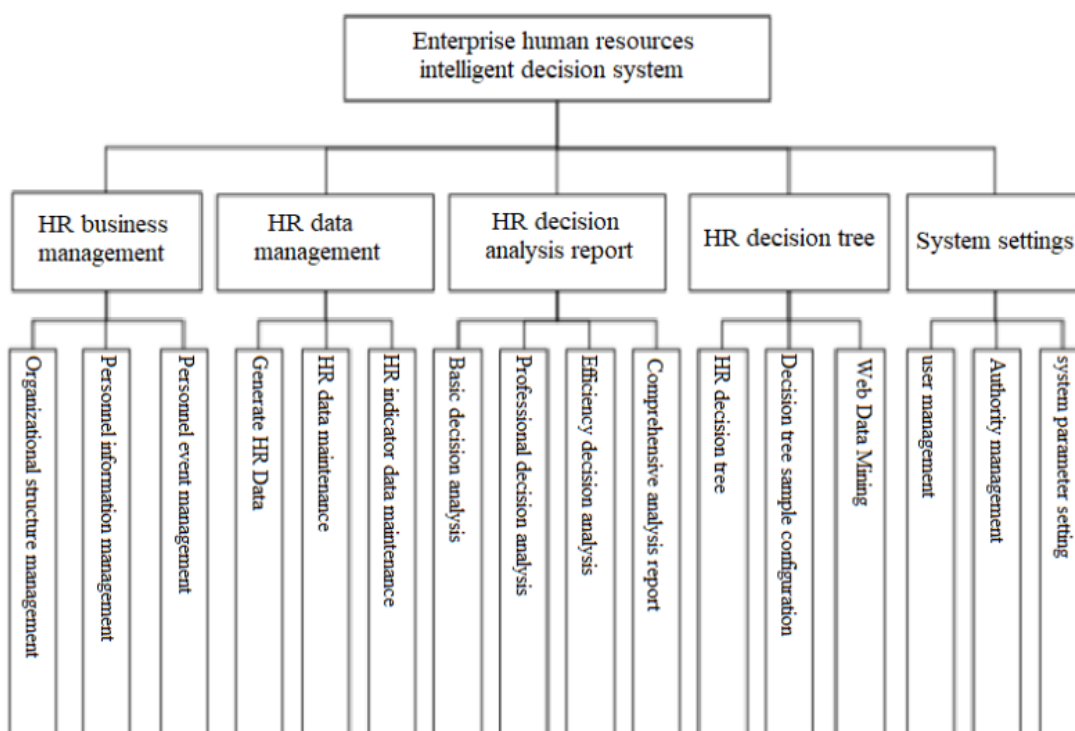


Рисунок 14 – Архитектура СППР для управления персоналом с использованием технологии Data mining

Для сравнения рассмотренных методологий используем таблицу 2.

Таблица 2 – Преимущества и недостатки методологий построения эффективных СППР для отбора персонала

Методология	Преимущества	Недостатки
Объектно-ориентированное моделирование	Предоставляет широкие возможности для моделирования различных аспектов СППР на UML	Слабая иерархия моделей и относительно высокие затраты на разработку
Интеллектуальный анализ данных	В СППР позволяет оптимизировать работу ЛПР снизить потери от возможных ошибок	Необходимость обеспечения высокого качества данных анализа

С учетом вышеизложенного для построения АСОС для кадрового резерва вуза выбираем комплексный подход, основанный на применении обеих рассмотренных методологий, каждая из которых будет использована для разработки моделей и алгоритмов системы.

Выводы по главе 2

Результаты проделанной работы позволили сделать следующие выводы:

- методология объектно-ориентированного моделирования предоставляет широкие возможности для моделирования различных аспектов СППР на UML. Недостатком данной методологии являются слабая иерархия моделей и относительно высокие затраты на разработку;
- как показывает практика, применение интеллектуального анализа данных в СППР позволяет оптимизировать работу ЛПР и снизить потери от возможных ошибок.

С учетом вышеизложенного для построения АСОС для кадрового резерва вуза выбираем комплексный подход, основанный на применении обеих рассмотренных методологий, каждая из которых будет использована для разработки моделей и алгоритмов системы.

Глава 3 Разработка моделей и алгоритмов автоматизированной системы отбора студентов для кадрового резерва вуза

3.1 Моделирование поведения студента на основе анализа его цифрового следа

Для построения СППР на основе Data mining необходимо выбрать источник данных для интеллектуального анализа.

Для АСОС для кадрового состава вуза в качестве такого источника выбран цифровой след студента.

Цифровой след студента – это запись его онлайн-активности в системах управления обучением и социальных сетях. Студенты оставляют цифровые следы в любое время, когда они находятся в сети, независимо от того, находятся ли они в вузе или за его пределами.

Существует две основные категории цифровых следов: активные следы и пассивные следы. Какой из них применяется, зависит от того, как информация получена [35].

Активные цифровые следы состоят из данных, которые остаются, когда пользователи Интернета делают преднамеренный выбор. Например, посты, сделанные в социальных сетях, являются типом активного следа.

Примеры активного цифрового следа студента: заполнение онлайн-форм (подписка на получение электронных писем или текстовых сообщений), публикации во всех формах социальных сетей (Facebook, Instagram, ВКонтакте и др.), работа в LMS и др.

Пассивные цифровые следы — это те, которые остаются без намерения пользователя Интернета, а иногда и без ведома об этом. Например, когда веб-сайт собирает информацию о том, сколько раз вы заходили на сайт в последнее время, это увеличивает пассивный цифровой след. Пассивные цифровые следы — это скрытый процесс, о котором пользователи Интернета могут вообще не знать.

Пример пассивного цифрового следа студента: образовательный портал университета.

Студенты оставляют цифровые следы в любое время, когда они находятся в сети, независимо от того, находятся ли они в вузе или за его пределами [18].

Большинство направлений в области образовательных исследований цифрового следа сосредоточены на поведении и успеваемости студентов [1].

Одним из таких направлений является моделирование поведения студента на основе анализа его цифрового следа.

Математически цифровой след студента FP можно представить в виде совокупности (2):

$$FP = (S, L, O) \quad (2)$$

где S – множество записей онлайн-активности студента в социальных сетях;

L – множество записей онлайн-активности студента в системах управления обучением;

O – множество записей онлайн-активности студента в других областях интернета-пространства.

Следует отметить, что данные о цифровом следе студента относятся к категории образовательных данных (Educational Data), которые в свою очередь относятся к категории больших данных (Big Data) [12].

Для получения качественных образовательных данных необходимо создать интегративную платформу больших данных образовательной деятельности вуза, которая включает функции сбора, хранения и глубокого анализа данных бизнес-системы вуза, включая данные цифровых следов студентов и др.

Архитектура такой платформы показана на рисунке 15.



Рисунок 15 – Архитектура платформы для сбора и анализа образовательных данных вуза

Данная платформа является основой для построения АСОС в кадровый резерв вуза.

Применение данного метода к модели (1) позволит проанализировать поведение студентов путем выделения групп студентов со сходными характеристиками по каждой совокупности индикаторов поведения.

Данный подход позволяет вырабатывать рекомендации, основанные на модели предшествующего поведения пользователей с учетом поведения других пользователей со сходными характеристиками.

Как было отмечено выше, анализ поведения студентов — это приложение интеллектуального анализа данных к университетской среде [42].

Общая структурная схема анализа поведения студента представлена на рисунке 16.

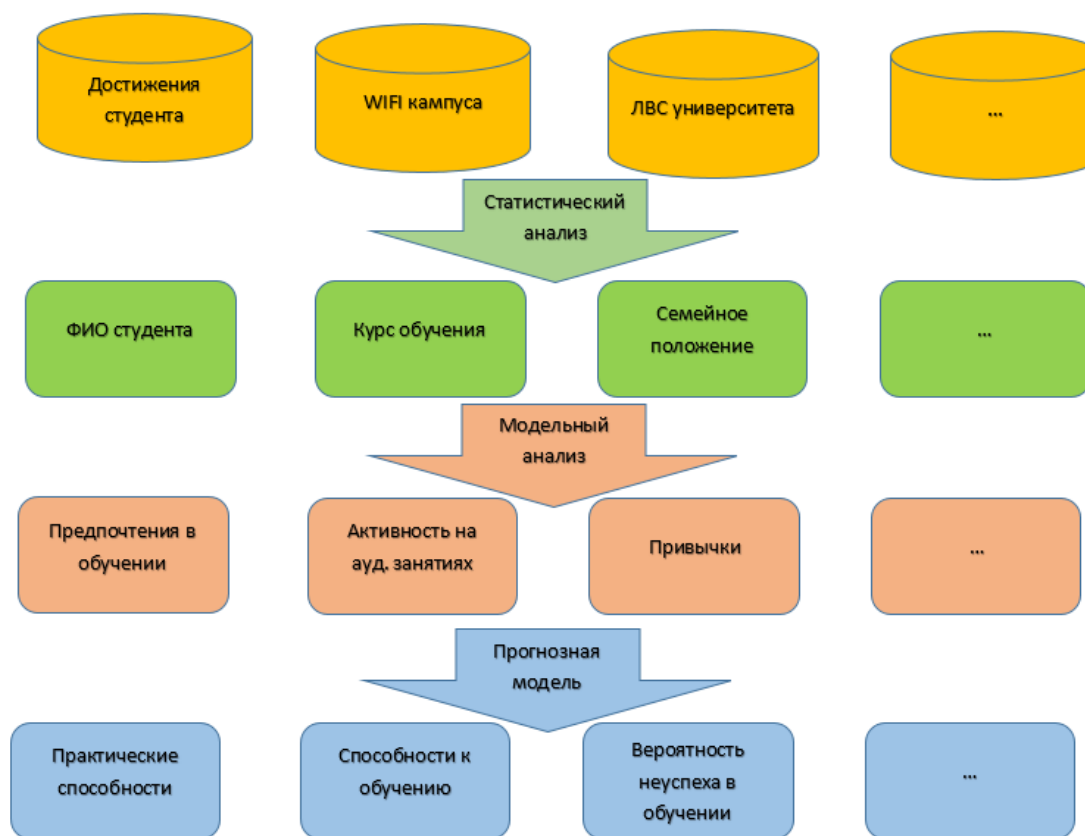


Рисунок 16 – Структурная схема анализа поведения студентов

Как следует из схемы, анализ поведения студентов состоит из следующих этапов [40]:

- сбор данных о цифровом следе студента;
- предварительная обработка исходных данных и выбор признаков для построения соответствующего индекса поведения студента;
- кластеризация, корреляционный анализ и другие связанные алгоритмы для создания соответствующей модели и проверки ее адекватности;
- использование результатов анализа в процессе обучения в университете.

Выделяются следующие области исследования поведения студентов на основе анализа цифрового следа:

- анализ потребительских предпочтений студентов. Карточная система доступа в кампус и оплаты в университетских столовых,

являющаяся ядром построения цифрового кампуса, генерирует большое количество данных о потреблении, использовании Интернета, траектории кампуса и др. Как правило, данные о потреблении студентов в университетах характеризуются большим количеством пользователей, большим ежедневным объемом транзакций и небольшой суммой. Эти данные содержат обширные знания, которые можно эффективно подсчитывать, анализировать, чтобы найти некоторые закономерности и потребительских предпочтений студентов [41];

- анализ образа жизни студентов. Образ жизни студента - это определенный способ интеграции потребностей и соответствующей деятельности, сопровождающийся переживаниями и действиями. От образа жизни студента зависят его здоровье, психологическое состояние и успеваемость [14];
- анализ процесса обучения студентов. Успеваемость студентов – это одна из важнейших характеристик образовательной деятельности вуза, по которой можно судить о достигнутых результатах или имеющихся проблемах. При этом, чем больший производится охват анализируемого контингента студентов, тем сложнее выявлять истинные причины проблем в образовании студентов [3].

Для выделения групп пользователей со сходными характеристиками, как правило, используются различные методы извлечения данных Data mining, которые, в свою очередь, используют алгоритмы кластеризации [39].

С учетом вышеизложенного модель анализа поведения студента на основе его цифрового следа будет иметь вид, представленный на рисунке 17.



Рисунок 17 – Модель системы отбора студентов для кадрового резерва вуза

Применение данной модели позволит проанализировать поведение студентов путем выделения поведенческих групп студентов со сходными характеристиками по каждой совокупности индикаторов поведения.

Прогнозирование вхождения студента в конкретную поведенческую группу осуществляется на основе его цифрового следа.

3.2 Алгоритмы кластеризации данных цифрового следа студента

3.2.1 Алгоритм k-means

«Алгоритм k-means – это итеративный алгоритм, который пытается разделить набор данных на заранее определенные k -отдельных неперекрывающихся групп-кластеров, где каждая точка данных принадлежит только одной группе.

Алгоритм k-means пытается сделать точки данных внутри кластера как можно более похожими, но при этом сохраняя кластеры как можно более разными.

Он назначает точки данных кластеру таким образом, чтобы сумма квадратов расстояния между точками данных и центроидом кластера (среднем арифметическим всех точек данных, принадлежащих этому кластеру) была минимальной.

Чем меньше вариаций внутри кластеров, тем более однородные (похожие) точки данных находятся в одном кластере.

Подход, который использует k-means для решения проблемы, называется EM (Expectation-Maximization)-алгоритмом.

EM-алгоритм состоит из итерационного повторения двух шагов. На E-шаге вычисляется ожидаемое значение (expectation) вектора скрытых переменных G по текущему приближению вектора параметров Θ . На M-шаге решается задача максимизации правдоподобия (maximization) и находится следующее приближение вектора Θ по текущим значениям векторов G и Θ » [8].

«Алгоритм k-means состоит из следующих шагов:

Шаг 1. Случайным образом выбирается k объектов обучающей выборки, которые будут служить начальными центрами кластеров.

Шаг 2. Для каждого объектов обучающей выборки определяется ближайший к ней центр кластера. Для этого вычисляется расстояние между объектами и центрами кластеров. Считается, что объект принадлежит тому

кластеру, к которому он ближе. В качестве формулы для оценки близости объектов в многомерном пространстве признаков используется одна из известных метрик» [8].

Шаг 3. Как только состав кластеров на данной итерации известен, производится расчёт новых центров кластеров. Это делается путем расчета средних значений для каждого числового признака по всем объектам рассматриваемого кластера. Например, в двухмерном пространстве координаты центр кластера на основе вошедших в него t объектов рассчитывается следующим образом (3):

$$(P_{2ц}, P_{1ц}) = \left(\frac{\sum_1^t P_1(t)}{t}, \frac{\sum_1^t P_2(t)}{t} \right) \quad (3)$$

«Шаг 4. Шаги 2 и 3 повторяются до тех пор, пока не выполнятся один из двух критериев остановки:

- границы кластеров и расположения центров кластеров не перестанет изменяться от итерации к итерации, т.е. на каждой итерации в каждом кластере будет оставаться один и тот же набор записей. На практике алгоритм k-means обычно находит набор стабильных кластеров за несколько десятков итераций;
- достигнут критерий сходимости. Чаще всего используется критерий суммы квадратов ошибок между центром кластера и всеми вошедшими в него объектами (4):

$$E = \sum_{i=1}^k \sum_{p \in C_i} (p - m_i)^2 \quad (4)$$

где $p \in C_i$ - произвольная точка данных, принадлежащая кластеру C_i , m_i – центр данного кластера. Иными словами, алгоритм остановится

тогда, когда ошибка E достигнет достаточно малого значения.

Преимуществом алгоритма k-means является высокая производительность, поскольку все, что мы на самом деле делаем, это вычисляем расстояния между точками и центроидами групп.

Алгоритм имеет линейную сложность $O(n)$.

С другой стороны, у k-means есть недостатки:

- необходимо выбрать количество групп/классов. Это не всегда тривиально, и в идеале с алгоритмом кластеризации мы хотели бы, чтобы он выяснил это за нас, потому что его цель - получить некоторое представление о данных;
- алгоритм начинается со случайного выбора центров кластеров и, следовательно, может давать разные результаты кластеризации при разных прогонах алгоритма.

Таким образом, результаты могут быть неповторимыми и противоречивыми.

Поэтому большая область исследований в области кластеризации была сосредоточена на улучшении процесса кластеризации в том числе с помощью адаптивных алгоритмов k-means» [8].

3.2.2 Адаптивный алгоритм k-means

Идея адаптивного алгоритма состоит в том, чтобы оптимизировать набор выборочных данных с помощью правила локтя для выявления и устранения выбросов.

Метод локтя (Elbow Rule) – один из самых известных методов, с помощью которого вы можете выбрать правильное значение k и повысить производительность модели. Этот эмпирический метод вычисляет сумму квадратов расстояний между точками и вычисляет среднее значение [24].

Когда алгоритм реализован, используется выборочный набор данных, который устраняет выбросы, а после завершения алгоритма окончательный выброс определяется в соответствии с сходством между выбросами и каждым кластером.

На основе адаптивной идеи после завершения каждой итерации значение автоматически корректируется в соответствии с ошибкой индекса оценки кластера для каждого кластера до тех пор, пока не будет достигнут диапазон ошибок.

Определение подобия классического алгоритма k-means основано на евклидовом расстоянии. Выбросы будут влиять на оценку значения, тем самым увеличивая временную сложность алгоритма.

Используем метод локтя для эффективного обнаружения выбросов в наборе данных для оптимизации алгоритма.

Конкретная реализация выглядит следующим образом.

Пусть есть множество данных $X = \{x_i \mid i=1,2,\dots,m\}$, где m – количество образцов. Каждый образец имеет n свойств ($n>0$).

Образцы разделены на категории $C = \{c_1, c_2, \dots, c_k\}$, где k – количество кластеров.

Изначально все образцы в X рассматриваются как один класс, а центр начального класса равен (5):

$$\mu_j = \frac{1}{N(C_j)} \sum_{x_i \in C_j} x_i, \quad (5)$$

где C_j – набор образцов, содержащихся в j -ом кластере;

$N(C_j)$ – количество образцов в j -ом кластере.

Вычисляем набор евклидовых расстояний от каждого образца до центра кластера по формуле (6):

$$d(x_i, \mu_j) = \sqrt{(x_i - \mu_j)^2}, \quad (6)$$

где x_i – это i -образец, $i \in [1, m]$, $D = \{d_1, d_2, \dots, d_m\}$.

Согласно правилу локтя, сортировка данных в D от меньшего к

большому и получение $x-d$ двухразмерного линейного графа, где d – расстояние от точки данных образца до центральной точки, а x – образец, соответствующий d .

По мере увеличения d значение, соответствующее положению, в котором эффект улучшения искажения увеличивается больше всего, будет локтем.

Таким образом, локоть отвечает следующим условиям (7):

$$\Delta d = \max (d_1 - d_{i-1}), i \in [1, m]. \quad (7)$$

Разделим локоть и временно определим данные, расстояние которых больше, чем соответствующее расстояние точки локтя, как выбросы.

Предположим, что w ($w \leq n$) – это образцы после устранения выбросов, и $(w-m)$ - это выбросы.

Сохраним невыбросы в наборе образцов X и перенумеруем образцы, чтобы получить $X' = \{x_1, x_2, \dots, x_w\}$.

Выбросы сохраним в наборе данных Y и перенумеруем, чтобы получить $Y' = \{x_{w+1}, x_{w+2}, \dots, x_m\}$.

Использование набора образцов X' при реализации алгоритма позволит устранить в некоторой степени влияние выбросов.

После реализации алгоритм получим кластер $C = \{c_1, c_2, \dots, c_z\}$, где z – количество кластеров, и i – максимальное расстояние от образца в кластере до центра j -го кластера - $\max d_j$ (8):

$$\max d_j = \max \{d_{j1}, d_{j2}, \dots\} \quad (8)$$

В соответствии с (4) рассчитываем расстояние $\{(d_{(w+1)1}, d_{(w+1)2}, \dots, d_{(w+1)z}), (d_{(w+2)1}, d_{(w+2)2}, \dots, d_{(w+2)z}), \dots, (d_{m1}, d_{m2}, \dots, d_{mz})\}$ от образца из Y до центра каждого кластера.

Если есть кластер ($a \in (0, z)$), создаем образец x_b ($b \in [w+1, m]$),

удовлетворяющий условию, что расстояние от x_b до центра кластера a будет меньше, чем максимальное расстояние от образца в кластере до центра кластера (9):

$$(d_{b1} < \max d_1) \parallel (d_{b2} < \max d_2) \parallel \dots \parallel (d_{ba} < \max d_a) \quad (9)$$

Тогда разделим x_b в ближайший среди a кластеров (10):

$$j(x_b) = \min (d_{b1}, d_{b2}, \dots, d_{ba}) \quad (10)$$

Если такого кластера нет, образец определяется как выброс. Пока все образцы в Y не будут пройдены, окончательные выбросы можно отделить.

Рассмотрим выбор адаптивных k -значения.

Соответствующий выбор k -значения необходимо оценивать на основе индекса оценки кластеризации. В рассматриваемом случае используется сумма квадратов ошибок прогнозирования (SSE) внутри кластера.

Формула расчета этого показателя выглядит следующим образом (11):

$$E = \sum_{j=1}^k \sum_{x \in C_j} |x - \mu_j|^2, \quad (11)$$

где k – количество кластеров, x – образцы, μ_j – центр j -го кластера, C_j – набор образцов, содержащихся в j -м кластере.

E описывает близость каждого кластерного: чем меньше E , тем лучше эффект кластеризации.

Обозначим через J_e сумму квадратов ошибок в каждом кластере (12):

$$J_e = \sum_{x \in C_j} \frac{1}{N(C_j)} |x_i - \mu_j|^2, \quad (12)$$

где x_i – образец в j -ом кластере;

$N(C_j)$ – количество образцов в j -ом кластере;

μ_j – центр j -го кластера.

Чем меньше J_e , тем лучше эффект кластеризации в j -ом кластере.

Сначала устанавливается J_e , и порог минимального количества образцов в кластере. После разделения каждого кластера можно получить количество образцов в каждом кластере (N_1, N_2, \dots, N_k) и вычислить значение J_e каждого кластера по формуле (11).

На рисунке 18 представлена блок-схема адаптивного алгоритма k-means.

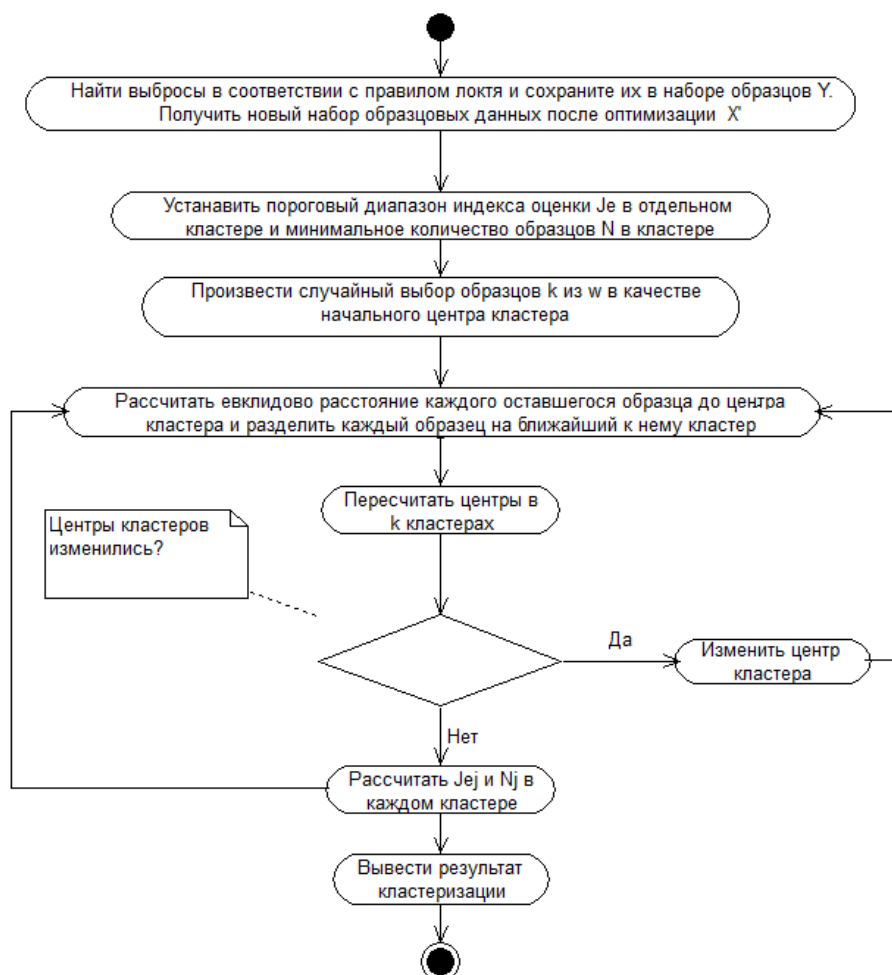


Рисунок 18 – Блок-схема адаптивного алгоритма k-means

Адаптивный алгоритм k-means состоит из следующих шагов:

Шаг 1. Оптимизация данных. Находим выбросы в соответствии с правилом локтя и сохраняем их в наборе образцов Y . Получаем новый набор образцовых данных после оптимизации $X' = \{x_1, x_2, \dots, x_w\}$.

Шаг 2. Устанавливаем пороговый диапазон индекса оценки J_e в отдельном кластере и минимальное количество образцов N в кластере.

Шаг 3. Случайно выбираем образцы k из w в качестве начального центра кластера, причем $1 < k \leq w$.

Шаг 4. Рассчитываем евклидово расстояние каждого оставшегося образца до центра кластера по формуле (7) и делим каждый образец на ближайший к нему кластер.

Шаг 5. Пересчитываем новый центр каждого кластера по формуле (6).

Шаг 6. Если новый центр кластера совпадает с исходным центром или меньше определенного порога, итерация завершается. Если новый центр кластера изменится, продолжайте повторять шаги 4 и 5 до сходимости.

Шаг 7. Вычисляем J_e каждого кластера и количество образцов N в каждом кластере. Сравниваем J_e и N с начальным пороговым диапазоном и вычисляем Δk_i по формуле (13):

$$\Delta k_i = \frac{\text{sgn}(\Delta N_j) - 1}{2} + \frac{\text{sgn}(\Delta N_j) + 1}{2} \cdot \prod (\log_w(\Delta J_e + 1)) \theta(\Delta J_e), \quad (13)$$

где w – количество образцов в X' ;

$N(C_j)$ – количество образцов в j -ом кластере;

$\text{sgn}()$ – символическая функция;

$\theta()$ – единичная ступенчатая функция;

$\Pi()$ – функция округления.

После обхода всех кластеров вычислите новое значение k и k' по формуле (14):

$$k' = k + \sum_{j=1}^k \Delta k_j, \quad (14)$$

Если какой-либо центр удален или добавлен, возвращайтесь к шагу 4 до тех пор, пока не останется новых или удаленных кластерных центров.

Шаг 8. Определяем окончательные разделенные образцы в соответствии со сходством между образцами в наборе образцов и центром каждого кластера.

Шаг 9. Вывод результата кластеризации $C = \{c_1, c_2, \dots, c_z\}$.

Таким образом, адаптивный алгоритм k-means позволит получить более точные результаты кластеризации данных для анализа цифрового следа студента.

3.3 Функциональная модель автоматизированной системы отбора студентов для кадрового резерва вуза

«С учетом вышеизложенного разработана диаграмма вариантов использования, которая представляет функциональную модель АСОС для кадрового резерва вуза.

Для разработки функциональной модели использована методология объектно-ориентированного анализа и проектирования.

Разработана диаграмма вариантов использования АСОС для кадрового резерва вуза.

Элементами диаграммы вариантов использования являются акторы – пользователи ИС или внешние взаимодействующие с ней системы.

Функции системы описываются с помощью вариантов использования (прецедентов), с которыми взаимодействуют акторы.

В результате анализа были выделены следующие акторы АСОС: ЛПР, Аналитик, Экспертная комиссия, СППР» [10].

Варианты использования АСОС описаны в таблицах 3-7.

Таблица 3 – Сформировать цифровой след студента

«Прецедент: Сформировать цифровой след студента
ID: 1
Краткое описание: Аналитик формирует цифровой след студента
Главный актер: Аналитик
Второстепенный актер: нет
Предусловие: нет
Основной поток: Аналитик формирует цифровой след студента для определения его поведенческой группы
Постусловие: нет
Альтернативные потоки: нет» [10]

Таблица 4 – Подготовить обучающую выборку

«Прецедент: Подготовить обучающую выборку
ID: 2
Краткое описание: Аналитик готовит обучающую выборку
Главный актер: Аналитик
Второстепенный актер: нет
Предусловие: нет
Основной поток: Аналитик готовит обучающую выборку для формирования поведенческих групп студентов
Постусловие: нет
Альтернативные потоки: нет» [10]

Таблица 5 – Сформировать поведенческие группы студентов

«Прецедент: Сформировать поведенческие группы студентов
ID: 3
Краткое описание: Экспертная комиссия формирует поведенческие группы студентов
Главный актер: Экспертная комиссия
Второстепенный актер: нет
Предусловие: нет
Основной поток: Экспертная комиссия формирует поведенческие группы студентов по результатам кластеризации
Постусловие: нет
Альтернативные потоки: нет» [10]

Таблица 6 – Выработка рекомендаций

«Прецедент: Выработка рекомендаций
ID: 4
Краткое описание: СППР формирует рекомендации для принятия решений
Главный актер: СППР
Второстепенный актер: нет
Предусловия: Подготовка обучающей выборки, Формирование цифрового следа студента, Формирование поведенческих групп
Основной поток: СППР формирует рекомендации для принятия решений ЛПР
Постусловие: нет
Альтернативные потоки: нет» [10]

Таблица 7 – Принятие решение о включении в кадровый резерв вуза

«Прецедент: Принятие решение о включении в кадровый резерв вуза
ID: 5
Краткое описание: ЛПР принимает решение о включении студента в кадровый резерв вуза
Главный актер: ЛПР
Второстепенный актер: СППР
Предусловие: нет
Основной поток: ЛПР принимает решение о включении студента в кадровый резерв вуза, используя рекомендации СППР
Постусловие: нет
Альтернативные потоки: нет» [10]

Между вариантами использования установлены отношения включения (include) и расширения (extend).

Диаграмма вариантов использования, отражающая функциональный аспект АСОС в кадровый резерв вуза, изображена на рисунке 19.

Данная диаграмма является функциональной моделью АСОС в кадровый резерв вуза.

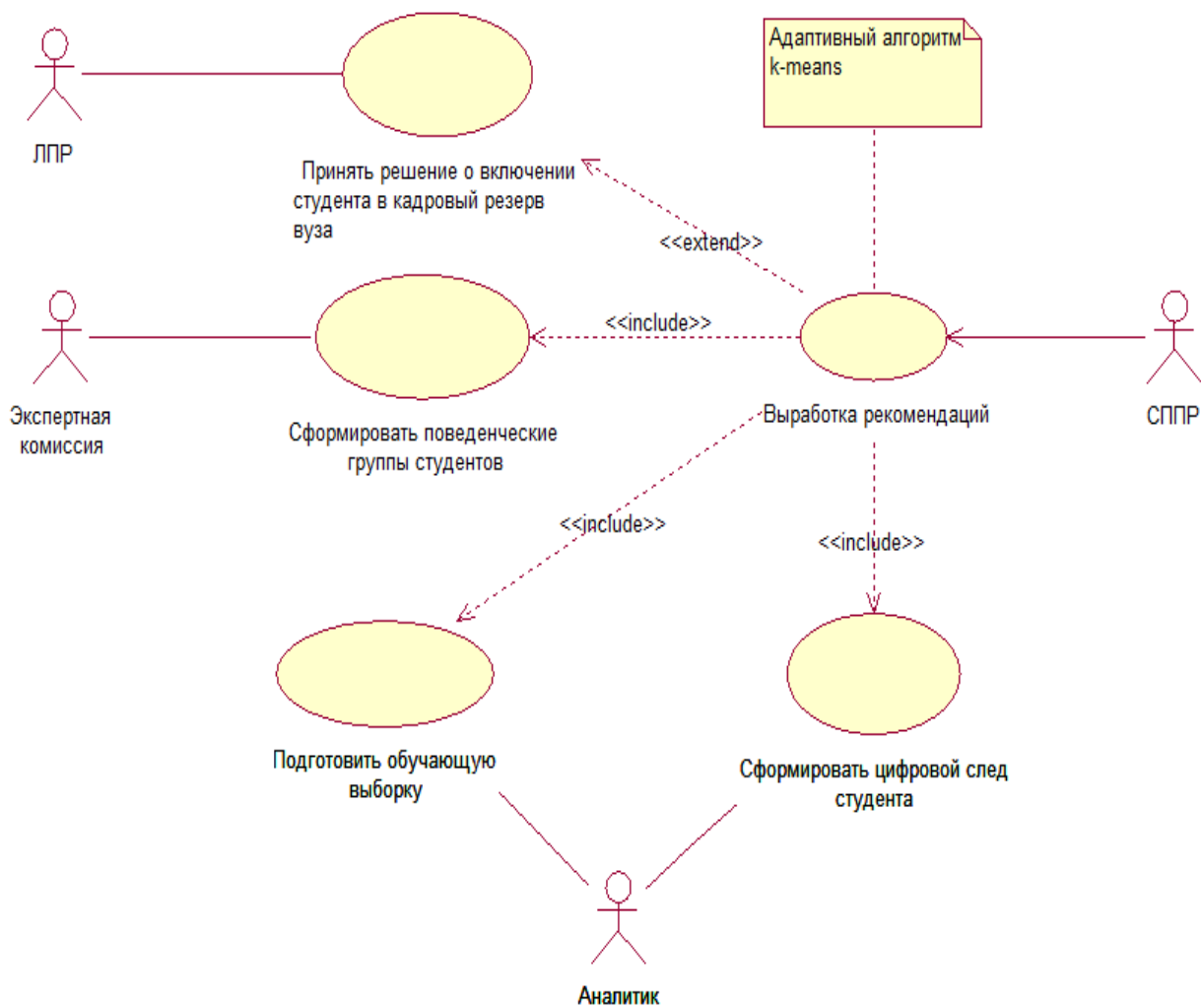


Рисунок 19 – Функциональную модель ACOS в кадровый резерв вуза

Для представления структуры системы и отражения ее элементного аспекта используется диаграмма классов.

Разработана диаграмма классов ACOS в кадровый резерв вуза, которая представлена на рисунке 20.

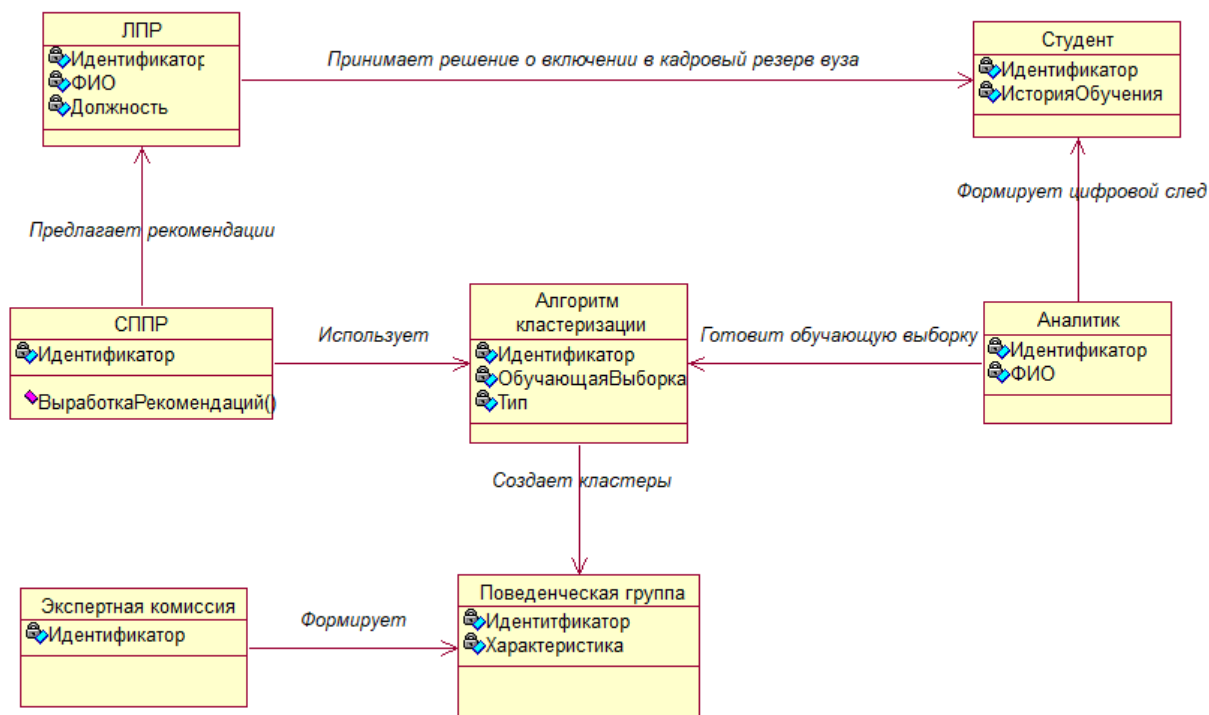


Рисунок 20 – Диаграмма классов АСОС в кадровый резерв вуза

Спецификация диаграммы классов представлена в таблице 8.

Таблица 8 – Спецификация диаграммы классов АСОС в кадровый резерв вуза

«Класс	Описание
ЛПР	Класс объектов, моделирующих на логическом уровне лиц, принимающих решение
Студент	Класс объектов, моделирующих на логическом уровне студентов
СППР	Класс объектов, моделирующих на логическом уровне СППР
Алгоритм кластеризации	Класс объектов, моделирующих на логическом уровне алгоритмы кластеризации. Наследник класса «Алгоритм Data mining»
Аналитик	Класс объектов, моделирующих на логическом уровне аналитиков
Экспертная комиссия	Класс объектов, моделирующих на логическом уровне экспертную комиссию
Поведенческая группа	Класс объектов, моделирующих на логическом уровне поведенческие группы студентов» [10]

Представленные модели и алгоритмы являются основой для построения эффективной АСОС в кадровый резерв вуза.

Выводы по главе 3

Результаты проделанной работы позволили сделать следующие выводы:

- в практических исследованиях и приложениях необходимо сочетать различные методы с реальной ситуацией для проведения исследований и анализа, чтобы улучшить работу в режиме реального времени и точность. В этой связи для анализа моделей поведения студентов выбран метод машинного обучения;
- выбор алгоритма кластеризации обусловлен характером решаемой задачи и функциональными особенностями АСОС в кадровый резерв вуза.

Представленные модели и алгоритмы являются основой для построения эффективной АСОС в кадровый резерв вуза.

Глава 4 Апробация проектных решений и оценка их эффективности

4.1 Апробация проектных решений

Для реализации алгоритма машинного обучения использованы язык Python и среда Jupyter Notebook (anaconda 3) с библиотекой sklearn [32].

Код адаптивного алгоритма кластеризации k-means представлен в листинге 1.

Листинг 1 – Код алгоритма k-means

```
«import copy
import math
import random
import numbers
import numpy as np
import matplotlib.pyplot as plt
%matplotlib notebook
def element_distance(elem1, elem2):
    """
    Euclidean distance between two elements.
    """
    diff = elem1 - elem2
    sum_squares = 0.
    for prop in diff.get_properties():
        sum_squares += prop ** 2
    distance = math.sqrt(sum_squares)
    return distance
def element_cluster_distance(element, cluster):» [7]
    """
```

```

    «Euclidean distance between an element and a cluster, which is just
    the distance between the element and the centroid of the cluster.
    """

    return element_distance(element, cluster.get_centroid())
def cluster_distance(cluster1, cluster2):
    """
    Euclidean distance between two clusters, which is just the distance
    between their centroids.
    """

    return element_distance(cluster1.get_centroid(), cluster2.get_centroid())
def find_closest_cluster(element, clusters):
    """
    Find the closest cluster to an element and the distance.
    """

    closest_cluster = None
    min_distance = float('inf')
    for cluster in clusters:
        distance = element_cluster_distance(element, cluster)
        if distance == 0:
            # Case 1 in the paper
            # "If the distance of the element from a
            # cluster is 0, assign the element to that
            # cluster, ..."
            return cluster, 0
        if distance < min_distance:
            min_distance = distance
            closest_cluster = cluster
    return closest_cluster, min_distance
def find_closest_clusters(clusters):» [7]
    """

```

```

«Find the two closest clusters and the distance.
"""
cluster1, cluster2 = None, None # C_{m_1} and C_{m_2} in the paper
min_distance = float('inf')
n = len(clusters)
for i in range(n - 1):
    for j in range(i + 1, n):
        distance = cluster_distance(clusters[i], clusters[j])
        if distance < min_distance:
            min_distance = distance
            cluster1, cluster2 = clusters[i], clusters[j]
return cluster1, cluster2, min_distance
def generate_clusters(num_elements, num_dims, num_clusters):
    """
    Generate specified number of elements with specified number of
dimensions in
specified number of clusters for the clustering algorithm.
"""
    assert num_elements >= num_clusters, \
        'Number of elements should not be smaller than number of clusters'
    clusters = []
    elements = []
    # sample cluster boundaries to determine the number of elements in each
cluster
    boundaries = [0, num_elements]
    for _ in range(num_clusters - 1):
        while True:
            boundary = np.random.randint(1, num_elements)
            if not boundary in boundaries:
                break» [7]

```

```

    <boundaries.append(boundary)
boundaries = sorted(boundaries)
# Generate elements for each cluster
for start, end in zip(boundaries[:-1], boundaries[1:]):
    # Generate an initial element, each following one is by adding a random
bias to it
    initial = np.random.randint(low=0, high=10, size=(2,))
    for _ in range(start, end):
        # Make the elements lie in a circle whose radius is random
        # and also proportional to the number of elements
        radius = 10. * (end - start) / num_elements * random.random()
        bias = (np.random.uniform(size=(2,)) * 2 - 1) * radius
        bias[1] = math.sqrt(radius ** 2 - bias[0] ** 2)
        if random.random() < 0.5:
            bias[1] *= -1
        elements.append(Element(initial + bias))
    # Generate one cluster using above elements
    cluster = Cluster(elements[start])
    for element in elements[start + 1: end]:
        cluster.add(element)
    clusters.append(cluster)
random.shuffle(elements)
return clusters, elements
def visualize(clusters, title):
    """
    Plot the clusters with a title.
    """
    plt.figure(figsize=(16, 8))
    for cluster in clusters:
        elements = cluster.get_elements()» [7]

```

```

<<x, y = [], []
    for element in elements:
        properties = element.get_properties()
        x.append(properties[0])
        y.append(properties[1])
    plt.scatter(x, y)
plt.title(title)
plt.show()
class Element(object):
    """
    Element represents a data point with multiple properties.
    """
    def __init__(self, properties):
        assert isinstance(properties, np.ndarray)
        self.properties = properties
    def __len__(self):
        return len(self.properties)
    def __str__(self):
        return '(' + ', '.join(map(str, self.properties)) + ')'
    def get_properties(self):
        return self.properties
    def __add__(self, other):
        assert isinstance(other, Element)
        return Element(self.get_properties() + other.get_properties())
    def __sub__(self, other):
        assert isinstance(other, Element)
        return Element(self.get_properties() - other.get_properties())
    def __mul__(self, other):
        assert isinstance(other, numbers.Number)»[7]

```

```

    <<return Element(self.get_properties() * other)
def __truediv__(self, other):
    assert isinstance(other, numbers.Number)
    return Element(self.get_properties() / other)
class Cluster(object):
    """
    Cluster represents a cluster which may include multiple elements and
    a computed centroid.
    """
    def __init__(self, element):
        assert isinstance(element, Element)
        self.elements = [element]
        self.centroid = copy.deepcopy(element)
    def __len__(self):
        return len(self.elements)
    def __str__(self):
        return 'elements: ' + ', '.join(map(str, self.elements)) + '; centroid: ' +
str(self.centroid)
    def add(self, element):
        self.centroid *= len(self)
        self.centroid += element
        self.elements.append(element)
        self.centroid /= len(self)
    def remove(self, element):
        self.centroid *= len(self)
        self.centroid -= element
        self.elements.remove(element)
        if len(self) == 0:
            self.centroid = Element(np.zeros_like(self.centroid.get_properties()))
        else:»[7]

```



```

        «self.centroid /= len(self)
def get_elements(self):
    return self.elements
def get_centroid(self):
    return self.centroid
class Clustering(object):
    """
    Adaptive K-Means Clustering.
    """
    def __init__(self, elements, num_clusters):
        assert len(elements) >= num_clusters, 'Number of elements should not
be smaller than number of clusters'
        self.initialize_clusters(elements, num_clusters)
        self.update()
    def __len__(self):
        return len(self.clusters)
    def initialize_clusters(self, elements, num_clusters):
        """
        Intialize the specified number of cluster seeds randomly.
        """
        random.shuffle(elements)
        seeds = elements[:num_clusters]
        self.clusters = [Cluster(seed) for seed in seeds]
        self.elements = elements[num_clusters:]

    def update(self):
        self.cluster1, self.cluster2, self.min_distance =
find_closest_clusters(self.clusters)
    def add(self, element):» [7]
    """

```

```

«Add an unclustered element into the clustering.
"""
    closest_cluster,    min_distance    =    find_closest_cluster(element,
self.clusters)

    if min_distance <= self.min_distance:
        # Case 2 in the paper
        #     "If the distance of the element from a cluster is less than the
distance
        #     d_{min}, assign this element to its closest cluster."
        closest_cluster.add(element)
    else:
        # Case 3 in the paper
        #     "merge C_{m_2} into C_{m_1}. Also, we destroy the cluster
C_{m_2} by removing
        #     all the elements from the cluster... we add the new element into
this now empty
        #     cluster."
        for element2 in self.cluster2.get_elements():
            self.cluster1.add(element2)
            self.cluster2.remove(element2)
        self.cluster2.add(element)
    self.update()
def cluster(self):
    while self.elements:
        self.add(self.elements.pop(0))
def get_clusters(self):
    return self.clusters
if __name__ == '__main__':
    num_elements = 1000
    num_dims = 2» [7]

```

```

«num_clusters = 10
clusters_gt, elements = generate_clusters(num_elements, num_dims,
num_clusters)
clustering = Clustering(elements, num_clusters)
clustering.cluster()
clusters_akm = clustering.get_clusters()
visualize(clusters_gt, 'Ground truth clusters')
visualize(clusters_akm, 'Clusters generated by adaptive k-means')» [7]

```

Для описания поведенческого типа студента использованы три признака:

- x, посещаемость занятий (в %);
- y, средний балл успеваемости студента;
- z, коэффициент участия в научно-проектной деятельности.

Для обучения алгоритма k-means разработан тестовый набор образовательных данных, загруженный в CSV-файл (рисунок 21).

1	x,y,z
2	95,5.0,10
3	94,4.9,9
4	95,4.8,10
5	96,5,9
6	95,5.0,5
7	94,4.9,5
8	95,4.8,5
9	96,5,4
10	60,4.8,10
11	55,4.9,9
12	55,5.0,10
13	50,4.8,9
14	50,4.1,4
15	55,4.1,5
16	55,4,3
17	50,4.2,4

Рисунок 21 – Структура файла обучающей выборки

Результат кластеризации по двум признакам (посещаемость и успеваемость) показан на рисунке 22.

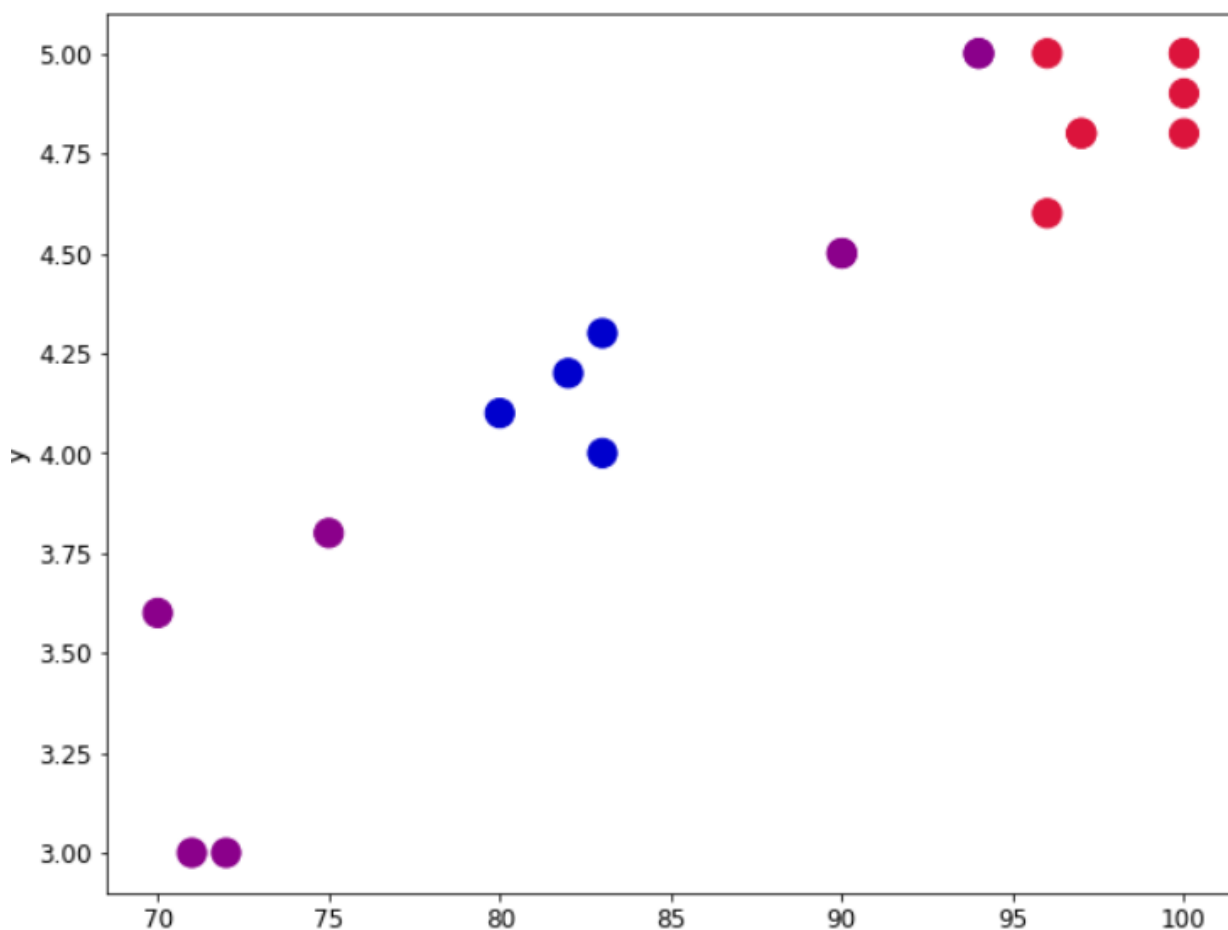


Рисунок 22 – График результата кластеризации по посещаемости и успеваемости студентов

Далее с помощью разработанной программы выполняем кластеризацию обучающей выборки и выделяем поведенческие группы студентов.

Для упрощения задаем 4 поведенческих типа (кластера).

На практике их может получиться больше или меньше данного количества.

Результат кластеризации показан на рисунке 23.

Введите количество кластеров:

5

[0 0 0 ... 0 0 0]

[[2.61442441 1.14875173 0.6074896]

[2.11320755 1.33962264 26.2754717]

[2.38319328 1.35630252 4.58386555]

[20. 4.57692308 0.97307692]

[2.31395349 1.38372093 12.38895349]]

Рисунок 23 – Результат кластеризации данных

В результате были выделены 4 поведенческие группы студентов, представленные в таблице 9.

Таблица 9 – Поведенческие группы студентов

Поведенческая группа	Посещаемость занятий (%)	Средний балл успеваемости студента	Степень участия в научно-проектной деятельности
0	95	4.9	9.5
1	95	4.9	4.8
2	56	4.9	9.7
3	52	4.2	5

В обучающей выборке были использованы данные студентов с успеваемостью более 4.0.

По рекомендации экспертов были сформированы следующие поведенческие группы студентов:

- группа 0: студенты с высокими показателями посещаемости, успеваемости и научно-проектной деятельности. Эти студенты могут быть рекомендованы для работы в качестве преподавателей вуза и сотрудников НИЧ;
- группа 1: студенты с высокими показателями посещаемости и успеваемости и средними показателями научно-проектной

деятельности. Эти студенты могут быть рекомендованы для работы преподавателями вуза;

- группа 2: студенты со средними показателями посещаемости, высокой успеваемостью и высокими показателями научно-проектной деятельности. Эти студенты могут быть рекомендованы для работы в НИЧ;
- группа 3: студенты со средними показателями посещаемости, хорошей успеваемостью и средними показателями проектной деятельности. Эти студенты не рассматриваются как кандидаты в кадровый резерв вуза.

«Для принятия решения по конкретному студенту необходимо спрогнозировать его вхождение в одну из представленных групп.

Входными данными для прогнозирования являются данные из цифрового следа студента.

Код для прогнозирования имеет вид:

```
# class predict
predicted_class = kmeans.predict([[90,5,9]])
print(predicted_class)» [41]
```

Результат выполнения кода показан на рисунке 24.

```
[0 0 0 0 3 3 3 3 1 1 1 2 2 2 2 2]
[[95.          4.925          9.5          ]
 [56.66666667  4.9          9.66666667]
 [52.          4.24          5.          ]
 [95.          4.925          4.75          ]]
[0]
```

Рисунок 24 – Результат прогнозирования группы студента

Таким образом, используя цифровой след конкретного студента, можно спрогнозировать его вхождение в одну из сформированных поведенческих групп, а, следовательно, определить перспективы его включения в кадровый

резерв вуза.

Окончательное решение о включении студента в кадровый резерв вуза принимается соответствующим коллегиальным органом.

4.2 Оценка эффективности управления автоматизированной системы

«Для оценки эффективности управления АСОС для кадрового резерва вуза предприятия используем формулу (14):

$$K_{\text{эу}} = \frac{\sum_{i=1}^n P_{yi}}{n} \quad (14)$$

где:

n - количество функций управления, реализуемых АСОС;

P_{yi} - вероятность выработки АСОС эффективного управляющего воздействия при реализации i -й функции управления.

Для поддержки принятия решения в АСОС используются следующие функции:

- выработка рекомендаций для принятия решения;
- принятие решения ЛПР.

Как показывает практика, на выполнение функции «Принятие решения ЛПР» может негативно повлиять человеческий фактор.

Пусть вероятность выработки эффективного управляющего воздействия для каждой функции равна 0.5.

В этом случае значение показателя функциональной эффективности управления АСУ андеррайтингом будет равно:

$$K_{\text{эу}} = 1.5/2 = 0,75$$

Таким образом, коэффициент эффективности управления предлагаемой АСОС $K_{эу} > 0,5$, что свидетельствует о высокой функциональной эффективности управления отбором студентов для кадрового резерва вуза» [2].

Выводы по главе 4

Результаты проделанной работы позволили сделать следующие выводы:

- для описания поведенческого типа студента использованы три признака: посещаемость занятий (в %), средний балл успеваемости студента и коэффициент участия в научно-проектной деятельности;
- для обучения алгоритма k-means разработан тестовый набор образовательных данных, загруженный в CSV-файл;
- как показали результаты апробации проектного решения, используя цифровой след конкретного студента, можно спрогнозировать его вхождение в одну из сформированных поведенческих групп, а, следовательно, определить перспективы его включения в кадровый резерв вуза.

Как показали расчеты, коэффициент эффективности управления предлагаемой АСОС $K_{эу} > 0,5$, что свидетельствует о высокой функциональной эффективности управления отбором студентов для кадрового резерва вуза.

Заключение

В настоящее время перед вузами поставлена задача создания внутреннего кадрового резерва в том числе из обучающихся в вузе студентов, используя эффективную методику отбора последних.

Неправильные решения при выборе студента для кадрового резерва вуза могут привести к значительным экономическим потерям, а, главное, могут негативно повлиять на будущую карьеру молодого специалиста, что соответствующим образом отразится на имидже вуза. Для решения данной проблемы на предприятиях используются автоматизированные системы отбора персонала.

Магистерская диссертация посвящена актуальной проблеме исследования и разработки моделей и алгоритмов эффективной автоматизированной системы отбора студентов для кадрового резерва вуза.

В процессе выполнения магистерской диссертации были решены следующие задачи:

- проведен анализ современного состояния исследований по отбору персонала для кадрового резерва вуза. Как показал анализ, для отбора кандидатов в кадровый резерв организации используются методы, основанные на модели компетенций и анализе поведенческой модели кандидатов. Широкие возможности для применения, точность и валидность методики DISC обусловили ее выбор для отбора студентов для кадрового резерва вуза. По своим функциональным и архитектурным особенностям автоматизированные системы отбора персонала для кадрового резерва относятся к категории СППР. Анализ литературы и источников подтвердил интерес ученых и специалистов к проблеме отбора персонала для различных социальных систем;
- проведен анализ методологий построения эффективных автоматизированных систем отбора персонала для кадрового резерва

вуза. Как показал анализ, методология объектно-ориентированного моделирования предоставляет широкие возможности для моделирования различных аспектов СППР на UML. Недостатком данной методологии являются слабая иерархия моделей и относительно высокие затраты на разработку. Как показывает практика, применение интеллектуального анализа данных в СППР позволяет оптимизировать работу ЛПР и снизить потери от возможных ошибок. С учетом вышеизложенного для построения АСОС для кадрового резерва вуза выбираем комплексный подход, позволяющий использовать обе рассмотренные методологии, каждая из которых будет использована для разработки моделей и алгоритмов системы;

- разработаны модели и алгоритмы автоматизированной системы отбора студентов для кадрового резерва вуза. В практических исследованиях и приложениях необходимо сочетать различные методы с реальной ситуацией для проведения исследований и анализа, чтобы улучшить работу в режиме реального времени и точность. В этой связи для анализа моделей поведения студентов выбран метод машинного обучения. Выбор алгоритма кластеризации обусловлен характером решаемой задачи и функциональными особенностями АСОС в кадровый резерв вуза. Представленные модели и алгоритмы являются основой для построения эффективной АСОС в кадровый резерв вуза;
- выполнены апробация и оценка эффективности проектных решений. Для описания поведенческого типа студента использованы три признака: посещаемость занятий (в %), средний балл успеваемости студента и коэффициент участия в научно-проектной деятельности. Для обучения алгоритма k-means разработан тестовый набор образовательных данных, загруженный в CSV-файл. Как показали результаты апробации проектного решения, используя цифровой

след конкретного студента, можно спрогнозировать его вхождение в одну из сформированных поведенческих групп, а, следовательно, определить перспективы его включения в кадровый резерв вуза. Как показали расчеты, коэффициент эффективности управления предлагаемой АСОС $K_{эу} > 0,5$, что свидетельствует о высокой функциональной эффективности управления отбором студентов для кадрового резерва вуза.

Таким образом, была решена проблема исследования и разработки моделей и алгоритмов эффективной автоматизированной системы отбора студентов для кадрового резерва вуза.

Гипотеза исследования подтверждена.

Работа может представлять интерес для бизнес-аналитиков и разработчиков систем управления персоналом организаций социальной сферы.

Список используемой литературы и используемых источников

1. Баранова Е. В., Швецов Г. В. Методы и инструменты для анализа цифрового следа студента при освоении образовательного маршрута // Перспективы науки и образования. 2021. № 2 (50). С. 415-430.
2. Вдовин В.М., Суркова Л.Е., Шурупов А.А. М. : Дашков и К, 2016. 388 с.
3. Жданов Д.Н. Анализ успеваемости студентов для оценки деятельности куратора // Гарантии качества профессионального образования. 2010. С. 317-320.
4. Ильченко С. В. Отбор персонала как составляющая кадровой политики организации // Бизнес и дизайн ревю. 2019. № 1 (13). С. 5.
5. Кадровый резерв [Электронный ресурс]. URL: https://www.audit-it.ru/terms/trud/kadrovuyu_rezerv.html (дата обращения: 15.03.2023).
6. Кальмай О.М., Юдина Е.С., Шерина О.А., Латышев А.С. Кадровый резерв как фактор повышения конкурентоспособности современного университета // Университетское управление: практика и анализ. 2016. № 101(1). С. 98-108.
7. Код адаптивного алгоритма k-means на языке Python [Электронный ресурс]. URL: <https://gist.github.com/jianchaoli/7065b510305d4df48063a801e6433213#file-adaptive-kmeans-ipynb> (дата обращения: 15.03.2023).
8. Котов К., Красильников Н. Кластеризация данных [Электронный ресурс]. URL: <https://logic.pdmi.ras.ru/~yura/internet/02ia-seminar-note.pdf> (дата обращения: 17.02.2023).
9. Кулагина Е. Модель компетенций: как разработать для разных должностей [Электронный ресурс]. URL: <https://www.hr-director.ru/article/66647-model-kompetentsiy-19-m4> (дата обращения: 15.03.2023).
10. Леоненков А. В. Объектно-ориентированный анализ и

проектирование с использованием UML и IBM Rational Rose : учебное пособие. М. : Интернет-Университет Информационных Технологий (ИНТУИТ), Ай Пи Ар Медиа, 2020. 317 с. [Электронный ресурс]. URL: <https://www.iprbookshop.ru/97554.html> (дата обращения: 24.03.2023).

11. Методы отбора соискателей [Электронный ресурс]. URL: Источник: https://www.jcat.ru/job_vacancy/blog/ocenka-i-otbor-personala/ (дата обращения: 15.03.2023).

12. Мкртычев С.В. Big data: подходы к определению и классификации // Сборник «Информационные технологии в моделировании и управлении: подходы, методы, решения». III Всероссийская научная конференция с международным участием: сборник докладов. Тольятти. 2021. С. 253-258.

13. Никитина Н.Ш., Бурмистрова Е.В. Методика отбора персонала на вакансию на основе нечетких показателей // Университетское управление. 2004. № 3(31). С. 98-103.

14. Парфенов А. С. Образ жизни и здоровья студентов: проблемы, анализ, опыт, рекомендации // Ученые записки ОГУ. Серия: Гуманитарные и социальные науки. 2012. №5. URL: <https://cyberleninka.ru/article/n/obraz-zhizni-i-zdorovya-studentov-problemy-analiz-opyt-rekomendatsii> (дата обращения: 20.03.2023).

15. Перфильев Д. А., Раевич К.В., Пятаева А.В. Интеллектуальные системы поддержки принятия решений : учебное пособие. Красноярск : Сибирский федеральный университет, 2018. 136 с. [Электронный ресурс]. URL: <https://www.iprbookshop.ru/84359.html> (дата обращения: 19.03.2023).

16. Психологические модели поведения личности в конфликте: найти, проанализировать, исправить [Электронный ресурс]. URL: <https://zen.yandex.ru/media/id/5b8e55db419fdb00abb9fc8c/psihologicheskie-modeli-povedeniia-lichnosti-v-konflikte-naiti-proanalizirovat-ispravit-5b96d87367918d00aa70bd88> (дата обращения: 20.03.2023).

17. Скитер Н.Н., Кетько Н.В. Методика отбора кадров с применением интеллектуального анализа письменной речи кандидатов // МНИЖ. 2020. №5-

2 (95). С. 58-63.

18. Степаненко А.А., Фещенко А.В. «Цифровой след студента»: поиск, анализ, интерпретация // Открытое и дистанционное образование. 2017. № 4(68). С. 58-62.

19. Узел Байесовская сеть [Электронный ресурс]. URL: <https://www.ibm.com/docs/ru/spss-modeler/SaaS?topic=models-bayesian-network-node> (дата обращения: 15.03.2023).

20. Фот Ю.Д. Модели и алгоритмы системы отбора персонала на основе повышения достоверности данных при принятии решения: автореферат диссертации на соискание ученой степени к.т.н. Пенза: ПГУ. 2013, 22 с.

21. Экзамен GRE: структура и специфика [Электронный ресурс]. URL: <https://www.iqconsultancy.ru/articles/ekzamen-gre-struktura-i-spetsifika/> (дата обращения: 15.03.2023).

22. Ashwin Kumar Galipalli, Haritha Jyothi Madyala “Process to build an efficient decision support system”, University of Boras, 2012. 60 p.

23. Azar A., Sebt M.V., Ahmadi P., Rajaeian A. A model for personnel selection with a data mining approach: A case study in a commercial bank. SA Journal of Human Resource Management/SA Tydskrif vir Menslikehulpbronbestuur, 11(1), Art. #449, 2013. 10 p.

24. Chen H. Adaptive K-means clustering algorithm [Электронный ресурс]. URL: <https://www.spiedigitallibrary.org/profile/Chen.Hailin-86165> (дата обращения: 20.03.2023).

25. DISC – технология для жизни и бизнеса [Электронный ресурс]. URL: https://ttisi.ru/about_disc (дата обращения: 15.03.2023).

26. Educational Data Mining [Электронный ресурс]. URL: <https://educationaldatamining.org/> (дата обращения: 15.03.2023).

27. Faerber F., Weitzel T., Keim T. An Automated Recommendation Approach to Selection in Personnel Recruitment, AMCIS 2003 Proceedings.

28. Filiberto Y. A New Method For Personnel Selection Based On Ranking Aggregation Using A Reinforcement Learning Approach, Comp. y

Sist. vol.22 no.2 Ciudad de México, 2018 Epub 21-Ene-2021.

29. Jian Li “Design of Enterprise Human Resources Decision Support System based on Data Mining”, 2021 [Электронный ресурс]. URL: <https://doi.org/10.21203/rs.3.rs-805084/v1> (дата обращения: 19.03.2023).

30. Mislevy R.J. et al. Leverage Points for Improving Educational Assessment, U.S. Department of Education Office of Educational Research and Improvement, 2000. 38 p.

31. Pampouktsi P. et al. Learning Techniques on Selection and Positioning of Human Resources in the Public Sector, Open Journal of Business and Management. 2021, Vol.9, No.2.

32. Project Jupyter [Электронный ресурс]. URL: <https://jupyter.org/> (дата обращения: 17.03.2023).

33. Rupnik R., Kukar M. Decision support system to support decision processes with Data mining, Journal of information and organizational sciences. 2007, Volume 31, Number 1, P. 217-232.

34. Success Insights® DISC - оценка поведения [Электронный ресурс]. URL: http://rucubes.com/assessment/instruments-technologiya/integrated_system/disc.php (дата обращения: 15.03.2023).

35. Understanding the Digital Footprint in the Classroom and Beyond [Электронный ресурс]. URL: <https://www.goguardian.com/blog/understanding-the-digital-footprint> (дата обращения: 20.03.2023).

36. Using DISC to Help Students Succeed [Электронный ресурс]. URL: <https://blog.extendeddisc.org/using-disc-help-students-succeed> (дата обращения: 15.03.2023).

37. Using task models in model -based testing [Электронный ресурс]. URL: https://ucaat.etsi.org/2015/presentations/NSPYRE_Kherrazi2.pdf (дата обращения: 15.03.2023).

38. Vasilakos S., Iacobellis G., Stylios C.D., Fanti M.P. Decision Support Systems Based on a UML Description Approach, 2012 IEEE 6th International Conference “Intelligent Systems”. P. 41-46.

39. Wang L. Analysis of Application of Big Data in College Education Management, 2019 J. Phys.: Conf. Ser. 1314 012220.
40. Wei-hong Wang et al. A Review of Behavior Analysis of College Students: 10th International Conference on Computer Science, Engineering and Applications, Vol. 10, No. 10, 2020.
41. Wenming Hou et al. Analysis and prediction of college students' consumption behavior, 2020 J. Phys.: Conf. Ser. 1437 012105.
42. Xiaoying Shen, Chao Yuan "A College Student Behavior Analysis and Management Method Based on Machine Learning Technology", Hindawi: Wireless Communications and Mobile Computing, Volume 2021, Article ID 3126347, 10 pages.
43. Yen-Ching Chuang et al. A New Method For Personnel Selection Based On Ranking Aggregation Using A Reinforcement Learning Approach, Technological and Economic Development of Economy, 2020, Volume 26, Issue 4: 751–784.