

**Департамент образования и науки города Москвы
Государственное автономное образовательное учреждение
высшего образования города Москвы
«Московский городской педагогический университет»**

**Институт цифрового образования
Департамент информатизации образования**

На правах рукописи

Белякова Анна Вячеславовна

**АЛГОРИТМЫ И ПРОГРАММНОЕ ОБЕСПЕЧЕНИЕ ПРОЦЕССОВ
ОБРАБОТКИ БОЛЬШИХ ДАННЫХ ДЛЯ ЗАДАЧ УПРАВЛЕНИЯ
В СЕТЕВЫХ ДИСТАНЦИОННЫХ СИСТЕМАХ
ВЫСШЕГО ОБРАЗОВАНИЯ**

**Направление подготовки 09.06.01 – Информатика и вычислительная
техника**

Направленность «Управление в социальных и экономических системах»

**Научный доклад об основных результатах научно-квалификационной
работы (диссертации)**

**Научный руководитель: доктор технических наук, профессор,
профессор департамента информатизации
образования института цифрового образования
Ромашкова Оксана Николаевна**

Москва
2024

ОБЩАЯ ХАРАКТЕРИСТИКА ДИССЕРТАЦИОННОЙ РАБОТЫ

Актуальность темы исследования обусловлена необходимостью обработки больших объёмов данных об обучающихся и упрощения постановки задач управления на их основе. Повышение эффективности управления в высших образовательных организациях может достигаться за счёт анализа показателей полученных в результате обработки больших данных об обучающихся и как следствие оценки эффективности образовательного процесса. Одним из наиболее эффективных показателей качества образования являются успехи обучающихся, для каждого из которых определённым образом рассчитывается рейтинг, по которому их можно достаточно полно оценить.

В сетевых дистанционных системах высшего образования накапливается большой объём данных, хранение которых обеспечивается ресурсами организации, и актуальная необходимость ответа на вызовы времени состоит в том, чтобы не просто хранить их, но извлекать из них информацию. В таком случае мы можем воспользоваться ими для принятия решений, сделав хранимые объёмы данных полезнее, а принятие решений – основанном на результатах аналитики больших данных.

Исходя из этого, **актуальными** являются разработка и внедрение алгоритмов и программного обеспечения обработки данных, способных повысить эффективность использования больших данных, которые помогут определить, какие специальности более востребованы, целесообразность открытия профильных направлений и определение целевой аудитории приёмной кампании разнопрофильных программ высшего профессионального образования.

Степень разработанности темы исследования. Вопросами анализа больших данных занимались многие российские учёные, в том числе Губернаторов А.М., Макшанов А.В., Николаев И.Н., Радченко И.А, Тесленко И.Б.

Вопросам автоматизации решения задач управления посвящены труды следующих ученых: Гущина О.М., Лаптева С.В., Пономаревой Л.А, Ромашковой О.Н., Саяпина О.В., Чискидова С.В. и др.

Тем не менее именно проблема использования больших данных, полученных в результате работы дистанционных систем высшего образования, оказалась недостаточно разработана.

Исследования на эту тему становятся всё более востребованными, поскольку объём больших данных вырастает с развитием информатизации.

Цель и задачи исследования. Целью работы является повышение уровня эффективности управления в сетевых дистанционных системах высшего образования на основе разработки алгоритмов и программного обеспечения процессов обработки больших данных обучающихся.

Для достижения цели диссертационного исследования были поставлены и решены следующие основные **задачи**:

1. Проанализировать организационно-структурные модели, информационные потоки и процессы в высшей образовательной организации по обеспечению управления сетевой дистанционной системой.

2. Выполнить анализ существующих методов оценки деятельности дистанционных систем высшего образования для задач управления, используемых критериев оценки, особенностей управления процессов обработки больших данных.

3. Выполнить анализ алгоритмов и программного обеспечения, применяемых для прогнозирования показателей эффективности функционирования образовательной организации.

4. Выполнить сравнительный анализ существующих методов и алгоритмов анализа данных для решения задачи классификации и оценки качества решения задачи классификации обучающихся.

5. Разработать алгоритм оценки и анализа достижений в образовании для повышения эффективности управленческих решений.

6. Разработать модель функционирования дистанционной системы высшего образования, объединённой с программным обеспечением анализа больших данных.

7. Разработка и реализация программного обеспечения процессов анализа данных для задач управления в виде веб-приложения.

8. Произвести моделирование и реализацию программного обеспечения оценки эффективности управления для задач принятия решения в социально-экономических системах в области образования.

9. Разработать методику применения веб-приложения для задач управления в сетевых дистанционных системах высшего образования.

Объектом исследования являются системные связи и закономерности функционирования процессов управления в социально-экономических системах – сетевых дистанционных системах высшего образования.

Предметом исследования являются управленческие алгоритмы, методы и модели обработки больших данных для решения задач управления в сетевых дистанционных системах высшего образования.

Научная новизна работы заключается в том, что:

– Разработаны модели и алгоритмы управления в сетевых дистанционных системах высшего образования путем проверки и профилирования обучающихся с использованием результатов обработки больших данных об успеваемости и самореализации обучающихся и предлагающих индивидуальные рекомендации по поступлению на следующую ступень высшего образования, предсказывающих востребованность специальностей, особенности целевой аудитории определенных направлений подготовки, а также дающих возможность оценивать эффективность образовательного процесса для принятия управленческих решений.

– Смоделированы и разработаны программное обеспечение и база данных, позволяющие из исходных больших данных получать

структурированные данные и хранить их для последующей независимой аналитики при помощи различных методов машинного обучения, в том числе – лесов решений и нейросети.

– Разработана математическая модель оценки эффективности обработки и хранения данных для решения задач управления в сетевых дистанционных системах высшего образования.

– Разработана методика сбора данных и работы с ними в web-приложении.

Методология и методы исследований включают математический анализ, методы обработки информации, включая методы анализа больших данных. При создании программного обеспечения использовался подход объектно-ориентированного программирования. В машинном обучении были использованы ансамблевые методы и нейронные сети.

Положения, выносимые на защиту:

1. Модель существующих информационных потоков и процессов в высшей образовательной организации по обеспечению управления дистанционной системой;

2. Алгоритм оценки эффективности обучающихся для управленческих решений;

3. Модели базы данных обучающихся, предназначенная для анализа данных, позволяющая корректно сохранить данные, отображающие учебный процесс в сетевых дистанционных системах высшего образования;

4. Имитационная модель разработанного алгоритма и программы автоматизации процесса оценки эффективности в продукте «Loginom»;

5. Методика применения веб-приложения для задач управления в сетевых дистанционных системах высшего образования;

6. Математическая модель для оценки эффективности обработки и хранения данных в образовательной организации;

7. Программное обеспечение процессов обработки больших данных для задач управления в сетевых дистанционных системах высшего образования в виде информационной системы.

Степень достоверности и апробация результатов. Степень обоснованности и достоверности полученных научных результатов и выводов подтверждается достаточным уровнем корреляции между теоретико-экспериментальными результатами научно-квалификационной работы, и результатами авторов смежных исследований.

Помимо этого, дополнительными факторами достоверности результатов являются практическая реализация моделей и алгоритмов, а также апробация на таких конференциях, форумах и семинарах, как: XI Международная научно-практической конференции студентов, аспирантов и молодых ученых, Россия, Москва в 2020, II Международная научно-практическая конференция, Россия, Краснодар в 2018, XXIV Всероссийская научно-технической конференции студентов, молодых ученых и специалистов, Россия, Рязань в 2019, IEEE conference of russian young researchers in electrical and electronic

engineering, Россия, Москва в 2021 году, студенческая конференции «#ScienceJuice», Россия, Москва в 2019, 2020, 2021 годах; международная научная конференцию «Цифровая трансформация реального сектора экономики», Россия, Москва в 2021 году; научная конференции с международным участием «Открытая наука-2021», Россия, Москва; общеинститутская конференция «Математика и информатика в образовании и бизнесе», Россия, Москва в 2020 году; всероссийская научная конференция «Открытая наука», Россия, Москва в 2022, 2023 годах.

Публикации. Всего по результатам исследования было опубликовано 18 работ, из них 3 статьи – в журналах, включенных в перечень ВАК, одна публикация индексируется в базе Scopus, и одна работа включена в ядро РИНЦ. Получены два свидетельства государственной регистрации программы для ЭВМ.

Личный вклад. Все представленные в диссертации экспериментальные данные и результаты исследований получены лично автором.

Соответствие паспорту специальности. Содержание диссертационной работы соответствует паспорту специальности 2.3.4 – Управление в организационных системах: п. 4 Разработка информационного и программного обеспечения систем управления и механизмов принятия решений в организационных системах; п. 5 Разработка методов получения данных и идентификации моделей, прогнозирования и управления организационными системами на основе ретроспективной, текущей и экспертной информации; п. 9 Разработка методов и алгоритмов интеллектуальной поддержки принятия управленческих решений в организационных системах.

Теоретическая значимость результатов исследования состоит в выявлении особенностей анализа и обработки больших данных в сетевых дистанционных системах высшего образования упрощая процесс управления изменениями; внесенном научном вкладе в качестве представлений о подходах к использованию больших данных сетевых дистанционных системах высшего образования и алгоритмов рекомендаций при принятии управленческих решений для задач стратегического, тактического и оперативного управления на основе анализа данных; кроме того данная работа проводит анализ существующих методов машинного обучения, предлагая новый способ внедрения результатов обработки данных и за счёт предоставления данной информации способствует улучшению организационных коммуникаций.

Практическая значимость результатов работы заключается в следующем. Разработано и внедрено веб-приложение со встроенным инструментарием анализа подгружаемых данных, позволяющим выявить связь достижений обучающегося с поступлением на определенный профиль, вследствие чего, лицо, принимающее решение может планировать количество групп, преподавательский состав на следующий год исходя из спроса на определенные программы. В связи с этим ожидается значительное снижение

издержек образовательной организации, поскольку выявляются наиболее востребованные специальности, обосновывается целесообразность открытия профильных направлений подготовки, определяется целевая аудитория приёмной кампании разнопрофильных программ высшего профессионального образования. Для обучающихся становятся доступны рекомендации по профильным программам и элективным курсам. Применение технологий обработки больших данных будет способствовать более точному определению рейтинговой оценки обучающихся образовательных организаций. Компоненты данного веб-приложения рекомендованы к использованию в образовательной организации ГАОУ ВО МГПУ, а также, без использования исходной базы данных – в других образовательных организациях высшего образования.

Внедрение результатов НКР осуществлено в ГАОУ ВО МГПУ, что подтверждено соответствующими актами.

Структура работы. НКР состоит из введения, четырёх глав, заключения, списка литературы и двух приложений.

ОСНОВНОЕ СОДЕРЖАНИЕ РАБОТЫ

Во введении обоснована актуальность темы научно-квалификационной работы, её цели и задачи, научная новизна, приведена апробация результатов, практическая и теоретическая значимость научной работы

Первая глава содержит проведённый анализ существующих моделей и процессов управления в сетевых дистанционных системах высшего образования.

В первую очередь исследована реорганизации систем высшего образования для осуществления дистанционного формата образования, обозначены как цели перехода на дистанционный формат, так и проблемы, возникающие в процессе перехода. Проведён анализ моделей осуществления дистанционного обучения и выделены три модели: модель задач, модели использующие радио или телевидение, сетевая модель. Исследован фрагмент организационной структуры высшей образовательной организации, напрямую отвечающий за сетевое дистанционное обучение, а также схема взаиморасположения участников организации дистанционного обучения.

Был проведён анализ информационных потоков в высшей образовательной организации, представленный на рисунке 1.



Рисунок 1 – Информационные потоки в высшей образовательной организации по обеспечению управления дистанционной системой

Были исследованы бизнес-процессы обработки больших данных обучающихся для задач управления в дистанционных системах высшего образования с точки зрения заместителя директора по учебной работе и качеству образования, разработка была произведена в CA ERWin Process Modeler, представленная на рисунке 2.



Рисунок 2 – Управлять деятельностью сотрудников института по обработке данных обучающихся

Затем были описаны методы оценки деятельности дистанционных систем высшего образования для задач управления и проведён анализ используемых критериев оценки а также рассмотрены существующие программы для сбора и обработки больших данных и их анализа, которые могут быть применены в образовательной организации.

В данной главе представлено обоснование задачи по созданию алгоритмов и программного обеспечения, которое учитывало достижения обучающихся, с последующим анализом нейросетью, деревом решений, композиционными алгоритмами.

Вторая глава содержит анализ существующего программного обеспечения обработки данных и построения нейросетей. Проведенный в первой главе анализ существующих образовательных дистанционных систем выявил особенности обеспечения процессов обработки больших данных для задач управления, появляющихся в процессе дистанционного обучения: не все виды деятельности обучающихся могут быть автоматически загружены в базу данных; требуется создание или использование существующего программного обеспечения, позволяющего хранить и обрабатывать большие данные об обучающихся и одновременно обладающего инструментами, для их последующего анализа - вследствие чего было необходимо провести сравнительный анализ возможных реализаций хранения данных перед анализом нейросетью, деревом решений, композиционными алгоритмами. Минимум функциональных особенностей подсистемы анализа данных отображён на рисунке 3.



Рисунок 3 – Особенности подсистемы анализа данных

Величину потерь, которые терпит организация высшего образования,

не используя большие данные можно вычислить как (1):

$$l = \int (1 - r)\eta(1 - r)d(1 - r), \quad (1)$$

Где d – это вероятность выявления критической для принятия решения закономерности, принимающая значения от 0 до 1, а $\eta(1-r)$ – это функция плотности распределения вероятности уровня затрат, где затраты r выступают в виде уровня количественного ресурса организации, который возможно возместить.

Для разработки нейросети в веб-приложении был выбран фреймворк «Django» на языке программирования «Python» так как именно этот язык является оптимальным для разработки нейросети. А в качестве среды разработки была выбрана интегрированная среда разработки «Visual Studio» как наиболее мощный инструмент. В итоге в качестве реляционной СУБД была выбрана «Microsoft SQL Server Management Studio», позволяющей более эффективно работать с хранилищем или озером больших данных образовательной организации и загружать их в базу данных для анализа. После чего был рассмотрен, собственно, сам алгоритм обработки больших данных в общем виде. Существующих программ, благодаря которым можно было бы строить нейросети (и при этом данная программа не являлась бы исключительно средством работы с фотографиями и картинками) не очень много. Можно рассмотреть три наиболее известные из них и поддерживаемые на данный момент, это:

- «STATISTICA Automated Neural Networks» - автоматизированные нейронные сети;
- «Deductor», который сменил «Loginom», от той же компании также, как и продукт, сменившей название с BaseGroup Labs на «Loginom Company». Поскольку «Loginom» является более совершенным продуктом той же компании, будем рассматривать его;
- нейросетевой инструментарий от Matlab.

Была проведена классификация обучения нейросетей в контексте решения задач управления в высшей образовательной организации.

Вариантом интеграции нейросети в программное обеспечение (ПО) для задач управления является разработка ПО с помощью языка Python, в котором существуют такие бесплатные библиотеки как «Sklearn», «Keras» и «TensorFlow», благодаря которым можно провести обучение нейросети. Было проведено их сравнение, вследствие которого была выбрана библиотека Sklearn, как наиболее доступная. Ранее выяснив, что для данной задачи нам необходимо решить именно задачу классификации обучающихся, была выявлена необходимость обозреть прикладные методы и алгоритмы решения задачи классификации, представленные на рисунке 4.

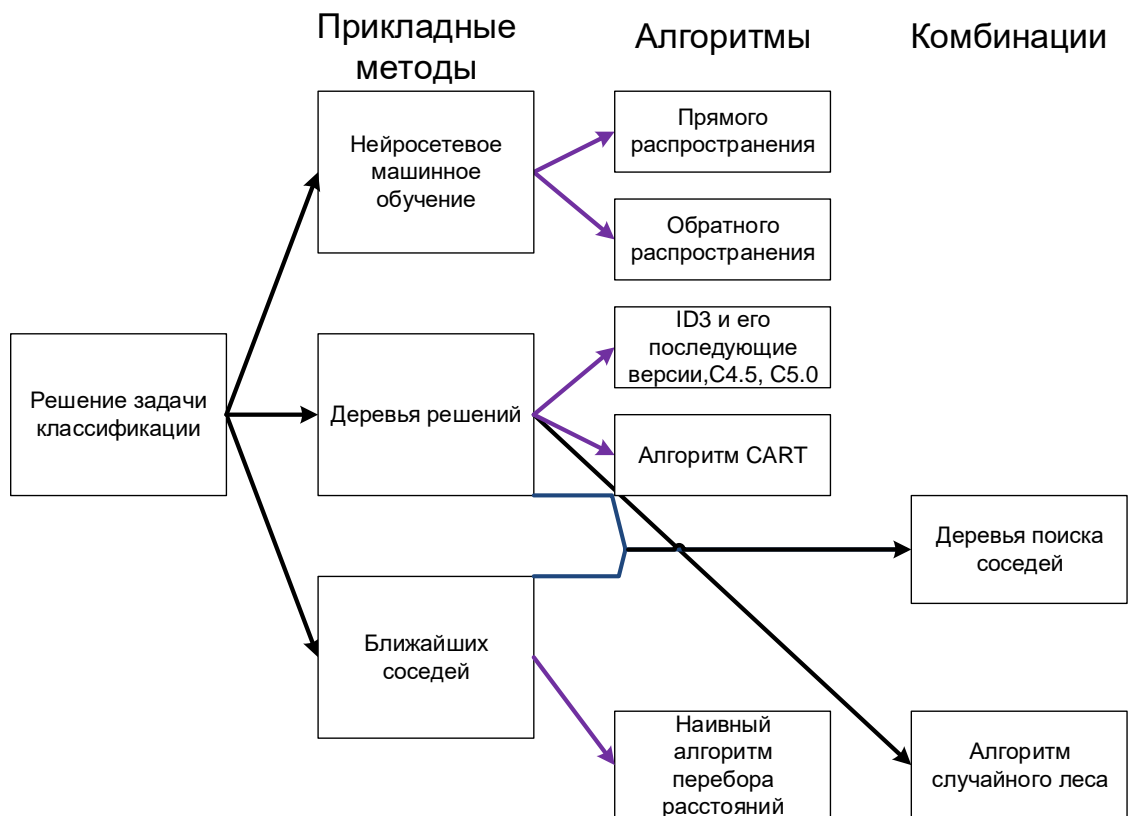


Рисунок 4 – Схема прикладных методов и алгоритмов решения задачи классификации

При рассмотрении всевозможных вариантов моделей были рассмотрены также и комбинации моделей. Помимо композиции алгоритмов, когда объединяются два или более метода решения задачи классификации, есть ансамблевые методы, говоря о которых, необходимо вернуться к вопросу о разбиении данных для обучения модели. Когда очевидно, какие наиболее прикладные методы лежат в основе решений задачи классификации можно перейти к их объединениям, в которых бывает много алгоритмов и много выборок. Что очевидным не является – это то, какой именно метод следует выбрать для классификации обучающихся.

Поскольку признаки являются достаточно разнородными, – рассмотрение классификаторов было начато с дерева решений. Был рассмотрен мета-алгоритм — такой алгоритм, который состоит из алгоритмов: в нём происходит обобщение нескольких алгоритмов в один. Именно таким алгоритмом является алгоритм агрегации начальной загрузки: деревья классификации объединяются в алгоритм случайного леса.

Ещё один композиционный мета-алгоритм — это классификация обучающихся последовательным объединением неглубоких деревьев, когда алгоритм не выбирает голосованием между независимыми деревьями итоговую классификацию, а строит деревья последовательно. Если в предыдущем мета-алгоритме в основном желательны были глубокие деревья, то тут необходимы неглубокие, так как они не независимы. У данного алгоритма много реализаций, но самые яркие из них это классификация

адаптивного усиления и классификация градиентного (по направлению наискорейшего роста функции) усиления.

Была рассмотрена классификация обучающихся многослойной нейронной сетью обратного распространения.

Предлагается не использовать реализовать отдельно нейросеть для определения количества обучающихся на каждом направлении и индивидуальной рекомендации по поступлению, а дерево решений – для определения особенностей целевой аудитории по направлениям, не усредняя все вышеозначенные особенности каждого метода классификации, а используя по максимуму преимущество каждого из них: ведь нейросеть это всего лишь один из подходов, не лучший из всех, для каждой задачи есть свой лучший подход и нейросеть является лучшим подходом только для части задач классификации.

Для того, чтобы замерить точность классификатора существует ряд методов, основанных на матрице неточностей, таких как вычисление точности и полноты. На основе такой матрицы можно вычислить метрики: метрика в контексте машинного обучения – это числовая мера, которая используется для оценки качества модели или алгоритма обучения. Метрика позволяет оценить, насколько хорошо модель выполняет свою задачу и дает представление о ее производительности. Метрика может быть выражена в виде числа, которое показывает определенный аспект модели, например точность предсказаний, полноту, среднюю ошибку и др. Метрик довольно много, от метрики аккуратной точности до метрики площади под кривой ошибок. Преимущества и недостатки разных метрик представлены на таблице 1.

В работе рассмотрено существующее программное обеспечение для прогнозирования оценкой эффективности показателей в сетевых дистанционных системах. Проанализированы алгоритмы и модели, подходящие для решения задачи классификации обучающихся, в том числе классификация обучающихся деревом решений, агрегацией начальной загрузки над деревьями решений, последовательным объединением неглубоких деревьев и нейронной сетью.

Т а б л и ц а 1 – Метрики оценки качества решения задачи классификации

Название/ Параметр	Преимущества	Недостатки
Аккуратная точность	Проста в интерпретации и позволяет оценить общую точность модели	Не учитывает дисбаланс классов
Точность	Помогает оценить, насколько модель точно предсказывает положительные результаты	Не учитывает ложно положительные случаи
Полнота	Позволяет оценить, насколько хорошо модель находит все положительные результаты	Не учитывает ложно отрицательные случаи
F_1 -мера	Учитывает и точность, и полноту, что полезно в случае, когда нужно достичь	Не учитывает вес различных ошибок

	баланса между обеими метриками	
ROC-AUC (Площадь под ROC-кривой)	Позволяет оценить общую производительность модели и ее способность различать между собой классы	Не учитывает вес различных ошибок
Логарифмическая потеря	Измеряет качество вероятностных предсказаний модели за счёт чего позволяет оценить, насколько хорошо модель оценивает вероятности для каждого класса	Не является интерпретируемой метрикой

Поскольку в каждом методе есть свои достоинства и недостатки, например деревья решений интерпретируемы, а нейросеть – нет, имеет смысл реализовать все рассмотренные способы классификации для их сравнительной оценки в рамках исследования и для решения специфических задач в процессе принятия управленческих решений.

В третьей главе было разработано веб-приложение. Для разработки нейросети в web-приложении был выбран фреймворк Django на языке программирования Python так как этот язык является оптимальным по своим характеристикам в этой области, в том числе за счёт качества и количества, подключаемых библиотек. Разработанный алгоритм анализа больших данных для прогнозирования учебного спроса для будущего года, сегментация и определение будущей профильной направленности обучающихся, реализуемый в данном программном обеспечении, позволяет более эффективно обрабатывать большие данные и принимать управленческие решения.

Разрабатываемый алгоритм оценки эффективности для управленческих решений должен приводить разнообразные данные, получаемые из больших данных, к однородному формату упрощая дальнейший анализ - используется шкала парных сравнений Саати для выставления весов достижений уже не попарно, а относительно всех остальных достижений в данной категории, что помогает учесть все достижения внутри каждой категории и определить их важность в контексте других достижений, все достижения разделяются на категории. Таким образом, обобщённый алгоритм оценки эффективности обучающихся состоит из следующих этапов: 1-й этап, выявление видов достижений; 2-й этап, присвоение весов; 3-й этап, консолидация данных; 4-й этап, выбор цели; 5-й этап, оценка с применением мета-алгоритмов и нейросети: используя мета-алгоритмы, а также нейросети для анализа данных и оценки эффективности обучающихся осуществляется их применения для агрегации и анализа рейтингов и весов достижений по каждой категории; 6-й этап, интерпретация результатов: производится анализ полученных оценок эффективности и делаются выводы о достижениях обучающихся. Возможно использование результатов для принятия управленческих решений, таких как идентификация наиболее успешных обучающихся, определение областей, требующих улучшений, или принятие решений о развитии образовательных программ.

Дистанционная система высшего образования, дополненная

программным обеспечением анализа больших данных, должна обладать функционалом, приведённым на диаграмме 5.

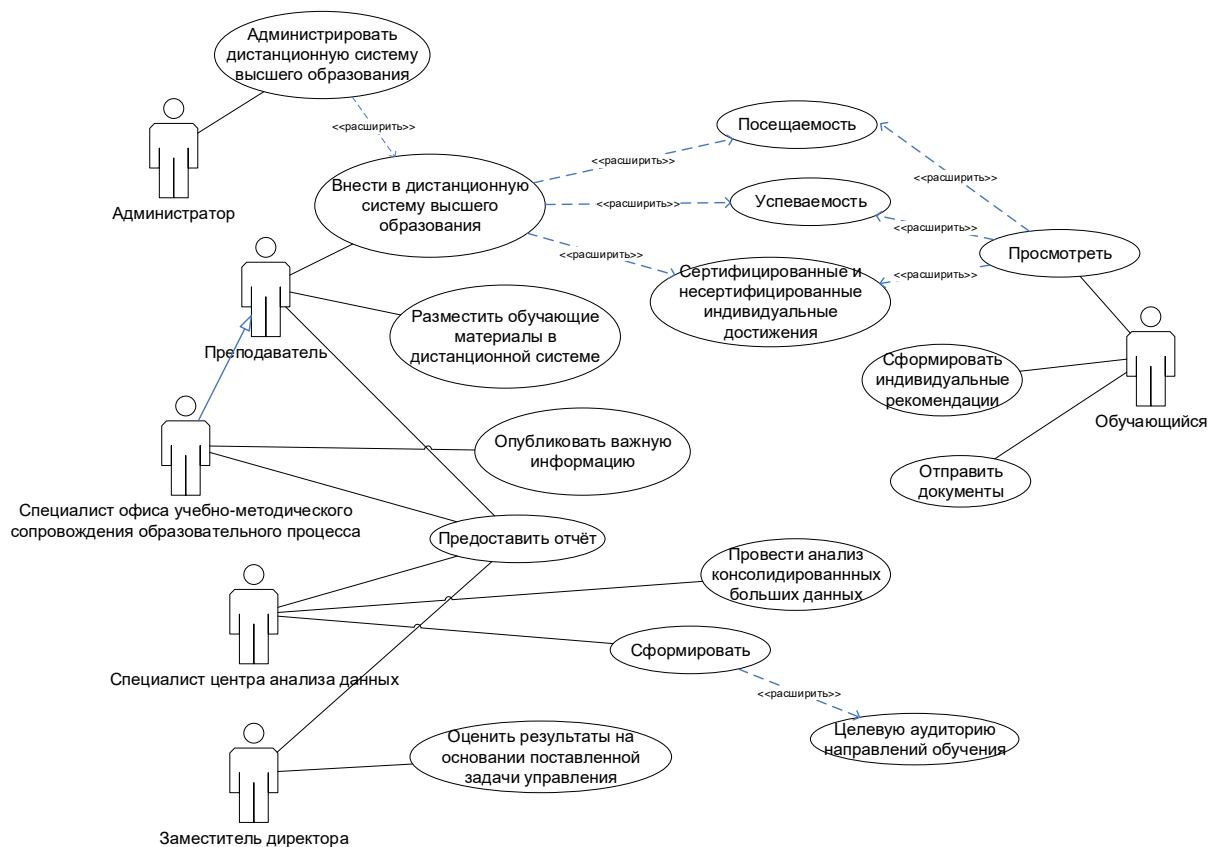


Рисунок 5 – Диаграмма вариантов использования

Схема процесса управления обработки больших данных в образовательной организации отображена на рисунках 6, 7.

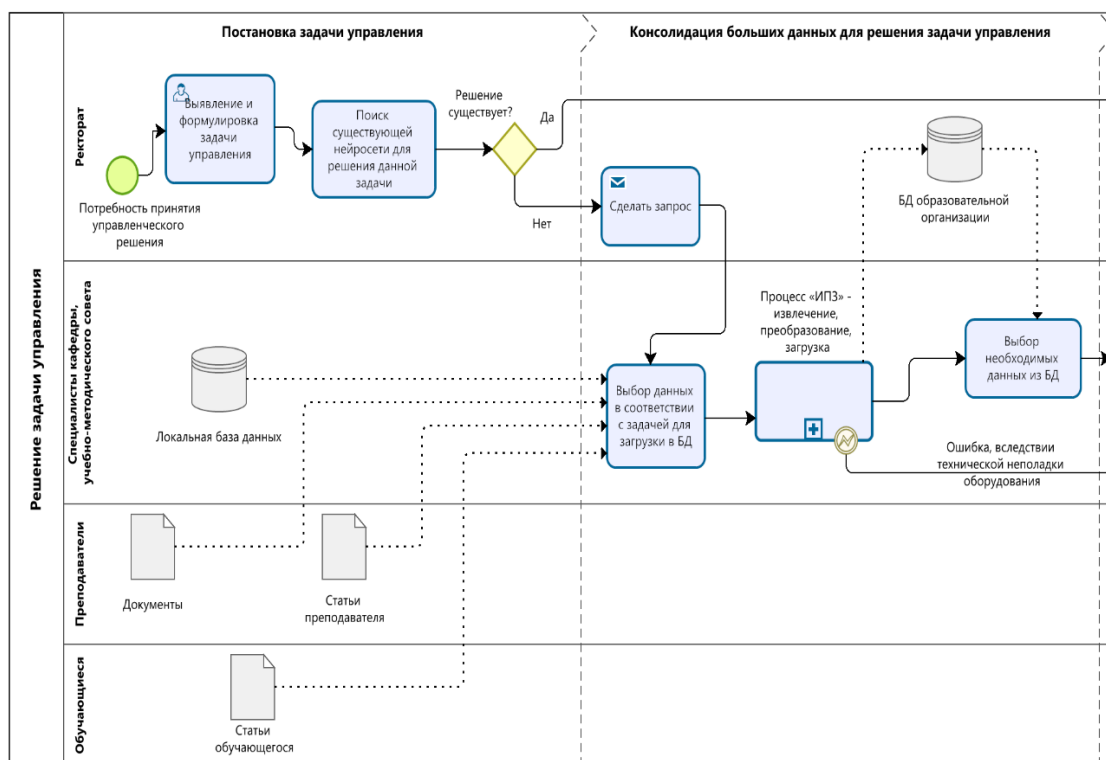


Рисунок 6 – Схема процесса решения задачи управления на этапах 1 и 2

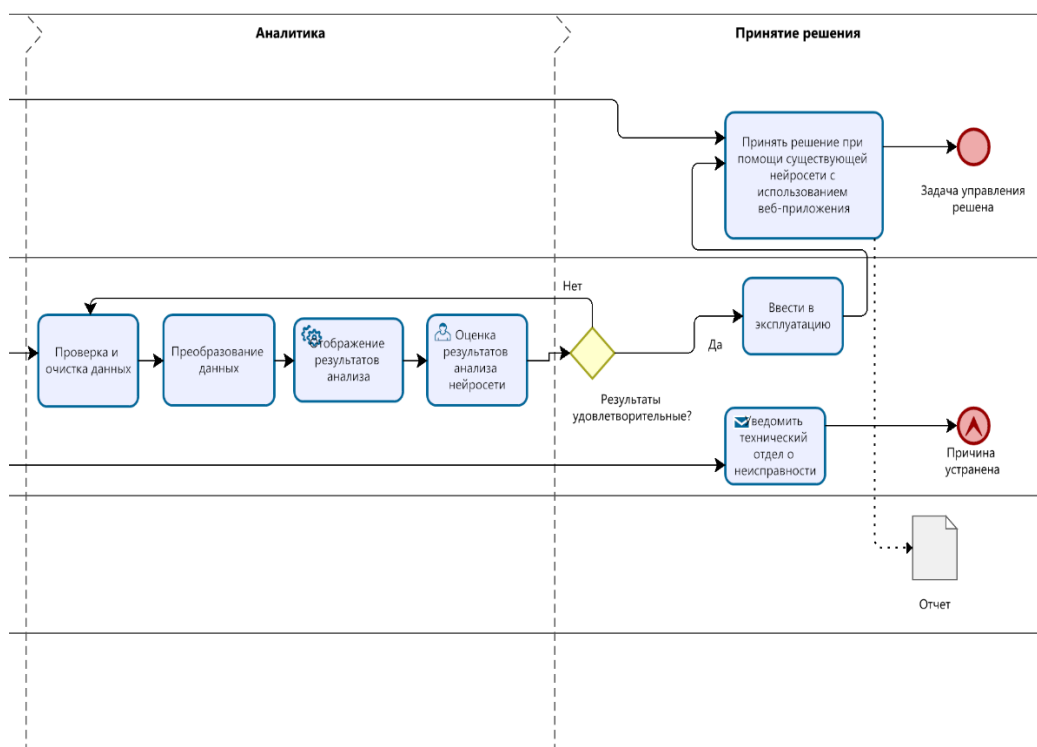


Рисунок 7 – Последние два этапа решения задачи управления

Если вычислительные мощности образовательной организации позволяют это сделать - то при поступлении новых данных они тут же должны проходить процесс «ИПЗ» - извлечения, преобразование и загрузки в БД, кроме того зачастую у локальной базы данных или хранилища или озера данных института есть срок хранения больших данных, после чего они

удаляются за ненадобностью и важно успеть вынуть из них именно актуальные данные и выгрузить в БД веб-приложения для дальнейшего анализа.

Была создана база данных, которая была дополнить данными обучающихся из «1С: Предприятие». После импорта данных база данных была усовершенствована. Схема базы данных, реализованная в Server Management Studio отображена на рисунке 8.

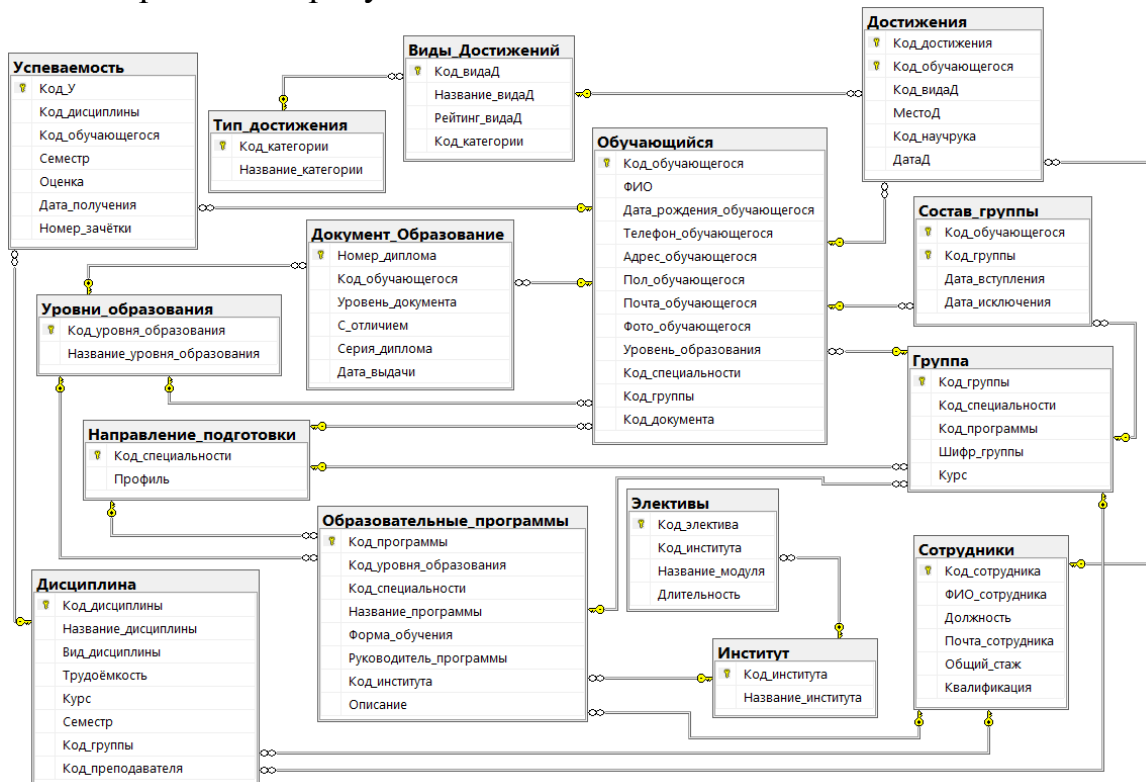


Рисунок 8 – Схема в MSSQL

В данной базе данных отображается динамика состава группы а уровни образования обозначаются цифрами: 0 – школа, 1 – колледж, 2 – бакалавриат, 3 – магистратура и так далее. Все типы достижений были вынесены в отдельную сущность «Тип_достижения» а все сущности видов достижений объединены в одну – «Виды_Достижений».

В среде «Microsoft Visual Studio» был создан веб-проект, ориентированный на использование фреймворка «Django», в результате одна из страниц в веб приложении выглядит следующим образом, показанным на рисунке 9.

Перейти к таблице обучающихся »

Справка об этой таблице

Show 25 entries Search:

Код_обучающегося	ФИО	Научные	Общественные	Учебные	Спортивные	Творческие
1	Абакумова Наталья Петровна	15	39	0	7	0
2	Абашин Алексей Антонович	6	0	13	13	9
3	Авдеев Александр Александрович	8	38	7	30	22
4	Авдеев Сергей Сергеевич	0	17	18	48	17
5	Авдеев Евгений Александрович	20	0	9	16	5
6	Адиловских Артур Петрович	6	31	0	5	0
7	Аксаненко Светлана Владимировна	15	16	2	8	14
8	Андреев Антон Валерьевич	24	23	9	16	1
9	Антиопова Лена Евгеньевна	22	33	5	10	9
10	Антоненко Виктория Константиновна	17	31	0	2	7

Рисунок 9 –Отображение рейтинга по категориям в веб-приложении

Классами являются направления магистратуры, доступные в МГПУ; в качестве же набора признаков выбраны все достижения обучающегося, а не только учебные достижения.

Данные об обучающихся из базы данных в свете особенностей этих данных проще и надежнее всего классифицировать комплексом мощным аналитических инструментов: мета-алгоритмами деревьев решений, нейросетью в виде многослойного персептрона. Сегментацию, которая будет выполнена вследствие классификации, можно использовать для того, чтобы понять, какие специальности более востребованы, целесообразность открытия профильных направленностей и определение целевой аудитории приёмной компании разнопрофильных программ высшего профессионального образования. То есть выявить связь успеваемость по дисциплинам, научных успехов, прочих достижений с поступлением на определенный профиль. Лицо принимающее решение может планировать количество групп, преподавательский состав на следующий год исходя из спроса на определенные программы.

О спросе важно знать как можно раньше для успешного выполнения всех функций управления. По результатам исследования больших данных об обучающихся можно понять - нужно ли дополнительно привлечь преподавателей или наоборот, спрос маленький, набрать группу по определенному направлению не получается, ведь пару человек обучать не рентабельно. В таком случае верным управленческим решением будет не открывать эту программу, но открыть новую, зато многочисленную - для кластера студентов, максимально близкую к их интересам и по этим интересам объединяющую этих студентов.

Предложенная рекомендация обучающемуся по направлению обучения содержит также рекомендацию *элективных курсов* либо *факультативов* на

основании выявленных интересов обучающихся, что было реализовано в веб-приложении как видно на рисунке 10.

*** Результат анализа данных обучающегося***
Для обучающегося с

Кодом обучающегося	ФИО
1	Абакумова Наталья Петровна

Рекомендуется следующая магистратура:

Код специальности	Форма обучения	Описание	Рекомендуемые элективные курсы
90403	очная	Программа направлена на подготовку разработчиков, аналитиков, исследователей, руководителей и менеджеров ИТ-проектов, способных развивать технологии создания, трансформации и внедрения цифровых образовательных систем.	Большие данные в образовании: управление и аналитика, Чат-боты в образовании

Разработка цифровых образовательных систем

Рисунок 10 – Индивидуальные рекомендации для обучающегося

Также в приложение были три мета-алгоритма – которые предсказывают количество обучающихся на каждом направлении и дерево решений (для выявления целевой аудитории оно является лучшим вариантом за счёт своей интерпретируемости), для выявления целевой аудитории по направлениям – как для организации в целом, так и для каждого института в отдельности, примеры отображаемого контента на веб-страницах приведены на рисунках 11, 12.

Веб-приложение

Анализ данных нейросетью, деревом решений, мета-алгоритмами

Организации »

Выбрать обучающегося для анализа »

Выбрать институт в организации для анализа »

Достижения

Рейтинг достижений »

Список обучающихся также доступен на отдельной странице

Обучающиеся »

© 2024 - Приложение обработки данных об обучающихся

Рисунок 11 – Главная страница веб-приложения

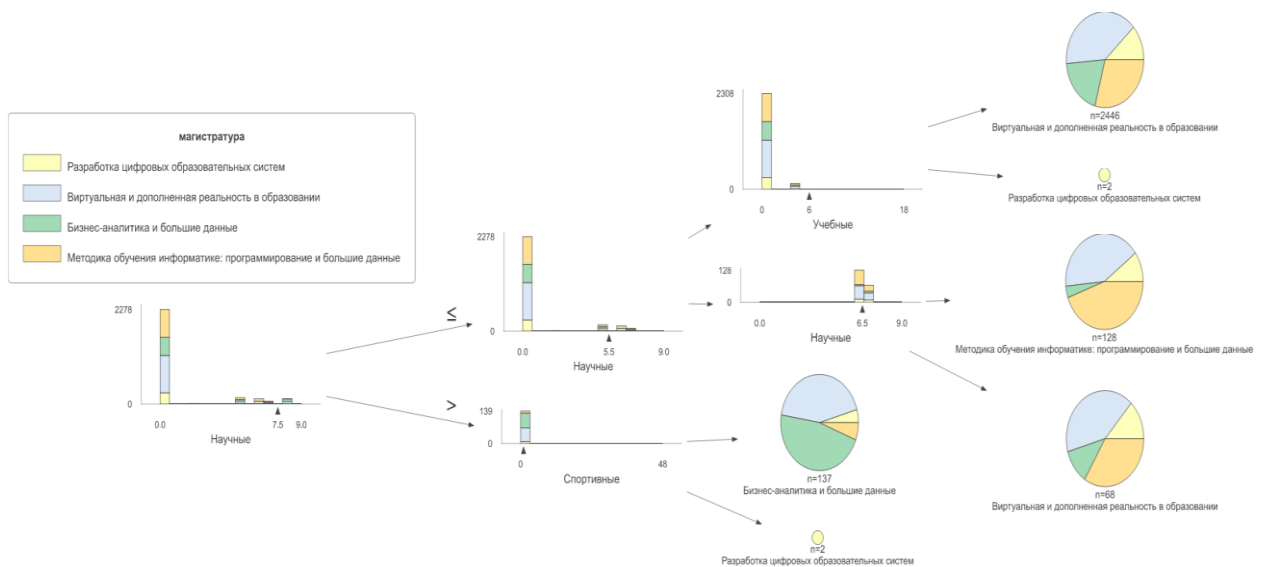


Рисунок 12 – Пример дерева решений для распределения целевой аудитории по достижениям для направлений обучения по институту цифрового образования

Полученный результат от нескольких классификаторов, замеряемый стандартными методами, был проанализирован несколькими метриками, результат отображен на рисунке 13.

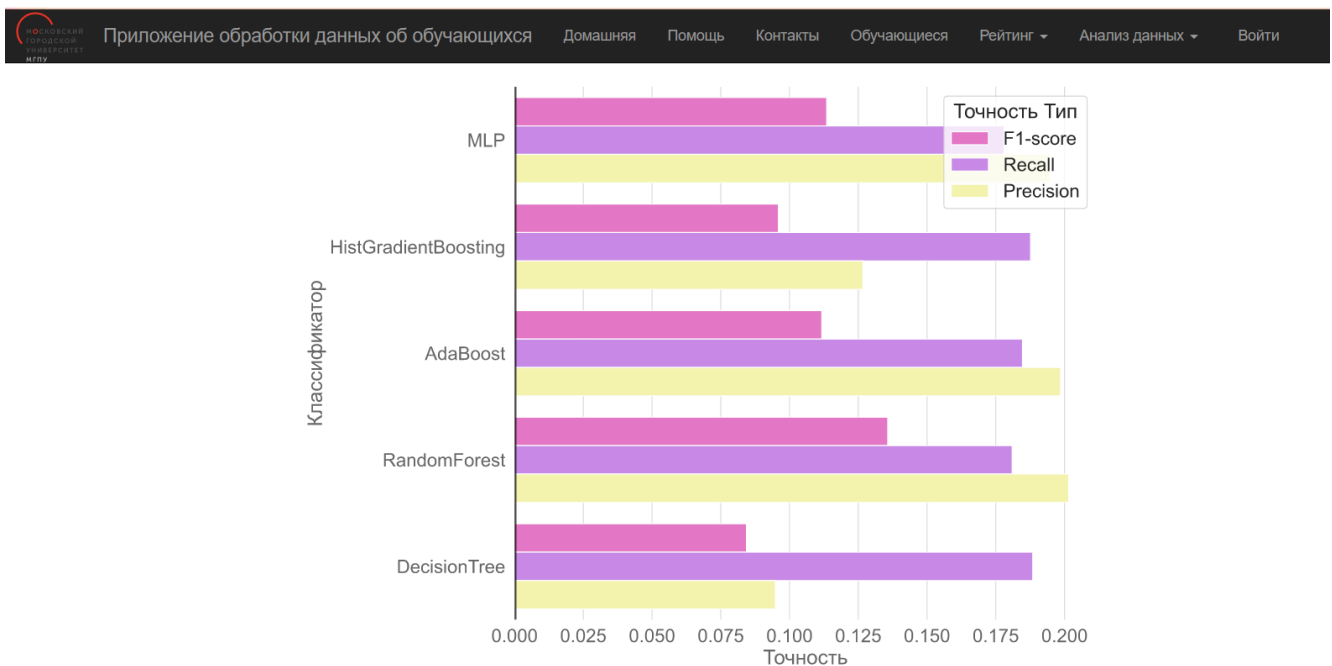


Рисунок 13 – Диаграмма оценки качества классификации

Замеры точности необходимы вследствие того, что большие данные, которые со временем после обработки подгружаются в БД для анализа, создают такую ситуацию, когда неизвестно до окончания анализа какой классификатор справится лучше с учетом появления новых, не анализированных ранее данных. Потому имеет смысл каждый раз проводить анализ данных несколькими классификаторами и на основании их оценки при

помощи метрик опираться в дальнейшем на анализ того классификатора, чьи результаты оказались наиболее точными.

В четвертой главе было проведено моделирование процессов принятия управленческих решений в образовательной организации на основе обработки больших данных.

Любой сценарий начинается с загрузки данных. В программном продукте «Loginom» можно загрузить данные как из SSMS, так и из 1С: Предприятие. Имитационная модель разработанного алгоритма и программы автоматизации процесса оценки эффективности в продукте «Loginom», реализованная в меру возможностей программного средства представлена на рисунке 14.

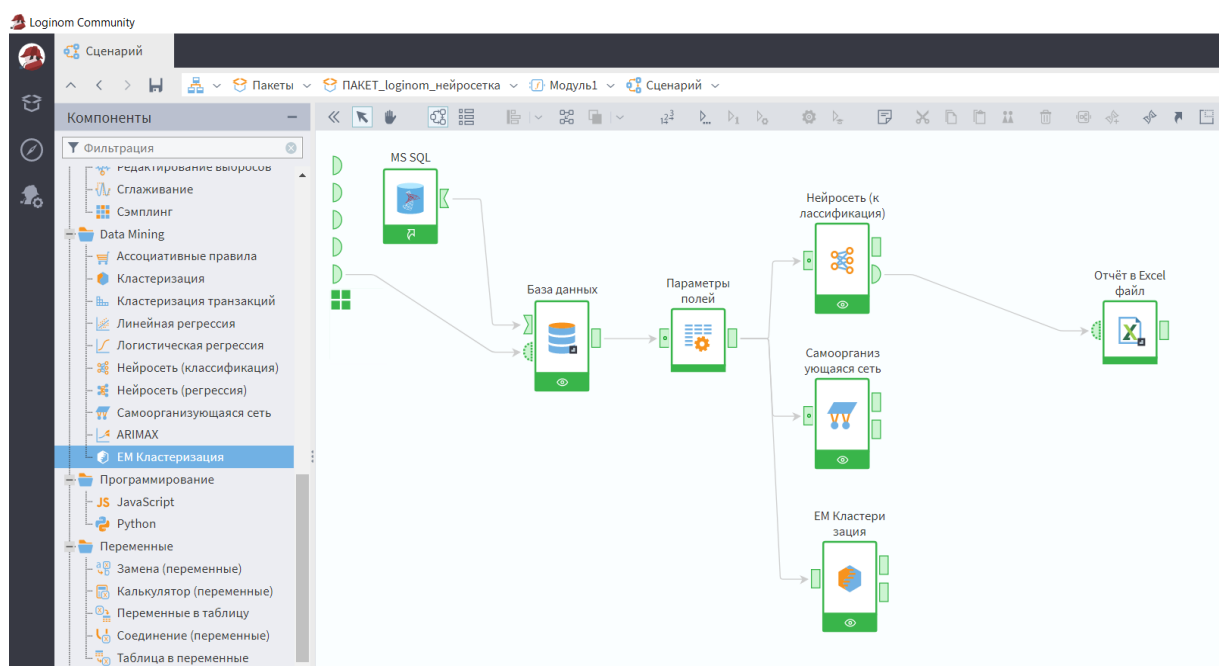


Рисунок 14 – Сценарий обработки данных

Подгрузить данные можно из разных источников, что без сомнения является плюсом, однако большим минусом оказалось невозможность построения ни дерева решений ни мета алгоритмов без создания подмодели с тем же программным кодом.

В работе описана методика применения веб-приложения для задач управления в сетевых дистанционных системах высшего образования, в которой, в частности, проводится опрос по составленной анкете про дисциплины, интересы, достижения чтобы дополнить существующие данные.

Предложена следующая математическая модель оценки эффективности обработки и хранения данных в образовательной организации (2):

$$P = \int_{t_0}^{t_1} \frac{D(t)}{K \cdot C(t) \cdot S} dt, \quad (2)$$

где P - эффективность обработки и хранения данных, D – функция объема обрабатываемых и хранимых данных в килобайт в зависимости от времени, C – функция стоимости хранения данных в единицу времени в рублях за месяц

в зависимости от времени, S - скорость обработки данных в единицу времени в килобайтах в секунду, t_0 – начало учебного года, t_1 – конец учебного года, K - сумма коэффициентов K_i - коэффициент разнообразия поступающих в БД источников данных, K_v - отражающий частоту и объем изменений в БД, K_a - коэффициент отражающий применение алгоритмов анализа данных: $K_a = \alpha_0 \circ \alpha_1 \circ \dots \circ \alpha_E$, E – все алгоритмы, а α_E – это классификации обучающихся (например, последовательным объединением неглубоких деревьев или многослойной нейронной сетью обратного распространения).

Эта модель позволяет оценить эффективность использования ресурсов обработки и хранения данных в образовательной организации, то есть чем больше значение P , тем эффективнее используются ресурсы.

В заключении описываются реализованные задачи диссертационного исследования.

В приложениях приведена анкета и фрагмент программы.

ЗАКЛЮЧЕНИЕ

Были получены следующие основные результаты:

1. За счёт анализа организационно-структурных моделей, информационных потоков и процессов в высшей образовательной организации по обеспечению управления сетевой дистанционной системой было выяснена какая реорганизации учебного процесса была произведена для внедрения сетевой дистанционной системы высшего образования, в результате была выявлена потребность в анализе больших данных, поступающих, в частности, от данной системы.

2. Выполнен анализ существующих актуальных методов оценки деятельности дистанционных систем высшего образования для задач управления, используемых критериев оценки, особенностей управления процессов обработки больших данных, которые можно использовать для создания программного обеспечения.

3. Выполнен анализ существующих алгоритмов и программного обеспечения прогнозирования показателей эффективности образовательной организации за счёт чего была выявлена необходимость создания таких алгоритмов и программного обеспечения, которые бы более эффективно обрабатывали бы большие данные об обучающихся.

4. Выполнен сравнительный анализ существующих методов и алгоритмов анализа данных для решения задачи классификации и оценки качества решения задачи классификации обучающихся, которые в дальнейшем были реализованы в программном обеспечении.

5. Разработан алгоритм оценки и анализа достижений для повышения эффективности управленческих решений в обобщенном для образовательных организаций виде, поддерживающих непосредственную реализацию идеи приводить разнообразные данные, получаемые из больших данных, к однородному формату упрощая дальнейший анализ.

6. Разработана модель функционирования дистанционной системы

высшего образования, дополненной программным обеспечением анализа больших данных, демонстрирующая требуемый функционал программного обеспечения.

7. Разработано и реализовано программное обеспечение процессов анализа данных для задач управления в виде веб-приложения, предназначенное для априорной оценки востребованности специальностей, целесообразности открытия профильных направлений и определения целевой аудитории приёмной кампании разнопрофильных программ высшего профессионального образования.

8. Произведено моделирование и реализация программного обеспечения оценок эффективности для задач принятия решения в социально-экономических системах в области образования за счёт чего было проведено сравнение разработанного программного продукта с «Loginom» на практике.

9. Разработана методика применения веб-приложения для задач управления в сетевых дистанционных системах высшего образования, позволяющее практически любой организации высшего образования внедрить в том или ином виде вышеописанные алгоритмы и программное обеспечение процессов обработки больших данных для задач управления.

ОСНОВНЫЕ ПУБЛИКАЦИИ ПО ТЕМЕ ДИССЕРТАЦИИ

Статьи в российских рецензируемых научных журналах, рекомендованных ВАК России для опубликования основных научных результатов:

1. Белякова, А.В Актуальные вопросы разработки нейросетевых модулей прогнозирования успеваемости обучающихся высших учебных заведений/ Белякова А.В., Ромашкова О.Н., Ермакова Т.Н., Чискидов С.В. // Современная наука: актуальные проблемы теории и практики. Серия: Естественные и технические науки. 2024. № 3-2. С. 33-38.

2. Белякова, А.В Методика анализа больших данных для задач управления в дистанционных системах высшего образования / Белякова А.В., Ромашкова О.Н. // Современная наука: актуальные проблемы теории и практики. Серия: Естественные и технические науки. 2023. № 2-2. С. 61-66.

3. Белякова, А.В Разработка требований к процессам обработки данных для задач управления в дистанционных системах высшего образования / Белякова А.В., Ромашкова О.Н., Ермакова Т.Н., Чискидов С.В. // Современная наука: актуальные проблемы теории и практики. Серия: Естественные и технические науки. 2021. № 12. С. 66-72.

Публикации, включённые в ядро РИНЦ:

1. Belyakova, A.V. Information model of data management in network online educational systems / Romashkova O.N., Belyakova A.V., Ponomareva L.A. // В сборнике: Proceedings of the 2021 IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering, ElConRus 2021 С. 2226-2229.

Свидетельства о регистрации программы для ЭВМ:

1. Белякова, А.В Программа для автоматизированного управления рейтинговыми показателями вузов / Белякова А.В., Пономарева Л.А., Чискидов С.В., Василюк И.П. // Свидетельство о регистрации программы для ЭВМ RU 2019611874, 05.02.2019. Заявка № 2018664941 от 24.12.2018.

2. Белякова, А.В Программа для автоматизированного управления рейтингом вуза / Чискидов С.В., Пономарева Л.А., Белякова А.В., Василюк И.П. // Свидетельство о регистрации программы для ЭВМ RU 2019613313, 13.03.2019. Заявка № 2019610625 от 24.01.2019.

Публикации в других журналах, сборниках научных трудов и материалах научных и научно-практических конференций:

1. Белякова, А.В Актуальные вопросы перспектив автоматизации использования больших данных для управления в образовании // В сборнике: Лига исследователей МГПУ: в 3 частях. – Москва, 2023. – С. 27-33.

2. Белякова, А.В Анализ возможностей интеграции нейросетей в программном обеспечении для задач управления // В книге: Сборник тезисов студенческой открытой конференции. – Москва, 2022. – С. 317-319.

3. Белякова, А.В Моделирование схем процессов обработки данных для задач управления в системах высшего образования // В сборнике: Открытая наука 2021. Сборник материалов научной конференции с международным участием. Москва, 2021. С. 89-91.

4. Белякова, А.В Моделирование схемы процесса получения рекомендаций нейросети для задач управления в системах высшего образования // В книге: Международная молодежная научная школа-конференция "Цифровая трансформация реального сектора экономики". Сборник тезисов докладов. Москва, 2021. С. 11-13.

5. Белякова, А.В Особенности разработки web-приложения для сетевых дистанционных систем высшего образования // В книге: #ScienceJuice2021. Сборник тезисов студенческой открытой конференции. Москва, 2021. С. 395-396.

6. Белякова, А.В Система прогнозирования рейтинга обучающихся с применением технологии глубокого обучения // В книге: #ScienceJuice2020. сборник статей и тезисов студенческой открытой онлайн-конференции. Москва, 2021. С. 53-54.

7. Белякова, А.В Информационная модель оценки личных достижений обучающихся // В сборнике: Информационные технологии в науке, бизнесе и образовании. сборник трудов XI Международной научно-практической конференции студентов, аспирантов и молодых ученых. 2020. С. 35-40.

8. Белякова, А.В Новый подход к анализу достижений обучающихся образовательных организаций // В сборнике: Математика и информатика в образовании и бизнесе. Сборник материалов международной

научно-практической конференции. 2020. С. 68-73.

9. Белякова, А.В. Перспективное прогнозирование личных достижений обучающихся образовательных организаций // В сборнике: #ScienceJuice2019. Сборник статей и тезисов студенческой открытой конференции. 2020. С. 351-356.

10. Белякова, А.В. Автоматизация процесса многокритериального ранжирования студентов с помощью электронного портфолио / Белякова А.В., Пономарева Л.А., Ромашкова О.Н., Заболотникова В.С. // Вестник Донского государственного технического университета. 2019. Т. 19. № 4. С. 382-388.

11. Белякова, А.В. Математическая модель оценки качества образовательного процесса / Белякова А.В., Пономарева Л.А., Ромашкова О.Н. // В книге: Новые информационные технологии в научных исследованиях. Материалы XXIV Всероссийской научно-технической конференции студентов, молодых ученых и специалистов. 2019. С. 43-44.

12. Белякова, А.В. Прототип информационной системы оценки качества учебного процесса в образовательной организации / Белякова А.В., Пономарева Л.А., Чискидов С.В. // В книге: Новые информационные технологии в научных исследованиях. Материалы XXIV Всероссийской научно-технической конференции студентов, молодых ученых и специалистов. 2019. С. 45-46.

13. Белякова, А.В. Новый подход к ранжированию образовательных организаций с применением электронного портфолио / Белякова А.В., Пономарева Л.А., Василюк И.П. // В сборнике: Новая наука: новые вызовы. II Международная научно-практическая конференция. 2018. С. 29-33.

14. Белякова, А.В. Разработка прототипа блока информационной системы для портфолио обучающихся в образовательной организации / Пономарева Л.А., Белякова А.В. // В сборнике: Наука и образование в XXI веке. сборник научных трудов по материалам международной научно-практической конференции. 2018. С. 42-45.

Белякова Анна Вячеславовна

**АЛГОРИТМЫ И ПРОГРАММНОЕ ОБЕСПЕЧЕНИЕ ПРОЦЕССОВ
ОБРАБОТКИ БОЛЬШИХ ДАННЫХ ДЛЯ ЗАДАЧ УПРАВЛЕНИЯ В
СЕТЕВЫХ ДИСТАНЦИОННЫХ СИСТЕМАХ ВЫСШЕГО
ОБРАЗОВАНИЯ**

09.06.01 – Информатика и вычислительная техника

Управление в социальных и экономических системах

Научный доклад

об основных результатах научно-квалификационной работы
(диссертации)