

## АННОТАЦИЯ

Выпускная работа бакалавра выполнена на тему «Разработка рекомендательной системы для веб-платформы взаимодействия работодателей со студентами».

Дипломная работа состоит из введения, трех содержательных разделов и заключения.

Во введении определена актуальность выбранной темы, также поставлена цель и определены основные задачи, сформулированные для достижения поставленной цели.

В первом разделе рассматриваются характерные особенности взаимодействия работодателей и студентов на российском рынке труда, описываются основные проблемы в этой сфере, проводится сравнительный анализ по сфере трудоустройства студентов и определены выводы по разделу.

Второй раздел посвящен разработке рекомендательной системы, где исследуются существующие решения в области рекомендательных систем и демонстрируется поэтапное решение DataScience проекта.

В третьем разделе предлагается улучшение полученной модели, описывается платформа взаимодействия работодателей со студентами, а также приведены оценки работы полученной рекомендательной системы.

В заключении приведены основные выводы и результаты по проделанной работе.

Работа выполнена на 73 страницах, содержит 31 рисунков, 3 таблицы, 8 формул, список литературы из 35 источников и 2 приложения.

## ОГЛАВЛЕНИЕ

АННОТАЦИЯ .....	4
ВВЕДЕНИЕ .....	6
ГЛАВА 1. ТЕОРЕТИЧЕСКИЕ ОСНОВЫ ФУНКЦИОНИРОВАНИЯ РАНКА ТРУДА .....	9
1.1 Характерные особенности взаимодействия работодателей и студентов на российском рынке труда .....	9
1.2 Основные проблемы взаимодействия работодателей и студентов .....	16
1.3 Сравнительный анализ по сфере трудоустройства студентов .....	22
1.4 Вывод.....	30
ГЛАВА 2. РАЗРАБОТКА РЕКОМЕНДАТЕЛЬНОЙ СИСТЕМЫ.....	31
2.1 Исследование существующих рекомендательных систем .....	31
2.2 Работа с данными.....	35
2.2.1 Сбор данных.....	36
2.2.2 Предобработка и очистка данных .....	37
2.3 Baseline-модель .....	39
2.4 Разработка модели .....	40
2.4.1 Выбор метода .....	41
2.4.2 Обработка текста.....	43
2.4.3 Content-based-модель .....	46
2.5 Вывод.....	51
ГЛАВА 3. ПЕРСПЕКТИВЫ РАЗВИТИЯ И РЕЗУЛЬТАТЫ РАЗВИТИЯ СИСТЕМЫ .....	53
3.1 Улучшение модели .....	53
3.2 О платформе «POLYGON» .....	54
3.3 Оценка работы рекомендательной системы .....	63
ГЛАВА 4. РЕЗУЛЬТАТИВНОСТЬ .....	65
ЗАКЛЮЧЕНИЕ .....	66
СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ .....	68
Приложение 1 .....	72
Приложение 2 .....	73

## ВВЕДЕНИЕ

Современный рынок труда является динамичной и сложной системой, которая требует от выпускников образовательных учреждений высшего образования высочайших профессиональных и личностных качеств. Развитие информационных технологий и интернет-технологий привели к возникновению новых возможностей для поиска работы и взаимодействия работодателей со студентами.

Взаимодействие работодателей со студентами — это процесс, в рамках которого представители работодателей вступают в контакт со студентами для проведения собеседований, организации практик и стажировок, заключения договоров о трудоустройстве и решения других вопросов, связанных с трудоустройством.

Одной из форм взаимодействия работодателей со студентами на рынке труда является использование веб-платформ. Данный формат позволяет работодателям быстро и эффективно находить необходимых им кандидатов и упрощает процесс принятия на работу студентов. Веб-платформы, обеспечивающие взаимодействие работодателей со студентами, позволяют работодателям размещать вакансии на сайте и получать от студентов резюме и заявки на вакансии, а студентам искать вакансии и отправлять свои резюме на размещенные вакансии.

Такие веб-платформы могут предоставлять работодателям и студентам дополнительные инструменты для более эффективного взаимодействия, например, чаты для общения, системы онлайн-встреч, тесты для проверки знаний и умений кандидатов, системы обратной связи и др. Поэтому разработка рекомендательных систем для веб-платформ взаимодействия работодателей со студентами может помочь упростить и оптимизировать процесс поиска работы для студентов, а также для работодателей — помочь найти подходящих кандидатов на вакансии.

В целом, поиск работы во время просмотра списков вакансий на сайтах по подбору персонала, что действительно отнимает много времени и денег, является раздражающим занятием, хотя в большинстве случаев эти вакансии не всегда подходят пользователям или пользователи не удовлетворяют им.

Поступая таким образом, рекрутеры тратят впустую свое время, проверяя, квалифицированы они или нет. Этот тезис направлен на решение очень важной проблемы в процессе подбора персонала, которая заключается в сопоставлении соискателей с предложениями о работе. В настоящее время процесс согласования между кандидатом и предложения о работе – это одна из основных проблем, с которыми приходится сталкиваться компаниям. Отбор кандидатов в короткий список и проверка резюме – это трудоемкие задачи для компании, особенно когда от 80 до 90 процентов резюме, полученных на ту или иную должность, неквалифицированные.

Цель данной дипломной работы – разработка рекомендательной системы для веб-платформы взаимодействия работодателей со студентами. Данная система будет выступать в качестве инструмента, облегчающего процесс поиска кандидатов и упрощающего выбор студентов работодателями.

Для достижения поставленной цели необходимо рассмотреть характерные особенности взаимодействия работодателей и студентов на российском рынке труда, выявить основные проблемы в этой сфере и провести сравнительный анализ по сфере трудоустройства студентов, а также изучить существующие решения в области рекомендательных систем, предложить свою модель рекомендаций и продумать дальнейшее развитие и внедрение своей собственной рекомендательной системы.

В работе будет использоваться методология, включающая в себя анализ научной литературы, методы, а также инструменты статистического анализа данных.

Обзор существующих методов и моделей для построения рекомендательных систем позволит определить наиболее эффективные решения

и использовать их для разработки рекомендательной системы для веб-платформы взаимодействия работодателей со студентами.

# ГЛАВА 1. ТЕОРЕТИЧЕСКИЕ ОСНОВЫ ФУНКЦИОНИРОВАНИЯ РАНКА ТРУДА

## 1.1 Характерные особенности взаимодействия работодателей и студентов на российском рынке труда

Взаимодействие работодателей и студентов является актуальной темой на современном российском рынке труда. Несмотря на то, что в последние годы были предприняты многочисленные усилия по улучшению взаимодействия и тесному сотрудничеству между работодателями и студентами, остаются многие проблемы и сложности для решения.

Основные этапы взаимодействия между работодателем и студентом:

- Поиск работы: студенты ищут работу на различных интернет-платформах, обращаются в центры занятости, используют услуги рекрутерских агентств, а также непосредственно обращаются к работодателям.
- Собеседование и отбор кандидатов: работодатель проводит собеседования для оценки квалификации и профессиональных навыков студентов, а также определяет, насколько они подходят для занимаемой должности.
- Трудоустройство: после прохождения всех этапов собеседования и отбора, работодатель и студент подписывают трудовой договор и начинают сотрудничество.

					ВКР-40461806-10.12-09.03.03-462-18-2023							
		№ докум.	Подп.	Дата								
Разраб.	Щетинин В.В.								Лит.	Лист	Листов	
Пров.	Кетько Н.В.										9	22
Н.контр.	Кузнецов С. Ю.											
Утв.	Скитер Н. Н.											
					ВолгГТУ ЭП-462							

Взаимодействие работодателей и студентов на российском рынке труда имеет свои характерные особенности:

1. Низкая интеграция университетов и компаний. На российском рынке труда наблюдается недостаточная связь между университетами и компаниями. Студенты не получают достаточной информации о компаниях и о возможностях трудоустройства, а компании не имеют достаточной информации о качестве подготовки выпускников.
2. Низкий уровень профессиональной подготовки. Российские вузы не всегда обеспечивают достаточную подготовку студентов к профессиональной работе. Это приводит к тому, что многие выпускники не могут найти работу по своей специальности.
3. Низкий уровень мотивации студентов. В России студенты не всегда заинтересованы в получении профессиональной подготовки. Это связано с тем, что рынок труда не всегда предлагает высокооплачиваемые и интересные должности, что не мотивирует студентов на учебу.
4. Невысокий уровень знаний работодателей о студентах. Компании не всегда знают, каких выпускников они хотят видеть в своих рядах, какие навыки, знания и качества им нужны. Это приводит к тому, что компании не могут подобрать квалифицированных и мотивированных сотрудников.
5. Недостаток информации о рынке труда. Студенты и компании не всегда имеют достаточный доступ к информации о рынке труда, о возможностях развития и о новых тенденциях. Это мешает им принимать правильные решения в выборе профессии или в поиске перспективной компании для трудоустройства.

Одним из главных проблем взаимодействия работодателей и студентов на российском рынке труда является несоответствие профессиональных навыков студентов требованиям работодателей. Зачастую у студентов нет возможности

приобретать необходимые практические навыки в соответствующих областях и проводить стажировку у работодателей.

Другими проблемами являются:

- Отсутствие прозрачности при поиске работы и обмене информацией между работодателями и студентами.
- Низкий уровень зарплат и неустойчивость на рынке труда.
- Отсутствие понимания требований работодателей со стороны студентов и обратно.

Служба исследований крупнейшей российской платформы онлайн-рекрутинга hh.ru выяснила, как работодатели оценивают уровень профессионализма соискателей, которые претендуют на вакансии сразу после вуза, насколько важен диплом о высшем образовании при трудоустройстве, а также в выпускниках каких вузов больше всего нуждаются работодатели. Опрос проводился с 27 июля по 9 августа 2021 г. среди 109 российских работодателей.

Основная часть работодателей (93%) заявила, что уровень профессионализма выпускников, которые претендуют на вакансии, средний и ниже. Также опрос показал, что наличие высшего образования важно для 61% опрошенных работодателей. Но диплом определенного вуза важен лишь для 41% респондентов.

Топ-3 востребованных направлений вузов, в которых нуждаются работодатели, составляют технические, социально-экономические и строительные направления [25].

На что нужно обращать внимание при выборе вуза, работают ли соискатели по специальности. Опрос проводился службой исследований hh.ru с 06 марта по 09 марта 2023 года среди 663 российских соискателей.





Рисунок 1.1 – Важность для работодателей

При выборе вуза соискатели с высшим образованием предлагают обращать внимание на наличие факультета или программы по желаемой специальности (75%), качество обучение (68%), квалификация преподавателей (47%), место вуза в рейтингах учебных заведений (44%), престижность вуза (41%) (Рис. 1.1).



Рисунок 1.2 – Работа по специальности

Говоря о работе по специальности, у 72% респондентов с высшим образованием был опыт работы, однако лишь 47% продолжают работать или искать работу по специальности, 19% - сейчас работают или ищут работу не по специальности, 6% - работают или ищут работу по смежной специальности. Также у 11% не было опыта работы по специальности, однако сейчас они ищут работу по тому направлению, на котором учились в вузе (Рис. 1.2).



Рисунок 1.3 – Работали по специальности

Среди тех, кто не работал/перестал работать по специальности, в качестве барьера указали сложность в поиске работы по специальности (36%), маленькая зарплата (35%), смутные перспективы карьерного роста (28%), «это не мое» - 26%, 18% - плохие условия работы (Рис 1.3) [23].

Рыночные тенденции также влияют на взаимодействие работодателей и студентов. Сегодня на рынке труда наблюдается тенденция к сокращению числа вакансий и повышению требований к кандидатам, в том числе к уровню образования и профессиональным навыкам. Это приводит к тому, что работодатели все более предпочитают опытных специалистов и не готовы рисковать с непроверенными кадрами.

Минобрнауки России активно работает над содействием занятости выпускников. С конца апреля 2022 года при министерстве заработал Экспертный центр карьеры и реализации профессиональных возможностей молодежи, который проводит бесплатное обучение для руководителей и сотрудников центров карьеры в вузах.

Показатели трудоустройства за 2021 года по сравнению с 2020 годом выросли больше всего выросли в Ленинградской и Брянской областях (на 13% в обоих регионах), в Курганской области и Камчатском крае (на 11% в обоих субъектах) и в Смоленской, Липецкой областях и Ставропольском крае (на 10% в каждом регионе).

Руководители российских вузов в регионах, где за год выросли показатели трудоустройства, отмечают, что в последнее время серьезно реформативирована работа с выпускниками. Сейчас она сосредоточена в Центре трудоустройства и развития карьеры вуза – это позволяет сделать ее системной и планомерно отслеживать результаты деятельности и ситуацию на рынке труда.

Взаимодействие университета с работодателями сейчас выходит на новый уровень, так как обучение становится все более практикоориентированным. В вузе проводятся различные чемпионаты, хакатоны, классические ярмарки вакансий, а также карьерные мастерские – сообщества молодых людей, уже более опытных в вопросах трудоустройства, которые помогают своим однокурсникам адаптироваться на рынке труда и найти работу. Также многие университеты имеют серьезные связи с индустриальными партнерами: это позволяет студентам в рамках обучения решать реальные практические кейсы действующих производств в тесной связке с сотрудниками компаний. Благодаря этому многие студенты получают предложения о трудоустройстве задолго до получения диплома. Создание базовых кафедр совместно с ключевыми партнерами, работодатели привлекаются к разработке образовательных программ, участвуют в создании кадровых резервов из числа обучающихся и так далее. Одна из главных задач при взаимодействии студентов и работодателей – это познакомить студентов с особенностями работы разных компаний

определенной отрасли. Чтобы студент еще на первом курсе смог получить понимание о том, где и кем он хочет работать по окончании обучения.

О значимости сотрудничества вузов и работодателей говорят и представители компаний. По словам Татьяны Авдеевой, бизнес-партнера по управлению персоналом Ставропольского отделения Юго-Западного банка Сбербанка, потенциальному нанимателю важно, чтобы студент, получая статус выпускника, обладал не просто багажом накопленных теоретических знаний, но и был погружен в профессиональный контекст, создавая серьезную конкуренцию на рынке труда.

«Именно поэтому Сбер заключил соглашения с ключевыми учебными заведениями Ставрополья, в рамках которых студенты, которые рассматривают Сбер как потенциальное место работы, могут попасть на практику или стажировку в крупнейший банк страны. Помимо этого, благодаря плотной работе с Центрами развития карьеры и молодежного предпринимательства учебных заведений в регионе растет число совместных проектов, а эксперты Сбера регулярно выступают на образовательных мероприятиях и делятся опытом с подрастающим поколением специалистов», – отмечает Татьяна Авдеева.

В Госкорпорации «Росатом» также уделяют большое внимание привлечению талантливых студентов на работу. Ежегодно в атомную отрасль трудоустраиваются около 2 тыс. выпускников вузов и колледжей. Как отмечает Юлия Ужакина, генеральный директор Корпоративной Академии Росатома, подбор потенциальных соискателей проводится путем проведения собственных мероприятий в вузах, участия в федеральных проектах для школьников и студентов и поиска новых форматов, таких как открытие диджитал-центров в ННГУ и ОЦ «Сириус». Также Росатом ежегодно организует стажировки и практики для более 5 тыс. студентов, так как невозможно понять специфику работы, не попробовав в нее погрузиться.

В марте 2022 года Минобрнауки России запустило Экспертный центр, в задачи которого входит трансформация центров карьеры в университетах. Он

осуществляет аналитическое, методологическое, экспертное и информационно-техническое сопровождение центров карьеры, а также координировать вопросы их цифровой трансформации, включая систему мониторинга, сбора и анализа данных [26].

В целом, характерные особенности взаимодействия работодателей и студентов на российском рынке труда показывают, что необходимы новые подходы и инструменты для решения сложностей и улучшения сотрудничества между сторонами. Одним из таких подходов может быть использование рекомендательных систем веб-платформы взаимодействия работодателей со студентами.

## 1.2 Основные проблемы взаимодействия работодателей и студентов

В настоящее время все более актуальными становятся проблемы взаимодействия работодателей и выпускников вуза, в частности по вопросам трудоустройства последних. По окончании высшего учебного заведения молодые выпускники должны рано или поздно устроиться на работу. Главный недостаток таких специалистов – отсутствие необходимого профессионального опыта, что является основной причиной их низкой конкурентоспособности на рынке труда по сравнению со специалистами, имеющими соответствующий опыт работы. Таким образом, сложность трудоустройства молодых выпускников связана с некоторыми факторами, например, такими, как:

- несоответствие уровня квалификации выпускников требованиям работодателей, недостаток опыта принятия самостоятельных решений, нехватка знаний в конкретной предметной области, отсутствие практических навыков по полученной специальности;
- не совсем верное представление молодых специалистов о путях адаптации в сфере труда и занятости.

Если рассматривать проблему взаимодействия с другой стороны, то работодатели часто ищут молодых сотрудников с определёнными нужными им знаниями и навыками, таким образом, и для них поиск может затянуться на

длительное время. И если крупные предприятия готовы взять молодого специалиста без некоторых навыков и обучить его, то небольшие компании не имеют такой возможности, и им необходимы сотрудники, уже владеющие определенными навыками и компетенциями.

Сложность трудоустройства молодых людей, только окончивших вуз, может привести к негативным последствиям, как в сфере экономики, так и в сфере социальной жизни общества [21].

Проблема дефицита кадров является глобальной. По данным международной исследовательской компании GlobalData, в 2022 г. во всем мире число работодателей, испытывающих нехватку рабочей силы, увеличилось по сравнению с 2021 г. на 28%.

Российские компании сталкиваются с такими же трудностями. Согласно исследованию лаборатории конъюнктурных опросов Института экономической политики, им. Гайдара, в октябре 2022 г. баланс оценок промышленными предприятиями достаточности персонала упал до минимального показателя за всю историю наблюдения с апреля 1996 г., пишет РБК. В первом полугодии 2022 г. российским промышленным предприятиям больше всего не хватало главных инженеров, технологов и механиков, пишут «Известия» со ссылкой на исследование сервиса по поиску работы Superjob. Чаще всего дефицит кадров наблюдался в таких отраслях, как металлургия, судостроение и радиоэлектроника, и таких регионах, как Сибирь, Урал и Дальний Восток, отмечается в исследовании.

Взаимодействие с учебными заведениями помогает бизнесу получить подготовленные кадры и избежать трат на переподготовку выпускников ссузов и вузов [24].

Кадровый дефицит – одна из острейших проблем современного бизнеса. Такие результаты показали исследования, проведенные Российским союзом промышленников и предпринимателей в 2016 – 2020 годах. И даже при сокращении количества рабочих мест дефицит рабочей силы по-прежнему остается. Источник проблем – разрыв между ожиданиями работодателей и тем

уровнем компетенций, с которыми выходят на рынок труда выпускники колледжей и вузов. Профессиональное образование сегодня не может развиваться как закрытая система, без учета потребностей работодателей. Диалог между сферой образования и сферой труда должен строиться на основании четких взаимных обязательств. Бизнес формулирует для образования своевременный заказ на кадры с определенными квалификациями и требования к содержанию их обучения, система образования отвечает на запрос качественными образовательными программами и соответствующими ожиданиям работодателей результатами подготовки выпускников.

Не случайно Стратегией развития национальной системы квалификаций до 2030 года предусмотрено системное вовлечение работодателей в управление качеством реализации программ профессионального образования и обучения. Должны быть сформированы условия, обеспечивающие взаимопонимание и, главное, - взаимодействие работодателей и образовательных организаций [22].



Рисунок 1.4 – Востребованность молодых специалистов по отраслям

Самые популярные вакансии от работодателей для молодых специалистов в большинстве – это не всегда вакансии, связанные с их профессией, так из-за нужды в деньгах или просто не знания о возможностях в сфере полученного

образования, молодые специалисты часто идут работать не по профессии. Пять самых популярных предложений для специалистов без опыта на сайте hh.ru: менеджер по продажам, ученик на пищевое производство, продавец-консультант, курьер и оператор колл-центра (Рис. 1.4).



Рисунок 1.5 – Молодые специалисты, которых ищут работодатели

Во все эти специальности низкий порог входа (рис. 1.5): работодатели не требуют обязательное наличие опыта и высшего образования, базовым вещам можно научиться уже на месте. Многие выпускники вузов мечтают сразу же после получения диплома устроиться на престижную и высокооплачиваемую работу, но это удастся лишь единицам. Обычно — тем, кто еще во время учебы думал о необходимости опыта работы.

При отсутствии опыта есть два пути поиска постоянной работы. Первый — искать вакансии, для которых не нужен опыт. Второй путь — получить опыт. Опыт можно приобрести в лагере летом, на стажировке, в университете, по волонтерской программе и т.д. [35].

Основные проблемы взаимодействия работодателей и студентов на российском рынке труда могут быть следующими:



1. Неточность описания требований к кандидатам и вакансиям. Работодатели не всегда указывают все необходимые требования к соискателям, что сложно для студентов-выпускников понимать, какие навыки они должны приобрести для успешной карьеры;
2. Недостаток информации о рынке труда. Также часто случается, что студентам не хватает объективных данных о том, какой специальности наиболее востребованы, какие компании предлагают лучшие условия труда и т.д.;
3. Неэффективность процесса подбора кандидатов для работодателей, а также поиск работы для выпускников. Процесс подбора должен быть прозрачным, чтобы с одной стороны - работодатель мог быстро и эффективно найти подходящего кандидата для своей вакансии, и с другой – студенты могли бы быстро и эффективно найти работу с учетом своих предпочтений и интересов;
4. Недостаток своевременной обратной связи. Работодатели нередко не отвечают на резюме студентов, что может разочаровать и вызвать недоверие к процессу подбора. Также, выпускникам не всегда сообщают об успехе или неудачах в их поиске работы, что было бы полезно для дальнейшей карьеры;
5. Ограниченное количество каналов взаимодействия. Студенты и работодатели могут столкнуться с проблемой выбора между различными онлайн-платформами, вакансиями от посредников или не прямой связи с компаниями. Это затрудняет и упрощает интерактивность процесса поиска работы.

Анализ трудностей в формировании профессиональных компетенций студентов. Современный работодатель требует от своих будущих сотрудников не только знаний по своей профессии, но и умения применять эти знания на практике. Однако студентам часто не хватает опыта работы, чтобы успешно справляться с задачами на предприятии. Кроме того, некоторые профессии, такие как IT и медицина, быстро развиваются, что требует от студентов более

глубоких знаний и навыков. Также существует проблема несоответствия знаний, полученных в учебных заведениях, требованиям реальной жизни. Это связано с тем, что многие учебные программы и методы обучения предназначены для формирования теоретических знаний, но не предоставляют достаточно практического опыта и тренировки необходимых профессиональных навыков. Кроме того, нередко студенты не имеют возможности получить опыт в своей области деятельности в связи с отсутствием практических занятий на учебном заведении или недостаточным количеством стажировок и практик во время обучения.

Рассмотрение проблем, связанных с получением практического опыта. Для большинства работодателей практический опыт является обязательным критерием при выборе кандидатов на должность. Многие студенты сталкиваются с трудностями в получении практического опыта. Это связано с различными факторами, например, с тем, что работодатели не заинтересованы в приеме стажеров из-за дополнительных затрат на обучение и контроль за персоналом, либо потому, что у студентов нет необходимых знаний и навыков для выполнения стоящих перед ними задач. Или потому что учебные заведения не всегда обеспечивают возможности для стажировок и практик. Однако, несмотря на это, получение студентами практического опыта необходимо для их дальнейшего развития и установления связей с работодателями.

Анализ влияния социально-экономических факторов на взаимодействие работодателей и студентов. В современном обществе существует проблема социально-экономических неравенств, которая может повлиять на возможности студентов для трудоустройства. В условиях неравномерного экономического развития многих регионов России, низкого уровня жизни, высокой безработицы и низкой оплаты труда молодежь испытывает определенные трудности в трудоустройстве. Кроме того, зачастую вакансии требуют наличия опыта работы, что осложняет ситуацию для выпускников. В связи с этим, имеются сложности и в установлении связей между работодателями и студентами. Студенты из бедных семей и/или малочисленных этнических групп часто не

имеют достаточных ресурсов и возможностей для получения хорошего образования, и успешного трудоустройства.

Оценка экономических проблем, возникающих у студентов при трудоустройстве. Как правило, ни у родителей, ни у самих студентов нет возможности выдержать длительный период безработицы. Для студентов, проживающих в другом городе, также представляется сложность организации переезда и проживания в новом месте работы. Все эти проблемы делают процесс трудоустройства непростым для студентов. Конечно, большинство этих проблем решаемые, но требуют серьезной организации со стороны учебных заведений и работодателей. Многие студенты сталкиваются с экономическими проблемами при трудоустройстве, такими как низкая зарплата, отсутствие социальных гарантий и долгий период ожидания начала работы. Это может стать причиной финансовых трудностей у студентов и побудить их искать другие варианты работы.

### 1.3 Сравнительный анализ по сфере трудоустройства студентов

В современном мире студенты сталкиваются с огромным количеством проблем связанных с трудоустройством. Это объясняется, в первую очередь, изменением экономической ситуации в стране, затруднением социального и экономического развития, высокой конкуренцией на рынке труда, а также перенасыщением рынка специалистами со схожими квалификациями. В данном разделе будет проведен сравнительный анализ по сфере трудоустройства студентов РФ.

Согласно отчету Минтруда РФ, за 2019 год, уровень занятости выпускников высших учебных заведений составил 89,9%, что на 4,4% больше, чем в 2018 году. Это означает, что почти 90% выпускников высших учебных заведений были трудоустроены в течение года после окончания обучения. Это является положительной динамикой в развитии рынка труда и может свидетельствовать о том, что выпускники получают достаточно

высококачественное образование, соответствующее требованиям современного рынка.

Наиболее массовой ступенью высшего образования в России является бакалавриат — первая образовательная ступень, предполагающая развитие общих профессиональных компетенций студентов. 69% выпускников, окончивших бакалавриат российских вузов в 2018 г., имели официальное трудоустройство осенью 2021 г. (Табл. 1.1).

Бакалавриат заменил специалитет в качестве базы высшего образования, но, несмотря на переход к двухступенчатой системе «бакалавриат — магистратура» в большинстве предметных направлений, подготовка кадров по программам специалитета до сих пор превалирует в некоторых сферах обучения. Наиболее значимые из них — медицина, отдельные профили образования в инженерных науках и искусстве и культуре. Показатель трудоустройства среди выпускников специалитета чуть выше, чем среди бакалавров (76%), но, скорее всего, разница обусловлена не различиями в качестве подготовки бакалавров и специалистов, а спецификой направлений обучения: современным рынком очень востребованы выпускники технических, инженерных и медицинских направлений подготовки.

Таблица 1.1 – Результат трудоустройства выпускников 2018 г. в разрезе уровней подготовки

	Доля трудоустроенных выпускников, %	Численность выпуска по форме ВПО-1, чел	Численность выпуска по данным МТВ, чел.
Бакалавриат	69,1	660 950	550 400
Из них:	76,2	—	113 004
закончили магистратуру в 2020/2021 г.			
не продолжили обучение в магистратуре	68,7	—	437 396
Магистратура	77,5	170 437	146 947
Специалитет	76,1	101 766	85 205

Лидерами по доле трудоустроенных являются выпускники с магистерской степенью (78%). Обучение в магистратуре нацелено на развитие социальных

навыков и углубление существующих профессиональных знаний. Кроме того, диплом магистра может являться важным барьером доступа к управленческим позициям, которые работодатели оплачивают более щедро. Ввиду тотального распространения практики совмещения учебы и работы среди студентов магистратуры, данный уровень образования может рассматриваться в качестве особого инструмента интеграции студентов на российском рынке труда. О преимуществе наличия магистерского диплома говорит и сравнение характеристик трудоустройства бакалавров-2018, получивших магистерскую степень в 2020–2021 гг. (таких 18%) и ограничившихся полученным бакалаврским дипломом. Доля трудоустроенных среди продолживших обучение выше на 7% (76% среди продолживших обучение и 69% среди остальных выпускников). Данное наблюдение дополнительно подтверждает высокую ценность магистерского диплома для российского рынка труда [18].

В 2021 году из российских вузов выпустилось 746 597 человек. По данным мониторинга на середину июля 2022 года, трудоустроились в течение года 89% из них [26].

Современный рынок труда рассматривает диплом о высшем образовании не как независимый сигнал о продуктивности работника, а как необходимый пререквизит для найма, который может быть дополнен другими информативными характеристиками. Наиболее показательная из них — репутация университета, выдавшего диплом. Неоднородное качество высшего образования, которое значительно разнится между университетами, вместе с неоднородным развитием регионов приводит к существенным различиям в возможностях трудоустройства выпускников. Заметную роль в массовизации высшего образования сыграло разрастание сектора негосударственных вузов. Хотя среди частных вузов в России есть достойные представители, конкурирующие с ведущими государственными образовательными организациями, в среднем негосударственные вузы в общественном представлении ассоциируются с более низким качеством подготовки кадров. 13% дипломов бакалавра и специалиста, а также 4% дипломов магистра,

выданных в 2018 г., относятся к негосударственным вузам. Государственные и негосударственные вузы привлекают разный контингент студентов: выпускники частных образовательных организаций в среднем старше выпускников государственных университетов на 3–4 года (табл. 1.2). При этом они демонстрируют более слабые показатели трудоустройства — средняя зарплата выпускника бакалавриата или специалитета частного вуза ниже на 8% (55,7 тыс. руб. у выпускника государственного университета, 51,7 тыс. руб. — негосударственного), магистратуры — на 3% (75,3 тыс. и 72,9 тыс. руб. соответственно). Небольшая величина различий может объясняться более длительным стажем работы выпускников частных вузов, который помогает компенсировать им отрыв в качестве образования.

Таблица 1.2 – Характеристики выпускников и их трудоустройства в зависимости от собственности образовательной организации

	Бакалавриат/специалитет			Магистратура		
	Доля выпуска в общем, %	Доля трудоустроенных выпускников, %	Возраст на момент выпуска, лет	Доля выпуска в общем, %	Доля трудоустроенных выпускников, %	Возраст на момент выпуска, лет
Государственные вузы	87,0	71,4	25	96,2	78,1	29
Негосударственные вузы	13,0	60,9	29	3,8	64,0	32

Другой источник дифференциации — особый статус университетов (НИУ, федеральные и опорные вузы, участники проекта «5-100», МГУ). Традиционно многие из вузов с особым статусом являются наиболее селективными образовательными организациями, которые отбирают лучших выпускников школ — золотых медалистов, «отличников» ЕГЭ и победителей олимпиад. Особый статус означает и больший объем финансовых средств, выделяемых

государством на развитие научного и инновационного потенциала образовательной организации. Это, в свою очередь, позволяет инвестировать в научную инфраструктуру, привлекать лучших преподавателей и улучшать тем самым качество образования в конкретной организации. Порядка 27% выпускников бакалавриата и специалитета, а также 43% выпускников магистратуры приходится на вузы с особым статусом. Более того, 14% выпускников бакалавриата и специалитета и 24% выпускников магистратуры приходится на ведущие вузы (НИУ, «5-100», МГУ). Выпускники ведущих вузов демонстрируют лучшие трудовые результаты по сравнению с остальными выпускниками [18].

На сегодняшний день наиболее востребованными специальностями являются связанные с IT-технологиями, математикой, механикой и физикой. Экономические специальности также остаются востребованными на рынке труда. Тем не менее, определенные отрасли, например, сфера производства или сфера журналистики, испытывают затруднения в трудоустройстве выпускников.

Так данные портала «Profi.ru» свидетельствуют о том, что среди наиболее востребованных специальностей в 2020 году преобладают IT-специалисты, финансисты, специалисты по логистике и снабжению, маркетологи, юристы и психологи. Такие профессии, как разработчик программного обеспечения, аналитик данных, инженер-проектировщик, специалист по цифровому маркетингу и т.д., оказались наиболее востребованными в России и за ее пределами. Это связано с ростом числа компаний, которые осуществляют цифровую трансформацию и автоматизацию бизнес-процессов, а также с ростом интереса к онлайн-платформам и сервисам в связи с COVID-19. С пандемией COVID-19 многие компании перешли на удаленную работу, что привело к увеличению спроса на сотрудников в сфере IT-технологий и дистанционного обучения. Одновременно специалисты по логистике и снабжению остаются важными для различных отраслей экономики, а финансисты и юристы являются необходимыми для устойчивого функционирования бизнеса.

Однако, в целом на рынке труда сохраняется дефицит квалифицированных кадров во многих отраслях экономики, что создает дополнительные возможности для трудоустройства студентов и молодых специалистов.

Хочется отметить, что темпы увеличения занятости ведущих регионов страны выше, чем в других районах. В Москве, например, официальный уровень безработицы находится на низком уровне, что обусловлено сильной экономикой города, присутствием в нем множества крупных корпораций и университетов. Выпускники в столице имеют больше возможностей для трудоустройства в крупных компаниях и получения высоких заработных плат.

Одной из основных причин проблем с трудоустройством, особенно в отдаленных регионах, является несоответствие требований работодателей и уровня образования выпускников. В связи с этим студентам стоит обратить внимание на сферы, в которых существует нехватка работников соответствующих специальностей. Кроме того, студентам необходимо работать над привлекательностью своего резюме и профессиональных навыков, для этого очень важно активно заниматься внеучебной деятельностью и проходить стажировки, практики.

Кроме того, учитывать следует, что определенные работодатели могут предоставлять льготы определенным категориям специалистов, включая студентов последних курсов. Такие возможности появляются в рамках программ обучения в специальных учебных центрах или на курсах повышения квалификации. Это важный момент для всех студентов, так как может помочь увеличить шансы на получение желаемой работы.

Рассмотрим основные способы трудоустройства студентов.

1. Трудоустройство через карьерные центры вузов. Данная программа предлагает студентам возможности для освоения новых навыков, приобретения опыта работы и знакомства с работодателями. Преимуществом этого способа является то, что многие карьерные центры привлекают к сотрудничеству крупные компании, что делает его более перспективным для студентов.



2. Участие в олимпиадах, хакатонах, чемпионатах и конкурсах. Студенты имеют возможность участвовать в различных мероприятиях, проводимых государством и крупными компаниями. Эти мероприятия направлены на выявление заинтересованных, мотивированных и способных молодых людей, которые могут быть полезны работодателям. Участие в олимпиадах и хакатонах позволяет студентам развивать свои профессиональные навыки, повышать квалификацию и улучшать знания в своей области. Для работодателей – возможность находить талантливых студентов с опытом работы в реальных проектах. Так же участие студентов в таких мероприятиях дает возможность работодателям оценить навыки, уровень знаний и профессионализм потенциальных кандидатов.

3. Работа на фриланс-площадках. Студентам доступна работа на фриланс-площадках, где они могут самостоятельно находить заказчиков и выполнять работу в онлайн-режиме. Однако данный способ имеет и недостатки: нестабильность заработка и иногда высокую конкуренцию.

4. Специализированные сервисы по трудоустройству выпускников. Эти сервисы предлагают студентам различные программы для трудоустройства, включая поиск работы, составление резюме и проведение собеседований. Минусом данного способа выступает тот факт, что подобные сервисы платные.

5. Без программ. Студенты часто занимаются поиском работы самостоятельно, через интернет и социальные сети. Этот способ, хоть и простой, но непредсказуем.

Данные о трудоустройстве выпускников свидетельствуют о том, что выпускники со степенью магистра обладают преимуществом при трудоустройстве перед бакалаврами и являются более предпочтительной рабочей силой. Тем не менее, количество образовательных программ и количество мест для набора абитуриентов по программам магистратуры ограничено по сравнению с численностью выпущенных бакалавров. В связи с этим, можно предположить, что работодатели часто завышают требования к рабочим местам, так как не все должности и профессии требуют более

углубленных знаний, полученных по программам магистратуры, для выполнения должностных обязанностей. Трудности при трудоустройстве бакалавров могут быть обусловлены (помимо других причин) тем, что работодатели часто руководствуются стереотипом о более качественной подготовке выпускника программы магистратуры, а не объективными требованиями к кандидату, определенными конкретным рабочим местом.

Основная проблема, в связи с которой выпускники не могут трудоустроиться по полученной специальности, заключается не столько в том, что на рынке труда отсутствуют вакансии для трудоустройства, сколько в несоответствии объемов выпусков образовательных организаций в разрезе специальностей реальному спросу на них со стороны работодателей. В связи с этим существует необходимость проведения профориентационной работы в процессе получения общего среднего образования. Это обусловлено тем, что абитуриенты не вполне осознанно подходят к выбору специальности для дальнейшего обучения, воспринимая эту ступень образования как продолжение общего среднего образования, а не как профессию, в соответствии с которой им предстоит трудоустраиваться в дальнейшем. Чаще всего потенциальные студенты руководствуются такими факторами, как престижность профессии, мнение родственников и друзей, результаты ЕГЭ и т.д. В то же время, только 38% всех выпускников 2016-2022 гг. работали во время прохождения обучения, а значит, хоть в какой-то степени были осведомлены о реальной ситуации на рынке труда, поэтому выпускники в большинстве своем имеют относительно размытое представление о своей будущей профессии, ее актуальности и востребованности, а также ничем не обусловленные ожидания в части условий труда и заработной платы.

#### 1.4 Вывод

В результате исследования были выявлены значимые проблемы взаимодействия работодателей и студентов на российском рынке труда, такие как недостаточная подготовленность студентов к требованиям работодателей,

отсутствие информации о вакансиях и возможностях стажировки, а также невысокая ценность для работодателей опыта работы во время учебы.

В то же время, были выявлены также положительные аспекты взаимодействия работодателей и студентов, такие как взаимоприемлемые ожидания сторон от входа в профессию, распространение практик стажировки и менторства, готовность работодателей к инвестированию в развитие навыков студентов.

Сравнительный анализ показал, что сфера IT трудоустройства студентов является наиболее успешной, так как в этой сфере работодатели активно сотрудничают с университетами и охотнее привлекают студентов на стажировки и проекты. В то же время, в сфере образования и науки возможности трудоустройства для студентов ограничены, что требует дополнительных усилий со стороны работодателей и образовательных учреждений.

Таким образом, взаимодействие работодателей и студентов на российском рынке труда представляет собой сложный процесс, который требует совместных усилий сторон для повышения качества подготовки кадров и стимулирования экономического роста.

## ГЛАВА 2. РАЗРАБОТКА РЕКОМЕНДАТЕЛЬНОЙ СИСТЕМЫ

### 2.1 Исследование существующих рекомендательных систем

В последнее время люди ежедневно сталкиваемся с огромным количеством информации из различных источников. Этот поток информации делает задачу поиска полезных или подходящих вещей, таких как фильмы, журналы, книги, веб-сайты, музыка, или даже рабочие места, большой проблемой. Отдельно взятый пользователь не в состоянии охватить взглядом весь перечень предметов, а значит он нуждается в подсказке со стороны, какие товары или услуги наиболее точно соответствуют его интересам и запросам. По этой причине все больше и больше приложений получают широкое развитие, а также появляются новые методы поддержки принятия человеческих решений, предлагающих клиентам услуги, продукты и различные типы информации [14]. Одной из областей исследований в этом направлении является направление рекомендательных систем.

					<i>ВКР-40461806-10.12-09.03.03-462-18-2023</i>			
		№ докум.	Подп.	Дата				
Разраб.	Щетинин В.В.				Разработка рекомендательной системы	Лит.	Лист	Листов
Пров.	Кетько Н.В.						31	22
						ВолгГТУ ЭП-462		
Н.контр.	Кузнецов С. Ю.							
Утв.	Скитер Н. Н.							

Рекомендательные системы – это программные инструменты, которые используют различные методы и алгоритмы, чтобы изолировать нерелевантную информацию от огромного количества данных и создать персонализированные предложения из небольшого подмножества, которые пользователь может изучить в разумные сроки. В действительности, рекомендации – это догадки, предположения, сделанные системой о предмете, который клиент, скорее всего, предпочитает. Для построения таких рекомендаций существует несколько различных методов, и для того, чтобы предоставлять персонализированные рекомендации, необходимы знания о интересах пользователя. Эти интересы могут быть собраны из оценок товаров или услуг, истории покупок или других взаимодействий с системой, сохраненных в профиле пользователя.

Рекомендательные системы имеют некоторые особенности, которые отличают их от других общих научных областей, таких как распознавание образов, искусственный интеллект, интеллектуальный анализ данных и т. д.

Известно множество примеров применения рекомендательных систем. Многие интернет-сервисы используют рекомендательные системы в целях предоставления персонализированных предложений пользователю. Наличие такого сервиса помогает увеличить популярность сайта, а также продажи, в случае площадок для электронной коммерции.

Рекомендательные системы уже интегрированы во множество вебприложений, которые широко используются каждый день миллионами пользователей. Рассмотрим примеры крупнейших ресурсов, использующие рекомендательные механизмы.

LinkedIn – бизнес-ориентированная социальная сеть. Встроенный рекомендательный механизм предлагает пользователю рекомендации людей, которых он, возможно, знает, вакансий, которые могли бы его привлечь, групп, в которые он мог бы захотеть вступить, компаний, которыми он мог бы заинтересоваться. Специализированная система коллаборативной фильтрации LinkedIn основана на технологии Apache Hadoop.

Amazon – одна из крупнейших площадок интернет-торговли — использует рекомендации на основе контента. Когда посетитель выбирает для покупки какой-либо товар, Amazon на основе этого исходного товара рекомендует посетителю другие товары, приобретенные другими пользователями (с помощью матрицы покупки следующего товара на основе его схожести с предыдущей покупкой). Компания Amazon запатентовала этот подход под названием *item-to-item collaborative filtering* (коллаборативная фильтрация от элемента к элементу).

Last.fm создает музыкальную «станцию» рекомендованных песен, наблюдая, какие группы и отдельные треки пользователь прослушивает на регулярной основе. Last.fm воспроизводит дорожки, которые не присутствуют в библиотеке пользователя, но часто воспроизводятся другими пользователями с аналогичными интересами. Поскольку этот подход использует поведение пользователей, он является примером совместной фильтрации.

Spotify – это сервис для стриминга различного контента (миллионов треков, подкастов и видеороликов) от авторов со всего мира. Spotify использует стратегию коллаборативной фильтрации, адаптированную к потребностям конкретного пользователя в зависимости от времени суток. Сервис Spotify поставил задачу разработки такой системы персональных рекомендаций, которая должна анализировать данные в реальном времени и историческую статистику, чтобы, соответственно понимая контекст и поведение пользователя, предлагать ему наиболее релевантные аудиозаписи. Для достижения поставленной бизнес-цели была разработана рекомендательная система на базе эффективного сочетания соответствующих Big Data технологий: Apache Cassandra, Kafka, Storm и Hadoop.

Среди сервисов, которые основываются на рекомендациях в области трудоустройства можно выделить следующие:

HeadHunter. Стоит выделить результаты компании и их опыт в разработке подобных систем.

Первое что сделали hh.ru: начали с создания рекомендательной системы, которая будет приносить соискателям подходящие вакансии сама – рассылки с подходящими вакансиями.

Появилась система ранжирования вакансий, которая проходила через несколько уровне фильтраций – это стало приносить нам около 1,2 миллиона дополнительных откликов в месяц, а это примерно 120 тысяч приглашённых на собеседование и 20 тысяч нанятых.

Разработали поиск по пустому запросу, поиск по не пустому запросу с применением машинного обучения. Было обучено две модели: линейная, XGBoost.

Также дорабатывалась рекомендательная система: были включены признаки по текстовым взаимодействиям, грейдированная целевую функция, признаки по «сырым» svd-векторам по текстам, метапризнаки по линейной регрессии над TF/IDF-векторами.

Повторное использование признаков из рекомендательной системы: статические (вычисляемые до выполнения запроса), текстовые, числовые и категориальные, а также динамические, которые считаются при обработке запроса. Добавлены к обычным признакам признаки, сравнивающие тексты с учётом текстовых взаимодействий.





центральное положение и вариабельность, выявить корреляции между признаками. После этого этапа можно приступать к подготовке данных. Как правило, этот этап самый трудоемкий и на него может уходить до 90% от всего времени проекта, но от того, насколько качественно он выполнен, зависит успех всего проекта.

### 2.2.1 Сбор данных

Так как на данный момент платформа «POLYGON» – платформа для взаимодействия работодателей со студентами, основной задачей которой является трудоустройство молодежи, не имеет достаточного объема данных о вакансиях для обучения модели, обратимся к уже существующим датасетам на площадке kaggle – система организации конкурсов по исследованию данных, а также социальная сеть специалистов по обработке данных и машинному обучению. Таким образом был найден датасет, сформированный из данных о вакансиях в области информационных технологиях, полученных с площадки по поиску работы hh.ru. Данный датасет (Приложение 1) содержит список с описанием IT вакансий по Москве и Санкт-Петербургу за период с 20.05.2021 по 12.02.2021. Датасет состоит из 48564 строк и 15 столбцов и содержит следующую информацию:

- Идентификатор вакансии
- Название компании, предлагающей работу
- Название должности
- Факт наличия заработной платы для вакансии
- Минимальная заработная ставка
- Максимальная заработная ставка
- Опыт
- График
- Ключевые навыки
- Подробное описание работы

- Город
- Профессиональная должность
- Специализация
- Профессиональная область
- Дата публикации вакансии

Подробную информацию о данных можно увидеть на рисунке 2.2. Все зарплаты в валютах, отличных от рубля, были пересчитаны в рубли по курсу Центрального банка России на 02.12.2021 годы. Таким образом, было получено 47330 уникальных вакансий (Приложение 1).

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 48564 entries, 0 to 48563
Data columns (total 15 columns):
#   Column                Non-Null Count  Dtype
---  -
0   Ids                    48564 non-null  int64
1   Employer               48564 non-null  object
2   Name                   48564 non-null  object
3   Salary                 48564 non-null  bool
4   From                   15399 non-null  float64
5   To                     10276 non-null  float64
6   Experience              48564 non-null  object
7   Schedule               48564 non-null  object
8   Keys                   48564 non-null  object
9   Description            48564 non-null  object
10  Area                   48564 non-null  object
11  Professional roles     48564 non-null  object
12  Specializations        48564 non-null  object
13  Profarea names         48564 non-null  object
14  Published at           48564 non-null  object
dtypes: bool(1), float64(2), int64(1), object(11)
memory usage: 5.2+ MB
```

Рисунок 2.2 – Информация о датасете

### 2.2.2 Предобработка и очистка данных

Определим наличие пустых значений в таблице датасета. В случае обнаружения пустых значений необходимо определить столбцы (признаки), к которым они относятся и количество пустых значений в них. При проверке видим, что в столбцах «From» и «To» есть пустые ячейки в количестве 33162 и 38288. Это очень большое количество пропусков для датасета с таким количеством строк, чтобы пытаться как-то восстановить данные. Следовательно,

данные столбцы никакой полезной информации нам не несет, и мы можем при обучении модели ими пренебречь.

Зарплата является важным фактором у соискателей при поиске работы, но учитывая, что работодатели сами не внесли уточнений в данные столбцы, придумывать или пытаться предсказать отсутствующие данные, которые могут с большой погрешностью отличаться от действительных, может сильно повлиять на репутацию компаний и платформы по поиску вакансии в том числе.

В случае равенства предпочтительности вакансий для конкретного пользователя (соискателя), если работодатель оставил информацию о заработной плате можно осуществлять ранжирование по предлагаемым значениям от большего к меньшему.

Платформа может стимулировать работодателей оставлять информацию о зарплатах, например, рассказать работодателям о ранжировании рекомендованных вакансий и/или не пропускать вакансии пока все поля (или важные поля включая данные о заработной плате).

Определим признаки для дальнейшей работы. Получим, что значимыми признаками могут являться: Компания (Employer), Должность (Name), Навыки (Keys), Описание (Description), Профессиональная должность (Professional roles), Специализация (Specializations), Профессиональная область (Profarea names). Для того чтобы модель взаимодействовала с данными, необходимо все признаки объединить в один текст и поместить в рабочий DataFrame, а именно перезаписать значение признака Description на объединённый текст. Для этого пропишем функцию объединения (Рис. 2.3).

```
def to_text(arr):
    """Конвертация списка слов в текст"""
    return arr.apply(lambda x: ' '.join(literal_eval(x))
                     if isinstance(x, str) else '')

def create_combined_data(x):
    """Объединение значимых признаков в один текст"""
    return x['Name'] + ' ' + to_text(x['Professional roles']) + ' ' + to_text(x['Keys']) + ' ' + to_text(x['Specializations']) + ' ' + to_text(x['Profarea names'])

data_small['Description'] = create_combined_data(data) # Перезаписываем признак Description новым описанием
```

Рисунок 2.3 – Функция объединения признаков

## 2.3 Baseline-модель

Перед началом создания модели машинного обучения, создадим baseline-модель (Рис. 2.4).

```
# Подготовка столбца для сортировки (Количество названий вакансий)
number_name = dict(df['Name'].value_counts())
df['Number name'] = df['Name'].apply(lambda x: number_name[x])

# Подготовка столбца для сортировки (Дата публикации)
from datetime import datetime

def format_date(x):
    try:
        return datetime.strptime(x.split(' ')[0], "%Y-%m-%d")
    except:
        return x
df['Published at'] = df['Published at'].apply(lambda x: format_date(x))

df.sort_values(["Published at", "Number name"], ascending=(False, False))[:5][['Ids', 'Employer', 'Name']]
```

Рисунок 2.4 – Код baseline-модели

Это простая модель, которая обычно не только не содержит машинного обучения, но часто даже не содержит вычислений. После построения базовой модели, приступим к созданию «настоящей» модели. Наша цель – улучшить качество модели по сравнению с базовой, а затем делать инкрементальные улучшения с каждой последующей моделью.

В данном случае для рекомендательной системы базовой моделью будет такой алгоритм, который на любой запрос выдаст топ (например, топ-5) самых свежих объектов (последние добавления) среди которых самые популярные вакансии от работодателей в базе данных. Популярность будет определяться со стороны работодателя, наибольшим количеством существующих вакансий (по названиям). Работа модели на рисунке 2.5

	Ids	Employer	Name
0	49313809	Space307	Golang Developer (Кипр)
1	48813842	Монополия	E-mail маркетолог
2	49413720	Eden Springs	Оператор call-центра (удаленно)
54	49844682	WebMasters Russia	Системный администратор
55	49844669	FINNTRAIL (ООО Финэкип)	Системный администратор

Рисунок 2.5 – Вывод baseline-модели

Этот алгоритм справляется с проблемой холодного старта, когда пользователь только зашел на сервис и мы не обладаем никакой информацией о том, что ему можно посоветовать.

Такой подход к созданию рекомендательных алгоритмов называется Popularity-based (основанный на популярности).

Для улучшения алгоритма, нужна дополнительная информация. В имеющихся данных нет информации о том, кто просматривал или откликнулся на какие вакансии: на основе такой информации можно было бы рекомендовать пользователю те вакансии, которыми интересовались другие пользователи, похожие на нашего.

Несмотря на это, есть возможность улучшить алгоритм. Так как у нас есть информация о, которую мы объединили выше. Теперь можно просто рекомендовать похожие вакансии на то, что пользователь смотрит сейчас.

## 2.4 Разработка модели

Разработка модели рекомендательной системы – это процесс создания алгоритма, который использует данные о поведении пользователей на платформе для рекомендации им контента (товаров, услуг, фильмов, музыки и т.д.). В разработку модели входит сбор данных, их анализ и выбор методов машинного обучения, которые наиболее точно предсказывают интересы пользователя. Основная цель рекомендательной системы – повысить удовлетворенность пользователей и увеличить прибыль платформы или магазина.

После того, как данные готовы, можно приступить собственно к разработке модели. Для этого нужно подобрать методы, алгоритмы для построения модели. Далее программируем модель, проверяем работоспособность и адекватность модели, если результат устраивает, то необходимо продумать внедрение модели и дальнейшее ее развитие.

### 2.4.1 Выбор метода

Говоря о создании рекомендательной системы нужно определить метод, алгоритм исходя из имеющихся данных, по которому будет производиться подбор рекомендаций.

Методы, используемые для построения рекомендательных систем: контент-ориентированные, коллаборативная фильтрация, кластеризация. Рассмотрим все методы по порядку.

Существует два основных подхода к созданию рекомендательных систем: контент-ориентированная и коллаборативная фильтрация.

Контент-ориентированные модели. Суть этого подхода заключается в том, что мы сопоставляем пользователей с тем контентом или товарами, которые им нравились или были ими куплены. Здесь важны атрибуты пользователей и продуктов. Например, для рекомендаций к фильмам мы используем такие признаки, как режиссер, актеры, продолжительность фильма, жанр и т.д., чтобы найти сходство между фильмами. Кроме того, мы можем извлечь такие характеристики, как оценка настроений и оценки TF-IDF из описаний фильмов и обзоров [16].

Целью данного метода является создание «профиля» для каждого пользователя и каждого предмета.

Коллаборативная фильтрация (совместная фильтрация). Метод использующий известные предпочтения (оценки) группы пользователей для прогнозирования неизвестных предпочтений другого пользователя. Его основное допущение состоит в следующем: те, кто одинаково оценивал какие-либо предметы в прошлом, склонны давать похожие оценки другим предметам и в будущем.

В отличие от контентно-ориентированного подхода, здесь нет признаков, соответствующих пользователям или предметам. Все, что у нас есть — это Матрица полезности (Табл. 2.1) [16].

Таблица 2.1 – Матрица полезности

	Object 1	Object 2	Object 3	Object 4	Object 5	Object 6	Object 7
A	4			5	1		
B	5	5	4				
C				2	4	1	
D		3					3

A, B, C, D — пользователи, а столбцы представляют объект (товар, фильм и т.д.). Значения представляют оценки (1–5), которые пользователи дали объекту.

Кластеризация. Кластеризация – задача группировки множества объектов на подмножества (кластеры) таким образом, чтобы объекты из одного кластера были более похожи друг на друга, чем на объекты из других кластеров по какому-либо критерию [19]. Кластеризация обычно используется, когда задача рекомендательной системы становится задачей без учителя. Это решение, конечно, не дает лучших результатов сразу, но является хорошей отправной точкой для таких случаев, пока не будет получено достаточно данных. Кластеризация также может быть использована для создания мета-признаков для объектов [16].

В данном случае был выбран алгоритм кластеризации. Преимущества использования этого алгоритма для данной задачи заключается в том, что задача кластеризации относится к классу задач обучения без учителя – работает с неразмеченными данными, у нас нет фиксированного набора классов, на которые мы можем разбить наши данные, а также критериев. Данные факторы идеально описывают проблем, с которыми сталкиваются при создании рекомендательных систем, когда имеем мало исторических/размеченных данных.

## 2.4.2 Обработка текста

Обработка текста будет осуществляться с помощью библиотек `nltk` и `sklearn`. Первая – библиотека нужна специально для работы с текстами, вторая – общая библиотека алгоритмов машинного обучения.

Цель обработки – векторизация текста для дальнейшего его использования в модели машинного обучения.

Необходимо сузить признаковое пространство перед векторизацией. Для этого проведем стандартный набор процедур:

- Приведение текстов к нижнему регистру;
- Исключение небуквенных символов — удаление из текста различных символов, знаков препинания, цифр;
- Лемматизация — один из ключевых этапов предобработки текстов, приведение слова к начальной форме, инфинитиву;
- Удаление стоп-слов — тоже очень важный этап, здесь идёт исключение неинформативных слов по словарю, который включает часть стандартного словаря и специальные словари, разработанные для конкретного типа текстов.

На выходе из такого пайплайна препроцессинга получаем предобработанный текст, готовый к вероятностной векторизации. Важно заметить, что для TF-IDF подхода исключение предобработки из пайплайна нерелевантно, так как это раздувает размерность латентного пространства и, соответственно, повышает вероятность переобучения тестовых моделей [15].

Токенизируем текст по словам, значит нужно очистить текст от лишнего и привести его к набору слов, каждое из которых состоит в начальной форме.

Получим набор стоп-слов, этот набор содержит самые популярные слова, которые содержатся во всех текстах и в больших объемах. Данные слова не несут важной смысловой нагрузки, поэтому мы можем ими пренебречь. Это могут быть предлоги, местоимения, союзы и т.п. В библиотеке `nltk`, есть модели наборов таких слов для разных языков. Воспользуемся такой моделью для



русского и английского языка (при дальнейшей токенизации текстов было выявлено большое количество используемых английских слов лишних для понимания содержания).

Перед тем как очистить текст от стоп-слов его необходимо привести к нижнему регистру, чтобы производить сравнения слов без учета регистра т.к. в этом нет необходимости, нас интересует только содержание текста и значения слов. Чтобы проще было удалять стоп-слова, необходимо из текста удалить все символы и знаки препинания, а также цифры.

После очистки и лемматизации текстов (Рис. 2.6) необходимо произвести их векторизацию. Векторизация будет осуществляться при помощи способа TF-IDF (формула 3) – это способ векторизации текста, отражающий важность слова в документе, а не только частоту его появления. Он означает «частота термина — обратная частота документа» и придает вес словам, которые в конечном счете являются более значимыми во всем корпусе слов, — словам, которые не появляются невероятно часто, а также словам, которые не являются чрезвычайно редкими.

```

STOPWORDS = set(stopwords.words('russian')) | set(stopwords.words('english')) # Множество русских и английских стоп слов
PATTERN_ALL_SIGNS = '[«»-.,!@#$$%^&*()_+=-?:;№""\|/<>{}-]' # Шаблон знаков

def del_all_signs(text):
    """Удаление всех знаков из текста"""
    return re.sub(PATTERN_ALL_SIGNS, ' ', text)

def del_stop_words(text):
    """Удаление стоп слова из текста"""
    text_split = text.split(' ')
    for word in text_split:
        if word in STOPWORDS:
            text_split.remove(word)
    return ' '.join(text_split)

morph = MorphAnalyzer()

def text2normal(text):
    """Лемматизация всех слов текста"""
    text = text.split()
    res = []
    for word in text:
        res.append(morph.parse(word)[0].normal_form)
    return ' '.join(res)

def text_clean(text):
    """Основная функция очистки текста,
    исполняет ранее описанные функции"""
    text = text.lower()
    text = del_all_signs(text)
    text = text2normal(text)
    text = del_stop_words(text)
    return text

data_small['description_normal'] = data_small['Description'].apply(text_clean) # Получаем текст пригодный для векторизации

```

Рисунок 2.6 – Набор функций для очистки текста

TF-IDF — применяется для анализа значимости слова в документе, который является частью большой коллекции документов либо корпуса. Мы сталкиваемся с задачами, где необходимо проанализировать или найти нужной информации в тексте. Для решения практически всегда используем статистическую меру – TF-IDF.

TF (или частота слова) – это отношение количества употреблений какого-либо слова к совокупному количеству слов документа. Следовательно, анализируется значимость слова в одном отдельном документе.

$$tf(t, d) = \frac{n_t}{\sum_k n_k}, \quad (1)$$

где  $n_t$  – количество слов  $t$  в текстовом документе, а в знаменателе общее количество слов.

IDF — это обратная частотность документов, с которой какое-либо слово упоминается в документах коллекции. Для любого уникального слова в пределах точной коллекции документов присутствует одно значение IDF.

$$idf(t, D) = \log \frac{|D|}{|\{d_i \in D \mid t \in d_i\}|}, \quad (2)$$

где  $|D|$  – количество документов ( $d$ ) в корпусе ( $D$ ). Знаменатель – это количество документов, в которых встречается термин  $t$ .

$$tfidf(t, d, D) = tf(t, d) * idf(t, D) \quad (3)$$

Чем выше оценка TF-IDF, тем более важным или релевантным является термин; по мере того как термин становится менее релевантным, его оценка TF-IDF приближается к 0 [32].

Для векторизации воспользуемся объектом `TfidfVectorizer` импортированным из модуля `sklearn` (Рис. 2.7).

```
documents = data_small['description_normal'].values.astype('U') # Получаем данные из ранее подготовленных текстов
vectorizer = TfidfVectorizer(stop_words=['english', 'russian']) # Инициализируем объект для векторизации текста
features = vectorizer.fit_transform(documents) # Применение TfidfVectorizer, чтобы преобразовать текст в векторы признаков
```

Рисунок 2.7 – Векторизация при помощи `TfidfVectorizer`

Векторизовав все необходимые текстовые значения, можно перейти к построению самой модели.

### 2.4.3 Content-based-модель

Основанные на контенте (content-based) — рекомендации основаны на данных, собранных о каждом конкретном потенциальном к рекомендациям объекте.

Этот тип лежит в основе многих рекомендательных систем. В отличие от коллаборативной фильтрации, как рассказывалось выше, этап знакомства с пользователем опускается. Объекты рекомендуются на основе знаний о них: жанр, производитель, конкретные функции и т.п. В общем, применяют любые данные, которые можно собрать.

По такому принципу работают системы интернет-магазинов, онлайн-кинотеатров и других сервисов.

Таким образом для данной модели были собраны и объединены необходимые данные в нужной форме и приведены в векторизованную (это было описано в пунктах выше) форму для дальнейшей работы.

Теперь, когда у нас есть наборы чисел, которые компьютер способен понимать.

Для того чтобы определить похожесть между двумя векторами, нужно уметь их сравнивать. Так будет возможность делать рекомендации.

Рассмотрим два подхода вероятностный, а именно метод k-means (k-средних) и перебор, сравнивая по cosine similarity (косинусное сходство). K-средних – алгоритм минимизирует суммарное квадратичное отклонение точек кластеров от их центров. Косинусное сходство – это мера сходства между двумя векторами предгильбертового пространства, которая используется для измерения косинуса угла между ними.

Задача стоит в том, чтобы узнать какой подход работает наиболее удачно. Однозначный ответ можно дать только вовремя А/В тестирования, непосредственно оценивая работу моделей во время эксплуатации на платформе. Сейчас получится дать субъективную оценку адекватности моделей.

Воспользуемся функцией KMeans из библиотеки sklearn для того чтобы осуществить кластеризацию при помощи метода k-средних. Так как данный метод способен кластеризовать данные только по известному количеству кластеров, то нужно определиться с их количеством. Для этого воспользуемся методом силуэта. Метод силуэта (Silhouette Method) вычисляет среднее расстояние между точками в своем кластере  $a_i$  и среднее расстояние от точек до следующего ближайшего кластера, называемого  $b_i$ . Чтобы найти оптимальное количество кластеров, запустим алгоритм кластеризации, импортировав модуль metrics из пакета sklearn (Рис. 2.8).

```
scores = []

for i in range(2, 41):
    model = KMeans(n_clusters=i, init='k-means++', max_iter=100, n_init=100).fit(features)
    score = metrics.silhouette_score(features, model.labels_, metric='euclidean', sample_size=features.shape[0])
    scores.append(score)

num_clusters = np.argmax(scores)
print('\nОптимальное количество кластеров =', num_clusters)
```

Оптимальное количество кластеров = 23

Рисунок 2.8 – Поиск оптимального количества кластеров методом силуэта

Так запустив цикл от 2 до 41 и на каждой итерации построив соответствующее количество кластеров модели равное  $i$  и посчитав оценку силуэта для такой модели можно получить оптимальное количество кластеров (Рис. 2.8). Получаем оптимальное количество кластеров для работы равный 23.

На известном подходящем количестве кластеров запустим модель кластеризации методом  $k$ -средних. Получим разбиение на кластеры, каждому вакансии присвоен номер определенного кластера, к которому он относится. В качестве проверки выведем кластеры по десятью ключевым словам и дадим название каждому кластеру. Получаем:

Cluster 0: Администраторы и специалисты по сетевым технологиям

Cluster 1: Интернет-разработчики и программисты

Cluster 2: Инженеры и программисты в области IT

Cluster 3: Аналитики и менеджеры проектов

Cluster 4: Работники в производстве и обслуживании станков

Cluster 5: Дизайнеры и художники в различных областях

Cluster 6: Разработчики систем управления и программного обеспечения для предприятий

Cluster 7: Разработчики и программисты в области iOS и веб-технологий

Cluster 8: Разработчики и программисты в области JavaScript и веб-технологий

Cluster 9: Менеджеры по продажам и телемаркетологи

Cluster 10: Сервисные и технические специалисты в области телекоммуникаций и оборудования

Cluster 11: Разработчики и программисты в области Java и связанных технологий

Cluster 12: Маркетологи и специалисты по рекламе в интернете

Cluster 13: Разработчики и программисты в области PHP и веб-технологий

Cluster 14: Техническая поддержка и специалисты-новички в IT

Cluster 15: Тестировщики и инженеры в области QA

Cluster 16: Разработчики и программисты в области игровых технологий

Cluster 17: Разработчики и программисты в области React и веб-технологий

Cluster 18: Системные администраторы и специалисты по Linux и DevOps

Cluster 19: Разработчики и программисты в области ERP и управления предприятием

Cluster 20: Менеджеры по продукту и управлению проектами

Cluster 21: Руководители и специалисты в проектной деятельности

Cluster 22: Специалисты по информационной безопасности и компьютерным технологиям

Подробнее узнать о содержании кластеров можно в приложении 2.

Кластеризовав данные мы сможем подбирать для каждой вакансии топ из максимально похожих на нее. Также в качестве одного из вариантов работы модели можно демонстрировать вакансии не только максимально похожие на данную, но и случайно выбранные вакансии из конкретного кластера. Это нужно только в качестве представления, чтобы пользователю предлагались разные вакансии по выбору (для просмотра) одной и той же вакансии.

Второй вариант модели – перебрать все вектора полученных описаний и при помощи косинусного сходства определить наиболее похожие вакансии для данной (формула 4).

$$Sim(A, B) = \cos(\theta) = \frac{A \cdot B}{||A|| ||B||} \quad (4)$$

Иллюстрация к формуле 4 изображена на рисунке 2.9.

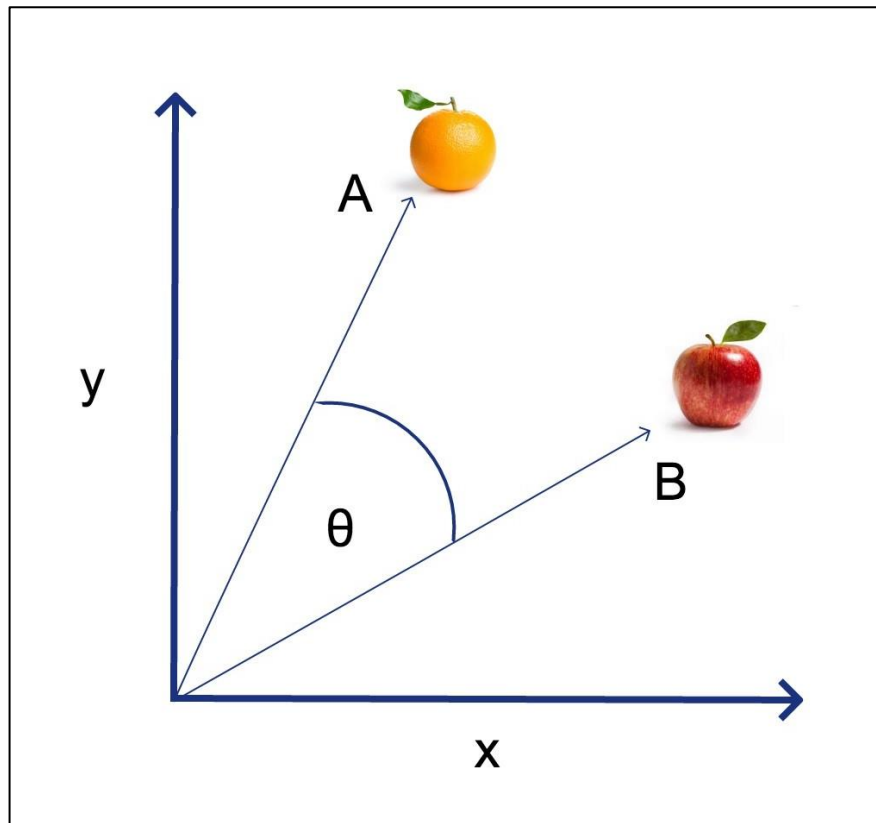


Рисунок 2.9 – Косинусное сходство

Для обоих методов были построены функции демонстрации наиболее подходящих вариантов для текущей вакансии (Рис. 2.10-2.11).

```
def predict_top(id, data, top):
    """Функция возвращает для конкретной вакансии по id топ подходящих вакансий в количестве top (Кластеризация k-means)"""
    tops = [x for x in range(len(data_small['cluster'])) if x != id and data_small['cluster'][x] == data_small['cluster'][id]]
    # return data.iloc[sample(tops, len(tops))[:top]] # Возвращает топ случайных вакансий из кластера
    return data.iloc[tops[:top]] # Возвращает топ ближайших вакансий
```

Рисунок 2.10 – Функция получения топа вакансий по методу кластеризации k-means

На рисунке 2.10 можно заметить закомментированную строчку кода – это есть один из вариантов рекомендации. Из сформированного кластера пользователю могут быть предложены в количестве top случайные вакансии.

```
def predict_top(id, data, top):
    """Функция возвращает для конкретной вакансии по id
    топ подходящих вакансий в количестве top (Косинусная схожесть)"""
    test_sentence = data.iloc[id]['Description']
    embed_mat = data['spacy_description'].values
    best_index = predict_spacy(nlp, test_sentence, embed_mat, top)
    return data.iloc[best_index]
```

Рисунок 2.11 – Функция получения топа вакансий по методу оценки косинусной схожести

Рассмотрим функцию `extract_best_indices` (Рис. 2.12), которая используется для извлечения лучших индексов на основе косинусного сходства в матрице `m`.

Если  $m$  имеет более одного измерения, функция вычисляет косинусное сходство между столбцами ( $\text{axis}=0$ ) матрицы  $m$ . Затем происходит сортировка косинусного сходства в порядке убывания и проверка наличия маски (если есть), которая используется для фильтрации результатов. Для элементов, которые равны 0, функция использует логическое «или» для объединения других значений маски. Необходима для получения лучших  $\text{topk}$  индексов.

```
def extract_best_indices(m, topk, mask=None):
    if len(m.shape) > 1:
        cos_sim = np.cosine_similarity(m, axis=0)
    else:
        cos_sim = m

    index = np.argsort(cos_sim)[::-1]
    if mask is not None:
        assert mask.shape == m.shape
        mask = mask[index]
    else:
        mask = np.ones(len(cos_sim))
    mask = np.logical_or(cos_sim[index] != 0, mask)
    best_index = index[mask][:topk]
    return best_index
```

Рисунок 2.12 – Функция извлечения лучших индексов

## 2.5 Вывод

Перебор с использованием косинусного сходства является наиболее подходящим решением для поставленной задачи, т.к. метод часто используется для коллаборативной фильтрации в рекомендательных системах и показывает хорошие результаты.

В свою очередь  $k$ -means полезен для сегментации пользователей по интересам и поведению, что может помочь в создании более точных рекомендаций. Однако, этот метод требует знания числа кластеров, что может быть сложно определить заранее. Также при изменении данных сами количество кластеров и их содержание может меняться, поэтому на обработку может уходить большое количество времени.

Важно отметить, что для эффективной работы рекомендательной системы, помимо выбора метода, важна качество и количество данных. Работу системы



можно улучшить путем обработки данных – попробовать использовать другие группы признаков, другие группы слов и т.п.

## ГЛАВА 3. ПЕРСПЕКТИВЫ РАЗВИТИЯ И РЕЗУЛЬТАТЫ РАЗВИТИЯ СИСТЕМЫ

### 3.1 Улучшение модели

В предыдущих моделях (content-based рекомендации) не учитывается поведение пользователей, так как модели смотрят только на похожесть объектов между собой.

Коллаборативная фильтрация работает лучше, так как можно использовать предыдущий подход и применить его к пользователям: собрать всех пользователей в одну матрицу и посчитать, какие пользователи похожи между собой.

Memory-based подход. Есть несколько способов применения коллаборативной фильтрации. Один из них – подход, основанный на памяти. Этот подход подразумевает, что никакие модели не используются для расчетов рекомендаций: все необходимые данные хранятся в одной таблице и они все должны находиться в памяти для создания рекомендации.

В данном подходе можно составить матрицу из действий пользователей, а затем сравнить пользователей с помощью какой-нибудь функции. Например, все той же метрике – косинусной схожести.

Недостатком такого метода является то, что при добавлении нового пользователя нужно заново строить матрицу. При этом, матрица будет разреженная (так как пользователей и вакансий может быть много), поэтому вычисления с ней будут более трудные.

					ВКР-40461806-10.12-09.03.03-462-18-2023							
		№ докум.	Подп.	Дата								
Разраб.	Щетинин В.В.				Перспективы развития и результаты развития системы				Лит.	Лист	Листов	
Пров.	Кетько Н.В.										53	12
									ВолгГТУ ЭП-462			
Н.контр.	Кузнецов С. Ю.											
Утв.	Скитер Н. Н.											

Выше уже были рассмотрены недостаток memory-based подхода: все приходится держать в памяти, поддерживать такую систему сложно.

Model-based подход заключается в построении модели, которая будет заниматься предсказанием. Алгоритмов, основанных на model-based подходе, много. Рассмотрим один из них.

Truncated SVD. Многие подходы так или иначе используют преобразования матриц (например, разложение или уменьшение размерности). Это помогает сократить вычисления и хранить меньше информации в памяти.

В данном случае лучше использовать TruncatedSVD в качестве алгоритма уменьшения размерности. SVD (Singular Value Decomposition) позволяет уменьшить размерность матрицы и эффективно работает с разреженными матрицами.

### 3.2 О платформе «POLYGON»

В третьей главе будет демонстрироваться сама веб-платформа для взаимодействия работодателей со студентами.

Платформа для поиска работы школьникам и студентам «POLYGON»<sup>1</sup>. Узкоспециализированная платформа, которая позволяет находить для работодателей молодую рабочую силу, а школьникам и студентам позволяет погрузиться в интересующую их профессию, давая им возможность отыскать интересную работу или стажировку.

Для данной платформы и разрабатывается рекомендательная система, чтобы улучшить подбор вакансий для школьников и студентов. На данный момент разработано ядро рекомендательной системы, но даже сейчас можно с ним взаимодействовать пользователю (Рис. 3.1), но к сожалению, пока не через платформу.

---

<sup>1</sup> [www.polygon-job.ru](http://www.polygon-job.ru)

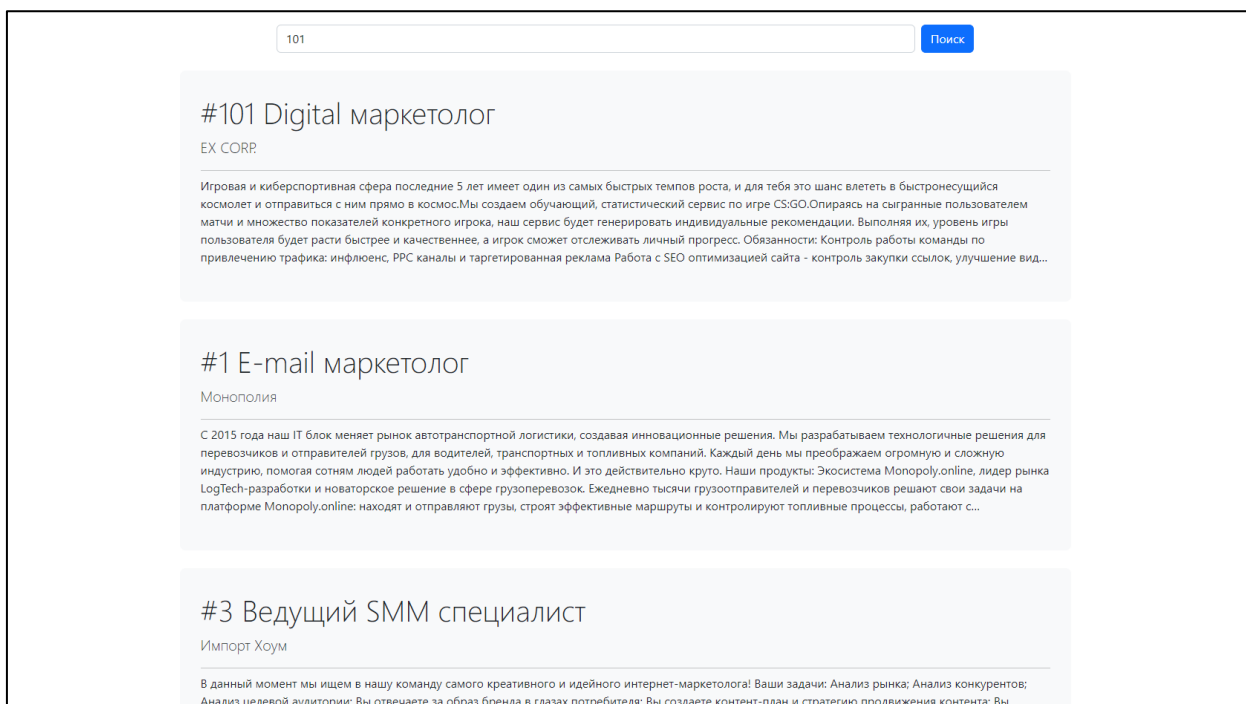


Рисунок 3.1 – Работа модели кластеризации k-means

Сама платформа представляет из себя сайт, на котором можно пройти процедуры регистрации, а в дальнейшем и авторизации (Рис. 3.2-3.3).

Рисунок 3.2 – Экранная форма страницы регистрации

Особенность данных страниц заключается во взаимодействии с пользовательскими данными и очень важно обеспечить необходимый уровень безопасности для хранения личных данных пользователей. На платформе используется защищенное соединение для обмена информацией и

осуществляется хранение данных на удаленном сервере, все пароли хранятся в хешированном виде.

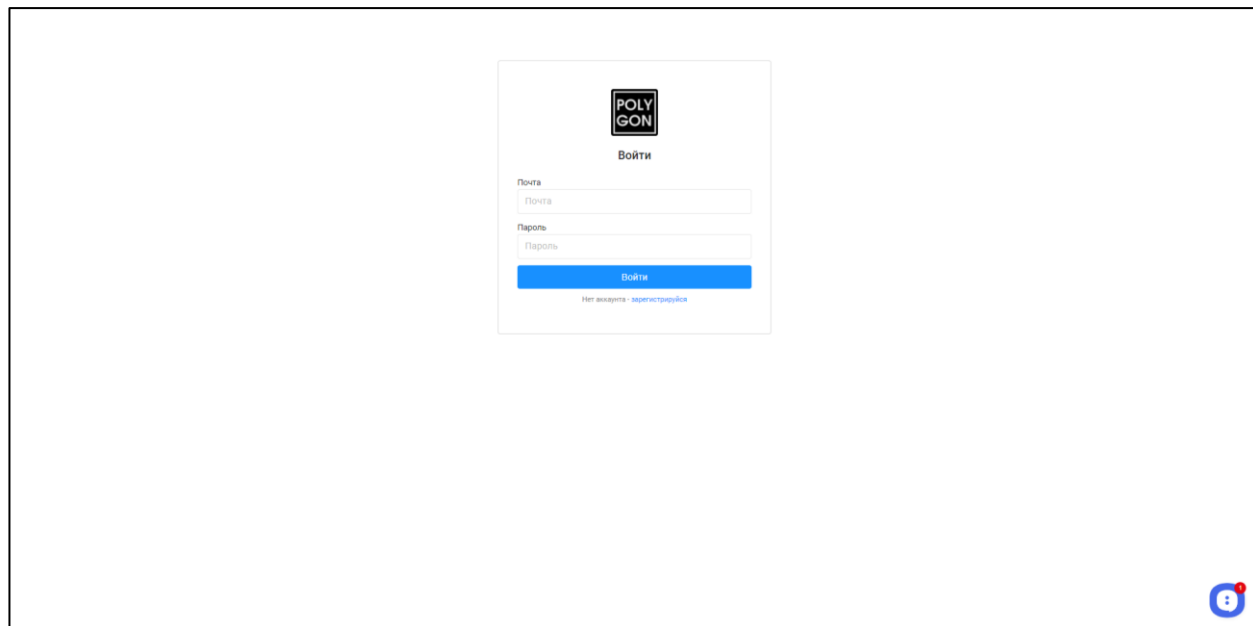


Рисунок 3.3 – Экранная форма страницы авторизации

Есть главная страница с возможностью искать вакансии при помощи различных фильтров, можно добавлять вакансии в избранные, чтобы не потерять. Также есть возможность посмотреть интересные статьи, написанные авторами платформы «POLYGON» (Рис. 3.4).

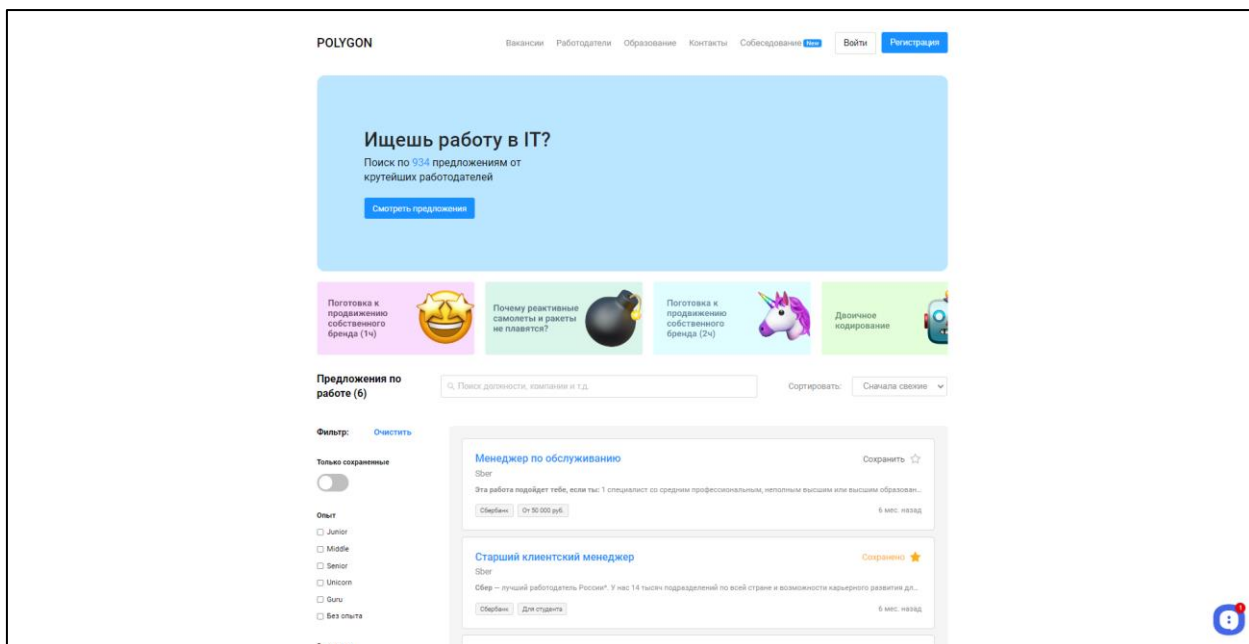


Рисунок 3.4 – Экранная форма главной страницы

На главном меню платформы, можно сделать переход на вкладку с вакансиями и можно перейти в мини-игру «Собеседование» (Рис. 3.5).

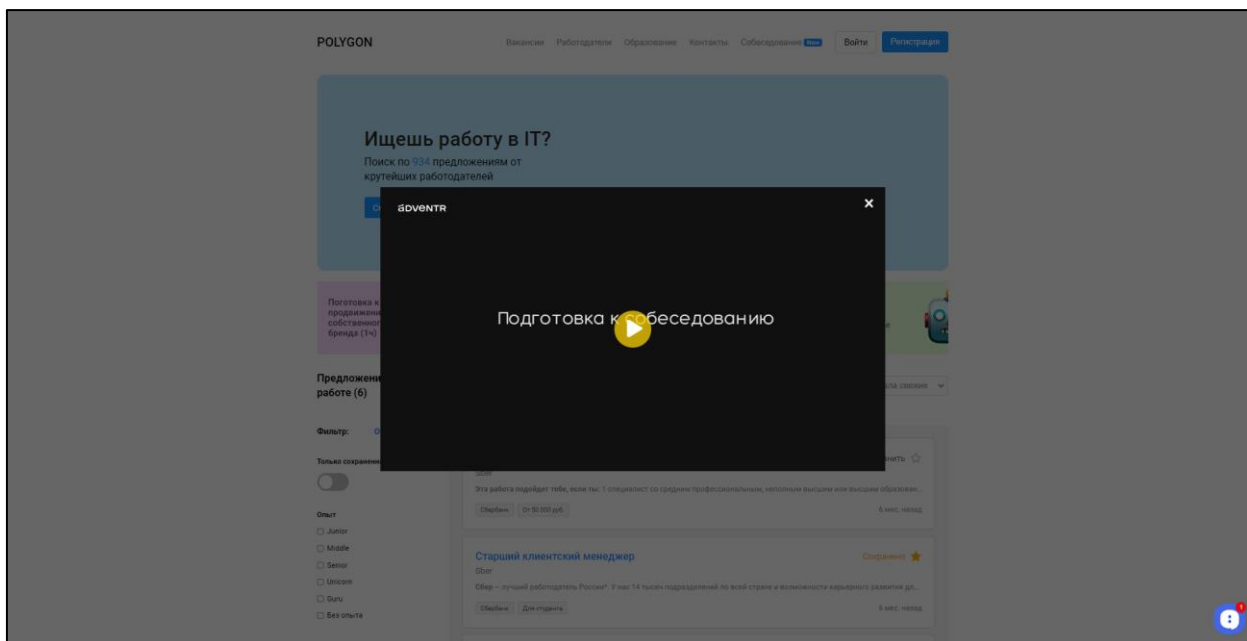


Рисунок 3.5 – Модальное окно мини-игры «Собеседование»

На сайте есть отдельная страничка с вакансиями, где нет лишней информации (Рис. 3.6)

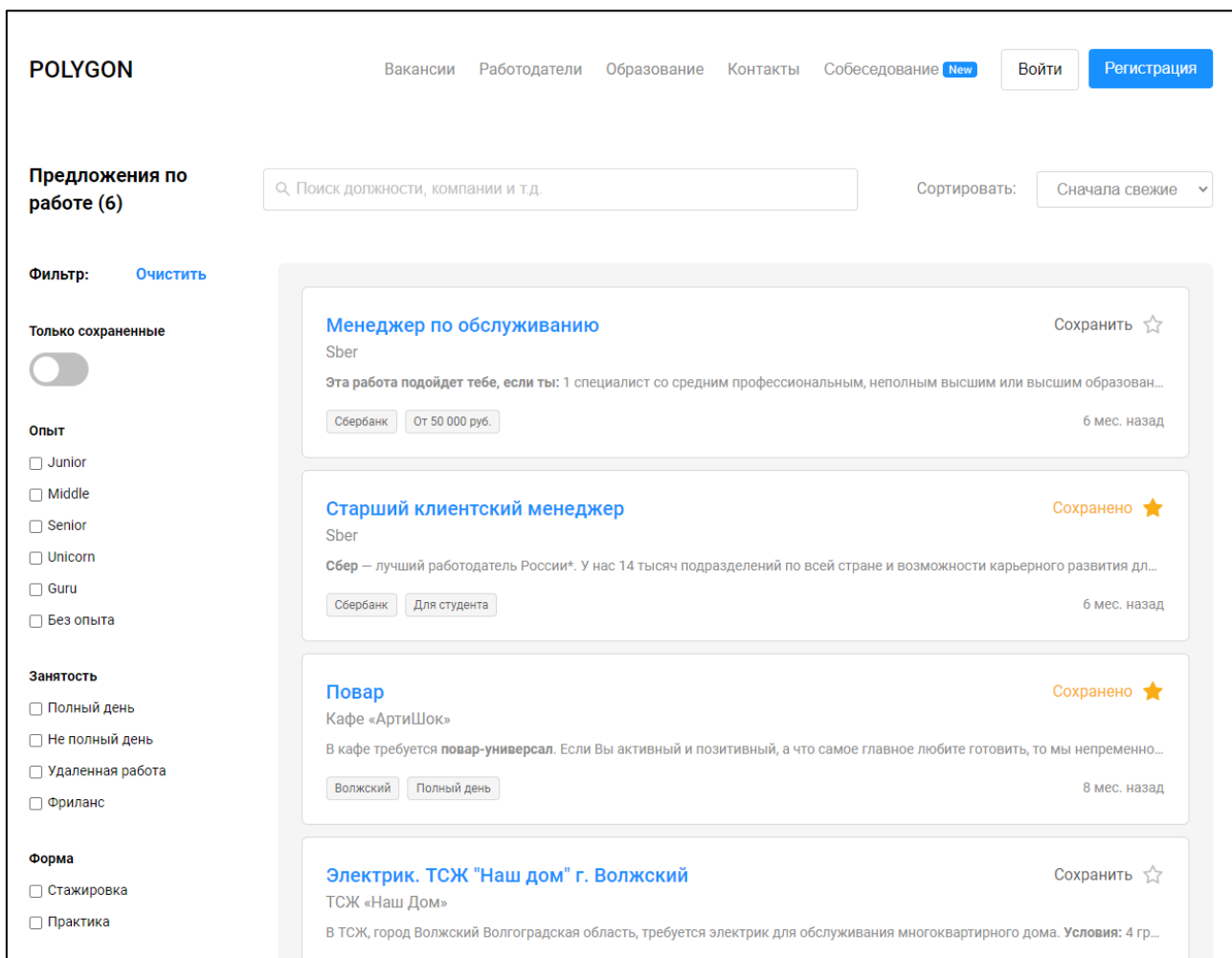


Рисунок 3.6 – Экранная форма раздела «Вакансии»

Что касается самих вакансий, то по клику на них можно открыть страницу с информацией, о конкретной вакансии, а также есть возможность откликнуться на нее, заполнив форму внизу страниц (Рис. 3.7).

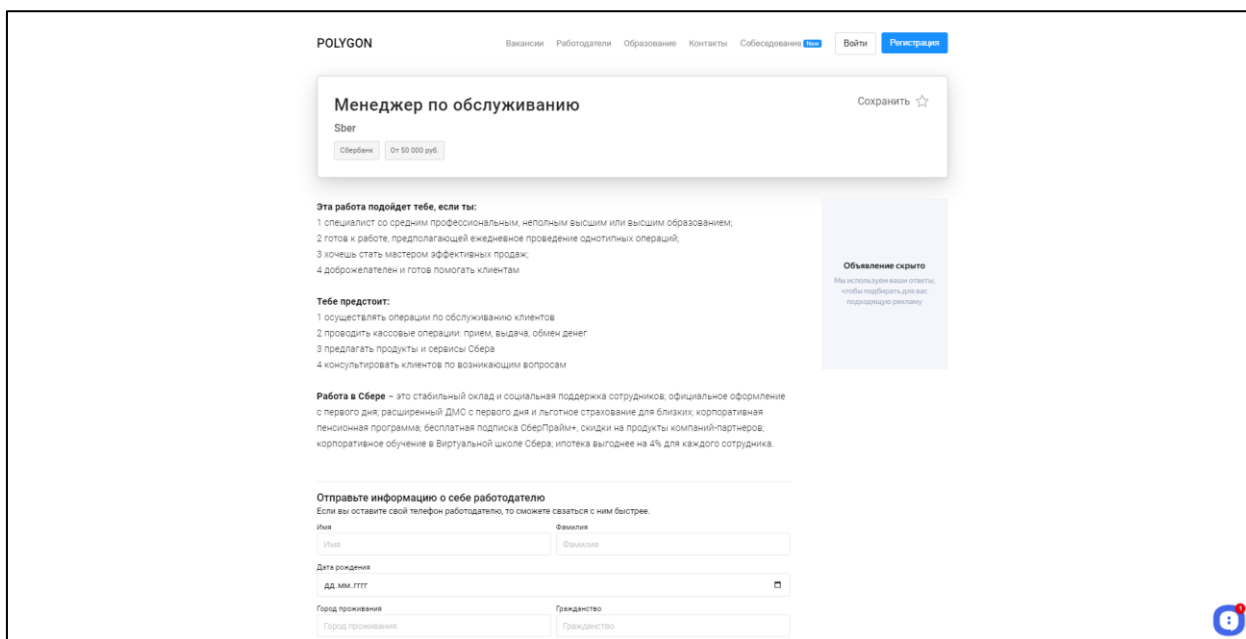


Рисунок 3.7 – Экранная форма страницы вакансии

Если пользователь представитель компании, то у него появляется возможность в своем личном кабинете добавлять, редактировать, удалять и просматривать предлагаемые его компанией вакансии (Рис. 3.8-3.10).

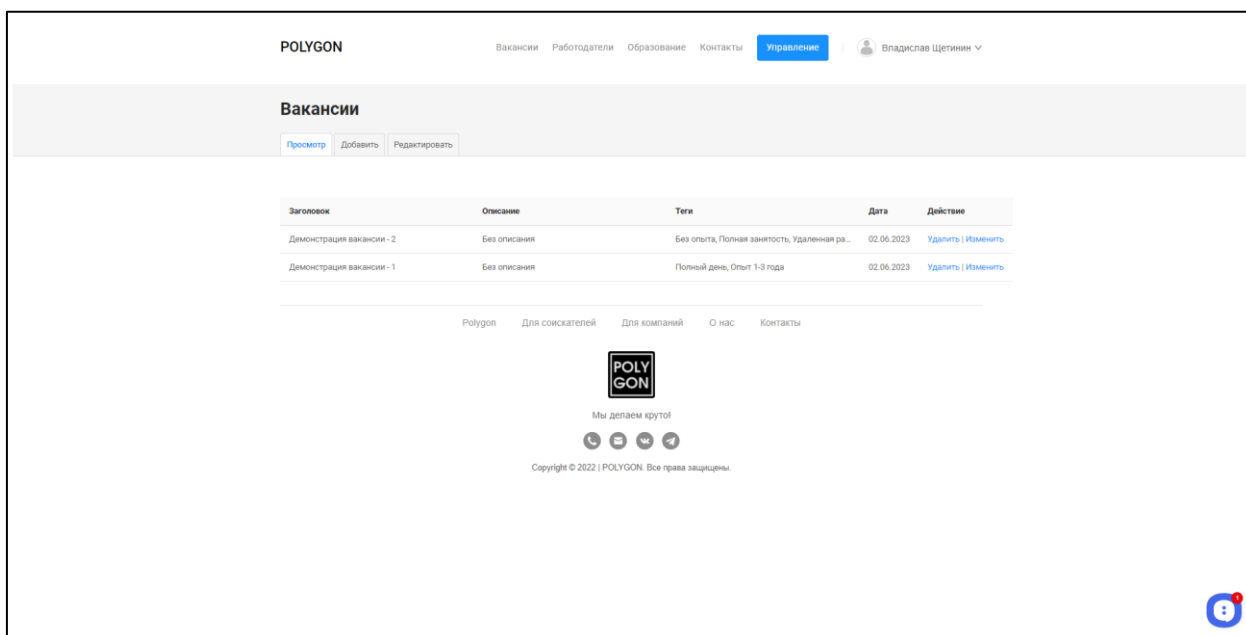


Рисунок 3.8 – Экранная форма админ-панели компании «Просмотр»

Страница на рисунке 3.8 дает работодателям просматривать созданные вакансии от их компаний, а также способность взаимодействовать с НИМИ.



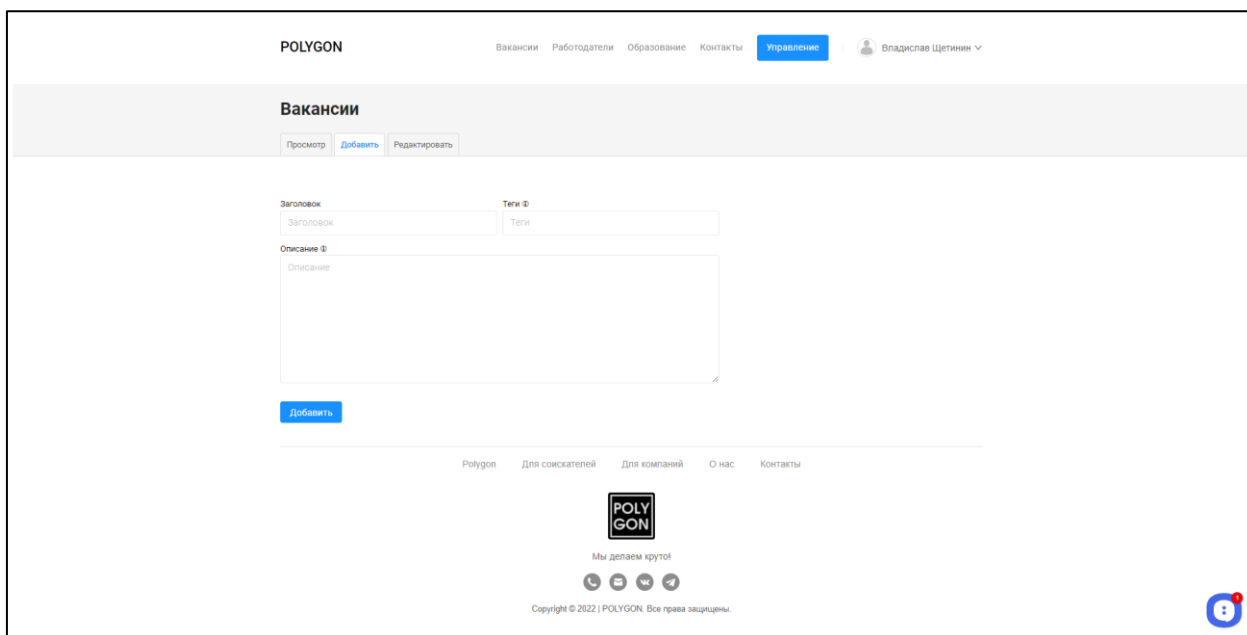


Рисунок 3.9 – Экранная форма админ-панели компании «Добавить»

На рисунке 3.9 представлена форма для добавления новой вакансии.

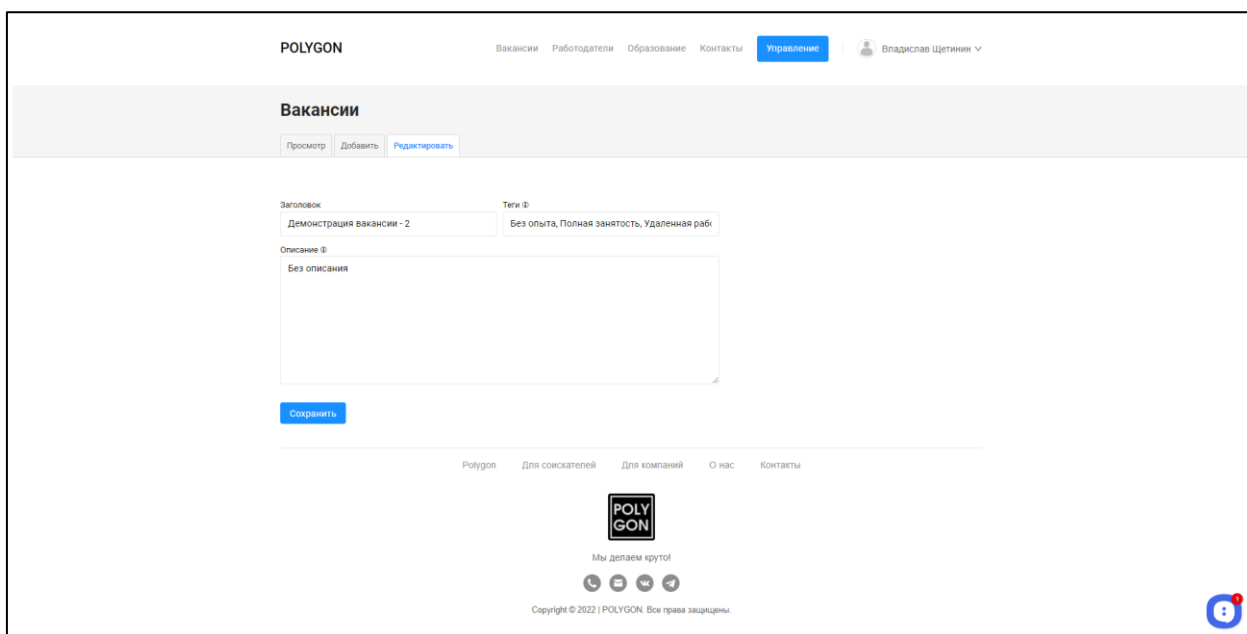


Рисунок 3.10 – Экранная форма админ-панели компании «Редактировать»

Рисунок 3.10 демонстрирует одну из возможности взаимодействия с вакансиями – редактирование (есть еще удаление). Особенность этой вкладки в том, что пользователь не может перейти в не напрямую т.к. не понятно какую вакансию нужно редактировать, поэтому переход осуществляется через ссылку «Изменить», которую можно найти при просмотре вакансий на соответствующей вкладке (Рис. 3.8).

Для всех пользователей есть возможность просматривать и редактировать свой профиль (Рис. 3.11-3.12).

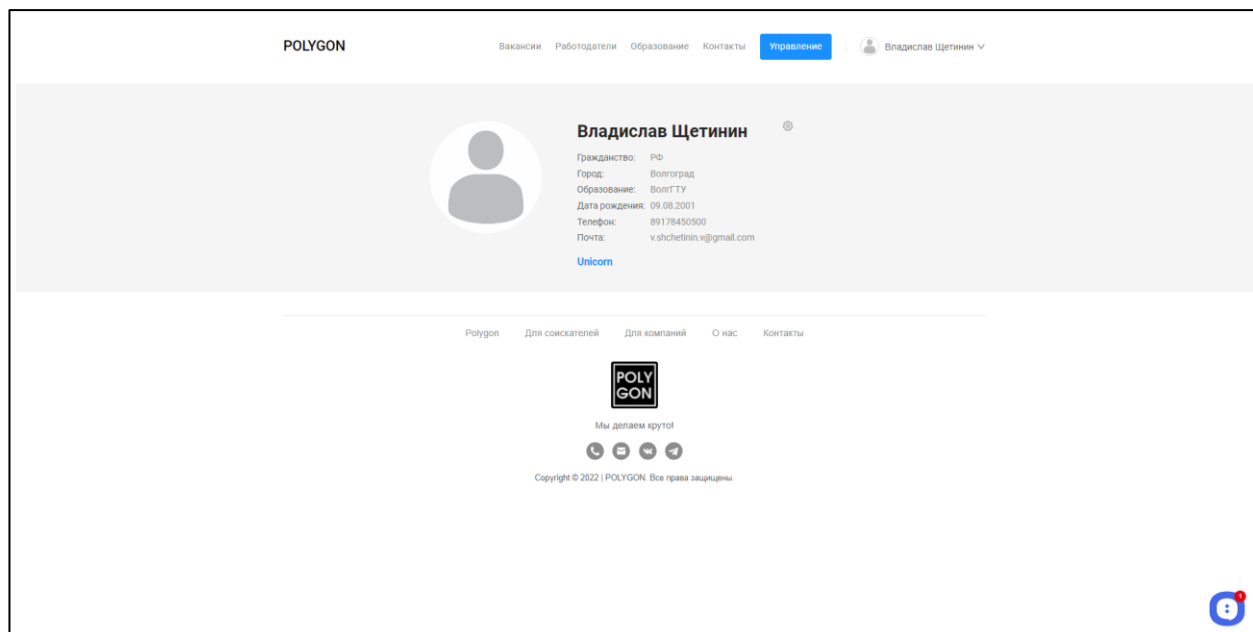


Рисунок 3.11 – Экранная форма страницы профиля

Страница профиля содержит необходимую информацию о пользователе. Также планируется на этой странице хранить резюме пользователя, на данный момент соискателям при каждом отклике на вакансию приходится по новой подгружать свое резюме (если соискатель зарегистрирован на платформе, то все необходимая информация автоматически подгружается для отправки заявки).

POLYGON

Вакансии Работодатели Образование Контакты Управление

Владислав Щетинин

### Редактирование профиля

Имя  
Владислав

Фамилия  
Щетинин

Гражданство  
РФ

Город  
Волгоград

Образование (сокращенно)  
ВолГУ

Дата рождения  
09.08.2001

Телефон  
89178450500

Сохранить Отмена

Polygon Для соискателей Для компаний О нас Контакты

POLYGON

Мы делаем круто!

Copyright © 2021 POLYGON. Все права защищены.

Рисунок 3.12 – Экранная форма страницы редактирования профиля

Также сайт имеет две посадочные страницы для двух целевых аудиторий: соискателей и компаний (Рис. 3.13-3.14).

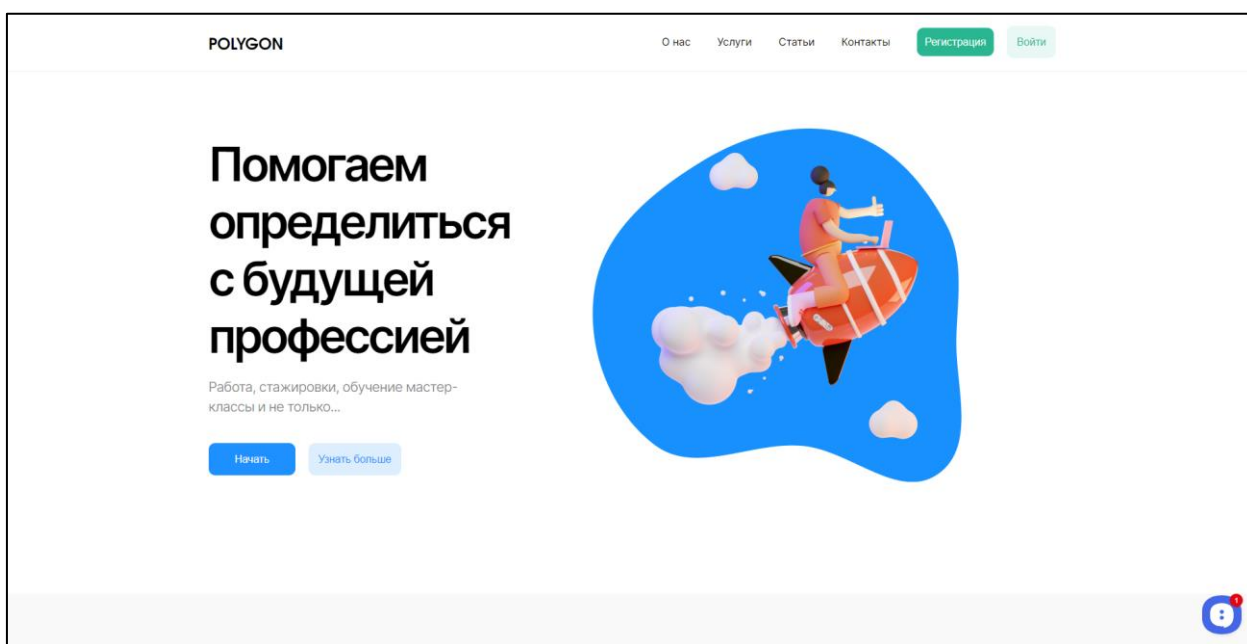


Рисунок 3.13 – Экранная форма посадочной страницы для соискателей

Посадочная страница для соискателей (Рис. 3.13) содержит необходимую информацию для одной из двух групп целевой аудитории – школьник/студенты, где описаны возможности платформы и есть ответы на возможные вопросы.

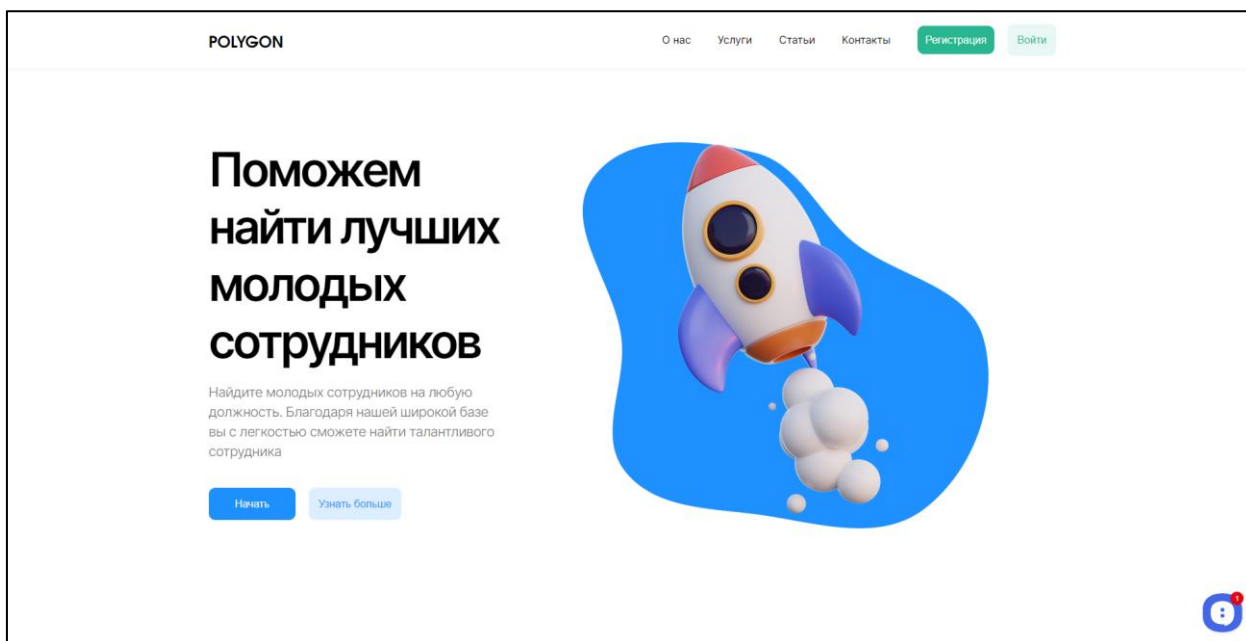


Рисунок 3.14 – Экранная форма посадочной страницы для компаний

Посадочная страница для компаний (Рис. 3.14) отвечает на все вопросы для второй группы – работодатели.

### 3.3 Оценка работы рекомендательной системы

Как показал практический опыт, запуск моделей на всем датасете (Приложение 1), является очень долгим и не оптимальным для работы. Было принято решение использовать выборку из 1000 элементов, что существенно дало результат в скорости. При эксплуатации данной модели также будет использоваться выборка из 1000 последних добавленных вакансий. Так новые вакансии будут иметь приоритет для пользователей, тем самым будут предлагаться различные и актуальные вакансии (в зависимости от количества добавлений вакансий выборка для рекомендаций может меняться).

Чтобы дать оценку правильности работы рекомендательной системы обратимся к теореме Клейнберга (теорема невозможности): Для множества объектов, состоящего из двух и более элементов, не существует алгоритма кластеризации, который был бы одновременно масштабно-инвариантным, согласованным и полным [19]. Следовательно, не существует оптимального алгоритма кластеризации. Если мы не можем математически доказать точность работы рекомендательной системы, то мы можем оценить ее влияние на

пользователя, а именно посмотреть метрики взаимодействия: просмотры предложенных вакансий, добавление их в избранное и т.п.

Также мы можем посмотреть адекватность модели, посмотрев содержание кластеров и соотнести их с профессиями, которые были определены для каждого кластера и в целом оценить схожесть вакансий. Еще необходимо сделать несколько запусков моделей, чтобы подтвердить их работоспособность и оценить качество вывода.

Анализ запусков моделей показал, что выходные данные моделей сильно отличаются при одинаковых входных данные, но при этом работа каждой модели не лишена смысла. Точность рекомендаций обеих моделей можно назвать высокой, но при этом некоторые примеры не всегда были точны, это может быть связано с недостатком вакансий на определенные темы. Но будем помнить, что это рекомендации, поэтому если они заинтересуют пользователя, то нельзя будет сказать, что данные модели не справились с задачей. Как следствие все упирается в проверку эксплуатации на платформе и оценку качества при помощи A/B тестирования.

## ГЛАВА 4. РЕЗУЛЬТАТИВНОСТЬ

Отдельные результаты выпускной (дипломной) работы публикаций не имеется, актов внедрения на предприятиях не получено.

Руководитель работы \_\_\_\_\_ (ФИО)  
   
подпись должность

					ВКР-40461806-10.12-09.03.03-462-18-2023				
		№ докум.	Подп.	Дата					
Разраб.	Щетинин В.В.				Результативность	Лит.	Лист	Листов	
Пров.	Кетько Н.В.						65	1	
						ВолгГТУ ЭП-462			
Н.контр.	Кузнецов С. Ю.								
Утв.	Скитер Н. Н.								

## ЗАКЛЮЧЕНИЕ

В данной работе была представлена рекомендательная система для рынка труда, которая подбирает наиболее подходящие вакансии для пользователя, на основе информации о просматриваемой вакансии.

В процессе проведения исследования предметной области, были рассмотрены характерные особенности взаимодействия работодателей и студентов на российском рынке труда, а также выявлены основные проблемы в этой сфере. Помимо этого, в предлагаемом исследовании представлен краткий обзор современных подходов к построению рекомендательных систем. Исходя из полученных знаний на этапе проектирования был сделан вывод использовать метод фильтрации на основе содержимого для построения рекомендательной системы.

Поскольку для проверки работы программного комплекса требуются тестовые данные, то были рассмотрены несколько вариантов источника таких данных. В результате, на основе ряда критериев в качестве ресурса с тестовыми данными (вакансий) был выбран сайт по онлайн-рекрутингу hh.ru.

Так как полученные из источника данные имеют избыточную структуру, в представленной работе была решена задача приведения их в вид, удобный для последующего анализа с помощью программных средств. В итоге, все данные, прошли процедуру первичной обработки и были преобразованы в векторное представление.

В результате для функционирования предлагаемой рекомендательной системы были спроектированы и реализованы следующие программные компоненты:

- Платформа рекрутинга среди школьников и студентов;
- Система обработки вакансий;
- Система генерации рекомендаций.

А также продуманы детали внедрения и развития модуля рекомендаций на платформе «POLYGON», включая решение проблемы холодного старта.

Использование ML технологий может помочь молодому проекту улучшить пользовательский опыт, увеличить вовлеченность пользователей, улучшить показатели активности пользователей и сделать платформу более конкурентоспособной на рынке.



## СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. Abt, B. phpmanual / B. Abt // php.net : [сайт]. – URL: <https://www.php.net/manual/en/index.php> (дата обращения: 10.06.2023).
2. DashBootstrapComponents // opensource.faculty.ai : [сайт]. – URL: <https://dash-bootstrap-components.opensource.faculty.ai/docs/> (дата обращения: 05.06.2023).
3. DashPythonUserGuide // plotly : [сайт]. – URL: <https://dash.plotly.com/> (дата обращения: 18.05.2023).
4. Documentation Natural Language Toolkit // nltk.org : [сайт]. – URL: <https://www.nltk.org/> (дата обращения: 21.05.2023).
5. Documentation of scikit-learn 0.21.3 // scikit-learn.org : [сайт]. – URL: <https://scikit-learn.org/0.21/documentation.html> (дата обращения: 22.05.2023).
6. HTML // mozilla.org : [сайт]. – URL: <https://developer.mozilla.org/ru/docs/Web/HTML> (дата обращения: 10.08.2022).
7. Industrial-Strength Natural Language Processing // spacy.io : [сайт]. – URL: <https://spacy.io/> (дата обращения: 19.06.2023).
8. JavaScript // mozilla.org : [сайт]. – URL: <https://developer.mozilla.org/ru/docs/Web/JavaScript> (дата обращения: 12.08.2022).
9. MySQL Documentation // mysql.com : [сайт]. – URL: <https://dev.mysql.com/doc/> (дата обращения: 20.08.2022).
10. NumPy Documentation // numpy.org : [сайт]. – URL: <https://numpy.org/doc/> (дата обращения: 18.05.2023).
11. pandas documentation // pydata.org : [сайт]. – URL: <https://pandas.pydata.org/docs/> (дата обращения: 21.05.2023).
12. Python 3.11.4 documentation // python.org : [сайт]. – URL: <https://docs.python.org/3/> (дата обращения: 19.05.2023).
13. scikit-learn // devdocs.io : [сайт]. – URL: [https://devdocs.io/scikit\\_learn/](https://devdocs.io/scikit_learn/) (дата обращения: 22.05.2023).

14. Shani, G. Evaluating Recommendation Systems / G. Shani, A. Gunawardana // Recommender Systems Handbook. – : Springer Nature, 2011. – С. 257–297.
15. Tfidfvectorizer, BERT, LASER: векторизация данных и кластерный анализ для улучшения рекомендательной системы // habr : [сайт]. – URL: <https://habr.com/ru/companies/lanit/articles/594759/> (дата обращения: 03.06.2023).
16. Арзамазов, Н. Как работают рекомендательные системы / Н. Арзамазов // Neurohivee : [сайт]. – URL: <https://neurohive.io/ru/osnovy-data-science/rekomendatelnye-sistemy-modeli-i-ocenka/> (дата обращения: 03.06.2023).
17. Грас, Д. DataScience. Наука о данных с нуля / Д. Грас. – 2-е изд. – Санкт-Петербург : БХВ-Петербург, 2021. – 416 с.
18. Выпускники высшего образования на российском рынке труда: тренды и вызовы : докл. к XXIII Ясинской (Апрельской) междунар. науч. конф. по проблемам развития экономики и общества, Москва, 2022 г. / Н. К. Емелина, К. В. Рожкова, С. Ю. Рошин, С. А. Солнцев, П. В. Травкин ; НИУ «Высшая школа экономики». – Москва : ИД ВШЭ, 2022. – 160 с.
19. Кластеризация // Викиконспекты : [сайт]. – URL: <http://neerc.ifmo.ru/wiki/index.php?title=Кластеризация> (дата обращения: 02.06.2023).
20. Милорадов, К. А. Проектирование интернет-приложений (курс лекций) / К. А. Милорадов // Международный журнал экспериментального образования. – 2015. – № 10. – С. 200–201.
21. Наумова, Е. Г. Организация взаимодействия работодателей и выпускников ДПИИГТУ им. Р. Е. Алексеева с целью трудоустройства / Е. Г. Наумова, Н. О. Кулигина, Н. А. Нажимова // Современные наукоемкие технологии. – 2021. – № 6. – С. 79–83.
22. Образование и бизнес: практики взаимодействия // Сетевое издание национальная система квалификаций России : [сайт]. – URL: <https://journal.nark.ru/articles/podgotovka-kadrov/obrazovanie-i-biznes-praktiki-vzaimodeystviya/> (дата обращения: 03.06.2023).

23. Опрос: работа по специальности // hh.ru : [сайт]. – URL: <https://hhcdn.ru/file/17463835.pdf> (дата обращения: 03.06.2023).
24. Ошанина, О. Как обучение студентов помогает компаниям справиться с дефицитом сотрудников / О. Ошанина // Ведомости : [сайт]. – URL: <https://www.vedomosti.ru/partner/articles/2023/03/22/967341-obuchenie-studentov> (дата обращения: 23.05.2023).
25. Польза дипломов вузов - миф или реальность? // hh.ru : [сайт]. – URL: <https://hhcdn.ru/file/17145079.pdf> (дата обращения: 23.05.2023).
26. Порядка 90% выпускников вузов устроились на работу // Министерство науки и высшего образования Российской Федерации : [сайт]. – URL: <https://minobrnauki.gov.ru/press-center/news/nauka-i-obrazovanie/56650/> (дата обращения: 23.05.2023).
27. Сидоров, А. Умный поиск: как искусственный интеллект hh.ru подбирает вакансии к резюме / А. Сидоров // habr : [сайт]. – URL: <https://habr.com/ru/companies/hh/articles/347276/> (дата обращения: 10.06.2023).
28. Скляр, Д. Изучаем PHP 7: руководство по созданию интерактивных веб-сайтов / Д. Скляр. – Москва : Диалектика-Вильямс, 2020. – 456 с.
29. Современный учебник JavaScript // javascript.ru : [сайт]. – URL: <https://learn.javascript.ru/> (дата обращения: 10.06.2023).
30. Справочник CSSй // htmlbook.ru : [сайт]. – URL: <http://htmlbook.ru/css> (дата обращения: 10.08.2022).
31. Справочник по HTML // htmlbook.ru : [сайт]. – URL: <http://htmlbook.ru/html> (дата обращения: 12.08.2022).
32. Ткаченко, А. Сравнение текста с помощью статистической меры TF-IDF / А. Ткаченко // nta : [сайт]. – URL: <https://newtechaudit.ru/kak-sverit-tekst-tf-idf/> (дата обращения: 04.06.2023).
33. Фальк, К. Рекомендательные системы на практике / К. Фальк. – 2-е изд. – Москва : ДМК Пресс, 2020. – 448 с.
34. Фримен, Э. Паттерны проектирования / Э. Фримен, Э. Фримен. – Санкт-Петербург : Питер, 2018. – 656 с.

35. Чего хотят работодатели от молодых специалистов? // i canchoose : [сайт]. – URL: <https://icanchoose.ru/blog/chego-hotyat-rabotodатели-ot-molodyh-specialistov/> (дата обращения: 23.05.2023).

	Ids	Employer	Name	Salary	From	To	Experience	Schedule	Keys	Description	Area	Professional roles	Specializations	Profarea names	Published at
0	49313809	Space307	Golang Developer (Клмп)	True	251322.0	NaN	От 3 до 6 лет	Полный день	[Docker, Golang, Redis, Английский язык...	Мы в Space307 разрабатываем международную тор...	Санкт-Петербург	[Программист, разработчик]	[Программирование, Разработка]	[Информационные технологии, интернет, телеком]	2021-12-02 12:15:37+03:00
1	48813842	Монополия	E-mail маркетолог	True	60900.0	NaN	От 1 года до 3 лет	Полный день	[Грамотность, Написание текстов, Грамотна...	С 2015 года наш IT блок меняет рынок рекламы автопарк...	Санкт-Петербург	Менеджер по маркетингу и рекламе]	[Маркетинг]	[Информационные технологии, интернет, телеком]	2021-12-02 10:33:15+03:00
2	49413720	Eden Springs	Оператор call-центра (удаленно)	False	NaN	NaN	От 1 года до 3 лет	Удаленная работа	[Клиентоориентированность, Ориентация на ре...	Что нужно будет делать: Принимать входящие зв...	Санкт-Петербург	[Оператор call-центра, специалист контактного ...]	[Маркетинг, Продажи по телефону, Телемаркет...	[Продажи, Информационные технологии, интернет, телеком]	2021-12-02 10:29:37+03:00
3	46460892	Import Hoyu	Ведущий SMM специалист	True	60000.0	80000.0	От 1 года до 3 лет	Полный день	[Продвижение бренда, Креативность, Adobe ...	В данный момент мы ищем в нашу команду самого ...	Санкт-Петербург	[SMM-менеджер, контент-менеджер]	[Управление маркетингом, PR, Маркетинговые ...	[Информационные технологии, интернет, телеком]	2021-12-01 16:57:02+03:00
4	49555567	Pride Games Studio	UX/UI Designer	False	NaN	NaN	От 1 года до 3 лет	Полный день	[UI, UX, gamedev, game design, проект...	Pride Games Studio — это команда единомышленни...	Санкт-Петербург	[Дизайнер, художник]	[Игровое ПО, Программирование, Разработка ...]	[Маркетинг, реклама, PR, Информационные тех...	2021-12-01 16:48:24+03:00
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
48559	49108967	A-N SECURITY	Монтажник слабых систем	True	60000.0	NaN	От 1 года до 3 лет	Полный день	[Мобильность, Работа в команде, Системы о...	Обязанности: ведение технической документации...	Москва	[Сервисный инженер, механик]	[Передача данных и доступ в интернет, Сетев...	[Рабочий персонал, Информационные технологии...]	2021-10-25 11:40:23+03:00
48560	49108929	Филиал ФКУ Нагор-Сервис ОПС России в г. Москва	Системный администратор	True	60000.0	NaN	От 1 года до 3 лет	Полный день	[Windows 7, Windows 8, Windows 10, Unix...	Функциональные обязанности: Сопровождение опе...	Москва	[Системный администратор]	[Системный администратор, Инженер, Поддер...	[Информационные технологии, интернет, телеком]	2021-10-25 11:39:41+03:00
48561	47462568	Market Warrior	Менеджер по работе с маркетинг-каси (офис)	True	50000.0	NaN	От 1 года до 3 лет	Полный день	[Работа с большими объемам информации, Управ...	Мы молодая команда, занимающаяся помощью в раб...	Москва	[Координатор отдела продаж]	[Управление проектами]	[Информационные технологии, интернет, телеком]	2021-10-25 11:36:52+03:00
48562	48713091	Кон Трон	Менеджер по продажам (Серверное оборудование, ...)	True	120000.0	300000.0	От 1 года до 3 лет	Полный день	[B2B Продажи, Ведение переговоров, CRM, ...	GAGARIN (ГК &quot;Cos. Trolis&quot;) - рос...	Москва	[Менеджер по продажам, менеджер по работе с к...	[Телекоммуникации, Сетевые решения, Услуги ...	[Информационные технологии, интернет, телеком]	2021-10-25 11:38:34+03:00
48563	49108761	Vivid Money	Product Analyst	False	NaN	NaN	От 3 до 6 лет	Удаленная работа	[SQL]	We are an innovative FinTech start-up with the...	Москва	[Менеджер продукта]	[Менеджмент продукта (Product manager), Упр...	[Информационные технологии, интернет, телеком]	2021-10-25 11:36:59+03:00

Cluster 0: технология информационный интернет работа управление телек персонал sap менеджер технический	Cluster 1: devops системный инженер администратор linux docker kubernetes ansible engineer технология	Cluster 2: проект управление руководитель project группа разработка технология информационный консультирование интернет	Cluster 3: инженер конструктор проектировщик производство сельский хозяйство станок оператор промышленность чпу	Cluster 4: данные база sql oracle администратор ms разработчик программист pl разработка
Cluster 5: php разработчик web mysql программист программирование разработка инженер git laravel	Cluster 6: поддержка технический специалист helpdesk начальный мало уровень опыт инженер администратор	Cluster 7: разработчик программист программирование разработка developer python технология интернет информационный телек	Cluster 8: маркетинг реклама pr менеджер маркетинговый интернет seo контент marketing коммуникация	Cluster 9: продажа менеджер клиент работа переговоры прямой b2b развитие активный ведение
Cluster 10: тестирование тестировщик qa engineer инженер sql automation технология интернет информационный	Cluster 11: администратор системный сетевой администрирование технология поддержка helpdesk linux интернет инженер	Cluster 12: дизайнер дизайн художник adobe графика живопись развлечение масса искусство медиа	Cluster 13: аналитик бизнес анализ sql ms технология информационный интернет телек данные	Cluster 14: сервисный инсталляция инженер оборудование настройка механик сервис монтажник телекоммуникация рабочий
Cluster 15: net разработчик asp framework программист core sql программирование разработка web	Cluster 16: java spring разработчик framework программист программирование разработка developer инженер hibernate	Cluster 17: продукт product manager менеджер проект управление менеджмент маркетинг pr реклама	Cluster 18: безопасность специалист информационный компьютерный инженер сетевой технология экономический системный администратор	Cluster 19: аналитик системный управление интеграция проект erp система предприятие разработка программирование
Cluster 20: web инженер мастер разработчик программист erp предприятие система управление программирование	Cluster 21: 1c программист предприятие управление разработчик программирование erp система разработка бухгалтерия	Cluster 22: react javascript js разработчик frontend программист node typescript web программирование		