Санкт-Петербургский государственный университет

***БРАЖНИКОВ Филипп Викторович***

**Выпускная квалификационная работа**

***Моделирование спроса на навыки и компетенции на рынке труда***

***Modeling the demand for skills and competences in the labor market***

Уровень образования: Магистратура

Направление 01.04.02 «Прикладная математика и информатика»

Основная образовательная программа   
ВМ.5754 «Математические методы цифровизации экономики»

Научный руководитель,

д.ф.-м.н., профессор

Прасолов А.В.

Рецензент,

ГУП «Санкт-Петербургский

информационно-аналитический центр»

Замураев К.А.

Санкт-Петербург

2022

Содержание

[Введение 3](#_Toc103004113)

[Постановка задачи 5](#_Toc103004114)

[Анализ литературы 8](#_Toc103004115)

[Глава 1. Предварительная обработка данных 16](#_Toc103004116)

[1.1. Набор данных 16](#_Toc103004117)

[1.2. Обработка пропусков в данных 22](#_Toc103004118)

[1.3. Оценка результатов 24](#_Toc103004119)

[Глава 2. Моделирование спроса на навыки и компетенции 28](#_Toc103004120)

[2.1. Количественная оценка спроса 28](#_Toc103004121)

[2.2. Моделирование спроса на навыки и компетенции 30](#_Toc103004122)

[2.3. Оценка результатов 33](#_Toc103004123)

[Глава 3. Анализ ситуации на рынке труда Санкт- Петербурга 36](#_Toc103004124)

[3.1. Анализ спроса на навыки и компетенции 36](#_Toc103004125)

[3.2. Рекомендации по совершенствованию государственной политики в области обеспечения занятости 43](#_Toc103004126)

[Выводы 47](#_Toc103004127)

[Заключение 50](#_Toc103004128)

[Список литературы 51](#_Toc103004129)

[Приложение 1 54](#_Toc103004130)

[Приложение 2 55](#_Toc103004131)

[Приложение 3 57](#_Toc103004132)

[Приложение 4 60](#_Toc103004133)

[Приложение 5 62](#_Toc103004134)

Введение

Цифровизация и проникновение технологий во все виды экономической деятельности сопряжены с быстрыми темпами технологических и структурных изменений, являются глобальными процессами и отражают современные тенденции в экономике и обществе. Важность происходящих процессов очевидна для любого государства, и Россия не исключение. Реализуемая национальная программа «Цифровая экономика Российской Федерации» направлена на «…решение задачи по обеспечению ускоренного внедрения цифровых технологий в экономике и социальной сфере» [11]. Одной из ключевых целей данной программы является подготовка кадров для цифровой экономики.

Риски нехватки квалифицированных кадров, обладающих необходимыми навыками, вызывают особый интерес к теме кадрового обеспечения трансформирующейся цифровой экономики. Во-первых, быстрые темпы разработки и совершенствования технологий способствуют формированию и увеличению разрыва в развитии – предприятия и работники не могут быстро адаптироваться к постоянно меняющимся условиям. Существующий цифровой разрыв во многом связан отсутствием знаний и навыков, необходимых для доступа и использования технологий. Во- вторых, цифровая трансформация рынка труда связана как с созданием новых рабочих мест, так и с уничтожением существующих. Это также означает, что значительная часть трудовых функций претерпит изменения либо перестанет существовать в привычном виде. В-третьих, в современном мире успех отдельного человека будет определяться сочетанием сложных технических навыков, специфичных компетенций и знаний, необходимых для эффективного выполнения работы.

Пандемия коронавирусной инфекции ускорила цифровую трансформацию и подчеркнула важность технологий для преодоления ограничений на рынке труда и обеспечения функционирования системы обучения [16]. Удаленная и дистанционная работа, освоение новых программных продуктов, необходимость всегда «быть онлайн» – доминанты, без которых нельзя представить современный процесс организации работы.

Происходящие процессы цифровой трансформации обуславливают актуальность настоящей работы. Важным исследовательским направлением выступает изучение существующего разрыва между навыками, необходимыми на рынке труда, и навыками, предлагаемыми рабочей силой. Проблемы изучения спроса и предложения навыков на рынке труда связанны как со сложностью обобщения результатов эмпирических тематических исследований, так и с отсутствием всеобъемлющих данных о навыках и трудностями их измерения.

Таким образом, построение моделирование спроса на навыки и компетенции является актуальной исследовательской задачей. Результаты представленного моделирования могут быть в дальнейшем использованы для прогнозирования потребности в навыках и минимизации несоответствий их спроса и предложения.

Постановка задачи

Происходящие процессы цифровой трансформации затрагивают как вопросы функционирования рынка труда, так и определения направлений государственной политики в области обеспечения занятости населения. Моделирование спроса на навыки и компетенции позволяет получить актуальную информацию о состоянии и тенденциях рынка труда, основными пользователями которой выступают государственные органы власти.

На сегодняшний день государственные органы власти обладают широким кругом полномочий в области содействия занятости населения, которые предусматривают, в том числе, определение направлений обучения и дополнительного профессионального образования граждан. В связи с этим, при планировании программ профессионального обучения полезно использовать информацию о наиболее востребованных на рынке навыках и профессиях.

Существует несколько подходов к моделированию и прогнозированию спроса на навыки, которые позволяют получить достаточно точные результаты и предоставить полезную информацию для прогнозирования будущей потребности в навыках. Наиболее интересным представляется подход, в рамках которого моделирование спроса осуществляется на основе навыков и компетенций, которые встречаются в онлайн-объявлениях о вакансиях.

Использование онлайн-объявлений о вакансиях в качестве источника данных является новым исследовательским направлением. Публикуемые вакансии содержат ключевые навыки и требования к работе, а их извлечение является задачей обработки естественного языка. В конечном счете мы имеем возможность наблюдать навыки на уровне каждой вакансии. Навыки и компетенции могут быть оценены по профессиональным группам, отраслям.

При использовании данных из вакансий для моделирования спроса на навыки возникают некоторые ограничения. Прежде всего, недостаточная проработанность подходов в литературе по схожим исследовательским задачам. Другим ограничением выступают сами данные: неточности, пропуски, формулировки. Несмотря на это, использование информации о спросе на навыки обеспечивает большую гибкость при прогнозировании потребности в навыках

Цель исследования – разработка модели оценки спроса на навыки и компетенции на рынке труда.

В рамках исследования планируется решение следующих задач:

1. анализ теоретической и эмпирической литературы, посвященной моделированию спроса на навыки и компетенции;
2. формирование информационной базы исследования;
3. предварительная обработка данных;
4. построение модели и проведение эксперимента по восстановлению пропущенных значений в данных;
5. построение модели спроса на навыки и компетенции, а также оценка результатов моделирования;
6. анализ ситуации на рынке труда Санкт-Петербурга и выработка рекомендаций.

Объект исследования – региональный рынок труда Санкт-Петербурга.

Гипотеза исследования – использование альтернативных источников данных о рынке труда вместе с современными методами обработки данных позволят получить потенциально полезную информацию, которую могут использовать государственные органы при реализации политики в области обеспечения занятости населения.

Настоящее исследование носит прикладной характер, научная новизна которого заключается в использовании информации из онлайн- объявлений о поиске работы в качестве аппроксимирующего показателя спроса. Использование альтернативных источников данных для моделирования во многом связано с проблемами качества и полноты имеющихся статистических данных о рынке труда.

Использование в качестве источника данных информации из онлайн- объявлений о вакансиях является новым направлением эмпирических исследований. Остаются открытыми вопросы обработки, обеспечения репрезентативности и сбалансированности исходных данных. Кроме того, не до конца проработаны методы оценки важности навыков и компетенций, а также способы учета этих показателей при моделировании. В рамках исследовательской работы будут предложены подходы, позволяющие ответить на наиболее актуальные вопросы.

Анализ литературы

Эмпирическую и методологическую базу настоящего исследования составляют современные работы, посвященные анализу и моделированию спроса и предложения на навыки на рынке труда. Основное внимание в работах уделяется вопросам трактовки понятия рынка труда, определению места навыков и компетенций в системе рыночных отношений, выбору способов классификации навыков, а также моделированию спроса на навыки.

В современной экономической литературе существует множество подходов к определению понятия рынка труда. С одной стороны, различия обусловлены историческими этапами развития экономической мысли, а с другой стороны, отдельными исследовательскими направлениями, охватывающими различные аспекты трудовой деятельности. Принято рассматривать понятие рынка труда в широком и узком смысле.

В узкой трактовке рынок труда рассматривается через особенности функционирования рыночного механизма. По определению Кубищина E.C «Рынок труда – это специфический вид рынка, представляющий собой форму проявления рыночных отношений в сфере обращения и функционирования рабочей силы (трудового фактора экономики)» [7, с. 29]. В данном определении подчеркивается универсальность рыночного механизма – существует и функционирует множество рынков, на каждом из которых есть покупатели и продавцы (покупатели –работодатели, а продавцы – работники). Также на рынке присутствует специфический товар – рабочая сила. Таким образом, рынок труда рассматривается на ряду с другими рынками (капитала, товаров, услуг, финансов и др.), которые обладают схожими функциями и вместе составляют основу любой современной экономической системы.

Усложнение трудовых отношений, повышение значимости социально-экономических факторов способствовали рассмотрению понятия в более широкой трактовке. К примеру, Дроздов О.А. определяет рынок труда как «…систему социально-экономических отношений, возникающих между работодателями, формирующими спрос на рабочую силу, и наемными работниками, предлагающими свои услуги труда» [2, c. 17]. Другими исследователями рынок труда определяется как «…совокупность социально-экономических отношений между государством, работодателями и работниками по вопросу купли-продажи рабочей силы, обучения работников и использования их в процессе производства» [9, с. 7]. Таким образом, в данных определениях подчеркивается важность экономических, социальных, правовых и институциональных факторов, которые имеют решающее значения для описания процессов и механизмов взаимодействия на рынке труда.

Для того, чтобы отразить различные подходы к определению рынка труда, отдельные исследователи пытаются дать «глобальное» определение понятию. Так по определению Рофе А. И. «Рынок труда – это система общественных отношений, связанных с наймом и предложением труда, т.е. с его куплей и продажей; это также экономическое и географическое пространство – сфера трудоустройства, в которой взаимодействуют покупатели и продавцы специфического товара – труда; наконец это механизм, обеспечивающий согласование цены и условий труда между работодателями и наемными работниками» [8, c. 138]. Такое определение позволяет выделить три существенных элемента понятия – систему отношений, некоторое экономическое пространство отношений и механизм функционирования.

В рамках настоящего исследования под рынком труда понимается система социально-экономических отношений между субъектами рынка (работодателями, работниками и государством), обеспечивающая согласование и координацию рыночного механизма – взаимодействия спроса и предложения рабочей силы.

На сегодняшний день существуют различные подходы к определению понятия навыков и их места в системе рыночных отношений. Существующие различия в трактовках понятия «навык» во многом связаны с близостью терминологии в смежных областях знания. Очевидно, что социологи и экономисты склонны по-разному трактовать данное понятие. Как следствие, в эмпирической литературе понятие навыка переплетается с понятиями знания, опыта, компетенции, квалификации. Среди российских исследователей рынка труда отсутствует единая точка зрения на данный вопрос.

Одно из первых определений понятия навыка принадлежит британскому инженеру Гансу Ренольду, который в 1928 г. определил понятие навыка как «…любую полезную для промышленности комбинацию умственных и физических качеств, для приобретения которых требуется значительная тренировка» [24]. В данном определение на ряду с когнитивными аспектами трудовой деятельности делается акцент на факторе обучения, хотя и игнорируются другие способы овладения навыками (например, практика). Можно сказать, что данное определение задало важный вектор трактовки определения понятия навыка. С точки зрения экономики классическая трактовка предполагает использование понятия навыка для обозначения уровня производительности, который характеризуется точностью и скоростью выполнения конкретных задач.

В то же время нельзя упускать из виду широкий пласт работ, в которых формируются концептуальные основы изучения знаний, навыков и компетенций. Резюмируя различные подходы, можно определить связи между понятиями в виде простой схемы «знания-навыки-компетенции». Интеллектуальные способности необходимы для развития знаний, практическая реализация знаний является частью развития навыков, а все в совокупности является предпосылками к развитию компетенций. Другим подходом является выделение ряда параметров, которые определяют степень компетенции человека. К примеру, в данной работе [29] выделяется семь параметров, отражающих степень компетенции: способности, знания, понимание, навыки, действия, опыт и мотивация.

В целях настоящей работы понятия навыков и компетенций рассматриваются в широкой трактовке. Навыки – это развитые способности, необходимые работнику для эффективного выполнения профессиональных функций и задач. Компетенции – это способности, которые позволяют использовать знания, навыки и личные качества в целях эффективного выполнения профессиональных функций и задач. Для упрощения и обобщения мы проводим незначительные различия между двумя понятиями, рассматривая навыки и компетенции как универсальную единицу.

Как и в случае с определением понятия навык, существуют различные способы выделения и классификации конкретных навыков и компетенций. Простая классификация предполагает разделение на универсальные и специфические. Универсальные навыки – это навыки, которые не зависят от контекста работы и применимы для решения широкого круга задач. К ним можно отнести базовые навыки – грамотность, коммуникативные навыки, навыки презентации и др. В свою очередь специфические навыки и компетенции зависят от контекста работы и предусматривают глубокое понимание предметной области. Проблема такой классификации состоит в том, что она не обеспечивает высокий уровень детализации при определении ключевых навыков и компетенций. Поэтому в современных эмпирических работах широко распространены два похода для классификации, обеспечивающие рассмотрение различных групп навыков и компетенций.

Первый подход базируется на использовании исходных данных для формирования списков навыков или построения иерархий. Большинство исследователей стремится выделить отличительные черты (ключевые слова и фразы) из профессиональных дескрипторов или объявлений о вакансиях, на основании которых происходит классификация. К примеру, Д. Деминг и Л. Б. Кан классифицируют навыки на основании описания 40 тыс. онлайн-объявлений о вакансиях [18]. В работе авторы выделяют ключевые слова и фразы из вакансии и на их основании формируют десять групп навыков, среди которых когнитивные, социальные, управленческие, компьютерные и др. Такой способ нашел широкое отражение в работах, посвященных анализу навыков на основании онлайн-объявлений о вакансиях [20, 21]. В то же время отдельные исследователи, использующие данные онлайн-объявлений о вакансиях, не формируют отдельные укрупненные группы навыков, а используют всю информацию для оценки и моделирования [18]. Преимуществами такого способа построения классификации навыков являются актуальность получаемых результатов и возможность использования инструментария интеллектуального анализа данных. В качестве недостатка можно выделить сложность обобщения полученных результатов для других данных.

Вторым подходом выступает использование существующих иерархических классификаций, которые разработаны государственными органами различных стран для осуществления политики в области прогнозирования потребностей в навыках. Особого внимания заслуживают иерархические классификации, разработанные Министерством труда США (O\*NET), Министерством занятости и социального развития Канады и Европейской комиссией (ESCO). O\*NET представляет собой иерархическую структуру, в основании которой лежит восемь основных категорий навыков, которые используются на рынке труда для описания требований к работе и личных характеристик соискателей [25]. Министерство занятости и социального развития Канады использует собственную иерархическую классификацию навыков и компетенций [26], которая также формируется с использованием различных профессиональных дескрипторов. В рамках классификации представлено описано свыше 250 навыков и компетенций, имеющих детальное описание. На территории Европейского союза используется многоязычная классификация европейских навыков, компетенций, квалификаций и занятий (European Skills, Competences, Qualifications and Occupations, ESCO), которая разработана на основе иерархий O\*NET и Министерства занятости и социального развития Канады [27]. В рамках иерархической классификации описано 13 485 навыков. Нужно сказать, что описанные выше иерархические классификации в основном используются государственными органами и организациями для моделирования потребности в навыках. Несомненным преимуществом данного подхода является легкость обобщения и сравнения результатов для различных исследований.

Моделирование спроса на навыки и компетенции является непростой задачей, так как на сегодняшний день отсутствуют устоявшиеся подходы к ее решению. В то же время существует несколько подходов к моделированию и прогнозированию спроса на навыки, которые позволяют получить достаточно точные результаты и предоставить полезную информацию для осуществления политики в области прогнозирования потребностей в навыках. На сегодняшний день выделяют три основных подхода к прогнозированию востребованных навыков [14]:

* прогнозирование занятости по профессиям и сопоставление этих данных с существующими иерархическими классификаторами навыков;
* прогнозирование на основе навыков, встречающихся в онлайн- объявлениях о вакансиях;
* экспертная оценка основных направлений развития навыков и изменения спроса.

Прогнозирование занятости по профессиям и сопоставление этих профессий с навыками основывается на хорошо зарекомендовавших себя методах прогнозирования. Их главное преимущество заключается в том, что уровень занятости в разбивке по профессиям отслеживается статистическими органами. Кроме того, прогнозные модели могут использовать другие экономические переменные, такие как ВВП, взаимосвязь которых с занятостью широко признана и понятна. Например, данный подход используется территории Европейского союза при прогнозировании потребности в навыках. Для прогнозирования занятости по профессиям используется мульти- секторальная макроэкономическая модель, после чего происходит сопоставлению профессий с иерархической классификацией навыков ESCO [15]. В результате любой прогнозируемый рост или снижение спроса на навыки в будущем полностью зависит от траекторий трудоустройства по основным профессиям. Недостатком данного подхода является то, что прогноз не отражает возможного изменения с течением времени требуемых навыков в рамках профессии.

Второй подход предполагает использование онлайн- объявлений о вакансиях в качестве нового источника данных. Необходимые навыки и требования к работе указаны в вакансиях, а их извлечение осуществляется с использованием различных алгоритмов обработки естественного языка (Natural Language Processing, NLP). Таким образом, навыки непосредственно наблюдаются на уровне каждой вакансии и могут быть оценены по профессиям, отраслям, регионам или экономике в целом. При использовании данных из вакансий при прогнозировании навыков важно понимать существующие ограничения. Во-первых, возможны ситуации, когда требования к работе очевидны и их описание не включается в вакансии. Во- вторых, возможны искажения в данных. В-третьих, выбор подходов к предварительной обработке данных во-много зависит от выбора конкретного исследователя. Несмотря на это, прямые наблюдения за востребованными навыками обеспечивают большую гибкость при прогнозировании потребности в навыках, а также выборе направлений реализации политики в области занятости и образования.

Последний подход базируется на прогнозах экспертов о будущей траектории занятости в отдельных профессиях. Экспертные прогнозы широко используются в качестве ключевого источника для моделирования более широких тенденций на рынке труда и в области навыков. Основная идея – эти прогнозы объединяют широкий спектр знаний, которые могут учитывать тенденции, которые трудно оценить количественно. Кроме того, качественные прогнозы могут учитывать экономические сбои, а также будущие структурные сдвиги. В качестве примера можно привести «Атлас новых компетенций» [4], который представляет собой альманах перспективных отраслей и профессий, которые будут востребованы на рынке труда в ближайшие 15-20 лет. Он сформирован на основе данных форсайт- сессий, в которых приняло участи более 4 тыс. экспертов. Несомненно, для построения точных прогнозов требуется высокая квалификация экспертов и проработанная методология, что является некоторым ограничением данного подхода.

Вышеописанные подходы к определению места навыков на рынке труда выступают в качестве методологических и концептуальных рамок настоящего исследования. Основные выводы, сформулированные на основании анализа источников, используются для решения поставленных исследовательских задач.

Глава 1. Предварительная обработка данных

1.1. Набор данных

В рамках настоящего исследования в качестве основного источника данных о рынке труда используется информация из онлайн-объявлений о вакансиях. Набор данных был сформирован с использованием существующих API и методов веб-скрапинга (web scraping). Данный набор содержит подробную информацию о 829 200 вакансиях, которые были размещены на платформах и онлайн-сервисах по поиску работы за период с 2016 по 2018 г. В наборе представлены только те вакансии, в которых в качестве региона работы указан Санкт- Петербург.

Данные содержат поля с полнотекстовым описанием вакансий и дополнительные атрибуты, такие как уровень заработной платы, необходимый опыт, ключевые навыки и компетенции и др. Вакансии в наборе представлены неравномерно, большинство вакансий рассчитано на специалистов, обладающих высокой квалификацией («беловоротничковые профессии»). Существующая несбалансированность в данных во многом обусловлена естественными причинами – наличием альтернативных вариантов размещения поиска рабочей силы (объявления в газетах, корпоративные порталы), а также доступность альтернативных сервисов по поиску работы.

Другими важными недостатками исходных данных являются неточности описания и пропуски (отсутствующие значения). Полнота и детализация в описании вакансий во многом зависят от потенциального объема информации, которые работодатели готовы открыть для внешнего пользователя. Все это можно считать ограничениями, возникающими при работе с данными вакансий.

Несбалансированность данных, неточности описания и отсутствующие значения являются существенными проблемами при работе с исходными данными и моделировании спроса на навыки и компетенции. Поэтому этап предварительной обработки данных направлен на устранение вышеописанных проблем.

Этап предварительной обработки данных можно представить в виде схемы (Рис. 1). Этап предобработки предусматривает решение нескольких задач. Во-первых, формирование выборки исследования, на основании которой будет осуществлен анализ и моделирование спроса. Во-вторых, очистка и устранение неточностей в тестовом описании вакансий, а также ключевых навыков и компетенций. В-третьих, ряд экспериментов, направленных на восстановление пропущенных значений в данных.

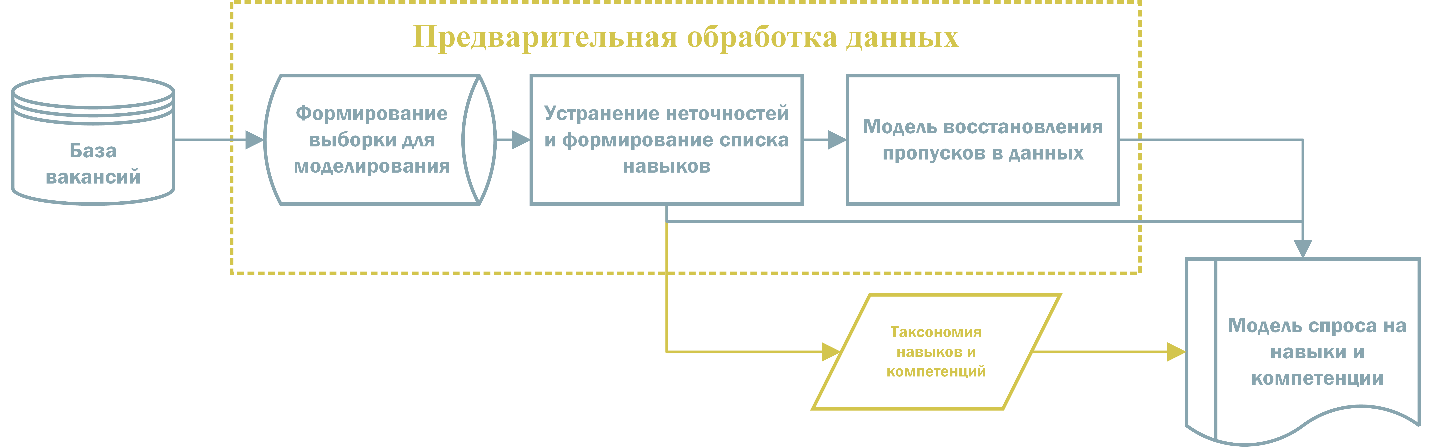


Рис. 1 – Основные этапы предварительной обработки данных

Источник: составлено автором

Одним из способов преодоления проблемы несбалансированности данных является рассмотрение некоторого подмножества исходных данных, при условии, что такое подмножество обеспечит достижение поставленных целей. Поэтому в дальнейшем мы будем рассматривать некоторую выборочную совокупность, в которую попали более однородные профессиональные группы.

На начальном этапе исследования была сформирована выборочная совокупность, в которой представлено 78 350 вакансий, относящихся к сфере информационных технологий, интернета и телекома. Данные представлены набором следующих полей*:* job\_id, job\_title, skills\_required, salary, employment\_type, min\_experience, job\_description, last\_update\_date.

Отраслевая выборка обеспечивает сбалансированность данных, упрощает процесс сравнения и обобщения информации из вакансий. В выборку были отобраны широкие группы профессий в сфере информационных технологий – вакансии разработчиков, специалистов по тестированию, специалистов в области данных, системных аналитиков, бизнес-аналитиков, менеджерские и другие. Не были включены вакансии специалистов в сфере продаж (продаж IT-продуктов), маркетинга, PR, рекламы, бухгалтерского учета и других «сквозных» профессий. Данные группы профессий сложно отнести к конкретной сфере деятельности, так как они являются универсальными и широко распространены во всех отраслях экономики.

На следующем шаге осуществлялась очистка и устранение неточностей в текстовых полях набора данных. Прежде всего обрабатывались описания указанных ключевых навыков и компетенций. Основная цель – сформировать список навыков и компетенций, который будет выступать в качестве некоторой таксономии для нашей выборки. Во-первых, исключались нерелевантные требования – «иметь голову», «уметь работать» и т. п. Во-вторых, устранение неточностей описания и приведение данных к единому шаблону. К примеру, навыки «1С: Предприятие 8.3» и «1С: Предприятие 8» объединялись в одну группу 1С: Предприятие. Другим примером является ситуация, когда одинаковые навыки описаны различными способами – «JavaScript» и «JS» объединялись в одну группу JavaScript.

Процесс обработки описания осуществлялся итерационно – на каждой итерации формировались шаблоны и паттерны, используемые для сопоставления последовательностей символов в строках (Регулярные выражения). На последней итерации производилась экспертная оценка полученных результатов и формировался список ключевых навыков и компетенций. Таким образом, был получен список из 2 409 уникальных навыков, которые характерны для сферы информационных технологий, интернета и телекома. На Рис. 2 представлена визуализация ключевых навыков и компетенций, характерных для сферы информационных технологий, интернета и телекома.

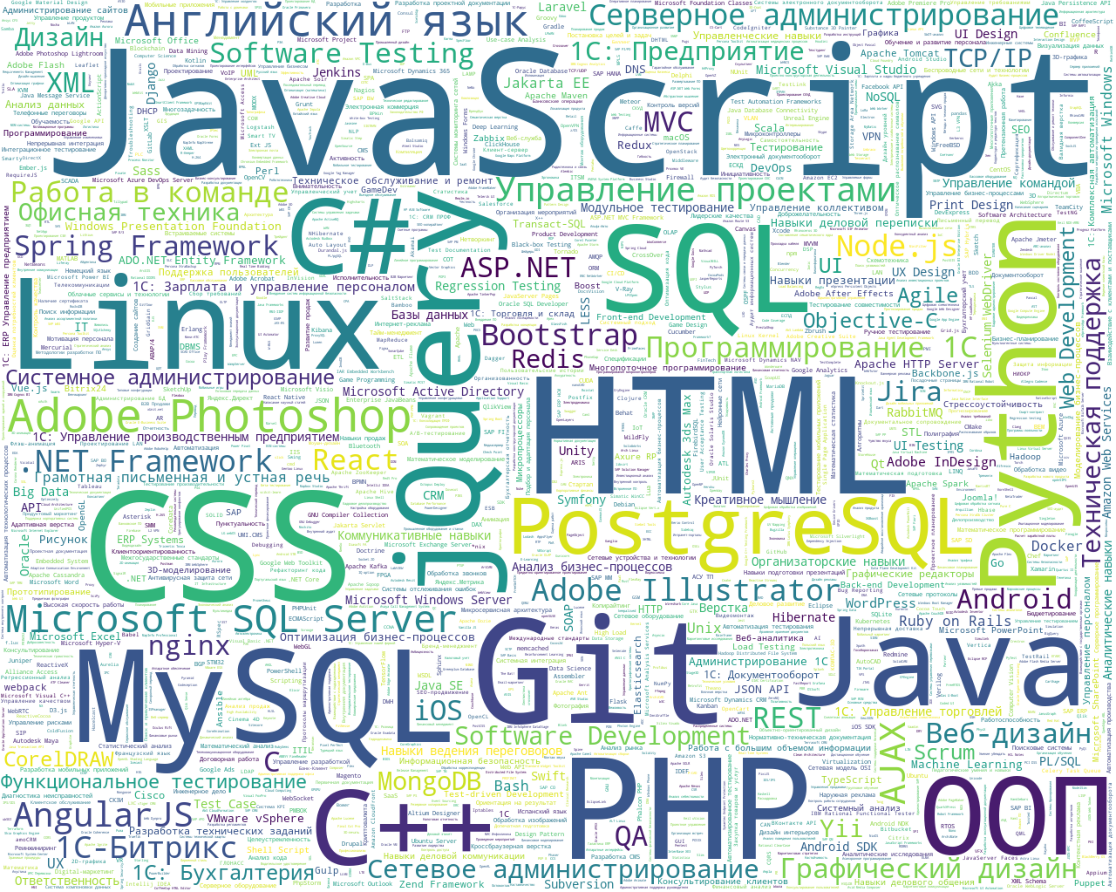


Рис. 2 – Ключевые навыки и компетенции, характерны для сферы информационных технологий, интернета и телекома

Источник: составлено автором

При этом в наборе данных 41,26% навыков упоминается 10 и более раз. В среднем на одну вакансию приходится около 6 навыков.

На Рис. 3 представлен пример описания вакансии и соответствующий ей список навыков. Как видно из Рис. 3, полнотекстовые описания вакансий и названия содержат отдельные символы, которые также необходимо исключить из описания.



Рис. 3 – Пример частичного описания вакансии в наборе данных (без обработки поля описания вакансии)

Источник: составлено автором

Процесс обработки текстового описания вакансии также осуществлялся c применением регулярных выражений. В целом процесс обработки включал следующие этапы: конвертация текста в нижний регистр; преобразование слов; удаление знаков препинания, специальных и повторяющихся символов. Также были удалены стоп-слова – это предлоги, суффиксы, междометия и др. Для предварительно обработанного поля с текстовым описанием средняя длина описания вакансии составила 1 556,58 символов. На Рис. 4 представлены наиболее часто встречающиеся слова.

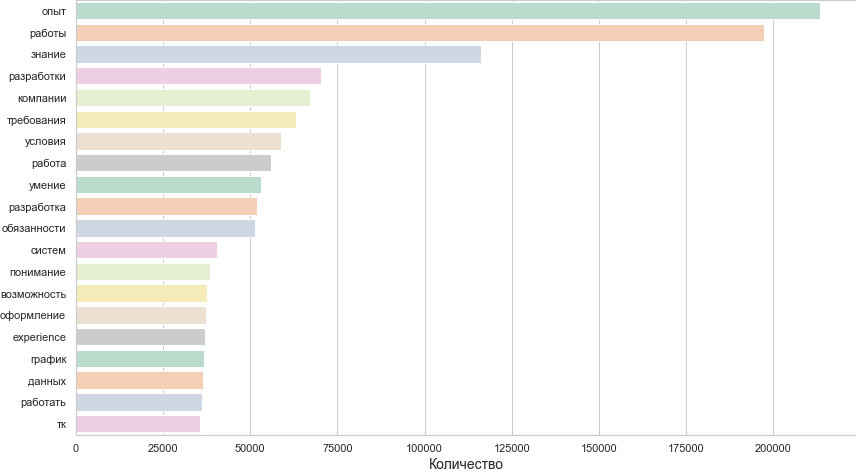


Рис. 4 – ТОП-20 наиболее часто встречающихся слов в наборе данных

Источник: составлено автором

Таким образом были получены предварительно обработанные данные, в которых устранены неточности описания и произведена нормализация текстовых полей. В Таблице 1 представлены основные характеристики предварительно обработанного набора данных.

Таблица 1. – Основные характеристика набора данных

|  |  |
| --- | --- |
| **Показатель** | **Значение** |
| Общее количество вакансий в наборе, ед. | 78 350 |
| Доля вакансий, в которых отсутствует информация о ключевых навыках, % | 66,45 |
| Общее количество навыков уникальных навыков, ед. | 2 409 |
| Количество навыков, которые упоминаются более 10 раз, ед. | 994 |
| Среднее количество навыков на одну вакансию, ед. | 6,05 |
| Средняя длина описания вакансии, символы | 1 556,58 |

Источник: Составлено автором

Важно отметить, что в наборе данных более чем у половины вакансий отсутствует информация о ключевых навыках. Поэтому очень важно попытаться восстановить пропущенные значения. Далее будет рассмотрен подход, в рамках которого будет предпринята попытка по восстановлению пропущенных данных.

1.2. Обработка пропусков в данных

Отсутствующие значения в данных являются проблемой для любого количественного исследования, так как они в итоге могут повлиять на результат. Одним из вариантов обработки пропусков выступает восстановление значений с использованием имеющейся информации из набора данных. Этот прием хорошо работает для количественных данных, однако обработка пропуска текстовых меток требует более внимательного рассмотрения проблемы.

В анализируемом наборе данных в более чем половине вакансий отсутствуют метки ключевых навыков. Исключение данных вакансий из рассмотрения ограничивает объем потенциально полезной информации, которую можно выделить. Поэтому необходимо рассмотреть возможность восстановления отсутствующих значений в данных. В настоящем параграфе будет разработана и апробирована модель прогнозирования набора требуемых меток (навыков) на основании текстового описания вакансий.

Постановка задачи: для некоторого текста  необходимо найти такое отображение , что , где  это общее количество меток (навыков) в наборе данных. Таким образом,  позволяет оценить вероятность для каждой метки  при заданном :



где ,  – это метка, соответствующая -му навыку.

Данную задачу следует рассматривать с точки зрения обработки естественного языка. Таким образом, в рамках параграфа рассматривается задача классификации текста с несколькими метками. Для описанной задачи мы пытаемся найти взаимосвязь между текстовым описанием вакансии и соответствующими текстовыми метками (навыками и компетенциями).

На сегодняшний день нейросетевые подходы позволяют решать широкий круг задач обработки естественного языка. Классификация текстов с несколькими метками представляет собой отдельную группу задач языкового моделирования. Отличительная черта такой классификации – количество меток может быть большим (сотни, тысячи).

Для решения поставленной будет использована языковая модель BERT (Bidirectional Encoder Representations from Transformers), в которой используется архитектура трансформера для решения широкого круга задач [19]. На основе архитектуры BERT была построена нейронная сеть, которая представлена в Приложении 1 (язык – Python).

Выбор метрик для оценки моделей базируется на предположении, что требуемые навыки являются неполными. К примеру, сотрудник, который заполнял поля вакансии, мог не внести важные характеристики. В условиях, когда отсутствует полная группа истинно положительных меток (True Positive), истинно отрицательные метки (True Negative) могут не быть такими. В связи с этим невозможно использовать классические метрики для задачи классификации с несколькими метками (например, Prevalence, F1 score и др.).

В контексте настоящей задачи для нас важно, чтобы модель могла предсказывать первые *N*-меток. Поэтому в качестве метрик для оценивания модели выбраны – *Recall@K* и *nDCG@K*, которые относятся к основным метрикам ранжирования. Данные метрики позволяют определить, насколько модель хорошо сортирует набор элементов, исходя из их релевантности. Рассмотрим данные метрики более подробно.

1. *Recall@K* – полнота на *K*-элементах, которая вычисляется как отношение найденных моделью релевантных меток для *K-*элементов к общему количеству релевантных меток:



Таким образом, данный показатель показывает долю релевантных меток, найденный для *K*-элементов. Показатель принимает значение от 0 до 1.

2. Normalized Discounted Cumulative Gain at K (*nDCG@K*) – метрика, оценивающая качество первых *K*-элементов [1, 23]:



где  – фактор нормализации, который равен максимально возможному (=1);  – средняя оценка релевантности, выставленная метке, расположенному на позиции *p* в списке результатов;  – дисконт за позицию метки (первые имеют больший вес).

*nDCG@K* принимает значения от 0 до 1, причем *nDCG@K*=1 только в случае, если система отранжировала метки в порядке убывания оценок.

Полученная модель использовалась для проведения серии экспериментов с различными конфигурациями. Результаты моделирования будут рассмотрены ниже.

1.3. Оценка результатов

Для классификации текста с несколькими метками мы используем предварительно обученные модели – ruBERT и mBERT. Первая из моделей обучалась на русскоязычной части Википедии и новостных данных [22], а вторая – это мультиязычная версия [19].

Используемый набор включает 26 291 вакансии, для которых указаны ключевые навыки. Данные были разбиты на тренировочный, тестовый и валидационный наборы – 80%, 10%, 10% соответственно.

Дополнительно при моделировании рассмотрен подход, при котором создается искусственный набор данных, в котором учитывается корреляция навыков на основе их совпадения:



где  – это оригинальная метка *k*-й выборки (двоичный вектор),  – набор всех навыков, упомянутых в *k*-й выборке, *M* – количество обучающих выборок.

Таким образом, было проведено четыре эксперимента со следующими основными параметрами:

* модели: ruBERT, mBERT, ruBERT\_C, mBERT\_C;
* количество эпох обучения: 7;
* batch size: 4;
* learning rate: 1.0e-4
* optimizer: Adam.

Основные гиперпараметры анализируемых моделей представлены в Приложении 2.

Для метрик *Recall@K* и *nDCG@K* оценка осуществлялась для *K* равного 5, 10, 30, 50, 100. Результаты экспериментов представлены в Таблице 2.

Таблица 2. – Результаты тестирования моделей

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Метрика** | **Модель** | | | |
| **ruBERT** | **ruBERT\_C\*** | **mBERT** | **mBERT\_C\*** |
| *Recall@5* | 0,2061 | 0,2159 | 0,1228 | 0,1167 |
| *Recall@10* | 0,3135 | 0,3319 | 0,2068 | 0,1935 |
| *Recall@30* | 0,5210 | 0,5407 | 0,4038 | 0,3894 |
| *Recall@50* | 0,6126 | 0,6346 | 0,5155 | 0,4961 |
| *Recall@100* | 0,7278 | 0,7548 | 0,6663 | 0,6474 |
| *nDCG@5* | 0,2303 | 0,2430 | 0,1260 | 0,1167 |
| *nDCG@10* | 0,2906 | 0,3079 | 0,1720 | 0,1186 |
| *nDCG@30* | 0,3767 | 0,3944 | 0,2534 | 0,1609 |
| *nDCG@50* | 0,4068 | 0,4254 | 0,2902 | 0,2413 |
| *nDCG@100* | 0,4397 | 0,4597 | 0,3329 | 0,2764 |

\* учитывается корреляция навыков

Источник: составлено автором

На основании полученных результатов можно сделать ряд выводов. Во- первых, русскоязычная модель BERT предпочтительнее мультиязычной, так как обеспечивает лучшие результаты по используемым метрикам. Во- вторых, использование дополнительной информации о корреляции навыков позволяет улучшить результаты. По показателю полноты (*Recall@K)* модель демонстрирует хорошие результаты только для высоких значений *K.* Для более низкого порядка, к примеру, для модели ruBERT\_C точность на уровне 10 элементов составляет 33,19%. По качеству ранжирования модель также демонстрирует низкие результаты.

Работа модели представлена на Рис. 5. Так явные истинные значения – это метки, которые присутствуют как в описании вакансии, так и в списке ключевых навыков. Неявные истинные значения – это те метки, которые присутствуют в списке ключевых навыков, но отсутствуют в описании вакансии.

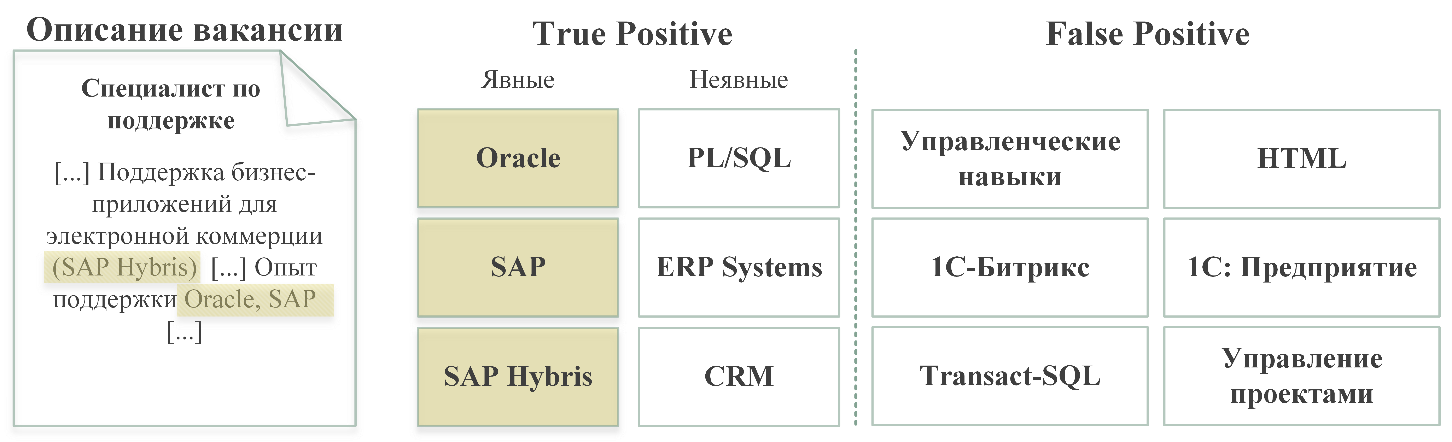


Рис. 5 – Пример результатов вывода модели для вакансии «Специалист по поддержке

Источник: составлено автором

Подводя итоги настоящего параграфа, нужно сказать, что рассмотренный подход для прогнозирования набора требуемых меток (навыков) на основании текстового описания вакансий имеет потенциал развития. Во-первых, необходимо обеспечить лучшую консистентность исходных данных. Когда требования к работе очевидны, то их описание не включается в вакансии. Искажения в данных также влияют на производительность модели. Таким образом, предпочтительно использование более сбалансированных выборок, в которых высокое среднее количество меток на одну вакансию. Современные нейросетевые подходы позволяют эффективно осуществлять тонкую настройку гиперпараметров моделей.

Глава 2. Моделирование спроса на навыки и компетенции

2.1. Количественная оценка спроса

Для проведения анализа информации о навыках и компетенциях необходимо выбрать способ оценки показателя спроса. В литературе широко распространен подход, при котором оценка спроса на навыки осуществляется на основании частоты их появления в объявлениях, вакансиях, относящихся к выбранной профессии или группе профессий.

В рамках настоящей работы частота появления навыка в вакансиях служит количественным показателем спроса – более высокие показатели свидетельствую о высоком спросе. Очевидно, что показатель частоты не является показателем спроса в традиционном представлении и выступает в качестве аппроксимирующего показателя.

Пусть *–* это показатель спроса на навык *s* для вакансии *j*. Показатель принимает значение , когда навык *s* требуется для вакансии *j*, в противном случае . Тогда количественная оценка важности (веса)  навыка равна:



Таким образом, на основании частоты появления навыка в вакансиях можно сформировать ранжированный список наиболее востребованных навыков и компетенций для группы профессий или экономики в целом.

При использовании частоты появления навыков в качестве показателя спроса возникает проблема нормализации оценок. Высокая частота необязательно свидетельствует о важности навыка, так как существуют навыки, которые могут встречаться во всех вакансиях с высокой частотой. Поэтому в качестве дополнительной оценки важности (веса) используется индекс выявленного сравнительного преимущества.

Индекс выявленного сравнительного преимущества (Revealed comparative advantage, RCA) или индекс Баласса [12] получил широкое распространение в работах, посвященных анализу конкурентного преимущества в международной торговле, сравнительному анализу экспорта и импорта экономик различных стран.

Показатель RCA также нашел применение в работах, посвященных оценке навыков и компетенций на рынке труда [11, 17]. В общем случае показатель позволяет сравнить относительную важность навыка для отдельной вакансии с ожидаемой относительной важностью навыка для всей совокупности вакансий.

RCA может быть рассчитан как соотношение между удельным весом спроса на навык в общем спросе на навыки для выбранной вакансии и удельным весом спроса на навык в общем объеме спроса среди всех вакансий:



Ситуация, когда , свидетельствует о том, что навык *s* более важен для вакансии *j*, чем ожидается в совокупности. Используя выражения (1), можем рассчитать веса навыков для отдельных профессиональных групп:



где – это вес (важность) навыка *s* в вакансиях, относящихся к профессии *o (*). Веса навыков для экономики в целом могут быть рассчитаны следующим образом:



Расчет показателя RCA позволяет оценить относительную важность навыка для отдельных групп вакансий в сравнении со всеми рассматриваемыми вакансиями. В зависимости от значения показателя RCA, навыки и компетенции можно распределить по трем группам: с низким RCA

(), со средним RCA () и с высоким RCA ().

Таким образом, для характеристики спроса на навыки и компетенции используется два показателя – востребованность и важность навыка. Данный подход позволяет учесть специфику анализа навыков на рынке труда и предоставить универсальный инструмент оценки спроса.

2.2. Моделирование спроса на навыки и компетенции

Моделирование спроса на навыки и компетенции может осуществляться по двум направлениям. Во-первых, это количественное моделирование структуры спроса на навыки и компетенции. Во-вторых, это прогнозирование уровня заработной платы на основе описания вакансий.

Моделирование структуры спроса предусматривает оценку показателей с использованием методики оценки, описанной в предыдущем параграфе. Структурные оценки формируются на основании частоты появления навыка в вакансиях. В условиях, когда мы можем наблюдать актуальную информацию о навыках на уровне каждой вакансии, задача прогнозирования структуры спроса не является актуальной. Структура спроса на навыки и компетенции будет рассмотрена в Главе 3.

Более интересной исследовательской задачей выступает определение заработной платы на основании полей, которые представлены в вакансиях. Классическая задача прогнозирования заработной платы рассматривается как задача множественной линейной регрессии. Однако данный подход не применим для прогнозирования заработной платы для нашего набора данных. Это связано с тем, что вакансии в сфере информационных технологий характеризуются длинным верхним хвостом заработных плат.

Чтобы преодолеть вышеописанную проблему, необходимо рассмотреть задачу классификации. Распределим все вакансии на 7 групп в зависимости от интервала заработной платы: «<70,000», «70,000-140,000», «140,000-210,000», «210,000-280,000», «280,000-350,000», «350,000-420,000», «>420,000».

Постановка задачи: пусть *X –* множество описаний объекта*, Y –* конечное множество меток классов. Тогда существует отображение *: ,* значения которого известны только на объектах конечной обучающей выборки. Требуется построить такой алгоритм *a*: *,* способный классифицировать произвольный объект *.*

Таким образом необходимо разработать модель, которая будет классифицировать вакансии по одному из семи выделенных классов. Для построения модели будут рассмотрены несколько классических подходов, позволяющие решать широкий круг задач текстовой классификации.

В рамках первого подхода моделирование будет осуществляться с использованием простого представления документов в векторном виде – мешка слов (Bag of Words, BoW). BoW позволяет представить документ (текст) в виде неупорядоченного набора слов (словаря) [28]:



где  – *j*-е слово в *i*-м документе, *m* – общее количество слов во всех документах.

На основе набора документов, которые входят в корпус, строится словарь из всех встречающихся признаков. Таким образом, документ можно представить набором признаков.

Для того, чтобы учесть важность слов в документе среди всего набора документов, используется показатель *TF-IDF* (TF – term frequency, IDF – inverse document frequency). Этот показатель основывается на предположении, что важность или релевантность слова пропорциональна частоте употребления этого слова в документе и обратно пропорционален частоте употребления слова во всех документах.



где  – частота слова  в документе ,  – набор документов,  – все документы в наборе, в которых встречается слово .

В целом использование *TF-IDF* позволяет представить документ (описание вакансии) в виде числовых векторов, которые используются для классификации. Для сравнительной характеристики производительности моделей используются различные алгоритмы классификации: Random Forests, Support Vector Machine, Stochastic Gradient Descen, XGBoost и Decision Tree.

В рамках другого направления для построения моделей рассматривались методы «Word embeddings», основанные на векторном представлении слов. В отличие от предыдущего подхода, векторное представление слов позволяет учитывать контекст документа – слова с одинаковым значением имеют сходное представление. Для моделирования используется библиотека FastText, разработанная Facebook [13]. Данная библиотека содержит предобученные векторные представления слов. Для задачи классификации были сформированы модели с различными спецификациями – FastText, CNN+ FastText, LSTM+FastText, BiLSTM+FastText.

Первая модель FastText предусматривает использование стандартной библиотеки без модификаций. В модели CNN+ FastText предусмотрено использование простой сверточной нейронной сети на вход которой будут подаваться предобученные векторные представления слов. Модель LSTM+FastText предусматривает использование архитектуры рекуррентной нейронной сети LSTM. Модель BiLSTM+FastText предусматривает использование двунаправленной рекуррентной сеты, содержащей две однонаправленные рекуррентные сети LSTM.

Описанные подходы являются классическими для задач классификации текстовых данных. Эти модели простые и легко интерпретируемы, что упрощает сравнительную характеристику результатов.

Для оценки результатов моделирования используется стандартный набор метрик классификации –*Accuracy, Precision, Recall и F1*.



где *TP* (True Positive) – ситуация, когда классификатор правильно отнес объект к рассматриваемому классу; *TN* (True Negative) – ситуация, когда классификатор верно определил, что объект не принадлежит к рассматриваемому классу; *FP* (False Positive) – ситуация, когда классификатор неверно отнес объект к рассматриваемому классу; *FN* (False Negative) – классификатор неверно определяет, что объект не принадлежит к рассматриваемому классу.

Таким образом, полученные модели использованы для классификации вакансий по уровню заработной платы. Основные результаты моделирования рассмотрены ниже.

2.3. Оценка результатов

Для классификации вакансий по уровню заработной платы был сформирован набор данных, содержащий 15 568. Данные были разбиты на тренировочный и тестовый наборы – 70% и 30% соответственно. Такая пропорция сохранялась для всех рассматриваемых моделей.

Классификация осуществляется по семи классам, которые были описаны в предыдущем параграфе. На Рис. 6 представлено распределение заработной платы по классам.

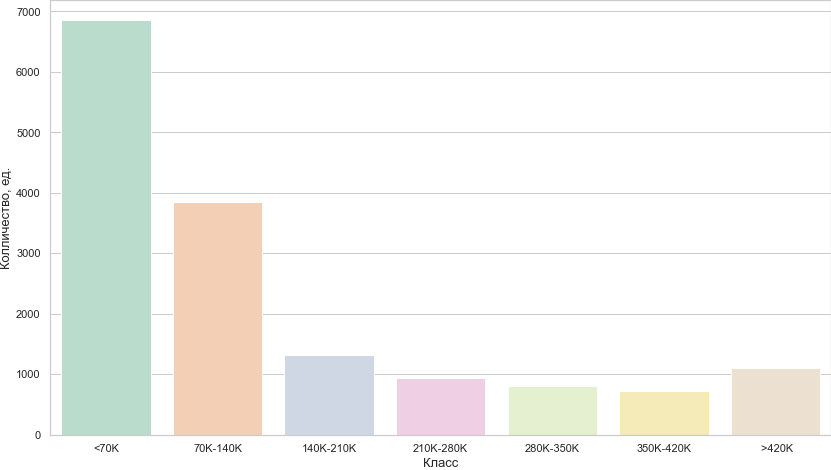


Рис. 6 – Распределение заработной платы по классам

Источник: составлено автором

Модели с использованием *TF-IDF* рассматриваются как базовые. Для моделирования использованы алгоритмы классификации Random Forests, Support Vector Machine, Stochastic Gradient Descen, XGBoost и Decision Tree. Основные средневзвешенные результаты моделирования представлены в Таблице 3., а в Приложении 3 представлены расширенные результаты моделирования.

Таблица 3. – Результаты тестирования моделей

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Модель** | **Метрика** | | | |
| Accuracy | Precision | Recall | F1 |
| Random Forests | 0.70 | 0.70 | 0.70 | 0.68 |
| Support Vector Machine | 0.71 | 0.71 | 0.71 | 0.70 |
| Stochastic Gradient Descen | 0.60 | 0.57 | 0.60 | 0.57 |
| XGBoost | 0.58 | 0.62 | 0.58 | 0.53 |
| Decision Tree | 0.62 | 0.61 | 0.62 | 0.62 |

Источник: составлено автором

Интерпретируем основные результаты проведенного моделирования. Наиболее эффективной оказалась модель, основанная на методе опорных векторов, точность классификации составила 71%. Модель достаточно хорошо показывает долю истинных объектов класса среди всех объектов, отнесенных классификатором к истинным (показатель Precision). Также классификатор хорошо справляется с задачей предсказания объектов класса среди всех объектов данного класса (показатель Recall). Схожие результаты были получены для модели, основанной на методе случайного леса. Другие рассматриваемые модели, к сожалению, показали низкие результаты.

В рамках второго наплавления тестировались модели на основе предобученных векторных представлений слов. Рассматриваются следующие модели: FastText, CNN+ FastText, LSTM+FastText, BiLSTM+FastText. Основные средневзвешенные результаты моделирования представлены в Таблице 4. В Приложении 5 представлены расширенные результаты моделирования.

Таблица 4. – Результаты тестирования моделей

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Модель** | **Метрика** | | | |
| Accuracy | Precision | Recall | F1 |
| FastText | 0.67 | 0.66 | 0.67 | 0.65 |
| CNN+ FastText | 0.69 | 0.68 | 0.69 | 0.67 |
| LSTM+FastTex | 0.42 | 0.44 | 0.42 | 0.31 |
| BiLSTM+FastText | 0.64 | 0.64 | 0.64 | 0.64 |

Источник: составлено автором

Использование предварительно обученных векторных представлений слов в целом не позволило улучшить результаты по сравнению с моделями, рассмотренными выше. Наибольшую точность показала модель простой сверточной нейронной сети с предобученными векторными представлениями (CNN+ FastText) – 69%. Таким образом, даже базовые модели классификации на основе текстовых данных показывают хорошие результаты. В то же время обеспечение сбалансированности данных в рассматриваемых классах позволит в будущем повысить точность моделирования.

Глава 3. Анализ ситуации на рынке труда Санкт- Петербурга

3.1. Анализ спроса на навыки и компетенции

В настоящем параграфе будет произведен анализ спроса на навыки и компетенции на рынке труда Санкт-Петербурга. Для проведения анализа были сформированы четыре укрупненных группы вакансий, относящихся к сфере информационных технологий:

* Python-разработка;
* UI/UX-дизайн;
* Анализ данных;
* Android/iOS-разработка.

Оценка спроса осуществлялась в соответствии с методологией, описанной в Главе 2. Для каждой группы, отобранной для анализа, формировался ранжированный список наиболее востребованных навыков и компетенций. Ранжирование осуществлялось на основании частоты появления навыка в вакансиях. Дополнительно рассчитывался показатель RCA, позволяющий сравнить относительную важность навыка для группы вакансий с ожидаемой относительной важностью навыка для всей совокупности вакансий (по всему набору данных).

Характерной особенностью сферы информационных технологий, интернета и телекома является сильная дифференциация заработной платы. Большинство профессиональных групп характеризуется длинным верхним хвостом заработных плат. Для отобранных групп вакансий данная тенденция сохраняется. На Рис. 7 представлена сравнительная характеристика распределения заработной платы по каждой группе. Так для группы «Python- разработка» в 2018 г. медианная заработная плата составила 145 000 руб. Для групп «Android/iOS-разработка», «Анализ данных», «UI/UX-дизайн» медианная заработная плата составила 180 000 руб., 185 000 руб. и 142 500 руб. соответственно. Более широкие участки скрипичного графика представляют более высокую вероятность того, что заработная плата примет данное значение.

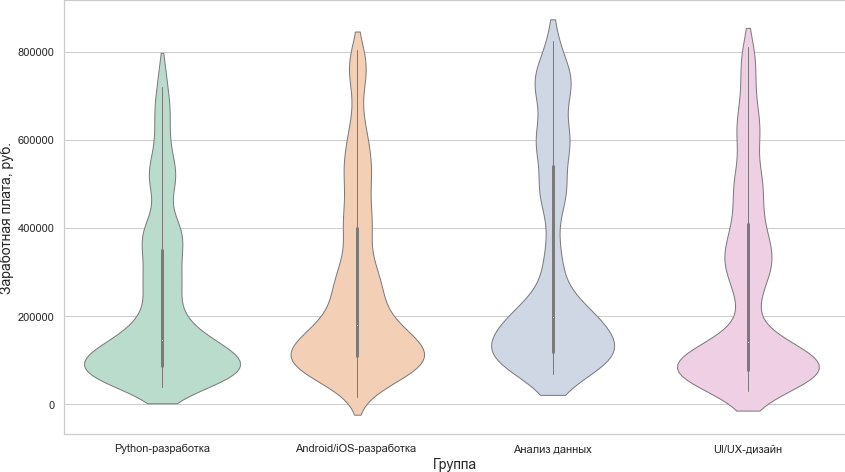


Рис. 7 – Сравнительная характеристика распределения заработной платы по анализируемым группам (в 2018 г.)

Источник: составлено автором

В целом наблюдаемые распределения заработных плат соответствуют рыночной ситуации – разрабатывать приложения и сервисы стоит дороже.

Рассмотрим каждую группу вакансий с точки зрения спроса на навыки и компетенции. В рамках группы «Python-разработка» рассматриваются вакансии специалистов, которые разрабатывают и программируют на языке Python. В то же время для данных специалистов язык программирования Python может выступать в качестве одного из основных. Здесь и далее анализ спроса осуществляется на данных 2017–2018 гг. На Рис. 8 представлен ранжированный список из двадцати наиболее востребованных навыков и компетенций для данной группы вакансий. Линии на рисунке отображают изменения в отчетном периоде по сравнению с базовым. Цвет линий показывает значения показателя RCA. Список из пятидесяти наиболее востребованных навыков представлен в Приложении 5.

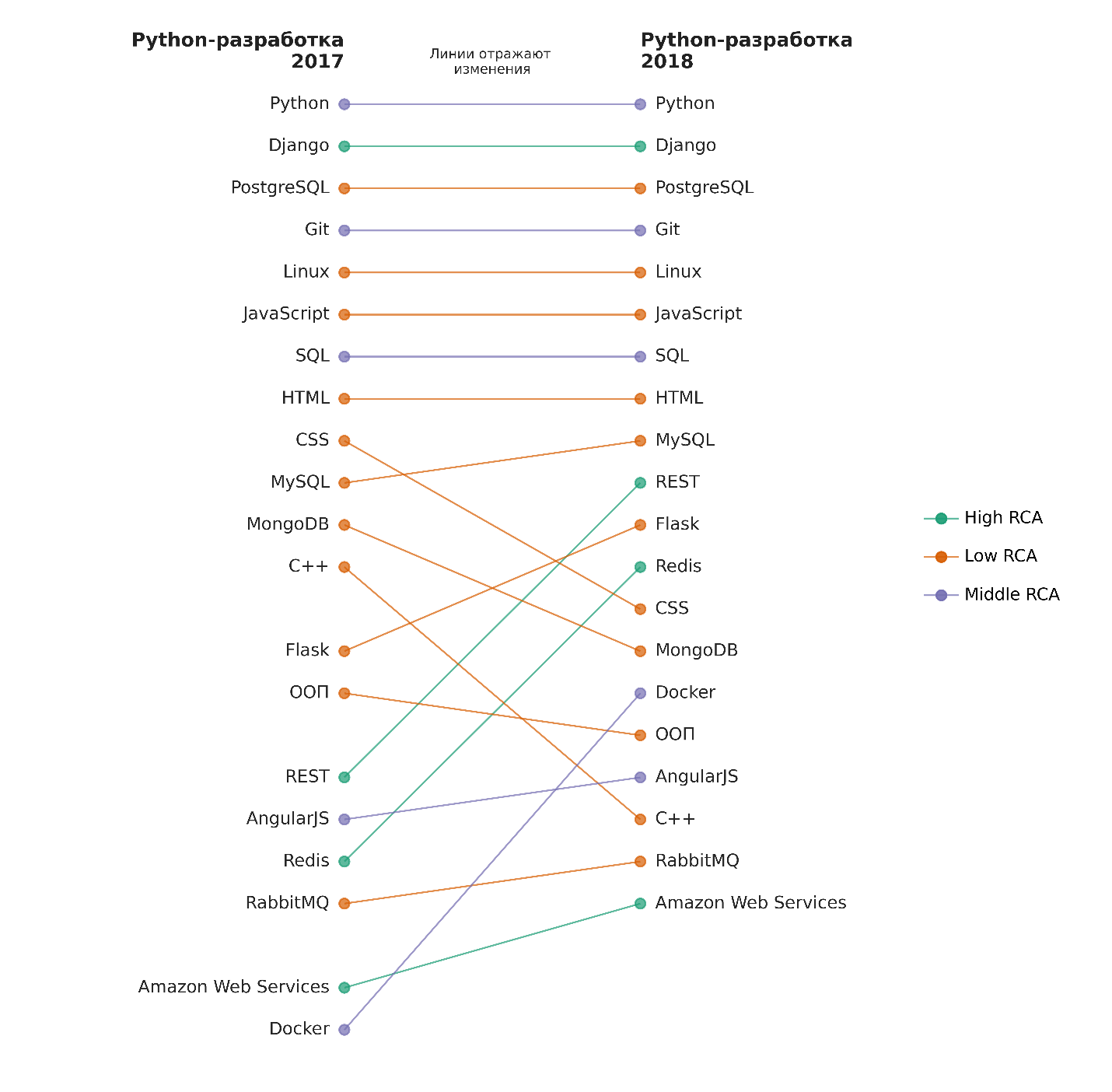


Рис. 8 – Спрос на навыки и компетенции по группе «Python-разработка» в 2017- 2018 гг.

Источник: составлено автором

В списке из ТОП-20 наиболее востребованных навыков и компетенций по группе «Python-разработка» представлены только «жесткие» навыки (hard skills). Это технические навыки и компетенции, которые включают знание различных языков программирования и фреймворков, необходимых для эффективного выполнения профессиональных обязанностей. В целом структура спроса в отчетном периоде претерпела незначительные изменения по сравнению с базовым – основные изменения наблюдаются за границей ТОП-10 востребованных навыков. Анализируя значение показателя RCA, можно выделить ряд навыков, которые более важны для данной группы вакансий, чем для всей совокупности. Это знание языков Python и SQL, фреймворка Django, СУБД PostgreSQL, а также умение работать с Git. Именно эти навыки являются наиболее востребованными и важными для анализируемой группы.

Группа «UI/UX-дизайн» объединяет в себе вакансии специалистов, которые осуществляют проектирование, разработку и дизайн пользовательских интерфейсов. Ранжированный список из двадцати наиболее востребованных навыков и компетенций для данной группы вакансий представлен на Рис. 9. На рынке труда Санкт-Петербурга от специалистов в области UI/UX-дизайна требуется как развитые прикладные навыки, так и широкие знания различных направлений в дизайне. В первую очередь это умение проектировать различные пользовательские интерфейсы, знание графического и веб-дизайна, а также владение графическими редакторами (Adobe Photoshop, Adobe Illustrator). В Приложении 5 представлен список из пятидесяти наиболее востребованных навыков и компетенций.

Характерной особенностью данной группы является высокое значение показателя RCA в рамках ТОП-10 востребованных навыков. Эти навыки являются, с одной стороны, востребованными на рынке, а с другой особенно важны для данной группы вакансий.

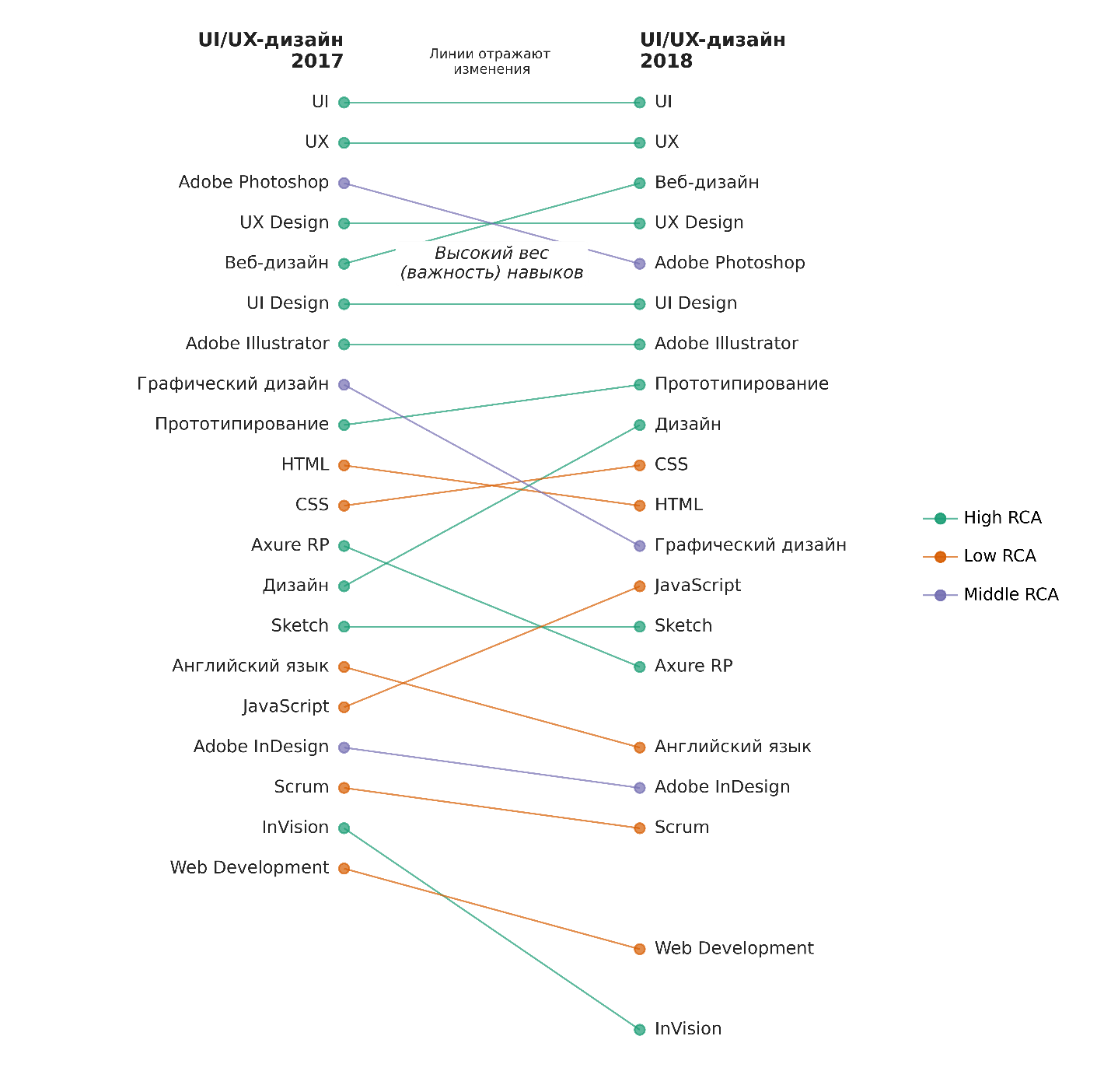


Рис. 9 – Спрос на навыки и компетенции по группе «UI/UX-дизайн» в 2017- 2018 гг.

Источник: составлено автором

В группу «Анализ данных» был включен широкий круг специалистов, включающий: аналитиков данных, специалистов в области «Data Science», BI- разработчиков, BI-аналитиков, специалистов по машинному обучению, специалистов по Big Data и др. На Рис. 10 представлен ранжированный список из двадцати наиболее востребованных навыков и компетенций для данной группы вакансий. Список из пятидесяти наиболее востребованных навыков представлен в Приложении 5.

Для данной группы в ТОП-10 востребованных навыков наблюдается устойчивая структура спроса. Подавляющее большинство навыков, представленных в ТОП-20, имеют высокое значение показателя RCA, что подтверждает важность этих навыков для рассматриваемых вакансий. Успешный специалист в области данных должен знать несколько языков программирования (Python, SQL, C++), иметь хорошую математическую подготовку (математическое моделирование, машинное обучение), а также владеть современными средствами работы с большими данными (Hadoop, Apache Spark). На сегодняшний день данная сфера активно развивается, поэтому навыки в 2017–2018 гг. могут не содержать новых распространенных технологий.

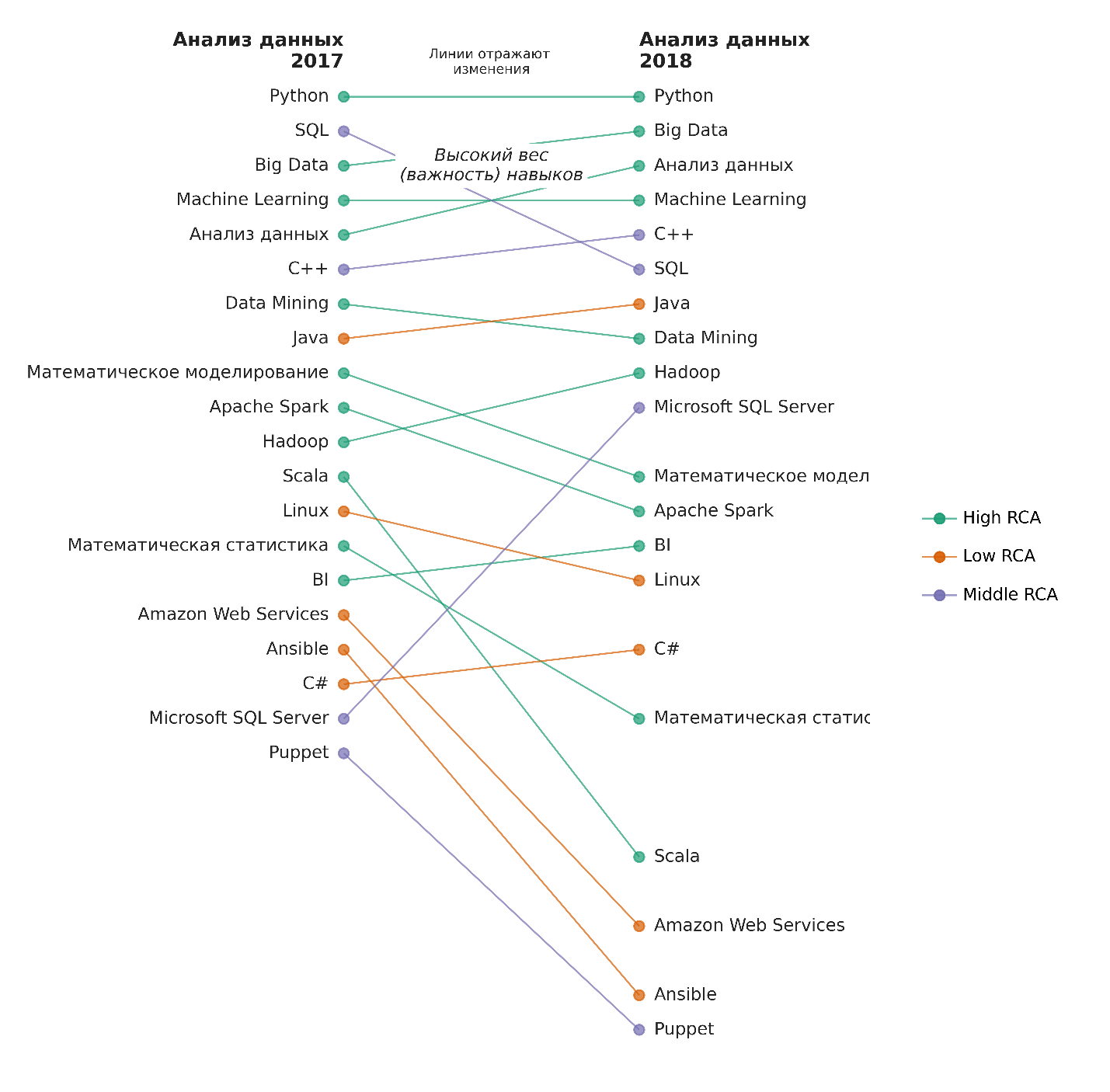


Рис. 10 – Спрос на навыки и компетенции по группе «Анализ данных» в 2017- 2018 гг.

Источник: составлено автором

В рамках последней рассматриваемой группы «Android/iOS-разработка» представлены разработчики приложений и сервисов для операционных систем на базе Android/iOS. Ранжированный список из двадцати наиболее востребованных навыков и компетенций представлен на Рис. 11. Пятьдесят наиболее востребованных навыков представлен в Приложении 5. Структура спроса на навыки и компетенции для данной группы устойчива в 2017–2018 гг. наблюдаются незначительные изменения. В целом наиболее востребованные навыки имеют высокое значение показателя RCA. Успех на рынке труда для представителей данной группы профессий определяется опытом работы с Android/iOS, знанием нескольких языков программирования (Java, Objective- C, Kotlin, Swift) и фреймворков для разработки.

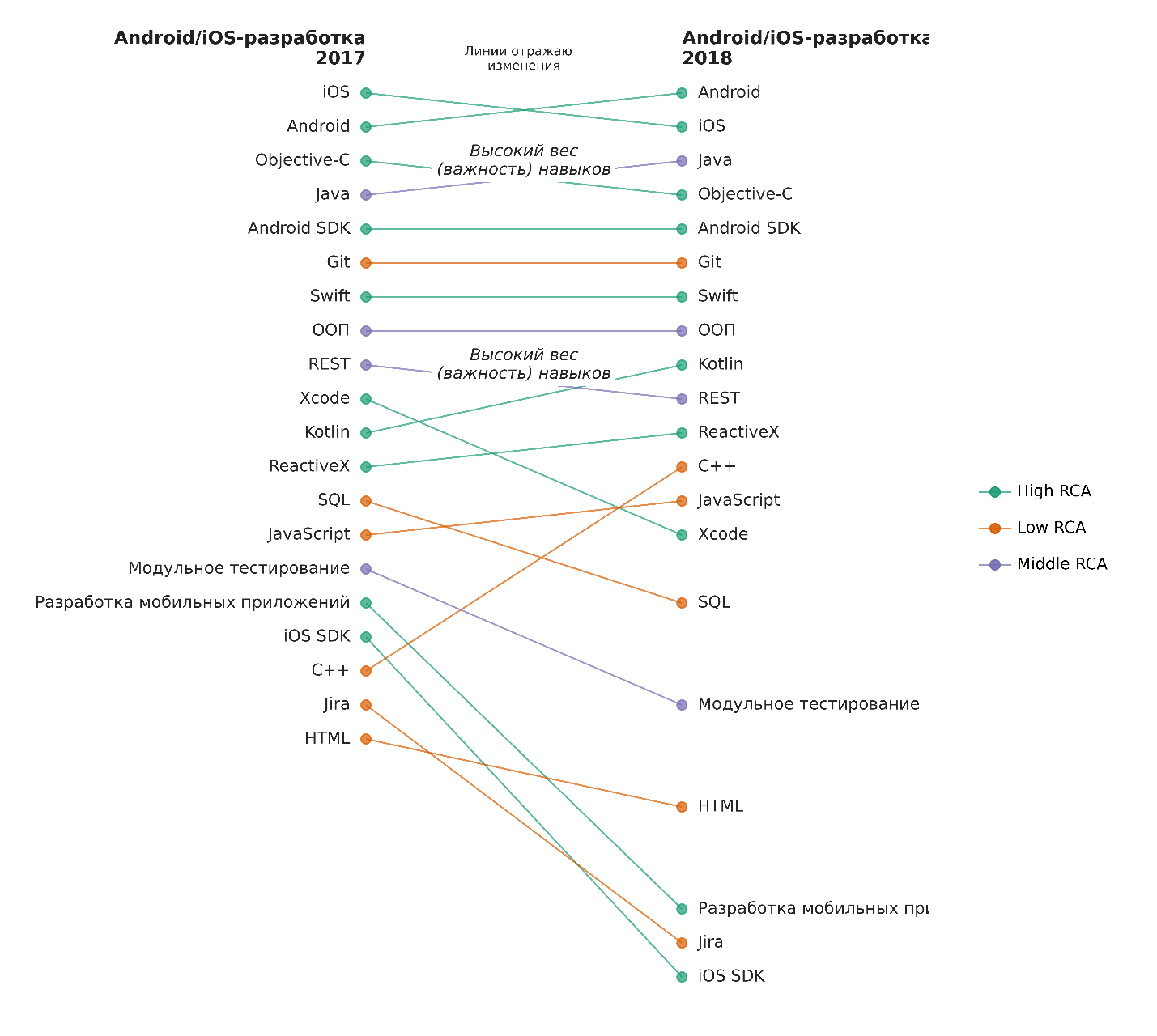


Рис. 11 – Спрос на навыки и компетенции по группе «Android/iOS-разработка» в 2017- 2018 гг.

Источник: составлено автором

Поводя итоги проведенного анализа спроса на навыки и компетенции на рынке труда Санкт-Петербурга можно сделать несколько основных выводов. Во-первых, использование предварительно обработанных данных позволяет эффективно представить и визуализировать ключевые навыки и компетенции, характерные для любой профессиональной группы.   
Во-вторых, предложенный подход по оценке спроса позволяет рассмотреть показатель с разных точек зрения – востребованности и важности навыков. В условиях, когда отсутствуют устоявшиеся подходы к анализу онлайн- объявлений о вакансиях, апробация различных подходов к оценке спроса выступает актуальной задачей. В-третьих, результаты анализа имеют практическую значимость для понимания происходящих изменений на рынке труда. Анализируя показатели спроса, можно сформировать различные траектории движения навыков, а также использовать результаты при планировании политики в области занятости.

3.2. Рекомендации по совершенствованию государственной политики в области обеспечения занятости

На сегодняшний день изучение тенденций и особенностей функционирования отраслевых рынков труда является актуальной исследовательской проблемой. В частности, открытыми остаются вопросы анализа и моделирования рыночных тенденций. Проблемы изучения российского рынка труда связаны с доступностью качественных и полных данных о рынке. Отсутствие систематических и надежных исходных данных затрудняет использование отдельных показателей для анализа и прогнозирования рыночных тенденций [5]. Поэтому важно разрабатывать подходы, позволяющие работать с альтернативными источниками данных о рынке труда [3].

Обобщая результаты, полученные в рамках настоящей исследовательской работы, можно сформировать аналитический инструментарий по анализу данных из онлайн- объявлений о вакансиях, который включает:

* подходы к предварительной обработке исходных данных;
* способы моделирования рыночного спроса на основе данных вакансий;
* способы количественной оценки спроса на навыки и компетенции.

Полученные результаты анализа и моделирования имеют потенциальную ценность для широкого круга пользователей, прежде всего для государственных органов, осуществляющих политику в области занятости. Спрос и предложение на навыки и компетенции может быть представлен в виде простой схемы (Рис. 12). Результатом взаимодействия спроса и предложения является информация о несоответствии текущих навыков необходимым, то есть о потребности в развитии новых навыков. При этом природа несоответствия спроса на навыки и компетенции различна. К основным типам несоответствия можно отнести: нехватку навыков, избыток навыков, разрыв в навыках, и чрезмерную/недостаточную квалификацию.



Рис. 12 – Взаимодействие спроса и предложения на навыки и компетенции

Источник: составлено автором

Таким образом, государственные органы заинтересованы в получении актуальной информации и рыночных сигналов при планировании политики в области обеспечения занятости населения.

В соответствии с Законом Российской Федерации от 19 апреля 1991 г. № 1032-1 «О занятости населения в Российской Федерации» государство реализует политику в области содействия занятости населения, направленную на развитие трудовых ресурсов, повышение их мобильности, защиту национального рынка труда [6]. Это предполагает широкий круг полномочий, часть которых передается органам государственной власти субъектов Российской Федерации.

Особый интерес представляют полномочия органов государственной власти субъектов Российской Федерации в области содействия занятости населения, направленные на организацию профессионального обучения и дополнительного профессионального образования безработных граждан. Реализация указанных полномочий в Санкт-Петербурге входит в обязанности Комитета по труду и занятости населения Санкт-Петербурга, который в свою очередь возлагает их на государственные службы занятости населения.

На сегодняшний день службы занятости населения осуществляют профессиональное обучение или дополнительное профессиональное образование по трем направлениям: подготовка, переподготовка и повышение квалификации. При этом выбор направлений реализации программ образования во многом имеет несистемный характер («по заявкам»).

Поэтому для формирования эффективной государственной политики в области обеспечения занятости и удовлетворения потребностей региональной экономики в квалифицированных кадрах, необходимо рассмотреть возможность использования актуальной информации о спросе на навыки и компетенции на рынке труда. Для этого сформирован ряд рекомендаций службам занятости населения при выборе направлений подготовки, переподготовки и повышению квалификации граждан.

Во-первых, при планировании и разработке программ профессионального обучения необходимо учитывать актуальные потребности рынка в квалифицированных кадрах. Это означает не только формирование программ обучения по наиболее востребованных специальностям на рынке, но и развитие аналитических инструментов прогнозирования потребности.

Во-вторых, в современных условиях специальные, узкие «жесткие» навыки (hard skills) приобретают все большее значение. Описание востребованного специалиста можно представить в виде списка навыков и компетенций, необходимых для работы. Поэтому результаты настоящей исследовательской работы носят практический характер, а представленные подходы к анализу данных онлайн- объявлений о вакансиях могут быть использованы службами занятости при планировании программ профессионального обучение или дополнительного профессионального образования.

В-третьих, описанные в рамках работы походы носят универсальный характер, то есть могут быть использованы для анализа и моделирования различных наборов данных, содержащих информацию из вакансий.

Поводя итоги, можно сказать, что государственные органы власти обладают широкими полномочиями в области содействия занятости населения. Для эффективной реализации этих полномочий требуется актуальные аналитические данные, отражающие текущую ситуацию и тенденции рынка труда. Анализ и моделирование спроса на навыки и компетенции выступает одним из инструментов, способствующих пониманию происходящих изменений на рынке труда.

Выводы

Выводы настоящего исследования представлены по разделу с анализом литературы и трем главам.

В разделе, посвященном анализу литературных источников, были описаны и проанализированы основные эмпирические и методологические источники. Были рассмотрены вопросы определения понятия рынка труда, трактовки понятий навыков и компетенций, классификации навыков и моделирования спроса на навыки. Различия в определениях понятия рынка труда обусловлены как историческими этапами развития экономической мысли, так и сложностью изучения существующих механизмов рыночного взаимодействия. Рынок труда рассматривается как сложная система социально-экономических отношений между субъектами рынка, обеспечивающая согласование и координацию рыночного механизма.

На сегодняшний день место навыков и компетенций в системе рыночных отношений до конца не определено. В общем случае интеллектуальные способности работника необходимы для развития знаний, практическая реализация знаний является частью развития навыков, а все в совокупности является предпосылками к развитию компетенций. Простая классификация предполагает разделение навыков и компетенций на универсальные и специфические. Современные же подходы к классификации предполагают выделение навыков и компетенций из описаний профессиональных дескрипторов или объявлений о вакансиях, а также использование иерархических классификаций. Моделирование спроса на навыки и компетенции позволяют получить полезную информацию, которая используется при прогнозировании потребностей в навыках и планировании политики в области обеспечения занятости населения.

В Главе 1 были рассмотрены подходы к предварительной обработке данных. Использование онлайн- объявлений о вакансиях в качестве основного источника данных является перспективным исследовательским направлением. Прежде всего потому, что необходимые навыки и требования к работе указаны в вакансиях. Навыки непосредственно наблюдаются на уровне каждой вакансии и могут быть оценены по профессиям, отраслям, регионам или экономике в целом. Современные подходы и алгоритмы обработки естественного языка значительно упрощают их извлечение и анализ. Результаты проведенной в рамках работы предварительной обработки исходных данных позволили сформировать подход, позволяющий устранить несбалансированность в данных и неточности в описании вакансий. Нейронная сеть на основе архитектуры BERT использовалась для прогнозирования набора метрик (пропущенных навыков в текстовом описании вакансии). Полученные результаты показали потенциал использования данного подхода. Стоит отметить, что консистентность исходных данных является одним из наиболее важных параметров.

В Главе 2 были рассмотрены вопросы количественной оценки и моделирования спроса на навыки и компетенции. В работе рассмотрены два показателя количественной оценки спроса – востребованность и важность навыка. Эти показатели позволяют отразить особенности анализа и оценки навыков на рынке труда. Моделирование спроса на навыки и компетенции на рынке труда предусматривает оценку структуры спроса (Глава 3), а также прогнозирование уровня заработной платы на основе описания вакансий. В рамках прогнозирования уровня заработной платы решалась проблема классификации вакансий по интервалам заработной платы. Для решения этой задачи использовались две группы моделей – простые модели на основе представления документов в векторном виде (TF-IDF), а также модели на основе векторного представления слов (Word embedding). Таким образом, осуществлялась сравнительная характеристика двух групп на основании одинаковых метрик. В рамках первой группы моделей были апробированы различные алгоритмы классификации, среди которых наилучшие результаты показали два алгоритма – Random Forests, Support Vector Machine. Вторая группа моделей предусматривала использование различных комбинаций нейронных сетей, которым на вход поступали предварительно обученные векторные представления слов. В конечном счете отдельные модели позволили классифицировать вакансии по интервалам заработной платы с точностью свыше 70%.

В Главе 3 был осуществлен анализ ситуации на региональном рынке труда Санкт-Петербурга. Анализ позволил увидеть динамику изменений спроса на конкретные навыки со стороны работодателей. Полученные результаты могут являться основой для работы в области прогнозирования востребованности отдельных навыков, а в последствии и профессий на рынке труда. Одним из наиболее значимых получателей такой информации являются компетентные органы власти субъектов, обладающие полномочиями в области содействия занятости населения.

Заключение

В своевременных условиях множество факторов оказывает влияние на рынок труда: цифровая трансформация и риски нехватки квалифицированных кадров, пандемия коронавирусной инфекции, внешнеполитическая ситуация. В этих условиях требуется реализация грамотной государственной политики в области содействия занятости и удовлетворения потребностей экономики в квалифицированных кадрах. Поэтому при планировании и реализации государственной политики необходимо опираться на актуальную информацию о состоянии и тенденциях рынка труда.

Существует множество факторов, оказывающих влияние на рынок труда и действующие рыночные механизмы. Современный рынок труда является сложной системой, которая характеризуются высокой степенью неопределенности. В условиях быстроизменяющейся рыночной конъюнктуры важно иметь инструментарий, позволяющий оперативно фиксировать и анализировать происходящие изменения.

Использование онлайн-объявлений о вакансиях в качестве источника данных является эффективным способом отразить изменения в краткосрочном периоде. Вакансии содержат широкий круг потенциально полезной информации, обеспечивающей большую гибкость при анализе рыночной ситуации. Поэтому основная задача состоит в разработке подходов, позволяющих работать с информацией из вакансий.

Разработанные и апробированные в рамках исследовательской работы подходы к обработке данных вакансий, оценке и моделированию спроса на навыки и компетенции представляют собой универсальный инструментарий.

Таким образом, поставленные задачи и цель исследования выполнены в полном объеме, а также подтверждена гипотеза об эффективности использования современных методов обработки данных для получения актуальной и полезной для государственных органов рыночной информации.

Список литературы

1. Агеев М., Кураленок И., Некрестьянов И. ПриложениеА. Официальные метрики РОМИП’2004 //Труды второго российского семинара по оценке методов информационного поиска. Под ред. ИС Некрестьянова-Санкт-Петербург: НИИ Химии СПбГУ. – 2004. – С. 142-150.
2. Базжина В. и др. Рынок труда 2-е изд., испр. и доп. Учебник и практикум для академического бакалавриата. – Litres, 2021.
3. Бражников Ф. В., Макарова О. А. Бенчмаркинг организаций для оценки профессионально-квалификационной структуры регионального рынка труда //Эффективность труда и качество трудовой жизни XXI века. – 2020. – С. 44-55.
4. Варламова Д. и др. Атлас новых профессий 3.0. – Альпина Паблишер. – 2016. – 473 с.
5. Гурвич Е. и др. Материалы экспертной дискуссии" Проблемы прогнозирования и моделирования рынка труда в России" //Научный вестник ИЭП им. Гайдара.ру. – 2016. – №. 1. – С. 40-61.
6. Закон РФ от 19.04.1991 N 1032-1 (ред. от 19.11.2021) «О занятости населения в Российской Федерации».
7. Кубишин, Е. С. Экономика рынка труда: учебное пособие для вузов. М.: Издательство Юрайт, 2021. – 127 с.
8. Рофе А. И. Экономика труда: учебник. М.: КНОРУС, 2010. – 400 с.
9. Рынок труда: учебник и практикум для вузов / под ред. Е. Б. Яковлевой. – 2-е изд., исп. и доп. М.: Издательство Юрайт, 2021. – 253 с.
10. Цифровая экономика РФ [Электронный ресурс] // Министерство цифрового развития, связи и массовых коммуникаций Российской Федерации/ URL: https://digital.gov.ru/ru/activity/directions/858/ (дата обращения: 25.11.2021).
11. Alabdulkareem A. et al. Unpacking the polarization of workplace skills //Science advances. – 2018. – Т. 4. – №. 7. – С. 777-780.
12. Balassa B. Trade liberalisation and «revealed» comparative advantage //The manchester school. – 1965. – Т. 33. – №. 2. – С. 99-123.
13. Bojanowski P. et al. Enriching word vectors with subword information //Transactions of the association for computational linguistics. – 2017. – Т. 5. – С. 135-146.
14. Bonen T., Loree J. How to forecast skills in demand: a primer // Working Paper – 2021. – 33 с.
15. Cedefop, Eurofound (2018). Skills forecast: trends and challenges to 2030. Luxembourg: Publications Office. Cedefop reference series; No 108. – 2018. – 140 с.
16. Changing demand for skills in digital economies and societies: Literature review and case studies from low- and middle-income countries [Электронный ресурс] // International Labour Organization. 2021. URL: https://www.ilo.org/wcmsp5 /groups/public/---ed\_emp/---ifp\_skills/documents/publication/wcms\_831372.pdf (дата обращения: 25.11.2021).
17. Dawson N. et al. Predicting skill shortages in labor markets: A machine learning approach //2020 IEEE International Conference on Big Data (Big Data). – IEEE, 2020. – С. 3052-3061.
18. Deming D., Kahn L. B. Skill requirements across firms and labor markets: Evidence from job postings for professionals //Journal of Labor Economics. – 2018. – Т. 36. – №. S1. – С. S337-S369.
19. Devlin J. et al. Bert: Pre-training of deep bidirectional transformers for language understanding //arXiv preprint arXiv:1810.04805. – 2018.
20. Hershbein B., Kahn L. B. Do recessions accelerate routine-biased technological change? Evidence from vacancy postings // American Economic Review. – 2018. – Т. 108. – №. 7. – С. 1737-1772.
21. Hershbein B., Macaluso C. Labor Market Concentration and the Demand for Skills // Research output: Working paper. – 2018. – 25 c.
22. Kuratov Y., Arkhipov M. Adaptation of deep bidirectional multilingual transformers for russian language //arXiv preprint arXiv:1905.07213. – 2019.
23. Liu T. Y. et al. Learning to rank for information retrieval //Foundations and Trends® in Information Retrieval. – 2009. – Т. 3. – №. 3. – С. 225-331.
24. More C. Skill and the English working class, 1870-1914. – Taylor & Francis, 1980.
25. O\*NET OnLine [Электронный ресурс] // The Occupational Information Network. URL: https://www.onetonline.org/ (дата обращения: 25.11.2021).
26. Skills and Competencies Taxonomy [Электронный ресурс] // Government of Canada. URL: https://noc.esdc.gc.ca/SkillsTaxonomy/Skills TaxonomyWelcome/f5ec9457d5a540eeb4529c2698acb19a (дата обращения: 25.11.2021).
27. Skills-OVATE [Электронный ресурс] // Cedefop URL: https://www.cedefop.europa.eu/en/tools/skills-online-vacancies (дата обращения: 28.12.2021).
28. V'yugin V. V. Matematicheskie osnovy teorii-mashinnogo obucheniya I prognozirovaniya [Mathematical foundations of the theory of machine learning and forecasting] //MCCME E. – 2018. – С. 384.
29. Weinert F. E. Concepts of Competence (Definition and Selection of Competencies: Theoretical and Conceptual Foundations–DeSeCo) //Max Planck Institute for Psychological Research. Munich, Germany. – 1999.

Приложение 1



Рис.  1 – Нейронная сеть, используемая для классификации текста с несколькими метками

Источник: составлено автором

Приложение 2



Рис.  1 – Основные гиперпараметры моделей ruBERT/ruBERT\_C

Источник: составлено автором



Рис.  2 – Основные гиперпараметры моделей mBERT/mBERT\_C

Источник: составлено автором

Приложение 3

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **precision** | **recall** | **f1-score** | **support** |
| <70K | 0.71 | 0.87 | 0.79 | 1689 |
| 70K-140K | 0.66 | 0.80 | 0.72 | 1426 |
| 140K-210K | 0.74 | 0.47 | 0.57 | 440 |
| 210K-280K | 0.78 | 0.41 | 0.53 | 292 |
| 280K-350K | 0.64 | 0.33 | 0.44 | 237 |
| 350K-420K | 0.68 | 0.40 | 0.51 | 225 |
| >420K | 0.74 | 0.43 | 0.55 | 362 |
|  |  |  |  |  |
|  |  |  |  |  |
| **accuracy** |  |  | 0.70 | 4671 |
| **macro avg** | 0.71 | 0.53 | 0.59 | 4671 |
| **weighted avg** | 0.70 | 0.70 | 0.68 | 4671 |

Рис.  1 – Random Forest

Источник: составлено автором

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **precision** | **recall** | **f1-score** | **support** |
| <70K | 0.76 | 0.87 | 0.81 | 1689 |
| 70K-140K | 0.68 | 0.80 | 0.74 | 1426 |
| 140K-210K | 0.69 | 0.51 | 0.59 | 440 |
| 210K-280K | 0.72 | 0.44 | 0.55 | 292 |
| 280K-350K | 0.62 | 0.38 | 0.47 | 237 |
| 350K-420K | 0.63 | 0.42 | 0.51 | 225 |
| >420K | 0.66 | 0.49 | 0.56 | 362 |
|  |  |  |  |  |
|  |  |  |  |  |
| **accuracy** |  |  | 0.71 | 4671 |
| **macro avg** | 0.68 | 0.56 | 0.60 | 4671 |
| **weighted avg** | 0.71 | 0.71 | 0.70 | 4671 |

Рис.  2 – Support Vector Machine

Источник: составлено автором

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **precision** | **recall** | **f1-score** | **support** |
| <70K | 0.65 | 0.83 | 0.73 | 1689 |
| 70K-140K | 0.58 | 0.69 | 0.63 | 1426 |
| 140K-210K | 0.48 | 0.25 | 0.33 | 440 |
| 210K-280K | 0.51 | 0.25 | 0.33 | 292 |
| 280K-350K | 0.45 | 0.17 | 0.25 | 237 |
| 350K-420K | 0.49 | 0.23 | 0.31 | 225 |
| >420K | 0.49 | 0.32 | 0.39 | 362 |
|  |  |  |  |  |
|  |  |  |  |  |
| **accuracy** |  |  | 0.60 | 4671 |
| **macro avg** | 0.52 | 0.39 | 0.42 | 4671 |
| **weighted avg** | 0.57 | 0.60 | 0.57 | 4671 |

Рис.  3 – Stochastic Gradient Descent

Источник: составлено автором

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **precision** | **recall** | **f1-score** | **support** |
| <70K | 0.59 | 0.84 | 0.70 | 1689 |
| 70K-140K | 0.53 | 0.70 | 0.60 | 1426 |
| 140K-210K | 0.62 | 0.20 | 0.30 | 440 |
| 210K-280K | 0.79 | 0.18 | 0.29 | 292 |
| 280K-350K | 0.78 | 0.14 | 0.23 | 237 |
| 350K-420K | 0.72 | 0.13 | 0.22 | 225 |
| >420K | 0.76 | 0.16 | 0.27 | 362 |
|  |  |  |  |  |
|  |  |  |  |  |
| **accuracy** |  |  | 0.58 | 4671 |
| **macro avg** | 0.68 | 0.34 | 0.37 | 4671 |
| **weighted avg** | 0.62 | 0.58 | 0.53 | 4671 |

Рис.  4 – XGBoost

Источник: составлено автором

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **precision** | **recall** | **f1-score** | **support** |
| <70K | 0.75 | 0.76 | 0.75 | 1689 |
| 70K-140K | 0.64 | 0.67 | 0.65 | 1426 |
| 140K-210K | 0.48 | 0.46 | 0.47 | 440 |
| 210K-280K | 0.46 | 0.39 | 0.42 | 292 |
| 280K-350K | 0.38 | 0.35 | 0.36 | 237 |
| 350K-420K | 0.40 | 0.36 | 0.28 | 225 |
| >420K | 0.46 | 0.48 | 0.47 | 362 |
|  |  |  |  |  |
|  |  |  |  |  |
| **accuracy** |  |  | 0.62 | 4671 |
| **macro avg** | 0.51 | 0.50 | 0.50 | 4671 |
| **weighted avg** | 0.61 | 0.62 | 0.62 | 4671 |

Рис.  5 – Decision Tree

Источник: составлено автором

Приложение 4

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **precision** | **recall** | **f1-score** | **support** |
| <70K | 0.71 | 0.87 | 0.78 | 1689 |
| 70K-140K | 0.64 | 0.78 | 0.71 | 1426 |
| 140K-210K | 0.61 | 0.41 | 0.49 | 422 |
| 210K-280K | 0.66 | 0.29 | 0.40 | 292 |
| 280K-350K | 0.60 | 0.21 | 0.31 | 237 |
| 350K-420K | 0.61 | 0.29 | 0.40 | 241 |
| >420K | 0.61 | 0.42 | 0.50 | 368 |
|  |  |  |  |  |
|  |  |  |  |  |
| **accuracy** |  |  | 0.67 | 4671 |
| **macro avg** | 0.64 | 0.47 | 0.51 | 4671 |
| **weighted avg** | 0.66 | 0.67 | 0.65 | 4671 |

Рис.  1 – FastText

Источник: составлено автором

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **precision** | **recall** | **f1-score** | **support** |
| <70K | 0.72 | 0.89 | 0.80 | 1689 |
| 70K-140K | 0.69 | 0.76 | 0.72 | 1426 |
| 140K-210K | 0.59 | 0.46 | 0.52 | 422 |
| 210K-280K | 0.65 | 0.38 | 0.48 | 292 |
| 280K-350K | 0.61 | 0.29 | 0.39 | 237 |
| 350K-420K | 0.70 | 0.41 | 0.52 | 241 |
| >420K | 0.63 | 0.48 | 0.54 | 368 |
|  |  |  |  |  |
|  |  |  |  |  |
| **accuracy** |  |  | 0.69 | 4671 |
| **macro avg** | 0.66 | 0.52 | 0.57 | 4671 |
| **weighted avg** | 0.68 | 0.69 | 0.67 | 4671 |

Рис.  2 – CNN+ FastText

Источник: составлено автором

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **precision** | **recall** | **f1-score** | **support** |
| <70K | 0.42 | 0.98 | 0.59 | 1689 |
| 70K-140K | 0.71 | 0.14 | 0.23 | 1426 |
| 140K-210K | 0.00 | 0.00 | 0.00 | 422 |
| 210K-280K | 0.89 | 0.03 | 0.06 | 292 |
| 280K-350K | 0.00 | 0.00 | 0.00 | 237 |
| 350K-420K | 0.00 | 0.00 | 0.00 | 241 |
| >420K | 0.18 | 0.23 | 0.20 | 368 |
|  |  |  |  |  |
|  |  |  |  |  |
| **accuracy** |  |  | 0.42 | 4671 |
| **macro avg** | 0.32 | 0.20 | 0.15 | 4671 |
| **weighted avg** | 0.44 | 0.42 | 0.31 | 4671 |

Рис.  3 – LSTM+FastTex

Источник: составлено автором

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **precision** | **recall** | **f1-score** | **support** |
| <70K | 0.76 | 0.78 | 0.77 | 1689 |
| 70K-140K | 0.69 | 0.68 | 0.69 | 1426 |
| 140K-210K | 0.44 | 0.55 | 0.49 | 422 |
| 210K-280K | 0.39 | 0.42 | 0.46 | 292 |
| 280K-350K | 0.41 | 0.37 | 0.39 | 237 |
| 350K-420K | 0.44 | 0.42 | 0.49 | 241 |
| >420K | 0.56 | 0.49 | 0.53 | 368 |
|  |  |  |  |  |
|  |  |  |  |  |
| **accuracy** |  |  | 0.64 | 4671 |
| **macro avg** | 0.54 | 0.53 | 0.53 | 4671 |
| **weighted avg** | 0.64 | 0.64 | 0.64 | 4671 |

Рис.  4 – BiLSTM+FastText

Источник: составлено автором

Приложение 5

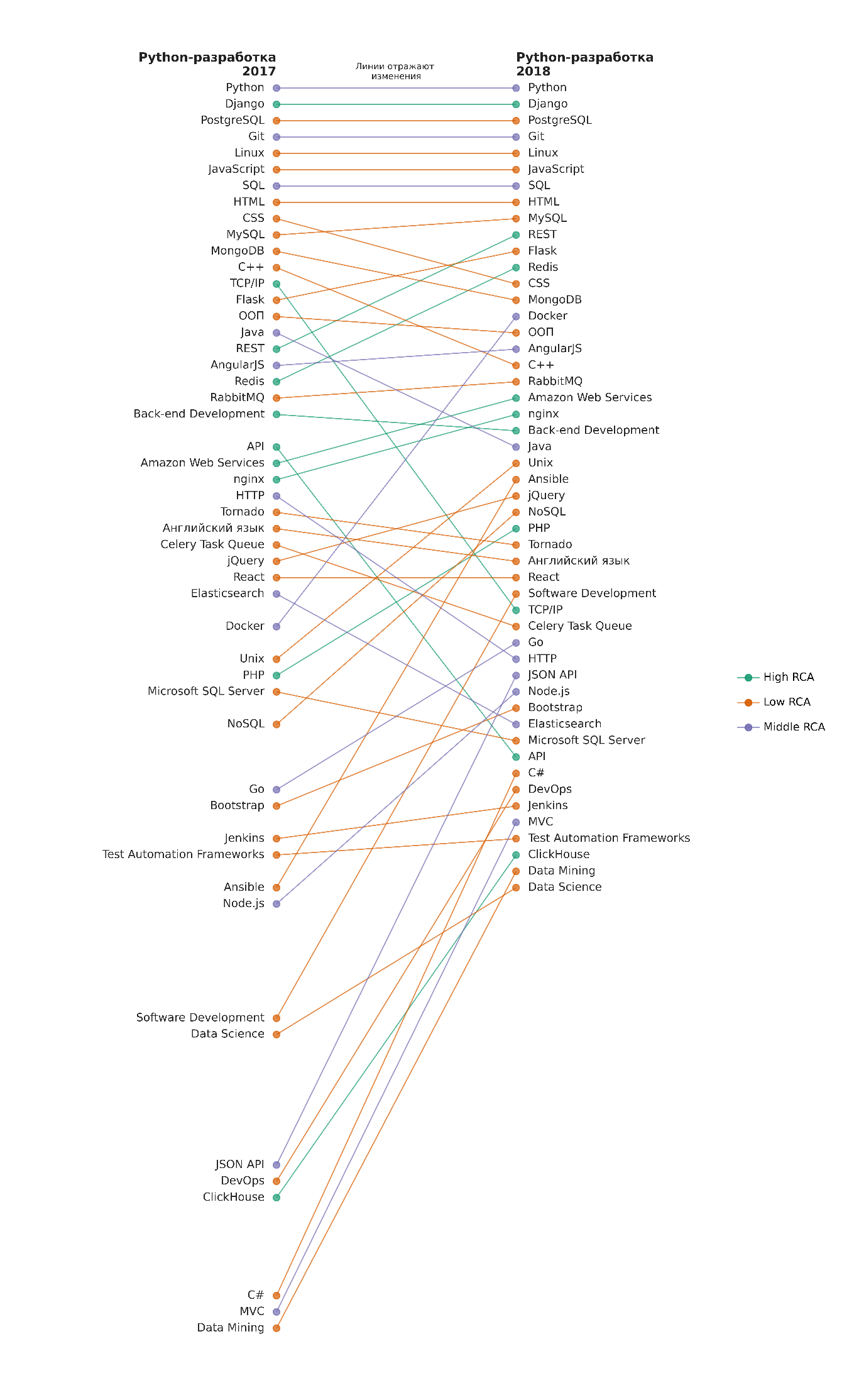


Рис.  1 – ТОП-50 навыков и компетенций для группы

«Python-разработка»

Источник: составлено автором

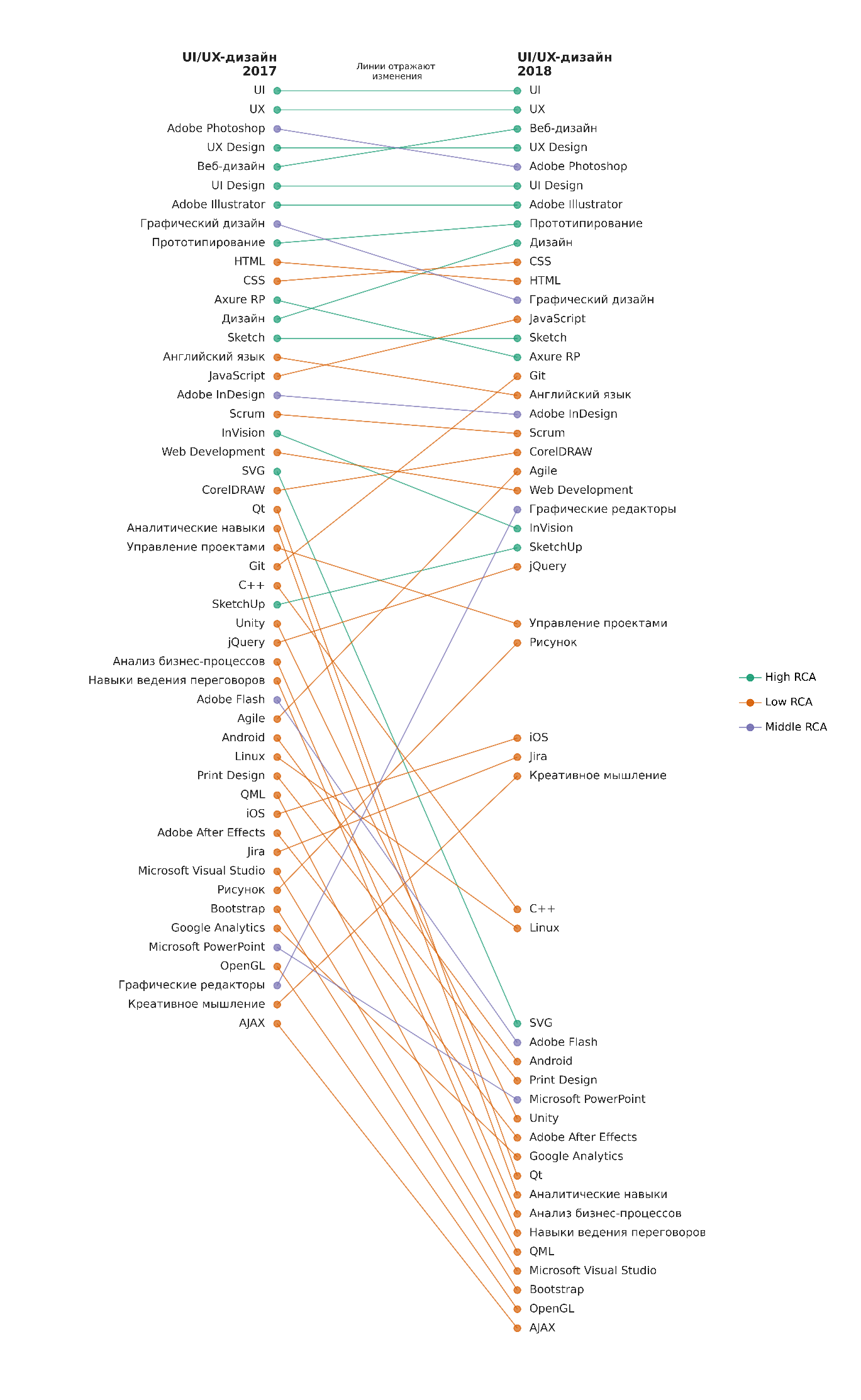


Рис.  2 – ТОП-50 навыков и компетенций для группы

«UI/UX-дизайн»

Источник: составлено автором

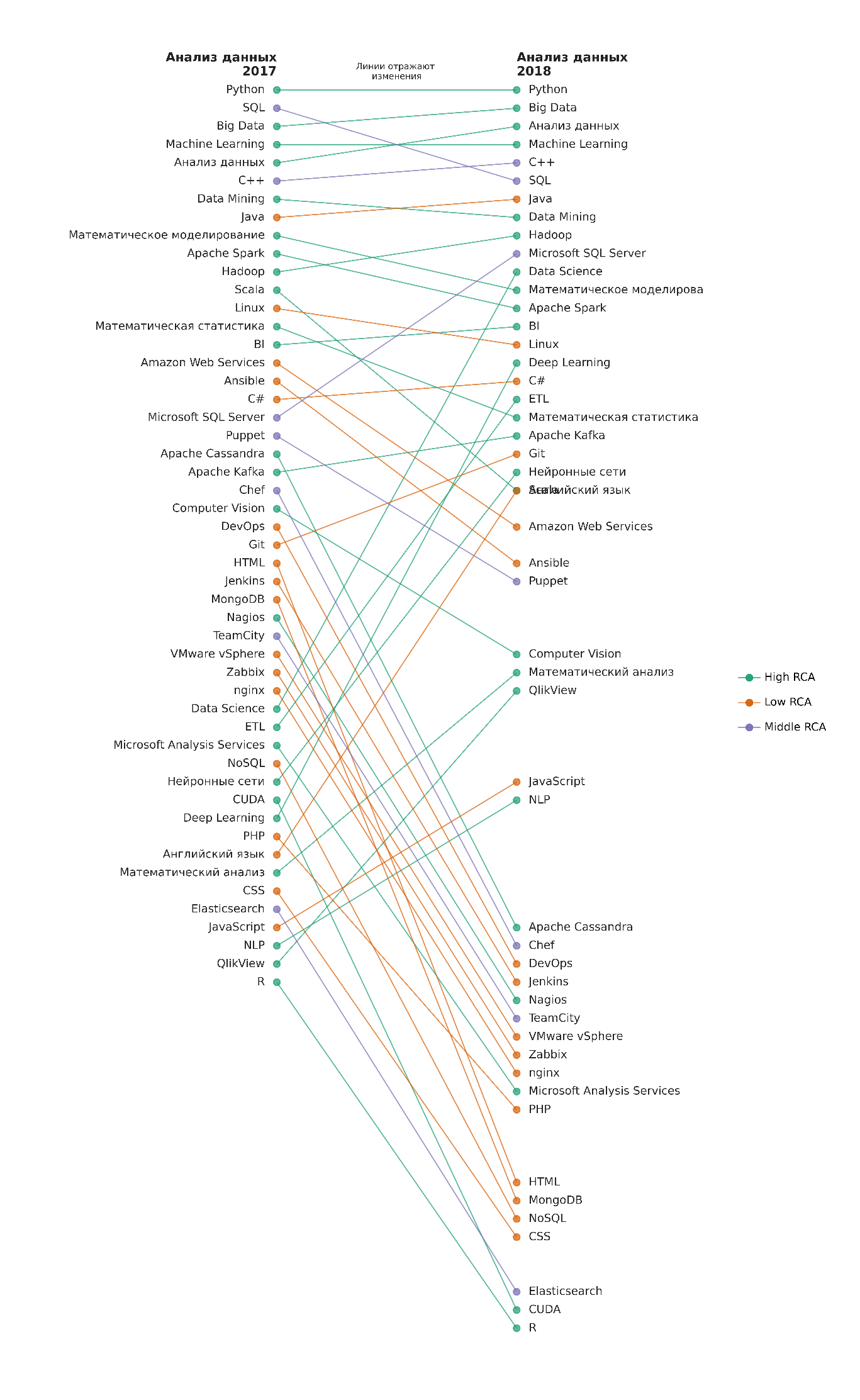


Рис.  3 – ТОП-50 навыков и компетенций для группы

«Анализ данных»

Источник: составлено автором

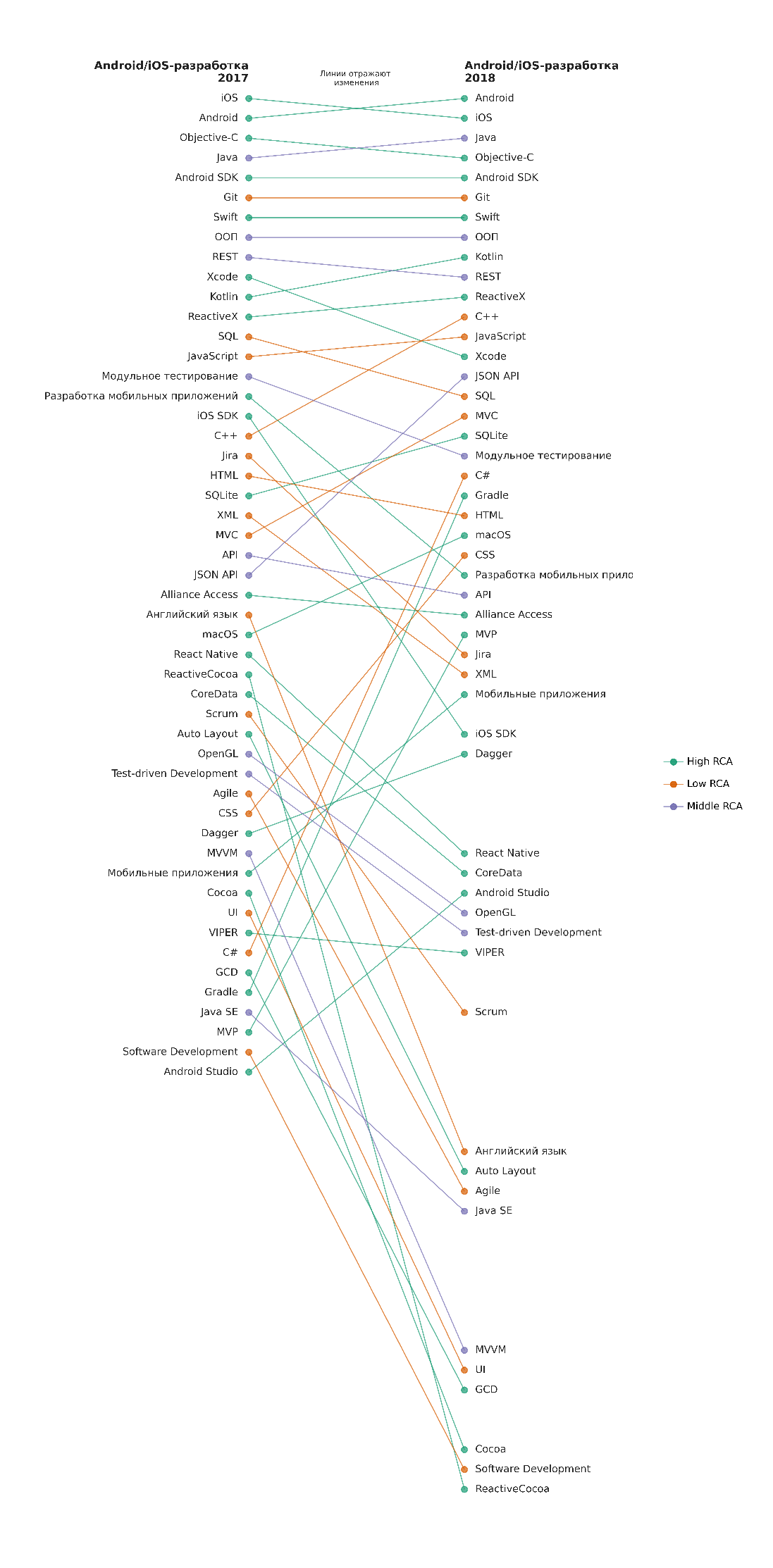


Рис.  4 – ТОП-50 навыков и компетенций для группы

«Android/iOS-разработка»

Источник: составлено автором