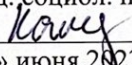


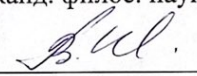
Министерство науки и высшего образования Российской Федерации
НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
ТОМСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ (НИ ТГУ)
Философский факультет

ДОПУСТИТЬ К ЗАЩИТЕ В ГЭК
Руководитель ООП
канд. социол. наук, доцент
 В.В. Кашпур
«15» июня 2023 г.

ВЫПУСКНАЯ КВАЛИФИКАЦИОННАЯ РАБОТА БАКАЛАВРА
БОЛЬШИЕ ДАННЫЕ В КОНТЕКСТЕ АНАЛИТИКИ РЫНКА ТРУДА

по направлению подготовки 39.03.01 Социология
направленность (профиль) «Социология»

Остапенко Мария Александровна

Руководитель ВКР
канд. филос. наук, доцент
 В.С. Иванова
«09» июня 2023 г.

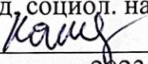
Автор работы
студент группы № 121903
 М.А. Остапенко
«09» июня 2023 г.

Министерство науки и высшего образования Российской Федерации
НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
ТОМСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ (НИ ТГУ)
Философский факультет

УТВЕРЖДАЮ

Руководитель ООП

канд. социол. наук, доцент

 В.В. Кашпур

«15» июня 2023 г.

ЗАДАНИЕ

по выполнению выпускной квалификационной работы бакалавра обучающегося
Остапенко Марии Александровне

(Ф.И.О. обучающегося)

по направлению подготовки 39.03.01 Социология, направленность (профиль) «Социология»

1 Тема выпускной квалификационной работы бакалавра

Большие данные в контексте аналитики рынка труда

2 Срок сдачи обучающимся выполненной выпускной квалификационной работы:

а) на кафедру – 09 июня 2023 г.

б) в ГЭК – 15 июня 2023 г.

3 Исходные данные к работе:

Объект исследования – Метод анализа больших данных

Предмет исследования – Возможности и ограничения метода анализа больших
в исследованиях рынка труда

Цель исследования – Выявить возможности и ограничения применения анализа
больших данных в исследованиях рынка труда

Задачи:

1) Проанализировать основные подходы к понятию, сущности рынка труда, а также его показатели в современной ситуации; 2) Описать особенности метода BigData, а также его возможности; 3) Проанализировать платформы по мониторингу рынка труда и выявить показатели рынка труда, доступные для анализа; 4) Описать специфику и проблематику применения метода анализа больших данных в исследованиях рынка труда; 5) Выявить перспективы развития анализа рынка труда с помощью метода Big Data

Методы исследования

Теоретический анализ литературы отечественных и зарубежных исследователей, метод сбора информации – полуструктурированное интервью, метод отбора единиц наблюдения - снежный ком.

Организация или отрасль, по тематике которой выполняется работа

Экономическая социология

4. Краткое содержание работы

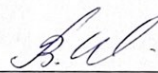
Глава 1. Теоретические основания изучения больших данных в контексте аналитики рынка труда. Подходы к изучению рынка труда. Показатели рынка труда в современной ситуации. Понятие и сущность больших данных. Глава 2. Специфика применения больших данных в исследованиях рынка труда

Руководитель выпускной квалификационной работы
Доцент кафедры социологии

(должность, место работы)

Задание принял к исполнению
студент гр. 121903

(должность, место работы)



(подпись)

В.С. Иванова

(И.О. Фамилия)



(подпись)

М.А. Остапенко

(И.О. Фамилия)

АННОТАЦИЯ

В контексте аналитики рынка труда большие данные представляют собой огромные объемы информации, которые собираются и анализируются для выявления тенденций и закономерностей на рынке труда. Данная работа направлена на изучение возможностей и ограничений применения больших данных (Big Data) в аналитике рынка труда. Проведен обзор платформ, предназначенных для аналитики рынка труда. Выявлено мнение экспертов по использованию метода анализа больших данных в исследованиях рынка труда. Приводится пример использования сервиса, предназначенного для мониторинга рынка труда – «РосНавык».

Одним из главных преимуществ использования больших данных является возможность получения более точной и актуальной информации о рынке труда. Анализ данных позволяет выявить тенденции и закономерности на рынке труда, определить спрос на различные профессии и специальности. Кроме того, использование больших данных позволяет определить наиболее востребованные профессии и специальности на рынке труда. Однако, использование больших данных также имеет свои ограничения. Например, данные могут быть неполными или неточными, что может привести к ошибкам в анализе. Кроме того, обработка больших объемов данных требует значительных ресурсов и времени.

Таким образом, анализ больших данных представляет собой важный инструмент для анализа рынка труда. Однако, для достижения максимальной эффективности необходимо учитывать ограничения и возможности данного метода.

ОГЛАВЛЕНИЕ

Введение.....	5
1 Теоретические основания изучения больших данных в исследованиях рынка труда	8
1.1 Подходы к изучению рынка труда и его основные показатели	8
1.2 Понятие и сущность больших данных.....	16
2 Применение метода анализа больших данных в исследованиях рынка труда	23
2.1 Платформы по мониторингу рынка труда	23
2.2 Специфика и проблематика применения анализа больших данных в исследованиях рынка труда.....	27
2.3 Перспективы развития анализа рынка труда с помощью методов Big Data.....	34
2.4 Пример анализа рынка труда с помощью больших данных	37
Заключение.....	42
Литература	44
Приложение А Программа исследования	48
Приложение Б Гайд экспертного интервью.....	50
Приложение В Список экспертов.....	51
Приложение Г Расшифровки интервью	52

ВВЕДЕНИЕ

Рынок труда – это система взаимоотношений между людьми, нуждающимися в работе, и организациями, выступающими в качестве работодателей, система, которая постоянно меняется, и для ее анализа требуются новые методы. Большие данные актуальны в этой сфере, потому что с их помощью можно обрабатывать огромное количество информации.

Большие данные (Big Data) – это огромный объем данных, который сложно обработать с помощью традиционных методов обработки данных. Большие данные могут быть структурированными и неструктурированными, могут храниться в различных форматах, таких как текстовые файлы, таблицы, изображения, видео и т.д. Анализа больших данных предполагает использование набора инструментов и технологий, которые используются для обработки и анализа огромных объемов данных. Он включает в себя различные алгоритмы машинного обучения, статистические методы, графовые алгоритмы и другие подходы. Цель метода анализа больших данных – выявить скрытые закономерности и взаимосвязи в данных, сделать прогнозы и принять обоснованные решения на основе полученных результатов.

Актуальность: Широкое распространение интернета и развитие цифровых технологий приводят к изменениям в социальных науках, появляются новые источники данных и развиваются методы для их сбора, обработки и анализа. Особенностью таких данных является то, что они созданы не для целей исследования, охватывают всю популяцию и производятся в режиме реального времени.

По данным оценки экспертов Fortune Business Insights объем мирового рынка аналитики больших данных в 2021 году оценивался в 240,56 млрд долларов и по прогнозам вырастет до 655,53 млрд долларов к 2029 году, при этом среднегодовой темп роста составит 13,4% в течение прогнозируемого периода. Анализ показал, что в среднем, рост в 2020 году составлял 12,5%, в сравнении с 2019 годом [39].

В России в 2021 г. технологии сбора, обработки и анализа больших данных применяли 25,8% организаций — на 3,4 п. п. больше, чем в 2020 г. Самым популярным источником оказались веб-сайты компаний: их данные собирают 9,2% компаний. Немного менее востребованы данные учетных систем организации (ERP, CRM и др.) и социальных сетей (8 и 7,2% соответственно). Помимо этого, анализируются данные операторов сотовой связи (6,7%), информация с цифровых датчиков и радиочастотных меток (6,3%) [13].

Что касается рынка труда, наиболее доступным источником информации о нем является Интернет-пространство. На специализированных сайтах по поиску работы, в

объявлениях о вакансиях содержатся требования к соискателю, необходимые навыки и перечень обязанностей. На данный момент в сети существует множество сайтов по трудоустройству и объёмы информации продолжают расти, поскольку ежедневно публикуются новые резюме и вакансии. Так, по оценкам hh.ru Россия, на конец сентября 2022 года количество открытых вакансий на сайте – 275393 [15].

Выделяют несколько подходов к изучению рынка труда, Классическая теория, авторами являются А. Смит и Д. Рикардо, доказывает, что рыночная система может поддерживать полную занятость благодаря стабильности спроса и предложения на рынке труда, что является основой для благосостояния нации. В неоклассической теории (А. Лаффер, Д. Гилд, Р. Холл) предполагается, что рынок труда является внутренне неоднородной и динамичной системой, где цена труда быстро реагирует на изменения спроса и предложения [30].

Таким образом, технологии Big Data (Большие данные) является одним из перспективных направлений в аналитике рынка труда, поскольку они предполагают работу с информацией огромного объема и разнообразного состава, которая постоянно обновляется и может находиться в различных источниках.

Проблемная ситуация: С одной стороны, метод анализа больших данных обладает такими преимуществами как скорость сбора информации, высокая достоверность и возможность выявления закономерностей, а с другой стороны, данный метод имеет ряд недостатков, например, таких как неструктурированная информация и сложность ее интерпретации. Преимущества и недостатки данного метода сказываются на возможностях и ограничениях применения методов Big Data в анализе рынка труда

Проблема: Каковы особенности применения анализа больших данных в исследованиях рынка труда? В чем заключаются возможности и ограничения больших данных исследованиях рынка труда?

Объект: В теоретическом аспекте: Метод анализа больших данных; В эмпирическом аспекте: Мнения экспертов в оценке применения больших данных в исследованиях рынка труда

Предмет: Возможности и ограничения анализа больших данных при исследовании рынка труда

Цель: Выявить возможности и ограничения применения анализа больших данных в исследованиях рынка труда

Задачи: В теоретическом аспекте:

- Проанализировать основные подходы к понятию, сущности рынка труда, а также его показатели в современной ситуации

- Описать особенности метода BigData, а также его возможности

В эмпирическом аспекте:

- Проанализировать платформы по мониторингу рынка труда и выявить показатели рынка труда, доступные для анализа
- Описать специфику и проблематику применения метода анализа больших данных в исследованиях рынка труда
- Выявить перспективы развития анализа рынка труда с помощью метода Big Data

Теоретико-методологическими основаниями для данной работы являются: подход к рынку труда новой экономической социологии Грановеттера, понятие человеческого капитала Шульца и Беккера, набор признаков больших данных – VVV (volume, velocity, variety) выработанных Meta Group.

Новизна: Большие данные позволяют собирать, обрабатывать и анализировать огромный объем информации о рынке труда, который раньше был недоступен. Это дает возможность более точно и полно изучать спрос и предложение, а также выявлять новые тенденции на рынке труда.

Практическая значимость: Данная работа направлена на изучение возможностей больших данных в анализе рынка труда, понимание которых позволит повысить эффективность принятия решений на рынке труда за счет более точного и глубокого анализа данных.

Структура работы: ВКР состоит из введения, двух глав, заключения, списка литературы и приложений (программа исследования, гайд экспертного интервью, список экспертов, расшифровки интервью). Первая глава состоит из двух параграфов. Вторая глава состоит из четырех параграфов.

1 Теоретические основания изучения больших данных в исследованиях рынка труда

1.1 Подходы к изучению рынка труда и его основные показатели

Под рынком труда принято понимать место, где сталкиваются спрос и предложение рабочей силы. Это означает, что между работодателями и работниками происходит взаимодействие на рынке труда, чтобы найти лучшие условия для трудоустройства. Рынок труда может быть местным, региональным, национальным или международным и может охватывать различные отрасли и профессии.

В рамках экономической теории рынок труда рассматривается как отношение наемных работников и работодателей и как динамическая система, которая включает, с одной стороны, работодателей (владельцев средств производства) и рабочих (наемных работников), с другой стороны, формирует размер, структуру и соотношение спроса и предложения на рабочую силу [9].

Выделяют несколько теоретических подходов в изучении рынка труда. Первый из них – это классическая теория (А. Смит и Д. Рикардо), согласно авторам, труд – это особый товар, который имеет свою естественную и рыночную стоимость. А. Смит уделял особое внимание изучению возможностей рынка труда, анализу компонентов заработной платы, причин безработицы, а также особенностей предложения труда и спроса на рабочую силу. Стоит отметить важность анализа формальных институтов на рынке труда. Важным аспектом является выделение производительного и непроизводительного типов труда. Классическая теория доказывает, что рыночная система может поддерживать полную занятость благодаря стабильности спроса и предложения на рынке труда, что является основой для благосостояния нации [30].

В неоклассической теории (А. Лаффер, Д. Гилд, Р. Холл) доход рассматривается как то, что создается всеми факторами производства, в отличие от классической теории, где все виды доходов создаются трудом. Теория предполагает, что рынок труда является внутренне неоднородной и динамичной системой, где цена труда быстро реагирует на изменения спроса и предложения. Неоклассические экономисты считают, что заработная плата может быть повышена или снижена в зависимости от спроса и предложения на рабочую силу, и это может привести к изменениям в уровне безработицы. Если предложение рабочей силы превышает спрос, то это может привести к снижению заработной платы и появлению безработицы, а если спрос превышает предложение, то заработная плата может повышаться. Неоклассическая теория также предполагает, что рыночный механизм является регулятором рынка труда, который обеспечивает равновесие между спросом и предложением на рабочую

силу. Если на рынке наблюдается избыток предложения рабочей силы, то цена труда будет снижаться, что приведет к уменьшению заработной платы и увеличению количества безработных. Если же спрос на рабочую силу превышает предложение, то цена на труд будет повышаться, что может привести к увеличению заработной платы и уменьшению числа безработных. Таким образом, рынок труда регулируется рыночным механизмом, который быстро реагирует на изменения в спросе и предложении на рабочую силу и обеспечивает равновесие между ними. Это помогает поддерживать стабильность на рынке труда и обеспечивать эффективность его функционирования [30].

В социологии рынок труда рассматривается как социальный институт, который регулирует отношения между работодателями и работниками. Рынок труда охватывает не только экономические, но и социальные и культурные аспекты, такие как социальный статус, профессиональная идентичность, возможности карьерного роста и т.д. В социологии рынок труда также рассматривается в контексте социальной стратификации, то есть как механизм, определяющий социальное положение и доходы людей в обществе.

Рынок труда изучается с помощью различных методов и подходов. Одним из основных подходов является анализ структуры рынка труда, который включает изучение различных категорий работников (например, по возрасту, полу, образованию, опыту работы и т. д.), а также виды занятости (например, полный или неполный рабочий день, временная или постоянная работа).

Другой подход заключается в анализе процессов, происходящих на рынке труда, таких как поиск работы, наем и увольнение сотрудников, формирование заработной платы и т. д. в рамках этого подхода социологи изучают, какие факторы влияют на эти процессы, как они относятся к другим социальным институтам (например, образованию, семье и т.д.) и как они влияют на жизнь работников и общество в целом.

В социологии также изучаются проблемы, связанные с неравенством на рынке труда, такие как дискриминация по признаку пола, возраста, национальности и т. д., а также проблемы безработицы и низкой заработной платы. В целом социологический подход к изучению рынка труда позволяет понять его роль в общественной жизни и выявить проблемы, связанные с его функционированием.

Институциональный подход к исследованию рынка труда включает анализ институтов, таких как законы и правила, регулирующие трудовые отношения, системы социального обеспечения, профсоюзы, работодатели и т. д. Институциональный подход также учитывает влияние культурных и исторических факторов на создание институтов и их влияние на рынок труда. Данный подход позволяет понять, как институты влияют на поведение работников и работодателей, как они формируют и поддерживают определенные

структуры и отношения на рынке труда и как они могут быть изменены для улучшения условий труда и занятости [24].

Рынок труда принято понимать в контексте отношений между покупателями и продавцами труда или в качестве экономических отношений купли-продажи рабочей силы [24]. Институциональный подход подразумевает представление рынка труда как системы взаимодействующих элементов. Согласно Д. Норту, институционализированные системы складываются из институтов разного уровня, таким образом, в рынок труда входят другие как формальные, так и неформальные институты [5]. Среди формальных институтов, которые влияют на функционирование рынка труда и регулируют его положение, во-первых, это политические институты. Политические институты, такие как правительство, законодательство и судебная система, могут оказывать значительное влияние на рынок труда [25]. Например, в некоторых странах правительство может устанавливать законы, которые ограничивают или запрещают определенные виды деятельности, что может повлиять на количество доступных рабочих мест и их качество. Также, правительство может принимать законы, направленные на улучшение условий труда и защиту прав работников. Законодательство может также регулировать трудовые отношения между работодателями и работниками, устанавливать правила найма и увольнения, определять минимальную заработную плату и другие условия труда. Кроме того, суды могут принимать решения, которые влияют на трудовые отношения, например, разрешать споры между работодателями и работниками или определять, какие законы были нарушены. Политические институты могут оказывать как положительное, так и отрицательное влияние на рынок труда, в зависимости от того, какие законы и правила они устанавливают и как они применяются на практике.

Во-вторых, к формальным институтам, которые оказывают влияние на рынок труда относятся экономические институты, такие как налоговая система, банковская система, система социального обеспечения и т.д. [24]. налоговая система может влиять на доходы населения и, соответственно, на спрос на рабочую силу. Если налоговая система предусматривает высокие налоги на заработную плату, то это может снизить спрос на работников с высокой заработной платой и стимулировать компании к найму работников с более низкой заработной платой. Банковская система также может влиять на рынок труда, особенно если она предоставляет кредиты на развитие бизнеса или инвестирование в новые технологии. Это может стимулировать компании нанимать больше работников для развития своего бизнеса и внедрения новых технологий, что в свою очередь может привести к росту спроса на рабочую силу в определенных отраслях. Система социального обеспечения, например, может влиять на предложение рабочей силы, поскольку это может увеличить

количество людей, которые готовы работать на низкооплачиваемых работах, чтобы получить социальное обеспечение.

Для выполнения функций формальных институтов, необходимы специальные структуры, такие как профсоюзы, объединения работников и работодателей; министерства; образовательные учреждения; службы занятости; кадровые агентства; биржи труда; трудовая инспекция; общественные организации; фирмы и другие организации [24].

Системный подход в социологии рынка труда предполагает изучение рынка труда как системы, состоящей из множества взаимосвязанных элементов. В рамках системного подхода рынок труда рассматривается как открытая система, которая взаимодействует с другими системами. Во-первых, рынок труда представляет собой социально – экономическую систему, центральный элемент которой – совокупность обменных отношений-определяется социальным производством и содержащимися в нем имущественными отношениями. Во-вторых, он выражает систему трудовых отношений (взаимосвязей и взаимодействий) субъектов различного уровня, управляемых объективно обусловленными мотивами и целями. В-третьих, рынок труда отражает определенную систему социальных отношений его субъектов, связанных с формированием способов вовлечения населения в хозяйственную деятельность. В-четвертых, рынок труда – это система институтов, имеющих свой сложный, самоуправляемый механизм, элементы которого взаимосвязаны, взаимозависимы и дополнены внешними регуляторными механизмами. Взгляд на рынок труда как систему взаимоотношений с формированием занятости и доходов от труда позволяет отнести рынок труда к особым социально-экономическим системам в связи с особыми экономическими интересами его субъектов. С точки зрения характеристик целостности, относительной обособленности, связей с окружающей средой, структурирования, подчинения одной цели - рынок труда не только удовлетворяет всем необходимым и достаточным свойствам сложной системы, но и обладает внутренней целостностью [10].

Основными элементами системного подхода к рынку труда являются: Рабочая сила – совокупность людей, обладающих определенными навыками и квалификацией, которые могут быть использованы для производства товаров и услуг. Спрос на рабочую силу – потребность работодателей в рабочей силе для выполнения определенных задач. Предложение рабочей силы – это количество людей, готовых работать за определенную заработную плату. Заработная плата – вознаграждение, которое работодатель платит работнику за выполнение определенных задач. Конкуренция на рынке труда – борьба между работодателями за привлечение наиболее квалифицированных работников. Регулирующие

органы – это государственные и общественные организации, которые устанавливают правила и нормы для регулирования рынка труда.

Системный подход к изучению рынка труда позволяет более глубоко понять его функционирование и взаимодействие между его элементами. Он также помогает разработать стратегии по улучшению условий труда и повышению его эффективности.

В 1980-е гг. в экономической социологии произошел подъем, связанный с появлением новой дисциплины, называемой «новой экономической социологией». Эта дисциплина возникла в ответ на необходимость более глубокого изучения экономических процессов и их влияния на общество. Импульсом являлись работы 1950-1970-х годов Парсонса и Смэлсера. В рамках новой экономической социологии были разработаны новые методы и подходы к изучению экономических явлений. Одним из главных принципов новой экономической социологии является понимание того, что экономика не существует изолированно от других сфер общества. Новая экономическая социология изучает экономические процессы с точки зрения социальных отношений и культурных ценностей и учитывает влияние политических и институциональных факторов на экономику. Это позволяет глубже понять экономические явления и разработать более эффективные стратегии развития экономики.

Началом новой экономической социологии считают работу М. Грановеттера «Экономическое действие и социальная структура: проблема укорененности». Новая экономическая социология основывается на трех основных направлениях: теория сетей, теория организаций и социология культуры [42].

Грановеттер считает, что необходимо изменить порядок слов в слогане «Economy and Society» на «Society and Economy», чтобы подчеркнуть социальную и экономическую интеграцию. Он противопоставляет структурно-функциональный анализ и предлагает новую экономическую социологию, основанную на теории социальной интеграции экономики К. Поланьи и на социальном построении экономических институтов (по аналогии с социальным конструированием Бергера и Лукмана). Одним из основных методологических инструментов новой экономической социологии является сетевой подход, который предполагает сеть социальных отношений и встроенных в нее экономических институтов. Несмотря на методологический коллаж в новой экономической социологии, он представляет собой новый и важный шаг по сравнению со структурно-функциональной традицией, открывая новые перспективы для экономико-социологических исследований [42].

Связь социального и экономического отражается в пересечении спроса и предложения на труд, так с рынком труда связано понятие «человеческого капитала». Эта концепция появилась во второй половине XX века в работах американских ученых-экономистов Т. Шульца и Г. Беккера. Человеческий капитал - это совокупность накопленных знаний,

навыков, умений, профессиональных компетенций, которыми человек обладает как работник и которые приобретаются им через общее, специальное среднее и высшее или дополнительное образование, а также через опыт работы. Человеческий капитал формируется в определенных культурных, экономических условиях: путем инвестиций в воспитание, образование, здоровье, науку, в интеллектуальную и духовную деятельности, в создание условий для предпринимательства, поддержку активных профессионалов, в формирование эффективной экономической, политической, научной и культурной элиты. Человеческий капитал формируется в результате притока из других стран или уменьшается в результате его оттока [36].

Отечественный социолог Радаев В.В. придерживался экономического понимания человеческого капитала – «совокупность накопленных профессиональных знаний, умений и навыков, получаемых в процессе образования и повышения квалификации, которые впоследствии могут приносить доход – в виде заработной платы, процента или прибыли» [28].

При устройстве на работу, работодатель учитывает соответствие кандидатов необходимым требованиям и наличие у него необходимых компетенций. Компетенции – это навыки и знания, овладев которыми человек становится более осведомленным в области их применения. Соискатель, который обладает большим числом различного рода компетенций, является более приоритетным кандидатом на должность, поскольку результат их деятельности будет приносить более значимый эффект.

По данным Сибирского федерального университета, проведенный в 2014 году опрос работодателей показал, что среди наиболее популярных требований к соискателям: ответственность (отметили 25% работодателей), инициативность (21%), законченное образование (18%) и высокий уровень обучаемости (18%). Опыт работы отметили в числе основных требований лишь 4% опрошенных работодателей.

По данным одной из крупнейших Российских компаний интернет-рекрутмента – HeadHunter, в 2022 году у работодателей основными требованиями к кандидату являются: умение работать в команде (32%), грамотная речь (22%) и навыки ПК (18%). Все навыки можно разделить на две группы: soft skills и hard skills. Мягкие навыки, также известные как "гибкие навыки", не связаны напрямую с профессиональной компетенцией, но они очень важны для выполнения работы и построения карьеры. Помимо навыков командной работы и грамотной речи, работодатели ценят деловое общение, телефонные звонки, грамотность, клиенториентированность, консультирование, а также способность работать с большим количеством информации и работу с людьми.

Hard skills – это профессиональные навыки и технические знания. Кроме знания ПК, работодатели часто обращают внимание на следующие навыки: знание английского языка; умение искать информацию в Интернете; обучение персонала; наличие водительских прав; знание PowerPoint; составление договоров; работа с оргтехникой [32].

В ходе исследований, проведенных исследовательской службой hh.ru, оказалось, что работодателям чаще всего не хватает развитого чувства ответственности, мотивации и исполнительности от кандидатов. Квалифицированным специалистам также часто не хватает таких качеств, как добросовестность, дисциплина и лидерские навыки. Для линейных менеджеров, то есть менеджеров среднего звена, важными качествами является способность работать с людьми, умение избегать конфликтов, предпринимательские и креативные навыки. При найме новых сотрудников работодатели обращают внимание на эти качества, а также на основную мотивацию к работе и способность вписаться в корпоративную культуру компании [32].

Что касается соискателей, по данным исследования Института социально-политических исследований отмечается, что значительная часть молодых специалистов не удовлетворены предоставляемые рынком труда условиями. Наиболее низкие оценки получили такие возможности как: служебное продвижение (в карьере), перспективы повышения заработной платы и занятие бизнесом. Возможность поиска работы оценивается молодыми специалистами, работающими в соответствии с полученной специальностью – 4,87 (из 7), работающими по близкой, схожей специальности – 4,30 и работающими по другой специальности – 4,08 [11].

Магистранты и выпускники, как потенциальные молодые специалисты, среди основных причин, которые затрудняют трудоустройство, называют высокий уровень конкуренции на рынке труда, отсутствие профессионального образования, а также отмечают отсутствие достойных предложений работы и то, что зачастую предложения по вакансиям не соответствуют ожиданиям [34].

Для более эффективного трудоустройства и поиска сотрудников важно отслеживать вызовы и тренды рынка труда и уметь на них реагировать. Мониторинг рынка труда – это систематический анализ и сбор информации о текущем состоянии рынка труда, его динамике и тенденциях. Он позволяет получать данные о количестве и качестве рабочих мест, уровне безработицы, заработной плате и условиях труда.

Мониторинг необходим для получения актуальной информации о рынке труда, что позволяет правительству, работодателям, работникам и другим заинтересованным сторонам принимать обоснованные решения о занятости, обучении и развитии человеческих ресурсов, а также о формировании социальной политики. Данный анализ позволяет: оценить

предложение и спрос на рабочую силу в различных секторах экономики и регионах страны; изучить динамику изменения уровня безработицы и занятости; определить среднюю заработную плату и уровень доходов населения; оценить качество рабочей силы и необходимость обучения и переподготовки; определить социально-экономические тенденции и предсказать их развитие. Наблюдение за рынком труда является важным инструментом для разработки и реализации государственной политики в области занятости и социальной защиты населения. Он осуществляется как на макроуровне (на уровне страны или региона), так и на микроуровне (на уровне отдельных предприятий или отраслей). Его можно проводить как с помощью официальной статистики, так и опросов работодателей и работников.

Мониторинг рынка труда позволяет выявить проблемы, с которыми сталкиваются работники и работодатели, и разработать соответствующие меры для их решения. Например, если мониторинг показывает высокий уровень безработицы, могут быть разработаны программы поддержки занятости и обучения, чтобы помочь людям найти работу. Если мониторинг показывает низкий уровень заработной платы, могут быть разработаны меры по увеличению заработной платы.

Таким образом, мониторинг является важным инструментом для планирования и управления рынком труда и создания условий для устойчивого экономического роста и социальной стабильности.

Что касается современной ситуации на Российском рынке труда, согласно данным Росстата, доля рабочей силы в возрасте 22 лет и старше, имеющей среднее профессиональное и высшее образование в конце 2022 года составляла более 80%. Численность безработных в России на 2022 год составляет 2951 тысяч человек и уровень безработицы равен 3,9%. По видам деятельности наибольший процент занятых в России находится в следующих отраслях: Торговля оптовая и розничная; ремонт автотранспортных средств и мотоциклов – около 20% и Обрабатывающие производства – 14%. Показатель среднемесячной начисленной заработной платы наёмных работников в организациях, у индивидуальных предпринимателей и физических лиц равен 50702 рублей [23].

Современная ситуация анализа рынка труда для работодателей и соискателей предполагает рассмотрение таких показателей как заработная плата, индекс востребованности профессии и индекс конкуренции среди соискателей, активность рынка труда, требования к профессионально важным качествам и компетенциям [12].

Одним из индикаторов рынка труда является индекс Фраучи, который показывает взаимосвязь между количеством рабочих мест и количеством резюме. Высокий индекс говорит о том, что на рынке труда больше предложений, чем соискателей, что говорит о

благоприятной ситуации для соискателей. Данный индекс отражает ситуацию как по стране, по регионам, так и по конкретным профессиональным ролям. Благодаря данному показателю, помимо актуальности профессии в регионах, можно также оценить насколько конкурентная среда в той или иной специальности. Среднее значение данного показателя по России – 4,02

Показателями активности рынка труда являются индекс активности работодателей и индекс активности соискателей. Индекс активности работодателей показывает количество вакансий в регионе по сравнению со средним показателем по стране. Это позволяет определить, есть ли в регионе избыток или дефицит вакансий. Если индекс превышает 100, это означает, что в регионе больше вакансий, чем в среднем по стране, и найти работу будет легче. Если индекс ниже 100, это означает, что в регионе меньше вакансий, чем в среднем по стране, и найти работу будет сложнее. Аналогичен и показатель соискателей, отражает количество соискателей в регионе относительно среднего значения по стране [12].

Еще одним из основных показателей рынка труда являются требования работодателей, то есть необходимые для той или иной профессии знания, навыки, а также личные качества. Как уже было отмечено навыки могут быть строго профессиональными и гибкими. Также к соискателю могут быть предъявлены требования в опыте работы в определенной сфере занятости.

Таким образом, существуют разные подходы к исследованиям рынка труда, наиболее современные обращают внимание на изучение рынка в непосредственном контакте социальных и экономических показателей. Для отслеживания трендов и вызовов применяется, мониторинг рынка труда – процесс анализа и сбора информации о рынке труда и его условиях. Это позволяет получить данные о количестве рабочих мест, уровнях безработицы, зарплате и других факторах, влияющих на рынок труда. С развитием цифровых технологий отслеживание трендов рынка труда выходит на новый уровень, рост технологий и информации о спросе и предложении рабочей силы позволяет принимать необходимые решения быстрее, это касается как работодателей, которые будут предлагать актуальные вакансии, соискателей, которые смогут подстраиваться под запросы компаний, так и в масштабе органов власти и образовательных учреждений.

1.2 Понятие и сущность больших данных

Переходя к цифровым технологиям важно начать с определения тех данных, с которыми приходится сталкиваться, изучая рынок труда. Большие данные (Big Data) – это термин, описывающий огромные объемы данных, которые нельзя обрабатывать и

анализировать с помощью традиционных методов и инструментов. Эти данные могут быть собраны из различных источников, таких как социальные сети, мобильные устройства, датчики IoT, банковские транзакции и т. д.

Для обработки больших данных используются специальные технологии и инструменты, такие как Hadoop, Spark, NoSQL базы данных и другие. Они позволяют обрабатывать и анализировать данные в режиме реального времени и извлекать ценную информацию из больших объемов данных.

Большие данные имеют огромный потенциал для бизнеса, науки и техники. Они могут помочь компаниям принимать более правильные решения, повышать эффективность и улучшать бизнес-процессы. В науке большие данные могут использоваться для изучения новых тенденций и открытий, а в технологиях – для создания новых продуктов и услуг. С развитием интернета и технологий количество информации постоянно растет. Так, на 2021 год размер рынка аналитики больших данных был оценен в 240,56 долларов США [38].

Согласно различным статьям, направленным на изучение больших данных, термин «Big Data» был введен в научный оборот в 2008 году редактором научного журнала «Nature» Клиффордом Линчем в специальном номере, который был посвящен росту объемов информации. Линч, используя этот термин, говорит, что данные по-разному большие, например, когда речь идет об их масштабах (международные проекты) или данных исследований, которые имеют постоянное значение — результат клинического испытания или наблюдение за уникальным событием. Поскольку цифровые данные так легко передаются, реплицируются и объединяются, они предоставляют огромные возможности для повторного использования, ускоряя уже проводимые исследования. [47]. Однако, в некоторых источниках говорится о более ранних употреблениях данного термина и популяризацию с 1990-х годов, широкое распространение связывают с именем Джона Р. Маши, доктора философии в компьютерных науках. [16].

В 2001 г. исследование Дуга Ланей (Doug Laney) для Meta Group определило три ключевых параметра больших данных: объем, скорость, разнообразие (так называемые три «V»: Volume, Velocity, Variety).

В Оксфордском английском словаре (The Oxford English dictionary, OED) представлено следующее определение больших данных: «Наборы информации, которые слишком велики или слишком сложны для обработки, анализа или использования стандартными методами» [52].

Перевод определения из Кембриджского словаря гласит, что Big Data - это «очень большие наборы данных, которые производятся людьми, использующими Интернет, и

которые могут быть сохранены, поняты и использованы только с помощью специальных инструментов и методов» [16].

В 2011 году Институт McKinsey опубликовал отчет о технологиях больших данных, в котором рассматривались три аспекта роста объема, скорости передачи данных и информационного разнообразия. В отчете определены пять способов создания ценности с использованием больших данных: во-первых, прозрачность и ускорение обработки информации; во-вторых, более точный сбор данных о производительности; в-третьих, сегментация и разработка рекламных кампаний; в-четвертых, улучшение процесса принятия решений; в-пятых, использование данных для создания новых продуктов.

Для понимания, что же такое большие данные, следует обратиться к их характеристикам. Во-первых, объем данных (Volume). Линч относит к Big Data любые массивы неоднородных данных, которые превышают объем обработки 150 Гб в сутки, однако в силу особенностей данных из разных источников, единого критерия относительно объема нет. Во-вторых, скорость сбора и обработки массивов данных (Velocity). Такие данные регулярно обновляются, поэтому необходимы такие интеллектуальные технологии, которые способны обновлять их в онлайн режиме. В-третьих, разнообразие типов данных (Variety). Большие данные могут быть структурированными, неструктурированными или полуструктурированными. Например, в социальных сетях поток данных не организован, там встречаются данные как в виде текстов, так и в виде фото или видео. В-четвертых, достоверность самого набора данных и результатов его анализа (Veracity). В-пятых, изменчивость (Variability). У потоков данных существуют пики и спады, которые меняются под влиянием сезонов и социальных явлений. Чем нестабильнее и изменчивее поток данных, тем сложнее он поддается анализу. В-шестых, значимость и ценность (Value). Как и любая информация большие данные могут быть простыми или сложными для анализа и восприятия. Посты в социальных сетях, например, являются простыми данными. Примером сложных можно считать банковские операции [6].

Таким образом, Big Data или большие данные – это структурированные или неструктурированные массивы данных большого объема, которые обрабатываются с помощью автоматизированных специализированных инструментов и используются для статистики, анализа, прогнозов и для принятия решения.

Анализ применения больших данных показывает, что их использование решает ряд основных задач, во-первых, хранение и управление данными, которые обычные реляционные базы данных не позволяют эффективно использовать. Во-вторых, Организация неструктурированной информации, состоящей из текстов, изображений, видео и других типов данных. В-третьих, анализ больших данных, который ставит вопрос о том, как

работать с неструктурированной информацией, генерировать аналитические отчеты, а также внедрять прогностические модели [2].

Проекты Big Data пересекаются с проектами бизнес-аналитики, мировой объем которого уже в 2012 году составлял около 100 млрд. долларов. Рынок бизнес-аналитики включает в себя элементы сетевых технологий, серверов, программного обеспечения, а также технических услуг. Технологии больших данных применяются для решения класса гарантирования доходов, которые предназначены для автоматизации деятельности компаний. Современные системы гарантирования доходов включают в себя инструменты для выявления несоответствий и углубленного анализа данных, которые позволяют своевременно выявлять возможные потери или искажения информации, которые могут привести к снижению финансовых результатов. На этом фоне российские компании, подтверждающие спрос на технологии Больших Данных на внутреннем рынке, отмечают, что факторами, стимулирующими развитие больших данных в России, являются рост данных, ускорение принятия управленческих решений и улучшение их качества [4].

Большие данные используются в различных областях, включая: Бизнес и экономика: анализ данных позволяет компаниям принимать более обоснованные решения, оптимизировать процессы и улучшать качество продуктов и услуг. Медицина: большие данные помогают в исследовании заболеваний, разработке новых лекарств и методов лечения, а также в улучшении качества медицинского обслуживания. Наука: большие данные используются в различных научных исследованиях, включая астрономию, физику, биологию и другие области. Государственное управление: анализ больших данных помогает правительствам принимать более эффективные решения в области экономики, здравоохранения, образования и других сферах. Транспорт: большие данные используются для оптимизации транспортных потоков, улучшения безопасности и комфорта пассажиров. Социальные сети: большие данные помогают социальным сетям анализировать поведение пользователей, предлагать персонализированный контент и улучшать качество сервиса. Финансы: анализ больших данных помогает банкам и финансовым компаниям принимать более обоснованные решения в области инвестирования, кредитования и управления рисками. Также технологии Big Data применяются в оценках поведения людей в социальных сетях, отслеживание GPS-сигналов, в банках, для отслеживания транзакций клиентов, данные о покупках.

Технологии больших данных включают в себя различные инструменты и методы сбора, хранения, обработки и анализа больших объемов данных. Некоторые из них: Hadoop - это платформа для распределенной обработки больших объемов данных. Он позволяет обрабатывать данные в кластерах сотен и тысяч узлов. NoSQL базы данных - это базы

данных, которые не используют традиционную реляционную модель. Они позволяют хранить и обрабатывать большие объемы неструктурированных данных. Apache Spark-это платформа для обработки больших объемов данных в памяти. Он позволяет обрабатывать данные в кластерах сотен и тысяч узлов. MapReduce-это модель программирования для распределенной обработки больших объемов данных. Он используется в Hadoop и других фреймворках. Машинное обучение – это методы анализа данных, которые позволяют компьютерам учиться на больших объемах данных, а также прогнозировать и рекомендовать. Интернет вещей (IoT) - это технология, которая позволяет собирать данные с различных устройств и датчиков и анализировать их для принятия решений. Бизнес-аналитика-это методы анализа данных, которые позволяют компаниям принимать решения на основе данных и оптимизировать свою деятельность.

Однако обработка больших данных также представляет собой ряд проблем, таких как сложность хранения и обработки данных, необходимость обеспечения безопасности и конфиденциальности данных, а также необходимость обучения и подготовки специалистов, которые могут работать с такими объемами данных. В исследованиях, направленных на изучение технологий Big Data, выделяется ряд проблем, связанных с данными. Первая группа проблем связана с классификацией данных. Большинство данных, которые можно отнести к Big Data, являются неструктурированными, поэтому каждый такой набор требует индивидуального подхода к его анализу. Такие данные имеют разный вид, источники и особенности заполнения, что отличает виды данных и, следовательно, анализировать их одним механизмом становится затруднительно.

Одной из основных проблем, связанных с обработкой больших объемов данных, является извлечение из них полезной информации. Анализ больших объемов данных может быть сопряжен с большой вычислительной сложностью, особенно при устранении несоответствий и неопределенностей в данных. Однако существуют способы снизить вычислительную сложность, особенно те, которые используют машинное обучение. Инструменты анализа должны быть достаточно мощными, чтобы справляться с несогласованностью данных и неопределенностями, а также вычислительной сложностью.

Еще одна проблема работы с большими данными связана с хранением и анализом данных, которая вызвана увеличением скорости создания новых данных, чему способствуют мобильные телефоны, интернет и его распространение. Данная проблема способствует увеличению значимости носителей информации, скорости записи и чтения для доступности больших данных.

Big Data – это огромный объем данных, которые невозможно обработать традиционными методами. В связи с этим возникают проблемы, связанные с хранением,

обработкой и анализом данных. Основные проблемы Big Data: Во-первых, большой объем данных требует дорогостоящих технологий для их хранения и обработки. Во-вторых, хранение данных должно обеспечивать их целостность, доступность и конфиденциальность. В-третьих, неструктурированный характер данных требует сложных алгоритмов для их анализа и структурирования. В-четвертых, низкая скорость обработки может привести к длительному времени ожидания ответа и устареванию данных в процессе обработки. Также среди проблем отмечается отсутствие эффективных алгоритмов учета большого объема данных и структуры может привести к ошибкам при анализе данных. Работу с такими данными осложняет наличие большого количества шумов и процесс их учета [22].

В последние годы термин «большие данные» стал настолько популярным, что вызвал негативную реакцию у некоторых экспертов. По данным Gartner, популярность технологических терминов обычно проходит несколько этапов: резкая волна, затем спад и разочарование, затем начинается настоящее массовое принятие. Сейчас "Большие данные" находятся на пике своей популярности, поэтому можно рассчитывать на массовое распространение этих технологий через 5-10 лет. Однако, как уже было отмечено, термин «большие данные» до сих пор не имеет четкого определения, и его часто называют новой методологией получения знаний. Некоторые считают, что термин «Big» не очень подходит для описания данных, поскольку они не могут быть «большими» в традиционном смысле, а бывают лишь «длинными» (long) и «широкими» (wide)». «Длина» данных определяется количеством наблюдений за конкретным фактором, а «ширина» определяется количеством функциональных зависимостей между различными факторами. Обе характеристики определяют, насколько "большими" данными располагает исследователь. Таким образом, "длина" и "ширина" данных вместе определяют, насколько хорошо исследователь может использовать эти данные для принятия решений [24].

Big Data - это подход, который позволяет получать новые знания из больших массивов данных, и уже сейчас он широко используется для прогнозирования в различных сферах деятельности. Взаимодействие между социологией и технологиями больших данных имеет огромное значение для получения новых знаний о социальных явлениях. Новый подход позволяет анализировать и обрабатывать огромные объемы данных, что дает более точные результаты и позволяет увидеть детали, недоступные при использовании выборочных данных. Кроме того, большие данные позволяют снизить требования к точности анализа, поскольку строгая точность может и не требоваться в условиях постоянно меняющихся данных. Новый подход также позволяет отказаться от поиска причинно-следственных связей, поскольку анализ больших данных позволяет проводить более всестороннее и многомерное изучение явлений. Все эти особенности нового подхода имеют

огромное значение для социологии, которая стремится к более глубокому пониманию социальных явлений и процессов [3].

Анализ больших данных часто критикуется за то, что не учитывает мнения старшего поколения, однако из-за своих размеров такой критерий неточности, перестает влиять на результаты. Использование Big Data при анализе субъективного мнения будет возрастать по мере уменьшения числа людей, не пользующихся Интернетом. В настоящее время, единственным ограничением для использования Big Data является выявление мнения старшего поколения, однако через одно-два поколения все население будет так или иначе использовать социальные медиа. Социальные медиа предоставляют богатый материал для анализа поведения людей, и теперь основным становится не столько умение строить выборку, сколько сбор и обработка объективных данных. Это уже не задача социолога, а программиста. Исследователям уже помогают различные программные продукты для анализа социологических данных.

2 Применение метода анализа больших данных в исследованиях рынка труда

2.1 Платформы по мониторингу рынка труда

В данной работе был проведен анализ возможностей, которыми обладает ряд зарубежных платформ, позволяющих проводить мониторинг рынка труда. Список платформ состоит из следующих площадок: Market Intelligence; Lightcast; WCC; Horsefly Analytics; TalentNeuron; Claro Analitics; CEDEFOP; International Labour Organization.

Первая платформа Market Intelligence. Целевой аудиторией данной платформы являются Владельцы бизнеса и HR-специалисты. Данный сервис позволяет проводить обзор глобального рынка труда, то есть аналитика может быть проведена сравнительно регионов. Ресурс предоставляет ценную информацию о тенденциях и изменениях в области найма и обучения персонала. Он включает сравнение с ориентированными на будущее компаниями, а также проверенные в будущем профили навыков, чтобы оптимизировать найм и развитие. Также сервис обладает автоматизированными панелями бизнес-аналитики, которые предоставляют ключевые показатели эффективности HR, которые можно использовать для принятия решений на основе актуальной информации. Еще одним аспектом аналитики являются тенденции, влияющие на бизнес и отрасль. В обзоре также может быть представлена информация о кадровых тенденциях конкурентов в отрасли, а также о том, какие компании ищут одни и те же навыки и какие навыки находятся в упадке. Кроме того, обзор может помочь определить, какие навыки востребованы на рынке, и определить, какие из них имеют отношение к конкретной отрасли занятости. Сервис также может предоставлять оценку динамики, то есть, могут быть выявлены специальности и навыки, которые будут становиться все более актуальными, а также оценивается влияние новых технологий на способы выполнения работы. Благодаря предоставляемой сервисом аналитики, владельцы бизнеса могут принимать решения относительно оптимизации развития, обучения и найма кандидатов, и решения будут соотноситься с требованиями будущего, а также сервис помогает в составлении описания вакансий, что сказывается на конкурентоспособности вакансии среди других [50].

Платформа Lightcast, предназначена для работодателей, государственных структур, учебных заведений и студентов. Данный сервис также предоставляет информацию о профессиональных ролях и навыках, а также анализ карьерных возможностей для соискателей, анализ требований к опыту и образованию от работодателей. Сервис предлагает несколько продуктов, один из них – Alumni Pathways, который использует реальные карьерные профили, чтобы показать учебным заведениям, запрос на какие навыки есть на

рынке труда и тем самым помогая школам адаптировать существующие программы и открывать новые возможности в обучении своих учеников. Таксономия навыков Lightcast ставит перед собой цель создания общего языка для сферы труда, помогая объединять работодателей, образовательные учреждения и отдельных лиц. Для студентов и соискателей на сервисе разработан Lightcast Skills, который помогает составлять более качественные резюме и узнавать о наиболее востребованных навыках. Колледжи и университеты также могут использовать Lightcast Skills, но уже для разработки и продвижения программ, связанных с работой, которые помогают студентам передавать свои навыки потенциальным работодателям. Работодателям данный сервис необходим для описания своих запросов к соискателям. Бесплатные инструменты Open Skills используют данные для оптимизации резюме и объявлений о вакансиях, помогая сопоставлять навыки с профессиями, а работников с вакансиями. Таким образом, данный сервис является инструментом, который дает обучающимся представление о рынке труда, помогает учебным заведениям соотносить программы обучения с запросами работодателей, способствует эффективному поиску соискателей и в целом формирует представление о тенденциях рынка труда [46].

WCC – это мировой поставщик решений по сопоставлению и анализу данных для различных государственных и частных организаций, таких как агентства по трудоустройству и рекрутинговые агентства, государственные службы, занимающиеся удостоверениями личности и охраной, а также предприятия. Программные решения платформы поиска и сопоставления обрабатывают ежедневно более 500 миллионов человек, позволяя организациям собирать, анализировать и использовать большие объемы разнородных данных из разных источников. Это дает возможность получать аналитическую информацию, необходимую для принятия обоснованных решений в различных областях. Платформа WCC способна обрабатывать данные любого типа – структурированные, неструктурированные, частные или общедоступные, и с любым форматом. У сервиса есть несколько модулей. Модуль EP Match позволяет государственным и кадровым агентствам сопоставлять данные соискателей и работодателей с помощью нескольких моделей сопоставления за одну операцию, учитывая требования обеих сторон. Администраторы могут легко и быстро настраивать модели в соответствии с их потребностями, пользователи могут применять фильтры и веса к различным характеристикам работы и соискателей, таким как заработная плата и местоположение. Это позволяет эффективно управлять результатами сопоставления. Одним из ключевых модулей является модуль профилирования, который предназначен для работы с разными категориями соискателей, например, социальные работники классифицируют ищущих работу в зависимости от их опыта работы и уровня безработицы. Сегментация помогает реализовать целенаправленную активную политику на рынке труда и

показывает подходящие вакансии, которые могут повысить шансы соискателя на успешное трудоустройство. Модуль регистрации пользовательских данных, таких как информация о соискателях, работодателях и вакансиях, имеет решающее значение для качественного поиска соответствий. Модуль EP Enrollment предоставляет два API, которые улучшают качество данных и расширяют возможности служб регистрации данных в приложениях для конечных пользователей. Модуль анализа пробелов EP Gap Analysis позволяет выявлять пробелы в своих навыках, компетенциях, сертификатах, опыте, зарплатных ожиданиях, образовании и местоположении – тех факторов, которые затрудняют поиск подходящих предложений. Модуль EP Referral предлагает индивидуальные планы обучения, курсы повышения квалификации, профессиональную подготовку для того, чтобы соискатель стал более востребованным на рынке труда. EP Analytics предоставляет аналитические данные и графики, благодаря которым специалисты могут отслеживать показатели рынка труда и их динамику и принимать соответствующие решения [41].

Следующая платформа – Horsefly Analytics, данные в реальном времени о кандидатах и вакансиях по миру. Данные собираются из нескольких источников, обрабатываются, очищаются, сервис поддерживает 39 языков, сравнивает гендерные различия, анализирует спрос и предложения для поиска скрытых человеческих ресурсов для должности или набора навыков. Для компаний это возможность прогнозировать бюджеты, сравнивать разнообразие с кадровым резервом, отслеживать последние тенденции найма, видеть ситуацию с набором кадров у конкурентов, планировать карьерные пути, улучшать стратегическое планирование и многое другое. Платформа позволяет сравнивать местоположения, чтобы найти лучший кадровый резерв. Ежедневно собирается миллионы социальных профилей для показателей предложений, которые можно фильтровать по опыту, полу, должности, ключевым словам и т. д. Целевая аудитория данного сервиса – владельцы бизнеса и HR-специалисты [44].

Аналитика данных TalentNeuron предоставляют необходимые инструменты для планирования работы, определения критически важных компетенций и разработки гибридной стратегии. Аналитический подход TalentNeuron включает анализ навыков, брендов, конкурентов, проверку данных по странам, производящим более 90% мировой экономики, стандартизацию глобальных данных и сравнение по местоположению, профессиям, компетенциям, организациям и должностям, а также 21 язык. Готовые аналитические отчеты для руководителей включают важные темы, такие как разнообразная рабочая сила, стратегические ресурсы и цифровые навыки, а также экспертный анализ ключевых вопросов, связанных с кадрами. Платформа самопомощи TalentNeuron предоставляет инструменты для сегментации данных по местоположению, компетенциям,

конкурентам и другим параметрам, чтобы обеспечить более точные результаты для HR-специалистов [51].

Claro Analytics — это платформа, которая анализирует рынок труда, она помогает определить тенденции найма и доступность талантов, а также сравнительный анализ между мужчинами и женщинами. В платформе используются данные из словаря данных, содержащие более 5 млрд. комбинаций имен и других параметров, таких как пол, раса и национальность. Платформа также предлагает API-доступ к данным. Claro Analytics предоставляет следующие функции: конкурентный бенчмаркинг; прогнозирование рисков для сотрудников; выявление возможной дискриминации; сравнение рисков увольнения с конкурентами; оценка баланса на рынке труда с учетом местоположения и навыков; разработка сценариев для удаленной работы и работы из дома. Целевая аудитория программы – бизнес- и HR-менеджеры [39].

Cedefop является одним из децентрализованных агентств Евросоюза. Cedefop поддерживает продвижение, разработку и реализацию политики Союза в области профессионального образования и обучения, а также политики в области навыков и квалификаций, работая вместе с Комиссией, государствами-членами и социальными партнерами. Анализ мегатрендов экономики и общества, таких как увеличение числа пожилых работников и переход на цифровые и экологически чистые технологии. Это создает новые потребности в навыках, но также приводит к несоответствию между ними. Эти мегатренды влияют на будущее рабочих мест в ЕС. Предметом исследований и обсуждений с политиками, специалистами, учебными заведениями, службами занятости и другими заинтересованными сторонами являются навыки и их возможности на рынке труда. Исследования включают в себя различные методы, такие как обследования предприятий, рабочей силы, прогнозы, тематические исследования и анализ вакансий. Цель исследований – выявить общие европейские и международные тенденции в требованиях к навыкам. Для этого используются целостные и инновационные подходы, которые учитывают временной разрыв между изменениями спроса на квалифицированную рабочую силу и мерами, принимаемыми в политике и практике. Кроме того, проводятся прогнозы спроса и предложения, анализ навыков в различных секторах и опрос работодателей [48].

International Labour Organization (Международная организация труда) – трехстороннее агентство Организации Объединенных Наций, которое объединяет правительства, работодателей и работников из 187 стран-членов с 1919 г. МОТ устанавливает трудовые нормы, разрабатывает политику и программы, чтобы обеспечить достойную работу для всех людей. В сферу деятельности организации входят статьи, посвященные анализу рынка труда

и навыков, которые размещены в открытом доступе и позволяют всем желающим принимать необходимые решения на основе анализа рынка труда [49].

Таким образом, платформы по анализу рынка труда необходимы для того, чтобы помочь работодателям и соискателям найти друг друга и создать взаимовыгодное сотрудничество. Они позволяют работодателям узнать о навыках и опыте соискателей, а соискателям – получить информацию о вакансиях и требованиях работодателей. Кроме того, платформы могут предоставлять возможность для обмена опытом и знаниями между участниками рынка труда, что способствует развитию профессиональных навыков и повышению конкурентоспособности. Платформы анализа рынка труда также могут быть полезны в более масштабном понимании, они могут использоваться для мониторинга уровня безработицы, определения потребностей в рабочей силе и оценки эффективности мер по поддержке занятости. Кроме того, анализ данных, собранных на платформах, может помочь правительству принимать более обоснованные решения в сфере экономики и социальной политики.

2.2 Специфика и проблематика применения анализа больших данных в исследованиях рынка труда

В рамках данной работы был проведен опрос экспертов, для выявления возможностей аналитики рынка труда с помощью больших данных. В качестве экспертов были отобраны сотрудники лаборатории прикладного анализа больших данных Томского государственного университета, которые в своих проектах, применяя методы Big Data анализируют рынок труда, а также было проведено интервью со специалистом из Высшей школы экономики.

Одним из блоков вопросов был про специфику метода больших данных. В своих определениях Big Data эксперты говорят о критериях, которые описывают большие данные, а именно объем данных, широкое разнообразие форм данных, скорость. Это те данные, которые окружают людей ежедневно. То есть, большие данные – это огромные объемы структурированных и неструктурированных данных, которые накапливаются в различных источниках, огромные объемы информации, которые создаются и обрабатываются в современном мире. Они могут состоять из разных типов данных, обработка которых требует специальных инструментов и подходов для эффективного анализа, что было отмечено экспертами:

- «...кто-то когда-то сказал мне, что большие данные – это то, что не помещается в Excel». (Ж, аналитик, ТГУ);

- «Второй момент - это *velocity*, скорость. Скорость их обработок очень важна. Ну, например, ты подходишь с банковской картой, тебе надо, чтобы ты сразу оплатил, а не ждал определенного дня, пока придет транзакция. Надо очень быстро обрабатывать эти данные». (М, заведующий лабораторией, ТГУ);

- «Большие данные они очень вариативные, в том смысле, что они бывают они текстовые, бывают они картинки, бывают цифровые следы, в общем, всякие разные форматы у них». (Ж, аналитик, ТГУ);

«...у нас сейчас информация в обществе растет по экспонентам. То есть каждый день количество информации удваивается изо дня в день, ну то есть каждый день становится больше в два раза, чем в предыдущий день. То есть это очень большое количество информации, это информация обо всем буквально». (Ж, аналитик, ТГУ).

Среди преимуществ метода анализа больших данных эксперты отмечают возможность отслеживать актуальную информацию, видеть тенденции развития, отслеживать изменения; возможность собирать данные для анализа из разных источников; скорость обработки информации, но важно отметить, что для быстрой обработки информации данные должны быть подготовлены к анализу. Еще одним преимуществом такого метода является более объективная оценка объекта исследования в сравнении с выборочными методами.

- «Преимущество в том, что это постоянно можно отслеживать необходимую информацию и видеть определенные тенденции <...> можно выгружать данные, смотреть как та или иная проблема была представлена там год назад или как проблема, или что-то, закономерность, как они меняются с течением времени». (Ж, старший научный сотрудник, Институт образования ВШЭ);

- «это возможность, во-первых, изучения какого-либо объекта в динамике». (Ж, аналитик, ТГУ);

- «Большие данные – это все-таки возможность проанализировать весь объект целиком, если мы анализируем пользователей социальных сетей, вот, бери социальную сеть и не выбирай оттуда 15 человек или 100 человек, или сколько у тебя там ресурсов, а бери их всех и анализируй, это, наверное, такое важное преимущество». (Ж, аналитик, ТГУ).

Но помимо преимуществ, данный метод анализа данных обладает недостатками. Среди них можно выделить неструктурированность данных, поскольку большие данные являются той информацией, которая находится везде, по всему интернет-пространству, она может иметь разные вид и форму, это и фото, и видео, это тексты, это данные о местоположении и т.д. и при обработке такой информации исследователи сталкиваются с тем, что данные нужно подготовить, привести в один вид, пригодный для последующего анализа. Еще одним недостатком эксперты называют необходимость дорогих

вычислительных ресурсов, тут опять же стоит обратить внимание на объем больших данных, которые для того, чтобы с ними работать, нужно где-то хранить и для этого нужны мощные ресурсы. Кроме ресурсов дорогими являются кадры – специалисты, которые умеют работать с такой информацией и такими вычислительными ресурсами. Также данные нужно собрать, данные могут находиться в разных источниках, нужно найти такие источники, сопоставить их данные между собой и правильно проанализировать. Отсюда вытекает еще одна сложность, которая заключается в интерпретации полученных результатов. Еще один момент, на который обращают внимания эксперты – это недостатки этического характера, этот недостаток проявляется в исследованиях, которые связаны со сбором информации о людях, с одной стороны информация находится в открытом доступе, но с другой ее можно использовать и против человека.

- *«Это дорого, не все научные коллективы могут себе позволить, или не научные, организации, бизнес, не все могут позволить себе ресурсы купить, для того чтобы большие данные хранить <...> нужны хорошие вычислительные ресурсы, чтобы их обрабатывать, плюс — это специалисты дорогостоящие, то есть, именно вот эксперты, кто может с этими большими данными работать...»*. (Ж, аналитик, ТГУ);

- *«это и большие серверные нагрузки, и большие, там, нагрузки, связанные с различными хранилищами данных, различными серверами...»* (Ж, аналитик, ТГУ);

- *«Конечно, необходимо вдумчивое использование информации, правильно интерпретировать – это тоже своего рода искусство, своего рода сложность»*. (Ж, старший научный сотрудник, Институт образования ВШЭ).

Область применения больших данных достаточно велика, что отражается в проектах, с которыми работают эксперты. Это проекты в области образования; проекты, связанные с изучением качества жизни; анализ дискурсов социально-значимых проблем; изучение такого феномена как научное волонтерство или гражданская наука; изучение коренных малочисленных народов Севера и Дальнего Востока; выявление мотивов для переезда в село

- *«...мы использовали данные социальной сети для набора абитуриентов»*. (Вячеслав, заведующий лабораторией, ТГУ);

- *«я стала заниматься аналитикой по вакцинации, там ковид как раз грянул, у нас была значит тема, связанная с анализом дискурсов о вакцинации, анализом радикалистов, антивакеров, которые пишут разные аргументы против вакцинации, провакеров, которые наоборот говорят, что давайте все будем вакцинироваться, искали лидеров, искали площадки, ну, и в принципе, продолжаем делать это по сей день»*. (Ж, аналитик, ТГУ);

- *«...мы занимались организационной культурой Томского государственного университета. Мы проводили опрос среди всех сотрудников Томского государственного*

университета и хотели построить организационную сеть. То есть кто с кем взаимодействует, по каким каналам, кто как с кем. Вообще дружит, не дружит, работает, не работает». (Ж, аналитик, ТГУ);

«...анализ мотивов и ценностей людей для переезда в село из города <...> мы брали данные соц. сетей и смотрели, а что собственно люди, которые переезжают в село, что для них важно, что они сами об этом пишут, какие у них дискурсы, какие проблемы они видят в селе, какие проблемы они видят в городе и давали рекомендации как строить политику, чтобы люди переезжали в село». (Ж, аналитик, ТГУ).

Работа с большими данными предполагает наличие различных знаний и навыков, но конкретные умения отличаются от этапа работы с данными. Есть 5 основных этапов: целеполагание, первичное исследование, сбор данных, первичная обработка данных, потом исследование данных и описание. На каждом этапе необходимы свои компетенции, обработка данных подразумевает работу с базами данных с помощью различных языков программирования, таких как Python или SQL, нужно понимать принципы алгоритмизации. Поскольку анализ больших данных чаще подразумевает количественную методологию, в обработке и анализе данных нужны знания математической статистики, понимание как работать с цифрами. Также эксперты выделяют такие профессиональные навыки для работы с большими данными как постановка и проверка гипотез. Также нужно уметь работать с программными обеспечениями, которые позволяют анализировать данные, например, Polyanalyst, Orange, Gephi. Также эксперты говорят о необходимости визуализировать полученные результаты и уметь их презентовать. Среди гибких навыков называются аналитическое мышление, нестандартное мышление, коммуникабельность, внимательность, а также, поскольку большие данные подразумевают несколько этапов, на каждом из этапов могут работать специалисты именно в этом этапе и поэтому, среди soft-skills отмечается умение работать в команде. Кроме перечисленных компетенций отмечается необходимость знаний в той отрасли, которая анализируется.

- «Ну, во-первых, мне кажется, я бы отметила, наверное, это скрупулезность, если говорить, внимательность, да, если из таких навыков, непрофессиональных навыков...» (Ж, старший научный сотрудник, Институт образования ВШЭ);

- «...изначально, и всецело, и сам вот этот термин дата аналитика, дата сейтисты, дата анализ, он предполагает работу с базами данных, предполагает работу с базами данных при помощи Python и запросов в SQL». (Ж, аналитик, ТГУ);

- «Конечно, во-первых, хотелось бы, чтобы ты понимал, что вообще это такое, сам концепт больших данных, что они из себя представляют, как их можно использовать, ну, такие какие-то базовые вещи, навыки программирования минимальные хотя бы, чтобы

имелось представление о базах данных тех же самых, например, или какие-то языки программирования на таком самом элементарном уровне, чтобы уметь запросы писать к этим базам данных». (Ж, аналитик, ТГУ);

- «Аналитическое мышление, коммуникабельность, умение работать в команде, еще умение нестандартно мыслить, не только аналитическим». (Ж, младший научный сотрудник, ТГУ).

Как было отмечено, работа с большими подразумевает несколько этапов и с эти эксперты связывают необходимость и важность команды профессионалов. Команда по работе с большими данными необходима для эффективного управления данными, их анализа, составления выводов и принятия решений на основе полученных результатов. Без команды невозможно собирать и обрабатывать большие объемы данных и интерпретировать результаты анализа.

- «должна быть, во-первых, для этого набрана хорошая, сильная команда, которая умеет с ними просто работать, потому что мне в одиночку, даже с поддержкой моего коллеги, работать с большими данными тяжело, потому что не могут с таким объемом данных справиться, как говорится, два человека». (Ж, старший научный сотрудник, Институт образования ВШЭ).

Еще одним блоком вопросов был мониторинг рынка труда и необходимость его аналитики. По мнению экспертов, анализировать рынок труда и проводить его мониторинги необходимо для ряда целей, от помощи соискателям находить работу и работодателям находить кадры, до того, что с помощью данной аналитики могут выстраиваться региональные и государственные политики развития. Мониторинг рынка труда позволяет соискателям отслеживать современные тренды на профессии и учитывать актуальность тех или иных навыков для того, чтобы быть востребованным специалистом на рынке труда. Работодатели с помощью результатов таких исследований могут подстраивать сферу своей занятости под актуальные тренды, например, учитывать размер заработной платы для тех или иных сотрудников, составлять актуальные требования к соискателям и в целом учитывать конкурентоспособность своего бизнеса. Органы власти с помощью данных исследований могут понимать какие сферы в каком направлении движутся, куда направлять необходимые ресурсы и в какую сторону выстраивать стратегии развития. Еще одним направлением, для которого нужен анализ рынка труда, является сфера образования, а именно программа образовательных курсов. С помощью аналитики рынка труда руководители образовательных программ могут выстраивать свой курс так, чтобы те, кто на этом курсе обучается, стали востребованными специалистами на рынке труда, имели необходимые и актуальные навыки и компетенции.

- «...это позволяет, в принципе, выявлять навыки и компетенции, необходимые работодателям, то есть то, что они заявляют в режиме реального времени для сотрудников разных должностей». (Ж, старший научный сотрудник, Институт образования ВШЭ);

- «Рынок труда - это, наверное, первое место, по которому видно изменения в экономике, в принципе, в стране. Какие должности умирают, какие зарплаты предлагают, в каких регионах что происходит, можно увидеть именно на примере рынка труда». (Ж, младший научный сотрудник, ТГУ);

- «Это во многом позволяет и властным структурам, правительству планировать стратегии развития, стратегии страны, регионов отдельных или отдельных отраслей деятельности, в зависимости от состояния этих регионов или отраслей <...> стратегии управления персоналом для бизнеса это важно и нужно понимать куда вообще движется их отрасль, рынок, смотреть кто им понадобится в их компании для того, чтобы быть успешными в ближайшем будущем». (Ж, аналитик, ТГУ);

- «Это позволит менять систему образований и более гибко ее настраивать на потребности вот этого рынка труда, потому что мы знаем, что рынок труда и образование, они оторваны друг от друга...» (Ж, старший научный сотрудник, Институт образования ВШЭ);

- «...для человека, мне кажется, тоже очень важно понимать, кем быть, чтобы быть, чтобы было более высокое качество жизни». (Вячеслав, заведующий лабораторией, ТГУ).

Для того, чтобы анализировать рынок труда есть несколько подходов, один из вариантов — это провести интервью или фокус-группы с работодателями, таким способом можно узнать какие запросы есть у работодателей определенной сферы, куда они направляют свой бизнес и какая ситуация в их сфере по их мнению. Еще один вариант изучения рынка труда — это опять же интервью и фокус-группы, но уже с экспертами, такими как сотрудники центров занятости, HR-специалисты, органы власти в сфере экономики. И третий вариант — это проанализировать рынок труда с помощью больших данных. Преимуществами такого метода в аналитике рынка труда является то, что при опросах работодателей присутствует доля субъективности, то есть каждому работодателю нужно своё, при исследовании с помощью больших данных такой проблемы нет, поскольку анализируются запросы сразу множества работодателей. Еще одним преимуществом является возможность отследить динамику, каких специалистов требуется больше или меньше в тот или иной период, какие навыки актуальны, а какие уже нет. Также анализ больших данных позволяет видеть реальную ситуацию на рынке труда здесь и сейчас.

Большие данные позволяют собирать информацию о рынке труда из разных источников, что позволяет говорить о более объективных результатах. Эксперты отмечают скорость анализа информации, когда заранее подготовлены необходимые операции по обработке данных, их можно анализировать сразу и получать необходимую информацию в реальном времени. Еще один плюс больших данных это то, что они позволяют рассмотреть ситуацию на рынке труда целиком, то есть по всей стране или по всей отрасли. Потому эксперты сходятся во мнении, что изучении рынка труда предполагает совмещение разных методов исследований.

- *«Преимущество в том, что это постоянно можно отслеживать необходимую информацию и видеть определенные тенденции, закономерности на различных, так скажем, этапах наблюдения и в то же время проводить такие сравнения».* (Ж, старший научный сотрудник, Институт образования ВШЭ);

- *«...когда мы спрашиваем работодателей, экспертов, у них достаточно зашоренный взгляд. Но это не только потому, что им нужен кто-то свой, а потому что они, скорее всего, и чаще всего, эксперты только в своей какой-то области, и не эксперты в остальных. Вот. А нам надо очень сильно избегать субъективизма...»* (Ж, младший научный сотрудник, ТГУ);

- *«...с их помощью можно получить доступ к самым разным уголкам нашей большой страны и посмотреть максимально насколько это возможно картину по региону или по отрасли. То есть собрать все-все вакансии в одну общую базу и агрегировать их, проанализировать вместе».* (Ж, аналитик, ТГУ);

- *«Независимость от того, как специалисты видят государство, независимость от того, как специалистов видит конкретный работодатель. То есть это позволяет построить, наверное, более общую картину о том, как этих специалистов в целом видит рынок».* (Полина, аналитик, ТГУ).

Но помимо перечисленных возможностей анализ рынка труда с помощью больших данных имеет ряд ограничений, это опять же неструктурированность данных, дорогие вычислительные ресурсы и специалисты, необходимость команды специалистов. интерпретация и визуализация полученных результатов, то есть все те недостатки, которые есть в работе с большими данными вне зависимости от отрасли применения. Одним из отраслевых недостатков применения больших данных эксперты отмечают недостаток экспертности, то есть требования, которые предъявляются к соискателям в вакансиях не всегда грамотно составлены, что также сказывается на таком ограничении как сложность обработки из-за отсутствия так называемой культуры заполнения вакансий, каждый работодатель пишет требования так, как это видится ему, что требует качественной обработки информации для ее анализа. Также есть сложность, связанная с тем, что вакансии

могут заполняться HR-специалистами, которые не разбираются в конкретной отрасли и поэтому заполняют вакансии, сами не понимая, что требуется от соискателя. Вакансии на сайтах-агрегаторах могут дублироваться, что искажает ситуацию на рынке труда, это требует более детальной обработки баз данных, вычищение дублей. Еще одним ограничением является то, что в некоторых отраслях может быть такая ситуация, когда работодатели ищут сотрудников через свои локальные каналы, которые не отражены в массовых источниках по поиску сотрудников.

- *«Но я, как аналитик, я не сварщик, я не знаю 15 это разных вариантов сварки или один и тот же, и мы, безусловно, мы собирали фокус-группы с экспертами, мы создавали и разрабатывали классификаторы навыков и компетенций вместе с экспертами, для того, чтобы понять, что из написанного вообще адекватно, а что неадекватно, но, все эти данные невозможно обработать всецело, полностью руками, все эти данные не могут обработать эксперты».* (Ж, аналитик, НИЛПАБД ТГУ);

- *«На мой взгляд, большое ограничение – это культура заполнения вакансий <...> мы вот с центрами занятости работаем, с департаментом труда занятости томским нашим, они нам об этом говорили еще год назад, что на самом деле работодателя сложно заставить написать корректно вакансию, они копируют друг у друга, они не отдают себе отчет часто, что они там пишут, они дублируют вакансии, ну в общем, большое количество текста получается нечестного, скажем так, который приходится дополнительно чистить...»* (Ж, аналитик, ТГУ);

- *«...вакансии заполняют HR, которые вообще не понимают ничего в отрасли, как правило, либо понимают прямо минимум».* (Ж, младший научный сотрудник, ТГУ)

Таким образом, специфика применения больших данных заключается в том, что они могут быть использованы для анализа различных аспектов рынка труда, таких как спрос на профессии, тенденции на рынке труда, эффективность найма и адаптации персонала, а также для разработки новых продуктов и услуг. Проблематика применения больших данных при анализе рынка труда заключается в том, что данные могут быть сложными для обработки, могут содержать ошибки и неточности, а также требовать значительных ресурсов для обработки.

2.3 Перспективы развития анализа рынка труда с помощью методов Big Data

Еще одним блоком исследование было выявление перспектив развития анализа рынка труда с применением больших данных, одним из перспективных направлений является разработка российских платформ по мониторингу рынка труда. Как уже было отмечено,

существуют различные платформы, позволяющие анализировать рынок труда, но такие платформы существуют только за рубежом, и они не позволяют анализировать российский рынок труда. На данную проблему обратили внимание аналитики прикладного анализа больших данных ТГУ и разработали такой сервис. Сервис «РосНавык» позволяет анализировать российский рынок труда, в нем есть интерактивный дашборд, который позволяет анализировать навыки и компетенции, которые запрашивают работодатели. В дашборде можно выбирать регион, специальность и отслеживать как устроен запрос рынка труда, какие есть предложения, от каких работодателей, какие зарплатные предложения, какие условия труда, какие требуются навыки. Также можно отдельно выбирать навык и смотреть, где он требуется, и каковы зарплатные предложения в вакансиях, где нужен этот навык. Также на сервисе есть отражение динамики открытия и закрытия вакансий. (Рисунок 1) Целевой аудиторией сервиса на данный момент являются по большей части разработчики образовательных программ, но информация будет полезна как работодателям, так и соискателям, в дальнейшем платформа будет оснащена профориентационным тестом и ссылками на образовательные курсы для соискателей. Для разработчиков образовательных программ команда аналитиков сервиса предлагает услуги по анализу рынка труда, для составления более подробных отчетов о рынке занятости определенных отраслей или регионов.

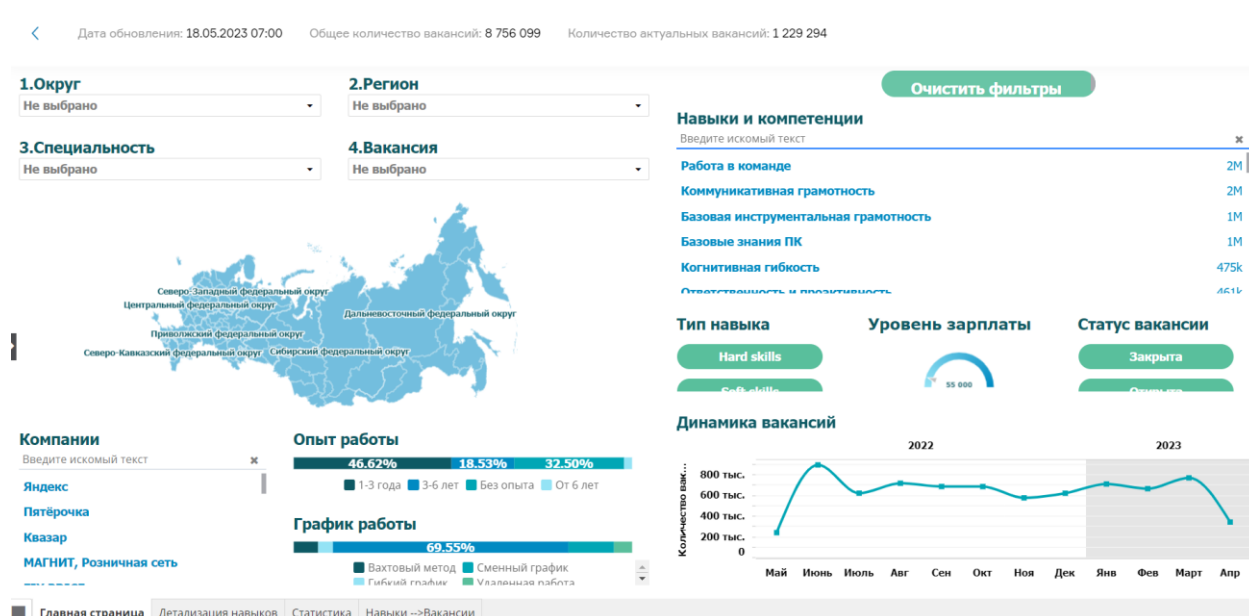


Рисунок 1 – Интерактивный дашборд сервиса «РосНавык»

- «На сегодняшний день в Российской Федерации у нас нет конкурентов. Это связано с тем, что специфика нашего сервиса заключается в анализе требуемых на рынке труда навыков и компетенций». (Ж, аналитик, ТГУ);

- «...сейчас у нас сервис сделан только для одной целевой аудитории, нашей основной – руководители образовательных продуктов. Хотим расширить, сделать еще для

соискателей, кадровых центров в ближайшее время, то есть для соискателей внедрить туда профориентационный тест на выявление тех или иных навыков компетенций, и если эти навыки компетенции тест выявляют, то есть они присутствуют у человека, можно будет перейти на дополнительную страницу и посмотреть, в каких компаниях могут люди пройти стажировку или подработке <...> Либо можно будет перейти на сторонний сервис по обучению». (Ж, аналитик, ТГУ).

Развитие изучения рынка труда с помощью больших данных эксперты связывают с развитием таких сервисов для аналитики рынка труда как «РосНавык», а также с тем, что качество заполнения вакансий будет становится лучше, что работодатели начнут грамотнее их составлять. Отмечается, что необходимо оснащать такие сервисы данными со всех доступных источников, сайтов-агрегаторов, социальных сетей и т.д, а также появление на таких сервисах профориентационных возможностей. Еще одной тенденцией является развитие алгоритмов, с помощью которых такая аналитика вообще возможна и инструментов для анализа данных, в которых для получения результата уже не так важны знания языков программирования.

- «Мне кажется, чтобы мы могли говорить о каких-либо перспективах, нам нужно просто получается создавать такие же системы, которые есть в Европе, Америке, когда просто не вручную естественно, а при помощи больших данных, при помощи различных методов IT-технологий, все данные по вакансиям они агрегируются в одном месте, и человек будучи работодателем или соискателем рабочего места, он может видеть вот эти компетенции в режиме реального времени, применимые в запрашиваемой, даже в каждой из должностей». (Ж, старший научный сотрудник, Институт образования ВШЭ);

- «...тенденция, которая давно появлялась, сейчас на самом деле я очень рада, это упрощение аналитики данных и переход к новому типу инструментов». (Ж, младший научный сотрудник ТГУ);

- «Перспективы развития, хотелось бы, конечно, чтобы вакансии переходили в хорошее качество». (Ж, аналитик, ТГУ);

- «...на основании такой аналитики предлагаться способы его дальнейшего трудоустройства <...> но, смотрите, вы с такой зарплатой, с таким графиком работы, вы можете трудоустроиться сюда. Но, смотрите, одновременно с этим, если вы к своим навыкам добавите еще такой навык, например, обучитесь ему где-нибудь, и тут же, например, появляется какой-нибудь список со всеми ресурсами, где человек может этим навыкам научиться, например, то вы сможете работать уже там-вот там-вот там, на таких-то условиях и так далее». (Ж, аналитик, ТГУ).

Таким образом, аналитика рынка труда с помощью больших данных обладает рядом возможностей и преимуществами такими как объективный анализ ситуаций на рынке труда, анализ спроса и предложений рабочей силы, возможность отследить актуальные тенденции в запросах определенных специалистов и навыков. Также преимуществами являются возможность видеть реальную ситуацию и отслеживать динамику. Возможность собирать информацию из разных источников. Но также есть и сложности, с которыми сталкиваются исследователи, это отсутствие единого стандарта к заполнению вакансий, что усложняет процесс обработки данных, также такие данные требуют дорогостоящие ресурсы для их хранения и обработки. Осложняют аналитику рынка труда и то, что некоторых специалистов есть тенденция искать по локальным каналам и поэтому их невозможно учесть.

2.3 Пример анализа рынка труда с помощью больших данных

В рамках данной работы было проведено практическое применение платформы РосНавык, которая предназначена для мониторинга рынка труда, разработанной экспертами лаборатории прикладного анализа больших данных Томского государственного университета.

Был проведен анализ вакансий Республики Башкортостан, опубликованных на сайте hh.ru с 1 января по 27 мая 2023 года, всего в регионе было опубликовано 88147 вакансий. В топ 10 самых востребованных специальностей региона попали: Менеджер по продажам, менеджер по работе с клиентами – 6543 вакансии; Продавец-консультант, продавец-кассир – 5766; Водитель – 4120; Машинист – 2889; Слесарь, сантехник – 2251; Инженер-конструктор, инженер-проектировщик – 1887; Повар, пекарь, кондитер – 1859; Сварщик – 1855; Бухгалтер – 1840; Разнорабочий – 1789.

Анализ зарплатных предложений показал, что медианное предложение равно 52500 рублей. Самыми высокооплачиваемыми из наиболее востребованных профессий являются: Машинист – 130.000 рублей; Сварщик – 110.555; Водитель – 84.000. (Рисунок 2)



Рисунок 2 – Зарплата по самым востребованным профессиям региона

Также выявлялись компании-работодатели, которые выражают запрос на наиболее востребованные специальности региона. Наибольшее число вакансий от таких компаний как: Пятёрочка; МАГНИТ, Розничная сеть; Тинькофф; Консалтинг Групп; Газстройпром; Газпром нефть; Холва; Стройтранснефтегаз; Старт, торговая компания; РН-Транспорт.

Основным направлением платформы является возможность выявления востребованных навыков, анализ показал, что в Республике Башкортостан среди самых запрашиваемых специальностей востребованными профессиональными навыками являются Базовые знания ПК, Поиск и привлечение новых клиентов, Ведение бухгалтерского учета, Управление грузовым автомобилем. (Рисунок 3)



Рисунок 3 – Наиболее востребованные профессиональные навыки (в %)

Помимо профессиональных навыков к сотрудникам предъявляют требования о наличии дополнительных компетенций, которые считаются гибкими, мягкими или надпрофессиональными. В востребованных профессиях республики в топ таких навыков попали Работа в команде, Коммуникативная грамотность, Базовая инструментальная грамотность. (Рисунок 4)



Рисунок 4 – Наиболее востребованные надпрофессиональные навыки (в %)

Также было выявлено сколько готовы платить за те или иные навыки. Таким образом, наиболее высокооплачиваемыми вакансиями являются те, в которых требуются такие профессиональные навыки как Управление грузовым автомобилем, Знание технических

характеристик, конструкции и устройства грузовых автомобилей и спецтехники, Управление легковыми автомобилями и небольшими грузовиками. (Рисунок 5)



Рисунок 5 – Заработные предложения по профессиональным навыкам

Вакансиями, где предлагается наиболее высокая медианная заработная плата, являются те, в которых в качестве надпрофессиональных навыков требуются Целеустремленность и работоспособность, Коммуникативная грамотность, Когнитивная гибкость, Внимательность. (Рисунок 6)



Рисунок 6 – Заработные предложения по надпрофессиональным навыкам

Таким образом, пример данной аналитики отражает возможности сервисов по мониторингу рынка труда и применение анализа больших данных. С помощью аналитики рынка труда можно выявлять актуальные профессиональные и надпрофессиональные навыки, специальности на которые есть спрос, проводить анализ зарплатных предложений и оценивать специфику отраслей и регионов.

ЗАКЛЮЧЕНИЕ

Большие данные (Big Data) – это термин, описывающий огромное количество структурированных и неструктурированных данных, которые накапливаются в результате деятельности человека или организации. Эти данные могут быть собраны из различных источников, включая социальные сети, интернет-трафик, медицинские записи, финансовые транзакции и многое другое. Большие данные обладают огромным потенциалом для анализа и извлечения ценной информации, которая может быть использована в различных областях, таких как бизнес, здравоохранение, наука и т.д.

Одной из областей применения больших данных являются исследования рынка труда, с помощью больших данных. Анализ рынка труда с использованием больших данных – это анализ больших объемов данных о рынке труда, таких как вакансии, резюме, информация о работодателях и соискателях работы и т.д. Этот анализ позволяет нам выявить тенденции и закономерности на рынке труда, выявить наиболее популярные профессии, выявить изменения в требованиях к кандидатам и т.д. Для выполнения такого анализа используются различные методы и инструменты, такие как машинное обучение и статистика. Анализ рынка труда с использованием больших данных имеет много преимуществ. Во-первых, это позволяет нам получать более точные и достоверные данные о рынке труда, чем традиционные методы исследования. Во-вторых, это позволяет быстро реагировать на изменения рынка, выявлять новые тенденции и возможности для развития бизнеса или отрасли. Наконец, анализ рынка труда с использованием больших данных помогает оптимизировать процесс найма и адаптации сотрудников, что способствует повышению эффективности.

Ограничения анализа больших данных зависят от многих факторов, таких как объем данных, скорость обработки, доступность инструментов и ресурсов, а также квалификация специалистов. Ограничения скорости обработки. Обработка больших объемов данных требует значительных вычислительных ресурсов. Это может привести к задержке получения результатов анализа и необходимости оптимизации процессов обработки данных. Ограничения на доступность ресурсов. Анализ больших данных требует специальных инструментов и программного обеспечения. Анализ больших данных требует высококвалифицированных специалистов в области статистики, программирования и машинного обучения, не все специалисты обладают необходимыми знаниями и навыками. Ограничения, связанные с объемом данных. Большие объемы данных могут быть слишком сложными для анализа или содержать слишком много мусора, что требует более тщательной обработки и делает результаты анализа менее точными.

Для проведения такой аналитики существуют различные платформы, на данный момент наиболее распространены такие сервисы за рубежом, однако команда аналитиков лаборатории прикладного анализа больших данных ТГУ разработали такой сервис. Он позволяет анализировать востребованные профессиональные и надпрофессиональные навыки и компетенции на рынке труда, выявлять какие специалисты востребованы в том или ином регионе, оценивать наиболее высокооплачиваемые профессии и наиболее ценные навыки.

Таким образом, большие данные играют важную роль в анализе рынка труда. Они позволяют собирать и обрабатывать огромный объем данных о вакансиях, резюме, демографических данных и других факторах, влияющих на спрос и предложение на рынке труда. Анализ этих данных позволяет выявлять тенденции и закономерности, определять наиболее востребованные профессии и изменения в требованиях работодателей к кандидатам. Например, с помощью анализа больших данных можно определить, какие профессии являются наиболее высокооплачиваемыми и перспективными, какие навыки наиболее востребованы на рынке труда и какие изменения происходят в требованиях работодателей. Также большие данные могут использоваться для прогнозирования изменений на рынке труда и принятия решений о найме и обучении сотрудников. Но также данный метод обладает рядом ограничений, которые требуют более тщательной обработки информации и совмещение такого метода с другими.

ЛИТЕРАТУРА

1. Богданов М.Б. Возможности и ограничения цифровых следов и методов машинного обучения в социологии / М.Б. Богданов, И.Б. Смирнов // Мониторинг общественного мнения: экономические и социальные перемены. – 2021. – № 1. – С. 304—328.
2. Большие Данные — новая теория и практика | Открытые системы. СУБД | Издательство «Открытые системы». - №10. - 2011. - URL: <https://www.osp.ru/os/2011/10/13010990?ysclid=lfuva6n2p7575183289> (Дата обращения: 23.03.2023).
3. Васянин М. С. Взаимодействие социологии и больших данных // V социологическая Грушинская конференция «БОЛЬШАЯ СОЦИОЛОГИЯ: расширение пространства данных» 12–13 марта 2015 г., МОСКВА – С. 18-21.
4. Волков В. В. Проблемы и перспективы исследований на основе Big Data (на примере социологии права) / В.В. Волков, Д.А. Скугаревский., К.Д. Титаев // Социологические исследования. – 2016. – № 1. – С. 48-58.
5. Волков С.И. Экономический механизм достижения равновесия интересов агентов российского рынка труда: автореферат дис. ... кандидата экономических наук: 08.00.01 // Тамбовский гос. ун-т им. Г. Р. Державина. – Тамбов. – 2000. – 20 с.
6. Где применяются технологии Big Data уже сегодня? - 2020. - URL: <https://vc.ru/s/productstar/130207-gde-primenyayutsya-tehnologii-big-data-uzhe-segodnya> (Дата обращения 16.02.2023).
7. Губа К.С. Большие данные в исследовании науки: новое исследовательское поле. // Социологические исследования. – 2021. – № 6. – С. 24-33.
8. Губа К.С. Большие данные в социологии: новые данные, новая социология? // Социологическое обозрение. – 2018. – №1. – С. 231-236.
9. Гударенко Ю. А. Рынок труда как социальный институт в современном российском обществе // Известия РГПУ им. А. И. Герцена. – 2010. – №124. – С. 330-336.
10. Ермолаева С.Г. Рынок труда: учебное пособие. — Екатеринбург : Изд-во Урал. ун-та. – 2015.— 108 с.
11. Зубок Ю. А. Молодые специалисты: проблема подготовки и положение на рынке труда / Ю. А. Зубок, В.И. Чупров // Социологические исследования. – 2015. – № 5. – С. 114-122.
12. Индикаторы и показатели рынка труда для работодателей и соискателей. - 2022. - URL: <https://crcg.ru/labor-market-indicators/> (Дата обращения: 10.05.2023).

13. Как в России используют технологии Big Data? - 2022. - URL: <https://issek.hse.ru/news/776383019.html> (Дата обращения: 19.03.2023).
14. Качала В.В. Подход к анализу рынка труда на базе технологий big data / В.В. Качала, Т.И. Зайнуллина // Перспективы развития современных математических и естественных наук. Сборник научных трудов по итогам международной научно-практической конференции. – Воронеж. – 2014. – С. 62-63.
15. Количество открытых вакансий на hh.ru за последний месяц снизилось на 20% — Карьера на vc.ru - 2022. - URL: <https://vc.ru/hr/403247-kolichestvo-otkrytyh-vakansiy-na-hh-ru-za-posledniy-mesyac-snizilos-na-20?ysclid=lf9ez9dyz5238885175> (Дата обращения: 19.03.2023).
16. Корнев М. С. История понятия "большие данные" (Big Data): словари, научная и деловая периодика // Вестник РГГУ. Серия: Литературоведение. Языкознание. Культурология. – 2018. – №1 (34). – С. 81-85.
17. Кравченко В.О. "Большие данные" - практические аспекты и особенности / Кравченко В.О., Крюкова А. А. // Academy. – 2016. – №6 (9). – С.65-67.
18. Крупеникова Л.Ш. Big Data и новые задачи социологии // Гуманитарий Юга России. – 2022. – №2. – С. 50-57.
19. Майер-Шенбергер В. Большие данные. Революция, которая изменит то, как мы живем, работаем и мыслим. / В. Майер-Шенбергер, М. Кукуер // М. : Манн, Иванов и Фербер – 2014. – 240 с.
20. Мальцева А. В. Big Data как источник социологической информации: пример анализа блога губернатора Петербурга / А. В. Мальцева, М. С. Матвеев, М. Б. Моисеева // Siberian Socium. – 2019. – Том 3. – № 3. – С. 74-84.
21. Мальцева А. В. Проблемы репрезентативности при работе с «большими данными» // Социальные практики и управление: проблемное поле социологии: материалы Сибирского социологического форума с международным участием / отв. ред. С. В. Ровбель, С. А. Ильиных. Новосибирск: НГУЭУ. – 2017. – С. 141-145.
22. Менщиков А. А. Основные проблемы использования больших данных в современных информационных системах / А. А. Менщиков, В.Э. Перфильев, М.Ю. Федосенко, И.Р. Фабзиев // Столыпинский вестник. – 2022. – №1. – С. 316-329.
23. Меццанцаника М. Большие данные в контексте аналитики рынка труда / М. Меццанцаника, Ф. Меркорио // Европейский фонд образования. – 2019. – 67 с.
24. Норт Д. Институты, институциональные изменения и функционирование экономики. // Фонд экономической книги “Начала”. – 1997. – 117 с.

25. Одинцов А. В. Социология общественного мнения и вызов Big Data // Мониторинг. – 2017. – №3 (139). – С. 30-43.
26. Платонова С.И. Большие данные: создание вызовов и возможностей в социальных науках // Манускрипт. – 2020. – №4. – С.119-123.
27. Радаев В. В. Понятие капитала, формы капиталов и их конвертация // Экономическая социология. – 2002. – №4. – С. 20-32.
28. Радаев В. В. Экономическая социология: учебное пособие для вузов // В. В. Радаев; Гос. Ун-т – Высшая школа экономики. – М.: Изд. Дом ГУ ВШЭ. – 2005. – 603 с.
29. Риккардо Д. Начало политической экономии и налогового обложения // Антология экономической классики. М. – 1993. – Т. 1. – С. 449-450.
30. Сведберг Р. Новая экономическая социология: что сделано и что впереди? // Экономическая социология. – 2004. – №4. – С.37-55.
31. Топ востребованных навыков молодого специалиста. - 2023. - URL: <https://tomsk.hh.ru/article/31231> (Дата обращения: 2.04.2023).
32. Трудовые ресурсы, занятость и безработица // Росстат. - URL: https://rosstat.gov.ru/labour_force (Дата обращения: 10.05.2023).
33. Хусяинов Т. М. Социологический анализ социального взаимодействия на рынке Интернет-труда // Политика и общество. – 2016. – № 8. – С. 1145-1151.
34. Цыганкова И. В. Трудовые ресурсы молодежи на рынке труда России и Франции / И.В. Цыганкова, Ю.В. Мячин // Экономика труда. – 2019. – №2. – С 735-746.
35. Шилина Н. В. Что такое Big Data? Основные проблемы: хранение и управление данными // Молодой ученый. — 2018. — № 28 (214). — С. 1-3.
36. Шульц Т. Инвестиции в человеческий капитал // Американское экономическое обозрение. – 1961. – №51. – С. 1-17.
37. Big Data Analytics | IBM URL: <https://www.ibm.com/analytics/big-data-analytics> (Дата обращения: 23.03.2023).
38. Big Data Analytics Market Size, Share | Growth Statistics [2030]. - 2023. - URL: <https://www.fortunebusinessinsights.com/big-data-analytics-market-106179> (Дата обращения: 10.03.2023).
39. De Mauro A. What is big data? A consensual definition and a review of key research topics / A. De Mauro, M. Greco, M. Grimaldi // Conference: 4th International Conference on Integrated Information. Madrid, 2014. – 11 с.
40. Granovetter, Mark. Economic Action and Social Structure: The Problem of Embeddedness // American Journal of Sociology. – 1985. – Vol. 91. – P. 481-510.

41. Hilbert M. The world's technological capacity to store, communicate, and compute information / M. Hilbert, P. Lopez // Science. – 2011. – Global. – 261 p.
42. Kitchin R. What makes Big Data, Big Data? Exploring the ontological characteristics of 26 datasets / R. Kitchin, G. McArdle // Big Data & Society. – 2016. – P. 1-10.
43. Lynch C. Big data: how do your data grow? // Nature. – 2008. – Vol. 455. – № 7209. – P. 28-29.
44. Tweet geekery and epic crowdsourcing: an Oxford English Dictionary update. URL: <http://blog.oxforddictionaries.com/2013/06/oed-june-2013-update/> (дата обращения: 05.02.2023).
45. smartData – People analytics platform | HRForecast. – URL: <https://hrforecast.com/market-intelligence/> (Дата обращения: 24.02.2023).
46. Lightcast Dataю – URL: <https://lightcast.io/about/data> (Дата обращения: 24.02.2023).
47. Employment Platform Analytics | WCC Group. – URL: <https://www.wcc-group.com/employment/products/employment-platform-analytics/> (Дата обращения: 24.02.2023).
48. Horsefly Analytics. – URL: <https://horseflyanalytics.com> (Дата обращения: 25.02.2023).
49. TalentNeuron™ | Global Labor Market Analytics for HR Leaders. – URL: <https://www.gartner.com/en/human-resources/research/talentneuron> (Дата обращения: 25.02.2023).
50. Claro Analytics. – URL: <https://www.claroanalytics.com> (Дата обращения: 25.02.2023).
51. Skills and labour market | CEDEFOP. – URL: <https://www.cedefop.europa.eu/en/themes/skills-labour-market> (Дата обращения: 26.02.2023).
52. Skills strategies for future labour markets. – URL: <https://www.ilo.org/skills/areas/skills-training-for-poverty-reduction/lang--en/index.htm> (Дата обращения: 26.02.2023).

ПРИЛОЖЕНИЕ А

Программа исследования

Актуальность: Широкое распространение интернета и развитие цифровых технологий приводят к изменениям в социальных науках, появляются новые источники данных и развиваются методы для их сбора, обработки и анализа. Особенностью таких данных является то, что они созданы не для целей исследования, охватывают всю популяцию и производятся в режиме реального времени.

По данным оценки экспертов Fortune Business Insights объем мирового рынка аналитики больших данных в 2021 году оценивался в 240,56 млрд долларов и по прогнозам вырастет до 655,53 млрд долларов к 2029 году, при этом среднегодовой темп роста составит 13,4% в течение прогнозируемого периода. Анализ показал, что в среднем, рост в 2020 году составлял 12,5%, в сравнении с 2019 годом [39].

В России в 2021 г. технологии сбора, обработки и анализа больших данных применяли 25,8% организаций — на 3,4 п. п. больше, чем в 2020 г. Самым популярным источником оказались веб-сайты компаний: их данные собирают 9,2% компаний. Немного менее востребованы данные учетных систем организации (ERP, CRM и др.) и социальных сетей (8 и 7,2% соответственно). Помимо этого, анализируются данные операторов сотовой связи (6,7%), информация с цифровых датчиков и радиочастотных меток (6,3%) [13].

Что касается рынка труда, наиболее доступным источником информации о нем является Интернет-пространство. На специализированных сайтах по поиску работы, в объявлениях о вакансиях содержатся требования к соискателю, необходимые навыки и перечень обязанностей. На данный момент в сети существует множество сайтов по трудоустройству и объёмы информации продолжают расти, поскольку ежедневно публикуются новые резюме и вакансии. Так, по оценкам hh.ru Россия, на конец сентября 2022 года количество открытых вакансий на сайте – 275393 [15].

Выделяют несколько подходов к изучению рынка труда, Классическая теория, авторами являются А. Смит и Д. Рикардо, доказывает, что рыночная система может поддерживать полную занятость благодаря стабильности спроса и предложения на рынке труда, что является основой для благосостояния нации. В неоклассической теории (А. Лаффер, Д. Гилд, Р. Холл) предполагается, что рынок труда является внутренне неоднородной и динамичной системой, где цена труда быстро реагирует на изменения спроса и предложения [30].

Таким образом, технологии Big Data (Большие данные) является одним из перспективных направлений в аналитике рынка труда, поскольку они предполагают работу с информацией огромного объема и разнообразного состава, которая постоянно обновляется и может находиться в различных источниках.

Проблемная ситуация: С одной стороны, метод анализа больших данных обладает такими преимуществами как скорость сбора информации, высокая достоверность и возможность выявления закономерностей, а с другой стороны, данный метод имеет ряд недостатков, например, таких как неструктурированная информация и сложность ее интерпретации. Преимущества и недостатки данного метода сказываются на возможностях и ограничениях применения методов Big Data в анализе рынка труда

Проблема: Каковы особенности применения анализа больших данных в исследованиях рынка труда? В чем заключаются возможности и ограничения больших данных исследованиях рынка труда?

Объект: Мнения экспертов в оценке применения больших данных в исследованиях рынка труда

Предмет: Возможности и ограничения анализа больших данных при исследовании рынка труда

Цель: Выявить возможности и ограничения применения анализа больших данных в исследованиях рынка труда

Задачи:

- Проанализировать платформы по мониторингу рынка труда и выявить показатели рынка труда, доступные для анализа
- Описать специфику и проблематику применения метода анализа больших данных в исследованиях рынка труда
- Выявить перспективы развития анализа рынка труда с помощью метода Big Data

Метод сбора информации – опрос экспертов, полуструктурированное интервью, метод отбора единиц наблюдения – снежный ком

Обоснование метода: Исследование предполагает получение специализированных знаний об объекте исследования. Данный метод позволит получить компетентные, профессиональные представления и оценки использования метода больших данных при исследованиях рынка труда.

ПРИЛОЖЕНИЕ Б

Гайд экспертного интервью

Вводный блок

1. Какая у Вас должность? Какими проектами Вы занимаетесь? Как давно вы работаете в данной сфере?

Что такое большие данные

2. Что такое Big Data? Какие данные относятся к большим? В чем особенности таких данных? Как с ними работать?

3. В чем заключается специфика метода анализа больших данных?

4. Какие бывают типы анализа больших данных?

5. Какими знаниями и навыками необходимо обладать, чтобы работать с большими данными?

6. В чем преимущества метода анализа больших данных?

7. Какие есть недостатки у данного метода?

Аналитика рынка труда

8. Зачем нужен мониторинг рынка труда?

9. Как меняется рынок труда и как эти изменения влияют на его анализ?

Специфика анализа больших данных в аналитике рынка труда

10. Расскажите о проектах, которые связаны с аналитикой рынка труда? Какие цели и задачи поставлены в проектах, направленных на исследования рынка труда?

11. Какими возможностями обладает метод анализа больших данных в аналитике рынка труда

12. Какие есть ограничения метода анализа больших данных в аналитике рынка труда?

13. В чем преимущества анализа больших данных в сравнении с другими методами в исследованиях рынка труда?

14. Какие недостатки метода больших данных в сравнении с другими методами в исследованиях рынка труда?

15. Какие вы видите перспективы развития больших данных? Как эти развития повлияют на анализ рынка труда?

ПРИЛОЖЕНИЕ В

Список экспертов

1. Ширкова Наталия, старший научный сотрудник института образования, лаборатория непрерывного образования взрослых, ВШЭ
2. Гойко Вячеслав, заведующий научно-исследовательской лабораторией прикладного анализа больших данных
3. Дунаева Дарья, аналитик, научно-исследовательская лаборатория прикладного анализа больших данных ТГУ, менеджер продукта «РосНавык»
4. Мундриевская Юлия, младший научный сотрудник, научно-исследовательская лаборатория прикладного анализа больших данных ТГУ, руководитель образовательных программ академии Data Diving
5. Плешкевич Ирина, аналитик, научно-исследовательская лаборатория прикладного анализа больших данных
6. Круглик Полина, аналитик, научно-исследовательская лаборатория прикладного анализа больших данных

ПРИЛОЖЕНИЕ Г

Расшифровки интервью

Интервью №1, Наталья, старший научный сотрудник института образования ВШЭ
Длительность: 20 мин.

Вопрос: Какая у вас должность? Кем работаете? Какими проектами занимаетесь?

Ответ: Я работаю в институте образования в Высшие школы экономики с 2018 года, в должности старшего научного сотрудника. Я кандидат педагогических наук. Проекты нашей лаборатории... я занята в лаборатории непрерывного образования взрослых, в которой основные наши проекты связаны с реализацией и запуском, стимулированием, так скажем, проектов в сфере дополнительного профессионального образования взрослого населения. Наши ключевые сейчас проекты - это проекты стратегического академического лидерства, Приоритет 2030 в Высшей школе экономики и, собственно, у нас есть, также мы являлись мониторами федерального проекта в 19-20 году, "Новые возможности для каждого" и, собственно, вот с Томским гос. университетом тоже сотрудничаем по линии содействия занятости и есть еще проект в рамках создания научного центра мирового уровня, который у нас называется как "Образование взрослых". Это именно исследовательские проекты. Если говорить более расширенно, то мы занимаемся также и реализацией собственной программы дополнительного проф. образования, по запуску системы ДПО в вузах. То есть, проводим такой консалтинг, поддержку вузов, региональных, ну, в том числе и московских, где мы сотрудничаем с преподавателями и с инициативным персоналом, который занимается запуском проектов ДПО.

В: Вы получаете работаете и с большими данными, правильно?

О: С большими данными, да, мы тоже работаем, то есть, в части вот именно аналитики рынка труда, в части поиска необходимых компетенций для разных должностей, для разных видов экономической деятельности, если можно так сказать, ну, прям подробно мы так глубоко в рынок труда стараемся не особо залезать, потому что это достаточно трудоемко, но чтобы понимать какие компетенции так или иначе нужны, да, и чтобы видеть эту палитру, так скажем, навыков, да и каких-то трендов мы тоже да, занимаемся.

В: А что вообще такое большие данные? Какие данные относятся к большим? В чем особенность таких данных?

О: Ну, если говорить непосредственно про нашу деятельность, то с чем, как бы, я сталкиваюсь в своей работе, это непосредственно аналитика баз данных о работе и и вот именно в части анализа вакансий, размещенных на базе HeadHunter и на базе trudvsem, я использую вот эти данные. Ну, то есть получается, что для меня большие данные - это неструктурированный, так скажем, хаотичный набор переменных, так скажем, да, и показателей, которые просто нужно постоянно обрабатывать, постоянно нужно за ними следить, потому что появляются новые тенденции, появляются новые вакансии, если говорить в части вакансий, новые вот эти все переменные, которые могут быть ну лично для меня, так скажем, полезными. Для меня большие данные - это хаотичный такой набор переменных.

В: А какие знания и навыки нужны, чтобы уметь работать с большими данными?

О: Ну, во-первых, мне кажется, я бы отметила, наверное, это скрупулезность, если говорить, внимательность, да, если из таких навыков, непрофессиональных навыков, вот, а в части... там нужно постоянно их отслеживать, ну, вот я выгружаю, говорю, что не могу приступить к работе с ними до тех пор пока я просто их не обработаю, не приведу в удобный для себя формат, вид. У меня условно на аналитику этих данных уходит гораздо меньше, чем разбиение их и приведение к нормальному виду. Если говорить про профессиональные навыки, то в целом, это нужно знать хорошо языки программирования, различные статистические пакеты и здесь, к сожалению, моих навыков, компетенций не всегда хватает и, поэтому у меня есть человек, который в команде работает по ГПХ и он работает с этими данными, как говорится, работать. То есть, я вижу какие то там тенденции, закономерности, то я просто адресую и говорю "Сергей, нам нужно проверить вот эту вот закономерность, пожалуйста, помоги", как говорится, вот, и Сергей через SPSS помогает с этим работать дальше.

В: В чем преимущество такого метода?

О: Использования больших данных? Преимущество в том, что это постоянно можно отслеживать необходимую информацию и видеть определенные тенденции, закономерности на различных, так скажем, этапах наблюдения и в то же время проводить такие сравнения, лонгитюдные даже, наверное, это можно выгружать данные, смотреть как та или иная проблема была представлена там год назад или как проблема или что-то, закономерность, как они меняются с течением времени. Плюс, конечно, это и сбор данных помогает, то что из разных источников, потому что даже база данных о работе, одно дело анализировать только HeadHunter, другое дело прибегать даже к анализу, не знаю, Работа в России другие дают вакансии, не знаю там, Вконтакте третьи вакансии, в телеграмм-каналах четвертые, и то есть, можно видеть разные такие группы, скажем, интересующих вопросов. Вот. Ну, плюс, конечно, это даст скорость обработки данных, если данные вычищенные, если они удобные для работы, то скорость данных, скорость обработки информации тоже имеет большой такой плюс. Плюс выявление, конечно, закономерностей. Единственное, я говорю, что здесь данные получаются неструктурированные, хранятся вразнобой, в разных таблицах, изображениях, с этим надо уметь работать, правильно интерпретировать в дальнейшем.

В: А помимо неструктурированности, какие бы вы еще недостатки отметили?

О: Это получается, тоже, наверное, что должна быть, во-первых, для этого набрана хорошая, сильная команда, которая умеет с ними просто работать, потому что мне в одиночку, даже с поддержкой моего коллеги, работать с большими данными тяжело, потому что не могут с таким объемом данных справиться, как говорится, два человека. Поэтому тут нужно, конечно, ну, да, хорошая сильная должна быть команда. Плюс... какие еще тут сложности... Конечно, необходимо вдумчивое использование информации, правильно интерпретировать - это тоже своего рода искусство, своего рода сложность. Одно дело найти закономерность на данных, но другое дело уже понять, почему это происходит так или иначе и, когда вот хаотичный набор данных, он не проливает, так скажем, на это свет. Поэтому вот. Плюс, конечно, я понимаю, что это может быть какая-то утечка информации, которая неправильно хранилась, еще что-то, в целом, поскольку у нас это базы данных о работе, то мы с этим, в принципе, не сталкиваемся. Кибербезопасность. Для меня, наверное, вот эти проблемы, больше организационные даже.

В: Если переходить к аналитике рынка труда, зачем вообще нужно анализировать рынок, проводить мониторинги? Что это дает?

О: Получается, что аналитика рынка труда - это такой своеобразный, уже новый, по крайней мере для России, новый тренд, потому что в той же там Европе и Америке этим уже занимаются достаточно долгое время, в России это выходит на такой новый уровень. Анализировать большие данные для рынка труда - это, конечно, позволяет видеть тренды, которые существуют, потому что мир меняется, мы за ним не всегда, как говорится, можем поспевать и нам нужно просто видеть, что, так скажем, существует на рынке труда с точки зрения исследований, то есть, для меня это такой механизм получается исследований. Если говорить, конечно, еще, то, в целом, это мы можем увидеть там опять какие вакансии новые открываются, и мы увидим какие новые там неочевидные зависимости и закономерности. Также это позволяет, в принципе, выявлять навыки и компетенции, необходимые работодателям, то есть то, что они заявляют в режиме реального времени для сотрудников разных должностей, получается, в разных отраслях экономической деятельности. В дальнейшем такой подробный анализ, если его проводить, это позволит менять систему образований и более гибко ее настраивать на потребности вот этого рынка труда, потому что мы знаем, что рынок труда и образование, они оторваны друг от друга и то, чему учат в университетах на бакалавриате, в магистратуре это не всегда, условно, соответствует запросам реальных работодателей, вот, поэтому есть такой профиль компетенций для должностей, для видов деятельности, это очень и очень важно. Для самого же соискателя рабочего места, наличие такой вот палитры навыков, это позволит составлять ему свой собственный профиль компетенций и видеть какие компетенции ему нужно развивать, для того чтобы быть успешным в этой отрасли, на данном даже месте работы, то есть, конкретно, такое своего рода кадровое планирование более грамотное именно с точки зрения компетенций. Плюс, конечно, можем увидеть насколько, так скажем, если смотреть с региональной повесткой, насколько запрос работодателей на специалистов того или иного профиля отвечает потребностям региона, то есть, если я вижу, что какой-то регион, условно, там связан с аграрной промышленностью, то логично было бы, что туда нужны специалисты понятные как бы, а когда я открываю вакансии, я смотрю эти базы данных о работе, я вижу, что, ну, условно, в Кабардино-Балкарии, там совершенно нет вакансий в сфере такого сельского хозяйства, хотя это основной их вид деятельности, который обеспечивает вклад в развитие этой территории, то есть, это именно аграрной промышленности называют. И это тоже своего рода вызов опять таки уже для государства

даже, что нужно стимулировать и создавать новые рабочие места, открывать новые предприятия и вкладывать деньги, инвестиции именно вот в эту отрасль, потому что она является основной, условно, для данного субъекта. Это интересный для меня кейс анализа, скажем так.

В: Какие вы видите перспективы развития больших данных и как это развитие повлияет на анализ рынка труда?

О: Мне кажется, чтобы мы могли говорить о каких-либо перспективах, нам нужно просто получается создавать такие же системы, которые есть в Европе, Америке, когда просто не вручную естественно, а при помощи больших данных, при помощи различных методов IT-технологий, все данные по вакансиям они агрегируются в одном месте, и человек будучи работодателем или соискателем рабочего места, он может видеть вот эти компетенции в режиме реального времени, применимые в запрашиваемой, даже в каждой из должностей. Для меня это очень интересный бы был кейс и анализ, чтобы я просто видела, что в графическом варианте я могу выбрать условную профессию для себя или должность и увидеть какие компетенции мне нужно развивать, что запрашивает работодатель в текущий момент. И это то, что бы мне хотелось видеть в России в том числе, чтобы это подтягивало автоматически из разных ресурсов, с разных баз данных о работе, из социальных сетей, откуда угодно. И это чтобы было постоянно наполняемым таким, возобновляемым ресурсом, чтобы информация не устаревала, чтобы были видны новые тренды, тогда и образование под это можно подтянуть и включать новые модули, системы дополнительного профессионального образования, по крайней мере проектировать новые что ли программы, которые будут отвечать действительно запросам рынка труда, а не просто вот реализовать программу, чтобы она была, плюс один к карме чтобы зачлось, нет, чтобы было именно полезно для связи рынка труда и такого образования и государственной даже политики. Вот такие тенденции. И большие данные тут помогут, потому что вручную мы это не можем все сделать, понятное дело, даже при всем желании, вручную мы будем только одну тысячу вакансий этих обрабатывать сколько, неделю, как говорится, а все это меняется стремительно. Поэтому тут традиционные методы, даже контент-анализа, они абсолютно не помогут, а биг дата вот как раз да, позволит получить такие скрытые, своего рода, знания об исследуемой области.

Интервью №2, Дарья, аналитик лаборатории прикладного анализа больших данных ТГУ, менеджер проекта «РосНавык»

Длительность: 47 мин.

Вопрос: Начнем с того какую должность занимаешь, какими проектами занимаешься?

Ответ: Так, в научно-исследовательской лаборатории прикладного анализа больших данных Томского государственного университета на текущий момент я называюсь, если меня забить в Персоне, то я буду лаборант, но по факту моя должность называется аналитик, ее всячески разно меняют в зависимости от того, где нужно меня презентовать, иногда я бываю менеджер проекта, иногда я бываю менеджер продукта, иногда я бываю аналитик, иногда я бываю ведущий аналитик, ну, в общем, вариаций много, но суть одна я руковожу некоторыми проектами, организую работу людей и делаю какую-то часть аналитическую сама.

В: А если говорить про проекты, то в каких они отраслях?

О: Проекты... С момента как я сюда трудоустроилась, три года что ли назад, три или четыре, я сначала вела направление, у меня было основное, изучение цифрового качества жизни, ну то есть какое качество жизни в разных регионах России по данным социальных медиа, это был мой большой такой проект, гигантский, он был грантовый, на несколько лет. Потом он закончился, и я стала заниматься аналитикой по вакцинации, там ковид как раз грянул, у нас была значит тема, связанная с анализом дискурсов о вакцинации, анализом радикалистов, антивакеров, которые пишут разные аргументы против вакцинации, провакеров, которые наоборот говорят, что давайте все будем вакцинироваться, искали лидеров, искали площадки, ну, и в принципе, продолжаем делать это по сей день, это направление, как и качество жизни, тянется долго со мной, такое постоянное. Были ситуативные разные проекты, такие типа по заказу или просто маленькие изначально, это типа гражданская наука, научное волонтерство, это коренные малочисленные народы севера Сибири и Дальнего востока, мы там атлас их составляли, цифровые следы, я там сети строила и, например, анализ мотивов и ценностей людей для переезда в село из города, это был заказ МинСельхоза, то есть прям большой и там тоже была наша маленькая часть, где мы брали данные соц. сетей и смотрели, а что собственно люди, которые переезжают в село, что для них важно, что они сами об этом пишут,

какие у них дискурсы, какие проблемы они видят в селе, какие проблемы они видят в городе и давали рекомендации как строить политику, чтобы люди переезжали в село. Ну, и бывают вообще маленькие коммерческие проекты типа анализ экспертов Казахстана в связи с кризисом в январе в прошлом году. То есть, какое-то событие произошло социально значимое, есть какой-то заказчик, которому это интересно, он к нам приходит и говорит: “Здравствуйте, нам бы вот на больших данных что-нибудь тут поизучать” и мы что-нибудь изучаем.

В: Так, хорошо. А что такое большие данные, как бы ты их описала?

О: Большие данные - это разная информация, которую нельзя изучать обычными методами, скажем так, то есть, у больших данных есть несколько характеристик, если так, то, во-первых, это их объем, какие-то там огромные терабайты и так далее данных, которые надо где-то хранить, как-то обрабатывать и как-то дальше анализировать. Большие данные они очень вариативные, в том смысле, что они бывают они текстовые, бывают они картинки, бывают цифровые следы, в общем, всякие разные форматы у них. Скорость, то есть, большие данные собираются сами собой, в большинстве случаев, допустим в социальной сети, люди заходят и по доброй воле чего-то постят, лайкают, комментируют и так далее и это такой снежный ком, то есть, в каждую секунду времени большие данные самонакапливаются, скажем так, самонакручиваются, если кто-то там чего-то запостил, то потом через день у этого будет еще в геометрической прогрессии цифровых следов и так далее по нарастающей. Поэтом вот объем, это скорость появления этих данных, вариативность их, ну, и, может быть, добровольность что ли, если можно это выделить в отдельную характеристику. То есть больше данные, вообще данные их обычно кто-то собирает зачем-то, а большие данные они существуют как отдельная сущность, живут сами себе, даже если они никому не нужны, они есть.

В: Что нужно знать и уметь, чтобы работать с такими данными?

О: По опыту, смотря, наверное, с какими и зачем, а в принципе, если ты закончил социологию, то можешь работать с большими данными в части аналитики, но если ты хочешь полный цикл, так сказать, работы освоить, то есть, и сбор, и хранение, и предобработку, и аналитику, и рекомендации, то на каждом из этих этапов у тебя должны быть определенные компетенции. Конечно, во-первых, хотелось бы, чтобы ты понимал, что вообще это такое, сам концепт больших данных, что они из себя представляют, как их можно использовать, ну, такие какие-то базовые вещи, навыки программирования минимальные хотя бы, чтобы имелось представление о базах данных тех же самых, например, или какие-то языки программирования на таком самом элементарном уровне, чтобы уметь запросы писать к этим базам данных. Аналитические методы, допустим, тоже зависит от того, какие большие данные ты хочешь анализировать, хочешь, например, анализировать тексты, значит, пожалуйста, Natural Language учи опросинг и все у тебя будет хорошо, то есть, будешь уметь анализировать именно текстовую информацию или, допустим, хочешь картинки - учи компьютерное зрение как применяется. То есть, вариантов много, для каждого этапа есть свои компетенции в аналитике больших данных и для каждого вида этих больших данных тоже они есть. Если в целом, то какие-то базовые навыки программирования, аналитики и общее представление о том, что это, откуда оно.

В: А в чем преимущество метода анализа больших данных?

О: Перед традиционными методами?

В: Да.

О: В целом, большие данные, они дают представление общее об объекте исследования, то есть, когда мы используем соц. опросы, мы все таки... есть некоторая фрагментарность присутствует и, наверное, субъективность больше, или в экспертизе той же, или фокус-группах. Большие данные - это все таки возможность проанализировать весь объект целиком, если мы анализируем пользователей социальных сетей, вот, бери социальную сеть и не выбирай оттуда 15 человек или 100 человек, или сколько у тебя там ресурсов, а бери их всех и анализируй, это, наверное, такое важное преимущество. Они, на мой взгляд, очень освобождают от такого субъективизма исследователя. С помощью больших данных можно фиксировать реальную ситуацию, что происходит, допустим, на том же рынке труда или в той же вакцинации, или качестве жизни, то есть, что есть здесь и сейчас, можно взять и зафиксировать, и можно спрогнозировать, а что будет потом, есть различные модели, которые позволяют прогнозировать развитие объекта исследования твоего, они на больших данных строятся, не скажу, что легко, но это хорошая почва для того, чтобы их строить, какие-то другие методы не дают такую почву, а здесь у тебя есть такая возможность. Ну, наверное, как-то так.

В: А какие тогда недостатки у такого метода?

О: Это дорого, не все научные коллективы могут себе позволить, или не научные, организации, бизнес, не все могут позволить себе ресурсы купить, для того чтобы большие данные хранить, как я

уже говорила, это очень большие объемы информации, соответственно нужны хорошие вычислительные ресурсы, чтобы их обрабатывать, плюс это специалисты дорогостоящие, то есть, именно вот эксперты, кто может с этими большими данными работать, они такие, как сказать помягче, не то чтобы на дороге не валяются, они знают себе цену и эта цена обычно немаленькая. Потом еще можно отметить, большие данные они как бы везде и нигде, то есть нет определенного места, куда я могу пойти и забрать все данные про рынок труда. Нет. Есть, допустим, в одном месте информация о рынке труда, в другом, в третьем, в пятом, то есть есть некоторая фрагментарность больших данных, и для того, чтобы мне их все собрать, нужно приложить дополнительные усилия, чтобы все источники захватить, потом их собрать в одну кучу, потом убрать дубли, чтобы они там не пересекались. То есть, агрегировать все эти данные, в разных источниках находящиеся, в одну большую базу данных - это тоже дополнительные усилия, и они часто довольно много ресурсов требуют дополнительных. Все ругают большие данные за репрезентативность, что типа вот вы изучаете вакцинацию в социальных сетях, а как же пенсионеры, они же вот не зарегистрированы ВКонтакте, это так, это такое вот допущение, но ровно за то же ругают другие разные методы исследования и традиционные, это как бы не ново. Из-за того, что большие данные как бы исчерпывающие что ли, то есть, они позволяют делать выводы... В общем, их настолько много, что проблема репрезентативности перестает существовать, их там миллиарды у тебя вакансий, она сама себя окупает, это решается. Недостатки... Дорого, долго, интеллектуальные ресурсы, репрезентативность, неструктурированность данных, есть большие данные, которые структурированных, а есть которые нет, текстовая информация, которая написана, кто как захотел, как попало, и, соответственно, большие ресурсы требуются на предобработку этой информации.

В: Ты начала говорить про рынок труда, а скажи зачем вообще нужен анализ рынка труда, его мониторинг?

О: Анализ рынка труда позволяет нам зафиксировать спрос на те или иные профессии, на те или иные навыки и компетенции, посмотреть зарплатные предложения по стране, это во многом позволяет и властным структурам, правительству планировать стратегии развития, стратегии страны, регионов отдельных или отдельных отраслей деятельности, в зависимости от состояния этих регионов или отраслей, стратегии управления персоналом для бизнеса это важно и нужно понимать куда вообще движется их отрасль, рынок, смотреть кто им понадобится в их компании для того, чтобы быть успешными в ближайшем будущем каком-то. Для людей тоже, для самих соискателей и для людей, которые просто хотят развиваться в уже имеющейся своей профессии, здесь аналитика рынка труда нужна для того, чтобы построить грамотно свою карьерную траекторию и решить для себя куда направить и вложить ресурсы, время, деньги, что учить, чтобы потом быть классным и востребованным на этом самом рынке. Много на самом заинтересованных сторон в этом на разных-разных уровнях, от власти до простых людей.

В: Как рынок труда меняется и как эти изменения влияют на его анализ?

О: На рынок труда влияют разные факторы, во-первых, это, как мы видим за последний год, это и внутрисполитическая и внешнеполитическая ситуация, то есть различные события, которые происходят, которые несвязанные напрямую с рынком труда, они могут повлиять на то, что все айтишники у тебя соберутся и уедут из России, и нужно будет срочно что-то делать, готовить дополнительные кадры, придумывать программы, которые бы стали аргументом для того, чтобы они вернулись эти самые айтишники или не уезжали, это в качестве примера, на самом деле для всех отраслей своя есть такая специфика, но у айтишников она наиболее явная и наиболее на слуху в последние месяцы. Цифровизация, автоматизация, то есть, некоторые задачи, которые раньше выполняли специалисты, теперь выполняют алгоритмы и, соответственно, они не нужны, те функции, которые они выполняли раньше, теперь не востребованы в компаниях и они вынуждены или переквалифицироваться, доказывать как-то свою валидность, что их на самом деле нельзя заменить полностью, они что-то доделывают или перепрофилируются, начинают какую-то другую нишу занимать. Может быть еще глобализация, то есть, в связи с еще одним внешним обстоятельством - пандемией, стало понятно, что на самом деле для большинства профессий не обязательно очное присутствие, можно работать удаленно, ну, вообще всем, кроме какого-то ручного труда, все, что касается интеллектуальной деятельности, она вполне происходит неплохо и дистанционно, и поэтому мир перемешался, люди работают, находясь в одном месте вообще в другом месте, иногда в нескольких местах и это тоже влияет на рынок труда, когда мы смотрим вакансии и резюме в одном конкретном регионе, нужно понимать, что это не исчерпывающая информация и на эти вакансии могут претендовать люди из другого региона и эти резюме могут уйти в другие регионы, то есть это не 1=1, вот такой тренд важный.

В: Расскажи про проекты лаборатории, которые связаны с аналитикой рынка труда.

О: Он один этот проект, ему год, он маленький, проект появился в связи с тем, что мы в рамках демографии реализуем свою программу по дата-аналитике и наш руководитель Юлия Мудриевская решила подойти к этому осознанно и посмотреть, а что вообще от дата-аналитиков требуется на рынке труда, и она искала, собственно, прям по HeadHunter(y) и другим агрегаторам вакансий этих самых дата-аналитиков и читала вакансии. И мы подумали, а что если сделать сервис, где можно тыкать кнопки, а не читать миллионы вакансий самостоятельно и там тебе будет какая-то аналитика выдаваться. Ну, и в общем, стартанули где-то в марте, с мая начали грузить вакансии и учиться их обрабатывать, нет стратегии в России, вообще нет сервисов, которые были бы на это нацелены, сеть HeadHunter, который делает анализ по вакансиям, но нас интересовали именно навыки, то есть содержание этих вакансий. По навыкам аналитику делают коллеги наши из Университета 2035, но у них там больше на айтишные именно профессии, вообще на айтишные профессии есть есть несколько таких локальных продуктов, которые смотрят, что там требуется от сеньора, что от мидла, от джуна в разных направлениях IT, у них это есть, но по всем регионам, по всем отраслям аналитика не предоставляется пользователям, нет такого инструмента, поэтому мы тут такие уникальные пока становимся, ну, и вот этот год мы занимались разработкой методологии поиском и складыванием в эти самые базы данных вакансий, сейчас у нас их уже 8 миллионов лежит и визуализировали это все дело, и в начале года разработали продукт, придумали ему название “РосНавык” и сейчас активно занимаемся его тестированием, привлекаем людей, основная целевая аудитория у нас - это руководители образовательных продуктов, такие же как мы, как Юлия Мудриевская, кому нужно собрать, условно, своих дата аналитиков. Но развитие, на самом деле, перспективы мы видим большие это и добавление к целевой аудитории соискателей и студентов, выпускников, кто уже вот-вот выйдет на рынок труда и тоже хочет проанализировать свои навыки и компетенции, и понять, куда ему пойти дальше работать или учиться. Это кадровые центры, то есть, РОИВы, органы власти которые заинтересованы в трудоустройстве в том чтобы снизить безработицу в своем регионе, это и работодатели могут быть, которые хотят получить сотрудника готового или почти готового и не учить его тому, что они должны знать при трудоустройстве, ну, или минимально учить, а не как сейчас, многие работодатели жалуются, что выпускники приходят сырые и всему их приходится обучать с нуля по сути. Вот, то есть, размах у нас такой на перспективу широкий, будем искать новые рынки, будем искать новые функции, которые этим рынкам нужны. Надеемся стать таким экспертным центром, площадкой, которая будет собирать на себе всех заинтересованных субъектов и помогать людям понимать рынок, рынку понимать людей, и всем вовремя учиться, прогнозировать что-то, набирать навыки прежде, чем они станут остро необходимыми, и трудоустраиваться туда куда захочется.

В: А что сервис уже умеет? Какие у него возможности?

О: Сейчас сервис может делать аналитику по регионам, по отраслям (около 170 отраслей), отраслей - это я имею в виду специализаций или проф. ролей, профессий, если совсем упрощать, которые защиты на сайтах-агрегаторах вакансий. Может анализировать навыки, которые требует работодатель, в разрезе на soft- и hard- skills, сейчас нельзя говорить soft и hard, мягкие и жесткие навыки. Может делать анализ по зарплатным предложениям от различных работодателей, в различных регионах и различных отраслях, может делать анализ по условиям труда, которые предлагают работодатели по тем или иным специалистам, то есть, людям, которые обладают какими-то навыками, допустим, вот за этот навык мы готовы платить вот столько и работать ты можешь на удаленке, а вот за этот мы готовы платить меньше и работай только очно, пожалуйста. Вот такие вещи можно легко тыкать в дашборде, который у нас сейчас уже представлен на сервисе. По динамике можно отсмотреть тренды навыков и вакансий, то есть, можно выбрать какой-то навык, например, я соискатель хочу учить Python, условно, и я тыкаю “Python” и смотрю сколько вакансий, в которых требовался Python было за последний год и вижу, что, допустим, в начале года Python был очень нужен, в середине стал не очень, а сейчас опять очень нужен, ну и принимаю уже окончательные решения, хочу ли я учить Python или может быть уже сейчас достаточно специалистов с таким навыком и мне стоит выбрать SQL и учить его. Есть анализ по сетам навыков, то есть надо было запросы от руководителей образовательных продуктов, что на самом деле им важно знать не только то, что я сейчас перечислила в привязке к конкретному одному навыку, им еще важно понимать такие комплексы навыков, то есть, что работодатели, которые просят навык А, просят еще какие навыки там, Б и так далее. Это важно для проектирования образовательных программ, потому что там они зашивают сразу комплексы этих курсов, которые каждый курс направлен на выработку того или иного навыка, вот поэтому мы сделали отдельный граф со связью

этих навыков. Также ведем регулярную статистику, выкатываем по тому, сколько вакансий публикуется, сколько в них навыков публикуется, сколько из этих навыков, ну скажем так, нерелевантные. Очень часто работодатели пишут в требованиях не навыки и компетенции, а разные слова, которые не интерпретируются вообще. Вот их довольно много, качество составления вакансий сейчас оставляет желать лучшего на российском рынке труда, поэтому, в общем, все, что делает наш алгоритм, мы подробненько описываем и тоже статистику в открытый доступ выкладываем, чтобы пользователи могли понимать, а на основании чего они вообще эти данные получают. Записали материалы обучающие, скринкасты, презентации, сделали все в куче разных форматах, и в pdf можно почитать текстом и видео можно посмотреть о том, как этим всем пользоваться. Проводим семинары, где презентуем проект целевой аудитории, в новостях о нас сейчас активно стали писать, то есть сейчас мы находимся на такой стадии продвижения, ну того, что я описала, чего мы уже научились делать.

В: А есть уже кейсы использования данного сервиса и что в них получилось, а что нет?

О: Да, уже есть несколько кейсов, мы в первую очередь конечно делали их с ТГУ, потому что как бы наши коллеги, все-таки мы разработали, знакомые и они очень активные и заинтересованные. Делали кейс для проектирования программы библиотечарей будущего, программа несколько лет существует, и руководитель программы смотрел, собственно, какие навыки от библиотечарей требуются и соотносил это с теми навыками, которые формирует его программа и обновлял курс в ней. Делали мы кейс по бизнес-консультированию, тоже магистерская программа собирается в ТГУ, заказ у нас был, какие навыки требуются от бизнес-консультантов в разрезе на европейскую, скажем, часть страны и за Уралом, сравнительный анализ делали. Потом с Уфой у нас есть сейчас в процессе реализуется кейс. То есть у них задача проанализировать вообще рынок труда республики Башкортостан и в зависимости от того, какие навыки требуются в этом регионе, то есть у них именно акцентный региональный аспект, формировать образовательные продукты. У них есть большой университет, ну прям гигантский, который в себе собирает там порядка 10 вузов и вот все эти вузы целены на то, чтобы обеспечить кадрами всю Большую Республику Башкортостан. Наши коллеги прямо из нашей лаборатории делают кейс, ну, как бы это даже был не заказной кейс, это был такой прикладной кейс, который в принципе мы не просили делать, он сам получился. У нас есть магистратура с Яндекс.Практикумом по BI-аналитике и руководитель программы, когда создавал эту магистратуру, уже пользовался сервисом Роснавок. Ну, в общем, сферы разные. По английскому языку тоже мы делали анализ, где, если знаешь английский язык, где ты можешь работать. Если по дизайну, тоже у нас очень популярные навыки, ну дизайнерские навыки популярные в России, вот мы по ним делали отдельный кейс. Их все можно посмотреть, они у нас в открытом доступе лежат на сайте rosnavyk.ru.

В: А есть ли какие-то моменты, которые не может сервис, которые планируют развиваться?

О: Да, таких моментов много. На самом деле, чем больше делаешь, тем больше хочется. И у нас запросы растут очень быстро, много чего хочется. У нас сейчас сервис сделан дашборд, то есть инфографика, которая представлена, она сделана в программе Polyanalyst - это наши партнеры, компания российская Megaputre Intelligence. Это очень хороший продукт, который позволил нам быстро визуализировать большие объемы данных, но, к сожалению, у него есть свои ограничения. Например, связи можно ставить только односторонние от элементов, то есть если мы, например, знаем в каких компаниях требуется тот или иной навык, то в обратную сторону посмотреть, какие навыки требуются в той или иной компании, мы уже не можем. И это значительное ограничение, в основном с этим связано пожелание коллег, и мы планируем за лето сделать дашборд самостоятельно, уже на своих ресурсах, и там ставить связи в любую сторону, в любом количестве, какие нам будут угодно. Вот это, наверное, важная часть развития. Плюс сейчас у нас сервис сделан только для одной целевой аудитории, наш основной руководитель образовательных продуктов. Хотим расширить, сделать еще для соискателей, кадровых центров в ближайшее время, то есть для соискателей внедрить туда профориентационный тест на выявление тех или иных навыков компетенций, и если эти навыки компетенции тест выявляют, то есть они присутствуют у человека, можно будет перейти на дополнительную страницу и посмотреть, в каких компаниях могут люди пройти стажировку или подработку, если они студенты, практику, если они выпускники, имея этот набор компетенций, который у них выделился. Либо можно будет перейти на сторонний сервис по обучению, то есть у нас есть, это даже не наш партнер, это наш соавтор, это академия Data Diving, которая занимается образовательной деятельностью. Также у нас в партнерах есть ИДО ТГУ, у них тоже большое количество образовательных материалов. Вот на такого рода ресурсы мы планируем сделать переадресацию для людей, которые прошли тест. Ну, условно, там я прошел тест и вижу, что у меня не хватает лидерских качеств. Я тыкаю на ссылочку и вижу там 5

курсов на выборы, где я могу повысить этот, прокачать этот свой навык. Ну или и с хардами точно также. Такая же ситуация. Наверное это вот из основных планов на ближайшее будущее все.

В: Если подытожить, что вообще могут большие данные в аналитике рынка труда?

О: Большие данные могут в аналитике рынка труда. Во-первых, с их помощью можно получить доступ к самым разным уголкам нашей большой страны и посмотреть максимально насколько это возможно картину по региону или по отрасли. То есть собрать все-все вакансии в одну общую базу и агрегировать их, проанализировать вместе. Наверное здесь стоит сказать, что нельзя говорить о том, что вообще абсолютно все вакансии есть на сайтах-агрегаторах, которые являются сейчас нашим источником данных. Например, есть у нас кейс с экскурсоводами. Мы там встречались с ассоциацией экскурсоводов и они нам сказали, что многие вакансии они не выкладывают, а ищут сотрудников по своим каналам. Конечно к таким вакансиям мы доступ не получим, этой информации у нас просто нет. Это такое значимое ограничение. Ну тут пока они не начнут их где-то публиковать, мы их никак и не найдем. Только точно брать экскурсоводов, как мы сейчас сделали, и анализировать прям их отдельно, всю их сферу. Большие данные дают возможность в режиме реального времени мониторить вакансии. То есть когда мы собираем экспертов, ну в качестве традиционной методы аналитики, спрашиваем у них, что будет с вашей сферой дальше или как сейчас у них дела обстоят, то мы не можем их собирать каждый день. А большие данные дают нам возможность каждый день смотреть, что меняется. И поэтому если какие-то изменения происходят, мы можем это зафиксировать в моменте. Вот буквально там за неделю посмотреть и увидеть, что на каких-нибудь людей в каком-то регионе резкий всплеск возник. Ну там, я не знаю, технопарк у них создали или какое-то производство в этот регион переехало. Мы это можем вот прям вот так зафиксировать очень быстро и соответственно настроить обучение в этой области и какую-то государственную политику, с этим связанную тоже, если мы РОИВ, орган власти какой-то. Большие объемы, фиксация в моменте мониторинговая. Это все,

что касается сбора данных. По поводу аналитики, большие данные, ну вот как сейчас у нас они реализованы в дашборде, это сделано с помощью текстовой аналитики вакансий. Если хорошо настроить методологию, проработать ее, то алгоритмы машинного обучения, нейросетки, они смогут делать всю работу за вас. То есть они будут просто собирать все вакансии, которые вы уже настроили, и раскладывать их по тем классам, которые вы уже задали. Вам нужно просто сидеть и смотреть, как это все красиво раскладывается по корзиночкам. То есть это на самом деле большие, я бы сказала, вложения, но это и большая прибыль на выходе в плане того, что когда у тебя уже все работает, ты просто это курируешь и особо никаких дополнительных ресурсов не вкладываешь, ну просто видишь, что творится. Как бы это начинает работать само, жить своей жизнью, без тебя, как автора скажем так. Вот, ну и большие данные - это классно, используйте все большие данные, потому что учитывая, какой мы сейчас живем век цифровизации, информатизации, сколько информации выливается в открытый доступ, это нам дает вообще все карты в руки, чтобы пользоваться ей, а не пытаться достать что-то сверх, дополнительно куда-то, собирая экспертов или, я не говорю, что эксперты плохо, эксперты это хорошо, но экспертов это не достаточно в наших текущих реалиях, поэтому они и искусственный интеллект, ну машинное обучение, алгоритмы, классификации, они должны работать, на мой взгляд, в связке для того, чтобы собирать максимальное возможное количество информации и максимально правильно его анализировать, потому что алгоритмы тоже не совершенны, тут эксперты, конечно, выигрывают, вот их задача - научить эти алгоритмы и просто наблюдать, как все будет работать само.

В: А в чем ограничение больших данных в исследовании рынка труда?

О: На мой взгляд, большое ограничение - это культура заполнения вакансий, дело в том, что работодатели не хотят как будто бы делать всем хорошо, себе в том числе, они пишут непонятно, они пишут не в те поля, то есть когда грузим, есть реляционные и нереляционные базы данных, реляционные, ну типа Oracle, MySQL, PostgreSQL - это такие базы данных, где данные структурировано хранятся, то есть есть какое-то поле, есть какой-то текст, ну условно там Key Skills и в Key Skills мы предполагаем, и создатели сайта HeadHunter, например, предполагают, что туда будут тезисно записаны, что нужно от человека, но туда написано вообще, что попало, то есть мы вот с центрами занятости работаем, с департаментом труда занятости томским нашим, они нам об этом говорили еще год назад, что на самом деле работодателя сложно заставить написать корректно вакансию, они копируют друг у друга, они не отдают себе отчет часто, что они там пишут, они дублируют вакансии, ну в общем, большое количество текста получается нечестного, скажем так, который приходится дополнительно чистить, и учитывая объемы информации, объемы мусора такие же огромные, и это сложно, отделить мусор от не мусора, на это время дополнительно уходит силы, и

в общем, если работодатели начнут грамотно заполнять вакансии, нам будет всем легче, потому что и нам будет понятно в плане аналитики, и соискателям будет понятно, чего от них хотят, и в итоге работодатели в первую очередь выиграют, потому что к ним будут приходить те люди, которых они зовут, а не все подряд. То же самое с зарплатой, часто какая-нибудь компания пишет от 20 тысяч, например, или наоборот от 20 до 200 тысяч, и мы такие, интересно, ну в общем, не может быть такого, на мой взгляд, чтобы такой был гигантский разброс, они видимо пишут, чтобы людей так привлекать, ну их есть какие-то свои очевидные цели, но получается сложно. В плане того, что вот я про экскурсоводов рассказывала, есть некоторые профессии, которые ну просто не проявляются, так как есть на самом деле, и это требует дополнительных усилий, это пока я не знаю, какую задачу решить, кроме как вот идти к экскурсоводам и сажать этих столов и спрашивать, что, ну собственно, традиционные методы применять, скажем так. Есть компании, которые, ну кадровые агентства, скажем так, которые дублируют вакансии, если зайти на HeadHunter или там на Superjob, можно увидеть, что одна и та же вакансия по одинаковому шаблону написана там в семи разных регионах, ну потому что глобализация, люди переезжают, они не боятся, плюс удалёнка, и эти вакансии, они вирусные, они просто рассылаются везде, конечно, их нельзя учитывать как отдельную единицу в общем зачете, то есть есть дополнительные сложности с обработкой такого рода данных, с их вычислением, с фильтрацией, но это тоже всё в сторону мусора, то есть, наверное, пока я такие вижу ограничения, вот связанные именно с обработкой этих вакансий. Ну, может быть, с доступом к данным, то есть у нас есть три крупных сайта агрегатора, это HeadHunter, Работа в России и Superjob, на который мы сейчас ориентируемся, и у них, ну у HeadHunter и Superjob открыты API, и мы можем грузить вакансии, то есть они не против и даже за, а вот РВР он выкладывает сам датасет и каждый день, то есть они тоже делятся данными, это не секрет, но если в какой-то момент они передумают и скажут нет, мы никому не дадим грузить наши данные, то у нас возникнут проблемы, то есть можно, конечно, писать краулеры, парсеры и таскать эти вакансии иначе, но это тоже дополнительные будут технические трудности, с этим связанные. У нас возникает несколько раз проблема, связанная с тем, что у нас вис дашборд, потому что в нем очень много вакансий и когда в него заходит большое количество людей, ну, вычислительные мощности наши не выводят, поэтому мы покупали сервера, увеличивали лицензию с полианалистом, чтобы нас допускало до 100 человек за раз, изначально было 10, это было сложно, но вот мы выросли количество пользователей и лицензию нам дали побольше и как бы с этим мы справились, хотя это, ну, все равно большое такое ограничение. Экспертиза тоже, мы собирали экспертные сессии по каждой отрасли и это вообще отдельная была работа гигантская и отдельные проблемы там были связаны с тем, что эксперты не могут договориться, то есть мы берем двух людей с одинаковой должностью из разных регионов, они говорят диаметрально противоположные вещи, ну, потому что у них вот разный опыт, они из разного опыта своего так делают, поэтому можно до бесконечности совершенствовать классификатор навыков в каждой отрасли, но в какой-то момент приходится сказать себе стоп, кажется, хватит и на какой-то более-менее хорошей версии остановиться и ей пользоваться, потому что на самом деле, если так вот искать каждый раз новых экспертов и каждый раз спрашивать у них про навыки, которые требуются в их сфере и показывать им, что пишут в вакансиях, у них у всех будет свое мнение и оно очень часто, оно очень редко будет совпадать вот так, наверное, оно всегда будет не совпадать. Экспертов сложно найти, именно вот настоящих экспертов, мы сначала привлекали руководителей в разных крутых компаниях в той или иной отрасли и поняли, что руководители нам не подходят, то есть нам нужен средний менеджер среднего звена или вообще люди, которые прямо работают на производстве, они намного лучше в этом всем разбираются, чем те же HR или тот менеджмент, вот здесь отдельные свои такие проблемки есть. Ну, наверное, из основного крупного, что нам испортил жизнь. Все.

В: А какие ты видишь перспективы развития больших данных и как они повлияют на анализ рынка труда?

О: Перспективы развития, хотелось бы, конечно, чтобы вакансии переходили в хорошее качество и чтобы, возможно, у нас уже была такая попытка в России, как раз у сервиса Работа в России, ну есть, допустим, город-работ, в общем, сделать такой агрегатор всех вакансий, где они действительно будут все. В городе работ они не совсем все, а вот в работе в России, ну как мы с коллегами разговаривали, они говорили, что у них как раз идея в том, чтобы собрать все-все сайты, агрегаторы, сложить все вакансии в одну базу, вычистить дубли и вот к ним открыть доступ. Вот это было бы очень классно, это бы очень упростило нам момент со сбором данных. Было бы здорово, если бы все-таки в России сформировалась работодатель культура заполнения вакансий, и они стали писать более осознанно текст в объявлениях о работе своих. Мне кажется, сейчас уже на самом деле на достаточно хорошем

уровне находится Big Data, и даже вот те возможности, которые сейчас она дает, не до конца используются, потому что, ну хотя бы потому что сервиса для мониторинга навыков в 2023 году не было, хотя сервисы-агрегаторы вакансий существует уже очень много лет, десятки некоторые. Поэтому сейчас самое основное это перейти от традиционных методов именно анализа к нетрадиционным и всячески задействовать автоматические алгоритмы. Ну в принципе тренд на это уже давно есть, и сейчас он очень такой на пике. Я думаю, надо подождать еще пару лет и тогда думать о том, куда дальше все пойдет. Пока все и так очень хорошо.

В: Хорошо, спасибо.

Интервью №3, Юлия, младший научный сотрудник лаборатории прикладного анализа данных ТГУ, руководитель образовательных программ академии Data Diving
Длительность: 30 мин.

В: Так, тогда начнем. Скажи, какую должность занимаешь? Какими проектами занимаешься?

О: У меня два места работы, две должности. Одно официально здесь я являюсь... По документам я младший научный сотрудник в лаборатории прикладного анализа больших данных. По факту я являюсь руководителем образовательных программ по анализу данных в университете. Веду программы ДПО и магистерской программой, ну, академическое руководство, по крайней мере, несу, программой магистерской дата-аналитика для бизнеса совместно с Яндексом. Также параллельно я являюсь сооснователем, руководителем Академии Data Diving. Это тоже образовательная организация. Она занимается разработкой и реализацией онлайн-курсов по анализу данных массового обучения людей.

В: А как давно работаешь в лаборатории и с большими данными?

О: С самого основания лаборатории. Это, по-моему, с 2016 года, когда он там 16-17 годов появился. Собственно, меня Кашпур привел туда и все.

В: А что вообще такое большие данные?

О: Вот вообще здесь я бы сказала, что про большие данные не то, чтобы стоит говорить. Причем, ну, большие данные, если мы откроем термина в любой википедии, это будут данные, которые... Ну, по-простому, это те данные, которые не поддаются компьютерной обработке либо ручной обработке обычными нашими техническими средствами. Это простое объяснение. Есть более сложные объяснения, которые говорят, что у больших данных есть куча критериев, типа там объемов, скорость обработки. Но по факту я бы не говорила вообще про большие данные. Причем большие данные существуют на уровне огромных корпораций. Это фанк-предприятия типа Яндекс, Google, Майк и так далее. В общем, такие корпораты, где как раз таки на каждом шагу генерируется некоторый объем данных. И, естественно, они их обрабатывают в больших дата-центрах. Когда мы говорим, если про образование, например, либо про какие-то просто предприятия, лучше говорить просто про данные. Потому что даже в нашей обычной деятельности мы не оперируем большими данными как таковыми. И в принципе в России предприятия, наверное, процентов 70-80 предприятий, которые про большие данные ничего не знают в принципе, потому что они их не обрабатывают. Ну и у них столько не копится. Потому что, почему я говорю про большие корпораты, один Яндекс взять, сколько там поисковых запросов, сколько почты ежедневно генерируется постоянно. Но это размер там ежедневно появляется более, там, 5 петабайт данных. А это прям огромный объем. Вот там, да, большие данные.

В: Так, а как работать с такими данными?

О: Ну опять же, если мы говорим непосредственно про большие данные, да, в принципе, про любые данные, давай, не буду ограничиваться. Есть некоторый цикл работы с данными. Это есть методология, прям, полноценная, CRISP-D. Это методология, как раз, которая диктует нам некоторый процесс работы с большими данными. Если по-простому, то у нас весь процесс делится на 5 основных этапов. Там, целеполагание, первичное исследование, сбор данных, первичная обработка данных, потом исследование данных, описание. Это то, что любят. А потом начинается внедрение, либо построение выводов. Вот, соответственно, так работают с данными. На каждом этапе существуют свои алгоритмы. Причем, я говорю не только про технические алгоритмы, есть некоторые аналитические модели, с помощью которых мы с ними работаем. Это либо какие-то сложные отраслевые модели, либо это модели больше не отраслевые, а универсальные типы математических моделей. Но опять же, здесь стоит сделать различие, когда мы говорим про работу, можно работу какую конкретно и для кого, с какой целью. Почему? Потому что, когда мы говорим про работу с большими данными, есть огромное количество отраслей, которые этим занимаются. И я

бы так сказала, видов направлений и специализаций, которые с ними работают. Это может быть дата-аналитика, это может быть дата-сайенс, дата-аналитика, да, когда мы анализируем для того, чтобы принести пользу там обществу, человеку, компании, принести прибыль, ну вот это все. Дата-сайенс, например, это направление работы с данными. Вот так, кстати, большие данные скорее могут больше появиться. Почему? Потому что дата-сайенс, это прямо наука практически направленная на то, чтобы не просто описать, что у нас происходит, а сделать некоторые прогнозы, построить, выявить повторяющиеся модели, чтобы потом их уже применить, создать некоторые приложения, например, прогнозное приложение, которое будет говорить, что с предприятием будет потом, например. Есть также разные нейросетевые методы и вот это все, которое тоже работает с данными, поэтому начинать нужно с того, какая у нас цель. В принципе, я из этого, да, у практически всех есть одна единая методология с помощью которой, с помощью которых мы работаем, но в принципе опять же это какая цель и какое направление работы с данными, их прямо очень много сейчас.

В: А если переходить к рынку труда, то какие проекты в лаборатории есть связанные с рынком труда и какие цели в этих проектах?

О: Ну, наверное, самый главный проект ключевой на данный момент - это проект "РосНавык", которым вот Даша руководит, который мониторинг труда. Это, ну, главный и единственный, но разве что стоит добавить мы, например, все свои программы. Я просто первая кто это делала, потом Даша просто подхватила. Я просто это попробовала пилотировать на своем проекте, потом из-за этого у Даши родилась идея со Славой, что это можно делать прямо в приложение классное и там дальше подзавертелось. Мы в принципе используем данные, как раз анализ данных для разработки программ, причем, ну, когда "РосНавык" еще разрабатывался, я периодически там заглядывала, и сейчас это делаю периодически, обращаюсь к данным для того, чтобы разрабатывать ряд программ. То есть, с одной стороны, данный рынок труда анализируем для того, чтобы мониторить рынок. Есть программное приложение, которое создан для этого мониторинга. Причем, у него, насколько мне известно, есть несколько направлений, одно направление, это его создали для того, чтобы помогать как раз руководителям образовательных программ, чтобы они создавали на этом свои программы и могли это делать быстро, гибко и более мобильно. А с другой стороны, ну, как я знаю, у него потенциал потом идти дальше в некоторые профориентационные направления. Ну и второе, именно где работал с данными, это непосредственно при конструировании программ.

В: А зачем вообще анализировать рынок труда, кроме для того, как создание программ? Для чего еще может быть это полезно?

О: Я бы сказала, что первичное, рынок анализировать полезно не просто для того, чтобы создавать образовательные программы, чтобы вообще смотреть, как изменяется мир, как изменяется экономика. Потому что рынок труда - это, наверное, первое место, по которому видно изменения в экономике, в принципе, в стране. Какие должности умирают, какие зарплаты предлагают, в каких регионах что происходит, можно увидеть именно на примере рынка труда.

В: А как рынок меняется и как изменения сказываются на его анализе?

О: Здесь очень сложный, на самом деле, вопрос. Я не то, чтобы сильный специалист в рынке труда, я знаю только тенденции, которые сейчас происходят в IT и, наверное, в каких-то смешанных. Если говорить, в принципе, про общие тенденции, мы все знаем, что рынок IT растет, особенно после событий прошлого года, когда у нас сентябрь грянул, даже не февраль, именно сентябрь. Когда началась история с релокацией и всем

остальным, у нас рынок IT очень сильно начал взрываться, я бы сказала. Сначала начался очень сильный кадровый голод. Но при этом здесь важно говорить, кадровый голод был всегда и будет он скорее всегда. Если раньше мы говорили про то, что кадровый голод значит на рынке не хватает людей, которые придут на должности, сейчас он совершенно иной. Здесь немножко камень в огород курсов, программ и всего остального, на рынке очень много людей, которые в потенциале могли бы закрыть весь

кадровый голод, который там есть. Людей достаточно, людей с дипломами достаточно, недостаточно хорошо подготовленных людей, особенно если IT брать. Я просто очень люблю этот рынок, он прям такой парадоксальный, на мой взгляд. Очень много людей, которые заканчивают кучу всяких обучалок доморощенных, я на YouTube посмотрю, научился, либо я пошел в skillbox научился, либо меня в универе научили. В универ камень бросать такая себе история, хотя я это тоже могу. Но фишка в чем? На рынок сейчас, особенно после пандемии вышло огромное количество людей, которые типа готовы работать, которые хотят работу, хотят пойти хоть куда готовы, по-любому зайти в IT. И это, кстати, не только IT-шки касается, я бы сказала, много их экономической специальности, типа менеджерские, а еще дизайнеры всякие, фотографы, они прямо сейчас наплодились именно

после курсов. И с этой точки зрения казалось бы, людей много, о каком кадровом голоде мы можем говорить, вот же кадры, их прям много. С другой стороны, есть такая парадоксальная ситуация, когда кадры есть, но они плохие. Потому что, как я сказала, много самоучек, много недоучек, много людей, которые считают, что им достаточно будет того, что дали им на курсе, а еще курсы не очень хорошие. И когда мы, по крайней мере, с работодателем тоже разговариваем, нам приходится это делать, потому что мы валидируем свои данные. То, что мы валидируем, типа вот это то, что мы на рынке увидели, это вот так или не так. Либо просто, когда выясняли, о каком должен быть более полно, чем это есть у нас, собственно, в данных, но говорили, что да, классно, джунов много, мы бы взяли, но они не очень. Почему не очень? Потому что практически каждый человек сейчас, будет грубо, но ленив, и, как я сказала, дальше курсов люди не учатся. У работодателей сейчас появились, это вот про изменения на рынке труда, очень много нестандартных задач, потому что а) импортозамещение, б) огромная конкуренция. Соответственно, всем нужны специалисты, не которые умеют все то же самое, что умеют все остальные. А уметь, в первых, немножко больше, уметь немного по другому. Соответственно, очень много людей, которые выходят на рынок труда и хотят на найти работу, они к этому не склонны, не способны, к сожалению. Вот, надеюсь, меня никто не убьет за такие слова. Но они пока еще не понимают, что выходя на рынок труда, им никто не должен. Это первое, наверное, изменение, которое сейчас и самое парадоксальное, что у нас огромный кадровый голод, но при этом у нас огромное, такое, и спрос, и предложение соотносятся между собой. Это очень интересно. Ну и второе, это, я бы сказала, трансформация рынка труда не в плане того, сколько людей, а в плане традиционного понимания профессий. Сейчас, особенно в эту заезженную эпоху цифровизации, будь она неладна, но очень многие профессии изменились. Соответственно, приходится за этим людям успевать. Изменились даже такие традиционные, ну, понятные для нас, вот социолог, когда я училась на социолога, какой социолог должен быть сейчас, это два разных уже, каким он будет через 10 лет, это уже будет не просто там компьютерный какой-то социолог, а какой-то, не знаю, мега-евангелист, там что-нибудь такое. Журналисты такие же, например. Сейчас очень многие редакторы, те, которые пишут статьи, уже не так сильно нужны. То есть это уже уход в копирайтинг, например. Хороший журналист - это журналист на основе данных. Начинается понимание, что журналист это исследователь, это вот прям человек, который будет копаться в данных, а не просто ходить, например, брать интервью или что-то искать. Мы это и в данных видим, мы это видим по разговорам с самими, там, издательствами, например. И это касается вообще многих очень профессий. То, что очень многие профессии начали меняться, и, кстати, в качестве рекомендации я даже тоже про это говорю, агентство стратегических инициатив буквально в прошлом году выпустил третью редакцию атласа профессий. Этот атлас профессий - это некоторая попытка спрогнозировать, каким будет рынок труда, и он как раз говорит о том, что в течение пяти лет, да, пяти, может десяти, не знаю, насколько это точный прогноз, но больше шестидесяти профессий уже не будут существовать. Соответственно, очень многие люди придут к тупику некоторому самому профессиональному, и надо будет очень быстро перестраиваться. Это вот еще одна из таких тенденций нашего рынка. А вот более точно, что там в конкретной сфере, я вот не скажу. Я просто могу общую динамику видеть.

В: А вот те проблемы, которые ты отметила, как их можно решить, могут ли помочь в этом большие данные как раз таки?

О: Честно, я не уверена. Почему? Потому что рынок, скажем так, с каким-то, да, с каким-то нет, быстро, гибко людям перестраиваться на рынке труда, переориентироваться, перепрофилироваться, я бы сказала, смотря на то, как динамичен рынок, как он изменяется, и пытаюсь строить прогнозы, кто будет нужен, тогда люди смогут успевать за трендами, в принципе. И уже сейчас мы можем примерно попытаться, но я, например, человек, который захочет учиться и получить крутую профессию через пять лет, я уже сейчас могу посмотреть, ну, например, этот Атлас, посмотреть на некоторые тренды и сказать, ага, возможно, через пять лет нужны будут вот эти люди. И начать к этому готовиться. С этой позиции, да, я как человек, допустим, соискатель, могу быть более востребован, а рынок сможет закрыть свои, там, кадровые дыры. С другой стороны, я бы сказала, что не совсем все-таки могут данные помочь, потому что если это проблема на рынке касается не каких-то экономических, я бы сказала, проблем, не проблема с именно, с его непосредственной динамикой, а именно качеством кадров, с одной стороны, помочь можно, а с другой стороны нет, потому что, ну, обучение, квалификация людей, которые будут закрывать и должны закрывать позиции на рынке, с этим данные не помогут, потому что это зависит непосредственно от самих людей. Вот, это же мы постоянно говорим, когда обучим людей, даже при том, что у нас классные программы, мы действительно готовим на основе данных, мы там очень много чего перелопачиваем для того, чтобы

создавать реально крутые программы, которые будут соответствовать рынку труда, и причем они у нас динамичные достаточно. Мы всегда объявляем людям, что мы вам не дадим 100% гарантии, что вы найдете работу, правда, даже как компанию, в которую могут к нам обратиться, мы не можем вам дать 100% гарантии, что человек, который придет к вам, от нас будет классный, мы сделаем для этого максимум, а остальное зависит от него. Вот, поэтому с чем-то данные помогут, ну, например, просматривать динамики изменения и попробовать подготовиться к ним, а с чем-то нет.

В: Так, а в чем преимущество анализа больших данных в сравнении с другими методами именно в исследованиях рынка труда?

О: Насколько мне известно, когда мы анализируем рынок труда, есть вообще несколько подходов. Ну, классический подход - пойти к работодателям в интервью, поспрашивать, ну, типа, а нет, экспертный, это еще отдельно, там пойдет, окей, вот. Я когда только начинала, там, когда нам говорили в penale, что программа должна быть востребованная, и как раз начали изучать там, да, первый вариант - пойти к работодателям, ну, опять же, сколько ты интервью возьмешь там за день, тебе надо все быстро еще их обработать, расшифровать, и вот этот весь кошмар, это одно. Второе - экспертное. Ну, там, во-первых, тоже самое, но и как с работодателем, каждому работодателю нужен свой классный эксперт, который, свой классный спец, который будет разбираться именно в том, что ему конкретно надо. Ну, в России по последним данным я вчера поднимала примерно более 4 миллионов компаний, и представим, что каждому нужен свой специалист, но мы же не можем готовить. Во-первых, ты их всех не опросишь, вот большие данные здесь гораздо проще спрашивать, именно там, не знаю, HeadHunter, Superjob, потому что там уже это есть, и мы не сможем физически опросить огромное количество людей. Второе - это то, что действительно каждому работодателю нужно что-то свое, а с помощью больших данных мы сможем немножко от этого отойти, как бы это не было хорошо или плохо, а посмотреть более объективно, типа какие функции повторяются, какие компетенции повторяются. Ну, вот. Ну, и третье, чем лучше, когда мы спрашиваем работодателей, экспертов, у них достаточно зашоренный взгляд. Но это не только потому, что им нужен кто-то свой, а потому что они, скорее всего, и чаще всего, эксперты только в своей какой-то области, и не эксперты в остальных. Вот. А нам надо очень сильно избегать субъективизма и некоторого в том, ну, при подготовке программ, при анализе рынка для того, чтобы это стало эффективным. Даже сейчас мы, допустим, когда магистратуру нашу разрабатываем, мало того, что мы смотрели данные, мы смотрели данные, потом мы обращались к рынку, потом мы обращались к стандартам, и до сих пор у нас нет четкого единого понимания на самом деле, кто там нужен, а какие навыки ключевые и так далее, потому что у нас на основе данных Яндекс по-другому работает. Соответственно, наши данные начинают расходиться, и вот это еще одна проблема.

В: А какие-то еще есть ограничения при использовании больших данных в рынке труда?

О: Знаешь, есть не ограничения, я бы сказала, есть сложности. Сложность первая в том, что каждый... От компании, например, вакансии, ну, когда мы анализируем рынок труда, мы же чаще всего анализируем либо резюме, либо вакансии. Чаще вакансии. Вакансии - это то, что нужно работодателю. Здесь есть проблема первая. Работодатель не всегда знает, что ему надо. Во-первых, не всегда знают, что человек его он просит, это именно аналитик, а не, там, не знаю, кто-нибудь, рабокоп какой-нибудь. Второе - это вакансии заполняют... Здесь, больше подчеркнуто, я бы сказала, вакансии заполняют HR, которые вообще не понимают ничего в отрасли, как правило, либо понимают прямо минимум. Это даже на нашем универе есть пример HR ТГУ, иногда обращаются ко мне, появляется какая-то техническая вакансия из разряда аналитик, разработчик, тестер, там, вот это вот всё. Они начинают приходить ко мне и спрашивать, а мы правильно всё описали? Либо они не могут долго закрыть. Ну, то есть, я представляю себе ситуацию, они не могут четко описать, что им надо, а как ты будешь это анализировать, если они сами не понимают, что они там имели в виду, и ты должен ещё на данных увидеть, что же они там имели в виду. Ну и действительно, каждый пишет своим языком, это 50 разных языков просто, которые надо сличать. Одного только аналитика напишут 500 разных способов, как обозвать аналитика, это и аналитик, и дата-аналист, и дата-анализист, и там вот это всё начинается, и ты сидишь, ну, и чё вот это за херня на самом деле? Я когда анализировала первый раз, это было ужасно, потому что я сидела, копалась в каждом навыке, о чём они там имеют в виду, а ещё они могут писать максимально обширно, типа, ну это вот с точки зрения образования программы, они тебе напишут, классный аналитик должен уметь анализировать данные, а ещё у него должно быть аналитическое мышление, навык аналитических исследований, и так далее. И ты начинаешь, что такое аналитическое мышление, давайте декомпозируем, с точки зрения, например, программ, это надо разбирать. Скорее всего, не только с точки зрения программ, это надо разбирать. А вот это типа, анализировать данные, что такое анализировать данные? Вот они

это так пишут, например, а когда мы хотим людей подготовить, вот мы представляем ситуацию, они написали, работодатели дружно с кубом, написали датамайнинг или анализировать данные. Ты такую собрал программу, которая учит анализировать данные, в ней там какой-то стек технологий, который ты чё-то там на гугле, вот так далее. А приходит, ты обучил на своей программе людей, тебе кажется, что твой человек подходит тому, что там было написано, а он приходит на работу, и ему говорят, ну нет, у нас стек технологий другой, а ещё у нас вот это, это такое, а какой нужен-то был? И вот здесь такая проблема неточности, очень большая неточность в том, как люди, в принципе, как компании, как своих сотрудников потенциальных, как они пишут, и соответственно, вот большая проблема с анализом данных, потому что мы не, каким бы мы специалистами не были, мы не всегда понимаем, что они там хотят. И это прям большая проблема при анализе, потому что и в конце концов потом оказывается, что то, кого мы готовим, да даже то, какие данные мы представляем, анализируем, неважно, нужно нам готовить или нет, они могут оказаться нерелевантными рынку реальному, они могут расходиться именно поэтому. И я знаю, что у Даша с этим встретила, со своим РосНавыком, когда я была просто при тех созвонах, несколько, пару-тройки, когда говорили, типа, вот наши регионы неправильно отразили, то есть данные, рынок труда нашего региона некорректный, ты спрашиваешь, почему, а потому что что-то не так было описано, что-то куда-то там подсмывается, там кто-то вообще не знал, как это написать, ну, то есть там такое возникает, и соответственно, огромная проблема в том, что не просто, чтобы обработать данные, так далее, чтобы оно реально соответствовало, и это прям огромная работа, которой сейчас нам только предстоит, именно экспертная работа, чтобы данные, которые есть, которые представляют нам работодатели для анализа, именно соответствовали этому рынку.

В: А как анализ больших данных будет меняться, и как это скажется на мониторинге рынка труда?

А что технологии, или что конкретно, или отрасль? Или то, или другое? Не совсем корректный вопрос, если честно, почему некорректный, потому что, ну, типа как меняется анализ больших данных, ну, он меняется, но просто что конкретно? Я скажу так, как бы так сказать, чтобы отредить, смотри, аналитика труда, ой, не аналитика труда, анализ данных, анализ больших данных, и все, что мы подразумеваем в этой отрасли, сейчас претерпевает огромные изменения, в общем-то, как и многие отрасли. Во-первых, появление в нейросети, в которой сейчас GPT, вот это вот все появилось, она начинает менять в принципе отрасль анализа данных, потому что она теперь может многое. Второе, это тенденция, которая давно появлялась, сейчас на самом деле я очень рада, это упрощение аналитики данных и переход к новому типу инструментов. Традиционно мы предполагаем, что чтобы работать с большими данными нам нужны сложные инструменты, а еще не каждому это дано, скорее всего, потому что там же надо уметь кодить, и вот это вот все, что мы не любим. Какова тенденция? Тенденция такова, что сейчас идет переход к лоукод инструментам, это вот у социологов из SPSS, например, это то, где не надо уметь программировать, чтобы получить результат. С появлением таких инструментов, соответственно, повышается доступность аналитики данных как скилла, и аналитики данных как возможности, потому что, если раньше аналитиками, чаще всего опыт разумеет, что это технари, действительно, технари с хорошим техническим бэкграундом, мы же, например, всегда пропагандировали, что аналитик данных это прежде всего именно либо представитель соцгумов, в общем, это какой-то отраслевой представитель в связке с техническим. Почему? Потому что в анализе данных первичные задачи, первичные мышления, чем само решение. Лоукод в этом плане хорош, потому что именно таким людям, как ты, как я, те, которые не бродеры до мозга костей, условно говоря, не технари, мы можем по-другому смотреть, в принципе, на данные, мы можем в этом видеть тренды, и мы знаем, как поставить задачу. А когда нам дается такой инструмент, мы можем гораздо больше увидеть в данных свои призмы, и мы можем делать исследования, тот анализ, который нам не был доступен до. И, собственно, здесь, переходя немножко к рынку труда, когда инструменты анализа данных, в принципе, начали меняться, когда они становятся более доступными, соответственно, к аналитике труда начинают подступаться некоторые компании, группы людей, эксперты, исследователи, которые прежде не могли собирать эти данные, потому что у них не было навыка. Сейчас навык сильно не нужен, появляются нужные инструменты. Соответственно, с точки зрения аналитики рынка труда, сейчас будут как раз, идут такие изменения, что больше стали понимать, становится более возможно в нем копаться, больше видеть тенденции, находить, в принципе, эти тенденции, соответственно, мы можем больше полезного узнавать. Я знаю, что в ВНИИ труда сейчас с этим параллельно с нами занимается. Тоже люди, у которых вообще никогда не было раньше инструментов, сейчас они у них появились и классно они это используют. В том числе не только для исследования, но и, в конце концов, наверное, какая-то польза от этого будет.

В: Помимо умения кодить, что еще нужно уметь и знать, чтобы работать с большими данными?

О: Есть две группы навыков, даже нет, три группы навыков, как я это люблю называть. Это известный типа Hard, Soft и что-то еще, называют отраслевыми. Hard - это как правило мы подразумеваем, что человек умеет кодить, этот человек знает мат-статистику, он может тестировать, проверять гипотезы, ставить, проверять гипотезы. Soft - это то, что аналитика, как правило, считается аналитическим. Аналитическое мышление, коммуникабельность, умение работать в команде, еще умение нестандартно мыслить, не только аналитическим. Третий набор навыков, я бы сказала отраслевые. Это очень хорошее понимание отрасли, в которой ты собираешься работать и в которой вообще работаешь. То есть, аналитики, ну, мы любим считать, что есть универсальный солдат, даже их социолог раньше считался универсальным солдатом. Но проблема в том, что аналитик не будет хороший без знания о какой-то конкретной отрасли, в которой он специализируется. И даже мы, при том, что мы готовим универсалов первоначально, мы всегда объясняем, что мы даем инструменты и даем инструменты для постановки мышления, такого мышления, которое вам позволит разобраться в отрасли, в которой вы не спецы, именно потому что аналитик должен там разбираться. С этой точки зрения, тот ключевой soft, который принято считать софтом, аналитическое мышление, у человека, который работает с данными, это hard становится. Это прям конкретный навык, который нужно ставить, ставить жестко, потому что аналитиков это не просто умение там думать критически, это умение, мыслить моделями, это умение думать в некоторых, с одной стороны, в некоторых рамках, с другой стороны выходя за эти рамки. Вот. Ну и там, я забыла, в хардах же это еще наша любимая, собирает, обрабатывает данные, анализирует данные, хранит данные, вот это вот все, что считается. Нужно работать с данными, ну вот так.

В: Все, спасибо!

Интервью №4, Вячеслав, заведующий научно-исследовательской лаборатории прикладного анализа больших данных ТГУ

Длительность: 28 мин.

В: Расскажи про лабораторию, как вообще она появилась.

О: Неожиданный вопрос. В 2015 году наш университет участвовал в программе 5-100, по вхождению 5 российских университетов в 100 лучших. Решил сделать ставку на большие данные. В конце 2015 года появился Михаил Мягков, бывший вице-президент Сколтех и американский профессор. И он эту идею донес до руководства. И в апреле 2016 года появилась лаборатория прикладного анализа больших данных в Томском гос. университете. А лаборатория изначально имела интересную концепцию. Мы алгоритмы не разрабатывали, мы их применяли для решения социальных значимых задач. С того момента она ведет свою историю, лаборатория наша. Далее я туда пришел техническим специалистом. В 2018 стал заведующим.

В: Расскажи о каких-то самых крупных проектах, которыми занимается лаборатория.

О: Начнем с того, что мы с 2016 года считаем показатели Московского международного рейтинга университетов. Рейтинговое агентство нравится аналитикам. Далее мы с 2016 года делали проекты про абитуриентов. В 21 году, когда Мишустин приезжал в Томск, и в 22 этому проекту особое внимание осветили. И мы, Томский гос. университет за нее премию получили большую, мы использовали данные социальной сети для набора абитуриентов. Был большой проект по качеству жизни, вот Женя Щекотин с Дашей Дунаевой, мы начинали. Мы и для минсельхоза делали проект по качеству жизни населения сельских территорий. Сейчас он видоизменился, и мы изучаем, например, как люди достигают собственного счастья. Очень классный проект. Есть большой проект с фондом Потанина по целевым капиталам, мы три года делали внешнее финансирование. Есть сейчас у нас РосНавык небезызвестная история, тоже большой проект. Есть двухгодичный проект, мы исследовали настроения российского студенчества, в интересах Минобрнауки, в социальных сетях, то есть мы анализировали, чего мы писали в ковид и оказалось, что на основе этих данных можно вычлнять проблемы университетов. Чего еще такого? Еще ряд проектов, интересов, внутренних и внешних заказчик.

В: А что вообще такое Big Data? Как бы ты описал это?

О: Большие данные, у меня даже есть презентация, с которой я обычно выступаю, когда рассказываю про большие данные. Начнем с того, что многие считают, что это просто данные большого объёма. А распространенное, которое определение слышу, это данные, большие данные начинаются тогда, когда ты в Excel не можешь их открыть или обработать. Очень-очень, на самом деле, такое,

приземленное понятное понятие, но... А вообще-то четкого понимания нет. Большие данные - это набор не только самих данных, но и методов их обработки. То есть, некоторые технологии обработки данных, которые имеют три основных свойства. Они реально большого объема. Второй момент - это velocity, скорость. Скорость их обработок очень важна. Ну, например, ты подходишь с банковской картой, тебе надо, чтобы ты сразу оплатил, а не ждал определенного дня, пока придет транзакция. Надо очень быстро обрабатывать эти данные. И третий - это Variety. То есть, это разнообразие. Они очень разные. Это, например, могут быть аудиофайлы, это фото данные, изображения, тексты, комментарии, лайки, всевозможные активности и так далее.

В: А в чем заключается тип анализа таких данных? Как их анализировать?

О: Ну, есть набор методов. Они называются "Интеллектуальный анализ данных". В английском это "Data mining". Английское определение более правильное, потому что mining - это извлечение полезных ископаемых, то есть ты из набора данных извлекаешь полезную информацию. Как это сделать? Ну, может быть руками, конечно, делать. Открыть тот же Excel и пытаться что-то сделать. Но есть всевозможные количественные, качественные методы. Например, машинное обучение, математическая статистика, искусственный интеллект, нейронные сети, социальный сетевой анализ, SNA. На основном это методы либо математики.

В: А что нужно уметь, чтобы работать с данными большими?

О: Ну, смотря кем ты хочешь быть. Ну, вообще всем людям, сейчас мое личное мнение, нужно некоторый ликбез по данным, по их анализу, потому что данные нас окружают везде, начиная от того, что твоими данными могут воспользоваться в мошеннических случаях, ну, это когда деньги воруются как-то. Заканчивая тем, что в любой профессии теперь мы работаем с данными. Врачи, учителя, социологи и так далее все работают с данными. Эти данные часто цифровые. Плюс мы с цифровыми платформами. С одной стороны, все должны понимать, как вообще данные выглядят и как с ними, для каких целей можно использовать. С точки зрения людей, которые управленцы, им важно там понимать, какие управленческие задачи решать на основе данных. И как правильно поставить задачи, это очень важно. Те люди, которые данные анализируют, понятно, что они должны понимать саму задачу, и подобрать дизайн исследования либо дизайн проекта, который задачу заказчика перенесет на данные. И работать с базами данных. Соответственно, аналитик должен иметь аналитическое мышление, набор скиллов хардовых, как например обработка текстовой информации, не обязательно программирование, но понимание алгоритмизации, это презентация данных, визуализация данных, это работа с базами данных Ну и наверное вот.

В: А в чем заключается преимущество анализа больших данных перед другими методами?

О: Ну преимущество в том, что можно получить некоторые зависимости, которые даже в головы не приходят, и хороший пример это сеть американских гипермаркетов Target, они одни из первых начали использовать большие данные, вообще эти карты лояльности, которые в ленте используешь, все думают, что они для скидок, на самом деле они для сбора информации, о тебе, обо мне и так далее. А покупка? Зачем? Потому что коммерческие компании могут получить определенную дополнительную информацию о человеке и продать ему что-то, либо сделать то, чтобы ты дольше находился в магазине, это тоже очень важно. Либо ты приходишь за определенными скидками, либо сезон из товаров, очень много данных. И вот Target они сделали, они одни из первых начали использовать данные в 60-х годах. И есть кейс, когда 16-летней девушке пришел набор будущей мамы, и папа там поднял бучу, какая она будущая мамы, а потом долго извинялся, потому что она реально беременная и два ребенка. То есть магазин по ее проведению понял, что она беременная, ну, и соответственно может продать ей дополнительные услуги. То же самое касается всего маркетинга и рекламы, то есть это понимание того, чего ты хочешь, чем ты интересуешься. С точки зрения логистики, например, есть аналог "Почта России" USPS американских, Они, например, уменьшили количество поворотов на левом автомобиле, построили логистику так, чтобы поворачивать направо, каждый поворот налево - это сложная ситуация для водителей, и они существенно сократили количество аварий и своих затрат на эти аварии. Казалось бы. Томские коллеги компании "Рубикон" сделали оптимизацию пути следования скорой медицинской помощи, чтобы успеть к нужным больным вовремя, а всем остальным, которые могут подождать, они в следующий раз чуть позже придут. С точки зрения образования, кейс с абитуриентами, мы же за ними наблюдаем, они после третьего курса на 50% меньше отчисляются, которые вышли по нашей методологии, чем обычные. Что это значит? Мы угадываем их обывательский интерес. Человек попадает туда, куда он реально хочет, а не туда, куда его привели до родителей, либо там, где баллы поменьше и так далее. То есть большие данные дают, на самом деле, колоссальные возможности понимания, во-первых, в сути объекта, его поведения, либо явления, и, например, некоторую прогнозирующую модель.

Например, комбайн, обычно в каждом фермерском хозяйстве есть несколько комбайнов, вот комбайн сломался, и ты разорился, поэтому тебе бы по-хорошему знать, что у него сломается. Современный комбайн, даже российский, вот Сельмаш, это 70 наборов датчиков. Мы работали с Омским государственным аграрным университетом, где я немножко больше узнал о комбайнах. И эти данные можно анализировать, и тебе датчики скажут, что скорее всего этим летом у тебя выйдут из строя вот эти элементы. Как ты это узнаешь? Потому что у тебя другие комбайны собирают, и с них эта информация, как они ломаются, нужно много наблюдений. Поэтому большие данные, они работают, когда реально данных много, либо наблюдений очень много, зависит от этого. Вот Яндекс.Музыка, казалось, все люди говорят, у меня уникальный музыкальный вкус, оказалось, что около 60-70 типов музыкальных, поэтому вся наша уникальность и да бог раскладывается. Вот.

В: А какие недостатки у метода?

О: Ну, во-первых, недостатки сугубо этического характера. Можно понять, например, что Facebook, секунду, и Instagram собирают данные в Wi-Fi сетях. Зачем? А вот если мы с тобой в единой комнате сидим, то у нас список Wi-Fi сетей будут одинаковыми. Значит, почему-то эти два пользователя находятся вместе, а если они очень часто постоянно вместе находятся, то тоже можно некоторые выводы сделать по поводу этих двух людей, и их использовать. Можно использовать как в целях тебе что-то продать или нет, но можно использовать достаточно интересные цели, политические предпочтения моделировать, манипулировать общественным мнением, ну, все возможности фейк-ньюс и так далее. Но такое большое оружие в руках определенных людей в том числе, поэтому здесь плюс мы нарушаем некую... То есть мы можем не спрашивать тебя, а к себе данные собирать, и это слабо регулируется с законом сейчас. Поэтому риски очень высокие именно такого этического характера, когда люди с не очень чистыми помыслами могут эту информацию использовать.

В: А еще какие-то недостатки есть, например, в работе с такими данными?

О: Я бы не сказал, что недостатки. Но во-первых, данные надо где-то брать, большой вопрос, а где данные взять? Второй момент – эти данные надо где-то обрабатывать, нужно использовать специфическое железо, сервера и так далее. Это дорого. Надо некоторые умения иметь, то есть должна быть команда, один человек должен себе тогда совмещать и аналитика, и айтишника, и еще лучше заказчика. Это сложная история, поэтому да, я бы сказал, что это минус. Это некоторые особенности работы с этими данными.

В: А если переходить к рынку труда, зачем нужен мониторинг рынка труда?

О: Это очень простой ответ для того, чтобы когда ты закончишь университет, ты поступил на социолога, если я не ошибаюсь, а знаешь, сколько социологов нужно в России каждый год?

В: Мало.

О: 500-600 человек. А знаешь, сколько выпусков социологов в год?

В: Много.

О: 4000 человек, и видимо, если бы ты эту информацию узнала чуть раньше, возможно, ты бы немножко подумала, например. Но в то же время, если ему дать знания работе с данными, он уже аналитиком становится, аналитиков требуется от 17 до 50 тысяч ежегодно, то есть уже недостаток специалистов. Поэтому в первую очередь, мониторить рынок труда для обычного человека нужно, чтобы понять, где тебе работать, какие перспективы самореализации и так далее. Для образовательного учреждения – это чему учить. Рынок меняется. Для работодателя получения именно того выпускника, который что-то умеет, не нужно тебе его доучивать. Ну а для экономики целыми общества – это очень важная история связана с развитием общества, потому что мы же хотим, чтобы у нас было меньше рутинной работы, больше интеллектуальной, творческой и так далее. Ну тогда надо эффективность труда повышать и так далее. Ну а так, в принципе, это очень классная карьерная перспектива и понимание того, что в твоём регионе у тебя по этой специальности крайне низкие перспективы, а вот то если ты переедешь, чтобы от этих перспектив, право или лево, открываются другие перспективы. Ну то есть в этом отношении для человека, мне кажется, тоже очень важно понимать, кем быть, чтобы быть, чтобы было более высокое качество жизни.

В: А как рынок труда меняется и влияют ли эти изменения на его анализ?

О: Ну меняется офигеть как. Приведем пример пандемии, ну вот, подрабатывал ты официантом, например, случилась пандемия и твой ресторан закрылся. В таком положении оказалось огромное количество студентов, и они там решились подработать, рынок труда сильно изменился, вообще в пандемию, зато тут усиливалась доставка. В принципе доставка еды и продуктов активно развивалась у них в пандемию. Онлайн образование, рынок труда может под влиянием внешних фактов меняться очень сильно. Это первый момент. Второй момент, ну он сначала спецоперации с нашими айтишниками, то есть такие политические процессы тоже сильно влияют, но не только политические,

еще влияют, ну как бы, эти наши развития в принципе общества. Представь завтра Яндекс, ну как бы, изобретает крутой беспилотник, он же всех таксистов после завтра уволит, и на рынок выйдут, например, несколько миллионов человек, которым нечего делать. Это огромные социальные потрясения для общества, для государства и так далее. Все действительно надо поработать, надо переучивать. То есть это высокие риски и возможности. То есть вообще в принципе может быть в далеком будущем работать нам нужно будет три дня в неделю, а все остальное будет делать робот, операции, а ты будешь заниматься творчеством и так далее. То есть есть некоторые профессии будущего, сейчас, например, с развитием интернет-технологии, я последнюю статистику смотрел, не помню какое агентство, они по онлайн образованию собирали, то есть на 23% сжался рынок онлайн образования в области IT, потому что в журналах очень много вышло народа. И если год назад ты могла пойти курсы по программированию и достаточно быстро найти рабочее место, теперь это почти нереально. Зато почти в 4 раза подскочил спрос на монтажеров, колосеров, дизайнеров, кто работает с контентом, потому что контента очень много генерируется. Блогеров очень много. В мероприятиях, которые в онлайн работают, очень много и нужны люди в рейтинге. Поэтому отслеживание рынка труда это очень важно.

В: А вот изменения, которые влияют на рынок труда, влияют ли на... как они влияют на анализ его?

О: Они усуждают анализ. То есть, нам эти изменения надо фиксировать. Плюс, ладно, изменения у нас... Где университеты публикуют свои вакансии? Как ты думаешь?

В: Университеты?

О: Да, где университеты, где можно занять вакансию младшим научным сотрудником.

В: Каких-то университетских платформах, чатах?

О: Да, университетских центрах. Понятно, что нет, ну, практически нет на хедхантере, не на суперджобе и так далее. Это вот туда. Ну, как бы это особенности рынка труда. Изменения рынка труда, но на аналитике, как мы эти изменения должны фиксировать. Это некоторые вызовы. Иначе наша аналитика будет мало кому нужна. С прогнозированием крайне сложно, потому что... Это некоторый кот в мешке, который нужно поймать, мало того, еще посмотреть, что там люди. Например, надо смотреть инвестиции в те или иные отрасли, куча огромных внешних фактов и так далее. Поэтому изменения, да, в общем и целом, оно существенно усложняет анализ. Но если бы он не менялся, то есть, какой смысл в этом анализе?

В: Ну да. Так, а если кратко рассказать про РосНавык? Зачем он, для кого, какие цели?

Ну, во-первых, это суперзамечательный продукт, мы его придумали не просто так, мы сначала его сделали сами для себя. У нас была программа, введение в аналитику данных в профессии, аналитик данных, начало карьеры, что-то такое, было в демографии, мы учили кучу людей, мы-то очень умные, мы считали, что мы знаем, чему учить, а потом, так как мы аналитики, мы выгрузили хедхантера вакансии аналитики, оказалось, что некоторые ставки наши не сыграли, надо было заменить, мы заменили и трудоустраиваемость наших выпускников на 30% повысилась. Мы почесали голову, подумали, блин, если это для нас так работает, почему не сделать сервис удобный для других. Мы провели фокус группы. Оказалось, что руководители образовательных программ вообще практически не используют данные, именно данные рынка труда, для составления образовательных программ. То есть у нас гигантский разрыв между тем, чему учат, и тем, что надо делать на рабочем месте. И поэтому здесь основная проблема, на которую направлен продукт, это как раз снижение разрыва, для всех целевых аудиторий, где студент получает более качественное образование и понимание, куда ему идти, профориентацию и так далее. Университеты и образовательные учреждения получают возможность, именно возможность, сделать или нет, они сделают уже на совести каждого. Сделать образовательную программу такой, которая будет получать запрос на рынке труда. Работодатель на выходе получает более крутого специалиста.

Есть ли у РосНавыка какие-то ограничения?

О: Их полно. Причем с того, что сейчас мы используем только одну платформу. Мы в скором будущем это исправим, мы работаем только на Headhunter, на данных Headhunter. Это далеко не весь рынок, мы добавим туда SuperJob, работа в России, и может быть Avito, тогда да, это станет больше. А если бы еще мы добавили, например, вакансии, которые публикуются на сайтах, ну, например, организации таких как университетов, то у нас еще больше будет. Понятно, что есть некоторая серая зона занятости, которую мы не сможем таким образом померить. Но в любом случае это очень хороший инструмент. Это первый. Второй момент, мы же там навыки извлекаем из вакансий, есть некоторая погрешность, и некоторая недостаточная глубина. Например, с точки зрения погрешности, мы думали автоматом все это изменить, но оказалось, что для алгоритмов ремонт двигателей

внутреннего сгорания, ремонт сотовых телефонов – это один и тот же навык, на мой взгляд, немножко разный навык.

Плюс вот работаем с коллегами, они говорят, у вас сварщики везде кучи, а нам нужны сварщики по титану, нам обычные сварщики не нужны. Тоже запрессификация. То есть в этом отношении нам, конечно, нужно будет некоторые отрасли доуточнять, для того чтобы конкретные, например, образовательные учреждения получили возможность получать аналитику именно по своему правлению.

А в чем преимущество анализа больших данных в сравнении с другими методами именно в рынке труда?

О: Ну а какие другие методы? Другие методы – это в основном опросные методы. Вот я просил 50 работодателей, а через данные Headhunter'a опрашиваются, примерно 500 000 работодателей сразу. И мы их опрашиваем, причем они голосуют деньгами, вакансии размещают за деньги. Самое простое, это то что они большие, это обмен информации и некоторых средств, плюс мы смотрим так как оно есть, то есть ты спрашиваешь, например, вот меня, задаешь конкретные вопросы, а я может в реальной жизни на эти вопросы никогда бы не ответил, а большие данные – они так называют нереативные данные. То есть я пишу что-то в соцсетях, потому что я хочу написать, либо у меня спросили, либо мы это обсуждаем. А например, в количественных методах или качественных ты меня спрашиваешь конкретный вопрос и это искажает серьезно картину реальную. Ну, и плюс, это же, например, можем из Томска смотреть сразу на всю Россию, отличие онлайн-данных. Ну, и плюс скорость, у нас каждый день грузятся десятки тысяч вакансий, мы их каждый день качаем, у нас каждый день быстрый анализ, скорость обработки.

В: А есть ли недостатки в сравнении с другими методами?

О: Ну опять же точность, ты можешь пойти экскурсоводов найти, у нас их просто физически может не быть, а ты можешь найти 10 компаний, где они нужны, если тебе нужен очень точный запрос, то учреждения можно найти, то есть это некоторый молоток, которым можно много гвоздей забивать, то есть если тебе нужно забить какой-то большой гвоздь - тебе нужен большой молоток, у нас его может не быть, или наоборот, может быть очень маленький гвоздик, тебе нужен какой-нибудь фигурный молоток и у нас его нет. Но в целом забивать можно, поэтому в этом отношении функционал, он конечно ограничен, но для небольших, но важных задач его достаточно.

В: А чем РосНавык отличается от других платформ для мониторинга рынка труда?

О: Мы проводили анализ, я думаю, ты с Дашей беседовала, она лучше меня об этом расскажет, на самом деле, только у хедхантера есть похожая история, но устроена в логике работодателя, то есть мы то представляем университетское сообщество, наша такая великая миссия - это чтобы образовательные учреждения делали качественный образовательный продукт, образование дает шанс будущему и надо, чтобы это образование имело некоторую аналитику того, чему надо учить, что там востребовано на рынке, понятно, что история инноваций, вот Стив Джобс изобрел айфон, новую индустрию построил, это, наверное, предсказать было тяжело, но когда мы говорим про массовые специальности, здесь без аналитики обойтись нельзя. На Западе есть подобные платформы, которые используют данные тоже, агрегаторы, поэтому здесь мы не едины, но на территории Российской Федерации подобного решения для организации образования просто нет.

В: Как будут развиваться большие данные и как это скажется на аналитике рынка труда?

О: Начнем с того, что данных становится больше, это факт, и если эти данные будут также открыты, то мы сможем более точно делать аналитику и предсказывать, это важно для развития и управлять этим проще. С точки зрения обработки, алгоритмы совершенствуются, тот же чат GPT показал, что нейронные сети шагнули вперед, то есть мы можем более точно выделять, мы сейчас это уже пробуем делать, или определять те или иные навыки в вакансиях. С точки зрения вообще рынка труда, ну например, было бы круто еще, у нас были проекты, я про них забыл, на определение психологических характеристик у человека. Например, есть надпрофессиональные навыки - Soft-skill, да, и для некоторых профессий, ну например, я вот никогда не смог бы работать кассиром в магазине, я ненавижу монотонную работу, я просто не способен к ней, и если я пойду кассиром - я с ума сойду, либо я там работать не буду, учитывать мои особенности личностные, например, фантазируем, для той или иной работы. Какие-нибудь дизайнеры требуют высокого уровня креативности, и к тому же рынок меняется, люди меняются, у нас в проекте содействие занятости очень много пенсионеров, которые хотят изменить свою жизнь, у нас есть кейсы интересные, когда бухгалтер, женщина 58 лет в Санкт-Петербурге, устроилась в компанию, одна мама, например, в декрете, прошла обучение и добилась продвижения в аналитическом отделе. Люди... Нужно, во-первых, понять их мотивацию, понять насколько они вообще способны к той или иной профессии, есть профессии, которые требуют

определенного набора личностных характеристик, и если мы, например, побольше узнаем о соискателе, то мы смогли бы, но опять же, есть риски, что с детства, например, ребенку говорят “ты будешь инженером”, “ты вот, Вася, будешь каменщик”. Сейчас вот профессиональных спортсменов очень часто тестируют для генетических выявлений, для большого спорта - это же соревнования, и для подходов к тренировке. И вот связывая с рынком труда, с одной стороны, технологии развиваются и можно выявлять характеристики соискателей.

В: А как это узнавать о соискателе?

О: Ну например, в центре занятости ты заполняешь анкету профориентационную, более того ты же резюме составляешь, и в резюме пишешь, есть свой стиль написания, есть некоторые ключевые слова и так далее. Это можно использовать.

В: Поняла. Это всё, спасибо!

Интервью №5, Полина, аналитик лаборатории прикладного анализа больших данных ТГУ

Длительность: 1 час.

В: Так, давай начнём с того, кем ты работаешь и какими проектами занимаешься.

О: Так, меня зовут Полина. Я работаю в центре прикладного анализа больших данных в Томском государственном университете. Какими проектами я занимаюсь? Я занимаюсь проектом, который связан с разработкой сервиса для аналитики и мониторинга рынка труда в России.

В: Так, а как давно ты работаешь?

О: Я работаю здесь с августа 2022 года, ну почти год.

В: Вот ты работаешь с большими данными. Расскажи, что это такое.

Ох, когда-то кто-то, наверное, Слава - это руководитель нашей лаборатории, по-моему, это Слава, ну в общем, кто-то когда-то сказал мне, что большие данные - это то, что не помещается в Excel. В Excel доступна, ну то есть там реально ограниченная число строчек, оно не бесконечное, и оно ограничено миллионами 119 тысячами и каким-то еще небольшим количеством, сотен десятков, ну и одиннадцать, ну там типа миллион сто, двенадцать тысяч, столько-то там строк. И вот все, что не помещается в Excel - это большие данные. Большие данные - это информация и информация большого объема. Мы сейчас живем в цифровом обществе и согласно исследованиям, если честно, я не помню у кого, у нас сейчас информация в обществе растет по экспонентам. То есть каждый день количество информации удваивается из дня в день, ну то есть каждый день становится больше в два раза, чем в предыдущий день. То есть это очень большое количество информации, это информация обо всем буквально. Я не знаю, вот ты едешь на такси из дома в университете, и у тебя есть информация о цене этой поездки, о протяженности этой поездки, о том, сколько минут она шла, о том, сколько минут ты ждала, пока тебе придет такси. То есть и все это данные, и их очень много, а это только одна твоя поездка на такси, таких поездок ежедневно совершается, до Томского государственного университета сотни тысяч, десятки тысяч, а для других университетов миллиард, ну и это только такси. И соответственно данных очень много. И их данные, большие данные, это, собственно, большое количество какой-то информации, вот так.

В: А кроме большого объема эти данные обладают какими-то особенностями?

О: Ой, да, там есть четыре каких-то основных критерия, описание этих данных... Блин, я, к сожалению, не помню, я тебе не скажу, я с теоретической точки зрения редко на это гляжу.

В: Так, а как работать с такими данными?

О: Ну, на сегодняшний день есть множество новых инструментов, то есть инструменты, которые не нужны, чтобы работать и знать какие-то такие программирования для работы с этими данными. Но вообще, изначально, и всецело, и сам вот этот термин дата аналитика, дата сейтисты, дата анализ, он предполагает работу с базами данных, предполагает работу с базами данных при помощи Python и запросов в SQL. Но на сегодняшний день существует масса различных новых инструментов, при помощи которых можно обрабатывать подобные объемы информации. Это программное обеспечение, например, разработанное компанией Megaruter, российского разработчика программного обеспечения. У них есть программный пакет, он называется Polyanalyst, и это инструмент, направленный на аналитику больших данных. То есть он обрабатывает большую массивную информацию, позволяет строить какие-то ампликации, все такое. То есть условно все функции, которые ты можешь прописать в Python или SQL, там сделаны в виде графического интерфейса, и тебе не нужно ничего самому писать, ты можешь просто накладывать галочки, и все расстроится, посчитается, будет классно. Еще есть из инструментов Orange. Это не чья-то конкретная

программа, наверное, это чья-то конкретная, но я просто не знаю чья. У него открытый код, и туда люди могут, если я правильно помню, дописывать различные новые выписки. То есть если в Polyanalyst функционал ограничен, и ограничен он ровно тем, что сделали разработчики, то в Orange открытый код, и пользователи самостоятельно могут дописывать и дорабатывать новые функции, делиться ими с друзьями, в смысле с сообществом пользователей других кодов. Что еще из инструментов можно использовать? Обработка информации же, Big Data, информация, она представлены не только в табличном виде, не только в структуре, но и в графическом виде. Графы, например, сетевые штуки, когда есть 2-3 тысячи точек, они между собой связаны с помощью ребер. Для обработки графов используется тоже non-bot программное обеспечение - Gephi. Но графы также можно строить и обрабатывать без использования non-bot инструментов, просто при помощи языков программирования.

В: А кроме знания таких программ, что еще надо уметь, чтобы работать с большими данными?

О: Я по образованию социолог, и нас как социологов учили, что такое мат-статистика, что такое обработка данных, как они между собой связаны. Я понимаю основы мат-статистики. Я знаю, что такое среднее, я знаю формулу, я знаю, что это Хи квадрат, я не просто могу поставить галочку в Хи квадрат и сделать какой-то вывод, но технически, если мне ее показать, я вспомню и смогу посчитать это руками, не только при помощи программного обеспечения. Я бы сказала, что нужно знать мат-статистику. Еще для работы с большими данными нужно знать алгоритмизацию. Это вся фигня, которую в школе на информатике учат, где алгоритмы делаете в 8 классе, когда у тебя окошечко, стрелочка, окошечко, потому что на самом деле это больше важно для программирования, нежели для анализов и так далее. Но если мы говорим об анализе через языки программирования, то есть через обращение к базам данных или через написание запросов, то там важно понимать алгоритм, как выглядит код, как структура кода, что цикл у тебя, если начинается какой-нибудь цикл, что он обязательно должен закончиться, что вот это один кусочек кода, там перед ним будет другой кусочек кода, и что они должны идти именно в таком порядке и так далее. Что еще нужно знать? Что еще? Наверное, для того чтобы, знаешь, обрабатывать данные, можно использовать что-нибудь, какие-нибудь софты, типа там, не знаю, логического мышления, технического склада ума, что-нибудь такое, потому что на самом деле я не раз сталкивалась с этим, правда, не в этом продукте, а просто когда социолог работал и, не знаю, когда диплом писала, я пыталась одноклассникам очень много статанализ объяснять, когда к ГОСам готовились, и у всех были беды с методологией, и в группе вообще никто не знает, как считать количественные данные. Это большая проблема у моей группы, конечно. Но я тогда, когда готовилась к ГОСу, я объясняла в группе, как работает какой-то метод, стат-критерии и так далее. И да, я сталкивалась с тем, что есть люди, которые реально не вообще... Это проблема не во мне, честно, в том, как я доношу информацию. Я пыталась разными путями, более того, это пытались сделать другие люди, то есть там потом Дина Олегова помогала, что-то объясняла, еще что-то. Нет, есть просто люди, которым не дается счет, не дается математика, не дается логика.

Ну, I don't know. Им просто очень тяжело. Я не говорю, что это плохо. Я говорю, что для работы с цифрами, не только с большими данными, с любыми цифрами нужно обладать каким-то таким счетно-логическим мышлением, я думаю. Ну и для расчетов, наверное, больше всего ничего не надо. Но я бы, возможно, добавила навык визуализации данных. На мой взгляд, это очень важный навык, которого у многих, да, аналитиков, сайентистов, в принципе, аналитиков, не знаю, социологов, математиков, множество, да, у множества людей этого навыка нет. Мне кажется, это очень важный навык, уметь представить данные и результаты своих исследований в графическом виде, причем так, чтобы это было, ну, фиг с ним красиво или нет, но чтобы это было хотя бы понятно, то есть чтобы ты мог посмотреть на график или там на диаграмму или на, не знаю, точечное рассеяние, как там, диаграмму рассеянных точек, боксплоты, ну, я не знаю, на любой, на любой тип графика что-то мог посмотреть и сделать из него, ну, какой-то вот, хотя бы наглядный. То есть там вот это больше, вот этого, и то, что больше, должно стоять сверху, а потом оно вот так должно было бы идти, потому что ты себе поможешь, не, ты бы наверх помещаешь, ну, было бы, что-то видно. Я не знаю, ну, в идеале, чтобы это еще красиво было, но красиво это уже дело десятое, но чтобы хотя бы было видно что-то на этих графиках.

В: Ага

О: Наверное, это все.

В: А если переходить к самому анализу данных, то в чем преимущество такого метода перед другими методами социологии?

О: А ты имеешь в виду...

В: Ну, чем анализ больших данных лучше других методов?

О: Он не лучший. Я не считаю, что он лучше.

В: Так, но есть ли у него какие-то преимущества в сравнении с другими методами, вот так?

О: Не знаю. Ну, как бы как. Вообще есть теория, недавно кто-то написал статью, я не помню даже кто, но если хочешь, скажу тебе. В общем, был чувак, который написал статью о том, что это классическая социология постепенно умирает. Я пока, на сегодняшний день, я не готова согласиться. Я не считаю, что классическая социология пока что умирает. Вот. Но в целом, правда в этом тоже есть. Методы анализа больших данных? Ну тут на самом деле не в больших данных дело, а скорее просто в анализе медиа. То есть, ну да так, просто данные, которые содержатся в, не знаю, в социальных сетях, в медиа, еще где-то, это не всегда прям большие данные. Они ассоциируются с большими данными, потому что они в интернете, но если ты анализируешь паблик, в котором, я не знаю, 900 постов при помощи контента анализа, это нельзя назвать анализом больших данных. Ну объективно. Так вот, я говорю про то, что анализ интернет пространства - это довольно перспективная история. Но я бы не сказала, что на сегодняшний день это как-то сильно выигрывает перед классическими методами социологии. Надо понимать, что пока что, пока что все методы, которые мы можем применять в интернет-пространство, это количественные методы. То есть мы не можем залезть в какой-то сверхмега глубинный смысл. Нет, мы можем использовать дискурс анализ, безусловно. То есть мы можем там вычитывать тексты, искать в них смысл. Искать в них смысл, еще такое. Но, преимущественно, на сегодняшний день это все же количественные. Это первое. Это то, что пока нельзя, например, сравнить. То есть нельзя сказать, что анализ интернет-пространства выигрывает перед классическими методами, потому что нельзя сравнить, потому что нет возможности. Если говорить про количественные методы, я бы тоже не сказала, что нет выигрыша или проигрыша. Мне кажется, что метод нужно выбирать, исходя из того, что ты хочешь достичь, исходя из целей и задач твоего анализа. Если у тебя есть задача, которая решается через разговор с людьми, решай ее через разговор с людьми. Если у тебя есть задача, которая решается через просмотр каких-то объявлений или комментарии в каком-то объявлении то иди собирай комментарии и смотри, что люди пишут. В этом плане на конференции в сентябре или октябре, в нашей лабе была конференция. Там был дядька, который сказал, что сегодня интернет - это жалобная книга. Из прикольного, люди в интернете очень любят жаловаться. У меня есть подруга, которая сейчас работает в центре городской среды, развитие чего-то там. Одним из методов исследования в городском пространстве был анализ комментариев с Google cards, 2ГИСа. И надо понимать, что очень много комментариев негативные. Это не потому, что место плохое, потому что когда люди говорят, что место хорошее, то в комментариях пишут обидевшиеся. С другой стороны, надо понимать, что сидят в этом пространстве люди, которым пространство нравится, и это тоже дает некоторое искажение. Мне кажется, что эти методы имеют и те, и другие, и классические социологические методы, и методы анализа интернет пространства имеют место быть.

Имеют место быть как в отдельности, так и в совокупности. Нужно понимать, для чего ты это используешь. Почему ты используешь именно это, а не классическую социологию. Почему ты используешь именно классические интервью, фокус группы, а не контент анализ или сетевой анализ из какого-то публика.

В: Есть какие-то еще недостатки больших данных?

О: О да. Я очень часто сталкиваюсь с этим, когда рассказываю где-то о своих исследованиях. Обычно у меня очень часто спрашивают, с чего вы вообще решили, что ваши данные реальные. С чего вы вообще решили, что вы можете делать какие-то выводы на основании этого. Дело в том, что в лаборатории по работе я занимаюсь проектом, связанным с монитором данных рынка труда, но мои личные исследовательские интересы в основном крутятся вокруг анализа социальных сетей, и поэтому все исследования, с которыми я выступаю на конференциях, это, как правило, про анализ сетей. У меня очень часто спрашивают, почему вы думаете, что распределение по возрасту, которое вы составили, это реальное распределение? Я говорю, что люди указали такой возраст. И они такие, ну вообще это все не так, это все стопроцентно. Вы все придумали, потому что в соцсетях все могут сделать. Все мы в контакте можем быть 98-летней школьницей из Токио. Каждый из нас может это сделать. Но мне кажется, что последнее время, последние 10 лет, наверное, общество постепенно приходит к тому, чтобы начать вести свои социальные сети плюс-минус адекватно. Там уже значительно меньше школьниц из Японии, чем в 2010-2011. Да, ну и вот такие школьницы из Японии, которым 98, я, как правило, считаю их за выбросы. И часть данных, когда я провожу свои исследования, я все же проверяю. То есть часть данных я протыкиваю руками, смотрю, что там

происходит и так далее. Да, ну очень много вопросов, валидности и репрезентативности результатов, которые я получаю. Потому что интернет - что-то такое, в интернете каждый может быть тем, кем ему хочется, писать что хочет, независимо от того, было это или нет. Вот так, наверное.

Если переходить к рынку труда, зачем вообще нужен анализ рынка труда?

О: Вообще, это, в целом, это очень большой пласт жизни каждого человека. То есть есть три основных вида в России, есть исследование, и вот эту тетю я помню, Елена Васильевна Глушко ее зовут. Она выделяла три типа в России. Это бытовые практики, досуговые практики и трудовые практики. То есть она условно считает, что всю нашу всю дневную жизнь можно поделить на три больших части - быт, досуд, труд. То есть это как минимум 30% повседневности каждого человека. Для чего нужен рынок труда? Для того, чтобы понимать... Нет, ладно. Я могу... Я могу зайти с позиции... с позиции, с которыми мы заходим в лабораторию. Мы в нашей лаборатории, да? Ну нет, это и так, на самом деле. Блин, подожди, ладно, я подумаю, мне нужно время, чтобы... Ладно, у меня есть, наверное, несколько ответов. Я даже не знаю, с какой позиции к этому подойти. Мониторинг рынка труда нужен много для чего, он нужен преподавателем, он нужен соискателем, он нужен ученым, он нужен политикам. Он нужен агрегаторам по поиску вакансий. Он нужен людям по разным причинам. Не знаю, абитуриентам он нужен для того, чтобы понять, куда поступать. Очень много абитуриентов стремятся поступать в те направления, на которых много платят, в перспективе. То есть не только потому, что вам папа сказал, не только потому, что вам что-то очень нравится, потому что многие после школы вообще не знают, что им нравится. И идут по принципу, ну, в IT много платят, нефтянке тоже, пойдем на таможню, не знаю. Работодателям он нужен для того, чтобы понимать конкурентный рынок. То есть ты условно смотришь, что происходит у твоих конкурентов, и от этого выстраиваешь собственную стратегию, например, поиска сотрудников. То есть ты знаешь, что другие компании в твоей отрасли платят сотрудникам, ну, вот, не знаю, платят уборщикам 40 тысяч рублей, а ты платишь уборщику 20 тысяч рублей, и вот сидишь и думаешь, что что же я два месяца никого не могу найти к себе уборщиком. Ну, ты посмотри, сколько люди платят денег уборщикам и соответствуй. Потому что на самом деле сейчас Россия переживает довольно сильный кадровый голод. Особенно сильно это ощущают у себя рабочие специальности типа заводов. То есть на заводах сейчас очень сильный недостаток кадров. И заводы сейчас буквально борются за каждого человека, которого они могут к себе поставить. Они стараются максимально много дать им бенефитов, максимально много позитивного настроения. То есть они стараются максимально много дать сотрудникам. И когда ты знаешь, что другие люди дают сотрудникам в этой сфере, ты можешь увеличить свою конкурентоспособность в плане как работать. Потому что государству нужно знать, куда вкладывать деньги. Какие отрасли у нас в России экономически прибыльные, какие неприбыльные, каким предприятиям нужно увеличить господдержку, каким-то сократить господдержку. Плюс надо понимать, что современная система образования такая большая, тяжелая и неповоротливая, менять ее сложно, и обычно меняют ее только тогда, когда все становится уже совсем плохо. Это та история, когда машинистки, профессия, они раньше были, потом в какой-то момент их заменили. Когда профессия машинистки померла, наверное потому что появились компьютеры, возможно, когда профессия машинистки померла, еще в какое-то количество времени на машинисток учили учебных заведениях. Можно было учиться, когда профессия машинистки померла. И для того, чтобы вовремя отслеживать тренды и обучение, для того чтобы люди не учились на графического дизайнера на скиллбоксе, а могли учиться на графического дизайнера в Томском государственном университете. Нужно понимать, что графические дизайнеры в Томске требуются недавно, плюс надо разработать программу для этих графических дизайнеров, попробовать связаться с компаниями, которые эти графические дизайнеры нанимают или выпускают, чтобы понять, как это работает, что-то новое, страшное. Нужно понимать, как учить, почему учить, для чего учить и куда распределять бюджет. Это я только вижу в этом случае. Может быть, журналистам тоже, чтобы писать какие-нибудь статьи, типа "У, в России безработица", "У, кошмар". Ну да, наверное, как-то так.

Так, хорошо. А теперь скажи про ваш сервис по аналитике рынка труда, какие цели у него.

О: Да, я справлюсь с этой задачей, как никто другой. Я люблю наш сервис. В общем, сервис, который мы разработали, направлен на аналитику и мониторинг рынка труда в Российской Федерации. На сегодняшний день в Российской Федерации у нас нет конкурентов. Это связано с тем, что специфика нашего сервиса заключается в анализе требуемых на рынке труда навыков и компетенций. То есть сервисы по мониторингу рынка труда в целом есть. Они есть у HeadHunter, у SuperJob, у Работы в России, есть какая-то своя история. Есть так называемая ВМ-карта, это разработка университета 2035. И все эти сервисы в целом позволяют анализировать рынок труда, но они не позволяют

анализировать компетенции, которые требуют на рынке. А это как раз то, что мега важно для системы образования. То есть для того, чтобы вовремя менять какие-то образовательные программы, необходимо понимать, что от сотрудников требуют на рынке труда. То есть условно, не знаю, условный, в том числе государственный университет, выпускает там ежегодно 10 агрономов, но почему-то эти 10 агрономов не могут трудоустроиться. Почему-то их не берут ни в одну компанию, и для того, чтобы так не происходило, нужно понять, почему эти агрономов не берут их, скорее всего, потому что они не обладают каким-то необходимым для агрономной сферы навыков. Нужно проанализировать агрономную сферу, посмотреть, какие навыки требуются, требуются в ней сейчас, и, соответственно, перестроить программу так, чтобы все навыки, которые требуют от агрономов на сегодняшний день, чтобы они все попадали в образовательную программу условного агронома условно в Томском государственном университете. Так вот наш сервис - замечателен и примечателен именно тем, что он позволяет анализировать навыки и компетенции. Ну и, собственно, да, поэтому у нас нет конкурентов... В России все еще нет такого сервиса, кроме нашего. Такие сервисы есть за рубежом. И за рубежом они активно применяются, причем применяются не только с точки зрения образования, но в том числе и HR различных компаний для поисков сотрудников, для привлечения сотрудников, ну да, но и соискателями, то есть школьниками для профориентирования, для домашней информации, чему они хотят поучиться, где они хотят поучиться, в какие компании они могут пойти, зная, ну да, условно, имея вот этот навык. Вот. И за рубежом это активно применяется, и уже наперехват. Но зарубежные сервисы не позволяют анализировать российский рынок. Мы пытались, мы на всех зарегистрировались, нигде невозможно выбрать регион Россию, и вот мы решили, что нам такое тоже надо, и разработали сервис. Но в самом деле разработали его по другой причине, потому что год назад в рамках федеральной программы, федерального проекта "Демография" нужно было научить энное количество людей, аналитики, дата аналитики. И в нашей лаборатории мы такие сели, подумали, о чему же учить людей, что вообще нужно знать аналитикам. Мы-то много чего знаем, мы еще из разных сфер сюда пришли. А что нужно знать дата аналитикам, которых учат для дата-аналитики. Сели, подумали, выгрузили данные с HeadHunter просто по аналитикам, потыкали, посмотрели, поняли, почему надо научить аналитиков, чтобы они были востребованы на рынке труда. Собрали под это образовательную программу, отучили дата аналитиков, и они там все, ну почти все, ну просто 90 с лишним процентов, они трудоустроились. И мы такие, вау, это работает, и решили, что другим руководителям и разработчикам образовательных программ такая штука будет супер полезна, потому что нам оказалась супер полезна. И мы решили сделать сервис, который будет анализировать рынок труда в реальном времени, то есть, ну, чтобы смотреть ситуацию на рынке на сегодняшний день. И чтобы в этом сервисе можно было смотреть много разных штук, начиная там от зарплаты, требуемого опыта, региональной специфики, потому что Россия большая, везде от всех требует разное. Там, не знаю, заканчивая вот какие навыками и компетенциями, тем, какие требуют определенных специалистов, тем, какие требуют пары, связи между собой, с какими компетенциями туда можно идти устраиваться, в какую компанию, в какую должность. На сегодняшний день нашего сервиса есть сайт и личный кабинет, в который можно зайти в ссылке gosnavyk.ru. И, зарегистрировавшись в личный кабинетик, подтвердив свою электронную почту, можно попасть в наш сервис. Там сегодняшний день есть две таких больших вкладки. Первая связана с дашбордом. У нас есть дашборд, это интерактивная аналитика, то есть графики, цифры, картинки, все в одном месте. Там можно потыкать галочки, выбрать определенные специальности или регионы, и автоматически дашборд перестроится, и там будет аналитика по определенным регионам или специальностям в зависимости от того, что ты выберешь. И вторая довольно крупная вкладка - данные. Она позволяет получить доступ к нашей базе данных. О боже, я не рассказывала про данные. Вся наша аналитика на сегодняшний день построена на данных в агрегаторах вакансии HeadHuntera. Это самый крупный агрегатор вакансий в России. В ближайшие месяцы мы планируем добавить в это данные агрегаторов Работа в России или другом, и SuperJob. На сегодняшний день мы работаем над данными от HeadHunter. Мы грузим данные из HeadHunter через API, мы их грузим каждый день. Мы на сегодняшний день раз в месяц обновляем дашборд. Планируем в будущем делать это чаще, но пока это технически довольно тяжело, там очень много данных, и дашборд тяжело переваривает. На сегодняшний день в дашборде 8,9 миллионов вакансий и 193 тысячи уникальных навыков. То есть того, как работодатели их написали. Сварку можно написать, как выяснилось 15 разными способами. В дашборде 193-194 тысячи уникальных навыков. В нашем сервисе есть вторая крупная вкладка, это вкладка данные, при помощи этой вкладки можно получить доступ к нашей базе данных непосредственно. То есть все эти вакансии, которые мы за весь год, все эти 8,9 миллионов мы спарсили, мы не парсили., мы получаем доступ

через API. Да, через вкладку данные получили доступ к нашей базе данных, там 8,9-9,9 миллиона вакансий, и получить оттуда в табличной форме, если по какой-то причине не получается пользоваться интерактивной графикой, которая представлена в дашборде, а удобнее работать с табличными, то вот можно получить данные в табличной форме. Там есть свои ограничения по количеству строк, которые можно выгрузить за раз, например, или там, по количеству доступных вакансий, или регионов, которые также за раз выбрались, но в целом для какой-то небольшой самостоятельной аналитики того объема данных и того функционала, который есть в нашем сервисе в виде дашборда и в виде доступа данных, этого достаточно. Если вам необходимо провести какую-то более сложную аналитику, более глубокую аналитику с множеством условий и не представляется возможностью сделать это с помощью дашборда или это с помощью самостоятельных выгрузок и анализов данных, можно обратиться к аналитической команде сервиса, аналитическая команда сервиса с радостью, поможет вам реализовать любую задачу, любой вопрос совместно с вами, разрабатывается техническое задание, выбрать данные, проанализировать данные и разработать отчет. Вот, я не знаю, все ли я рассказала про наш сервис. Наверное все.

В: Так, хорошо, а в чем преимущество анализа больших данных в аналитике рынка труда? Какие возможности есть?

О: Наверное, независимость, потому что при проектировании образовательных продуктов есть три пути, как понять, что нужно вложить в образовательные данные. Есть три пути, первый путь - это идти по ГОСТам, по государственным образовательным стандартам и смотреть, что есть в ГОСТах и что нужно делать. Но беда в том, что не под все современные профессии разработаны ГОСТы, то есть часть современных профессий до сих пор не имеет ни государственных стандартов, ни каких-либо образовательных стандартов, это первый момент, а второй путь, как можно подойти к разработке программы - это пойти к работодателям, пойти к работодателям, провести фокус-группу из всяких агрономов. Да, вторая путь - это пойти к работодателям, сесть с ними и проверить, типа, от чего вам нужно вообще от людей, которых вы хотите видеть на работе, и они тебе расскажут, что мы хотим агрономов, чтобы вы могли бы сажать и полоть, но еще и чистить картошку. Я просто не знаю, чем заниматься агрономы. Не важно, дизайнеров, которые умеют не только садить и полоть, но в том числе чистить картошку. И люди такие "О, я знаю, чистить картошку, добавим сюда, добавить в нее приправу". Но есть нюанс, и он связан с узостью, с тем, что, возможно, таким образом вы сохраняете специалистов для определенной позиции и определенного работодателя. То есть, условно, возможно, в течение 10 компаний, которых вы собирали на фокус-группу, действительно нужно чистить картошку, а всем остальным это не нужно. Но так случилось, что дальше выбор попали те, кому нужна картошка. Поэтому вы учите уметь чистить картошку, а надо было все эти годы учить чистить апельсины. Вот. Да, это второй путь разработки образовательных программ. И третий путь, собственно, с опорой на какие-то независимые источники, на данные, на данные и это уж неважно, на большие, на маленькие, какие-то они будут, и опираться на данные мониторинга рынка труда - это хорошая удобная история. Но так или иначе нельзя опираться на что-то одно, естественно, лучше всего, наверное, комбинировать эти штуки. То есть что-то брать из ГОСТов, что-то брать из работодателей. То есть работодатель тоже могут много что полезного рассказать. Ну и да, какую-то часть брать исходя из аналитики данных. Наверное, я не скажу, что у этого есть преимущество. Наверное, у этого преимущество только независимость. Независимость от того, как специалисты видят государство, независимость от того, как специалистов видит конкретный работодатель. То есть это позволяет построить, наверное, более общую картину о том, как этих специалистов в целом видит рынок. Вот. Ну, в остальном я не могу сказать, что это прям какое-то конкретное преимущество, кроме независимости. Наверное, нет. Но мне кажется, что с точки зрения именно образовательных программ лучше комбинировать все эти три штуки и не опираться на что-то одно. Вот.

В: А какие ограничения у такого метода в аналитике рынка труда?

О: Наверное, недостаток экспертности. Мы столкнулись с этим, когда разрабатывали сервис. Навыки и компетенции, в которых сварку, можно записать 15 разными способами. Но я, как аналитик, я не сварщик, я не знаю 15 это разных вариантов сварки или один и тот же, и мы, безусловно, мы собирали фокус-группы с экспертами, мы создавали и разрабатывали классификаторы навыков и компетенций вместе с экспертами, для того, чтобы понять, что из написанного вообще адекватно, а что неадекватно, но, все эти данные невозможно обработать всецело, полностью руками, все эти данные не могут обработать эксперты.

Мы не можем обработать мы, даже основу ВАИС НАТО, высказываем эксперт.

193 тысячи уникальных навыков это только сегодня, а завтра их будет больше, ведь данные растут по экспоненте, и после завтра их будет больше, их из раза в раз будет больше. Физически люди такое

количество данных обработать не могут, поэтому мы обучаем искусственный интеллект, ну не искусственный интеллект, а модельку, которая на основании уже какого-то массива, размеченного экспертами, то есть эксперт, там разметили, я не знаю, три, десять тысяч навыков, и вот на основании этих размеченных десяти тысяч навыков моделька учится, и дальше она начинает самостоятельно эти навыки расфасовывать. Но моделька не способна выделять новые смысловые категории, то есть если в какой-то момент на рынке труда начнут требовать абракадабру, а в модели у нас нет абракадабры, то модель эти навыки постепенно будет складывать куда-то отдельно, то есть она будет говорить, я не знаю, куда это деть, я не знаю, куда это деть. И если мы как разработчики сервиса вовремя не обратим на этот момент внимание, просто отложим абракадабру в долгий ящик, она будет там лежать, то в этом все перестает иметь смысл, потому что как раз в тот момент, когда появляется какая-то новая компетенция, которую требуют и под которую нужно перестроить образовательную программу довольно срочно, или там разработать какую-нибудь ДПОшку, чтобы люди с такими навыками появились на рынке труда, мы не сможем этого сделать, потому что мы эту абракадабру откладываем в долгий ящик. Потому что мы не все сильные модели... мы так не делаем, пока что мы не сталкивались с такой ситуацией. То есть на сегодняшний день разработаны нами классификаторы, там, открывают 70... нет, 70, можно же соврать, сколько, ну проще, открывают там энное количество массива, и вот это то, что лежит в ящике, оно, конечно, пока лежит, но мы им занимаемся. Нет, ну, наверное. Возможно, да, возможно, неточность... Ну, да, поскольку этим занимается модель, она не всегда способна что-то поделить. То есть модель представляет это по близкой паре слов или символов и иногда, когда символы или слова схожи, они могут иметь разный смысл. Ну, там замок и замок, условно. Это как бы одинаковое написание. Для модели это одно и то же, для нас это разная вещь, поэтому некоторые вещи модель классифицирует, естественно, неверно. Но поскольку это объем данных очень большой, руками и человеческими ресурсами это невозможно обработать, и иногда в огромном массиве данных это тяжело заметить. Вот, что еще? Ну, возможно, вычислительные мощности. Объективно, далеко не каждый дома сможет такое проверить. Для аналитики каждый день на Headhunter открыто миллион вакансий. То есть, если в один день прийти исключать все вакансии из Headhunter, потом пытаться как-то их проанализировать, то существует вероятность, что практически любое домашнее железо умрет. То есть это действительно большие вычислительные мощности, и не каждый дома сможет себе такое проверить. Ну, не знаю. Ну, опять же, наверное... Наверное, вы знаете, что все может сводиться к необъективности. Ну, то, о чем я вначале говорила про то, что... А с чего это вы решили, что ваши данные вообще валидные? С чего это вы решили, что ваши данные репрезентативные? Как это вы себе решили? Да, мы собирали группу с экспертов. Это были разные эксперты. Ну, то есть там по одной отрасли, допустим, не знаю, отрасли, горнодобывающая отрасль у нас была, и там были эксперты из разных горнодобывающих компаний, добыча нефти, добыча газа и так далее. Эти эксперты были из разных компаний, да, мы уже разработали с ними совместно, то есть они там все согласились с разработанными классификаторами навыков. Но это не значит, что вы не совершили ошибку. Там вот эти три-пять экспертов, которые у нас были, они составили мегаверный классификатор. Придет обязательно... Вот сейчас у нас пришел эксперт, неожиданно пользовался в нашем сервисе, эксперт в сфере нанoeлектроники и он высказал нам свое мнение по поводу того, что у нас там очень неправильно написаны навыки в дашборде, потому что с него нанoeлектроническое дозвучение - это супер неверно и по его мнению там должны быть какие-то другие штуки. Мы работаем с этим. С этим мы тоже работаем, мы продолжаем привлекать экспертов, и мы всегда рады, если какой-то эксперт придет к нам и скажет, давайте мы с вами там потусим, что-нибудь сделаем. Классно разработаем конституцию. Мы поможем вам доработать какие-то области, если они у вас работают. Мы открыты к соперничеству и людям.

Я начинаю отказываться, что я минусов перечислила больше, чем плюсов.

Ну, что поделать.

В: Какие перспективы развития у аналитики рынка труда с помощью больших данных?

О: Перспективы?

В: Как изучение будет меняться? Или может быть рынок труда будет меняться и как подстроиться с помощью больших данных?

О: Я не уверена, что я могу и что я компетентна строить подобные прогнозы. В идеале хотелось бы, чтобы люди перестали писать, чтобы люди внимательно писали, какие данные они продюсируют в мире. Возможно, если люди поймут, что это полезно, что они могут что-то свое из этого вынести, что свое это научить, они начнут внимательнее относиться к тем данным, которое они продюсируют. В поле ключевых навыков в стране они не будут писать "звездочка, звездочка, звездочка, классный

дизайнер, звездочка, звездочка, звездочка". Это тоже окей. Это здорово, креативно. Очевидно, что креативно. Но я к тому, что масса каких-то данных, она не овладни массой.

Иногда мы сталкиваемся с очень странными вещами во время аналитики. В как-то раз мы нашли вакансию, это любимый пример, который мы часто приводим в своем проекте, как-то раз мы нашли вакансию дворника, которому в навыках нужен был PowerPoint. Почему так? Не понятно. Но мы можем, люди будут внимательнее относиться к данным, который они производят, данным, с которыми они работают.

Но как это повлияет на рынок тогда, мне, наверное, тяжело сказать. Может быть, все супер классно перестроится. И на самом деле сейчас на это есть тренд, что у нас очень много образования стало неформального, типа того же самого Skillbox, Skin Factory, еще куча всяких skills. Как это называется? Yandex? Yandex Academy? Куча онлайн школ, где можно получать навыки, которых у тебя нет. Может быть, где-то кем-то вводят. Условно, ходя с ВУЗа, даже отучившись на графического дизайнера, тебе может не хватать скиллов. И ты пойдешь на тот же Skill Factory, и эти скиллы себе получишь. Наверное, в перспективе, возможно, открыть какого-то ДПО, дополнительное образование сильно проще, чем изменить существующую программу высшего образования. Или открыть новую программу высшего образования. Возможно, рынок ДПО захватит мир, люди перестанут получать высшее образование, все перейдут к ДПОшке, потому что так проще жить. Не знаю.

В: Хорошо, все, спасибо.

Интервью №6, Ирина, аналитик лаборатории прикладного анализа больших ТГУ

Длительность: 35 мин

Вопрос: Расскажите про работу в лаборатории. Какими проектами занимались и занимаетесь сейчас, и какая должность?

Ответ: Вообще в целом.

В: Да.

О: То есть про лабораторию? В лаборатории я работаю аналитиком, должность аналитика. Соответственно, проекты были разные. Был, например, проект связанный с изучением коренных молочистостных народов с севера Сибири и Дальнего Востока. Там мы анализировали разные сети, хотели ещё анализировать разные проблемы, которые вообще в целом, которые волнуют коренные молочистостные народы, но до этого этапа, к сожалению, проекта дошёл, мы остановились на таком сетевом анализе. Плюс сейчас, например, идёт проект по мониторингу труда. Мы создали платформу «Роснавыка», которая позволяет анализировать разные навыки, вообще все вакансии, навыки в том числе, которые есть в этих вакансиях, выкригированного в общем виде в одном месте. Это очень удобно. И вот сейчас мы активно занимаемся продвижением этого проекта. Поэтому я не так давно работаю в лаборатории. Поэтому, в принципе, наверное, из основных проектов это всё. Есть ещё некоторые, ну, такие были точечные проекты. Например, мы занимались организационной культурой Томского государственного университета. Мы проводили опрос среди всех сотрудников Томского государственного университета и хотели построить организационную сеть. То есть кто с кем взаимодействует, по каким каналам, кто как с кем. Вообще дружит, не дружит, работает, не работает. Но тоже возникли некоторые проблемы, скорее даже со сбором данных трудности возникли, поэтому проект тоже был на паузе. Поэтому вот таких больше проектов было несколько, но основные это вот такие.

В: А как давно работаете в лаборатории?

О: Слушайте, сложный вопрос, если честно. Ну, вот прям чтобы работаю, ну, наверное, около года.

В: А вот лаборатория больших данных. А что такое большие данные? Какие это данные?

Большие данные, это данные и информацию о которых очень сложно анализировать в каких-то привычных нам объёмах. То есть это информация, которая пополняется постоянно, она не конечна. Она, то есть база этих данных постоянно пополняется. Это данные, которые соответственно, ну, действительно, которые, например, сложно обрабатываются какими-то привычными нам инструментами. И соответственно, для их обработки нужно привлекать искусственный интеллект, машинное обучение и прочие инструменты для такой аналитики. Плюс большие данные – это данные, которые сейчас в настоящее время становятся одними из таких самых, наверное, популярных, востребованных данных, в принципе, по современному обществу. Потому что в целом

наше общество, оно очень сильно сейчас завязано на информации, на любого рода данных. Поэтому не могу сказать, что есть, например, если есть какие-то большие данные, есть, например, там маленькие какие-то данные. Нет, но вот большие данные – это то, что стоит сейчас, как так во главе, в принципе, информационной эпохи, скажем так.

А в чем преимущество такого метода? Метода анализа больших данных.

О: Анализа больших данных? А преимущество по сравнению с чем?

В: С другими методами. В социологии, например.

О: А преимущество с другими методами в социологии, да. Ну, на самом деле у разных методов разные задачи, в принципе, всегда. То есть для социологических методов, если мы говорим про опросные методы, у них совсем другие задачи, они выявляют совсем другие... Ну, преследуют совсем другие цели. То есть там мы больше акцентируем, например, на людей, внимание, если мы говорим про методы интервью, например, то мы говорим про изучение там жизненного мира человека. А большие данные, например, у них абсолютно тоже своя совсем другая специфика. То есть, например, анализировать большие данные можно разными способами, в принципе. То есть можно под большими данными, например, понимать данные, которые находятся там, не знаю, в поисковиках, например, какие-то данные. Либо это могут быть какие-то... Ну, на самом деле, вообще любые данные могут быть абсолютно, не знаю, начиная от данных о сводках погоды, например, за последние года по всему миру, например, заканчивая, не знаю, тем же социальными сетями и всей информацией, например, которая содержится в этих социальных сетях вариантами. Поэтому анализ больших данных проводится всегда по-разному и в зависимости от того, что какие цели. То есть, на те преимущества и будут. Но если говорить в общем, про такую аналитику, то это возможность, во-первых, изучения какого-либо объекта в динамике. То есть мы всегда видим предпосредств. Масса всегда и в целом в наше общество предполагает, что мы эти данные постоянно накапливаем. То есть если мы их накапливаем, мы всегда можем обратиться в какие-то респективы. Плюс ко всему мы видим естественность этих данных. То есть эти данные не собираются, ну, прям специально. Они есть, и мы занимаемся их обработкой, соответственно, ну, естественной их среде обитания, скажем так. Поэтому это всегда показатель, ну, с точки зрения социальных данных, показатель того, как вообще общество себя чувствует, например. В каком состоянии оно находится, какие у него проблемы, боли, ожидания и так далее. То есть в целом это такой срез. Большие данные, анализ больших данных дает нам возможность такого среза, скажем так, такого рода информации. Что еще? Ну, большие данные – это данные, которые, если социологически сравнивать, это данные, которые, по большому счету, не зависят от внешних условий. Ну, по внешним условиям я имею в виду субъекты действия. То есть, например, если социология – это человек, мы зависим от человека, мы должны его найти, мы должны его опросить специально, про какие-то темы, начать с ним разговаривать, то большие данные, ну, этого не требуют в принципе. Мы можем вне зависимости от желания человека, ну, даже не желание, если так можно сказать, а вне зависимости от, ну, в целом, его отношения, например, к тому, хочет он как-то или не хочет, ну, можем проанализировать какую-либо информацию, то есть больше объективности, можно сказать, дают нам большие данные. Тут, в принципе, достаточно вспомнить, например, знаменитого социолога, Пьера Бурдьё, который там писал про общественное мнение, у него была замечательная работа, которая называется "Общественное мнение и не существует", в котором он критиковал вообще процедуру вопросного метода, не только он, еще, например, Патрик Шампань тоже про это писал, французский социолог, вообще про критику вопросного метода, ну, если мы говорим про социологию. Вот, поэтому большие данные, они как раз таки позволяют вот такой срез провести, который не так субъективен, как, например, большинство социологических методов. Вот. Что еще? Ну, наверное, пока все, если я еще вспомню что-то, я, по ходу, могу сказать.

Хорошо. Какие тогда недостатки у метода?

О: Недостатки. На мой взгляд, это трудоемкость такой процедуры, потому что для того, чтобы, ну, давайте так, в целом, чтобы организовать любое исследование, анализ больших данных, ну, как метод, да, и в целом, это все равно предполагается, что это какое-то исследование, для исследования, ну, чаще всего, всегда предполагается наличие какой-то команды для выполнения одной или иной задачи. Соответственно, если мы сравниваем, например, социологическими методами, ну, с опросами давайте, удобнее мне почему-то с ними сравнить, если мы, например, с ними сравниваем, то для того, чтобы организовать социологическое исследование, для того, чтобы организовать исследование с использованием метода больших данных, на самом деле, как бы это ни казалось, ну, странным, ну, я здесь не могу говорить про все исследования, но, на мой взгляд, все-таки большие данные требуют порой даже большей команды, чем организация социологического исследования, и я говорю именно

про организацию, то есть не про полевой этап, а именно про вот такую, скажем так, подготовку, потому что большие данные, там, это и большие серверные нагрузки, и большие, там, нагрузки, связанные с различными хранилищами данных, различными серверами, ну, в общем, если честно, вот в плане технической части я не очень хорошо в этом разбираюсь, но в любом случае, то есть с технической точки зрения организации этого процесса, на мой взгляд, метод больших данных требует гораздо больших затрат. Что еще? Ну, если говорить про преимущество, кстати говоря, он будет, скорее всего, дешевле, чем, например, те же публичные, опросные методы. Если что, про недостатки, да, еще, соответственно, мы можем сказать. Недостатки. Информация - это такая вещь, которая, ну, это ресурс, скажем так, то есть это некий ресурс, это некоторый капитал, обладание информацией - это некоторый капитал. Соответственно, за капиталы люди обычно борются. Почему? Потому что капитал дает тебе возможность какого-то социального роста, ну, изменения своего социального статуса. Но в связи с этим может возникнуть, крайней мере, такая проблема, связанная с закрытостью информации. То есть люди, многие могут просто, ну, наверное, это банальный какой-то пример, но, тем не менее, люди могут бояться делиться какой-то информацией, либо не разглашать какую-то информацию. И, соответственно, если эта информация недоступна, например, то чтобы ее добыть каким-то образом, нужно приложить либо очень много усилий, либо потратить на это, опять же, очень много ресурсов для того, чтобы она стала доступной, и так далее. Я это к тому, что недостаток использования больших данных может заключаться в том, что информация, которую мы хотим найти, она может быть всегда доступной. Соответственно, вот такая закрытость информации - это некоторый барьер, некоторая такая преграда для использования этого метода. Ну, если мы говорим, например, про анализ тех же самых социальных сетей, то что мы там можем анализировать? Мы там можем анализировать, например, данные, которые находятся на странице людей, какая-то их личная информация. Соответственно, если человек закрывает эту информацию, то мы никаким образом это не проанализируем. Вот. Еще один недостаток случается, тоже связан, на самом деле, с понятием того, что информация - это некоторый ресурс. Это, ну, громко, возможно, скажу, но это подделка этой информации. То есть, в принципе, люди, публикуя какую-то информацию, не всегда могут задумываться о том, насколько эта информация правдива. Соответственно, мы, когда анализируем какую-либо информацию, мы всегда должны понимать, то есть отличать, например, критический взгляд, в общем, смотреть эту информацию, понимать, насколько она действительно, например, достоверная или нет. И, соответственно, мы можем столкнуться с такой ситуацией, когда мы анализируем что-либо какое-то явление, какой-то феномен с помощью больших данных, но получаем, например, результаты, которые не соответствуют реальности. Почему? Потому что эта информация была, ну, эта вот информация не соответствовала настроенности. Поэтому здесь, вот, наверное, в этом я бы еще увидела более достаток. Ну, это из-за глобальных вещей.

В: Так, а что нужно знать и уметь, чтобы работать с большими данными? Какие навыки нужны?

О: Много. Много навыков на самом деле надо. Ну, во-первых, это действительно критическое мышление. То есть, во-первых, это всегда такой очень критический взгляд на вещи должен быть, потому что нужно понимать, какого рода вообще информацию человек анализирует. То есть нужно разбираться в источниках информации, в том, насколько они релевантны. Нужно понимать про, ну, вообще, в целом, содержание этой самой информации. Поэтому такой критический взгляд, он очень полезен всегда при работе с большими данными. Не могу сказать, что этот навык специфичен для использования больших данных. Он нужен, в принципе, и везде практически. Но, мне кажется, он один из самых важных для этой сферы. Что еще? Нужно обладать знаниями в области аналитики, условно, конечно же. Но понятие аналитика – это очень большое и очень пространное понятие. В аналитику входит абсолютно все, начиная от анализа каких-то данных и заканчивая вообще там, в честь линии разработки, то есть какие-то аналитические схемы, придумывания и прочих вещей. Но, в целом, аналитика, ну, давайте так, два аспекта аналитики выделить. Первый – это именно обработка данных, то есть это непосредственно аналитическое мышление. Два аспекта аналитики. Первое – аналитическое мышление непосредственно. То есть, когда мы берем информацию, можем ее анализировать каким-то образом. Мы строим какие-то схемы, мы продумываем какую-то логику. То есть, это не только описательная какая-то идея. Мы не только описываем какое-то явление, но и мы можем его проанализировать. То есть, мы можем его проинтерпретировать и аналитически на него посмотреть. Вот. А второй аспект аналитики, он касается непосредственно уже, ну, такой технической части. Может быть, это не совсем стоит аналитикой, конечно, назвать, но здесь я имею в виду больше про инструменты. Непосредственно те инструменты, которыми мы можем обрабатывать большие данные. То есть, это начиная от инструментов, связанных с обработкой этих данных. То

есть, там, ну, даже вот то, что для социологии, например, неминимо, да, это программа из PSS и прочие какие-то методы анализа. Так и заканчивая разными вещами, там, по типу Python, по типу знания языков SQL, по типу, там, ну, и всех других прочих языков, там, программирование и так далее. То есть, умение работать вообще вот с таким В.У., да, программами обеспечения. В принципе, нужно для анализа больших данных вообще понимать понятие базы данных. Вообще понятие данных, на самом деле, тоже очень важно. Но про базы данных это, наверное, ключевой, потому что любые большие данные, это, ну, вот именно в изначальном, да, в первичном таком виде своем, это некоторые базы данных. Действительно, некоторые таблицы, массивы данных, которые содержат в себе совершенно разные информации. И вот умение работать с базой данных - это очень важная навыка. Что еще? Ну, про ПО, на самом деле, здесь можно тоже еще говорить про много-много разного всякого. Ну, как я уже говорила, анализировать большие данные можно разными способами. То есть, можно, например, строить там сетевой анализ, да, то есть строить социальные сети. Соответственно, нужно обладать навыками, которые позволяют тебе построить этот сетевой анализ совершить, да, то есть, там, программное обеспечение соответствующее, знать-знать, понятие там сетей и так далее. То есть, если мы говорим про обработку, например, там, естественного языка, это совершенно другие наборы, да, навыков и компетенций. То есть, нужно понимать, каким образом анализировать, там, есть, проводить анализ естественного языка, да, то есть, нужно создать методы обработки текстов, да, начиная там от того же самого контента анализа и заканчивая там, ну, вот, действительно, каких-то более серьезными вещами, серьезными, я имею в виду, с технической точки зрения, там, тоже каким-то программным обеспечением, по типу, огае или более, анализа. То есть, ну, это тоже совсем другие навыки появляются. И, в целом, ну, вот, основные это вот критическое мышление, аналитика, знание программного обеспечения, и они, соответственно, ну, вот дальше они уже, на мой взгляд, такие отраслевые становятся, в зависимости от того, какой метод используется.

В: А если переходить к аналитике рынка труда, зачем нужен мониторинг рынка труда? Что это дает?

О: Вообще, в целом или...

В: В целом.

О: Аналитика рынка труда, ну, вообще, если мы говорим про проблематику и про область рынка труда, это очень такая интересная область. В самом деле, ее истоки исходят еще там, начиная от даже социологической, если смотреть социологическое точки зрения, там еще и Дюргейм, например, и, ну, вот, такие классики социологии, да, говорили, например, например, про способ разделения труда, о разные виды солидарности и так далее. То есть, в целом, проблема труда как таковая, разделения труда, она была актуальна, в принципе, всегда. Здесь, ну, можно много говорить про это, но, я думаю, вопрос был не совсем про конкретно прошлую ситуацию. Соответственно, сейчас, если мы говорим про современное состояние, сейчас рынок труда - это такое место, такое пространство, в котором происходят, ну, порой очень интересные и очень разные и странные события, скажем так, которые делают рынок труда интересным для изучений. Почему? Потому что в современном обществе проблема трудоустройства и, в принципе, проблема занятости людей, она встает достаточно остро. Начиная от занятости, например, несовершеннолетних, например, что сейчас вышли разные законы, которые касаются того, что можно предостраиваться и до достижения совершеннолетнего возраста, там уже школьники стали работать и так далее, и заканчивая теми же самыми пенсионерами, предпенсионного возраста и пенсионного возраста, чтобы люди разным вообще категориям каким-то образом трудоустраиваться и работать. И, соответственно, вот эта тема, она, изучение этой темы вообще позволяет понять, насколько рынок туда и вот насколько разные виды деятельности, которые существуют в нашем обществе, они востребованы. Ну, здесь тоже не секрет, я думаю, для кого, что мы с течением времени, с исторические периоды всегда можем видеть смены основных видов деятельности. В принципе, я думаю, вряд ли кто-то 100-200 лет назад мог сказать бы, что сейчас, например, вот через какое-то время будет очень популярна сфера, например, тех же самых информационных технологий и все, что с этим связано, у нас появляются новые профессии, появляются новые отрасли, появляются новые виды занятости, формы занятости, то, что там фрилансов, занятости и прочие вещи. Это все происходит потому, что люди понимают, что вот их занятость, вот это разделение труда, которое всегда на протяжении всего периода существования истории происходит, эти формы занятости, они всегда претерпевают некоторые изменения и для того, чтобы эти изменения своевременно отслеживать, для того, чтобы эти тренды понимать, как развивается это, например, в каких отраслях, в каких формах, в каких видов, в каких, не знаю, если мы говорим о рамках России, например, в каких регионах и так далее, то есть какие-то мы видим различия, для того, чтобы понимать, насколько и каким образом вообще это развитие происходит,

нужно всегда анализировать в очень общем виде. Поэтому вот такая аналитика, она всегда очень полезна, то есть не только в рамках всей России, даже в рамках конкретного региона, даже на самом деле в рамках конкретного подразделения, например, мы можем в университете, часто история, когда собирается информация о трудоустройстве выпускников, чтобы посмотреть, как рынок труда ведет себя в рамках какой-то специальности, по направлению подготовки. Поэтому в первую очередь аналитика рынка труда дает нам возможность проследить вот эти тренды, не упустить их и дать возможность понимать дальнейшие перспективы, то есть понимать, в каком направлении рынок труда может меняться в дальнейшем, расстраивать другие продукты. Вот. Надеюсь, ответила на вопрос.

В: А как большие данные помогают в таких исследованиях?

О: Большие данные очень хорошо помогают в таких исследованиях, потому что рынок труда – это очень большая область. Здесь мы никогда не ограничиваемся одним, двумя, тремя, десятью людьми, это всегда сотни, тысячи, даже миллионы занятых людей, поэтому рынок труда с использованием методов анализа больших данных – это неотъемлемая часть этого анализа, потому что с помощью больших данных мы можем как раз-таки агрегировать все данные, которые у нас есть в какой-то удобный вид, скажем так. То есть, например, сайты агрегаторов вакансий, что мы, например, делаем, мы берем просто обычные сайты, где выпускается информация о вакансии, для простого человека это выглядит как набор некоторых страничек с разными вакансиями, он ищет себе нужные вакансии, если он, например, соискатель либо наоборот, он работодатель, он может своих конкурентов посмотреть. Но в любом случае обычно, как правило, это просто несколько таких страниц, которые можно просмотреть вручную. Большие данные – это не те данные, которые можно просмотреть вручную глазами. Поэтому здесь нам нужно всегда понимать, во-первых, что мы анализируем, какой объем данных мы берем, а во-вторых, зачем. Если мы анализируем рынок труда в целом, даже в рамках города, тем более в рамках страны, то эти данные – они очень большого объема. Поэтому здесь большие данные являются ключевым инструментом для вообще возможности этой аналитики, потому что нужно выгрузить данные, нужно их каким-то образом унифицировать, собрать их в некоторые базы, потом из этих баз доставать нужную информацию, уметь вообще это делать. То есть здесь на самом деле большие данные – это, наверное, один из ключевых инструментов для этого анализа, только так мы можем увидеть картину в целом.

В: А какие есть недостатки в применении небольших данных в исследованиях рынка труда?

О: Недостатки? На самом деле здесь все те же самые недостатки, что и я перечисляла для просто использования методов больших данных. Не думаю, что здесь принципиально от сферы что-то зависит, но мы, например, сталкиваемся с тем, что информация этих данных, которые мы получаем, она либо является неполной, то есть какие-то данные просто отсутствуют, из-за чего мы не можем провести аналитику. Например, история, когда мы берем сайты агрегаторов вакансий, вытаскиваем оттуда вакансии, их анализируем, но большинство работодателей, например, не указывают уровень заработной платы для соискателей, или не указывают ключевые навыки, например, которые требуются, или еще что-то не указывают. И, соответственно, из-за того, что этой информации у нас нет, мы не можем в общем виде проанализировать рынок труда, и мы не можем составить вот такую прям стопроцентную картинку, скажем так, очень полную. То есть мы можем делать какие-то некоторые прогнозы, но они не всегда могут являться точными, так как часть данных отсутствует. Какие еще есть недостатки? Есть недостаток, связанный с тем, что работодатели, например, часто... ну даже не так, есть проблема в способах составления вакансий, и работодатели, когда выкладывают свои вакансии, они очень часто, грубо говоря, могут написать какую-то ерунду, то есть они могут особо не заморачиваться и просто написать, как им кажется. То есть там очень такая показательная, наверное, картина, когда мы начинали только вести этот проект, и мы вот вручную некоторые вакансии отсматривали, и вот действительно наткнулись на вакансии, в которых, например, написано "вакансия там охранника" или "вакансия дворника", и в какой-то описание вакансии внизу были написаны какие-то ключевые навыки для этой вакансии, ключевые навыки, например, умения работать в Powerpoint, то есть там в обязанностях написано "уборка территории", при этом в ключевых навыках написано "умение работать в Powerpoint". И вот и сиди и гадай, зачем охраннику или дворнику при таком раскладе уметь пользоваться в Powerpoint. Можно говорить о том, что это естественные ошибки, да, такие случайные ошибки, да, это бывает. Никто от них не застрахован, но если мы, например, сравним, опять же, с изучением рынка труда с использованием, например, тех же самых социологических опросных методов, то такая проблема немножко отпадает, потому что человек нам скажет, что "мне не нужен никакой Powerpoint, у меня есть другие требования какие-то, зачем он мне?" А здесь мы полагаемся на то, что он написал. Поэтому здесь такие ошибки, которые

случаются, к сожалению, они могут привести не совсем к правильной аналитике, не совсем к правильным результатам, но мы стараемся вот действительно такую проблему избежать, насколько это возможно, предугадать её как-то, предостеречься. Но не всегда это может получиться.

В: А как будет меняться анализ рынка труда с помощью больших данных?

О: Как будет меняться с помощью больших данных?

В: Да, перспективы развития анализа данных. Какие с помощью больших данных? Перспективы развития анализа рынка труда с помощью анализа больших данных.

О: На самом деле, я думаю, что это было бы здорово, я не могу сказать, что будет это или нет, но вообще интересной перспективой является идея о том, что все кадровые агентства, процедура поиска сотрудников и прочих вещей, она станет достаточно автоматизированной, когда, например, любой человек сможет буквально вбить три-четыре галочки, поставить несколько полей, и ему будет также на основании такой аналитики предлагаться способы его дальнейшего трудоустройства. Например, он вводит свои какие-то навыки, вводит свои умения, что он имеет, например, и ему какая-нибудь программа, какой-нибудь ресурс говорит. Потому что, смотрите, вы можете трудоустроиться сюда-сюда-сюда, например, но, смотрите, вы с такой зарплатой, с таким графиком работы, вы можете трудоустроиться сюда. Но, смотрите, одновременно с этим, если вы к своим навыкам добавите еще такой навык, например, обучитесь ему где-нибудь, и тут же, например, появляется какой-нибудь список со всеми ресурсами, где человек может этим навыкам научиться, например, то вы сможете работать уже там-вот там-вот там, на таких-то условиях и так далее. То есть, на мой взгляд, вот создание подобных сервисов и создание подобных площадок, где это могло бы быть возможно, это очень здорово, очень перспективно, потому что это очень сильно упрощает и снижает проблему обучения. То есть, когда студенты, например, приходят в университет, они... Ну, это частный вопрос, когда спрашивают, кем я буду работать, кем я смогу трудоустроиться, что я получу, какие навыки, там, смогу ли я научиться еще чему-то и пойти куда-то еще работать. Вот на данном этапе развития нашего общества, ну, я, по крайней мере, не знакома, возможно, где-то есть такие ресурсы, но вот пока я не наткнулась в такое, когда бы я посмотрела бы и увидела, что, ага, значит, вот был бы такой сервис, была бы такая возможность это все автоматизировать. Это было бы очень здорово, и я думаю, что в рамках нашего, например, проекта, мы могли бы такую идею реализовать, то есть, когда бы это все было бы очень классно или нездорово реализовывалось. Вот, наверное, ну, вот на данном этапе, мне кажется, вот это самая такая ближайшая перспектива, которая могла бы быть реализована.

В: Так, ну все, вопросы закончились, спасибо.

Отчет о проверке на заимствования №1



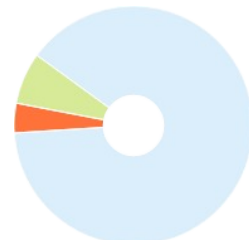
Автор: Остапенко Мария Александровна
Проверяющий: Плешкевич Ирина Борисовна
Организация: Томский Государственный Университет
Отчет предоставлен сервисом «Антиплагиат» - <http://tsu.antiplagiat.ru>

ИНФОРМАЦИЯ О ДОКУМЕНТЕ

№ документа: 329
Начало загрузки: 16.06.2023 05:28:46
Длительность загрузки: 00:00:17
Имя исходного файла:
VKR_Ostapenko_Maria.docx
Название документа: VKR_Ostapenko_Maria
Размер текста: 260 кБ
Символов в тексте: 266008
Слов в тексте: 37594
Число предложений: 2082

ИНФОРМАЦИЯ ОБ ОТЧЕТЕ

Начало проверки: 16.06.2023 05:29:04
Длительность проверки: 00:00:46
Комментарии: не указано
Поиск с учетом редактирования: да
Проверенные разделы: титульный лист с. 1, основная часть с. 2-3,5-40, содержание с. 4, библиография с. 41-44, приложение с. 45-80
Модули поиска: ИПС Адилет, Библиография, Сводная коллекция ЭБС, Интернет Плюс*, Сводная коллекция РГБ, Цитирование, Переводные заимствования (RuEn), Переводные заимствования по eLIBRARY.RU (EnRu), Переводные заимствования по коллекции Гарант: аналитика, Переводные заимствования по коллекции Интернет в английском сегменте, Переводные заимствования по Интернету (EnRu), Переводные заимствования по коллекции Интернет в русском сегменте, Переводные заимствования издательства Wiley, eLIBRARY.RU, СПС ГАРАНТ: аналитика, СПС ГАРАНТ: нормативно-правовая документация, Медицина, Диссертации НББ, Коллекция НБУ, Перефразирования по eLIBRARY.RU, Перефразирования по СПС ГАРАНТ: аналитика, Перефразирования по Интернету, Перефразирования по Интернету (EN), Перефразированные заимствования по коллекции Интернет в английском сегменте, Перефразированные заимствования по коллекции Интернет в русском сегменте, Перефразирования по коллекции издательства Wiley, Патенты СССР, РФ, СНГ, СМИ России и СНГ, Шаблонные фразы, Модуль поиска "tsu", Кольцо вузов, Издательство Wiley, Переводные заимствования



СОВПАДЕНИЯ	САМОЦИТИРОВАНИЯ	ЦИТИРОВАНИЯ	ОРИГИНАЛЬНОСТЬ
4,29%	0%	7,07%	88,64%

Совпадения — фрагменты проверяемого текста, полностью или частично сходные с найденными источниками, за исключением фрагментов, которые система отнесла к цитированию или самоцитированию. Показатель «Совпадения» — это доля фрагментов проверяемого текста, отнесенных к совпадениям, в общем объеме текста.

Самоцитирование — фрагменты проверяемого текста, совпадающие или почти совпадающие с фрагментом текста источника, автором или соавтором которого является автор проверяемого документа. Показатель «Самоцитирования» — это доля фрагментов текста, отнесенных к самоцитированию, в общем объеме текста.

Цитирование — фрагменты проверяемого текста, которые не являются авторскими, но которые система отнесла к корректно оформленным. К цитированиям относятся также шаблонные фразы; библиография; фрагменты текста, найденные модулем поиска «СПС Гарант: нормативно-правовая документация». Показатель «Цитирования» — это доля фрагментов проверяемого текста, отнесенных к цитированию, в общем объеме текста.

Текстовое пересечение — фрагмент текста проверяемого документа, совпадающий или почти совпадающий с фрагментом текста источника.

Источник — документ, проиндексированный в системе и содержащийся в модуле поиска, по которому проводится проверка.

Оригинальный текст — фрагменты проверяемого текста, не обнаруженные ни в одном источнике и не отмеченные ни одним из модулей поиска. Показатель «Оригинальность» — это доля фрагментов проверяемого текста, отнесенных к оригинальному тексту, в общем объеме текста.

«Совпадения», «Цитирования», «Самоцитирования», «Оригинальность» являются отдельными показателями, отображаются в процентах и в сумме дают 100%, что соответствует полному тексту проверяемого документа.

Обращаем Ваше внимание, что система находит текстовые совпадения проверяемого документа с проиндексированными в системе источниками. При этом система является вспомогательным инструментом, определение корректности и правомерности совпадений или цитирований, а также авторства текстовых фрагментов проверяемого документа остается в компетенции проверяющего.

№	Доля в тексте	Источник	Актуален на	Модуль поиска	Комментарии
[01]	3,87%	не указано	13 Янв 2022	Цитирование	
[02]	3,2%	не указано	29 Сен 2022	Библиография	
[03]	0,73%	не указано	29 Сен 2022	Шаблонные фразы	Источник исключен. Причина: Маленький процент пересечения.
[04]	0,68%	СТРУКТУРА ЛАНДШАФТОВ И КАЧЕСТВЕННАЯ ОЦЕНКА ТЕРРИТОРИ...	09 Июн 2022	Модуль поиска "tsu"	
[05]	0,66%	Big date and social networks how it works. How to use the big data sy... https://halzen.ru	21 Окт 2022	Переводные заимствования по коллекции Интернет в английском сегменте	
[06]	0,66%	диплом.pdf	02 Июн 2022	Модуль поиска "tsu"	
[07]	0,58%	Programs for opening big files. How to open a .BIG file? How it's done https://halzen.ru	31 Окт 2022	Переводные заимствования по коллекции Интернет в английском сегменте	