

Прогнозирование трендов формирования предложения на рынке труда по открытым данным абитуриентов в социальных сетях

Азарнова Татьяна Васильевна

Воронежский государственный университет, Россия, e-mail: ivdas92@mail.ru

Дашкова Екатерина Сергеевна

Воронежский государственный университет, Россия, e-mail: dashkova-82@mail.ru

Дорохова Наталья Васильевна

Воронежский государственный университет, Россия, e-mail: nv_dorohova@mail.ru

Каширина Ирина Леонидовна

МИРЭА – Российский технологический университет, Москва, Россия, e-mail: kash.irina@mail.ru

Цитирование: Азарнова Т.В., Дашкова Е.С., Дорохова Н.В., Каширина И.Л. (2025). Прогнозирование трендов формирования предложения на рынке труда по открытым данным абитуриентов в социальных сетях. *Тerra Economicus* **23**(2), 43–61. DOI: 10.18522/2073-6606-2025-23-2-43-61

В статье исследованы тренды формирования предложения на рынке труда, в числе которых сокращение предложения, «старение» рабочей силы, усиление разрыва между профессионально-квалификационными характеристиками соискателей рабочих мест и имеющимися вакансиями. Использование методологии экономической социологии позволило выявить и проанализировать влияние качественных переменных на современное состояние и тенденции развития предложения на рынке труда, а также разработать аналитический инструментарий прогнозирования такого предложения. Доказана целесообразность использования социальных сетей как источника информации об осуществлении профессионального выбора. При помощи специальных методов получения открытой информации из маршрутных карт абитуриентов в социальных сетях и обработки ее широким спектром моделей машинного обучения построен аналитический инструментарий для выявления значимых социальных и поведенческих характеристик и прогнозирования будущего профессионального выбора. В статье приведены метрики качества по всем используемым моделям машинного обучения и осуществлен их сравнительный анализ. Результаты свидетельствуют о высоком качестве полученных инструментов прогнозирования, которые в дальнейшем можно использовать для регулирования предложения на рынке труда. Практическая значимость предложенного аналитического инструментария состоит в возможности повышения эффективности профориентационной работы, прогнозирования профессионального выбора абитуриентов, оптимизации распределения мест в вузах, в оценке индекса напряженности при поступлении на различные направления. Предложенный инструментарий позволяет также выявить, насколько планируемые показатели распределения бюджетных мест соответствуют востребованности образовательных направлений со стороны абитуриентов, и проанализировать действенность механизма информирования будущих студентов о тенденциях и перспективах развития ситуации на рынке труда.

Ключевые слова: рынок труда; предложение на рынке труда; профессиональный выбор; профориентация; аналитический инструментарий; машинное обучение

Predicting labor market trends through the analysis of open data from job applicants in social networks

Tatiana V. Azarnova

Voronezh State University, Russia, e-mail: ivdas92@mail.ru

Ekaterina S. Dashkova

Voronezh State University, Russia, e-mail: dashkova-82@mail.ru

Natalia V. Dorokhova

Voronezh State University, Russia, e-mail: nv_dorohova@mail.ru

Irina L. Kashirina

MIREA – Russian Technological University, Moscow, Russia, e-mail: kash.irina@mail.ru

Citation: Azarnova T.V., Dashkova E.S., Dorokhova N.V., Kashirina I.L. (2025). Predicting labor market trends through the analysis of open data from job applicants in social networks. *Terra Economicus* 23(2), 43–61 (in Russian). DOI: 10.18522/2073-6606-2025-23-2-43-61

The article examines trends in the supply of labor market, including the reduction of supply, “aging” workforce, and widening gap between professional and qualifications of job seekers and vacancies. Economic sociology methodology allowed to identify and analyze the influence of quality variables on current state and development trends of supply. Social networks are proved to be a valuable source of information about professional choice. Special methods for obtaining information from applicant’s route maps and processing with machine learning models allow to build analytical tools for identifying social and behavioral characteristics, predicting future choices. Quality metrics for machine learning models are presented and compared. Results indicate high quality of forecasting tools, which can be used to adjust supply. Practical significance lies in improving career guidance, prediction of applicants’ choice, optimization of the allocation of university places to students, assessment of admission stress index. Our analytical tools allow to assess how the planned indicators correspond with demand for education and to analyze the effectiveness of the mechanism for informing students about market trends.

Keywords: labor market; supply in the labor market; professional choice; career guidance; analytical tools; machine learning

JEL codes: J21, J24

Введение

Трудовые ресурсы являются основной созидательной силой экономики любого государства, их количественные и качественные характеристики определяют потенциал социально-экономического развития. Усиление трансформационных процессов во всех сферах общественной жизни в конце XX в. способствовало существенному повышению важности человеческого фактора в экономике. Сегодня именно человек выступает ключевым производственным ресурсом, позволяющим обеспечить конкурентоспособность и инновационное развитие экономики.

Современный этап развития рынка труда и формирования предложения на нем отличается высокой скоростью изменений в условиях масштабного роста количества информации. Это обусловлено распространением новых цифровых технологий предоставления и поиска информации.

Наша работа базируется на методологических подходах, применяемых в таких междисциплинарных областях исследования как социэкономика и экономическая социология (Этциони, 2002; Шабанова, 2005; Радаев, 2002). С позиции последней между экономической и социальной сферами существует тесная взаимосвязь, которая в современных условиях становится все более выраженной и сложной. Формирование предложения на рынке труда в рамках данного методологического подхода выступает зависимой переменной, а в роли формирующих ее независимых качественных переменных нами рассматриваются экономические, демографические, ценностные, психологические параметры.

Выбранная методология исследования позволила:

- во-первых, установить влияние качественных переменных на основные тренды формирования предложения на современном рынке труда;
- во-вторых, обосновать целесообразность включения качественных переменных в аналитический инструментарий прогнозирования предложения на рынке труда по открытым данным абитуриентов в социальных сетях;
- в-третьих, определить воздействие ценностных и поведенческих характеристик абитуриентов на формирование предложения на рынке труда в перспективе.

Цель работы состоит в исследовании трендов формирования предложения на рынке труда и разработке аналитического инструментария их прогнозирования по открытым данным абитуриентов в социальных сетях.

Основные тренды формирования предложения на современном рынке труда

В настоящее время как мировой рынок труда, так и большинство рынков труда развитых стран сталкиваются с проблемой дефицита рабочей силы, которая, по прогнозам исследователей, с течением времени будет обостряться. Среди причин – процесс цифровизации, рост благосостояния населения, демографический кризис, трансформация системы ценностей и трудового поведения молодого поколения и др. (Graetz et al., 2022; Fossen and Sorgner, 2019; Frey and Osborne, 2017; Балацкий, Екимова, 2023). Аналогичная ситуация складывается и на российском рынке труда, о чем свидетельствуют как статистические данные, так и многочисленные работы отечественных ученых (Кашепов и др., 2021; Ларионова и др., 2022; Тарасьев, 2020). Важнейшим изменением в динамике предложения труда стала смена повышательного тренда (2005–2016 гг.) понижательным (2017–2022 гг.) (Капелюшников, 2023: 7). На наш взгляд, понижательный тренд формирования предложения на рынке труда обусловлен следующими обстоятельствами:

1) демографический кризис, спровоцировавший количественное сокращение трудовых ресурсов и изменение их возрастной структуры (Капелюшников, 2023: 12). Так, в период с 2017 по 2022 г. сокращение численности рабочей силы составило 1361,2 тыс. чел. Ожидаемого эффекта для «смягчения» ситуации на рынке труда не обеспечила и пенсионная реформа 2018 г., которая смогла лишь частично компенсировать последствия демографического кризиса;

2) рост численности *NEET*-молодежи, доля которой по результатам эмпирических исследований в 2020 г. достигла 28% (Буланова, Артамонова, 2022). Представители данной категории отличаются отсутствием потребностей в обучении и трудовой деятельности, нигилистическим отношением к общепринятой системе моральных норм и ценностей (Zudina, 2017). Вероятность изменения их ценностных ориентиров и отношения к труду в будущем крайне низка, что создает существенные риски для формирования предложения на рынке труда;

3) специальная военная операция, в ходе которой часть трудоспособного населения была мобилизована (300 тыс. чел.), а часть эмигрировала за пределы страны (500–800 тыс. чел.¹).

Еще одним трендом формирования предложения на рынке труда, отмеченным зарубежными и отечественными исследователями (Кудиньш и др., 2022; Børing and Grøgaard, 2023; Borsch-Supan and Weiss, 2016; Калабина, Шадрина, 2022; Лукичев и др., 2023), является «старение» рабочей силы. Так, в Российской Федерации за период с 2008 по 2022 г. доля лиц старшего возраста (50 лет и старше) в составе рабочей силы увеличилась на 4,2 п.п. (с 23,8% до 28,0%) при одно-

¹ Три тенденции: как СВО и релокация меняют демографию. РБК. 29 января 2023. https://www.rbc.ru/spb_sz/29/01/2023/63ce99cc9a794787b4a2e9de (дата обращения: 04.04.2024)

временном сокращении доли молодых людей в возрасте до 35 лет на 7,3 п.п. (с 38,1 до 30,8%)². Тенденцию обусловили:

- 1) демографический кризис, упомянутый выше: за период с 2008 по 2022 г. доля лиц старшего возраста (50 лет и старше) в общей численности населения возросла на 3,8 п.п. (с 32,0% до 35,8%) при одновременном сокращении доли молодых людей в возрасте до 35 лет на 5,9 п.п. (с 47,0% до 41,1%);
- 2) повышение трудовой активности людей старшего возраста, что может быть обусловлено пенсионной реформой 2019–2028 гг., низким уровнем пенсий, а также увеличением продолжительности жизни (с 67,99 в 2008 г. до 72,73 года в 2022 г.).

К числу трендов формирования предложения как на мировом, так и на российском рынке труда относится усиление разрыва между профессионально-квалификационными характеристиками соискателей рабочих мест и имеющимися вакансиями. Такой дисбаланс выступает серьезным препятствием на пути внедрения цифровых технологий и перехода к инновационному развитию экономики. Обозначенную тенденцию отмечают в своих трудах многие отечественные ученые (Мирзабалаева, Антонова, 2023; Соболева, 2022; Симонова и др., 2023; Вольчик, Маслюкова, 2023). В свою очередь, Росстат зафиксировал показатель соответствия работы трудоустроенных выпускников 2019–2021 гг. по специальности, полученной в образовательной организации высшего образования, на уровне 72%.

Причины такой динамики заключаются в следующем:

- 1) дисбаланс рынка труда и рынка образовательных услуг, присутствующий в экономике Российской Федерации на протяжении длительного периода. Нивелирование данной проблемы позволит «смягчить» количественный дефицит рабочей силы за счет обеспечения сбалансированности спроса и предложения на рынке труда в профессионально-квалификационном срезе;
- 2) проблема профессионального выбора, вызванная недостаточно активной и масштабной профориентационной работой. Долгое время профориентационная работа в нашей стране не носила системного характера и выступала по большей части инструментом конкурентной борьбы организаций профессионального образования (высшего и среднего) за абитуриентов. Однако Приказ Министерства просвещения РФ от 31 августа 2023 г. № 650 «Об утверждении Порядка осуществления мероприятий по профессиональной ориентации обучающихся по образовательным программам основного общего и среднего общего образования» в перспективе позволит снизить остроту обозначенной проблемы.

Социальные сети как источник сведений о формировании предложения на рынке труда

По мнению ученых и практиков, важным инструментом решения проблемы формирования предложения на рынке труда является профориентация (Кудринская и др., 2020; Никулина, 2020). Анализируя факторы, определяющие профессиональные предпочтения в условиях цифровизации (Зинич и др., 2022), следует отметить влияние общения в социальных сетях на выбор направлений подготовки (для абитуриентов) и переобучения (для людей, уже имеющих профессию), выбор работы. Социальные сети могут служить навигатором профессионального выбора и источником агрегированной информации о нем.

Исследуется влияние социальных сетей на получение работы соискателями, приводится статистика, показывающая, что около половины занятых нашли работу благодаря социальным контактам (Calvó-Armengol and Matthew, 2004). На основе различных моделей авторы доказывают, что социальные сети не только способствуют эффективности поиска работы, но и дают некоторые представления о различиях в заработной плате и других видах неравенства среди работников.

В другом исследовании по итогам обработки результатов опросов автор делает вывод, что социальные сети, являющиеся неформальным источником данных о рынке труда, могут, с одной стороны, улучшить качество индивидуальной трудовой деятельности и общую эффективность, а с другой – способствовать усугублению неравенства на рынке труда (Schmutte, 2016). Иными словами, социальные сети уменьшают проблему доступа к информации, но в то же время укрепляют существующие модели социального расслоения.

В условиях роста интереса к исследованию социальных сетей важными становятся цифровые технологии, способные агрегировать информацию о поведении людей в социальных сетях и строить на ее

² Рабочая сила, занятость и безработица в России (по результатам выборочных обследований рабочей силы). 2022. Стат. сб. М.: Росстат, 2022, 151 с.

основе различные индикаторы профессионального соответствия, успешности в выбранной сфере деятельности, а также предсказывать профессиональный выбор (Shchekotin et al., 2021; Перова и др., 2023).

Прогнозирование трендов формирования предложения на рынке труда: разработка и апробация аналитического инструментария

Эффективность применения цифровых технологий (обработки больших данных, машинного обучения) при анализе и прогнозировании различных аспектов формирования предложения на рынке труда отражена в работах российских (Каширина и др., 2022) и зарубежных ученых. Следует особо выделить масштабное международное исследование 2018 г., ставившее задачу разработки концептуальных, методологических и организационных аспектов использования больших данных в аналитике рынка труда (Mezzanzanica and Mercori, 2018). В работе было обосновано, что современное состояние рынка труда характеризуется ростом количества информации в геометрической прогрессии, что связано с распространением интернет-сервисов. Продемонстрированы возможности технологии обработки больших данных для решения таких задач, как мониторинг динамики рынка труда, предсказание профессий будущего, прогнозирование востребованности профессиональных компетенций в ближайшие годы.

Крупное исследование источников и технологий получения информации абитуриентами при выборе профиля обучения было проведено *Digital*-агентством TAGREE³. Выделены основные ресурсы, которые приводят абитуриентов на сайт того или иного вуза (*VK* – с долей 50–80%, *Telegram*, *WhatsApp* – с долей 1–8%), и основные площадки, которые используют вузы для продвижения (*YouTube* – 17–45%). В работе Н.И. Архиповой, А.Л. Абаева, А.Г. Голова, М.Т. Гуриевой (Архипова и др., 2022) оценивается эффективность изучения характеристик абитуриентов через формирование их маршрутной карты потребителя в цифровой среде *Customer Journey Map (CJM)*, с помощью которой можно определить потенциальные точки пересечения с абитуриентом и стратегии эффективного воздействия для достижения требуемых целей.

В работе Г.В. Можаяевой, А.В. Слободской, А.В. Фещенко (Можаяева и др., 2017) статистическими методами исследуется информационный потенциал социальных сетей для выявления образовательных потребностей и психолого-педагогических особенностей школьников, а также для поиска одаренных детей. Статья коллектива авторов под руководством А.В. Фещенко (Фещенко и др., 2017) посвящена вопросам обнаружения вузами перспективных абитуриентов на основе анализа их профилей в социальных сетях. Исследование, проведенное А.А. Степаненко, К.С. Шиляевым, З.И. Резановым (Степаненко и др., 2018), раскрывает вопросы определения профессиональных интересов пользователей социальной сети «ВКонтакте» на основе текстов тематических групп и персональных страниц. Ряд работ посвящен изучению различных аспектов выявления профессиональных интересов респондентов на основе анализа их профилей в социальных сетях (Можаяева и др., 2018; Feshchenko et al., 2018; Васюков и др., 2018; Кашпур и др., 2021).

Таким образом, в настоящее время зарубежными и отечественными учеными активно применяются цифровые технологии для анализа и прогнозирования различных аспектов формирования предложения на рынке труда, а также изучения проблем в данной сфере.

В рамках настоящего исследования предложен аналитический инструментарий, позволяющий прогнозировать тренды формирования предложения на рынке труда на основе открытых данных абитуриентов в социальных сетях. Его разработка и применение подразумевают несколько этапов.

1. Сбор данных: с использованием *API* социальных сетей собирается информация о будущих абитуриентах.
2. Анализ профилей: собранные профили обрабатываются, чтобы выявить увлечения, предпочтения и другие данные абитуриентов, так или иначе указывающие на потенциальные направления в выборе профессии.
3. Прогнозирование будущей профессии: на основе анализа данных из социальных сетей с использованием алгоритмов машинного обучения строятся прогнозы будущих профессий абитуриентов (Азарнова и др., 2018).

³ Университеты через призму *digital*. Как абитуриенты выбирают вуз и как в этом помогают *digital*-инструменты. <https://vc.ru/marketing/791406-university-cherез-prizmu-digital-kak-abiturienty-vybirayut-vuz-i-kak-v-etom-pomogayut-digital-instrumenty> (дата обращения: 15.04.2024)

4. Сравнение с данными рынка труда: предпочтения абитуриентов, полученные на основе прогнозов, сопоставляются с требованиями и трендами на рынке труда, что позволяет определить соответствие выбранной профессии потребностям рынка труда.

5. Формирование рекомендаций: на основе анализа данных из социальных сетей абитуриентам предлагаются профессиональные направления, соответствующие их интересам и требованиям рынка труда.

6. Поддержка принятия решений: абитуриентам может быть предоставлена информация о перспективах развития в выбранной профессии, возможностях карьерного роста, уровне заработной платы и других важных аспектах, знание которых поможет им принять обоснованное решение.

7. Мониторинг и анализ результатов: отслеживаются дальнейшие действия абитуриентов и результаты их профессионального выбора, чтобы оценить эффективность использования информации из социальных сетей и внести необходимые коррективы в методику.

Предлагаемый подход позволит абитуриентам принять информированное решение о выборе профессии на основе их интересов, потребностей рынка труда и трендов формирования предложения в определенных областях. Важно отметить, что анализ формирования предложения на рынке труда с использованием информации из социальных сетей требует постоянного отслеживания и обновления данных, так как трендам свойственно быстро меняться.

Апробация описанного аналитического инструментария базировалась на анализе профилей абитуриентов в социальной сети «ВКонтакте» и включала следующие шаги.

1. Сбор данных из открытых профилей абитуриентов «ВКонтакте» с использованием *VK_API*.
2. Выделение групп абитуриентов:
 - а) школьники со склонностью к гуманитарным наукам;
 - б) школьники со склонностью к точным наукам (математика, информатика, физика, технические науки);
 - с) школьники со склонностью к естественным наукам (биология, химия, медицина);
 - д) другое (интерес к строительству, архитектуре).
3. Разработка методов для прогнозирования выбора абитуриентами образовательных программ.
4. Анализ результатов.

На третьем шаге были построены модели машинного обучения, предсказывающие наиболее вероятную предметную область и наиболее вероятное направление поступления абитуриента по его профилю из социальной сети «ВКонтакте».

Информация, собранная со страниц абитуриентов в социальной сети, включала пол, дату рождения, количество подписчиков, город проживания и родной город, список предпочтений (музыкальных, политических, религиозных и др.), отношение к алкоголю и курению, список подписок на различные сообщества. В течение 2019–2020 учебного года мы получили и структурировали данные о 30 тыс. одиннадцатиклассников из 37 крупнейших городов (с населением ≥ 300 тыс. чел.) разных регионов России. В сентябре–декабре 2020 г. профили выбранных школьников проанализировали повторно на предмет сведений о факультетах, на которые они поступили. Такую информацию указали 17 407 первокурсников, и на ее основе мы построили модели машинного обучения, на входе принимающие данные абитуриента из его открытого профиля в социальной сети, а на выходе формирующие прогноз выбора этим абитуриентом одного из направлений подготовки. Обучение моделей проводилось на 80% от всего объема исходных данных, а тестирование – на 20% данных, не участвовавших в процессе обучения модели. Для тестирования моделей формировались случайные стратифицированные (по направлениям подготовки) выборки.

Далее, опираясь на построенные модели машинного обучения, мы проанализировали значимость признаков, определяющих выбор абитуриентом того или иного направления подготовки.

В процессе разведочного анализа и подготовки данных производилось распределение факультетов, указанных в профилях «ВКонтакте» (около 2,5 тыс. наименований), по предметным областям, а также областям наук. К направлениям обучения, которые будет прогнозировать алгоритм, были отнесены: биология/медицина, география, химия, математика, техника и технологии, физика, филология, философия, искусство/культура, психология/педагогика, маркетинг/PR, юриспруденция, иностранные языки, история/политология, обществознание/социология,

менеджмент/управление, экономика/финансы, бизнес/предпринимательство, строительство и др. Предметные области, как было отмечено выше, были разделены на четыре группы: гуманитарные, точные, естественные, а также другое.

Для обработки данных использовался язык программирования *Python*, а непосредственная работа с данными производилась с помощью библиотеки *Pandas*. Так как различных сообществ социальной сети «ВКонтакте» в этой библиотеке оказалось представлено свыше 300 тыс., было принято решение использовать для анализа только те из них, что встречались хотя бы 100 раз, т.е. хотя бы 100 выбранных пользователей были на них подписаны (таких сообществ оказалось около 4 тыс.). На рис. 1 представлена схема сбора и обработки данных.



Рис. 1. Схема сбора и обработки данных для обучения моделей

Источник: составлено авторами.

Рассмотрим результаты разведочного анализа данных, осуществленного для решения первой задачи – предсказания предметной области, которую выберет абитуриент. Вначале был проведен первичный анализ признаков: так, например, исследовалась связь числа подписчиков в социальной сети и выбора абитуриента. Тест Краскела – Уоллиса показал, что есть различия в медианах групп. Медиана выборки по точным наукам отличается в меньшую сторону от медианы по естественным/гуманитарным, т.е. у абитуриентов, выбирающих точные науки, подписчиков часто меньше (рис. 2).

Также показательным признаком оказался город рождения. На рис. 3 видно, как отличаются предпочтения старшеклассников в разных городах. Так, в Самаре наибольшей популярностью пользуются точные направления, а в Иркутске и Казани – гуманитарные. Но в Иркутске, в отличие от Казани, совсем непопулярны естественные направления. Подобный анализ был проведен для каждого из 37 выбранных крупных городов.

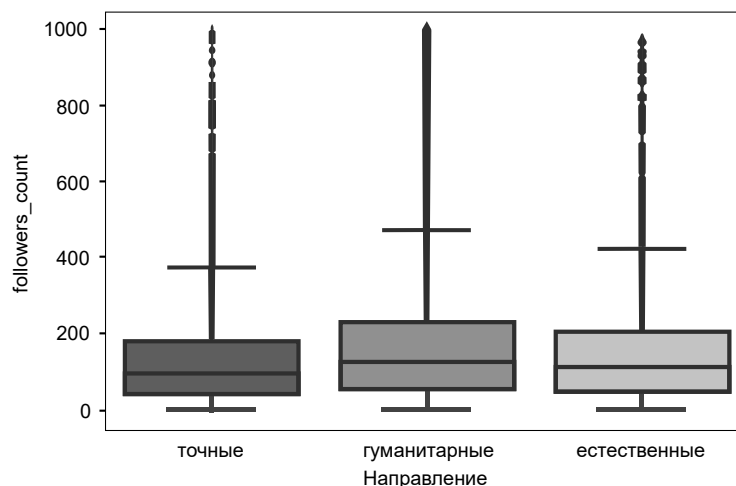


Рис. 2. Количественное распределение подписчиков по предметным областям

Источник: составлено авторами.

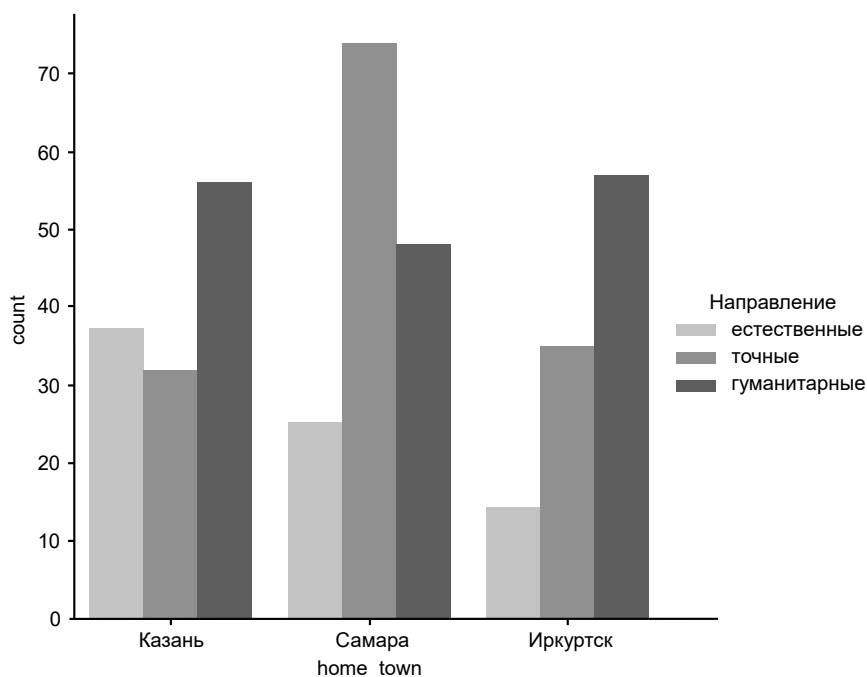


Рис. 3. Предпочтения абитуриентов в разных городах

Источник: составлено авторами.

Наравне с политическими, социальными и иными предпочтениями рассматривался и вопрос религиозной идентичности. Люди неохотно указывают свое отношение к религии, однако любопытно было отметить, что последователи ислама одинаково склонны выбирать все направления, в то время как православные больше склоняются к гуманитарным. Это можно наблюдать на рис. 4.

Гендерные особенности профессионального выбора представлены на рис. 5, из которого следует, что женщины более склонны к гуманитарным наукам, в то время как мужчины тяготеют к точным.

На рис. 6 заметно, что те, кто имеет негативное отношение к курению (0 на графике), чаще склоняются к гуманитарным направлениям.

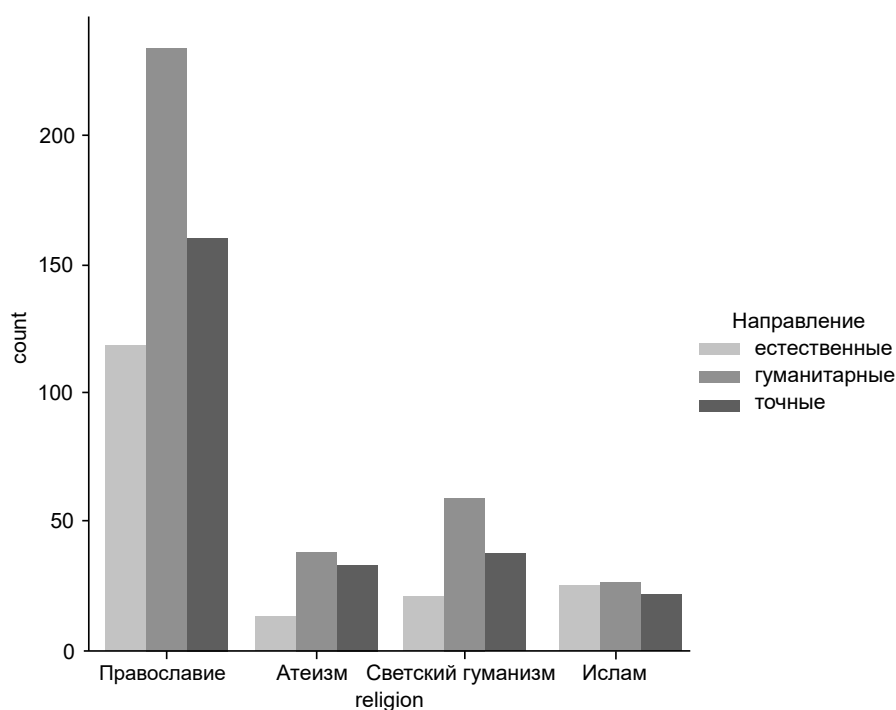


Рис. 4. Предпочтения абитуриентов, указавших отношение к религии

Источник: составлено авторами.

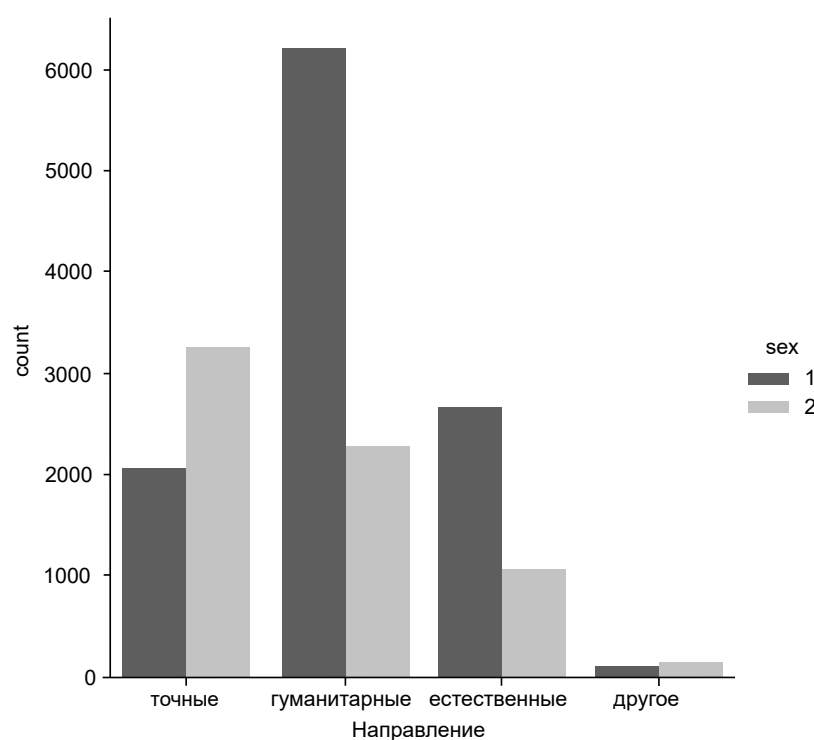


Рис. 5. Распределение предметных областей в зависимости от пола
(1 – женщины, 2 – мужчины)

Источник: составлено авторами.

На рис. 7 представлена диаграмма численности классов в исходных данных по предметным областям. Заметим, что существует дисбаланс классов: абитуриенты-гуманитарии встречаются гораздо чаще, чем абитуриенты, выбирающие точные и естественные науки. В связи с этим при

дальнейшем построении модели машинного обучения помимо классической метрики точности в задачах классификации *Accuracy* (доля правильных ответов алгоритма) был выбран нечувствительный к дисбалансу классов показатель метрики *AUC-ROC* (площадь под кривой ошибок) в качестве более объективной оценки точности модели. Кроме того, для улучшения предсказаний моделей была использована техника *imbalanced-learn* (He and Ma, 2013).

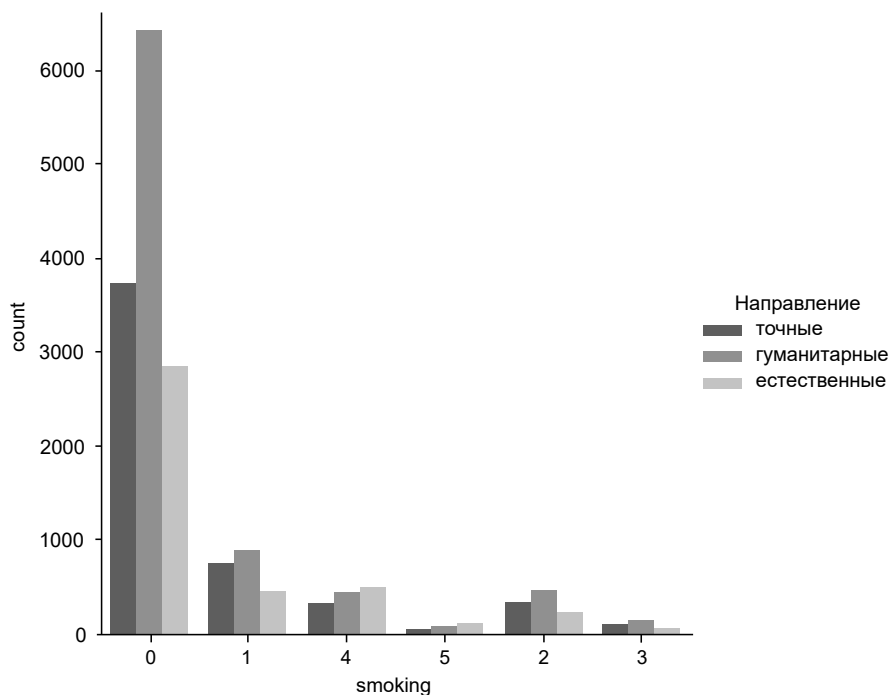


Рис. 6. Склонность к определенным наукам и отношение к курению

Источник: составлено авторами.

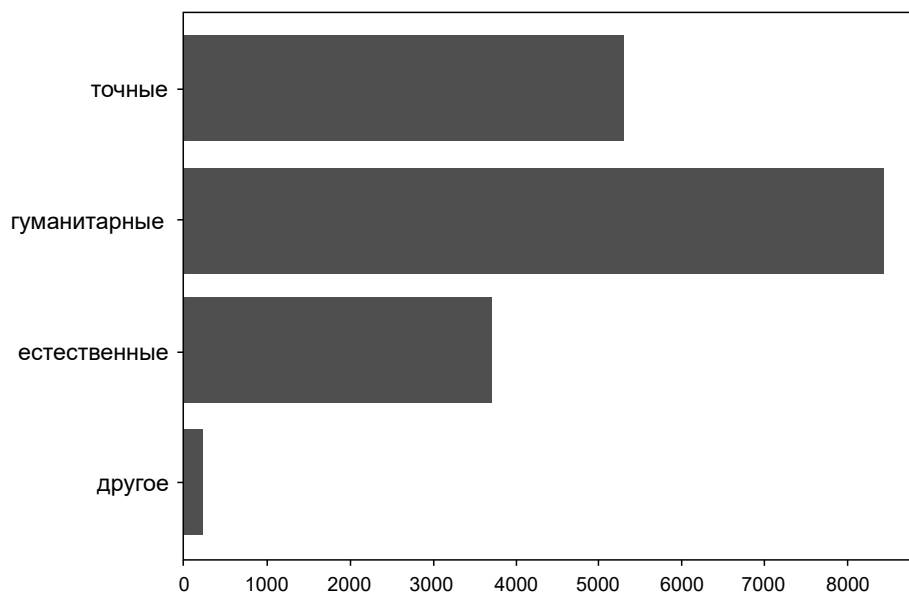


Рис. 7. Распределение целевого признака для предсказания предметной области

Источник: составлено авторами.

Для решения второй задачи – предсказания конкретного направления – также был произведен разведочный анализ данных. Распределение целевого признака продемонстрировано на рис. 8.

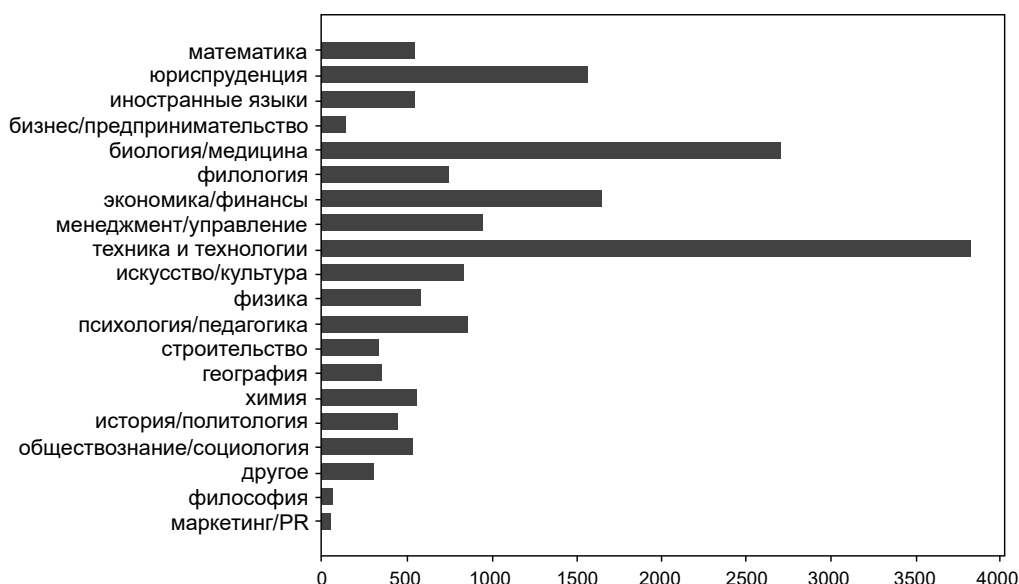


Рис. 8. Распределение целевого признака для предсказания конкретного направления

Источник: составлено авторами.

После сбора и анализа данных абитуриентов решалась задача прогнозирования предметной области при помощи моделей машинного обучения. Для проведения сравнительного анализа и выбора наилучшего варианта решения данной задачи использовались четыре наиболее популярные модели машинного обучения.

1. Логистическая регрессия (Плотников, Богатов, 2020). Это простая линейная модель для решения задач классификации, которая легко интерпретируется, быстро работает даже на больших объемах данных, но плохо подходит для прогнозирования сложных нелинейных зависимостей.

2. Случайный лес (Jin et al., 2024). Представляет собой независимый ансамбль деревьев решений. Имеет хорошую точность, устойчив к переобучению, может работать с нелинейными зависимостями, хорошо интерпретируется. Медленнее логистической регрессии, долго работает с большими объемами данных.

3. *XGBoost* (Iranzad and Liu, 2024). Представляет собой ансамбль деревьев решений, построенный по принципу градиентного бустинга, когда каждое дерево, добавляемое в ансамбль, старается исправить ошибки, сделанные ранее добавленными деревьями, с помощью различных градиентных оптимизационных процедур. Характеризуется очень высокой точностью, но сложной настройкой гиперпараметров.

4. *CatBoost* (Iranzad and Liu, 2024). Еще одна модификация метода градиентного бустинга, разработанная в компании Яндекс. Эффективнее всего работает с категориальными (нечисловыми) входными признаками. Имеет высокую точность, сравнимую с *XGBoost*, но проще в настройке.

Первой построенной моделью стала многоклассовая модель логистической регрессии. Для настройки точности этой модели был оптимизирован параметр ее регуляризации C . Регуляризация предотвращает эффект переобучения модели, т.е. снижает вероятность ухудшения ее обобщающей способности при применении к новым данным. Максимальное значение метрики точности *Accuracy* на тестовых данных составило 0,695 и было достигнуто при $C = 0,005$. При этом значение метрики качества *AUC-ROC* достигло 0,803, что свидетельствует о том, что предсказательная способность модели на новых данных (полученных из профилей абитуриентов, не использовавшихся при обучении модели) по данной метрике превысила 80%.

Следующей протестированной моделью машинного обучения был случайный лес (*RF*). Ансамблевые модели на большом объеме данных обучаются довольно долго, поэтому сначала с помощью метода кросс-валидации были найдены примерные диапазоны лучших значений гиперпараметров модели (количество деревьев леса, их максимальная глубина и т.п.), а в дальнейшем их наилучшая комбинация подбиралась с помощью построения кривых ошибок (*ROC-кривых*) (Feurer and Hutter, 2019). Лучшими гиперпараметрами модели случайного леса оказались: количество деревьев – 140, максимальная

глубина деревьев – 85, минимальное количество объектов в листе – 7. Метрика *Accuracy* при таких значениях параметров достигла значения 0,73, а *AUC-ROC* – 0,851, что выше, чем в модели логистической регрессии. Однако модель случайного леса очень плохо распознавала класс «другое».

Далее были построены две модели градиентного бустинга – *XGBoost* и *CatBoost*. В настоящее время модели градиентного бустинга демонстрируют наиболее высокую точность для большинства задач обработки табличных данных. Однако модель *XGBoost* показала результат метрики *AUC-ROC* равным 0,830, что хуже, чем аналогичный показатель в модели случайного леса. В то же время *CatBoost* оказалась лучшей моделью (показатель метрики *AUC-ROC* – 0,858, метрики *Accuracy* – 0,75). Важной положительной чертой этой модели оказалось то, что она неплохо предсказывала класс «гуманитарный» и отлично справилась с прогнозом класса «другое». Лучшими параметрами для обучения *CatBoost* оказались: глубина дерева – 5, параметр регуляризации – 3, шаг градиентного спуска – 0,05, число деревьев – 822.

В то же время выяснилось, что модель *CatBoost* на этапе обучения повышает качество предсказания класса «другое» в ущерб правильным предсказаниям класса «естественные». В связи с этим было принято решение объединить прогнозы модели *CatBoost*, хорошо предсказывающей класс «другое», и модели случайного леса, плохо предсказывающей класс «другое», но лучше предсказывающей класс «естественные». Прогнозы данных моделей были поданы на вход модели *XGBoost* в качестве признаков дополнительно к остальным переменным. *XGBoost* стала лучшей моделью с показателями метрик *AUC-ROC* = 0,88 и *Accuracy* = 0,79. На рис. 9 дана схема ее построения, в табл. 1 приведено сравнение точности всех построенных моделей.

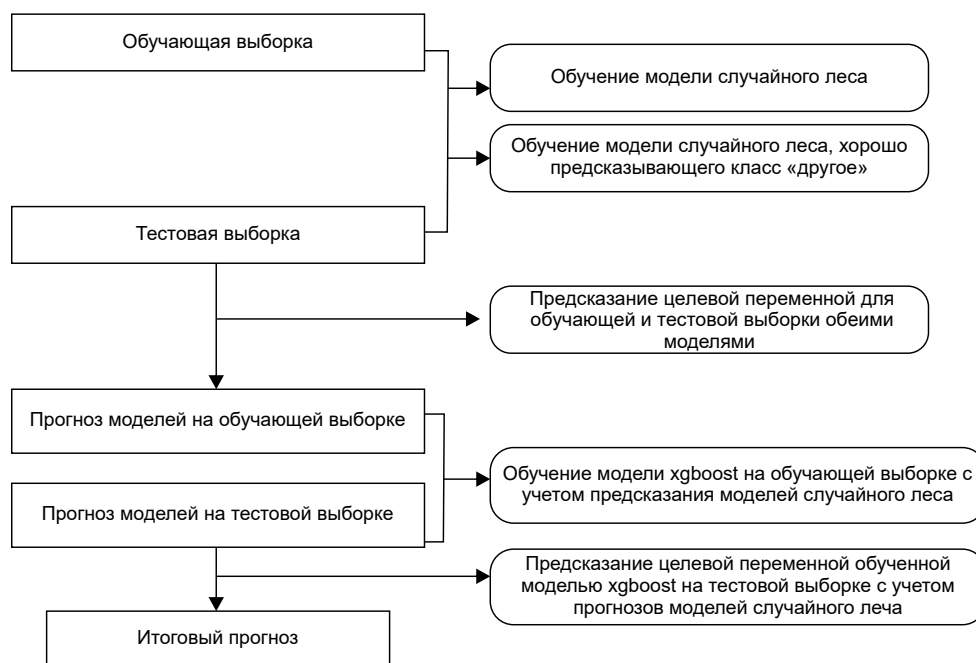


Рис. 9. Схема построения итоговой ансамблевой модели

Источник: составлено авторами.

Таблица 1

Сравнение моделей

Модель	<i>Accuracy</i>	<i>AUC-ROC</i>	«Другое»
Логистическая регрессия	0,695	0,803	–
Случайный лес (RF)	0,73	0,851	–
<i>XGBoost</i>	0,74	0,830	–
<i>CatBoost</i>	0,75	0,858	+
<i>XGBoost</i> + RF+ <i>CatBoost</i>	0,79	0,88	+

Источник: составлено авторами.

Далее решалась задача предсказания конкретного направления обучения, которое выберет абитуриент. Были построены и протестированы те же модели машинного обучения, и лучшей оказалась *CatBoost* со значениями гиперпараметров: глубина деревьев – 7, параметр регуляризации – 3, скорость обучения – 0,1, функция ошибок – *Multiclass*.

Подробный отчет о результатах обучения модели *CatBoost* представлен в табл. 2.

Таблица 2

Точность моделей предсказания направления обучения

Направление	<i>Accuracy</i>	<i>AUC-ROC</i>
Биология/медицина	0,674	0,732
География	0,345	0,534
Химия	0,419	0,541
Математика	0,595	0,703
Техника и технологии	0,701	0,778
Физика	0,643	0,718
Филология	0,638	0,711
Философия	0,315	0,531
Искусство/культура	0,654	0,731
Психология/педагогика	0,623	0,724
Маркетинг/PR	0,289	0,514
Юриспруденция	0,655	0,723
Иностранные языки	0,631	0,719
История/политология	0,607	0,708
Обществознание/социология	0,329	0,582
Менеджмент/управление	0,576	0,698
Экономика/финансы	0,589	0,701
Строительство	0,252	0,508

Источник: составлено авторами.

Однако некоторые направления, которые выбирало незначительное число абитуриентов (география, философия, строительство), плохо предсказывались моделью (в отличие от биологии и медицины, техники и технологии, юриспруденции). Таким образом, точность этой модели можно назвать удовлетворительной, что является основанием для ее дальнейшего совершенствования.

Для оценки работы модели и степени осознанности прогнозов алгоритма важно ответить на вопрос о том, какие признаки оказались первостепенными, или что влияет на профессиональный выбор будущего абитуриента. Так как решение второй задачи требует дополнительных исследований, анализу подвергались признаки, ставшие основополагающими для решения первой задачи.

В теории машинного обучения принято считать, что наиболее обоснованные веса важности признаков формирует метод случайного леса. При расчете этих весов для каждого признака оценивается его вклад в уменьшение информационной энтропии (неопределенности) значений прогнозируемой величины. Сумма весов всех признаков нормируется к единице.

Для модели случайного леса наиболее значимыми (внесшими максимальный вклад в прогнозирование направления подготовки) оказались признаки, связанные с гендерной принадлежностью, городом проживания, а также интересами. За последние отвечают подписки на различные тематические сообщества, посвященные наукам (в порядке убывания значимости: «Физика ЕГЭ», «Химия ЕГЭ», «Биология ЕГЭ», «Анатомичка», «Обществознание ЕГЭ», сообщество подготовки к ЕГЭ по химии и биологии *BuboUnicus* и т.д.); также важными оказались подписки на странички, выражающие интерес к определенным сферам («Биохимические мемы», «Исторические мемы», «Наука и техника», сообщество интернет-игр *IGM* и др.). Кроме того, весовыми признаками, влияющими на выбор абитуриента, оказались количество подписчиков, религиозные предпочтения и отношение к курению.

Как отмечалось выше, всего признаков оказалось около 4 тыс. (что связано с большим числом различных сообществ, представленных в подписках абитуриентов). В табл. 3 приведены веса наиболее важных признаков, выделенных моделью случайного леса.

Таблица 3

Веса значимости наиболее важных признаков

Признак	Вес	Признак	Вес
Пол	0,045	«Анатомичка»	0,019
Город проживания	0,035	«Обществознание ЕГЭ»	0,019
«Физика ЕГЭ»	0,030	<i>MDK</i>	0,015
«Химия ЕГЭ»	0,029	«Медицина, которую мы заслужили»	0,013
<i>BuboUnicus</i>	0,028	<i>IGM</i>	0,012
«Биология ЕГЭ»	0,026	«Информатика ЕГЭ»	0,011
«ЕГЭ история»	0,024	«Английский язык ЕГЭ»	0,009

Источник: составлено авторами.

Для модели *CatBoost* все перечисленные признаки также оказались важны, однако в числе особо значимых она выделила еще подписки на сообщества, связанные с изучением иностранных языков («Мой британский юмор», «Словарный запас» и др.), а также литературные сообщества («Бездна книжного уюта», «Журнал о блогерах», *Nothing Book* и др.).

На рис. 10 в качестве примера показано, какие предметные области выбрали для поступления подписчики сообщества «Мой компьютер». Видно, что информация о том, является ли абитуриент подписчиком сообщества, может оказаться полезной при прогнозировании его профессиональных предпочтений.

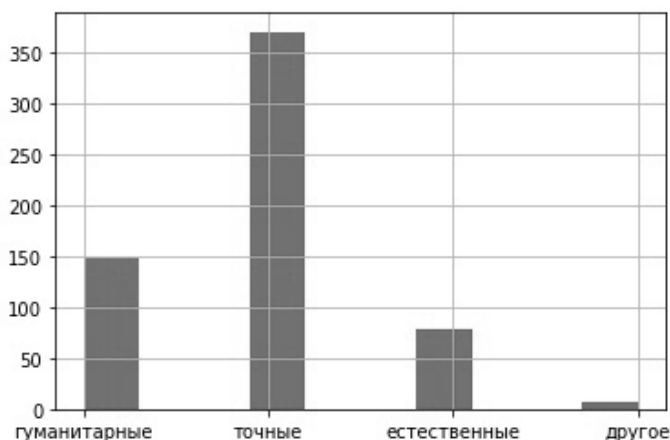


Рис. 10. Распределение предметных областей для подписчиков сообщества «Мой компьютер»

Источник: составлено авторами.

Проведенное исследование показывает, что сведения, представленные на странице абитуриента в социальной сети «ВКонтакте» в начале обучения в выпускном классе, позволяют оценить его склонность к той или иной сфере деятельности и с высокой точностью спрогнозировать предметную область и направление подготовки, которые он выберет при поступлении. Код программы доступен на портале *ResearchGate*⁴.

Перспективы использования предложенного инструментария

Предложенный аналитический инструментарий имеет следующие перспективы практического применения:

⁴ Прогнозирование трендов формирования предложения на рынке труда по открытым данным абитуриентов в социальных сетях. *ResearchGate*. DOI: 10.13140/RG.2.2.19083.81448/1

- 1) в процессе профориентационной работы – позволит выявить неочевидные закономерности и создать наглядные образы (профили) поведения абитуриента в социальных сетях по каждому классу профессионального выбора;
- 2) при прогнозировании профессионального выбора абитуриентов – даст возможность анализировать результативность используемых ранее и в текущее время стратегий профориентационной работы, направленных на решение проблем формирования предложения на рынке труда;
- 3) для оценки индекса напряженности при поступлении на различные направления на основе предикативной классификации профессионального выбора абитуриентов.

При помощи данного аналитического инструментария можно также отследить, насколько планируемые показатели распределения бюджетных мест соответствуют востребованности образовательных направлений со стороны абитуриентов, и проанализировать действенность механизма информирования будущих студентов о тенденциях и перспективах развития ситуации на рынке труда. Это может быть использовано для совершенствования системы распределения бюджетных мест в вузах, которая должна непосредственно опираться на планирование, прогнозирование и формирование потребности экономики в трудовых ресурсах. В масштабах крупных регионов планирование возможно только на базе специализированных информационных систем, включающих мощные инструменты сбора, обработки и анализа информации.

Заключение

Подытожив проведенное исследование, целью которого стало выявление трендов формирования предложения на рынке труда и разработка аналитического инструментария их прогнозирования по открытым данным абитуриентов в социальных сетях, мы сформулировали следующие выводы.

Использование методологических подходов, применяемых в экономической социологии, позволило расширить представления о детерминантах формирования предложения на рынке труда и обосновать важность таких качественных переменных, как экономические, демографические, ценностные, психологические. Были выявлены основные тренды формирования предложения на российском рынке труда, в числе которых сокращение предложения, «старение» рабочей силы, усиление разрыва между профессионально-квалификационными характеристиками соискателей рабочих мест и имеющимися вакансиями.

Разработан цифровой инструментарий, базирующийся на использовании данных социальных сетей и обработки их широким спектром моделей машинного обучения. Использование этого инструментария для решения задачи прогнозирования профессионального выбора молодежи осуществлено впервые в отечественной исследовательской практике.

Апробация цифрового инструментария показала, что сведения, представленные на странице абитуриента в социальной сети «ВКонтакте» в начале обучения в выпускном классе, позволяют оценить его склонность к той или иной сфере деятельности и с высокой точностью спрогнозировать предметную область и направление подготовки, которые он выберет при поступлении в высшее учебное заведение.

Перспективы применения данного инструментария состоят в его использовании для профориентационной работы с абитуриентами, прогнозирования их профессионального выбора и предикативной оценки конкурсной ситуации по различным направлениям подготовки.

Литература / References

- Азарнова Т.В., Каширина И.Л., Швиндт А.Н. (2018). Нейросетевое моделирование взаимодействия субъектов рынка труда и образовательных услуг. *Моделирование, оптимизация и информационные технологии* 6(4), 225–243. [Azarnova, T., Kashirina, I., Schwindt, A. (2018). Neural network modeling of interaction between subjects of the labor market and educational services. *Modeling, Optimization and Information Technologies* 6(4), 225–243 (in Russian)].
- Архипова Н.И., Абаев А.Л., Голова А.Г., Гуриева М.Т. (2022). Сопряжение СЖМ абитуриента с профилем вуза в цифровой среде как управленческая задача. *E-Management* 5(3), 106–116. [Arkhipova, N.,

- Abaev, A., Golova, A., Gurieva, M. (2022). Pairing the applicant's CJM with the profile of the university in the digital environment as a managerial task. *E-Management* 5(3), 106–116 (in Russian)].
- Балацкий Е.В., Екимова Н.А. (2023). Перспективы демографической экспансии России: экономика, институты, культура. *Terra Economicus* 21(2), 23–37. [Balatsky, E., Ekimova, N. (2023). Prospects for Russia's demographic expansion: Economics, institutions, culture. *Terra Economicus* 21(2), 23–37 (in Russian)].
- Буланова М.Б., Артамонова Е.А. (2022). NEET-молодежь: потребительское поведение в новой реальности. *Вестник Российского университета дружбы народов. Серия: Социология* 22(1), 113–125. [Bulanova, M., Artamonova, E. (2022). NEET-youth: Consumer behavior in the new reality. *Bulletin of the Peoples' Friendship University of Russia. Series: Sociology* 22(1), 113–125 (in Russian)].
- Васюков К.Л., Орлов С.А., Ошева М.С., Фещенко А.В. (2018). Университет в поисках своего абитуриента в социальных сетях: маркетинговые и технологические задачи. *Гуманитарная информатика* (14), 67–76. [Vasyukov, K., Orlov, S., Osheva, M., Feshchenko, A. (2018). The university in search of its entrant in social networks: Marketing and technological tasks. *Humanitarian Informatics* (14), 67–76 (in Russian)].
- Вольчик В.В., Маслюкова Е.В. (2023). Социальное государство и кадры для российской инновационной системы. *Журнал исследований социальной политики* 21(3), 449–466. [Volchik, V., Maslyukova, E. (2023). The social state and personnel for the Russian innovation system. *Journal of Social Policy Research* 21(3), 449–466 (in Russian)].
- Зинич А.В., Ревякина Ю.Н., Ревякин П.И. (2022). Молодежь на рынке труда в цифровую эпоху: социально-профессиональный аспект. *Экономика труда* 9(10), 1605–1616. [Zinich, A., Revyakina, Yu., Revyakin, P. (2022). Youth on the labor market in the digital age: socio-professional aspect. *Labor Economics* 9(10), 1605–1616 (in Russian)].
- Калабина Е.Г., Шадрина Е.А. (2022). Трансформация занятости возрастных работников в современной России: причины и последствия. *Экономика труда* 9(10), 1577–1590. [Kalabina, E., Shadrina, E. (2022). Transformation of employment of age-related workers in modern Russia: Causes and consequences. *Labor Economics* 9(10), 1577–1590 (in Russian)].
- Капелюшников Р.И. (2023). *Российский рынок труда: статистический портрет на фоне кризисов*. М.: Изд. дом Высшей школы экономики. [Kapelyushnikov, R. (2023). *The Russian Labor Market: A Statistical Portrait Against the Background of Crises*. Moscow: Publishing House of the Higher School of Economics (in Russian)].
- Кашепов А.В., Афонина К.В., Головачёв Н.В. (2021). Рынок труда РФ в 2020–2021 гг.: безработица и структурные изменения. *Социально-трудовые исследования* 43(2), 33–44. [Kashepov, A., Afonina, K., Golovachev, N. (2021). The Russian labor market in 2020–2021: Unemployment and structural changes. *Social and Labor Research* 43(2), 33–44 (in Russian)].
- Каширина И.Л., Азарнова Т.В., Бондаренко Ю.В. (2022). Разработка методов оценки эффективности человеческих ресурсов на основе алгоритмов глубокого обучения. *Инженерный вестник Дона* (2), 156–166. [Kashirina, I., Azarnova, T., Bondarenko, Yu. (2022). Development of methods for evaluating the effectiveness of human resources based on deep learning algorithms. *Engineering Bulletin of the Don* (2), 156–166 (in Russian)].
- Кашпур В.В., Петров Е.Ю., Гойко В.Л., Фещенко А.В. (2021). Возможности использования цифровых следов для прогнозирования образовательных достижений студентов. *Вестник Томского государственного университета. Философия. Социология. Политология* (64), 140–150. [Kashpur, V., Petrov, E., Goiko, V., Feshchenko, A. (2021). Possibilities of using digital traces to predict students' educational achievements. *Bulletin of Tomsk State University. Philosophy. Sociology. Political Science* (64), 140–150 (in Russian)].
- Кудиньш Я., Комарова В., Чижо Э., Кокаревича А. (2022). Влияние старения рабочей силы на производительность экономики. *Вестник Витебского государственного технологического университета* (1), 181–196. [Kudins, Ya., Komarova, V., Chizho, E., Kokarevich, A. (2022). The impact of an aging workforce on economic productivity. *Bulletin of the Vitebsk State Technological University* (1), 181–196 (in Russian)].

- Кудринская И.В., Кидинов А.В., Кабкова Е.П. и др. (2020). Проблема профориентации молодежи в отечественной педагогической теории и практике. *Евразийский журнал биологических наук* (14), 3815–3821. [Kudrinskaya, I., Kidinov, A., Kabkova, E. et al. (2020). The problem of youth career guidance in Russian pedagogical theory and practice. *Eurasian Journal of Biological Sciences* (14), 3815–3821 (in Russian)].
- Ларионова Н.И., Юрьева О.В., Бурганова Л.А. (2022). Рынок труда в условиях цифровой трансформации экономики. *Вестник экономики, права и социологии* (4), 90–97. [Larionova, N., Yurieva, O., Burganova, L. (2022). The labor market in the context of the digital transformation of the economy. *Bulletin of Economics, Law and Sociology* (4), 90–97 (in Russian)].
- Лукичѳ П.М., Чекмарев О.П., Конев П.А. (2023). Новые вызовы современного рынка труда: работники старших возрастов vs пенсионеры. *Экономика труда* **10**(4), 525–542. [Lukichev, P., Chekmarev, O., Konev, P. (2023). New challenges of the modern labor market: older workers vs pensioners. *Labor Economics* **10**(4), 525–542 (in Russian)].
- Мирзабалаева Ф.И., Антонова Г.В. (2023). Структурные диспропорции спроса и предложения на рынке труда в отраслевом и профессионально-квалификационном разрезах. *Экономика труда* **10**(8), 1145–1168. [Mirzabalayeva, F., Antonova, G. (2023). Structural imbalances of supply and demand in the labor market in the sectoral and vocational-qualification sections. *Labor Economics* **10**(8), 1145–1168 (in Russian)].
- Можаева Г.В., Суханова Е.А., Фещенко А.В. (2018). Привлечение и удержание университетами абитуриентов с высоким образовательным потенциалом с помощью анализа открытых пользовательских данных социальной сети «ВКонтакте». *Открытое и дистанционное образование* (4), 52–58. [Mozhaeva, G., Sukhanova, E., Feshchenko, A. (2018). Attracting and retaining university applicants with high educational potential by analyzing open user data of the VKontakte social network. *Open and Distance Education* (4), 52–58 (in Russian)].
- Можаева Г.В., Фещенко А.В., Слободская А.В. (2017). Информационный потенциал социальных сетей для выявления образовательных потребностей школьников. *Открытое и дистанционное образование* (3), 25–30. [Mozhaeva, G., Feshchenko, A., Slobodskaya, A. (2017). The information potential of social networks to identify the educational needs of schoolchildren. *Open and Distance Education* (3), 25–30 (in Russian)].
- Никулина Ю.Н. (2020). Профессиональная ориентация молодежи в системе кадрового обеспечения экономики региона. *Экономика, предпринимательство и право* **10**(4), 1263–1280. [Nikulina, Yu. (2020). Professional orientation of youth in the system of personnel support of the regional economy. *Economics, Entrepreneurship and Law* **10**(4), 1263–1280 (in Russian)].
- Перова Ю.П., Григорьев В.Р., Жуков Д.О. (2023). Модели и методы анализа сложных сетей и социальных сетевых структур. *Russian Technological Journal* **11**(2), 33–49. [Perova, Yu., Grigoriev, V., Zhukov, D. (2023). Models and methods for analyzing complex networks and social network structures. *Russian Technological Journal* **11**(2), 33–49 (in Russian)].
- Плотников Д.В., Богатов А.В. (2020). Обзор алгоритмов машинного обучения. *Информационные технологии: межвузовский сборник научных трудов*. Рязань: ИП Коняхин А.В. (Book Jet), с. 112–114. [Plotnikov, D., Bogatov, A. (2020). Review of machine learning algorithms. *Information Technologies. Collection of papers*. Ryazan: Konyakhin A.V. Publ. (Book Jet), pp. 112–114 (in Russian)].
- Радаев В.В. (2002). Еще раз о предмете экономической социологии. *Экономическая социология* **3**(3), 21–34. [Radaev, V. (2002). Once again about the subject of economic sociology. *Economic Sociology* **3**(3), 21–34 (in Russian)].
- Симонова М.В., Санкова Л.В., Мирзабалаева Ф.И. (2023). Стратегическое планирование кадрового обеспечения социально значимых отраслей экономики регионов. *Креативная экономика* **17**(8), 2815–2838. [Simonova, M., Sankova, L., Mirzabalayeva, F. (2023). Strategic planning of personnel support for socially significant sectors of the regional economy. *Creative Economics* **17**(8), 2815–2838 (in Russian)].
- Соболева И.В. (2022). Профессионально-квалификационный дисбаланс как вызов экономической и социальной безопасности. *Экономическая безопасность* **5**(3), 989–1008. [Soboleva, I.

- (2022). Professional and qualification imbalance as a challenge to economic and social security. *Economic Security* 5(3), 989–1008 (in Russian)].
- Степаненко А.А., Шиляев К.С., Резанова З.И. (2018). Атрибуция профессиональных интересов пользователей социальной сети «ВКонтакте» на основе текстов тематических групп и персональных страниц. *Вестник Томского государственного университета. Филология* (52), 130–144. [Stepanenko, A., Shilyaev, K., Rezanova, Z. (2018). Attribution of professional interests of users of the VKontakte social network based on texts of thematic groups and personal pages. *Bulletin of Tomsk State University. Philology* (52), 130–144 (in Russian)].
- Тарасьев А.А. (2020). *Оценка и прогнозирование развития российского рынка труда в условиях динамики трудовой миграции*. Диссертация на соискание ученой степени кандидата экономических наук. Екатеринбург. [Tarasyev, A. (2020). *Assessment and Forecasting of the Development of the Russian Labor Market in the Context of Labor Migration Dynamics*. A Dissertation Presented in Partial Fulfillment of the Requirements for the Degree of Candidate of Economic Sciences. Yekaterinburg (in Russian)].
- Фещенко А.В., Гойко В.Л., Степаненко А.А., Суханова Е.А., Мацута В.В., Киселев П.Б. (2017). Методы и инструменты выявления перспективных абитуриентов в социальных сетях. *Открытое и дистанционное образование* (4), 45–52. [Feshchenko, A., Goiko, V., Stepanenko, A., Sukhanova, E., Matsuta, V., Kiselev, P. (2017). Methods and tools for identifying promising applicants in social networks. *Open and Distance Education* (4), 45–52 (in Russian)].
- Шабанова М.А. (2005). Социэкономика и экономическая социология: точки размежевания и интеграции. *Экономическая социология* 6(5), 12–27. [Shabanova, M. (2005). Socio-economics and economic sociology: Points of separation and integration. *Economic Sociology* 6(5), 12–27 (in Russian)].
- Этциони А. (2002). Социэкономика: дальнейшие шаги. *Экономическая социология* 3(1), 65–71. [Etzioni, A. (2002). Socio-economics: Further steps. *Economic Sociology* 3(1), 65–71 (in Russian)].
- Børing, P., Grøgaard, J. (2023). Do older employees have a lower individual productivity potential than younger employees? *Journal of Population Ageing* 16, 369–397. DOI: <https://doi.org/10.1007/s12062-020-09323-1>
- Borsch-Supan, A., Weiss, M. (2016). Productivity and age: Evidence from work teams at the assembly line. *Journal of the Economics of Ageing* (7), 30–42.
- Calvó-Armengol, A., Matthews, O. (2004). The effects of social networks on employment and inequality. *American Economic Review* 94(3), 426–454.
- Feshchenko, A., Goiko, V., Matsuta, V. et al. (2018). Modelling of an educational profile of a student by analyzing public user data from social networks. *INTED 2018: Proceedings of the 12th International Technology, Education and Development Conference, Valencia*. Ed. by Gómez Chova, L., López Martínez, A., Candel Torres, I. Valencia: IATED Academy.
- Feurer, M., Hutter, F. (2019). Hyperparameter optimization. In: Hutter, F., Kotthoff, L., Vanschoren, J. (eds.) *Automated Machine Learning*. The Springer Series on Challenges in Machine Learning. Cham: Springer, pp. 3–33. DOI: https://doi.org/10.1007/978-3-030-05318-5_1
- Fossen, F., Sorgner, A. (2019). Mapping the future of occupations: Transformative and destructive effects of new digital technologies on jobs. *Foresight and STI Governance* 13(2), 10–18.
- Frey, C., Osborne, M. (2017). The future of employment: How susceptible are jobs to computerisation? *Technological Forecasting and Social Change* (114), 254–280.
- Graetz, G., Restrepo, P., Skans, O. (2022). Technology and the labor market. *Labour Economics* (76), 102177.
- He, H., Ma, Y. (2013). *Imbalanced Learning: Foundations, Algorithms, and Applications*. Wiley-IEEE Press.
- Iranzad, R., Liu, X. (2024). A review of random forest-based feature selection methods for data science education and applications. *International Journal of Data Science and Analytics*. DOI: <https://doi.org/10.1007/s41060-024-00509-w>
- Jin, K., Zhong, Z., Zhao, E. (2024). Sustainable digital marketing under big data: An AI random forest model approach. *IEEE Transactions on Engineering Management* (71), 3566–3579.

- Mezzanzanica, M., Mercurio, F. (2018). Big data enables labor market intelligence. In: Sakr, S., Zomaya, A. (a cura di). *Encyclopedia of Big Data Technologies*. Springer, pp. 1–11.
- Schmutte, I. (2016). How do social networks affect labor markets? *IZA World of Labor* 2016, 304.
- Shchekotin, E., Goiko, V., Myagkov, M., Dunaeva, D. (2021). Assessment of quality of life in regions of Russia based on social media data. *Journal of Eurasian Studies* **12**(2), 182–198.
- Zhou, Z. (2012). *Ensemble Methods: Foundations and Algorithms*. London: Chapman and Hall.
- Zudina, A. (2017). What makes youth become NEET? The evidence from Russian LFS. NRU Higher School of Economics Series Working Paper BRP Economics/EC. № WP BRP 177/EC/2017.