

УДК 331.554

DOI: 10.34670/AR.2026.58.55.011

Возможности применения инструментов машинного обучения и открытых датасетов для прогнозирования портрета соискателя на рынке труда Санкт-Петербурга

Кузнецов Юрий Викторович

Доктор экономических наук,
профессор кафедры управления и планирования
социально-экономических процессов,
Санкт-Петербургский государственный университет,
199034, Российская Федерация,
Санкт-Петербург, Университетская наб., 7-9;
e-mail: demteammates@gmail.com

Беляков Максим

Аспирант 3 курса,
Санкт-Петербургский государственный университет,
199034, Российская Федерация,
Санкт-Петербург, Университетская наб., 7-9;
e-mail: demteammates@gmail.com

Аннотация

В статье рассматриваются возможности использования методов анализа больших данных и машинного обучения для описания и прогнозирования характеристик соискателей на рынке труда. В качестве базы для исследования использованы данные цифровой платформы «Работа России» по соискателям из Санкт-Петербурга. Сформирован обширный датафрейм, включающий ключевые характеристики резюме: пол, возраст, стаж, образование, желаемую должность, условия занятости и др. Анализ выявил значительную неоднородность и смещённость выборки в сторону административных и неквалифицированных профессий, а также высокую долю незаполненных или нерелевантных данных. Проведено машинное обучение с целью предсказания отдельных характеристик соискателей на основе желаемой должности. Результаты демонстрируют высокую предсказательную способность по бинарным и слабо категоризированным признакам (тип занятости, готовность к переезду), но низкую точность предсказаний по признакам с большим числом уникальных значений (специальность, предыдущее место работы). Работа подчеркивает необходимость предварительной очистки и агрегации данных для повышения эффективности применения ИИ-технологий в задачах анализа рынка труда.

Для цитирования в научных исследованиях

Кузнецов Ю.В., Беляков М. Возможности применения инструментов машинного обучения и открытых датасетов для прогнозирования портрета соискателя на рынке труда Санкт-Петербурга // Экономика: вчера, сегодня, завтра. 2026. Том 16. № 4А. С. 115-124. DOI: 10.34670/AR.2026.58.55.011

Ключевые слова

Анализ рынка труда, большие данные, машинное обучение, соискатели, «Работа России», резюме, профессиональная мобильность, искусственный интеллект, Санкт-Петербург.

Введение

Современный рынок труда стремительно трансформируется под воздействием цифровых технологий, автоматизации и изменения структуры занятости. Эти процессы создают новые вызовы как для соискателей, так и для работодателей, требуя более точного и системного понимания структуры и динамики трудовых ресурсов. В этих условиях возрастает значимость аналитических инструментов, способных выявлять закономерности и фрикции на рынке труда, а также формировать обоснованные прогнозы относительно предпочтений и поведения соискателей.

Одним из перспективных направлений в этом контексте выступает использование методов машинного обучения и анализа больших данных, позволяющих на основе открытых источников статистики создавать модели, описывающие поведение участников рынка труда. В особенности, это актуально для оценки профессиональной мобильности, соответствия между квалификацией и предпочтениями соискателей и востребованностью тех или иных компетенций у работодателей.

Целью статьи является анализ состава и характеристик соискателей, зарегистрированных на цифровой платформе «Работа России», а также определение возможностей применения методов машинного обучения для предсказания их профессиональных и демографических признаков. В качестве фокусной территории рассматривается рынок труда Санкт-Петербурга – одного из крупнейших мегаполисов России, обладающего развитой экономикой и высокой концентрацией кадровых ресурсов.

Основная часть

Для определения специфических характеристик соискателей, представленных на рынке труда, могут быть использованы инструменты анализа больших данных и машинного обучения. Основным открытым источником статистики, описывающей характеристики соискателей, выступает сервис «Работа России», агрегирующий вакансии 1 192 002 работодателей и резюме жителей 89 субъектов Российской Федерации.

Рассматривая данные о ранее размещенных и актуальных резюме в Санкт-Петербурге, может быть выделен следующий датафрейм (набор данных), содержащий ряд характеристик о соискателях, включая их пол, возраст, стаж и предыдущий опыт работы, место и специальность обучения, предпочтительные условия труда (график, формат), а также иные характеристики, определяющие потенциал соискателей на рынке труда. В результате набор данных содержит характеристики о 689 666 соискателях, указавших желаемую должность и обладающих хотя бы минимальным набором характеристик, характеризующих соискателя (Таблица 1).

При этом важно отметить специфический характер соискателей, представленных в базе данных, связанный с преимущественной представленностью государственных и крупных

работодателей на платформе, осуществляющих массовый набор персонала на базовые позиции.

**Таблица 1 – Характеристики соискателей,
включенные в исследуемый датафрейм**

Номер	Характеристика соискателя	Число соискателей с указанием характеристики
1	Пол соискателя (gender)	689 708
2	Возраст соискателя (age)	689 708
3	Стаж соискателя (experience_age)	567 347
4	Категория в/у (categories)	689 708
5	Заработная плата (salary)	689 659
6	Факультет обучения (faculty)	689 707
7	Образовательное учреждение (institute_name)	599 542
8	Специальность (speciality)	316 696
9	Готовность к переезду (relocation)	689 707
10	График работы (employment_type)	689 707
11	Тип занятости (employment_status)	639 830
12	Готовность к переобучению (retraining)	689 707
13	Знание языков (languages)	689 707
14	Предыдущая профессия (job_title)	689 707
15	Предыдущее место работы (company_name)	599 348
16	Гражданство (country)	689 708
17	Желаемая должность (profession)	689 666

Источник: составлено авторами по данным сервиса «Работа России»

Из Рисунка 1 видно, что основными профессиями, на которые претендуют соискатели, данные которых представлены в датафрейме, относятся к административной, финансовой и иной обеспечивающей деятельности (администратор, менеджеры различных типов, бухгалтер, юрист), торговле (продавец-кассир, продавец) и других направлениях, относящихся к сфере услуг. Также в выборке представлено большое число соискателей, относящихся к неквалифицированной рабочей силе, ориентированных на поиск соответствующей деятельности – временное трудоустройство несовершеннолетних, общественные работы, подсобный рабочий и т.д.

Специфический характер выборки, представленной в датафрейме, может быть продемонстрирован и распределением числа соискателей по предыдущим профессиям, которые они занимали на предыдущих местах работы (Рисунок 2). Наиболее распространенными профессиями на предшествующих поиску местах работы являются бухгалтер, администратор, менеджер по продажам, менеджер, продавец-консультант. Наиболее широко в выборке представлены профессии, связанные с финансово-экономической и административной деятельностью, часто предусматривающие массовый наём персонала. Иные широко представленные профессии часто характеризуются низким квалификационным порогом (официант, менеджер, кладовщик, офис-менеджер), а также носят достаточно широкий характер (инженер, директор).

Тем не менее, несмотря на особенности, связанные с представленностью тех или иных профессиональных групп в выборке, стоит отметить её достаточно широкий характер – датафрейм содержит 78 371 уникальную запись, характеризующую желаемую соискателем должность, что позволяет использовать указанный набор данных для анализа рынка труда с ограничениями, связанными с имеющимися и указанными выше особенностями выборки.



Рисунок 1 – Распределение соискателей по характеристике «желаемая должность», топ-30 (число соискателей в датафрейме)

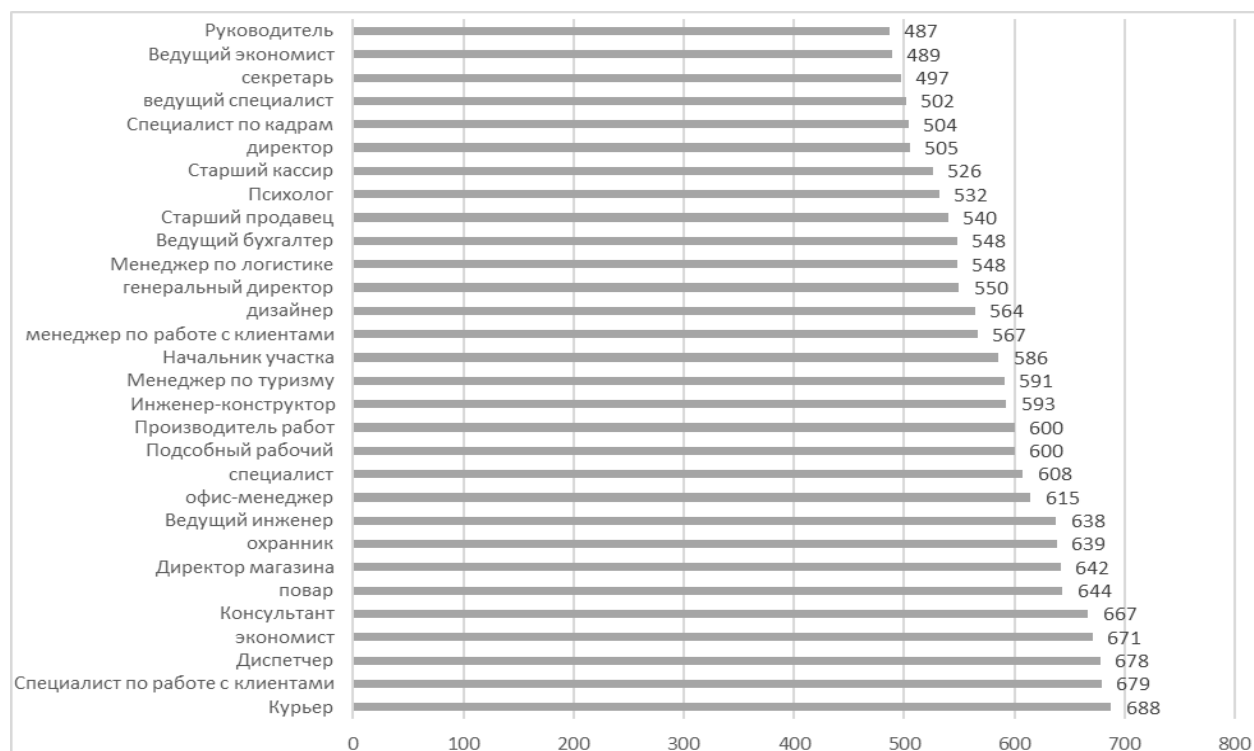


Рисунок 2 – Распределение соискателей по характеристике «предыдущая поиску работы должность», топ-30 (число соискателей в датафрейме)

Распределение соискателей по укрупненным направлениям обучения (характеристика faculty), включенных в датафрейм, также демонстрируют высокий уровень концентрации отдельных преимущественно социо-гуманитарных направлений. Наиболее распространены среди соискателей, включенных в выборку такие направления, как «Экономика и управление», «Юриспруденция», «Психология» (Рисунок 3). Это может как свидетельствовать о возможной специфике датафрейма, так и указывать на направления обучения, выпускники которых наиболее часто сталкиваются с необходимостью поиска работы в Санкт-Петербурге.

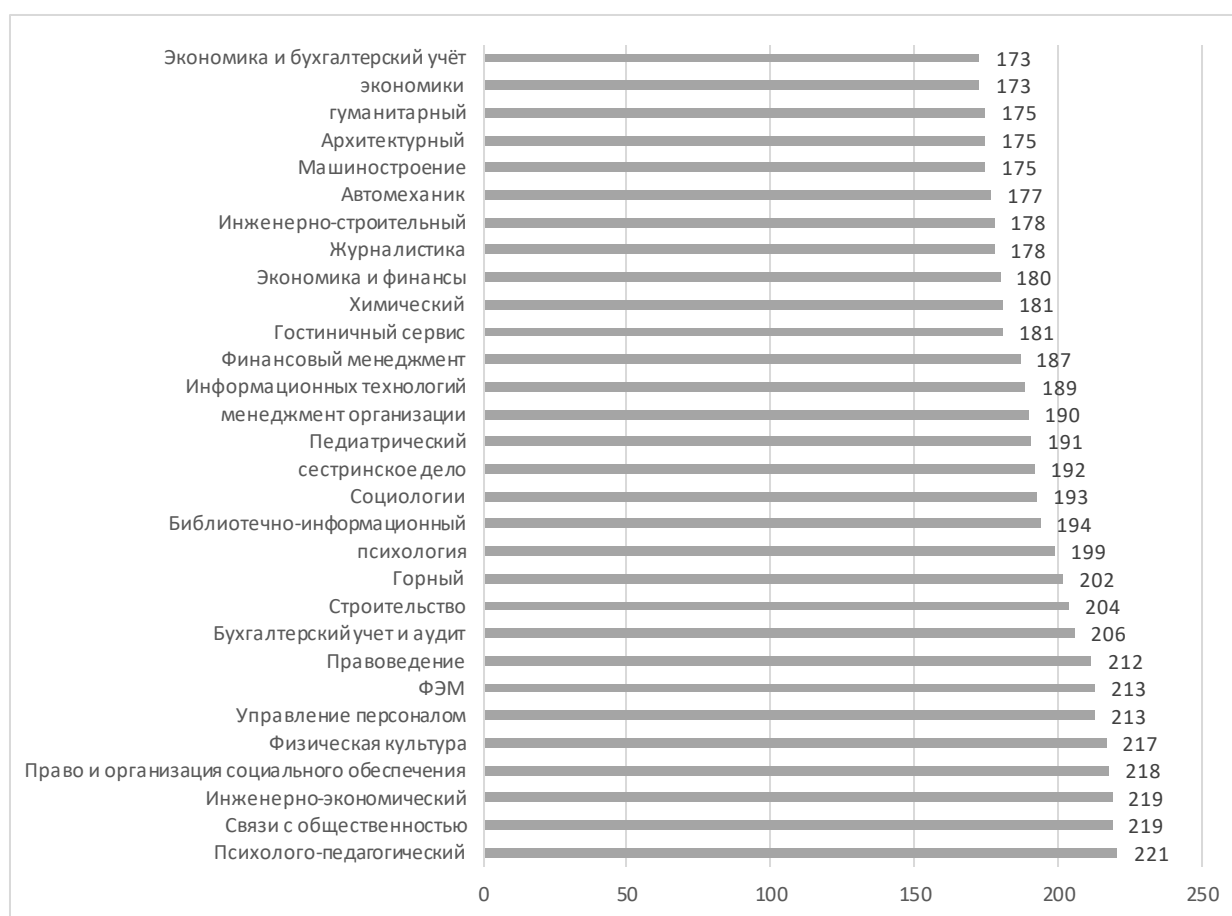


Рисунок 3 – Распределение соискателей по характеристике «факультет (укрупненное направление) обучения», топ-30 (число соискателей)

Аналогично можно выделить и образовательные учреждения, выпускники которых наиболее часто встречаются в датафрейме. Наиболее распространенными у соискателей учебными заведениями являются Санкт-Петербургский государственный университет им. Петра Великого (СПбПУ), Санкт-Петербургский государственный университет (СПбГУ), Российский государственный педагогический университет (РГПУ), Санкт-Петербургский государственный экономический университет (СПбГЭУ), а также Санкт-Петербургский государственный университет аэрокосмического приборостроения (ГУАП).

Неравномерность распределения вызвана неравномерным распределением выпускников из данных высших учебных заведений, связанных с различиями в объемах контингента и контрольных цифрах приема, однако также может свидетельствовать и о большей активности выпускников отдельных высших учебных заведений на рынке труда.



Рисунок 4 – Распределение соискателей по характеристике «образовательное учреждение», топ-30 (число соискателей по вузу)

Таким образом, можно выделить ряд особенностей датафрейма, ограничивающих его применимость для обучения моделей искусственного интеллекта, а именно:

а) отсутствие укрупнённой группировки по ключевым характеристикам соискателей, обеспечивающей снижение число уникальных значений и применимость данных для обучения моделей машинного обучения;

б) смещение выборки в пользу профессий и образовательных программ (факультетов, специальностей), связанных с административной деятельностью и сферой услуг;

в) наличие «пробелов» в заполнении резюме, а также граф «не указано».

В результате была получена следующая таблица, демонстрирующая результаты обучения моделей машинного обучения на данных соискателей, включенных в датафрейм (Таблица 2). В качестве объясняющей переменной для обучения моделей использовался параметр «Желаемая должность» (profession), а в качестве объясняемой – остальные характеристики (столбцы датафрейма) по отдельности. Данный подход позволил установить взаимосвязи между выбираемой соискателем должностью и другими характеристиками соискателя, которые могли бы свидетельствовать о наличии некоторых закономерностей на рынке труда, являющихся основой для возникновения фрикций перераспределения на рынке труда.

Таблица 2 – Результаты обучения моделей машинного обучения

Номер	Независимая переменная	Accuracy	F1-тест
y ₁	Пол соискателя (gender)	0,6913	0,6475
y ₂	Возраст соискателя (age)	MAE = 8,8074	
y ₃	Стаж соискателя (experience_age)	MAE = 5,2102	
y ₄	Категория в/у (categories)	0,6436	0,2763
y ₅	Заработная плата (salary)	MAE = 18 876,57	
y ₆	Факультет обучения (faculty)	0,3884	0,0022
y ₇	Образовательное учреждение (institute_name)	0,3920	0,0068
y ₈	Специальность (speciality)	0,0608	0,0086
y ₉	Готовность к переезду (relocation)	0,9291	0,4897
y ₁₀	График работы (employment_type)	0,6255	0,4673
y ₁₁	Тип занятости (employment_status)	0,8673	0,6537
y ₁₂	Готовность к переобучению (retraining)	0,7345	0,5536
y ₁₃	Знание языков (languages)	0,3255	0,0057
	Знание языков без «не указано»	0,2706	0,0092
y ₁₄	Предыдущая профессия (job_title)	0,1682	0,0033
	Предыдущая профессия без «не указано»	0,0752	0,0044

Источник: составлено авторами по данным сервиса «Работа России»

Из Таблицы 2 видно, что наилучшим образом на основе данных о профессии получается предсказывать бинарные характеристики (пол, готовность к переезду, готовность к переобучению), а также характеристики с небольшим числом потенциальных значений (категория водительского удостоверения, график работы, тип занятости и др.).

Наилучшим образом инструменты машинного обучения предсказывают готовность к переезду, предсказывая между значениями «Готов к переезду» или «Не готов к переезду» – точность модели составляет 0,93 (максимальное значение точности – 1), а также тип занятости, предсказывая такие значения, как «Временная занятость», «Полная занятость», «Частичная занятость», «Сезонная занятость», «Частичная занятость» и «Стажировка» – точность равняется 0,87.

Наиболее низкие значения точности демонстрируют характеристики, обладающие значительным числом уникальных значений, такие как «Специальность» (speciality) и «Предыдущая профессия» (job_title). Так, характеристика «Специальность» имеет 26 796 уникальных значений, а «Предыдущая профессия» – 121 390. Число уникальных значений по каждой из характеристик отражено в Таблице 3.

Таблица 3 – Число уникальных значений по характеристикам соискателя

Номер	Характеристика соискателя	Число уникальных значений в датафрейме
y ₁	Пол соискателя (gender)	2
y ₂	Возраст соискателя (age)	116
y ₃	Категория в/у (categories)	30
y ₄	Заработная плата (salary)	1 557
y ₅	Факультет обучения (faculty)	34 964
y ₆	Образовательное учреждение (institute_name)	80 281
y ₇	Специальность (speciality)	26 796
y ₈	Готовность к переезду (relocation)	2
y ₉	График работы (employment_type)	99
y ₁₀	Тип занятости (employment_status)	6
y ₁₁	Готовность к переобучению (retraining)	3
y ₁₂	Знание языков (languages)	4 297
y ₁₃	Предыдущая профессия (job_title)	121 390
x	Желаемая профессия соискателя (profession)	78 370

Источник: составлено авторами по данным сервиса «Работа России»

Данный фактор в условиях отсутствия группировки значений или иной разметки уникальных категорий существенно затрудняет машинное обучение, не позволяя с достаточно высокой точностью предсказывать значения, используя данные о желаемой должности, указанной в резюме. По этой причине отсутствие существенной и статистически значимой точности обучения моделей, предсказывающих данные параметры, не позволяют судить об отсутствии взаимосвязей между образованием и желаемой должностью соискателя, указанной в резюме, а также предыдущей работой и желаемой должностью соискателя, что могло бы свидетельствовать о высокой мобильности рабочей силы на рынке труда Санкт-Петербурга.

Заключение

Таким образом, видно, что в существующем виде датафреймы сервиса «Работа России» имеют ограниченную применимость для проектирования портрета соискателя для той или иной профессии. Тем не менее, дополнительная обработка данных, включенных в датафрейм, связанная с группировкой характеристик соискателей и устранением смещения выборки, должна обеспечить более высокую предсказательную способность моделей машинного обучения в предсказании характеристик соискателей.

Библиография

1. Беляков М. Фрикции перераспределения на рынке труда Санкт-Петербурга в условиях структурного дисбаланса кадров // Экономика и управление: проблемы, решения. 2023. Т. 4, № 6 (138). С. 78-88.
2. Кайсарова В.П., Беляков М.О. Влияние фактора образования на сбалансированность рынка труда Санкт-Петербурга // Менеджмент XXI века: формирование суверенной системы образования Российской Федерации: Материалы XXII международной научно-практической конференции, Санкт-Петербург, 22-23 ноября 2023 года. СПб.: РГПУ им. А.И. Герцена, 2024. С. 95-100.
3. Резник Г.А. Закономерности формирования и тенденции развития национального рынка труда в условиях транзитивной экономики: дис. ... д-ра экон. наук: 08.00.01. СПб., 1999. 324 с.
4. Reznikova O.S., Ganieva A.K., Verna V.V., Korolenko J.N., Shelygov A.V. Determinants of the Russian labor market model // Revista Inclusiones. 2020. Vol. 7. P. 260-267.
5. Stoikov V., Raimon R.L. Determinants of Differences in the Quit Rate among Industries // The American Economic Review. 1968. Vol. 58, No. 5. P. 1283-1298.

Possibilities of Applying Machine Learning Tools and Open Datasets for Predicting the Portrait of a Job Seeker in the Labor Market of Saint Petersburg

Yurii V. Kuznetsov

Doctor of Economic Sciences,
Professor of the Department of Management
and Planning of Socio-Economic Processes,
Saint Petersburg State University,
199034, 7-9, Universitetskaya nab., Saint Petersburg, Russian Federation;
e-mail: demteammates@gmail.com

Maksim Belyakov

3rd-year Postgraduate Student,
Saint Petersburg State University,
199034, 7-9, Universitetskaya nab., Saint Petersburg, Russian Federation;
e-mail: demteammates@gmail.com

Abstract

The article examines the possibilities of using big data analysis and machine learning methods for describing and predicting the characteristics of job seekers in the labor market. Data from the digital platform "Work of Russia" on job seekers from Saint Petersburg were used as the research base. An extensive dataframe was formed, including key resume characteristics: gender, age, work experience, education, desired position, employment conditions, etc. The analysis revealed significant heterogeneity and bias of the sample towards administrative and unskilled professions, as well as a high proportion of unfilled or irrelevant data. Machine learning was conducted in order to predict individual characteristics of job seekers based on the desired position. The results demonstrate high predictive ability for binary and weakly categorized features (type of employment, willingness to relocate), but low prediction accuracy for features with a large number of unique values (specialty, previous place of work). The work emphasizes the need for preliminary data cleaning and aggregation to improve the efficiency of applying AI technologies in labor market analysis tasks.

For citation

Kuznetsov Yu.V., Belyakov M. (2026) Vozmozhnosti primeneniya instrumentov mashinnogo obucheniya i otkrytykh datasetov dlya prognozirovaniya portreta soiskatelya na rynke truda Sankt-Peterburga [Possibilities of Applying Machine Learning Tools and Open Datasets for Predicting the Portrait of a Job Seeker in the Labor Market of Saint Petersburg]. *Ekonomika: vchera, segodnya, zavtra* [Economics: Yesterday, Today and Tomorrow], 16 (4A), pp. 115-124. DOI: 10.34670/AR.2026.58.55.011

Keywords

Labor market analysis, big data, machine learning, job seekers, "Work of Russia", resumes, professional mobility, artificial intelligence, Saint Petersburg.

References

1. Belyakov, M. (2023). Friksii pereraspredeleniya na rynke truda Sankt-Peterburga v usloviyakh strukturnogo disbalansa kadrov [Redistribution frictions in the St. Petersburg labor market under conditions of structural personnel imbalance]. *Economics and Management: Problems, Solutions*, 4(6), 78-88.
2. Kaisarova, V. P., & Belyakov, M. O. (2024). Vliyanie faktora obrazovaniya na sbalansirovannost rynka truda Sankt-Peterburga [The influence of the education factor on the balance of the St. Petersburg labor market]. In *Management of the 21st century: formation of a sovereign education system of the Russian Federation* (pp. 95-100). Herzen State Pedagogical University of Russia.
3. Reznik, G. A. (1999). *Zakonomernosti formirovaniya i tendentsii razvitiya natsionalnogo rynka truda v usloviyakh tranzitivnoy ekonomiki* [Patterns of formation and development trends of the national labor market in a transitive economy] (Doctoral dissertation).
4. Reznikova, O. S., Ganieva, A. K., Verna, V. V., Korolenko, J. N., & Shelygov, A. V. (2020). Determinants of the Russian labor market model. *Revista Inclusiones*, 7, 260-267.
5. Stoikov, V., & Raimon, R. L. (1968). Determinants of differences in the quit rate among industries. *The American Economic Review*, 58(5), 1283-1298.