Разработка алгоритма анализа вакансий на рынке труда по данным из открытых источников

Оксана Анатольевна Хохлова^{а)}, Александра Николаевна Хохлова⁶⁾, Аюна Цыденовна Чойжалсанова^{а)}

Во вводном разделе статьи авторы обосновывают актуальность разработки методического инструментария анализа вакансий на рынке труда в условиях современной технологической революции, существенно повышающей требования к профессиональным знаниям и опыту работающего персонала и меняющей соотношение между традиционными и новыми профессиями.

В основном разделе публикуемых результатов исследования для оценки текущей ситуации на рынке труда и востребованности ныне существующих профессий приведен алгоритм анализа вакансий по большим массивам данных из открытых источников при помощи математико-статистического инструментария и методов машинного обучения с использованием языка программирования Руthоп и аналитической платформы IBM SPSS modeler. Алгоритм включает в себя: парсинг данных о вакансиях, анализ вакансий по основным критериям, кластеризация вакансий по уровню заработной платы и построение нейросетевой модели — многослойного персептрона зависимости заработной платы от ряда предикторов. Следует отметить, что разработанный алгоритм, по мнению авторов, является универсальным, так как может быть использован для анализа больших данных из любого открытого источника (на определенный момент времени).

Результаты анализа позволят исследователям и специалистам управленческих структур более реалистично оценить текущую ситуацию на рынке труда, образовательным учреждениям корректировать программы обучения в соответствии с современными требованиями работодателей, работодателям принимать решения по вопросам развития компетенций в своей сфере деятельности и проводить сравнительный анализ имеющих спрос вакансий по количественным и качественным характеристикам, а соискателю увидеть востребованность вакансий на рынке труда и необходимость развития новых навыков.

Ключевые слова: рынок труда, статистика труда, математико-статистические методы анализа данных, методы машинного обучения, программный язык Python.

JEL: C10, C45, J01, J23, J44. *doi*: https://doi.org/10.34023/2313-6383-2022-29-4-33-41.

Для цитирования: Хохлова О.А., Хохлова А.Н., Чойжалсанова А.Ц. Разработка алгоритма анализа вакансий на рынке труда по данным из открытых источников. Вопросы статистики. 2022;29(4):33—41.

Development of an Algorithm to Analyze Vacancies in the Labor Market Based on Open-Source Data

Oksana A. Khokhlova^{a)}, Alexandra N. Khokhlova^{b)}, Ayuna T. Choyzhalsanova^{a)}

a) East Siberia State University of Technology and Management, Ulan-Ude, Russia;

In the introductory part of the article, the authors substantiate the relevance of developing methodological tools for analyzing job vacancies in the labor market in the context of the modern technological revolution, which significantly increases requirements for professional knowledge and experience of working personnel and changes the ratio between traditional and new professions.

To assess the current situation on the labor market and the demand for currently existing professions, the main section of the published results of the study presents the algorithm for analyzing vacancies using large data arrays from open sources using mathematical and statistical tools and machine learning methods using the Python programming language and the IBM SPSS modeler analytical platform. The algorithm includes: parsing data on vacancies, analyzing vacancies by the main criteria, clustering vacancies by salary level and building a neural network model — a multilayer perceptron of the dependence of salary on a number of predictors. It should be noted that the developed algorithm is universal, because it can be used to analyze big data from any open source at a certain point in time.

^{а)} Восточно-Сибирский государственный университет технологий и управления, г. Улан-Удэ, Россия;

^{б)} АО «Тинькофф Банк», г. Москва, Россия

b) JSC «TinkoffBank», Moscow, Russia

The results of the analysis will allow researchers and specialists of management structures to more realistically assess the current situation on the labor market, educational institutions will be able to adjust training programs in accordance with the modern requirements of employers, employers will make decisions on the development of competencies in their field of activity and conduct a comparative analysis of demanded vacancies in terms of quantitative and qualitative characteristics, and for the applicant it will be easier to see the demand for vacancies in the labor market and develop new skills.

Keywords: labor market, labor statistics, mathematical and statistical methods of data analysis, machine learning methods, Python programming language.

JEL: C10, C45, J01, J23, J44. *doi*: https://doi.org/10.34023/2313-6383-2022-29-4-33-41.

For citation: Khokhlova O.A., Khokhlova A.N., Choyzhalsanova A.T. Development of an Algorithm to Analyze Vacancies in the Labor Market Based on Open-Source Data. Voprosy Statistiki. 2022;29(4):33–41. (In Russ.)

Введение

Стремительное развитие технологий в секторах экономики и сопутствующие социальные изменения в долгосрочной перспективе приведут к прорывному росту экономики и высокой производительности труда, но в краткосрочном периоде могут привести к глобальной структурной безработице. Многие процессы будут автоматизированы или исчезнут из-за кардинальной смены экономического уклада, поэтому возникнут значительные сбои на рынке труда, и предприятия во всем мире могут столкнуться с рисками, вызванными нехваткой квалифицированных трудовых ресурсов. Жизненный цикл профессий будет сокращаться. В этой ситуации чрезвычайно важно знать, какие сейчас востребованы профессии, какие новые профессии и навыки ожидаются в будущем, чтобы своевременно принимать эффективные решения в подготовке кадров [1].

В целом государство должно комплексно оценивать потенциальные перспективы на рынке труда при обновлении своей образовательной стратегии, систематически прогнозировать эволюцию навыков в будущем для формирования профессиональных компетенций, актуальных на рынке труда.

В связи с этим одна из важных задач — это оценка текущей ситуации на рынке вакансий с помощью современного математико-статистического инструментария и методов машинного обучения, позволяющих провести обработку больших данных и выявить существующие стохастические связи и закономерности.

Методологические принципы формирования алгоритма анализа вакансий

В данном исследовании использовались такие средства анализа, как высокоуровневый язык программирования Python (версия 3.8.5) и аналитическая платформа IBM SPSS modeler, представляющая собой широкий набор инструментов интеллектуального анализа данных.

IBM SPSS modeler предлагает множество методов моделирования, в частности, методы машинного обучения, искусственного интеллекта и статистики. С помощью SPSS modeler можно быстро и интуитивно строить качественные прогнозные модели без программирования, используя уникальный визуальный интерфейс. А благодаря поддержке расширенной аналитики, встроенной в продукт, обнаруживать в данных ранее скрытые закономерности и тенденции¹.

Основные этапы алгоритма анализа вакансий представлены на рис. 1.

На первом этапе поставлены цель и задачи исследования, дано описание объекта и предметной области, позволяющее разработать алгоритм анализа и моделирования данных, выбрать методы и средства обработки исходной информации.

Второй этап — получение данных о вакансиях из открытых источников. Сбор данных отличается высокой трудоемкостью, здесь важны доступность и качество данных, их смысл и формат. В качестве открытого источника в исследовании выбран сайт вакансий на рынке труда — headhunter (hh.ru). По данным рейтинга Similarweb, это один из самых крупных

¹ URL: https://www.ibm.com/docs/en/spss-modeler/SaaS?topic=tutorial-about-spss-modeler.

сайтов по поиску работы и сотрудников в мире. Сайт обрабатывает до 3000 запросов в секунду. Каждый месяц на сайте появляется более полумиллиона вакансий, а технологии для работы с персоналом насчитывают более 30 позиций².

Постановка общей задачи исследования (актуальность, цель, задачи, видение, практическая значимость). Выбор методов и средств для проведения исследования



Разработка парсера на языке Python для получения данных о вакансиях из открытых источников (сбор данных, оценка сущности, характеристик и качества данных)



Подготовка данных (очистка данных, форматирование, обработка пропущенных данных и экстремальных значений, выбор данных для дальнейшего анализа) на языке Python



Обработка данных (трансформация и группировка данных, сводка, расчет новых показателей на основе исходных данных: средних, относительных величин и т. д.) на языке Python



Описательный статистический анализ вакансий по текущим количественным данным — визуализация результатов в виде аналитических графиков на языке Python



Кластеризация вакансий по заработной плате в разрезе специализаций при помощи метода k-средних с использованием IBM SPSS modeler



Построение нелинейной регрессионной модели — влияние предикторов на заработную плату при помощи нейросетевой модели — многомерного перцептрона в IBM SPSS modeler. Оценка качества модели



Интерпретация результатов исследования (оценка результатов анализа, оценка моделей, контроль)



Сопровождение и обновление алгоритма (обновление данных, обучающей и тестовых выборок и др.)

Рис. 1. Основные этапы алгоритма анализа

На этапе подготовки данных проводится очистка данных, их форматирование, обработка пропущенных данных и экстремальных значений на языке Python. На этом этапе возможна корректировка процесса сбора исходной информации, которая может привести к необходимости его повтора.

Далее необходима обработка данных для дальнейшего статистического анализа: трансформация и группировка данных, сводка, расчет новых показателей на основе исходных данных, таких как средние величины, относительные показатели и т. д. на языке Python.

На следующем этапе проводится анализ базы данных о вакансиях по городам, работодателям, графику работы, ежемесячной заработной плате. Результаты анализа позволяют увидеть целостную картину состояния рынка вакансий на текущий момент времени.

Анализ данных с сайта hh.ru заключается в изучении вакансий в разрезе разных сфер деятельности, таких как автомобильный бизнес; административный персонал; банки, инвестиции, лизинг; безопасность; бухгалтерия, управленческий учет, финансы предприятия; менеджмент; государственная служба, некоммерческие организации; маркетинг, реклама, PR; медицина, фармацевтика; наука, образование; продажи и др. по следующим характеристикам:

- уровень заработной платы;
- профобласть;
- специализация;
- условия работы: полный день, удаленная работа, гибкий график, сменный график, вахтовый метод;
 - город.

Также в описании вакансии рассматриваются компетенции и умения, выделяются ключевые навыки.

Далее проводится статистический анализ вакансий по основным количественным показателям и визуализация результатов в виде аналитических графиков на языке Python.

Затем осуществляется кластеризация вакансий по размеру ежемесячной заработной платы в разрезе специализаций и построение многослойного персептрона, характеризующего зависимость заработной платы от ряда предикторов при помощи аналитической платформы IBM SPSS modeler.

² URL: https://hh.ru/article/28?from=footer_new.

Интерпретация результатов исследования — важный этап исследования, на котором дается оценка результатов анализа, построенных моделей, описание дальнейших действий для принятия решений.

Сопровождение алгоритма. На данном этапе необходимо оценивать качество принимаемых решений, проводить обновление исходных данных, повторное обучение моделей, подбирать выборки и т. д.

Этапы носят итеративный характер, т. е. возможен возврат к каждому этапу, дополнение и их повторение.

Следует отметить, что ввиду больших массивов исходных данных как количественных, так и описательных, статистические методы и методы машинного обучения позволяют быстро и качественно провести комплексный анализ рынка вакансий по различным сферам деятельности на любой момент времени.

Описание парсинга вакансий

Парсинг — это процесс сбора данных, их обработки и систематизации. Этот метод используется тогда, когда нужно собрать и обработать большой массив информации. Программа, которая собирает и анализирует данные, называется парсером³.

Для написания парсера о вакансиях использовался язык программирования Python (версия 3.8.5). Программные коды разрабатывались в интегрированной среде программирования Python — Anaconda и в среде программирования PyCharm [2].

Используемые библиотеки Python⁴:

- os данная библиотека позволяет взаимодействовать с операционной системой: узнавать/ менять файловую структуру, переменные среды, узнавать имя и права пользователя и др.;
- requests библиотека, выполняющая http-запросы, позволяет легко и с минимальным количеством кода взаимодействовать с веб-приложениями;
- pandas библиотека предназначена для обработки и анализа данных, оформления данных в табличную форму;
- питру библиотека поддерживает матрицы, многомерные массивы, а также высокоуровневые математические функции, работающие с многомерными массивами.

Разработанный в исследовании парсер позволил получить информацию о вакансиях с $hh.ru^5$ из 738 286 записей. В результате обработки и очистки набор данных составил уже 688 382 записи, а число атрибутов (компонентов) — 29.

Описательный статистический анализ вакансий

Далее на языке Python проводился статистический анализ по следующим параметрам: города, высокооплачиваемые профессии, работодатели, график работы. Были использованы такие статистические методы, как сводка и группировка, ранжирование значения показателей, статистические графики.

На рис. 2 представлены результаты анализа топ-10 городов по числу вакансий. Москва и Санкт-Петербург занимают доминирующее положение, соответственно, 150 024 и 69 706 вакансий на рассматриваемый момент времени. Далее идет Екатеринбург с 20 828 вакансиями, затем Новосибирск и Краснодар — более 17 000 вакансий (17 857 и 17 004, соответственно), Казань — 15 441, Нижний Новгород — 15 254 и т. д. Топ-10 городов замыкает Уфа с 10 955 вакансиями.

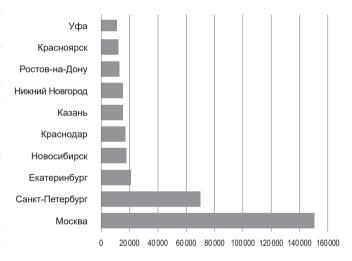


Рис. 2. Топ-10 городов по количеству вакансий

По высокооплачиваемым профессиям выделились следующие вакансии (см. рис. 3): коммерческий директор (автоматизация процессов хранения, обработки и передвижения грузов) с ежемесячной заработной платой 1 200 000 руб.; торговый представитель по продаже гуминовых удобрений, франчайзи-партнер, руководитель отдела

³ URL: https://ipipe.ru/info/parsing#

⁴ URL: https://habr.com/ru/post/481432/

⁵ Анализ проводился по данным с hh.ru на 10.05.2021 г.

продаж — 1 000 000 руб.; региональный управляющий — 800 000 руб.; директор данных инвестиционно-финансового холдинга, специалист по продаже элитной недвижимости — 700 000 руб. и т. д. Замыкает топ-15 высокооплачиваемых про-

фессий брокер по работе с клиентами, менеджер по работе с ключевыми клиентами, генеральный директор с зарплатой в 600 000 руб. в месяц. Анализ показал, что в основном высокооплачиваемые вакансии представлены в сфере услуг.

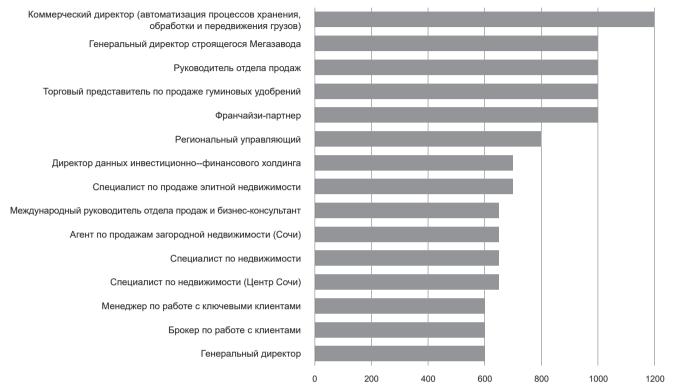


Рис. 3. Топ-15 высокооплачиваемых вакансий (тыс. рублей)

На рассматриваемую дату больше всего вакансий было выставлено на hh.ru от следующих организаций (см. рис. 4): Сбер — 5168, ЦентрПроф — 5073, Тинькофф — 4003, Росгосстрах — 3275, JCat.ru — 2666, Яндекс — 2606 и т. д. В этом же списке 2-й Специальный полк полиции ГУ МВД России по г. Москве — с 2272 вакансиями.

На рис. 5 показано распределение вакансий по занятости (полный рабочий день, удаленная работа, гибкий график, сменный график, вахтовый метод). Большинство вакансий, 498 175, — на полный рабочий день, со сменным графиком работы — 83 600, вахтовый метод — 36 285, удаленная работа — 35 799, гибкий график — 33 371.

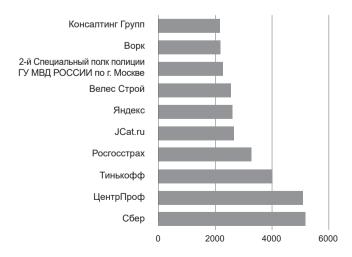


Рис. 4. Распределение вакансий по работодателям (топ-10)

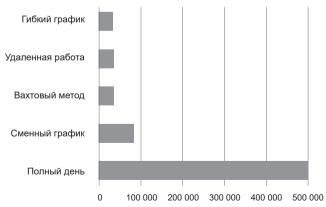


Рис. 5. Распределение вакансий по типу занятости

Проведенный статистический анализ показал, что большая часть вакансий на hh.ru с высокой заработной платой представлена в сфере услуг, в основном в компаниях, внедряющих цифровые технологии, методы искусственного интеллекта; в центральных городах — Москва, Санкт-Петербург, Екатеринбург, Новосибирск, Казань и др.; график работы — полный рабочий день.

Анализ заработной платы по данным о вакансиях: кластеризация и построение нейросетевой модели при помощи IBM SPSS modeler

Для общей характеристики вакансий, охватывающей множество признаков, отобран показатель «Ежемесячная заработная плата». На наш взгляд, данный показатель отражает ценность работника и его конкурентоспособность на рынке труда, а со стороны работодателя возможность выбрать лучшего кандидата на определенную вакансию. Поэтому далее в работе проанализирован выбранный показатель на примере вакансий специализации «Аналитик» в среде IBM SPSS modeler, а именно, проведена кластеризация вакансий по ежемесяч-

ной заработной плате и построена нейросеть в виде многослойного персептрона [3-5].

Исходные данные по специализации «Аналитик» представляют собой массив данных из 1560 вакансий на рассматриваемую дату, каждая из которых описывается следующими характеристиками: название вакансии, описание вакансии (навыки, умения, компетенции), размер заработной платы, страна, город, условия работы (режим работы), работодатель.

Модель анализа в среде SPSS modeler включает следующие связанные между собой узлы: «Загрузка данных из текстового файла», «Выбор набора записей на основе заданного условия», «Определение метаданных полей и управление ими», «Автокластер: создание и сравнение моделей кластеров», «Модель кластеризации», «Построение нейронной сети», «Модель нейросети», «Анализ точности модели», «Вывод данных в таблицу».

Разработанный в исследовании сценарий потока SPSS modeler (см. рис. 6) дополняет алгоритм анализа вакансий по заработной плате, представленный на рис. 1. В сценарии пошагово отражен весь процесс моделирования, начиная с импорта данных и заканчивая выводом результатов по кластеризации и нейросети.

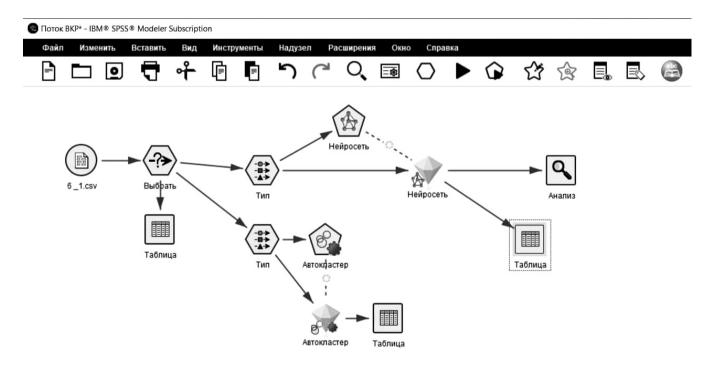


Рис. 6. Сценарий потока SPSS modeler «Модель анализа заработной платы по данным о вакансиях из открытого источника hh.ru»

 $^{^{6}}$ Анализ проводился по данным с hh.ru на $10.05.2021~\mathrm{r}$.

Кластерный анализ вакансий

Кластерный анализ вакансий проведен методом k-средних, в результате которого было получено 5 групп вакансий по размеру заработной платы (см. рис. 7).

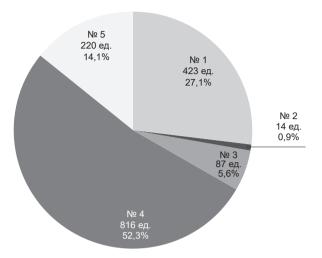


Рис. 7. Размеры кластеров

Таким образом, в кластер № 4, размером 52,3% от всех вакансий, или 816 единиц, попали вакансии в интервале 15 400 — 67 086 руб. со средней заработной платой 41 112 руб. и медианой — 40 302 руб.; соответственно, в кластер № 1, размером 27,1%, или 423 вакансии: 67 086 — 132 755 руб., 84 788 руб., 80 111 руб.; в кластер № 5, размером 14,1%, или 220 вакансий: 132 755 — 199 924 руб., 136 168 руб., 139 876 руб.; кластер № 3, размером 5,6%, или 87 вакансий: 199 924 — 240 000 руб., 191 390 руб., 200 012 руб.; кластер № 2, размером 0,9%, или 14 вакансий: 240 000 — 280 000 руб., 250 714 руб., 249 934 руб. Размер наибольшего кластера в 52,29 раза превышает размер

наименьшего кластера. По результатам кластеризации вакансий можно дать подробное описание состава каждого кластера.

Качество кластеризации хорошее, о чем свидетельствует силуэтная мера связности и разделения кластеров, равная 0,682. Распределение вакансий внутри каждого кластера близко к нормальному закону распределения.

Нейросеть - многослойный персептрон

Далее в работе построена предиктивная регрессионная модель в виде нейросетевой модели при помощи IBM SPSS modeler (см. таблицу 1). Построенная нейросетевая модель представляет собой многослойный персептрон (MLP) с одним скрытым слоем и 15 нейронами (см. рис. 8) [6].

Таблица 1 **Характеристика нейросети**

| Группа вакансий | «Аналитик» | |
|--|-------------------------------------|--------------------------------|
| Нейросеть | Многослойный персептрон | |
| Целевой параметр | Ежемесячная заработная плата | |
| Входные параметры (предикторы) | Название вакансии | Важность предиктора 0,27 |
| | Работодатель | 0,24 |
| | Описание вакансии (навыки и умения) | 0,23 |
| | Город | 0,20 |
| | Условия работы | 0,05 |
| Нейроны скрытого слоя | 15 | |
| Обучающая выборка | 73% | |
| Тестовая выборка | 27% | |
| Линейный коэффициент корреляции между предсказанными и наблюдаемыми значениями | 0,751 | |

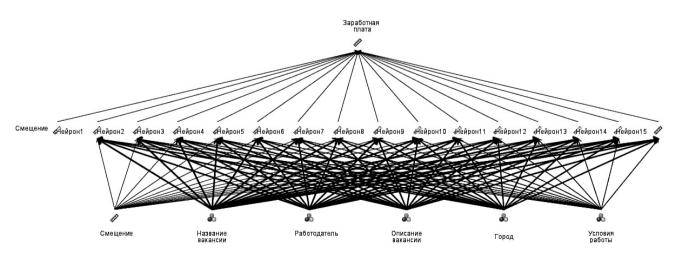


Рис. 8. Многослойный персептрон

Целевым параметром явилась заработная плата, входными параметрами — предикторами, представляющими основные характеристики вакансии за рассматриваемый период: название вакансии, описание вакансии, город, условия работы, работодатель. Для обучения нейросети было отобрано 1140 (73%) записей, остальные 420 вакансий (27%) — тестовая выборка.

В итоге была рассчитана важность каждого отобранного предиктора, в той или иной мере влияющего на размер заработной платы (см. рис. 9). Наиболее важными явились: название вакансии (должность) и работодатель (компания). Наименее важный фактор — условия работы (полный рабочий день, удаленная работа и т. д.), это говорит о том, что режим работы особо не влияет на размер заработной платы.

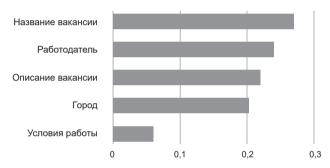


Рис. 9. Важность предикторов нейросети

Количество и состав предикторов зависит от имеющейся базы данных по вакансиям на определенный момент времени.

В результате построения нейросетевой модели средняя ошибка между предсказанными и наблюдаемыми значениями заработной платы составила 2497,5, а ошибка аппроксимации — 9,1% (меньше 10%), что свидетельствует о хорошем подборе модели к исходным данным. Также тесную связь между рассматриваемыми значениями подтверждают поле корреляции (см. рис. 10) и линейный коэффициент корреляции, равный 0,751.

Вышеназванные характеристики подтверждают качество построенной предиктивной регрессионной модели.

Итак, при помощи аналитической платформы IBM SPSS modeler проведен кластерный анализ вакансий на примере специализации «Аналитик», позволивший выявить многомерные группы по размеру средней заработной платы и дать подробное описание состава каждого полученного кластера вакансий.

Предсказанные против наблюденных Целевой элемент: Заработная плата

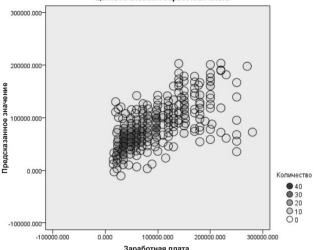


Рис. 10. Корреляционная связь между наблюдаемыми и предсказанными значениями заработной платы

Построенная предиктивная регрессионная модель в виде многослойного персептрона выявила важность каждого предиктора, влияющего на размер ежемесячной заработной платы. Следует отметить, что можно менять состав предикторов нейросети в виде количественных и описательных показателей. Это будет зависеть от исходных данных, т. е. спектра характеристик вакансий на рассматриваемый момент времени.

Выводы

Таким образом, разработанный алгоритм анализа вакансий по данным из открытых источников при помощи статистических методов и методов машинного обучения позволит исследователям, работодателям, органам государственной власти, образовательным учреждениям проводить анализ вакансий по количественным и качественным характеристикам из открытых источников на рынке труда по любой специализации на определенный момент времени, быстро обработав большие массивы данных, а соискателю увидеть текущую востребованность вакансий на рынке труда в разрезе специализаций, городов, заработной платы и другим характеристикам.

Литература

- 1. Атлас новых профессий 3.0./ под ред. Д. Варламовой, Д. Судакова. М.: Интеллектуальная Литература, 2020. 456 с.
- 2. **Хохлова А.Н.** Парсинг вакансий на рынке труда из открытых источников / X Международная науч-

- но-практическая конференция имени А.И. Китова «Информационные технологии и математические методы в экономике и управлении» (ИТиММ-2020). 15—16 октября 2020 г.: сборник статей. Москва: ФГБОУ ВО «РЭУ им. Г.В. Плеханова», 2020. С. 253—259.
- 3. Факторный, дискриминантный и кластерный анализ: Пер. с англ. / Дж.-О. Ким, Ч.У. Мьюллер, У.Р. Клекка и др. /под ред. И.С. Енюкова. М.: Финансы и статистика, 1989. 215 с.
- 4. **Dangeti P.** Statistics for Machine Learning. Packt Publishing Ltd., 2017. 442 p.
- 5. **Gutierrez D.D.** Machine Learning and Data Science: An Introduction to Statistical Learning Methods with R. Technics Publications, 2015. 282 p.
- 6. **Кун М., Джонсон К.** Предиктивное моделирование на практике. СПб: Питер, 2019. 640 с.

Информация об авторах

Хохлова Оксана Анатольевна — д-р экон. наук, профессор, заведующий кафедрой «Макроэкономика, экономическая информатика и статистика», Восточно-Сибирский государственный университет технологий и управления. 670013, г. Улан-Удэ, ул. Ключевская, д. 40B, стр. 1. E-mail: hohlovao@mail.ru. ORCID: https://orcid.org/0000-0002-0851-7587.

Хохлова Александра Николаевна — аналитик Департамента Тинькофф бизнес, АО «Тинькофф Банк». 125212, г. Москва, Головинское шоссе, д. 5. E-mail: alexandra.khokhlova@mail.ru. ORCID: https://orcid.org/0000-0002-3984-5022.

Чойжалсанова Аюна Цыденовна — канд. экон. наук, старший преподаватель кафедры «Макроэкономика, экономическая информатика и статистика», Восточно-Сибирский государственный университет технологий и управления. 670013, г. Улан-Удэ, ул. Ключевская, д. 40В, стр. 1. E-mail: ayuna-777@mail.ru. ORCID: https://orcid.org/0000-0001-9402-2507.

References

- 1. Varlamova D., Sudakov D. (eds.) *Atlas of New Professions 3.0.* Moscow: Intellectual Literature; 2020. 456 p. (In Russ.)
- 2. **Khokhlova A.N.** Parsing Vacancies in the Labor Market from Open Sources. In: *Proceedings of the X International Scientific and Practical Conference Named after A.I. Kitov «Information Technologies and Mathematical Methods in Economics and Management» (IT&MM-2020), 15–16 October, 2020. Moscow: Plekhanov Russian University of Economics; 2020. pp. 253–259. (In Russ.)*
- 3. **Kim J.-O., Muller Ch.W.** Factor Analysis: Statistical Methods and Practical Issues (Eleventh Printing). SAGE Publications; 1986; **Klecka W.K.** Discriminant Analysis (Seventh Printing). SAGE Publications; 1986; **Aldender-**
- fer M.S., Blashfield R.K. Cluster Analysis (Second Printing). SAGE Publications; 1985. (Russ. ed.: Kim J.-O. et al.; Enyukov I.S. (ed.) Faktornyi, diskriminantnyi i klasternyi analiz. Moscow: Finansy i statistika Publ.; 1989. 215 p.).
- 4. **Dangeti P.** *Statistics for Machine Learning*. Birmingham, United Kingdom: Packt Publishing Ltd.; 2017. 442 p.
- 5. **Gutierrez D.D.** *Machine Learning and Data Science: An Introduction to Statistical Learning Methods with R.* Technics Publications; 2015. 282 p.
- 6. **Kuhn M., Johnson K.** *Applied Predictive Modeling*. New York: Springer; 2013. 600 p. Available from: https://doi.org/10.1007/978-1-4614-6849-3 (Rus. ed.: Kun M., Dzhonson K. *Prediktivnoe modelirovanie na praktike*. St. Petersburg: Peter; 2019, 640 p.).

About the authors

Oksana A. Khokhlova — Dr. Sci. (Econ.), Professor, Head, Department of Macroeconomics, Economic Informatics and Statistics, East Siberia State University of Technology and Management. 40V, Klyuchevskaya St., Bldg. 1, Ulan-Ude, 670013, Russia. E-mail: hohlovao@mail.ru. ORCID: https://orcid.org/0000-0002-0851-7587.

Alexandra N. Khokhlova — Analyst, Tinkoff Business Department, JSC «Tinkoff Bank». 5, Golovinskoye Shosse, Moscow, 125212, Russia. E-mail: alexandra.khokhlova@mail.ru. ORCID: https://orcid.org/0000-0002-3984-5022.

Ayuna T. Choyzhalsanova — Cand. Sci. (Econ.), Senior Lecturer of the Department of Macroeconomics, Economic Informatics and Statistics, East Siberia State University of Technology and Management. 40V, Klyuchevskaya St., Bldg. 1, Ulan-Ude, 670013, Russia. E-mail: ayuna-777@mail.ru. ORCID: https://orcid.org/0000-0001-9402-2507.