



Известия Саратовского университета. Новая серия. Серия: Математика. Механика. Информатика. 2022. Т. 22, вып. 1. С. 123–129
Izvestiya of Saratov University. Mathematics. Mechanics. Informatics, 2022, vol. 22, iss. 1, pp. 123–129
<https://mmi.sgu.ru> <https://doi.org/10.18500/1816-9791-2022-22-1-123-129>

Article

Analysis of technological trends to identify skills that will be in demand in the labor market with open-source data using machine learning methods

O. A. Khokhlova^{1✉}, A. N. Khokhlova²

¹East Siberian State University of Technology and Management, 40V Klyuchevskaya St., Ulan-Ude 670013, Russia

²JSC Tinkoff Bank, 5 Golovinskoe Highway, Moscow 125212, Russia

Oksana A. Khokhlova, hohlovao@mail.ru, <https://orcid.org/0000-0002-0851-7587>

Alexandra N. Khokhlova, alexandra.khokhlova@mail.ru, <https://orcid.org/0000-0002-3984-5022>

Abstract. The further development of society directly depends on the use of technologies connected with processing data arrays and identifying patterns with the help of computer means. In this study, machine learning methods allowed us to analyze technological trends using large open-source data on patents, which enable predicting future skills in demand in the labor market. It is of major importance in the context of the rapid development of technology, leading to large-scale technological changes that transform the social conditions of human life as a whole, the requirements for the skills of people, which in the future will cause the emergence of new specialties and the disappearance of existing professions. For this purpose, predictive regression models of groups of patents according to the International Patent Classification are built using machine-learning methods — classical forecasting methods, such as naive forecasting, simple exponential smoothing, and ARIMA. As a result of comparing the quality of the constructed models and choosing the best one, ARIMA models were identified, showing “fading” technologies if there is a decrease in the number of patents; promising technological directions if the growth is stable; or “breakthrough” technologies if there has been a sharp increase in recent years. The input variables of the models were the series of dynamics of patents of different classes in the form of historical data, the output variables were the predicted values of the number of patents of these classes of a certain technological trend. The algorithm was implemented in the high-level Python programming language. The research results will enable authorities, employers, educational institutions, etc. to make a forecast of the demand for existing, as well as new professional skills and competencies in the labor market.

Keywords: machine learning methods, ARIMA, technology trend, patent, labor market

For citation: Khokhlova O. A., Khokhlova A. N. Analysis of technological trends to identify skills that will be in demand in the labor market with open-source data using machine learning methods. *Izvestiya of Saratov University. Mathematics. Mechanics. Informatics*, 2022, vol. 22, iss. 1, pp. 123–129. <https://doi.org/10.18500/1816-9791-2022-22-1-123-129>

This is an open access article distributed under the terms of Creative Commons Attribution 4.0 International License (CC-BY 4.0)



Научная статья
УДК 004.85

Анализ технологических трендов для выявления востребованных в будущем навыков на рынке труда по данным из открытого источника с использованием методов машинного обучения

О. А. Хохлова^{1✉}, А. Н. Хохлова²

¹Восточно-Сибирский государственный университет технологий и управления, Россия, 670013, г. Улан-Удэ, ул. Ключевская, д. 40В

²АО «Тинькофф Банк», Россия, 125212, г. Москва, Головинское шоссе, д. 5

Хохлова Оксана Анатольевна, доктор экономических наук, профессор, заведующий кафедрой «Макроэкономика, экономическая информатика и статистика», <https://orcid.org/0000-0002-0851-7587>, hohlova.o@mail.ru

Хохлова Александра Николаевна, аналитик Департамента Тинькофф бизнес, <https://orcid.org/0000-0002-3984-5022>, alexandra.khokhlova@mail.ru,

Аннотация. Дальнейшее развитие общества напрямую зависит от использования технологий, связанных с обработкой массивов данных и выявления закономерностей компьютерными средствами. В данном исследовании методы машинного обучения позволили провести анализ технологических трендов по большим данным из открытого источника о патентах, позволяющих предсказать в будущем навыки, востребованные на рынке труда. Это имеет важное значение в условиях стремительного развития технологий, приводящих к масштабным технологическим изменениям, меняющим социальные условия жизни человечества в целом, требования к навыкам людей, которые в дальнейшем вызовут возникновение новых специальностей и исчезновение существующих ныне профессий. С этой целью в работе построены предиктивные регрессионные модели групп патентов согласно Международной патентной классификации при помощи методов машинного обучения — классических методов прогнозирования, таких как наивное прогнозирование, простое экспоненциальное сглаживание и ARIMA. В результате сравнения качества построенных моделей и выбора лучшей были выявлены модели ARIMA, показывающие: «угасающие» технологии, если происходит снижение числа патентов; перспективные технологические направления, если наблюдается стабильный рост; или «прорывные» технологии, если произошел резкий рост за последние годы. Входными переменными моделей явились ряды динамики патентов разных классов в виде исторических данных, выходными — прогнозные значения числа патентов этих классов определенного технологического тренда. Алгоритм реализовывался на высокоуровневом языке программирования Python. Результаты исследования позволят органам власти, работодателям, образовательным учреждениям и т. д. сделать прогноз востребованности ныне существующих, а также новых профессиональных навыков и компетенций на рынке труда.

Ключевые слова: методы машинного обучения, ARIMA, технологический тренд, патент, рынок труда

Для цитирования: *Khokhlova O. A., Khokhlova A. N. Analysis of technological trends to identify skills that will be in demand in the labor market with open-source data using machine learning methods [Хохлова О. А., Хохлова А. Н. Анализ технологических трендов для выявления востребованных в будущем навыков на рынке труда по данным из открытого источника с использованием методов машинного обучения] // Известия Саратовского университета. Новая серия. Серия: Математика. Механика. Информатика. 2022. Т. 22, вып. 1. С. 123–129. <https://doi.org/10.18500/1816-9791-2022-22-1-123-129>*

Статья опубликована на условиях лицензии Creative Commons Attribution 4.0 International (CC-BY 4.0)



Introduction

Under the conditions of rapidly advancing technology, many of the workforce processes will be automated or disappear due to a dramatic economic shift, significant disruptions in the labor market will occur, and businesses around the world may face risks from a shortage of skilled labor. In this regard, the state should comprehensively assess the potential prospects in the labor market, in accordance with this, update the educational strategy, systematically predict the evolution of skills in the future for the timely formation of professional competencies relevant to the labor market.

In our opinion, the demanded professions and skills can also be seen by focusing on technological trends emerging in the patent market. Patents usually protect what should appear on the market in five to seven years. Accordingly, if an increase in the number of patents in a certain area is visible, then we can conclude what will be in trend in the next few years, what knowledge and skills will be needed. For this, the study developed an algorithm for constructing predictive regression models for the number of patents using machine learning methods in the Python programming language.

1. Description of the algorithm and the used tools

At the first stage, data collecting was carried out. The primary initial data for the analysis were time series for different classes of patents according to the International Patent Classification (hereinafter – IPC) from the open portal <https://www.lens.org> for the period 2010–2020 which makes 2,250,000 records.

Further, a preliminary analysis of the data was made. Description of the source data and data cleaning and preparation for further analysis were accomplished. After this stage, 1,123,189 records remained, as well as 31 attributes.

Then the grouping of patents was carried out – from the set of data, patents were selected in the context of subsections according to the IPC and visualization of data by groups of patents, i.e. building time series in the form of line charts to identify technology trends:

- “dying” technologies (if there is a decrease in the number of patents);
- promising technological directions (if the growth is stable);
- “breakthrough” technologies (if there is exponential development, that is, a sharp increase in recent years).

At the next stage, predictive regression models of subsections / classes of patents according to the IPC were built on the basis of time series using machine learning methods – classical forecasting methods: naive forecasting, simple exponential smoothing, and ARIMA.

Time series of patents for 2010–2016 were selected as a training sample, patents for 2017–2019 were selected as a test sample. The forecast was based on 2020–2022. Next, we compared the results for assessing the quality of the models – the root mean square error (RMSE), which is necessary to check the accuracy of the models on test data, and select the best model.

The tool was the Python 3.8.5 programming language in Python. When developing the program code, the Python libraries were used: OS, Requests, NumPy, Matplotlib, Scikit-Learn.



2. Predictive modeling of technology trends to identify future skills in the labor market

According to experts, simple classical methods such as linear methods and exponential smoothing outperform complex methods such as decision trees, multilayer perceptrons (MLP), and long-term short-term memory (LSTM) network models. The research was carried out on a diverse set of more than 1000 one-dimensional time series forecasting problems. The results showed that deep learning methods have not yet met their expectations for univariate time series forecasting, and there is still a lot of work to be done in their development [1]. Therefore, in this work, predictive regression models are developed using classical forecasting methods, such as naive forecasting, simple exponential smoothing, and ARIMA [2].

As an implementation of the algorithm, the article presents examples of different technological trends by groups of patents according to the IPC. Therefore, in Figures 1, 2, 3, there are groups of patents that can be attributed to technological trends of future development, since they show a steady increase in the number of patents. The best model was the ARIMA model, as evidenced by the low RMSE (Table 1).

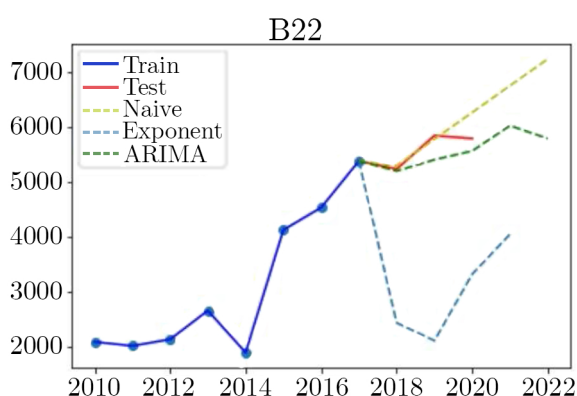


Fig. 1. Modeling the number of patents in subsection B22 “Foundry production; powder metallurgy” (color online)

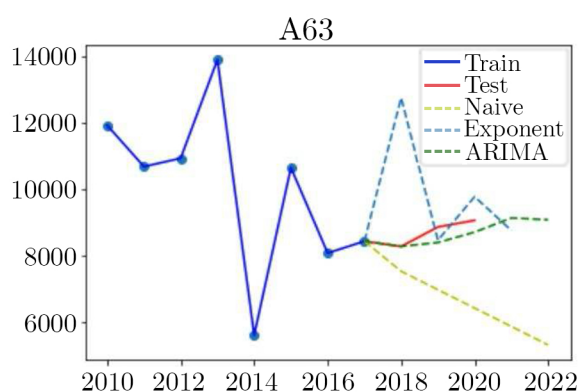


Fig. 2. Modeling the number of patents in subsection A63 “Sports; games; entertainment” (color online)

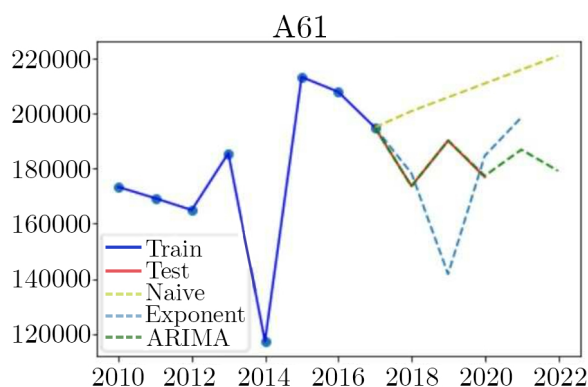


Fig. 3. Modeling the number of patents in subsection A61 “Medicine and veterinary medicine; hygiene” (color online)



Table 1

Root mean square error (RMSE) of patent subsections B22, A63 and A61

IPC patent subsections	Naive Forecasting	Simple Exponential Smoothing	ARIMA
B22 “Foundry production; powder metallurgy”	242.0	2647.4	221.6
A63 “Sports; games; entertainment”	1672.7	2282.5	129.8
A61 “Medicine and Veterinary Medicine; hygiene”	23103.9	24688.9	15724.4

Examples of “dying” technology trends are shown in Table 2 and Figures 4, 5.

In recent years, there has been a strong decline in patents for bookbinding, animal and vegetable oils. In the first case, from 2015, in the second, from 2017.

Examples of “breakthrough” technologies include patents of subsections B64 “Aeronautics; aviation; cosmonautics”; B33 “Layer-by-layer synthesis technology” which refer to patents related to the manufacture of three-dimensional (3D) objects with the help of additive deposition, additive agglomeration or additional layering, for example,

Table 2

Root mean square error (RMSE) of patent subsections B42 and C11

IPC patent subsections	Naive Forecasting	Simple Exponential Smoothing	ARIMA
B42 Bookbinding; albums; means of classification and storage of documents, etc.; special types of printed materials	331.2	217.7	368.2
C11 “Animal and vegetable oils; fats, fatty substances and waxes derived from them fatty acids; detergents; candles”	780.1	1160.7	129.8

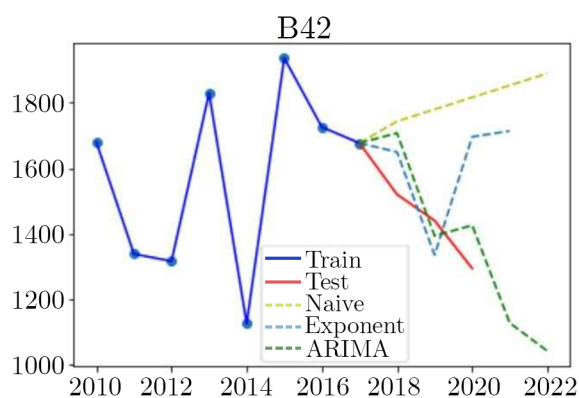


Fig. 4. Modeling of the number of patents for subsection B42 “Bookbinding; albums; means of classification and storage of documents, etc.; special types of printed materials” (color online)

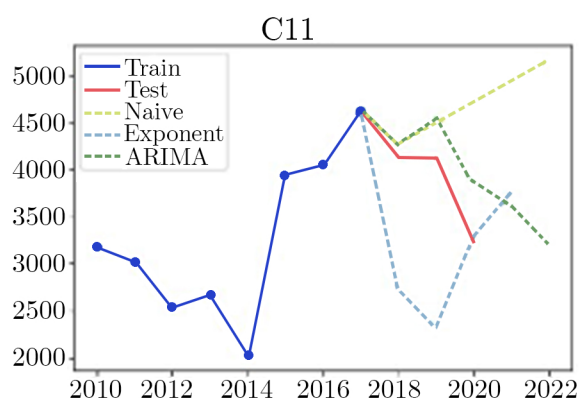


Fig. 5. Modeling the number of patents in subsection C11 “Animal and vegetable oils; fats, fatty substances and waxes derived from them fatty acids; detergents; candles” (color online)



through 3D printing, stereolithography, or selective laser sintering; class of patents — B22F “Powder metallurgy, production of products from metal powders” (Table 3 and Figures 6, 7, 8). Therefore, we can conclude that the labor market will continue to have a high demand for specialists and skills in these areas of activity.

Table 3

Root mean square error (RMSE) of subsections /classes of patents B64, B33, B22F

IPC patent subsections	Naive Forecasting	Simple Exponential Smoothing	ARIMA
B64 “Aeronautics; aviation; cosmonautics”	653.2	4089.9	577.5
B33 “Layer-by-layer synthesis technology”	1192.4	2136.6	720.2
B22F “Powder metallurgy, production of products from metal powders”	306.8	2252.4	173.3

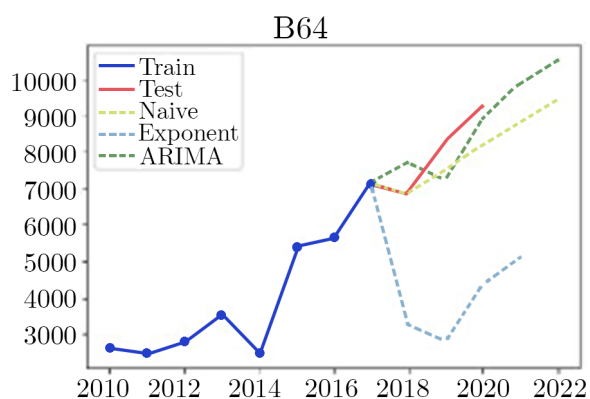


Fig. 6. Modeling the number of patents of subsection B64 “Aeronautics; aviation; cosmonautics” (color online)

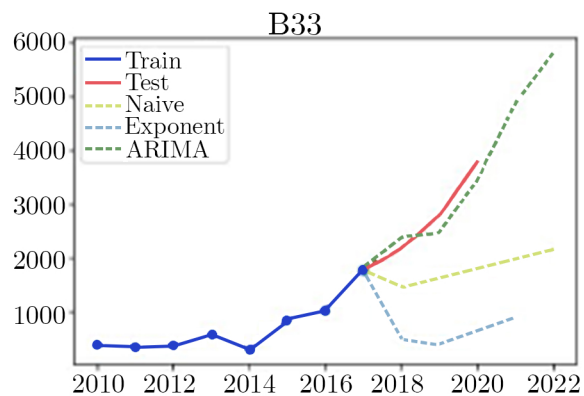


Fig. 7. Modeling the number of patents of subsection B33 “Layer-by-layer synthesis technology” (color online)

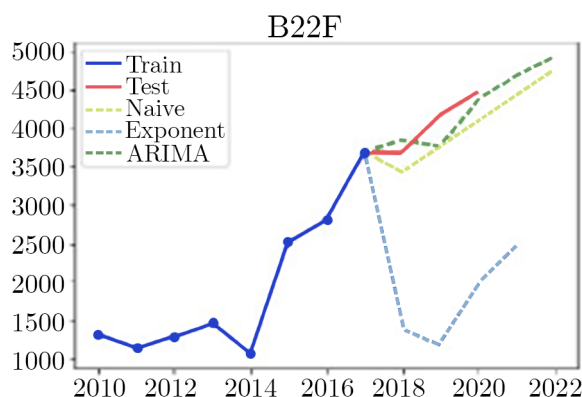


Fig. 8. Modeling the number of patents of subsection B22F “Powder metallurgy, production of products from metal powders” (color online)



In all of the abovementioned examples of predicting the number of patents in the context of patent subsections according to IPC by machine learning methods — naive forecasting, simple exponential smoothing, and ARIMA — the best model is the ARIMA model, as evidenced by the low root mean square error (RMSE).

Thus, technological trends show that, depending on their development, it is possible to predict the demand for existing or new skills and professions, as well as their training, i.e. having big data, it is possible to form certain patterns in certain directions, the results of which will allow generalizations and certain conclusions to be drawn.

Conclusion

A study based on the developed algorithm for analyzing patents using machine learning in the Python programming language showed that if professions are currently in demand in the labor market, and the technological trend based on patents in this area has a promising or “breakthrough” development, then it is possible to do the conclusion that specialists and skills in this area are needed now and in the future, their training is important for the development of the economy.

It should also be noted that the proposed algorithm allows you to quickly and efficiently process large data on patents from open sources and build predictive regression models of technological trends for various classes of patents.

The results will enable researchers, public authorities to explore future professions and skills in the labor market; educational institutions to adjust training programs in accordance with modern employers’ requirements and future skills; employers to make decisions on the formation of new competencies in their field of activity, based on big data analytics and the use of machine learning methods; carry out a comparative analysis of in-demand vacancies in terms of quantitative and qualitative characteristics; for the applicant to see the demand for vacancies in the labor market and the development of new skills.

References

1. Brownlee J. *Comparing Classical and Machine Learning Algorithms for Time Series Forecasting*. Available at: <https://machinelearningmastery.com/findings-comparing-classical-and-machine-learning-methods-for-time-series-forecasting/> (accessed 28 August 2021).
2. Kuhn M., Johnson K. *Applied Predictive Modeling*. New York, Springer Science+Business Media, 2013. 600 p. <https://doi.org/10.1007/978-1-4614-6849-3>

Поступила в редакцию / Received 07.12.2021

Принята к публикации / Accepted 21.12.2021

Опубликована / Published 31.03.2022