MSC 91B84, 62P20 DOI: 10.14529/jcem240205

# A SYSTEM FOR FORECASTING THE VOLUME OF SALES OF RESIDENTIAL REAL ESTATE BASED ON MACHINE LEARNING

***O.V. Korobkova,*** South Ural State University, Chelyabinsk, Russian Federation, ufimtcevaov@susu.ru

Any construction project contains a cash flow model, the purpose of which is to assess the ability of an enterprise to generate cash in the required amounts and within the time required for planned costs, to calculate revenue, profit/loss. The income from the implementation of the construction project, as well as its profitability, directly depend on the volume of real estate sales. The article describes a system for forecasting the volume of real estate sales by a construction company in the regional market. This system is built on a neural network basis using the domestic Loginom Community analytical platform. To train the system, three groups of factors were used that can be quantified from official sources: external macroeconomic factors determined at the federal level, external regional and retrospective data downloaded from the corporate database of a construction company and characterizing the dynamics of residential real estate sales. The system has a modular structure. The modular structure gives the system a universal character by allocating independent modules in the structure, which allow taking into account regional, federal and corporate input factors. The system is trained and has good forecast properties. The average relative error of forecasting is 6.89%.

*Keywords: equity construction; cash flow model; Loginom analytical platform; neural network forecasting.*

## 1. Introduction

The Fourth Industrial Revolution (Industry 4.0) is characterized by the widespread use of information technologies in all industries, large-scale automation of business processes, and the use of data analysis methods [1–6]. Intelligent data analysis methods (data mining) using machine learning algorithms allow us to identify patterns that are implicitly contained in the analyzed data, which increases the effectiveness of managerial decision-making. Digital technologies are in demand in all sectors of the economy, and the construction industry is no exception.

Construction project management is a complex and time-consuming process that is possible only with huge financial investments and a high level of professionalism of specialists. One of the key components of such management is cash flow management, and building a cash flow model is a prerequisite for achieving successful results. The purpose of building a cash flow model is to assess the ability of an enterprise to generate cash in the required amounts and within the time required for the planned costs, to calculate revenue, profit/loss.

With the introduction of Federal Law No. 214 in 2019, companies are required to open escrow accounts for shared-equity construction. The money of the shareholders is not transferred to the company's account, but is frozen in a special bank account, which is disclosed after the house is rented. Accordingly, until the housing is rented, the company does not receive any profit. The developer company is forced to build on loans taken

from the bank, the interest on which depends on the amount accumulated in the escrow account. And this amount, in turn, depends on the volume of sales of real estate in a new building, the size of which is unknown in advance during planning and at the beginning of construction.

Econometric, heuristic and intuitive methods are used to solve forecasting problems. Econometric methods require the construction of formal mathematical models. The forecast is based on time series analysis, and various regression models are also widely used [2–12]. Such methods require assumptions about the distribution of time series values, so there are significant errors in the forecast results. Regional construction companies usually use an intuitive subjective assessment of future sales volumes, which only approximately takes into account changing environmental factors. Therefore, the planned financial indicators are inaccurate and largely indicative.

An alternative approach to solving the forecasting problem is the use of neural network (NS) forecasting models, which are based on machine learning methods and allow using a variety of factors that influence the behavior of the process under study [3, 4, 11, 12].

## 2. Construction of Neural Network Forecasting System

Let's describe the advantages of models based on neural networks [3, 4]. A simple neural network are used for constructs the forecasting system. The structure of such neural network is shown in Fig. 1.

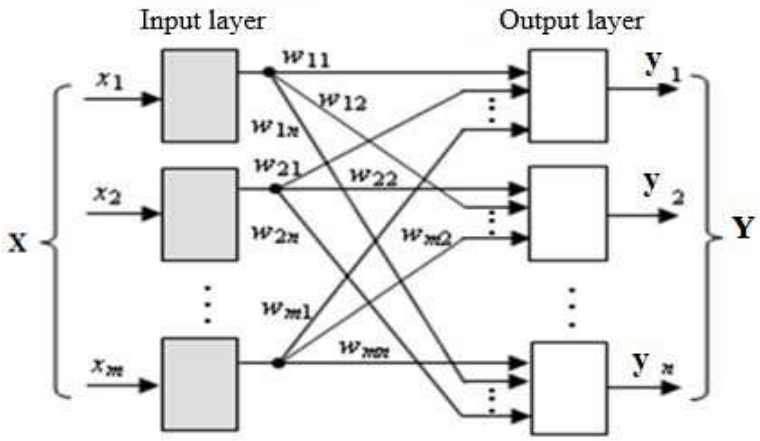


**Fig. 1**. The structure of a simple neural network

The output signal of a neural network is determined from the expression

$$Y = X \cdot W, \tag{1}$$

where $Y$ and $X$ are string vectors and $W$ is a matrix of weighting coefficients of form

$$W = \begin{pmatrix} w_{11} & \cdots & w_{1n} \\ \vdots & \ddots & \vdots \\ w_{m1} & \cdots & w_{mn} \end{pmatrix}.$$

Forecasting is related to obtaining future values of an object's state based on its current state. Therefore, to train a neural network, a training sample is used, given in tabular pairs

"known input – known output" $\left[(X^{(1)}, Y^{(1)}), (X^{(2)}, Y^{(2)}), ..., (X^{(N)}, Y^{(N)})\right]$. The purpose of the training is to find such values of the weight matrix $W$ that minimize the standard error $E$ of form

$$E = \frac{1}{N} \sum_{i=1}^{N} \left(Y^{(i)} - D^{(i)}\right)^2, \tag{2}$$

where $N$ is the size of the training sample; $Y^{(i)}$ is the value at the output of the model for the $i$-th example; $D^{(i)}$ is the target value. The size of the input vector $X^{(i)} = (x_1^i, x_2^i, ...x_m^i)$ is determined by the amount of input data, the size of the output vector $Y^{(i)} = (y_1^i, y_2^i, ...y_n^i)$ is determined by the problem in (1).

The system uses 3 groups of factors that influence the dynamics of the volume of real estate purchases: macroeconomic indicators determined at the federal level (federal), regional macroeconomic indicators and individual factors related to buyer preferences.

Regional factors can be quantified by sources representing official information and which are typical for the Chelyabinsk region. For example, the number of housing and mortgage loans, the number of people who used maternity capital to buy real estate in new buildings, the average per capita income of the population, the average monthly salary.

External macroeconomic factors established at the state level, which significantly affect the size of housing loans, household incomes and the average monthly salary at work, are attributed to: the amount of the Central Bank's % rate, the inflation rate. The values of the listed factors can be obtained from official sources [5, 7–10, 13].

Internal factors are retrospective data on real estate sales obtained from the corporate database of a construction company. These data represent a time series of sales for an already implemented similar project with similar technical and economic indicators. Therefore, the proposed system can be applied in an area where the uniformity properties of an object are fulfilled, for example, apartments in a certain area or city built according to projects with similar technical and economic indicators.

Individual factors have a great influence on the decision to purchase residential real estate. Buyers usually evaluate the location of the property, the development of social institutions: kindergartens, schools, shops, accessibility of transport, the prestige of the developer and many others. However, such information is usually inaccessible to the researcher, and it is impossible to quantify it.

The system has a modular structure (Fig. 2).

The modular structure gives the system a universal character

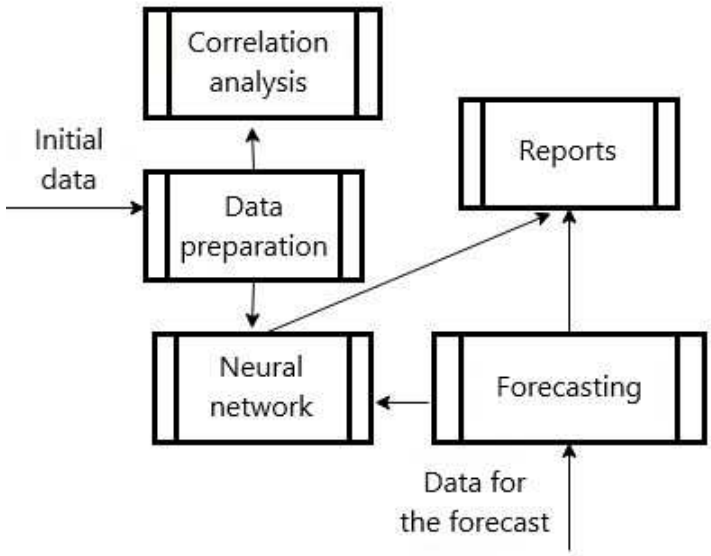


**Fig. 2**. The modular structure of system

by allocating independent modules by allocating independent modules in the structure, which allow taking into account regional, federal and corporate input factors.

In the Data Preparation module, input and enrichment of the source data is performed. The Correlation Analysis module checks the statistical significance of the selected factors. The Neural Network module is a neural network that is trained on the retrospective data entered in the Data Preparation module.

The neural network forecasting model was built on the domestic Loginom Community platform (Fig. 3). The Loginom analytical low-code platform contains a visual constructor that allows you to build analysis models without programming knowledge (Fig. 3).

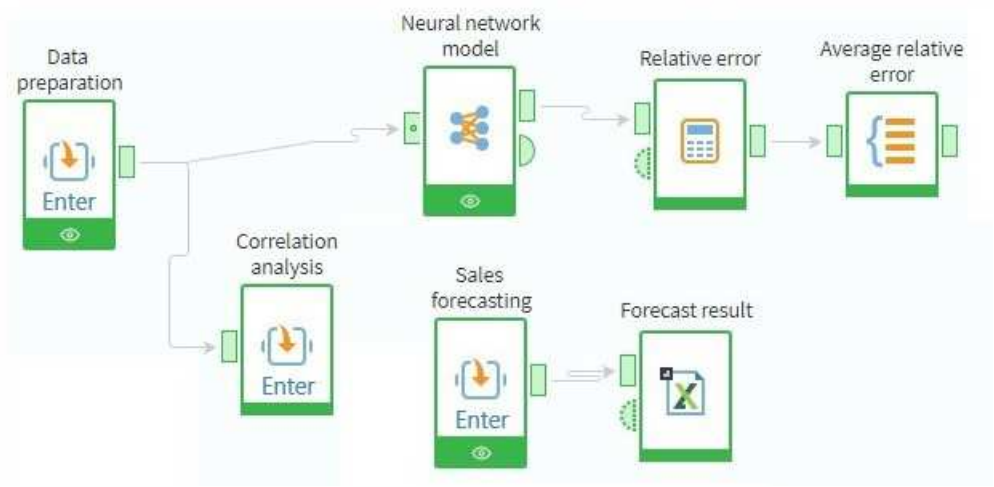


**Fig. 3**. Implementation of the system on the Loginom platform

The accuracy of forecasting is significantly influenced by the quality of the data on which the neural network is trained. Data that is structured and has a certain quality is used for analysis. The list of input data is shown in Figure 4.

| Label | Name | Data kind | Usage type |
|---|---|---|---|
| 12 Year | Data_Y_1 | Discrete | Input |
| 12 Month | Data_M_1 | Discrete | Input |
| 12 Number of sales | Vyruchka_Count | Continuous | Output |
| 9.0 Average price sq.m | tsena_za_kvadrat | Continuous | Input |
| 9.0 Total revenue | Vyruchka_Sum | Continuous | Input |
| 12 Number of housing loans | Kolichestvo_zhilischnykh_kreditov | Continuous | Input |
| 12 Number of mortgage housing loans | Kolichestvo__ipotechnykh_zhilischnykh_kre... | Continuous | Input |
| 9.0 Interest rate % per annum | Razmer_stavki___godovykh | Continuous | Input |
| 12 Number of persons maternity capital | CHislennost_lits__rasporyadivshikhsya_chas... | Continuous | Input |
| 9.0 % inflation | __inflyatsii | Continuous | Input |

**Fig. 4**. Input data used to train the NS

So, the neural network has 9 input neurons and one output, which generates a forecasted value.

All real data has been normalized and brought to the range $0-1$. Using normalization increases the quality and speed of neural network learning.

In the system, sales volume forecasting was performed for one of the regional construction companies based on real retrospective monthly sales data for 2018–2022. A graphic illustration of the learning process of the model is shown in Fig. 5.

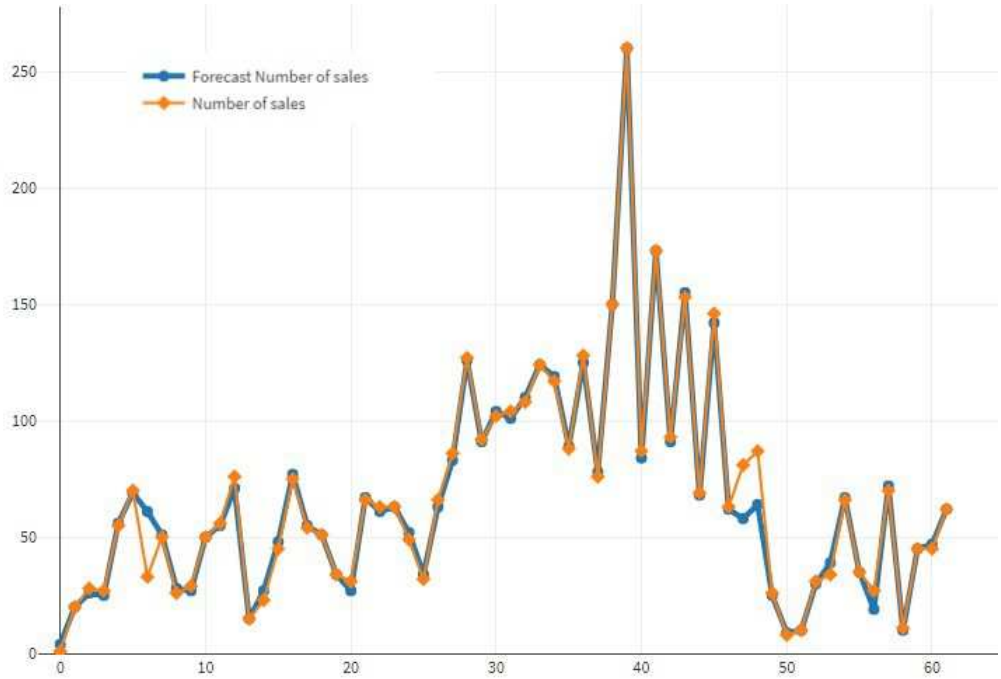


**Fig. 5**. A graphic illustration of the learning process of the system

A comparison of real and forecast values for January – April 2023 is shown in Fig. 6.

| 12 year | 12 month | 12 sales quantity (forecast) | 12 sales quantity |
|---|---|---|---|
| 2 023 | 1 | 10 | 11 |
| 2 023 | 2 | 42 | 45 |
| 2 023 | 3 | 46 | 45 |
| 2 023 | 4 | 60 | 62 |

**Fig. 6**. Comparison of real and forecast values

The average relative error of forecasting is 6.89%. More information about it can be founded in [14].

## 3. Conclusions

1. The model is trained and has good forecast properties. The average relative error of forecasting is 6.89%.

2. The factors influencing the volume of sales of residential real estate in the primary market of the Chelyabinsk region are presented.

3. The model is universal and allows you to make a forecast with changing macroeconomic indicators within the framework of three different scenarios of economic development: optimistic, pessimistic and real.

The input data characterizing the change in the economic situation are entered in the "Data preparation" module.

The accuracy of the forecasting result significantly depends on the completeness and adequacy of the information collected. However, the approximate estimates obtained will be determined by external economic factors, which is of practical importance for any construction company.

## References

1. Barinova N.V., Barinov V.R. Digital Economy, Industry 4.0 and Artificial Intellect. *Vestnik of the Plekhanov Russian University of Economics*, 2021, vol. 18, no. 3 (117), pp. 82–91. DOI: 10.21686/2413-2829-2021-3-82-91 (in Russian)

2. Bogdanova T.K., Kamalova A.R., Kravchenko T.K., Poltorak A.I. Problems of Modeling the Valuation of Residential Properties. *Business Informatics*, 2020, vol. 14, no. 3, pp. 7–23. DOI: 10.17323/2587-814X.2020.3.7.23

3. Lyubimova T.V., Gorelova A.V. Solving the Forecasting Problem Using Neural Networks. *Innovative Science – Innovacionnaya nauka*, 2015, no. 4, pp. 39–43. (in Russian)

4. Morozova V.I., Logunova D.I. Forecasting by Machine Learning. *Young scientist – Molodoy uchenyi*, 2022, no. 21 (416), pp. 202–203. – https://moluch.ru/archive/416/92048 (date of access: 12/13/2023) (in Russian)

5. Ministry of Economic Development of the Russian Federation [Electronic resource]. – https://www.economy.gov.ru (date of access: 12/13/2023)

6. Neskoromnyi S.V., Varyanova A.A. Industry 4.0. Digital Economy in Russia. *Current Research – Aktual'nye issledovaniya*, 2020, no. 21 (24), pp. 112–115. – https://apni.ru/article/1362-industriya-40-tsifrovaya-ekonomika-v-rossii (date of access: 12/13/2023) (in Russian)

7. SBER, official cite [Electronic resource]. – https://www.sberbank.ru (date of access: 12/13/2023)

8. Statistical Yearbook "Housing (Mortgage) Loan Market in Russia" [Electronic resource]. – https://cbr.ru/eng/statistics/bank_sector/mortgage (date of access: 12/13/2023)

9. Federal State Statistics Service of Russian Federation [Electronic resource]. – https://eng.rosstat.gov.ru/ (date of access: 12/13/2023)

10. Bank of Russia [Electronic resource]. – https://cbr.ru/eng (date of access: 12/13/2023)

11. Shagalova P.A., Lyakhmanov D.A. Neural Network Technologies for Solving of Forecasting Problem. *Modern Problems of Science and Education*, 2014, no. 6. – https://science-education.ru/en/article/view?id=16494 (date of access: 12/13/2023)

12. Yasnitskiy V.L. *Neural Network Modeling of the Processes of Mass Valuation and Scenario Forecasting of the Market Value of Residential Real Estate – Neirosetevoe modelirovanie processov massovoi ocenki i scenarnogo prognozirovaniya rynochnoi stoimosti zhiloi nedvizhimosti*, Abstract of PhD (Economic) Thesis. Perm, 2018, 24 p. – http://www.psu.ru/files/docs/science/dissertatsionnye-sovety/Yasnickiy/avtoreferat_yasnickyi_vl.pdf (date of access: 12/13/2023) (in Russian)

13. Territorial Office of the Federal State Statistics Service for the Chelyabinsk region [Electronic resource]. – https://74.rosstat.gov.ru (date of access: 12/13/2023) (in Russian)

14. Pollack G.A., Korobkova O.V., Pollak I.Yu. Neural Network System of Sales Volume Forecasting of Residential Real Estate in the Primary Regional Market. *Journal of Computational and Engineering Mathematics*, 2023, vol. 10, no. 3, pp. 39–53. DOI: 10.14529/jcem230304

*Olga V. Korobkova, PhD Student, Department of Mathematical and Computational Modelling, South Ural State University (Chelyabinsk, Russian Federation), ufimtcevaov@susu.ru*

*Received* December 15, 2023

**УДК 004.032.26** **DOI: 10.14529/jcem240205**

# СИСТЕМА ПРОГНОЗИРОВАНИЯ ОБЪЕМА ПРОДАЖ ЖИЛОЙ НЕДВИЖИМОСТИ НА ОСНОВЕ МАШИННОГО ОБУЧЕНИЯ

***О.В. Коробкова***

Любой строительный проект содержит модель движения денежных средств, цель построения которой заключается в оценке способности предприятия генерировать денежные средства в необходимых размерах и в сроки, необходимые для планируемых затрат, выполнить расчет выручки, прибыли/убытков. Доходы от реализации строительного проекта, а также его рентабельность напрямую зависят от объема продаж недвижимости. В статье описывается система прогнозирования объема продаж недвижимости строительной компанией на региональном рынке, построенная в нейросетевом базисе с использованием отечественной аналитической платформы Loginom Community. Для обучения системы использованы три группы факторов, которые можно определить количественно из официальных источников: внешние макроэкономические факторы, определяемые на федеральном уровне, внешние региональные и ретроспективные данные, выгружаемые из корпоративной базы данных строительной фирмы и характеризующие динамику продаж жилой недвижимости. Система имеет модульную структуру. Модульная структура придает системе универсальный характер за счет выделения в структуре независимых модулей, которые позволяют учесть региональные, федеральные и корпоративные входные факторы. Система обучена и имеет хорошие прогностические свойства. Средняя относительная погрешность прогнозирования 6,89%.

*Ключевые слова: долевое строительство; модель движения денежных средств; аналитическая платформа Loginom; нейросетевое прогнозирование.*

## Литература

1. Баринова, Н.В. Цифровая экономика, индустрия 4.0 и искусственный интеллект / Н.В. Баринова, В.Р. Баринов // Вестник Российского экономического университета им. Г.В. Плеханова. – 2021. – Т. 18, № 3 (117). – С. 82–91.

2. Богданова, Т.К. Проблемы моделирования оценки стоимости жилой недвижимости / Т.К. Богданова, А.Р. Камалова, Т.К. Кравченко, А.И. Полторак // Бизнес-информатика. – 2020. – Т. 14, № 3. – С. 7–23.

3. Любимова, Т.В. Решение задачи прогнозирования с помощью нейронных сетей / Т.В. Любимова, А.В. Горелова // Инновационная наука. – 2015. – № 4. – С. 39–43.

4. Морозова, В.И. Прогнозирование методом машинного обучения / В.И. Морозова, Д.И. Логунова // Молодой ученый. – 2022. – № 21 (416). – С. 202–203. https://moluch.ru/archive/416/92048 (дата обращения: 13.12.2023)

5. Министерство экономического развития Российской Федерации [Электронный ресурс]. – https://www.economy.gov.ru (дата обращения: 13.12.2023)

6. Нескоромный, С.В. Индустрия 4.0. Цифровая экономика в России / С.В. Нескоромный, А.А. Варянова // Актуальные исследования. – 2020. –

№ 21 (24). – С. 112–115. – https://apni.ru/article/1362-industriya-40-tsifrovaya-ekonomika-v-rossii (дата обращения: 13.12.2023)

7. СБЕР, официальный сайт [Электронный ресурс]. – http://www.sberbank.ru (дата обращения: 13.12.2023)

8. Статистический сборник "Сведения о рынке жилищного (ипотечного жилищного) кредитования в России" (Интернет-версия) [Электронный ресурс]. – https://cbr.ru/statistics/bank_sector/bank_sector_review/ (дата обращения: 13.12.2023)

9. Федеральная служба государственной статистики Российской Федерации [Электронный ресурс]. – https://rosstat.gov.ru/ (дата обращения: 13.12.2023)

10. Центральный банк России [Электронный ресурс]. – https://cbr.ru (дата обращения: 13.12.2023)

11. Шагалова, П.А., Нейросетевые технологии в решении задач прогнозирования / П.А. Шагалова, Д.А. Ляхманов // Современные проблемы науки и образования. – 2014. – № 6. – https://science-education.ru/ru/article/view?id=16494 (дата обращения: 13.12.2023)

12. Ясницкий, В.Л. Нейросетевое моделирование процессов массовой оценки и сценарного прогнозирования рыночной стоимости жилой недвижимости: автореферат дисс. ... эконом. наук / В.Л. Ясницкий. – Пермь, 2018. – 24 с. – http://www.psu.ru/files/docs/science/dissertatsionnye-sovety/Yasnickiy/avtoreferat_yasnickyi_vl.pdf (дата обращения: 13.12.2023)

13. Территориальный орган Федеральной службы государственной статистики по Челябинской области [Электронный ресурс]. – https://74.rosstat.gov.ru (дата обращения: 13.12.2023)

14. Pollack, G.A. Neural Network System of Sales Volume Forecasting of Residential Real Estate in the Primary Regional Market / G.A. Pollack, O.V. Korobkova, I.Yu. Pollak // Journal of Computational and Engineering Mathematics. – 2023. – V. 10, № 3. – P. 39–53.

*Коробкова Ольга Викторовна, аспирант, кафедра математического и компьютерного моделирования, Южно-Уральский государственный университет (г. Челябинск, Российская Федерация), ufimtcevaov@susu.ru*

*Поступила в редакцию* 15 декабря 2023 г.