ҚАЗАҚСТАН РЕСПУБЛИКАСЫНЫҢ ҒЫЛЫМ ЖӘНЕ ЖОҒАРЫ БІЛІМ МИНИСТРЛІГІ
МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РЕСПУБЛИКИ КАЗАХСТАН
MINISTRY OF SCIENCE AND HIGHER EDUCATION OF THE REPUBLIC OF KAZAKHSTAN

ХАЛЫҚАРАЛЫҚ УНИВЕРСИТЕТІ INTERNATIONAL UNIVERSITY SINCE 2009 МЕЖДУНАРОДНЫЙ УНИВЕРСИТЕТ

ХАЛЫҚАРАЛЫҚ АҚПАРАТТЫҚ ЖӘНЕ КОММУНИКАЦИЯЛЫҚ ТЕХНОЛОГИЯЛАР ЖУРНАЛЫ

МЕЖДУНАРОДНЫЙ ЖУРНАЛ ИНФОРМАЦИОННЫХ И КОММУНИКАЦИОННЫХ ТЕХНОЛОГИЙ

INTERNATIONAL JOURNAL OF INFORMATION AND COMMUNICATION TECHNOLOGIES

2024 (20) 4

қазан - желтоқсан

ISSN 2708–2032 (print)
ISSN 2708–2040 (online)

Халықаралық ақпараттық және коммуникациялық технологиялар журналы
ISSN 2708–2032 (print)
ISSN 2708–2040 (online)
Меншіктенуші: «Халықаралық ақпараттық технологиялар университеті» АҚ (Алматы қ.).
Қазақстан Республикасы Ақпарат және әлеуметтік даму министрлігінің Ақпарат комитетінде – **20.02.2020** жылы берілген.
№ KZ82VPY00020475 мерзімдік басылым тіркеуіне қойылу туралы куәлік.
Тақырыптық бағыты: ақпараттық технологиялар, әлеуметтік-экономикалық жүйелерді дамытудағы цифрлық технологиялар, ақпараттық қауіпсіздік және коммуникациялық технологияларға арналған.
Мерзімділігі: жылына 4 рет.
Тиражы: 100 дана
Редакцияның мекенжайы: 050040, Алматы қ-сы, Манас к-сі, 34/1, 709-кабинет, тел: +7 (727) 244-51-09.
E-mail: ijict@iitu.edu.kz Журнал сайты: https://journal.iitu.edu.kz

---

Международный журнал информационных и коммуникационных технологий
ISSN 2708–2032 (print)
ISSN 2708–2040 (online)

Собственник: АО «Международный университет информационных технологий» (г. Алматы).

Свидетельство о постановке на учет периодического печатного издания в Министерство информации и общественного развития Республики Казахстан **№ KZ82VPY00020475,** выданное от **20.02.2020 г.**

Тематическая направленность: информационные технологии, информационная безопасность и коммуникационные технологии, цифровые технологии в развитии социо-экономических систем**.**

Периодичность: 4 раза в год.

Тираж: 100 экземпляров.

Адрес редакции: 050040 г. Алматы, ул. Манаса 34/1, каб. 709, тел: +7 (727) 244-51-09.

E-mail: ijict@iitu.edu.kz Сайт журнала: https://journal.iitu.edu.kz

---

«International Journal of Information and Communication Technologies»
ISSN 2708–2032 (print)
ISSN 2708–2040 (online)
Owner: International Information Technology University JSC (Almaty).
The certificate of registration of a periodical printed publication in the Ministry of Information and Social Development of the Republic of Kazakhstan, Information Committee No. KZ82VPY00020475, issued on 20.02.2020.
Thematic focus: information technology, digital technologies in the development of socio-economic systems, information security and communication technologies
Periodicity: 4 times a year.
Circulation: 100 copies.
Editorial address: 050040. Manas st. 34/1, Almaty. +7 (727) 244-51-09. E-mail: ijict@iitu.edu.kz
Journal website: https://journal.iitu.edu.kz

МАЗМҰНЫ

ӘЛЕУМЕТТІК-ЭКОНОМИКАЛЫҚ ЖҮЙЕЛЕРДІ ДАМЫТУДАҒЫ ЦИФРЛЫҚ ТЕХНОЛОГИЯЛАР

АҚПАРАТТЫҚ ТЕХНОЛОГИЯЛАР

АҚПАРАТТЫҚ ҚАУІПСІЗДІК ЖӘНЕ КОММУНИКАЦИЯЛЫҚ ТЕХНОЛОГИЯЛАРҒА АРНАЛҒАН

СОДЕРЖАНИЕ

ЦИФРОВЫЕ ТЕХНОЛОГИИ В РАЗВИТИИ СОЦИО-ЭКОНОМИЧЕСКИХ СИСТЕМ

ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ

ИНФОРМАЦИОННАЯ БЕЗОПАСНОСТЬ И КОММУНИКАЦИОННЫЕ ТЕХНОЛОГИИ

CONTENT

DIGITAL TECHNOLOGIES IN THE DEVELOPMENT OF SOCIO-ECONOMIC SYSTEMS

INFORMATION TECHNOLOGY

INFORMATION SECURITY AND COMMUNICATION TECHNOLOGIES

INTERNATIONAL JOURNAL OF INFORMATION AND COMMUNICATION TECHNOLOGIES
ISSN 2708–2032 (print)
ISSN 2708–2040 (online)
Vol. 5. Is. 4. Number 20 (2024). Pp. 105–117
Journal homepage: https://journal.iitu.edu.kz
https://doi.org/10.54309/IJICT.2024.20.4.009

MODELS AND INFORMATION TECHNOLOGY FOR REAL ESTATE VALUATION USING MACHINE LEARNING ALGORITHMS

I.L. Khlevna*, A.O. Buzyurova, A.O. Khlevnyi
Taras Shevchenko National University of Kyiv, Kyiv, Ukraine.
E-mail: yuliia.khlevna@knu.ua

Iulia Khlevna — Doctor of Technical Science, professor, Department of Technology Management, Taras Shevchenko National University of Kyiv, Kyiv, Ukraine.
E-mail: yuliia.khlevna@knu.ua, https://orcid.org/0000-0002-1874-1961;
Anna Buzyurova — student, Department of Technology Management, Taras Shevchenko National University of Kyiv, Kyiv, Ukraine.
Andrii Khlevnyi — PhD, associate professor, Department of Technology Management, Taras Shevchenko National University of Kyiv, Kyiv, Ukraine.
https://orcid.org/0000-0001-8942-6670.

Abstract. It has been established that one of the main challenges in the field of real estate valuation is processing large volumes of information. Traditional valuation methods, which rely on expert experience, can be limited and prone to human error. The application of Data Science enables the automation of data collection, analysis, and processing, leading to more accurate and objective results. The article explores the development of Data Science technology for real estate valuation, including data analysis, the use of statistical methods and machine learning techniques, and the creation of a valuation model based on the obtained results. Existing approaches to the use of analytical methods in real estate valuation tasks are examined. A new methodology for their application is developed, along with a justification of the feasibility and necessity of implementing the proposed methodology. Recommendations for the practical implementation of the methodology are provided.

Keywords: forecasting, machine learning, real estate market, Data Science
For citation: I.L. Khlevna, A.O. Buzyurova, A.O. Khlevnyi MODELS AND INFORMATION TECHNOLOGY FOR REAL ESTATE VALUATION USING MACHINE LEARNING ALGORITHMS. 2024. Vol. 5. No. 20. Pp. 105–117 (In Eng.). https://doi. org/10.54309/IJICT.2024.20.4.009.

МОДЕЛЬДЕР ЖӘНЕ ЖЫЛЖЫМАЙТЫН МҮЛІКТІ БАҒАЛАУ ҮШІН АҚПАРАТТЫҚ ТЕХНОЛОГИЯЛАР МАШИНАЛЫҚ ОҚЫТУ АЛГОРИТМДЕРІН ҚОЛДАНУМЕН

Ю.Л. Хлевна*, А.О. Бузюрова, А.О. Хлевный
Киев қаласындағы Тарас Шевченко атындағы Ұлттық университеті, Украина.
E-mail: yuliia.khlevna@knu.ua

Юлия Хлевна — техника ғылымдарының докторы, профессор, Технологияларды басқару кафедрасы, Тарас Шевченко атындағы Ұлттық университеті, Киев, Украина.
E-mail: yuliia.khlevna@knu.ua, https://orcid.org/0000-0002-1874-1961;
Анна Бузюрова — студентка, Технологияларды басқару кафедрасы, Тарас Шевченко атындағы Ұлттық университеті, Киев, Украина.
Андрей Хлевной — PhD, доцент, Технологияларды басқару кафедрасы, Тарас Шевченко атындағы Ұлттық университеті, Киев, Украина.
https://orcid.org/0000-0001-8942-6670.

Аннотация. Жылжымайтын мүлікті бағалау саласындағы басты мәселелердің бірі – үлкен көлемдегі ақпаратты өңдеу екені анықталды. Сарапшылардың тәжірибесіне негізделген дәстүрлі бағалау әдістері шектеулі және адам қателіктеріне бейім болуы мүмкін. Data Science қолдану деректерді жинау, талдау және өңдеу процесін автоматтандыруға мүмкіндік береді, бұл дәлірек және объективті нәтижелер алуға жол ашады. Мақалада жылжымайтын мүлікті бағалауға арналған Data Science технологиясын әзірлеу, соның ішінде деректерді талдау, статистикалық әдістер мен машиналық оқыту әдістерін пайдалану, алынған нәтижелер негізінде бағалау моделін құру мәселелері қарастырылған. Жылжымайтын мүлікті бағалау міндеттерінде аналитикалық әдістерді қолданудың қолданыстағы тәсілдері зерттелген. Олардың қолданылуына арналған жаңа әдістеме әзірленіп, ұсынылған әдістемені енгізудің орындылығы мен қажеттілігі негізделді. Әдістемені практикалық іске асыру бойынша ұсыныстар берілген.

Түйін сөздер: болжау, машиналық оқыту, жылжымайтын мүлік нарығы, Data Science

Дәйексөздер үшін: Ю.Л. Хлевна, А.О. Бузюрова, А.О. Хлевный. МОДЕЛЬДЕР ЖӘНЕ ЖЫЛЖЫМАЙТЫН МҮЛІКТІ БАҒАЛАУ ҮШІН АҚПАРАТТЫҚ ТЕХНОЛОГИЯЛАР МАШИНАЛЫҚ ОҚЫТУ АЛГОРИТМДЕРІН ҚОЛДАНУМЕН. 2024. Vol. 5. No. 20. Pp. 105–117 (Ағылшын тілінде). https://doi.org/10.54309/IJICT.2024.20.4.009.

МОДЕЛИ И ИНФОРМАЦИОННАЯ ТЕХНОЛОГИЯ ОЦЕНКИ НЕДВИЖИМОСТИ С ИСПОЛЬЗОВАНИЕМ АЛГОРИТМОВ МАШИННОГО ОБУЧЕНИЯ

И.Л. Хлевна*, А.О. Бузюрова, А.О. Хлевный
Киевский национальный университет имени Тараса Шевченко, Киев, Украина.
E-mail: yuliia.khlevna@knu.ua

Юлия Хлевна — доктор технических наук, профессор, кафедра управления технологиями, Киевский национальный университет имени Тараса Шевченко, Киев, Украина.
E-mail: yuliia.khlevna@knu.ua, https://orcid.org/0000-0002-1874-1961;
Анна Бузюрова — студентка, кафедра управления технологиями, Киевский национальный университет имени Тараса Шевченко, Киев, Украина.
Андрей Хлевной — кандидат технических наук, доцент, кафедра управления технологиями, Киевский национальный университет имени Тараса Шевченко, Киев, Украина.
https://orcid.org/http://orcid.org/0000-0001-8942-6670.

Аннотация. Установлено, что одной из основных проблем в области оценки недвижимости является обработка больших объемов информации. Традиционные методы оценки, основанные на экспертном опыте, могут быть ограничены и подвержены человеческим ошибкам. Применение технологий Data Science позволяет автоматизировать процесс сбора, анализа и обработки данных, что обеспечивает получение более точных и объективных результатов. В статье рассматривается разработка технологии Data Science для оценки недвижимости, включая анализ данных, использование статистических методов и методов машинного обучения, создание модели оценки на основе полученных результатов. Исследуются существующие подходы к применению аналитических методов в задачах оценки объектов недвижимости. Разработана новая методика их применения, обоснована целесообразность и необходимость внедрения предложенной методики. Даны рекомендации по практической реализации методики.

Ключевые слова: прогнозирование, машинное обучение, рынок недвижимости, Data Science

Для цитирования: Ю.Л. Хлевна, А.О. Бузюрова, А.О. Хлевной. МОДЕЛИ И ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ ОЦЕНКИ НЕДВИЖИМОСТИ С ИСПОЛЬЗОВАНИЕМ АЛГОРИТМОВ МАШИННОГО ОБУЧЕНИЯ. 2024. Vol. 5. No. 20. Pp. 105–117 (На англ.). https://doi.org/10.54309/IJICT.2024.20.4.009.

Introduction

In the modern world, the increasing volume of data and the growing need for its analysis are becoming more relevant than ever. In the real estate sector, this process is particularly significant, as property valuation is a critical step in various processes, such as sales, purchases, rentals, and more. The development of Data Science technology in this field can enhance the accuracy and efficiency of valuations, which, in turn, can have a positive impact on the real estate market.

One of the main challenges in real estate appraisal is managing large amounts of information. Traditional valuation methods, which rely on expert experience, can be limited and prone to human error. By leveraging Data Science, it is possible to automate the processes of data collection, analysis, and processing, leading to more accurate and objective resu

Data Science methods can be applied to develop real estate valuation models based on statistical analyses and machine learning. These models enable the prediction of property values based on various factors, such as house size, the number of rooms, location, the availability of amenities, and more. This approach ensures more objective and precise valuations that take into account a wide range of factors. Automated algorithms and software solutions can significantly reduce the time and effort required for valuation, making the process faster and more efficient, while allowing more time for data analysis and informed decision-making.

The use of Data Science methods in real estate valuation holds great potential for improving the accuracy, objectivity, and efficiency of the appraisal process. Considering the vast amount of available data and analytical capabilities, Data Science facilitates more well-informed decisions regarding property purchases, sales, rentals, or investments. The data analysis process relies on objective criteria and algorithms, minimizing the influence of subjective factors or personal biases. This is crucial in real estate, where valuations can often be subjective and influenced by the assessor's personal opinions and experience.

Given the above, the development of Data Science technologies for real estate valuation, including data analysis, the use of statistical and machine learning methods, and the creation of valuation models based on these results, is a highly relevant and promising area of research.

Material and Methods

An analysis of literary sources demonstrates that the application of data analysis models is a crucial component of real estate valuation projects. Different models can be utilized for various aspects of such projects, including data collection, processing, and analysis.

For instance, in the study (Miroshnychenko et al., 2022), the authors explore machine learning methods for developing models to predict the prices of suburban real estate. They identify factors influencing suburban property valuation and build four predictive models: linear regression, ridge regression, random forest, and XGBoost. A comparison of results revealed that the XGBoost algorithm was the most effective (RMSE = 67669.15, MAE = 47063.49, MAPE = 0.1034903). Compared to multiple linear regression, it was concluded that the XGBoost model better explains the data, providing a higher probability of accurately forecasting actual property prices.

In another study (Pashkevych et al., 2022), a statistical model was developed to predict housing prices using linear regression. To evaluate the relationship between various property characteristics, 3D modeling was employed. The authors performed data clustering and testing on real-world data, achieving satisfactory predictive results. The model was tested on real properties, achieving an accuracy rate of 76 %.

In study (Velytchenko, 2020), a neural network was built to estimate property values. The author used both numerical and categorical attributes, as well as four images of the property. The resulting mean absolute percentage error (MAPE) on the test set was 17.52 %.

In work (Ivashchenko, 2022), the author compared two models for price prediction: decision tree and random forest. While the errors appeared relatively large, this was attributed to extreme deviations in predictions. The author noted that such deviations should be considered in further research. The random forest model was found to deliver more accurate and reliable predictions than the decision tree model.

In article (Khlevna et al., 2021), a variety of models were analyzed for predicting real estate market prices, including Lasso regression, Elastic Net regression, Ridge regression, Gradient Boosting regression, and XGBoost. It was determined that aggregated models could improve results. The best-performing model was an ensemble of Lasso regression, Extreme Gradient Boosting, Elastic Net, and Ridge regression, achieving RMSE = 0.1091 and a standard deviation of 0.0075.

Another intriguing approach, discussed in study (Quanzeng et al., 2017), involved the use of image-based valuation. The authors investigated how visual features, reflecting property characteristics, could aid in estimating property values. They developed algorithms that relied solely on location and photo attributes. Recurrent neural networks were built and analyzed, achieving MAE = 66.3 and MAPE = 16.11 % on one dataset, and MAE = 13.32 and MAPE = 22.69 % on another.

In study (Yaroshenko et al., 2012), neural networks were also developed to assess real estate, focusing on quantitative and qualitative attributes. The model utilized real estate agency data over five months. The mean absolute error in predictions was 3.41 %. The authors speculated that this error was due to an insufficient number of factors affecting price formation.

In work (Alisha Kuvalekar, 2020), the authors collected data from various real estate websites in Mumbai and applied methods such as SVM, Random Forest, Linear Regression, Multiple Linear Regression, Decision Tree Regressor, and KNN. The best results were achieved by the Decision Tree Regressor, with an accuracy of 89 %.

In article (Kintzel Joseph, 2019), models were trained on a triad of data types: numerical, spatial, and image-based. Comparisons included models such as OLS, Neural Network, Random Forest, and Gradient Boosting. The Random Forest and Gradient Boosting regression algorithms performed best in predicting prices, with average error coefficients of 0.0758 and 0.0756, respectively. When combined with image evaluations (PCA-generated features), Gradient Boosting further reduced the error coefficient to 0.0728. Neural networks with PCA features had an error coefficient of 0.0930.

Finally, study (Kovpak et al., 2016: 56–60) explored a type of multiple regression using dummy variables to describe the current condition and location characteristics of real estate. The resulting average relative approximation error for the model was 11.72 %.

A range of specialized software and technologies is available for predicting real estate prices, including web services and mobile applications. These programs offer users a broad array of tools for analyzing the real estate market, such as access to historical price data, visualization of geographic trends, and the ability to create forecasts based on various machine learning models.

Many of these platforms integrate interactive maps, enabling users to explore different neighborhoods and their market characteristics, such as average property prices, price dynamics, infrastructure, and more. Some services also provide comparative analyses of property prices across various locations and property types.

These technologies are not only helpful for investors and buyers in making informed decisions about real estate investments but also serve as valuable tools for real estate agents, enabling them to provide more objective and detailed information to their clients.

Notable Examples of Real Estate Technologies:

Zillow (https://www.zillow.com): One of the most popular real estate price forecasting tools developed by an American real estate services company. Using an integrated algorithm, Zillow estimates property values by considering factors like location, size, condition, and market trends. Forecasts are made for up to 12 months, incorporating county-level economic data and property-specific details.

CoreLogic (https://www.corelogic.com): This service offers robust analytical tools for predicting real estate prices at national, state, and metropolitan levels. It provides monthly updates, with forecasts spanning up to 30 years depending on the subscription level. Data sets are updated monthly, and results are published five weeks after the end of each month.

PropMix (propmix.io): Using a 25-terabyte database with diverse information, including images, PropMix employs artificial intelligence and computer vision to analyze real estate prices and generate forecasts.

Realyse (www.realyse.com): A UK-based platform providing detailed real estate market insights, helping investors, developers, and agents make informed price forecasts. It collects data from multiple sources and offers APIs or downloadable CSV files for forecast-related information.

Redfin Estimate (www.redfin.com/redfin-estimate): An online tool leveraging a variety of data sources to determine the approximate value of a residential property. Known for its accuracy, with an average error rate of 2.06 %, Redfin Estimate uses MLS data from recently sold homes to calculate current market values.

HouseCanary (www.housecanary.com): This solution predicts property prices—both for individual homes and the broader market—over a period of up to three years using machine learning and time series analysis. The platform also provides comprehensive analytics evaluating market conditions, desirability, and stability.

State Property Fund of Ukraine (www.spfu.gov.ua): This service analyzes property data by comparing multiple sources (Lun, OLX, contracts, reports) and provides official valuation certificates. However, as of 2021, the system does not account for property or apartment conditions, which can significantly affect valuations.

Features of Real Estate Technologies:

These tools utilize large databases containing information on market prices, geographic data, historical transaction data, and other factors influencing property values. Machine learning and AI algorithms analyze this data to produce highly accurate forecasts.

Limitations and Considerations:

Data Updates: Continuous updates are crucial to maintain accuracy. Real estate markets are dynamic, with changes in transactions, legal environments, and other factors affecting property values. Automated updating mechanisms are essential for these systems.

Market Focus: Most services target specific markets, such as the U.S. or U.K., due to the unique characteristics and data availability of these regions. This geographic focus enables a deeper analysis of local market conditions, providing more precise forecasts.

Cost Barriers: Accessing forecasts often requires paid subscriptions. For some users, these costs may be prohibitive, limiting their ability to utilize these tools effectively. Moreover, users might be reluctant to pay for a service they perceive as insufficiently valuable, creating a barrier to widespread adoption.

Despite these limitations, such technologies play a pivotal role in improving the precision and accessibility of real estate market analysis, benefiting investors, buyers, and agents alike.

The analysis of the reviewed literature indicates that the best-performing models are gradient boosting, random forest, and neural networks. Therefore, this study will focus on these models using a new dataset. Additionally, one of the ensemble models (gradient boosting or random forest) with the best results will be selected, and a combined model with a neural network will be developed to evaluate the potential for improving prediction accuracy through this approach.

The aim of this research is to enhance the effectiveness of real estate price prediction by formalizing data processing methods.

Results and Discussion

To predict real estate prices, the following models were selected for development: gradient boosting, random forest, and neural networks.

Gradient Boosting is an ensemble machine learning method, a form of bagging, which iteratively builds weak models to improve predictive accuracy. The goal of gradient boosting, like any supervised learning algorithm, is to define a loss function and minimize it (Trevor, 2017). This method is known for its high accuracy and is widely used in various fields, including image recognition, time series forecasting, and recommendation systems.

The optimization function for gradient boosting is represented by the following formula:

$$L(t) = \sum^{n}{}_{i}{=}^{1}1l(y_i, f^{-1}{}_i{}'t(x_i)) + \Omega(f_i), \quad [1]$$

where

l – loss function,

$y_i, f^{-1}1$ – the value of the i-th element of the training sample and the sum of the predictions of the first t trees, respectively,

x_i – feature set of the i-th element of the training sample,

f_i – the function (in our case, a tree) that we want to train at step t,

$f_i(x_i)$ – prediction on the i-th element of the training sample,

$\Omega(f)$ – regularization of function f. $\Omega(f) = \gamma T + ½\lambda\|w\|^2$, where T – number of tree vertices, w – values at the leaves, and γ and λ – regularization parameters.

Using the Taylor expansion to the second term, the optimization function L(t) can be approximated by:

$$L(t) = \sum^{n}{}_{i}{=}^{1}1l(y_i, f^{-1}1_i + g_i f_i(x_i) + 0.5h_i f^2{}_i(x_i)) + \Omega(f_i), \quad [2]$$

where

$g_i = \partial l(y_i, f^{-1}1)/\partial f^{-1}1$,

$h_i = \partial 2l(y_i, f^{-1}1)/\partial 2f^{-1}1$.

Since the goal is to minimize model error on the training set, the minimum of L(t) must be found for each t.

Random Forest is also an ensemble machine learning method, similar to gradient boosting, which employs a combination of multiple decision trees to develop predictions. It is one of the most popular methods for classification and regression, known for its flexibility, ease of use, and efficiency.

Neural Networks are a machine learning method inspired by the functioning of the human brain. This method is based on the use of «neurons» – computational units that can learn to recognize patterns in data.

In this work, a deep neural network will be constructed. This is a type of artificial neural network with two or more hidden layers between the input and output layers. These additional layers enable the network to learn more complex and abstract functions, translating simple input data into more intricate structures. For example, in image processing tasks, the initial layers may detect simple shapes, such as lines and circles, while deeper layers combine these simple shapes to detect more complex patterns, such as human faces.

Deep Neural Networks (DNN) are often associated with deep learning, a branch of machine learning focused on using deep networks. The training process for deep networks typically utilizes algorithms based on the gradient descent method, such as stochastic gradient descent (SGD), RMSprop, Adam, and others.

The architecture of the constructed neural network is shown in Figure 1.

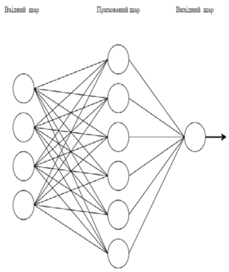

Fig. 1 - Neural Network Diagram

To activate the layers, the ReLU activation function was chosen, which has the following mathematical formula:

$$f(x) = \max(x, 0) \quad [3]$$

For optimizing the training process, the Adam function was selected. Adam combines the approaches of the AdaGrad and RMSProp methods. It also adjusts the learning steps individually for parameters, taking into account the «history» of gradient changes on them (Marusyk, 2019).

Let

$$v = \beta 1v + (1 - \beta 1)d\theta,$$
$$s = \beta 2s + (1 - \beta 2)d\theta 2\ ,$$
$$\hat{w_v} = w_v/(1 - \beta 1^n),$$
$$\hat{s}_v = s_v/(1 - \beta 2^n), \quad [3]$$

where v and s are referred to as the first and second moment estimates, $\hat{w_v}$ and $\hat{s}_v$ – bias-corrected moment estimates. Then parameter changes occur as follows:

$$\theta = \theta - \lambda \hat{w_v}/\sqrt{\hat{s}_v} + \varepsilon \quad [4]$$

Mean Squared Error (MSE) is used to calculate losses. This loss function computes the square of the difference between predicted and true values and then averages these squares:

. [5]

To evaluate the obtained predictions and compare the mentioned methods, metrics such as MAE, MAPE, RMSE, and the coefficient of determination will be used.

MAE is a metric for evaluating the accuracy of a predictive model. It calculates the difference between predicted values and actual values, taking the absolute value of these differences and averaging across all data points. MAE measures the absolute error of the forecast, regardless of direction. The smaller the MAE value, the higher the model's accuracy.

. [6]

MAPE is another metric for evaluating the accuracy of a predictive model. It calculates the percentage difference between predicted and actual values, taking the absolute value of these differences and averaging the percentage across all data points. MAPE measures the percentage error of the forecast. MAPE errors serve to evaluate the percentage deviation of the forecast relative to the actual value. Typically expressed as a percentage, the smaller the MAPE value, the higher the model's accuracy.

. [7]

RMSE is yet another metric for evaluating the accuracy of a predictive model. It computes the square root of the mean squared error, which is the sum of the squared differences between predicted and actual values, divided by the number of data points, and then takes the square root of the resulting value. RMSE measures the root mean square error of the forecast. The smaller the RMSE value, the higher the model's accuracy.

Coefficient of Determination is a metric that measures how well the model predicts actual data. It represents the ratio of the variation in model predictions to the total variation in the data. The coefficient of determination ranges from 0 to 1, where a value of 0 means the model does not explain any variation in the data, and a value of 1 means the model fully explains the data variation. The higher the R-squared value, the better the model predicts the data. However, this metric can be sensitive to overfitting and the number of variables.

- [8]

All these metrics are used to assess and compare the accuracy of predictive models.

The chosen programming language for implementing the described methods was Python (docs.python.org/3.9). The decision to use this language was driven by its robustness and widespread popularity in the fields of data analysis and Data Science.

The input data was decided to be sourced from a real estate information portal (rieltor.ua). The data was preprocessed and cleaned beforehand.

The XGBoost library was used to build the gradient boosting model. Experiments were conducted with various parameter sets, but the best-performing parameters are shown in Figure 2. A model was created to train on the training dataset using the fit method and predict prices on the test dataset using the predict method.

```
model = XGBRegressor(n_estimators=100,
                     learning_rate=0.1,
                     subsample=0.7,
                     colsample_bytree=0.9,
                     alpha=0.5)
model.fit(X_train, y_train)
y_pred = model.predict(X_test)
```

Fig. 2 - Code snippet of the XGBoost model

The random forest model was built using the sklearn library. Experiments were also conducted with different numbers of estimators, and ultimately, two hundred estimators were chosen (Fig. 3).

```
model = RandomForestRegressor(n_estimators=200,
                              random_state=42)
model.fit(X_train, y_train)
y_pred = model.predict(X_test)
```

Fig.3 - Code snippet of the Random Forest model

Similarly to the previous models, the best parameters were selected for the neural network. In this case, attempts were made to use different activation functions (sigmoid, linear, ReLU), various numbers of input and hidden layers, the number of epochs, and so on.

The Keras library was used for the construction. The code for building the model is shown in Figure 4.

```
model = Sequential([
    Dense(64, activation='relu', input_shape=(df[features].shape[1],)),

    Dense(64, activation='relu'),
    Dense(64, activation='relu'),
    Dense(64, activation='relu'),
    Dropout(0.2),
    Dense(128, activation='relu'),
    Dropout(0.2),
    Dense(256, activation='relu'),
    Dropout(0.2),
    Dense(128, activation='relu'),
    Dropout(0.2),
    Dense(64, activation='relu'),
    Dropout(0.2),

    Dense(1)
])

model.compile(optimizer='adam', loss='mean_squared_error', metrics=['mae'])
early_stopping = EarlyStopping(monitor='val_mae', patience=15, mode='min', restore_best_weights=True)
model.fit(X_train, y_train, epochs=100, batch_size=64, callbacks=[early_stopping], validation_data=(X_test, y_test))
y_pred = model.predict(X_test)
```

Fig.4 - Code snippet with the neural network model

The approach of combining gradient boosting and neural network models opens up significant opportunities for improving prediction quality. Since the gradient boosting model has already demonstrated better results compared to the random forest, leveraging this model in combination with neural network predictions can be a highly effective strategy.

The quality assessments of this combined model are presented in Table 1 and compared to standalone gradient boosting.

Table 1

Quality Assessments of the Constructed Models

Model	MAE	MAPE	RMSE	R2 score
Gradient Boosting	39498.78	0.261	92911.20	0.81
Random Forest	38738.20	0.245	101943.28	0.77
Neural Network	49712.98	0.286	141427.70	0.56
Combined Model	10064.71	0.107	14337.08	0.99

To utilize the developed model, it was decided to create an information system, specifically a web service for property valuation. To build a web service based on the developed model, a user-friendly and intuitive interface needs to be designed. Users should be able to easily input property details, such as location, area, number of rooms, and so on.

The concept of an information system for property valuation includes the following components:

1. Database: This is the central element of the system, containing a vast amount of information about real estate in Kyiv. It holds extensive data on the real estate market, including apartment sales records. The database stores various apartment attributes used for value estimation, such as location, proximity to the metro, area, floor, and the total number

of floors in the building. These data points are stored in corresponding database tables and are utilized for analyzing and predicting property values using machine learning models. The database is continuously updated to include new apartment sales data and their characteristics, ensuring the relevance and accuracy of valuations.

2. Analytical Tools: The information system should include analytical tools for data processing and analysis, specifically using the predictive pricing model developed in this project. This model analyzes apartment parameters stored in the database and uses them to generate price forecasts. The predictive pricing model is maintained and enhanced with the latest analytical methods and algorithms to ensure its accuracy and reliability. It can be periodically updated based on new data to account for real estate market changes and improve forecast quality.

3. User Interface: The information system must have a user-friendly and intuitive interface. This could be, for example, a web application or service that allows users to quickly and conveniently access real estate information, calculate property values, and perform other actions.

The concept of a property valuation information system for the city of Kyiv can be implemented by combining a database, analytical tools, and a user-friendly interface. This system will simplify and make the property valuation process in Kyiv more objective.

Relational databases such as PostgreSQL, MySQL, or SQLite are commonly used to store real estate data. They provide structured data storage with the ability to perform complex queries while ensuring data reliability and consistency. Therefore, it was decided to use one of these databases for this project, specifically a MySQL database.

The logical model of the database (Figure 5) represents the structure of the data and the relationships between them without being tied to a specific database management system (DBMS). It describes the structure of the database at the level of concepts and entities, without focusing on technical implementation aspects. It includes entities, their attributes, relationships, and keys—both primary and foreign.

Creating a web service for real estate price prediction that includes a form for apartment price calculation opens up broad opportunities in various fields.

Firstly, it provides convenient access to predictions anytime and anywhere. Users can access the service from any device with an Internet connection, such as a computer, tablet, or smartphone. This allows users to access price forecasts even on the go or in places without access to a stationary computer. The web service also features an intuitive and user-friendly interface, enabling users to navigate and utilize its features quickly. The price calculation form is straightforward and easy to fill out. Furthermore, users will be able to save the prediction results or send them to their email for further analysis or comparison, making it convenient to use the data in their work.

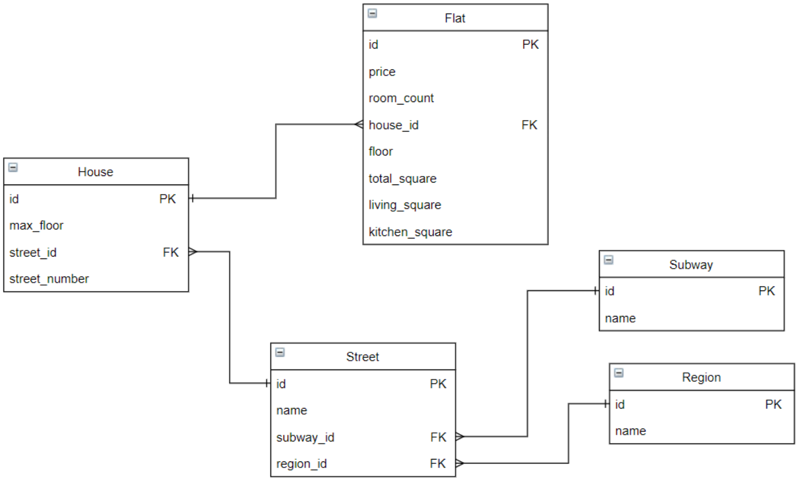

Fig. 5 - Logical Database Model

Secondly, the price calculation form allows users to receive personalized predictions by considering the unique characteristics of each apartment. Each parameter is crucial for determining the property value, and users can freely adjust them according to their needs and requirements.

This web service provides users with valuable information for making informed decisions regarding the purchase, sale, or rental of real estate in the city.

The collected data from the web service can be used for further analysis of the real estate market, allowing for the observation of trends and conducting research in this field. For example, the data provided to users for price calculations can be aggregated and used for market trend analysis. Alternatively, collecting data on real estate prices enables the analysis of supply and demand in the market. This helps to understand which types of properties are most popular among buyers and which areas have the highest demand.

The accumulated data can also be used to prepare analytical reports and market condition summaries. This can be valuable for real estate agencies, urban planning development, and other market participants.

Conclusion

Three models for predicting real estate prices were developed. These models included gradient boosting, neural networks, and random forest. Each of these models was carefully fine-tuned and trained using the prepared data.

After analyzing their performance, the best ensemble model was selected to construct a combined model, which ultimately provided significantly better predictions than the standalone models. This highlights the value of combining different models to improve the accuracy and efficiency of predictions.

The combined model can be more versatile and precise in predicting real estate prices, enabling better decision-making in the real estate market. A conceptual description of the developed information system and database schema has been provided

REFERENCES

Alisha Kuvalekar, Shivani Manchewar, Sidhika Mahadik, Shila Jawale (2020). «House Price Forecasting Using Machine Learning». — 2020, URL: https://deliverypdf.ssrn.com/delivery.php?

CoreLogic.com: Website. URL: https://www.corelogic.com/

HouseCanary.com: Website. URL: https://www.housecanary.com/

Ivashchenko O. (2022). «Study and Implementation of a Classification Method for Real Estate Price Prediction». — 2022. — URL: https://openarchive.nure.ua/server/api/core/bitstreams/128e53e3-4220-495f-993a-a8cd4a9519f0/content

Khlevna Y., Bura Y. (2021). «Information Support for Real Estate Price Prediction Using Machine Learning Methods». —2021. — URL: https://ir.library.knu.ua/knurepo/bitstream/handle/123456789/823/Bura_magistr.pdf?sequence=1&isAllowed=y

Kintzel Joseph (2019). «Price Prediction and Computer Vision in the Real Estate Marketplace». — 2019. — URL: https://dash.lib.harvard.edu/bitstream/handle/1/37365260/KINTZEL-DOCUMENT-2019.pdf?sequence=1&isAllowed=y

Kovpak E.O. & Malets V.V. (2016). Modeling Apartment Prices Using Multiple Regression, *Economy and Society*. — 2016. — Pp. 56–60.

Miroshnychenko I.V., Krasheninnikova O.V. (2022). «Real Estate Price Prediction Using Machine Learning Algorithms». —2022. — URL: http://www.economy.nayka.com.ua/pdf/1_2022/83.pdf

Marusyk O.M. (2019). Optimal Tuning System for Deep Belief Neural Networks. — 2019. — URL: https://ela.kpi.ua/server/api/core/bitstreams/a9e64d6f-147d-4ab8-99bb-2745ee03cd60/content

Pashkevych O., Vashchyshchak S., Boychuk A., Styslo T., Demchyna M. (2022). «Application of Machine Learning Models for Real Estate Market Price Prediction». — 2022. — URL: http://journals.khnu.km.ua/vestnik/wp-content/uploads/2023/01/315-1-170-178.pdf

PropMix.io: Website. URL: https://propmix.io/

Python documentation: URL: https://docs.python.org/3.9/

State Property Fund of Ukraine: Website. — URL: https://www.spfu.gov.ua/

Trevor Hastie Robert (2017). Tibshirani Jerome Friedman, The Elements of Statistical Learning. — 2017. — URL: https://hastie.su.domains/Papers/ESLII.pdf

Quanzeng You, Ran Pang, Liangliang Cao, and Jiebo Luo. (2017). «Image-Based Appraisal of Real Estate Properties». —2017. — URL: https://arxiv.org/pdf/1611.09180.pdf

Realyze.com: Website. URL: https://www.realyse.com/

RedFin.com: Website. URL: https://www.redfin.com/redfin-estimate

Rieltor ua: URL: https://rieltor.ua/

Velytchenko Y.V. (2020). «Intelligent System for Assessing the Market Value of Housing». — 2020. — URL: https://ela.kpi.ua/bitstream/123456789/38428/1/Velitchenko_magistr.pdf

Yaroshenko O., Hryhorkiv V., Filipchuk N. (2012). «Neural Networks and Their Use for Predicting Real Estate Market Trends». — 2012. URL: http://www.irbis-nbuv.gov.ua/cgi-bin/irbis_nbuv/cgiirbis_64.exe?I21DBN=LINK&P21DBN=UJRN&Z21ID=&S21REF=10&S21CNR=20&S21STN=1&S21FMT=ASP_meta&C21COM=S&2_S21P03=FILA=&2_S21STR=nvnltu_2012_22.5_58

Zillow.com: Website. URL: https://www.zillow.com/ ID=48506402706507 4 0 7 6 0 9 4 0 7 7 0 9 7 0 2 1 1 1 6 1 0 7 0 0 9 0 4 8 0 7 2 0 4 3 0 5 7 0 2 6 1 1 0 0 9 4 1 1 4 0 6 9 1 1 9 0 0 5 0 0 9 1 0 6 0 7 1 0 7 3 0 3 8 0 1 0 1 1 1 0 4 8 0 5 3 0 3 0 0 1 1 0 2 7 1 0 7 0 9 0 0 8 8 1 2 7 0 6 4 0 7 0 1 2 1 0 7 5 0 6 1 0 1 3 0 8 1 1 -

ХАЛЫҚАРАЛЫҚ АҚПАРАТТЫҚ ЖӘНЕ КОММУНИКАЦИЯЛЫҚ ТЕХНОЛОГИЯЛАР ЖУРНАЛЫ

МЕЖДУНАРОДНЫЙ ЖУРНАЛ ИНФОРМАЦИОННЫХ И КОММУНИКАЦИОННЫХ ТЕХНОЛОГИЙ

INTERNATIONAL JOURNAL OF INFORMATION AND COMMUNICATION TECHNOLOGIES

Правила оформления статьи для публикации в журнале на сайте:
https://journal.iitu.edu.kz
ISSN 2708–2032 (print)
ISSN 2708–2040 (online)
Собственник: АО «Международный университет информационных технологий» (Казахстан, Алматы)

ОТВЕТСТВЕННЫЙ РЕДАКТОР
Мрзабаева Раушан Жалиқызы

КОМПЬЮТЕРНАЯ ВЕРСТКА
Асанова Жадыра

Подписано в печать 15.12.2024.
Формат 60x881/8. Бумага офсетная. Печать - ризограф. 9,0 п.л. Тираж 100
050040 г. Алматы, ул. Манаса 34/1, каб. 709, тел: +7 (727) 244-51-09).

Издание Международного университета информационных технологий
Издательский центр КБТУ, Алматы, ул. Толе би, 59