UDC 004.67

V. Konchynskyi , T. Likhouzova

MODELS FOR FORMING THE MARKET VALUE OF A REAL ESTATE

Abstract: The problem of a high-quality and quick assessment of real estate is considered. It is proposed to use machine learning models and technologies to create a real estate evaluation service based on the entered characteristics. To evaluate the effectiveness of the proposed models, data on already concluded agreements were used.

Keywords: machine learning, cost forecasting, cloud computing.

Introduction

The value of real estate objects is affected by many factors that can be directly related to these objects, for example, their characteristics, or not be directly related, but affect the situation in the real estate market, for example, regional factors or the exchange rate currencies [1,2]. While the former is obvious to the average person looking to sell their property, the latter is not so obvious, their impact is easily overlooked and difficult to calculate.

When housing is put up for sale, the seller determines the price intuitively rather than reasonedly. But all factors take numerical values or take one option from a limited number of predefined ones that can also be specified as numbers. Since in this case, it is worth using a comparative valuation method, it is possible to find a relationship between the values of these factors and the value of the real estate object based on information about recent transactions and it is possible to create software that will calculate the value. This software will allow sellers to get the best price, at which they can close the deal and get the maximum profit, simply by entering all the necessary information about their home.

Analysis of existing solutions

In Ukraine, there are several services that allow you to evaluate housing for sale. For example, such a service was created by the State Property Fund of Ukraine [3], in which you can evaluate housing by entering the necessary information and get a certificate of value. The consulting company "Uvekon" [4] provides the same opportunity. You can also estimate the cost of housing quickly on the website Dom.ria.com, which, unlike the previous examples, does not require registration in the system.

The listed implementations perform the evaluation by searching for deals and advertisements with real estate objects that are close to the evaluated object and have similar characteristics. This approach has several disadvantages.

ISSN 1560-8956

The first disadvantage is that sometimes there may be very few such objects or none at all since we are only interested in recent transactions. Due to the lack of information, the result of the assessment can be very far from the real value. For example, Uvekon's "Appraiser's Calculator" will look for properties for comparison from other areas, which may be very different, if there is a shortage of close options.

The second disadvantage is the limitation of parameter values. For comparison, a sample with a limited number of options is created, so there will be limitations for the parameters as well. If the entered parameter will go beyond these limits, the evaluation result will be incorrect.

These shortcomings can be eliminated by finding a relationship between the values of the input parameters and the cost. If there aren't enough samples in an area to compare, we can identify regional factors that affect cost. Having analyzed these factors and determined their influence, it is enough for us to know the value of these parameters for this region. This dependence can be found using machine learning technologies.

The purpose of the work is to optimize the process of real estate evaluation, in particular, to improve the quality of the forecast of the value of the real estate object with a limited period of sale. The main task is to create a software solution for estimating the value of real estate using machine learning technologies. For evaluation, users will only need the address of the real estate object and its characteristics and the date of sale if the owner wants to close the deal sooner. The regional parameters for the given address will be extracted from the database, which will be filled in and updated by the developers.

Materials and methods

All input parameters can be set with numerical values, the output value is also numerical, so you can find the relationship between them. Therefore, it is advisable to use forecasting technologies to solve the problem.

Forecasting is the process of obtaining a certain value, which is the result of the work of an algorithm that is trained on a set of data prepared in advance, with new data that has not been used before. A vector of parameters of fixed length is supplied to the input, on the basis of which the output value is obtained. The very process of training this algorithm is called machine learning [5].

It is necessary to build a model that, after training on the training data set, will have the smallest value of the loss function for an arbitrary sample of input data.

The inputs are data describing the real estate object and data describing the region in which it is located. Since several objects are located in the same region, data about the region will often be repeated. Because of this, it is advisable to divide the machine learning model into two parts: the first part will receive input data about the region and return a single valid value; the second part accepts data about the real estate object together with the result of the work of the first part and as a result, returns an estimate of the value.

Regional features are mostly numerical values, so the first part of the model is a neural network. The information about the real estate object is given mainly through categorical features or through integers from a limited range, so both a neural network and a random forest method can be used here, which can perform training and evaluation faster with little difference in quality [6, 7].

Two models are implemented: a model with two neural networks and a model with a neural network and a random forest. MSE is chosen as the loss function.

A model with two neural networks

Learning algorithm

Since the model consists of two neural networks, one composite network can be created from them by connecting the output layer of the first network to the input layer of the second. The algorithm of operation and learning will not differ from the algorithm of a simple network.

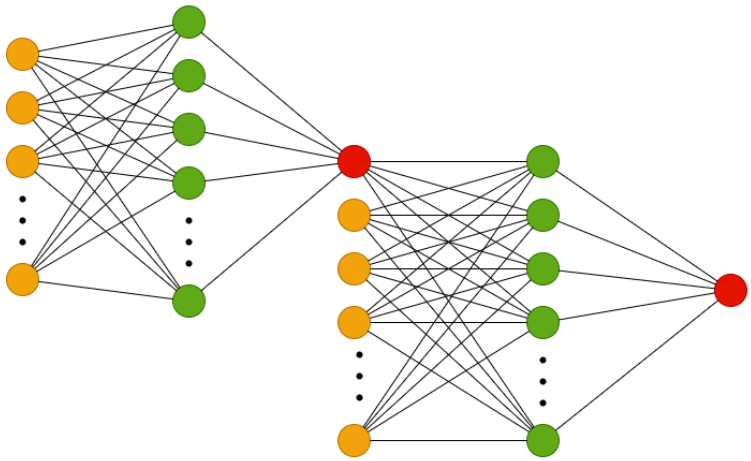

Figure 1. A model with two neural networks

Evaluation algorithm

The neural network has already been trained and the most optimal values of all coefficients have been selected. It can be used immediately after training or saved in a separate file and read each time the program is started again.

To obtain the evaluation result, data about the region is transmitted to the first network, after which direct propagation is performed and the initial value of the region is obtained, which, together with data about the housing, is transmitted to the input layer of the second neural network. Direct distribution is again performed and the cost of housing is returned.

ISSN 1560-8956

A model with a neural network and a random forest

Learning algorithm

The first part of the model, which is a neural network, is already trained and can be obtained from the first version of the model. Through this network, data about the region from the training dataset is passed and the result of its work is written into a new training dataset together with data about real estate objects.

The resulting set is then used to train decision trees. Since only one pass is sufficient for learning a decision tree and it is not necessary to repeat one algorithm many times as in the case of neural networks, this will allow training to be performed much faster.

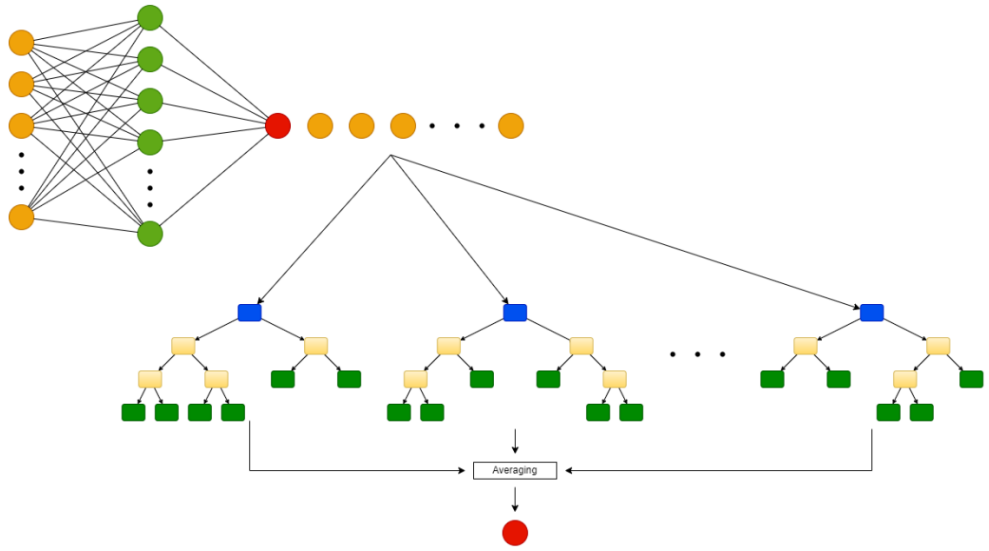

Figure 2. A model with a neural network and a random forest

Evaluation algorithm

As with the first model, the region data is fed to the first part of the model, which is the neural network. The obtained result is combined with the housing data and passed to the random forest, which consists of several decision trees. No computation is performed in trees, only a comparison of the input values with those stored in the nodes.

After performing several comparisons, we reach one of the final nodes and get the result. Since there are several decision trees, several different results will be obtained. The final result of the model will be the average value of the results of all trees. The number of calculations when evaluating using the random forest method is less than when using a neural network, so the speed of evaluation as well as learning will be higher.

Results

The proposed system is a web service. Users can use this service through a browser. This solution is the best for several reasons. The first is the presence of regional factors,

which are used to calculate the value of a real estate object, and of which almost all users are unaware. These factors will be stored in the database on the server and will be updated by the administrator. The second reason is that the real estate market changes, causing prices to fall or rise. Administrators will update the model used for scoring on the server by training it on a new data set from recent sales transactions. In addition, this solution allows to reduce of system requirements for the user, since all calculations will take place on the server.

The service has a three-level architecture, so it consists of a client part, a server, and a database server. The structure of the service is depicted in Figure 3 through a deployment diagram.

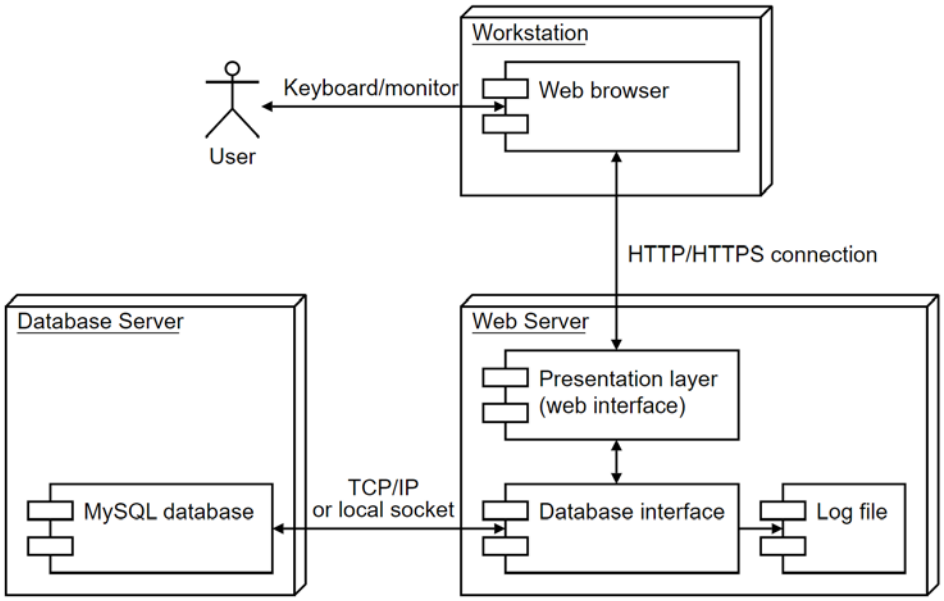

Figure 3. Deployment diagram

The Python programming language was chosen for the development of the server part. This language is the most popular for working with machine learning algorithms because it is very simple and has a large number of libraries for both mathematical operations and working with various machine learning models. In addition, libraries and frameworks for working with databases and for creating web applications have been implemented in the Python language.

The PostgreSQL database management system was chosen to create and work with a database that will store information about users, real estate owned by them, and regions. PostgreSQL is a relational DBMS and is based on the SQL language.

The client part of the service is developed using the JavaScript programming language and the React library, which is used to develop single-page web applications, that is, those that can update data without refreshing the page.

ISSN 1560-8956

Analysis of algorithms

For training, a set of data from real estate objects located on the sites: Lun.ua, Olx.ua, and Dom.ria.com was created. In order to train a model and to use it further, the input data must first be transformed into a form with which the model can work efficiently. One way to do this is through standardization. Since among the parameters of the apartment, there are those that are optional, and some mandatory parameters may have empty values in the training set, the quality of the model may deteriorate. To avoid this, standardization is performed only for non-empty values. And the empty values must be replaced with the average, thanks to which this defect will not greatly affect the effectiveness of training.

To select the best algorithm, both the proposed solutions and the conventional neural network were implemented to make sure that the proposed solutions are more effective in solving the given problem. Training and testing took place with the same data sets and, in the case of neural networks, the same sample size. The results of the implemented models are shown in Table 1.

Table 1.

№	Model	Training time, c	Number of epochs	Test data processing time, c	Test data evaluation error
1	Neural network	15.383	112	0.065	8.034
2	Two neural networks	18.097	141	0.043	5.125
3	Neural network and random forest	12.451	-	0.021	5.851

As can be seen from the table, the best efficiency is in the model consisting of two neural networks. But the model with a neural network and a random forest works much faster and learns with slightly lower accuracy. Speed is also important because the system has to handle a large number of requests at the same time. Considering this, it will be more appropriate to use the latest model.

Conclusion

The developed system is aimed at owners of residential real estate who want to sell or rent it, and buyers who want to check the fairness of the price. Therefore, in order to generate a value, users are required to provide a detailed description for an accurate result, but some parameters are optional in order to obtain an approximate value for guidance. In addition, unlike other implementations, this service offers to choose a period during which the seller wants to find a buyer.

References

1. Why do we need a Property Valuation? URL: https://bakertillyks.com/why-do-we-need-a-property-valuation/.

2. Abdallah Allabadi. Three Main Property Valuation Methods for Real Estate Investors. URL: https://www.mashvisor.com/blog/property-valuation-methods-real-estate-investors/.

3. Svitlana Tartasyuk. How to estimate the cost of an apartment online for free: step-by-step instructions. 2021. URL: https://minfin.com.ua/ua/realty/articles/kak-besplatno-ocenit-stoimost-kvartiry-onlayn/.

4. Description of the architecture and principles of operation of the "Estimator Calculator".URL:
https://www.uvecon.ua/images/tinymce/file/Robotic%20calculator%20principle.pdf

5. Iqbal H. Sarker. Machine Learning: Algorithms, Real-World Applications and Research Directions. SN Computer Science. Vol. 2. 2021. Article number: 160.

6. Oludare Isaac Abiodun, Aman Jantan, Abiodun Esther Omolara, Kemi Victoria Dada, Nachaat Mohamed, Humaira Arshad. State-of-the-art in artificial neural network applications: A survey. Heliyon. Vol. 4. Issue 11. November 2018. Article number: e00938.

Mark R. Segal. Machine Learning Benchmarks and Random Forest Regression. 2004. URL: https://escholarship.org/uc/item/35x3v9t4.

ISSN 1560-8956