Online-Journals.org

iJIM International Journal of Interactive Mobile Technologies

iJIM | eISSN: 1865-7923 | Vol. 17 No. 23 (2023) | OPEN ACCESS

https://doi.org/10.3991/ijim.v17i23.38073

PAPER

Web Application with Machine Learning for House Price Prediction

Raúl Jáuregui-Velarde[1], Laberiano Andrade-Arenas[1], Domingo Hernández Celis[2], Roberto Carlos Dávila-Morán[3], Michael Cabanillas-Carbonell[4](✉)

[1]Universidad Privada Norbert Wiener, Lima, Perú

[2]Universidad Nacional Federico Villarreal, Lima, Perú

[3]Universidad Continental, Huancayo, Perú

[4]Universidad Privada del Norte, Lima, Perú

mcabanillas@ieee.org

ABSTRACT

Every year, the price of a house changes due to different aspects, so accurately estimating the buying and selling price is a problem for real estate agencies. Therefore, the research work aims to build a Machine Learning (ML) model in Azure ML Studio and a web application to predict the buying and selling price of two types of houses: urban and rural houses, according to their characteristics, to minimize the forecast error in prediction. Following the basic stages of machine learning construction, we build the prediction model and the Rational Unified Process (RUP) methodology to build the web application. As a result, we obtained a model trained with a linear regression algorithm and a predictive ML model with a coefficient of determination of 95% and a web application that consumes the prediction model through an Application Programming Interface (API) that facilitates price prediction to customers. The quality of the prediction system was evaluated by expert judgment; they evaluated efficiency, usability, and functionality. After the calculation, they obtained an average quality of 4.88, which indicates that the quality is very high. In conclusion, the developed prediction system facilitates real estate agencies and their customers the accurate prediction of the price of urban and rural housing, minimizing accuracy errors in price prediction. Benefiting all people interested in the real estate world.

KEYWORDS

house price, linear regression, machine learning, price prediction, web application

1 INTRODUCTION

Everyone wants to buy and live in a house with features that suit their lifestyle and offer amenities that meet their needs, but predicting the price of a house is very difficult, as it is constantly changing [1]. Therefore, setting a price for the purchase and sale of a house is a process that must be analyzed in depth. The price at which the house is marketed directly affects the profitability of real estate agencies. In that sense, the price of the house plays an important role in the economy and is of great importance to the various interacting stakeholders, including homeowners, buyers,

Jáuregui-Velarde, R., Andrade-Arenas, L., Celis, D.H., Dávila-Morán, R.C., Cabanillas-Carbonell, M. (2023). Web Application with Machine Learning for House Price Prediction. *International Journal of Interactive Mobile Technologies (iJIM)*, 17(23), pp. 85–104. https://doi.org/10.3991/ijim.v17i23.38073

Article submitted 2023-01-07. Revision uploaded 2023-02-13. Final acceptance 2023-02-15.

banks, real estate developers, real estate agents, among others [1], [2]. Moreover, since price is a crucial factor in the sale of a house, it is advisable to set the right purchase and sale price from the beginning to obtain the desired results.

The price of housing increases every year, due to some potential factors such as location, area, facilities, etc. Therefore, housing price prediction is an important issue in the real estate industry and is beneficial for buyers in making decisions [3]. Consequently, there is a need for a system or mechanisms to predict house prices in the future [4], mechanisms such as ML to improve and predict house prices with high accuracy [5].

For this reason, the importance of the research is to build an ML model and a web application for the prediction of the price of a house and make it available to the customers of real estate agencies so that they have the facility to estimate the price of the house online according to the characteristics of their preference. In other words, the person who wishes to buy a house will be able to specify in the web application the characteristics or aspects such as location, type of house, size, year of construction, among others, of their preference to determine the price of the house. This will help real estate agents to determine the selling price of a house and will help the client to plan the ideal moment to buy a house.

The objective is to build an ML model in Azure ML Studio and a prototype web application to predict the buying and selling price of two types of houses: urban and rural housing, according to their characteristics. To achieve the goal, a data scientist obtains and defines the database fields for the training model and to generate a web services API. Finally, build the web application to consume the API and make it available to real estate agencies and customers.

2 LITERATURE REVIEW

In this section, the topic of machine learning and web application was addressed. For this reason, different scientific articles related to the research work were investigated, which provide ML models and results achieved.

The authors [6] claim that determining a reasonable house price involves a computational process. Therefore, to predict the price of a house, they propose advanced ML approaches. They compared two advanced ML algorithms, such as LightGBM and XGBoost, with two traditional approaches: multiple regression analysis and ridge regression. They used mean absolute error (MAE), root-mean-square error (RMSE) and adjusted r-squared value to evaluate the performance. The findings show that the XGBoost model showed the highest performance by generating the lowest MAE and RMSE, and the adjusted r-squared value closest to one. The proposed study is helpful in predicting future housing prices.

To determine the price of a house, price prediction with ML can help. Therefore, their research aims to predict house prices based on NJOP houses in Malang city using regression analysis and particle swarm optimization (PSO). To carry out variable selection, PSO is used, and to determine the optimal coefficient for prediction, regression analysis is performed [7]. The findings show that the combination of regression and PSO is suitable, as it obtained the minimum prediction error of IDR 14.186.

Moreover, accurately predicting the price of a house can help buyers and other decision makers to make the right decisions. Therefore, in their research, they propose machine learning models to predict house prices using linear regression, decision tree, XGBoost, additional trees and random forest [8]. They then evaluated using the root-mean-square error of metrics (RMSE), mean absolute error (MAE) and coefficient of determination (R2).

On the other hand, the authors [9] in their research aim to predict the price of housing for non-homeowners according to their financial resources and aspirations. To determine the price from collected data, they used techniques such as ML, Artificial Neural Network (ANN) and Chatbot. This model helps in the prediction process for buyers as well as sellers.

Predicting housing prices is beneficial for investors as well as for the real estate market. Therefore, the purpose of the research is to build a prediction model to forecast housing prices in Miami [10]. To achieve this, they used several machine learning methods and deep learning models, including SVR, linear regression, random forest, neural network, and XGBoost. However, of all the models, Random Forest and XGBoost produce better results. With these powerful learning models, they successfully forecast housing prices.

Similarly, the authors [11] state that the rapid development of the real estate industry and the growth of housing prices have attracted attention. Therefore, in their research, they propose the integration of genetic algorithm and particle swarm algorithm to optimize BP neural network to establish a prediction model. Through training prediction and simulation, the error between the predicted value and the actual value was within 0.5, thus proving the validity and accuracy of the model. Similarly, the proposed model predicts the average price of residential commercial housing in Chongqing in 2021, thereby providing a reference for the macro-control of the government and sellers to realize residential commercial housing.

They propose a machine learning techniques (MLT) model to accurately predict the price of a house. In the model, they integrated the XGBoost algorithm with the outlier sum statistics (OS) approach. In addition, the house price is predicted using the extreme gradient boosting (XG) algorithm and hedonic regression pricing [12]. For this purpose, they used 13 variables as input for both the hedonic pricing model and XGBoost. The contribution of this research lies in the practicality of using the XGboost technique to predict housing prices. Finally, the accuracy of the prediction algorithms is reported with XGBoosting, which shows a higher accuracy than the hedonic regression algorithm.

The authors [13] argue that to help the problem of predicting the price of a house, the use of machine learning systems is of vital importance. Therefore, they propose a machine learning model that uses CRISP-DM as a framework and analyzes using linear regression and random forest for the best possible accuracy. In addition to that, to find a potential market, they used K-Means as a clustering method. With the proposed model, it will be easier to predict the housing price successfully.

After analyzing the literature review of all the authors under study, it is concluded that the study conducted of ML with a web application in the real estate sector is still at a nascent level, as it has not yet been developed. The use of a sophisticated mechanism such as ML with the web application to predict the price of a house has yet to be investigated. Therefore, the research work will investigate all these deficiencies and limitations that have been found in the literature review and thus contribute with this research to the real estate sector.

3 METHODOLOGY

This section explains the stages that were carried out in the construction of the ML model. Likewise, the RUP methodology used in the development of the prototype web application for the prediction of the price of a house is explained. Furthermore, the development tools are explained; thus, achieving the objective of the study.

3.1 Machine learning model building stages

Four stages were followed in building the experimental model. These are divided into six basic steps to create, train and test the experiment model in Azure ML Studio. The steps and procedures to be followed for building the ML model are detailed in Figure 1.

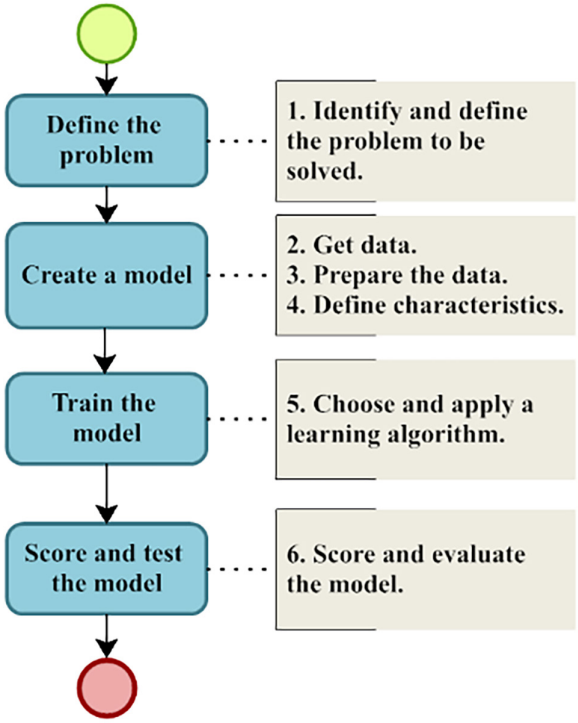

Fig. 1. Stages of machine learning model building

Define the problem: In this first stage, step 1 Identification and definition of the problem is developed. The problems to be solved with the ML model must be considered clearly and concisely. Similarly, the beneficiaries are identified.

Create model: In this second stage, step 2, data collection is followed. The historical data is obtained to feed the training model. Step 3, prepare the data. The data that will not be useful to us, such as incomplete data and unnecessary data, are purged. Step 4, define features: define and select the features suggested by the expert to predict.

Train the model: In this third stage, step 5 choose and apply, the learning algorithm is developed. The appropriate algorithm is chosen to train and test the model. Algorithms can be supervised and unsupervised. It is important to evaluate what data and data types we have available. Algorithms should be adapted to the needs of the project.

Score and test the model: In this fourth stage, step 6, scoring and evaluation of the experiment model is developed. The model is scored to see how well it works. Finally, evaluate and test the quality of the results, important to ensure the quality of the results to be obtained.

3.2 RUP methodology

It's an object-oriented software development approach that can be tailored to the specific context and needs to meet the unique requirements of each organization. It also uses UML for visual representation [14]. It's divided into four phases as shown in Figure 2, each representing a software development cycle.

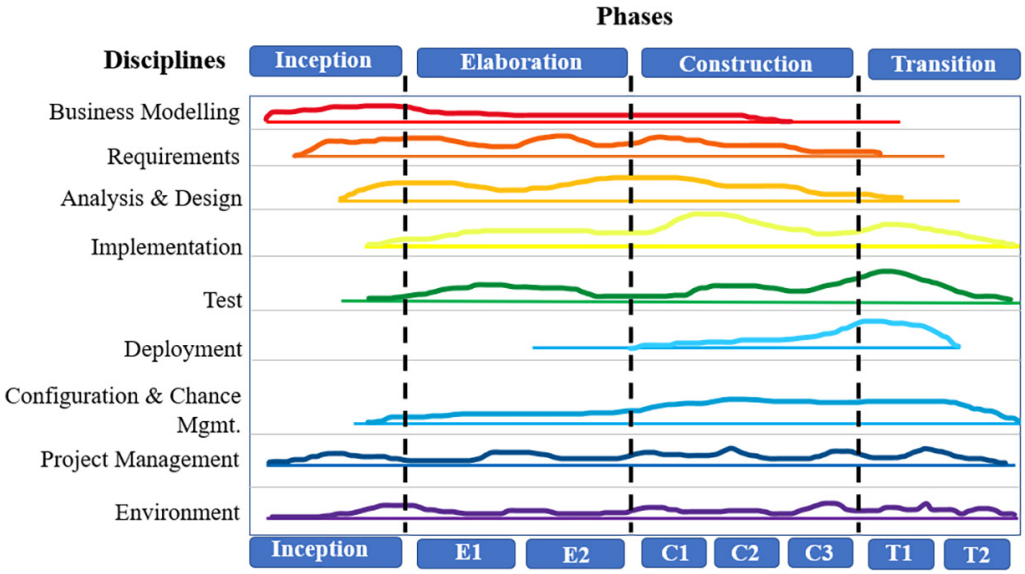

Fig. 2. Phases of the RUP methodology

Inception: In this first phase, the vision, objectives, and scope of the project are defined. A list of functional and non-functional system requirements is obtained. It focuses on business modeling. Likewise, the software architecture is detailed.

Elaboration: In this second phase, the use case analysis and the first analysis of the problem domain are completed. The specification of the selected use cases is performed. Furthermore, the base architecture of the system is defined. In addition, in this phase, an executable application that responds to the selected or defined use cases, is obtained.

Construction: This third phase begins with the construction and coding of the software's operational functionality. In addition, it successively incorporates the use cases. The implementation is ready for the testing stage.

Transition: In this last phase, tests are performed to ensure that the software is available to end users. In addition, bugs and defects found in the acceptance testing stage are adjusted. Finally, end users are released.

3.3 Development tools

Azure ML Studio: It is a Microsoft tool, an online development environment that allows you to create models in a very visual way. It is designed for developing and working with predictive analytics models. According to the authors [15], it's used for machine learning model creation and evaluation.

Figma: It's an online tool that allows you to design web and application interfaces (UX/UI). Likewise, it allows live collaboration. According to the authors [16] and [17] Figma is an application designed specifically for visual design and prototyping. A useful tool for editing victory graphics.

StarUML: This is a tool designed to create UML diagrams. In other words, it is a tool used for software modeling based on UML standards. According to the authors [18] it's an open-source software modeling tool for developing UML diagrams. It provides users with expressive modeling.

PHP: Hypertext Preprocessor (PHP) is an open-source programming language specially designed for developing dynamic web pages and applications. It is easy to embed in HTML. According to the authors [19], PHP is a dynamic server-side

scripting language especially suitable for web development. In addition, it can be used as a fast and easy development tool to create [20].

JavaScript: It's a programming language designed to give interactive functionality to web pages and applications. In addition, it allows creating dynamic content, controlling multimedia, animating images, etc. According to the authors [21], JavaScript is one of the most widely used programming languages for interactive web application development; to improve the user experience.

Bootstrap: It is an open-source CSS front-end framework designed to make clean and responsive web development easier and faster. According to the authors [22] and [23] Bootstrap is an open-source CSS framework that provides CSS classes that are already integrated with JQuery, with the ability to create a web that easily adapts to all device specifications.

HTML: Hypertext Markup Language (HTML) is a markup language designed to indicate document structure using tags. That is, it's used to specify the structure and display of a web page and its contents [24].

CSS: Cascading Style Sheets (CSS) is a rule-based language for the design and presentation of web pages. That is, to give the presentation style and user interface design of the web page. According to the author [25], cascading style sheet selectors are patterns that are used to select HTML elements. In addition, CSS technology is useful to adopt font size in a new responsive web suitable for different screen sizes, including tiny devices [26].

4 CASE STUDY

In this section, each of the steps of building the ML model and the RUP methodology for the construction of the web application for predicting the price of a house were developed. Before starting the development, we first detail the general vision of the architecture of the house price prediction system and the communication processes between the web application and the machine learning model, as shown in Figure 3, where the expert is the one who builds the prediction model. The user is the one who makes use of the model through a web application, which connects to the prediction model through an API generated by Azure.

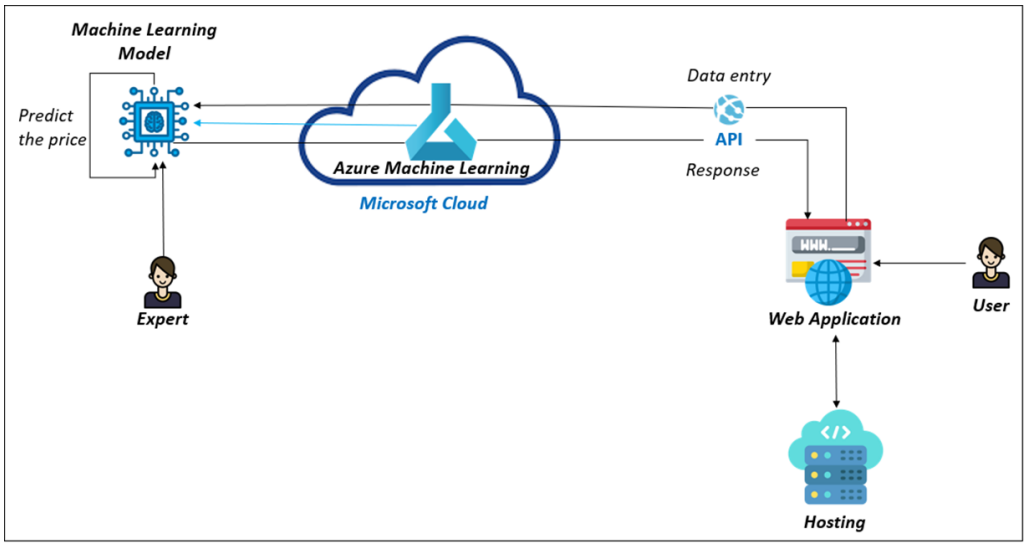

Fig. 3. Price prediction system architecture

4.1 Define the problem

There are problems in predicting the purchase and sale price of a home due to constant change. These changes affect the utility of real estate. Therefore, minimizing the accuracy error in predicting the purchase and sale prices of each type of home is a necessity for the real estate industry. Accurately predicting or estimating the selling price of a property is of great help when making decisions in the real estate sector.

4.2 Get data

The dataset for this model was obtained from historical real estate agency data. Collaborated by experienced real estate agents, the data was collected by a data scientist considering the variables (features or aspects) as shown in Table 1. The dataset was uploaded to the Azure ML Studio (classic) workspace in csv format with 2000 records.

Table 1. Data variables

Characteristics	Definition
House type	Urban or rural house
Year of construction	Year the house was built
Area	From 50 to 200 square meters
Location	The area in which a property is located
Typology	multi-family, single-family, isolated, semi-detached, semi-detached, in a block, in a block
Panoramic view	Yes: Facing spaces with nature, be they parks, gardens or, where appropriate, the sea, which allow you to enjoy and connect with nature. No: without panoramic view.
Floor	Number of floors the house has
Bedroom	Number of bedrooms in the house
Room	Number of rooms in the house
Kitchen	Kitchen area in square meters
Bathroom	Number of bathrooms in the house
Terrace	House terrace area
Chimneys	Number of chimneys in the house
Garage	Garage area of the house in square meters
Car's garage	Number of cars in garage
Pool type	Made of concrete, with a polyester hull, stainless steel, etc.
Pool area	Pool area in square meters
Garden	Yes: has a garden. No: It does not have a garden
Roof type	Concrete, asphalt shingles, rubber slate, clay tiles, etc. Metal roofing, slate.
Condition	Good, regular, repair
Price	Sale price of the house

4.3 Define the characteristics

The Select column in dataset element was added. A component that allowed us to choose a subset of columns (variables) as shown in Figure 4. This is for use in subsequent model training operations.

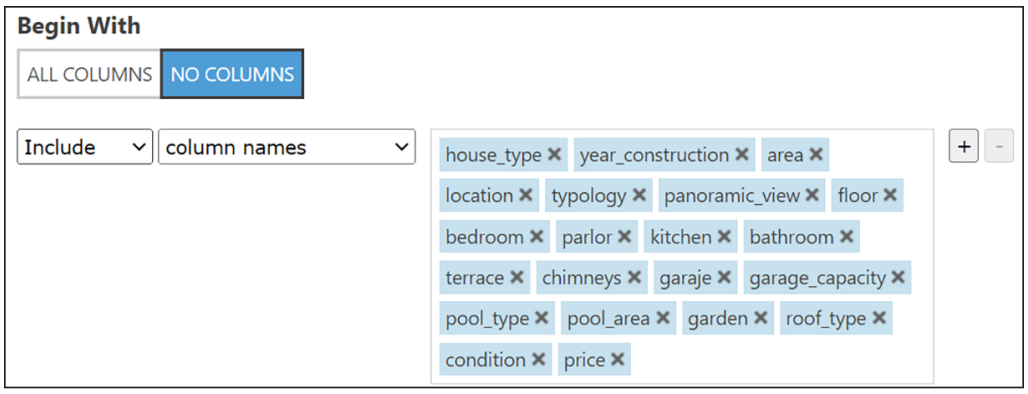

Fig. 4. Selected variables

4.4 Select and apply a learning algorithm

As shown in Figure 5, the Linear Regression algorithm was chosen to train the model, since the result that the model will give is a number. In addition, the Split data element, a component that allows you to split a data set into two different sets, one for training and one for testing, has been added. We split 70% of the data for training and 30% of the data for testing. With the Train Model element, the variable Price to be predicted was selected, and the regression model was trained.

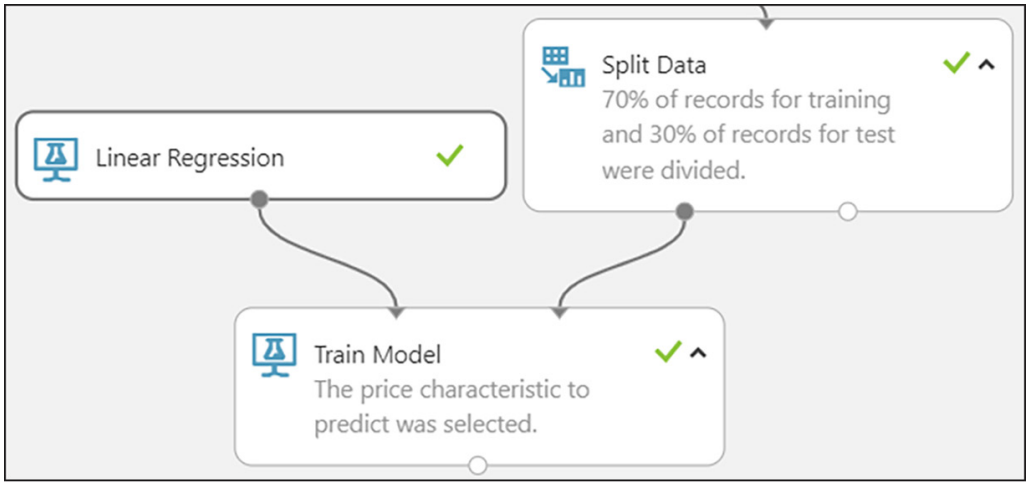

Fig. 5. Selection of the algorithm and training of the model

4.5 Scoring and evaluation

As shown in Figure 6, the item score model generates house price predictions using a trained regression model. In addition, the Evaluate Model element evaluates

the accuracy of the trained model. This evaluates how accurate the model is in delivering an accurate result. The trained model is accurate with a coefficient of determination of 95%.

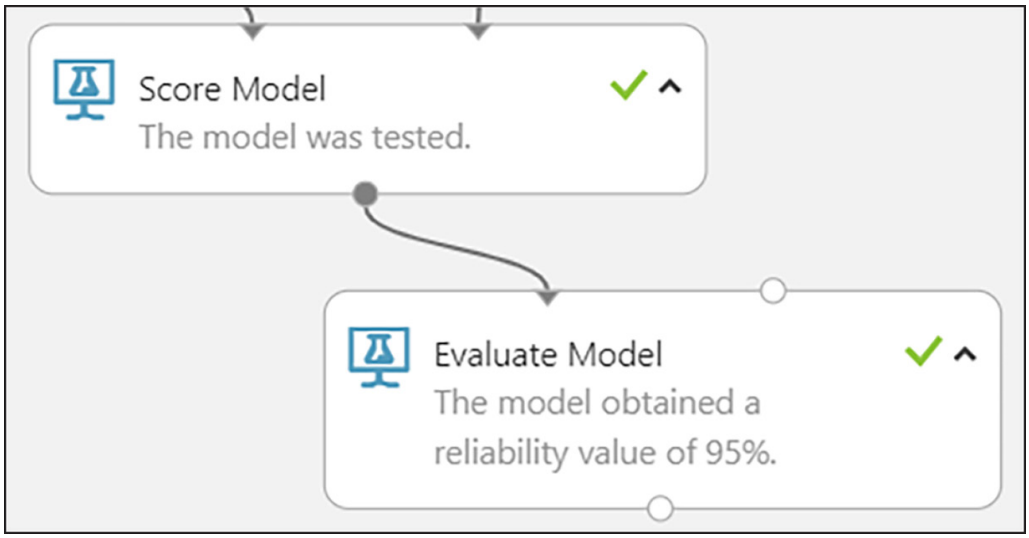

Fig. 6. Scoring and evaluation of the model

4.6 System use case diagram

Figure 7 shows the use case diagram of the system. It details the functionality of the web application to predict the price of a house. To predict the price, the client, as well as the real estate agent, must enter the data in the form, and then obtain the result according to the characteristics or aspects of their preference. Finally, the result can be printed if desired.

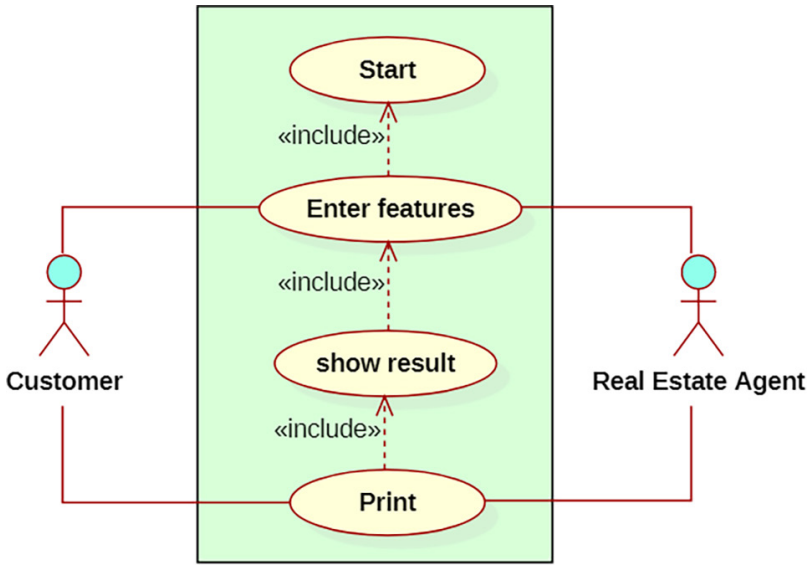

Fig. 7. Use case diagram

4.7 Web application construction

In this phase of the RUP methodology, the operational functionality of the web application as well as the user interface are coded to give a friendly and intuitive presentation to the user.

Connect to Azure ml Studio: PHP programming language is used to establish the connection of the web application with the predictive ML model. The model is located in the azure ml studio web service hosted in the cloud. The connection is made through the API generated in Azure. Likewise, to establish the connection, it's authorized through the API key, also generated by Azure ml studio as shown in Figure 8.

```php
<?php
require_once 'HTTP/Request2.php';
require_once 'HTTP/URL2.php';
require_once 'HTTP/Request2/Adapter.php';
require_once 'HTTP/Request2/SocketWrapper.php';
require_once 'HTTP/Request2/Response.php';

$request = new HTTP_Request2();

$request->setUrl('https://ussouthcentral.services.azureml.net/workspaces/b8938033f51743e4bdbebc687b32f1d5/services/c44c901802214e3a9f15db7f16beb9ed/execute?api-version=2.0&format=swagger');
$request->setMethod(HTTP_Request2::METHOD_POST);
$request->setConfig(array(
    'follow_redirects' => TRUE
));
$request->setHeader(array(
    'Content-Type' => 'application/json',
    'Authorization' => 'Bearer 1zqdrPmBBskRpm
    +gH3rnbunoXWUNOU66NOCeiT0I7A/Ny9K93MmkghyBczqQ7h
    +cBkl2ff8JnsNt+AMC0uEFpA=='
));
```

Fig. 8. Connection to machine learning model

Data input and output coding: To predict the price of the house, the data is entered through the web application form, from where each of the variables takes the corresponding value with the POST method as shown in Figure 9. To display the result, it is coded in JavaScript language. Each of the identifiers (id) of the HTML tag takes and assigns the corresponding value returned by the predictive machine learning model, as shown in Figure 10.

```php
$price = 0; //We initialize the price to zero.
$request->setBody('{
    "Inputs": {
            "input1":
            [
                {
                    "house_type": "'.$house_type=$_POST['house_type'].'",
                    "year_construction": "'.$year_construction=$_POST['year_construction'].'",
                    "area": "'.$area=$_POST['area'].'",
                    "location": "'.$location=$_POST['location'].'",
                    "typology": "'.$typology=$_POST['typology'].'",
                    "panoramic_view": "'.$panoramic_view=$_POST['panoramic_view'].'",
                    "floor": "'.$floor=$_POST['floor'].'",
                    "bedroom": "'.$bedroom=$_POST['bedroom'].'",
                    "parlor": "'.$parlor=$_POST['parlor'].'",
                    "kitchen": "'.$kitchen=$_POST['kitchen'].'",
                    "bathroom": "'.$bathroom=$_POST['bathroom'].'",
                    "terrace": "'.$terrace=$_POST['terrace'].'",
                    "garage": "'.$garage=$_POST['garage'].'",
                    "garage_capacity": "'.$garage_capacity=$_POST['garage_capacity'].'",
                    "pool_type": "'.$pool_type=$_POST['pool_type'].'",
                    "pool_area": "'.$pool_area=$_POST['pool_area'].'",
                    "garden": "'.$garden=$_POST['garden'].'",
                    "roof_type": "'.$roof_type=$_POST['roof_type'].'",
                    "condition": "'.$condition=$_POST['condition'].'",
                    "price": "'.$price.'",
                }
            ],
    },
"GlobalParameters":  {
}
}');
try {
    $response = $request->send();
    if ($response->getStatus()==200) {
        $response = $response->getBody();
        $obj = $response;
    }
    else{
        echo 'Unexpected HTTP status: ' . $response->getStatus().''.
        $response->getReasonPhrase();
    }
}
catch(HTTP_Request2_Exception $e){
    echo 'Error; ' . $e->getMessage();
}
?>
```

Fig. 9. Data input coding

```
var data = '<?php echo $obj; ?>';
var obj = jQuery.parseJSON(data);
var result = obj.Results.output1[0];

$('#house_type').val(result.house_type);
$('#year_construction').val(result.year_construction);
$('#area').val(result.area);
$('#location').val(result.location);
$('#typology').val(result.typology);
$('#panoramic_view').val(result.panoramic_view);
$('#floor').val(result.floor);
$('#bedroom').val(result.bedroom);
$('#parlor').val(result.parlor);
$('#kitchen').val(result.kitchen);
$('#bathroom').val(result.bathroom);
$('#terrace').val(result.terrace);
$('#chimneys').val(result.chimneys);
$('#garage').val(result.garage);
$('#garage_capacity').val(result.garage_capacity);
$('#pool_type').val(result.pool_type);
$('#pool_area').val(result.pool_area);
$('#garden').val(result.garden);
$('#roof_type').val(result.roof_type);
$('#condition').val(result.condition);
$('#price').val(Math.round(result['Scored Labels']));
```

Fig. 10. Data output encoding

5 RESULTS

This section shows the results of the case study. About the ML model, about the web application prototype, about the evaluation by expert judgment, and about the software development methodology used in the study.

5.1 About the ML model

Figure 11 shows the training model, a model to provide training data from which the ML algorithm learns. The model is composed of elements such as Select column in dataset, to select the (features) to train; Split data, to split the dataset into training and test data; Linear regression, algorithm with which the model is trained; Train model, to train the regression model; Score model, to generate predictions using the trained model; Evaluate model, to verify the prediction accuracy of the trained model. From the training model, the predictive model is generated, as shown in Figure 12, a model that allows predicting the housing price using the training model. In addition, a web service is generated from this model to subsequently generate an API that's used to establish the connection between the model and the web application.

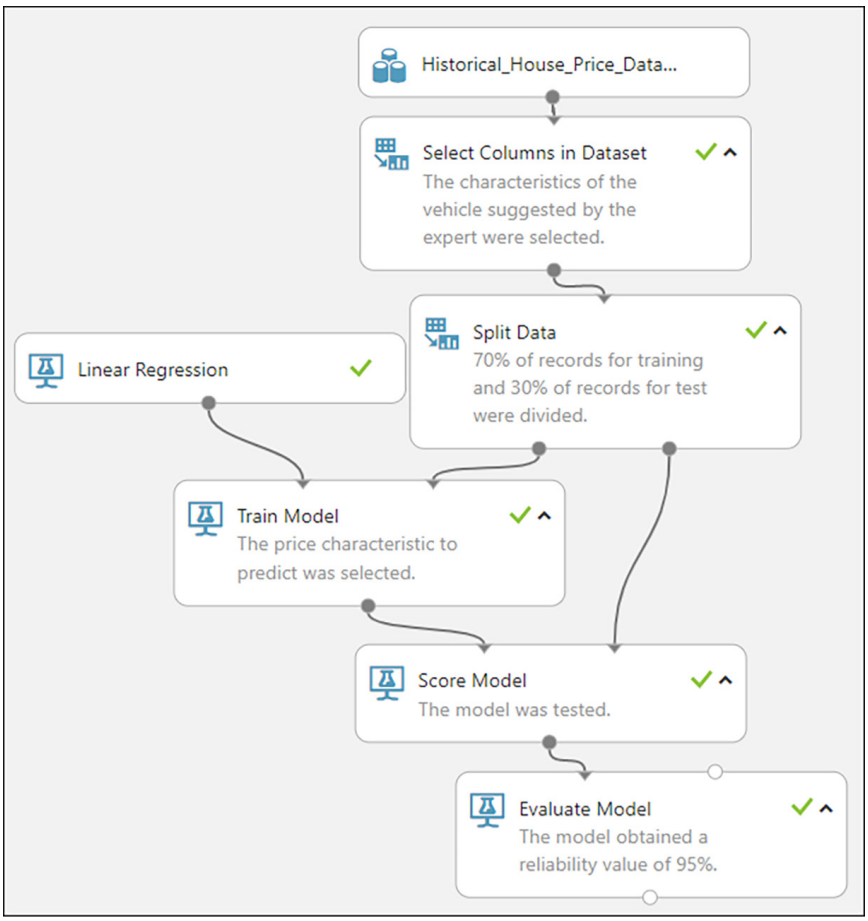

Fig. 11. Training model

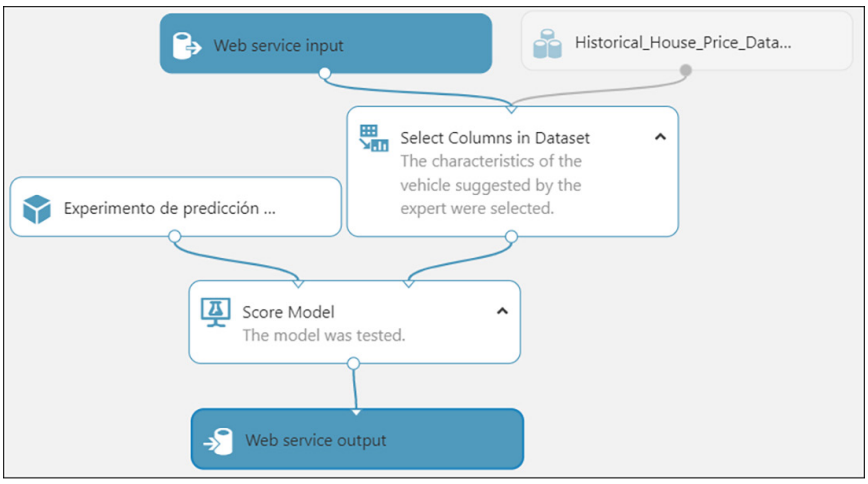

Fig. 12. Predictive model

5.2 About the prototypes

Figure 13 shows the main screen of the web application, where it details or explains what the application can do. To predict the price of the house, you must go to the form screen by pressing the “start” button or the “estimate price” button found in the navigation menu and a screen will open as shown in Figure 14, where you must

enter the data (characteristics or aspects) of the house of your choice such as type of house, year of construction, area, location, typology, panoramic view, floor number, bedroom number, living room number, kitchen area, number of bathrooms, terrace area, number of fireplaces, garage capacity, garage area, pool material type, pool area, garden, roof material type and house condition. To display the result, simply press the "Estimate" button and a model will open as shown in Figure 15, where the indicated data and the estimated price with ML according to the data indicated in the form are displayed. The user can print the result if desired with the "Print" button.

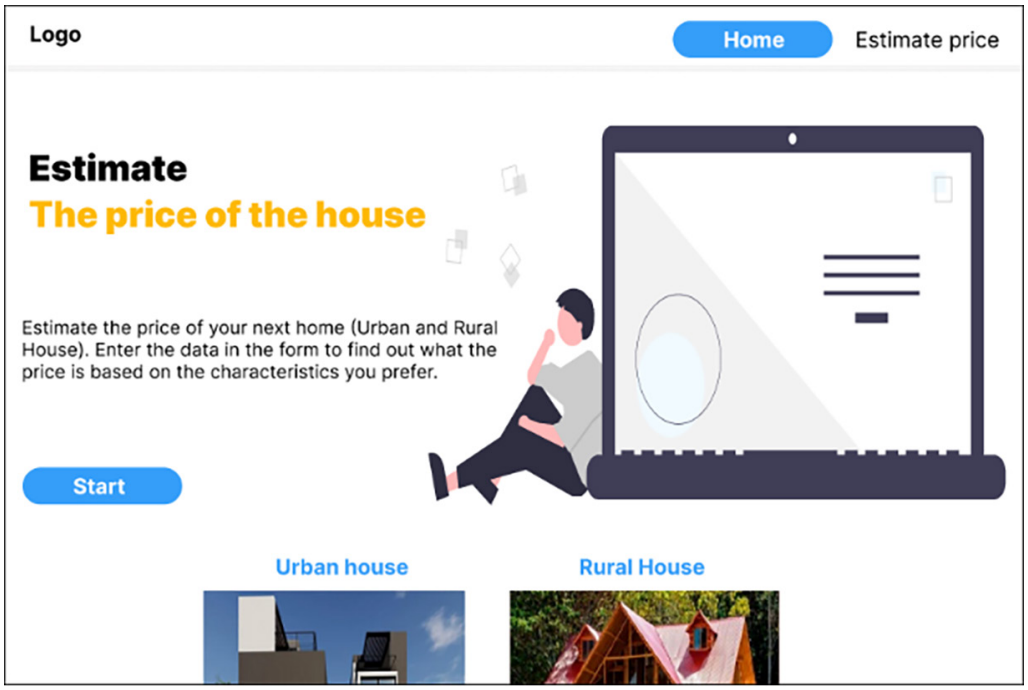

Fig. 13. Startup prototype

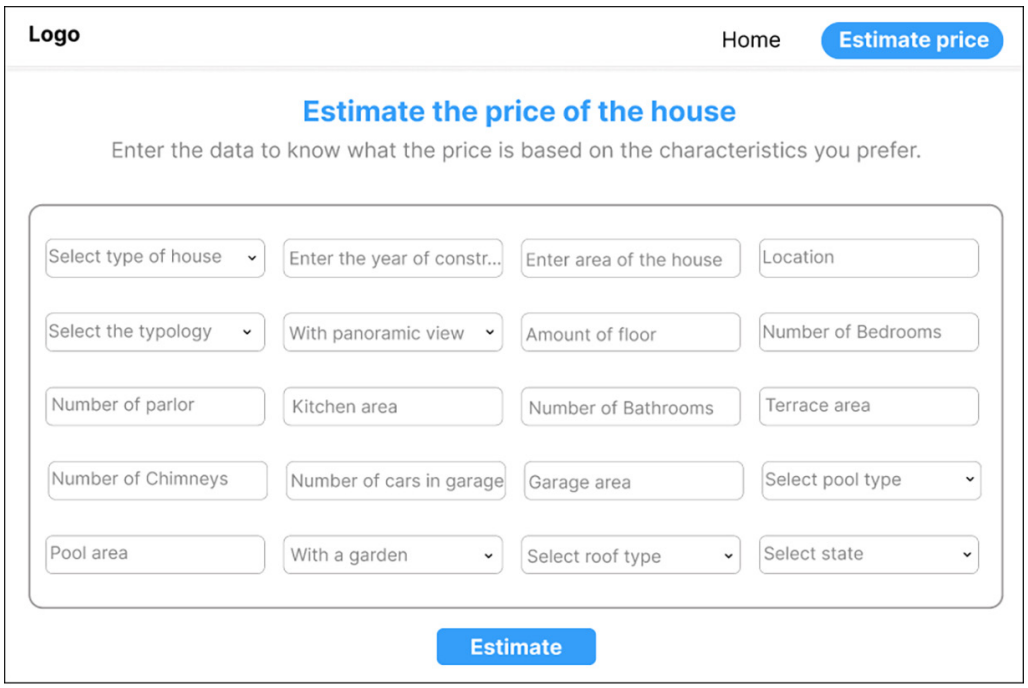

Fig. 14. Data entry prototype

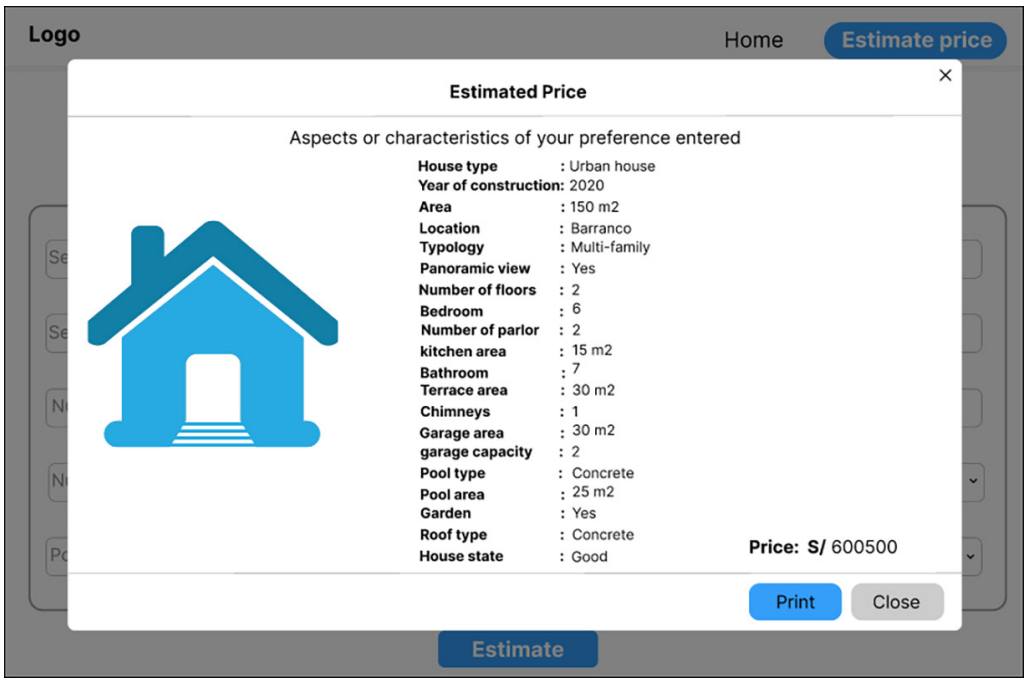

Fig. 15. Result prototype

5.3 About expert judgment

The evaluation of the ML model and the web application prototype was performed by expert judgment. Table 2 shows the results of the validation of the quality level of the predictive model and the prototype of the web application for predicting the price of a house, which was validated by ten experts in ML and the real estate industry. Criteria such as efficiency, usability, and functionality were used for the evaluation. The questions were asked based on a Likert scale including 1 very low, 2 low, 3 regular, 4 high and 5 very high. The experts' responses were analyzed using SPSS Statistics software, where the mean and standard deviation (SD) were calculated. After calculating the mean and SD of each question with the software, the quality level of the ML model and the prototype was rated according to the quality range previously established, which comprises: 0.00 to 1.00 Very low, 1.01 to 2.00 Low, 2.01 to 3.00 fairs, 3.01 to 4.00 high and 4.01 to 5.00 very high. According to the result, the total mean is 4.88. This means that the quality level of the model and prototype web application for predicting the price of a house is very high, as the total average is within the quality range.

Table 2. Evaluation by expert judgment

Criterion	Mean	SD	Quality
Efficiency			
The ML model accurately predicts the price of housing.	4.40	.516	Very high
The web application provides a fast response time.	4.80	.422	Very high
The web application shows the precise result.	4.70	.483	Very high

(Continued)

Table 2. Evaluation by expert judgment *(Continued)*

Criterion	Mean	SD	Quality
Usability			
The web application is easy to use.	5.00	.000	Very high
The web application interface is intuitive and simple.	5.00	.000	Very high
The web application is easy to access and navigate.	5.00	.000	Very high
Functionality			
The web application is readable.	5.00	.000	Very high
The web application makes it easy for the user to interact.	5.00	.000	Very high
The web application is fast loading.	5.00	.000	Very high
Average and total standard deviation of quality level	4.88	.328	Very high

5.4 About the methodology

Table 3 shows the comparison between the different methodologies such as RUP, Cascade and Rapid Application Development (RAD) to determine which of the methodologies is adequate to develop the prototype of the proposed web application. For the evaluation, the numbers from 1 to 5 were used as scores to qualify the methodology according to the evaluative condition, where the number 1 indicates that to carry the evaluative condition, the methodology is inadequate while 5 indicates that the methodology is favorable on the evaluative condition to carry out the development of the project. After making the sum of the scores of each of the methodologies, the RUP methodology has a total score of 22, the waterfall methodology has a score of 16 and the RAD methodology has a score of 20. In summary, the RUP methodology is the most suitable for the development of the web application, since it has a higher score than the rest of the methodologies.

Table 3. Comparison of methodologies

Evaluative Condition	RUP	Waterfall	RAD
Cost of risk at a single increment.	4	3	3
Adaptability to changes.	4	3	4
Documentation required.	5	4	4
Accelerate the pace of development.	4	2	5
Adaptable to customer needs.	5	4	4
Total score	22	16	20

6 DISCUSSION

In our research, a predictive machine learning model was built with Azure ml studio (classic) to predict the price of a house. Moreover, the RUP software development methodology was used to build the prototype of the web application to facilitate real estate agents and customers to predict the price according to the aspects or characteristics of the house, making use of the ML model through an API.

However, the authors [9] used techniques such as ML and Artificial Neural Network (ANN) to create the model and predict the price of the house, and to facilitate the use of the predictive model for users, they built a chatbot. In our research, linear regression algorithm was used to train the model. However, authors [8] used linear regression, decision tree, XGBoost, additional trees, and random forests to provide higher accuracy in price prediction. Authors [10] used algorithms such as linear regression, random forest, neural network and XGBoost, obtaining a highly effective predictive model in price prediction. On the other hand, authors [12] used algorithms such as extreme gradient gain (XG) and hedonic regression to obtain price predictions with quality accuracy. But all the algorithms used in the research have a common goal, which is to accurately predict the price of a house.

7 CONCLUSIONS AND FUTURE WORK

In conclusion, the machine learning model and the prototype of the web application for predicting the price of a house have been successfully developed. Thus, it will help real estate agencies and their customers who want to buy a house to determine the price with high accuracy since the coefficient of determination of the model has 95% accuracy. The stages and steps of building the machine learning and the software development methodology used, RUP, was very efficient for the development of the model, with the price prediction prototype being key to achieve the stated objective. In future research, the machine learning model could be developed with different algorithms. This is with the aim of improving the accuracy of the prediction of the price of a house. Furthermore, implement a mobile application to provide greater access and use of the ML model to users. This will achieve greater confidence in customers when buying a home.

8 REFERENCES

[1] S. Sanyal, S. Kumar Biswas, D. Das, M. Chakraborty, and B. Purkayastha, "Boston house price prediction using regression models," in *2022 2nd International Conference on Intelligent Technologies, CONIT 2022*, Hubli, India, 2022, pp. 1–6. https://doi.org/10.1109/CONIT55038.2022.9848309

[2] Q. Gao, V. Shi, C. Pettit, and H. Han, "Property valuation using machine learning algorithms on statistical areas in greater Sydney, Australia," *Land Use Policy*, vol. 123, p. 106409, 2022. https://doi.org/10.1016/j.landusepol.2022.106409

[3] W. Weng, "Research on the house price forecast based on machine learning algorithm," *BCP Business & Management*, vol. 32, pp. 134–147, 2022. https://doi.org/10.54691/bcpbm.v32i.2881

[4] V. V. B. Reddy, B. Chandana, R. S. Reddy, T. Charitha, and S. Siripragada, "House price prediction," *YMER*, vol. 21, no. 5, pp. 762–767, 2022. https://doi.org/10.37896/YMER21.05/87

[5] G. G. Priya, Ch. N. Akhil, E. S. Mahija, G. Teja, and D. RagaVamsi, "House price prediction using machine learning techniques," *International Journal for Research in Applied Science & Engineering Technology*, vol. 9, no. VI, pp. 3645–3650, 2021. https://doi.org/10.22214/ijraset.2021.35831

[6] S. Abdul-Rahman, S. Mutalib, S. Alam, M. Nor, H. Zulkifley, and M. I. Ibrahim, "Advanced machine learning algorithms for house price prediction: Case study in Kuala Lumpur," *International Journal of Advanced Computer Science and Applications*, vol. 12, no. 12, pp. 736–745, 2021. https://doi.org/10.14569/IJACSA.2021.0121291

[7] A. N. Alfiyatin, R. E. Febrita, H. Taufiq, and W. Firdaus Mahmudy, "Modeling house price prediction using regression analysis and particle swarm optimization case study: Malang, East Java, Indonesia," *International Journal of Advanced Computer Science and Applications*, vol. 8, no. 10, pp. 323–326, 2017. https://doi.org/10.14569/IJACSA.2017.081042

[8] M. Cekic, K. N. Korkmaz, H. Mukus, A. A. Hameed, A. Jamil, and F. Soleimani, "Artificial intelligence approach for modeling house price prediction," in *2022 2nd International Conference on Computing and Machine Intelligence, ICMI 2022*, Istanbul, Turkey, 2022, pp. 1–5. https://doi.org/10.1109/ICMI55296.2022.9873784

[9] C. Chee Kin, Z. Arabee Bin Abdul Salam, and K. Batcha Nowshath, "Machine learning based house price prediction model," in *2022 International Conference on Edge Computing and Applications (ICECAA)*, Tamilnadu, India, 2022, pp. 1423–1426. https://doi.org/10.1109/ICECAA55415.2022.9936336

[10] X. Wu and B. Yang, "Ensemble learning based models for house price prediction, case study: Miami, U.S.," in *Proceedings – 2022 5th International Conference on Advanced Electronic Materials, Computers and Software Engineering, AEMCSE 2022*, Wuhan, China, 2022, pp. 449–458. https://doi.org/10.1109/AEMCSE55572.2022.00095

[11] Z. Sun and J. Zhang, "Research on prediction of housing prices based on GA-PSO-BP neural network model: Evidence from Chongqing, China," *International Journal of Foundations of Computer Science*, vol. 33, nos. 6–7, pp. 805–818, 2022. https://doi.org/10.1142/S0129054122420163

[12] J. Zaki, A. Nayyar, S. Dalal, and Z. H. Ali, "House price prediction using hedonic pricing model and machine learning techniques," *Concurrency and Computation*, vol. 34, no. 27, p. e7342, 2022. https://doi.org/10.1002/cpe.7342

[13] J. Wijaya, H. P. Ipung, and M. A. Soetomo, "Design an agile of machine learning to predictive house pricing and targeting segmented market," in *ACM International Conference Proceeding Series*, 2022. https://doi.org/10.1145/3557738.3557856

[14] A. Nurliyani, Rachmansyah, and Yoannita, "Designing applications for dictionary of Indonesian law based on Android," *Journal of Information Technology Education: Research*, vol. 3, no. 1, 2019. https://doi.org/10.31289/jite.v3i1.2624

[15] M. Owais Raza, N. Pathan, A. Umar, and R. Bux, "Activity recognition and creation of web service for activity recognition using mobile sensor data using azure machine learning studio," *Review of Computer Engineering Research*, vol. 8, no. 1, pp. 1–7, 2021. https://doi.org/10.18488/journal.76.2021.81.1.7

[16] D. Dwi Aulia, S. Aminah, and D. Sundari, "Perancangan Prototype Tampilan Antarmuka Berbasis Web Mobile Pada Toko Amira Kosmetik," vol. 5, no. 1, pp. 2621–4962, 2022. https://doi.org/10.47324/ilkominfo.v5i1.134

[17] O. D. Alao, E. A. Priscilla, R. C. Amanze, S. O. Kuyoro, and A. O. Adebayo, "User-centered/user experience Uc/Ux design thinking approach for designing a university information management system," *Ingenierie des Systemes d'Information*, vol. 27, no. 4, pp. 577–590, 2022. https://doi.org/10.18280/isi.270407

[18] S. Suriya, R. Asmitha, V. G. Darshanaa, S. Priyadarshini, K. Priyanga, and J. R. Sanjeetha, "Design of UML diagrams for intervention for autism children (IAC System)," *Communications in Computer and Information Science*, vol. 1393, pp. 13–28, 2021. https://doi.org/10.1007/978-981-16-3660-8_2

[19] A. H. Odeh, "Analytical and comparison study of main web programming languages-ASP and PHP," *TEM Journal*, vol. 8, no. 4, pp. 1517–1522, 2019. https://doi.org/10.18421/TEM84-58

[20] M. A. A. Hammoudeh and A. S. Al-Ajlan, "Implementing web services using PHP soap approach," *International Journal of Interactive Mobile Technologies*, vol. 14, no. 10, pp. 35–45, 2020. https://doi.org/10.3991/ijim.v14i10.14391

[21] F. D. A. Farzat, M. D. O. Barros, and G. H. Travassos, "Evolving JavaScript code to reduce load time," *IEEE Transactions on Software Engineering*, vol. 47, no. 8, pp. 1544–1558, 2021. https://doi.org/10.1109/TSE.2019.2928293

[22] R. Kaban and F. Fajrillah, "Pengembangan sistem informasi perpustakaan dengan framework CSS bootstrap dan web development life cycle," *Jurnal Ilmiah Informatika*, vol. 2, no. 1, pp. 83–89, 2017. https://doi.org/10.35316/jimi.v2i1.454

[23] R. Kaban, M. H. Simbolon, and A. Abdullah, "Aplikasi e-archiving dan monitoring Surat Menyurat," *MEANS (Media Informasi Analisa dan Sistem)*, vol. 4, no. 2, pp. 112–119, 2019. https://doi.org/10.54367/means.v4i2.544

[24] T. Wongkhamdi, N. Cooharojananone, and J. Khlaisang, "E-commerce competence assessment mobile application development for SMEs in Thailand," *International Journal of Interactive Mobile Technologies*, vol. 14, no. 11, pp. 48–75, 2020. https://doi.org/10.3991/ijim.v14i11.11358

[25] E. Uzun, "A regular expression generator based on CSS selectors for efficient extraction from HTML pages," *Turkish Journal of Electrical Engineering and Computer Sciences*, vol. 28, no. 6, pp. 3389–3401, 2020. https://doi.org/10.3906/elk-2004-67

[26] D. Bratić, N. S. Loknar, and T. K. Ivančević, "Fuzzy logic-based evaluation model of handwritten font sizes readability on extra small devices," *International Journal of Interactive Mobile Technologies*, vol. 16, no. 9, pp. 93–106, 2022. https://doi.org/10.3991/ijim.v16i09.31135

9 AUTHORS

Raúl Jáuregui-Velarde is a student at the Faculty of Engineering and Business at the Norbert Wiener Private University. Professional Academic School, Systems Engineering and Informatics. He focused on expert systems, data science, networks and communications, information security, software development and database management (E-mail: jaureguivraul@gmail.com).

Laberiano Andrade-Arenas, Doctor in Systems and Computer Engineering (UNFV). Master in Systems Engineering (UNFV). Graduated from the Master's Degree in University Teaching (UNE). Graduated from the Master's degree in accreditation and evaluation of educational quality (UNE). Systems Engineer (UTP). ITILV3 Fundamentals International Course (Zonngo – Peru / IMLAD – Mexico). Scrum Fundamentals Certified, Research Professor with publications in SCOPUS indexed journals. He has extensive experience in the University Chair in face-to-face and blended classes at different undergraduate and postgraduate universities in Lima (E-mail: landradearenas@gmail.com).

Domingo Hernández Celis is a member of the International Association of Online Engineering (IAOE), Doctor of Accounting; Doctor of Economics; Doctor of Administration; Master in Accounting and Financial Auditing; Certified Public Accountant; Independent Auditor, General Manager of Microconsult-DHC Associates, normal, remote and virtual undergraduate teacher; master's teacher; doctoral professor; Financial Advisor. In research he is a teacher, advisor, reviewer, jury. More than 30 years of professional practice and more than 20 years in teaching work. Teaching experience at: Federico Villarreal National University; University of San Martin de Porres (E-mail: dhernandez@unfv.edu.pe).

Roberto Carlos Dávila-Morán is research professor with various publications indexed in databases such as Scopus, Web Of Science and Scielo (E-mail: rdavila430@gmail.com).

Michael Cabanillas-Carbonell is Systems Engineer. A qualified researcher by the National Council of Science, Technology, and Innovation – Peru. Research professor at the Universidad Privada Norbert Wiener and Universidad Privada del Norte. President of the IEEE-Peru Education Society Chapter. President of the EIRCON Conference (Engineering International Research Conference) (E-mail: mcabanillas@ieee.org).