Systematic Approach to Housing Price Prediction And Recommendations-Case Study Based on Sri Lankan Housing Prices

Sulochana Ekanayake
Department of Information Technology
Faculty of Computing
General Sir John Kotelawala Defence University,Ratmalane,Sri Lanka.
emskekanayake96@gmail.com

Dushyanthi Vidanagama
Department of Information Technology
Faculty of Computing
General John Kotelawala Defence University
Ratmalane, Sri Lanka
udeshika@kdu.ac.lk

Abstract—**Machine learning is a technology that is currently being popular. It can be used to increase effectiveness in every field. Real estate has had a big impact on all aspects of our society. As today's society relies more and more on technology, its dependability for precise prediction and application recommendations has become more important. Most people prefer to purchase a home rather than build one. By manually calculating the influencing parameters necessary for estimating the rate of property, it is challenging to determine the price of a property. Customers who rely on real estate agents are taken advantage of because agents may quote prices that are significantly higher than the going rate. People with a budget for a home purchase are unable to purchase due to the disparity in prices their agent provided. The project's goal is to improve communication between buyers and sellers. As the Main improvement of the project, the solution was given to the better Searching Options. After reviewing various ML models, the Random Forest Regression model was chosen to train the data set. The Housing Price Prediction and Recommendation System were developed using a dataset of Sri Lankan real estate prices. RPA was used to extract the data set. Future home price predictions were made using the ARIMA model. Housing suggestions were generated using Hybrid recommendations. Collaborative filtering and Content-based filtering have been identified as major recommendation methods. In our proposed system hybrid recommendation has been used to increase the accuracy. The buying and selling platforms have integrated prediction, forecasting, and recommendation features together. This system will aid in the development of trust between buyers and Sellers**

Keywords— Housing Price Prediction, Forecasting, ARIMA Model, Random Forest Regression, Hybrid Recommendations

I. INTRODUCTION

As is well known, in addition to necessities like food and water, a person's basic needs also include a place to call home. Many people now use their homes as an investment or pieces of real estate as living standards have improved. On the other hand, some people all over the world are buying homes. People frequently use the internet to look up news stories and other pertinent information when they are first thinking about buying a house or other piece of real estate. People do this to look for a house with all the amenities they need. The cost of these homes is taken into consideration by those who are doing this. The average person, however, lacks in-depth knowledge and precise information about what the real price should be. As a result, they may receive inaccurate information because they think the prices listed online are real. Contacting various estate agents is the second thought

that comes to mind when looking for a property. The issue with this is that these agents should only be paid a small amount for looking at houses and determining a price for you. Most of the time, people blindly accept this price because they have no other choice. In some instances, agents and sellers may have a covert agreement, and the buyer may be unknowingly sold home at an inflated price.

Considering house sellers Contacting various estate agents is the second thought that comes to mind when looking for a property. The issue with this is that these agents should only be paid a small amount for looking at houses and determining a price for you. Most of the time, people blindly accept this price because they have no other choice. In some instances, agents and sellers may have a covert agreement, and the buyer may be unknowingly sold home at an inflated price.[1], [2].

Currently, Sri Lanka forecasts housing prices using a manual valuation process. The entire process, however, usually takes two to three weeks, and the finished product is subject to many variations for several reasons. The government rates are significantly different from the initial market rates, which is the main justification. Given the rapid expansion of housing projects and schemes, Sri Lanka urgently needs a system for predicting and forecasting housing prices. It is especially crucial to predict future property prices accurately. As a result, county-level economic conditions are reflected in real estate prices, which are of great interest to both buyers and sellers.

By determining the preferences of users and offering tailored recommendations that meet their needs, a recommender system (RS) can help users and online platforms. RSs are intelligent models that use data mining and machine learning techniques to personalize user experiences by learning from users' past interactions with the system. RSs are widely used by the real estate market, online retailers, booking services, music and movie streaming services, and online merchants. One of a person's fundamental needs is to be accommodated. People rarely buy or rent properties throughout their lifetime, which makes the housing decision a particularly difficult one to make. Users consider several criteria and decision factors when looking for the desired property. These criteria include things like location, cost, and property size, which are important considerations for users. Others are minor elements that may have less of an impact on users' decisions, such as particular amenities, and closeness to markets, schools, and public transportation. Since users can specify their

preferences and find appropriate items that best match their criteria among many other irrelevant properties, there is currently growing interested in using online platforms to search and find real estate items (properties)RSs assist a real estate platform user by making tailored recommendations based on the user's profile. Different ways exist for users to interact with these platforms. Typically, they formally state their requirements by establishing certain search parameters, such as location, cost, and the number of bedrooms. They can also interact with some of the properties on the website by clicking on the links to the individual properties, bookmarking them, requesting more information, or scheduling a visit. An RS can better infer user preferences and make more pertinent recommendations with the aid of all these pulses.[3], [4].

Considering these problems, the main objectives of this research paper is
• Provide a successful conclusion for housing price forecasting and thoroughly validate several regression modeling implementation techniques.
• create a website to implement a machine learning model that is effective and can calculate the price of a house accurately based on the features provided.
• Assess the effectiveness of price prediction algorithms that use machine learning.
• Housing Price Predictions
• User-based house listings and recommendations
So, our proposed niwahana.lk is an integrated buying and selling platform for Sri Lanka with predicting, Forecasting, and recommendation feature.

II. LITERATURE REVIEW

House Price Prediction

The methodology for creating real estate price indices presented in this article is model-independent. The development of this methodology aims to extend the library of algorithms used to construct price indices to non-linear, non-parametric models such as Machine Learning (ML). The main innovation is the use of individual out-of-time prediction errors to assess price changes.[5]
In this study, several regression models have been created to predict a house's price based on a variety of its attributes. They compared and evaluated each model to ascertain which one performed the best. They also looked at how some models ranked the significance of the features. In this paper, the data science methodology was used, starting with data collection, followed by cleaning and preprocessing, data exploration, model building, results from the evaluation, and result in visualization.[6]
The House Price Index (HPI) is frequently used to gauge changes in housing costs. To predict a person's housing price, information other than the HPI is needed because housing price is strongly correlated with other factors like location, area, and population. There have been a lot of papers that use traditional machine learning techniques to predict housing prices accurately, but they rarely pay attention to how well each model performs individually and ignore the less well-known but more complex models. This paper will use both conventional and advanced machine learning approaches to investigate the differences between several advanced models to explore various impacts of features on prediction methods.

This study will provide a positive outcome for housing price prediction as well as thoroughly validate numerous modeling methodologies.[7]
According to this paper, machine learning can make predictions as accurately as reputable institutions. Elastic Net produces the most accurate quarterly predictions out of all the machine learning models. Elastic Net's predictions are more accurate than Norges Bank's in 29,6% of the quarters, but less accurate overall. Significant deviations during 2018 and 2019. following the implementation of new mortgage regulations from Finanstilsynet are what account for the poor performance. Norges Bank outperforms Random Forest, which makes the most accurate yearly predictions. Nevertheless, Random Forest outperforms both DNB and SSB in every evaluation step.[8]
This paper demonstrates the use of machine learning algorithms in the prediction of real estate/house prices. This research will be beneficial, to find the most important attributes to decide house values, especially for housing developers and academics and to recognize the most effective machine learning model for conducting research in this field.[9]

B. House Price forecasting

An exercise in forecasting was done using six different models. These include the Random Acceleration (RA), ARMA-Generalized Autoregressive Conditional Heteroskedastic (GARCH), Self-Exciting Threshold Autoregressive (SETAR), Smooth Transition Autoregressive (STAR), and Model with Structural Breaks in the Trend Function. These models were chosen after considering models that other researchers had already used to analyze house prices, as well as after including nonlinear dependencies and structural breaks in the analysis. [10]

C. House Recommendations

The most popular and widely used recommendation techniques in recommender systems today are the three general techniques listed below: content-based filtering, collaborative filtering, and case-based filtering. Depending on the components that are applied and the stages of action, each method has distinct effectiveness and accuracy regarding recommendations.[11]

Collaborative filtering

To generate new recommendations, collaborative methods for recommender systems rely solely on the historical interactions between users and items that have been recorded. The fundamental tenet of collaborative methods is those past user-item interactions are sufficient to identify similar users and/or similar items, and predictions based on these estimated proximities, are the mainstay of these methods. [12]

Content-based Filtering

A content-based recommender uses information that the user provides, whether consciously (like rating something) or unconsciously (like clicking a link). This information is used to create a user profile, which is then used to provide suggestions to the user.[13]

D. Novelty of the Proposed System

According to the literature, more papers were discussed about Machine learning algorithms for predicting house prices. But few were discussed about price forecasting features. It was difficult to find research papers related to house recommendations. We can find real estate salespeople or websites in Sri Lanka (Lankaproperty.lk, House.lk, Saleme.lk), but there is no system in place to predict the prices of those homes through automation. There are only investment analysis professionals, and they only offer information about investments. So, our proposed system is an result, a computer program for forecasting housing prices will be in demand across the country.

integrated web system with these three features in recent years, Sri Lanka's economy has paid a lot of attention to the issue of housing price forecasting. Because it provides many employment opportunities, it has a significant trend in both academia and the economy. Sri Lanka's social and economic statistics can be significantly impacted by a minor error in the housing market. In Sri Lanka, factors that affect housing price forecasting have evolved. It takes a long time to manually insert those changes for the prediction, though. These elements cast doubt on the reliability of the findings.

Table I. System Features of the previous Research Papers

Research Paper	Technology	House Price Prediction	House Price Forecasting	House Recommendations	Buying and Selling Platform
[14]]	Extreme Gradient Boosting, Gradient Boosting Regression, Random Forest regression, Light Gradient Boosting Machine regression, support vector regression	X	✓	X	X
[13], [15]]	Hybrid recommendation	X	X	✓	X
[3]	Collaborative filtering	X	X	✓	X
[16]	Random forests	✓	X	X	X
[17]	GIS Technology Machine learning Algorithms	✓	X	X	X
[6]	ML techniques Random forest SVM Linear Regression	✓	X	X	X
[7]	Random Forest, XGBoost, and LightGBM and two techniques in machine learning including hybrid Regression and Stacked Generalization Regressionlgorithms	✓	X	X	X
[8]	Elastic Net, Random Forest, and Neural Network	✓	X	X	X
[18]	CNN Model	✓	✓		✓

Table II. Features of Existing House Selling Platforms in Sri Lanka

Existing Systems in Sri Lanka	Features
Lankapropertyweb.com	• E-commerce website. • Buying and selling Properties. • Post Advertisements and view Advertisements are enabled. • Search by keywords. • (Bedrooms/Bathrooms/Price range) • Contact individual Buyers through Email.
Saleme	• Buy, sell, rent anything, and post Advertisements. • Not specified for houses.
Ikman.lk	• Sri Lanka's household name for buying and selling anything to everything online. • Buy and sell anything. • Eble to contact advertiser through call or Email directly.
House.lk	• Sell or Buy Any type of house.
Homelands	• House selling Agency. • Individual sellers cannot post Advertisements. • Buyer Should contact Agency directly.

III.   METHODOLOGY

A.   Design
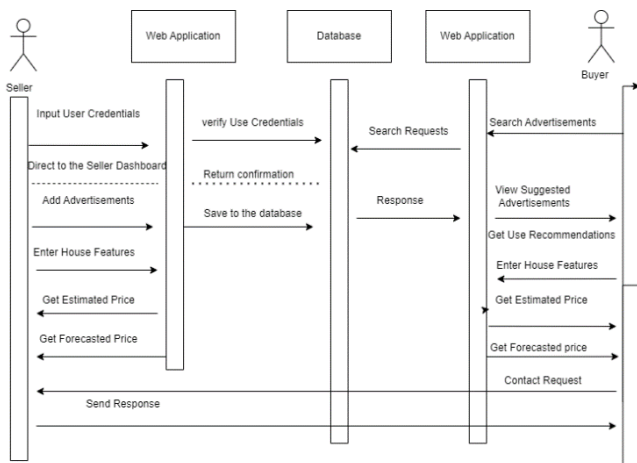
Figure 1. Sequence Diagram of the Proposed System
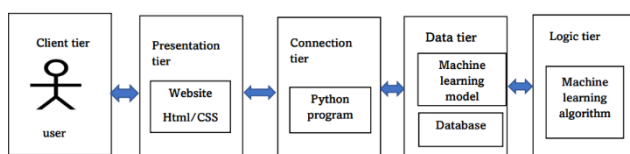
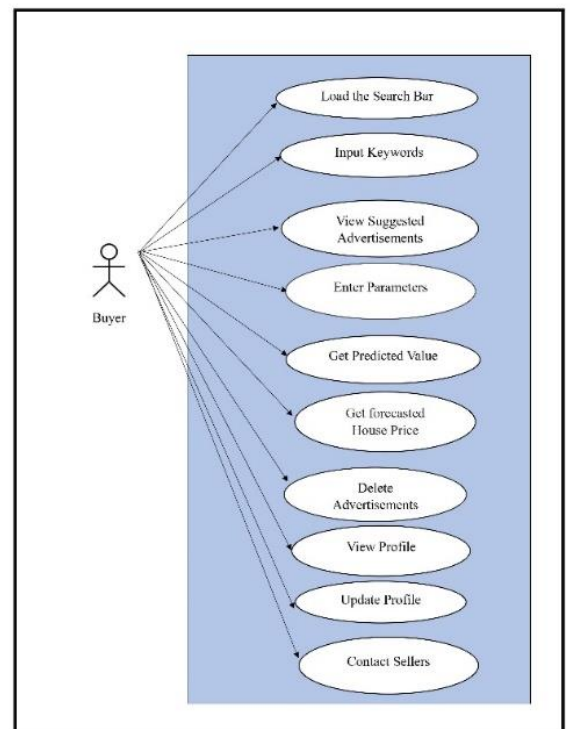Figure 2. Overall Architecture of the System

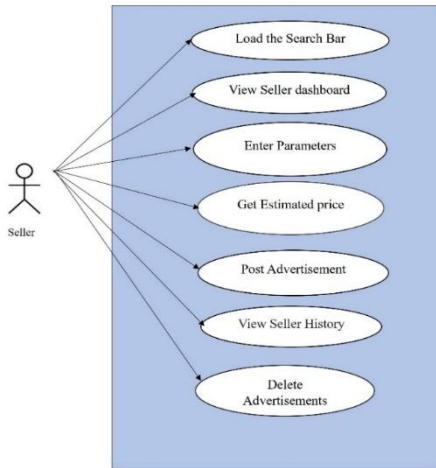Figure 3. Use case Diagram of the Buyer Profile

Figure 4. Use case Diagram for Seller Profile

B. Implementation

Data sets were extracted from the online house-selling platform house.lk using robotic process Automation. It includes the house selling details from 2018 to 2022.

Figure 5. Data Set

Correlation reveals the association between the means of two attributes. Any statistical relationship can be described by correlation, but it is most frequently used to describe how strongly two variables are correlated linearly. The data set's attribute correlation is depicted in the diagram below.
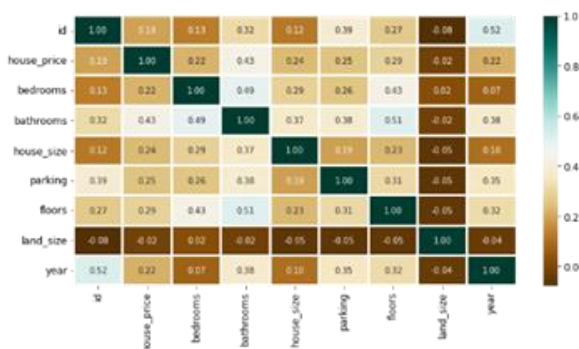

Figure 6. Correlation Matrix

To predict housing prices a few of Regression Models were tried. To find the model that best fits our data, particularly the test datasets, we trained several ML models. Starting with the most fundamental one, Linear Regression. After Comparing Linear Regression, SVM, Decision Tree and Random Forest Random Forest were selected for our proposed System. Building multiple trees using a random subset of features and averaging their predictions is how random forest works.

House Price Forecasting

ARIMA (Autoregressive Integrated Moving Average) only requires one variable [5]. ARIMA models use real values and lagged data to generate the residuals, or values that represent the gap between reality and expectations. When working with this kind of data, creating a time series object is the best way to get univariate data and R to cooperate. The ARIMA model requires complex statistical calculations. The p, d, and q values are related to them. It is typically written as ARIMA. The entered date will serve as the user-input for the ARIMA model, which analyzes the autoregressive component of the model and typically analyzes the order in which the linear component of the model integrates with earlier points.

House Recommendations

- Collaborative filtering: Collaborative filtering methods create a model from past user behavior, such as the items that users have searched for or purchased, as well as similar choices that other users have made.
- Content-based filtering: This approach uses several discrete characteristics of an item to suggest additional items with related characteristics.

To provide the best Recommendations hybrid recommendation were used.

The system must quickly collect and analyze visitor-specific data during the first session to be useful. The developer team created an algorithm that instantly creates a user profile, enabling property recommendations to be customized even for a single visit immediately following the initial interactions. The engine employs a content-based filtering technique, which means that it compares information from user profiles with attributes for each property profile.

Content-based Filtering and Collaborative Filtering were taken into consideration. hybrid Recommendations were chosen for this system's recommendation function from among these recommendation techniques. It was challenging to anticipate what the user would want to see next when using traditional approaches (collaborative filtering, content-based filtering).

IV. RESULTS AND DISCUSSION

Evaluation

To find the best model to train the dataset Linear regressionism, Decision tree, and Random Forest were compared. Least RMSE can be identified as Random Forest. As well as best accuracy is also displayed in Random forests.

Mean Absolute Percentage Error (MAPE) can be used to calculate the loss. Considering the RMSE.MAPE and Accuracy Random Forest were identified as the best-fitted Models.

To Evaluate the performance of collaborative filtering and Content-based filtering following two measurements are used.

How many of our top k recommendations are visited by the user, according to precision @ k (in the test set)?

Recall @ k: How many of them with eld listings that the user looked through (test set) were correctly suggested using the training data?

Compared to content-based filtering collaborative filtering had a high recall and precision value. But Using Hybrid mode increase the accuracy of the best viewing advertisements. So, hybrid regression was used for recommendation purpose in our system

System Functions

According To input, keyword advertisements can be searched by everyone. Without searching using the hybrid filtering method advertisements are displayed on some pages according to the preferences of individual users.

When the user (Maybe a Buyer/Seller) enters the home page Advertisements can be viewed without logging in to the system. The filter by keyword option is displayed in the search bar.
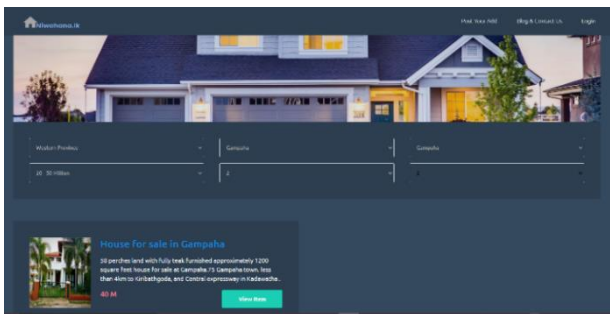

Figure 7 Search by keyword

After login in with Buyer credentials and entering by preferred house features, buyers can predict and forecast the price for the preferred house.

Figure 8 System Interface for Buyer Profile

After login into the seller account sellers can predict and forecast prices for their selling house and Add Advertisement option is Enabled in the interface. When posting an Advertisement both the selling price of the house and the predicted price by the system will be displayed near the advertisements. Users can view both the selling price and the predicted price according to the Advertisements.

Figure 9 System Interface for seller Profile

V. CONCLUSION

Machine learning Model comparison is done by starting from simple machine learning regression Algorithms to Advanced Machine Learning Algorithms. Random forests Regression has been identified as the most accurate model. By introducing new features or by creating new features, this model can be enhanced and expanded with more features to increase accuracy. This model can be used to forecast home values anywhere in the world by simply changing the features and parameters. Although there are many house-selling platforms in Sri Lanka I couldn't find a website to estimate a selling and buying price for a house. Forecasting the Future Price of the house is a very important factor for both buyers and Sellers. The ARIMA Model can be used to forecast future prices of the house and saving time and viewing a proper Advertisement for a Buyer can Enhance the Efficiency of the Web System. Collaborative filtering and Content-based can be used for user recommendations. The hybrid recommendation can be identified as the most suitable method for user recommendations. This proposed system will help both sellers and buyers.

REFERENCES

[1] Chawda, and M. Techscholar, "Issue 6 www.jetir.org (ISSN-2349-5162)," JETIR, 2021. [Online]. Available: www.jetir.orgc710

[2] P. Lawhale, Y. Ramatkar, S. Dakhore, M. Gumble, A. Tawlare, and Y. A. Thakare, "House Price Prediction using Machine Learning," International Journal of Advanced Research in Science, Communication, and Technology (IJARSCT, vol. 2, no. 7, 2022, doi: 10.48175/IJARSCT-4404.

[3] A. Gharahighehi, K. Pliakos, and C. Vens, "Recommender systems in the real estate market—a survey," Applied Sciences (Switzerland), vol. 11, no. 16. MDPI AG, Aug. 02, 2021. doi: 10.3390/app11167502.

[4] K. Kumarasinghe, "AN IMPROVED JOB RECOMMENDATION WEB APPLICATION," 2021.

[5] F. D. Calainho, A. M. van de Minne, and M. K. Francke, "A Machine Learning Approach to Price Indices: Applications in Commercial Real Estate," Journal of Real Estate Finance and Economics, 2022, doi: 10.1007/s11146-022-09893-1.

[6] A. Alyousfi, "House Price Prediction Using Machine Learning Techniques," 2018.

[7] T. Quang, N. Minh, D. Hy, and M. Bo, "Housing Price Prediction via Improved Machine Learning Techniques," in Procedia Computer Science, 2020, vol. 174, pp. 433–442. doi: 10.1016/j.procs.2020.06.111.

[8] M. Dimoski and M. Pettersen, "Predicting Housing Prices with Machine Learning NORWEGIAN SCHOOL OF ECONOMICS."

[9] K. S. Liyaqatullah, U. Malik, and Q. M. Basheeruddin, "This work is licensed under a Creative Commons Attribution 4.0 International License HOUSE PRICE PREDICTION USING MACHINE LEARNING," International Journal of Advanced Research in Computer and Communication Engineering, vol. 10, 2021, doi: 10.17148/IJARCCE.2021.10468.

[10] M. Barari, N. Sarkar, S. Kundu, and K. B. Chowdhury, "Forecasting House Price in the United States: A Time Series Study with Focus on Multiple Structural Breaks."

[11] S. M. Satapathy, R. Jhaveri, U. Khanna, and A. K. Dwivedi, "Smart Rent Portal using Recommendation System Visualized by Augmented Reality," in Procedia Computer Science, 2020, vol. 171, pp. 197–206. doi: 10.1016/j.procs.2020.04.021.

[12] M. Pratik Kamble, S. Bavdhankar, A. Patil, A. Joshi, A. Nalawade, and M. Deshpande, "Real Estate Recommendation using Hybrid Recommendation Algorithms," 2021.

[13] M. Mubarak et al., "A Map-Based Recommendation System and House Price Prediction Model for Real Estate," ISPRS Int J Geoinf, vol. 11, no. 3, Mar. 2022, doi: 10.3390/ijgi11030178.

[14] R. Monika, J. Nithyasree, V. Valarmathi, M. G. R. Hemalakshmi, and N. B. Prakash, "House Price Forecasting Using Machine Learning Methods," 2021.

[15] M. Pratik Kamble, S. Bavdhankar, A. Patil, A. Joshi, A. Nalawade, and M. Deshpande, "Real Estate Recommendation using Hybrid Recommendation Algorithms," 2021.

[16] B. Klaus De Aquino Afonso, L. Carvalho Melo, W. D. Gomes De Oliveira, S. Bruno, S. Sousa, and L. Berton, "Housing Prices Prediction with a Deep Learning and Random Forest Ensemble." [Online]. Available: https://www.kaggle.com/c/data-science-challenge-at-eef-2019/

[17] C. Xue, Y. Ju, S. Li, Q. Zhou, and Q. Liu, "Research on accurate house price analysis by using gis technology and transport accessibility: A case study of xi'an, china," Symmetry (Basel), vol. 12, no. 8, pp. 1–21, Aug. 2020, doi: 10.3390/sym12081329.

[18] A. v Farooq P and C. Rose Paul, "Web Application for House Price Prediction," International Research Journal of Engineering and Technology, 2022, [Online]. Available: www.irjet.net