ISSN (Online): 2455-3662
EPRA International Journal of Multidisciplinary Research (IJMR) - Peer Reviewed Journal
Volume: 12| Issue: 4| April 2026|| Journal DOI: 10.36713/epra2013 || SJIF Impact Factor 2026: 8.753 || ISI Value: 1.188

REAL ESTATE LAND PRICE PREDICTION

Dr. V. Udhayakumar[1], V.S. Hari Haran[2]
[1]*Associate Professor, Department of Master Computer Application, Sri Manakula Vinayagar Engineering College, Pondicherry, India.*
[2] *Student, Department of Master Computer Application, Sri Manakula Vinayagar Engineering College, Pondicherry, India*

ABSTRACT

The Land Price Predicting System is a Web-Based Application tool built for predicting land prices through Machine Learning. Predicted property values can be established through input variables like location, geographical Area type, and District. The dataset used in this analysis was created using Real Estate datasets compiled from 10 different Districts to guarantee accurate predictive power. The system uses a MySQL database, which is a relational database management system, to store the required data digitally and efficiently. The use of various ML algorithms such as Linear Regression, Decision Trees, and Random Forest will increase the accuracy of the predictions. Generating and saving summary results as PDF files are also included as features in the proposed system for documentation.

The proposed system was developed to deliver an efficient, reliable, and user-friendly means of estimating land values. The system is intended to assist buyers/sellers/Real Estate analysts in making better-informed decisions through data-driven insight. Future enhancements of the system will include added real-time data sources and improved prediction models for better overall performance.

KEY WORDS: *Real Estate, Land Price Prediction, Machine Learning, Flask, Python, Web Application, MySQL, Regression Models, Random Forest, Decision Tree, Property Valuation, Predictive Modeling*

1. INTRODUCTION

There is no doubt that as our economy continues to grow rapidly, the Real Estate industry has a large impact on that growth. Thus, estimating the value of land before buying, selling, or investing is very important. But, this task is not an easy one and can be very challenging to do correctly; the price of land can depend on many different factors, including location, surface area, available infrastructure, and many types of market conditions. Oftentimes, a person will either rely upon manually estimating a property's value, or upon utilizing an agent's estimation; both may not be very reliable methods for accurately determining the price of a land parcel, leading to a potential for erroneous valuations.

As technology has rapidly advanced, the ability of machine learning to aid with complex prediction problems and produce accurate predictions has grown as well. Machine learning allows a system to "learn" from the previous data so that it will be able to make predictions without any direct human involvement. Therefore, by examining historic data *(in this case from the Real Estate industry*), the machine learning model will be able to identify patterns and trends, which would allow them to more accurately estimate future values for a land parcel. Human error has a much lower chance of occurring when the model is making price estimations than if a human had manually made a price estimation, which will ultimately achieve better and effective decisions.

In this case, the project will create the Real Estate Land Price Prediction System that will provide an easier process to obtain an estimate of the value of land by providing a way for a person to obtain a quick and easy *(automatically calculated and based on inputs made*) price estimate from a provided input. This system is a web-based application where users can enter the district, location and land area to receive an estimate of the land price almost immediately. The backend of the system was created using the Python programming language with the Flask framework while the frontend of the system was created using HTML, CSS and JavaScript allowing for a user-friendly interface. The data used to make these estimates is stored in and managed by a MySQL database.

In order to generate more accurate estimates, the system uses machine learning techniques such as linear regression, decision trees and random forests respectively. Additionally, estimates can be downloaded as a PDF. The main motivation for this system is to provide users with a quick and reliable method of estimating the prices of land.

2. LITERATURE SURVEY

Numerous different methods have been presented by researchers for predicting real estate prices. Previously, many systems used simple methods such as Linear Regression, providing fundamental estimates based on identifiable factors (*i.e., Location and Size*). These early approaches tend to be constrained and may fail to accurately predict complex relationships between various attributes. The use of Machine Learning continues to evolve from simple models to sophisticated algorithms including Decision Trees and Random Forests now offer enhanced data pattern recognition capabilities and greater flexibility for utilizing additional input fields. Additionally, there has been increased effort by researchers in developing web-based systems for providing users access to available predictions.

ISSN (Online): 2455-3662
EPRA International Journal of Multidisciplinary Research (IJMR) - Peer Reviewed Journal
Volume: 12| Issue: 4| April 2026|| Journal DOI: 10.36713/epra2013 || SJIF Impact Factor 2026: 8.753 || ISI Value: 1.188

Studies have also included other unique predictive variables (*e.g., adding distance from the subject property to various local features, e.g., bus stops, schools, markets, etc.; the existence of specific amenities at the subject property*). Additionally, proper data pre-processing techniques and feature selection processes have also been identified by researchers as key components for improving prediction accuracy. As a result of these studies, it has been established that Machine Learning is the most technology-efficient and dependable.

3. THEORETICAL FRAMEWORK

This project was developed using a Machine Learning framework. It takes advantage of the extensive amount of accumulated real estate data, to identify trends and correlations among many different variables. As a result, it combines inputs such as the city, location, distance to a bus stop, distance to a school, distance to a market, as well as the availability of water and roads into variables that have value when developing an accurate model.

Once the input has been determined, machine learning techniques such as linear regression, decision trees and random forests will be utilized to process these inputs. During the training phase, the model learns from the historical data, so that the model can leverage this knowledge when predicting land prices of the new inputs. In addition, data preprocessing techniques were used to clean up and organize the historical dataset to provide additional performance improvements.

In conclusion, the structured framework of the data will have a predictable flow from input, processing, prediction to output. This structure provides greater accuracy and creates reliability to use this system for real-time evaluation.

4. METHODOLOGY

The methodology of the project includes Real Estate Data Collection from 10 districts (cities), with some essential characteristics like the city name, location, distance to major transportation (*bus, school, and market*) locations, and availability of water and road facilities. Collected data is then pre-processed, to identify and eliminate any missing values, or errors in the data, and inconsistencies so that the end result is a more accurate representation of actual data values.

Pre-cleaned data undergoes a feature selection process to identify which data points are most useful predictive factors; after the dataset is segmented into training and test sets, Machine Learning algorithms (*Linear Regression, Decision Tree, and Random Forest*) are utilized to train the predictive Model with the training data.

The effectiveness of each model is assessed against the test set, and the model that performed best (based on accuracy) is selected; once the most effective Model has been selected, it is connected to a Flask Web-based Application solution.

Users can submit data inputs via the front-end interface (*HTML, CSS, and JavaScript*), and data provided, is processed in the back-end, returning a prediction of land price, input into the application for immediate display. .

Table: Comparison of Current System vs Proposed System

Feature	Current System	Proposed System
Core Functions	Land price is determined through overall guestimation of land price, or through the opinion of a broker	The price of land will be predicted using the Machine Learning.
Accuracy	According to experience, this pricing can be hit or miss, thus not always accurate	Data will provide much more specific pricing for land than without.
Data Usage	Historically, very little or no proper data for pricing was collected.	Pricing will now use a data set of 10 districts to help with predicting pricing for land.
Processing Speed	Pricing will historically take several hours and days to be valued.	Pricing will now be determined almost instantaneously.
User Interaction	typically offline pricing would use agents.	With new version of web application, pricing will be much simpler.
Data Storage	Data has never been stored properly.	After pricing is determined, MySQL will be used to store the data.
Accessibility	Limited method exist for obtaining today's pricing.	Pricing will always be available through use of the website.
Report Generation	No method exists for saving reports of land price from historical records.	After pricing is determined, users will have the ability to download as a PDF format.

5. Existing System

Typically most valuation and pricing of land is performed by people through the usage of real estate agents, brokers, or local knowledge. The method relies on experience and assumptions rather than on thorough data analysis; therefore, the method is variously inaccurate.

Many times sellers and buyers do not have access to the real market valuation of the land that they are buying/selling. Rather they rely on recommendations from friends or relatives which can lead to inflated or deflated prices for the land. As a result, some users will feel the process is not just an essential but also fair.

ISSN (Online): 2455-3662
EPRA International Journal of Multidisciplinary Research (IJMR) - Peer Reviewed Journal
Volume: 12| Issue: 4| April 2026|| Journal DOI: 10.36713/epra2013 || SJIF Impact Factor 2026: 8.753 || ISI Value: 1.188

The current system does not utilise more modern technological approaches such as Machine Learning for data analysis. The data consideration process tends to focus on only the most fundamental factors (ie location or area). In comparison, other critical factors may have a significant effect, including accessibility or proximity of other services & public transportation & the quality of the infrastructure associated with the land.

A further negative consequence of the current system is that the time a would-be seller/buyer must be preparing to purchase would require many hours of travel or discussion with various agents. In addition, there is not an adequate means of storing and providing proper access to the real estate data. In addition to being unable to provide any kind of report or documentation related to prior calculations, virtually all of this work takes place "off-line" and thus suffers from poor accessibility and convenience. To summarise, today's land valuation process has many issues. Some examples of these deficiencies include: they lack speed and accuracy, rely on manual processes too much, and they do not produce accurate or intelligent automatable estimates. The above list of deficiencies shows there is a real requirement for an alternative solution providing fast, factual,/results driven land value assessments.

6. Proposed System

By integrating machine learning technology into our systems and utilizing data analysis instead of manual processes to make land price predictions, we are creating an effective, accurate method that can be used worldwide. The new web-based application allows for easier user access and consumption of the land price prediction model. The primary focuses of this model include improving accuracy, speed and overall user experience. The land price prediction model considers multiple input factors so all users have access to more accurate and realistic estimates of land prices.

6.1 Data Collection Input

In this system, the acquisition of data is critical to improving the accuracy of predictions. The acquisition will include ten distinct regions in order to create a much larger and more varied dataset that includes real estate records; thereby allowing the system to recognize differences in the ways the values of land have been determined. The dataset represents sampled data as well as detailed records about where the sample was taken; to whom it Applied; where they were located; etc.; The data should be very detailed.

Users must also provide certain data elements that include the city they wish to investigate; the district (*based on where the potential property is located*); the distance to a bus stop; the distance to a school; and the distance to a grocery store (*or other markets*). The above distances will be expressed in kilometers so users can better understand how far away the location of their property is from the above locations. The following are additional factors that will be considered when making property valuations: access to water supply; or access by road are very important to property makers.

6.2 Machine Learning Model

After the data has been gathered it can then be processed and trained by creating a Machine Learning model. First, the data will undergo a cleaning process before being entered into the Machine Learning model; this will help to improve the overall quality of the data set by removing any missing values, errors and/or unwanted information from the data set. Next, each input feature's relevancy will be evaluated in order to identify which features should be included in the Machine Learning model.

The data set will then be broken up into two different sets: a training set and a test set. The training set will use one of several different types of Machine Learning algorithms *(i.e., Decision Trees, Linear Regression, Random Forest*) to train the model on that set of data. Each algorithm will learn how to analyze and interpret the data, but will interpret that data differently and will learn the relationship between input feature(s) and land values differently too.

Finally, the performance of each model will be evaluated based upon how well it was able to create a prediction when given the testing data. The model that creates the best prediction based on accuracy from the testing data will be selected as the final Machine Learning model. This insures that the system will deliver accurate and dependable results.

6.3 Collaborative Virtual Assistant (CVA)

The system that has been proposed will be available in a form that will allow all users and interested parties to easily access, use, and benefit from the application, as it will be web-based. The backend of the proposed system will be built using the Python programming language and the Flask framework, which is responsible for performing all data processing and prediction operations on behalf of the user by connecting the user to the Machine Learning model through the user interface located within the proposed application.

The frontend of the proposed application will be developed using HTML, CSS, and Javascript to provide a simple user interface that users will be able to navigate with ease, as well as provide the user with the ability to supply input to the application without needing to possess any kind of technical knowledge, thus allowing all system users to operate the system. The proposed application will connect to and use a MySQL database as the primary means for storing, managing, and maintaining the real estate data that is utilized by the proposed application in order to ensure that the application performs correctly and optimally; integrating the frontend and backend of the proposed application with the MySQL database allows all necessary parts within the proposed application to work together smoothly.

7. SYSTEM ARCHITECTURE

The input layer is responsible for collecting the information you want to analyze; this includes the city, location, distance from the nearest bus station, as well as any other property data such as the availability of water-based infrastructure, public transportation, etc. This data is transferred to the processing layer, where it can be cleaned and made ready for analysis.

ISSN (Online): 2455-3662
EPRA International Journal of Multidisciplinary Research (IJMR) - Peer Reviewed Journal
Volume: 12| Issue: 4| April 2026|| Journal DOI: 10.36713/epra2013 || SJIF Impact Factor 2026: 8.753 || ISI Value: 1.188

The ML model then uses this information for training purposes based on previous training. Finally, using algorithms such as linear regression models, decision trees or random forests, the model makes predictions regarding the price of the land for all users within that dataset.

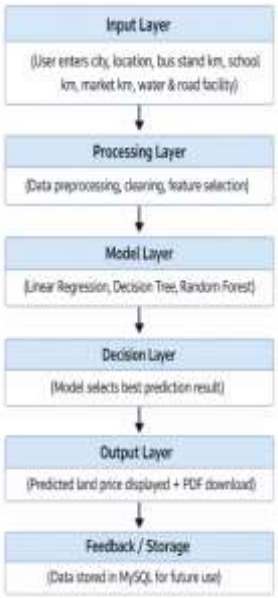

Fig: Architecture Diagram

8. APPLICATIONS

Before buying land, the buyer can utilize this www system to generate an estimate of the appropriate value of the property. This system can also be utilized by a seller to price his/her property with a competitive, yet reasonable selling price. Additionally, agents utilize this system to make quick and data-based recommendations for their clients.

This system can be useful to the investor by allowing him/her to analyze the values of properties in several different locations in order to make better investment choices. Similarly, researchers and analysts working in the real estate business have found this system to be very beneficial and helpful. In general, the entire process of estimating the price of land can be made faster, simpler and more accurate using this www system.

9. EVALUATION METRICS

Evaluation metrics are utilized to assess the accuracy of Machine Learning (ML) models when predicting land prices. They allow us to understand how accurately and reliably the model predicts land prices. One commonly used evaluation metric is Mean Absolute Error (MAE), which measures the average absolute error of predicted (P) prices from their actual (A) prices. Another commonly used evaluation metric is Mean Squared Error (MSE) that calculates the sum of the squares of each absolute error, where larger errors will receive more weight than smaller errors.

Lastly, the Root Mean Squared Error (RMSE) provides another way to determine the quality of the P values your P model has predicted to the same degree of accuracy as the A values.R squared (R^2) is also used to measure how well your model has accounted for the variability present within a dependent variable. In other words, as you get closer to an R^2 of 1, your model can account for greater amounts of variability within the dependent variable.Furthermore, these metrics will allow an assessment of which of the three models (Linear regression Decision Tree) provided the highest accuracy. Thus, evaluation metrics ultimately facilitate the selection of a machine learning model with the highest level of prediction capabilities and an overall higher-performance machine learning system via the processes mentioned above.

10. BENEFITS

This platform will be able to accurately forecast property values through detailed algorithms rather than guessing randomly. Instantaneous accessibility via online web platforms allows users immediate access to result data without requiring extensive technical expertise or experience.

In turn, both buyers and sellers will have improved decision-making abilities regarding real estate transactions through this method. Furthermore, this offer reduces manual processing and increases efficiency within existing systems. Lastly, allowing downloadable forms of report results in PDF format makes them easy to make use of as both documents and reference materials later on.

11. ETHICAL AND PRACTICAL CONSIDERATIONS

Although this system has a good way to predict future land price fluctuations; it is very important that you have correct and unbiased information. If the information is incorrect, or not complete, that could result in you making poor purchasing decisions. Your user's privacy will be protected, so their personal information will always be kept private. I do not believe that there will ever be anyone with the ability to abuse the system to create a financial advantage.

Finally, I would suggest that you continuously add new data to the model for better accuracy. You should also make sure to provide users with accurate predictions regarding land prices, and that users will be able to find those predictions to be a close approximation of the actual value of the land. The company must also provide on-going maintenance and improvements to support the company's commitment to remain a reliable source of information.

12. FUTURE DIRECTIONS

This system could be enhanced through future implementations of realtime data for improved forecasting accuracy. It is possible that new attributes can be added such as adjacent hospitals, shopping areas, and transit facilities upon which further analysis can be conducted.
Advanced Machine Learning and Deep Learning techniques can be applied to enhance system performance. Additionally, there are multiple options for expanding the system to different districts and/or states.

Map service integration would allow the user to view property locations with greater clarity. With future enhancements, a mobile application version of this system would improve accessibility for users. These enhancements would make this system better at serving the user and create additional value for

ISSN (Online): 2455-3662
EPRA International Journal of Multidisciplinary Research (IJMR) - Peer Reviewed Journal
Volume: 12| Issue: 4| April 2026|| Journal DOI: 10.36713/epra2013 || SJIF Impact Factor 2026: 8.753 || ISI Value: 1.188

all involved.

13.CONCLUSION

The project introduces a new machine-learning-based solution designed to help predict land prices quickly and accurately using various input factors, such as location, proximity to amenities, etc. The current technique is more accurate than a manual approach. The project features a simple website that is easy to navigate and provides users with instant feedback about their land valuation requests.

By utilizing multiple input parameters and trained models, the project will provide an enhanced level of confidence in price estimation for land. It will also reduce the amount of time and effort required to estimate land values. The ability to download the estimated pricing documents in PDF format will add a significant benefit for documenting the findings of his or her estimated land value and referencing it in the future. This project also shows how utilizing technology can solve everyday issues within the real estate profession. Additionally, the proposed project provides an excellent platform for future, more advanced features.

14.REFERENCES

1. *T. M. Mitchell,* Machine Learning. *New York: McGraw-Hill, 1997.*
2. *C. M. Bishop,* Pattern Recognition and Machine Learning. *New York: Springer, 2006.*
3. *S. Raschka and V. Mirjalili,* Python Machine Learning. *Birmingham: Packt Publishing, 2019.*
4. *F. Pedregosa et al., "Scikit-learn: Machine Learning in Python,"* Journal of Machine Learning Research, *2011.*
5. *L. Breiman, "Random Forests,"* Machine Learning Journal, *2001.*
6. *J. R. Quinlan, "Induction of Decision Trees,"* Machine Learning Journal, *1986.*
7. *A. Géron,* Hands-On Machine Learning with Scikit-Learn and TensorFlow. *California: O'Reilly, 2019.*
8. *I. Goodfellow, Y. Bengio, and A. Courville,* Deep Learning. *Cambridge: MIT Press, 2016.*
9. *Flask Documentation, "Flask Web Framework." Available: https://flask.palletsprojects.com*
10. *MySQL Documentation, "MySQL Database System." Available: https://www.mysql.com*
11. *J. Brownlee, "Machine Learning Algorithms with Python,"* Machine Learning Mastery, *2016.*
12. *H. Zhao, "Real Estate Price Prediction using Machine Learning,"* International Journal, *2019.*
13. *S. Kok, "House Price Prediction using Machine Learning,"* International Journal, *2020.*
14. *R. Singh and P. Gupta, "Property Price Prediction using ML Techniques,"* International Journal, *2020.*