Predictive Analytics for Real Estate Valuation

Ashishdeep
Computer Science and Engineering
Chandigarh University
Mohali,Punjab
Deepc7340@gmail.com

Dr.Vinod Kumar
Computer Scince and Engineering
Chandigarh University
Mohali,Punjab
Vinod.e16460@cumail.com

JasveerKaur
Computer Science and Engineering
Chandigarh University
Mohali,Punjab
jasveerjassowal@gmail.com

Navdeep chawla
Computer Science and Engineering
Chandigarh University
Mohali,Punjab
Navdeepchawla181@gmail.com

***Abstract*—** A subfield of artificial intelligence called machine learning makes it possible for software programs to forecast data more accurately, to forecast performance in the present, and to get better at the future. The application of machine learning techniques for price prediction and appraisal is reviewed in this work. In order to assist policymakers in evaluating the state of the economy as a whole, the authors research the best ways to forecast property price indexes. Additionally, this study investigates the application of machine learning to property valuation in order to determine the most effective model for forecasting property prices based on attributes like location, land size number of rooms, and others.

Keywords—Real estate, machine learning, forecasting property and valuation.

I. INTRODUCTION

Since its discovery in the early years, machine learning(ML) has advanced and been used extensively. According to ja'afar and Mohamad(2021), machine learning technologies have expanded and improved their capabilities across a range of applications .A computer progam and subfield of artificial intelligence, machine learning(ML) is used to recognize ,gather ,and enhance data performance in order to fulfill its functions as a prediction model(Kamaloy & Gurrib,2021) . Put differently, machine learning(ML) gains knowledge from past experiences to forecast performances in the present and enhance it for data in the future. There are several types of learning in machine learning ,including supervised and unsupervised learning category that examine various patterns. Ml operates by analyzing past patterns using specfic algorithims and forecasting futures outcomes based on observations by analyzing past pattterns using certain algorithims (Oladunni,2016).

One advantage of machine learning is its ability to enhance iterative algorithms efficiency by caching previously visited datasets. By avoiding overfitting on datasets containing noise or numerous additional variables, Park & Kwon(2015) are able to reasonably predict volatile and uncertain markets. Manufacturing ,education, financial modeling, policing Jordan(2015),Sarip(2015), healthcare Ghassemi(2018), engineering (Bege),2019, medical (Christodoulau,2019), and transportation system (Maalel,2011).

II. LITERATURE REVIEW

In accordance with the Systematic Review Preferred Reporting items and Reviewing technique for meta-analyses involves four steps, such as participation (Lalu, Li, & Loder, 2021). In the identification process, authors use a range of keywords in electronic databases to search the literature for relevant papers. PRISMA's goal is to make the literature review easier to identify. To create a thorough literature evaluation, the authors looked at number of earlier journals. The Lalu et al, (2021) checklist includes topics that need to be examined, including establishing precise research objectives, setting inclusion and exclusion criteria, and looking through a few databases for scientific literature. Additionally, researches employed two peer-reviewed publication databases – Scopus and Web Science – for their literature assessment. These databases are the most comprehensive and widely used, serving as the main rival database for journal rating statistics and citation analysis. This subheading is categorized as the first step in the literature review identification process. When it came to penetrating keywords and queries string information strategy were employed, TITLE-ABS- KEY ("machine learning") AND "real estate" AND "price" AND "price prediction" AND "predict" AND "price predict" OR "property" OR "house" OR "housing"),("Machine learning" AND "price AND predict"), are among the query string search items used by authors from Scopus. In the meantime, TITLE-ABS-KEY "machine learning" OR "housing" are used for the Web of Science query string search ("Machine learning" AND "Real Estate" AND "price AND predict").

Main analyses

Numerous fields have found modern machine learning to be a useful application (Jordan,2015) define machine learning(ML) as a statistical learning of predictive analysis, a subset of artificial intelligence that is applied to a variety of tasks such modeling, designing, programming, and recognition. In addition, machine learning(ML) is the process of obtaining knowledge from input data in order to produce output. According to a study by Kilibarda (2018), scientists have determined that machine learning(ML) is a substitute for model- predicting algorithms in the twenty-first century. Due to distinct approaches to learning processes have different algorithms. According to earlier research, supervised machine learning is the most often used learning dataset is . The application of ML must provide complete dataset for further prediction, ML will learn the given dataset in the entire system from start entire system from start until the produce result. Previous study has applied ML techniques to predict various study to observe the possibility to get an accurate result (Varma & Sarma,2018). The hedonic price regression is mainly been used for inferences. In contrast, ML has a great potential in prediction.

The market value of real estate is assessed through the valuation methods by following the existing procedures to reflect the nature and circumstances of property to meet the market value definition .Every country has a different cultural and environment backgrounds, thus it has dissimilarity in determining the appropriate method for each particular property(Pagourtzi, Assimakopoulos, & Thomos,2003). In the valuation process, there are several methods such as traditional and advanced method has been practiced in Malaysia and it has several methods as stated in the diagram. Due to the nature of the method which has several limitations and restrictions to produce accurate value, the advance method has been

adopted in carrying out valuation prediction (Olanrewaju & Lim,2018). Even with these developments, there are still gaps in the literature. For instance, real-time data integration which is essential in volatile real estate markets, is absent from the majority of models. Furthermore, the predicted accuracy of many models is constrained by incomplete feature sets, such as missing information on crime rates, geographical characteristics, or accessibility to facilities. In order for predictions to the transferable between regions, models that generalize well across various real estate markets are also required. By filling in these gaps, this study intends to advance the area by utilizing a comparative analysis of machine learning and hybrid models and providing a novel feature engineering approach. With this strategy, the study hopes to provide insightful information about AI-driven real estate value, utilizing cutting-edge methods to improve precision and flexibility in practical.

III. Objective

This study paper's main goal is to create and assess an AI-based model that can forecast real estate market values and prices more accurately, flexible, and effectively than conventional valuation techniques .Numerous influencing elements, including location, property qualities, market conditions, and external socio-economic indicators, contribute to the complexity of real estate assessment. In order to determine the best prediction model for real estate price, this study uses sophisticated machine learning and deep learning models, such as ensemble approaches, neural networks, and regression-based techniques. Furthermore ,by employing a novel feature engineering methodology, this stud seeks to close current gaps by enabling a deeper integration of region-specific characteristics and real-time data elements that are frequently absent in traditional methods. By comparing various AI models and analyzing their performance metrics, this study will not only evaluate model accuracy and adaptability but will also contribute valuable insights into how AI can enhance real estate price prediction. This research ultimately seeks to provide a robust, data-driven solution for property valuation that can benefit investors, buyers, and sellers by enabling more informed and accurate decision- making in the real estate market. Furthermore, the study will evaluate the generalizability of these AI models with the goal of developing a prediction system that is not limited to a specific geographic area but can adjust to many regional markets. For real estate stakeholders and investors who work in many markets and need consistent valuation models, this flexibility is essential. By accomplishing these goals, this paper hopes to offer a through, data driven solution that enable different real estate industry stakeholders, including investors, analysts, purchasers, and sellers, to make defensible judgements based on precise, scalable and real-time property assessments.

IV. Methodology

This research paper technique is focused on creating, evaluating, and verifying AI-based models to provide highly accurate and flexible real estate properly price predictions. The process start with gathering data, which is sourced from government property records and publicly accessible databases like Zillow, Redfin, and other real estate portals. These records include important details like property features (like square footage, age, and number of rooms), location details(like neighbourhood quality, proximity to amenities, and school districts), and market trends (like past price trends

And interest rates). Additional socioeconomic variable, such as local crime rates and demographic trends, are added to this dataset to increase its depth, This provides a holistic view that allows for deeper model insights.

To make sure the dataset is clean, consistent, and appropriate for machine learning models, data preparation is done after data collection. This stage involves scaling numerical features, encoding categorical variables (e.g., converting neighbourhood names to numerical codes), addressing missing values using imputation techniques, and eliminating outliers that could distort the predictions. Furthermore, this process relies heavily on feature engineering, which creates new variables by merging or altering preexisting features. For example, factors such as "price per square foot" are designed to increase forecast accuracy and model interpretability. To reduce dimensionality and increase model efficiency, only the most significant features are kept by using feature selection strategies including correlation analysis and feature importance ranking.

Model selection and training are the following steps after the following steps after the dataset has been improved. Because of their distinct capacities to manage structured real estate data, a number of machine learning and deep learning algorithms are evaluated, including XG Boost, decision trees, random forests, linear regression, and deep neural network(DNNs). Several performance metrics, such as R-squared, Mean Absolute Error(MAE), and Root Mean Squared Error (RMSE) are used to evaluate models. These metrics together offer a thorough examination of each models are not just suited to the training data but also have good generalization, a cross-validation technique is used. Comparative evaluation of these models aids in identifying the best algorithm for the purpose of predicting real estate prices, In order to determine which property and location attributes have the most predictive capacity, feature importance analysis is also carried out to evaluate the aspects that have the biggest effects on price forecasts. Last but not least, model optimization is done by employing strategies like random search and grid search for hyperparameter tuning, which improve model performance by identifying the best setups for every algorithm. Results are evaluated in terms of predicted accuracy, computational efficiency , and adaptability across various regional marketplaces after the final model is validated on a test dataset to verify its practicality. This methodology seeks to provide a solid, scalable AI solution that can provide precise, dynamic, and context- sensitive property price prediction for the real estate sector by methodically tackling data quality, model selection, evaluation, and optimization. A crucial element is feature engineering which aims to create significant characteristics that improve predictive power. Creating derived features that offer more information, including "price per square foot," 'average neighborhood price," and "distance to the city center," is part of this process. To enable the model the capture intricate connections, interaction terms are also produced, such as the relationship between location desirability and property size. In order to reduce model complexity and improve interpretability, the dataset is further refined using feature selection techniques, such as correlation analysis and recursive feature elimination, to keep just the most significant variables and prevent multicollinearity. To guarantee consistency and quality, data preparation is done after collection. In order to improve model convergence and performance, this step entails encoding categorical data, such as property type or neighborhood using one-hot or ordinal encoding based on feature requirements scaling numerical data with normalization or standardization ro bring all features to a similar scale; and handling missing values using techniques like mean, median mode imputation for numerical data and frequency- based attribution for numerical data and frequency- based estimation for categorical features. Unnecessary or irrelevant features are eliminated to lower noise and boost model efficiency. While data outliers are found and handled using statistical techniques or robust scaling to avoid skewed findings.

Identify applicable funding agency here. If none, delete this text box.

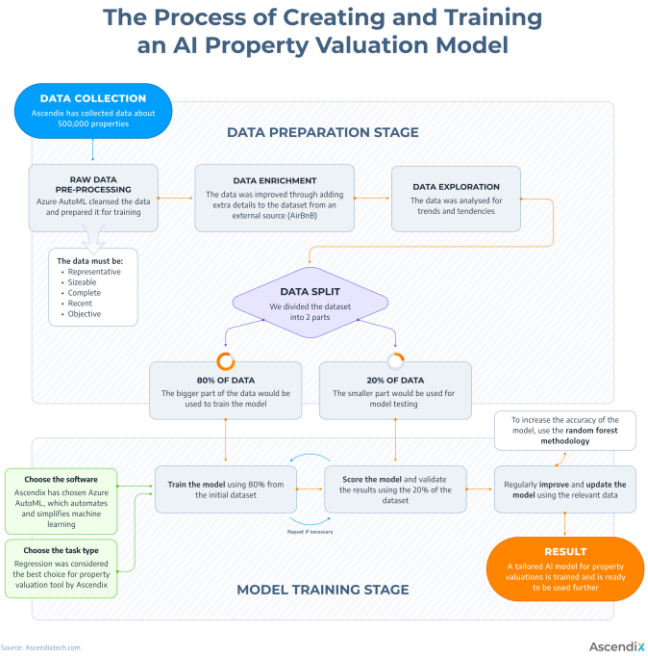
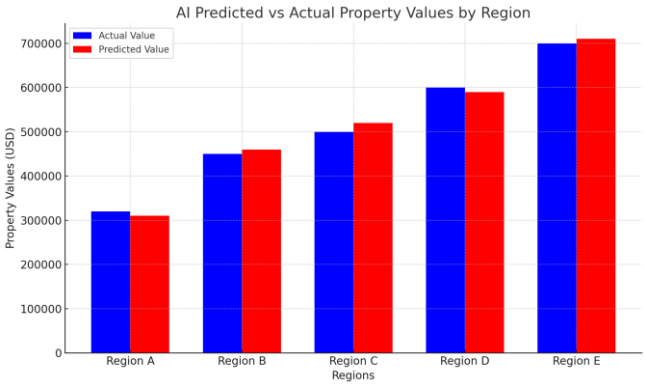


Fig. 1. Example of a figure caption. (*figure caption*)

The real estate market's actual and AI-predicted property prices from 2020 to 2024 are contrasted in this bar graph.The vertical axis shows real estate prices in US dillars, and the horizontal axis shows prices projected by the AI model, and a sky- blue bar shows prices projected by the AI model, and sky-blue shows real market values for each year.



V Software Requirement specification

With the use of machine learning and deep learning models, an AI- powered system will analyze real estate data and forecast property values. Data intake, preprocessing, model training, evaluation, and a prediction user interface will all be part of the system.

- Python: Python is the main language used for developing APIs, machine learning, and data processing.
- Jupyter Notebook: For iterative model building. Data exploration, and prototyping.
- IDE: PyCharm or VS Code for efficient code management and debugging.
- Pandas and NumPy: Both are crucial for numerical calculations, data cleansing, and data manipulation.
- MongoDB : Used to store external market data, historical pricing and property data.
- TensorFlow : For developing deep learning models which may call for intricate, non-linear correlations in price forecasts, TensorFlow is suitable options.
- DVC : Data version Control, also known as ML flow is used to track performance metrics across iterations, manage model versions, and oversee machine learning studies.
- Matplotlib and Seaborn : For typical data visualization of feature distributions and model performance indicators, use Matplotlib and Seaborn.
- Web Framework : Flask or Django for backend development, handling user requests, and serving predictions.

System Design.

Linear Regression is a great place to start for a simple AI-based real estate property price prediction project. By assuming a linear relationship between property values and their salient characteristics, Linear Regression provides a simple method. It's especially helpful because it offers an interpretable basis that enables.

EXERIMENTAL RESULTS *(Heading 5)*

An organized workflow for setting up and carrying out an AI-based real estate property price prediction project from beginning to end is provided by this set of instructions. Navigating to the project directory and setting up a virtual environment – an isolated configuration that guarantees dependencies are maintained without disrupting other projects on your computer- is the first step in the process.

```
(c) Microsoft Corporation. All rights reserved.

C:\Users\HP># 1. Navigate to the project directory
'#' is not recognized as an internal or external command,
operable program or batch file.

C:\Users\HP>cd path/to/your/project
The system cannot find the path specified.

C:\Users\HP>
C:\Users\HP># 2. Create a virtual environment (optional, but recommended)
'#' is not recognized as an internal or external command,
operable program or batch file.

C:\Users\HP>python -m venv venv

C:\Users\HP>
C:\Users\HP># 3. Activate the virtual environment
'#' is not recognized as an internal or external command,
operable program or batch file.

C:\Users\HP># On Windows:
'#' is not recognized as an internal or external command,
operable program or batch file.

C:\Users\HP>venv\Scripts\activate

(venv) C:\Users\HP># On macOS/Linux:
'#' is not recognized as an internal or external command,
operable program or batch file.

(venv) C:\Users\HP>source venv/bin/activate
```

References

Begel, A. (2019). Software Engineering for Machine Learning : A Case Study. *Microsoft Research*, 1–10.

Borde, S., & Rane, A. (2017). Real Estate Investment Advising Using Machine Learning. *Journal of Engineering and Technology (IRJET)*, *04*(03), 1821–1825.

Chardon, I., & Javier, F. (2018). *Housing Prices: Testing Machines Learning Methods*. 2–15.

Christodoulou, E. (2019). A Systematic Review shows no Performance Benefit of Machine Learning over Logistics Regression for Clinical Prediction Models.

Journal of Clinical Epidemiology, *110*, 12–22.

Crosby, H., & Davis, P. (2016). *A Spattio -Temporal, Gaussian Process Regressio , RealEstate Price Predictor*. 3–6.

We suggest that you use a text box to insert a graphic (which is ideally a 300 dpi TIFF or EPS file, with all fonts embedded) because, in an MSW document, this method is somewhat more stable than directly inserting a picture.

To have non-visible rules on your frame, use the MSWord "Format" pull-down menu, select Text Box > Colors and Lines to choose No Fill and No Line.

[1] Dellstad, M. (2018). *Comparing Three Machine Learning Algorithms in the task of Appraising Commercial Real Estate*. KTH Royal Institute of Technology.

[2] Di, N. F. M., Satari, S. Z., & Zakaria, R. (2017). *Real estate value prediction using multivariate regression models*

[3] Ghassemi, M. (2018). *A Review of Challenges and Opportunities in Machine Learning for Health*.

[4] Gu, G., & Xu, B. (2017). Housing Market Hedonic Price Study Based on Boosting Regression Tree. *Advanced Computational Intelligence and Informatics*, *21*(6).

[5] Lalu, M. M., Li, T., & Loder, E. W. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *Journal of Surgery*, *88*, 1–11.

IEEE conference templates contain guidance text for composing and formatting conference papers. Please ensure that all template text is removed from your conference paper prior to submission to the conference. Failure to remove template text from your paper may result in your paper not being published.