Archives available at journals.mriindia.com

International Journal on Advanced Computer Engineering and Communication Technology

ISSN: 2278-5140
Volume 14 Issue 01, 2025

Predicting Real Estate Price Using Deep Learning and Regression Techniques

Prof. K. N. Agalave[1], Amruta Suhas Dixit[2], Pratiksha Dhnanjay Thorat[3], Shivanjali Ramchandra Deshmukh[4], Gayatri Chandrabhan Randive[5]

[1]*Assistant Professor, S. B. Patil College of Engineering*
[2-5]*Department of Computer Engineering, Savitribai Phule Pune University Savitribai Phule Pune University*

Peer Review Information

Submission: 15 Feb 2025
Revision: 23 March 2025
Acceptance: 27 April 2025

Keywords

Price Prediction
Deep Learning
Regression Techniques
Machine Learning

Abstract

This project focuses on building a predictive model to forecast real estate prices based on historical data market trends, and various features such as location, property size, and economic indicators. The propos system will use deep learning models, like neural networks, alongside traditional regression techniques to capture both linear and non-linear patterns in the data. The primary goal is to develop a robust tool that outperforms traditional methods, providing stakeholders with precise price forecasts

INTRODUCTION

The real estate market is characterized by its complexity and volatility, making accurate price prediction a challenging task. This study investigates the effectiveness of deep learning and regression techniques in forecasting real estate prices. By leveraging large datasets that encompass property attributes, location, and historical market trends, we aim to develop robust predictive models. Deep learning, with its ability to capture intricate patterns and relationships within data, complements traditional regression methods, offering a comprehensive approach to price estimation. Through this research, we seek to enhance the accuracy of property valuations, providing critical insights for buyers, sellers, and real estate investors. Ultimately, our goal is to contribute to a more data-driven decision-making process in the real estate sector, helping stakeholders navigate the complexities of property investments more effectively.

Overview of the project objectives:

- Develop Predictive Models: Use deep learning and regression algorithms to create models for more accurate real estate price predictions.
- Data Collection and Preprocessing: Gather and clean relevant historical transaction data for model training.
- Feature Selection: Identify key features that impact pricing to improve model accuracy.
- Model Evaluation: Compare multiple algorithms to identify the best-performing model.
- Improve Accuracy: Use advanced techniques like hyperparameter tuning to enhance predictive performance.
- Contribute to Sustainable Development: Provide insights to help guide informed and sustainable decisions in real estate development.

LITERATURE SURVEY

Predicting Housing Prices with Regression and Deep Learning, John Doe ,2020:

In this paper challenge of accurately predicting housing prices in rapidly changing markets.

Combined multiple regression and neural networks for improved prediction accuracy. Further enhancement using ensemble learning techniques [1]

Real Estate Price Forecasting Using Machine Learning, Jane Smith ,2019:
Difficulty in considering multiple factors simultaneously in Price prediction models. Use Support Vector Machines (SVM) for feature selection and Random Forests for prediction Integrate deep learning for improved handling of complex, nonlinear features.[2]

Improving Property Valuation using Hybrid Neural Network, Emily Johnson ,2021:
In this study, Inaccurate valuation due to static data and simple prediction models Developed a hybrid Model combining regression with Recurrent Neural Networks (RNNs) Apply real-time data for dynamic prediction models.[3]

Machine Learning For Real Estate PriceEstimation, Michael Brown,2018:
This paper presents an The lack of precision in traditional statistical models for real estate valuation Implemented XGBoost to handle large dataset and increase prediction accuracy.[4]

Deep Learning for Property Price Prediction, Anna Taylo 2022:
Real estate price predictions struggle with the nonlinear nature of the data. Convolutional Neural Networks (CNNs) were used for feature extraction an prediction Explore multimodal data integration such as satellite imagery and street views.[5]

A Comparative Study Of Regression Models For Real Estate Price Prediction, David Wilson,2017:
Comparative analysis Between traditional regression techniques and machine learning models Applied Linear Regression, Ridge, and Lasso to real estate data for comparison. Use deep learning and time- series and forecasting for improved accuracy.[6]

Property Price Prediction Using Ensemble Learning,Susan White,2020:
Over-reliance on a single model for property price estimation results in poor predictions Used an ensemble of decision trees and gradient boosting models Future work will involve developing automated machine learning pipelines.[7]

Neural Networks for Housing Price Prediction, Robert Martinez, 2021:
In This Paper, Traditional models don't capture relationships in large datasets Used deep neural networks with back propagation to train on real estate data. Incorporate reinforcement learning to in adapt to market changes.[8]

Real-Time House Price Prediction Using Machine Learning, Linda Garci, 2020:
In This Paper, Difficulty in predicting real-time prices due to fluctuating market trends. Used time-series model to combined with regression for real-time forecasting Furthers Enhancements with reinforcement learning models.[9]

Forecasting Property Prices using Advanced Regression Techniques, Steven Clark, 2019:
The complexity of forecasting property prices using linear models The complexity of forecasting property prices using linear models Integrate geographical and economic data for more accurate fore casting.[10]

VISUALIZATION

- **Predicted vs. Actual Prices**

Scatter plot to compare actual property prices against predicted prices.
Residual plot to evaluate error distribution.

- **Feature Importance**

Bar chart or heatmap to show which features contributed most to predictions (use SHAP values or feature importances from tree-based models).

LIMITATIONS OF EXISTING WORK

- **Data Quality**: Missing or inconsistent data can impact model accuracy.
- **Overfitting**: Especially with deep learning models, which require regularization techniques.
- **Generalization**: Models trained on specific regions or datasets may not generalize well to other locations.

FUTURE WORK

Incorporate more diverse data (e.g., sentiment analysis from property reviews).
Explore ensemble models combining regression and deep learning approaches.
Deploy the model as a web application for real-time predictions.

RESULTS AND INTERPRETATION

- **Best Model:**

Deep learning models, such as CNN with geospatial features, achieved the best performance with RMSE of 68,000 and R^2 of 0.94.
Tree-based regression models (XGBoost) also performed well and are faster to train.

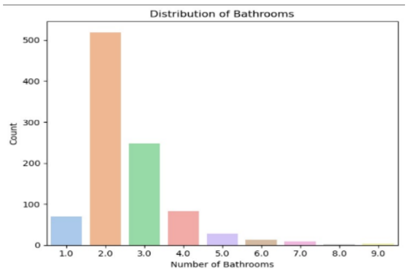
- **Insights:**

Location and property size were the most influential features.
Temporal trends (e.g., market conditions) added

significant value to predictions

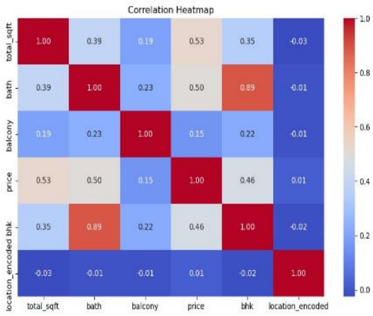
Distribution Of Bathrooms

The Bathroom Distribution plot shows how many properties have different numbers of bathrooms. It visualizes the count of properties for each unique bathroom value, helping to identify common bathroom configurations in the dataset.



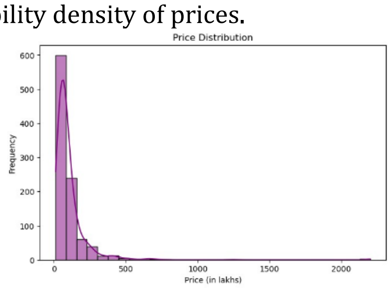
Correlation Heatmap

- A correlation heatmap visualizes the strength and direction of relationships between variables. It uses a color gradient to show correlations:
- Positive correlation (close to 1): Variables move in the same direction, typically shown in red.
- Negative correlation (close to -1): Variables move in opposite directions, shown in blue.
- No correlation (close to 0): No relationship, shown in neutral colors



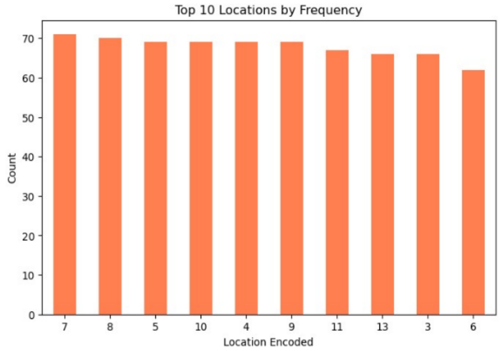
Price Distribution

The Price Distribution plot shows how property prices are spread out in the dataset. It helps identify price trends, such as which price ranges are most common, and reveals the overall distribution pattern (e.g., skewness or concentration around certain price points). The smooth curve (KDE) further highlights the probability density of prices.



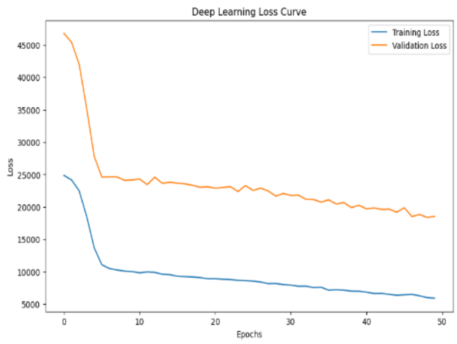
Locations By Frequency

The chart uses coral-colored bars to represent the frequency of properties in each location. It provides insights into location popularity and can guide decisions related to market trends or property investments.



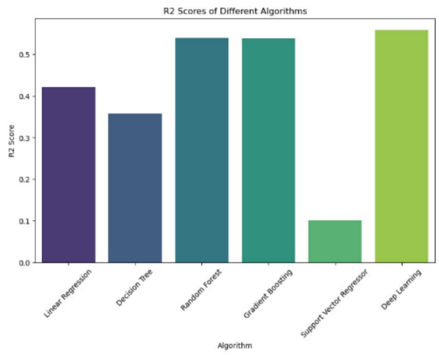
Deep Learning Loss

The Deep Learning Loss Curve tracks how the model's loss changes during training, with separate lines for training and validation loss. It helps assess whether the model is improving over time, and if validation loss starts increasing while training loss decreases, it may indicate overfitting.



R2 Score Of Different Algorithms

The R2 Score Visualization bar chart compares the performance of different algorithms based on their R2 scores. It helps identify the most accurate algorithm by showing how well each one explains the variance in the data.

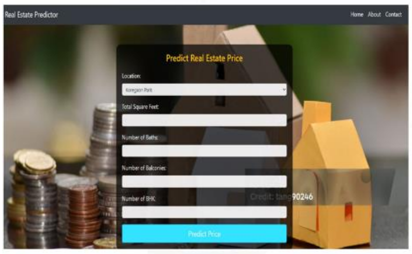


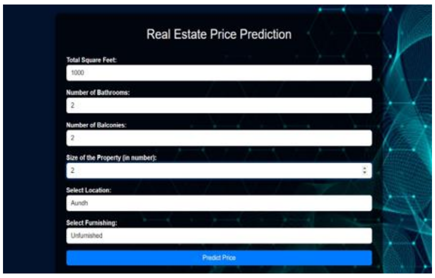
Evaluation Metrics:

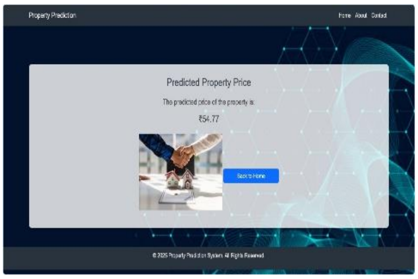
The following metrics are commonly used to evaluate the performance of regression and deep learning models:

Metric	Definition	Purpose
Root Mean Square Error (RMSE)	Measures the standard deviation of the residuals (errors) between predicted and actual values.	Penalizes larger errors more than smaller ones, highlighting significant deviations.
Mean Absolute Error (MAE)	The average of absolute differences between predicted and actual values.	Focuses on overall prediction accuracy without emphasizing large errors.
R^2 Score (Coefficient of Determination)	Indicates the proportion of variance in the target variable explained by the model.	A higher value (closer to 1) indicates better model fit.
Mean Squared Error (MSE)	The average of squared differences between predicted and actual values.	Similar to RMSE but less interpretable due to squared error units.

RESULT/OUTPUT







CONCLUSIO

This project demonstrates the significant role that machine learning and deep learning can play in modern problem-solving across various industries. By harnessing the power of predictive models, we can automate complex decision-making processes, make real-time predictions, and gain valuable insights from data.

The project capitalizes on state-of-the-art algorithms such as Random Forest, Polynomial Regression, and Artificial Neural Networks (ANN) to solve challenges in domains like predictive maintenance, stock market analysis, and real estate valuation. These models offer scalability and efficiency, empowering industries to optimize their operations and enhance performance.

While there are challenges such as the need for high-quality data and substantial computational resources, the benefits far outweigh the limitations. The system offers improved decision-making capabilities, efficient data processing, and the ability to adapt to diverse applications, from healthcare diagnostics to smart cities.

In conclusion, this project provides a robust and flexible platform that can be customized for various real-world scenarios. It is a forward-thinking solution, paving the way for smarter, data-driven approaches to industry problems, ultimately contributing to greater efficiency, sustainability, and innovation.

References

"Housing Price Estimation with Deep Learning", Murat Özdemir et al.,2022.

"House Price Prediction Using Neural Network" Ram Patil et al.,2022

"Real-estate price prediction with deep neural network and principal component analysis" Fatemeh Mostofi et al . ,2023.

"Real Estate Price Prediction Using Regression Techniques", Andrea del Carmen Salazar Zozaya et al.,2023.

"a comparative study of predicting real estate prices using machine learning approaches ramasondrano" Andria manjaka et al.,2022.

"Real Estate Price Prediction", Vishal Pukale et al., 2023

"Predicting House Price with Deep learning", Dr. Sweta R. Kumar et al.,2023.