JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

House Price Prediction

[11]Pinnamraju T S priya,[2]Jemimah Hebzibah Ambati,

[1]Assistant Professor, [2]MCA Final Semester,
[1]Masters of Computer Applications,
[1]Sanketika Vidya Parishad Engineering College, Visakhapatnam, Andhra Pradesh, India

Abstract : Usually, House price index represents the summarized price changes of residential housing. While for a single-family house price prediction, it needs more accurate method based on location, house type, size, build year, local amenities, and some other factors which could affect house demand and supply. With limited dataset and data features, a practical and composite data pre-processing.

Index terms : Dataset, Data collection, Data analysis, Train data set MSE, RMSE, MAE,

I. INTRODUCTION

- People looking to buy a new home tend to be more conservative with their budgets and market strategies. This project aims to analyses various parameters like average income, average area etc. and predict the house price accordingly. This application will help customers to invest in an estate without approaching an agent To provide a better and fast way of forming operations. To provide proper house price to the customers[13]. To eliminate need of real estate agent to gain information regarding house prices. To provide best price to user without getting cheated[16]. To enable user to search home as per the budget. The aim is to predict the efficient house pricing for real estate customers with respect to their budgets and priorities. By analysing previous market trends and price ranges, and also upcoming developments future prices will be predicted. House prices increase every year, so there is a need for a system to predict house prices in the future. House price prediction can help the developer determine the selling price of a house and can help the customer to arrange the right time to purchase a house. We use linear regression algorithm in machine learning for predicting the house price trend

1.1 Existing System

In existing system,Redfin offers a similar tool that provides estimates of home values. It considers factors such as property details, location, recent sales in the area, and user-provided data. Machine learning models are used to analyze and predict prices based on these inputs. Platforms like CoreLogic and ATTOM Data Solutions provide analytics and predictive modeling services for real estate professionals. They leverage AI and machine learning to analyze market trends, property data, and economic indicators to forecast home prices. Many data scientists and researchers participate in Kaggle competitions focused on house price prediction. Open-source models developed in these competitions often use advanced machine learning algorithms such as gradient boosting, neural networks, and ensemble methods to achieve accurate predictions. Large real estate companies often develop custom AI- driven systems for pricing properties. These systems integrate historical sales data, property features, neighborhood characteristics, and external factors like economic trends to generate price estimates. **Regression Models**: Linear regression, polynomial regression, ridge regression, etc., are commonly used for predicting continuous values like house prices.

Ensemble Methods: Random Forest, Gradient Boosting Machines (GBM), XGBoost, and LightGBM are popular for their ability to handle complex interactions and nonlinearities in data. **Neural Networks**: Deep learning models can capture intricate patterns in data but require more computational resources and data for training[15]. **Feature Engineering**: Creating meaningful features from raw data, such as calculating derived variables (e.g., price per square foot) or encoding categorical variables. **Geospatial Analysis**: Incorporating spatial features like proximity to amenities, neighborhood characteristics, and geographic trends. Overall, AI and machine learning play a crucial role in improving the accuracy and efficiency of house price predictions, benefiting both buyers and sellers in the real estate market by providing valuable insights into property values[14].

1.1.1 Challenges:

- **Data Quality and Quantity**: Ensuring datasets are comprehensive, accurate, and up-to-date.
- **Model Interpretability**: Understanding how models arrive at their predictions, especially for stakeholders like real estate agents and homeowners.
- **Scalability**: Handling large datasets and ensuring models can process data efficiently.
- **Ethical Considerations**: Ensuring fairness and transparency in predictions, especially regarding factors like neighbourhood demographics and socioeconomic data.
- **Economic Factors**: Housing prices are influenced by economic indicators such as interest rates, GDP growth, and unemployment rates, which may require incorporating external data sources and economic models.
- **Local Market Trends**: Real estate markets vary significantly by location, and capturing local market dynamics (e.g., neighbourhood trends, zoning regulations) is crucial but challenging[10].
- **Fairness and Bias**: Ensuring models do not exhibit biases against certain demographic groups or neighbourhoods, which could perpetuate inequalities.
- **Privacy Concerns**: Handling sensitive data (e.g., personal information about homeowners) in compliance with regulations such as GDPR or CCPA.
- **Transparency**: Providing explanations for predictions and ensuring users understand the limitations and uncertainties associated with AI-driven predictions[9].

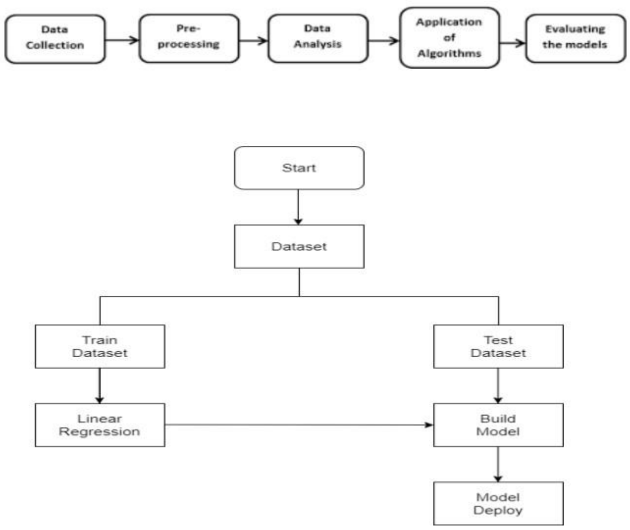

Figure 1: Block diagram

1.2 Proposed System

Proposing a robust system for house price prediction using AI and machine learning involves designing a comprehensive framework that addresses key challenges while leveraging advanced techniques for accurate predictions. Here's a structured outline for a proposed system:

1. **Data Collection and Preprocessing**

Data Sources: Collect data from various sources such as real estate listings, historical sales data, property characteristics, economic indicators, and demographic information.
Data Cleaning: Handle missing values, outliers, and inconsistencies in the data.
Feature Engineering: Create relevant features from raw data (e.g., age of property, price per square foot, proximity to amenities) to enhance predictive power[17].

2. **Feature Selection and Dimensionality Reduction**

Feature Selection: Use techniques like correlation analysis, feature importance from machine learning models, and domain knowledge to select the most informative features[2].
Dimensionality Reduction: Apply techniques such as Principal Component Analysis (PCA) or feature selection algorithms to reduce the number of features while preserving information.

3. **Model Selection and Training**

Regression Models: Experiment with various regression algorithms such as Linear Regression, Ridge Regression, Lasso Regression, and Elastic Net Regression.
Ensemble Methods: Utilize ensemble techniques like Random Forest, Gradient Boosting Machines (GBM), and XGBoost to capture complex relationships and improve prediction accuracy.
Deep Learning: Explore deep neural networks, such as feedforward networks or convolutional neural networks (CNNs), for learning intricate patterns in data (especially for image or spatial data).

4. **Model Evaluation and Optimization**

Cross-Validation: Use techniques like k-fold cross-validation to assess model performance and ensure generalizability.

Hyperparameter Tuning: Optimize model parameters using techniques like grid search or Bayesian optimization to improve performance metrics (e.g., RMSE, MAE)[3].

5. **Deployment and Integration**

API Development: Develop APIs using frameworks like Flask or Django to serve predictions based on user inputs (property features).

Scalability: Deploy models on cloud platforms (e.g., AWS, Google Cloud) for scalability and accessibility.

Real-Time Updates: Implement mechanisms for updating models with new data to ensure predictions remain accurate over time.

6. **Monitoring and Maintenance**

Performance Monitoring: Continuously monitor model performance and drift in data distributions to trigger retraining when necessary.

Model Explainability: Provide insights into how the model arrives at predictions to enhance transparency and trust among users and stakeholders.

Security and Compliance: Ensure compliance with data privacy regulations (e.g., GDPR, CCPA) and ethical guidelines in handling sensitive information.

7. **User Interface and Visualization**

Dashboard Development: Create interactive dashboards using tools like Plotly or Tableau to visualize trends, predictions, and model performance metrics. By implementing such a proposed system, stakeholders in the real estate industry can benefit from more accurate and transparent house price predictions, facilitating informed decision-making and enhancing overall market efficiency

1.2.1 Advantages:

House price prediction using AI and machine learning (ML) offers several advantages over traditional methods, providing more accurate, efficient, and data-driven insights for stakeholders in the real estate industry. Here are the key advantages:

- **Improved Accuracy**:

AI and ML models can analyze vast amounts of data, including property characteristics, location data, economic indicators, and historical sales data, to predict house prices more accurately[18].

These models can capture complex relationships and non-linearities in data that traditional statistical methods might miss, leading to more precise predictions.

- **Better Decision Making**:

Real estate professionals, buyers, and sellers can make informed decisions based on reliable predictions generated by AI and ML models.

Predictions help in setting competitive listing prices, negotiating offers, and making investment decisions, thereby optimizing outcomes.

- **Automation and Efficiency**:

Automation of data collection, preprocessing, model training, and prediction deployment reduces manual effort and saves time for real estate agents and analysts.

Predictive models can handle large datasets efficiently, providing quick insights into market trends and property valuations.

- **Handling Complex Data**:

AI and ML techniques excel at handling high-dimensional and heterogeneous data types commonly found in real estate datasets.

They can incorporate spatial data (e.g., proximity to schools, crime rates) and temporal trends (e.g., seasonal variations, economic cycles) effectively into predictions.

- **Scalability**:

ML models deployed on cloud platforms can scale easily to handle large volumes of data and increasing computational demands.

This scalability ensures that models can adapt to changes in data and market conditions over time without compromising performance.

- **Support for Real-Time Decision Making:**

Real-time prediction capabilities enable stakeholders to react promptly to market changes and make timely decisions[1].

This agility is particularly valuable in competitive real estate markets where timing can significantly impact outcomes.

LITERATURE REVIEW

1. Real Estate Price Prediction with Regression and Classification

This paper, house costs are anticipated utilizing logical factors that cover numerous parts of private houses. House costs are anticipated with different relapse procedures Lasso, Ridge, SVM relapse, and Random Forest are just a few examples. According to this article, the best-performing model for a relapse issue is SVR with Gaussian bit, which has an RMSE of 0.5271. in any case, representation for SVR was troublesome because of its high dimensionality. As per its examination, residing region rectangular feet, the material of the rooftop, and locality have the best measurable importance in foreseeing a family's deal cost[23].

2. An SVR based forecasting approach for real estate price prediction
The help vector machine (SVM) has been applied successfully to order, bunch, and gauge. The support vector relapse (SVR) method is proposed in this paper to estimate land costs in China. The point of this newspaper was to analyze the attainability of SVR in land cost expectation[22]. The test results were determined because of the mean outright mistake (MAE), the mean outright rate mistake (MAPE) and the root mean squared error (RMSE), as well as the SVR-based method was an effective device to gauge land costs.
3. House Price Prediction Using Machine Learning Algorithms The demand for renting and owning homes has increased as a result of increased urbanisation. As a result, figuring out a better technique to compute property prices that truly represent market prices has become a popular issue.[19] The research focuses on applying machine learning methods such as simple linear regression (SLR), multiple linear regression (MLR), and neural networks to properly determine the house price (NN). The algorithm with the lowest Mean Square Error (MSE) is picked as the best for estimating the price of a property. This will assist both sellers and buyers in determining the optimal price for a home.
4. House Price Prediction Using Machine Learning and Neural Networks In this paper means to make assessments in light of each essential boundary that is considered while deciding the cost. This model involved different relapse strategies in it is pathway, and the outcomes are still up in the air through single strategy instead, it's the weighted average of different procedures to give the most dependable outcomes. The outcomes demonstrated that this method yields the least mistake and most extreme precision than separate calculations useful.
5. Vision-based real estate price estimation The subject of automated calculation of market prices for properties has gotten a lot of attention with the development of online real estate database businesses like Zillow, Trulia, and Redfin. Several real estate websites use a proprietary methodology to offer such estimations. Although these predictions are frequently close to the actual sale prices, they can be quite incorrect in other circumstances. The inside and external look of a home are important aspects that are not taken into account when creating computerized value estimations. The influence of a house's aesthetic qualities on its market value is examined in this research. We build a method for predicting the luxury level of real estate photographs by using deep convolutional neural networks on a huge dataset of photos of property interiors and exteriors[6]. We also create a new framework for automatic value evaluation that incorporates the above photographs as well as property parameters such as size, proposed price, and number of bedrooms. Finally, we demonstrate that our proposed technique for price estimation surpasses Zillow's estimations by applying it to a fresh dataset of real estate photographs and information.
6. Traditional Features: Most studies utilize traditional property features such as size, number of bedrooms/bathrooms, location (via coordinates or ZIP code), and historical sales data.
Advanced Features: Some studies incorporate more complex features like proximity to amenities (schools, parks), crime rates, economic indicators (interest rates, GDP), and neighborhood characteristics derived from geospatial data.
7. Data Quality: Ensuring data accuracy, completeness, and consistency across diverse sources remains a challenge. **Model** Interpretability: Despite advancements, interpreting complex AI and ML models for real estate stakeholders remains a concern[10]. Ethical Considerations: Addressing fairness and bias issues, particularly in automated decision-making processes, is gaining attention.
8. The literature on house price prediction using AI and ML underscores the field's evolution from traditional statistical models to sophisticated machine learning techniques. Ongoing research focuses on improving model accuracy, interpretability, and ethical considerations to meet the diverse needs of stakeholders in the dynamic real estate market. Future directions include integrating more diverse data sources, enhancing model transparency, and addressing regulatory challenges to further advance the field.
9. Spatial Analysis: Studies incorporate spatial autocorrelation, spatial weights matrices, and distance metrics to account for spatial dependencies and neighbourhood effects. Temporal Dynamics: Time series models and seasonal adjustments are used to capture seasonal variations and trends in housing market data.

II. METHODOLOGY

- STUDIED ALGORITHMS:

During the time spent encouraging this model, different backslide computations were thought about. Straight backslide, Multiple straight backslide, Decision Tree Regressor, and KNN are all examples of machine learning techniques. were attempted upon the planning dataset. In any case, the decision tree regressor gave the most raised accuracy to the extent that expecting the house costs[7]. The decision to pick the computation particularly depends on the angles and the type of data in the data that was used For our dataset, the decision tree computation is the best option.

- REGRESSOR FOR DECISION TREE:

The decision tree regressor recognizes quality components and trains a model like a tree to forecast data in the future to provide a massive result. The highest significance and minimum significance of a chart are gained by the decision tree regressor, which then separates the data as demonstrated by the system. Network Search CV is a strategy for overseeing limit tuning that will beneficially create and study a model for each mix of calculation limits exhibited in a cross-section.

System Lookup In this calculation, CV is utilized to determine the optimal impetus for max significance, which is then used to construct the decision tree

SYSTEM DESIGN AND ARCHITECTURE

Step 1: **Collection of data**:

Information handling strategies and cycles are various. We gathered the information for Mumbai's land properties from different land sites[4]. The information would have traits, for example, Location, cover region, developed region, age of the property, postal district, and so forth We should gather the quantitative information which is organized and ordered. Information assortment is required before any sort of AI research is completed. Dataset legitimacy is an unquestionable requirement in any case it is a waste of time to break down the information.

	CRIM	ZN	INDUS	CHAS	NOX	RM	AGE	DIS	RAD	TAX	PTRATIO	B	LSTAT	price
0	0.00632	18.0	2.31	0.0	0.538	6.575	65.2	4.0900	1.0	296.0	15.3	396.90	4.98	24.0
1	0.02731	0.0	7.07	0.0	0.469	6.421	78.9	4.9671	2.0	242.0	17.8	396.90	9.14	21.6
2	0.02729	0.0	7.07	0.0	0.469	7.185	61.1	4.9671	2.0	242.0	17.8	392.83	4.03	34.7
3	0.03237	0.0	2.18	0.0	0.458	6.998	45.8	6.0622	3.0	222.0	18.7	394.63	2.94	33.4
4	0.06905	0.0	2.18	0.0	0.458	7.147	54.2	6.0622	3.0	222.0	18.7	396.90	5.33	36.2

Figure 2:Sample Data

we have taken first 5 index/instance of data and printed them. In total there are 506 rows of data from the dataset , of which we have printed first 5 rows using head() function. There are 14 columns in total, i.e, 13 colums containing data of the place, and the 14th column is the target column which contains the house prices.

Here the data parameters are explained as follows:

```
1. CRIM      per capita crime rate by town
2. ZN        proportion of residential land zoned for lots over
             25,000 sq.ft.
3. INDUS     proportion of non-retail business acres per town
4. CHAS      Charles River dummy variable (= 1 if tract bounds
             river; 0 otherwise)
5. NOX       nitric oxides concentration (parts per 10 million)
6. RM        average number of rooms per dwelling
7. AGE       proportion of owner-occupied units built prior to 1940
8. DIS       weighted distances to five Boston employment centres
9. RAD       index of accessibility to radial highways
10. TAX      full-value property-tax rate per $10,000
11. PTRATIO  pupil-teacher ratio by town
12. B        1000(Bk - 0.63)^2 where Bk is the proportion of blacks
             by town
13. LSTAT    % lower status of the population
14. MEDV     Median value of owner-occupied homes in $1000's
```

Figure 3: explanation of data parameters

Step 2: **Data Preprocessing and Exploration**

Data preprocessing is the most well-known approach to cleaning our instructive file. There might be missing characteristics or irregularities in the dataset[5]. Data cleansing can help with these issues. describe the data in such a way so that both people and machine find it easy to understand the given data . In order to do this we use the describe() function.

	CRIM	ZN	INDUS	CHAS	NOX	RM	AGE	DIS	RAD	TAX	PTRATIO	B	LSTAT	price
count	506.000000	506.000000	506.000000	506.000000	506.000000	506.000000	506.000000	506.000000	506.000000	506.000000	506.000000	506.000000	506.000000	506.000000
mean	3.613524	11.363636	11.136779	0.069170	0.554695	6.284634	68.574901	3.795043	9.549407	408.237154	18.455534	356.674032	12.653063	22.532806
std	8.601545	23.322453	6.860353	0.253994	0.115878	0.702617	28.148861	2.105710	8.707259	168.537116	2.164946	91.294864	7.141062	9.197104
min	0.006320	0.000000	0.460000	0.000000	0.385000	3.561000	2.900000	1.129600	1.000000	187.000000	12.600000	0.320000	1.730000	5.000000
25%	0.082045	0.000000	5.190000	0.000000	0.449000	5.885500	45.025000	2.100175	4.000000	279.000000	17.400000	375.377500	6.950000	17.025000
50%	0.256510	0.000000	9.690000	0.000000	0.538000	6.208500	77.500000	3.207450	5.000000	330.000000	19.050000	391.440000	11.360000	21.200000
75%	3.677083	12.500000	18.100000	0.000000	0.624000	6.623500	94.075000	5.188425	24.000000	666.000000	20.200000	396.225000	16.955000	25.000000
max	88.976200	100.000000	27.740000	1.000000	0.871000	8.780000	100.000000	12.126500	24.000000	711.000000	22.000000	396.900000	37.970000	50.000000

Figure 4: Describing the data

Counts refers to the number of instances of data in each column i.e 506 since there are 506 rows of data for each column Mean refers to mean value of data in given column[11]. Std means the standard value i.e the most common value in given set of data for a particular column. Next we try to understand the correlation between the different values, in order to do that, the best way is by using heat map. Heat map is a representation of data in the form of a map or diagram in which data values are represented as colors. Correlation is a statistical measure that expresses the extent to which two variables are linearly related (meaning they change together at a constant rate)

Step 3: **Feature Selection and Dimensionality Reduction:**
Feature Selection: Select the most relevant features using techniques like correlation analysis and feature importance. Dimensionality Reduction: Apply methods such as Principal Component Analysis (PCA) to reduce the number of features while retaining important information[8].
HEATMAP – for better understanding of which place is best suited for individual personal preference based on given dataset. This uses correlation concept

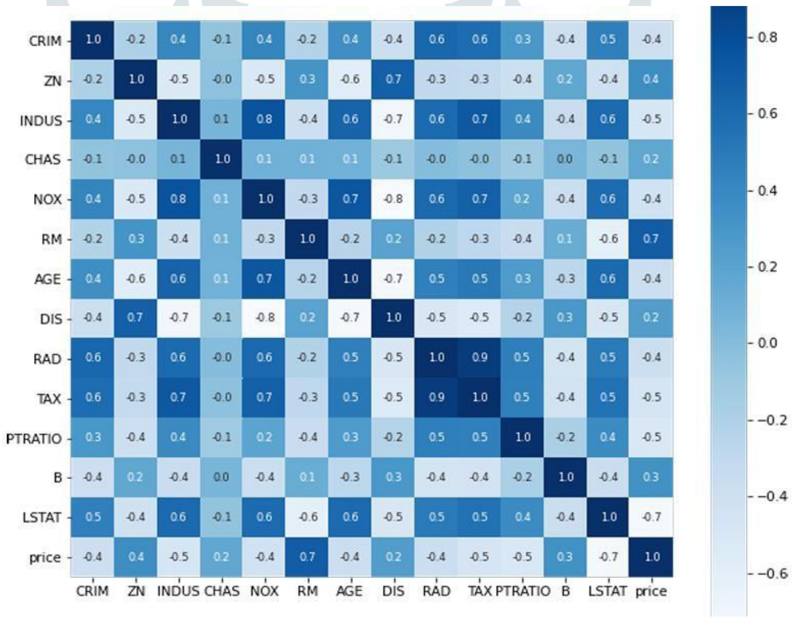

Figure 5: Heat map

Step 4: **Model Selection and Training:**
Model Selection: Choose appropriate regression models (e.g., Linear Regression, Random Forest) based on the problem requirements and data characteristics. Model Training: Train the selected models using the training dataset and evaluate their performance**.** Evaluation Metrics: Evaluate model performance using metrics such as Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and R-squared.
Next we split our data into variables x and y , in order to train our model to predict data

	CRIM	ZN	INDUS	CHAS	NOX	RM	AGE	DIS		
0	0.00632	18.0	2.31	0.0	0.538	6.575	65.2	4.0900	1.0	296.0
1	0.02731	0.0	7.07	0.0	0.469	6.421	78.9	4.9671	2.0	242.0
2	0.02729	0.0	7.07	0.0	0.469	7.185	61.1	4.9671	2.0	242.0
3	0.03237	0.0	2.18	0.0	0.458	6.998	45.8	6.0622	3.0	222.0
4	0.06905	0.0	2.18	0.0	0.458	7.147	54.2	6.0622	3.0	222.0
..	...	...	...	...	...	...	...	...	...	...
501	0.06263	0.0	11.93	0.0	0.573	6.593	69.1	2.4786	1.0	273.0
502	0.04527	0.0	11.93	0.0	0.573	6.120	76.7	2.2875	1.0	273.0
503	0.06076	0.0	11.93	0.0	0.573	6.976	91.0	2.1675	1.0	273.0
504	0.10959	0.0	11.93	0.0	0.573	6.794	89.3	2.3889	1.0	273.0
505	0.04741	0.0	11.93	0.0	0.573	6.030	80.8	2.5050	1.0	273.0

	PTRATIO	B	LSTAT
0	15.3	396.90	4.98
1	17.8	396.90	9.14
2	17.8	392.83	4.03
3	18.7	394.63	2.94
4	18.7	396.90	5.33
..	...	...	...
501	21.0	391.99	9.67
502	21.0	396.90	9.08
503	21.0	396.90	5.64
504	21.0	393.45	6.48
505	21.0	396.90	7.88

[506 rows x 13 columns]

0	24.0
1	21.6
2	34.7
3	33.4
4	36.2
	...
501	22.4
502	20.6
503	23.9
504	22.0
505	11.9

Name: price, Length: 506, dtype: float64

Figure 6: Training the data

Here the variable x contains the value of the first 13 columns i.e the parameters that are required for calculating and predicting the house prices. The variable y contains the 14th column values which are the house prices.

First we predict the values in y using the values in x . Then we compare the actual prices and predicted prices by using scatter plot. Then we find the r square error and mean square error between them . If the errors is less enough then we proceed for testing of the model since the training phase is over. If the error is large , then we use optimizers like Adam, and repeat drop and fitting process for a set number of epochs to reduce the error[12]. The r square error or mean square error for good accuracy of the model in predicting the data is indicated numerically also. A model is good if these error values are less then 5. Then during testing process we predict the future house prices using present and past data parameters of houses in an location. Then we plot this graphically as a house price over time graph.

Step 5: **Deployment and Integration:**

For training the model , the error needs to be minimum for greater accuracy of model. The error between the actual and predicted price is plotted graphically using scatter plot. Here we can see that error is minimum since the data points of actual and predicted value are close to each other

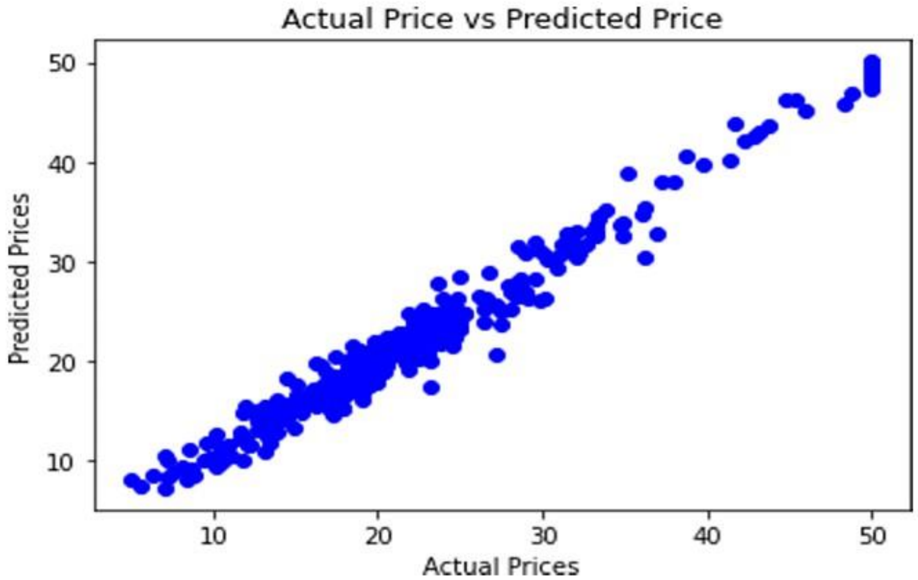

Figure 7: Predicted prices

III. RESULTS

Leveraging these results, we computed an aggregate ranking of feature importance by assigning weights based on the accuracy of each algorithm. our analysis reveals that the three most influential features affecting house prices in our dataset are in decreasing order of significance: "lat", "waterfront" and the "grade" of the house. Thus the machine learning model to predict the house price based on given dataset is executed successfully using xg regressor (a upgraded/ slighted boosted form of regular linear regression, this gives lesser error). This model further helps people understand whether this place is more suited for them based on heatmap correlation. It also helps people looking to sell a house at best time for greater profit. Any house price in any location can be predicted with minimum error by giving appropriate dataset.

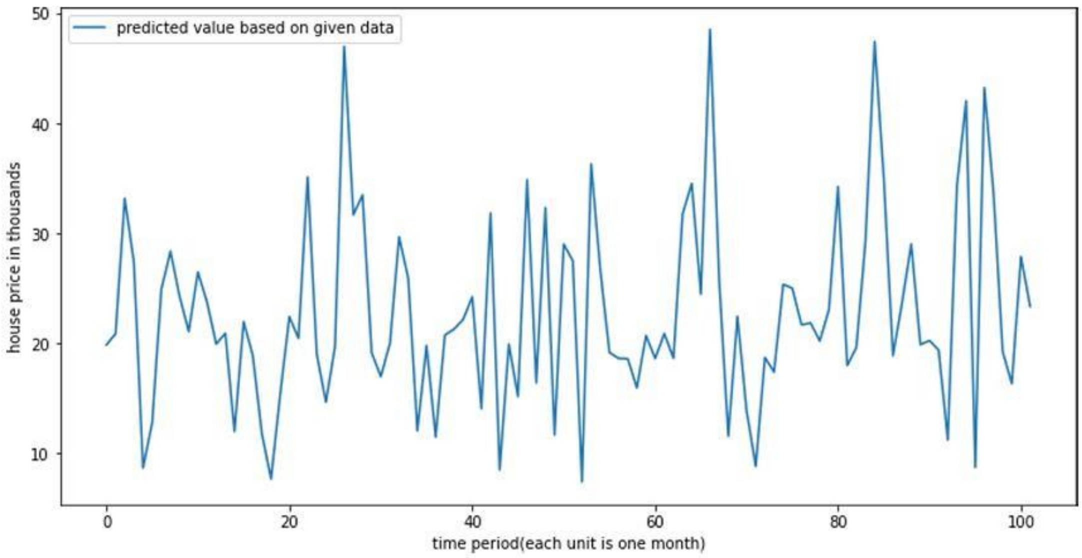

Figure 8: PREDICTED VALUE OF HOUSE PRICE BASED ON TEST SAMPLE DATA

IV. DISCUSSIONS

This system's aim was to make a model that can give us a good house price prediction based on other variables. They used the Linear Regression for Ames dataset and hence it gave good accuracy. The house price prediction project had two modules namely, Admin and the User. Admin can add location and view the location. Admin had the authority to add density on the basis of per unit area. Users can view the location and see the predicted housing price for that particular location.

IX. CONCLUSION

This article examines three different house price prediction models and compares their accuracy. They are linear regression, random forest and network neural. This article explores the impact of each feature on price at same time. In the experiment, we can see that XGBoost has the highest accuracy, linear regression has the lowest. Therefore, for predicting house prices, considering accuracy, the XGBoost method would be more suitable. In comparison with neural networks, random forests and XGBoost although linear regression has the lowest accuracy, its accuracy is relatively close to the other three methods.

Considering that the linear regression method is relatively simple and straightforward, when time is limited and accuracy requirements are not so high, the linear regression method will be an option. More research is needed in this part. Regarding the importance of features, we can see that the three most influential features affecting house prices in our dataset are lat, waterfront and grade in the experiment. In other words, for the houses we choose, their latitude, waterfront condition and house grade can most affect their prices. This result may hold true for other houses and be helpful in their price predictions. However, more research is needed to confirm this..

X. FUTURE SCOPE

The future scope of house price prediction using ML and AI is promising, driven by advancements in technology, data availability, and evolving consumer demands. Here are several key aspects shaping the future of this field:

1. **Improved Accuracy**: As algorithms become more sophisticated and datasets richer, the accuracy of house price predictions is expected to increase. ML techniques such as ensemble methods, deep learning, and reinforcement learning are being applied to capture complex relationships in data more effectively.
2. **Integration of Big Data**: ML algorithms benefit from large and diverse datasets. Integration of big data sources including satellite imagery, IoT devices (for smart homes), and social media sentiment analysis can provide deeper insights into property values[20].
3. **Risk Assessment**: ML techniques can enhance risk assessment in real estate investments by predicting market volatility, identifying potential property value fluctuations, and evaluating economic indicators that impact housing prices[21].

4. **Ethical Considerations and Fairness**: There's growing emphasis on developing AI models that are transparent, fair, and unbiased. Addressing issues of bias in data and algorithms is crucial to ensure equitable predictions and avoid perpetuating inequalities in housing markets.

5. **Predictive Analytics for Policy Making**: Governments and urban planners can leverage ML predictions to formulate housing policies, urban development strategies, and infrastructure investments that are responsive to future housing needs and market dynamic

6. **Cross-Domain Integration**: ML models in house price prediction are increasingly integrating insights from related domains such as urban planning, environmental impact assessments, and infrastructure development, offering a holistic view of property valuation[24].

7. **Emerging Technologies**: Advancements in augmented reality (AR) and virtual reality (VR) may enhance property valuation by offering immersive property tours and visualizations, enriching the decision-making process for buyers and investors.

8. **Personalization**: ML allows for personalized predictions based on individual preferences and behaviors. Predictive models can account for specific buyer profiles, market segments, and evolving consumer trends to tailor recommendations.

9. **Automated Valuation Models (AVMs)**: AVMs, powered by AI, are becoming more prevalent in real estate transactions. These models use ML to automatically assess property values based on real-time data inputs, reducing reliance on traditional appraisal methods.

XI. ACKNOWLEDGEMENT

Pinnamaraju.p.t. priya working as Assistant professor in Master of computer application (MCA) Sankethika vidya parishad engineering college, Visakhapatnam, Andhra Pradesh. with 6 years of experience in master of computer applications(MCA),Accredited by NAAC. with her area of interests in c, computer organization, software engineering, IOT

Ambati Jemimah is perusing her final semester MCA in Sankethika Vidya Parishad Engineering College, accredited with A grade by NAAC, affiliated by Andhra University and approved by AICTE. With interest in Artificial intelligence Ambati Jemimah has taken up his PG project on House price prediction and published the paper in connect to the project under the guidance of Pinnamaraju.p.t.priya as Assistant professor

References

Book references:

1. Real Estate Price Prediction with Regression and Classification, CS 229 Autumn 2016 Project Final Report
2. Gongzhu Hu, Jinping Wang, and Wenying Feng Multivariate Regression Modelling for Home Value Estimates with Evaluation using Maximum Information Coefficient
3. Byeonghwa Park , Jae Kwon Bae (2015). Using machine learning algorithms for housing price prediction , Volume 42, Pages 2928-2934 [4] Douglas C. Montgomery, Elizabeth A. Peck, G. Geoffrey Vining, 2015. Introduction to Linear Regression Analysis.
4. Iain Pardoe, 2008, Modelling Home Prices Using Realtor Data
5. Aaron Ng, 2015, Machine Learning for a London Housing Price Prediction Mobile Application
6. Wang, X., Wen, J., Zhang, Y.Wang, Y. (2014). Real estate price forecasting based on SVM optimized by PSO. Optik-International Journal for Light and Electron Optics, 125(3), 14391443
7. R. J. Shiller, “Understanding recent trends in house prices and home ownership,” National Bureau of Economic Research, Working Paper 13553, Oct. 2007. DOI: 10.3386/w13553. [Online]. Available: http://www.nber.org/papers/w13553.
8. D. Belsley, E. Kuh, and R. Welsch, Regression Diagnostics: Identifying Influential Data and Source of Collinearity. New York: John Wiley, 1980.
9. J. R. Quinlan, “Combining instance-based and model-based learning,” Morgan Kaufmann, 1993, pp. 236–243.
10. S. C. Bourassa, E. Cantoni, and M. Hoesli, “Predicting house prices with spatial dependence: a comparison of alternative methods,” Journal of Real Estate Research, vol. 32, no. 2, pp. 139–160, 2010. [Online] Available:http://EconPapers.repec.org/RePEc:jre:issued:v:32: n:2:2010:p:139-160.
11. S. C. Bourassa, E. Cantoni, and M. E. Hoesli, “Spatial dependence, housing submarkets and house price prediction,” eng, 330; 332/658, 2007, ID: unige:5737.[Online]. Available: http:// archive - ouverte. unige. ch/unige:5737
12. Pow, Nissan, Emil Janulewicz, and L. Liu. "Applied Machine Learning Project 4 Prediction of real estate property prices in Montréal." (2014).
13. Limsombunchai, Visit. "House price prediction: hedonic price model vs. artificial neural network."New Zealand Agricultural and Resource Economics Society Conference. 2004.
14. Arietta, SeanM., etal."City forensics Using visual rudiments to prognosticatenon-visual megacity attributes."IEEE deals on visualization and computer plates20.12 (2014) 2624-2633.
15. Yu, H., andJ. Wu." Real estate price vaticination with retrogression and bracket CS 229 Afterlife 2016 Project Final Report 1 – 5." (2016).
16. Banerjee, Debanjan, and Suchibrota Dutta."Predicting the casing price direction using machine literacy ways."2017 IEEE International Conference on Power, Control, Signals and Instrumentation Engineering (ICPCSI). IEEE, 2017.
17. Sampathkumar,V., Santhi,M.H., & Vanjinathan,J. (2015). Vaticinating the land price using statistical and neural network software. Procedia Computer Science, 57, 112-121.
18. Pow, Nissan, Emil Janulewicz, and Liu Dave Liu."Applied Machine Learning Project 4 Vaticination of real estate property prices in Montréal." Course design, Presentation-598, Fall/ 2014, McGill University (2014).
19. N.N. Ghosalkar andS.N. Dhage, "Real Estate Value Prediction Using Linear Retrogression, "2018 Fourth International Conference on Computing Communication Control and Robotization (ICCUBEA), Pune, India, 2018, pp. 1-5.
20. Nihar Bhagat, Ankit Mohokar, Shreyash Mane" House Price Soothsaying using Data Mining"International Journal of Computer Applications,
21. Li, Li, and Kai-Hsuan Chu." Vaticination of real estate price variation grounded on profitable parameters."2017 International Conference on Applied System Innovation (ICASI). IEEE, 2017.
22. A. Varma etal., “ House Price Prediction Using Machine Learning And Neural Networks,” 2018 Second International Conference on Inventive Communication and Computational Technologies, pp. 1936 – 1939, 1936.
23. Manjula, R., etal." Real estate value vaticination using multivariate retrogression models." Accoutrements Science and Engineering Conference Series.Vol. 263. No. 4. 2017.
24. Lakshmi, B.N., and G.H. Raghunandhan."An abstract overview of data mining."2011 National Conference on Inventions in Emerging Technology. IEEE, 2011.