



Comparing Machine Learning Algorithms Performance in House Price Prediction

Business Analytics

REMOVED PROF NAME

Project Management

BUSA 521-01W

Group Members: -

Arbaz

Sri

Ravi

Karthikeya

Hitesh



Table of Contents

Project Planning	4
Background Problem	5
Objective of the Report	4
Research Questions	5
Potential Contributions	6
Theoretical Background	7
References	9
Project Implementation Report	10
Abstract & Introduction	10
Hypothesis	10
Methodology	11
Data Collection & Analysis	12
Appendix	13
Work Breakdown Structure	14
Risk Breakdown Structure	14
Final Report	
Executive Summary	<u>155</u>
Hypothesis Development	<u>166</u>
Hypothesis_1	<u>166</u>
Hypothesis_2	<u>17</u>
Hypothesis_3	<u>177</u>
Hypothesis_4	<u>188</u>
Hypothesis_5	<u>188</u>
Methodology	19
Results and Discussions	<u>222</u>
Random forest model	<u>23</u>



Gradient Booster algorithm.....	233
Decision tree	233
Hypothetical testing	2424
Conclusions.....	255
References.....	255
Appendices.....	266
Time-Series Analysis	Error! Bookmark not defined.8
Linear Regression Models	265
Evaluation Metrics	266
Random Forest	277
Gradient Booster Algorithm	277
Decision Tree	278



Project Planning

Background problem

Day to day house prices in urban areas increasing rapidly, and people face a lot of challenges in identifying a house with a particular price range in a particular region. The development of technology gives a hand in predicting the price of a house in a particular location based on the price history. As well, it helps in predicting the price of a house in the future based on the price increment of houses in past. Manually analysing the dataset and understanding it to make the prediction on house price is a challenging task. Machine Learning (ML) models can learn from the data which is available to predict the price increment based on the history. ML models helps in making the accurate prediction about the house price.

Dataset: <https://www.kaggle.com/datasets/farhankarim1/usa-house-prices>

Objectives of the project

The objectives of the project are as follow,

- To identify the dataset which contains various properties to extract features from it to train the model.
- To clean the dataset by removing the duplicate data, and noisy data.
- To develop a machine learning model to train it with the train data to predict the house price.
- To evaluate the model accuracy and loss in the prediction using the test data

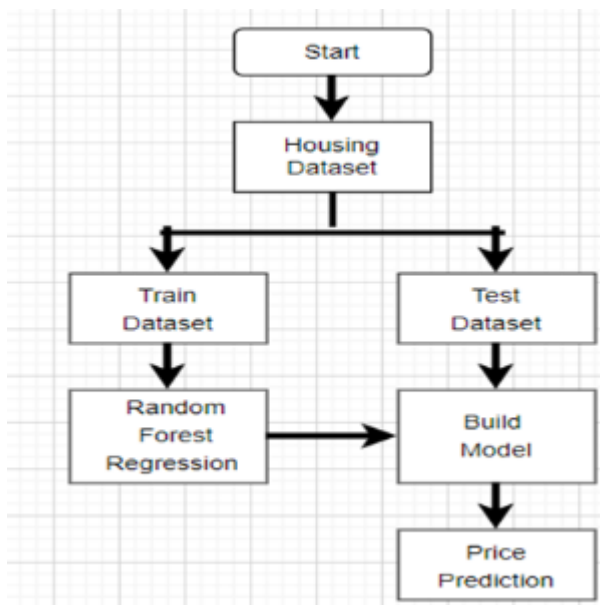
Research questions

- How do machine learning algorithms performs in predicting the house prices in future based on the available data?
- Evaluate which machine learning algorithm achieves less error in the prediction?



Potential contributions

This project aims in analysing the selected USA data to predict the house price because day to day in the US the house prices are increasing based on the location and type of the property. The dataset which is identified in this project work contains number of attributes including avg.Area income, Avg AreaHouse Age, Avg.Area Number of Rooms, Avg. Area Number of Bedrooms, Area Population, Price, and Address.

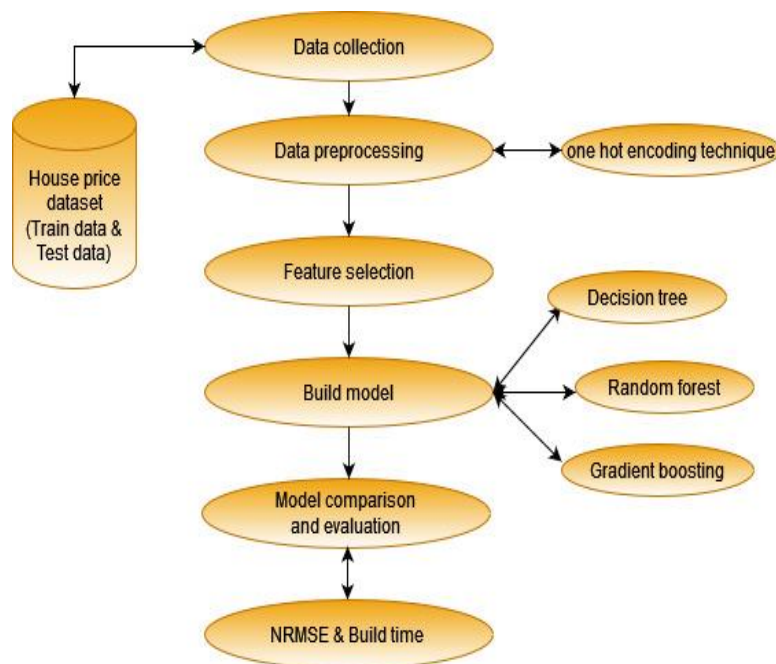


By analysing the dataset gives clear idea about the features to extract it from the dataset and to train the model. At the model development the data will be divided into two category such as training and testing. Here training data is allocated 80% from the original dataset and remaining 20% data will be allocated to test the model. So finally, developing a random forest regression model the prediction accuracy will be evaluated.

Once the model development is done then the input data will be given to the model to train it. Based on the learning outcome such as model loss and accuracy. If the model loss is high and

accuracy is very less than the model will be trained with more number of iterations to extract the unique features from the dataset and to learn to predict more accurately.

Theoretical background/Related Study



The above diagram shows how the data extraction will be done to perform the data pre-processing. In the data pre-processing the null values, duplicated data, and unwanted columns from the dataset will be removed. Then the feature selection will be done to extract the unique feature from the dataset to make the algorithms to learn to predict the price in future.



References

Ahtesham, M., Bawany, N. Z. & Fatima, K., 2020. House Price Prediction using Machine Learning Algorithm - The Case of Karachi City, Pakistan. *IEEE International Arab Conference on Information technology*, pp. 1-5.

Crossman, A., 2020. *Thought Co.* [Online]
Available at: <https://www.thoughtco.com/qualitative-research-methods-3026555>
[Accessed 7 Jan 2022].

Dagar, A. & Kapoor, S., 2020. A Comparative Study on House Price Prediction. *International Journal for Modern Trends in Science and Technology*, 6(12), pp. 103-107.

Hoa, W. K., Tangb, B.-S. & Wong, S. W., 2021. Predicting property prices with machine learning algorithms. *Journal of Property Research*, 38(1), pp. 48-70.

Hunter, J. D., 2007 . Matplotlib: A 2D Graphics Environment. *Computing in Science Engineering*, 9(3), pp. 1-5.

Ippolito, P. P., 2019. *Towards Data Science.* [Online]
Available at: <https://towardsdatascience.com/feature-extraction-techniques-d619b56e31be>
[Accessed 7 Jan 2022].

Korstanje, J., 2020. *Towards Data Science.* [Online]
Available at: <https://towardsdatascience.com/gridsearch-the-ultimate-machine-learning-tool-6cd5fb93d07>

[Accessed 7 Jan 2022].



Kumar, A. V. et al., 2020. House cost prediction using data science and machine learning. *International Research Journal of Engineering and Technology*, 7(1), pp. 1786-1788.

Kumar, D., 2018. *Towards Data Science*. [Online] Available at: <https://towardsdatascience.com/introduction-to-data-preprocessing-in-machine-learning-a9fa83a5dc9d>

[Accessed 7 Jan 2022].

Kumar, H., 2018. Python Libraries, Development Frameworks and Algorithms for Machine Learning Applications. *International Journal of Engineering Research & Techno*, 7(4), pp. 210-217.

Patil, P., Shah, D., Rajput, H. & Chheda, J., 2020. House Price Prediction Using Machine Learning and RPA. *International Research Journal of Engineering and Technology*, 7(3), pp. 5560-5563.

Peng, Z., Huang, Q. & Han, Y., 2019. Model Research on Forecast of Second-Hand House Price in Chengdu Based on XGboost Algorithm. *IEEE 11th International Conference on Advanced Infocomm Technology*, pp. 168-172.

Phan, D., 2018. Housing Price Prediction using Machine Learning Algorithms: The Case of Melbourne City, Australia. *IEEE International Conference on Machine Learning and Data Engineering*, pp. 35-42.

Satish, N., Raghavendran, Rao, M. & Ch.Srinivasulu, 2019. House Price Prediction Using Machine Learning. *International Journal of Innovative Technology and Exploring Engineering*, 8(9), pp. 717-722.



Shinde, N. & Gawande, K., 2018. Valuation of house prices using predictive techniques. *International Journal of Advances in Electronics and Computer Science*, 5(6), pp. 34-40.

Sinha, A., 2020. Utilization Of Machine Learning Models In Real Estate House Price Prediction. *Amity Journal of Computational Sciences*, 4(1), pp. 18-23.

Sivasankar, B., Ashok, A. P., Madhu, G. & Fousiya, 2020. House Price Prediction. *International Journal of Computer Sciences and Engineering*, 8(7), pp. 98-102.

Thakur, A. & Satish, M., 2021. Bangalore House price prediction. *International Research Journal of Engineering and Technology*, 8(9), pp. 193-196.

Tiwari, P. & Thakur, V. S., 2020. Review on House Price Prediction through Regression Techniques. *International Journal of Scientific Progress and Research*, 73(173), pp. 62-65.

Truong, Q., Nguyen, M., Dang, H. & Mei, B., 2020. Housing Price Prediction via Improved Machine Learning Techniques. *Procedia Computer Science: 2019 International Conference on Identification, Information and Knowledge in the Internet of Things*, Volume 174, pp. 433-442.

Kaggle : <https://www.kaggle.com/datasets/farhankarim1/usa-house-prices>

Project Implementation Report



Abstract

House: It is now one of the most essential place to live. A place where a person can relax, eat, sleep. Houses are built such that an individual can live inside a personal space. An individual can cook their own food and have their privacy in a house. There are multiple database's available for houses. Here we have used the house data for Dallas city which is situated in the State of Texas in United States of America. We have used multiple sources to collect the dataset from Internet to get multiple attributes instead of using one website with limited number of attributes.

Introduction

In USA house ownership was not a big concern in early days. It is estimated that only around 65% people used to own houses in earlier days. The demand for owning the houses has been rising in since after 2020 as it has been a down fall in during that period of Covid-19. There has been a significant recovery of house prices increase. Due to Covid-19 people were afraid of selling their homes or buying new ones which impacted the prices of houses. This also caused a slowdown on building of new houses are people were not likely to add a huge amount by thinking that there will be no end to the pandemic.

As soon as the pandemic ended people started moving out of their houses and started looking for ways to invest and earn livelihood. It was more likely that the house's sold in 2021 were high by 6.2% in the whole state of Texas whereas the inventory was lowered approximately by year end at 1.2 months. There was a constant median price going above \$3,00,000 which is almost 16% more than previous year. The house price from 2021 to present date has increased at a drastic level and has broken the historical records till date at around median price of \$3,35,000 as of March 2022 that is 20% higher than the previous year.

According to Reports there has been an assumption that the house prices will drastically increase around 33% by the year 2027.

Hypothesis

Hypothesis 1: Future house prices can be predicted through machine learning algorithms with maximum accuracy.

Hypothesis 2: The house prices significantly vary based on the area of the house and the type of house

Hypothesis 3: The gradient boosting algorithm provides higher accuracy when compared to the decision tree and random forest algorithms.

The research is carried out based on the above three hypotheses and the hypothesis are expected to be true at the end of the analysis. Hypothesis 1 focuses on the overall purpose of the research.



Hypothesis 2 highlights how various factors influence the price of houses. Hypothesis 3 focuses on comparing the build model such as decision tree, random forest, and gradient boosting in which gradient boosting is expected to provide higher accuracy than the other two algorithms (Truong & Nguyen, Housing Price Prediction via Improved Machine Learning Technique, 2020).

Methodology

Research Design

Descriptive research design is followed for developing machine learning models to predict house prices. In this research Design we are using multiple attributes as variables are not required. We have used all the dependent variables so that the prediction of house price is appropriate. The multiple factors that accompany this project is the Date from 2020 until the present day which is monthly. Our research will not have variables being changed so we use other attributes like Median Price which will show the overall prices of houses. Then we have the consumer price index, or we can say CPI, it is basically used to measure the inflation of the household prices. We have used Federal Interest Rate which will show the prediction of the houses based on the loans or the mortgage according to percentage. To see the market rate, we have used Housing Supply which will indicate how many houses have been sold overall and will give a glimpse of sold houses in Dallas. Unemployment Rate is one of the crucial factor used as more number of people are unemployed then less number of houses are sold in Dallas, Texas, USA. The quantitative research method is utilized in this research because the research mainly deals with numerical data.

Data Collection

Quantitative data collection methods (Existing data) have been utilized. Publicly available data sets have been collected for this research from multiple sources. The dataset contains various attributes such as date from the year 2020 until the present data which is distributed monthly, Median Price, Consumer Price Index or CPI, Federal Interest Rate in Percentage, Housing Supply, Unemployment Rate



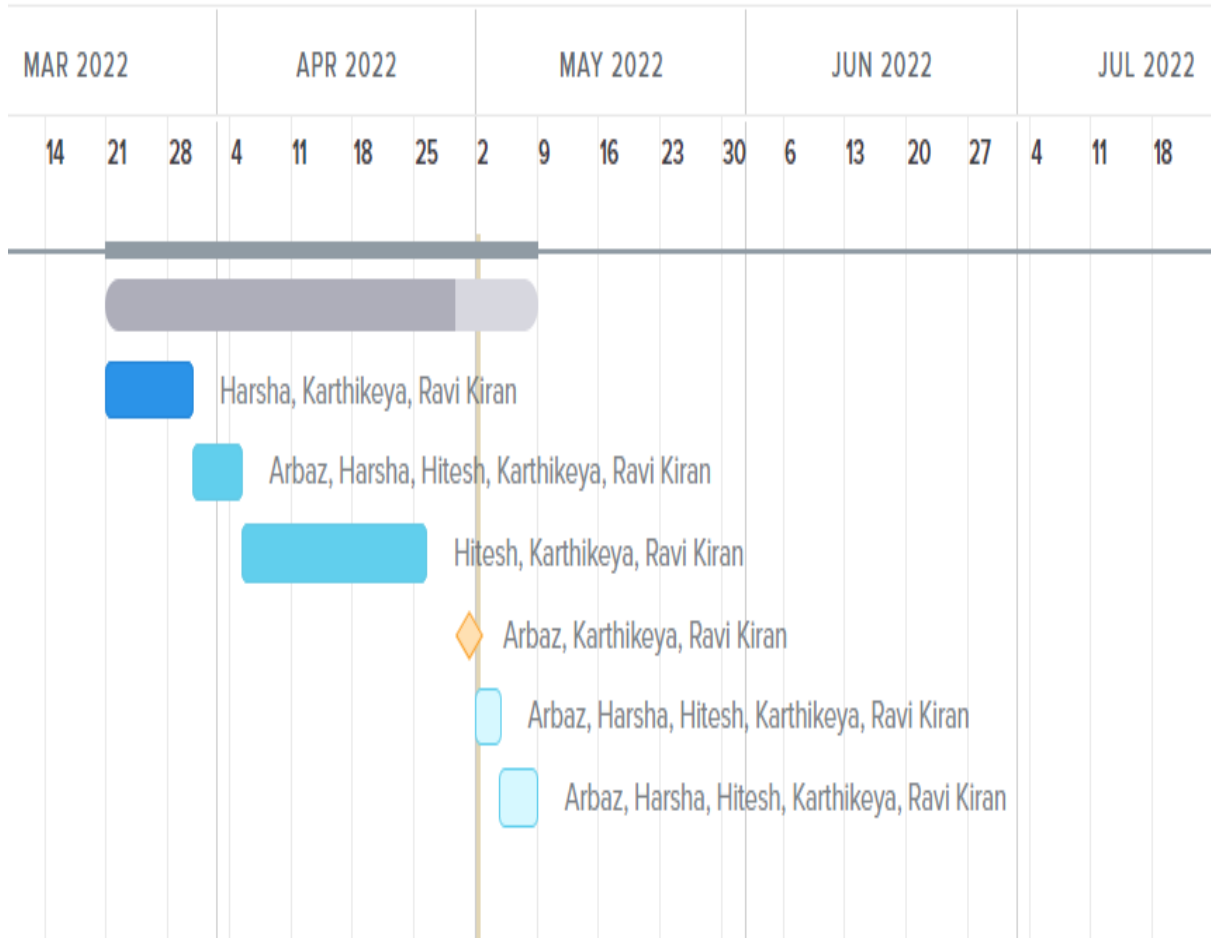
Data Analysis

Quantitative data analysis is utilized in this research for analyzing the numerical data collected from the dataset. The descriptive statistics method is used for the first level of analysis which involves summarizing the data to find patterns. The research is limited and it doesn't need to generalize to larger data (Zulkifley, et al., 2020).

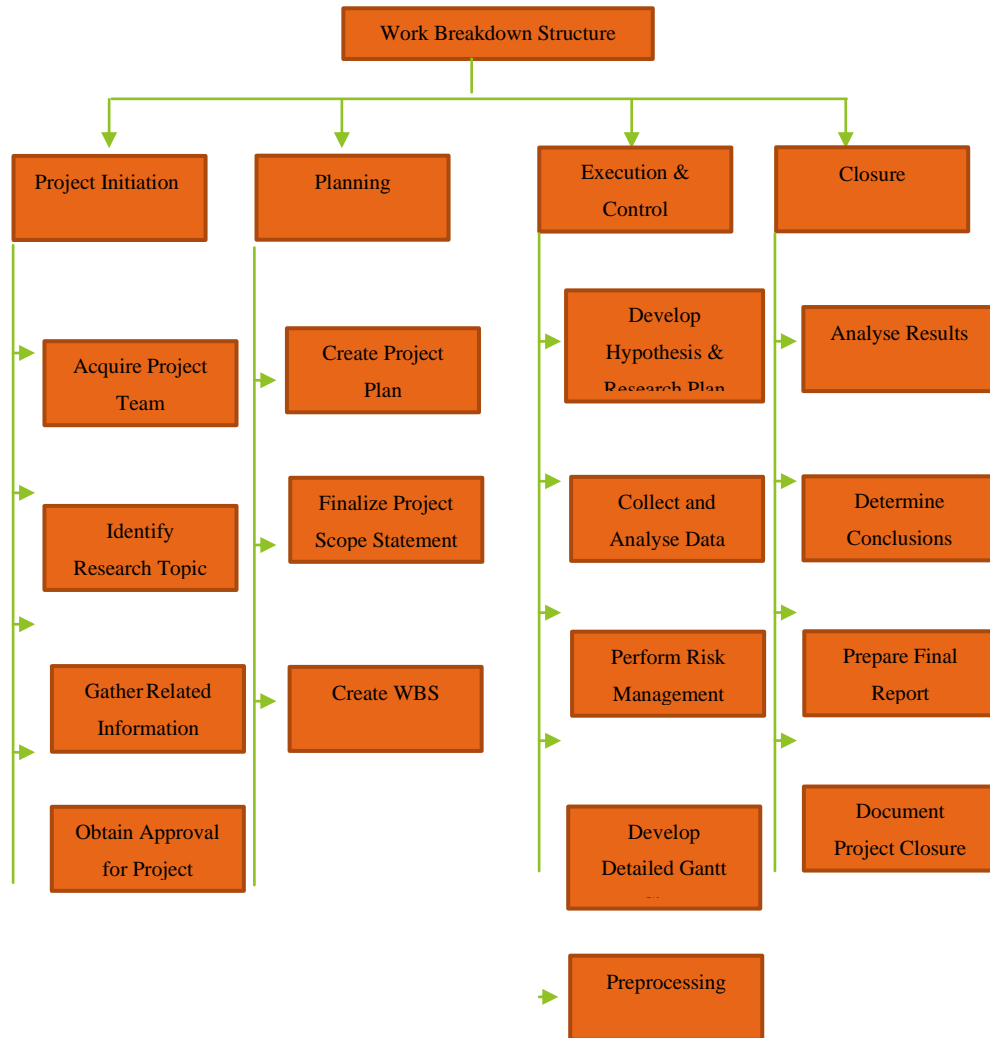
Appendix

Gantt chart

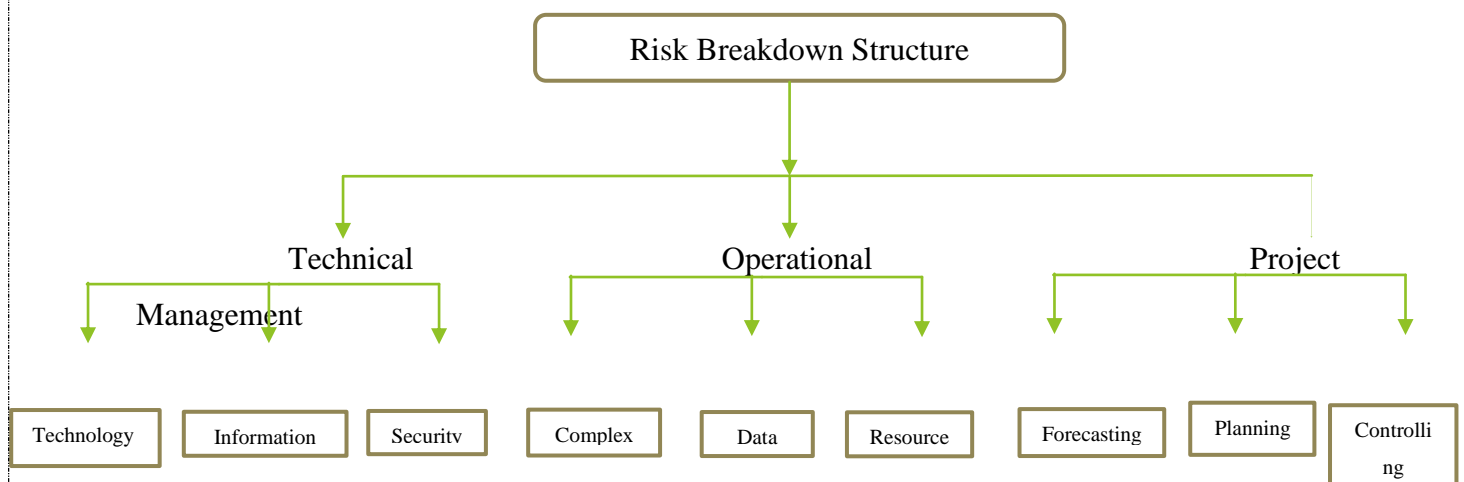
TOPIC: Comparing Machine Learning Algorithms Performance in House Pri...	
▼ Comparing Machine Learning Algorithms Performance for House Price Predic...	
Topic Selection & Approval	Harsha, Karthikeya,
Defining the aim and objective of the research and problem statement	Arbaz, Harsha, Hites
Collecting data sets and analyzing data using quantitative method	Hitesh, Karthikeya, F
Implementation of machine learning models using Python	Arbaz, Karthikeya, R
Experimenting with results by comparing each model	Arbaz, Harsha, Hites
Result Analysis & Outcome	Arbaz, Harsha, Hites



Work Breakdown Structure:



Risk Breakdown Structure:





Executive Summary

There are economic models for the real estate market where the sellers attempt to sell the houses and the buyers attempt to purchase the homes with the best prices and within their budget. But the house prediction result obtained from the traditional estimation techniques may not predict accurate result and, they are time consuming, Accurate estimation and prediction of the value of real estate is essential and considered as a significant problem for various stakeholders including house owners, buyers, agents, creditors and also the investors. As the property prices are changing rapidly with the change in market demands for each situation, it is needed for effective prediction of house price (Pathak, 2021). For our projects, Machine learning algorithm would be applied for predictive analysis of the house price prediction problems and property demands. The project aim focus on predicting the real state process by the time series forecast analysis and the various algorithm would be compared to evaluate the best performance model. Hence, with the ML models accurate prediction of house price could be made without any error. The study also aims to helps the property owner or any of the real housing agents to make rational decision while undertaking the property transaction. Aim and Objective Ultimate Goal of this project is to build a prediction engine with machine learning model that are capable to predict the price of houses. The main aim of this project is to use linear regression models where it provides the continues ranges instead of classifying the features or classes in categories. So, the advanced machine learning models such as regression-based model another ML model like gradient boosting, decision tree and random forest were built to check the performance of model in how accurately they were able to predict the house prices. The main objective of the project is to: • Collect the house dataset and pre-process it by preforming EDA. • Selection of hypothesis statement • Building ML based model and implementing • Comparison of result regarding which model gives best accuracy in predicting house prices The main significant of this project is to house price prediction that allows the developer to the selling price of houses, and this enable the customers in arranging time to buy house. Theoretical Background Machine learning model for House Price Predictions According to research conducted by Pathak (2021), various machine learning algorithm are compared for house price prediction. The study focused on using the real time factors for predicting the house prices and they tend to use the “Regression technique” of the machine learning. The Real time dataset of house from the

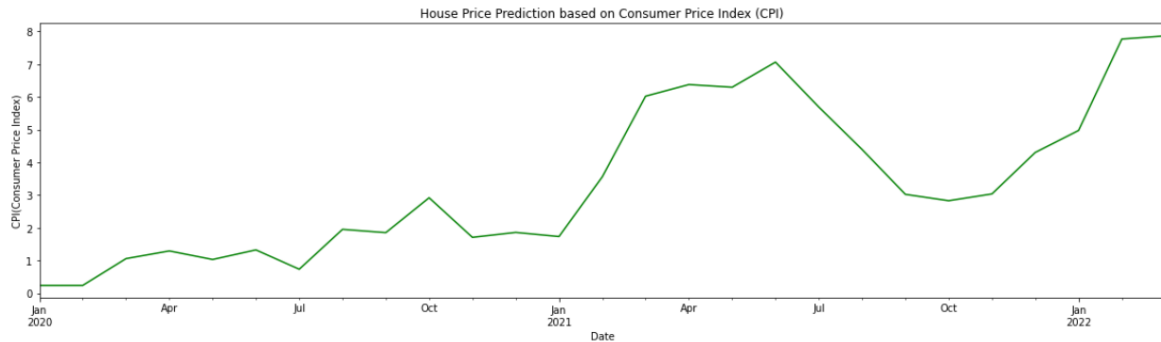


Pune city was employed and was split to train the data, The regression adopted was Decision Tree Regression, Support Vector & Linear Regression. For presenting the output interface the Flask framework was used through which the web application was built. The research performed well by implementing various regression models and this research has illustrated how to select the most effective ML model for the house prediction problems. According to the research proposed by Ho (2020), three machine learning model such as Gradient Boosting Machine (GBM), Random Forest (RF) and SVM are used for predicting house prices. The method examined the housing transaction in Hong Kong and evaluated the models result using the performance metrics like MSE, RMSE and the mean absolute percentage error. The final experimental result concluded was that for the time constraint problems as in this house price transaction for 18 years, the SVM algorithm found to make accurate prediction and outperformed the other two ML algorithms. Therefore, for the property researchers, advanced ML algorithms have been a promising tool in house price prediction. According to research proposed by Truong (2019), the house price prediction problems was resolved through machine learning techniques. For the research, “house price in Beijing” dataset was used to predict the prices of houses in the year 2009 and 2018. The steps undertaken in the research was data pre-processing, data analysis, model selection (RF, XGBoost, LightGBM, Hybrid regression and Stacked generalization regression). The RF generates desirable result in house price prediction but prone to overfitting problems. Hybrid regression model performed better because of generalization. When accuracy is compared, stacked generalization model gives best result but had complicated architecture. In future the combination of different model can be used to check the model performance (Truong, 2020).

Hypothesis Development

Hypothesis_1

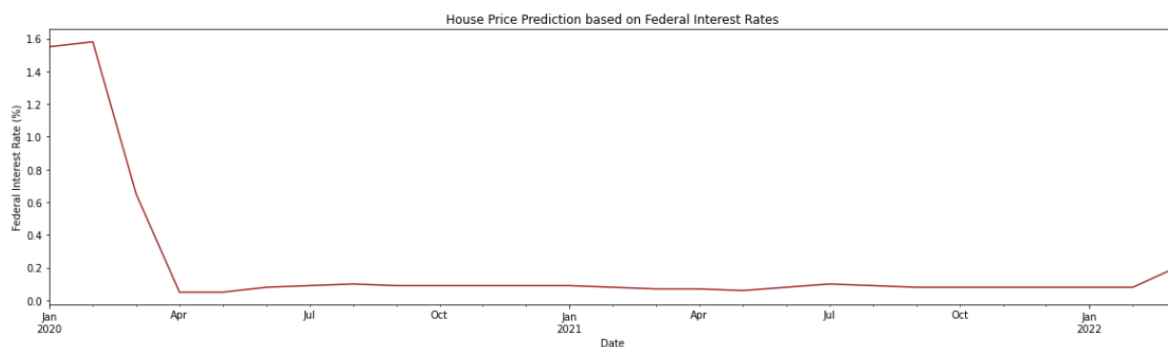
The First hypothesis states that with the increase in house price the Consumer Price Index (CPI) will also increase.



The above graph shows that with the increase in CPI, the house price has also increased.

Hypothesis_2

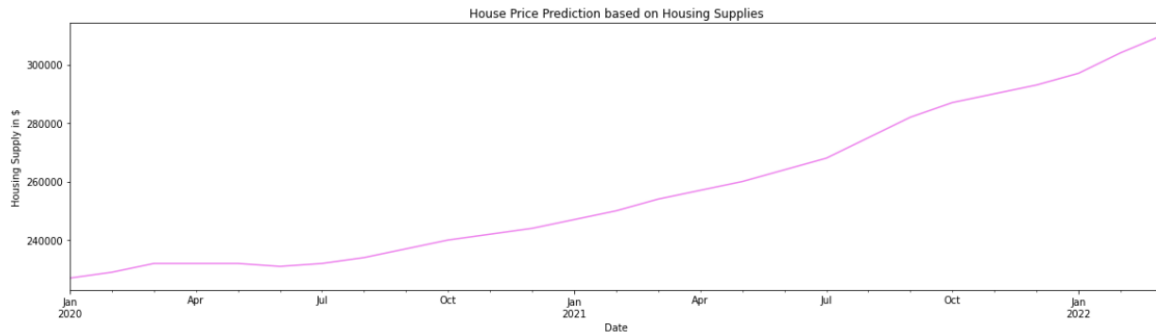
The hypothesis_2 states that, when the house price increases then the Federal Interests Rate decreases.



This hypothesis is true from the plotted graph. We can see that Federal interest rate have been decreased when the house value has increased.

Hypothesis_3

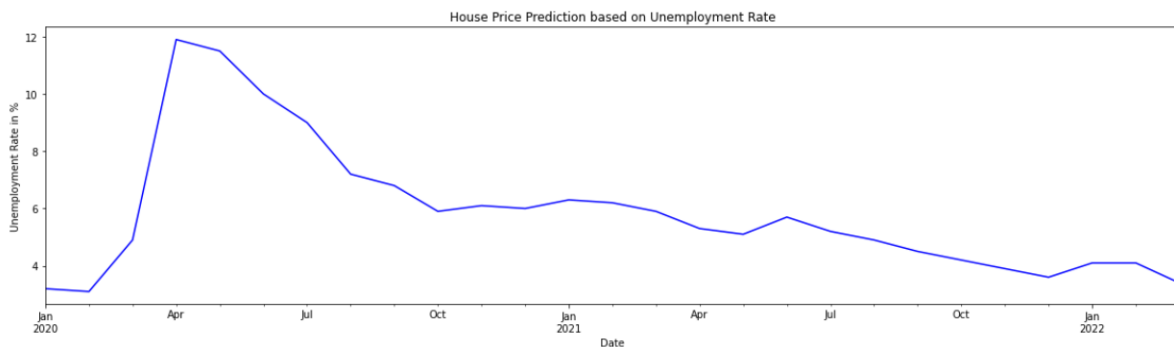
According to hypothesis_3, the increase in house price would decrease the housing supply demand.



But from the obtained result this hypothesis is not true as with the increase in house price we get increase in housing supply. The graph shows a continues increase in curve when the house supplies increased. This is because with the housing supply increases then the house owners would automatically raise the price of houses.

Hypothesis_4

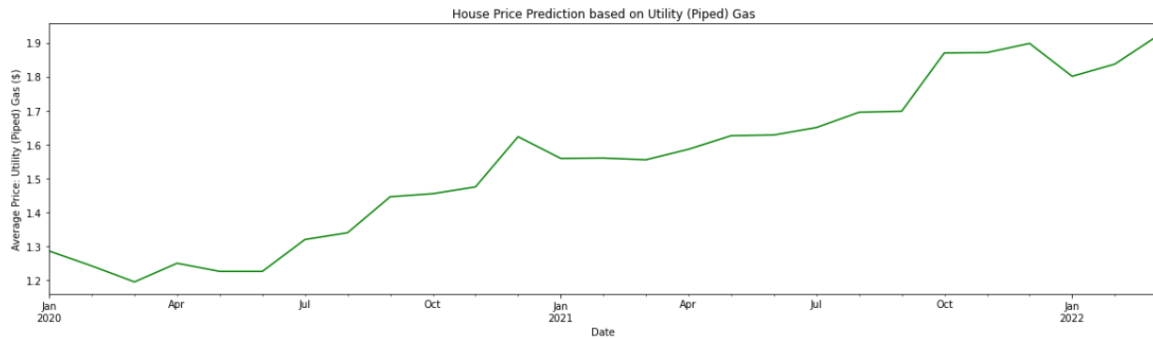
The hypothesis_4 states that when the unemployment increases, housing price also increases. That is when more employment is there, people will invest in new houses thus the house would be in demand make their process higher. But still when there is no work the people tend to buy houses.



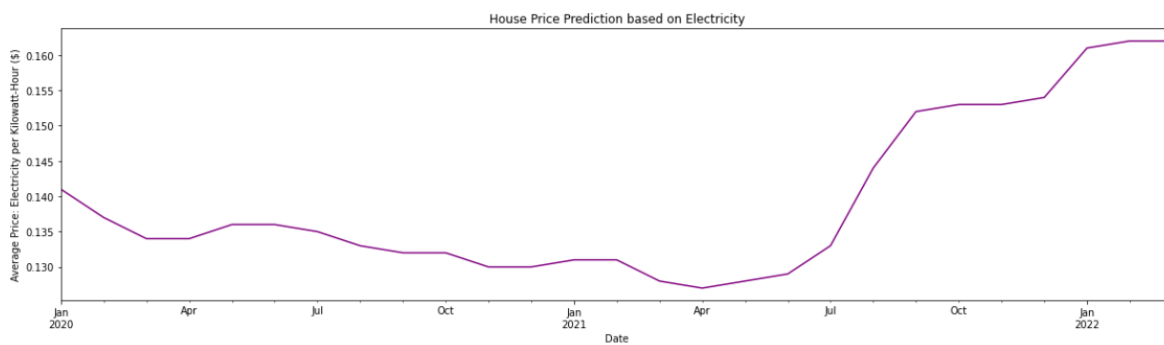
We can see from above graph that when the employment is still not there people tend to invest in house and the house price would increase.

Hypothesis_5

According to Hypothesis_5, when the utilities such as gas and electricity bill increases the house price is assumed to decrease. But this is not the case. Though the utility price increases it does not impact the house price and so house are always in demand so the price also increase.



The above graph depicts that though the utility such as gas increases, still the house price is predicted to increase.



Similarly, for the utility electricity, the house price has shown to increase since April 2021. So even the electricity rate increase does not reduce the house value.

Methodology

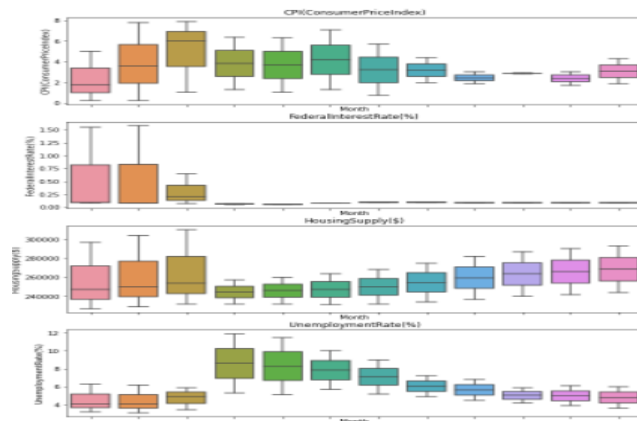
The research methodology undertaken for the house price prediction project is quantitative research. The real-world dataset is collected from multiple websites. The data is cleaned and pre-processed to be fed into the machine learning model such as gradient boosting algorithm, decision tree, random forest algorithm, linear regression models like Ridge Regression, Bayesian Regression model, like OLS model, Lasso Model lastly the Elastic Net Regression Model. Data collection Publicly available data sets have been collected for this research from multiple sources. The dataset contains various attributes such as date from the year 2020 until the present data which is distributed monthly, median price, CPI, federal interest rates, housing supply, unemployment rate, housing inventory, utility prices and other attributes. Data analysis Data analysis involves the cleaning of categories data that has missing values or redundant values so as to effectively help the model in prediction. The inconsistent data are



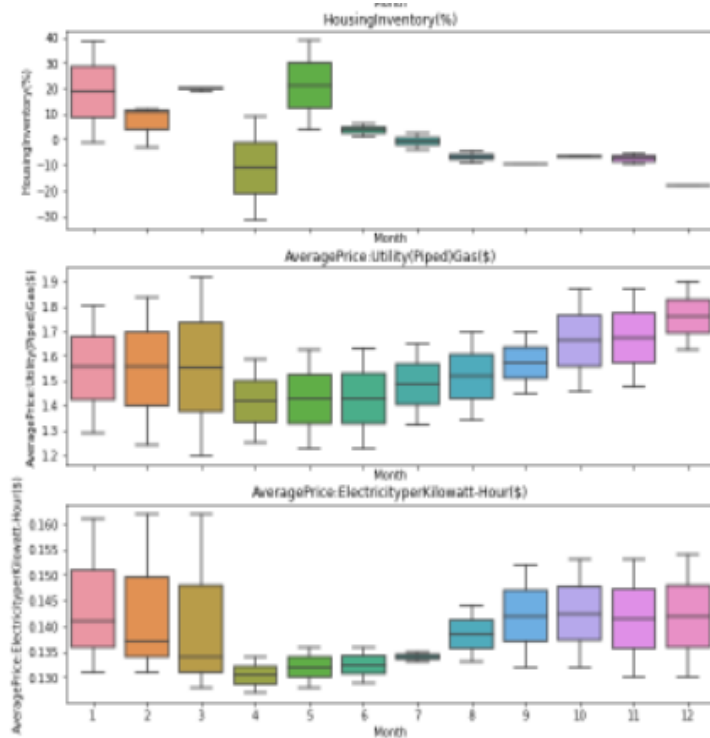
removed and only the necessary categorical features are used to predict the house price and this increase accuracy rate. Proposed methods The extracted features are fed into the model (Kudavkar, 2019) to train the data and the house price is predicted based on time series forecasting. The time series forecasting is referred as the process of analyzing the time series data by using the statistics and then modelling them to effectively predict the outcomes. In our project we analyze the time series data of house prices from the year 2020 to 2022 and based on this time series the house prices are predicted according to the change in demand. XGBoost The Gradient Boosting algorithm is used for forecasting the house prices and this builds the model with a forward stage wise fashion. XGBoost provides the optimization of arbitrary based differentiable loss function. In that case, boosting involves making a weak learner to become more better in prediction and minimize the model loss. Decision tree Decision tree model are tree shaped figures that determines the course of any action. That is to specify each branch of the tree would point a possible decision and reaction. This algorithm makes effective classification decision and only less effort is needed for data preparation (Zhang Z. , 2021). Random forest Random forest regression model would develop decision tress that is based on randomly selected variables that is to provide the class of dependent variable depending on many trees. Most of the decision tree predicts the class accurately for most part of data (Adetunji, 2021). Linear regression The linear regression in house price predicts the scores of single variables from the score or value of the second variable. OLS regression – This is an Ordinary Least Square regression model where we look for the best fitting line in a diverse cloud of points. Ridge regression - Ridge regression model allows to alter the loss or errors by addition of penalty that is equivalent to the square of the magnitude of coefficients. For reducing the complexity, the ridge regression is employed. Lasso linear regression - Lasso regression takes the magnitudes into account and this regression leads to zero coefficients. That is, it avoids some features to generate the output and they reduce loss of model and greatly help in feature selection (SHINDE, 2018). Bayesian regression – The Bayesian regression models are used to estimate the distribution when it comes to prediction problems. It allows the natural mechanism to to survive the poorly distributed data. Elastic net regression model – This elastic net model is referred as penalized linear regression model where it involves both L1 and L2 penalties caried out during training. This regression model would combine the feature elimination to improve the



model's prediction of house price (Imran, 2021). Boxplot explanation The boxplot has been used to show the prediction of housing price based on many attributes such as CPI, federal interest rate, housing supply, utilities, etc.



In the above boxplot, the CPI, federal interest rate, housing supply and unemployment. In March month alone there is high increase in CPI with the increase in house price. The federal interest for starting two months were increased and the remaining month showed declined in federal interest rate. The housing supply increases for tree month and shows a decrease then it increased slightly over the month. Though these attributes have shown an increase and decrease factor, the housing price tends to increase. Similarly, for the unemployment, the month April and may showed the greatest increase abut still the house prices increased.



The housing inventory shows variance despite the increase in house pricing, The housing utility such as gas and electricity show increase and the house pricing also increased along with that.

Results and Discussions

The different machine learning model implemented were evaluated using the performance metric such as explained variance score, and R squared. These two measures are applied of rt regression problems to check their accuracy in predicting the house price for our project. The explained variance score is applied as a metric to measure the discrepancy between the regression model and the actual house data (Park, 2015). The higher percentage of the explained variance value would indicate that stronger strength of association. That means the model perform better predictions.

<i>Model</i>	<i>Explained variance score</i>	<i>R-Squared</i>
--------------	---------------------------------	------------------



OLS model	0.8579	0.8489
Ridge model	0.9123	0.9109
Lasso model	0.8580	0.8490
Bayesian model	0.8088	0.8061
Elastic Net model	0.9056	0.9034

Table 1: Comparison table of Performance metrics

The explained variance for the regression model “Ridge model” is high with 0.9123 when compared with other four model. The next better prediction of house price was achieved by using elastic net model showing 0.9056. The other three models such as OLS, Lasso and Bayesian model perform averagely when compared with the rigs and elastic net model (Zhang Q. , 2021).

The “R-squared” is defined as the statistical measure and in our project, we applied R-squared to measure how close the house data are to be fit in regression line. It determines the measure of goodness of fit of model. The ridge model has highest r-squared score of 0.9109 and next one is elastic net model with 0.9034. The other model has less r-squared values such as 0.8489 for OLS model, 0.8490 for lasso model and 0.8061 for Bayesian model.

Random forest model

Th Out of Bag (OOB)score is a method for validating the random forest model. The OOB score for Random Forest is 0.94444. The accuracy of the random forest classifier is found to be 89%.,

Gradient Booster algorithm

The gradient booster algorithm has been used to predict the house price and it is found that it gives highest accuracy rate of 89% when the leaning rate is 0.1. Because the XGBoost helps to predict not only the continuous target variable but they also used to predict the categorical target variables (Kumar, 2020). The learning rate is decreased from 1.0 to 0.05 but still there is no improvement in accuracy rate.

Decision tree

The decision tree performance is estimated and the accuracy is found to be 89% similar to that of random forest. The explained variance score for decision trees os 0.600001.



Hypothetical testing

The hypothetical testing is performed for the house price prediction where it draws conclusion regarding the probability of distribution. The confidence of null hypothesis is explained with the estimated pvalue for each hypothesis. That is p-value would assume that the possibility of the assumption being truer that depends on null hypothesis. After determining the “pvalue” we can accept or reject our hypothesis. For the significance level of initial assumption made pvalue has the range between “0 and 1”. The pvalue can be less than 0.0 and greater than 0.1.

```
Ttest_indResult(statistic=50.90270678221037, pvalue=4.866303706519669e-46)
```

In our project for the hypothesis_1, the pvalue is very less and below 0.5. That is the change in CPI has no effect on house reprice.

```
Ttest_indResult(statistic=50.90327520245471, pvalue=4.863533341314597e-46)
```

Similarly, for hypothesis_2 the value is very less and so it has no affect on house prices. The federal interest rate whether it increases or decrease does not affect the house price.

```
Ttest_indResult(statistic=3.4498848642826205, pvalue=0.0011208063522833965)
```

But for the hypothesis_3, the value is almost near and so it affects the house price at a slighter rate.

```
Ttest_indResult(statistic=50.90226220163623, pvalue=4.86847162273805e-46)
```

For hypothesis 4 when unemployment rate is considered, it also shows decrease in Pvalue and unemployment rate has no effect on house price. Because it was assumed that though unemployment rate decreases, the house price would increase. But from the result it is not true.

As most pf the hypothesis does not meet the criteria of pvalue of greater than 0.05 then the null hypothesis is not accepted for the house price prediction (G. Naga Satish, 2019).



Conclusions

The different ML algorithm have been applied to predict the house prices from the given dataset. The hypothesis statement was valuated and illustrated using boxplot. The various regression-based methods and other ML methods such as decision tree, RF and XGBoost were applied and their performance metrics were evaluated. The explained variance score and r-squire peromace metric were applied to check the regressionsl mdoels performance in predicting the hosue prices. Most of the algorithm is found to accurately predict the house prices and the model perform better when compared with other linear regression models.

References

- Adetunji, A. B. (2021). House Price Prediction using Random Forest Machine Learning Technique. *The 8th International Conference on Information Technology and Quantitative Management*.
- Babu, A. (2019). Literature Review on Real Estate Value Prediction Using Machine Learning. *International Journal of Computer Science and Mobile Applications*.
- G. Naga Satish. (2019). House Price Prediction Using Machine Learning. *International Journal of Innovative Technology and Exploring Engineering (IJITEE)*.
- Imran. (2021). Using Machine Learning Algorithms for Housing Price. *Soft Computing and Machine Intelligence Journal*.
- Kudavkar, A. A. (2019). House Price Prediction Using Machine Learning Algorithm. *INTERNATIONAL RESEARCH JOURNAL OF ENGINEERING AND MANAGEMENT STUDIES (IRJEMS)*.
- Kumar, B. (2020). HOUSE PRICE PREDICTION USING GRADIENT BOOST REGRESSION MODEL. *International Journal of Research and Analytical Reviews (IJRAR)*.
- Park, B. (2015). Using machine learning algorithms for housing price prediction: The case of Fairfax County, Virginia housing data. *Expert systems with applications*.



Pathak, S. M. (2021). Comparison of Machine Learning Algorithms for House Price Prediction using Real Time Data. *INTERNATIONAL JOURNAL OF ENGINEERING RESEARCH & TECHNOLOGY (IJERT)*.

SHINDE, N. (2018). VALUATION OF HOUSE PRICES USING PREDICTIVE TECHNIQUES. *International Journal of Advances in Electronics and Computer Science*.

Truong, Q. (2020). Housing Price Prediction via Improved Machine Learning Techniques. *Procedia Computer Science*.

Zhang, Q. (2021). Housing Price Prediction Based on Multiple Linear Regression. *Scientific Programming*.

Zhang, Z. (2021). Decision Trees for Objective House Price Prediction. *2021 3rd International Conference on Machine Learning, Big Data and Business Intelligence (MLBDBI)*.

Fred : <https://fred.stlouisfed.org/series/ATNHPIUS19124Q>

Zillow : <https://www.zillow.com/dallas-tx/home-values/>

Appendices

**CODE HAS BEEN REMOVED FOR PRIVACY & CONFIDENTIALITY
CONCERNS**

Linear Regression Models

Evaluation Metrics

Explained Variance Score



R-Squared

Random Forest

Gradient Booster Algorithm

Decision Tree

----- END -----