HOUSE RENT PREDICTION

UE20CS302-Machine Intelligence
Project Title : HOUSE RENT PREDICTION
Team No:
Project Guide : DR GOKUL K
Project Team with SRNs :
ARJUN VINAY AVADHANI - PES2UG20CS901
DISHA SINGH - PES2UG20CS906
PARTH SINGH- PES2UG20CS914

Problem Statement

Renting, also known as hiring or letting, is an agreement where a payment is made for the temporary use of a good, service, or property owned by another. In this Dataset, we have information on almost 4700+ Houses/Apartments/Flats Available for Rent with different parameters like BHK, Rent, Size, No. of Floors, Area Type, Area Locality, City, Furnishing Status, Type of Tenant Preferred, No. of Bathrooms, Point of Contact. On the basis of these indicators we are going to predict house rent values .

In the contemporary digital era, useful information of the society can be retrieved from a wide variety of sources and stored in the form of structured, unstructured and semi-structured formats. In the analysis of economic phenomena or social observations, advancement of innovative technology makes it possible to systematically extract the relevant information, transform them into complex data formats and structures, and then perform suitable

Application and Uses

House rent prediction finds multiple applications in the real world . The value of rent depends on multiple factors and all of them need to be factored in , manually doing this is close to impossible and hence using MI algorithms we can efficiently predict the rent of any house . Many MI algorithms have already been implemented in the past to predict the house rent values with promising accuracy values . MI algorithms use various factors and give us accurate results taking into consideration what the user wants and according to his/her specifications . It also gives us factors which are more influential in predicting rental values and what factors play an important role . The rental industry is a growing one and will continue to remain so . Factors influencing rental value keep changing and hence new MI algorithms help in accurate predictions with the trend and hence house rent prediction finds numerous applications .

Attributes in our dataset

BHK

Number of Bedrooms, Hall, Kitchen

Area Locality

Locality of the Houses/Apartments/Flats

Furnishing Status

Furnishing Status of the Houses/Apartments/Flats, either it is Furnished or Semi-Furnished or Unfurnished

City

City where the Houses/Apartments/Flats are Located

Point of Contact

Whom should you contact for more information regarding the Houses/Apartments/Flats

Rent

Rent of the Houses/Apartment Flats

House_Rent_Prediction|EDA|11_Models

Area Type

Size of the Houses/Apartments/Flats calculated on either Super Area or Carpet Area or Build Are

Size

Size of the Houses/Apartments/Flats in Square Feet

Tentant Prefered

Type of Tenant Preferred by the Owner or Agent

Bathroom

Number of Bathrooms

Floor

Houses/Apartments/Flats situated in which Floor and Total Number of Floors (Example: Ground out of 2, 3 out of 5, etc.)

Literature Survey

PAPER ANALYSIS BY: ARJUN AVADHANI

Title of the paper	Advantages	Limitations
Predicting the rental value of houses in household surveys in Tanzania, Uganda and Malawi : Evaluations of hedonic pricing and ML approaches.	Uses a very intuitve method of calculating the overall performance of different models which ranks the ML approaches in relation to the OLS approach	Journal does not calculate error,accuracy,precision etc and hence does not include it in its findings.
Housing Price Prediction via Improved ML techniques	For Hybrid Regression and Stacking methods, performance tuning was notrequired since both methods were combinations of the best regressions. Instead, architecture implementation could beconsidered to further enhance the prediction	For Hybrid Regression and Stacking methods, performance tuning was notrequired since both methods were combinations of the best regressions. Instead, architecture implementation could beconsidered to further enhance the prediction

Literature Survey

LITERATURE SURVEY DONE BY : DISHA SINGH

Title of the paper	Advantages	Limitations
monitoring house rental price based on social media	it uses a combination of hedonic regression with mi algorithms which is a growing area of research	no cons can be seen as such because we are using MI algorithms to create a more efficient house rent prediction model
Modelling house rent in atlanta metropolitan area using textual information and deep learning	we are also including textual data in our analysis process which is a huge plus .	our analysis is restricted to only atlanta and we have not inluded pictures in our study

Literature Survey

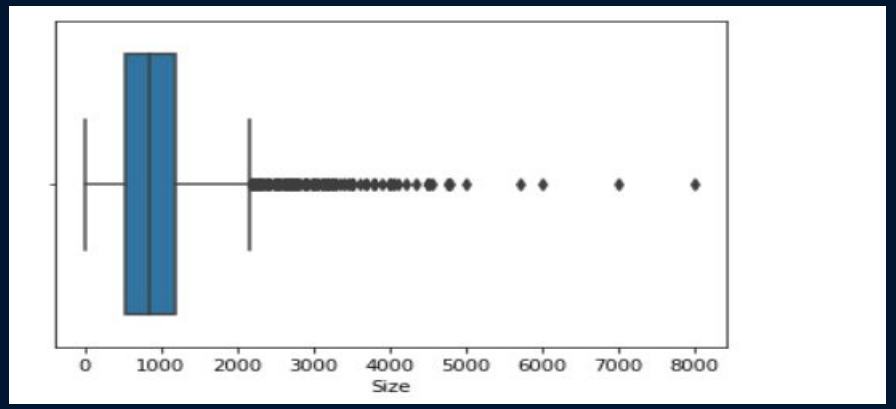
LITERATURE SURVEY DONE BY: PARTH SINGH

Title of the paper	Advantages	Limitations
House Price Prediction using Random Forest Machine Learning Technique	Potential buyers can buy the huse based on the prediction made which considers physical conditions, styles and location, the three main factors that influence the house's price	A comparison of the predicted and actual prices reveal that the model achieves a prediction difference of ±5. There are other models that give accurate prediction.
House Price Prediction Using LSTM	Prediction is much better than baseline ARIMA, about 90% reduce on MSE. Deeper neural network of stacked LSTM is supposed to give better results with the right selection of structure and parameters to improve accuracy	Stateful LSTM model has bad prediction on validation data set.Which indicate that the essence of the problem may not suitable for the application of stateful LSTM.The result in some districts are not ideal. The possible reasons are low frequency of data and the loss of data in several months.

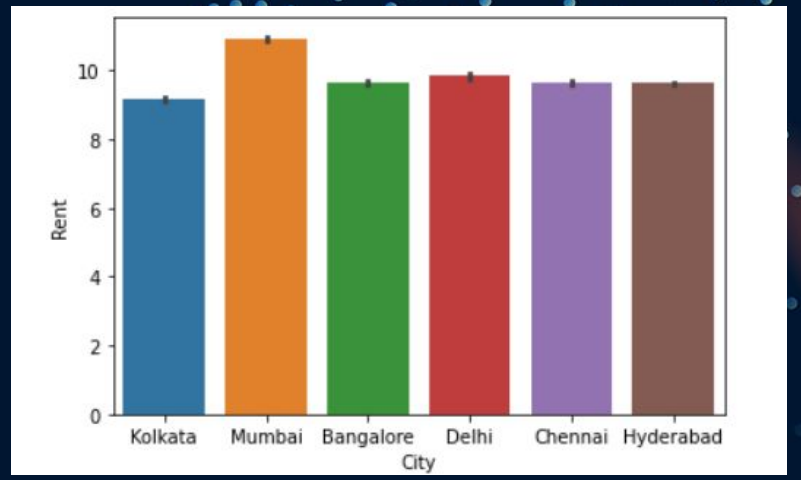
Proposed Approach

1. After performing a thorough EDA we found out our dataset had a lot of outliers . Extreme outliers were dropped and on the others we used log transformations . Our dataset had no NULL values present so we did not have to deal with missing data . Correlation between variables was found to analyse which of the variables were strongly related using correlation heatmap and the variables which were not important were dropped . The dataset was then divided into training and test data .
2. As mentioned earlier we trained our models using 11 different algorithms which are Linear algorithm,ridge algorithm,xgboost algorithm ,catboost algorithm ,lightgbm algorithm,gradient boosting algorithm,lasso algorithm,random forest algorithm,bayesian ridge algorithm,support vector algorithm,knn trained algorithm

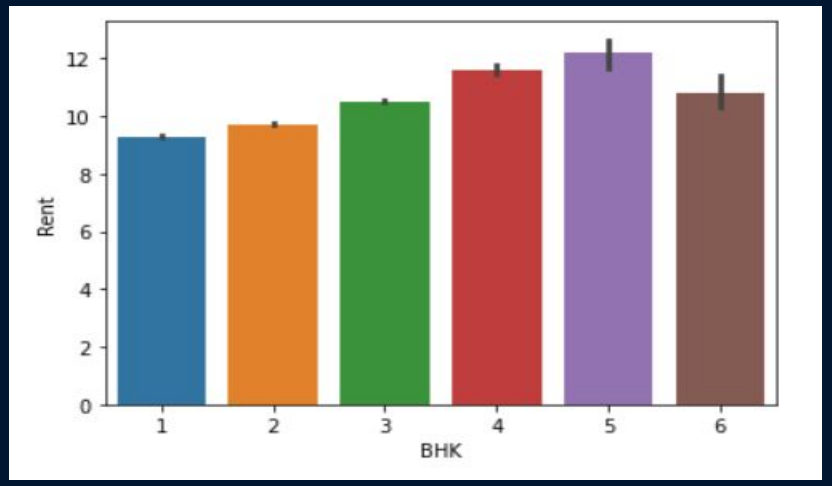
Some important results



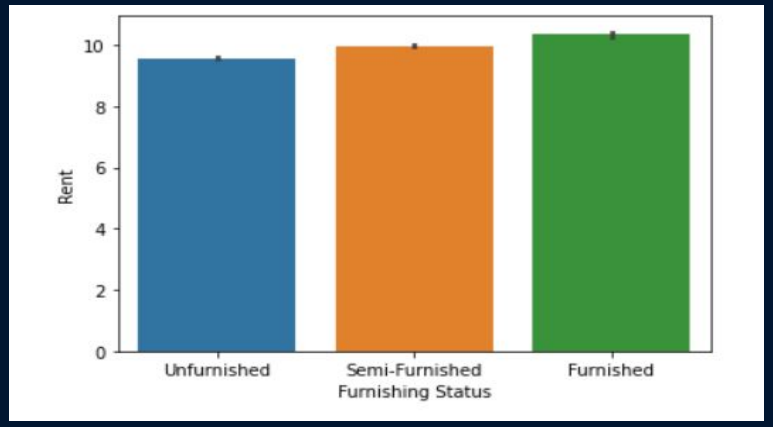
boxplot for outlier analysis (BHK against rent)



from the barplot ,we can conclude Mumbai has the highest rent , which implies that the feature ‘city’ also plays an important role in analysing rent prices

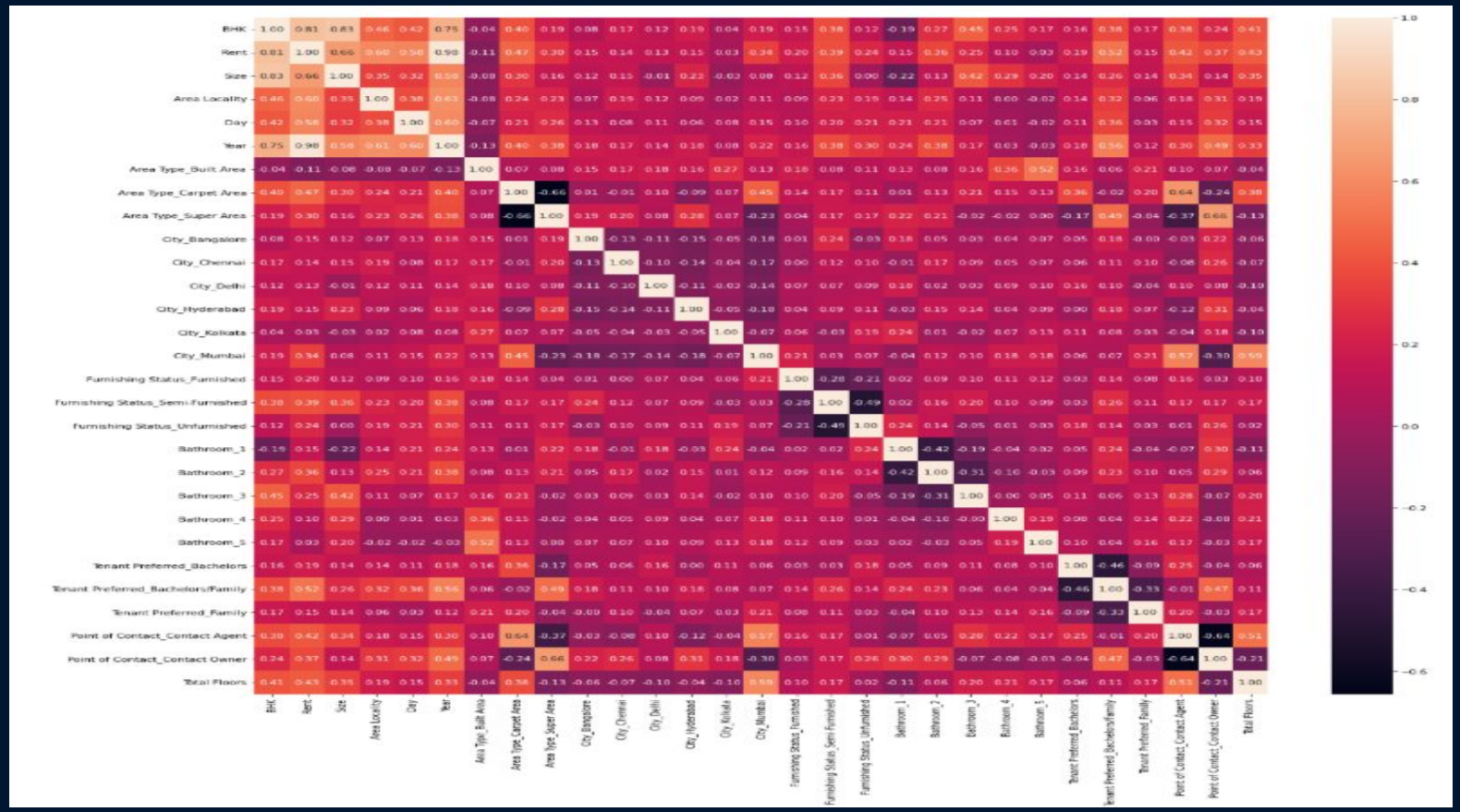


from the barplot ,we can conclude that the rent progressively increases with increase in BHK and in our dataset 5BHK has the highest rent with the exception of 6BHK



From the above graph we can conclude furnishing status also effects rent prices and furnished houses have higher rents .

Correlation heatmap



The shown figure is a Correlation heatmap and it gives us information about linear and non linear relationships between variables . It gives us a measure about how strongly correlated two variables are on a two dimensional plane. This correlation heatmap has been formed after dropping some features to reduce probability of multi-collinearity.

LIGHT GBM ALGORITHM

1. LightGBM which is a fast, distributed, high-performance gradient boosting framework that is based on the decision tree algorithm and is used for ranking, classification and many other machine learning tasks. LightGBM splits the tree leaf-wise as opposed to other boosting algorithms that grow tree level-wise. It chooses the leaf with maximum delta loss to grow. Since the leaf is fixed, the leaf-wise algorithm has lower loss compared to the level-wise algorithm.

2. Some of the important hyper parameters of this method are :

3. num_leaves (controls complexity of tree model), min-data-in-leaf (avoids over fitting in a leaf wise tree) and max_depth(limits the tree depth explicitly).

LIGHT GBM

#Lightgbm is giving the r2_score of more than 0.71

#Hence,We Select the LightGBM

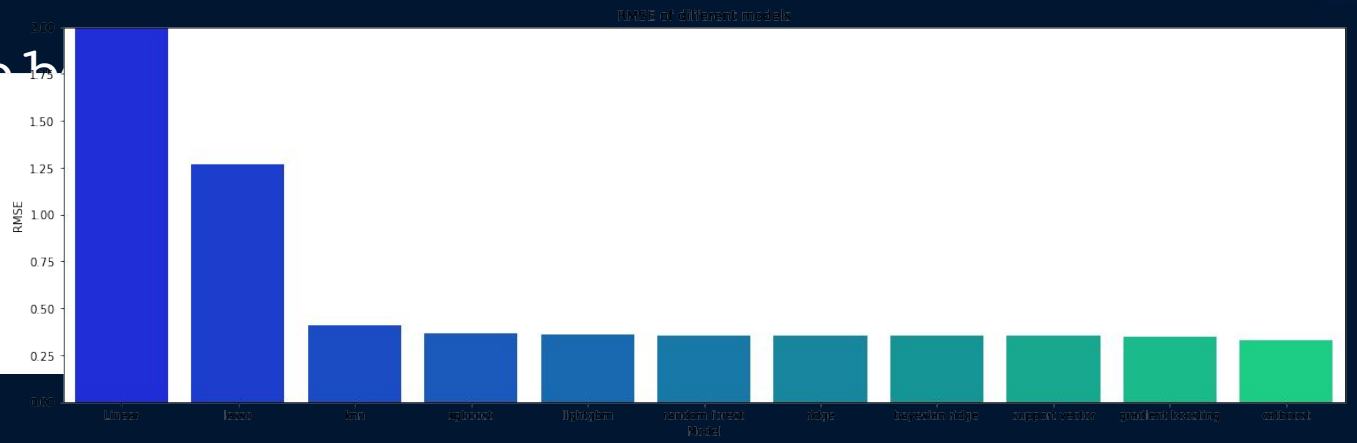
#Light GBM is a fast, distributed, high-performance gradient boosting framework

#based on decision tree algorithm, used for ranking, classification and many other machine learning tasks.

Results and Discussion

After running our models we inferred that the LightGBM algorithm suits our dataset the best and gives us more accurate and efficient results .
Evaluation metrics used in our project are R squared and Root mean squared error .
Using the Light gbm method we have found that:
1. The R squared value is greater than 0.71 which is higher compared to all the other tested models.
2. The RMSE value is found to b

```
Linear r2_score is 0.667356035807981
ridge r2_score is 0.6687001695717816
xgboost r2_score is 0.6782093439768724
catboost r2_score is 0.6699485536735215
lightgbm r2_score is 0.7130113622736785
gradient boosting r2_score is 0.6799024730300449
lasso r2_score is -0.03677438959812562
random forest r2_score is 0.6903932534504511
bayesian ridge r2_score is 0.6689207974172564
support vector r2_score is 0.6502212669411573
knn r2_score is 0.5753463137414174
```



Refrences

Paper1:
https://www.researchgate.net/publication/349227005_Predicting_the_rental_value_of_houses_in_household_surveys_in_Tanzania_Uganda_and_Malawi_Evaluations_of_hedonic_pricing_and_machine_learning_approaches Paper 2: (PDF) Housing Price Prediction via Improved Machine Learning Techniques (researchgate.net) Paper 3:
Monitoring housing rental prices based on social media:An integrated approach of machine-learning algorithms and hedonic modeling to inform equitable housing policies - ScienceDirect Paper 4:
Monitoring housing rental prices based on social media:An integrated approach of machine-learning algorithms and hedonic modeling to inform equitable housing policies - ScienceDirect Paper 5:
https://paperswithcode.com/paper/house-price-prediction-using-lstm Paper 6: (PDF) House Price Prediction Using Random Forest Machine Learning Technique. (sciencedirect.com) Dataset used :
https://www.kaggle.com/datasets/iamsouravbanerjee/house-rent-prediction-dataset

THANK YOU