Housing Rent Prediction Using Machine Learning

Amritha Puravelil Manoharan

3050687

Submitted in partial fulfillment for the degree of

Master of Science in Big Data Management and Analysis

Griffith College Dublin
January, 2023

Under the supervision of Aqeel Kazmi

Disclaimer

I hereby certify that this material, which I now submit for assessment on the programme of study leading to the Degree of Master of Science in Applied Digital Media at Griffith College Dublin, is entirely my own work and has not been submitted for assessment for an academic purpose at this or any other academic institution other than in partial fulfilment of the requirements of that stated above.

Signed: _____   **Date: 30/01/2023**

Acknowledgements

I would like to thank my supervisor Dr. Aqeel Kazmi for guiding me throughout the project. He helped me with ideas and guided me in improving my presentation. I would also like to thank my other professors who helped me thought out my course. I would also like to thank my friends and family as they have been a constant support. Last but not least I would like to thank Griffith college for providing me with resources.

Table of Contents

List of Equations

List of Figures

List of Tables

Abstract

A basic human requirement is housing. However, a shortage of affordable properties that may be rented or purchased affects towards the world's housing crisis. According to the World Bank, 1.6 billion people would be impacted by the global housing crisis by 2025. Housing costs have increased faster than earnings in the majority of countries. Right now, the major cities in the world are suffering housing crisis. This project here tries to understand what are the factors affecting this price. What algorithm is best for predicting the housing rent.

The data for the project is taken from Kaggle. It has attributes like number of beds, baths, type of parking, area etc. The project makes use of machine learning to predict the housing rent by analysing these factors. Here the paper also uses the coordinate location to predict the price, hence understanding how the location can influence the price. The paper compares Random forest, linear regression, support vector machine, and artificial neural network algorithms to predict housing rent. The pre-processing of the data is done in R programming, the visualization on Power BI, and the prediction in python.

Chapter 1. Introduction

Different factors affect housing prices around the world. Population growth and economic expansion are two major factors that contribute to rising housing costs. The housing crisis has limited people from buying houses to renting them in major cities. The increase in population is one of the keys causes pushing up housing costs. The need for housing rises as more people settle in cities and metropolitan areas. As more individuals compete for a small number of available homes, prices may rise as a result. In addition, when more people migrate into cities, the price of land and building supplies tends to increase as well, further pushing up housing costs.

 The rent of these apartment is increasing in rapidly as the supply can't satisfy the demand. The rent of these apartments is influenced by factors like location and features of the property. As housing market is crucial to the nation's economy its beneficial for the renters, real estate agents and the economic professional. The dynamics of the overall economy will also be impacted by the excessive growth of housing prices, in addition to the quality of life.

The housing price changes depending on the area or the location. Hence the high demand and thriving economic nations like USA, Ireland, and UK tend to have higher housing prices and rents. In the case of developing countries even though the prices are not as that of the above-mentioned countries the price has increased compared to past years. For comparison, the median price of a property in the US is around $250,000 and in India, that cost is $50,000.

It is difficult for middle-class people to purchase and rent residences since real estate prices are rising significantly. Decision-makers and the government should be informed of the current trends and how they might affect future prices. If real estate prices keep increasing at this rate, several analysts predict that the housing market will eventually collapse.

Machine learning has emerged as a crucial prediction method due to the growing interest in Big Data since it can forecast property prices more correctly based on their qualities, regardless of the data from prior years. Numerous research looked into this issue and demonstrated the effectiveness of the machine learning technique.

There are a lot of studies that tried to understand and predict housing prices usually from data available from the official sites. The data for this project has been taken from

Craigslist and is available on Kaggle. Craigslist is an American classified advertisement website where you can post your advertisements of anything, housing to unwanted items and Kaggle is a community of data enthusiasts that share their code, data, and thoughts on machine learning.

The goal of this project is to analyse this data using various machine learning algorithms and find the best one. The understanding of an efficient algorithm that can predict the price will help people in real estate, economists, and policymakers in general. Which contributes to affordable homes for renters.

In this paper, we compared Random Forest, Support vector machine, multi-variant regression, and Neural network to find a good prediction model. Hyperparameter tuning was used hence a better output for each algorithm is produced. Two types of hyperparameter tuning were tested one being random search for random forest and grid-search for the neural network.

Chapter 2. Background

2.1   Literature Review

There has been a lot of research done on housing price prediction. A lot of initial machine learning research has been done for housing price prediction. These studies mainly focus on the basic statistical values as the attributes. The number of rooms, bathrooms, area of the property, and location.

In the housing market, machine learning has been utilized to identify prospective buyers as well as potential sellers. Finding wise investments has benefited from it as well. This is accomplished by looking at the cost of the location, the occupancy rates, and the capital growth. The housing market has become much more comprehensible due to machine learning.

Data mining algorithms used to predict housing price has been based on different factors such as the property features, the people whose interested, location. The data mining algorithms analyse these factors. Some common data mining algorithms used as linear regression, random forest, decision forest, and neural network.

The study by Truong used traditional and advanced machine learning algorithms to predict the housing price. They have considered the housing price index and other property features for prediction. As only using the housing price index doesn't predict individual housing prices. They have also created attributes from existing attributes so as to enrich the dataset. They have also set a minimum value for the attributes like price and area. After cleaning the data was standardized and hot-encoded. The evaluation function used in this paper is Root mean squared logarithmic Error (RMSLE). The algorithms used here are Random Forrest, Extreme Gradient Boosting (XGBoost), Hybrid Regression, and Stacked Generalization. They used Grid Search to tune the algorithms and Random forest produced the lowest error.

Housing price prediction in Melbourn city had data similar to the above paper but they have property features, location data, and type of property. Columns with more than 55% of missing data were deleted as imputation is difficult for those cases. The locational data was imputed using google map API. The limit was set for land size. After cleaning important predictors were selected using techniques like stepwise and boosting. The algorithms used here are linear regression, polynomial regression,

regression tree, support vector machine, and neural network. The evaluation function used here is mean squared error. PCA was also used for selecting important attributes for prediction. SVM produced the lowest error combined with stepwise.

Other papers have also struggled with feature selection creating new attributes is the popular way of handling this issue. Followed by adding extra data depending on the market is also added. The study by Rushab Sawant has added data like vastu since its an important factor affecting the Indian housing market. Vastu is the ancient science of architecture in India, which people in India still follows. The algorithm used here is also decision tree and random forest. With random forest producing less error. Here the evaluation functions used are R2 value, mae, mse, and msle.

Paper by Ayush Varma also considered the location of the property as an import factor and calculated the distance of each property to the nearest schools, railway stations, etc. this is done using the google map API. The algorithms used here are linear regression, forest regression, and neural networks.

By the research we have done the main algorithms used here are linear regression,

In a paper by Fan they have also used SVR and linear regression. Followed by RSME value for evaluation. This study contained more than 300 features so feature selection was done using 3 methods: filter, wrapper and embedded. Then they compare the finding and selected the best method.

In the paper Housing Value Forecasting Based on Machine Learning Methods by Jingyi Mu performs machine algorithms like SVR, least squares support veactor machine and partial least squares. They found SVR and LSSVM to perform much better than PLS.

2.2 Data mining techniques and applications

Data mining is the process of analyzing data to produce meaningful knowledge from a huge amount of data. This is done using statistics and machine learning algorithms. Through this, we can identify hidden patterns in the data that a human being cannot understand.

Data mining is used in various industries, from health care to business, from scientific research to social media analysis.

Business analytics is a study of itself it covers topics like customer relationships, marketing analysis, supply chain predictions, fraud detection, risk management, etc. the goal of business analytics is to increase the performance of a business. These can be done by studying sales and marketing, finding trends in the finance part of the business, and understanding the customers using customer behavior analysis.

Rick management is also an important part of business hence analysis of financial markets, credit risk can be done using machine learning.

In the case of health care data mining is used to find the potential health risks in a patient by analyzing the historical data of the patient. This in term help in diagnosis. During the time of the pandemic, data mining techniques were used to predict and understand the spread of the virus. It is also used as a tool for creating a new drug. Algorithms like KNN and neural network is used to classify and identify diseases. Image analysis is used to track diseases like cancer.

An interesting use of data mining is in astrophysics where they analyze a large amount of astronomical data. The researchers try to understand the formation of celestial bodies. DNA sequencing, material science, and climate science are the other fields in which data mining is used in scientific research.

Social media analysis is another application of data mining. Here it is used to analyse data from Instagram, Facebook, Twitter, etc. data from these sites are used to understand the market and identify the demographic type of a person.

Sentimental analysis is used to understand the emotional context of a text. This can help in marketing, politics, customer feedback, etc. it uses natural language processing to identify these emotions.

Chapter 3. Methodology

This paper tries to predict the rent of housing we'll be using machine learning models. The initial steps of this process are the same: data collection, pre-processing, and EDA then comes each model is tested. After producing output by the evaluation function, we compare the best model.

Here the data is taken from Kaggle. Kaggle is a website that shares datasets, skills, and projects based on machine learning. Mostly data scientists and engineering solve data-related competitions. Kaggle also provides an environment for this competition to analyze the data. The housing rent data is a dataset that is taken from craigslist. Which listed advertisements for each property.

The pre-processing of the dataset is done in R programming. R is a statistical data analysis and visualization tool. It supports machine learning, data manipulation, and statistical analysis.

For EDA and visualization Power BI is used. It is a visualization tool by Microsoft. It user friendly as it can work with a wide variety of data sources, e.g.: CSV, Excel sheets, databases, etc. The dashboards used here are interactive hence providing reports that are also interactive.

Python programming is used for implementing machine learning algorithms. Python has large libraries and tools for machine learning like sci-kit-learn. Using its simple coding techniques, it is easier for data analysis.
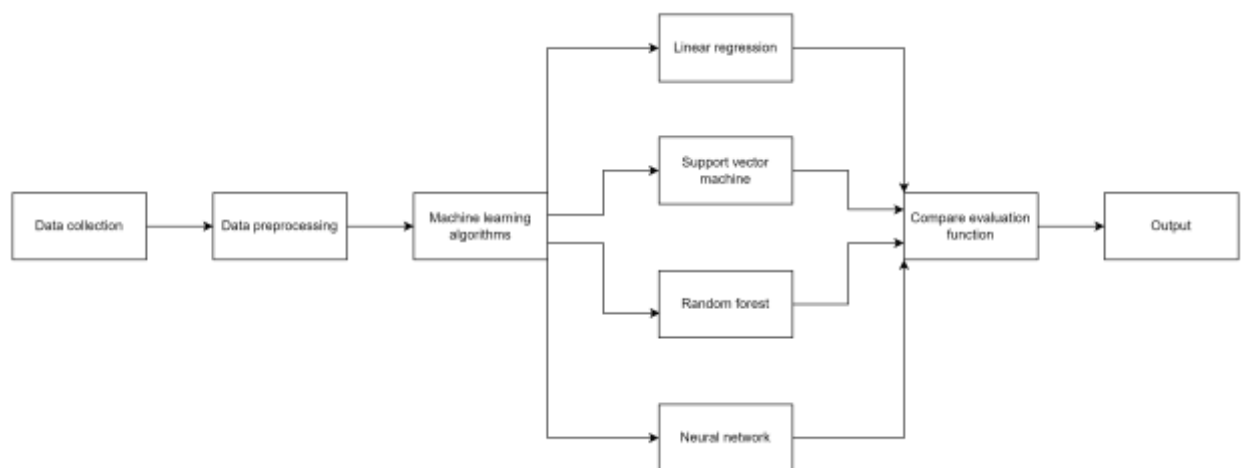

Figure 1: architecture of the proposed system.

Chapter 4. System Design and Specifications

4.1   Hardware Requirement

- PROCESSOR: this project was done on an 11th Gen Intel(R) Core(TM) i7-1165G7 @ 2.80GHz   2.80 GHz. Anything above Intel i5 is fine.

- RAM: 8.00 GB but anything above 4 GB is required.

- HARD DISK: Space up to 15GB would be required.

4.2   Software Requirement

- OS: Windows 10

- Python version: 3.10

- Google Chrome for accessing Colab.

4.3   Technologies used

- Pre-processing: R programming

- Visualization: Power BI

- Model building and performance prediction: Google Colab.

Chapter 5. Implementation

The data set contains 50,000 records with 22 features. These features can be divided into 3 different groups: Locational, structural, facilities, target variable and others.

Type of attribute	Attribute name	Data type	Description
Locational	region	String	Craigslist region
	region_url	String	Region URL
	lat	Float	Latitude
	long	Float	Longitude
	state	String	State of listing
Structural	type	String	Housing type
	sqfeet	Float	Total square footage
	beds	Integer	Number of beds
	baths	Integer	Number of bathrooms
	wheelchair_access	Boolean	Has wheelchair access
	electric_vechile_charge	Boolean	Has electric vehicle charger
	comes_furnished	Boolean	Comes with furniture
	laundry_options	String	Laundry options available
	parking_options	string	Parking options available
Facilities	cats_allowed	Boolean	Cats allowed
	dogs_allowed	Boolean	Dogs allowed
	smoking_allowed	Boolean	Smoking allowed
Target variable	price	float	Rent per month
Others	id	Integer	Listing id
	url	String	Listing URL
	image_url	String	Image URL
	description	String	Description by poster

Table 1: Attribute table

5.1 Data Pre-processing and EDA

Exploratory Data Analysis is the process of understanding a dataset. It helps to find patterns in the dataset, remove or handle unclean data, and find and summarize the characteristics of the dataset.

The data pre-processing in done in R. The data is first loaded onto the R environment. The figure below shows the summary of the raw data.

Figure 2: Summary of Dataset

5.1.1 Data Cleaning

Before using any machine learning algorithms, we must first clean the data. In data cleaning raw data is converted to more reliable values. This is done by handling missing values, handling outliers, handling duplicate vales etc.

- Missing values in lat and long were deleted as number of rows where less.

- One to many and many to one mapping between region and region_url was removed.

- Some of the missing values in parking_options and laundry_options were entered as space values. These missing values were imputed using MICE cart. MICE or Multiple imputation by chained Equation uses regression model to predict the missing values. It also considers the relation between variables and can produce efficient estimations. Here since the variables are categorical, we use MICE-CART. CART stands for classification and regression trees. It is a type of decision tree-based algorithm.
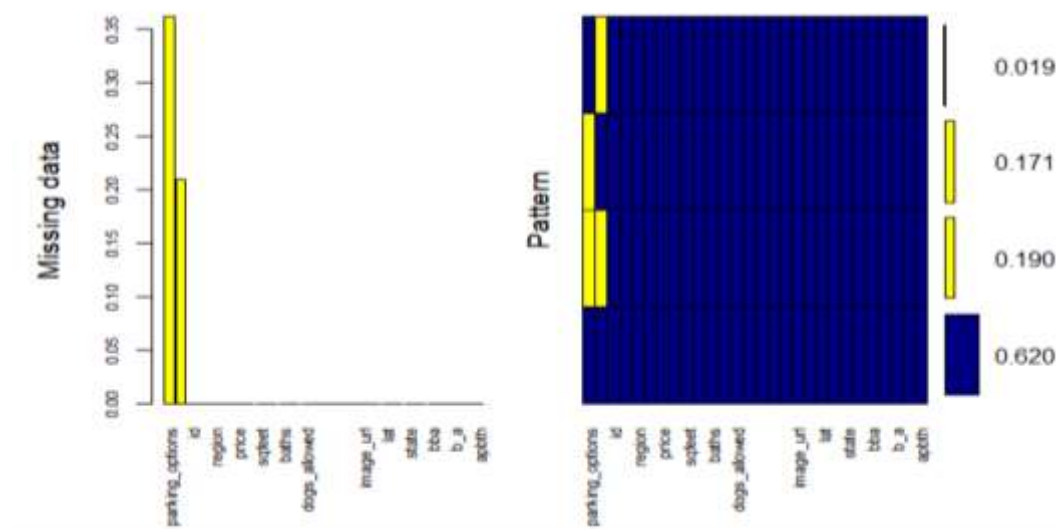
Figure 3: Imputing Dataset

- Deleting 0 valued rows for variable such as beds, baths, price, and sqfeet.

After doing the above EDA steps we try to check the correlation hetmap of the dataset. Figure below shows than target variable price has zero correlation with other variables.

Figure 4: Correlation heatmap 1

Removing outliers from price and sqfeet variables. The lower limit of these variables is set as 150 and upper limit as the 1.5 * Q3. Since the minimum rent and size of the property won't be below 150 units and Q3 is the third quantile value.

Inter Quartile Range is used to handle outliers in a dataset. Usually it is defined by the range between the 25th quartile and 75th quartile. Formula being

$$IQR = Q3 - Q1.$$

So, all the values between

$$Q1 - 1.5*IQR$$

and

$$Q3 + 1.5*IQR.$$

But here we have taken Q3 as 99th quantile value.

After removing outlier the heatmap becomes the following. There is a small corelation seen between price and other variables.

Hence corelation values are low we create new variables from existing variables:

Attributes	Data type	Description
ppa	float	Price per area
bba	float	Beds per area
apb	float	Area per beds
b_a	float	Beds times area
b_b	float	Beds times baths
apbth	float	Area times baths

Table 2: New attributes.

- During visualization of coordinate points in power BI, there were some co-ordinates outside the US borders these were identified and deleted.
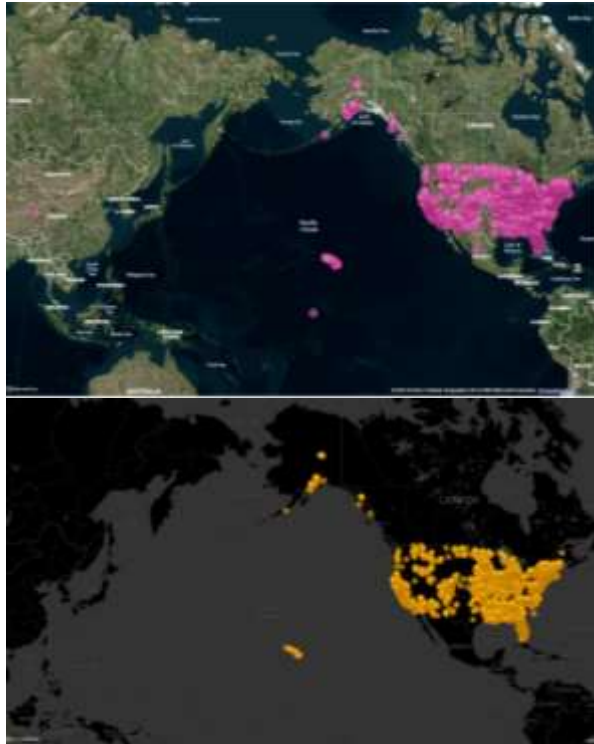
Figure 6: Map cleaning

Now the correlation of variables with price becomes the following:
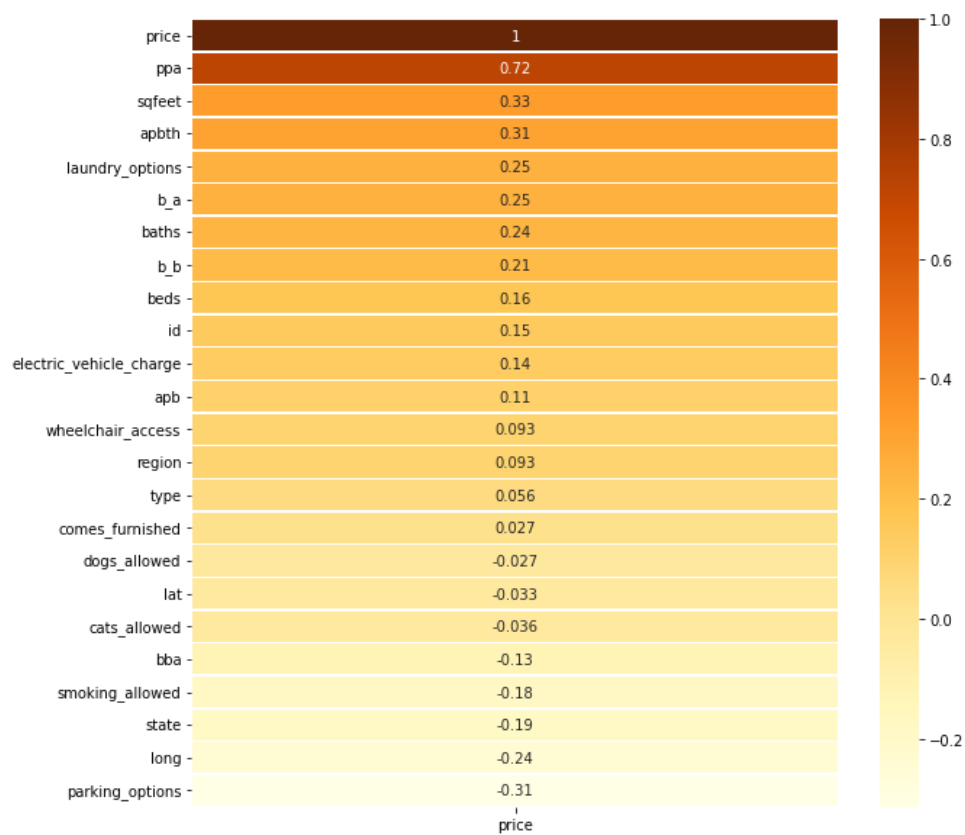
Figure 7: Correlation of price variable.

5.1.1 Data Visualization

- Correlation between price and sqfeet by parking_options and also by type.
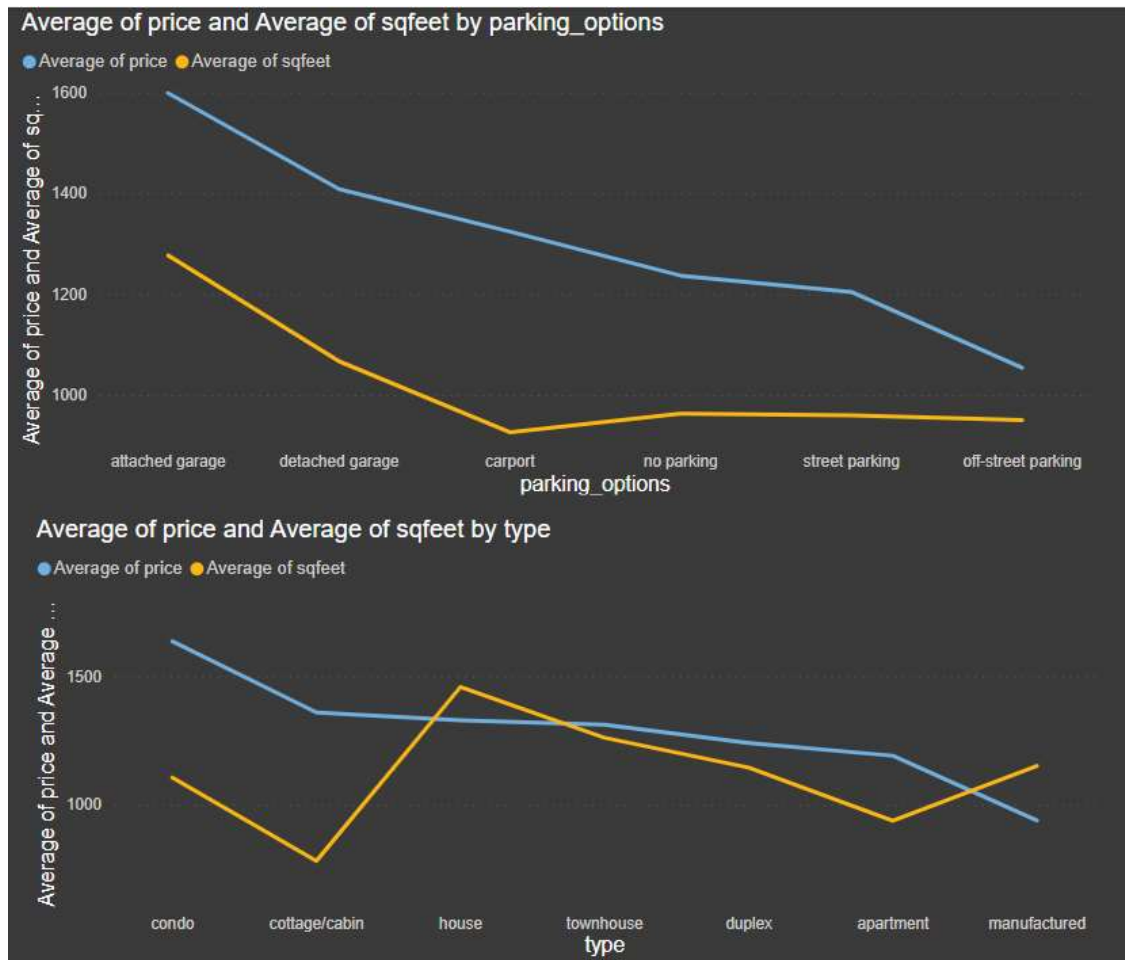
Figure 8: Average price and sqfeet across type and and parking_options

- Correlation between beds and baths by type.
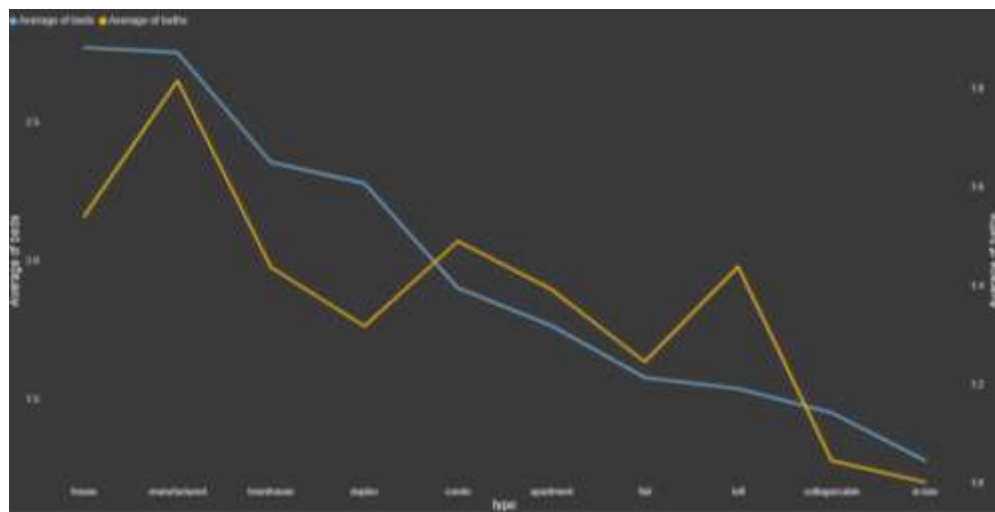
Figure 9 Average beds and baths across type.

- Correlation between cats_allowed and dogs_allowed by type.
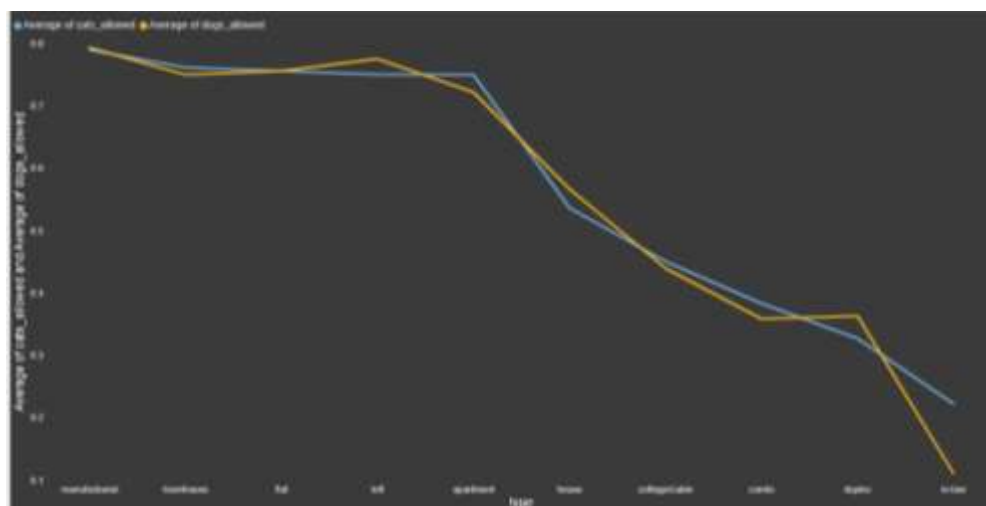
Figure 10: Average cats_allowed and dogs_alloed across type.

Chapter 6. Model Building

After analysing the data, the cleaned data is used for building the model. Python is used for model building and the main library used for machine learning in python is sklearn. It is an open-source library which contains algorithms for clustering, regression, classification etc. It also provides tools for pre-processing like normalization, evaluating a model, cross-validation and performance analysis.

6.1 Splitting of Dataset.

Before applying any machine learning algorithm, the data must be divided into testing and training datasets. After splitting the algorithm is applied to the training set and the effienency of the model is tested on test set.

train_test_split is used from sklearn to split the dataset.

6.2 Data Normalization.

Data Normalization is the process of transforming the dataset to a uniform standard scale. There are several methods like Min-Max method, Z-score normalization etc. here we have used Min-Max metho.

In Min-Max method data is scaled between 0 and 1 by subtracting the value by the minimum value of that column and dividing by the range of that column.

$$x\_normalized = (x\_orginal - x\_min) / (x\_max - x\_min)$$
Equation 1: min-max normalization

MinMaxScaler is used from sklearn.

6.3 Machine Learning Model.

In this project, we have used 4 machine-learning algorithms that use regression to predict the housing price. The Algorithms used are:

- Linear Regression:

Linear regression is the model than find the linear relationship between the variables. The formula for linear regression is:

$$y = mx + c$$

Equation 2: Linear Regression

where y is the dependent variable, x is the independent variable, c is the y intercept and m is the slop.

Advantages:
1. Efficient: Its can handle large volume of data.
2. Best for simple relationship between independent and dependent variable.

Disadvantages:
1. Limited to linear relationship.
2. Outliers can easily affect the prediction

- Support vector Machine

Support vector machine is used for classification and regression. This algorithm creates a hyperplane between different classes. Then the algorithm finds the best hyperplane separator that is closest to the data points. The closest data points to the hyperplane are called support vector.

SVM has four basic kernels: linear, poly, rgb and sigmoid. In this project we have used grid search to find the best kernel, that is rgb.

Grid search is a type of hyper parameter tunning that systematically find the optimal combination of hyperparameters. The models created by the grid search is trained using the performance metric. Hence providing the highest score models. We can use it to find the best learning rate and regularization strength.

Advantages:

1. Handle non-linear relationship
2. Less impact by outliers due to variable margin between data points and boundary
3. Handle large set of data
4. Can handle multiple outputs by training each pair of outputs.

Disadvantages:

1. Requires tunning as there is a lot of hyperparameters.
2. Output depends on the kernel we use.
3. Expensive in the case of computational time and space.

- Random forest

Random forest uses ensemble learning technique which allows it to combine multiple learning algorithms to predict the best single model. Here we have used Random Search for hyperparameter tunning. Here hyperparameters are randomly searched until the best combination of parameters are found. Its is simples and less expensive alternative for grid search as it can handle complex hyperparameters.

Advantages:

1. Less impact from outliers.

2. Data can be predicted in parallel hence faster and better for a large set of data.

3. Can handle missing data.

4. Feature importance is used to identify important variables for modeling.

5. Hyperparameter testing is required.

Disadvantages:

1. Require a Large Amount of Memory

2. There is a risk of overfitting.

3. Takes a long time to train

- Neural network

Neural network machine learning model is a model similar to the structure of a human brain. It has neuron interconnected in layers. The independent variables are passed through the input layer in the network which goes through different hidden layer to predict the targeted variable in the output layer.

An activation function is used in each neuron these can be: sigmoid, relu, etc. Figure shows the general structure of a neural network.

input layer   hidden layer 1   hidden layer 2   output layer
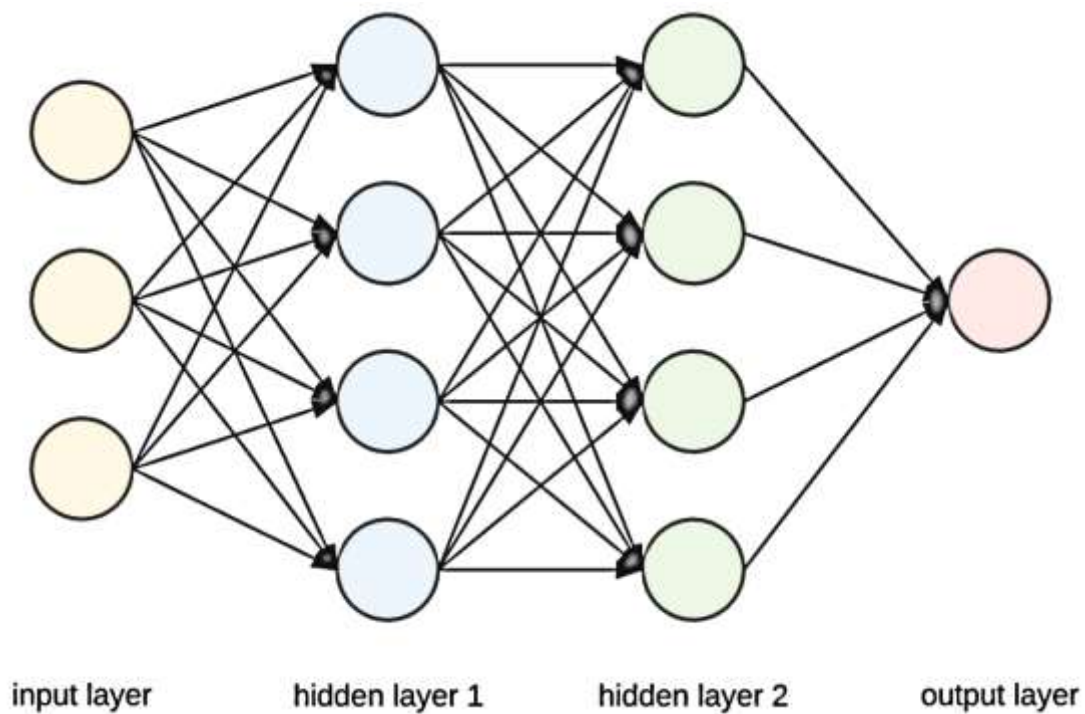
Figure 11: Architecture of Neural Network.

Advantages:

1. Suitable for complex non-linear regression problems.
2. Handle larger set of data as prediction is parallelized
3. can handle missing data.
4. Automate feature engineering. It learns complex relationships own its own.

Disadvantages:

1. Overfitting
2. High memory and execution time require.
3. Required hyperparameter tuning.

In this paper we have used grid search for hyperparameter tuning where we found that the best classifier, we can use is adam. The activation function used here is relu.

- Evaluation Function:

  The evaluation function used here is Mean Squared Error. It is the average squared difference between the test values and the predicted values. It is also known as mean squared deviation.

$$MSE = \frac{\sum(y_i - \hat{y}_i)^2}{n}$$

Equation 3: Mean Squared Error

Advantages:

1. Easy to compute.
2. Suitable for regression models.
3. Measure the magnitude of error.

Chapter 7. Results and Evaluations

Different algorithms were used which predicted a continuous target variable price. Linear regression, SVM, Random Forest and neural network algorithms were tuned using grid search or random search function provide by scikit-learn. The predicted results are as follows:

Model	Accuracy	MSE
Linear Regression	0.8982	0.0017
Support vector machine	0.9173	0.0014
Random forest	0.9935	0.0001
Neural network	0.9996	5.4327e-06

Table 3: Model performance

The experiment result show neural network has the highest accuracy and less mean squared error. Followed by random forests with similar results. Linear Regression and SVM had similar results with linear regression performing the worst. Hence Artificial neural network model proposed by the project can effectively predict the housing price. Even though the neural network has the highest accuracy and least MSE values it is prone to overfitting. Hence by our study, we can conclude that Random Forest is the best algorithm for predicting housing rent.

Chapter 7. Conclusion

This paper tried to discuss the various machine learning algorithms that were used to predict housing price in the past by different studies. We took necessary the steps during EDA to transform raw data into useful data for prediction. The model that predicted the price with the highest accuracy is Neural Network with 99% of accuracy. Hence for the particular dataset we considered neural network is the best model.

In Future there is space for text analysis from which we can extract more data and use machine learning to predict much more accurate data. We can also use Google Map API to find the nearest between facilities like schools, hospitals and railway station in future studies.

References

1. Truong, Q. et al. (2020) 'Housing Price Prediction via Improved Machine Learning Techniques', Procedia Computer Science, 174, pp. 433–442. Available at: https://doi.org/10.1016/j.procs.2020.06.111.
2. Piao, Y., Chen, A. and Shang, Z. (2019) 'Housing Price Prediction Based on CNN', in 2019 9th International Conference on Information Science and Technology (ICIST). 2019 9th International Conference on Information Science and Technology (ICIST), Hulunbuir, China: IEEE, pp. 491–495. Available at: https://doi.org/10.1109/ICIST.2019.8836731.
3. Phan, T.D. (2018) 'Housing Price Prediction Using Machine Learning Algorithms: The Case of Melbourne City, Australia', in 2018 International Conference on Machine Learning and Data Engineering (iCMLDE). 2018 International Conference on Machine Learning and Data Engineering (iCMLDE), Sydney, Australia: IEEE, pp. 35–42. Available at: https://doi.org/10.1109/iCMLDE.2018.00017.
4. Zhang, Q. (2021) 'Housing Price Prediction Based on Multiple Linear Regression', Scientific Programming, 2021, p. e7678931. Available at: https://doi.org/10.1155/2021/7678931.

5. Zhan, C. et al. (2020) 'Housing prices prediction with deep learning: an application for the real estate market in Taiwan', in 2020 IEEE 18th International Conference on Industrial Informatics (INDIN). 2020 IEEE 18th International Conference on Industrial Informatics (INDIN), Warwick, United Kingdom: IEEE, pp. 719–724. Available at: https://doi.org/10.1109/INDIN45582.2020.9442244.

6. Yu, L. et al. (2018) 'Prediction on Housing Price Based on Deep Learning', International Journal of Computer and Information Engineering, 12(2), pp. 90–99.

7. Xue, C. et al. (2020) 'Research on the Sustainable Development of Urban Housing Price Based on Transport Accessibility: A Case Study of Xi'an, China', Sustainability, 12(4), p. 1497. Available at: https://doi.org/10.3390/su12041497.

8. Varma, A. et al. (2018) 'House Price Prediction Using Machine Learning and Neural Networks', in 2018 Second International Conference on Inventive Communication and Computational Technologies (ICICCT). 2018 Second International Conference on Inventive Communication and Computational Technologies (ICICCT), Coimbatore: IEEE, pp. 1936–1939. Available at: https://doi.org/10.1109/ICICCT.2018.8473231.

9. Sawant, R. et al. (2018) 'Comprehensive Analysis of Housing Price Prediction in Pune Using Multi-Featured Random Forest Approach', in 2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA). 2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA), Pune, India: IEEE, pp. 1–5. Available at: https://doi.org/10.1109/ICCUBEA.2018.8697402.

10. Park, B. and Bae, J.K. (2015) 'Using machine learning algorithms for housing price prediction: The case of Fairfax County, Virginia housing data', Expert Systems with Applications, 42(6), pp. 2928–2934. Available at: https://doi.org/10.1016/j.eswa.2014.11.040.

11. Mu, J., Wu, F. and Zhang, A. (2014) 'Housing Value Forecasting Based on Machine Learning Methods', Abstract and Applied Analysis, 2014, pp. 1–7. Available at: https://doi.org/10.1155/2014/648047.

12. Manrique, M.A.C. et al. (2021) 'Housing-Price Prediction in Colombia using Machine Learning'. OSF Preprints. Available at: https://doi.org/10.31219/osf.io/w85z2.

13. Komagome-Towne, A. (no date) 'Models and Visualizations for Housing Price Prediction'.

14. Fan, C., Cui, Z. and Zhong, X. (2018) 'House Prices Prediction with Machine Learning Algorithms', in Proceedings of the 2018 10th International Conference on Machine Learning and Computing. ICMLC 2018: 2018 10th International Conference on Machine Learning and Computing, Macau China: ACM, pp. 6–10. Available at: https://doi.org/10.1145/3195106.3195133.

15. Banerjee, D. and Dutta, S. (2017) 'Predicting the housing price direction using machine learning techniques', in 2017 IEEE International Conference on Power, Control, Signals and Instrumentation Engineering (ICPCSI). 2017 IEEE International Conference on Power, Control, Signals and Instrumentation Engineering (ICPCSI), Chennai: IEEE, pp. 2998–3000. Available at: https://doi.org/10.1109/ICPCSI.2017.8392275.

16. Ming, Y. et al. (2020) 'Prediction and Analysis of Chengdu Housing Rent Based on XGBoost Algorithm', in Proceedings of the 2020 3rd International Conference on Big Data Technologies. ICBDT 2020: 2020 3rd International Conference on Big Data Technologies, Qingdao China: ACM, pp. 1–5. Available at: https://doi.org/10.1145/3422713.3422720.

17. Chen, Y. et al. (2016) 'Mapping the fine-scale spatial pattern of housing rent in the metropolitan area by using online rental listings and ensemble learning', Applied Geography, 75, pp. 200–212. Available at: https://doi.org/10.1016/j.apgeog.2016.08.011.

18. Zhou, X., Tong, W. and Li, D. (2019) 'Modeling Housing Rent in the Atlanta Metropolitan Area Using Textual Information and Deep Learning', ISPRS International Journal of Geo-Information, 8(8), p. 349. Available at: https://doi.org/10.3390/ijgi8080349.

19. Hu, L. et al. (2019) 'Monitoring housing rental prices based on social media:An integrated approach of machine-learning algorithms and hedonic modeling to inform equitable housing policies', Land Use Policy, 82, pp. 657–673. Available at: https://doi.org/10.1016/j.landusepol.2018.12.030.

20. Neloy, A.A., Haque, H.M.S. and Ul Islam, Md.M. (2019) 'Ensemble Learning Based Rental Apartment Price Prediction Model by Categorical Features Factoring', in Proceedings of the 2019 11th International Conference on Machine Learning and Computing. ICMLC '19: 2019 11th International Conference on Machine Learning and Computing, Zhuhai China: ACM, pp. 350–356. Available at: https://doi.org/10.1145/3318299.3318377.

21. Müller Andreas C., and Sarah Guido. Introduction to Machine Learning with Python: A Guide for Data Scientists. O'Reilly Media, Inc., 2018.