DrillBit

The Report is Generated by DrillBit Plagiarism Detection Software

Submission Information

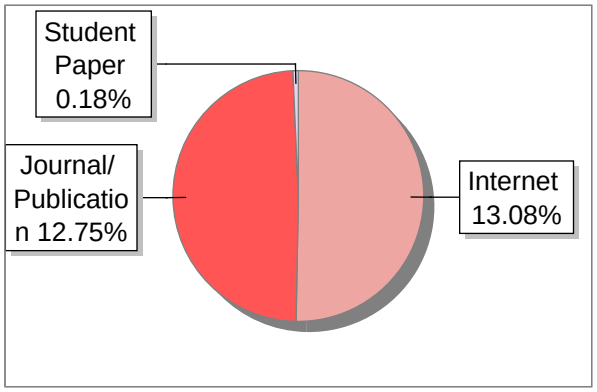
Author Name	Bhuvika A M
Title	property price pridiction
Paper/Submission ID	741613
Submission Date	2023-05-13 18:46:15
Total Pages	50
Document type	Project Work

Result Information

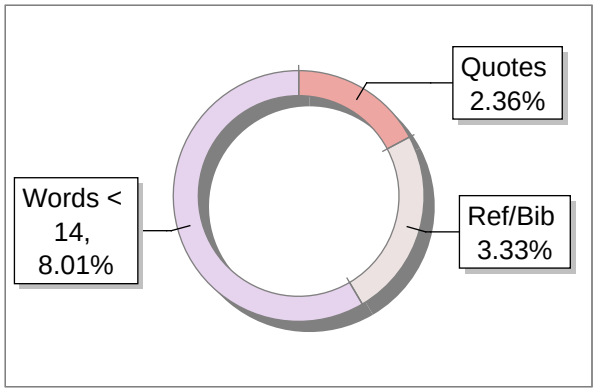
Similarity **26 %**

1 10 20 30 40 50 60 70 80 90

Sources Type



Report Content



Exclude Information

Quotes	Excluded
References/Bibliography	Excluded
Sources: Less than 14 Words Similarity	Excluded
Excluded Source	**0 %**
Excluded Phrases	Not Excluded

A Unique QR Code use to View/Download/Share Pdf File

DrillBit

DrillBit Similarity Report

26	43	B	A-Satisfactory (0-10%) B-Upgrade (11-40%) C-Poor (41-60%) D-Unacceptable (61-100%)
SIMILARITY %	MATCHED SOURCES	GRADE	

LOCATION	MATCHED DOMAIN	%	SOURCE TYPE
1	expertsystem.com	3	Internet Data
2	pdfcoffee.com	2	Internet Data
3	Using machine learning algorithms for housing price prediction The case of Fair by Park-2015	2	Publication
4	House Price Prediction using Machine Learning Algorithm - The Case of Karachi Ci by Ahtesham-2020	2	Publication
5	Predicting property prices with machine learning algorithms by Ho-2020	2	Publication
6	Housing Price Prediction via Improved Machine Learning Techniques by Truong-2020	2	Publication
7	www.researchgate.net	1	Internet Data
8	IEEE 2020 International Conference on Electronics and Sustainable Communicatio by Jain-2020	1	Publication
9	www.meninavaidosacosmeticos.com.br	1	Internet Data
10	en.wikipedia.org	1	Internet Data
11	IEEE 2013 IEEE International Conference on Big Data - Silicon Valley by	1	Publication
12	cse.anits.edu.in	1	Publication

13	www.tandfonline.com	1	Internet Data
14	matplotlib.org	1	Internet Data
15	cse.anits.edu.in	<1	Publication
16	avxlive.icu	<1	Internet Data
17	gecgudlavalleru.ac.in	<1	Publication
18	people.cs.umass.edu	<1	Publication
19	dspace.daffodilvarsity.edu.bd 8080	<1	Publication
20	Customer Churn Prediction In Telecommunication Industry Using Random Forest Clas by Geetha-2020	<1	Publication
21	www.rvitm.edu.in	<1	Internet Data
22	moam.info	<1	Internet Data
23	www.simplilearn.com	<1	Internet Data
24	ijitee.org	<1	Publication
25	www.ijrte.org	<1	Publication
26	researchers.mq.edu.au	<1	Internet Data
28	gecgudlavalleru.ac.in	<1	Publication
29	moam.info	<1	Internet Data
30	The main causes of risk in residential real estate projects by Pereira-2020	<1	Publication
31	citeseerx.ist.psu.edu	<1	Publication
32	www.jetbrains.com	<1	Internet Data

33	A Data-Driven Approach for Driving Safety Risk Prediction Using Driver Behavior by Arbabzadeh-2017	<1	Publication
34	www.projectpro.io	<1	Internet Data
35	cse.anits.edu.in	<1	Publication
36	diversityhealthcare.imedpub.com	<1	Publication
37	sjcit.ac.in	<1	Publication
38	18.182.206.220	<1	Internet Data
39	bjp.org.br	<1	Internet Data
41	jtdo.velaccessibile.it	<1	Internet Data
42	moam.info	<1	Internet Data
43	pdfcookie.com	<1	Internet Data
44	www.internationaljournalssrg.org	<1	Publication
45	Hazardous Traffic Event Detection Using Markov Blanketand Sequential Minimal Optimization(M B- S M O)	<1	Student Paper

Chapter-1

INTRODUCTION

Machine Learning is a field of Artificial Intelligence which enables PC frameworks to learn and improve in execution with the assistance of information. It is used to study the construction of algorithms that make predictions on data. The evolve of Internet and better hardware and software has opened many doors such as we can now use Machine Learning to do prediction here houses so that it can address all the concern of buyers and the sellers.

House is a basic necessity for a person and their prices vary from location to location based on the facilities available like parking space, locality, etc. Property investment has also increased significantly. Buying of a house is one of the greatest and significant choice of a family as it expands the entirety of their investment funds and now and again covers them under loans. It is the difficult task to predict the accurate values of house pricing. Our proposed model would make it possible to predict the exact prices of houses. Housing price trends are not only the concern of buyers and the sellers, it also indicates the current economic situation. There are many factors which has impact on house prices, such as location, BHK, floor etc. Also, a location with a great accessibility to highways, expressways, schools, shopping malls and local employment opportunities contributes to the rise in house price. Manual house prediction becomes difficult, hence there are many systems developed for house price prediction. The aim of this system is to create a website through which the user can give his house requirements as input which is then passed on to the most accurate model among Naïve Bayes, Linear Regression with SVM. Random Forest Regression, for predicting the house price. The website allows user to predict the house prices to a particular place, price-range and other amenities as specified by the user.

1.1 About Domain

Machine learning is an application of artificial intelligence (AI) that provides systems the ability to automatically learn and improve from experience without being explicitly programmed. Machine learning focuses on the development of computer programs that can access data and use it learn for themselves.

The process of learning begins with observations or data, such as examples, direct experience, or instruction, in order to look for patterns in data and make better decisions in the future based on the examples that we provide. The primary aim is to allow the computers learn automatically without human intervention or assistance and adjust actions accordingly.

Some machine learning methods

Machine learning algorithms are often categorized as supervised or unsupervised.

Supervised machine learning algorithms can apply what has been learned in the past to new data using labeled examples to predict future events. Starting from the analysis of a known training dataset, the learning algorithm produces an inferred function to make predictions about the output values. The system is able to provide targets for any new input after sufficient training. The learning algorithm can also compare its output with the correct, intended output and find errors in order to modify the model accordingly.

Unsupervised machine learning algorithms are used when the information used to train is neither classified nor labeled. Unsupervised learning studies how systems can infer a function to describe a hidden structure from unlabeled data. The system doesn't figure out the right output, but it explores the data and can draw inferences from datasets to describe hidden structures from unlabeled data.

Semi-supervised machine learning algorithms fall somewhere in between supervised and unsupervised learning, since they use both labeled and unlabeled data for training – typically a small amount of labeled data and a large amount of unlabeled data. The systems that use this method are able to considerably improve learning accuracy. Usually, semi-supervised learning is chosen when the acquired labeled data requires skilled and relevant resources in order to train it / learn from it. Otherwise, acquiring unlabeled data generally doesn't require additional resources.

Reinforcement machine learning algorithms is a learning method that interacts with its environment by producing actions and discovers errors or rewards. Trial and error search and delayed reward are the most relevant characteristics of reinforcement learning. This method

allows machines and software agents to automatically determine the ideal behavior within a specific context in order to maximize its performance. Simple reward feedback is required for the agent to learn which action is best; this is known as the reinforcement signal.

Machine learning enables analysis of massive quantities of data. While it generally delivers faster, more accurate results in order to identify profitable opportunities or dangerous risks, it may also require additional time and resources to train it properly. Combining machine learning with AI and cognitive technologies can make it even more effective in processing large volumes of information.

1.2 Overview

Ridge and Lasso are powerful techniques generally used for creating parsimonious models in presence of a 'large' number of features. Here 'large' can typically mean either of two things:

1. Large enough to enhance the tendency of a model to overfit (as low as 10 variables might cause overfitting)
2. Large enough to cause computational challenges. With modern systems, this situation might arise in case of millions or billions of features

Though Ridge and Lasso might appear to work towards a common goal, the inherent properties and practical use cases differ substantially. If you've heard of them before, you must know that they work by penalizing the magnitude of coefficients of features along with minimizing the error between predicted and actual observations. These are called 'regularization' techniques. The key difference is in how they assign penalty to the coefficients:

1. **Ridge Regression:**

    - Performs L2 regularization, adds penalty equivalent to square of the magnitude of coefficients
    - Minimization objective = LS Obj + α * (sum of square of coefficients)

2. **Lasso Regression:**

    - Performs L1 regularization, adds penalty equivalent to absolute value of the magnitude of coefficients
    - Minimization objective = LS Obj + α * (sum of absolute value of coefficients)

1.3 Objectives

- Predict the house price of different localities.
- Using different models in terms of minimizing the difference between predicted and actual rating.
- Comparing the accuracy of the algorithms to get the most accurate result.
- Recommending houses based on the user's search pattern

1.4 Problem Statement

The goal is to predict the house price with considering the different parameters as entered by the customer and build a house recommendation system based on the search pattern.

1.4.1 Existing System features

- Existing system does not consider taxes but only considers location and other features which makes a huge difference in accurate prediction.
- Using other regression techniques gives comparatively low accuracy.

The existing models have used classification algorithms that predicts based on location and do not include external features.

1.4.2 Proposed System features

- Proposed system mainly concentrates on tax price prediction in different localities.
- Increasing the accuracy of the model than the existing one.
- Combining of machine learning techniques decreases the large variation in the price prediction compared to actual price.

1.5 Organization of Report

- **Chapter:2, Literature survey-** An overview of 10 different literature papers on the same topic written by different authors.
- **Chapter:3, Summary table-** An overall summary of the 10 different literature papers mentioned in the above section.
- **Chapter:4, System requirements-** This section contains hardware, software requirements along with functional and nonfunctional requirements.
- **Chapter:5, System design-** Contains the software architecture, data-flow diagram and system design of the project.
- **Chapter:6, Possible outcomes-** This section contains the possible outcomes and references of the project.

Chapter-2

Literature Survey

1. [2020] Machine Learning based Predicting House Prices using Regression Techniques, Manasa J, Radha Gupta, Narahari N S

The primary location that is taken into consideration is Bengaluru, the prospective home buyer considers several factors such as location, size of the land, proximity to parking schools, hospitals, power generation facilities and most importantly the house price gets affected by all these factors. Five prediction models had taken into consideration-ordinary least squares model, Lasso and Ridge regression models, SVR model, and also XG Boost regression model. Authors evaluated model's performance by using metrics: the coefficient of determination, 2 R adjusted 2 R and RMSE (Root Means Square Error), RMLSE (Root Mean Squared Logarithmic Error). The lower RMSE values are indicative of a better fit model and the higher the R-squared, the better the model fits the data. For linear regression model R-square value obtained is 0.418(train set), -2.12 (test set) and RMSE value obtained is 0.0912(train set), 0.2077(test set). For ridge regression model R-square value obtained is 0.4345 (train set), 0.4345 (test set) and RMSE value obtained is 0.5415 (train set), 0.5224 (test set). For lasso regression model R-square value obtained is 0.4341 (train set), 0.4430 (test set) and RMSE value obtained is 0.5416 (train set), 0.5224 (test set). For support vector regression model R-square value obtained is 0.799 (train set), 0.6630 (test set) and RMSE value obtained is 0.0256 (train set), 0.0317(test set). For XG Boost regression model R-square value obtained is 0.7868(train set), 0.7584(test set) and RMSE value obtained is 0.3309 (train set), 0.3462 (test set). Ridge regression model gave the better accuracy than the other models.

2. [2021] House Price Prediction using Machine Learning, Anand G. Rawool, Dattatary V. Rogye, Sainath G. Rane, DR. Vinayak A. Bharadi

In this project, Lucknow is the primary location and the house price prediction of the house is done using different Machine Learning algorithms like Leaner Regression, Decision Tree Regression, K-Means Regression and Random Forest Regression. 80% of data from the dataset is used for training purpose and remaining 20% of data used for testing purpose. The work applies various techniques such as features, labels, reduction techniques and transformation techniques such as attribute combinations, set missing attributes as well as looking for new correlations. First, Data Processing is implemented. In this phase, the missing attribute is handle by using mean value. The target is feature is drop out. By using Pandas library, the operation is performed. For visualization of dataset graph use Matplotlib python function. After that author try to catch some attribute combination and set the missing values. Once data processing done, then created suitable pipeline for execution of model, then missing attributes are being filled. Final RMSE-2.9131988953. Out of all the Random Forest is predict better accuracy than other models.

3. [2020] House Price Prediction Modeling using Machine Learning, Dr. M. Thamarai, Dr. S P. Malarvizhi

These model helps the customer of west Godavari district of Andra Pradesh to purchase a house suitable for their need. Proposed work makes use of the attributes or features of the houses such as number of bedrooms available in the house, age of the house, travelling facility from the location, school facility available nearby the houses and shopping malls available nearby the house location. The work involves decision tree classification, decision tree regression and multiple linear regression and is implemented using Scikit-Learn Machine Learning Tool. Decision Tree Classifier is used to predict the availability of houses as per the users' requirement constraints and it produces responses like yes or no respectively to tell whether a house is available or not and Decision tree regression and Multiple Linear Regression methods are used to predict the prices of the houses. The performance metrics of the decision tree regression model- Mean Absolute Error: 2.125, Mean Squared Error: 6.625, Root Mean Squared

Error: 2.57390753524675.The performance metrics for multiple linear regression model- Mean Absolute Error: 1.9527234112192413, Mean Squared Error: 6.0653477870232635, Root Mean Squared Error: 2.462792680479472. The performance of the decision tree classifier is measured.

4. [2019] Prediction of House Pricing using Machine Learning with Python, Namit Jain, Parikshay Goel, Purushottam Sharma, Vikas Deep

The dataset consists of data of Ames, lowa (United States) at several locations and the factors affecting the house prices such as number of floors, garden area, total area, carpet area, utilities available etc. to a total of 79 factors that affect the house prices. First Data Cleaning has done, Data cleaning is the process of detecting and correcting inaccurate records from a record set, table or database. It is the process of identifying incomplete data and then replacing the dirty data. The data is altered to make sure that it is accurate and correct. It is used to make a dataset consistent. The main goal of data cleaning is to detect and remove errors to increase the value of data in decision making. Then 5 data models are deployed one by one-Lasso Regression, Elastic Net Regression, Kernel Ridge Regression, Gradient Boost Regression, XG Boost and found their RMSE values. To improve the accuracy of the algorithms, authors took a simple stacking approach which begin with averaging base models first. Then took 4 base algorithms and put them in the class to find out their average stacked model accuracy. After averaging the base models, stacked them together to improve their accuracy and to get a more reliable outcome. Average base model score=0.2407(0.0634) and Stacking Average model score-0.1087(0.0070). By this the sales prices have been calculated with better accuracy and precision.

5. [2020] Using Machine Learning Algorithms for Housing Price Prediction: The Case of Fairfax County, Virginia Housing Data, Byeonghwa, and Jae Kwon Bae.

House sales are determined based on the Standard & Poor's Case-Shiller home price indices and the housing price index of the Office of Federal Housing Enterprise Oversight (OFHEO). These reflect the trends of the US housing market. In addition to these housing price indices, the development of a housing price prediction model can greatly assist in the prediction of future housing prices and the establishment of real estate policies. This study uses machine learning algorithms as a research methodology to develop a housing price prediction model. To improve the accuracy of housing price prediction, this paper analyzes the housing data of 5359 townhouses in Fairfax County, Virginia, gathered by the Multiple Listing Service (MLS) of the Metropolitan Regional Information Systems (MRIS). By developing a house price prediction model based on machine learning algorithms such as C4.5, RIPPER, Naïve Bayesian, and AdaBoost and compare their classification accuracy performance. Here they propose an improved housing price prediction model to assist a house seller or a real estate agent make better informed decisions based on house price valuation. The experiments demonstrate that the RIPPER algorithm, based on accuracy, consistently outperforms the other models in the performance of housing price prediction with an error rate of 25% whereas Naïve Bayesian has the highest error rate of 30% and Decision Tree, AdaBoost also has their error rate of 27% and 26% respectively.

6. [2020] Housing Price Prediction via Improved Machine Learning Techniques, Truong, Quang, Minh Nguyen, Hy Dang, and Bo Mei.

House Price Index (HPI) is commonly used to estimate the changes in housing price. Since housing price is strongly correlated to other factors such as location, area, population, it requires other information apart from HPI to predict individual housing price. "Housing Price in Beijing" is a dataset containing more than 300, 000 data with 26 variables representing housing prices traded between 2009 and 2018. These variables, which served as features of the dataset, were then used to predict the average price. This paper will apply both traditional and advanced machine learning approaches

to investigate the difference among several advanced models. The paper will also comprehensively validate multiple techniques like XGBoost, Random Forest, Light GBM, Hybrid Regression in model implementation, compare and provide an optimistic result for housing price prediction. The best results belong to Random Forest with a RMSLE value of 0.12980 for the training set whereas the RMSLE value for Extreme Gradient Boosting 0.16118 Light Gradient Boosting Machine 0.16687 Hybrid Regression 0.14969 Stacked Generalization Regression 0.16404 and the for the test set, Stacke Generalization Regression with an RMSLE value of 0.16350 whereas Random Forez 0.16568.

7. [2019] Predicting Sale Prices of the Houses using Regression Methods of Machine Learning, Viktorovich, P.A., Aleksandrovich, P.V., Leopoldovich, K.I. and Vasilevna

They aimed to apply data imputation, feature engineering and machine learning modelling to achieve a better predictive accuracy on the housing price. Each house description consists of such attributes as house area (called 'GrLivArea'), garage capacity('GarageCars'), overall quality estimation of house and kitchen ('OverallQual','KitchenQual'), distinct of the city (MS_Zoning), data about neighborhood, type of sale (*SaleType'), year of building (YearBuilt'), and another similar attributes. The training set also has the sale price as response while the test set does not. Solutions are evaluated on Root-Mean-Squared-Error (RMSE) between the logarithm of the predicted value and the logarithm of the observed sales price. (Taking logs means that errors in predicting expensive houses and cheap houses will affect the result equally). Our solution used Python programming language with Pandas, NumPy, Sklearn and XGBoost libraries. Cross-Validation Results for some variants of our solutions: Algorithm, CV score and CV score standard deviation- Lasso 0.11139 0.0106, XGBoost 0.13058 0.0108, XGBoost with logit transform 0.12986 0.0107, ElasticNet 0.11203 0.0107, Neural network 0.11787.

8. [2020] Predicting Property Prices with Machine Learning Algorithms, Ho, W.K, Tang,B.S. and Wong, S.W

This study uses three machine learning algorithms including, support vector machine (SVM), random forest (RF) and gradient boosting machine (GBM) in the appraisal of property prices. It applies these methods to examine a data sample of about 40,000 housing transactions in a period of over 18 years in Hong Kong, and then compares the results of these algorithms. In terms of predictive power, RF and GBM have achieved better performance when compared to SVM. The three-performance metrics including mean squared error (MSE), root mean squared error (RMSE) and mean absolute percentage error (MAPE) associated with these two algorithms also unambiguously outperform those of SVM. However, our study has found that SVM is still a useful algorithm in data fitting because it can produce reasonably accurate predictions within a tight time constraint. Our conclusion is that machine learning offers a promising, alternative technique in property valuation and appraisal research especially in relation to property price prediction. Estimated results based on SVM, RF and GBM: R2, MSE, RMSE and MAPE- Support Vector Machine 0.82715 0.01422 0.11925 0.54467%, Random Forest 0.90333 0.00795 0.08918 0.32270% and Gradient Boosting Machine 0.90365 0.00793 0.08903 0.32251%.

9.[2019] Housing Price Prediction based on CNN, Piao, Y Chen A. and Shang Z Housing price has been one of the most concerned issues to the public all over the world.

The excessive growth of housing price will affect not merely the quality of life, but also he business cycle dynamics. However, the factors influencing residential real estate prices are complex and the selection of effective features is vague, which leads to a lower accuracy in many of the traditional housing price prediction approaches. Accordingly, a novel prediction model based on CNN is proposed for prediction of housing price as well as the process of feature selection. Compared with other traditional methods, our work can obtain a better performance through experiments using actual data of property transaction. XGBoost model and GM model as the two baseline models to compare the effectiveness.The parameters of XGBoost model are

optimized by grid searching and its accuracy is evaluated through cross checking method. The results of the three methods are provided respectively. COMPARISON OF EXPERIMENTAL RESULTS: Prediction algorithm mean-square error and Accuracy 1) GM Model 0.435 90 % 2) XGBoost Model 0.1096.5% 3) CNN model 0.01057 98.68%.

10. [2021] House Price Prediction using Machine Learning Algorithm - The Case of Karachi City, Pakistan, Maida Ahtesham, Narmeen Zakaria Bawany, Kiran Fatima.

House prices are a significant impression of the economy, and its value ranges are of great concerns for the clients and property dealers. Housing price escalate every year that eventually reinforced the need of strategy or technique that could predict house prices in future. Machine learning has been used for image recognition, spam reorganization, medical diagnosis for more than a decade. Machine Learning basedpredictions achieve better results when put in practice. Almost every economic domainnow benefits from machine learning prediction models. In this research paper, Houseprice prediction has been performed using machine learning technique XGBoost. Areal state property contains various factors. In order to predict house prices, machinelearning algorithms are considered to be efficient techniques. This paper providesinsight of XGBoost algorithm as one of the useful algorithm for house price predictionand to provide flexible and efficient results. Dataset used for the experiment wereobtained from Open Data Pakistan website. A real state property data set of Karachicity that contain 38,961 records and features including location, city, property type(Flat, House), number of bedrooms, baths, longitude, latitude, area, price. Train testdata of the work is as follows: Train/Test Ratio, Learning Rate, Model Accuracy, Mean,Absolute Error- 1, 60/40, 0.01, 98%,22502.0824694; 2, 50/50, 0.01, 98%,
22502.0824694; 3, 70/30, 0.01, 98%
22502.0824694.

Chapter 3

Summary Table

SL.NO	Title of the Paper	Problem Addressed	Authors Approach/ Method	Results
1	Machine Learning based Predicting House Prices Using Regression Techniques [2020]	Predictive models for determining the sale price of houses in cities like Bangalore is still remaining as more challenging and trickier task.	An analytical approach has been carried out by considering the data set of only Bangalore city housing the available housing properties using various regression techniques.	Both time and frequency domain features can be extracted for training the classifiers.
2	Prediction of House Pricing Using Machine Learning with Python [2019].	The objective is to predict house price based on various features on given dataset of Lucknow.	To fix this problem and to achieve greater accuracy author used a simple stacking algorithm to improve the results produced by different algorithms	Using the simple stacking algorithm, the accuracy of the outputs was improved
3	House Price Prediction Using Machine Learning [2021]	The objective is to predict house price based on various features on given dataset of Lucknow.	Machine learning algorithms like Linear Regression, Decision Tree Regression, K-Means and Random Forest Regression used here.	Random Forest Regression gives maximum accuracy.

SL.NO	Title of the Paper	Problem Addressed	Authors Approach/ Method	Results
4	House Price Prediction Modelling Using Machine Learning [2020]	The proposed work and the model are constructed for as small town in West Godavari district of Andhra Pradesh.	The work involves decision tree classification, decision tree regression and multiple linear regression and is implemented using Scikit-Learn Machine Learning Tool.	Performance of multiple linear regressions is better than decision tree regression for predicting house price.
5	Using Machine Learning Algorithms for housing price Prediction :The case of Fairfax County , Virginia housing data	To analyze the housing data of 5359 town houses in Fairfax County, Virginia ,gathered by the Multiple Listing Service (ML S] of the Metro Politian Regional Information Systems (MRIS)	In this study, several machine learning algorithms are used to develop a prediction model for housing prices C4.S, RIPPER. Naive Bayesian and AdaBoost and Campare their accuracy.	The experiments demonstrate that the RIPPER algorithm, based on accuracy. Consistently out performs the other models in the performance of housing price prediction
6	House price prediction using machine learning Algorithms: The Case of Melbourne City, Australia,[2019].	Machine learning techniques are applied to analyze prediction and Recommendation using Machine Learning property transactions in Australia particularly Melbourne City to discover useful models for house buyers and sellers.	Different regression models are implemented to Historical property transaction in Australia particularly Melbourne City to discover useful model for house buyers and sellers.	The combination of stepwise and tuned SVM .Which produces the lowest error on this dataset, is most competitive models.

7	Predicting Sale Prices of the Houses using Regression Methods of Machine Learning.[2020]	Predicting house sales price by attributes like house area, year of building etc.	Here different techniques of machine leaning like lasso, XG Boost, Elastic net, Neural Network etc. are applied.	According to author the best way to improve result is to train more base models and one of the ways to achieve model's decorrelation is to learn same models on different feature sets.
8	House price prediction based on CNN [2019]	Most of the Web application scanners can not detect the security problems with HTML5 features.	Customizing scanning modules and scripts and designing a Web application security scanning service.	Detection of vulnerabilities brought by HTML5 and providing efficient Web application security scanning and evaluation services.
9	Predicting sale prices of the houses using regression methods of machine learning.	Predicting house sales price by attributes like house area year of building.	Here , different techniques of machine leaning like lasso , XGBoost, Elastic net, Neural Network etc are applied.	Performance of multiple linear regressions is better than decision tree regression for predicting house price.
10	Prediction of House Pricing Using Machine Learning with Python[2020]	It provides an overview about how to predict house costs utilizing different regression methods with the assistance of python libraries.	Various Python libraries such as Pandas, NumPy, SciPy, Matplotlib etc were used. It supports various packages such as Xlsx Writer and XI Rd.	prediction. Stacking algorithm was used to improve the accuracies of the various regression algorithms.

Chapter 4

SYSTEM REQUIREMENT

4.1 FUNCTIONAL REQUIREMENTS

- Residential property price prediction model can help the buyer determine the buyingprice of the property based on their requirement.
- This can help the customer to arrange money to purchase a house.
- This also helps to compare prices of the property in different locations.
- Require accurate prediction.
- Adaptable for all data set.
- The collected data should be cleaned and preprocessed to handle missing values, outliers, and inconsistencies. This process may involve data normalization, feature scaling, and handling categorical variables.
- The system should perform feature selection to identify the most important variables that impact house prices. Additionally, it may involve creating new features based on domain knowledge or data transformation techniques.
- The system should be able to train a predictive model using the collected and preprocessed data. This may involve using machine learning algorithms such as lasso regression and ridge regression.
- The system should evaluate the performance of the trained model using appropriate metrics, such as mean squared error (MSE), root mean squared error (RMSE), or coefficient of determination (R-squared). Cross-validation techniques can be employed to assess the model's generalization ability.
- Once the model is trained and evaluated, the system should allow users to input relevant features of a house (e.g., number of bathrooms, number of bhk, square footage, location) and generate a predicted price based on the trained model.
- The system should have a user-friendly interface that allows users to interact with the prediction model easily. This can be in the form of a web application.
- The system can be designed to periodically update the predictive model with new data to ensure the predictions remain accurate and up to date.

- The predictive model should provide accurate house price estimates based on the available data. The level of accuracy should be defined, and the system should strive to achieve it.

4.2 NON-FUNCTIONAL REQUIREMENTS

- Maintaining and updating the large data to achieve accuracy.
- Algorithm to achieve accuracy in prediction.
- Achieving the speed of working of the application.
- The system should be able to handle large datasets and perform predictions within a reasonable time frame. It should be capable of processing multiple requests concurrently without significant delays.
- The predictive model should provide accurate house price estimates based on the available data. The level of accuracy should be defined, and the system should strive to achieve it.
- The system should be reliable and available for use whenever needed. It should minimize downtime and handle errors gracefully to ensure uninterrupted service.
- The system should be designed to handle increasing data volumes and user load. It should be scalable to accommodate growth and future requirements without compromising performance.
- The user interface should be intuitive, easy to navigate, and user-friendly. Users should be able to input relevant features and obtain predicted prices without needing extensive technical knowledge.
- The system should be designed and implemented in a way that facilitates maintenance and updates. It should have well-structured code, clear documentation, and follow coding best practices to allow for easy maintenance by developers.
- The system should be portable across different platforms and environments. It should be compatible with various operating systems, databases, and web browsers.
- The system should make efficient use of computational resources such as memory, CPU, and storage. It should optimize the processing and prediction algorithms to minimize resource usage.
- The system should ensure the quality and integrity of the input data. It should include data validation mechanisms and handle data anomalies or inconsistencies appropriately.
- These non-functional requirements contribute to the overall usability, performance, reliability, and security of the house price prediction system, ensuring that it meets the expectations and needs of its users.

4.3 Hardware requirements

- Processor : core i3.
- Hard Disk : 250GB. ·
- RAM : 4GB
- Mouse : Optical
- Keyboard : Multimedia

4.4 Software Requirements

- Anaconda 2018.12
- Jupyter notebook
- Pycharm

Chapter – 5

SYSTEM DESIGN

5.1 Block Diagram

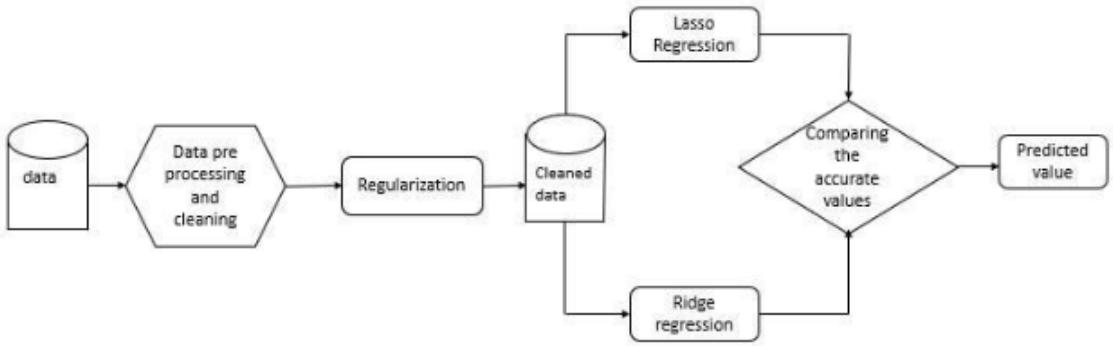


Fig 5.1 : Block Diagram for House Price Prediction

- **Data:** the collected data set for residential property price prediction with location values and price per sq feet
- **Data preprocessing and cleaning:** clean and pre-process the data by removing noise, null values and repeated values.
- **Regularization:** this includes feature selection which selects the mainly used features.
- **Lasso regression:** apply lasso regression to the cleaned data set to obtain accuracy
- **Ridge regression:** apply ridge regression to the dataset to achieve more accuracy.
- **Comparing the accuracy:** compare the accuracy of both lasso and ridge regression technique
- **Predict value:** after comparing the most accurate regression model to predict the accuracy

5.2 METHODOLOGY

- Regression algorithms are used to predict the continuous value which enhances the prediction accuracy.
- On the model building phase, we apply three methods Linear regression, Lasso regression and Ridge regression.The data preprocessing phase is a crucial step that involves preparing the raw data for regression analysis. This phase typically consists of the following steps

Data cleaning Phase: Identify and handle missing values, outliers, and errors in the dataset. Missing values can be imputed using techniques such as mean imputation, regression imputation, or using a machine learning algorithm like k-nearest neighbors (KNN). Outliers can be detected and treated by methods like z-score or interquartile range (IQR) filtering.

Feature Extraction: Feature extraction in the context of house price prediction involves transforming the original set of features into a new set of features that captures the most relevant information for the task at hand. Unlike feature reduction, which selects a subset of existing features, feature extraction creates new features by combining or representing the original features in a different way. Here are some common techniques for feature extraction in house price prediction:

Feature Reduction/Feature Selection:

Feature reduction or feature selection is a process of selecting a subset of relevant features from a larger set of available features in order to improve the efficiency and effectiveness of a regression model for house price prediction. Feature reduction can be beneficial in the following ways:

Improved Model Performance: Including irrelevant or redundant features in the regression model may lead to overfitting and reduced generalization ability. By selecting only the most informative features, you can enhance the model's performance and prevent overfitting.

Interpretability: Having a smaller set of features makes it easier to interpret the relationship between the predictors and the house prices. It allows you to focus on the most influential factors and understand their impact on the target variable.

There are several approaches to feature reduction in house price prediction:

Correlation Analysis: Calculate the correlation between each feature and the target variable (house prices) and select features with high correlation values. Features with low correlation may not contribute significantly to the prediction task.

Regularization Techniques: Techniques like Lasso regression (L1 regularization) or Ridge regression (L2 regularization) can be employed to automatically shrink the coefficients of irrelevant or less important features towards zero, effectively reducing the impact of those features in the model.

Domain Knowledge: Leverage expert knowledge or domain expertise to identify and select features that are known to be relevant in the context of house price prediction. This approach can be particularly useful when dealing with specific domain-specific datasets.

Dimensionality Reduction: Techniques like Principal Component Analysis (PCA) or Linear Discriminant Analysis (LDA) can be used to transform the original set of features into a reduced set of orthogonal features while preserving the most important information. These methods are particularly useful when dealing with high-dimensional datasets.

Algorithms used:

1. **Linear Regression:** It is an algorithm of supervised machine learning in which the predicted output is continuous which is having a constant slope. It is used to predict the values in a continuous range instead of classifying the values in the categories. Linear regression is used for performing different tasks such as house price prediction.

2. **Lasso Regression:**

LASSO means Least Absolute Shrinkage and Selection Operator. Lasso Regression is one of the type of linear Regression which uses the technique of shrinkage. As the name suggest, it is a regression analysis method that performs both variable selection and regularization. Lasso

regression selects only a subset of the provided covariates for use in the final model.

When using lasso regression for house price prediction, you typically start by collecting a dataset that contains various features or variables that can potentially influence house prices. These features might include the number of bedrooms, square footage, location, bathrooms etc. The target variable would be the house price you want to predict.

Fit the lasso regression model using the training data, specifying the chosen alpha value. The model will automatically perform feature selection and assign coefficients to the remaining features. Evaluate the performance of the model using the test data. Common evaluation metrics for regression models include mean squared error (MSE), root mean squared error (RMSE), mean absolute error (MAE), and R-squared (coefficient of determination). If necessary, perform hyperparameter tuning to find the optimal value for alpha. This can be done using techniques like cross-validation, where the data is split into multiple folds to estimate the model's performance across different hyperparameter values.

3. Ridge Regression:

It is a tool for analysis of Multiple Regression on the data that have multicollinearity(mcl). Multicollinearity(mcl) is existence of near–linear relationships among the variables which are independent.

When Multicollinearity occurs, least squares estimates are unbiased. Ridge regression reduces the standard errors by adding a degree of bias to the regression estimates. The formula for Ridge regression is ,adding a degree of bias to the regression estimates. Ridge regression is an extension of ordinary least squares (OLS) regression that introduces a regularization term to the loss function. This term penalizes the magnitudes of the regression coefficients, encouraging them to be smaller.

By shrinking the coefficients, ridge regression helps mitigate the effects of multicollinearity and reduces the risk of overfitting. Ridge regression introduces a tuning parameter, commonly denoted as λ (lambda), which controls the amount of regularization applied. A larger λ value results in stronger regularization and greater coefficient shrinkage. The optimal value of λ can be determined through techniques such as cross-validation or using information criteria like Akaike information criterion (AIC) or Bayesian information criterion (BIC). The ridge regression equation is an extension of the ordinary linear regression equation, where the least squares loss function is modified to include the regularization term. The ridge regression

equation can be represented as: $y = \beta 0 + \beta 1x1 + \beta 2x2 + ... + \beta nxn + \varepsilon + \lambda\Sigma(\beta i^2)$

In this equation, y represents the dependent variable (house price), β0 to βn are the regression coefficients, x1 to xn are the independent variables (features), ε is the error term, and $\lambda\Sigma(\beta i^2)$ represents the regularization term.

Ridge regression is effective in dealing with highly correlated independent variables, which is common in house price prediction datasets. By shrinking the coefficients, it reduces the impact of multicollinearity on the regression model's stability and interpretability.

Ridge regression introduces regularization, which prevents the model from becoming too complex and overfitting the training data. This leads to better generalization performance when predicting house prices on unseen data.

Unlike feature selection techniques that discard some variables, ridge regression retains all the variables in the model while shrinking their coefficients. This can be useful when you want to retain all the available information without eliminating any features.

Ridge regression can be implemented using various statistical software packages or programming languages that offer regression functionality. Common implementations include Python with libraries such as scikit-learn, R with packages like glmnet, or MATLAB with the Statistics and Machine Learning Toolbox.

When using ridge regression for house price prediction, it is important to evaluate the model's performance. Common evaluation metrics include R-squared (coefficient of determination), adjusted R-squared, root mean squared error (RMSE), or mean absolute error (MAE). These metrics assess the goodness-of-fit and predictive accuracy of the ridge regression model.

The choice of the regularization parameter λ is crucial in ridge regression. It is important to find the optimal value that balances between fitting the training data well and avoiding excessive shrinkage. Cross-validation techniques can be employed to search for the best λ value or to perform a grid search over a range of λ values.

Chapter – 6

DESIGN

6.1 SYSTEM ARCHITECTURE

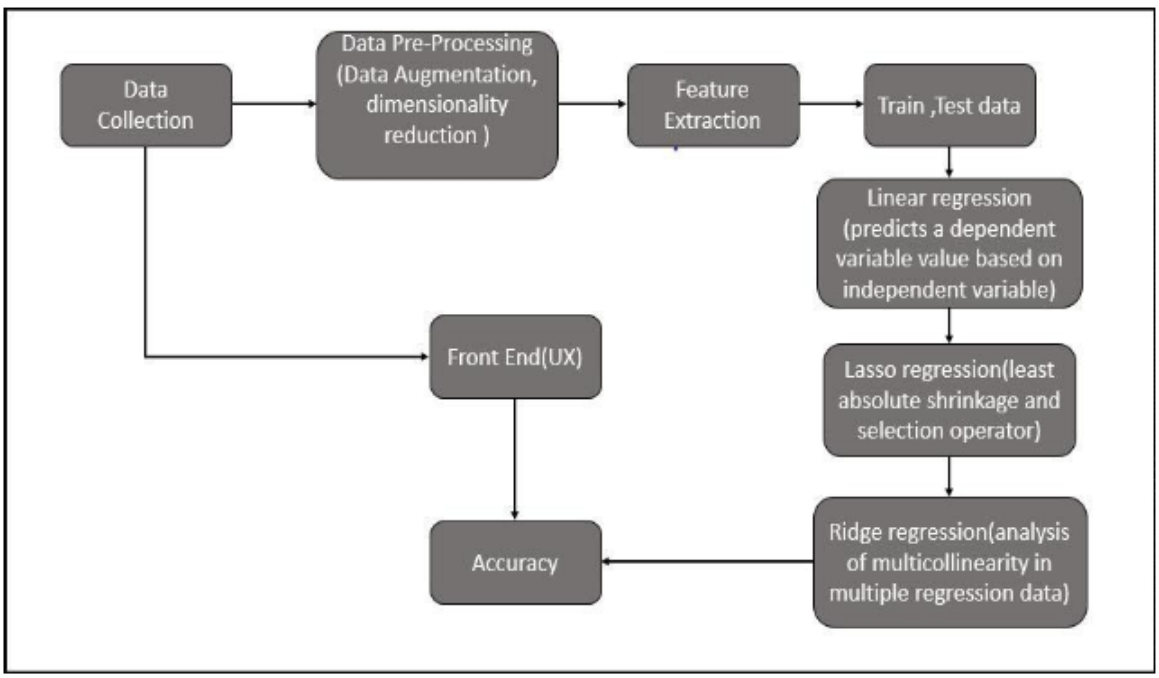


Fig 6.1 : System Architecture of House Price Prediction.

- System Architecture gives an overview of requirement collection and the machine learningtechniques best suited for prediction and accuracy.
- Collecting a accurate and error free data sets.

6.2 FLOW CHART

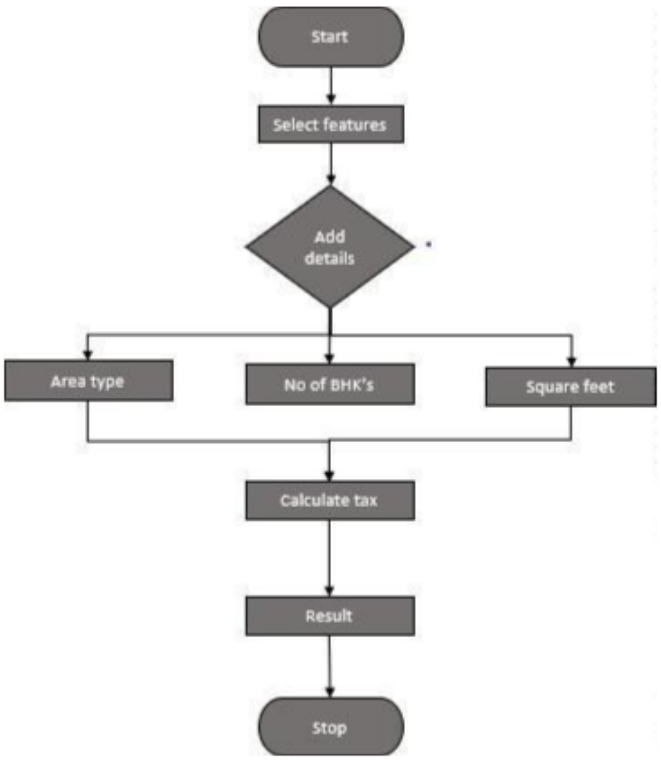


Fig 6.2 : Flow Chart Diagram for House Price Prediction

A flowchart for data acquisition in a machine learning project typically involves several key steps. The first step is to identify the data sources, which could include databases, APIs, web scraping, or manual data collection. Once the data sources have been identified, the next step is to retrieve the data and store it in a suitable format.

After the data has been collected and stored, the next step is to perform data cleaning, which involves removing any missing or erroneous data points, standardizing the format of the data, and resolving any inconsistencies or discrepancies in the data. This step is crucial in ensuring the quality and reliability of the data used for the machine learning project.

Select features: four parameters need to be selected for the accurate prediction.

Select Area: select the desired area from the drop down .

Number of BHK: Type the number of BHK's.

Square feet: Enter the square feet in the range.

calculate tax:Tax is calculated based on area type.

Result :price is displayed on the above parameters.

Once the data has been cleaned and analyzed, the next step is to split the data into training and testing sets. This is necessary to evaluate the performance of the machine learning model and prevent overfitting. The training set is used to train the model, while the testing set is used to evaluate its performance on unseen data.

Finally, the preprocessed data is fed into the machine learning model, which learns from the data and makes predictions or classifications based on its input. The output of the model can then be evaluated and refined as necessary to improve its accuracy and reliability.

Overall, the flowchart for data acquisition in a machine learning project is a critical step in ensuring the quality and relevance of the data used for the project, and can significantly impact the performance of the machine learning model.

6.3 DATA FLOW DIAGRAM

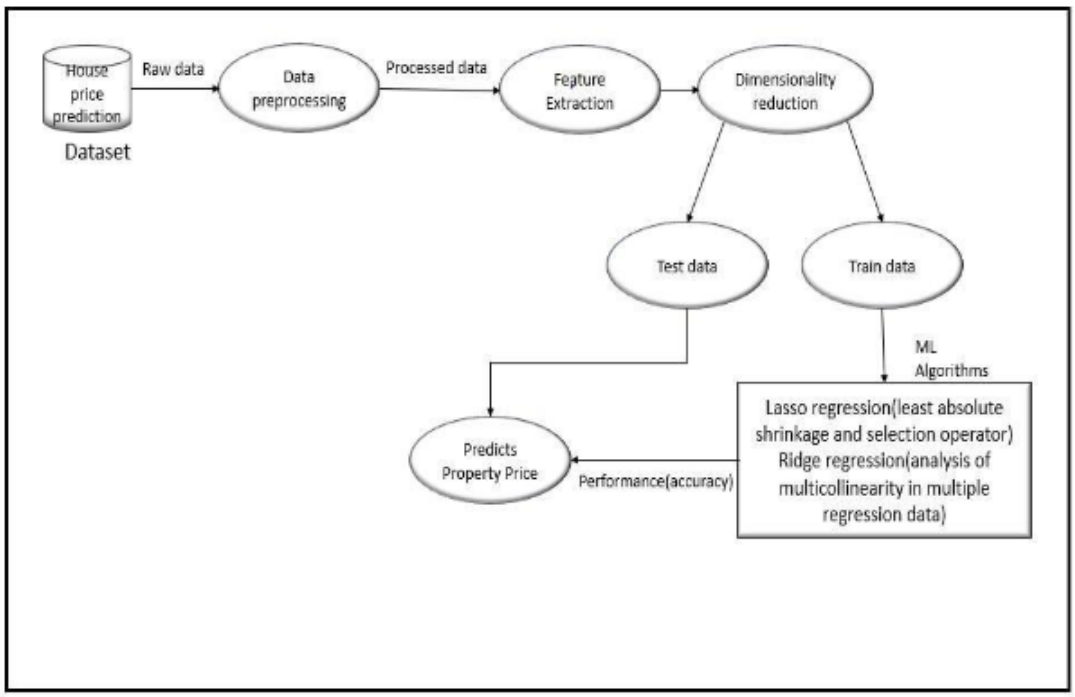


Fig 6.3 : Data Flow Diagram for House Price Prediction

- Dataset is preprocessed along with the techniques of feature extraction and dimensionality reduction.
- Preprocessed data is further categorized into train and test data for accuracy rate.

6.4 CLASS DIAGRAM

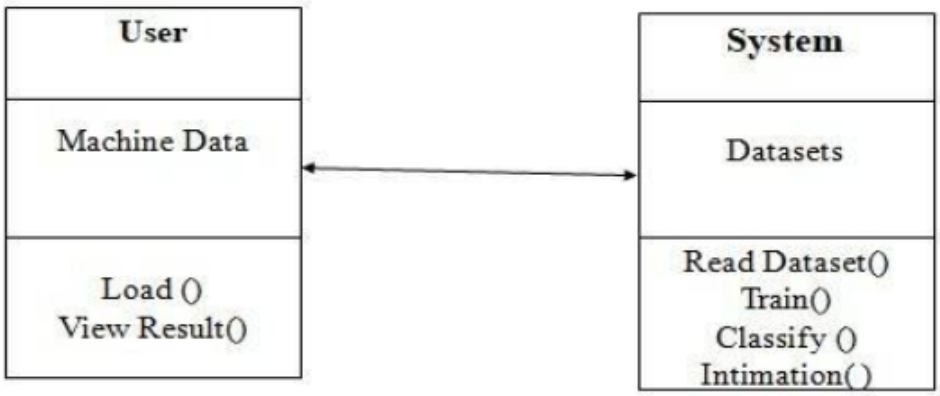


Fig 6.4: Class Diagram for House Price Prediction

- This figure shows relationship between user and system.
- Datasets are preprocessed and multicollinearity was removed.

6.4 USECASE DIAGRAM

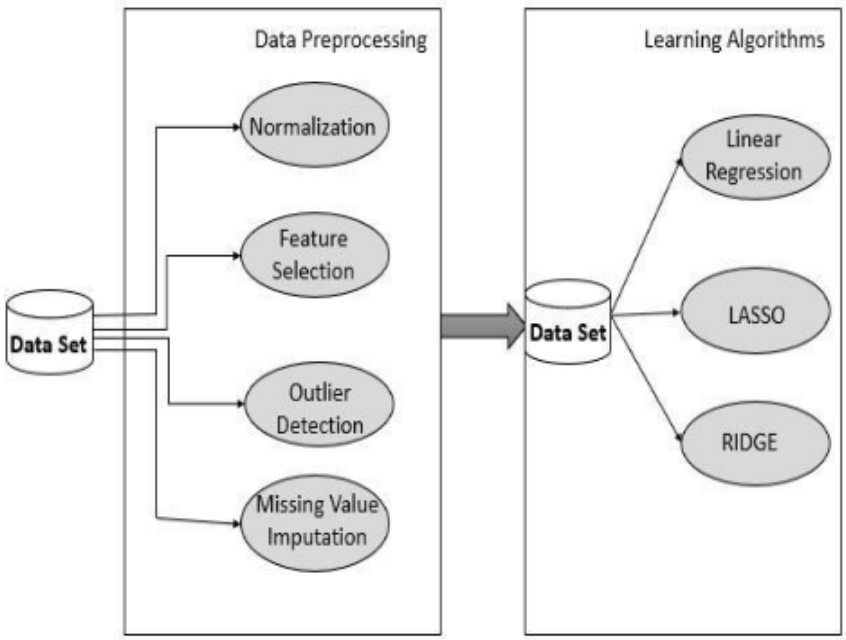


Fig 6.5: Usecase Diagram for House Price Prediction

- This figure shows the data preprocessing and algorithm applied.
- In data preprocessing process include the normalization, Feature selection, outlier detection and missing value imputation.
- In this project linear, lasso and ridge regression is used.

6.5 SEQUENCE DIAGRAM

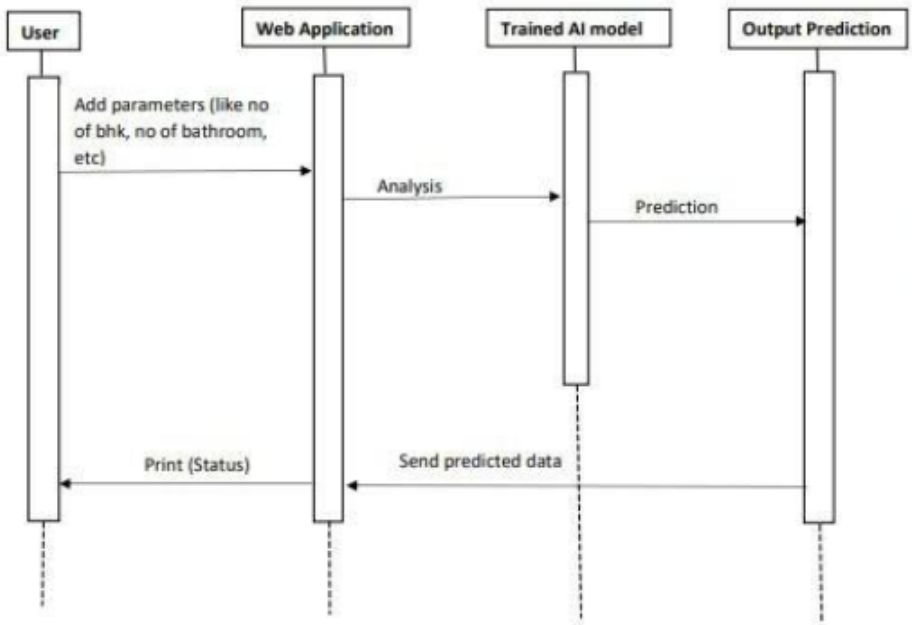


Fig 6.5: Sequence Diagram for House Price Prediction

- The Sequence diagram shows the relationship between web application and trained AI model.
- In this figure we can see the working flow of this model.

6.6 STATE CHART DIAGRAM

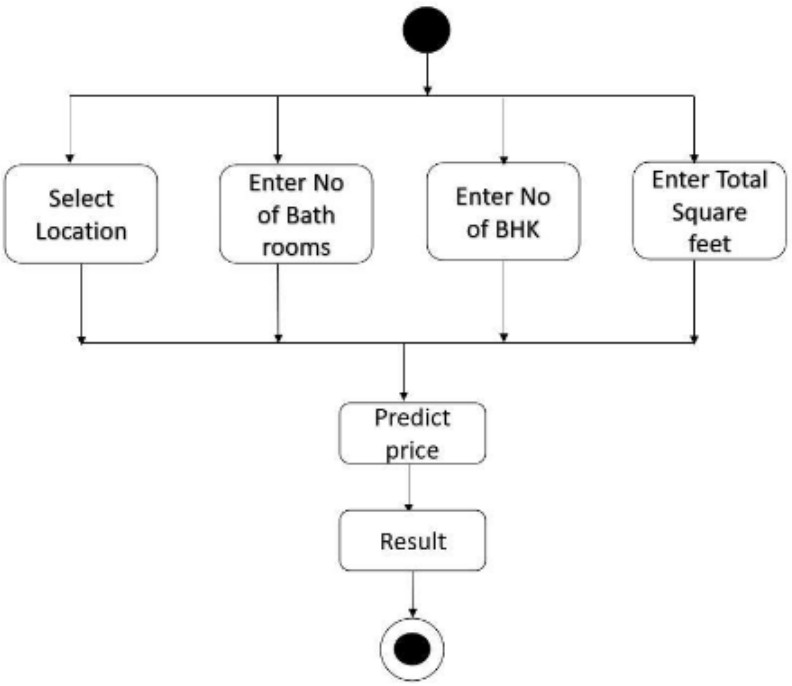


Fig 6.6: State Chart Diagram for House Price Prediction

- State chart diagram shows the workflow of web application.
- User need to enter the required parameters like location, no of BHK etc to get price.
- After getting this user data regression algorithm compare and predict the price with respect to dataset.
- Then predicted price will be printed.

6.8 ACTIVITY DIAGRAM

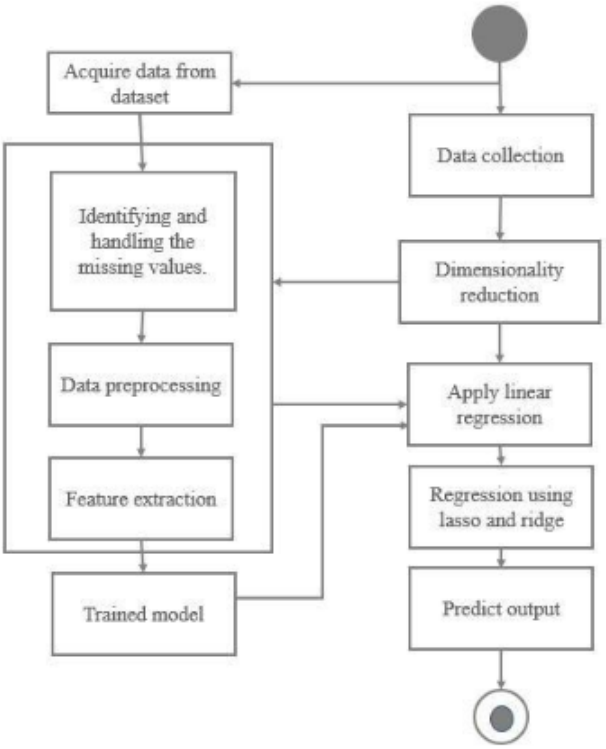


Fig 6.8: Activity Diagram for House Price Prediction

Chapter - 7

IMPLEMENTATION

To implementation of house price prediction model to predicts the accurate price software's used are

- Language used to code the project is Python.
- Operating System-Windows 10.

7.1 pycharm

PyCharm is an integrated development environment (IDE) used for programming in Python. It provides code analysis, a graphical debugger, an integrated unit tester, integration with version control systems, and supports web development with Django. PyCharm is developed by the Czech company JetBrains.

It is cross-platform, working on Microsoft Windows, macOS and Linux. PyCharm has a Professional Edition, released under a proprietary license and a Community Edition released under the Apache License.[2] PyCharm Community Edition is less extensive than the Professional Edition.

Coding assistance and analysis, with code completion, syntax and error highlighting, linter integration, and quick fixes Project and code navigation: specialized project views, file structure views and quick jumping between files, classes, methods and usages Python code refactoring: including rename, extract method, introduce variable, introduce constant, pull up, push down and others Support for web frameworks: Django, web2py and Flask Integrated Python debugger Integrated unit testing, with line-by-line coverage Google App Engine Python development Version control integration: unified user interface for Mercurial, Git, Subversion, Perforce and CVS with changelists and merge Scientific tools integration: integrates with IPython Notebook, has an interactive Python console, and supports Anaconda as well as multiple scientific packages including Matplotlib and NumPy.

7.2 Programming Language Used

Programming language used to design the proposed method is Python. Python is a high-level programming language with dynamic semantics. It is an interpreted language i.e. interpreter executes the code line by line at a time, thus makes debugging easyPython Imaging Library (PIL) is one of the popular libraries used for image processing. PIL can be used to display image, create thumbnails, resize,

rotation, convert between file formats, contrast enhancement, filter and apply other digital image processing techniques etc.

Python is often used as a support language for software developers, for build control and management, testing, and in many other ways. Python is designed by Guido van Rossum. It is very easy for user to learn this language because of its simpler coding. It provides an easy environment to furnish computation, programming visualization. Python supports modules and packages, which encourages program modularity and code reuse. It has various built-in commands and functions which will allow the user to perform functional programming. Apart from being an open-source programming language, developers use it extensively for application development and system development programming. It is a highly extensible language. Python contains many inbuilt functions which helps beginners to learn easily. Some of the most commonly used functions are: data.head() – display all columns and 5 rows. data.isna.sum() – detect missing values for an array like object. Lambda – a keyword in python for defining the anonymous function.

7.2.1 Key Features of Python

• Python is an interpreted language i.e. interpreter executes the code line by line at a time, thus makes debugging easy.

• Python is more expressive language, since it is more understandable and readable.

• To design and solve problems it offers an interactive atmosphere.

• Python has a large and broad library and provides rich set of module and functions for rapid application development.

• Built in graphics for conception of data and tools is also supported. To implementation of house price prediction model to predicts the accurate price.

• It can be easily integrated with languages like C, C++, JAVA etc.

7.2.2 Packages

Packages are the namespaces which consists of multiple packages and module themselves. Each package in Python is a directory which must contain a special file called _init_py. This file can be empty, and it indicates that the directory it contains is a Python package, so it can be imported in the same way that the module can be imported.

As our application program grows larger in size with a lot of modules, one can place the similar

modules in one package and different modules in different packages. This makes a program easy to manage and conceptually clear. We can import the modules from packages using the dot (.) operator.

1. Pandas

Pandas is a powerful and widely used Python library for data manipulation and analysis. It provides a high-level interface for working with structured data, such as tabular or timeseries data, and is built on top of NumPy, another popular Python library for scientific computing. With Pandas, you can easily load, manipulate, and transform data in various formats such as CSV, Excel, SQL databases, or even HTML tables. Pandas data structures include Series, which is a one-dimensional array with labels, and DataFrame, which is a two-dimensional table with columns and rows.

One of the most significant advantages of using Pandas is its ability to handle missing or incomplete data, which is a common problem in real-world datasets. It provides a variety of functions and methods for data cleaning, such as dropping or filling missing values, removing duplicates, and transforming data types. Pandas also supports data filtering, grouping, and aggregation operations that allow you to analyze and summarize large datasets efficiently.

Another notable feature of Pandas is its ability to perform time-series analysis and manipulation, making it a popular tool in finance, economics, and other fields where timeseries data is common. With Pandas, you can easily manipulate and resample time-series data, perform date/time calculations, and plot time-series data.

Overall, Pandas is a versatile and powerful library for data manipulation and analysis in Python, making it an essential tool for data scientists, analysts, and developers working with structured data. Its easy-to-use interface and extensive documentation make it accessible to both beginners and advanced users.

2. NumPy

NumPy is a python library used for working with arrays. It also has functions for working in domain of linear algebra, fourier transform, and matrices. NumPy was created in 2005 by Travis Oliphant. It is an open source project and you can use it freely. NumPy stands for Classification of Abnormal Red Blood Cells using Deep Learning. AIT/ISE/2022-23 Page 47 Numerical Python. In Python we have lists that serve the purpose of arrays, but they are slow to process. NumPy aims to provide an array object that is

up to 50x faster than traditional Python lists. The array object in NumPy is called ndarray, it provides a

lot of supporting functions that make working with ndarray very easy. Arrays are very frequently used in

data science, where speed and resources are very important. NumPy arrays are stored at one continuous place in memory unlike lists, so processes can access and manipulate them very efficiently. This behaviour is called locality of reference in computer science. This is the main reason why NumPy is faster than lists. Also, it is optimized to work with latest CPU architectures.

3.Matplotlib

Matplotlib is a comprehensive library for creating static, animated, and interactive visualizations in Python. Matplotlib produces publication-quality figures in a variety of hardcopy formats and interactive environments across platforms. Matplotlib can be used in Python scripts, the Python and IPython shell, web application servers, and various graphical user interface toolkits. Matplotlib. pyplot is a collection of command style functions that make matplotlib work like MATLAB. Each pyplot function makes some change to a figure: e.g., creates a figure, creates a plotting area in a figure, plots some lines in a plotting area, decorates the plot with labels, etc. Some people embed matplotlib into graphical user interfaces like wxpython or pygtk to build rich applications. Others use matplotlib in batch scripts to generate postscript images from some numerical simulations, and still others in web application servers to dynamically serve up graphs. Matplotlib is the brainchild of John Hunter (1968-2012), who, along with its many contributors, have put an immeasurable amount of time and effort into producing a piece of software utilized by thousands of scientists worldwide.

4. Python Imaging Library

Python Imaging Library (PIL) is a popular open-source library for working with image files in Python. It provides a set of powerful tools and functions for creating, editing, and processing images. PIL supports a wide range of image file formats, including JPEG, PNG, BMP, GIF, and many others. One of the most significant advantages of PIL is its simplicity and ease of use.

With just a few lines of code, you can perform basic image operations such as resizing, cropping, rotating, and flipping images. Additionally, PIL offers more advanced image processing techniques, such as filtering, blending, and color adjustment, making it a versatile tool for image manipulation. PIL

also includes support for working with image metadata, such as EXIF data, allowing you to extract useful information from image files. Overall, Python Imaging Library is a powerful tool for working with images in Python, providing a comprehensive set of features that can help you accomplish a wide range of image-related tasks with ease.

5. Scikit-learn

Scikit-learn, also known as sklearn, is a popular Python library for machine learning. It provides simple and efficient tools for data mining and data analysis. Scikit-learn is built on top of NumPy, SciPy, and matplotlib and offers a wide range of supervised and unsupervised learning algorithms for tasks such as classification, regression, clustering, and dimensionality reduction.

One of the key features of scikit-learn is its ease of use. It provides a consistent interface for all the algorithms, making it easy to switch between them. It also provides extensive documentation, tutorials, and examples to help users get started with machine learning.

Scikit-learn provides a wide range of machine learning algorithms, including decision trees, support vector machines, k-nearest neighbors, random forests, and many more. It also provides tools for data preprocessing, such as scaling, normalization, and feature extraction.

Another useful feature of scikit-learn is its model selection tools, which help users to choose the best model for their data. This includes tools for cross-validation, grid search, and performance metrics.

Scikit-learn is widely used in both academia and industry, and has a large community of users and developers. Its open-source nature makes it easy to contribute to and customize, and it has been integrated with many other Python libraries and frameworks.

Overall, scikit-learn is a powerful and flexible library for machine learning in Python, and is a valuable tool for data scientists, machine learning engineers, and researchers alike.

7.3 Pseudocodes for Preprocessing Techniques

7.3.1 Pseudocode for Prediction

Step1: prints total number of values

```
for column in data.columns:
print(data[column].value_counts())
        print("*"*20)
```

Step2: checking null values

```
data.isna().sum()
```

Step3: removing less dependent features

```
data.drop(columns=['area_type','availability','society','balcony'],inplace=True)
```

Step4: total number of unique location names

```
data['location'].value_counts()
```

Step5: converting the values in to float for calculation

```
def convertRange(x):
temp = x.split('-')
if len(temp) == 2:
return(float(temp[0])+ float(temp[1]))/2
try:
return float(x)
except:
return None
```

Step6: calculating price per square feet

```
data['price_per_sqft']=data['price']*100000/data['total_sqft']
```

Step7: stripping the values of location

```
data['location']=data['location'].apply(lambda x: x.strip())
location_count=data['location'].value_counts()
```

Step8: removing the outlier from dataset

```
def bhk_outlier_remover(df):
exclude_indices=np.array([])
for location,location_df in df.groupby('location'):
bhk_stats={}
for bhk,bhk_df in location_df.groupby('bhk'):
bhk_stats[bhk]={
'mean':np.mean(bhk_df.price_per_sqft),
'std':np.std(bhk_df.price_per_sqft),
'count':bhk_df.shape[0]
}
for bhk,bhk_df in location_df.groupby('bhk'):
stats=bhk_stats.get(bhk-1)
if stats and stats['count']>5:
exclude_indices=np.append(exclude_indices,bhk_df[bhk_df.price_per_sqft<(stats['mean'])].index.
values)
return df.drop(exclude_indices,axis='index')
```

Step9: transferring the cleaned data to the file

```
data.to_csv("cleaned_data.csv")
```

Step10: splitting the dataset in to train and test data

```
X_train,X_test,y_train,y_test=train_test_split(X,y,test_size=0.2,random_state=0)
```

Chapter - 8

TESTING

Testing is the significant phase in the process or application development life cycle. Testing is final phase where the application is tested for the expected outcomes. Testing of the system is done to identify the faults or prerequisite missing. Therefore, testing plays a vital role for superiority assertion and confirming the consistency of the software. Software testing is essential for correcting errors and improving the quality of the software system. The software testing process starts once the program is written and the documentation and related data structures are designed. Without proper testing or with incomplete testing, the program or the project is said to be incomplete.

Throughout the testing phase, the procedure will be implemented with the group of test circumstances and the outcome of a procedure for the test case will be appraised to identify whether the program is executing as projected. Faults found during testing will be modified using the testing steps and modification will be recorded for forthcoming reference. Some of the significant aim of the system testing is,

- To confirm the superiority of the project.
- To discover and eradicate any residual errors from prior stages.
- To authenticate the software as a result to the original problem.
- To offer effective consistency of the system.

8.1 Test Procedures

Test cases are the key aspect for the success of any system. A test case is a manuscript which has a set of test data, prerequisites, predictable results and post conditions, designed for a specific test scenario in imperative to validate acquiescence against a specific requirement. Performance of a system is based on the cases written for each and every modules of a system. Test cases are precarious for the prosperous performance of the system. If the predicted Classification of Abnormal Red Blood Cells using Deep Learning. outcome doesn't match with the real outcome, then error log is displayed. There are two elementary forms of test cases, namely

- Formal test case
- Informal test case

8.1.1 Formal Test Cases

Test cases which are authored as per the test case format. It has all the information like preconditions, input data, output data, post conditions, etc. It has a defined set of inputs which will provide the expected output.

8.1.2 Informal Test Cases

This test cases are based on user requirements where exact input and output were not know. User can specifies the requirement so that applied algorithm in the model compare and process the price in the given dataset and provide the accurate prediction of house price.

Chapter - 9

TEST CASES AND RESULTS

Test Case ID	Description	Input	Expected output	Actual Output	Status
1	Desired Features	Selecting the required location in the dropdown menu	Select location	Location selected	PASS

Table 8.1: Test case for selecting Location

The drop down menu appears with all the locations displayed and the selected location should be considered for the further process.this successful test case for the location is shown in the table 8.1, the drop down menu contains all the locations of Bangalore present in the data set.here it makes the user easy to select locations from the list of locations displayed.

The value selected should be used for the further process of prediction, as different locations may have different price for square feet

Test Case ID	Description	Input				Expected output	Actual Output	Status
		location	bhk	bathrooms	sqft			
1	Try different combinations of features	Electronic city Phase ll	2	2	1056	3631095.29	365232.50	PASS
2	Try different combinations of features	Amrutahalli	3	2	1200	6164189.44	6164189.44	PASS
3	Try different combinations of features	Anandapura	1	1	1000	3747703.67	3747703.67	PASS
4	Try different combinations of features	Bhoganahalli	4	1	1500	8467731.96	8467731.96	PASS
5	Try different combinations of features	Choodasandra	2	1	1250	6389071.11	5432031.01	FAIL
6	Try different combinations of features	EPIP Zone	3	2	1450	8888643.22	8888643.22	PASS
7	Try different combinations of features	Girinagar	1	1	850	28222775.63	28222775.63	PASS
8	Try different combinations of features	HSR layout	2	1	1300	6171970.93	7216425.63	FAIL
9	Try different combinations of features	Jakkur	4	2	1800	11492908.54	11492908.54	PASS
10	Try different combinations of features	NRI layout	6	2	2000	8937873.39	8937873.39	PASS

Chapter - 10

SNAPSHOTS

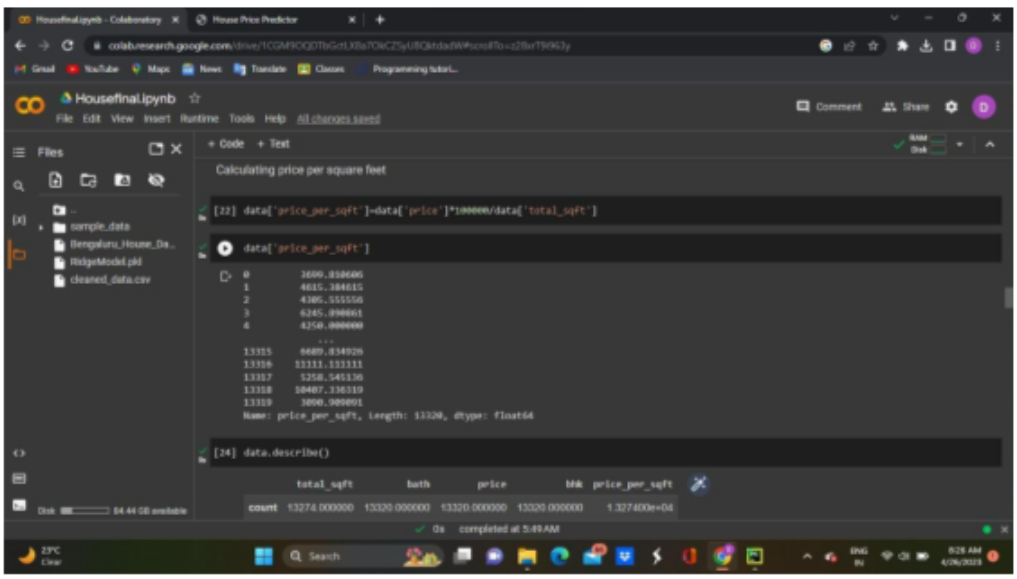


Fig 10.1: Calculating price

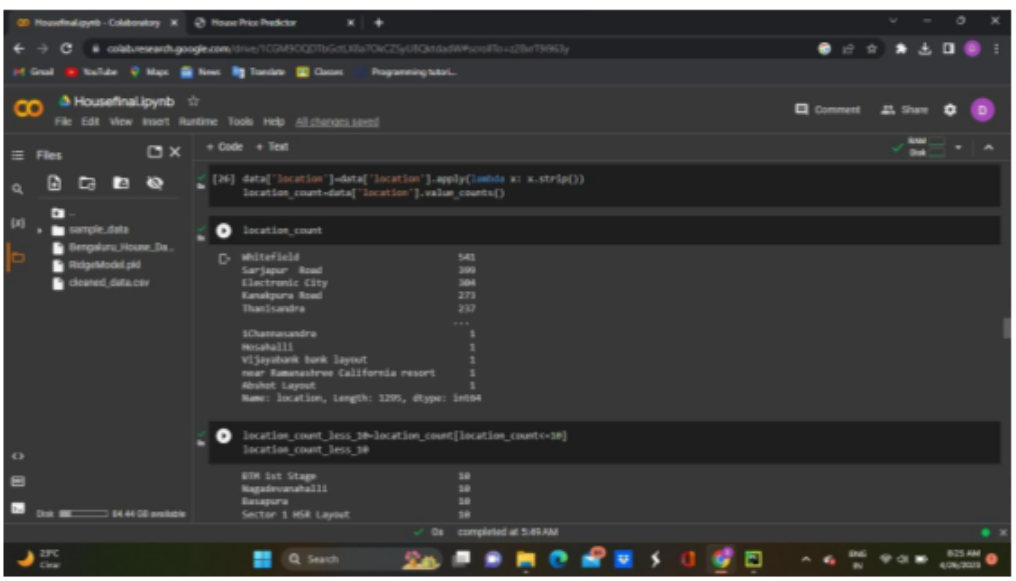


Fig 10.2: Dataset location

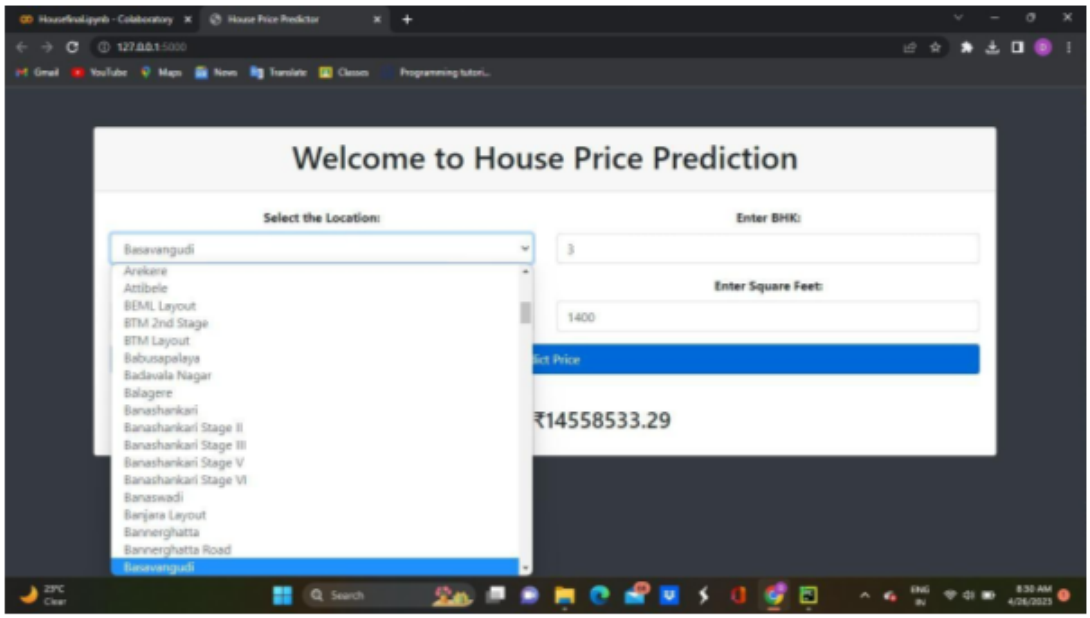


Fig 10.3: Drop down for location selection

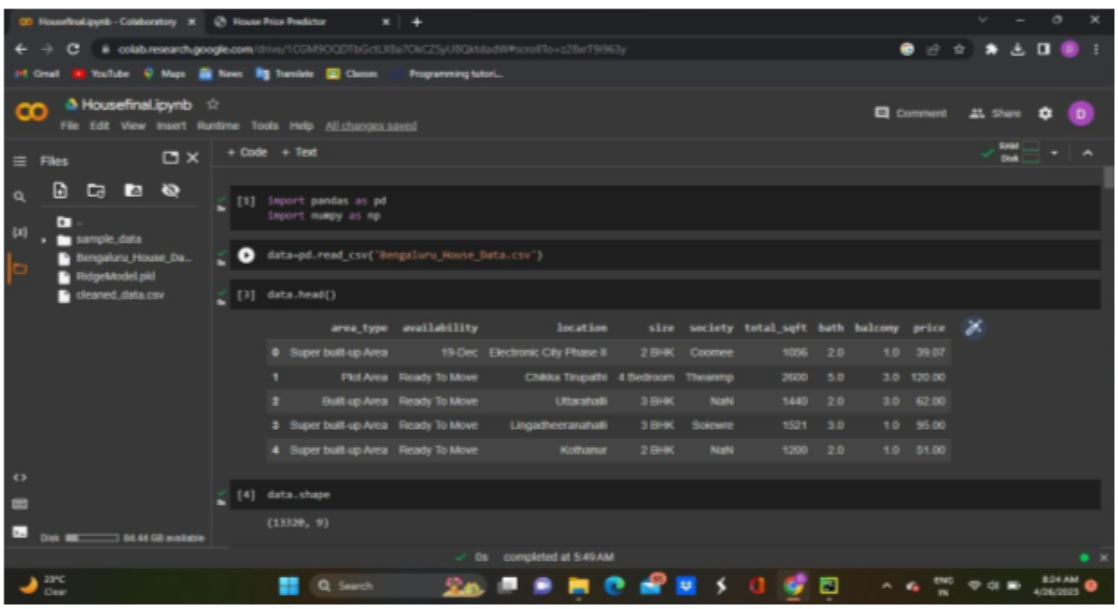


Fig 10.4: Displaying dataset

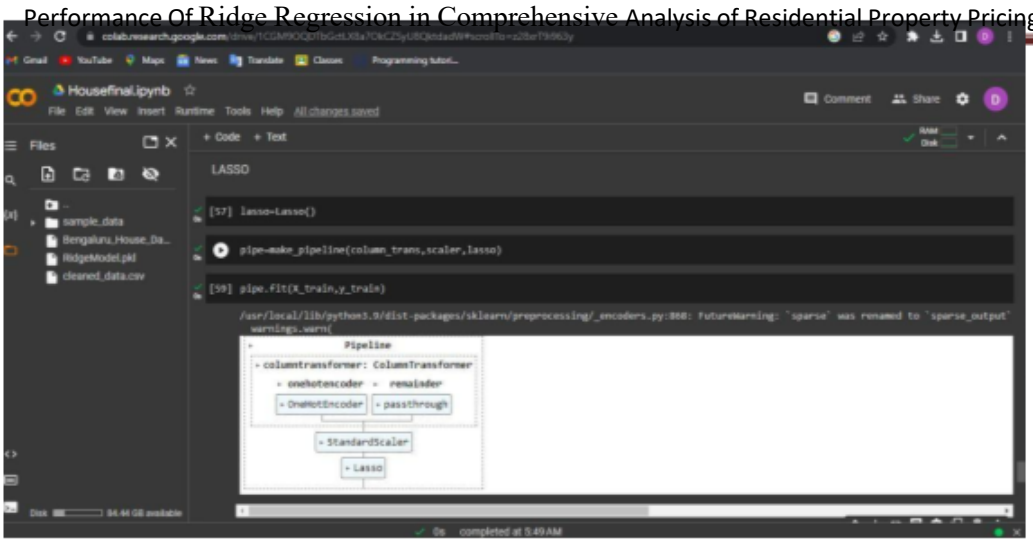


Fig 10.5: Applying lasso to the dataset

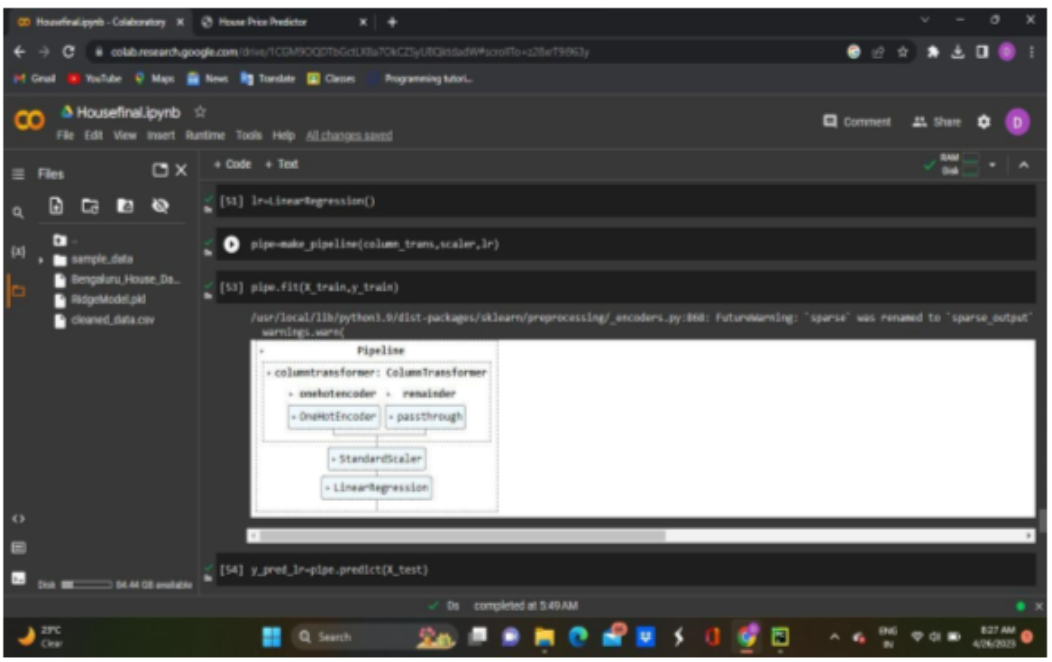


Fig 10.6: Applying linear regression to the dataset

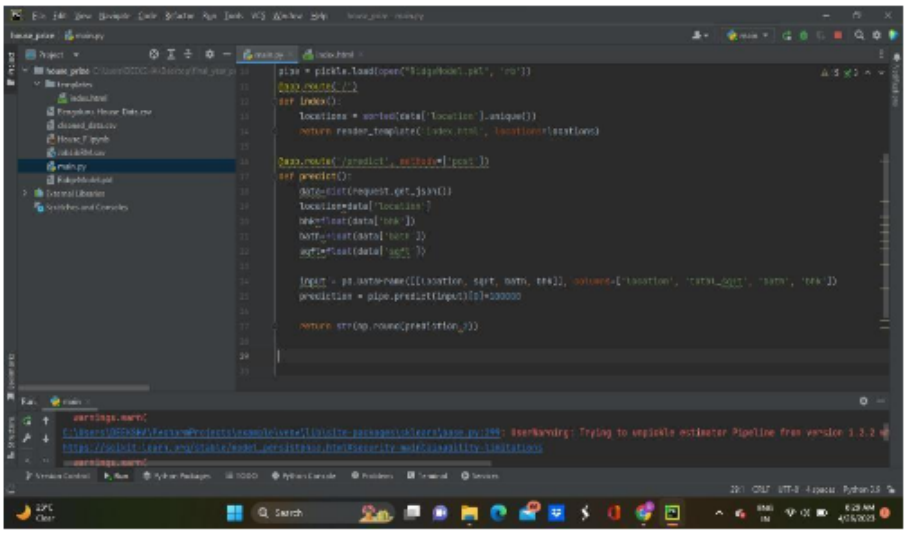


Fig 10.7 Applying lasso to the dataset

Chapter-10

Possible Outcomes

1. Help the developer determine the selling price of a house and can help the customer to arrange the right time to purchase a house.
2. Ridge and lasso regression method used to analyze the machine performance and effective prediction.
3. Saves the customers money by predicting the prices of house in different areas.
4. Majorly Considers the tax Price of the house calculated based on updated Government formula.

References

[1] [2020] Machine Learning based Predicting House prices using Regression Techniques , Manasa j, Radha Gupta , Narahari N S.

[2] [2019] Using Machine Learning Algorithms for Housing Price Prediction : The Case of Fairfax county, Virginia Housing Data, Byeonghwa, and Jae Kwon Bae .

[3] [2019] Prediction of House Pricing using Machine Learning with Python, Namit Jain, Parikshay Purushottam Sharma, Vikas Deep.

[4] [2020] House Price Prediction Modelling using Machine Learning, Dr. M. Thamarai,Dr. S P. Malar vizhi.

[5] Khan Sohil Liyaqatullah1, Usman Malik2, Qureshi Mohammed Basheeruddin3, "International Journal of Advanced Research in Computer and Communication Engineering“, Vol. 10, Issue 4, April 2021,DOI 10.17148/IJARCCE.2021.10468.

[6] Vignesh M ,Vijay V, Krishna S, Satya moorthy, “International Research Journal of Modernization in Engineering Technology and Science,Volume:04/Issue:04/April-2022. eISSN: 2582-5208.

[7] Aldrin Fernandes, Abdullah, Qureshi , Amit Narote, Amir shaik ,“International Journal Of Advance Research ,Ideas And Innovation In Technology ”,Vol 7, Issue 6-V716-1155.

[8] Amena Begum, Nishad Jahan Kheya, Md. Zahidur Rahman, “International Journal of Innovative Technology and Exploring Engineering (IJITEE)”, ISSN: 2278-3075 (Online), Volume-11 Issue-3, January 2022.

[9] Bindu Sivasankar1 , Arun P. Ashok2* , Gouri Madhu3 , Fousiya S. 4 JCSE International Journal of Computer Sciences and Engineering. Vol.8, Issue.7, July 2020 E-ISSN: 2347-2693

[10] International journal of creative research, Aniket Singh1 , Adarsh Kumar Singh2 , Aditya Raj3 , Harshit Jain4 ,Mrs. Asha M S 5 , House Price Prediction Analysis using Machine Learning,ISSN 2320-2882.

[11] Anand G Rawool, Dattatary V,Rogye, Sainath G,Rane , DR Vinayak A, Bharathi, “House price prediction using machine learning”. Iconic research and engineering” Vol 4, No11,pp 29-33,2021.

[12] Park, Byeonghwa, and joe Kwon Bae, “Using Machine Learning Algorithms for Housing price prediction: The case of FairFax country ,Virginia Housing Data”,pp.2928-2934, 2019.

Textbooks:

[1] Tom M Mitchell, “Machine Lerning”,1st Edition, McGraw Hill Education, 2017

[2] AurÈlienGÈron,"Hands-On Machine Learning with Scikit-Learn and Tensor Flow: Concepts, Tools, and Techniques to Build Intelligent Systems", 1st Edition, Shroff/O'Reilly Media, 2017

Websites:

https://www.geeksforgeeks.org/machine-learning/

https://en.m.wikipedia.org/wiki/Machine_learning

https://www.ybifoundation.org/course/machine-learning?previouspage=home&isenrolled=no