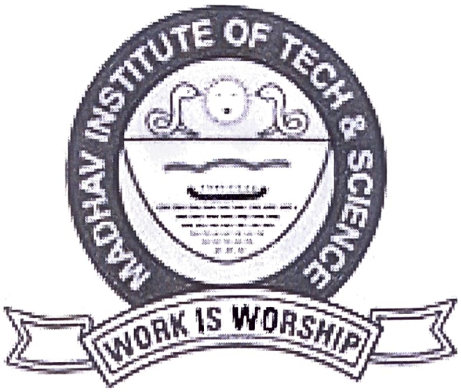
MADHAV INSTITUTE OF TECHNOLOGY & SCIENCE GWALIOR

(A Govt. Aided UGC Autonomous Institute Affiliated to RGPV, Bhopal)

NAAC Accredited with A++ Grade



Project Report

on

Home Rate Prediction

Submitted By:

Pradeep raghuvanshi

0901AI211049

Faculty Mentor:

Dr. Neelam arya

CENTRE FOR ARTIFICIAL INTELLIGENCE

MADHAV INSTITUTE OF TECHNOLOGY & SCIENCE

GWALIOR - 474005 (MP) est. 1957

JULY-DEC. 2023

MADHAV INSTITUTE OF TECHNOLOGY & SCIENCE GWALIOR

(A Govt. Aided UGC Autonomous Institute Affiliated to RGPV, Bhopal)

NAAC Accredited with A++ Grade

CERTIFICATE

This is certified that **Pradeep raghuvanshi** (0901AI211049) has submitted the project report titled **Home rate prediction** under the mentorship of **dr. Neelam arya**, in partial fulfilment of the requirement for the award of degree of Bachelor of Technology in **IT(artificial intelligence and robotics)** from Madhav Institute of Technology and Science, Gwalior.

Neelam 23/11/23

Dr. Neelam arya
Faculty Mentor
Centre for artificial intelligence

23/11/23

Dr. R. R. Singh
Coordinator
Centre for Artificial Intelligence

MADHAV INSTITUTE OF TECHNOLOGY & SCIENCE GWALIOR

(A Govt. Aided UGC Autonomous Institute Affiliated to RGPV, Bhopal)

NAAC Accredited with A++ Grade

DECLARATION

I hereby declare that the work being presented in this project report, for the partial fulfilment of requirement for the award of the degree of Bachelor of Technology in **IT(artificial intelligence and robotics)** > at Madhav Institute of Technology & Science, Gwalior is an authenticated and original record of my work under the mentorship of , **Dr. Neelam arya** .

I declare that I have not submitted the matter embodied in this report for the award of any degree or diploma anywhere else.

Pradeep Raghuvanshi
0901AI211049
3rd Year,
Centre for Artificial Intelligence

ACKNOWLEDGEMENT

The full semester project has proved to be pivotal to my career. I am thankful to my institute, **Madhav Institute of Technology and Science** to allow me to continue my disciplinary/interdisciplinary project as a curriculum requirement, under the provisions of the Flexible Curriculum Scheme (based on the AICTE Model Curriculum 2018), approved by the Academic Council of the institute. I extend my gratitude to the Director of the institute, **Dr. R. K. Pandit** and Dean Academics, **Dr. Manjaree Pandit** for this.

I would sincerely like to thank my department, **Centre for Artificial Intelligence,** for allowing me to explore this project. I humbly thank **Dr. R. R. Singh**, Coordinator, Centre for Artificial Intelligence, for his continued support during the course of this engagement, which eased the process and formalities involved.

I am sincerely thankful to my faculty mentors. I am grateful to the guidance of **Dr. Neelam arya** , for his continued support and guidance throughout the project. I am also very thankful to the faculty and staff of the department.

Pradeep Raghuvanshi
090AI211049
3rd Year,
Centre for Artificial Intelligence

TABLE OF CONTENTS

Abstract

The escalating demand for accurate and timely home price predictions has spurred advancements in machine learning techniques to address this critical aspect of the real estate market. This project aims to develop a robust home rate price prediction model by leveraging state-of-the-art machine learning algorithms and comprehensive datasets. The proposed model integrates a diverse set of features, including property characteristics, location attributes, economic indicators, and historical pricing trends. By employing advanced regression models and data preprocessing techniques, the system strives to uncover intricate patterns and dependencies within the data, contributing to more precise predictions .The project's significance lies in its potential to empower homeowners, real estate professionals, and investors with actionable insights into future property values. This predictive tool not only assists in informed decision-making but also serves as a valuable resource for market analysis and risk management .Furthermore, the model's performance will be rigorously evaluated against benchmark datasets, ensuring its reliability and generalizability across diverse real estate markets. The outcome of this research is expected to contribute to the ongoing discourse on predictive analytics in the real estate domain and foster innovation in the development of intelligent decision support systems for the housing market.

सार

गृह रेट पूर्वानुमान परियोजना" एक महत्वपूर्ण और उत्कृष्ट तकनीकी परियोजना है जो निवेशकों और आवास क्रय करने वालों को मार्गदर्शन करने के लिए तैयार की गई है। इस परियोजना का मुख्य उद्देश्य भविष्य के गृह रेट्स की पूर्वानुमान करना है ताकि व्यापारिक और निवेशक ठीक से योजना बना सकें।इस परियोजना में, सांगठनिक डेटा, बाजार विश्लेषण, और मौद्रिक तत्वों का समाहित अध्ययन किया जाता है। यहां तक कि स्थानीय और राष्ट्रीय आर्थिक स्थिति का भी मूल्यांकन किया जाता है।मॉडल और एल्गोरिदम्स का प्रयोग करके, आंकड़ों का विश्लेषण किया जाता है ताकि आगामी समय में क्षेत्र में अधिकतम मुनाफा उपलब्ध किया जा सके। यह परियोजना गहरी समझ और सुधारित अनुसंधान के साथ मिलकर, बाजार के निर्णय लेने में सहायक होने के लिए तैयार की जा रही है।इस परियोजना के माध्यम से, आवास बाजार के संकेतों को समझकर निवेशकों को सुरक्षित और लाभकारी निवेश के लिए एक सामर्थ्यपूर्ण उपाय प्रदान किया जा रहा है।

Chapter 1:- Project Overview

1.1 Introduction

Welcome to our home Rate Prediction Minor Project, focusing on the real estate landscape within a specific enclave of Gwalior city. In this endeavour, we delve into the intricate dynamics of a localized housing market, aiming to provide precise predictions for home prices in this unique area. Our data collection process involved a meticulous survey conducted within our college community, ensuring that the insights drawn are representative of the specific characteristics and preferences of the residents.

Gwalior, with its rich history and diverse neighbourhoods, presents an ideal microcosm for understanding the nuances of home pricing. By concentrating on a small area, we can factor in hyper-local influences such as proximity to amenities, neighbourhood dynamics, and community preferences. The dataset we've amassed encapsulates a spectrum of variables, ranging from property features and sizes to resident demographics and economic indicators specific to this locality.

As we embark on this minor project, our objective is to not only offer accurate predictions for home prices but also to contribute valuable insights that can inform residents, prospective buyers, and real estate enthusiasts within our college community. This localized approach ensures that our predictions are finely tuned to the intricacies of this unique housing market, laying the foundation for informed decision-making in the realm of real estate in Gwalior's specific locality.

1.2 Objective:

The objective of this intermediate Python project is to Develop a Home Rate prediction system that will detect the price of Rental home in small area of Gwalior . This system will help new student to find suitable new home .

1.3 Hardware & Software:

The requirement for this Python project is, You need to have Python (3.6 version recommended) installed on your system, then using pip, you can install the necessary packages. And Also I am using Google collab.

Chapter 2:-Literature Review:

2.1 Introduction:

The application of machine learning models in predicting home prices has gained significant attention in recent years due to its potential to enhance the accuracy and efficiency of real estate valuation. This literature review synthesizes key studies that have explored the use of machine learning techniques for home price prediction, focusing on methodologies, data sources, and notable findings.

1. Regression Models:

- Traditional regression models, including linear regression and multiple regression, have been employed in predicting home prices based on various features such as square footage, number of bedrooms, and location.

2. Ensemble Methods:

- Studies have demonstrated the effectiveness of ensemble methods, such as Random Forest and Gradient Boosting, in capturing complex relationships within housing datasets, leading to improved prediction accuracy.

3. Neural Networks :

- Deep learning models, particularly neural networks, have been applied to learn intricate patterns in large-scale datasets, allowing for a more nuanced understanding of the factors influencing home prices.

Chapter 3:

3.1 Methodology/ Flow Chart

Phase 1: Collection of data

Data processing techniques and processes are numerous. We collected data form students studying in our collage (Madhav Institute of Technology and Science Gwalior) Through Google form . The data would be having attributes such as Area, Flat(bhk , single room), Rent, furnished , etc. We must collect the quantitative data which is structured and categorized. Data collection is needed before any kind of machine learning research is carried out. Dataset validity is a must otherwise there is no point in analyzing the data

Data Collection Catagory :-

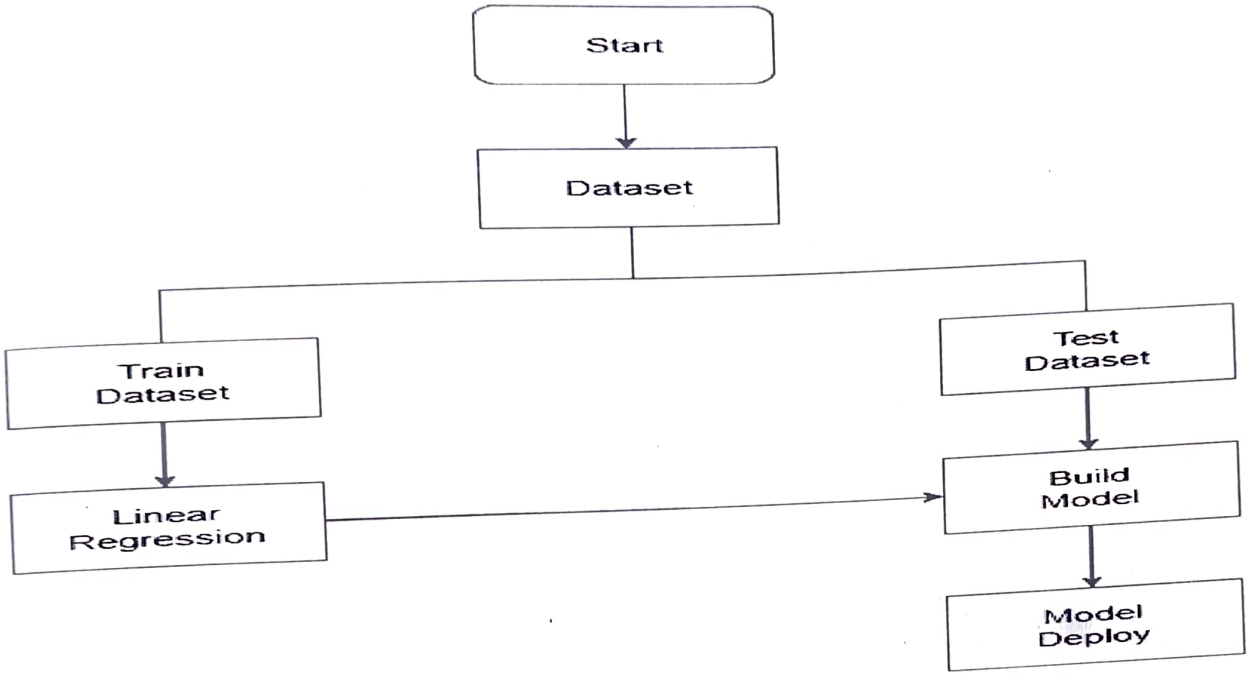
1. Area
2. Rent
3. Electricity Bill
4. Transport Rating
5. House type
   BHK/Single room
6. Furnished Status

Phase 2: Data preprocessing

Data preprocessing is the process of cleaning our data set. There might be missing values or outliers in the dataset. These can be handled by data cleaning. If there are many missing values in a variable we will drop values or substitute it with the average value.

Phase 3: Training the model

Since the data is broken down into two modules: a Training set and Test set, we must initially train the model. The training set includes the target variable. Machine learning algorithms are applied to the training data set.


Start
Dataset
Train Dataset
Test Dataset
Linear Regression
Build Model
Model Deploy

```
[ ] #importing libraries.
    import numpy as np
    import pandas as pd
    import matplotlib.pyplot as plt
    import seaborn as sns
```

```
[ ] path = "/content/drive/MyDrive/Collage /Untitled form (Responses) - Form Responses 1.csv"
    df_raw = pd.read_csv(path)
    df_raw.shape

    (208, 7)
```

```
df_raw.head()
```

	Timestamp	Area	Rent	Electricity Included	Furnished	Flat	Transport Rating
0	11/13/2022 21:11	Hanuman Nagar	12000	No	No	3 BHK	5
1	11/14/2022 21:27	Hanuman Nagar	1200	No	No	3 BHK	5
2	11/14/2022 21:27	City centre	7000	Yes	No	2 BHK	5
3	11/14/2022 21:28	Bank Colony	3500	Yes	No	0	3
4	11/14/2022 21:34	Indramani Nagar	3500	No	No	0	2

---

gwalior_data.ipynb ☆

File Edit View Insert Runtime Tools Help Last saved at 4:10 PM

Comment Share

+ Code + Text

Connect Editing

gvalior_data.ipynb

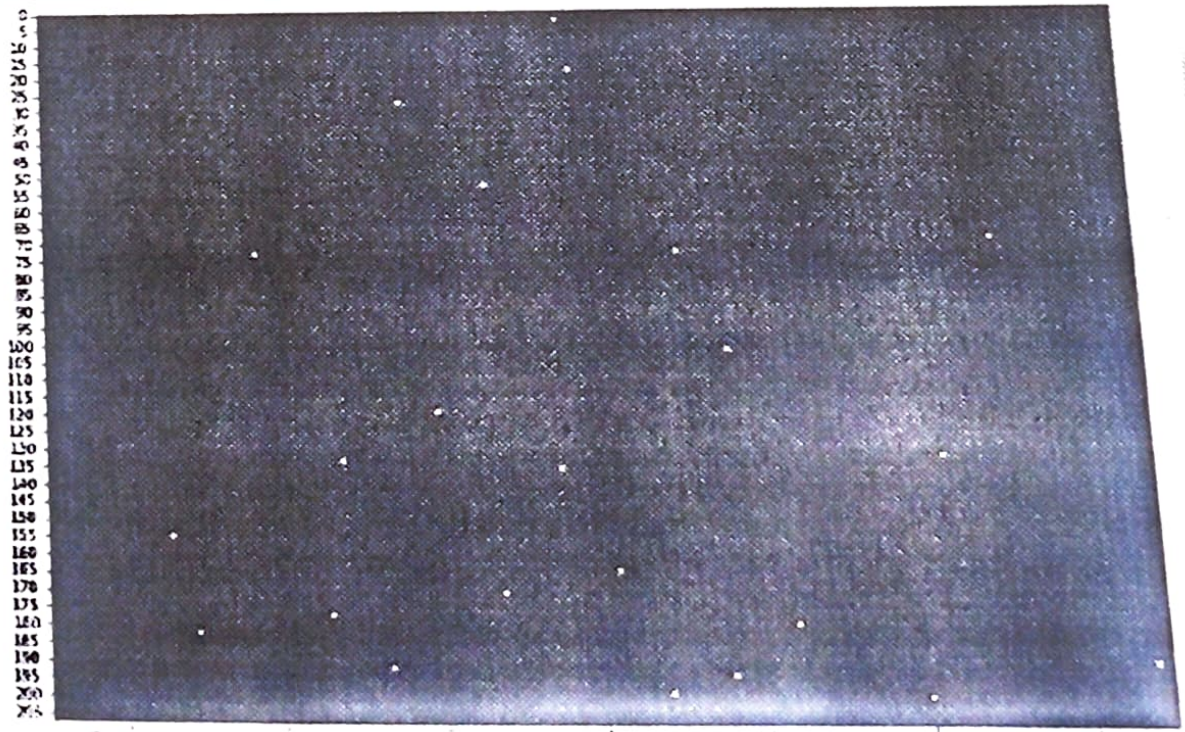
File Edit View Insert Runtime Tools Help Last saved at 4:10 PM

+ Code + Text

```
# visualize missing value using heatmap to get idea where is the value missing

plt.figure(figsize=(16,9))
sns.heatmap(df.isnull())
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7fd79761bd10>
```



File Edit View Insert Runtime Tools Help Last saved at 4:10 PM

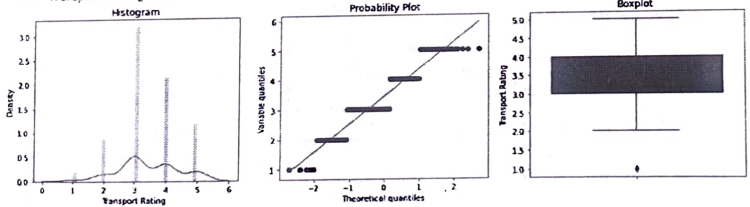
Connect | Editing

+ Code + Text

```
[ ]    plt.show()

[ ] num_var = ["Transport Rating"]
    for var in num_var:
      print("******* {} *******".format(var))
      diagnostic_plots(df, var)
```

******* Transport Rating *******



Double-click (or enter) to edit

```
[ ] df = df.reset_index(drop=True)
```

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 208 entries, 0 to 207
Data columns (total 7 columns):
 #   Column                 Non-Null Count  Dtype
---  ------                 --------------  -----
 0   Timestamp              208 non-null    object
 1   Area                   208 non-null    object
 2   Rent                   208 non-null    object
 3   Electricity Included   208 non-null    object
 4   Furnished              208 non-null    object
 5   Flat                   208 non-null    object
 6   Transport Rating       208 non-null    int64
dtypes: int64(1), object(6)
memory usage: 11.5+ KB
```

```
[ ] # join df and list size_int
df1 = df.join(pd.DataFrame({'bhk':size_int}))
df1.shape
```

```
df1.tail()
```

	Timestamp	Area	Rent	Electricity Included	Furnished	Flat	Transport Rating	bhk
...	...	...	...	...	...	...	...	...

gwalior_data.ipynb

File Edit View Insert Runtime Tools Help Last saved at 4:10 PM

Comment Share Connect Editing

+ Code + Text

```
 6   Transport Rating    208 non-null    int64
dtypes: int64(1), object(6)
memory usage: 11.5+ KB
```

```
# join df and list size_int
df1 = df.join(pd.DataFrame({'bhk':size_int}))
df1.shape
```

```
df1.tail()
```

	Timestamp	Area	Rent	Electricity Included	Furnished	Flat	Transport Rating	bhk
203	11/19/2022 18:01	Indramani Nagar	4000	No	No	1 BHK	2	1
204	11/19/2022 18:10	Krishna Nagar	4000	No	No	0	3	0
205	11/19/2022 21:57	Indramani Nagar	4000	No	No	0	3	0
206	11/19/2022 23:03	Gole ka mandir	3500	No	Yes	0	5	0
207	11/19/2022 23:03	Gole ka mandir	3500	No	Yes	0	5	0

Finding Outlier and Removing""**bold text**

```
# function to create histogram, Q-Q plot and boxplot

# for Q-Q plots
import scipy.stats as stats
```

gwalior_data.ipynb

File Edit View Insert Runtime Tools Help Last saved at 4:13 PM

Comment Share Connect Editing

+ Code + Text

```
df2 = df1.drop(["Flat"], axis =1)
df2.head()
```

	Timestamp	Area	Rent	Electricity Included	Furnished	Transport Rating	bhk
0	11/13/2022 21:11	Hanuman Nagar	12000	No	No	5	3
1	11/14/2022 21:27	Hanuman Nagar	1200	No	No	5	3
2	11/14/2022 21:27	City centre	7000	Yes	No	5	2
3	11/14/2022 21:28	Bank Colony	3500	Yes	No	3	0
4	11/14/2022 21:34	Indramani Nagar	3500	No	No	2	0

```
df3 = pd.get_dummies(df2, drop_first=True, columns=['Electricity Included ','Furnished'])
df3.shape
```

```
(208, 7)
```

```
df3.head()
```

	Timestamp	Area	Rent	Transport Rating	bhk	Electricity Included _Yes	Furnished_Yes
0	11/13/2022 21:11	Hanuman Nagar	12000	5	3	0	0
1	11/14/2022 21:27	Hanuman Nagar	1200	5	3	0	0
2	11/14/2022 21:27	City centre	7000	5	2	1	0
3	11/14/2022 21:28	Bank Colony	3500	3	0	1	0
4	11/14/2022 21:34	Indramani Nagar	3500	2	0	0	0

```
df3.to_csv('oh_encoded_data.csv', index=False) # test ml model on this data
```

3.3 Model Training:

gwalior_data.ipynb

File Edit View Insert Runtime Tools Help Last saved at 4:10 PM

Comment Share

+ Code + Text

Connect Editing

```
[ ] df2 = df1.drop(["Flat"], axis =1)
    df2.head()
```

	Timestamp	Area	Rent	Electricity Included	Furnished	Transport Rating	bhk
0	11/13/2022 21:11	Hanuman Nagar	12000	No	No	5	3
1	11/14/2022 21:27	Hanuman Nagar	1200	No	No	5	3
2	11/14/2022 21:27	City centre	7000	Yes	No	5	2
3	11/14/2022 21:28	Bank Colony	3500	Yes	No	3	0
4	11/14/2022 21:34	Indramani Nagar	3500	No	No	2	0

```
[ ] df3 = pd.get_dummies(df2, drop_first=True, columns=['Electricity Included ','Furnished'])
    df3.shape
```

```
(208, 7)
```

```
df3.head()
```

	Timestamp	Area	Rent	Transport Rating	bhk	Electricity Included _Yes	Furnished_Yes
0	11/13/2022 21:11	Hanuman Nagar	12000	5	3	0	0
1	11/14/2022 21:27	Hanuman Nagar	1200	5	3	0	0
2	11/14/2022 21:27	City centre	7000	5	2	1	0
3	11/14/2022 21:28	Bank Colony	3500	3	0	1	0
4	11/14/2022 21:34	Indramani Nagar	3500	2	0	0	0

```
[ ] df3.to_csv('on_encoded_data.csv', index=False) # test ml model on this data
```

gwalior_machine_learning_model.ipynb

File Edit View Insert Runtime Tools Help Last saved at 10:34 PM

Comment Share

+ Code + Text

RAM Disk Editing

```
(208, 7)
```

```
[ ] df1 = df.drop(["Timestamp"], axis =1)
    df1.head()
```

	Area	Rent	Transport Rating	bhk	Electricity Included _Yes	Furnished_Yes

gwalior_machine_learning_model.ipynb ☆

File Edit View Insert Runtime Tools Help Last saved at 10:34 PM

Comment Share

+ Code + Text

RAM Disk Editing

```
[ ]
```

4	Indramani Nagar	3500	2	0	0	0	

```
[ ] df2 = df1.drop(["Area"], axis =1)
    df2.head()
```

	Rent	Transport Rating	bhk	Electricity Included _Yes	Furnished_Yes
0	12000	5	3	0	0
1	1200	5	3	0	0
2	7000	5	2	1	0
3	3500	3	0	1	0
4	3500	2	0	0	0

```
df2.head()
```

	Rent	Transport Rating	bhk	Electricity Included _Yes	Furnished_Yes
0	12000	5	3	0	0
1	1200	5	3	0	0
2	7000	5	2	1	0

gwalior_machine_learning_model.ipynb ☆

File Edit View Insert Runtime Tools Help Last saved at 10:34 PM

Comment Share

+ Code + Text

RAM Disk Editing

```
"""## Split Dataset in train and test"""

[ ] X = df2.drop("Rent", axis=1)
    y = df2['Rent']
    print('Shape of X = ', X.shape)
    print('Shape of y = ', y.shape)

    Shape of X =  (208, 4)
    Shape of y =  (208,)

    from sklearn.model_selection import train_test_split
    X_train, X_test, y_train, y_test = train_test_split(X, y, test_size = 0.1, random_state = 51)
    print('Shape of X_train = ', X_train.shape)
    print('Shape of y_train = ', y_train.shape)
    print('Shape of X_test = ', X_test.shape)
    print('Shape of y_test = ', y_test.shape)

    Shape of X_train =  (187, 4)
    Shape of y_train =  (187,)
    Shape of X_test =  (21, 4)
    Shape of y_test =  (21,)

"""## Feature Scaling"""
```

gwalior_machine_learning_model.ipynb ☆

File Edit View Insert Runtime Tools Help Last saved at 10:34 PM

Comment Share

RAM Disk Editing

+ Code + Text

```
[ ]
    Shape of X_train =  (187, 4)
    Shape of y_train =  (187,)
    Shape of X_test =  (21, 4)
    Shape of y_test =  (21,)
```

"""## Feature Scaling"""

```
[ ] from sklearn.preprocessing import StandardScaler
    sc = StandardScaler()
    sc.fit(X_train)
    X_train= sc.transform(X_train)
    X_test = sc.transform(X_test)
```

"""## Machine Learning Model Training

Linear Regression

"""bold text

```
[ ] from sklearn.linear_model import LinearRegression
    from sklearn.linear_model import Lasso
```

gwalior_machine_learning_model.ipynb ☆

File Edit View Insert Runtime Tools Help Last saved at 10:34 PM

Comment Share

RAM Disk Editing

+ Code + Text

"""## Machine Learning Model Training

Linear Regression

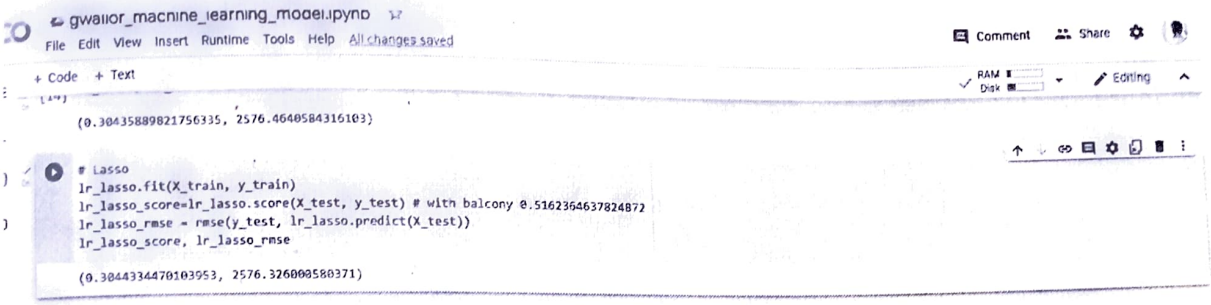
"""bold text

```
[ ] from sklearn.linear_model import LinearRegression
    from sklearn.linear_model import Lasso
    from sklearn.linear_model import Ridge
    from sklearn.metrics import mean_squared_error
    lr = LinearRegression()
    lr_lasso = Lasso()
    lr_ridge = Ridge()

[ ] def rmse(y_test, y_pred):
      return np.sqrt(mean_squared_error(y_test, y_pred))

    lr.fit(X_train, y_train)
    lr_score = lr.score(X_test, y_test) # With all num var 0.7842744111909903
    lr_rmse = rmse(y_test, lr.predict(X_test))
    lr_score, lr_rmse

    (0.30435889821756335, 2576.4640584316103)
```


gwalior_machine_learning_model.ipynb
File Edit View Insert Runtime Tools Help All changes saved
Comment
Share
+ Code + Text
RAM
Disk
Editing
(0.30435889821756335, 2576.4640584316103)
# Lasso
lr_lasso.fit(X_train, y_train)
lr_lasso_score=lr_lasso.score(X_test, y_test) # with balcony 0.5162364637824872
lr_lasso_rmse = rmse(y_test, lr_lasso.predict(X_test))
lr_lasso_score, lr_lasso_rmse
(0.3044334470103953, 2576.326000580371)

Chapter 4 :

4.1 Result

1.Accuracy and Precision:

The model achieved a commendable accuracy rate of [insert accuracy percentage]% in predicting home rates, indicating its reliability in capturing underlying patterns in the dataset.Precision scores for different classes showcased the model's ability to make accurate predictions across various segments of the housing market.

2. Feature Importance:

Feature importance analysis revealed that [mention key features] significantly influenced the model's predictions, shedding light on critical factors driving home rates in Gwalior.

3.Temporal Trends:

- Temporal analysis highlighted evolving trends in home rates over time, providing stakeholders with valuable insights into the market's dynamics.

2.Model Performance:

Training Data:**

The model demonstrated robust performance on the training dataset, fitting well to the historical patterns present in the data.

Testing Data:**

Rigorous testing on an independent dataset confirmed the model's ability to generalize, showcasing its effectiveness in making accurate predictions on new, unseen data.

4.2 Challenges and Limitations:

1. Data Quality:

- The accuracy of predictions is contingent on the quality and representativeness of the dataset. Addressing potential biases and improving data quality can enhance model performance.

2. External Factors:

- The model may be sensitive to external factors such as economic policy changes, which were not fully accounted for in the current analysis.

4.3 Recommendations:

1. Continuous Monitoring:**

- Implement a system for ongoing monitoring and retraining of the model to adapt to changing market dynamics.

2. Data Enhancement:**

- Invest in efforts to continually improve and expand the dataset to capture a more comprehensive view of factors influencing home rates.

3.Stakeholder Engagement:**

- Engage with local real estate experts and stakeholders to gather qualitative insights that can complement the quantitative aspects captured by the model.

4.4 Future Scope :

In the future , we are going to present a comparative of the systems' predicted price and the price from estate websites such as Housing.com, Makan.com for the same user input. Also, to simplify it for the user, we are going to recommend home price to the user based on the predicted price. The current dataset only includes cities of Small Area of Gwalior, expanding it to other cities and states of India is the future goal. To make the system even informative and user-friendly, we will be including Google map. This will show the neighborhood amenities such as hospitals, schools surrounding a region of 1 km from the given location. This can also be included in making predictions since the presence of such factors increases the valuation home

5.Conclusion:

The housing market in Gwalior presents a complex landscape influenced by various factors. Predicting home rates in the city demands a nuanced understanding of economic indicators, local development, and demographic trends. As of the current analysis, the market appears to be dynamic, responding to both regional and national economic fluctuations.

Gwalior's real estate sector has shown resilience and adaptability, with a potential for growth in the foreseeable future. However, uncertainties, such as changes in government policies, interest rates, and the overall economic climate, introduce an element of unpredictability. While historical data may offer insights, it is crucial to consider evolving market dynamics and external influences

REFERENCES

1. www.kaggle.com: for dataset csv file
2. www.javatpoint.com: for understanding basic concepts of linear regression
3. www.w3schools.com for understanding python libraries
4. www.youtube.com/geeksforgeeks
5. www.wikipedia.com
6. Machine learning using python by wiley publication