

Data Driven Insights: Leveraging Machine Learning in House Price Prediction.

Abdulmalik Umar, Yakubu Ibrahim Galadima, Firdausi Buhari Dalhat
Al Qalam-University, Katsina.

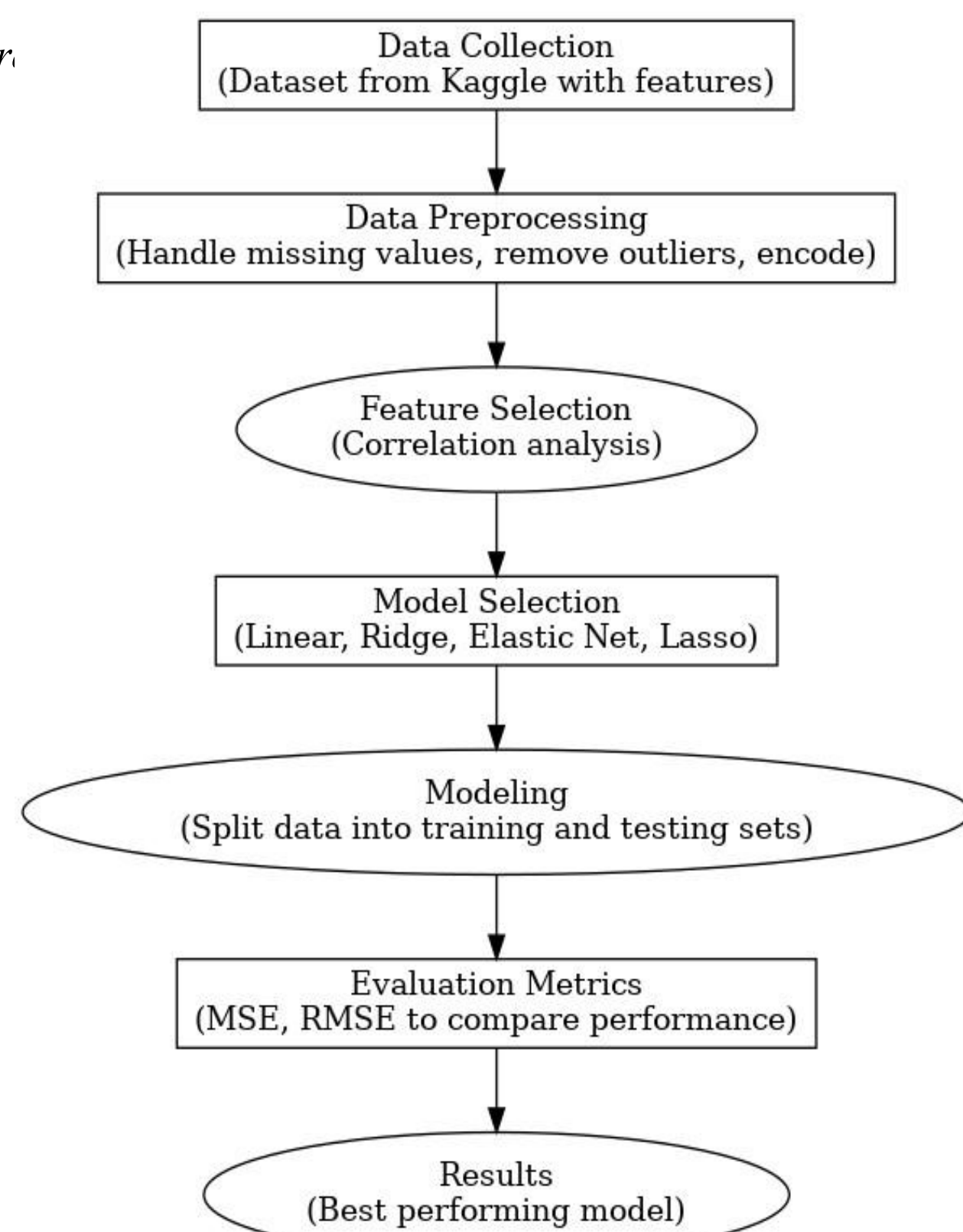
INTRODUCTION & AIM

- The real estate market is complex, influenced by a variety of factors such as location, property size, economic trends, and market conditions. Accurately predicting house prices is a critical task for buyers, sellers, and investors, enabling informed decisions and better market efficiency. This study explores the use of machine learning algorithms, to predict house prices based on factors like location, square footage, and number of bedrooms. The ultimate goal of this study is to develop a robust model that can reliably assist buyers, sellers and investors, in making informed decisions.
- The aim of this study is to investigate and compare the effectiveness of various machine learning models in predicting house prices. Specifically, it seeks to:
 - To analyze the key factors influencing house prices .
 - To build a machine learning model that delivers accurate predictions.
 - To Evaluate and compare the performance of different algorithms to determine the most reliable model for use in real estate decision-making and investment strategies.

METHOD

- Data Collection and preprocessing:** Dataset consists of real world data from King County, Washington, USA. The cities listed (e.g., Seattle, Bellevue, Redmond, Kent, Shoreline, Renton) have features like location, square footage, and number of bedrooms. We cleaned the data by handling missing values, removing outliers, and encoding categorical variables.
- Feature Selection:** Selected relevant features using correlation analysis.
- Modeling:** Splitting the data into two, the first part (training sample) is used to the train the algorithm and the second part (testing sample) is used to evaluate the performance of the algorithm.
- Evaluation Metrics:** Evaluated models using MSE and RMSE to compare performance.

Figure 1
Methodical Fr.



RESULTS & DISCUSSION

Figure 2
Understanding Variable Interactions in our housing Data

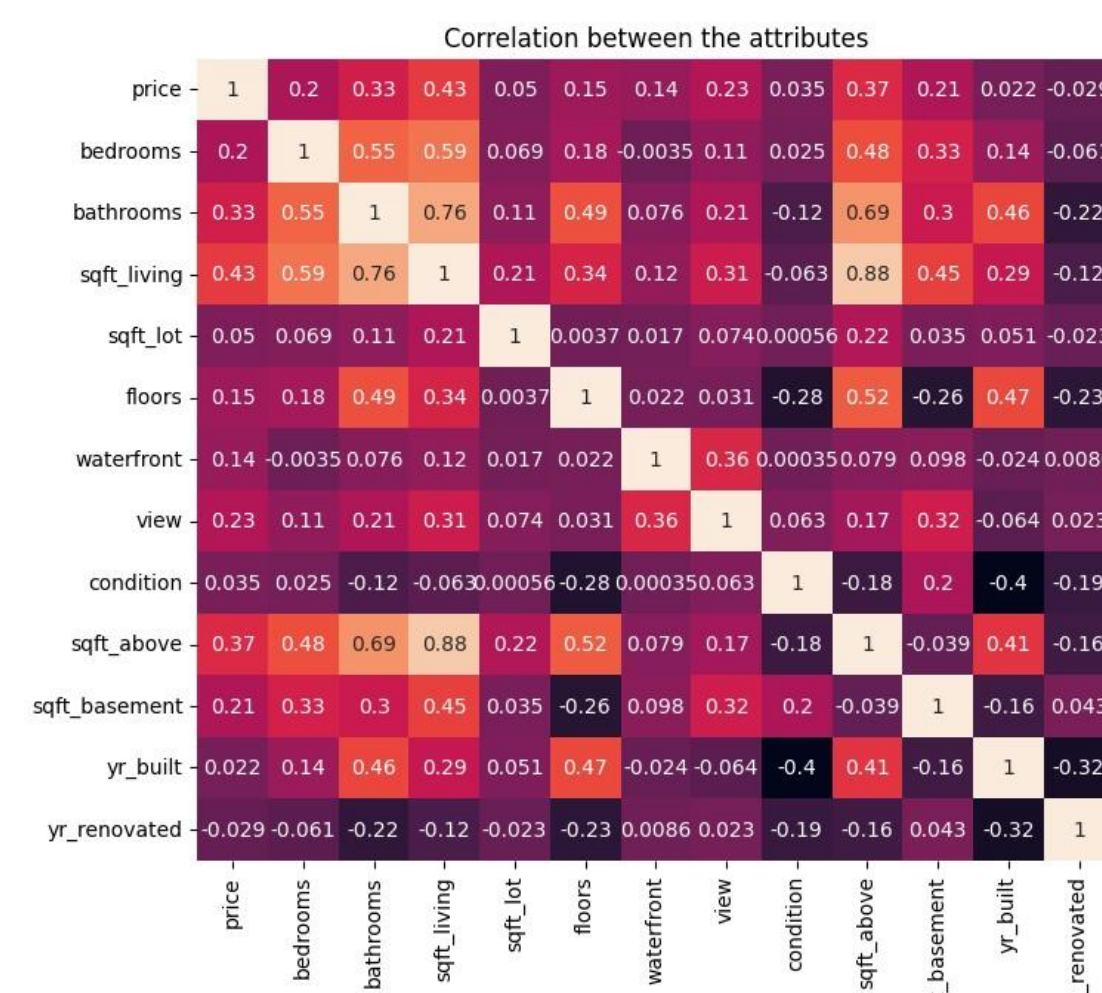


Figure 3
Price trend per Sqft during EDA

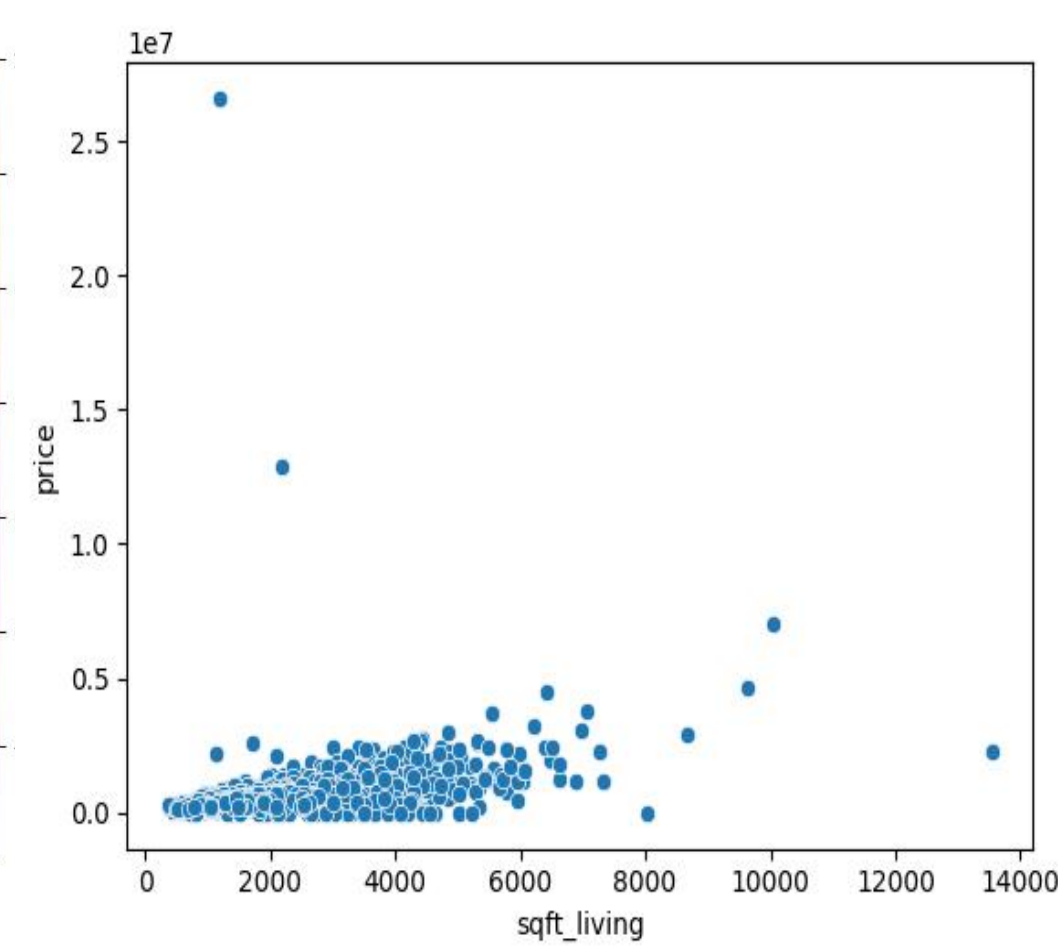


Figure 4
Assessment of Prediction Accuracy

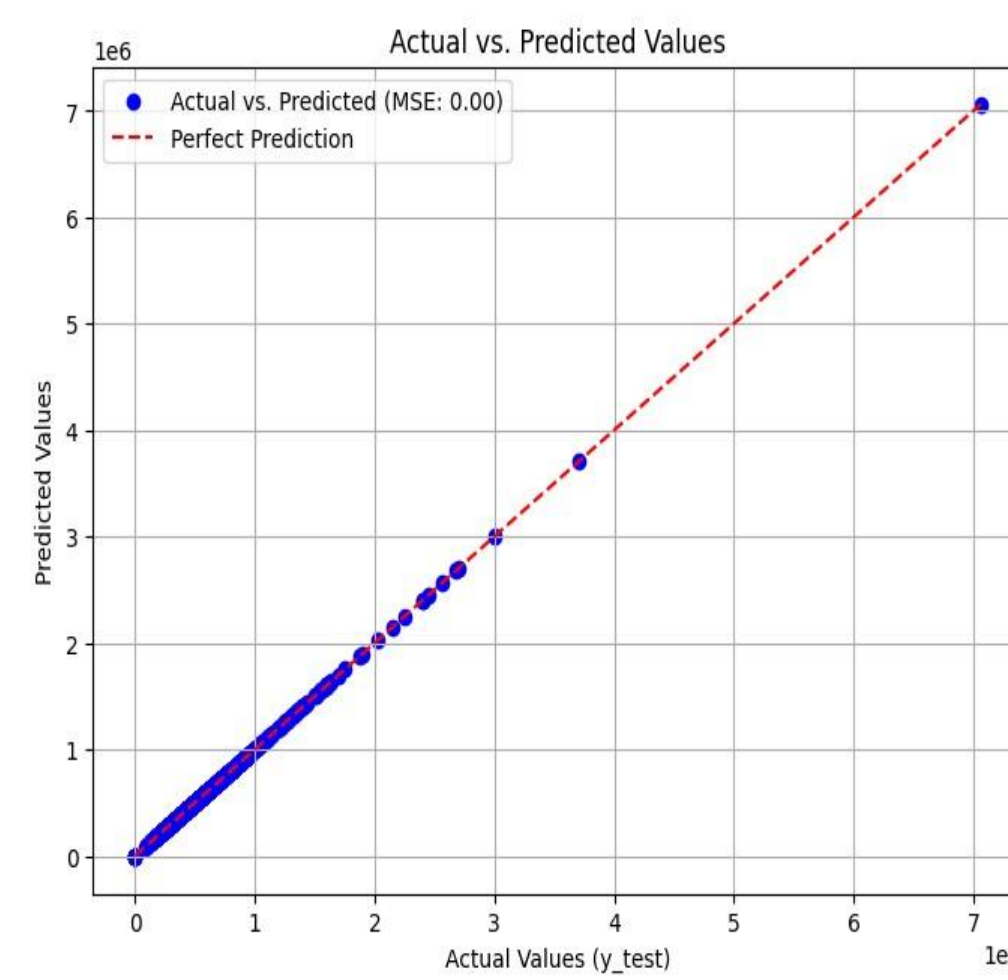
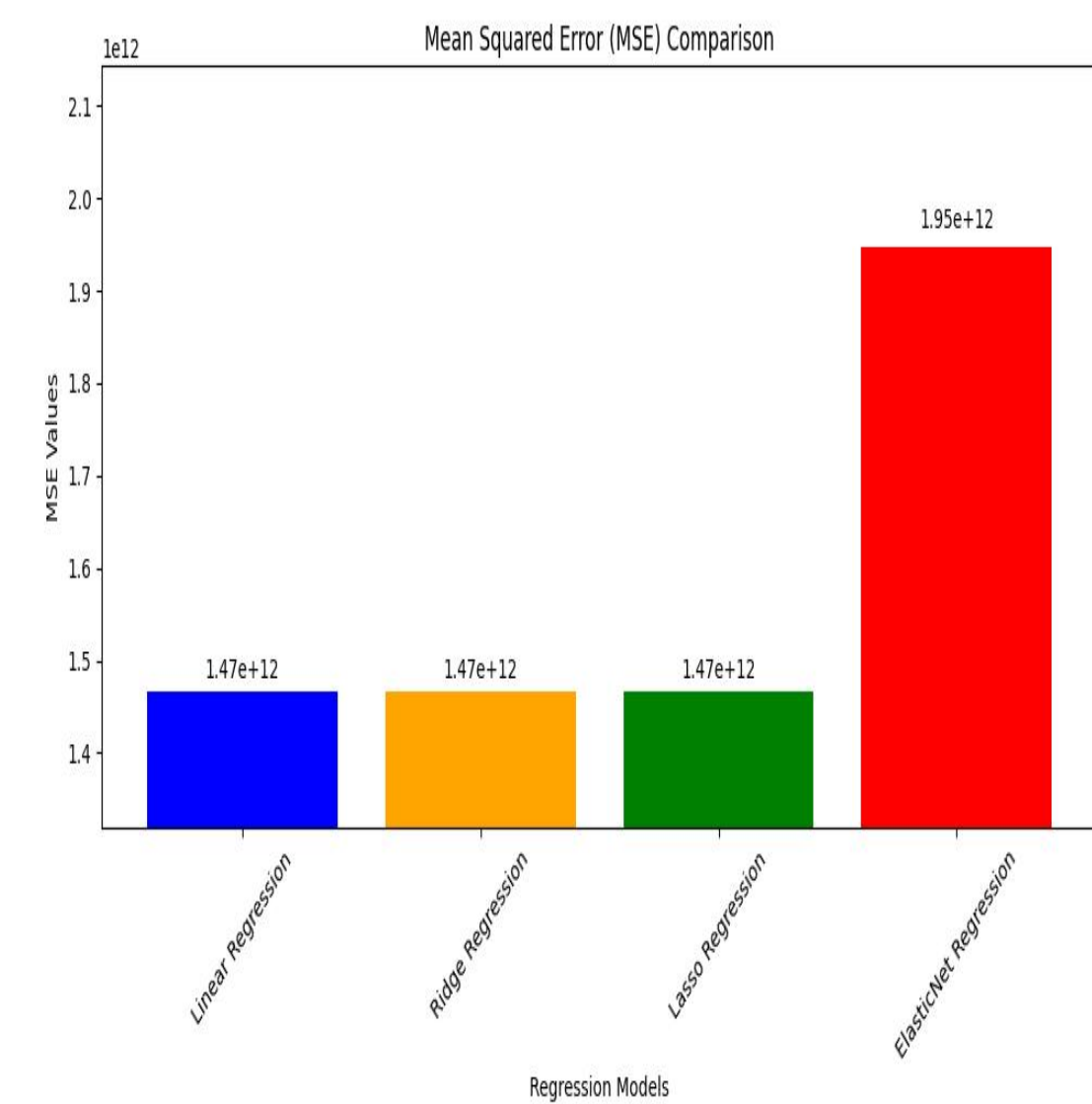


Figure 5
MeanSE Comparison across Regression Models



Model Performance Metrics based on the MSE values:

- Linear Regression** [MSE: 2.2808456841205377e-20 and RMSE: 1.5102468950872031e-10]
- Lasso Regression** [MSE: 1.474381685231829e-12 and RMSE: 1.2142411973046496e-06] and **ElasticNet Regression** [MSE: 1.474381685231829e-12 and RMSE: 1.2142411973046496e-06] have slightly higher MSE values.
- Ridge Regression** [MSE: 7.995445367364421e-19 and RMSE: 8.941725430454919e-10]: shows a slightly higher MSE compared to Ridge and ElasticNet, indicating it might be less effective in the house price prediction or for regression problems.

CONCLUSION

The study demonstrated that machine learning algorithms, particularly Linear Regression, are highly effective in predicting house prices when trained on a well preprocessed data-set. These findings from the Feature selection underscore the importance of considering the property features in real estate valuation.

FUTURE WORK / REFERENCES

Future research should investigate the use of more advanced machine learning techniques, Expand the dataset to include more diverse features and analyze house prices over time and across different regions to help develop models that are more generalizable and applicable in various contexts.