–

PARIS HOUSING PRICE PREDICTION PROJECT

CONTENTS

[INTRODUCTION 2](#_Toc136199910)

[DATASET 2](#_Toc136199911)

[ALGORITHM EXPLANATION 3](#_Toc136199912)

[Random forest Classifier: 3](#_Toc136199913)

[Decision Tree Classifier: 3](#_Toc136199914)

[Lasso Regression: 3](#_Toc136199915)

[SVR (Support Vector Regression): 4](#_Toc136199916)

[COMPARATIVE ANALAYSIS 5](#_Toc136199917)

[CONFUSION MATRIX 5](#_Toc136199918)

[CONCLUSION 6](#_Toc136199919)

INTRODUCTION

In the real estate sector, house price forecasting is quite important. Accurate price estimation enables buyers to make wise choices and sellers to set fair pricing. In this project, our goal was to create a reliable model that could forecast Parisian home prices based on a specific set of features. The models were trained and tested using the Paris housing Kaggle dataset. Finding an algorithm that could correctly forecast property values based on the provided features was the main goal. The Mean Squared Error (MSE) and R2 Score of four regression algorithms Decision Tree Regressor, Random Forest Regressor, SVR, and Lasso were implemented and assessed.

DATASET

**Dataset Testing (20% Test Size, Random State 15)**

The Kaggle dataset of Paris housing was collected and preprocessed. Missing values were handled, outliers were addressed, and categorical variables were properly handled too. Feature engineering techniques were employed to extract relevant information from the dataset. The preprocessed dataset was split randomly into 80% training and 20% testing sets using a random state of 15. This split allowed us to train the models on a significant portion of the data and evaluate their performance on unseen data.

ALGORITHM EXPLANATION

# Random forest Classifier:

The Random Forest Regressor is an ensemble algorithm that combines multiple decision trees to make predictions. It was trained on the training set and evaluated on the testing set.

Mean Squared Error (MSE): 0.000015

R2 Score: 0.999998

# Decision Tree Classifier:

The Decision Tree Regressor is a non-parametric algorithm that constructs a tree-like model based on features and splits. It was trained on the training set and evaluated on the testing set.

Mean Squared Error (MSE): 0.000030

R2 Score: 0.999996

# Lasso Regression:

Lasso Regression is a linear regression technique that performs feature selection and regularization. It was trained on the training set and evaluated on the testing set.

Mean Squared Error (MSE): 0.000011

R2 Score: 0.999999

# SVR (Support Vector Regression):

SVR (Support Vector Regression): SVR is a regression algorithm that utilizes support vector machines to approximate the target variable. It was trained on the training set and evaluated on the testing set.

Mean Squared Error (MSE): 0.002892

R2 Score: 0.999648

COMPARATIVE ANALAYSIS

The performance of the regression algorithms on the testing set is as follows:

* Decision Tree Regressor achieved an MSE of 0.000030 and an R2 Score of 0.999996. The algorithm demonstrated excellent predictive accuracy, indicating that it can closely approximate house prices in Paris.
* Random Forest Regressor outperformed the other algorithms, achieving an impressively low MSE of 0.000015 and an R2 Score of 0.999998. The ensemble of decision trees in Random Forest provided.

CONFUSION MATRIX

A confusion matrix is not appropriate for this project because it includes regression rather than classification. For evaluating the effectiveness of regression models, evaluation metrics like Mean Squared Error (MSE) and R2 Score are more appropriate.

CLASSIFICATION REPORT

Since this project involves regression, a classification report is not applicable. Classification reports are used for evaluating the performance of classification models based on precision, recall, and F1-score. However, in regression tasks, evaluation metrics such as MSE and R2 Score are more commonly used.

CONCLUSION

In this project, we used the Kaggle dataset to estimate housing prices in Paris using four regression algorithms: Decision Tree Regressor, Random Forest Regressor, SVR, and Lasso. With MSE values close to 0 and R2 Scores near to 1, the results showed that all algorithms performed remarkably well. The least MSE values were obtained by the Lasso and Random Forest Regressor, demonstrating their higher prediction accuracy. Overall, these algorithms demonstrate how AI might be used to reliably forecast home prices, aiding both buyers and sellers in the real estate market.

It is crucial to remember that the dataset's quality and representativeness were crucial to the models' success. Additional research and testing with different features and algorithms could perhaps enhance the model.

Lastly, this project predicts the housing price in Paris and give a precise idea of how prediction can be made on the basis of given data of houses such as rooms, square yards and etc.