**HOUSE PRICE PREDICTION USING MACHINE LEARNING TECHNIQUES**

Basit Ali CS201177, Usman Saeed Motiwala CS201157, Abdul Moiz CS201004

Department of Computer Science, Dha Suffa University

Artificial Intelligence Lab, CS-3102L

Ms Aniqa Hussain

June 20, 2023

**ABSTRACT**

This research paper aims to predict house prices using two machine learning algorithms, Support Vector Machine (SVM) and Random Forest Classifier (RFC). The purpose of this research is to compare the performance of SVM and RFC in predicting house prices and determine which algorithm yields accurate results. The study focuses on utilizing the number of bedrooms and bathrooms as the input features for the prediction task. The findings of this study can provide valuable insights for real estate agents and homeowners in estimating house prices more accurately.

**KEYWORDS:** Machine Learning, House Prices, Support Vector Machine, Random Forest Classifier, Accuracy Comparison.

1. **INTRODUCTION**

Predicting house prices accurately has been a significant research area in the field of machine learning. Accurate house price prediction is crucial for various purposes, such as real estate investments, mortgage approvals, and property valuation. House prices trends are not only the concerns for buyers and sellers, but they also indicate the current economic situations [1]. That is why it is important to predict accurate house prices without bias to help buyers and sellers.

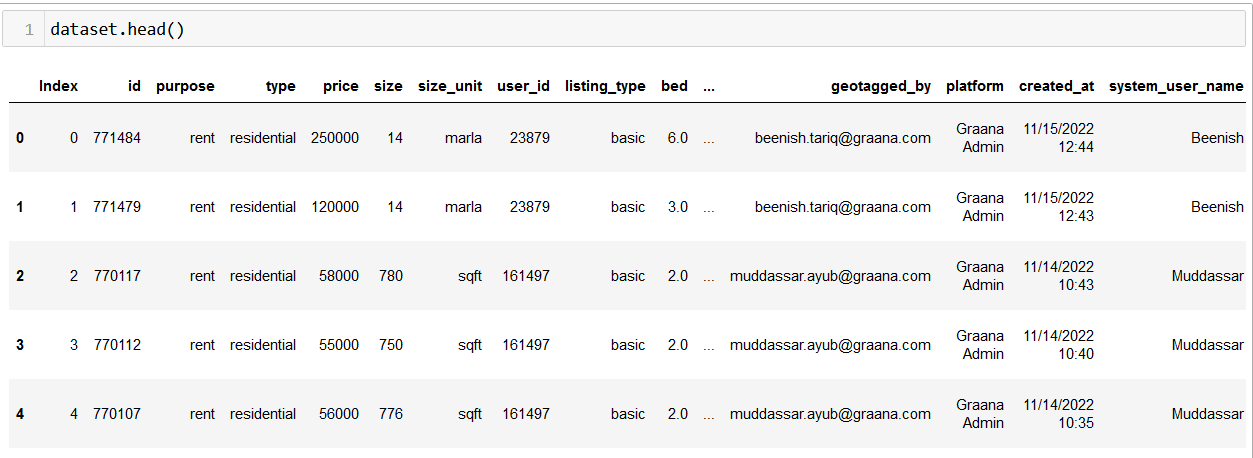
The research compares the performance of Support Vector Classifier (SVC) and Random Forest Classifier (RFC) algorithms in this prediction task. There are many factors which affect house prices such as the number of bedrooms and bathrooms. This study is to compare which among these two algorithms provides more accurate prediction. The research is being conducted on houses of Pakistan and the dataset is used from Kaggle of Graana to train the machine learning model.

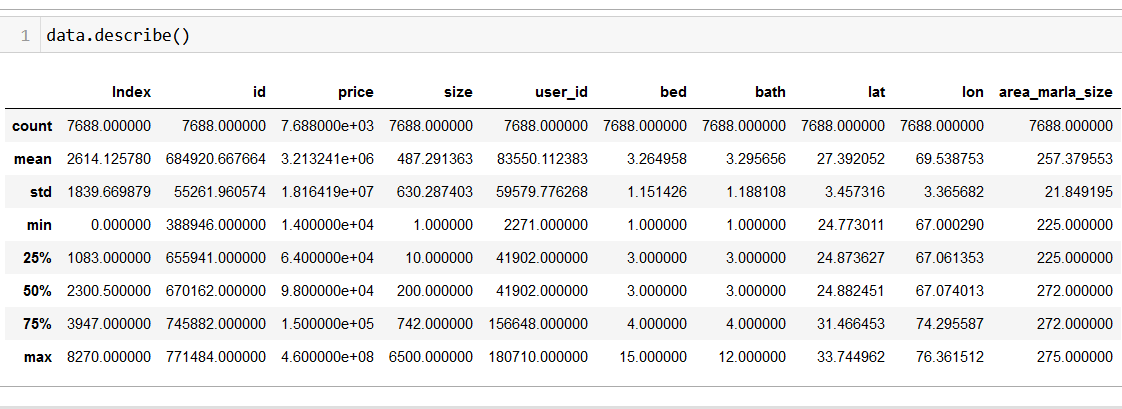
1. **LITERATURE REVIEW**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Name** | **Description** | **Pros** | **Cons** | **Reference** |
| **Support Vector Machine** | SVM is a machine learning algorithm that classifies data by finding the best hyperplane to separate different classes, maximizing the margin between them | Effective for high-dimensional data.  Robust against overfitting.  Versatile kernel functions.  Memory-efficient. | Computationally expensive.  Sensitivity to parameter tuning.  Lack of probabilistic outputs.  Limited effectiveness with noisy data. | Auria, L., & Moro, R. A. (2008). Support vector machines (SVM) as a technique for solvency analysis. [2] |
| **Random Forest Classifier** | Random Forest is an ensemble learning algorithm that combines multiple decision trees and combines their prediction through voting or averaging to make accurate predictions. | Good performance on large datasets.  Reduces overfitting.  Robust to outliers and noise.  Deals with irrelevant inputs.  Computationally Scalability. | Low prediction accuracy.  High variance  Computationally expensive training.  Lack of interpretability.  Requires careful parameter tuning.  Difficult to visualize the decision-making process. | Montillo, A. A. (2009). Random forests. Lecture in Statistical Foundations of Data Analysis. [3] |

1. **CONCEPT/ANALYSIS**

**Data Set:**

****Figure 1:** *Dataset first 5 records or rows.*

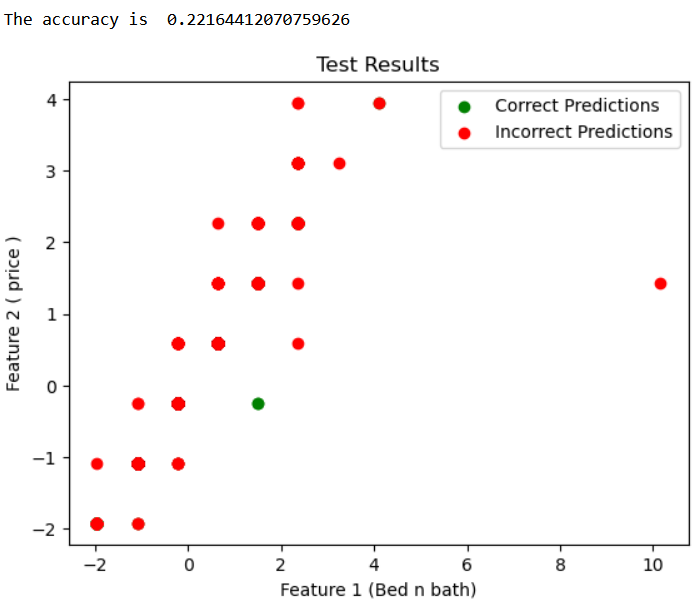
**Figure 2:** *Dataset description.*

To predict house prices using Support Vector Classifier and Random Forest Classifier, we first removed any data with null values, then we used two features from the dataset to predict house price: number of bedrooms and bathrooms. The dataset was divided into training and testing sets (80% training set and 20% testing set) to evaluate the performance of the algorithms. SVC was implemented using a radial basis function kernel, while RFC was trained with 100 decision trees. The house price of test data is predicted using bedrooms and bathrooms and is then compared with its actual value of test data to calculate Accuracy Score. The models were evaluated on bases of Accuracy score to assess their predictive accuracy.

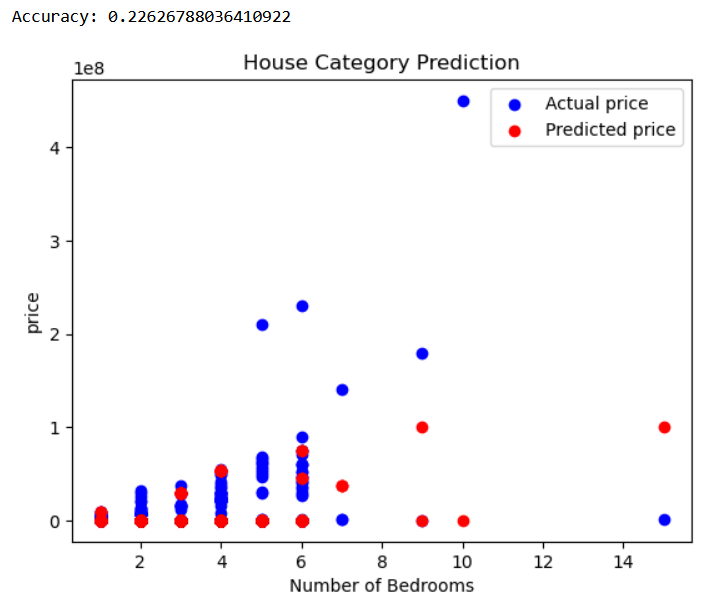
1. **RESULTS**

**Screenshot:**

**Figure 3:** *SVM Graph of number of correct and incorrect predictions.*

****

**Figure 4:** *Random Forest Classifier Prediction Graph.*

****

The results of our analysis indicate that both Support Vector Classification and RFC performed well in predicting house prices. However, RFC exhibited slightly better performance compared to SVC. The Accuracy Score for RFC were higher than those of SVC, indicating that SVC produced more accurate predictions. The feature importance analysis revealed that the number of bedrooms and bathrooms were the most significant factors influencing house prices, as captured by both algorithms. From the result of the Accuracy Score we can see that for this particular dataset RFC is better than SVC.

1. **CONCLUSION**

In this research paper, we compared the performance of two machine learning algorithms, Support Vector Classifier and Random Forest Classifier, in predicting house prices. The results demonstrated that RFC outperformed SVC in terms of predictive accuracy, as indicated by their accuracy score. These findings suggest that both algorithms have similar results for house price prediction, with RFC slightly more accurate and better suited for this dataset than SVC. However, further research can explore other machine learning algorithms and features to improve the accuracy of house price predictions.

|  |
| --- |
|  |
|  |

|  |  |
| --- | --- |
| [1] | D. Kumar, S. and K. , "HOUSE PRICE PREDICTION USING RANDOM FOREST AND CNN," *International Research Journal of Modernization in Engineering Technology and Science,* vol. 4, no. 8, p. 5, 2022. |
| [2] | A. Laura and R. A. Moro, "Support vector machines (SVM) as a technique for solvency analysis.," *Discussion Papers,* p. 18, 2008. |
| [3] | M. Albert A, "Random forests," *Lecture in Statistical Foundations of Data Analysis,* p. 28, 2009. |
| [4] | W. and J. Yang, "Housing price prediction using support vector regression," *Master's Projects,* p. 58, 2017. |

1. **REFERENCE**