International Journal of Multidisciplinary Engineering in Current Research - IJMEC
Volume 9, Issue 1, January-2024, http://ijmec.com/, ISSN: 2456-4265

HOUSE RENT PREDICTION OF MAIJDEE TOWN-NOAKHALI, BANGLADESH, A STUDY OF SUPERVISED MACHINE LEARNING

Forhad Mahmud[1], Jannatul Naime[2], Md. Rayhan[3]
[1, 2, 3]Dept. of Applied Mathematics, Noakhali Science and Technology University, Bangladesh

Abstract-

In this article, we proposed a machine learning-based approach to predict house rent prices. We use a dataset of real estate listings containing various features such as main-road, area, number of rooms, gas-line and other advantage to train our model. We first pre-process the data to remove missing values, handle outliers, and convert categorical variables to numerical ones. We then explore and analyze the data using various visualization techniques. We use multiple linear regression, random forest regression models and decision tree to predict house rents based on the available features. We evaluate the performance of the models using various metrics such as mean squared error, root mean squared error, and R-squared. Our results show that the linear regression model outperforms the random forest regression model and decision tree in terms of prediction accuracy. Overall, this article demonstrates the effectiveness of machine learning techniques in predicting house rent prices. The proposed approach can help landowners and renters make informed decisions on pricing and rental contracts, as well as assist real estate, agents and property managers in setting the right rent prices for their properties.

Keywords: Rental Price, Machine Learning, Linear Regression, Random Forest, Decision Tree.

I. INTRODUCTION

Machine learning is a technology that allows computers to learn from past data and use that learning to make predictions. It uses algorithms to build models and is used for tasks like predicting house rent, recognizing images and voice recognition, filtering emails, tagging photos on Facebook and recommending products or services. The three types of machine learning language which are supervised, unsupervised, and reinforcement learning. A division of machine learning and artificial intelligence is supervised machine learning. It is unique in that it uses labeled datasets to teach computers how to effectively categorize data or predict outcomes. Unsupervised learning refers to a class of algorithms that discovers patterns from unlabeled data. The idea is to drive the machine to create a clear representation of its surroundings and then produce inventive material from it by forcing it to imitate, which is a key method of learning in humans. In the field of machine learning known as

DOI: https://doi-ds.org/doilink/01.2024-53988668

reinforcement learning, behaviors that intelligent agents should execute in a given environment to maximize the concept of cumulative reward are analyzed.

Machine learning has a wide range of applications across various fields, and it is being used to solve complex problems and automate tasks in various industries. Here are some of the applicable fields of machine learning such as Traffic alerts, Health care, social media, Transportation and Commuting, Products Recommendations, Virtual Personal Assistants, Self-Driving Cars, Dynamic Pricing, Finance, Retail and E-commerce, Fraud Detection etc. The house rent price prediction model using linear regression is a machine learning model that is designed to predict the rental price of a house based on its features. Thinking the importance of this prediction a number of authors [1-16] have tried to solve such models.

We discussed "House Rent Prediction" where different types of machine learning classifier algorithms and data mining techniques are used to predict the house rent price according to desired features. All training data and test data are used to develop a machine learning classification algorithm with the best algorithm being selected via machinelearning techniques. The dataset has 200 record house samples where each of them having 9 exclusive features which was collected from Housing State at Maijdee Town in Noakhali. This dataset is tested by a machine learning classification algorithm, and there are no missing elements in the dataset. We have run several types of classification algorithms such as Multiple Linear Regression (MLR), Random Forest Classifier (RFC), Decision Tree etc. To predict the accuracy rate of house price according to features. The goal of this paper is to provide landowners and real estate companies with a tool to make informed decisions about setting the rental price for their properties, and to give them insights into the factors that influence the rental price of a house. This could help them avoid setting prices that are too high or too low, leading to better returns on their investments.

II. METHODOLOGY

Here we will discuss the methodology in step by step to determine the house price prediction. Our process can be divided into several steps. The First step is the data collection phase, during which we have collected some raw data for our dataset from Housing State at Maijdee town in Noakhali. This will be utilized to train the machine-learning model. The dataset generated at this step is entirely made up of unstructured raw data. There are 200 rows and 9 columns in the dataset. According to the dataset, the prices are listed in taka (Tk) and the plot size is listed in square feet (sq. ft.). The price column in the dataset is the dependent variable, whereas the remaining columns are independent variables (also called features). Here we use "pandas" which is a function found in the scikit-learn python library which is used for reading and analyzing the dataset.

DOI: https://doi-ds.org/doilink/01.2024-53988668

	price	area	bedrooms	bathrooms	stories	Main-road	Dining-room	Gas-line	Furnishing-status
0	10,000	900	3	2	3	Yes	Yes	Yes	Furnished
1	9000	800	2	3	4	Yes	Yes	Yes	Furnished
2	9000	950	3	2	2	Yes	Yes	No	Semi-furnished
3	7500	750	2	2	2	No	Yes	Yes	Furnished
4	8300	720	2	2	2	Yes	Yes	Yes	Furnished
5	7600	750	2	3	1	No	Yes	Yes	Semi-furnished
6	7000	710	2	3	4	No	Yes	Yes	Semi-furnished
7	8000	800	3	3	2	Yes	Yes	No	Unfurnished
8	6500	700	2	1	2	Yes	No	Yes	Furnished
9	7000	900	3	2	4	Yes	Yes	Yes	Unfurnished

Table. 1. Data

Cleaning

As part of the data cleaning process, we looked to see if any rows in the raw dataset had any missing values. However, no empty rows were found in our dataset sample. So, we moved on to the next phase, data pre-processing.

2. Data Pre-processing

In this phase, we formatted our unstructured dataset such that it could be used to train a machine learning model. All independent variables must store data as numbers rather than text since we must apply a multivariate regression model, and this model must be trained using our dataset.

However, in our dataset, the columns named for main-road, dining-room, and gas-line contained text data in the form of yes or no. We used the "binary map" function found in the scikit-learn python library to convert this into numerical data, where "yes" is represented by the number 1 and "no" by the number 0. The information given in the column named "furnishing-status" represents the furnishing status of the house which is text data. We applied the concept of "one hot encoding" to turn this text data into numerical data. We subdivided the "furnishing-status" column into three new columns, "furnished," "semi-furnished," and "unfurnished."

The data will now be stored in the new columns as binary numbers, where 1 will stand for "true" or "yes" and 0 will stand for "false" or "no." The original “furnishing-status” column is then removed because it is no longer needed. After completing the above processes, the dataset includes only data that can be represented numerically, which makes it qualified to train the model.

	price	area	bedrooms	bathrooms	stories	Main-road	Dining-room	Gas-line	Furnished	Semi-furnished	Unfurnished
0	10,000	900	3	2	3	1	1	1	1	0	0

DOI: https://doi-ds.org/doilink/01.2024-53988668

1	9000	800	2	3	4	1	1	1	1	0	0
2	9000	950	3	2	2	1	1	0	0	1	0
3	7500	750	2	2	2	0	1	1	1	0	0
4	8300	720	2	2	2	1	1	1	1	0	0
5	7600	750	2	3	1	0	1	1	0	1	0
6	7000	710	2	3	4	0	1	1	0	1	0
7	8000	800	3	3	2	1	1	0	0	0	1
8	6500	700	2	1	2	1	0	1	1	0	0
9	7000	900	3	2	4	1	1	1	0	0	1

Table. 2: Dataset after Pre-processing

3. Data Visualization

Data visualization is an essential part of data processing in data prediction. It is the process of representing data and information in a graphical or visual format. It has various effects, including improved understanding, increased engagement, better decision-making, enhanced communication, identification of outliers, fluctuations, and improved data quality. We can visualize data on different perspective such as numerical values of different features, string values of dataset, furnishing status and heat maps etc.

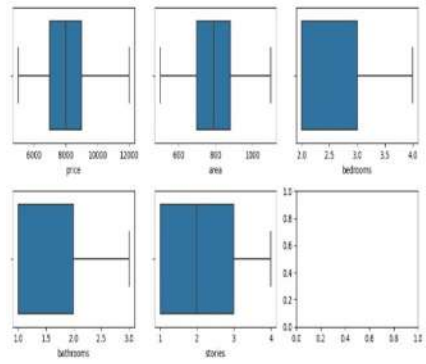


Fig. 1: Data Visualization (Numerical

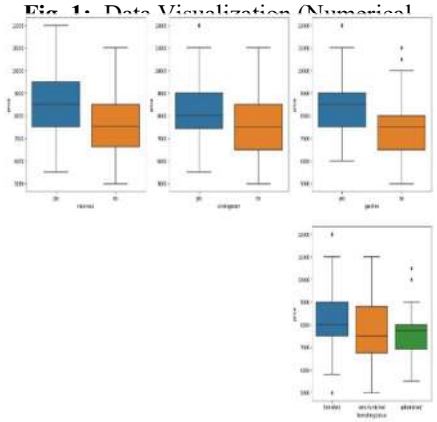


Fig. 2: Data Visualization (String Values)

DOI: https://doi-ds.org/doilink/01.2024-53988668

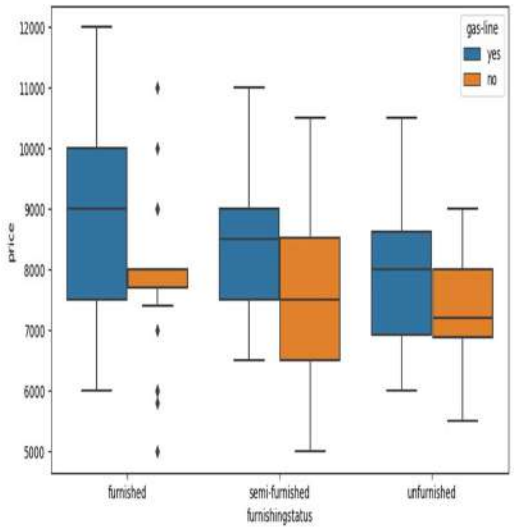


Fig. 3: Data Visualization (Furnishing Status)

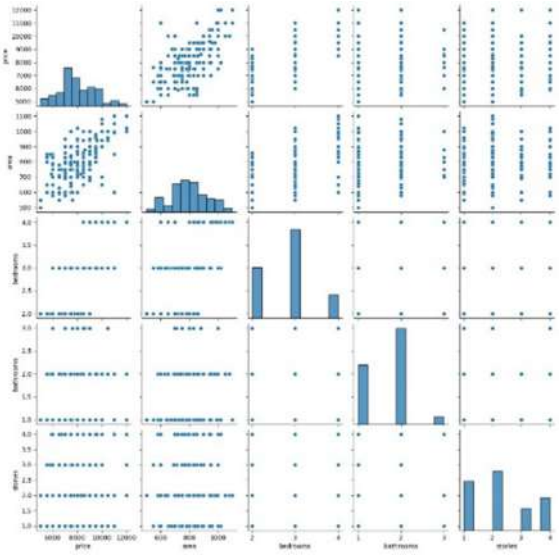


Fig. 4: Data Visualization (Pair Plot)

DOI: https://doi-ds.org/doilink/01.2024-53988668

We decided to visualize the correlation of the features after encoding them, which can be seen in the heat map below.

Fig. 5: Correlation between Features

4. Model Selection

In order to predict the house, rent for users three classification algorithms were utilized in this work objective. The applying classifiers are Multiple Linear Regression, Decision Tree, Random Forest Classifier. The house rent data set was collected and applied to each model to predict or target the house rent, and the success of the classification algorithms is measured based on the accuracy of success.

5. Evaluating the Model

In order to determine the model's performance, we developed a scatter plot that compares the actual house prices listed in the dataset with the prices projected by the model.

DOI: https://doi-ds.org/doilink/01.2024-53988668

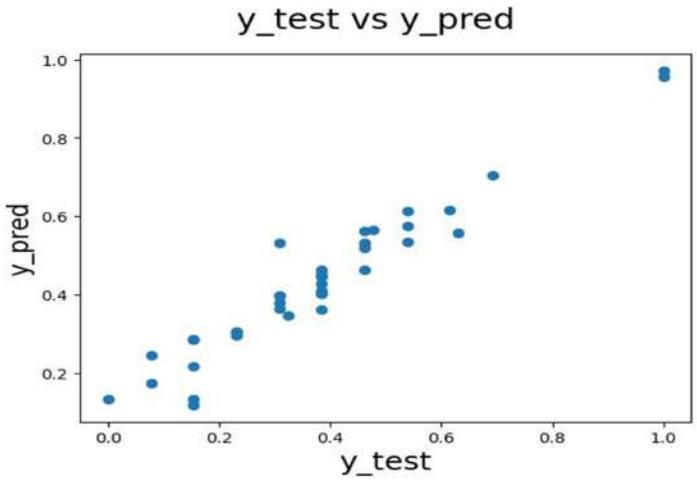
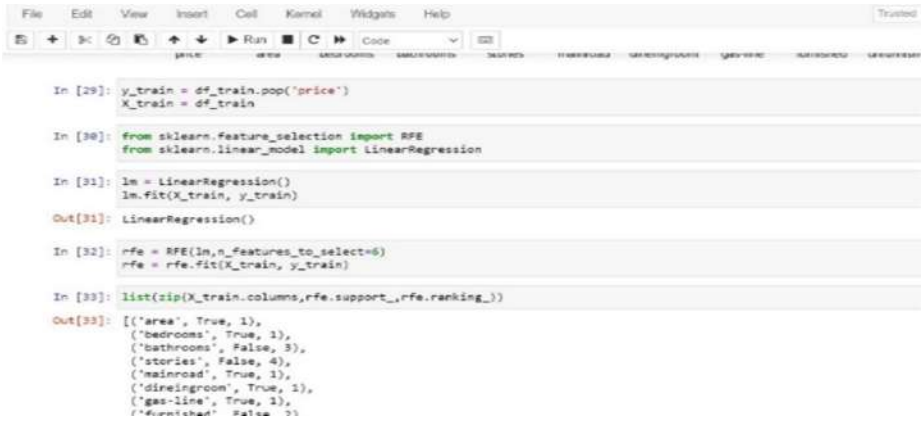


Fig. 6: Scatter Plot Showing Difference between Actual Price(y_test) and Predicted Price(y_pred).

III. Results and Discussion

In this analysis, we have used three machine learning algorithms to predict house rent prices based on various features. The algorithms used were linear regression, random forest, and decision tree. We use a dataset which contain 200 house samples and 9 exclusive features which was collected from Housing State at Maijdee Town in Noakhali. For Linear Regression Model the price Y will be our predicted value and all the features belong X to the dataset will be our targeting value. Then we must train our dataset by Linear Regression Model where "train size = 0.8'' and "test size=0.2". Now the testing process is given below.



```
In [29]: y_train = df_train.pop('price')
         X_train = df_train

In [30]: from sklearn.feature_selection import RFE
         from sklearn.linear_model import LinearRegression

In [31]: lm = LinearRegression()
         lm.fit(X_train, y_train)
Out[31]: LinearRegression()

In [32]: rfe = RFE(lm,n_features_to_select=6)
         rfe = rfe.fit(X_train, y_train)

In [33]: list(zip(X_train.columns,rfe.support_,rfe.ranking_))
Out[33]: [('area', True, 1),
          ('bedrooms', True, 1),
          ('bathrooms', False, 3),
          ('stories', False, 4),
          ('mainroad', True, 1),
          ('diningroom', True, 1),
          ('gas-line', True, 1),
```

Fig. 7: Testing (Multiple Linear Regression Model)

DOI: https://doi-ds.org/doilink/01.2024-53988668

Now the accuracy rate of Multiple Linear Regression Model is given below.

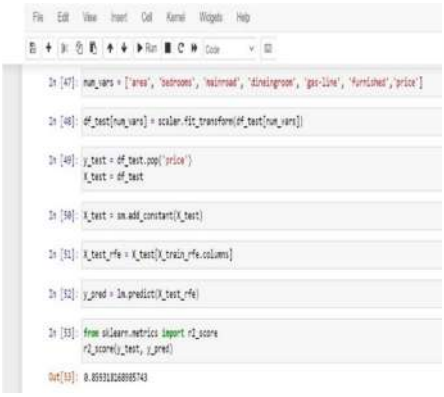


Fig. 8: Accuracy Rate (Multiple Linear Regression Model)

For checking the working process of our Multiple Linear Regression Model let's. For example, let's take a house which has area (825 sq. ft.), three bedrooms, two bathrooms and which is situated on third floor of a random building. Then the price of the house is 9356.1177 Tk which is shown in the figure below.

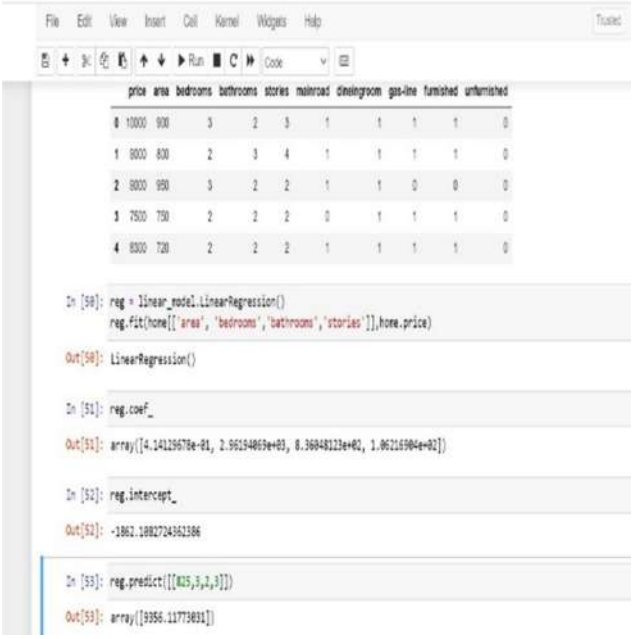


Fig. 9: Prediction of (Multiple Linear Regression Model)

DOI: https://doi-ds.org/doilink/01.2024-53988668

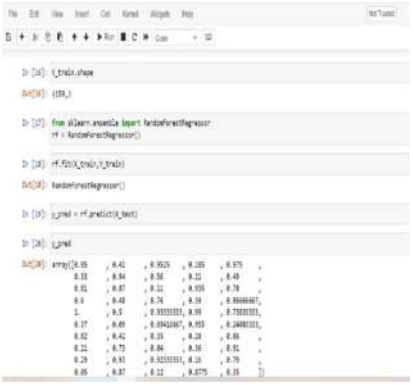


Fig. 10: Testing (Random Forest)

Now for checking the accuracy rate through Random Forest Model. Firstly, we select random samples from trained data. Secondly the algorithm will construct a decision tree for every trained data and then voting will take place by averaging the decision tree. Finally, we select the most voted result as a final prediction result. Now the accuracy rate of the Random Forest Model is given below.

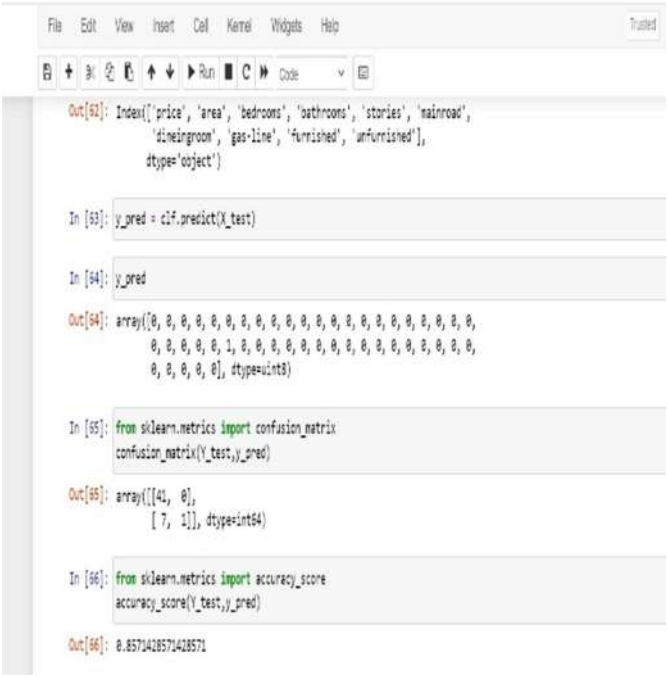


Fig. 11: Accuracy Rate (Random Forest)

DOI: https://doi-ds.org/doilink/01.2024-53988668

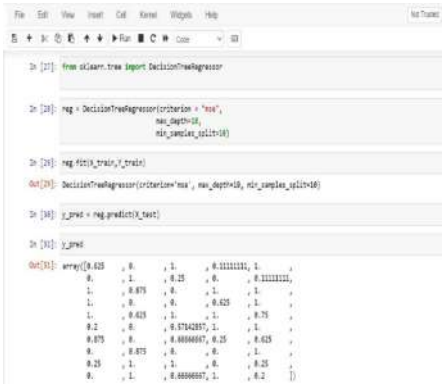


Fig. 12: Testing (Decision Tree)

In order to calculate accuracy rate through a decision tree we must instantiate the Decision Tree Model. Then we must split the trained data where train size is 0.8 and test size is 0.2. And then we must find the mean square error and root mean square error. After doing all these steps we get the mean square error is 0.2865 and the root mean square error is 0.53529.

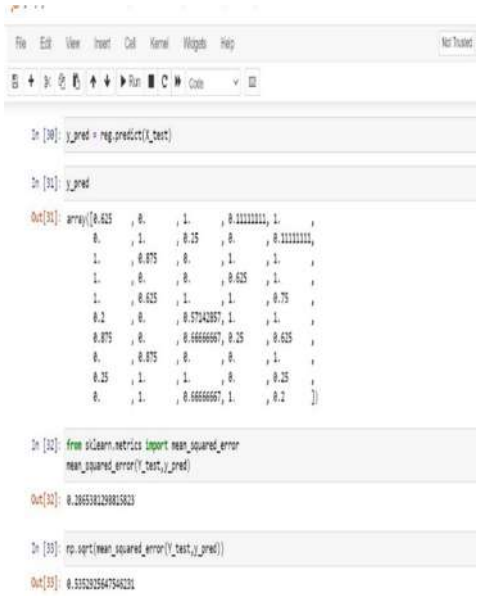


Fig. 13: Accuracy Rate (Decision Tree)

Now for checking the highest and lowest accuracy rate we must train our data through cross value score. Hence the accuracy rate of the cross-value section is given below.

DOI: https://doi-ds.org/doilink/01.2024-53988668

Fig. 14: Accuracy Rate (Cross Value Score)

The results of our analysis show that the linear regression algorithm achieved an accuracy rate of 85.9%, while the random forest algorithm had a slightly lower accuracy rate of 85.7%. The cross-validation technique had the highest score of 93.33% and the lowest score of 60%, which indicates that this technique is more reliable and effective than the individual algorithms.

We also evaluated the decision tree algorithm using mean square error and root mean square error. The mean square error was 0.2865 and the root mean square error was 0.53529. These results suggest that the decision tree algorithm is not as accurate as the other two algorithms and may require further optimization.

Overall, our results indicate that machine learning algorithms can be useful in predicting house rent prices. Based on the analysis, it was found that the multiple linear regression model is well-suited for predicting house prices.

CONCLUSION

Our analysis highlights the importance of using machine learning algorithms in predicting house rent prices. The results obtained can help landowners and renter estimate the cost of renting a property in a specific area. There are several ways to improve the accuracy and reliability of a house rent prediction model, including increasing the size and quality of the dataset, utilizing feature engineering to capture additional variables, using more advanced machine learning algorithms, regularly updating the model to reflect changes in the housing market, and collaborating with domain experts to gain valuable

DOI: https://doi-ds.org/doilink/01.2024-53988668

insights. By taking these steps, the model can provide more accurate and useful predictions for a wider range of scenarios.

There are many opportunities for future work around house rent prediction for example time series analysis techniques can be applied to historical rent data to predict future rent prices based on past trends and patterns, experimenting with different feature selection techniques can help identify the most important variables for predicting rent prices, simplifying models and reducing the risk of overfitting. Overall, there are many opportunities for future work around house rent prediction, and continued research could have significant implications for the real estate industry and rental markets worldwide.

REFERENCES

[1] Qadir, M. a. (2020). House Rent Prediction using Machine Learning Techniques in Lahore, Pakistan.

[2] Reddy, S. a. (2019). House Rent Prediction using Random Forest and Gradient Boosting in Python.

[3] Sharma, S. &. Singh (2020). House rent prediction using machine learning techniques. In 2020 6th International Conference on Advanced Computing and Communication Systems (ICACCS) (pp. 1476 1480). IEEE.

[4] Zhang, Q. & Zhang, C. (2020). House Rent Prediction using Neural Network and Support Vector Machine in Hangzhou, China.

[5] Qu, C. L. (2021). House Rent Prediction in Beijing Based on Machine Learning.

[6] Singh, D. K. (2021). Comparative analysis of different machine learning algorithms for house rent prediction.

[7] Roy, & Mahmud. (2019). Machine learning based prediction of house rent prices in Dhaka.

[8] Saleem, M. &. Hussain (2019). Comparison of machine learning algorithms for prediction house rent prices in Islamabad, Pakistan.

[9] Rehman, A., Hussain., & Tufail (2020). Analysis of machine learning techniques for house rent prediction. In 2020 IEEE 8th International Conference on Future Internet of Things and Cloud (iCloud) (pp. 226-231). IEEE.

[10] Adeyemi. & Eyehole (2021). House Rent Prediction Using Decision Tree Algorithm in Lagos, Nigeria. Journal of Physics: Conference Series, 1769(1), 012019.

[11] Dhiman, R. G. (2021). House Rent Prediction using Machine Learning Techniques.

[12] Miah, A. R. (2020). House rent prediction using machine learning techniques. In 2020 IEEE International Conference on Informatics, Electronics & Vision (ICIEV) (pp. 739-744). IEEE.

DOI: https://doi-ds.org/doilink/01.2024-53988668

[13] Anh, N. a. (2020). House Rent Prediction using Machine Learning Techniques in Vietnam. Proceedings of the International Conference on Advanced Computing and Intelligent Engineering.

[14] Singh, S. A. (2020). House Rent Prediction using Machine Learning Techniques in Mumbai, India. Proceedings of the International Conference on Computer Networks, Big Data and IoT.

[15] Vohra, P. (2020). House Rent Prediction using Linear Regression and Random Forest in Python.

[16] Bera, D. a. (2019). House Rent Prediction in Kolkata using Machine Learning Algorithms. Proceedings of the International Conference on Advanced Computing and Intelligent Engineering.

DOI: https://doi-ds.org/doilink/01.2024-53988668