Advances and Applications in Mathematical Sciences
Volume 21, Issue 7, May 2022, Pages 3805-3819

PROPERTY PRICE PREDICTION AND POSSIBILITY PREDICTION IN REAL ESTATE USING LINEAR REGRESSION

P. PRIYA[1], G. MAHALAKSHMI[2], K. MANOJKUMAR[3] and V. KUMARARAJA[4]

[1]Department of Computer Science and Engineering
M. Kumarasamy College of Engineering
Thalavapalayam, Karur – 639113, India
E-mail: priyap.cse@mkce.ac.in

[2,3]Assistant Professor
Department of Computer Science and Engineering
Government College of Engineering
Thanjavur, Tamilnadu, India
E-mail: gmahalakshmi@gces.edu.in
kmk_manoj@yahoo.com

[4]Assistant Professor
Department of Computer Science and Engineering
K. Ramakrishnan College of Engineering
Trichy, Tamilnadu, India
E-mail: mail2kumarcs@gmail.com

Abstract

The prediction model is more significant among the research areas. As there is a growing demand for forecasts, it has become a more popular and useful instrument. As a result, the majority of us rely on predictions the majority of the time. The majority of us do not consider what will occur after the prediction has been made. Prediction is the process of determining a rough outcome, but the anticipated outcome can also be saved for later use. The more we employ the projected value, the higher our result's future usability will be. People are growing increasingly interested in property values and their forecasts. Real estate value is unusual

2020 Mathematics Subject Classification: 34Bxx, 76-10, 80A30.

Keywords: Prediction model, forecasts, possibilities.
Received December 15, 2021; Accepted January 14, 2022

among them because it is a fixed asset. Real estate is concerned with the land that is most important to everyone. Land and house ownership will have a direct impact on the country's economy, both in terms of growth and decline. Home is where we all live, and thus it is a mandatory place for all people on the planet, and economic growth is anticipated using real estate as one of its predictive qualities. The child develops positive attributes as a result of where he lives and how his environment influences him. In order to predict its worth, the environment and the environment are also key factors to examine. However, most of us will just analyse the structure and its physical elements, ignoring the fact that the true values of price rises and falls will be determined by the additional surrounding environments that are located nearby. The idea's major purpose is to not only anticipate the price, but also to find some possibilities after the prediction is correct, so that the prediction is not the conclusion of the idea, but rather the start of a new one.

1. Introduction

Not only in our country, but around the world, the real estate market is a booming and appealing investment option. It is one of the ways to assess a country's and an individual's wealth. It's a major issue for all parties involved, including homebuyers, homeowners, brokers, real estate agents, and investors.

The price of real estate will continue to climb in a haphazard manner. As a result, the price is determined by certain qualities and parameters like as the property's square footage [m2. ft], the number of rooms and balconies, and so on. The location of the property is crucial in predicting its value. Additional functions, such as the availability of basic requirements and the valuation of people, can also influence the property's value. These characteristics fluctuate over time and are not fixed. As a result, if we have all of this data in the correct format with likely real values, the prediction will be more accurate. Data collection and cleaning are the most important aspects of forecasting. It will assist us in obtaining the true values of data and in converting irrelevant data into helpful or real data. As a result, price forecasting is influenced by a number of factors. All of these characteristics aid in the prediction of individual outputs, such as price. Figure 1 depicts the various adjacent features which influence the predictors for the property.

Figure 1. Features for Property prediction.

Several works are intertwined with our implementation. Figure 2 depicts the supply and demand relationship. However, the most important aspect of our execution is to make excellent use of the predicted outcomes. The outcomes are influenced by a number of variables. Other associated data can also influence the output. As a result, when the anticipated values are not accurate, the cost of a property or habitat can be forecasted using Machine Learning ideas in our application. They may also be able to change throughout time as a result of other factors including the environment and current events. To anticipate housing prices, a variety of techniques are utilised, including linear regression, random forest, gradient improvement and others.

Figure 2. Supply and Demand.

This research [2] employs a hybrid regression strategy to find house

prices by focusing on the novel feature in order to discover the optimal feature and its relationship with the selling price. To improve precision, feature engineering boosted the data's normality and linearity. They devised a strategy that integrated a number of algorithms. They showed that working with the Ames housing data was simple, and that using hybrid algorithms (64 percent and 36 percent, respectively) yielded better results in terms of house prices than using just one. A group of students presented this paper [5] by integrating linear and augmented algorithms with neural networks, so that the prediction accuracy might be improved. In this case, data with a variety of features is employed, which is crucial for forecasting. It was not cleaned at first, but it was cleaned well later for better performance. In this implementation, they used regression methods. The dataset was processed using neural network techniques, which will return all of the results. He'll examine the data and present the findings. As a result, using a neural network to improve regression improves accuracy. This article proposal [8] is about using an optimization strategy and regression analysis to improve the model's outcome and obtain a highly predictable value. NJOP pricing is projected using hedonic price and regression approaches based on NJOP construction price, land area, and land price. A stochastic optimization (PSO) technique is employed to select the effect variables. As a result, the RMSE 14.186 was found to be the minimal predicted error. This paper [11] was used to anticipate an unstable value, with certain machine learning algorithms doing better than others. Data purification, data analysis, and data evaluation are some of the intermediary phases. Finally, for easier understanding, the output result was saved as a CSV file. The Decision Tree was used to attain the highest precision of roughly 84 percent in this implementation, and a classification method was used to test the regression problem effectively. There are 80 features in an open source (Gaggle) data set. As a result, 37 factors that influenced the cost of homeownership are examined. The prediction of two regression methods is compared in this study proposed by Limsombunchai, Christopher Fan, and Minnie Lee [14] and implemented empirically. A considerable amount of household data was put to the test. Hedonic pricing models backed up the aforesaid findings. Even with an R2 value of more than 75%, the pricing model does not outperform the neural network. The price projection will vary due to anomalies between the attributes of the home and the price due to insufficient data for

surrounding features and fewer data. The model's shortcoming is that, due to a lack of genuine market data, the anticipated price is an estimate rather than the actual price of the house. The model's weakness is that it does not account for adjacent extra features when making predictions. To anticipate price, this article [17] proposes employing algorithms (ML) based on residential property and crime data. Property appreciation is a prevalent problem in today's market. Many elements are now playing a larger role in deciding the price of homes and real estate in the aftermath of the US crisis, which was triggered by stock losses in recent days. Agencies, finance, real estate specialists, and brokers have all played a role in price prediction, which has outperformed the traditional method of projecting sales. This one employs the various factors of the house's attributes to produce and present forecasts, as opposed to pricing, where the only variable sets the price. If the home has outstanding interiors and views, its worth is higher than normal; however, investors or buyers will have to make a decision based on other data, so that all of these factors are taken into account when predicting the price. To forecast the price represented in Figure 3, buyers must consider time and quality.

Figure 3. Time, quality Vs price.

The goal of this study [20] was to mimic ozone in the lower atmosphere. Environmental contaminants and weather variables were used to make the forecast. In this suggestion, principal component analysis was combined with ANN. It is demonstrated that combining multiple model predictions improves accuracy when compared to a single model prediction. The PCR variable selection technique was used to identify the subset of variables. The KNN was then used to fix the non-linearity. ANN is better at matching than PCR,

but when the two are combined, the result is much better. The reason behind this research [23] is to deal with the evaluation of the value of a house in Virginia. The value of OFHEO is used to determine housing offers. It also aids in the forecasting of future pricing. The accuracy of this model was increased by training on 5, 359 town dwellings. The property's value is determined by the crime rate in the area depicted in Figure 4. To compare the precision for improved performance, the model was constructed by utilising methods such as Naive Bayesian, Ada Boost, and RIPPER. As a result, the RIPPER algorithm performs better than the other algorithms in this scenario. Recent research has attempted to use algorithms like Neural Networks and Support Vector Machine, but in the above research, the algorithms' accuracy was compared to find the best algorithm.

Figure 4. Factors for the property price prediction.

The purpose of this study [26] is to depict the direction of stock movement and the stock price index for the Indian stock market. It works in two ways and is compared to three algorithm methods: For starters, commercial data can be analysed using technical parameters. The efficiency was enhanced when the parameters were represented as deterministic trend data, according to the findings. The focus of this model is on short-term prediction, however when real-time data, such as *B*. time-based data, is required, long-term prediction will be a challenge. As a result, the success of our plan is dependent on human behaviour in terms of spending and making judgments in numerous areas.

3. Existing System

Real-Estate Prediction was performed using various Machine learning algorithms like Regression Algorithms and some other commonly used classification Algorithms [27].

3.1. Random Forest Algorithm. Classification and regression issues can be solved by Random Forest Algorithm in Machine Learning. It's based on the ensemble learning principle, which combines several classifiers to provide solution for a complex issues and increase model performance. It averages a sequence of decision trees from distinct subsets of a data set to enhance the data set's prediction accuracy.

3.2. Stepwise Regression. In stepwise regression, the attribute that contributes the most to the result has a high priority; it ignores the attribute that did not contribute to the result, even if it is of high value and high availability; therefore, only important attributes for model prediction are used and ensured.

3.3. Linear Regression. It's an algorithm that depicts the x value based on multiple values assigned to y. The line of best fit is determined using the mathematical formula $Y = mx + c$, but the data points are dispersed. As a result, drawing the perfect line is impossible, but a form of best-fit line is achievable. Simple linear regression and multiple linear regression are the two forms of linear regression. A value with only one independent variable is predicted by Simple Linear, whereas a value with one or more independent variables is predicted by Multiple Linear. The data processing and analysis steps are the most important implementation process. The analysis steps are shown in Figure 5.

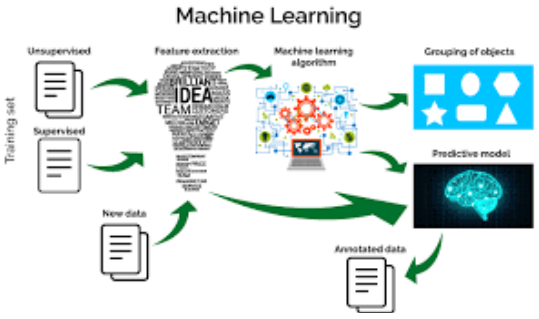

Figure 5. Analysing and processing of data.

4. Linear Regression for property and possibility prediction. House price prediction can assist a builder in determining the selling price of a home in a specific location, as well as assist a client in determining the best time to buy a home. House prices are predicted with a new set of parameters using a different technique. We also forecast settlement compensation. Mathematical relationships help us understand many aspects of daily life. When such relationships are expressed with accurate numbers, we gain additional clarity of variables. A linear regression based on hedonic prices is used as the method. Gharehchopogh's previous research, which used the linear regression method, discovered 0.929 inaccuracies with the real pricing. The least squares coefficient determination is commonly used in linear regression, although finding the optimum formula takes a long time.

4.1. Algorithm. Used After using several algorithms, thus at last we got the high accuracy in one of the algorithm used. Thus the algorithm with highest accuracy is mentioned below,

- Linear Regression

4.2. Linear Regression. It's a type of supervised learning in which we know both the X and Y forms and can get y- >output for the X given as data collecting input. In linear regression, the mathematical formula $Y = MX + c$ is used to identify the best fit line that runs across the data points. However, because the data points are distributed, a perfect line can't be drawn; only the best-fit line can be drawn. Simple Linear predicts a value with only one independent variable and Multiple Linear predicts a value with one or more independent variables.

4.3. Process Flow. The process implementation starts from data gathering and ends with prediction. Thus, we planned to use those predicted value for further insight prediction and the flow is depicted in figure 6.

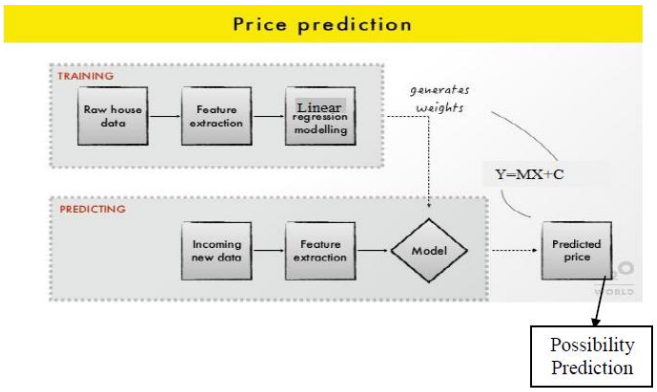

Figure 6. Implementation Process Flow.

4.4. Dataset Used. The size of the property, the number of bathrooms, the number of bedrooms, the balcony, and the type of location are not factors in our projects. Instead, we employed a large amount of data that is critical for the price prediction's contribution. We used approximately 10,000 rows of data with approximately 8 columns of price prediction features. Because the first data we acquire may or may not be erroneous, the first and most crucial step is to clean up the data so that we can better comprehend it. When data is adequately cleansed, it helps us anticipate the value more precisely. The data properties are depicted in Figure 8 below. Data cleansing is one of the more time-consuming steps, yet it is necessary for the process to progress. Keep yourself safe and productive. Each trait has a priority and a role to play in the forecast. The function with the highest priority has a bigger impact on the output, whereas the function with the lowest priority has a smaller impact on the predicted result. The prediction procedure begins after the data has been cleaned up. The feature's meanings are based on the proposal we implemented, as illustrated in Figure 7. This is time series data, which means that it varies over time. As a result, this data is frequently controlled in order to make the process more efficient.

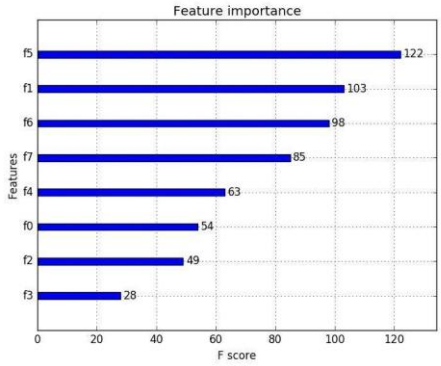

Figure 7. Feature importance.

4. 5. Dataset Attributes.

	area_type	availability	location	size	society	total_sqft	bath	balcony	price
0	Super built-up Area	19-Dec	Electronic City Phase II	2 BHK	Coomee	1056	2.0	1.0	39.07
1	Plot Area	Ready To Move	Chikka Tirupathi	4 Bedroom	Theanmp	2600	5.0	3.0	120.00
2	Built-up Area	Ready To Move	Uttarahalli	3 BHK	NaN	1440	2.0	3.0	62.00
3	Super built-up Area	Ready To Move	Lingadheeranahalli	3 BHK	Soiewre	1521	3.0	1.0	95.00
4	Super built-up Area	Ready To Move	Kothanur	2 BHK	NaN	1200	2.0	1.0	51.00

Figure 8. Dataset Attributes.

5. Experimental Result and Design Analysis

5.1. System Analysis and Design.

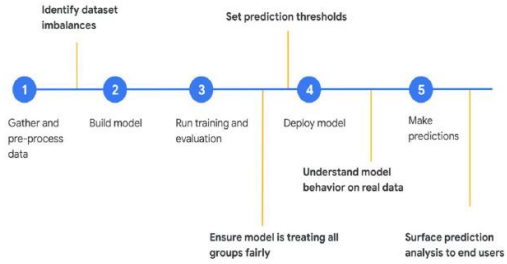

Figure 9. System Analysis and Design.

5.3. Algorithm's Accuracy. Thus the accuracy of the algorithm shows that how correctly our implementation predicts the outcome of the data. Thus these predicted values are compared with predefined outcome values and produce the algorithm's accuracy as shown in figure 10.

```
from sklearn.linear_model import LinearRegression
obj = LinearRegression()
obj.fit(xtrain,ytrain)
obj.score(xtest,ytest)
```

```
0.9100072236121388
```

```
y_predd = obj.predict(xtest)
df2 = pd.DataFrame({'Actual': ytest, 'Predicted': y_predd})
df3 = df2.head(10)
df3
```

	Actual	Predicted
3447	73.85	81.505634
629	90.75	93.969875
1807	32.49	34.004864
4805	145.00	111.563622
2497	38.00	42.047739
1785	41.00	43.440139
4953	74.93	99.867304
6288	50.00	46.178020
6103	52.00	75.088943
3664	95.00	101.583934

```
from sklearn import metrics

print('Mean Absolute Error: {:.2f}'.format(metrics.mean_absolute_error(ytest, y_predd)))
# print('Mean Squared Error:{:.2f}'.format(metrics.mean_squared_error(ytest, y_predd)))
# print('Root Mean Squared Error:{:.2f}'.format(np.sqrt(metrics.mean_squared_error(ytest, y_predd))))
print('Variance score is: {:.2f}'.format(metrics.explained_variance_score(ytest,y_predd)))
```

```
Mean Absolute Error: 10.67
Mean Squared Error:355.15
Root Mean Squared Error:18.85
Variance score is: 0.91
```

Figure 10. Algorithm Accuracy.

5.4. Crime Possibility. The crime report possibility is generated by analysing the past 10 years' crime data. These data are manipulated and produce a graph to understand the crime rate for a particular area as shown in figure 11.

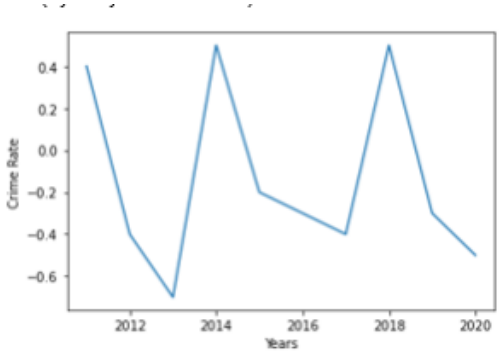

Figure 11. Crime Report Chart.

5.5. Price Prediction. The price prediction module in our implementation will predict the price for the user inputs. These inputs are taken and converted into an acceptable format and then it will be given to the Regression model for prediction process. These process are shown in figure 12.

```
def predict_price(location,sqft,bath,size):
    loc_index = np.where(x1.columns==location)[0][0]
    X = np.zeros(len(x1.columns))
    X[1] = sqft
    X[2] = bath
    X[0] = size
    if loc_index >= 0:
        X[loc_index] = 1

    return obj.predict([X])[0]
```

```
fi_price=predict_price('1st Phase JP Nagar',1000, 2, 3)
print(fi_price)
```

```
92.09822845458984
```

Figure 12. Price Prediction.

6. Conclusion

In summary, the price of the property and homeownership depends on

the different factors around the property. The idea of using the efficient algorithm is to create a useful model which drops the implementation to the higher end. Hence, the final predictions are used not only to predict the price but also to find the possibilities which helps the buyers to take decision in buying a property [27]. In the future, implementation will not be based on the accuracy of the prediction, but on how we will make better use of those predictions. The big challenge is to collect additional information about the crime rate around the property, so that the task of predicting the value of a property can be made accurate.

References

[1] Samuel Hameau, Tri KaviniawanWijaya, MatteoVasirani and Karl Aberer, Electricity load forecasting for Residential customer: Exploiting Aggregation and Correlation between Households profiles, Sustainable Internet and ICT for Sustainability IEEE, (2013).

[2] Sifei Lu, Zengxiang Li, Zheng Qin, Xulei Yang, Rick SiowMongGoh A Hybrid Regression Technique for House Price Prediction, (2017).

[3] Byeonghwa Park and Jae Kwon Bae, Using machine learning algorithms for housing Price prediction: The Case of Fair fax country Virginia housing data, Expert Systems with Applications 42(6) (2015), 2928-2934.

[4] Jooyong Shim, Okmyun Bin and Changha Hwang, Semi-parametric partial effects kernel minimum squared error model for predicting housing sales price, Neurocomputing 124 (2014), 81-88.

[5] Ayush Varma, Abhijit Sharma, Sagar Doshi and Rohini Nair, House Price Prediction Using Machine Learning and Neural Networks, INSPEC number (2018), 18116205.

[6] Rushab Sawant, Saurabh Jain, Tushar Tiwari, Yashwant Jangid and Anita Gupta, A multi feature-based housing price prediction for Indian market using Machine Learning, International Journal of Computer and Mathematical Science, (2017).

[7] Azme Bin, Khamis and Nur Khalidah Khalilah Binti Kamarudin Comparative Study on estimate house price using statistical and neural network model, International Journal of Science and Technology Research 3 (2014).

[8] Adyan Nur Alfiyatin, Hilman Taufiq, Ruth Ema Febrita and Wayan Firdaus Mahmudy, Modeling house price prediction using regression analysis and particle swarm optimization, (IJACSA) International Journal of Advanced Computer Science and Applications 8(10) (2017).

[9] T. M. Oshiro, P. S. Perez and J. A. Baranauskas, How many trees in a random forest, In Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), vol. 7376 LNAI, (2012), pp. 154-168, ISBN: 9783642315367. DOI: 10. 1007 / 978 - 3 - 642 -31537-4\13.

[10] S. C. Bourassa, E. Cantoni, and M. Hoesli Predicting house prices with spatial dependence: a comparison of alternative methods," Journal of Real Estate Research (32)2 (2010), 139-160.

[11] NeelamShinde, KiranGawande, Valuation of House Price Using Predictive Techniques, International Journal of Advances in Electronics and Computer Science, ISSN: 2393-2835(IJAECS) 5(6) (2018).

[12] Rohan Bafna, Anirudh Dhole, Ankit Jagtap, Asif Kazi, Arbaz Kazi, Prediction of Residential Property Prices – A State of the Art, International Advanced Research Journal in Science, Engineering and Technology 5(3) March 2018.

[13] Visit Limsombunchai, Christopher Gan and Minsoo Lee, House Price Prediction: Hedonic Price Model vs. Artificial Neural Network, American Journal of Applied Sciences ISSN 1546-9239 1(3) (2004), 193-201.

[14] J. M. Davis, B. K. Eder and P. Bloomfield, Modeling ozone in the Chicago urban area, In: Nychka, D., Piegorsch, W.W., Cox, L.H. (Eds.), Lecture Notes in Statistics, Case Studies in Environmental Statistics, Springer, New York, 132 (1998), 5e26.

[15] E. Hromada, Mapping of Real Estate Prices Using Data Mining Techniques, Procedia Engineering (2015), 233-240.

[16] Y. Bin and L. Weijun, Research on Prediction Methods of Residential Real Estate Price Based on Improved BPNN, International Conference on Smart Grid and Electrical Automation (ICSGEA), Zhangjiajie, (2016), 150-156, doi:10.1109/ICSGEA.2016.45.

[17] Muhammad A. Razi and Kuriakose Athappilly, A comparative predictive analysis of neural networks (NNs), nonlinear regression and classification and regression tree (CART) models, Expert Systems with Applications 29 (2005), 6574. DOI=http://dx.doi.org/10.1016/j.eswa.2005.01.006.

[18] S. M. Al-Alawi, S. A. Abdul Wahab and C. S. Bakheit, Combining principal component regression and artificial neural networks for more accurate predictions of ground-level ozone, Environmental Modelling and Software 23(4) (2008), 396-403. Doi:10.1016/j.envsoft.2006.08.007.

[19] L. Li and K. H. Chu, Prediction of real estate price variation based on economic parameters, International Conference on Applied System Innovation (ICASI), Sapporo, (2017), 8790, Doi:10.1109/ICASI.2017.7988353.

[20] Z. X. li, Using Fuzzy Neural Network in Real Estate Prices Prediction, Chinese Control Conference, Hunan, (2007), 399-402, Doi:10.1109/CHICC.2006.4347291.

[21] Park Byeonghwa and Jae Kwon Bae, Using machine learning algorithms for housing price prediction: The case of Fairfax County, Virginia housing data, Expert Systems with Applications 42(6) (2015), 2928-2934.

[22] W. T. Lim, L. Wang, Y. Wang and Q. Chang Housing price prediction using neural networks, 12th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery, Changsha (2016), 518-522, Doi:10.1109/FSKD.2016.7603227.

[23] Y. Ganjisaffar, R. Caruana and C. V. Lopes, Bagging gradient-boosted trees for high precision, low variance ranking models, Proceedings of the 34th international ACM

SIGIR conference on Research and development in Information Retrieval, ACM, (2011), 85-94.

[24] J. Patel, S. Shah, P. Thakkar and K. Kotecha, Predicting stock and stock price index movement using trend deterministic data preparation and machine learning techniques, Expert Systems with Applications 42(1) (2015), 259-268.

[25] Yang Yonghui, Research on House Price Prediction Based on Multi-Dimensional Data Fusion [J], International Journal of Advanced Network, Monitoring and Controls 05(01) 2020 Xi'an Technological University Xi'an, 710021, China.

[26] P. Priya, M. Arul Kumaran, K. Dhinesh Kumar, S. Nivas Singha and K. Rajkumar, Prediction of Property Price and Possibility Prediction Using Machine Learning, Annals of R. S. C. B. 25(4) (2021), 3870-3882.

[27] Nair, How To Use XGBoost To Predict Housing Prices In Bengaluru: A Practical Guide, 2019. [Online] Available:https://www.analyticsindiamag.com/how-to-use-xgboost-to-predicthousing-prices-in-bengaluru-a-practical-guide/ [Accessed: 25-Mar-2019].