PAPER • **OPEN ACCESS**

Empirical analysis of regression techniques by house price and salary prediction

To cite this article: U Bansal *et al* 2021 *IOP Conf. Ser.: Mater. Sci. Eng.* **1022** 012110

Empirical analysis of regression techniques by house price and salary prediction

U Bansal[1], A Narang[2], A Sachdeva[3], I Kashyap[4] and S P Panda[5]

[1]Department of Computer Science, Manav Rachna International Institute of Research and Studies, Faridabad, India.
[2]Department of Computer Science, Manav Rachna International Institute of Research and Studies, Faridabad, India
[3]Department of Computer Science, Manav Rachna International Institute of Research and Studies, Faridabad, India
[4]Department of Computer Science, Manav Rachna International Institute of Research and Studies, Faridabad, India
[5]Department of Computer Science, Manav Rachna International Institute of Research and Studies, Faridabad, India

Email: utkarshbansal0203@gmail.com

Abstract. Regression analysis is extensively used for prediction and prognostication, and its use has substantial overlap with the domain of machine learning. The main objective of this paper is to compare the performance of two regression techniques namely Simple Linear Regression (SLR) and Multiple Linear Regression (MLR) algorithms by two cases: predicting the salary of employees after certain years and predicting the prices of real estates. An employee's salary depends on numerous factors, such as total employee experience, certifications, and overall experience as a lead and manager. The factors in predicting house prices are the area of land (sqft_living), condition, waterfront, number of bedrooms, and so on. The dataset used in this experiment is an open-source dataset from KaggleInc. The algorithms were compared using parameters like R-squared value, Mean absolute error (MAE), Mean Squared Error (MSE), Median Absolute Error (MDAE), Variance Score, and Root Mean Square Error (RMSE). Results have shown that MLR provides the better efficiency in comparison to SLR.

1. Introduction

A prediction is an expectation about a future event. As the Internet is growing day by day and data is produced at a prodigious rate, for example, Twitter produces the data at a volume of 12 Terabytes(TB) per day, Facebook produces the data at a volume of 4 Petabytes(PB) per day in recent years. So it is essential to assemble, inspect, and model this huge data to predict future events in numerous fields [1].

As future events are not known to anyone, so finding accurate data about future events is impossible sometimes. However, to improve the accuracy of results, a suitable algorithm and model can be selected. In this paper, two algorithms of regression namely SLR and MLR for prediction are

analyzed. The main aim is to compare these two algorithms using two datasets: the salary of employees and house prices.

As it is known accurate hiring of employees is important for any company's growth. It would be an intimidating task for the company to recruit and keep top talents. The Salary of an employee is predicted by taking into account the following factor(s) - certification, the total experience of the employee, total experience as a lead.

On the other hand, calculating the price of a property is highly important for real estate, the stock market, the tax sector, the economy, etc. The researchers show that house prices are mainly dependent upon the size of the house [2]. The distinct parameters such as the number of bedrooms, living area, condition of the house, view, and many other factors have also been considered.

Machine Learning plays a vital role to predict the value in such cases. It offers the power to machines to predict and perform clustering based on past experiences. Its accuracy is dependent upon the dataset pattern, the parameter tuning, and the feature selections. Regression techniques that come under Supervised Machine Learning Algorithms have been used in this research. Supervised algorithms take labeled data as input for prediction.

Regression is a Supervised Machine Learning technique that predicts numeric or continuous value(s) statistically. SLR is a type of regression modeling technique, which is used to analyze the relationship between only two variables (one is dependent and the other is independent) whereas, MLR Regression is a linear relationship between a dependent variable and various independent variables.

The following sections are arranged as follows: Section 2 comprises related studies, Section 3 contains research methods, Section 4 describes experimentation, Section 5 involves outcomes and review, and the concluding remarks are accomplished by section 6.

2. Related Research

Rong S and Zhang Bao-wen [1] have performed empirical analysis using the Linear Regression technique on the sale of iced products of a company and the effect of temperature variation on the sale. They have concluded that the variation in temperature of iced products is of great commercial value.

Khamis A B et al.[3] have compared the performance of MLR and Artificial Neural Network(ANN) models using R-squared and MSE to estimate the house prices. The factors they have considered include the number of bathrooms, number of bedrooms, living area, lot size, and year of built. The value of R-squared and MSE were compared to select a preferred model. By using ANN, the R-squared value was higher than the MLR. It was concluded that the ANN model is preferred over the MLR model.

Navyashree et al. [4] accomplished a research study to establish the salary prediction in the IT job sector using different data mining techniques like Random Forest Regression, Decision Trees Regression, and Support Vector Machine (SVM) implemented through different python libraries like Matplotlib, NumPy, pandas in the Jupyter notebook Anaconda Navigator. In the current scenario, the salary of employees has become one of the major fields which enforced the researchers to determine the way to predict the salary. They have concluded that Random Forest Regression gives a more accurate result as compared to SVM and Decision Tree Regression.

Khongchai P and Songmuang [5] illustrated the interface of salary prediction which contains several attributes like gender, job training, certification, and GPA which the system compared and displayed the predicted salary of 3 graduated people. They have collated the different data mining techniques in which the highest accuracy was predicted for K-nearest neighbor and lowest for Multilayer perceptron.

In the last two decades, it is forecasted that the demand for property has increased which challenges the researchers to predict all the minute factors that affect the price evaluation of real estates and make a predictive model by considering all these features. Shinde N and Gawande K [2] performed different predictive techniques – SVM, Lasso Regression, Logistic Regression, and Decision Trees. By comparing them they concluded that the decision trees give the best results with high accuracy.

In [6], the authors have mainly focused on feature selection that plays a vital role in machine learning prediction. They have correlated the features and observed the data distribution between the cost price and selling price of the property. The best accuracy is provided by the decision tree according to them.

3. Research Methodology

The basic process to identify and analyze information here are six-folds:

a) Data Collection: Dataset is a collection of dependent and independent attributes or factors. As this paper is based on supervised machine learning algorithms so these attributes are called features [6].
b) Splitting dataset: The dataset is split into a training set and test set. The training set is used for training the model whereas the test set is used for testing or prediction of result. Usually, the training set and test set are split into an 80-20 ratio, respectively because the greater the data in the training phase, the better will be the results.
c) Model Training: Generally, the algorithms are selected according to the data but in this research, both algorithms are used to compare their predictions. The selected model is trained on the training set to discover the predicted results optimally.
d) Predicting Results: The trained model is then applied to the test set of the respective dataset, which gives the predicted results.
e) Visualizing dataset: Various graph types can be used to plot the predicted results. The libraries called Matplotlib.pyplot and graphlab in python is used to generate various types of graphs such as scatter plot, box and whisker plot, etc.
f) Model Comparison: A simplistic approach that is applied to compare the model results is accuracy determination. The various techniques are considered for accuracy determination which is discussed later in this paper.

3.1. SIMPLE LINEAR REGRESSION

Regression is a statistic based path to find affiliation between two or more variables. The SLR Regression model is predictive in nature which investigates the dependent variable(y) and the independent variable(x). It establishes a linear relationship between both variables. It is an example of parametric learning [7].

The two main factors that must be considered while applying the algorithms to the dataset are:

a) The choice of the independent attribute should be made thoughtfully to predict an outcome accurately. There can be multiple attributes influencing the dependent variable. But in SLR, only one independent variable is possible.
b) This model attempts to explain the relationship between two variables using a straight line graph.

The Equation of the Simple Linear Regression is represented as:

$$y = \beta_\circ + \beta_1 x + \varepsilon \quad (1)$$

Where y is the dependent variable, x is the independent variable, $\beta_\circ$ is constant or y-intercept of the regression line, β_1 is the slope of the regression line and ε is the random error which is the difference between the predicted value and actual value.

3.2. MULTIPLE LINEAR REGRESSION

MLR is a statistical analysis that determines the explanatory variables which include one dependent and more than one independent variables i.e. (x_1, x_2, ……, x_n) where (n>1). This model is structured as:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots\ldots\ldots \beta_p x_{ip} + \varepsilon \quad (2)$$

y_i is a predicted variable, β_0 is the y-intercept, β_1 and β_2 are regression coefficients, β_p is the slope for each independent variable and ε is the random error in the model.

The assumptions of this model are:

a) The relationship between the dependent variable(y) and the p-vector regressor is linear.
b) When the explanatory variables (independent variables) are highly correlated to each other, then the data shows the multi-collinearity.
c) The scenario homoscedasticity, which means that the amount of errors in the residual is similar at each point of the linear model. To test this assumption, the data should be plotted on a scatterplot.
d) When the residuals are normally distributed, the multivariate normality occurs.

For prediction, we need to find the regression line for both the algorithms.

To discover the regression line $\beta_\circ$ and β_i has to be calculated by applying the following formula:

$$\beta_i = \frac{\sum(x-\bar{x})(y-\bar{y})}{\sum(x-\bar{x})^2} \quad (3)$$

Now, calculate $\beta_\circ$ using the following formula:

$$\beta_\circ = \bar{y} - \beta_i \bar{x} \quad (4)$$

Once the regression line has been fitted then the performance of the models can be estimated.

Linear Regression generally uses the root mean square to calculate errors in the model. Errors are useful in estimating quality of the coefficient and substantiate the model. The standard error of estimation (S_{est}) gives the estimated distance between actual values and the regression line i.e.

$$S_{est} = \sqrt{\frac{\sum(\hat{y}-y)^2}{n-2}} \quad (5)$$

Where $\hat{y}$ represents the estimated value, y is the actual value and n indicates the number of observations.

The scattering of data values around the fitted regression line is evaluated by R-squared(R^2), which is also known as the coefficient of determination. R^2 is calculated by taking the ratio of the distance between actual and mean values by the distance between estimated and mean values.

$$R^2 = \frac{\sum(\hat{y}_i-\bar{y})^2}{\sum(y_i-\bar{y})^2} \quad (6)$$

Where y_i are the experimental values, $\bar{y}$ is the mean and $\hat{y}_i$ is the fitted value.

R^2 measures the goodness of fit for linear regression models, the range of the coefficient of determination lies between 0 and 1. The more closely the value is to 1, the more accurate is the model, as the value tends to 0 the model's performance decreases.

It is said by statisticians that a regression model fits the data well only when the difference between predicted values and actual values are small and unbiased, where unbiased means the fitted values should be systematically too low or too high.

4. Experimentation

This prediction on both the datasets is executed through the following approaches:

a) Pandas: Pandas is an open-source library written for python programming language which is used for data analysis. This library is used when dealing with tabular data and time-series data. It helps to refine, explore, and process the data.
b) Matplotlib.pyplot: Matplotlib.pyplot is a wide-ranged library for creating animated, static, and interactive visualization which can be used in python, python script, and various graphical user interface toolkit.
c) NumPy: NumPy is a package used for fast scientific computation [8]. It is a fundamental library which provides multidimensional array object and other types of operation.
d) Graphlab: It is a python library for providing big scale, high-performance data objects. Its main features are to analyze huge scale data at a prodigious rate and provide production monitoring.
e) read_csv: It is a python function defined under the Pandas package to read data from comma-separated values (CSV) files. Specifically, it is used to import CSV formatted datasets.
f) SFrame: It is an out of core data frame which allows working with a dataset that is larger than the amount of volatile memory in the systems. It is declared inside the graph lab library and it is open source under the BSD license.
g) LinearRegression: It is a pre-defined function under sklearn library which is used to implement SLR and MLR and make predictions accordingly.

The experiment is performed like:

Firstly, a dataset is created in an excel file or downloaded from web sources (KaggleInc), then it is further functioned in Jupyter notebook from the Anaconda Navigator platform.

Now, libraries such as NumPy, Matplotlib.pyplot, Pandas, and graphlab are imported. The Employee Salary and House Pricing datasets are then read using pandas library through the read_csv (dataset name) function and graphlab library through SFrame (dataset name) function respectively.

Through the data points in both the dataset, graphs are plotted using Matplotlib.pyplot and graphlab libraries. Now if data points are collinear then the LinearRegression function is used otherwise other algorithmic functions are used.

The dataset is now split into two sets – training set and test set in the 80-20 ratio respectively.

Then the models are trained by applying Simple Linear and Multiple Linear Regression algorithms on the training set. The house price and salary of an employee are set as target variables for the House Pricing and Employee Salary dataset respectively. After this, the models are ready to predict the results, and the two models are compared in the following section using these parameters.

5. Result and discussion

In the following section, the results of two algorithms implemented on both the datasets are discussed.

Table 1. Dataset of employee salary.

	Total	lead	manager	certification	salary
0	1.0	0	0	0	20000
1	1.5	0	0	0	23000
2	2.0	0	0	0	25000
3	2.0	0	0	1	30000
4	2.5	0	0	0	27000
5	3.0	0	0	2	30000
6	3.5	1	0	0	33000
7	3.5	1	0	1	35000
8	3.5	1	0	2	40000
9	4.0	1	0	0	35000
10	4.0	2	0	0	40000
11	4.0	2	0	2	43000

In house prices, data includes house ID, date of purchase, price, bedrooms, bathrooms, square feet of living, condition, zip code, view, and 11 other columns.

In employee salary prediction, data includes the total experience of the employee, total experience as a lead and manager, total certifications, and salary class.

The paper will use python 3.6 and python 2.0 versions to set up a linear regression analysis model targeted on two different datasets. The models are compared using the tool – Jupyter notebook.

Table 2. Dataset of House prices

id	price	bedrooms	bathrooms	sqft_living	sqft_lot	floors
7129300520	221900.0	3.0	1.0	1180.0	5650.0	1.0
6414100192	538000.0	3.0	2.25	2570.0	7242.0	2.0
5631500400	180000.0	2.0	1.0	770.0	10000.0	1.0
2487200875	604000.0	4.0	3.0	1960.0	5000.0	1.0
1954400510	510000.0	3.0	2.0	1680.0	8080.0	1.0
7237550310	1225000.0	4.0	4.5	5420.0	101930.0	1.0
1321400060	257500.0	3.0	2.25	1715.0	6819.0	2.0
2008000270	291850.0	3.0	1.5	1060.0	9711.0	1.0
2414600126	229500.0	3.0	1.0	1780.0	7470.0	1.0
3793500160	323000.0	3.0	2.5	1890.0	6560.0	2.0

sqft_above	sqft_basement	yr_built	yr_renovated	lat	long
1180.0	0.0	1955.0	0.0	47.51123398	-122.25677536
2170.0	400.0	1951.0	1991.0	47.72102274	-122.3188624
770.0	0.0	1933.0	0.0	47.73792661	-122.23319601
1050.0	910.0	1965.0	0.0	47.52082	-122.39318505
1680.0	0.0	1987.0	0.0	47.61681228	-122.04490059
3890.0	1530.0	2001.0	0.0	47.65611835	-122.00528655
1715.0	0.0	1995.0	0.0	47.30972002	-122.32704857
1060.0	0.0	1963.0	0.0	47.40949984	-122.31457273
1050.0	730.0	1960.0	0.0	47.51229381	-122.33659507
1890.0	0.0	2003.0	0.0	47.36840673	-122.0308176

On code execution, plotted results for prediction are achieved. These plots help us to figure out the correlation between the target variables (price and salary) and the predictor variable [9]. These plots also help to predict the salary and house prices automatically.

Different performance metrics such as R-squared value, Root Mean Squared Value (RMSE), Mean Absolute Value (MAE), and Mean Squared Value (MSE), Median absolute error (MDAE), and Variance score. The prediction results of models in table 1 and table 2 are evaluated using these metrics.

5.1. Results on employee salary dataset

5.1.1.The results of MLR applied to the Salary Prediction dataset

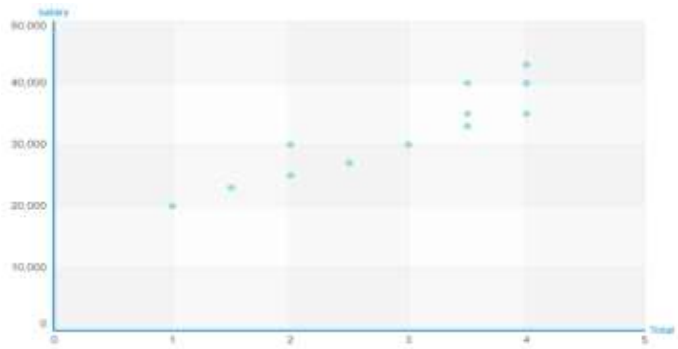

Figure 1. Salary of the employees vs. total experience of the employees.

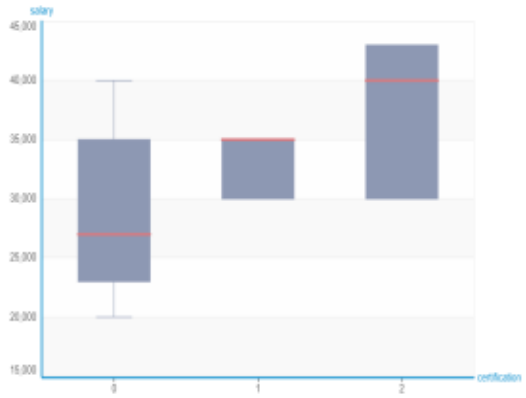

Figure 2. Salary of the employees vs. Certifications of the employees.

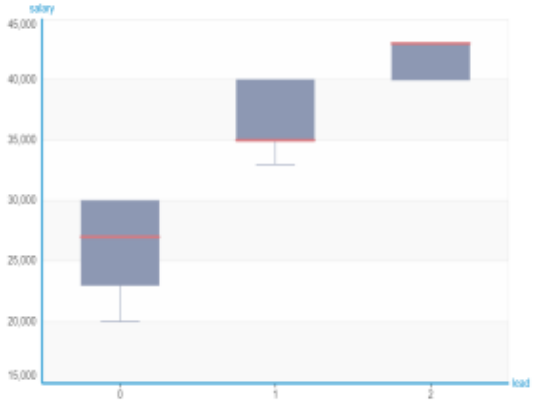

Figure 3. Salary of the employees vs. Experience of the employees as a lead.

In Figure 1 the blue points represent the interrelationship between the salary and the total experience of the employee.

In Figures 2 and 3, the box and whisker plot represents the range of salaries corresponding to the number of certifications and total experience as a lead respectively and the red line represents the median value of salary corresponding to certifications and experience of employee as a lead respectively.

It was found that the correlation coefficient indicated that the salary of an employee is positively and strongly associated with the total experience, certification and total experience as a lead while zero significance to total experience as a manager because its value is constant throughout the dataset.

The MLR equation for salary prediction can be written as

*Salary of Employee=19455.9+total*2389.7+lead*5775.73+manager*0+certification*2672.8* (7)

5.1.2. The results of SLR applied to the Salary Prediction dataset

Figure 4. Salary of the employees vs. Total Experience of the employees (Training Set).

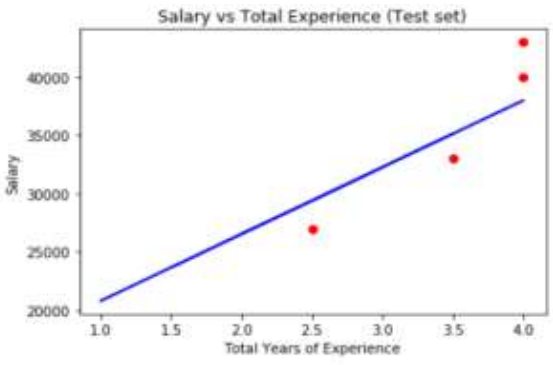

Figure 5. Salary of the employees vs. Total Experience of the employees (Test Set).

Figure 4 and 5 represents Salary in correspondence with Total Years of experience. These reflect that the Total Years of experience has a strong relationship with Salary because its p-value is higher than other factors justifying the selection as the independent variable in the SLR model.

The SLR equation for salary prediction can be written as

*Salary of Employee =15057.034+total*5733.84* (8)

Table 3. Error values for Employee Salary Prediction

ERRORS	SIMPLE LINEAR	MULTIPLE LINEAR
MAE	2883.08	1410.85
MSE	9836035.65	2904523.28
MDAE	2258.56	1082.72
VARIANCE	0.76	0.94
R-SQUARE	0.75	0.92
RMSE	3136.24	1704.26

From the above table, it is found that MLR gives higher accuracy and low errors as compared to SLR.

5.2. Results of house prices prediction

5.2.1. The results of MLR applied to the House Prediction dataset

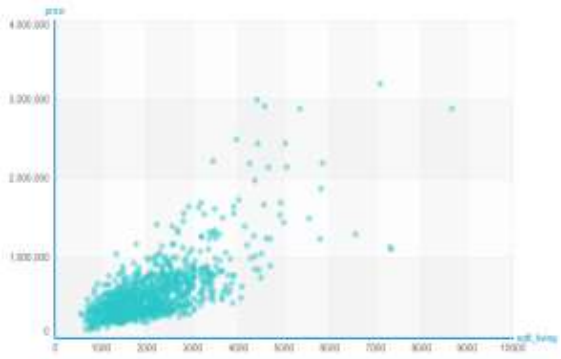

Figure 6. Price of the houses vs sqft_living of the houses.

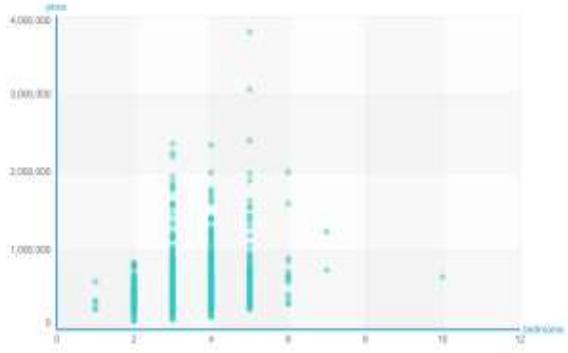

Figure 7. Price of the houses vs. No. of Bedrooms in the houses.

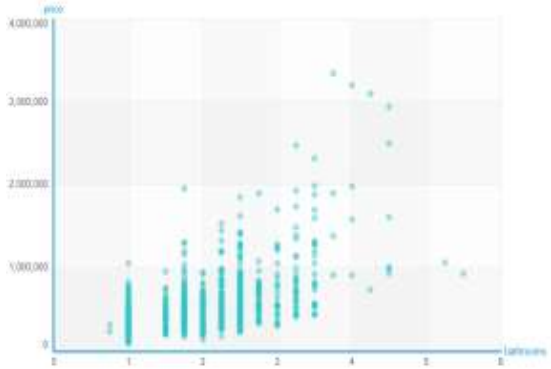

Figure 8. Price of the houses vs. No. of Bathrooms in the houses.

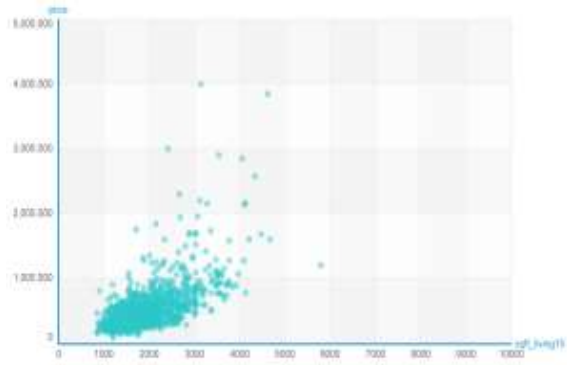

Figure 9. Price of the houses vs sqft_living15 of the houses.

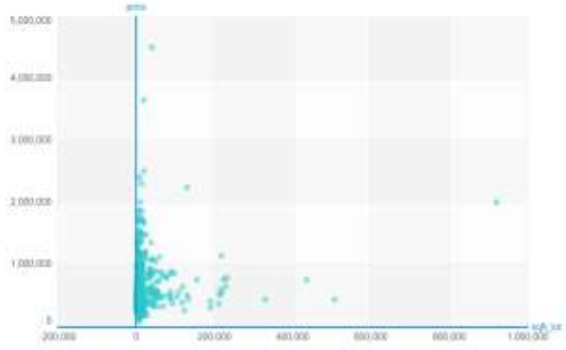

Figure 10. Price of the houses vs. sqft_lot of the houses.

Figure 11. Price of the houses vs. sqft_above of the houses.

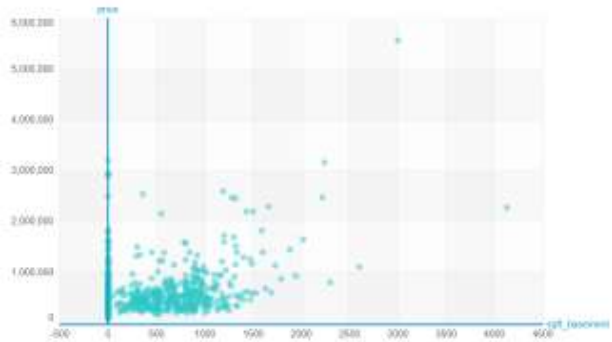

Figure 12. Price of the houses vs. sqft_basement of the houses.

In Figure 8 to 12 the scatter plot represents the relationship between factors such as sqft_living, number of bedrooms, number of bathrooms, sqft_living15, sqft_above, sqft_basement, and sqft_lot.

The MLR equation for House Prediction can be written as

House Price= -48730906.3336 -33531.9294822 bedrooms + 998.780636267* bathrooms + 87.3738340124* sqft_living + 0.219198360181* sqft_lot -23040.3869423* floors + 56510.2831637* condition + 85374.3910196* grade + 110.119200776* sqft_above + 87.6049760458* sqft_basement*

+ 64.5062911433 yr_renovated + 23.6666053538* sqft_living15 -0.542683183718* sqft_lot15 + 54529.5607631* view + 616595.268622* waterfront + 602758.590596* lat -159000.257566* long* (9)

5.2.2. The results of SLR applied to the House Prediction dataset

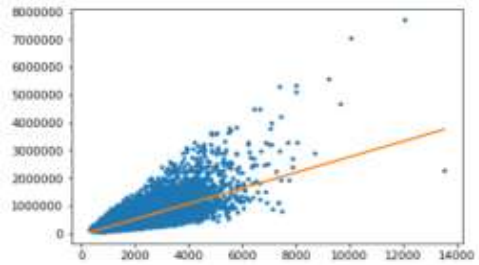

Figure 13. Price of the houses vs. sqft_living of the houses(Training Set)

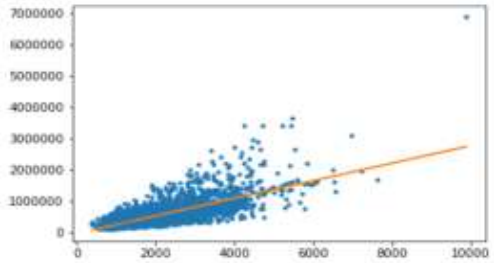

Figure 14. Price of the houses vs. sqft_living of the houses(Test Set).

Figures 13 and 14 show the scatter plot between Price and sqft_living for Training and Test Set respectively. The area of the real estate has the most major impact on the price of the house, hence it has been selected as the independent variable among various other factors in the SLR model.

The SLR equation for House Prediction can be written as

*House Price =280.919461011*sqft_living -44959.9285106* *(10)*

Table 4. Error values for House Price Prediction.

ERRORS	SIMPLE LINEAR	MULTIPLE LINEAR
MAE	171751.57	130879.73
MSE	65111435997	42017575314
MDAE	129638.55	93933.2
VARIANCE	0.49	0.67
R-SQUARE	0.49	0.67
RMSE	255169.4261	204981.8902

From the above table, it can be concluded that MLR gives higher accuracy and low errors as compared to SLR.

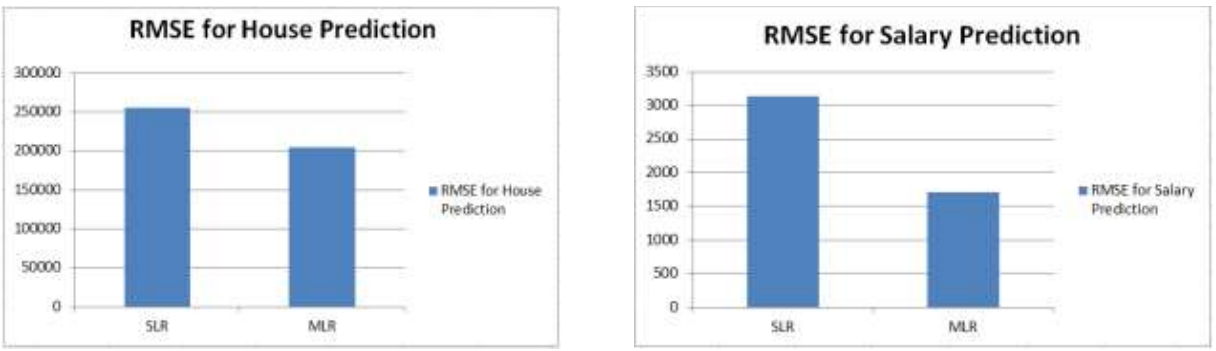

Figure 15. RMSE for Salary and House Prediction.

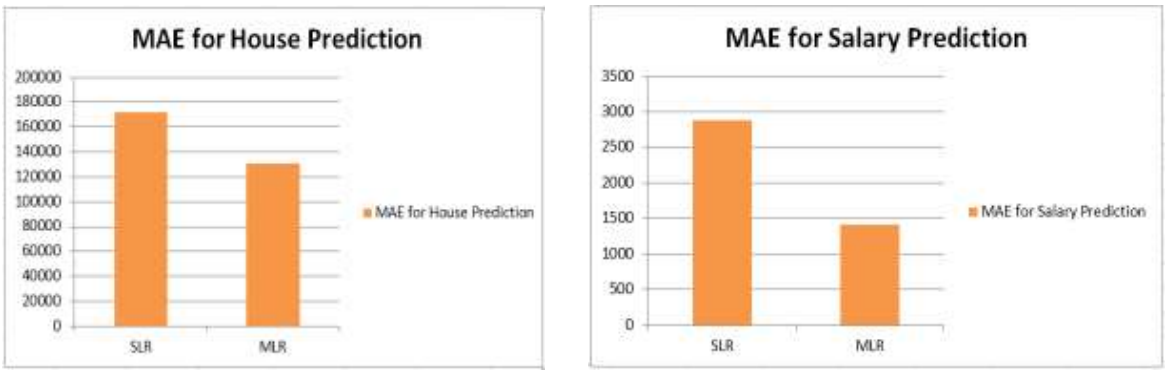

Figure 16. MAE for Salary and House Prediction.

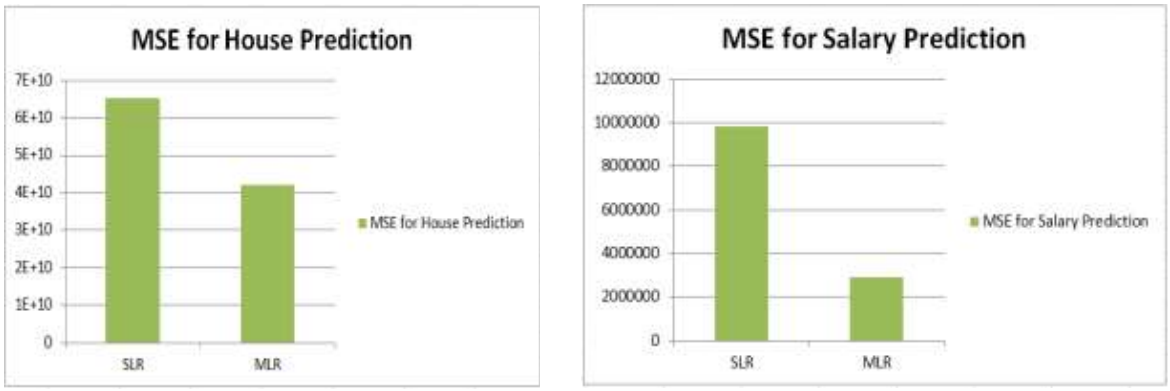

Figure 17. MSE for Salary and House Prediction.

Figure 18. MDAE for Salary and House Prediction.

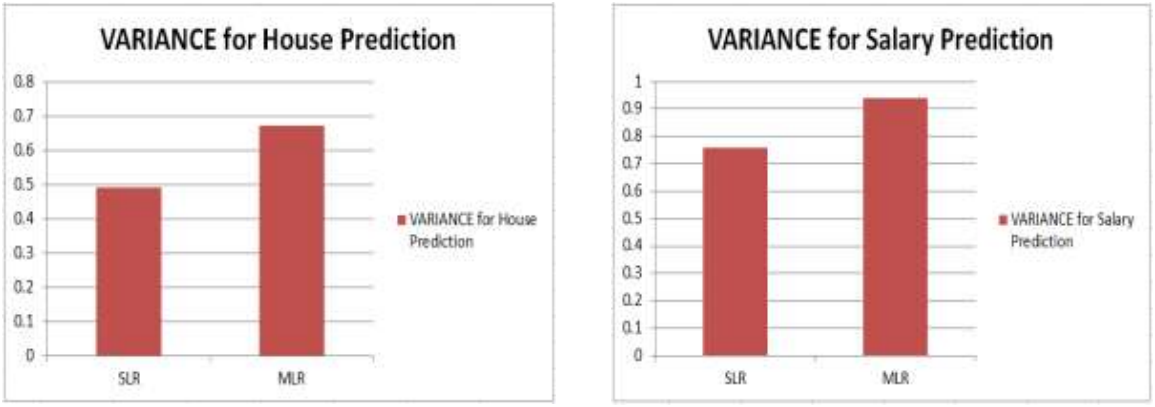

Figure 19. VARIANCE for Salary and House Prediction.

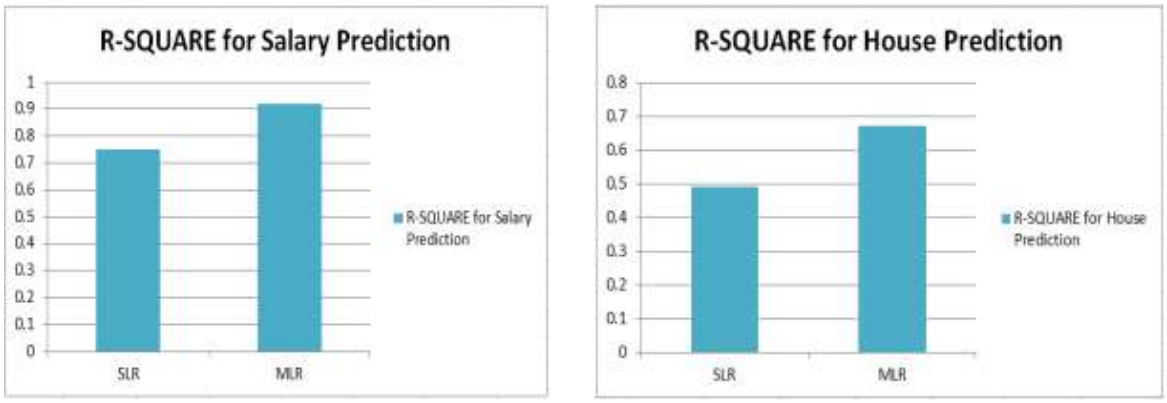

Figure 20. R-SQUARE for Salary and House Prediction.

Figures 15 to 20, show that higher accuracy can be achieved using MLR as compared to SLR because it has a lower value of errors as well as high value for the coefficient of determination.

6. Conclusion

On comparing the two algorithms SLR and MLR based on different performance metrics as RMSE, coefficient of determination (R^2), variance score, MAE, MSE, and MDAE the following results are drawn. For the House Pricing data, MLR has the R-squared value of 0.67 and that for SLR is 0.49. For the Salary Prediction data, MLR has the R-squared value of 0.92 and that for SLR the value is 0.75. Hence it is concluded that MLR gives better performance than SLR. It is recommended that it would be better to work with a large dataset to draw a better picture of the model.

References

[1] Rong S and Bao-wen Z 2018 The research of regression model in machine learning field *MATEC Web of Conf.* **vol. 176**

[2] Shinde N and Gawande K Jun 2018 Valuation of house prices using predictive techniques *Int. Journal of Advances in Electronics and Comp. Sci.* **vol. 5** Issue 6

[3] Khamis A B and Kamarudin N K K B Dec 2014 Comparative study on estimate house price using statistical and neural network model *International Journal of Scientific & Technology Research* **vol 3** Issue 12

[4] Navyashree M, Navyashree M K, Neetu M, Pooja G R and Arun Biradar May 2019 Salary

prediction in IT job market *Int. Journal of Comp. Sci. and Engineering* **vol. 7**, Special Issue 15

[5] Khongchai P and Songmuang P 2016 Implement of salary prediction system to improve student motivation using data mining techniques *Int. Conf. on Knowledge, Info. and Creativity Support Systems*

[6] Mohd T, Masrom S and Johari N Sep 2019 Machine learning house price prediction in petaling Jaya , Selangor, Malaysia *Int. Journal of Recent Tech. and Engineering*, **vol. 8**, Issue 2S11

[7] Uysal I and Guvenir H A 1999 An overview of regression techniques for knowledge discovery *The Knowledge Engineering Review* **vol. 14:4** pp. 319-40

[8] Das S, Barik R and Mukherjee A Sep 2019 Salary prediction using regression techniques *Int. Conf. On Industry Interactive Innovations in Sci. , Engineering and Tech*

[9] Satish G N, Raghavendran C V, Rao M D S and Srinivasulu C July 2019 House price prediction using machine learning *Int. Journal of Innovative Tech. and Exploring Engineering*, **vol. 8** Issue 9

[10] Rao S S Feb 2020 Stock prediction analysis by using linear regression machine learning algorithm *Int. Journal of Innovative Tech. and Exploring Engineering* **vol. 9** Issue 4

[11] Chakraborty, Satadruti, and Mitra D 2018 A study on consumers adoption intention for digital wallets in India *International Journal on Customer Relations* **vol. 6** Issue 1 pp. 38-57

[12] Kumar R and Dr. Rajesh Verma Aug 2012 Classification algorithm for data mining *Int. Journal of Innovations in Engineering and Tech.* **vol. 1** Issue 2

[13] Al A and Ami M Z 2019 A deep level understanding of linear regression with practical implementation in Scikit Learn *Wavy AI Research Foundation Linear Regression*

[14] Kumari K and Yadav S 2018 Linear regression analysis study *Journal of the Practice of Cardiovascular Sci.* **vol. 4** Issue 1 pp. 33-6

[15] Yu L 2014 A study of english reading ability based on multiple linear regression analysis *Journal of Chemical and Pharmaceutical Research* **vol. 6** Issue 6 pp. 1870-77

[16] Uyanik G K and Güler N 2013 A study on multiple linear regression analysis *Int. Conference on New Horizons in Education* pp. 234-40

[17] Kartal B and Denli H H 2019 Estimation of housing sales prices by multiple linear regression analysis(MLRA) for the determination of economically effective earthquake based urban transformation location *Fresenius Environmental Bulletin* **vol. 28** pp. 937-40.

[18] Kadam, A K, Wagh, V M, Muley, A A et al. 2019 Prediction of water quality index using artificial neural network and multiple linear regression modelling approach in Shivganga River basin, India *Model. Earth Syst. Environ.* **5**, pp. 951–62

[19] García-Bárzana, M Ramos-Guajardo, A B Colubi et al. 2020 Multiple linear regression models for random intervals: a set arithmetic approach *Comput. Stat.* **35** pp. 755–73

[20] https://www.kaggle.com/shree1992/housedata