International Journal of Innovative Technology and Exploring Engineering (IJITEE)
ISSN: 2278-3075 (Online), Volume-8 Issue-12, October 2019

Exploration of Multiple Linear Regression with Ensembling Schemes for Roof Fall Assessment using Machine Learning

M. Shyamala Devi, Shakila Basheer, Rincy Merlin Mathew

Abstract*: *Roof fall of the building is the major threat to the society as it results in severe damages to the life of the people. Recently, engineers are focusing on the prediction of roof fall of the building in order to avoid the damage to the environment and people. Early prediction of Roof fall is the social responsibility of the engineers towards existence of health and wealth of the nation. This paper attempts to identify the essential attributes of the Roof fall dataset that are taken from the UCI Machine learning repository for predicting the existence of roof fall. In this paper, the important features are extorted from the various ensembling methods like Gradient Boosting Regressor, Random Forest Regressor, AdaBoost Regressor and Extra Trees Regressor. The extracted feature importance of each of the ensembling methods is then fitted with multiple linear regression to analyze the performance. The same extracted feature importance of each of the ensembling methods are subjected to feature scaling and then fitted with multiple linear regression to analyze the performance. The Performance analysis is done with the performance parameters such as Mean Squared Log Error (MSLE), Mean Absolute error (MAE), R2 Score, Mean Squared error (MSE) and Explained Variance Score (EVS). The execution is carried out using python code in Spyder Anaconda Navigator IP Console. Experimental results shows that before feature scaling, Extra Tree Regressor is found to be effective with the MSE of 0.06, MAE of 0.07, R2 Score of 87%, EVS of 0.89 and MSLE of 0.02 as compared to other ensembling methods. In the same way, after applying feature scaling, the feature importance extracted from the Extra Tree Regressor is found to be effective with the MSE of 0.04, MAE of 0.03, R2 Score of 96%, EVS of 0.9 and MSLE of 0.01 as compared to other ensembling methods.

Index Terms*: *Machine Learning, Regression, MSLE, MAE, MSE and EVS.

I. INTRODUCTION

In machine learning dimensionality reduction, the final prediction results are based on number of input components.

Revised Manuscript Received on October 30, 2019.
***** Correspondence Author

M. Shyamala Devi*, Associate Professor, Computer Science and Engineering, Vel Tech Rangarajan Dr. Saguntthala R&D Institute of Science and Technology, Avadi, Chennai, TamilNadu, India.

Shakila Basheer, Assistant Professor, Department of Information Systems, College of Computer and Information Sciences, Princess Nourah Bint Abdul Rahman university, Riyadh-11671, Saudi Arabia

Rincy Merlin Mathew, Lecturer, Department of Computer Science, College of Science and Arts, Khamis Mushayt, King Khalid university, Abha, Asir, Saudi Arabia.

In general many of the attributes in the dataset are correlated and they may be redundant. This attribute duplication enlarges the memory space and diminishes the execution time.

The prediction of the target variable will be difficult when the number of attribute is large. As well, when the number of attribute is more, it is tough to visualize the features to predict the dependent variable. This guarantees the requirement of dimensionality reduction algorithms.

The paper is organized in such a way that Section 2 deals with the related works. Section 3 discuss about dimensionality Reduction. Proposed work is discussed in Section 4 followed by the implementation and Performance Analysis in Section 5. The paper is concluded with Section 6.

II. RELATED WORK

A. Literature Review

Boundary layer wind tunnel dataset [1] is subjected to predict the mean, root mean square and peak pressure coefficient of low rise building patterns. It also predicts the magnitude and pressure distribution of roof in the building based on the free stream turbulent flow conditions. Automatic labeling of roof shape for the building [2] is done from the GIS data. The roof top detection is done by using ROC analysis [3] to estimate the cost of error from the aerial images of building. It also analyzes the performance of both training and testing image datasets. Roof fall accidents are predicted using binary logistic regression model [4]. Roof fall risk is identified using risk assessment classical methodology [5] with fuzzy based approach. It projects that the roof fall is due to various features like low resolution, fixed weighting, sharp class boundaries of the building.

The fuzzy approach prioritizes the features and analyzes the roof fall susceptibility. The estimation of combined building frame and roof frame bolting is done in [6] to predict the roof fall of the building region. The falling speed of the roof is tensile failure parallel to the heading direction and the prediction of unsupported instantaneous roof [7] is done by using the excavation features of the building. The roof fall of the large road is due to the horizontal tensile stress [8] and it is predicted by using the surface stress and fault slip of the surface. The roof fall detection fails due to the low quality of the digital surface model of the building [9].

A critical review on various feature selection, feature extraction methods, classification methods and the performances parameters are examined for predicting the wine quality [10]-[26].

III. PROPOSED WORK

In our proposed work, machine learning algorithms are used to predict the disease of Roof Fall data set. Our contribution in this paper is folded in four ways.

(i) Firstly, the important features are extorted from the Roof Fall dataset using various ensembling methods like Gradient Boosting Regressor, Random Forest Regressor, AdaBoost Regressor and Extra Trees Regressor.

(ii) Secondly, the extracted feature importance of each of the ensembling methods is then fitted with multiple linear regression to analyze the performance.

(iii) Thirdly, the same extracted feature importance of each of the ensembling methods are subjected to feature scaling and then fitted with multiple linear regression to analyze the performance.

(iv) Fourth, the Performance analysis is done with the performance parameters such as Mean Squared Log Error (MSLE), Mean Absolute error (MAE), R2 Score, Mean Squared error (MSE) and Explained Variance Score (EVS).

A. System Architecture

The overall design of this paper work is shown in Fig. 1

IV. IMPLEMENTATION AND PERFORMANCE ANALYSIS

A. Roof Fall Prediction for Feature Extraction

The Roof fall dataset from UCL ML Repository is used for implementation with 19 independent attribute and 1 roof fall rate dependent attribute. The attribute are shown below.

1) CMRR
2) BOLT_LENGTH_ft
3) BOLT_TENSION
4) BOLT_GROUT_COLUMN
5) BOLT_CAPACITY_Kips
6) BOLTS_PER_ROW
7) ROW_SPACING_ft
8) ENTRY_WIDTH_ft
9) PRSUP
10) INTERSECTION_DIAGONAL_ft
11) DEPTH_of_COVER_ft
12) MINING_HEIGHT_ft
13) DRIVAGE_10000_ft
14) No._of_3WAY
15) No._of_4Way
16) No._of_SEGMENTS
17) FALLS_3W
18) FALLS_4W
19) SEGMENT_FALLS
20) ROOF_FALL_RATE - Dependent Attribute

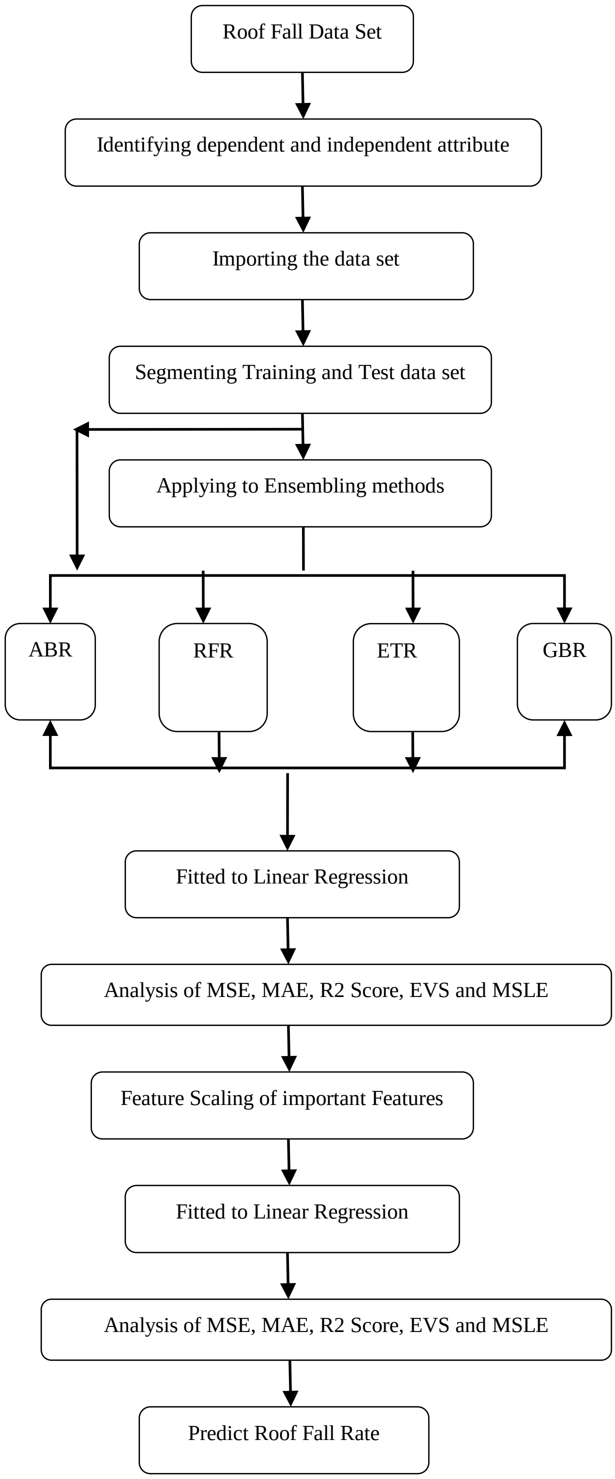


Fig. 1 System Architecture

B. Performance Analysis

The attribute and component relationship between the features of the dataset is shown in the fig 2.

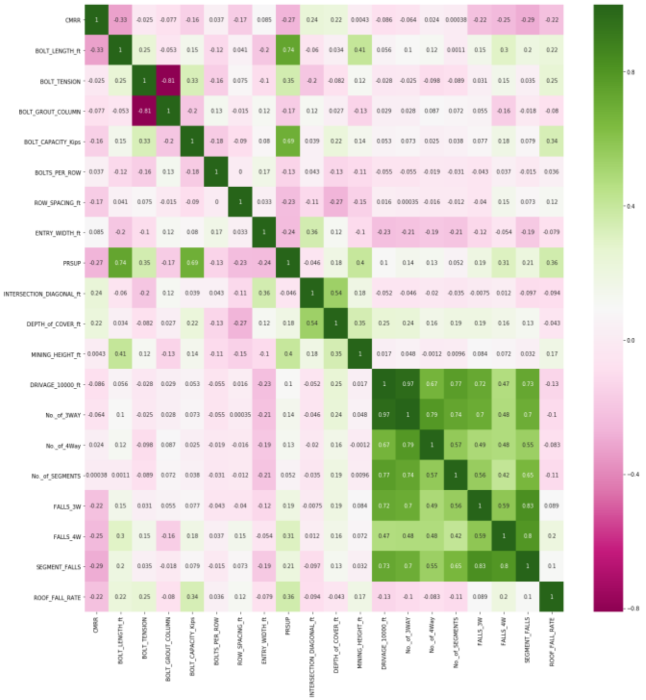


Fig. 2. Correlation Matrix of Roof Fall data set

The important features are extorted from the Roof fall dataset using AdaBoost Regressor and are shown in fig 3.

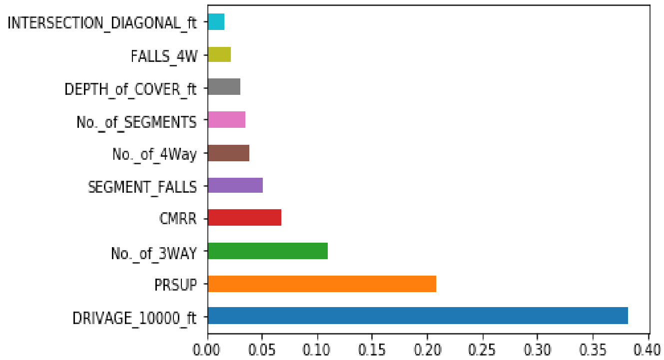


Fig. 3. Feature Importance of Ada Boost Regressor

The important features are extorted from the Roof fall dataset using Extra Trees Regressor and are shown in fig 4.

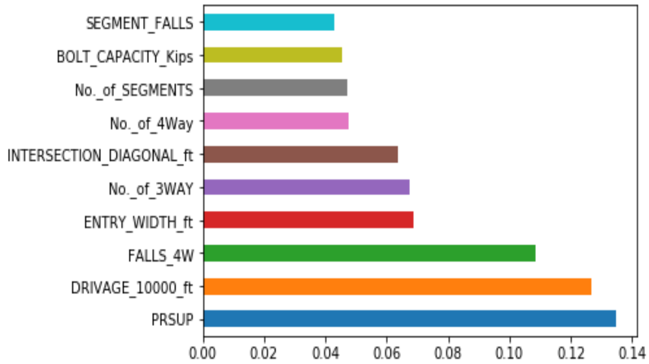
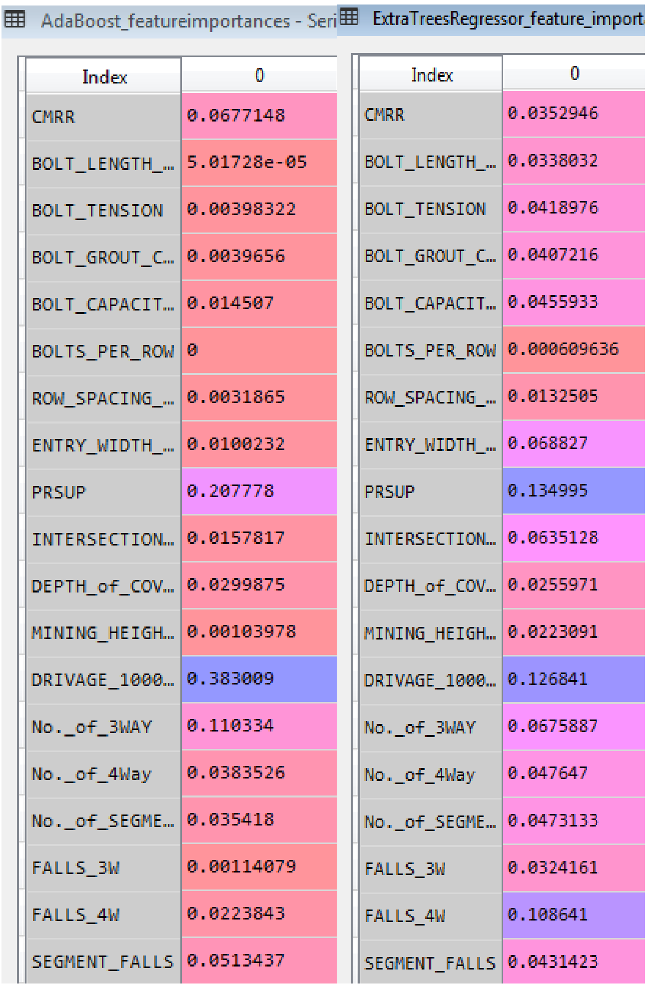


Fig. 3. Feature Importance of Extra Trees Regressor

The important features are extorted from the Roof Fall dataset using AdaBoost Regressor and Extra Trees Regressor and the features is shown in fig 5.



AdaBoost_featureimportances - Seri

Index	0
CMRR	0.0677148
BOLT_LENGTH_...	5.01728e-05
BOLT_TENSION	0.00398322
BOLT_GROUT_C...	0.0039656
BOLT_CAPACIT...	0.014507
BOLTS_PER_ROW	0
ROW_SPACING_...	0.0031865
ENTRY_WIDTH_...	0.0100232
PRSUP	0.207778
INTERSECTION...	0.0157817
DEPTH_of_COV...	0.0299875
MINING_HEIGH...	0.00103978
DRIVAGE_1000...	0.383009
No._of_3WAY	0.110334
No._of_4Way	0.0383526
No._of_SEGME...	0.035418
FALLS_3W	0.00114079
FALLS_4W	0.0223843
SEGMENT_FALLS	0.0513437

ExtraTreesRegressor_feature_import

Index	0
CMRR	0.0352946
BOLT_LENGTH_...	0.0338032
BOLT_TENSION	0.0418976
BOLT_GROUT_C...	0.0407216
BOLT_CAPACIT...	0.0455933
BOLTS_PER_ROW	0.000609636
ROW_SPACING_...	0.0132505
ENTRY_WIDTH_...	0.068827
PRSUP	0.134995
INTERSECTION...	0.0635128
DEPTH_of_COV...	0.0255971
MINING_HEIGH...	0.0223091
DRIVAGE_1000...	0.126841
No._of_3WAY	0.0675887
No._of_4Way	0.047647
No._of_SEGME...	0.0473133
FALLS_3W	0.0324161
FALLS_4W	0.108641
SEGMENT_FALLS	0.0431423

Fig. 5. Feature Important Attibutes of Ada Boost and Extra Trees Regressor

The important features are extorted from the Roof Fall dataset using Gradient Boosting Regressor and are shown in fig 6.

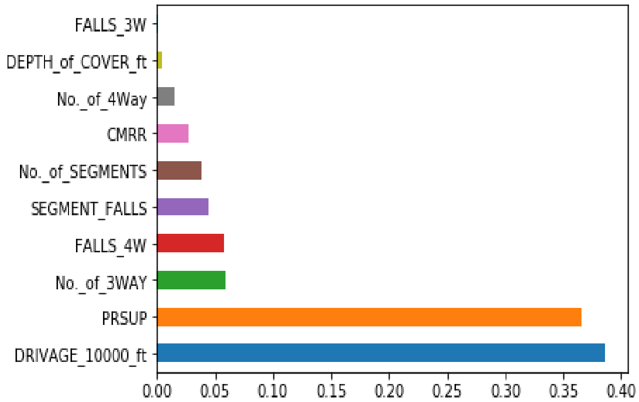


Fig 6. Feature Importance of Gradient Boost Regressor

The important features are extorted from Roof Fall e dataset using Random Forest Regressor and are shown in fig 7.

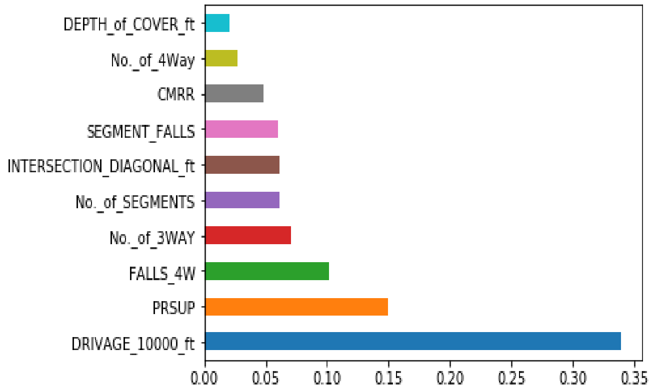
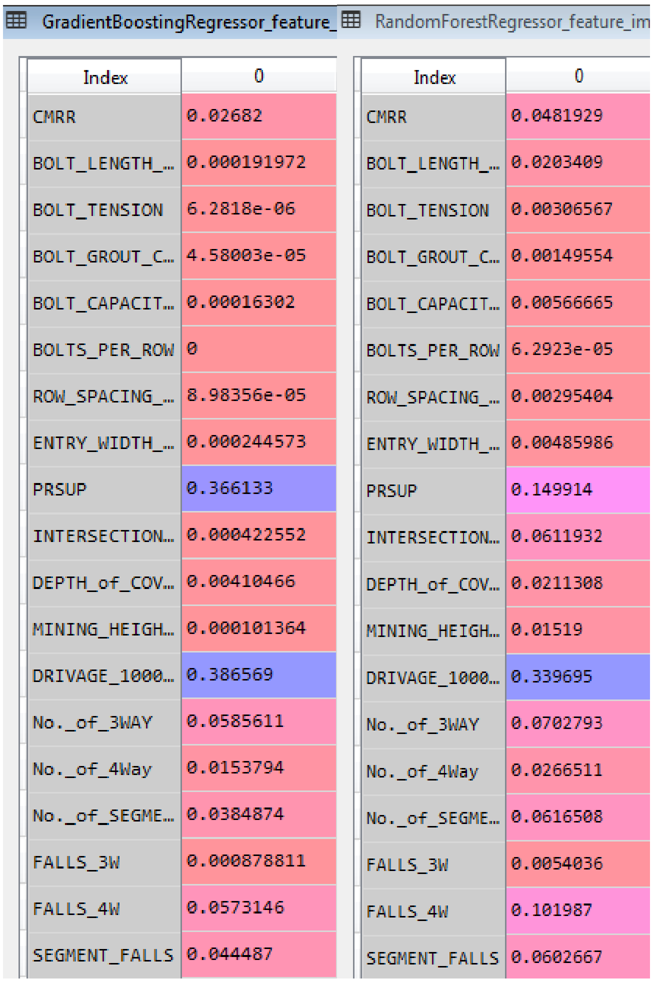


Fig. 8. Feature Importance of Random Forest Regressor

The important features are extorted from the Roof Fall dataset using Gradient Boosting Regressor and Random Forest Regressor and the features is shown in fig 9.



GradientBoostingRegressor_feature_

Index	0
CMRR	0.02682
BOLT_LENGTH_...	0.000191972
BOLT_TENSION	6.2818e-06
BOLT_GROUT_C...	4.58003e-05
BOLT_CAPACIT...	0.00016302
BOLTS_PER_ROW	0
ROW_SPACING_...	8.98356e-05
ENTRY_WIDTH_...	0.000244573
PRSUP	0.366133
INTERSECTION...	0.000422552
DEPTH_of_COV...	0.00410466
MINING_HEIGH...	0.000101364
DRIVAGE_1000...	0.386569
No._of_3WAY	0.0585611
No._of_4Way	0.0153794
No._of_SEGME...	0.0384874
FALLS_3W	0.000878811
FALLS_4W	0.0573146
SEGMENT_FALLS	0.044487

RandomForestRegressor_feature_im

Index	0
CMRR	0.0481929
BOLT_LENGTH_...	0.0203409
BOLT_TENSION	0.00306567
BOLT_GROUT_C...	0.00149554
BOLT_CAPACIT...	0.00566665
BOLTS_PER_ROW	6.2923e-05
ROW_SPACING_...	0.00295404
ENTRY_WIDTH_...	0.00485986
PRSUP	0.149914
INTERSECTION...	0.0611932
DEPTH_of_COV...	0.0211308
MINING_HEIGH...	0.01519
DRIVAGE_1000...	0.339695
No._of_3WAY	0.0702793
No._of_4Way	0.0266511
No._of_SEGME...	0.0616508
FALLS_3W	0.0054036
FALLS_4W	0.101987
SEGMENT_FALLS	0.0602667

Fig. 9. Feature Important Attibutes of Gradient Boost and Random Foresr Regressor

The Performance analysis is done with the performance parameters such as Mean Squared Log Error (MSLE), Mean Absolute error (MAE), R2 Score, Mean Squared error (MSE) and Explained Variance Score (EVS).and is shown in the Table 1 -Table 5.

Table. 1 Analysis of MSE, MAE and R2 Score parameters

Ensembling Methods	Fitting to Linear Regression Before Feature Scaling		
	MSE	MAE	R2 Score
AdaBoost Regressor	0.14	0.13	0.82
Random Forest Regressor	0.12	0.12	0.81
Extra Trees Regressor	0.06	0.07	0.87
Gradient Boosting Regressor	0.09	0.11	0.78

Table. 2 Analysis of EVS and MSLE parameters

Ensembling Methods	Fitting to Linear Regression Before Feature Scaling	
	EVS	MSLE
AdaBoost Regressor	0.81	0.16
Random Forest Regressor	0.78	0.18
Extra Trees Regressor	0.89	0.02
Gradient Boosting Regressor	0.79	0.17

Table. 3 Analysis of MSE, MAE and R2 Score parameters after Feature Scaling

Ensembling Methods	Fitting to Linear Regression after Feature Scaling		
	MSE	MAE	R2 Score
AdaBoost Regressor	0.26	0.26	0.87
Random Forest Regressor	0.24	0.24	0.91
Extra Trees Regressor	0.04	0.03	0.96
Gradient Boosting Regressor	0.19	0.20	0.93

Table. 4 Analysis of EVS and MSLE parameters after Scaling

Ensembling Methods	Fitting to Linear Regression after Feature Scaling	
	EVS	MSLE
AdaBoost Regressor	0.84	0.06
Random Forest Regressor	0.83	0.05
Extra Trees Regressor	0.92	0.01
Gradient Boosting Regressor	0.82	0.04

International Journal of Innovative Technology and Exploring Engineering (IJITEE)
ISSN: 2278-3075 (Online), Volume-8 Issue-12, October 2019

Table. 5 Analysis of MSE, MAE, R2 Score EVS and MSLE Score for Multiple Linear Regression without Ensembling.

Metrics	Fitting To Linear Regression
MSE	0.18
MAE	0.14
R2 Score	0.79
EVS	0.89
MSLE	0.18

The analysis of the performance parameter without ensembling, with ensembling before and after applying feature scaling is shown in the Fig 10 – Fig 14.

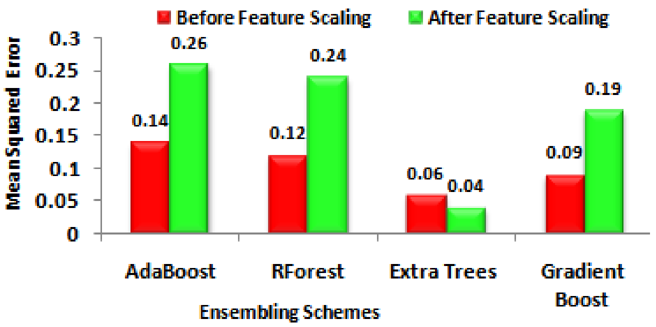


Fig.10. Analysis of MSE and Ensembling methods

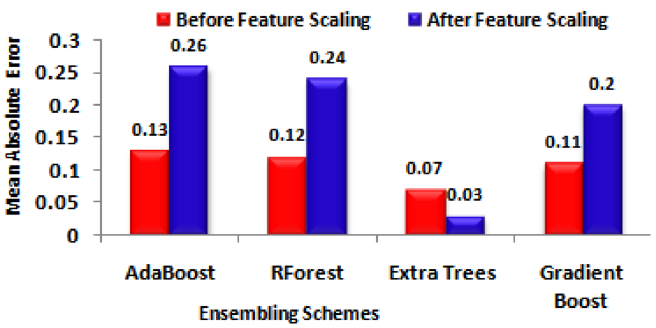


Fig.11. Analysis of MAE and Ensembling methods

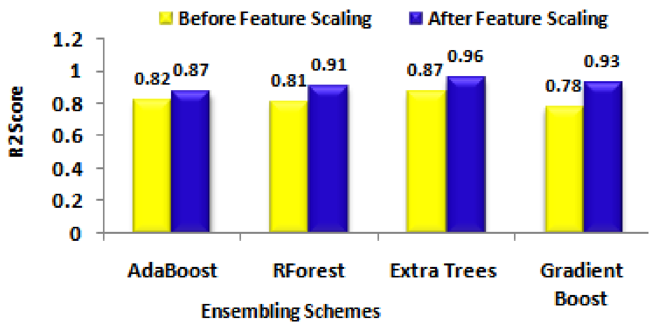


Fig.12. Analysis of R2Score and Ensembling methods

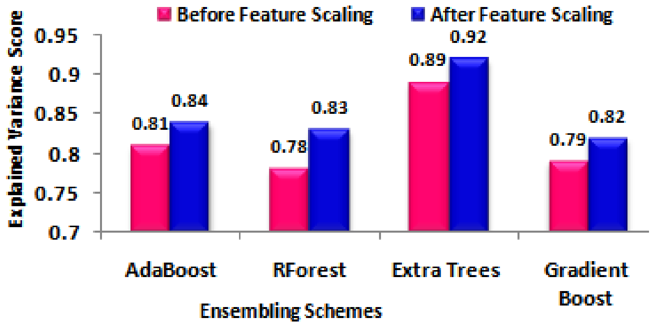


Fig.13. Analysis of EVS and Ensembling methods

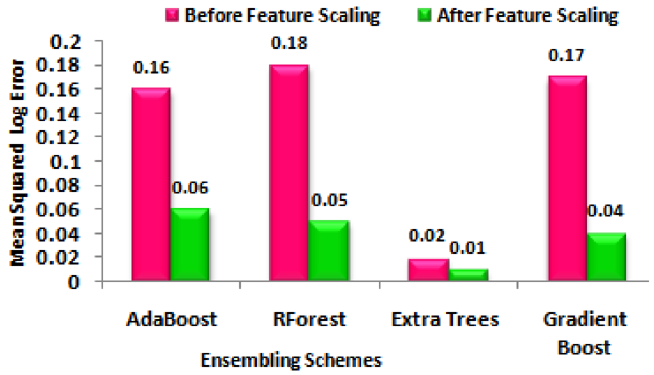


Fig.13. Analysis of MSLE and Ensembling methods

V. CONCLUSION

This paper analyzes the prediction of the existence of Roof Fall of the Roof Fall dataset through the ensembling methods before and after applying attribute scaling method. Experimental results shows that before feature scaling, Extra Tree Regressor is found to be effective with the MSE of 0.06, MAE of 0.07, R2 Score of 87%, EVS of 0.89 and MSLE of 0.02 as compared to other ensembling methods. In the same way, after applying feature scaling, the feature importance extracted from the Extra Tree Regressor is found to be effective with the MSE of 0.04, MAE of 0.03, R2 Score of 96%, EVS of 0.9 and MSLE of 0.01 as compared to other ensembling methods

REFERENCES

1. L. Fernández Cabán Pedro, J. Masters Forrest, and M. Phillips Brian, "Predicting Roof Pressures on a Low-Rise Structure From Freestream Turbulence Using Artificial Neural Networks ", Frontiers in Built Environment., vol. 4, Nov. 2018, pp. 68.
2. J. Castagno, and E. Atkins, "Roof Shape Classification from LiDAR and Satellite Image Data Fusion Using Supervised Learning". Journal of Sensors (Basel, Switzerland)., vol. 18(11), , Nov. 2018, pp. 3960.
3. M.A. Maloof, P. Langley, T.O. Binford, R. Nevatia, and S. Sage,"Improved Rooftop Detection in Aerial Images with Machine Learning", Journal of Machine Learning., vol 53, Oct. 2003, pp. 157-191.
4. Sanjay Kumar Palei, and Samir Kumar Das, "Logistic regression model for prediction of roof fall risks in bord and pillar workings in coal mines: An approach", Safety Science., vol 47(1), Jan. 2009, pp. 88-96.
5. Ebrahim Ghasemi, Mohammad Ataei, and Kourosh Shahriar, " Improving the Method of Roof Fall Susceptibility Assessment Based on Fuzzy Approach", Archives of Mining Sciences., vol. 62(1), March 2017, pp. 13-32.

6. S. I. Ivanov, N. V. Titov, A. A. Privalov, I. T. Trunov, and V. I. Sarychev, " Calculation of parameters of combined frame and roof bolting", Earth and Environmental Science., vol. 87, 2017. doi :10.1088/1755-1315/87/5/052009
7. Yang Sen Zhang, Nong Feng, Xiaowei Pan, Dongjiang Qian, and Deyu, "The Influence of Heading Rate on Roof Stability in Coal Entry Excavation", Advances in Civil Engineering., 2018.pp. 1-15.
8. Wang, Hongwei Xue, Sheng Jiang, Yaodong Deng, Daixin Shi, Suzhen Zhang, and Dengqian, "Field Investigation of a Roof Fall Accident and Large Roadway Deformation Under Geologically Complex Conditions in an Underground Coal Mine", Rock Mechanics and Rock Engineering.. vol. 51, 2018.
9. T. Partovi, F. Fraundorfer, S. Azimi, D. Marmanis, and P. Reinartz, " Roof Type Selection Based on Patch-Based Classification Using Deep Learning For High Resolution Satellite Imagery", The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, vol. XLII-1, Jun. 2017.
10. R. Suguna, M. Shyamala Devi, and Rincy Merlin Mathew, " Customer Churn Predictive Analysis by Component Minimization using Machine Learning", International Journal of Innovative Technology and Exploring Engineering, vol. 8, no.8, June 2019, pp. 2329-2333.
11. Suguna Ramadass, and Shyamala Devi Munisamy, Praveen Kumar P, Naresh P, "Prediction of Customer Attrition using Feature Extraction Techniques and its Performance Assessment through dissimilar Classifiers", Springer's book series "Learning and Analytics in Intelligent Systems, Springer, , LAIS vol. 3, pp. 613-620, 2019.
12. R.Suguna, M. Shyamala Devi, Rupali Amit Bagate, and Aparna Shashikant Joshi, "Assessment of Feature Selection for Student Academic Performance through Machine Learning Classification", Journal of Statistics and Management Systems, Taylor Francis, vol. 22, no. 4, 25 June 2019, pp. 729-739.
13. R. Suguna, M. Shyamala Devi, Rincy Merlin Mathew, "Customer Segment Prognostic System by Machine Learning using Principal Component and Linear Discriminant Analysis", International Journal of Recent Technology and Engineering, Volume-8 Issue-2, 30 July 2019. pp. 6198-6203.
14. M. Shyamala Devi, Rincy Merlin Mathew, and R. Suguna,"Attribute Heaving Extraction and Performance Analysis for the Prophesy of Roof Fall Rate using Principal Component Analysis", International Journal of Innovative Technology and Exploring Engineering, vol. 8, no.8, June 2019, pp. 2319-2323.
15. Shyamala Devi Munisamy, and Suguna Ramadass Aparna Joshi, "Cultivar Prediction of Target Consumer Class using Feature Selection with Machine Learning Classification", Springer's book series "Learning and Analytics in Intelligent Systems, Springer, LAIS vol. 3, pp. 604-612, 2019.
16. M. Shyamala Devi, Rincy Merlin Mathew, R. Suguna, "Feature Snatching and Performance Analysis for Connoting the Admittance Likelihood of student using Principal Component Analysis", International Journal of Recent Technology and Engineering, Volume-8 Issue-2, 30 July 2019.pp. 4800-4807.
17. M. Shyamala Devi, Rincy Merlin Mathew, R. Suguna, " Regressor Fitting of Feature Importance for Customer Segment Prediction with Ensembling Schemes using Machine Learning", International Journal of Engineering and Advanced Technology, Volume-8 Issue-6, pp. 952 – 956, 30 August 2019
18. R. Suguna, M. Shyamala Devi, Rincy Merlin Mathew, "Integrating Ensembling Schemes with Classification for Customer Group Prediction using Machine Learning", International Journal of Engineering and Advanced Technology, Volume-8 Issue-6, pp. 957 – 961, 30 August 2019.
19. Rincy Merlin Mathew, R. Suguna, M. Shyamala Devi, "Composite Model Fabrication of Classification with Transformed Target Regressor for Customer Segmentation using Machine Learning", International Journal of Engineering and Advanced Technology, Volume-8 Issue-6, pp. 962 – 966, 30 August 2019.
20. M. Shyamala Devi, Shefali Dewangan, Satwat Kumar Ambashta, Anjali Jaiswal, Nariboyena Vijaya Sai Ram, "Backward Eliminated Formulation of Fire Area Coverage using Machine Learning Regression", International Journal of Innovative Technology and Exploring Engineering, vol. 8, no.12, 10 October 2019. (Accepted for Publication)
21. M. Shyamala Devi, Ankita Shil, Prakhar Katyayan, Tanmay Surana, "Constituent Depletion and Divination of Hypothyroid Prevalance using Machine Learning Classification", International Journal of Innovative Technology and Exploring Engineering, vol. 8, no.12, 10 October 2019. (Accepted for Publication)
22. M. Shyamala Devi, Shefali Dewangan, Satwat Kumar Ambashta, Anjali Jaiswal, Sairam Kondapalli, "Recognition of Forest Fire Spruce Type Tagging using Machine Learning Classification", International Journal of Recent Technology and Engineering, Volume-8 Issue-3, pp. 4309 – 4313, 16 September 2019.
23. M. Shyamala Devi, Usha Vudatha, Sukriti Mukherjee, Bhavya Reddy Donthiri, S B Adhiyan, Nallareddy Jishnu, " Linear Attribute Projection and Performance Assessment for Signifying the Absenteeism at Work using Machine Learning", International Journal of Recent Technology and Engineering, Volume-8 Issue-3, pp. 1262 – 1267, 16 September 2019.
24. M. Shyamala Devi, Mothe Sunil Goud, G. Sai Teja, MallyPally Sai Bharath, "Heart Disease Prediction and Performance Assessment through Attribute Element Diminution using Machine Learning", International Journal of Innovative Technology and Exploring Engineering, vol. 8, no.11, pp. 604 – 609, 30 September 2019.
25. Shakila Basheer, Rincy Merlin Mathew, M. Shyamala Devi, "Ensembling Coalesce of Logistic Regression Classifier for Heart Disease Prediction using Machine Learning", International Journal of Innovative Technology and Exploring Engineering, vol. 8, no.12, 10 October 2019, pp. (Accepted for Publication)
26. Rincy Merlin Mathew, M. Shyamala Devi, Shakila Basheer," Exploration of Neighbor Kernels and Feature Estimators for Heart Disease Prediction using Machine Learning", International Journal of Innovative Technology and Exploring Engineering, vol. 8, no.12, 10 October 2019. (Accepted for Publication)