Attribute Heaving Extraction and Performance Analysis for the Prophesy of Roof Fall Rate using Principal Component Analysis

M. Shyamala Devi, Rincy Merlin Mathew, R. Suguna

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. By considering these aspects, this paper proposes the usage of machine learning algorithms for predicting the roof fall rate of the building. This paper uses Rooffall data set extracted from UCI machine learning repository and is subjected to the feature extraction methods like Principal Component Analysis (PCA), Kernel Principal Component Analysis, Sparse Principal Component Analysis, Mini Batch Sparse Principal Component Analysis and Incremental Principal Component Analysis. The optimized dimensionality reduced dataset from each of the above methods are then processed to find the mean squared error (MSE), Mean Absolute Error (MAE) and R2 Score. We have achieved the prediction of roof fall rate in two ways. Firstly, the dimensionality reduction is done using five feature extraction methods which results in the survival of sensible attribute to predict the roof fall rate. Secondly, the comparison of each method is done by the accuracy parameters. The performance analysis is done by implementing python scripts in Anaconda Spyder Navigator. Experimental Result shows that the Incremental PCA have achieved the effective prediction of roof fall rate with minimum MSE of 17.08, MAE of 3.19 and reasonable R2 Score of 0.20.

Index Terms*: *Machine Learning, Feature Extraction, PCA, MSE, MAE, R2 Score.

I. INTRODUCTION

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

Revised Manuscript Received on June 13, 2019

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

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

R. Suguna, Professor, Computer Science and Engineering, Vel Tech Rangarajan Dr. Sagunthala R&D Institute of Science and Technology, Avadi, Chennai, TamilNadu, India.

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]. It uses the convolution neural network method for the classification of the roof type. K-Modes, RBF network and Decision Trees [10] are used to predict the falling risk of the person due to the roof fall. Component based approach is used for designing the building [11] with two features as construction level components and zone level components which increases the prediction of roof fall. The machine learning feature selection and extraction methods can be used for the prediction of any factor in different application can be learnt through this article [12]–[14].

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

III. DIMENSIONALITY REDUCTION

Dimensionality reduction performs the transformation of high dimensional data to lower dimensional data without any loss of information. There are two grouping in dimensionality reduction: Feature selection and Feature Extraction. Feature selection seek at maintenance of only appropriate attributes in the given data set whereas feature extraction focus on finding an optimal set of attributes usually a combination of input attribute without loss of information.

IV. PROPOSED WORK

In our proposed work, machine learning algorithms are used to predict the roof fall rate of the building. Our contribution in this paper is folded in two ways.

(i) Firstly, the dimensionality reduction is done using five feature extraction methods PCA, Kernel PCA, Sparse PCA, Mini Batch Sparse PCA and Incremental PCA that results in the existence of sensible attributes to predict the dependent variable roof fall rate.

(ii) Secondly, the accuracy and the performance of the five methods is done by comparing the mean squared error, mean absolute error and R2 Score.

A. System Architecture

The system architecture of our proposed work is shown in Fig. 1

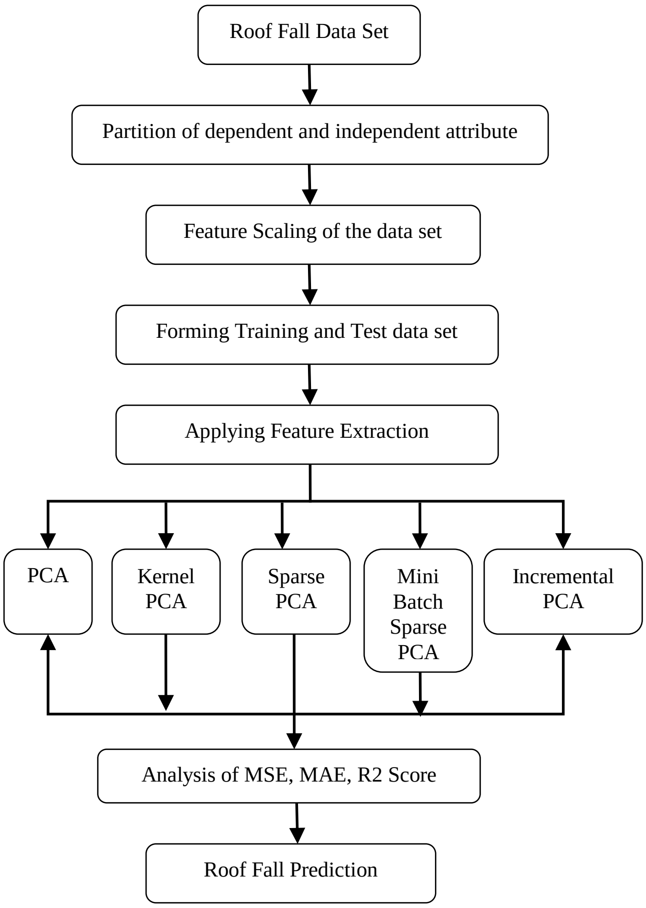

Fig. 1 System Architecture

B. Principal Component Analysis

It is the feature extraction method which extracts the new set of variables from the large set of variables in the dataset through linear transformation. The selected variables are the principal components and the steps of PCA are given below.

Step 1: Construct the Covariance matrix of the data set
Step 2: Calculate the Eigen vectors of that matrix
Step 3: The Eigen vectors with high Eigen values are used to reconstruct the original data set.
Step 4: The attributes with high variance are selected as principal components.

C. Kernel PCA

It is the feature extraction method which extracts the new set of variables from the large set of variables in the dataset through non linear transformation. The selected variables are the principal components and the steps of Kernel PCA are given below.

Step 1: Choose a Kernel Mapping
Step 2: Construct the normalized kernel matrix K of the training data set
Step 3: Construct the Covariance matrix of the data set
Step 4: Calculate the Eigen vectors of that matrix
Step 5: The Eigen vectors with high Eigen values are used to reconstruct the original data set.
Step 6: Solve the Eigen value of K to get the principal components.
Step 7: The attributes with high variance are selected as principal components.

D. Sparse PCA

It is the feature extraction method which extracts the new set of variables from the large set of variables in the dataset through non linear transformation with large variance and sparsity. The selected variables are the principal components such that they have less non zero values in their vector coefficients and the steps of sparse PCA are given below.

Step 1: Construct the normalized Sparse matrix S of the training data set
Step 2: Construct the Covariance matrix of the data set
Step 3: Calculate the Eigen vectors of that matrix
Step 4: Perform minimization problem for Sparse matrix.
Step 5: The Eigen vectors with high Eigen values are used to reconstruct the original data set.
Step 6: Solve the singular value decomposition for S and Eigen vectors with high variance and sparsity.
Step 7: The attributes with high variance and sparsity are selected as principal components.

E. Mini Batch Sparse PCA

It is the feature extraction method which extracts the new set of variables from the large set of variables in the dataset through non linear transformation of multiple batches with large variance and sparsity. The selected variables are the principal components and the steps are given below.

Step 1: Divide the data set into ‘n’ predefined batches
Step 2: Construct normalized Sparse matrix S for ‘n’ batches
Step 3: Construct the Covariance matrix of the data set
Step 4: Calculate the Eigen vectors of that matrix
Step 5: Perform minimization problem for ‘n’ Sparse matrix.

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

Step 6: The Eigen vectors with high Eigen values are used to reconstruct the original data set.
Step 7: Solve the singular value decomposition for S and Eigen vectors with high variance and sparsity.
Step 8: The attributes with high variance and sparsity are selected as principal components.

F. Incremental PCA

It is the feature extraction method which extracts the new set of variables from the large set of variables in the dataset through non linear transformation with low rank components. The selected variables are the principal components and the steps of incremental PCA are given below.

Step 1: Construct the Rank matrix R of the training data set
Step 2: Construct the Covariance matrix of the data set
Step 3: Calculate the Eigen vectors of that matrix
Step 4: Perform low rank approximation for rank matrix.
Step 5: The Eigen vectors with high Eigen values are used to reconstruct the original data set.
Step 6: Solve the singular value decomposition for R and Eigen vectors with high variance and low rank.
Step 7: The attributes with high variance and low rank are selected as principal components.

V. 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

Roof Fall dataset is applied to Feature extraction methods and the results of the roof fall prediction of components for each method is shown in Fig. 2 - Fig 6. The Roof Fall dataset is implemented in python and applied with feature extraction methods and the optimized components are further processed to find the accuracy parameters.

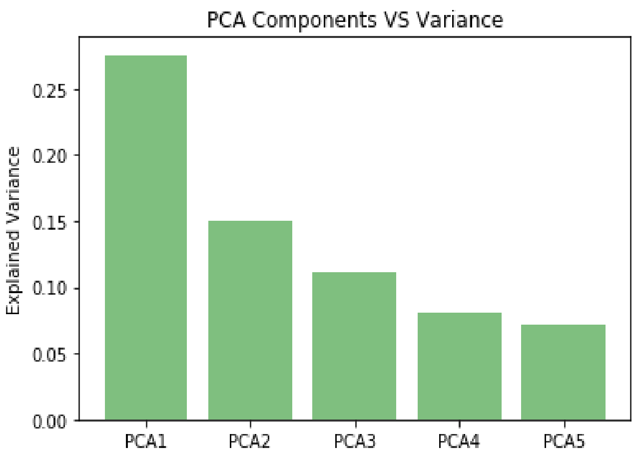

Fig. 2 Variance VS Principal Components

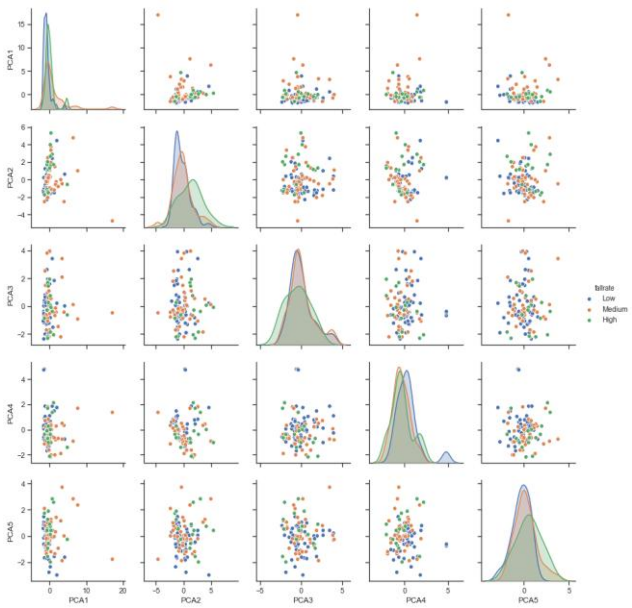

Fig. 3 Roof fall Prediction of PCA Components

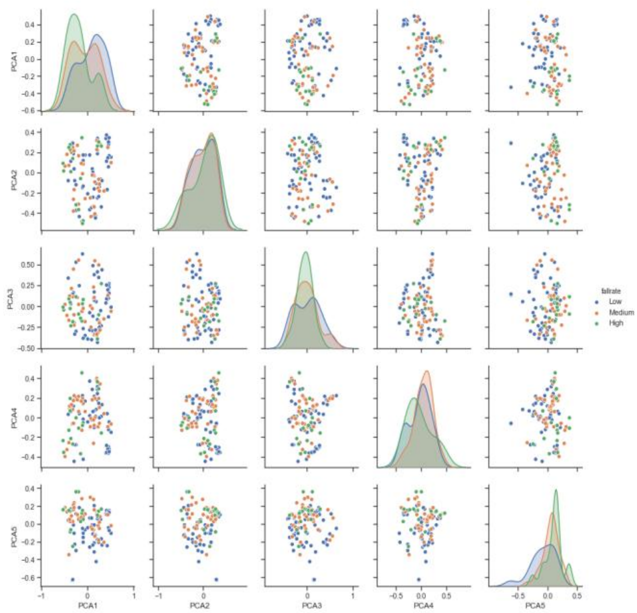

Fig. 4 Roof fall Prediction of Kernel PCA Components

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

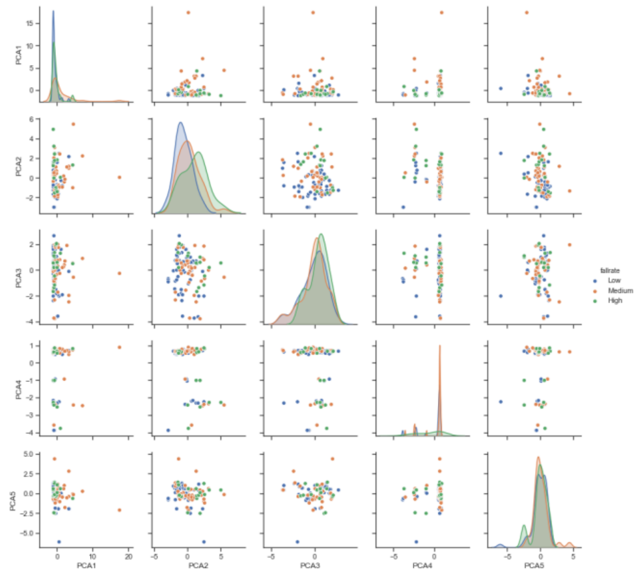

Fig. 5 Roof fall Prediction of Sparse PCA Components

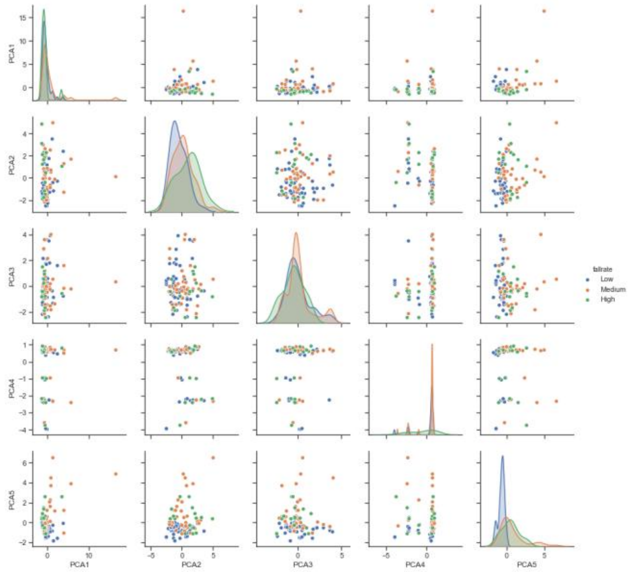

Fig. 6 Roof fall Prediction - MiniBatch SparsePCA Components

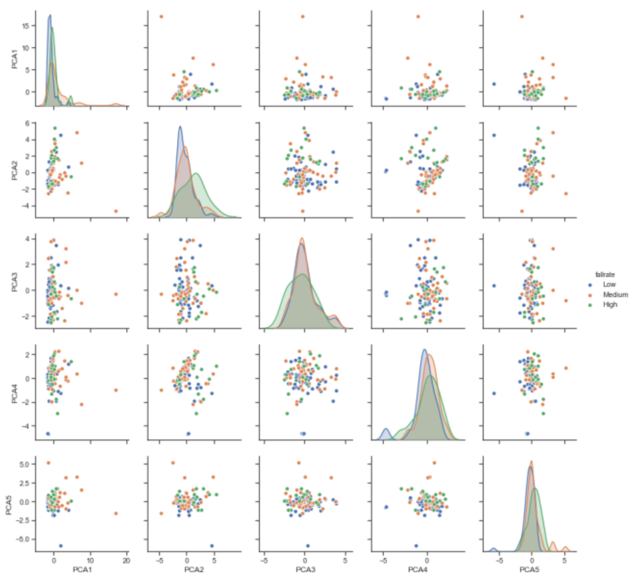

Fig. 7 Roof fall Prediction of Incremental PCA Components

B. Performance Comparison of PCA

The Performance analysis is done by comparing the mean squared error, mean absolute error and R2 score of each method. After implementing the feature extraction, the obtained evaluation parameters are shown in Table. 1.

Table. 1. Comparison of Evaluation Parameters

Feature Extraction Methods	Mean Squared Error	Mean Absolute Error	R2 Score
PCA	16.66	3.28	0.21
Kernel PCA	18.63	3.24	0.12
Sparse PCA	19.40	3.50	0.09
Mini Batch Sparse PCA	16.57	3.24	0.22
Incremental PCA	17.08	3.19	0.20

VI. CONCLUSION

This paper proposes a method to predict the fall rate for the Roof fall dataset which decreases the manual computation time thereby improving the early prediction. An attempt is made to implement the feature extraction for the Roof fall dataset using methods like PCA, Kernel PCA, Sparse PCA, Mini Batch Sparse PCA and Incremental PCA. The obtained optimized dataset from each of the above mentioned feature extraction is further compared and analyzed its performance of prediction through parameters like mean squared error, mean absolute error, R2 Score. Experimental Result shows that the Incremental PCA have achieved the effective prediction of roof fall rate with minimum MSE of 17.08, MAE of 3.19 and reasonable R2 Score of 0.20.

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. Hainan Chen, and Xiaowei Luo, " Severity Prediction Models of Falling Risk for Workers at Height", Creative Construction, June 2016, pp. 539-545.
11. Philipp Geyer, and Sundaravelpandian Singaravel, "Component-based machine learning for performance prediction in building design", Applied Energy, Elsevier., vol. 228,Oct 2018, pp. 1439-1453.
12. Shyamala Devi Munisamy, and Suguna Ramadass Aparna Joshi, "Cultivar Prediction of Target Consumer Class using Feature Selection with Machine Learning Classification (Accepted for publication)", Springer's book series entitled "Learning and Analytics in Intelligent Systems, Springer, 2019 to be published.
13. 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(Accepted for publication)", Springer's book series entitled "Learning and Analytics in Intelligent Systems, Springer, 2019 to be published.
14. R.Suguna, M. Shyamala Devi, Rupali Amit Bagate, and Aparna Shashikant Joshi, "Assessment of Feature Selection for Student Academic Performance through Machine Learning Classification (Accepted for publication)", Journal of Statistics and Management Systems, Taylor Francis, 2019 to be published.