International Journal of Recent Technology and Engineering (IJRTE)
ISSN: 2277-3878, Volume-8 Issue-3, September 2019

Recognition of Forest Fire Spruce Type Tagging using Machine Learning Classification

M. Shyamala Devi, Shefali Dewangan, Satwat Kumar Ambashta, Anjali Jaiswal, Sairam Kondapalli

Abstract*: In recent times, the natural resources are demolished due to the technological growth. The agricultural and the forest area are transformed to industries, Storage Warehouse and Container logistics companies to facilitate the living standards. This leads to scarcity of natural resources for the people to live a comfortable life. Due to the change of natural environment and fluctuations in the climate conditions, the forest has the chance of occurrence of fire. The forest fire is the resultant of high temperature, land mine, flight crashes and satellite damages from the environment. The precaution must be taken in advance to protect the coverage of fire. The less attention to fire control may lead to entire damage of the forest and the spreading of fire occurs due to the high wind blow. This makes researchers to focus on helping the forest area to overcome from the fire attack. The detection of fire type is a challenging task after the occurrence of the damage. With this view, we address the prediction of fire type classification using machine learning classification algorithms. The Forest Cover Type dataset is downloaded from UCI Data warehouse repository and done with classification analysis. The prediction of absent hours is achieved in the methodology of four steps. At first, the important feature attributes are found and depicted as a chart. Secondly, the raw dataset is applied to all the classification models like Logistic regression, Kernel SVM, KNN, Decision Tree, Naïve Bayes and Random Forest. Thirdly, the dataset is reduced with PCA and then the reduced dataset is fitted to all the classifiers. Fourth, Performance analysis is done by analyzing the performance metrics like Accuracy, FScore, Recall and Precision. The real time execution is performed by python in Anaconda Spyder Navigator Integrated Development Environment. Experimental Result shows that the Random Forest classifier is obtained with the accuracy of 92% before applying PCA. After applying PCA, the classifier namely Random forest is analyzed to be having the accuracy of 78% for 15 components, 83% for 20 components and 89% for 25 components.

***Index Terms*: Machine Learning, Precision, Recall, FScore and Accuracy.**

Manuscript published on 30 September 2019

* Correspondence Author

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.

Shefali Dewangan, II Year B.Tech Student, Computer Science and Engineering, Vel Tech Rangarajan Dr. Sagunthala R&D Institute of Science and Technology, Avadi, Chennai, TamilNadu, India.

Satwat Kumar Ambastha, II Year B.Tech Student, Computer Science and Engineering, Vel Tech Rangarajan Dr. Sagunthala R&D Institute of Science and Technology, Avadi, Chennai, TamilNadu, India.

Anjali Jaiswal, II Year B.Tech Student, Computer Science and Engineering, Vel Tech Rangarajan Dr. Sagunthala R&D Institute of Science and Technology, Avadi, Chennai, TamilNadu, India.

Sairam Kondapalli, II Year B.Tech Student, Computer Science and Engineering, Vel Tech Rangarajan Dr. Sagunthala R&D Institute of Science and Technology, Avadi, Chennai, TamilNadu, India.

I. INTRODUCTION

Natural calamity and Storm is the recent real time challenging platform which can be extended by using machine learning algorithms for forecasting. In machine learning techniques, the detection of dependent attribute is performed using regression or classification. However, it is essential to analyze and interpret the damages caused by any natural calamity. The existence of human beings is due to the availability of trees and water. The forest is the source for all animals to live in this earth. So the damages that are caused by the forest fire must be investigated in order to save the natural resources. The paper is prepared in which the review of literatures is discussed with Section 2 go after by the proposed work in the Section 3. Implementation and Performance Analysis is discussed in Section 4 followed by the conclusion of the paper in Section 5.

II. RELATED WORK

A. Literature Survey

The forest fire storm can be analyzed with the machine learning algorithms in SAMOA which is similar to Mahout for Hadoop. Vertical Hoeffding parallelizing streaming decision tree induction is included in SAMOA API technique for predicting forest cover types from the cartographic variables [1]. Abstract Mapping of forest attributes and connected variables is fundamental for forest planning and prediction design. The wall-to-wall mapping of forest attributes are done for northern Europe by using k-nearest neighbor's method for enhancing the wall-to-wall basal area, volume, and cover type estimation Some of the parameters like distance metric, weighting function, feature weighting parameters, and number of neighbors are also considered for prediction of the forest cover type [2]. The forest cover map is used for finding the cover type based on the Resources Renewable Planning Act (RPA). The High Resolution Radiometer (AVHRR) data of the forest were used to predict the forest cover type by using the multitemporal and multisource remote sensing data analysis methods [3]. A multi-source account method for the forest area is developed based on the k-nearest neighbor rule for creating the maps of selected forest variables.

It uses ground plan from the NFI, satellite image data and ground envelop maps from Landsat data[4]. The label noise is one of the major concern in classification. The review on label noise and its impact on classification is done by using label noise-robust, label noise cleansing, and label noise-tolerant algorithms [5].

The machine learning attribute collection and removal methods can be used for the detection of dependent attribute for various real time application can be understood through this article [6]–[15].

III. PROPOSED WORK

In this proposed work, we have used the classification algorithms for predicting the type of the fire in the forest. Our input of detection of dependent attribute fire type is accomplished in four ways.

(i) Firstly, the important feature attributes are found and depicted as a chart.

(ii) Secondly, the raw dataset is applied to all the classification models like Logistic regression, Kernel SVM, KNN, Decision Tree, Naive Bayes and Random Forest.

(iii) Thirdly, the dataset is reduced with PCA and then the reduced dataset is fitted to all the classifiers.

(iv) Fourth, Performance analysis is done by analyzing the performance metrics like Accuracy, FScore, Recall and Precision.

A. System Architecture

The overall design of the proposed work is shown in Fig. 1

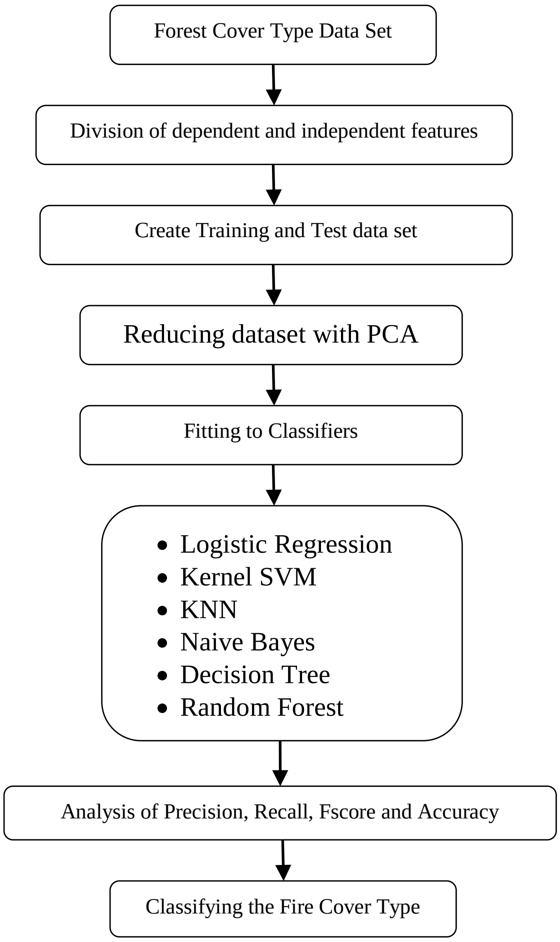

Fig. 1 System Architecture

IV. IMPLEMENTATION AND PERFORMANCE ANALYSIS

A. Data Set Information

The Forest Cover type dataset consist of 54 attributes each of which is contributed towards predicting the type of forest fire. The Forest Cover type dataset consist of “5, 81,012” number of Rows and 54 columns. The Forest Fire Cover Type dataset is downloaded from UCI Data warehouse repository and done with classification analysis. with 53 self-determining attribute and 1 Forest Cover type target attribute. The dataset attributes are as follows,

(1) Elevation Aspect
(2) Slope
(3) Hydrology Horizontal Space
(4) Hydrology Vertical Space
(5) Horizontal_Distance_To_Roadways
(6) Hillshade_9am
(7) Hillshade_Noon
(8) Hillshade_3pm
(9) Horizontal_Distance_To_Fire_Points
(10) Wilderness_Area1
(11) Wilderness_Area2
(12) Wilderness_Area3
(13) Wilderness_Area4
(14) Type1 Soil
(15) Type 2 Soil
(16) Type3 Soil
(17) Type4 Soil
(18) Type5 Soil
(19) Type6 Soil
(20) Type7 Soil
(21) Type8 Soil
(22) Type9 Soil
(23) Type10 Soil
(24) Type11 Soil
(25) Type12 Soil
(26) Type13 Soil
(27) Type14 Soil
(28) Type15 Soil
(29) Type16 Soil
(30) Type17 Soil
(31) Type18 Soil
(32) Type19 Soil
(33) Type20 Soil
(34) Type21 Soil
(35) Type22 Soil
(36) Type23 Soil
(37) Type24 Soil
(38) Type25 Soil
(39) Type26 Soil
(40) Type27 Soil
(41) Type28 Soil
(42) Type29 Soil
(43) Type30 Soil
(44) Type31 Soil
(45) Type32 Soil
(46) Type33 Soil
(47) Type34 Soil
(48) Type35 Soil
(49) Type36 Soil
(50) Type37 Soil
(51) Type38 Soil
(52) Type39 Soil

International Journal of Recent Technology and Engineering (IJRTE)
ISSN: 2277-3878, Volume-8 Issue-3, September 2019

(53) Type40 Soil
(54) Fire Cover_Type (Target- Dependent Attribute)

The target variable Fire Cover_Type may be falling under the category between 1 to 7. The type of fire is given in table. 1.

Table. 1. Fire Cover Type Description in the Dataset

Fire Cover Type	Type Description
1	Trim Fire
2	Shore Pine
3	Spruce Pine
4	Salix Fire
5	Nervous Fire
6	Conifer Fire
7	Meadow Fire

B. Prediction of Fire Type

The Data Set for Forest Cover Type is implemented for analyzing the feature importance from the set of all attributes and is shown in Fig. 2.

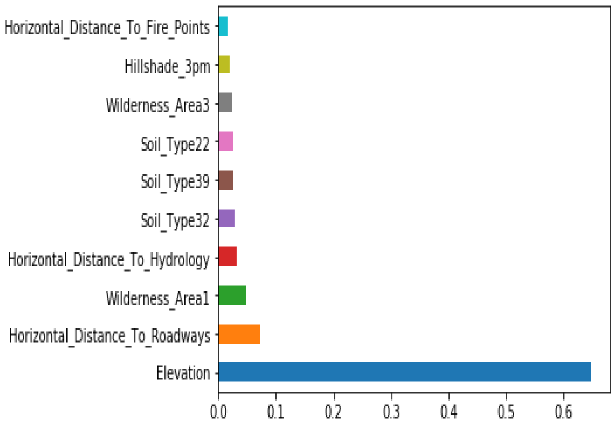

Fig. 2 Feature importance of attributes in the dataset

The raw dataset is applied to all the classification models like Logistic regression, Kernel SVM, KNN, Decision Tree, Random Forest classifier and Naive Bayes. The performance metrics such as FScore, recall and precision is analysed and the results are shown in Table. 2.

Table. 2. Performance measure for the classifiers before subjecting to PCA

Classification	Precision	Recall	FScore
Logistic Regression	0.73	0.73	0.72
Kernel SVM	0.82	0.83	0.82
KNN	0.65	0.65	0.62
DTree	0.88	0.88	0.88
Random Forest	0.92	0.92	0.92
NBayes	0.31	0.19	0.09

The dataset is subjected to principal component analysis with 15 components. The 15 component PCA reduced dataset is applied to all the classification models like Logistic regression, Kernel SVM, KNN, Decision Tree, Naive Bayes and Random Forest. The performance metrics such as precision, recall and FScore is analysed and the results are shown in Table. 3.

Table. 3. Performance measure for the classifiers after subjecting to PCA with 15 components

Classification	PCA No. of. Components = 15		
	Precision	Recall	FScore
Log Regression	0.48	0.49	0.48
Kernel SVM	0.51	0.48	0.49
KNN	0.52	0.53	0.51
DTree	0.44	0.41	0.41
RForest	0.39	0.32	0.32
Naive Bayes	0.45	0.47	0.45

The dataset is subjected to principal component analysis with 20 components. The 20 component PCA reduced dataset is pertained to all the classification models like Logistic regression, Kernel SVM, KNN, Decision Tree, Naive Bayes and Random Forest. The performance metrics such as precision, recall and FScore is analysed and the results are shown in Table. 4.

Table. 4. . Performance measure for the classifiers after subjecting to PCA with 20 components

Classification	PCA Number of Components = 20		
	Precision	Recall	FScore
Log Regression	0.42	0.45	0.42
Kernel SVM	0.53	0.52	0.52
KNN	0.49	0.52	0.49
DTree	0.46	0.42	0.42
RForest	0.49	0.41	0.40
Naive Bayes	0.56	0.54	0.53

The dataset is subjected to principal component analysis with 25 components. The 25 component PCA reduced dataset is applied to all the classification models like Logistic regression, Kernel SVM, KNN, Decision Tree, Naive Bayes and Random Forest. The performance metrics such as precision, recall and FScore is analysed and the results are shown in Table. 5.

Table. 5. Performance measure for the classifiers after PCA with 25 components

Classification	PCA Number of Components = 25		
	Precision	Recall	FScore
Logistic Regression	0.36	0.35	0.35
Kernel SVM	0.47	0.42	0.43
KNN	0.43	0.45	0.42
DTree	0.42	0.38	0.38
RForest	0.37	0.32	0.31
N Bayes	0.48	0.48	0.45

The performance metric accuracy is analyzed for the dataset before and after applying principal component analysis and is shown in Table. 6.

Table. 6. Comparison of Accuracy for all Classification before and after applying PCA

Classification	Accuracy			
	No PCA	15PCA	20PCA	25PCA
Log Reg	0.73	0.49	0.45	0.35
KSVM	0.83	0.48	0.52	0.42
KNN	0.65	0.53	0.52	0.45
DTree	0.88	0.40	0.42	0.38
RForest	0.92	0.78	0.83	0.89
NBayes	0.19	0.47	0.54	0.48

The Performance Analysis of all the metrics for all the classification algorithms before and after applying principal component analysis is shown from figure. 3.to Figure. 7.

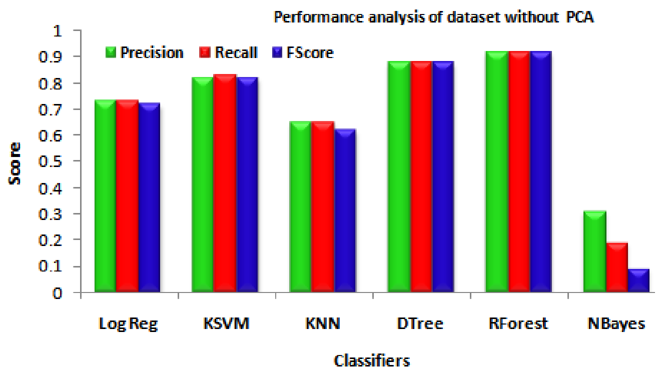

Fig. 5 Performance Analysis without PCA

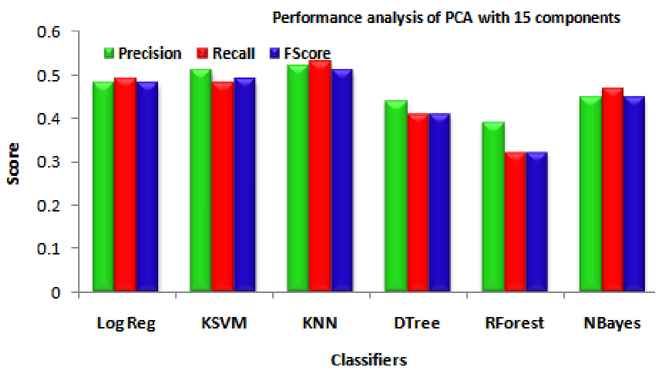

Fig. 5 PCA Performance Analysis with 15 components

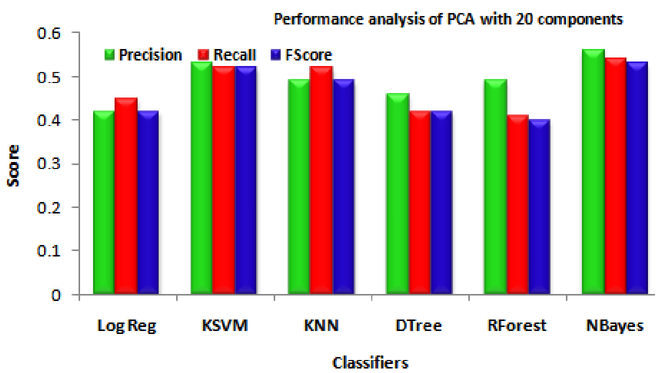

Fig. 5 PCA Performance Analysis with 20 components

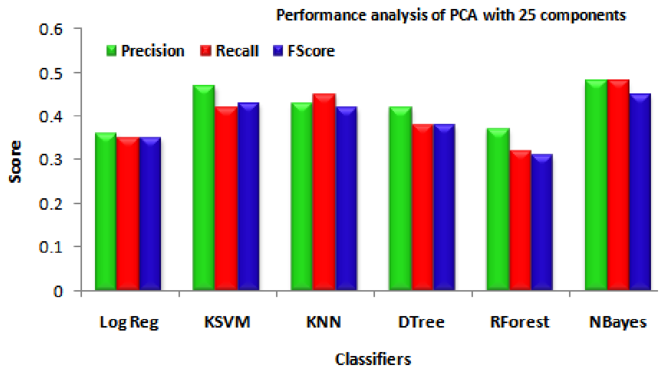

Fig. 5 Performance Analysis for PCA with 25 components

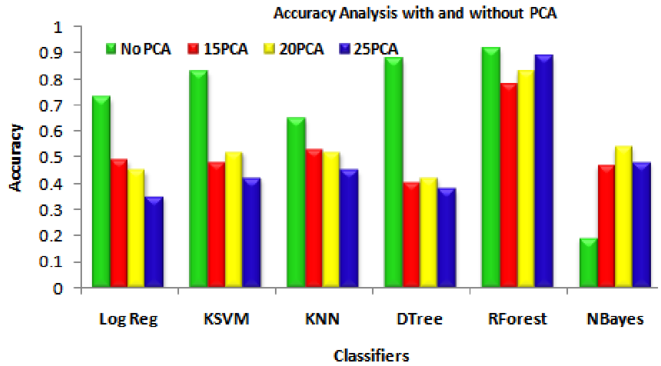

Fig. 5 Accuracy Analysis before and after applying PCA

V. CONCLUSION

This paper addresses to detect the type of fire in the forest area by analyzing the different parameters. The important feature attributes are found and depicted as a chart. The raw dataset is applied to all the classifiers like Logistic regression, KSVM, KNN, Decision Tree, NBayes and RForest. The dataset is reduced with PCA and then the reduced dataset is fitted to all the classifiers. The Performance analysis is implemented and compared with the performance measures like FScore, Recall, Accuracy and Precision. Experimental Result shows that the Random Forest classifier is obtained with the accuracy of 92% before applying PCA. After applying PCA, the Rforest is analyzed to be performing ideally having the accuracy of 78% for 15 components, 83% for 20 components and 89% for 25 components.

REFERENCES

1. 1.G. D. Li et al., "Forest Cover Types Classification Based on Online Machine Learning on Distributed Cloud Computing Platforms of Storm and SAMOA", Advanced Materials Research., Vol. 955-959, 2014, pp. 3803-3812.
2. H. Franco-Lopez, A. R. Ek and M. E. Bauer, "Estimation and mapping of forest stand density volume and cover type using the k-nearest neighbors method", Remote Sensing of Environment, vol. 77, no. 3, 2001, pp. 251-274.
3. Zhu Z and D. L. .Evans, “US forest types and predicted percent forest cover from AVHRR data” Photogrammetric Engineering and Remote Sensing., Vol. 60(5), 1994, pp. 525-532.
4. B. T. Wilson, A. J. Lister and R. I. Riemann, "A nearest-neighbor imputation approach to mapping tree species over large areas using forest inventory plots and moderate resolution raster data", Forest Ecology and Management, vol. 271, 2012, pp. 182-198.
5. K. Gjertsen, "Accuracy of forest mapping based on landsat TM data and a knn-based method", Remote Sensing of Environment, vol. 110, no. 4, 2007, pp. 420-430.

6. 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.
7. SSuguna 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.
8. 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.
9. 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.
10. 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, 30 August 2019. (Accepted for Publication)
11. 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, 30 August 2019. (Accepted for Publication)
12. 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.
13. hyamala 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.
14. 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.
15. 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, 30 August 2019. (Accepted for Publication)