FIRE PREDICTION ANALYSIS BASED ON ENSEMBLE MACHINE LEARNING ALGORITHMS

Abstract:

A fire accident is the most tragic incident in human life. Particularly environmental hazards such as forest fireslead loss of wildlife, economy, wealth, human lives and pollution. our research purpose of predict the occurrence of fire incidents using ensemble machine learning models.The goal is to develop an accurate and reliable model that can forecast the occurrence of forest fires based on various environmental factors. The best performance is obtained by the ensemble machine learning model for this work. Comparative study of individual model and ensemble model. If you check all models Decision tree predicts 75.4%, the Random Forest tree predicts 83.2%, the Support Vector Machine predicts 71.8%, and the K nearest neighbour predicts 82.1%. Ensemble models with two combinations of decision tree and random forest tree predicts accuracy is 80.8%. Support vector machine and KNN predicts the accuracy rate is 73.4%. Compared to the ensemble learning model, the individual model predicts more accuracy.

Keywords: Fire Prediction Analysis, Hybrid Machine Learning Models, accuracy

I. INTRODUCTION

Forest fires are the most important environmental and social issues causing huge damage, wildlife loss and human life loss [1]. Forest fire prediction is a good lesson for taking precautions against forest fires in future. Number of fire detection systems available for every strategy. The affected locations were estimated with the support of satellite images [2].

Figure 1: Sample file

Forest fires most frequently occurring disasters in the current time. One important reason for fire occurrence in forests is global warming due to the temperature of the earth. Some other reasons like human negligence, lightning and thunderstorms [3]. Forestfires can lead to deforestation, which negative impact on human society. It is reported that every year lakhs of hectaresare destroyed. Forest fires combine with weather conditions, dryness of flame items and terrain [4]. Few forest authorities use human observers as detectors and reporters of forest fires. Fire accident is the most tragic incident in human life [5]. Particularly environmental hazards such as forest fireslead loss of wildlife, economy, wealth, human lives and pollution. Fire prediction risks everything in its vicinity. Preliminary detection of such fires can help to control the blowout and protect nearby locations from loss [6].

The following paper continues with the proposed system and architecture in section two. Section three discusses with results analysis and comparative study. The final section concludes the paper

II. PROPOSED SYSTEM AND ARCHITECTURE

2.1 Proposed System

Environmental hazards such as forest fires lead loss of wildlife, economy, wealth, human lives and pollution. Fire prediction risks everything in its vicinity. our research purpose of predict the occurrence of fire incidents using ensemble machine learning models [16]. The best performance is obtained by the ensemble machine learning model for this work. This research proposesto predict the incidence of fires using ensemble machine learning models. If you check all models Decision tree, the Random Forest tree, the Support Vector Machine, and the K nearest neighbour. Our research proposed Ensemble models with two combinations,(i) a decision tree and random forest tree and (ii) a Support vector machine and KNN [7].

2.1.1 Data integration

Data integration is finished to make the data into an entire file. Hence, it is required to mix the data into a file [14].

2.1.2 Data cleaning

Data cleaning refers to the discovery of imperfect, incorrect, imprecise data components. We simply alter the improper format in weather data to carry out the accurate analysis [8].

2.1.3 Data Reduction

The data reduction reduces the raw data into a more useful format. But weather data already include useful data for analysis [15].

2.1.4 Data transformation

Data transformation for altering the scale of measurement of unique data into other forms so that the analysis can read weather data [9].

2.2 Classification model

Numerous models of forest fire prediction using machine learning have been developed. Machine learning integrates informatics and statistical analysis to progress prediction, hence extensively used to resolve uncertainty issues [10].

2.3. Testing and evaluation

For classification modelling, each experiment was performed using the data set split training and test data. The parameters to test every modeming's output are as follows: accuracy Root Mean Square Error and confusion matrix [11].

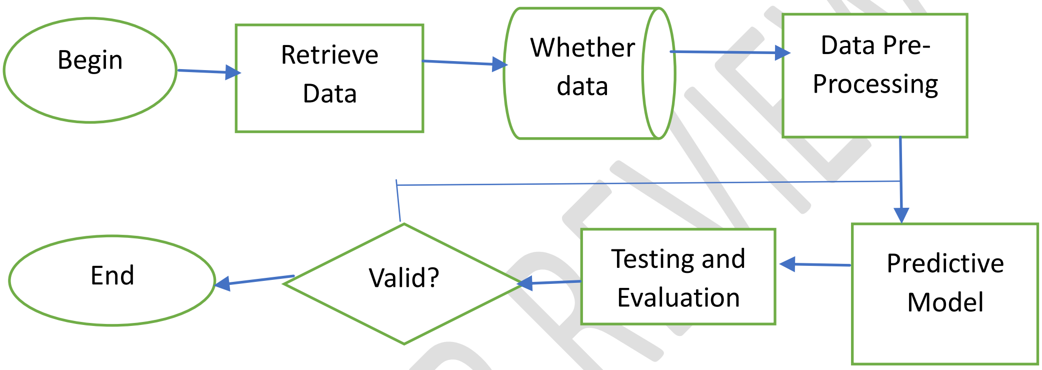


Figure 2: The architecture of Fire prediction using ML models

Figure 2 demonstrates the architecture of the projected structure from start to finish of the paper. So many phases from input to output as a procedure of machine learning[12]. The forest fires most frequently occurring disasters in the current time. One important reason for fire occurrence in forests is global warming due to the temperature of the earth—some other reasons like human negligence, lightning and thunderstorms[13].

III. RESULTS AND ANALYSIS

3.1 Dataset Description

The goal is to develop an accurate and reliable model that can forecast the occurrence of forest fires based on various environmental factors. Upload a dataset of 55 attributes in the current browser session.

Table 1: dataset

	Elevation	Aspect	Slope	Horizontal_Distance_To_Hydrology	Vertical_Distance_To_Hydrology	Horizontal_Distance_To_Roadways	Hillshade_9am	Hillshade_Noon	Hillshade_3pm	Horizontal_Distance_To_Fire_Points
0	2596	51	3	258	0	510	221	232	148	6279
1	2590	56	2	212	-6	390	220	235	151	6225
2	2804	139	9	268	65	3180	234	238	135	6121
3	2785	155	18	242	118	3090	238	238	122	6211
4	2595	45	2	153	-1	391	220	234	150	6172

Soil_Type32	Soil_Type33	Soil_Type34	Soil_Type35	Soil_Type36	Soil_Type37	Soil_Type38	Soil_Type39	Soil_Type40	Cover_Type
0	0	0	0	0	0	0	0	0	5
0	0	0	0	0	0	0	0	0	5
0	0	0	0	0	0	0	0	0	2
0	0	0	0	0	0	0	0	0	2
0	0	0	0	0	0	0	0	0	5

3.2 Data Preprocessing

Check missing values and fill those values using dissimilar methods otherwise ignore those values. Remove abnormal values also. Table 2 data represents the after-pre-processing dataset.

Table 2: Dataset after pre-processing

	Elevation	Aspect	Slope	Horizontal_Distance_To_Hydrology	Vertical_Distance_To_Hydrology	Horizontal_Distance_To_Roadways	Hillshade_9am	Hillshade_Noon	Hillshade_3pm	Horizontal_Distance_To_Fire_Points
count	2000.000000	2000.000000	2000.000000	2000.000000	2000.000000	2000.000000	2000.00000	2000.000000	2000.000000	2000.000000
mean	2894.813000	133.926000	13.902000	222.661500	38.874000	2743.085500	219.25700	217.723000	128.357500	2470.935000
std	233.116052	105.504944	8.484185	172.381674	49.748443	1929.001563	22.37459	23.644509	43.823555	1518.617143
min	2000.000000	0.000000	0.000000	0.000000	-134.000000	67.000000	100.00000	99.000000	0.000000	60.000000
25%	2744.000000	57.000000	7.000000	90.000000	5.000000	849.000000	209.00000	208.000000	107.000000	1489.000000
50%	2907.000000	90.000000	12.000000	190.000000	23.000000	2758.000000	224.00000	223.000000	134.000000	2140.000000
75%	3052.500000	198.500000	18.000000	319.000000	60.000000	4542.750000	234.00000	234.000000	156.000000	2919.500000
max	3404.000000	359.000000	49.000000	997.000000	554.000000	6890.000000	254.00000	254.000000	246.000000	6853.000000

8 rows × 55 columns

Soil_Type32	Soil_Type33	Soil_Type34	Soil_Type35	Soil_Type36	Soil_Type37	Soil_Type38	Soil_Type39	Soil_Type40	Cover_Type
2000.000000	2000.000000	2000.0	2000.0	2000.0	2000.0	2000.000000	2000.00000	2000.000000	2000.000000
0.005500	0.018000	0.0	0.0	0.0	0.0	0.017500	0.01250	0.003500	2.843500
0.073976	0.132984	0.0	0.0	0.0	0.0	0.131158	0.11113	0.059072	1.803783
0.000000	0.000000	0.0	0.0	0.0	0.0	0.000000	0.00000	0.000000	1.000000
0.000000	0.000000	0.0	0.0	0.0	0.0	0.000000	0.00000	0.000000	2.000000
0.000000	0.000000	0.0	0.0	0.0	0.0	0.000000	0.00000	0.000000	2.000000
0.000000	0.000000	0.0	0.0	0.0	0.0	0.000000	0.00000	0.000000	5.000000
1.000000	1.000000	0.0	0.0	0.0	0.0	1.000000	1.00000	1.000000	7.000000

After pre-processing complete the list of the variables with the count.

Table 3: Variable count

```
2       827
5       518
1       488
7        86
3        43
6        36
4         2
Name: Cover_Type, dtype: int64
```

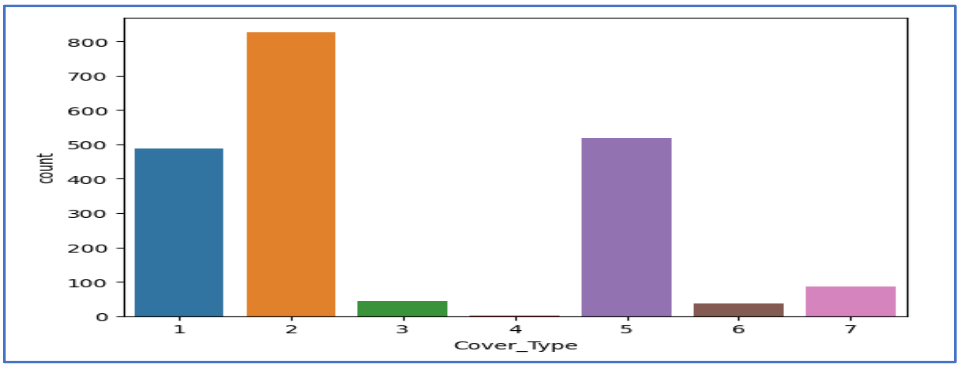


Figure 3: Chart for Variable count

Figure 3 shows the sample graph for the variable count. It also shows the count in table 3.

3.3 Visualization

Based on the given dataset after pre-processing of data, it can be shown in different graphs with multiple time slots with fire points. The following figure 4 shows the different time slots of fire points.

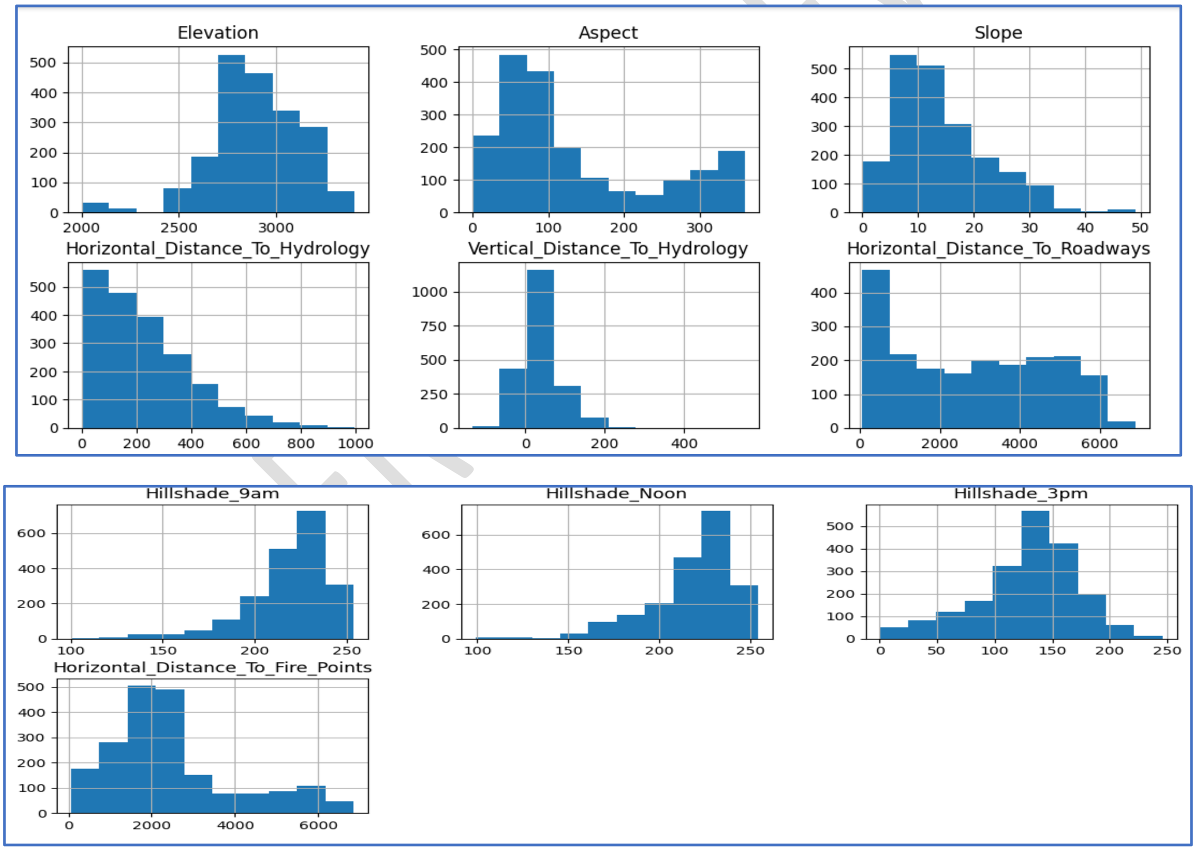
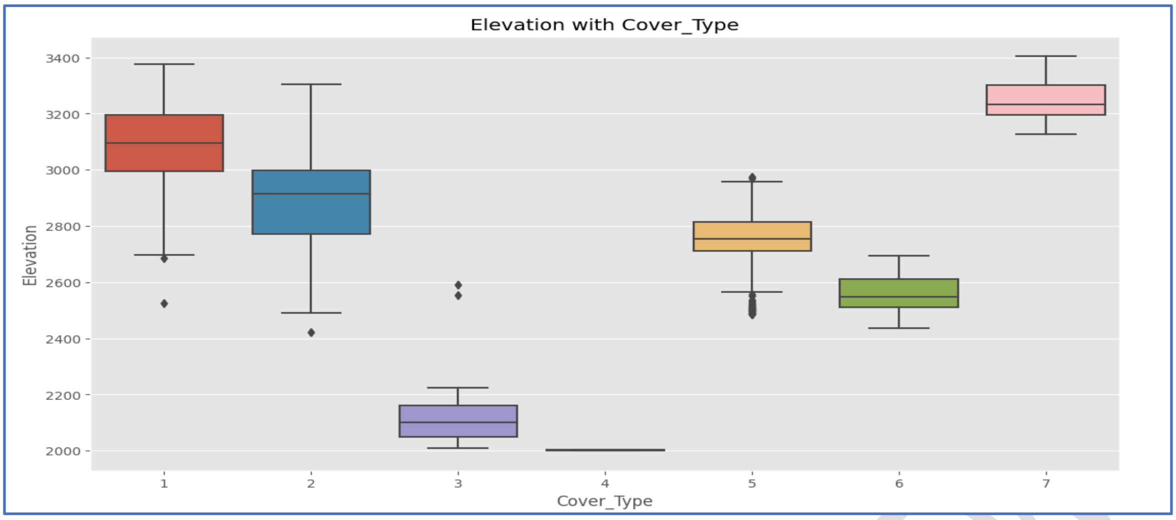


Figure 4: Data visualization @ multiple time slots with different fire points

The following Figure 3 shows the detailed visualization of the given dataset with corresponding Figure 4.



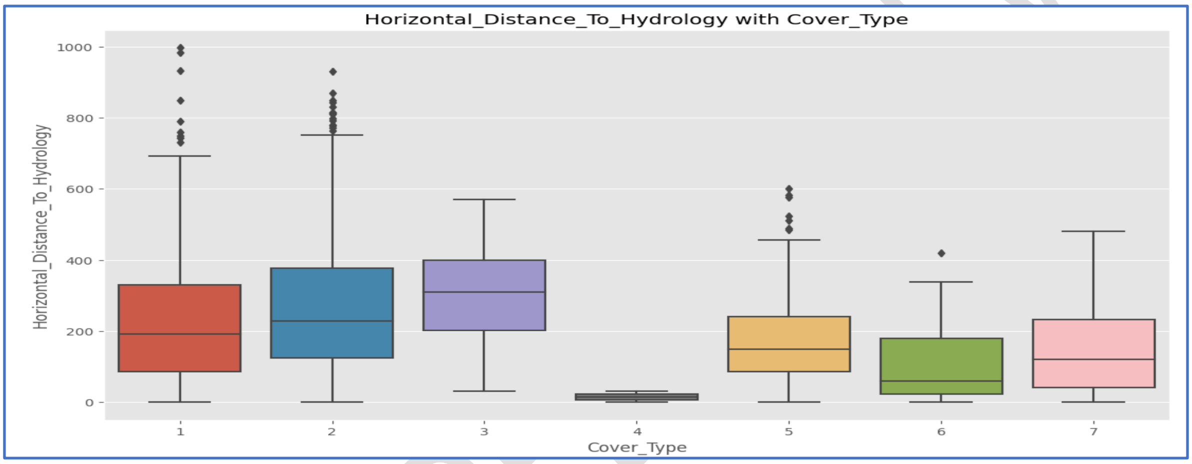


Figure 5: Detaileddata visualization

Figure 5 shows the detailed data visualization of every attribute. Few samples only describe in this paper. Figure 6 describes the Heatmaps are used in numerous forms of analytics but are most normally used to demonstrate models.

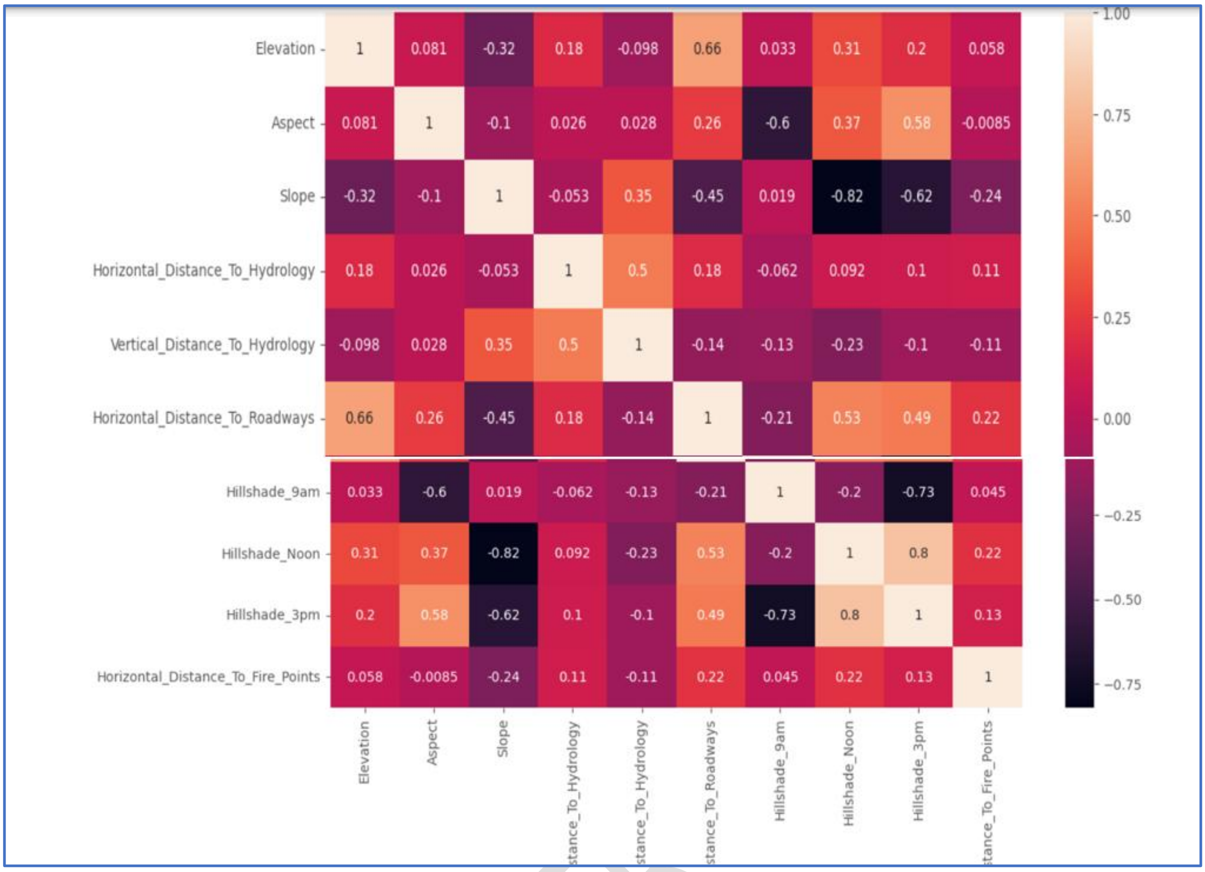


Figure 6: Heatmap for correlation of dataset

3.4 Feature Selection

In feature selection, initially separate the features and target. After that reduce the feature using dimensionality reduction algorithms and then shape the novel features. Finally split the data into test and train.

3.5 Ensemble Modelling

In past most of the research did the individual model of the data. Now we propose an ensemble model that means a combination of two or more models. Predict the results using these hybrid models. The following description for different models.

3.5.1 Decision Tree

The decision tree predicts the accuracy rate is 75.4 per cent. It shows in the below box.

```
[ ]  from sklearn.tree import DecisionTreeClassifier
     dt = DecisionTreeClassifier()

     dt.fit(X_train, y_train)

     y_pred = dt.predict(X_test)
     print("Accuracy -- ", dt.score(X_test, y_test)*100)

     Accuracy --  75.4
```

3.5.2 Random Forest

```
[ ]  #Random Forest
     from sklearn.ensemble import RandomForestClassifier
     rf = RandomForestClassifier(n_estimators=100)

     #fit
     rf.fit(X_train, y_train)

     #prediction
     y_pred = rf.predict(X_test)

     #score
     print("Accuracy -- ", rf.score(X_test, y_test)*100)

     Accuracy --  83.2
```

Random forest tree predicts the accuracy rate is 83.2 per cent.

3.5.3 Decision tree and Random Forest tree (Ensemble)

We propose ensemble learning with the combination of a decision tree and a Random Forest tree. This model predicts the accuracy rate is 80.8 per cent. This ensemble model performance is better than the decision tree and low performance compared to the random forest tree. The ensemble model is not good compared to random forest tree performance because Random Forest trees give more performance at 83.2%. In this comparison,the Individual model is better than the ensemble model.

```
from sklearn.ensemble import VotingClassifier

ensemble_model1 = VotingClassifier(estimators=[('decision_tree', dt), ('random_forest', rf)], voting='hard')

#fit
ensemble_model1.fit(X_train, y_train)

#prediction
y_pred = ensemble_model1.predict(X_test)

#score
print("Accuracy -- ", ensemble_model1.score(X_test, y_test)*100)

#confusion
cm = confusion_matrix(y_pred, y_test)
plt.figure(figsize=(10, 8))
sb.set(font_scale=1.2)
sb.heatmap(cm, annot=True, fmt='g')
plt.show()

Accuracy --  80.80000000000001
```

In predictive analytics of figure 7, a table of confusion is a table with two rows and two columns that reports the number of true positives, false negatives, false positives, and true negatives.

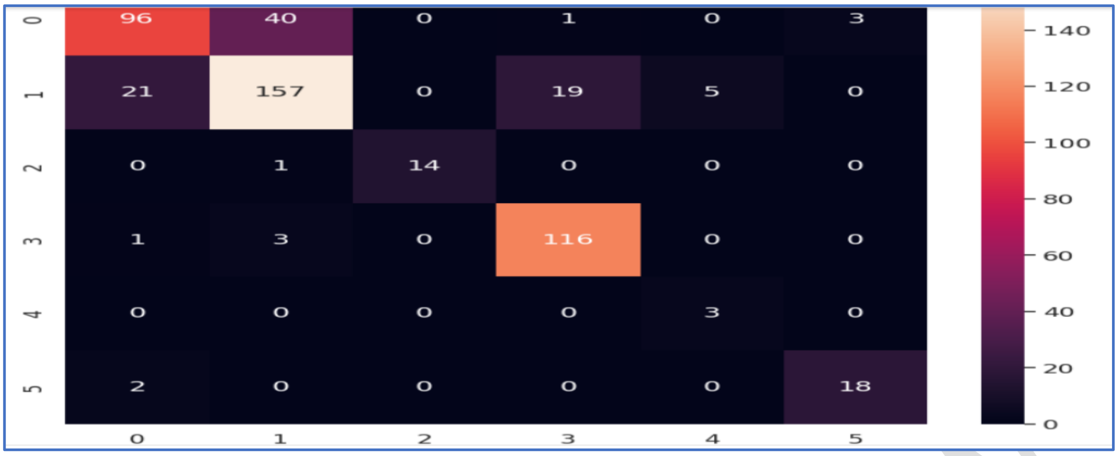


Figure 7: confusion matrix for ensemble learning (DT+RF)

5.6Ensemble model

5.6.1 Support Vector Machine

The support vector machine predicts the accuracy rate is 71.8%.

```
from sklearn.svm import SVC

svm_classifier = SVC(kernel='linear', C=1.0, random_state=42)

#fit
svm_classifier.fit(X_train, y_train)

#prediction
y_pred = svm_classifier.predict(X_test)

#score
print("Accuracy -- ", svm_classifier.score(X_test, y_test)*100)

Accuracy --  71.8
```

5.6.2 K Nearest Neighbour

The KNN model predicts the accuracy rate is 82.1%.

```
[ ]  from sklearn.neighbors import KNeighborsClassifier

     # Create a KNN classifier
     knn_classifier = KNeighborsClassifier(n_neighbors=3)

     #fit
     knn_classifier.fit(X_train, y_train)

     #prediction
     y_pred = knn_classifier.predict(X_test)

     #score
     print("Accuracy -- ", knn_classifier.score(X_test, y_test)*100)

     Accuracy --  82.19999999999999
```

6.2 Ensemble Model SVM and KNN

The ensemble model predicts the accuracy rate is 73.4%. This model predicts less accuracy compared to the individual model of machine learning. KNN model predicts 82.1% accuracy, it is better performance compared to ensemble learning (SVM+KNN). In predictive analytics of Figure 8, a table of confusion is a table with two rows and two columns that reports the number of true positives, false negatives, false positives, and true negatives.

```
from sklearn.ensemble import VotingClassifier

ensemble_model2 = VotingClassifier(estimators=[('svm', svm_classifier), ('knn', knn_classifier)], voting='hard')

#fit
ensemble_model2.fit(X_train, y_train)

#prediction
y_pred = ensemble_model2.predict(X_test)

#score
print("Accuracy -- ", ensemble_model2.score(X_test, y_test)*100)

#confusion
cm = confusion_matrix(y_pred, y_test)
plt.figure(figsize=(10, 8))
sb.set(font_scale=1.2)
sb.heatmap(cm, annot=True, fmt='g')
plt.show()

Accuracy --  73.4
```

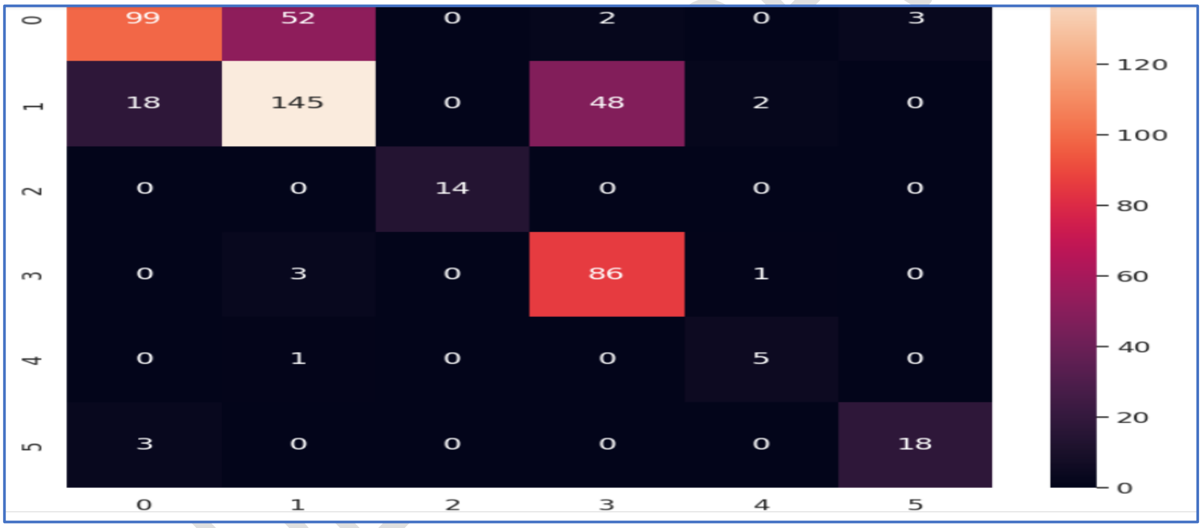


Figure 8: confusion matrix for ensemble learning (SVM+KNN)

5.7 Comparative study

The following Table4is for a comparative study of the individual model and ensemble model. If you check all models Decision tree predicts 75.4%, the Random Forest tree predicts 83.2%, the Support Vector Machine predicts 71.8%, and the K nearest neighbour predicts 82.1%. Ensemble models with two combinations of decision tree and random forest tree predicts accuracy is 80.8%. Support vector machine and KNN predicts the accuracy rate is 73.4%. Compared to the ensemble learning model, the individual model predicts more accuracy.

Table 4: Comparative study of models

S. No.	Model Name	Accuracy Rate (%)
1	Decision tree	75.4
2	**Random forest tree**	**83.2**
3	SVM	71.8
4	**KNN**	**82.1**
5	DT+RF	80.8
6	SVM+KNN	73.4

IV. CONCLUSION

Our research purposes of predicting the occurrence of fire incidents using ensemble machine learning models. The best performance is gotten by the ensemble machine learning model for this work. The statistical report shows that, all models Decision tree predicts 75.4%, Random Forest tree predicts 83.2%, Support Vector machine predicts 71.8%, and K nearest neighbour predicts 82.1%. Ensemble models with two combinations of decision tree and random forest tree predicts accuracy is 80.8%. Support vector machine and KNN predicts the accuracy rate is 73.4%. Compare to ensemble learning model, individual model predicts more accuracy.

References

[1] A. Karouni et al., "Applying decision tree algorithm and neural networks to predict forest" fires in Lebanon", *J. Theor. Appl. Inf. Technol.*, vol. 63, no. 2, pp. 282-291, 2014.

[2] T. Preeti et al., "Forest Fire Prediction Using Machine Learning Techniques", *2021 International Conference on Intelligent Technologies (CONIT)*, pp. 1-6, 2021.

[3] F. Abid and N. Izeboudjen, "Predicting Forest Fire in Algeria Using Data Mining Techniques: Case Study of the Decision Tree Algorithm" in Advanced Intelligent Systems for Sustainable Development (AI2SD'2019). AI2SD 2019. Advances in Intelligent Systems and Computing, Cham: Springer, vol. 1105, 2020.

[4] Dieu Tien Bui, Hung Van Le and Nhat-Duc Hoang, "GIS-based spatial prediction of tropical forest fire danger using a new hybrid machine learning method", *Ecological Informatics*, 2018.

[5] B. S. Negara, R. Kurniawan, M. Z. A. Nazri, S. N. H. S. Abdullah, R. W. Saputra and A. Ismanto, "Riau Forest Fire Prediction using Supervised Machine Learning", *Journal of Physics Conference Series*, vol. 1566, no. 1, 2020.

[6] Binh Pham, Abolfazl Jaafari, Mohammad Taghi Avand, Nadhir Al-Ansari, Tran Du, Hoang Yen, et al., "Performance Evaluation of Machine Learning Methods for Forest Fire Modelling and Prediction", *Symmetry.*, 2020.

[7] Subhani Shaik and Dr. Uppu Ravibabu "Classification of EMG Signal Analysis based on Curvelet Transform and Random Forest tree Method" Paper selected for Journal of Theoretical and Applied Information Technology (JATIT), Vol. 95, December 2017.

[8] Ch. Shravya, Pravallika and Subhani Shaik," Brest cancer prediction using machine learning Techniques", International Journal of Innovative Technology and Exploring Engineering, Vol. 8, Issue 6, 2019.

[9] Shiva Keertan J and Subhani Shaik," Machine Learning Algorithms for Oil Price Prediction", International Journal of Innovative Technology and Exploring Engineering, Volume-8 Issue-8, 2019.

[10] K P Surya Teja, Vigneswaran Reddy and Subhani Shaik,"Flight Delay Prediction Using Machine Learning Algorithm XGBoost",Jour of Adv Research in Dynamical & Control Systems, Vol. 11, No. 5, 2019.

[11] Mr. Sujan Reddy, Ms. Renu Sri andSubhani Shaik," Sentimental Analysis using Logistic

Regression",International Journal of Engineering Research and Applications (IJERA), Vol.11, Series-2, July 2021.
[12] Ms. Mamatha, Srinivasa Datta and Subhani Shaik," Fake Profile Identification using Machine Learning Algorithms", International Journal of Engineering Research and Applications (IJERA), Vol.11, Series-2, July-2021.
[13] R. Vijaya Kumar Reddy, Subhani Shaik,B. Srinivasa Rao, "Machine learning based outlier detection for medical data" Indonesian Journal of Electrical Engineering and Computer Science, Vol. 24, No. 1, October 2021.
[14] Eskandari, S., Amiri, M., Sãdhasivam, N., & Pourghasemi, H. R. (2020). Comparison of new individual and hybrid machine learning algorithms for modelling and mapping fire hazard: a supplementary analysis of fire hazard in different counties of Golestan Province in Iran. *Natural Hazards*, *104*, 305-327.
[15] Abid, F. (2021). A survey of machine learning algorithms based forest fires prediction and detection systems. *Fire Technology*, *57*(2), 559-590.
[16] Jana, S., & Shome, S. K. (2023). Hybrid ensemble-based machine learning for smart building fire detection using multi-modal sensor data. *Fire Technology*, *59*(2), 473-496.