International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395 -0056
Volume: 02 Issue: 04 | July-2015 www.irjet.net p-ISSN: 2395-0072

Envisagation and Analysis of Air Pollution Caused by Forest fire using Machine Learning Algorithm

Divya TL[1], Dr.Vijayalakshmi MN[2]

[1]VTU Research Scholar, Department of MCA, RVCE, Karnataka, India
[2] Associate Professor, Department of MCA, RVCE, Karnataka, India

---

Abstract- *Forest fire is a natural hazard which plays an important role in causing air pollution. The fire affected tree liberates gas (smoke) which increases the air pollutants. The forest fire liberates the pollutants such as Carbon Monoxide (CO), Sulphur dioxide (SO_2), Nitrogen oxide (NO) and soon. The contribution of forest fire pollutant emissions to the ecosystem can be calculated as Particulate Matter (PM). Air pollution levels in forest regions will spread to kilometers away from the fire source. This paper aims for analysis of air quality indices from Good to Very poor quality air in fire affected place. J48 classification algorithm is used for air analysis.*

Key Words: *Forest fire, Natural hazard, air pollution, air pollutants, Carbon Monoxide (CO), Sulphur dioxide (SO_2), Nitrogen oxide (NO), J48 classification algorithm*

1. INTRODUCTION

Forest fire emits significant amount of gas/smoke that increases Particular Matter (PM) in the atmosphere. Every year due to forest fire 2% of forest is under degradation, the smoke/gas liberated will spread to more than ten kilometers from fire source. The major pollutants liberated are nitrogen oxides, sulphur oxides, carbon monoxides, ozone and these organic pollutants easily dissolve in the atmosphere [2].The statistical information of Indian forest fire with chemical is given in Table1.

Table-1. Observation of fire occurrences in different states using Advanced Along Track Scanning Radiometer (AATSR) night-time data between 2000 and 2010 for the months of January to June [6] in Ha.

Sl. No	State	Total Number of Forest fires	Total Area Burnt (KM^2)	Carbon pollutants(CO) mg/m^3	Nitrogen Pollutants (NO_2) $\mu g/m^3$	Ozone (O_3) $\mu g/m^3$	Sulphur Pollutants (SO_2) $\mu g/m^3$	Total Particulate Matter (PM) $\mu g/m^3$
1	Madhya Pradesh	910	550	10-12	301-400	225 - 250	251-350	4.5
2	Chhattisgarh	759	430	8-9	201-300	200 - 224	200-250	5
3	Maharashtra	739	419	8-9	201-300	200 - 224	200-250	6
4	Karnataka	57	300	1-3	1-39	0-35	1-40	5.5
5	Kerala	41	298	1-3	1-39	0-35	1-40	5
6	Andhra Pradesh	283	368	4-5	40-200	125 - 180	50-100	5

In Table 1 the forest fire emissions were calculated to find the increase in average Particulate Matter (PM) concentration in atmosphere. The increase in PM concentration will cause respiratory problems for wild animals and nearby village people.

2. RELATED WORK

Forest fire emissions are important for local air pollution levels .Forest fires contributes 0.2% of Nitrogen dioxide (NO_2), 0.5% of Carbon dioxide (CO_2),1.2% of CO. Each time the chemicals which are introduced by fire will increase the Particulate Matter more than 2.5[1].

Forest fires will affect the chemical properties of the atmosphere, by the release of significant amounts of particulate matter, which interact with solar radiation. Biomass burning is responsible for as much as 45% of the emissions of carbon compounds with larger scale. The release of Nitrogen compounds also plays a role for air pollution [2].

Data mining is domain which helps in Prediction and cause effects of patterns. Artificial neural networks with back propagation learning rule and k- nearest neighbor classifiers were used, in order to predict future peaks of carbon monoxide [3]. Athanasiadis, Karatzas and Mitch's made effort to analyze Ozone in fire area [4].

K-Nearest Neighbor is used for the classification of air quality. An adaptive method of nearest neighbor classification (DANN) was analyzed by Hastie and Tibshirani for air pollution data analysis [6]. The probabilistic approach can give increase in Sulphur dioxide (SO_2) concentration in the atmosphere [7]. Decision trees helps in the prediction of cause effects in health dataset, meteorological data set and pollution data set. They provide the target classes by divide and conquer method [8].Here an effort is made to identify different pollutants responsible for air pollution with J48 Machine learning algorithm.

3. EXPERIMENT WITH RESULTS

Classification is a process of constructing a model of available target classes from asset of records. The algorithm generates rule for the prediction of target classes. J48 is the Java implementation of Classification 4.5(C4.5) algorithm.

The summarized data during forest fire is monitored and collected in the data base. The time of fire is monitored along with vegetation type and duration of fire. The duration helps in identification of different pollutants concentration in air. During the analysis, quantified gases are CO mg/m3, NO_2ug/m3, O_3ug/m3, SO2ug/m3and Particulate Matter concentration. The values of pollutants are represented in Table 2.

Table-2: The general air quality index takes the higher value of the pollutants indexes

Carbon pollutants (CO) mg/m^3	Nitrogen Pollutants(NO_2) $\mu g/m^3$	Ozone (O_3) $\mu g/m^3$	Sulphur Pollutants(SO_2) $\mu g/m^3$	Total Particulate Matter(PM) $\mu g/m^3$	Target class For Air quality Index
0-2	0-49	0-35	0-45	<=2.5	Good
3-6	50-125	36-80	46-100	3-5	Medium
6-9	126-250	81-150	101-150	6-9	Poor
>=10	>250	>150	>150	>=10	Very Poor

Table 2. Shows Air quality index categorization .Air quality is categorized into Good when pollutants concentration gives PM<2.5, Medium when PM is between 3 to 5,Poor When PM between 6 to 9.Very poor air quality when PM>=10.

Table-3: Sample data of Forest fire Smoke Concentration

Duration of fire (Hours)	Type of vegetation	Temp (Celsius)	CO mg/m^3	O_3 µg/m^3	NO_2 µg/m^3	SO_2 µg/m^3	PM µg/m^3	Target class
4	Herbs	36	5	48	37	67	2.5	good
2	Sherbs	34	7	59	38	69	6	medium
5	Pine	37	4	68	50	128	7	poor
6	Herbs	35	6	80	108	130	5	medium
6	Pine	30	8	126	143	150	6	poor
7	herbs	30	7	139	200	168	7	poor
5	Sherbs	28	6	178	220	199	9.5	very poor
8	Pine	29	7	250	201	200	10	very poor
5	herbs	32	4	43	35	65	3	good
8	pine	36	6	109	152	200	5	medium
4	Herbs	36	5	48	37	67	2.5	good

Table- 3, Shows the sample data set of Carbon, Nitrogen, Sulphur Particulate Matter concentrations with target air quality indices.

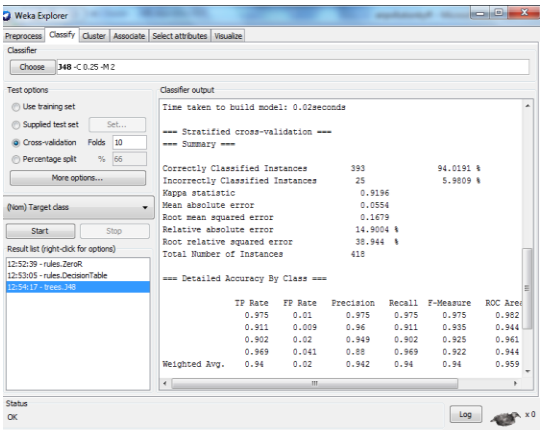

Fig -2: Cross validation of J48 Algorithm.

Figure 2 shows total instances are 418.In that 393 correctly classified instances and 25 incorrectly classified instances. The mean absolute error 0.06.Kappa statistics is measured for agreement between predicted and observed categories of air pollutant data set is agreeable that is 0.91.

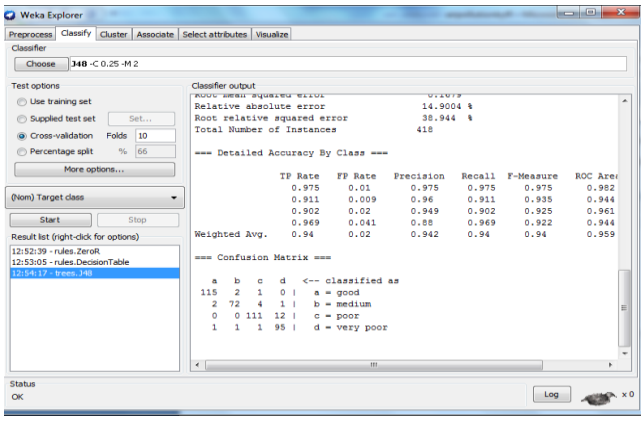

Fig-3: Cross validation with Confusion matrix of data set.

Figure 3 shows the obtained true and false classification of data with Confusion matrix**.** From the matrix 115 records are classified to good quality air , 72 records are classified as medium , 111 records are poor and 95 records as very poor quality of air.

Table-4: Confusion Matrix Air Quality index classification during forest fire.

Classification	Good(a)	Medium(b)	Poor(c)	Very poor(d)
Good(a)	115	2	1	0
Medium(b)	2	72	4	1
Poor(c)	0	0	111	12
Very poor(d)	1	1	1	95

Table 4 Shows confusion matrix for air quality classification. The data set has 418 instances. The main diagonal values shows 115+ 72+111+95=393 are correct classified data. Hence the success rate for classification of data is 393/418=0.94that is 94%.

Table -5: Accuracy rate for the J48 algorithm

Detailed accuracy	TP Rate	FP Rate	Precision	Recall
Good(a)	0.97	0.01	0.975	0.975
Medium(b)	0.912	0.09	0.91	0.93
Poor(c)	0.906	0.02	0.902	0.92
Very poor(d)	0.96	0.4	0.96	0.92
Weighted average index	0.94	0.02	0.94	0.94

Table5. Gives the result data for classification of Air quality at place of forest fire. Precision and Recall are the measures of performance analysis. Higher Precision means less number of misclassification of data. Precision 0(Zero) means the data set is completely misclassified. Here the weighted average Precision measure is 0.94 which is close to 1.Recall shows complete classification of data. Higher the recall value show less number of missing data during classification. The weighted recall value is 0.94.

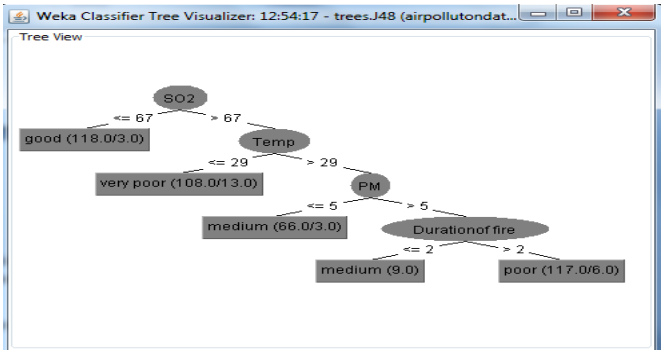

Fig- 4: Pruned Tree from J48 algorithm

Figure4 Shows the visualization tree which helps in understanding classification of data. Tree is created based on Gain information

$$GAIN_{split} = Entropy(p) - \left(\sum_{i=1}^{k} \frac{n_i}{n} Entropy(i) \right)$$

J48 pruned tree based on information gain. It has 5 leaves and size of tree is 9.It gives first split at SO_2 which has <=67µg/m3 are classified as good air quality index.SO_2>67 and Temp > 29 and PM<-5 and fire duration <=2 hours is classified as Medium air quality index .Remaining with Poor and Very poor air quality index.

The tree has pruned the nodes (attributes) Ozone (O_3) and Type of Vegetation. Because the Information gain from the nodes is zero. The SO_2 has highest information gain value hence, it is considered with first split node followed the Temperature, Particulate Matter (PM) and duration of fire.

The pruned tree analysis helped in prediction of air quality is poor. The rule generated for poor quality air during forest fire is, "If SO_2>=67µg/m3and temperature >29degree centigrade and PM>5µg/m3and fire duration is >2 hours" then it will be poor quality of air.

4. CONCLUSION

The forest fire which is hazard causing tremendous loss to the ecosystem. The work has provided an analysis of forest fire causing the air pollution using J48 Classification algorithm. The algorithm has proved that increase in Sulphur components more than 67µg/m3 in the air and temperature more than 29^0celcius causes more pollution than other pollutants.

REFERENCES

[1] Xiaolian Li , Weiguo Song,Liping Lian and Xiaoge Wei"Forest Fire Smoke Detection Using Back-Propagation Neural Network Based on MODIS Data" , Remote Sens. 2015, 7, 4473-4498; ISSN 2072-4292 Received: 14 December 2014 / Accepted: 15 March 2015 / Published: 15 April 2015. www.mdpi.com/journal/remotesensing doi:10.3390/rs70404473.

[2] M. Lazaridis, M. Latos& V. Aleksandropoulou, Hov & A. Papayannis & K. Tørseth "Contribution of forest fire emissions to atmospheric pollution in Greece "Received: 4 September 2008 /Accepted: 31 October 2008 / Published online: 18 November 2008 . This article is published with open access at Springerlink.com .DOI 10.1007/s11869-008-0020-0

[3] M a r t in, M. L. Prediction of CO maximum ground level concentrations in the Bay of Algeciras, Spain using artificial neural networks, Chemosphere, volume 70, Issue 7, January 2008, pp. 1190 – 1195

[4] A t h a n a s i a d is, I .N., K a r a t z a s, K. , M i t k a s , P . - Classification techniques forair quality

forecasting, BESAI 2006 Workshop on Binding Environmental Sciences and Artificial Intelligence, part of the 17th European Conference on Artificial Intelligence, 2006

[5] Hart, P.E., Cover, T.M. - Nearest neighbor pattern classification, IEEE Transactions on Information Theory, IT-13, 1967

[6] Hastie, T.Tibshirani, R. - Discriminant adaptive nearest neighbor classification, IEEE Trans. Pattern Anal. Mach. Intell. 18(6), 607–616 (1996)

[7] Brunelli, U. et al. - "Three hours ahead prevision of SO2 pollutant concentration using an Elman neural based forecaster", Building and Environment, Volume 43, Issue 3, March 2008, pp. 304- 314.

[8] Gaga jot Kaur, Amith Chabra, "Improved J48 Classification Algorithm for the Prediction of Diabetes", International Journal of Computer Applications(0975-8887)Vol98,No.22July2014.pp13-17

[9] Martin, M.L.et al. - Prediction of CO maximum ground level concentrations in the Bay of

[10] Algeciras, Spain using artificial neural networks, Chemosphere, volume 70, Issue 7, January 2008, pp. 1190 – 1195

[11] Oprea,M.Nichita, C., Dunea, D. - Aplicatii ale inteligentei artificiale in protectia mediului, Editura Universitatii Petrol – Gaze din Ploiesti, 2008

[12] Petre, E.G.- A Decision Tree for Weather Prediction, Bulletin of PG University of Ploiesti, Series Mathematics, Informatics, Physics, vol. LXI, nr. 1/2009, pp. 53-58

[13] Song, Y.,Huang, J.,Zhou, D., Zha, H., Giles, C.L. - IKNN: Informative K-Nearest Neighbor Pattern Classification, Proceedings of the 11th European conference on Principles and Practice of Knowledge Discovery in Databases, Warsaw, Poland, 2007, pp. 248 - 264

BIOGRAPHIES

Divya TL is currently working as a Assistant professor in Dept of MCA R.V. College of Engineering. She is pursuing PhD in VTU. Her research interest includes Data Mining, image processing.

Dr.Vijayalakshmi M.Nreceived her Ph.D from Mother Theresa Women's University, Kodaikanal, India in 2010, currently working as Associate Professor in Department of MCA R.V College of Engineering. Her current research interest includes Data mining, Image Processing, Wireless Sensor Networks.

