International Journal of Innovative Research in Computer Science & Technology (IJIRCST)
ISSN: 2347-5552, Volume-7, Issue-3, May 2019
DOI: 10.21276/ijircst.2019.7.3.2

Urban Air Computing: For Air Quality Detection

Nikhitha N, Dr. Rajashekara Murthy S

Abstract— **Air pollution is one of the biggest challenges that every metro-political areas is facing today. There are several tools and techniques evolved to predict pollution level which in turn helps in controlling and mitigating pollution. The three areas for computing pollution level are feature analysis, interpolation and prediction of fine grained air quality. These areas are providing extremely useful information so that one can take steps to mitigate pollution level, thus it also generates big societal impacts. Currently, there are individual models to address these issues separately. This paper proposes proposes single efficient framework by combining interpolation, prediction and feature analysis for air quality detection. This framework evaluates the different machine learning approaches to predict the air pollution components based on real data sets obtained from Bangalore. The main idea of the Urban Air Computing(UAC) is to gather the data of air quality and showing the feature analysis of air quality, interpolation and prediction.**

Index Terms- **Machine Learning, Urban Air Computing, Interpolation, Prediction, Feature Analysis.**

I. INTRODUCTION

The feature evaluation of quality, interpolation and prediction of air satisfactory are the vital subjects in the area of Urban Air Computing (UAC). The answers to these subjects can offer extraordinarily beneficial data to detect pollutants, measure pollution level and consequently generate high-quality societal impacts. Maximum of the current models addresses the three issues one at a time via different models. These tasks recommend a preferred and powerful technique to clear up the three troubles in a single model namely Urban Air Computing.The primary concept of Urban Air Computing involves choosing feature which has major contribution to pollution and builds semi-supervised model.

Manuscript received May 4, 2019.

Nikhitha N, M.Tech., Software Engineering, RV College of Engineering®, Bengaluru -59.
Email : nikhithan.sse17@rvce.edu.in

Dr. Rajashekara Murthy S, Associate Professor, Dept. of ISE,RV College of Engineering®, Bengaluru -59.
Email : rajashekaramurthys@rvce.edu.in

The recommended technique makes use of the records, related to the unlabeled spatial-worldly statistics, in turn enhancing the overall performance of the prediction, interpolation and plays feature selection. It also shows affiliation evaluation to expose the main applicable capabilities to the version of the air best. The Urban Air Computing predicts output by implementing the Random Forest algorithm. The algorithm works with the decision tress to get random class from the data sets and then gives out the decision to the system.

II. RELATEDWORK

Air pollution is the main reason for diseases. So an effective air quality monitoring system is the indeed system to look on the health of the human beings which is of the high value [1]. Inclination organize connected to the town quality assessment of air quality attributes, build up a exploratory outcomes harmonize with the real circumstance [2].

The interjection, forecast and highlight examination of fine-picked up air quality are three critical subjects in the coordinated non-cyclic chart, and the air fine information of for preparing and approval, the exploration demonstrates that the two strategies accomplish comparable assessment results [3]. In paper [4] authors have given insights on air quality measurement techniques used in China. The technique discharged by the Ministry of Environmental Protection(MEP) of China uses the most extreme individual air quality list as air quality file. The greater part of the procedures is AI (ML) based as it has turned out to be well known examination device on account of its different particular highlights [5]. The analysis of industrial air pollutants concentrations helps in taking preventive measures for example decreasing poison emission within the environment. The purpose of this work is to create pollutants awareness (NOx and CO) in business websites via using a Non-Linear Auto Regressive Version (NARX) based Artificial Neural Network (ANN) [6].

Semi-supervised schooling is a means for lowering the attempt had to put together the schooling set by way of education the model with a slight variety of completely categorized samples and an additional set of unlabeled or weakly labeled samples [7].

The increase in the framework to comprise unlabeled facts within the Errors-Correcting Output Coding (ECOC) setup by way of decomposing multiple class troubles into more than one problems and then use co-training to analyze the person one class problems [8].

Using multimedia annotation, results in time constrained and tediousness of guide tagging, it's far quite common place to utilize each tagged and untagged facts to improve the overall performance of supervised gaining knowledge of while simplest confined tagged training statistics [9].

III. METHODOLOGY

The admin will be having all rights to collect the data and upload the data which has been altered. The data is specifically location in the excel sheet through internet as this is an online website.

The data will be loaded to the main server by admin. In this stage preprocessing of data with the data cleaning and data modeling will be undergone. After this the data in main server will be consider with the Excel API, Data Mining techniques, location details and data sets for further operation. All these operation will be serviced through the web services [11].

The user is connected through this web services itself. In this section only feature selection and feature extraction will be taken place. The user will connected to the web services through internet. The user can be reached to the location server by selecting the location button in the web page [13]. The location server is connected to the current environment data and display system. The graphical user interface will be helping out to reach the activities. The admin and user will be connected through the web service. The same is shown in the figure 1.

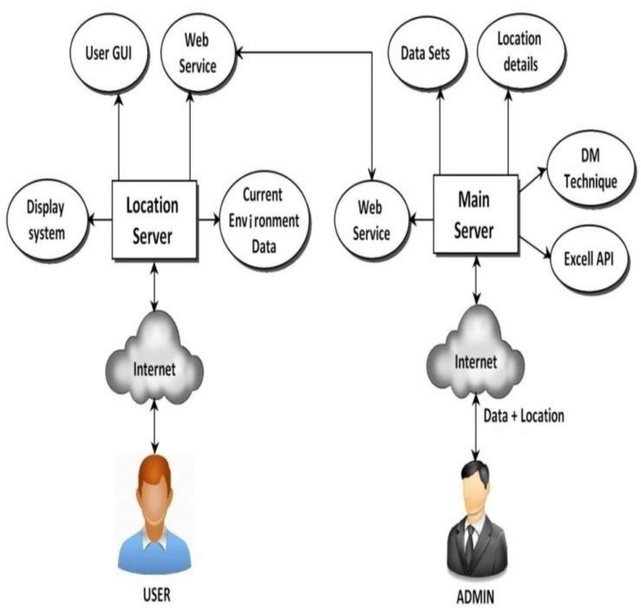


Fig 1: Proposed system architecture

Figure 2 Interaction of objects with different components in Urban Air Computing. The prediction process request data to database then data will be fetching from the dataset DB. The data will be retrieving from the dataset DB to prediction process. The main operation of random forest algorithm is to make unmanned prediction bout the favors of the users by gathering the data from many users [12]. Then those data will be given to location server for the user to display the data.

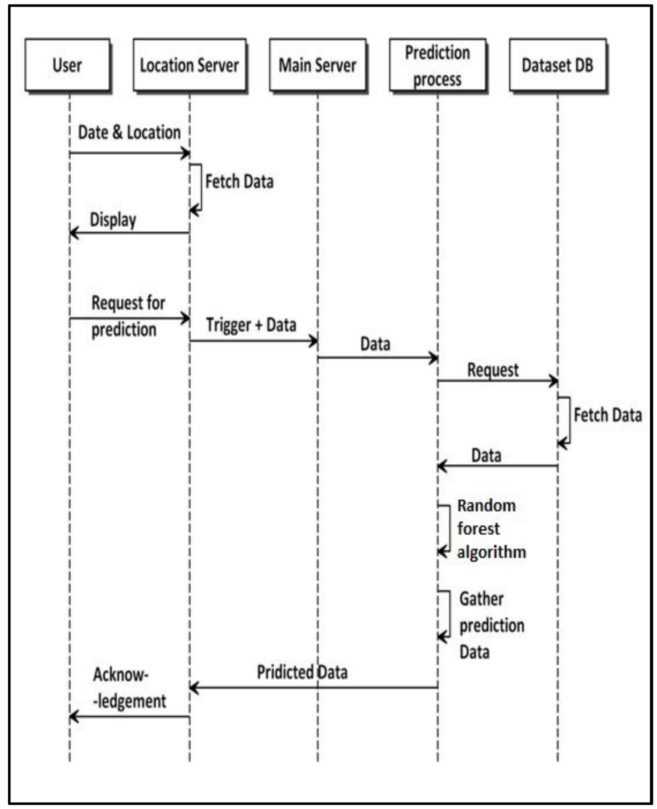


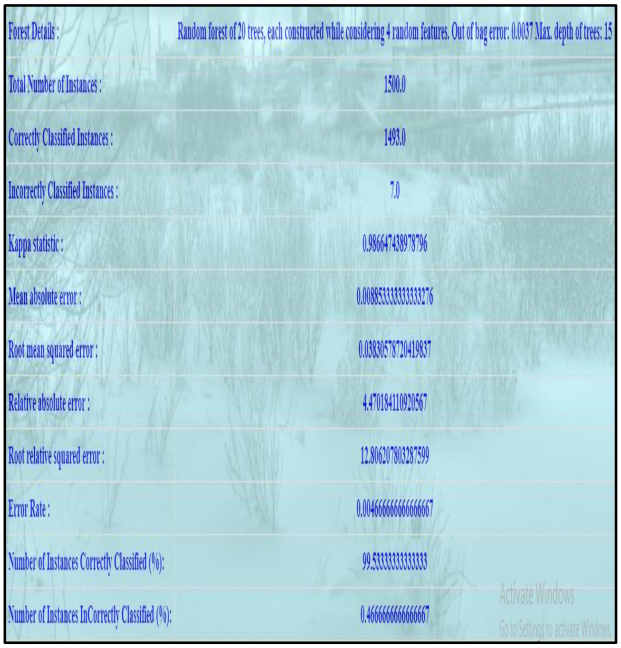
Fig 2: Data transmission in sequence diagram

As depicted in the Figure 2, there are four modules location server, where all the location and date will be stored. Location server fetches data from the user and it will display the data.

IV. ALGORITHM ANALYSIS

Random forests is an ensemble studying technique for regression, classification, prediction and other responsibilities that operates by using building a large number of choice bushes at training time and outputting the elegance that is the method of the instructions or on the other hand mean forecast of the person tree. Random decision forests accurate for decision timber dependency of over turning into to their training set. By using this algorithm the complete details about the data sets can be know. This helps in predicting the amount of pollution taken place in particular area. The sequence diagram of Urban Air Computing is shown in the Figure 2. The depicted Figure indicating that, the modules are interacting and communicating over some time period and the messages or the information is transferred between the modules [16]. For specific time period the event occurs and communicates with the other module with the same amount of time period. If the time varies then there may be the chances of more than one event to occur [14].

ISSN: 2347-5552, Volume-7, Issue-3, May 2019
DOI: 10.21276/ijircst.2019.7.3.2



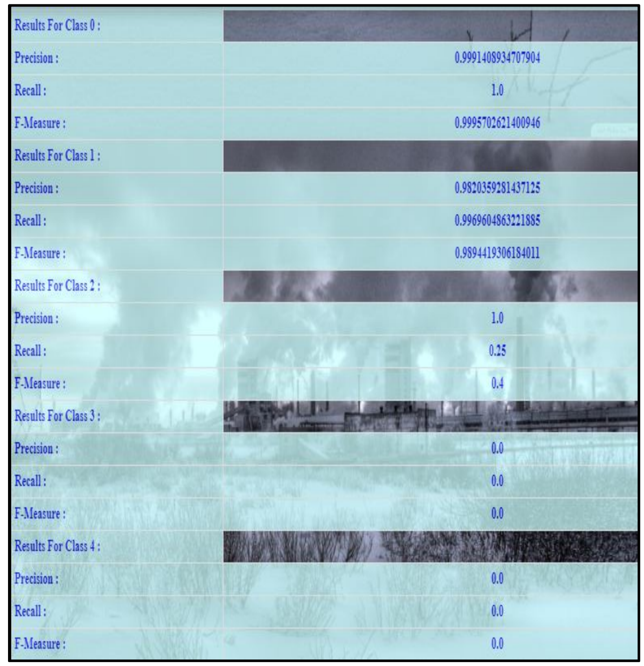
Forest Details :	Random forest of 20 trees, each constructed while considering 4 random features. Out of bag error: 0.0037 Max. depth of trees: 15
Total Number of Instances :	1500.0
Correctly Classified Instances :	1493.0
Incorrectly Classified Instances :	7.0
Kappa statistic :	0.986647438978796
Mean absolute error :	0.008853333333333276
Root mean squared error :	0.03830578720419837
Relative absolute error :	4.470184110920567
Root relative squared error :	12.806207803287599
Error Rate :	0.004666666666666667
Number of Instances Correctly Classified (%):	99.53333333333333
Number of Instances InCorrectly Classified (%):	0.4666666666666667

Fig 3: Interpolations by using random forest algorithm

The Figure 3 shows the interpolation result for the given datasets and with respect to the split ratio. It shows the random forest algorithm details, total number of entries in the data set, correctly classified data samples, incorrect classification of data samples, Kappa statistics, mean absolute error, relative error, root mean squared error, root relative squared error, error rate, total instances correctly classified and number of incorrectly classified data samples [17].

In forest details is shows about the out of how much trees the calculation takes place, bag error and the max length of the tree. Number of instances refers to the how many times the variable has been defined. Correctly classified is the total of the true positives and true negatives. Incorrectly classified is the addition of the false positives and false negatives [19]. The Kappa statistic (or value) is a metric that compares a found Accuracy with an anticipated Accuracy. Mean absolute error gives the pure value of the difference between the fore casted value and the actual value. The Root-Mean-Square Error (RMSE) is a regularly applied degree of the variations among numbers expected by means of a 2Version and the values actually observed. It also shows the relative errors. The root relative squared error regards to the prediction model [15].

The error rate shows the number of errors occurred. The number of correctly classified samples shows the exact classification of the number of samples. The number of samples incorrectly classified shows the remaining instances.



Results For Class 0 :	
Precision :	0.9991408934707904
Recall :	1.0
F-Measure :	0.9995702621400946
Results For Class 1 :	
Precision :	0.9820359281437125
Recall :	0.9969604863221885
F-Measure :	0.9894419306184011
Results For Class 2 :	
Precision :	1.0
Recall :	0.25
F-Measure :	0.4
Results For Class 3 :	
Precision :	0.0
Recall :	0.0
F-Measure :	0.0
Results For Class 4 :	
Precision :	0.0
Recall :	0.0
F-Measure :	0.0

Fig 4: Decision tree results

Figure 4 shows the decision tree results for the prior four classes. It shows the top four classes which have high rate results. The results consist of the precision, recall and F_Measure. Precision is the fraction of applicable times most of the retrieved instances where as recall is the fraction of relevant instances that have been retrieved over the entire quantity of applicable times. Both precision and consider are therefore based totally on an knowledge and degree of relevance. F_Measure is the weighted harmonic mean of recall and precision.

V. EXPERIMENTAL ANALYSIS AND RESULTS

This section shows the results of the urban air computing in implementation.

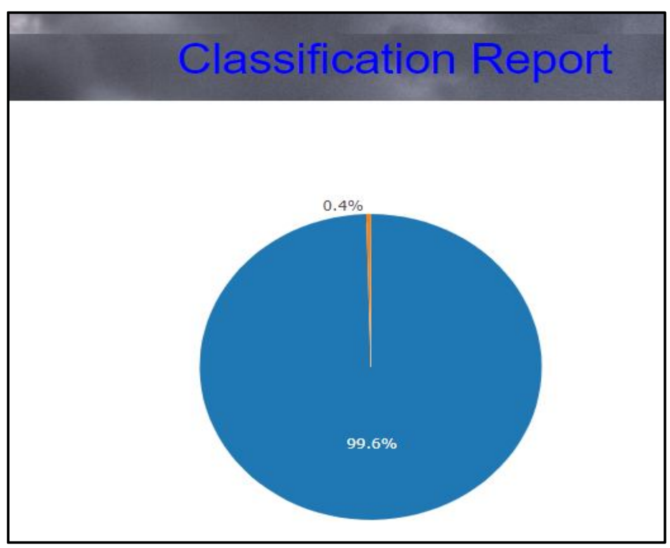


Fig 5: Pie chart shows the correctly classified errors and incorrectly classified errors.

The figure 3 shows the random forest as a train model. The figure 4 shows the pie chart for the prediction model. Figure 5 shows the pie chart with the correctly classified errors and incorrectly classified errors. It takes the input has data sets and output will be in the pictorial representation. The graph is showing 99.6 percentages of correctly classified errors and 0.4 percentage of incorrectly classified error. It is not a static one it changes according to the data sets.

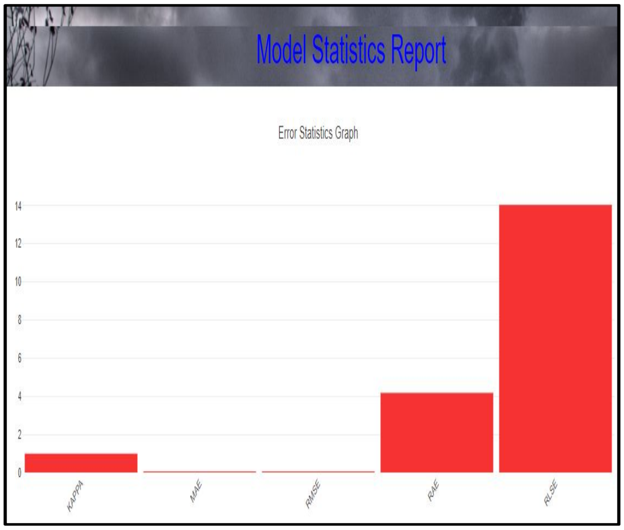


Fig 6: Bar graph shows model statistics

Figure 6 shows the error statistic graph with model statistic report. The bar graph shows the kappa statistics, Mean squared error, random mean squared error, randomized error are the statistics. The data sets will be considered with the random forest algorithm and above results will be calculated are given as output [20]. This will be shown in pictorial representation by using bar graph so that everyone can easily understand by seeing the result.

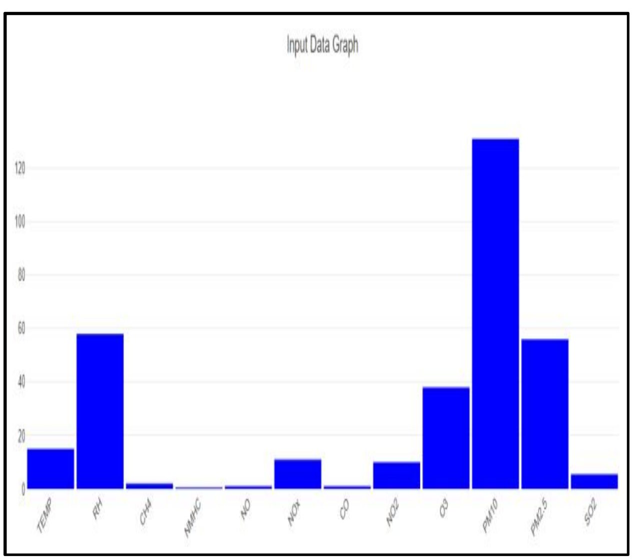


Fig 7: Predictions in graphical representation

Figure 7 shows the graphical representation of the urban air computing of the BTM layout. In the graph it shows the quality of the air along with components [18]. The components shown are temperature, relative humidity, methane, non methane hydrocarbons, nitric oxide, nitrogen oxide, carbon monoxide, nitrogen dioxide, ground level ozone, particulate matter 10, particulate matter 2.5, sulfur dioxide.

VI. CONCULUSION AND FUTURE WORK

The smart urban air computing model is proposed by using random forest algorithm. It works with the provided datasets as input and output will be the predicted data which will be shown using graphical representation. The main idea is to show the result in graphical representation so that any user can understand the result easily. It shows the clear cut information about the air quality by undergoing the random forest algorithm. It shows the amount of the components in air in graphical representation. If user is having the datasets of any area then they can predict the results, by simply uploading the data. It shows the results in bar graph and pie chart which helps in analyze the result in feasible way.

The future works inculcate the study of air pollution control system with the real time data. The control system has to be present everywhere so that data can be collected easily which helps in the working of the project in real time.

ACKNOWLEDGMENT

This project work is supported as part of major project for M.Tech, Software Engineering Post Graduate Program in the R.V.College of Engineering®, Bengaluru-560059.

REFERENCES

[1] Avijoychakma, Ben vizena, Tingting Cao, Jerry Ling, and Jing Zhang. “Image based air quality analysis using deep convolutional neural network”. IEEE International Conference on Image Processing (ICIP).

[2] Ruijun Yang, Feng Yan, Nan Zhao “Urban Air Quality Based on Bayesian Network“. IEEE 9th International Conference on Communication Software and Networks (ICCSN).

[3] Zhongang Qi, Tianchun Wang, Guojie Song, Weisong Hu, Xi Li and Zhongfei (Mark) Zhang “Deep Air Learning: Interpolation, Prediction, and Feature Analysis of Fine-grained Air Quality”IEEE Transactions on Knowledge and Data Engineering Volume: 30 , Issue: 12 , Dec. 1 2018.

[4] Meng Wei, Zeng Bo, Liu Si-feng, XieNai-ming“A grey incidence evaluation on air quality” Proceedings of 2013 IEEE International Conference on Grey systems and Intelligent Services (GSIS)

[5] VarshaHable-Khandekar, Pravin Srinath “Machine Learning Techniques for Air Quality Forecasting and Study on Real-Time Air Quality Monitoring” 2017 International Conference on Computing, Communication, Control and Automation (ICCUBEA).

ISSN: 2347-5552, Volume-7, Issue-3, May 2019
DOI: 10.21276/ijircst.2019.7.3.2

[6] NadjetDjebbri; MouniraRouainia "Artificial neural networks based air pollution monitoring in industrial sites" 2017 International Conference on Engineering and Technology (ICET).

[7] C.Rosenberg, M.Hebert and H.Schneiderman, "Semi-supervised self-training of object detection models" Seventh IEEE Workshops on Applications of Computer Vision (WACV/MOTION'05) - Volume 1.

[8] R. Ghani "Combining labeled and unlabeled data for text classification with a large number of categories" Proceedings IEEE International Conference on Data Mining.

[9] Lianli Gao, Jingkuan Song, FeipingNie, Yan Yan, NicuSebe, Heng Tao Shen "Optimal graph learning with partial tags and multiple features for image and video annotation" IEEE Conference on Computer Vision and Pattern Recognition (CVPR).

[10] Richard Socher, Jeffrey Pennington, Eric H. Huang, Andrew Y. Ng, Christopher D "Semi-Supervised Recursive Auto encoders for Predicting Sentiment Distributions" Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing.

[11] Yvan Saeys IñakiInza Pedro Larrañaga "A review of feature selection techniques in bioinformatics" Bioinformatics, Volume 23, Issue 19, 1 October 2007, Pages 2507–2517.

[12] HaoXu ; David J. Eis ; Peter J. Ramadge "The generalized lasso is reducible to a subspace constrained lasso" IEEE International Conference on Acoustics, Speech and Signal Processing.

[13] B. Krishnapuram, L. Carin, M. Figueiredo, and A. Hartemink, "Sparse multinomial logistic regression: fast algorithms and generalization bounds" IEEE Transactions on Pattern Analysis and Machine Intelligence. Volume: 27 , Issue: 6.

[14] Tzu-Hsuan Yang ; Tzu-Hsuan Tseng ; Chia-Ping Chen "Recurrent neural network-based language models with variation in net topology, language, and granularity" 2016 International Conference on Asian Language Processing (IALP).

[15] Mingxia Liu , Daoqiang Zhang "Pair wise Constraint-Guided Sparse Learning for Feature Selection" IEEE Transactions on Cybernetics (Volume: 46 , Issue: 1 , Jan. 2016).

[16] Xiabing Zhou, Wenhao Huang, Ni Zhang, Weisong Hu, Sizhen Du, Guojie Song, KunqingXie "Probabilistic dynamic causal model for temporal data" 2015 International Joint Conference on Neural Networks (IJCNN).

[17] SudjitKaruchit ; PongpatSukkasem "Application of AERMOD Model with Clean Technology Principles for Industrial Air Pollution Reduction" 2018 Third International Conference on Engineering Science and Innovative Technology (ESIT).

[18] Hua Yang, Qingjun Liu, Ke Pu, Yanju Liu "Purifying effects on high concentration of benzene, toluene and xylene from air cleaners under ventilation" International Conference on Remote Sensing, Environment and Transportation Engineering.

[19] YaliLv ; Shizhong Liao "Learning Temporal Qualitative Probabilistic Networks from Data" Second International Conference on Intelligent Networks and Intelligent Systems.

[20] Guiliang Liu "Seemingly unrelated regression modeling of urban air quality by direct Monte Carlo algorithm" International Conference on Wavelet Analysis and Pattern Recognition (ICWAPR).

