EFFECTIVE ANALYSIS OF AIR POLLUTION USING DECISION TREE, NAIVE BAYES AND ZEROR CLASSIFIERS

[1]Alaba, O. B., [2]Ogunsanwo, G. O., [3]Olayiwola, O. R., [4]Taiwo E. O., and [5]Abass O. A.
[1,2,3,4]Department of Computer Science, Tai Solarin University of Education, Ijagun, Ogun State
[5]Department of Computer Science, Tai Solarin College of Education, Omu-Ijebu, Ogun State
[1]alabaob@tasued.edu.ng, [2]ogunsanwogo@tasued.edu.ng, [3]phumisola@gmail.com,
[4]taemol19@gmail.com, [5]abassoa@tasce.edu.ng

Abstract

One of the leading factors for survival of living organisms on the earth is air. In the last twenty years, the world has been advancing in technology which has impacted negatively on the atmosphere around us thereby polluting the air. This explained why many researchers placed value on accurately forecasting the levels of pollution in the air. Also, an effective air quality management greatly depends on accurate air prediction. Recently, machine learning techniques are widely used in knowledge discovery. The study, therefore, analyzed the effectiveness of three machine learning techniques on air pollution dataset that was downloaded from Kaggle repository. The dataset consists of 15 attributes and 29532 instances which were further divided into 70% for training and 30% for testing. The metrics used for evaluation include: classification accuracy, error rate, execution time, mean absolute error, Root mean squared error (RMSE), confusion matrix and area under curve (AUC). The stimulation was done using WEKA statistical tool. The results showed error rate value of 0, 0.7, and 64.5% for Decision tree, Naive Bayes and ZeroR respectively. Area under curve value of 1, 0.992 and 0.499; Mean absolute error value 0, 0.0034 and 0.2457; Root mean squared error (RMSE) value of 0, 0.0427 and 0.3505; and Kappa statistic value of 1, 0.9903 and 0 for Decision tree, Naive Bayes and ZeroR respectively. Based on the analysis, the study concluded that decision tree algorithm recorded the highest prediction accuracy followed by Naive Bayes and ZeroR based on the datasets used. The study, therefore, recommends that the performance of other classification algorithms could be tested on the datasets.

Keywords: Air pollution, Classification algorithm, Decision tree, Naïve Bayes, ZeroR

Introduction

The existence of man can be seen to depend majorly on three things which are water, air and information, amongst these things mentioned air is very important. One of the leading factors for survival of living organisms on the earth is air. In the last twenty years, the world has been experiencing technology progression and transformation and this has counter effect on air. Air pollution is growing due to urbanization, industrial enterprise, machine, power plants, chemical action, and some of the other artificial activities such as agricultural burning and mountain eruptions. Pollution in the air has a straight effect on human well-being through the exposure of pollutants and particulates, this has led to high increase in air pollution research and its effects among the scientific gathering (Hvidtfeldt et. al., 2018; Gonzalez et. al., 2017; Pimpin et al. 2018).

The World Health Organization (WHO) estimated that around 6.9 million of deaths that occurred are due to air contamination throughout the world which has turned into a serious issue on the earth. It is evident from researches carried out around the world, that cities due to technological advancement have the leading danger zone. Many people are dying each year because of air pollution in major cities (i.e. urban areas) compared to rural areas. Air pollutants are categorized as solid particles, liquid droplets, or gases, which are grouped into the following: primary pollutants and secondary pollutants. The primary pollutants are discharge from the source straightaway into the air. The sources can be either physical processes, like sandstorms or human-affiliated, such as manufacture and vehicle discharge. The most common primary pollutants include the following: particulate matter (PM), sulfur dioxide (SO_2), carbon monoxide (CO) and nitrogen dioxide (NO_2). Secondary pollutants are air pollutants created in the atmosphere, resulting from the chemical or physical interactions between primary generated pollutants. Some of the examples of secondary pollutants include: photochemical oxidants and particulate matter. The most frequent air pollutants are known as the criteria pollutants, which agree to the most distributed health menace, e.g., PM, CO, lead, ground-level ozone (O_3), NO_2, and SO_2. Scientific investigation has proved a correlation between short-term vulnerability to this kind of pollutants and many health challenges, like limited ability to react to enlarged oxygen demands when sweating (particularly for people with heart issues), airway inflammation with people that have good health condition and increased respiratory indication for people with asthma, respiratory emergencies particularly for children and the elderly, and so on (US EPA, 2015). Developing a predictive system based on the levels of distribution of individual pollutants, which can predict air quality, will be a valuable and welcome research area that will promote population's health. Systems that can bring forth warnings based on air quality are consequently needed and essential for the world. They may play crucial roles in health vigilance when air pollution levels might surpass the specified levels; also, they may incorporate existing emission control programs, for example, by assisting environmental regulators the option of "on-demand" emission reductions, functional planning, or even emergency response (CERN, 2001).

In the face of progressively serious environmental pollution difficulty, scholars have conducted so many related studies, and in those works, the forecasting of air pollution has been of high importance. Thus, with the understanding of the increasing pollution caused problems, the importance of accurately forecasting the levels of air pollutants has equally increased, playing a vital role in air quality management and population prevention against pollution jinx. In order to solve all these problems, predictive model for forecasting the air quality in smart cities seems promising using machine learning techniques. The main goal of this paper is to analyze the efficiency of three different machine learning techniques for forecasting air pollution. Thus the concurrence of air pollutants like PM2.5, PM10, NO, NO_2, NO_X, NH_3, CO, SO_2, O_3, Benzene, Toluene, and Xylene, as a measure of air quality index (AQI), can be predicted by this model on time, with the least possible errors in the amount of air pollutants in the atmosphere to create an alert when specific threshold values are reached. The study therefore intends to use Decision Tree Algorithm, Naive Bayes Algorithm and ZeroR Algorithm for the air prediction model. The experiments would be conducted on the air pollution dataset using statistical WEKA tools. The rest of the paper is organized as follows: Section 2 contained the review of the related works on air quality prediction. Section 3 introduced study areas, models for predicting the air pollution and methodology in details. Section 4 introduced the performance criteria and the result discussion of the research. Section 5 presented concluding remarks.

Factors Responsible for Air Pollution

Air pollution resulted from both human and natural actions. Forest fires, volcanic eruptions, wind erosion, pollen dispersal, evaporation of organic compounds and natural radioactivity are natural events that pollute air (Pai et al., 2011). The burning of different kinds of fuel which include smoke stacks of power plants, manufacturing facilities (factories) and waste furnace, as well as furnaces and other types of fuel-burning heating devices. Also Waste product from paint, hair spray, coat, aerosol sprays and other dissolving agent make up air pollution. Waste products on landfills generate methane which is highly flammable, may turn to explosive mixtures with air. Nuclear weapons, toxic gases, germ and rocketry, also made up air pollutants. Vegetation in some areas discharge environmentally significant amounts of volatile organic compounds (VOCs) on sunny days. These VOCs react with main anthropogenic pollutants majorly, NO_X, SO_2, and anthropogenic organic carbon compound to produce a seasonal haze of secondary pollutants.

Naïve Bayesian model

Naïve Bayesian model is easy to develop, with no complex iterative parameter estimation, which makes it particularly useful for large datasets. Despite its simplicity, the Naïve

$$\mathbf{P(A|B)} = \frac{\mathbf{P(B|A)P(A)}}{\mathbf{P(B)}} \qquad (1)$$

Bayesian classifier often does brilliantly well and is commonly used because it regularly outperforms other classification techniques. The Bayes theorem is expressed in Equation 1**.**

Decision Tree Regression

Decision tree (DT) algorithm is a well-known machine learning model that falls under the category of supervised learning. As a good tool it can also be used for classification and predictions tasks (or activities). The basic idea of DT is to use the tree representation to solve the proposed problem in which each leaf/node corresponds to a specific class/label and columns are represented on the internal nodes of the created tree (Dixian et. al. 2018). The main problem of applying DT is to select the attribute for the root node in each level. To overcome this problem (attribute selection) there are two attributes selection measures namely information gain and gini index as shown in Equation 2.

$$\text{Information Gain}\,(S, A) = \text{Entropy}(S) - \sum_{j=1}^{z} (p_j)^2 \quad v \in \text{Values} \qquad (2)$$

ZeroR Classifier

ZeroR is one of the rule based classifiers which depends on the target and neglect all predictors. It merely predicts the bulk class. The ZeroR classifier examines the target attribute and its possible values. It builds the frequency table and pick out its most common value. It will output the value that is most often found for the reference point attribute in the given dataset. ZeroR as its names implies, does not consist of any rule that function on the non-target attributes. The overall goal is to predict the mean (for a numeric type target attribute) or the mode (for a nominal type attribute).

Related Works

Numerous techniques have been proposed in order to apply machine learning and data mining to predict air pollution control in recent literature. Cai et al. (2009) carried out a study to compare the results obtained from the prediction using a multilinear regression model (MRM) to the ones achieved by an artificial neural network (ANN) when predicting hourly air pollutant concentration, the outcome of the study found out that ANNs produced more robust results than MRM. Pires et al. (2010) conducted a research using geometric progression (GP) to predict the daily averages of PM10 concentrations, comparing it with partial least square regression (PLSR), and it was concluded that GP was able to predict the pollution. Tikhe et al. (2013) developed an air predictive model. The study used both ANNs and GP to predict air quality in India. The study concluded that both approaches obtained reasonable performance when predicting the air pollutant concentrations, but the study found out that GP obtained better results when short-term forecasting was considered. Castelli et al. (2016) also carried out a research to predict the quality of air. The study presented an evolutionary system to predict ozone concentrations an hour earlier with GP, based on other pollutant concentrations. The study found out that the approach achieved accurate results, better than the state-of-the-art machine learning (ML) techniques. Vineeta *et. al*. (2019) conducted a study to estimate the air pollution for any day/time in Bengaluru city using a trained machine learning model with pollution data gathered from government sites and static sensors. The study used linear regression (LR), decision tree regression (DTR) and random forest regression (RFR). The study concluded that ML technique predicted accurately the air pollution. Amrutha C and Prasad B. G. (2018) developed a multiple individual prediction models including autoregressive integrated moving average (ARIMA), multilayer perception (MLP), and multiple linear regression (MLR) for PM2.5 pollutant. The study came up with air pollutant concentration prediction using ensemble of machine learning techniques based on the three individual prediction algorithms ARIMA, MLR and MLP. Castelli et al. (2020) employed a popular machine learning method, support vector regression (SVR), to forecast pollutant and particulate levels and to predict the air quality index (AQI). Among the various tested alternatives, they found out that radial basis function (RBF) was the type of kernel that allowed SVR to obtain the most accurate predictions. The research concluded based on the results that SVR with RBF kernel allow to accurately predict hourly pollutant concentrations like carbon monoxide, sulfur dioxide, nitrogen dioxide etc. Doreswamy et al. (2020) developed machine learning predictive models for forecasting particulate matter concentration in atmospheric air. They investigated on Taiwan Air Quality Monitoring data sets, which were obtained from 2012 to 2017. Their models developed were compared with the existing traditional models and the result was better in predictive performance. The performance of these models was evaluated with statistical measures such as root mean square error (RMSE), mean absolute error (MAE), mean square error (MSE), and coefficient of determination (R2). Based on the literature consulted so far little or no research was carried out to analyze the effect of air pollution using Decision tree, Naïve Bayes and ZeroR classifiers on air pollution dataset.

Methodology

Three data mining classifications algorithms, namely Decision tree, Naïve Bayes and ZeroR were employed in the prediction of the air pollution. Figure 1 shows the framework for this study.

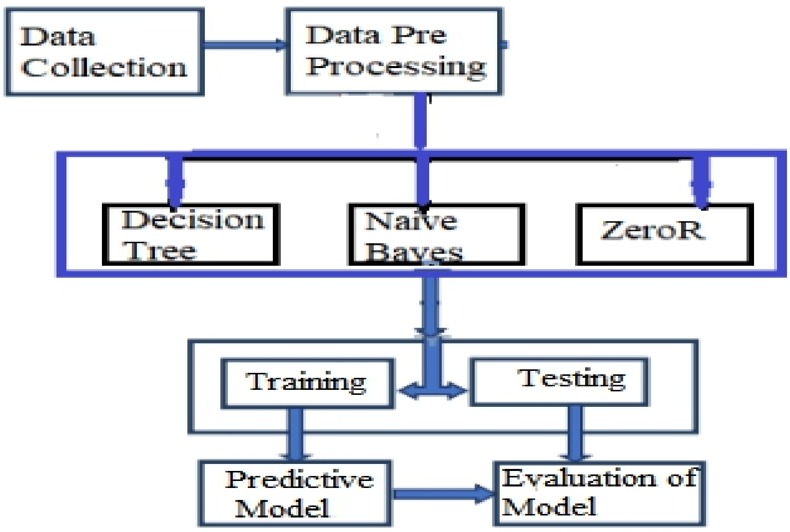

Figure 1:

Framework for the study

Datasets and Attributes

The datasets were obtained from Kaggle dataset Repository. The dataset consists of 15 attributes namely: Date, PM2.5, PM10, NO, NO_2, NO_X, NH_3, CO, SO_2, O_3, Benzene, Toluene, Xylene, AQI, and AQI_Bucket. The data consists of information collected from 29532 instances. The values of the first 15 features were used to define the input variables of the prediction model. The percentage splits of 70% training and 30% testing were used for the prediction. Table 1 shows the presentation of the characteristics of the datasets

.

Table 1: Air Pollution Dataset

Training Data	Testing Data
70%	30%

3.2 Models

The predictive model for the air pollution was formulated using the specified dataset. The mathematical expressions called mapping functions were used to express the process of model development (and loss function) following which the description of the selected Decision Tree, Naïve Bayesian and ZeroR algorithms were selected for the purpose of this study. The training dataset S which consisted of the original features identified at the point of data identification and collection was represented by i, where i is the number of features existing in the original dataset whose record were collected number of instances. If n datasets are selected for training the predictive model using a supervised machine learning (SML) to formulate the model using the relevant variables is given by the mapping:

$\varphi: X_{in} \rightarrow Y_n;$ (3)

where; $X_{in} = Y_n$ for all attributes , $_{in}$ is the set of n instances of the Kaggle datasets collected for the i air pollution variables and Y_n is the n instances of prediction of the datasets as correctly classified or incorrectly classified. Hence, the selected SMLs determine the best fit for $\varphi\varepsilon\mathbb{R}$ (the set of all possible mappings) based on the minimization of the loss function defined for each SML as the mapping below:

$$\mathbb{L}: \mathbb{Y} \times \mathbb{Y} \rightarrow \mathbb{R}+;\ defined\ as\ \mathbb{L}\,(Y_a\, Y_p) \quad (4)$$

where $\mathbb{R}+$ is a positive real number and Y_a, Y_p are the actual and predicted values of the target class of air pollution prediction respectively. Hence, the optimal predictive model is formulated when $\lim_{n\rightarrow j}\mathbb{L}_n=0$.

Hence, the prediction of the air is thus:

$$\mathcal{L}(Y_a,Y_p) = \begin{Bmatrix} correct\ classification; = 0 \\ incorrect\ classication; \neq 0 \end{Bmatrix} \quad (5)$$

Data Pre-processing

Data quality and its representation are the first and foremost points to guarantee the successful building of forecasting models. The data preprocessing step often impacts the generalization ability of a machine learning algorithm (Kotsiantis et. al. 2006). The data pre-processing was carried out using multi filter as shown in Figure 2 and Figure 3.

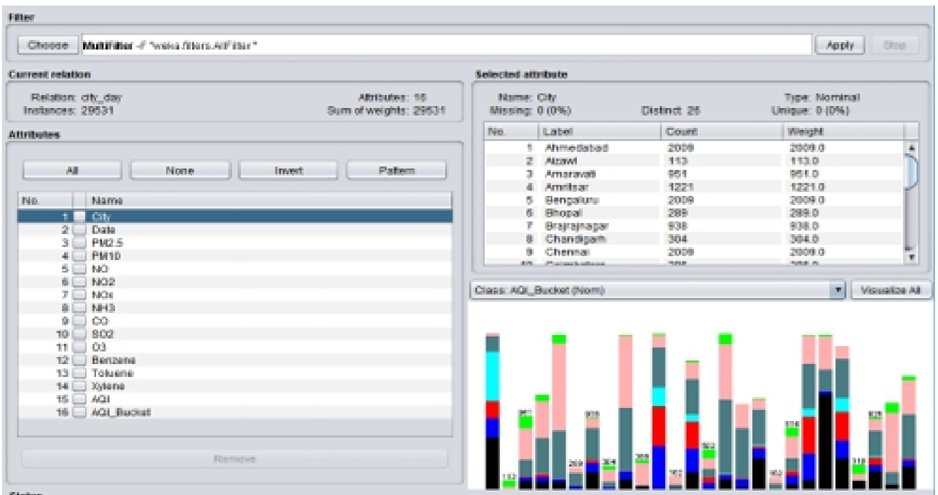

Figure 2: Data Prep processing

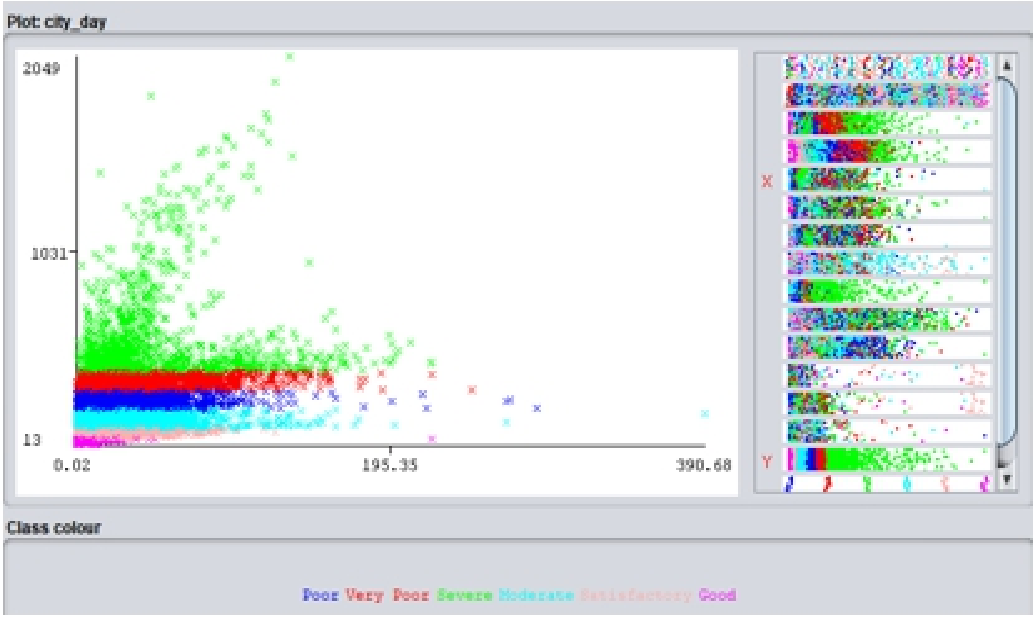

Figure 3: Graphical View of the Pre-processing

Evaluation Parameters

The classification techniques selected were evaluated based on six parameters namely: Classification Accuracy, Execution Time (Speed), Error Rate, Confusion Matrix, Kappa statistic and Area Under Curve (AUC).

Classification Accuracy

Classification accuracy is the ratio of several correct predictions to the number of input samples. The algorithm can correctly predict the class label of new or previously unseen data.

$$\text{Accuracy} = \frac{\text{Number of Correct Predictions}}{\text{Total Number of Predictions Made}} \quad (6)$$

Execution Time

Execution time is the time taken by the WEKA tool to classify the dataset using a classification algorithm. In measuring execution time, implementation-defined mechanism was employed. Execution time pertains to the computational cost involved in generating and using the algorithm.

Error Rate

The Error rate is measured in terms of the Mean Absolute Error and Mean Squared Error.

True Positive

True Positive rates which mean (sensitivity/recall) implies proportion of positive cases correctly classified**.**

$$\text{TP rate}_{\text{No}} = \frac{\text{TP}}{\text{TP+FN}} \quad (7)$$

$$\text{TP rate}_{\text{Yes}} = \frac{\text{TN}}{\text{FP+TN}} \quad (8)$$

Mean Absolute Error

The mean of the difference between the original values and the predicted values is referred to as the Mean Absolute Error (MAE). This statistic gives the measure of how far the predictions were from the actual output as shown in Figures 4 to 16. Figure 4, Figure 9 and

Figure14 show the model output of the training experiment of the Decision Tree Algorithm, Naïve Bayes classifier and ZeroR algorithm respectively on the full training set on Air pollution using 70%, 30% split. The result showed a correctly classified instances of 100%, 99.3% and 35.5% for Decision Tree Algorithm, Naïve Bayes classifier and ZeroR algorithm respectively and incorrectly classified instances of 0%, 0. 7% and 64.5% for Decision Tree Algorithm, Naïve Bayes classifier and ZeroR algorithm respectively.

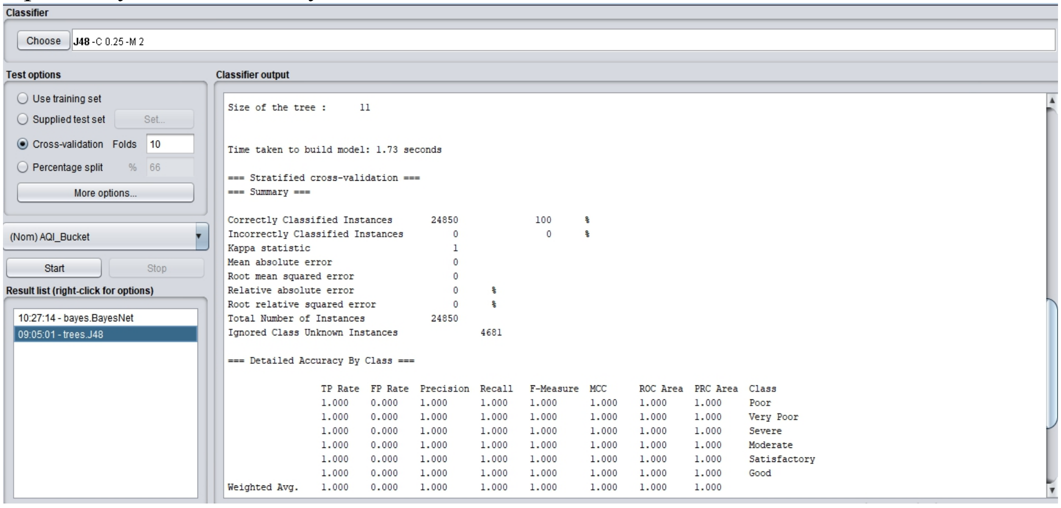

Figure 4: Decision Tree Algorithm Model

```
=== Confusion Matrix ===

    a    b    c    d    e    f   <-- classified as
 2781    0    0    0    0    0 |    a = Poor
    0 2337    0    0    0    0 |    b = Very Poor
    0    0 1338    0    0    0 |    c = Severe
    0    0    0 8829    0    0 |    d = Moderate
    0    0    0    0 8224    0 |    e = Satisfactory
    0    0    0    0    0 1341 |    f = Good
```

Figure 5: Confusion Matrix for Decision Tree Algorithm Model

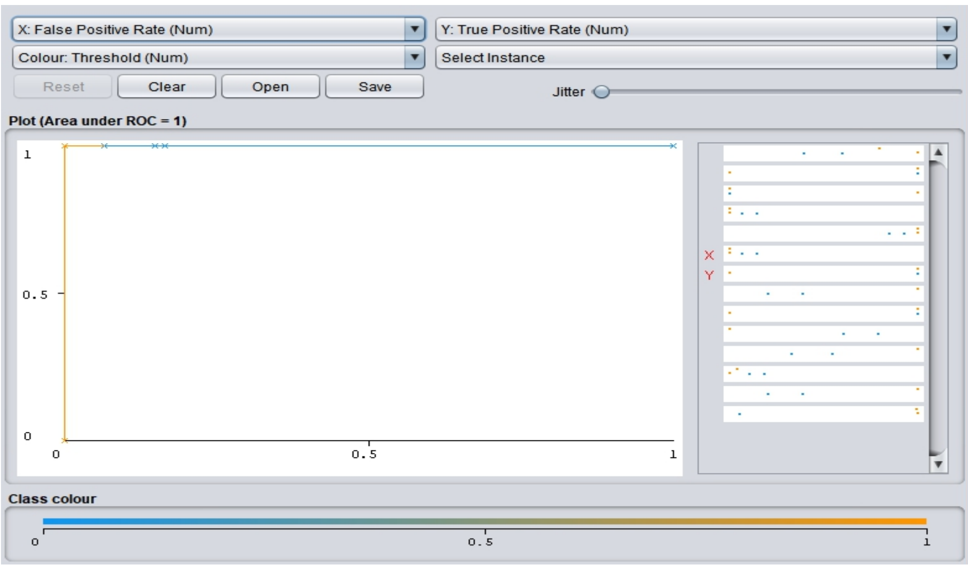

Figure 6: AUC for Decision Tree Algorithm Model

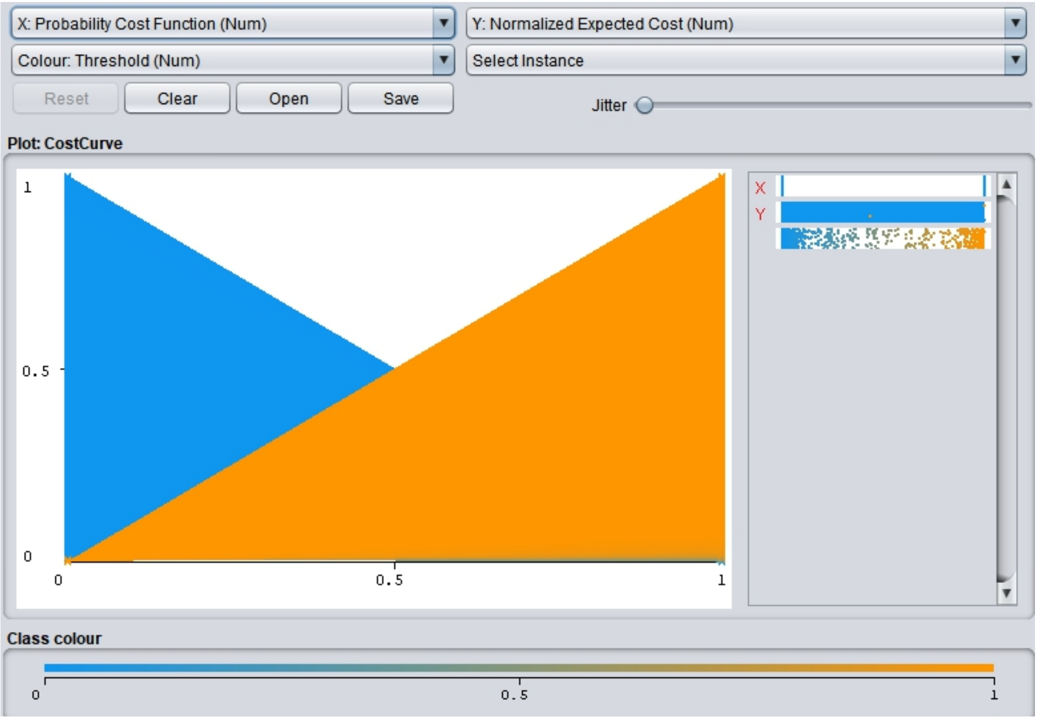

Figure 7: Cost Function for Decision Tree Algorithm Model

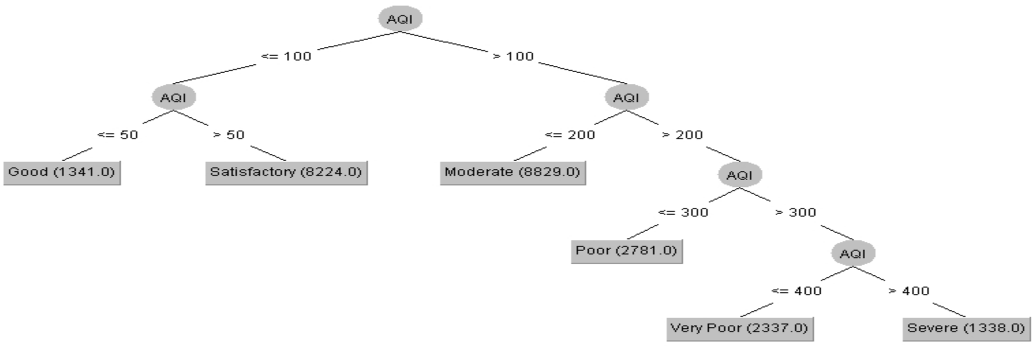

Figure 8: Tree View for Decision Tree Algorithm Model

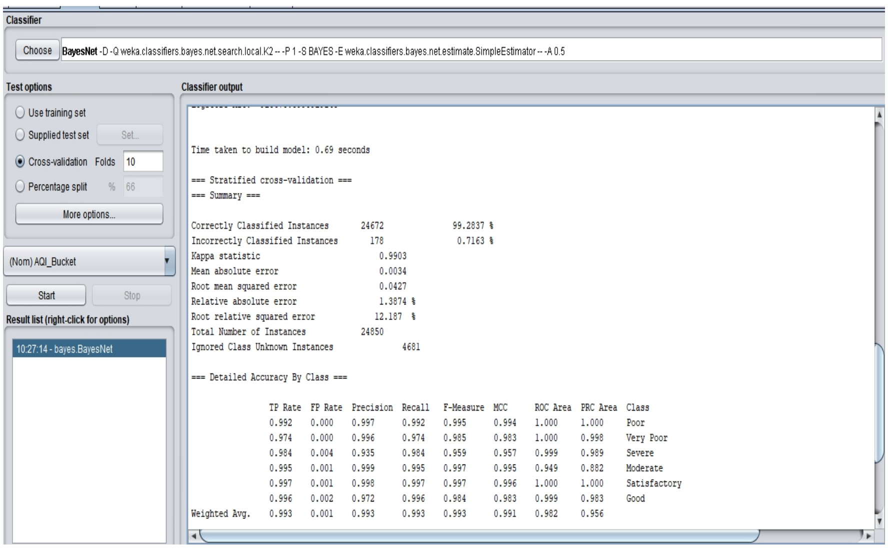

Figure 9: Model for Naïve Bayesian Algorithm

```
=== Confusion Matrix ===

    a    b    c    d    e    f   <-- classified as
 2760    0    9    1    5    6 |    a = Poor
    0 2276   52    3    5    1 |    b = Very Poor
    3    6 1317    4    5    3 |    c = Severe
    2    3   17 8787    1   19 |    d = Moderate
    3    0   14    1 8196   10 |    e = Satisfactory
    1    0    0    1    3 1336 |    f = Good
```

Figure 10: Confusion Matrix for Naïve Bayes

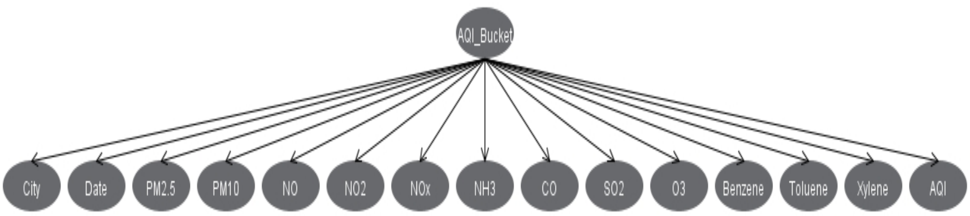

Figure 11: Graphical View for Naïve Bayesian Algorithm

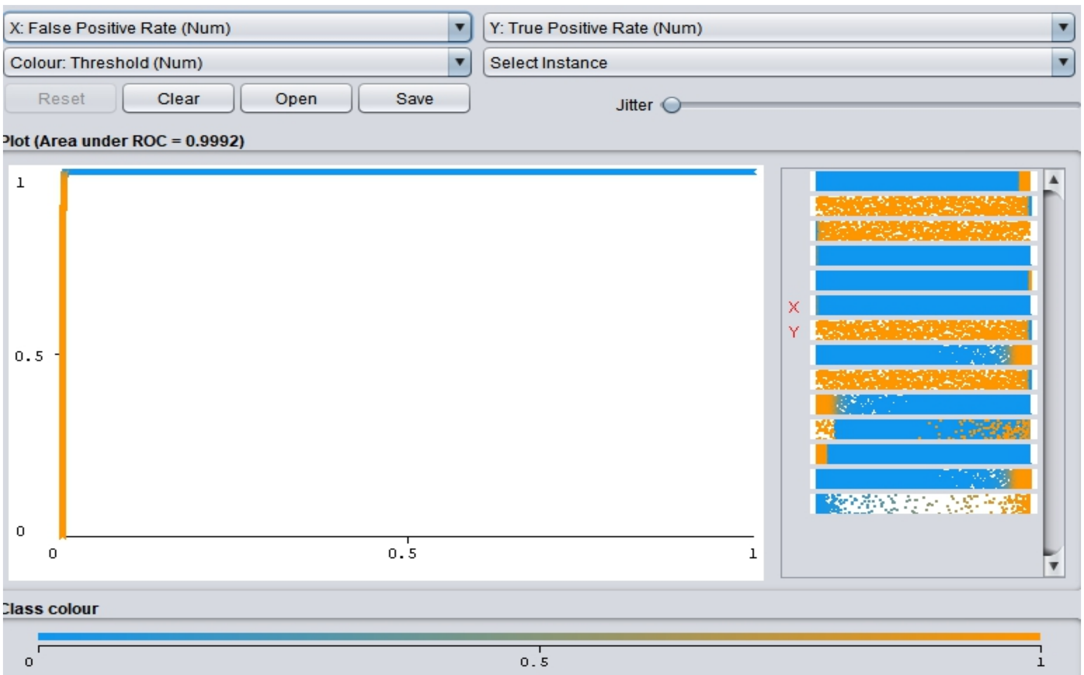

Figure 12: AUC for Naive Bayesian Algorithm

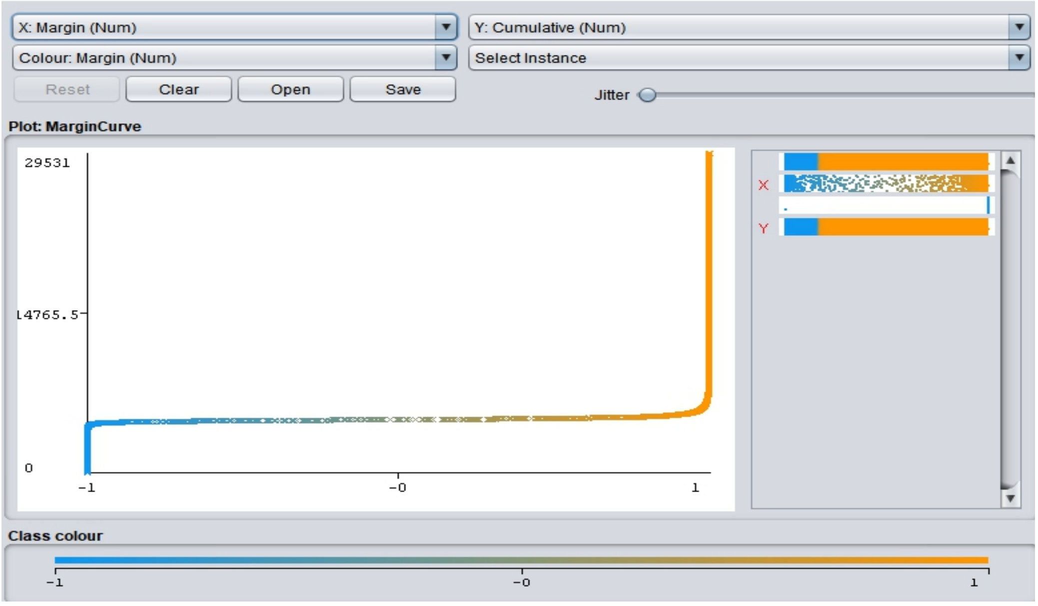

Figure 13: Model for ZeroR classifier

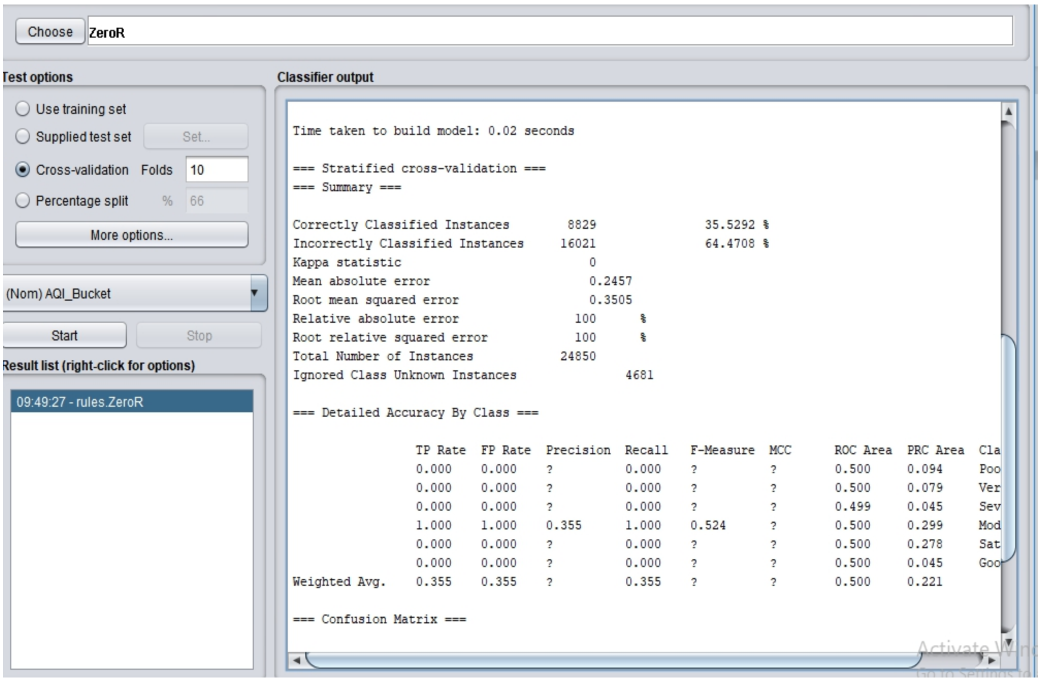

Figure 14 Model for ZeroR classifier

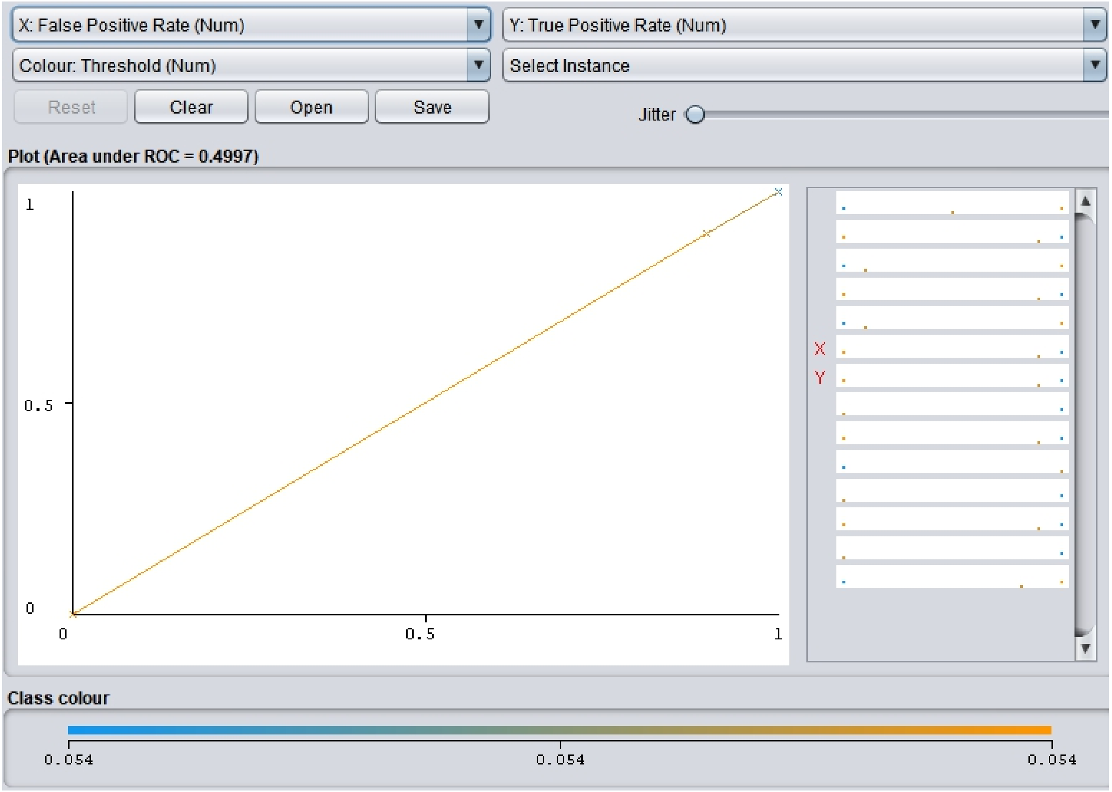

Figure: 15 AUC for ZeroR classifier

```
=== Confusion Matrix ===

    a    b    c    d    e    f   <-- classified as
    0    0    0 2781    0    0 |    a = Poor
    0    0    0 2337    0    0 |    b = Very Poor
    0    0    0 1338    0    0 |    c = Severe
    0    0    0 8829    0    0 |    d = Moderate
    0    0    0 8224    0    0 |    e = Satisfactory
    0    0    0 1341    0    0 |    f = Good
```

Figure 16: Confusion Matrix for zeroR classifier

Results and Discussion

Results

The results obtained from the analysis of each data split are presented in Table 2 and Figure 17.

Discussion

The accuracy of the Decision Tree algorithm is higher compared to the other two algorithms Naïve Bayes algorithm and ZeroR algorithm for the air pollution dataset split into 70% 30%. The execution time of the ZeroR algorithm is faster on 70% 30% compared to the other two algorithms Decision Tree algorithm and Naïve Bayes algorithm. With regards to the error rate, 70% 30% data split ZeroR algorithm has more percentage of recorded errors compare to Decision Tree and Naïve Bayes algorithm. Kappa statistic of Decision Tree is 1 which is high compare to Bayesian algorithm which is 0.9903 and 0 for ZeroR algorithm. The MAE is 0, 0.0034, 0.2457 Decision Tree algorithm, Naive Bayes and ZeroR respectively. The RMSE is 0, 0.0427, 0.3505 for Decision Tree algorithm, Naive Bayes and ZeroR respectively. The Area under the curve (ROC) is 1, 0.992, 0.499, Decision Tree algorithm, Naive Bayes and ZeroR respectively

Table 2: Results of the algorithms on the datasets

	Decision Tree Algorithm	Naive Bayesian Algorithm	ZeroR Algorithm
Air Pollution Dataset	**70% 30%**	**70% 30%**	**70% 30%**
Correctly Classified instances	2485	2467	8828
In correctly Classified instances	0	170	16021
Classification Accuracy (%)	100	99.3	35.5
Execution Time (seconds)	1.73	0.69	0.02
Error Rate (%)	0	0.7	64.5%
Area Under Curve	1	0.992	0.499
Mean absolute error(MAE)	0	0.0034	0.2457
Root mean squared error (RMSE)	0	0.0427	0.3505
Kappa statistic	1	0.9903	0

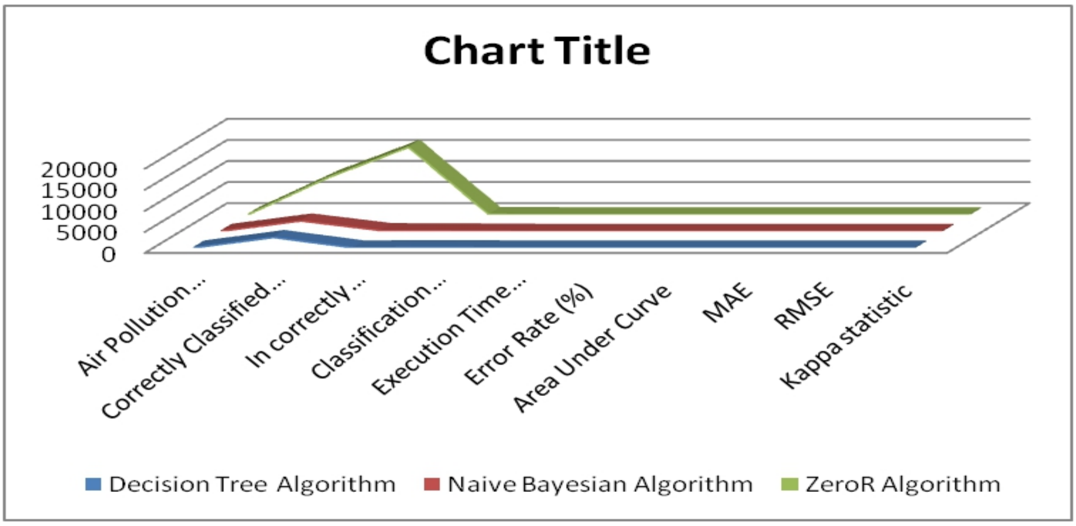

Figure 17: Comparison of the Three Algorithms

Conclusion

Air quality prediction has been and will remain a complex activity as a result of the dynamic nature, volatility, and high variability in space and time of pollutants and particulates. At the same time, being able to model, predict, and monitor air quality is becoming more and more important, especially in urban areas, due to the observed critical impacts of air pollution for populations and the environment. The datasets were obtained from the https://www.kaggle.com/muthuj7/airpollution-dataset repository. Each dataset was split into two, namely: 70% training and 30% testing. The study examined the performance of three data mining classification algorithms: Decision Tree, Naïve Bayes and ZeroR in predicting the chances of air pollution occurrence. Decision tree algorithm recorded the highest prediction accuracy followed by Naive Bayes and ZeroR based on the datasets used. The study, therefore, recommends that the performance of other classification algorithms could be tested on the datasets. For further future work, we hope to improve on what has been done so far and look into how SVR can be used to forecast air quality.

REFERENCES

Amrutha C. and Prasad B. G. (2018). Air Pollutant Concentration Prediction using Ensemble of Machine Learning Techniques. *International Journal of Research in Electronics and Computer Engineering* (IJRECE), *6*(3) version 2, 406-412.

Cai M., Yin Y., and Xie M. (2009). Prediction of hourly air pollutant concentrations near urban arterials using artificial neural network approach. *Transportation Research Part D: Transport and Environment*, *14*(1), 32–41.

Castelli M., Goncalves P. A., Trujillo L. and Popovic A. (2016). *An evolutionary system for ozone concentration forecasting*, pp. 1123– 1132, Springer Science+Business Media, New York, NY, USA.

Castelli M., Clemente F. M., Popovic A., Silva S. and Vanneschi L. (2020). A Machine Learning Approach to Predict Air Quality in California Hindawi Complexity Volume 2020, Article ID 8049504, 23 pages https://doi.org/10.1155/2020/8049504

CERN, (2001) Air Quality Forecasting, CERN, Geneva, Switzerland.

Dixian Z., Changjie C., Tianbao Y. and Xun Z. (2018). A machine learning approach for air quality prediction: Model regularization and optimization. *Big data and cognitive computing* **2**, 1(22): p.15.

Doreswamy, Harishkumar K. S., Yogesh K. M., Ibrahim G. (2020). Forecasting Air Pollution Particulate Matter (PM2). Procedia Computer Science 00 (2019) 000–000 www.elsevier.com/locate/procedia. Third International Conference on Computing and Network Communications (CoCoNet'19)

Gonzalez Y., Carranza C., Iniguez M.,Torres M., Quintana R., Osornio-Vargas A. R., Gardner C., Sarkar S. and Schwander S. (2017). Inhaled air pollution particulate matter in alveolar macrophages alters local pro-inflflammatory cytokine and peripheral IFN production in response to mycobacterium tuberculosis. *American Journal of Respiratory and Critical Care Medicine*, 195: A2901-A2901

Kotsiantis, S., Kanellopoulos, D., and Pintelas, P. (2007). Data Preprocessing for Supervised Leaning. World Academy of Science, Engineering and Technology, Open Science Index 12, *International Journal of Computer and Information Engineering*, *1*(12), 4104 - 4109.

Pai T. Y., Ho C. L., Chen S. W., Lo H. M., Sung P. J., Lin S. W., Lai W-J, Tseng S. C. and Ciou S-P (2011). Using seven types of GM (1.1) model to forecast hourly particulate matter concentration in Banciao City of Taiwan. *Water Air Soil Pollut 217*:25–33.

Pimpin L., Retat L., Fecht D., Preux L., Sassi F., Gulliver J., Belloni A. and Ferguson B. (2018). Estimating the costs of air pollution to the National Health Service and social care: an assessment and forecast up to 2035. *PLoS Medicine*, *15*(7), Article ID e1002602, 1–16.

Pires J. C. M., Alvim–Ferraz M. C. M., Pereira M. C., and Martins F. G. (2010). Prediction of PM10 concentrations through multi-gene genetic programming. *Atmospheric Pollution Research, Science Direct 1*(4), 305–310.

Vineeta, Ajit B., Asha S. M. and Vineeta M. (2019). Machine Learning based Prediction System for Detecting Air Pollution. *International Journal of Engineering Research & Technology* (IJERT) ISSN:2278-0181 http://www.ijert.org IJERTV8IS090050 *8*(9).

Hvidtfeldt, U., Ketzel, M., Sørensen, M., Hertel, O., Khan, J., Brandt, J., and Raaschou-Nielsen, O. (2018). Evaluation of the Danish AirGIS air pollution modeling system against measured concentrations of PM2.5, PM10, and black carbon. *Environmental Epidemiology, 2*, e014.

Tikhe S. S., Khare K. C., Londhe S. N. (2013). Forecasting criteria air pollutants using data driven Approaches; an Indian case study. *IOSR Journal Of Environmental Science,* *Toxicology And Food Technology* (IOSR-JESTFT), *3*(5), 01–08, http://www.iosrjournals.org/iosr-jestft/pages/v3i5.html.

US Environmental Protection Agency (US EPA), "Criteria air pollutants," *America's Children and the Environment*, US EPA, Washington, DC, USA, 2015. vol. 195, p. S29.