International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 4
DOI: https://doi.org/10.17762/ijritcc.v11i4.6442
***Article Received**: 06 February 2023 **Revised**: 04 March 2023 **Accepted**: 18 March 2023*

Air Pollution Detection and Control System Using ML Techniques

[1]Anuradha Sanapala, [2]B. Jaya Lakshmi, [3]K SandhyaRani Kundra, [4]K. B. Madhuri
[1]Research Scholar, Duke Training Center, Abu Dhabi, United Arab Emirates
[2,3,4]Department of Information Technology
Gayatri Vidya Parishad College of Engineering (A)
Visakhapatnam, India
e-mail: anuradha.sanapala@gmail.com
e-mail: meet_jaya200@gvpce.ac.in
e-mail: sandhyaranikk@gvpce.ac.in
e-mail: drkbmadhuri@gvpce.ac.in

Abstract—In present times, air pollution is increasing day by day, depriving the health of many people due to the various toxic components in air. So, it is necessary to monitor and detect the levels of pollution in various areas and try to control it by taking precautionary actions. Air pollution detection and control system is all about detecting the level of pollution in a particular area based on the amount of polluting components and proposing the measures to control the pollution. Analysis is made on the regions of Visakhapatnam city in Andhra Pradesh, India and grouped based on their pollution and displayed along with each component level, reasons for the pollution depending on each component and measures that can be followed. Apart from this, we also display list of top 10 regions with the highest values for each component which can be used to identify the harmful regions based on the toxic components

Keywords- Air Pollution, Machine Learning algorithms, Accuracy

I. INTRODUCTION

Air pollution is a mixture of solid particles and gases in the air. Car emissions, chemicals from factories, dust, pollen, and mold spores may be suspended as particles [1]. Ozone, a gas, is a major part of air pollution in cities. When ozone forms air pollution, it's also called smog. Some of these air pollutants are poisonous. In short, air pollution means the undesirable presence of impurities or the abnormal rise in the proportion of some constituents of the atmosphere. Air pollution is caused by the presence in the atmosphere of toxic substances, mainly produced by human activities, even though sometimes it can result from phenomena such as volcanic eruptions, dust storms and wildfires, also depleting quality of the air. Air pollution has a major impact on the process of plant evolution by preventing photosynthesis. On top of that, air pollution is a major contributor to global warming and climate change [2]. Human health deteriorates as a result of constant exposure to air pollutants. So, it is necessary to monitor and detect the levels of pollution in various areas and try to control it by taking precautionary actions. Air pollution detection and control system is all about detecting the level of pollution in a particular area based on the amount of polluting components and proposing the measures to control the pollution. Further analysis is also made grouping the areas based on their pollution which can be useful to take precautionary actions by proposing them along with the groups.

A. Objective

- Predicting the air pollution level in a particular area.
- Indicating the major components that are responsible for the pollution.
- Proposing the measures to be followed to reduce the level of intensity of the pollution based on the values of components.
- Analyzing and categorizing the areas depending on their level of pollution and displaying them along with the precautions that can be followed in the regions.
- Displaying the top 10 areas with the highest values of each component along with effects and precautions.

B. Purpose

- To create an effective environment for predicting the level of air pollution in a particular area.
- Understand the components that are mainly causing pollution in an area.
- Follow the precautions that are stated based on the pollutants' pollution to control the same.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 4
DOI: https://doi.org/10.17762/ijritcc.v11i4.6442
***Article Received**: 06 February 2023 **Revised**: 04 March 2023 **Accepted**: 18 March 2023*

- Based on the historical data sets, analysis is made and areas are grouped based on the pollution along with reasons, effects and control measures.
- Retrieving areas with higher values of each component.

II. EXISTING SYSTEM

A. *Display Boards*

1) *Functionality*

Display boards offer to show current time and date from satellite, temperature, and humidity along with various air quality parameters such as CO, PM2.5, and PM10 along with calculated Air Quality Index (AQI) depending upon the above parameters [3]. It is important and it's a social responsibility to communicate the effects of pollution to the public and improve general awareness [4].

2) *Disadvantages*

- It may not be understood by common people.
- There is no comprehensive information.
- It shows too much simple data.
- AQI may not be shown completely.
- It is not interactive.

B. *AQI SCORE*

1) *Functionality*

The AQI (Air Quality Index) is an index for reporting daily air quality. It tells you how pure or polluted the air is, and what associated health effects might be a concern. The AQI focuses on health effects one may experience within a few hours or days after breathing polluted air. 6 National Air Quality Index (NAQI) scale.

2) *Disadvantages:*

- It does not suggest preventive measures.
- It may not display the highest polluting component.
- There is a possibility of AQI measurement failure.
- It may be relying on subjective judgements.
- It may be overly generalized and unscientific.

C. *Laser Egg*

The Laser Egg is powered by highly accurate optical sensors that count individual particles in the air, and Kaiterra's innovative cloud-based calibration [5].

Sensedge: The Sensedge is provided with modular sensors that indicate performance reports and receive notifications when air changes. The sensors can be removed and replaced with no additional need of recalibration.

Flow: Two tiny molecule-capturing membranes are heated up to 250 degrees-one for VOC sensor, the other for NO2 sensor. As they capture molecules, the membranes encounter a variation in their electronic characteristics and measure the concentration.

D. *Motivation*

Our motive is to give a more advanced and transpicuous idea about air pollution which is understandable even by the common people. Generally, the existing AQI models either display the scale or percentage of concentration of pollutants in the local area using the display boards updated by the pollution control boards of the local government or a private organization. Else, we could be looking at the websites or mobile applications that provide the air quality index of a particular area along with the weather updates such as temperature, humidity etc. Mobile app Website. But the proposed model not only specifies the AQI scale and concentration of pollutants but also the highest polluting component, the sources of pollution, steps to be taken to reduce its production, precautions and preventive measures for good health.

III. PROPOSED WORK

The work is an interface where people are given three different options that are featured in three modules. The first module depends on analysis, where the user can retrieve the areas of Visakhapatnam grouped in the increasing order of the pollution along with the measures that can be followed by the pollution control committees and government to reduce it.

The second module depends on prediction where the user can retrieve of the pollution level of a particular region based on the available data of the pollutants along with reasons and effects of the pollution and the precautionary actions that can be followed by individuals.

In the third module the user can retrieve the top 10 areas with highest values of each component that can be used to identify the places with harmful pollutants among all and try to control them in the primary place with the help of the government and this can also help many pollution control organizations to control the harmful gases in various locations.

A. *Modules of proposed system*

- Collect the data of various pollutant gases and their levels in various locations in different atmosphere conditions.
- Train a model using machine learning algorithms based on the collected data.
- Build an interface that illustrates the levels of Sulphur Dioxide, Nitrogen Dioxide, Particulate Matter (PM10), Particulate Matter (PM2.5) and other pollutants.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 4
DOI: https://doi.org/10.17762/ijritcc.v11i4.6442
***Article Received**: 06 February 2023 **Revised**: 04 March 2023 **Accepted**: 18 March 2023*

- Indicate the level of pollution i.e., level-1(low), level-2(medium), level-3(high) of various areas using HTML pages by taking the pollutant values of an area as input parameters. 10 • State the precautions based on the levels of gases in air of a particular location to reduce the air pollution.
- Perform real time analysis on the pollution levels of various locations based on the pollutants.
- Display top 10 locations with highest values of each pollutant that causes air pollution.

B. *Advantages of proposed system*

- High accuracy compared to existing algorithms.
- It is highly reliable.
- True Prediction in most cases.
- Not restricted to a single scan knowledge extraction.

C. *Block Diagram*

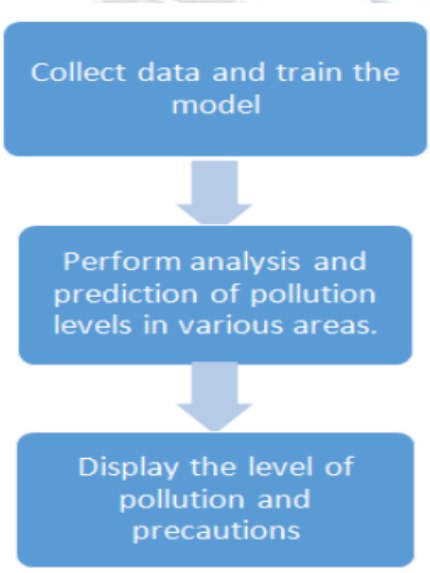

Fig1: Block diagram of Air Pollution Detection and Control System

The above diagram describes the Work Flow of the system where initially the data about the pollution levels of each pollutant component in various locations are collected. Pre-processing has to be done on the collected data and a model is trained using this data to predict the pollution level of an area depending on the percentage of given pollutants in that area. Apart from this analysis is also performed on the collected data and top 10 areas with highest values of each component are displayed to take precautionary actions accordingly.

D. *Machine Learning Algorithms*

Based on the requirements of the proposed work we have 3 modules. Several algorithms are used for the development of each module in “Air Pollution Detection and Control System”.

1) *Module-1: Prediction*

We used several algorithms for the prediction of level of pollution of each area using the pollutant values as input attributes. The algorithms used are as following:

- Logistic Regression
- Expectation Maximization (Gaussian NB)
- Neural Networks (MLP Classifier)
- Decision tree
- SVM (Support Vector Machine)
    - SVM Polynomial Kernel
    - SVM Gaussian Kernel
    - SVM Sigmoid Kernel

a) *Regression*

Regression is a machine learning technique that allows us to estimate continuous or real-valued outputs such as temperature, stock prices, and the like. It relies on a hypothesis that can be linear, quadratic, polynomial, non-linear, and so on. By utilizing independent variables, regression can model predicted values of a target to a great degree, enabling researchers to uncover relationships between variables and make predictions. Logistic regression, on the other hand, is a classification algorithm that employs regression modeling to predict the likelihood that a given data point belongs to a specific category, making it a powerful tool in supervised learning. It is a technique for classifying data that may also be used to categorise observations into a set of discrete classes. One of the most fundamental and well-known algorithms for solving a classification problem is logistic regression. Because the fundamental methodology is very similar to that of linear regression, it is known as "logistic regression." The Logit function, which is utilised in this categorization method, is where the term "Logistic" originates [8]. A statistical method known as logistic regression is used to forecast the likelihood of a binary response based on one or more independent variables. This indicates that logistic regression is used to forecast outcomes with two possible values, such as 0 or 1, pass or fail, yes or no, etc., given a set of parameters.

b) *Expectation Maximization*

In a sample where the model depends on unobserved latent variables, the expectation maximisation approach is an iterative method to determine the maximum likelihood or maximum a posteriori estimates of parameters that can be used to fill in the missing data. In the Expectation Maximization Algorithm, GaussianNB was utilised. This algorithm's goal is to look at the

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 4
DOI: https://doi.org/10.17762/ijritcc.v11i4.6442
***Article Received**: 06 February 2023 **Revised**: 04 March 2023 **Accepted**: 18 March 2023*

dataset's accessible data, estimate (guess) the values of the missing data, and then construct the full dataset once the expectation step is applied to update the parameters until convergence. The concentration values of attributes in the dataset of the air quality index are typically among the dataset's missing values. Pre-processed values are used to fill in the missing values in this data. [9]. A unique variety of NB algorithm is the Gaussian Naive Bayes algorithm. When the features contain continuous values, it is specially employed. Additionally, it is presumed that all of the characteristics follow a normal Gaussian distribution. In Gaussian Naive Bayes, it is presumed that the continuous values connected to each feature are dispersed according to a gaussian distribution. Normal distribution is another name for a gaussian distribution. Based on the naive assumptions about concentration values of the variables that happen to be the features, we are using this classifier to classify the level of pollution of a certain area.

c) Neural Network

The goal of a neural network is to identify underlying links in a batch of data using a process that closely resembles how the human brain thinks [10]. In this context, neural networks are systems of neurons that can be either organic or synthetic in origin. In a neural network, MLP Classifier was utilised. Multi-Layer Perceptron Classifier When data is linearly separable into 'n' dimensions, where 'n' is the number of features in the dataset, a single layer perceptron can solve straightforward problems. The single-layer perceptron's accuracy drastically declines in non-linearly separable data. On the other hand, multilayer perceptrons can effectively process non-linearly separable input. Multilayer perceptrons, or artificial neural networks more generally, are a combination of an input layer, one or more hidden layers, and an output layer.

In this paper, a module is about making predictions using the attributes SO2, NO2, PM2.5, PM10 from the dataset. The above attributes are used as input parameters with which we train the model. Depending on these parameters predictions are made regarding the level of pollution it belongs to. Along with the level of pollution, measures to control the pollution are also suggested depending on each input attribute levels.

d) Decision tree

A model that predicts the value of a target variable by learning straightforward choice rules derived from the data features and identifying ways to split a data set based on various conditions [11] is created using the decision tree algorithm, a non-parametric supervised learning technique. The decision-making rules often take the form of if-then sentences. A decision tree can be used to openly and aesthetically express making decisions. With a straightforward linear decision surface, decision trees are utilised in non-linear decision making. The examples are categorised by sorting them from the tree's root to a leaf node, with the leaf node giving the example its categorization. This process is recursive in nature and is repeated for every sub-tree rooted at the new nodes. The data available to train the decision tree will be split into a training set and test set and trees with various maximum depths (length of the longest path from the tree root to a leaf) will be created based on the training set and tested against the test set. Classification decision tree: A classification tree labels, records, and assigns variables to discrete classes. It can also provide a measure of confidence that the classification is correct. It is built through a process known as binary recursive partitioning. This is an iterative process of splitting the data into partitions, and then splitting it up further on each of the branches.

e) Support Vector Machine

Support Vector Machine (SVM) is a supervised machine learning algorithm which can be used for both classification and regression challenges [12]. However, it is mostly used in classification problems. In this algorithm, we plot each data item as a point in n-dimensional space (where n is number of features you have) with the value of each feature being the value of a particular coordinate. Then, we perform classification by finding the hyper-plane that differentiates the two classes very well. Support Vectors are simply the co-ordinates of individual observation. Support Vector Machine is a frontier which best segregates the two classes (hyper-plane/ line). In our work we are using SVM's for classification of areas to its pollution levels.

2) Module-2: Analysis

We used algorithms for analyzing and categorizing the areas depending on their level of pollution and displaying them along with the precautions that can be followed in the regions. Following algorithms are used for analysis 1. K-Means Clustering 2. DBSCAN

Clustering is the task of dividing the population or data points into a number of groups such that data points in the same groups are more similar to other data points in the same group and dissimilar to the data points in other groups.

a) K-Means Clustering

K-means clustering is one of the simplest and most popular partitional based unsupervised machine learning algorithms. Typically, unsupervised algorithms make inferences from datasets using only input vectors without referring to known or labeled outcomes. A cluster refers to a collection of data points aggregated together because of certain similarities. We'll define a target number k, which refers to the number of centroids we need in the dataset. A centroid is the imaginary or real location representing the center of the cluster. In other words, the K-means algorithm identifies k number of centroids, and then allocates every data point to the nearest cluster, while keeping

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 4
DOI: https://doi.org/10.17762/ijritcc.v11i4.6442
***Article Received**: 06 February 2023 **Revised**: 04 March 2023 **Accepted**: 18 March 2023*

the centroids as small as possible. The 'means' in the K-means refers to averaging of the data; that is, finding the centroid.

The way k-means algorithm works is as follows: 1. Specify the number of clusters K. 2. Initialize centroids by first shuffling the dataset and then randomly selecting K data points for the centroids without replacement. 3. Keep iterating until there is no change to the centroids. i.e., assignment of data points to clusters isn't changing. The approach K-means follows to solve problem is called Expectation maximization In this paper, one module is about analyzing the data and grouping them based on their pollution level. For this, we used K-Means clustering with input attributes SO2, NO2, PM 2.5, PM10. These are the components primly responsible for pollution. Depending on these input attributes various locations are clustered and displayed based on their level of pollution. Along with these we also display the levels of the four input attributes and suggest measures to control them.

b) DBSCAN

Density-based spatial clustering of applications with noise (DBSCAN) is a well-known density-based data clustering algorithm that is commonly used in data mining and machine learning. Based on a set of points, DBSCAN groups together points that are close to each other based on a distance measurement (usually Euclidean distance) and a minimum number of points and hence generates natural and true clusters that exist in data. It also marks as outliers the points that are in low-density regions.

Parameters: The DBSCAN algorithm basically requires two parameters:

(i) epsilon (eps): It specifies how close points should be to each other to be considered a part of a cluster. It means that if the distance between two points is lower or equal to this value (eps), these points are considered neighbors.

(ii) minpoints: The minimum number of points to form a dense region. For example, if we set the minpoints parameter as 5, then we need at least 5 points to form a dense region. Parameter estimation: The parameter estimation is a problem for every data mining task. To choose good parameters we need to understand how they are used and have at least a basic previous knowledge about the data set that will be used.

If the eps value chosen is too small, a large part of the data will not be clustered. It will be considered outliers because don't satisfy the number of points to create a dense region. On the other hand, if the value that was chosen is too high, clusters will merge, and the majority of objects will be in the same cluster. The eps should be chosen based on the distance of the dataset (we can use a k-distance graph to find it), but in general small eps values are preferable.

Minpoints: As a general rule, a minimum of minpoints can be derived from a number of dimensions (D) in the data set, as minpoints ≥ D + 1. Larger values are usually better for data sets with noise and will form more significant clusters. The minimum value for the minpoints must be 3, but the larger the data set, the larger the minpoints value that should be chosen. In this paper, one module is about analyzing the data and grouping them based on their pollution level. For this, we used DBSCAN clustering with input attributes SO2, NO2, PM 2.5, PM10 [6-7]. These are the components primly responsible for pollution. We gave parameter minPoints with the number of the clusters required and the range considered as parameter eps.

Depending on these input attributes and parameters, various locations are clustered and displayed based on their level of pollution. Along with these we also display the levels of the four input attributes and suggest measures to control them. The algorithms are developed on Jupyter notebook. It is developed using python machine learning programming which is an extension of the python language which is specially designed for deep learning algorithm development. It is developed for classifying the level of pollution for any given location provided data required.

IV. RESULT ANALYSIS

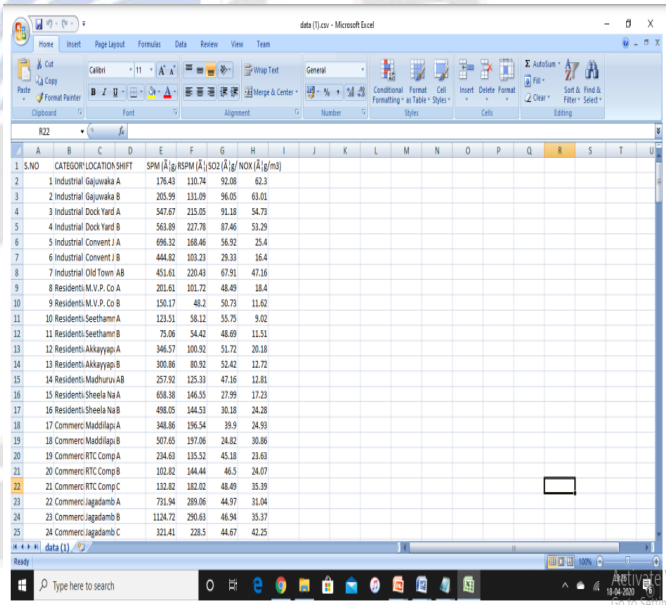

Fig2: Sample dataset used for Analysis.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 4
DOI: https://doi.org/10.17762/ijritcc.v11i4.6442
***Article Received**: 06 February 2023 **Revised**: 04 March 2023 **Accepted**: 18 March 2023*

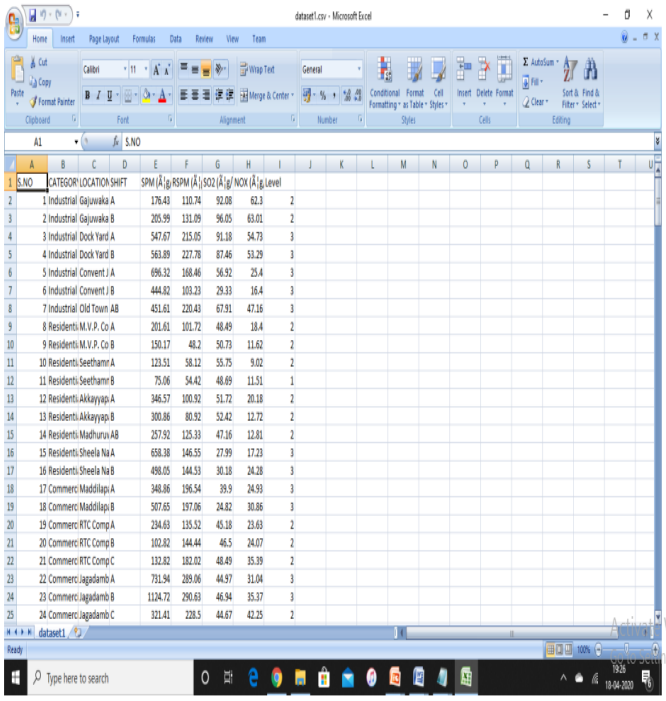

Fig3: Sample dataset used for Prediction.

A. APPLYING DATASET TO ML ALGORITHMS

All the components that cause air pollution are read using different sensors in real time and collected as a working dataset. The dataset is pre-processed so as to remove missing data and anomalies, handling outliers and applied machine learning algorithms to predict levels of pollution prevailing in various places of the city. The performance of ML algorithms is evaluated and compared based on Accuracy metric. Among all the algorithms, it is found that decision tree classifier is able to predict better which is witnessed by better accuracy score.

B. ACCURACY SCORE

Accuracy score is defined as the ratio of number of correct predictions to the total number of input samples.

1) Logistic Regression Algorithm

Logistic Regression is one of the most popular machine learning algorithms, which comes under supervised learning. It is used for predicting the categorical dependent variable using a given set of independent variables. In this module the user is allowed to enter the values of the four pollutants value i.e., PM10, PM2.5, NO2, SO2 of and the level of pollution of the particular area is predicted. The accuracy obtained is 90.9%.

2) Expectation Maximization Algorithm

Maximum Likelihood estimation estimates a datasets density by searching across probability distributions and their parameters. Using Maximum likelihood becomes intractable if there are variables that interact with those in the dataset but were hidden or unobserved. The observed accuracy for the dataset is 93.9%.

3) Neural Network

MLP (Multi-Layer Perceptron) classifier from Neural Network is used to implement prediction. Artificial neural networks, also known as neural networks or simulated neural networks, are a subset of machine learning that form the core of deep learning algorithms. Their name and structure are based on the human brain, which imitates the way biological neurons communicate with each other. These networks use training data to improve their accuracy over time. With fine-tuned learning algorithms, we can use them in computer science and artificial intelligence to swiftly classify and cluster data. The accuracy obtained is 68.7%.

4) Decision Tree Algorithm

Decision Tree is a method used for both classification and regression issues, known as supervised learning. Nonetheless, it is widely used for classification purposes. It is a type of classifier based on a tree structure where the data features are represented by internal nodes, the decision rules by branches, and each outcome by a leaf node. Within a Decision tree, there are two main types of nodes, namely the Decision Node and Leaf Node. Decision nodes serve to make a decision and have multiple branches, while Leaf nodes represent the output of those decisions and do not have any subsequent branches. The accuracy obtained is 98.4%.

5) Support Vector Machine

SVM, short for Support Vector Machine, is a commonly used algorithm in Supervised Learning that is capable of tackling both Classification and Regression problems. However, it is mainly utilized for Classification purposes in the field of Machine Learning. The aim of the SVM algorithm is to construct an optimal line or decision boundary that can divide an n-dimensional space into distinct categories, allowing for easy categorization of new data points in the future. This optimal boundary is referred to as a hyperplane. The Support Vector Machine (SVM) algorithm selects the extreme vectors, also known as support vectors, to create the hyperplane. The algorithm is named after these support vectors. The accuracy obtained is 98%.

6) K-Means Clustering Algorithm

K-Means Clustering Algorithm is used for data analysis. Using K-Means Clustering the areas in the dataset are grouped into clusters where each cluster has the areas with similar level of pollution and pollutant components. Using this we displayed the areas in the dataset in the increasing order of pollution. Along

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 4
DOI: https://doi.org/10.17762/ijritcc.v11i4.6442
***Article Received**: 06 February 2023 **Revised**: 04 March 2023 **Accepted**: 18 March 2023*

with this we also displayed the top 10 areas with the highest values of each pollutant.

Table1: Comparison of Accuracy Scores of Different Machine Learning Algorithms

Algorithm	**Accuracy Score**
Logistic Regression	90.9%
Expectation Maximization	93.9%
Neural Network	68.7%
Decision Tree Induction	**98.4%**
Support Vector Machine	98%

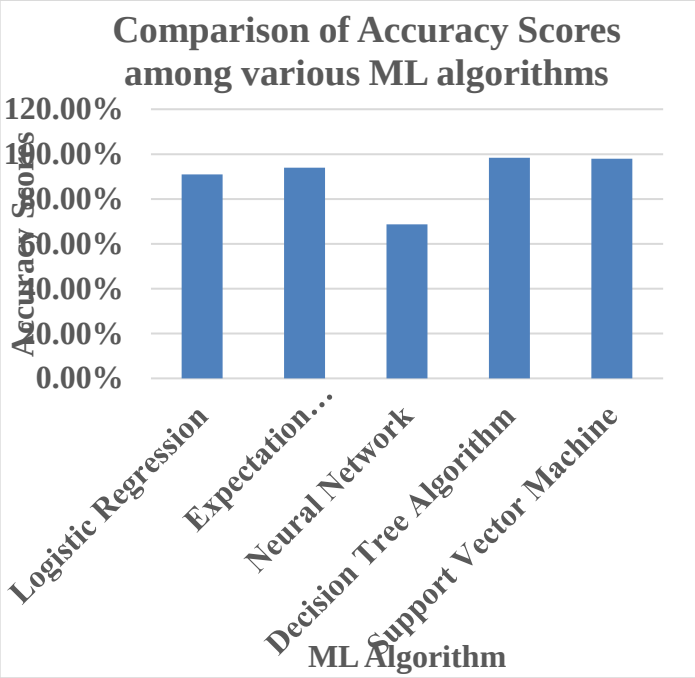

Fig4: Comparison of Accuracy Scores among various ML algorithms
Footnotes

V. CONCLUSION

Many governments and organizations all over the world are recognizing the alarming rates of air pollution and its consequences. New technologies and advanced methods are rolling up so as to create awareness and enlighten the need of abating the air pollution as much as possible. Our work is intended to outstretch the issue which can be apprehended even by the common people. We also suggest the best ways to minimize air pollution on an individual level by integrating the sources and factors of air pollution along with the solutions to curtail it. The accuracy of different Machine learning algorithms is compared and the Decision tree Induction has given better accuracy. This work may be extended to apply deep learning techniques as future scope.

REFERENCES

[1] Soumyadeep Sur;Rohit Ghosal;Rittik Mondal, "Air Pollution Hotspot Identification and Pollution Level Prediction in the City of Delhi", 2020 IEEE 1st International Conference for Convergence in Engineering (ICCE), 2020.

[2] Anandharajan TRV, Vignajeth KK, Hariharan GA, Jijendiran R. Identification of outliers in pollution concentration levels using anomaly detection. In: 2016 International Conference on Computational Techniques in Information and Communication Technologies (ICCTICT). New Delhi (2016). p. 433–8. doi: 10.1109/ICCTICT.2016.7514620.

[3] Wang P, Chen K, Zhu S, Wang P, Zhang H. Severe air pollution events not avoided by reduced anthropogenic activities during COVID-19 outbreak. Resour Conserva Recycl. (2020) 158:104814. doi: 10.1016/j.resconrec.2020.104814.

[4] A. Gupta, A. Gupta, K. Jain and S. Gupta, "Noise pollution and impact on children health", The Indian Journal of Pediatrics, vol. 85, no. 4, pp. 300-306, 2018.

[5] S. K. Shah, Z. Tariq, J. Lee and Y. Lee, "Real-Time Machine Learning for Air Quality and Environmental Noise Detection," 2020 IEEE International Conference on Big Data (Big Data), 2020, pp. 3506-3515, doi: 10.1109/BigData50022.2020.9377939.

[6] S. Zhang, X. Li, Y. Li and J. Mei, "Prediction of Urban PM2.5 Concentration Based on Wavelet Neural Network," 2018 Chinese Control And Decision Conference (CCDC), 2018, pp. 5514-5519, doi: 10.1109/CCDC.2018.8408092.

[7] B. R. Subrahmanyam, A. Gautam Singh and P. Tiwari, "Air Purification System for Street Level Air Pollution and Roadside Air Pollution," 2018 International Conference on Computing, Power and Communication Technologies (GUCON), 2018, pp. 518-522, doi: 10.1109/GUCON.2018.8674934.

[8] S. Jiyal and R. K. Saini, "Prediction and Monitoring of Air Pollution Using Internet of Things (IoT)," 2020 Sixth International Conference on Parallel, Distributed and Grid Computing (PDGC), 2020, pp. 57-60, doi: 10.1109/PDGC50313.2020.9315831.

[9] H. ALTINÇÖP and A. B. OKTAY, "Air Pollution Forecasting with Random Forest Time Series Analysis," 2018 International Conference on Artificial Intelligence and Data Processing (IDAP), 2018, pp. 1-5, doi: 10.1109/IDAP.2018.8620768.

[10] M. Etkind, "Air pollution-an overview," IEE Colloquium on Pollution of Land, Sea and Air: An Overview for Engineers, 1995, pp. 10/1-10/6, doi: 10.1049/ic:19951017.

[11] T. Hastie, R. Tibshirani, J. Friedman. The Elements of Statistical Learning, 2nd edition, 2008.

[12] Dietel and Dietel : "Internet and World Wide Web – How to Program", 5th Edition, PHI/Pearson Education, 2011.