IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Survey Paper On Different Types Of PredictionAlgorithm For Air Quality Index And ComparativeStudy On Types Of algorithms Used

[1] Diana George; [2] Jayakrishnan.R; [3] Sam.G.Benjamin; [4] ManoLekshmi.I;

[1] CSE Department, BMCE
APJ Abdul Kalam Technological University, Sasthamcotta, Kerala, India.

[2] CSE Department, BMCE
APJ Abdul Kalam Technological University, Sasthamcotta, Kerala, India.

[3] CSE Department, BMCE
APJ Abdul Kalam Technological University, Sasthamcotta, Kerala, India.

[4] CSE Department, BMCE
APJ Abdul Kalam Technological University, Sasthamcotta, Kerala, India.

ABSTRACT

Air pollution is considered to be a major concern of the day.Many researches has been demonstrated on various fields on regarding air pollution and seeking a stable solution for the problem.Early prediction of air quality and to make it avail and aware to the society being an important aspect.This survey paper present about various machine learning algorithms for Air Quality Index Prediction and in-depth analysis of most relevant machine leaning algorithms available on the literature survey .The discussed techniques are Neural Networks, fuzzy a system, Support Vector Machine, Support Vector Machine for regression, fuzzy logic, Decision Trees, K-Nearest Neighbor,Naive Bayes,Random forest and Decision Trees.We are working on the algorithm with best accuracy.This will help the Government ,environmentalist and the general public to make the plans best.

Keywords— Machine Learning, Prediction, Classification Technique, Decision Tree, Accuracy,Random forest.

I. INTRODUCTION

Air quality has increasingly concern from environment managers and citizens all over the world.New tools continue to evolve to raise quality of air and awareness worldwide.Continuous improvements in air quality monitoring are occurring along with the advancements of smart cities.As a natural consequence, air pollution forecasting has become a hot topic,aiming the prediction of the atmospheric composition of pollutants at a given time and location.With an accurate air quality forecast, individuals can take remedial measures to reduce the possible adverse effects of air pollution on their health.Enhancing air quality forecasting is one of the prime targets for the society for reducing the adverse impact of it on directly to humans and other living and non living creations.Different algorithms and tools such as Neural

Networks, fuzzy a system, Support Vector Machine, Support Vector Machine for regression, fuzzy logic, Decision Trees, K-Nearest Neighbour has been used previously for different use cases.

II. LITERATURE STUDY

Usha mahalingam, kirthiga elangovan, himanshu dobhal, chocko valliappa, sindhu shrestha, and giriprasad kedam [1] proposed real time air quality monitoring system using back propagation ANN and Support Vector Machine algorithm, which is a supervised learning algorithm.The model considered several parameters.The model deploy pollution monitoring sensor for monitoring air.The measured value is provided to an artificial neural network which is a natural model that handle information and make decision like how it fascinating the human brain.At the first stage it predicts the future data and in the second stage,feeding the artificial neural networks results to support vector machines that classified data into more classes.

Chuanting zhang, dongfeng yuan [2] proposed a method to analyse and predict real time air quality Index based on the spark implementation of distributed random forest algorithm. The proposed method is implemented and it shows the methods is very accurate in predicting the concentration level of pm 2.5 with an average accuracy greater than 70 %. The model works by a loading data in the database and select the data according to the need of prediction algorithm.For the working of the model create the spark context object for running application under distributed platform application.It primarily need three parts:master node ,cluster manager and worker node. master node communicate with worker node through cluster manager .The driver node program present in the master note take care of data partition ,node statistics computing and execute information.After that training and predictions were constructed using constructive random forest tree.

Dr .j. Gnana jeslin[3] propose a method for predicting air quality index using multi variable regression method. In this proposed model calculating the individual index of pollutant for every available data points using AQI formulae. To know about air quality level in particular location using gradient descent and box plot analysis and predict the activity of upcoming year using this present values air quality values.

Kostandina veljanovska1 and angel dimoski [4] proposed several machine learning algorithm for predicting air quality index.In the proposed methods they utilise this the capacity of neural networks to mimicking the human brain by interconnecting large number of neurons positioned as several layers which work together to solve complex problems.They proposed another supervised learning algorithms among them KNN and Decision Tree are the most prominent. In knn output of air quality categorisation is a member of range of classes.The quality of air is classified according to the votes of most of the nearest neighbour.They also implement decision tree algorithm for predicting fine AQI.In this trees can be constructed and used for visual or explicit presentation of decision or for making decisions.The working tree at each individual node is test for some attributes and branch the tree based on the corresponding inner states. Tree expand at each level and make predictions at the leaf node.

Venkat rao pasupuleti and uhasri [5] proposed a device that predicts air quality with the help of past pollutant rate.For for that different regression technique are utilized .Based on the independent variables regressor perform target prediction which is most likely used for find out relationship between variables and forecasting .for better performance they suggest to combine a decision tree along with progressive regression.

Shajulin benedict [6] proposed a cloud oriented approach to predict air quality index.The collected data are transmitted to powerful sensor gateways and sensed the data to cloud database server for storing.The model deploy random forest algorithm.The algorithm construct trees during the bagging process ,so that the air quality parameter value were dispersed with high variance.During the prediction stage random forest create regression point based on independent variables of the test set and examine the trees closely based on incoming air quality values.

Sachin bhoite [7] the proposed a supervised machine learning approach for linear regression to determine the air quality.The developed machine perform two important task firstly it detect the pollutant based on given atmospheric values and secondly it predict the level of pollutant for a particular date. Here,Autoregression is a time series model that made prediction based on the previous observation as input to regression equation, to predict the value at the next time step.

Mauro castelli , fabiana martins clemente, alespopovic and leonardo vanneschi [8], propose a model for forecast pollution level and predict corresponding air quality index value using support vector machine regression technique. support vector machine where mainly applicable to classification type problem and the objective of the algorithm is for finding an optimal separating hyperplane between classes and the points lying on class boundaries were called as support vectors. In this model data points scattered on higher dimensional space, that is known linearly separable points converted to linearly separable using kernel functions.for the purpose, here they employ a radial bias kernel function.

Jose antonio moscoso-l´opez1(b),daniel urda, javier gonz´alez-enrique, juan jesus ruiz-aguilar1, and ignacio j. Turias[9] describe a model to predict the air quality index using long short term memory neural network(LSTM).The observed data concentration were converted into numerical index by using linear interpolation. For the mentioned long short term memory neural network, network topologies include a sequence of input layer, a LSTM layer , a dropout layer, a fully connected layer and a regression output layer.Number of neurons present in the hidden layer the initial learning rate and the drop out layer were learned by performing Bayesian optimisation. The forecasting phase where divided into two stages, in first stage pollutant concentrations were obtained independently for each pollutant and at the second stage future air quality index was obtained.

Kehu,Ashfaqur,Rahman,Hari Bhrugubanda and vijay sivaraman developed a machine learning model hazeest ,that combines sparse fixed station data ,that trains a machine learning model using existing historical pollution data from fixed and mobile sensors, and uses the model to estimates the air pollution surface for any day/time in specific places.For the aim,they select nine input features to feed in to the model.The target is to estimate pollutant value for the particular time and location.In the training and implementation phase,the regression models were implemented.The proposed system comprises three major phases.They are data collection, cloud server and web-based application for users to visualize the pollution contour map data from the sensor network is collected and uploaded using node sensors and a mobile application .

Agar V Belavadia,Sreenidhi Rajagopala,Ranjani Ra , Rajasekar Mohana[11] conceptualised a scalable architecture to processed real time air pollutant concentration data from two sources.The first being a wireless sensor topology that monitor and sends pollutant concentrations to a server.The second, is the real-time air quality from an open source initiative.Both sources provide average concentration ions of various air pollutants on an hourly basis. Due to the efficiency of long short-term memory (LSTM) recurrent neural network in case of dealing with time series data it is finalized for the purpose.A long short-term memory,recurrent neural network architecture is used for the forecasting model. In the training process minimizes the mean squared error between the predicted value and the actual value.The crucial network weights are updated in an iterative manner based on the training data.The learning rate is separately maintained for each network parameter and it is adapted as learning happens.

Andreaslepperod,haithanhnguyen,sigmundakselsen,leendert wienhofen[12] proposed an air quality monitoring and forecast system using nb-iot sensors and machine learning principles.This system uses micro-sensor devices at stationary locations as well as on the moving vehicles to provide a comprehensive overview of air quality in the city. For the proposed project they modelled a device consisting of a board with sensors and a communication modem together.Multi layer perceptron and random forest had been used to get reliable results within air quality prediction problems.

.

TABLE I. A COMPARATIVE STUDY OF VARIOUS ALGORITHMS IN LITERATURE REVIEW

YEAR	AUTHOR	PURPOSE	METHODS MENTIONED	INFERENCE
2016	Usha Mahalingam, Kirthiga Elangovan, Himanshu Dobhal, Chocko Valliappa, Sindhu Shrestha, and Giriprasad Kedam	A Machine Learning Model For Air Quality Prediction For Smart Cities	ANN and SVM	ANN is more efficient than SVM . Because SVM posses following disadvantages SVM works by putting data points, above and below the classifying hyperplane there is no probabilistic explanation for the classification Accuracy depends upon the kernel functions.[13]
2015	Chuanting Zhang, Dongfeng Yuan.	Fast Fine-Grained Air Quality Index Level Prediction Using Random Forestalgorithm On Cluster Computing Of Spark	Spark Implementation of Random forest	Random forest is a versatile algorithm,it can be used for both regression and classification task.. In random forest the default hyper parameters it uses often produce a good prediction result. Understanding the hyper parameters is pretty straightforward.[14]
2019	DR .J. Gnana Jeslin p	Indian Air Quality Prediction And Analysis Using Machine Learning	Regression methodologies.	Advantages of Regression: Linear Regression performs well when the data set is linearly separable. Easier to implement, interpret and very efficient to train. It is prone to over-fitting but it can be easily avoided.
2018	Kostandina Veljanovska1 and Angel Dimosk	Machine Learning Algorithms In Air Quality Index Prediction	ANN,KNN and DT	DT is more efficient as compared to ANN and KNN. Disadvantages of KNN: KNN need feature scaling otherwise it may generate wrong predictions. KNN is sensitive to noise in the dataset. Disadvantages of ANN: Determination of proper network structure is doubtful. The duration of the network is not determined previously.
2017	ShajulinBenedict,	Revenue Oriented Air Quality Predictionmicroservices For Smart Cities	Cloud oriented Random Forest approach.	Highly scalable and powerful. More accurate and efficient prediction by overcoming machine constraints.

2021	Jose Antonio Moscoso-L´opez1(B) , Daniel Urda, Javier Gonz´alez-Enrique, Juan Jesus Ruiz-Aguilar1, and Ignacio J. Turias,Hourly	Air Quality Index (AQI) Forecasting Using Machine Learning Methods	LSTM	LSTM overcome the issue of ANN by creating both a short-term and a long-term memory component.
2017	KeHu, Ashfaqur Rahman, Hari Bhrugubanda,and VijaySivaraman	Hazeest Machine Learning Based Metropolitan Air Pollution Estimation From Fixed and Mobile Sensors	Sensor based Linear Regression.	It helps to find Air quality index based on live data. Root level implementation is easy. .
2020	Doreswamy, Harishkumar K S1, Yogesh KM, Ibrahim Ga	Forecasting Air Pollution Particulate Matter (PM2.5) Using Machine Learning Regression Models	MLP Regression Decision Tree Random Forest Gradient Boosting	Random forest is best among all other methods. Disadvantage of MLP: The number of total parameters can grow to very high which is inefficient because there is redundancy in such high dimensions. Another disadvantage is that it disregards spatial information. It takes flattened vectors as inputs.
2016	Ilias Bougoudis, Konstantinos Demertzi	Semi-Supervised Hybrid Modeling of Atmospheric Pollution in Urban centers	Naïve Bayesian algorithm Yatsi algorithm	Naive Bayes is more efficient than yatsi because yatsi is based on a sample of few pre-classified data vectors, something that incorporates the hidden knowledge and the correlations between the features.
2020	Agar V Belavadia,,Sreenidhi Rajagopala , Ranjani Ra , Rajasekar Mohana	Air Quality Forecasting Using Lstm Rnn And Wireless Sensor Networks	LSTM And RNN	RNN and LSTM more efficient and faster as compared to traditional time series models. Both deal with larger data sets better. RNN combines its feed back ability with LSTM made to solve the traditional issues of memory capability.
2017	ShajulinBenedict,	Revenue Oriented Air Quality Prediction microservices For Smart Cities	Cloud oriented Random Forest approach.	Highly scalable and powerful. More accurate and efficient prediction by overcoming machine constraints.

III. CONCLUSION AND FUTURE WORK

We have summarized different types of machine learning algorithms for prediction of Air Quality Index. We elaborated various machine learning algorithms and worked towards finding the best algorithm by analysing their features. Every algorithm has given different result in different situations. Further it is analysed a marginal accuracy is achieved by Random forest among others. Hence more complex models are needed to increase the accuracy of predicting early Air quality index following the current trend. In future we will propose methodology for early prediction of Air quality index with high accuracy and minimum cost and complexity.

REFERENCE

[1]. Usha Mahalingam, Kirthiga Elangovan, Himanshu Dobhal, Chocko Valliappa, Sindhu Shrestha, and Giriprasad Kedam, A Machine Learning Model For Air Quality Prediction For Smart Cities,2016.

[2]. Chuanting Zhang, Dongfeng Yuan,FAst Fine-Grained Air Quality Index Level Prediction Using Random Forestalgorithm On Cluster Computing Of Spark,2015.

[3].DR .J. Gnana Jeslin p,Indian Air Quality Prediction And Analysis Using Machine Learning,2019.

[4]..Kostandina Veljanovska1 and Angel Dimoski,Machine Learning Algorithms In Air Quality Index Prediction,2018.

[5].Venkat Rao Pasupuleti and Uhasri , Air Quality Prediction Of Data Log By Machine Learning,2020.

[6].ShajulinBenedict,Revenue Oriented Air Quality Predictionmicroservices For Smart Cities,2017

[7].Sachin Bhoite,Air Quality Prediction Using Machine Learning Algorithm,2019.

[8].Mauro Castelli , Fabiana Martins Clemente, AlesPopovic and Leonardo Vanneschi ,A Machine Learning Approach To Predict Air Quality In California,2020

[9].Jose Antonio Moscoso-L´opez1(B) , Daniel Urda, Javier Gonz´alez-Enrique, Juan Jesus Ruiz-Aguilar1, and Ignacio J. Turias,Hourly Air Quality Index (Aqi) Forecasting Using Machine Learning Methods ,2021.

[10].KeHu,AshfaqurRahman,HariBhrugubanda,andVijaySivaraman ,Hazeest:Machine Learning Based Metropolitan Air Pollution Estimation From Fixedand Mobile Sensors,2017

[11].Agar V Belavadia,,Sreenidhi Rajagopala , Ranjani Ra , Rajasekar Mohana,Air Quality Forecasting Using Lstm Rnn And Wireless Sensor Networks,2020.

[12].Andreas Lepperød,Hai Thanh Nguyen,Sigmund Akselsen, Leendert Wienhofen,Air Quality Monitor And Forecast In Norway Using Nb-Iot And Machine Learning,2019.

[13].https://dhirajkumarblog.medium.com/top-4-advantages-and-disadvantages-of-support-vector-machine-or-svm-

[14].https://builtin.com/data-science/random-forest-algorithm

[15]Zhang, Y., Bocquet, M., Mallet, V., Seigneur, C., and Baklanov, A. Real-time air quality forecasting, part I: History, techniques, and current status,

[16].G. Petnehazi, “Recurrent neural networks for time series forecasting.

[17]..Real time Air Quality Index data.gov.in/catalog/real-time-air-quality-index.

[18].S. R. Jino Ramson and D. Jackuline Moni. ”Applications of wireless sensor networks — A survey”. International Conference on Innovations in Electrical, Electronics, Instrumentation and Media Technology (ICEEIMT), 2017.

[19].Adenugba, et al. ”Smart irrigation system for environmental sustainability in Africa: An Internet of Everything (IoE) approach”. Mathematical Biosciences and Engineering, 16(5), 5490-5503. doi:10.3934/mbe.2019273