

## ABSTRACT

The prediction of accurate pollutant in any dynamic nature provides different benefits for humans or other living creatures on the planet. The machine learning process and classification technique are developed for forecasting the air pollution that attains enhanced result on cloud computing environment. The accurate air pollutant forecasting performance of data is an important issue which affects the air quality. During the prediction of air pollution, feature selection and data classification is the most considerable task for accurate forecasting results. The performance of feature selection helps to give enhanced results in prediction with higher accuracy. Here, irrelevant features are eliminated effectively. By removing negative features of air data, people's health conditions are managed to prevent the dangerous impact.

Several machine learning methods are performed on air data for estimating pollution forecasting performance. But the accurate forecasting with minimum time and minimum memory consumption is a challenging issue. The accurate performance of pollution prediction is an important goal in an environmental issue. Accurate pollution forecasting is a demanding problem to predict air quality index at particulate levels. Hence, three different models, namely, Linear Regression and Multiclass Support Vector (LR-MSV) model 1, Bilateral Transformative Broken Stick Regression and Quadratic Weighted Emphasis Boost Classification (BTBSR-QWEBC) model 2 and Discretized Regression and Least Square Support Vector (DR-LSSV) model 3 are developed with various air quality data from dataset to attain efficient air pollution forecasting for monitoring and controlling the air pollution.

The LR-MSV model is proposed at first for forecasting air pollution with accurate results. The performance of pre-processing is handled on each input data through sliding window to remove noise data for reducing time of classification. Subsequently, LRC (LRC) is utilized to select the relevant features of data. Following this, the correlation coefficient is estimated for selecting the relevant feature subset. With selected features, multiclass support vector classification is developed to classify the data with accurate result of pollution forecasting. From classification result, air pollution prediction is obtained with higher accuracy and minimized error. The accuracy level can still be

maximized, and also, the memory consumption for implementation can also be considered for air pollution forecasting. Hence, model 2 is developed to give an efficient solution.

An efficient BTBSR-QWEBC model is designed with Otsuka Indexive Brokenstick Regression (OIBR) and Quadratic Weighted Emphasis Boost (QWEB) technique to forecast the air pollution with lesser time consumption. Initially, pre-processing is presented, using wavelet transform, to eliminate the noise. Next, OIBR is implied to select significant data for classification. Finally, classification is performed with QWEB technique. Here, weighted function attains data classification with minimized quadratic error. Hence, forecasting air pollution data with minimum error rate and memory consumption is obtained.

Finally, to provide more enhancement than model 1 and model 2, DR-LSSV model is developed based on Constrained Maximum Likelihood Linear Regression (CMLLR) - based feature selection and Concordance Correlative Least Square Support Vector (CCLSSV) classification to perform air pollution prediction. The number of air quality data is gathered from dataset and specified to the input unit. In the first hidden unit, Discretized Hartley Transformation (DHT) -based data pre-processing is performed for removing noise. After that, the relevant feature selection is performed using CMLLR technique at third hidden layer. With selected features, CCLSSV is applied to obtain data classification results in the output unit. It helps to forecast the air pollution more accurately than model 1 and model 2, with lesser error rate and minimized time consumption.

The simulation work on different LR-MSV, BTBSR-QWEBC and DR-LSSV models are carried out by implementing in Java JDK 1.8 language. For experimental work, air quality India dataset is taken from air-quality-data-in-India at Kaggle. Here, different number of air data with different features is considered for forecasting air pollution with higher accuracy. From the simulation results, proposed models attain higher accuracy on pollution forecasting with minimum time and memory consumption.