CHAPTER 7

CONCLUSION AND FUTURE WORK

7.1 Conclusion

Machine learning-based classification models are developed for forecasting air pollution to prevent health issues in urban metropolises. In air quality databases, accurate air pollution forecast is the consistent problem, and the various pollutants affect the air quality. Air pollution is a considerable problem in all developed and developing countries because of fast industrialization. Thus, air quality management is important to forecast pollution of the environment. However, the early prediction of pollutant data in air databases was still a demanding problem. However, accuracy and time complexity of data forecasting was not at the required level. In the current study, there are three different models developed for achieving enhanced accuracy on air pollution forecasting with better performance level.

Initially, the LR-MSV model was introduced based on data pre-processing, feature selection, and classification. The pre-processing performs noise data removal process using sliding window. The linear regression function is carried out next to select relevant features. Following this, multiclass support vector classification is utilized to classify the air data based on air quality index value. Thus, it provides an accurate classification with minimum error and higher pollution forecasting accuracy.

Next, the BTBSR-QWEBC model is developed to improve air pollution prediction accuracy based on data classification. Data pre-processing is used to eliminate noise data from the dataset. With obtained pre-processed data, Otsuka Indexive Broken-stick regression process is applied to extract significant relevant features. The similarity coefficient value is estimated between data features to identify relevant and irrelevant features. Based on selected features, classification is performed using weighted emphasis boost classification process. Additionally, weighted emphasis function is carried on data classification result to minimize the quadratic error. Thus, it enhances accuracy and reduces time with memory consumption on forecasting air pollutant data.

Finally, the DR-LSSV model is introduced for pollution monitoring on air quality data from dataset. The discretized Hartley transformation is applied for identifying noise data and removing data. With obtained pre-processed data, likelihood linear regression process is applied for identifying relevant features for data classification. The estimation of target feature projection matching helps to identify features for better results. A concordance correlative least square support vector is applied to classify data for forecasting air pollutant data. As a result, the proposed model minimizes the error and improves the accuracy of pollution forecasting.

The experimental evaluation of the proposed LR-MSV model, BTBSR-QWEBC model, and DR-LSSV model is carried out by using JAVA JDK 1.8 language. For conducting experimental work, air quality India dataset is considered and taken from https://www.kaggle.com/rohanrao/air-quality-data-in-india. From dataset, with various air quality data in the range of 20,000 to 2,00,000, data is considered for experimental purpose. The experimental result shows that the proposed model achieves improved air pollution forecasting accuracy by classifying data with minimum time and error rate.

7.2 Future Work

The proposed models are designed for achieving effective performance for forecasting air pollutant data. Some relevant features are missed during the selection process due to the presence of numerous air data on environment. Thus, future work may be developed with advanced feature selection techniques to attain an enhanced result. When the number of features from dataset is varied, it failed to achieve better performance of classification. It leads higher time to classify data for efficient performance prediction. A future enhancement will focus on overcoming the proposed limitations with minimum performance forecasting time by classifying data. In addition, future work may be extended to provide better results of air pollutant data classification with reduced memory consumption at an early stage of development. Then in future, Product based implementation can be carried out by using sensors. More significantly, the predicted data can be stored in cloud server and the status of air pollution index can be updated, and using IoT technologies, the predicted data can be sent to the user's dashboard and an alert message can pop out when the air pollution threshold limit exceeds.