Air Quality Index Prediction

1st Md. Tahiadur Rahman
Id: 20200204003
Group: A1 (3)
Department of Computer Science and Engineering

2nd Adiba Amin
Id: 20200204012
Group: A1 (3)
Department of Computer Science and Engineering

3rd Md. Rafiu Alam Rafi
Id: 20200204051
Group: A1 (3)
Department of Computer Science and Engineering

***Abstract*—In this study, we perform air quality prediction to estimate the AQI based on various pollutant concentrations. Utilizing regression analysis techniques, we analyze data to predict AQI levels from pollutants such as PM2.5, PM10, NO2, NOx, NH3, CO, SO2, O3, Benzene, Toluene, and Xylene. This research aims to provide accurate air quality predictions and enhance the understanding of environmental pollution trends.**

I. Introduction

Air quality prediction is a vital tool in environmental monitoring that helps assess pollution levels. By analyzing pollutant data, we aim to predict the AQI, providing insights into air quality trends. This research focuses on pollutants such as PM2.5, PM10, NO2, NOx, NH3, CO, SO2, O3, Benzene, Toluene, and Xylene. Accurate AQI predictions can aid policymakers, environmental agencies, and the public in making informed decisions to improve air quality and public health.

II. Motivation

The escalating concerns about air pollution underscore the necessity for precise air quality monitoring and prediction. While regression analysis for AQI forecasting is extensively studied in major regions, there is a notable gap in research and tools for many localized environments. Accurately understanding air quality trends can provide valuable data to policymakers and environmental agencies, aiding in pollution mitigation and public health protection. Reliable AQI predictions can greatly influence environmental strategies and public awareness. Extending these benefits to underserved regions can significantly enhance environmental monitoring efforts and promote global health equity. With growing urban populations and increasing industrial activities, tools for precise AQI prediction in diverse settings become crucial. This research fills an important gap and contributes to global efforts to improve air quality monitoring and management.

III. Literature Review

There is existing research on air quality index (AQI) prediction for various regions that has primarily utilized regression analysis along with statistical and machine learning techniques. These studies focus on forecasting AQI levels by analyzing historical data and incorporating meteorological factors and pollutant concentrations through regression modeling.

This paper [1] introduces a hybrid deep learning model integrating Attention Convolutional Neural Networks (ACNN), Quantum Particle Swarm Optimization (QPSO)-LSTM, and XGBoost to predict Air Quality Index (AQI) in Seoul. The model combines linear and nonlinear data components for enhanced accuracy. It outperforms traditional methods, showing superior performance in metrics like MSE and MAE.

This paper [2] presents an air pollution forecasting model combining Long Short-Term Memory (LSTM) with a Genetic Algorithm (GA) to optimize critical hyperparameters such as window size and the number of LSTM units. The model is trained to predict next-day pollution levels for four pollutants: PM10, PM2.5, CO, and NOx. By leveraging GA, the model improves prediction accuracy and efficiency over traditional LSTM models, effectively handling noisy data and optimizing parameter selection. This approach proves more effective for long-term pollution pattern forecasting.

This paper [3] employs regression techniques to predict carbon monoxide concentrations using a dataset with hourly air pollutant data. It evaluates the relationship between meteorological factors and air pollutants to understand how these variables affect carbon monoxide levels. The study aims to raise public awareness about air quality, assist environmentalists, and guide government policies on air pollution and health standards.

This paper [4] estimates a multiple linear regression model for AQI in Velachery, Chennai, using data from a CPCB monitoring station. Key pollutants influencing AQI are CO, PM2.5, O3, and SO2. Meteorological factors showed minimal impact. Residual analysis confirms the model's validity and fitness.

This paper [5] proposes an AQI prediction model using CNN-ILSTM, which combines Convolutional Neural Networks (CNN) for feature extraction and an Improved Long Short-Term Memory (ILSTM) network. ILSTM removes LSTM's output gate, enhances the input and forget gates, and adds a Conversion Information Module (CIM) to prevent supersaturation. Compared with SVR, RFR, LSTM, GRU, and CNN-LSTM, CNN-ILSTM shows superior prediction accuracy, with lower MAE, MSE, and training time.

This paper [6] explores various regression models to forecast the Air Quality Index (AQI) in Delhi and Houston, focusing on pollutants like NO2, CO, O3, PM2.5, PM10, and SO2. Models such as support vector regression (SVR) and multiple linear regression (with gradient descent variations) were tested. SVR demonstrated superior performance in predicting AQI.

IV. Methodology

We employed both traditional machine learning and ensemble learning techniques for predicting air quality index (AQI) due to their effectiveness in handling complex datasets. Specifically, supervised learning was chosen for its capability to model and adapt to dynamic environmental data. Our study proposes three methods, Linear Regression, LSTM and HMM for predicting AQI from various pollutant concentrations, including PM2.5, PM10, NO2, NOx, NH3, CO, SO2, O3, Benzene, Toluene, and Xylene, to provide accurate and reliable air quality predictions. Figure1 represent the simple working procedure for this study-

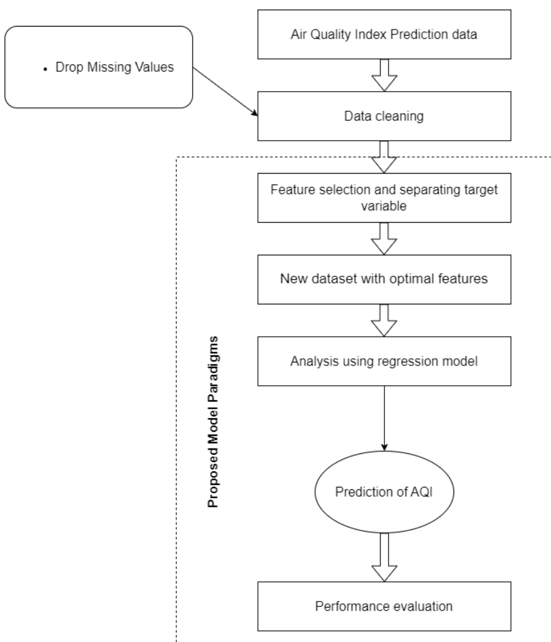

Fig. 1. Model of working mechanism

We utilized a dataset [7] comprising 29,532 samples for predicting Air Quality Index (AQI) based on pollutant levels. Our approach involved preprocessing the data by imputing missing values with column means. We then selected relevant features such as PM2.5, PM10, NO2, NOx, NH3, CO, SO2, O3, Benzene, Toluene, and Xylene. Splitting the data into training and testing sets facilitated the training of three models: Linear Regression, LSTM and HMM. These models were evaluated using Mean Squared Error and R-squared metrics to assess their predictive performance. Visualizations of actual versus predicted AQI values provided insights into model accuracy and effectiveness.

V. Result Analysis

Table 1 presents the performance of Linear Regression, LSTM, and HMM models based on Mean Squared Error (MSE) and R-squared metrics. The analysis identifies LSTM as excelling in predictive accuracy with the lowest MSE of 1585.19 and a strong R-squared value of 0.818, closely matching the performance of Linear Regression. In contrast, the HMM model demonstrated significantly lower predictive performance, with a high MSE of 16743.23 and a negative R-squared value, indicating that it failed to capture the underlying data patterns effectively. This comparison highlights the trade-off between model complexity and suitability for continuous data prediction, with LSTM demonstrating the ability to handle time-series air quality data more accurately, while HMM, typically suited for discrete states, performs poorly in this context.

TABLE I
PERFORMANCE METRICS FOR DIFFERENT MODELS

Model	**R-squared**	**MSE**
Linear Regression	0.82	2810.90
LSTM	0.82	1585.19
HMM	-0.005	16743.23

References

[1] Anh Tuan Nguyen, Duy Hoang Pham1, Bee Lan Oo , Yonghan Ahn and Benson T. H. Lim, 'Predicting air quality index using attentionhybrid deep learning and quantum-inspiredparticle swarm optimization'

[2] Ghufran Isam Drewil, Riyadh Jabbar Al-Bahadili, 'Air pollution prediction using LSTM deep learning and metaheuristics algorithms', Volume 24 , December 2022, 100546

[3] A.Aarthi, P.Gayathri, N.R.Gomathi , S.Kalaiselvi , Dr.V.Gomathi, 'Air Quality Prediction Through Regression Model', INTERNATIONAL JOURNAL OF SCIENTIFIC AND TECHNOLOGY RESEARCH VOLUME 9, ISSUE 03, MARCH 2020

[4] A. Loganathan , P. Sumithra, V. Deneshkumar, 'Estimation of Air Quality Index Using Multiple Linear Regression', Applied Ecology and Environmental Sciences, 2022, Vol. 10, No. 12, 717-722

[5] Jingyang Wang1,Xiaolei Li.Lukai Jin,Jiazheng Li, Qiuhong Sun1 and Haiyao Wang,'An air quality index predictionmodel based on CNN-ILSTM',(2022) 12:8373

[6] S. Ganesh, Sri Harsha Modali et al, 'Forecasting air quality index using regression models: A case study on Delhi and Houston', International Conference on Trends in Electronics and Informatics (ICEI) 2017

[7] Anju Sunilkumar et al, 'Air Quality Index prediction', 2022