IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

AIR POLLUTION PREDICTION USING MACHINE LEARNING

Mr. Pinku Padhy[*1], Mr. CH. Srinivasa Reddy[*2]

[1]MCA Student, Department of Master of Computer Applications,
Vignan's Institute of Information Technology(A), Beside VSEZ,Duvvada,Vadlapudi Post,
Gajuwaka, Visakhapatnam-530049.
[2]Assistant Professor, Department of Information Technology,
Vignan's Institute of Information Technology(A), Beside VSEZ,Duvvada,Vadlapudi Post,
Gajuwaka, Visakhapatnam-530049.
vignaniit.edu.in

Abstract:

The amount of pollution caused by humans on the planet has increased dramatically since the industrial revolution. Many of the pollutants in the environment are visible, such as those in the air, water, and soil. Some people, particularly those who reside in big industrial cities, will be aware of air pollution. Since air quality is becoming one of the main factors affecting human health. Air pollution has become a major concern worldwide due to its detrimental effects on human health and the environment. Accurate prediction of air quality is crucial for implementing effective mitigation strategies and safeguarding public health. This study focuses on employing machine learning techniques, specifically the Long Short-Term Memory (LSTM) algorithm, for air quality prediction. The LSTM algorithm, a type of recurrent neural network, is known for its ability to capture temporal dependencies in sequential data. The methodology involves collecting historical air quality data, including pollutant concentrations, meteorological variables, and other relevant factors. These data are preprocessed and used to train the LSTM model, which learns the complex relationships between the input variables and the air quality outcomes. The trained model is then used to make predictions for future air quality conditions. The performance of the LSTM model is evaluated using various evaluation metrics, such as mean absolute error (MAE) and root mean square error (RMSE), to assess its accuracy in predicting air quality.

Keywords: Long Short-Term Memory (LSTM) algorithm,Air Pollution,Quality Outcomes,Human Health,Environment.

1. INTRODUCTION

In today's rapidly advancing world, human progress in the realms of technology is elevating standards across various industries, including the automobile sector and factories. However, this progress is accompanied by a dark side as societal behavior tends toward selfishness, leading to environmental issues such as deforestation, reduced oxygen levels, and air pollution. Air pollution, a consequence of both human activities and natural processes, is marked by increased aerosols that adversely impact living organisms' life cycles. Weather factors, including wind direction, speed, pressure, temperature, and cumulative rain and snowfall, play a crucial role in air pollution dynamics.Several factors contribute to air pollution, including the busy movement of vehicles, acid rain, industrial activities located far from residential areas, particulate contamination, greenhouse effects, and excessive ultraviolet radiation. Particulate matter, indicated as PM2.5, consists of suspended liquid droplets and solid particles in the atmosphere, causing respiratory issues, lung cancer, and heart problems. The impacts of air pollution are extensive, affecting major organs such as the lungs and heart, leading to heart disease, lung cancer, bronchitis, and respiratory problems.

Short-term exposure to pollutants can significantly impair lung and cardiovascular system functions. To mitigate the impact of air pollution on public health, awareness about air quality is crucial, and technology should be harnessed for the greater good. Research on air pollution analysis and prediction is essential to make informed decisions and take effective actions against air pollution.In the realm of technology and science, air pollution forecasting emerges as a vital application, predicting pollutant concentrations and providing an air quality index for specific times and locations. Machine learning, a technology that enables computers to learn from experience, is instrumental in air pollution forecasting by predicting timely pollutant concentrations. Large-scale optimization algorithms facilitate effective model training on big data.

Two types of time series forecasting methods, univariate and multivariate, are employed. Univariate time series involves forecasting based on variables like date-time, while multivariate time series considers additional variables. Deep learning techniques, particularly Recurrent Neural Networks (RNN) and its enhanced version, Long Short-Term Memory (LSTM), are crucial for time series forecasting, overcoming limitations associated with memory constraints in traditional RNNs. An LSTM model is developed for air pollution forecasting using the Keras deep learning library. This approach integrates technology and science to address the critical issue of air pollution and its detrimental effects on human health.

2. LITERATURE SURVEY

1) Application of BP Neural Network Optimized by Genetic Simulated Annealing Algorithm to Prediction of Air Quality Index in Lanzhou

Authors: Zhou Kang; Zhiyi Qu

Conference: 2017 2nd IEEE International Conference on Computational Intelligence and Applications (ICCIA)

Predicting air quality in urban areas is crucial for preventing air pollution and improving the living environment. This paper focuses on Lanzhou's air quality, utilizing data from China's air quality online monitoring and analysis platform. Three air quality index (AQI) prediction models are established: one based on Back Propagation (BP) neural network, another using genetic algorithm optimization, and the third integrating genetic simulated annealing algorithm optimization with BP neural network. The comparative analysis reveals that the BP neural network based on genetic simulated annealing algorithm exhibits strong generalization and global search capabilities, resulting in higher accuracy rates.

2) Air Quality Prediction Method in Urban Residential Area

Authors: RuiJun Yang; HaiLong Zhou; DanFeng Ding

Conference: 2018 11th International Symposium on Computational Intelligence and Design (ISCID)

This study employs hedonic price theory to establish an internal mapping relationship between housing prices and air quality in urban residential areas. Using machine learning classification algorithms, including SVM, Naive Bayesian, and KNN, the research predicts air quality based on feature variables. The experiments conducted on the dataset of Tianhe residential district in Guangzhou city demonstrate high accuracy. The proposed method, relying on machine learning algorithms, proves practical and stable for predicting air quality in urban residential areas. The approach offers valuable insights for potential buyers as a reference method with strong practical applicability.

3. EXISTING SYSTEM

Air quality prediction plays a vital role in assessing and managing environmental conditions. This abstract presents an approach to predict air quality using DHT (Digital Humidity and Temperature) sensors and microcontrollers. The system utilizes the DHT sensors to measure temperature and humidity, which are key factors in evaluating air quality. A microcontroller board, such as Arduino or Raspberry Pi, is employed to collect sensor data, process it, and generate air quality predictions. The collected data is processed, and air quality metrics are calculated based on established indexes or formulas. These sensors are used to calculate air quality in particular areas only. So to calculate whole area like states we require more number of sensors. So to gather whole information from the sensors we require a lot of time to predict the air quality.

Problems in the Existing System:

1. Limited Spatial Coverage: The system relies on sensors to calculate air quality in specific areas, limiting its ability to provide comprehensive coverage for larger regions like states.

2. Dependency on Sensor Density: To assess air quality over broader areas, a higher number of sensors is needed, leading to increased implementation costs and logistical challenges.

3. Time-Consuming Data Collection: Gathering information from numerous sensors for predicting air quality across a larger region is a time-intensive process, delaying real-time predictions.

4. Lack of Real-Time Monitoring: The system may not offer real-time air quality predictions due to the time required for data collection and processing from multiple sensors.

5. Limited Scalability: The current setup may face scalability issues when attempting to expand its coverage to encompass larger geographic areas, making it less adaptable to diverse environmental settings.

4. PROPOSED SYSTEM

We propose a system for air pollution forecasting using an LSTM model. The system collects historical air pollution data and related meteorological information from reliable sources. The collected data is cleaned, missing values are handled, and outliers are removed. Important features that affect air pollution, such as pollutant levels, temperature, humidity, wind speed, and time information, are selected. An LSTM model, a type of machine learning model suitable for capturing temporal patterns, is developed and trained using the selected features. The trained model is then used to forecast future air pollution levels based on real-time meteorological data. The system provides user-friendly visualizations and reports to present the forecasted results. Regular updates and maintenance ensure the accuracy and reliability of the forecasting system.

5. EXPERIMENTAL RESULTS

From the below figures it can be seen that proposed model is more accurate in order to prove our proposed system.

Main Window:

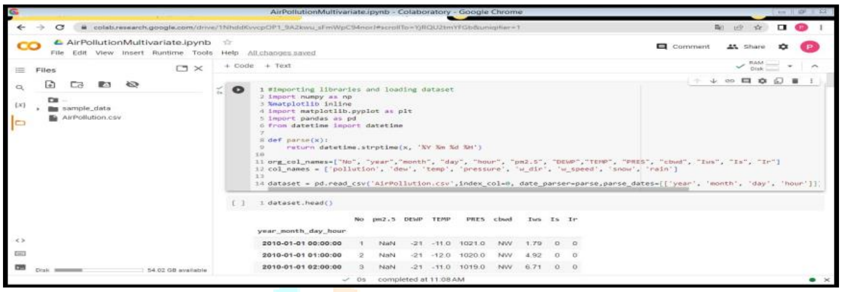

Figure1: Reading a dataset

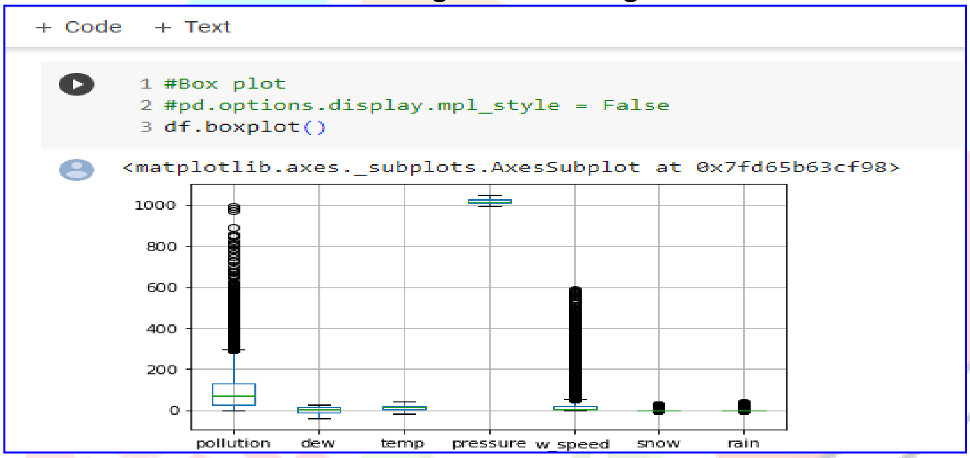

Figure2: Box plot shows distribution of a dataset

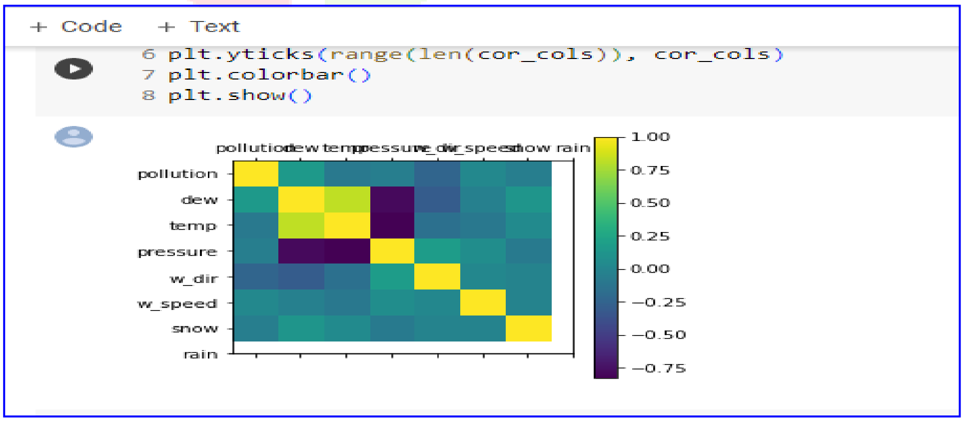

Figure 3: Shows the correlation matrix

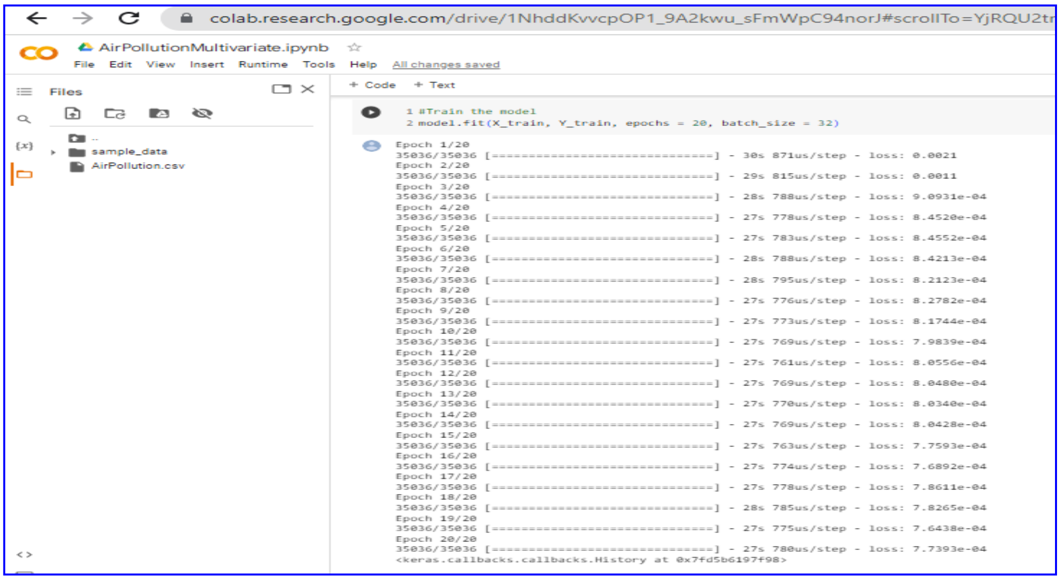

Figure 4: Shows the Training of Model

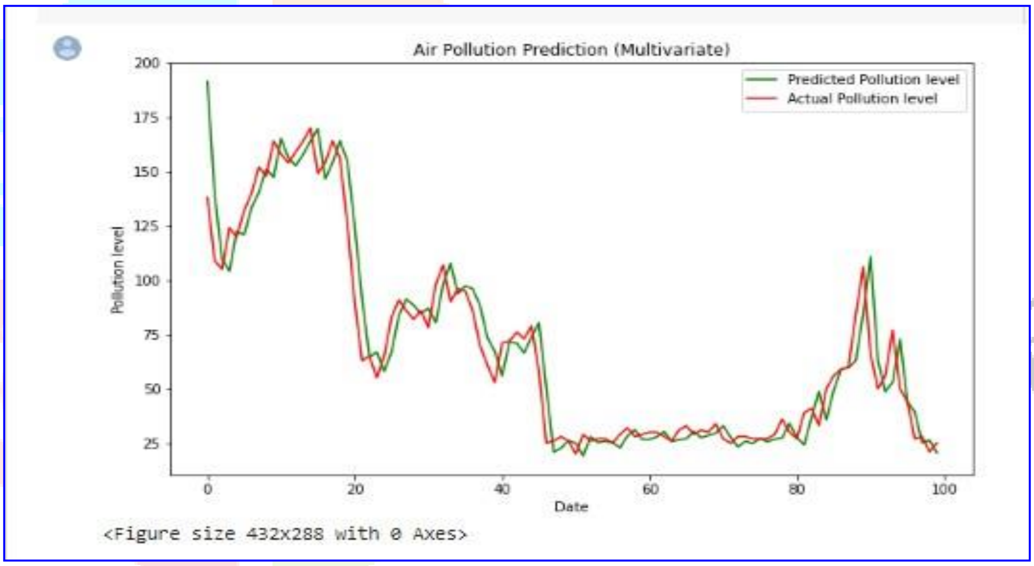

Figure 5: Represents the comparison of the actual and predicted pollution level.

6. CONCLUSION

In conclusion, addressing the issue of air quality is paramount due to its profound impact on human health, particularly causing problems such as heart and lung issues. Raising awareness about air quality among the general populace is imperative. Leveraging advanced technologies such as deep learning and machine learning, especially utilizing multivariate time series, proves instrumental in predicting and forecasting air pollution. The integration of autoencoders in the forecasting process enhances accuracy by reducing errors. The utilization of Long Short-Term Memory (LSTM) for multivariate time series forecasting in air pollution data further improves prediction capabilities. To broaden the system's reach and enhance its accessibility, it can be extended to IOS and Android platforms, allowing for real-time alerts about atmospheric air quality levels. Additionally, the system can be adapted for use on wearable devices, offering a practical and proactive solution for individuals to safeguard both nature and living beings from the adverse effects of air pollution.

References

1. Aditya CR, Chandana R Deshmukh, Nayana D K, Praveen Gandhi Vidyavastu, "Detection and Prediction of Air Pollution using Machine Learning Models," Department of Computer Science and Engineering, Vidya Vikas Institute of Engineering and Technology, Mysuru, Karnataka, India, 570028.

2. Zhongang Qi, Tianchun Wang, Guojie Song, Weisong Hu, Xi Li, Zhongfei (Mark) Zhang, "DeepAir Learning: Interpolation, Prediction, and Feature Analysis of Fine-grained Air Quality."

3. Kang, Gaganjot Kaur, et al. "Air quality prediction: Big data and machine learning approaches." Int. J. Environ. Sci. Dev 9.1 (2018): 8-16.

4. Yi-Ting Tsai, Yu-Ren Zeng, Yue-Shan Chang, "Air pollution forecasting using RNN with LSTM," Dept. of Computer Science and Information Engineering, National Taipei University, New Taipei, Taiwan, R.O.C.

5. Venkat Rao Pasupuleti, Uhasri, Pavan Kalyan, "Air Quality Prediction of Data Log By Machine Learning," 2020, IEEE.