e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:06/June-2024 Impact Factor- 7.868 www.irjmets.com

AIR QUALITY INDEX PREDICTION USING ML & DL IN PYTHON

Deepak. S[*1], Sarala Devi. V[*2], Ilam Chezhian. J[*3]

[*1]MCA Student, Department Of Computer Applications, Dr. M.G.R Educational And Research Institute, Chennai, India.

[*2,3]Asst. Professor, Department Of Computer Applications Dr. M.G.R Educational And Research Institute, Chennai, India.

ABSTRACT

In this suggested approach, we employ two distinct types of artificial intelligence models, namely machine learning (ML) and deep learning (DL). Within the ML framework, we evaluate three various classifiers, namely Decision Trees, Random Forests, and Support Vector Machines (SVMs), aiming to achieve an optimal prediction success rate. Concurrently, in the DL framework, we assess three additional classifiers, which include Auto-regressive integrated moving average (ARIMA), Long Short-Term Memory (LSTM), and Artificial Neural Networks (ANN), with the same objective. After determining the final prediction success rate, the process moves to Django for the development of the web application. Within this application, the conclusive prediction outcomes are presented to the user.

I. INTRODUCTION

Air pollution occurs as a result of fossil fuel combustion in power generation, transportation, and industrial processes. To mitigate and halt the worsening of global warming, it's essential to monitor the air quality index to identify the pollutants in the atmosphere.

Air pollutants are invisible to the naked eye, and the origins of rising pollution levels often go unnoticed. To grasp the origins of air pollution, it's necessary to first explore the fundamental causes.

PM2.5 particles pose a significant threat to the environment worldwide. Prolonged exposure to these minuscule particles, which can penetrate deep into the respiratory system, can lead to various health issues. PM2.5 particles, measuring 2.5 microns in size, are significantly smaller than a human hair. Consequently, the risk of exposure to these particles on the human body is greatly increased, resulting in a variety of diseases. This poses a serious threat to human health, including the development of cancer and various endemic and epidemic illnesses. The primary objective is to raise awareness among the public about the causes and effects of PM2.5 exposure in order to better regulate energy consumption.

II. LITERATURE SURVEY

According to Patil.et al., conducted a thorough examination of various methodologies and techniques for assessing air pollution concentration and predicting Air Quality Index (AQI). Their study emphasized the effectiveness of these analytical approaches and underscored the significance of calculating AQI as a crucial indicator for evaluating pollution levels and its profound impact on human health and the environment. In a similar vein, Oliveri et al. (2015) reviewed air quality models and explored the impact of air pollution concentration on human health.

According to Ameer.et al., examined the effectiveness of four regression techniques, namely Decision Tree, Gradient Boosting, Multilayer Perceptron, and Artificial Neural Network (ANN), in forecasting air quality levels. These techniques were assessed by monitoring PM2.5 levels in the air and calculating the AQI. The results of this study indicated that the Random Forest regression method outperformed the others, achieving an adjusted MAE of 16% for Beijing City. Additionally, this method reduced the computational time compared to Gradient Boosting and Multilayer Perceptron. Similarly, Maleki et al. employed the ANN approach to predict the concentration levels of various air pollutants such as NO2 and SO2. This study encompassed multiple monitoring areas including Naderi, Havashenasi, Behdasht, MohiteZist, and Iran. The authors of this study took into account the impact of parameters such as time, date, and meteorological data to develop a robust predictive model for air quality[2].

According to Zhang.et al., employed the long short-term memory (LSTM) to propose a deep learning approach for the detection of air pollution. This research conducted a series of experiments using Detrended Cross-

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:06/June-2024 Impact Factor- 7.868 www.irjmets.com

Correlation Analysis (DCCA) to investigate the connection between predicting levels of various air pollutants and meteorological factors such as temperature and humidity. The findings of this study revealed a negative correlation between AQI and meteorological data (temperature, humidity, and wind speed), while a significant positive correlation was observed between pressure and AQI. Moreover, Bougoudis developed a hybrid computational method to identify the correlation between air pollutants and weather conditions in order to determine the actual cause of pollution. The study utilized ANN and Random Forest as ensemble learning methods, claiming improved accuracy. However, the feedforward neural network encountered difficulties in predicting continuous values due to insufficient data [3].

According to Gore.et al., introduced a classification methodology to investigate the impact of air pollutant levels on human health using classification machine learning algorithms. They utilized Naive Bayes and Decision Tree algorithms in their approach and achieved a remarkable accuracy with the Decision Tree model. Additionally, Simu et al. conducted a comparative study to assess the performance of various machine learning algorithms, including Random Forest and Multi-linear Regression, in analyzing air pollutants and predicting air pollution levels. The findings of the study indicated that the Multilayer Perceptron algorithm outperformed the others [4].

According to Peng.et al., employed Multilayer Perceptron to improve the accuracy of air quality predictions. However, they highlighted constraints related to data expansion and the significant computational expenses due to the periodic model updates. Mahalingam and team suggested the utilization of ANN and SVM algorithms for forecasting the AQI in the intelligent city of Deldi, achieving remarkable accuracies, particularly with the Medium Gaussian SVM function. In order to forecast the AQI and air pollution levels, Sharma et al. considered various algorithms such as Linear regression, ANNs, Lasso regression, and XGBoost regression. The study concentrated on monitoring the concentrations of multiple pollutants, including NO2, SO2, PM2.5, PM10, CO, and O3. The research outcomes revealed that the Random Forest algorithm surpassed the other algorithms, showcasing its superior performance in predicting the AQI and air pollution levels [5].

EXISTING SYSTEM:

The current air quality control system utilizes a single classifier, resulting in lower accuracy compared to the proposed method. The dataset employed in the current model is also smaller than that of the proposed model, leading to decreased performance.

DISADVANTAGES:

- The existing air quality control system relies on a lone classifier, leading to reduced accuracy in contrast to the suggested approach.
- Additionally, the dataset utilized in the current model is less extensive than that of the proposed model, resulting in diminished performance.

III. PROPOSED SYSTEM

The suggested approach involves three distinct classifiers in both machine learning and deep learning, including decision tree learning, random forests, support vector machines (SVMs), auto-regressive integrated moving average (ARIMA), long short-term memory (LSTM), and artificial neural networks (ANN).

The most effective classifier with superior prediction capabilities in both machine learning and deep learning is integrated with Django for the user interface.

Django is utilized for creating websites and applications, through which the application delivers the final predicted outcomes.

The proposed technique aims to offer an alternative solution for quality analysis that reduces the necessary time and expenses.

ADVANTAGES:

- Using two models ensures a high level of precision and efficiency.
- By closely monitoring the air quality in industrial settings, we can minimize the impact on workers and effectively manage the release of gases.
- The implementation of the suggested approach can significantly reduce the emission of PM2.5, which is known to contribute to respiratory diseases and even cancer.

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:06/June-2024 Impact Factor- 7.868 www.irjmets.com

ARCHITECTURE DIAGRAM:

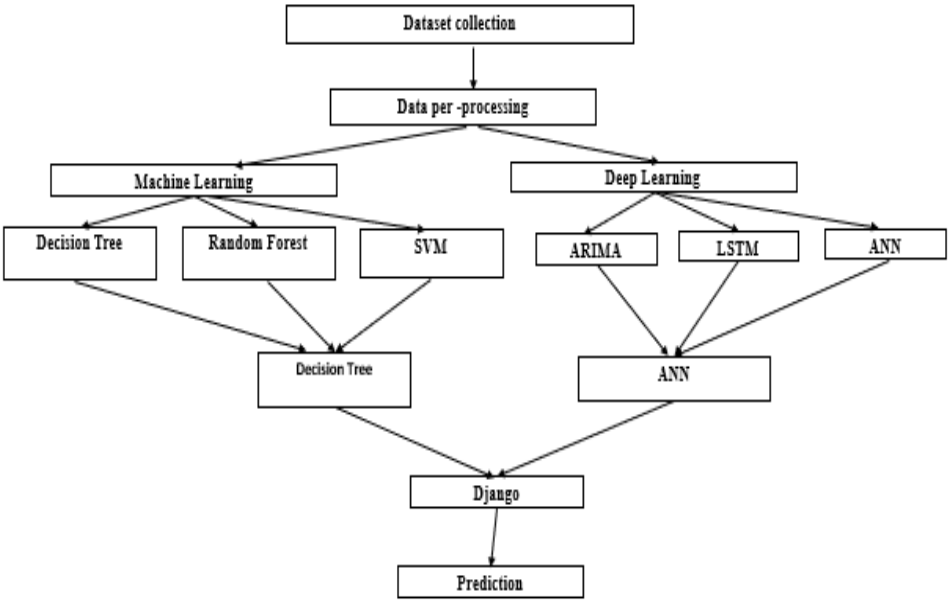

Fig 3.1: SYSTEM ARCHITECTURE

SYSTEM MODULE:

1. **Dataset collection**
2. **Data preprocessing**
3. **Algorithm**

- **Decision tree learning**
- **Random forests**
- **Support vector machines (SVMS)**
- **Auto-regressive integrated moving average (ARIMA)**
- **Long short-term memory (LSTM)**
- **Artificial Neural Networks (ANN)**

4. **Prediction**

Django

MODULE DESCRIPTION:

1. Dataset collection:

Data collection involves gathering data from a predetermined and validated source, using measurable methods, and analyzing various types of information. The primary objective of data collection is to obtain informative and dependable data. The PM2.5 dataset comprises 1092 sets. The gathered information is then subjected to analysis, followed by processing to identify any missing components that require further pre-processing.

2. Data preprocessing:

Preprocessing is an essential technique in data mining that aims to convert raw data into a format that is both useful and efficient. The collected data undergoes a thorough process of data cleaning and data processing. Data cleaning involves removing noise from the raw data and filling in missing parts to obtain useful data. Data transformation is another crucial step that helps in converting the data into a suitable format for the mining process. Additionally, data reduction techniques are employed to mitigate the processing time required for data mining, especially when dealing with large datasets.

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:06/June-2024 Impact Factor- 7.868 www.irjmets.com

3. Algorithm:

- **Decision tree learning:**

Decision tree learning, also known as decision tree induction, is a technique used in statistics, data mining, and machine learning for predictive modeling. It involves the use of a decision tree as a predictive model to derive conclusions about an item's target value based on observations represented in the branches. The decision tree consists of two types of nodes: decision nodes and leaf nodes. Decision nodes are responsible for making decisions and have multiple branches, while leaf nodes represent the output of those decisions and do not have any further branches. The decisions or tests performed in a decision tree are based on the features of the given dataset.

- **Support vector machines (SVMS):**

Support Vector Machines (SVMs) are specific linear classifiers that rely on the principle of maximizing the margin. They aim to minimize structural risk, thereby enhancing the complexity of the classifier and achieving superior generalization performance. The objective of the SVM algorithm is to establish an optimal line or decision boundary, known as a hyperplane, that effectively separates classes in an n-dimensional space. This enables easy categorization of new data points in the future.

- **Random forests:**

Random forests, also known as random decision forests, are a type of ensemble learning technique used for classification, regression, and other tasks. This method involves creating numerous decision trees during the training phase. For classification tasks, the random forest outputs the class that is chosen by the majority of the trees. For regression tasks, the random forest returns the mean or average prediction made by the individual trees. When compared to other machine learning algorithms, decision tree learning is considered one of the top classifiers.

- **ARIMA:**

ARIMA is a technique employed for time series forecasting by combining auto regression and moving average. It predicts future values by analyzing past values, taking into account lags and lagged errors in forecasting. This method allows for the modeling of non-seasonal time series patterns, excluding random white noise.

- **LSTM:**

The artificial recurrent neural network architecture known as Long short-term memory (LSTM) was introduced in 1997 by Sepp Hochreiter and Jurgen schmidhuber. LSTM utilizes feedback connections, enabling it to process both individual data points and sequences of data. This makes it particularly effective for tasks such as classifying, processing, and making predictions based on time series data, where there may be unpredictable delays between significant events.

- **ANN:**

Artificial neural networks (ANN) are structured as multi-layer fully-connected neural nets, as depicted in the diagram provided. These networks comprise an input layer, several hidden layers, and an output layer. Each node in a given layer is interconnected with every other node in the subsequent layer. This architecture enables ANN to effectively represent intricate patterns and address prediction problems. When it comes to deep learning, Artificial neural networks (ANN) stand out as the most proficient classifier.

4. Prediction:

Django is a sophisticated web framework for Python that promotes efficient development and elegant design, ensuring the creation of secure and easily maintainable websites. It is capable of producing accurate predictions by utilizing historical data and applying algorithms to forecast the probability of specific outcomes. The web application developed using Django showcases the ultimate result of air quality prediction.

IV. CONCLUSION

Django, a sophisticated Python web framework, facilitates efficient development and elegant design, guaranteeing the creation of secure and easily maintainable websites. By leveraging historical data and employing algorithms, it excels at generating precise predictions and forecasting the likelihood of specific

outcomes. The web application built with Django beautifully demonstrates the end result of air quality prediction.

V. REFERENCES

[1] S. Ameer, M. A. Shah, A. Khan, H. Song, C. Maple, S. U. Islam, and M. N. Asghar, "Comparative analysis of machine learning techniques for predicting air quality in smart cities," IEEE Access, vol. 7, pp. 128325–128338, 2019.

[2] M. Batty, K. W. Axhausen, F. Giannotti, A. Pozdnoukhov, A. Bazzani, M. Wachowicz, G. Ouzounis, and Y. Portugali, "Smart cities of the future," Eur. Phys. J. Special Topics, vol. 214, no. 1, pp. 481–518, Nov. 2012.

[3] I. Bougoudis, K. Demertzis, and L. Iliadis, "HISYCOL a hybrid computational intelligence system for combined machine learning: The case of air pollution modeling in Athens," Neural Comput. Appl., vol. 27, no. 5, pp. 1191–1206, Jul. 2016.

[4] D. Ganeshkumar, "Air and sound pollution monitoring system using cloud computing," Int. J. Eng. Res., vol. V9, no. 6, Jun. 2020.

[5] R. W. Gore and D. S. Deshpande, "An approach for classification of health risks based on air quality levels," in Proc. 1st Int. Conf. Intell. Syst. Inf. Manage. (ICISIM), Oct. 2017, pp. 58–61.

[6] B.-J. He, L. Ding, and D. Prasad, "Enhancing urban ventilation performance through the development of precinct ventilation zones: A case study based on the greater sydney, Australia," Sustain. Cities Soc., vol. 47, May 2019, Art. no. 101472.

[7] G. R. Kingsy, R. Manimegalai, D. M. S. Geetha, S. Rajathi, K. Usha, and B. N. Raabiathul, "Air pollution analysis using enhanced K-means clustering algorithm for real time sensor data," in Proc. IEEE Region 10 Conf. (TENCON), Nov. 2016, pp. 1945–1949.

[8] C. G. Kirwan and F. Zhiyong, Smart Cities and Artificial Intelligence: Convergent Systems for Planning, Design, and Operations. Amsterdam, The Netherlands: Elsevier, 2020.

[9] Z. Lv, D. Chen, R. Lou, and Q. Wang, "Intelligent edge computing based on machine learning for smart city," Future Gener. Comput. Syst., vol. 115, pp. 90–99, Feb. 2021.