DETECTION OF AIR POLLUTANTS

Project Reference No.: 45S_BE_0706

College   : *Rajarajeswari College of Engineering, Bengaluru*
Branch   : *Department of Computer Science Engineering*
Guide(s)   : *Ms. Chaitra S*
Student(S)   : *Ms. Likitha K.S*
                       Ms. Aishwarya S
                       Ms. Chethana K.R
                       Ms. Ganavi D

Introduction:

Pollution is a common issue in India. Artificial intelligence and machine learning have been two of the most significant technological advances in recent years. Air poisons might be "standards contaminations" or "perilous air toxins". Measures poisons are utilized to decide if a locale is fulfilling air quality guidelines, for example is in "achievement" status. Perilous air poisons, known as "air toxics", are substance compounds associated with causing disease.Some are exceptionally neighborhood in character, with the ecological effect of the contamination confined to the prompt area of, for instance, a street or a manufacturing plant. Different issues are provincial in character because of the long-range transport of poisons like corrosive testimony and tropospheric ozone. As SML techniques, we Make use of decision trees, linear regression , Artificial neural network.

Many of the algorithms we employ were developed by numerous researchers, but none of them examines the findings of all three investigations as if they were one a single research with similar data and circumstances. Various models have been exercised in the literature to predict AQI, like statistical, deterministic, physical, and Machine Learning (ML) models. The traditional techniques based on probability, and statistics are very complex and less efficient. The ML-based AQI prediction models have been proved to be more reliable and consistent. Advanced technologies and sensors made data collection easy and precise. The accurate and reliable predictions through such huge environmental data require rigorous analysis which only ML algorithms can deal with efficiently. The SMLT captures a lot of information from the dataset, such as Missing price treatments, techniques, and knowledge analysis. The record revolves around six major contaminates.

Background work:

1. The present system can only identify air quality with a low degree of precision, and it cannot proper decisions or problems
2. In the current system, a smaller dataset is still utilized, which is incompatible with all algorithm.

Disadvantages of Existing System:

1. Accuracy is reduced.
2. The cost of implementation is expensive, and the process is more complicated.
3. Accurate measures of contaminating gaseous cannot be detected in ppm.

Objectives:
1. The main objective of this research is to develop air pollution prediction system.
2. The system can discover and extract hidden knowledge associated with pollution from a historical air pollution data set air pollution prediction system aims to exploit ML techniques on data set to assist in the prediction of the air pollution
3. The main purpose is to examine the air quality index by building ML models and predicting the best results from different ML approaches based on accuracy
4. We use supervised machine learning techniques to calculate the Air Quality Index.
5. Remaining web providers will get analyzed data via a front-end API and display the visualized data in the browser to present projections and forecasts.
6. The comparison of AQIs between cities and polluted cities or areas is shown.
7. To develop a model that shows air quality index of next 10 days.

Methodology:

Dataset collection and Data pre-processing:
It is the method involved with changing crude information into a justifiable organization. It is additionally a significant stage in information mining as we can't work with crude information. The nature of the information ought to be checked prior to applying AI or information mining calculations. The data pre-processing mainly concludes:

1. Data Gathering- Gathering the specific dataset required for the project. Gathering information is the main move toward taking care of any directed AI issue. Your text classifier must be just about as great as the dataset it is worked from.
2. Data integration- Information Integration is an information pre-processing procedure that includes consolidating information from different heterogeneous information sources into a cognizant information store and give a brought together perspective on the information. These sources might incorporate different information blocks, data sets, or level records.
3. •Data cleaning- Information cleaning is the most common way of fixing or eliminating mistaken, ruined, inaccurately designed, copy, or fragmented information inside a dataset. While consolidating different information sources, there are numerous potential open doors for information to be copied or mislabelled.

Data train and test model:
1. Train/Test is a technique to gauge the precision of your model.
2. Before splitting train & test — we need to define dependent and independent variables.

3. The dependent variable is also known as X and the independent variable is known as y
4. X = data.drop('Class', axis = 1).valuesy = data['Class'].values Now, let split our train and test data.X_train, X_test, y_train, y_test = train_test_split(X, y, test_size = 0.25, random_state = 1).
5. We now have two different data set — Train data we will be used for training our model and the data which is unseen will be used for testing.

Data prediction:
1. Once the dataset has been pre-processed, we need to train them using machine learning algorithms to the system.
2. The accuracy will be predicted for the model.

Result And Conclusions:

In this study, detection of air pollutants using supervised machine learning techniques like decision tree, linear regression and artificial neural network. The system is user friendly and it gives highest accuracy. In this system we get air quality index of next ten days. In this system we get the bar graph for prediction of air quality in a state overall year, output for states which has highest air quality each year, output for air quality level with highest prediction in a state each year and also, we get the prediction of future which state will have highest air quality pollutants and we get the prediction of air quality pollutants which will have highest prediction. We analyze the toxins and air quality record contingent upon the year and month. We have looked at the toxins and air quality file in different urban communities. A lot of endeavors are being made by both state and focal legislatures to think about the air quality file. This system will be useful to control the pollution in cities by its future prediction.

References:

1. M. Huang, T. Zhang, J. Wang and L. Zhu, "A new air quality forecasting model using data mining and artificial neural network", 6th IEEE International Conference on Software A.
2. Xia Xi, Zhao Wei, Rui Xiaoguang, Wang Yijie, Bai Xinxin, Yin Wenjun, et al., "A Comprehensive Evaluation of Air Pollution Prediction Improvement by a Machine Learning Method", IEEE International Conference, 2015
3. K. Hu, A. Rahman, H. Bhrugubanda and V. Sivaraman, "HazeEst: Machine learning based metropolitan air pollution estimation from fixed and mobile sensors", IEEE Sens. J, vol. 17, no. 11, pp. 3517-3525, 2017.
4. Li, S., Song, S., and Fei, X. (2011). "Spatial characteristics of air pollution in the main city area of chengdu, china." 2011 19th International Conference on Geoinformatics, 1–4.
5. Qin, D., Yu, J., Zou, G., Yong, R., Zhao, Q., and Zhang, B. (2019). "A novel combined prediction scheme based on cnn and lstm for urban pm2.5 concentration." IEEE Access, 7, 20050– 20059.

6. J. Gao, C.-L. Xie, and C.-Q. Tao, "Big data validation and quality assurance - issues, challenges, and needs," IEEE Symposium on Service-Oriented System and Engineering, IEEE Computer Society, Oxford, UK, April, 2016.

Scope For the Future Work:

The machine learning in its components has become the key focus areas of all companies from the early stages to major platform vendors. Machine learning is a field in which an artificial intelligence device collects sensor data and learns to act. One of those factors, the propose work has chosen to| learn and forecast the air quality index with the opportunity to adapt to the machine learning algorithms. Air Quality Index is a device used for determining the air quality and the causes of air pollution. The proposed work will be beneficial for the governments to take effective measures and to control the pollution and to reduce the air quality index. As we get the prediction about future air quality index and the pollutants it will be helpful for the government to take measure to control the air pollution and to reduce the air pollutants which is contaminating the air. In future it will be helpful to easily detect the pollutants and the air quality index.