SREENIVASA INSTITUTE OF TECHNOLOGY AND MANAGEMENT STUDIES
(Autonomous)
Murukambattu, Chittoor – 517127.

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

Zeroth Review

PREDICTING AIR QUALITY INDEX USING MACHINE LEARNING

B KRISHNA MOHAN YADAV -20751A0529
A STEPHEN RAJ - 20751A0508
B SONU - 20751A0517
G TEJESWARA RAO - 20751A0516

Supervisor Comment **Signature of the Supervisor**

Panel Member Comments

Panel Member Signature

1.
2.

Abstract

The degradation of air quality has become a pressing issue globally, impacting human health, ecosystems, and climate. In response, this project aims to develop a robust machine learning model to predict Air Quality Index (AQI) values based on various environmental parameters. Utilizing historical air quality data, meteorological data, and geographical information, a comprehensive dataset is constructed. Feature engineering techniques are applied to extract relevant features, followed by rigorous data preprocessing to handle missing values and outliers.

Using machine learning techniques like CatBoost, Neural Network and Random Forest to predict air quality index (AQI) is a great approach. CatBoost is particularly effective for handling categorical features, while Random Forest is robust and handles non-linear relationships well.

To predict AQI, you'd typically collect features like temperature, humidity, wind speed, and pollutants such as PM2.5, PM10, CO, etc. Then, you'd train your models using historical data where AQI values are known.

For evaluation, you'd use mean squared error (MSE) to quantify the difference between predicted AQI values and the actual ones. A lower MSE indicates a better fit of the model to the data.

Random Forest works by creating multiple decision trees during training and outputs the average prediction of the individual trees for regression tasks like AQI prediction. It's robust against overfitting and performs well with large datasets.

In summary, combining CatBoost and Random Forest for predicting AQI can provide accurate predictions by leveraging their strengths in handling different types of data and capturing complex relationships between features and AQI values.

KEYWORDS: Air Quality Index(AQI),Machine Learning, CatBoost, Neural Network, Mean Squred Error(MSE), Temperature, Humidity, Pollutants, Historical data.

Problem statement/objectives

Develop a machine learning model to predict the Air Quality Index (AQI) using CatBoost, Mean Squared Error (MSE) evaluation, and Neural Networks (NN). The goal is to accurately forecast AQI levels based on various environmental factors such as temperature, humidity, wind speed, and pollutant concentrations (e.g., PM2.5, PM10, CO).

Key Objectives:

1. *Data Collection:* Gather historical data on AQI levels and corresponding environmental factors from reliable sources.

2. *Feature Selection:* Identify relevant features that influence AQI levels, including temperature, humidity, wind speed, and pollutant concentrations.

3. *Model Training:* Utilize CatBoost, a gradient boosting algorithm known for its effectiveness with categorical features, and Neural Networks, which excel in capturing complex relationships, to train predictive models.

4. *Evaluation:* Assess model performance using Mean Squared Error (MSE) as the evaluation metric. Lower MSE values indicate better model accuracy in predicting AQI levels.

5. *Optimization:* Fine-tune model parameters and architecture to optimize performance and minimize prediction errors.

6. *Deployment:* Deploy the trained model in real-time or batch processing systems to provide timely and accurate AQI predictions for stakeholders and decision-makers.

Expected Outcome:

A robust machine learning model capable of accurately predicting AQI levels using CatBoost and Neural Networks, validated through MSE evaluation. This model will assist in monitoring and managing air quality, contributing to public health and environmental sustainability efforts.

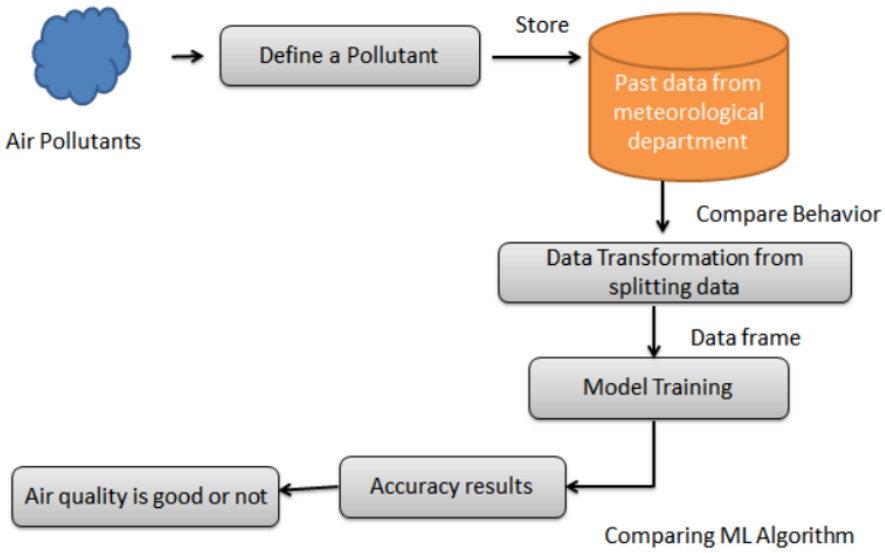
Existing system

Random Forest:

- Using random forest in predicting air quality index (AQI) through machine learning is a common approach. Random forest is advantageous for its ability to handle large datasets, high dimensionality, and nonlinear relationships between features and the target variable. By training on historical air quality data along with various environmental and meteorological features, a random forest model can predict AQI levels for future time periods or locations. It's important to preprocess the data, select relevant features, tune model parameters, and evaluate model performance to ensure accurate predictions.

Support Vector Machines:

- Maximizing Margin:SVM aims to find the hyperplane that maximally separates classes, achieving a wide margin of separation to improve generalization and classification accuracy.
- With SVMs, you'd preprocess the data, select relevant features, and then train the model on historical air quality data along with various environmental and meteorological features. SVMs can handle nonlinear relationships between features and the target variable, making them suitable for capturing complex patterns in AQI data.



Proposed system

Various modern techniques like artificial neural network

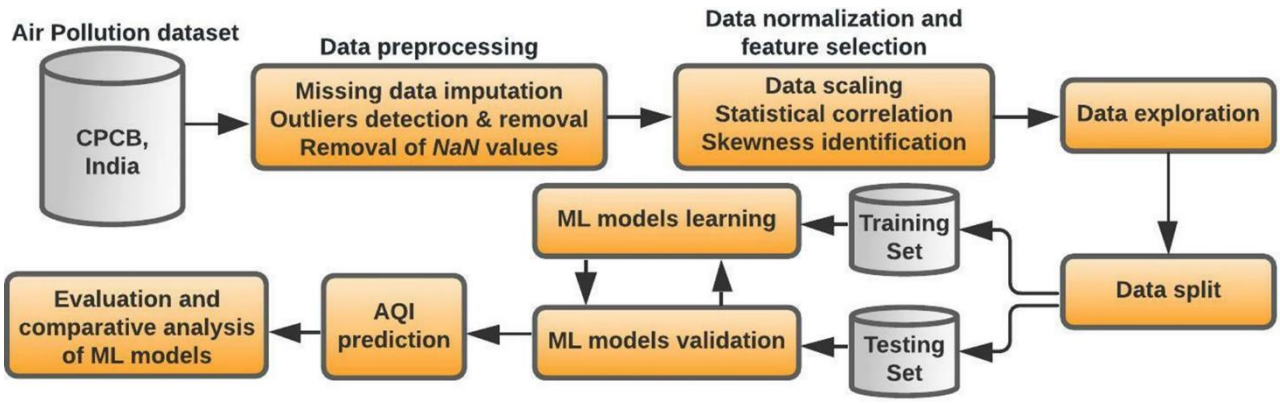
Different machine learning algorithms are compared, including Auto Encoder, Local Outlier Factor, Kmeans Clustering.

• This project uses various algorithm, and neural network which comprises of techniques for finding optimal solution for the problem and implicitly generating the result of the fraudulent transaction.

The main aim is to detect the fraudulent transaction and to develop a method of generating test data.

This algorithm is a heuristic approach used to solve high complexity computational problems.

The implementation of an efficient fraud detection system is imperative for all UPI issuing companies and their clients to minimize their losses.



References

- Machine Learning Approaches for Prediction of Air Quality Parameters: A Comprehensive Review" by K. S. Deepika and S. M. Akhil:
  https://ieeexplore.ieee.org/document/9356583
- "Air Quality Prediction Based on Machine Learning and Deep Learning: A Review" by R. Patidar and R. K. Agrawal:
  Link: https://link.springer.com/article/10.1007/s10586-021-03364-4
- "A Review of Machine Learning Techniques for Air Quality Prediction" by S. R. Algamal, M. A. A. El-Sherbeny, and K. E. El-Kilany:
  Link:https://www.researchgate.net/publication/348275529_A_Review_of_Machine_Learning_Techniques_for_Air_Quality_Prediction
- "Deep Learning for Air Quality Prediction: A Review" by X. Zhang et al.:
  Link: https://www.mdpi.com/2073-4433/10/8/521
- "Air Quality Prediction Based on Machine Learning Algorithms" by A. Chatterjee and S. Dutta:
  Link:https://www.researchgate.net/publication/352442996_Air_Quality_Prediction_Based_on_Machine_Learning_Algorithms

Learning Techniques for Fraud Detection in UPI Transactions." In Soft Computing Applications, Vol. 24, Issue 2, pp. 399-412. Available at: https://www.researchgate.net/publication/349935115_A_Comparative_Study_on