

Road Traffic Forecasting Using Air Pollution and Atmospheric Data

Sivamurugan S¹, Nithya P², Vedapriya N³, Vijayashree K⁴

¹Assistant Professor, Department of Artificial Intelligence and Data Science, Sri Sairam Engineering College, West Tambaram, Chennai, India

^{2,3,4}Student, Department of Artificial Intelligence and Data Science, Sri Sairam Engineering College, West Tambaram, Chennai, India

Abstract

Traffic congestion in urban areas is a persistent challenge with significant implications for public health, economic productivity and environmental sustainability. Effective management of traffic flow requires accurate forecasting to anticipate congestion patterns and implement timely interventions. Traditional traffic prediction models often overlook the influence of environmental factors such as air pollution and atmospheric condition on traffic dynamics. In this paper, we propose a novel machine learning based approach that integrates environmental data into road traffic for forecasting models. By collecting historical data on road traffic, air pollution levels and atmospheric parameters, we develop predictive models capable of capturing the intricate interplay between traffic patterns and environmental conditions. Our experimental findings demonstrate the efficiency of our approach in enhancing the accuracy of traffic forecast compared to conventional methods. Through this research, we contribute to advancing the understanding of how environmental factors impact flow dynamics and provide practical insights for improving urban transportation management strategies.

Keywords: Machine Learning, Air Pollutant, Pollution, Traffic, Traffic pollution, Atmospheric data, Accuracy, Human Health, Sustainable Environment.

1. Introduction

In this modern era of advancements and urbanization, one of the most crucial problems in society is air pollution. Air pollution is caused by any physical, chemical or biological agents that change the characteristics of the natural form of atmosphere. It is a pressing global issue that poses significant risk to human health and ecosystem and the overall well-being of the planet. Household combustion devices, automobile smoke emission, industries and forest fires are the most common sources of air pollution that release Carbon monoxide, Carbon dioxide, Nitrogen dioxide, Sulphur oxide, Chlorofluorocarbons, Particulate Matter, and other air pollutants that cause air pollution into the environment. WHO data show that almost 99% people are breathing air that crosses the WHO guideline limits and is exposed to large amounts of pollutants. The low and middle income countries are found to be affected the most. During several billion years of chemical and biological evolution, the composition of earth's atmosphere has changed. Ambient air quality standards are permissible exposure of all living and nonliving things for 24 hours per day, 7 days per week.

Air pollution poses significant damage to both humans and the environment, so monitoring pollutants level is crucial. We can do this with the help of Machine Learning models. Machine learning is a subset of Artificial Intelligence that helps the computer to learn how to build models based on training data. Machine Learning can inspect a wide range of data and recognize particular trends and patterns. Machine learning is the ability given to a computer program to do a task without any external programming and this task is achieved by using some statistical and advanced mathematical algorithms. Previous research in traffic forecasting has explored various approaches, including time-series models, machine learning algorithms and hybrid methods. Some studies have incorporated environmental data such as weather conditions into traffic prediction models but relatively few have considered air pollution as predictor. Our work builds upon this research by investigating the impact of air pollution on road traffic and developing models that leverage this information for improved forecasting. This project, presents a machine – learning based approach to road traffic forecasting that leverages air pollution and atmospheric data for improved accuracy.

2. Literature Survey

In [1] authors proposed that the Machine Learning models are showing very good accuracy and efficiency in terms of training the model. Only Machine Learning models can handle and train the rigorous dataset collected with advanced techniques and sensors. The Machine Learning algorithm KNN is showing accuracy of 99.1071% in their air pollution prediction.

In [2] the authors concluded their work by saying that concentration of air pollutants in ambient air is governed by the various parameters such as wind speed, wind direction, relative humidity, and temperature. Air Quality Index(AQI), is used to measure the quality of air. The proposed work is a supervised learning approach using various algorithms such as LR, SVM, DT and RF. The result has shown that AQI predictions obtained through RF are promising and which are analysed with results.

In [3] the authors intend to develop models based on past data and use them to make future decisions. The future is evaluated or forecasted in accordance with the past. The Time series supplements an additional time order dependence among observations. This dependency provides both a knowledge source and a knowledge barrier. According to the authors of this review, the majority of research has concentrated on evaluating or forecasting the AQI and pollutant concentration levels, which will provide a precise idea of AQI. Several researchers opt for Artificial Neural Network (ANN), ARIMA Model, Linear Regression, and Logistic Regression for forecasting of AQI and air pollutants concentration. When protruding the AQI or the subsequent concentration level of several pollutants, the future needs may take attributes into the picture, including meteorological framework and air contaminants. As the data switches at particular periods of time, it is also possible to use real-time data analysis through the cloud to get better outcomes for increased performance.

In [4] the authors have examined the application of machine learning algorithms in the forecasting and prediction of air pollution. The review has also analysed the proliferation of pollutants and their effects and level of concentration in the places away from the source. The dispersion module has made use of the Gaussian air dispersion model that was carried out using python in spyder IDE. For air pollution forecasting and prediction, different machine learning algorithms were applied on the data and differentiated with, which comprise Random Forest, Multi-layer Perceptron, K-Nearest Neighbour, Support Vector Regression, and Multi-linear Regression. The outcome of the result confirms that the Multi-layer Perceptron algorithm has shown to give least mean squared error compared to the other

machine learning algorithms. Their work suggests that Future work can be done in comparing several air dispersion models to predict and analyse the spread of air pollution.

In [5] the authors present a spatial temporal predictive model to overcome the limitation where it conducted several experiments using different models. The data set provides an information on the city NO₂, O₃, SO₂ levels for through 10 years. Relation between the pollutants to their geographical locations translates the problem into a classification issue. In order to predict the continuous values, it have used SVM, SVR, LSTM, ARIMA model, k-means clustering algorithms and determined a low cost-complexity combination of models.

In [6] the authors proposed a system that will focus on the monitoring of air pollutants with the combination of IOT with a machine learning algorithm called Recurrent Neural Network, specifically Long short term memory. In this paper they monitor the air quality with the help of IOT devices. The data used in this work is collected from the DHT11 sensor for generating real-time digital temperature and humidity. The system utilises air sensors to detect and transmit this data to microcontroller. Then the microcontroller stores the data into the web server. For predicting the LSTM is implemented.

In [7] the authors done a research to determine the air quality index and their conclusion is various researchers collect the data set from the kaggle repository and air quality monitoring sites and divided into two training and testing where used various machine learning algorithm are compared irrespective of pollutants. The algorithms are Linear Regression, Decision tree, Random forest, Artificial Neural Network and Support Vector Machine. In this paper they compared the analysis of result obtained by various researchers with various algorithms had taken meteorological data like temperature, wind speed, humidity in predicting accurately the upcoming pollutant level.

In [8] the authors developed powerful machine training techniques to prevent air pollution. We discussed the use of pollution estimation machine-learning algorithms and the Indian air quality index in turn (AQI). We noted that that the decision tree Algorithm gave the best result among all the algorithms, with an overall accuracy of 99.8%. The number of model parameters and optimized output was reduced with structure regularization, which in turn alleviated model complexity.

In [9] the authors predict the two pollutants concentration NO_x and CO in industrial sites by the use of a nonlinear Auto Regressive model (NARX) based Artificial Neural Network (ANN). Database used to train the neural network corresponds to historical time series of meteorological variables (wind speed, wind direction, temperature and relative humidity) and concentrations of pollutants in the petrochemical plant of Skikda site. The estimation performance is determined using the Roots Mean Square Error (RMSE) and Mean Absolute Error (MAE). Results will show the importance of the meteorological variable set on the prediction of pollutants concentrations and the neural network efficiency.

Methodology

The first step in data collection involves identifying reliable sources for road traffic, air pollution and atmospheric data, This may include government agencies responsible for transportation and environmental monitoring, weather stations and research institutions. The collected data should cover a significant time period and geographic area to capture variations in traffic patterns and environmental conditions. Ensuring the quality of data is crucial for accurate forecasting. Steps are taken to access data quality including checking for completeness, consistency and accuracy. Before training machine learning models, it is common practice to normalize the feature values to standard scale. Techniques such as Min-Max scaling or Standard scaling are applied to rescale feature values to a specified range,

ensuring that all features contribute equality to the model training process . The preprocessed data is split into training and validation sets . The training set is used to train the forecasting models , while the validation set is used to evaluate model performance . Care is taken to ensure the validation set contains time period than the training set to simulate real world forecasting scenarios .

The feature selection of feature engineering may include Traffic related features such as traffic volume , speed , congestion levels , accident data and road infrastructure . Environmental features add air pollutant concentrations , meteorological parameters such as temperature , humidity , windspeed and precipitation . Temporal features includes Time of day , day of the week , month , season , holidays , special events . Transform raw features into more meaningful representations that better capture underlying patterns and relationships . Common feature transformation includes Logarithmic transformation , Normalization , Binning / binarization and Seasonal decomposition . Generation of new features derived from the existing ones to capture additional information or interactions between variables . This may involve creating lagged versions of time – series variables to incorporate historical information known as lagged features . Multiplying or combining two or more variables to capture non-linear relationships known as interaction terms . Polynomial features and Rolling statistics are also included in feature creation . In dimensionality reduction if the feature space is high dimensional or contains redundant information , apply dimensionality reduction techniques such as principal component analysis or feature selection algorithms to reduce the number of features while preserving the most relevant features . For validation and iteration , first validate the engineered features using appropriate evaluation metrics and model performance benchmark . Iterate on the feature engineering process based on model performance feedback , domain insight and any further data exploration.

In this project , “Road Traffic Forecasting using Air Pollution and Atmospheric data “ , data scaling plays a crucial role in preparing the input features for machine learning models . Since the project involves integrating air pollution and atmospheric data along with road traffic data , scaling becomes more important to ensure that all features contribute effectively to the model training process . After feature selection and data pre-processing , data scaling is done . It is done by applying data scaling techniques to normalize the feature values and ensure that all features contribute equally to the model training process . Min Max scaling , Standardization or Z-score normalization and Robust Scaling are the widely used data scaling techniques in machine learning . Min-Max Scaling is suitable for scaling features to a specified range , thus it can be applied to features like traffic volume and air pollutant concentrations , the Standardization or Z-Score normalization can be applicable to meteorological parameters like temperature , humidity and windspeed . These scaling techniques must be performed separately for the training and testing datasets to prevent data leakage and maintain the integrity of the model evaluation process .

Since this project is a time series forecasting problem , its important to ensure the training dataset and test dataset . Depending upon the dataset , we can split the data as 70% is considered as training data and remaining 30% is to be testing data . Time series forecasting models such as Auto Regressive Integrated Moving Average (ARIMA) or Seasonal Auto Regressive Integrated Moving Average (SARIMA) are well suited for implementing temporal patterns in data. Machine Learning Algorithms such as Long Short Term Memory (LSTM) and Recurrent Neural Network (RNN) can effectively model sequential data and capture long-term dependencies . During dataset training , the models learn to identify patterns and relationship between the input features such as air pollution levels , atmospheric conditions and the target variables such as traffic flow . Hyperparameters may be tuned to optimize model performance ,

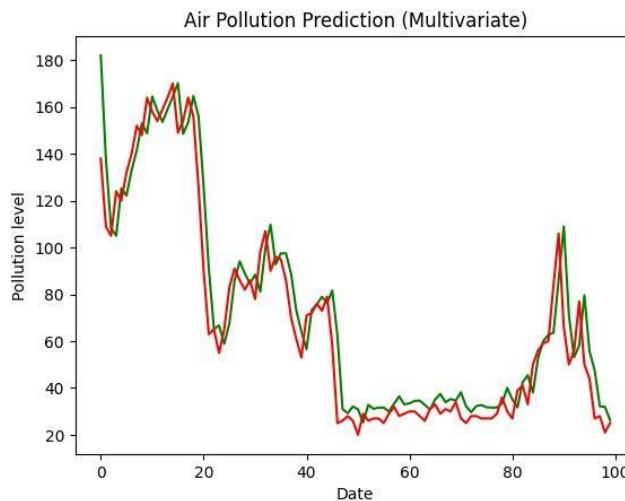
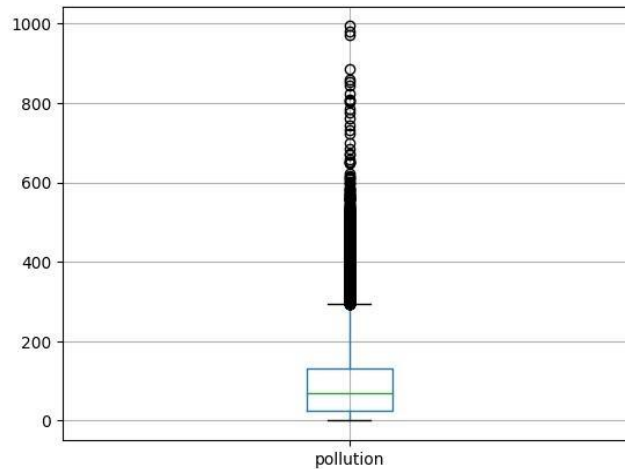
typically through methods like random search or grid search . Once the model is trained it should be evaluated using validation dataset to assess their performance and generalization ability . The most commonly used evaluation metrics are Mean Absolute Error (MAE) , Mean Squared Error (MSE) and Root Mean Squares Error (RMSE) are calculated to quantify the discrepancy between the predicted value and actual value of the model .The visual inspection of the model predictions against the actual traffic data can provide insights into the model's strengths and weakness , by helping to identify the potential areas for improvement .

For infrastructure setup , this may include cloud based platforms such as Google Cloud or Azure or AWS .It is mandatory to ensure that the infrastructure can accommodate the computational requirements of the deployed models and handle incoming traffic for predictions . Model serialization technique is also implemented to serialize the trained machine learning models into a format that can be easily located and executed by the deployment infrastructure . Common serialization methods include pickle for Python based models or Open Neural Network Exchange (ONNX) for interoperability across different frameworks . Also develop an Application Programming Interface (API) or any web service that exposes endpoints for making predictions using deployed models . Implementation of error handling and validation to ensure whether that input data meets the expected format and range . Designing the deployment architecture to be scalable and capable of handling varying levels of traffic and computational load . For model monitoring , set up monitoring and logging mechanism to track the performance and health of the deployed models in real time . Monitor key metrics such as response time , throughput and error rates to detect anomalies and performance degradation . The log prediction requests and responses for auditing , debugging and troubleshooting purposes .

3. Implementation

The project “ Road Traffic Forecasting Using Air Pollution and Atmospheric Data “ aimed to develop a machine Learning model that forecasts road traffic patterns based on environmental data , specially air pollution and atmospheric conditions . Air Pollution and atmospheric datasets were collected from reliable sources . The datasets contained information on air pollutant concentrations and meteorological parameters such as temperature , humidity . The collected datasets were preprocessed to handle missing values , outliers and inconsistency . Min-max scaling was applied to normalize the feature values to a range of [0,1] . Related features were extracted from the datasets, including both traffic related features and environmental variables. Various machine learning models were explored including sequential, dense ,dropout and Long Short Term (LSTM) neural networks . The preprocessed dataset was split into training and validation sets . The LSTM model was trained using the training dataset and validated using the validation dataset . Root Mean Square Error (RMSE) was calculated as the evaluation metric to assess the model's performance in forecasting road traffic patterns . Once trained and validated , the LSTM model was serialized and deployed to a web application . An API endpoint was created to accept input data , including air pollutant concentrations and atmospheric parameters . The deployed model processed the input data and provided traffic forecasts as output , categorizing the traffic level into categories such as poor , moderate , normal and good based on predicted traffic flow.

The trained and validated model was integrated into a website , allowing users to input air pollutant data for their location. The website provided real-time traffic forecasts based on the input data , helping users understand the traffic conditions in their area and make informed decisions about travel routes and timings.



4. Conclusion

In conclusion, our project “Road Traffic Forecasting using air pollution and Atmospheric Data “ demonstrates that machine learning models can be used to forecast air quality with a high degree of accuracy. This model has the potential to help individuals and organizations take informed action to reduce their exposure to harmful pollutants and improve public health. Overall this project showcased the potential of machine learning in addressing complex urban transportation challenges by integrating environmental data into traffic forecasting models . By providing accessible and actionable insights into traffic conditions , the project contributes to improved urban mobility and environmental sustainability . Further iterations and enhancements to the model and web application could potentially enhance accuracy and user experience , paving the way for more effective transportation management strategies in urban areas .

5. Reference

1. Deepu B P, Dr. Ravindra P Rajput, “Air Pollution Prediction using Machine Learning”, International Research Journal of Engineering and Technology (IRJET), Volume: 09 Issue: 07 | July 2022
2. Madhuri VM, Samyama Gunjal GH, Savitha Kamalapurkar, “Air Pollution Prediction Using Machine Learning Supervised Learning Approach”, INTERNATIONAL JOURNAL OF

SCIENTIFIC & TECHNOLOGY RESEARCH VOLUME 9, ISSUE 04, APRIL 2020.

3. Vidit Kumar, Sparsh Singh, Zaid Ahmed, Ms. Nikita Verma, “Air Pollution Prediction using Machine Learning Algorithms: A Systematic Review”, International Journal of Engineering Research & Technology (IJERT), Vol. 11 Issue 12, December 2022
4. Shreyas Simu*, Varsha Turkar*, Rohit Martires*, Vranda Asolkar*, Swizel Monteiro*, Vaylon Fernandes*, and Vassant Salgaoncar, “Air Pollution Prediction using Machine Learning”, ETC Department, Don Bosco College Engineering, Fatorda, Goa, India
5. K. Rajakumari, V. Priyanka, “Air Pollution Prediction in Smart Cities by using Machine Learning Techniques”, 2020, International Journal of Innovative Technology and Exploring Engineering (IJITEE), Volume 9, Issue 05.
6. Agele, Temesgen Walegin and Rutuvik Mehta. “Air Pollution monitoring and prediction using IOT” . In 2018 Second International Conference on Inventive Communication 6 w,r,t CO and Computational Technologies (ICICCT), pp. 1741-1745. IEEE 2018
7. Venkat Rao Pasupuleti, Uhasri, Pavan Kalyan. “Air Quality Prediction of Data Log By Machine Learning”, 2020, IEEE.
8. Sri Krishna Yarragunta, Mohammed Abdul Nabi, Jeyanthi P, “Prediction of Air Pollutants using Supervised Machine Learning”, 2021, IEEE.
9. Nadejet Djebri and Mounira Rouainia. “Artificial neural networks based air pollution monitoring in industrial sites” In 2017 International Conference on engineering and Technology (ICET), pp.1-5. IEEE, 2017.
10. Jiang Ningbo and Mathew L. Riley. “Exploring the utility of the random forest method for forecasting ozone pollution in SYDNEY” Journal of Environment Protection and Sustainable Development 1.5 (2015): 245-254 \
11. Noor Ullah Khan, Munam Ali Shah, Carsten Maple, Ejar Ahmed, Nabeel Asghar “Traffic Flow Prediction :An Intelligent scheme for Forecasting Traffic Flow using Air Pollution data in Smart Cities with Bagging Ensemble”, 2022.
12. Ismail Abiodun Sulaimon, Hafiz A. Alaka, “Traffic-Related Air Pollutant (TRAP) Prediction using Big Data and Machine Learning”, Conference Environmental Design and Management International Conference at Bristol, United Kingdom, July 2021.
13. Nimra Shahid, Munam Ali Shah, Abid Khan, Carsten Maple, Gwanggil Leon, “Towards greener smart cities and road traffic forecasting using air pollution data”, Sustainable Cities and Society, Volume 72, September 2021, 103062.
14. Faraz Malik Awan, Roberto Minerva, Noel Crespi, “Improving Road Traffic Forecasting Using Air Pollution and Atmospheric Data : Experiments based on LSTM Recurrent Neural Networks”, July 2020. S