World Journal of Advanced Research and Reviews

eISSN: 2581-9615 CODEN (USA): WJARAI

Cross Ref DOI: 10.30574/wjarr

Journal homepage: https://wjarr.com/

[redacted]

Air quality index prediction for Gorakhpur city using k-nearest neighbors: Model evaluation and analysis

Mandvi *, Hrishikesh Kumar Singh and Vipin Kumar

Department of Civil Engineering, Institute of Engineering and Technology, Lucknow,226021, Uttar Pradesh, India.

World Journal of Advanced Research and Reviews, 2024, 23(02), 444–454

Publication history: Received on 25 June 2024; revised on 03 August 2024; accepted on 06 August 2024

Article DOI: https://doi.org/10.30574/wjarr.2024.23.2.2373

Abstract

Emissions have increased and city air quality requirements have decreased due to rapid urbanization. Living in a city is negatively impacted by the increasing levels of toxins in the air. In these cities, the ambient air quality is measured and reported by the CAAQMS based there. This work delves deeper into applying models based on machine learning for AQI prediction. Gorakhpur City, Uttar Pradesh, India's CPCB provided the source data for this study. NO_2, SO_2, and particle matter (PM_{10} and $PM_{2.5}$) were the primary AQI pollutant measurements. This study examines the effects of Gorakhpur City's AQI on health using the K-Nearest Neighbors (KNN) method. The model finds patterns and relationships between emissions and respiratory ailments by combining temporal and spatial data on traffic density, pollutant concentrations, and climatic conditions. Recognized for its simplicity and efficacy, the KNN model forecasts possible health hazards and classes areas with high pollution levels. The results provide useful information about the KNN model that it can develop a robust model as it generates lower evaluation metrics and higher coefficient of determination, so they may put specific pollution reduction and public health protection policies into action. The model shows higher accuracy with an R^2 value of 0.985. which indicates the model's capability to recognize the larger variance in the dataset. With its machine learning tool, we can develop a robust model to forecast AQI even with a small dataset.

Keywords: Air Pollution; Machine Learning models; Air Quality Index; K-Nearest Neighbors (KNN); Air Pollution Monitoring

1. Introduction

Long-term effects of climate change on air temperatures and weather patterns may result from natural or man-made processes. The burning of fossil fuels like coal and gas, industrial development, and transportation are some of the main industries that cause air pollution. Emissions of greenhouse gases lead to global warming, which in turn leads to climate change (1).

Developing countries like India have to contend with a wide range of problems related to air pollution and its negative impacts on the environment and public health. The primary resource that allows life on Earth to exist is air. Air is an essential element of life on Earth and must be available for survival. Air quality has declined as a result of population growth, manufacturing, the combustion of fossil fuels, subpar farming methods, and automobile emissions. Particulate matter 2.5 and 10 ($PM_{2.5}$ and PM_{10}), carbon dioxide (CO_2), sulfur dioxide (SO_X), nitrogen oxide (NOx), nitrogen dioxide (NO_2), ammonia, which benzene, volatile organic compounds (VOCs), carbon monoxide (CO), and ozone are some of the air pollutants that are most frequently released from these activities (2).

Air pollution is caused by these contaminants, which are released directly from the previously indicated source. Particulate matter and gaseous pollutants are the two main categories of air pollutants. The amount and size of the

* Corresponding author: Mandvi

pollutants are determined by meteorological variables such as temperature, solar radiation, wind direction, relative humidity, wind speed, and severity of the rainfall. The air quality index (AQI) of a city or region is determined using these air pollutants. In addition to causing air pollution, these pollutants also contribute to the depletion of the ozone layer, global warming, an increase in the average earth temperature, climate change, and acid rain (3).

Pollution of the air is likely the most genuine natural problem that is resisting modern solutions. It is typically the outcome of human activities like mining, transportation, development, and modern work. The introduction of contaminants, such as poisons that threaten human health, into the atmosphere is recognized as air pollution. Affecting the general environment and health. Approximately seven million individuals die from air pollution, according to WHO estimates every year, all throughout the world (4).

Nine out of ten people now exceed the WHO's pollution guidelines, and the majority of them inhabit in nations with middle or low incomes. However, natural activities can also contribute to air pollution, such as wildfires and volcanic eruptions. Both gaseous and solid forms of air pollution are bad for the ecosystem (as airborne spreading particle matter). Air pollution's short-term health effects are just as dangerous as its long-term ones. The long-term impact of air pollution on health. Lung cancer, liver damage, and kidney damage are long-term negative consequences (5). Problems with breathing, brain damage, and damage. Among the short-term, transient side effects are nose inflammation, skin, eyes, and throat. Some examples of air pollutants are:

Carbon monoxide is released when fossil fuels, diesel, and petrol are burned. Vehicles release this gas, which cannot be seen or detected. It hurts the human lungs, making breathing more challenging.

When fossil fuels are burned, toxic air pollution is released. These are the most common causes of congenital abnormalities and cancers. Ozone, a poison, is surrounded by unstable natural chemicals when exposed to sunlight. Breathing issues, asthma, and reduced lung function are the outcomes (6).

Nitrogen dioxide and wood are two pollutants that are produced when fuels like petrol are burned in industrial boilers. These pollutants have been related to pulmonary disorders. This sulfur and oxygen-based gaseous air pollution is known as sulfur dioxide (7).

In their study, the authors provide a data analysis engine using open-source technologies and business intelligence. The air pollution study incorporates additional data sources, such as traffic and metrological information, to provide a more complete view of the issue. This data is gathered in smart city sensor networks. The combined data set is regularly assessed to create educational dashboards applying a selection of pertinent KPIs, or performance indicators. applying publicly available data as an example on air pollution in a major station's urban area use case, the suggested method will be demonstrated in a real-world smart city setting. Thus, this essay aids in becoming aware of the methods for gathering information for the suggested metropolitan city model. It also discusses how and where to collect the data (8).

Delhi is among the Indian cities that are among the most polluted in the world. The serious issues Delhi is currently facing due to high air pollution are not new. Government agencies use data on air quality to inform the public about the present and projected levels of air pollution. As a result, the prediction for air quality is a significant and effective methodology that aids in the analysis and forecasting of air quality. In the project Initially, raw data is collected from a specific city (Bangalore), and then data preparation is used methods for handling data. The sci-kit learn module is used in the following step to partition the data and test the data. When an unknown input is presented, the machine learning model is designed to be able to generalize the observed data using the training data, roughly predicting the output. Using training and testing data, the accuracy method is utilized to analyze the classifiers and, once scored, determine which classifier is the best (9).

Our objective in this effort was to create an enhanced system that could surpass all current systems in use. Throughout this process, we have examined numerous studies and come across a wide range of findings from the systems that are now in use. We have gone through a range of papers and research theses, analyzed the findings of other researchers, and implemented several tactics based on their studies and conclusions. The goal of the last few years air quality research has been to develop the best possible technique for air quality prediction (10).

Through this effort, a better system was developed that could outperform all other systems in terms of performance. Many discoveries about the current systems were made during this procedure. Numerous journals and research theses were examined in order to examine the findings of different researchers, who then used their findings to adopt a variety of strategies. For the past few years, research on air quality has been conducted to create an ideal method for predicting air quality. Numerous articles have been written outlining the necessity of air quality prediction in addition to the steps involved in making various system modifications to increase accuracy compared to earlier systems (11).

Numerous academics have presented a variety of strategies for using air quality prediction to overcome a range of obstacles. The papers that served as the foundation for the construction of our thesis are reviewed and validated by our study. By citing numerous theses and publications that have been suggested by different studies, the review clarifies how learning is done in this situation (9).

The author conducted research for the work in Beijing and Italy, two different cities. He predicted Beijing's AQI and the amounts of NOx in Italian cities using two publicly accessible statistics. On both points, he was correct. The fundamental dataset, which includes hourly averages for AQI, $PM_{2.5}$, O_3, SO_2, PM_{10}, and Beijing NO_2, spans 1738 occurrences from December 2013 to August 2018. It was contributed by the Beijing Municipal Environmental Centre. Between March 2004 and February 2005, Italian municipalities were visited to collect the second batch of 9358 cases. This dataset contains amounts of NOx, NO_2, benzene, and non-methane volatiles. Since NOx prediction is one of the most important indicators of air quality, we focused only on it (12).

CPCB, India, kindly supplied data on air pollution for Gorakhpur City, Uttar Pradesh, India, for this study. The dataset was gathered between July 2021 and December 2023. To estimate the AQI, we create machine learning models called K-Nearest Neighbors (KNN). In the comparison study, evaluation metrics such as R^2, RMSE, MAE, and MSE were compared to assess how well the machine learning model performed. One of the most widely used machine learning models is the KNN model because of its excellent prediction accuracy (2).

2. Material and methods

2.1. Study area

The selected location for the study is the city of Gorakhpur. It is situated in the state of Uttar Pradesh, India, on the bank of the Rapti River. From a geographic perspective, Gorakhpur is situated about approximately 26.76° N and 83.37°E. The research region has a specific landscape and encompasses an urban area of approximately 3483.8 sq. km. The city has become especially vulnerable to air pollution because of its high population, and the rapid industrial and development of the area only make things worse. Vehicle emissions, burning of agricultural waste, and other sources that lead to abnormally high concentrations of PM_{10}, $PM_{2.5}$, SO_2, and NO_2, are the main causes of air pollution in the city. Located at the MMMUT, Deoria Road, Singhariya, Kunraghat, Gorakhpur, is the CPCB monitoring station (13).

2.2. Methodology

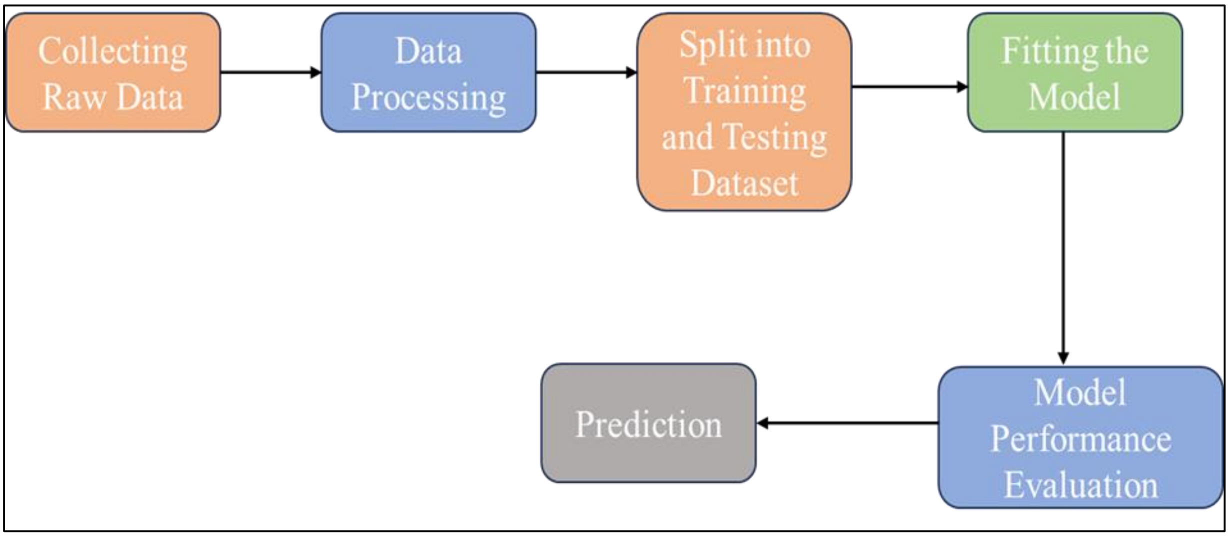

Figure 1 Flowchart of methodology

The data collection process for Gorakhpur city is now complete. Stage one in the data preprocessing technique was an exhaustive review of the dataset to ensure that no null values existed. Following feature selection, exploratory data analysis was performed to assess the relationship between the features. After that, the dataset was split into training and testing datasets, with 20% of the dataset used for testing and the remaining 80% being used to train the model. This process involves choosing a model, hyperparameter modifying, model training, and cross-validation. A model was applied to the test dataset to produce predictions after being trained on the training dataset (14).

2.3. Pre-Processing of Data

A K-Nearest Neighbor (KNN) model has been utilized in the application of predictive analysis. The dataset underwent pre-processing to fix any null or missing values. The training datasets, X and Y trains, were used to generate and train instances of the scikit-learn Linear Regression class. Cross-validation was used to evaluate the model's performance. For visual purposes, the distribution and scatter plots of the actual value against the expected value were made. Data preprocessing is the process of converting unprocessed data into a format that machine learning algorithms may use to learn from or forecast results. The search for missing values is the data processing method applied in this research (15). Obtaining all the data points for each record in a dataset is challenging. Fill in the values in any empty cells that may exist. The dataset used for the study contained no missing values. The most crucial need for efficient data visualisation and machine learning model building is ensuring data quality. To increase processing speed and the generalisability of machine learning algorithms, data preparation is necessary to minimize noise in the data.

2.4. Fitting the model

Fitting is the process of adjusting a model's parameters to increase its accuracy. An algorithm is run on data for which the target variable is known to build a machine-learning model. A model's accuracy needs to be evaluated by contrasting it with actual, observed values of the target variable. Model fitting is the process by which a machine learning model can generalize data that it was trained on. A well-fitting model is one that accurately forecasts the outcome with unknown inputs.

2.5. Air quality index

The issue has been evaluated using the Air Quality Index (AQI). According to standards set forth by the CPCB, India, the average AQI score falls between 0 and 500. The highest possible index value denotes acute air pollutants, which have detrimental impacts on both human health and ecology. Similarly, the purest air is indicated by the lowest AQI index number. The lowest AQI readings indicate the concentration of various air pollutants in the atmosphere within the allowed limits for each pollutant. The average 24-hour period was used to collect Together with AQI readings, CPCB provides information on each pollutant's health impact. Values of AQI 0–50 indicate the least influence on health. 51–100 may result in minor respiratory issues in vulnerable individuals. For people who have respiratory issues or lung diseases, 101–200 could be difficult. In patients with cardiac conditions, brief exposure to 201-300 may cause pain. Extended exposure has been linked to respiratory illnesses (301–400) (17). Those with pre-existing heart or lung conditions may experience a more pronounced impact. Patients with lung and cardiovascular diseases may experience substantial adverse effects from the extreme level of AQI (401–500), even in healthy persons the air quality data for this investigation (18).

2.6. AQI prediction using machine learning technique

Although deep learning models have proven useful in many applications, classical machine learning models might be more appropriate and simpler to understand for a relatively small dataset, such as the AQI dataset utilized in this study. The training of the model was done with the best settings and scores were reported, along with the fit () procedure. Lastly, we assessed the model's effectiveness using the test set. To guarantee the for the model's robustness, we employed 5-fold cross-validation, which calculated for every fold, the mean, MAE, MSE, RMSE, and R^2 scores were reported, and each metric's standard deviation over all folds. When a model possesses a lower MAE and MSE score and a higher R^2 score, it is typically thought to be operating more effectively.

2.7. Exploratory data analysis (EDA)

One of the most important steps in creating a K-Nearest Neighbors (KNN) model is exploratory data analysis. It entails analyzing the dataset to determine its primary attributes before utilizing any machine learning methods. To obtain an understanding of the data's central tendency and variability, we begin the EDA process by summarising the data using descriptive statistics like mean, median, mode, and standard deviation. To find trends, correlations, and possible outliers, visualisation tools including scatter plots, box plots, and histograms are used. Understanding the data distribution and the relationships between variables is made easier with the aid of these visual aids, which is essential for distance-based algorithms such as KNN.

Another important component of EDA is correlation analysis, which looks at the direction and strength of correlations between variables. Due to the possibility of redundant information provided by strongly correlated features, this aids in the selection of the most pertinent characteristics for the KNN model. EDA also aids in the detection and management of categorical variables and missing values, guaranteeing that the data is clear and appropriate for modelling. To guarantee that every feature contributes equally to the distance computations in KNN, normalization or scaling of features is frequently carried out during EDA. All things considered, EDA offers a thorough comprehension of the dataset, guaranteeing that the KNN algorithm's application, later on, is founded on organized and perceptive facts, producing predictions that are more accurate and trustworthy.

3. Result

In this study, our regression objective was the Air quality index (AQI) for Gorakhpur City. For this research objective, we have employed K-Nearest Neighbor (KNN) for Air quality prediction. The dataset used in this study was collected from CPCB. The dataset for evaluation purposes is split into two, one for training the model and second for the testing the model. The whole dataset is divided into a ratio of 80:20 means 80 % of the data is used for training the model and the rest 20% of the data for testing purposes. For processing the dataset, Python modules such as NumPy, Seaborn, Pandas, and, Scikit-learn were used.

For the determination of the major influence of the pollutant concentration on Air quality value, we explored the dataset in different ways. The correlation analysis was performed to determine the relationship between different pollutants and meteorological factors, it also showed the connection between different climatic data with AQI values. For analysis of the model results, we used evaluation metrics like mean absolute error (MAE), root mean square error (RMSE), mean squared error (MSE), and coefficient of determination (R^2).

3.1. Data Investigation and Visualization

To develop a reliable and accurate machine-learning model for forecasting AQI, data exploration and visualization methods are very important as they provide valuable understanding into what factors influence more to the AQI value. Methods such as pairplot and feature importance were used for this objective. The pairplot, python Seaborn library feature depicts the relationship between every set of variables in the dataset. It also offers a thorough analysis of scatterplots and histograms. In correlation metrics, every variable is plotted against each other, and the connection between the pairs of variable factors in the lower and upper triangle of the grid is depicted by scattered plots and histograms and shows the distribution of individual variables along the diagonal.

This improves the assessment of patterns, trends, and relationships between the different air pollutants and other variables.

3.2. Machine learning model to predict AQI

In the optimization stage of analysis, cross-validation is used to determine the best performance over different data subsets. To ensure a more accurate evaluation of the model's capabilities, this approach includes training and testing of the model over various folds of the data. by this method, it was found that the model was not overfitted or under-fitted over the training and testing of the data across all the folds. the model's performance was evaluated based on the error metrics like MAE, MSE, RMSE, and R^2.

3.3. Result of KNN model

The results show that the K-Nearest Neighbors (KNN) model performs very well. MAE, MSE, RMSE, and R^2 are the metrics that are offered. The value of the error metrics is shown in Fig.2, Fig.3, and Table 1.

3.3.1. Mean absolute error (MAE)

This matrix represents the average absolute difference between the actual and predicted AQI values. The average error in this case is approximately 4.061 units. In this instance, the MAE shows that on average there is a difference of 4.061 units between actual and predicted readings.

3.3.2. Mean squared error (MSE)

MSE calculates the average of the squared differences among actual and predicted values. The bigger errors get hit more than the smaller ones by the squaring process. The MSE value for this study is observed as 34.567 which indicates that the actual and predicted values have significant differences.

3.3.3. Root Mean Squared Error (RMSE)

RMSE is the square root of the MSE and reduces the error metric to the target variable's actual unit which is AQI. In this study, the RMSE value is 5.879, which denotes that there is an average of 5.879 AQI value that deviates from the actual value of AQI. With lower values, a greater prediction is shown which is why it is an important indicator of model performance.

3.3.4. Coefficient of determination (R^2)

For the training set the R^2 value is observed as 1.0 and for testing the R^2 value is observed as 0.985. In the training phase, it reflects 100% accuracy and proves that the model is capable of explaining 100%vatiance in the training dataset. The R^2 value for the testing data set suggests that the model is capable of explaining 98.5%of variance in the test dataset. this great value indicates that the model generalizes well to unseen data and shows a robust predicting power. Results shown in Fig.1.

Table 1 Evaluation metrics for the KNN model

MAE	**4.061**
MSE	34.567
RMSE	5.879
R^2	0.985

```
44]: # Coefficient of determination R^2
     print("Coefficient of determination R^2 <-- on train set:", regressor.score(X_train, y_train))

     Coefficient of determination R^2 <-- on train set: 1.0

45]: print("Coefficient of determination R^2 <-- on test set:", regressor.score(X_test, y_test))

     Coefficient of determination R^2 <-- on test set: 0.9852411714393856
```

Figure 2 Coefficient of determination values on test and train datasets

```
[50]: # Evaluation metrics
      print('MAE:', metrics.mean_absolute_error(y_test, prediction))
      print('MSE:', metrics.mean_squared_error(y_test, prediction))
      print('RMSE:', np.sqrt(metrics.mean_squared_error(y_test, prediction)))

      MAE: 4.061032645817183
      MSE: 34.56794888942638
      RMSE: 5.879451410584697
```

Figure 3 Evaluation metrics values on the test dataset

4. Discussion

4.1. Model accuracy

The KNN model performance is quantitively evaluated by the evaluation metrics like MAE, MSE, RMSE, and, R^2. A smaller MAE value shows that the predicted values by the model are approximately the same as those of actual AQI values. The RMSE value penalizes higher errors more than the MAE which indicates the presence of certain larger errors in the forecasting of AQI.

The R^2 shows that the KNN model has a greater degree of accuracy in forecasting AQI, on both training and testing data sets. the perfect R^2 value on the training set depicts an excellent fit and with the R^2 value on the test set, it is confirmed that the model is free from overfitting. Fig.4 indicates the accuracy of the model.

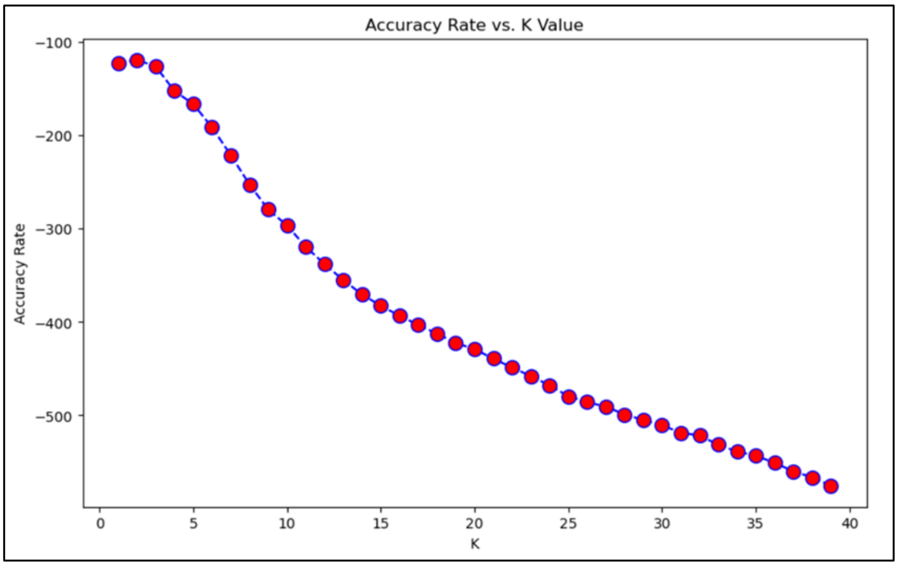

Figure 4 KNN model Accuracy rate vs k value

4.2. Model suitability

These evaluation metrics give an overview of the model's performance. the results of cross-validation provide insight on how strong the model is across various dataset. the model suitability can be checked by comparing different models together with this KNN mode.

4.3. Outcome analysis and visualization

4.3.1. Scattered plot

A scatter plot graphically depicts the connections between actual and predicted values in regression. The predicted values are shown on the y-axis and the actual values on the x-axis. the predicting errors are indicated by points that deviate from the diagonal line, which indicates the more accurate forecast. In general, the scattered plots are used to evaluate the correctness of the model, identify the differences between actual and predicted values, and represent the patterns as represented in Fig.5.

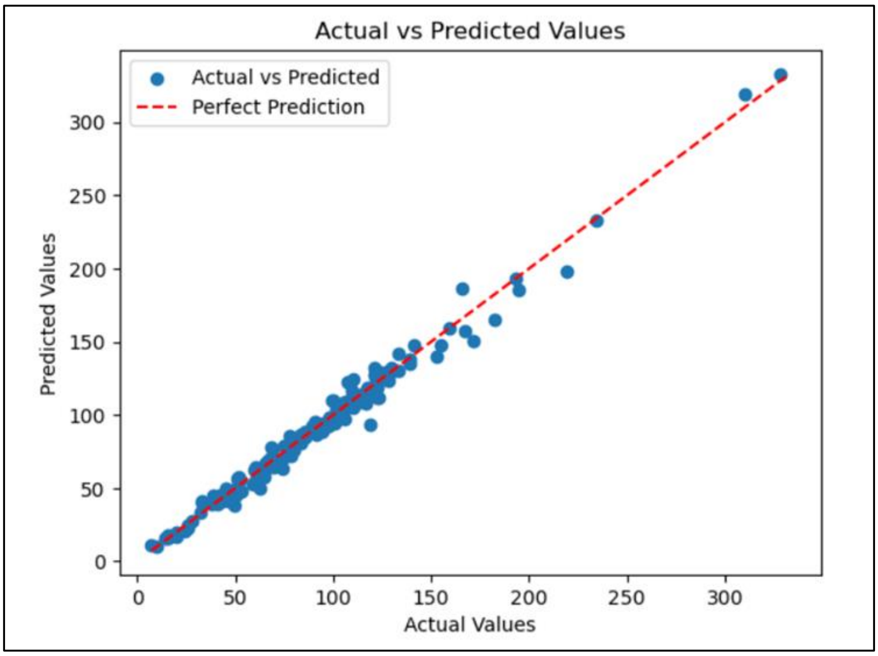

Figure 5 Scattered plot for KNN model

4.3.2. Residual analysis

One of the very important stages in evaluating a regression model's performance is residual assessment, additionally referred to as residual analysis. the discrepancies between actual and predicted values generated by the models are known as residuals.

To identify whether the regression algorithm satisfies the underlying assumption and to determine any patterns or trends in the model's mistakes, the residual analysis includes exploring these residuals as shown in Fig 6.

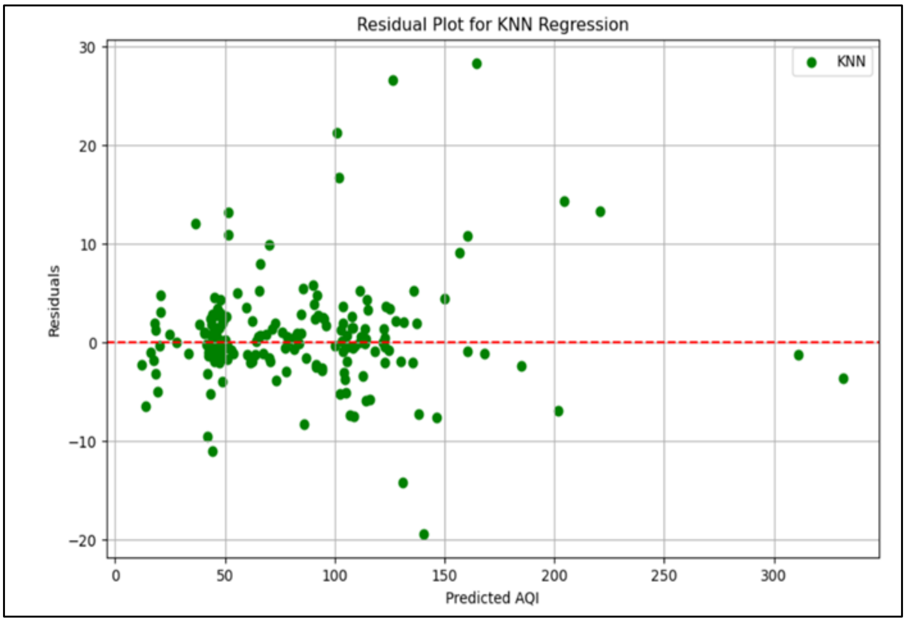

Figure 6 Residual plot for KNN model

4.3.3. Distplot histogram analysis

A kernel density estimates (KDE) plot and histogram are utilized to generate distribution plots (distplots). this offers a visual interpretation of a dataset distribution providing details on the data skewness, outliers, data distribution, and density estimates represented in Fig.7.

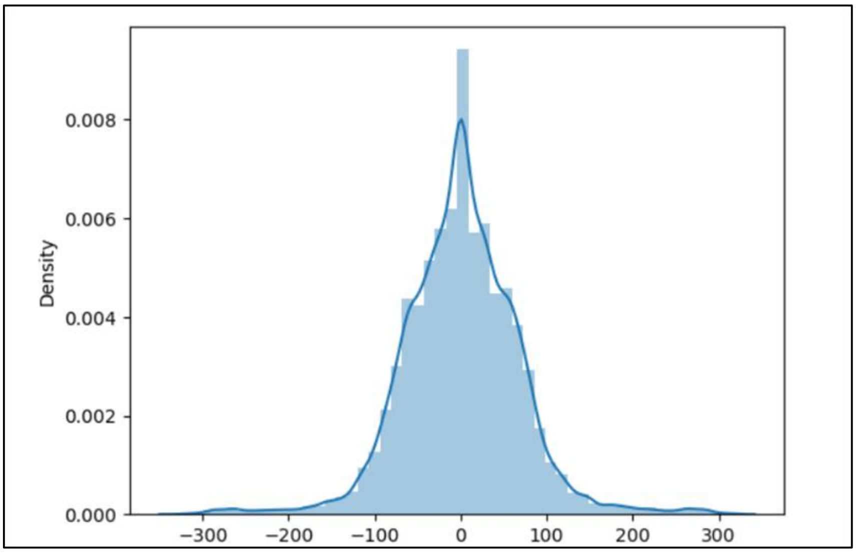

Figure 7 Distplot histogram for the KNN model

Abbreviation

- AQI Air Quality Index
- ML Machine Learning
- MAE Mean Absolute Error
- MSE Mean Squared Error
- RMSE Root Mean Squared Error
- CPCB Central Pollution Control Board
- SO_2 Sulfur dioxide
- NO_2 Nitrogen dioxide
- PM Particulate Matter
- CO Carbon monoxide
- MMMUT Madan Mohan Malaviya University of Technology.
- CAAQMS Continuous Ambient Air Quality Monitoring Station.

5. Conclusion

The evaluation metrics and research represent that the KNN model predicts the AQI value for Gorakhpur city with good accuracy. The MAE, MSE, RMSE, and, R^2 depict a low value indicating the model's good precision. The low MAE value of 4.061, MSE value of 34.567, RMSE value of 5.879, and R^2 value of 0.985 on the testing set and 1.0 on the training set indicates that the model is free from overfitting and well generalized to a new data and an outstanding match.

The model predictions are verified to be accurate and dependable by scatter plots, residual plots, and distplot histogram assessment. the residual shows a normal distribution and does not exhibit insignificant patterns.

Overall, the KNN model is significantly reliable and produces accurate AQI predictions Maintaining optimal performance necessitates ongoing validation and comparison with other models

Compliance with ethical standards

Acknowledgments

The authors would like to thank the Civil Engineering Department, Institute of Engineering and Technology, Lucknow, and CPCB, India for providing guidance, data, and support required to conduct the study.

Disclosure of conflict of interest

The authors confirmed that they had no conflicting financial interests or personal connections that might have appeared to have influenced the work reported in this paper.

Authors contribution

Mandvi: Data curation, Writing-Original draft preparation, methodology, software, Investigation, formal analysis, **Hrishikesh Kumar Singh:** Supervision, Conceptualization, **Vipin Kumar**: Writing- review & editing, Visualization, Validation.

Data Availability Statement

The data was collected from CPCB, India.

References

[1] Ravindiran G, Rajamanickam S, Kanagarathinam K, Hayder G, Janardhan G, Arunkumar P, et al. Impact of air pollutants on climate change and prediction of air quality index using machine learning models. Environ Res. Academic Press Inc.; 2023; 239.

[2] Atmakuri KC, Prasad K V. Urban Air Quality Analysis And Aqi Prediction Using Improved Knn Classifier. Journal of Pharmaceutical Negative Results ¦. [date unknown]; 13:2022.

[3] Bodor, Z., Bodor, K., Keresztesi, Á., & Szép, R. (2020). Major air pollutants seasonal variation analysis and long-range transport of PM10 in an urban environment with specific climate condition in Transylvania (Romania). Environmental Science and Pollution Research, 27(30), 38181–38199. https://doi.org/10.1007/s11356-020-09838-2

[4] Dewan, S., & Lakhani, A. (2022, December 15). Tropospheric ozone and its natural precursors impacted by climatic changes in emission and dynamics. Frontiers in Environmental Science. Frontiers Media S.A. https://doi.org/10.3389/fenvs.2022.1007942

[5] Janarthanan, R., Partheeban, P., Somasundaram, K., & Navin Elamparithi, P. (2021). A deep learning approach for prediction of air quality index in a metropolitan city. Sustainable Cities and Society, 67. https://doi.org/10.1016/j.scs.2021.102720

[6] Kumar, K., & Pande, B. P. (2023a). Air pollution prediction with machine learning: a case study of Indian cities. International Journal of Environmental Science and Technology, 20(5), 5333–5348. https://doi.org/10.1007/s13762-022-04241-5

[7] Kumar, K., & Pande, B. P. (2023b). Air pollution prediction with machine learning: a case study of Indian cities. International Journal of Environmental Science and Technology, 20(5), 5333–5348. https://doi.org/10.1007/s13762-022-04241-5

[8] Liu, H., Li, Q., Yu, D., & Gu, Y. (2019). Air quality index and air pollutant concentration prediction based on machine learning algorithms. Applied Sciences (Switzerland), 9(19). https://doi.org/10.3390/app9194069

[9] Vipin Kumar, Mandvi, Shashank Pandey, Vishvanath Pratap Singh. Assessment of ambient air quality of Lucknow city post monsoon 2023. World Journal of Advanced Research and Reviews [Internet]. 2024; 23(1):188–201. Available from: https://wjarr.com/node/13281.

[10] Baidya S, Borken-Kleefeld J. Atmospheric emissions from road transportation in India. Energy Policy. 2009; 37(10):3812–22.

[11] Duan M, Sun Y, Zhang B, Chen C, Tan T, Zhu Y. PM2.5 Concentration Prediction in Six Major Chinese Urban Agglomerations: A Comparative Study of Various Machine Learning Methods Based on Meteorological Data. Atmosphere (Basel). MDPI; 2023; 14(5).

[12] Vo LHT, Yoneda M, Nghiem TD, Sekiguchi K, Fujitani Y, Shimada Y. Seasonal variation of size-fractionated particulate matter in residential houses in urban area in Vietnam: relationship of indoor and outdoor particulate matter and mass size distribution. E3S Web of Conferences. EDP Sciences; 2022.

[13] Vo LHT, Yoneda M, Nghiem TD, Sekiguchi K, Fujitani Y, Shimada Y. Seasonal variation of size-fractionated particulate matter in residential houses in urban area in Vietnam: relationship of indoor and outdoor particulate matter and mass size distribution. E3S Web of Conferences. EDP Sciences; 2022.

[14] Fabregat A, Vernet A, Vernet M, Vázquez L, Ferré JA. Using Machine Learning to estimate the impact of different modes of transport and traffic restriction strategies on urban air quality. Urban Clim. Elsevier B.V.; 2022; 45.

[15] Rodriguez-Rey D, Guevara M, Linares MP, Casanovas J, Salmerón J, Soret A, et al. A coupled macroscopic traffic and pollutant emission modelling system for Barcelona. Transp Res D Transp Environ. Elsevier Ltd; 2021; 92.

[16] Wang Y, Liang X, Wang Y, Yu H. Effects of Viscosity Index Improver on Morphology and Graphitization Degree of Diesel Particulate Matter. Energy Procedia. Elsevier Ltd; 2017.

[17] Singh N, Mishra T, Banerjee R. Emissions inventory for road transport in India in 2020: Framework and post facto policy impact assessment [Internet]. [date unknown]. Available from: https://doi.org/10.21203/rs.3.rs-297185/v1.

[18] Kumar V, Kumar Patel P. Change in the concentration of pollutants in the air over the city of Lucknow, together with HYSPLIT4.0's trajectory and dispersion analysis [Internet]. 2024. Available from: https://doi.org/10.21203/rs.3.rs-4295589/v1.