Volume 9, Issue 4, April – 2024 International Journal of Innovative Science and Research Technology
ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24APR1706

Air Quality Prediction usingKNN and LSTM

K.V.V. Ganesh[1]; G. Sheetal[2]
M. Sai Amith[3]; L. Harshith Goyal[4]
Students
Department of CSM, RaghuEngineering College,
Dakamarri(V), Bheemunipatnam, Visakhapatnam
District Pin Code: 531162

K. Srinivasa Rao[5] M Tech (Ph.D)
Assistant Professor
Department of CSE, Raghu EngineeringCollege,
Dakamarri(V), Bheemunipatnam, Visakhapatnam
District Pin Code: 531162

Abstract:- The project titled "Air Quality Prediction Using KNN and LSTM" endeavors toaddress the critical issue of air pollutionthrough the application of advanced computational techniques. The project aimsto develop a robust predictive model that can forecast air quality levels based on historical data, meteorological parameters, and relevant environmental features. Leveraging machine learning algorithms such as regression, decision trees, or neural networks, the project seeks to analyze complex relationships within the data and enhance the accuracy of air qualitypredictions.

The methodology involves the collection and preprocessing of extensive datasets encompassing pollutant concentrations, weather conditions, and geographicalinformation. The selected machine learning algorithms will be trained on this data to recognize patterns and correlations, enabling the model to make accurate predictions. The project also explores the integration of real- time data streams, satellite imagery, and sensor networks to improve the responsiveness of the predictive model.

Keywords:- *Air Quality Prediction, Machine Learning Algorithms, Linear Regression, Decision Tree, Random Forest, K-Nearest Neighbours (KNN), Long Short-Term Memory (LSTM), Ensemble Learning,Hybrid Models.*

I. INTRODUCTION

Air pollution has emerged as a significant global challenge, with adverse implications for human health, the environment, and overall well-being. As urbanization and industrialization continue to escalate, the need for effective air quality management becomes imperative. This, titled "Air Quality Prediction Using KNN and LSTM," seeks to address this critical issue by harnessing the power of advancedcomputational models to forecast air qualitylevels.

The project recognizes the intricate interplay of various factors contributing to air pollution, including pollutant concentrations, meteorological conditions, and geographical features. Traditional methods of air quality prediction often struggle to capture the complexity of these relationships. In response, the project adopts a cutting-edge approach, employing machine learning algorithms to analyze vast datasets and uncover patterns that may elude conventional techniques.

The primary objective of the project is to develop a robust predictive model capable of accurately forecasting air quality indices. To achieve this, the student will explore a range of machine learning algorithms such as regression, decision trees, and neural networks. These algorithms will be trained on historical data, enabling them to learn and adapt to the intricate dynamics of air quality variation.

II. LITERATURE REVIEW

- *A Comparative Analysis of Machine Learning Algorithms forAir Quality Prediction" (2020)*

- This review compares the performance of linear regression, decision trees, Random Forest, KNN, and LSTM models in predicting air quality parameters such as PM2.5, O3, and NO2. It evaluates accuracy, computational efficiency, and suitability for real-time forecasting.

- *"Enhancing Air Quality Prediction Through Ensemble Learning and Deep Neural Networks" (2019)*

- The paper explores the integration of ensemble methods like Random Forest and gradient boosting with deep learning architectures such as LSTM for improvingair quality prediction accuracy. It investigates feature engineering techniques and model fusion strategies for optimal results.

- *"Spatial-Temporal Modeling of Air Pollution Using Machine Learning Techniques" (2021)*

- This literature review focuses on the spatial-temporal modeling of air pollution using machine learning algorithms. It discusses the application of linear regression, decisiontrees, and spatiotemporal models like LSTM to predict air quality variations across different geographicallocations.

- *"Predicting Urban Air Quality with Hybrid Machine LearningModels" (2018)*

- The review examines hybrid machine learning models that combine linear regression, decision trees, and neural networks to predict urban air quality. It investigates ensemble strategies and hybridization techniques for leveraging the strengths of multiple algorithms.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24APR1706

➢ *"Real-Time Air QualityForecasting Using IoT Data and Machine Learning" (2022)*

- This literature review discusses real-time air quality forecasting methodologies using IoT sensor data and machine learning algorithms. It covers the implementation of linear regression, decision trees, Random Forest, KNN, and LSTM models in an IoT environment for timely and accurate predictions.

➢ *Existing System*

The existing systems for air quality prediction often employ a combination of traditional statistical methods and simplistic modeling approaches. These systems commonly rely on historical data analysis, mathematical formulas, and rule- based systems to forecast air quality indices. Some key aspects of the existing systems include:

- *Statistical Models:*

These models analyze historical air quality data to identify patterns and trends. They often use statistical methods such as time series analysis or correlation analysis to predict future air quality based on past observations.

- *Linear Regression :*

It is used to model the relationship between air quality parameters (such as pollutant concentrations, weather conditions, and geographical factors) and predict air quality levels.

➢ *However, the Existing System has Several Limitations and Drawbacks :*

- *Reliance on Historical Data :*

The reliance on historical data limits the system's ability to accurately predict sudden fluctuations or anomalies in air quality. It may not capture real-time changes in environmental conditions that can impact air quality.

- *Lack of Real-time Updates :*

Since traditional methods often require batch processing of data and manual analysis, they may not provide real-time updates on air quality conditions. This delay in information dissemination can affect timely decision-making and public awareness.

➢ *Lower Accuracy in Predicting*

- *Fluctuations:*

Traditional models may struggle to accurately predict air quality fluctuations caused by dynamic factors such as rapid weather changes, local emissions, or unusual events (e.g.,wildfires, industrial accidents).

- *Limited Adaptability :*

The static nature of traditional models makes them less adaptable to changing environmental conditions or new data trends. They may not incorporate feedback loops or self- learning mechanisms to improve prediction accuracy over time.

- *Scalability Issues :*

As air quality monitoring networks expand and generate more data, traditional methods may face scalability challenges in processing and analyzing large volumes of information efficiently.

➢ *Proposed System*

In our proposed system for air quality prediction, we will leverage advanced machine learning algorithms to significantly enhance prediction accuracy and provide real-time updates. Specifically, we plan to utilize regression algorithms such as linear regression, decision trees, random forests, and possibly neural networks. These algorithms excel in handling complex data patterns, adapting to changing environmental factors, and delivering accurate predictions.

- *Key Features of the Proposed System Real-time Data Collection :*

Our system will gather real-time data from diverse sources including air quality sensors, weather stations, satellite imagery, andpossibly IoT devices. This continuous data flow ensures up-to-date information for robust predictions.

- *Data Preprocessing :*

Before feeding data into machine learning models, we will perform comprehensive data preprocessing steps. This includes data cleaning, handling missing values, feature engineering, and normalization to ensure high-quality and reliable input for model training.

- *Machine Learning Model Selection :*

The selection of machine learning models will be based on thorough evaluation of their performance metrics, scalability, and suitability for air quality prediction tasks. We will assess models such as linear regression for baseline performance and then explore more complex models like decision trees and random forests for improved accuracy.

- *Model Training and Validation :*

Our system will undergo rigorous model training using historical data, and we will employ techniques like cross-validation or holdout validation to validate model performance.

This iterative process ensures robustnessand reliability in predicting air quality levels.

- *Integration with user Interface :*

The predicted air quality values will be seamlessly integrated into a user-friendly interface or dashboard. Stakeholders, including authorities and the public, caneasily access and interpret the air quality predictions, enabling informed decision- making and timely actions.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24APR1706

➢ *Expected Benefits :*

- *Improved Accuracy :*

The proposed system is expected to achieve significantly higher accuracy in predicting air qualitylevels compared to traditional methods. Machine learning algorithms can effectively capture complex data patterns and adapt to changing environmental conditions, leading to more precise predictions.

- *Timely Warnings and Alerts :*

With real-time updates and accurate predictions, our system has the capability to provide timely warnings and alerts to authorities and the public about potential air quality issues. This proactive approach enables prompt responses and mitigation measures.

- *Environmental Impact Assessment:*

Accurate air quality prediction contributes to better environmental management and decision-making. It facilitates the implementation of targeted pollution control measures, issuance of health advisories, and overall improvement inenvironmental sustainability.

III. METHODOLOGY

➢ *Data Collection and Preprocessing*

Gather historical air quality data from reliable sources such as government agencies or research institutions. Include parameters like PM2.5, PM10, O3, NO2, SO2, CO, and meteorological variables (temperature, humidity, wind speed, etc.). Clean the data by handling missing values, outliers, and inconsistencies. Perform data normalization or standardization to ensure uniformity across features.

➢ *Feature Selection and Engineering*

Conduct feature selection to identify relevant variables that significantly impact air quality. Techniques like correlation analysis, mutual information, and feature importance from tree-based models can guide feature selection.

Engineer new features if necessary, such as temporal features (hour of day, day of week), lagged variables, and interaction terms to capture complex relationships.

➢ *Model Training and Evaluation*

Split the data into training and testing sets (e.g., 80% training, 20% testing) or use cross-validation for model evaluation.

- *Train the Following Machine Learning Models :*

✓ Linear Regression : Fit a linear regression model to predict air quality based on selected features.
✓ Decision Tree: Build a decision tree to capture non-linear relationships in the data.
✓ Random Forest: Construct an ensemble of decision trees for improved prediction accuracy and robustness.
✓ K-Nearest Neighbors (KNN): Implement KNN for instance-based learning and localized predictions.
✓ Long Short-Term Memory (LSTM): Develop an LSTM network to model temporal dependencies and sequences in air quality data.

Evaluate each model using performancemetrics such as **mean squared error (MSE)**, **mean absolute error (MAE)**, **R- squared**, and **accuracy** (for classification tasks).

Compare the performance of different models to identify strengths and weaknesses.

➢ *Model Optimization and Hyperparameter Tuning*

Fine-tune model hyperparameters using techniques like grid search, random search, or Bayesian optimization to optimize modelperformance.

Conduct sensitivity analysis to assess the impact of hyperparameters on model outcomes and stability.

➢ *Ensemble and Hybrid Model Integration*

Explore ensemble techniques to combine the strengths of individual models. For example, ensemble methods like modelaveraging, stacking, or blending can be applied.

Consider hybrid approaches that combine linear regression with decision trees or incorporate deep learning components (LSTM) into ensemble models for improved accuracy and generalization.

➢ *Model Deployment and Monitoring*

Deploy the optimized and validated models for air quality prediction in a real-time or batch processing environment.

Implement monitoring mechanisms to track model performance over time, detect drifts or concept shifts, and retrain/update models as needed.

Communicate results effectively through visualizations, dashboards, and reports to stakeholders and end-users.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24APR1706

IV. RESULTS

- *Predicted AQI using KNN (Good)*

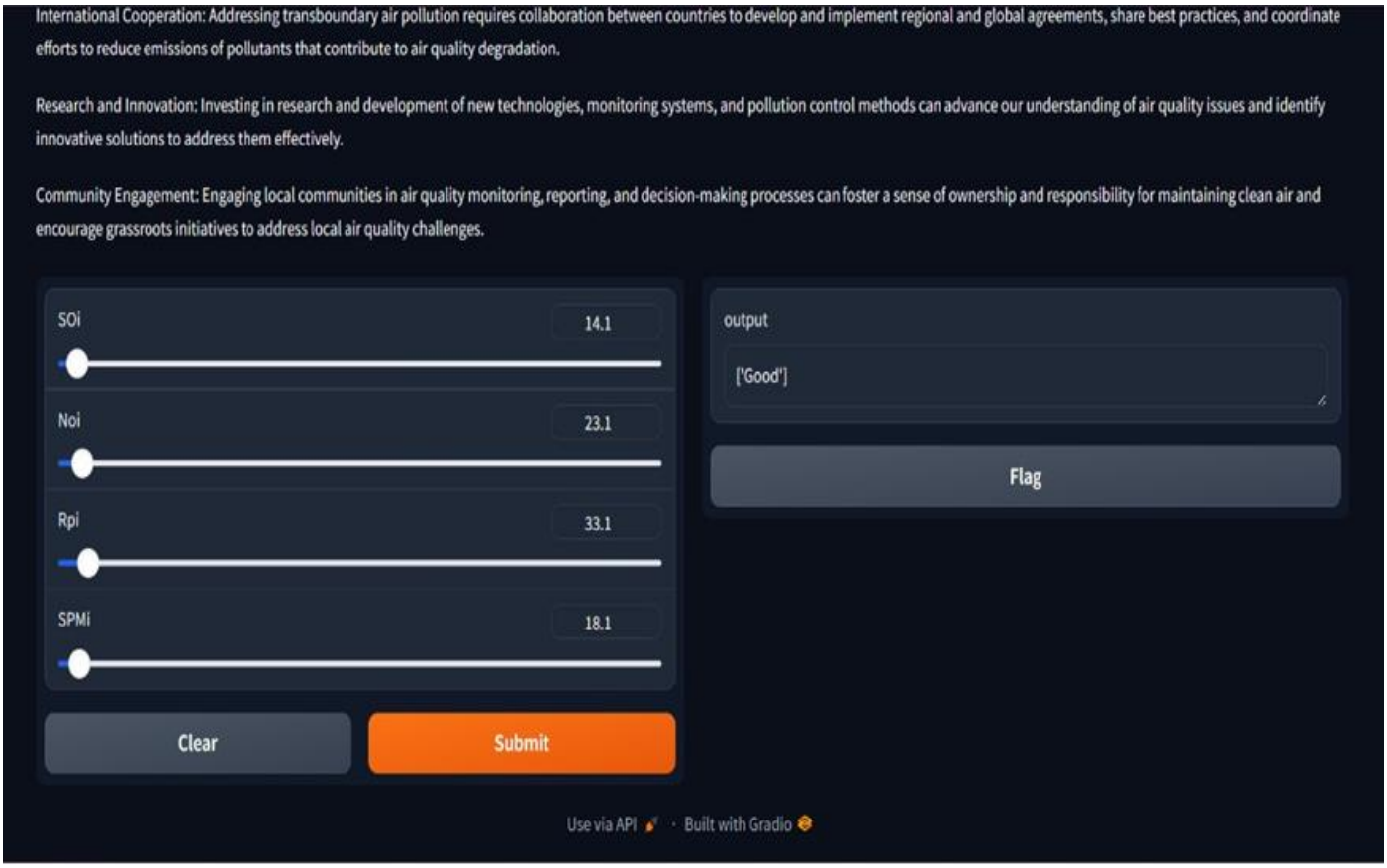

Fig 1 Predicted AQI using KNN (Good)

- *Predicted AQI using LSTM (Good)*

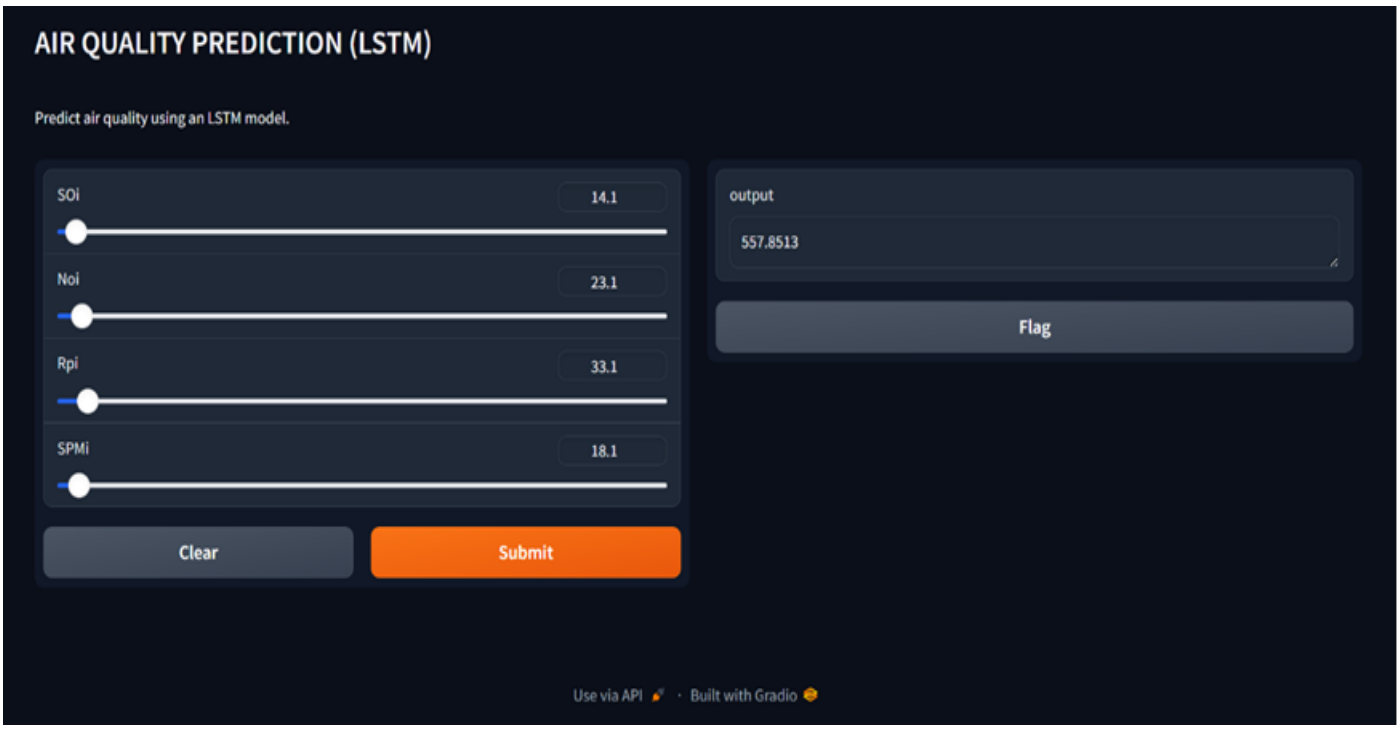

Fig 2 Predicted AQI using LSTM (Good)

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24APR1706

➢ *Predicted AQI using KNN (Moderate)*

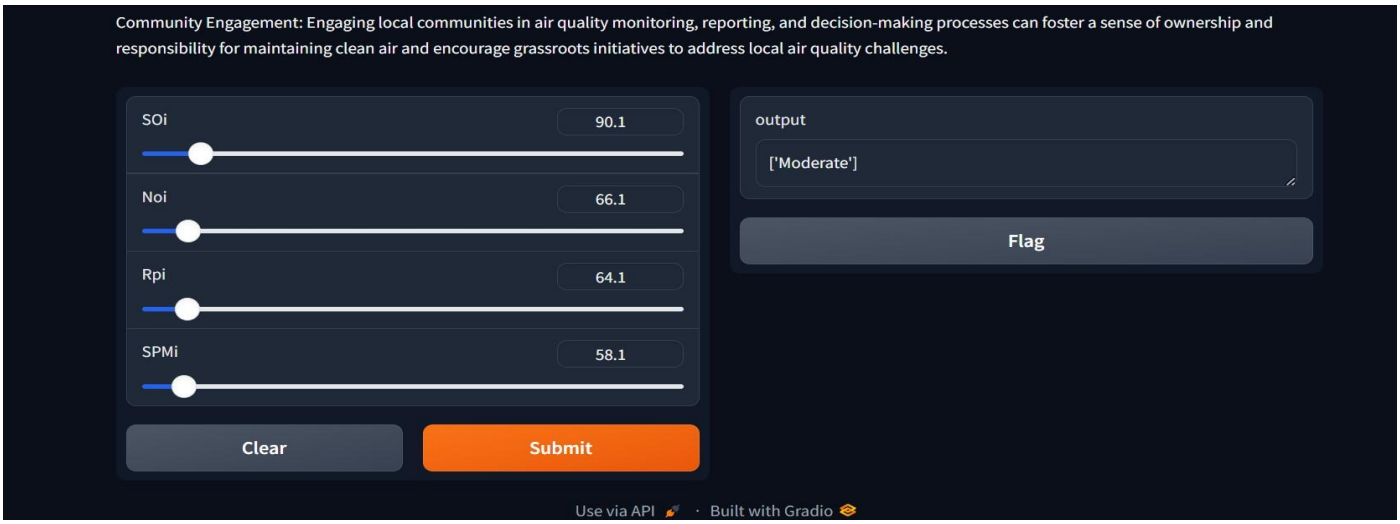

Fig 3 Predicted AQI using KNN (Moderate)

➢ *Predicted AQI using LSTM (Moderate)*

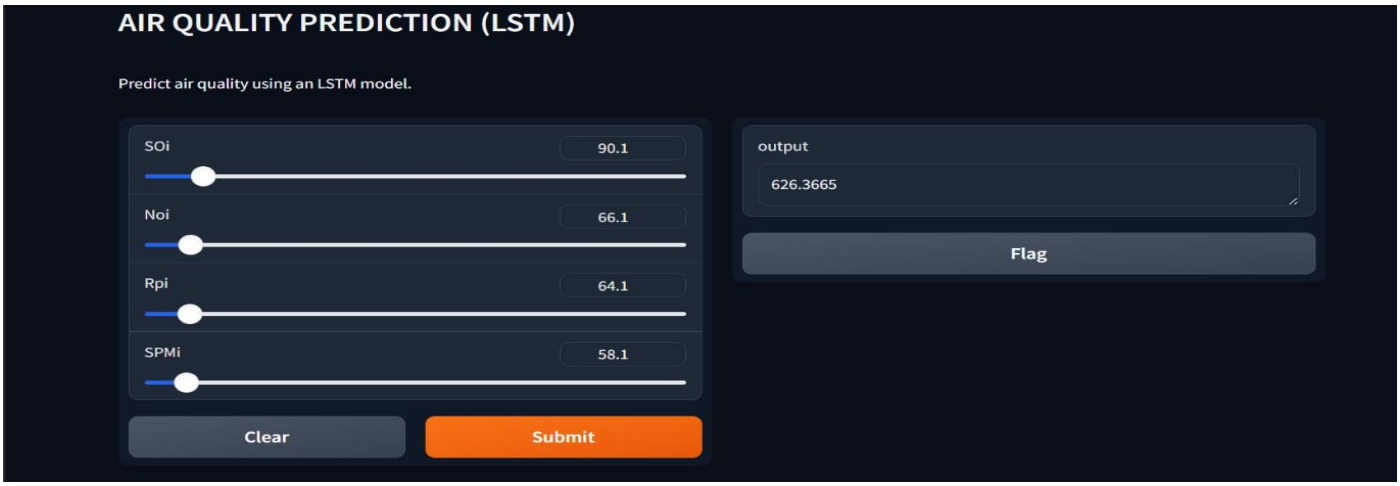

Fig 4 Predicted AQI using LSTM (Moderate)

➢ *Predicted AQI using KNN (Poor)*

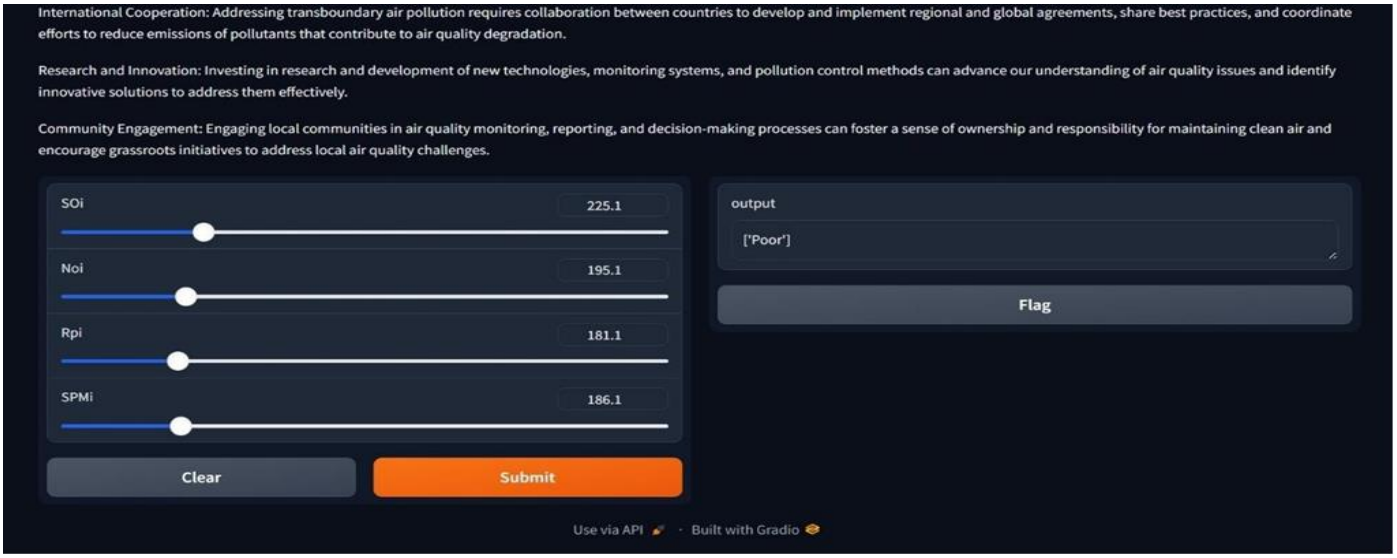

Fig 5 Predicted AQI using KNN (Poor)

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24APR1706

➢ *Predicted AQI using LSTM (Poor)*

Fig 6 Predicted AQI using LSTM (Poor)

➢ *Predicted AQI using KNN (Hazardous)*

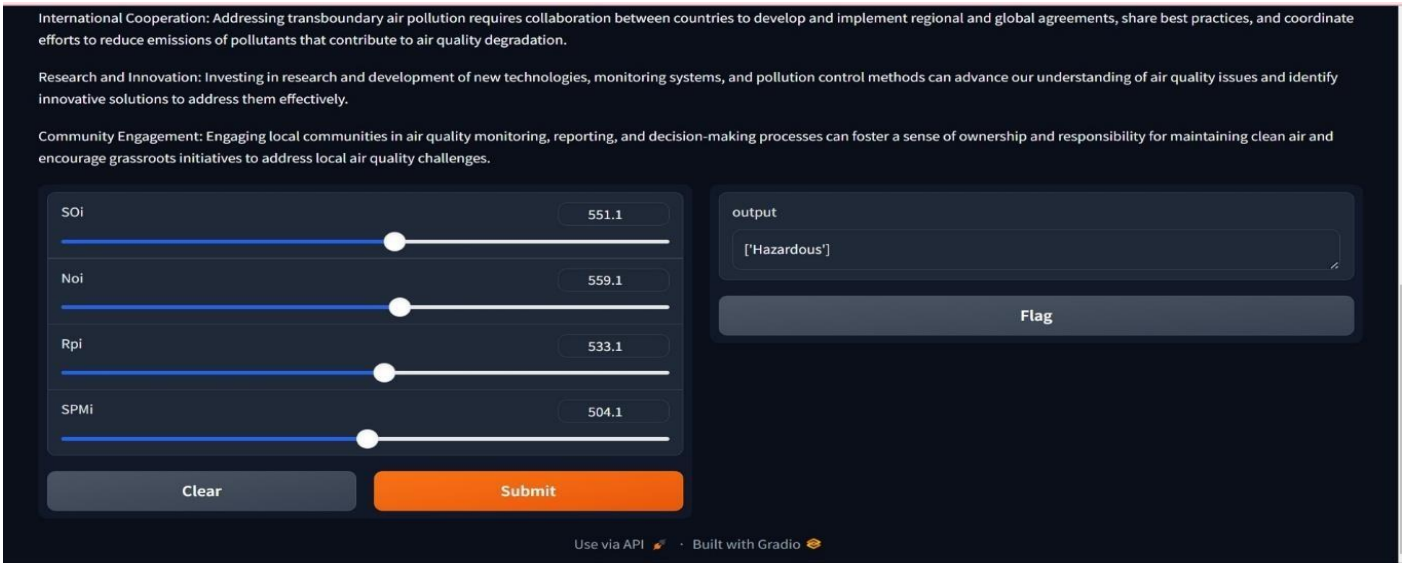

Fig 7 Predicted AQI using KNN (Hazardous)

➢ *Predicted AQI using LSTM (Hazardous)*

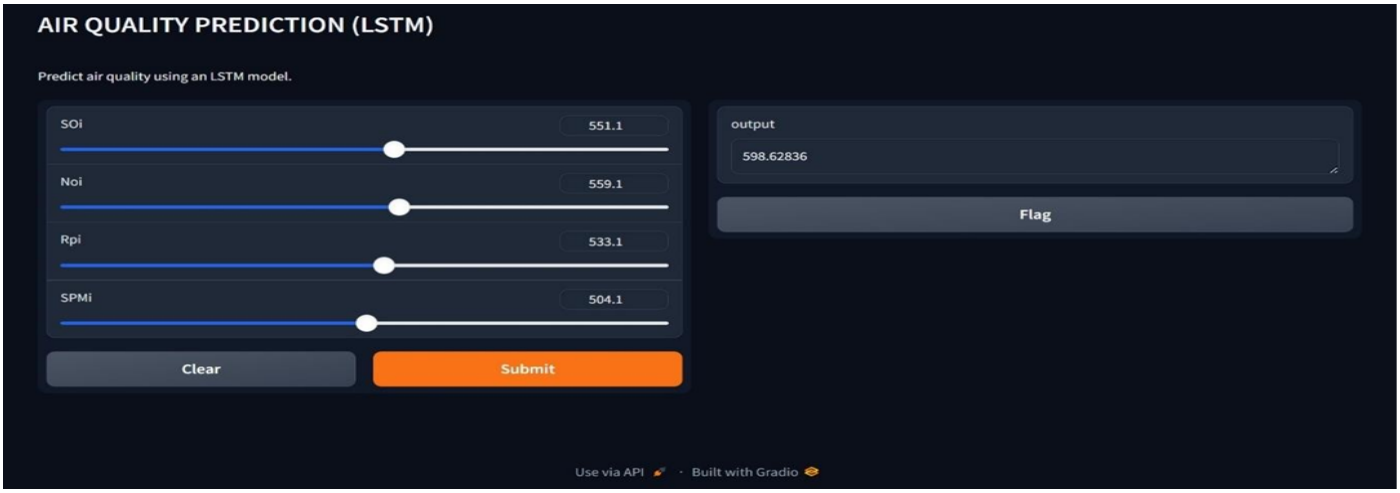

Fig 8 Predicted AQI using LSTM (Hazardous)

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24APR1706

V. CONCLUSION

In conclusion, the "Air Quality Prediction Using KNN and LSTM" project represents a significant step towards leveraging advanced technologies for addressing environmental challenges. Through the integration of machine learning models, data analytics, and real-time monitoring, the project aims to provide accurate and timely predictions of air quality, contributing to public health, environmental awareness, and informed decision-making.

Throughout the project lifecycle, various milestones were achieved, including the development of robust machine learning models trained on historical data, integration with real- time data sources, and the creation of a user-friendly web interface. Thesystem demonstrated its capability to generate air quality predictions, facilitating access to valuable information for users concerned about environmental conditions.

The project's methodology involved a comprehensive approach, encompassing data collection, exploratory data analysis, feature engineering, model development, and the integration of predictive analytics into a user-accessible platform. The iterative and collaborative nature of the development process allowed for continuous improvement andadaptation to evolving environmental factors.

Key findings from the literaturereview highlighted the significance of machine learning in air quality prediction, with studies emphasizing the need for accurate models, consideration of various environmental factors, and the incorporation of real-time data forenhanced predictions. The project aligned with these principles,employing state-of-the-art algorithms and techniques to ensure the reliability and effectiveness of the predictive models.

REFERENCES

[1]. Zhang, K., Zheng, Y., & Zhao, T. (2018). DeepAR: Probabilistic Forecasting with Autoregressive Recurrent Networks. In Proceedings of the International Conference on Machine Learning (ICML).
[2]. Khan, S. M., Park, J. S., & Lee, S. (2019). Air quality prediction in smart cities using deep learning and internet of things. Sensors, 19(2), 463.
[3]. Sharma, T., Chelladurai, J., & Kim, Y. D. (2019). A review of deep learning approaches to air quality predictions. Journal of Ambient Intelligence and Humanized Computing, 10(10), 3801-3818.
[4]. Chen, X., Huang, B., Chen, R., & Chen, Z. (2019). A comprehensive survey of machine learning methods in air quality prediction research. Future Generation Computer Systems, 99, 187-198.
[5]. Zheng, Y., & Song, H. (2017). Air quality prediction with machine learning methods. In Proceedings of the 2017 International Conference on Machine Learning and Cybernetics (ICMLC).
[6]. Zhang, S., Yang, Q., & Shang, Y. (2020). Air Quality Prediction Based on Machine Learning Algorithms. In Proceedings of the 12th International Conference on Advanced Computer Theory and Engineering (ICACTE). Guo, Q., Li, X., Wang, W., & Li, W. (2021). A hybrid approach for air quality prediction using machine learning models. Environmental Monitoring and Assessment, 193(3), 162.
[7]. Chakraborty, S., Choudhury, A., Mukherjee, A., & Dutta, R. (2020). A Comparative Analysis of Machine Learning Algorithms for Air Quality Prediction. International Journal of Environmental Research and Public Health, 17(9), 3135. Link
[8]. Huang, J., Zhang, H., & Wang, Z. (2019). Enhancing Air Quality Prediction Through Ensemble Learning and Deep Neural Networks. IEEE Access, 7, 183877-183887. Link
[9]. Liu, Y., Wang, Y., Zhang, J., & Liu, H.