Parshvanath Charitable Trust's
A. P. SHAH INSTITUTE OF TECHNOLOGY
(Approved by AICTE New Delhi & Govt. of Maharashtra, Affiliated to University of Mumbai)
(Religious Jain Minority)

Department of Computer Science & Engineering (AI & ML)

Air Quality Prediction For Indian Cities Using Machine Learning Techniques

Nishant Hire (21106060)
Shipra Asthana (21106039)
Maviya Bubere (21106022)
Ayush Khargutkar (21106042)

Project Guide
Ms. Vijaya Bharathi J

Outline

- Introduction
- Literature Survey of the existing systems
- Limitations of the existing systems
- Problem statement
- System Design
- Technologies and methodologies
- Implementation
- Conclusion
- References

Introduction

- The Air Quality Index, or AQI, is the system used to predict the air pollution of a region
- The main goal of this project is to develop a predictive model for forecasting the Air Quality Index
- This project can analyze real-time environmental data prediction
- This model incorporates key pollutants and meteorological factors, providing an accurate tool for proactive air quality monitoring and intervention

Objectives

- To provide accurate prediction in forecasting air quality levels
- To identify patterns and trends in air pollution data
- To raise engagement in air quality issues
- To provide real-time graphical representation of temperature, wind speed and humidity for each state

Literature Survey of the existing system

Sr. No	Title	Author Name	Description
1.	Prediction of Air Quality Index Based on LSTM.	Yu Jiao, Zhifeng Wang and Yang Zhang	This paper proposes a prediction model of environmental quality based on Long Short Term Memory (LSTM). This paper uses data provided by the environmental protection department to predict Air Quality Index (AQI) through temperature, PM2.5, PM10, SO_2, wind direction, NO_2, CO and O_3

Sr. No	Title	Author Name	Description
2.	Prediction of Air Quality Index Based on Improved Neural Network	Wang Zhenghua; Tian Zhihui	The model uses the characteristics of nonlinear fitting approximation of BP neural network to solve the problem that air quality has many influencing factors and is nonlinear and difficult to predict. Aiming at the problem of slow convergence and easy to fall into local optimal solution of BP neural network, genetic algorithm is used to optimize

Sr. No	Title	Author Name	Description
3.	An Adaptive Kalman Filtering Approach to Sensing and Predicting Air Quality Index Values	Chen Ding; Guizhi Wang; Qi Liu	In this paper, an Auto-Regressive (AR) prediction model based on sensed AQI values is proposed, where an adaptive Kalman Filtering (KF) approach is fitted to achieve efficient prediction of the AQI values

Sr. No	Title	Author Name	Description
4.	Air Quality Index Forecasting via Genetic Algorithm	Chunhao Liu; Guangyuan Pan	By analyzing SO2 , NO2 , PM10 , CO , O3 , PM2.5 concentration and AQI with comparative experiments based CMAQ (Community Multiscale Air Quality), SVM (Support Vector Machines) and DBN-BP (Deep Belief Networks with Back-Propagation). The results show that the proposed model trains faster and makes more accurate prediction

Sr. No	Title	Author Name	Description
5.	Prediction of Air Quality Index Using Machine Learning Techniques	N. Srinivasa Gupta, Yashvi Mohta, Khyati Heda.	Data mining techniques are one of the most interesting approaches to forecast AQI and analyze it. The aim of this paper is to find the most effective way for AQI prediction to assist in climate control. The most effective method can be improved upon to find the most optimal solution. Hence, the work in this paper involves intensive research and the addition of novel techniques such as SMOTE to make sure that the best possible solution to the air quality problem is obtained

Limitations of Existing Systems

From the literature review of existing systems, we find that,

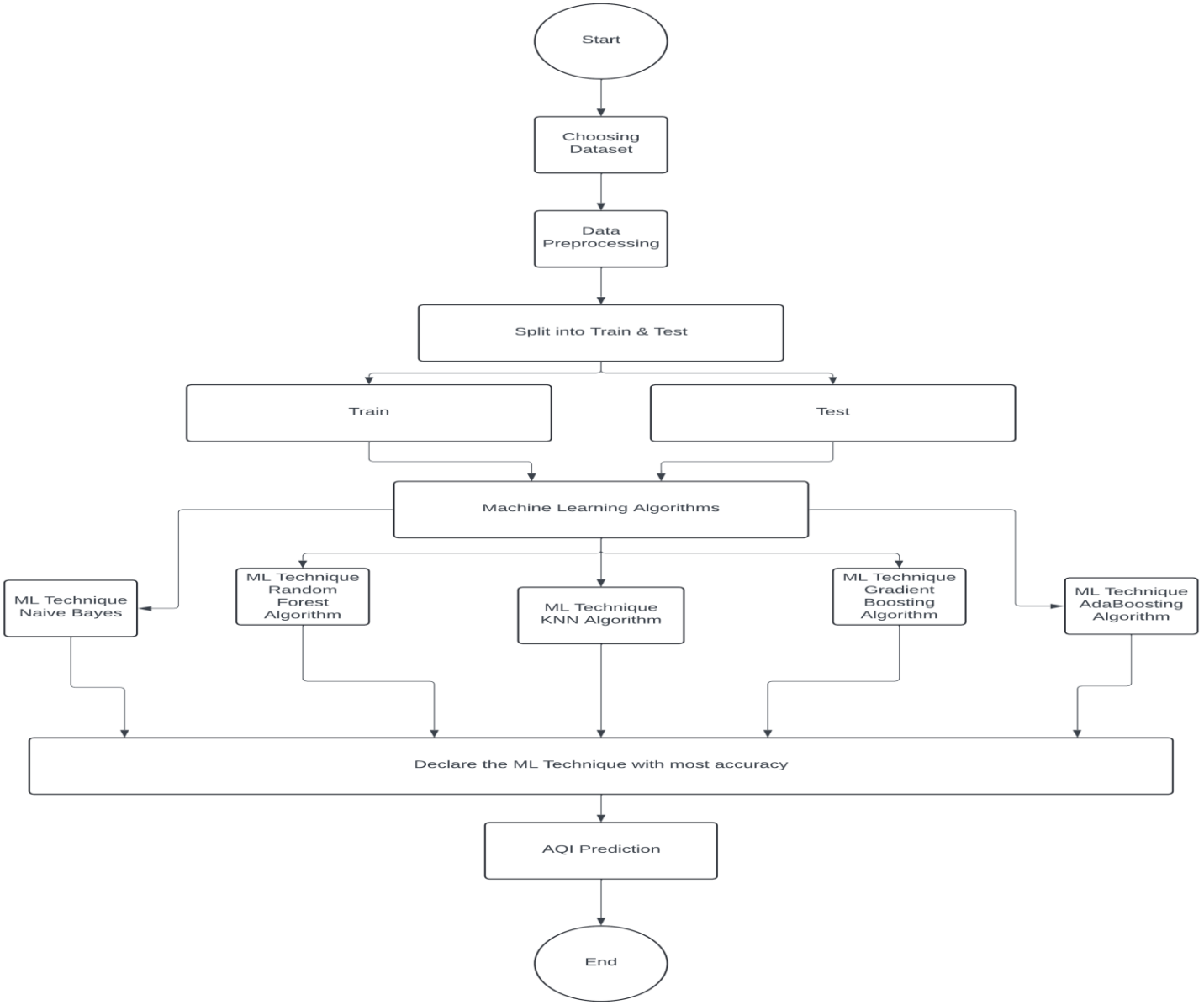
- Generalizability: SMOTE-balanced dataset's effectiveness may vary outside major Indian cities due to diverse pollution and population dynamics
- Health Focus: Existing systems often lack specificity in addressing health outcomes, hindering tailored preventive measures against air pollution
- Implementation Challenges: Urgent pollution control measures face hurdles like resource availability and community engagement, especially in residential areas like NEERI, Nagpur

Problem statement

- Air pollution poses a significant threat to public health and the environment, requiring accurate and timely monitoring and prediction

- Leveraging machine learning techniques, such as regression and classification algorithms, offers the potential to improve the accuracy and reliability of air quality forecasts

System Design



- **Source Data Identification**: Obtain relevant data including historical air quality measurements and meteorological data
- **Data Preprocessing**: Clean and preprocess the data to address missing values, outliers, and inconsistencies
- **Model Training and Testing**: Train various ML algorithms like linear regression and KNN on the preprocessed data
- **Model Selection**: Evaluate and compare the performance of the trained models to select the best-performing one
- **Model Deployment**: Deploy the selected model for AQI forecasting in a production environment
- **Monitoring and Evaluation**: Continuously monitor and evaluate the model's performance for accuracy and reliability

Technologies and methodologies

Technology Stack:

Editor :

- PyCharm
- Visual Studio Code
- Google Collab

Libraries :

1) Python :

- NumPy
- Pandas

2) Machine Learning :

- Scikit-learn
- TensorFlow or PyTorch

Methodology :

- Data Collection
- Data Preprocessing
- Model Selection
- Training & Testing
- Performance Evaluation
- Model Optimization
- Implementation
- Documentation

Implementation

Importing Datset

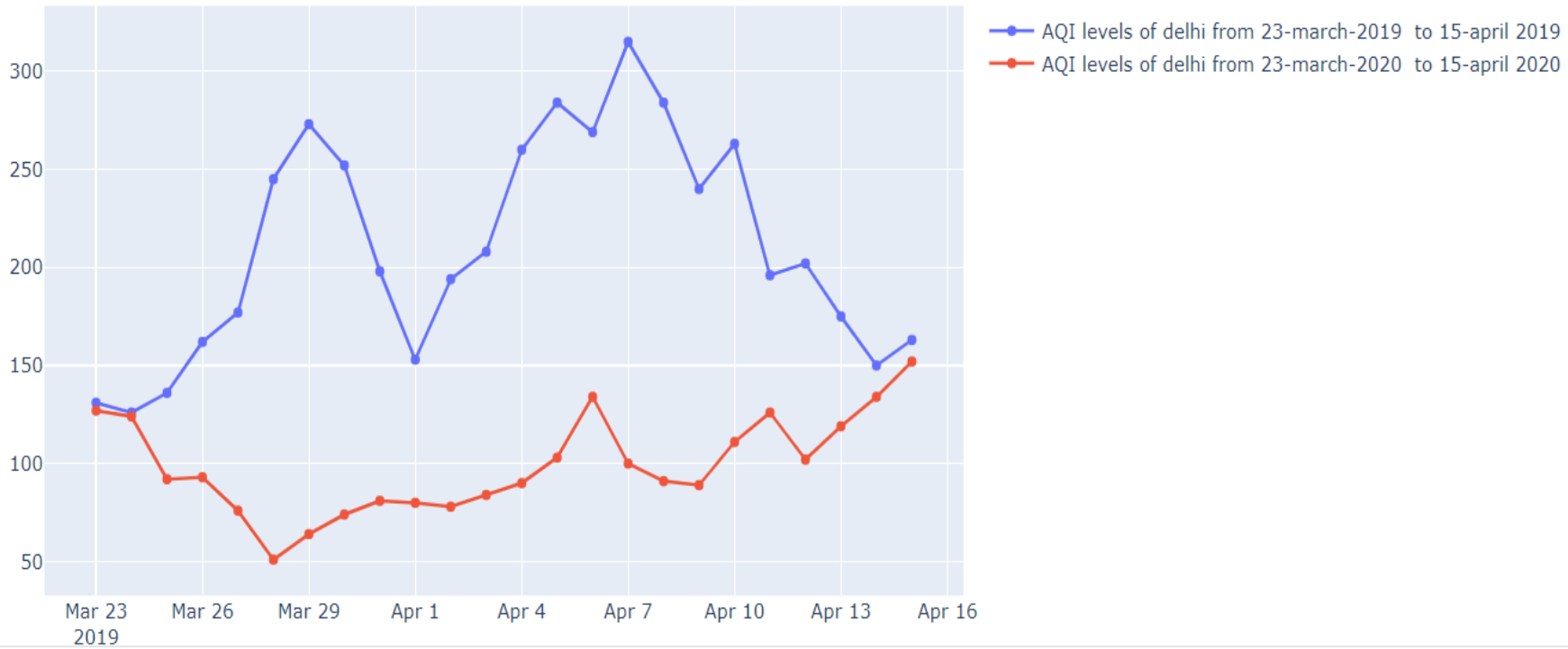
```
[ ]  url='/content/sample_data/city_day.csv'
     city_day_data=pd.read_csv(url)

     # Extract delhi's data

     delhi_data=city_day_data.groupby('City').get_group('Delhi')
```

```
[ ]  delhi_data.head()
```

	City	Date	PM2.5	PM10	NO	NO2	NOx	NH3	CO	SO2	O3	Benzene	Toluene	Xylene	AQI	AQI_Bucket
10229	Delhi	2015-01-01	313.22	607.98	69.16	36.39	110.59	33.85	15.20	9.25	41.68	14.36	24.86	9.84	472.0	Severe
10230	Delhi	2015-01-02	186.18	269.55	62.09	32.87	88.14	31.83	9.54	6.65	29.97	10.55	20.09	4.29	454.0	Severe
10231	Delhi	2015-01-03	87.18	131.90	25.73	30.31	47.95	69.55	10.61	2.65	19.71	3.91	10.23	1.99	143.0	Moderate
10232	Delhi	2015-01-04	151.84	241.84	25.01	36.91	48.62	130.36	11.54	4.63	25.36	4.26	9.71	3.34	319.0	Very Poor
10233	Delhi	2015-01-05	146.60	219.13	14.01	34.92	38.25	122.88	9.20	3.33	23.20	2.80	6.21	2.96	325.0	Very Poor

Dataset

```
fig.add_trace(go.Scatter(x=march24_2019['Date'], y=march24_2020['AQI'],
                    mode='lines+markers',
                    name='AQI levels of delhi from 23-march-2020  to 15-april 2020'))
fig.show()
```



Effect on AQI levels of Delhi due to lockdown

```python
#fit the model on train data
KNN = KNeighborsClassifier().fit(X_train2, Y_train2)

#predict on train
train_preds5 = KNN.predict(X_train2)
#accuracy on train
print("Model accuracy on train is: ", accuracy_score(Y_train2, train_preds5))

#predict on test
test_preds5 = KNN.predict(X_test2)
#accuracy on train
print("Model accuracy on test is: ", accuracy_score(Y_test2, test_preds5))
print('-'*50)

#Kappa Score
print('KappaScore is: ', metrics.cohen_kappa_score(Y_test2,test_preds5))
```

```
Model accuracy on train is:  0.9983079004607032
Model accuracy on test is:  0.9968218423578175
--------------------------------------------------
KappaScore is:  0.9952893818649885
```

Conclusion

• In conclusion, the AQI forecast project aims to provide accurate and reliable predictions of Air Quality Index (AQI) levels by leveraging machine learning algorithms
• By preprocessing and analyzing relevant data, training various models, and selecting the best-performing one, the project seeks to address challenges in air quality prediction
• Through continuous monitoring and evaluation, the deployed model will contribute to informed decision-making for public health and environmental management, ultimately improving air quality and safeguarding community well-being

References

Journal Papers

[1] Yu Jiao, Zhifeng Wang and Yang Zhang "Prediction of Air Quality Index Based on LSTM" 2019 IEEE 8th Joint International Information Technology and Artificial Intelligence Conference (ITAIC)

[2] Wang Zhenghua; Tian Zhihui "Prediction of Air Quality Index Based on Improved Neural Network" 2017 International Conference on Trends in Electronics and Informatics (ICEI)

[3] Chen Ding; Guizhi Wang; Qi Liu "An Adaptive Kalman Filtering Approach to Sensing and Predicting Air Quality Index Values" IEEE Access (Volume: 8)

[4] Chunhao Liu; Guangyuan Pan, "Air Quality Index Forecasting via Genetic Algorithm" IEEE Access (Volume: 11)

[5] N. Srinivasa Gupta, Yashvi Mohta, Khyati Heda, "Prediction of Air Quality Index Using Machine Learning Techniques" Journal of Environmental and Public Health, vol. 2023

Useful Links:

https://www.coursera.org/specializations/machine-learning-introduction

Thank You...!!