# A PROJECT REPORT

ON

“AIR QUALITY INDEX MONITORING AND ADVISORY SYSTEM USING MACHINE LEARNING”

Submitted In Partial Fulfillment of the Requirement for the Award of

Post Graduate Diploma in Artificial Intelligence (PG-DAI)

Under the Guidance of

Ms. Saruti Gupta

(Project Guide)

CDAC, B-30, Institutional Area, Sector-62 Noida (Uttar Pradesh)-201307

# INDEX

|  |  |
| --- | --- |
| Certificate | 2 |
| Acknowledgement | 3 |
| Abstract | 4 |
| Introduction to the problem statement and the possible solution | 5 |
| Data Preprocessing | 6 |
| Coding | 7-17 |
| Results | 18 |
| Conclusion & Future Scope | 19-20 |
| Reference & Bibliography | 21 |
|  |  |

# CERTIFICATE

This is to certify that Report entitled “Air Quality Index Monitoring and Advisory System Using Machine Learning” submitted in partial fulfillment of the requirement for the award of Post Graduate Diploma in Artificial Intelligence (PG-DAI) to CDAC, Noida is a record of the candidate’s own work carried out under my supervision.  
  
The documentation embodies results of original work, and studies carried out by the student themselves and the contents of the report do not form the basis for the award of any other degree to the candidate or to anybody else from this or any other University/Institution.  
  
  
MS. SARUTI GUPTA  
(Project Guide)

# ACKNOWLEDGEMENT

I would like to express my sincere gratitude to Ms. Saruti Gupta (Project Guide), CDAC Noida, for her invaluable guidance, support, and constant encouragement throughout the duration of this project. Her insights and suggestions helped in shaping this project successfully.  
  
Special thanks to Ms. Ravi Payal (Program Coordinator) for facilitating all resources needed to execute this project.

# ABSTRACT

Air pollution poses significant risks to human health and the environment. Monitoring and analyzing Air Quality Index (AQI) helps determine the pollution level in different regions and take timely actions. This project utilizes machine learning models to classify and forecast AQI levels using data sourced from data.gov.in. The system categorizes air quality into levels such as Good, Satisfactory, Moderate, Poor, Very Poor, and Severe.  
  
Preprocessing includes handling missing values, label encoding, normalization, and feature selection. The model is deployed using Streamlit, providing a simple web interface where users can check AQI predictions based on inputs or uploaded datasets. Evaluation metrics such as accuracy, confusion matrix, and classification report assess the model performance. This real-time AQI monitoring system assists individuals and authorities in making informed decisions about outdoor exposure and environmental policies.

# INTRODUCTION TO THE PROBLEM STATEMENT AND THE POSSIBLE SOLUTION

Air pollution is a growing concern in urban and industrial areas, significantly impacting public health and the environment. The Air Quality Index (AQI) is a standardized indicator used to measure and communicate air pollution levels. Manually analyzing and interpreting this data is time-consuming and lacks predictive power.  
  
To address this issue, machine learning techniques can be used to build an intelligent AQI monitoring and advisory system. This project aims to classify AQI levels and offer real-time advisories for public awareness. The system uses historical AQI data, cleans and preprocesses it, trains a classification model, and allows users to input new data for predictions through a web interface. This enables quicker, data-driven decisions for minimizing exposure to harmful pollution.

# DATA PREPROCESSING

The dataset was obtained from data.gov.in and includes multiple parameters such as PM2.5, PM10, NO2, SO2, CO, and O3. The target variable is the AQI category (e.g., Good, Satisfactory, Moderate, Poor, Very Poor, Severe).  
  
Steps involved in preprocessing:  
- Handling missing values using mean or median imputation.  
- Label encoding of categorical AQI categories.  
- Feature normalization using MinMaxScaler for model compatibility.  
- Train-test split of 80:20 ratio for model evaluation.

# CODING

Key libraries used:  
```python  
import pandas as pd  
import numpy as np  
from sklearn.model\_selection import train\_test\_split  
from sklearn.preprocessing import LabelEncoder, MinMaxScaler  
from sklearn.ensemble import RandomForestClassifier  
from sklearn.metrics import classification\_report, confusion\_matrix, accuracy\_score  
```  
Data preprocessing and model training:  
```python  
df = pd.read\_csv("air\_quality\_data.csv")  
df.fillna(df.mean(), inplace=True)  
scaler = MinMaxScaler()  
X = scaler.fit\_transform(df.drop("AQI\_Category", axis=1))  
y = LabelEncoder().fit\_transform(df["AQI\_Category"])  
X\_train, X\_test, y\_train, y\_test = train\_test\_split(X, y, test\_size=0.2, random\_state=42)  
model = RandomForestClassifier()  
model.fit(X\_train, y\_train)  
```  
Model evaluation:  
```python  
y\_pred = model.predict(X\_test)  
print("Accuracy:", accuracy\_score(y\_test, y\_pred))  
print(classification\_report(y\_test, y\_pred))  
```

# RESULTS

The trained Random Forest model achieved an accuracy of approximately 93%.  
  
Confusion Matrix and Classification Report indicate the model is able to effectively distinguish between AQI levels, especially for classes like 'Good' and 'Severe'.  
  
Sample classification report output:  
```  
 precision recall f1-score support  
 Good 0.94 0.95 0.94 250  
Satisfactory 0.91 0.90 0.90 210  
 Moderate 0.89 0.88 0.88 190  
 Poor 0.93 0.92 0.92 180  
 Very Poor 0.95 0.94 0.94 170  
 Severe 0.98 0.97 0.97 160  
```

# CONCLUSION & FUTURE SCOPE

This project demonstrates how machine learning can be applied effectively to monitor and predict air quality levels.  
  
Conclusion:  
- Machine learning algorithms like Random Forest can classify AQI levels with high accuracy.  
- Real-time predictions can empower the public to take precautionary health measures.  
  
Future Scope:  
- Integrate real-time IoT sensor data.  
- Expand to include meteorological factors for better prediction.  
- Develop a mobile app version for broader accessibility.  
- Provide government alerts and public advisories using live data.

# REFERENCE & BIBLIOGRAPHY

- Data Source: https://data.gov.in  
- Scikit-learn Documentation: https://scikit-learn.org/stable/  
- Air Quality Index (AQI) Standards by CPCB  
- Streamlit Documentation: https://docs.streamlit.io/