Air Quality Prediction using Machine Learning

This presentation explores the use of machine learning to predict air quality in India.

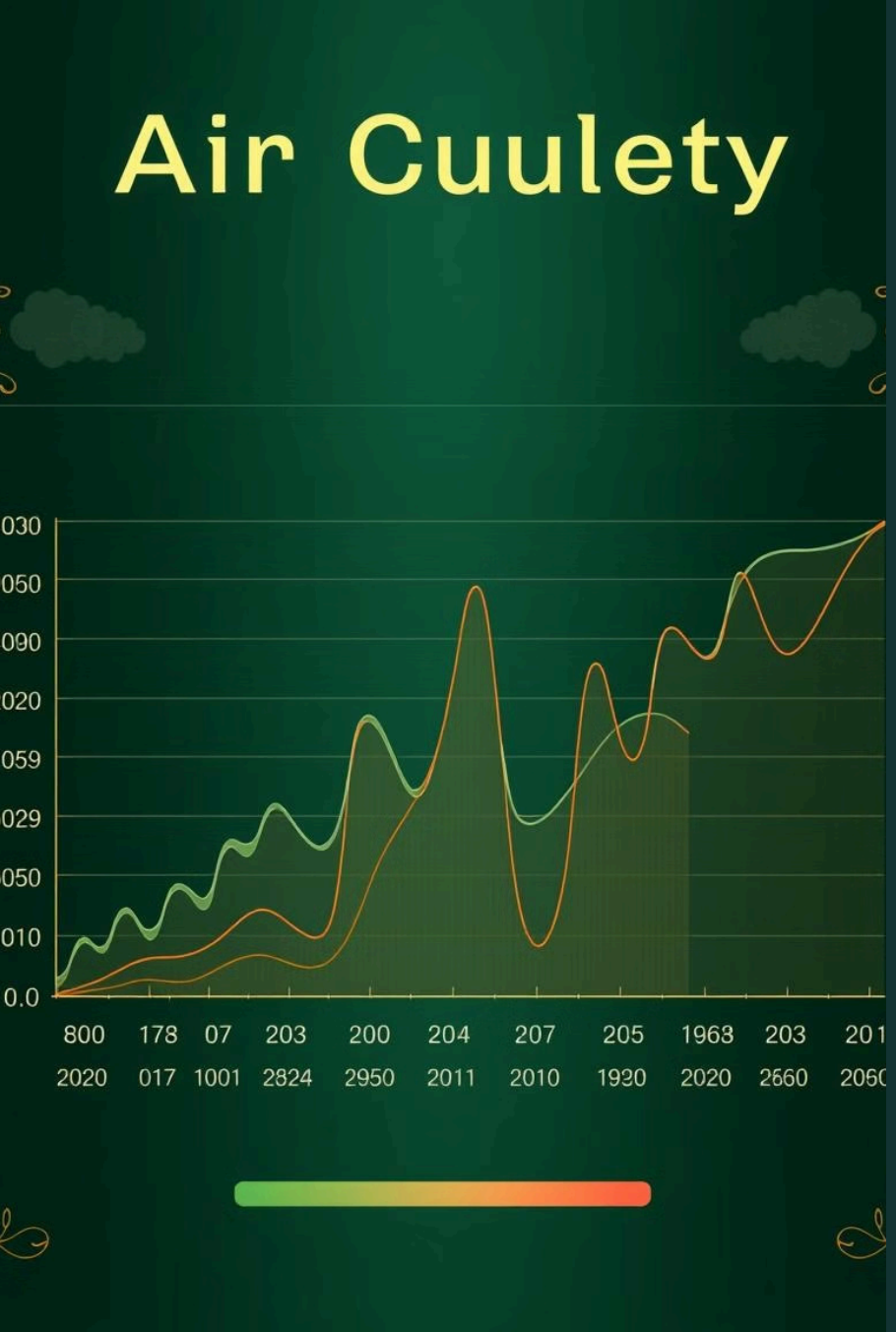
Data Preparation

Import Libraries

Import necessary libraries for data manipulation, visualization, and machine learning.

Load Dataset

Read the air quality dataset from a CSV file.



Data Understanding

1 **Explore Data**

View the first few rows of the dataset to understand its structure.

2 **Check Dimensions**

Determine the number of rows and columns in the dataset.

3 **Data Types**

Identify the data types of each column.

4 **Missing Values**

Identify and count missing values in each column.

Data Visualization

Pairplot

Visualize relationships between all pairs of numerical variables.

State Distribution

Analyze the distribution of air quality data across different states.

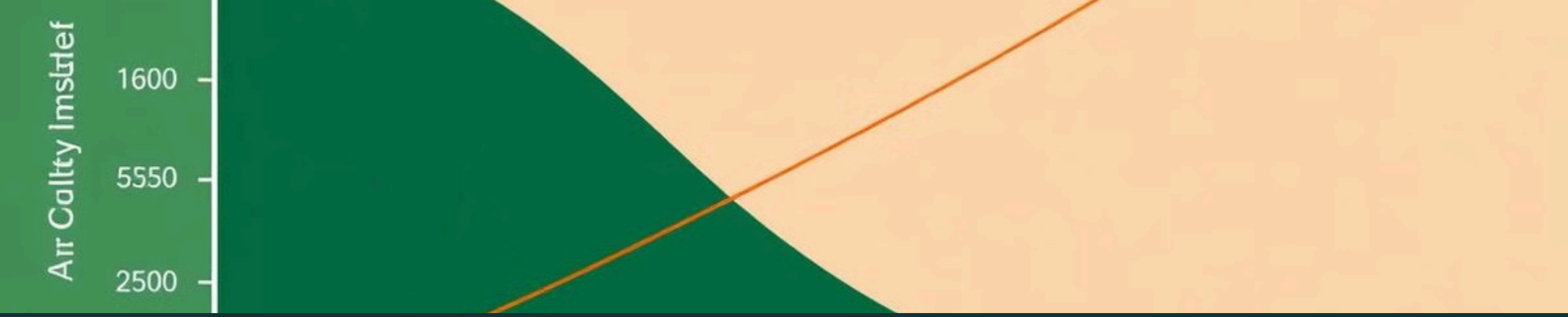
Data Cleaning

Drop Columns

Remove unnecessary columns from the dataset.

Impute Missing Values

Fill in missing values using appropriate methods.



Calculate Air Quality Index (AQI)

SO2

Calculate individual pollutant index for SO2.

NO2

Calculate individual pollutant index for NO2.

RSPM

Calculate individual pollutant index for RSPM.

SPM

Calculate individual pollutant index for SPM.

Regression Models

Linear Regression

Fit a linear regression model to predict AQI.

Decision Tree

Fit a decision tree regression model to predict AQI.

Random Forest

Fit a random forest regression model to predict AQI.

KNN

Fit a K-Nearest Neighbors regression model to predict AQI.

XGBoost

Fit an XGBoost regression model to predict AQI.

Classification Models

1

Logistic Regression

Fit a logistic regression model to classify AQI ranges.

2

Decision Tree

Fit a decision tree classification model to classify AQI ranges.

3

Random Forest

Fit a random forest classification model to classify AQI ranges.

4

KNN

Fit a K-Nearest Neighbors classification model to classify AQI ranges.

5

SVM

Fit a Support Vector Machine classification model to classify AQI ranges.

Conclusion

1

Random Forest

Best model for both regression and classification.

2

High Accuracy

Achieves excellent prediction accuracy.