**AIR QUALITY ANALYSIS AND PREDICTION IN TAMIL NADU**

**PROJNUMBER:PROJ\_220061\_TEAM\_1**

**PHASE 2 SUBMISSION DOCUMENT**

**Project:Air quality analysis and prediction in Tamil nadu**

****

Project Introduction:

* Air pollution is a critical environmental concern that affects the health and well-being of millions of people worldwide. Tamil Nadu, a state in India, is no exception to this issue.
* With rapid urbanization and industrialization, understanding and forecasting air quality in Tamil Nadu has become imperative for public health, urban planning, and environmental management.
* The "Air Quality Analysis and Prediction in Tamil Nadu" project aims to analyze historical air quality data in the region, develop predictive models, and provide insights and forecasts for air quality conditions.
* This project seeks to empower local authorities, researchers, and citizens with the information necessary to make informed decisions and take preventive measures to mitigate the adverse effects of air pollution.

Data source:

* Government Agencies: Many governments around the world monitor and report air quality data. In the United States, for example, the Environmental Protection Agency (EPA) provides air quality data through its Air Quality Index (AQI) program.
* Research Institutions: Universities and research institutions often conduct air quality studies and make their data available to the public.
* Satellite Data: Satellites equipped with remote sensing instruments can provide information on air quality at a regional or global scale, including the concentration of certain pollutants and aerosols in the atmosphere.
* Environmental Monitoring Stations: Air quality data is often collected by specialized monitoring stations equipped with instruments that measure pollutants such as particulate matter (PM2.5 and PM10), ozone (O3), nitrogen dioxide (NO2), sulfur dioxide (SO2), and carbon monoxide (CO)

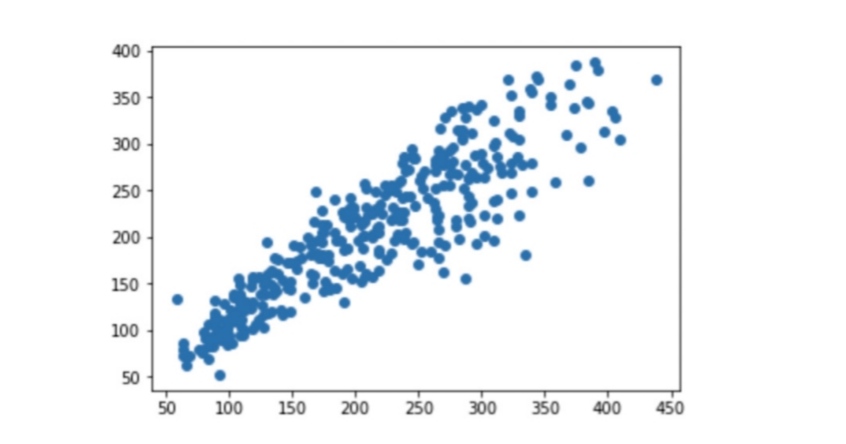
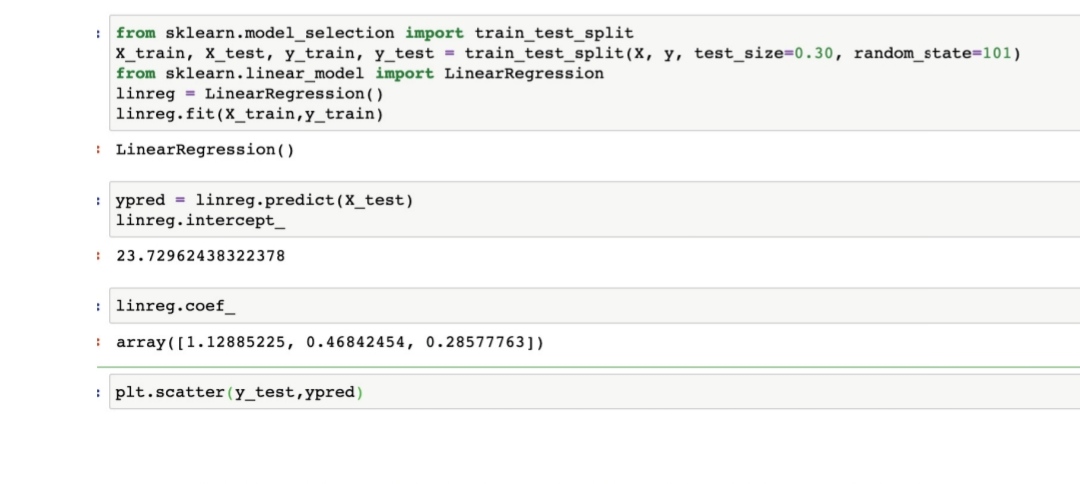
**Dataset Link**: https://tn.data.gov.in/resource/location-wise-daily-ambient-air-quality-tamil-nadu-year-2014

Project Objectives:

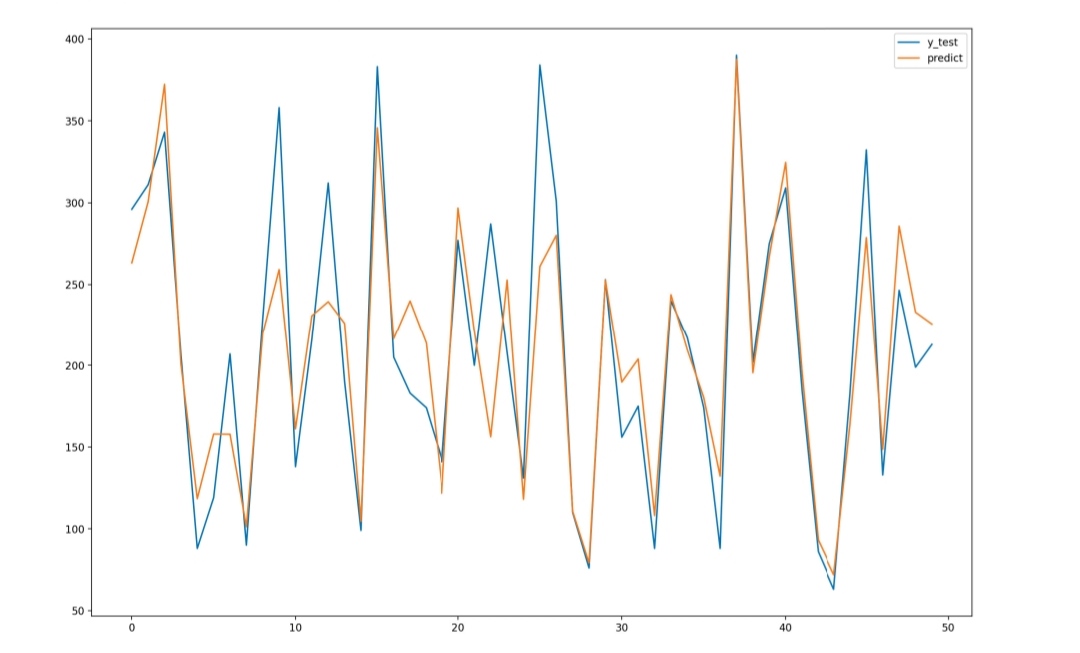
* Data Collection: Gather historical air quality data from multiple monitoring stations across Tamil Nadu. This data will include parameters such as PM2.5, PM10, NO2, SO2, CO, O3, temperature, humidity, and wind speed.
* Data Preprocessing: Clean and preprocess the collected data, handling missing values, outliers, and formatting it for analysis.
* Exploratory Data Analysis (EDA): Conduct EDA to understand temporal and spatial patterns in air quality, identify trends, and visualize pollution levels in different regions of Tamil Nadu.
* Feature Engineering: Create relevant features such as seasonality, meteorological data, and geographical factors to improve prediction accuracy.
* Model Selection: Choose appropriate machine learning and time series forecasting models (e.g., ARIMA, LSTM, Random Forest) for air quality prediction.
* Model Training: Train selected models on historical air quality data, optimizing model hyperparameters.
* Prediction: Generate real-time or short-term air quality forecasts for different locations in Tamil Nadu.
* Model Evaluation: Assess the accuracy of the predictive models using metrics like Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and R-squared.
* Visualization: Develop interactive dashboards and visualizations to communicate air quality information effectively to stakeholders and the public.
* Deployment: If applicable, deploy the prediction model as a web application or mobile app for widespread accessibility.
* Public Awareness: Raise awareness about air quality issues and educate citizens about the importance of air quality management and pollution reduction.
* Policy Recommendations: Provide insights and recommendations to policymakers for implementing measures to improve air quality in Tamil Nadu.

**Linear regression**

Prediction of air quality index using linear regression

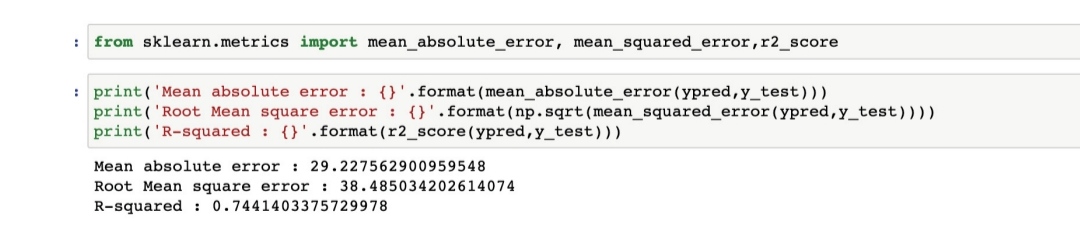
****

Scatter plot graph of linear regression model

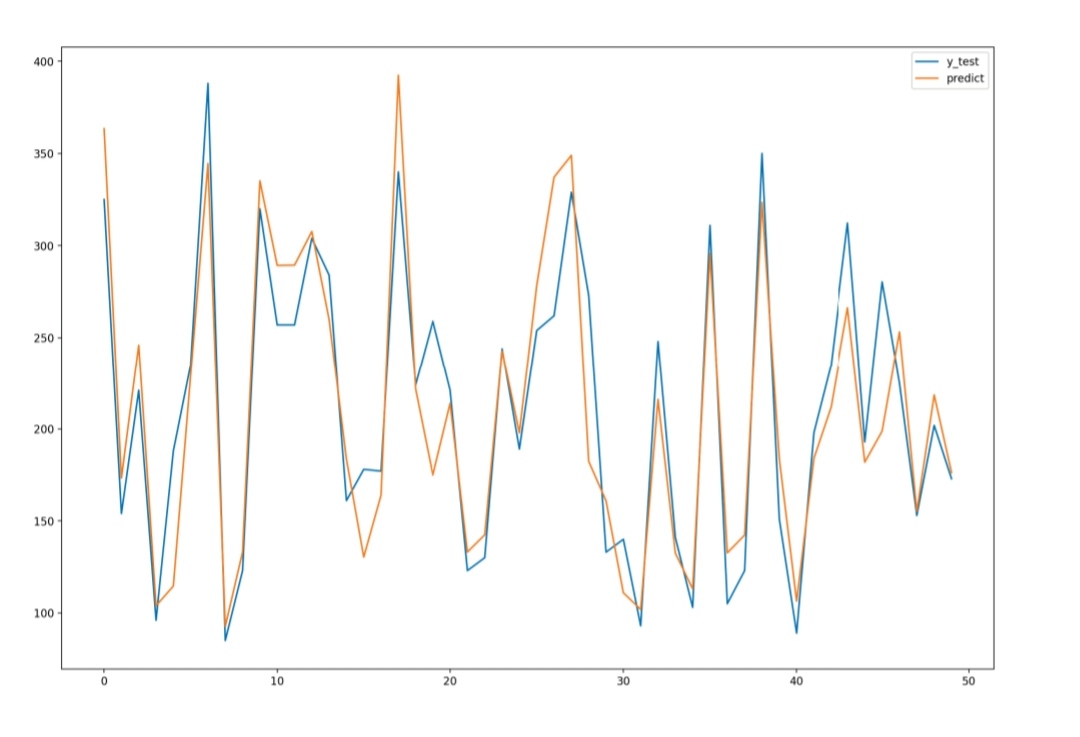


Predicted vs observed (linear regression)

Performance metrics (Linear regression)

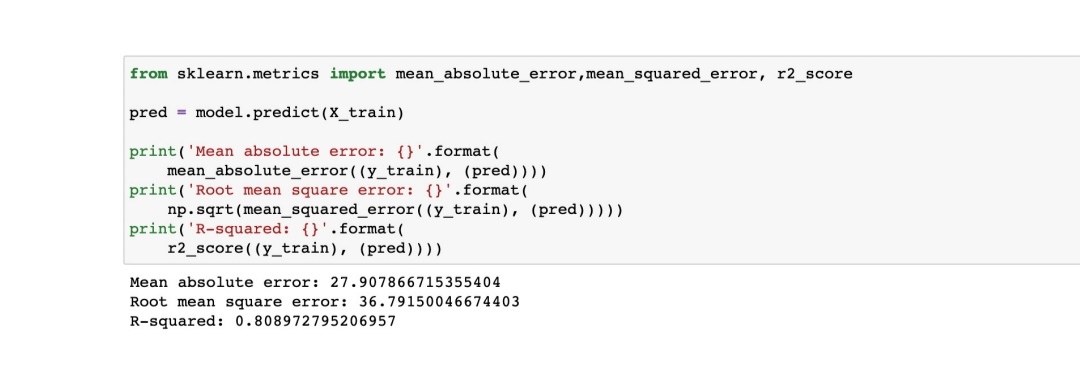


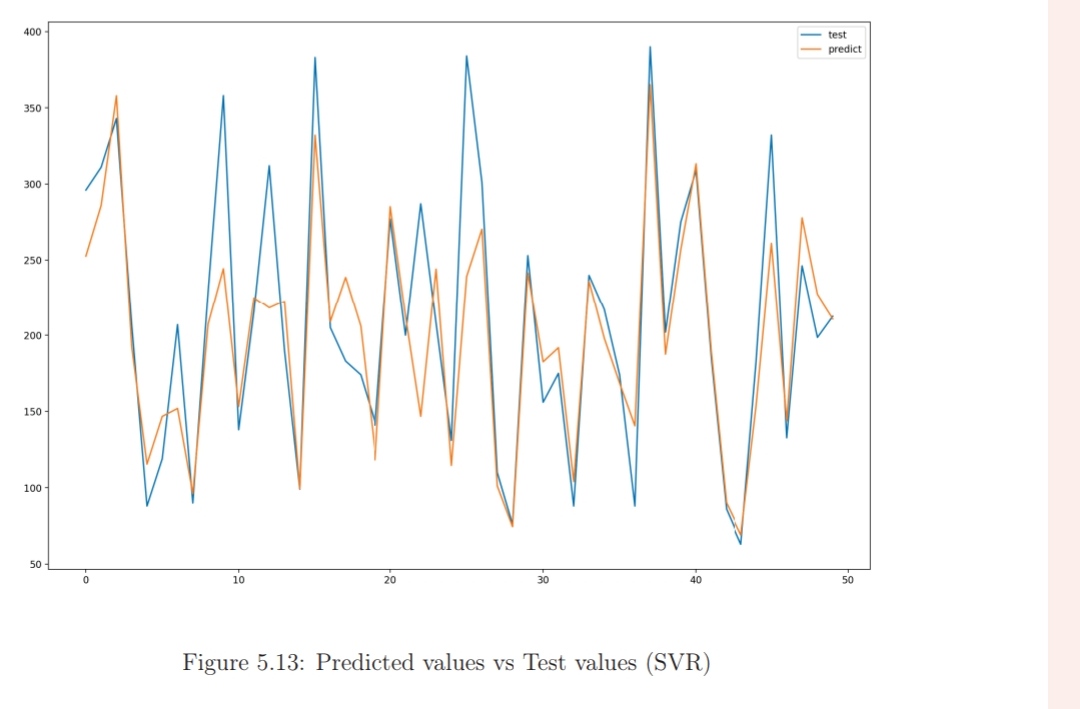
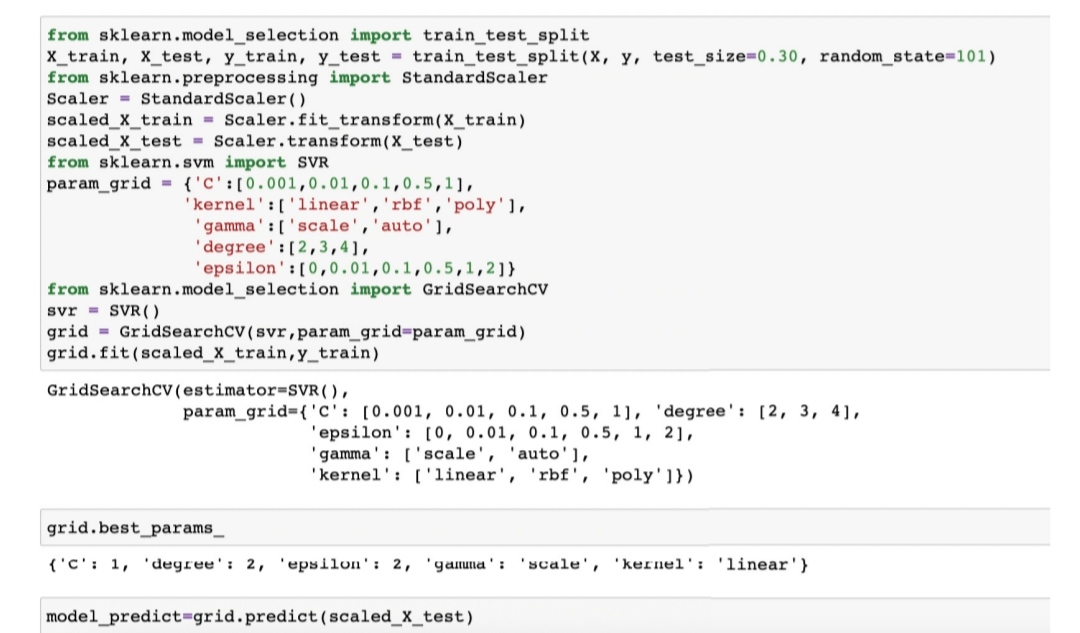
Performance metrics(lasso regression)



Predicted vs observed(lasso regression)

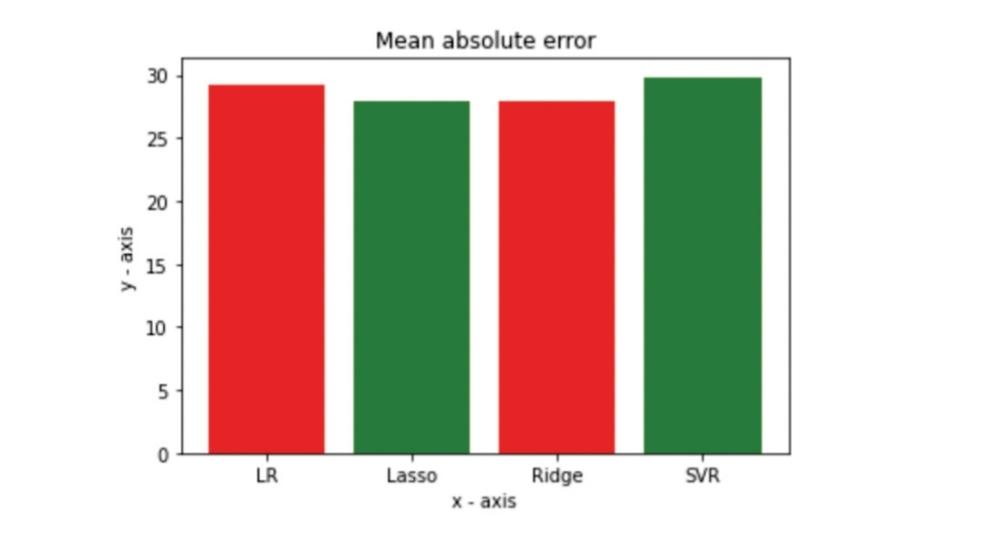
**Lasso regression**



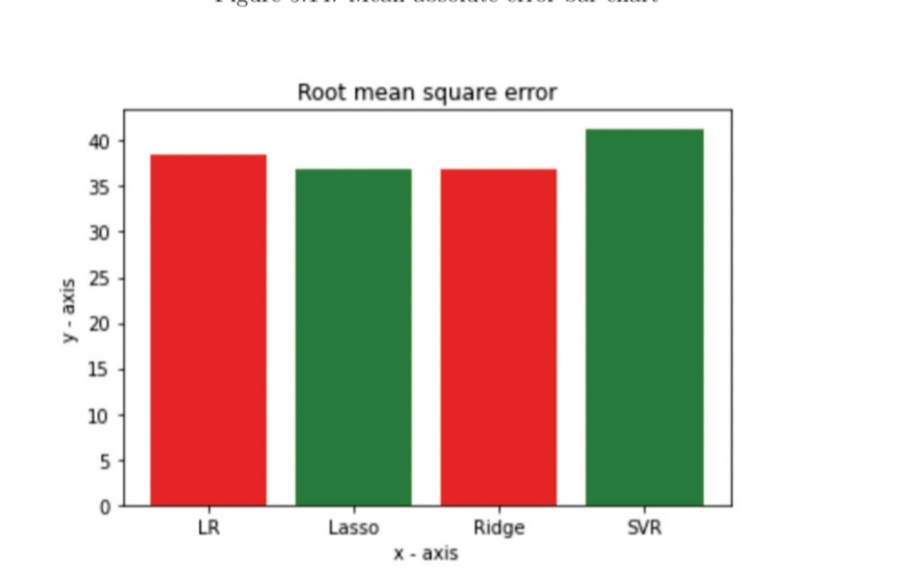


Support vector regression:

Predicted value vs test value (SVR)



Root mean sqaure error bar chart



Mean absolute error bar chart

**Comparsion of result**

Project Conclusion:

* Summary of Findings: Recap the main results and insights obtained from your air quality analysis and prediction efforts.
* Contributions: Highlight the project's contributions to the field of air quality research, such as improved prediction accuracy, new insights into pollution sources, or the development of user-friendly tools.
* Impact: Discuss the potential impact of your work on public health, environmental policies, or other relevant areas.
* Challenges and Limitations: Acknowledge any challenges faced during the project and the limitations of your analysis. This could include data availability issues, model limitations, or uncertainties in predictions.
* Future Work:
* Improved Models: Identify areas where predictive models could be enhanced. Consider exploring more advanced machine learning techniques or incorporating additional data sources, such as weather patterns or traffic data, to improve accuracy.
* Real-time Monitoring: Explore the feasibility of real-time air quality monitoring and prediction systems, which could provide timely information to the public and policymakers.
* Data Quality: Work on improving data quality and consistency, which is crucial for accurate predictions. Collaborate with data providers to ensure data integrity.
* Environmental Interventions: Investigate potential interventions or policies that can help mitigate air quality issues based on your findings. This might involve urban planning, emissions reductions, or public awareness campaigns.