INFO 7390 ADVANCED DATA SCIENCE FINAL PROJECT

Air Quality Index Prediction

Team Members:

1. Aishwarya Mishra (002127206)
2. Sri Sai Amulya Nittala (002194776)
3. Zeenia Singla (002980842)

Background:

With the quick advancement of urbanization and industrialization, air pollution has become a serious issue now more than ever before. It has become one of the primary reasons for global warming and climate change. There is a dire need for reducing pollution caused by various factors such as transportation vehicles, industries, household combustion devices etc. This has posed a serious threat to the lives of all living organisms including humans. Hence, there is a dire need to address this issue and create a sound balance between development and pollution levels. But first we need to understand how it affects us now and, in the future, and for this we need to predict it.

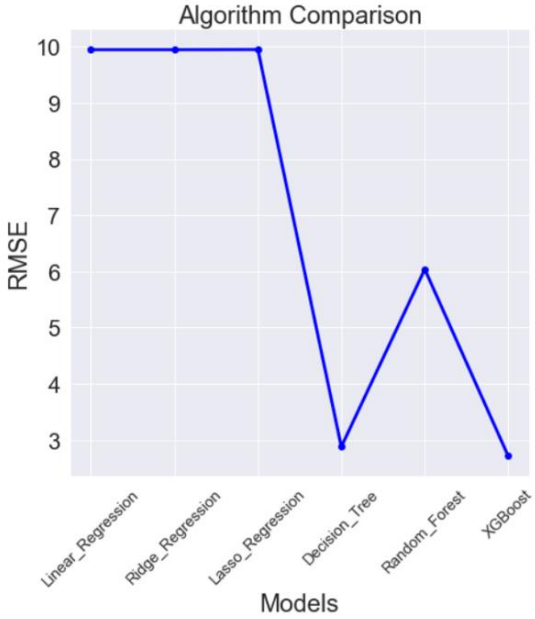
Problem Statement:

In our project we aim to predict the levels of Air Quality Index (AQI) using 'India Air Quality' dataset and regression models. Early prediction of the AQI can help authorities to plan and take steps to curb pollution and maintain a healthy and sustainable environment. We have taken into consideration chemical pollutants which affect air quality. These are Sulphur Dioxide (so2), Nitrogen Dioxide (no2), Respirable Suspended Particulate Matter (rspm), Suspended Particulate Matter (spm) and Suspended Particulate Matter with diameter less than 2.5micrometers (pm2_5) and have predicted how each pollutant contributes to AQI. Since we are predicting AQI, we have used regression models for our prediction.

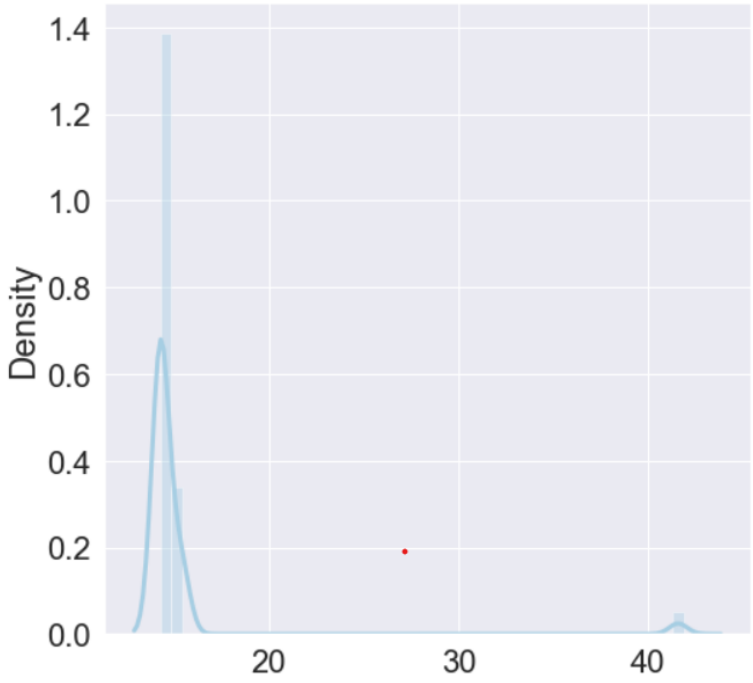
Results:

From the graph:

- **XGBoost** performs best for our dataset.
- Linear, Ridge, Lasso Regression are not the best models for our prediction and have higher RMSE as compared to XGBoost.
- Decision Tree is also a good model at predicting AQI.



Running XGBoost on test data and plotting the predicted values:



Citations:

- https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LinearRegression.html
- https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.Ridge.html
- https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.Lasso.html
- https://scikit-learn.org/stable/modules/tree.html
- https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestRegressor.html
- https://xgboost.readthedocs.io/en/stable/python/python_api.html