Analysis And Prediction Of Delhi Climate Using Machine Learning

Gaurav Lodhi [MT19063] , Kaamraan Khan [MT19064] , Vedant Desai [MT19074]

Motivation :

Delhi is infamous for air pollution and erratic weather conditions like fog, hail, smog, etc. It causes huge damage to life and resources in significant sectors like agriculture, transport system, infrastructure, etc. And the climatic conditions are only going to get worse in the near future. Thus there is an urgent need for accurate climate analysis and prediction. Unlike typical weather forecasting approaches, which have traditionally relied on physical simulations, we take a data-centric approach using machine learning algorithms to derive our solution.

Data Acquisition :

We are using two dataset as follows-
1.Dellhi weather dataset
2.Delhi air quality data
We obtained weather data for Delhi from from wunderground Delhi for the years 1997-2016. It contains various features such as temperature, pressure, humidity, rain, precipitation, etc. Second dataset is obtained from Historical Daily Ambient Air Quality Data released by the Ministry of Environment and Forests and Central Pollution Control Board of India under the National Data Sharing and Accessibility Policy (NDSAP).It contains features such as type of area ,SO2,NO2,etc.

Data Pre-processing :

We will merge the necessary features of second dataset to our first dataset by using pandas library. We will also manage noise like NAN value in our dataset. Further we will apply feature scaling using normalization on our dataset for uniformity and faster training. To deal with categorical data we will use techniques like one-hot encoding and label encoding.

Learning Techniques :

We will first implement a basic linear regression model for predicting temperature, Air pollution (Content of N02,SO2 in air), cumulative climatic condition in advance. We will then apply advance machine learning techniques like SVR, NN, Random Forest to predict the same .
We will then implement a basic logistic regression model for classifying extreme weather conditions such as thunder, hail, etc. We will then apply advance machine learning techniques like SVM, NN, Decision Tree to predict the same .

Evaluation Metric :

For our regression problem ,we will use root mean square error and for our classification problem, we will use accuracy score. Since the data for extreme weather conditions is skewed, we will use precision, recall and roc as our evaluation metric.

Deliverables :

Classification model – Gaurav Lodhi
Regression model - Kaamraan Khan and Vedant Desai
Skewed data prediction model- Kaamraan Khan and Vedant Desai
Data Acquisition- Kaamraan Khan and Vedant Desai
Data Pre-processing- Gaurav Lodhi
Data Analysis- Kaamraan Khan and Vedant Desai