Weather, Pollution Prediction and Ozone Level Detection Using ML

[1]Muskan Tanna, [2]Diksha Ramnani, [3]Nikita Samtrai, [4]Kajal Jewani
[1, 2, 3] Student, [4] Assistant Professor
Department of Computer Engineering,
Vivekanand Education Society's Institute of Technology, Chembur, Mumbai-400074, India

Abstract: In this paper, prediction of weather and air pollution along with ozone level detection is to be done. Although many factors combine to influence weather, but the major contribution is done by parameters such as temperature, heat index, dew point etc. Air quality is majorly determined by pollutants such as SO2, NO2, PM2.5, PM10 etc. The prediction is done on the basis of historical data. Ozone variability and trends are dependent on temperature, wind speed, humidity, precipitation to determine whether a day is Ozone day or Non- Ozone day. The results will show the effectiveness of Machine Learning techniques when predicting the weather and pollution.

Index Terms **- Weather, Machine learning, prediction, ozone, pollution, temperature, windspeed.**

I. INTRODUCTION

Weather is a part of our daily experience. With global warming and climate change, weather patterns are changing faster than ever. One major contributor to global warming and climate change is air pollution. Air pollution and weather co-exist in our atmosphere. The trace gas ozone plays multiple roles in the Earth system. Besides being an important greenhouse gas, it is the only absorber of harmful solar UV-B radiation which would otherwise make life on Earth impossible (WMO 2011). However, ozone's distribution in the atmosphere is subject to change. Anthropogenic and natural factors force variability and trends in its concentrations. Traditionally weather forecasting has been done by physical models. These models aren't reliable for longer durations which leads to inaccuracy**.** Machine learning over the time has proved a better solution for prediction models.In this prediction model we present a method that uses historical data of India that is to be trained by simple machine learning algorithms, which can provide usable forecast for near future. A simulated system is developed to predict various weather conditions using Data analysis and Machine learning techniques such as Recurrent Neural Networks (RNN), SVM (Support Vector Machine), KNN (K- Nearest Neighbor), Auto Regressive Integrated Moving Average (ARIMA). The model helps to predict weather, pollution details based on basic parameters such as temperature, heat index, humidity, SO2, NO2, PM2.5,etc and forecasts future weather, pollution. Along with this ozone level detection is done having factors like Wind speed (WSR), Temperature, Precipitation, Humidity.

II. LITERATURE REVIEW

Sidhart S Bhatkande, Rupali G. Hubballi (2016): In there work the authors have used data mining techniques and decision tree algorithm as a means to classify weather parameters like maximum temperature, minimum temperature in terms of day month and year.

E.B. Abrahamsen, O.M Brastian, B.Lie (2018):In there work, python API are used for collecting weather data the data have been used to train and tune several auto regressive artificial neural networks(AR-ANN) by using tensor flow from python initially only temperature of the city from past is considered for determining the future temperature. The example ANN is then extended with precipitation data and compare to the initial AR-ANN, auto regressive structure with exogenous input (ARX) i.e (ARX-ANN). Introducing precipitation as an input in the ARX model was shown to slightly improve the prediction performance.

Aditya C R , Chandana R Deshmukh , Nayana D K , Praveen Gandhi Vidyavastu (2018): In this paper, Logistic regression is employed to detect whether a data sample is either polluted or not polluted. Autoregression is employed to predict future values of PM2.5 based on the previous PM2.5 readings. Knowledge of level of PM2.5 nearing years, month or week, enables us to reduce its level to lesser than the harmful range. This system attempts to predict PM2.5 level and detect air quality based on a data set consisting of daily atmospheric conditions in a specific city.The proposed system will help common people as well as those in the meteorological department to detect and predict pollution levels and take the necessary action in accordance with that. Also, this will help people establish a data source for small localities which are usually left out in comparison to the large cities.

K. Mahesh Babu, J. Rene Beulah (2019): Machine learning is to predict the future from past data. Machine Learning (ML) is a style of artificial intelligence (AI) that delivers computers the capability to gain knowledge of without being explicitly programmed. Machine finding out can be roughly separated in to three classes. There are supervised learning, unsupervised finding out and reinforcement finding out. Supervised studying software is each given the input knowledge and the corresponding labeling to be trained data must be labeled with the aid of a person previously.

Peer Nowack , Peter Braesicke , Joanna Haigh , Nathan Luke Abraham , John Pyle and Apostolos Voulgarakis (2018): In this paper,the main processes driving ozone variability and trends are in some way be correlated with atmospheric temperature. Even though this does not always represent direct cause-effect relationships, it implies that once a certain temperature state of the atmosphere has been associated with a specific ozone distribution, such information can be used to predict future ozone distributions.

III. METHODOLOGY

The following steps are used to achieve the objectives of this paper-

A. Setup
B. Data collection
C. Data preprocessing
D. Training models.

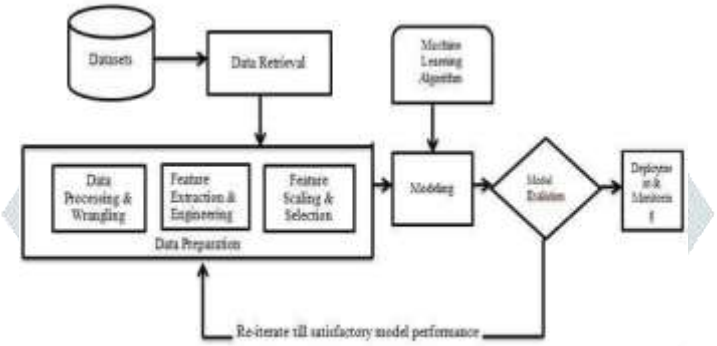

Fig. 1 Flow chart of research methodology

III-A Setup:

Complete data analysis and preprocessing in jupyter Python 3.7 environment. Using libraries like Pandas, Sklearn, NumPy, Matplotlib, Tensorflow, Keras. For training purposes, use Python jupyter notebook.

III-B Data collection:

In order to analyze different machine learning techniques. We have collected data from Kaggle for pollution, UCI repository for Ozone data and API for generating weather data. The pollution dataset includes various attributes like AQI, CO , SO2, NO, NO2, PM2.5, PM10 etc. The Ozone dataset includes wind speed (WSR) and temperature. The weather dataset includes maxtemp, mintemp, temp in Celsius, heatindex, humidity, dew point etc.

III-C Data preprocessing:

Data preprocessing in Machine Learning refers to the technique of preparing (cleaning and organizing) the raw data to make it suitable for building and training Machine Learning models. Initially, the datasets contained valueless attributes, missing instances, inadequate attributes data types and other problems that raise the necessity of preparing it first before feeding it to the analysis phase. Therefore, the datasets were passed through the following preparation stages.

III-D Training models:

For Ozone:

Support Vector Machine:

Support Vector Machine (SVM) is a supervised machine learning algorithm which can be used for both classification or regression challenges. For this method, we used temperature, wind speed, humidity, precipitation as input features and prepared a Correlation heat map for the same.

k-nearest neighbors (KNN):

K-Nearest Neighbor is one of the simplest Machine Learning algorithms based on Supervised Learning technique. K-NN algorithm assumes the similarity between the new case/data and available cases and put the new case into the category that is most similar to the available categories. We used K Neighbors Classifier with count of neighbors as 5, with attributes temperature, wind speed, humidity, precipitation. Then accordingly trained and fit the data.

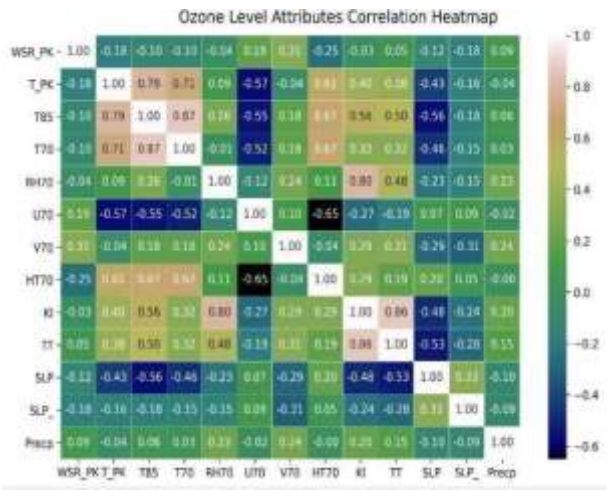

Fig. 2 CORRELATION HEATMAP

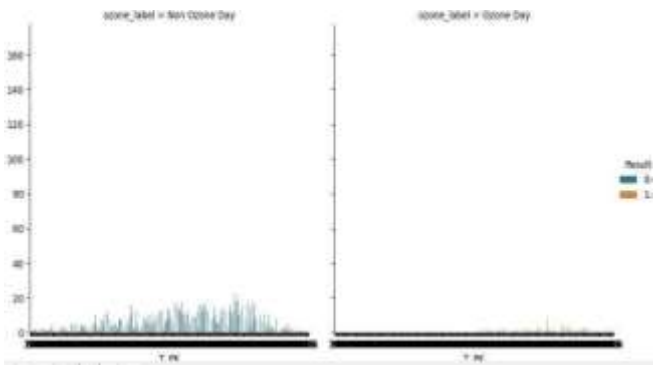

Fig. 3 LABELING OZONE AND NON-OZONE DAY

Table 3.1: COMPARISON OF MODEL FOR OZONE

ALGORITHM USED	ACCURACY (IN %)
SVM	94.9
KNN	93.7

For Weather:

Recurrent Neural Networks:

Recurrent neural networks (RNN) are a class of neural networks that are helpful in modeling sequence data. Derived from feedforward networks, RNNs exhibit similar behavior to how human brains function. Simply put: recurrent neural networks produce predictive

results in sequential data that other algorithms can't. Using RNN we have predicted temperature, humidity, dew point.

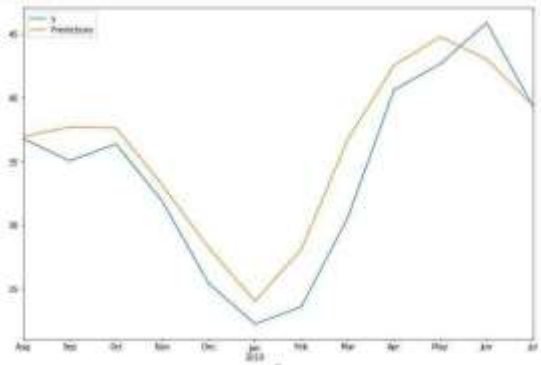

Fig. 4 ACTUAL VS PREDICTION FOR TEMPERATURE.

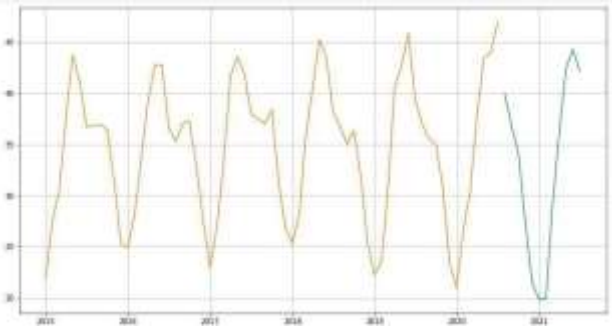

Fig. 5 FUTURE PREDICTION FOR TEMPERATURE.

The input range was from 2015 – 2020.
The RMSE obtained for temperature using RNN is 2.8270150441559694.

For Pollution:

ARIMA:

ARIMA is an acronym that stands for AutoRegressive Integrated Moving Average. It is a class of model that captures a suite of different standard temporal structures in time series data. AR: Autoregression. A model that uses the dependent relationship between an observation and some number of lagged observations. I: Integrated. The use of differencing of raw observations (e.g. subtracting an observation from an observation at the previous time step) in order to make the time series stationary. MA: Moving Average. A model that uses the dependency between an observation and a residual error from a moving average model applied to lagged observations.

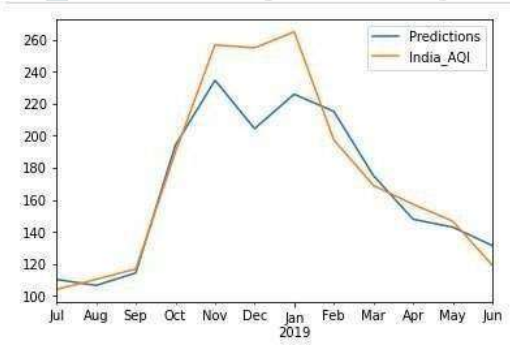

Fig. 6 ACTUAL VS PREDICTION FOR AQI

RNN:

A recurrent neural network (RNN) is a class of artificial neural networks where connections between nodes form a directed graph along a temporal sequence. In pollution AQI, PM2.5, PM10, SO2, NO2 are predicted.

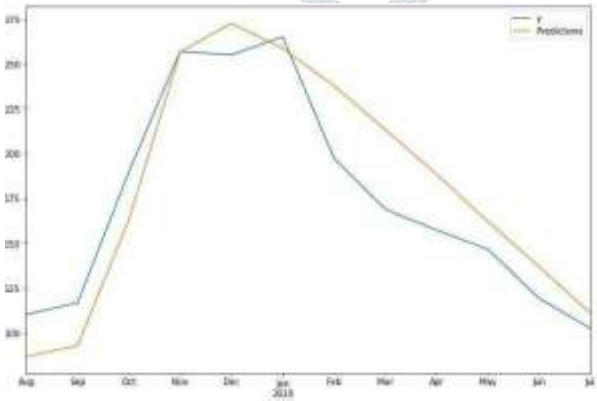

Fig. 7 ACTUAL VS PREDICTED FOR AQI

The input range was from 2015 – 2020.

Table 3.2: COMPARISON OF MODEL FOR POLLUTION.

ALGORITHM USED	RMSE
ARIMA	20.89
RNN	24.78

IV. CONCLUSION

In this work, SVM and KNN are used to detect Ozone. Two separate models were trained to detect the ozone. We saw that both the models were quite good while SVM showed greater accuracy than KNN. For weather we used RNN model for predicting various attributes such as temperature, heat index, humidity, dew point. In this paper, weather data is considered with different attributes for weather forecasting. The pollution forecasting experiment was carried out to analyze the performance of different machine learning techniques. We trained two different models on this data ARIMA, and time series RNN. We then used these models to predict AQ, PM2.5, PM10 and calculated root mean square error. From observation of this project, we found out that time series using RNN is a better method for weather and pollution forecasting.

V. FUTURE SCOPE

In this paper, we have predicted the values of the attributes based on their past values. The future work involved is to find the dependencies between various attributes so as to increase the accuracy of the model, whether introduction of data from other geographical locations can improve the prediction results.

REFERENCES

[1] https://www.kaggle.com/rohanrao/air-quality-data in-india

[2] https://archive.ics.uci.edu/ml/machine-learning- databases/ozone/

[3] https://www.worldweatheronline.com/developer/api/ historical-weather-api.aspx

[4] Imran Maqsood, Muhammad Riaz Khan, and Ajith Abraham, "An ensemble of neural networkfor weather forecasting", Neural Comput & Applic (2004) 13: 112–122

[5] Siddharth S. Bhatkande1 ,Roopa G. Hubballi2 "Weather Prediction Based on Decision Tree Algorithm Using Data Mining Techniques", Belgaum India: International Journal of Advanced Research in Computer and Communication Engineering, 2016.

[6] Radhika Y, Shashi M. Atmospheric temperature prediction using support vector machines. International Journal of Computer Theory and Engineering. 2009 Apr; 1(1):1793– 8201.

[7] Gokhale Sharad and Namita Raokhande, "Performance evaluation of air quality models for predicting PM10 and PM2.5 concentrations at urban traffic intersection during winter period", Science of the total environment 394.1(2008): 9- 24.

[8] Kumar, Anikender and P. Goyal, " Forecasting of daily air quality index in Delhi", Science of the Total Environment 409, no. 24(2011): 5517- 5523..

[9] Young P Jet al 2018 Tropospheric Ozone Assessment Report (TOAR): Assessment of global- scale model performance for global and regional ozone distributions, variability, and trends Elem. Sci. Anth.

[10] Singh Kunwar P., et al. "Linear and nonlinear modelling approaches for urban air quality prediction, " Science of the Total Environment 426(2012):244-255.