KCES’s College of Engineering and Management Jalgaon Department of Computer Engineering

2022-23 (SEM – VII)

Group Id: M4

Name of Group Members:

1. Pratham Mahajan
2. Chetan Chaudhari
3. Kunal Bhangale
4. Aditya Mahajan Project Title :

Weather Forecasting

# Introduction :

The forecasting of weather conditions and in particular the prediction of precipitation is important for hydro-power operation and flood management. Mechanistic meteorol-ogy prediction models based on 3D CFD/Navier Stokes equations (Thibault and Senocak, 2009) is extremely de-manding wrt. computing power. Generating a 14 day weather forecast can easily take 12 hours even on fast computers. Machine Learning (ML), Big Data, and use of Internet of Things (IoT) are receiving increased interest from the industry. It is well known that large amounts of data coupled with novel ML methods can produce results on par with traditional physics based models.

Due to an interest in weather monitoring in the gen-eral public, today a large number of weather stations are connected to the internet, and are thus available as cheap, distributed sensors. Additionally, several organizations that are involved in collection of meteorological data offer online data servers with accessible Application Program-ming Interfaces (API) such as the HTTP based GET/REST protocols. In order to simplify experimentation with sev-eral sources of meteorological data it is of interest to de-velop a unified API, hence facilitating the extraction of data from different sources. With large quantities of data, both historical and current measurements, it is an attrac-tive solution to use machine learning in order to predict weather conditions based on these relatively simple data sources. Using a large amount of data together with novel machine learning algorithms can then compensate for lack of complex meteorological models and yield usable fore-casts with less computing time.

Simple ML models would base predictions on auto re-gressive (AR) structures, where, say the current temper-ature in a location is correlated with several past temper-atures in the same location. In a slightly more advanced auto regressive structure, a set of properties, e.g., the tuple (temperature, humidity, and precipitation) could be corre-lated with several past values of the same tuple. An even more advanced structure is the auto regressive structure with exogenous input (ARX). In such a model, the current (local) set of properties is correlated with both past values of the same (local) set, but also with other values from the same location or values of the same properties from other locations at *current time*. Finally, in ARMAX structures, exogenous inputs at *different times* (= moving average) are used in the correlation.

# Problem Definition :

# Period of time depending on the initial state of the atmosphere. As a consequence, the numerical weather prediction community began to consider the use of probabilistic methods for forecasting, especially beyond the deterministic limit of one week or so

# Early implementation of probabilistic methods for numerical weather prediction was based on applying small, random perturbations to the atmospheric state variables (temperature, humidity, winds and pressure) in the analysed initial condition.

# Literature survey :

Weather forecasting has been one of the most challenging difficulties around the world because of both its practical value in popular scope for scientific study and meteorology. Weather is a continuous, dynamic ,multidimensional chaotic process, and data-intensive and these properties make weather forecasting a stimulating challenge. It is one of the most imperious and demanding operational responsibilities that must be carried out by many meteorological services all over the globe. Various organizations / workers in India and abroad have done demonstrating using supported time series data manipulation. The various methodologies viz. statistic decomposition models, Exponential smoothing models, ARIMA models and their dissimilarities like seasonal ARIMA models, vector ARIMA models using flexible time series, ARMAX models i.e. ARIMA with following informative variables etc., which has been used for forecasting purposes. Many trainings have taken place within the analysis of pattern and circulation of rainfall in many regions of the world. Totally altered time series methods with different purposes are used to investigate weather information in many different literatures. Accurate and timely weather forecasting is a major challenge for the scientific research. Weather prediction modelling involves a combination of many computer models, observations and acquaintance of trends and designs. Using these methods, practically accurate forecasts can be made up. Regression is a statistical experimental technique and it must be widely used in many business, the behavioural sciences, social and climate recasting and many other areas. (1980) explained the phenomena for time series regression models for forecasting the yield of rice in Raipur district on weekly data using weather parameters . In the author Kuo and Sun, (1993) was used to Associate in having intervention model for average10 days stream flow forecast and synthesis that was investigated by to effect the extraordinary phenomena caused by typhoons and different serious irregularities of the weather of the Tanshui geographical area in Taiwan. In Chiew et al, (1993) conducted a comparison of six rainfall-runoff modeling

approaches to pretend daily, monthly and annual flows in eight tolerant catchments. They concluded that time-series approaches will agreement adequate estimates of monthly and annual yields within the water capitals of the catchments. In Langu, (1993) is another approach which used statistical analysis to observe changes in weather and runoff patterns to go and look for important changes inside the parts of variety of weather statistic. Box and Jenkins (1994), in early 1970's, pioneered in developing methodologies for statistic indicating within the univariate case often known to Univariate Box-Jenkins (UBJ) ARIMA modeling in this approach of the author.The variables which describes weather conditions vary continuously with time in this explanation we found that the author , describing time series of each and every parameter and can be used to develop a forecasting model either statistically or using some new means that uses this time series data (Chatfield 1994; Montgomery and Lynwood 1996). Several authors have discussed the fuzziness associated with the weather systems.

# Proposed System :

Python API for Data Collection:

A Python API wrapper is an easy way to obtain free weather data from APIs and open data. A wrapper was designed to support multiple weather data suppliers, so it is possible to add more suppliers in the future. The API does not support the use of multiple suppliers at the same time. Currently the Norwegian Meteorological Institute data service frost.met.no and *Netatmo* are supported. The API will request hourly data for a given date, either at the station nearest to the specified latitude and longitude coordinate or within a specified rectangle as specified in kilometers centered on a given latitude and longitude co-ordinate. The wrapper uses HTTP GET requests to obtain the data from the data suppliers and returns a list where each element is a 3 item list with stationID, timestamp, and measured value. The returned data can then be saved to a file or database.

**Data Set**

The dataset is a public weather dataset from Austin, Texas available on Kaggle.

austin\_weather.csv

Columns:

Date-

The date of the collection (YYYY-MM-DD)

**TempHighF-**

High temperature, in degrees Fahrenheit

**TempAvgF-**

Average temperature, in degrees Fahrenheit

**TempLowF-**

Low temperature, in degrees Fahrenheit

**DewPointHighF-**

High dew point, in degrees Fahrenheit

**DewPointAvgF-**

Average dew point, in degrees Fahrenheit

**DewPointLowF**-

Low dew point, in degrees Fahrenheit

**HumidityHighPercent-**

High humidity, as a percentage

**HumidityAvgPercent-**

Average humidity, as a percentage

**HumidityLowPercent-**

Low humidity, as a percentage

**SeaLevelPressureHighInches-**

High sea level pressure, in inches of mercury

**SeaLevelPressureAvgInches-**

Average sea level pressure, in inches of mercury

**SeaLevelPressureLowInches-**

Low sea level pressure, in inches of mercury

**VisibilityHighMiles-**

High visibility, in miles

**VisibilityAvgMiles-**

Average visibility, in miles

**VisibilityLowMiles-**

Low visibility, in miles

**WindHighMPH-**

High wind speed, in miles per hour

**WindAvgMPH**-

Average wind speed, in miles per hour

**WindGustMPH-**

Highest wind speed gust, in miles per hour

**PrecipitationSumInches-**

Total precipitation, in inches ('T' if trace)

Events-

Adverse weather events (' ' if None)

# Expected Results :

The goal of this work is to predict the temperature using an Existing weather dataset. We will get 4 charts based on our dataset one will show average temperature 2nd will show precipitation level other charts will present humidity, visibility and wind pressure.

# Software and hardware requirements:

Software tools required :

1. Python
2. Jupyter Notebook
3. MS Excel

# References :

Divya Chauhan and Jawahar Thakur. Data mining techniques for weather prediction: A review. *International Journal on Recent and Innovation Trends in Computing and Communi-cation*, 2(8):2184–2189, 2014.

Dan Claudiu Ciresan, Ueli Meier, Luca Maria Gambardella, and Jürgen Schmidhuber. Deep big simple neural nets excel on handwritten digit recognition, 2010. *Cited on*, 80, 2010.

Ian Goodfellow, Yoshua Bengio, Aaron Courville, and Yoshua Bengio. *Deep learning*, volume

1. MIT press Cambridge, 2016.

Jeff Heaton. The number of hidden layers, 2017.

Max Kuhn and Kjell Johnson. *Applied predictive modeling*, vol-ume 26. Springer, 2013.