

Implication of Machine Learning towards Weather Prediction

Shivangi Raj¹, Deepak Arora², Pooja Khanna³, Bramah Hazela⁴

Department of Computer Science & Engineering,

Amity School of Engineering and Technology,

Amity University, Uttar Pradesh, Lucknow Campus, INDIA

shivangi.raj13october@gmail.com¹, deepakarorainbox@gmail.com², pkhanna@lko.amity.edu³, bhazela@lko.amity.edu⁴

Abstract-the following paper is about weather prediction using python programming language and concepts of machine learning. This report is an attempt to predict the weather of next five days at the interval of three hours. The concepts of python help in extracting the required data from the API (open weather API), wherein API refers to Application Programming Interface. Through the coding in python and using the library requests we have requested the site to get the predicted weather and we have extracted and clean the data to make the correct use. The basic data types list and dictionary are used to extract the required data from the raw API. The libraries 'Requests', 'Matplotlib', 'Pandas' are used for the coding in python to implement the prediction of weather. The demonstration and use of these libraries are clearly depicted in the report. These libraries play a very important role in the execution of the program.

1. INTRODUCTION

Weather prediction or forecasting is the topic that has to be worked upon. Since, ancient era, weather prediction is being considered as the most interesting and fascinating activity. Weather is predicted making use of past data. Weather forecasting provides critical information about information about weather using computer science.[1] Initially, in the ancient times people tried to predict weather by looking at the sky and analyzing the activities of some animals as it is considered that they have a better sense than humans. But now, we have different technologies to predict weather. There are different apps that give easy access to future weather details. In today's time weather prediction plays a very important role in every field, may it be in airways, railways or day-to-day activities of common people. People are using weather prediction sites or apps to decide their every day's actions. So, learning the technology and computer science behind weather prediction is an attempt to better the forecasting details in future Prediction of weather is an application of machine learning. Machine learning can be done making using of python. Firstly, the question arises what is machine learning?

The answer is, Machine learning is a type of AI (artificial intelligence) that gives the ability to perform tasks without being explicitly programmed, and it takes the help of prior experience.[2] Python is open source, high level programming language that is interpreted and the most trending language in current years. It is convenient and easy to use. This programming language provides an easy platform to implement machine learning codes. Machine learning has many applications; some of these applications include

Weather forecasting, Face recognition- this is widely used these days in smart devices for the security reasons, Credit card approval, Handwriting recognition, Speech recognition, Web advertising, Pattern recognition that is again used for security purpose, Shopping suggestions and advice. All of

the above are its application wherein weather forecasting is discussed in detail

These applications are possible either through supervised learning, unsupervised learning or reinforcement learning. [11] Both R and python are open source programming languages but python provides a more general approach to data analytics and data science. Unlike R, Python is a complete language and platform that is used both for research and developing production systems. A basic knowledge of python is required for the implementation of machine learning.

2. PYTHON

Python is open source, high level programming language that is interpreted and the most trending language in current years. It is convenient and easy to use. This programming language provides a easy platform to implement machine learning codes. Python is created by Guido van Rossum and released in 1991. Some characteristics of the programming language PYTHON: Open source programming language (It means it follow open source protocol, with certain provisions can be modified by the public itself), It is interpreted, It is high level object oriented programming language. It is a partial object oriented programming language. Python has no primitive data type, Portable, Every python file has a .py extension, and Integration with other components is easy. For example, C and Python – Cython, Java and Python-Python.

There are about 5000 libraries in Python.

3. STANDARD LIBRARIES USED IN PROPOSED WORK

Some standard libraries that are being used in this project:

1. Pandas
2. Matplotlib
3. Requests

In order to work on the project, firstly you need to install python and pycharm community for the installation of libraries run the command prompt of the desktop as administrator and use as follows:

```
C:\windows\system32>pip3 install numpy (and press enter, note internet access is Compulsory)
```

Using this command any library can be installed.

All these libraries are discussed well with small code examples.

A. PANDAS

Pandas are the open source, BSD licensed Python library. It is most widely used in memory representation of complex data frames, series, data collections in Python and provides high performance in the field of analyzing, cleaning, munging, modelling, organizing the result of analysis that in turn can be used for plotting, graphical display and tabular display.

```
import pandas as pd
```

```
df=pd.DataFrame()
```

```
print (df)
```

output:

```
Empty DataFrame
```

```
Columns: []
```

```
Index: []
```

[Note: these demos are run in pycharm community, and exact output are shown]

```
import pandas as pd
```

```
list=[[1,2,3,4],[3,4,5,6]]
```

```
df=pd.DataFrame(list,columns=['column A', 'column B','column C','column D'])
```

```
print (df)
```

The output of the following code will be as follows-

```
column A column B column C column D
0      1      2      3      4
1      3      4      5      6
```

This example will be used in the original program where there will be a list of many lists. After passing the list parameter 'columns' keyword is used for assigning the names to various columns. Similarly the names to various rows can be assigned using the keyword 'index' in the same way as the keyword 'columns' is used.

```
import pandas as pd
```

```
list=[[1,2,3,4],[3,4,5,6]]
```

```
df=pd.DataFrame(list, index=['row A', 'row B'])
```

```
print (df)
```

the output of the following code will be as follows-

```
0 1 2 3
row A 1 2 3 4
row B 3 4 5 6
```

The 'pandas' library has many functions. One of them is 'DataFrame'. In the previous examples the working of pandas dataframe is demonstrated. A pandas DataFrame can be created using the constructor:

```
pandas.DataFrame( data, index, columns, dtype, copy)
```

B. MATPLOTLIB

This library is used frequently in Python programs. 'Matplotlib' is a plotting library for the python programming language and its numerical extension 'Numpy', this library provides us with the ability to make our project eye-catching and more demonstrating. The DataFrame that is formed using Pandas library can be made more demonstrating by plotting it, as plotting a graph gives the user an ease of understanding.

Pyplot provides a convenient platform for the Matplotlib plotting library. Here matplotlib.pyplot is imported as plt. This means that instead of writing the whole name again and again we can write 'plt' instead of 'matplotlib.pyplot'.

For example:

```
import matplotlib.pyplot as plt
```

```
import pandas as pd
```

```
import numpy
```

```
x=numpy.linspace(-numpy.pi,+numpy.pi)
```

```
c=numpy.cos(x)
```

```
s=numpy.sin(x)
```

```
plt.plot(x,c,color='red')
```

```
plt.plot(x,s,color='violet')
```

```
plt.show()
```

The output of the following code is a graph shown as:

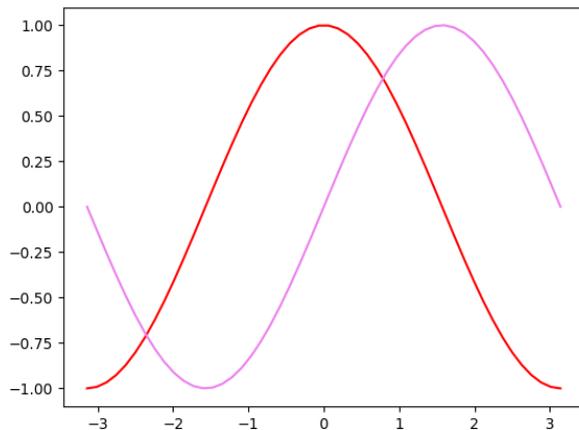


Fig.1. Sine and cosine curve using Matplotlib. The red color demonstrates cosine curve and violet color depicts sine curve.

Example - this is the example how a DataFrame is plotted choosing particular columns.

```
import matplotlib.pyplot as plt
import pandas as pd
#import numpy
list=[[1,2,3,4],[5,6,7,8]]
df=pd.DataFrame(list,columns=['aa','bb','cc','dd'])
print (df)
df[['aa','bb']].plot(kind='bar')
plt.show()
```

The output plot will be as follows:

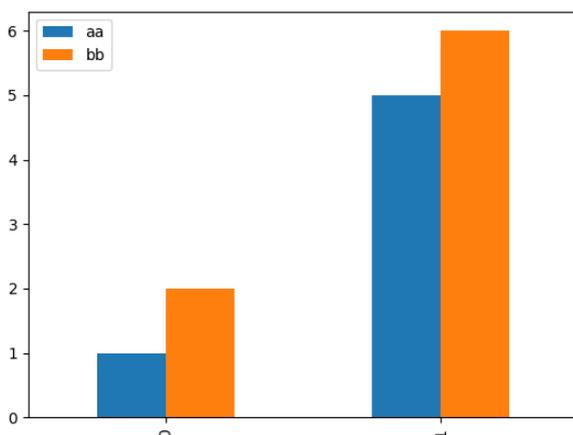


Fig. 2. Plotting two columns of the dataframe using 'Matplotlib'

C. REQUESTS

This is also a library in python which helps the coder to send requests to various sites or HTML and acquire the contents of the site page.[3] This library helps us in using the contents of a site in an authorised form. It is not hacking but it is

simply requesting for the codes that are prevailing in the requested sites. The use of this python library is simple. Here in the weather prediction coding we are making request to the open weather API website to give us the dictionary of the weather forecast. The following code is a simple example of using the 'Requests' library in python .

```
import urllib.request as u
def main():
    web=u.urlopen("http://www.precursorinfo.org/")
    for code in web:
        print (str(code,encoding="utf_8"))
    url=u.urlopen()
if __name__=="__main__":main()
```

4. APPROACH

Moving to the original program of weather prediction using the coding of Python. The first step is to subscribe for the open weather API from <https://openweathermap.org/api> and subscribe for 3 hour 5 day forecast. After signing in a unique API key is allotted to every subscriber. We use this API key to get the dictionary containing the forecasted weather. The API is present is mostly in the form of json or XML. [4]Care is being taken while fetching the required features of weather that we want to make use of. The API key is called making use of the city id. The features that has been taken into account are Humidity, Weather,

Weather description, Temperature, Maximum Temperature, Minimum temperature, pressure. There may be many features that can be taken into account depending on one's choice. The features chosen are a pure matter of choice.

These features are counted for different hours of different dates. To fetch all these features a basic concept of list and dictionary in Python is needed to be known. The API got is in the form of a big dictionary and in that dictionary there are different keys and values and for some keys the values is again a list. 'List' and 'Dictionary' are standard data types in the python programming language.

a. LIST AND DICTIONARY

List is a sequence which contains elements at a particular index positions. In a list, the elements can be duplicated. It is like an array but can contain dissimilar data type values. Most importantly a list is mutable and there are many functions that can be used to alter the list. For example: if

```
Fruitlist= ["banana","apple","guava"]
```

This is a list and if we want to fetch out the 'banana' and print it then the following code is to be executed.

```
print(Fruitlist[0])
if there is a list containing a dictionary. For example:
list=[{fruit:'banana',vegetable:'ladyfinger'}]
in this case use
print(list[0]['fruit'])
```

This will give output as 'banana' in the same way, access to a list element is made possible .

Dictionary is another standard python data types. It works in key value pair. It is not in ordered form. Items are always enclosed in { }.

Value is accessed using the key.

For example,

```
Collegedic={123: "xyz",456: "abc"}
```

```
print(Collegedic[123])
```

```
print(Collegedic.get(123))
```

both these statements will give the output as xyz

5. CODING EXPLANATION

Moving on with the coding firstly it is needed to import the libraries that are, pandas and from pandas json is also imported, 'matplotlib.pyplot' is imported and 'urllib.request' is imported. As other programming language the first function to be executed is 'main' where the other functions are called. The functions in python are defined using the 'def' keyword. 'try' and 'catch' are used to prevent the unexpected termination of the program. Three functions are called and used that perform different tasks. The first one is 'url_builder' which builds the full url so that the request can be sent and API can be collected. [5]The API key is initialised form and modes are added. City Id is passed at the time of calling the main function. The second function is 'data_fetch' which sends request to the server and gets the full dictionary containing all the data of different time and which is fetched in the third function i.e. 'data_oraganiser' which will organize the data and fetch it from the API and used the 'for' loop to fetch all the features at different time and date using the basic concepts of 'list' and 'dictionary' as described above. The data fetched in variables are collected inside the loop in a list. [7-10]

In this case many lists same as number of times the loop will run are formed which in turn are appended to a second list, which will contain all the lists. Using this list a 'DataFrame' is created using 'pandas' library. Now the dataframe formed contains the predicted weather containing all the features-1. Humidity, 2. Weather 3.weather description 4. Temperature 5. Maximum temperature 6. Minimum temperature 7.pressure 8. Date and time. Dataframe formed is in form of rows and columns. Then the graphs are plotted using 'Matplotlib' library. The graph gives a better understanding of the weather. Weather is forecasted for all the places just by changing the city id that is passed in the main function for calling the function 'url_builder'. The main function of the program is called in this way:

If there is no internet access then there is a possibility of the unexpected termination of the program so the calling statement is placed under the 'try' block so that if there is an 'IOError' it is would tell the user about it printing "no internet".

6. RESULTS

Three graphs are plotted, the graphs plotted are easy to understand and that is the reason graphs are plotted so that that is eye-catching.

The first one (Fig. 3.) is between minimum and maximum temperature as per different dates and timings. It is bar plot and is from the dataframe that contain all the features as mentioned above and the data frame contains the date and time also. If we see there are 38 rows and 8 columns so the 38 rows are indicated through x axis whereas the maximum and minimum temperature on the y-axis .The blue colour shows maximum temperature whereas the orange colour shows minimum temperature.

The second graph (Fig.4.) is again a bar plot of the data frame considering the humidity and pressure at different timings of the different dates. Again the x-axis represents the different rows that in turn represent the different timings and dates and the y-axis showing the pressure and humidity. Pressure is represented with blue colour and humidity with orange colour.

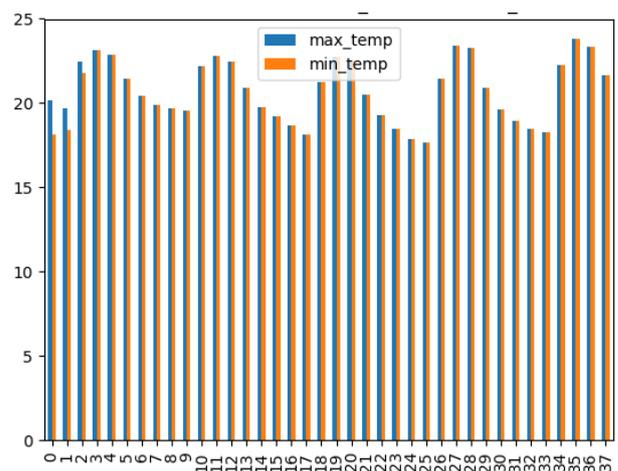


Fig. 3. Bar plot for minimum and maximum temperatures

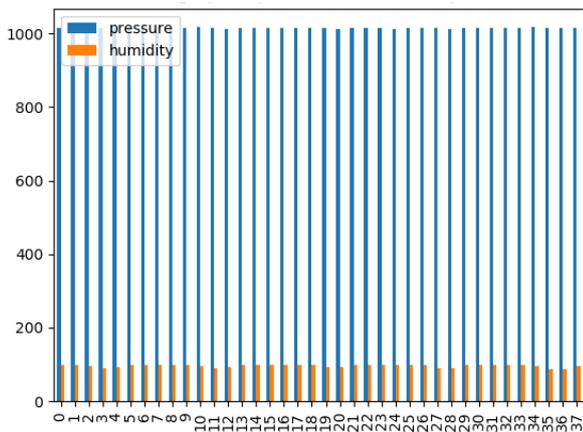


Fig. 4. Bar plot for pressure and humidity

The third graph (Fig. 5) is a gist of all the features of the weather predicted, i.e. humidity, temperature, maximum temperature, minimum temperature, pressure.

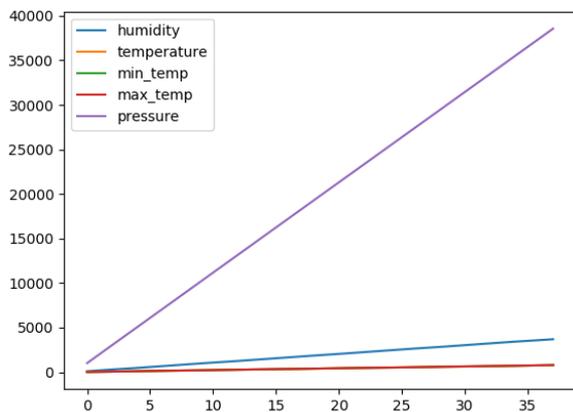


Fig. 5. The gist of all features in a single plot.

7. CONCLUSION

This paper was an attempt to study the basic elements of machine learning and the programming language Python. A trivial program is created attempting to predict the weather of any city for the next five days using the city id. A simple program has been worked upon to understand the concepts of PYTHON and Machine learning.[6] Coding is run on the Pycharm 3.6 environment where we need to import the required libraries using the command prompt before starting the necessary coding, and internet access is necessary.

The coding of the program is thoroughly explained in the report; the basic libraries are explained in detail whereas the basic python tools or data types that are used to fetch the data from the API are also explained. Machine learning has a great scope in present as well as in near future. The coding is run on the Pycharm environment and three graphs are also retrieved using the Matplotlib library. This paper is a small attempt to start working on the trending topics like Python, machine learning, Artificial intelligence, deep learning and neural learning and will surely work on these topics in future.

REFERENCES

- [1] Mark Holmstrom, Dylan Liu, Christopher Vo, "Machine Learning Applied to Weather Forecasting" Stanford University, 2016. Available: <http://cs229.stanford.edu/proj2016/report/HolmstromLiuVo-MachineLearningAppliedToWeatherForecasting-report.pdf>
- [2] Piyush Kapoor and Sarabjeet Singh Bedi "Weather Forecasting Using Sliding Window Algorithm", Kvantum Inc., Gurgaon 122001, India MJP Rohilkhand University, Bareilly 243006, India, 2013. [Online] Available: <https://www.hindawi.com/journals/isrn/2013/156540/>
- [3] Divya Chauhan, Jawahar Thakur "Techniques for Weather Prediction": A Review 5, India: ISSN, 2013. [4] Qing Yi Feng1 , Ruggero Vasile, Marc Segond4 ,
- [4] Avi Gozolchiani, Yang Wang , Markus Abel , Shilomo Havlin, Armin Bunde, and Henk A. Dijkstra1 "ClimateLearn: A machine-learning approach for climate prediction using network measures," , Germany, 2016.
- [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] Sanyam Gupta, Indumathy , Govind Singhal "Weather Prediction Using Normal Equation Method and Linear regression Techniques" , Vellore, Tamil Nadu, India: International Journal of Computer Science and Information Technologies, 2016.
- [7] Muthulakshmi A, ME (SE), Dr.S.Baghavathi Priya "A survey on weather forecasting to predict rainfall using big data analytics", Chennai.: IJISSET , 2015.
- [8] K. Mohammadi, S. Shamshirband, S. Motamedi, D. Petković, R. Hashim, M. Gocic, "Extreme learning machine based prediction of daily dew point temperature", Computers and Electronics in Agriculture, vol. 117, pp. 214-225, 2015.
- [9] Abramson, Bruce, et al "Hailfinder: A Bayesian System for forecasting severe weather". International journal of forecasting. 12.1(1996):57-71
- [10] "Stanford ,CA" in weather underground, The weather company, 2016. [online]. Available: <https://www.wunderground.com/us/ca/palo-alto/zmw.94305.1.99999>
- [11] Ng, Andrew. "CS229 Lecture Notes Supervised Learning", 2016