**WEATHER FORECASTING APPLICATION USING PYTHON**

**ABSTRACT**

 Weather forecasting is the application of science and technology to predict the state of the atmosphere for a given location. Ancient weather forecasting methods usually relied on observed patterns of events, also termed pattern recognition. For example, it might be

observed that if the sunset was particularly red, the following day often brought fair weather.

However, not all of these predictions prove reliable. System will take this parameter and will predict weather (rainfall in inches) from previous data in database (dataset). The role of the admin is to add

previous weather data in database, so that system will calculate weather (estimated rainfall in

inches) based on these data. This system can be used in Air Traffic, Marine,

Agriculture, Forestry, Military, and Navy etc.

**I. INTRODUCTION**

Weather forecasting is the prediction of the state of the atmosphere for a given location using

the application of science and technology. This includes temperature, rain, cloudiness, wind

speed, and humidity. Weather warnings are a special kind of short-range forecast carried out

for the protection of human life.

Today there are no certain methods by using which we can predict whether there will be rainfall today or not. Even the meteorological department’s prediction fails sometimes. Here we will learn how to build a machine-learning model which can predict whether there will be rainfall today or not based on some atmospheric factors. This problem is related to Rainfall Prediction using Machine Learning Machine Learning  models tend to perform better on the previously known task which needed highly skilled individuals to do so

Features Weather Forecast Project In Python :

• Time to time update weather

• Temperature Update

• Last 7 days data Predict

• change weather in every hours as according to weather changes.

• provide accurate data information about weather.

• user can search weather anytime and anywhere.

• any places data can be search and provide information as according to weather.

• help user to travel.

• help User to future plans for holidays.

**II. EXISTING SYSTEMAND ITS DRAWBACKS**

* Temperature Update
* Last 7 days data update
* Change weather in every Hour
* Help User to travel & Plan their Holidays

The traditional forecast process employed by most NMHSs involves forecasters producing

text-based, sensible, weather-element forecast products (e.g. maximum/minimum

temperature, cloud cover) using numerical weather prediction (NWP) output as guidance.

The process is typically schedule-driven, product-oriented and labour-intensive. Over the last

decade, technological advances and scientific breakthroughs have allowed NMHSs’ hydro

meteorological forecasts and warnings to become much more specific and accurate. As

computer technology and high-speed dissemination systems evolved (e.g. Internet), National

Weather Service (NWS) customers/partners were demanding detailed forecasts in gridded,

digital and graphic formats. Traditional NWS text forecast products limit the amount of

additional information that can be conveyed to the user community. The concept of digital

database forecasting provides the capability to meet customer/partner demands for more

accurate, detailed hydro meteorological forecasts. Digital database forecasting also offers one

of the most exciting opportunities to integrate PWS forecast dissemination and service

delivery, which most effectively serves the user community.

**III. PROBLEM STATEMENT**

The traditional forecast process employed by most NMHSs involves forecasters producing

text-based, sensible, weather-element forecast products (e.g. maximum/minimum

temperature, cloud cover) using numerical weather prediction (NWP) output as guidance.

The process is typically schedule-driven, product-oriented and labour-intensive. Over the last

decade, technological advances and scientific breakthroughs have allowed NMHSs’

hydrometeorological forecasts and warnings to become much more specific and accurate.

**IV. PROPOSED METHOD**

User will enter current temperature; humidity and wind, System will take this parameter and

will predict weather from previous data in database. The role of the admin is to add previous

weather data in database, so that system will calculate weather based on these data. Weather

forecasting system takes parameters such as temperature, humidity, and wind and will

forecast weather based on previous record therefore this prediction will prove reliable.

**V. MODULES AND PROJECT DESCRIPTION**

In this Module we have two Modules

1. Data Gathering and pre-processing
2. Applying Algorithm for Prediction

Explanation:

1) In this module we first gather the data(dataset) for our prediction model.Data comes in all

forms, most of it being very messy and unstructured. They rarely come ready to use.

Datasets, large and small, come with a variety of issues- invalid fields, missing and additional

values, and values that are in forms different from the one we require. In order to bring it to

workable or structured form, we need to “clean” our data, and make it ready to use. Some

common cleaning includes parsing, converting to one-hot, removing unnecessary data, etc.

In our case, our data has some days where some factors weren’t recorded. And the rainfall in

cm was marked as T if there was trace precipitation. Our algorithm requires numbers, so we

can’t work with alphabets popping up in our data. so we need to clean the data before

applying it on our model.

2) Once the data is cleaned, In this module that cleaned data can be used as an input to our

Linear regression model. Linear regression is a linear approach to form a relationship

between a dependent variable and many independent explanatory variables. This is done by

plotting a line that fits our scatter plot the best, ie, with the least errors. This gives value

predictions, ie, how much, by substituting the independent values in the line equation.

We will use Scikit-learn’s linear regression model to train our dataset. Once the model is

trained, we can give our own inputs for the various columns such as temperature, dew point,

pressure, etc. to predict the weather based on these attributes.

Module Outcomes:

1) By the end of the first module the fully cleaned and useful data is available for the apply

the algorithm for the prediction.

2) By the end of the second module the actual prediction will be happen the outcome is the

amount of rainfall in inches based upon the users input.

**VI.ALOGORITHM AND ITS TECHNIQUES**

Linear Regression is a machine learning algorithm based on supervised learning. It performs

a regression task. Regression models a target prediction value based on independent

variables. It is mostly used for finding out the relationship between variables and forecasting.

Different regression models differ based on – the kind of relationship between dependent and

independent variables, they are considering and the number of independent variables being

used.

A graph of a line with dots

Description automatically generated

Linear regression performs the task to predict a dependent variable value (y) based on a given

independent variable (x). So, this regression technique finds out a linear relationship between

x (input) and y(output). Hence, the name is Linear Regression.In the figure above, X (input)

is the work experience and Y (output) is the salary of a person. The regression line is the best

fit line for our model.

Hypothesis function for Linear Regression :

Y = MX+C

Where Y is the response variable. X is the predictor variable.

m and c are constants which are called the coefficients.

Exploratory Data Analysis

[EDA](https://www.geeksforgeeks.org/what-is-exploratory-data-analysis/) is an approach to analyzing the data using visual techniques. It is used to discover trends, and patterns, or to check assumptions with the help of statistical summaries and graphical representations. Here we will see how to check the [data imbalance](https://www.geeksforgeeks.org/imbalanced-learn-module-in-python/) and skewness of the data.

A blue and orange pie chart

Description automatically generated

Here we can clearly draw some observations:

* maxtemp is relatively lower on days of rainfall.
* dewpoint value is higher on days of rainfall.
* humidity is high on the days when rainfall is expected.
* Obviously, clouds must be there for rainfall.
* sunshine is also less on days of rainfall.
* windspeed is higher on days of rainfall.

**VII. IMPLEMENTATION**

A day (in red) having precipitation of about 2 inches is tracked across multiple parameters

(the same day is tracker across multiple features such as temperature, pressure, etc). The x-

axis denotes the days and the y-axis denotes the magnitude of the feature such as temperature,

pressure, etc. From the graph, it can be observed that rainfall can be expected to be high when

the temperature is high and humidity is high.

**Let’s draw boxplots for the continuous variable to detect the outliers present in the data.**

**Model Training**

**Now we will separate the features and target variables and split them into training and testing data by using which we will select the model which is performing best on the validation data.**

A group of blue boxes with numbers

Description automatically generated

**Sometimes there are highly correlated features that just increase the dimensionality of the feature space and do not good for the model’s performance. So we must check whether there are highly correlated features in this dataset or not.**

A screenshot of a graph

Description automatically generated

**VIII.CONCLUSION**

We successfully predicted the rainfall using the linear regression but here this is not very

accurate only sometimes any way it depends upon the climate changes to season to season.

Here we are taking only summer season weather data set it only useful to predict rainfall in

summer season.

