*A project report on*

RAINFALL PREDICTION

*By*

# MYTHILY S

Under the guidance of

# Dr. CHANDRA MOULISWARAN S

**Associate Professor**

# School of Information Technology & Engineering

**VIT University, Vellore**

## DECLARATION

I hereby declare that the project entitled “**RAINFALL PREDICTION USING MACHINE LEARNING**” submitted by **S. MYTHILY** for the award of the degree of BCA to VIT is a record of bonafide work carried out by me under the supervision of **Dr. CHANDRA MOULISWARAN S.**

I further declare that the work reported in this project has not been submitted and will not be submitted, either in part or in full, for the award of any other degree or diploma in this institute or any other institute or university.

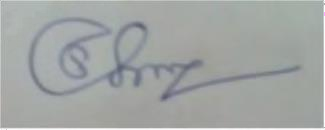
Place: Vellore Date:

Signature of the Candidate

## CERTIFICATE

This is to certify that the project entitled “RAINFALL PREDICTION USING MACHINE LEARNING” submitted by **S. MYTHILY** School of Information Technology and Engineering, VIT, for the award of the degree of **BCA** is a record of bonafide work carried out by him/her under my supervision.

The contents of this report have not been submitted and will not be submitted either in part or in full, for the award of any other degree or diploma in this institute or any other institute or university. The Project report fulfils the requirements and regulations of VIT and in my opinion meets the necessary standards for submission.



**Signature of the Guide**

**ABSTRACT**

The idea logy of this project is outfit for rainfall prediction by multiple models of machine learning techniques. Rainfall forecasting plays a vital role in prediction of torrential rain and erratic rainfall with prior planning to avoid natural calamities and other consequences, including impacts on agriculture and public and personal assets. As an country’s economy is far more based on agriculture, upgraded forecasting models n techniques will enhance the forecasting of rainfall for irrigation and harvesting purposes. This will be a better pathway and stand as a pillar for the country's economical and farming growth

Henceforth four algorithm of ML for prior prediction of rainfall and will go for one of the finest algorithm to exhibit the net output on the website with the help of Anaconda flask.

Typical rainfall prediction is done by methods of Artificial Intelligence, neural networks, geo-statistical techniques. It enriches the supply of spatial data. Unfortunately the subsidy budget of geo-statistical technique is much above. Since the complication in the datasets of historical rainfall grows, the existing approach gets lagged. Machine learning prediction will support economically and Machine learning algorithms based on the information we provide. With each new piece of data added, the model's accuracy and efficiency in making decisions is expected to be improved.

**Keywords:** Prediction, Machine Learning, Anaconda, Algorithms.

# CONTENTS

LIST OF ACRONYMS… VII

**CHAPTER 1**

**INTRODUCTION**

1.1 GENERAL 1

CHAPTER 2

PROJECT DESCRIPTION

* 1. [DATASET DESCRIPTION 5](#_TOC_250003)
  2. [ALGORITHM DESCRIPTION 5](#_TOC_250002)

CHAPTER 3

3.1 LITERATURE SURVEY 7

CHAPTER 4

REQUIREMENT ENGINEERING

* 1. HARDWARE REQUIREMENT 11
  2. SOFTWARE REQUIREMENTS 11

CHAPTER 5

* 1. [SYSTEM ARCHITECTURE 12](#_TOC_250001)
  2. [FLOW CHART 13](#_TOC_250000)

## CHAPTER 6

* 1. IMPLEMENTATION 14
  2. WEB APP IMPLEMENTATION 43

## CHAPTER 7

* 1. CONCLUSION 55
  2. FUTURE WORK 55

## CHAPTER 8

8.1 REFERENCE 56

**LIST OF ACRONYMS**

|  |  |
| --- | --- |
| KNN | K- Nearest Neighbour |
| AUC | Area Under the Curve |
| MLP | Multi-Layer Perceptron |
| SVM | Support Vector Machine |
| WHO | World Health Organisation |
| ANN | Artificial Neural Networks |
| PCA | Principal Component Analysis |
| MICE | Multivariate Imputation by Chained Equations |
| ROC | Receiver Operating Characteristics |
| ML | Machine Learning |
| EBP | Error Back Propagation |

## CHAPTER 1

**INTRODUCTION**

The data of rainfall extracted from the Kaggle is taken for experimental study. A decade's rainfall measurements is taken as database. The impact of water scarcity throws more focus on rainfall forecasting for the entire world. With this case study, we are going to specify a model that suits cent percent for rainfall prediction.

This forecast favours farmers, in addition to the optimal use of water resources. As it is challenging task for Prediction of rainfall, the accuracy matters in all aspects. As much as electrical and mechanical machines and devices are available in the prediction of rainfall depending on weather statistics of temperature, humidity, and pressure. As the accuracy of traditional techniques is inefficient, machine learning algorithms will rectify this by procuring accurate outputs. This can be easily done with previous rainfall data and forecasting of rainfall for future seasons.

We can use a variety of methodologies based on the requirements, and we can also calculate the real forecast as well as the accuracy. Different strategies provide varying degrees of accuracy, therefore it's very important to pick the proper algorithm and model it to meet the needs.

**Keywords:** Prediction, Machine Learning, Rainfall, Algorithms, Accuracy.

## OVERVIEW OF PROJECT

The precise and accurate prediction of rain continues to be lacking, which could aid in a variety of fields such as agriculture, water-way preservation, and swamp projection. The difficulty is in formulating the calculations for rainfall prediction that can be supported by previous findings and similarities and can provide reliable and appropriate output predictions.

Imprecise and inaccurate predictions appear to be not only a waste of your time, but also a waste of resources, resulting in inefficient crisis management, such as less agriculture cultivation, poor water reservoirs, and poor flood management. As a result, the requirement is not only to develop a rainfall prediction system, but also one that is more accurate and precise than the current rainfall prediction system.

## PROBLEM STATEMENT

Predicting precipitation accurately and precisely remains a challenge, which could aid in a variety of fields such as agriculture, water-way preservation, and swamp projection. The challenge is speculating on a precipitation prediction computation that is supported by previous findings and similarities and can provide square measure reliable and applicable output predictions.

Inexact and inaccurate predictions appear to be not only a waste of your time, but also a waste of resources and a source of inefficient crisis management, such as poor agricultural cultivation, inadequate water reservoirs, and inadequate flood management. As a result, the requirement is not only to develop a failure predicting system, but also one that is more correct and precise than the current predictors.

## OBJECTIVE OF THE PROJECT

India is a "agriculture country," and the "economy" is based on crop cultivation and moderate rainfall. Predicting rain-fall is essential and required for the analysis of "crop production" for all people or any people who do farming.

This technology is important in for predicting atmospheric conditions. Prior prediction of rainfall will helps in the prevention of wasting water. And also it results in better use of water resourcing agents, crop productivity, and water resource agent planning. Rainfall forecasting is improving in accuracy and predictability.

## SCOPE OF THE PROJECT

Nowadays, technologies are rapidly evolving, and a recent technology known as machine learning (ML) aids in making this evolution more efficient. The primary goal of rainfall prediction was to provide trustworthy assistance to humanity or organisations that can be used to reduce climatic losses and regrowth society's profits, essential needs, and so on. As a result, agriculture is the primary occupation of the majority of the country's population; however, this may change in small-scale business.

## PROPOSED SYSTEM

The prompt of moving to the specifics of AI & here even the simple things are also maintained with the assistance of machines Another AI process in which machines think for themselves in the absence of any other related sources is machine learning (ML). It teaches the appliances to be told from history or data in ML, and then checks out and boosts them, thereby evaluating from its previous process. Predictive analysis using ML can provide answers to more complex problems with confidence. In this work, we used a rainfall dataset from the Kaggle repository.

And in this case, I used the ml algorithm to forecast the RF and aid in the search for the best suitable algorithm's for forbidden usage, which means ease, precision, and so on. We used a rainfall dataset from the "Kaggle repository" in this work. And this uses ML algorithms to predict and ease ought to produce better algorithm in all means of factors such as ease, accuracy, and so on.

**STEPS**

* Here the balancing of data is worked at first
* The encoding is worked for “categorical variables”.
* This uses the “sophisticated imputation like MICE”.
* Outliers can be detected and which will be excluded from dataset.
* Feature selection is done using filter and wrapper method.
* speed and performance are compared between various popular models
* This chooses better one for judging the data set: factors “Accuracy, F1-Score or Cohen's Kappa, time”
* By using these, approaches we can easily interpret the result.

## CHAPTER 2 MODULE DESCRIPTION

## DATASET DESCRIPTION

Dataset is taken from Kaggle. The dataset contains 142193 rows and 26 columns. Here in this proposed work we are using the 21 attributes of the dataset in order to predict the result. The 21 attributes helps the application to find whether it has rain tomorrow or not. In this dataset some has numerical value and has categorical values and some other in order to predict the result the whole data has changed to float value so that their will be no complexity to fetch. Many methods are used to predict the because of its high data value.

## ALGORITHM DESCRIPTION

In this project, I planned to implement five different ML algorithms to analyse its performance as well as the accuracy of the result on this dataset in order to predict the result for the data which we are getting from the user. So by implementing these algorithm I will get one best algorithm, with the best algorithm I am, to predict my resultant values. The algorithms are:

## RANDOM FOREST

The Random Forest Algorithm's ability to handle data sets with both continuous and categorical variables, as in regression and classification, is one of its most important features. It outperforms the competition when it comes to classification problems

**XGBOOST**

Extreme Gradient Boosting, this algorithm gives decision tree in which subsequent tree will reduce the error of previous tree. It is able to do parallel on a single machine computing because it is linear model learning and tree learning algorithm. This algorithm is specially designed for speed and performance.

## DECISION TREE

Decision tree technique is used for classification and regression. The outcome will be in the form of tree, in which internal nodes are the features of tree, branches are the rules and then leaf is the outcome. It is a supervised machine learning algorithm.

## LOGISTIC REGRESSION

It is used to predict categorical dependent variable with the use of independent variables. Our project is helpful in prediction of rainfall like whether if it rains tomorrow or not. But the output can be a categorical value or discrete value like yes/no, 0 or 1.

## CHAPTER 3 LITERATURE SURVEY

In [1]. Venkatesan C, he used at “Multilayer Feed Forward Neutral Networks (MLFNN)”. During this paper he expected period monsoon precipitation. For predicting period monsoon he used “Error Back Propagation (EBP) algorithm”. He conjointly targeting 3 network models with 2, 3 and ten input parameters. As a conclusion he matched the output result to applied math models furthermore. In [2]. N. Philip, K. Joseph et al. They examined yearly precipitation foretelling in Kerala. They used ABF neutral network. In their analysis they found that ABFNN outperforms analysis.In [3]. Chattopadhyay S and Chattopadhyay M, They used 2 parameters. They think about “minimum temperature and most temperature”. In [4]. Chattopadhyaya S and Chattopadhyaya G did analyse on precipitation foretelling. For predicting the precipitation they used the “Conjugate Gradient good (CGD)” and “Levenberg– Marquardt (LM)” learning algorithms. As a conclusion for the prediction task, each algorithms performed equally well. In [5]. Htike K and Khalifa O, They used yearly, biannually, quarterly, and monthly precipitation information. For precipitation foretelling, they trained four distinct “Focused Time Delay Neural Networks (FTDNN)”. Whereas doing the analysis they found that once victimization yearly precipitation information for coaching, the FTDNN model provided the best prediction accuracy.

In [6]. Kannan S and Ghosh S contributed to the event of the “CART method”, that uses the “K- mean cluster technique” in conjunction with the choice tree algorithmic rule to get precipitation states from massive scale meteoric information in a very geographical region. The daily precipitation state is calculated victimization K-mean cluster victimization historical daily multi-site precipitation information. In [7]. Abhishek Kumar, adopt a synthetic neural network with learning via “back-propagation algorithm” configuration. Back-propagation is expected to be utilized in regarding

eightieth of all neural network comes in development. Within the learning cycle of a back-propagation algorithmic rule, there are 2 phases: one for propagating input patterns through the network and another for adapting the output by dynamical the weights within the network. In [8]. Sahai A K et al used the “Error Back Propagation Algorithm” to estimate India's Summer Monsoon precipitation on monthly and seasonal statistic. For precipitation prediction, they used information from the previous 5 years' monthly and seasonal mean precipitation values. In [9]. Somvanshi V K et al used AN “ANN model” to predict precipitation within the Hyderabad, Asian nation region. They conjointly compared the ANN methodology to the ARIMA methodology. They fed precipitation information from the previous four months into a neural network model. In [10]. Wu C et al. projected Asian nation and China's precipitation in their analysis paper. “Modular Artificial Neural Network was used (MANN)”. They conjointly compared the performance of MANN with LR, K-NN, and ANN.

In [11]. Kannan M et al forecasted short downfall in their paper. Associate empirical approach technique is applied for prediction. For a particular region, knowledge from 3 distinct months over the course of 5 years is analysed. Bunch is employed for grouping things along. In [12]. Kumar A et al. They mentioned average downfall within the Karnataka district of Udipi. They used “ANN models” for the downfall prediction challenge. Back Propagation rule (BPA) was found to be superior to layer perennial and cascaded back propagation. In [13]. Nanda S et al used an advanced applied mathematics model referred to as “ARIMA”, still as 3 ANN models referred to as “MLP, LPE (Legendre Polynomial Equation)”, and :FLANN (Functional-Link Artificial Neural Network)”. Whereas doing the analysis they found that, within the comparison to the ARIMA model, FLANN has superior forecast accuracy. In [14]. Thirumalai, Chandrasegar, et al evaluated the amount of downfall received in previous years supported crop seasons and forecast downfall for future years. “Rabi, Kharif, and Zaid square measure the 3 crop seasons”. For early prediction, the “linear regression” approach is employed. Rabi and kharif were used as variables during this study, and if one was provided, the opposite might be expected victimisation rectilinear regression.

The quality deviation and mean were additionally determined so as to forecast crop seasons within the future. Farmers can apply this implementation to work out that crops to reap supported agricultural seasons. In [15]. Parmar, Aakash, Kinjal Mistree, and Mithila Sompura, et al mentioned numerous ways used for downfall prediction for prediction.” various neural networks algorithms” for prediction square measure examined intimately, along side their steps, and categorises numerous methodologies and algorithms used for downfall prediction by numerous teachers within the era. Finally, the paper's conclusion is conferred. Background analysis on some machine learning models, as well as the ARIMA model, artificial neural networks, and types like Back-Propagation Neural Networks - Cascade Forward Back Propagation Network Layer perennial Network, Self-Organizing Map, and Support Vector Machine. The info was gathered, assessed, and a table was created to categorize the varied techniques to downfall prediction.

In [16]. For predicting monthly precipitation within the urban center region, G. Geetha and R. Selvaraj et al. used associate ANN model. Rainfall, most and minimum temperatures, and ratio were all factors evaluated by M. Sharma and J. Singh. Over the Pantnagar region, they projected weekly precipitation. ANN outperformed the multiple rectilinear regression model in terms of prediction accuracy. In [17]. Hourly precipitation was forecasted by Soo-Yeon JI et al. For prediction, CART and C4.5 area unit utilized, which can reveal hidden key patterns and their causes. The observation post provided eighteen variables. Validation is completed employing a 10-fold cross validation approach. CART was superior to C4.5 in terms of performance. In [18]. finger Saikia Dutta and Hitesh Tahbilder, et al used a regular applied math technique known as Multiple rectilinear regression to estimate monthly precipitation in state. The model's parameters embody minimum and most temperatures, mean water level pressure, wind speed, and precipitation. Acceptable accuracy is given by prediction model supported multiple rectilinear regression.

In [19]. A model developed by Geetha, A., and G. M. Nasira, et al forecasts meteorologic things like precipitation, fog, thunderstorms, and cyclones, permitting people to require precautional precautions. The choice trees were modelled exploitation data processing techniques and an information mining programme known as fast manual Labourer. Trivandrum knowledge set together with parameters like day, temperature, dew point, pressure, and so on. S call tree formula is employed to partition the dataset into coaching and testing sets. The exactitude is set, and also the actual and projected values area unit compared. The exactitude is eighty.67, and it may be improved by exploitation soft computing approaches like mathematical logic and organic process algorithms. In [20]. in line with Singh, Gurpreet, and Deepak Kumar et al., several machine learning algorithms area unit used for precipitation prediction, and that they used a hybrid approach that mixes 2 techniques, Random forest and Gradient boosting, with several machine learning techniques like ADA boost, K-Nearest Neighbor(KNN), Support vector machine(SVM), and Neural Networks (NN). These were applied to North geographic region precipitation knowledge from 2007 to 2017, and also the performance was calculated exploitation many measures like F-score, precision, accuracy, and recall.Finally, eight hybrid models were conferred, with Gradient boosting-Ada boost rising because the superior model with promising results.

## CHAPTER 4 REQUIREMENTS ENGINEERING

* 1. **SOFTWARE REQUIREMENTS**

Used Jupyter notebook and Anaconda Flask to visualize our content and various Machine Learning algorithms.

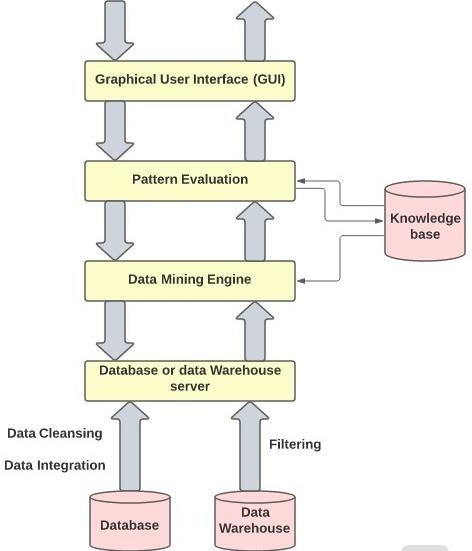
**PROGRAMMING LANGUAGES USED**

## Python 3.8

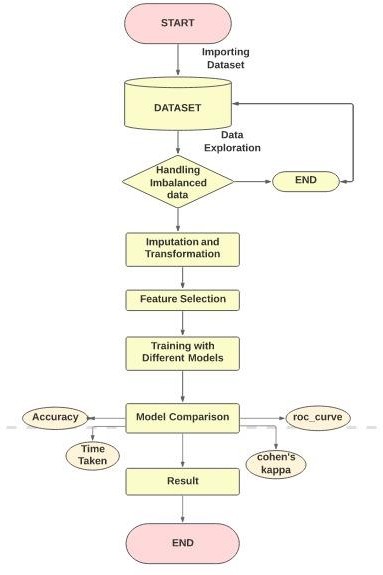
* 1. **HARDWARE REQUIREMENTS**
     + **OS**: windows.
     + **Monitor:** 15” color
     + **Processor:** Intel Core i5 for efficient use
     + **RAM** — 8GB min, 16GB max RAM.
     + **Hard disk**: 50GB

## CHAPTER 5 DETAILED DESIGN

## SYSTEM ARCHITECTURE



## FLOWCHART



**CHAPTER 6 IMPLEMENTATION**

## IMPLEMENTATION

**BASIC NEEDS**

## CREATE FOLDER

Create folder with any name and any location in the system environment.

* + 1. **INSTALL PYTHON AND PIP MODULES**

Make sure to install current version of python and import all modules of pip.

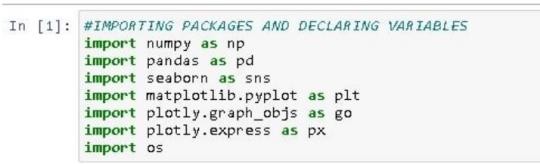
* + 1. **IMPORT DATASET FILE INTO FOLDER AND CREATE NEW FILE**

Import the dataset (which is taken from Kaggle repository) to the folder already created by using Jupyter notebook. And also create the new file with any name of extension of

.ipynb.

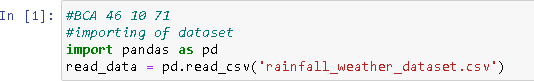
## IMPORTING PACKAGES

We are importing some of the packages that we will need to implement in our project, so we will need numpy to perform mathematical operations, pandas for data analysis, seaborn for data visualisation, matplotlib for graphical plotting, OS for system libraries, and plotty for graphical plotting.



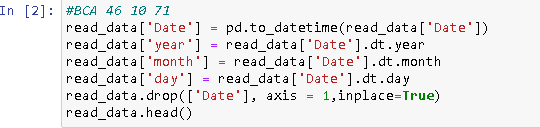
## IMPORTING DATASET

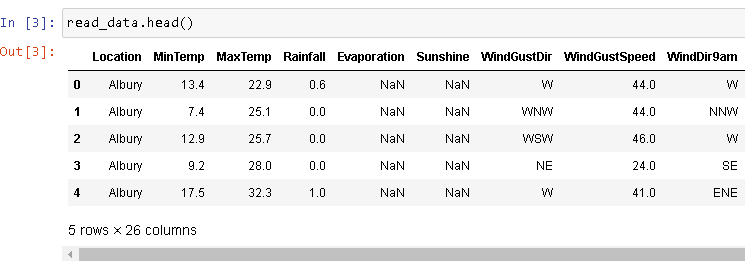
In this case, we are reading the dataset to complete the implementation process. We are retrieving the dataset from the location where the file (.csv) is stored using Python code.



## DATA\_EXPLORE:

Here, we are viewing the dataset with the help of head function.



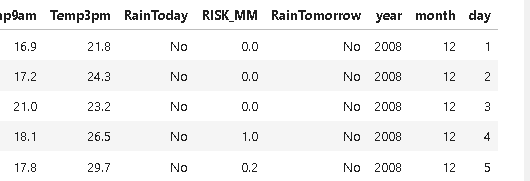


With the help of shape command, We found that there are 142193 row and 26 column of datum in the imported dataset.



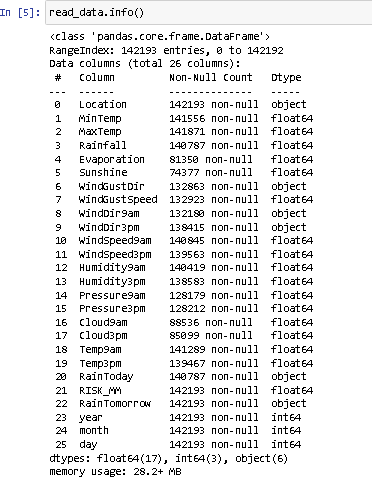
To make all the data to a float value because many of the attributes are in float and also it is easy to keep the rainfall prediction in float value because of its point accuracy measure of rainfall. So first we are splitting the date attribute to day, month and year separately for our use.





**DATASET DESCRIPTION**

Here, fetching out the total number of attributes and its datatype by using the info method.



## REPLACING VALUES

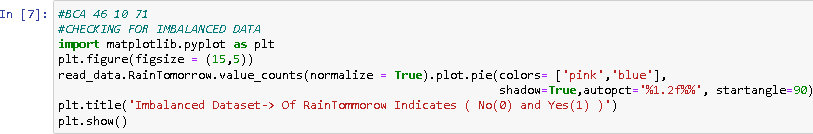
Here, we are taking into account two factors, "RainToday" and "RainTomorrow," and replacing the categorical type with a numerical type.

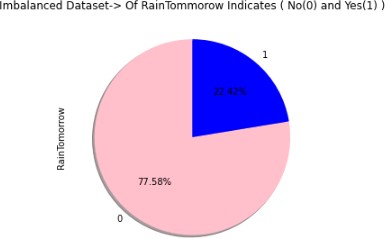
We replaced YES with 1 and NO with 0.



## BALANCING AN IMBALANCED DATA

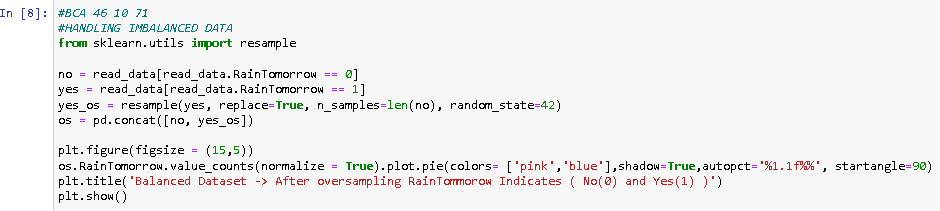
We are getting RainTomorrow as a "TARGET VARIBLE" and visualising it with a "pie chart" here.

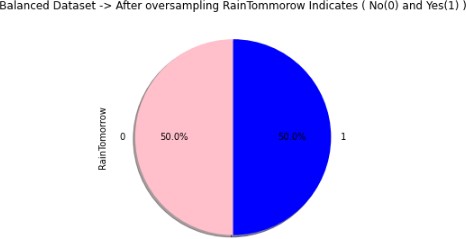




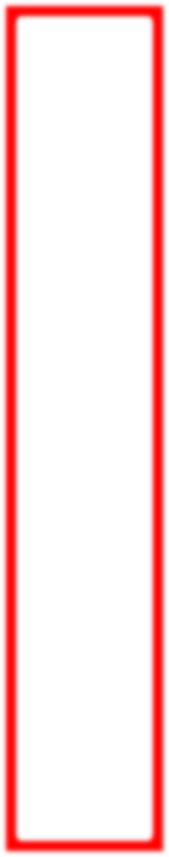
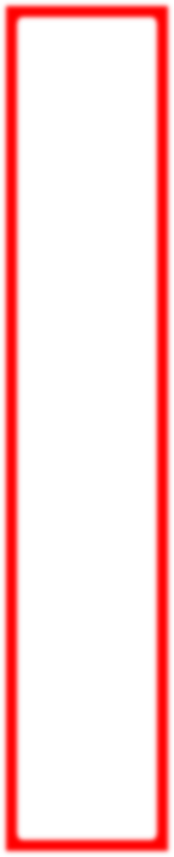
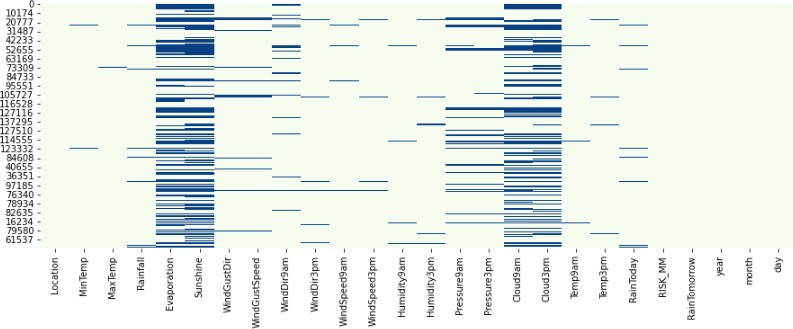
Based on the preceding example, we can now easily comprehend the visual and conclude that the dataset is typically imbalanced.

Using the "over sampling of the minority class" technique, we increased the YES (22.42%) to 50%, and both were the same.

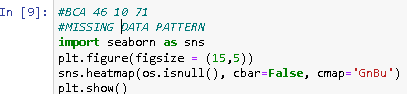




## MISSING VALUES

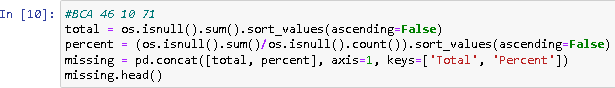


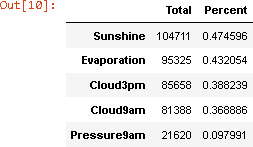
In this case, we are attempting to retrieve the misplaced values from the data-set. We going to use the "seaborn" for this. Attributes such as "evaporation, sunshine, cloud 3 am & 3pm" have the most missing data in the "heatmap" visual.



## TOTAL % OF MISSING DATA

We currently working on recovering a missing pattern of data in math imputation contours. As a result, basic imputations like "total, count, and sum" are employed here.

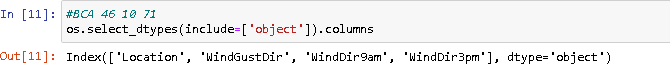




All of the features listed below have a percentage of greater than 50%, as we can clearly see. So, before we get rid of them entirely, we can easily convert them to proper imputations.

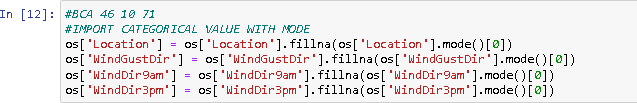
## FETCHING OBJECT DATATYPE

We tracing the non-numerical data from my dataset. We discovered that "Location, WindGustDir, WindDir9am, WindDir3pm" are categorical data.

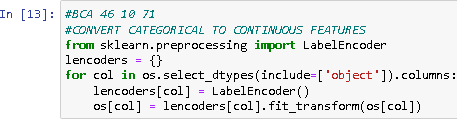


## CONVERTING CATEGORICAL DATA TO NUMERICAL DATA

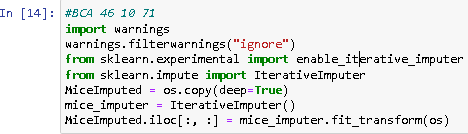
From the preceding, we derived the category data. It is now time to translate them into a numerical process that a machine can comprehend. So we are going to use mode in conjunction with the attributes here. So we will be able to get zeros(0) in that spot.



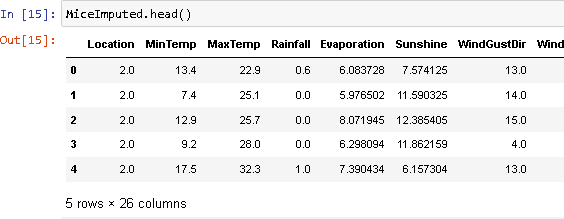
Now by here we are using label coder to convert them into numeric once.



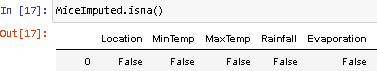
Here we can easily see that all the categorical columns has been replaced by the numeric ones.



## DATA EXPLORE



Here, we are looking for "null values" or "missing values." The image below clearly shows that there are no null values in the dataset for this process.

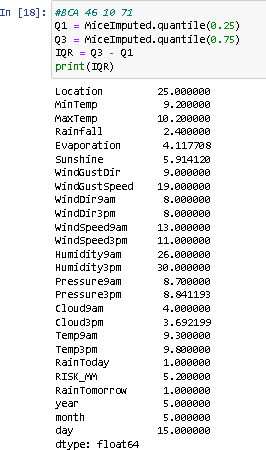


## DETECTING AND EXCLUDING OUTLIERS

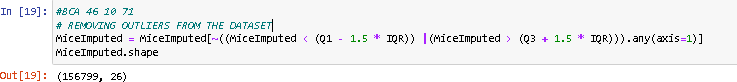
We are looking for "outliers," which means nothing in this context, but I can exclude data from the dataset that is outside the range bounds or outside the class boundary.

“IQR=Q3-Q1”

“Outliers=1.5 \* IQR”



We are here excluding outliers using math imputations in this case, and after removing them, my dataset contains 156703 data points.

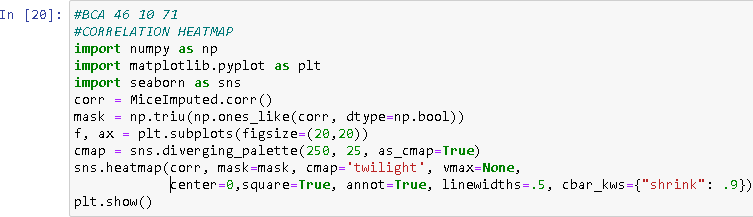


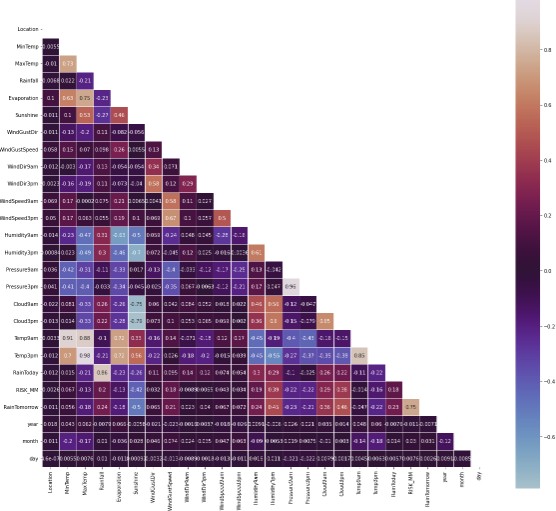
## CORRELATION CHECKING

After the outliers have been removed, the new correlation is displayed. The following characteristics are highly correlated.

* “MaxTemp and MinTemp”
* “Pressure9am and Pressure3pm”
* “Temp9am and Temp3pm”
* “Evaporation and MaxTemp”
* “MaxTemp and Temp3pm”

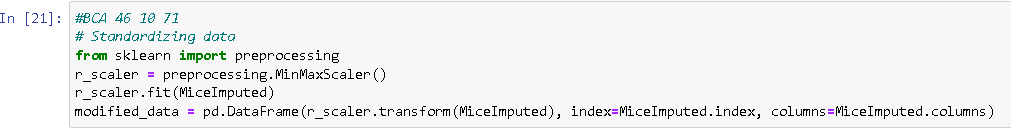
But in “NO” case, the correlation value is equal to a perfect "1". So we are not discarding any feature.





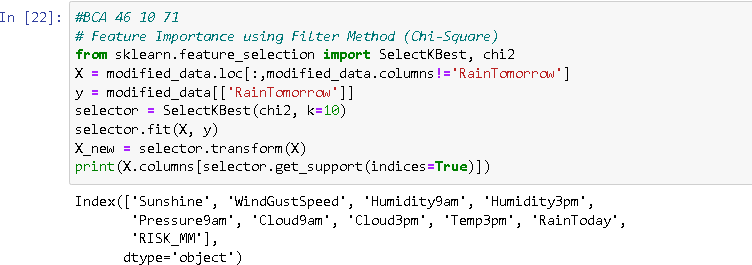
## PREPROCESSING

The term "pre-processing" refers to the process of standardising data. We also used the "MinMaxScaler" by importing the "preprocessing," and the modified data is saved in the new variable. And we will be using the modified (new) variable from now on.



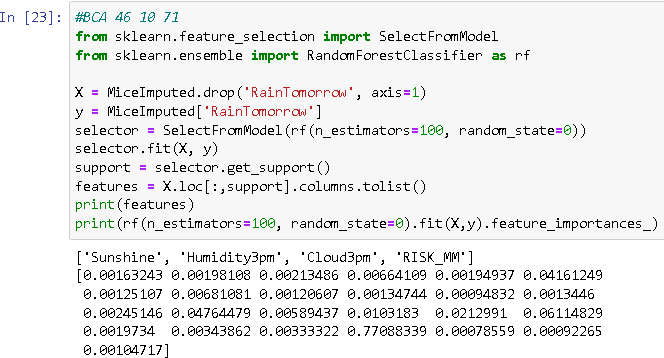
## FEATURE SELECTION

Here we used the “filter method” which uses the “chi-square method”.



Here we used the “Wrapper method” which uses the “random forest classifier”. By adding all the values in the below image

We are excluding the “RISK\_MM” factor which is near to one. And according to the “dataset publisher” that factors will be result in leakage of info.



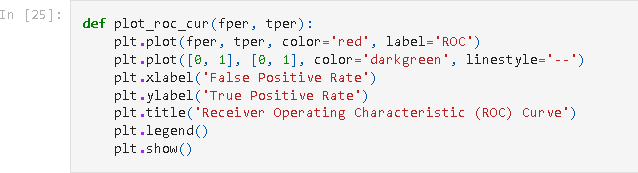
## TRAINING AND TESTING

The dataset will divide into two one is training and another is testing. So we have taken 75% for training and 25% for testing to compare the results in the basis of time taken, accuracy, roc\_curve, cohen’s kappa.



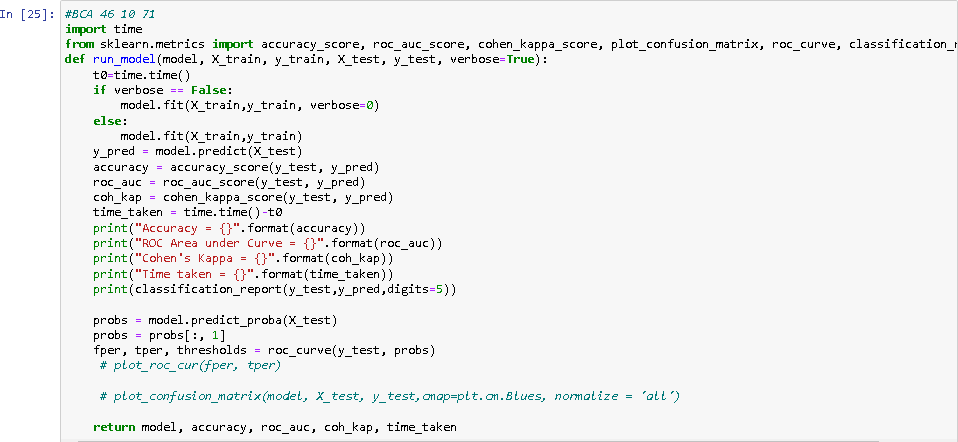
## PLOTTING ROC\_CURVE FOR THE MODELS

To visualize the roc\_curve for our model the below code is introduced.

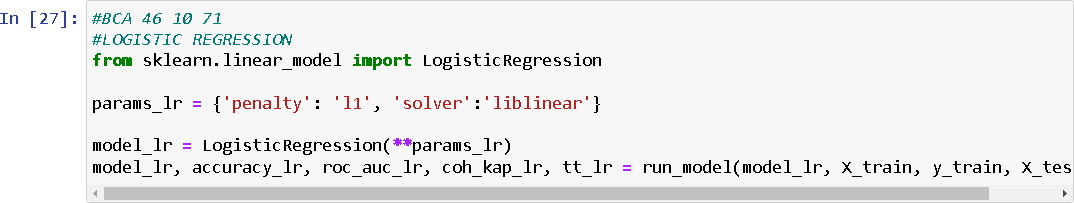


## WORKING WITH DIFFERENT MODELS

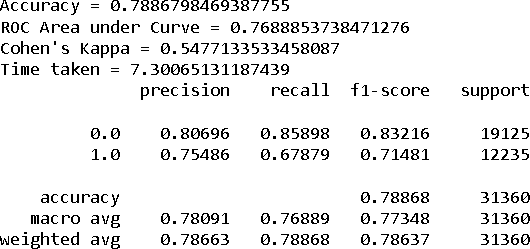
The models used by us are logistic regression, random forest, descision tree, XGboost.



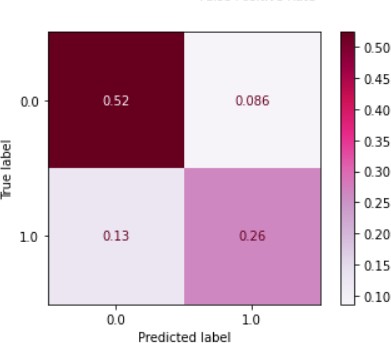
## LOGISTIC REGRESSION



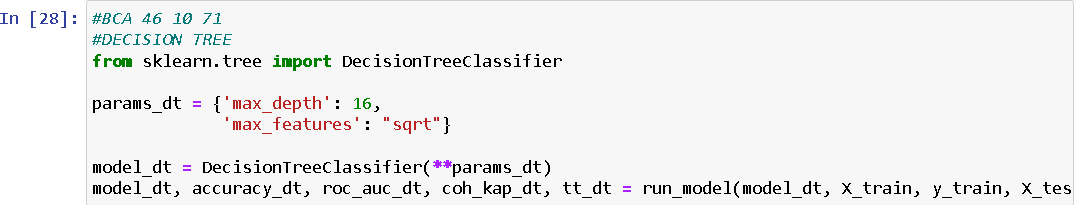
**Output:**



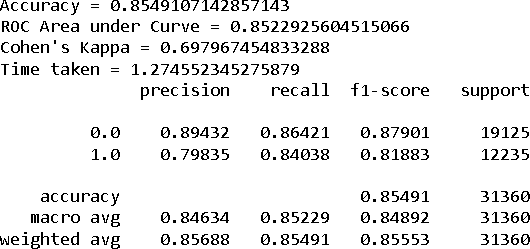
## Confusion matrix



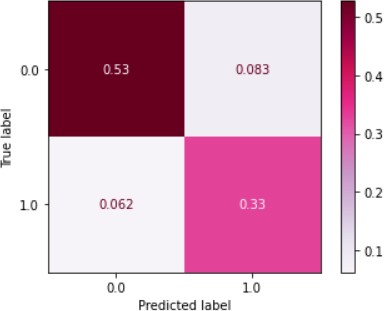
**DECISION TREE**



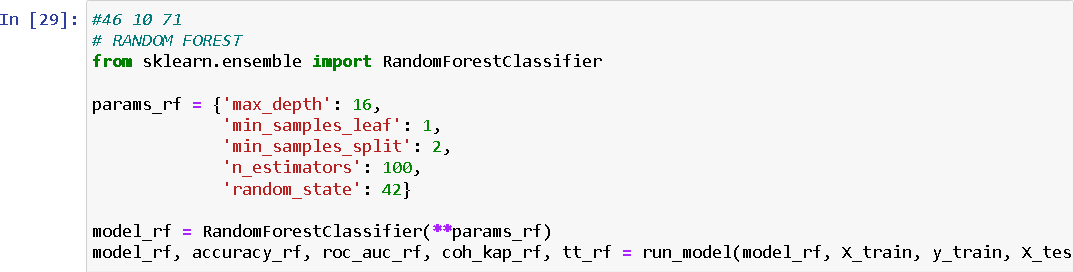
**Output**



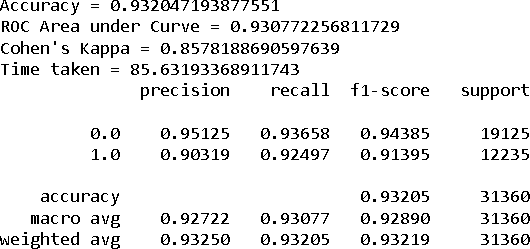
**Confusion matrix**



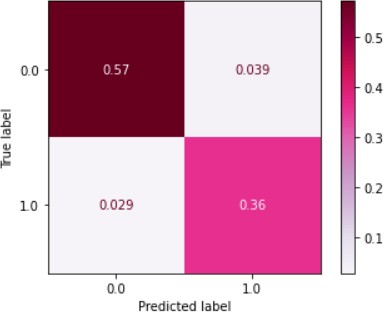
## RANDOM FOREST



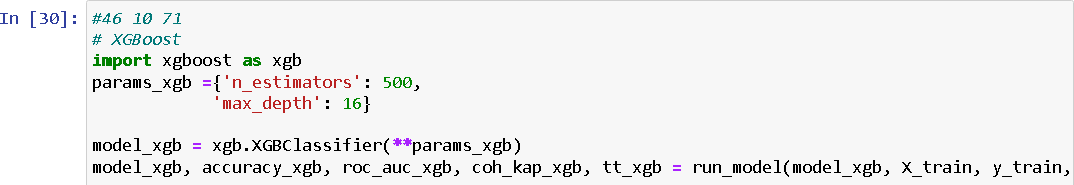
**Output**



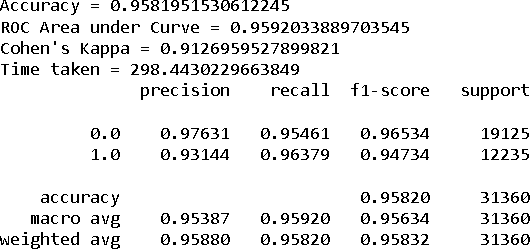
**Confusion matrix**



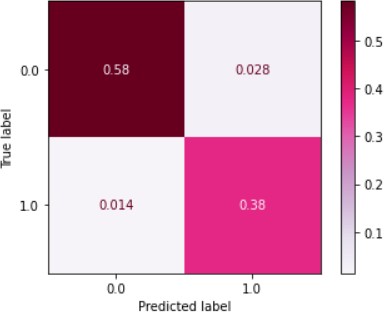
**XG-BOOST**



**Output**

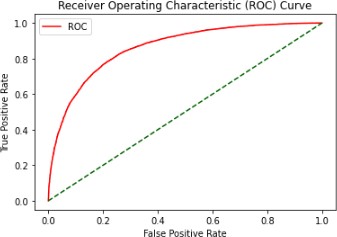


**Confusion matrix**

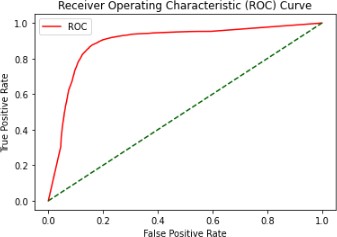


# Roc\_curve

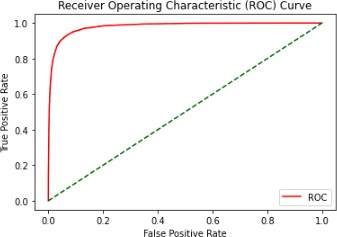
* Logistc Regression



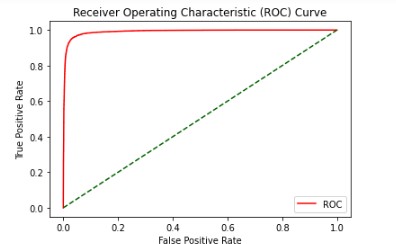
* Decision Tree



* Random forest

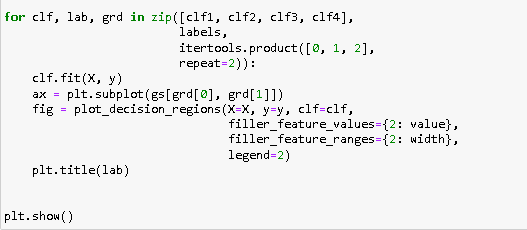


* XG-Boost

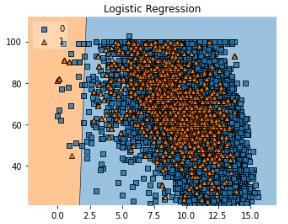


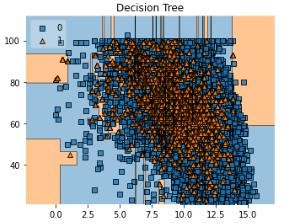
**PLOT DECISION REGIONS CODE**

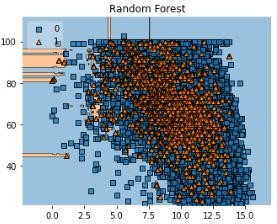


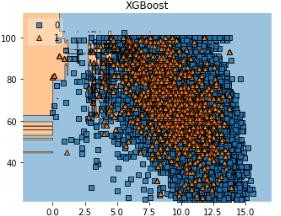


The output for plotting the decision regions is given below however it is considered based on the training dataset. We can observe based on the regional borders.





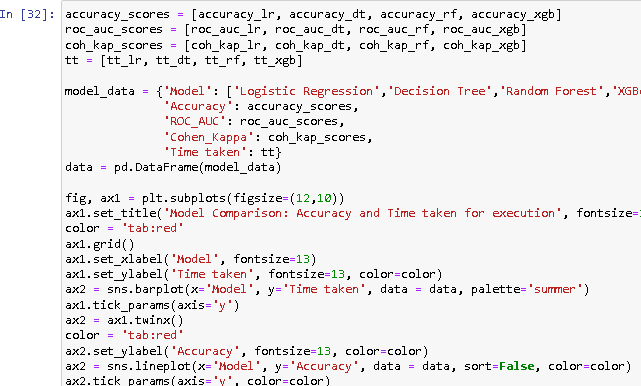


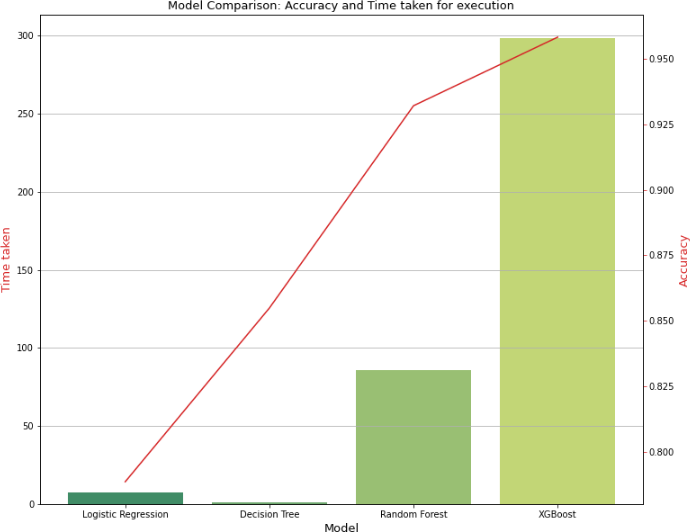


As compared to all other models it seems that random forest and xgboost has little number of mis-classified data compared to other.

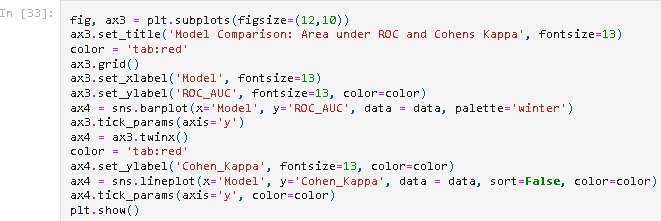
## COMPARISON OF ALGORITHMS

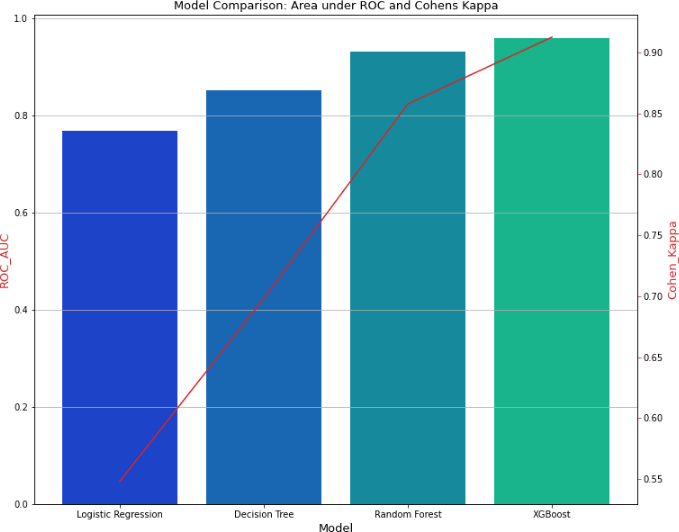
Comparing models with the accuracy and time taken:





Comparing models with the cohen’s kappa and roc\_curve:





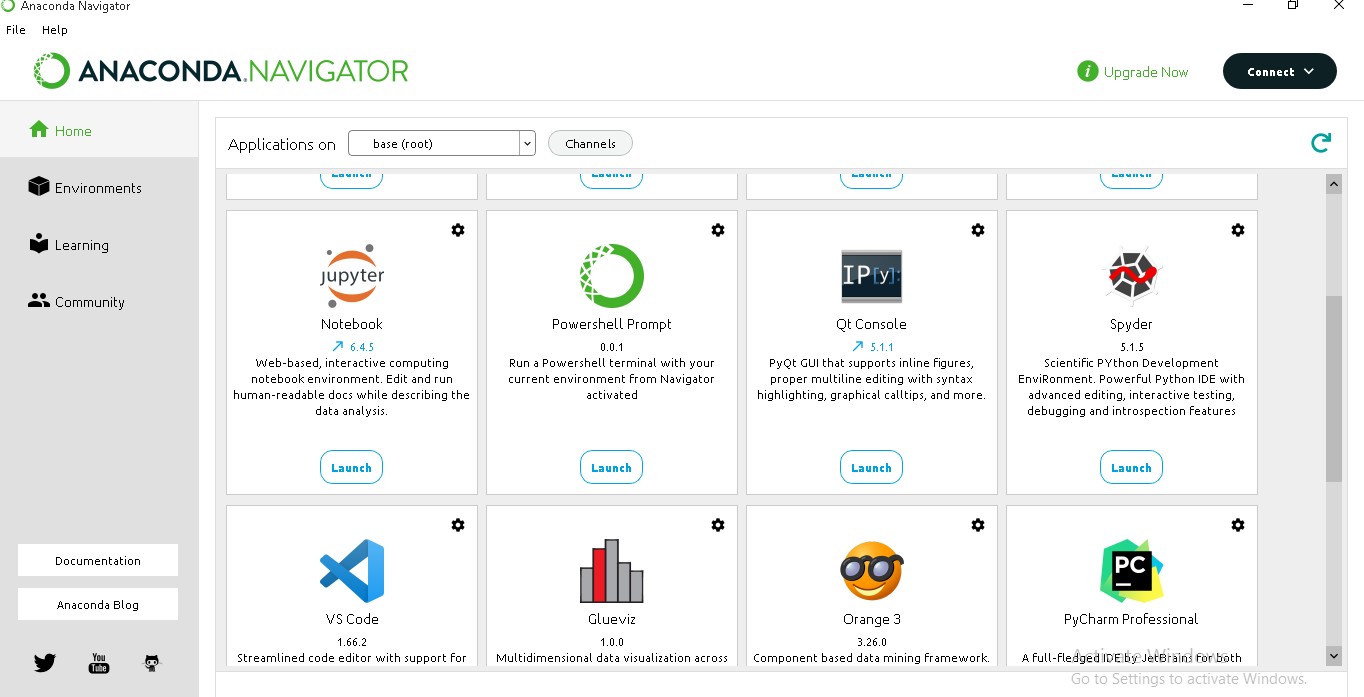
### EXPERIMENTAL RESULT

As compared to all we found that the random forest has less time taken compared to all and also XG-boost shows high accuracy keeping all this then result will be concluded. As compared to time taken random forest will best suits hence the web application is create by following below steps using anaconda has a framework.

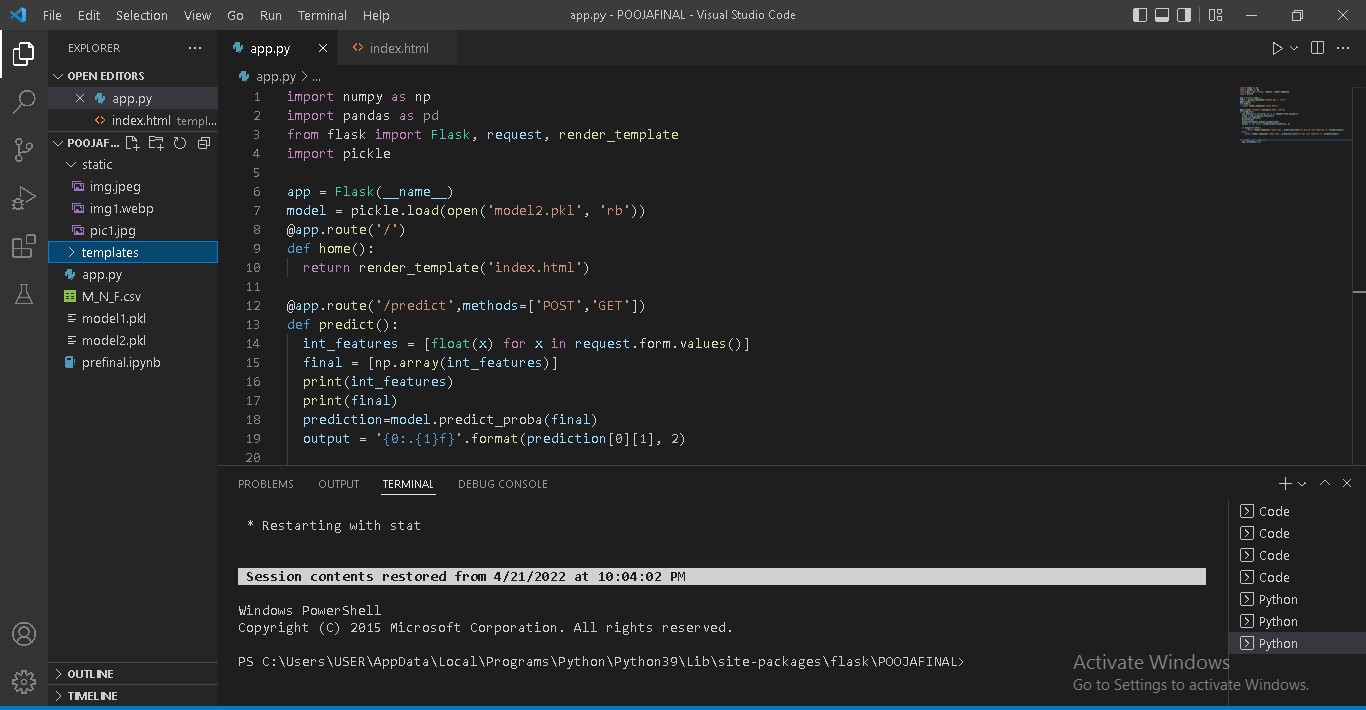
## WEB APP IMPLEMENTATION

**FRAME WORK SCREENSHOT**

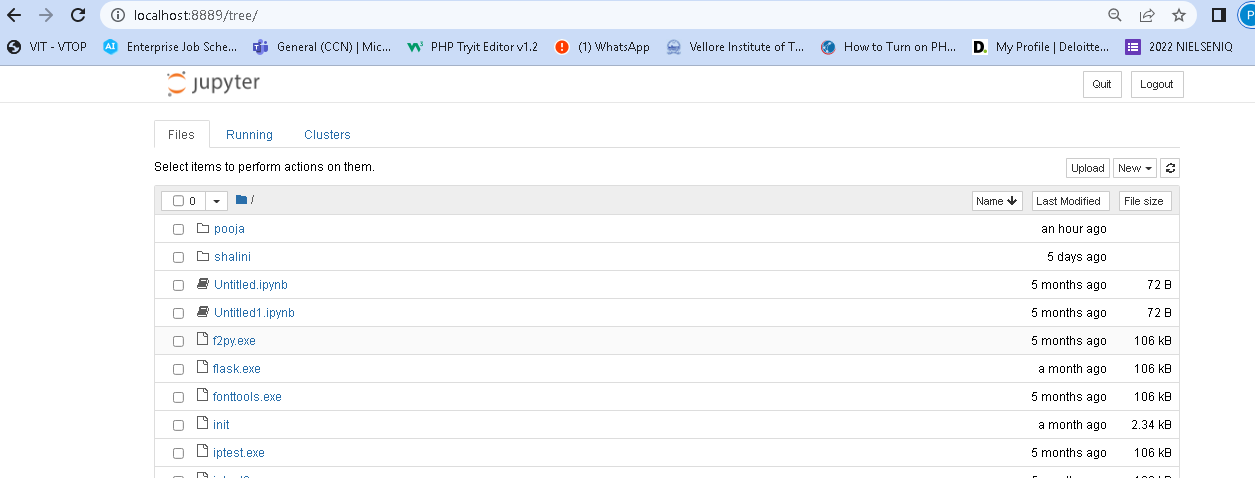
For this project, we used Anaconda framework, in that we used Jupyter Notebook for model implementation, Flask for web implementation, vs code for implementation of html and python for our website.



VS Launch



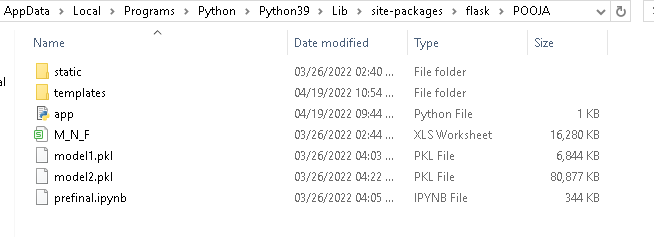
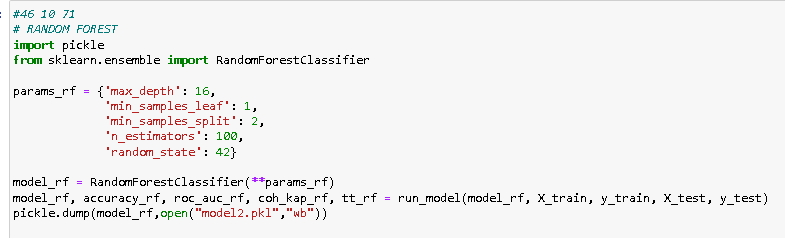
Jupyter Notebook



## WEB APP IMPLEMENTATION

Here the Anaconda framework is used to provide a GUI. In that VStudio acts as a code editor for implementation. We used flask, a python web framework to create a web application, where html is used to design the website and python file is used to get the details and attributes of our data’s so that if the user enters the value in the web it will predict the required solution. Pickle used for serialize our model for that pickle dump() method is used to train our algorithms like ( ML algorithms etc..). In our project we used Random Forest, because it provides good accuracy comparing with other algorithms. In order to predict the new entry value result we used Random Forest algorithm which is stored in the pickle file for prediction. So here we created a web page in the form format using html and collected 21 attributes values has an integer value for predicting that there has a symptoms for rainfall or not. The trained algorithm will find the result and provide the result in the same web page in string format.

**IMPORTING PICKLE FILE**



CODE

app.py

import numpy as np import pandas as pd

from flask import Flask, request, render\_template import pickle

app = Flask( name )

model = pickle.load(open('model2.pkl', 'rb')) @app.route('/')

def home():

return render\_template('index.html')

@app.route('/predict',methods=['POST','GET']) def predict():

int\_features = [float(x) for x in request.form.values()] final = [np.array(int\_features)]

print(int\_features) print(final)

prediction=model.predict\_proba(final)

output = '{0:.{1}f}'.format(prediction[0][1], 2)

if output>str(0.5):

return render\_template('index.html', prediction\_text='It has no rain tomorrow

{}'.format(output)) else:

return render\_template('index.html', prediction\_text='It has rain tomorrow

{}'.format(output))

if name\_\_ == "\_\_main ": app.run(debug=True)

index.html

<!--GUI for Breast Cancer Detection Application using SVM-->

<!DOCTYPE html>

<html>

<head>

<meta charset="UTF-8">

<!-- Bootstrap CSS -->

<link rel="stylesheet" href="https://stackpath.bootstrapcdn.com/bootstrap/4.3.1/css/bootstrap.min.css" integrity="sha384- ggOyR0iXCbMQv3Xipma34MD+dH/1fQ784/j6cY/iJTQUOhcWr7x9JvoRxT2MZw1 T" crossorigin="anonymous">

<title>Rainfall Prediction</title>

<style>

/\*just bg and body style\*/ body {

margin: 40px;

background-color: #808080;

background-image: linear-gradient(315deg, #de5499 19%, #a0c5ba 85%);

}

.container {

border-radius: 3000px; text-align: center;

}

.btn-container { background: white;

box-shadow: 0 19px 38px rgba(0, 0, 0, 0.30), 0 15px 12px rgba(0, 0, 0, 0.22); border-radius: 300px;

padding: 100px;

}

.head {

font-weight: bolder;

}

.btn-primary { color: #ffffff;

text-shadow: 0 -1px 0 rgba(0, 0, 0, 0.25); background-color: #de5499 !important; border-color: #de5499 !important; padding: 10px;

}

label { width: 42%;

}

#predict { display: none;

}

.form-group { padding: 2px;

}

</style>

<!--Font Awesome-->

<script src="https://kit.fontawesome.com/a076d05399.js"></script>

<link rel="stylesheet" href="https://stackpath.bootstrapcdn.com/font- awesome/4.7.0/css/font-awesome.min.css" integrity="sha384- wvfXpqpZZVQGK6TAh5PVlGOfQNHSoD2xbE+QkPxCAFlNEevoEH3Sl0sibVcO QVnN" crossorigin="anonymous">

<script src="https://ajax.googleapis.com/ajax/libs/jquery/3.5.1/jquery.min.js"></script>

<!-- jQuery first, then Popper.js, then Bootstrap JS -->

<script src="https://cdnjs.cloudflare.com/ajax/libs/jquery/3.5.1/jquery.min.js" integrity="sha512- bLT0Qm9VnAYZDflyKcBaQ2gg0hSYNQrJ8RilYldYQ1FxQYoCLtUjuuRuZo+fjqh x/qtq/1itJ0C2ejDxltZVFg==" crossorigin="anonymous"></script>

<script src="https://code.jquery.com/jquery-3.5.1.slim.min.js" integrity="sha384- DfXdz2htPH0lsSSs5nCTpuj/zy4C+OGpamoFVy38MVBnE+IbbVYUew+OrCXaRkf j" crossorigin="anonymous"></script>

<script src=["https://cdn.jsdelivr.net/npm/popper.js@1.16.0/dist/um](https://cdn.jsdelivr.net/npm/popper.js%401.16.0/dist/umd/popper.min.js)d[/popper.min.js"](https://cdn.jsdelivr.net/npm/popper.js%401.16.0/dist/umd/popper.min.js) integrity="sha384- Q6E9RHvbIyZFJoft+2mJbHaEWldlvI9IOYy5n3zV9zzTtmI3UksdQRVvoxMfooAo" crossorigin="anonymous"></script>

<script src="https://stackpath.bootstrapcdn.com/bootstrap/4.5.0/js/bootstrap.min.js" integrity="sha384- OgVRvuATP1z7JjHLkuOU7Xw704+h835Lr+6QL9UvYjZE3Ipu6Tp75j7Bh/kR0JKI

" crossorigin="anonymous"></script>

</head>

<body>

<div class="container">

<div class="row">

<div class="col-md-12">

<h1 class="head">Rainfall Prediction</h1>

</div>

</div>

<div class="row">

<div class="col-md-12">

<div class="btn-container">

<!-- Main Input For Receiving Query to our ML -->

<form action="{{ url\_for('predict')}}" method="post" class="form-inline">

<div class="row">

<div class="col-md-6">

<br/></br>

<img class="image" src="{{url\_for('static', filename='pic1.jpg')}}" width="112%" height="86%" />

</div>

<div class="col-md-6">

<div class="container">

<h4>Enter Cell Details</h4>

<div class="form-group">

<label for="Location">Location </label>

<input type="text" class="form-control" name="Location" required="required">

</div>

<div class="form-group">

<label for="MinTemp">MinTemp </label>

<input type="text" class="form-control" name="MinTemp" required="required">

</div>

<div class="form-group">

<label for="MaxTemp">MaxTemp </label>

<input type="text" class="form-control" name="MaxTemp" required="required">

</div>

<div class="form-group">

<label for="Rainfall">Rainfall</label>

<input type="text" class="form-control" name="Rainfall" required="required">

</div>

<div class="form-group">

<label for="Evaporation">Evaporation</label>

<input type="text" class="form-control" name="Evaporation" required="required" />

</div>

<div class="form-group">

<label for="Sunshine">Sunshine</label>

<input type="text" class="form-control" name="Sunshine" required="required" />

</div>

<div class="form-group">

<label for="WindGustDir">WindGustDir</label>

<input type="text" class="form-control" name="WindGustDir" required="required" />

</div>

<div class="form-group">

<label for="WindGustSpeed">WindGustSpeed</label>

<input type="text" class="form-control" name="WindGustSpeed" required="required" />

</div>

<div class="form-group">

<label for="WindDir9am">WindDir9am</label>

<input type="text" class="form-control" name="WindDir9am" required="required" />

</div>

<div class="form-group">

<label for="WindDir3pm">WindDir3pm</label>

<input type="text" class="form-control" name="WindDir3pm" required="required" />

</div>

<div class="form-group">

<label for="WindSpeed9am">WindSpeed9am</label>

<input type="text" class="form-control" name="WindSpeed9am" required="required" />

</div>

<div class="form-group">

<label for="WindSpeed3pm">WindSpeed3pm</label>

<input type="text" class="form-control" name="WindSpeed3pm" required="required" />

</div>

<div class="form-group">

<label for="Humidity9am">Humidity9am</label>

<input type="text" class="form-control" name="Humidity9am" required="required" />

</div>

<div class="form-group">

<label for="Humidity3pm">Humidity3pm</label>

<input type="text" class="form-control" name="Humidity3pm" required="required" />

</div>

<div class="form-group">

<label for="Pressure9am">Pressure9am</label>

<input type="text" class="form-control" name="Pressure9am" required="required" />

</div>

<div class="form-group">

<label for="Pressure3pm">Pressure3pm </label>

<input type="text" class="form-control" name="Pressure3pm" required="required">

</div>

<div class="form-group">

<label for="Cloud3pm">Cloud3pm </label>

<input type="text" class="form-control" name="Cloud3pm" required="required">

</div>

<div class="form-group">

<label for="Temp9am">Temp9am </label>

<input type="text" class="form-control" name="Temp9am" required="required">

</div>

<div class="form-group">

<label for="Temp3pm">Temp3pm </label>

<input type="text" class="form-control" name="Temp3pm" required="required">

</div>

<div class="form-group">

<label for="RainToday">RainToday </label>

<input type="text" class="form-control" name="RainToday" required="required">

</div>

<div class="form-group">

<label for="RISK\_MM">RISK\_MM </label>

<input type="text" class="form-control" name="RISK\_MM" required="required">

</div>

</br>

<button type="submit" class="btn btn-primary btn-lg">Predict Rainfall</button>

</div>

</div>

</div>

</form>

<br />

<center>

<h1 style="background:#de5499">{{prediction\_text}}</h1>

</center>

<br />

</body>

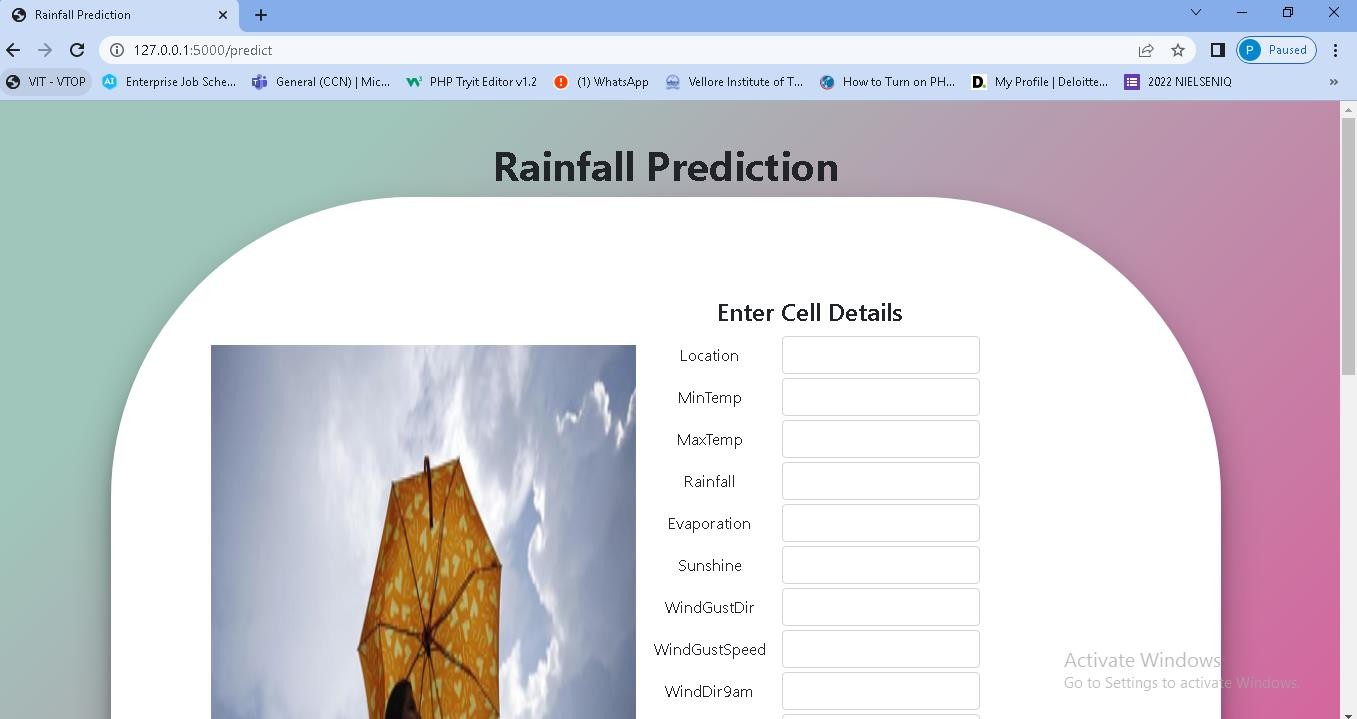
</html>

## RESULT OF THE PROJECT

As a result of the project, Random Forest is trained for result prediction. So here by getting the 21 attributes new value the trained model provides the test result as it will rain tomorrow or not. This is done with the help of the trained algorithm. Here we are not directly training the algorithm, the algorithm is trained and serialized has the pickle file. So in the python code of our project, I am calling the pickle file and assign the test value to it for prediction. Based on the trained algorithm result it will also predict the test result and provide the result for our assigned values also. Then with help of the python code, I am displaying the result in the web page.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **MODEL** | **TIME TAKEN** | **ACCURACY** | **COHENS KAPPA** | **ROC** |
| **Logistic Regression** | 7.3 | 78 | 54 | 76 |
| **Decision Tree** | **1.27** | 85 | 69 | 85 |
| **Random Forest** | 93 | 93 | 85 | 93 |
| **XGBoost** | 298 | **95** | **91** | **95** |

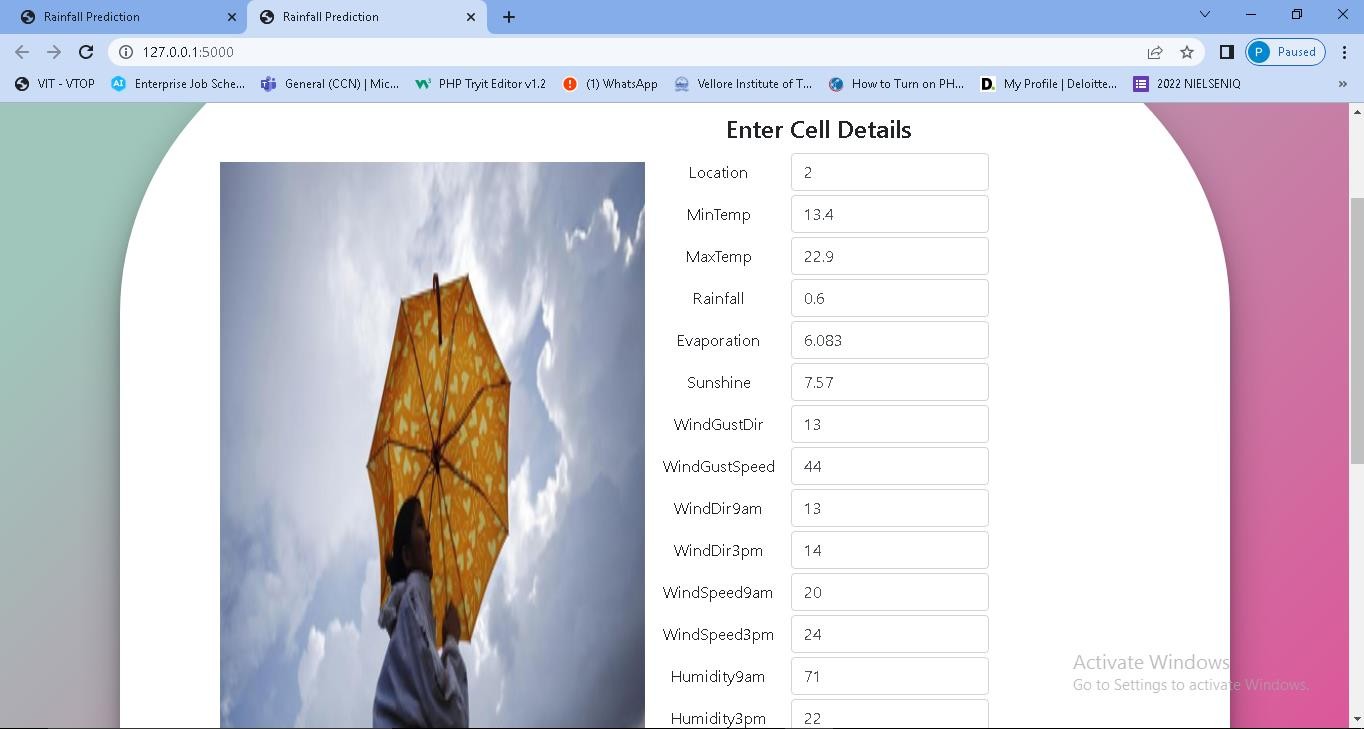
# WEB PAGE

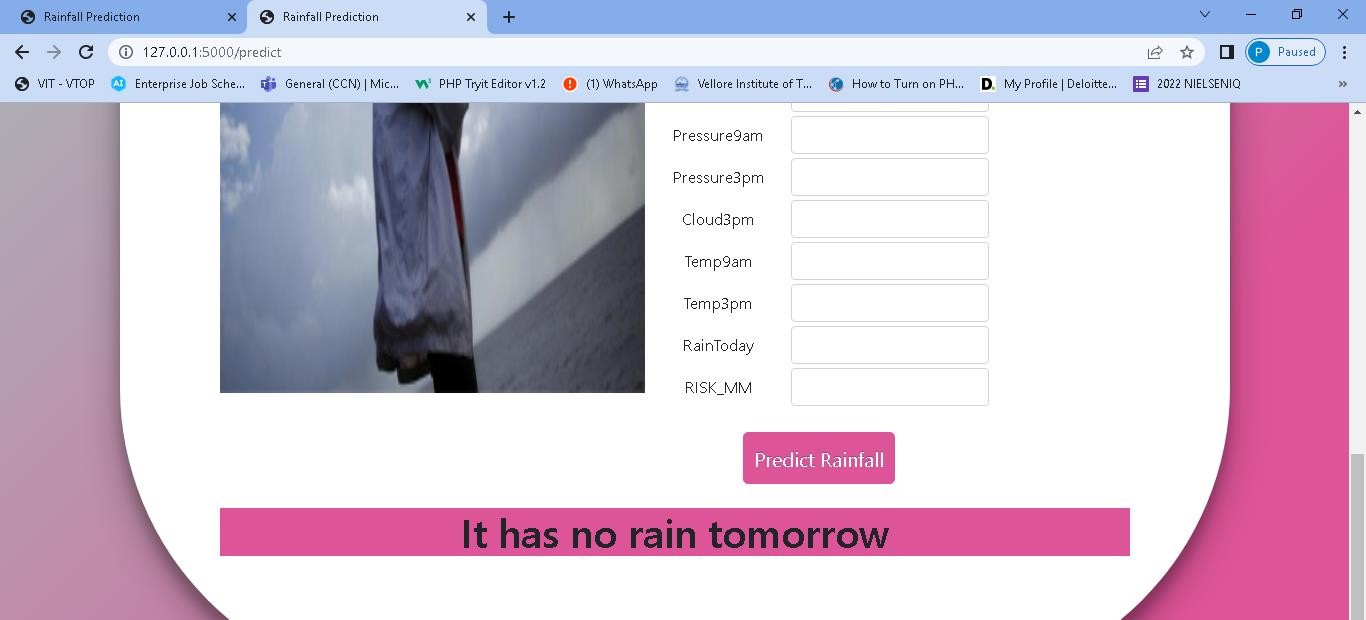


Result for It will not rain Tomorrow

Values: 2, 13.4, 22 .9, 0.6, 6.083, 7.57, 13, 44, 13, 14, 20, 24, 71, 22,

1007.7, 1007.1, 4.78, 16.9, 21.8, 0, 0.

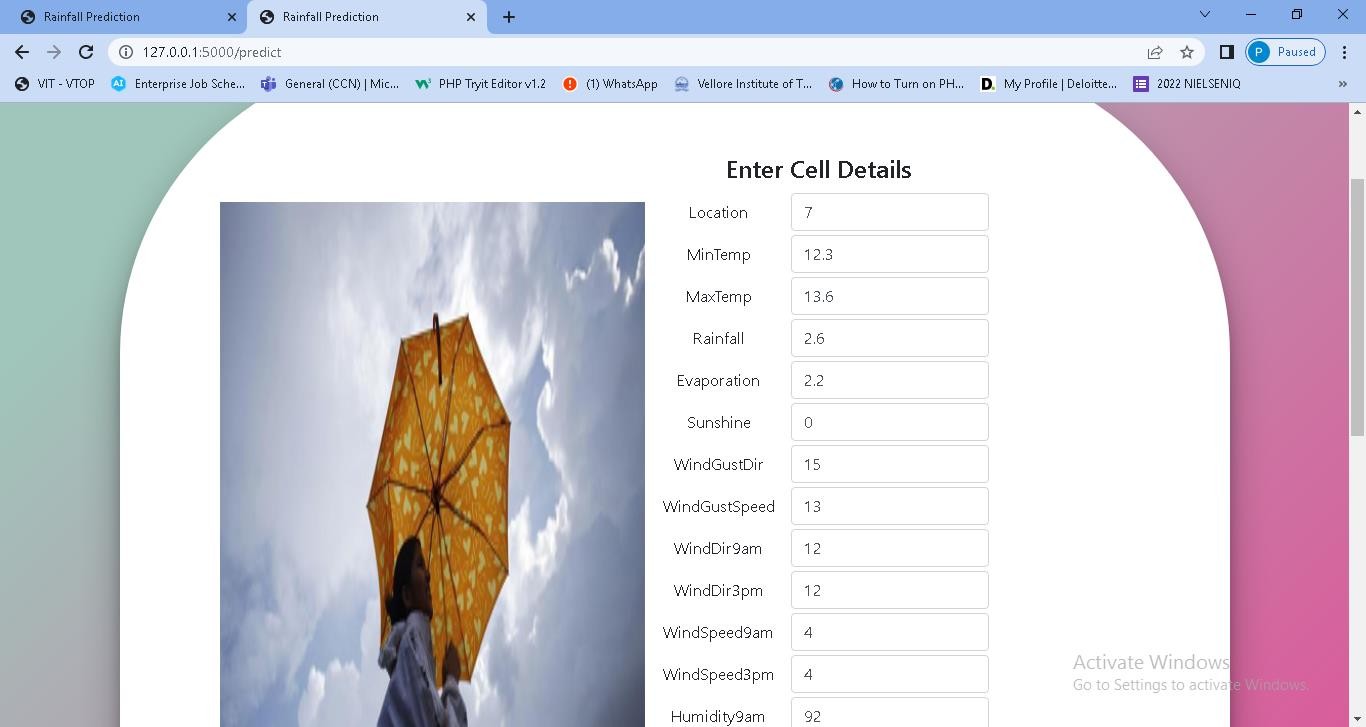


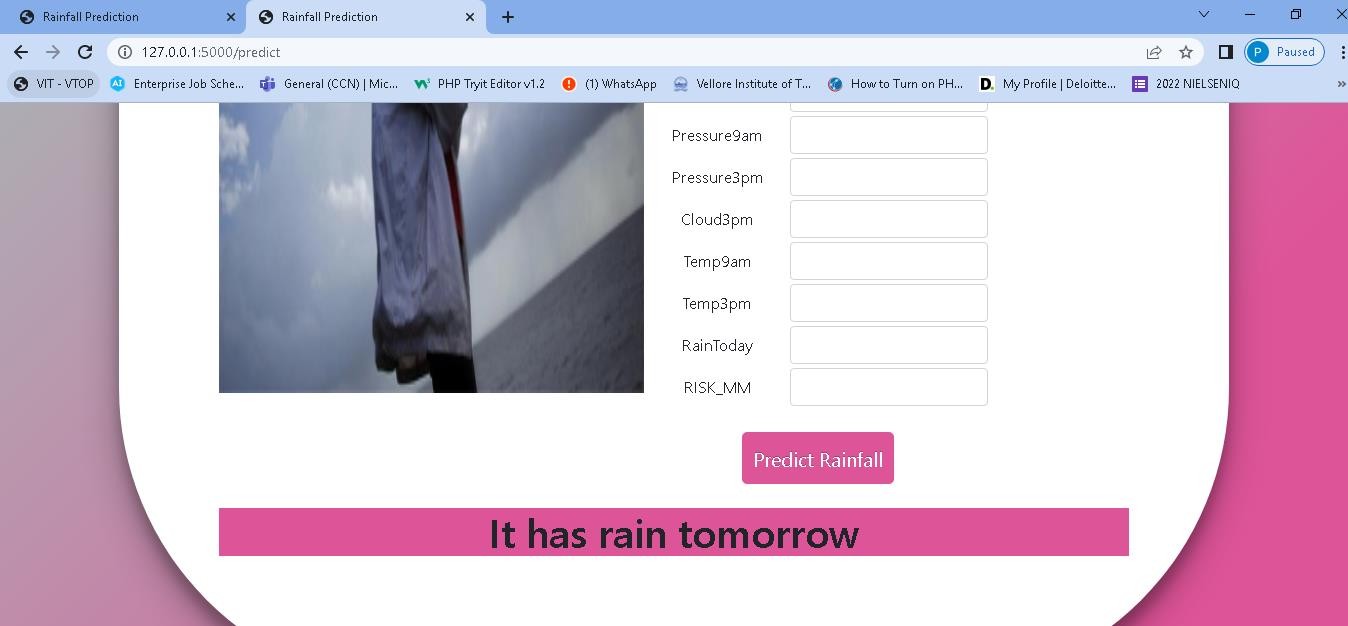


Result for It will rain Tomorrow

Values: 7, 12.3, 13.6, 2.6, 2.2, 0, 15, 13, 12, 12, 4, 4, 92, 90, 1019.8,

1017.3, 8, 12.7, 13.4, 1, 6.





* 1. **CONCLUSION**

# CHAPTER 7

The result of this rainfall prediction model is based upon the comparison of the proposed algorithms, which is going to provide the better solution on the consideration of the above aspects, **Random forest** provides better speed and accuracy for execution in- place of cat-boost and XG-boost.

Not only from considering above table, but also this can be accomplish by making combinations of the above factors and also the classification report of each predicted models (the precision, the recall, the f1-score and the support) as concluded.

## FUTURE WORK

As this dataset contains huge amount of data in it, we can use deep learning models and make comparison among machine learning vs deep learning (or) it can be incorporate with other trending web apps or native applications like flutter to make it more ease from user end. Also we can compare more algorithms in order to check which algorithm is best suit for rainfall prediction. Also we can use another platform for the creation and deployment of the web application. In this project I used Anaconda platform for implementation. In my project mostly I am using classification type of supervised learning algorithms, further we can also implement with unsupervised learning also. We can also use advanced ML type of algorithms or deep learning or also we can use neural networks, fuzzy logics for further implementation and for prediction.

# CHAPTER 8 REFERENCES

1. C. Venkatesanet, S. D. Raskar , S. S. Tambe , B. D. Kulkarni , and R. N. Keshavamurty , “Prediction of all India summer monsoon rainfall using ErrorBack- Propagation Neural Networks,” Meteorology and Atmospheric Physics, pp. 225-240, 2017.
2. N. S. Philip and K. B. Joseph, “On the predictability of rainfall in Kerala. An application of ABF neural network,” Computational ScienceICCS, Springer Berlin Heidelberg, pp. 1-12, 2015.
3. S. Chattopadhyay and M. Chattopadhyay, “A Soft Computing technique in rainfall forecasting,” Int. Conf. on IT, HIT, pp. 19-21, 2017.
4. S. Chattopadhyay and G. Chattopadhyay, “Comparative study among different neural net learning algorithms applied to rainfall time series, "Meteorological applicat., vol. 15, no. 2, pp. 273-280, 2018.
5. K. K. Htike and O. O. Khalifa, “Rainfall forecasting models using Focused Time- Delay Neural Networks,” Comput. and Commun. Eng. (ICCCE), Int. Conf. on IEEE, 2019.
6. S. Kannan, Subimal Ghosh, “Prediction of daily rainfall state in a river basin using statistical downscaling from GCM output”, Springer-Verlag, July2011.

[7].Abhishek, Kumar, et al. "A rainfall prediction model using artificial neural network." IEEE Control and System Graduate Research Colloquium. IEEE, 2018.

1. A. K. Sahai, M. K. Soman, and V. Satyan, “All India summer monsoon rainfall prediction using an Artificial Neural Network,” Climate dynamics, vol. 16, no. 4, pp. 291-302, 2018.
2. V. K. Somvanshi, O. P. Pandey, P. K. Agrawal, N.V.Kalanker1, M.Ravi Prakash, and Ramesh Chand, “Modeling and prediction of rainfall using Artificial Neural Network and ARIMA techniques,” J. Ind. Geophys. Union, vol. 10, no. 2, pp. 141-151, 2016.
3. C. L. Wu, K. W. Chau, and C. Fan, “Prediction of rainfall time series using Modular Artificial Neural Networks coupled with datapreprocessing techniques," J. of hydrology, vol. 389, no. 1, pp. 146- 167, 2010.
4. M.Kannan, S.Prabhakaran, P.Ramachandran, “Rainfall Forecasting Using Data Mining Technique”, International Journal of Engineering and Technology Vol.2 (6), 397-401, 2010.
5. A. Kumar, A. Kumar, R. Ranjan, and S. Kumar, “A rainfall prediction model using artificial neural network,” Control and Syst. Graduate Research Colloq. (ICSGRC), pp. 82-87, 2012.
6. S. K. Nanda, D. P. Tripathy, S. K. Nayak, and S. Mohapatra, “Prediction of rainfall in India using Artificial Neural Network (ANN) models,” Int. J. of Intell. Syst. and Applicat., vol. 5, no. 12, pp. 1-22, 2013.
7. Thirumalai, Chandrasegar, et al. "Heuristic prediction of rainfall using machine learning techniques." 2017 International Conference on Trends in Electronics and Informatics (ICEI). IEEE, 2017.
8. Parmar, Aakash, Kinjal Mistree, and Mithila Sompura. "Machine learning techniques for rainfall prediction: A review." 2017 International Conference on Innovations in information Embedded and Communication Systems. 2017.
9. G. Geetha and R. S. Selvaraj, “Prediction of monthly rainfall in Chennai using Back Propagation Neural Network model,” Int. J. of Eng. Sci. and Technology, vol. 3, no. 1, pp. 211 213, 2011.
10. Soo-Yeon Ji, Sharad Sharma, Byunggu Yu, Dong Hyun Jeong, “Designing a Rule- Based Hourly Rainfall Prediction Model”, IEEE IRI 2012, August – 2012.
11. Pinky Saikia Dutta, Hitesh Tahbilder, “Prediction Of Rainfall Using Data mining Technique Over Assam”, Indian Journal of Computer Science and Engineering (IJCSE), Vol. 5 No.2 Apr May 2014.
12. Geetha, A., and G. M. Nasira. "Data mining for meteorological applications: Decision trees for modeling rainfall prediction." 2014 IEEE International Conference on Computational Intelligence and Computing Research. IEEE, 2014.
13. Singh, Gurpreet, and Deepak Kumar. "Hybrid Prediction Models for Rainfall Forecasting." 2019 9th International Conference on Cloud Computing, Data Science & Engineering (Confluence). IEEE, 2019.