Weather Analysis Using Machine Learning

1. Introduction:

1.1 Weather Forecasting

Weather forecasting is the task of predicting the state of the atmosphere at a future time and a specified location. Traditionally, this has been done through physical simulations in which the atmosphere is modeled as a fluid. The present state of the atmosphere is sampled, and the future state is computed by numerically solving the equations of fluid dynamics and thermodynamics. However, the system of ordinary differential equations that govern this physical model is unstable under perturbations, and uncertainties in the initial measurements of the atmospheric conditions and an incomplete understanding of complex atmospheric processes restrict the extent of accurate weather forecasting to a 10 day period, beyond which weather forecasts are significantly unreliable.

Machine learning, on the contrary, is relatively robust to perturbations and doesn't require a complete understanding of the physical processes that govern the atmosphere. Therefore, machine learning may represent a viable alternative to physical models in weather forecasting. Two machine learning algorithms were implemented:

1. **Linear regression:** A corpus of historical weather data for Delhi, India was obtained and used to train this algorithm. The input to these algorithms was the weather data of the past two days, which include the temperature, mean humidity, mean atmospheric pressure, and weather classification for rainy season, winter season and summer season. The output was the graphical prediction of the weather

2. **Decision Tree algorithm**: Decision Tree algorithm belongs to the family of supervised learning algorithms. The goal of using a Decision Tree is to create a training model that can use to predict the class or value of the target variable by learning simple decision rules inferred from training data. The input to these algorithms was the weather data of the past two days, which include the temperature, mean humidity, mean atmospheric pressure, and weather classification for rainy season, winter season and summer season. The output was the graphical prediction of the weather as decision tree.

1.2 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.

Agrawal et al. (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.

2. Purpose:

An analysis model is an idealized representation of a weather system that helps you as a meteorologist visualize that system, its associated weather, and its evolution. It provides a distribution in space and time of the typical three-dimensional system structure.

By using deep learning mathematical models, AI could learn from past weather records to predict the future. One example is the Numerical Weather Prediction (NWP).This model studies and analyses vast data sets from satellites and other sensors to provide short term weather forecasts and long term climate predictions.

3. Intended Audience:

- Usually peoples need information about weather for their daily work such as prepare for how to dress (i.e. warm weather, cold weather, windy weather, rainy weather)
- For business peoples information about weather helps to plan for power production and how much power to use (i.e. power companies, where to set thermostat)
- Helps people plan outdoor activities (i.e. to see if rain/storms/cold weather will impact outdoor event)
- Weather predictions helps people know if they need to leave early for work.
- Helps people involved with certain activities to know if conditions will be good (i.e. skiing, boating, ballooning)
- People who wants to know what sort of weather can be expected (i.e. a snow on the way, severe storms)
- Helps people with health related issues to plan the day (i.e. allergies, asthma, heat stress)
- To achieve all the goals of the intended audience weather forecasting is useful.
- Forecasting the weather begins by continuously observing the state of the atmosphere, the ocean, and land surface. These analyses serve as the foundation for weather prediction on scales from individual clouds to regional severe weather events and global patterns.

4. Intended devices:

Weather forecasting can be achieve through mobile as well as computer.

5. Proposed work

5.1 Problem Definition:

In this project, we tried to predict weather using different weather predicting/forecasting algorithms such as Linear Regression and Decision Tree Algorithm.

5.2 Final objective:

Final objective of this project is to predict the weather using different prediction algorithm, build models for different algorithm, calculate accuracy for each model and implement best model for weather forecasting

5.3 Constraint:

5.3.1 Operating Environment/Dependency:

This project is platform independent, and can be implemented on any operating environment like Linux, Windows and Mac OS.

5.3.2 Tools:

Tools used in this project are as follows

- Anaconda's Jupyter notebook
- Python libraries like Numpy, Pandas, Sklearn etc.

5.4 Technology:

5.4.1 Language:

Python version (3.7)

C++

5.4.2 Framework:

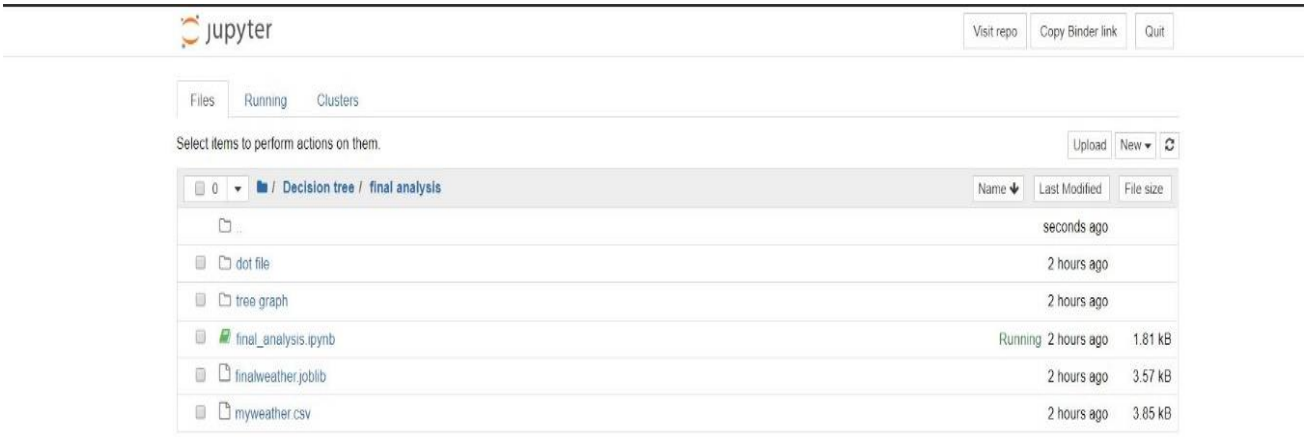
Anaconda (Jupyter Notebook)

5.4.3 Database:

For weather forecasting records from weather sites are used for e.g. (kaggle.com, towardsdatascience.com)

The data is collected in the form of Excel sheets (.cvs file) and imported directly into project

6. User Interface Basic Wireframe



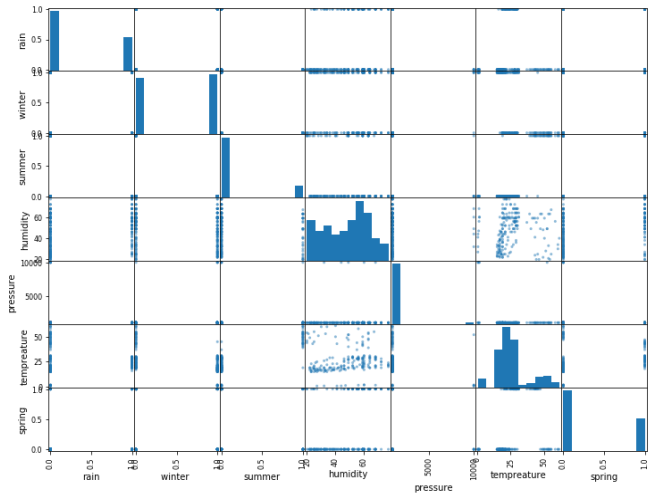
7. Algorithm

7.1 Linear Regression Algorithm:

In statistics, linear regression is a linear approach to modeling the relationship between a scalar response (or dependent variable) and one or more explanatory variables (or independent variables). The case of one explanatory variable is called simple linear regression. For more than one explanatory variable, the process is called multiple linear regression. This term is distinct from multivariate linear regression, where multiple correlated dependent variables are predicted, rather than a single scalar variable.

Linear regression has many practical uses. Most applications fall into one of the following two broad categories:

- If the goal is prediction, forecasting, or error reduction linear regression can be used to fit a predictive model to an observed data set of values of the response and explanatory variables. After developing such a model, if additional values of the explanatory variables are collected without an accompanying response value, the fitted model can be used to make a prediction of the response.
- If the goal is to explain variation in the response variable that can be attributed to variation in the explanatory variables, linear regression analysis can be applied to quantify the strength of the relationship between the response and the explanatory variables, and in particular to determine whether some explanatory variables may have no linear relationship with the response at all, or to identify which subsets of explanatory variables may contain redundant information about the response.

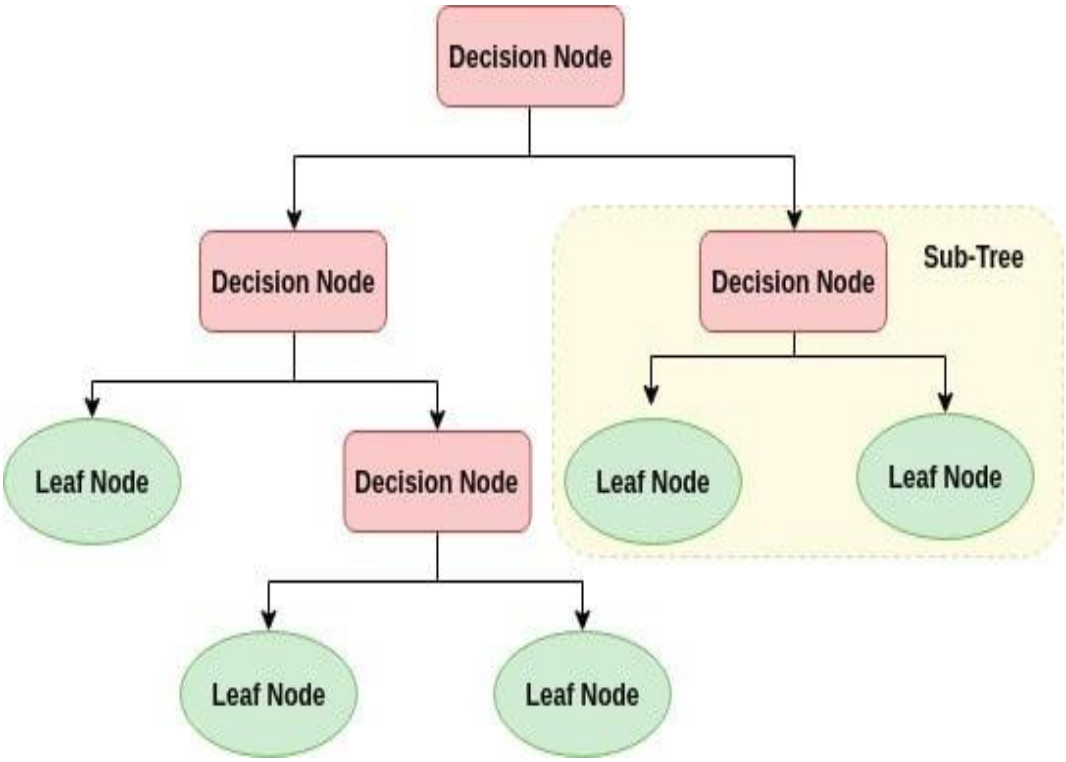


7.2 Decision Tree Algorithm:

Decision Tree algorithm belongs to the family of supervised learning algorithms. Unlike other supervised learning algorithms, the decision tree algorithm can be used for solving regression and classification problems too.

The goal of using a Decision Tree is to create a training model that can use to predict the class or value of the target variable by learning simple decision rules inferred from prior data (training data).

In Decision Trees, for predicting a class label for a record we start from the root of the tree. We compare the values of the root attribute with the record's attribute. On the basis of comparison, we follow the branch corresponding to that value and jump to the next node.



8. Implementation

8.1 Pseudocode

8.1.1 Pseudocode for Decision Tree Algorithm and Linear Regression:

1. Import all the libraries.
    - import pandas as pd
    - from sklearn.tree import DecisionTreeClassifier
    - from sklearn.model_selection import train_test_split
    - from sklearn.metrics import accuracy_score
    - from sklearn.externals import joblib
    - from sklearn import tree
    - import numpy as np
2. Import data base file “weather-data” into jupyter notebook using pandas library
    - weatherdata = pd.read_csv('myweather.csv')
3. split input and output data from the cells of the myweather file using drop() method
    - X = weatherdata.values[:,3:7]
    - Y = weatherdata.values[:,3:7]

4. Create DecisionTreeClassifier object
    - model = DecisionTreeClassifier()
5. Give our ipuut data i.e. X to the decision tree classifier
    - model.fit(X, Y)

6. now our model is created, and we have to train our model usin train_test_split method from sklearn python library
    - X_train, X_test, Y_train, Y_test = train_test_split(X, Y, test_size=0.8)
    - Here we can split data into two parts X_train, Y_train to train and X_test and Y_test to test out model, and test_size parameter is used to define size of data to be tasted

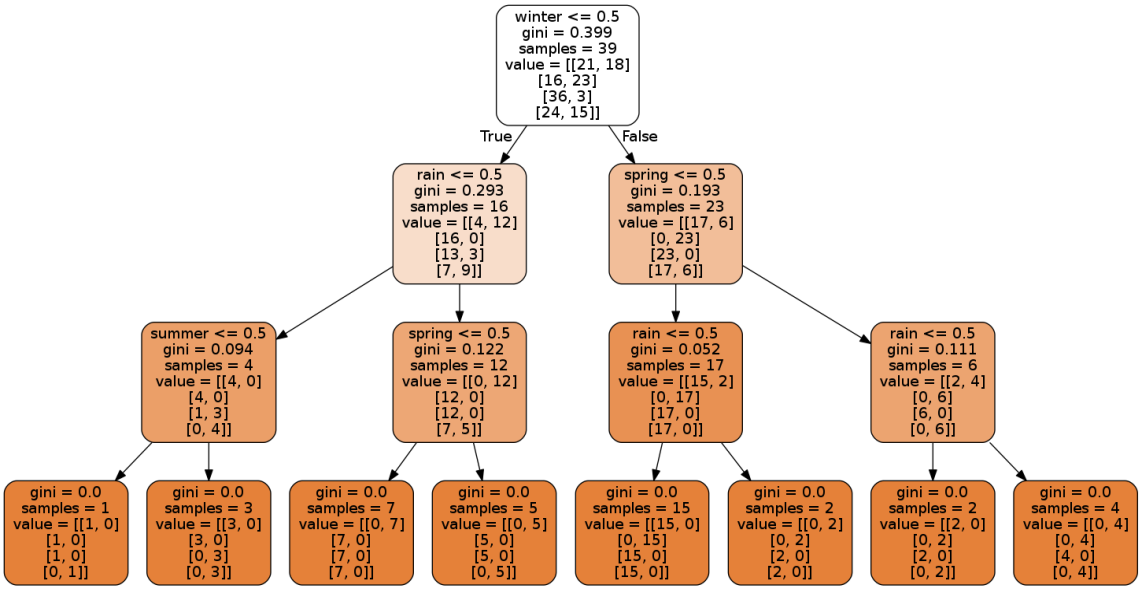
7. After training and resting model we have to check how accurate our mode is by calling accuracy_score() method from sklearn.matrics library.
    - predictions = model.predict(X_test)
    - score = accuracy_score(Y_test,predictions)

8. After model creation, training model, testing model, accuracy check we have to dump our model for predictions using sklearn’s joblib function
    - joblib.dump(model,'finalweather.joblib')
    - after successful dumping it creates .joblib file to your pc, and now we have to load our dumped model
    - joblib.load('finalweather.joblib')

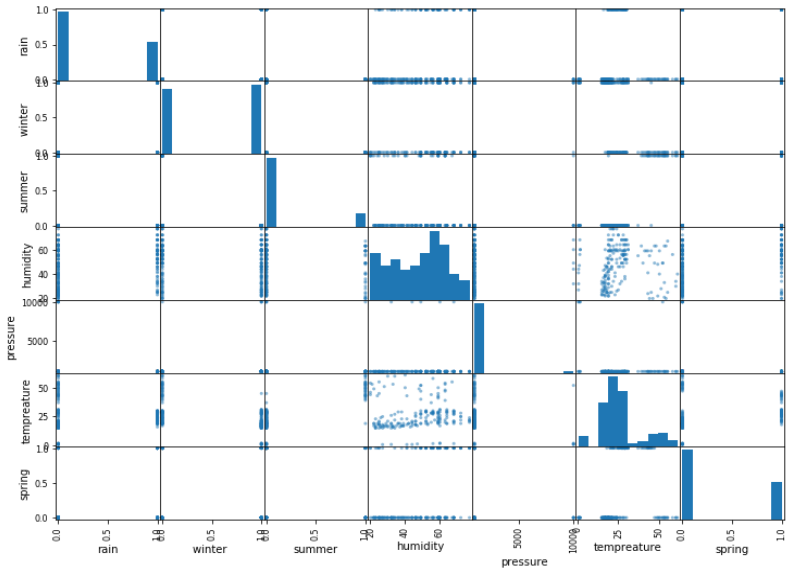
9. After that we have to represent our model in graphical form this is done by using sklearn's graphviz()

    - tree.export_graphviz(model, out_file='weather-analysis-fin.dot',feature_names=['rain','winter','summer','spring'],label='all',rounded=True,filled=True)

10. Stop



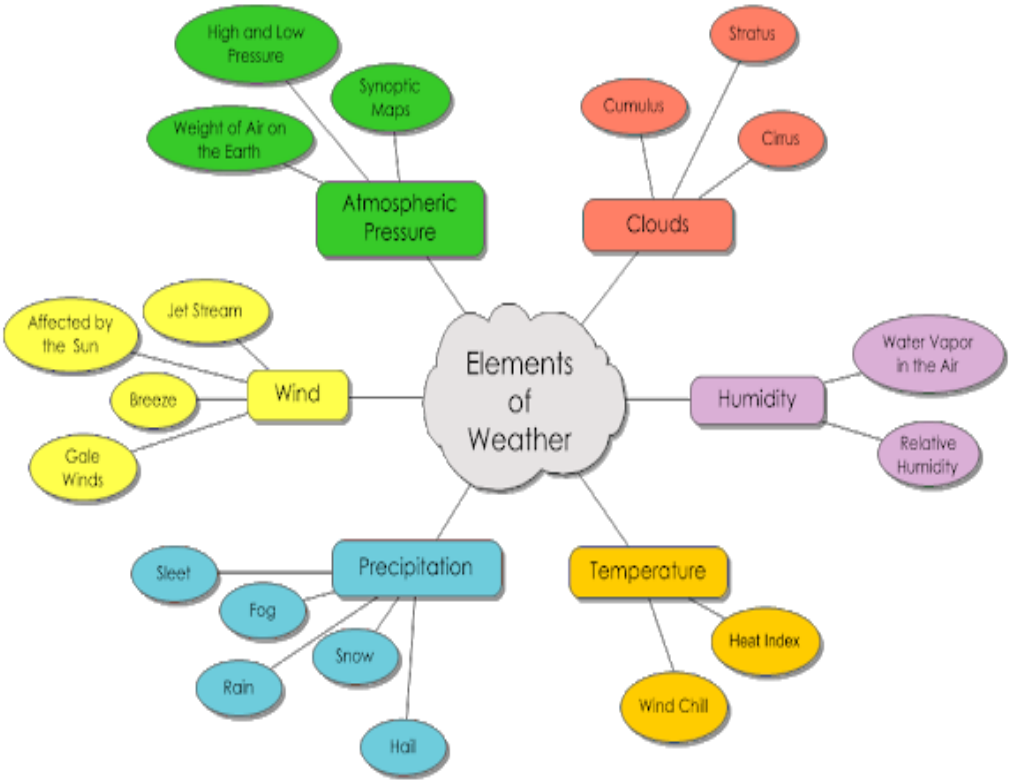
Tree Decision algorithm output



Linear regression algorithm output

8.2 Mind map

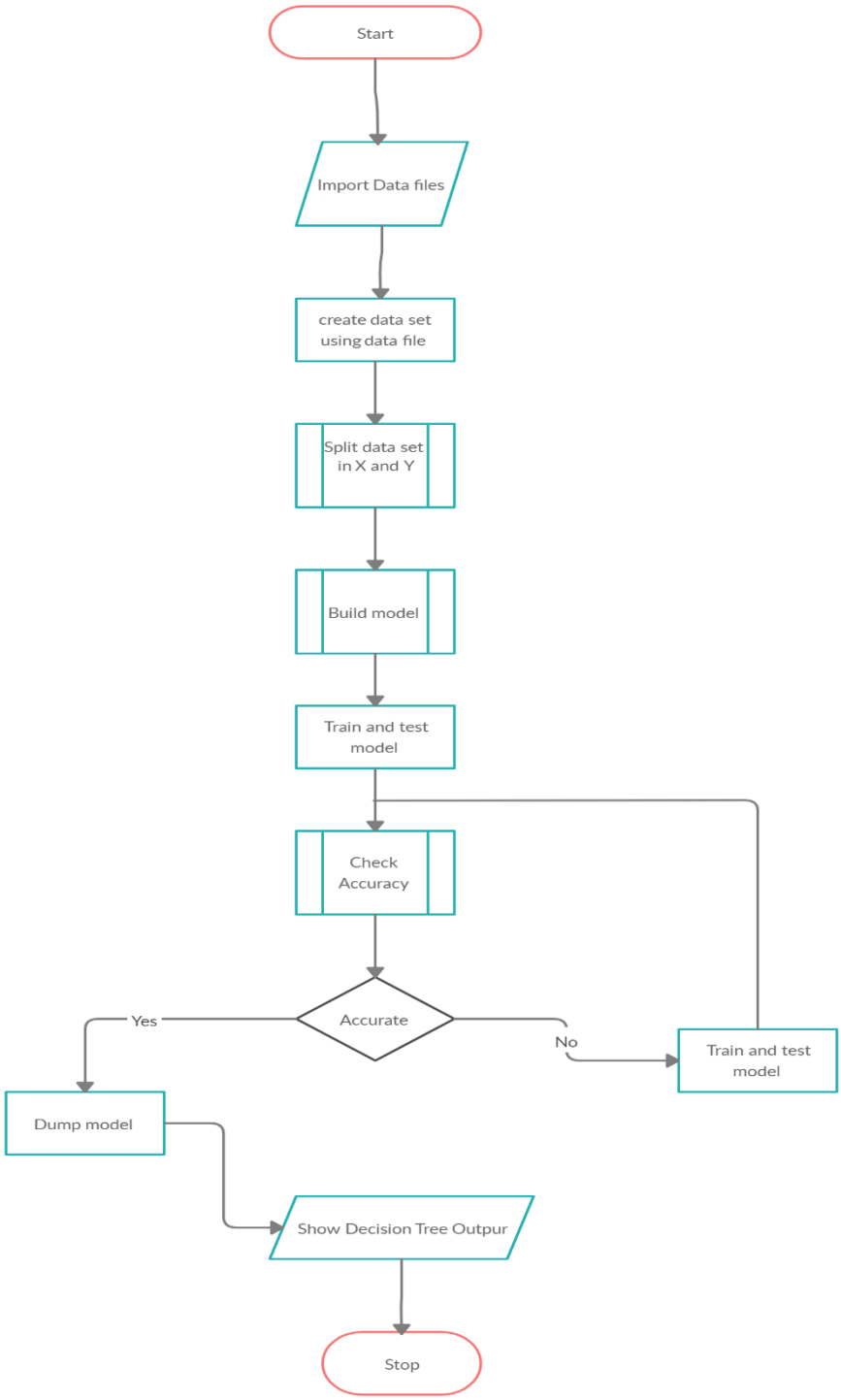
Mind map: A mind map is a graphical way to represent ideas and concepts. It is a visual thinking tool that helps structuring information, helping you to better analyse, comprehend, synthesize, recall and generate new ideas. Just as in every great idea, its power lies in its simplicity.



Mind map for weather analysis

8.3 Workflow diagram:

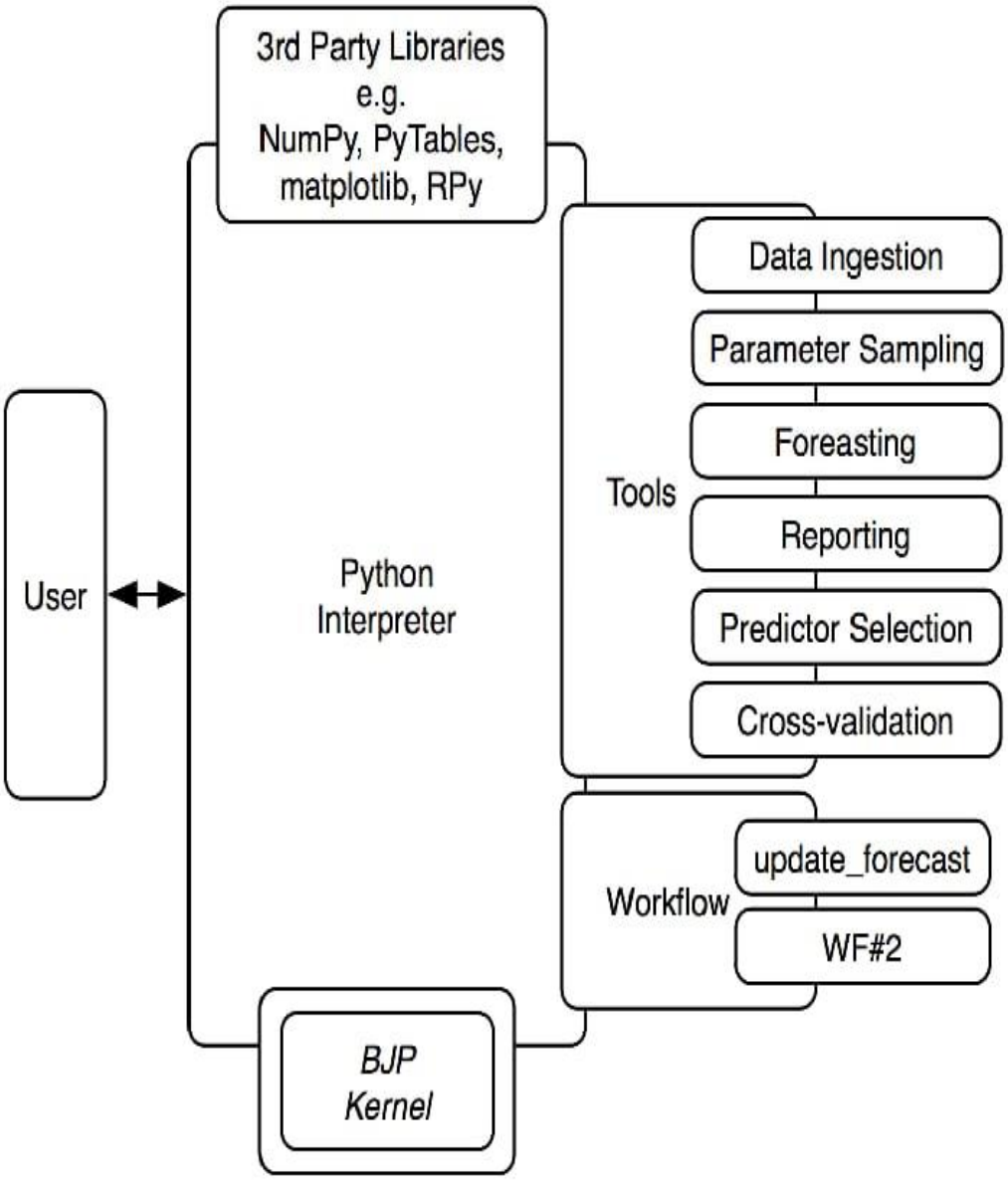
Workflow diagram: a diagram of the sequence of movements or actions of people or things involved in a complex system or activity.



Work flow diagram

8.3 System Architecture Diagram:

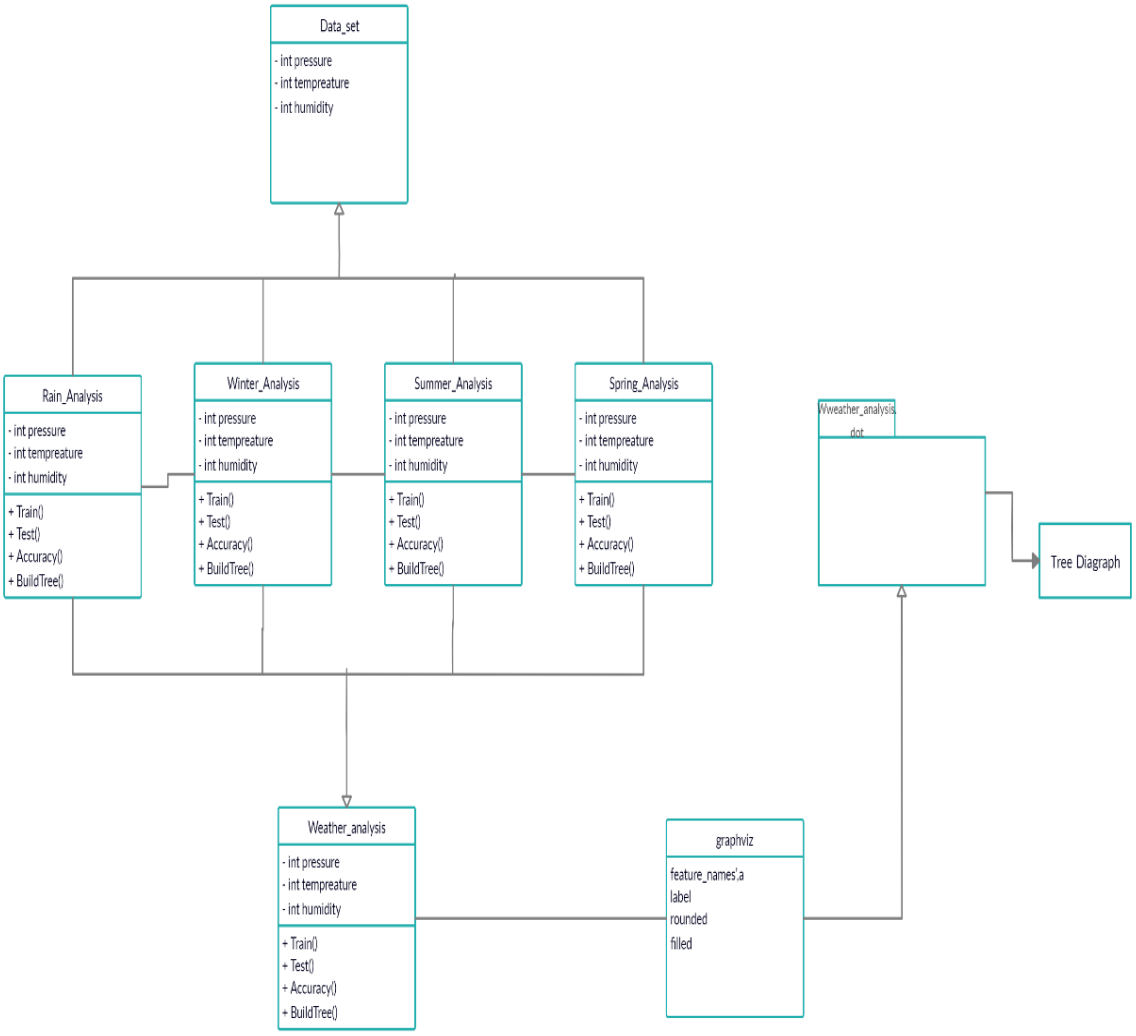
System Architecture Diagram: A system architecture diagram would be used to show the relationship between different components. Usually they are created for systems which include hardware and software and these are represented in the diagram to show the interaction between them.



System Architecture Diagram

8.4 Class Diagram

Class Diagram: A class diagram in the Unified Modeling Language (UML) is **a** type of static structure diagram that describes the structure of a system by showing the system's classes, their attributes, operations (or methods), and the relationships among objects.



Class diagram

Conclusion

Human-induced climate change has contributed to changing patterns of extreme weather across the globe, from longer and hotter heat waves to heavier rains. Extreme weather is on the rise, and the indications are that it will continue to increase, in both predictable and unpredictable ways. We tried predicting weather using two different algorithms Linear Regression and Decision Tree algorithm, and we found out, prediction accuracies for both models are different. And we conclude that due to decision node accuracy for Tree Decision algorithm can predict weather more accurately than Linear Regression algorithm.