# A PROJECT REPORT ON

**WEATHER FORECASTING USING DATA MINING**

**Submitted in partial fulfillment of the requirement for the award of the degree of**

Master of Computer Applications

# Submitted By

*BOPPEPALLI KULASEKHAR REDDY*

**HT. NO: 11400221004**

**Batch: 2021-2023**

**(CAMPUS STUDENT)**

Under the guidance of

# Prof.K.Ammulu

# (Internal Guide)

**B.Tech(CSE).,M.Tech(CSE).,Ph.D.,**

Dept. of Computer Science& Technology Dravidian University, Kuppam.

Department of Computer Science& Technology

School of Science and Technology

DRAVIDIAN UNIVERSITY KUPPAM – 517426, A.P., INDIA.

**Department of Computer Science& Technology Dravidian University**

**Kuppam – 517426**



**CERTIFICATE**

This is to certify that the project report entitled “***WEATHER FORECASTING USING DATA MINING****”* being submitted by **BOPPEPALLI KULASEKHAR REDDY, HT.NO:11400221004,** in partial fulfilment of the requirements for the award of the Degree of M.C.A from the **Dravidian University**, is a record of bonafide work.

|  |  |
| --- | --- |
| **Prof. K. Ammulu** | **Prof.V. Kiran kumar** |
| **B.Tech(CSE),M.Tech(CSE), Ph.D.,** | **M.Sc.(cs), Ph.D.,** |
| **(Internal Guide)** | **(Head of the Department)** |
| **Department of Computer Science& Technology** | |
| **Examiners** |  |
| **1.** | **2.** |

Submitted for the University Examination the University Campus, held on…...

**DECLARATION**

I am Boppepalli Kulasekhar Reddy (HTNO :11400221004) hereby declare that the project entitled “***WEATHER FORECASTING USING DATA MINNING*”** is done under the guidance of **prof.K.Ammulu.,B.Tech(CSE).,M.Tech(CSE).,Ph.D.,** submittedin the partial fulfilment of the requirements for the award of degree of master of computer Applications in Dravidian university located at Srinivasavanam, Kuppam during the academic year 2021-2023.This projected has not been submitted by anybody towards the award of degree.

Signature of the student

(B. Kulasekhar Reddy)

Date :

Place :

**ACKNOWLEDGEMENT**

It is my pleasure to be indebted to various people, who directly or indirectly contributed in the development of this work and who influenced my thinking, behaviour, and acts during the course of my training. I express my sincere gratitude to **Prof.K.Ammulu, B.Tech(CSE).,M.Tech(CSE).,Ph.D.,** Department of Computer Science, worthy mentor, guide and a great teacher who influenced and inspired me in many ways. Lastly, I would like to thank our **Head of the** **Department Prof.V. Kiran Kumar, M.Sc.(CSE).,Ph.D.,** who gave the golden opportunity to do this wonderful “***Weather Forecasting Using Data Mining”*** project which is also helped me in doing a lot of Research and I came to know about so many things.

TABLE OF CONTENTS

1.INTRODUCTION------------------------------------------------------------------ \*\*

2.SYSTEM ANALYSIS---------------------------------------------------------------- \*\*

EXISTING SYSTEM

PROPOSED SYSTEM

3.METHODOLOGY & ALGORITHMS--------------------------------------------- \*\*

DECISION TREE

LOGISTIC REGRESSION

RANDOM FOREST CLASSIFIER

4.SOFTWARE DEVELOPMENT LIFE CYCLE------------------------------------- \*\*

5.FEASABILITY STUDY-------------------------------------------------------------- \*\*

ECONOMICAL FEASABILITY

TECHNICAL FEASABILITY

SOCIAL FEASABILITY

6.SYSTEM REQUIREMENTS SPECIFICATION---------------------------------- \*\*

FUNCTIONAL REQUIREMENTS

NON-FUNCTIONAL REQUIREMENTS

SOFTWARE AND HARDWARE REQUIREMENTS

7.SYSTEM DESIGN------------------------------------------------------------------ \*\*

INPUT DESIGN

OUTPUT DESIGN

UML DIAGRAMS

USECASE DIAGRAMS

CLASS DIAGRAM

SEQUENCE DIAGRAMS

COLLABORATION DIAGRAM

DEPLOYMENT DIAGRAM

ACTIVITY DIAGRAM

COMPONENT DIAGRAM

ER DIAGRAM

DFD DIAGRAM

8.SOURCE CODE--------------------------------------------------------------------- \*\*

9.OUTPUT SCREENS---------------------------------------------------------------- \*\*

10.BIBLIOGRAPHY------------------------------------------------------------------ \*\*

11.CONCLUSION-------------------------------------------------------------------- \*\*

12.REFERENCES--------------------------------------------------------------------- \*\*

**WEATHER FORECASTING USING DATA MINING**

**OBJECTIVE**

The main objective of this research is to classify the weather forecasting using data mining. Here we are going to predict the types of weather like drizzle, rain, sun, rain and fog.

**ABSTRACT**

Weather forecasting is a method to predict what the atmosphere will be like in a particular place by using scientific knowledge to make the weather observations. Weather forecasting is a challenging task due to the dynamic and complex nature of atmospheric conditions. Recently, data mining techniques have been applied to predict weather patterns using machine learning algorithms. In this study, we propose a weather forecasting model that predicts weather types based on historical weather data. The dataset used in this study includes precipitation, temperature, wind speed, and direction collected from various weather stations. To predict weather types, we used algorithms, which is a popular machine learning technique for classification tasks. The model was trained on historical weather data and tested on a separate set of data to evaluate its accuracy. The results showed that our proposed model achieved a high accuracy rate of over 90%, indicating that it could be a valuable tool for weather forecasting. The study demonstrated that data mining techniques can be used to predict weather patterns accurately. The proposed model can provide valuable insights to weather forecasters and decision-makers in various industries such as agriculture, transportation, and aviation, which rely heavily on accurate weather predictions. The model can be improved by incorporating more weather variables, such as precipitation, cloud cover, and solar radiation, and by applying more sophisticated machine learning techniques, such as ensemble methods.

**Keywords**: Machine Learning, Random Forest, Logistic Regression, Decision Tree, ML techniques, evaluation.

**INTRODUCTION**

Weather Forecasting is area of current technology that predicts the atmospheric conditions for a particular location by collecting dynamic data related to weather. The knowledge of current atmospheric state is updating by various weather forecasting tools such as satellites, Balloons and aircraft, Buoys and land stations and Radar System. The data collected from states are distorted into Numerical representation is known as assimilation. Weather Predictions are essential for various applications such as climate monitoring, drought detection, agriculture and production, energy industry, aviation industry, communication, pollution dispersal etc. The technique used for weather prediction is data mining. Data mining refers to the extraction of given set of data that is useful for making predictions and correlation between different parameters. Data mining offers that to analyze datas and derive rules for weather predictions. In database the datas are stored and organized .Data mining techniques are divided into classification and clustering techniques. Classification is a data mining technique designed for classifying unknown samples. Rainfall can be easily predict using classification techniques. Clustering is a technique that grouping the objects based on information. Weather forecasting is the prediction of future weather conditions, such as temperature, humidity, precipitation, wind speed, and direction, based on past and present weather data. Accurate weather forecasting is critical for various sectors such as agriculture, aviation, transportation, energy, and public safety. Traditional forecasting methods rely on meteorological models, which are complex mathematical simulations of atmospheric conditions, and historical data. However, these models can be inaccurate due to the complexity of the weather system, the limited availability of data, and the uncertainty of the atmosphere's behavior.Data mining is an advanced analytical method that involves extracting patterns and knowledge from large datasets. It has become increasingly popular in weather forecasting due to its ability to handle massive amounts of data, identify hidden patterns, and provide accurate predictions. Data mining techniques can be used to analyze various weather data sources such as satellite imagery, radar data, and weather station observations.One of the most common data mining techniques used in weather forecasting is machine learning. Machine learning algorithms can be trained on large amounts of historical weather data to learn the relationships between different weather variables and predict future weather conditions. For example, a machine learning model can be trained to predict the probability of rainfall based on factors such as temperature, humidity, and wind speed. The model can then be used to make accurate forecasts in real-time.Another data mining technique used in weather forecasting is cluster analysis. Cluster analysis involves grouping similar weather patterns based on their characteristics. For example, a cluster analysis algorithm can group similar weather patterns based on temperature, humidity, and pressure. This technique can help meteorologists identify common weather patterns and make accurate predictions based on past weather patterns.Data mining can also be used to analyze the impact of weather on various sectors such as agriculture, transportation, and energy. For example, data mining techniques can be used to analyze the relationship between weather conditions and crop yields. This information can help farmers make informed decisions about planting and harvesting crops. Similarly, data mining can be used to predict the impact of severe weather events such as hurricanes and tornadoes on energy infrastructure and transportation networks.

**LITERATURE SURVEY**

**[1]. Pushpa Mohan and Dr. Kiran KumariPatil: “Survey on Crop and Weather Forecasting based on Agriculture related Statistical Data”, International Journal of Innovative Research in Computer and Communication Engineering,Volume 5, Issue 2…**

Weather is the most important factor in terms of farming and agriculture. It continuous, data-intensive, multidimensional, and chaotic process. These properties of weather make its forecasting a formidable challenge. The most technologically challenged problems of the last century are weather forecasting. The harvest of crops is dependent on this factor. To make an accurate weather prediction is one of the major challenges that is being faced all over the world. Scientists have tried their best to forecast environmental characteristics using a number of methods and some of these methods are more accurate than others. Weather forecasts provide critical information about future weather. Every year notorious weather harms the life, property and many government activities which is usually heavily funded is destroyed, as a result weather forecasting would help government to plan out things in advance to prepare its citizens for the worst of the weather. There are many different methodologies that have come into observation regarding weather prediction. This paper describes one of the many techniques used for prediction of weather which will be beneficial for the farmers, agricultural and scientists. It will help them to better understand the weather for yielding crops and for studying environment too.

**[2]** **Amruta A. Taksande and P. S. Mohod: “Applications of Data Mining in Weather Forecasting Using Frequent Pattern Growth Algorithm”, International Journal of Science and Research**

Rainfall forecasting or Weather forecasting has been one of the most challenging problems around the world because it consists of multidimensional and nonlinear data such as in the field of agriculture to determine initial growing season. Recently, climate change causes much trouble in rainfall forecasting. Our Project describes five data mining algorithms namely neural network (NN), random forest, classification and regression tree (CRT), support vector machine (SVM) and k-nearest neighbour. Generally these algorithms are used for the prediction. Generally these algorithms used for prediction. In this project we use five years previous data from Jan 2010-Jan 2014 for Nagpur station. On available datasets we apply the Frequent Pattern Growth Algorithm for deleting the inappropriate data. Generally temperature, humidity, wind speed are mainly responsible for the rainfall prediction. On the percentage of these parameters we predict there is a rainfall or not. Based on experiment result, it can be concluded that the combination of GA and FP growth algorithm weather data can gives prediction with higher than 90 % accuracy with several population size and crossover probability.

**[3] Fahad Sheikh and S. Karthick: “Analysis of Data Mining Techniques for Weather Prediction”,Indian Journal of Science and Technology, Volume 9**

To forecast weather, which is one of the greatest challenges in meteorological department. Weather prediction is necessary so as to inform people and prepare them in advance about the current and upcoming weather condition. This helps in reduction in loss of human life and loss of resources and minimizing the mitigation steps that are expected to be taken after a natural disaster occurs. Methods/Statistical analysis: This study makes a mention of various techniques and algorithms that are likely to be chosen for weather prediction and highlights the performance analysis of these algorithms. Various other ensemble techniques are also discussed that are used to boost the performance of the application. Findings: After a comparison between the data mining algorithms and corresponding ensemble technique used to boost the performance, a classifier is obtained that will be further used to predict weather. Applications: Used to Predict and forecast the weather condition of specific region based on the available pre historical data which helps to save resources and prepare for the changes forth coming.

**[4] Prashant Biradar, Sarfraz Ansari, Yashavant Paradkar and Savita Lohiya:“Weather Prediction Using Data Mining”,International Journal of engineering research and development,Volume 5, Issue 2**

Weather forecasting is the application of science and technology to predict the state of the atmosphere for a given locati on. Here this system will predict weather based on parameters such as temperature, humidity and wind. This system is a web application with effective graphical user interface. To predict the future’s weather condition, the variation in the conditions in past years must be utilized. The probability that it will match within the span of adjacent fortnight of previous year is very high .We have proposed the use of K-medoids and Naive Bayes algorithm for weather forecasting system with parameters such as temperature, humidity, and wind. It will forecast weather based on previous record therefore this prediction will prove reliable. This system can be used in Air Traffic, Marine, Agriculture, Forestry,

Military, and Navy etc.

**SYSTEM ANALYSIS & FEASIBILITY STUDY**

**EXISTING SYSTEM**

Weather forecasting is a critical application of data mining and machine learning. The existing system of weather forecasting using data mining and machine learning involves the use of historical weather data to predict future weather patterns. They have implemented Support vector machine, Naïve baye’s. To overcome all this, we use machine learning packages available in the scikit-learn library.

**Disadvantages:**

* High complexity.
* Time consuming.

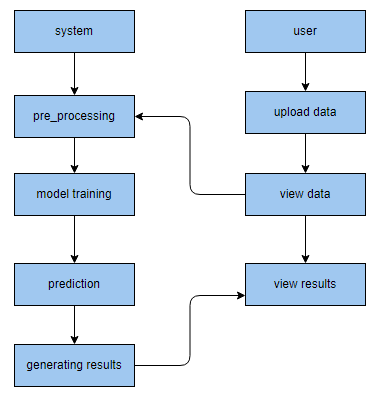
**PROPOSED SYSTEM**

The proposed system of weather forecasting using data mining and machine learning involves collecting and analyzing large amounts of weather data, such as temperature, humidity, wind speed, precipitation, and atmospheric pressure. This data can be obtained from various sources, including weather stations, satellites, and other sensors. Therefore, we propose a Random Forest, Logistic Regression and Decision Tree machine Classifier to predict the weather types.

**Advantages**:

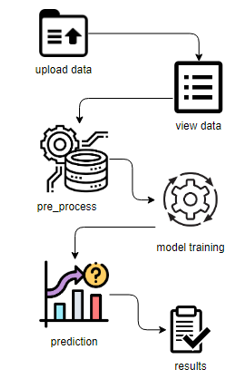
* Highest accuracy
* Reduces time complexity.
* Easy to use

**BLCOK DIAGRAM**



**Fig**: Block Diagram

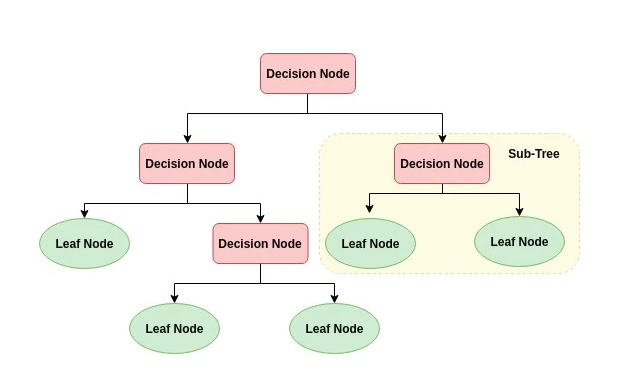
**Architecture:**

****

**METHODOLOGY AND ALGORITHMS:**

**1. DECISION TREE:**

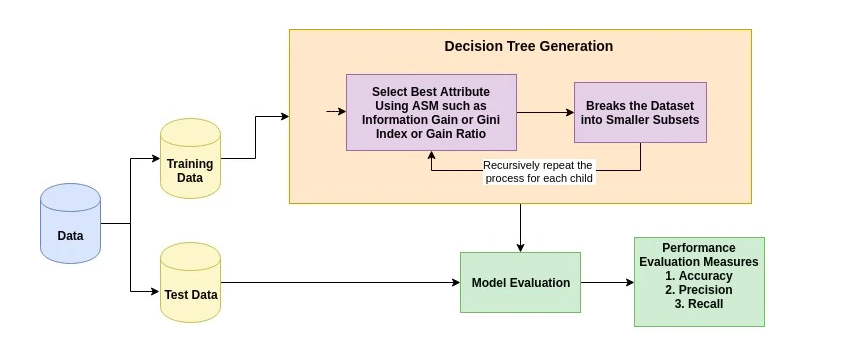
Decision tree is a flowchart-like tree structure where an internal node represents feature(or attribute), the branch represents a decision rule, and each leaf node represents the outcome. The topmost node in a decision tree is known as the root node. It learns to partition on the basis of the attribute value. It partitions the tree in recursively manner call recursive partitioning. This flowchart-like structure helps you in decision making. It's visualization like a flowchart diagram which easily mimics the human level thinking. That is why decision trees are easy to understand and interpret.



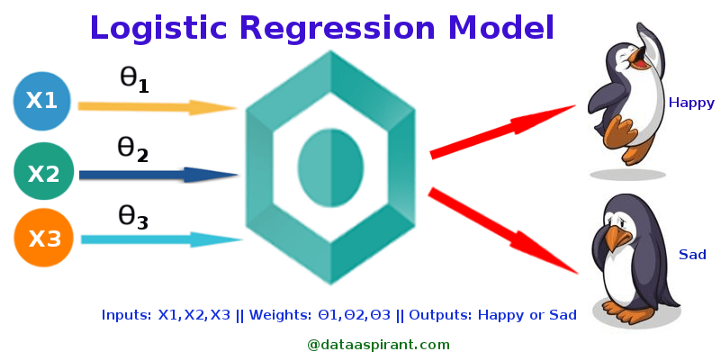
The basic idea behind any decision tree algorithm is as follows:

1. Select the best attribute using Attribute Selection Measures (ASM) to split the records.
2. Make that attribute a decision node and breaks the dataset into smaller subsets.
3. Starts tree building by repeating this process recursively for each child until one of the conditions will match:

* All the tuples belong to the same attribute value.
* There are no more remaining attributes.
* There are no more instances.



**2. Logistic Regression:**



Logistic Regression was used in the biological sciences in early twentieth century. It was then used in many social science applications. Logistic Regression is used when the dependent variable(target) is categorical. For example,

* To predict whether an email is spam (1) or (0)
* Whether the tumor is malignant (1) or not (0)
* Consider a scenario where we need to classify whether an email is spam or not. If we use linear regression for this problem, there is a need for setting up a threshold based on which classification can be done. Say if the actual class is malignant, predicted continuous value 0.4 and the threshold value is 0.5, the data point will be classified as not malignant which can lead to serious consequence in real time.
* From this example, it can be inferred that linear regression is not suitable for classification problem. Linear regression is unbounded, and this brings logistic regression into picture. Their value strictly ranges from 0 to 1.

## **Where to use logistic regression**

Logistic regression is used to solve classification problems, and the most common use case is [binary logistic regression](https://en.wikipedia.org/wiki/Logistic_regression#:~:text=Binary%20logistic%20regression%20is%20used%20to%20predict%20the%20odds%20of%20being%20a%20case%20based%20on%20the%20values%20of%20the%20independent%20variables%20(predictors).%20The%20odds%20are%20defined%20as%20the%20probability%20that%20a%20part), where the outcome is binary (yes or no). In the real world, you can see logistic regression applied across multiple areas and fields.

* In health care, logistic regression can be used to predict if a tumor is likely to be benign or malignant.
* In the financial industry, logistic regression can be used to predict if a transaction is fraudulent or not.
* In marketing, logistic regression can be used to predict if a targeted audience will respond or not.

**The three types of logistic regression**

1. **Binary logistic regression** - When we have two possible outcomes, like our original example of whether a person is likely to be infected with COVID-19 or not.
2. **Multinomial logistic regression** - When we have multiple outcomes, say if we build out our original example to predict whether someone may have the flu, an allergy, a cold, or COVID-19.
3. **Ordinal logistic regression** - When the outcome is ordered, like if we build out our original example to also help determine the severity of a COVID-19 infection, sorting it into mild, moderate, and severe cases.

**3. Random Forest Classifier:**

A random forest is a machine learning technique that’s used to solve regression and classification problems. It utilizes ensemble learning, which is a technique that combines many classifiers to provide solutions to complex problems.

A random forest algorithm consists of many decision trees. The ‘forest’ generated by the random forest algorithm is trained through bagging or bootstrap aggregating. Bagging is an ensemble meta-algorithm that improves the accuracy of machine learning algorithms.

The (random forest) algorithm establishes the outcome based on the predictions of the decision trees. It predicts by taking the average or mean of the output from various trees. Increasing the number of trees increases the precision of the outcome.

A random forest eradicates the limitations of a decision tree algorithm. It reduces the over fitting of datasets and increases precision. It generates predictions without requiring many configurations in packages (like [Scikit-learn](https://en.wikipedia.org/wiki/Scikit-learn)).

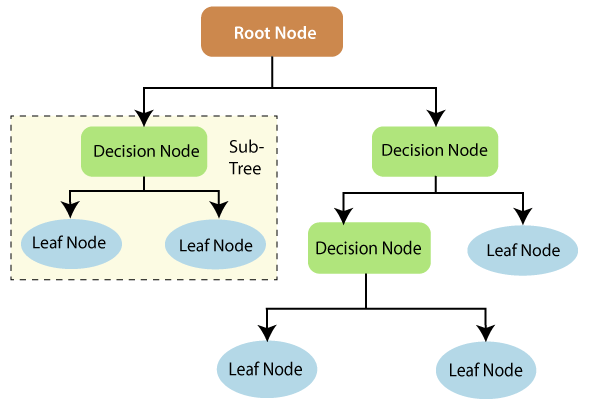
Features of a Random Forest Algorithm:

* It’s more accurate than the decision tree algorithm.
* It provides an effective way of handling missing data.
* It can produce a reasonable prediction without hyper-parameter tuning.
* It solves the issue of over fitting in decision trees.
* In every random forest tree, a subset of features is selected randomly at the node’s splitting point.

Decision trees are the building blocks of a random forest algorithm. A decision tree is a decision support technique that forms a tree-like structure. An overview of decision trees will help us understand how random forest algorithms work.

A decision tree consists of three components: decision nodes, leaf nodes, and a root node. A decision tree algorithm divides a training dataset into branches, which further segregate into other branches. This sequence continues until a leaf node is attained. The leaf node cannot be segregated further.

The nodes in the decision tree represent attributes that are used for predicting the outcome. Decision nodes provide a link to the leaves. The following diagram shows the three types of nodes in a decision tree.



The information theory can provide more information on how decision trees work. Entropy and information gain are the building blocks of decision trees. An overview of these fundamental concepts will improve our understanding of how decision trees are built.

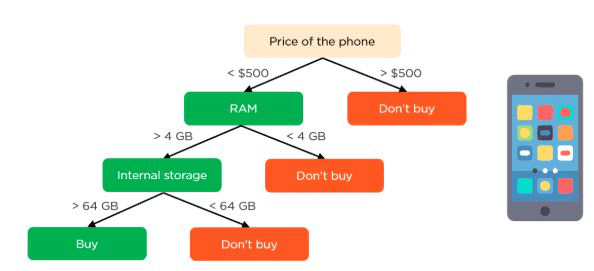
Entropy is a metric for calculating uncertainty. Information gain is a measure of how uncertainty in the target variable is reduced, given a set of independent variables.

The information gain concept involves using independent variables (features) to gain information about a target variable (class). The entropy of the target variable (Y) and the [conditional entropy](https://en.wikipedia.org/wiki/Conditional_entropy) of Y (given X) are used to estimate the information gain. In this case, the conditional entropy is subtracted from the entropy of Y.

Information gain is used in the training of decision trees. It helps in reducing uncertainty in these trees. A high information gain means that a high degree of uncertainty (information entropy) has been removed. Entropy and information gain are important in splitting branches, which is an important activity in the construction of decision trees.

Let’s take a simple example of how a decision tree works. Suppose we want to predict if a customer will purchase a mobile phone or not. The features of the phone form the basis of his decision. This analysis can be presented in a decision tree diagram.

The root node and decision nodes of the decision represent the features of the phone mentioned above. The leaf node represents the final output, either *buying* or *not buying*. The main features that determine the choice include the price, internal storage, and Random Access Memory (RAM). The decision tree will appear as follows.



Applying decision trees in random forest

The main difference between the decision tree algorithm and the random forest algorithm is that establishing root nodes and segregating nodes is done randomly in the latter. The random forest employs the bagging method to generate the required prediction.

Bagging involves using different samples of data (training data) rather than just one sample. A training dataset comprises observations and features that are used for making predictions. The decision trees produce different outputs, depending on the training data fed to the random forest algorithm. These outputs will be ranked, and the highest will be selected as the final output.

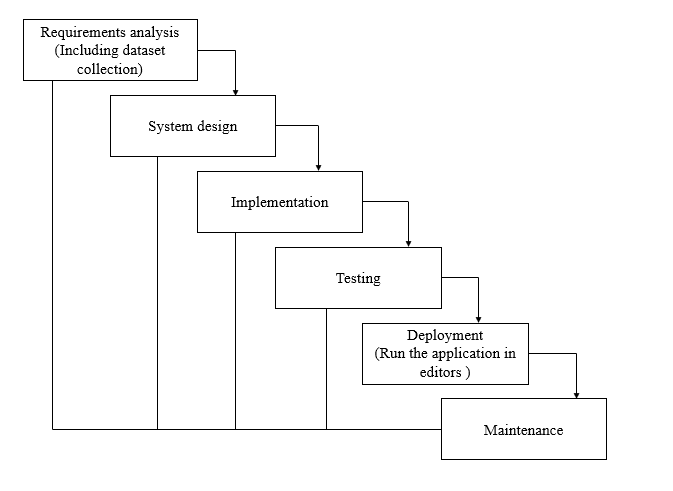
Our first example can still be used to explain how random forests work. Instead of having a single decision tree, the random forest will have many decision trees. Let’s assume we have only four decision trees. In this case, the training data comprising the phone’s observations and features will be divided into four root nodes.

The root nodes could represent four features that could influence the customer’s choice (price, internal storage, camera, and RAM). The random forest will split the nodes by selecting features randomly. The final prediction will be selected based on the outcome of the four trees.

The outcome chosen by most decision trees will be the final choice. If three trees predict *buying*, and one tree predicts *not buying*, then the final prediction will be *buying*. In this case, it’s predicted that the customer will buy the phone.

**SOFTWARE DEVELOPMENT LIFE CYCLE – SDLC:**

In our project we use waterfall model as our software development cycle because of its step-by-step procedure while implementing.



**Fig1**: Waterfall Model

* **Requirement Gathering and analysis** − All possible requirements of the system to be developed are captured in this phase and documented in a requirement specification document.
* **System Design** − the requirement specifications from first phase are studied in this phase and the system design is prepared. This system design helps in specifying hardware and system requirements and helps in defining the overall system architecture.
* **Implementation** − with inputs from the system design, the system is first developed in small programs called units, which are integrated in the next phase. Each unit is developed and tested for its functionality, which is referred to as Unit Testing.
* **Integration and Testing** − All the units developed in the implementation phase are integrated into a system after testing of each unit. Post integration the entire system is tested for any faults and failures.
* **Deployment of system** − Once the functional and non-functional testing is done; the product is deployed in the customer environment or released into the market.
* **Maintenance** − There are some issues which come up in the client environment. To fix those issues, patches are released. Also, to enhance the product some better versions are released. Maintenance is done to deliver these changes in the customer environment.

**FEASIBILITY STUDY**

The feasibility of the project is analysed in this phase and business proposal is put forth with a very general plan for the project and some cost estimates. During system analysis the feasibility study of the proposed system is to be carried out. This is to ensure that the proposed system is not a burden to the company. For feasibility analysis, some understanding of the major requirements for the system is essential.

Three key considerations involved in the feasibility analysis are

* ECONOMICAL FEASIBILITY
* TECHNICAL FEASIBILITY
* SOCIAL FEASIBILITY

**Economic feasibility:**

This study is carried out to check the economic impact that the system will have on the organization. The amount of fund that the company can pour into the research and development of the system is limited. The expenditures must be justified. Thus, the developed system as well within the budget and this was achieved because most of the technologies used are freely available. Only the customized products had to be purchased.

### Technical feasibility:

This study is carried out to check the technical feasibility, that is, the technical requirements of the system. Any system developed must not have a high demand on the available technical resources. This will lead to high demands on the available technical resources. This will lead to high demands being placed on the client. The developed system must have a modest requirement, as only minimal or null changes are required for implementing this system.

**Social feasibility:**

The aspect of study is to check the level of acceptance of the system by the user. This includes the process of training the user to use the system efficiently. The user must not feel threatened by the system, instead must accept it as a necessity. The level of acceptance by the users solely depends on the methods that are employed to educate the user about the system and to make him familiar with it. His level of confidence must be raised so that he is also able to make some constructive criticism, which is welcomed, as he is the final user of the system.

**SYSTEM REQUIREMENTS SPECIFICATION**

**Functional and non-functional requirements:**

Requirement’s analysis is very critical process that enables the success of a system or software project to be assessed. Requirements are generally split into two types: Functional and non-functional requirements.

**Functional Requirements**: These are the requirements that the end user specifically demands as basic facilities that the system should offer. All these functionalities need to be necessarily incorporated into the system as a part of the contract. These are represented or stated in the form of input to be given to the system, the operation performed and the output expected. They are basically the requirements stated by the user which one can see directly in the final product, unlike the non-functional requirements.

Examples of functional requirements:

1. Authentication of user whenever he/she logs into the system
2. System shutdown in case of a cyber-attack
3. A verification email is sent to user whenever he/she register for the first time on some software system.

**Non-functional requirements**: These are basically the quality constraints that the system must satisfy according to the project contract. The priority or extent to which these factors are implemented varies from one project to other. They are also called non-behavioral requirements.  
They basically deal with issues like:

* Portability
* Security
* Maintainability
* Reliability
* Scalability
* Performance
* Reusability
* Flexibility

Examples of non-functional requirements:

1. Emails should be sent with a latency of no greater than 12 hours from such an activity.
2. The processing of each request should be done within 10 seconds
3. The site should load in 3 seconds whenever of simultaneous users are > 10000

**SOFTWARE AND HARDWARE REQUIREMENTS**

**H/W Configuration:**

Operating system : Windows 7 or 7+

RAM : 8 GB

Hard disc or SSD : More than 500 GB

Processor : Intel 3rd generation or high or Ryzen with 8 GB Ram

**S/W Configuration:**

Software’s : Python 3.6 or high version

IDE : PyCharm.

Framework : Django, pandas, numpy and Scikit-Learn

**Learning Outcome:**

1. Regarding Machine Learning
2. Supervised Machine Learning
3. Classification Technique
4. About Pycharm
5. Flask Frame work

**SYSTEM DESIGN:**

## **Input Design:**

In an information system, input is the raw data that is processed to produce output. During the input design, the developers must consider the input devices such as PC, MICR, OMR, etc.

Therefore, the quality of system input determines the quality of system output. Well-designed input forms and screens have following properties −

* It should serve specific purpose effectively such as storing, recording, and retrieving the information.
* It ensures proper completion with accuracy.
* It should be easy to fill and straightforward.
* It should focus on user’s attention, consistency, and simplicity.
* All these objectives are obtained using the knowledge of basic design principles regarding −
  + What are the inputs needed for the system?
  + How end users respond to different elements of forms and screens.

### Objectives for Input Design:

The objectives of input design are −

* To design data entry and input procedures
* To reduce input volume
* To design source documents for data capture or devise other data capture methods
* To design input data records, data entry screens, user interface screens, etc.
* To use validation checks and develop effective input controls.

**Output Design:**

The design of output is the most important task of any system. During output design, developers identify the type of outputs needed, and consider the necessary output controls and prototype report layouts.

### Objectives of Output Design:

The objectives of input design are:

* To develop output design that serves the intended purpose and eliminates the production of unwanted output.
* To develop the output design that meets the end user’s requirements.
* To deliver the appropriate quantity of output.
* To form the output in appropriate format and direct it to the right person.
* To make the output available on time for making good decisions.

**MODULES:**

1. **User**:
   1. **View Home page:**

Here user view the home page of the Weather application.

* 1. **View about page:**

In the about page, users can learn more about the weather platform.

* 1. **View load page:**

In the load\_data page , the user will load the dataset for modelling.

**View page :**

In view page, the user will see the uploaded dataset.

* 1. **Input Model:**

The user must provide input values for the certain fields in order to get results.

* 1. **View Results:**

User view’s the generated results from the model.

* 1. **View score:**

Here user have ability to view the accuracy score in %

1. **System**
   1. **Working on dataset:**

System checks for data whether it is available or not and load the data in csv files.

* 1. **Pre-processing:**

Data need to be pre-processed according the models it helps to increase the accuracy of the model and better information about the data.

* 1. **Training the data:**

After pre-processing the data will split into two parts as train and test data before training with the given algorithms.

* 1. **Model Building**

To create a model that predicts the personality with better accuracy, this module will help user.

* 1. **Generated Score:**
  2. Here user view the score in %
  3. **Generate Results:**

We train the machine learning algorithm and predict the weather types.

**UML DIAGRAMS**

UML stands for Unified Modelling Language. UML is a standardized general-purpose modelling language in the field of object-oriented software engineering. The standard is managed, and was created by, the Object Management Group.

The goal is for UML to become a common language for creating models of object-oriented computer software. In its current form UML is comprised of two major components: a Meta-model and a notation. In the future, some form of method or process may also be added to; or associated with, UML.

The Unified Modelling Language is a standard language for specifying, Visualization, Constructing and documenting the artefacts of software system, as well as for business modelling and other non-software systems.

The UML represents a collection of best engineering practices that have proven successful in the modelling of large and complex systems.

The UML is a very important part of developing objects-oriented software and the software development process. The UML uses mostly graphical notations to express the design of software projects.

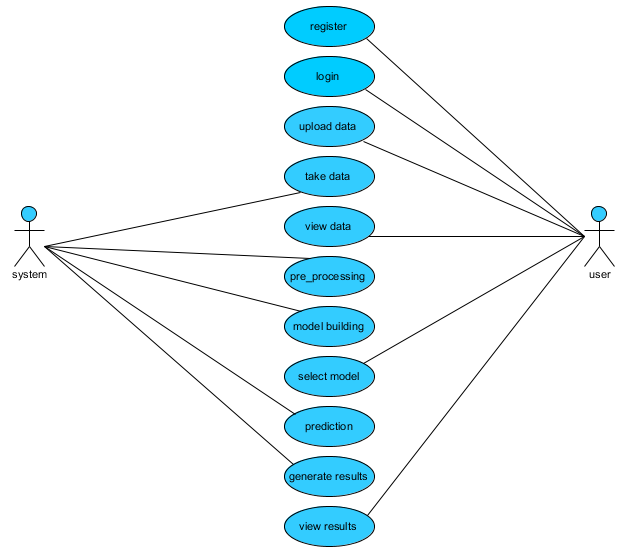
**GOALS:**

The Primary goals in the design of the UML are as follows:

1. Provide users a ready-to-use, expressive visual modelling Language so that they can develop and exchange meaningful models.
2. Provide extendibility and specialization mechanisms to extend the core concepts.
3. Be independent of particular programming languages and development process.
4. Provide a formal basis for understanding the modelling language.
5. Encourage the growth of OO tools market.
6. Support higher level development concepts such as collaborations, frameworks, patterns and components.
7. Integrate best practices.

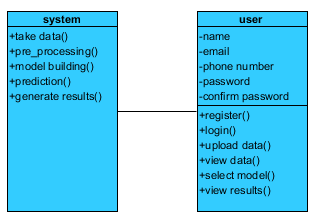
**USE CASE DIAGRAM**

* A use case diagram in the Unified Modeling Language (UML) is a type of behavioral diagram defined by and created from a Use-case analysis.
* Its purpose is to present a graphical overview of the functionality provided by a system in terms of actors, their goals (represented as use cases), and any dependencies between those use cases.
* The main purpose of a use case diagram is to show what system functions are performed for which actor. Roles of the actors in the system can be depicted.

****

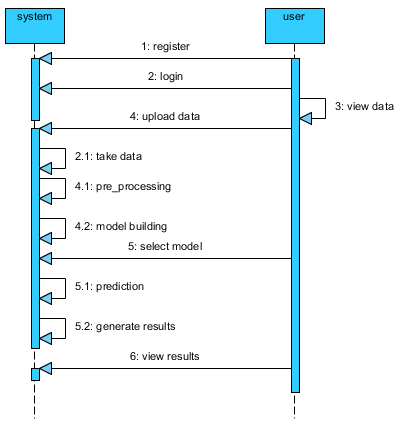
**CLASS DIAGRAM**

In software engineering, 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 the classes. It explains which class contains information



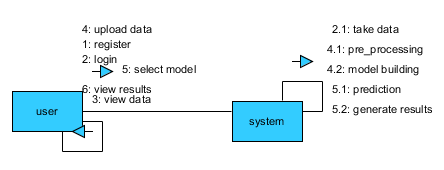
**SEQUENCE DIAGRAM**

* A sequence diagram in Unified Modeling Language (UML) is a kind of interaction diagram that shows how processes operate with one another and in what order.
* It is a construct of a Message Sequence Chart. Sequence diagrams are sometimes called event diagrams, event scenarios, and timing diagrams



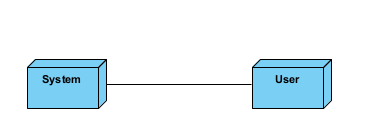
**COLLABORATION DIAGRAM:**

In collaboration diagram the method call sequence is indicated by some numbering technique as shown below. The number indicates how the methods are called one after another. We have taken the same order management system to describe the collaboration diagram. The method calls are similar to that of a sequence diagram. But the difference is that the sequence diagram does not describe the object organization whereas the collaboration diagram shows the object organization.



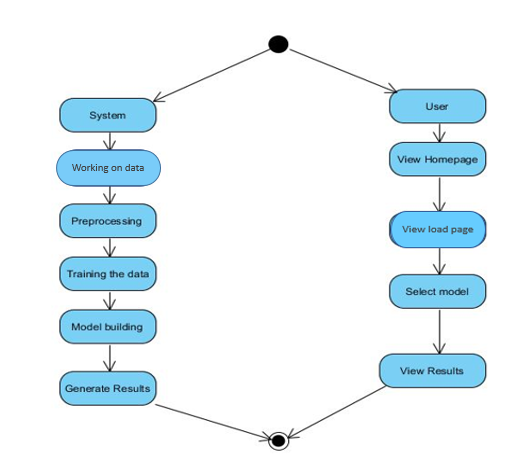
**DEPLOYMENT DIAGRAM**

Deployment diagram represents the deployment view of a system. It is related to the component diagram. Because the components are deployed using the deployment diagrams. A deployment diagram consists of nodes. Nodes are nothing but physical hardware’s used to deploy the application.



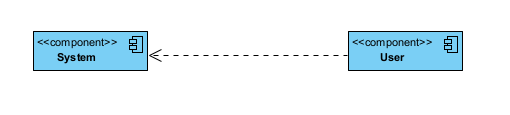
**ACTIVITY DIAGRAM:**

Activity diagrams are graphical representations of workflows of stepwise activities and actions with support for choice, iteration and concurrency. In the Unified Modelling Language, activity diagrams can be used to describe the business and operational step-by-step workflows of components in a system. An activity diagram shows the overall flow of control.



**COMPONENT DIAGRAM**:

A component diagram, also known as a UML component diagram, describes the organization and wiring of the physical **c**omponents in a system. Component diagrams are often drawn to help model implementation details and double-check that every aspect of the system's required function is covered by planned development.



**ER DIAGRAM:**

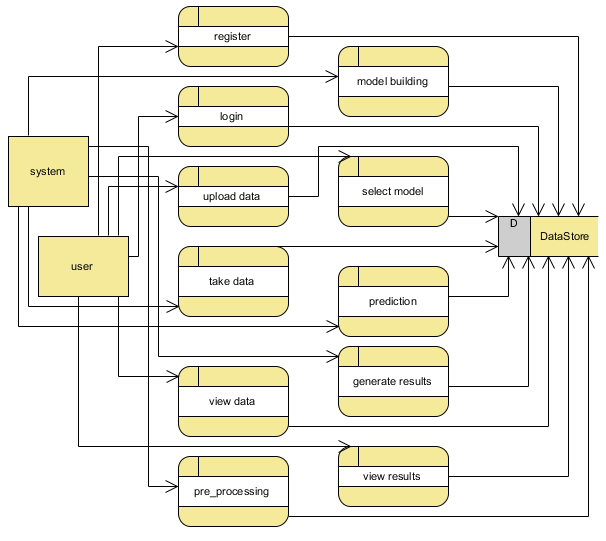
An Entity–relationship model (ER model) describes the structure of a database with the help of a diagram, which is known as Entity Relationship Diagram (ER Diagram). An ER model is a design or blueprint of a database that can later be implemented as a database. The main components of E-R model are: entity set and relationship set.

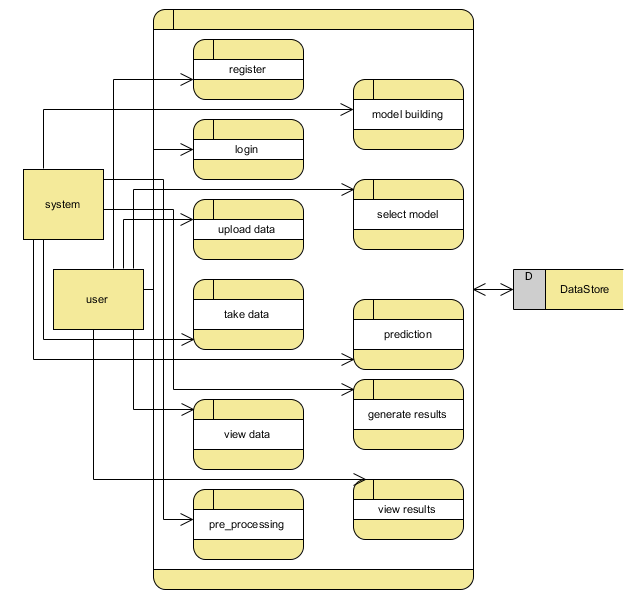
An ER diagram shows the relationship among entity sets. An entity set is a group of similar entities and these entities can have attributes. In terms of DBMS, an entity is a table or attribute of a table in database, so by showing relationship among tables and their attributes, ER diagram shows the complete logical structure of a database. Let’s have a look at a simple ER diagram to understand this concept.

# 

**DFD DIAGRAM:**

A Data Flow Diagram (DFD) is a traditional way to visualize the information flows within a system. A neat and clear DFD can depict a good amount of the system requirements graphically. It can be manual, automated, or a combination of both. It shows how information enters and leaves the system, what changes the information and where information is stored. The purpose of a DFD is to show the scope and boundaries of a system as a whole. It may be used as a communications tool between a systems analyst and any person who plays a part in the system that acts as the starting point for redesigning a system.





**SOURCE CODE:**

**from django.shortcuts import render, redirect**

**from django.contrib.auth.models import User**

**# Create your views here.**

**from django.contrib import messages**

**import pandas as pd**

**import numpy as np**

**from sklearn.preprocessing import LabelEncoder**

**from imblearn.over\_sampling import SMOTE**

**from imblearn.under\_sampling import RandomUnderSampler**

**from sklearn.pipeline import Pipeline**

**from sklearn.metrics import accuracy\_score,classification\_report,confusion\_matrix**

**from sklearn.model\_selection import train\_test\_split**

**from collections.abc import Iterable**

**from . models import \***

**def index(request):**

**return render(request,'index.html')**

**def about(request):**

**return render(request,'about.html')**

**def login(request):**

**if request.method=='POST':**

**lemail=request.POST['email']**

**lpassword=request.POST['password']**

**d=Register.objects.filter(email=lemail,password=lpassword).exists()**

**print(d)**

**return redirect('userhome')**

**else:**

**return render(request,'login.html')**

**def registration(request):**

**if request.method=='POST':**

**Name = request.POST['Name']**

**email=request.POST['email']**

**password=request.POST['password']**

**conpassword=request.POST['conpassword']**

**print(Name,email,password,conpassword)**

**if password==conpassword:**

**rdata=Register(name=Name,email=email,password=password)**

**rdata.save()**

**return render(request,'login.html')**

**else:**

**msg='Register failed!!'**

**return render(request,'registration.html')**

**return render(request,'registration.html')**

**# return render(request,'registration.html')**

**def userhome(request):**

**return render(request,'userhome.html')**

**def load(request):**

**if request.method=="POST":**

**file=request.FILES['file']**

**global df**

**df=pd.read\_csv(file)**

**messages.info(request,"Data Uploaded Successfully")**

**return render(request,'load.html')**

**def view(request):**

**col=df.to\_html**

**dummy=df.head(100)**

**col=dummy.columns**

**rows=dummy.values.tolist()**

**return render(request, 'view.html',{'col':col,'rows':rows})**

**# return render(request,'viewdata.html', {'columns':df.columns.values, 'rows':df.values.tolist()})**

**def preprocessing(request):**

**global X\_train,X\_test,y\_train,y\_test,X,y**

**if request.method == "POST":**

**# size = request.POST['split']**

**size = int(request.POST['split'])**

**size = size / 100**

**df.drop('date',axis=1,inplace=True)**

**from sklearn.preprocessing import LabelEncoder**

**le = LabelEncoder()**

**df['weather'] = le.fit\_transform(df['weather'])**

**#Preprocess Data for Machine Learning Development**

**X = df.drop(['weather'], axis = 1)**

**y = df['weather']**

**over\_strategy = {0 : 1000, 1 : 1000, 2 : 1000, 3 : 1000, 4 : 2000}**

**under\_strategy = {0 : 1000, 1 : 1000, 2 : 1000, 3 : 1000, 4 : 2000}**

**oversample = SMOTE(sampling\_strategy = over\_strategy)**

**undersample = RandomUnderSampler(sampling\_strategy = under\_strategy)**

**X\_final,y = oversample.fit\_resample(X,y)**

**X\_final,y = undersample.fit\_resample(X\_final, y)**

**X\_train,X\_test,y\_train,y\_test = train\_test\_split(X\_final,y,random\_state = 10, test\_size = 0.2)**

**messages.info(request,"Data Preprocessed and It Splits Succesfully")**

**return render(request,'preprocessing.html')**

**def model(request):**

**if request.method == "POST":**

**model = request.POST['algo']**

**if model == "0":**

**from sklearn.ensemble import RandomForestClassifier**

**rf = RandomForestClassifier(n\_estimators=52)**

**rf = rf.fit(X\_train,y\_train)**

**y\_pred = rf.predict(X\_test)**

**acc\_rf=accuracy\_score(y\_test,y\_pred)**

**msg = 'Accuracy of RandomForestClassifier : ' + str(acc\_rf)**

**return render(request,'model.html',{'msg':msg})**

**elif model == "1":**

**from sklearn.tree import DecisionTreeClassifier**

**dt=DecisionTreeClassifier(criterion="entropy",max\_depth=3,random\_state=1245)**

**dt = dt.fit(X\_train,y\_train)**

**y\_pred = dt.predict(X\_test)**

**acc\_dt=accuracy\_score(y\_test,y\_pred)**

**msg = 'Accuracy of DecisionTreeClassifier : ' + str(acc\_dt)**

**return render(request,'model.html',{'msg':msg})**

**elif model == "2":**

**from sklearn.linear\_model import LogisticRegression**

**lr = LogisticRegression(random\_state=100)**

**lr = lr.fit(X\_train,y\_train)**

**y\_pred = lr.predict(X\_test)**

**acc\_lr=accuracy\_score(y\_test,y\_pred)**

**msg = 'Accuracy of LogisticRegression : ' + str(acc\_lr)**

**return render(request,'model.html',{'msg':msg})**

**return render(request,'model.html')**

**def prediction(request):**

**global X\_train,X\_test,y\_train,y\_test,X,y**

**if request.method == 'POST':**

**f1 = float(request.POST['precipitation'])**

**f2 = float(request.POST['temp\_max'])**

**f3 = float(request.POST['temp\_min'])**

**f4 = float(request.POST['wind'])**

**PRED = [[f1,f2,f3,f4]]**

**from sklearn.tree import DecisionTreeClassifier**

**model = DecisionTreeClassifier()**

**model.fit(X\_train,y\_train)**

**xgp = np.array(model.predict(PRED))**

**if xgp==0:**

**msg = ' <span style = color:white;>The Weather is going to be : <span style = color:green;><b>Drizzle</b></span></span>'**

**elif xgp==1:**

**msg = ' <span style = color:white;>The Weather is going to be: <span style = color:red;><b>Fog</b></span></span>'**

**elif xgp==2:**

**msg = ' <span style = color:white;>The Weather is going to be: <span style = color:red;><b>Rain</b></span></span>'**

**elif xgp==3:**

**msg = ' <span style = color:white;>The Weather is going to be: <span style = color:red;><b>Snow</b></span></span>'**

**else :**

**msg = ' <span style = color:white;>The Weather is going to be: <span style = color:red;><b>Sun</b></span></span>'**

**return render(request,'prediction.html',{'msg':msg})**

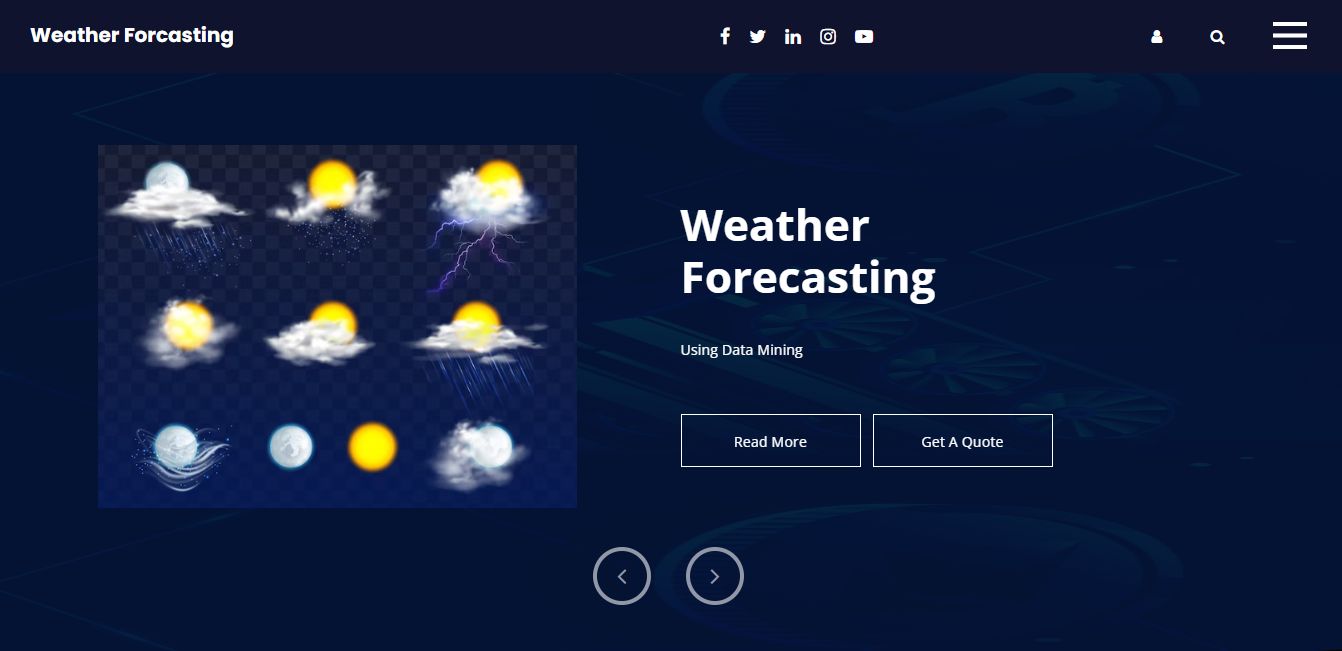
**return render(request,'prediction.html')**

**OUTPUT SCREENS:**

OUTPUT SCREEN SHOTS WITH DESCRIPTION.

**Home Page:**

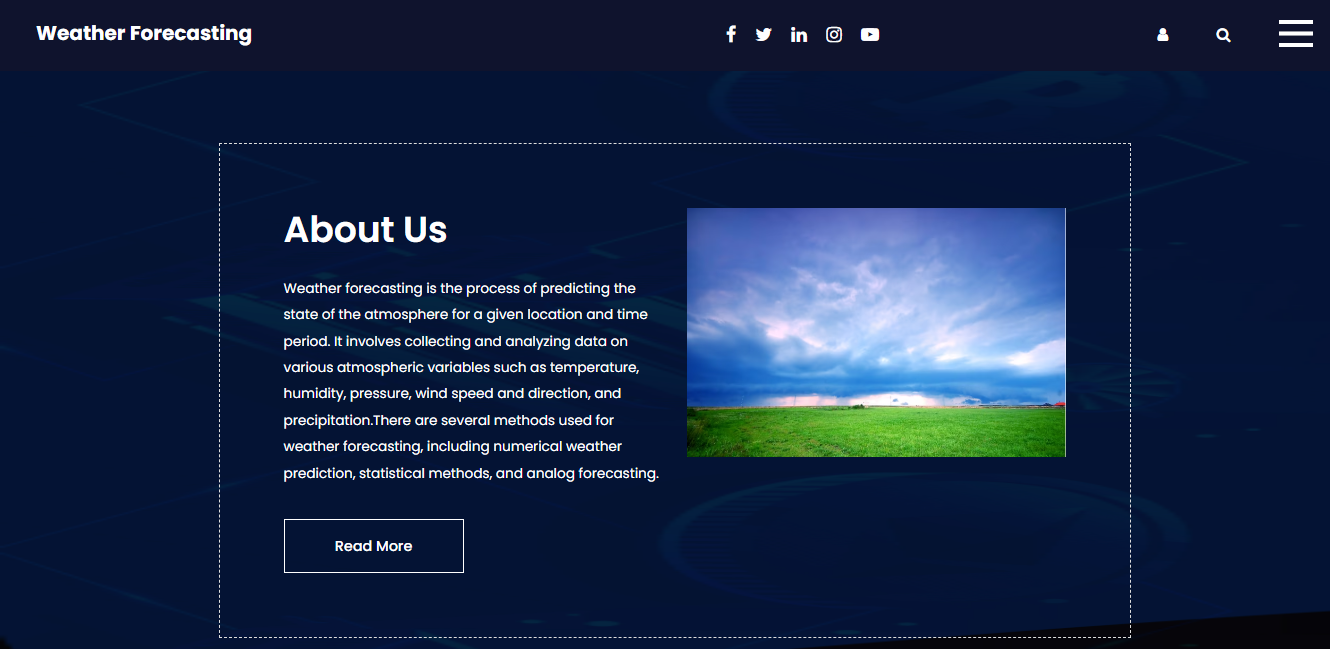
Here user view the home page of web application.



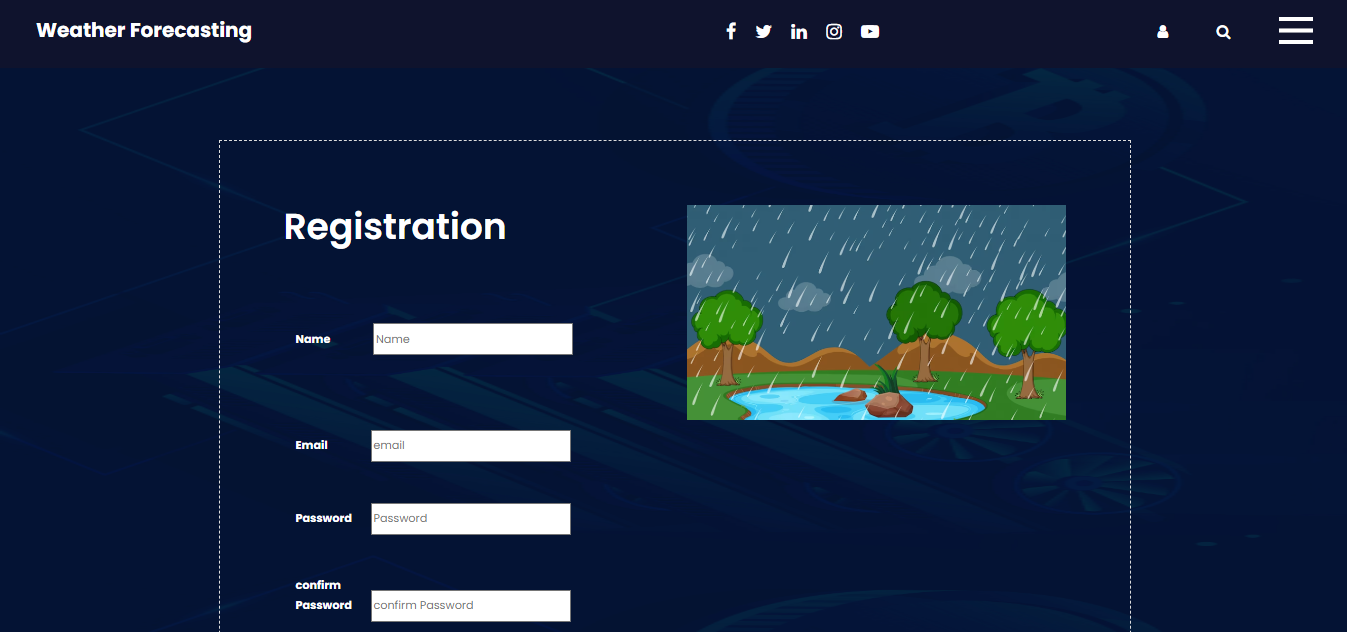
**Fig1**: Home Page

**ABOUT:**

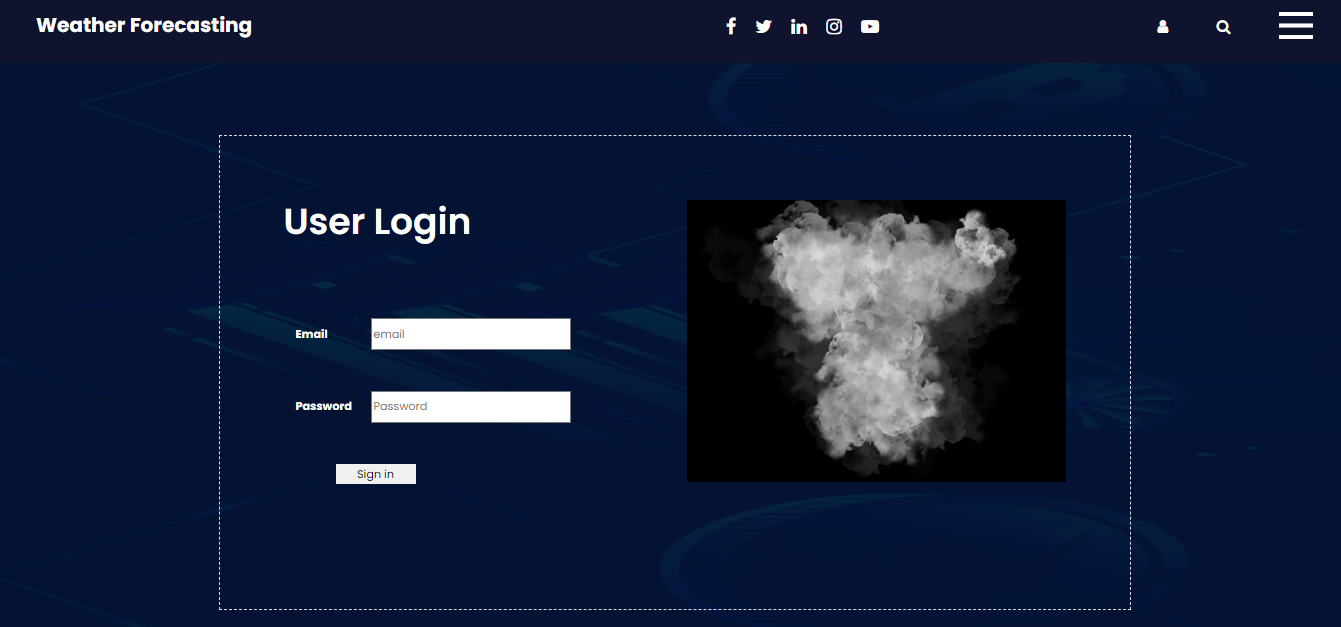
Here we can read about our project.



**Register:**

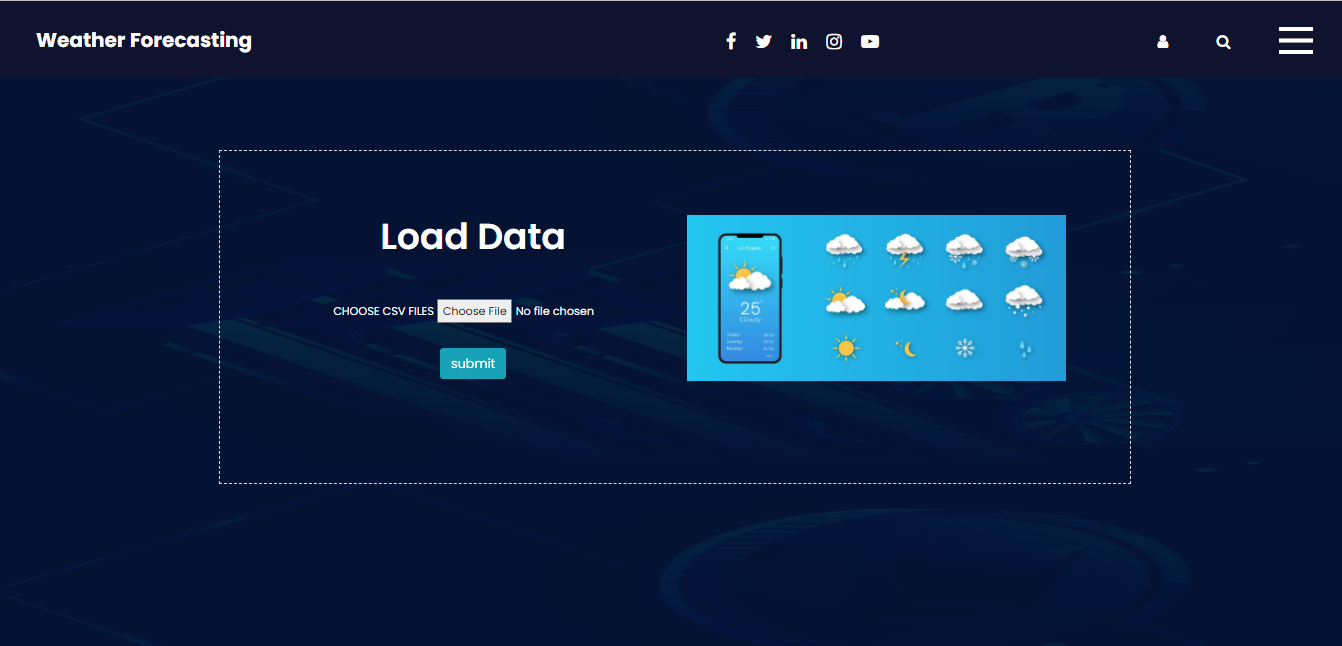
****

**Login:**

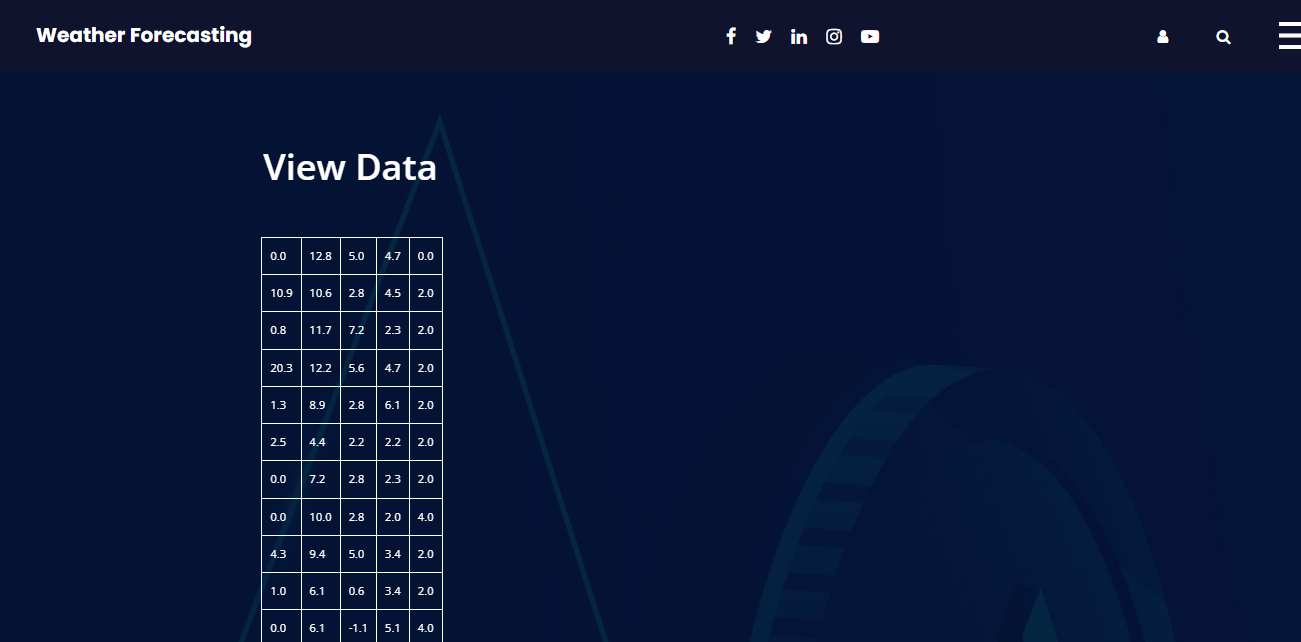
****

**Load:**

In the load page, users can load the dataset.

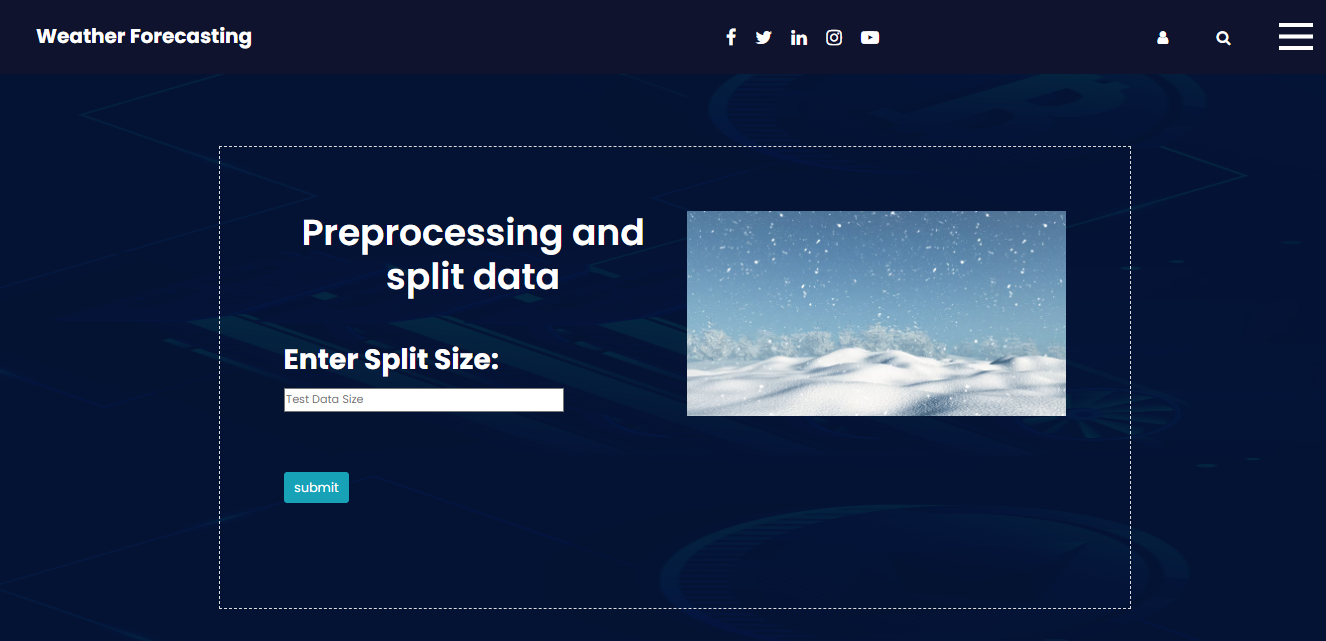


**View:**

****

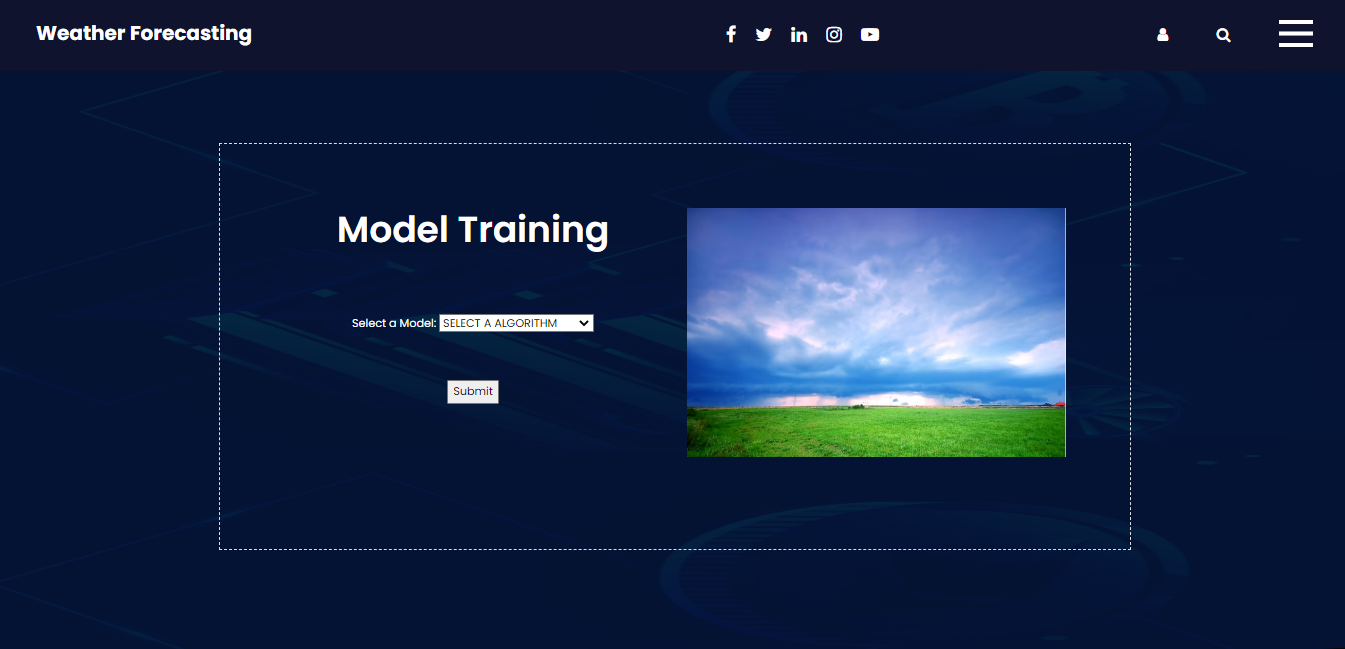
**Pre-process:**

Here we can pre-process and split our data into train and test.



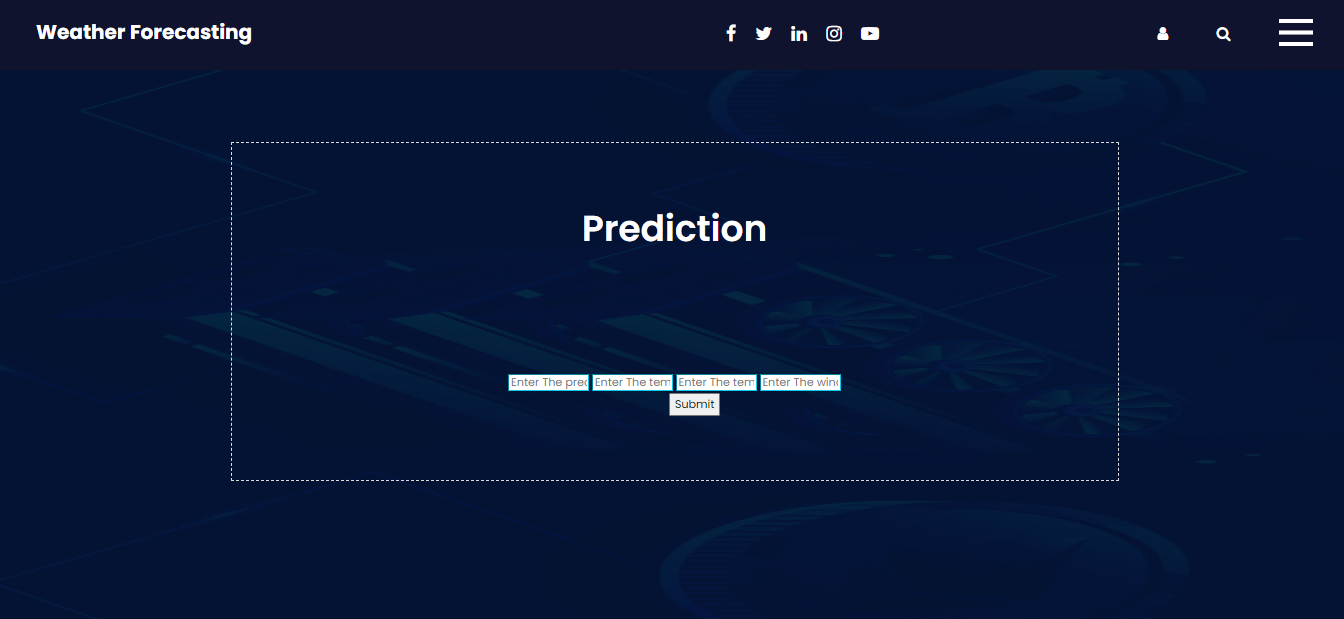
**Model:**

Here we train our data with different ML algorithms.



**Prediction:**

This page show the result of the data.



**BIBLIOGRAPHY:**

**SOFTWARE INSTALLATION FOR MACHINE LEARNING PROJECTS:**

**Installing Python:**

1. To download and install Python visit the official website of Python <https://www.python.org/downloads/> and choose your version.



1. Once the download is complete, run the exe for install Python. Now click on Install Now.
2. You can see Python installing at this point.
3. When it finishes, you can see a screen that says the Setup was successful. Now click on "Close".

**Installing PyCharm:**

1. To download PyCharm visit the website <https://www.jetbrains.com/pycharm/download/> and click the "DOWNLOAD" link under the Community Section.



1. Once the download is complete, run the exe for install PyCharm. The setup wizard should have started. Click “Next”.
2. On the next screen, Change the installation path if required. Click “Next”.
3. On the next screen, you can create a desktop shortcut if you want and click on “Next”.
4. Choose the start menu folder. Keep selected JetBrains and click on “Install”.
5. Wait for the installation to finish.
6. Once installation finished, you should receive a message screen that PyCharm is installed. If you want to go ahead and run it, click the “Run PyCharm Community Edition” box first and click “Finish”.
7. After you click on "Finish," the Following screen will appear.



9. You need to install some packages to execute your project in a proper way.

10. Open the command prompt/ anaconda prompt or terminal as administrator.

11. The prompt will get open, with specified path, type “pip install package name” which you want to install (like NumPy, pandas, sea born, scikit-learn, Matplotlib, Pyplot)

Ex: Pip install NumPy



# **INTRODUCTION TO PYTHON**

* Python

### What Is a Script?

Up to this point, I have concentrated on the interactive programming capability of Python.  This is a very useful capability that allows you to type in a program and to have it executed immediately in an interactive mode

Scripts are reusable

Basically, a script is a text file containing the statements that comprise a Python program.  Once you have created the script, you can execute it over and over without having to retype it each time.

Scripts are editable

Perhaps, more importantly, you can make different versions of the script by modifying the statements from one file to the next using a text editor.  Then you can execute each of the individual versions.  In this way, it is easy to create different programs with a minimum amount of typing.

You will need a text editor

Just about any text editor will suffice for creating Python script files.

You can use *Microsoft Notepad, Microsoft WordPad, Microsoft Word,*or just about any word processor if you want to.

Difference between a script and a program

Script:

Scripts are distinct from the core code of the application, which is usually written in a different language, and are often created or at least modified by the end-user. Scripts are often interpreted from source code or byte code, whereas the applications they control are traditionally compiled to native machine code.

Program:

The program has an executable form that the computer can use directly to execute the instructions.

The same program in its human-readable source code form, from which executable programs are derived (e.g., compiled)

Python

What is Python? Chances you are asking yourself this. You may have found this book because you want to learn to program but don’t know anything about programming languages. Or you may have heard of programming languages like C, C++, C#, or Java and want to know what Python is and how it compares to “big name” languages. Hopefully I can explain it for you.

Python concepts

If you’re not interested in the how’s and whys of Python, feel free to skip to the next chapter. In this chapter I will try to explain to the reader why I think Python is one of the best languages available and why it’s a great one to start programming with.

• Open-source general-purpose language.

• Object Oriented, Procedural, Functional

• Easy to interface with C/ObjC/Java/Fortran

• Easy-is to interface with C++ (via SWIG)

• Great interactive environment

• Great interactive environment

Python is a high-level, interpreted, interactive and object-oriented scripting language. Python is designed to be highly readable. It uses English keywords frequently where as other languages use punctuation, and it has fewer syntactical constructions than other languages.

* Python is Interpreted − Python is processed at runtime by the interpreter. You do not need to compile your program before executing it. This is similar to PERL and PHP.
* Python is Interactive − you can actually sit at a Python prompt and interact with the interpreter directly to write your programs.
* Python is Object-Oriented − Python supports Object-Oriented style or technique of programming that encapsulates code within objects.
* Python is a Beginner's Language − Python is a great language for the beginner-level programmers and supports the development of a wide range of applications from simple text processing to WWW browsers to games.

History of Python

Python was developed by Guido van Possum in the late eighties and early nineties at the National Research Institute for Mathematics and Computer Science in the Netherlands.

Python is derived from many other languages, including ABC, Modula-3, C, C++, Algol-68, Smalltalk, and UNIX shell and other scripting languages.

Python is copyrighted. Like Perl, Python source code is now available under the GNU General Public License (GPL).

Python is now maintained by a core development team at the institute, although Guido van Possum still holds a vital role in directing its progress.

Python Features

Python's features include −

* Easy-to-learn − Python has few keywords, simple structure, and a clearly defined syntax. This allows the student to pick up the language quickly.
* Easy-to-read − Python code is more clearly defined and visible to the eyes.
* Easy-to-maintain − Python's source code is fairly easy-to-maintained.
* A broad standard library − Python's bulk of the library is very portable and cross-platform compatible on UNIX, Windows, and Macintosh.
* Interactive Mode − Python has support for an interactive mode which allows interactive testing and debugging of snippets of code.
* Portable − Python can run on a wide variety of hardware platforms and has the same interface on all platforms.
* Extendable − you can add low-level modules to the Python interpreter. These modules enable programmers to add to or customize their tools to be more efficient.
* Databases − Python provides interfaces to all major commercial databases.
* GUI Programming − Python supports GUI applications that can be created and ported to many system calls, libraries and windows systems, such as Windows MFC, Macintosh, and the X Window system of Unix.
* Scalable − Python provides a better structure and support for large programs than shell scripting.

Apart from the above-mentioned features, Python has a big list of good features, few are listed below −

* It supports functional and structured programming methods as well as OOP.
* It can be used as a scripting language or can be compiled to byte-code for building large applications.
* It provides very high-level dynamic data types and supports dynamic type checking.
* IT supports automatic garbage collection.
* It can be easily integrated with C, C++, COM, ActiveX, CORBA, and Java.

Dynamic vs. Static

Types Python is a dynamic-typed language. Many other languages are static typed, such as C/C++ and Java. A static typed language requires the programmer to explicitly tell the computer what type of “thing” each data value is.

For example, in C if you had a variable that was to contain the price of something, you would have to declare the variable as a “float” type.

This tells the compiler that the only data that can be used for that variable must be a floating point number, i.e. a number with a decimal point.

If any other data value was assigned to that variable, the compiler would give an error when trying to compile the program.

Python, however, doesn’t require this. You simply give your variables names and assign values to them. The interpreter takes care of keeping track of what kinds of objects your program is using. This also means that you can change the size of the values as you develop the program. Say you have another decimal number (a.k.a. a floating point number) you need in your program.

With a static typed language, you have to decide the memory size the variable can take when you first initialize that variable. A double is a floating point value that can handle a much larger number than a normal float (the actual memory sizes depend on the operating environment).

If you declare a variable to be a float but later on assign a value that is too big to it, your program will fail; you will have to go back and change that variable to be a double.

With Python, it doesn’t matter. You simply give it whatever number you want and Python will take care of manipulating it as needed. It even works for derived values.

For example, say you are dividing two numbers. One is a floating point number and one is an integer. Python realizes that it’s more accurate to keep track of decimals so it automatically calculates the result as a floating point number

Variables

Variables are nothing but reserved memory locations to store values. This means that when you create a variable you reserve some space in memory.

Based on the data type of a variable, the interpreter allocates memory and decides what can be stored in the reserved memory. Therefore, by assigning different data types to variables, you can store integers, decimals or characters in these variables.

Standard Data Types

The data stored in memory can be of many types. For example, a person's age is stored as a numeric value and his or her address is stored as alphanumeric characters. Python has various standard data types that are used to define the operations possible on them and the storage method for each of them.

Python has five standard data types −

* Numbers
* String
* List
* Tuple
* Dictionary

## Python Numbers

Number data types store numeric values. Number objects are created when you assign a value to them

## Python Strings

Strings in Python are identified as a contiguous set of characters represented in the quotation marks. Python allows for either pairs of single or double quotes. Subsets of strings can be taken using the slice operator ([ ] and [:]) with indexes starting at 0 in the beginning of the string and working their way from -1 at the end.

## Python Lists

Lists are the most versatile of Python's compound data types. A list contains items separated by commas and enclosed within square brackets ([]). To some extent, lists are similar to arrays in C. One difference between them is that all the items belonging to a list can be of different data type.

The values stored in a list can be accessed using the slice operator ([ ] and [:]) with indexes starting at 0 in the beginning of the list and working their way to end -1. The plus (+) sign is the list concatenation operator, and the asterisk (\*) is the repetition operator.

## Python Tuples

A tuple is another sequence data type that is similar to the list. A tuple consists of a number of values separated by commas. Unlike lists, however, tuples are enclosed within parentheses.

The main differences between lists and tuples are: Lists are enclosed in brackets ([ ]) and their elements and size can be changed, while tuples are enclosed in parentheses (( )) and cannot be updated. Tuples can be thought of as read-only lists.

## Python Dictionary

Python's dictionaries are kind of hash table type. They work like associative arrays or hashes found in Perl and consist of key-value pairs. A dictionary key can be almost any Python type, but are usually numbers or strings. Values, on the other hand, can be any arbitrary Python object.

Dictionaries are enclosed by curly braces ({ }) and values can be assigned and accessed using square braces ([]).

Different modes in python

Python has two basic modes: normal and interactive.

The normal mode is the mode where the scripted and finished .pie files are run in the Python interpreter.

Interactive mode is a command line shell which gives immediate feedback for each statement, while running previously fed statements in active memory. As new lines are fed into the interpreter, the fed program is evaluated both in part and in whole

# 20 Python libraries

1. Requests. The most famous http library written by Kenneth remits. It’s a must have for every python developer.

2. Scrappy. If you are involved in web scraping then this is a must have library for you. After using this library you won’t use any other.

3. Python. A guy toolkit for python. I have primarily used it in place of tinder. You will really love it.

4. Pillow. A friendly fork of PIL (Python Imaging Library). It is more user friendly than PIL and is a must have for anyone who works with images.

5. SQL Alchemy. A database library. Many love it and many hate it. The choice is yours.

6. Beautiful Soup. I know it’s slow but this xml and html parsing library is very useful for beginners.

7. Twisted. The most important tool for any network application developer. It has a very beautiful ape and is used by a lot of famous python developers.

8. Numbly. How can we leave this very important library? It provides some advance math functionalities to python.

9. Skippy. When we talk about numbly then we have to talk about spicy. It is a library of algorithms and mathematical tools for python and has caused many scientists to switch from ruby to python.

10. Matplotlib. A numerical plotting library. It is very useful for any data scientist or any data analyser.

11. Pygmy. Which developer does not like to play games and develop them? This library will help you achieve your goal of 2d game development.

12. Piglet. A 3d animation and game creation engine. This is the engine in which the famous [python port](https://github.com/fogleman/Minecraft) of mine craft was made

13. Pit. A GUI toolkit for python. It is my second choice after python for developing GUI’s for my python scripts.

14. Pit. Another python GUI library. It is the same library in which the famous Bit torrent client is created.

15. Scaly. A packet sniffer and analyser for python made in python.

16. Pywin32. A python library which provides some useful methods and classes for interacting with windows.

17. Notch. Natural Language Toolkit – I realize most people won’t be using this one, but it’s generic enough. It is a very useful library if you want to manipulate strings. But its capacity is beyond that. Do check it out.

18. Nose. A testing framework for python. It is used by millions of python developers. It is a must have if you do test driven development.

19. Simply. Simply can-do algebraic evaluation, differentiation, expansion, complex numbers, etc. It is contained in a pure Python distribution.

20. I Python. I just can’t stress enough how useful this tool is. It is a python prompt on steroids. It has completion, history, shell capabilities, and a lot more. Make sure that you take a look at it.

NumPy

Humpy’s main object is the homogeneous multidimensional array. It is a table of elements (usually numbers), all of the same type, indexed by a tuple of positive integers. In numbly dimensions are called axes. The number of axes is rank.

• Offers Matlab-ish capabilities within Python

• Fast array operations

• 2D arrays, multi-D arrays, linear algebra etc.

Matplotlib

• High quality plotting library.

Python class and objects

These are the building blocks of OOP. Class creates a new object. This object can be anything, whether an abstract data concept or a model of a physical object, e.g. a chair. Each class has individual characteristics unique to that class, including variables and methods. Classes are very powerful and currently “the big thing” in most programming languages. Hence, there are several chapters dedicated to OOP later in the book.

The class is the most basic component of object-oriented programming. Previously, you learned how to use functions to make your program do something.

Now will move into the big, scary world of Object-Oriented Programming (OOP). To be honest, it took me several months to get a handle on objects.

When I first learned C and C++, I did great; functions just made sense for me.

Having messed around with BASIC in the early ’90s, I realized functions were just like subroutines so there wasn’t much new to learn.

However, when my C++ course started talking about objects, classes, and all the new features of OOP, my grades definitely suffered.

Once you learn OOP, you’ll realize that it’s actually a pretty powerful tool. Plus many Python libraries and APIs use classes, so you should at least be able to understand what the code is doing.

One thing to note about Python and OOP: it’s not mandatory to use objects in your code in a way that works best; maybe you don’t need to have a full-blown class with initialization code and methods to just return a calculation. With Python, you can get as technical as you want.

As you’ve already seen, Python can do just fine with functions. Unlike languages such as Java, you aren’t tied down to a single way of doing things; you can mix functions and classes as necessary in the same program. This lets you build the code

Objects are an encapsulation of variables and functions into a single entity. Objects get their variables and functions from classes. Classes are essentially a template to create your objects.

Here’s a brief list of Python OOP ideas:

• The class statement creates a class object and gives it a name. This creates a new namespace.

• Assignments within the class create class attributes. These attributes are accessed by qualifying the name using dot syntax: ClassName.Attribute.

• Class attributes export the state of an object and its associated behaviour. These attributes are shared by all instances of a class.

• Calling a class (just like a function) creates a new instance of the class.

This is where the multiple copy’s part comes in.

• Each instance gets ("inherits") the default class attributes and gets its own namespace. This prevents instance objects from overlapping and confusing the program.

• Using the term self identifies a particular instance, allowing for per-instance attributes. This allows items such as variables to be associated with a particular instance.

Inheritance

First off, classes allow you to modify a program without really making changes to it.

To elaborate, by sub classing a class, you can change the behaviour of the program by simply adding new components to it rather than rewriting the existing components.

As we’ve seen, an instance of a class inherits the attributes of that class.

However, classes can also inherit attributes from other classes. Hence, a subclass inherits from a superclass allowing you to make a generic superclass that is specialized via subclasses.

The subclasses can override the logic in a superclass, allowing you to change the behaviour of your classes without changing the superclass at all.

Operator Overloads

Operator overloading simply means that objects that you create from classes can respond to actions (operations) that are already defined within Python, such as addition, slicing, printing, etc.

Even though these actions can be implemented via class methods, using overloading ties the behavior closer to Python’s object model and the object interfaces are more consistent to Python’s built-in objects, hence overloading is easier to learn and use.

User-made classes can override nearly all of Python’s built-in operation methods

Exceptions

I’ve talked about exceptions before but now I will talk about them in depth. Essentially, exceptions are events that modify program’s flow, either intentionally or due to errors.

They are special events that can occur due to an error, e.g. trying to open a file that doesn’t exist, or when the program reaches a marker, such as the completion of a loop.

Exceptions, by definition, don’t occur very often; hence, they are the "exception to the rule" and a special class has been created for them. Exceptions are everywhere in Python.

Virtually every module in the standard Python library uses them, and Python itself will raise them in a lot of different circumstances.

Here are just a few examples:

• Accessing a non−existent dictionary key will raise a Key Error exception.

• Searching a list for a non−existent value will raise a Value Error exception

. • Calling a non−existent method will raise an Attribute Error exception.

• Referencing a non−existent variable will raise a Name Error exception.

• Mixing data types without coercion will raise a Type Error exception.

One use of exceptions is to catch a fault and allow the program to continue working; we have seen this before when we talked about files.

This is the most common way to use exceptions. When programming with the Python command line interpreter, you don’t need to worry about catching exceptions.

Your program is usually short enough to not be hurt too much if an exception occurs.

Plus, having the exception occur at the command line is a quick and easy way to tell if your code logic has a problem.

However, if the same error occurred in your real program, it will fail and stop working. Exceptions can be created manually in the code by raising an exception.

It operates exactly as a system-caused exceptions, except that the programmer is doing it on purpose. This can be for a number of reasons. One of the benefits of using exceptions is that, by their nature, they don’t put any overhead on the code processing.

Because exceptions aren’t supposed to happen very often, they aren’t processed until they occur.

Exceptions can be thought of as a special form of the if/elf statements. You can realistically do the same thing with if blocks as you can with exceptions.

However, as already mentioned, exceptions aren’t processed until they occur; if blocks are processed all the time.

Proper use of exceptions can help the performance of your program.

The more infrequent the error might occur, the better off you are to use exceptions; using if blocks requires Python to always test extra conditions before continuing.

Exceptions also make code management easier: if your programming logic is mixed in with error-handling if statements, it can be difficult to read, modify, and debug your program.

User-Defined Exceptions

I won’t spend too much time talking about this, but Python does allow for a programmer to create his own exceptions.

You probably won’t have to do this very often but it’s nice to have the option when necessary.

However, before making your own exceptions, make sure there isn’t one of the built-in exceptions that will work for you.

They have been "tested by fire" over the years and not only work effectively, they have been optimized for performance and are bug-free.

Making your own exceptions involves object-oriented programming, which will be covered in the next chapter

. To make a custom exception, the programmer determines which base exception to use as the class to inherit from, e.g. making an exception for negative numbers or one for imaginary numbers would probably fall under the Arithmetic Error exception class.

To make a custom exception, simply inherit the base exception and define what it will do.

Python modules

Python allows us to store our code in files (also called modules). This is very useful for more serious programming, where we do not want to retype a long function definition from the very beginning just to change one mistake. In doing this, we are essentially defining our own modules, just like the modules defined already in the Python library.

To support this, Python has a way to put definitions in a file and use them in a script or in an interactive instance of the interpreter. Such a file is called a module; definitions from a module can be imported into other modules or into the main module.

Testing code

As indicated above, code is usually developed in a file using an editor.

To test the code, import it into a Python session and try to run it.

Usually there is an error, so you go back to the file, make a correction, and test again.

This process is repeated until you are satisfied that the code works. T

His entire process is known as the development cycle.

There are two types of errors that you will encounter. Syntax errors occur when the form of some command is invalid.

This happens when you make typing errors such as misspellings, or call something by the wrong name, and for many other reasons. Python will always give an error message for a syntax error.

Functions in Python

It is possible, and very useful, to define our own functions in Python. Generally speaking, if you need to do a calculation only once, then use the interpreter. But when you or others have need to perform a certain type of calculation many times, then define a function.

You use functions in programming to bundle a set of instructions that you want to use repeatedly or that, because of their complexity, are better self-contained in a sub-program and called when needed. That means that a function is a piece of code written to carry out a specified task.

## To carry out that specific task, the function might or might not need multiple inputs. When the task is carved out, the function can or cannot return one or more values.

## There are three types of functions in python:

## Help (), min (), print ().

Namespaces in Python are implemented as Python dictionaries, this means it is a mapping from names (keys) to objects (values). The user doesn't have to know this to write a Python program and when using namespaces.

Some namespaces in Python:

* global names of a module
* local names in a function or method invocation
* built-in names: this namespace contains built-in functions (e.g. abs(), camp(), ...) and built-in exception names

Garbage Collection

Garbage Collector exposes the underlying memory management mechanism of Python, the automatic garbage collector. The module includes functions for controlling how the collector operates and to examine the objects known to the system, either pending collection or stuck in reference cycles and unable to be freed.

Python XML Parser

XML is a portable, open source language that allows programmers to develop applications that can be read by other applications, regardless of operating system and/or developmental language.

What is XML? The Extensible Markup Language XML is a markup language much like HTML or SGML.

This is recommended by the World Wide Web Consortium and available as an open standard.

XML is extremely useful for keeping track of small to medium amounts of data without requiring a SQL-based backbone.

XML Parser Architectures and APIs the Python standard library provides a minimal but useful set of interfaces to work with XML.

The two most basic and broadly used APIs to XML data are the SAX and DOM interfaces.

Simple API for XML SAX: Here, you register callbacks for events of interest and then let the parser proceed through the document.

This is useful when your documents are large or you have memory limitations, it parses the file as it reads it from disk and the entire file is never stored in memory.

Document Object Model DOM API : This is a World Wide Web Consortium recommendation wherein the entire file is read into memory and stored in a hierarchical tree − based form to represent all the features of an XML document.

SAX obviously cannot process information as fast as DOM can when working with large files. On the other hand, using DOM exclusively can really kill your resources, especially if used on a lot of small files.

SAX is read-only, while DOM allows changes to the XML file. Since these two different APIs literally complement each other, there is no reason why you cannot use them both for large projects.

Python Web Frameworks

A web framework is a code library that makes a developer's life easier when building reliable, scalable and maintainable web applications.

## Why are web frameworks useful?

Web frameworks encapsulate what developers have learned over the past twenty years while programming sites and applications for the web. Frameworks make it easier to reuse code for common HTTP operations and to structure projects so other developers with knowledge of the framework can quickly build and maintain the application.

Common web framework functionality

Frameworks provide functionality in their code or through extensions to perform common operations required to run web applications. These common operations include:

1. URL routing
2. HTML, XML, JSON, and other output format tinplating
3. Database manipulation
4. Security against Cross-site request forgery (CSRF) and other attacks
5. Session storage and retrieval

Not all web frameworks include code for all of the above functionality. Frameworks fall on the spectrum from executing a single use case to providing every known web framework feature to every developer. Some frameworks take the "batteries-included" approach where everything possible comes bundled with the framework while others have a minimal core package that is amenable to extensions provided by other packages.

## Comparing web frameworks

There is also a repository called [compare-python-web-frameworks](https://github.com/mattmakai/compare-python-web-frameworks) where the same web application is being coded with varying Python web frameworks, tinplating engines and object.

## Web framework resources

* When you are learning how to use one or more web frameworks it's helpful to have an idea of what the code under the covers is doing.
* Frameworks is a really well done short video that explains how to choose between web frameworks. The author has some particular opinions about what should be in a framework. For the most part I agree although I've found sessions and database ORMs to be a helpful part of a framework when done well.
* What is a web framework? Is an in-depth explanation of what web frameworks are and their relation to web servers?
* Jingo vs. Flash vs. Pyramid: Choosing a Python web framework contains background information and code comparisons for similar web applications built in these three big Python frameworks.
* This fascinating blog post takes a look at the code complexity of several Python web frameworks by providing visualizations based on their code bases.
* Python’s web frameworks benchmarks  is a test of the responsiveness of a framework with encoding an object to JSON and returning it as a response as well as retrieving data from the database and rendering it in a template. There were no conclusive results but the output is fun to read about nonetheless.
* What web frameworks do you use and why are they awesome? Is a language agnostic Reedit discussion on web frameworks? It's interesting to see what programmers in other languages like and dislike about their suite of web frameworks compared to the main Python frameworks.
* This user-voted question & answer site asked "What are the best general purpose Python web frameworks usable in production?” The votes aren't as important as the list of the many frameworks that are available to Python developers.

## Web frameworks learning checklist

1. Choose a major Python web framework (Jingo or Flask are recommended) and stick with it. When you're just starting it's best to learn one framework first instead of bouncing around trying to understand every framework.
2. Work through a detailed tutorial found within the resources links on the framework's page.
3. Study open source examples built with your framework of choice so you can take parts of those projects and reuse the code in your application.
4. Build the first simple iteration of your web application then go to the [deployment](https://www.fullstackpython.com/deployment.html) section to make it accessible on the web.

### SYSTEM TESTING

The purpose of testing is to discover errors. Testing is the process of trying to discover every conceivable fault or weakness in a work product. It provides a way to check the functionality of components, sub-assemblies, assemblies and/or a finished product It is the process of exercising software with the intent of ensuring that the

Software system meets its requirements and user expectations and does not fail in an unacceptable manner. There are various types of test. Each test type addresses a specific testing requirement.

TYPES OF TESTS

Unit testing

Unit testing involves the design of test cases that validate that the internal program logic is functioning properly, and that program inputs produce valid outputs. All decision branches and internal code flow should be validated. It is the testing of individual software units of the application .it is done after the completion of an individual unit before integration. This is a structural testing, that relies on knowledge of its construction and is invasive. Unit tests perform basic tests at component level and test a specific business process, application, and/or system configuration. Unit tests ensure that each unique path of a business process performs accurately to the documented specifications and contains clearly defined inputs and expected results.

Integration testing

Integration tests are designed to test integrated software components to determine if they actually run as one program. Testing is event driven and is more concerned with the basic outcome of screens or fields. Integration tests demonstrate that although the components were individually satisfaction, as shown by successfully unit testing, the combination of components is correct and consistent. Integration testing is specifically aimed at exposing the problems that arise from the combination of components.

Functional test

Functional tests provide systematic demonstrations that functions tested are available as specified by the business and technical requirements, system documentation, and user manuals.

Functional testing is centered on the following items:

Valid Input : identified classes of valid input must be accepted.

Invalid Input : identified classes of invalid input must be rejected.

Functions : identified functions must be exercised.

Output : identified classes of application outputs must be exercised.

Systems/Procedures: interfacing systems or procedures must be invoked.

Organization and preparation of functional tests is focused on requirements, key functions, or special test cases. In addition, systematic coverage pertaining to identify Business process flows; data fields, predefined processes, and successive processes must be considered for testing. Before functional testing is complete, additional tests are identified and the effective value of current tests is determined.

SYSTEM TEST

System testing ensures that the entire integrated software system meets requirements. It tests a configuration to ensure known and predictable results. An example of system testing is the configuration-oriented system integration test. System testing is based on process descriptions and flows, emphasizing pre-driven process links and integration points.

White Box Testing

White Box Testing is a testing in which in which the software tester has knowledge of the inner workings, structure and language of the software, or at least its purpose. It is purpose. It is used to test areas that cannot be reached from a black box level.

Black Box Testing

Black Box Testing is testing the software without any knowledge of the inner workings, structure or language of the module being tested. Black box tests, as most other kinds of tests, must be written from a definitive source document, such as specification or requirements document, such as specification or requirements document. It is a testing in which the software under test is treated, as a black box. you cannot “see” into it. The test provides inputs and responds to outputs without considering how the software works.

6.1 Unit Testing:

Unit testing is usually conducted as part of a combined code and unit test phase of the software lifecycle, although it is not uncommon for coding and unit testing to be conducted as two distinct phases.

Test strategy and approach

Field testing will be performed manually and functional tests will be written in detail.

Test objectives

* All field entries must work properly.
* Pages must be activated from the identified link.
* The entry screen, messages and responses must not be delayed.

Features to be tested

* Verify that the entries are of the correct format
* No duplicate entries should be allowed
* All links should take the user to the correct page.

# 6.2 Integration Testing

Software integration testing is the incremental integration testing of two or more integrated software components on a single platform to produce failures caused by interface defects.

The task of the integration test is to check that components or software applications, e.g. components in a software system or – one step up – software applications at the company level – interact without error.

Test Results: All the test cases mentioned above passed successfully. No defects encountered.

6.3 Acceptance Testing

User Acceptance Testing is a critical phase of any project and requires significant participation by the end user. It also ensures that the system meets the functional requirements.

Test Results: All the test cases mentioned above passed successfully. No defects encountered.

**CONCLUSION:**

In conclusion, machine learning has revolutionized the field of weather forecasting. By using machine learning algorithms to analyse vast amounts of historical weather data, meteorologists can now make more accurate and timely predictions of future weather conditions. Machine learning models can learn complex relationships between different weather variables and use that knowledge to make accurate predictions in real-time. This is critical for various sectors such as agriculture, transportation, and public safety, where accurate weather forecasting can help save lives, reduce costs, and increase efficiency. Machine learning is also helping to advance the field of weather forecasting by providing new insights into the behavior of the atmosphere and helping to identify new weather patterns. Overall, data mining using machine learning is a powerful tool for weather forecasting and will continue to play an essential role in providing accurate and timely weather information to society.

**REFERENCES:**

[1] Pushpa Mohan and Dr. Kiran KumariPatil: “Survey on Crop and Weather Forecasting based on Agriculture related Statistical Data”, International Journal of Innovative Research in Computer and Communication Engineering, Volume 5, Issue 2

[2] Amruta A. Taksande and P. S. Mohod: “Applications of Data Mining in Weather Forecasting Using Frequent Pattern Growth Algorithm”, International Journal of Science and Research

[3] Fahad Sheikh and S. Karthick: “Analysis of Data Mining Techniques for Weather Prediction”,Indian Journal of Science and Technology, Volume 9

[4] PrashantBiradar, Sarfraz Ansari, YashavantParadkar and Savita Lohiya:“Weather Prediction Using Data Mining”,International Journal of engineering research and development,Volume 5, Issue 2

[5] AishwaryaDhore, AnaghaByakude, Bhagyashri Sonar and Mansi Waste: “ Weather prediction using the data mining Technique”