MACHINE LEARNING BASED FLOOD PREDICTION

Team Members:

M.Maneesh - 19BCE7192

K.Kamal - 19BCE7324

K.Lokesh - 19BCN7242

Guided by:

Dr. Snehal Bankatrao Shinde

INDEX

ABSTRACT

- Flooding is the most common natural disaster on the planet, affecting hundreds of millions of people and causing between 6,000 and 18,000 fatalities every year – of which 20 percent are in India.
- Reliable early warning systems have been shown to prevent a significant fraction of fatalities and economic damage, but many people don’t have access to those types of warning systems. So, we're building Flood prediction system Based on ML or AI.
- This advancement of the prediction system provides cost-effective solutions and better performance. In this, a prediction model is constructed using rainfall data to predict the occurrence of floods due to rainfall.
- The model predicts whether “flood may happen or not” based on the rainfall range for particular locations. Indian district rainfall data is used to build the prediction model. The dataset is trained with various algorithms like K-Nearest Neighbors, XGBoost etc.,

INTRODUCTION

- Every year, India is the topmost flood-prone disaster place in the world. Mostly water logging in urban cities occurs in low-lying areas. Moreover, the increase in water logging is due to some fundamental points such as surface runoff, relative altitude, and not enough path of the water to drainage So, flood forecasting is essential at these places.
- In a recent year, there were many parts of countries which are prone to flood like Assam, Bihar, Goa, Odisha, Pune, Maharashtra, Tamil Nadu, Karnataka, Kerala, and Gujarat.
- In the year 2015 rainfall, Chennai received 1049 millimeters (mm) of rainfall in November. Since 1918, 1088 mm of precipitation was the best recorded in November. Between October and December, the average rainfall in Kanchipuram district is 64 cm. It received the heaviest rainfall of 181.5 cm, which is 183% higher against average precipitation. In the Tiruvallur district, the average rainfall is 59 cm but recorded 146 cm of rain

- There was much research for prediction of flood ahead, but not many methods give the estimate with high accuracy. The flood prediction analysis majorly uses Machine Learning (ML). There are many methods in machine learning to predict the problem with higher accuracy.

- In this work, we have proposed to estimate the flash flood to prevent places that are prone to flood risk. The approach is to the establishment of the ML algorithm model. It incorporates the flood factor to estimate short term prediction in an urban area with higher accuracy.

LITERATURE SURVEY

S.No	Title	Authors	Journal Type and year	Algorithms used	Outcome
1.	A Hybrid Machine Learning Approach for Classifying Aerial Images of Flood-Hit Areas	J. Akshya and P. L. K. Priyadarsini	(ICCIDS) 2019	SVM and K-means clustering	Accuracy of classifying 92% of flooded areas.
2.	Different Techniques of Flood Forecasting and Their Applications	A. B. Ranit and P. V. Durge.	(RICE) 2018	Statistical Approach ANN Approach Clustering Approach	Flood Forecasting and Warning system is implemented as a aspects of technical and non-technical and they are seamlessly integrated.

S.No	Title	Authors	Journal Type and year	Algorithms used	Outcome
3	Multiple Input Single Output (MISO) ARX and ARMAX model of flood prediction system: Case study Pahang	F. A. Ruslan, K. Haron, A. M. Samad and R. Adnan	(CSPA) 2017	ARX "Autoregressive with Exogenous input" model structure. ARMAX "Autoregressive Moving Average with Exogenous Input" model structure.	Best Fit (%) ARX: 50.41 ARMAX 63.06
4	Flood Prediction Using Multi-Layer Artificial Neural Network in Monitoring System with Rain Gauge, Water Level, Soil Moisture Sensors	F. R. G. Cruz, M. G. Binag, M. R. G. Ga and F. A. A. Uy	IEEE Region 10 Conference, 2018	Neural Network Tool in MATLAB was used to implement artificial neural network. Supervised learning was implemented in training. The gathered data is divided into three parts. 70 % of the data were used for training, 15 % was used for testing and the other 15 % was used for validation.	The root mean square deviation (RMSDwas used to determine the degree of difference between the predicted water level and the actual water level values. The flood prediction model had a small deviation with reference to the actual water level, with RMSD of 2.2648.

S.No	Title	Authors	Journal Type and year	Algorithms used	Outcome
5	An Efficient Automated Hybrid Algorithm to Predict Floods in Cloud Environment	G. Kaur and A. Bala,	IEEE (CCECE) 2019	Genetic Algorithm (GA) is a meta-heuristic algorithm for selecting the optimal set of features in an efficient way. GA-SVM, GA-BPM, GA-NN and GA-DF	standalone machine model is 86.36%, 85.23%, 82.39% and 77.27% respectively. cloud environment is 88.57%, 88.57%, 88%,and 84.57% respectively.
6	Developing a Flood Risk Assessment Using Support Vector Machine and Convolutional Neural Network: A Conceptual Framework	J. M. A. Opella and A. A. Hernandez	IEEE 15th International (CSPA) 2019	Convolutional neural network (CNN) and support vector machine (SVM)	The fusion of these two distinct network architecture will produce an effective and robust flood map .

EXISTING METHOD

- The increasing growth of machine learning, computer techniques divided into traditional methods and machine learning methods.
- This section describes the related works of flood predictions and how machine learning methods are better than traditional methods.
- The existing method in this project have a certain flow and also SVM is used for model development. But it requires large memory and result is not accurate.

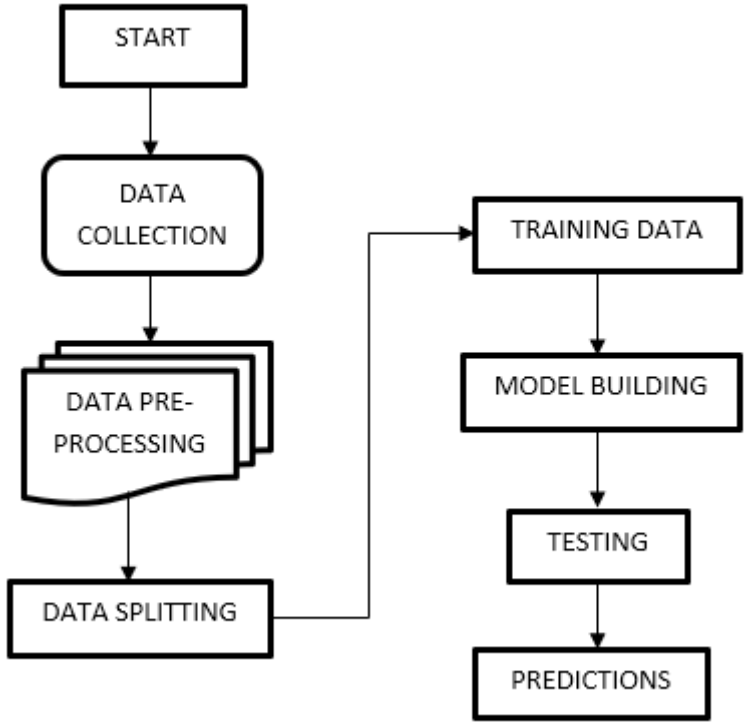
DISADVANTAGES

- More Computation memory.
- Time Consuming.
- Difficult to handle.

PROPOSED SYSTEM

- We propose this application that can be considered a useful system since it helps to reduce the limitations obtained from traditional and other existing methods.
- In proposed system, we implement a Machine Learning algorithms for getting insights from the complex patterns in the data in order to predict the floods.
- This technique is computationally inexpensive because of its simple architecture.

BLOCK DIAGRAM



ADVANTAGES

- High accuracy.
- Time Saving.
- Computationally inexpensive.
- Low complexities

IMPLEMENTATION

- Firstly, Collect the datasets available in open resources.
- Perform Machine Learning algorithms to the publicly available dataset
- Choose the best algorithm based on the performance of the above two models.
- Apply the best model to our target data.
- After the model building, based on the input values provided by the user our model predicts whether the customer will continue with this network or not.

MODULES

- **1. User:**
- **1.1 Upload:**

  User has ability to upload the dataset for the model building.
- **1.2 Model Selection:**

  User should selects the machine learning model for training.
- **1.2 prediction:**

  User needs to enter input in order to detect the desire output
- **1.3 View Results:**

  User has ability to view the results generated by the system.

- **2. System**

- **2.1 Take the dataset:**

  System works with the dataset provided to it for model building.

- **2.2 Preprocessing:**

  In preprocessing step system works with to impute any disorders in the data set and extract the features.

- **2.3 Model Training:**

  In training phase system generates the model from the dataset by using python modules.

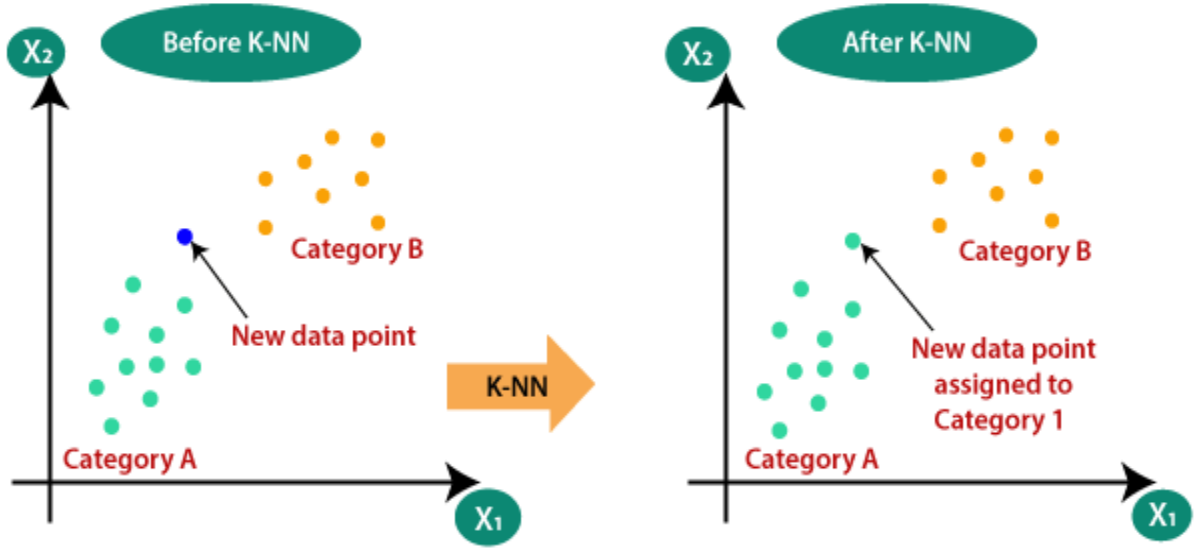
- **2.4 Generate Results:**

  System generates the detection results from the model whether the there is a chance of floods occurring or not.

ALGORITHMS

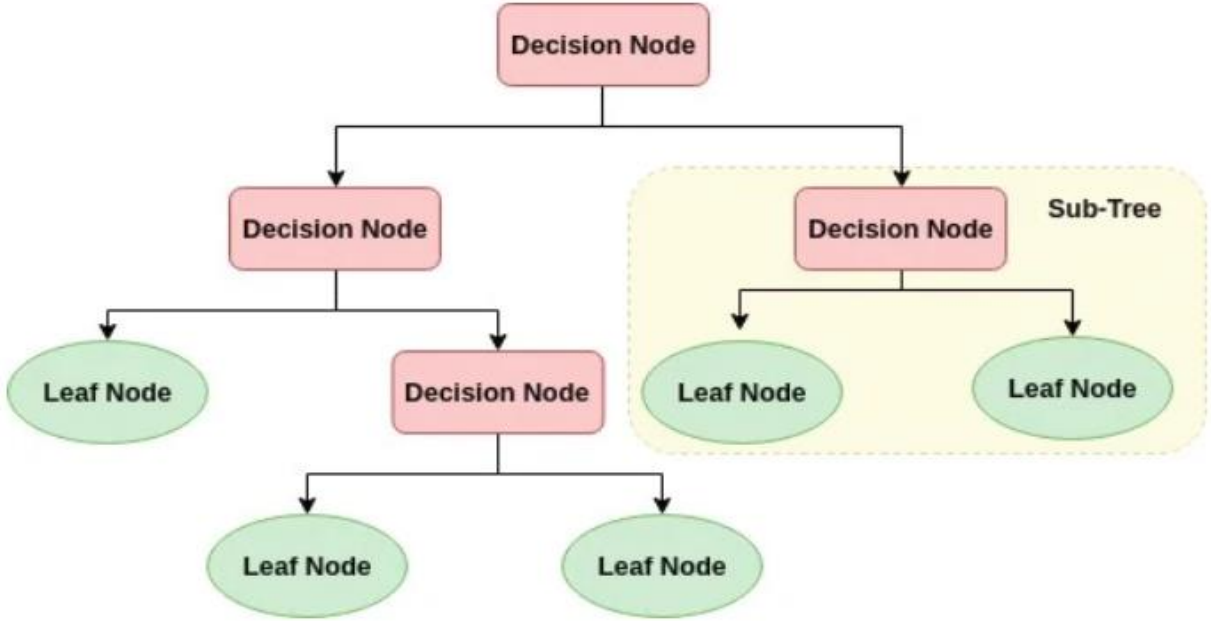
KNN Algorithm.

- The k-nearest neighbours (**KNN**) **algorithm** is a simple, supervised machine learning **algorithm** that can be used to solve both classification and regression problems. It's easy to implement and understand, but has a major drawback of becoming significantly slows as the size of that data in use grows.



Decision Tree Classifier:

A **decision tree** is a diagram or chart that people use to determine a course of action or show a statistical probability. It forms the outline of the namesake woody plant, usually upright but sometimes lying on its side. Each branch of the **decision tree** represents a possible **decision**, outcome, or reaction



XGBoost Classifier:

XGBoost is a popular and efficient open-source implementation of the gradient boosted trees algorithm. When using gradient boosting for regression, the weak learners are regression trees, and each regression tree maps an input data point to one of its leafs that contains a continuous score.

XGBoost is a scalable and accurate implementation of gradient boosting machines and it has proven to push the limits of computing power for boosted trees algorithms as it was built and developed for the sole purpose of model performance and computational speed.

Each tree learns from its predecessors and updates the residual errors. Hence, the tree that grows next in the sequence will learn from an updated version of the residuals. The base learners in boosting are weak learners in which the bias is high, and the predictive power is just a tad better than random guessing.

Logistic Regression:

Logistic Regression is a supervised machine learning algorithm which is used to solve classification problems particularly binary classification.

Logistic regression is a statistical model that in its basic form uses a logistic function to model a binary dependent variable, although many more complex extensions exist. In regression analysis, logistic regression (or logit regression) is estimating the parameters of a logistic model (a form of binary regression).

Like all regression analyses, the logistic regression is a predictive analysis. Logistic regression is used to describe data and to explain the relationship between one dependent binary variable and one or more nominal, ordinal, interval or ratio-level independent variables.

HARDWARE AND SOFTWARE REQUIREMENTS

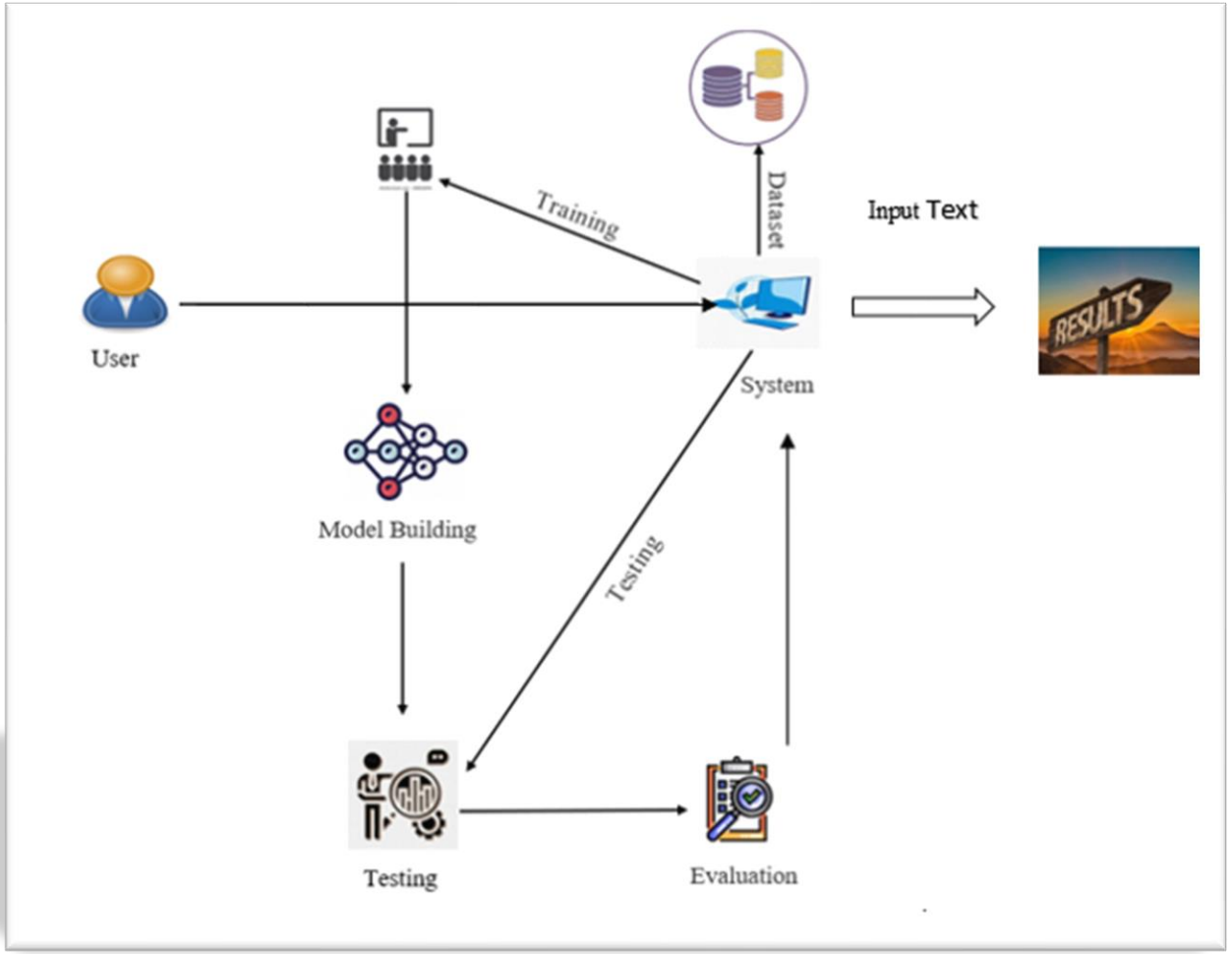
H/W Configuration

Processor	- I3/Intel Processor
Hard Disk	- 160GB
RAM	- 8Gb

S/W Configuration:

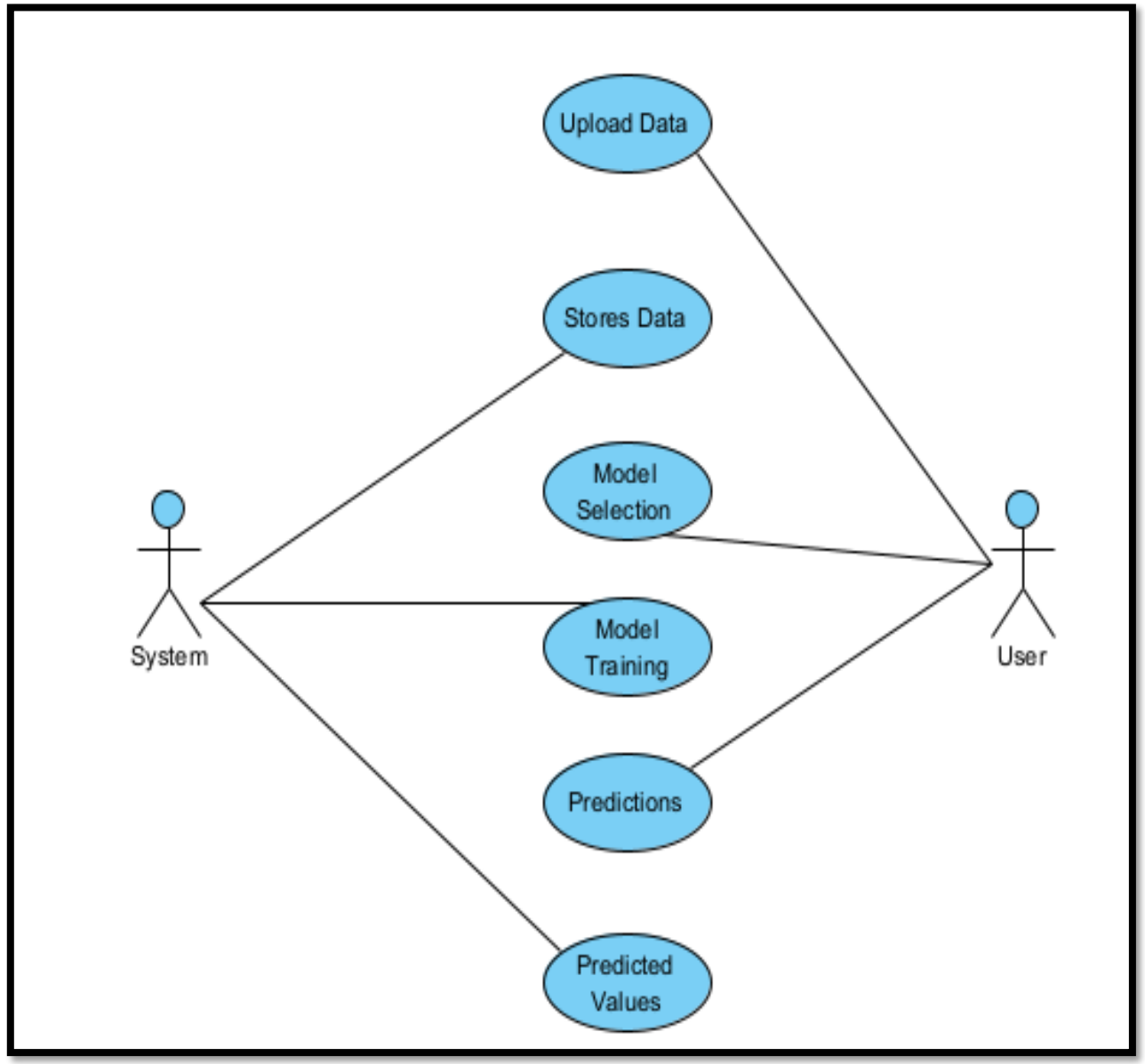
Operating System	: Windows 7/8/10
Server side Script	: HTML, CSS & JS
IDE	: Pycharm / Google colab.
Libraries Used	:Numpy, OS, Scikit-Learn, Pandas.
Technology	: Python 3.6+

ARCHIECTURE



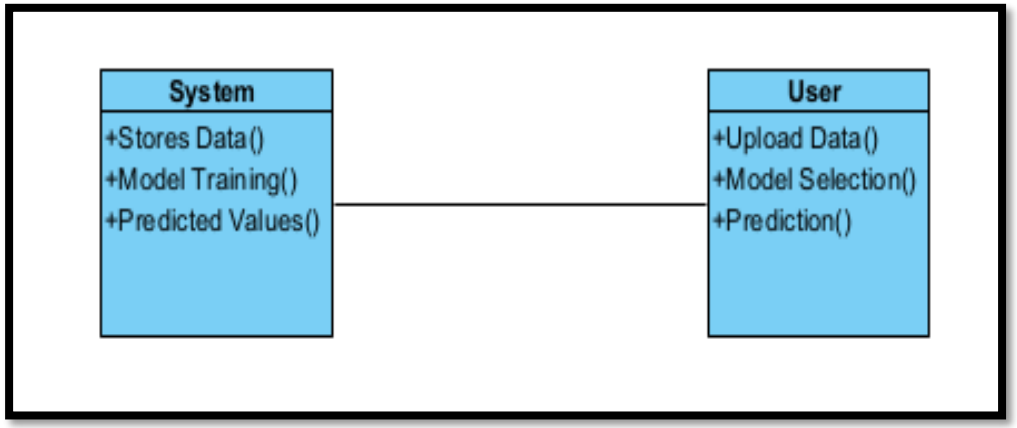
USE CASE DIAGRAM

- 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.


Upload Data
Stores Data
Model Selection
Model Training
Predictions
Predicted Values
System
User

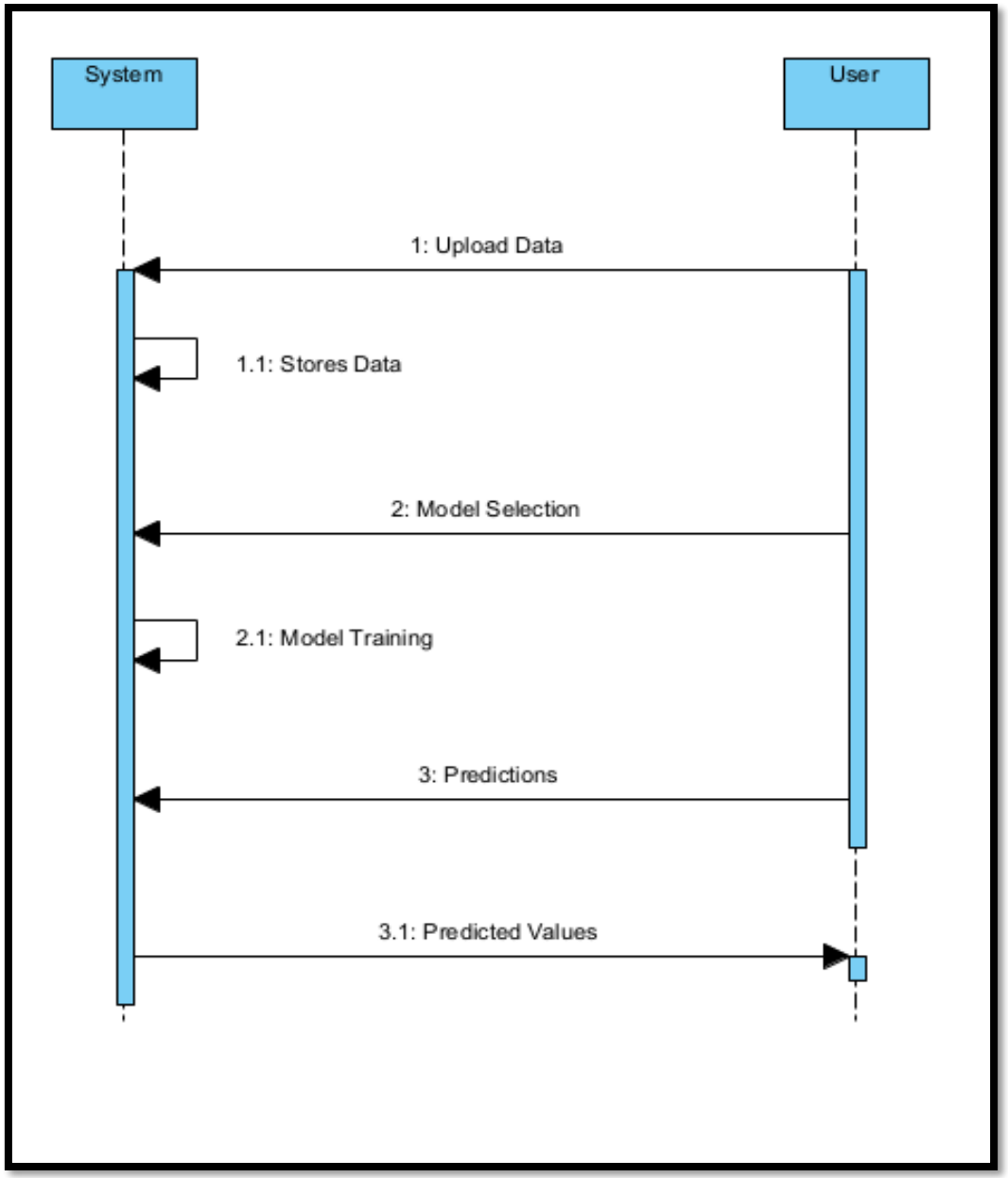
CLASS DIAGRAM

- 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.



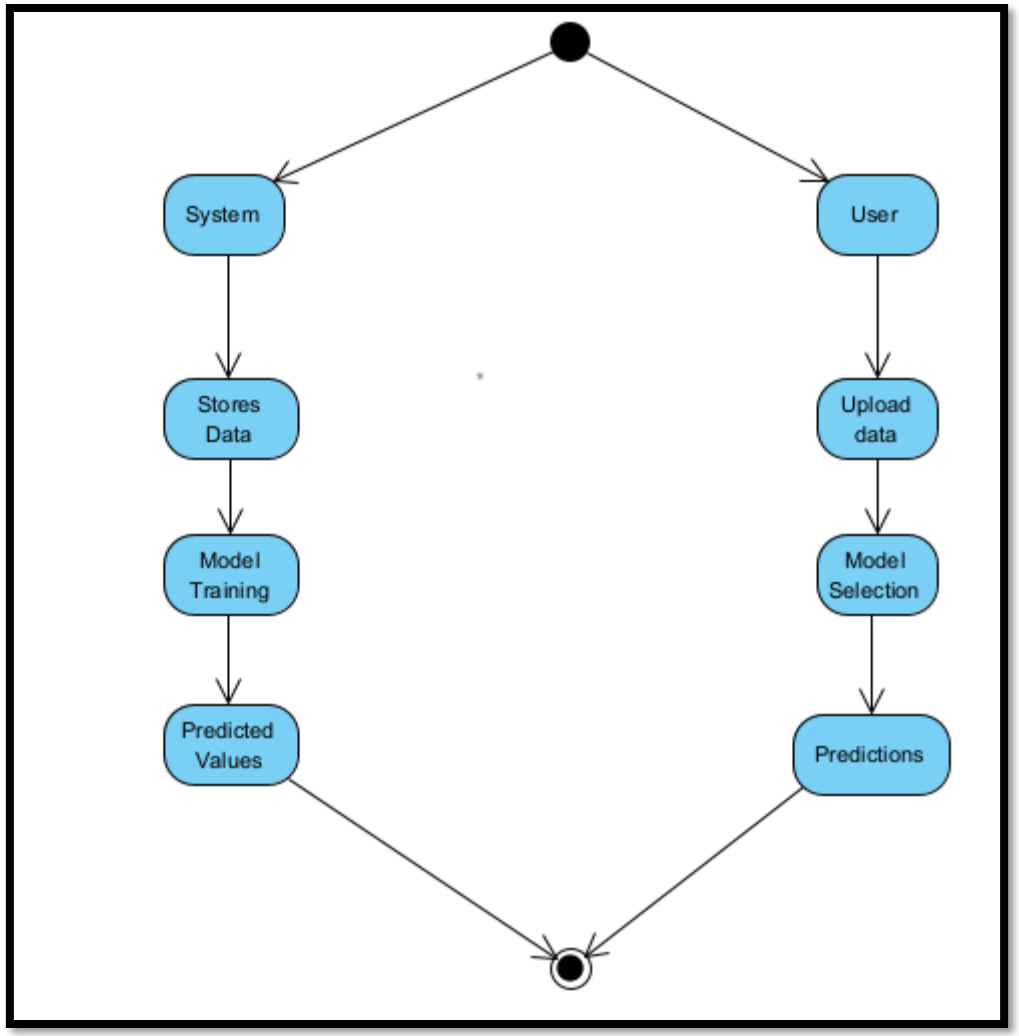
SEQUENCE DIAGRAM

- 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.


System
User
1: Upload Data
1.1: Stores Data
2: Model Selection
2.1: Model Training
3: Predictions
3.1: Predicted Values

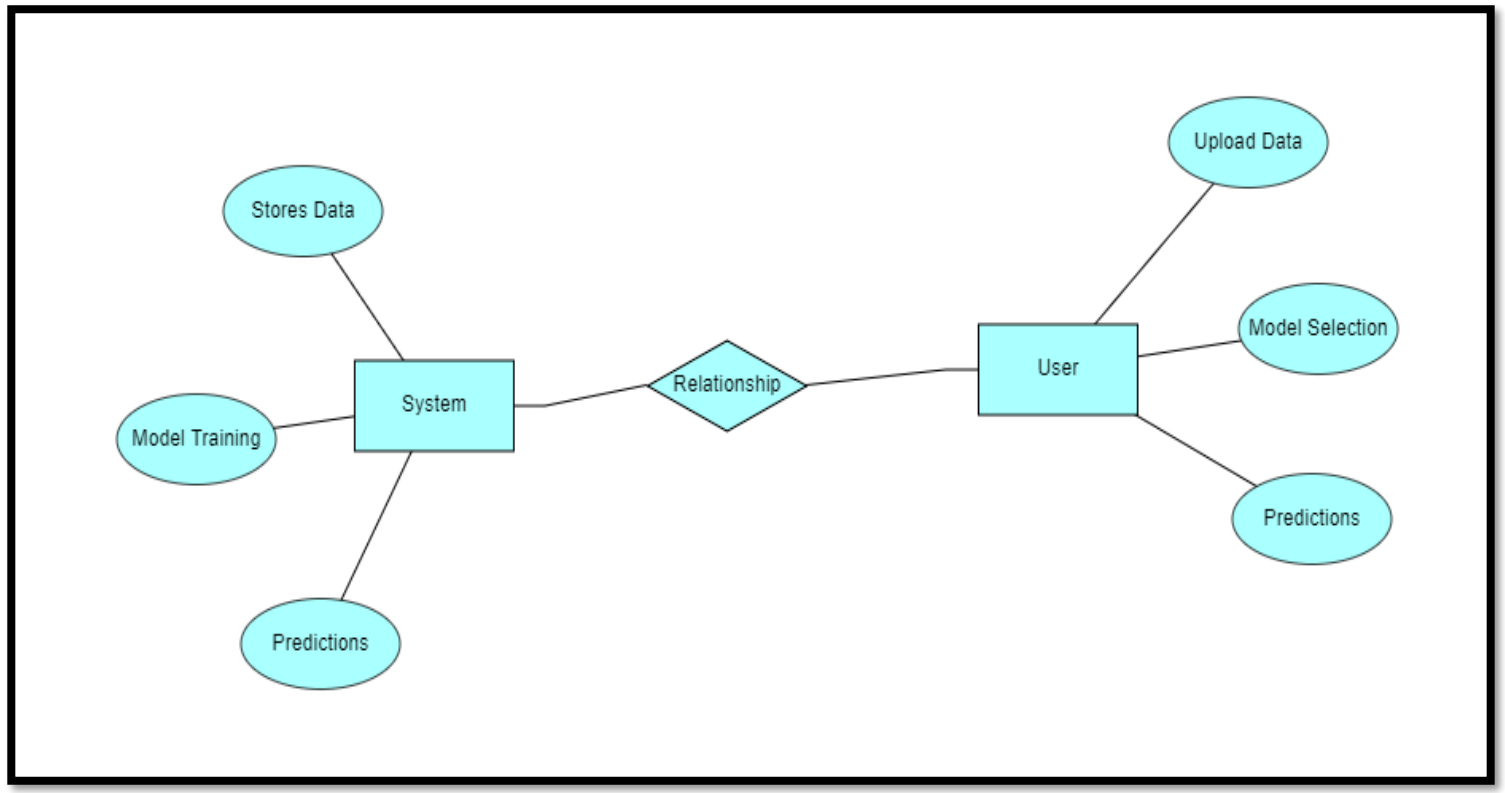
ACTIVITY DIAGRAM

- Activity diagrams are graphical representations of workflows of stepwise activities and actions with support for choice, iteration and concurrency.
- In the Unified Modeling 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.


System
User
Stores Data
Upload data
Model Training
Model Selection
Predicted Values
Predictions

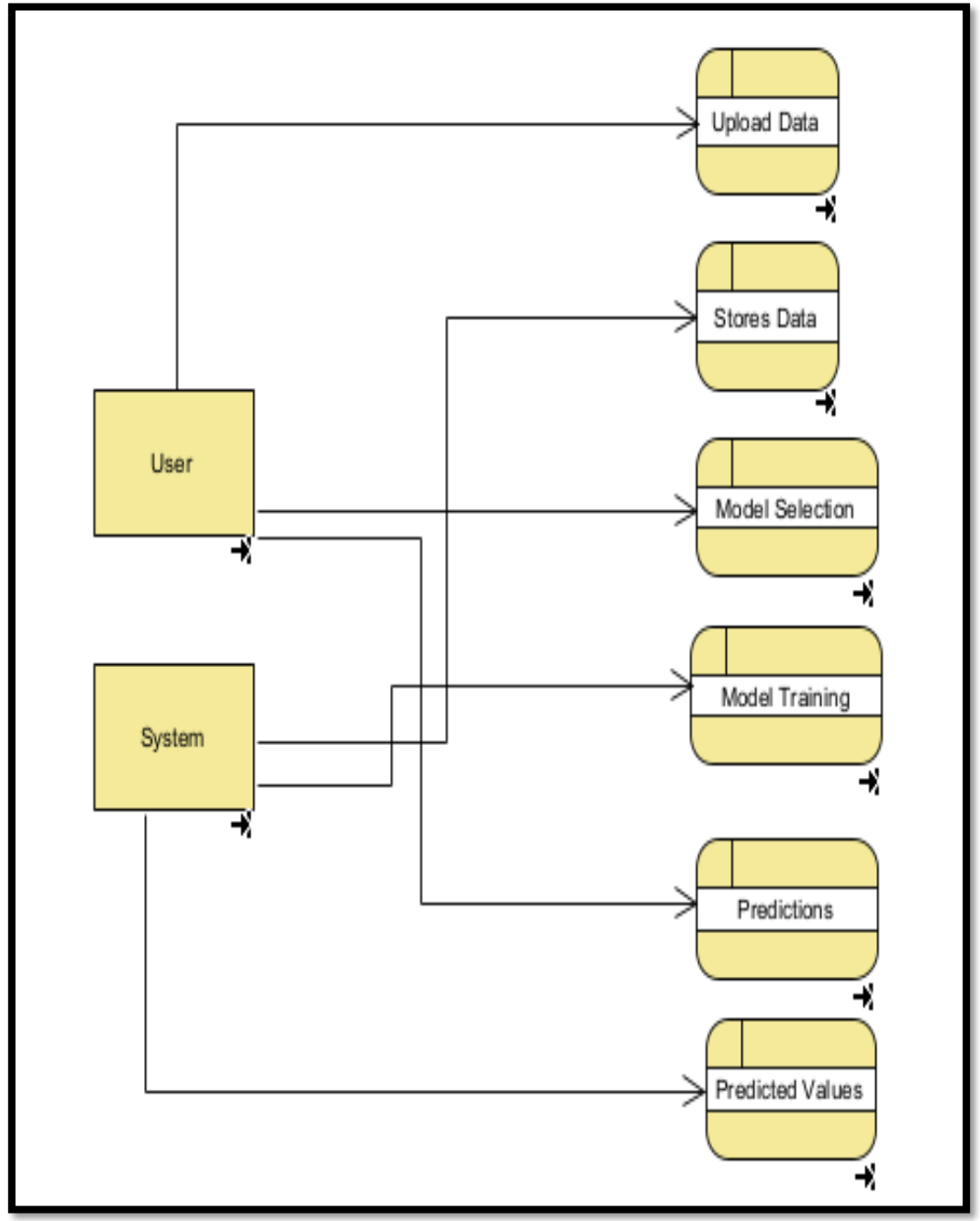
ER DIAGRAM

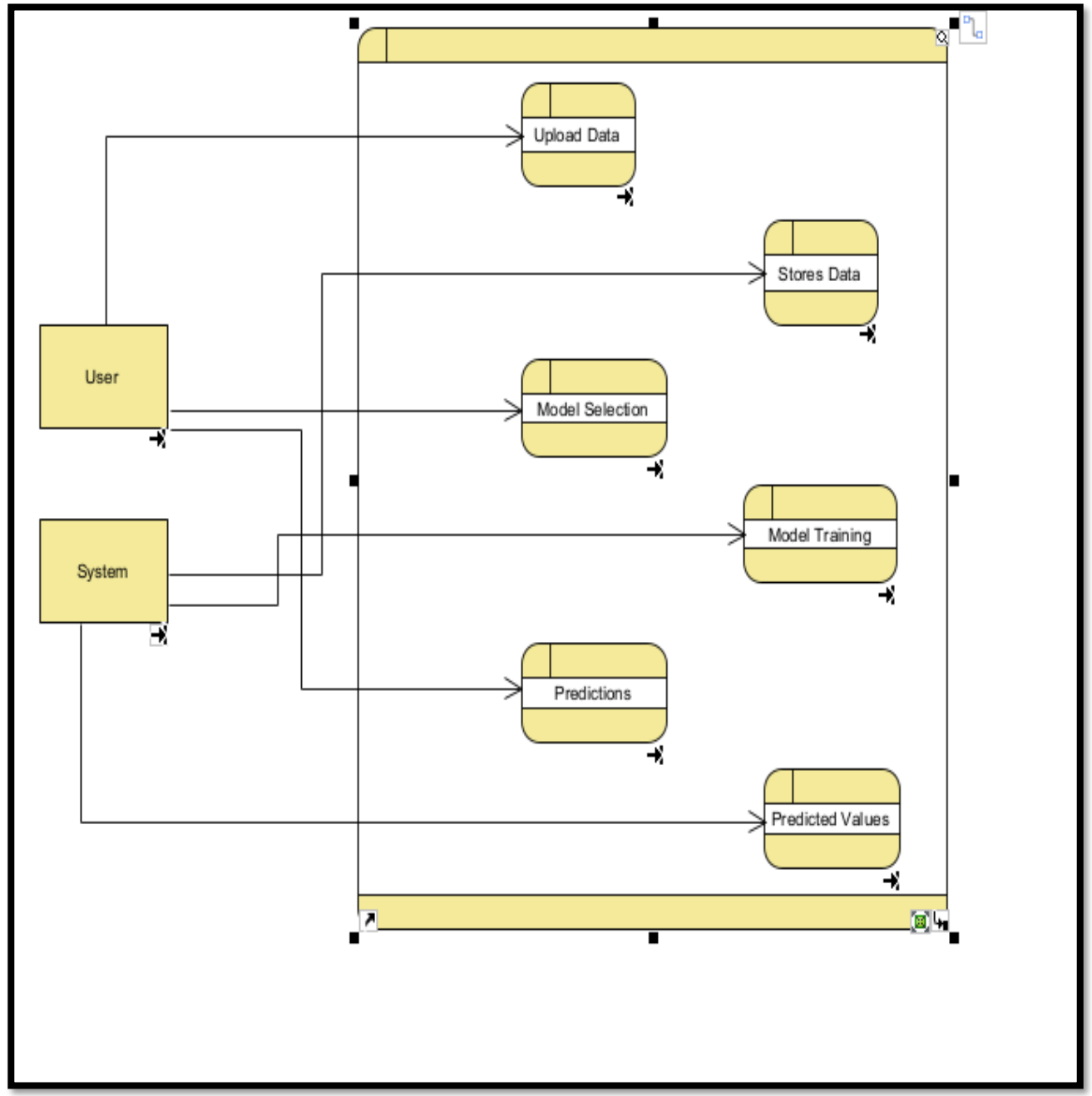
- An Entity–relationship model (ER model) describes the structure of a database with the help of a diagram, which is known as Entity Relationship Diagram (ER Diagram).
- An ER diagram shows the relationship among entity sets. An entity set is a group of similar entities and these entities can have attributes.
- 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.


Stores Data
Model Training
System
Predictions
Relationship
User
Upload Data
Model Selection
Predictions

DF DIAGRAM

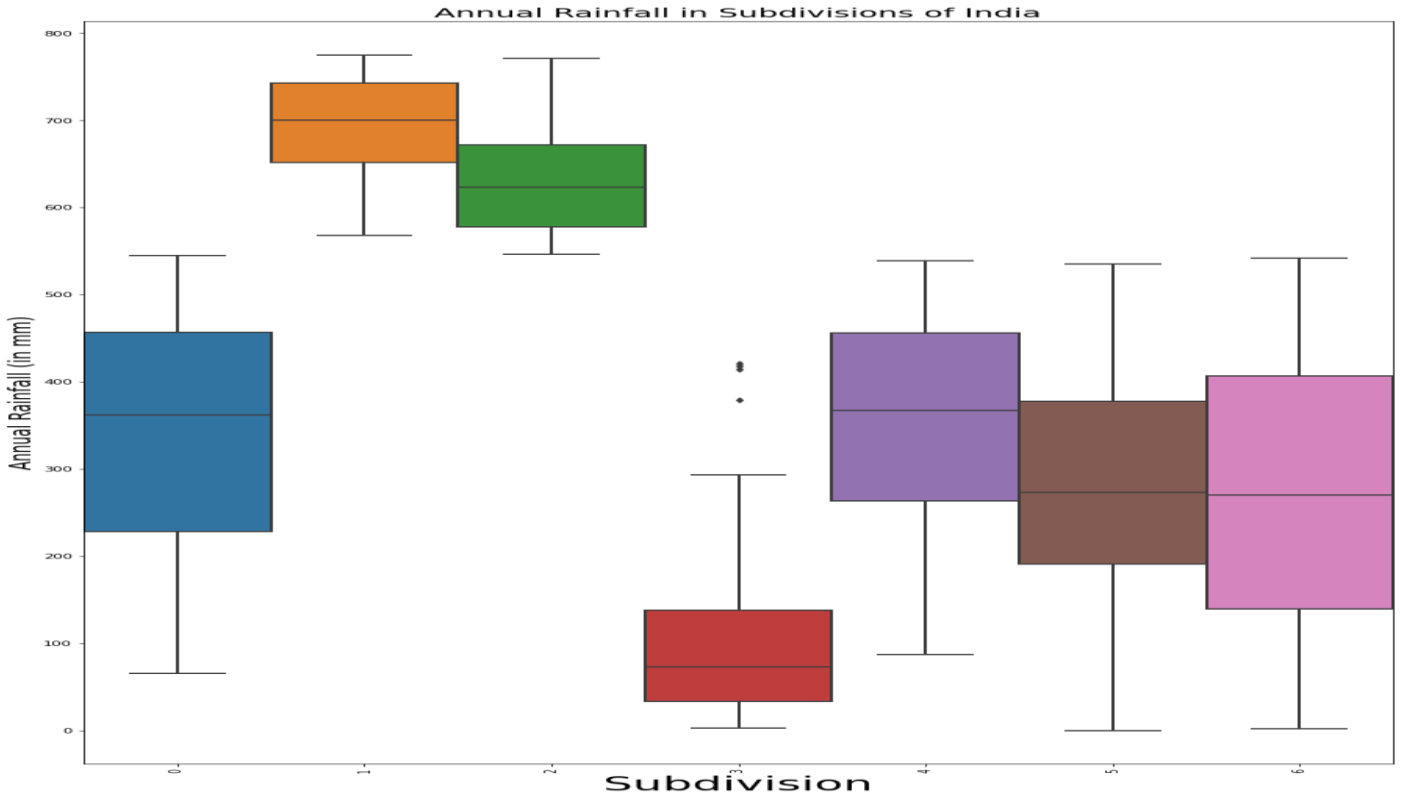
- A Data Flow Diagram (DFD) is a traditional way to visualize the information flows within a system.
- A neat and clear DFD can depict a good amount of the system requirements graphically. It can be manual, automated, or a combination of both.
- It shows how information enters and leaves the system, 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.


Upload Data
Stores Data
User
Model Selection
Model Training
System
Predictions
Predicted Values

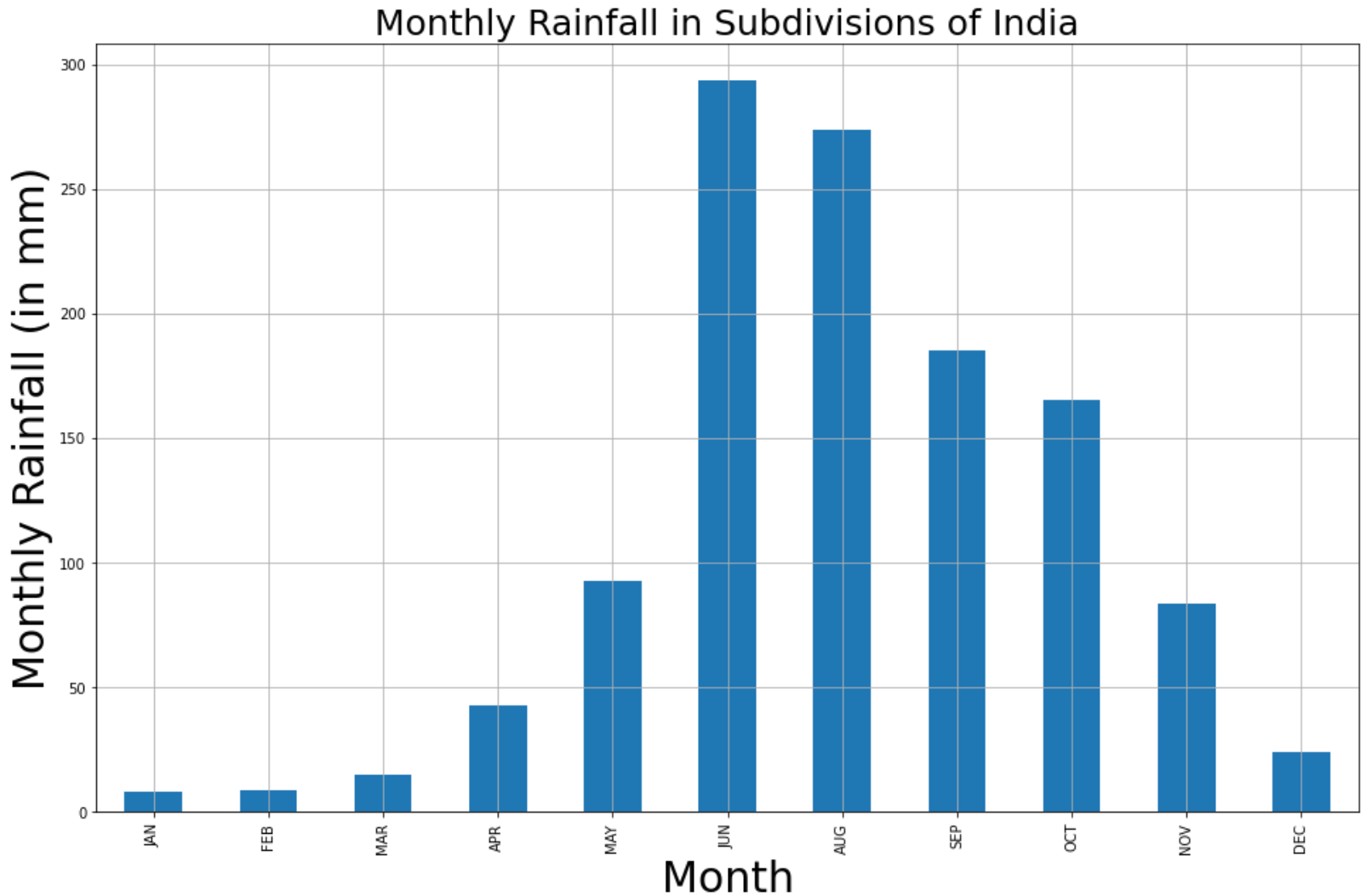

Upload Data
Stores Data
User
Model Selection
Model Training
System
Predictions
Predicted Values

Plots

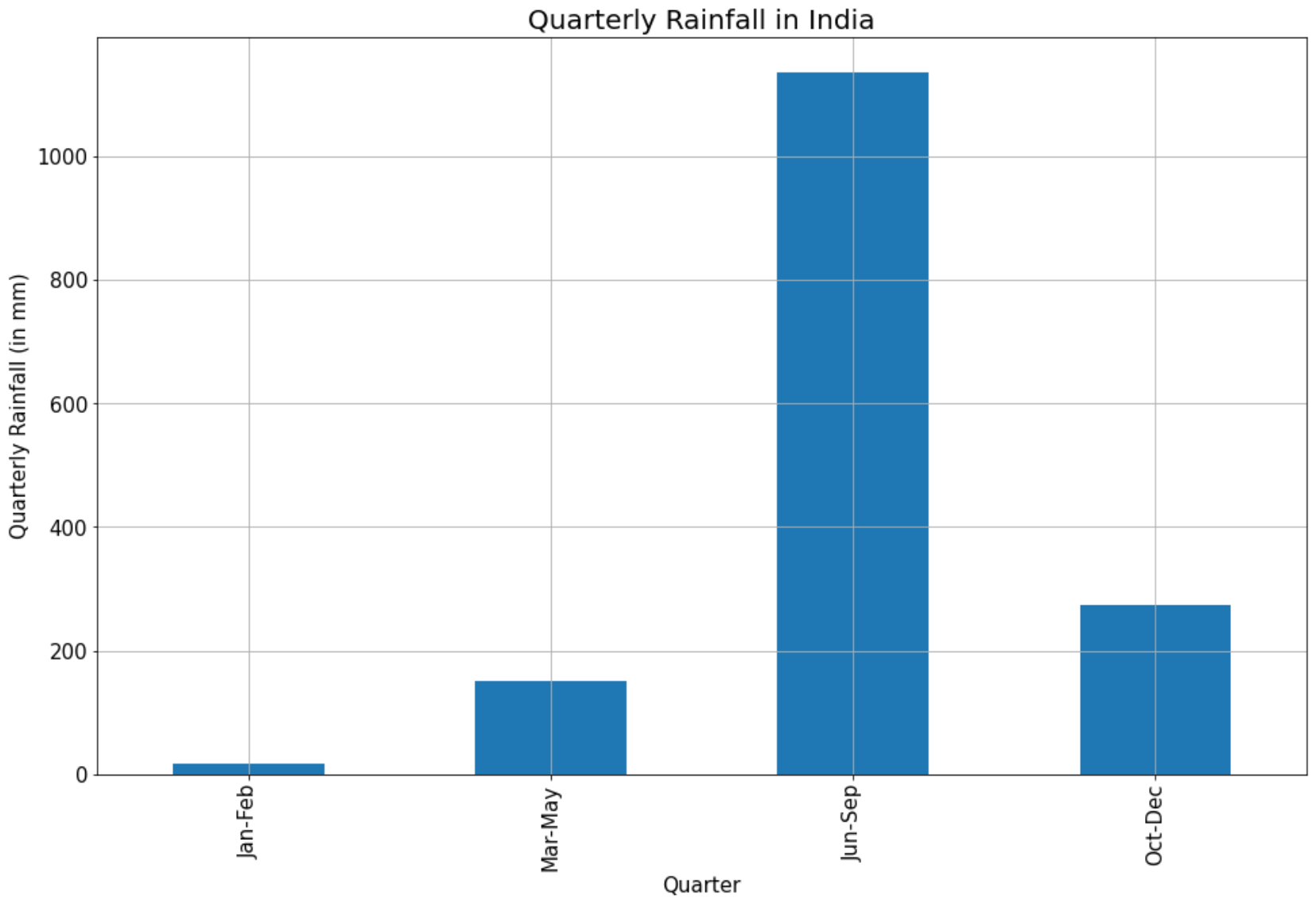
Annual rainfall in subdivisions of India



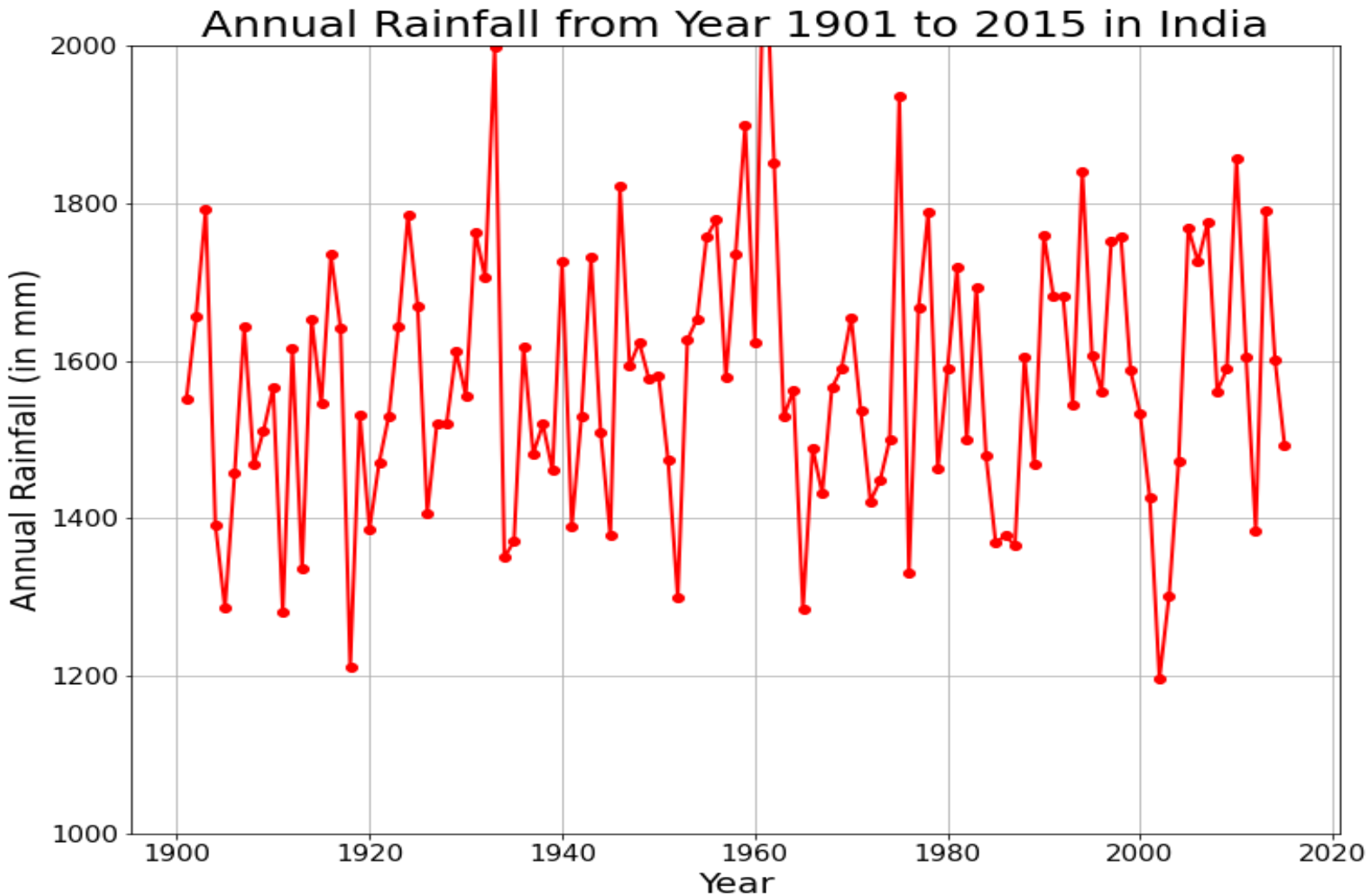
Average monthly rainfall in India



Average Quartely rainfall in India



Visualizing annual rainfall over the years(1901-2015) in India



Accuracy

Without using Hyper Parameter Tuning

	Name	Score
0	svm	0.987578
1	rf	0.993789
2	knn	0.975155
3	dt	0.993789
4	lr	0.993789
5	xg	0.993789

Accuracy

By using Hyper Parameter Tuning

SVM

```
{'C': 10, 'kernel': 'rbf'}
Accuracy:0.9913212746675545
```

RF

```
{'criterion': 'gini', 'max_depth': 15, 'n_estimators': 20}
Accuracy:1.0
```

KNN

```
{'n_neighbors': 10}
Accuracy:0.9913166509460134
```

DT

```
Fitting 4 folds for each of 50 candidates, totalling 200 fits
{'criterion': 'gini', 'max_depth': 2, 'min_samples_leaf': 5}
Accuracy:1.0
```

LR

```
Accuracy - : 0.996
```

XG Boost

```
Fitting 4 folds for each of 405 candidates, totalling 1620 fits
{'colsample_bytree': 0.6, 'gamma': 0.5, 'max_depth': 3, 'min_child_weight': 1, 'subsample': 1.0}
Accuracy:1.0
```

Accuracy

Some Ensemble Techniques

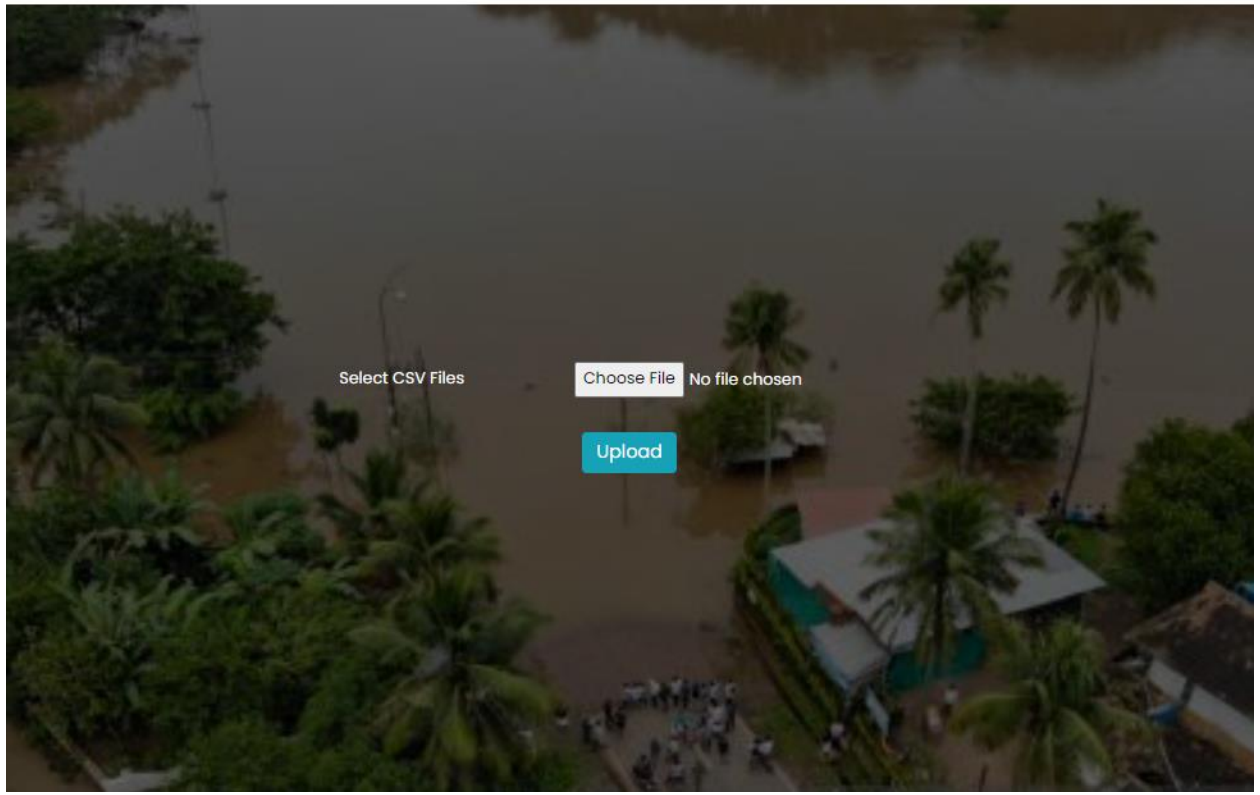
Max Voting	1.0
Bagging meta-estimator	1.0
ADA Boost	1.0
Gradient Boost	1.0

Result

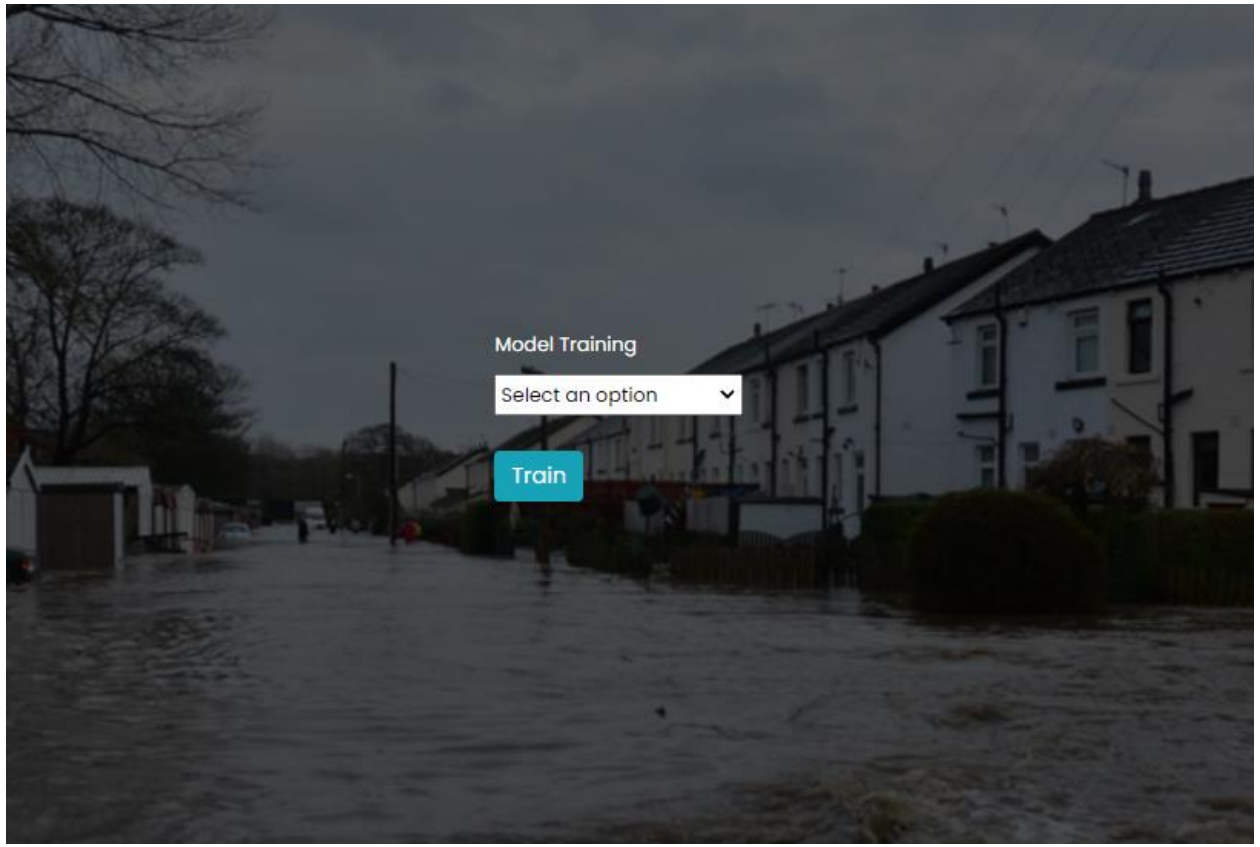
Home:

In our project, we are detecting whether the customer will be using the same network or not.

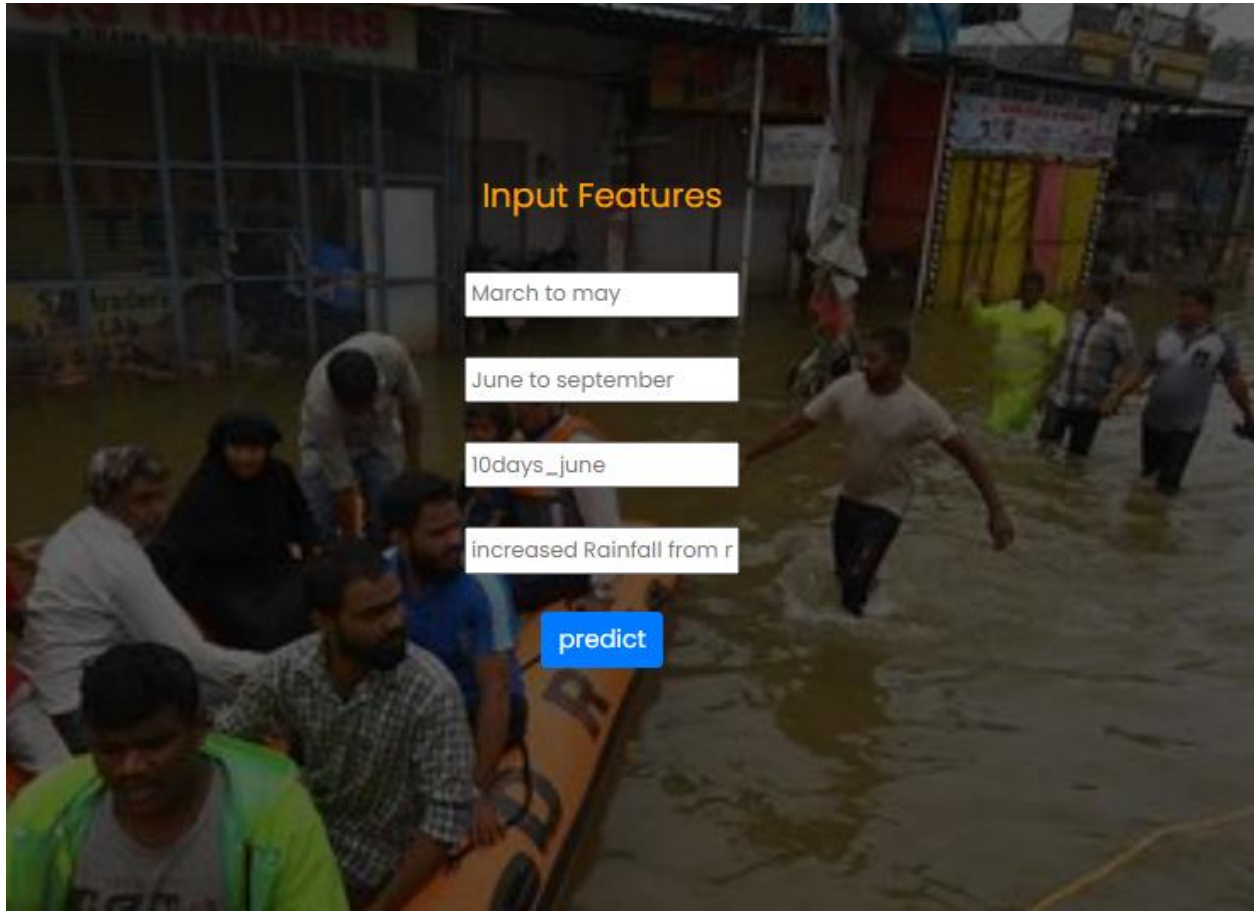
Upload Data



Model Training



Prediction


Input Features
March to may
June to september
10days_june
increased Rainfall from r
predict

Conclusion

- We have successfully developed a system to predict whether the floods will occur or not in this application.
- This is created in a user-friendly environment with Python programming and Flask.
- The system is likely to gather data from the user in order to predict whether there is a chance of flood occurring or not.

Future Scope

- We intend to investigate prediction approach with the revised data set and employ the most accurate and relevant machine learning algorithms for detection.

Thank you