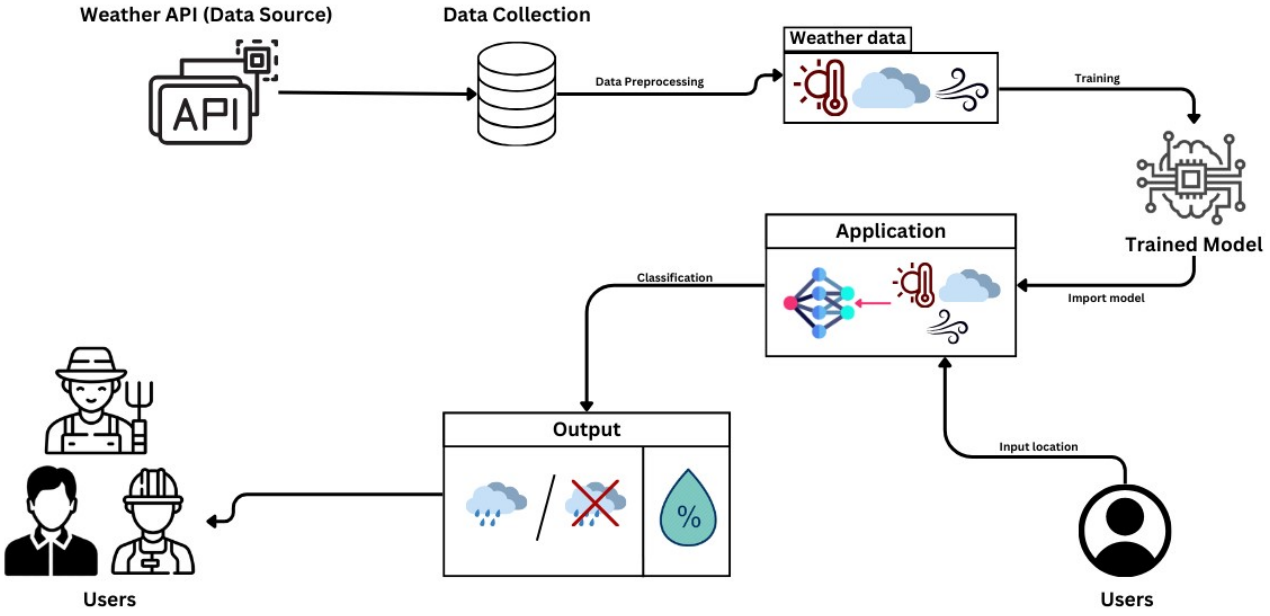
Project Design Phase-II
Data Flow Diagram & User Stories

Date:	26-10-2023
Team ID:	PNT2023TMID-592801
Project Name:	Machine Learning Approach for Predicting Rainfall
Maximum Marks:	4 Marks

Data Flow Diagrams:

A Data Flow Diagram (DFD) is a traditional visual representation of the information flows within a system. A neat and clear DFD can depict the right amount of the system requirement graphically. It shows how data enters and leaves the system, what changes the information, and where data is stored.



User Type	Functional Requirement (Epic)	User Story Number	User Story/Task	Acceptance Criteria	Priority	Release
	Data Sourcing	USN-1	As a **data scientist**, I want to **source a reliable dataset** for rainfall prediction so that the machine learning model can be trained effectively.	Obtaining a dataset with relevant features for rainfall prediction, either from Kaggle, other repositories, cloud sourcing, or APIs	High	First Sprint
	Data Preprocessing	USN-2	As a **data scientist**, I need to **preprocess the sourced data** to ensure it is clean and suitable for model development.	Having a dataset free of null values, outliers handled, and data normalized if required	High	First Sprint
	Model Development	USN-3	As a **data scientist**, I want to **develop a machine learning model** using classification and regression techniques to predict rainfall.	The acceptance criteria would be having a working ML model that can take in the preprocessed data and output a prediction	High	Second Sprint
	Model Training	USN-4	As a **data scientist**, I need to **train the developed ML model** to ensure it learns from data and performs well while testing	The model should achieve a certain threshold of performance on a validation dataset during training and should not overfit or underfit the training data.	High	Second Sprint
	Model Testing	USN-5	As a **data scientist**, I need to **test the developed ML model** to ensure its accuracy in predicting rainfall.	Achieving an acceptable level of accuracy, precision,	High	Second Sprint

				recall, or any other decided metric on the test data		
	Model Deployment	USN-6	As a **developer**, I want to **deploy the trained ML model** into a web application so that users can use it to predict rainfall.	The acceptance criteria would be having a working web application where users can input data and receive rainfall predictions	High	Third Sprint
	Web App Deployment	USN-7	As a **developer**, I need to develop a simple web application using Flask that integrates with the deployed ML model so that users (general public, farmers, other stakeholders) can predict rainfall	The acceptance criteria would be having a user-friendly web application where users can easily input data and receive rainfall predictions.	High	Third Sprint