Project Planning Phase

Project Planning Template (Product Backlog, Sprint Planning, Stories, Story points)

Date	15 February 2026
Team ID	LTVIP2026TMIDS64419
Project Name	Rising Waters: A Machine Learning Approach to Flood Prediction
Maximum Marks	5 Marks

Project Planning

The development of the Flood Prediction System was carried out in a structured and systematic manner to ensure accuracy, reliability, and scalability. The project planning phase included problem analysis, data preparation, model development, evaluation, and deployment.

---

1. Requirement Analysis

The first stage involved understanding the problem of flood prediction and identifying the key environmental factors that influence flood occurrence. Based on domain research, parameters such as rainfall, temperature, humidity, and related climatic features were selected as input variables. The goal was defined as building a classification model that predicts whether flood risk is High or Low.

---

2. Data Collection

Historical environmental and flood-related datasets were collected from reliable sources. The dataset consisted of multiple records containing climatic attributes along with corresponding flood labels. The data was stored in structured CSV format for further processing.

---

3. Data Cleaning

Data cleaning was performed to ensure quality and consistency. This step included checking for missing values, removing duplicate entries, and correcting inconsistent data formats. Missing values were handled using appropriate techniques such as removal of

incomplete rows or statistical replacement methods. Outliers were identified and analyzed to prevent misleading model training. This stage ensured that the dataset was accurate and reliable before proceeding to model development.

---

4. Data Preprocessing

After cleaning, preprocessing was performed to prepare the data for machine learning algorithms. Feature selection was conducted to identify the most relevant environmental parameters affecting flood prediction. If categorical features were present, they were encoded into numerical format. Feature scaling and normalization techniques were applied to ensure uniform value ranges, improving model performance. The dataset was then divided into training and testing sets to evaluate model generalization.

---

5. Exploratory Data Analysis (EDA)

Exploratory analysis was conducted to understand relationships between variables. Statistical summaries, correlation analysis, and visualization techniques such as histograms and boxplots were used to identify patterns and trends in rainfall and other climatic parameters. This step helped in selecting suitable algorithms and understanding feature importance.

---

6. Model Development

Multiple classification algorithms including Decision Tree, Random Forest, SVM, Extra Trees, and XGBoost were implemented. Each model was trained using the training dataset to learn patterns associated with flood occurrence.

---

7. Model Evaluation and Selection

The performance of all models was evaluated using metrics such as accuracy, precision, recall, F1-score, and confusion matrix. After comparison, XGBoost was selected as the final model due to its superior predictive performance and ability to handle nonlinear relationships effectively.

---

8. Model Deployment

The trained XGBoost model was saved using serialization techniques and integrated into a Flask-based web application. The system allows users to input environmental parameters and receive instant flood risk predictions through a user-friendly interface.

---

9. Testing and Validation

The application was tested with various input scenarios to ensure consistent and accurate performance. Edge cases and boundary conditions were also verified to confirm system stability.

This structured project planning ensured that the flood prediction system was developed systematically, with strong emphasis on data quality, model accuracy, and practical usability.