Rising Waters - A Machine Learning Approach to Flood Prediction

Milestone 1: Project Initialization and Planning Phase

The Rising Waters project addresses the growing challenge of flood preparedness by developing a machine learning-based prediction system. The project aims to leverage historical environmental data to provide early warnings for potential flood events, helping communities take preventive measures. The project team planned the workflow, assigned roles, identified resource needs, and defined clear objectives aligned with disaster management efforts.

Activity 1: Define Problem Statement
Floods cause significant loss of life, property damage, and disruption to communities, especially in regions with limited early warning systems. The lack of reliable flood prediction increases vulnerability and prevents effective disaster preparedness. This project addresses the need for a machine learning-based flood prediction system using environmental and seasonal data to enable early alerts and minimize risks.

Ref. template: Click Here
Rising Waters Problem Statement Report: Click Here

Activity 2: Project Proposal (Proposed Solution)
Rising Waters proposes a machine learning approach to predict flood occurrence based on historical weather and rainfall data. By analyzing features such as temperature, humidity, cloud cover, and seasonal rainfall, the project aims to build an accurate and reliable classification model for flood prediction. This system will help authorities, residents, and disaster response teams take proactive measures to reduce flood-related damage.

Ref. template: Click Here
Rising Waters Project Proposal Report: Click Here

Activity 3: Initial Project Planning
The project begins with data collection, followed by preprocessing, model development, evaluation, and deployment. The team will use Decision Tree, Random Forest, KNN, and XGBoost models for flood prediction. Responsibilities are assigned for data handling, model building, and system deployment. The project plan ensures a structured approach with clear milestones and deliverables.

Ref. template: Click Here
Rising Waters Project Planning Report: Click Here

Milestone 2: Data Collection and Preprocessing Phase

Relevant flood-related data, including rainfall, temperature, humidity, and cloud cover, was collected from reliable meteorological sources. The dataset was verified for missing values and inconsistencies, ensuring high-quality inputs for model development. Preprocessing steps included scaling numerical features and preparing the target variable for binary classification. This phase ensured the dataset was clean, structured, and ready for machine learning applications.

Activity 1: Data Collection Plan, Raw Data Sources Identified, Data Quality Report
The dataset for this project contains historical weather data, including temperature, humidity, cloud cover, seasonal rainfall, and flood occurrence labels. The data was sourced from reliable meteorological records and stored in an Excel format for analysis. This dataset provides the foundation for training accurate flood prediction models.

Ref. template: Click Here
Rising Waters Data Collection Report: Click Here

Activity 2: Data Quality Report
A thorough data quality check was conducted to identify missing values, inconsistencies, and scaling issues. The dataset was clean with no missing values. Standardization techniques were applied to scale numerical features, ensuring uniformity and improving model performance.

Ref. template: Click Here
Rising Waters Data Quality Report: Click Here

Activity 3: Data Exploration and Preprocessing
Exploratory analysis highlighted seasonal rainfall patterns and humidity as major contributors to flood events. Preprocessing steps included feature scaling using standardization, encoding the target variable, and preparing the dataset for machine learning models. These steps ensured clean, consistent, and meaningful input for the models.

Ref. template: Click Here
Rising Waters Data Exploration and Preprocessing Report: Click Here

Milestone 3: Model Development Phase

Exploratory analysis highlighted key features influencing flood risk, such as seasonal rainfall patterns and humidity levels. Four machine learning models — Decision Tree, Random Forest, K-Nearest Neighbors (KNN), and XGBoost — were selected based on their ability to handle classification tasks. These models were trained, evaluated, and compared to determine the most effective approach for predicting flood occurrence.

Activity 1: Feature Selection Report
Based on domain knowledge and data exploration, key features like temperature, humidity, cloud cover, and seasonal rainfall were selected for model development. These features showed strong relationships with flood occurrence and were retained to maximize predictive accuracy.

Ref. template: [Click Here]
Rising Waters Feature Selection Report: [Click Here]

Activity 2: Model Selection Report
Decision Tree, Random Forest, K-Nearest Neighbors (KNN), and XGBoost were chosen for this project. These models were selected for their proven effectiveness in classification tasks, with a balance between interpretability, robustness, and handling complex patterns in the data.

Ref. template: [Click Here]
Rising Waters Model Selection Report: [Click Here]

Activity 3: Initial Model Training Code, Model Validation and Evaluation Report
All four models were trained on the preprocessed dataset. Their performance was evaluated using accuracy, classification reports, and confusion matrices on both training and test data. This ensured that the models were assessed for both predictive accuracy and generalization capability.

Ref. template: [Click Here]
Rising Waters Model Development Phase Template: [Click Here]

Milestone 4: Model Optimization and Tuning Phase

Hyperparameter tuning was performed for all four models to maximize predictive performance. While Decision Tree, Random Forest, and XGBoost showed very high accuracy, they risked overfitting due to perfect results on both training and test data. KNN, however, demonstrated a more realistic and balanced performance with 91.3% test accuracy. Its simplicity and ability to handle non-linear patterns made it the preferred final model for deployment.

Activity 1: Hyperparameter Tuning Documentation
Hyperparameter tuning was performed to enhance model performance. Parameters such as tree depth, number of neighbors, and learning rates were optimized for each model. The tuning process helped improve accuracy while minimizing overfitting.

Activity 2: Performance Metrics Comparison Report
After tuning, all models were compared based on their accuracy and overall performance. While some models showed perfect accuracy, realistic generalization was a key factor considered during comparison to avoid overfitting.

Activity 3: Final Model Selection Justification
The K-Nearest Neighbors (KNN) model was selected as the final model due to its balanced performance and realistic accuracy of 91.3%. Unlike other models that risked overfitting with perfect accuracy, KNN demonstrated good generalization and was well-suited for environmental data patterns.

Ref. template: Click Here
Rising Waters Model Optimization and Tuning Phase Report:

Milestone 5: Project Files Submission and Documentation

The complete project, including cleaned datasets, model training code, evaluation reports, and project documentation, was organized and submitted. The system is now prepared for deployment, with the capability to provide real-time flood risk predictions based on environmental data.
For project file submission in Github, Kindly click the link and refer to the flow. **Click Here**
For the documentation, Kindly refer to the link. **Click Here**

Milestone 6: Project Demonstration

A project demonstration was recorded, showcasing the end-to-end process — from data handling to model development and flood prediction output. The demo highlights the system's potential to contribute to disaster preparedness by providing timely, accurate flood risk assessments.