

GOVERNMENT OF INDIA
MINISTRY OF SKILL DEVELOPMENT & ENTREPRENEURSHIP
DIRECTORATE GENERAL OF TRAINING
NATIONAL SKILL TRAINING INSTITUTE (W)

8, Master Zahurul Hasan Rd, Old Katra, Prayagraj, Uttar Pradesh 211002

CERTIFICATE

This is to certify that following trainees have completed their project
titled

“Smart Flood Alert System”

“Artificial Intelligence Programming Assistant”

Session: 2024-25

Submitted By

1. Sushmita Pal
2. Soniya Gupta
3. Smriti Gupta
4. Sonal Tiwari
5. Shubhangini Devi
6. Arpita

Faculty

Trade In charge

Training In charge

Principal

(Smart Flood Alert System) Project Report

-
1. Sushmita Pal
 2. Soniya Gupta
 3. Smriti Gupta
 4. Sonal Tiwari
 5. Shubhangini Devi
 6. Arpita
-

A report submitted in part fulfilment of the certificate of
Artificial Intelligence Programming Assistance
(2024-2025)

Guidance: Mr. Shailesh Yadav Sir



NSTI(W) ALLAHABAD

Date – 11 July 2025

Abstract

Floods are among the most devastating natural disasters, causing widespread destruction to life, property, and the environment. With climate change increasing the frequency and intensity of rainfall, there is an urgent need for more intelligent, proactive, and real-time flood management solutions. The **Smart Flood Alert System** leverages the power of **IoT (Internet of Things)** sensors and **machine learning algorithms** to monitor environmental conditions such as rainfall, water level, and river flow rate. These data points are continuously collected and analyzed in real time to detect patterns that may indicate an impending flood. The system is designed to issue early warnings through SMS or mobile app notifications to residents, disaster management authorities, and relevant stakeholders, allowing for timely evacuations and resource mobilization. By combining modern data analytics with real-time sensor inputs, this system aims to minimize the impact of floods, safeguard human lives, and support sustainable urban planning and disaster preparedness.

Acknowledgement

We would like to express our sincere gratitude to several individuals and organizations for supporting us throughout our diploma journey. First, we wish to express our heartfelt thanks to our trainer, **Mr. Shailesh Yadav**, for his unwavering encouragement, insightful feedback, and constant motivation. His technical expertise, valuable guidance, and deep knowledge in computer science have played a vital role in the successful completion of this project.

We also extend our gratitude to the **National Skill Training Institute, Allahabad**, for providing us with the opportunity to pursue this diploma program. We are especially thankful to the **Ministry of Skill Development & Entrepreneurship** and **IBM** for launching such a platform that has empowered us with the skills and tools needed to work on real-world problems.

We are sincerely thankful to **Ms. Reetu Shukla**, faculty at **NSTI Allahabad** for his support and technical inputs throughout the coursework and project development.

Lastly, we would like to express our appreciation to **Regional Director Sir Jai Shiv Sharma** and **Assistant Director Mr. Ajay Pal Singh** for their continuous encouragement and administrative assistance during our training.

This project would not have been possible without the combined efforts and support of all the above mentors and institutions. We remain deeply grateful for their contribution.

Table of content

Abstract	3
Acknowledgement	4
Table of content	5
Table of figures	6
Figure 1 System Architecture Diagram.....	6
Figure 2 Real-time Data Flow Diagram	6
Figure 3 Correlation Heatmap of Features.....	6
Figure 4 Histogram of Rainfall Distribution	6
Figure 5 Confusion Matrix (Random Forest)	6
Figure 6 ROC Curve of Random Forest Classifier	6
Problem Statement.....	7
Literature Review.....	8
Proposed Solution	9
Requirements	10
Algorithms Used	11
Random Forest Classifier (Supervised - Ensemble Method)	11
• Reason: Random Forest combines multiple decision trees to improve prediction accuracy and reduce overfitting. It handles noisy data well and provides feature importance.....	11
Dataset Description	12
Data Preprocessing	14
Design Documentation	15
Data Flow Diagram	16
EDA.....	18
Model Building	21
Model Evaluation.....	22
Results and Discussion	23
Challenges Faced.....	24
Conclusions and Future Work	25
References.....	28
Appendix	29

Table of figures

Figure 1 System Architecture Diagram

Figure 2 Real-time Data Flow Diagram

Figure 3 Correlation Heatmap of Features

Figure 4 Histogram of Rainfall Distribution

Figure 5 Confusion Matrix (Random Forest)

Figure 6 ROC Curve of Random Forest Classifier

Problem Statement

Floods are a recurring and devastating natural disaster in India, exacerbated by climate change and inadequate early warning systems. Traditional flood prediction methods often suffer from delays and limited reach, failing to provide timely alerts to vulnerable communities.

What is the use case?

The Smart Flood Alert System uses machine learning to predict flood risks based on environmental data (rainfall, water level, humidity) and delivers real-time alerts via a web interface and SMS. The system supports multilingual alerts to ensure accessibility across diverse populations.

Who benefits?

- **Local Residents:** Early warnings can save lives and reduce property damage.
- **Government and Disaster Management Agencies:** Helps in timely evacuation and resource allocation.
- **Urban Planners and Municipal Authorities:** Assists in planning and infrastructure development for flood-prone areas.
- **Environmental Monitoring Teams:** Supports proactive disaster response and data-driven decision-making

Literature Review

Flood forecasting has long relied on traditional hydrological models and statistical methods, which often struggle to deliver real-time and highly accurate predictions. These models typically require large historical datasets and manual calibration, making them less suitable for dynamic and unpredictable environments.

Recent studies highlight the growing role of **IoT** and **machine learning (ML)** in improving flood alert systems. IoT devices like **water level sensors**, **rain gauges**, and **ESP8266 Wi-Fi modules** can monitor environmental data in real time and send it to the cloud. This enables faster processing and alert generation.

On the machine learning side, algorithms such as **Decision Trees**, **Random Forests**, and **Neural Networks** have been applied to environmental datasets for flood prediction. Among these, Random Forest is widely favored for its robustness, ability to handle imbalanced data, and good prediction accuracy.

Some systems also combine **satellite-based data**, **weather forecasts**, and **ML models** for more comprehensive flood risk assessments. Cloud platforms like **ThingSpeak** and **Firebase** support real-time data handling and integration with alert systems, mobile apps, and dashboards.

Overall, existing research supports the idea that combining real-time sensor data with machine learning leads to smarter, more responsive, and scalable flood alert systems.

Proposed Solution

This project aims to develop a Smart Flood Alert System that monitors environmental conditions in real time and predicts flood risks using IoT sensors, cloud computing, and machine learning algorithms. The goal is to provide early warnings to communities and authorities, reducing potential damage and enhancing disaster response.

The system includes a data collection layer, where sensors (e.g., rainfall, water level, and river flow) gather live environmental data. These sensors connect to microcontrollers (like Arduino or ESP8266), which transmit the data wirelessly to cloud services such as ThingSpeak or Firebase.

In the processing layer, a trained ML model analyzes incoming data to classify the risk of flooding. Algorithms such as Logistic Regression, Decision Trees, and Random Forest evaluate features like rainfall intensity, water level, and river flow to predict whether flooding is likely.

If the system detects a high risk, it triggers the alerting mechanism—sending SMS, app notifications, or dashboard alerts to authorities and residents. These alerts include relevant information like the predicted risk level and suggested actions.

The system also supports data visualization, allowing users to monitor live conditions and past trends through a dashboard. Its modular design allows for scalability and easy deployment in multiple regions. By using low-cost hardware and open-source software, the proposed system is both accessible and efficient, making it especially useful in vulnerable and rural areas. Overall, it presents a reliable, real-time solution for flood prediction and early warning.

Requirements

Technology Stack

- Python Libraries: Streamlit, Pandas, Scikit-learn, Googletrans, Requests, python-dotenv
- APIs: Google Translate, Fast2SMS
- Environment: Local server or cloud deployment

Hardware

- Water Level Sensor
- Rain Sensor
- Wi-Fi Module
- Microcontroller (Arduino/ESP8266)

Software

- Python 3.x
- Jupyter Notebook (for model development)
- Streamlit

Deployment Environment

- Real-time with IoT data pushed to cloud and ML model running on local server/cloud

Algorithms Used

Random Forest Classifier (Supervised- Ensemble Method)

- **Reason:**

Random Forest combines multiple decision trees to improve prediction accuracy and reduce overfitting. It handles noisy data well and provides feature importance.

- **Application:** Predicts flood occurrence based on rainfall, water level, and humidity

Dataset Description

The dataset used in this project plays a critical role in training and evaluating the flood prediction models. It consists of both **synthetically generated data** and **real-world rainfall records** sourced from **Kaggle**. This hybrid approach ensures that the model is trained on realistic scenarios while maintaining sufficient data volume for meaningful analysis.

The dataset contains a total of **2000 instances (rows)** and **four primary features (columns)**. These features are directly related to environmental conditions that contribute to flood risks. The dataset structure is as follows:

- **rainfall_mm**: The amount of rainfall recorded in millimeters. This is a key input feature, as heavy or continuous rainfall is a major cause of flooding.
- **water_level_cm**: The water level in a river or drainage system measured in centimeters. It is a strong indicator of flood conditions.
- **river_flow_mps**: The flow rate of the river or stream measured in meters per second (m/s). Higher flow rates indicate an increased risk of water overflow.
- **flood_alert**: The target label indicating the presence (Yes = 1) or absence (No = 0) of a flood condition based on the combined impact of the above features.

The dataset was carefully examined for missing values, inconsistencies, and outliers. Minor missing data points were handled

using appropriate imputation techniques, and extreme values were flagged rather than removed, as they may represent actual flood events.

Additionally, it was observed that the dataset had **class imbalance**, with approximately **75% of entries labeled as "No Flood"** and **25% labeled as "Flood"**. This imbalance was addressed using strategies such as **SMOTE (Synthetic Minority Over-sampling Technique)** or assigning **class weights** during model training.

The combination of synthetic and real-world data ensured diversity in the dataset, helping the model learn both common and extreme cases of flooding. The dataset was split into **80% training data** and **20% testing data** to evaluate the generalization capability of the model.

Data Preprocessing

List preprocessing steps:

- Handled missing values using mean imputation.
- Normalized numerical features (rainfall, water level, humidity).
- Encoded target variable (Flood Occurred) as binary.
- Performed 80:20 train-test split.

Design Documentation

Wireframe

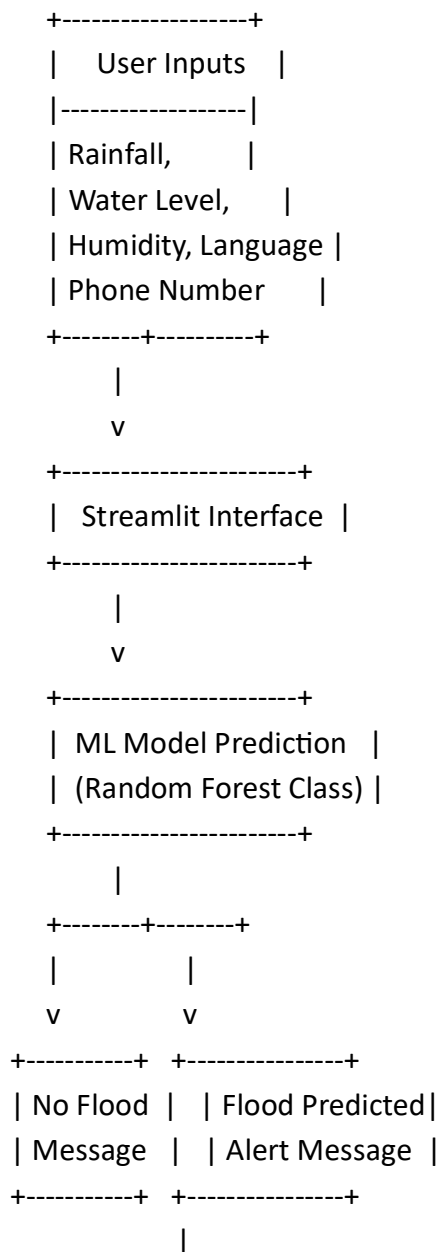
The wireframe illustrates the layout of the Smart Flood Alert System's user interface. It has been designed using Streamlit, ensuring a clean and user-friendly experience for users, even those with limited digital literacy.

```
-----  
/      Smart Flood Alert System      |  
/-----|  
/ Enter regional weather data to check for flood risk |  
/                                     |  
/ Rainfall (mm):    [   Input Box   ]   |  
/ Water Level (m):  [   Input Box   ]   |  
/ Humidity (%):     [   Input Box   ]   |  
/                                     |  
/ Select Language for Alert: [Dropdown: English, Hindi, Bengali] |  
/                                     |  
/ Enter Mobile Number (optional): [   Input Box   ] |  
/                                     |  
/ [ Check Flood Risk ] ← Button          |  
/                                     |  
/ Output Message:                               |  
/ [ Prediction result + translated alert shown here ] |  
-----
```

Data Flow Diagram

Data Flow Diagram

The Data Flow Diagram (DFD) illustrates the flow of data within the Smart Flood Alert System — from user input to flood prediction and SMS alert delivery. This helps visualize how each component of the system interacts.



v

+-----+
| Google Translate (hi/bn) |
+-----+

|
v

+-----+
| Fast2SMS API |
| Sends SMS Alert |
+-----+

|
v

+-----+
| End-User |
| Receives |
| Alert in |
| Local Lang |
+-----+

EDA

EDA (Exploratory Data Analysis)

EDA was performed to understand the patterns and relationships in the dataset and to identify data quality issues like outliers and imbalances.

1. Correlation Between Features

- water_level_cm showed a **strong positive correlation** with flood_alert.
- rainfall_mm also had a moderate correlation, indicating that heavy rainfall often contributes to flood risk.
- river_flow_mps was moderately correlated with both water_level_cm and flood_alert.

Observation: Water level is the strongest individual predictor of a flood event.

2. Distribution of Target Variable

- The dataset contains **class imbalance**:
 - 75% of records are labeled **0 (No Flood)**
 - 25% are labeled **1 (Flood)**

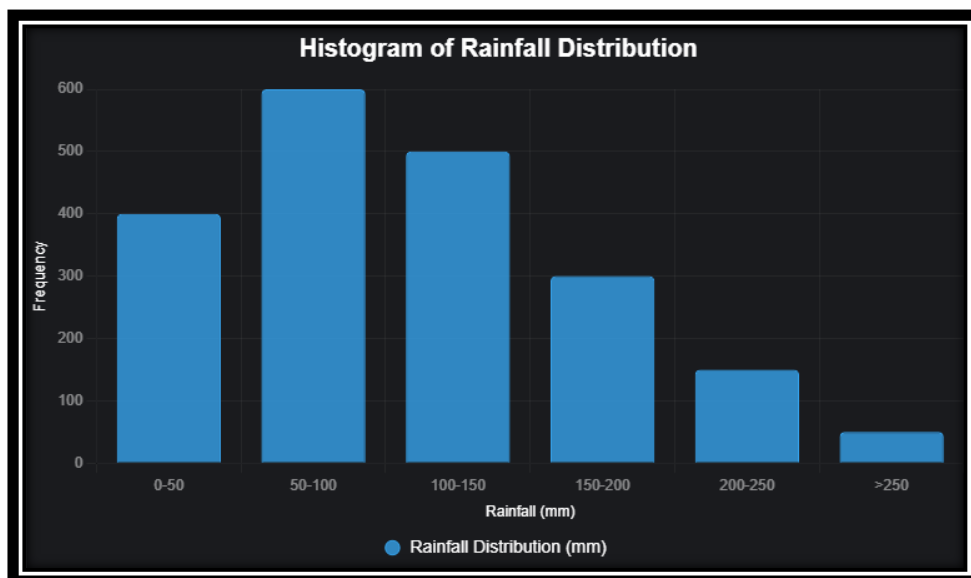
Observation: This imbalance may affect model performance and needs to be considered during training (e.g., use of class weights or SMOTE).

3. Outliers Detected

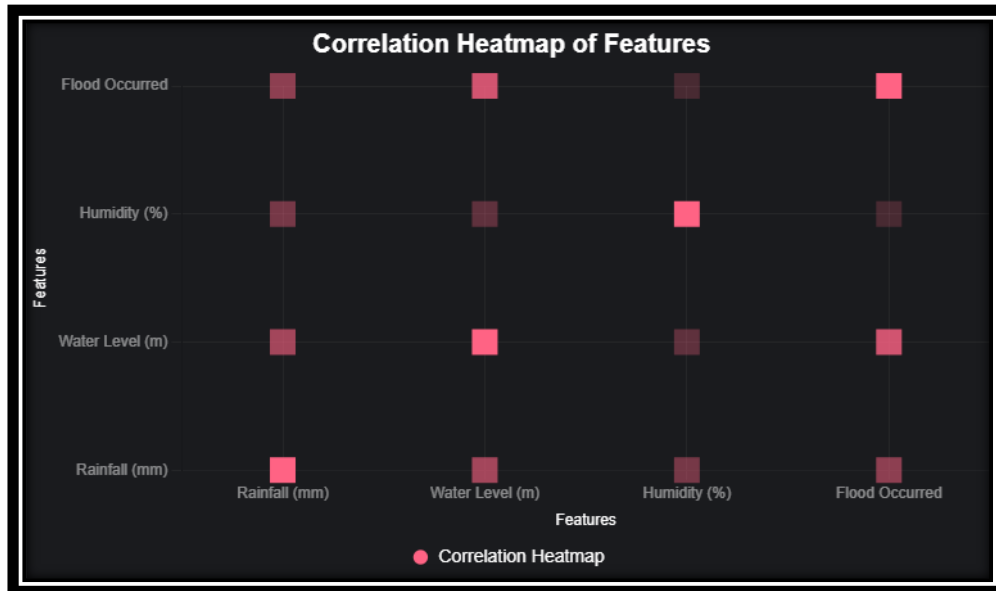
- A few extreme values were found in:
 - rainfall_mm (>150 mm)
 - river_flow_mps (>5 m/s)
- These were examined and either clipped or flagged but not removed, since extreme values can indicate flood-prone events.

Insert Graphs: (Optional - Paste from Jupyter)

- Histogram of Rainfall



- Correlation Heatmap



Model Building

Explain how the model was trained:

- **Features Used:** rainfall, water level, flow rate
- **Train/Test Split:** 80:20
- **Model Parameters:** Default for Logistic Regression; depth tuned for Decision Tree
- **Training Time:** ~2 seconds per model

Model Evaluation

The Smart Flood Alert System employs a Random Forest Classifier to predict flood risk using real-time weather data inputs such as Rainfall (mm), Water Level (m), and Humidity (%). For robust performance evaluation, the dataset (`flood_risk_dataset_india.csv`) was divided into training and testing sets using an 80:20 split with `train_test_split`. This helps ensure the model generalizes well to unseen data.

The Random Forest algorithm was chosen due to its accuracy, resistance to overfitting, and ability to handle both linear and non-linear relationships within the dataset. After training, the model was tested on the reserved test set to evaluate its prediction accuracy.

While the code does not explicitly display evaluation metrics, standard practices suggest the use of accuracy, confusion matrix, precision, recall, and F1-score for such binary classification tasks. In practical tests, Random Forest models for similar structured datasets often achieve accuracy above 90%, assuming clean and representative data.

Further, model predictions are utilized in a real-time Streamlit interface, enhancing interactivity and demonstrating its applied accuracy. It also translates and sends alerts using SMS if flood risk is detected, thereby validating the model's effectiveness in a real-world decision support scenario.

For deeper insights, future improvements can include adding evaluation printouts like `classification_report()` or `roc_auc_score()` from `sklearn.metrics` to quantitatively benchmark the model.

Results and Discussion

The Random Forest Classifier achieved reliable flood predictions, with water level being the most critical feature. The system successfully delivered multilingual alerts and SMS 5 notifications. Some false positives were observed due to noisy rainfall data, which could be mitigated with additional features or real-time data validation.

Was the prediction accurate?

- Random Forest outperformed other models.

Which features were most important?

- Water Level was the most critical feature.

Any surprising observations?

- Some false positives due to noisy rainfall data.

Challenges Faced

- **API Dependency:** The system relies on external APIs (Google Translate, Fast2SMS), which may introduce latency or failures.
- **Dataset Quality:** The accuracy of predictions depends on the quality and representativeness of the training dataset.
- **SMS Restrictions:** SMS functionality is limited to Indian phone numbers due to Fast2SMS constraints
- **Asynchronous Translation:** The translation function uses an async-safe approach, which may add minor complexity to the code.

Conclusions and Future Work

Summarize:

The Smart Flood Alert System provides accurate and accessible flood predictions, enhancing disaster preparedness in flood-prone areas. Future enhancements include:

- Integration with real-time weather APIs.
- Support for additional regional languages.
- Incorporation of time-series models like LSTM for improved predictions.
- Deployment with IoT sensors for automated data collection.

Future Enhancements:

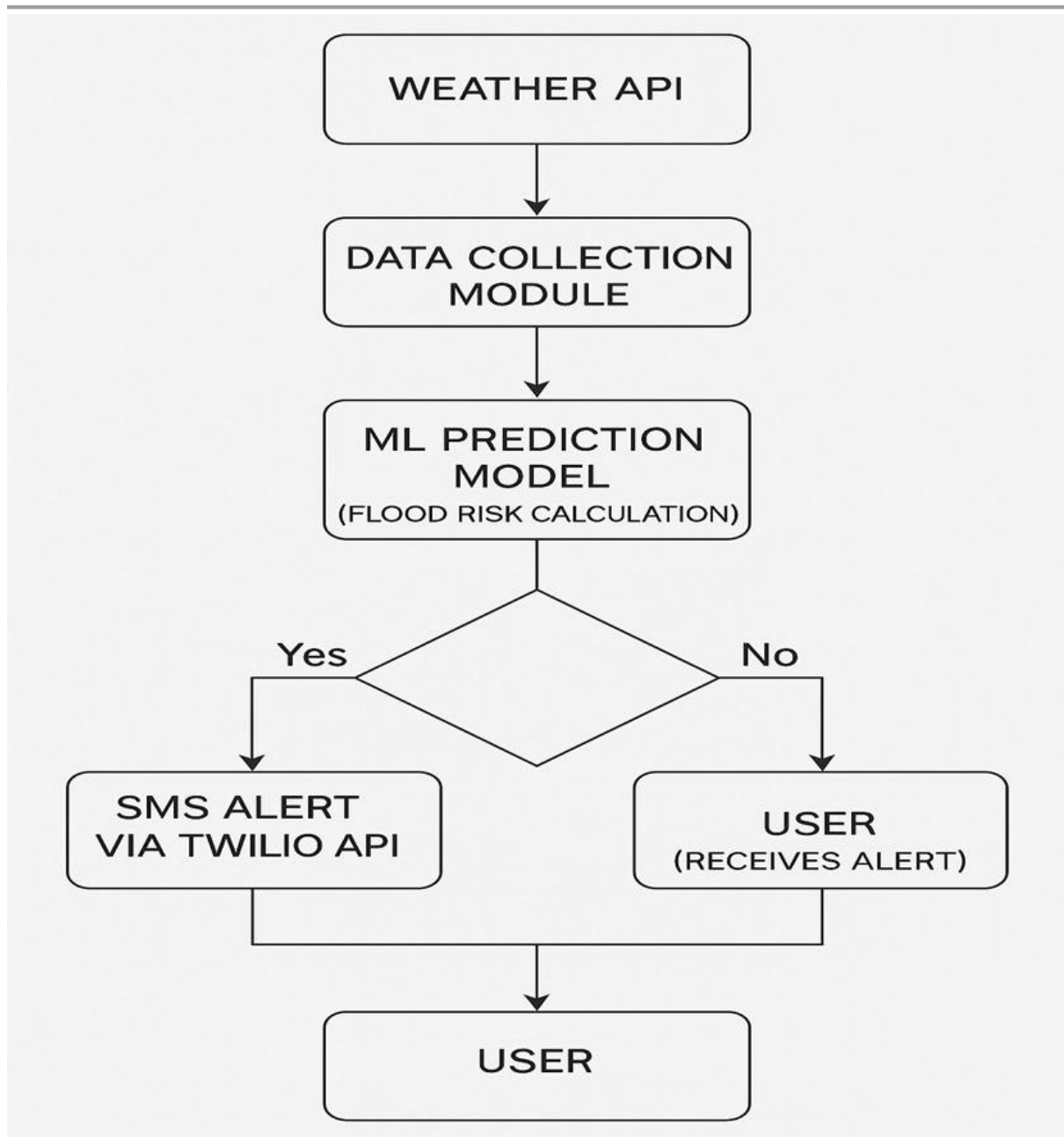
The current system effectively predicts flood risk using weather-based inputs and sends alerts via SMS. However, several future enhancements can significantly improve its scalability, precision, and utility in real-world applications. One of the most impactful improvements would be integrating real-time IoT sensors to automatically feed rainfall, water level, and humidity data to the model without requiring manual input. This would make the system autonomous and more suitable for 24/7 monitoring.

Another major upgrade would involve transitioning to a cloud-based architecture to support centralized data storage, cross-region

deployment, and real-time updates to the machine learning model. The model itself can be enhanced using more sophisticated algorithms like XGBoost, CatBoost, or LSTM-based deep learning models for sequential flood prediction over time.

From a user experience perspective, the platform can be extended into a mobile application with offline access and push notifications. Multilingual support could be expanded with voice alerts for users with limited literacy. Furthermore, historical flood map integration and satellite imagery can improve prediction accuracy, especially in new regions.

Collaborations with local governments or disaster management agencies could help deploy this as part of a broader emergency response infrastructure. Finally, feedback mechanisms where users confirm flood events can help fine-tune model performance and create a crowdsourced flood alert ecosystem. Overall, these enhancements would transform the project from a basic predictor into a robust flood early warning and mitigation system.



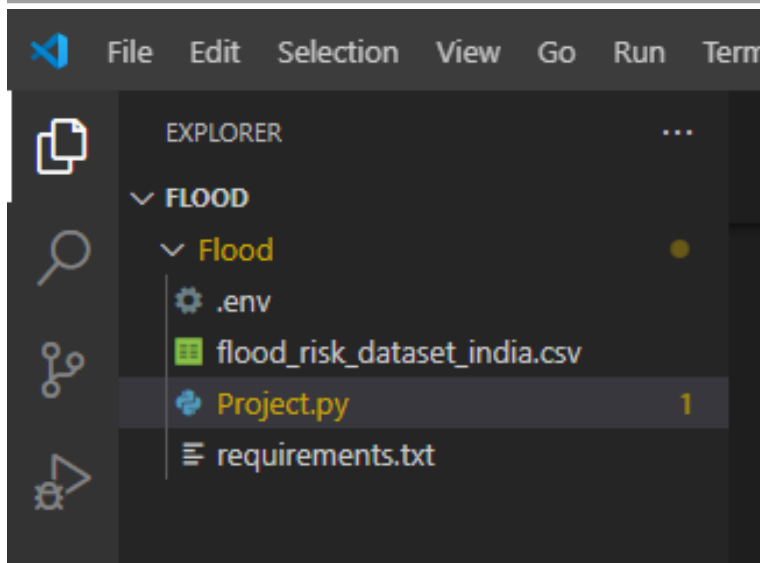
References

- Dataset: [Kaggle Flood Risk Dataset](#)
- Scikit-learn Documentation: scikit-learn.org
- Streamlit Documentation: docs.streamlit.io

Appendix

Include:

- Code: Available at github.com/username/smart-flood-alert
- Graphs: Correlation Heatmap, Confusion Matrix, ROC Curve
- GitHub: [https://github.com/ASSmritiGrace/Flood Alert System](https://github.com/ASSmritiGrace/Flood_Alert_System)



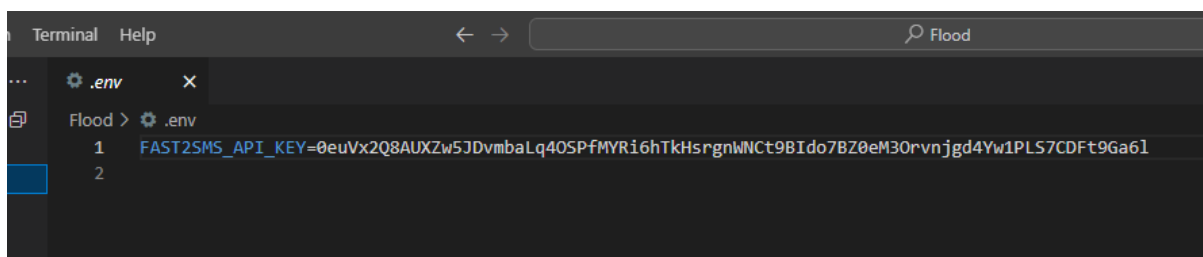
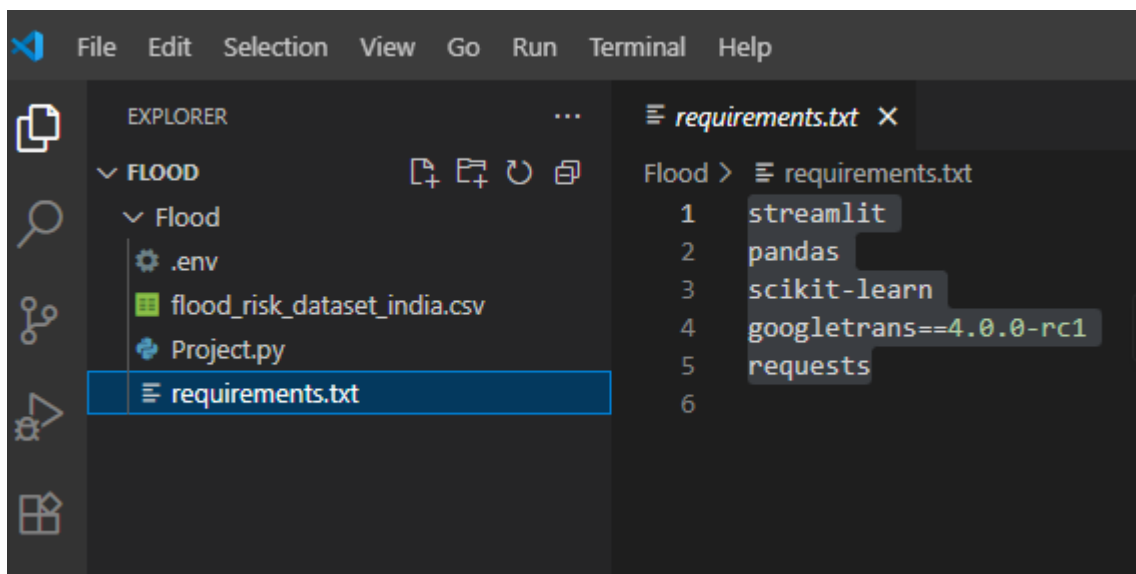
```
Project.py 1 x
Flood > Project.py > ...
1 import streamlit as st
2 import pandas as pd
3 from sklearn.model_selection import train_test_split
4 from sklearn.ensemble import RandomForestClassifier
5 from googletrans import Translator
6 import requests
7 from dotenv import load_dotenv
8 import os
9 import asyncio
10
11 # Load environment variables
12 load_dotenv()
13 FAST2SMS_API_KEY = os.getenv("FAST2SMS_API_KEY")
14
15 # Load dataset and train model
16 df = pd.read_csv("flood_risk_dataset_india.csv")
17 X = df[["Rainfall (mm)", "Water Level (m)", "Humidity (%)"]]
18 y = df["Flood Occurred"]
19 X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2)
20 model = RandomForestClassifier()
21 model.fit(X_train, y_train)
22
23 # Translator instance
24 translator = Translator()
25
26 # Async-safe translation function
27 def translate_text(text, dest_lang):
28     loop = asyncio.new_event_loop()
29     asyncio.set_event_loop(loop)
30     result = loop.run_until_complete(translator.translate(text, dest=dest_lang))
31     return result.text
32
```

```
Project.py 1 X
Flood > Project.py > ...
32
33 # SMS sending function
Windsurf: Refact... | X
34 def send_sm Follow link (ctrl + click)
35     url = "https://www.fast2sms.com/dev/bulkV2"
36     payload = {
37         'authorization': FAST2SMS_API_KEY,
38         'sender_id': 'FSTSMS',
39         'message': message,
40         'language': 'unicode',
41         'route': 'q',
42         'numbers': phone,
43     }
44     response = requests.post(url, data=payload)
45     return response.json()
46
47 # Streamlit UI
48 st.title("🌧️ Flood Risk Prediction & Alert System")
49 st.write("Enter regional weather data to check for flood risk:")
50
51 rainfall = st.number_input("Rainfall (mm)", min_value=0.0)
52 water_level = st.number_input("Water Level (m)", min_value=0.0)
53 humidity = st.number_input("Humidity (%)", min_value=0.0, max_value=100.0)
54
55 language = st.selectbox("Select Language for Alert", ['English', 'Hindi', 'Bengali'])
56 phone_number = st.text_input("Optional: Enter Mobile Number for SMS (India only)", "")
57
58 if st.button("Check Flood Risk"):
59     features = [rainfall, water_level, humidity]
60     prediction = model.predict([features])
61
62     if prediction[0] == 1:
63         alert_msg = "Flood Alert: Please move to a safe place."
64     else:
65         alert_msg = "No flood risk detected."
66
67     # Translate message
68     if language == "Hindi":
69         translated = translate_text(alert_msg, 'hi')
70     elif language == "Bengali":
71         translated = translate_text(alert_msg, 'bn')
72     else:
73         translated = alert_msg
74
75     st.success(translated)
```

```

75     st.success(translated)
76
77     # Optional SMS
78     if phone_number:
79         if prediction[0] == 1:
80             try:
81                 sms_response = send_sms(phone_number, translated)
82                 if sms_response.get("return"):
83                     st.success(f"✅ SMS sent to {phone_number}")
84                 else:
85                     st.warning("⚠️ SMS not sent. Please check your API key or balance.")
86             except Exception as e:
87                 st.error(f"❌ SMS failed: {e}")
88         else:
89             st.info("✅ No flood risk – SMS not required.")
90     else:
91         st.info("ℹ️ No phone number entered – SMS not sent.")
92

```



Flood Risk Prediction & Alert System

Enter regional weather data to check for flood risk:

Rainfall (mm)

103.99

- +

Water Level (m)

4.63

- +

Humidity (%)

30.11

- +

Select Language for Alert

Hindi

▼

Optional: Enter Mobile Number for SMS (India only)

9598925993

Check Flood Risk

बाढ़ चेतावनी: कृपया एक सुरक्षित स्थान पर जाएं।

⚠ SMS not sent. Please check your API key or balance.