Report On,

Data-Driven Flood Vulnerability Assessment: Integrating Environmental Parameters in Machine Learning

Course: Mastering Python Programming

Dec 6, 2024

EDGE: BU-CSE Digital Skills Training

UNIVERSITY OF BARISHAL

Submitted to:
Md Mahbub E Noor
Lecturer,
Computer Science and Engineering

Submitted by:
Md. Omar Faruk
Roll: 09-004-16
Batch: DS 4

Abstract

Flooding is a critical challenge, particularly in coastal and low-lying areas like Chittagong, Bangladesh, where urbanization and climate change exacerbate the risk. This report presents a flood vulnerability prediction model using machine learning, specifically a Random Forest, Decision Tree, Logistic Regression Classifier to assess flood susceptibility based on environmental parameters such as rainfall, elevation, slope, land use/land cover (LULC), and soil texture. Decision Tree Classifier was performed best for better output. By integrating both numerical and categorical features, the model achieves high accuracy in predicting flood vulnerability levels, offering valuable insights for disaster management, urban planning, and resource allocation. This data-driven tool enhances decision-making by providing real-time predictions and can be adapted to other flood-prone regions, helping to improve resilience and mitigate flood impacts.

Table of Contents

1. Introduction

Flooding poses a significant risk to communities worldwide, particularly in coastal areas like Chittagong, Bangladesh, where urbanization, complex river networks, and varied land use exacerbate the problem (Dey, 2023; Miah et al., 2020). Understanding flood vulnerability is critical to developing effective mitigation strategies. This report outlines the development of a machine learning-based model that predicts flood vulnerability levels based on key environmental and geographical criteria. The model employs a Decision Tree Classifier to analyze the relationship between critical factors such as rainfall, elevation, slope, proximity to rivers, land use, and soil texture, offering a systematic approach to assess vulnerability and support flood risk management.

2. Rationale of the Study

Chittagong, being one of Bangladesh's most flood-prone districts, faces recurrent flooding due to its geographical location, dynamic river systems, and climate variability (Mazumdar, 2023). The region's vulnerability is further exacerbated by urban expansion, population growth, and unsustainable land use practices (Tanim & Goharian, 2021). Despite existing flood risk studies, there is a gap in predictive models that integrate spatial, environmental, and land-use factors to generate actionable insights. This study provides a novel approach by using a machine learning model to classify areas into flood vulnerability levels based on key indicators. Such a tool can significantly aid in disaster preparedness, infrastructure planning, and resource allocation. Furthermore, the model's ability to predict flood vulnerability based on user input makes it a valuable decision support system for stakeholders, ensuring localized, data-driven, and efficient flood risk management.

3. Objectives

3.1 Key Goals of the Study

- Develop a flood vulnerability prediction model using machine learning.
- Analyze the significance of key environmental parameters in determining flood risks.
- Provide an interactive tool for real-time flood vulnerability prediction based on user input.

3.2 Target Applications of the Model

- Disaster management and preparedness planning.
- Urban development policies to minimize flood risks.

4. Main Contribution

This project introduces a data-driven approach to flood risk assessment, demonstrating the following contributions:

- Integration of numerical and categorical environmental features in a predictive model.

- Utilization of Random Forest, Decision Tree, Logistic Regression Classifier to achieve a balance between accuracy and interpretability.
- Development of an interactive prediction system for user-specified environmental conditions.

5. Methodology

5.1 Data Sources

- This study mostly utilized secondary data calculated from ArcGIS 10.8 software, encompassing satellite imagery (2024), Digital Elevation Model (DEM), major river locations, and soil characteristics to achieve the study objectives. Additionally, interviews were taken with six urban flood management specialists, including engineers, planners, and technical officials from flood management institutions. Table 1 provides a summary of the data sources used.
- Categorical data such as LULC and soil texture are encoded using Label Encoding to make them suitable for machine learning analysis.

Table 1 Data sources of the study.

Data Type	Format	Date	Source	Output
Satellite image (Landsat 8) (10m resolution)	Raster	2024	United States Geological Survey (USGS) Earth explorer	Land use and land cover
Digital elevation model (DEM) (30m resolution)	Raster	2014	United States Geological Survey (USGS) Earth explorer	Elevation, Slope
Location of Rivers	Vector	2024	Local Government Engineering Department (LGED)	Distance from river
Soil Texture	Vector	2014	Bangladesh Agricultural Research Council (BARC)	Soil texture
Documented interviews	Text	2024	Field (Study area)	-----

5.2 Data Preprocessing:

- Data cleaning, including handling missing values and encoding categorical variables.
- Normalization of numerical features for effective model training.

5.3 Exploratory Data Analysis (EDA):

- Statistical summarization and visualization of features (e.g., histograms, count plots).

5.4 Model Development:

- Feature-target segregation and dataset splitting into training and test sets.
- Implementation of Random Forest, Decision Tree, Logistic Regression Classifier for prediction.

5.5 Evaluation Metrics:

- Assessment using accuracy, precision, recall, and F1 score.

5.6 Interactive Prediction Tool:

- A user-input-based prediction system integrates preprocessing and a trained model.

6. Experiment

6.1 Dataset Characteristics

- Numerical features: Rainfall, Elevation, Slope, Distance to River.
- Categorical features: LULC, Soil_Texture.
- Target variable: Flood Vulnerability Level.

6.2 Implementation Steps

- Data preprocessing, feature encoding, and normalization.
- Training and testing the Random Forest, Decision Tree, Logistic Regression Classifier.

6.3 Interactive Prediction Framework

A user-input interface was developed to enable real-time flood vulnerability predictions.

7. Results and Discussions

7.1 Performance Metrics of the Model

Logistic Regression, Decision Tree, and Random Forest models have been applied. Each model's accuracy, precision, recall, and F1 Score are mentioned in Table 2.

Table 2: Performance Metrics of the Model

Model	Accuracy	Precision	Recall	F1 Score
Logistic Regression	0.54	0.6	0.54	0.55
Decision Tree	0.85	0.9	0.85	0.84
Random Forest	0.77	0.8	0.77	0.77

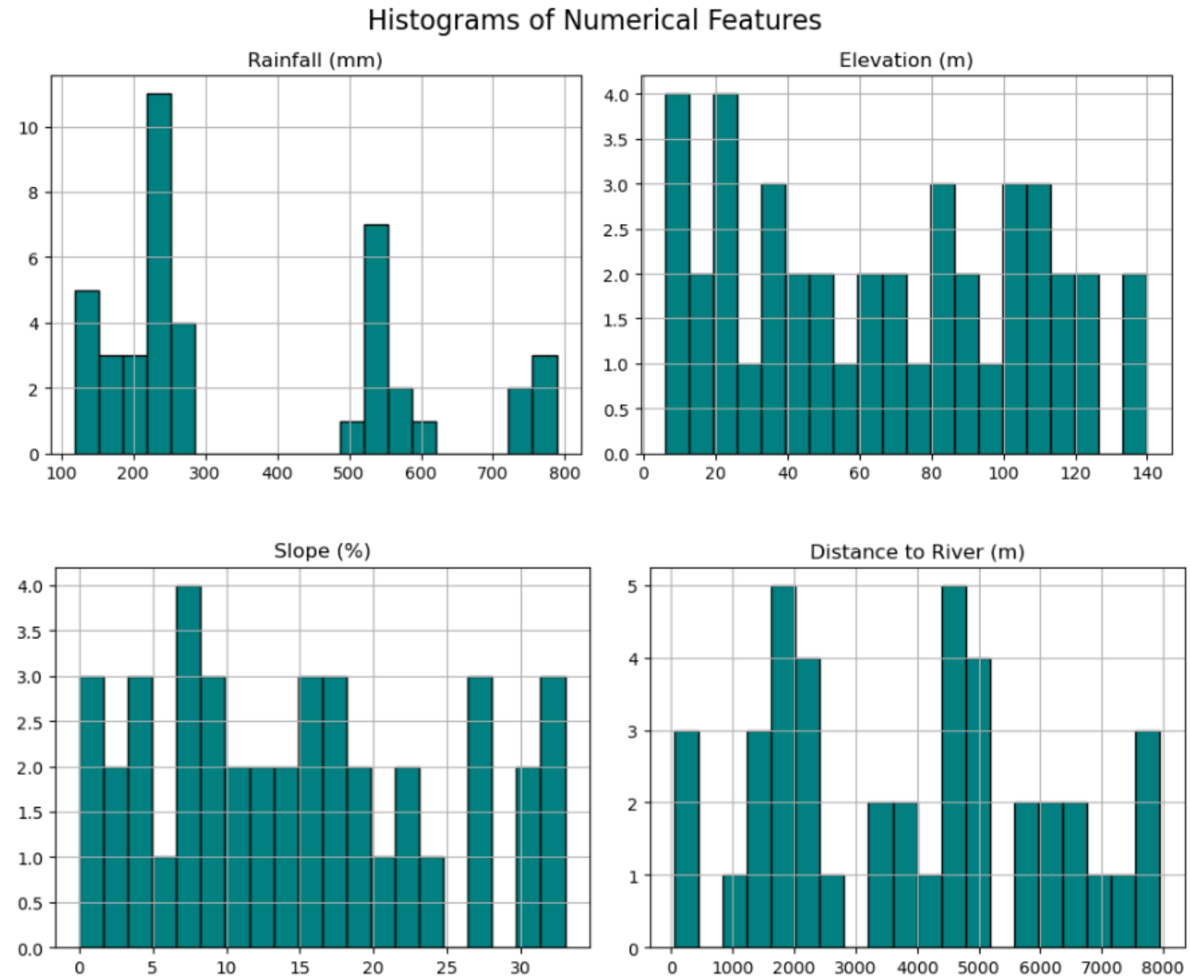
7.2 Key Observations and Interpretations

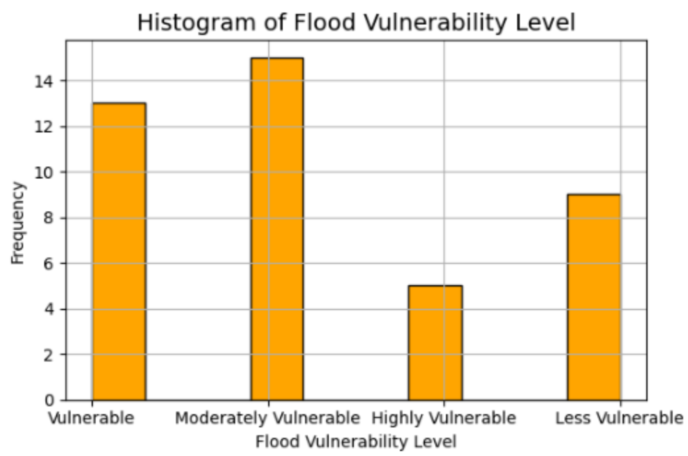
- Rainfall and elevation emerged as critical predictors of flood risk.
- Misclassifications highlight areas for future refinement.

7.3 Visualization Insights

- Histograms demonstrated skewness in rainfall distribution.
- Count plots emphasized variability in LULC and soil texture across regions.

The following histograms visually summarize data distributions, highlight patterns, and make complex numerical information easier to interpret, enhancing clarity and impact.





8. Future Works

8.1 Enhancements to the Model

- Incorporate spatial data like floodplain maps and drainage patterns.
- Explore advanced machine learning models, such as ensemble techniques or neural networks.

8.2 Broader Applications

- Real-time flood monitoring systems.
- Decision-support tools for policymakers and urban planners.

9. Conclusion

This proposal demonstrates a novel application of machine learning to predict flood vulnerability in the Chittagong district. By leveraging critical environmental factors, the model provides actionable insights, enabling stakeholders to mitigate the impacts of floods effectively. The development of this decision-support tool will enhance resilience in flood-prone regions, aligning with global goals for sustainable disaster risk reduction.

10. References

Dey, A. B. (2023, September 24). *Chattogram city condemned to waterlogging*. The Daily Star. https://www.thedailystar.net/news/bangladesh/news/chattogram-city-condemned-waterlogging-3426161

Mazumdar, M. D. H. (2023, August 11). *What caused the catastrophic floods in Chattogram?* The Daily Star. https://www.thedailystar.net/opinion/views/news/what-caused-the-catastrophic-floods-chattogram-3391791

Miah, J., Hossain, K. T., Hossain, M. A., & Najia, S. I. (2020). Assessing coastal vulnerability of Chittagong District, Bangladesh using geospatial techniques. *Journal of Coastal Conservation*, *24*(6), 66. https://doi.org/10.1007/s11852-020-00784-2

Tanim, A. hasan, & Goharian, E. (2021). Developing a hybrid modeling and multivariate analysis framework for storm surge and runoff interactions in urban coastal flooding. *Journal of Hydrology*, *595*, 125670. https://doi.org/10.1016/j.jhydrol.2020.125670