Flood Forecasting using **Machine Learning**

B.Tech. Project

Presented By: Saurabh Tiwari (2020CE10295)
Supervised By: Prof. A. K. Keshari

Introduction

- The most frequent natural disaster in the world, flooding affects hundreds of millions of people and kills between **6,000** and **18,000** people annually, with **20%** of those deaths occurring in India.

- India is the country with the **highest** annual risk of flooding in the entire world. An area of more than **40 million ha.** in India has been identified as flood prone. On an average, floods have affected about **33 million** persons between 1953-2000.

- Floods cause an economic loss of nearly **40 billion USD** annually worldwide, and **15 billion USD** per year in India alone.

Problem Statement

Due to the **large amounts of flood-related data** available for the US, creating a model to predict and determine future floods in the **US** is a relatively simple task.

However, given the **extremely limited data** available regarding floods in India, the problem is to innovate a method to create a prediction model for the **Indian Subcontinent**.

Objectives

1. Data Collection

Collect and integrate various data sources, including meteorological data, and historical flood records, particularly in the challenging context of the Indian Subcontinent.

2. Machine Learning Model

Develop and implement machine learning models like logistic regression and Random Forest Classification to enhance flood prediction accuracy.

3. Data Visualization

The objective is to employ advanced data visualization techniques to enhance the interpretability and accessibility of flood-related insights.

4. Web App Development

Create a web app capable of providing real-time flood forecasts to aid in disaster preparedness and response.

Methodology

Data Collection

1. My goal was to predict floods from weather data using machine learning.
2. Obtaining flood-related data at a sufficient scale over the Indian Subcontinent proved to be a challenging task, with limited available data.
3. As such, I innovated a two-step method of obtaining this data, which first mines flood reports from news sources, and then retrieves historical weather conditions based on temperature, maximum temperature, humidity, cloud cover, windspeed, and precipitation in accordance with the extracted flood data.
4. This method aims to form a connection between environmental factors and floods.
5. Thus, for the dataset, I first **scraped** the website **http://floodlist.com/tag/india** using the **python beautiful-soup 4 library**.
6. This website provided me with information about past and current floods India, as well as their date and location.
7. I then used the **Visual Crossing weather API** to obtain historic weather data such as precipitation, humidity, temperature, cloud cover, and windspeed in those areas and during those times.

Methodology

Machine Learning Model

1. My machine learning model is based on the python **sci-kit learn** library.
2. I used **pandas** to generate a data-frame for the dataset, and then tried various machine learning models from Logistic Regression to K-Nearest Neighbours to **Random Forest Classification**.
3. I proceeded to save the model with highest accuracy in a pickle file.

Methodology

Data Visualization

1. I first obtained a dataset of the major cities and towns in India (around 200 of them) along with their latitude, longitude, and population.
2. I then obtained the numerous weather factors in each city using the **weather API** and ran the data into my machine learning model.
3. Next, I plotted the data from the model on various different types of maps, using **Plotly chart studio**.
4. The maps represent various data such as flood prediction, precipitation analysis, and damage estimates, in the form of **scatter plot**s, **bubble plots** and **density maps**.
5. Quantifying flood damage is challenging due to the complex nature of such events. To better understand the impact, I've adopted a method that involves **multiplying precipitation by population**.
6. This approach aims to offer a reference for estimating potential flood damage, considering the interaction of meteorological factors and affected population density.

Methodology

Web app Frontend and Hosting

1. My web app is based on the **Flask** python framework.
2. I rendered **HTML** templates – with **CSS** for styling and **JavaScript** for added functionality – and integrated it with my machine learning models and datasets via the flask back-end.
3. I then used **AWS Elastic Beanstalk** hosting service to host my web application for everyone to try.

Literature Survey

1. Ambore, Charan, et al. (2023): The authors presents an examination of diverse machine learning algorithms for predicting floods based on rainfall data in Indian districts. Previous works on flood prediction using techniques like support vector machines, neural networks, image processing, and big data analytics are reviewed. The employed machine learning algorithms in this study include XGBoost, K Nearest Neighbours, Decision Trees, Logistic Regression, and Random Forest.

2. Mosavi, Ozturk, et al. (2018): The report presents the challenges and importance of flood prediction for risk reduction and water resource management. The advantages and disadvantages of different types of models, such as physically based, statistical, and data-driven models.

3. Xiong & O'Connor (2002): The author presents a comparison of four updating models for real-time river flow forecasting, based on the simulation errors of a rainfall-runoff model. The author recommends the use of the AR model for error-forecast updating in real-time river flow forecasting.

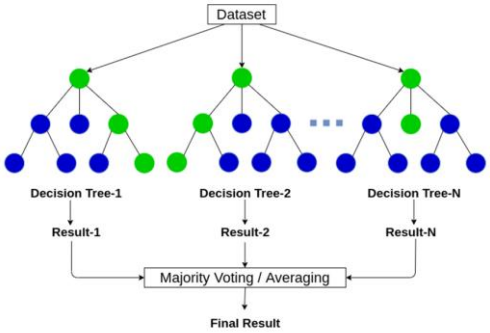
4. Nevo, Morin et al. (2018): Google's operational flood warning system uses machine learning models to provide accurate and timely alerts to agencies and the public in India and Bangladesh. The system consists of four subsystems: data validation, stage forecasting, inundation modelling, and alert distribution.

Results and Discussion

Data used for model training and plotting

Temp	Max_Temp	Wind_Speed	Cloudcover	Precip	Humidity	Class
29.90	37.00	23.00	58.60	146.00	83.86	1
93.04	82.41	63.44	102.42	192.07	129.30	1
87.13	98.82	72.29	122.12	191.80	142.25	1
92.12	72.39	73.84	96.50	196.53	150.42	1
92.64	76.03	67.13	92.61	191.76	126.61	1
			...			
61.43	77.45	80.24	77.38	81.32	124.24	0
72.26	78.98	47.48	73.92	67.72	113.09	0
76.40	82.69	79.23	58.95	65.84	122.76	0
61.03	74.17	70.44	61.39	64.09	106.96	0
85.19	64.06	73.84	56.96	71.45	109.65	0

[5040 rows x 7 columns]



City	Lat	Lon	Precip	Class	Population	Damage
Mumbai	18.987807	72.836447	1.57	0	12691836	1992.62
Delhi	28.651952	77.231495	0.00	0	10927986	0.00
Kolkata	22.562627	88.363044	0.00	0	4631392	0.00
Chennai	13.084622	80.248357	1.63	0	4681087	763.02
Bengaluru	12.977063	77.587106	0.15	0	8443675	126.66
			...			
Calicut	11.248016	75.780402	3.07	0	0	0.00
Kagaznagar	19.331589	79.466051	0.44	0	0	0.00
Jaipur	26.913312	75.787872	0.00	0	2711758	0.00
Ghandinagar	23.216667	72.683333	1.34	0	0	0.00
Panchkula	30.691512	76.853736	0.00	0	200000	0.00

[212 rows x 7 columns]

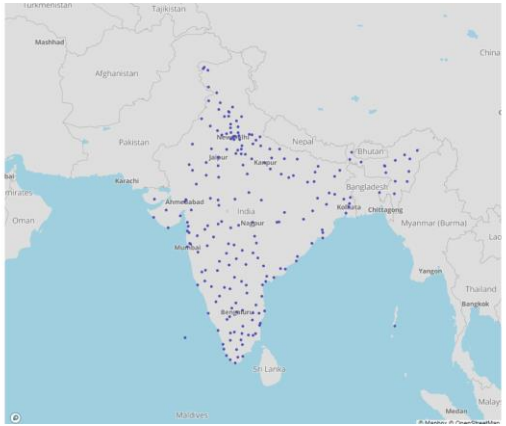
1. Final data required for ML training.

2. Random Forest Classifier gave the highest accuracy of 98.71% on the test set.

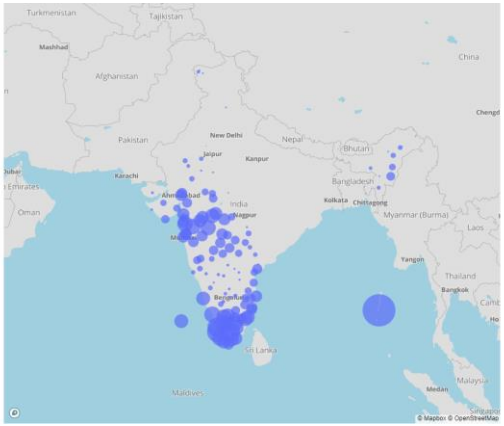
3. Final data for plotting.

Results and Discussion

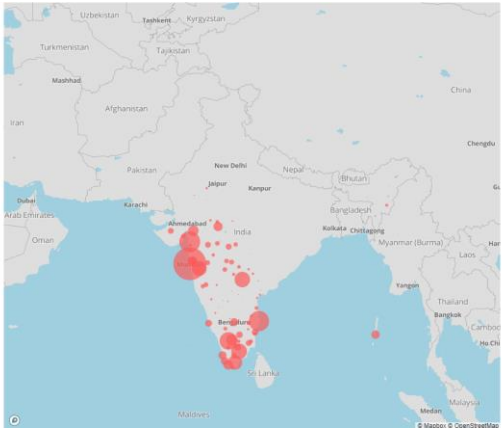
Scatter plot and bubble plots



1. Flood prediction plot, flood marked by red dots.



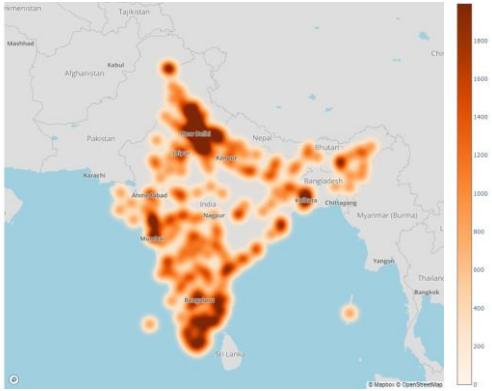
2. Precipitation plot, larger bubbles indicating more precipitation.



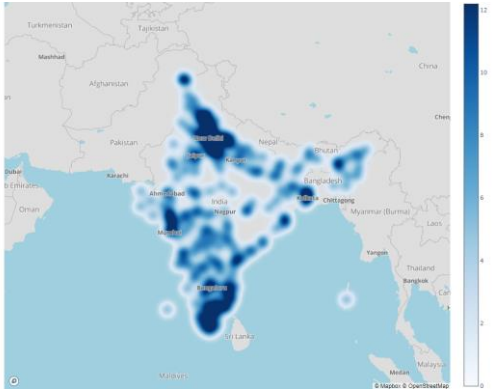
3. Damage analysis plot, size of the bubbles indicates damage.

Results and Discussion

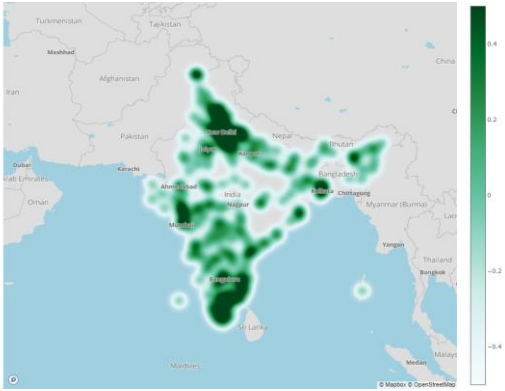
Density Maps



1. Damage analysis density map, colour scale of the density map indicates damage.



2. Precipitation density map, darker areas indicating greater precipitation.

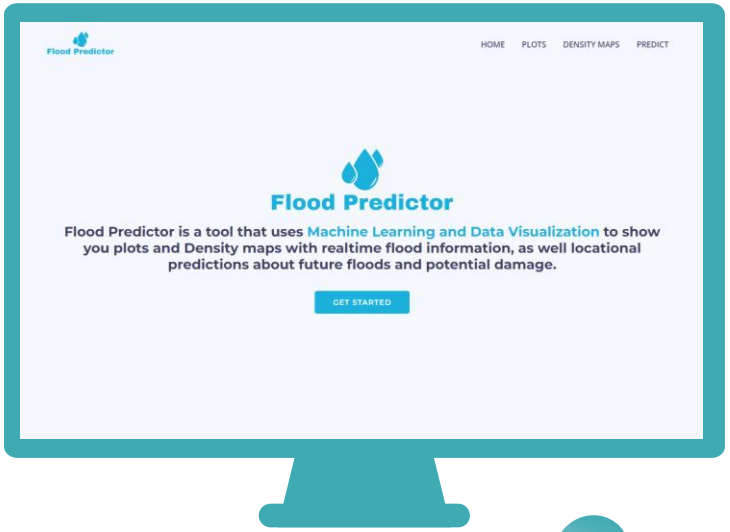


3. Flood prediction map, flood is most likely marked by darker spots.

Results and Discussion

Web app!

- **GitHub Repo**: GitHub - SaurabhTiwari4093/Flood-Forecasting
- **URL**: https://flood-predictor.iccntsd.in/



Conclusion

Data Challenges

Overcame limited data obstacles in flood-related research through innovative data mining and **scraping** techniques, ensuring a comprehensive dataset.

Modelling Success

Proudly achieved a machine learning model accuracy **exceeding 98%**, showcasing substantial growth in modelling skills and effective integration of data manipulation techniques.

Continuous Learning

Expanded expertise in web scraping, data mining, and Plotly, marking a continuous learning journey, with a specific focus on enhancing machine learning skills. Additionally, learned **LaTeX** to effectively create and format report.

Plan for future

- I would like to expand my web application to cover cities in **countries all around the world**, and to make my visualizations and predictions accessible to people across all nations.

- Facilitate **real-time updates** to the plot and density map on the web application seamlessly, eliminating the need for manual intervention.

- Furthermore, I recognize the potential of **incorporating satellite data** for enhanced flood detection and plan to develop an image classification model for this purpose.

References

1. Amazon Web Service: https://aws.amazon.com/

2. Anil Kumar Ambore, T. Sri Sai Charan, U. Rohit Reddy, T. Samara Simha Reddy, Tarun. G (2023), Flood Prediction using Machine Learning, International Journal for Research in Applied Science & Engineering Technology (IJRASET) ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538 Volume 11 Issue V May 2023

3. Beautiful soup 4: https://pypi.org/project/beautifulsoup4/

4. Flask: https://flask.palletsprojects.com/en/3.0.x/

5. FloodList: https://floodlist.com/

6. LIHUA XIONG & KIERAN M. O'CONNOR (2002) Comparison of four updating models for real-time river flow forecasting, Hydrological Sciences Journal, 47:4, 621-639, DOI: 10.1080/02626660209492964

References

7. Mosavi, A., Ozturk, P., & Chau, K.-w. (2018). Flood prediction using machine learning models: Literature review. Department of Computer Science (IDI), Norwegian University of Science and Technology (NTNU), NO-7491 Trondheim, Norway; Department of Civil and Environmental Engineering, Hong Kong Polytechnic University, Hong Kong, China; dr.kwok-wing.chau@polyu.edu.hk. Correspondence: amir.mosavi@ntnu.no (A.M.); pinar@ntnu.no (P.O.). Received: 1 September 2018; Accepted: 17 October 2018; Published: 27 October 2018.

8. Nevo, S., Morin, E., Rosenthal, A. G., Metzger, A., Barshai, C., Weitzner, D., ... Matias, Y. (2023). Flood forecasting with machine learning models in an operational framework. Google Research, Yigal Alon 96, Tel-Aviv 6789141, Israel; Hebrew University of Jerusalem, Safra Campus, Jerusalem 91904, Israel.

9. Plotly chart studio: https://chart-studio.plotly.com/

10. Python: https://www.python.org/

11. Population Data: https://public.opendatasoft.com/

12. Weather Data & API Global Forecast & History Data: https://www.visualcrossing.com

Thanks!

Do you have any questions?