

Extracting Historic Flood Information from News Sources

Payal Chodha, Alicja Ligas, Beth
Padera, Dominika Twardowski



Problem Statement

Can we extract meaningful information about historic flood events in Punjab, India from online news sources?

Property damage, agricultural losses and human deaths caused by floods are calamities for developing cities. This project seeks to identify useful information relating to economic loss caused by floods in Punjab, India from online news sources.



Flood Impact

- **Floods were the most frequently occurring natural disaster between 2006-2015, accounting for 47% of all disasters**
- **Recurrent flooding has severe impacts on livelihood and health. Particularly rural regions dependent on farming**
- **Globally 2.4 billion people have been affected by floods between 1994 and 2013**

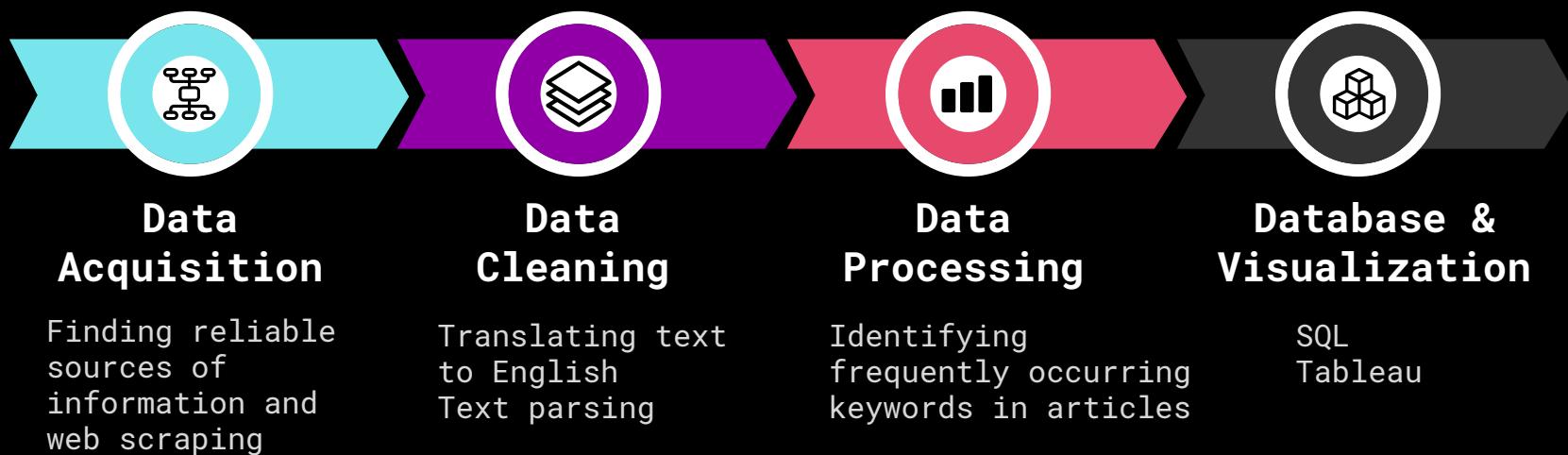


Punjab, India

- Mortality rates caused by floods have dropped in the last two decades, except for certain countries, one being India
- This August, an estimated 1.2 million people in India were displaced to government relief camps after widespread flooding



Process Overview



Data Sources:

1

News Sources

- ReliefWeb
- India Tribune
- Dainick Tribune
- Other regional sources

2

Weather

- Weather Company API

3

Economic

- Open Government Data India



Gathering Data

2008 - 2017



Web Scraping

- Sending requests via BeautifulSoup
- Collecting HTML code to extract text



Weather Company API

- Precipitation, runoff data
- Time Series Analysis



Open Source Data

- Open Government Data of India
- Economic Impacts and lives lost



Data Cleaning

01 | Translating text to English via TextBlob



02 | Checking for missing values and duplicate rows



03 | Replacing line breaks with spaces and parsing sentences using regex
Converting text to lowercase

04 | Categorized sentences using keywords and numeric values



Data Analysis



Lives Affected

Human lives affected and the damages associated



Death

Overview of lives lost due to floods



Agriculture

Impact on agriculture, livestock and crops



Infrastructure

Structural devastation

Interactive Tools

+tableau public

Impact of Floods in Punjab, India 2008 -2017

Population Affected Over Flood Dates

Human Lives Lost

Population Affected

3 397,294

Precip (cm)

16.06

Punjab

Crop Value (1000 R)

179,879

Punjab

Flood Precipitation & Runoff (cm)

Crops & Cattle

News Sources - Important Sentences

Date Location Sentence Topic Source

008 Ferozepur subrahmanyam), 7, race course road, new d.. house <https://reliefweb.int/node/27016>

2008 Ferozepur a woman along with her son died in house death <https://reliefweb.int/node/27016>

collapse last night in jalandhar while three .. house <https://reliefweb.int/node/27016>

a woman drowned and five houses were death <https://reliefweb.int/node/27016>

damaged in doda and reasi districts of house <https://reliefweb.int/node/27016>

jammu and kashmir following heavy rains a.. lives <https://reliefweb.int/node/27016>

india: assam reels under floods, orissa agri <https://reliefweb.int/node/27016>

sounds alert new delhi: as people, particula.. lives <https://reliefweb.int/node/27016>

six persons were killed in separate incident.. death <https://reliefweb.int/node/27016>

the official death count was four in the state death <https://reliefweb.int/node/27016>

- 07 relief camps have been opened in the fl.. lives <https://reliefweb.int/node/27589>

- 199 medical teams continued to be deploy.. lives <https://reliefweb.int/node/27589>

Keep Only Exclude

Location: Ferozepur

Flood Dates: 6/11/2008 to 8/13/2008

Population Affected: 50,000

Human Lives Lost: 4

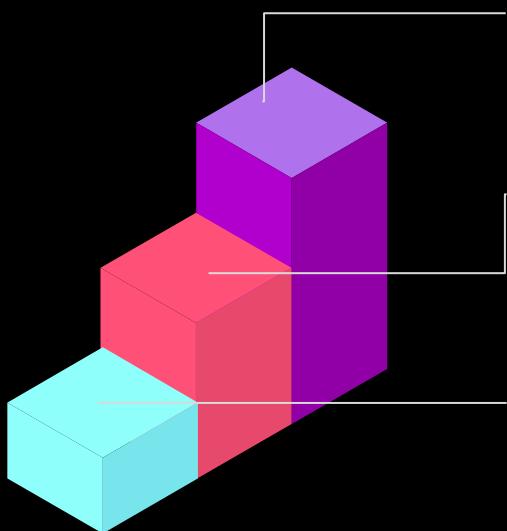
Houses Damaged Value (1,000R): 13,474

Houses Damaged (number): 1,678

© 2019 Mapbox © OpenStreetMap



Interactive Tools



Front End

HTML tables and
Javascript

Flask

Web Application
Framework

SQLite

Database Engine

<http://127.0.0.1:5000/>



Conclusions

Challenges

- Freely available data
- Unevenly distributed data

Next Steps

- Growing database through other news sources, social media and paid APIs
- Expanding beyond Punjab

App Development

- UX Design
- Software updates

The background of the slide features a vibrant, abstract pattern of ink or paint swirling in water against a black background. The colors include deep purple, bright blue, magenta, pink, and yellow, creating a dynamic and organic feel.

Thank you!

Questions?