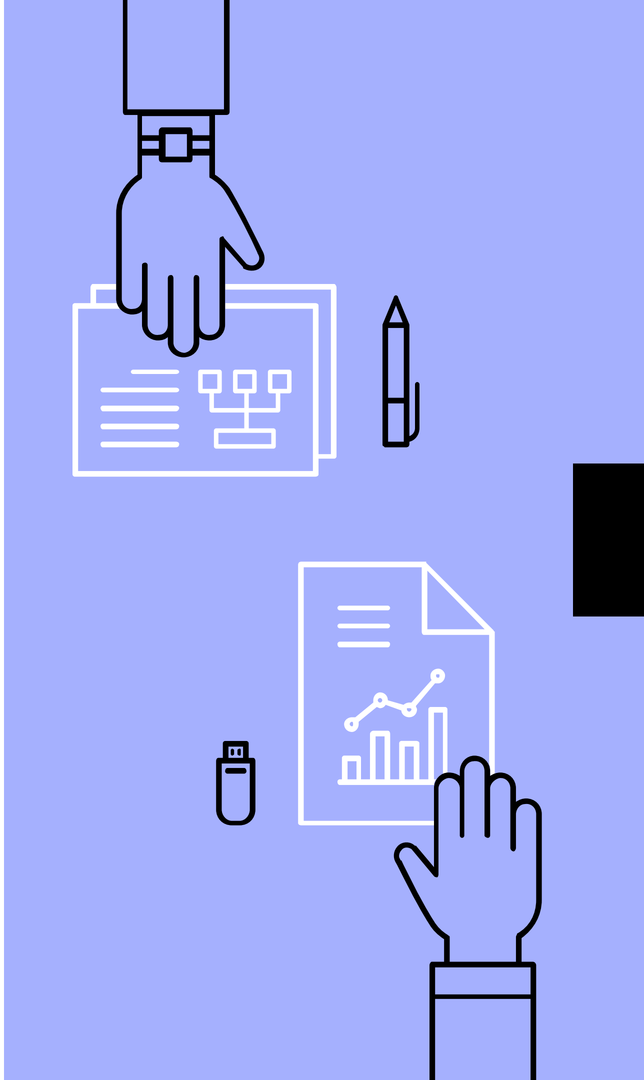
Maharashtra Rainfall Forecast

Overview

- Introduction
- Problem Statement
- Literature
- Objective
- Previous Work Done
- Data
- Methodology
- Website
- Plan on Improvement
- References

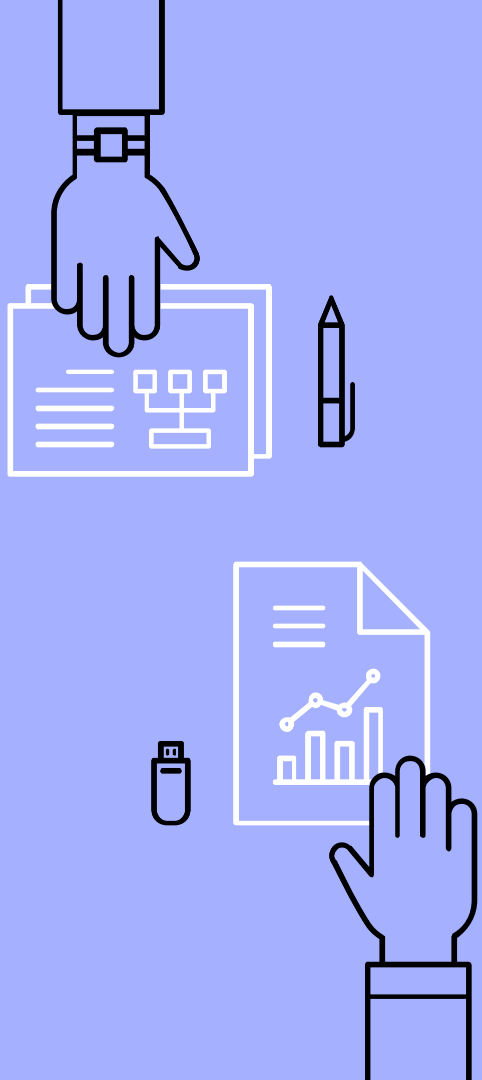


Introduction

- Rainfall forecasting is still a concern for researchers considering the increase in uncertainty of weather conditions.

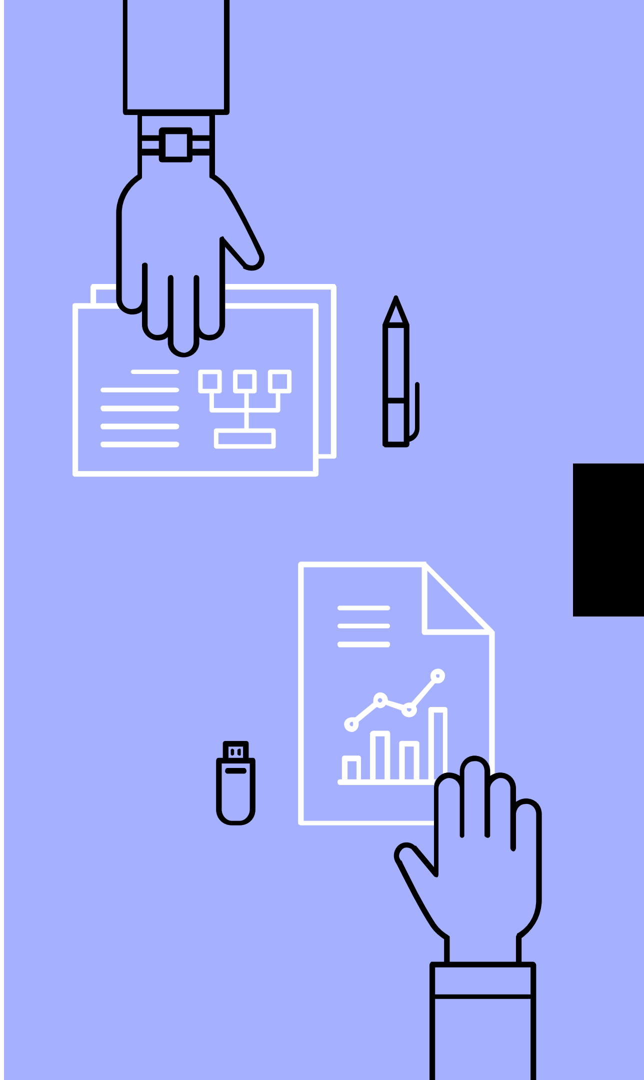
- As the climate models are unreliable, we need a further improvement on the prediction models.

- Simulated probabilities are said to be reliable if they truly reflect the observed frequencies (within uncertainties).



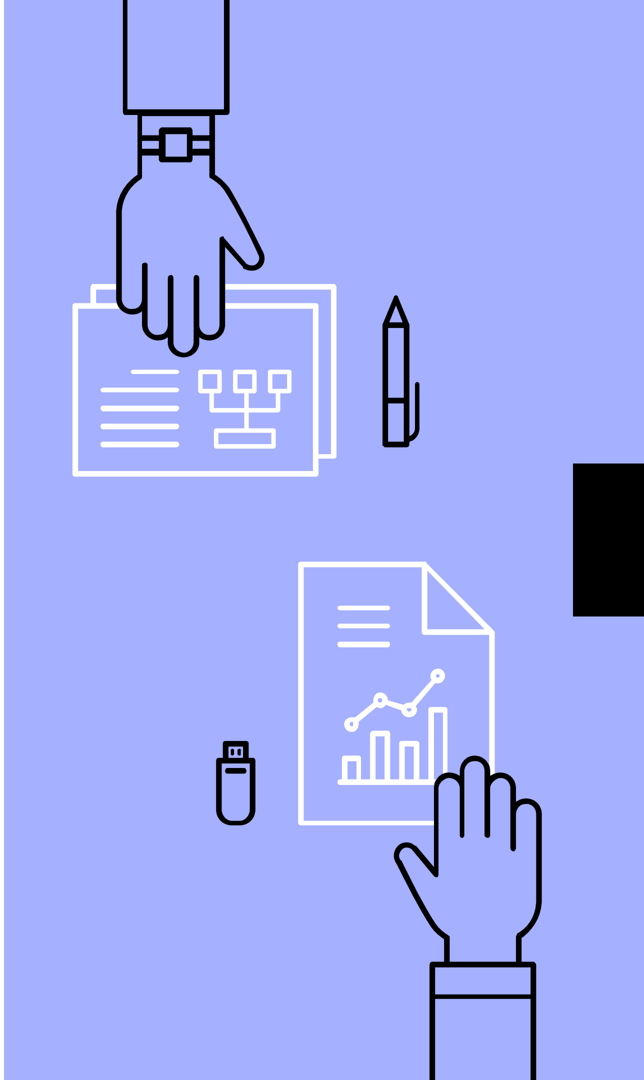
Introduction(Contd.)

- In recent days, Deep Learning enabled the self-learning data labels which allows to create a data-driven model for a time series dataset.
- As the climate models are unreliable, we need a further improvement on the prediction models.
- Simulated probabilities are said to be reliable if they truly reflect the observed frequencies (within uncertainties).



Problem Statement

- The main aim of our project is to build a deep learning model which performs with higher accuracy in rainfall forecasting and predicting future events' data with respect to the events occurred in the past.



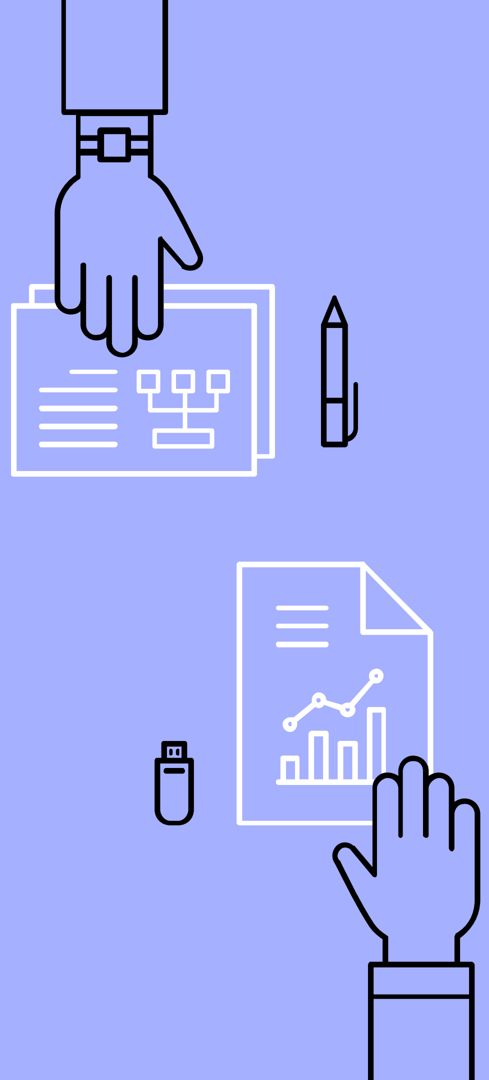
Literature

- The paper introduces different approaches for designing various model, both statistical and deep learning, to forecast rainfall.

  Yoo, Ji-Young & Kwon, Hyun-Han & Kim, Tae-Woong & Ahn, Jae-Hyun. (2012). Drought frequency analysis using cluster analysis and bivariate probability distribution. Journal of Hydrology. 420. 102-111. 10.1016/j.jhydrol.2011.11.046
  Link

- The paper uses statistical method of drought analysis using threshold rainfall value for a month to decide the occurrence of a drought event .

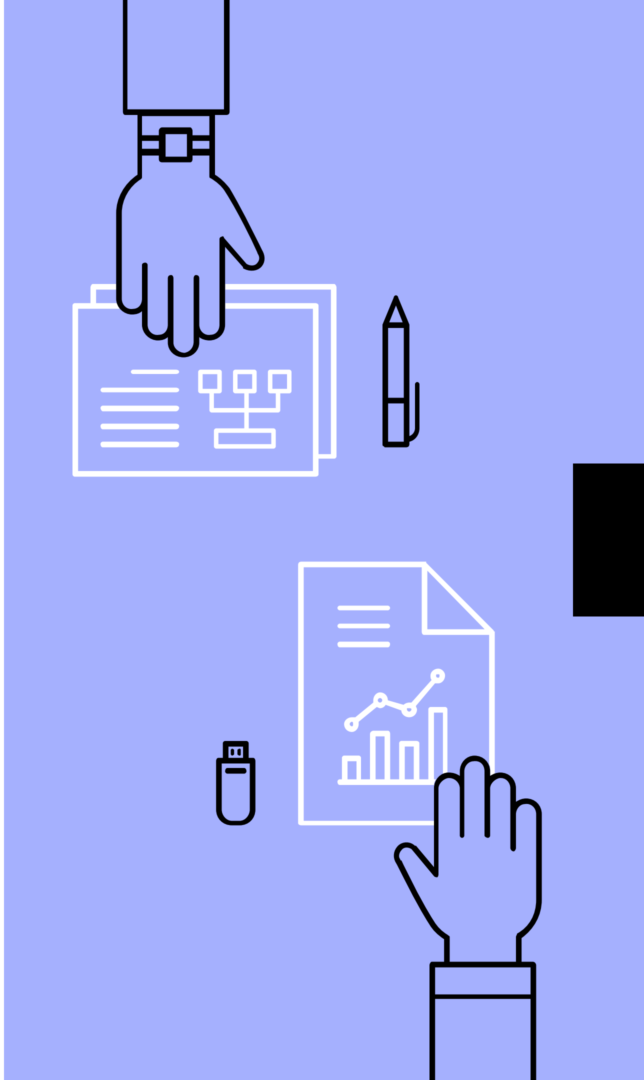
  Dhawal Hirani , Nitin Mishra ,"A Survey On Rainfall Prediction Techniques", International Journal of Computer Application (2250-1797). Link



Literature(Contd.)

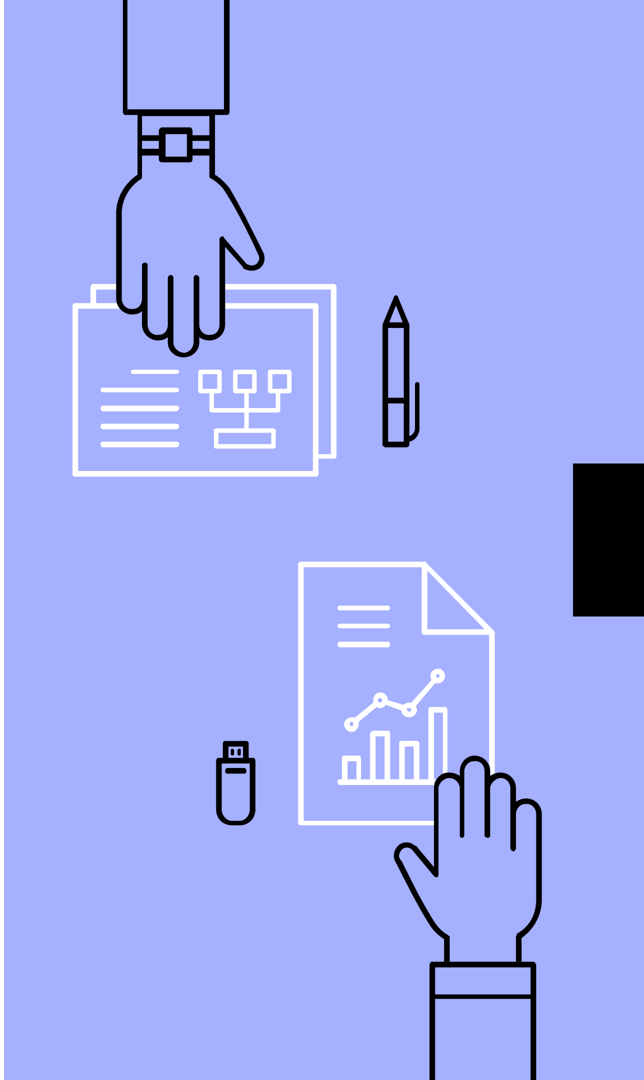
- The paper talks about density estimation for functions of correlated random variables.

  Jeffrey P. Kharoufeh,Density Estimation For Functions Of Correlated Random Variables,1997,PP 26-30. [Link]



Previous Work Done

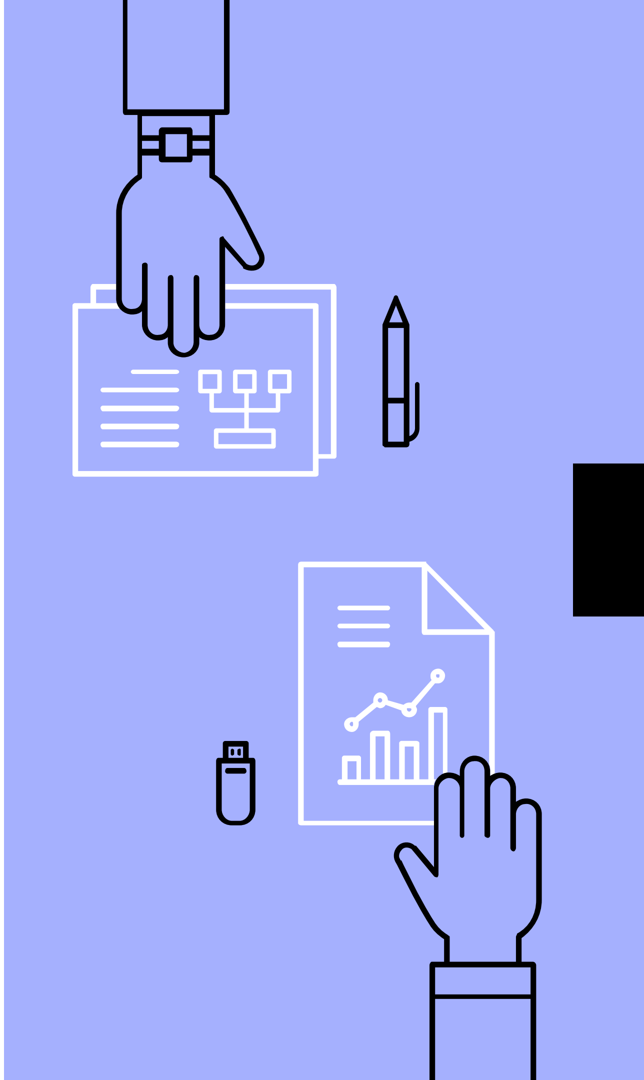
- Data Collection
- Analysis over the data(Seasonal,Monthly,Yearly).
- Analysis over each variable(Temperature,Soil Moisture,Rainfall).
- Website Front-End(Implemented map with filters).
- Building different models and finding the best one.



Data

- ➢ We have the data of 123 coordinates of Maharashtra.
- ➢ Our dataset contains time series data of rainfall(in mm), relative humidity and surface temperature(in Kelvin) with temporal resolution of 1 day for the coordinates in a time period of 12 years(2008-2020).
- ➢ Total number of days in the series is 4116.

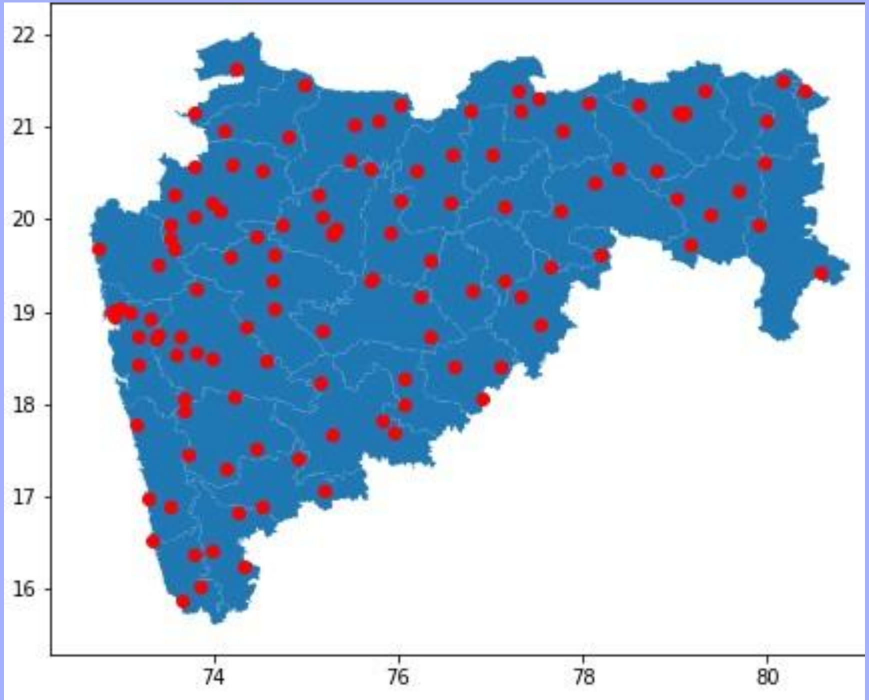
https://www.mosdac.gov.in/



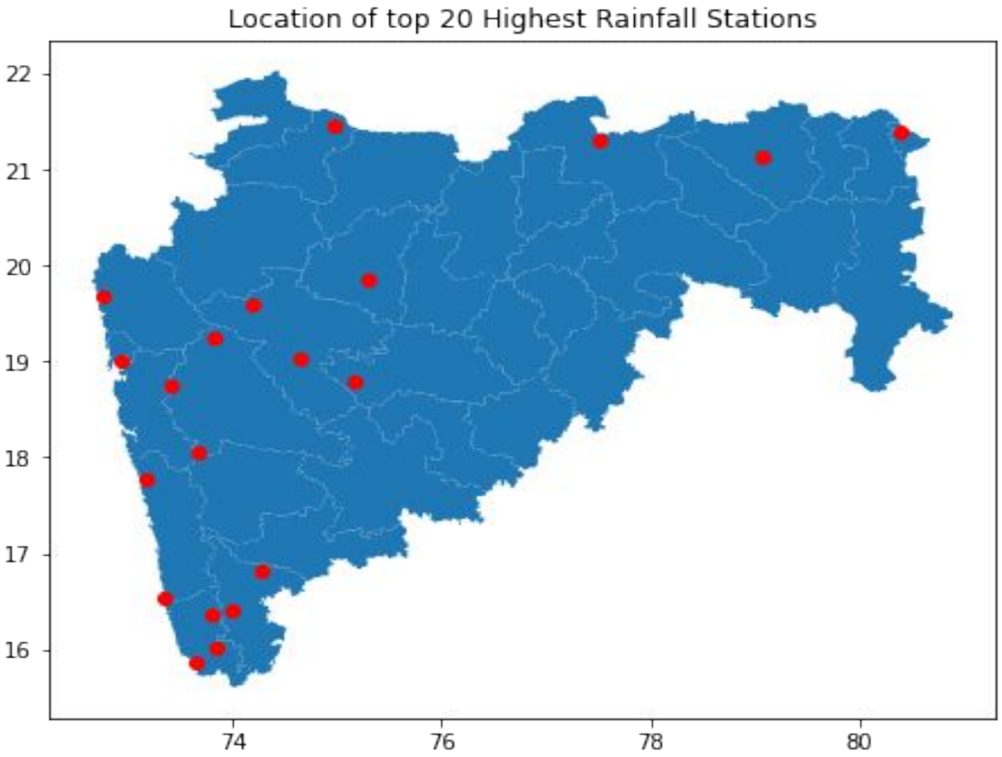
Data(Contd..)

- For handling the null values, we used the average of temporal interpolation and IDW(Inverse Distance Weighted).

- As the dates were inconsistent in between, we fixed most common occurring dates, 2008-08-24 to 2019-12-31.

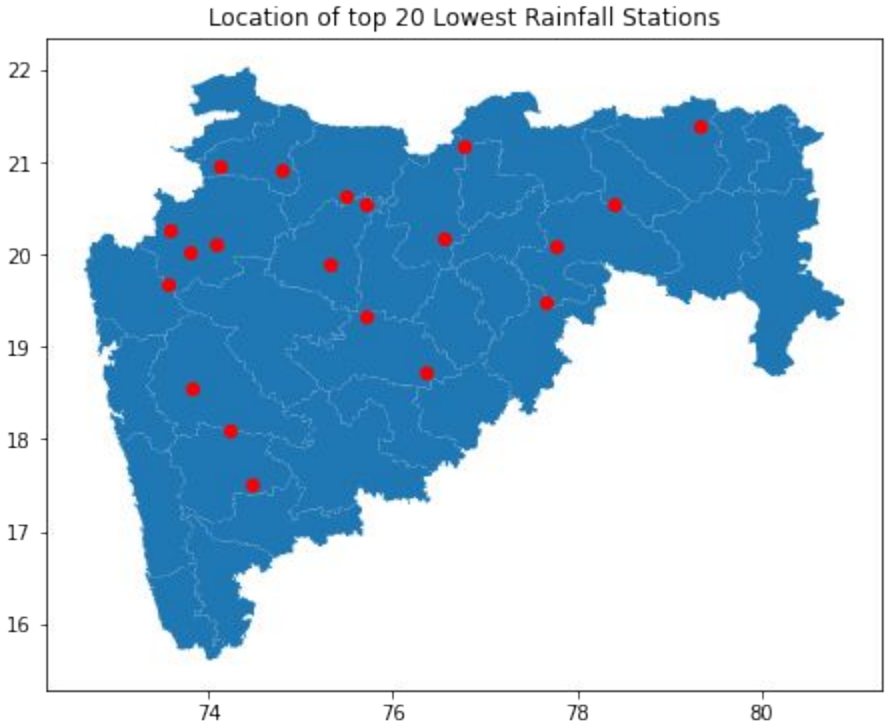


Data Analysis



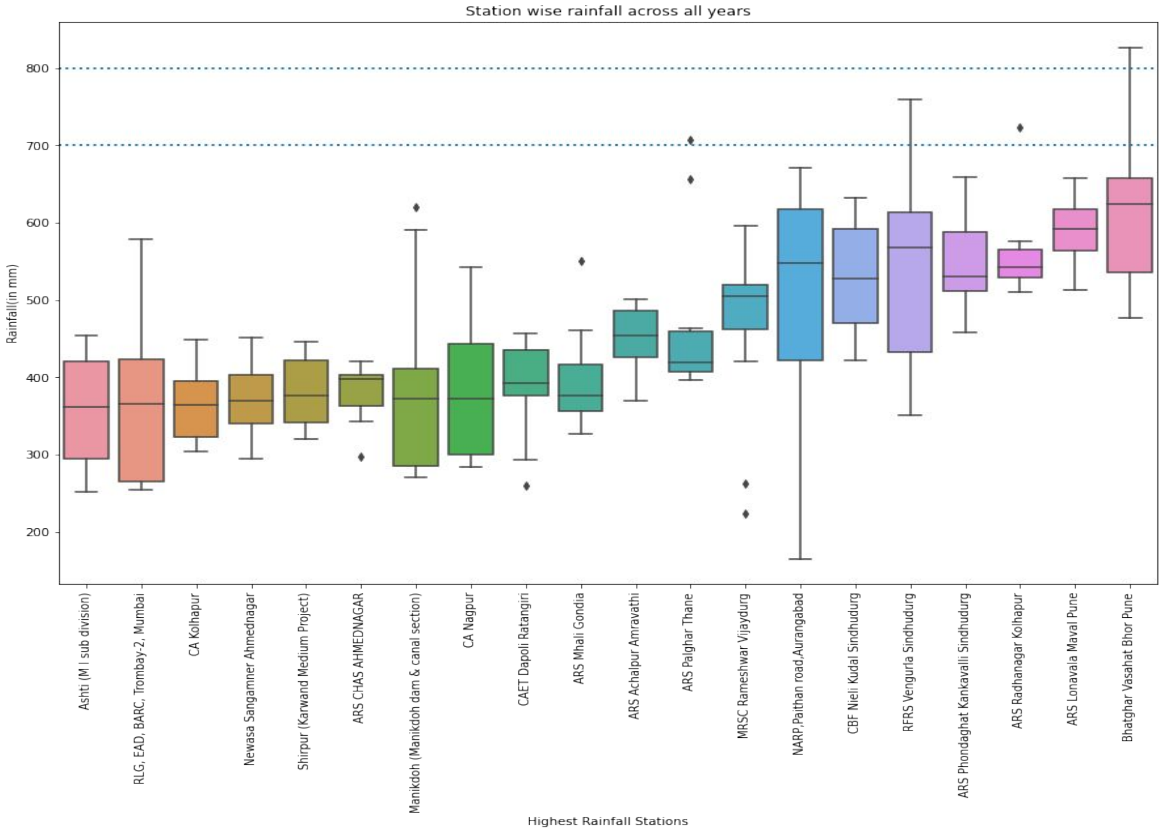
- The figure shows the top 20 highest rainfall stations.

Data Analysis



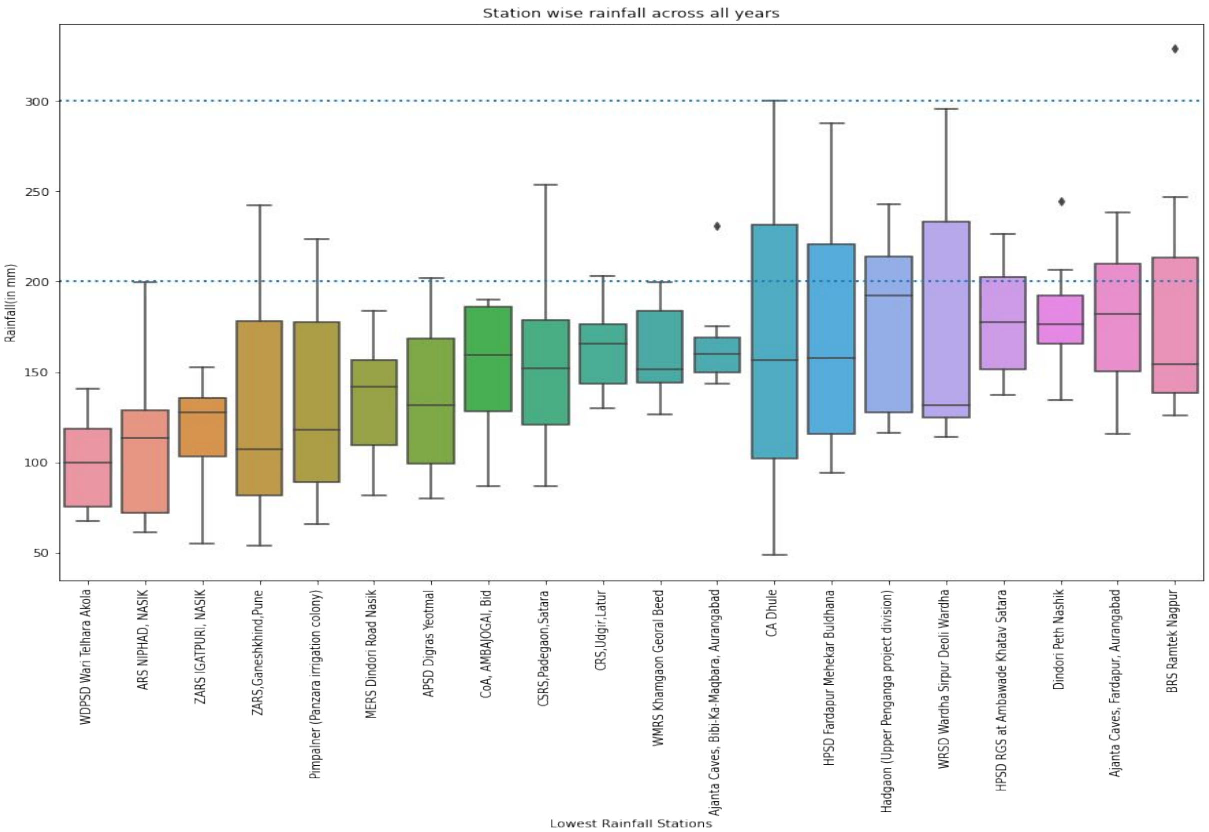
➢ The figure shows the top 20 lowest rainfall stations.

Data Analysis



- The figure shows the top 20 highest rainfall stations.

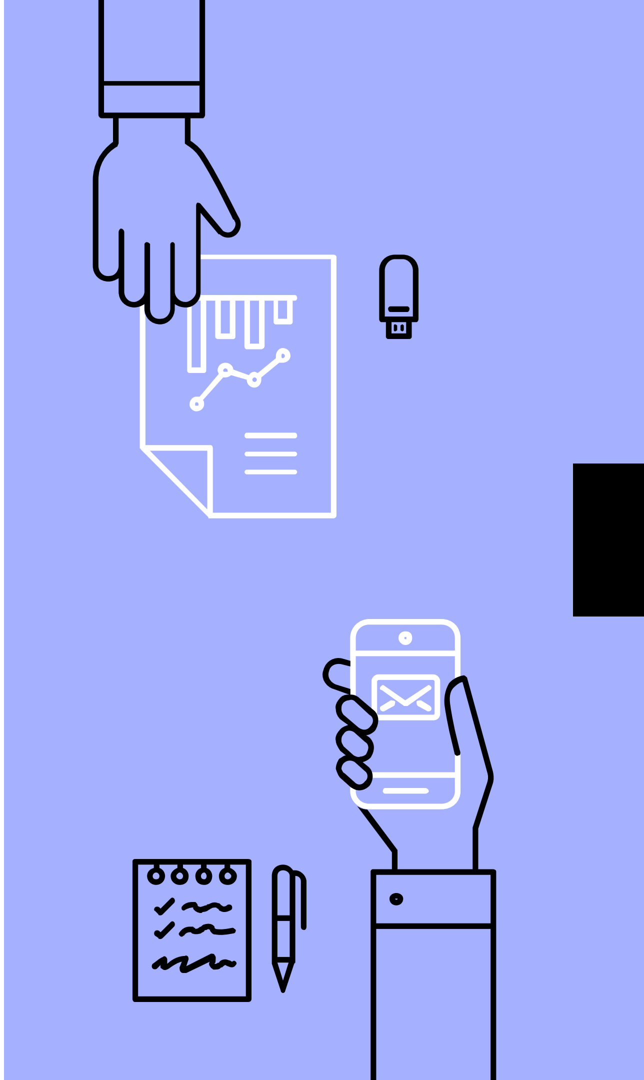
Data Analysis



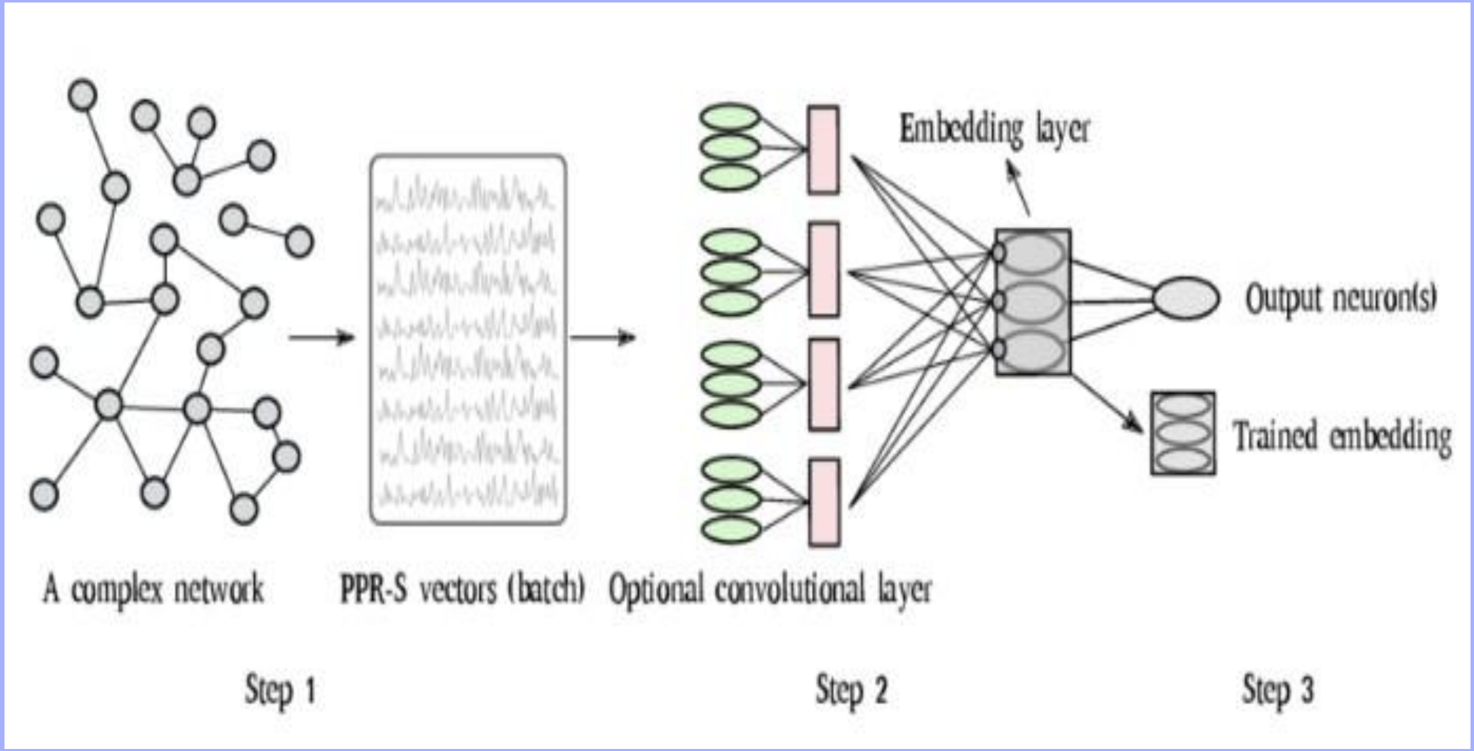
- ➢ The figure shows the top 20 lowest rainfall stations.

Methodology

➢ **Graph neural networks** (GNNs) are connectionist models that capture the dependence of **graphs** via message passing between the nodes of **graphs**. Unlike standard **neural networks**, **graph neural networks** retain a state that can represent information from its neighborhood with arbitrary depth

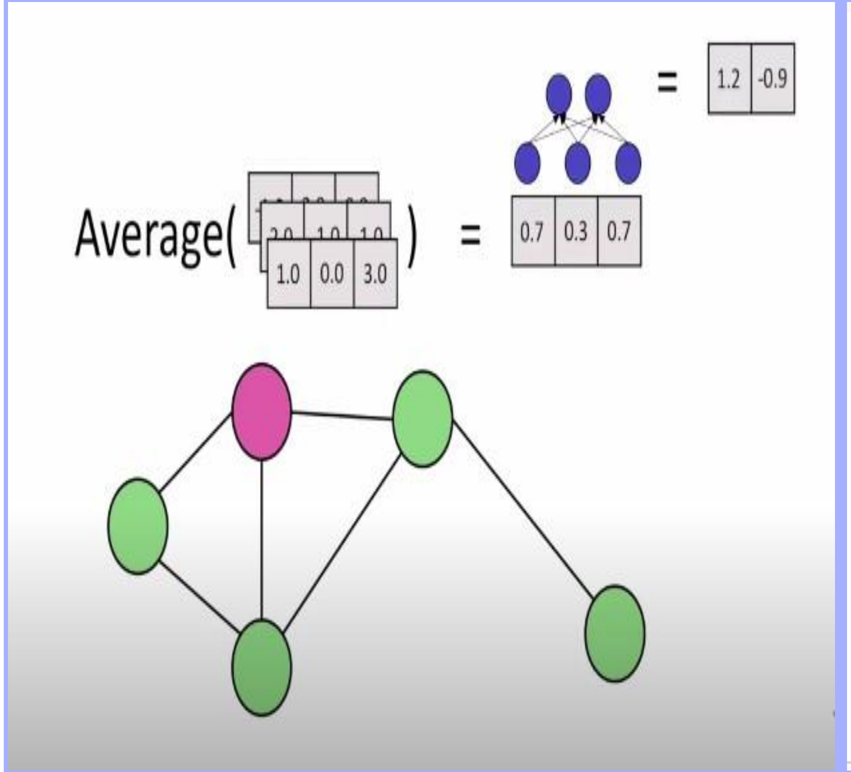


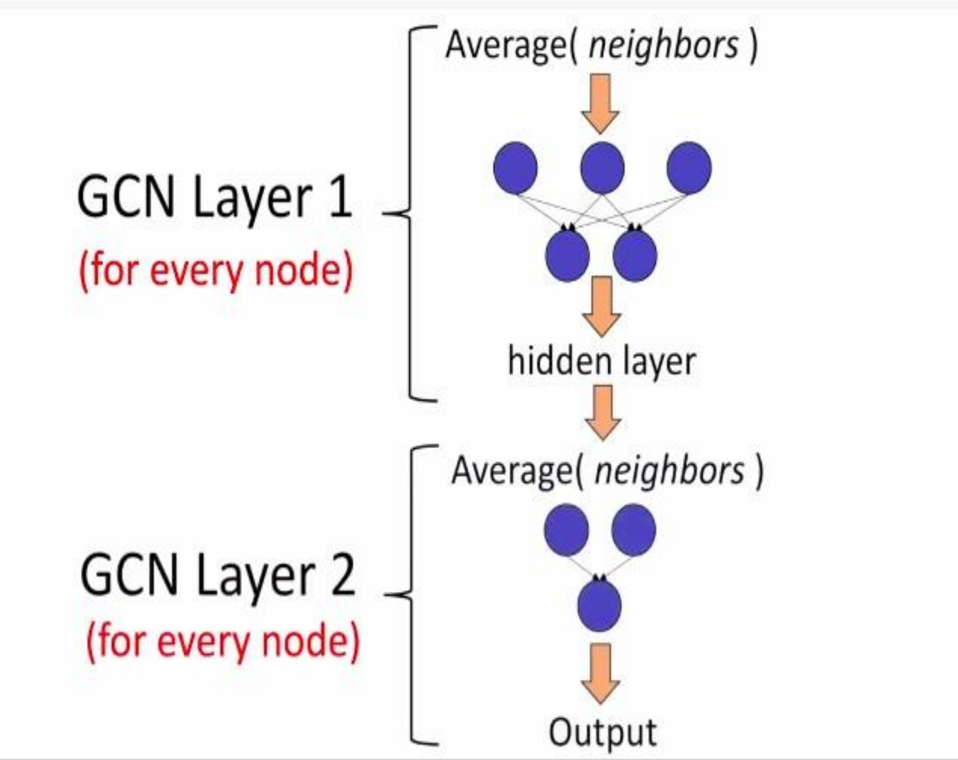
Methodology



Source

Methodology

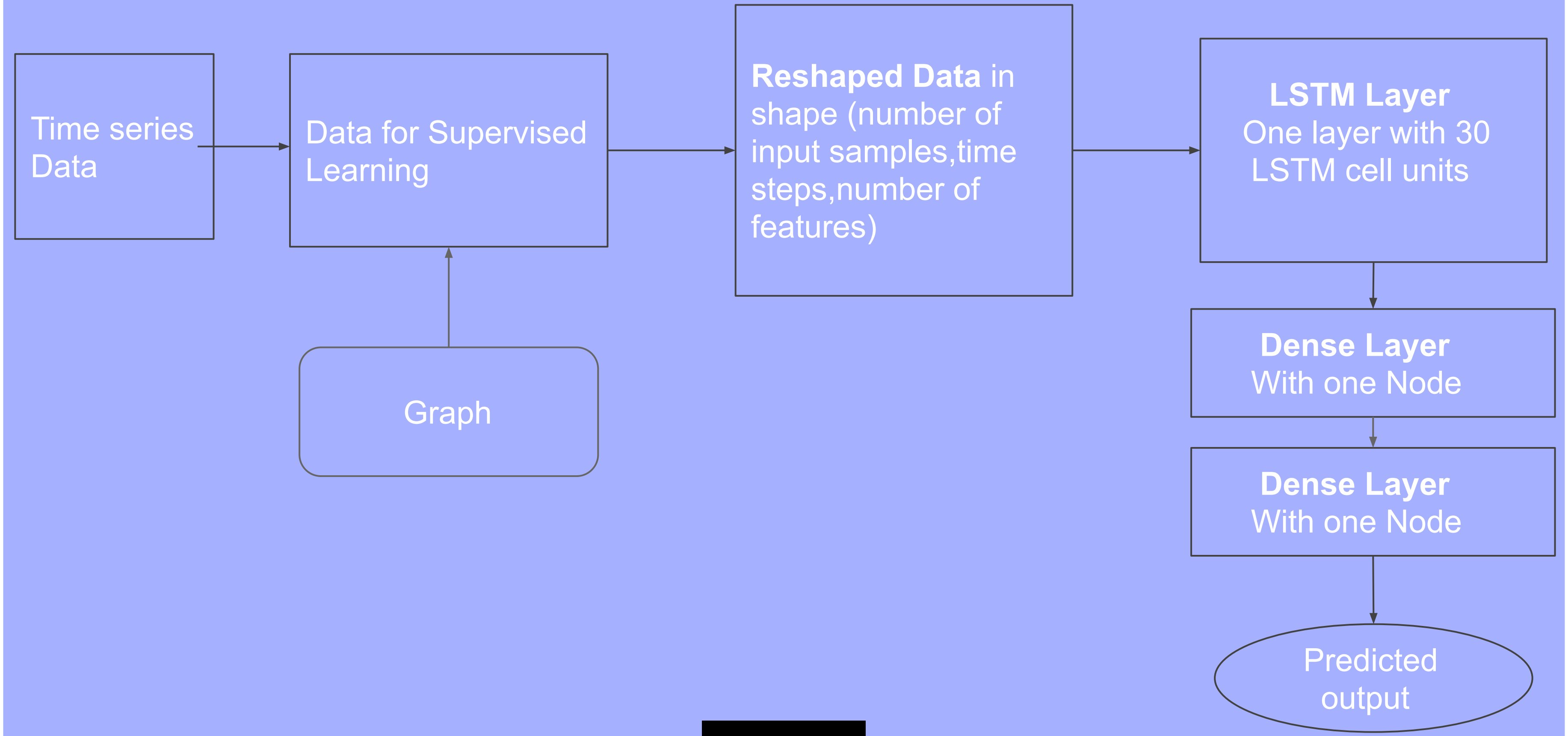




Reference

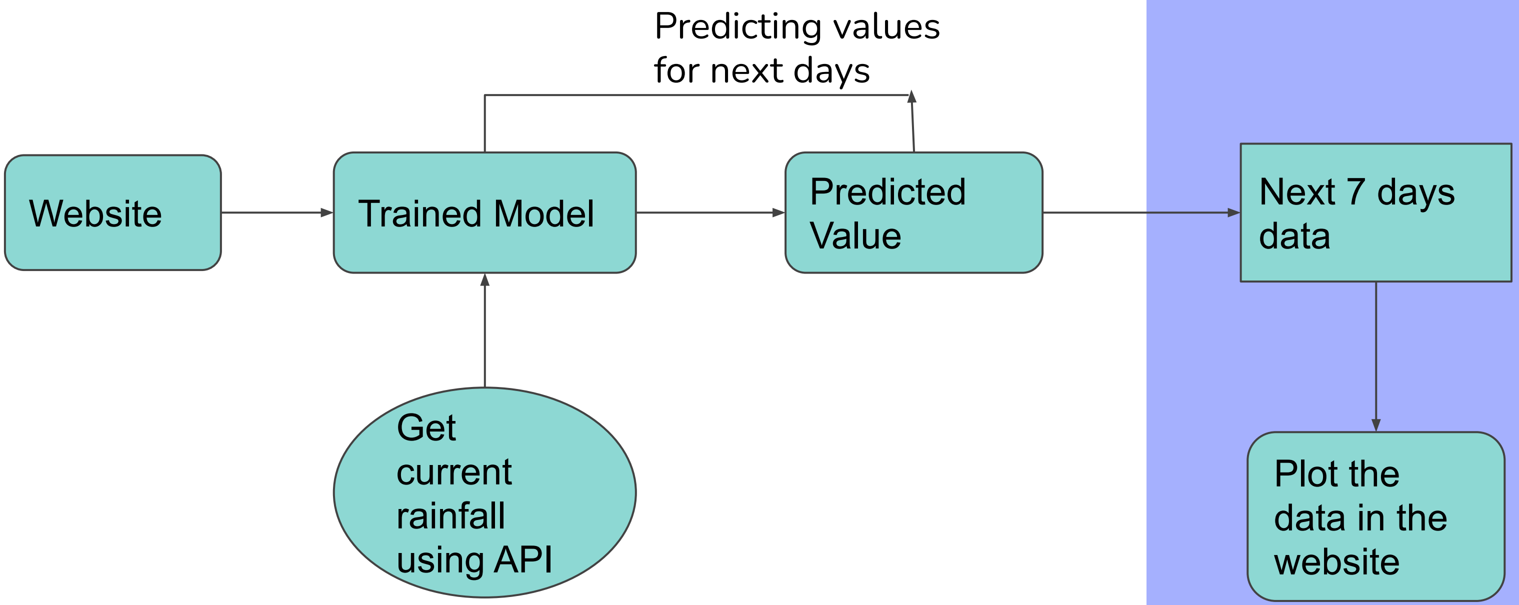
[redacted]

Model Architecture



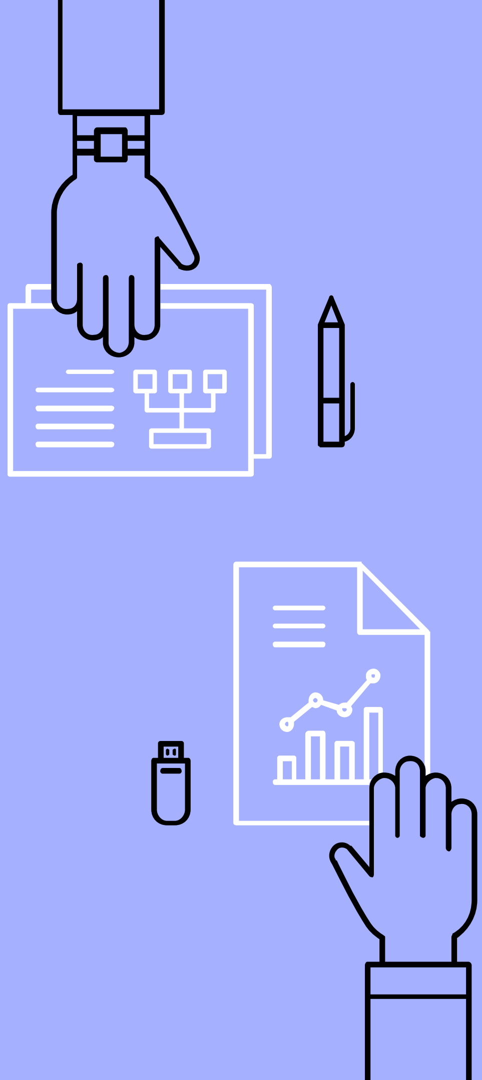
[redacted]

Workflow


Predicting values
for next days
Website
Trained Model
Predicted
Value
Next 7 days
data
Get
current
rainfall
using API
Plot the
data in the
website

Probability Estimation

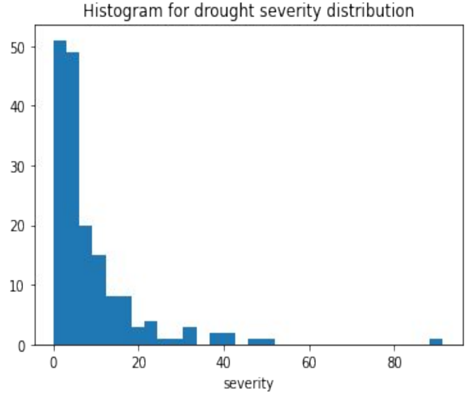
- We did extreme rainfall events analysis of all the stations.
- Here we are demonstrating the analysis for Chandrapur as it experiences both drought and flood.
- Severity calculation of an extreme event:
    - Threshold rainfall for extreme event occurrence for a month is calculated as the mean of average monthly rainfalls in the particular month for all previous years.
    - An extreme rainfall event is said to have occurred if an area receives monthly average rainfall greater or less than threshold rainfall.
    - Severity value for a month is the difference between threshold value and actual rainfall for the month.
    - Event duration is the duration for which the extreme event occurs.

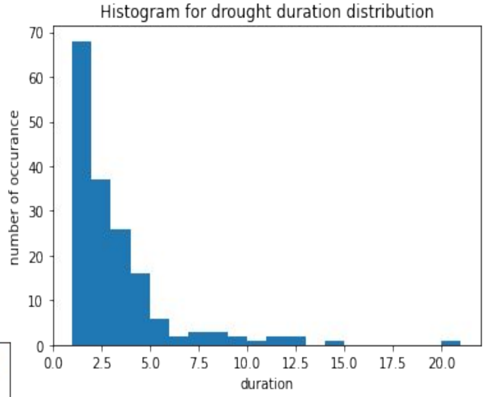


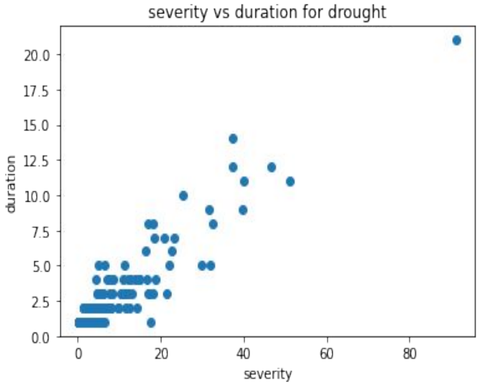
Probability Estimation(contd...)

- Joint Kernel Density Estimation for bivariate distribution(duration,severity) for drought and flood:
    - We have used non-parametric kernel density estimation because:
        - The joint distribution did not resemble any standard bivariate distribution.
    - Severity is considered in mm and durations are considered in months.

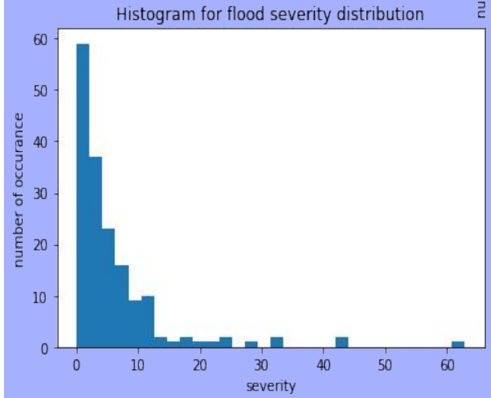
Probability Estimation(drought)

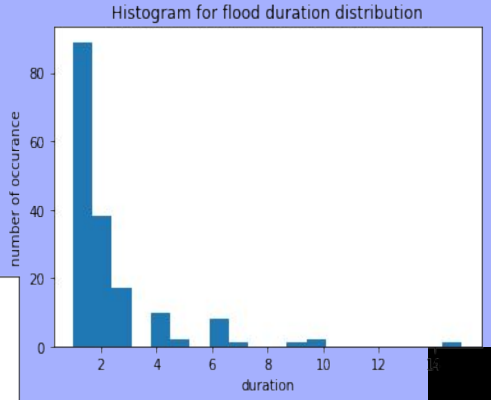


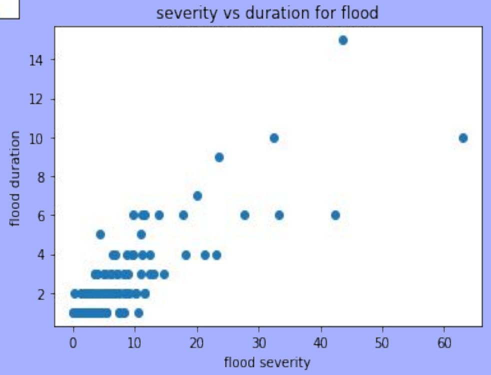




Probability Estimation(flood)







Probability Estimation(contd...)

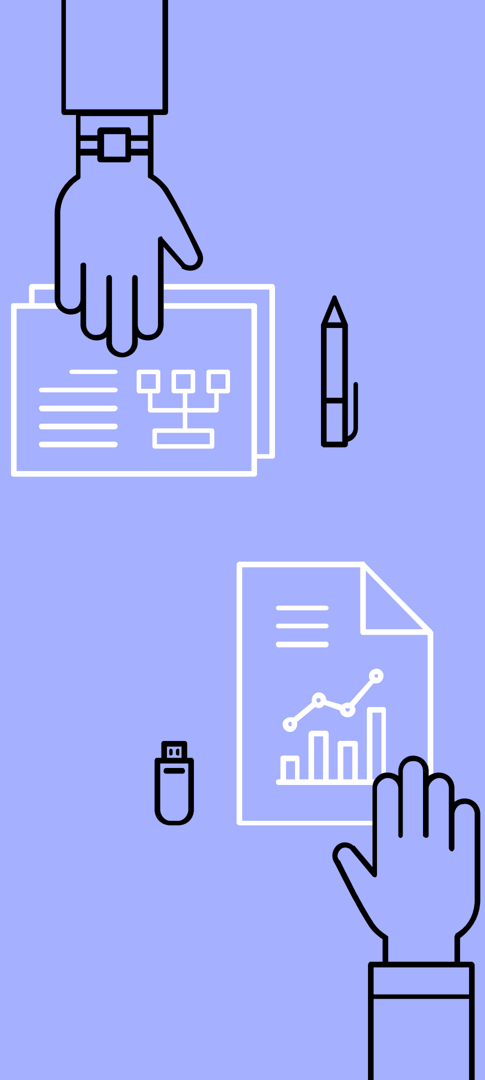
- Joint Kernel Density Estimation for bivariate distribution (duration,severity):
  - Our random variables are dependent. So we needed to use multivariate kernel density estimation unlike in the case of independent random variables where joint probability distribution could be given by the product of univariate density estimations.
  - The correlation coefficient between severity and duration was higher than the upper limit for the application of parametric distributions.
    - For Drought:

      Correlation coefficient = 0.81320855

      P-value = 2.4029×10^{-41}
    - For Flood:

      Correlation coefficient = 0.7468217

      P-value = 2.1309×10^{-31}

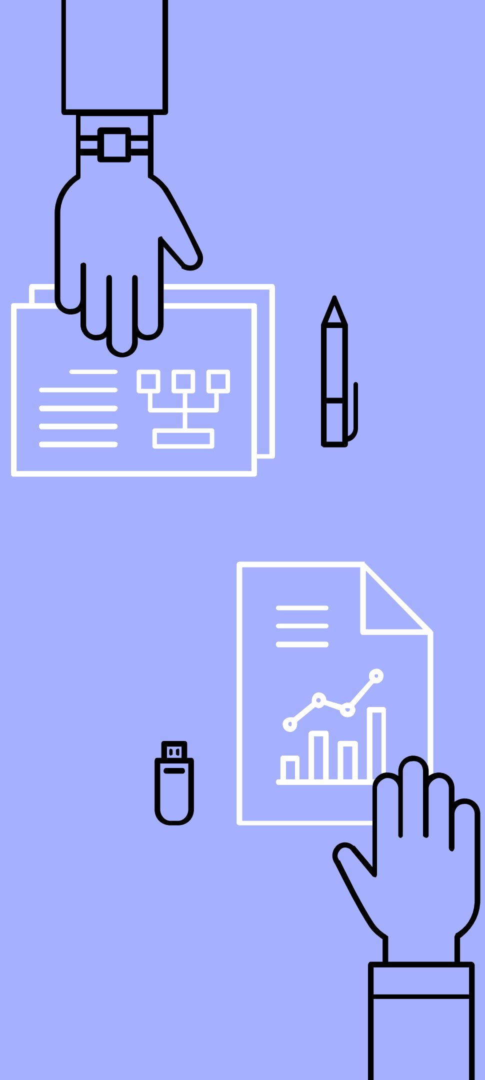


Probability Estimation(contd...)

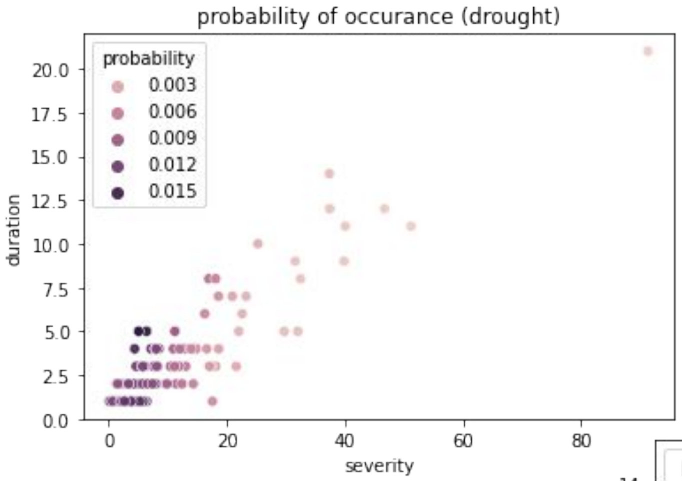
- Probability estimation of predicted severity values:
    - The joint probability distribution function for the bivariate distribution using non-parametric method is given by:

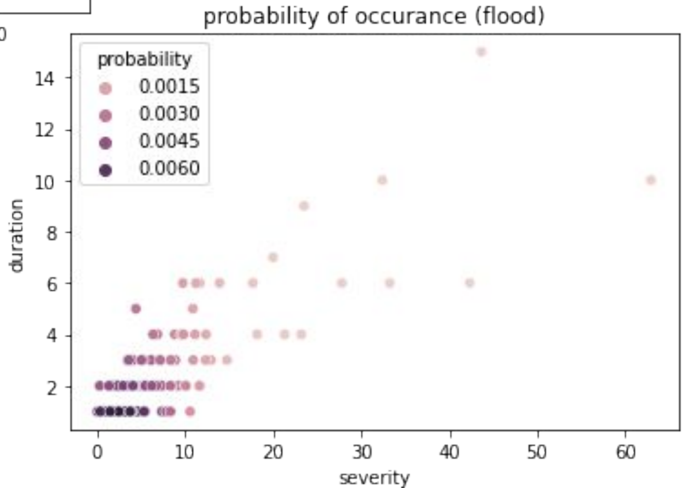
$$K_h(X_i, X_j) = \prod_{s=1}^{q} h_s^{-1} k\left(\frac{X_{is} - X_{js}}{h_s}\right)$$

    - Here $\boldsymbol{X_i}$ is event duration, $\boldsymbol{X_j}$ is event severity.
    - $\boldsymbol{k}$ is the kernel function and $\boldsymbol{h}$ is the bandwidth matrix.
    - Bandwidth matrix controls the amount and orientation of smoothing induced while estimating the kernel density.



Probability Estimation(contd...)





- Integrated mean squared error obtained with the best bandwidth matrix for the estimation of *pdf* for drought and flood distribution were 5.203664 and 0.0078951 respectively.

Risk Estimation

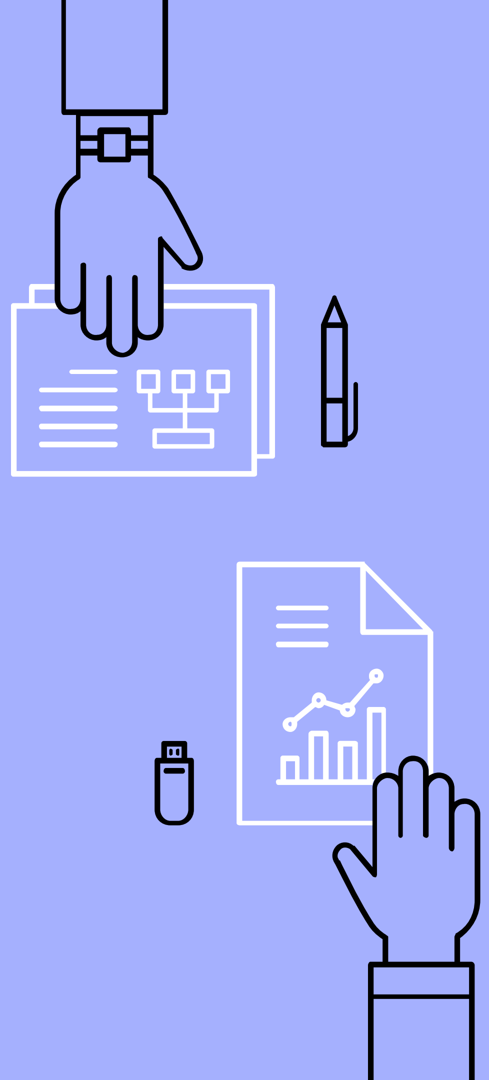
- The risk of an event associated with the bivariate distribution is given by:

  $$R = 1 - (1 - 1/T)^N$$

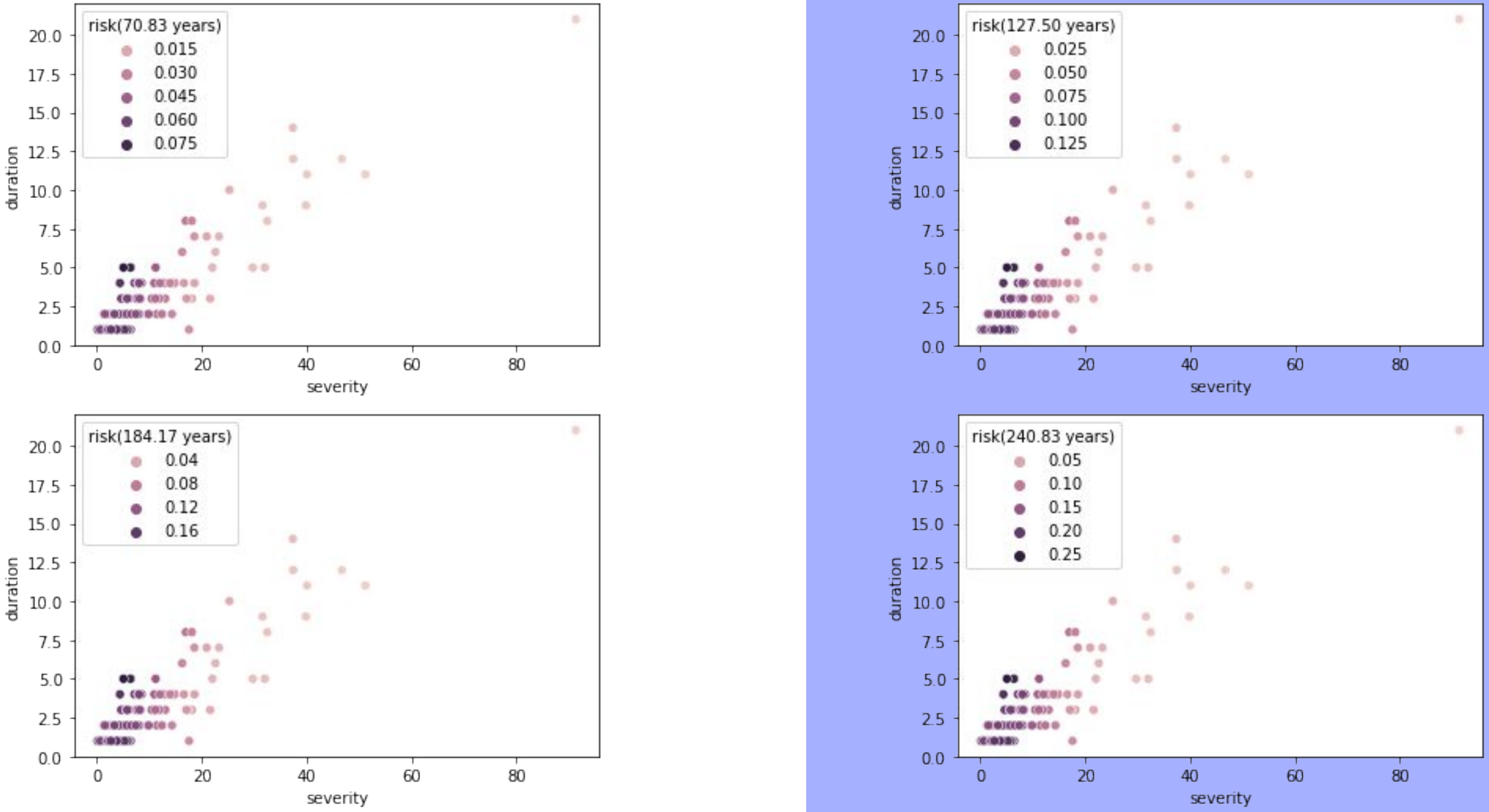
  Here T is the return period of a severity value and the duration associated with it.

  N is the time span for which the risk is being calculated.

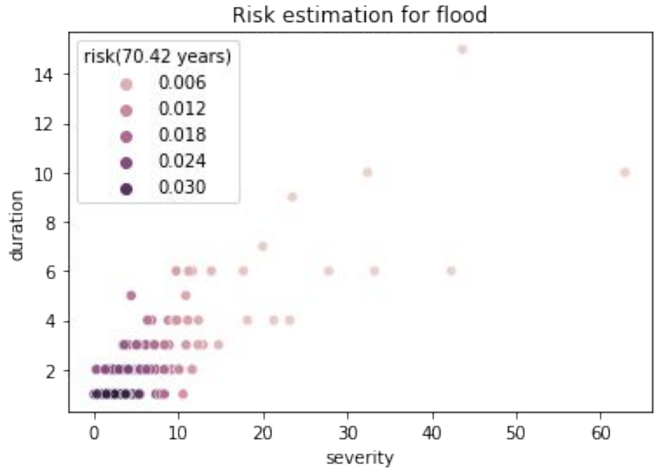
- Return period:
  - The estimated average time between two similar events.

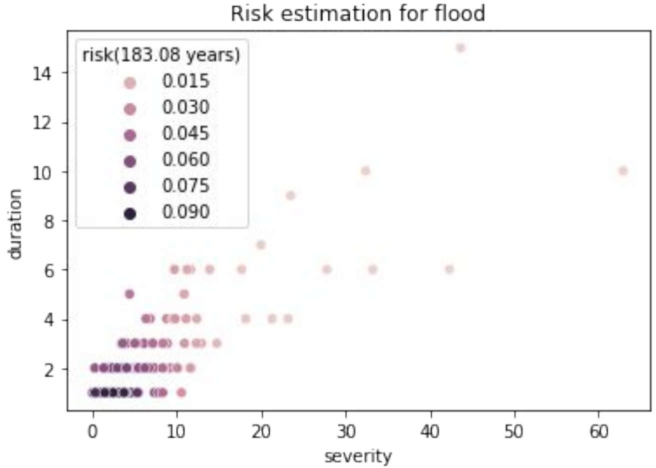


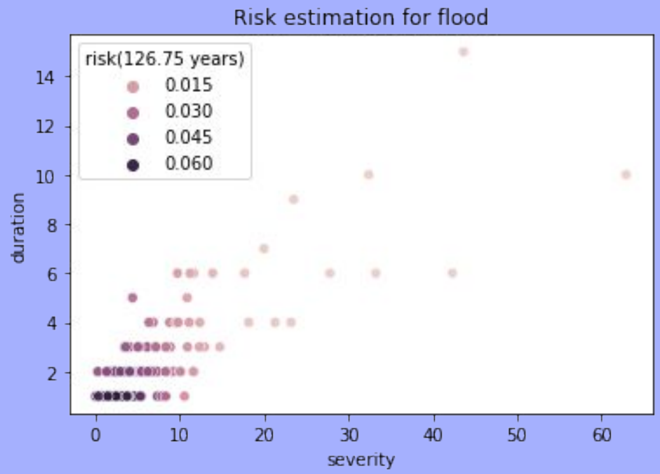
Risk Estimation (drought)

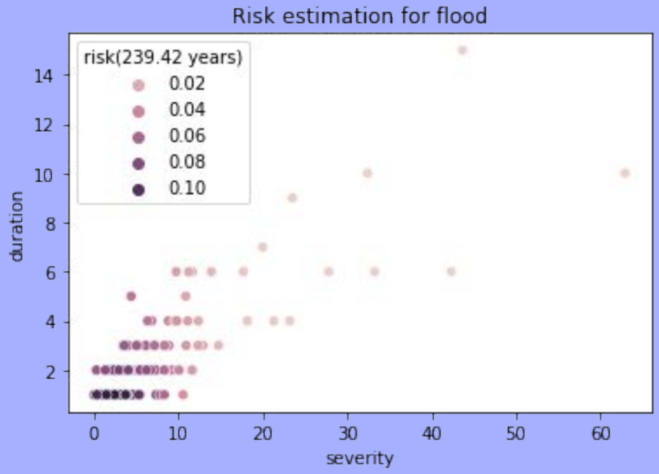


Risk Estimation (flood)

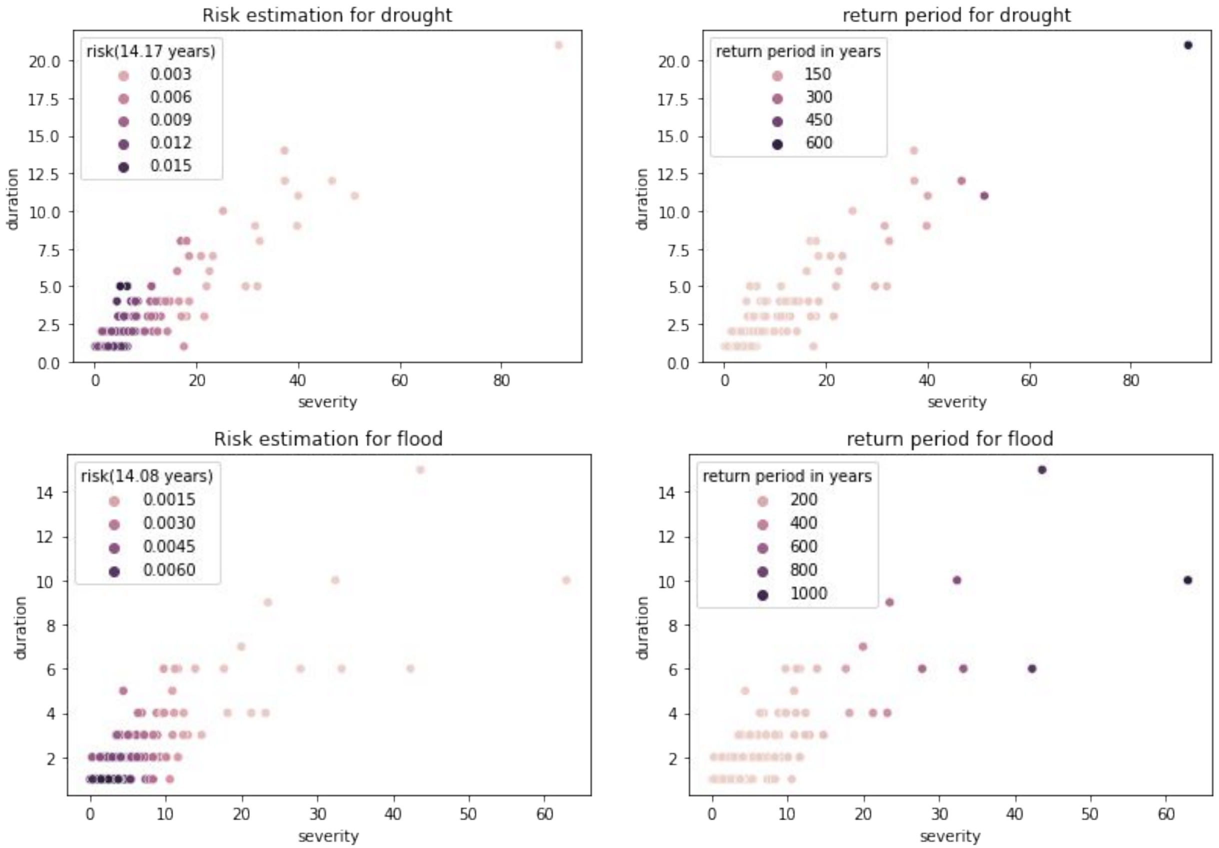








Risk estimation



Risk shows high negative correlation with return period for both flood and drought *i.e.* lower the return period, higher will be the risk associated with the corresponding severity and duration value.

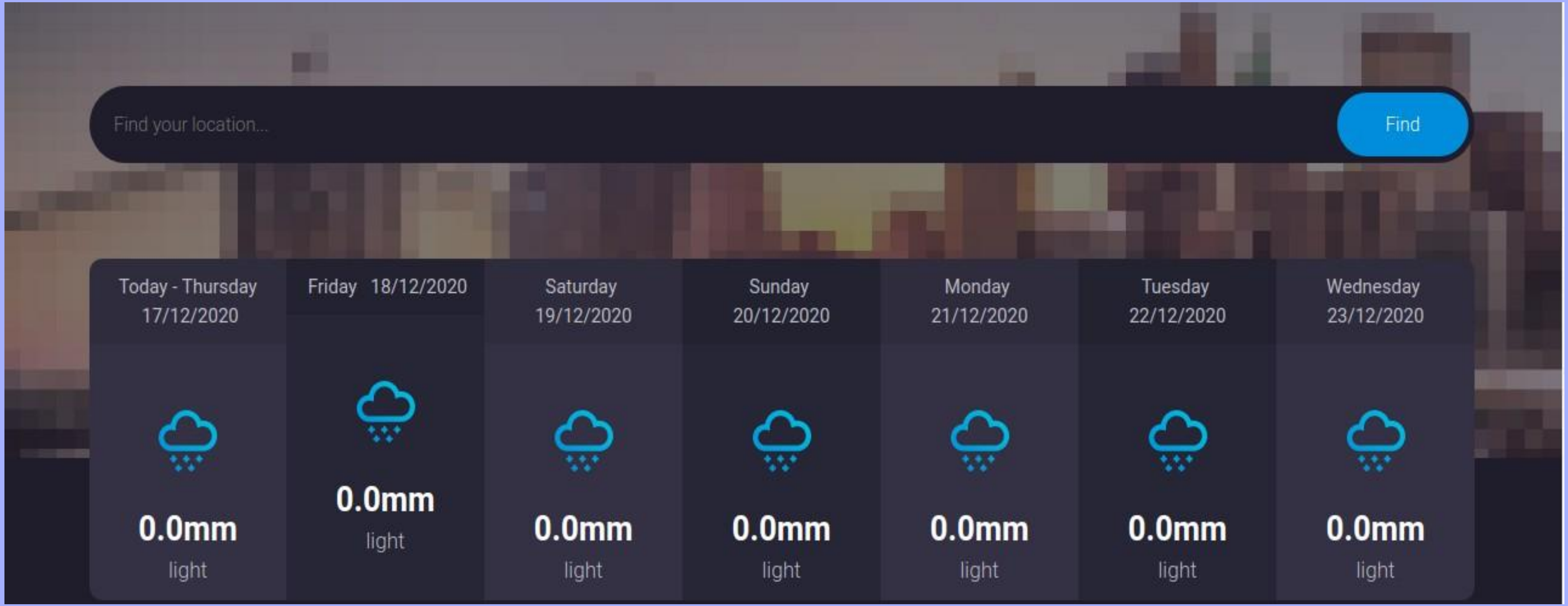
Working of Website

- We've built a website in which we can show the rainfall forecast values predicted by the model for the coming six days of a particular place .
- We've also implemented map using javascript library called Leaflet.js
- As our project is concerned with maharashtra we displayed Maharashtra region highlighted.
- In the Map we implemented a search function which zooms into the place searched for.

Working of Website(Contd..)

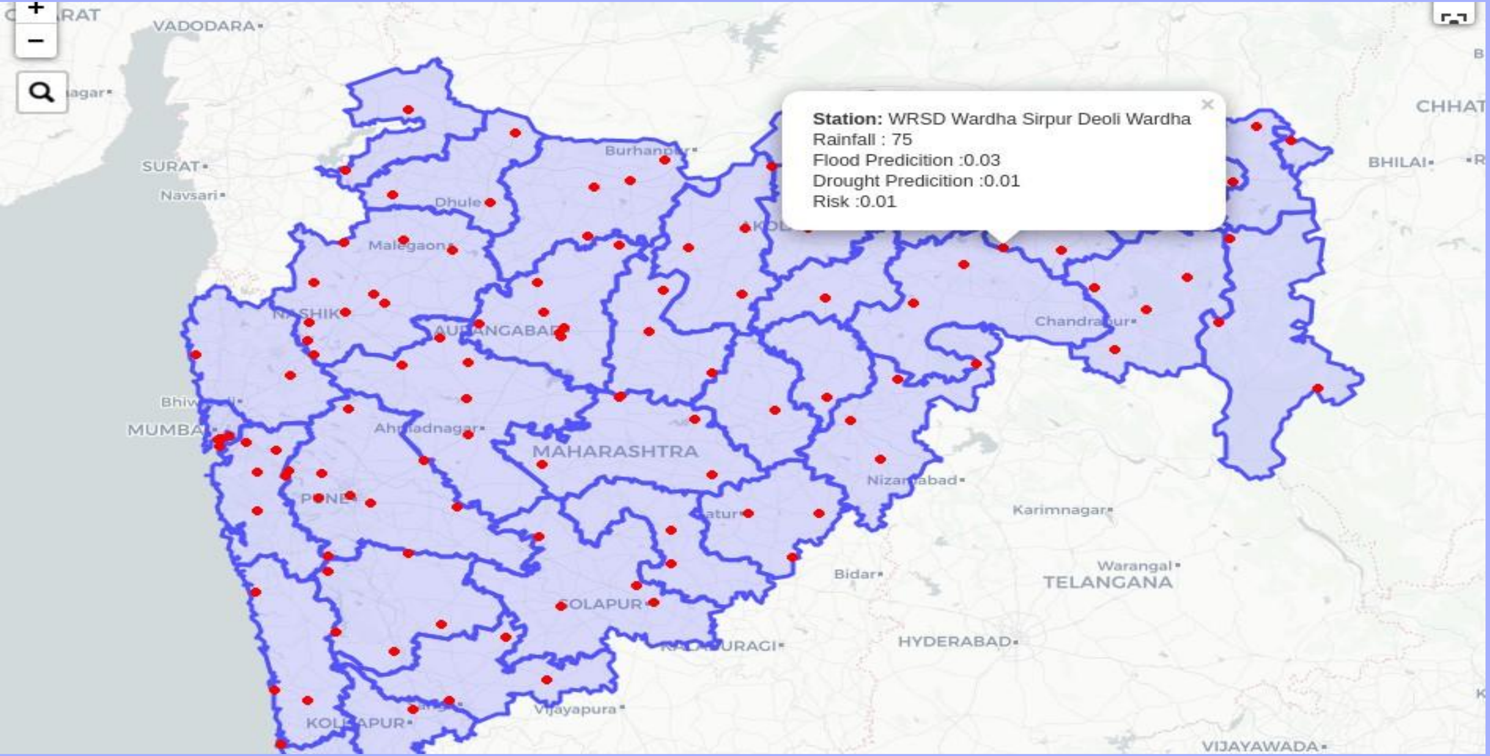
- We've integrated the model trained into website.
- We used weather API to get the current rainfall data at all the 123 locations.
- place and use the model to predict next days data.
- We keep on appending the predicted value to the current data list.
- We keep iterating over predicted list till we get the next 7 days forecast.
- We show the next 7 days forecast in the website

Working of Website(Contd..)

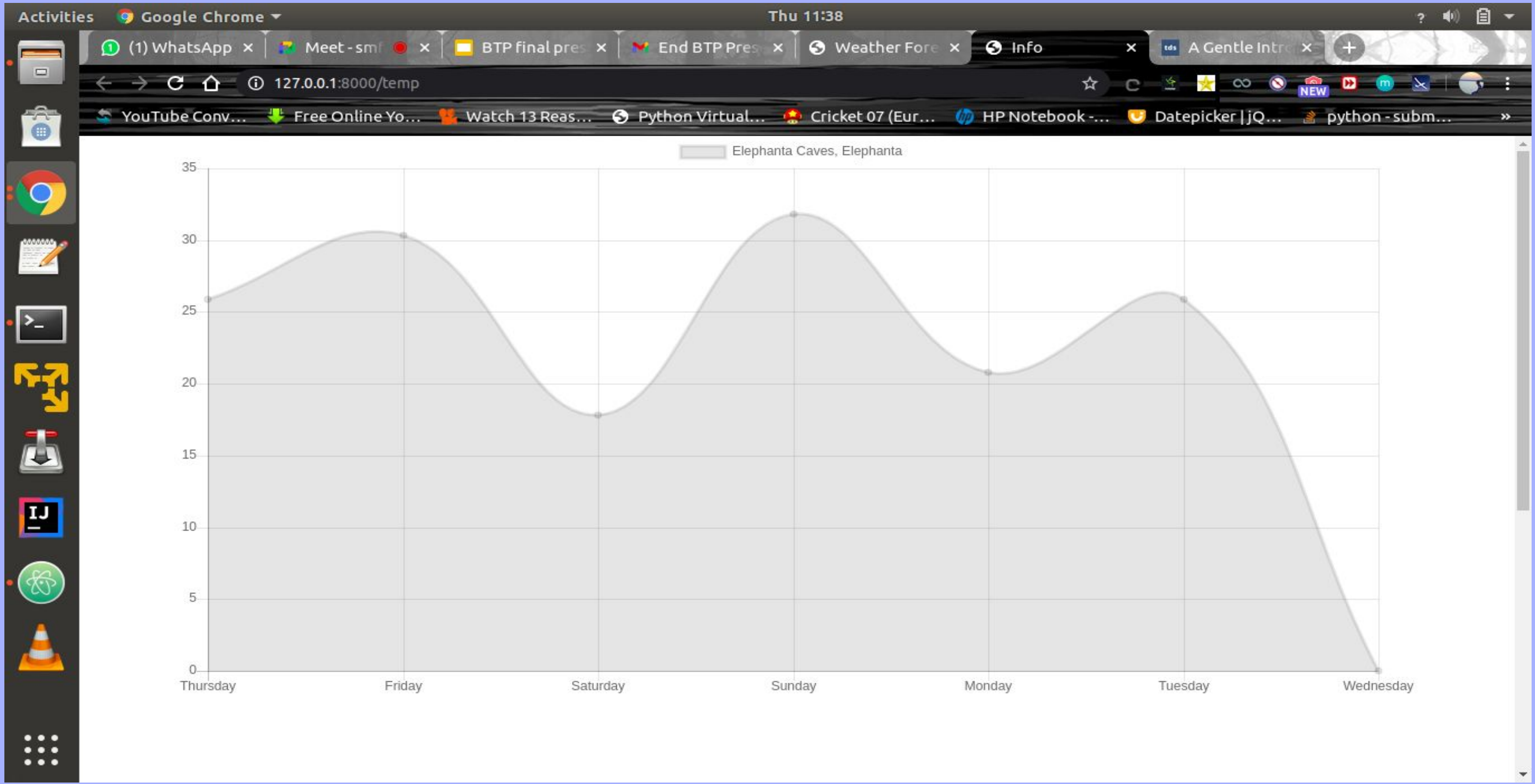


[redacted]

Map visualization

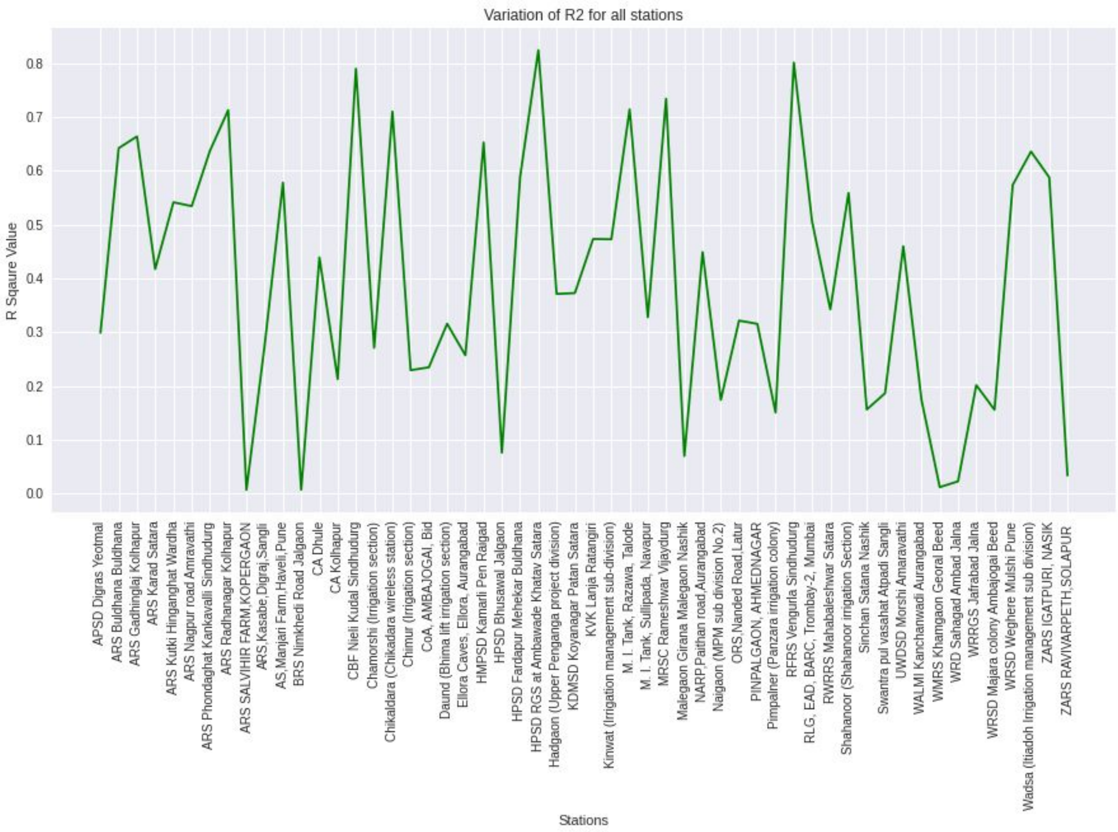


Map visualization

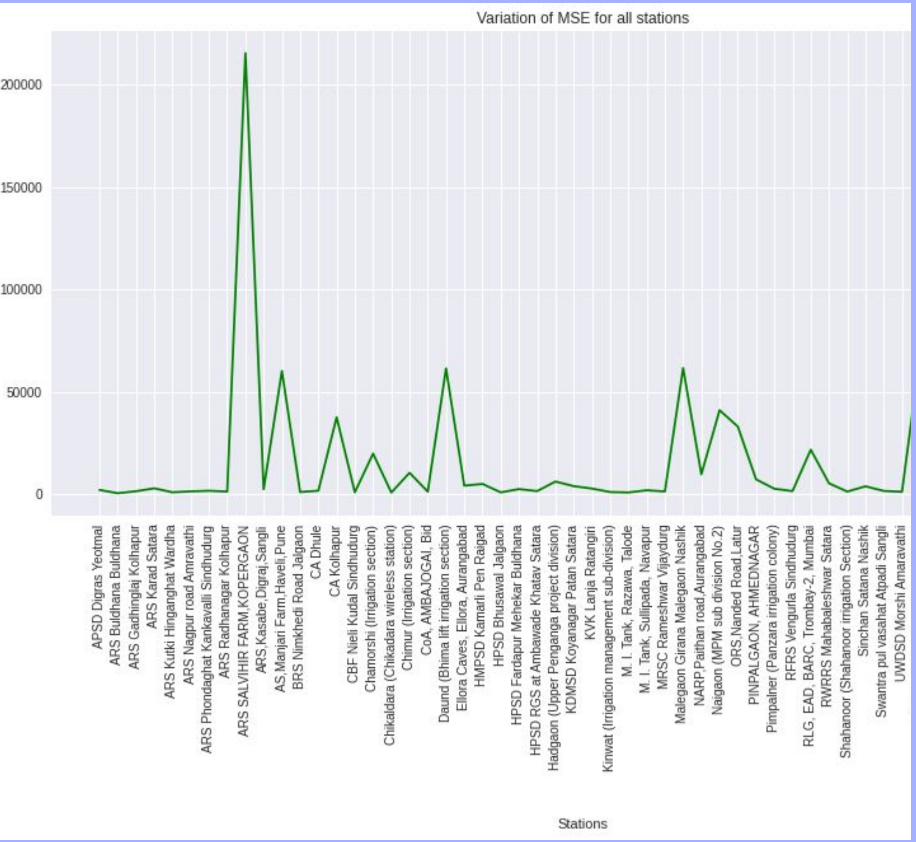


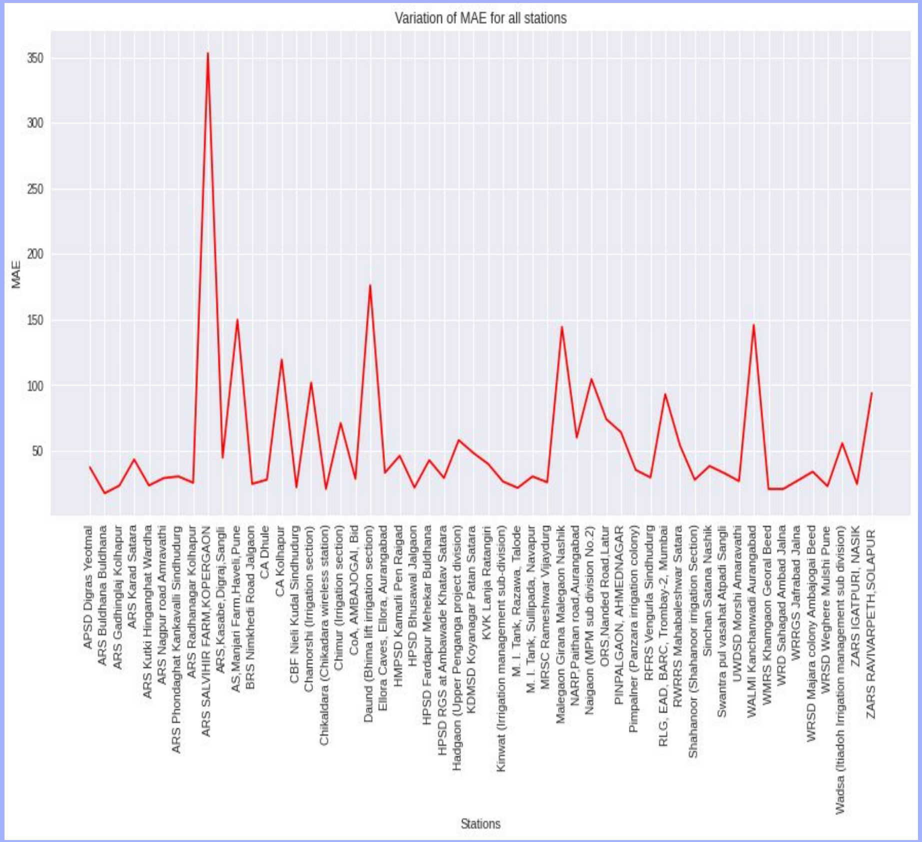
[redacted]

Results

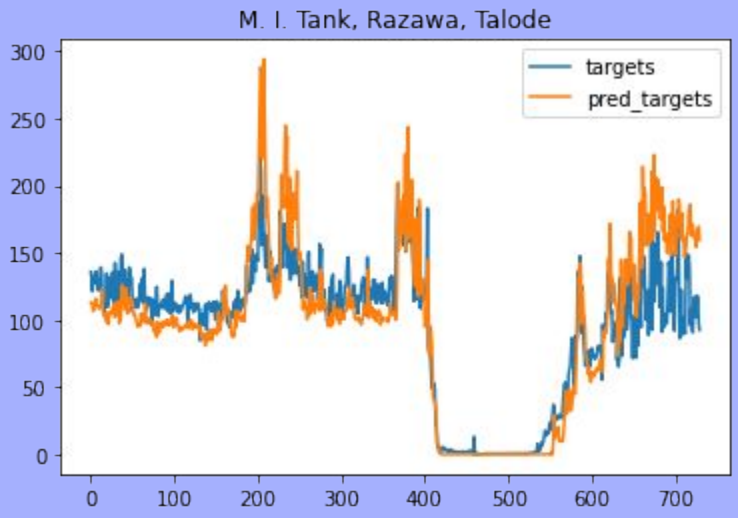


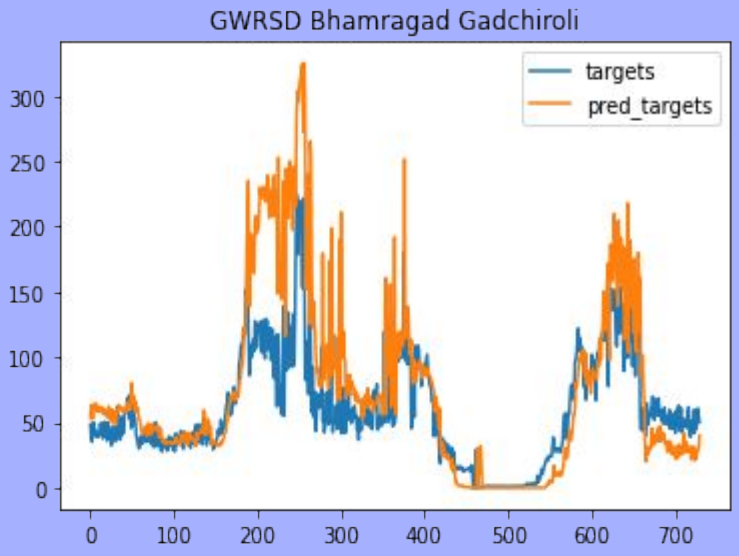
- ➢ The figure shows R2 value of the top predicted stations.





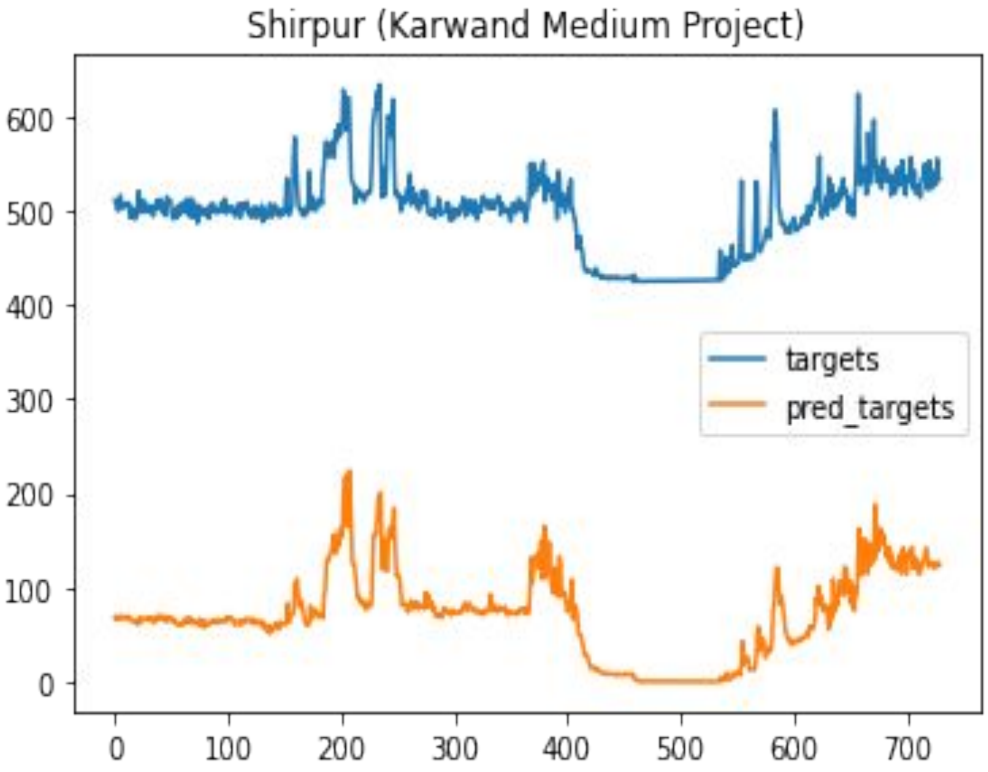
- ➢ The values show the MAE and MSE for all the top predicted stations.





- The above figures show how the predictions and actual values vary.

Results

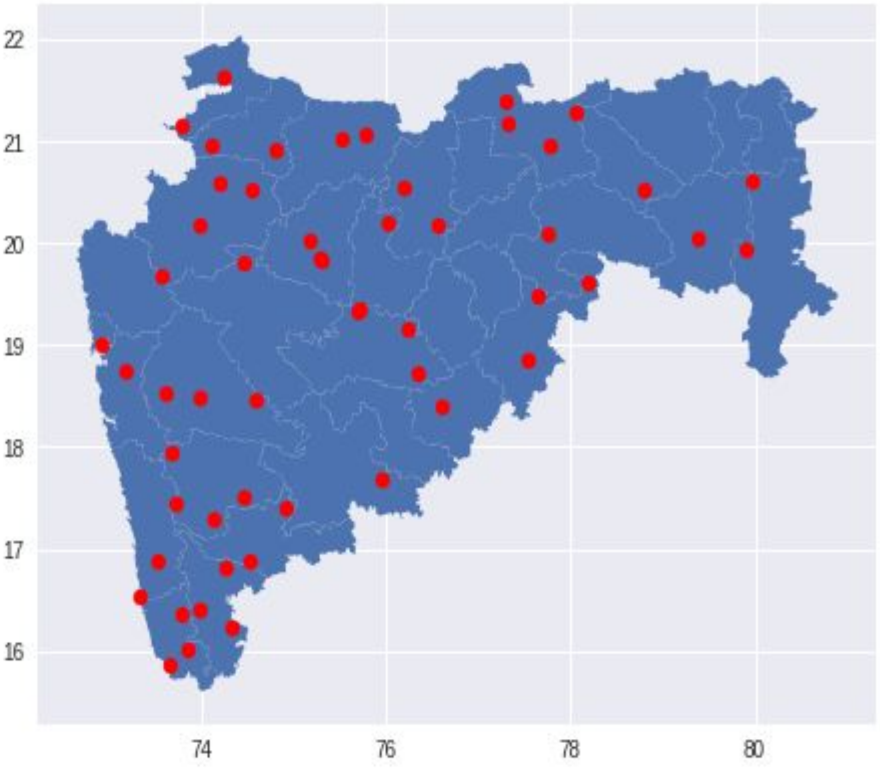


- For some of the places the prediction was different. The heightening and lowering of the curves matched.
- The reason is less amount of data is available, therefore there are certain bumps while filling NaN values.
- Though the rise and fall are equal in both the graphs, sudden increase causes the damage.

Result

- ➢ The places where rainfall didn't show much variation over the years had best results.
- ➢ The places which showed quite high variation in rainfall pattern over the years had bad prediction.
- ➢ This variation is because of filling NaN values by taking the average of temporal and spacial values, there is sudden increase in some of the places on certain days.
- ➢ For these places LSTM would require more data to understand the pattern of variation in rainfall.
- ➢ As shown in the previous slide, predicted rainfall at these places showed resemblance in pattern but values were either higher or lower than the observed rainfall.

Results



- ➢ The figure to the left shows the points plotted at places with good results.

Future Work

- ➢ As we have seen the model is not giving right prediction for some of the places, there a is scope for improvement with a better dataset.
- ➢ Including other meteorological variables like temperature, wind, vegetation index etc.

Thank you

Team Members (MT-03):

Ranjan kumar(S20170010123)
M.Hemanth Reddy(S20170010089)
Vikramadithya(S20170010020)

Mentor:
Dr.Mainak