Maharashtra Rainfall Forecast

Team Members (MT-03):

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

Mentor:

Dr.Mainak Thakur

Overview

- Introduction
- Problem Statement
- Literature
- Objective
- Previous Work done
- Data
- Methodology
- Website Overview
- Future Plan
- 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.
- Maharashtra(19.7515° N, 75.7139° E), one of the fastest growing Indian states and a hub of industries, suffers rainfall events quite frequently.
- We can’t avoid an extreme rainfall event but we can prepare to minimise its effect if we have a rough prior knowledge of its occurrence.

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.

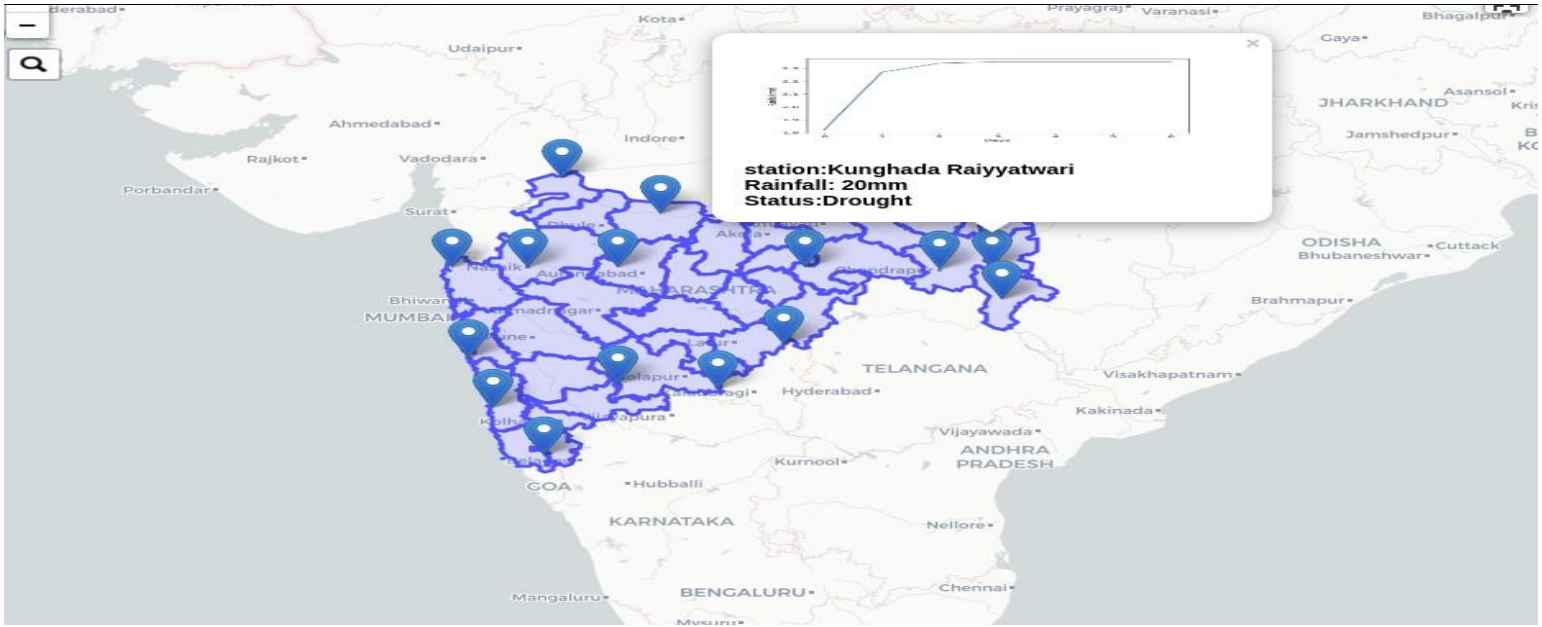
Data

- We have the data of 16 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 72 years(1948-2019).
- The spatial resolution of the values in dataset is 2.5^{o} X 2.5^{o} (~ 277.5 X 277.5 km^2).
- Total number of days in the series is 26298.

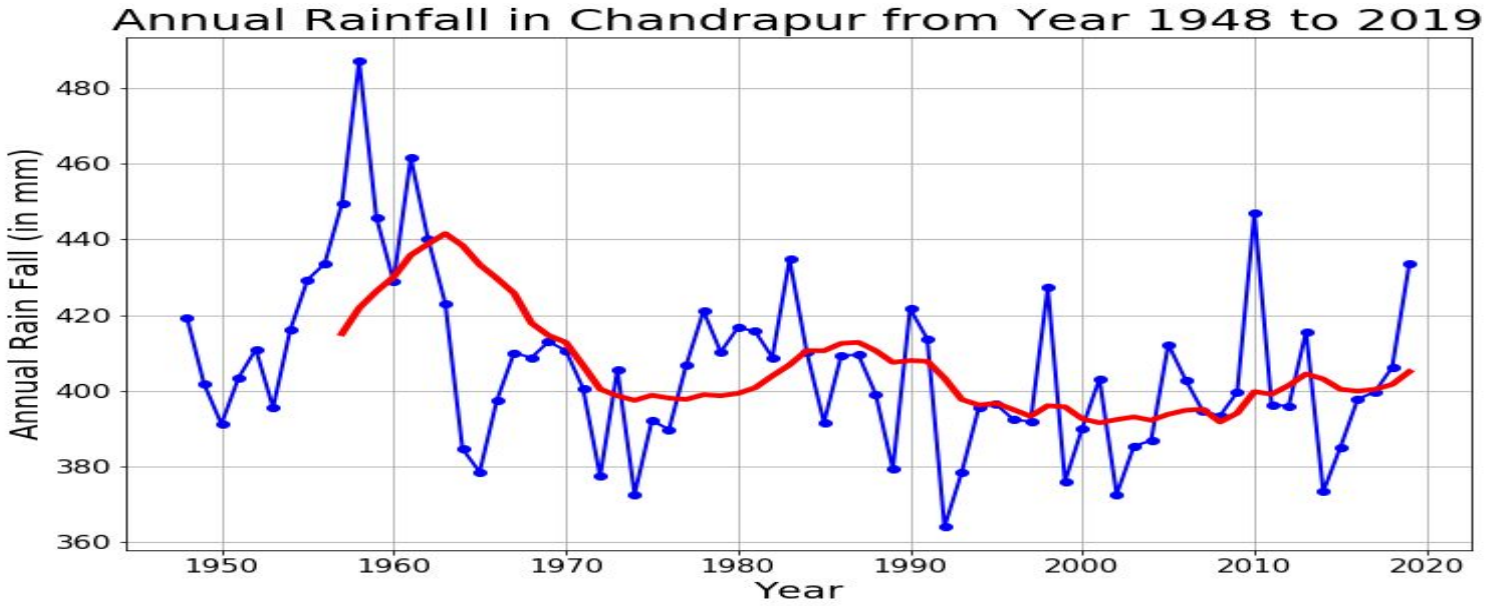
https://psl.noaa.gov/data/gridded/data.cpcsoil.html

Data(Contd..)

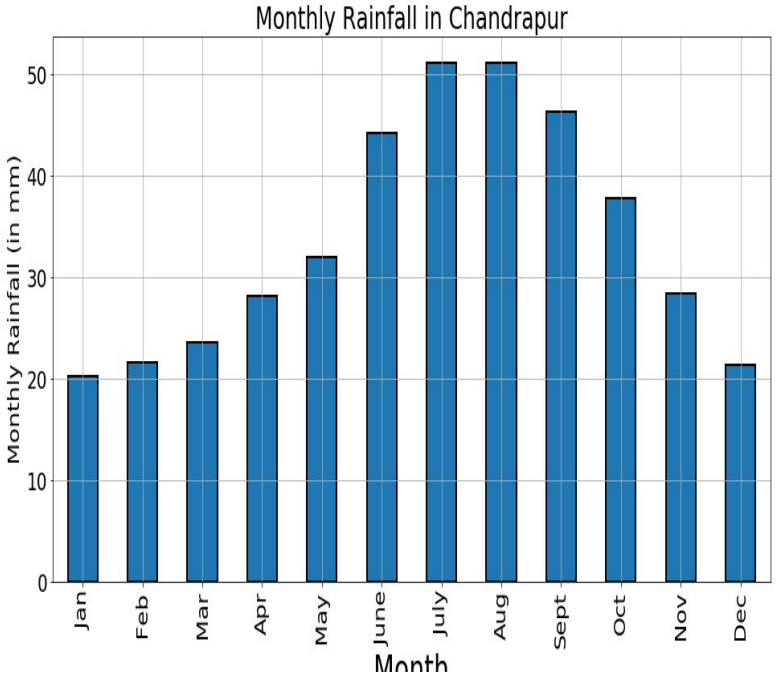
- Currently we are modelling the data for a single district Chandrapur and will expand it in coming days.

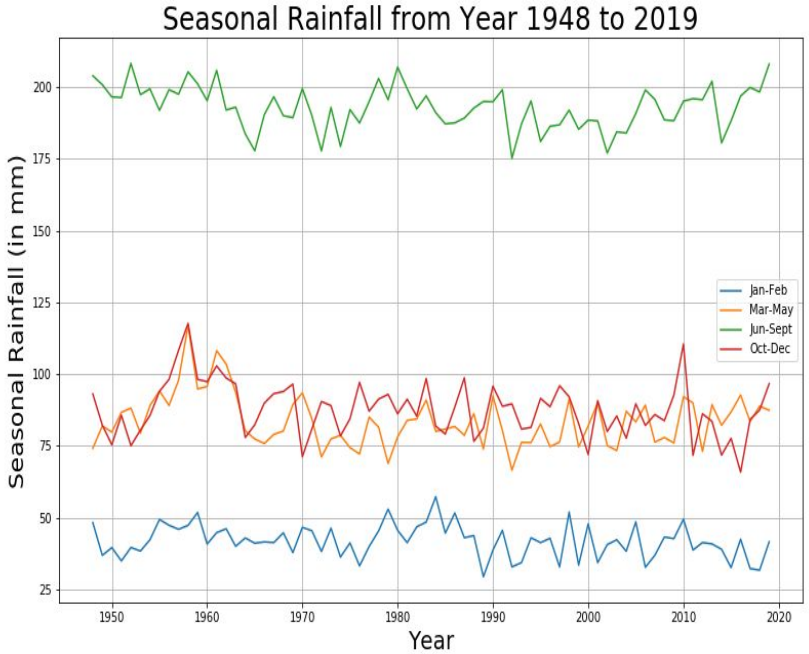


Rainfall

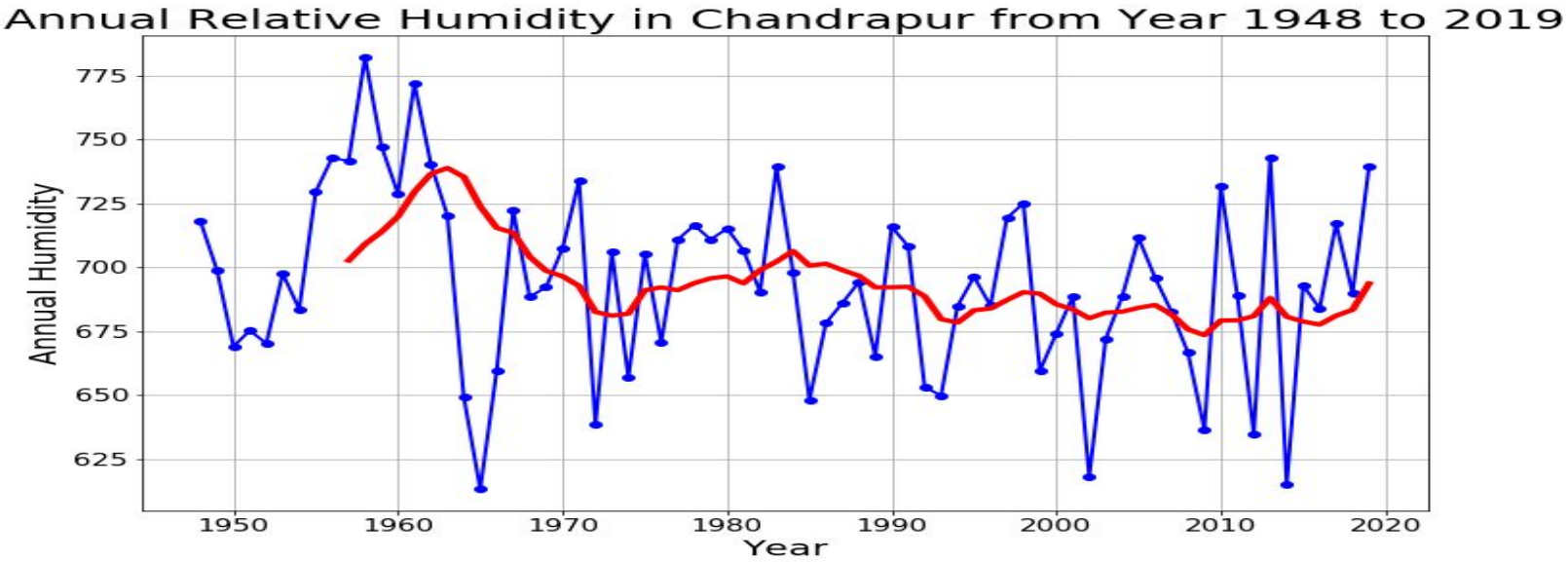

Annual Rainfall in Chandrapur from Year 1948 to 2019
Annual Rain Fall (in mm)
480
460
440
420
400
380
360
1950
1960
1970
1980
1990
2000
2010
2020
Year

Rainfall

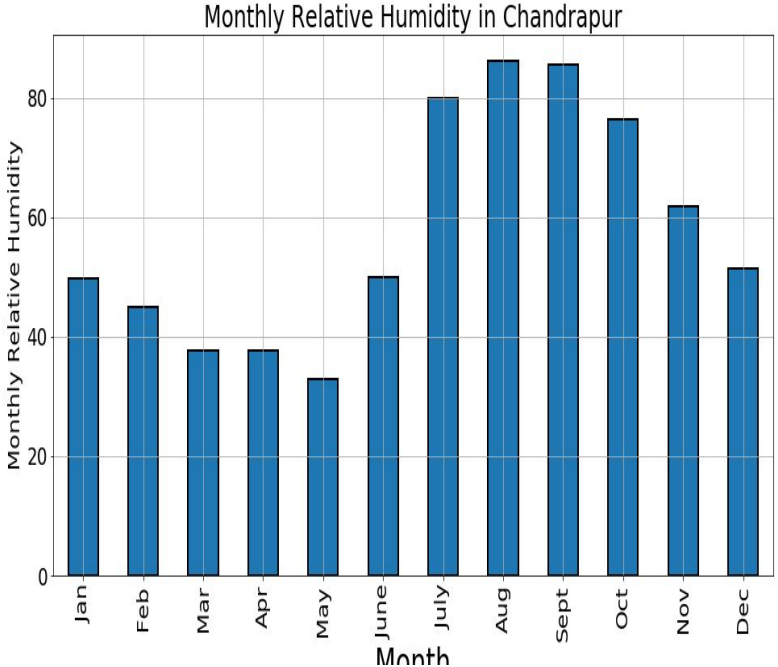

Monthly Rainfall in Chandrapur
Monthly Rainfall (in mm)
0
10
20
30
40
50
Jan
Feb
Mar
Apr
May
June
July
Aug
Sept
Oct
Nov
Dec
Month

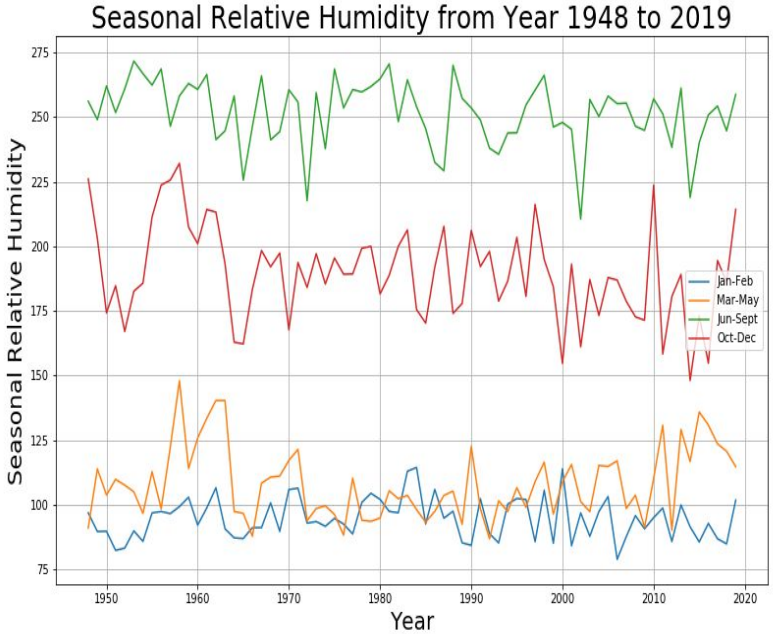

Seasonal Rainfall from Year 1948 to 2019
Seasonal Rainfall (in mm)
25
50
75
100
125
150
175
200
1950
1960
1970
1980
1990
2000
2010
2020
Year
Jan-Feb
Mar-May
Jun-Sept
Oct-Dec

Humidity


Annual Relative Humidity in Chandrapur from Year 1948 to 2019
Annual Humidity
775
750
725
700
675
650
625
1950
1960
1970
1980
1990
2000
2010
2020
Year

Humidity

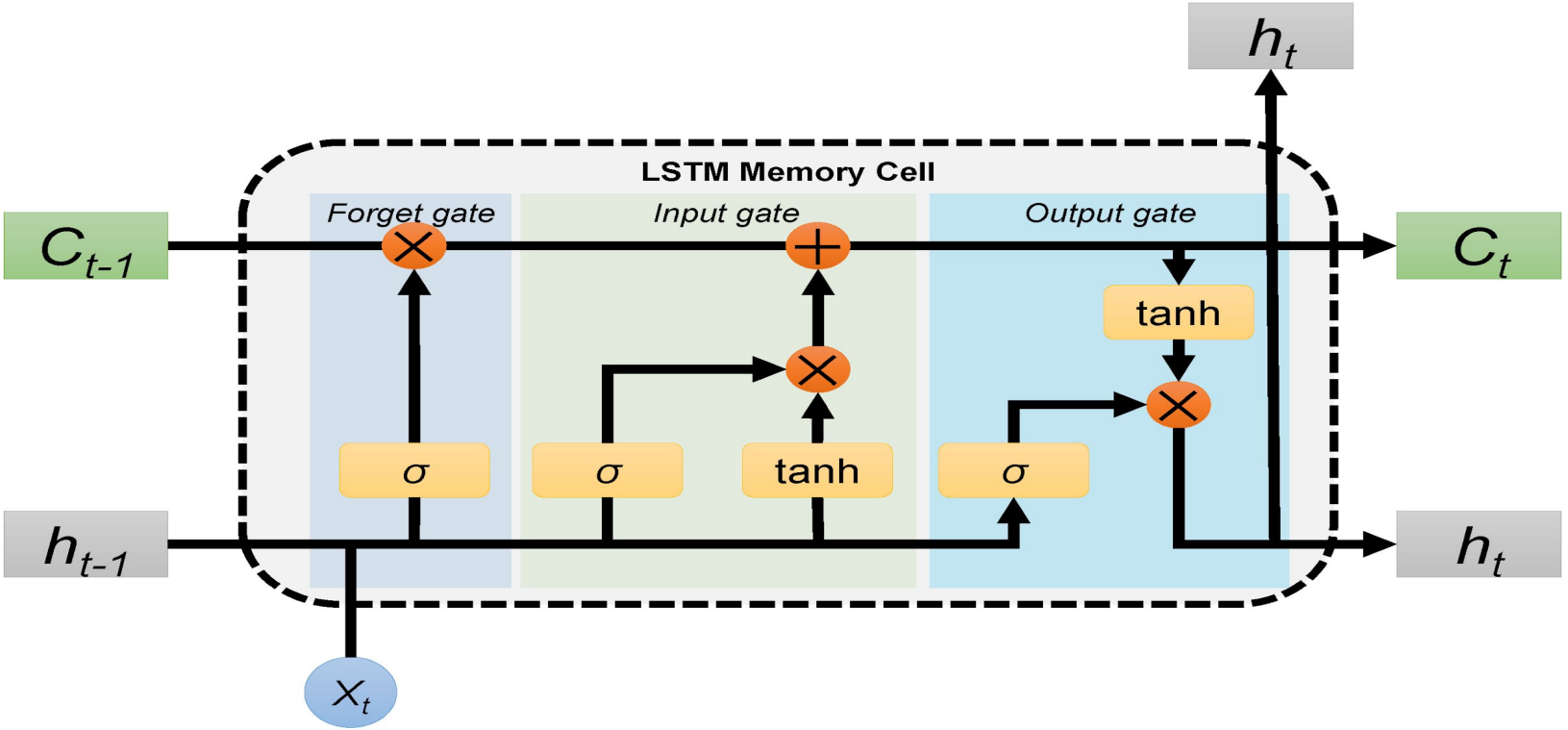

Monthly Relative Humidity in Chandrapur
Monthly Relative Humidity
0
20
40
60
80
Jan
Feb
Mar
Apr
May
June
July
Aug
Sept
Oct
Nov
Dec
Month


Seasonal Relative Humidity from Year 1948 to 2019
Seasonal Relative Humidity
75
100
125
150
175
200
225
250
275
1950
1960
1970
1980
1990
2000
2010
2020
Year
Jan-Feb
Mar-May
Jun-Sept
Oct-Dec

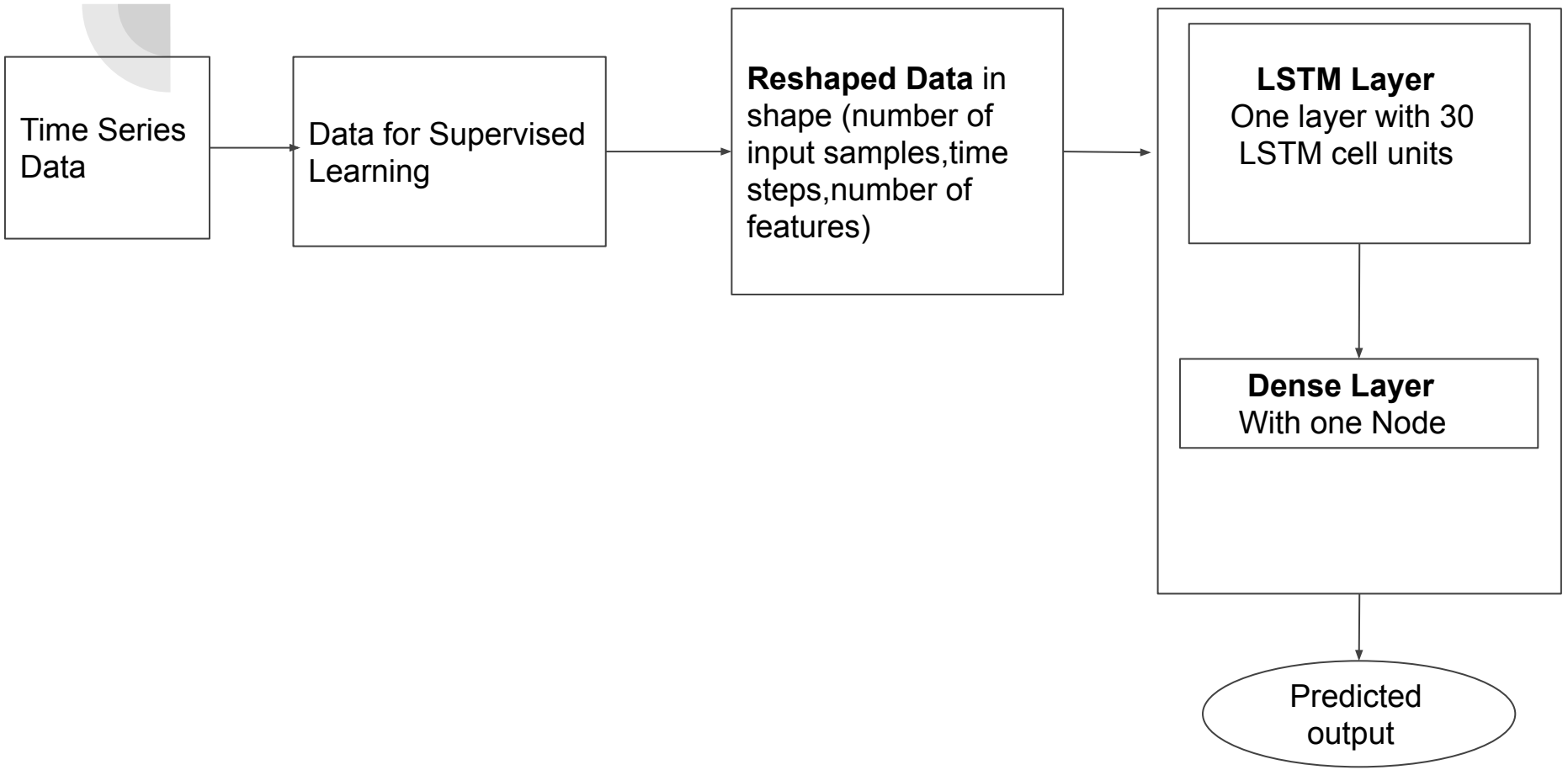
Methodology

- LSTM - Long short-term memory is an artificial recurrent neural network architecture used in the field of deep learning. Unlike standard feedforward neural networks, LSTM has feedback connections. It can not only process single data points, but also entire sequences of data.

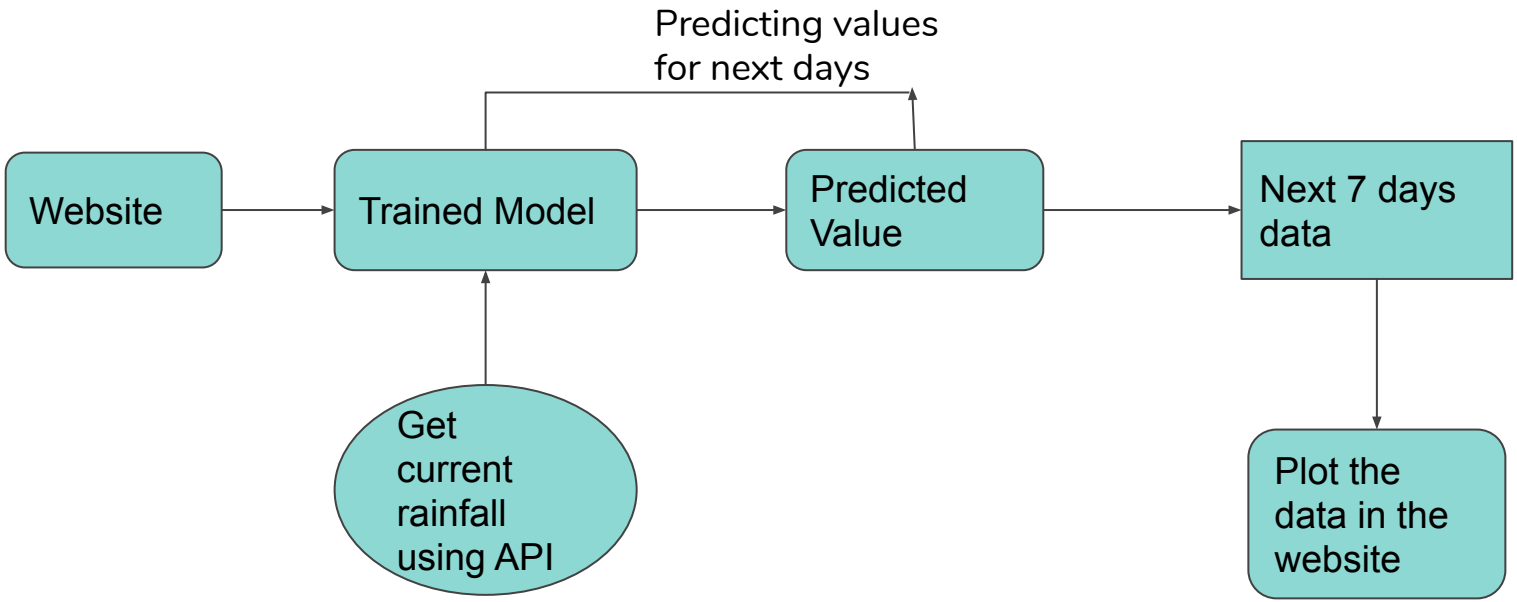
Illustration of LSTM



Model Architecture



Workflow

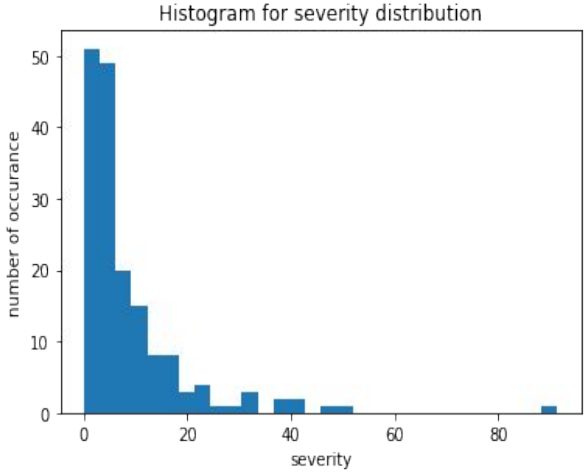

Predicting values
for next days
Website
Trained Model
Predicted
Value
Next 7 days
data
Get
current
rainfall
using API
Plot the
data in the
website

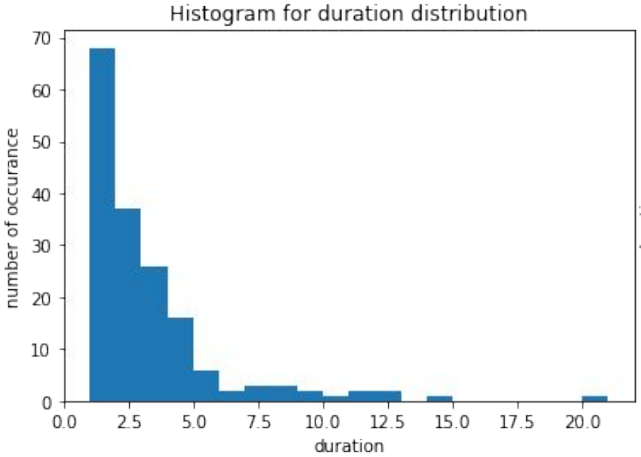
Probability Estimation

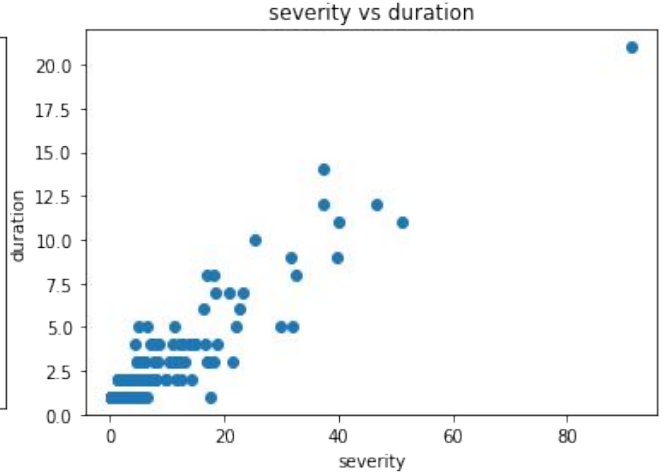
- We did extreme events analysis of Chandrapur for a period of 72 years (1948-2019) *i.e.* 864 months.
- Severity calculation of an extreme event:
  - Threshold rainfall for extreme event occurrence for a month is calculated as the mean of average monthly rainfalls in the particular month for all previous years.
  - Drought event is said to have occurred if an area receives monthly average rainfall less than threshold rainfall.
  - Severity value for a month is the difference between threshold value and actual rainfall for the month.
  - Drought duration is the duration for which a drought period occurs.

Probability Estimation(contd...)

- Joint Kernel Density Estimation for bivariate distribution(duration,severity):
  - We have used non-parametric kernel density estimation because:
    - The joint distribution did not resemble any standard bivariate distribution.







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.

Correlation coefficient = 0.81320855

P-value = 2.4029×10^{-41}

Probability Estimation(contd...)

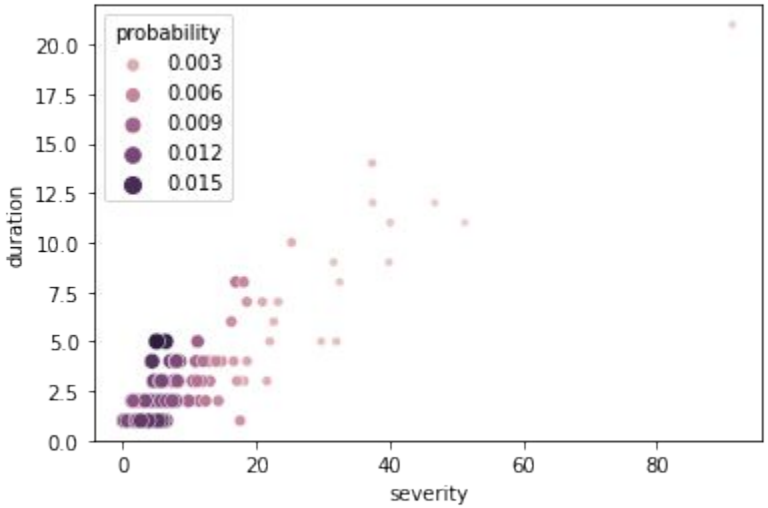
- Probability estimation of predicted severity values:
  - The joint probability distribution function for the bivariate distribution using non-parametric method is given by:

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

  - Here X_i is drought duration, X_j is drought severity.
  - k is the kernel function and 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 was 5.203664.



Risk Estimation

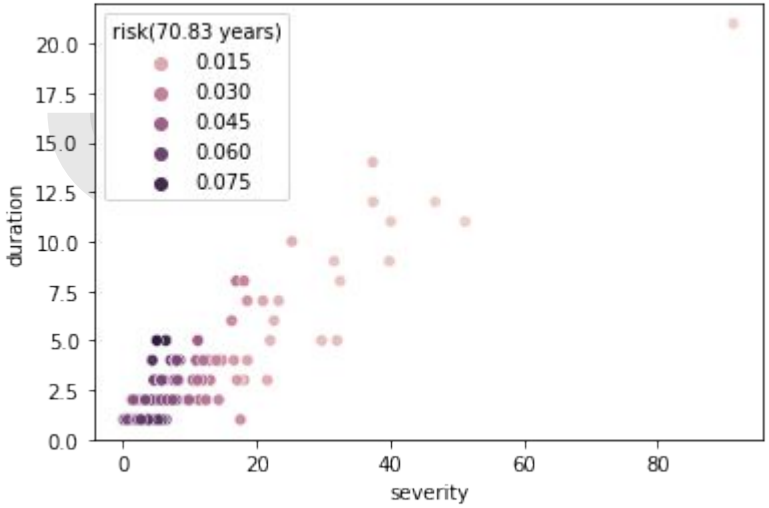
- The risk of drought associated with the bivariate distribution is given by:

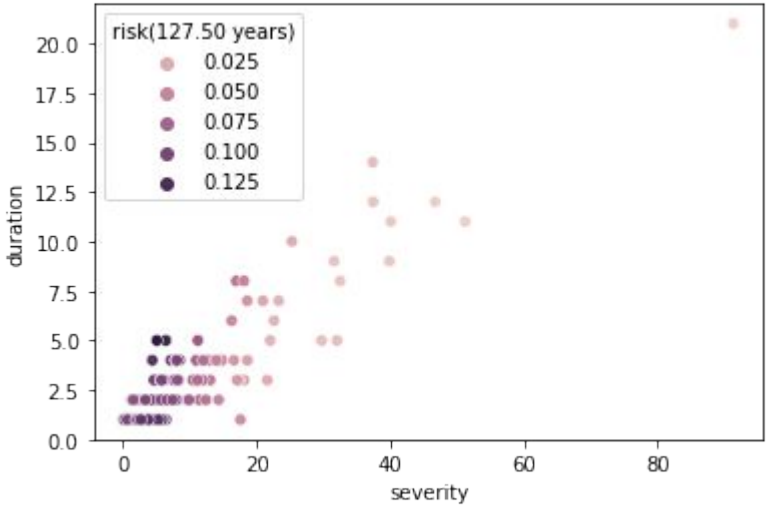
$$R = 1 - (1 - 1/T)^N$$

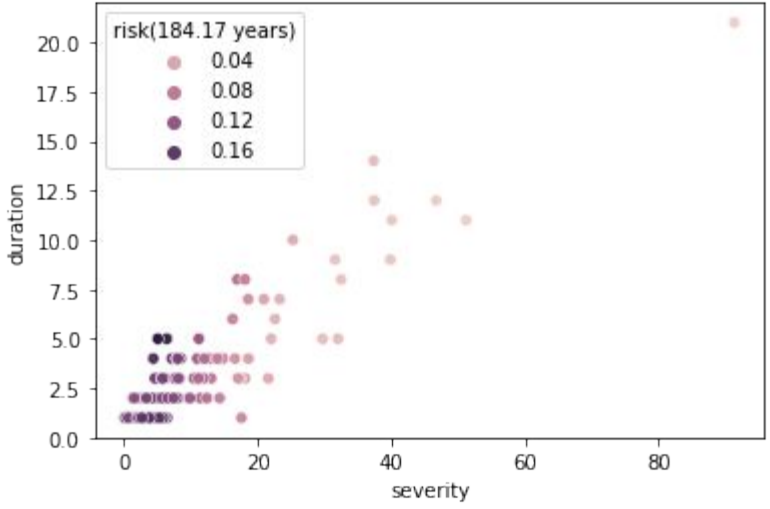
Here T is the return period of a severity value and the drought duration associated with it.

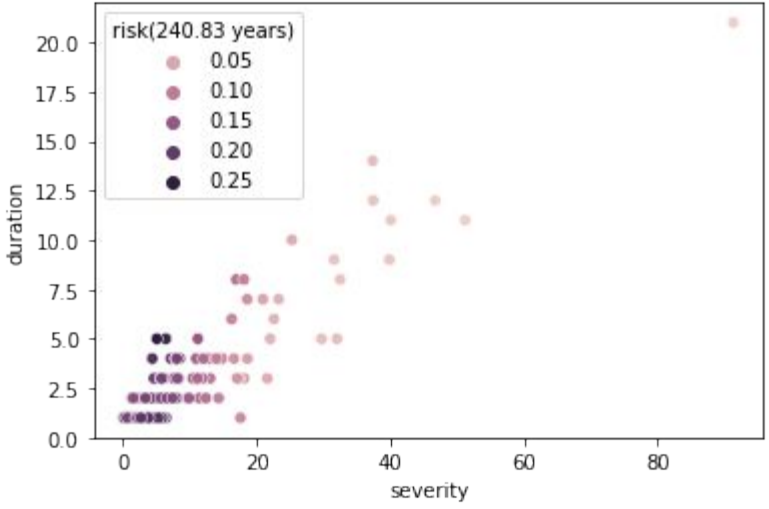
N is the time span for which the risk is being calculated.

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


risk(70.83 years)
0.015
0.030
0.045
0.060
0.075
duration
severity

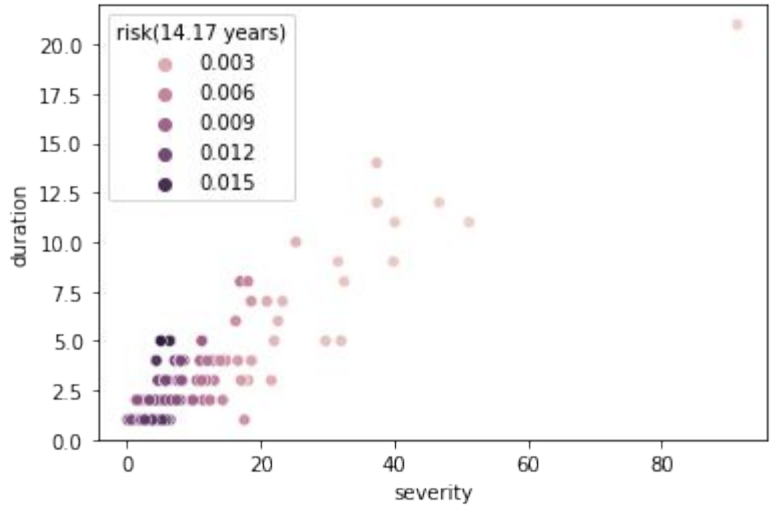

risk(127.50 years)
0.025
0.050
0.075
0.100
0.125
duration
severity

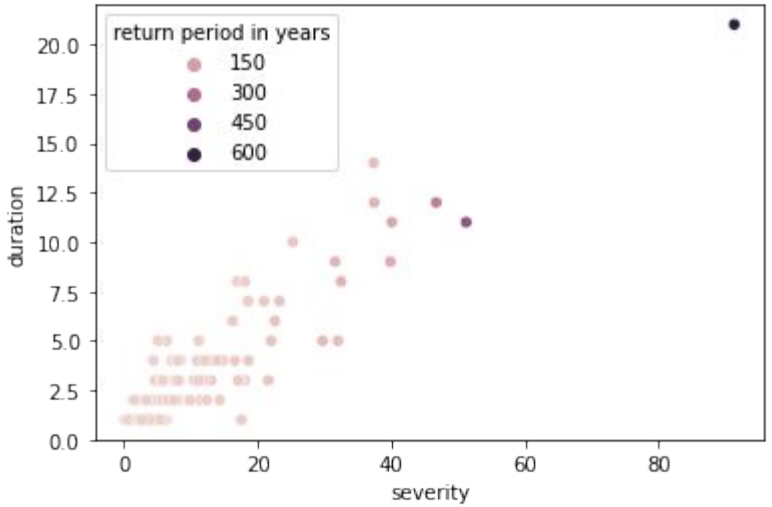

risk(184.17 years)
0.04
0.08
0.12
0.16
duration
severity


risk(240.83 years)
0.05
0.10
0.15
0.20
0.25
duration
severity

Risk estimation

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





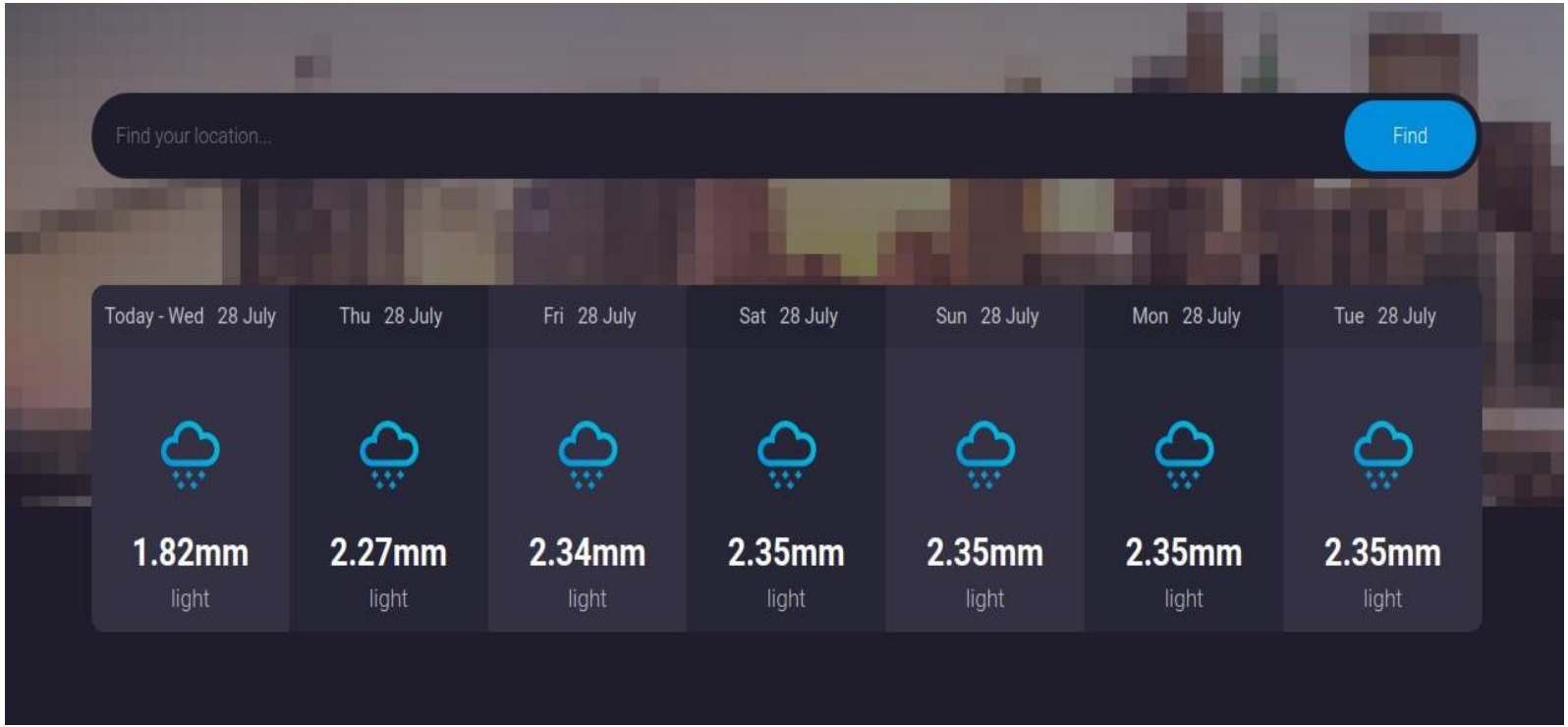
Working Of Website

- We’ve built a website in which we can show the rainfall forecast values predicted by the model for the coming six days of the particular place as of now .
- 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.

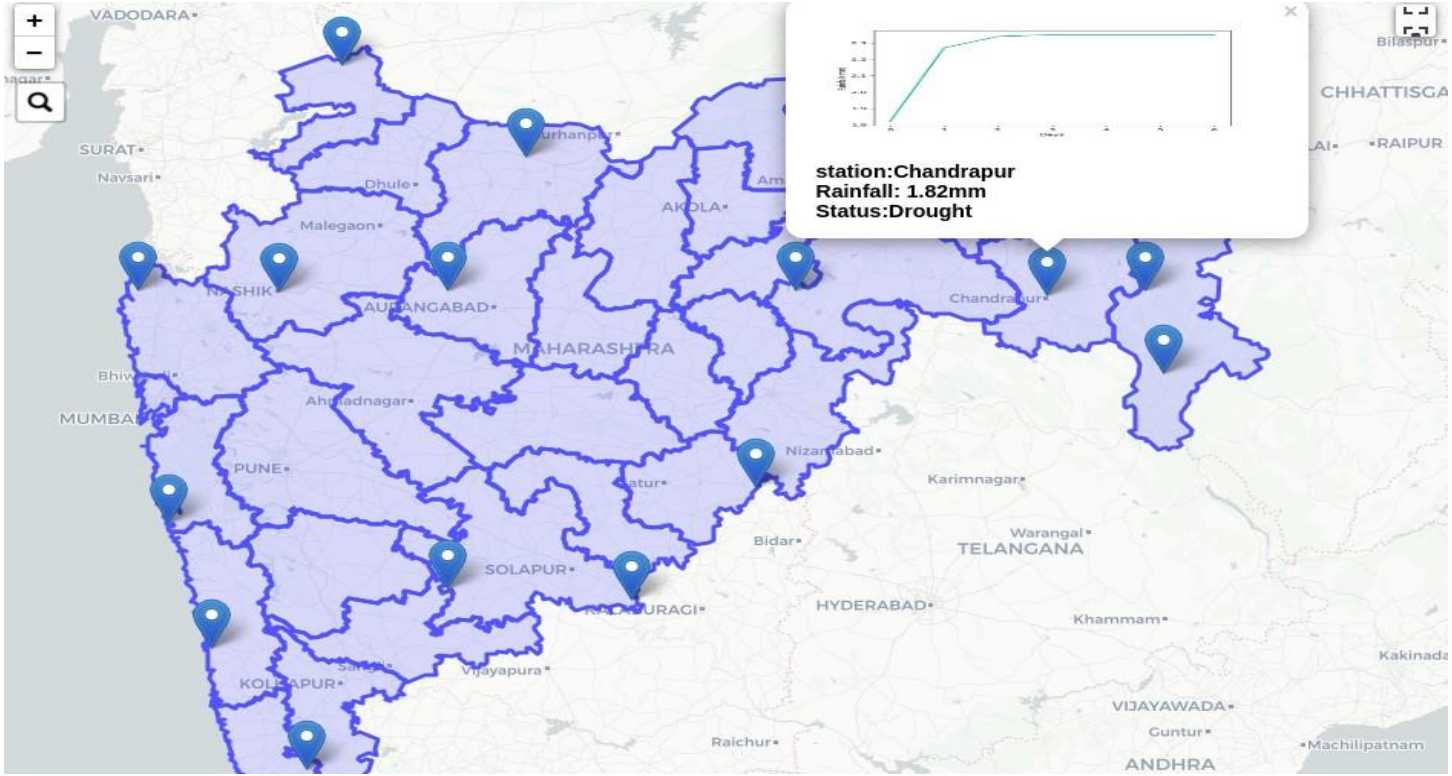
Working Of the Website(Contd..)

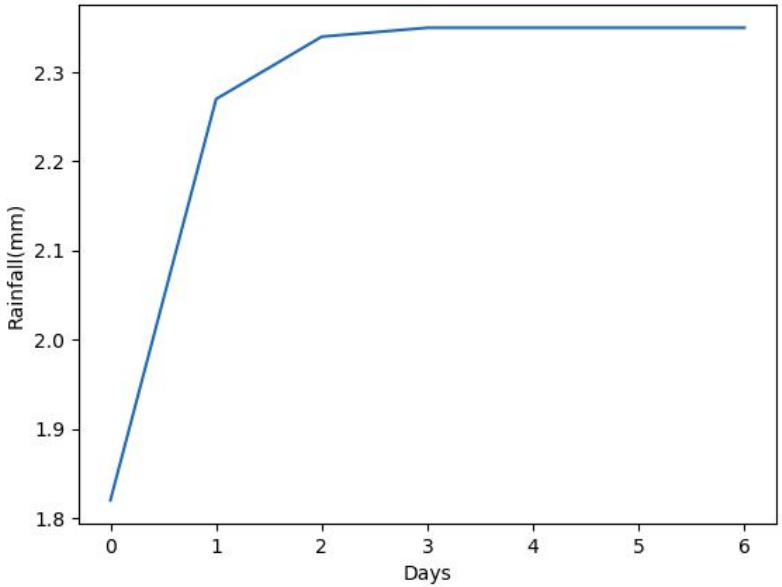
- We've integrated the model trained into website.
- We used weather API to get the current rainfall and humidity data at the place and use the model to predict next days data.
- We keep on appending the predicted value to the current data list.
- We keep iterating over predicted list till we get the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

Working of Website(Contd..)



Map visualization





Place	Rainfall(in mm)	Condition
Chandrapur	1.86	Low Rainfall

Date	Day	Rainfall(mm)
7/10	Wed	1.82
8/10	Thu	2.27
9/10	Fri	2.34
10/10	Sat	2.35
11/10	Sun	2.35
12/10	Mon	2.35
13/10	Tue	2.35

Future Plan

- Estimating Flood risks using accurate statistical techniques .
- Integration of rainfall events risk estimations with LSTM model.
- Representation and visualization of statistics related to rainfall events on the web application.
- Expand the representation for all 16 locations in Maharashtra.