**CLIMATE-INDUCED DISASTER PREDICTION**

TEAM:-

ADARSH KUMAR

ADARSH PATHAK

ANURAG TIWARI

**INTRODUCTION**

The project aims to predict climate-induced disasters and floods using machine learning and deep learning models.

The work timeline can be divided into 3 parts.

· For this project, we read many papers, about general climate disaster predictions.

· Before finalizing the direction of work, we also worked on ArcGIS, tried to visualize the points, worked on raster data points, and worked on DEM models; as earlier, we tried to work on landslide prediction.

· Still, upon reading the papers, we concluded that it requires many parameters and is tough to build upon, so we, under the supervision of our mentor, decided to work on floods and chose Kerela as it has a lot of data points thus, predictions can be valid.

**LITERARY SURVEY**

1.<https://www.sciencedirect.com/science/article/pii/S0040162522001949?casa_token=LzcRsc2ah08AAAAA:DsW5uKHpWSAuU8D1yZWF6oQzxIgw-55BcfG1_1cOow9TPh_Kz--D6zcyJCMXUFRgM3jA1NloHg>

2.<https://www.sciencedirect.com/science/article/pii/S1044028322001077?casa_token=YMbiiAyLUOQAAAAA:jmRiOeGP5GPtFfiHNX59lRi9Ctx-fMrjbhdXSO9lzKd93915NZVMhOHAUCoA9IKNef7wzViRAA>

3.<https://www.sciencedirect.com/science/article/pii/S2212420922001030?casa_token=AmnvTQWNe84AAAAA:d17H8gt-jCFa5RQuCEAedcwAjyvXeHiFSBzDZml7hAnvLJ2eyUmmlAPwxm9KWGODci47srTFGw>

4.<https://www.mdpi.com/2072-4292/13/9/1819>

5.<https://www.sciencedirect.com/science/article/pii/S2665972721000283>

6.<https://www.sciencedirect.com/science/article/pii/S1364815221001791?casa_token=JiO8KPdYk_8AAAAA:taGL9QS5ewr-NzE0DLeBF6shMfvietF8EIoyHUt4yCGdEp9T3Lk2GSe9AJtVColFOyPT0bNSAw>

7.<https://www.diva-portal.org/smash/get/diva2:1680594/FULLTEXT01.pdf>

8.<https://www.researchgate.net/publication/349365397_A_deep_learning_model_for_predicting_climate-induced_disasters>

9.<https://www.mdpi.com/2504-4990/4/2/20>

10.<https://iopscience.iop.org/article/10.1088/1748-9326/ab4e55>

For our work the paper implemented is by **May Haggag**.<https://doi.org/10.1007/s11069-021-04620-0>

The model in the paper aimed to predict floods in the Ontario region of Canada by using some climate parameters to train the model to be used.

In this paper, a deep learning (neural network) model for CID prediction is developed by linking historical disaster records to different climate change indices.

The paper is divided into two main parts,

1. The first part involves the general model structure that is generic enough to be employed to predict any class of CID in any location, given the availability of the influencing spatiotemporal climate data.

2. The second part demonstrates the applicability of the developed model using Ontario’s disaster rerecords and relevant climate change indices data.

This work is considered the first step in CID prediction, based on historical disaster data, global climate models, and climate change metrics, to maximize urban resilience and mitigate CID impacts on cities worldwide.

**MOTIVATION AND PROBLEM FORMULATION**

The motivation behind our work was to use the artificial intelligence techniques for the good of the nation.Since our work was mainly on floods prediction,it is important as occurrence of floods effect the economy of the state or country so it becomes important to know whether the given climatic conditions will affect flood or not.

In this respect, a deep learning model was developed for spatial–temporal disaster occurrence prediction. To demonstrate its application, food disaster data from the Kerala Disaster Database was linked to climate change indices data in order to train, test and validate the developed model.Moreover a similar paper has been implemented for Ontario,Canada so the main motivation of the work was to implement the same thing for indian scenario. In addition to its association with precipitation indices, the study results affirm that flood disasters are closely linked to temperature-related features including the daily temperature gradient, and the number of days with minimum temperature below zero. This work introduces a new perspective in CID prediction, based on historical disaster data, global climate models, and climate change metrics, in an attempt to enhance urban resilience and mitigate CID risks on cities worldwide.

To minimize the impacts of Climate-Induced Disasters (CID), cities must be resilient— able to absorb disturbance and retain their basic functions, during and following such disturbances . One way of addressing the preparedness aspect of resilience is through disaster forecasting which would enable adequate planning and proactive risk management. In this respect, machine learning can be employed because of its ability to as an efficient modelling tool for the prediction of different processes including extreme weather driven hazard realizations . Originating from artifcial intelligence, machine learning assumes that machines (i.e., computational models) can be trained using data records and subsequently can learn to efficiently predict different complex phenomena. Machine learning was recently used for wildfire event prediction; specifically anthropogenic wildfire events were predicted using random forests, boosting and support vector machines.

**DATASET CREATION**

So on similar lines, we attempted to obtain the similar parameters used in the paper and apply them to the Indian scenario in the state of Kerala. Since the dataset was maximum for Kerala, it was preferred. Although with sufficient training data, any region of India could be used.

Now there were many parameters used in the paper, but we decided to use the following set of parameters :-

Now the parameters used by us are listed:

● DTR (Daily Temp Range)- this was the parameter which depicted the extremism of the weather and is relevant in the flood scenarios

● CWD-this parameter related to rainfall and thus was important to consider

● TN10p

● TX10p

Some parametrs like FROST DAYS were irrelavant for KERELA as its temperature never drops below 0 so no point of counting days below 0C.

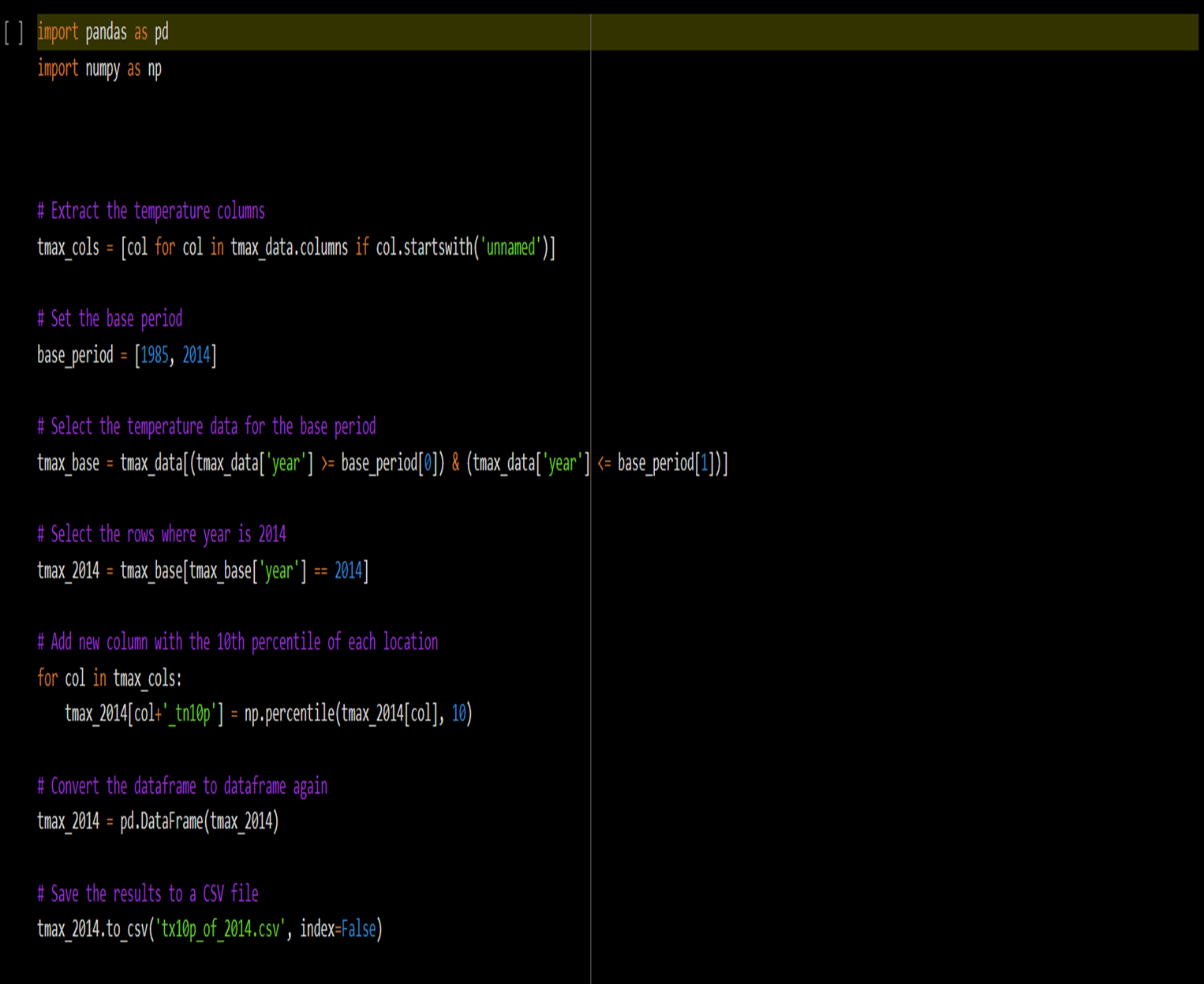
1) DTR (daily temperature range) (Expert Team on Climate Change Detection and Indices 2009; Wazneh, Arain and Coulibaly 2017) represented by Eq. 1, where TXij and TNij are the daily maxima and minimum temperature, respectively, on day i in period j.

2) CWD (maximum length of the wet spell) (Expert Team on Climate Change Detection and Indices, 2009; Wazneh et al. 2017) which is calculated by counting the largest number of consecutive days where RRij≥1 mm, where RRij is the daily precipitation amount on day i in period j.

3) TN10P (Expert Team on Climate Change Detection and Indices 2009; Wazneh et al. 2017) refers to the percentage of days when TNij<TNin10, where TNij is the daily minimum temperature on the day i in period j and TNin10 is the calendar day 10th percentile centred on a 5-day window for the base period 1961–1990.

4) TX10p: Percentage of days when TX<10th percentile. Now, for calculating these parameters’ python was used and the codes will be attached below. Also, the use of excel was done.

**The way to calculate the value of different parameters is given in the readme file where an explanation of how it can be expanded for the different regions is also given.**

****

A sample code is provided for reference.

To incorporate all weather conditions, regions surrounding Kerala were taken, and some points of Rajasthan were included to incorporate extreme conditions.

This was done to remove the bias of the model towards Kerela.

The attempt was made to collect sufficient yes and no points for the flood to ensure data is not skewed. Also, in the Canada Model, some other parameters were used, but they were irrelevant to the Indian scenario, so they were dropped.

**ALGORITHMS USED**

The algos,results and complete code are in this colab notebook.

<https://colab.research.google.com/drive/1m18GnPM47RqbfS9IS0oTHc9IeTayDABN?usp=sharing>

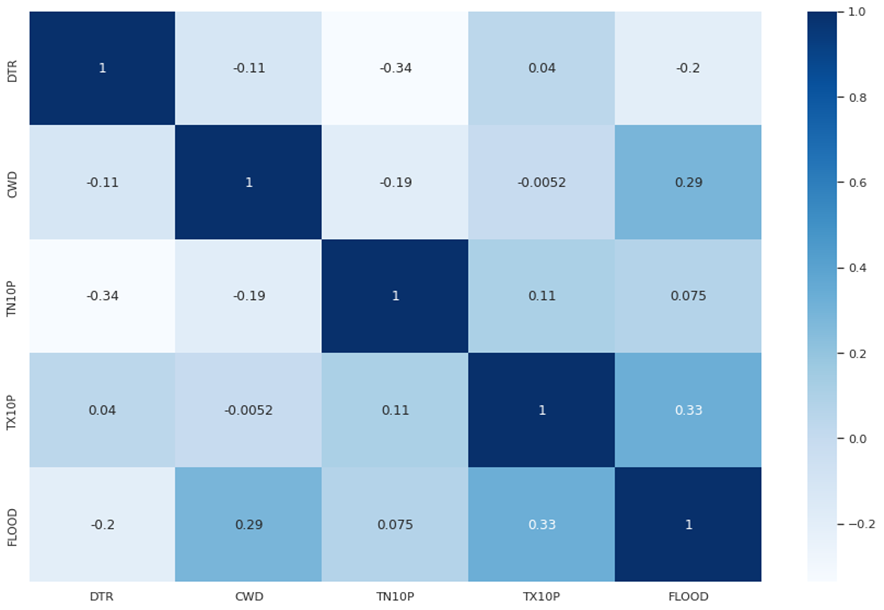
**RESULT**

The results and complete code are in this colab notebook.

[**https://colab.research.google.com/drive/1m18GnPM47RqbfS9IS0oTHc9IeTayDABN?usp=sharing**](https://colab.research.google.com/drive/1m18GnPM47RqbfS9IS0oTHc9IeTayDABN?usp=sharing)

The snippets of the correlation of parameters with output and results of some models are provided below.

**CORRELATION VALUES: -**



We observed that TX10P and CWD give the maximum correlation with the output variable (FLOOD)

**P-VALUES: -**

In statistics, the p-value is the probability of obtaining a test statistic as extreme as the one observed, assuming the null hypothesis is true. The null hypothesis is typically a statement that no effect or relationship exists between two variables or groups being compared.

The p-value is used to determine the statistical significance of the results of a hypothesis test. If the p-value is very small (usually less than 0.05 or 0.01), it suggests that the observed result is unlikely to have occurred by chance, and the null hypothesis can be rejected. On the other hand, if the p-value is relatively large, it suggests that the observed result is not unusual, and the null hypothesis cannot be rejected.

So our null hypothesis was that there was no discrepancy in the data at the time of it being collected and it was independent of all the parameters.

● DTR P VALUE: 0.00715

● CWD P VALUE: 0.00215

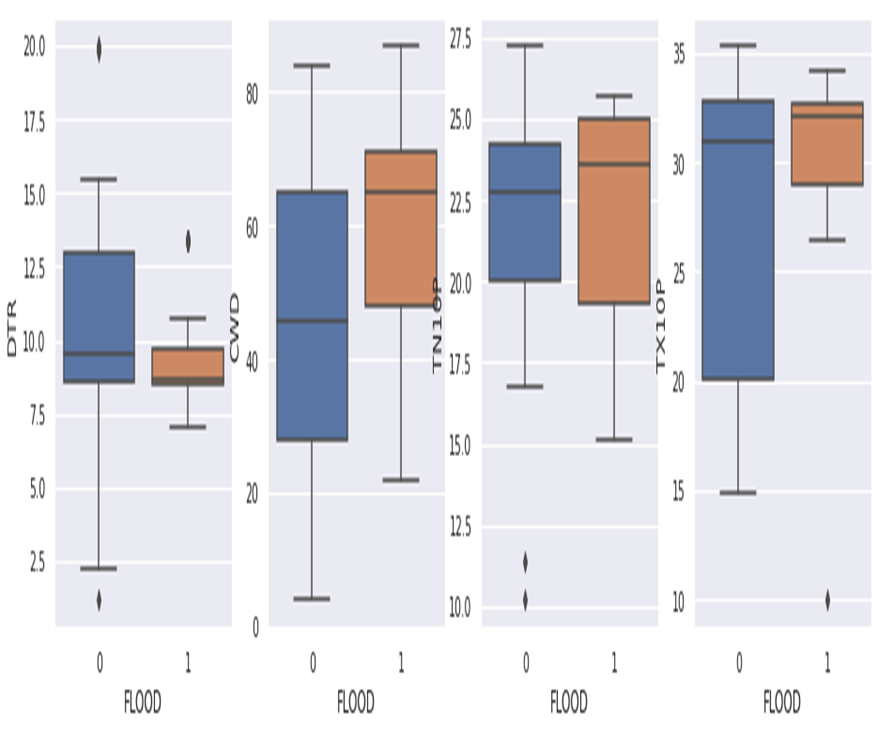
● TN10P P VALUE: 0.19678

● TX10P P VALUE: 0.01739

So we can see the only TN10P had a bit high p value but since it was not very high we will use it currently.

**BOX PLOTS AND OUTLIER ANALYSIS: -**

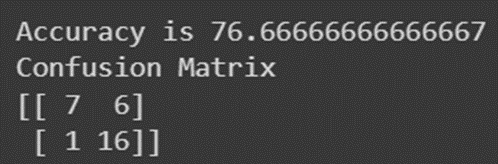
This shows if there are outlier or not.



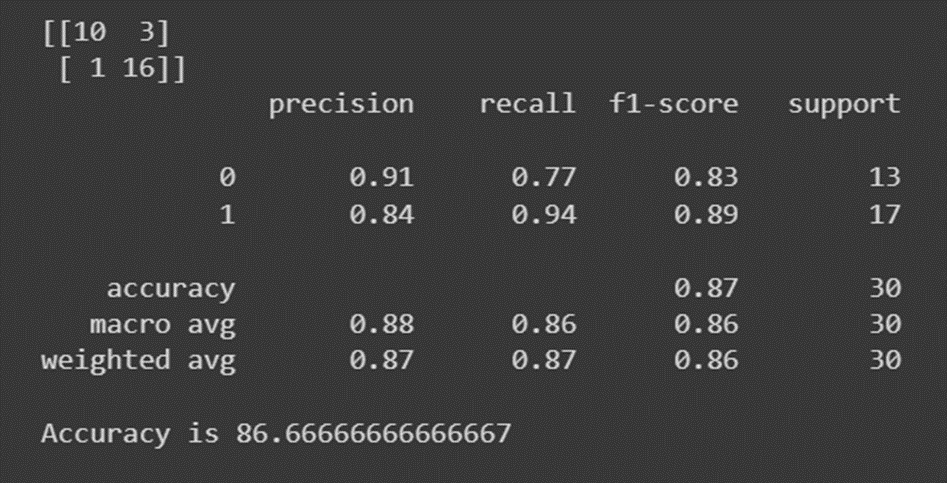
There were some outliers in DTR and TN10P but were less in quantity.

**# ACCURACY AND CLASSIFICATION REPORTS OF MODELS: -**

**LOGISTIC REGRESSION: -**



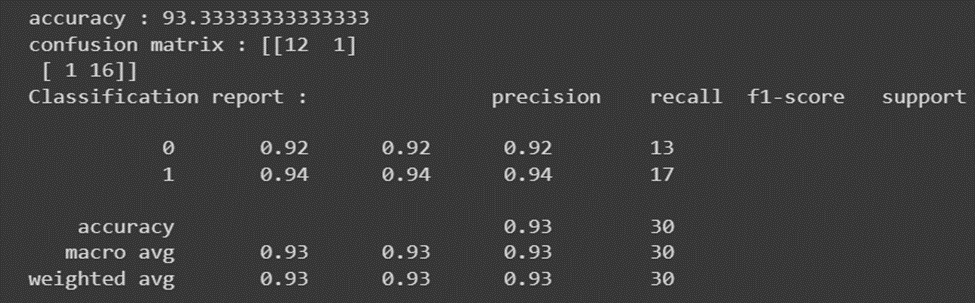
**DECISION TREE: -**

****

**XGBOOST (WITH UNSAMPLING): -**

Upsampling is a technique used in statistics and machine learning to address imbalanced datasets, by increasing the number of samples in the minority class.

Upsampling involves generating new samples for the minority class by replicating existing samples or creating synthetic samples based on the existing ones. This is often done when the dataset has a significant class imbalance, where one class has many fewer samples than the other.

****

**CONCLUSION :-**

• Our model did well and achieved high accuracy and the reason for not using NEURAL NETWORKS WAS THAT THERE WERE LESS DATA POINTS SO IT CAN OVERFIT EASILY AS COMPARED TO OTHER ML ALGOS.

• Our models and techniques to calculate and fill the data worked well, as seen from the p-values.

• Our choice of parameters also worked well, as seen from the correlation values.

• Also, from upsampling, the accuracy increased with XGBoost, so it also worked well.

• Thus, our work can be expanded for different regions also.

**FUTURE WORK: -**

For this ,first one needs to get the historic data from one of the The 12 Global climate models and needs to assume the RCP = 2.6 for data of year>2014. Now one needs to decide the timeline in which one is working like we worked on 1985 to 2014 so we made a separate excel file for these years of Tmax , Tmin , precipitation. Then one has to select the regions and also should have the iunformation about whether there was flood in that region int hat specific year or not and also have to select the regions where there was no flood for y=0 points also. \*\*We have also uploaded our final excel file which contained the values of features and the locations and both flood and non flood points.

**#DTR**:- It is calculated annually and for that say one needs to calculate the DTR FOR 2010 , select the enteries of 2010 in both Tmax and Tmin then take the difference and then average using the inbuild functions in excel.

**#CWD**:- \*\*The code for cwd is given in the colab notebook , ["CWD.ipynb".](https://github.com/Adda2003/climate_disaster_predicton/blob/main/CWD.ipynb)

**#TN10p and TX10P:-** For this first select the base period for which you want to calculate the features , we selected 1985-2014 as our base period and then per day values were stored in a excel file using the code then we took the average for the whole year and calculated the value. \*\*The code for cwd is given in the colab notebook , ["1985-2014.ipynb".](https://github.com/Adda2003/climate_disaster_predicton/blob/main/1985_2014_tn10p.ipynb)

**#Final prediction:-**

We then filled the excel file with the 4 features of a specific location and the 5th feature was whether there was flood in that location in that year or not thats why we calculated all the features annually. Now there are 4 input features and 1 output feature. Apply the binary classification models like logistic regression , random forest , XGBoost Classifier , etc,. \*\*We have uploaded a notebook where we have made the predicitons using different models and also calculated the p-values and correlations in ["flood\_prediction.ipynb".](https://github.com/Adda2003/climate_disaster_predicton/blob/main/flood_prediction.ipynb)