

**Flood prediction using machine learning approaches**

**Supervisor:** Dr. Stish Sarna

**By:** Md Musa Nayem (21104291)

MSc Advanced Computer Science

Dissertation Report

16 January 2023

**Abstract**

In order to forecast floods or droughts, this study analyses climate data using machine learning techniques. The project offers background information and a summary of the literature on floods in Bangladesh, a country with a largely agricultural economy. In order to address this issue, the initiative will offer recommendations to farmers who are planning crops and the national disaster management authority those are ready to tackle issues. The machine learning prototype that will be used for this project will deliver the results of forecasts that have been advised by study. Data are collated using appropriate technique and pre-processed before applying algorithms. The given result from different approaches compared and analysed. In order to plan and make a decision regarding the growth of crops, the outcome of the prediction will be displayed through a graphical interface.

**Keyword:** Machine learning, Algorithm, API (Application programming Interface), pre-processing, Regression, Linear regression, Logistic regression, Support vector machine.

# Acknowledgements

Firstly, I would like to express my sincere gratitude to my supervisors Dr. Stish Sarana for their continuous support for this master’s dissertation, for their patient guidance and immense knowledge. I would like to thank them not only for the insightful feedbacks and encouragement, but also for posing hard questions which motivated me to widen my project from various perspectives.

Last but not the least, I am indebted to all my friends and my family: to my father, my sister, my lovely children and all my extended family for supporting me throughout this project.

**Table of content**

|  |  |
| --- | --- |
| Abstract | 2 |
| Acknowledgement | 3 |
| Table of Content | 4 |
| Table of figures | 5 |
| Chapter 1. Introduction | 6 |
| 1.1 Background | 6 |
| 1.2 Aim | 8 |
| 1.3 Objective | 8 |
| 1.4 Overview of dissertation and result | 8 |
| 1.5 How to read this dissertation | 9 |
| Chapter 2 Literature Review | 10 |
| 2.1 How can machine learning aid in flood prediction | 10 |
| 2.2 Utilising machine learning methods | 11 |
| 2.3 Machine learning as an efficient prediction of flood severity | 13 |
| 2.4 Possible ethical issues of using machine learning in flood prediction | 15 |
| Chapter 3 Methodology | 16 |
| 3.1 Overview of system development | 16 |
| 3.1.1 Data collection | 16 |
| 3.1.2. Data Preparation | 16 |
| 3.1.3 Data Analysis. | 16 |
| 3.2 Data Collection Apprach | 17 |
| 3.3 Theoretical Background | 18 |
| 3.3.1 Transfer learning | 18 |
| 3.3.2 Types of machine learning | 19 |
| 3.4 Data preparation | 19 |
| 3.4.1 Dataset description | 19 |
| 3.4.2 Data pre-processing | 20 |
| 3.4.3 feature selection | 21 |
| 3.5 Modeling | 21 |
| Chapter 4 Result and Discussion | 22 |
| Chapter 5 Conclusion and future Work | 23 |
| References | 24 |

**Table of Figures**

|  |  |
| --- | --- |
| Fig 1: Rate of use of different machine learning tools for flood prediction over the years | 15 |
| Fig 2: python API for real time weather dataset | 18 |
| Fig 3: Weather dataset description | 20 |
| Fig 4: Feature Scaling | 21 |

**Chapter 1. Introduction**

* 1. **Background**

Bangladesh is an agricultural country. Most of the economy and GDP(Gross domestic product) depends on agriculture. Water plays an important role in the production of agricultural products, but if flood or drought happens then the economy will fall in risk. In Bangladesh around 70 % of total population live is rural areas (Yousuf Zaman,2018). Majority of people of rural areas are depends on agriculture. Sometime the farmers fall in lose when unacceptable rain fall happens that cause damage crops. In this year 7.2 million of people affected by flood and damaged their houses, which will impact on food production, also floods may harm soil fertility. The production of rice is decrease by 350,000 tonnes due to the flood of May and June in Bangladesh (Ifthekar Mahmud, 31 july 2022).

The following concern is not only for Bangladesh it’s a major issue for all other countries in the world. In the following year the flood happened in Pakistan and India and its damaged lots houses and crops and also died lots of peoples. In 2022, 1000 villages and 22 districts of India also affected by unwanted flood which impact on the economy of India (RICHARD DAVIS, 1 September 2022).

From this problem high volume of crops and damaging and people are falling treble every year. The farmer and government body could lead this problem if the now the possibility of flood due to measuring rain falls. They can active other bodies (like civil service or disasters management authority) to take actions to pretend the damages.

It takes time to comprehend the challenges faced by other people and the significance of people's lives. Many individuals lost their homes as a result of the flood, and many more passed away from illnesses like leptospirosis, which is brought on by contaminated water, hepatitis E, and gastrointestinal problems. When a flood occurs, all of the rivers and ponds are filled, and all types of water are mixed together, becoming poisonous and fostering the growth of bacteria that are dangerous to human health. The flood caused property losses in the USA as well, and it can also alter societal structure.

For the reason of drought the global economy also suffered, From the survey authority every year the global economy lose 6 to 8 billion U.S dollars (Qiang Zhang, 2019). In news papers and others social media they also published the impact of drought and the disaster management authority also showed their methods to tackle this environmental issue but it can be successful if they can predict that disasters.

The another things which is can be the biggest factor which is due to sudden flood, the food price and other commodities price become higher then those are living on poverty they become more poverty after that they could start doing crimes like theft and robbery, As seen by them, others are also will engage in this type of work. The social structure could be changed which will be biggest issue for upcoming generation.

Due to this concern, hydrological researchers attempted to resolve the problem, but in the present era, Artificial Intelligence (AI) would be more successful in doing so. People's lives are made simpler and faster support is provided via technology. In machine learning, a computer may analyze a dataset and use mathematical formulas to provide the best results. The hydrological system is ageing as a result of bettering infrastructure to deal with water levels, but it does not foresee floods. Machine learning allows for the analysis of historical weather data, and the algorithms' output predicts how much precipitation will fall. Some researchers employed the decision tree approach to address regression-type issues, but after thorough consideration, they discovered that it is ideal for applying to classification issues.Using deep learning technique, it allow to make model to predict target data trend (Yann LeCun, 2015).

* 1. **Aim**

Utilize machine learning principles to forecast rain or floods, helping farmers grow crops and governmental entities take action to stop flooding and preserve lives.

**Research Questions:**

The following research question must be addressed as part of this project:

Q1: How flood can be predicted by using machine learning technique?

Q2: Why you think ML is efficient in predicting flood?

* 1. **Objectives**
* Look up earlier research on analysing rainfall data using a machine learning framework.
* Collect live dataset from valid source using advance technical approaches
* To explore dataset to get insights as to what the dataset is like
* Visualize any relationship between variables
* Pre-process the given dataset using appropriate technique
* Identify possible method to be utilized for this project
* Design, develop and test machine learning prototype on the given data set.
* Apply some regression algorithms on the dataset
* Compare , analyse and evaluate the result
* To use further (test set) data to see how the model would perform in real world
  1. **Overview of Dissertation and Result**

This dissertation presents a systematic approach to utilize different methodologies to design, develop and evaluate a prototype which would support the people or farmer who can get forecast about flood and can take initiative to tackle.

The brief overview of the chapters is provided here, with detailed explanation later in the respective chapters. The key message of this dissertation is: Data Mining,Machine Learning can be utilized effectively to develop systems capable of supporting farmers by climate data.

The dissertation is organized as follows

• Chapter 2 provides the essential research to describe our work and its relationship to the literature, and describes the literature review.

• Chapter 3 discusses the Project Proposal's addition, the Methodology. This also incorporates the System Development Methodology, which has particular advantages and restrictions. Additionally, the theoretical backdrop offers a thorough explanation of the suitable methods and machine learning algorithms employed in the project.

• Chapter 4 explains the entire technique and how it was compared to the conclusions of other researchers.

• Chapter 5 outlines the conclusion and further work.

* 1. **How to read this dissertation**

Every chapter is briefly summarised in the introduction chapter, which also provides a thorough explanation of how the outcomes of other chapters connect to one another. Each chapter can be read independently of the other chapters in this thesis because they are all independent and stand on their own. A rudimentary understanding of optimization, data mining, flood analysis, and machine learning is also expected of the reader..

**Chapter 2. Literature Review**

* 1. **How can machine learning aid in flood prediction:**

As a result of over rain fall or less rain fall, there has grown an idea of using machine learning methods to predict the flood and by the use of previous climate data. In machine learning which is workby some statistical concept and some some set of algorithms. Thisconcept helps to classify riskarea and can predict where and how much rain fall could be in the following or next year.The flood nonlinearity can be numerically formulated by ML models merely using historical data, without the need for understanding the underlying physical processes. In comparison to conventional statistical models, machine learning (ML) models are more accurate, evolve more quickly, and require less inputs. Using this method, a variety of binary classification algorithms were taken into consideration because, depending on their underlying statistical presumptions (such as linear separability, normality, and homoscedasticity), some algorithms are more suitable than others and necessitate some initial testing and model evaluation. As a result, a set of six classification models was deployed and assessed using their corresponding scikit-learn implementations in order to cover a wide range of heuristics used for ML. According to reports and reviews by Mosavi et al., these ML models were built on commonly used algorithms suitable for flood predication.

• Logistic Regression (LR): a linear, maximum-entropy classifier that estimates a linear function to determine the probability for the class variable;

• Support Vector Machine (SVM): a gaussian, kernel-based classifier that determines a separation hyper-plane between the two classes and then returns the projected class;

• Gaussian Naïve-Bayes (NB): a non-linear Bayesian classifier that calculates the conditional posterior probability of the class variable under the assumption that the predictors have a gaussian distribution.

• Random Forest (RF): a non-linear ensemble classifier that uses rule inference to calculate the probability for the class variable and is obtained by a group of decision trees;

• K-Nearest Neighbours (KNN): a non-linear, instance-based classifier that determines a class variable's probability based on the k-nearest observed instances;

•Multi-Layer Perceptron (MLP): a feedforward, nonlinear artificial neural network that determines the likelihood of the class variable.

* 1. **Utilising machine learning methods:**

There are many methods can be used for this project. Clustering, classification and regression are popular but regression would be fit for this problem to findout solution. Regression usually used to predict the trends according to previous data or trends like the possibility of flood due to the amount of rain fall. Data preparation is the initial step in machine learning, which in this case is getting the datasets and cleaning them. Remove dead data, such as null values and other examples. This is done so that the machine can train using all of the data (Tompson, L., et al 2015). In the recent years, more machine learning algorithms have been developed to keep track of numerous datasets across various application areas. These algorithms have a wide range of qualities they may assess, and each object is represented by a unique label. In essence, the mapping between the input space and the output space is performed by a mechanism that the machine learning method model learns to construct. Several classification algorithms, including ADASVM, ADANAIVE, J48, and IMLP classification, are employed for the analysis and classification of flood sensor data.

**ADASVM classification**

The Support Vector Machine is the method employed in pattern recognition (SVM). Utilizing this technology will reduce structural risk. The closeness of data points in the hyperplane indicates support vectors. SVM seeks to create a model using a training dataset. Based on the properties given in the test data, it lends destination values. When dealing with linearly separable data, the hyperplane's biggest margin, the SVM, is used to partition the data. It is used to analyze a dataset's testing and training results.

**ADANAIVE classification**

The Bayesian theorem serves as the foundation for the Naive Bayes classifier's operation [3]. When the input has a high value, it is appropriate. This theorem can be used more readily if the attribute only has one value. The computation is done using a formula in accordance with the Naive Bayes theory under the presumption that attributes must have independent distributions. Financial forecasting employs it. It operates on discrete, real, and streaming data and is quick to classify and train.

**J48 Classification**

J48 Tree has been used to select the objective esteem based on several dataset aspects and to describe their accuracy. The outcome is based on 10 cross-validations, with a mean absolute error of 0.0227 and an accuracy of 94.10%. The J48 add-on features include decision trees, continuous attribute value ranges, and running consideration for missing values, among others. J48 is an expanded version of ID3 (Iterative Dichotomiser 3). An algorithm called ID3 is employed to create a decision tree from a dataset. The ID3 algorithm is related to the C4.5 algorithm. It is frequently employed in the fields of machine learning and natural language processing.

**IMLP classification**

A supervised learning model is used by Improved Multi-Layer Perceptron. It is employed in the training of synthetic neural networks. The algorithm operates as though mistakes come from the source nodes back to the destination nodes. The network is already aware of the anticipated outcome. A machine learning technique called ensemble learning combines multiple base models to create a single, ideal predictive model. The optimal model is produced by combining the IMLP model with K-means clustering. The k number of centroids are recognized using the K-means method. It assigns each data point to the closest cluster while maintaining a small number of centroids. Finding the data's average is what the term "means" alludes to.

**2.3 Machine learning as an efficient prediction of flood severity:**

The most common and widespread natural disasters are floods, which can have devastating effects on both human society and the environment (Qi et al., 2022). Floods considerably alter the hydro-morphodynamics of floodplain ecosystems and alter river shape, transport sediment, and leach minerals (Eslamian et al., 2018a; Moghaddam et al., 2020; Zhu et al., 2020). (Guo et al., 2021, Wang et al., 2022). Diverse ecosystems in floodplains are quick to react to flood conditions and severe hydrological disturbances (Doering et al., 2021).Flooding happens unexpectedly in time and place in arid and semi-arid regions, giving floodplain ecosystems valuable components (such water and sediment) at sporadic and irregular intervals (e.g., periodical and seasonal). In arid and semi-arid areas, river flooding of floodplains can also have an impact on bird migration, fish breeding, vegetation productivity, soil moisture, and nutrients (Eslamian et al., 2018b, Mohammadi et al., 2017, Torabi Haghighi et al., 2020). Flood mapping requires a thorough understanding of the dynamic dynamics affecting flood duration and inundation zones (Mason et al., 2014).

As they incorporate concepts and ideas from computer science, statistics, and optimization techniques, machine-learning algorithms constitute a multidisciplinary approach. They are commonly employed in remote sensing image classification to divide information from a given dataset into desired groups, typically, types of land use and land cover (Cervantes et al., 2020, Du et al., 2019, Ostad-Ali-Askari and Shayan, 2021, Tian et al., 2020, Xie and Sun, 2020). The benefits of machine learning techniques have been established by prior research, which also shows how they have helped to improve the spatiotemporal study of flood dynamics.

The application of machine learning (ML), which is based on big data science and artificial intelligence, can be a substitute for and addition to those models. Agriculture (Wolanin et al., 2020), hydrology (Qiu et al., 2017; Zhang et al., 2018), and other fields have made extensive use of machine learning (ML). Fuzzy systems, neural networks, and evolutionary algorithms are a few examples of computational intelligence techniques that have been applied to forecasting rainfall time series and flood prediction (Fotovatikhah et al., 2018). (Wu and Chau, 2013).In addition, deep learning (DL) with a multi-layer structure, such as convolutional neural networks (CNN) and recurrent neural networks (RNN), has advanced quickly in recent years and has been successfully used in crop yield prediction (Kamilaris and Prenafeta-Bold, 2018), streamflow forecasting (Ha et al., 2021), and other fields (Li et al., 2021). The studies on machine learning (ML) applications in waterlogging prediction, however, are still scarce and largely concentrate on mapping flood vulnerability (Avand et al., 2021b) or image or text identification of urban waterlogging (Liu et al., 2021b). It's crucial to use the right statistical or machine learning algorithm to guarantee precision and computational effectiveness. The method should ideally be as straightforward as it is accurate. As a result, the model is guaranteed to be interpretable, both in terms of output and in terms of understanding how different factors affect the final model.

A thorough examination of the different kind of machine learning techniques for flood prediction published in the literature is provided in (Fig. 1).

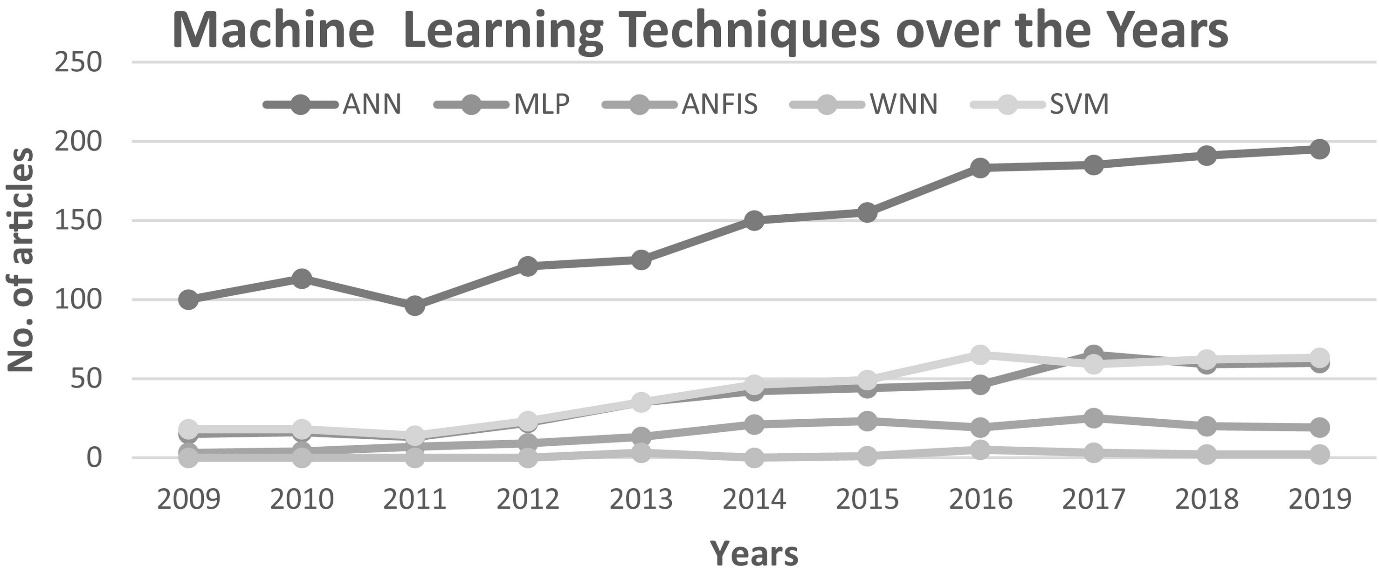


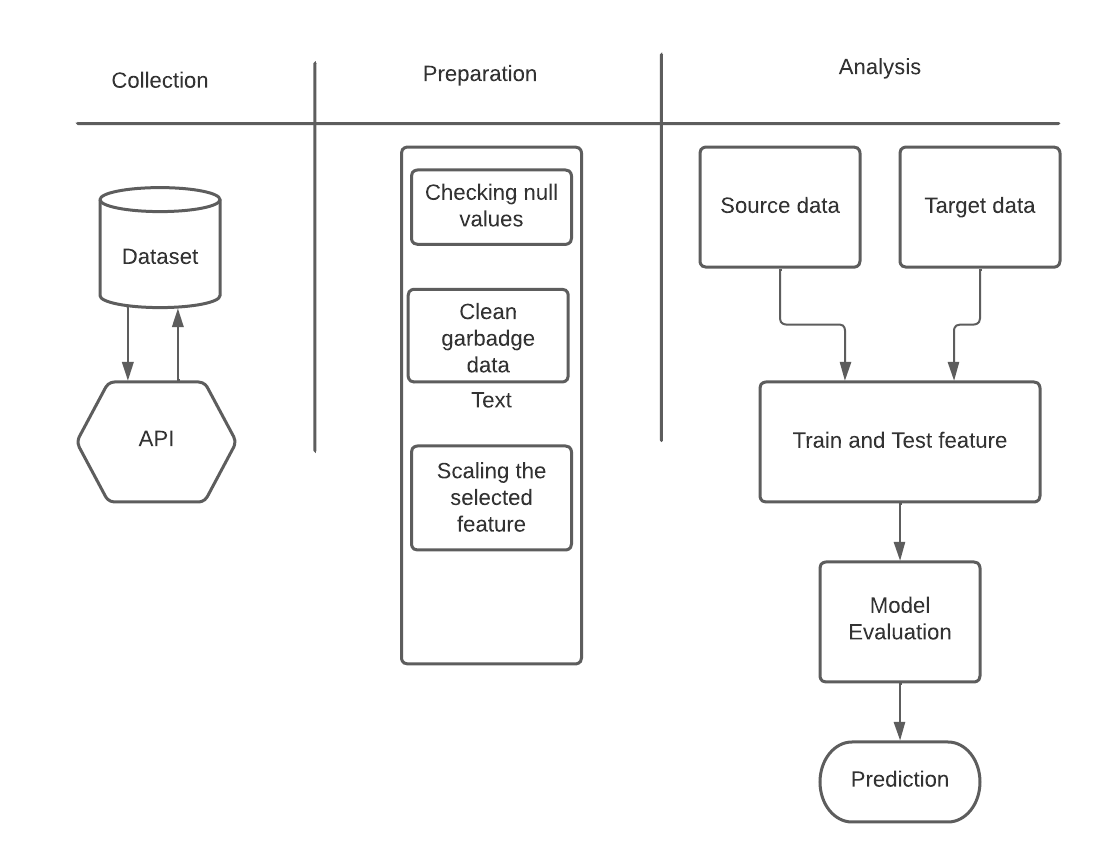
Fig. 1. Rate of use of different machine learning tools for flood prediction over the years.

**2.4 Possible ethical issues of using machine learning in flood prediction**

There are numerous advantages to using machine learning to predict floods. There are reasons to be concerned about using machine learning to predict floods. This when it comes to racial discrimination, privacy concerns and so on, this is as machine learning mainly constructed on class-based features in the datasets. This dataset could carry the climate situation happened previous year like wind flow, humidity and temperature. Although it is not directly related to the racial things but the data could be of any country, but it suppose not to be an issue.

**Chapter 3. Methodologies**

**3.1 Overview of the system development**

****

**3.1.1 Data collection**

All data for the analysis were collected from the openly available resources using best- practices. For this research live dataset are used by the help of API (Application programming Interface) technology from renowned live weather data provider. For the forecast evaluation real time data set performed better than historical data for other studies (Dean Croushore and Tom Stark, 2001). Using the wetherstack API , live data were collected based on specific area where the data is wind speed, humidity, pressure ,visibility and precipitation.

**3.1.2 Data Preparation**

Garbage was removed from the dataset before it was used by the system. For better output, it is crucial to remove duplicate and unnecessary data. This weather data includes information about wind direction that was eliminated by thinking extraneous data. For data preparation python pandas are used to preparing this dataset.

**3.1.3 Data Analysis**

On the dataset, a number of statistical significance tests were run. After preparing the dataset some features are selected by using feature selection technique. Model was created by using python panda’s code and implemented few machine learning algorithms to predict the precipitin of rain fall. In this operation, the target data wasprecipitation of rain fall and other data like humidity, visibility and pressure were train data. For better performance, feature scaling technique was used.

**3.2 Data Collection Approach**

Data are collected from the world renowned weather data source which is providing real time live weather data. Using weatherstack API (Application programming Interface) which enabled to collect data according to city such as in this dataset which is based on New York City climate information like wind pressure, humidity and cloud cover. The user will determine which city's weather information is needed. In this research the real time dataset are collected by using python API technology and all data gathered by following python JSON format. The below code is used to collect live dataset



Fig 2: python API for real time weather dataset

The code above calls the weather stack API, which sends live updated data with timestamps based on the city selection, although in this case, data for the city of "New York" is used. By refreshing the code, this dataset will be updated and the update time will be displayed. Dataset will save as csv file in the system for every update and it will contain previous data and updated data as well. For this study the several weather information is required that’s why those data information collected through JSON format. Weather stack API offering some valuable live climate information which is wind degree, fells like, cloud covering, humidity, wind pressure, temperature and precipitation.

**3.3 Theoretical background**

**3.3.1 Transfer learning**

The machine learning technique known as transfer learning is used to reuse a model that has been trained for one task for another activity. (Lisa Torrey and Jude shavlik, 2010). This approach is works for classification, regression and clustering problems with labelled data. There are three types of transfer learning which is based on problem category.

* Inductive transfer learning: it works for when source domain and target domain labels available and used for Classification and regression type problem.
* Transductive transfer learning: Describe the setting when only source domain are available and used for regression and classification problems.
* Unsupervised learning: When both source and domain data are not available. Used to complete Dimensionality reduction and Clustering task.

**3.3.2 Types of machine Learning**

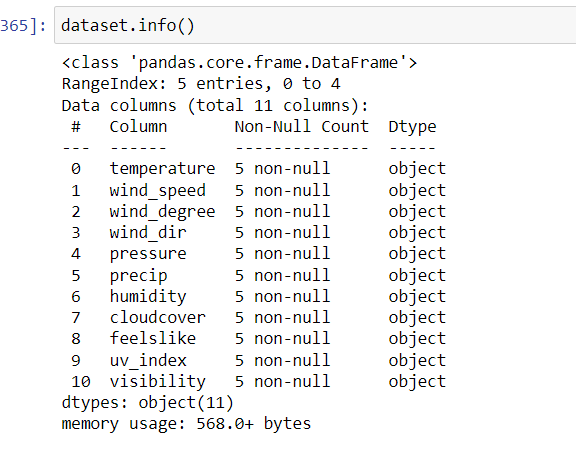
There are three type of machine learning which is

* Supervised learning: when machine has already the data or machine already trained with knowledge. Supervised learning algorithms are trained with labelled datasets. Random forest, Logistic regression and decision tress algorithms are used to predict and classifying target data. In this research flood prediction technique is based on datasets which is solved by supervised learning algorithms
* Unsupervised learning: When machine has no labelled data only input data provided to the model. Clustering and dimensionality reduction technique are used to solving machine learning problems.
* Reinforcement learning: machine can learn from feedback. When a machine receives a bad signal, it will record it as feedback or a lesson to apply to the next task. Reinforcement learning used for robotics task and resource management.

**3.4 Data Preparation**

**3.4.1 Dataset description**

The set of weather data is gathered from the weatherstack website utilising the API (Application Programming Interface), which offers extensive climatic data. Eleven columns are present. The columns include temperature, pressure, precipitation, wind speed, wind degree, wind direction, humidity, cloud cover, and visibility. The raw data will increase any single refreshment of page with timestamp. Here the dataset information is

****

**Fig 3: Weather dataset description**

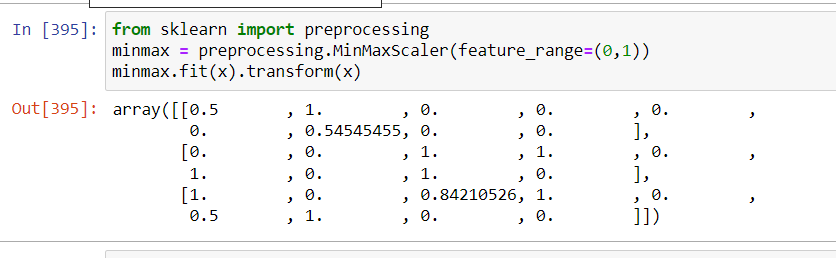
There were no null values in the provided dataset. Python's null function is used to identify null values. If any null values exist, they must be filled in with other values in order for the machine to function properly. In this dataset, each column is equally crucial for predicting the values of the target column. The feels like column contains temperature values that are actually feeling but not the actual temperature, and theprecipitation column is the percentage of the total amount of rain fall. The temperature columns contain temperature values that can be minus temperature or higher than thirty degrees depending on the weather. The information in the columns for cloud cover and visibility is nearly identical. The wind direction data is a character that describes the direction of the wind flow and this data are not an integer.

**3.4.2 Data pre-processing**

A group of techniques utilised before a data mining strategy is used is referred to as data pre-processing for data mining (Salvador Garcia, 2016). There are certain functions in Python that can be used to pre-process data. Pre-processing the data is important for better results before feeding it into algorithms. If the dataset contains any extraneous data or junk, the results will be incorrect. Before adding algorithms, the wind direction in this weather dataset was deleted because it was thought to be an unneeded feature.

**3.4.3 Feature selection**

The best features should be chosen in order to improve output predictions. Precipitation is the target data in this weather dataset whereas other elements are the source data. Using the pandas function in Python, features are chosen. For fitting this for training and testing, the features are contained in the x and y variables. After choosing the features, the scaling was done using the minmax scalar function. Here is the outcome after scaling.

Fig 4: feature scaling

**3.5 Modeling**

Models are increasingly being built from data using machine learning techniques. Such models necessitate a deep understanding of the data as well as a familiarity with the available ML tools and their features. A rigorous approach that ensures that all important aspects of the modelling process are taken into account can help you build high-quality models. One of the most vital steps in the design process is reviewing models, regardless of their purpose.. Study will evaluate whether the stated goal of the modelling effort has been achieved. It allows for the comparison of different modelling philosophies and directs further research.. In this research the regression analysis methods will use for predicting the target value. Regression is good fit for predicting future values and it also works for multivariable’s (Eduardo Núñez, Ewout W. Steyerberg, 2011). Regression is useful for identifying trends in data, and it has various techniques that can be used to forecast precipitation from weather datasets. Which algorithm is perfect and which result is accurate is determined by the correctness of the results from various algorithms. Here is the mathematical equation for measuring accuracy.

Here TP means true positive value, TN means true negative value and FN means false positive value. It is necessary to compare the predictive accuracy of the algorithms used for evaluation of results. For finding best methods, some regression algorithms such as linear regression, logistic regression, Support vector regression and Gaussian nb applied to the given dataset and got some result. Here is the accuracy result of some algorithms

|  |  |
| --- | --- |
| Algorithms | Accuracy |
| Linear Regression | 0.6156 |
| Logistic regression | 0.7267 |
| Support vector machine | 0.5678 |
| Gaussian nb | 0.6045 |
| Polynomial regression | 0.5989 |

Table 1: Accuracy result of some algorithms

Regression problems are best solved by applying all machine learning techniques. Scikit Learning technology was used to implement these algorithms.

**Chapter 4. Result and Discussion**

In the above accuracy table, got outstanding result from different approaches among them logistic regression performed better for this real time weather dataset. Using deep learning technology flood are predicted using historical data through predicting Indian rainfall, here area are classified by risk area ([Suresh Sankaranarayanan](javascript:;); [Malavika Prabhakar](javascript:;), 2020), But it’s not effective for predicting current rainfall amount. Machine learning technique are used on hydrological modelling to solve urban flooding in thiland and applied five machine learning algorithms which measured water level for forecasting flood disease (Jeerana Noymanee, Thanaruk Theeramunkong, 2019), but it has some limitation where the result will deploy lately to the authority. The logistic regression used generally for dependent variables but in some cases it also worked for other categories such as multinomial and polytomous variables (L. G. Grimm, 1995). Support vector machine also effective to use predicting the data trend and performed better for text categorization problems (M. A. Hearst, S. T. Dumais, E. Osuna, 1998) but in this dataset it is not performed better. \support vector machine performed better for historical data. To tackle this problem it is declared that machine learning technology is performed better than using hydrological approaches (Marcel Motta, Miguel de Castro Neto, 2021) where machine can calculate result depend on data and can give proper output. Whatever others researcher got different advantageous outputs by using other approaches, but in my opinion, it is better to use machine learning methods and collect live data instead of historical data. Few researchers classified the area according to climate data and predicted which area is in danger zone. In this modern era, advanced technologies are using to getting better output. In this research the modern technique are used to getting data and applied advanced machine learning methods for getting better result. Here user can get live update of precipitation rate and can measure the flood happening chances. Despite of some limitations this strategy better than other old process of flood prediction.

**Chapter 5. Conclusion and Future Work**

Flood prediction technique is Breakthrough step for this generation as we see most of the countries are suffering loses every year and its impact on global economy. Weather data is gathered and advanced technological methods are used to solve this problem. It has been pre-processed using contemporary programming languages and machine learning techniques. To address this issue in real life, machine learning techniques are utilised and produced the greatest results.

On the other hand, other machine learning algorithms which are fit for solving regression problem will be applied to this dataset and will compare those for better accuracy. A portable web application will develop to show the precipitation percentage with proper user interface so that it will be usable for every people. Other advance pre-processing technique will be applied to prepare the dataset for making model. This research project can be further enhanced as future work by considering the adding timestamp for every prediction.

**Refference:**

Mosavi, Amir & Ozturk, Pinar & Chau, Kwok. (2018). Flood Prediction Using Machine Learning Models: Literature Review. Water. 10. 1536. 10.3390/w10111536.

B. Maschler and M. Weyrich, "Deep Transfer Learning for Industrial Automation: A Review and Discussion of New Techniques for Data-Driven Machine Learning," in IEEE Industrial Electronics Magazine, vol. 15, no. 2, pp. 65-75, June 2021, doi: 10.1109/MIE.2020.3034884.

A. Singh, N. Thakur and A. Sharma, "A review of supervised machine learning algorithms," 2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom), New Delhi, India, 2016, pp. 1310-1315.

García, S., Ramírez-Gallego, S., Luengo, J. *et al.* Big data preprocessing: methods and prospects. *Big Data Anal* **1**, 9 (2016).

Lisa Torrey and Jude Shavlik , (2010). Machine learning Applications and trends.

Dallas Thornton, Roland M. Mueller, 2013. Predicting Healthcare Fraud in Medicaid: A Multidimensional Data Model and Analysis Techniques for Fraud Detection.

Yoram Reich, S.V. Barai, 1999. Evaluating machine learning models for engineering problems.

Eduardo Núñez, Ewout W. Steyerberg, 2011. Regression Modeling Strategies.

Suresh Sankaranarayanan, Malavika Prabhakar, Sreesta Satish, Prerna Jain, Anjali Ramprasad, Aiswarya Krishnan; Flood prediction based on weather parameters using deep learning. Journal of Water and Climate Change 1 December 2020; 11 (4): 1766–1783.

Jeerana Noymanee, Thanaruk Theeramunkong, 2019. Flood Forecasting with Machine Learning Technique on Hydrological Modeling.

Wright, R. E. (1995). Logistic regression. In L. G. Grimm & P. R. Yarnold (Eds.), Reading and understanding multivariate statistics (pp. 217–244). American Psychological Association.

M. A. Hearst, S. T. Dumais, E. Osuna, J. Platt and B. Scholkopf, "Support vector machines," in IEEE Intelligent Systems and their Applications, vol. 13, no. 4, pp. 18-28, July-Aug. 1998, doi: 10.1109/5254.708428.

S. Abraham, J. V R, S. Thomas and B. Jose, "Comparative Analysis of Various Machine Learning Techniques for Flood Prediction," 2022 International Conference on Innovative Trends in Information Technology (ICITIIT), 2022, pp. 1-5, doi: 10.1109/ICITIIT54346.2022.9744177.

H. Riza, E. W. Santoso, I. GunawanTejakusuma and F. Prawiradisastra, "Advancing Flood Disaster Mitigation in Indonesia Using Machine Learning Methods," 2020 International Conference on ICT for Smart Society (ICISS), 2020, pp. 1-4, doi: 10.1109/ICISS50791.2020.9307561.

P. Ghorpade et al., "Flood Forecasting Using Machine Learning: A Review," 2021 8th International Conference on Smart Computing and Communications (ICSCC), 2021, pp. 32-36, doi: 10.1109/ICSCC51209.2021.9528099.

Yue Ma, Xiaomeng Liu, Xiaojuan Li, Yonghua Sun and Xiaomeng Li, "Rapid assessment of flood disaster loss in Sind and Punjab province, Pakistan based on RS and GIS," 2011 International Conference on Multimedia Technology, 2011, pp. 646-649, doi: 10.1109/ICMT.2011.6003024.

Amir Mosavi, PinarOzturk, “Flood Prediction Using Machine Learning Models: literature Review”, 27 October 2018.

Tom Sharp, “An introduction of support vector regression (SVR)”, 3 march 2020.

KatarzynaAlderman,Shilu Tong “Flood and human health: A systematic review” Environmental International, 2012.

Samuel Brody, Wesley E. Highfield, Russell Blessing, Understanding the impacts of the built environment on flood loss, Coastal Flood Risk Reduction, 10.1016/B978-0-323-85251-7.00013-5, (167-176), (2022).

C. Z. Basha, N. Bhavana, P. Bhavya and S. V, "Rainfall Prediction using Machine Learning & Deep Learning Techniques," 2020 International Conference on Electronics and Sustainable Communication Systems (ICESC), 2020, pp. 92-97, doi: 10.1109/ICESC48915.2020.9155896.

R. Gowthamani, K. Sasi Kala Rani, S.R. Abishek, G. Akash, A. Kavin Kumar, Efficient detection and prediction of flood severity using machine learning algorithm, Materials Today: Proceedings 47 (2021) 376–380.

J. Men, G. Xu, Z. Han, Z. Sun, X. Zhou, W. Lian, X. Cheng, Finding sands in the eyes: Vulnerabilities discovery in IoT with EUFuzzer on human machine interface, IEEE Access 7 (2019) 103751–103759.

A. Moraru, M. Pesko, M. Porcius, C. Fortuna, D. Mladenic, Using machine learning on sensor data, J. Comput. Inf. Technol. 18 (4) (2010) 341, https://doi. org/10.2498/cit.1001913.

E. Borgia, The Internet of Things vision: Key features applications and open issues, Comput. Commun. 54 (2014) 1–31.

A.Y. Sun, B.R. Scanlon, How can big data and machine learning benefit environment and water management: A survey of methods, applications and future directions, Environ. Res. Lett. 14 (7) (2019) 28.

W. Qi, C. Ma, H. Xu, K. Zhao, Z. Chen, A comprehensive analysis method of spatial prioritization for urban flood management based on source tracking, Ecol. Indic., 135 (2022), Article 108565, [10.1016/j.ecolind.2022.108565](https://doi.org/10.1016/j.ecolind.2022.108565).

Eslamian, S., Parvizi, S., Ostad-Ali-Askari, K., Talebmorad, H., 2018a. Water, in: Bobrowsky, P.T., Marker, B. (Eds.), Encyclopedia of Engineering Geology, Encyclopedia of Earth Sciences Series. Springer International Publishing, Cham. doi: 10.1007/978-3-319-12127-7.

D.D. Moghaddam, O. Rahmati, M. Panahi, J. Tiefenbacher, H. Darabi, A. Haghizadeh, A.T. Haghighi, O.A. Nalivan, D. Tien Bui, The effect of sample size on different machine learning models for groundwater potential mapping in mountain bedrock aquifers, Catena, 187 (2020), p. 104421.

L. Zhu, L. Huang, L. Fan, J. Huang, F. Huang, J. Chen, Z. Zhang, Y. Wang, Landslide susceptibility prediction modeling based on remote sensing and a novel deep learning algorithm of a cascade-parallel recurrent neural network, Sensors (Switzerland), 20 (2020), p. 1576, [10.3390/s20061576](https://doi.org/10.3390/s20061576).

L. Guo, C. Zhu, W. Xie, F. Xu, H. Wu, Y. Wan, Z. Wang, W. Zhang, J. Shen, Z.B. Wang, Q.He, Changjiang Delta in the Anthropocene: Multi-scale hydro-morphodynamics and management challenges, Earth-Sci. Rev., 223 (2021), Article 103850, [10.1016/j.earscirev.2021.103850](https://doi.org/10.1016/j.earscirev.2021.103850).

J. Wang, Z. Dai, S. Fagherazzi, X. Zhang, X. Liu, Hydro-morphodynamics triggered by extreme riverine floods in a mega fluvial-tidal delta, Sci. Total Environ., 809 (2022), Article 152076, [10.1016/j.scitotenv.2021.152076](https://doi.org/10.1016/j.scitotenv.2021.152076).

M. Doering, R. Freimann, N. Antenen, A. Roschi, C.T. Robinson, F. Rezzonico, T.H.M. Smits, D. Tonolla, Microbial communities in floodplain ecosystems in relation to altered flow regimes and experimental flooding, Sci. Total Environ., 788 (2021), Article 147497, [10.1016/j.scitotenv.2021.147497](https://doi.org/10.1016/j.scitotenv.2021.147497).

Eslamian, S., Sayahi, M., Ostad-Ali-Askari, K., Basirat, S., Ghane, M., Matouq, M., 2018b. Saturation, in: Bobrowsky, P.T., Marker, B. (Eds.), Encyclopedia of Engineering Geology. pp. 1–2. doi: 10.1007/978-3-319-12127-7\_251-1.

A. Mohammadi, J.F. Costelloe, D. Ryu, Application of time series of remotely sensed normalized difference water, vegetation and moisture indices in characterizing flood dynamics of large-scale arid zone floodplains, Remote Sens. Environ., 190 (2017), pp. 70-82, [10.1016/j.rse.2016.12.003](https://doi.org/10.1016/j.rse.2016.12.003).

A. TorabiHaghighi, M. Sadegh, S. BehroozKoohenjani, A.A. Hekmatzadeh, A. Karimi, B. Kløve, The mirage water concept and an index-based approach to quantify causes of hydrological changes in semi-arid regions, Hydrol. Sci. J., 65 (2020), pp. 311-324, [10.1080/02626667.2019.1691728](https://doi.org/10.1080/02626667.2019.1691728).

D.C. Mason, L. Giustarini, J. Garcia-Pintado, H.L. Cloke, Detection of flooded urban areas in high resolution Synthetic Aperture Radar images using double scattering, Int. J. Appl. Earth Obs. Geoinf., 28 (2014), pp. 150-159, [10.1016/j.jag.2013.12.002](https://doi.org/10.1016/j.jag.2013.12.002).

J. Cervantes, F. Garcia-Lamont, L. Rodríguez-Mazahua, A. Lopez, A comprehensive survey on support vector machine classification: Applications, challenges and trends, Neurocomputing, 408 (2020), pp. 189-215, [10.1016/j.neucom.2019.10.118](https://doi.org/10.1016/j.neucom.2019.10.118).

Y. Du, Y. Fan, X. Liu, Y. Luo, J. Tang, P. Liu, Multiscale cooperative differential evolution algorithm, Comput. Intell. Neurosci., 2019 (2019), pp. 1-17, [10.1155/2019/5259129](https://doi.org/10.1155/2019/5259129).

K. Ostad-Ali-Askari, M. Shayan, Subsurface drain spacing in the unsteady conditions by HYDRUS-3D and artificial neural networks, Arab. J. Geosci., 14 (2021), p. 1936, [10.1007/s12517-021-08336-0](https://doi.org/10.1007/s12517-021-08336-0).

Y. Tian, C. Chen, X. Chen, Q. Zhang, R. Sun, Research on real-time analysis technology of urban land use based on support vector machine, Pattern Recognit. Lett., 133 (2020), pp. 320-326, [10.1016/j.patrec.2020.03.022](https://doi.org/10.1016/j.patrec.2020.03.022).

X. Xie, S. Sun, General multi-view semi-supervised least squares support vector machines with multi-manifold regularization, Inf. Fusion, 62 (2020), pp. 63-72, [10.1016/j.inffus.2020.04.005](https://doi.org/10.1016/j.inffus.2020.04.005).

Marcel Motta, Miguel de Castro Neto, Pedro Sarmento, A mixed approach for urban flood prediction using Machine Learning and GIS, International Journal of Disaster Risk Reduction 56 (2021) 102154.

LeCun, Y., Bengio, Y. & Hinton, G. Deep learning. *Nature* **521**, 436–444 (2015). https://doi.org/10.1038/nature14539