Project Report

Monthly Rainfall Measurement Prediction using Machine learning

Submitted by

Dhendup Ghishing(12190048)

Dorji Yangzom(12190050)

Kushal Chetrii(12190064)

Sourav Rai(12190085)

Under the guidance of:

Mr. Yonten Jamtsho

GYALPOZHING COLLEGE OF INFORMATION TECHNOLOGY

MONGAR, BHUTAN, 2022

ROYAL UNIVERSITY OF BHUTAN

GYALPOZHING COLLEGE OF INFORMATION TECHNOLOGY

CERTIFICATE

This is to certify that the BSc.IT project report titled “Monthly Rainfall Measurement Prediction using Machine learning”, which is being submitted by Dorji Yangzom(12190050), Dhendup Ghishing(12190048), Kushal Chetrii(12190064), Sourav Rai(12190085), the students of Bachelors of Science in Information Technology, prepared during the academic year 2022 in partial fulfillment of the requirement for the award of the degree of Bachelor of Science in Information Technology is a record of the students’ work carried out at the Gyalpozhing College of Information Technology, Royal University of Bhutan, Gyalpozhing under my supervision and guidance.

Mr. Yonten Jamtsho

(Project Guide)

BSc.IT

Gyalpozhing College of Information Technology

ACKNOWLEDGMENT

Any achievement, be it scholastic or otherwise does not depend solely on the individual efforts but on the guidance, encouragement and cooperation of intellectuals, elders and friends. A number of personalities, in their own capacities have helped us in carrying out this project. We would like to take an opportunity to thank them all.

This work would not have been possible without the constant supervision and guidance of our guide Mr. Yonten Jamtsho who stimulated encouragement and rendered immeasurable help during the entire project in 2022.

We are thankful to College of Information Technology for providing us with such a good course and letting us to develop a wonderful website integrated with machine learning algorithms and in addition to we are more grateful to National Hydrology and Meteorology Center for granting with a dataset.

Moreover, we are immensely thankful to Panel members for guiding us and reviewing progress with constructive feedback and valuable suggestion which helped us a lot while working with the project.

We are also indebted to Ms. Rebecca Tirwa for coordinating the project reviews providing necessary advice and material.Last, but not the least, We would like to thank our peers and friends who provided us with valuable suggestions to improve our project.

ABSTRACT

Rainfall prediction is one of the challenging tasks in weather forecasting. Accurate and timely rainfall prediction can be very helpful to take effective security measures in advance regarding: on-going construction projects, transportation activities, agricultural tasks, flight operations and flood situation, etc. Machine learning techniques can effectively predict the rainfall by extracting the hidden patterns among available features of past weather data. This project contributes by providing a critical analysis and review of machine learning techniques, used for rainfall prediction using the past data. This review will serve the researchers to analyze the latest work on rainfall prediction with the focus on machine learning techniques and also will provide a baseline for future directions and comparisons.

ABBREVIATIONS

Term	**Abbreviation**
ML	**Machine Learning**
SVM	**Support Vector Machine**
ANN	**Artificial Neural Network**
KNN	**K-nearest Neighbors**
MSE	**Mean Squared Error**
RMSE	**Root Mean Squared Error**
Tmax	**Maximum Temperature**
Tmin	**Minimum Temperature**

Table 1: Abbreviation table

LIST OF TABLES

Table 1: Abbreviation table
Table 2: Dataset information
Table 3: Train-test accuracy
Table 4: MSE
Table 5: RMSE

LIST OF FIGURES

Figure 1: System Architecture
Figure 2: UseCase Diagram
Figure 3: Activity Diagram
Figure 4:Sequence Diagram
Figure 5: Train-test Comparison
Figure 6: MSE Comparison
Figure 7: RMSE Comparison
Figure 8:Home
Figure 9: About
Figure 10: Team
Figure 11: Form
Figure 12: Dashboard

ACKNOWLEDGMENT 3

ABSTRACT 4

ABBREVIATIONS 5

LIST OF TABLES 6

LIST OF FIGURES 6

CHAPTER 1 9

BACKGROUND AND INTRODUCTION 9

1.1. Background and Introduction 10
1.2 Problem Statement 11
1.3 Aim 11
1.4 Goals of the project 11
1.5 Objectives 12
1.6 Scope 12

CHAPTER 2 13

LITERATURE SURVEY 13

3.1 Domain based Literature Survey 14
3.1 Algorithm based Literature Survey 16

CHAPTER 2 17

PROJECT REQUIREMENTS 17

2.1 Requirements 18
Python 18
Pandas 18
Seaborn 18
Matplotlib 19

CHAPTER 4 21

DESIGN AND METHODOLOGY 21

4.1 System Architecture 22
4.2 System Design 23
Preliminary Design 23
Use case Diagram 24
Activity Diagram 25
Sequence Diagram 25
4.3 Methods and Algorithms 27

Data collection 27
Data preprocessing 27
Machine Learning Algorithms 28
Multiple regression 28
Random Forest(RF) 28
Gradient Boosting Algorithm 29
KNN (K- Nearest Neighbors) Algorithm 30
Decision Tree 30
Artificial Neural Networks(ANN) 31
Support Vector Machine 31

CHAPTER 5 32

RESULT AND DISCUSSION 32

5.1 Result 33
Algorithm comparison based on Train Test Accuracy 33
Algorithm comparison based on MSE 34
Algorithm comparison based on RMSE 36
5.2 Discussion 37

CHAPTER 6 38

USER INTERFACE 38

6.1 Website overview 39

CHAPTER 7 43

CONCLUSION AND FUTURE WORK 43

REFERENCES 45

CHAPTER 1

BACKGROUND AND INTRODUCTION

1.1. Background and Introduction

Rainfall is one the most significant atmospheric occurrence that is not only useful for the environment itself but for all the living beings on the earth. It affects everything directly or indirectly and because it is one of the most important natural phenomena; it is also important for human beings to ponder on how precipitation changes with the change in climate. The rainfall helps in balancing the increasing temperature and in the survival of human beings . The increasing temperature of the world is associated with global warming and water is one of the scarce and most useful resources which in the result of this increasing temperature are evaporating from the reserves. Rainfall is also compensation to all these reserves and it is necessary for agriculture and its production as well.

This study is focusing on predicting monthly rainfall using machine learning over some identified places in bhutan. The rainfall prediction will not just assist in analyzing the changing patterns of rainfall but it will also help in organizing the precautionary measures in case of disaster and its management. The rainfall prediction would also assist in planning the policies and strategies to deal with the increasing global issue of ozone depletion. The rainfall prediction and weather updates not only help in managing the macro-level problems like flood and agricultural issues because of poor or extreme rainfall (Lima & Guedes, 2015). The rainfall prediction could also contribute to the well-being and comfort of the people by keeping them informed by tracking the rainfall patterns and predicting the rainfall using machine learning. The rainfall predictions help people to deal with hot and humid weather. The technological development in the modern world has expanded the space for innovation and revolution. Although the issues concerned are probably associated with these technological advancements, one needs to consider the range of possibilities and opportunities that this technological evolution has opened to human beings.

Over the past few decades, major landslide events and water shortage issues were reported over the places in Bhutan. In order to overcome those issues and to reach out information beforehand we are going to develop a model using machine learning(python) where rainfall can be predicted precisely and accurately by studying the datasets of the past 5 years.

Bhutan is a small Himalayan country landlocked between India to the south and China to the north. The sensor data used in this project were collected from a weather station located in various places in Bhutan. The data for the study will be obtained from the National Center For Hydrology and Meteorology. The dataset provides the intensity of rainfall received by specified places during the period of 2010 - 2021 and the dataset has records of rainfall received daily, monthly and yearly, we will use the records of monthly rainfall with 4 different parameters. These parameters had either zero or very few missing values that will be handled during data preprocessing. The weather parameters were extracted from monthly records by taking the mean of tmax(℃), tmin(℃), relative_humidity(%), Rainfall(mm), and wind speed(m/s).

1.2 Problem Statement

The accurate and precise rainfall prediction is still lacking which could assist in diverse fields like agriculture, water reservation, and flood prediction. The issue is to formulate the calculations for the rainfall prediction that would be based on the previous findings and similarities and will give the output predictions that are reliable and appropriate. The imprecise and inaccurate predictions are not only a waste of time but also the loss of resources and lead to inefficient management of crises like poor agriculture, poor water reserves, and poor management of floods. It also helps to solve problems related to road transportation and aviation as the travelers will be informed accurately and efficiently. Therefore, the need is not to formulate only a rainfall predicting system but also a system that is more accurate and precise as compared to the existing rainfall predictors.

1.3 Aim

The aim of the project is to study the features for the prediction of monthly rainfall using machine learning algorithms.

1.4 Goals of the project

The main purpose of the precipitation forecasting model is to predict the amount of precipitation in a specific well or division in advance, using various regression techniques and determining which one is best for precipitation forecasting. This model also helps the agricultural farmer to decide the crop, helps the watershed department with water storage, and also helps to analyze the groundwater level, as well as to predict the current road conditions for the mobile vehicles and aviation weather.

1.5 Objectives

Rainfall Prediction Model has a main objective in prediction of the amount of rain in a specific well or division in advance by using various regression techniques and finding out which one is best for rainfall prediction.

1.6 Scope

a) System scope:

The project is capable of using the maximum, minimum and average temperature, to predict the monthly rainfall.

b) User scope:

The scope is set to some identified places of Bhutan.

CHAPTER 2

LITERATURE SURVEY

3.1 Domain based Literature Survey

1) Monthly Rainfall Prediction Using Various ML Algorithms for Early Warning of Landslide Occurrence

Majority of landslides are precipitated by means of extended or heavy Rainfall forecasting enabled in figuring out the precipitation situations accountable for landslide incidence. linear regression, returned propagation neural network (BPNN), network (LSTM) used to forecast rainfall that can be compared with the rainfall thresholds to expect landslide rainfall information acquired from Narendra Nagar, a small city in Uttarakhand. The proposed algorithms use pre-fashions to have the capacity to expect rainfall depth one month estimating rainfall and therefore predicting the opportunity of landslide incidence properly earlier. The study concludes that the BPNNs are able to outperform and provide optimal inferences stating the aptness of artificial neural networks (ANNs) in estimating rainfall and hence predicting the possibility of landslide occurrence well in advance. The study is conducted explicitly for regions highly vulnerable to landslides near Narendra Nagar but may be implemented to any landslide prone area.

Natural calamities like landslides cause major human casualties and severe damage to infrastructure and natural resources. Preventing landslides is beyond the human
capabilities but their impact can be minimized if they can be predicted prior to their occurrence.producing optimal forecast.machine learning algorithms for predicting monthly rainfall that monthly forecast are more accurate than weekly or daily forecast when compared with the actual rainfall data. Hence, the daily rainfall data was converted into monthly months' rainfall data with the past three years' average forecast monthly rainfall well in advance. .Based on the results obtained, it can be inferred that BPNN is the best machine learning algorithm for forecasting rainfall followed by LSTM. Coupling previous three months' rainfall data with the past three years' average rainfall of the targeted month aided the network in between the forecasted rainfall intensity and the optimum predicting the occurrence of rainfall-induced landslides. The study aims at developing a feasible machine learning model for forecasting rainfall which can be used for predicting rainfall-induced landslides. models are capable of predicting low as well as medium-intensity rainfalls effectively, however under performed in

mapping high-intensity rainfalls accurately. The predictive accuracy of the models can be increased by introducing other input variables such as humidity, temperature, wind speed etc. The study is conducted specifically for Narendra Nagar region of Uttarakhand but can be generalized to any area vulnerable to rainfall-induced landslides.

2) Deep BLSTM-GRU Model for Monthly Rainfall Prediction: A Case Study of Simtokha, Bhutan

Rainfall prediction is an important task due to the dependence of many people on it. In this study, we carry out monthly rainfall prediction over Simtokha, a region in the capital of Bhutan, Thimphu. and the Meteorology Department (NCHM) of Bhutan. Memory (LSTM), Gated Recurrent Unit (GRU), and Bidirectional Long Short Term Memory (BLSTM) this paper proposes a BLSTM-GRU based model which outperforms the existing machine and deep From the six different existing models under study, LSTM recorded the best Mean The proposed BLSTM-GRU model outperformed LSTM by 41.1% model can achieve lower MSE in rainfall prediction systems. The study of deep learning methods for rainfall prediction is presented in this paper, and a BLSTM-GRU based model is proposed for rainfall prediction over the Simtokha region in Thimphu, 2 value of 0.50), which is widely used for rainfall prediction, did not perform well in comparison to the recent deep learning models on weather station data. 1024 neurons performed better than the others, with an MSE score of 0.013, a correlation value of 0.90, Furthermore, the proposed model presented an improved correlation value of 0.93 and R. Predicting actual rainfall values has become more challenging due to the changing weather patterns. In the future, we aim to improve the performance of our prediction model by incorporating patterns of global and regional weather such as sea surface temperature, global wind circulation, etc. We also intend to explore the predictive use of climate indices and study the effects of climate change on rainfall patterns.

3.1 Algorithm based Literature Survey

1) Prediction of Rainfall using Machine Learning Algorithms for Different Districts of Meghalaya.

In Prediction of Rainfall using Machine Learning Algorithms for Different Districts of Meghalaya byShabbir Ahmed Osmani , Foysol Mahmud , and Md. Abu Zafor , Machine learning algorithms have been used to find fine and smooth prediction models for different time series and the efficiency of the models are varying depending on the trends and unusual shapes of the dataset. This study has focused on a case study of rainfall and other climate parameters of Meghalaya state due to its versatile climate and geographical position in the world. To build a perfect machine learning model for prediction of rainfall in Meghalaya, total five machine learning algorithms, i.e. Linear Regression (LR), Regression Trees (RT), Gaussian Process Regression (GPR), Support Vector Machines (SVM) and Ensembles of Trees (ET) were used. The performance of each model was assessed on the basis of root mean squared error (RMSE). The study also checked the optimization of the parameters as predictors to impart in the training dataset. The results found LR, SVM and GPR as very good result-oriented models for almost all datasets of different districts of Meghalaya. However, LR and SVM algorithms have most effective predictions in almost all cases. The analyses found wet day frequency as the most sensitive and efficient single parameter as predictor which might produce the prediction models to operate the study for a better prediction.

CHAPTER 3

PROJECT REQUIREMENTS

2.1 Requirements

a) Hardware Requirement

CPU/GPU

b) Software Requirement

VS Code, Google Colab

2.2 Technology

Python

Python is a high-level, general-purpose programming language designed by Guido Van Rossum. Its design philosophy emphasizes code readability with the use of significant indentation. Its language constructs an object-oriented approach to help programmers write clear,logical code for small and large scale projects. This project also uses Python programming language in collaboration with Django web framework.

Pandas

Panda is a Python package which provides data structure designed to make working with relational or labeled data both easy and intuitive. In this project, pandas would be used for analyzing and manipulating the tabular data set of rainfall in identified places of Bhutan. Additionally panda can be used in handling missing values, inserting and deleting data, intelligent slicing, indexing, and reshaping

Seaborn

Seaborn is a library for making statistical graphics in python. It builds on top of matplotlib and integrates closely with pandas data structures. Seaborn helps explore and understand the data. Its plotting operates on data frames and arrays containing whole datasets and internally performs the necessary semantic mapping and statistical aggregation to produce informative plots.

Scikit-learn

Scikit-learn is a software machine learning library for python programming language. It features various classification, regression and clustering algorithms including support-vector machine, random forest, etc.

Matplotlib

It is a cross-platform used for data visualization and graphical plotting library for Python and it is imported as NumPy. It will be used to visualize the data of rainfall for better understanding. This would make things easy and helps to create quality plots while doing the project

CHAPTER 4

DESIGN AND METHODOLOGY

4.1 System Architecture

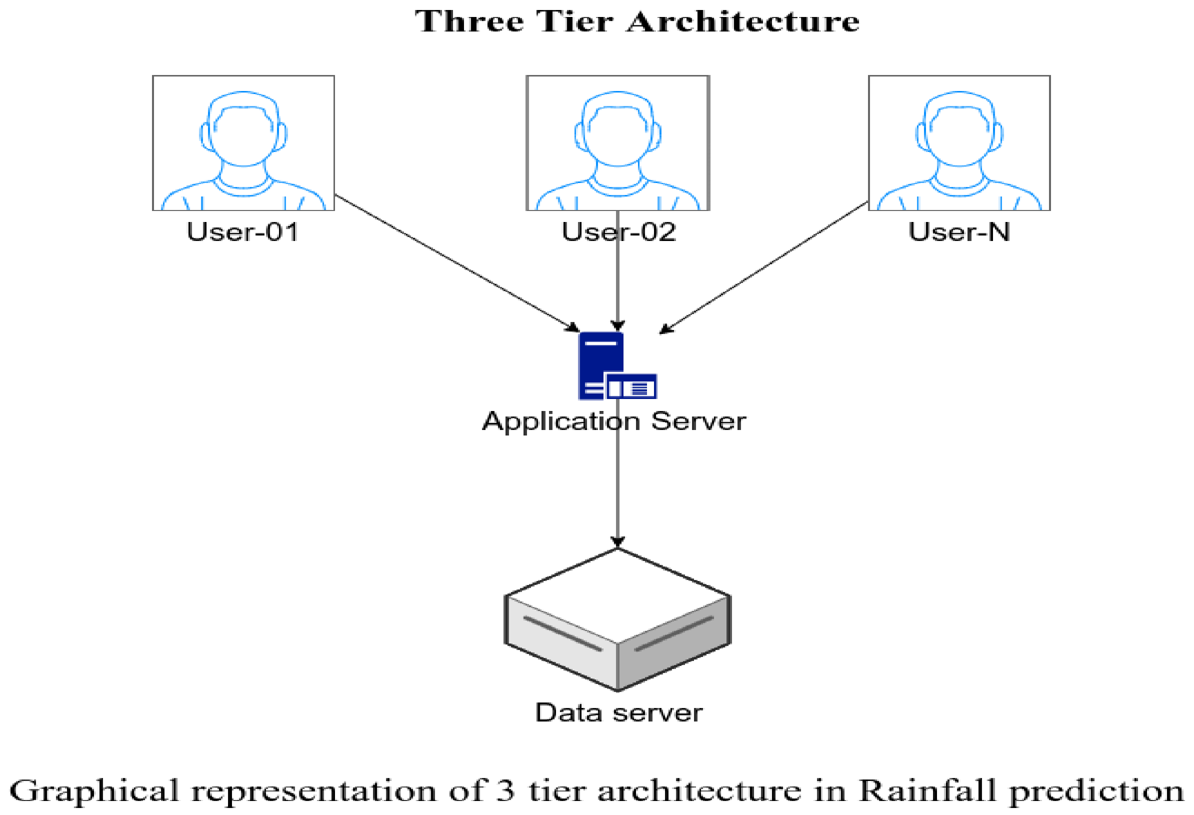


Figure 1: System Architecture

Three-tier architecture is a well-established software application architecture that organizes applications into three logical and physical computing tiers: the presentation tier, or user interface of this project will allow the users to view to rainfall prediction of particular duration; the application tier, where data is processed and it acts as a bridge between the client and the trained rainfall model; and the data tier, where the data associated with the application will be managed from that model.

The chief benefit of three-tier architecture is that because each tier runs on its own infrastructure, each tier can be developed simultaneously by a separate development team, and can be updated or scaled as needed without impacting the other tiers.

Presentation tier

The presentation tier is the user interface and communication layer of the application, where the end user interacts with the application. Its main purpose is to display information to and collect information from the user. This top-level tier can run on a web browser, as desktop application, or a graphical user interface (GUI).

When the user enters the website they will be able to choose a particular month and a year, after that they will be guided to a page which shows the rainfall level.

Application tier

The application tier, also known as the logic tier or middle tier, is the heart of the application. In this tier, information collected in the presentation tier is processed - sometimes against other information in the data tier - using business logic, a specific set of business rules. The application tier can also add, delete or modify data in the data tier.

Data tier

The data tier, sometimes called database tier, data access tier or back-end, is where the information processed by the application is stored and managed. In a three-tier application, all communication goes through the application tier. The presentation tier and the data tier cannot communicate directly with one another.

4.2 System Design

Preliminary Design

System development is done in many different ways. It forms the basis of all methodologies. The approach that is being implemented for this project is a structured approach. Structured programming, structured analysis, structured design are the techniques for structured approach. This is implemented for this system development.

Structured analysis defines system-processing requirements by identifying all of the events that will cause a system to react in some way. Each event leads to a different system activity. The most important development activity is preparation of computer programs needed for the system. The system flowcharts, input charts, output charts, are transferred into the program. In each stage of preparation, the program has been tested and errors are corrected if any. All accuracy measures fall into account while testing the program.

Use case Diagram

A use case models the functionality of a system as perceived by outside agents, called actors that interact with the system from a particular viewpoint. Its main purpose is to help the development team visualize the functional requirements of a system, including the relationship of actors to essential processes and also among different use cases.

In the monthly rainfall prediction use case, the actor/user can view web information. The actor can also fill up the form/provide the parameters(place, month, year, tmax, tmin, humidity, rainfall and wind speed) if he/she wants to check the rainfall. After providing the data, validation will be done. if the provided data are correct then the data are taken as an input for a model for prediction. If the provided data is incorrect then the user can reenter the data once again. Once the data is sent to the model then the model will generate the prediction, where user can view the prediction as shown in figure.

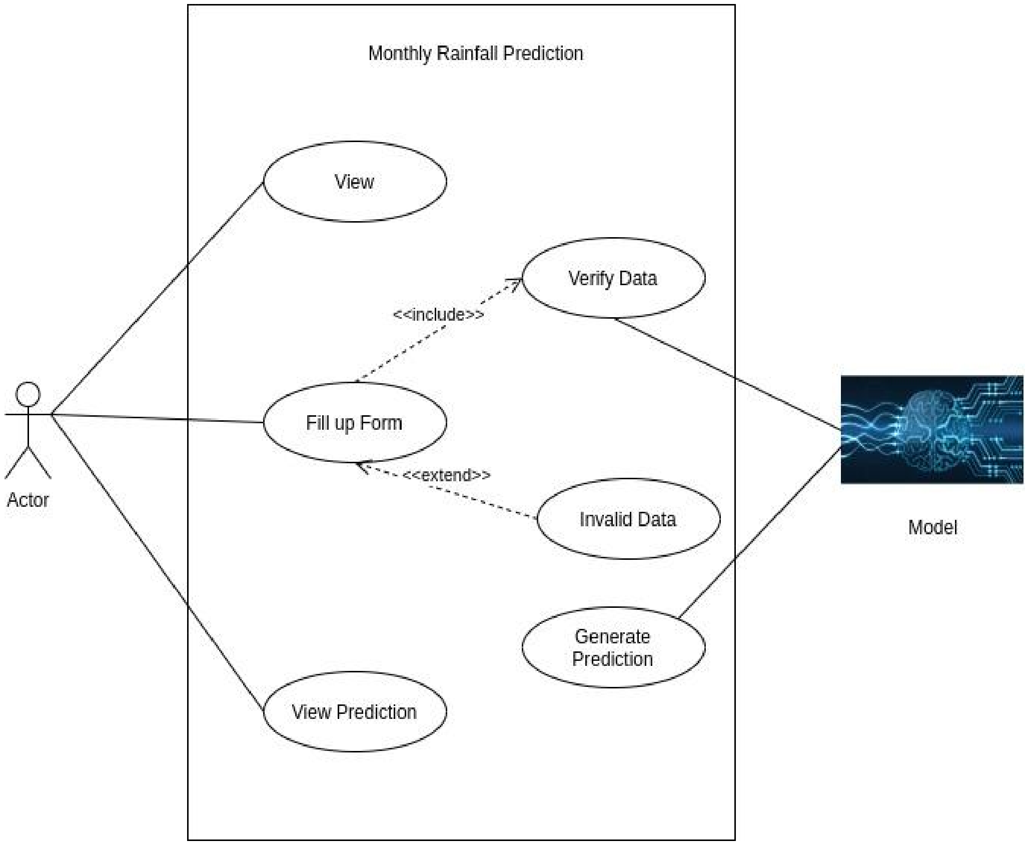


Figure 2: UseCase Diagram

Activity Diagram

Activity Diagrams to illustrate the flow of control in a system and refer to the steps involved in the execution of a use case. We model sequential and concurrent activities using activity diagrams. So, we basically depict workflows visually using an activity diagram. An activity diagram focuses on the condition of flow and the sequence in which it happens. We describe or depict what causes a particular event using an activity diagram.

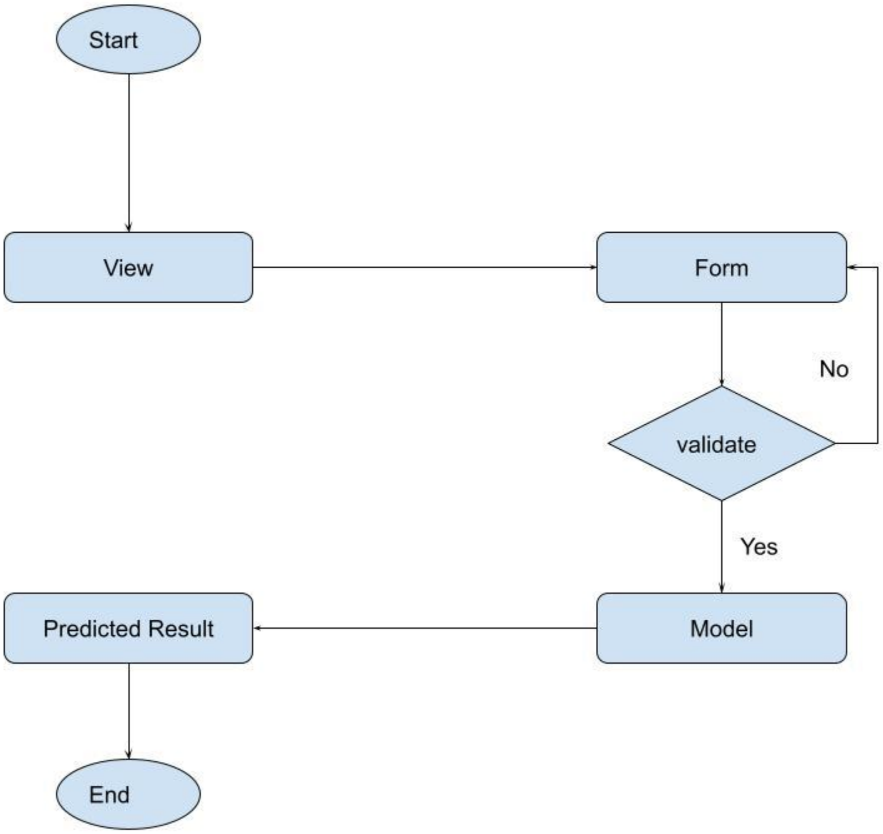


Figure 3: Activity Diagram

Sequence Diagram

Sequence Diagrams are used to focus the time as well as to show the order of the interaction visually with the help of a vertical axis that will represent the time in order to send messages. It validates the logic of the working of the model and the website within itself and also the user who uses it.

Dataset is loaded and pre-processed. Models are build and they are trained using the pre-processed dataset. Later the same models are used for predicting the rainfall. User has options to select Places and months from the ui. After the place is selected, user enters the observed parameter values. Later the data is pre-processed and given as input to the model. Model predicts the rainfall values and displays it on the ui. The detailed sequence diagram of the rainfall prediction model is depicted in figure 3.

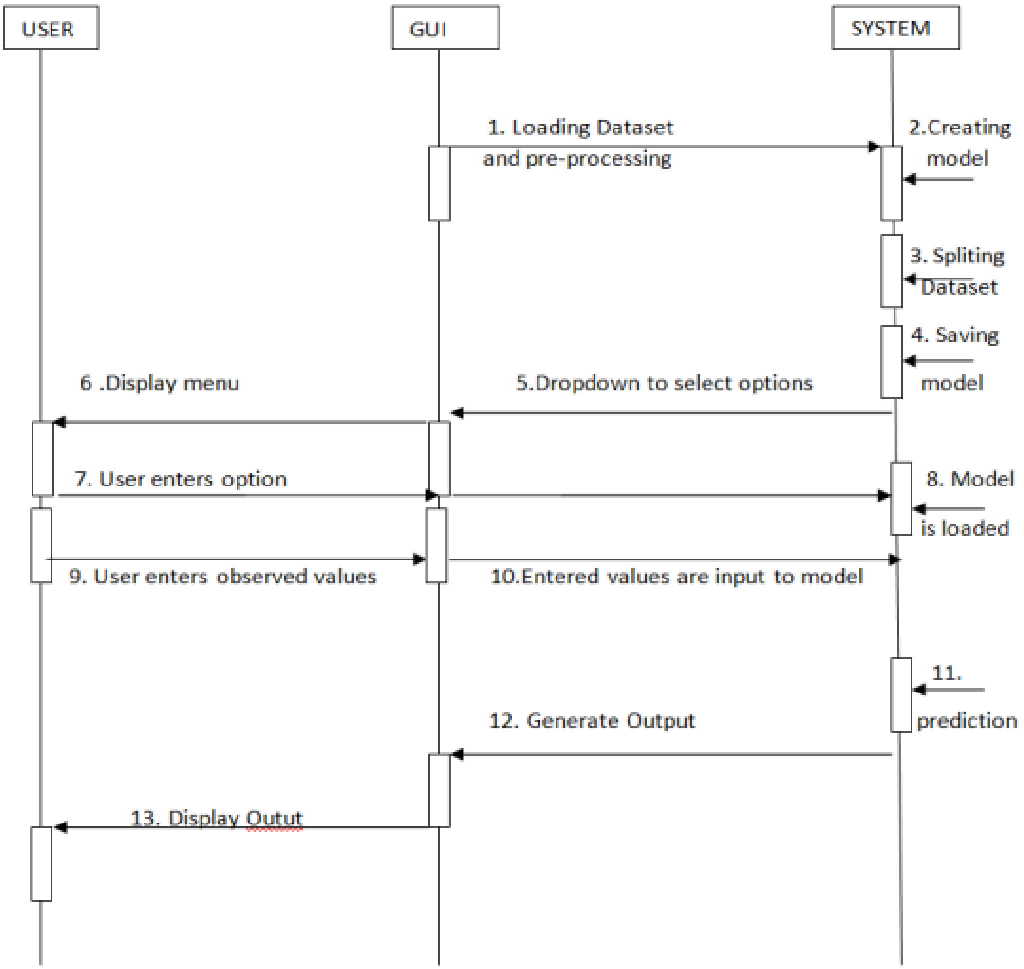


Figure 4:Sequence Diagram

4.3 Methods and Algorithms

Data collection

For this project, the raw data were collected from the National Center of Meteorology and Hydrology. 8 data features such as place, year, month, date, maximum temperature, minimum temperature, humidity, wind speed, and rainfall were included as shown in table 1. The raw data was recorded from 2020 - 2021. The meteorology station records the values of the environmental variable every day for each year directly from the devices in the station.

Parameters	Measurement (unit)
Place	-
Year	-
Month	-
Max. Temperature	℃
Min. Temperature	℃
Relative Humidity	%
Wind speed	m/sec
Rainfall	mm

Table 2:Dataset information

Data preprocessing

The data preprocessing step includes the conversion of data (daily data to monthly data), handling the missing values, categorical encoding, and splitting dataset into train and test sets. A total of 11 years (2010–2021) data were collected from the meteorology office. Since the data were raw, they contained missing values, and wrongly encoded values. Therefore, the missing value will be replaced with the mean of the data. Since the machine learning algorithms handle the numerical value, those columns having categorical value need to be encoded with the numerical values.

Machine Learning Algorithms

Machine learning uses two types of techniques: supervised learning, which trains a model on known input and output data so that it can predict future outputs, and unsupervised learning, which finds hidden patterns or intrinsic structures in input data. A supervised learning algorithm takes a known set of input data and known responses to the data (output) and trains a model to generate reasonable predictions for the response to new data. Supervised learning uses classification and regression techniques to develop predictive models

Multiple regression

Multiple regression is a statistical technique that can be used to analyze the relationship between a single dependent variable and several independent variables. The objective of multiple regression analysis is to use the independent variables whose values are known to predict the value of the single dependent value. Each predictor value is weighed, the weights denoting their relative contribution to the overall prediction

$Y=a+b1X1+b2X3+\ldots+bnXn$

Here Y is the dependent variable, and X1,…,Xn are the *n* independent variables. In calculating the weights, a, b1,…,bn, regression analysis ensures maximal prediction of the dependent variable from the set of independent variables. This is usually done by least squares estimation.

Why Multiple regression in rainfall prediction?

- It will help us to predict the value of rainfall by looking at several parameters such as tmax, tmin, humidity, wind speed.

Random Forest(RF)

A Random Forest Regression model is powerful and accurate. It usually performs great on many problems, including features with non-linear relationships. Random forest regression is a supervised machine learning algorithm that uses the ensemble learning method for regression.

RF works by building several decision trees during training time and outputting the mean of the classes as the prediction of all the trees.

The RF algorithm works on the following steps:

- Take at random p data points from the training set
- Build a decision tree associated with these p data points
- Take the number N of trees to build and repeat a and b steps
- For a new data point, make each one of the N tree trees predict the value of y for the data point and assign the new data point to the average of all of the predicted y values.

According to the RF algorithm, it is efficient for large datasets and a good experimental result is obtained using large datasets having a large proportion of the data missing.

Why random forest in rainfall prediction?

- Accurate and efficient when running on large data
- Resistant to overfitting
- Can handle thousands of input variables without variable deletion
- Can estimate what variables are important in classification
- Provides effective methods for estimating missing data
- Maintains accuracy when a large proportion of the data is missing

Gradient Boosting Algorithm

The boosting algorithm is used when massive loads of data have to be handled to make predictions with high accuracy. Boosting is an ensemble learning algorithm that combines the predictive power of several base estimators to improve robustness. In short, it combines multiple weak or average predictors to build a strong predictor. These are the most preferred machine learning algorithms today. Use them, along with Python and R Codes, to achieve accurate outcomes.

Why Gradient Boosting Algorithm

- Often provides predictive accuracy that cannot be trumped.
- Lots of flexibility - can optimize on different loss functions and provides several hyper parameter tuning options that make the function fit very flexible.
- No data pre-processing required - often works great with categorical and numerical values as is.
- Handles missing data - imputation not required.

KNN (K- Nearest Neighbors) Algorithm

This algorithm can be applied to both classification and regression problems. Apparently, within the Data Science industry, it's more widely used to solve classification problems. It’s a simple algorithm that stores all available cases and classifies any new cases by taking a majority vote of its k neighbors. The case is then assigned to the class with which it has the most in common. A distance function performs this measurement.

KNN can be easily understood by comparing it to real life. For example, if you want information about a person, it makes sense to talk to his or her friends and colleagues!

Why KNN?

- Simple implementation
- can learn non-linear functions
- Evolves with new data

Decision Tree

Decision Tree algorithm in machine learning is one of the most popular algorithms in use today; this is a supervised learning algorithm that is used for classifying problems. It works well classifying for both categorical and continuous dependent variables. In this algorithm, we split the population into two or more homogeneous sets based on the most significant attributes/ independent variables.

Why Decision Tree?

- Compared to other algorithms, decision trees require less effort for data preparation during pre-processing.

- A decision tree does not require normalization of data.
- A decision tree does not require scaling of data as well.
- Missing values in the data also do NOT affect the process of building a decision tree to any considerable extent.
- A Decision tree model is very intuitive and easy to explain to technical teams as well as stakeholders.

Artificial Neural Networks(ANN)

Artificial Neural Networks are a special type of machine learning algorithms that are modeled after the human brain. That is, just like how the neurons in our nervous system are able to learn from the past data, similarly, the ANN is able to learn from the data and provide responses in the form of predictions or classifications.

Support Vector Machine

Support vector machines (SVMs) are a set of supervised learning methods used for classification, regression and outlier detection. The advantages of support vector machines are: Effective in high dimensional spaces. Still effective in cases where the number of dimensions is greater than the number of samples.

CHAPTER 5

RESULT AND DISCUSSION

5.1 Result

The proposed framework is implemented on a real-time rainfall dataset of the 10 Dzongkhags of Bhutan, Granted by NHMC,The dataset used in this project spans over 21 years (2000 to 2021) and consists of 2640 instances and 8 features. First, 7 features are the independent features, which are given as input to the proposed framework in order to predict the 8th feature, which is the output class (dependent feature). The output class indicates how much rainfall will occur in a given month. The dataset is divided into two parts: 80% of the data is reserved for training(2065), and 20% of the data is reserved for testing (517). The activities of the pre-processing stage, including cleaning and normalization, are performed on the rainfall dataset before building a model. To predict, Seven machine learning techniques are used: Linear regression, KNN, Decision Tree, Random Forest, ANN and SVM. These Machine Learning techniques are optimized iteratively until maximum accuracy is achieved and finally the best model is selected as the best technique to predict the monthly rainfall.

Algorithm comparison based on Train Test Accuracy

After implementing all the proposed algorithms we have drawn a test and train accuracy from every algorithm as given in table __ and figure __ .The best train accuracy we have got are 95% in Random forest and 84 and 78% in ANN and SVM respectively. Even if the train accuracy of random forest and ANN are higher we cannot consider them as a good algorithm for this dataset because their test accuracy is relatively low which will cause overfitting in the later part of prediction. Based on the train and test accuracy result, it can be seen that the level of accuracy using SVM is the best algorithm for a provided dataset with 78% of train accuracy and 76% of test accuracy.

Algorithm	Linear	KNN	Decision Tree	Gradient boosting	Random forest	ANN	SVM
Train accuracy	38	75	75	65	95	84	78
Test accuracy	28	57	49	46	52	54	76

Table 3:Train-test accuracy

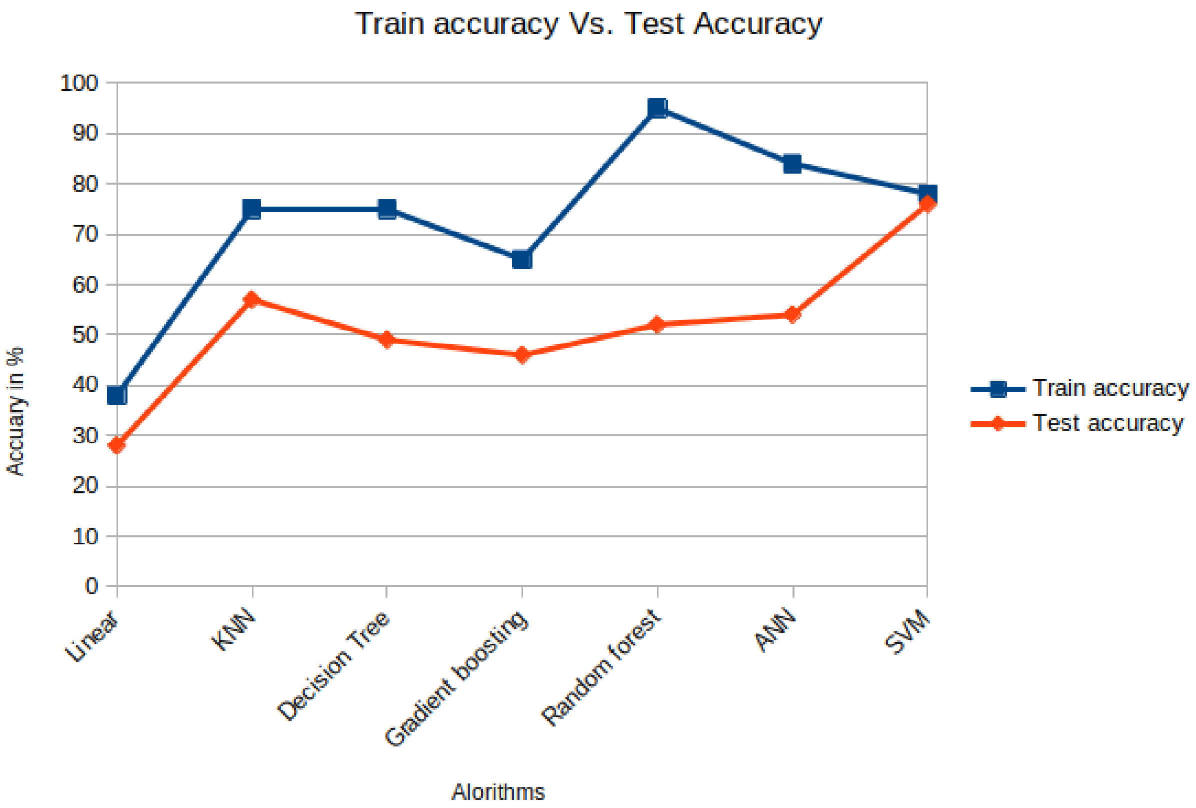


Figure 5: Train-test Comparison

Algorithm comparison based on MSE

Mean squared error(MSE) aims to determine the average of error squares i.e. the average squared difference between the predicted values and true value. By looking at table __ and graph __ we can say that KNN and ANN got the least error compared to the mse values of other algorithms. Linear regression has got the maximum mse for the provided dataset with 0.92 mse value.

Algorithm	MSE
Linear	0.92
KNN	0.54
Decision Tree	0.64
Gradient boosting	0.69
Random Forest	0.61
ANN	0.58
SVM	0.78

Table 4:MSE

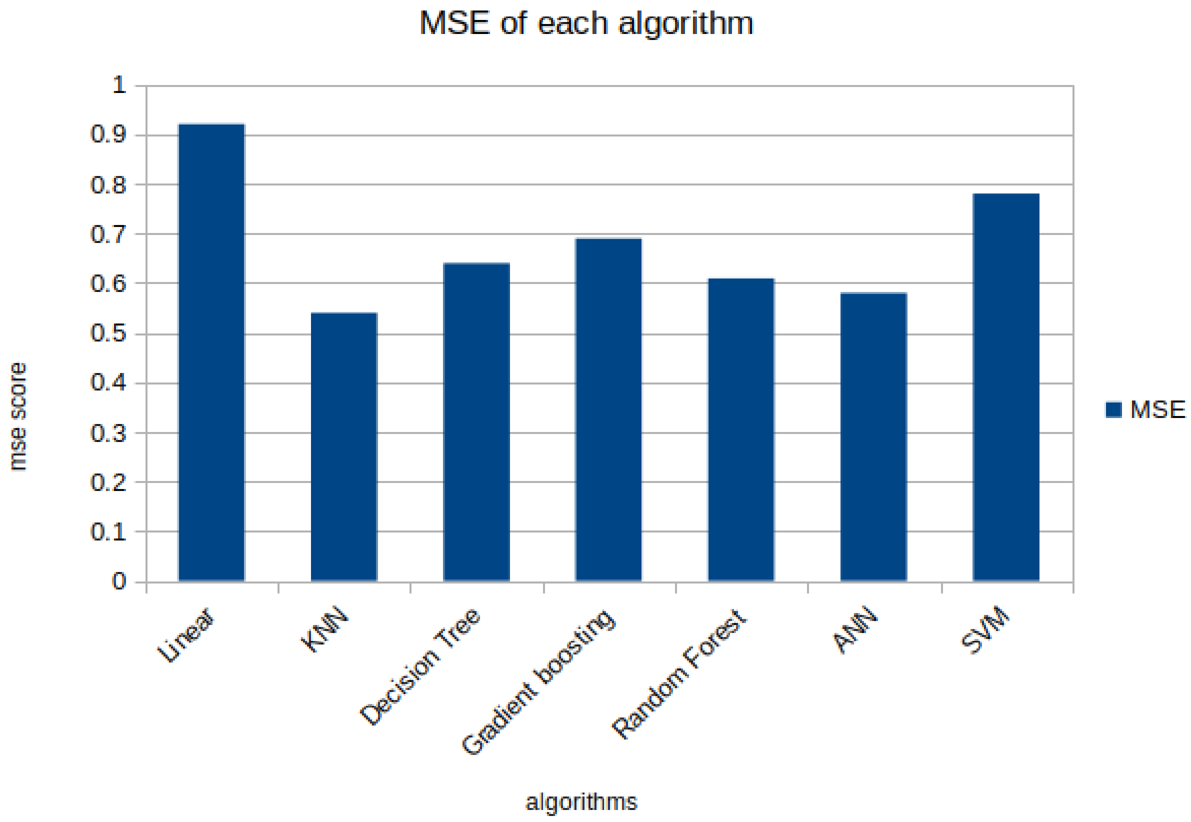


Figure 6: MSE Comparison

Algorithm comparison based on RMSE

Root Mean Square Error is a well-known and commonly used evaluation process for regression models.

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{n} (yt-\widehat{yt})^{2}}{n}}$$

Here, yt is the original value of a point for a given time period t, n is the total number of fitted points, and ^yt is the fitted forecast value for the time period t.

Table __ and figure __ gives the RMSE of each algorithm. From the table and graph given below we can conclude that SVM and ANN has got the least rmse with 0.73 andw 0.58 respectively which means that these two algorithms are predicting more accurately with less error than other algorithms.

Algorithm	rmse
Linear	0.96
KNN	0.79
Decision Tree	0.8
Gradient boosting	0.83
Random Forest	0.78
ANN	0.58
SVM	0.73

Table 5: RMSE

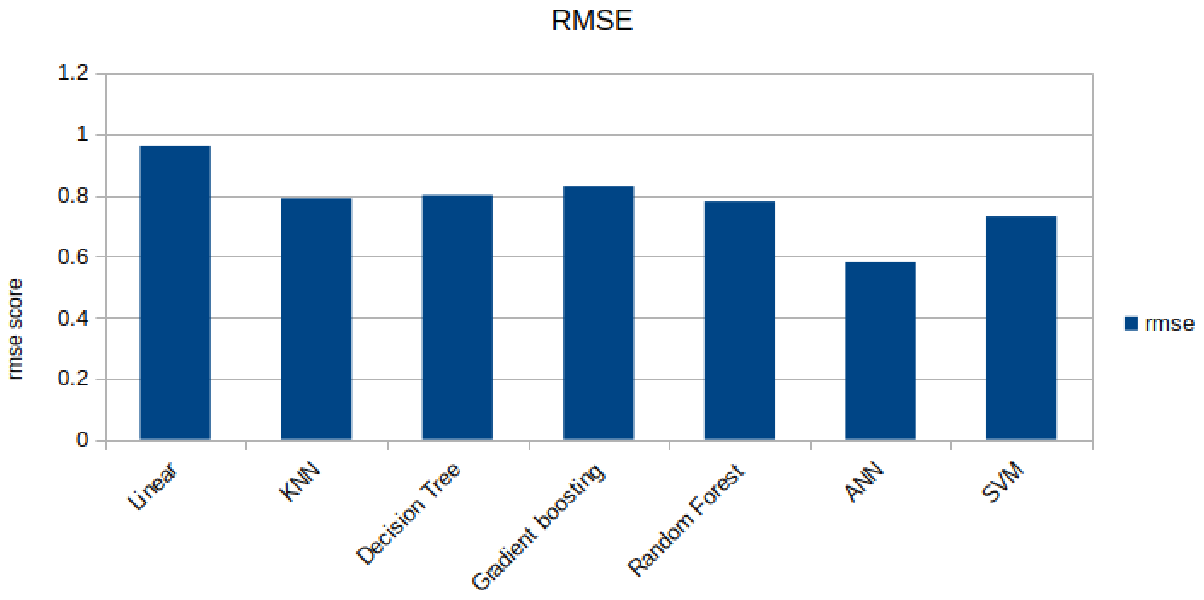


Figure 7: rMSE Comparison

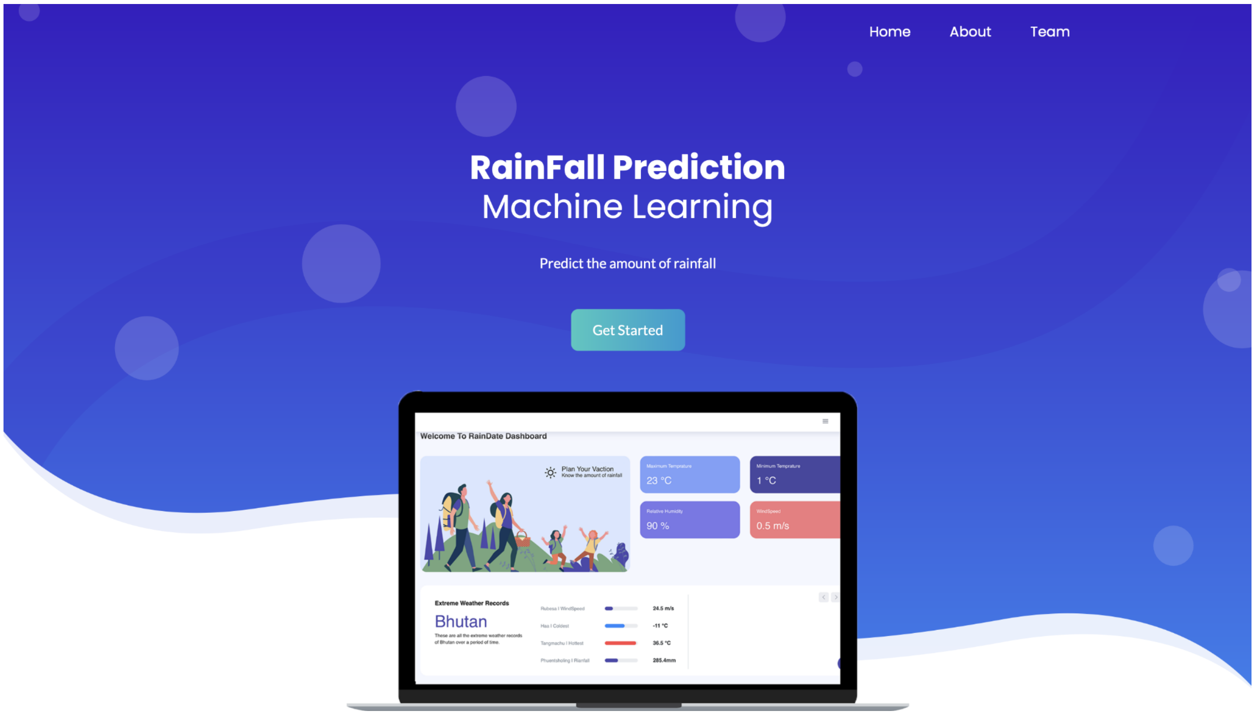
5.2 Discussion

After going through all the result analysis above we concluded that SVM is the best machine learning algorithm for monthly rainfall prediction because it has equal train and test accuracy which will avoid overfitting during prediction and it has less RMSE as compared to other algorithms which helps us to get more accurate results with less error.

CHAPTER 6

USER INTERFACE

6.1 Website overview



RainFall Prediction
Machine Learning

Predict the amount of rainfall

Welcome To RainDate Dashboard

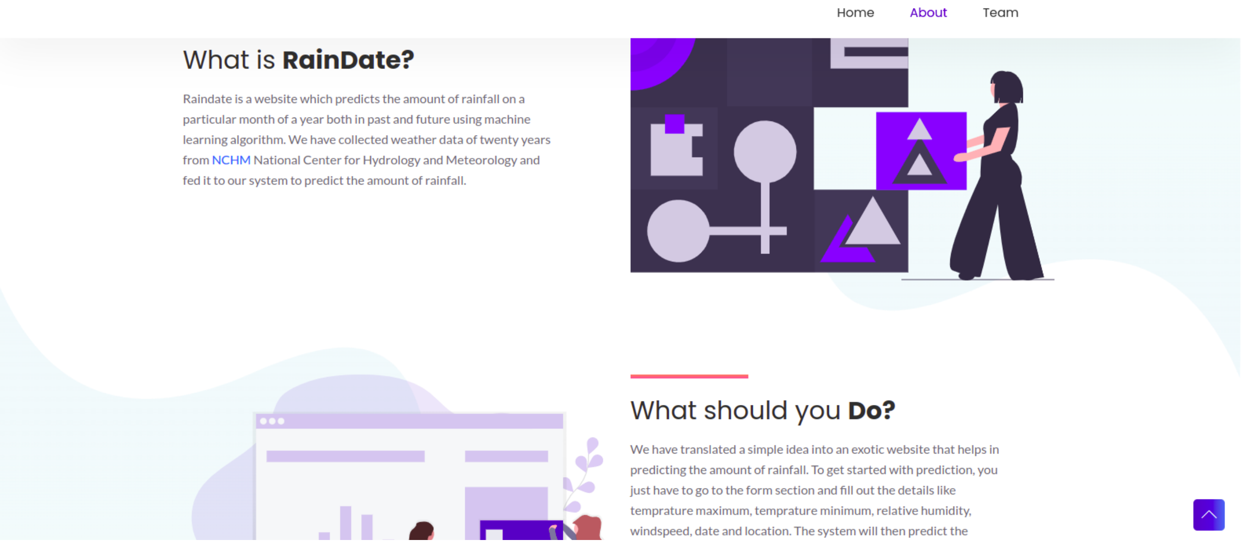
Extreme Weather Records

Bhutan

These are all the extreme weather records of Bhutan over a period of time.

Rubesa I WindSpeed	24.5 m/s
Haa I Coldest	-11 °C
Tangmachu I Hottest	36.5 °C
Phuentsholing I Rianfall	285.4mm

Figure 8:Home



What is **RainDate?**

Raindate is a website which predicts the amount of rainfall on a particular month of a year both in past and future using machine learning algorithm. We have collected weather data of twenty years from NCHM National Center for Hydrology and Meteorology and fed it to our system to predict the amount of rainfall.

What should you **Do?**

We have translated a simple idea into an exotic website that helps in predicting the amount of rainfall. To get started with prediction, you just have to go to the form section and fill out the details like temprature maximum, temprature minimum, relative humidity, windspeed, date and location. The system will then predict the

Figure 9: About

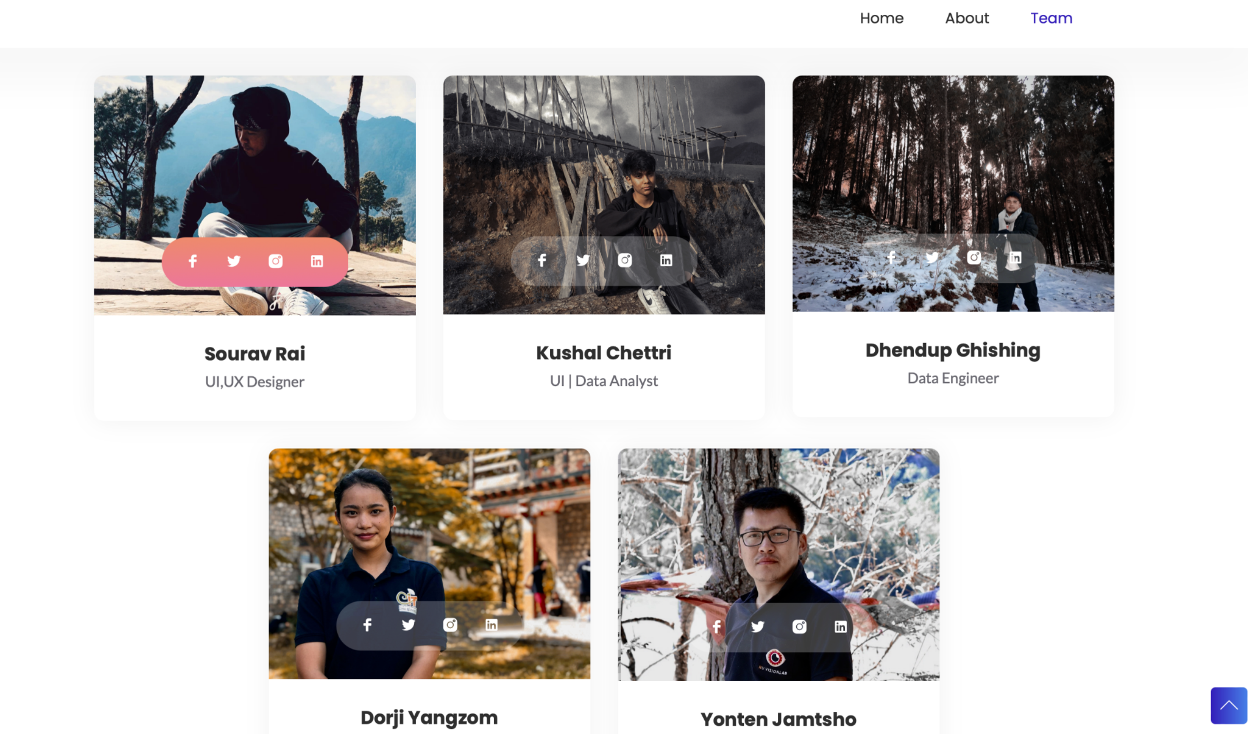
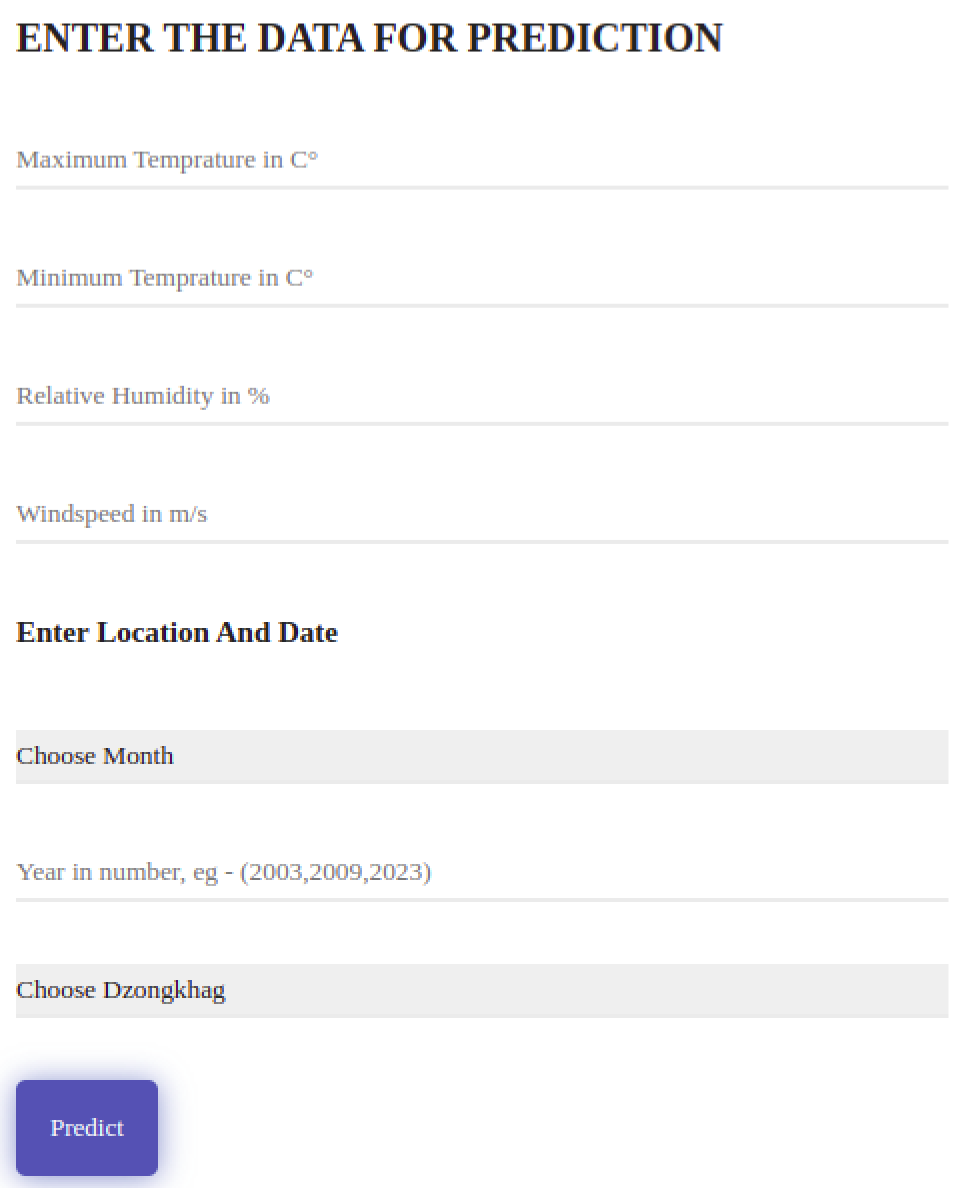


Figure 10: Team



ENTER THE DATA FOR PREDICTION

Maximum Temprature in C°

Minimum Temprature in C°

Relative Humidity in %

Windspeed in m/s

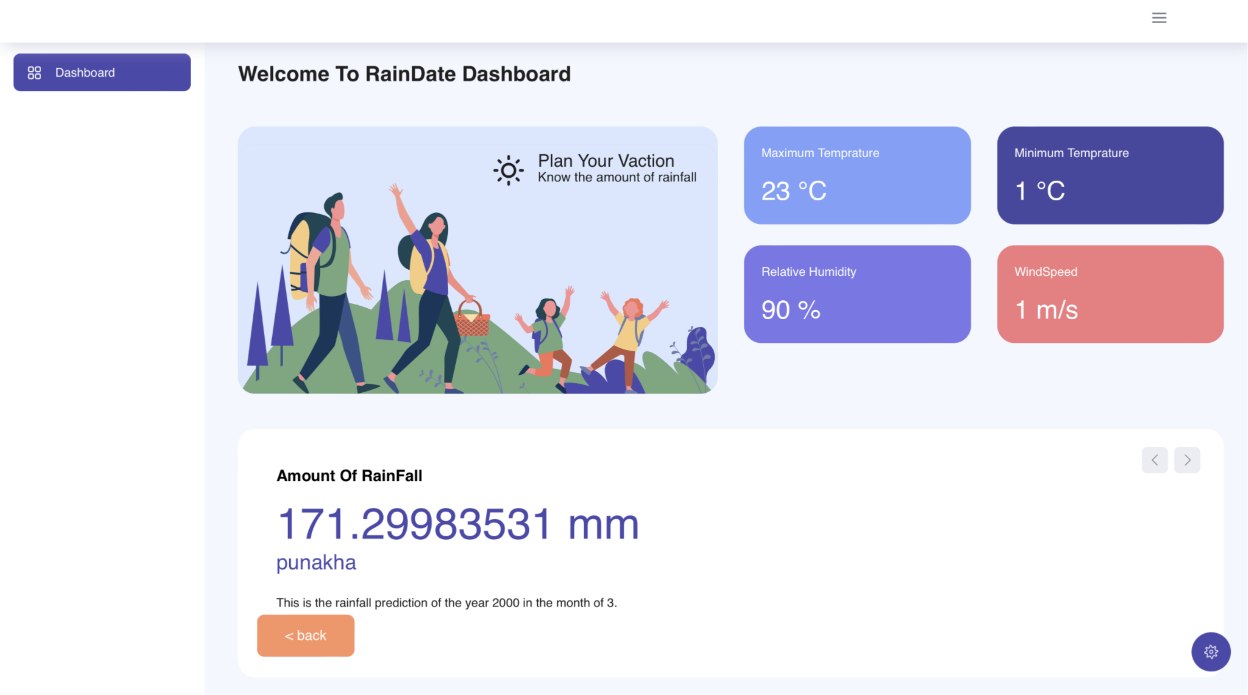
Enter Location And Date

Choose Month

Year in number, eg - (2003,2009,2023)

Choose Dzongkhag

Figure 11: Form



Welcome To RainDate Dashboard

Plan Your Vaction
Know the amount of rainfall

Maximum Temprature
23 °C

Minimum Temprature
1 °C

Relative Humidity
90 %

WindSpeed
1 m/s

Amount Of RainFall

171.29983531 mm

punakha

This is the rainfall prediction of the year 2000 in the month of 3.

Figure 12: Dashboard

CHAPTER 7

CONCLUSION AND FUTURE WORK

A detailed survey on rainfall predictions using different machine learning algorithms over twenty-one years is done. From the survey it has been found that most of the researchers used different models for rainfall prediction, but SVM gives significant results. SVM is the model with least root mean squared error and accurate prediction. The survey also gives a conclusion that forecasting techniques like Linear regression,Random forest,KNN,Decision tree,ANN and Gradient Boosting are suitable to predict rainfall. However, some limitations of those methods have been found.The study of machine learning methods for rainfall prediction is presented in this paper, and a SVM based model is proposed for rainfall prediction over the 10 places of bhutan. The sensor data are collected from the meteorology department of Bhutan, which contain daily records of weather parameters from 2000 to 2021. Finally the SVM model performed much better than all the other models under study for this dataset. Its MSE score of 0.78 and furthermore, the proposed model presented an improved accuracy of 76% and RMSE score of 0.73. Predicting actual rainfall values has become more challenging due to the changing weather patterns caused by climate change.

In the future, we aim to improve the performance of our prediction model by incorporating patterns of global and regional weather such as sea surface temperature, global wind circulation, etc.We also intend to explore the predictive use of climate indices and study the effects of climate change on rainfall patterns.

REFERENCES

AKSHATHA, M. (2019). RAINFALL PREDICTION USING MACHINE LEARNING TECHNIQUES.

Chhetri, M., Kumar, S., Pratim Roy, P., & Kim, B. G. (2020). Deep BLSTM-GRU model for monthly rainfall prediction: A case study of Simtokha, Bhutan. *Remote sensing*, *12*(19), 3174.

Hasan, N., Nath, N. C., & Rasel, R. I. (2015, December). A support vector regression model for forecasting rainfall. In *2015 2nd international conference on electrical information and communication technologies (EICT)* (pp. 554-559). IEEE.

Lima, P. M., & Guedes, E. B. (2015). Rainfall Prediction for Manaus, Amazons with Artificial Neural Networks. Computational Intelligence (pp. 1-5). Curitiba, Brazil: IEEE.

Liyew, C. M., & Melese, H. A. (2021). Machine learning techniques to predict daily rainfall amount. *Journal of Big Data*, *8*(1), 1-11.

Osmani, S. A., Mahmud, F., & Zafor, M. A. Prediction of Rainfall using Machine Learning Algorithms for Different Districts of Meghalaya.

Shaw, R. (2017). Top 10 machine learning algorithms for beginners. *URl: https://www.kdnuggets. com/2017/10/top-10-machine-learning-algorithmsbeginners. html*.

Srivastava, S., Anand, N., Sharma, S., Dhar, S., & Sinha, L. K. (2020, June). Monthly rainfall prediction using various machine learning algorithms for early warning of landslide occurrence. In *2020 International Conference for Emerging Technology (INCET)* (pp. 1-7). IEEE.