FINDING WEATHER PREDICTION ACCURACY USING LINEAR REGRESSION

BY

ARAFARTUR RAHMAN SOIKAT
ID: 171-15-9330

G M FARADUZZAMAN
ID: 171-15-9322

AND

MD AL AMIN
ID: 171-15-9302

This Report Presented in Partial Fulfillment of the Requirements for the Degree of Bachelor of Science in Computer Science and Engineering

Supervised By

Shah Md. Tanvir Siddiquee
Assistant Professor
Department of CSE
Daffodil International University

Co-Supervised By

Mr. Narayan Ranjan Chakraborty
Assistant Professor
Department of CSE
Daffodil International University

DAFFODIL INTERNATIONAL UNIVERSITY

DHAKA, BANGLADESH

JANUARY 2021

APPROVAL

This Project titled "**FINDING WEATHER PREDICTION ACCURACY USING LINEAR REGRESSION"**, submitted by **Arafartur Rahman Soikat**, ID No: 171-15-9330, **G M FARADUZZAMAN**, ID No: 171-15-9322 and **Md Al Amin**, ID No: 171-15-9302 to the Department of Computer Science and Engineering, Daffodil International University, has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of B.Sc. in Computer Science and Engineering and approved as to its style and contents. The presentation has been held on **Wednesday**, 27 January 2021.

BOARD OF EXAMINERS

Dr. Touhid Bhuiyan **Chairman**

Professor and Head

Department of Computer Science and Engineering

Faculty of Science & Information Technology

Daffodil International University

Gazi Zahirul Islam **Internal Examiner**

Assistant Professor

Department of Computer Science and Engineering

Faculty of Science & Information Technology

Daffodil International University

Raja Tariqul Hasan Tusher **Internal Examiner**

Senior Lecturer

Department of Computer Science and Engineering

Faculty of Science & Information Technology

Daffodil International University

Dr. Dewan Md. Farid **External Examiner**

Associate Professor

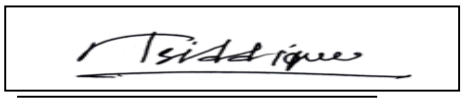
Department of Computer Science and Engineering

United International University

DECLARATION

We hereby declare that, this project has been done by us under the supervision of **Shah Md. Tanvir Siddiquee, Assistant Professor, and Department of CSE** Daffodil International University. We also declare that neither this project nor any part of this project has been submitted elsewhere for award of any degree or diploma.

Supervised by:

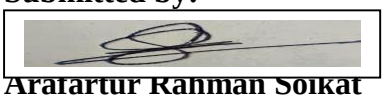


Name: Shah Md. Tanvir Siddiquee
Designation: Assistant Professor
Department of CSE
Daffodil International University

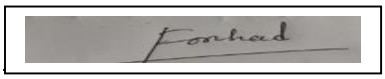
Co-Supervised by:

Name: Mr. Narayan Ranjan Chakraborty
Designation: Assistant Professor
Department of CSE
Daffodil International University

Submitted by:

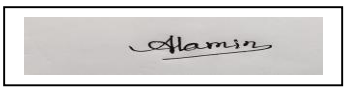


Arafartur Rahman Soikat
ID: - 171-15-9330
Department of CSE
Daffodil International University



G M FARADUZZAMAN

ID: - 171-15-9322
Department of CSE
Daffodil International University



Md Al Amin
ID: - 171-15-9302
Department of CSE

ACKNOWLEDGEMENT

First we express our heartiest thanks and gratefulness to almighty God for His divine blessing makes us possible to complete the final year project successfully.

We really grateful and wish our profound our indebtedness to our Supervisor **Shah Md. Tanvir Siddiquee, Assistant Professor** and Co-supervisor **Mr. Narayan Ranjan Chakraborty,** Department of CSE Daffodil International University, Dhaka. Deep Knowledge & keen interest of our Supervisor and Co-supervisor in the field of "*Machine Learning*" helped us to carry out this project. The entire time they have upheld and enlivened us and demonstrated the correct way. Their endless patience, scholarly guidance ,continual encouragement , constant and energetic supervision, constructive criticism , valuable advice ,reading many inferior draft and correcting them at all stage have made it possible to complete this project.

We would like to express our heartiest gratitude to **Prof. Dr. Touhid Bhuiyan,** Professor and Head**,** Department of CSE, for his kind help to finish our project and also to other faculty member and the staff of CSE department of Daffodil International University.

We would like to thank our entire course mate in Daffodil International University, who took part in this discuss while completing the course work.

Finally, we must acknowledge with due respect the constant support and patients of our parents.

ABSTRACT

Weather forecasting is a technology to found the future condition of atmosphere for the selected location. This research field is the most sensitive for real time issues. We propose linear regression algorithm for predict “Rainfall”, “Max Temp”, “Min Temp” and “Relative Humidity”. We are using here Jupyter Notebook and library was Numpy, Pandas for finding prediction accuracy. We find this accuracy from Bangladesh weather data perspective and finding annual accuracy of weather prediction using linear regression from 1949-2013.proper weather forecasting is very important for our daily life and our country. It affects our daily lives and our economy. The weather prediction that we have can be used to accurately predict the weather. To deliver better weather forecasting, we have to learn machine learning technique perfectly. Machine learning technique will help us to analyze weather prediction pattern from the dataset. In this research paper we can use machine learning based weather prediction method and lead dataset to analyze real time humidity, maximum temperature, minimum temperature and rainfall. In this paper we use direct linear regression method t calculate the accuracy of rainfall, maximum temperature, minimum temperature and humidity. Based on this method we calculate not only relation between dataset and prediction accuracy of weather forecasting but also the relation between dataset modernism and prediction accuracy.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF Table

CHAPTER 1

Introduction

1.1 Introduction

Forecasting is an issue of delivering a different, scope of numbers or a couple of important measurements that are viable with fate occurrence. Inverse to expectation, past measurements are the premise of estimating which is identified with instinct and presumption. The principle thought process of climate anticipating is to caution the individuals of a looming hazard so that proper advances can be taken to guard their lives and homes. Different utilizes determining to include air traffic, marine, agribusiness, ranger service, programming organizations, and private areas, armed force, etc. Accumulating the quantitative. Realities about the overarching nation of air and the utilization of the methodical and consistent acknowledgment of the environmental strategy, Climate figures are finished. Bangladesh is a calamity willing spot of the field where limit of the disappointments are brought about by wonders of the environmental factors.

A couple of extraordinary climate exercises were liable for some losses and harms including the top notch Backerganj storm of 1876, the destroyer of November 1970 Bhola Cyclone twister, the Urirchar hurricane 1985; and the typhoons of 1969 Demra, Dhaka, 1974 Manikganj, 1977 Madaripur, 1989 Saturia of Manikganj, 1995 Laohajong, and 1996 Tangail. 14- 15 May Cyclone Akash speeds up to 115 km, 2007, Chittagong. 15 November Cyclone Sidr with 260 km/hour southern Bangladesh, 26–27 October Rashm Cyclone (2008) ,19–21 April Bijli Cyclone 2009, 27–29 May Cyclone Ail 2009, 16–17 May Cyclone Viyaru 2013, 29 July Cyclone Komen 2015, 21 May Cyclone Roanu 2016 , 20 August 2016, 29–31 May Cyclone Mora 2017, 4 May Cyclone Fani 2019, 9 November Cyclone Bulb 2019, 20 May Cyclone Amphan 2020.

If we want to do weather prediction analysis in different places and in the future then we need to know what the atmosphere is like in different places. In this way climatology a run of the mill learn about climate conditions experimentally and anticipating the future atmosphere. In late years, climate determining confronting a few specialized gives everywhere on the world based on the weather problem. In climatology, climate forecasting assumes a fundamental job, particularly for the meteorologist is consider as the significant difficulty in precise expectation [1]. A forecast is a number, a range of

numbers, or some relevant information that is subject to production consistent with future events. In contrast to predictions, past information is the premise of forecasting which is identified with instinct and suspicion [2].

The primary motivation behind climate estimating is to warn people of impending danger so that suitable advances can be taken to secure their lives and properties. For useful of human existence, prediction of nature positively climate alerts is more significant in guaranteeing their life and property. In cultivating, the climate conjecture is crucial and the central point that rely upon determining are wind, temperature, mugginess, and viewpoint. The vast majority of the service organizations using temperature forecasting for assessing what's to come. The abrupt atmosphere changes bring about generous downpour, day off wind chill which seriously affects open air exercises. Such future forecasts empower to plan and plan action as indicated by the events which reflect in their stabilities. Evaluating the main problem of weather forecasting and processing huge amounts of weather datasets. Run it implements several data mining methods. Data mining strategies apply to data analysis, briefly classify and recognize connections. The data mining process is related to three assignments inside phrases as classification, education and prediction. Data mining, taxonomy is a machine learning process application for forecasting and data field collection [2].

If we give a example, on whether prediction datasets classification is of “rainfall”, “maximum temperature” or “minimum temperature”, ‘’humidity’’ according to the weather on a specific day. The process of training and mapping datasets to obtain specific information related to studies. Learning are supervised in two ways, such as supervised and unsupervised learning. Supervised learning algorithms use classifieds to analyze and distribute the power generated from training data. One of the most emerging issues in the past decade is climate change and forecasting in general. The way to focus on big climate data is big data analysis. The most needed and acceptable technology for big data analytics is a myriad of issues. The huge amount of climate data that is kept growing in our daily lives [2].

Bangladesh positions fifth in the Global Climate Risk Index Ranking of the 170 nations most in danger for environmental Change. The nation is especially in danger since it is an immense deltaic plain 230 rivers, a significant number of which swell unpredictably

during the storm. Toward the north of the Himalayas is an ice sheet and an undetectable Bay of Bengal. In the south, the district is inclined to serious flooding. The circumstance is exacerbated by the impacts of more extreme tempests, one marker of climatic pressure. Sidr, a twister in four divisions that crushed southern Bangladesh in November 2004, slaughtered around 3,500 individuals, dislodged 2 million and cleared out paddy fields. Around 1,500 individuals kicked the bucket and around 2 million tons of food were harmed after two more floods than expected after Sidr. The United Nations has cautioned that if ocean levels rise three feet in the following 50 years, a fourth of Bangladesh's coastline could sink. , Relocating 30 million Bangladeshis from their homes and homesteads. On the off chance that this occurs, the capital Dhaka, presently in the focal point of the nation, will have its own ocean take off (CCC, 2009) [3].

1.2 Motivation

Bangladesh arguably have the dangerous weather in the world. We know Bangladesh is a riverine country and Bangladesh is a country of conspiracies. There are many disaster have gone over through Bangladesh like sidr, ayla, flood, nargis, amphan, droughts, heat waves. Dangerous weather is the reason behind the national weather service have established.

Different weather prediction reduce casualties and weather-related hazards related to property damage, and economic impact. Climate prediction see current climate and estimate guides and come with a forecast and the climate figure encourages a great deal to give alerts.

It is very easy to predict the temperature of the weather after the invention of the computer. We can quickly and quickly analyze wind, temperature, rainfall and other data. Utilizing computers we can anticipate the best models that are expected to precisely figure the climate in our nation. Already we were unable to anticipate the specific climate in our conventional framework. It was long and tedious, so the current course of action is an extraordinary gift for us. Atmosphere anticipating is critical on the grounds that our economy is absolutely subject to environmental change, we can clarify that everything in our farming, ware shipments is identified with environmental change.

1.3 Rational of the study

If there is no destination, we can't even imagine to expect the result of related work. So the essential objective of our work is to find the correct atmosphere technique with the utilization of current AI. The force based AI technique for man-made mind encourages us to actually gauge atmosphere and anticipate atmosphere temperature. Our hypothesis endeavors to utilize a coherent based procedure utilizing machine learning technique to deal with things.

Different assurance methods for temperature assessment for various measures have been proposed and detailed. In this activity, we propose new extravagant strategies for month temperature far off months utilizing AI counts. These strategies rely upon the manufacture of the pseudo-sensory system and the pseudo-neural frameworks. Two kinds of checks are the focal point of this activity: mathematical what not. The desire for numbers is the best approach to direct real estimations of temperature. In spite of the free desire, precipitation is shipped off the inverse (typically ordinary) class. To see the temperature, exceptional and close by atmosphere qualities from around the globe were assembled as possible pointers. These markers incorporated elite of temperature, sunflower properties, and air.

Moreover, we proposed assorted gathering strategies for temperature gauge. The purpose of the gatherings was to meld parts arranged assortments. This included improving the blend strategies and picking outfit parts.

1.4 Research Questions

There are many questions that we are trying to answer in the research paper.

- Has any algorithm to provide better weather prediction forecasting?
- Has any medium by which we can get the weather forecast in advance?
- Why do we care about weather forecasting?
- What about people who grow food?
- How do people know about predicted weather?

1.5 Expected Output

Here we will discuss about what output we got from this paper.

- Get help to know about the weather forecasting.
- Get information about future scope of weather prediction.
- Get help to know about weather condition of Bangladesh.
- This paper will help us to know about how climate change.

- ❖ This paper will help us to know about machine learning (Linear Regression).
- ❖ Get help to know about how can we calculate or get perfect accuracy for weather prediction.
- ❖ This will raise awareness among all.
- ❖ The number of people who die every year as a result of disasters because they do not know in advance what will happen tomorrow. It will help us to know and decrease dying issue.

1.6 Report Layout

The report is divided into five chapters. Each chapter deals with the different aspects of.

"Finding Weather Prediction Accuracy using Linear Regression". Each chapter has different part and this parts is describe briefly.

- **Chapter 1: Introduction**

  This part examines the significant theoretical ideas behind our research paper. Here we also discuss about motivation, rational of study, research question and expected output.
- **Chapter 2: Background**

  In this chapter we describe about related works, research summary, scope of the problem and challenges.
- **Chapter 3: Research Methodology**

  In this chapter we describe about research subject & instrumentation, procedure of data collection, statistical analysis implementation requirements.
- **Chapter 4: Experimental Results and Discussion**

  In this chapter we describe about experimental results, descriptive analysis.
- **Chapter 5: Summary, Conclusion, Recommendation and Implication for Future Research**

  In this chapter we describe about summary, conclusions recommendations, further study.

CHAPTER 02
Background

2.1 Introduction

Finding Weather Prediction Accuracy using Linear Regression is a process for Finding Weather Prediction Accuracy. For prediction, first of all select some data basis on some criteria, like Data Collection, Data preprocessing, Data Selection, and Accuracy. Different Decision Framework will be discussed for different applications in this section that are related in this paper.

This chapter is having details work present, related work, research summary. Details about Scope of the problem. Our target and challenges that we faced are described here.

2.2 Related Works

In recent times, weather prediction researcher has been developed several systems to finding weather prediction. Jarrett Booz e analysis-based weather forecast system for data volume. Weather prediction is essential to our everyday life. To give weather figure, machines learning methods can be utilized for comprehension and investigating whether design. We use Keras Sequential model as our profound learning model to learn and anticipate the weather information. The information forecast depends on a sliding window matrix-based mechanism. We utilize this present reality informational index from the National Maritime and Atmospheric Administration (NOAA) for assessments. We discovered that expanding the volume of information used to prepare a profound learning model utilized for weather forecasts could decidedly the model presentation [4]. Tanjid Rahman et al. (2014) presents two algorithms to improve make weather forecasting. Two algorithms are Markov-Fourier, Numerical Weather Prediction (NWP) model to predict forecasting. At first, this model Fuzzy Interface take linguistic Statements, after take statement prepare for Neural Network then Learning algorithms make the decisions [2]. MAI Navid, NH Niloy used two algorithm and data mining for data preprocessing. They used Multiple Linear Regression and Rainfall Prediction for data analysis and prediction. Bangladesh is a wealthy and deeply populated nation that has been frequent experience disasters cyclones floods saltiness interruptions, dry spells and so forth. This paper combines ultimate climatic occasions in Bangladesh with regards to the atmosphere displaying information. The displaying the next effect of Offensive occasions demonstrated huge patterns in Bangladesh because of environmental change. The analyses were performed to assess the precision

of rainfall prediction using multiple linear regressions [3]. Therefore, Dipak Kumar analysis climate change on rainfall in Northwestern located beside "Teesta River". Analysis climate change using LARS- WG weather simulator. This simulator good performer tested in demonstrated and diverse climates. Climate data are analyzed by this model also present Global Climate Models (GCM)for IPCC. Then make application for observed daily rainfall and provide statistical properties of each site [5]. S.M. Taohidul Islam and Sajal Saha used ANN model, seasonal variations, weather parameters, Pre disaster prediction, for data analysis, prediction and other calculation. There are many kinds of disasters like storm, flood, heavy rain, earthquake etc. cause an immediate effect in security and economy of Bangladesh. This paper examines the usage of ANN model for month-to-month premise dependable climate observing with occasional debacle forecast. The consequence of given model utilized for month-to-month precipitation forecast shows that a decent execution and sensible forecast exactness for Bangladesh [5]. In addition, Md Rashid Mahmood present weather forecast used to prediction to knowing the condition of weather atmospheric like rainfall, humidity, Heat etc. For predicted weather condition using data mining to handle data and using some algorithms like Naïve Bayesian, Decision Tree etc. Presenting ECDF model for analysis monthly low and high temperature and represent statistical graphical bar [6]. Md. Rejaur Rahman and Habibah Lateh they both are research about on the prediction of the weather period of Bangladesh and in this paper the month-to-month dataset of the base and most extreme temperatures also, precipitation from many places in Bangladesh by the period 1949–2013 and that was utilized and broke down to this examination is given by the Bangladesh Meteorological Department (BMD 2013). Linear regression, coefficient of variety, opposite distance weighted addition methods and geological data frameworks performed to investigate the patterns, changeability and another examples of temperature and precipitation. Autoregressive coordinated moving normal time arrangement model was utilized to reproduce the temperature and precipitation information. The outcomes affirm an especially solid and late atmosphere change in Bangladesh [7]. Fahad Sheikh analysis of data mining techniques for weather prediction. Weather forecast is vital in order to educate individuals and set them up ahead of time about the current and impending weather condition. This aides in decrease in loss of human existence and loss of resources and limiting the alleviation steps that are relied upon to be taken after a natural disaster occurs. Analysis weather prediction using various data mining techniques and

comparison between algorithms CART, C4.5, ANN, k-means clustering and also comparison between C4.5 and Naïve Bayes algorithm.C4.5 algorithm gives a better result. Then the data helps to prepare current and upcoming changes [8].

To improve and finding weather prediction Accuracy assure the training accreditation training process is required. A multi criteria decision framework which can minimize the accreditation revisions process. The method analyzes the internal and external criteria using the tool, and prioritize the coefficients required to take an effective decision. It will also help the authority to design innovative training programs based on identify the value of different decision elements. Normalized procedure is used with this method to assess individual measures and combined them to get the final prediction. Unfortunately, it generates different results for the same data. Therefore, a multiplicative approach with aggregation function overcomes the difficulties to predict weather prediction with additive approach.

2.3 Comparative Analysis & Summary

Exact weather prediction is essential to our everyday life and have both monetary and environment impact. Weather predication model, can be foresee the weather. To give weather figure, machines learning strategies can be utilized for comprehension and examining whether designs. At the present time, one of the most extensively used methods for climate forecast is information mining. Data mining method of dissecting information factually and infer such guidelines that can be utilized for expectations. The main outcomes are rainfall, maximum temperature, minimum temperature and relative humidity.

This paper uses some process. We select criteria and every criterion have some sub criteria. We can calculate data selection, data preprocessing, data visualization, and data analysis and data classification.

2.4 Scope of the Problem

- As there is less previous documentation or research paper, so data was not much more.
- As raw material for our research as our research data was kind of sources from Bangladesh Metrological Department (BMD), so many problems have been faced to collect data from there.

- Our data set has 65 years stations wise weather of data, so many problems have been faced to collect data from there.
- The implementation of this system was quite difficult.
- All data are not actual or fixed. Those are changed day by day for this reason finding the accuracy is quite difficult.
- Processing of the raw data and the calculations of the equations in this research were lengthy.
- Changes in the surface highlights of a zone influence can numerous variables. For instance, they can influence precipitation, temperature, and even breezes. Enormous networks can likewise make it hard for meteorologists to precisely anticipate limited scope weather occasions.
- Changes in the surface highlights of a territory influence can numerous components. For instance, they can influence precipitation, temperature, and even breezes. Enormous networks can likewise make it hard for meteorologists to precisely foresee limited scope weather occasions.

2.5 Challenges

- We need to contact with some department such as Bangladesh Metrological Department (BMD to collect data, also we have to understand that what actually we want to do in our research, what the importance of our work is.
- The most significant difficulties are successful handling of availability, accessibility and dramatically examination of significant information.
- This dataset taken from different kind of sources from Bangladesh Metrological Department (BMD). Our data set has 65 years stations wise weather of data and data collect was very difficult.
- Many data are secured by weather stations.
- Attribute selection is a big challenge.
- Each decision is given an importance.
- All data are not actual or fixed. Those are changed day by day.
- Data synchronization was also taking time to plan.
- Select decision framework in this system is a challenge.
- There are not many works have been made with large information investigation particularly in the weather figure.
- Finally, we said that actual data collection is main challenge.

CHAPTER 03

Research Methodology

3.1 Introduction

Computer science has lots of algorithm, we choose the linear regression algorithm for prediction the weather. This report discusses the theory and predict the weather accuracy like “Max Temp”, “Min Temp”, and “Relative Humidity”, “Rainfall”. Linear regression method key terms has response, independent variable, intercept, record, regression coefficient, fitted values, residuals and least squares. Weather is a broad domain forecast, which is extremely necessary for humans, however, only limited research for this domain. To work on forecasting weather conditions we are doing a study on it. Some of the work already has some valuable observations.

3.2 Data Collection Procedure

Our dataset collection from several datasets. Maximum data collection from different combined and sources together. This dataset taken from different kind of sources from Bangladesh Metrological Department (BMD). Our data set has 65 years stations wise weather of data. Stations are “Barisal”, “Bhola”, “Bogra”, “Chadpur”, “Chittagong”, “Chuadanga”, “Comilla”, “Cox’s Bazar”, “Dhaka”, “Dinajpur”, “Faridpur”, “Feni”, “Hatiya”, “Ishurdi”, “Jessore” “Khepupara”, “Khulan”, “Kutubdia”, “Madaripur”, “Maijdee Count”, “Mongla”, “Mymensingh”, “Patuakhali”, “Rajshahi”, “Rangamati”, “Rangpur”, “Satkhira”, “Sitakunda”, “Srimangal”, “Syedpur”, “Sylhet”, “Tangail”, and “Teknaf”. Dataset attributes has “Station Names”, “YEAR”, “Month”, “Max Temp”, “Min Temp”, “Rainfall”, “Relative Humidity”, “Wind Speed”, “Cloud Coverage”, “Bright Sunshine”, “Station Number” etc. But Our focusing attributes are “Max Temp”, “Min Temp”, “Rainfall”, “Relative Humidity”. We take dataset from Kuggle (https://www.kaggle.com/emonreza/65-years-of-weather-data-bangladesh-preprocessed) which is open source for any kind of research project area.

3.3 Proposed Methodology

Weather is a broad domain forecast, which is extremely necessary for humans, however, only limited research for this domain. To work on forecasting weather conditions we are doing a study on it. Some of the work already has some valuable observations. We used the Jupiter notebook to determine our contribution to this study.

In this paper we use advanced technology for weather forecasting prediction. In this paper we use linear regression which is used in Jupiter Notebook tool for analyzing the accuracy of data. We take dataset from Kuggle (https://www.kaggle.com/emonreza/65-years-of-weather-data-bangladesh-preprocessed) which is open source for any kind of research project area. To explain the ability of our proposed work, performance is compared with existing work. Let us discuss the submitted process in detail.

3.3.1 Data Mining Tool

At present world is become digital and world is dependent on technology. In this Technological age there are different type of data mining tool which is used for data analyze. All of this tools are available. Analyzing data for weather prediction we collect data from kuggle which is open source for any kind of research area. We use Jupiter notebook for data analyze. We can calculate data selection, data preprocessing, data visualization, data analysis, classification. Based on characteristics and their relationship from the data Jupiter Notebook predict the possible attribute and possible assumption for forecasting weather. Jupiter Notebook can support multiple data format and also numeric and nominal attribute. Jupiter Notebook can analyze the data directly from dataset and response to the result quickly. Using linear regression we can merge data in table and separate software by Jupiter Notebook. The Jupiter notebook is easy to use and learn which is considered to be one of the best mining tools in real-time situations.

3.3.2 Data Preprocessing

Data preprocessing is a data mining method that includes changing raw data into a justifiable format. Real data is frequently fragmented, conflicting, as well as ailing in specific practices or drifts, and is probably going to contain numerous mistakes. Data preprocessing is a demonstrated strategy for settling such issues. In reality data are for the most part incomplete: lacking qualities, without specific credits of interest, or containing just total data. Containing inconsistencies in codes or names. In data mining preprocessing is the initial step advance.

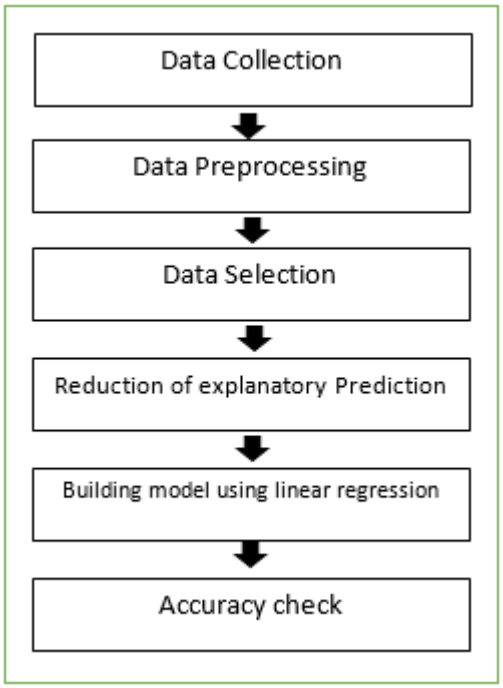


Figure 3.1: Data Preprocessing Technique

Which saves times and accommodating in extraction of details. In detail during preprocessing, it deficient, conflicting and blunder data are written down also as eliminated. Because of this process, an unmistakable useful data is gotten and will be used for additional process.

3.3.3 Linear Regression

Linear regression works for making relationship between two variables by consistent a linear equation to analyst data. One variables is explanatory other variables is dependent variable.

Linear regression is equation $Y = a + bX$

Here,

Y = Dependent

X = Independent (explanatory)

Line of slope is b and a is intercept

3.3.4 Regression and Prediction Model

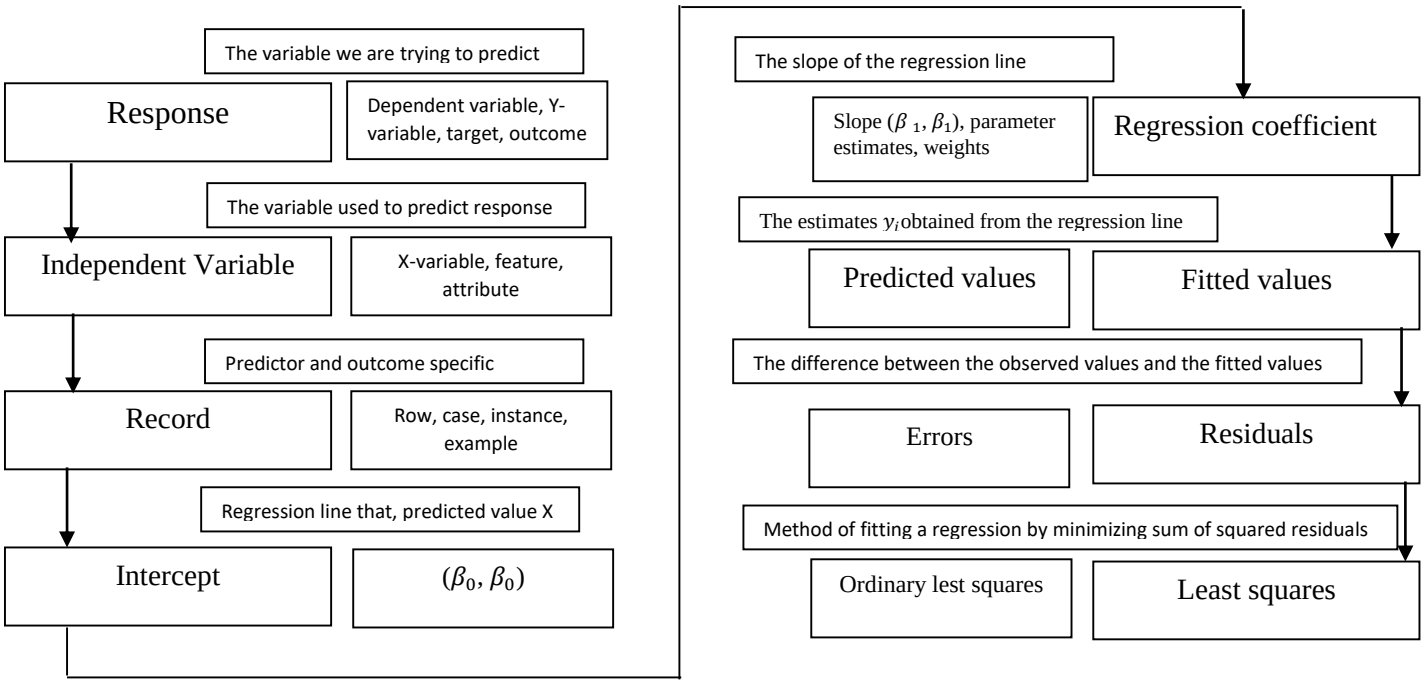


Figure 3.2: Regression and Prediction Model

3.3.5 Logical Data Model

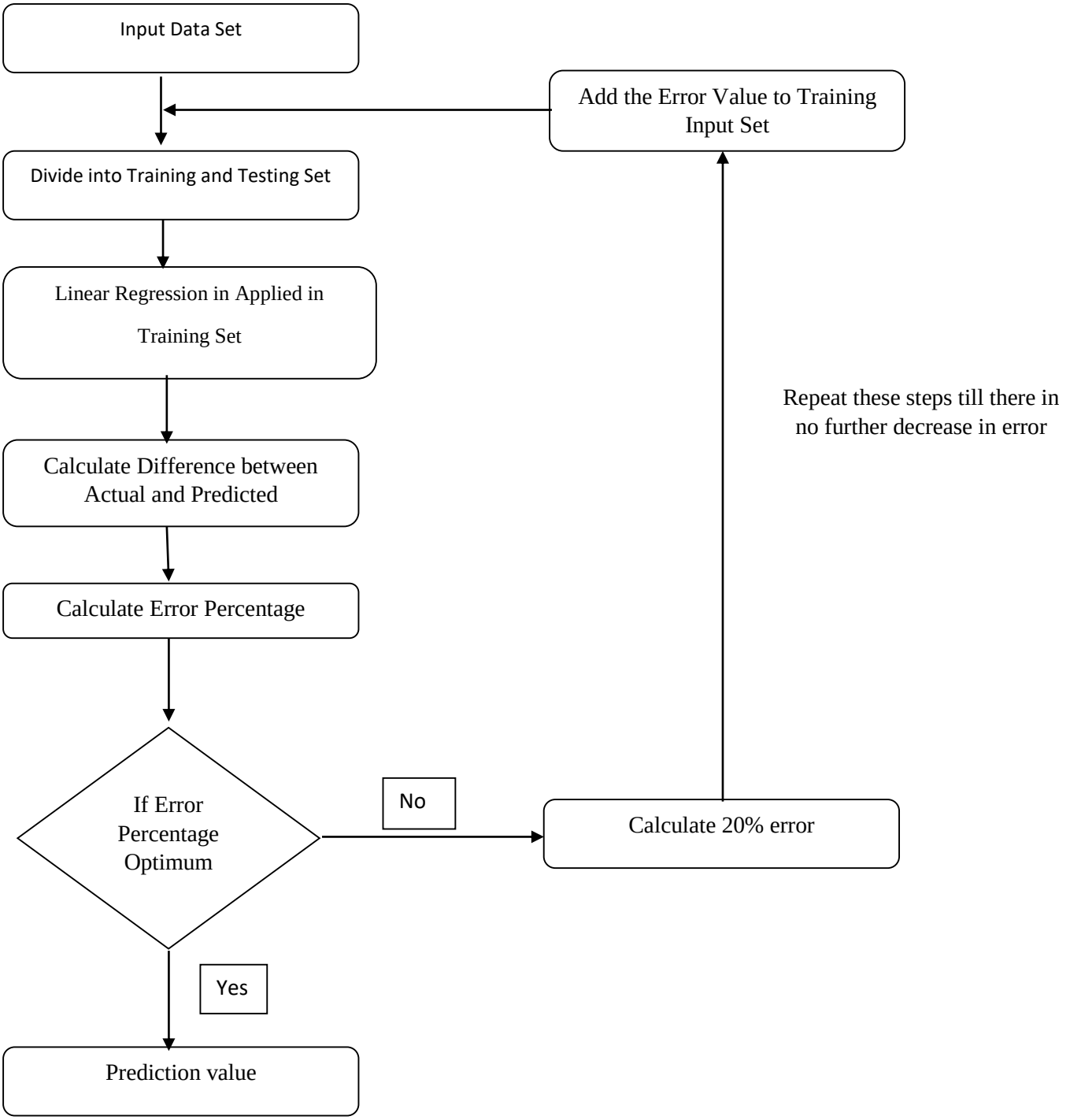


Figure 3.3: Logical data Model

3.3.6 Direct Regression Method

The direct regression method is known as ordinary least squares estimation. Let a set of n time monitoring on (x_i,y_i), i=1, 2,….,n which gratify direct linear regression model y=β_0+$\beta_1 x_i$+ε_i, (i=1,2,…,n).

Direct regression method can minimize the sum of squares

S (β_0, β_1) = $\sum_{i=1}^{n} \varepsilon_i^2$= $\sum_{i=1}^{n}$ (y_i - β_0 -$\beta_1 x_i$) 2

With the respect of β_0 and β_1.

Partial derivations of S (β_0, β_1) with the respect of β_0.

$$\frac{\partial S\,(\beta_0,\beta_1)}{\partial \beta_0} = -2\sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i).$$

And the partial derivations of S (β_0, β_1) with the respect of β_1.

$$\frac{\partial S\,(\beta_0,\beta_1)}{\partial \beta_1} = -2\sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)x_i.$$

The solution of the equation are obtained by β_0 and β_1

$$\frac{\partial S\,(\beta_0,\beta_1)}{\partial \beta_0} = 0$$

$$\frac{\partial S\,(\beta_0,\beta_1)}{\partial \beta_1} = 0$$

3.3.7 Algorithm Decision Structure

Algorithm Structure Part shows us How Linear Regression work and how the output comes to it. Basically we can start our work in Jupiter notebook from dataset. When linear Regression is applied on dataset for weather prediction and how it works. Linear regression analyze the data according to the data structure and information which is given in dataset. Generally in Jupiter notebook there are some rule and relationship and a tree structure formed. Based on the data there are four kind of prediction such as rainfall, max temperature, min temperature and humidity. As can be seen from the figure there is a relationship between rainfalls humidity, min temperature, max temperature and rainfall. In the first row of graph in the figure shows the relationship between rainfall and humidity, min temperature, max temperature and rainfall. The rest of the graph is like this kind of relationship between them. Here second row of this graph shows us a relation between max temperature and humidity, min temperature, max temperature, rainfall. Third row of this graph shows us a relation between min

temperature and humidity, min temperature, max temperature, rainfall. Like this fourth row of the pair plot graph show that humidity in a relation with humidity, min temperature, max temperature and rainfall. This is how linear regression algorithm work for weather prediction.

3.3.8 Pair plot graph between each Entities

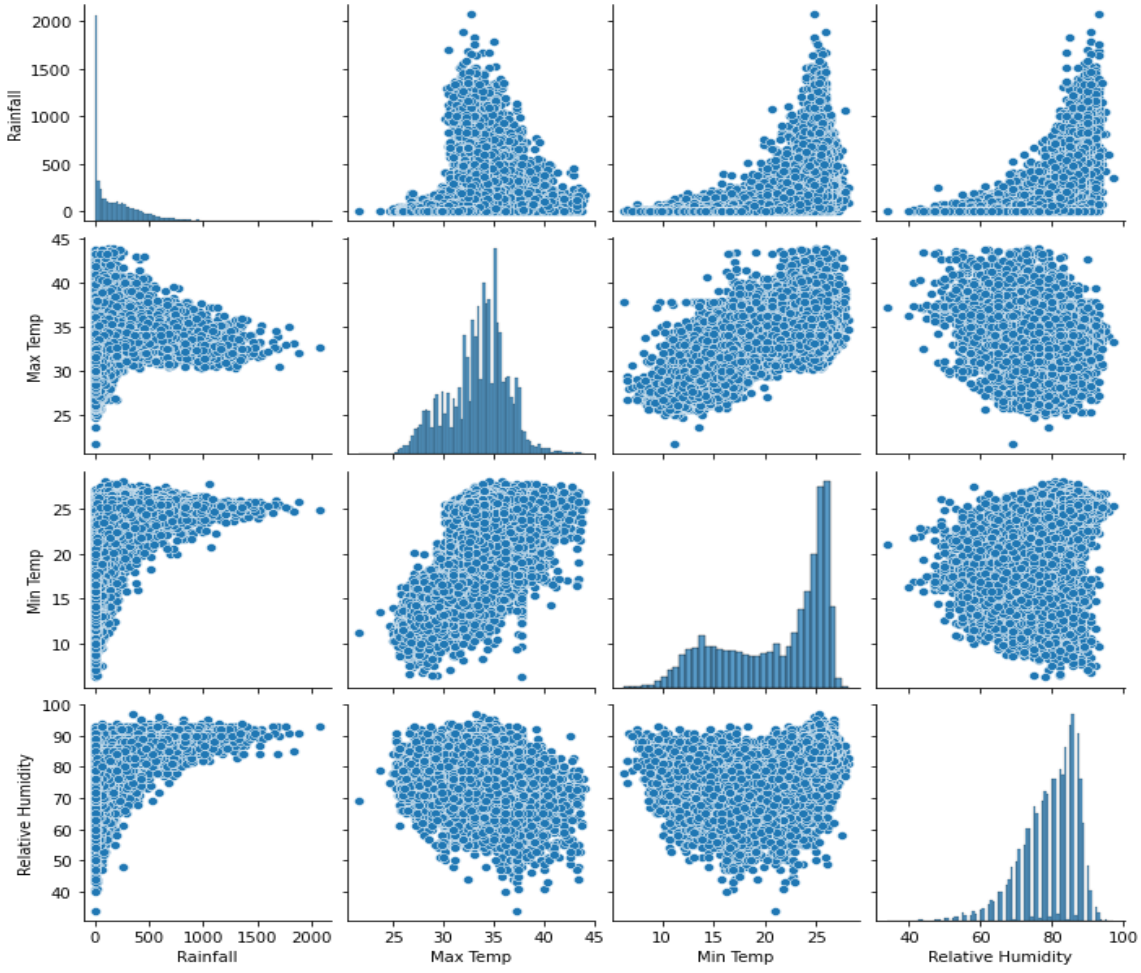


Figure 3.4: Pair plot graph between each Entities

CHAPTER 04

Implementation and Result Analysis

4.1 Introduction

In this research, we select the linear algorithm for finding prediction accuracy of data set. Here, Rainfall depends on Relative Humidity, and Relative Humidity depends on Max and Min Temperature. So, When the Max Temp and Min Temp will change then Relative Humidity and Rainfall will effected. We collect our dataset from Bangladesh weather perspective and target the 4 attributes of the dataset and find the prediction accuracy.

4.2 Linear Regression Algorithm Applying Procedure

First we will utilize the temperature data marker. By then the information marker will be cleared and all lacking information will be taken out. Since we have really experienced methods of figuring out significant pointers (features), how might we make a gauge model and utilize the science-picture to test its capacity to foresee the normal temperature. Psychite-Learning is an eventually elaborate AI library that is ordinarily utilized in both the mechanical and insightful networks. Something outstanding about Psychit-Learn is that it has an entirely steady programming interface of "dominant part", "Persian" and "test" crosswars that makes it uncommonly direct to utilize. Notwithstanding this anticipated programming interface plan, Psychit-Learn will give extra assistance to many AI adventures, including a couple of extra gadgets to deal with essential data. The accompanying advance is to make a reconnection show utilizing the arrangement dataset. To do this I will import and utilize the Direct Relays class from the sklearn. linear_model module. As referenced before, Skykit-Learn utilizes various highlights and a fit programming interface in various manners that make the library particularly straightforward.

4.3 Experimental Result and Analysis

Finding the prediction accuracy using linear regression algorithm working in Jupyter notebook. Jupyter notebook is an open source platform for any kind of research area. Here, we using two libraries are Numpy and Pandas. Four type of attributes accuracy we are finding like “Rainfall”, “Max Temp”, “Min Temp”, and ‘Relative Humidity”.

We find this accuracy from Bangladesh weather data perspective. We find months and annuals accuracy of weather prediction using linear regression from 1949-2013.

Table 4.1: Accuracy Table

Attributes Name	Accuracy
Max Temp	71%
Min Temp	86%
Relative Humidity	67%
Rainfall	64%

"Relative Humidity" accuracy depends on "Max Temp" and "Min Temp" accuracy. "Rainfall" Accuracy depends on "Relative Humidity". The expected result we get are plotted on the percentage. This result can show the prediction of weather forecasting of Bangladesh.

4.4 Description of Our Work

We use a modified variation of Direct Relapse to play out the estimate of temperature in our system. The system of this strategy is explained in these after advances:

1. From the start we gather the information and afterward examined them. We make a month long arrangement to how to execute our task.

2. The readiness and test data are outlined from the data instructive lists.

3. We will clean all the information and afterward will stack the dataset.

4. At that point we will sum up the dataset.

5. We will imagine the information to show our information current shape.

6. At that point we will assess our direct relapse calculation to anticipate the temperature.

7.We will approve our dataset, we will part our informational collection into preparing and testing dataset.80% of our information we will utilize train our dataset and 20%of our information will be utilized for the testing the approval of our dataset.

8. We will us 10 cross crease approval for assessing exactness.

9. We will utilize straight relapse test calculation for our information precision forecast.

CHAPTER 5

Conclusion & Future Work

5.1 Conclusion

Temperature is the critical purpose behind countless the calamitous occasions like glint floods, dry seasons, waves. In order to keep these trademark calamities, we should be prepared to anticipate the explanation behind the source. The proposed system can be used to check the temperature over the necessary time frame with the objective that the individual experts can take careful steps to keep the loss of life and property. The proposed structure uses changed straight backslide approach to manage predict the temperature that has less slip-up rate than stood out from most data mining strategies like clustering, back spread which surrenders the added characteristics rather than measure regards.

In this paper, we discussed the latest 64 years (1949–2013) of environmental change in Bangladesh reliant on temperature and precipitation data. The based future desire for ecological change for the 2013–2020 period was surveyed and attested a particularly strong progressing natural change in Bangladesh subject to temperature and precipitation changes. This examination similarly supports more restricted time scale air diversions and can without a doubt be applied to more limited climate data. In any case, this model is only established on a quantifiable philosophy, it is limited the extent that crazy and inconsistent events since this can't envision phenomenal events as a result of any external force or trademark events. For example, if a strong overall external intensity of the climate, like a critical volcanic discharge occurs, and this might be a sound inspiration to discredit the measure for that particular time. Therefore. For the climate of Bangladesh, the movements of temperature reflect a warming all in all and since 1949, the air of Bangladesh is warming at much higher movement of overall typical warming (0.20 versus 0.13 °C consistently). Desire for the temperature shows that in the hour of 2013–2020, the mean of the greatest temperature is 33°C, exhibiting the temperature will be 1.0 °C more blazing in Bangladesh by 2020 diverged from 1949. An essentially more vital rising is foreseen for the mean of greatest temperature (33.4 °C) than the mean of least temperature (21.14 °C) temperature. The base temperature warmed more in the northern, northwestern, northeastern, central and central southern parts while the most extraordinary temperature warmed more in the southern, southeastern and northeastern parts during the 1949–2013 period. In a segment of these

parts, the mean least and mean most extraordinary temperatures rose by more than 2.0 °C (0.50 °C consistently), which will likely present troubles to the general population in those bits of the country. Of course, the fundamental results are precipitation, greatest temperature, least temperature and relative dampness. The ultimate result of precipitation in our exploration were related to the exactness of 64%. The most extreme temperature gives precision in our examination is 71%. The exactness of least temperature is 86% and the relative stickiness precision in this paper is 67%. High precipitation changeability is a marker of drought, and henceforth, the zones under amazingly high vacillation with low precipitation, particularly the north western regions, are slanted to dry season risk. Desires for precipitation reveal that declining precipitation will continue and a drying condition will suffer during 2013–2020 (153 mm diminishing of yearly precipitation), especially during the pre-and post-storm seasons. Spatial instances of example and vacillation of temperature and precipitation show that the northwestern, western and southwestern bits of the country are more feeble to natural change with respect to rising temperature,

High alterability and storm deficiencies, particularly for the pre-and post-rainstorm deluges. It is typical that. This assessment won't simply help with depicting legitimate methodologies and expecting to fight the impact of natural change in Bangladesh yet also help to understand the nearby ecological change in this bit of the South Asia. The example direction, significance and spatial models perceived for both temperature and precipitation may similarly give strong information on a risky environmental deviation on a nearby/country level scale. Improved appreciation of progressing natural change helps with clarifying the impacts and shortcoming of the local people to execute the most fitting practices to adjust to ecological change and manage the changing condition in an unrivaled way.

This paper explores the concerns of climate forecasters under big data analysis. For perfect climate determining, we have applied our proposed direct Linear Regression algorithm to the atmosphere dataset which is taken from an open source Kuggle atmosphere dataset from the authority Kuggle site. The choice to assemble and develop a tree model dependent on related data works with tree standards. To improve the exploration the perception is finished with Kuggle, a high level data mining device that encourages direct access from data sets. This proposed technique is reasonable for looking at the verifiable information. The proposed structure is successful in breaking

down the relationship of climatic boundaries, for example, temperature, wind speed, stickiness, and so on an exhibition correlation is made with the direct Linear Regression algorithm to demonstrate the effectiveness of our proposed cycle. The outcomes acquired are far superior to the current strategy demonstrating the exactness of the expectation by the straight relapse calculation. In this work atmosphere chronicled information is utilized distinctly for the examination of atmosphere information. Later on, the calculation for breaking down current information was changed.

We have chosen a method for predicting rainfall after analyzing the rainfall dataset of Bangladesh which is generated by some data mining techniques such as first analyzing the related relationship, then applying regression analysis. The effect of rainfall is incredible in Bangladesh, yet additionally in farming and economy everywhere on the world. With the goal that we can predict rainfall later on knowing the reasons for atmosphere are very effective for farmers in their agriculture. This is the solitary conjecture of precipitation yet not exact because of climatic elements. As we realize that atmosphere factors change for various reasons and here. We have utilized a few components so other leftover elements can influence precipitation.

5.2 Further Work

We use temperature data for forecast Relative Humidity and Rainfall of weather in Bangladesh. We use the maximum, minimum, dew point temperature of a particular district for our mounts and annuals analysis from 1949-2013. We only use linear regression techniques; we can also support multiple regression, vector machines and artificial neuron networks. We can then make a comparison as to which algorithm is the best suit for our data. In the future we can improve our system by adding many more factors to make wind speed, cloud coverage and bright sunshine forecasts and our projects more accurate to weather forecasts.

Artificial intelligence Might Be the Future for Weather Forecasting. Weather prediction has made some amazing progress in the course of recent years. AI help improve meteorologist’s capacity to anticipate future weather prediction.

References

[1] ,. A. L. H. Tamjid Rahmana, "A Fuzzy-Neuro Based Weather Prediction System for," ScienceDirect, vol. 36, p. 606 – 611, 2014.

[2] M. R. Dastagir, "Weather and Climate Extremes," ScienceDirect, pp. 49-60, 2015.

[3] N. N. MAI Navid, "Multiple Linear Regressions for Predicting Rainfall for Bangladesh," SciencePG, vol. 6, no. 1, pp. 1-4, 2018.

[4] D. S. A. A. R. M. a. M. M. R. Dipak Kumar, "Impact of climate change on rainfall in Northwestern Bangladesh using multi-GCM ensembles," RMetS, p. 1395–1404, 2013.

[5] S. S. A. A. E. N. H. S. C. D. Taohidul Islam, "Monthly Weather Forecasting through ANN Model: A Case Study in Barisal, Bangladesh," ResearechGate, vol. 5, no. 6, pp. 2278-1021, 2016.

[6] R. K. P. R. R. a. G. R. S. Md Rashid Mahmood, "A Novel Approach for Weather Prediction Using Forecasting Analysis and Data Mining Techniques," ResearchGat, pp. 479-489, 2019.

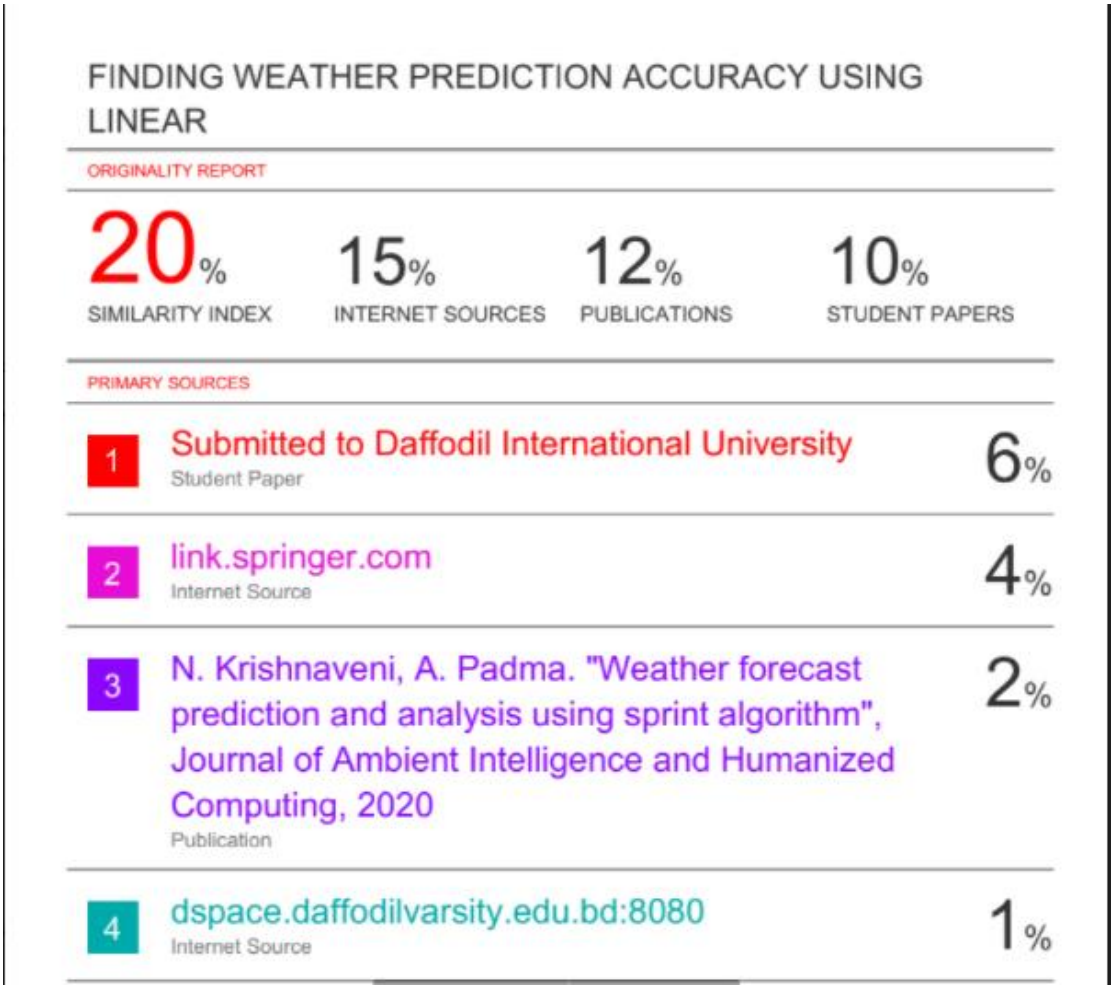
[7] 2. &. H. L. Md. Rejaur Rahman1, "Climate change in Bangladesh: a spatio-temporal analysis and simulation of recent temperature and rainfall data using GIS and time series analysis model," ResearchGate, 2015 .

[8] S. S. a. S. B. Anisha Datta, "Complete Statistical Analysis to Weather Forecasting," Springer, vol. 999, pp. 751-763, 2019.

[9] W. Y. G. X. D. G. a. N. G. Jarrett Booz, "A Deep Learning-Based Weather Forecast System for Data Volume and Recency Analysis," IEEE, pp. 2325-2626, 2019.

[10] N. K. A. Padma, "Weather forecast prediction and analysis using sprint algorithm," Springer, 2020.

[11] S. K. D. M. S. S. C. A. Fahad Sheikh, "Analysis of Data Mining Techniques for Weather Prediction," ResearchGate, p. 9(38), 2016.



FINDING WEATHER PREDICTION ACCURACY USING LINEAR

ORIGINALITY REPORT

20%	15%	12%	10%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to Daffodil International University Student Paper	6%
2	link.springer.com Internet Source	4%
3	N. Krishnaveni, A. Padma. "Weather forecast prediction and analysis using sprint algorithm", Journal of Ambient Intelligence and Humanized Computing, 2020 Publication	2%
4	dspace.daffodilvarsity.edu.bd:8080 Internet Source	1%