WEATHER FORECASTING USING DATA MINING

BY

MEHEDI HASAN SAGOR

ID: 151-15-4933

AND

TARIKUL ISLAM

ID: 151-15-5144

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

Supervised By

Most. Hasna Hena
Senior Lecturer
Department of CSE
Daffodil International University

Co-Supervised By

Mr. Saiful Islam
Lecturer
Department of CSE
Daffodil International University

DAFFODIL INTERNATIONAL UNIVERSITY

DHAKA, BANGLADESH

DECEMBER 2018

APPROVAL

This Project/internship titled **"Weather prediction using data mining**", submitted by **Mehedi Hasan Sagor**, ID No: 151-15-4933 and **Tarikul Islam** ID No: 151-15-5144 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 9th December 2018.

BOARD OF EXAMINERS

_____

Dr. Syed Akhter Hossain — **Chairman**
Professor and Head
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

_____

Dr. Sheak Rashed Haider Noori — **Internal Examiner**
Associate Professor & Associate Head
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

_____

Md. Zahid Hasan — **Internal Examiner**
Assistant Professor
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

_____

Dr. Mohammad Shorif Uddin — **External Examiner**
Professor
Department of Computer Science and Engineering
Jahangirnagar University

DECLARATION

We hereby declare that, this project has been done by us under the supervision of **Most. Hasna Hena, Senior Lecturer, 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:

_____

Most. Hasna Hena
Senior Lecturer
Department of CSE
Daffodil International University

Submitted by:

_____

Mehedi Hasan Sagor
Department of CSE
Daffodil International University

_____

Tarikul Islam
Department of CSE
Daffodil International University

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/internship successfully.

We really grateful and wish our profound our indebtedness to **Most. Hasna Hena**, **Senior Lecturer**, Department of CSE Daffodil International University, Dhaka. Deep Knowledge & keen interest of our supervisor in the field of "*Data mining*" to carry out this project. His 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 Most. Hasna Hena, Mr. Saiful Islam, 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

Forecasting weather is one of the greatest challenges in meteorological department. Weather prediction is necessary so as to inform people and prepare them in advance about the current and upcoming weather condition. This helps in reduction in loss of human life and loss of resources and minimizing the justification steps that are expected to be taken after a natural disaster occurs. So we want to develop a system which can predict the weather condition including temperature, humidity and wind speed. We study various techniques and algorithms that are likely to be chosen for weather prediction and highlight the performance analysis of those algorithms like as ANN, Linear Regression, Decision tree, Time series analysis. From those algorithms we selected Linear Regression as the outcome is very efficient close to real time data. These techniques are discussed that are used to boost the performance of the application. After a comparison between the data mining algorithms and corresponding collective technique used this algorithm to boost the performance, a classifier (Decision tree) is obtained that will be further used to predict weather outcome. Applications used to Predict and forecast the weather condition of specific region based on the available pre historical data which helps to save resources and prepare for the changes forth coming. And finally we develop a web based system using these algorithms to visualize and so that people can use it as a predictive system. Our system has some limitation too such as real time data set in weather for Bangladesh is not available everywhere. Therefore, we think this system will help efficiently in the era of weather prediction.

TABLE OF CONTENTS

CONTENS	PAGE
Board of examiners	ii
Declaration	iii
Acknowledgements	iv
Abstract	v
Chapter 1: Introduction	**1-6**
1.1 Introduction	1
1.2 Motivation	2
1.3 Rationale of the Study	3
1.4 Research Questions	4
1.5 Expected Output	5
1.6 Report Layout	6
Chapter 2: Background	**7-11**
2.1 Introduction	7
2.2 Related Works	8
2.3 Research Summary	9
2.4 Scope of the Problem	10
2.5 Challenges	11
Chapter 3: Research Methodology	**12-20**
3.1 Introduction	12
3.2 Research Subject and Instrumentation	13-15
3.3 Data Collection Procedure	16-17
3.4 Statistical Analysis	18-19
3.5 Implementation Requirements	20

Chapter 4: Experimental Results and Discussion **21-29**

4.1 Introduction 21

4.2 Experimental Results 22

4.3 Descriptive Analysis 22-28

4.4 Summary 29

Chapter 5: Summary, Conclusion, Recommendation and Implication for Future Research **30-31**

5.1 Summary of the Study 30

5.2 Conclusions 30

5.3 Recommendations 31

5.4 Implication for Further Study 31

Appendix **32**

References **33**

LIST OF FIGURES

FIGURES	PAGE NO
Figure - 3.4.2.1 Decision tree	19
Figure - 4.3.1: Autumn season dataset	23
Figure - 4.3.2: Spring season dataset	24
Figure - 4.3.3: Summer season dataset	25
Figure - 4.3.4: Winter season dataset	26
Figure - 4.3.5: winter season using linear regression	27
Figure - 4.3.7: Summer season using linear regression	28

LIST OF TABLES

Table - 3.3.1: Attributes of prediction system 17

Table - 4.2.1: Seasonal data as input 22

CHAPTER 1
Introduction

1.1 Introduction

Weather prediction is a challenging problem in meteorological department since years. Even after the technological and scientific improvement, the accuracy in prediction of weather never enough. In current date this field remains as a research topic in which scientists and mathematics are working to make a model or an algorithm. which will exactly predict weather. There need huge improvements in the sensors and that dependable for recording the data from the environment and cancel the noise present in them. We make new models have been planned with different attributes related to weather to compose accurate prediction. Now one of the most commonly used data mining techniques for weather prediction. Data mining gives a way of analyze data statistically and extract. it derives many rules that can be used for predictions. Now it is used in many fields. There is stock market, sports, banking section and many other sectors. Scientists have present realized that data mining can be used in the field of weather prediction as well. The basic thing of data mining is data itself. It is defined as raw set of information which can be used as Keywords: Data Mining, Decision Tree, Ensemble Technique, Pre-Processing, Weather Prediction to extract meaningful information depending upon the requirements of the application. Data can be stored in an organized manner which is known as database. The word data mining refers to the methods that are used to extract the required material from the given set of data that might be useful for statistical purpose or making predictions by learning patterns in data and correlation between different parameters. Data mining has now been assumed by many domains such as sports, banking, meteorological department, etc., and because of this, scientists, mathematicians and researchers have come up with a wide range of algorithms for finding solution.

1.2 Motivation

Our real time motivation when I see my mobile weather application then we see that a list of one weak weather Prediction are there. I was so more curious about it. And we want to know how it is work, how we can make a better on of the application. Sowed work with the data mining technique using Weka.

In our future life we want to work with big data analysis. So, we want to make something that help to success in life.

Data mining is the field where huge amount of data is collected and being processed to extract some useful data i.e. information. Everyone wants the concise and precise information which is possible through it, as it is not an easy task but it becomes possible through series of process and science.

1.3 Rationale of the Study

The aim of data mining is for finding structure through unstructured data. Data mining pull out meaning from noisy data and it's found out patterns in apparently with random data. It also uses all in order to better understand trends, patterns, correlations, and at last predict the customer behavior and market and competition trends, so there for we get better capable algorithm to make the better prediction of weather.

In the history of weather prediction, attempts have often been made to devise numerical and objective methods for producing the forecast. The understandable ultimate goal of forecasting weather is to enable the forecaster to increase the accuracy of forecasts made routinely. We analysis data using mining tools and algorithm to find the better outcome. The results of the mining are very useful and accurate which gives the way to appliance that information in many areas and contributes in development.

1.4 Research Questions

We study with the weather prediction that's why we study a lot of prediction algorithm. For a batter prediction result we need to find collect a huge amount of data. Data collection is challenge but we collect it properly. The more we collect data the more we gets better prediction. We are wanting to predict the weather temperature, humidity, wind speed and the outlook of next day.

The questions we have faced:

1. Collection of raw data.
2. Validity of data.
3. Select the algorithm for prediction.
4. Get the expected output.
5. Errors of the algorithms.

1.5 Expected Output

- Weather forecasting entails predicting how the present state of the atmosphere will change.
- Temperature, humidity, wind speed of the next day will predict.
- We can easily find out Weather condition by using this system.
- It provides the business with valuable information that in business, that helps to make decisions about the future production of the organization.
- This method can be used in Air Traffic, Marine, Agriculture, Forestry, Military, and Navy etc.

1.6 Report Layout

Chapter 1: Is all about Introduction of our research.
In this chapter Introduction, Motivation, Rationale of the Study, Research Questions, Expected Output, Report Layout are described.

Chapter 2: Is all about Background of the research. In this chapter Related Works, Research Summary, Scope of the Problem and Challenges are described.

Chapter 3: Is all about Research Methodology of the research.
In the chapter Research Subject and Instrumentation, Data Collection Procedure, Statistical Analysis, Implementation Requirements are described.

Chapter 4: Is all about Experimental Results and Discussion of the research.
In the chapter Experimental Results, Descriptive Analysis and Summary are described.

Chapter 5: In this chapter Summary, Conclusion, Recommendation and Implication for Future Study are described.

CHAPTER 2
Background

2.1 Introduction

Before starting this choice, we study a lot about weather forecasting that needs predicting weather of our country, its necessity, existing systems and a lot of marketing analysis as well as the challenges of the thesis. As we think, Weather forecasting system is needed for convenience and safety. This system is invented to keep safe our life and goods. So we have to prepare our self-first for this and that's why we need background studies for successfully complete our recourse.

2.1.1 Ancient forecasting method:

The time of 650 BC in Babylon they are predict weather by cloud prototypes. Then the time of 350 BC the Aristotle selected or weather prediction with the patterns Meteorologica[1]

2.1.2Modern methods:

The now days the weather prediction started in 1835. Before that days the isolated weather information's travel with approximately 100 miles in a day (160 km/d), but was more naturally 40–75 miles in a day (60–120 km/day).

2.2 Related Works

In scope of climate expectation framework there are numerous analysts have attempted by utilizing a few data mining strategies. There is good amount of literature is check which basing on Neural Networks approach [2]. These are failed to find out the irregular patterns of the weather. Those models are based on the Support Vector Machine and the stochastic methods come close to also shown in writing [3], The algorithm of Genetic which based on the advanced with Neural Networks are also projected with it. [4]. With the models recommended for the researchers, [5]. The successful prediction study could not accomplish for the difficult data systems of the weather unqualified and constant model of the weather. It depends on raucous data and high dimensionality of data for prediction. for that reason, successful model for forecasting the weather are examining. Now with the help of powerful super computer, scientists are running vast amount global weather simulation. For the reason scientists ran 10,240 simulations through the model and the global environment divided with 112-km divisions. Then they used data adjustment and numerical methods for come up with a model that closely apt the real data for historical time era. The time of 1 November to between 8 November, 2011. [6]. Many international organizations lead by Takamasa Miyoshi of RIKEN Advanced Center for Computational Science (AICS) they used the great super computer and highly developed radar surveillance data to correctly predict the happening of driving rains of the localized areas. [7]

2.3 Research Summary

We study many research papers, in those we get many information about the weather prediction system of modern world. In Bangladesh this type of research work is very rear. So, we choose the topic. In much research paper we have seen that they work Neural Networks approach for weather prediction. The of use prediction analysis con not done for the complex data systems of the weather that definite and constant model of the weather. The raucous data and high length of data. So, we divided dataset in four sub datasets according to season. We study various algorithms, and select the best of them for Bangladesh historical dataset.

2.4 Scope of the Problem

Here we in this paper a weather prediction model is assemble for predicting the weather successful. Our model the weather data with the form of time series is careful and those data are transformed with information where useful information’s are derivative with the help the concepts of data mining techniques. The data mining system are used for discover the hidden patterns and connect a linkage with the different attributes which associated with the weather conditions. Now, the data exhibits a series over a period of time and weather prediction can be well consider by using time series mining. Some automated prediction systems in the market are run away for giving false system to customer. There are many prediction systems in the market which don't have a features and suitable interface designs. So, people think every prediction system are same. In order to remove this thinks our system will provide appropriate levels of prediction while occupants are away from their residence. This system will particularly helpful for those who put in irregular hours at their place of employment those who vacation on a regular basis.

Finally, bad prediction system is harmful for the people asset; it's become a drama with people trust. So, we take it as a scope for working on this platform.

2.5 Challenges

While working on this we face many challenges which is too tough to solve but we solve those challenges. Because without solve those challenges we can't reach to our goal. We took those challenges very seriously and solve those challenges.

There are some major challenges:

1. Making the whole system user friendly.
2. Huge collection of data.
3. Different kinds of algorithm analysis.
4. Weka loading problem.
5. Taking data perfectly.
6. Select the better algorithm on from all those algorithms.

CHAPTER 3
Research Methodology

3.1Introduction

The first work in predicting the weather is to collect data. We must to know the history of weather in the area because one of the key values of weather forecasting is that past patterns can indicate future events. Conflict of meteorological aspects, mainly those weather parameters which are responsible for Local monsoonal precipitation forces us to develop an approach which can recognize such an inconsistent pattern and use it for future prediction of weather which will be very much realistic.

The easy policy for determining the climate, ingenuity, works with the present conditions to figure the situation the next day. This can be a substantial method to be hopeful of the environment when it is in a long-term state, such as, amid the year season divided in four seasons in the tropics. This strategy for gauging firmly relies on the nearness of a sleeping type of weather design. Along these lines, when in a fluctuating climate design, this strategy for estimating ends up erroneous. It very well may be helpful in both short-range conjectures and long-range gauges. The primary Problem of this vital work is to test the capacity of the Model to look forward to typical weather marvels for lesser blunder for the long-term evaluation.

3.2 Research Subject and Instrumentation:

We are working with weather forecasting analysis using data mining technique.
We worked with Weka as data mining tools. We use linear regression and decision tree algorithms in it. In weka we add forecast from weka tools. Then we apply there linear regression algorithm. then we get the prediction. Then we classify the dataset in weka classification with decision tree algorithm. Which, give the output of the tree that happened past. Depends on the tree we give the outlook result.

With Microsoft excel we implement our data for the prediction. Then by weka we get the outlook from decision tree and temperature, humidity, wind speed from linear regression analysis.

3.3 Materials and Methods

3.3.1 Linear regression:

Regression is a system of modeling and that target value based on autonomous variable. This technique is mostly used for prediction and finding output. Its gives real connection between two factors. Relapse process dependent on the quantity of autonomous factors [8]

Regression tries to find out the strength of the relationship between one dependent variable usually signify by Y and a series of other changing variables known as independent variables. Here regression there are only two variables where one is the dependent variable and other is the independent variable. the relation among them is of kind as below.

This is known as the deterministic model -

$$Y=\alpha + \beta X+e$$

Here

Y= Dependent variable for regression.

X= independent variable for regression.

α, β = Regression parameters for regression.

E = Error in regression

E is distributed normally zero.

In Multiple regressions there are more than two independent variable and one dependent variable and the equation look like this:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_n x_n$$

Where:

β =Model coefficients.

β0 = Intercept

β1 = Coefficient for X1 (the first feature)

βn = Coefficient for Xn (the nth feature)

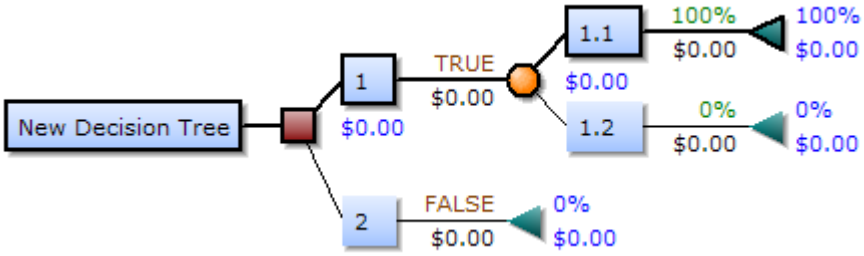
Here coefficient of determination is r^2,

$$r^2 = \frac{\text{Variance explained}}{\text{Total Variance}}$$

A r^2 = high value, it shows that there exists a linear relationship between the two variables. If r^2=1, that indicate that the linear regression is perfect relationship.

3.3.2 Decision tree:

Decision tree is the tool that makes graph which like a tree. That's make a classic of decisions and their possible outcome. It is a way to display an algorithm which only contains qualified statements. Decision trees are usually used for research, particularly in decision analysis it's help make out a strategy likely to make a goal, and it also a popular tool in data mining.



3.3.2.1: Decision tree

3.3 Data Collection Procedure

Here we use the paper which is contain of the steps which are depends on data mining techniques the steps are followings all the steps. [9, 10]

3.3.1 Data Collection:

First, we can't find data from anywhere and the data was not fulfilling, there was gap in daily data set. Missing data can't give good prediction. That's why we take data from another site. And there has daily complete data.

We use the data from the website that obtained from the website wunderground.com [11]

The format of data was in CSV format and included parameters like humidity, temperature, cloud cover, wind speed, etc.

3.3.2 Data Transformation:

The CSV document was first changed over to. arff organization to nourish it into the Data Mining apparatus - WEKA. The transformation to. arff format was implemented through code written in java in weka we make divide our weather dataset in four new datasets, through the season.

3.3.3 Data Pre-processing:

The information document was utilized as a source record and after that procedure was connected to the dataset. Here the method contains choosing illustration from dataset on a random basis with or without substitute the data.

3.3.4 Feature Extraction:

From all the limitation we considered which make of max temperature, min temperature, mean temperature, max humidity, min humidity, mean humidity, wind speed. Rest of the constraint were used for further giving out in application. They were mutually exclusive and no redundancy was present and between them.

3.3.4.1 Data mining:

We here analyzing the given data- set with the algorithms Linear Regression and C4.5 algorithm. Then finding the better prediction for further. Then the dataset was opening in training set. And testing the data into machine learn and dataset along with cross justification. From the patterns that were verification for make further predictions. besides we take different attribute for the weather which are predict by the algorithm. The attribute which was shown in the table 3.3.1.

Table 3.3.1: Attributes of prediction system

Date	Numeric
TemAvg	Numeric
TemMin	Numeric
HuMax	Numeric
HuMin	Numeric
WindMax	Numeric
winMin	Numeric
Outlook	Boolean

3.4 Statistical Analysis

3.4.1 Linear regression:

The Linear regression use for the prediction of the temperature, humidity, wind speed, from the given dataset. Linear regression gives the output based on previous value. Previous value will be taken from our dataset. Here the dependent variable is the all value we want to predict like temperature, humidity, wind speed and the independent variable is the time. The dependent variable are change through the time. Here time is predictor and all other variable will predict depend on time. For calculate the predicted value first we need the intercept and slope. In Linear regression the intercept is only the expected mean value of dependent variable. The slope of a regression line symbolizes the rate of change in dependent variable as independent variable changes. If the value of dependent variable change it will change the value of independent variable. By regression analysis we get the graph of temperature.

Here like temperature, humidity, wind speed is numeric value but the outlook non-numeric value so linear regression not work on the non-numeric value. So, we use decision tree for calculate the outlook.

3.4.2 Decision Tree

Here the decision tree algorithm uses to find the tree for given data set which is develop on the information entropy system. The decision tree to make the better prediction attribute that make the simple split set. The characteristic having the most astounding entropy contrast or standardized data gain is chosen as the part criteria for that specific hub. The characteristic having the most noteworthy entropy distinction or standardized data gain is chosen as the part criteria for that specific hub. Comparative form is pursued and hubs are added to the choice tree. Every penultimate hub conveys the last characteristic or various qualities for settling on an official choice of the issue in figure 3.4.2.1

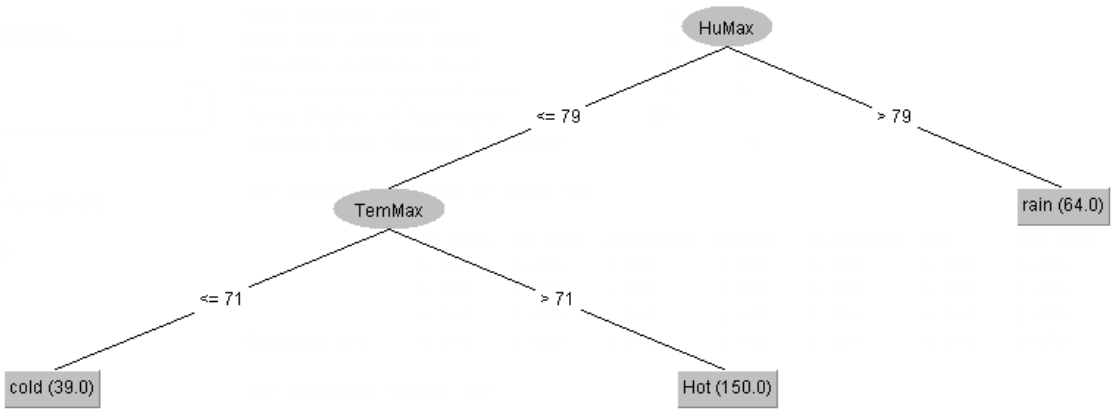


Figure -3.4.2.1 Decision tree

After the execution correlation, the J48 calculation was decided for further usage which included investigation of the heritage information about the climate. The resample channel was utilized for information pre-handling, besides for the information choice advance, from a sum of 8 parameters the most extreme mugginess was dismissed because of commotion present in it which was affecting the accuracy of the system. This brought about thought of straight parameters like max temperature, min temperature, avg temperature, min mugginess, max humidity, min wind speed, max wind speed.

After the performance comparison, the decision tree algorithm was chosen for further achievement which involved study to the legacy data about the weather. The resample filter was used for data pre-processing, furthermore for the data selection step, from a total of 9 parameters. This resulted in consideration of linear parameters like max temperature, min temperature, avg temperature, min humidity, max humidity, min wind speed, max wind speed, outlook.

3.5 Implementation Requirements

Nowadays, the huge need of power for the prediction application used of many supercomputers for control the data. For every case, the parallel development technique aces different problem for compute data for the need of power of hardware of data.

For the high amount of character of the data caught up in the problems higher than explain. It is essential to analyze and make simpler the data before scheduled with other analysis. There for some data mining system are proper in this context. The unsupervised self-organizing maps (SOM) groups of techniques allow to partition the global recheck databases and make producing practical model for weather prediction applications. That data archive are used to develop the models able to pick up direct impressive model output. We are planned to use smooth of data with standard error bands. Them use bootstrap technique to complete achievement for show the nonparametric or parametric maximum model inference. We use the php language to develop the application.

CHAPTER 4

Experimental Results and Discussion

4.1 Introduction

Weather is the state of the environment which describing for which it is hot or cold, wet or dry, calm or stormy and clear or cloudy. In this chapter the results of the study are presented and discussed with reference to the aim of the study, which was to determine the weather forecasting. In this project we work with Weka data mining tools and used various algorithms to show the differences. At the end we got our expected outcome what we want to find out in our project. Here we will show the results we have got and will discuss briefly with our findings. To make better predictions and forecasts about what the weather we will do some update in future.

4.2 Experimental Results

In this study, Linear regression and Decision tree models are designed to predict the weather attribute. Weather attributes are predicted maximum temperature, minimum temperature, maximum humidity, minimum humidity, minimum wind speed, maximum wind speed. For both model's yearly data is divided into four seasons as shown in Table:4.2.1. and models are trained using this seasonal data as input.

Table- 4.2.1: Seasonal data as input

S NO	Season	Duration
1	Winter	1 December – 28 February
2	Spring	1 march – 31 May
3	Summer	1 June – 31 August
4	Autumn	1 September – 30 November

4.3 Descriptive Analysis

Here we make the dataset four-part as winter, spring, summer and autumn. We take two-year data January 2017 to December 2018. In winter we take from 1 December to 28 February, In spring from 1 March to 31 May, In Summer 1 June to 31 August and in autumn from 1 September to 30 November. In each Season we take two-year data in CSV file. First, we train the file in weka under forecast with linear regression analysis. From this we get the three-day weather of maximum temperature, minimum

temperature, average temperature, maximum humidity, minimum humidity, maximum wind speed, minimum wind speed. Then we use the decision tree on weka for each dataset and we get the tree for each season. From the all output we make a model for weather prediction. By the model we make a web application which can predict the weather for three day based on the historical dataset depends on season.

Seasonal sample data are given bellow:

Database: final_project, Table: autumn, Purpose: Dumping data

day	date	TemMax	TemAvg	TemMin	HuMax	HuMin	WindMax	WindMin	outlook
1	2017-09-01	102	90	78	66	14	14	0	Hot
2	2017-09-02	101	92	82	70	33	14	0	Hot
3	2017-09-03	102	92	82	74	33	16	0	rain
4	2017-09-04	100	93	86	75	36	14	0	rain
5	2017-09-05	100	94	87	70	43	16	4	Hot
6	2017-09-06	98	90	82	79	41	14	0	rain
7	2017-09-07	100	92	84	70	35	16	0	Hot
8	2017-09-08	100	92	84	74	35	16	0	rain
9	2017-09-09	100	92	85	70	43	16	0	Hot
10	2017-09-10	102	94	86	66	29	14	0	Hot
11	2017-09-11	100	92	84	74	44	16	0	rain
12	2017-09-12	100	93	86	75	37	18	0	rain
13	2017-09-13	99	92	84	84	47	16	0	rain
14	2017-09-14	100	94	87	70	42	20	2	Hot
15	2017-09-15	99	93	87	75	36	14	0	rain
16	2017-09-16	97	91	85	70	46	12	0	Hot
17	2017-09-17	98	90	82	79	39	12	0	rain
18	2017-09-18	100	94	87	70	29	14	0	Hot
19	2017-09-19	99	90	81	62	23	14	0	Hot
20	2017-09-20	100	92	85	70	36	16	0	Hot
21	2017-09-21	100	89	78	65	29	14	0	Hot
22	2017-09-22	98	88	78	66	30	14	0	Hot
23	2017-09-23	99	88	78	78	39	12	0	rain
24	2017-09-24	100	92	83	79	22	14	0	rain
25	2017-09-25	98	89	80	66	14	12	0	Hot
26	2017-09-26	97	88	78	69	34	12	0	Hot
27	2017-09-27	98	90	82	70	39	14	0	Hot
28	2017-09-28	97	90	83	84	42	9	0	rain
29	2017-09-29	99	90	80	84	27	18	0	rain
30	2017-09-30	100	89	78	66	19	12	0	Hot
31	2017-10-01	96	89	82	79	40	12	0	rain
32	2017-10-02	95	86	78	78	34	16	0	rain
33	2017-10-03	95	90	85	43	21	20	5	Hot
34	2017-10-04	94	88	81	49	19	22	0	Hot
35	2017-10-05	96	86	77	61	21	22	0	Hot
36	2017-10-06	91	84	76	48	8	18	0	Hot
37	2017-10-07	96	82	69	62	15	7	0	Hot
38	2017-10-08	95	86	78	74	37	12	0	rain
39	2017-10-09	92	87	82	79	49	14	0	rain
40	2017-10-10	93	84	75	74	25	12	0	rain
41	2017-10-11	96	84	73	66	16	16	0	Hot

Figure 4.3.1: Autumn season dataset

day	date	TemMax	TemAvg	TemMin	HuMax	HuMin	WindMax	WindMin	outlook
1	2017-03-01	71	61	51	88	34	12	0	cold
2	2017-03-02	81	71	61	72	16	22	0	rain
3	2017-03-03	79	70	62	88	25	12	0	rain
4	2017-03-04	75	68	62	100	35	12	4	rain
5	2017-03-05	77	67	57	94	36	16	0	rain
6	2017-03-06	69	66	62	64	37	24	7	cold
7	2017-03-07	72	62	51	76	36	16	0	rain
8	2017-03-08	74	66	57	77	29	14	0	rain
9	2017-03-09	73	68	62	73	26	18	7	rain
10	2017-03-10	76	64	51	82	21	12	0	rain
11	2017-03-11	79	72	65	56	16	24	4	Hot
12	2017-03-12	82	74	66	50	18	22	0	Hot
13	2017-03-13	80	72	63	64	13	22	0	Hot
14	2017-03-14	72	68	64	88	41	20	0	rain
15	2017-03-15	74	68	61	64	35	16	0	Hot
16	2017-03-16	84	72	61	62	14	29	0	Hot
17	2017-03-17	74	70	66	78	40	20	0	rain
18	2017-03-18	79	72	66	73	36	12	0	rain
19	2017-03-19	69	66	64	83	40	38	5	cold
20	2017-03-20	72	68	64	83	58	20	7	rain
21	2017-03-21	76	66	57	94	19	12	0	rain
22	2017-03-22	75	66	58	72	34	18	0	rain
23	2017-03-23	73	69	65	73	35	20	0	rain
24	2017-03-24	71	68	64	88	47	14	0	cold
25	2017-03-25	71	66	60	94	56	18	0	cold
26	2017-03-26	73	68	62	94	57	14	0	rain
27	2017-03-27	74	69	64	83	47	14	4	rain
28	2017-03-28	75	67	59	94	43	16	0	rain
29	2017-03-29	78	73	68	73	32	14	0	rain
30	2017-03-30	81	72	64	88	45	14	0	rain
31	2017-03-31	81	73	65	83	30	18	0	rain
32	2017-04-01	88	78	69	69	28	9	0	Hot
33	2017-04-02	86	80	74	74	40	12	0	rain
34	2017-04-03	78	75	72	69	35	25	0	Hot

Figure 4.3.2: Spring season dataset

Database: final_project, Table: summer, Purpose: Dumping data

day	date	TemMax	TemAvg	TemMin	HuMax	HuMin	WindMax	WindMin	outlook
1	2017-06-01	88	86	83	48	28	25	13	Hot
2	2017-06-02	93	88	82	51	23	23	13	Hot
3	2017-06-03	91	86	80	61	14	22	7	Hot
4	2017-06-04	89	82	75	73	26	16	0	rain
5	2017-06-05	92	82	72	70	29	12	0	Hot
6	2017-06-06	93	83	73	65	30	14	0	Hot
7	2017-06-07	96	86	75	66	22	12	0	Hot
8	2017-06-08	96	86	75	61	18	12	0	Hot
9	2017-06-09	95	88	82	58	24	16	0	Hot
10	2017-06-10	101	88	74	61	10	16	0	Hot
11	2017-06-11	100	88	75	70	19	16	0	Hot
12	2017-06-12	96	88	79	74	32	18	0	rain
13	2017-06-13	96	90	85	55	26	22	0	Hot
14	2017-06-14	102	92	83	70	25	14	0	Hot
15	2017-06-15	104	92	80	70	14	12	0	Hot
16	2017-06-16	100	92	83	79	25	116	0	rain
17	2017-06-17	104	97	90	49	10	21	7	Hot
18	2017-06-18	100	95	90	52	21	20	7	Hot
19	2017-06-19	98	92	85	59	20	20	0	Hot
20	2017-06-20	96	92	88	52	34	22	11	Hot
21	2017-06-21	96	92	88	55	32	27	11	Hot
22	2017-06-22	96	92	87	55	25	25	11	Hot
23	2017-06-23	92	89	86	52	31	22	11	Hot
24	2017-06-24	92	87	82	66	37	21	0	Hot
25	2017-06-25	92	84	77	69	39	16	0	Hot
26	2017-06-26	92	85	78	74	31	14	0	rain
27	2017-06-27	97	86	75	65	17	16	0	Hot
28	2017-06-28	104	90	76	59	12	14	0	Hot
29	2017-06-29	106	94	81	89	7	12	0	rain
30	2017-06-30	98	91	84	79	31	14	0	rain
31	2017-07-01	106	95	84	70	16	18	0	Hot
32	2017-07-02	101	92	82	58	14	18	0	Hot
33	2017-07-03	101	92	82	52	12	14	0	Hot
34	2017-07-04	105	96	86	55	11	9	0	Hot
35	2017-07-05	102	93	84	62	14	12	0	Hot
36	2017-07-06	105	96	86	75	19	12	0	rain

Figure 4.3.3: Summer season dataset

Database: final_project, Table: winter, Purpose: Dumping data

day	date	TemMax	TemAvg	TemMin	HuMax	HuMin	WindMax	WindMin	outlook
1	2017-12-01	73	68	64	77	43	16	2	rain
2	2017-12-02	70	66	61	68	42	16	5	cold
3	2017-12-03	72	66	60	73	41	16	0	rain
4	2017-12-04	74	66	59	82	43	12	0	rain
5	2017-12-05	72	64	55	94	36	7	0	rain
6	2017-12-06	75	64	54	88	17	9	0	rain
7	2017-12-07	78	71	64	60	28	20	7	Hot
8	2017-12-08	69	64	59	59	33	26	7	cold
9	2017-12-09	65	59	53	63	38	20	9	cold
10	2017-12-10	65	60	55	59	39	16	7	cold
11	2017-12-11	66	60	55	67	34	18	7	cold
12	2017-12-12	65	59	53	63	36	16	2	cold
13	2017-12-13	68	57	46	72	21	14	0	cold
14	2017-12-14	74	63	52	82	31	16	0	rain
15	2017-12-15	75	68	62	64	31	24	7	Hot
16	2017-12-16	73	68	62	72	24	34	7	rain
17	2017-12-17	71	64	57	73	34	12	0	cold
18	2017-12-18	70	65	60	77	34	18	0	cold
19	2017-12-19	69	63	57	64	32	14	4	cold
20	2017-12-20	72	62	53	77	31	9	0	rain
21	2017-12-21	74	64	53	94	39	7	0	rain
22	2017-12-22	74	65	56	88	45	12	0	rain
23	2017-12-23	73	63	53	94	27	9	0	rain
24	2017-12-24	78	64	51	94	25	7	0	rain
25	2017-12-25	85	70	56	72	8	22	0	rain
26	2017-12-26	69	66	64	73	31	20	0	cold
27	2017-12-27	68	62	55	82	45	12	0	cold
28	2017-12-28	72	64	57	88	56	9	0	rain
29	2017-12-29	73	63	53	100	31	7	0	rain
30	2017-12-30	73	62	52	88	37	7	0	rain
31	2017-12-31	73	62	51	88	33	9	0	rain
32	2018-01-01	75	62	49	82	33	55	0	rain
33	2018-01-02	69	63	57	73	33	16	0	cold
34	2018-01-03	68	64	60	59	33	22	9	cold
35	2018-01-04	67	62	57	59	24	16	0	cold
36	2018-01-05	72	60	48	66	18	9	0	Hot

Figure 4.3.4: Winter season dataset

The testing of the dataset:

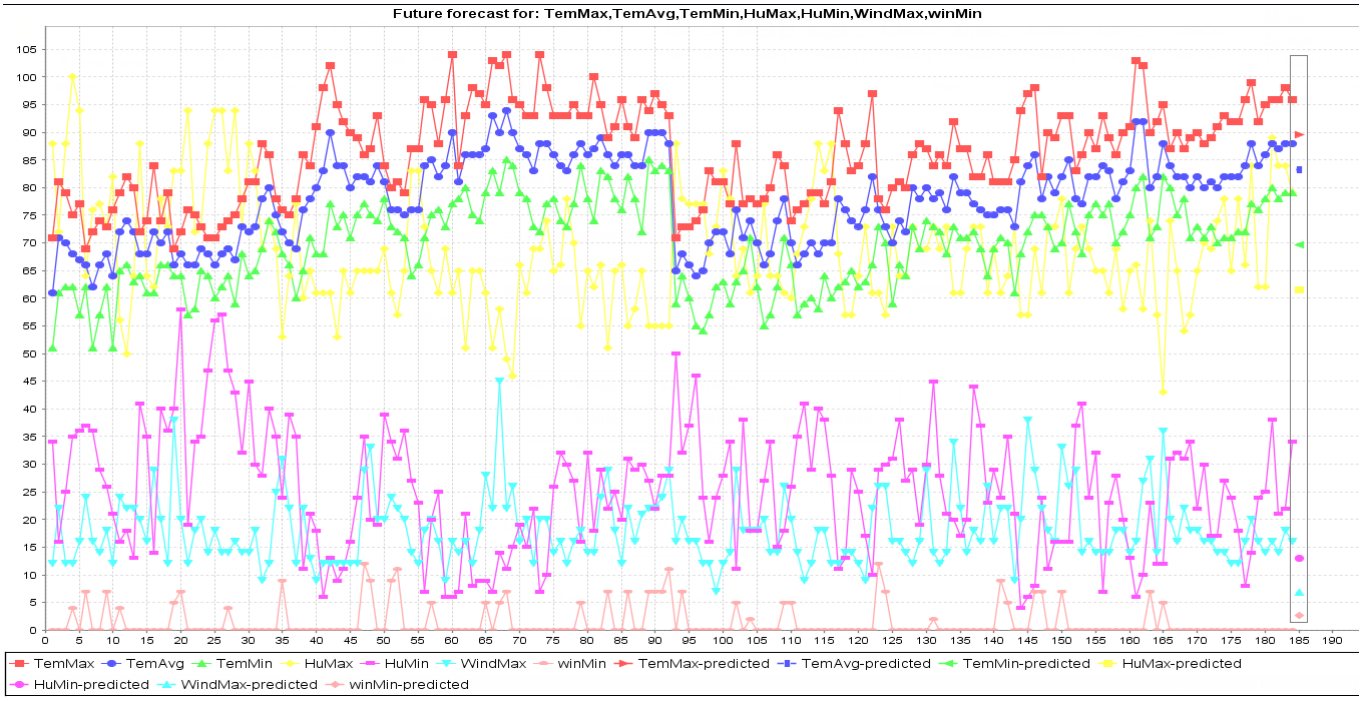


Figure 4.3.5: winter season using linear regression

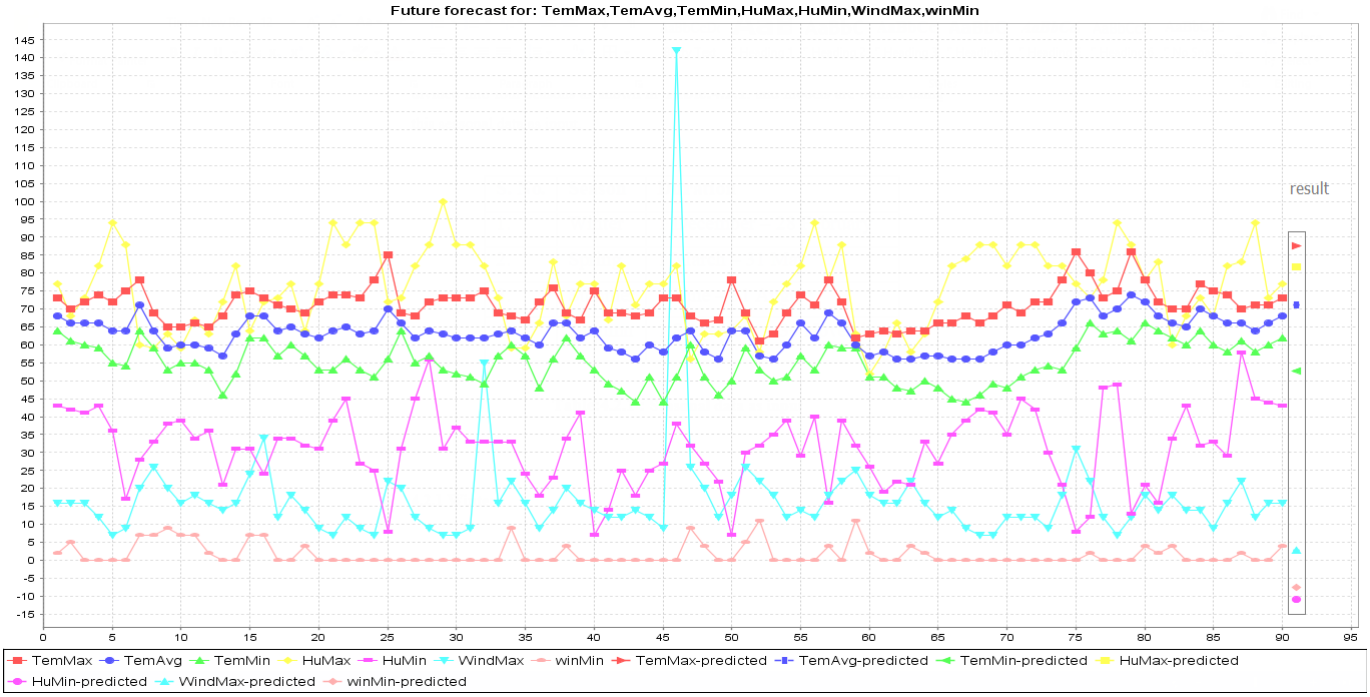


Figure 4.3.6: Spring season using linear regression

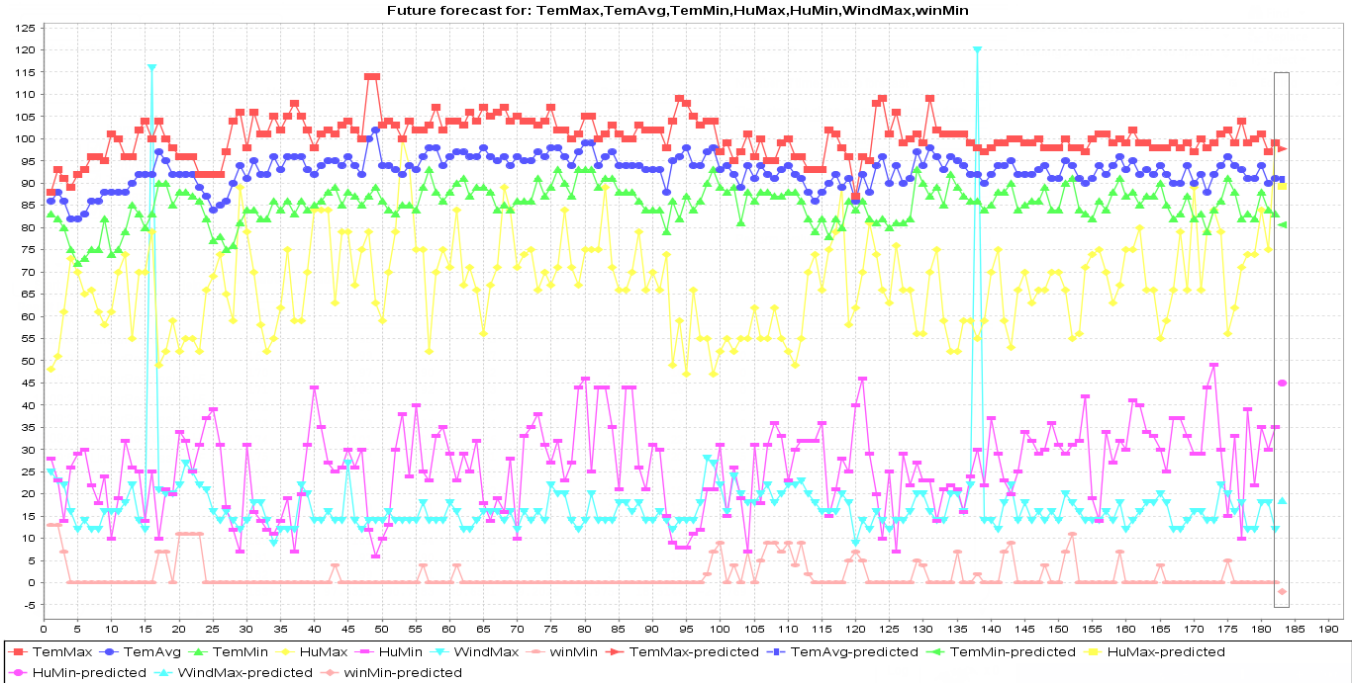


Figure 4.3.7: Summer season using linear regression

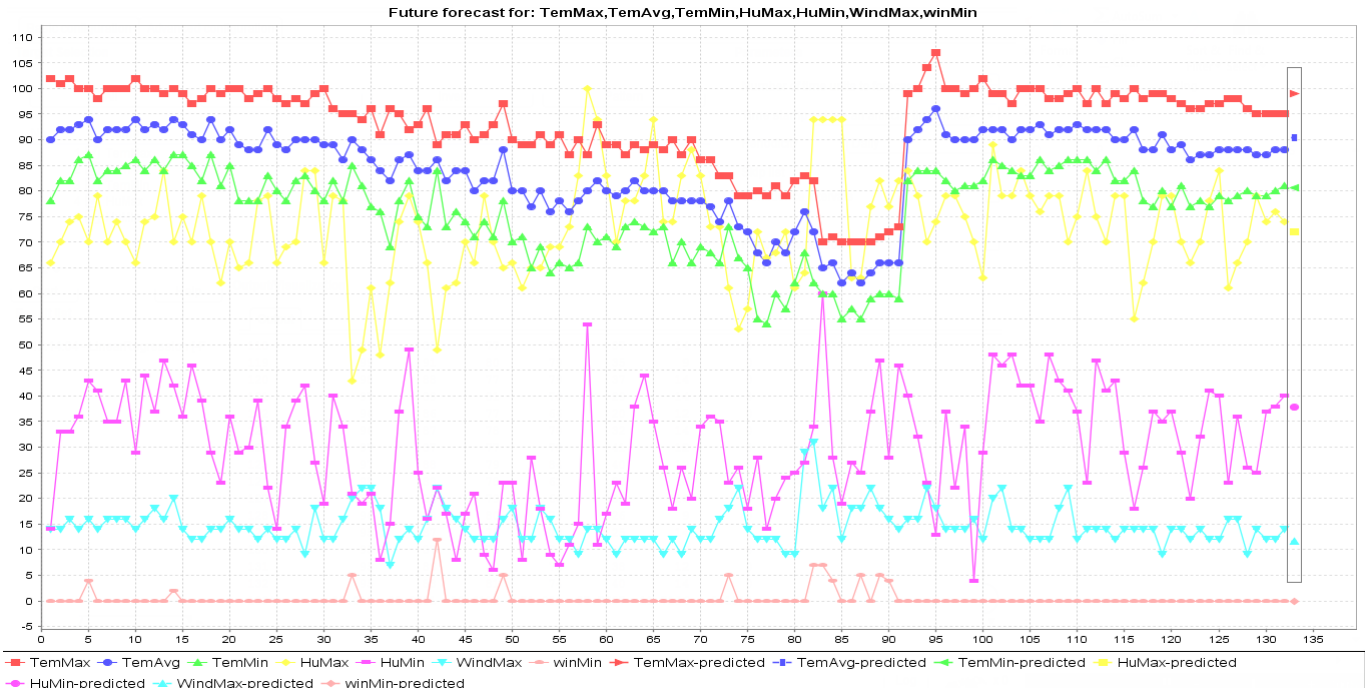


Figure 4.3.8: Autumn season using linear regression

Through divided the dataset we get the better result for each season of a year.

4.4 Summary

From 19th century, scientist is trying for predicting the weather. But they face plenty of problems. Weather prediction calculate through the problems. Moreover, all the parameters we work upon like temperature, humidity and wind speed can change at any time. So, all the time our prediction may not work as the weather act on. But yet we will develop it gradually. In the consequence of Bangladesh this type of research is not fully developed. We are trying our level best to build up a fruitful terminology for the country so that we can fulfill our necessity. And here is some limitation such as the weather data is not available for all the location. And we all know machine always work on algorithms and predefined instructions. Matching doesn't know is it wrong or right, it just takes a decision from the previous trained parameter. Here the prediction we have done will be best if the weather remains in general flow. It's not our duty to control the weather, but we have tried our level best to make a similarity by trying again and again. Further if we notify about any problem in it, we will check that and update that from our own passion.

CHAPTER 5
Summary, Conclusion, Recommendation and Implication for Future Research

5.1 Summary of the Study

Any prediction can't give the absolute result. Every prediction some limitation. But we have tried our level best to improve and get the better result from the research.
In this Study Forecasting Models algorithm are used for classifying weather parameters such as maximum temperature, minimum temperature, and wind speed in terms of the month and years. The implementation of data mining approach to solve the wind forecasting problems for wind farm production, in particular, for predicting wind speed. the data mining prediction algorithm are used for prediction here. We study such as decision tree, linear regression, naive Bayes, Artificial neural network, from all of them we use the decision tree and linear regression algorithm.

5.2 Conclusions

This research suggests and offers a well-organized and correct weather prediction and forecasting model which is using linear regression ideas. We also use normal equation model. All these concepts are a part of machine learning. The linear regression equation is a very capable weather prediction model. We are using the parameters such as temperature, humidity and wind speed. This can be used to make trustworthy weather predictions. This model also helps decision making in everyday life. It can give better results when be valid to cleaner and larger datasets. Before processing of the datasets can valuable in the prediction and the unprocessed data can also affect the efficiency of the model.

5.3 Recommendations

Here, we done the prediction by use of historical data, but the weather of the earth changing day by day simultaneously. So, by predict using historical data not so much accurate. That's for better and dependable prediction we need to add some systemic component. Here we use historical data an algorithm in future we will work with the system and develop it as possible we can.

5.4 Implication for Further Study

Prediction of weather is not a easy task. From the output of the research we will make a system for predict weather such temperature, humidity, wind speed, outlook. It is not only software base thing. It depends on hardware. So only by data mining tools is not possible to give accurate result. In future we will work with software and hardware both to get the better and real time result.

APPENDIX

Appendix A: Project Reflection

We started our project from Fall 2017 to implement this weather prediction platform. We have tried to make a user-friendly interface of our project. The main feature of our project is to predict the weather efficiently. It's very essential to keep the information of the previous weather data set in our database to predict efficiently. This web app will help to view our prediction system more successfully. We used weka as our data mining tool. To implement the web application first we build a model of our application then we implement our web interface step by step. After a long journey and hard work finally, we able to reach our goal.

Appendix B: Related Issues

To implement our proposed application at first, we read a number of paper and journal known as literature review. Weather forecasting is a complex and often challenging skill that involves observing and processing vast amounts of data. We use decision tree algorithm in weka to analyze how the outlook change. And linear regression uses to find pattern in the attributes. We implement our application using web platform and MySQL.

References:

[1] Learn about Weather forecasting, available at <<https://en.wikipedia.org/wiki/Weather_forecasting>>, last accessed on 01-06-2018 at 04:53 pm.

[2] C. Zhang, W.-B. Chen, X. Chen, R. Tiwari, L. Yang and G. Warner, "A Multimodal Data Mining Framework for Revealing Common Sources of Spam Images", Journal of multimedia, vol. 4, no. 5, pp. 313-320, October (2009).

[3] J. Jiang, B. Guo, W. Mo and K. Fan, "Block-Based Parallel Intra Prediction Scheme for HEVC", Journal of Multimedia, vol. 7, no. 4, pp. 289-294, August 2012

[4] Z. Fu, J. Bai and Q. Wang, "A Novel Dynamic Bandwidth Allocation Algorithm with Correction-based the Multiple Traffic Prediction in EPON", Journal of Networks, vol. 7, no. 10, pp. 1554-1560, October (2012).

[5] S.-Y. Yang, C.-M. Chao, P.-Z. Chen and C.-Hao, "SunIncremental Mining of Closed Sequential Patterns in Multiple Data Streams", Journal of Networks, vol. 6, no. 5, pp. 728-735, May (2011).

[6] Takemasa Miyoshi, Keiichi Kondo, and Koji Terasaki. Big Ensemble Data Assimilation in Numerical Weather Prediction. Computer (Volume: 48 , Issue: 11 ,pp 15-21, Nov.(2015).

[7] Takemasa Miyoshi, Masaru Kunii, Juan Ruiz, Guo-Yuan Lien, Shinsuke Satoh, Tomoo Ushio, Kotaro Bessho, Hiromu Seko, Hirofumi Tomita, Yutaka Ishikawa. "Big Data Assimilation" Revolutionizing Severe Weather Prediction. Bulletin of the American Meteorological Society, 2016.

[8] Learn about Linear_regression, available at<<https://en.wikipedia.org/wiki/Linear_regression>>, last accessed on 06-06-2018 at 05:56 pm.

[9] Taksande AA, Mohod PS. Applications of data mining in weather forecasting using frequent pattern growth algo- rithm. IJSR.; 4(6):3048–51, Jun 2015.

[10] Olaiya F. Application of data mining techniques in weather prediction and climate change studies. I. J. Information Engineering and Electronic Business; 1:51–9, Jul 2012.

[11] Learn about geographic, available at https://www.wunderground.com/weather/bd/dhaka >>, last accessed on 20-11- 2018 at 9:32pm.

Document Viewer

Turnitin Originality Report

Processed on: 04-Nov-2018 12:03 +06
ID: 1032390810
Word Count: 4457
Submitted: 1

151-15-4933 By Mehedi Hasan Sagor

Similarity Index	Similarity by Source	
20%	Internet Sources:	15%
	Publications:	4%
	Student Papers:	2%

include quoted include bibliography excluding matches < 1% download
refresh print mode: quickview (classic) report

11% match (Internet from 18-Jun-2017)
http://www.i-scholar.in

2% match (Internet from 06-May-2016)
http://research.ijcaonline.org

2% match (publications)
Zwieflhofer, Walter, and Norbert Kreitz. "IMPLEMENTATION OF DATA MINING TECHNIQUES FOR METEOROLOGICAL APPLICATIONS", Realizing Teracomputing, 2003.

1% match (student papers from 16-Oct-2018)
Submitted to Amity University on 2018-10-16

1% match (publications)
Byers, H. R., H. E. Landsberg, H. Wexler, B. Haurwitz, A. F. Spilhaus, H. C. Willett, and H. G. Houghton. "Objective Weather Forecasting", Compendium of Meteorology, 1951.

1% match (Internet from 24-Jan-2017)
http://www.ijarcce.com

1% match (Internet from 01-Mar-2014)
http://en.wikipedia.org

1% match (publications)
de Rooy, D. P. C., M. L. E. Andersson, R. Knevel, T. W. J. Huizinga, B. Svensson, and A. H. M. van der Helm-van Mil. "Does the season at symptom onset influence the long-term severity of radiographic joint destruction in rheumatoid arthritis?", Annals of the Rheumatic Diseases, 2012.

1% match (Internet from 25-Oct-2018)