WEATHER TEMPERATURE PREDICTION BASED ON DATA MINING TECHNIQUES

BY

MD. FARHAD RAIHAN
ID: 161-15-7080

MD. MEHEDI HASAN SHEJAN
ID: 161-15-6928
AND

MD. SHAHRIAR HOSSAIN
ID: 161-15-7523

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

Supervised By

MD. FIROZ HASAN
Lecturer
Department of CSE
Daffodil International University

Co-Supervised By

DEWAN MAMUN RAZA
Lecturer
Department of CSE
Daffodil International University

DAFFODIL INTERNATIONAL UNIVERSITY

DHAKA, BANGLADESH

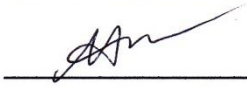
DECEMBER 2019

APPROVAL

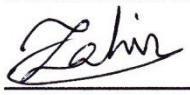
This Research titled "**Weather Temperature Prediction Based on Data Mining Techniques**", submitted by Md. Farhad Raihan and Md. Shahriar Hossain and Md. Mehedi Hasan Shejan, ID No: 161-15-7080, 161-15-7523, 161-15-6928 to the Department of Computer Science and Engineering, Daffodil International University, 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 had been held on December 07, 2019.

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



Nazmun Nessa Moon — **Internal Examiner**
Assistant Professor
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

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

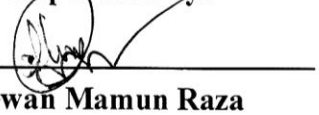
DECLARATION

We hereby declare that, this research has been done by us under the supervision of **Md. Firoz Hasan, Lecturer, Department of CSE** and co-supervision of **Dewan Mamun Raza, 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:

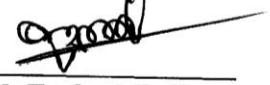
Md. Firoz Hasan
Lecturer
Department of CSE
Daffodil International University

Co-Supervised by:



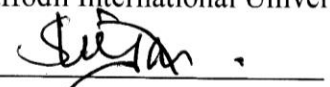
Dewan Mamun Raza
Lecturer
Department of CSE
Daffodil International University

Submitted by:



Md. Farhad Raihan
ID: 161-15-7080
Department of CSE
Daffodil International University

Md. Shahriar Hossain
ID: 161-15-7523
Department of CSE
Daffodil International University



Md. Mehedi Hasan Shejan
ID: 161-15-6928
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 research successfully.

We really grateful and wish our profound our indebtedness to **Md. Firoz Hasan**, **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 **Prof. Dr. Syed Akhter Hossain, 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 prediction is an anxious work for researchers and has completed a lot of research interest in the last few years. It has been a demand application of climatology and one of the specialized challenging problems in the world. Meteorological Department are working very hard day by day to improve the accuracy. But none of them become hundred percent successful. So, it is a continuous process and we decided to involve ourselves in this challenge. In our case, we have investigated about different techniques in weather prediction. Finally we intend to use data mining technique in our research. This paper comes up with a current method to increase a fitness-oriented architecture for the weather notification systems which predict whether using these data mining techniques. Our main approach to apply different classification algorithm on a bulk of realistic data and watch any major difference among their output of prediction. More precisely we always tried to choose the best algorithm. We observe their behavior using graphical interface. Most of the case we got Support Vector Machine to fulfil our satisfied output. Then we have studied how the Support Vector Machine works in classification and what is the internal infrastructure of it. We were trying to improve this algorithm and proposed a better version.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1
Introduction

1.1 Introduction

As the application of science and technology, weather prediction is an art to predict climatic changes and behavior in the subject of time for an appropriate zone. Weather prediction was one of the most investigated problems all over the world in the previous several decades scientifically and technologically. Prediction is one of the basic application of data mining. Data Mining is used to explore and analyze large amount of data, generate patterns, gather meaningful information.

Weather prediction brings about computing how the expressed situation will alternate, change or transform. In this era of electronics weather circumstances are observed and recorded by sea measurement, aircraft measurement, satellite, sensor, radio station and radar. Rendered information are sent to weather department where the data are possessed by itself, figured out, and made graphical representation using graphs, charts, and maps. Powerful and modern super computers are used to calculate thousands of measurements and obvers. Predicting weather using a computer is known as algorithmic weather prediction.

Weather prediction is a complex and lengthy process not only for general people but also for scientist and researchers. It demands skillfulness in versatile regulation. Atmospheric parameters is necessary for this kind of applications. Accurate scientific prediction is terrible job because of the anomalous habit of the atmosphere. Scientists use various numerical method like neural network, regression, radial basis function network to forecast atmospheric parameter like temperature, humidity, rains, sunrise, sunset etc.

The task of certain the best basic surrounding for numerical weather prediction (NWP) is of the best constructive necessity and has been the subject of various studies by people for kind of qualification. Not especially these studies result in at first sight rather different programs. However, on clause, it reviews they can normally be shown to be closely

associated. In this paper, we expedition an ordinary form for the great analysis, articulate the individual component of the practical NWP analysis problem, and we describe this to other released methods.

1.2 Motivation

First of all, when I see my mobile weather application then I see that a list of one weak weather Prediction is there. I was so questioning about it. And we shortage to learn how it works, how we can make even better arrangements of the application. Driven into the data mining technique using Weka.

In the future, we want to appoint ourselves with big data analysis. So, we want to make something that helps us to succeed in life.

Data mining is the field where a vast amount of data is collected and being processed to compost some useful data i.e. information. Everyone wants the summary and appropriate information which is possible by it, as it is not an easy ought but it turns out possible by a series of methods and science.

1.3 Rational of the study

The point of information digging is for discovering structure through unstructured information. Information mining hauls out the importance of boisterous information and its discovered examples are obviously with arbitrary information. It likewise utilizes from all to all the more likely get patterns, examples, connections, and finally foresee the client conduct and market and rivalry patterns, in this way, along these lines, we improve able calculation to improve a forecast of climate.

Throughout the entire existence of climate expectations, endeavors have frequently been made to devise numerical and target strategies for creating the conjecture. The reasonable extreme objective of the gauging climate is to empower the forecaster to build the precision of figures made routinely. We investigation information utilizing mining apparatuses and calculations to locate a superior result. The consequences of the mining are valuable and exact which gives the route to an apparatus that data in numerous regions and add to advancement.

1.4 Research Questions

We study with climate expectations that are the reason we study a ton of forecast calculations. For a hitter expectation result, we have to discover gather a gigantic measure of information. Information accumulation is a test yet we gather it appropriately. The more we gather information the more we show signs of improvement forecast. We are needing to foresee the climate temperature, moistness, wind speed and the viewpoint of the following day.

The inquiries we have confronted:

1. Accumulation of crude information.
2. Legitimacy of information.
3. Select the calculation for the forecast.
4. Get the normal yield.
5. Mistakes of the calculations.

1.5 Expected Output

1. We will get an approximate idea about how the current situation of the environment will change. Temperature, humidity, wind velocity of the next day will calculate.
2. We can easily find out the Weather situation by using this system.
3. It gives the business with valuable information that in business, that helps to make decisions about the future outcome of the establishment.
4. This process can be used in Air Traffic, Marine, Agriculture, Forestry, Military, and Navy, etc.

1.6 Report Layout

Chapter 1: This is about the Introduction of our exploration.
In this Chapter Introduction, Motivation, Rationale of the Study, Research Questions, Expected Output, Report Layout are depicted.

Chapter 2: Is about the Background of the examination. In this section Related Works, Research Summary, Scope of the Problem and Challenges are portrayed.

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

Chapter 4: Is about Architectural Design of the examination.
In the chapter Planning, Flow of Working Procedure, Database and Implementation are explored

Chapter 5: It is about Experimental Results and Discussion of the examination.
In the chapter Experimental Results, Descriptive Analysis and Summary are depicted.

Chapter 6: In this part Summary, Conclusion, Recommendation, and Implication for Future Study are depicted.

CHAPTER 2
Background

2.1 Introduction

There is no doubt, weather prediction is very important and necessary for lots of sector such as air traffic, agriculture, admiralty, fishermen, awareness about natural disaster, climate change etc. But, causes of chaotic nature of weather, it is a very difficult job.

Thousands of research and investigation had done about weather prediction. By doing lots of study, we have known that many metrological department are trying to improve the accuracy of weather prediction but none of them are hundred percent accurate. So, we have decided ourselves to work in this field.

2.1.1 Ancient methods: For thousands of years people are trying to understand the nature and predict the weather. In 650 BC, people of Babylonia forecast the weather from observation of cloud patterns as well as astrology. During 350 BC, Aristotle said about weather patterns in his book *Meteorologica* in details. Theophrastus wrote a book on weather prediction named which is known as *Book of Signs*.

In a nut shell, ancient weather forecasting was totally depend on pattern of events. For example, if the sunset is particularly red then it is said that the next day brings fair weather. Another observation is that, if the amount of rain is much little in one year then there will be much winter in next year. [1]

2.1.2 Modern methods:

After the invention of electric telegraph, the modern chapter of weather prediction was began. The two man named Royal Navy Francis Beaufort his protégé Robert Fitzroy are memorable for treating weather forecasting as a science. Modern forecasting method were started in 1835. [1]

2.1.3 Broadcasts:

The first daily weather forecasts were published in *The Times* on August, 1861. BBC broadcasted the first weather forecast in 1936 in television experimentally which was included weather map, graph. [1]

2.2 Related Works

In diversity of climate expectation framework there are few analyst have attempted by utilizing a few data mining strategies. There are lots of publication which has work with weather prediction using Conventional Neural Network (CNN). These are related to predict and described the weather situation or status from a network of lots of images about cloud, thunderstorm, tornadoes, earthquake, flood, snowfall etc. These are fail to find out irregular patterns of weather. These model are based on Support Vector Machine regression and stochastic method. There are many types of algorithm to predict something such as Gaussian Naïve Bayes (GNB), Decision Tree, Support Vector Machine (SVM). But most of the cases SVM perform better results. The successful prediction could not gain for the difficult data system of weather unqualified and constant model of weather. It depends on vast and well organized weather data to get better output. A power-full supercomputer is needed to run bulk amount of simulation. For this reason, scientist ran 10,240 simulations over the model and divide global environment with 112 kilometers divisions. [2]

2.3 Research Summary

From research paper, books, different article we got a lots of information about weather prediction related works. There are completed a tiny work about this topic in Bangladesh. In almost research paper, they work with only CNN and predict the behavior of weather from a trained model generated using image. Which is not suitable for large and complex data set and model. We collect a bulk of historical data from different websites and also collect daily from a website using API. Then we create data model for specific day from dataset and apply different algorithm to get desired output.

2.4 Scope of the Problem

In this paper a weather forecasting model is compiled for predicting the weather successfully. Our model of the weather data with consist of time series is very careful and those data are transformed with information where the transformation are done using useful data mining technique. The data mining technique are used to uncover the invisible pattern of data and create a link among the different attributes with the weather condition.

It is true that there are many automated weather prediction system with attractive and smart looking graphical user interface in market. Lots of websites, android apps, IOS apps are available. People are cheated by thinking that all of them are very accurate. But it is not true. We find many of them are fraud and give unrealistic results. In order to reduce this hassle our system will provide appropriate level of prediction. This system will help specially those people who stay irregular hour in their place of employment.
Lastly, bad prediction is very harmful for human daily and make the necessary decision. So, we take this opportunity to work in this platform.

2.5 Challenges

During working on this research we face enormous problem which we had deal very carefully. Some of those major challenges are mentioned bellow.

1. Finding lots of data approximately 50 years historical data.
2. Selecting a good API to get real time data and updating the daily data.
3. Challenge of choosing a good looking graphical interface by matching our system while market place offers a lots.
4. Making the data set suitable for our system and correcting the huge amount of errors.
5. Apply the relative algorithm.
6. Choosing the best algorithm and improve the internal workflow of it.

CHAPTER 3

Research Methodology

3.1 Introduction

The primary work of our research is collecting data. We must know about the history of weather in one area because one of the key value of weather forecasting indicates the future events. Modern method are not nearly be so accurate without numerical prediction, which uses mathematical formula, equations for weather prediction. Such prediction requires most powerful super computer and real time observational data (using temperature sensor, barometer, compass etc.) from land, air, sea and ground. It is not a single person job. So, we need a company which consists of much small station to collect data locally and store all of them in a prime data center.

3.2 Research Subject and Instrumentation

We are working on weather prediction using data mining technique. First of all, we collect the historical data from different meteorological websites. The data was appeared in various form such as csv, excel, JSON. We collected all them together and merge in a structural database management system to work easily. We also done adding or deleting row, editing cell, correcting data types, avoiding unrealistic value or attribute to get a clean data set. After doing this, we choose a portion of dataset according to desired date for making model. After that we split the model into two parts i.e. Train model and
Target model and then applying different classification algorithm such as SVM, GNB, Decision Tree.

After getting output we compere the predicted value with actual situation or data. We calculated the accuracy for choosing the best algorithm suitable for our goal.

3.3 Materials and Methods

3.3.1 Classification:

Classification is related to data mining, the technique of discovering a pattern, model and similar object that details and discriminates data categories, classes, ranges, groups and concepts. Classification is a problem of determining the category of a set of observation.

Example: Before going to play a game in outside, we need to check the weather that we will go to play or not. In this circumstances, a classification is needed to predict the class labels such as play Yes or No. It is done in two steps:

1. Learning Step: Building of classification model. There are many types of algorithm that are used to create a classifier using the model generated by a certain dataset. This model will be trained to perform accurate prediction.
2. Classification Step: Now the trained model is used to classify a test data. We enter a test data of certain category and get a class as our expected output [3]

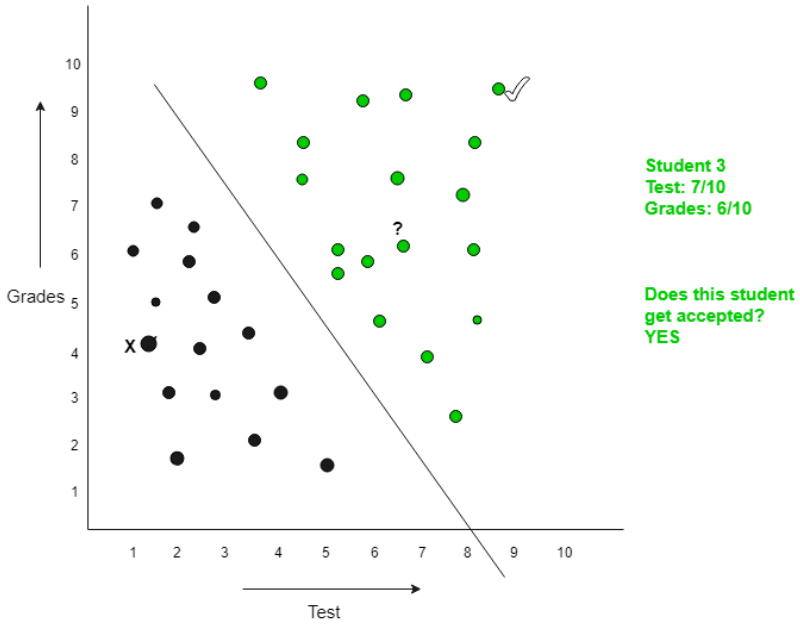


Figure 3.3.1: Classification

3.3.2 Gaussian Naïve Byes:

Bayes' Theorem: Bayes' Theorem is developed by Thomas Bayes describes the probability of an event based on acquired knowledge of conditions.

$$P(A|B) = \frac{P(B|A)(P(A)}{P(B)}$$

Figure 3.3.2: Gaussian Naïve Byes formula

Let's explain the equation.

- A and B are two instance.
- P (A) and P (B) are the possibilities of event which is independent from each other.
- P (B|A) is the possibility of B under the condition of A and vice versa.

Naïve Bayes has two classifier i.e. Bernoulli naïve Bayes and Gaussian naïve Bayes. We uses Gaussian naïve Bayes because it is most popular and simple. [4]

3.3.3 Decision Tree:

Decision Tree, one of the most popular and powerful tool is used for classification and prediction. It is a graph like tree which helps to make decision. In decision tree each internal node indicates a test on attribute, each branch denotes an outcome of test and each leaf node holds a class label. [5]

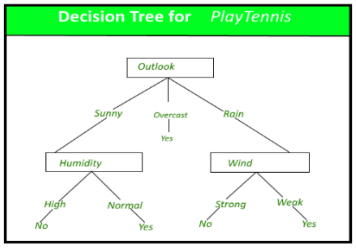


Figure 3.3.3: Decision Tree

3.3.4 Support Vector Machine:

Support Vector Machine are established on the concept of decision pane that define the boundaries. A decision pane or line are used which separates a set of object according to their different class membership. A graphical representation is given below:

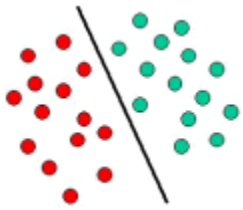


Figure 3.3.4: Support Vector Machine

In the above picture there are a bunch of red and green and balls. We put a pane in the middle. We sort the balls one by one by keeping the red balls in left side of the pane and the green balls in right side of pane. So, we said there are two class Green and Red. If a new ball like blue colored is placed in left side of the pane then it is classified as Red or if it is placed in right side of the pane then it is classified as Green. [6]

3.4 Data Collection Procedure

We had followed some steps to collect data which is used in data mining techniques.

3.4.1 Data Collection:

First of all it is very difficult to find a suitable dataset that satisfy us to reach our goal. We found a small resource of historical data. Fortunately, we got a website named *OpenWeatherMap* which provides daily weather data and five day forecast data freely in JSON format. We can access these data by calling a simple API. [7]

We also collect historical data from another website named *Meteostat*. It provides us a huge amount of previous data such as daily forecast, hourly forecast, geography map etc.

3.4.2 Data Transformation:

The data was in CSV and JSON format. We read the every cell of data row wise programmatically and then converted it to Sqlite3 format. We define a function by ourselves in python for doing this job.

3.4.3 Data Pre-processing:

Data pre-processing is involved with different task like removing invalid cell, adding new cell, changing data type, correcting null or unrealistic value.

3.4.4 Feature Extraction:

Considering all the limitations, we make of average temperature, minimum temperature, maximum temperature, humidity, air pressure, wind flow, wind direction, clouds. Rest of the constraint were used for additional outcome in our application. They were mutually exclusive and no redundancy were found between them.

3.4.4.1 Data Mining:

Here we analyzing the proceeds dataset by using GNB, SVM, Decision Tree, KNN algorithm. Then we finding the better predicted result comparing with the real data. We calculate the accuracy for knowing which algorithm works better over the time. We define some attributes from the dataset that are predicted by algorithm. This attribute are shown in table next page:

Table 3.4.4.1: Name of the attributes and their data types

Attribute Name	Data Type
Temperature	Float
Minimum Temperature	Float
Maximum Temperature	Float
Humidity	Integer
Cloud	Float
Wind Speed	Float
Wind Direction	Integer

3.5 Statistical Analysis

We used classification technique which is very classic method on data mining and it is based on machine learning. Details were said in lesson 3.4. Our dataset has many attributes like temperature, humidity, pressure, clouds, speed of wind and its direction. In classification method two things are very important. These are independent attribute and

dependent attribute. Suppose we want to predict the tomorrow's temperature from the given dataset. So, here the temperature attribute would be considered as dependent and other attributes like humidity, pressure, clouds, wind speed would be considered as independent variable. Similarly, to predict humidity it is considered as dependent attribute and others as independent attribute. By repeating this process for all other attribute we get our desired output model.

3.6 Implementations Requirements

We used some development tools to implement our system. We use python 3.7 as programming language.

- JetBrains PyCharm community edition 2018
- Spyder (Anaconda 3)

Python has a lots of built-in class, package and library to implement an algorithm avoiding details and a bunch of line of code. It makes our work easy. We imported sklearn and create an object of Gaussian Naïve Bayes, SVMC, Decision Tree class. We split the data set into train and test portion. Then, we passed them as parameter of function of different classification.

We display our output in a web page. To connect python with web there are many frame work. We used Django for Graphical User Interface (GUI). Taking function returned data we used chart.js, a javaScript framework to make graph like bar chart, line, pie chart, doughnut chart etc.

CHAPTER 4

Architectural Design

4.1 Planning

Use Case Diagram: In Unified Modelling Language (UML) use case diagram preliminary structure of system properties, modules and requirements for a new system design which is under developed. Use case focus on the expected behavior of a system and not anything about the real method of build. A notable features of use case diagram is that it helps us to design a system from end user‘s perspective. Use case diagram is very simple. It shows a system at a glance without detailing anything.

It is useful for visiting the system behavior subject to user term. It only summarize the internal architecture of a system. In earlier, it is said that a use case diagram is the blue-print of a system. Here is our blue print: [8]

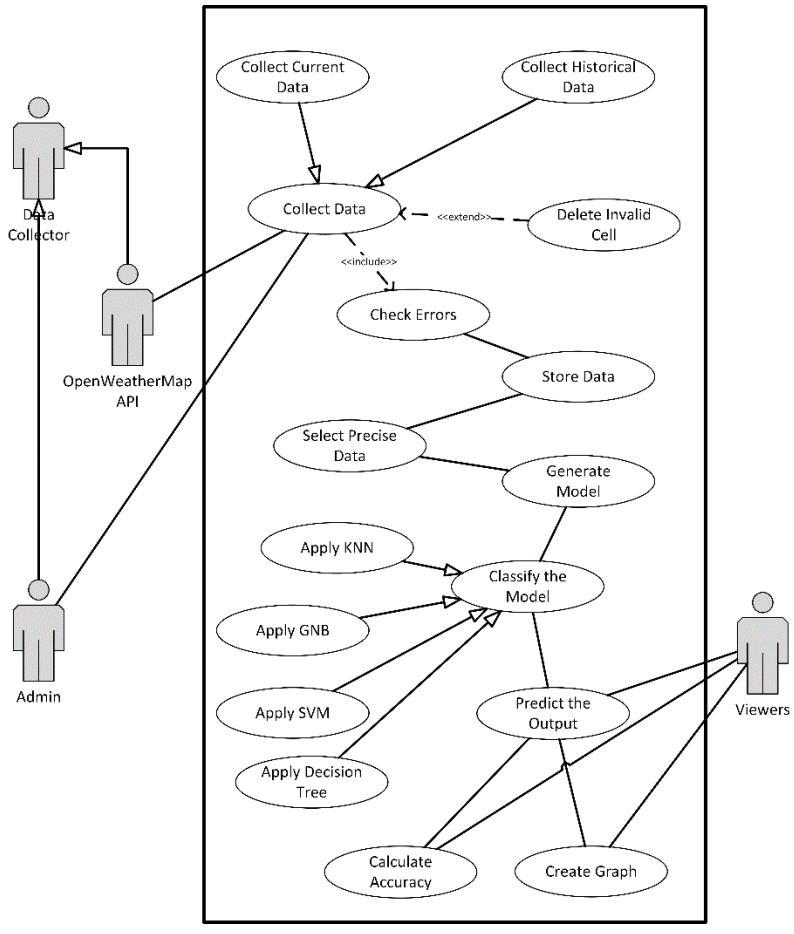


Figure 4.1: Use case diagram

4.2 Flow of Working Procedure

Data Flow Diagram: A Data Flow Diagram (DFD) shows the flow of data in any process or system. It uses some definite symbols like rectangular, circle and arrow. It shows input-output and both entities of front-end and back-end. It describes how input or raw data is taken from user and every steps of internal processing before reaching admin. Data Flow Diagram can be low level, mid-level and in-depth level. They can analyze the exact system. DFD is the one of the easiest way to describe something that is difficult to describe in word. It works for both technical, non-technical, from developers to CEO. [9]

DFD represents the function, module, process, which manipulate, capture, store, retrieve, implement, distributes data between a system and its environment and between its internal components. It is often used for following reason mentioned below:

- Uses very simple and less notation.
- Flow of logical information of the system.
- Make a list or determine of physical requirements of system. [10]

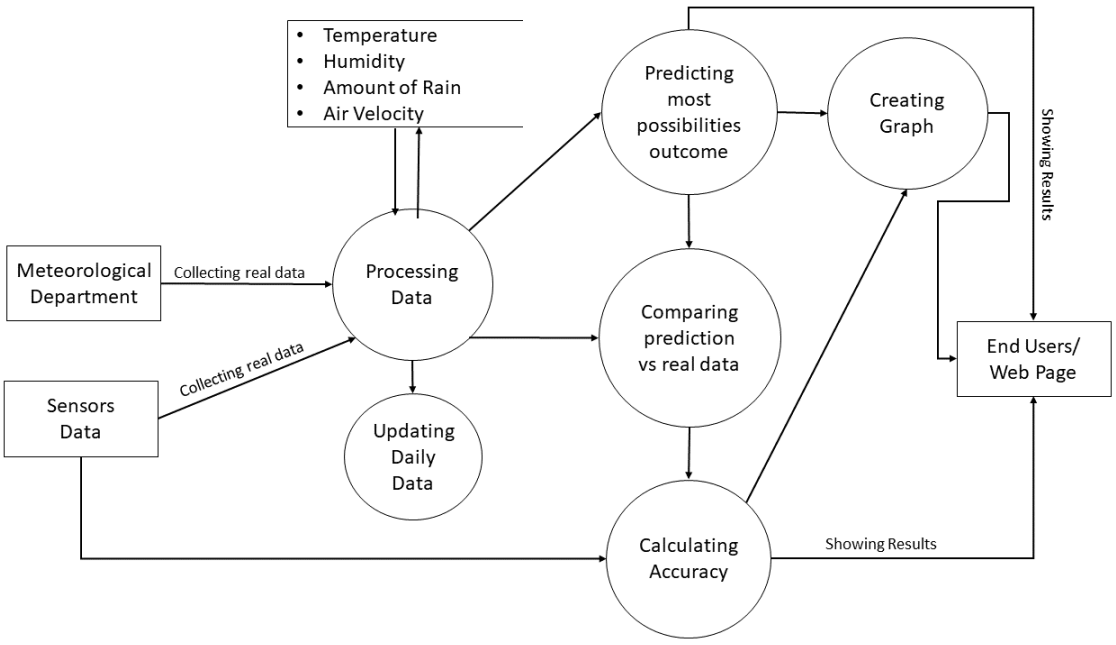


Figure 4.2.1: Data Flow Diagram 1

This is Data Flow Diagram 2 and it is more details of Data Flow Diagram 1. Here we describe the internal working of our system and data flow.

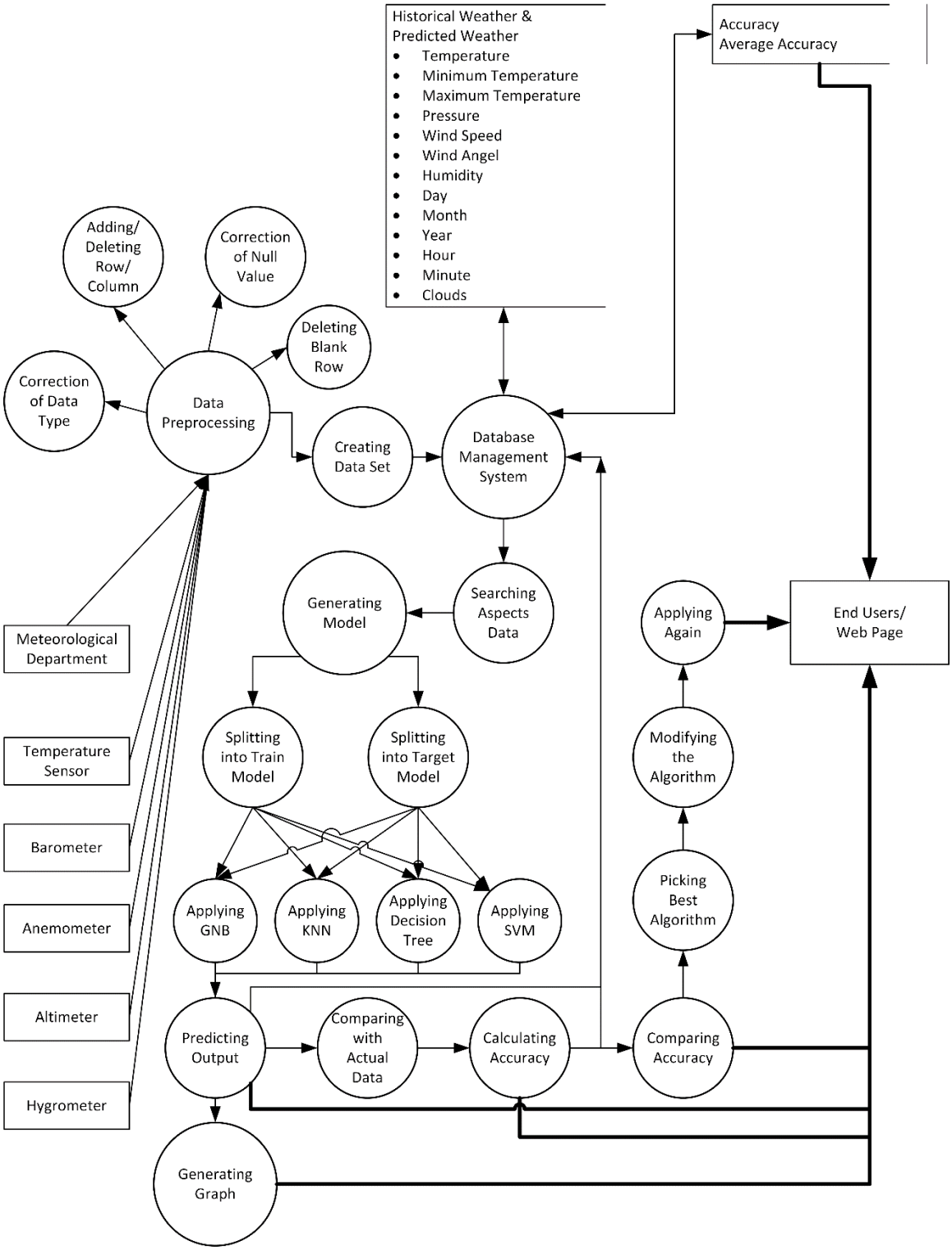


Figure 4.2.2: Data Flow Diagram 2

4.3 Database

Entity Relationship Model (E-R Diagram): Now a days, database is an essential part of a software system. For utilizing E-R Diagram in database engineering make sure us to produce high quality database design. An E-R Diagram shows relation among the entities. Tables are known as entity. A table has primary key, foreign key, derived attribute, multivalued attribute, compound attribute etc. A table has partial or full dependency, 1 to 1, 1 to many relationship with other table. Database also has week entity set and a week entity set must have a week relationship.

E-R Diagram represents all of these relation graphically. It is very useful to a database engineer to design a database and understand its infrastructure from E-R Diagram. Our system has a database. It contains three table named currentweather, forecastweather, accuracy. The table has attributes like temp, min_temp, max_temp, humidity, clouds, date, time etc.

This is our Entity Relationship model of our database. There are four tables in the database i.e. current_weather, city, predicted_weather, accuracy. Each table has different attribute which some of them are complex, some of them are derived etc.

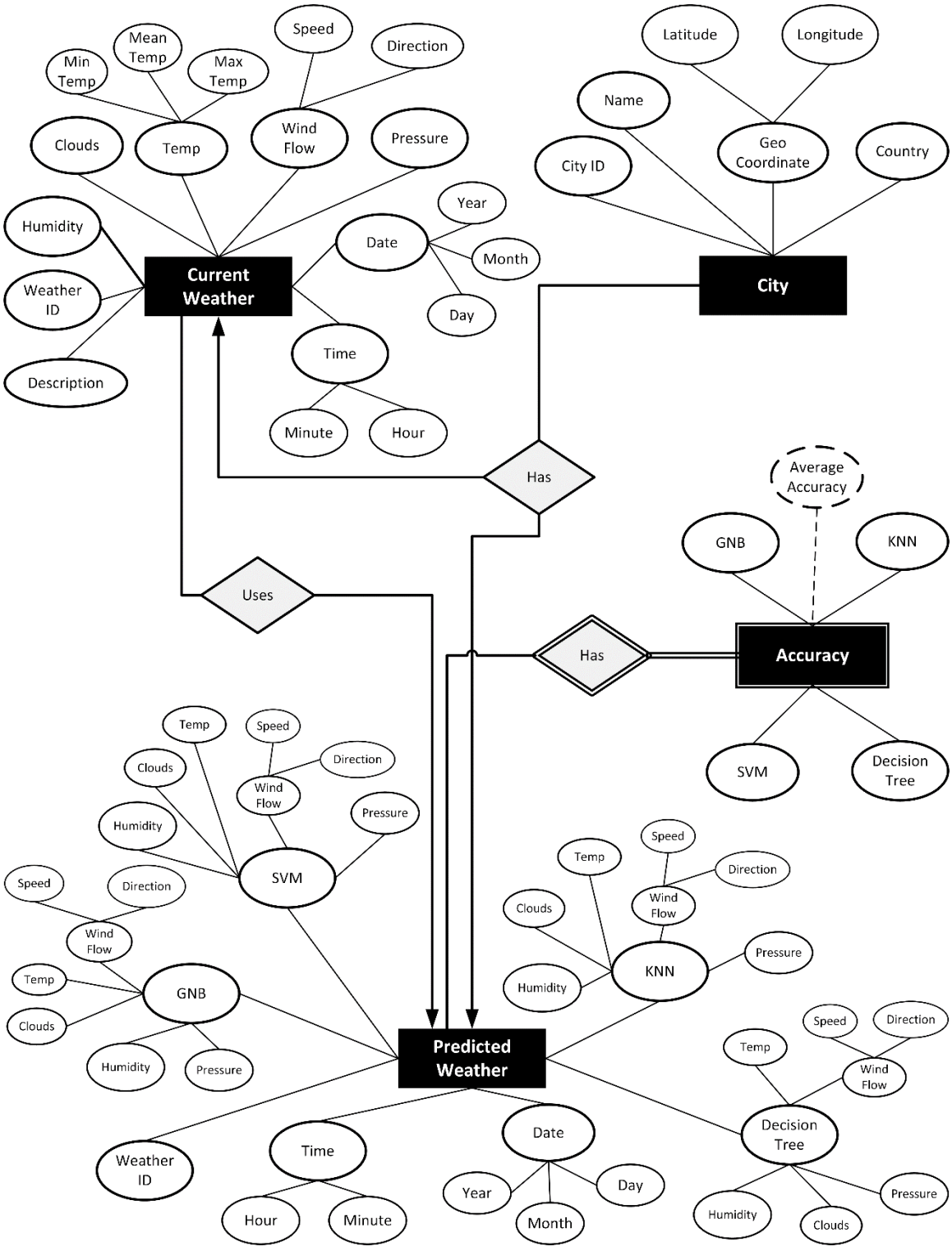


Figure 4.3.1: Entity Relationship Model

This is relational schema of our database. Relational schema is similar to the E-R diagram. But it also describes data types of attributes as well as their name, relation.

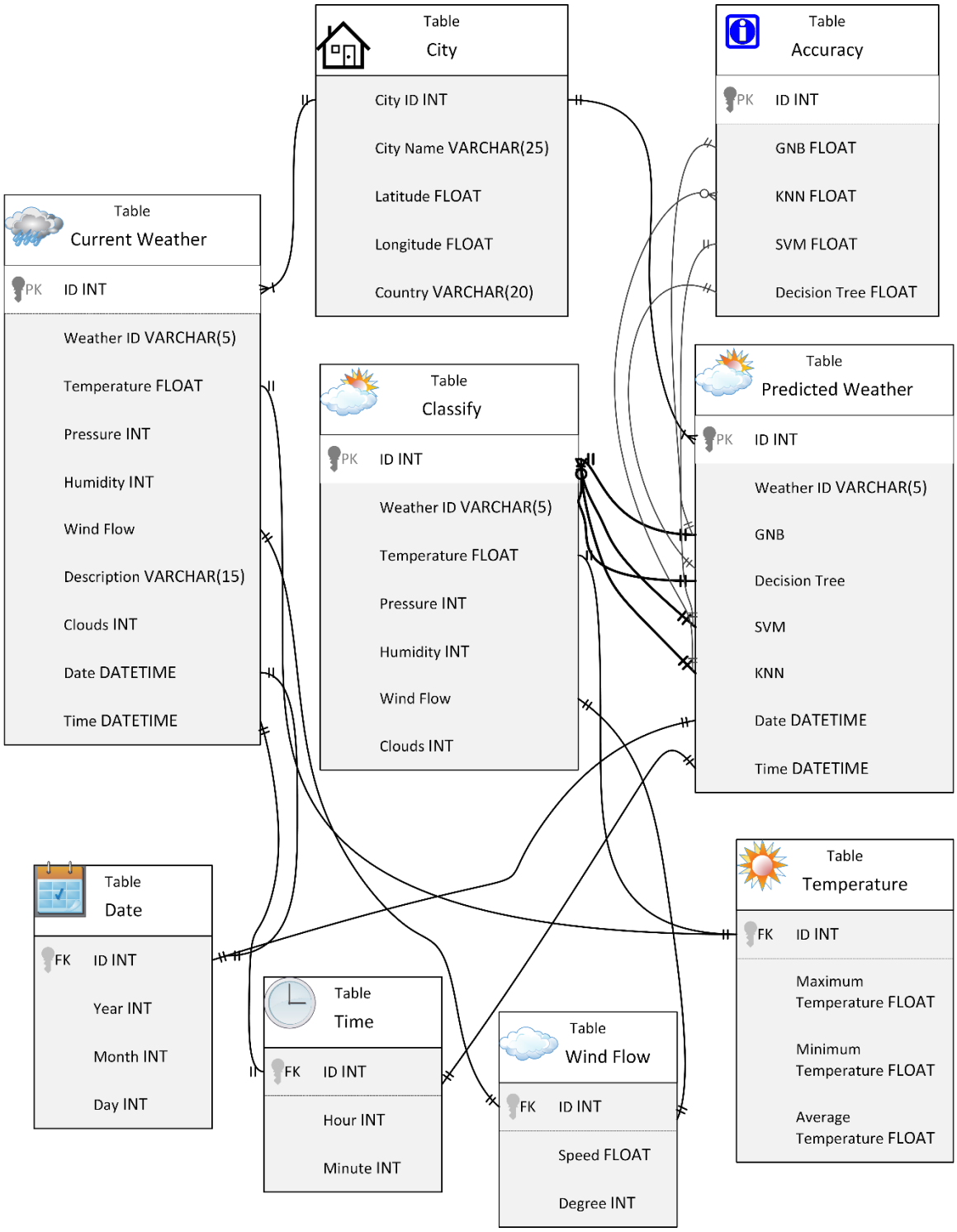


Figure 4.3.2: Relational Schema

4.4 Implementation

Class Diagram: Class diagram shows us a static view of a system. It explains all kind of attributes, methods of different class and make a graphical representation about their relation among themselves. It is more near to implementation of a system. Almost all object oriented language support class diagram. Object oriented programming is the collection of class. We can easily write necessary code from a class diagram.

In class diagram every class represented by a rectangular shape. It rectangular has three parts. The upper portion indicates the class name. The middle portion is list of all attribute with their data types. The last part hold all function or method with their parameter and return data type. An attribute or method can be public, private and protected. It can be represented using three notation that is +, -, # respectively.

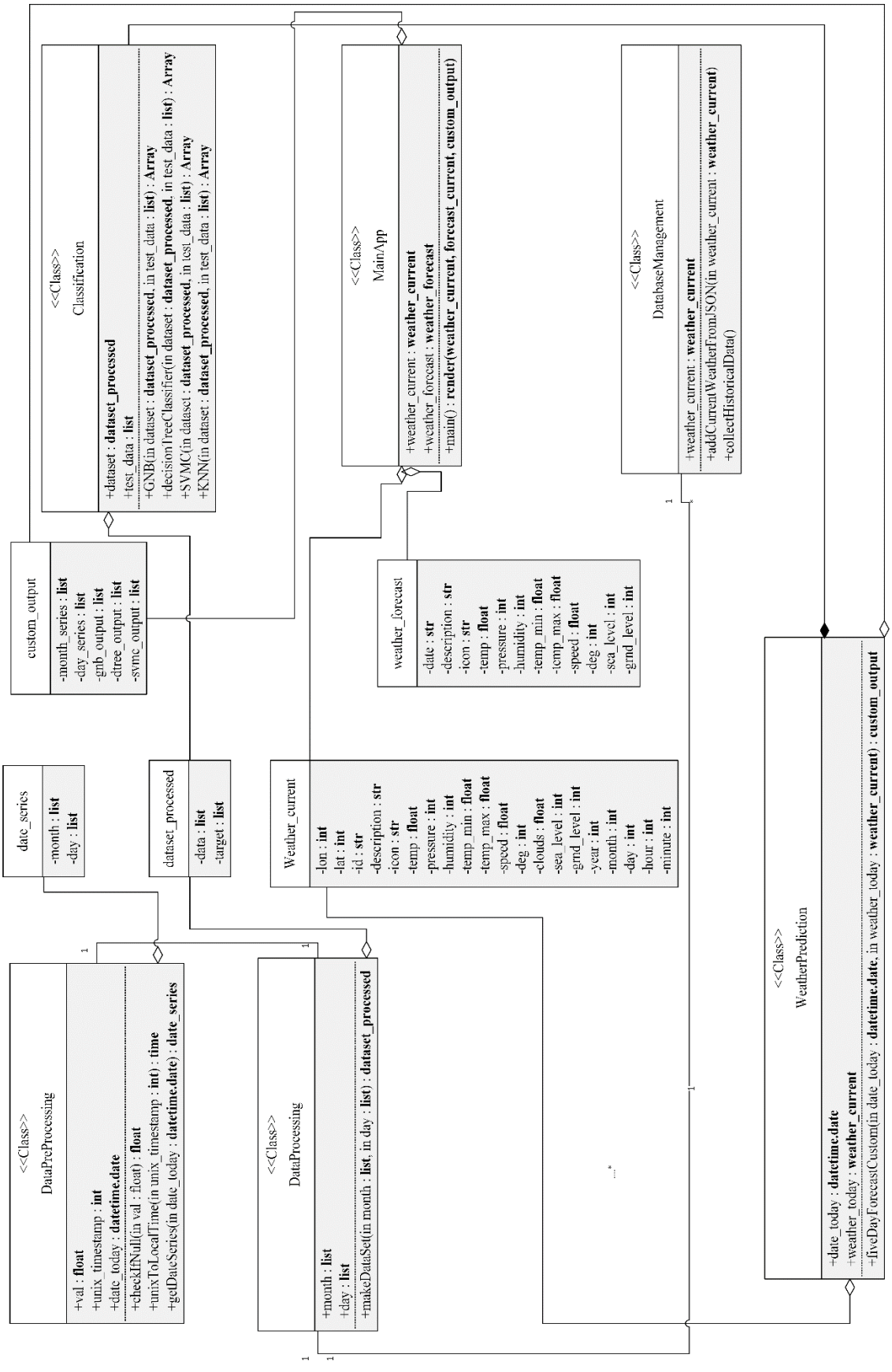


Figure 4.4: Class Diagram

CHAPTER 5

Experimental Results and Discussion

5.1 Introduction

Weather is the status of our environment which explain whiter it is cloudy or clear sky, thunderstorm is coming, hot or cold, if their possibilities of rain or flood or tornadoes. This chapter will describe the results of our study as well as described about the goal of study which was determine to predict the weather. At last what we expected from our project and actually what we got. Here we will display our output that we had got and briefly describe about it. To get more efficient result about predictions and forecasts we have to do some update in future.

5.2 Experimental Results

In this study, SVM, GNB, Decision Tree models are designed to forecast the weather attribute. List of weather attributes which to be predicted temperature, humidity, pressure, clouds and wind flow speed. For all models data set are generated as which date of weather would be predicted. Models are trained using 10 years historical data and daily real time data update.

1	Address	Date time	Minimum	Maximum	Temperati	Dew Point	Relative H	Heat Inde	Wind Spe	Wind Gus	Wind Dire	Precipitati	Precipitati	Snow Dep	Visibility	Cloud Cov	Sea Level	Weather T	Latitu
2	Dhaka	12/31/2018 18:00	64.5	64.5	64.5	53.7	67.98		3.4		330	0	0		2	0		Smoke Or	23.7
3	Dhaka	12/31/2018 19:00	64.5	64.5	64.5	55.5	72.59		3.4		310	0	0		2	0		Smoke Or	23.7
4	Dhaka	12/31/2018 20:00	62.7	62.7	62.7	53.7	72.41		0			0	0		2	0		Smoke Or	23.7
5	Dhaka	12/31/2018 21:00	59.3	59.3	59.3	53.7	81.87		0			0	0		1.9	0	1017	Smoke Or	23.7
6	Dhaka	12/31/2018 22:00	60.9	60.9	60.9	53.7	77.16		0			0	0		2	0		Smoke Or	23.7
7	Dhaka	12/31/2018 23:00	60.9	60.9	60.9	53.7	77.16		0			0	0		2	0		Smoke Or	23.7
8	Dhaka	01-01-19 0:00	60.9	60.9	60.9	55.5	82.39		0			0	0		1.6	0		Mist	23.7
9	Dhaka	01-01-19 1:00	59.1	59.1	59.1	55.5	87.84		0			0	0		0.9	0		Mist	23.7
10	Dhaka	01-01-19 2:00	62.7	62.7	62.7	55.5	77.32		0			0	0		0.9	0		Mist	23.7
11	Dhaka	01-01-19 3:00	66.4	66.4	66.4	54.8	66.4		0		340	0	0		1.7	0	1019.7	Mist, Smo	23.7
12	Dhaka	01-01-19 4:00	69.9	69.9	69.9	57.3	64.3		5.8		60	0	0		1.6	0		Smoke Or	23.7
13	Dhaka	01-01-19 5:00	73.5	73.5	73.5	55.5	53.33		4.7		70	0	0		2	0		Smoke Or	23.7
14	Dhaka	01-01-19 6:00	78.5	78.5	78.5	47.7	33.91		4.1		360	0	0		2.4	0	1018.4	Smoke Or	23.7
15	Dhaka	01-01-19 7:00	80.7	80.7	80.7	48.3	32.22	79.8	4.7		40	0	0		2.2	0		Smoke Or	23.7
16	Dhaka	01-01-19 8:00	80.7	80.7	80.7	44.7	28.13	79.5	5.8		60	0	0		2.2	0		Smoke Or	23.7
17	Dhaka	01-01-19 9:00	80.7	80.7	80.7	47.3	31	79.7	1.2		300	0	0		2.7	23.6	1015.4	Smoke Or	23.7
18	Dhaka	01-01-19 10:00	80.7	80.7	80.7	53.7	39.35	80.4	4.7		70	0	0		2.2	0		Smoke Or	23.7
19	Dhaka	01-01-19 11:00	77.1	77.1	77.1	55.5	47.3		3.4		90	0	0		2.2	0		Smoke Or	23.7
20	Dhaka	01-01-19 12:00	74.8	74.8	74.8	57.5	54.82		3.2		320	0	0		2.2	0	1016.8	Smoke Or	23.7
21	Dhaka	01-01-19 13:00	71.7	71.7	71.7	57.3	60.48		3.4		310	0	0		2.2	0		Smoke Or	23.7
22	Dhaka	01-01-19 14:00	69.9	69.9	69.9	59.1	68.59		0			0	0		2.2	0		Smoke Or	23.7

Figure 5.2.1: Dataset in CSV format

Figure 5.2.2: Dataset in JSON format

5.3 Descriptive Analysis

We made a sqlite3 database containing three tables i.e. one for current weather, another for predicted weather output and the last one for accuracy of prediction. In current weather table we put all historical data collected from different websites. We sort out this table as day and hour. We also store daily data from a weather API that means our data being reach day by day.

The second table is for predicted data. We stored daily prediction of four types of algorithm separately. That means we have predicted temperature, humidity, air pressure, wind speed, wind direction for all classification.

Lastly, we compared our predicted output with actual data and calculated accuracy. Then we stored all of these accuracy in accuracy table for comparing the frequent accuracy among them.

Some screenshots of our database table. We made our database in sqlite3 and viewed by sqlite3 browser.

Table: currentweather

New Record | Delete Record

	temp_min	temp_max	pressure	humidity	speed	deg	year	month
	Filter	Filter	Filter	Filter	Filter	Filter	Filter	Filter
1	21.6	21.6	1013	80	0.0	0	2017	1
2	18.0	18.0	1014	88	0.0	0	2017	1
3	16.4	16.4	1014	91	0.0	0	2017	1
4	19.8	19.8	0	0	0.0	0	2017	1
5	26.0	26.0	1014	54	0.0	0	2017	1
6	22.2	22.2	1013	64	0.0	0	2017	1
7	17.4	17.4	1013	80	0.0	0	2017	1
8	21.7	21.7	1014	66	0.0	0	2017	1
9	25.2	25.2	1014	57	5.4	330	2017	1
10	23.2	23.2	1011	64	0.0	0	2017	1
11	19.8	19.8	1013	79	3.6	270	2017	1
12	16.7	16.7	1013	84	3.6	360	2017	1
13	24.2	24.2	1013	53	5.4	330	2017	1
14	26.2	26.2	1011	39	0.0	0	2017	1
15	27.2	27.2	1012	47	3.6	340	2017	1
16	15.7	15.7	1011	96	3.6	350	2017	1
17	16.4	16.4	1014	91	3.6	310	2017	1
18	19.8	19.8	1013	79	3.6	270	2017	1

Table: currentweather

New Record | Delete Reco

	speed	deg	year	month	day	hour	minute	description	
		Filter	Filter	Filter	Filter	Filter	Filter	Filter	Filte
1		0	2017	1	1	12	0	null	0
2		0	2017	1	1	21	0	null	0
3		0	2017	1	2	0	0	null	0
4		0	2017	1	2	3	0	null	0
5		0	2017	1	2	6	0	null	0
6		0	2017	1	2	12	0	null	0
7		0	2017	1	2	21	0	null	0
8		0	2017	1	3	3	0	null	0
9		330	2017	1	3	6	0	null	0
10		0	2017	1	3	12	0	null	0
11		270	2017	1	3	18	0	null	0
12		360	2017	1	4	0	0	null	0
13		330	2017	1	4	6	0	null	0
14		0	2017	1	4	9	0	null	0
15		340	2017	1	4	21	0	null	0
16		350	2017	1	5	0	0	null	0
17		310	2017	1	5	3	0	null	0
18		270	2017	1	5	6	0	null	0

Figure 5.3.1: Table of database

This is home page of our system. Here we displayed the five day forecast fetching from API.

Figure 5.3.2: Home page view

Here the forecast data using three different classification algorithm which is done by ourselves.

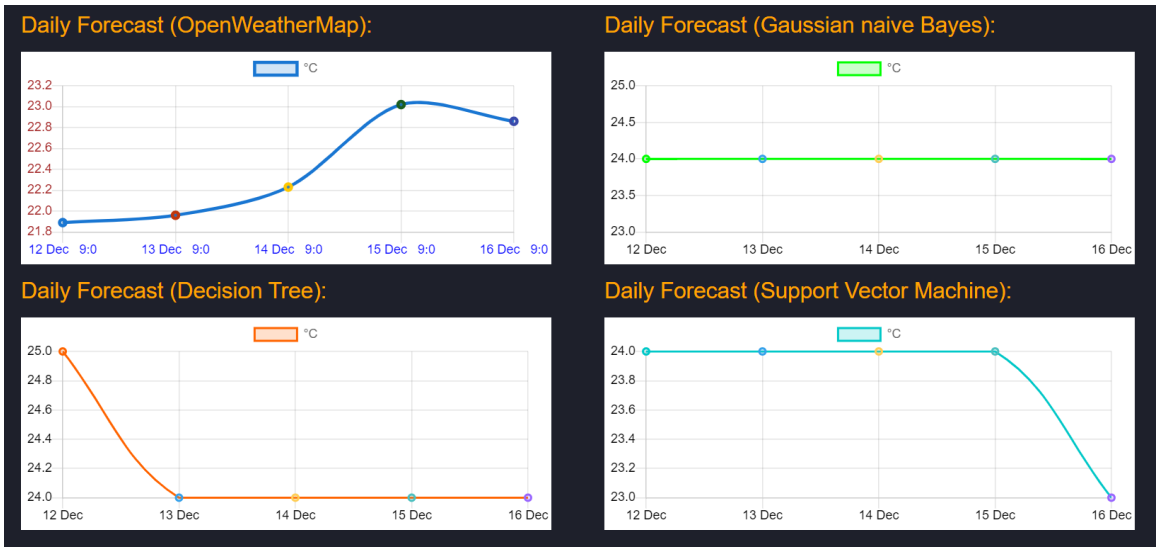


Figure 5.3.3: Output of different algorithms (5 day forecast)

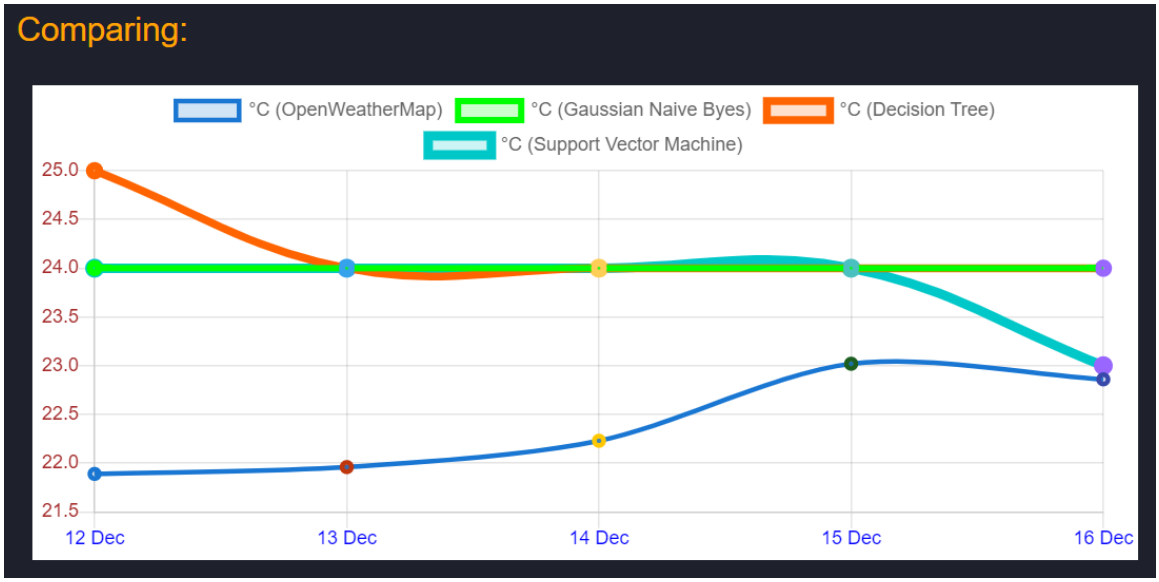


Figure 5.3.4: Comparison among the algorithmic output

Accuracy Table:

Date	GNB	Decision Tree	SVM
12-12-2019	90.36	85.79	90.36
13-12-2019	90.71	90.71	90.71
14-12-2019	92.04	92.04	92.04
15-12-2019	95.74	95.74	95.74
16-12-2019	95.01	95.01	99.39
Average Accuracy	92.77%	91.86%	93.65%

Figure 5.3.5: Accuracy table

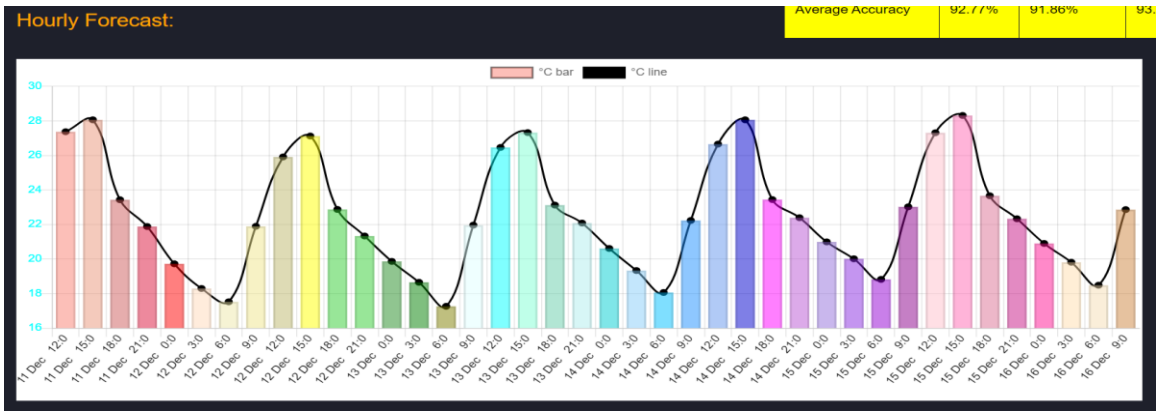


Figure 5.3.6: A graph of hourly forecast

5.4 Summary

Though scientist are trying to forecast weather from 19th century but there is a plenty of problems. Weather is predicted through the problems. Moreover, parameters of weather like temperature, humidity, pressure, clouds, rains etc. are not stable. They can change at any time. Weather is not predicted as well for it chaotic behavior. So, all time prediction is not good as what we expected. But yet we develop it generally. In the perspective of Bangladesh, this type of research is not sufficient and well developed. We are hardly trying from our level best to fulfill the requirements so that it help us to fill our necessity. Unfortunately there are some limitations such as data is not available for all the locations. All of us know that machine or computer works by depending on instruction or algorithm. It does not take any decision by its own. It only follows what we gives as program. Machine can decide what is wrong or right. It only takes a decision from trained parameter or previous data. We can no control the weather, climate or any natural events. But we are trying our level best to make a similarity by trying more and more times. Therefore, if anyone feel any lacking or improvement of our system, feel free to notify us.

CHAPTER 6

Summary, Conclusion, Recommendation and Implication for Future Research

6.1 Summery of the Study

There is no such prediction which gives us the absolute result and all of prediction has some lacks, limitations, errors. But we were trying our level best according to our ability to improve the prediction result. There are huge amount of prediction application, department, research team and websites whose work continuously to improve the result. We have also included ourselves in this journey. We study about lots of technique, methods, technology, device related to weather prediction. From all of them we had focus only on the improvement of algorithmic application. We choose data mining for this purpose. The implementation of this approach to improve the prediction of different weather parameters such as temperature, sea level, ground level, humidity, status of clouds, amount of rain, air pressure specially wind flow and temperature. Geographically, our country is so much natural disaster prone. In a year there are happened cyclone, flood and tornado for several time. We found 4 different classification algorithm such as Gaussian Naïve Bayes, Support Vector Machine, KNN, Decision Tree which are suitable to for us.

6.2 Conclusion

Our research and system offers a better and well organized weather prediction model using classification algorithm. Classification is such topic that is used in data mining and machine learning widely. This can be help us to get a trustworthy weather prediction. Out of the technical subject, our system has good impact in our daily life and our daily works. We can take perfect decision in our life with help of this system. Day by day our model will give better results. Because of updating our data sets regularly more precisely daily, the depth of our model is increasing day by day. We hope after a certain time, our system will give the output more near to the actual value.

6.3 Recommendations

Here we used historical data to do the prediction. But there is a serious drawback of the historical data specially the procedure of collecting and recording of data. There were many errors in historical data that made us disappoint so much. Wrong dataset causes bad prediction. So, we recommend some additional requirements. If we add hardware instruments such as sensor, barometer, compass for data collection purpose then it will be so realistic and help us to get best prediction.

6.4 Implication for Further Study

Weather prediction is not as a simple task as like other data mining and data prediction sector. From the outcome of this research we will make a better system. Weather prediction is not only software or programming jobs. It also depends much on hardware. Because for calculation of huge data we need a higher configuration computer or super computer and a big data center. So, this is all of our future plans.

APPENDIX

Appendix A: Project Reflection

We initiated our project work from Fall-2018. Our main goal was to improve the accuracy. We search, download and read hundreds of research paper. We found a lot of article about weather prediction process and read them with conscious mind. We had most problem to collect data. There were a few site which provided us a huge amount data in free. Our next challenge was to modify, editing, correction of data. This process was included to data preprocessing. Our next challenge was to convert this system to a web platform for making this user friendly for our viewers. After a long time hard working and our hearted efforts we complete our purpose.

Appendix B: Related Issues

Weather prediction is very complex and it demands a challenging skills and big amount of data. We use classification to classify the different weather attributes according to their class. As a result, our output will be always integer value because classification doesn't support floating type number. We used sqlite3 database management system to store data in database and web as GUI.

REFERENCES

[1] Weather Forecasting - Wikipedia, available at <<https://en.wikipedia.org/wiki/Weather_forecasting>>, last accessed on 31-10-2019 at 12:00 AM.

[2] 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).

[3] Basic Concept of Classification (Data Mining) – GeeksforGeeks, available at << https://www.geeksforgeeks.org/basic-concept-classification-data-mining/>>, last accessed on 31-10-2019 at 6:00 AM.

[4] Gaussian Naive Bayes - Lasse Schultebraucks – Medium, available at << https://medium.com/@LSchultebraucks/gaussian-naive-bayes-19156306079b>>, last accessed on 31-10-2019 at 6:45 AM.

[5] Decision Tree – GeeksforGeeks, available at <<https://www.geeksforgeeks.org/decision-tree/>>, last accessed on 31-10-2019 at 7:00 AM.

[6] Support Vector Machines (SVM), available at <<http://www.statsoft.com/textbook/support-vector-machines>>, last accessed on 31-10-2019 at 7:30 AM.

[7] Current weather and forecast – OpenWeatherMap, available at <<https://openweathermap.org>>, last accessed on 31-10-2019 at 7:55 AM.

[8] What is Use Case Diagram?, available at <<https://www.visual-paradigm.com/guide/uml-unified-modeling-language/what-is-use-case-diagram/ >>, last accessed on 31-10-2019 at 9:04 AM.

[9] What is a Data Flow Diagram | Lucidchart, available at <<https://www.lucidchart.com/pages/data-flow-diagram >>, last accessed on 31-1-2019 at 9:20 AM.

[10] What is Data Flow Diagram?, available at <<https://www.visual-paradigm.com/guide/data-flow-diagram/what-is-data-flow-diagram/ >>, last accessed on 31-10-2019 at 9:40 AM.

Weather Prediction

ORIGINALITY REPORT

6%	2%	0%	5%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to Daffodil International University Student Paper	1%
2	Submitted to City University Student Paper	1%
3	Submitted to University of Greenwich Student Paper	<1%
4	Submitted to The University of the South Pacific Student Paper	<1%
5	ijarcsms.com Internet Source	<1%
6	Submitted to Carlisle College Student Paper	<1%
7	iaescore.com Internet Source	<1%
8	Submitted to Graz University of Technology Student Paper	<1%
9	repository.um.edu.my Internet Source	<1%

10	ir.nctu.edu.tw Internet Source	<1%
11	Submitted to American Intercontinental University Online Student Paper	<1%
12	academic.oup.com Internet Source	<1%
13	Submitted to KDU College Sdn Bhd Student Paper	<1%
14	Submitted to University of Newcastle Student Paper	<1%
15	Submitted to Midlands State University Student Paper	<1%
16	Submitted to VIT University Student Paper	<1%
17	Submitted to The University of Manchester Student Paper	<1%

Exclude quotes	On	Exclude matches	Off
Exclude bibliography	On		