# Weather Status Prediction of Dhaka City with Machine Learning Algorithms

**A project report**

**Submitted by**

| | |
|---|---|
| Muhtasim Shafi Kader | ID : 181-15-10809 |
| Md.Shihab Mahmud | ID : 181-15-10961 |
| Jobeda Akter | ID : 181-15-10816 |

## Department of Computer Science and Engineering

## Daffodil International University

**Submitted to**

### Dr. Fizar Ahmed

PhD, Assistant Professor
Department of Computer Science and Engineering
Daffodil International University

**Submission date: April 18, 2021**

# TABLE OF CONTENTS

**TITLE**

# I.   Introduction

The weather has an important effect on daily life. It is something that can change any time without any further notice. Because of some changes in the atmosphere, this instability happens. Weather prediction, however, is a vital part of daily life. So, weather forecasting accuracy is very critical. Nowadays, the weather data are interpreted by supercomputers. They obtain raw data from space-launched satellite. But the data collected in raw format do not contain any insightful information. So, cleaning of those data is needed for giving input in the mathematical model.

This process is known as data mining. After cleaning the data, it is used to input into the mathematical model and predict the weather. In this research paper, the data of the previous months are collected and a dataset is created. For avoiding complexity, them on this are sorted in three seasons like summer, fall, and winter. Each of these seasons contains 4 months. Then, algorithms like linear regression, logistic regression, and Gaussian Naïve Bayes are implemented on those datasets.

Machine leaning algorithms that used to make this project efficient.
Here we are using machine learning algorithms like KNN (K Nearest Neighbour) algorithm which can be imported by KneighborsClassifier. We are using Using Decision Tree Classifie algorithm which can be imported by DecisionTreeClassifier.We are using Using RandomForest Classifier which can be imported by RandomForestClassifier. We are using Gradient Boosting Classifier which can be imported by GradientBoostingClassifier. And finally, we are using Support Vector Machine Classifier which can be imported by SVC.

All of these Machine learning Algorithm can be helped us to get the perfect prediction of a complete weather status of Dhaka City in a day. In the dataset we are taking data of Some Machine Learning algorithms have used to build models to predict weather status by analyzing Temperature (°c), Wind (km/h), Gust (km/h), Rain (mm), Humidity (%), Pressure (mb) data. By taking these data from the data set we will able to find out a complete multiple weather status of Dhaka City. After processing the data, we will find out the particular accuracy of different Machine learning approach.

We will try to collect the most accuracy rate among all of these Machine Learning algorisms. In the weather prediction, the prediction of rain is mostly common structure. But we are gathering information about a complete weather status with all this particular weather information. Here the accuracy will not remain the same in every terms. It will change by the algorithms. Which algorithm performs the best accuracy that will the best case for this prediction method.

# II. Methodology

The goal of this work is to use Machine Learning techniques to predict the weather of the next day in Dhaka City. So, in this section, there will be some detailed descriptions of the research work. For more clarification, research subject and instrumentation will be explained shortly. Data processing is a very important part of machine learning, so it will be described after that.

**1. Research Subject and Instrumentation:** Because this is a research job, it needs to be very well understood. Not only that, but analysis will also vary from the study because it can alter the result at any time. So work is really effective in interpreting those variations correctly. And instrumentation refers to the instruments or devices used in this investigation.

**2.Data Processing:** In machine learning, no work can be done without data. So, the key part of this research was collecting data, and it is difficult part too. Since locating or obtaining data is not as straight forward as it would seem. No source for all of the data was available. Here the data is about 2019 to 2020 weather report of Dhaka that we are using. But this work is on the "Summer." The techniques can be extended to the remaining datasets as well. Prediction can be done on any other variables like humidity, pressure, and wind speed from here. But here the research has done only for predicting temperature.

**3.Statistical Analysis:** When working with data, there were some errors regarding some missing data in the dataset. Those errors are needed to fix because the successful implementation of machine learning algorithms is dependent on correct pre-processing data. So, fixing the dataset becomes the main responsibility then. In Fig. 1, there is a flowchart which is about the working process of the research.
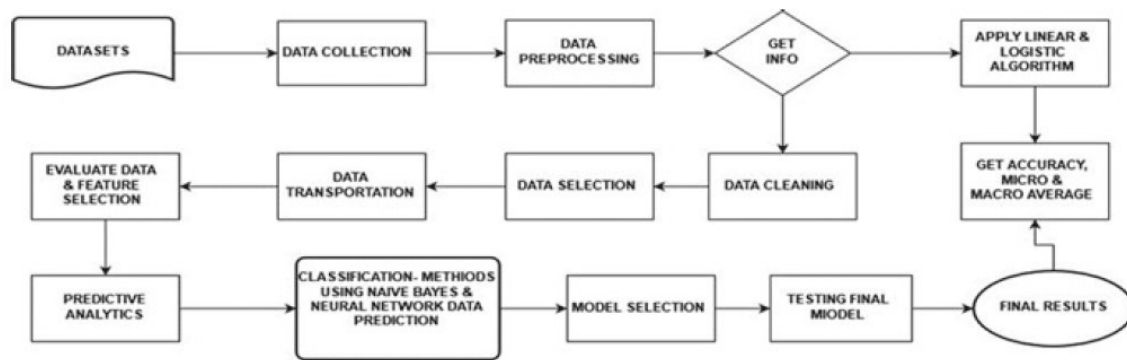


Fig 1 Flowchart for analysis

# III. Result and discussion

After training the data with four algorithms, the experimental model is built. Dataset was missing some values and needed to get them filled by using panda's method.

So that data can be more accurate. To build the regressor model, the dataset was separated into two parts:

- ✓ Training Dataset
- ✓ Testing dataset.

There is plot which is used for building the model. The four portions of the dataset are used as train data and rest 1 portion is used as test data. There were 122 data points and 15 attributes in the dataset. For training, 97 data points were taken and rest 25 data points were used for testing. To build the desired model, three different algorithms were used: linear regression, logistic regression, and Naïve Bayes. Also, with the help of,

1. Support Vector Machine (SVM),

2. Random Forest,

3. Decision Tree and

4. KNN

These are the algorithms that used to make the prediction of a complete weather information system.

## Experimental Result Between Various Algorithms:

After building models Accuracy, predicted the weather of different algorithms is added below. Algorithm details are given below:

## 1.Linear Regression:

Linear regression is one of the basic and renowned machine learning algorithms. It is a way of shaping relationships among variables. Linear regression comprises two kinds of variables: continuous variable and independent variable. The line's slope is m and c are the intercept (the y value if x = 0). MaxTemp and MinTemp appear in Fig. 2. In Fig. 3, the data are plotted in scatterplot so that the relationship can be visualize better.

Weather Status Prediction of Dhaka City Using …
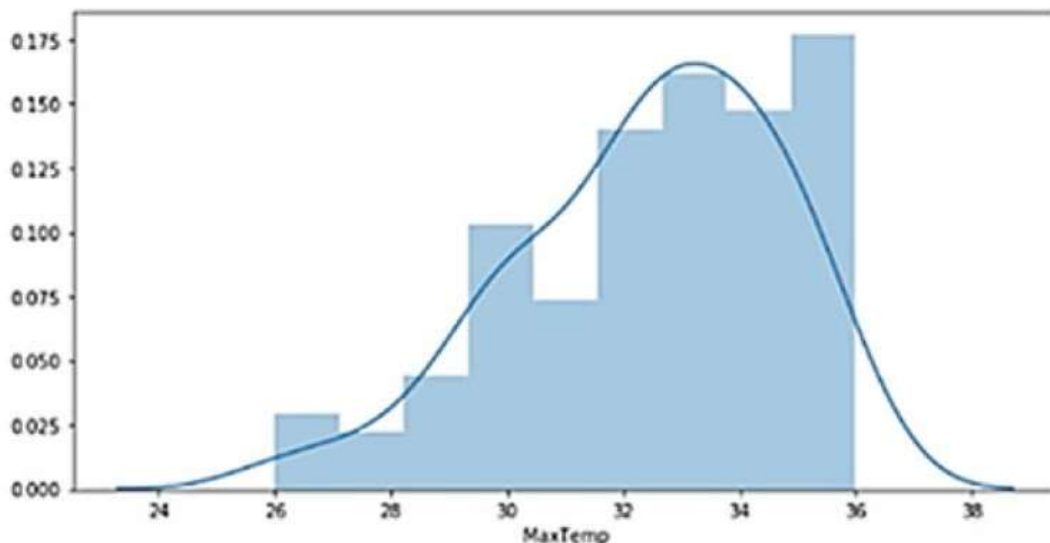
## 2.Logistic Regression:

Logistic regression is also a popular and frequently used algorithm for solving classification problem which linear regression cannot handle. Like if someone wants to separate positive and negative values from some given random values, then they need to use logistic regression. In this research, logistic classifier is used for classifying a mid-value from MaxTemp and MinTemp that is shown in Fig. 5. The logistic function takes any value between 0 and 1. The function is,

$$\sigma(t) = \frac{e^t}{e^t + 1} = \frac{1}{1 + e^{-t}} \tag{1}$$

In a univariate regression model, let us consider t as a linear function.

$$t = \beta o + \beta 1 x \tag{2}$$

The logistic equation would then become,

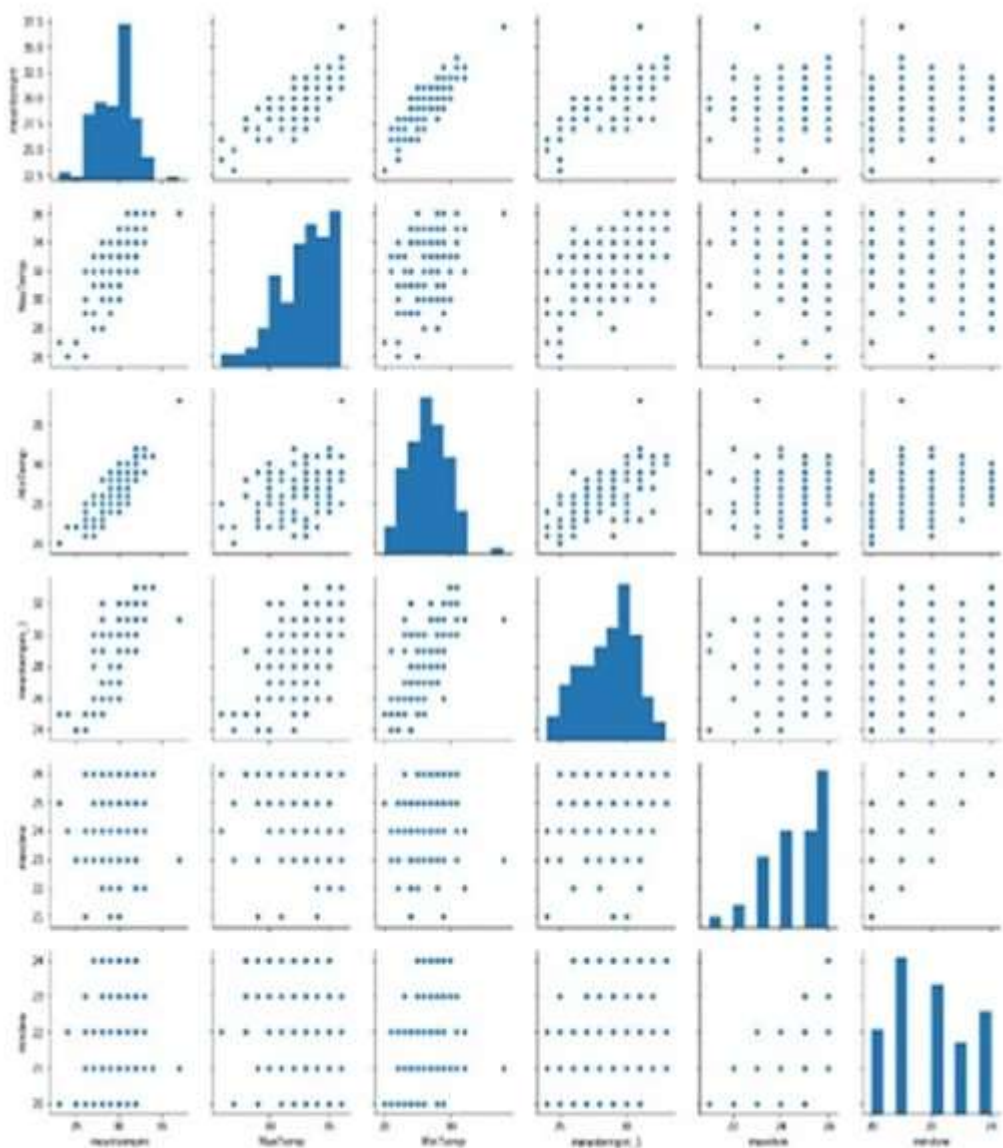$$p(x) = \frac{1}{1 + e^{-(\beta o + \beta 1 x)}} \tag{3}$$

Fig. 3   Scatterplot result

## 3.Naive Bayes:

It is a very popular classification technique based on Bayes probability theorem. Bayes theorem assumes that every parameter is independent in the analysis. This theorem is very useful for large datasets.
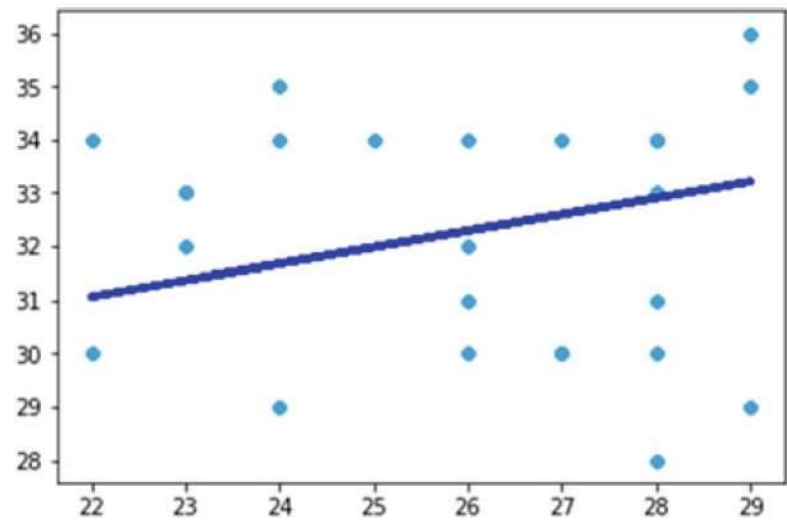
Bayes theorem measures P(a) from P(a), P(b), and P(b) as the likelihood. Here is the Bayes theorem equation:

$$P(a|b) = \frac{P(b|a)P(a)}{P(b)} \qquad\qquad (4)$$

Fig. 4 Visualization about prediction



Fig. 4 Visualization about prediction

Table 1 Actual value versus predicted value

|   | Actual value | Predicted value |
|---|---|---|
| 0 | 36 | 33.212321 |
| 1 | 30 | 32.597770 |
| 2 | 33 | 31.368670 |
| 3 | 29 | 33.212321 |
| 4 | 33 | 33.905045 |

**Fig. 5** Logistic regression classifier
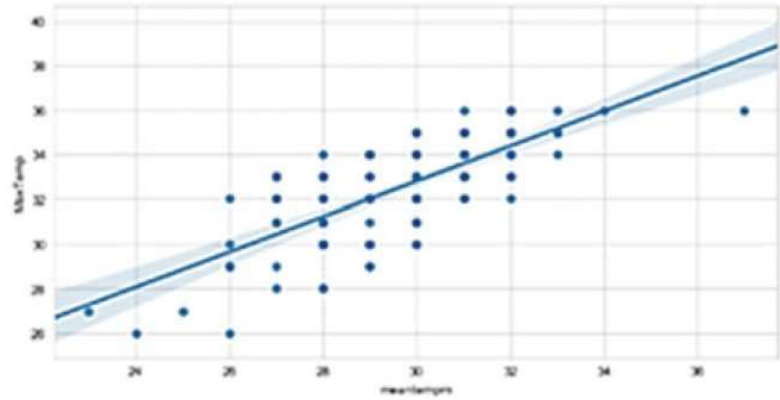
Then,

P(a|b) = P(b1|a) × P(b2|a)×· · ·× P(b|a) × P(a) (5)

To predict accuracy, a similar Naïve Bayes approach is used to predict the likelihood of different groups based on different data, as this algorithm is mainly used in the classification of text and with multiple class problems. The column was measured by the precap and is visualized in Fig. 6. The Naïve Bayes algorithm is visualized. In Table 2, there is the performance measurement of the Naïve Bayes classifier. The precision of a classifier defines the correctness and recall defines the entirety of the model. So, here the precision is 59% and recall is 25%.

**Fig. 6** Implementation of Naïve Bayes algorithm

**Table 2** Naive Bayes accuracy score

| Matric | Score (%) |
|---|---|
| Precision | 58.70 |
| Recall | 25 |
| Accuracy | 29 |

## Accuracy Comparison:

In Fig. 7, there is an accuracy comparison between three algorithms, which are used to build the model. It is shown that linear regression gives the best accuracy compared to the rest

# IV. Conclusion

The objective of this research is to create a model of weather prediction which will offer a forecast of the weather in Dhaka City tomorrow. The dataset is currently being introduced into the model as it is now different weather in Dhaka. Four algorithms were applied in this study. Support Vector Machine (SVM), Random Forest, Decision Tree and KNN provided Weather Status Prediction of Dhaka City Using adequate accuracy compared to others.
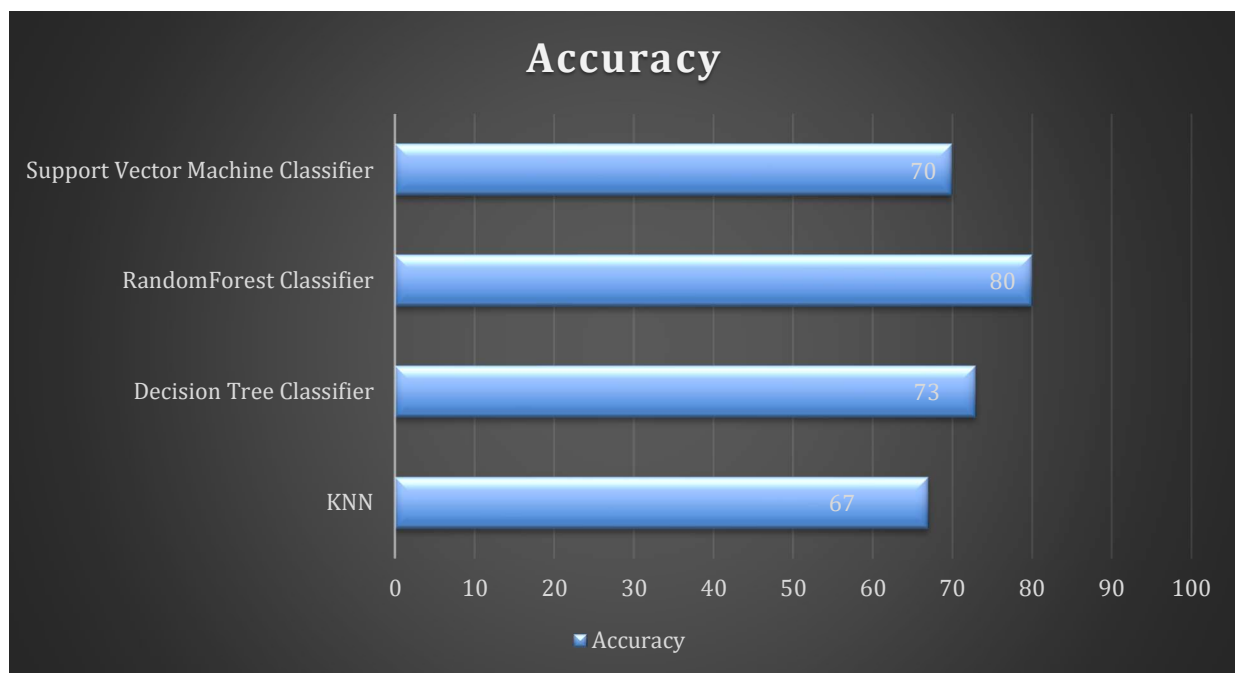


Fig.7 Accuracy comparison between algorithms

For the future, the key concern will be the deployment of the remaining three datasets (summer, fall and winter). Maybe along with these five algorithms, there will be a few other algorithms too, so that we can consider better weather forecasting techniques. We believe it would be important for potential weather data researchers to use more machine learning techniques.