

An hourly weather forecasting system for specific coordinates in Bangladesh using MERRA-2 reanalysis and machine learning

Project Thesis

Submitted By

16-32253-2	Mohammad Uzzal Hosen
18-36096-1	Rakib Hasan
18-36138-1	Abir-Al-Arafat
18-36094-1	SM Emran Hossain

Department of Computer Science

Faculty of Science & IT

American International University Bangladesh

July, 2022

Declaration

We declare that this thesis is our original work and has not been submitted in any form for another degree or diploma at any university or other institute of tertiary education. Information derived from the published and unpublished work of others has been acknowledged in the text and a list of references is given.



Mohammad Uzzal Hosen
16-32253-2
Department of Computer Science



Rakib Hasan
18-36096-1
Department of Computer Science



Abir-Al-Arafat
18-36138-1
Department of Computer Science



SM Emran Hossain
18-36094-1
Department of Computer Science

Approval

The thesis titled "An hourly weather forecasting system for specific coordinates in Bangladesh using MERRA-2 reanalysis and machine learning" has been submitted to the following respected members of the board of examiners of the department of computer science in partial fulfillment of the requirements for the degree of Bachelor of Science in Computer Science on 9th October, 2022 and has been accepted as satisfactory.

Sajib Hasan
Assistant Professor & Supervisor
Department of Computer Science
American International University-Bangladesh

Dr. M M Manjurul Islam
Assistant Professor & External
Department of Computer Science
American International University-Bangladesh

DR. Md. Abdullah- Al - Jubair
Assistant Professor & Head
Department of Computer Science
American International University-Bangladesh

Professor Dr. Tafazzal Hossain
Dean
Faculty of Science & Information Technology
American International University-Bangladesh

Dr. Carmen Z. Lamagna
Vice Chancellor
American International University-Bangladesh

Table of Contents

Table of Contents **3**

List of Figures **4**

List of Tables **5**

Acknowledgement **6**

Chapter 1: Introduction **7**

1.1 Background of Study 7

1.2 Statement of Problem 8

1.3 Goals & Objective 9

1.4 Significance of Study 9

1.5 Chapter Overview 9

Chapter 2: Literature Review **10**

2.1 Weather 10

2.2 Climate 10

2.3 Forecast and Prediction 12

2.4 Numerical Weather Prediction 13

2.5 Machine Learning 14

2.6 MERRA-2 15

2.7 Chapter Overview 15

Chapter 3: Methodology **16**

3.1 Data Source 16

3.2 Sample 17

3.3 Pre-Processing 18

3.4 Data Visualizations and Statistics 19

3.5 Algorithm Selection 21

3.6 Definition of terms 23

3.7 Chapter Overview 24

Chapter 4: Result Analysis **25**

Chapter 5: Conclusion **28**

References **29**

List of Figures

1.1 Bangladesh, showing meteorological stations used in the present study 11
3.1 Data access viewer by MERRA 2 16
3.2 Correlation Matrix 18
3.3 Pair plot for QV2M, RH2M, PS, WS2M, and T2M 19
3.4 QV2M in time series, representing T2M 20
3.5 RH2M in time series, representing T2M 20
3.6 WS2M in time series, representing T2M 20
3.7 PS in time series, representing T2M 20
4.1 Support Vector Machine Algorithm Confusion matrix demonstration with 72.51% accuracy 25
4.2 Random Forest Classifier Confusion matrix demonstration with 89.37% accuracy 25
4.3 Logistic Regression Confusion matrix demonstration with 27.03% accuracy 26
4.4 Decision Tree Confusion matrix demonstration with 92.67% accuracy 26
4.5 K Neighbor Classifier Confusion matrix demonstration with 91.77% accuracy 26
4.6 Gaussian Naive Bayes Confusion matrix demonstration with 17.06% accuracy 27

List of Tables

1.1 Major natural disaster in Bangladesh, 1991–2013 8
3.2 Collected Indices from MERRA 2 17
3.3 Sample of collected data from MERRA 2 after applying Coordinates 17
3.4 Preprocessing applied in MERRA2 data 18
3.5 Results 22

Acknowledgement

This research paper is on application of machine learning in weather forecasting. The main focus of the research was to develop a model that predicts weather parameters accurately. We thank all who directly or indirectly contributed to the completion of this thesis.

We are grateful to American International University - Bangladesh for creating this opportunity. Our heartfelt thanks to our honorable supervisor **Sajib Hasan,** Assistant Professor, American International University - Bangladesh. His instruction has contributed to every aspect of this research.

Also we are grateful to the Global Modeling and Assimilation Office (GMAO), NASA Goddard Space Flight Center for their dataset.

Chapter 1: Introduction

1.1 Background of Study

There has been a good amount of study throughout the years by meteorologists to observe, understand and predict the cyclic progression of weather systems. Using physical constraints like air temperature, humidity, air pressure, and sunlight density meteorologists usually can forecast the weather several days into the future. The utmost limiting factors to predictability beyond several days are the dynamic properties of the atmosphere.

In the 1960s metrologist, Edward Norton Lorenz discovered that minimal modifications to the basic state of the variables can produce a contrasting forecast. Chaos theory is about systems that are deterministic and whose behavior is theoretically predictable.Chaotic systems are predictable for a while, eventually, after some time they diverge into chaos. The temporal span over which the behavior of a chaotic system may be anticipated is determined by three factors: the level of uncertainty that can be tolerated in the prediction, the precision of the present state, and a time scale determined by the system's dynamics.

Here, The current weather condition predicts the future weather, while the approximate weather does not forecast the forthcoming weather. This is so-called the butterfly effect, the movement of a butterfly's wings in one place can create a tiny change in the atmosphere that might alter the path of a tornado or delay, accelerate or even prevent it. Of course, the butterfly’s wing flap does not have the power to create a tornado, but it is a part of the initial conditions, from which a set of conditions leads to a tornado [1].

In weather forecasts, such small effects as the flap of a wing, or lack thereof, have small effects when there are other factors such as high humidity, warm ocean temperature, and winds with low wind shear. Due to those factors, it's difficult to forecast weather parameters accurately. In consequence, more research is ongoing to improve forecast accuracy.

1.2 Statement of Problem

Many consequential weather events have affected mankind over the decades. We know because their effects have become part of the history of Bangladesh [Table 1.1]. Bangladesh is now at the top of the global climate risk ranking. Bangladesh is susceptible to climate change since it relies significantly on agriculture and natural resources for economic growth. [5]

Agriculture is Bangladesh's most important economic sector, with 62% of the country's population is employed in agriculture to augment their income. Approximately 87% of the population lives in rural areas and is either directly or indirectly dependent on agricultural production. Agriculture production is impacted by weather and climate. Even the weather has a significant impact on agricultural productivity. It has the potential to affect crop growth, overall yield, insect incidence, water and fertilizer requirements, and other agricultural activities during the growing season. [6]

Year	Disaster	People Killed	People Affected	Economic Damage
1991	Cyclone	138866	15438849	1780
1998	Flood	1050	15000050	4300
2007	Cyclone and flood	5344	13771380	2300
2009	Cyclone	509	12322220	1500
2012	Flood	344	11000380	1300
2013	Cyclone	50	100000	300

Table 1.1: Some of the notable natural disaster in Bangladesh, 1991–2013
Source: Ministry of Food and Disaster Management

Temperature is the most important element in determining rice cultivation time among numerous agricultural climatic variables. Rice is a summer crop in general, and when the temperature rises, the area accessible for rice production expands northward. The variety and growing method, in turn, must adapt to cope with the temperature shift. [7]

It is obvious that forecasting atmospheric parameters is critical for a variety of applications such as weather monitoring, drought detection, farming and production, energy sector planning, and so on. As a result, it is critical that we establish an appropriate model configuration that will allow us to anticipate those characteristics more precisely.

1.3 Goals & Objective

Initially hourly data for the recent 20 years from MERRA-2 (Modern-Era Retrospective analysis for Research and Applications, V-2) was collected. In order to comprehend space-based measurements of aerosols and describe their interactions with other physical processes in the climate system, the first long-term global reanalysis was conducted. In this project, trends from the data will be found, analyzed, and used to train a model that can predict weather changes in certain Bangladeshi coordinates over an hourly basis.

1.4 Significance of Study

Over the years there have been many efforts to forecast the weather more precisely and accurately. This study is to improve the ability to forecast, especially weather fluctuations, and improve the quality of warnings with high reliability. Which can be significantly effective for the country's sustainable development.

Since determining how to adapt to and plan for climate change is one of the most crucial aspects of the present study, it is believed that it will help guide initiatives in Bangladesh that are centered on agriculture, water resources, and disaster management. In Bangladesh, both the mean temperature and the average annual rainfall have increased significantly. The accumulation of moisture may have been

brought on by altered wind patterns brought about by the elevated sea surface temperature. The data and analyses provided by the current study, however, are insufficient to make any judgments on Bangladesh's exposure to global climate change.

1.5 Chapter Overview

Over the years weather forecasting has become an interesting topic for researchers. To predict more accurately a lot of studies are ongoing. As this can impact a country's economy and lives, it is important. The ultimate purpose of this study is to integrate a system that can predict the weather variables on an hourly basis.

Chapter 2: Literature Review

2.1 Weather

The atmospheric state which includes wind, temperature, humidity, state of the cloud, and pressure of the atmosphere is known as weather. These states always keep changing and influencing the weather which eventually causes the weather to fluctuate. The air gets heat from the sun that makes the air warm and causes it to rise and while in the process the air leaves an empty space which gets filled by the cold air later on. This whole process causes wind. Afterward, this wind and vapor water which is held by the air ends up making clouds. The state of these clouds is very important as it helps with rainfall and sometimes snowfall. Besides that, clouds can also form storms and help them to move. Climate change also has a huge impact on weather. Like when global temperature rises, global warming changes the weather patterns. The weather behaved differently than usual at that time. Variations in the weather like during the winter season, the temperature does not get much lower or during the rainy season, a lack of rainfall can be seen. These variations in weather can affect many aspects of human life in disruptive ways. Hence extreme variations can become a threat to the human race. Weather also has a huge impact on agriculture. Based on the change in weather, crops are grown and harvested in Bangladesh. Because photosynthetic activity, respiration rate, and natural disasters depend on weather which controls the lifespan of the crops. [13-14]

2.2 Climate

Climate is different weather conditions over a period of time in different locations. It includes conditions of a lower atmosphere for longer periods of time. These atmospheric conditions are rain, wind, snow, temperature, and humidity. Weather also includes these lower atmospheric conditions and coverage of overall weather conditions over a longer period of time. This period of time may vary from decades to longer time periods [13]. Based on different atmospheric conditions, the climate behaves differently in different regions of Bangladesh like some regions face abnormally heavy rainfall while other regions face drought, and there are also other regions that face heat waves. Besides that, severe fires in forests, water scarcity, melting polar ice, rising sea levels, declining biodiversity, flooding, and catastrophic storms happen due to climate change all around the world. These changes leave an impact on agriculture and human life. There are records that also show that the climate has been warm and relatively stable over the past 10,000 years [13]. One of the main reasons for these huge changes is global warming. The average increase in global temperature is known as global warming [7]. Burning fossil fuels, such as coal, oil, and gas, releases greenhouse gas emissions that can surround the globe and trap solar heat, causing the environment's temperature to rise. This is a major causes of global warming. The change in climate in one place or region which can be seen due to greenhouse emissions also affects the climate of other places in the world. To be more specific, it affects the health of people, food production, housing, safety, and work. Climate change solutions and climate prediction techniques can help to improve lives and protect the environment. [15]

Bangladesh is one of the nations most at risk from climate change in the globe, according to the IPCC (2007). Climate change is expected to result in a gradual increase in temperature and a shift in rainfall patterns, which might have a variety of consequences for agriculture, water resources, and public health.

Shamsuddin Shahid described recent changes in the climate of Bangladesh. 27 weather stations operated by the Bangladesh Meteorological Department (BMD) monitor daily rainfall and other meteorological factors. His study used 50 years (1958–2007) of rainfall records from those 17 stations to evaluate Bangladesh's recent climate change. Only 17 of these locations possessed long-term (>50 years) daily temperature and rainfall data. The 17 stations are dispersed around the country. [Figure 1.1] [18]

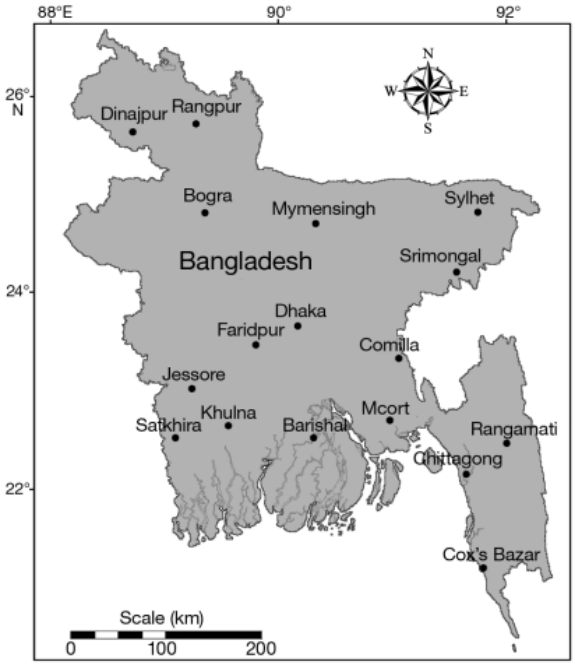


Figure 1.1: Bangladesh, showing meteorological stations used in the present study

Mohammad Shakhawat Hossain and others teffects of climate change on Bangladeshi agriculture were examined using the Ricardian model. Using household-level data from seven selected agroecological zones in Bangladesh, relationships between farmland value and long-term meteorological, farm household, and policy characteristics were assessed. The results show that farmers' current land values are vulnerable to climate change. [17]

2.3 Forecast and Prediction

The profession of weather forecasting involves utilizing science and technology to foresee the atmospheric conditions for a certain location and time. Weather predicting has been attempted by humans for millennia, and formal weather forecasting originates from the eighteenth century. It entails acquiring information about the atmosphere's current state and forecasting atmospheric advancement using a technical grasp of atmospheric dynamics. For individuals, weather forecasting is important. It assists in making better informed everyday judgments and avoiding risk. One of the most difficult tasks in the world is accurate weather forecasting.

To predict the weather, atmospheric data such as wind speed, wind direction, precipitation, temperature, humidity etcetera are needed. But these data are unstructured in most cases.Forecasting utilizing atmospheric data is difficult since there are many variables in weather forecasting that might alter depending on the state of the atmosphere. Meteorologists use different methods to forecast the weather. Some of the methods are based on the current atmospheric conditions whereas other methods use predictive principles, records of average weather data, and so on. These methods are chosen based on the amount of data available, the degree of accuracy, and the requirement of the prediction. These methods are known as statistical forecasting, persistence forecasting, synoptic forecasting and computer forecasting.

Persistence forecasting: It is considered the simplest way of forecasting. In this method, the weather condition of the future is based on the present condition. For example, if the weather is rainy today then it will be rainy tomorrow. It means this method works well only when there is a slight change in the weather.

Synoptic forecasting: This technique uses essential principles to analyze the set of synoptic charts. The primary concern is to use a common point to view the weather. Different atmospheric data are analyzed on a synoptic scale to develop large-scale weather patterns.

Statistical forecasting: In statistical forecasting, statistical data such as records of average temperatures along with other atmospheric data is used to understand the weather of a particular area during a particular period.

Computer forecasting: Different observations which can be found from different atmospheric data are applied to complex mathematical equations. These equations are run on different computers to make models. These models are used to give predictions of the weather for the next few days. Typically, the model's correctness and the equation's dependence on the model determine the outcome. A better equation might therefore result in a better forecast, and a better model depends on it.

By using these strategies, the condition of the weather for the next few days can be determined which may lead to many safety benefits and financial advantages. [8, 27]

2.4 Numerical Weather Prediction

In the last several decades, the relevance of weather forecasting as a study topic has increased. Forecasts for weather alerts are crucial because they protect lives and property. Traders rely on vital agricultural forecasts based on temperature and precipitation. In many regions of the world, people utilize weather predictions to make daily weather-related decisions. People use weather forecasts to decide what to do about the weather on a day-to-day basis in many parts of the world. Y.Radhika and M.Shashi described a weather prediction program using Support Vector Machines (SVMs). Based on daily maximum temperatures over the preceding n days, time series data of daily maximum temperatures at a location are analyzed to predict the highest temperature for the next day at that site (order of the input). They claimed that SVM outperformed MLP for various orders. For all orders, the findings reveal that SVM outperforms MLP trained with the backpropagation algorithm. Additionally, it was shown that the performance of the SVM model is significantly influenced by the choice of parameters. It might be claimed that Support Vector Machines can take part of the place of some neural network-based models in weather prediction applications with the right parameter choices. [20]

Maqsood described a modelthat provides a 24-hour forecast of the temperature, wind speed, and relative humidity at hourly intervals. Winter, spring, summer, and fall training and testing are done independently.The authors compared the Hopfield Model (HFM), Multilayer Perceptron Networks (MLP), Elman Recurrent Neural Networks (ERNN), Radial Basis Function Networks (RBFN), and ensembles of these networks. The recommended neural networks cannot only learn better but also generalize better than the established MLPN, ERNN, HFM, and RBFN models. We evaluated the forecasting reliability of these models based on several statistical measures. The modeling findings generally indicate that decent forecast accuracy was reached. The WTA ensemble had the most accurate predictions since it was based on the four different learning paradigms. The MLPN and ERNN networks both performed well in their forecasting of temperature, humidity, and wind speed. The HFM model had the lowest accuracy, while the RBFN model fared better than the MLPN and ERNN models. In terms of predicting weather parameters, ensemble networks are more accurate than regression models. [19]

A technique for predicting weather conditions using a meteorological dataset obtained from the MERRA database and combining a DT algorithm with a k-means clustering approach. The dataset is initially processed, with the attributes selected being temperature, pressure, humidity, wind speed, rainfall, snowfall, and snow depth. The C5.0 technique is the DT classifier used for the classification and prediction model as well as k-means clustering. The results of the proposed model are not substantially better in all the parameters. The recommended model fared better than the earlier methods in terms of accuracy, sensitivity, and specificity by 1.8 % to 6.9 %, 2.1 % to 7.2 %, and 1.7 % to 4.6 % respectively. Only short-term forecasts are appropriate for this research. [8]

F. Mekanik and M. A. Imteaz discovered that these major forms of complicated climatic factors impact Australian rainfall as well. They also discovered that major forms of complicated climatic factors impact Australian rainfall as well. Using Artificial Neural Networks, scientists identified a nonlinear connection between Victorian rainfall and lagged indicators impacting the region (ANN) (ANN). For the three case study stations of Horsham, Melbourne, and Orbost in Victoria, Australia, incorporating these indicators in an ANN model enhanced model correlation to 99 percent, 98 %, and 43 %, respectively. [22]

2.5 Machine Learning

Out of five papers on machine learning for weather prediction and climate change we examined, support vector machine was used in two of them and neural networks were used in the remaining. Neural networks' ability to capture nonlinear dependencies seems the reason to be used most.

Mark Holmstrom and others from Stanford University trained a model using two machine learning algorithms: linear regression and a variant of functional regression to represent an alternative to physical models in weather forecasting. They used the last two days of data to predict the next seven days of weather. Though the professional weather forecasting service outdated their models, they claimed that machine learning algorithms are robust to perturbations in initial conditions, and over time the model will do better.

Nitin Singh and others presented a system, and a Raspberry Pi 3 B application was made for it. This program forecasts the likelihood of rain on any given day using real-time information from humidity, temperature, and pressure sensors. Among the information are temperature, humidity, and pressure. The accuracy of this machine learning model might reach 87.90%. [10]

Ashfaq Ali Shafin and colleagues used machine learning techniques including Linear Regression, Polynomial Regression, Isotonic Regression, and Support Vector Regressor to identify the seasonal and annual patterns of Bangladesh's average temperature. While the Polynomial Regression and Support Vector Regressor algorithms successfully estimate the future average temperature, the Isotonic Regression approach successfully forecasts the training dataset. The average temperature for each month from 1901 to 2018 is included in the dataset used in this article. They suggest utilizing Polynomial Regression or SVR of greater degrees to forecast future temperatures. [9]

Some models have been used in the past for prediction such as RAMS [11] which is a cluster-based weather prediction system. Also, a different approach was shown by G.Vamsi Krishna that the prediction can be done using a backpropagation algorithm (Auto-Regressive Integrated Moving Average). [12]

An artificial neural network model has been created by researchers in Calcutta, Chattopadhyay and Chaudhuri, to forecast the minimum surface temperature and relative humidity needed for severe thunderstorm genesis over Calcutta. The findings demonstrate that single-layer neural networks outperform autoregressive process forecasting in terms of predicting yield. The outcome demonstrates the effectiveness of the single hidden layer ANN. [21]

2.6 MERRA-2

NASA Global Modeling and Assimilation Office (GMAO) produced the first global reanalysis of the satellite era to use satellite-based aerosol measurements. The interconnections between these physical processes and other processes in the climate system are represented by the Modern-Era Retrospective analysis for Research and Applications, Version 2 (MERRA-2). MERRA-2 is intended to replace MERRA; it is an improvement over MERRA and takes into account current developments in atmospheric modeling and data assimilation. [25-26]

2.7 Chapter Overview

There are many existing research papers available on weather prediction and the application of machine learning. Out of many relevant papers,the mentioned above were reviewed and found relevant to the goal of this research. Also, the data source of this paper MERRA-2 was discussed at the end of the chapter.

Chapter 3: Methodology

3.1 Data Source

In this study, we have used MERRA-2 data from the Global Modeling and Assimilation Office (GMAO) at NASA Goddard Space Flight Center for this research. Data was obtained from the https://power.larc.nasa.gov/ website, which provides meteorological data sets to support agricultural concerns, renewable energy needs, and building energy efficiency needs.



Figure 3.1: Data access viewer by MERRA 2, Image source: https://power.larc.nasa.gov/

A program was built and used to collect data from the website. Including visual representation [Figure 3.1], this site provides data in several other formats like GeoJSON, CSV, ASCII, and NetCDF format for specific coordinates for each hour from the year 2000. We have collected the data for each hour from 2000, January 1st, 6 AM to 2021, January 5th, 5 AM for each of the district coordinates of Bangladesh.

3.2 Sample

MERRA 2 reanalysis provides several points of data on solar and metrology, to apply in research for different communities of researchers in renewable energy, sustainable buildings, and agricultural climatology. We have collected only the ones we needed to use. Abbreviation with elaboration and unit available in [Table 3.2] and samples of data available in [Table 3.3].

Abbreviation	Elaboration	Unit
T2MDEW	MERRA-2 Dew/Frost Point (2 Meters)	C
T2MWET	MERRA-2 Wet Bulb Temperature (2 Meters)	C
QV2M	MERRA-2 Specific Humidity (2 Meters)	g/kg
RH2M	MERRA-2 Relative Humidity (2 Meters)	%
PRECTOTCORR	MERRA-2 Precipitation Corrected	mm/hour
PS	MERRA-2 Surface Pressure	kPa
WS2M	MERRA-2 Wind Speed (2 Meters)	m/s
T2M	MERRA-2 Temperature (2 Meters)	C
WD2M	MERRA-2 Wind Direction (2 Meters)	Degrees

Table 3.2: Collected Indices from MERRA 2

QV2M	RH2M	PS	WS2M	WD2M	T2M	T2MDEW	T2MWET	HR	DY	MO	YEAR
8.42	91.44	101.19	0.80	26.06	12.88	11.54	12.21	6	1	1	2001
8.73	86.50	101.25	0.74	34.20	14.27	12.08	13.18	7	1	1	2001
8.91	77.56	101.29	0.85	43.51	16.34	12.45	14.40	8	1	1	2001
9.03	63.75	101.31	0.84	21.31	19.69	12.65	16.16	9	1	1	2001
8.73	51.75	101.27	1.16	7.36	22.48	12.09	17.29	10	1	1	2001

Table 3.3: Sample of collected data from MERRA 2 after applying Coordinates

3.3 Pre-Processing

From the dataset we have collected from MERRA2 we first take the data set of dhaka district,in coordinate [90.4149,23.7302]. We have found 21 years of hourly data for the fields mentioned in [Table 3.3]. To understand the data better we have divided the data into yearly data from the census. Also, we have converted the day, month, and hour fields into one single field as the hour of the year is represented by column HY.

QV2M	RH2M	PS	WS2M	WD2M	T2M	T2MDEW	T2MWET	HY
8.42	91.44	101.19	0.80	26.06	12.88	11.54	12.21	6
8.73	86.50	101.25	0.74	34.20	14.27	12.08	13.18	7
8.91	77.56	101.29	0.85	43.51	16.34	12.45	14.40	8
9.03	63.75	101.31	0.84	21.31	19.69	12.65	16.16	9
8.73	51.75	101.27	1.16	7.36	22.48	12.09	17.29	10

Table 3.4: Preprocessing applied in MERRA2 data

3.4 Data Visualizations and Statistics

At first we calculated the correlation between the indices with respect to each other to find who has a higher impact for our target Temperature from the dataset for a specific coordinate of the last 20 years of Data. In the figure correlation matrix [Figure 3.2] it is absorbed that Temperature is positively correlated with QV2M (Specific Humidity 2 Meters (g/kg)) and negatively correlated with PS (Surface Pressure (kPa)).

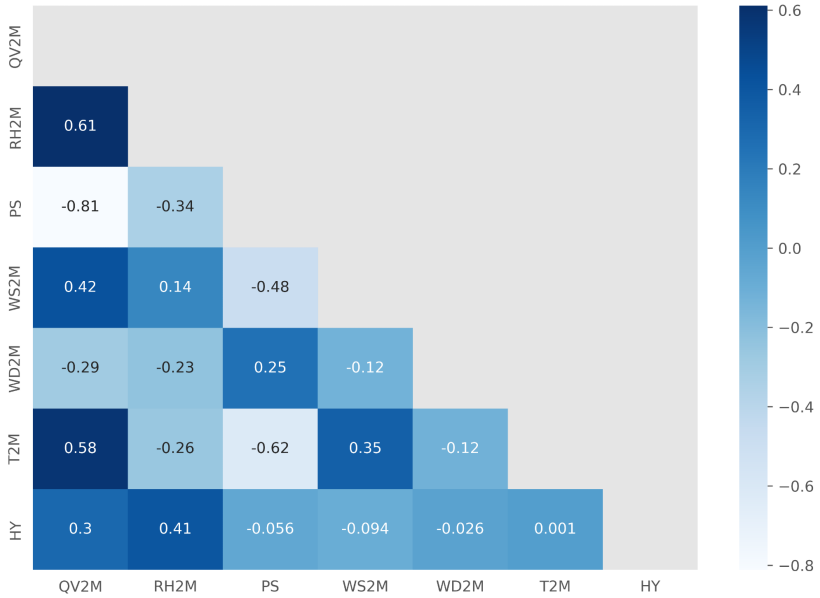


Figure 3.2: Correlation Matrix

It is also noticeable that QV2M is highly negatively correlated with PS and positively correlated with RH2M (Relative Humidity (2 Meters) %). There are also WS2M (Wind Speed at (2 Meter)) and WD2M (Wind Direction (2 Meter)) which are correlated but not significantly in this case.

We will be using this output from the correlation matrix in the selection of the estimator or the algorithm. The most correlated indexes will help to make a more accurate model and the less correlated indexes will increase the error rate.

From the correlation matrix observations, the selected most correlated items were visualized from the data of the year 2001 in Figure [3.3]. It can be observed that RH2M and PS have maintained a pattern in the change in temperature.

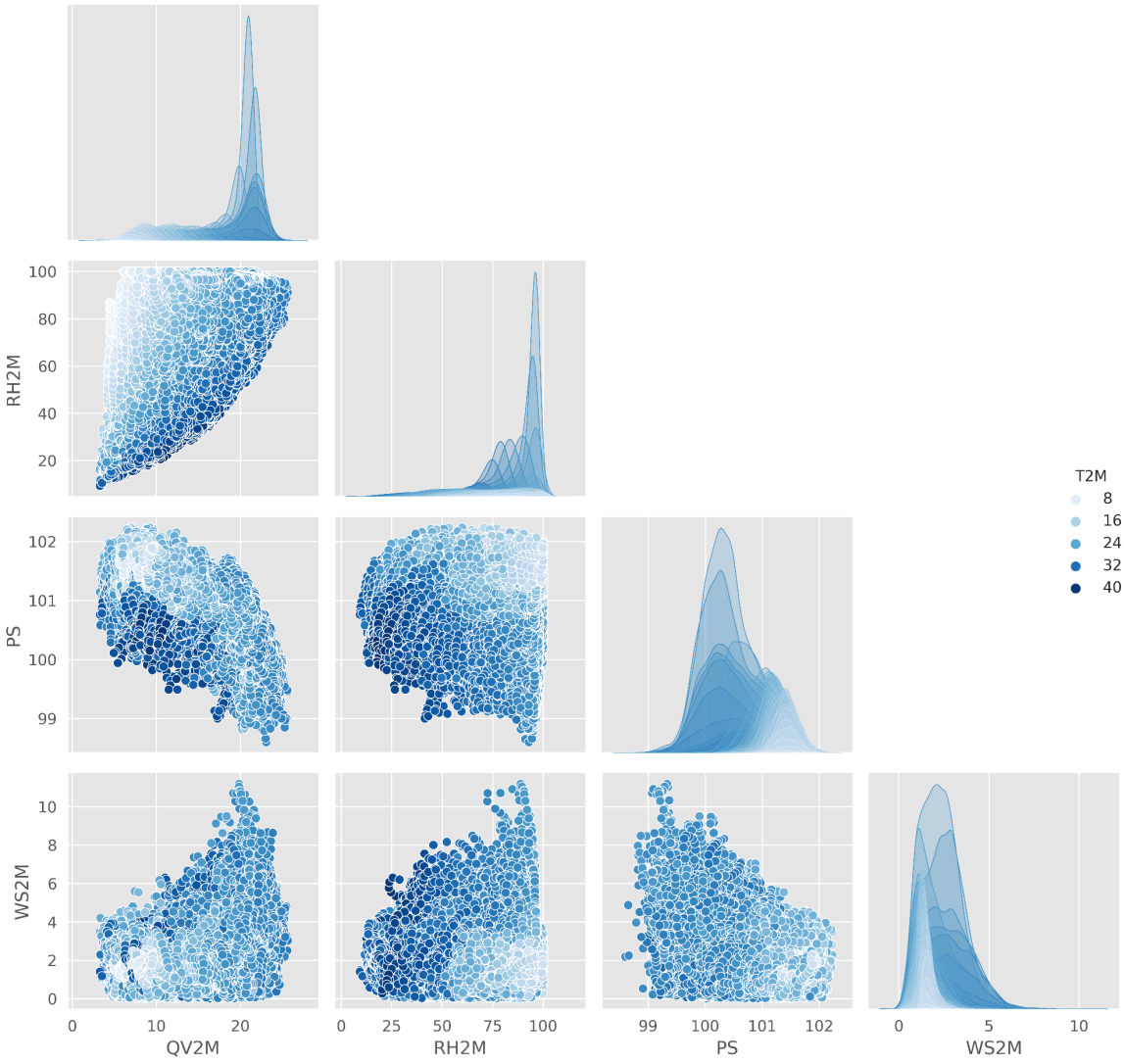


Figure 3.3: Pair plot for QV2M, RH2M, PS, WS2M, and T2M

After plotting the valuable changes in terms of time in the time series line plot, it can be observed how the temperature is maintained for each scope of that particular variable from Figure [3.4-3.7] dating from 2001 to 2021.

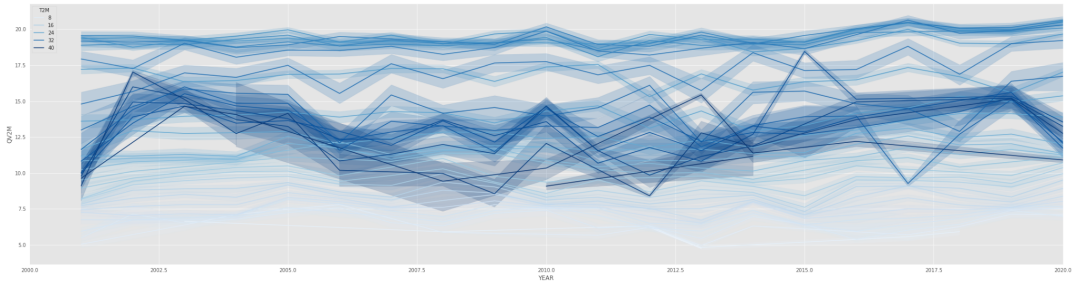


Figure 3.4: QV2M in time series, representing T2M

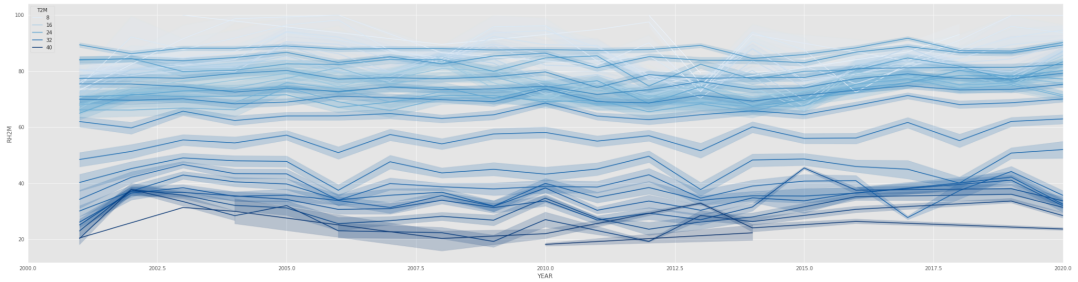


Figure 3.5: RH2M in time series, representing T2M

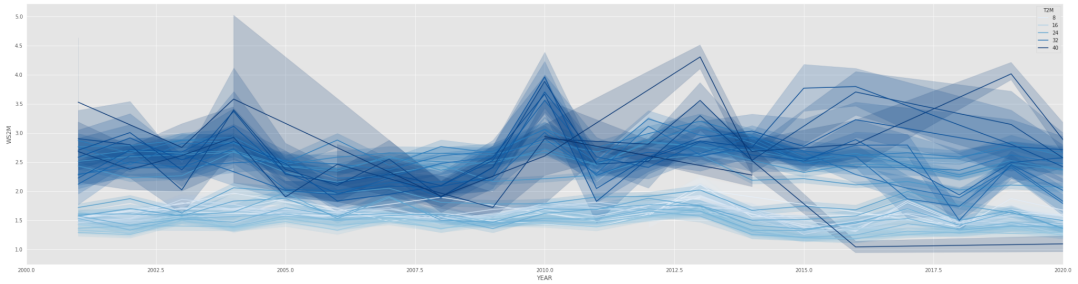


Figure 3.6: WS2M in time series, representing T2M

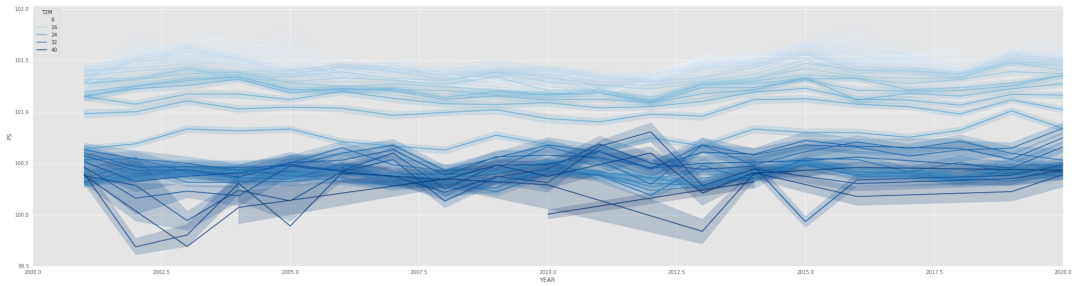


Figure 3.7: PS in time series, representing T2M

3.5 Algorithm Selection

Support Vector Machine

Support vector machine (SVM) examines data for regression and classification purposes. A supervised learning technique called SVM organizes data into one of two groups after looking at it. The sorted data are produced as a map by an SVM, with the margins between the two being as far away as feasible. SVMs are employed in the sciences of image classification, handwriting recognition, and text categorization. Supervised learning systems in AI and machine learning give input and intended output data that are labeled for categorization. A learning foundation for upcoming data processing is provided by the categorization. Two data groups are sorted using support vector machines according to similar categorization. The algorithms create hyperplanes (lines) to divide the groups into various configurations.

Decision Tree Classifier

The decision tree algorithm is a member of the family of supervised learning algorithms. Unlike other supervised learning techniques, the decision tree methodology can handle classification and regression problems. A Choice Tree is used to create a training model that may be used to predict the class or value of the target variable by learning simple decision rules obtained from prior data (training data). In decision trees, we start at the root when predicting the class label of a record. We compare the values of the attribute on the record with those of the root attribute. Based on the comparison, we go to the next node by following the branch that corresponds to that value.

Random Forest Classifier

Widely used machine learning algorithm Random Forest is an element of the supervised learning process. Regression and classification are only two of the many tasks that may be performed with the reliable machine learning algorithm Random Forest. It is an ensemble approach, which means that a random forest model is composed of several little decision trees, known as estimators, each of which generates a separate set of predictions. The estimators' estimates are combined with the random forest model to yield a more precise forecast.

Traditional decision tree classifiers are susceptible to overfitting the training set, which is a negative. This may be taken into consideration, and the random forest generalizes well to unseen data, including unobserved data with missing values, thanks to its ensemble architecture. Furthermore, random forests are excellent at handling huge datasets with a wide range of feature types and high dimensions.

Logistic Regression

An algorithm for classifying data is logistic regression. On the basis of a number of independent factors, it is used to forecast a binary result. What does this mean, then? There are only two outcomes that may occur in a binary situation: either the event occurs (1) or it doesn't (0). Independent variables are those variables or elements that might have an impact on the result (or dependent variable). Therefore, while working with binary data, you should utilize logistic regression as your kind of analysis. When the output or dependent variable is dichotomous or categorical in nature, or if it falls into one of two categories (such as "yes" or "no," "pass" or "fail," and so on), you know you are working with binary data.

K-Nearest Neighbor Classifier

The k-nearest neighbor's algorithm, sometimes referred to as KNN or k-NN, is a supervised learning classifier that employs proximity to produce classifications or predictions about the grouping of a single data point. Although it may be applied to classification or regression issues, it is commonly employed as a classification method since it relies on the idea that comparable points can be discovered close to one another.

Gaussian Naive Bayes

Naive Bayes classifiers are a subset of classification algorithms based on the Bayes Theorem. Instead of being a single technique, it is a family of algorithms, and they are all predicated on the notion that every pair of qualities being classified is independent of the other.

We projected T2M using the machine learning techniques mentioned above, and the forecasting results were excellent.

Algorithm	Score	Test Data
Support Vector Machine	0.7251	30%
Random Forest Classifier	0.8942	30%
Logistic Regression	0.2704	30%
Decision Tree Classifier	0.9267	30%
K Neighbor Classifier	0.9177	30%
Gaussian Naive Bayes	0.1706	30%

Table 3.5: Results

3.6 Definition of terms

Dew/Frost Point

The temperature at which airborne water vapor becomes saturated over a liquid surface is known as the dew point. When the temperature reaches the dew point, the relative humidity is 100 percent. The most popular methods for achieving 100% relative humidity are to chill the air to the dew point, evaporate moisture into the air until it is saturated, and elevate the air until it adiabatically cools to the dew point.

The frost point is the temperature at which air over an ice surface becomes saturated with water vapor. Because ice has a stronger connection between surrounding water molecules than a liquid surface, it is more difficult for new water molecules to escape. As a result, the temperature of the frost point is higher than that of the dew point. This fact is critical for cloud precipitation increase. Because the vapor pressure over an ice surface is lower than over a supercooled liquid surface at the same temperature, the relative

humidity over the ice surface will be more than 100 percent when the relative humidity is 100 percent with regard to water vapor.

Wet-Bulb Temperature

The temperature of adiabatic saturation is known as the wet-bulb temperature. A wet thermometer bulb exposed to the airflow indicates this temperature. With the bulb wrapped with wet muslin and a thermometer, one may check the temperature of a wet bulb. The air has a "wet-bulb temperature" that is lower than the "dry bulb temperature," which illustrates the cooling effect of the adiabatic evaporation of water from the thermometer. The rate of evaporation from the wet bandage on the bulb and the difference in temperature between the dry and wet bulbs are both influenced by the air's humidity. Evaporation is slowed down when there is more water vapor in the atmosphere. The wet-bulb temperature is always lower than the dry-bulb temperature when the relative humidity is 100%. (the air is at the saturation line).

Specific Humidity

Specific humidity is the ratio of water vapor mass to the air parcel's total (i.e., including dry) mass and is sometimes referred to as the humidity ratio. Specific humidity is approximately equal to the "mixing ratio", which is defined as the ratio of the mass of water vapor in an air parcel to the mass of dry air for the same parcel.

Relative Humidity

The ratio of an air-water mixture's partial pressure of water vapor (H20) to the saturated vapor pressure of water at a particular temperature is defined as relative humidity. As a result, the relative humidity of air is affected by both water content and temperature.

Surface Pressure

Surface pressure, also commonly known as atmospheric pressure, is the force exerted by air on the Earth's surface. Air has mass, and the gravitational influences on that mass result in air pressure. The pressure in the air varies depending on where you are and what time of year it is. Changes in air pressure, which occur throughout the year and in various places of the world, are substantially responsible for weather variations. The most common way to measure atmospheric wet-bulb is with a barometer, and the units used are Pascals (Pa). The atmosphere of the Earth exerts a pressure of 100,000 Pascals (100 kilopascals) on the Earth's surface as a whole (kPa). The average surface pressure of the Earth's atmosphere is represented by this value. In actuality, surface pressure varies a lot depending on where you are.

Wind Speed

The velocity of the wind is commonly used to determine wind speed. The majority of air movement measurements are collected from outside air, which can be influenced by several things. An anemometer is commonly used to calculate average wind speed, which is then classified using a defined measuring system known as the Beaufort Scale.

The pressure gradient, which is caused by a graded discrepancy in air pressure that occurs in various regions, is the most important component that influences wind speed. There is low pressure in certain locations and high pressure in others. A valley, for example, may have higher air pressure than a mountain top only a few miles distant. In most cases, the pressure between the two sites rises gradually.

Wind Direction

The wind direction is the direction in which the wind blows; it is also the direction in which the air moves. It is not the direction in which the wind is blowing. Wind direction is represented to the closest degree in a clockwise direction, or any of the 16 points on a compass, regarding 360 degrees on the compass (true North) (N, S, SE, WSE, etc...). [28]

3.7 Chapter Overview

In this chapter data source, collection, and processing of the data were interpolated including the definition of different data portions of the data set. Also, different forecasting algorithms were applied to find the best forecasting results.

Chapter 4: Result Analysis

After applying the algorithms mentioned in Table [3.5] in the data set we have achieved different result scores based on the accuracy of the prediction. To demonstrate the performance of the models which were trained to forecast weather, more clearly we have applied confusion matrices and plotted the results in heatmaps for all the machine learning algorithms mentioned in Table [3.5] and those are represented in Figure [4.1-4.6] where upper diagonal corner represents actual data and lower diagonal represents predicted data.

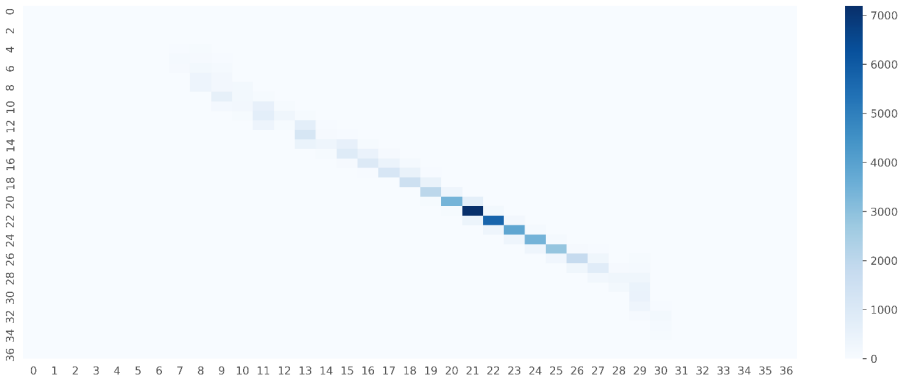


Figure 4.1: Support Vector Machine Algorithm
Confusion matrix demonstration with 72.51% accuracy

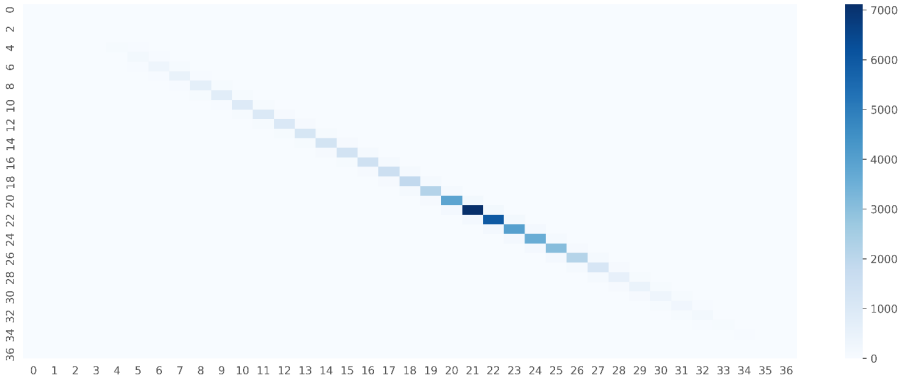


Figure 4.2: Random Forest Classifier
Confusion matrix demonstration with 89.37% accuracy

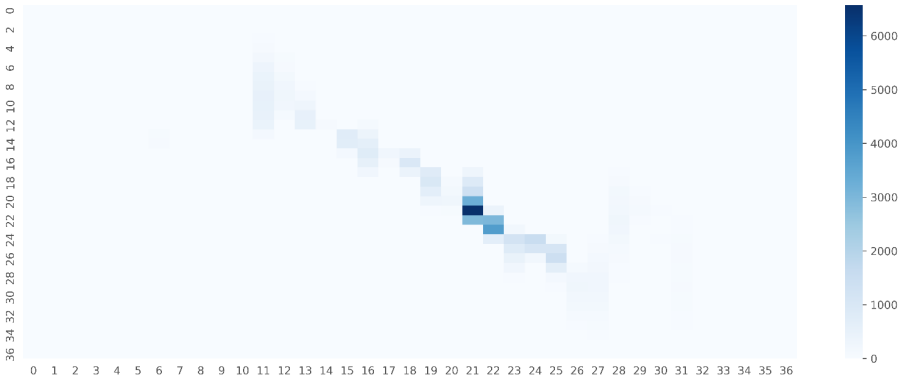


Figure 4.3: Logistic Regression
Confusion matrix demonstration with 27.03% accuracy

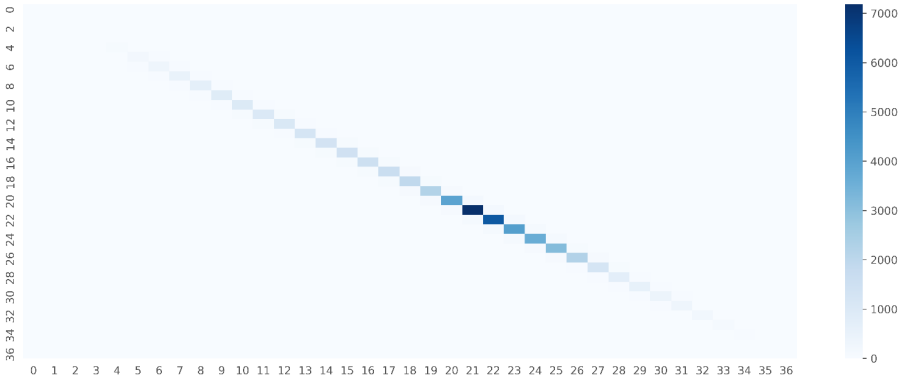


Figure 4.4: Decision Tree
Confusion matrix demonstration with 92.67% accuracy

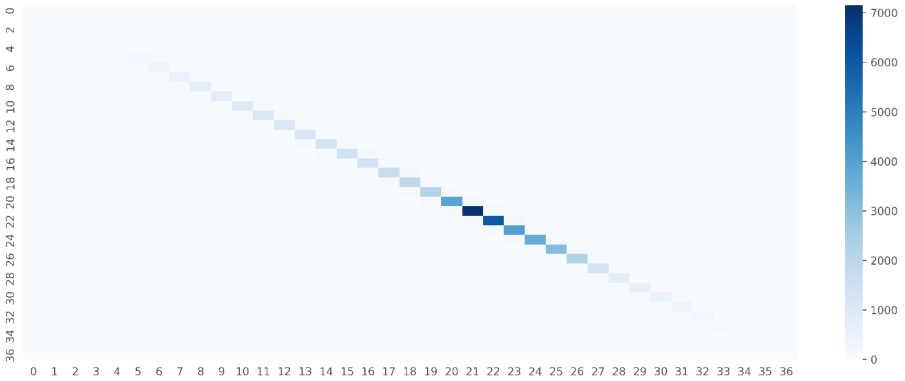


Figure 4.5: K Neighbor Classifier
Confusion matrix demonstration with 91.77% accuracy

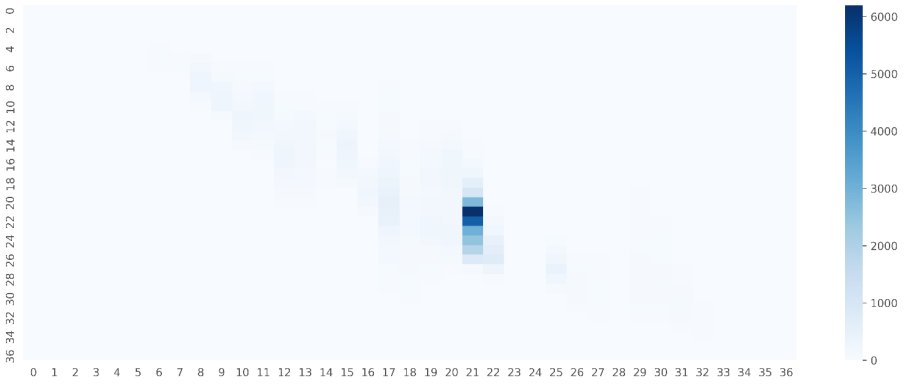


Figure 4.6: Gaussian Naive Bayes
Confusion matrix demonstration with 17.06% accuracy

From the score of 30% test data and the visual representation of the confusing matrix it can be observed that the Decision Tree and the K Neighbor Classifier machine learning algorithm forecasted T2M more than 90% accurately. Among those two, Decision Tree obtained the highest accuracy 92.67%. Also, Random Forest Classifier and Support Vector Machine forecasted 89.37% and 72.51% accurately. Gaussian Naive Bayes and Logistic Regression forecasted very poorly which was less than 30% accurate. This is reasonable because Logistic Regression is compatible with binary results, but here weather data is continuous. And to work with Naive Bayes, data needs to be independent of one another, here data is temperature dependent on multiple variables on multiple occasions.

Chapter 5: Conclusion

Several hypothesized weather prediction models, including Support Vector Machine, K Neighbors Classifier, Decision Tree, Random Forest Classifier, Logistic Regression, and Gaussian Naive Bayes, were developed to forecast weather conditions. This prediction model was constructed using the following steps: the first stage was to analyze the dataset, and then attributes such as temperature, pressure, humidity, and wind speed were considered. The attributes that were chosen are utilized for training and testing using the suggested model. The classifier was therefore trained utilizing 70% of the data, and the remaining 30% was utilized to test the model. In the last phase, the classifier predicted the temperature based on the training and testing.

At the end of this research, we can conclude by extrapolating that Decision Tree and K Nearest Neighbor performed very well in terms of MERRA2 data as they both performed forecasting very well. Also, Random Forest Classifier gave an accuracy near 90%.

Temperature forecasting alone is not particularly effective for weather-dependent aspects in Bangladesh. Other factors like as rain, flood, cyclones, and drought have a significant impact on the economy and population. This research can be expanded in the future to anticipate those phrases more precisely and to improve current capabilities to generate more accurate forecasts.

References

1. Lorenz, E. (1972). *Predictability: does the flap of a butterfly's wing in Brazil set off a tornado in Texas?* (p. 181). na.
2. Gelaro, R., McCarty, W., Suárez, M. J., Todling, R., Molod, A., Takacs, L., ... & Zhao, B. (2017). The modern-era retrospective analysis for research and applications, version 2 (MERRA-2). *Journal of climate*, *30*(14), 5419-5454.
3. R Usha Rani, T.K. Rama Krishna Rao, R. Kiran Kumar Reddy, "An Efficient Machine Learning Regression Model for Rainfall Prediction", International Journal of Computer Applications (0975 – 8887) Volume 116 – No. 23, April 2015
4. Hong S-Y, Lee J-W (2009) Assessment of the WRF model in reproducing a flash-flood heavy rainfall event over Korea. Atmos Res 93:818–831
5. Ali, Isahaque; Hatta, Zulkarnain A.; Azman, Azlinda (2014). Transforming the Local Capacity on Natural Disaster Risk Reduction in Bangladeshi Communities: A Social Work Perspective. Asian Social Work and Policy Review, 8(1), 34–42.
6. Hatfield, J. L., Boote, K. J., Kimball, B. A., Ziska, L. H., Izaurralde, R. C., Ort, D. R., ... & Wolfe, D. (2011). Climate impacts on agriculture: implications for crop production
7. Kim, C. G. (2012). The impact of climate change on the agricultural sector: implications of the agro‐industry for low carbon, green growth strategy and roadmap for the East Asian Region.
8. Murugan Bhagavathi, S., Thavasimuthu, A., Murugesan, A., George Rajendran, C. P. L., Raja, L., & Thavasimuthu, R. (2021). Weather forecasting and prediction using hybrid C5. 0 machine learning algorithm. International Journal of Communication Systems, 34(10), e4805
9. Shafin, A. A. (2019). Machine learning approach to forecast average weather temperature of Bangladesh. Global Journal of Computer Science and Technology
10. Singh, N., Chaturvedi, S., & Akhter, S. (2019, March). Weather forecasting using machine learning algorithms. In 2019 International Conference on Signal Processing and Communication (ICSC) (pp. 171-174). IEEE
11. Pielke, R.A., 1992. A comprehensive meteorological modeling system RAMS. Meteorology and Atmospheric Physics, Springer-Verlag, Vol.49, pp.69–91.
12. Vamsi Krishna, G., 2015. An integrated approach for weather forecasting based on data mining and forecasting analysis. International Journal of Computer Applications, 120(11), pp.26–29.
13. Ebi, K. L., Mearns, L. O., & Nyenzi, B. (2003). Weather and climate: changing human exposures. Climate Change and Health: Risks and Responses (McMichael AJ, Campbell-Lendrum DH, Corvalan CF, Ebi KL, Githeko A, et al., eds). Geneva: World Health Organization.
14. Weather or Climate. What's the Difference? (2019, October 22). www.Nationalgeographic.Org. Retrieved April 16, 2022, from https://bit.ly/3OsWZ90
15. What Is Climate Change? (2021, August 4). https://www.Un.Org/En/. Retrieved April 22, 2022, from https://www.un.org/en/climatechange/what-is-climate-change

16. Sharma, N., Sharma, P., Irwin, D., & Shenoy, P. (2011, October). Predicting solar generation from weather forecasts using machine learning. In *2011 IEEE international conference on smart grid communications (SmartGridComm)* (pp. 528-533). IEEE.
17. Hossain, M. S., Arshad, M., Qian, L., Kächele, H., Khan, I., Islam, M. D. I., & Mahboob, M. G. (2020). Climate change impacts on farmland value in Bangladesh. *Ecological indicators*, *112*, 106181.
18. Shahid, S. (2010). Recent trends in the climate of Bangladesh. *Climate Research*, *42*(3), 185-193.
19. Maqsood, I., Khan, M. R., & Abraham, A. (2004). An ensemble of neural networks for weather forecasting. Neural Computing & Applications, 13(2), 112-122.
20. Radhika, Y., & Shashi, M. (2009). Atmospheric temperature prediction using support vector machines. International journal of computer theory and engineering, 1(1), 55.
21. Chaudhuri, S., & Chattopadhyay, S. (2005). Neuro-computing based short range prediction of some meteorological parameters during the pre-monsoon season. Soft Computing, 9(5), 349-354.
22. Mekanik, F., & Imteaz, M. A. (2012). A multivariate artificial neural network approach for rainfall forecasting: case study of Victoria, Australia. In Proceedings of the world congress on engineering and computer Science (Vol. 1, pp. 557-561).
23. El Naqa, Issam & Murphy, Martin. (2015). What Is Machine Learning?. 10.1007/978-3-319-18305-3_1.
24. Alpaydin E. Introduction to machine learning. 3rd ed. Cambridge, MA: The MIT Press; 2014.
25. Harvard Wiki, (2022, June 20). MERRA-2,http://wiki.seas.harvard.edu/geos-chem/index.php/MERRA-2
26. NASA, GMAO, (2022, June 20). Modern-Era Retrospective analysis for Research and Applications, Version 2 . https://gmao.gsfc.nasa.gov/reanalysis/MERRA-2/
27. Naveen L, Mohan HS. Atmospheric weather prediction using various machine learning techniques: a survey, In: Proceedings of the Third International Conference on Computing Methodologies and Communication (ICCMC), IEEE; 2019:422-428. https://doi.org/10.1109/ICCMC.2019.8819643
28. National Weather Service, USA. (2022, June 20). https://www.weather.gov