Somalia Weather Forecast Using Machine Learning

Abdikarin Ibrahim Awale
Student Id: 012201029

A Project

in

The Department

of

Computer Science and Engineering

Presented in Partial Fulfillment of the Requirements

For the Degree of Master of Science in Computer Science and Engineering

United International University

Dhaka, Bangladesh

October 2021

Approval Certificate

This project titled **“Somalia Weather Forecast Using Machine Learning** submitted by **Abdikarin Ibrahim Awale”**, Student ID: **012201029** has been accepted as Satisfactory in fulfillment of the requirement for the degree of Master of Science in Computer Science and Engineering on 24 October 2021.

Board of Examiners

1.

____________________________ Supervisor

Prof. Dr. Mohammad Nurul Huda
Professor & Director-MSCSE
Department of Computer Science & Engineering (CSE)
United International University (UIU)
Dhaka-1212, Bangladesh

2.

____________________________ Examiner

Rubaiya Rahtin Khan
Assistant Professor, CSE
Department of Computer Science & Engineering (CSE)
United International University (UIU)
Dhaka-1212, Bangladesh

3.

____________________________ Ex-Officio

Prof. Dr. Mohammad Nurul Huda
Professor & Director-MSCSE
Department of Computer Science & Engineering (CSE)
United International University (UIU)
Dhaka-1212, Bangladesh

Declaration

This is to certify that the work entitled **“Somalia Weather Forecast Using Machine Learning”** is the outcome of the research carried out by me under the supervision of **Prof. Dr. Mohammad Nurul Huda**.

Abdikarin Ibrahim Awale
Student ID:012201029
MSCSE Program
Department of Computer Science and Engineering (CSE)
United International University (UIU)
Dhaka-1212, Bangladesh

In my capacity as supervisor of the candidate’s project, I certify that the above statements are true to the best of my knowledge.

Prof. Dr. Mohammad Nurul Huda
Professor & Director-MSCSE
Department of Computer Science & Engineering (CSE)
United International University (UIU)
Dhaka-1212, Bangladesh

Abstract

Generally, weather affects us in different ways. Sometimes, too much hot weather for a long time without any rain can cause drought, also huge storms with strong winds, such as tornadoes can do a great number of damages to the buildings, farms, etc. However, hot weather prevails all year in east Africa particularly Somalia. In Somalia there are different sports that take place every year. Weather conditions have a significant impact on outdoor sports such as track and field, football, and tennis. Those outdoor games occur in some cities and some in regions. For outdoor sports, weather conditions are an important aspect to consider while designing a game.

Machine learning represents a far bigger possibility in the growth of weather forecasting, so I try five (SVR, DT, RF, GB, MLP) models to test that. Those models/algorithms can predict the weather of three cities of Somalia such as Mogadishu, Kismayo, and Hargeisa. The model's reliability forecast was calculated between real and forecasted values by calculating the MAE, MSE and its accuracies. The results suggest that the Random Forest model might be beneficial for Somalia weather to forecast.

Acknowledgment

First and foremost, I want to thank Allah the almighty, most gracious and most merciful for the blessings he has bestowed upon me by completing this endeavor.

After that, I'd like to offer our heartfelt gratitude to **Prof. Dr. Mohammad Nurul Huda**, Director, Master of Science in Computer Science and Engineering (MSCSE) of United International University (UIU), who gently led me through the difficulties of this project with unmatched generosity. He also demonstrated unmistakable interest and considerable assistance in the field of Machine Learning for the improvement of our job work. Despite his hectic schedule managing corporate matters, he was always a great source of support and assistance. Also, I would like to extend our sincere esteems to examiner **Mrs. Rubaiya Rahtin Khan**, Assistant Professor for her timely support.

Finally, I'd like to thank my parents for their love and support, for which I would have never had so many possibilities without, and my friend's support.

Table of Contents

LIST OF TABLES

LIST OF FIGURES

Chapter 1

Introduction

Weather defines the condition of the atmosphere, such as how hot or cold it is, how wet or dry it is, how quiet or stormy it is, and how clear or overcast it is. Most weather events occur on Earth in the troposphere, the planet's lowest layer of atmosphere. Immediately beneath the stratosphere. Weather is the daily temperature, precipitation, and other atmospheric factors, whereas climate is the average of atmospheric conditions over longer periods. When used without a qualifier, the term "weather" is widely believed to refer to Earth's weather. Weather is everything that impacts our lives daily and can permanently alter people's lives, societies, economies, and environments. The amount to which weather affects our life is determined by a variety of elements such as the type of occurrence, timing, length, intensity, location, and so on [1]. Temperature, humidity, and wind speed are weather parameters [2].

Weather Forecasting is a space of current technology that predicts the atmospheric conditions for a selected location by collecting dynamic information associated with weather. There is a unit of varied approaches out there in foretelling, from comparatively easy observation of the sky to extremely complicated processed mathematical models. Among the foretelling, the prediction of temperature variation conditions is crucial for varied applications. A number of them are unit climate observance, drought detection, severe weather prediction, agriculture, and production, coming up within energy trade, aviation trade, communication, pollution diffusion, and so forth [3] [4]. The most important service given by the earth science profession is temperature forecasting, which helps protect people's lives and property while also increasing the efficiency of operations. It also helps people plan a wide range of everyday activities. As an example, though we can forecast tomorrow’s weather with a definite degree of accuracy, it's extremely difficult to forecast monthly temperature accurately even for professionally trained meteorologists. The information of current weather conditions is changed by varied foretelling tools like satellites, Balloons and craft, Buoys and land stations, and radiolocation Systems. The info collected from states area units distorted into Numerical illustration is understood as assimilation. Weather Predictions are an essential unit for varied applications like climate observance, drought detection, agriculture and production, energy trade, aviation trade, communication, pollution diffusion, etc. [3].

Due to Somalia's proximity to the equator, there is not a great deal of seasonal variation in its climate. But some unexpected precipitation generally occurs. Hot weather prevails all year, accompanied by monsoon (periodic seasonal reversal wind in the middle of matching variations in precipitation) winds and infrequent rain. Mean daily most temperatures vary from thirty to forty °C (86 to 104 °F), except at higher elevations and on the Nipponese shore, where the results of a cold offshore current are going to be felt. In Mogadishu, for instance, average afternoon highs vary from twenty-eight °C (82 °F) to thirty 2 °C (90 °F) in April. A variety of the simplest mean annual temperatures at intervals the globe is recorded at intervals the country; Berbera on the northwestern coast contains a day high that averages quite thirty-eight °C (100 °F) from new style calendar month through Sept. Nationally, mean daily minimums generally vary from relating to fifteen to thirty °C (59 to eighty-six °F). The simplest direct climate happens in a northern African country, where temperatures typically surpass 45 °C (113 °F) in new style calendar month on the littoral plains and drop below the temperature throughout December at intervals in the highlands. throughout this region, magnitude relation ranges from relating to forty percent at intervals the mid-afternoon to eighty 5% at the hours of darkness, dynamic somewhat in step with the season [4].

1.1 Problem Statement

Any game requires strategy before it can be played. Scheduling the game, constructing surroundings, collecting equipment, giving transport facilities to participants, selecting venues, offering lodgings to players, and so on are all examples of planning. These are the items that must be controlled. For outdoor sports, weather conditions are an important aspect to consider while designing a game. It has a significant impact on outdoor sports such as track and field, football, and tennis. When arranging outdoor sports, the weather conditions are used to determine whether or not to play [5]. Hot weather prevails all year in east Africa particularly Somalia [4]. In Somalia there are different sports that take place every year. Due to adverse weather, the summer and winter seasons make it difficult to play sports, particularly track and field, football, tennis, and golf. Those outdoor games occur in some cities and some in regions. Inclement weather, such as too much hot, rain and wind, makes deciding how to approach shots while golfing, playing tennis and football [6]. Extreme heat or hot, fog, rain, or wind can all detract from the enjoyment of outdoor activities. To solve the problems that are disrupting the outdoor games I need to predict the weather using machine learning techniques.

1.2 Objectives

My goal is to anticipate weather effects on outdoor games in Somalia, focusing on three major cities (Mogadishu, Kismayo, Hargeisa).

- I'm going to predict three weather parameters using the regression approach.
- I will try to experiment with several algorithms to get an accurate result.
- The sports decision-makers make decisions based on reports they have received from the weather forecasted.

Chapter 2

Background and Literature Review

In many economic activities, including athletic events and the sector as a whole, weather and weather conditions are key elements. Significant organizational and financial issues occur as a result of unexpected weather disturbances that cannot always be precisely foreseen. Even though climatic kinds do not inherently affect tourist activities, they are typically served as significant elements of both cash inflows predicted by travel operators and individual visitor experiences [7]. Machine learning approaches have been used to predict weather forecasting effect for many years, but some of those techniques have been used to anticipate how weather affects According to outdoor sports, the interface between outdoor sports activities and weather conditions is more complicated in the sense that we must incorporate meteorological elements that we cannot precisely anticipate or rigorously regulate. Because various sporting activities are tolerant to distinct ranges of each weather variable, the weather effect on outdoor sports may be broken down into: Wetter sports are divided into three different types: (1) Weather sport specialists, (2) sport weather and (3) sports for the benefit of weather (1) (including baseball). According to the same author's rating the dependence on the weather conditions of outdoor and indoor sport events could also be divided by three types: those that are based on specific weather conditions (e.g., skiing or sailing); those which could have unfair advantages in weather circumstances (e.g., golf, athletics) [8] [9].

For improved categorization results, [10] presented suitable classification approaches such as Support Vector Machine (SVM) and neural networks. These approaches will aid in the prediction of rainfall, agricultural production forecasting, and crop cost prediction. In this study, a unique prediction model for the optimal cereal crop cultivation time was presented, which strongly connected crop life cycle to the accumulation of specified quantities of heat, measured as thermal time or GDD. Based on past meteorological data, it used several ML regression techniques to forecast the daily minimum and maximum air temperature [11]. Examined machine learning methods for weather prediction using data mining approaches for rainfall prediction. The FP Growth Algorithm was used to build decision trees and rules for classifying meteorological characteristics such as maximum temperature, lowest temperature, rainfall, humidity, and wind speed in terms of the month and year [1]. suggested weather prediction utilizing a big data environment, and their approach is Hadoop with map reductions to analyze sensor data held in the National Climatic Center and deliver improved results [12].

Using K-medoids with the Naive Bayes algorithm for a weather forecasting system that includes factors such as temperature, humidity, and wind. This technology may be utilized in areas such as air traffic control, military and navy operations, marine operations, agriculture, and forestry, among others [11]. The decision tree algorithm was proposed [13]. This algorithm was used to build decision trees and rules for categorizing meteorological characteristics such as maximum and minimum temperatures, among other things. Utilizes different statistical models to estimate the hourly PV electricity output for the next day at certain power facilities in France, with the random forest outperforming the others. Random Forests (RF) is one of the most widely used data mining algorithms. The technique is widely utilized in a variety of time series forecasting disciplines, including biostatistics, climate monitoring, energy sector planning, and weather forecasting. Random forest (RF) is an ensemble learning method capable of handling both high dimension classification and regression [14] [10]. In contrast to previous regression models, the Gradient boosting technique was presented as the best outcome to anticipate both low and big wind speed levels and was effective in managing high wind speed values. The performance of gradient boosting in comparison to other models is observed to have the best performance [15].

2.1 Three Main Somalia Cities

Hargeisa, positioned at 1.300 m north-west (4.25 m), may get frigid during the winter night; from November to March lowest temperatures can drops under 5°C (41°F). Daytime temperatures of around 30°C (86°F) are convenient in winter and bearable in summer [16].

Figure 0.1 Hargeisa located northern of Somalia

Hargeisa's climate is semi-arid. Warm winters and scorching summers are typical in the city. Despite its position in the tropics, Hargeisa seldom suffers extreme hot or cold weather due to its high elevation. This is a characteristic that is uncommon in semi-arid environments. The city receives the majority of its precipitation between April and September, with an annual rainfall of slightly under 400 millimeters (16 in). Hargeisa's average monthly temperatures range from 18 °C (64 °F) in December and January to 24 °C (75 °F) in June [17].

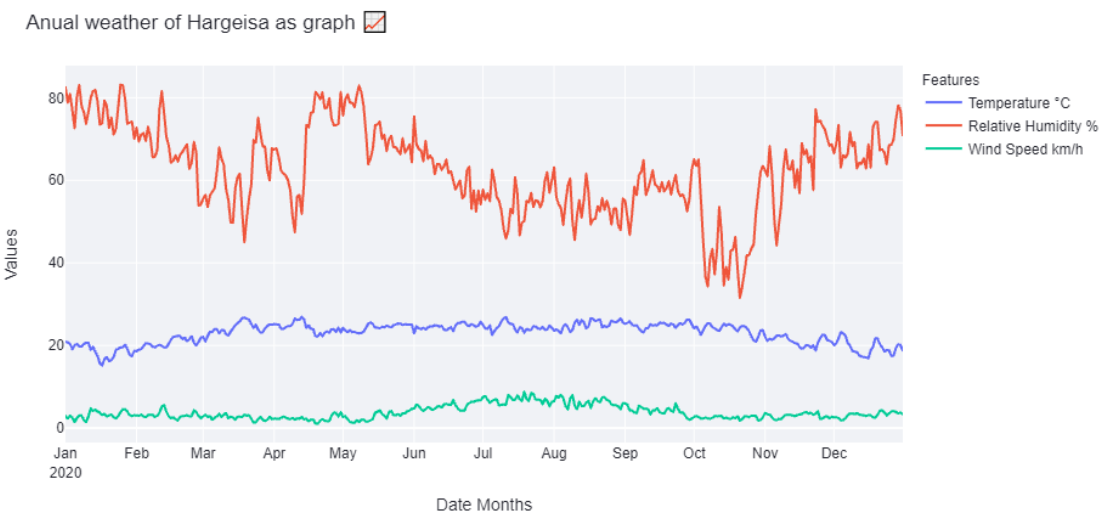


Figure 0.2: Last year weather of Hargeisa weather

Mogadishu lies on the coast of the Horn of Africa on the south shore of Banaadir, Somalia. The territory covers the city and is somewhat smaller than the historical province of Benadir. Bondoh, Hamar-jajab, Hamar-Wadag, Hamara-Weyne, Wardhigley, Heliwa, Hodan, Shangani, Howl-Wadag and Karan as well as Abdiaziz are all districts of the town's administrative areas. The features of the city are the old city of Hamar-Weyne, Bakaara Market and the beach of Gezira. The sandy beaches of Mogadishu have flourished coral reefs, making them the first tourist destinations of the city in many years. The Shebelle River, which comes from central Ethiopia, flows south-west towards Mogadishu, near the Indian Ocean, before reverting to Mogadishu. The river, normally dry in February and March, provides water for sugarcane, cotton and bananas to flourish [18].

Figure 0.3: Mogadishu is the capital of Somalia, which is located in the southern

Mogadishu has a surprisingly dry climate for a city so close to the equator. As with much of southern Somalia, it is classed as hot and semi-arid. Towns in northern Somalia, on the other hand, have a hot, dry environment. While the city is generally dry, relative humidity is very high due to its seaside position, averaging 79 percent for the year.

Mogadishu sits in or close to the tropical thorn woodland biome of the Holdridge global bioclimatic regime. The year-round average temperature in the city is 27°C and the highest temperature is 30°C (86°F) and the minimum temperature is 24°C (75°F). The middle monthly temperature fluctuates by 3°C (5.4°F), which corresponds to a hyper-oceanic type and the hyper-oceanic type. The yearly rainfall totaled 429,2 mm (16.9 in). Every year there are 47 rainy days with a daily chance of precipitation of 12 percent each year. The average sunshine in the city is 3 066 hours a year, and the average sunlight is

8.4 hours a day. The average daylight time is 8 hours and 24 minutes each day. The potential sunshine percentage is 70%. On the 21st of the month, average solar height at noon is 75% [19].

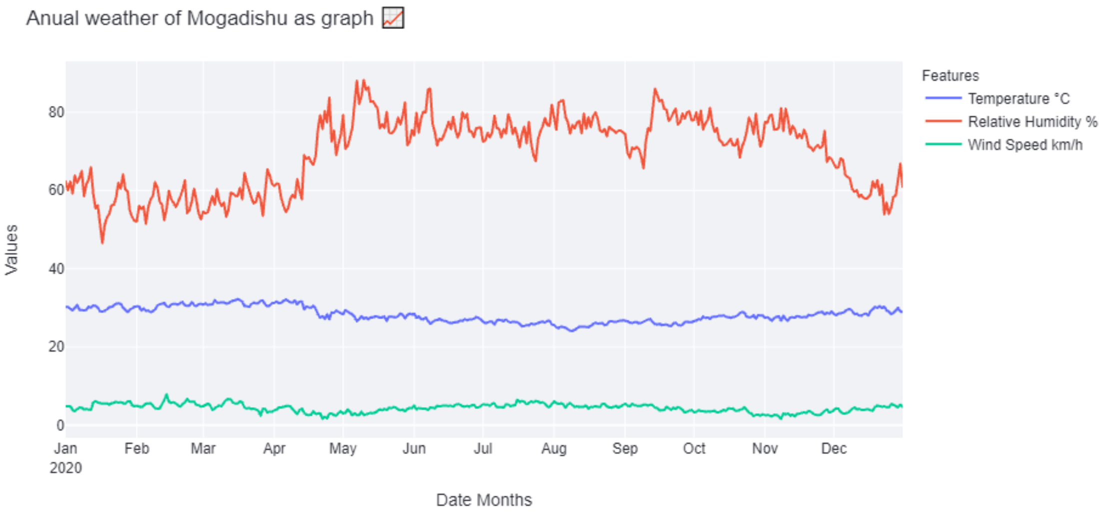


Figure 0.4: Last year weather of Mogadishu weather

In southern Somalia's rich Juba Valley, Kismayo is located on the Somali Sea coast. Qandal is to the south (6.5 Nm), Qeyla Dheere to the north-west (6.4 Nm) and Xamarey (5.0 Nm), Iak Bulle (10.0 Nm) to the south-west, Saamogo to the west (0.9 Nm). The most important towns of Kismayo are Jamaame (52 Km), Jilib (97 Km) and Merca (337 km) [20].

Figure 0.5: Kismayo is situated in southern Somalia's lush Juba valley

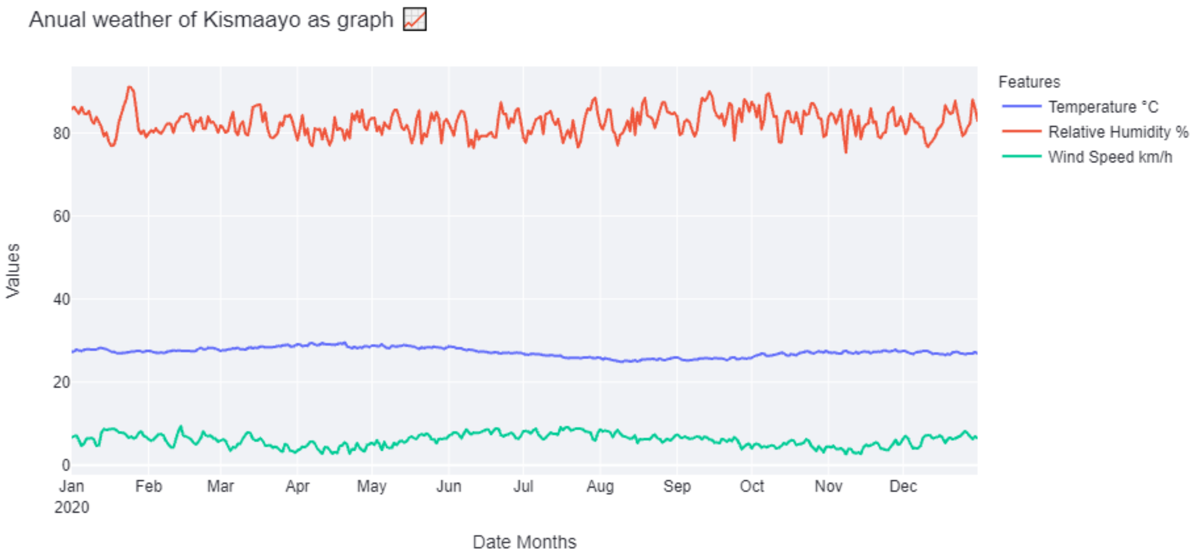


Figure 0.6: Last year weather of Kismayo weather

The climate in Kismayo is semi-hot and hot. The weather is warm throughout the year, with seasonal monsoon winds and occasional precipitation and periodic droughts. The rainfall, known as Southwest Monsoons, begins in April and goes through the whole of July. It gives rise to plenty of rain and lush plants. The dry season of xagaa is followed by the gu (hagaa) [2].

2.2 Machine Learning

Machine learning is a kind of artificial intelligence (AI) that allows computers to learn and develop without being explicitly programmed. The construction of computer programs that can access data and learn on their own is what machine learning is all about.

The learning process begins with observations or data, such as examples, direct experience, or teaching, so that we can look for patterns in data and make better decisions in the future based on our examples. The main goal is for computers to be able to learn on their own, without the need for human intervention, and to adapt their behavior accordingly.

However, text is treated as a series of keywords when using traditional machine learning algorithms; instead, a semantic analysis technique mimics the human ability to comprehend the meaning of a document.

2.2.1 Methods of Machine Learning

There are two types of machine learning algorithms: supervised and unsupervised.

- **Machine learning algorithms that are supervised** can use labeled examples to apply what they've learned in the past to fresh data and anticipate future events. The learning method generates an inferred function to make predictions about output values based on the study of a known training dataset. Following adequate training, the system may provide objectives for any new input. The learning algorithm can also compare its output to the correct, intended output and detect faults, allowing the model to be adjusted as needed.

- **Unsupervised machine learning** techniques, on the other hand, are utilized if the data used for training is neither categorized nor labeled. Unattended learning explores how systems might deduce a function from unmarked input to characterize the underlying structure. The system does not find the appropriate result, but it examines the data and can make inferences from data sets to characterize unlabeled structures.

- **Semi-supervised machine learning** techniques lie between supervised and unsupervised learning, as both labeled and unlabeled data for training – often a small quantity of labeled data and a big quantity of unlabeled data. The systems that apply this strategy can significantly increase the accuracy of learning. Semi-supervised education is usually used if the collected labeled data needs expert and appropriate resources to be trained and learned from it. Otherwise, it does not often take more efforts to get unstated data.
- **Reinforcement machine learning** is a learning strategy that works via actions and identifies faults or rewards in connection with the environment. The most important elements of reinforcement learning consist of trial-and-error search and delayed reward. In order to maximize their performance, this technology allows machines and software agents to autonomously decide on the best behavior in a certain scenario. For an agent to learn what action is optimal; this is known as the strengthening signal, simple reward feedback is needed.

2.3 Algorithms of Machine Learning

Because of the time series, there are various algorithms appropriate for weather forecasting in machine learning supervised methods, including regression and certain neural networks such as decision tree, support vector machine, random forest, gradient boosting, and MLP regressor.

2.3.1 Support Vector Regression

It supports the same notion as SVMs for vector regression. The fundamental principle of SVR is to find the highest fit. The hyperplane with the highest number of points is the best fit line in SVR. Unlike other regression models, the SVR tries to fit the best line into a given range of values to minimize the gap between the real and forecasted values. The threshold value is the distance between the hyperplane and the frontier line. The time complexity of SVR is more than quadratic with the sample number, which makes it more than tens of thousands of samples difficult to scale up to the datasets. For large datasets, a linear SVR or SGD regressor is employed. Linear SVR is a faster execution than SVR, but it takes only the linear kernel into consideration. Since the cost function misses samples whose forecast is close to its objective, the model generated by Support Vector Regression depends only on a fraction of the training data [21]. There is a model in the Sklearn package that can do linear regression on multiple outputs.

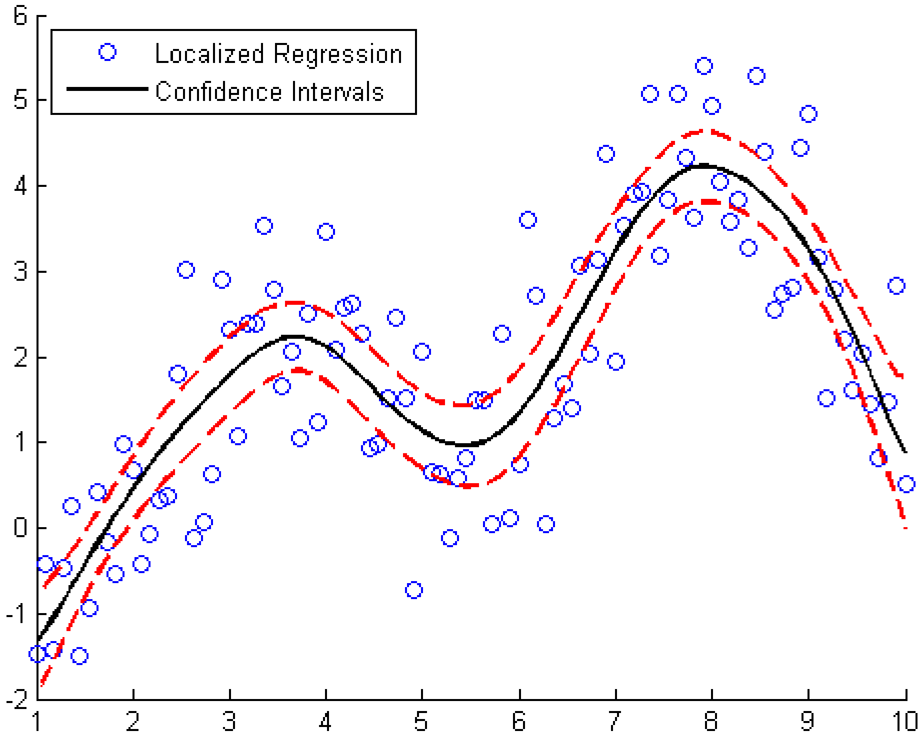


Figure 0.7: Support Vector Machine Regression

2.3.2 Decision Tree Regression

The Decision Tree is one of the most used and useful approaches for supervised learning (DT). It may be used to address both problems of regression and classification, the latter being more practical. It has three different node types and is a tree structured classifier. The Root Node is the first node representing the complete sample and split into nodes later. Inner nodes are properties of data sets, whereas branches are decision rules. Finally, the result is the Leaf Nodes. This is a very useful approach for handling policy concerns [22]. There is a model in the sklearn package that is capable of a decision tree regressor for multioutput.

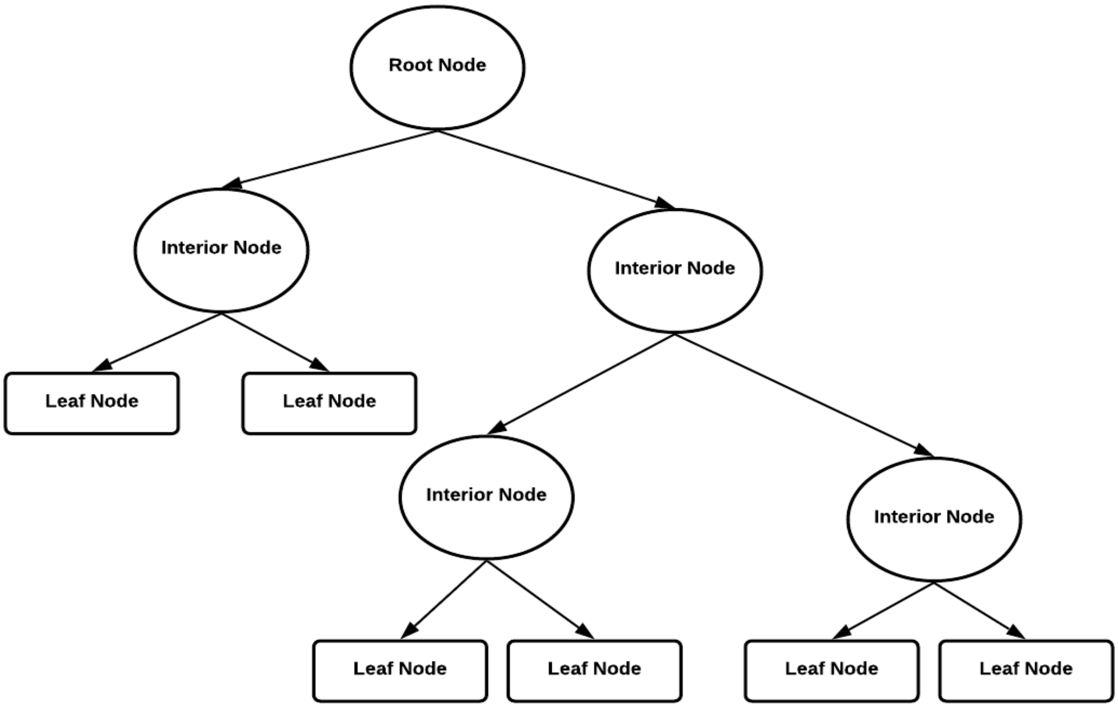


Figure 0.8: Decision Tree Regression

2.3.3 Random Forest Regression

Due to the easy understanding and understanding of the approach of the Decision Tree, a single tree might not be adequate to teach the model about the features. On the other hand, Random Forest is an algorithm based on the Tree that takes decisions by mixing the qualities of the many Decision Trees. This might thus be defined as a "forest of trees," which is why it is known as the "random forest." The moniker 'random' is because this approach is a forest of 'random Decision Trees.' There is a major disadvantage in the decision tree technique, which results in over-supply. The Random Forest Regression rather than the Decision Tree Regression can alleviate this issue. More than traditional regression models, Random Forest technology is considerably quicker and more robust [23]. There is a model in the sklearn package that is capable of random forest tree regressor for multi output.

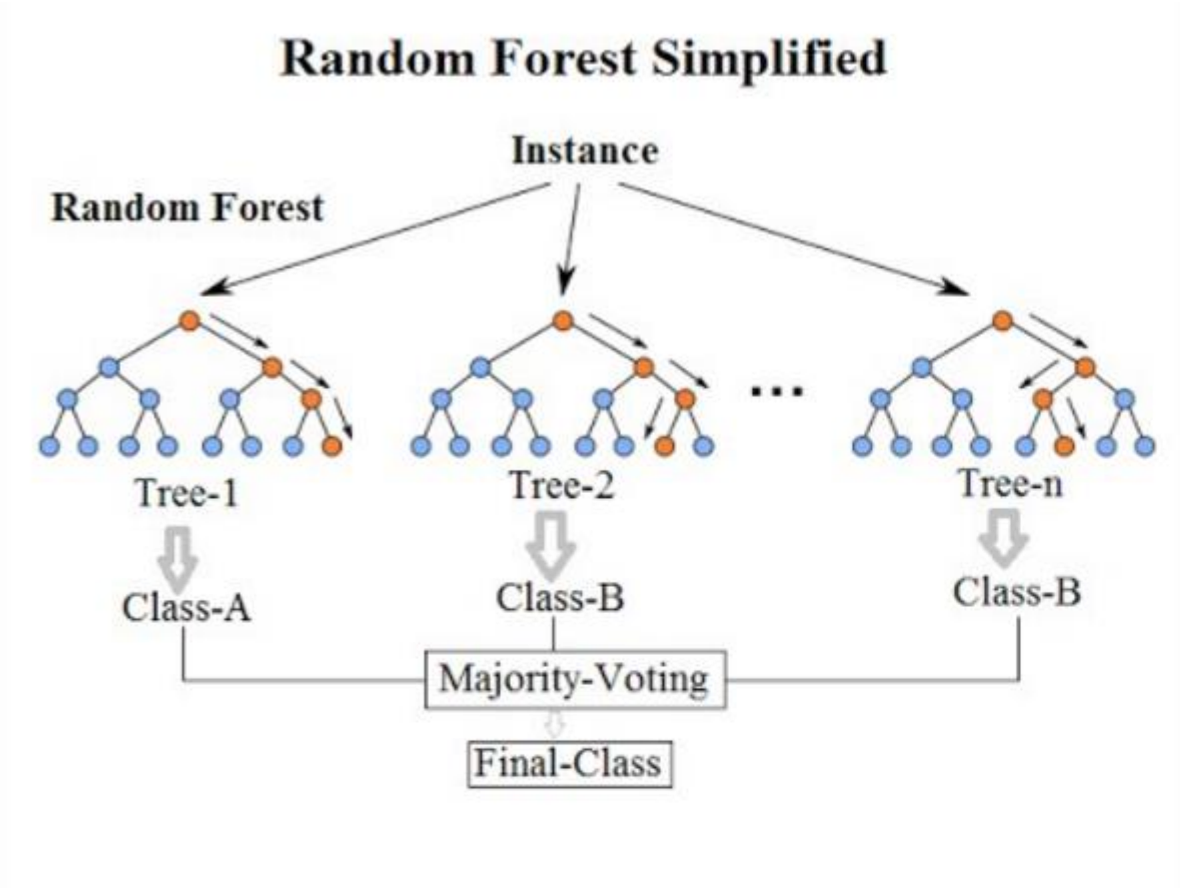


Figure 0.9: Random Forest Tree

2.3.4 MLP Regression

A regression MLP (multilateral perceptron) is a neural network that connects several layers of the diagram, suggesting that transmission of signals between nodes is only one direction. Apart from the input nodes, each node is non-linear. Backpropagation is an MLP-controlled technique of learning. A MLP is used as a supervised learning approach through back propagation. MLP is often used in supervised apprenticeships and neuroscience in the computer sciences and in parallel computer research. Spoken words, picture recognition and machine translation are among the applications [24].

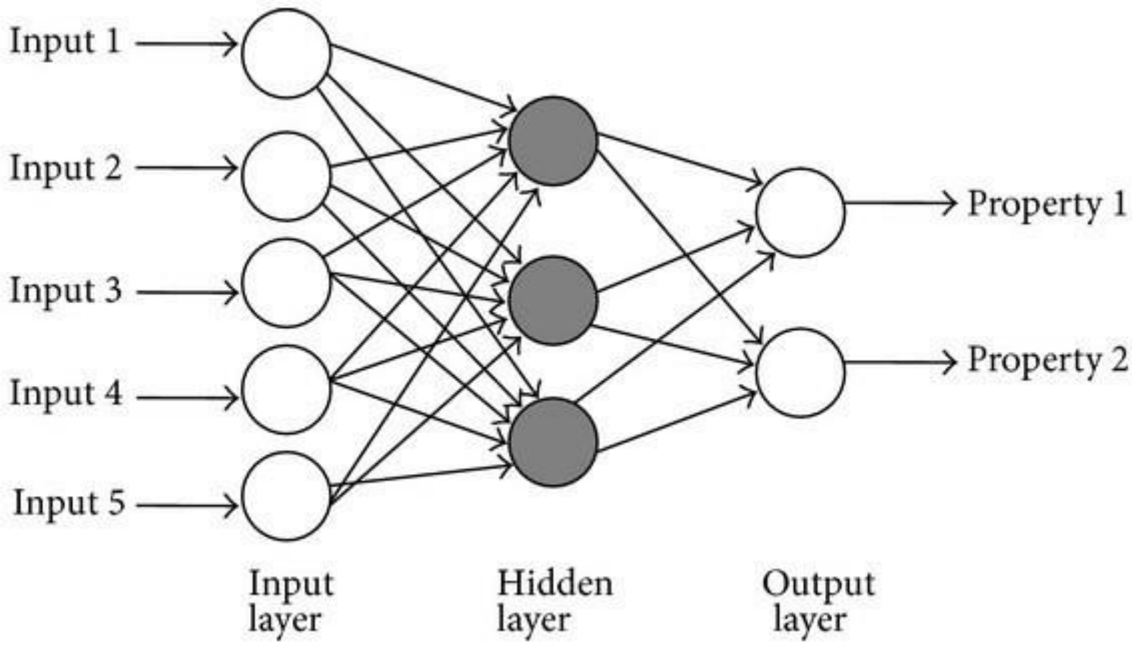


Figure 0.10: Multilayer Perceptron Regression

2.3.5 Gradient Boosting

Gradient-boosting regression tree methods include a whole learning strategy in which strong prediction models are created by the integration of multiple separate regression trees, known as weak learners. This reduces misleading models. Such an algorithm (regressors or classifiers). Weakly trained models consist of those whose outputs are little improved compared with random assumptions that show a significant bias towards the training dataset, low variance and regularization. The boost algorithms usually comprise three parts, namely a model additive, weak learners and a loss function. The algorithm may be used to simulate non-linear interactions such as the wind curve, and iterations between input features may include various differentiable loss functions. GBM (Gradient Boosting Machines) works by detecting weak models' limits using gradients. This may be achieved through an iterative strategy where the objective is to eventually unite basic learners in decreasing predictive mistakes where decision trees are coupled with an additive model and reduce the loss function by descending gradients.

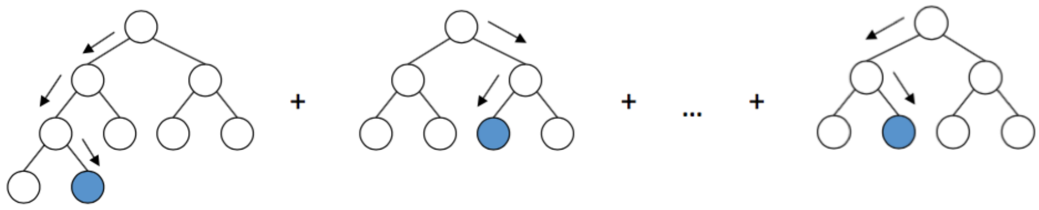


Figure 0.11: Gradient Boosting

Chapter 3

System Analysis and Design

To compare and contrast various machine learning techniques. I gathered information from various Somali cities. Temperature (2 meters) above the earth's surface, Wind speed (2 meters) above the earth's surface, and Relative humidity (2 meters) above the earth's surface are all included in these datasets. This dataset contains 1982-2020.

3.1 Dataset

There are several forms of weather, such as temperature, relative humidity, wind, precipitation, and pressure. The project's dataset includes four features: city, date, and three weather characteristics (temperature, relative humidity, wind speed).

3.1.2 Collecting Dataset

To gather weather data from data access viewer (DAV) - Nasa power, which delivers solar and meteorological data sets from Nasa research for renewable energy, building energy efficiency, and agricultural purposes. It provides annual, daily, and hourly statistics from 1981 to the present. It will take you a few minutes to choose a city as latitude and longitude, date, and a few weather characteristics, after which you may download as CSV, JSON format, and so on. Each city has 8 columns and 14262 rows in the dataset.

Table 0.1: Hargeisa dataset from data access viewer

LAT	LON	YEAR	MO	DY	T2M	RH2M	WS2M
9.56241	44.07701	1982	1	1	20.61	63	2.31
9.56241	44.07701	1982	1	2	20.94	65.81	2.66
9.56241	44.07701	1982	1	3	20.76	67.5	2.67
9.56241	44.07701	1982	1	4	20.38	66.88	2.25
9.56241	44.07701	1982	1	5	20.33	62.12	2.22
9.56241	44.07701	1982	1	6	20.3	62.62	2.2
9.56241	44.07701	1982	1	7	19.55	68.06	2.34
9.56241	44.07701	1982	1	8	19.54	65.12	1.54
9.56241	44.07701	1982	1	9	20.11	65.31	2.16

9.56241	44.07701	1982	1	10	20.12	68.19	2.38

Table 0.2: Mogadishu dataset from data access viewer

LAT	LON	YEAR	MO	DY	T2M	RH2M	WS2M
2.04691	45.31821	1982	1	1	26.57	81.81	6.1
2.04691	45.31821	1982	1	2	26.54	81.19	6.31
2.04691	45.31821	1982	1	3	26.4	81.5	6.75
2.04691	45.31821	1982	1	4	26.44	79.44	6.99
2.04691	45.31821	1982	1	5	26.3	78.94	7.01
2.04691	45.31821	1982	1	6	26.19	81	6.3
2.04691	45.31821	1982	1	7	26.69	80.81	6.54
2.04691	45.31821	1982	1	8	26.67	84.31	6.13
2.04691	45.31821	1982	1	9	26.3	86.81	5.28
2.04691	45.31821	1982	1	10	26.14	86.06	5.02

Table 0.3: Kismayo dataset from data access viewer

LAT	LON	YEAR	MO	DY	T2M	RH2M	WS2M
0.35601	42.54611	1982	1	1	29.07	64.75	3.86
0.35601	42.54611	1982	1	2	29.12	64.38	3.97
0.35601	42.54611	1982	1	3	28.83	65.69	4.08
0.35601	42.54611	1982	1	4	29.04	64	4.4
0.35601	42.54611	1982	1	5	29.29	61.44	4.38
0.35601	42.54611	1982	1	6	28.51	64.5	3.91
0.35601	42.54611	1982	1	7	29.15	64.69	3.92
0.35601	42.54611	1982	1	8	27.8	74.44	3.58
0.35601	42.54611	1982	1	9	27.19	76.81	3.1

3.2 Proposed System Diagrams

3.2.1 Software Architecture Diagram

The software architecture diagram for weather forecasts is a system diagram used to abstract the overall software system framework and the component connections, limitations and restrictions.

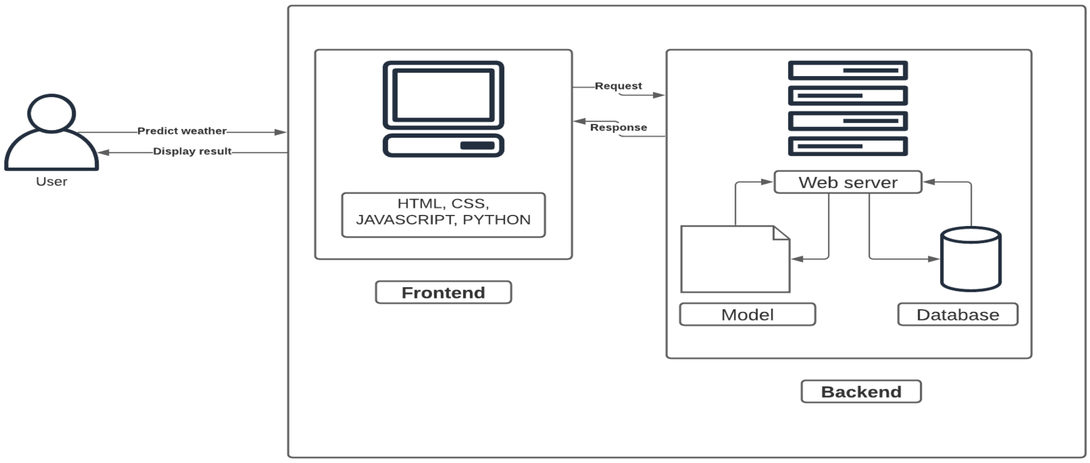


Figure 0.1: Software Architecture Diagram

3.3 Unified Modeling Language (UML)

3.3.1. Class Diagram

The primary building element of an object-oriented solution is the weather forecasts class schemes. It displays the classes of each class in one system, their properties and functions and their connections.

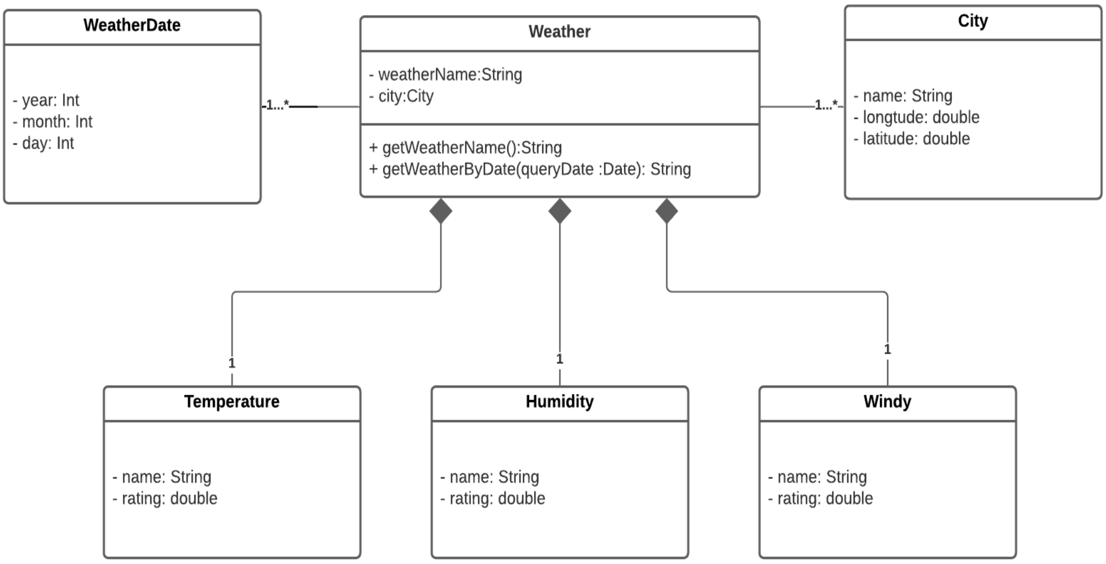


Figure 0.2: Class Diagram

3.3.2 Use Case Diagram

The weather forecast of use case diagram provide a graphical picture of the participating actors in a system that can able to select different types and then predict the weather, the different functions that these actors demand and their interactions.

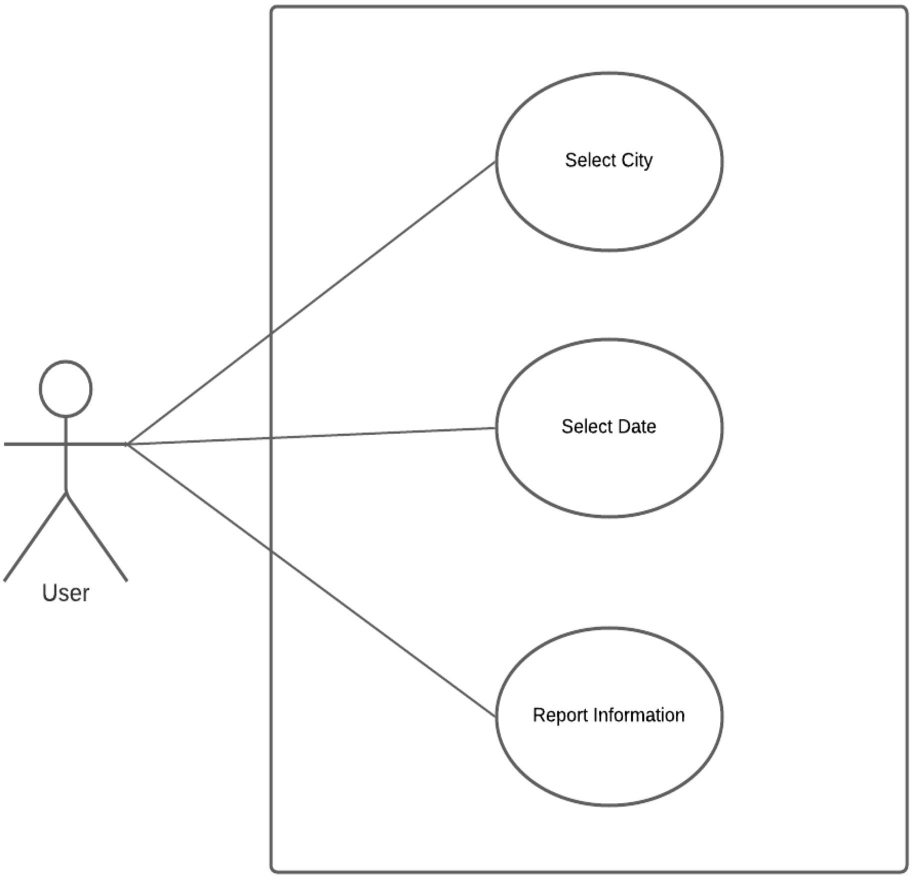


Figure 0.3: Use Case Diagram

3.3.3. Sequence Diagram

The UML sequence diagram weather forecast shows how items interact and how these interactions occur. It is crucial to notice that the interactions for a certain situation are shown. The processes are vertically depicted and the interactions as arrows are displayed.

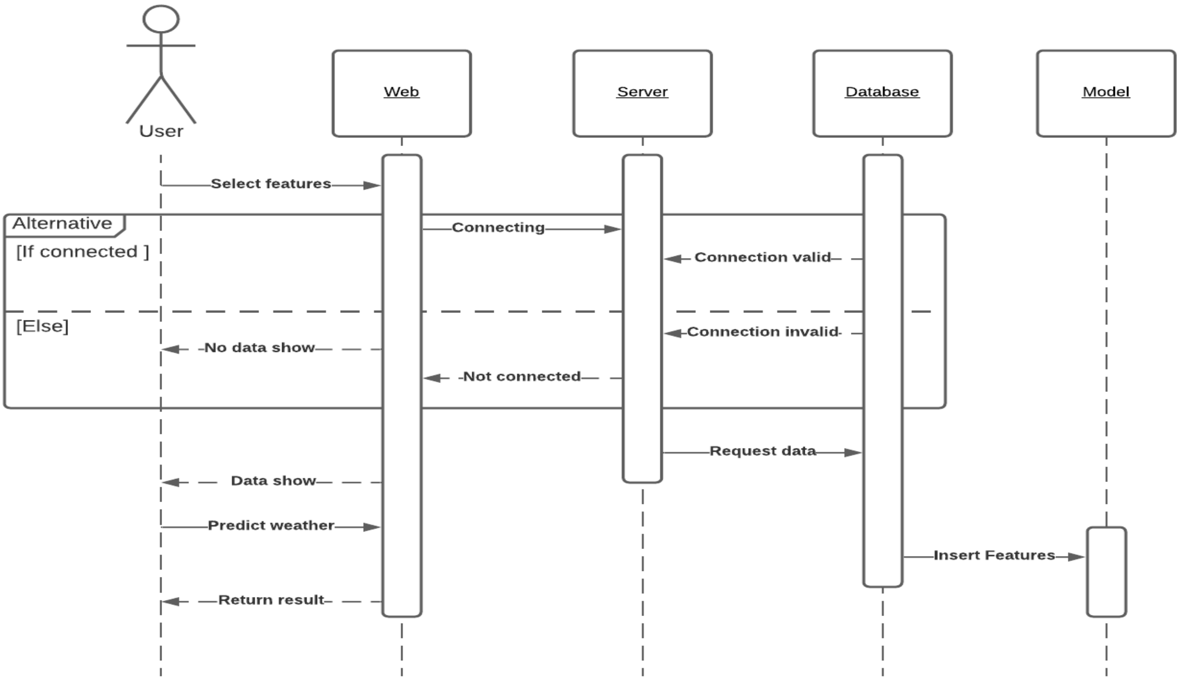


Figure 0.4: Sequence Diagram

3.4 Data Flow Diagram

The Data Flow Diagram (DFD) weather prediction represents the data flow in the system graphically. As it can show incoming data stream, exiting data stream and store data.

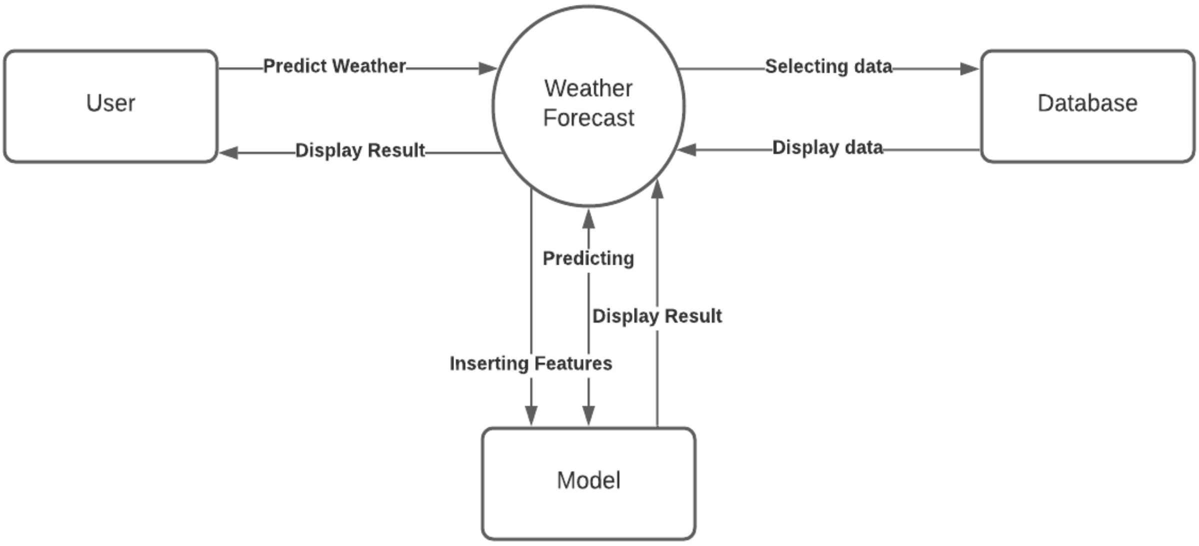


Figure 0.5: Data Flow Diagram

3.5 Waterfall Model

The waterfall model's weather prediction is a sequential model, dividing software development into predefined phases. Before the following step may begin, each phase must be completed without overlapping.

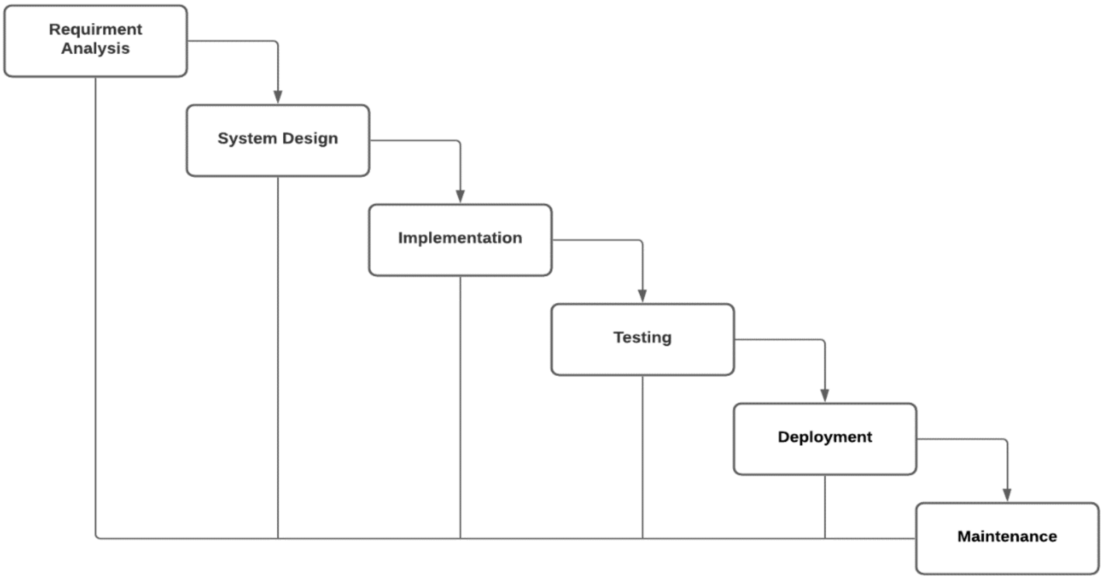


Figure 0.6: Waterfall Model Design

3.4 Hardware/Software Specifications

Table 0.4: Hardware and Software Specifications

Operating System for Computer	**Windows 10**
Browser	Google Chrome
Code Editor	Visual Studio Code (VS Code), PyTorch
Frontend	HTML, CSS, Streamlit, Plotly
Backend	Python, Streamlit, MySQL

3.5 User Interface Design

Using Streamlit, I designed and developed the UI. Streamlit is an open-source Python framework for building machine learning and data science web applications. Streamlit lets you construct an application in the same way Python code is written. Streamlit makes it simpler to work on the interactive cycle of code and monitor results in the web app.

3.5.1 Prediction Form

The first section of the UI comprises a prediction form. When the user clicks the predict button, he or she may select the city name and date that he or she wants to forecast on the right side of the picture that appears.

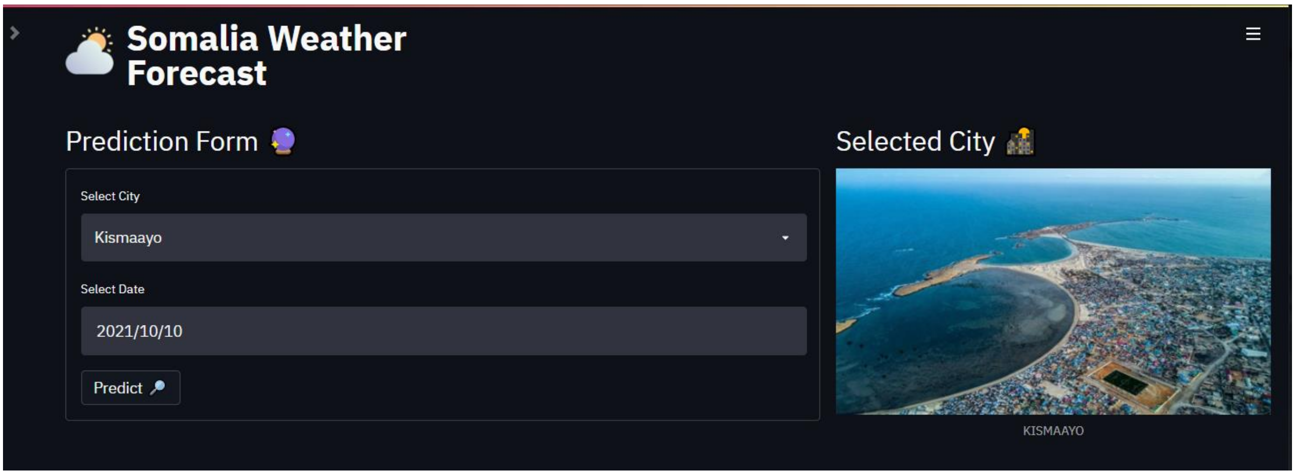
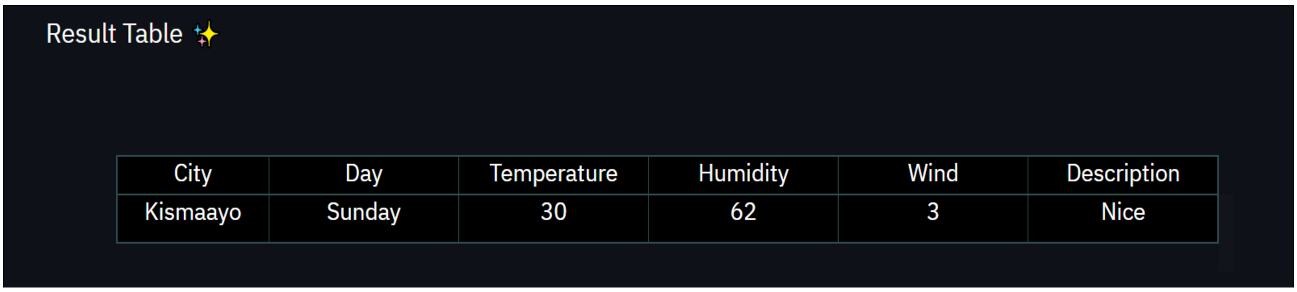


Figure 0.7: Prediction Form of Users

3.5.2 Result Table

This is the result table, which shows city name, day name, temperature, humidity, wind, and also weather condition so, the user can able see clearly the result of he/she prediction.


Result Table

City	Day	Temperature	Humidity	Wind	Description
Kismaayo	Sunday	30	62	3	Nice

Figure 0.8: Result Table of Users

3.5.3 Report

After the result table, there is a report that concludes the prediction table and gives the user report or decision-makers.

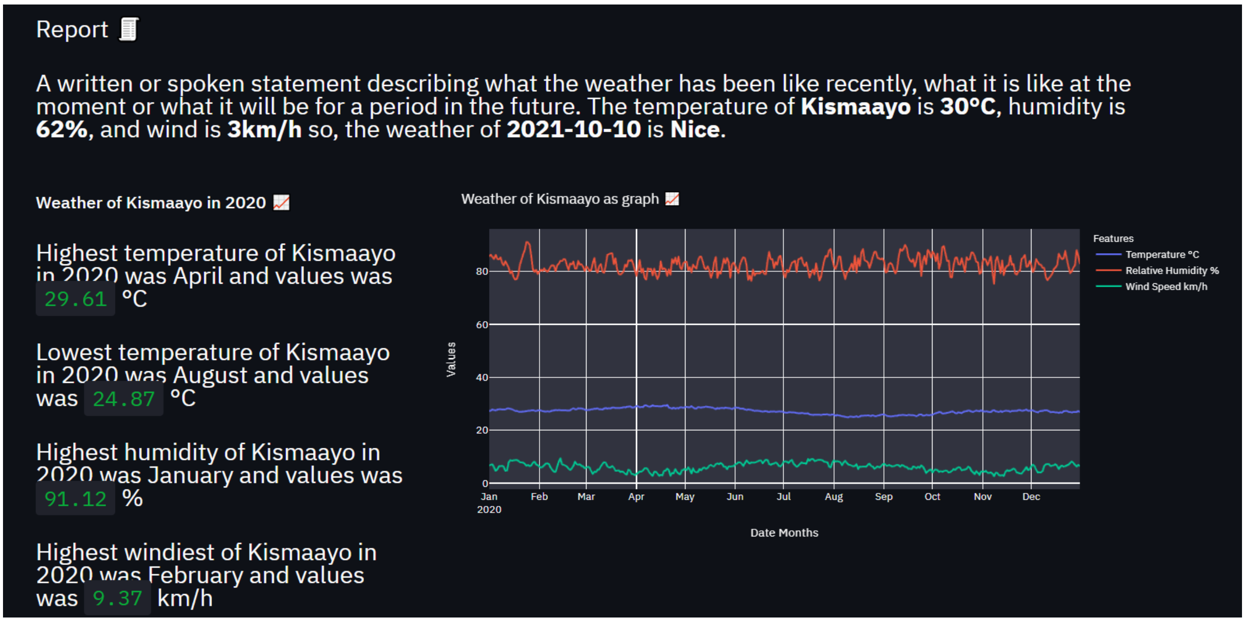


Figure 0.9: Report for the user that concludes for the prediction

3.5.4 Annual Weather Conclusion

Also, the user can ably see the annual weather of the last year which gives useful information about last year. That he/she can ably see last year the hottest month, or highest humidity also which month who have the highest wind.

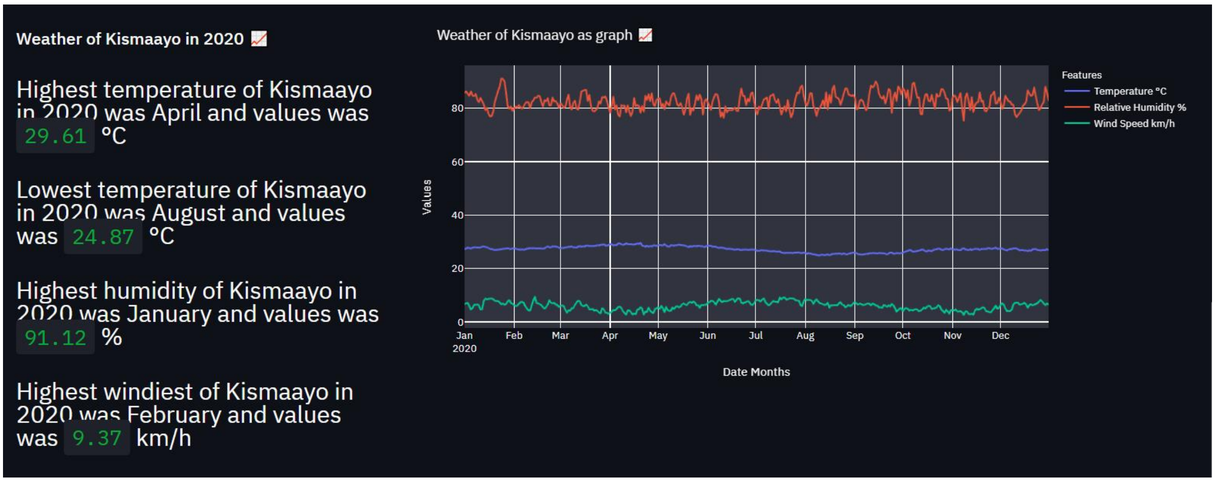


Figure 0.10: Report for Mogadishu of Annual weather from last year also graph

Chapter 4

Results and Output

4.1 Models

I tried four machine learning regression models (Algorithms) suited for weather forecasting. The project's implementation outcomes are shown below.

4.1.1 Support Vector Regression

This SVR model has shown to be the least accurate. A glimpse of the real result of the SVR project implementation is shown below. The accuracy of the SVR evaluated r2 score is 0.72, MAE is 2.15 which is the highest error compare with other models also MSE is 17.06 highest error.

Table 0.1 SVR for comparing actual value and predicted value

	Actual			Predicted			Difference		
No	**Temperature**	**Humidity**	**Wind**	**Temperature**	**Humidity**	**Wind**	**Temperature**	**Humidity**	**Wind**
1	23.26	70.19	4.35	26.50	68.60	4.17	-3.24	1.59	0.18
2	24.73	49.5	6.74	26.38	68.57	4.17	-1.65	-19.07	2.57
3	26.42	79.62	5.05	26.61	68.65	4.18	-0.19	10.97	0.87
4	26.53	82.06	6.12	26.58	68.60	4.20	-0.05	13.46	1.92
5	25.91	75.56	7.19	26.36	68.55	4.19	-0.45	7.01	3.00
6	27.98	61.5	4.16	26.62	68.62	4.16	1.36	-7.12	0.00
7	29.94	60.62	3.97	26.53	68.57	4.18	3.41	-7.95	-0.21
8	20.26	59.06	2.02	26.31	68.58	4.15	-6.05	-9.52	-2.13
9	27.45	60.94	4.89	26.58	68.61	4.16	0.87	-7.67	0.73
10	29.05	63.75	3.58	26.44	68.59	4.15	2.61	-4.84	-0.57

A line plot compares the actual values and projected values of the first ten rows of the data set weather forecast. This provides a geometric explanation of the performance of the model in the test set.

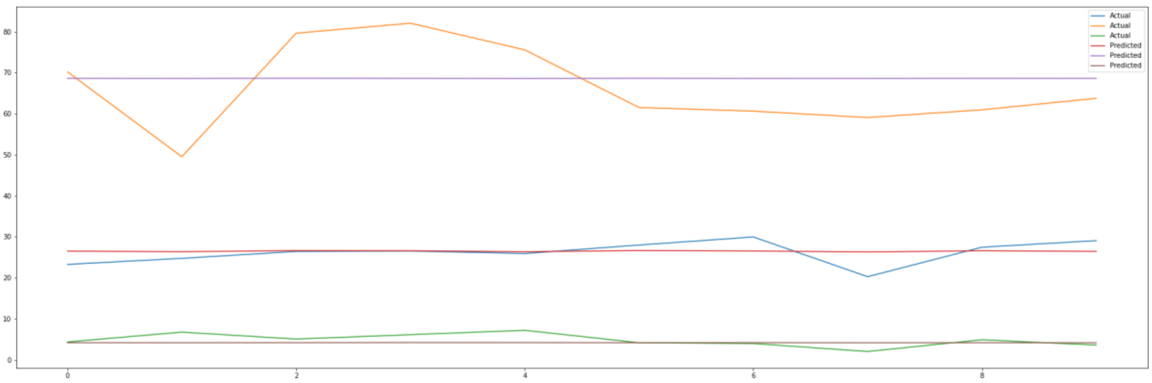


Figure 0.1: Plot showing comparison of actual value and predicted value in SVR

4.1.2 Decision Tree

This DT is a conceivable accurate model. Given below is a snapshot of the actual result from the project implementation of the DT tree. The accuracy r2 score is 0.88, MAE is 1.32 which is less error compare with other models also MSE is 6.87 less error.

Table 0.2: Decision tree for comparing actual value and predicted value

	Actual			Predicted			Difference		
No	**Temperature**	**Humidity**	**Wind**	**Temperature**	**Humidity**	**Wind**	**Temperature**	**Humidity**	**Wind**
1	23.26	70.19	4.35	22.40	67.50	6.95	0.86	2.69	-2.60
2	24.73	49.50	6.74	23.88	51.31	6.93	0.85	-1.81	-0.19
3	26.42	79.62	5.05	26.61	78.06	4.73	-0.19	1.56	0.32
4	26.53	82.06	6.12	26.28	81.88	6.45	0.25	0.18	-0.33
5	25.91	75.56	7.19	26.81	83.31	7.07	-0.90	-7.75	0.12
6	27.98	61.50	4.16	28.33	58.06	4.12	-0.35	3.44	0.04
7	29.94	60.62	3.97	30.05	57.81	3.86	-0.11	2.81	0.11
8	20.26	59.06	2.02	19.65	60.81	1.95	0.61	-1.75	0.07
9	27.45	60.94	4.89	27.51	64.12	4.88	-0.06	-3.18	0.01
10	29.05	63.75	3.58	29.01	58.25	3.95	0.04	5.50	-0.37

A line plot compares the actual values and projected values for the first 10 rows of the dataset's weather forecast. This provides a geometric explanation of the performance of the model in the trial set.

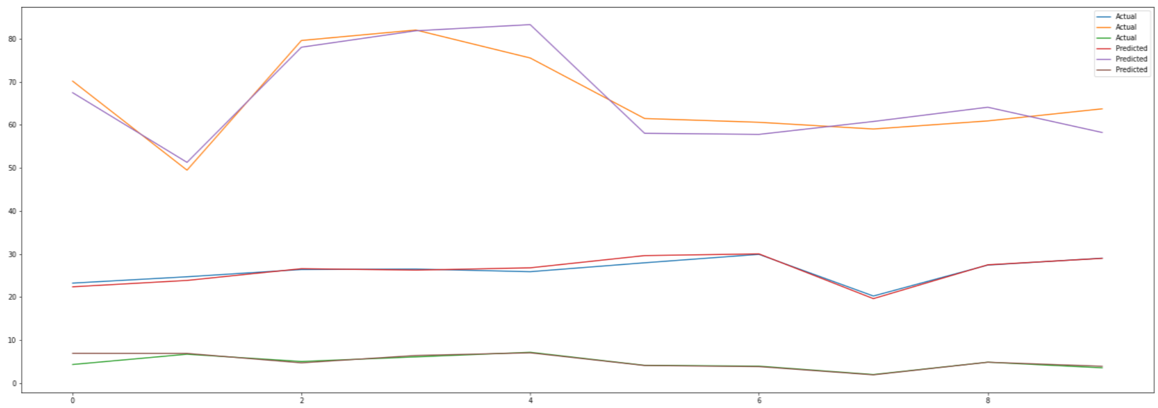


Figure 0.2: Plot showing comparison of actual value and predicted value in decion tree

4.1.3 Random Forest Tree

This random forest has high variance, hence turned out to be the uppermost accurate model. Given below is a snapshot of the actual result from the project implementation of Random Forest Tree. The accuracy r2 score is 0.92, MAE is 1.11 which is less error compare with other models also MSE is 4.56 less error.

Table 0.3: Random Forest tree for comparing actual value and predicted value

	Actual			Predicted			Difference		
No	**Temperature**	**Humidity**	**Wind**	**Temperature**	**Humidity**	**Wind**	**Temperature**	**Humidity**	**Wind**
1	23.26	70.19	4.35	22.46	68.11	6.59	0.80	2.08	-2.24
2	24.73	49.50	6.74	24.46	49.67	6.92	0.27	-0.17	-0.18
3	26.42	79.62	5.05	26.65	78.76	4.33	-0.23	0.86	0.72
4	26.53	82.06	6.12	26.48	79.32	6.81	0.05	2.74	-0.69
5	25.91	75.56	7.19	26.35	79.18	6.55	-0.44	-3.62	0.64
6	27.98	61.50	4.16	28.85	59.37	4.01	-0.87	2.13	0.15
7	29.94	60.62	3.97	30.10	58.62	4.17	-0.16	2.00	-0.20
8	20.26	59.06	2.02	20.06	61.66	2.62	0.20	-2.60	-0.60
9	27.45	60.94	4.89	27.48	63.76	4.80	-0.03	-2.82	0.09
10	29.05	63.75	3.58	29.09	60.05	3.74	-0.04	3.70	-0.16

A line plot compares the actual values with the expected values for the first 10 rows of the dataset weather forecast. This offers a geometric explanation of the performance of the model in the test set.

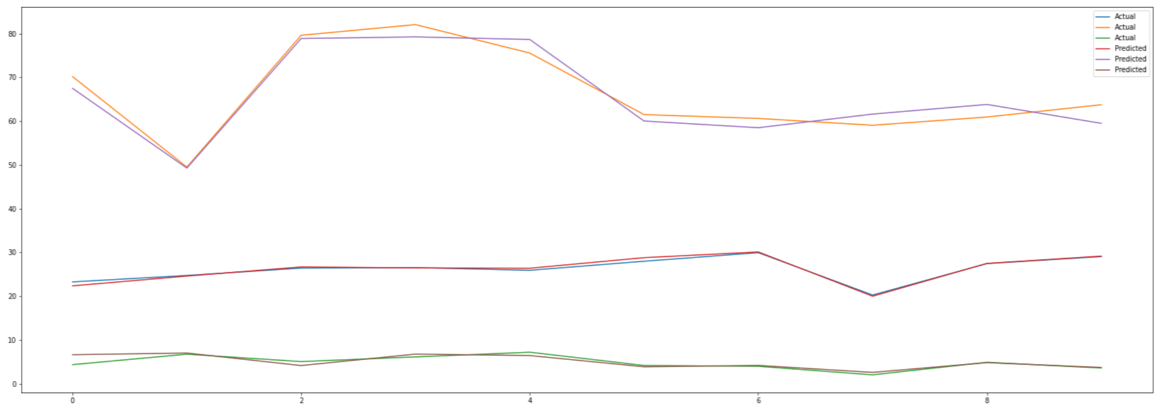


Figure 0.3: Plot showing comparison of actual value and predicted value in RF

4.1.4 Multilayer Perceptron Regression

This MLP has high variance, hence turned out to be the uppermost accurate model. Given below is a snapshot of the actual result from the project implementation of the MLP tree. The accuracy r2 score is 0.79, MAE is 1.98 which is less error compare with other models also MSE is 14.37 less error.

Table 0.4: MLP regression for comparing actual value and predicted value

	Actual			Predicted			Difference		
No	**Temperature**	**Humidity**	**Wind**	**Temperature**	**Humidity**	**Wind**	**Temperature**	**Humidity**	**Wind**
1	23.26	70.19	4.35	26.49	66.78	4.30	-3.23	3.41	0.05
2	24.73	49.5	6.74	26.41	65.99	4.29	-1.68	-16.49	2.45
3	26.42	79.62	5.05	25.75	74.06	5.21	0.67	5.56	-0.16
4	26.53	82.06	6.12	26.38	71.44	5.45	0.15	10.62	0.67
5	25.91	75.56	7.19	26.09	70.58	5.37	-0.18	4.98	1.82
6	27.98	61.5	4.16	27.25	61.61	3.35	0.73	-0.11	0.81
7	29.94	60.62	3.97	27.64	59.33	3.51	2.30	1.29	0.46
8	20.26	59.06	2.02	26.03	66.85	4.16	-5.77	-7.79	-2.14
9	27.45	60.94	4.89	27.33	60.94	3.39	0.12	0.00	1.50
10	29.05	63.75	3.58	26.87	61.39	3.24	2.18	2.36	0.34

A line plot compares the actual values with the expected values for the first 10 rows of the dataset weather forecast. This offers a geometrical understanding of the performances of the model on the test set.

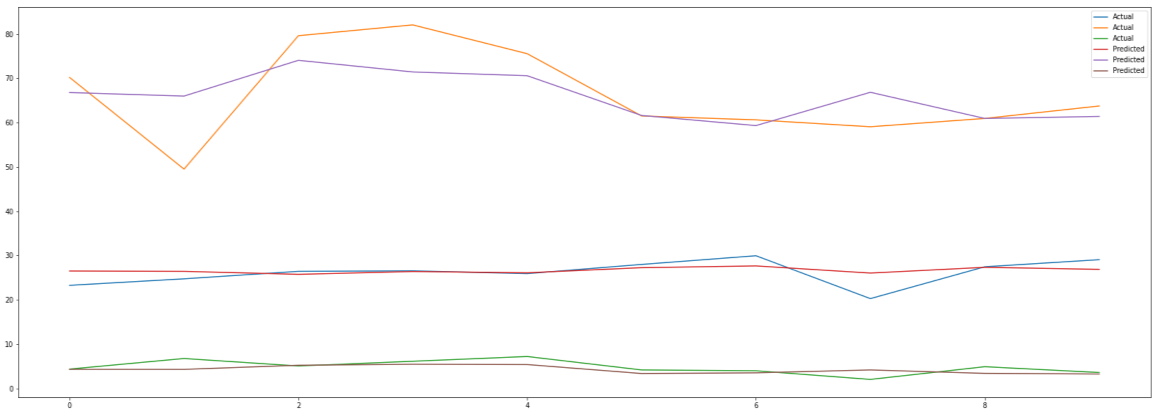


Figure 0.4: Plot showing comparison of actual value and predicted value in MLP

4.1.5 Gradient Boosting

Gradient boosting has high variance, hence turned out to be the uppermost accurate model. Given below is a snapshot of the actual result from the project implementation of the GB. The accuracy r2 score is 0.77, MAE is 2.0 which is less error compare with other models also MSE is 14.13 less error.

Table 0.5: GB comparing actual value and predicted value

	Actual			Predicted			Difference		
No	**Temperature**	**Humidity**	**Wind**	**Temperature**	**Humidity**	**Wind**	**Temperature**	**Humidity**	**Wind**
1	23.26	70.19	4.35	24.32	53.60	6.04	-1.06	16.59	-1.69
2	24.73	49.5	6.74	23.94	54.65	5.99	0.79	-5.15	0.75
3	26.42	79.62	5.05	27.23	77.83	4.35	-0.81	1.79	0.70
4	26.53	82.06	6.12	26.54	76.74	6.72	-0.01	5.32	-0.60
5	25.91	75.56	7.19	26.11	77.95	6.70	-0.20	-2.39	0.49
6	27.98	61.5	4.16	28.60	65.39	3.36	-0.62	-3.89	0.80
7	29.94	60.62	3.97	30.02	60.39	3.74	-0.08	0.23	0.23
8	20.26	59.06	2.02	20.72	63.15	2.74	-0.46	-4.09	-0.72
9	27.45	60.94	4.89	26.73	65.07	4.69	0.72	-4.13	0.20
10	29.05	63.75	3.58	28.22	69.39	2.90	0.83	-5.64	0.68

A line plot compares the actual values with the expected values for the first 10 rows of the dataset weather forecast. This offers a geometrical understanding of the performances of the model on the test set.

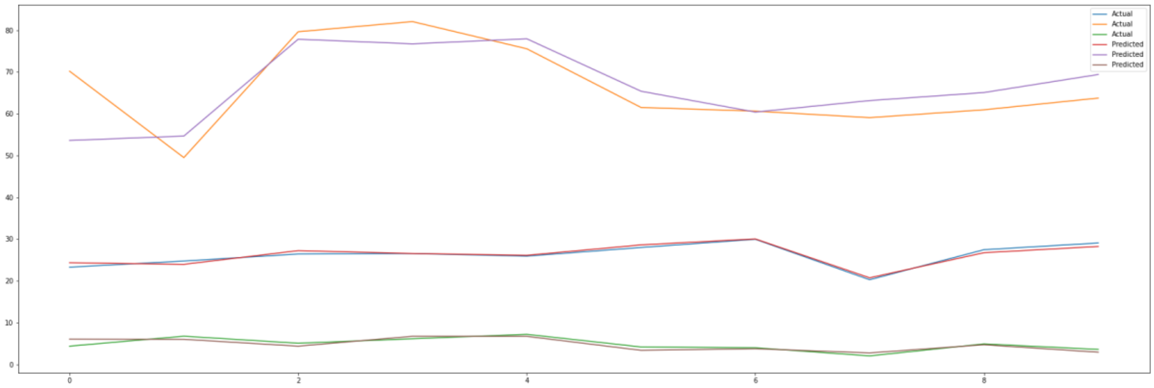


Figure 1: Plot showing comparison of actual value and predicted value in gradient boosting

4.2 Model Evaluation

Weather data is a time series. I utilize machine learning to forecast weather data. Because there is a suitable model for using weather data when the data is time series, I utilize the regression model. To determine which model is more accurate, I utilize the metrics function from sklearn, which estimates the accuracies and errors.

4.2.1 Metrics I use are

1. **R^2 Score** to evaluate the accuracy of the model.
2. **Mean Absolute Error** (MAE) is a measure of errors between paired observations expressing the same phenomenon.
3. **Mean Squared Error** (MSE) is an estimator that measures the average of the squares of the errors.

Table 0.6: Comparing machine learning models for weather forecast

No	Model	R2	MAE	MSE	Vote
1	Random Forest	0.92	1.11	4.56	1
2	Decision Tree	0.88	1.32	6.87	2
3	MLP Regressor	0.79	1.98	14.37	3
4	Gradient Boosting	0.77	2.00	14.13	4
5	Support Vector Machine	0.72	2.15	17.06	5

In the table above, I express five models which I tried this type of meteorological data. When it comes to accuracy, MAE, and MSE, those fore models provide distinct outcomes; however, the random forest tree produces highest accuracy and few errors for prediction.

The results of the RF model to forecast weather (temperature, humidity, wind) based on date and also Somalia cities (Mogadishu, Hargeisa, and Kismayo) demonstrate that the RF Model has a high performance and fair prediction accuracy. The forecasting reliabilities of the model were assessed by calculating the MAE (mean absolute error) and MSE (mean squared error) between the actual and projected values. The findings indicate that the random forests model might be a useful tool for temperature forecasting.

Chapter 5

Conclusion, Limitation and Future Work

I will concise in conclusion the main parts of Somali: weather forecast using machine learning. In limitation I will mention my project confines. Lastly not least in future work I will discuss what can be added in future in my project for taking further forecasting of Somalia using advance machine learning.

5.1 Conclusion

Due to Somalia's proximity to the equator, the climate is not very seasonal. However, from time to time, some very irregular precipitation occurs. The heat persists throughout the year with sporadic monsoon breezes and precipitation (periodic seasonal reverse winds, followed by corresponding precipitating fluctuations). The average maximum daily temperature ranges from 30 to 40°C (86 to 104°F) except at the higher altitudes and the east coast, where impacts of a cold offshore current can be felt. Participating in and watching athletic activities is anticipated to rise in popularity as the population ages but becomes more health concerned and desires to maintain a fit and trim physique.

As changes in the supply side occur, climate change will present both risks and possibilities to the whole sports-providing business. The public's tastes and preferences are expected to vary, and there will be a significant need to monitor and forecast which facilities will be most in demand. Any game requires strategy before it can be played. Scheduling the game, constructing surroundings, collecting equipment, giving transport facilities to participants, selecting venues, offering lodgings to players, and so on are all examples of planning. These are the items that must be controlled. For outdoor sports, weather conditions are an important aspect to consider while designing a game. Weather conditions have a significant impact on outdoor sports such as track and field, football, and tennis. To avoid this, the seasons have an impact on the games, particularly in the summer and winter. To offer information to decision-makers, I construct a machine learning model that can predict whether or not the game will take place at the specified time. The data will be based on previous years since 1982.

I have tested several methods for machine learning to predict weather; those trained weather forecast services have been outperformed, however the performance gap has fallen substantially for later days showing that our machines may outperform professional ones over lengthy time periods. Random forest has a high variation and it turned out to be the most accurate model. A screenshot of the real outcome of the random forest tree project implementation is provided below. The accuracy R2 score is 0.92, MAE is 1.11, which is less error than other models, and MSE is 4.56. In the first ten rows of the data set weather prediction, a line plot is produced by comparing the current and projected sequence of values. This shows how well the model performed on the test set is geometrically close to current and anticipated values.

The random forest tree model was shown to have high variance, whereas functional regression was found to be high bias and low variance. Because the random forest tree is essentially a high variance model due to its vulnerability to outliers, one method to enhance the random forest tree and other models is to collect additional data.

The findings of the RF model based on date and Somalia (Mogadishu and Hargeisa as well as Kismayo) demonstrate that the RF model has a high performance and an acceptable predictor accuracy. MAE (meantime absolute error) and MSE (mean squared error) were calculated between actual and predicted values to evaluate the models for calculation reliability. The results suggest that the random forests model might be beneficial for Somalia cities' weather forecasting so, decision-makers of outdoor games can decide whether they can play at that time or not.

5.2 Limitations

5.2.1 Benefits

This web-based software system can predict Somalia's weather using machine learning. Users may forecast three of Somalia's biggest towns, such Hargeisa, Mogadishu and Kismayo. Every time they attempt to predict the weather of three cities, they can also get each report.

5.2.2 Time

This web-based software system was launched 5 months ago, but still requires more time, to add other functionality. The major aspects of the system are accomplished at this time.

5.2.3 User Satisfaction

This web-based software is easy to understand, it's cost effective and it's help full for those who interest the weather forecast systems.

5.3.4 Sustainability

In future, the system will be easier for the user interface to traverse. It can be updated or modified for developers in the future. It might be in keeping with current and emerging technology trends.

5.3 Future Work

Because of a lack of time, a number of various adjustments, tests and experiments have been left for the future. In future study, we shall examine in depth certain mechanisms, suggest new techniques or just curiosity.

5.3.1 Recommend Depth Research

To acquire accurate information to utilize the forecast of weather, this system needs to do more profound weather study. I'd like to suggest individuals interested in participating in this study since the weather in Somalia still needs to be explored.

5.3.2 Advanced Machine Learning

Using weather forecast advanced machine learning like neural network models will be more accurate than the other algorithms used for this project. To use advance machine learning in the future needs more data. In deep learning algorithms like convolution neural network (CNN) algorithm, an artificial neural network (RNN) specifically long short-term memory (LSTM), are more beneficial for weather forecast.

References

[1] S. M. T. D. T. M. A. Subashini, "Advanced Weather Forecasting Prediction using Deep Learning," *International Journal for Research in Applied Science & Engineering Technology (IJRASET),* vol. 7, no. 8, August 2019.

[2] "Baseline climate means (1961–1990) from stations all over the world (in German)," Deutscher Wetterdienst, 2016.

[3] S. S. N. D. R. S. Christy Kunjumon, "Survey on Weather Forecasting Using Data Mining," in *IEEE Conference on Emerging Devices and Smart Systems (ICEDSS 2018),* Tamilnadu, 2018.

[4] R. L. Hadden, "The Geology of Somalia: A Selected Bibliography of Somalian Geology, Geography and Earth Science," Engineer Research and Development Laboratories, Topographic Engineering Center, 2007.

[5] A. M. M. K. a. D. M. S. H. K. Noreen Akram, "Implementation of Decision Support System for Outdoor Sports Using Machine Learning Techniques," *Journal Of Computing,* vol. 3, no. 9, September 20111.

[6] O. G. a. G. P. M. Franck Brocherie, "Emerging Environmental and Weather Challenges in Outdoor Sports," *Climate Impacts on Health,* vol. 3, no. 3, pp. 492-521, July 2015.

[7] A. Perry, "Sport Tourism and Climate Variablity," *ResearchGate,* January 2004.

[8] T. P. J. S. A. P. L. &. S. P. B. Kathryn L. Van Pelt, "Combating the Climate in British Sport," 2006.

[9] J. E. Thornes, "The Effect of Weather on Sport," vol. 32, no. 7, pp. 258-268, July 1977.

[10] B. C. Mohamed Abuella, "Random forest ensemble of support vector regression models for solar power forecasting," in *IEEE Power & Energy Society Innovative Smart Grid Technologies Conference (ISGT),* Washington, 2017.

[11] D. R. V. JAYASRI. R, "Weather Prediction Using K- means Clustering and Naive Bayes Agorithms," *JETIR,* vol. 6, no. 6, June 2019.

[12] A. D. Seema1, "Literature Review on Big Data Application, Suryansh Dabas," *JETIR,* vol. 5, no. 3, March 2018.

[13] R.Kumar, "Decision Tree for the Weather Forecasting," *International Journal of Computer Applications,* 2013.

[14] W. Y. N. N. a. Z. Z. Htike, "Forecasting of monthly temperature variations using random forests," *ARPN Journal of Engineering and Applied Sciences,* November 2015.

[15] M. R. M. A. a. I. A. Upma Singh, "A Machine Learning-Based Gradient Boosting Rgression Approach for Wind Power Production Forecasting: A step towards Smart Grid Enviroments," *Energies,* vol. 14, p. 5196, 2021.

[16] "Hargeisa geography," 2012.

[17] "Climate of Somalia," Food and Agriculture Organization., 2016.

[18] "Districts of Somalia," 2012.

[19] "Mogadishu Climate & Temperature," 2014.

[20] "Weather Forecast," 2013.

[21] A. Raj, "Unlocking the True Power of Support Vector Regression," 2020.
[22] G. M. K, "Machine Learning Basics: Decision Tree Regression," 2020.
[23] G. M. K, "Machine Learning Basics: Random Forest Regression," 2020.
[24] P. Mathur, "A Simple Multilayer Perceptron with TensorFlow," 2016.

Appendix A

The dataset of Somalia weather which contains Cities (Mogadishu, Kismayo, Hargeisa), Date (Year, Month, Day) from 1982 to 2020, also weather types temperature 2-meter, relative humidity 2-meter, wind speed 2-meter and it is 7 columns and 43279 rows.

City	Year	Month	Day	T2M	RH2M	WS2M
1	1982	1	1	26.57	81.81	6.1
1	1982	1	2	26.54	81.19	6.31
1	1982	1	3	26.4	81.5	6.75
1	1982	1	4	26.44	79.44	6.99
1	1982	1	5	26.3	78.94	7.01
1	1982	1	6	26.19	81	6.3
1	1982	1	7	26.69	80.81	6.54
1	1982	1	8	26.67	84.31	6.13
1	1982	1	9	26.3	86.81	5.28
1	1982	1	10	26.14	86.06	5.02
1	1982	1	11	25.8	80.31	5.97
1	1982	1	12	25.83	77.5	6.12
1	1982	1	13	26.15	76.06	5.76
1	1982	1	14	26.09	76.88	7.1
1	1982	1	15	26.11	77.88	7.55
1	1982	1	16	25.98	78.94	6.46
1	1982	1	17	25.89	80.69	5.83
1	1982	1	18	25.9	82.75	6.61
1	1982	1	19	26.22	82.06	6.25
1	1982	1	20	25.98	80.75	6.9
1	1982	1	21	26.26	80.94	7.48
1	1982	1	22	26.04	77.81	7.9
1	1982	1	23	26.08	77.12	7.41
1	1982	1	24	25.95	77.19	6.94
1	1982	1	25	25.87	80.38	6.4
1	1982	1	26	25.94	80	7.05
1	1982	1	27	25.87	78	6.23
1	1982	1	28	26	77.5	6.59
1	1982	1	29	25.78	75.94	6.7
1	1982	1	30	25.67	78	6.09
1	1982	1	31	25.71	81.75	5.66
1	1982	2	1	26.13	80.75	6.04
1	1982	2	2	26.29	77.94	6.2
1	1982	2	3	26.16	77	6.51
1	1982	2	4	26.05	75.56	6.28

1	1982	2	5	25.93	79.44	6.25
1	1982	2	6	26.37	79.25	6.39
1	1982	2	7	26.31	77.69	6.98
1	1982	2	8	26.36	76.75	6.39