MACHINE LEARNING APPROACHES TO ADDRESSING CLIMATE CHANGE HEALTH IMPACTS

BY

Shamiha Mosharof Smriti

ID: 232-25-015

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

Supervised By
Md. Sadekur Rahman
Assistant Professor
Department of Computer Science and Engineering
Daffodil International University

Co-Supervised By
Abdus Sattar
Assistant Professor & Coordinator M.Sc.
Department of Computer Science and Engineering
Daffodil International University

DAFFODIL INTERNATIONAL UNIVERSITY

DHAKA, BANGLADESH

JANUARY, 2025

APPROVAL

This Project/Thesis titled **"Machine Learning Approaches to Addressing Climate Change Health Impacts"**, submitted by **Shamiha Mosharof Smriti**, ID No: **232-25-015** to the Department of Computer Science and Engineering, Daffodil International University has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of M.Sc. in Computer Science and Engineering and approved as to its style and contents. The presentation has been held on **11-01-2025**.

BOARD OF EXAMINERS

Chairman

Dr. S.M Aminul Haque
Professor and Associate Head
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Internal Examiner

Dr. Fizar Ahmed
Associate Professor
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Internal Examiner

Dr. Md Alamgir Kabir
Assistant Professor
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

External Examiner

Mr. Sadat Hasan
Data Scientist
Risk Management Division,
BRAC Bank Limited

DECLARATION

We hereby declare that this research has been done by us under the supervision of **Md.Sadekur Rahman, Assistant Professor, Department of Computer Science and Engineering** Faculty of Science and Information Technology, Daffodil International University. I also declare that neither this research nor any part of this research has been submitted elsewhere for the award of any degree.

Supervised by:

Md. Sadekur Rahman
Assistant Professor
Department of CSE
Daffodil International University

Submitted by:

ShamihaMosharof Smriti
ID: 232-25-015
Department of CSE
Daffodil International University

ACKNOWLEDGEMENT

At First, I would like to express my heartiest thanks and gratefulness to almighty God for His divine blessing makes us possible to complete the final year thesis successfully.

I am really grateful and wish my profound my indebtedness to **Md. Sadekur Rahman, Assistant**

Professor, Department of Computer Science and Engineering, Department of CSE, Faculty of Science and Information Technology, Daffodil International University, Dhaka. His Deep Knowledge & keen interest with supportive instructions helped us in the field of deep learning-based research, finally I completed my work on "**Machine Learning Approaches to Addressing Climate Change Health Impacts.**". Their endless patience, scholarly guidance, continual encouragement, constant and energetic supervision, constructive criticism, valuable advice, reading many inferior drafts and correcting them at all stage have made it possible to complete this project.

I would like to express my heartiest gratitude to **Dr. Sheak Rashed Haider Noori, Professor and Head,** Department of Computer Science and Engineering, Faculty of Science and Information Technology, DIU, for his valuable support and advice to finish my project and also heartiest thanks to other faculty member and the staff of department of CSE, Daffodil International University.

At last, again I want to thank all the good wishers, friends, family, seniors for all the help and inspirations. This research is a result of hard work and all those inspirations and assistance.

Finally, I must acknowledge with due respect the constant support and patience of my parents.

ABSTRACT

With far-reaching effects, climate change is a pressing global issue. To make educated decisions and work to lessen the effects of climate change, it is essential to accurately predict future trends. Throughout human history, weather and climate forecasting have been essential for enabling efficient agricultural planning, protecting against natural disasters, and facilitating strategic decision-making across a variety of sectors. In this regard, it is imperative that forecasts be accurate and timely, and machine learning holds promise for increasing the precision and speed of prediction. This study uses online data from 2000 to 2023 to estimate future climate fluctuations using machine learning techniques, specifically advanced learning algorithms. I collect my data from Kaggle, and this dataset consists 48,000 data and 13 features have been used in this dataset. 'Date', 'Country', 'Temperature Anomaly (°C)', 'CO2 Level (ppm)', 'Extreme Weather Event', 'Economic Impact (USD)', 'Population Affected', 'Hour', 'NowCast Conc.' Etc. are features. Maintaining indoor air quality requires regular forecasting and monitoring of air pollution. As a result, machine learning (ML) has demonstrated potential in surpassing conventional methods in the prediction of the air quality index (AQI). The condition of the atmosphere is gauged by the air quality index (AQI). It determines how short-term health effects of modest exposure will manifest. Public education on the harmful effects of ambient pollutants on health is the aim of the AQI. After gathering the data and processing it all, we created a processed dataset. Using the previously processed dataset, we applied machine learning techniques. Using all 13 features, analysis between of all the related features. We use naïve Bayes, XG Boost, decision trees, random forests, and support vector machines (SVM). Applying 3 method without sampling, random under-sampling and SMOTE techniques among models outperformed the other five algorithms in our experiment in terms of accuracy; the random forest classifier's accuracy was 99.72%. AQI has 6 categories for apply the model.

TABLE OF CONTENT

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

Introduction

1.1 Introduction

Throughout human history, forecasting the weather and climate has been crucial. Weather forecasting is an essential instrument that supports many aspects of human existence and society functions, influencing anything from personal choices to extensive industrial planning. From avoiding dangerous outdoor activities during bad weather to taking preventative measures in extremely hot or cold climates, its ability to direct personal safety actions demonstrates its importance at the individual level. Forecasts are used in the agricultural sector to determine when to sow, harvest, and water crops, which eventually results in higher crop yields and more reliable sources of food [1]. Accurate forecasting has a knock-on impact on the energy industry as well, helping to effectively manage demand changes and enabling optimal power generation and delivery. This efficiency is mirrored in the transportation sector, where weather forecasts determine how aircraft, train lines, and sea operations are planned and scheduled. Accurate weather forecasts are essential for reducing delays and improving safety procedures [2]. Beyond these industries, weather forecasting is crucial to the building and infrastructure development sectors. Since unfavorable circumstances can result in project delays and quality degradation, precise projections are essential to efficient project management. Furthermore, the ability to predict extreme weather phenomena, such as hurricanes and typhoons, is crucial for disaster management because it allows for early warnings, which reduces property damage and fatalities [3]. The gradual alteration of the typical weather patterns that reflect the local, regional, and global climates of the Earth is known as climate change. This phenomenon has been connected to a number of observed outcomes. Human activity has been changing the Earth's climate since the mid-1900s, mostly as a result of the use of nonrenewable resources. This process raises the average surface temperature, which in turn raises the amount of greenhouse gasses in the atmosphere [4]. To track and look into changes in the Earth's climate, scientists use computer models and measurements from many locations, including the planet's surface, atmosphere, and outer space. Experts can spot significant indicators of climate change by analyzing

meteorological data from the past, present, and future. They include changes in cloud and plant cover, as well as variations in the frequency and intensity of extreme weather events including cyclones, heat waves, forest fires, dry spells, floods, and variations in rainfall patterns. These indicators also include rising sea levels, melting ice in the polar parts of the planet, mountain ice sheets, and increases in global land and ocean temperatures. Climate prediction has a tighter association with life on Earth, even if humans tend to overlook it in the near term. Sea level rise brought on by global warming is a serious issue that will have a significant impact on our planet's future. We may build focused mitigation plans by using advanced climate modeling and forecasting approaches to provide important insights into the possible consequences of these events. For example, realistic urban planning and catastrophe avoidance strategies in coastal communities might be informed by accurate projections of sea level fluctuations in the coming decades. Climate change has the potential to cause significant changes in the geographic distribution of many species over a long period of time, endangering biodiversity. Modern climate models incorporate a variety of factors, such as atmospheric conditions, ocean currents, ecosystems on land, and biospheric interactions, to provide a sophisticated understanding of environmental changes.

The only ingredient required for human life is air. We need to look at it and understand its quality for our own wellness. Airborne pollution causes respiratory diseases and other health problems for many people across the world. The quality of the air has significantly declined as a result of unchecked pollution. The Air Quality Index, or AQI for short, is another term for a technique used to assess and report pollution levels. The intangible quality index (AQI) is made up of twelve components, or airborne pollutants. The primary source of global warming is greenhouse gas emissions, or GHGs. Additionally, they have an impact on temperature and plant-soil dynamics, which negatively impacts farming, the environment, and human activities (Malhi et al., 2021). Using data from 2010 to 2019, the World Medical Association (WHO) published a research on air quality worldwide in 2022. Based on analyses of 6743 sites across 117 countries, this study examined a range of air pollutants as previously indicated. An elevated AQI score indicates a place where deaths are more likely to occur. Accordingly, AQI forecasting and monitoring have emerged as essential tools for ensuring long-term prosperity on a worldwide scale (Rybarczyk and Zalakeviciute, 2021). The AQI is reliably and accurately calculated by a machine learning algorithm in every situation. We can now produce AQI

estimates that are more accurate than ever thanks to machine learning (ML). The increasing amount of historical data accessible for analysis made machine learning possible. Because the dynamics of the highly unpredictable systems governing pollution levels are not well understood, it is difficult to develop a statistical model to anticipate pollution levels. A machine learning model is an excellent illustration of a symmetric and nonlinear strategy as it just requires historical data to create the connection between the uncorrelated variables. Consequently, we are able to create a more accurate forecasting model.

In conclusion, this study will look at how climate change occurred between 2000 and 2023, elucidating historical trends, pinpointing significant causes, and evaluating results and shortcomings, as well as guiding decision-making and policies. By being mindful of our objective during this pivotal moment is to improve comprehension of the principles underlying climate change and aid in developing plans to deal with this pressing global issue.

1.2 Motivation

A major global concern at the moment is how climate change is affecting human wellbeing, environmental systems, and the economy. The way humans utilize land and the emission of gasses that trap heat are two examples of human activities that are causing the Earth's climate system to undergo unprecedented changes. In order to fully comprehend the effects of climate change and the pressing need for action, research that examines its many facets is imperative. One important topic that has had a big influence on our globe recently is climate change. Extreme weather events, rising sea levels, and other weather pattern modifications have had a significant impact on ecosystems, human communities, and the planet's general capacity to support life. Finding solutions to this pressing problem requires understanding historical occurrences pertaining to climate change and how it has changed throughout time. The goal of my study in this area is to classify air pollution according to AQI values using ML and AL. Thus, we requested assistance from some of my favorite teachers. Given that several bird species are supported by the current ecology, it was proposed that we look into a relevant concept in relation to this issue. This served as the justification for my choice to write a paper with the title "Public Health Approaches To Addressing Climate Change Health Impacts." We also see how society is developing via new perspectives and how scholars are delving deeper into these subjects in an

effort to enhance my own. The following motivated this kind of research-based work. Machine learning is essential since intelligent technology has connected things inside of me.

1.3 Rationale of the study

Recent years have seen a notable increase in interest in the study of climate change and how it affects weather patterns. The phrase "climate change" refers to long-term changes in the Earth's climate system, including variations in temperature, precipitation, wind patterns, etc. Understanding how climate change affects weather forecasting is essential for prediction models to be accurate and dependable. This justification summarizes the main justifications for carrying out research in this area and highlights the significance of examining climate change in connection to weather forecasting. The frequency and intensity of severe weather events, such as cyclones, extreme heatwaves, dry spells, and heavy precipitation, have been linked to global warming. These occurrences have a significant impact on the environment, economy, and society. Researchers can learn more about the fundamental dynamics and causes that lead to catastrophic events by examining how climate change affects weather patterns. This data can improve weather prediction models' accuracy, enabling authorities to promptly issue alerts and implement efficient disaster management strategies. The rate at which things happen of extreme weather occurrences is impacted by climate change, as are general weather patterns and their geographical variability. Seasonal fluctuations, anomalies, and modified weather patterns can be caused by variations in wind, precipitation, and temperature. Researchers may identify and quantify these changes by examining climate change, which provides valuable information on for a long time trends and patterns. Forecasters may more accurately predict and communicate expected weather changes with the use of this knowledge, which is essential for improving weather prediction models [5].

There are several reasons to investigate climate change in relation to weather forecasting. By investigating the connection among climate change and weather trends, researchers may improve forecasting model accuracy, bolster readiness for disasters, ensure food security, and develop effective adaptation and mitigation strategies. Societies can more successfully handle the issues brought on by climate change and its effects on weather forecasting if individuals utilize the information obtained from this study to make educated decisions.

1.4 Research Question

Global warming and climate change have gained significant attention in the past decade, as everyone is aware. We will look at temperature variations to see whether this is the case. We'll examine how the Earth's and each continent's temperatures have evolved. We may have a better understanding of the following queries with this one:

- How to preprocess climate data?
- Which model perform best predicting AQI ?

1.5 Expected Output

The goal of the project is to improve techniques for weather prediction and climate modeling while examining how climate change affects CO2 level predictions. Through the analysis of CO2 levels, demographic effects, economic effects, factors related to climate change, and the use of sophisticated modeling techniques, the paper is anticipated to improve more reliable and accurate weather forecasts. This result will help prepare for and respond to climate change-related extreme weather events such storms, dry periods, and high heats [6]. The study will evaluate how local and worldwide weather patterns are impacted by climate change. It will look at how climate change affects temperature variations, rainfall patterns, air movements, and the frequency and severity of extreme weather occurrences.

Consequently, the following is a list of each of the previously mentioned expected outcomes:

- Analysis all the features of climate change.
- According to all the climate of each country have been analyzed.

1.6 Project Management and Finance

One project that uses machine learning is connected to ours. This section must provide a portion of our project's financial analysis. The approximate expenditures for each of the machine learning project's components are listed in Table 1.1 below.

Table 1.1: Estimated Cost for climate change

SN	Components	Estimated Cost (BDT)
01.	Visiting Stakeholders	500-1000
02.	Software and Tools	1500-2000
03.	Data Collection and Processing	500-1000
04.	Documentation and Report Writing	500-1000
05.	Contingency (10% of total)	1500-2000
	Total Estimated Cost	4,500-7,500

1.7 Report Organization

The first chapter provided an overview of the study's methodology, including its objectives, sources of inspiration, purpose, and expected outcomes. The overall structure of the investigation is also described in this section.

All of the earlier research in this field is covered in Chapter 2. Later in the next section, they provide an example of the depth that results from their limitation of this research issue. The last topic discussed addressed the primary difficulties or obstacles to this study. This chapter includes parts on pertinent studies and research abstracts as well as a discussion of the difficulties encountered throughout the project's development.

A conceptual assessment of this study's endeavor is detailed in Chapter 3. Additional details on the statistical methods specifically used in the investigation's arithmetic section are given in this chapter. This section also provides examples of practical uses for machine learning techniques. The next section discusses the methods for obtaining statistics and the system used to prepare them. A single-family confused matrix assessment is also utilized in the last section of this to assess the model and offer a suitable tag to identify the classifier. To ensure actual consistency while using machine learning methods, application evaluation is required. The study issue and methodology, operational efficacy, data collection strategy, data processing, suggested technique, teaching style, and performance criteria that must be reached in order to move this

project forward are all included in this section. Each approach to climate change is explained in full by the machine learning technology and categorization used in this work.

Chapter 4 presents the experiment's results, outcomes discussion, and outcome assessment. This chapter includes a few test photos to aid in the project's implementation. A demonstration of ML techniques and a summary of the results conclude this chapter. To identify air pollution, use the AQI number and all of the climate change characteristics.

Chapters 5 and 6 provided an overview of the research, information on the planned activities, and an outcome. A verified example that shows the report's construction complies with all standards is given in the next section. effects on the environment, society as a whole, and the sustainable development goal The chapter's final section emphasizes the limitations on our work, which might affect next generations of experts in this field.

CHAPTER 2

Background

2.1 Terminologies

We began this inquiry by reviewing a large number of relevant literature surveys. To do this, we are using artificial intelligence. Machine learning comes in three flavors: unsupervised, semi-supervised, and supervised. The term "climate change" describes a long-term, slow alteration of the Earth's climate system, which includes variations in temperature, precipitation, wind patterns, and other elements. It is mostly brought on by human actions that raise the atmospheric concentration of greenhouse gases, such as burning fossil fuels and cutting forests. The process of predicting the state of the outside world at a certain location and time is known as weather forecasting. It is essential to analyze the current atmospheric circumstances using mathematical models and observational data in order to anticipate future weather conditions such as Temperature Anomaly (°C), extreme weather event, and air quality index. AQI forecasting is influenced by a number of factors, including the initial quality of the air, air-related activities, and the availability of precise recorded information. It can be predicted by some climate models. Computer-generated numerical simulations of the interactions and regulation of the climate of the planet by the atmosphere, seas, land surface, and ice are known as climate models. climatic models are used to forecast future climatic trends and assess the effects of climate change. They provide the models used to predict the weather useful information.

The main topics of this part include the study's findings, challenges, pertinent literature, and a research summary. I will examine research articles written by other writers under "Related Works," discussing the connections between their thoughts, accuracy, and methods. In the section on comparable works, I will discuss the articles, approaches, and credibility of other scholarly publications that are relevant to my research. The section on research descriptions will offer a summary of my pertinent work. I describe how I overcame every challenge I ran into during the research and how I increased the dependability of each stratum, even in the most difficult parts. Everything has already been investigated.

2.2 Related Works

Climate change is already having an influence on Earth, and it is predicted that these effects will only get worse in the future (Vishvakarma2022) [7]. Unprecedented consequences of global climate change include rising sea levels that increase the probability of catastrophic floods and

shifting weather patterns that jeopardize agricultural output. Future adaptation will be simpler and less expensive if preventive measures are taken now rather than later. We want to define "Global Climate Change" in this introduction, as well as discuss some related concepts, possible effects on human welfare, and possible solutions to this problem. It highlights how urgent action is required to stop the serious accumulation of atmospheric greenhouse gases (GHGs) and the ensuing increase in global temperatures, which might have disastrous effects on economies and societies throughout the world [7]. They offer a thorough overview of the numerous advancements in weather and climate research and forecasting, as covered in Stocker's (2014) work (Krishnan 2020) [8] . It's common knowledge that "Climate is what you expect and precipitation is what you see. " Climate is a component of the air system that changes gradually and is determined by the typical atmospheric conditions, whereas weather mostly refers to short-term (a few days to a few minutes) variations in atmospheric variables including air motions, moisture content, and temperature. Forecasting the weather with numerical techniques (NWP):In 1904, the Norwegian weather scientist Vilhelm Bjerknes (1862–1951) was the initial individual to recognize that forecasting the state of the atmosphere can be done by identifying a set of algebraic formulas (more precisely, nonlinear partial differential equations) that govern the laws regarding atmospheric dynamics.

In 1917, Lewis Fry Richardson, an Englishman (1881–1953), tried to create the first numerical weather forecast (NWP) by mixing equations mathematically. Richardson used the primitive equations, sometimes referred to as the fundamental gas equation and the formulae for preserving momentum, energy, and mass, and began with temperature and pressure data from stations located around Europe. "Weather Prediction by Numerical Operations (Richardson 1922)" described the novel approach of attempting to manually estimate the surface pressure change over the following six hours, but it turned out to be incredibly impracticable. Much later, it was discovered that Richardson's inaccurate forecasts resulted from unstable meteorological conditions at the beginning of the computer program. Furthermore, the need to make sure that the model time interval is compatible with the mesh size or model grid dimensions was cited as the reason for the numerical partitioning breaches (Courant et al. 1928). Neumann created the first-ever supporting numerical weather prediction (NWP) model for height fields in the intermediate layers of the atmosphere after introducing a more straightforward set of formulae. This model established the foundation for next developments in NWP.NWP has undergone a

quiet transformation in recent years due to the ongoing collection of scientific data and technical advancements over time (Bauer et al. 2015) [9]. Because it involves predicting the future atmospheric conditions by solving a connected non-linear structure for differential equations (PDEs), modern computerized weather forecasting (NWP) requires powerful computers. For every time increment, from the present to many weeks or months in the future, this method requires billions of computations. In 2020, Asha, S., S., and P. Reena[10] The author approach, which is used to forecast time series data, is based on a model that incorporates non-linear patterns with regular daily, weekly, and annual changes while accounting for holiday impacts. The Prophet is accessible in versions for the Python and R programming languages.

In the fight against climate change, machine learning has become a potent instrument that aids in resource optimization, creative problem solving, and improved forecasting by researchers and decision-makers. Machine learning models are used mostly in climate modeling, which requires them to analyze large databases of historical weather data and climate patterns to increase forecast accuracy. More accurate predictions of temperature changes, the process of precipitation and extreme weather occurrences are made possible by these models, which include deep learning and neural networks and improve our knowledge of complex climate systems (Rolnick et al., 2019) [11].

Algorithms for machine learning are used in renewable energy management, another application field, to improve the generation and utilization of energy, especially in solar and wind energy systems, in order to maximize demand and minimize waste (Kumar et al., 2020)[12]. In the shift to cleaner energy systems, machine learning is essential because it makes it possible for renewable sources to be better integrated into the energy grid.

Additionally, machine learning helps in carbon capture and sequestration by enhancing greenhouse gas emission monitoring and detection. Machine learning helps to mitigate climate change by reducing atmospheric carbon through the identification of emission hotspots and the optimization of carbon capture techniques. Precision farming, which optimizes crop yields and utilization of resources to reduce environmental consequences, is another way that machine learning supports sustainable agriculture. Machine learning supports climate-smart agriculture by helping with crop planning, pesticide application, and water use efficiency by utilizing data from soil sensors, variations in the weather, and crop health (Shakoor et al., 2021)[13].

These developments demonstrate how machine learning may be used intelligently to combat climate change from a variety of perspectives. To properly utilize machine learning in climate action, however, issues with data accessibility, model understanding, and scalability still need to be resolved (Huntingford et al., 2019)[14].

2.3 Research summary

Our present effort focuses mostly on the range of approaches that the larger community may offer. Climate change can analysis all the features of all data. Overall, we employed five distinct methods and added more strategies to our dataset. Our publicly accessible online Kaggle collection served as the primary source of data for this example. As mentioned before, the material in our compilation includes both recently collected and previously used data. We will be able to evaluate the impact of the additional data we supplied from the same source as well as the efficacy of the five strategies we employed. It implies that there can be several comparable classes and types of labeling. Python was used to develop most of the machine learning techniques and related text classification algorithms used in the data collecting and feature collection procedures.

2.4 Scope of the Problem

In order to provide projections and estimations of future climate changes, the research article primarily focuses on integrating and analyzing existing climate models, scholarly publications, and observational data. The objective is to look at potential changes in global weather patterns, including variations in temperature anomalies (°C), extreme weather events, CO2 levels, and the frequency and intensity of severe weather occurrences. Our dataset essentially collects temperature data. Additionally, it focuses more on forecasting temperature changes. It acknowledges that inherent uncertainties pertaining to complex weather systems affect how accurate these forecasts are. Our research mostly focuses on building models using data analysis and machine learning techniques. These days, using machine learning and artificial intelligence to different tasks like air pollution prediction and object identification yields quite decent results. Therefore, we decided to use machine learning to create a model for predicting climate change.

2.5 Challenges

The primary issue with this study seems to be gathering and analyzing all of these data points, given how difficult it was to examine a single large data file. To clean and normalize the data set, we employed a variety of tools and techniques. Because of the quantity of the data files and the

variety of values they included from different historical periods, it required some time to arrive at the appropriate conclusions. The development of NLP (natural language processing) techniques for automatically categorizing instances of the AQI category has attracted a lot of attention due to the potential use in identifying and decreasing such behavior. However, this undertaking is fraught with difficulties. AQI-based climate models are helpful tools for predicting future climate change in a country. They do, however, have inherent uncertainties. Uncertainties in model results arise from the complexity of climatic systems, limitations in data collection, and various modeling methodologies. Accurately comprehending and disseminating the projections requires the identification and measurement of these uncertainties. The quality and accessibility of information are critical to the accuracy and reliability of climate forecasts. Biases and ambiguities in the AQI projections may result from missing data in historical climate records, sparse data in certain regions, and discrepancies between several sources. Reliable forecasting requires addressing data constraints and making sure that robust procedures for collecting and evaluating data are in place.

- Data collection
- Importing of data
- Preprocessing of data.
- Analysis dataset climate change features.
- Applying without sampling, random under-sampling, SMOTE techniques.
- Reaching the greatest accuracy levels of more than 90%,
- Selecting an ML model according to test outcomes.

CHAPTER 3

Research Methodology

3.1 Introduction

The next part discusses the research methodology, including how to gather datasets, execute each test, and employ each model to increase accuracy. This chapter also included recommendations for an all-data research and approach. Therefore, in an effort to improve and simplify the facts offered in this chapter. This research component offers a thorough overview of the complete process while concentrating on the techniques used to assess climate change in statistics. This section will detail the entire methodology of the study. Every given analysis can be solved in a variety of ways. The next step is to choose an ML approach. As we have already seen, we are employing five distinct machine learning methodologies, thus developing a data store is essential to developing the framework and executing the algorithm. The model is then trained using the gathered data. The data were then used to build training and testing sets. There is a common misunderstanding between "training dataset" and "testing dataset." After the input is taken out of the data set that must be created and fitted into various ML technique models, we are limited to access to a significant amount of the data required to evaluate our model. We then went on and explained everything using our simple process flow diagram.

3.2 Research subject and Instrumentation

A research area is one that is examined and explored to clarify ideas for managing modeling, data collection, task completion, and proving teaching, in along with execution. As measuring specialists, we discuss the tools and techniques we employ. Their Windows operating system, NumPy, Scikit Learn, and OpenCV were among the technologies we used, along with Python as our programming language. The software is executed on a GPU. Google Colab was the platform of choice for all instruction and assessment. Python programmers may use Google Colab to build code for machine learning and data science techniques. These methods employ statistical techniques linked to machine learning to identify clusters, or the six AQI categories for country-based climate change type classification.

Used libraries of all sorts:

- **Matplotlib**: This collection of Matplotlib tools makes it easy to plot, score, organize, and graph. To highlight certain fictional points of viewpoint or the story's limits of plausibility, this might be used in form-building.
- **NumPy**: Vector processing has been made simpler by the NumPy library in the language. These subject covers conversion indices, matrix computations, and the inverted convert of a Fourier series in depth. There are several tools and methods for working with various types of matrices in the NumPy Python packages. With the help of NumPy, the process of making devices might become more logical and practical. The Python module NumPy was created primarily for mathematical evaluation, to put it simply. This can alternatively be represented with "estimates. Py".
- **Scikit-learn**: For modeling data and analysis, Scikit-learn is an effective and user-friendly tool. The layout made use of three Python tools: Matplotlib, SciPy, and NumPy. Anyone can utilize these publicly available, open-source technologies.
- **Seaborn:** This well-liked Python data visualization tool will soon incorporate Matplotlib in its next version. This is an easy-to-use program for visualizing data artistically.
- **H5Py:** The h5py module in Python enables users to read HDF5 code that has been fragmented. HDF5 is used to store the data after NumPy has processed a sizable portion of it, primarily integers.
- **Pandas**: Pandas offers a freeware toolbox for dealing with and analyzing language-specific data. Statistical analysis techniques and basic data types can assist guarantee systematic data management, especially when employing brief information.
- **Python OS**: Developers may communicate with one other using the same vernacular thanks to the different functionalities offered by the Py OS element.

The data, software, and hardware resources utilized in this configuration for "TITLE." have been carefully selected to conduct in-depth study. Computing systems with a processing core composed of strong CPUs and potent GPUs are known as high-performance computers. I understand that no invention can provide perfect results. In a similar way, we may adjust our model's parameters as it's being trained to increase accuracy.

3.3 Proposed Methodology

This study most likely employed a variety of methods or strategies to achieve its goals. The workflow had to be selected, the written material and all values had to be cleaned up, the medication labels had to be gathered and altered, and the effectiveness of the classifier was evaluated using the outcomes of a randomly selected forest classification approach.

Step 1: Data Collection: From Kaggle is the source of our whole dataset have been collected. To predict climate change, the entire data set is utilized from 13 features. Since it is challenging to gather data for the particular AQI values using top-notch analyzers and classification type, this organization lacks a sizable, complete dataset.

Step 2: Prepare the data: After every feasible method of gathering data had been exhausted, each piece of information was examined independently. We are surrounded by numerous examples of poor and unclear language. Before using the final element of the dataset, we should review it.

Step 3: Data processing: Even after the dataset has been assembled, the "AQI" and "AQI category" still guide the creation and processing of the data for prediction. Sorting the data, removing null values, and displaying it are all necessary for training. There has been insufficient preparation of the data for separation.

Step 4: Model Selection: After selecting a prediction strategy, we train it with my data and evaluate it to boost dependability. There are a number of filters used in machine learning. Although many designs were employed to improve the design of the parts and allow the machine learning algorithm to identify the type of climate change, just one instrument was eventually chosen to assess the reliability of the data.

Step 5: Evaluation of Performance: The subsequent portions of this step cover every consequence. Applying 3 techniques like without sampling, random under-sampling, SMOTE techniques. Following the training and assessment stage, these approaches provided us with a limited level of consistency for the label groups of the six different AQI category datasets. The confusion matrices were supported by f1 scores and accuracy statistics. Each outcome is

described in this section. Even after testing and training, these strategies did not provide us enough reliability for the following two courses. They created visual aids for f1 measurement, memory, efficiency, and confusion matrix in addition to a system for classifying different country climate change & air pollution.

Step 6: Final Thoughts and Future Plans: The development plan for this field will be summarized.

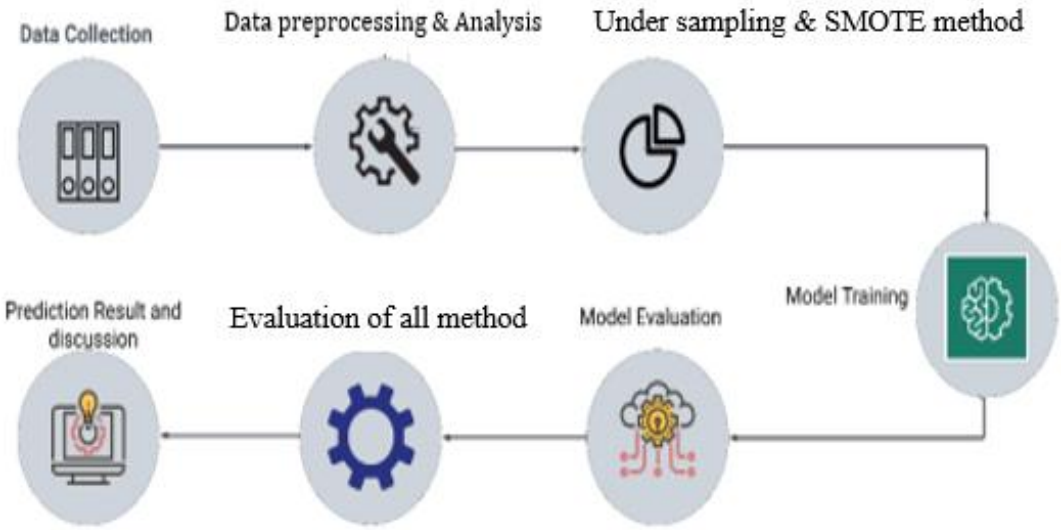

Fig 3.1: The suggested model for the entire study endeavor.

Figure 3.1 illustrates the phases of our study methodology, which can be used to calculate the AQI scores for the contaminants that comprise each index of air quality category as well as climate change. Since our web data was collected from a variety of sources, it is reliable and genuine. Each kind of air was assigned an AQI value in order to construct the dataset. We have reviewed this data collection, removed any unnecessary numbers, and cleared up the wording to ensure that the entire data set only includes accurate and relevant information. To investigate machine learning methods, we build and improve models using pre-existing data. To address the issue of class heterogeneity, we additionally use permutation approaches to guarantee fair inclusion and improve the overall effectiveness of the model. Our approach not only provides a summary but also searches for ways to reduce the quantity of erroneous results. We are able to offer a comprehensive method that efficiently manages the values in the air quality index numbers used for air pollution forecasting by combining machine learning technology with linguistic understanding techniques.

3.4 Data Collection

A vastly scalable collection of pertinent and conveniently available coordinates makes up the data set. I collect my dataset from kaggle source. This dataset has year of 2000 to 2023 data of 10 country. About 48000 data and 13 features have been consists in this dataset.

Table 3.1 displays all of the data elements present in each of these files, classifying them into 13 primary categories:

Table 3.1: Columns of Description in the Dataset

Column's Name	**Description of the Column's**
Date	Year of 2000 to 2023
Country	10 countries
Temperature Anomaly (°C)	Difference from average temperature (°C) in country.
CO2 Level (ppm)	Level of CO2 in every country
Extreme weather event	Weather event of every country
Economic Impact (USD)	Every country economic impact by climate change
Population Affected	Every country pullulation affected impact by climate change
Date (LT)	Date
Hour	Hour 0-23
Nowcast Conc.	Nowcast Concentration in micrograms/cubic meter.
Raw Conc.	Raw Concentration in micrograms/cubic meter.
Conc. Unit	Concentration Unit- micrograms/cubic meter
AQI	Air Quality Index
AQI Category	"Hazardous," "Very Unhealthy," "Unhealthy," "Unhealthy for Sensitive Groups," "Moderate," and "Good."

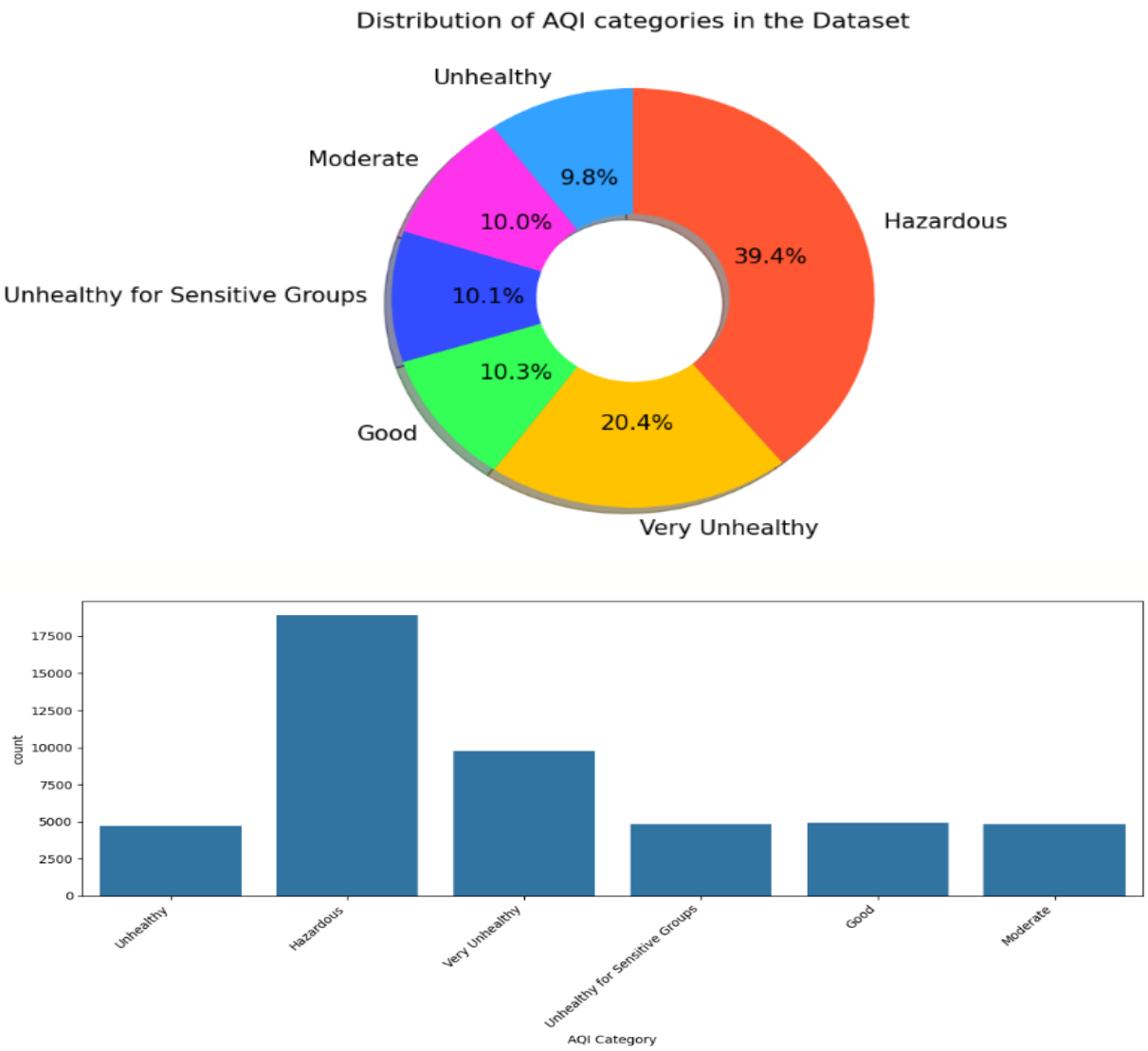

Fig. 3.2: 6 classes of AQI data contents

Label Encoder:

Regardless of the dataset's line count, machine learning experts frequently work with datasets that contain many features. These unique IDs can contain both letters and integers. Keywords are frequently used in training to categorize data in order to improve user comfort and comprehension. Encoding the labels is one method of transforming signals into computer-readable text or numerical values. To create a structure that conveys numerical values, identifiers must be translated within the framework of the aforementioned method. Ultimately, the creators

of the machine learning algorithms determine how to use these labels. After the data has been supplied, a preliminary analysis has to be carried out while taking the modifications under observation into consideration.

3.5 Statistical Analytics

3.5.1 Data Preprocessing

The process of turning primary data become well-structured data is known as data preparation. Data processing is a crucial step in doing climate change predicting research. It entails modifying, classifying, and analyzing gathered data in order to derive important insights and conclusions. Combining several datasets can yield data that can be used for training and testing. We started by eliminating any unnecessary characters and symbols from the database's contents that had null values in order to start correcting any errors. Tokenization is another technique that converts words into numerical vectors that may be readily concatenated and used in machine learning models. Below is a summary of the data processing procedures that were employed in our research article:

- Since they are not anymore needed, remove the many columns that contain code.
- Normalize the data to a standard scale to provide interoperability among models.
- In order to prepare the data for analysis, we have made the required changes.
- Converting formats for information, aggregating data into suitable time periods (daily, monthly, etc.), or creating new variables from preexisting ones (calculating anomalies, trends, etc.) have all been part of this process.

3.5.2 Duplicate and Null value remove

A crucial step in the production of datasets is the modification of numerical values. This strategy's two primary tactics are intended to improve the caliber and applicability of written material. Additionally, a text correction approach is used to eliminate them. Both null values and duplicate values have been cleaned up. To guarantee that every piece of information we collect is prepared for a thorough analysis, we employ a painstaking preservation procedure. Three rows, one associated with every statistic, reflect each of the six types of AQI.

3.5.3 Text Cleaning

Our use of text alteration is a crucial component in the dataset's creation. This strategy uses two main methods to raise the standard and relevance of written content. News articles are vetted before publication and are only kept for a set 100 words. Filtering techniques protect our pledge to deliver excellent, law-abiding, and instructive information. To alter the material in a systematic way, a text correction technique is also used to eliminate superfluous symbols like sections, stop categories, emoji removers, special markers, etc. A number of common symbols, line breaks, and English letters have been eliminated from the text. To ensure that everything we gather is ready for a careful examination, we use a thorough preservation process.

3.5.4 Tokenization and Padding

Two essential elements of modern data utilization are tokenization and padding. For our system to comprehend English, words must be tokenized, or converted into number sequences. The connection between the written word and symbolic numbers is established by giving each word a unique number. Padding increases the likelihood of agreement by ensuring that each sequence has a set time throughout a planned training session. In order to evaluate the symptoms model, our computers translate sentences into dollars and cents and confirm that there is sufficient variety in the total word lengths. For our computer systems to properly evaluate the data and differentiate between the six primary categories that are reflected in the climate change of AQI & AQI category, this step is essential.

Fig 3.3: Example of Tokenization.

3.5.5 Data Preparation

In order to evaluate the model and training throughout the data preparation process, we did not randomly partition the data, even after eliminating null values and duplicate data. The climate change dataset has 48,000 records, and we choose the most significant "AQI" and "AQI

Category" variables. Train and Test are the two components that make up the dataset. 9,600 data points are in the test, and 38,400 data points are in the train.

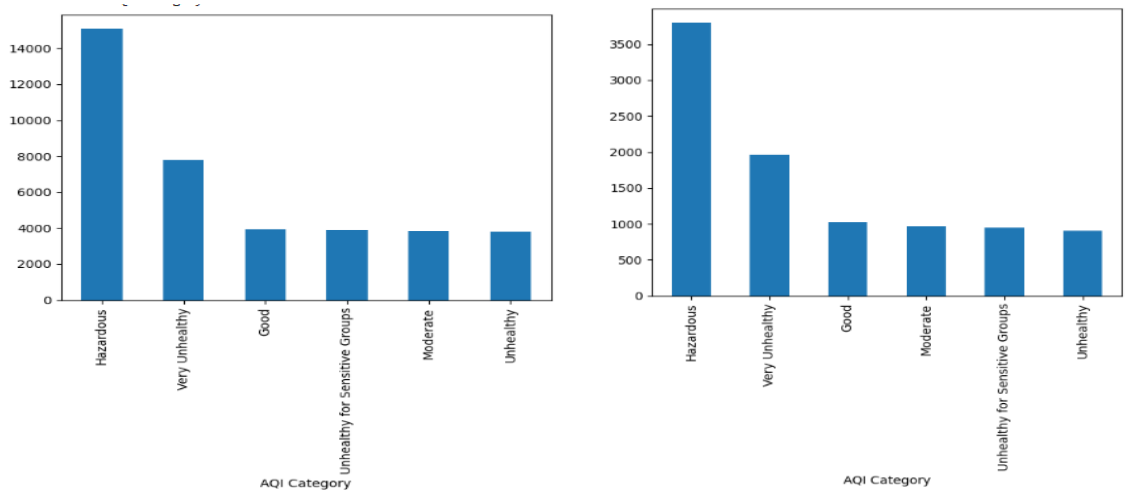

Fig 3.4: Climate change AQI Data Count for Train and Test.

3.5.6 Statistical Analysis

This dataset consists of 13 attributes or columns. Thus, each feature's count plot is shown in the figure below.

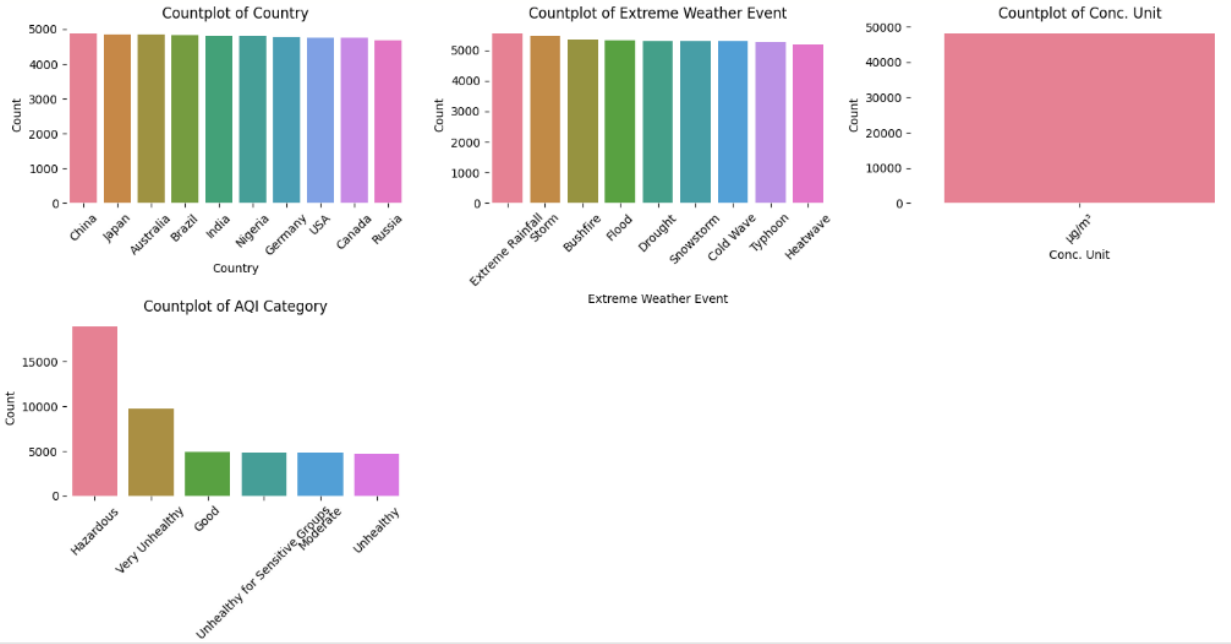

Fig 3.5: All the countable features' data count.

Here all columns have their own class shows as count plot. This fig 3.10. shows all count of each column.

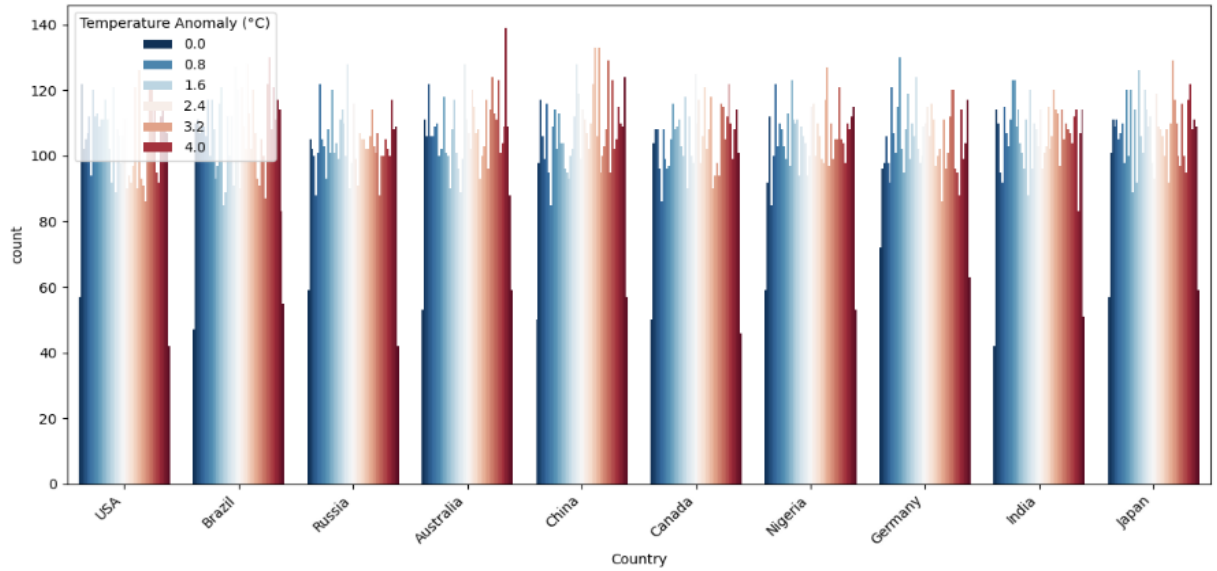

Fig 3.6: Temperature Anomaly (°C) by country based

From the fig 3.8 we can see temperature anomaly of 10 countries. In Australia temperature anomaly (°c) is high (4.0) around 140.

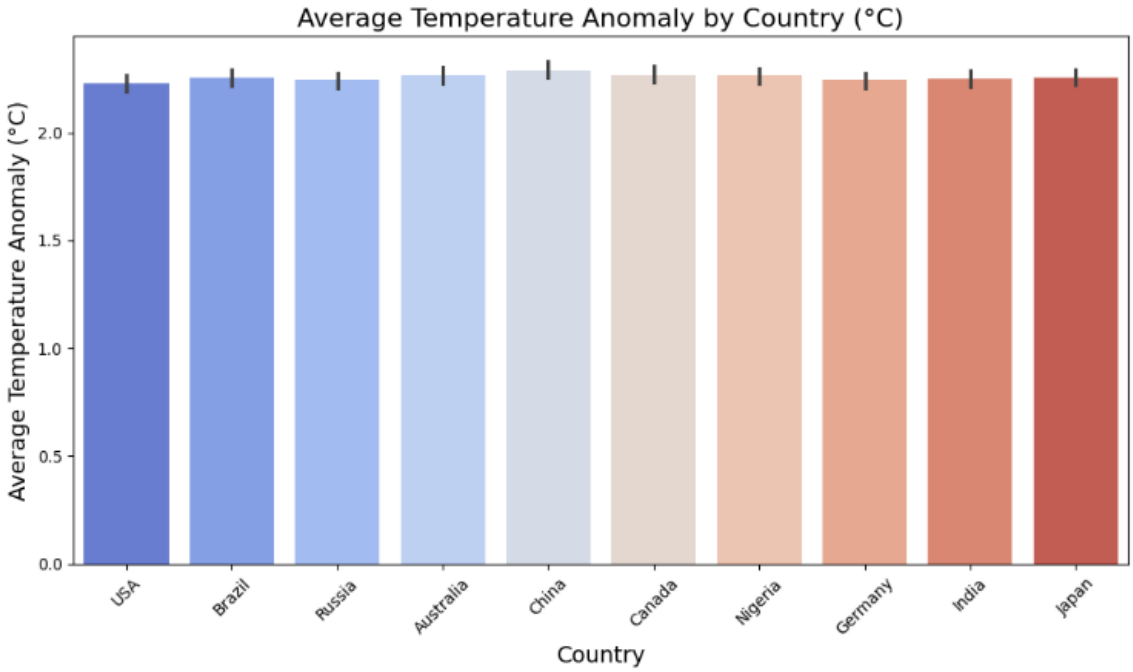

Fig 3.7: Average Temperature Anomaly (°C) by country based

The fig 3.9 shows Average Temperature Anomaly by country(°c). All countries range of average temperature over 2.0 and China achieve the highest range.

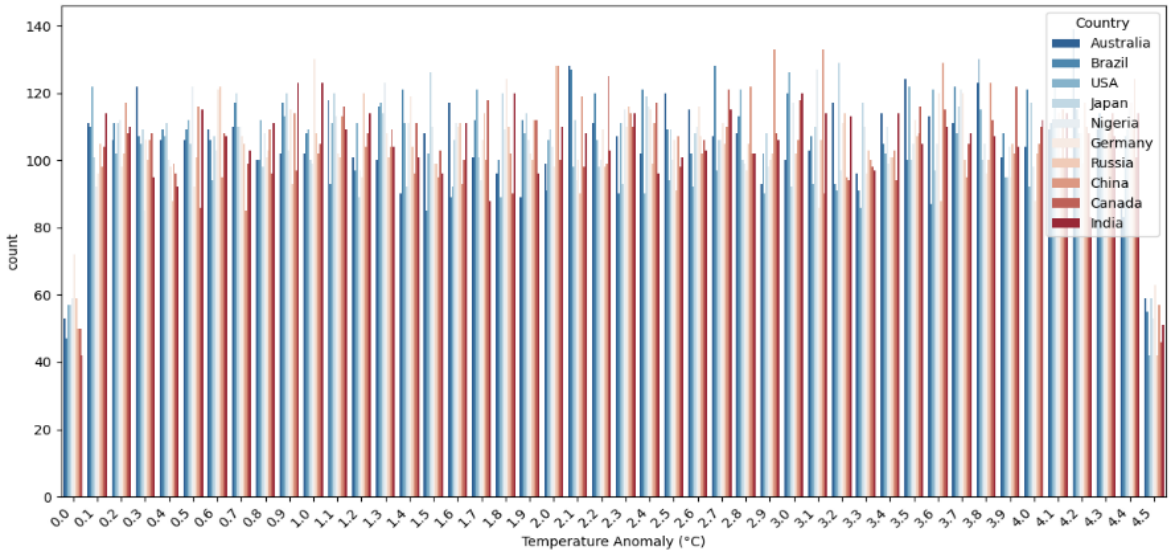

Fig 3.8: Country by Temperature Anomaly (°C) based

In this fig 3.10. shows country-based Temperature Anomaly (°c). Temperature anomaly 4.2° for the Australia and Brazil country count for almost 140. Then all the country temperature are average. 0.0° are the less for all country.

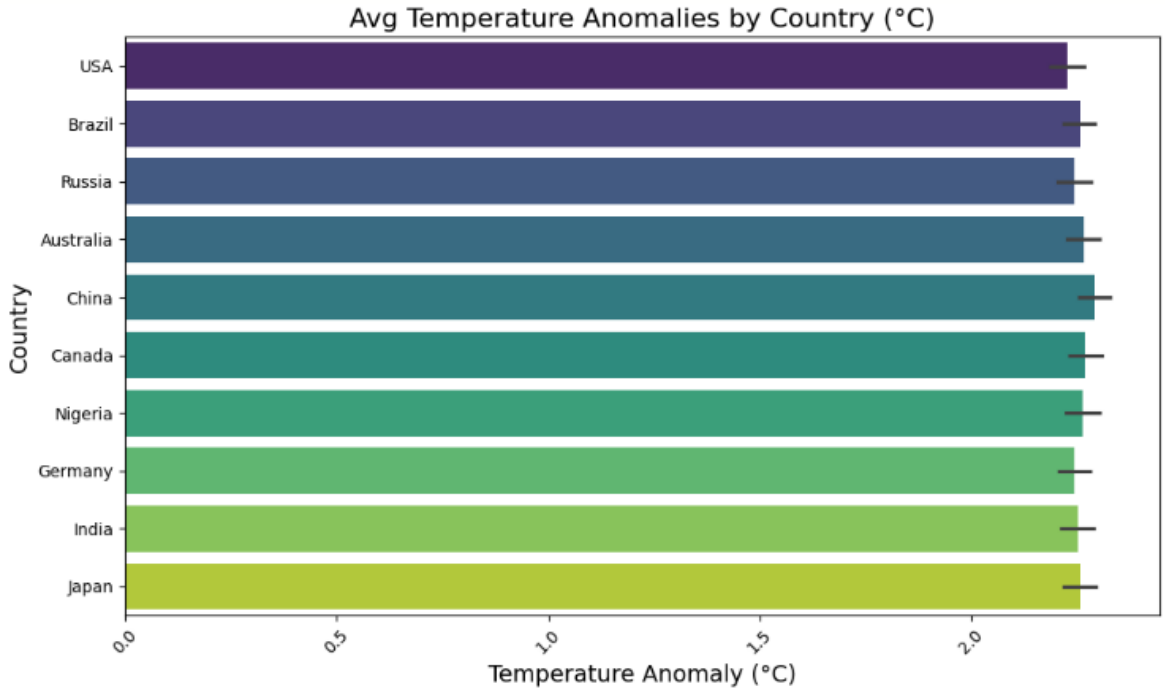

Fig 3.9: Average Temperature Anomaly (°C) by country based

The fig 3.11 shows the Average Temperature Anomaly by country (°c). China have taken the top position and USA has taken the last position in this graph.

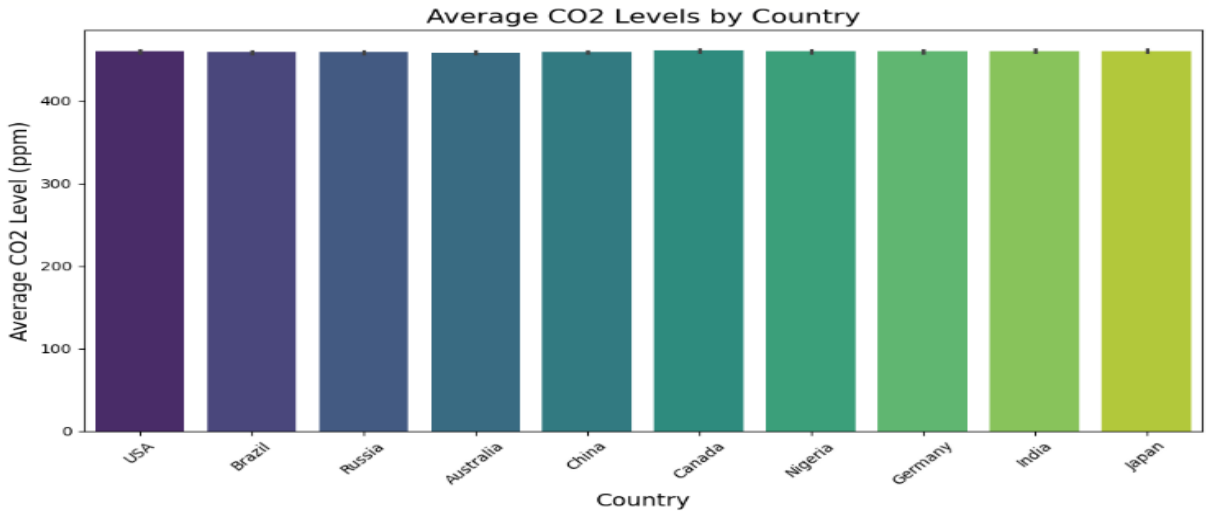

Fig 3.10: Average CO2 levels by country based

From this 3.12 we can see the Average CO2 level(ppm) by country. The range of CO2 level is over 400 ppm.

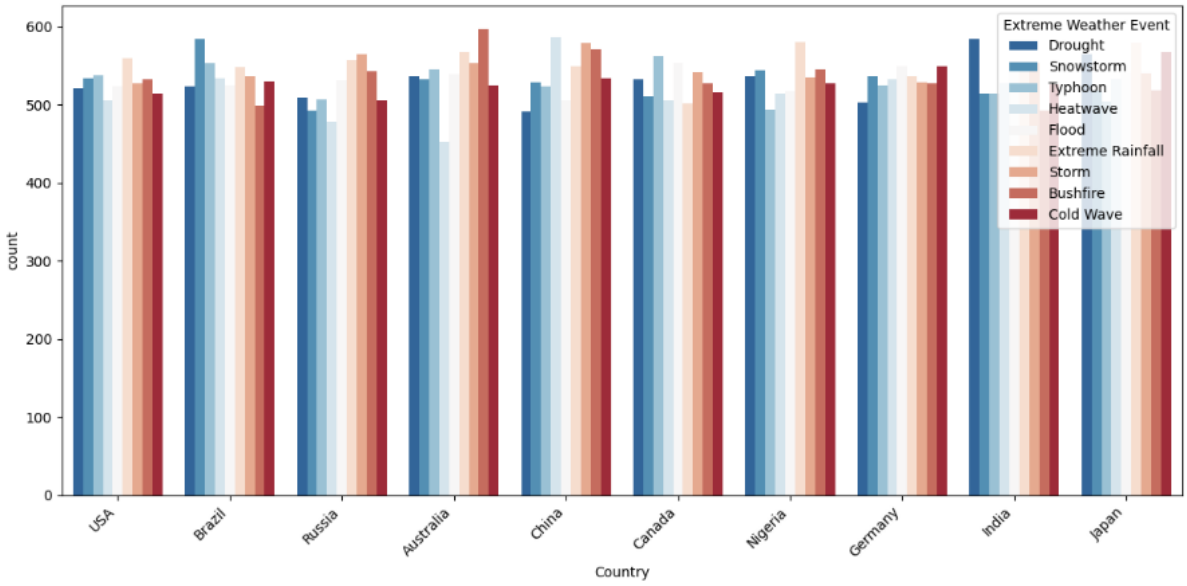

Fig 3.11: Extreme weather event by country based

The fig 3.13. shows the Extreme Weather Event. In this graph we can see most drought area is India almost 600, most snowstorm area is Brazil almost same, extreme brushfire in Australia ,cold wave in Japan etc. all the same almost 600.

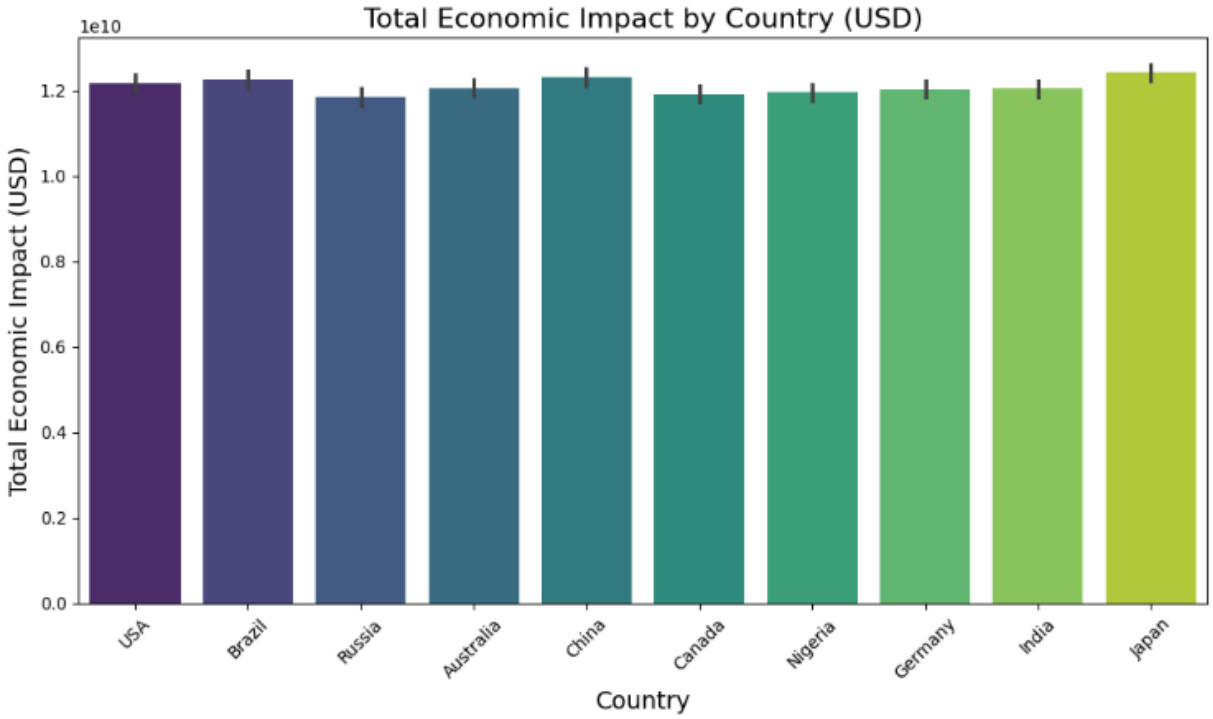

Fig 3.12: Total economic impact by country (USD)

The fig 3.14 shows the Total Economic Impact by country(USD). Highest position is taken by Japan and it's over 1.2 USD. Russia is in lowest position.

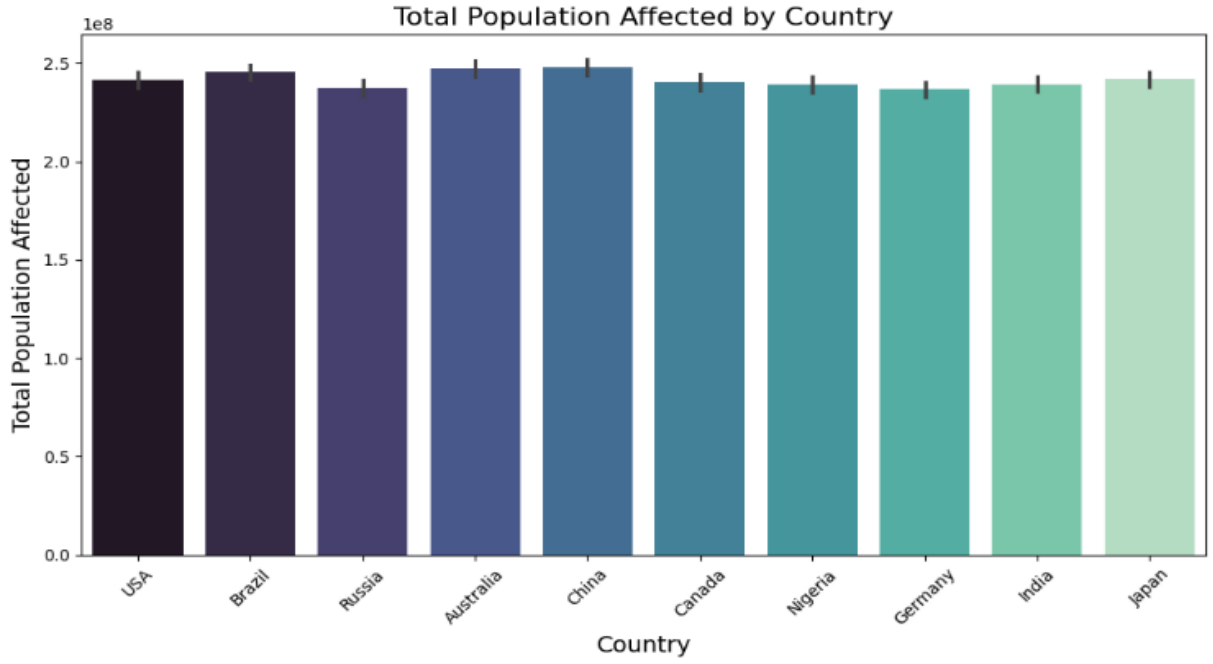

Fig 3.13: Total population affected by country based

The fig 3.15 shows the Total population affected by country. Total population affected is highest in Australia 2.5 avg ,China and lowest in Russia.

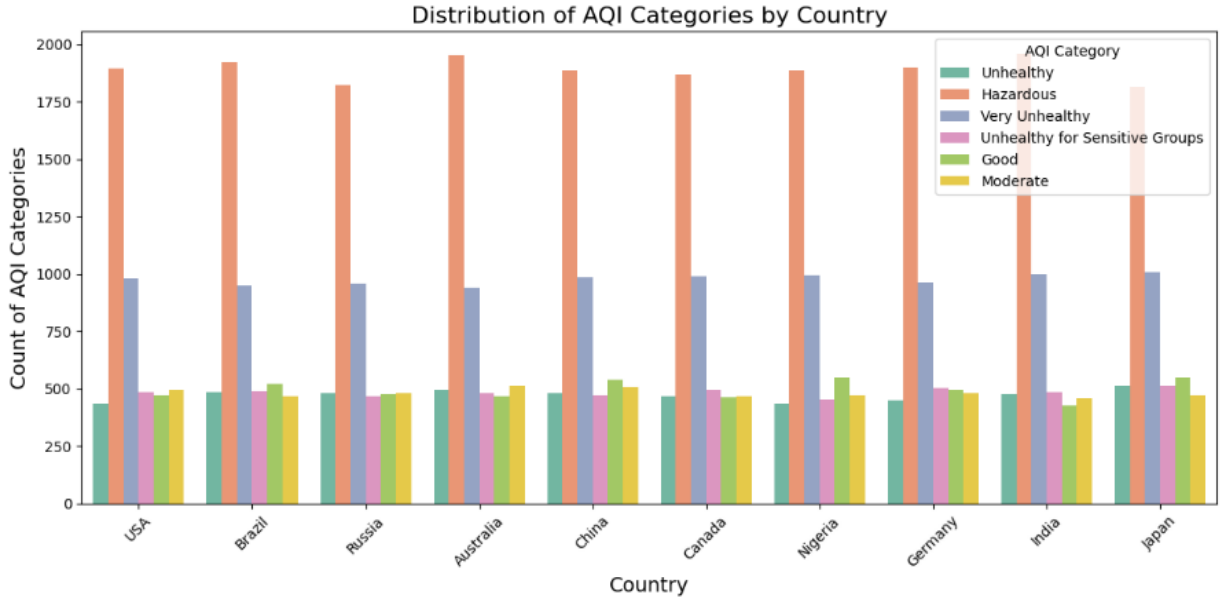

Fig 3.14: Distribution of AQI categories by country.

The fig 3.16 shows the Distribution of AQI Categories by country. In this chart Hazardous level is high and highest position in India 1900. Other level of AQI Categories is in similar range.

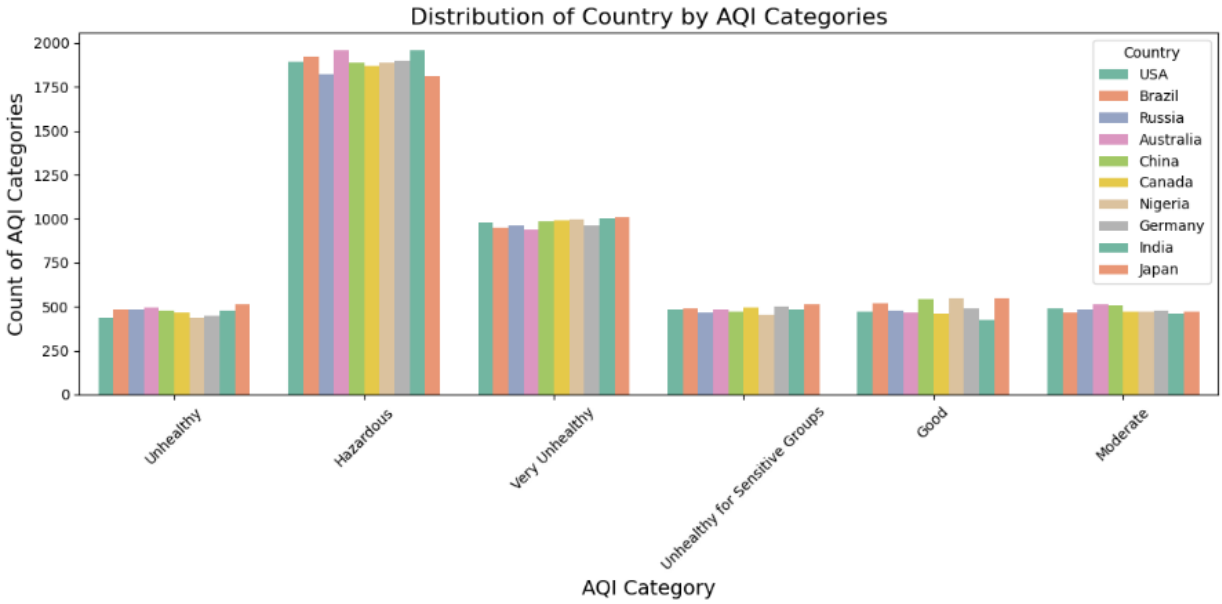

Fig 3.15: Distribution of country by AQI categories.

The fig 3.17 shows the Distribution of country by AQI Categories. Hazardous range is much high and it's range almost 2000. Second high range is Very unhealthy and it's range between 750-

1000.

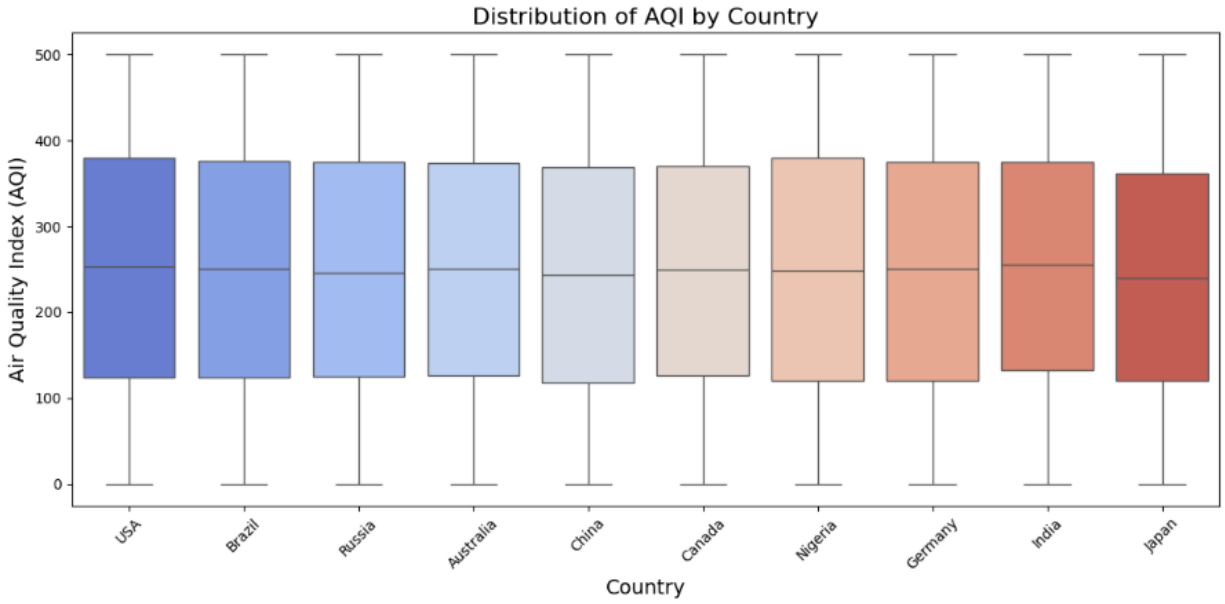

Fig 3.16: Distribution of AQI by country.

The fig 3.18 shows the Distribution of AQI by country. 0-500 shows the AQI range. All country AQI range is over 100 to under 400 and it's middle point around 350. But middle point of Japan is quite low than other countries.

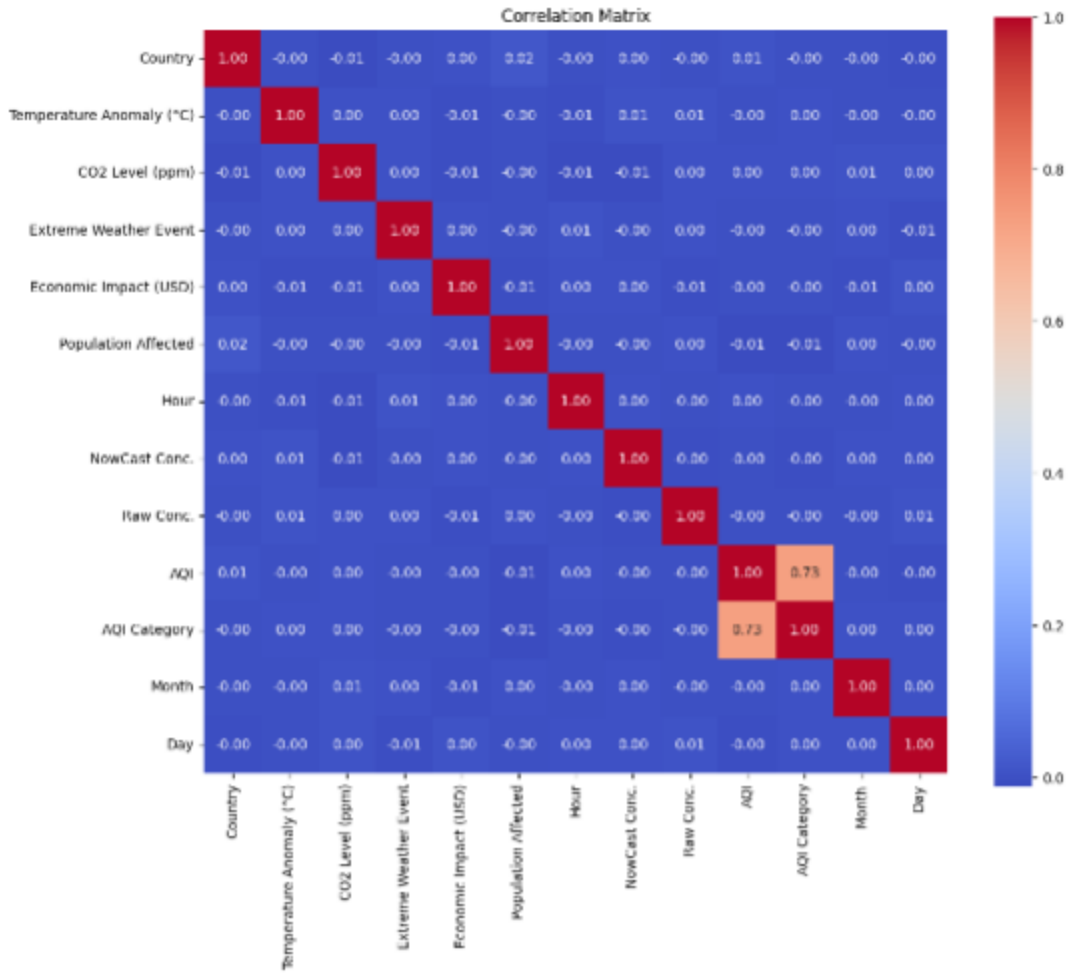

Fig 3.17: Correlation matrix analysis.

3.5.7 Model Evaluation

Several machine learning techniques, including Random Forest, XG Boost, SVM, Decision Tree, and Naive Bayes, were investigated. We make use of our models' capacity to represent context, patterns, and linguistic variances across a range of models. This thorough analysis aids us in achieving our objective of developing a trustworthy categorization scheme for AQI climate change.

1. **Random Forest:** The tree-based approach of the RF Separator may be used for the two categorization groups. A hierarchical tree is produced by an ML algorithm. Artificial intelligence makes hierarchical "decision trees" using this technique. Using a combination mechanism, the randomized forest classification approach constructs many decision trees before combining them together. This clarifies the problems with

overfitting. Because of its versatility and potential to be applied anyplace there is a large amount of data, machine learning is currently one of the most popular topics in business. Machine learning-based RF encoders are typically preferred over other methods because of their numerous advantages. When the method was first created in 1997, it was intended to function with extremely large datasets.

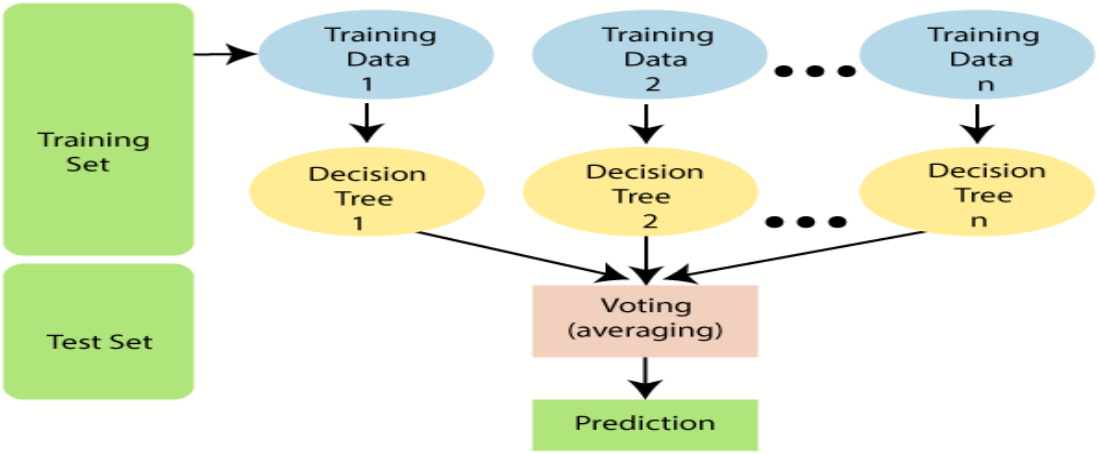

Fig 3.18: The process of the Random Forest classifier.

2. Naïve Bayes: Because the naïve Bayes approach can handle vast volumes of data, it is advised to utilize it for machine learning even when there are many records. Natural language processing (NLP) activities, such AQI classification in texts, are among its strong qualities. The filtering process is quick and simple to use. The Bayes theorem must be fully understood in order to fully comprehend the naive Bayes classifier. In the first exchange, we discuss the Bayes theory. The concept of conditional probability serves as the foundation for this argument. The likelihood that one event will occur given the potential of another is known as contingency probability. To ascertain the possibility of an event, we can apply the conditional probability and our past knowledge.

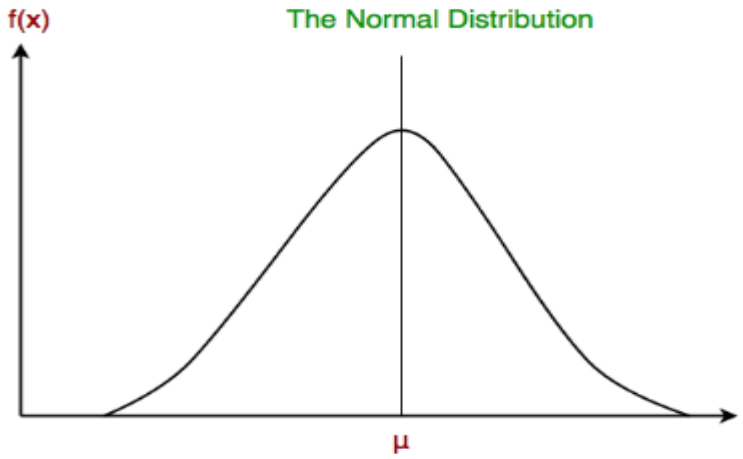

Fig 3.19: Naive Bayes models.

3. XG Boost: Extreme Gradients Boost, often known as "XG Boost," is a popular and effective machine learning method that may be applied to clustering issues like regression. This method may be used to generate slope-enhanced decision-tree structures more efficiently and effectively. In order to improve a model, XG Boost is an ensemble learning technique that aggregates the predictions of several weak learners, usually decision trees. By learning from their less fortunate counterparts, each failing student gradually atones for past errors. Several normalizing penalty strategies are integrated into XG Boost to prevent overfitting. XG Boost may be used to recognize and train structures on unexpected input. To be clear, everything is assembled and working straight out of the box.

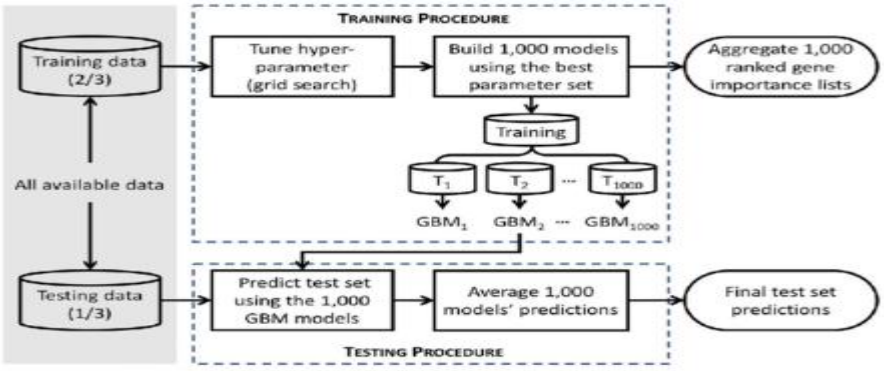

Fig 3.20: XG Boost model construction.

4. **SVM**: Data collection and preparation, including standardization, are the first steps in building a Support Vector Machine (SVM) classifier. The data is then separated into training and testing sets. In order to determine the optimal hyperplane that separates classes during training, SVM uses support vectors to optimize the margin between them. When the data cannot be separated linearly, SVM employs kernel functions to transfer the data into a space with additional dimensions for better separation. Following training, the model's performance on the testing set is evaluated using metrics like accuracy and F1-score. Following verification, the model may be used to predict new data, and the results can be improved by adjusting the hyperparameters.

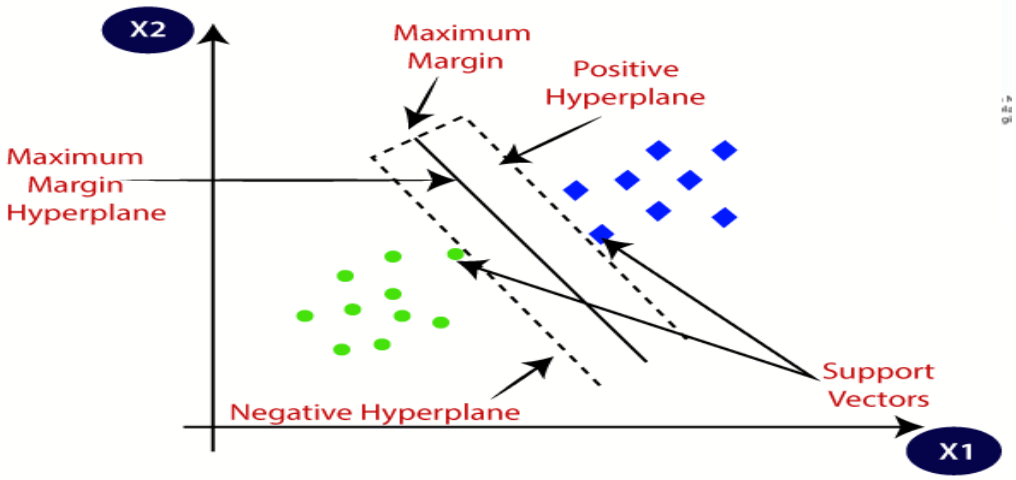

Fig 3.21: A demonstration of SVC

5. **Decision Tree:** Workflow for the Decision Tree classifier begins with data collection and preparation, which includes handling missing values and encoding categorical variables. From the dataset, subsets are then produced for training and testing. A tree-like model is created during the training phase when the algorithm iteratively divides the data according to feature values, producing branches that eventually lead to make-a-choice nodes and leaf nodes, which stand in for class labels. The goal at each split is to either maximize information acquisition or decrease impurity. The effectiveness of the tree is assessed once it has been constructed using the testing data by computing performance metrics like as accuracy, precision, and recall. Pruning techniques can be used to avoid overfitting by eliminating branches that have little predictive value.

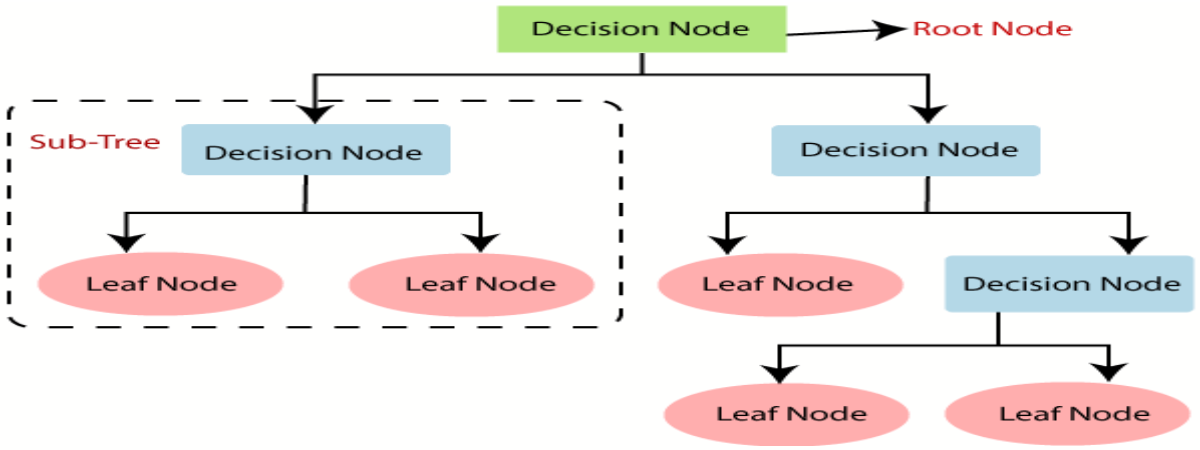

Fig 3.22: Workflow of Decision Tree

3.6 Implementation Requirements

Details must be obtained after the prior stages are completed in order to guarantee correctness. The main concept of our project needed 8 components to be finished. We must finish all of these tasks if we are to meet our objective.

- Data collection.
- Preparation of the data.
- Preprocessing data.
- Getting rid of null values and duplicates in the data.
- How tokenization works.
- Using models is recommended for all five strategies.
- Using SMOTE techniques and random sampling;
- Analyzing and discussing the results.

For the notion to work, we had to begin writing its code. We assessed the precision of five distinct approaches. After the procedure was finished, we evaluated its accuracy. We determined that the above-mentioned design would be more suitable for our needs after assessing the accuracy. After a careful analysis of the pertinent theoretical and numerical methodologies and concepts, a set of requirements for each attempt at classifying the quality of air pollution has been established. The following are some of the possibly notable outcomes:

1. Hardware and Software Requirements

- Operating Systems: Windows 7 or later;
- Hard Drive: 1 TB or more;
- RAM: 4 GB or less

2. Tool Development

- Environment of python
- PyCharm.
- Google Colab.

CHAPTER 4

Experiment Results and Discussion

4.1 Introduction

The process of allocating an AQI Category, a measure used to gauge the quality of air pollution, as well as an examination of climate change, are covered in this part. Selecting a model, collecting and assessing data, The experiment's findings are illustrated here, and the section that follows provides an analysis of them. Selecting a model, collecting and analyzing data, removing null values, adding more columns to the text, and assessing the model's performance in light of the results regarding drug addiction level assessment were the stages involved in constructing a model. This section of the article presents and analyzes the experiment's results. These visualizations show the RAM, precision, F1 score, supports, heat maps, and other information.

4.2 Experimental Setup

The hardware, software, and data resources used in this study's setup—"Public Health Approaches To Addressing Climate Change Health Impacts**.**"—have been thoughtfully planned to enable thorough investigation. With a strong CPU and powerful GPUs that function as a computational core, computers that perform well are computing devices. I am aware that there will never be a flawless invention. Similarly, we may change the parameters of our model to improve accuracy while training. But using a range of techniques, we find that the dependability level appears to be high. This provides an aesthetically pleasing overview of everything that we are now engaged in. Data including heat maps, F1 score, recall, accuracy, and supports are displayed in these infographics. Arrays, material classification methods, and older Python modules were used. One of the data collection's notable aspects is the machine learning assessment of a dataset made up of internet data with thirteen distinct feature categories. In this work, AQI was utilized to analyze climate change using machine learning. I made use of Python libraries, content classification strategies, and pre-existing dictionaries. The deep learning method analyzes the several kinds of meteorological events, including the air quality index, which makes the dataset pertinent. This study uses Python machine learning (ML) models for classification, which is another use of machine learning to identify the correct types of aqi categories.

Accuracy: A proportion of the total quantity of responses to all the samples was used to compute the accuracy assessing classification, which serves as a general indicator of the model's efficacy.

Confusion matrix: Each class's occurrences—false positives, false negatives, real positives, and true negatives—were identified and forecasted by using the confusion matrix to examine the model's behavior. This matrix presented the study directions and possible obstacles in the investigation of drug addiction.

4.3 Experimental Results and Discussion

The method we employed determined the various outcomes we obtained. We were able to accurately calculate the different countries' climate change from 2000 to 2023 with the use of five machine learning algorithms. The relative efficacy of each component of the overall structure was assessed using these methodologies, and the outcome of the forecast was then ascertained by a series of verification procedures. After the whole set of data was chosen, each model utilized a single file containing information from both publicly available web sources and our own study. Following the completion of the data gathering process, we used MATLAB and related pre-configured libraries to assess the algorithms' output. Using a similar dataset, the second step determines if the substance qualifies for the air quality index (AQI). These include, among others, Unhealthy, Good, Moderate, Very Unhealthy, Hazardous, and Unhealthy for Sensitive Groups. Here, we thoroughly examined a number of models using relevant performance criteria. Metrics including recall, accuracy, precision, and total F1 score offer a thorough understanding of the methods' effectiveness. For accuracy, we use the SMOTE random splitting method, random splitting, and without splitting.

Table 4.1: Accuracy after random under-sampling

Classifier	Accuracy Score (AUC)	Precision	Recall	F1-Score
Decision Tree	100%	100%	100%	100%
Random Forest	92.14%	95.48%	92.14%	92.77%
SVC	92.29%	95.53%	92.29%	92.90%

Naïve Bayes	98.19%	98.45%	98.19%	98.25%
XG Boost	63.60%	70.82%	63.59%	60.56%

Table 4.2: Accuracy after SMOTE over-sampling

Classifier	**Accuracy Score (AUC)**	**Precision**	**Recall**	**F1-Score**
Decision Tree	100%	100%	100%	100%
Random Forest	98.63%	98.68%	98.63%	98.63%
SVC	98.64%	98.68%	98.63%	98.63%
Naïve Bayes	98.19%	98.45%	98.19%	98.25%
XG Boost	64.41%	71.19%	64.41%	61.34%

Table 4.3: Table of Accuracy without splitting

Classifier	**Accuracy Score (AUC)**	**Precision**	**Recall**	**F1-Score**
Decision Tree	95.62%	96.06%	95.62%	95.56%
Random Forest	99.72%	99.73%	99.72%	99.73%
SVC	94.14%	94.89%	94.14%	94.07%
Naïve Bayes	97.93%	98.03%	97.93%	97.85%
XG Boost	97.94%	98.00%	97.95%	97.86%

The next section shows the efficacy for a number of models. Two open-source programs, PyCharm and CoLab, were used in the procedure. There were five models in total: SVC, XG Boost, Random Forest, Decision Tree, and Naive Bayes. With an accuracy of 99.72%, the Random Forest models were the most accurate.

	precision	recall	f1-score	support
Good	0.98	1.00	0.99	1019
Hazardous	1.00	0.99	1.00	3804
Moderate	1.00	1.00	1.00	969
Unhealthy	1.00	1.00	1.00	904
Unhealthy for Sensitive Groups	1.00	1.00	1.00	944
Very Unhealthy	1.00	1.00	1.00	1960
accuracy			1.00	9600
macro avg	1.00	1.00	1.00	9600
weighted avg	1.00	1.00	1.00	9600

Fig 4.1: Classification Report of Random forest's.

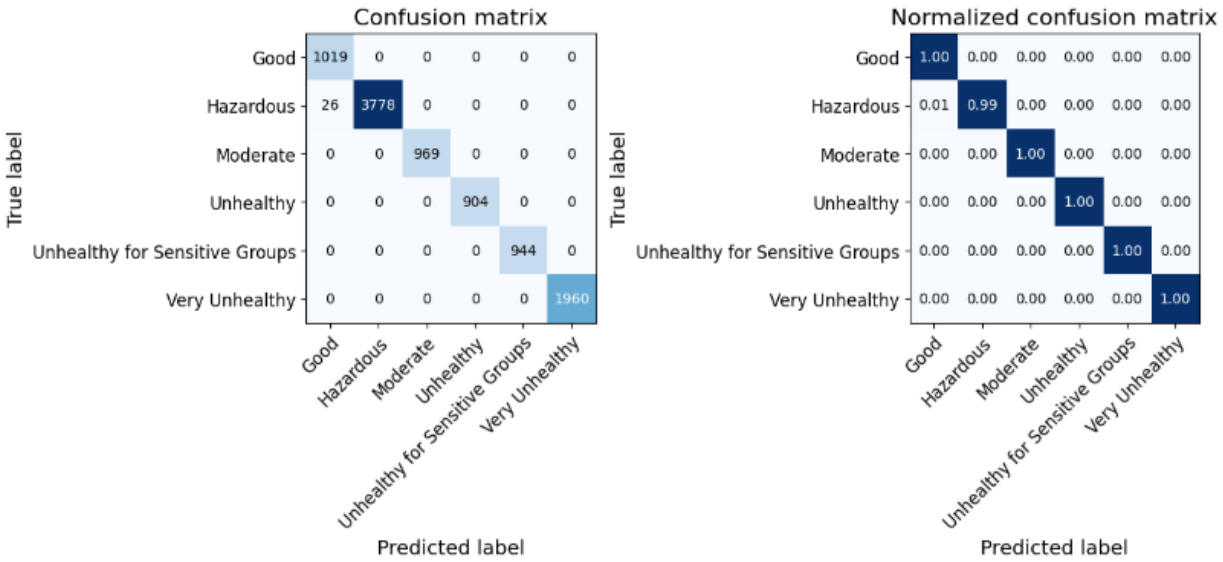

Fig 4.2: Confusion matrix & Normalization OF RF.

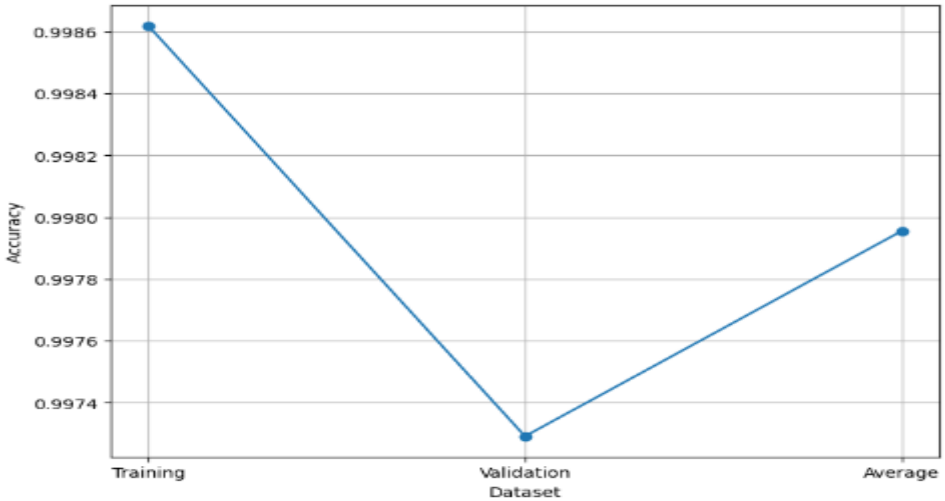

Fig 4.3: RF (Random Forest) Training and Validation Accuracy.

Simply displaying the entire classification result for Random Forest models without sampling yields the highest accuracy.

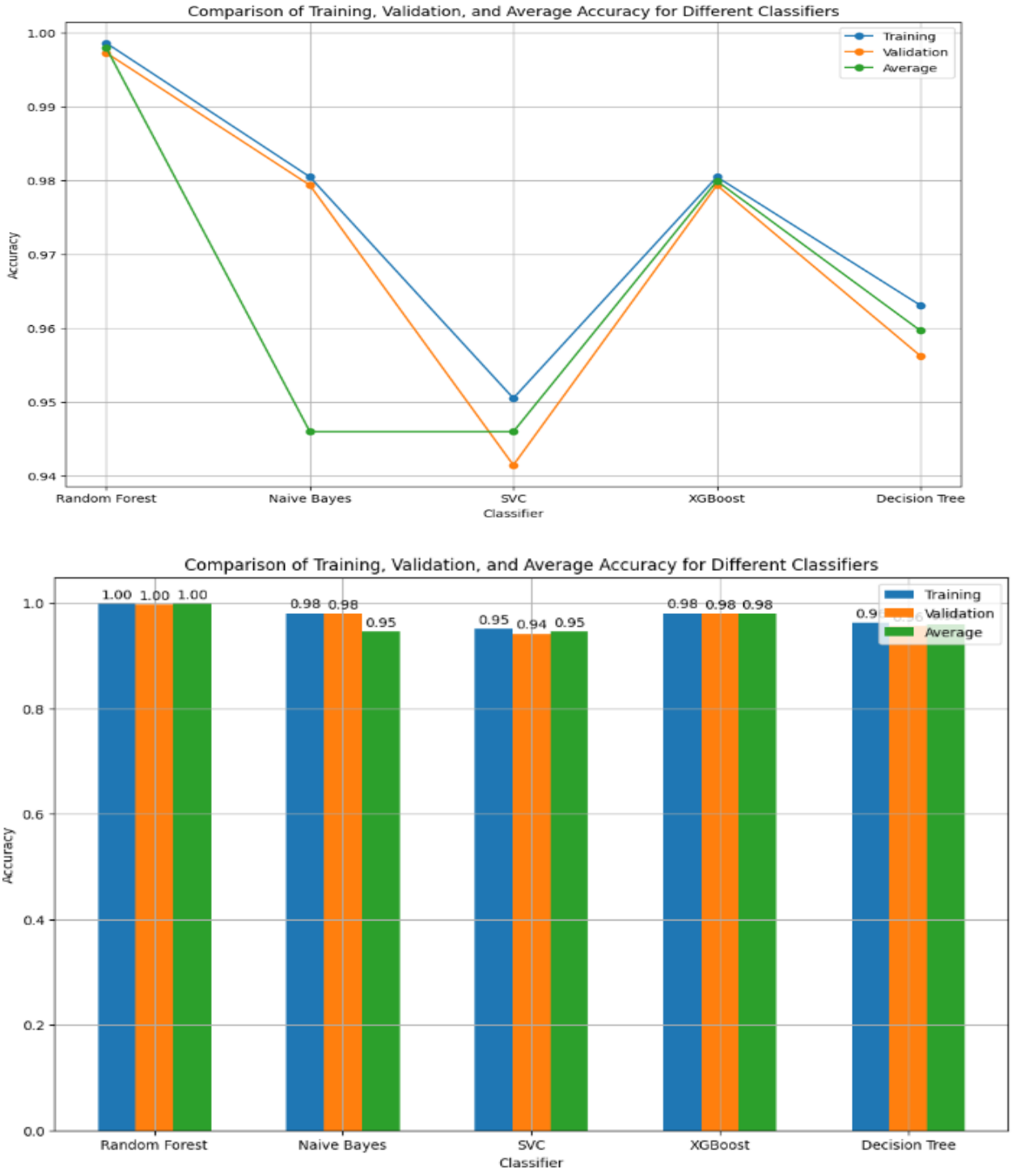

Fig 4.5: Comparison of training, validation, and average accuracy for five classifiers

4.4 Discussion

In this investigation, five different algorithms were used. All things considered, these findings provide us a better understanding of how various machine learning models forecast climate change and provide valuable new data to researchers and practitioners working in the domains of environmental science and climate science.They needed a lot of things to find before we could start working. Yes, we did select the algorithm and start developing it. The accuracy of each technique was then evaluated. Using random split as an approximate estimate for all five models, the random forest classifier employed methods to achieve an ultimate accuracy of 99.72%.

Precision: One popular criterion to evaluate the model's effectiveness is the accuracy with which the algorithm produces forecasts. The total number of true positives multiplied by the total number of accurate forecasts can be used to determine efficiency [13].

$$\text{Precision} = \frac{TP}{TP+FP}$$

Recall: Retrieval is the percentage of suitable cases that were ultimately found and retrieved, regardless of all relevant cases. A high recall rate indicates that the strategy yielded the most pertinent results [14].

$$\text{Recall} = \frac{TP}{TP+FN}$$

F1-Score: The validity of a test is assessed based on its recall and accuracy. Recall and accuracy function best together [15].

$$\text{F1} = \frac{2 \times precision \times recall}{precision + recall}$$

Accuracy: It is feasible to evaluate the efficacy of the model by combining recall and dependability, as the F1 score does. You may compute it using the following equation [16]:

$$\text{Accuracy} = \frac{TP+TN}{TP + FN + TN + FP}$$

CHAPTER 5

Impact on Society, Environment and Sustainability

5.1 Impact on Society

The study contributes to a better understanding of climate change by providing insightful data on future temperature trends. It serves as a scientific resource for educators, policymakers, and everyday residents, promoting climate change awareness and assisting in informed decision-making. Communities may use the projections to develop effective plans to mitigate and adapt to adverse effects. Decision-makers may reduce the adverse consequences of climate change on society by understanding the predicted temperature changes and developing plans, such as targets to reduce emissions, transitions to sustainable energy sources, and city development initiatives. There are significant effects of climate change on the general public's well-being. It can also assist authorities and healthcare professionals in preparing for and managing climate-related health hazards. It makes it possible to identify vulnerable regions and facilitates the implementation of measures to protect people from severe heat events, infectious diseases, and other health problems brought on by climate change.

5.2 Impact on Environment

The diversity of living things and ecosystems are greatly impacted by temperature fluctuations. The article's predictions can aid in understanding potential alterations to habitats, patterns of species dispersal, and the functioning of ecosystems. This data may be used to inform ecosystem management plans, nature preservation initiatives, and the identification of regions in need of renewal or conservation. Temperature variations can affect the water flow and the quantity of water available. By predicting shifts in rainfall patterns, glacier melting rates, and regions experiencing water scarcity, the research findings help guide the management of water resources. This information may be used to develop sustainable water management strategies, such as infrastructure development and water conservation. An increase in extreme weather events is associated with global warming, including severe heat waves, dry spells, and cyclones. Understanding the imminence, intensity, and dissemination of these catastrophes is made easier by the predictions. Policymakers and communities may use this knowledge to implement

measures that will reduce vulnerability, increase adaptation, and improve preparedness for disasters.

5.3 Ethical Aspects

There are possible advantages and moral dilemmas when using machine learning (ML) to combat climate change. On the plus side, machine learning (ML) can power data-driven approaches to climate change, including forecasting natural disasters, optimizing renewable energy systems, and evaluating climate data to provide policymakers insights. But there are important ethical ramifications. If not handled with green methods, machine learning algorithms' high computing demands can result in enormous energy consumption and emissions. Furthermore, policy suggestions resulting from machine learning models educated on biased data may be unjust or unequal, possibly disadvantageous to disadvantaged populations. Due to the possibility of recording location-based or personal data, privacy concerns can surface when applying machine learning for climate monitoring. In order to reduce these risks, ML systems should incorporate transparency, environmentally friendly modeling, and ethical concerns to make sure they promote equitable, sustainable, and inclusive climate change mitigation.

5.4 Sustainability Plan

A machine learning (ML)-based climate change sustainability strategy incorporates technology to minimize carbon footprints and maximize environmental results. Setting specific objectives, such reducing emissions, increasing energy efficiency, and strengthening climate resilience, is the first step in the strategy. More precise forecasts of weather occurrences and resource requirements are possible through the use of machine learning (ML) to examine massive datasets and simulate climate trends. The strategy is to lower total energy usage and increase the use of renewable sources using machine learning (ML)-driven optimizations in energy networks, transportation, and industrial processes. By making sure that machine learning applications are created with equity and inclusion, eliminating prejudices that can have an uneven effect on marginalized populations, ethical behavior is given top priority. In order to reduce the negative environmental effects of computation needs, the strategy calls for investment in energy-efficient computer resources and effective machine learning techniques. Emissions, water, and energy savings will be tracked through regular monitoring using key performance indicators (KPIs), and openness will be ensured through reporting procedures. This strategy promotes social and

environmental sustainability in addition to technological growth by positioning machine learning as a tool for aggressive climate action.

CHAPTER 6
Conclusion and Future Research

6.1 Summary of the Study

This work makes detailed predictions on climate change using a range of machine learning models, including SVC, XG Boost, Random Forest, Decision Tree, and Naive Bayes. We have learned a great deal about the problem from our inquiry. Predictions of air quality remain a contentious issue. According to the air quality index, this has a major impact on the yearly decline in patient usage of programs for detecting air pollution. Additionally, based just on looks, the ML approach has been utilized to evaluate the possibility of filthy air AQI values being transmitted using 3 methods. This study improves our knowledge of climate prediction models and guides future research toward ecologically acceptable climate solutions.

6.2 Conclusion

Predicting the weather and climate has always been essential for efficient agricultural planning, protecting against natural calamities, and promoting strategic decision-making across a range of industries. Forecasting that is both accurate and timely is crucial in this situation. Weather and climate prediction might become more accurate and faster using machine learning. This work use machine learning approaches, notably advanced learning algorithms, to anticipate future climatic changes using online data from 2000 to 2023. This dataset, which includes 48,000 data points and 13 features, is what I get from Kaggle. Frequent air pollution forecasts and monitoring are necessary to maintain interior air quality. Analyze all of the associated characteristics using all 13 features like: 'Date', 'Country', 'Temperature Anomaly (°C)', 'CO2 Level (ppm)', 'Extreme Weather Event', 'Economic Impact (USD)', 'Population Affected', 'Hour', 'NowCast Conc.' Etc. We employ support vector machines (SVM), decision trees, random forests, naïve Bayes, and XG Boost. The accuracy of the random forest classifier was 99.72%, which was higher than the accuracy of the other five algorithms in our experiment when three methods were used: random under-sampling, random without sampling, and SMOTE approaches.

6.3 Implication for Further Study

Our lives are now quicker and easier thanks to technology and contemporary science. We wish to continue using information technology and the internet in our nation by incorporating our model into software, a web application, or an Android application in the future. With a larger database, we will be able to improve our model's accuracy in the future. Furthermore, the model's software may be made accessible to the public by developing graphical user interfaces. The model may be improved in the future by introducing new features, altering its settings, and adding new algorithms. By gathering information from various groups of individuals based on the district, a strong database may be produced in the future.

Reference

[1] C. Abbe, "The physical basis of long-range weather forecasts," Mon. Weath. Rev., 29 (1901), pp. 551–561

[2] Y. Zheng, L. Capra, O. Wolfson, H. Yang, "Urban computing: concepts, methodologies, and applications,"ACM Transactions on Intelligent Systems and Technology (TIST), 5, no. 3 (2014), pp. 1–55.

[3] T. Gneiting, A.E. Raftery, "Weather forecasting with ensemble methods," Science, 310, no. 5746 (2005), pp.248–249

[4] H. (n.d.). Shaftel," Overview: Weather, Global Warming and Climate Change. Climate Change: Vital Signs of the Planet", https://climate.nasa.gov/global-warming-vs-climate-change.

[5] N. Milojevic-Dupont, & F. Creutzig, "Machine learning for geographically differentiated climate change mitigation in urban areas' Sustainable Cities and Society, 64, 102526, (2021). https://doi.org/10.1016/j.scs.2020.102526

[6] B. Bochenek, & Z. Ustrnul, "Machine Learning in Weather Prediction and Climate Analyses—Applications and Perspectives. Atmosphere," 13(2), (2022). 180. https://doi.org/10.3390/atmos13020180

[7] T. Vishvakarma, "Climate Change Forecasting using Machine Learning Algorithms." International Journal for Research in Applied Science and Engineering Technology, 2022, 10(10), 881–888. https://doi.org/10.22214/ijraset.2022.47098

[8] R. Krishnan, "Progress and Prospects in Weather and Climate Modelling", 2020, arXiv.org. https://arxiv.org/abs/2011.11353

[9] P. Bauer, A. J. Thorpe, & G. Brunet, "The quiet revolution of numerical weather prediction", Nature, 525(7567), 2015, 47–55. https://doi.org/10.1038/nature14956

[10] J. Asha, R. S, S. S, & P. Reena, "Analysis of Temperature Prediction Using Random Forest and Facebook Prophet Algorithms", 2020, In Lecture notes on data engineering and communications technologies. Springer International Publishing. https://doi.org/10.1007/978-3-030-38040-3_49

[11] Huntingford, C., Jeffers, E. S., Bonsall, M. B., Christensen, H. M., Lees, T., & Yang, H. (2019). Machine learning and artificial intelligence to aid climate change research and understanding. *Nature Climate Change*, 9(9), 629-637.

[12] Kumar, R., Malhotra, R., & Jain, M. (2020). Artificial intelligence applications in renewable energy: A review. *Materials Today: Proceedings*, 26, 2889-2893.

[13] Rolnick, D., Donti, P. L., Kaack, L. H., Kochanski, K., Lacoste, A., Sankaran, K., ... & Bengio, Y. (2019). Tackling climate change with machine learning. *arXiv preprint arXiv:1906.05433*.

[14] Shakoor, U., Qureshi, R., Shahid, M. N., & Alam, S. (2021). Role of machine learning in climate smart agriculture: A review. *Sustainable Computing: Informatics and Systems*, 30, 100547.

[15] Wu, C., Chen, Y., Peng, C., Li, Z., & Hong, X. (2019). Modeling and estimating aboveground biomass of Dacrydium pierrei in China using machine learning with climate change. Journal of Environmental Management, 234, 167–179.

[16] Rolnick, D., Donti, P. L., Kaack, L. H., Kochanski, K., Lacoste, A., Sankaran, K., Ross, A., MilojevicDupont, N., Jaques, N., Waldman-Brown, A., Luccioni, A., Maharaj, T., Sherwin, E. D., Mukkavilli, S. K., Kording, K. P., Gomes, C. P., Ng, A. Y., Hassabis, D., Platt, J., . . . Bengio, Y. (2023). Tackling Climate Change with Machine Learning. ACM Computing Surveys, 55(2), 1–96.

[17] Mansfield, L., Nowack, P., Kasoar, M., Everitt, R. G., Collins, W. J., & Voulgarakis, A. (2020). Predicting global patterns of long-term climate change from short-term simulations using machine learning. Npj Climate and Atmospheric Science, 3(1).

[18] Huntingford, C., Jeffers, E. S., Bonsall, M. B., Christensen, H., Lees, T., & Yang, H. (2019). Machine learning and artificial intelligence to aid climate change research and preparedness. Environmental Research Letters, 14(12), 124007.

[19] Crane-Droesch, A. (2018). Machine learning methods for crop yield prediction and climate change impact assessment in agriculture. Environmental Research Letters, 13(11), 114003.

[20] Bochenek, B., & Ustrnul, Z. (2022). Machine Learning in Weather Prediction and Climate Analyses— Applications and Perspectives. Atmosphere, 13(2), 180.

[21] Kadkhodazadeh, M., Anaraki, M. V., Morshed-Bozorgdel, A., & Farzin, S. (2022). A New Methodology for Reference Evapotranspiration Prediction and Uncertainty Analysis under Climate Change Conditions Based on Machine Learning, Multi Criteria Decision Making and Monte Carlo Methods. Sustainability, 14(5), 2601.

[22] Malakouti, S. M. (2023). Utilizing time series data from 1961 to 2019 recorded around the world and machine learning to create a Global Temperature Change Prediction Model. Case Studies in Chemical and Environmental Engineering, 7, 100312.

[23] Milojevic-Dupont, N., & Creutzig, F. (2021). Machine learning for geographically differentiated climate change mitigation in urban areas. Sustainable Cities and Society, 64, 102526. https://doi.org/10.1016/j.scs.2020.102526

Smrity

ORIGINALITY REPORT

15%	11%	7%	11%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	dspace.daffodilvarsity.edu.bd:8080 Internet Source	6%
2	Submitted to BRAC University Student Paper	4%
3	Sachi Nandan Mohanty, Bibhuti Bhusan Dash, G. Shanmugasundar, Johar MGM, Inakollu Aswani, Ajith Sundaram, Issac K. Varghese. "UAV Databased Temperature Patterns Analysis with Carbon Emission Detection Using Deep Neural Network", Remote Sensing in Earth Systems Sciences, 2024 Publication	1%
4	Saravanan Parthasarathy, Vaishnavi Jayaraman, Viji Vijayan, Aaswin Raja. "Stack-ClimaBoost: A Model for Analysing the Patterns of Global Warming Across the Continents", International Journal of Sustainable Development and Planning, 2024 Publication	1%
5	Submitted to University of Southern Mississippi	1%

Student Paper

6	Tan Yigitcanlar. "Urban Artificial Intelligence - A Guidebook for Understanding Concepts and Technologies", CRC Press, 2024 Publication	1 %
7	Submitted to University of Colorado, Denver Student Paper	1 %
8	Submitted to Jacksonville University Student Paper	1 %

Exclude quotes	On	Exclude matches	< 1%
Exclude bibliography	Off		