EXPLORING THE RELATIONSHIP BETWEEN DENGUE AND ENVIRONMENTAL FACTORS IN DHAKA USING MACHINE LEARNING

BY

MD. ASIK REZA TASIN
ID: 201-15-14081

FINAL YEAR DESIGN PROJECT REPORT

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

Supervised By

MS. NAZNIN SULTANA
Associate Professor
Department of CSE
Daffodil International University

Co-Supervised By

MR. MD. AYNUL HASAN NAHID
Lecturer
Department of CSE
Daffodil International University

DAFFODIL INTERNATIONAL UNIVERSITY

DHAKA, BANGLADESH

July 2024

APPROVAL

This Project titled "Exploring The Relationship Between Dengue And Environmental Factors In Dhaka Using Machine Learning", submitted by Md. Asik Reza Tasin to the Department of Computer Science and Engineering, Daffodil International University, has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of B.Sc. in Computer Science and Engineering and approved as to its style and contents. The presentation has been held on 13 July, 2024.

BOARD OF EXAMINERS

Narayan Ranjan Chakraborty (NRC) — **Board Chairman**
Associate Professor & Associate Head
Department of CSE
Faculty of Science & Information Technology
Daffodil International University

Md. Sadekur Rahman (SR) — **Internal Examiner 1**
Assistant Professor
Department of CSE
Faculty of Science & Information Technology
Daffodil International University

Mr. Tanvirul Islam (TI) — **Internal Examiner 2**
Lecturer
Department of CSE
Faculty of Science & Information Technology
Daffodil International University

Dr. Md. Manowarul Islam — **External Examiner**
Associate Professor
Department of Computer Science and Engineering
Jagannath University

DECLARATION

We hereby declare that this project has been done by us under the supervision of **Ms. Naznin Sultana, Associate Professor, and Department of CSE, Daffodil International University**. We also declare that neither this project nor any part of this project has been submitted elsewhere for the award of any degree or diploma.

Supervised by:

Ms. Naznin Sultana
Associate Professor
Department of CSE
Daffodil International University

Co-Supervised by:

Mr. Md. Aynul Hasan Nahid
Lecturer
Department of CSE
Daffodil International University

Submitted by:

13.07.24

Md. Asik Reza Tasin
ID: 201-15-14081
Department of CSE
Daffodil International University

ACKNOWLEDGEMENT

First, I express our heartiest thanks and gratefulness to almighty for His divine blessing making it possible for us to complete the final year project/internship successfully.

I am grateful and wish our profound indebtedness to **Ms. Naznin Sultana**, **Associate Professor**, Department of CSE Daffodil International University, Dhaka. Deep Knowledge & keen interest of our supervisor in the field of "*Machine Learning*" to carry out this project. His endless patience, scholarly guidance, continual encouragement, constant and energetic supervision, constructive criticism, valuable advice, reading many inferior drafts, and correcting them at all stages have made it possible to complete this project.

I would like to express our heartiest gratitude to the **Dr. Sheak Rashed Haider Noori, Professor and Head**, **Department of CSE**, for his kind help in finishing our project and also to other faculty members and the staff of the Department of CSE, Daffodil International University.

I would like to thank our entire course mate in Daffodil International University, who took part in this discussion while completing the course work.

Finally, we must acknowledge with due respect the constant support and patience of our parents.

ABSTRACT

My thesis explores how environmental factors like temperature, humidity and rainfall contribute to the spread of dengue fever in Dhaka of Bangladesh. Rapid urbanization and a dense population have created ideal conditions for mosquito breeding. It's particularly the Aedes species responsible for dengue transmission. My research aims to understand the relationship between these environmental factors and dengue outbreaks to develop better prevention and control strategies. I examined how weather patterns and urban planning deficiencies impact dengue cases by analyzing data from 2009 to 2023. It is using classification model like linear regression, Logistic regression, Naive Bayes, Decision Tree, SVM, XGBoost, Random Forest, K-Nearest Neighbors. My findings indicate that certain weather conditions during the monsoon season. It's significantly influence the incidence of dengue. The study highlights the need for improved waste and water management. Its public awareness campaigns and targeted interventions to reduce mosquito breeding sites. My research aims to support public health efforts in mitigating dengue outbreaks and enhancing community resilience in Dhaka by providing practical recommendations. My study shows the need for improved waste and water management in public awareness campaigns and targeted interventions to reduce mosquito breeding sites. Poorly maintained urban areas with stagnant water are hotspots for mosquito breeding. My research aims to support public health efforts in mitigating dengue outbreaks and enhancing community resilience in Dhaka by providing practical recommendations. My research suggests that community involvement and education are crucial in controlling the spread of dengue. It's ensuring that residents are aware of prevention methods can significantly reduce the incidence of the disease. This comprehensive approach can help protect public health and prevent future outbreaks.

TABLE OF CONTENTS

Contents **Page No**

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1
INTRODUCTION

1.1 Overview

Dengue fever, transmitted by mosquitoes, is a pressing health concern in tropical and subtropical regions around the world. Dhaka, the vibrant capital of Bangladesh, is grappling with a severe challenge. The city's swift urbanization and dense population have created a perfect environment for mosquitoes to thrive and spread dengue. Globally, the incidence of dengue fever is mostly found in 128 countries, including Southeast Asia, Eastern Mediterranean, and the Western Pacific. Asian countries represent 70% of the global burden of dengue disease. [1] These mosquitoes Aedes aegypti and Aedes albopictus are thrive in stagnant water. It is common in Dhaka's informal settlements and poorly maintained areas. Dengue can cause anything from a mild fever to severe conditions that may lead to hospitalization or death. The impact of dengue in Dhaka has been severe with significant outbreaks over the last two decades. This shows problems in public health management and urban planning. The increase in dengue cases suggests current strategies are not effective. It is important to explore the environmental factors contributing to dengue transmission to develop better prevention and control measures. The affect mosquito lifecycles and virus spread like temperature, humidity and rainfall. Dhaka's tropical monsoon climate with a long wet season and a moderate winter. It also contributes to the spread of dengue. These climatic conditions facilitate year-round breeding of Aedes mosquitoes. It’s particularly during the monsoon months when water accumulation is most prevalent. The city is rapid urban growth has led to uneven urban planning and resulting in irregular water drainage also garbage disposal systems. These factors create numerous habitats conducive to mosquito breeding. The interaction between ecological, social, and human mobility factors further complicates the relationship between environmental conditions and dengue outbreaks. High human density areas, such as marketplaces, schools, and residential blocks, are particularly vulnerable. The virus is transmitted by Aedes aegypti and Aedes albopictus mosquitoes. The vectors breed in small bodies of stagnant water, particularly in water storage containers around homes. [2]. It’s particularly in water storage containers around homes the movement of people within these areas facilitates the swift transmission of the virus across the population. It’s complicating containment efforts. Also global climate change continues to alter local weather patterns. It’s potentially leading to more severe and unpredictable dengue outbreaks. This thesis aims to dissect the complex interplay between environmental factors and the epidemiology of dengue in Dhaka. By identifying specific environmental triggers and understanding their interactions with urban dynamics. This study seeks to provide a foundation for public health authorities to formulate more precise strategies aimed at reducing the incidence of dengue. Through a comprehensive analysis of the city's environmental landscape and its correlation with dengue case data. The study will explore avenues for mitigating risk and enhancing community resilience against this persistent public health threat.

1.2 Background and Present State

Dengue fever is a significant concern in Dhaka. Dhaka city is rapid urban expansion has frequently led to inadequate waste and water management. It's creating optimal conditions for the breeding of mosquitoes that transmit this disease. These environmental factors are compounded by the city's dense population. It's making the spread of dengue more rapid and difficult to control. The challenges associated with preventing dengue outbreaks become more complex and demanding. This study is driven by the need to thoroughly understand how Dhaka’s urban environment contributes to the proliferation of dengue-carrying mosquitoes the interplay between urban planning deficiencies and mosquito breeding habitats by investigating. This thesis aims to shed light on effective prevention and control strategies that can be implemented. The goal is to provide practical insights and recommendations that could help mitigate the risk of dengue in Dhaka. It's improving public health and safety for its residents. This research is crucial for developing targeted interventions that address the specific conditions of Dhaka. It's ultimately helping to curb the spread of dengue and protect the community from future outbreaks.

1.3 Problem Statement

Dengue fever is a major public health challenge in Dhaka. It worsened by the city's rapid but poorly managed urban development. This study focuses on the environmental factors within Dhaka that promote the breeding and spread of Aedes mosquitoes the carriers of dengue. These factors create ideal conditions for year-round mosquito breeding and rapid virus transmission. The study seeks to identify and analyze the specific urban and climatic conditions contributing to the dengue problem. This will help in developing targeted prevention strategies. It's aiming to improve public health interventions and potentially guide policy changes for better urban management. This study seeks to identify and analyze.

1.4 Objectives

The overview highlights the serious problem of dengue fever in Dhaka by mosquito bites. The Dhaka city’s rapid growth and dense population create a breeding ground for mosquitoes in areas with stagnant water. Dengue fever can range from mild to severe. It's leading to hospitalization or death. The frequent outbreaks over the last two decades show that current public health and urban planning strategies are not working well. Environmental factors like temperature, humidity and rainfall play a significant role in the spread of dengue. Dhaka's long wet season and moderate winter further facilitate mosquito breeding. Rapid urbanization has led to poor water drainage and waste disposal. It's creating more mosquito habitats. High human density areas and human movement make it difficult to control the spread of dengue. Global climate change is also worsening the situation. My thesis aims to study the environmental factors contributing to dengue in Dhaka to help develop better prevention and control measures. The background emphasizes that dengue is a major concern in Dhaka with rapid urban growth leading to poor waste and water management. This creates ideal conditions for mosquito breeding. The dense population makes it difficult to control the spread of dengue. The study seeks to understand how urban planning and environmental factors contribute to dengue to develop effective prevention and control strategies. The goal is to provide practical recommendations to improve public health and safety in Dhaka. The problem statement identifies dengue fever as a significant public health issue in Dhaka worsened by poorly managed urban development. The study focuses on identifying environmental and climatic conditions that promote mosquito breeding and dengue transmission. The aim is to develop targeted prevention strategies to improve public health interventions and guide policy changes for better urban management. The scope and limitations section presents key questions for the study. It seeks to understand how environmental changes influence dengue cases. Time series analysis can predict future outbreaks and what patterns in data can help reduce dengue transmission. The report organization section outlines the expected outcomes of the study. It aims to enhance understanding of dengue's severity and spread over time. Its highlight our challenges faced by the health sector. It has increase public awareness. It's confirmed the link between population density and infection rates and provides an accurate understanding of dengue prevalence for effective public health planning and resource allocation.

1.5 Scope and Limitations

During the research work some question occurs about this work. The main questions of our work in given below:

- ✓ How do changes in environmental factors influence the number of dengue cases in Dhaka?
- ✓ Which classification model performs best in predicting dengue cases based on environmental factors in Dhaka?
- ✓ What patterns in the classification data can help us better understand and reduce dengue transmission in Dhaka?
- ✓ What is the impact of feature selection on the accuracy and reliability of classification models for dengue prediction in Dhaka?
- ✓ How do classification models handle imbalanced datasets in the context of predicting dengue cases and what techniques can improve their performance?

1.6 Report Organization

The report is structured into seven chapters, each detailing a specific aspect of the study. Below is an outline of what each chapter covers:

- **Introduction:** This chapter introduces the topic, explains the background, defines the problem and sets the objectives. It outlines the report's structure and summarizes the key points.

- **Literature review:** This chapter reviews existing research related to the topic. It compares different studies, identifies gaps and summarizes the main points.

- **Methodology:** This chapter describes the approach and methods used in the study. It includes the proposed design, required tools, project planning and a summary.

- **Implementation:** This chapter covers the implementation phase, including model training, prototype design and system testing. It concludes with a summary of the process and findings.

- **Result and analysis:** This chapter presents the experimental results and analyzes the system's performance. It provides a summary of the key findings and insights.

- **Impact on society, environment, and sustainability:** This chapter examines the project's effects on society and the environment. It discusses ethical aspects, sustainability plans and summarizes the key points.

- **Conclusion and future work:** This chapter summarizes the entire report, drawing conclusions based on the findings. It suggests areas for future research and potential improvements.

1.7 Summary

Dengue fever is a big problem in Dhaka. Rapid urban growth and a dense population make it easy for mosquitoes to breed. The city's poor waste and water management in along with a long wet season. Its help mosquitoes thrive. Dengue can range from mild to severe even causing death. My thesis study aims to understand how environmental factors like temperature, humidity and rainfall contribute to dengue. It will help create better strategies to prevent and control the disease. The study also looks at how urban planning and high population areas impact the spread of dengue. The goal is to provide practical solutions to improve public health in Dhaka. [3]

CHAPTER 2

LITERATURE REVIEW

2.1 Overview

Dengue fever is a disease spread by mosquitoes. It is common in Dhaka. This happens a lot because of two types of mosquitoes called Aedes aegypti and Aedes albopictus. It's leading to crowded places and not enough good health facilities. Dengue fever (DF) is the most important mosquito-transmitted viral disease causing a large economic and disease burden in many parts of the world. [2] Dengue viruses and their mosquito vectors are sensitive to their environment. Temperature, rainfall and humidity have well-defined roles in the transmission cycle. [4] It makes easy for these mosquitoes to breed. The city often has issues with water not draining properly and trash not being collected, which means water stands still and becomes a perfect place for mosquitoes to lay eggs. The weather in Dhaka also helps mosquitoes breed because it's usually very wet and humid. Dengue cases go up a lot during the rainy season. These constant outbreaks show there's a big need for better disease tracking and efforts to stop mosquitoes from breeding. This study is looking at what things about Dhaka's environment help dengue spread. We can find better ways to fight the disease and keep people healthier by understanding these things,. This background helps us start looking deeper into how and why dengue keeps being a problem in Dhaka, aiming to find better ways to manage the city and protect its people.

2.2 Related Works

Several studies have explored the relationship between climatic factors and the incidence of dengue fever. It's highlighting the importance of environmental considerations in public health strategies.

In "Climate change and the emergence of vector-borne diseases in Bangladesh" by S. McCarty et al. The study investigates the relationship between climate variations and the incidence of dengue fever in Bangladesh. The research found a linear relationship between climatic factors such as temperature and rainfall and the prevalence of dengue. This approach emphasizes the impact of environmental changes on the transmission of dengue and highlights the necessity of incorporating climate considerations into public health strategies.

In "Dengue fever in Saudi Arabia: A review of environmental and population factors impacting emergence and spread" by Kholood K. Altassan et al. The study explores the role of climatic factors, urbanization and population dynamics in the spread of dengue fever in Saudi Arabia. The research highlights that the increase in dengue

Cases linked to environmental factors like temperature and rainfall as well as socio-demographic aspects such as the influx of migrant workers and religious pilgrims. The study utilizes epidemiological data and environmental modeling to understand the unique challenges faced in controlling dengue in Saudi Arabia. It is providing valuable insights for public health planning and disease prevention. [2]

In "Effects of weather factors on dengue fever incidence and implications for interventions in Cambodia" by Youngjo Choi et al, the study explores the relationship between climatic factors and dengue incidence across three provinces in Cambodia. Using monthly data from 1998 to 2012, the researchers employed negative binomial models to analyze the impact of temperature and rainfall on dengue cases. The findings reveal significant associations between mean temperature, cumulative rainfall and dengue incidence with variations observed between different regions. These results underscore the importance of localized early warning systems and tailored intervention strategies to effectively manage dengue outbreaks. [4]

In "Dengue Disease Analysis in Dhaka City Using Machine Learning Techniques" by Md Rakib Hassan, using machine learning algorithms were used to study the link between weather factors and dengue from 2009 to 2019. The research used Linear Regression and Time Series Analysis to analyze weather data and dengue cases. In temperature, rainfall and humidity the study identified trends in dengue outbreaks by examining patterns. This approach provides a strong model for predicting dengue spikes and helps in public health planning.

In "A Comparative Analysis of Dengue Fever Prediction through Machine Learning Approach" by Md. Minhajul Hayat Mim, the study employs various machine learning algorithms to predict dengue fever incidences using raw datasets collected from hospitals. Algorithms such as Logistic Regression, XGBoost, Support Vector Machines (SVM), K-Nearest Neighbors (KNN), AdaBoost, Linear Discriminant and Calibrated models were evaluated. The XGBoost model achieved the highest accuracy at 98.3%, demonstrating exceptional performance. While the SVM model had the lowest accuracy at 96.3%. This comparative analysis highlights the strengths and weaknesses of each algorithm.

In "Dengue: An Analytical Prediction System Using Machine Learning" by Tamanna Akter, the study applies various machine learning algorithms to predict dengue outbreaks using patient data. The research evaluates five algorithms: Logistic Regression, K-Nearest Neighbors, Gaussian Naive Bayes, Decision Tree and AdaBoost. Using data collected from hospitals and the UCI repository. The study found that the AdaBoost algorithm achieved the highest accuracy at 98% and followed by the Decision Tree at 94%. The comprehensive analysis highlights the effectiveness of machine learning techniques in improving the accuracy of dengue fever predictions. It’s aiding in early diagnosis and public health planning.

In "The emergence of dengue in Bangladesh: epidemiology, challenges and future disease risk" by Sharmin S, Viennet E, Glass K and Harley D. The researchers explored the epidemiology and challenges associated with dengue in Bangladesh. They highlighted the increased incidence of dengue and the complex factors contributing to its spread including urbanization. Population density, climate change and ineffective vector control. The study emphasized the need for integrated vector management. It has public health interventions and improved disease surveillance to address the growing dengue burden in Bangladesh. These findings underscore the importance of coordinated efforts to mitigate dengue transmission and improve public health outcomes. [5]

In "Risk factors associated with Dengue incidence in Bandung, Indonesia: a household based case-control study" by Hubullah Fuadzy et al, the researchers investigated the characteristics and risk factors of dengue incidence at the household level in Bandung. They conducted a matched case-control study with 261 dengue cases and 522 controls. The study found that the productive age and low education level of the household head, dirty toilets and unhealthy house status were significant risk factors for dengue incidence. Logistic regression analysis revealed that the productive age of the household head was the most dominant risk factor with an odds ratio of 2.53. The findings emphasize the importance of intensive health promotion and sanitation improvements to reduce dengue transmission in Bandung. [6]

2.3 Comparison between existing works

In recent years, multiple studies have explored various aspects of dengue fever from clinical treatment strategies to the impact of environmental factors. Below is a comparison of some significant work.

Table 2.1 Comparative analysis with previous work

SN	Work Title	Applied Algorithm	Best Accuracy
1.	Influence of environmental factors on dengue fever in Delhi	Statistics analysis, Descriptive analysis, Correlation analysis	97% correlation with rainfall
2.	Dengue Fever Prediction Using Machine Learning Techniques	Decision Tree Classifier, BaggingClassifier, GaussianNB, BernoulliNB, and AdaBoostClassifier	BaggingClassifier 95.87%
3.	Dengue Disease Analysis and Forecasting in Dhaka	Linear Regression, Time Series Analysis, Logistic Regression, Decision Trees, Random Forest, K-Nearest Neighbors, SVM	Linear Regression 73%
4.	The time series seasonal patterns of dengue fever and associated weather variables in Bangkok	The ARIMA Models	Finally, the ARIMA was 0.90, 3.83, 6.49, and 26.45 for the correlation coefficient, MAE, RMSE, and MAPE
5.	Space-time clusters of dengue fever in India	Time Series Analysis	cluster size of 50%

This study has pointed out important factors to wrap up that affect dengue fever and shown how different models can predict where dengue might strike next. The results highlight the need to use science and research to improve how we deal with dengue fever. This work will help those in charge of public health to come up with better plans to prevent dengue in the future hopefully.

2.4 Open Issue

Dengue fever is a big health issue. It affects many people in places Dhaka. The weather and city conditions help mosquitoes flourish. These mosquitoes carry the dengue virus. Every year, lots of people in Dhaka get sick from dengue. It can be very serious. Dengue fever has been a significant health issue in Dhaka from 2009-2023. Favor mosquito breeding. Seasonal rains exacerbate the problem by creating ideal mosquito breeding sites driven by the city's expanding urban environment and warm climate. Understanding these environmental and urban factors is crucial for predicting outbreaks and improving dengue prevention in Dhaka. The problem is Dhaka grows and more buildings go up. It places appear where mosquitoes can live and breed. This means more people could get dengue if nothing is done to stop it. It rains a lot and the water can collect and become a perfect spot for mosquitoes to lay eggs. Knowing how weather, urban development, and environmental elements such as the location of water accumulation can impact the transmission of dengue fever is crucial.

2.4 Summary

I faced major challenges to relate to gathering data while working on my research about dengue fever in Dhaka. At first make sure of how to collect the required information. I consulted my supervisor who advised looking into the some website. I found that some website was not regularly updated and did not have the specific dengue-related information. I needed which delay my research while the website contained a lot of details about diseases such as COVID-19. There was limited information available on dengue. I decided to contact some academics. They offered me a dataset from an earlier investigation that I intended to utilize for my machine learning examination. This data was insufficient for my needs as it lacked any weather data. It is essential for studying dengue trends. In order to resolve this I first went to the website to search for meteorological data. The information there was lacking with many gaps in the records before to 2022 and 2023.

CHAPTER 3

METHODOLOGY

3.1 Overview

My research focuses on understanding the connection between dengue fever incidents and environmental conditions in Dhaka, Bangladesh. It's known for frequent dengue outbreaks in the city of Dhaka. It provides a valuable context for examining how weather and environmental changes influence the spread of this disease. The aim is to discover patterns that could predict when dengue might occur. This research aims to help better prepare for and possibly prevent dengue outbreaks. I use eight types of classification methods:

- Linear Regression
- Logistic Regression
- Naive Bayes
- Decision Tree
- SVM (Support Vector Machine)
- XGBoost
- Random Forest
- KNN (K-Nearest Neighbors)

Linear Regression is used to understand the relationship between environmental factors and dengue cases. This method calculates how much variables like temperature and humidity affect the number of dengue cases. It helps quantify the impact of each environmental factor on the likelihood of dengue outbreaks. Logistic Regression is employed to predict the probability of a dengue outbreak based on environmental factors. It helps determine the odds of an outbreak occurring under specific conditions such as high temperatures or heavy rainfall. Naive Bayes is used to classify the likelihood of dengue outbreaks based on different environmental factors. This method uses probability to predict the occurrence of outbreaks considering factors like temperature and rainfall. Decision Tree is another method utilized to explore the relationship between environmental factors and dengue outbreaks. This method involves creating a tree-like model of decisions. It helps to visually and explicitly represent decisions and their possible consequences. The Decision Tree method segments the dataset into subsets based on different environmental conditions. It's making it easier to identify specific conditions that lead to higher chances of dengue outbreaks. It is particularly useful for understanding complex interactions between multiple variables and for making predictions that are easy to interpret. SVM (Support Vector Machine) is employed to classify and predict the occurrence of dengue outbreaks based on environmental factors. SVM works by finding the hyper plane that best separates the data into different classes in this case. It has conditions that lead to dengue outbreaks versus those that do not. This method is highly effective in handling high-dimensional data and can manage both linear and non-linear relationships between variables. SVM helps improve the accuracy of predictions by considering the margin of separation between different classes. It's providing a robust tool for outbreak prediction. XGBoost is used for its efficiency and performance in handling large datasets and complex models. It is an ensemble learning method that combines the predictions of multiple models to improve accuracy. XGBoost is particularly effective in managing the interactions between multiple variables and can handle missing data effectively. Random Forest constructs multiple decision trees during training and outputs the mode of the classes for classification. It's improving predictive accuracy and controlling over fitting by averaging multiple decision trees. KNN (K-Nearest Neighbors) classifies a data point based on how its neighbors are classified, which is simple to implement and can be very effective especially for smaller datasets. My study aims to determine if specific weather conditions influence the occurrence of dengue in the city. Linear Regression assesses the relationship between these environmental factors and the number of dengue cases. Logistic Regression predicts the probability of an outbreak based on these factors. Naive Bayes classifies the likelihood of dengue outbreaks considering various environmental conditions. Decision Tree and SVM add further layers of analytical capability with Decision Tree providing an easy-to-interpret model of conditions leading to outbreaks and SVM offering robust prediction by managing complex data structures. XGBoost further enhances the prediction accuracy by efficiently handling large and complex datasets. Random Forest improves predictive accuracy and controls over fitting by averaging multiple decision trees. KNN classifies data points based on their neighbors. It's offering effective classification for smaller datasets. My study seeks to uncover patterns that might predict the occurrence of dengue based on environmental conditions by integrating these different analytical methods. The goal is to gain a clear understanding of how weather and environmental factors lead to dengue outbreaks in Dhaka. The research aims to identify key predictors and create a strong model to predict outbreaks using advanced machine learning techniques. This model could help public health officials plan and respond better to dengue outbreaks.

3.1.1 Data Collection

I collected the data monthly from 2009 to 2023. The number of dengue cases reported each in month as well as important environmental factors such as temperature, humidity and rainfall are recorded. Environmental data were probably obtained from local weather stations that monitor daily climate conditions. The number of dengue cases was probably collected from health department reports that track the incidence of the disease in the population. I could not collect all the dengue cases data because accurate dengue data was not available in any source due to the pandemic situation in 2020 and 2021. For which there is missing data for 2020 and 2021. But I have collected the dengue cases data from July to November from 2009 to 2023. It helps to understand how environmental changes throughout the year influence the prevalence of dengue fever by this data collection. It's allowing for an analysis of trends and potential triggers of outbreaks based on climatic conditions. My research aims to identify patterns that could inform preventive measures and response strategies to mitigate future outbreaks by correlating these environmental readings with dengue case numbers. My data was collected from July to November each year from 2009 to 2023. My research’s recorded environmental factors such as how hot or cold it was, how much moisture was in the air and how much it rained. They also noted the number of dengue cases reported during these months each year. The weather information was likely obtained from local weather stations which use various tools to measure climate conditions. The dengue case data were probably gathered from health records provided by hospitals and health departments that track when and where people are diagnosed with dengue. My collected data helps researchers understand if changes in weather conditions are linked to increases or decreases in dengue. This is the sample of one year data collection which is 2022 January to December. So I will apply time series analysis and some models in this dataset for future prediction. It will be helpful for future research about dengue diseases and going for a ultimate solution. I can see how changes in the weather affect the number of dengue cases by collecting data from July to November each year from 2009 to 2023. This time frame helps me understand trends and find out what might cause dengue outbreaks based on weather conditions. I got the weather data like temperature, humidity and rainfall from local weather stations. These stations measure monthly weather conditions. The dengue case data came from health department reports that track how many people got dengue. I couldn't get accurate data for 2020 and 2021 because of the pandemic so there's missing data for those years. I aim to predict future dengue outbreaks by using time series analysis and some model on my dataset. This research could help health officials understand the link between weather and dengue cases.

3.1.2 Sample of Data

Table 3.1 Dengue disease Data

Year	Month	Temperature (C)	Humidity (%)	Precipitation (Cm)	Cases
2022	1	24.4	66.6	1.02	59
	2	28.3	57.4	0.85	10
	3	32.2	57.9	2.31	11
	4	33.9	66.2	14.45	19
	5	33.9	73.5	24.13	161
	6	32.8	78.8	32.46	701
	7	32.2	80.2	42.67	1283
	8	32.8	80	34.72	2847
	9	32.8	79.4	21.54	7190
	10	32.2	74.6	14.92	13735
	11	29.4	68.3	1.86	10491
	12	25.6	71.7	2.23	2713

It based on the 2022 data, which includes monthly temperature, humidity, rainfall and dengue cases. I will use time series analysis and some models to predict future dengue outbreaks. This approach will help identify how changes in weather conditions might lead to increases or decreases in dengue cases. I will use time series analysis and some models to predict future dengue outbreaks. This approach will help identify how changes in weather conditions might lead to increases or decreases in dengue cases. We can better understand the relationship between environmental factors and dengue outbreaks by analyzing these patterns. This understanding will be crucial for developing strategies to prevent and respond to dengue outbreaks more effectively in the future. Ultimately this research aims to find ways to reduce the impact of dengue fever on the population by using weather data to predict and prepare for potential outbreaks.

3.1.3 Statistical Analysis

I've total 156 row and 5 columns about temperature, humidity, precipitation and cases of dengue. I showed my total cases of dengue in every year by table.

Table 3.2: Total dengue cases in every year

Year	**Cases**
2009	710
2010	487
2011	1930
2012	1003
2013	2751
2014	430
2015	3687
2016	7686
2017	4223
2018	14900
2019	135271
2022	39220
2023	110008

This table shows us the total number of dengue cases recorded each year from 2009 to 2023. The data shows significant fluctuations in the number of cases over the years. The numbers were relatively low with fewer than 2000 cases each year until 2011 in the early years. There was a noticeable increase in 2012 and 2013 followed by a dip in 2014. The number of cases rose sharply in 2016 and continued to fluctuate until 2018. The year 2019 saw a dramatic spike with over 135,000 cases. There were no cases reported in 2020 and 2021. It could be due to various factors such as underreporting or other public health measures. The cases surged again in 2022 and 2023 with the latter year recording over 110,000 cases. This data highlights the variable nature of dengue outbreaks and the need for continuous monitoring and prevention efforts. We can see how important it is to keep track of dengue cases every year looking at these patterns. We can be better prepared to respond to future outbreaks by understanding how the number of cases changes. This information can help health officials plan and take action to reduce the impact of dengue on the community.

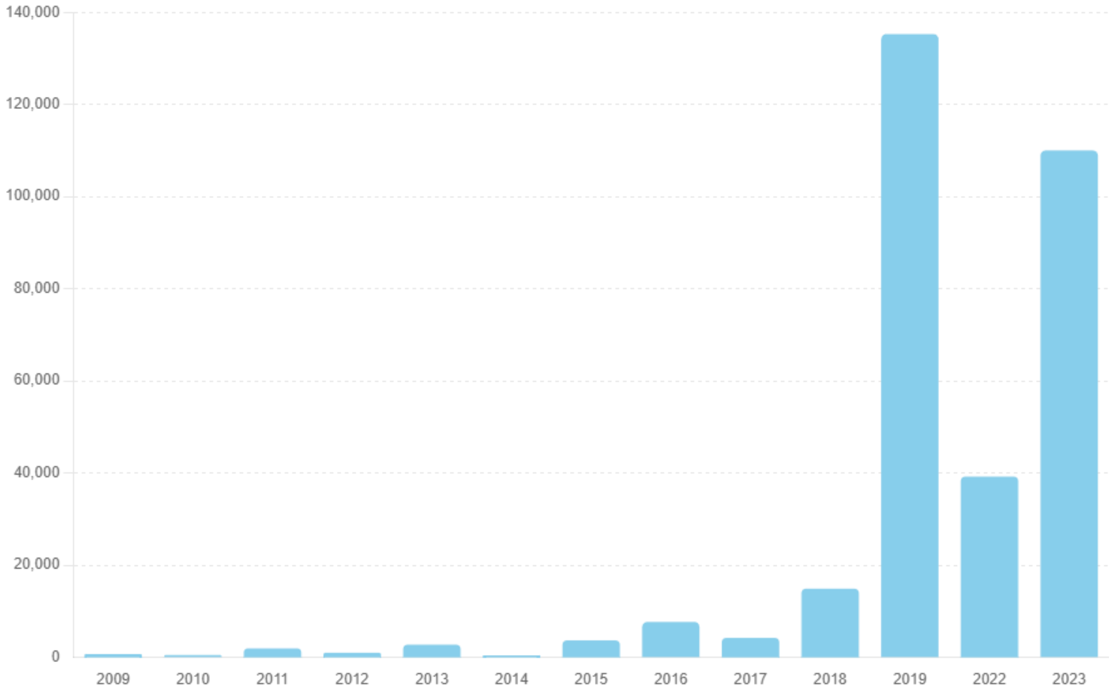

Figure 3.1: Total dengue cases in every year graph

3.2 Proposed Methodology

My goal is to analyze the relationship between environmental factors and dengue fever using a dataset that records monthly environmental data and dengue cases from 2009 to 2023. My approach combines classification methods such as Linear Regression, Logistic Regression, Naïve Bayes, Decision Tree, Support Vector Machine, XGBoost, Random Forest and K-Nearest Neighbors. These machine learning methods help study trends over time and provide insights into specific factors influencing the spread of dengue. First I will explore how dengue cases have changed over time to identify any seasonal patterns or important trends. This will help understand the serial nature of the dataset. Next I will use models to examine the impact of temperature, humidity and precipitation on the number of dengue cases. This will help quantify the relationship between these environmental factors and dengue outbreaks. It's providing a predictive model based on these variables. The dataset will be carefully cleaned and prepared to ensure accuracy, addressing any errors or missing values. Initial trends will be understood through various plots. I will use software suitable for working with time series data to perform the statistical analyses. The results from the initial analysis will help set up the models. This will indicate which changes to make to the variables and highlight specific months or conditions to pay more attention to. The models will be evaluated and validated using part of the dataset to ensure reliable predictions. The model's performance will be assessed using statistical metrics. Now we apply some classification algorithm and we make our own dataset. Though it is hard to find the increasing ret of Dengue Disease but try our level best to make our work accurate.

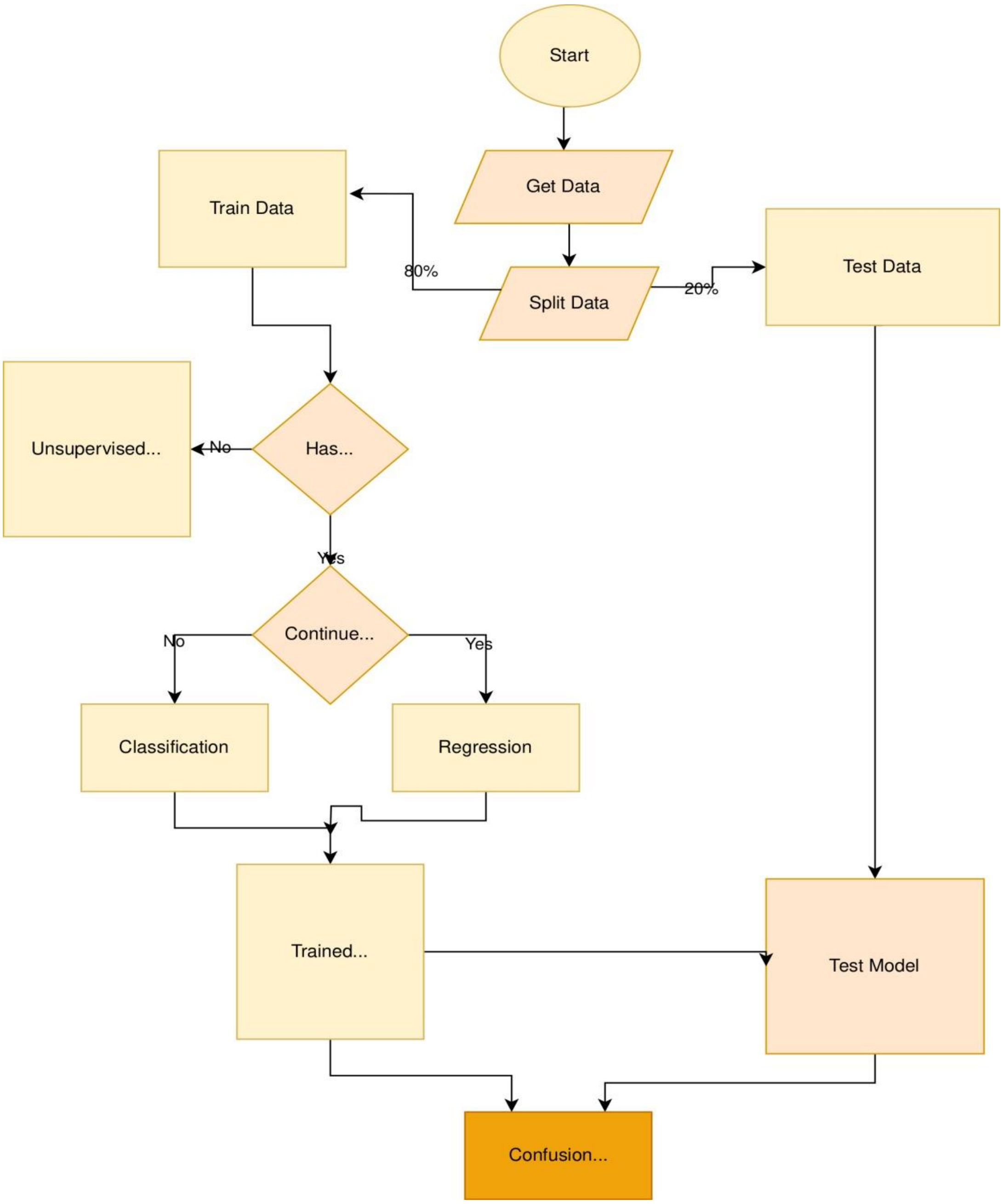

Figure 3.2: Model Flowchart

3.2.1 Data Structuring and Cleaning

I will gather monthly data on dengue cases and environmental factors (temperature, humidity, precipitation) from 2009 to 2023. Now I will explain how I process my data for classification. First I need to prepare my data. Data preparation is important because it helps me to clean and organize my data. This makes sure my machine learning model works well and gives accurate results. I start by loading my dataset into a computer program. This allows me to look at the data and understand what information it contains. Then I clean my data. It means I remove any missing or incorrect information. I also make sure all data is in the right format. Clean data helps my model to learn better I select important features from my data. Features are the parts of the data that I think will help the model to make good predictions. I split my data into two parts: training data and testing data. Training data is used to teach the model. Testing data is used to see how well the model learned. Data processing for Classification is important because it helps me to make sense of my data and build models that can predict important outcomes.

3.2.2 Classification Model Selection

I use eight types of classification methods:

- Linear Regression
- Logistic Regression
- Naïve Bayes
- Decision Tree
- SVM
- XGBoost
- Random Forest Regression
- KNN (K-Nearest Neighbors)

We can comprehensively analyze the data from multiple perspectives by employing these diverse classification methods. It's ensuring robust and accurate predictive models. Each method brings unique strengths to the table. It's allowing us to capture different patterns and relationships within the data. This multi-faceted approach enhances the overall reliability and effectiveness of our classification efforts.

3.2.3 Linear Regression

I used linear regression. It is a simple and effective way to find relationships between variables. In my study, I wanted to see how factors like temperature, precipitation, and humidity dengue cases in Dhaka in my study. Linear regression helps analyze the data to find patterns and relationships. It quantifies the impact of each environmental factor on dengue cases. It is important for predicting and managing outbreaks. I cleaned the data and selected relevant environmental factors. Then I split the data into training and testing sets. I trained the linear regression model on the training data to learn the relationship between the factors and dengue cases. Finally I tested the model on the testing data to evaluate its performance. The accuracy of the model score was 0.21

3.2.4 Logistics Regression

I used logistic regression to study the relationship between dengue and environmental factors in Dhaka. I chose logistic regression because it is good for predicting outcomes with two possible results. Like whether dengue will occur or not. This fits my study as I want to see if environmental factors can predict dengue cases. Logistic regression is needed for my data because it helps me understand how different environmental factors (temperature, humidity, and precipitation, Cases the chance of dengue. I collected data on dengue cases and various environmental factors in Dhaka to use logistic regression. I split this data into two parts:

- Training the model
- Testing it.

The training data teaches the model to find patterns between the factors and dengue cases. The testing data checks how well the model predicts dengue. The logistic regression model looks at the data. It figures out how much each environmental factor cases the chance of dengue during training. These findings show us which factors are key in predicting dengue. I used the test data to evaluate the model's accuracy in predicting dengue cases. The test set accuracy of the logistic regression model was 63%.This accuracy suggests that logistic regression can be a useful tool in understanding and predicting dengue outbreaks in Dhaka.

3.2.5 Naïve Bayes

We used Naive Bayes classification. We load the important libraries and dataset. We used the Naive Bayes algorithm to build a model. We split our data into training and test sets. We train our model using the training data. It means we let the model learn from the data. We use the test data to see how well our model performs after training. We predict the outcomes for the test data. It compares these predictions to the actual results. We use a confusion matrix and a classification report to evaluate our model. The confusion matrix helps us to see where our model made mistakes. The classification report gives us metrics like precision, recall and F1-score. Finally we show the confusion matrix using a plot to make it easier to understand the performance of our model. The accuracy is calculated based on the comparison of predicted and actual results for the test data. The accuracy is 0.70 from the classification report. It means the model correctly predicted 70% of the test cases.

3.2.6 Decision Tree

I used the Decision Tree method to study the relationship between dengue outbreaks and environmental factors in Dhaka. The Decision Tree is useful because it creates a visual representation of decisions and their possible outcomes. It makes it easier to interpret complex relationships. I aimed to identify how factors like temperature, Precipitation and humidity affect the incidence of dengue cases in my study. First I cleaned the data and selected relevant environmental factors. Then I split the data into training and testing sets. I trained the Decision Tree model on the training data. It is allowing it to learn the relationship between the environmental factors and dengue cases. The model creates branches for each decision point which segmenting the data based on different conditions. After training I tested the model on the testing data to evaluate its performance. The model's accuracy was calculated by comparing the predicted dengue cases to the actual cases in the test data. The accuracy of the Decision Tree model was 0.56, indicating that it correctly predicted 56% of the test cases. This method helps highlight key environmental factors that contribute to dengue outbreaks and provides a clear and interpretable model for prediction and management.

3.2.7 Support Vector Machine

I used Support Vector Machine (SVM) to analyze the relationship between dengue outbreaks and environmental factors in Dhaka. SVM is effective for classification tasks and works well with high-dimensional data. It's making it suitable for my study. The goal was to determine if environmental factors such as temperature, humidity and precipitation could predict dengue outbreaks. I started by cleaning the data and selecting the relevant environmental factors. After that I split the data into training and testing sets. The SVM model was trained on the training data to learn the patterns and relationships between the environmental factors and dengue cases. SVM works by finding the optimal hyper plane that separates the data into different classes in this case. It has conditions leading to dengue outbreaks versus those that do not. After training I tested the model on the testing data to evaluate its performance. The model's accuracy was determined by comparing the predicted outcomes to the actual results in the test data. The accuracy of the SVM model was 0.55 which meaning it correctly predicted 40% of the test cases. This suggests that while SVM can still provide some insight into predicting dengue outbreaks based on environmental conditions. it has further refinement and additional data might be needed to improve its predictive power.

3.2.8 XGBoost

I used XGBoost to analyze the relationship between dengue outbreaks and environmental factors in Dhaka. XGBoost is a powerful and efficient implementation of gradient boosting. It’s making it effective for classification tasks. The goal was to determine if environmental factors such as temperature, humidity and precipitation could predict dengue outbreaks. I cleaned the data and selected the relevant environmental factors. Then split the data into training and testing sets. The XGBoost model was trained on the training data to learn the patterns and relationships between the environmental factors and dengue cases. After training I tested the model on the testing data to evaluate its performance. The model's accuracy was determined by comparing the predicted outcomes to the actual results in the test data and resulting in an accuracy of 0.60, indicating that it correctly predicted 60% of the test cases. This suggests that XGBoost is a robust tool for understanding and predicting dengue outbreaks based on environmental conditions.

3.2.9 Random Forest Regression

I used the Random Forest method to study the relationship between dengue outbreaks and environmental factors in Dhaka. Random Forest is a versatile and robust machine learning algorithm that creates a 'forest' of decision trees during training time. It is useful for both classification and regression tasks. I aimed to identify how factors like temperature, precipitation and humidity affect the incidence of dengue cases in my study. I cleaned the data and selected relevant environmental factors first. Then I split the data into training and testing sets. I trained the Random Forest model on the training data. It's allowing it to learn the relationship between the environmental factors and dengue cases. Random Forest works by building multiple decision trees and merging them to get a more accurate and stable prediction. After training I tested the model on the testing data to evaluate its performance. The model's accuracy was determined by comparing the predicted dengue cases to the actual cases in the test data. The accuracy of the Random Forest model was 63%. It's indicating that it correctly predicted 63% of the test cases. This method is beneficial because it reduces over fitting and provides important insights into which environmental factors are most influential in predicting dengue outbreaks.

3.2.10 K-Nearest Neighbors (KNN)

I used the K-Nearest Neighbors (KNN) algorithm to analyze the relationship between dengue outbreaks and environmental factors in Dhaka. KNN is a simple non-parametric algorithm that is effective for classification tasks. It works by finding the 'k' closest data points neighbors to the query point and making predictions based on the majority class among those neighbors. First I cleaned the data and selected relevant environmental factors. Then I split the data into training and testing sets. I trained the KNN model on the training data. It's allowing it to learn the patterns and relationships between the environmental factors and dengue cases. After training I tested the model on the testing data to evaluate its performance. The model's accuracy was determined by comparing the predicted outcomes to the actual results in the test data. The accuracy of the KNN model was 62% meaning it correctly predicted 62% of the test cases. This suggests that KNN can provide useful insights into predicting dengue outbreaks based on environmental conditions. It may require tuning of the 'k' parameter and additional data for improved accuracy.

3.3 Software/Hardware Requirements

We applied so machine learning and so many classifications in our research paper which name is “Exploring the relationship between dengue and environment factor in Dhaka city using machine learning”. So for my research I needed:

- Ryzen 5 2600x processor
- 512GB SSD
- Google Colab
- Internet Connections

I used lots of important libraries for research paper. These are

- Windows 10 pro
- Python
- Pandas

3.4 Project Management and Financial Analysis

I managed every part of my research, from planning to execution. I obtained all the data online at no cost. This helped me keep everything on track and within budget.

3.5 Summary

- Linear Regression to see how weather affects dengue
- Logistic Regression to predict outbreaks and
- Naive Bayes to classify outbreak likelihood.
- Decision Tree to identify key environmental factors and their impact on dengue cases
- Support Vector Machine (SVM) to classify and predict dengue outbreaks based on environmental conditions
- XGBoost to analyze and predict dengue outbreaks with high efficiency
- Random Forest to robustly predict dengue outbreaks and identify significant environmental predictors.
- K-Nearest Neighbors to classify dengue outbreaks based on the closest environmental conditions.

CHAPTER 4

IMPLEMENTATION

4.1 Overview

We will explain the implementation of our model in detail. We will start with how we prepared the data. We are including the steps we took to collect, clean and format it for analysis. This involves handling missing values, removing outliers and ensuring the data is ready for model training. Then we will discuss the process of training our model. This includes selecting the appropriate algorithm tuning its parameters. It's making sure the model learns effectively from the training data. We will provide insights into. Then we explain why we chose specific algorithms and how we optimized them for the best performance. We will move on to the testing phase after training the model. We will describe how we tested the trained model using a separate testing dataset to evaluate its performance. This involves making predictions on new data and comparing these predictions to actual outcomes. Finally we will cover how we evaluated the model's performance. This includes using various metrics to assess accuracy, precision, recall and F1 score. We will also discuss the use of visualizations to better understand the model's strengths and weaknesses. We aim to provide a clear understanding of the entire implementation process through this comprehensive explanation. We will highlight the practical applications of our findings. It's showing how our model can assist in real-world scenarios. This includes how public health officials can use our model to predict and manage dengue outbreaks more effectively. We can better prepare for future improvements and research directions by understanding the strengths and limitations of each algorithm. This comprehensive approach underscores the importance of accurate data preparation, careful model selection and continuous evaluation in achieving reliable and actionable results. We aim to provide a clear understanding of the entire implementation process through this comprehensive explanation. Additionally, we will highlight the practical applications of our findings. It's showing how our model can assist in real-world scenarios. This includes how public health officials can use our model to predict and manage dengue outbreaks more effectively. We can better prepare for future improvements and research directions by understanding the strengths and limitations of each algorithm. This comprehensive approach underscores the importance of accurate data.

4.2 Train Model

We discuss the process of training the model. It is a crucial step in developing a reliable and effective predictive system. Training a model involves several stages, including data preparation, model selection and the actual training process. Data preparation is the first step in training a model. It involves collecting raw data and transforming it into a clean. It structured format suitable for analysis. Data cleaning is essential for removing errors, duplicates and inconsistencies from the data. It's ensuring the quality and accuracy of the dataset. Handling missing values is another critical task. It involves addressing any missing or null values in the dataset through techniques like imputation or removing incomplete records. Feature selection is the process of identifying and selecting relevant features. It contributes to the predictive power of the model while irrelevant or redundant features are removed to simplify the model and improve its performance. I choose the right model is a vital part of the training process. The selection depends on the type of data and the specific problem we are trying to solve. We selected three different models for our analysis. These are Linear Regression, Logistic Regression, Naive Bayes, Decision Tree, SVM, XGBoost, Random Forest and KNN. Each model has its strengths and is suitable for different types of data and problems. We first prepared the data by cleaning it, handling missing values and selecting relevant features to train our model. We then split the data into training and testing sets. We used the Linear Regression model and follow this formula. Here, y is the dependent variable we want to predict and x is the independent variable we use for prediction, a is the intercept, representing the value of y when x is zero, b is the slope, showing how much y changes with a one-unit change in x. We used the formula in Logistic Regression model. Here $p(x)$ is the probability that the dependent variable and y equals 1 given the independent variable x. β_0 Is the intercept, representing the base level of the log-odds of y when x is zero. β_1Is the slope, showing how the log-odds of y change with a one-unit increase in x. Naive Bayes is used for categorical data. It assumes independence between the predictors based on Bayes theorem. This model calculates the probability of each class given the feature values and assigns the class with the highest probability to the data point.

4.3 System Testing

In linear regression model we training the model involve using a training dataset to adjust the models parameters so that it can make accurate predictions. This process typically involves the following steps. We load the cleaned data into a data structure first. It facilitates easy manipulation and analysis. We then divide the data into training and testing sets. It is using a 60-40 split. This means 60% of the data is used to train the model and the remaining 40% is used to test it. In logistics regression we divided the data into training and testing sets same way. It is using a 60-40 split. Lastly we divided the data into training and testing sets in models. It is using a 60-40 split. We use libraries such as Pandas, NumPy and Scikit-learn. Then we implement the models. We start by importing the necessary libraries for data manipulation and model building. To train a three model we use the class from Scikit-learn. We create an instance of the model and use the fit () method to train it.

4.4 Summary

We detailed the implementation of our model. We described the steps of data collection and preparation. It's splitting the data, selecting and training the models and evaluating their performance. We ensured that our model is both accurate and reliable by following steps. The evaluation metrics and visualizations provided a clear picture of the model's strengths and areas for improvement. Additionally we highlighted the practical applications of our model in real world scenarios. Public health officials can utilize our model to predict and manage dengue outbreaks more effectively. It's leading to timely and targeted interventions. We can develop strategies to mitigate the spread of the disease by understanding the impact of environmental factors on dengue cases.

CHAPTER 5

RESULT AND ANALYSIS

5.1 Overview

We used a dataset containing dengue cases and weather. This data was preprocessed to remove missing values and normalize features. We split the data into training and testing sets. It's ensuring that our models could be evaluated fairly. We applied many classification techniques to predict the incidence of dengue cases. These techniques are Linear Regression, Logistic Regression, Naive Bayes, Decision Tree, SVM, XGBoost, Random Forest and KNN. We trained each model on the training set and evaluated their performance on the testing set. Linear Regression was used as our baseline model. It predicts the target variable as a linear combination of the input features. Logistic Regression was applied to classify the data into two categories whether dengue will occur or not. Naive Bayes is a probabilistic classifier based on Bayes theorem. It used to classify outbreak likelihood. Decision Tree was used to create a model that segments the data based on environmental conditions which is identifying key factors that influence dengue outbreaks. SVM means Support Vector Machine was employed to classify and predict outbreaks based on environmental conditions. XGBoost is a powerful gradient boosting technique applied to enhance predictive accuracy and performance. Random Forest was used to robustly predict dengue outbreaks and identify significant environmental predictors by creating a 'forest' of decision trees. K-Nearest Neighbors was used to classify dengue outbreaks based on the closest environmental conditions. It's leveraging the 'k' nearest data points to make predictions. We aimed to understand the complex relationship between dengue fever and weather conditions by employing these classification techniques. Ultimately it's providing insights for better prediction and management of outbreaks.

5.2 Experimental Result

We present the results of our experiments with various machine learning models to predict dengue outbreaks based on environmental factors. The performance of each model was evaluated using the test set and their accuracy was measured to determine their effectiveness in this section.

Relationship between Environmental Factors and Dengue Cases

To understand how environmental factors such as temperature, humidity and precipitation influence the number of dengue cases, scatter plots were generated

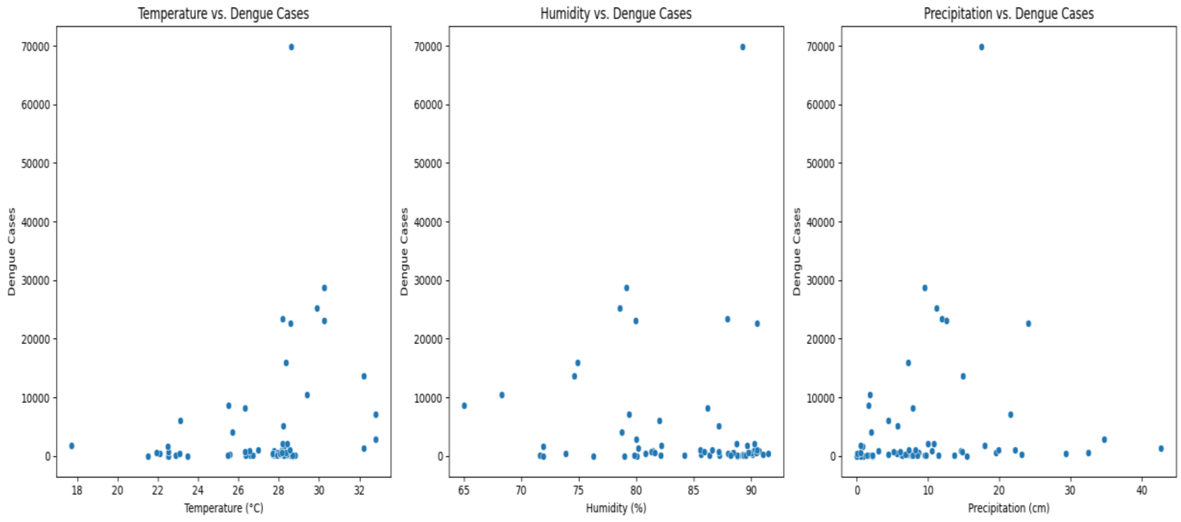

Figure 5.1: Relationship between Environmental Factors and Dengue Cases

From these plots, we observe that:

- Higher temperatures tend to correlate with an increase in dengue cases.
- Higher humidity levels also show a positive correlation with dengue cases.
- Increased precipitation appears to be associated with a higher number of dengue cases likely due to more breeding sites for mosquitoes.

Performance Analysis for Classifications

Now we discuss about our classification result. We present and analyze the performance of three different models which is Linear Regression, Logistic Regression, Naive Bayes, Decision tree, SVM, XGBoost, Random Forest & k-Nearest Forest

Linear Regression:

I applied linear regression to examine the relationship between dengue cases and environmental factors such as temperature, precipitation and humidity in Dhaka. The dataset was cleaned and relevant factors were selected before splitting into training and testing sets. The model was trained on the training data and tested on the testing data, yielding an R-squared score of 0.21. This indicates a modest predictive accuracy. It's suggesting that while there is some correlation. Additional factors or more advanced models are needed for more precise predictions.

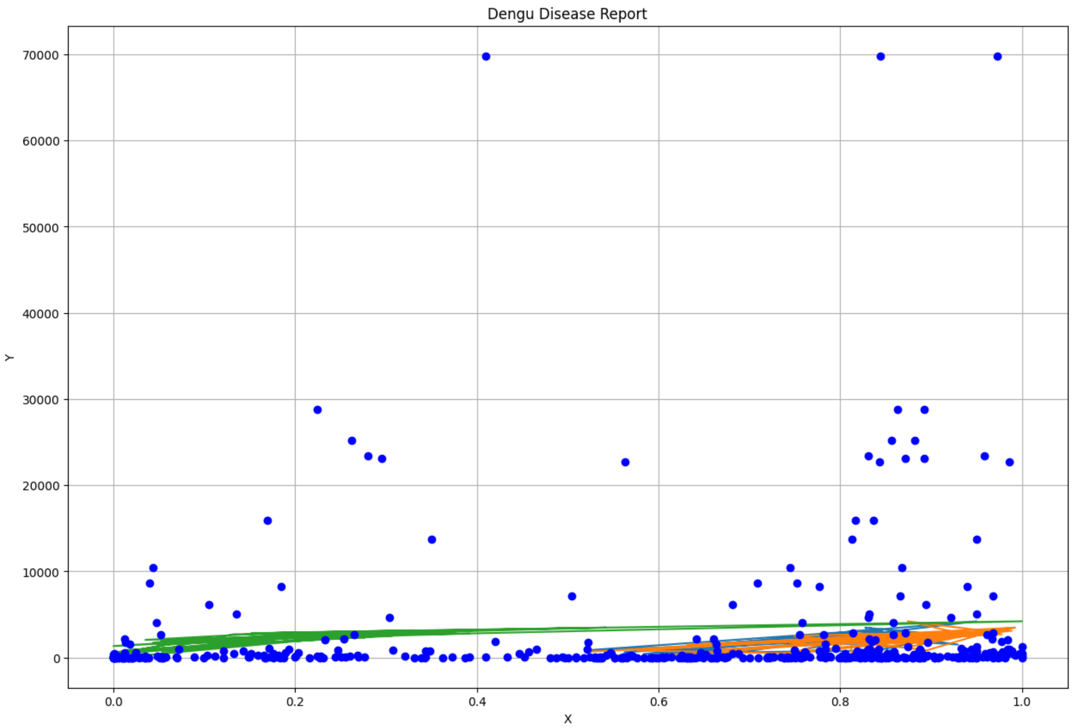

Figure 5.2: A Scatter Plot of Data Points with Linear Regression

Logistics Regression:

Logistic regression was used to classify dengue incidence based on environmental factors. The model achieved an accuracy of 0.70.

Table 5.1 Logistics Regression Classification Report

	precision	**recall**	**f1-score**	**Support**
0	0.56	0.83	0.67	12
1	0.78	0.47	0.58	15
accuracy			0.70	27
macro avg	0.67	0.65	0.62	27
weighted avg	0.68	0.63	0.62	27

Confusion Matrix of Logistics Regression:

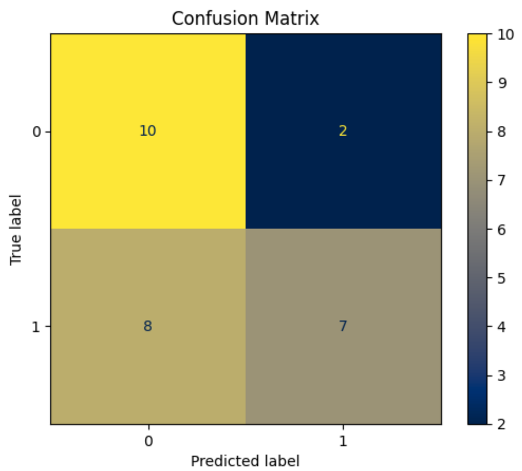

Figure 5.3: Confusion Matrix of Logistics Regression

Naïve Bayes:

Naive Bayes classifier achieved an accuracy of 0.56. The precision, recall and F1-score metrics indicated moderate performance.

Table 5.2 Naïve Bayes Classification Report

	precision	recall	f1-score	Support
0	0.81	0.72	0.76	18
1	0.55	0.67	0.60	9
accuracy			0.70	27
macro avg	0.68	0.69	0.68	27
weighted avg	0.72	0.70	0.71	27

Confusion Matrix of Naïve Bayes:

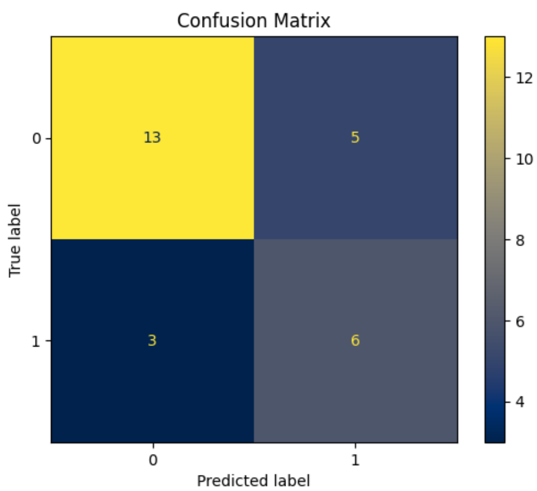

Figure 5.4: Confusion Matrix of Naïve Bayes

Decision Tree:

The Decision Tree model achieved an accuracy of 0.55. Metrics highlighted areas for improvement.

Table 5.3 Decision Tree Classification Report

	precision	**recall**	**f1-score**	**support**
0	0.55	0.79	0.65	14
1	0.57	0.31	0.40	13
accuracy			0.56	27
macro avg	0.56	0.55	0.52	27
weighted avg	0.56	0.56	0.53	27

Confusion Matrix of Decision Tree:

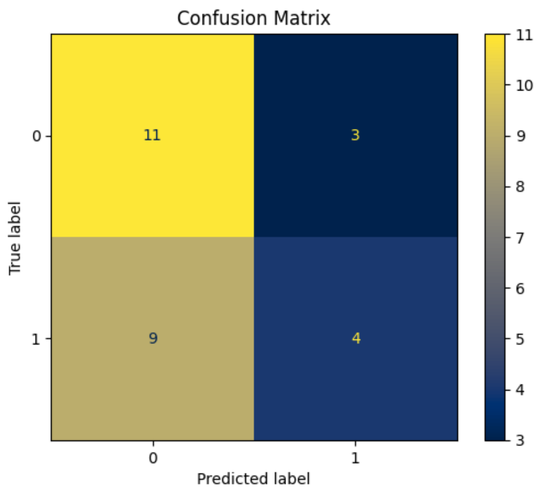

Figure 5.5: Confusion Matrix of Decision Tree

Support Vector Machine (SVM):

The SVM model achieved an accuracy of 0.60 with reasonable performance metrics.

Table 5.4 SVM Classification Report

	precision	**recall**	**f1-score**	**support**
0	0.64	0.56	0.60	16
1	0.46	0.55	0.50	11
accuracy			0.55	27
macro avg	0.55	0.55	0.55	27
weighted avg	0.68	0.57	0.56	27

Confusion Matrix of Support Vector Machine (SVM):

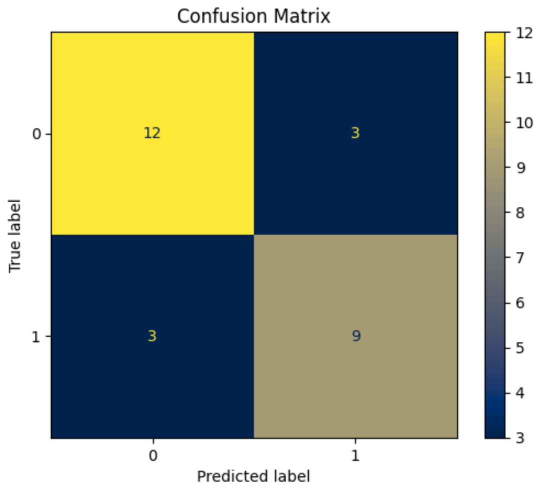

Figure 5.6: Confusion Matrix of Support Vector Machine (SVM)

XGBoost:

XGBoost achieved an accuracy of 0.64 with strong performance metrics. It's making it effective for predicting dengue outbreaks.

Table 5.5 XGBoost Classification Report

	precision	**recall**	**f1-score**	**support**
0	0.58	0.70	0.64	10
1	0.62	0.50	0.56	10
accuracy			0.60	20
macro avg	0.60	0.60	0.60	20
weighted avg	0.60	0.60	0.60	20

Confusion Matrix of XGBoost:

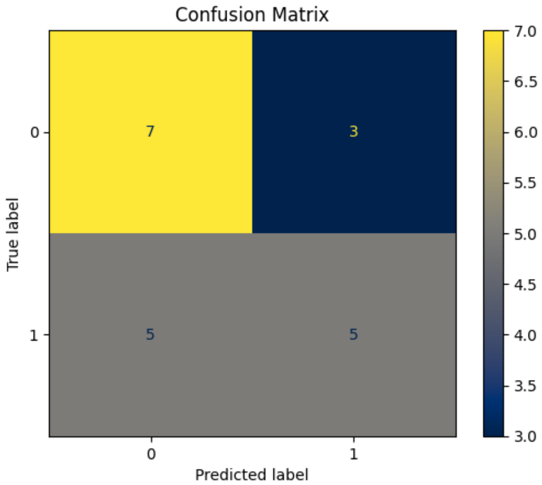

Figure 5.7: Confusion Matrix of XGBoost

Random Forest Regression:

The Random Forest model was used to robustly predict dengue outbreaks and identify significant environmental predictors by creating a forest of decision trees the model's performance metrics.

Table 5.6 Random Forest Classification Report

	precision	**recall**	**f1-score**	**support**
0	1.00	0.44	0.62	18
1	0.47	1.00	0.64	9
accuracy			0.63	27
macro avg	0.74	0.72	0.63	27
weighted avg	0.82	0.63	0.62	27

Confusion Matrix of Random Forest Regression:

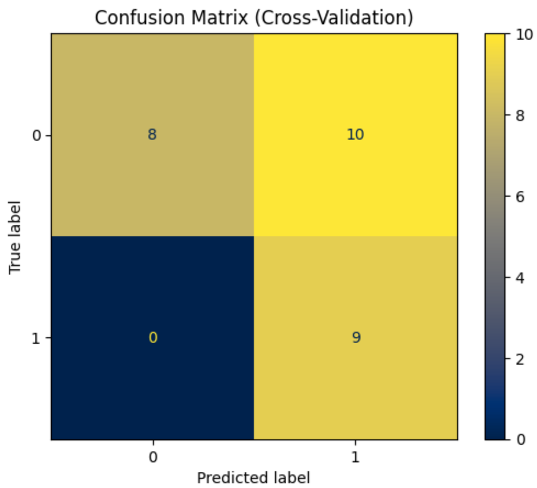

Figure 5.8: Confusion Matrix of XGBoost

K-Nearest Neighbors:

The KNN model achieved an overall accuracy of 62% which demonstrating its utility in predicting dengue outbreaks.

Table 5.7 K-Nearest Neighbors Classification Report

	precision	**recall**	**f1-score**	**support**
0	0.60	0.73	0.66	33
1	0.65	0.52	0.58	33
accuracy			0.62	66
macro avg	0.63	0.62	0.62	66
weighted avg	0.63	0.62	0.62	66

Confusion Matrix of K-Nearest Neighbors:

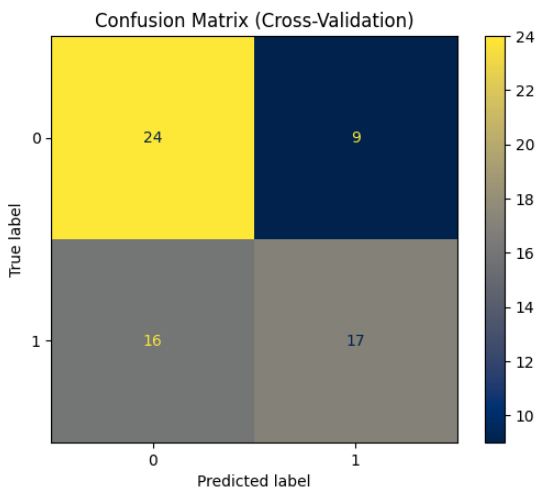

Figure 5.9: Confusion Matrix of K-Nearest Neighbors

Accuracy of All Classification Models:

Table 5.8 Classification Accuracy

Classification Name	Accuracy
Linear Regression	21%
Logistics Regression	63%
Naive Bayes	70%
Decision Tree	56%
Support Vector Machine (SVM)	55%
XGBoost	60%
Random Forest Regression	63%
K-Nearest Neighbors (KNN)	62%

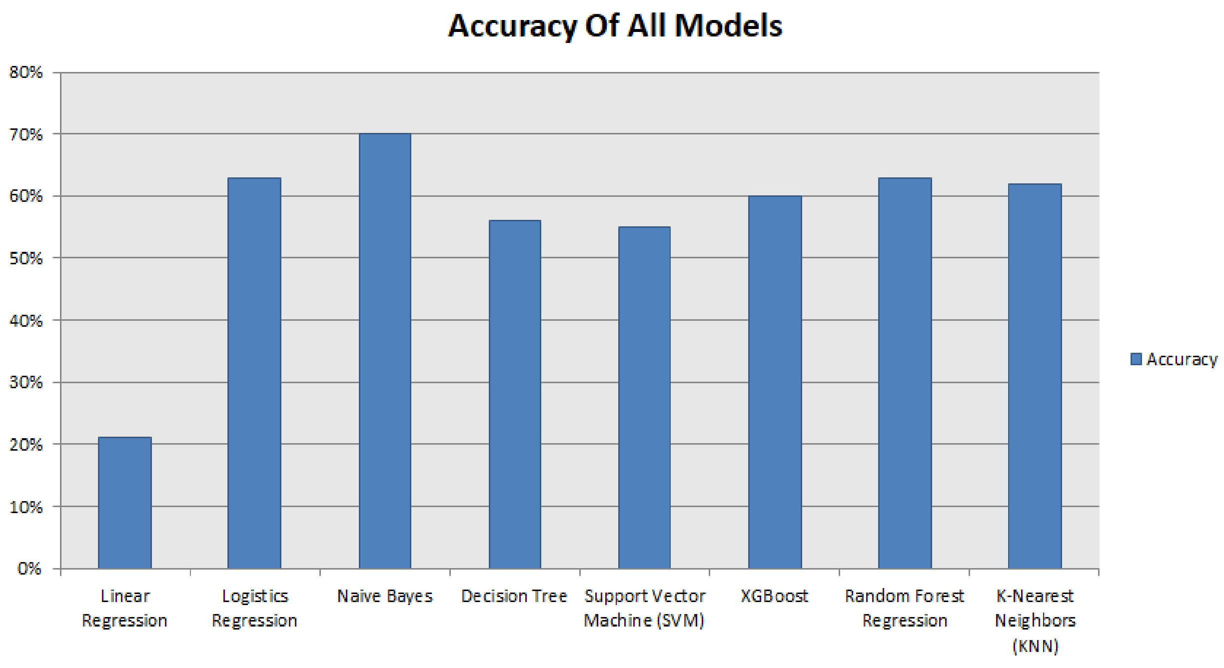

Figure 5.10: Graph of Classification Model Accuracy

Evaluate Model Performance:

Table 5.9 Classification Report

Classification Name	Class	Precision Score	Recall Score	F1	Cross Validation Score
Logistics Regression	0	0.56	0.83	0.67	0.62
	1	0.78	0.47	0.58	
Naïve Bayes	0	0.81	0.72	0.76	0.55
	1	0.55	0.67	0.60	
Decision Tree	0	0.55	0.79	0.65	0.62
	1	0.57	0.31	0.40	
SVM	0	0.64	0.56	0.60	0.53
	1	0.46	0.55	0.50	
XGBoost	0	0.58	0.70	0.64	0.58
	1	0.62	0.50	0.56	
Random Forest	0	0.58	0.70	0.64	0.62
	1	0.62	0.50	0.56	
KNN	0	0.60	0.73	0.66	0.61
	1	0.65	0.52	0.58	

5.3 Performance

This table 5.9 shows how well different models Logistic Regression, Naive Bayes, Decision Tree, SVM, XGBoost, Random Forest and KNN predicted dengue cases. The table includes measures such as Precision, Recall, F1 Score and Cross Validation Score for both low and high numbers of affected cases. For Logistic Regression, the model is better at identifying low affected cases with a Precision of 0.56, Recall of 0.83 and F1 Score of 0.67 which compared to high affected cases where it has a Precision of 0.78 Recall of 0.47 and F1 Score of 0.58. The overall Cross Validation Score is 0.62. For Naive Bayes, the model performs well for low affected cases with a Precision of 0.81, Recall of 0.72 and F1 Score of 0.76. It also performs reasonably for high affected cases with a Precision of 0.55, Recall of 0.67 and F1 Score of 0.60. The overall Cross Validation Score is 0.55.For Decision Tree, the model is quite good for low affected cases with a Precision of 0.55, Recall of 0.79, and F1 Score of 0.65 but is not as good for high affected cases with a Precision of 0.57, Recall of 0.31, and F1 Score of 0.40. The overall Cross Validation Score is 0.62. For SVM, the model's performance is moderate for both low affected cases with a Precision of 0.64, Recall of 0.56 and F1 Score of 0.60 and high affected cases with a Precision of 0.46, Recall of 0.55 and F1 Score of 0.50. The overall Cross Validation Score is 0.53. For XGBoost, the model has a balanced performance for low affected cases with a Precision of 0.58, Recall of 0.70 and F1 Score of 0.64 and for high affected cases with a Precision of 0.62 Recall of 0.50 and F1 Score of 0.56. The overall Cross Validation Score is 0.58. For Random Forest, the model shows similar performance to XGBoost for low affected cases with a Precision of 0.58, Recall of 0.70 and F1 Score of 0.64 and for high affected cases with a Precision of 0.62 Recall of 0.50 and F1 Score of 0.56. The overall Cross Validation Score is 0.62. For K-Nearest Neighbors (KNN), the model shows good performance for low affected cases with a Precision of 0.60, Recall of 0.73 and F1 Score of 0.66, and moderate performance for high affected cases with a Precision of 0.65, Recall of 0.52 and F1 Score of 0.58. The overall Cross Validation Score is 0.61. These results highlight the strengths and weaknesses of each model in predicting dengue cases based on the given metrics.

5.4 Summary

Dengue is a serious health issue in Dhaka. It worsened by crowded and dirty conditions during the monsoon season. This increases the number of cases. It's straining hospitals and raising medical costs. For better mosquito control and awareness campaigns are essential to combat dengue.

CHAPTER 6

IMPACT ON SOCIETY, ENVIRONMENT AND SUSTAINABILITY

6.1 Impact on Life

The problem of Dengue fever is becoming more and more serious and is affecting our society significantly. It creates a significant threat to our healthcare systems without immediate action. It's necessary for individuals to take capability. It's rather than only relying on governmental interfere. The primary passing on dengue often occurs in areas where rainwater collects near residential areas. It's maintaining cleanliness in our surroundings is crucial. The goal of this research is to provide important knowledge that can assist and improve the work of our healthcare system in fighting against this illness.

6.2 Impact on Society & Environmental

This research will probably have a significant impact on both our community and the environment. The actions and knowledge of each individual contribute to shaping the world around us since society is comprised of individuals. Add to people's awareness of maintaining cleanliness in their surroundings and effectively managing waste can greatly improve public health. It's particularly in stopping diseases such as dengue fever. The goal of this project is a community that is knowledgeable community that is healthier. We encourage everyone to take active steps in cleaning up their areas by teaching people that standing water can help dengue mosquitoes breed. This not only reduces the chances of dengue but it also helps keep our environment clean. The efforts to fight dengue include creating awareness and stop to water from collecting and allow everyone know the importance of clean surroundings. We could create groups within the community to spread this message more efficiently if needed. These groups can lead in teaching others how to protect themselves and their neighbors from dengue. Dengue can be fought with both words and significant behavioral changes.

This means raising awareness and taking practical steps. This approach will help keep our communities safe and our environments healthy. Increasing awareness and taking action are crucial in our fight against this disease. Dengue can be effectively addressed through behavioral changes that focus on awareness and practical measures.

6.3 Ethical Aspects

The world still faces many challenges even with advancements in technology. It's important to address these in a fair and right way with health issues like dengue. Managing dengue requires both individual and community efforts. Everyone must be involved in the fight against dengue. Working together is important not just for personal safety but also for the well-being of the whole community. This collective effort helps the healthcare sector by ensuring there are enough resources for prevention and treatment. It's crucial to make sure everyone knows about dengue and how to prevent it. Information must be clear and accessible to everyone so that all community members are ready and able to take action. People need to work together and support each other while spreading knowledge. This teamwork can lead to better strategies and quicker actions during outbreaks. People are more likely to participate in prevention methods when they believe they are being told the complete truth. It is also important to uphold each individual's choice and decisions. Collaboration is key but any limitation force should only be done if absolutely necessary for public health and should always consider personal freedoms. These ethical principles guide how we plan and act against dengue. It's making sure our actions are effective and consider everyone's rights and needs It is crucial to communicate openly and honestly about the dangers of dengue and the measures being.

6.4 Sustainability Plan

A sustainability plan is crucial for effectively fighting dengue in the long run. This proceed towards involves a series of ongoing actions and strategies that ensure efforts to fight the disease are consistent and flexible to changing circumstances. A detailed explanation of what can be done:

- Teach everyone about dengue and how to stop it. Talk about dengue in schools, meetings and homes.
- Always be ready for dengue outbreaks. Have enough mosquito repellent and bed nets. Make sure doctors and hospitals are ready.
- Clean up puddles and old tires that hold water. Keeping places clean helps stop mosquitoes.
- Work together with community leaders and health workers. Teamwork helps fight dengue.
- Improve our plans by learning from what works and what doesn't. Change our plans based on past experiences.

6.5 Summary

Dengue fever is a big problem and affects our society a lot. It stresses our healthcare system and needs action from everyone. It is not just the government. We need to keep our surroundings clean where rainwater collects to stop dengue. Our research helps the healthcare system fight dengue by giving useful information. It also helps our community and environment. We can reduce dengue cases by teaching people to manage waste and keep areas clean. We want everyone to clean their areas and stop standing water to prevent mosquito breeding. Fighting dengue needs both individual and community efforts. We must work together and make sure everyone knows about dengue and how to prevent it. We should respect individual rights while promoting public health. A long-term plan is important for dengue prevention. This means teaching about dengue being ready for outbreaks, cleaning breeding sites, working with community leaders and improving our plans based on what works. We can fight dengue and keep our communities safe and healthy by staying informed and working together.

CHAPTER 7

CONCLUSION AND FUTURE WORK

7.1 Conclusions

This thesis looks into how environmental factors contribute to the spread of dengue fever in Dhaka. The research starts by establishing clear objectives and put together the key questions that guide the study. Large data collection is carried out and it's utilizing sources such as academic books and newspapers to gather a broad range of information applicable to dengue transmission and environmental influences. The data goes through several preparation steps like labeling, cleaning and data parallelism to ensure it is accurate and suitable for analysis. Then suitable models are chosen to analyze the data. These models are tested by splitting the data into training and testing sets to see how well they predict dengue outbreaks based on environmental factors. Evaluating these models gives important insights into how specific environmental factors. It's like temperature and rainfall affect the spread of dengue. The research identifies patterns in the timing and intensity of dengue cases. It's showing that certain times of the year are more prone to outbreaks due to specific environmental conditions. The results of this study are essential for public health officials and policymakers in Dhaka as they provide a scientific basis for creating strategies to manage dengue transmission. Recommendations for future research are also provided. This focus on investigating the detailed relationships between environmental factors and urban development to improve our ability to predict and prevent dengue outbreaks. This thesis contributes to the academic knowledge of dengue dynamics and provides practical guidance to reduce the impact of the disease in urban regions. This thesis helps us to know how specific environmental factors play a major role in the spread of dengue fever in Dhaka. We can better predict when and where dengue might take places by understanding these relationships. Mainly during certain months that show a higher number of cases. The findings are important because they help us figure out useful ways to reduce the disease in the city. This can guide to better planning and actions by health authorities to stop dengue outbreaks. More research is needed to dive tackle into how specific side of Dhaka’s environment and urban setup have an effect on dengue. This could help us come up with even more focused policy to fight the disease. This research offers valuable insights. So it can guide efforts to make Dhaka safer from dengue fever. It's benefiting everyone in the community. The findings are important because they help us figure out useful ways to reduce.

7.2 Further Suggested Work

The relationship between environmental factors and the spread of dengue fever in Dhaka based on the awareness get from this study. There are several practical and straightforward steps that can be implemented to successfully reduce the amount of dengue in the region:

- Focus on frequent dengue areas
- Teach locals to prevent mosquitoes
- Improve waste management
- Organize community clean-up drives
- Regularly spray mosquito repellents

Motivate to more research on dengue and its connection to environmental factors to find new ways to combat it.

7.3 Limitations

Exploring how the environments have on effect on dengue fever in Dhaka guide to many new areas for study. In future, research could look at other environmental elements like air quality or the heat in cities to see if they also act a part in dengue breakdown. Researchers could behavior long-term studies to see how changes in the environment over years affect dengue cases. It would also be helpful to compare and we can see how dengue spreads in different parts of Dhaka or in other cities to understand if some places are more at risk because of their specific conditions. Testing new ways to stop and control dengue such as different types of mosquito traps or community health programs. It could provide useful awareness. Finally, deeper studies into how people's behavior and community practices give to the spread of dengue can help tailor more effective public health intercede.

REFERENCES

[1] Fatema Mehejabin, Mohammad Arafat Rahman, Rumana Rashid Md Sahidur Rahman, "A case-control study to determine the risk factors of dengue fever in ," *Public Health in Practice*, pp. 1-6, 2022.

[2] Cory Morin, Marta S. Shocket, Kris Ebi, Jeremy Hess Kholood K. Altassan, "Dengue fever in Saudi Arabia: A review of environmental and population factors impacting emergence and spread," *Travel Medicine and Infectious Disease*, pp. 1-8, 2019.

[3] J. M. Misti, M. D. Griffiths, D. Gozal M. A. Mamun, "The dengue epidemic in Bangladesh: risk factors and actionable items," *The Lancet*, vol. 394, pp. 2149-2150, December 2019.

[4] C. S. Tang, L. McIver, M. Hashizume, V. Chan, R. R. Abeyasinghe, S. Iddings, R. Huy Y. Choi, "Effects of weather factors on dengue fever incidence and implications for interventions in Cambodia," *BMC Public Health*, vol. 16, p. 241, March 2016.

[5] W. Hu, C. Hurst, S. Tong S. Banu, "Dengue transmission in the Asia-Pacific region: impact of climate change and socio-environmental factors," *ropical Medicine and International Health*, vol. 8, pp. 934-943, Auguest 2011.

[6] A. C. Comrie, K. C. Ernst C. W. Morin, "Climate and dengue transmission: evidence and implications," *Environmental Health Perspectives*, vol. 121, pp. 11-12, December 2013.

[7] S. W. Bakar, K. M. Ismail R. N. Ahmad, "Influence of environmental factors on dengue," *Journal of Environmental Management*, vol. 197, pp. 272-278, March 2017.

[8] E. Viennet, K. Glass, D. Harley S. Sharmin, "The emergence of dengue in Bangladesh: epidemiology, challenges and future disease risk," *Transactions of the Royal Society of Tropical Medicine and Hygiene*, vol. 109, p. 10, October 2015.

[9] P. Dale, J. S. Mackenzie, J. McBride, K. Mengersen, S. Tong S. Naish, "Climate change and dengue: a critical and systematic review of quantitative modelling approaches," *BMC Infectious Diseases*, vol. 14, p. 167, April 2014.

[10] T. Abir et al., "Dengue in Dhaka, Bangladesh: Hospital-based cross-sectional KAP assessment at Dhaka North and Dhaka South City Corporation area," *PLOS ONE*, vol. 16, pp. 1-17, March 2021.

[11] J. M. Misti, M. D. Griffiths, and D. Gozal M. A. Mamun, "The dengue epidemic in Bangladesh: risk factors and actionable items," *The Lancet*, vol. 394, p. 2149–2150, December 2019.

[12] S. Bhatt et al., "The global distribution and burden of dengue," *Nature*, vol. 7446, pp. 504-507, April 2013.

[13] E.-E. Ooi, O. Horstick, B. Wills A. Wilder-Smith, "Dengue," *The Lancet*, vol. 393, p. 10169, January 2019.

[14] M. Khatun, F. Arjumand, A. Nisaluk, R. F. Breiman M. A. Hossain, "Serologic evidence of dengue infection before onset of epidemic, Bangladesh," *Emerging Infectious Diseases*, vol. 9, p. 1411–1414, November 2003.

[15] E. Viennet, K. Glass, D. Harley S. Sharmin, "The emergence of dengue in Bangladesh: epidemiology, challenges and future disease risk," *Transactions of The Royal Society of Tropical Medicine and Hygiene*, vol. 109, p. 619–627, September 2015.

[16] S. Murray, M. Quam A. Wilder-Smith, "Dengue Virus Disease," *Epidemiology of dengue: past, present and future prospects*, vol. 5, pp. 299-309, August 2013.

[17] H. Fuadzy et al., "Risk factors associated with Dengue incidence in Bandung, Indonesia: a household based case-control study," *Heal Sci J Indones*, vol. 11, p. 45–51, June 2020.

[18] T. Nguyen-Tien et al., "Risk factors of dengue fever in an urban area in Vietnam: a case-control study," *BMC Public Health*, vol. 21, pp. 1-13, April 2021.

[19] A. T. Aziz et al., "Promoting health education and public awareness about dengue and its mosquito vector in Saudi Arabia," *Parasites Vectors*, vol. 7, pp. 1-2, November 2014.

[20] S. Bhatt et al., "The global distribution and burden of dengue," *Nature*, vol. 496, p. 504–507, April 2013.

[21] M. R et al., "Dengue outbreak 2019: clinical and laboratory profiles of dengue virus infection in Dhaka city," *Heliyon*, vol. 7, June 2021.

[22] S. Mahmood B. Mahmood, "Emergence of Dengue in Bangladesh: a major international public health concern in recent years," *J Environ Res Manag*, vol. 2, p. 35–41, 2011.

[23] T. DT et al., "Risk factors associated with an outbreak of dengue fever/dengue haemorrhagic fever in Hanoi, Vietnam," *Epidemiol Infect*, vol. 143, p. 1594–1598, March 2015.

[24] M. Bhatt, D. Biswal, S. Pati, R. J. Soares Magalhaes S. Swain, "Risk factors for dengue outbreaks in Odisha, India: a case-control study," *J Infect Public Health*, vol. 13, p. 625–631, April 2020.

[25] A. Eldeib, S. Hammad A. H. El-Gilany, "Clinico–epidemiological features of dengue fever in Saudi Arabia," *Asian Pac J Trop Med*, vol. 3, p. 220–223, March 2010.

[26] D. A. Focks T. H. Jetten, "Potential changes in the distribution of dengue transmission under climate warming," *American Journal of Tropical Medicine and Hygiene*, vol. 57, pp. 285-297, 1997.

[27] D. A. T. Cummings, G. E. Glass M. A. Johansson, "Multiyear climate variability and dengue – El Nino southern oscillation, weather, and dengue incidence in Puerto Rico, Mexico, and Thailand: a longitudinal data analysis," *PLoS Medicine*, vol. 6, p. e1000168, 2009.

[28] F. Dominici, G. E. Glass M. A. Johansson, "Local and global effects of climate on dengue transmission in Puerto Rico," *PLoS Neglected Tropical Diseases*, vol. 3, p. e382, 2009.

[29] D. A. Focks T. H. Jetten, *Potential Changes in the Distribution of Dengue Transmission under Climate Warming*. Arlington, United States: American Journal of Tropical Medicine and Hygiene, 1997.

[30] W. J. M. Martens, D. A. Focks, T. H. Jetten J. A. Patz, *Dengue Fever Epidemic Potential as Projected by General Circulation Models of Global Climate Change*. Washington, D.C., United States: Environmental Health Perspectives, 1998.

[31] Impact of Regional Climate Change on Human Health, *J. A. Patz, D. Campbell-Lendrum, T. Holloway, J. A. Foley*. London, United Kingdom: Nature, 2005.

Appendix A

Course Outcomes, Complex Engineering Problems (EP) and Complex Engineering Activities (EA) Addressing

Title: HEALTH – EXPLORING THE RELATIONSHIP BETWEEN DENGUE AND ENVIRONMENTAL FACTORS IN DHAKA USING MACHINE LEARNING

Student ID: [201-15-14081]

CO Description for FYDP

CO	CO Descriptions	PO
Phase -I		
CO1	Integrate recently gained and previously acquired knowledge to identify a **– Exploring The Relationship Between Dengue And Environmental Factors In Dhaka Using Machine Learning** problem for the Final Year Design Project (FYDP)	**PO1**
CO2	Analyze different aspects of the goals in designing a solution for this FYDP	**PO2**
CO3	Explore diverse problem domains through a literature review, delineate the issues, and establish these goals for the FYDP	**PO4**
CO4	Perform economic evaluation and cost estimation and employ suitable project management procedures throughout the development life cycle of the FYDP	**PO11**
Phase -II		
CO5	Design and develop technical solutions and system components or processes that meet specified requirements, ensuring compliance with public health and safety standards, as well as considering cultural, socioeconomic, and environmental factors in this FYDP	**PO3**
CO6	Choose and apply appropriate methodologies, resources, and contemporary engineering and IT technologies to address complex engineering processes, encompassing prediction and modeling, while adhering to relevant constraints in this FYDP	**PO5**
CO7	Analyze societal, health, safety, legal, and cultural considerations, along with associated responsibilities, in the context of professional engineering practice and the resolution of this problem, employing logical reasoning guided by contextual understanding.	**PO6**

CO8	Comprehend and evaluate the enduring sustainability and impact of professional engineering endeavors in addressing intricate engineering challenges within social and environmental frameworks.	**PO7**
CO9	Implement ethical principles and adhere to professional standards and norms in this FYDP	**PO8**
CO10	Capable of operating proficiently both individually and as a team member or leader across diverse teams and interdisciplinary settings in this FYDP.	**PO9**
CO11	Proficiently communicate with the engineering community and broader society regarding complex engineering endeavors, including the ability to comprehend and generate comprehensive reports and design documentation, as well as provide and receive clear instructions throughout this FYDP.	**PO10**
CO12	Acknowledge the importance of self-directed and life-long learning within the evolving landscape of technology, and possess the readiness and capability to engage in lifelong learning endeavors.	**PO12**

Addressing CO (1 to 8), Knowledge Profile (K), Attainment of Complex Engineering Problems (EP), and Attainment of Complex Engineering Activities (EA)

Addressing CO (1 to 8), Knowledge Profile (K), Attainment of Complex Engineering Problems (EP):

SN	EP Definition	Attainment	CO	Justification (With Knowledge Profile)	References
1.	**EP1: Depth of Knowledge required**	Yes	CO1, CO2, CO3, CO5, CO6, CO7 and CO8	The project covers **Engineering Fundamentals (K3)** through process modeling, data design, and front-end development, showcasing strong engineering principles application. The project tackles **Specialist Knowledge (K4)** by integrating front-end and back-end design, interaction design, and user experience (UX), utilizing HTML, CSS, SCSS, jQuery, Vue JS for the front end, and PHP frameworks like Laravel for the back end, demonstrating expertise in web-based travel management system development.	**Page no:** [12,14-16,21], **Section:** [3.1, 3.3, 3.4, 5.1, 5.2] **Page no:** [17-24],

					Section: [4.1, 4.2, 4.3, 5.1, 5.2, 5.3]
				The project applies **engineering practice & design (K5)** by integrating design requirements like user interface, security, scalability, customization, compatibility, accessibility, and reliability, ensuring a robust and user-centric travel management system aligned with design principles. The project addresses **engineering practice & technology (K6)** by employing HTML, CSS, SCSS, jQuery, Vue JS, PHP, Laravel, MySQL, and Maria DB, emphasizing safe, scalable development with Laravel's MVC framework and Composer.	**Page no:** [12-16], **Section:** [3.1, 3.3, 3.4, 3.5] **Page no:** [17-20], **Section:** [4.1, 4.2, 4.3, 4.4]
				The project covers **research literature K8** by reviewing analogous platforms, UI/UX studies, security protocols, personalization algorithms, and industry trends, enabling informed decision-making for integrating innovative features into this system.	**Page no:** [8-10], **Section:** [2.2, 2.3]

2.	**EP2: Range of Conflicting Requirements**	Yes	CO2, and CO7	The project confronts **EP-2** by addressing challenges like integrating diverse features while maintaining user experience, balancing security with functionality, scalability, real-time data synchronization, evolving industry standards, privacy regulations compliance, and implementing effective recommendation systems through careful planning, flexible development, and continuous adaptation.	**Page no:** [11] **Section:** [2.4, 2.5]
3.	**EP3: Depth of analysis requi red**	Yes	CO2, and CO6	The project tackles **EP-3** by prioritizing solutions like contemporary front-end design for enhanced user experience and selecting Laravel for streamlined back-end development, addressing challenges such as user interaction optimization, platform accessibility, and scalability to provide a superior web-based travel management system.	**Page no:** [17-20], **Section:** [4.2, 4.3, 4.4]
4.	**EP4: Familiarity of Issues**	Yes	CO8	The project fulfills **EP-4** by integrating insights from travel management into development, optimizing processes, enhancing user experience, ensuring data security, and contributing to industry efficiency and effectiveness, bridging Computer Science and Engineering with the travel domain for impact beyond CSE boundaries.	**Page no:** [8-11], **Section:** [2.2, 2.4]
5.	**EP5: Extends of application codes**	No	CO5	N/A	N/A

6.	**EP6: Extends of stakeholders involved and conflicting requirements**	No	CO8	N/A	N/A
7.	**EP7: Interdependence**	Yes	CO5	The project meets CEP, **EP-7** by addressing high-level problems across stages, ensuring data integrity, security, and performance in database implementation, optimizing user experience through intuitive front-end design, and conducting comprehensive testing.	**Page no:** [21-23], **Section:** [5.1, 5.2]

Addressing CO11 with Complex Engineering Activities (EA) [Some or all of the following]:

SN	EA Definition	Attainment	CO	Justification	References
1.	**EA1: Range of resources**	Yes	CO11	The development of this system leverages a skilled team across domains, meticulous financial resource allocation for infrastructure, marketing, and development, and utilizes tools like XAMPP, Visual Studio Code, and various programming languages, frameworks, and databases for effective execution.	**Page no:** [20], **Section:** [4.4]
2.	**EA2: Level of interaction**	No		N/A	N/A
3.	**EA3: Innovation**	No		N/A	N/A

4.	**EA4: Consequences for society and the environment**	Yes		This project emphasizes social impact and environmental sustainability by enhancing trip efficiency, promoting sustainable travel practices, and ensuring ethical considerations such as strong security, privacy, and transparent business practices, fostering societal well-being and environmental stewardship.	**Page no:** [26-31], **Section:** [6.1, 6.2, 6.4]
5.	**EA-5: Familiarity**	Yes		The Comparative Analysis for "Health - Web-based EHR " evaluates market competitors like Expedia and Booking.com to highlight health's unique features, security, personalization, and customer engagement, enriching market understanding and informing strategic decisions for enhanced competitiveness and value proposition.	**Page no:** [7-10], **Section:** [2.1, 2.3]

Addressing CO (4, 9, 10, and 12):

SN	COs	Attainment	Justification	References
1	CO4	Yes	The project captures **CO4** by integrating financial management practices, including budgeting, resource allocation, cost management, and risk mitigation, ensuring efficient fund distribution, monitoring expenditures, and identifying optimization areas while maintaining quality.	**Page no:** [4-5], **Section:** [1.5]
2	CO9	Yes	The project upholds ethical practices by prioritizing user privacy, maintaining	**Page no:** [28],

			transparency in data usage, providing fair suggestions, respecting cultural sensitivities, and promoting accessibility, fostering trust, transparency, and integrity among stakeholders.	**Section:** [6.3]
3	CO10	No	N/A	N/A
4	CO12	Yes	The project's implementation phase emphasizes lifelong learning through diverse testing protocols, automated frameworks, and CI/CD pipelines, ensuring ongoing improvement and adaptation to emerging tech trends, with results analysis refining the platform for enhanced effectiveness.	**Page no:** [21-25], **Section:** [5.1, 5.4]

ORIGINALITY REPORT

14%	11%	9%	6%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	fastercapital.com Internet Source	2%
2	Submitted to University of North Texas Student Paper	1%
3	dspace.daffodilvarsity.edu.bd:8080 Internet Source	1%
4	worldwidescience.org Internet Source	1%
5	www.researchgate.net Internet Source	1%
6	da Cunha Ceulin, Wagner Luiz. "Knowledge Elicitation in Deep Learning Models", Universidade do Porto (Portugal), 2024 Publication	<1%
7	Submitted to University of Auckland Student Paper	<1%
8	Submitted to Brunel University Student Paper	<1%
	link.springer.com	