

DENGUE FEVER PREDICTION USING MACHINE LEARNING TECHNIQUES

BY

MD: Kamrul Hasan Mozumder

ID: 201-15-3248

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

Supervised By

Ms. Lameya Islam

Lecturer

Department of CSE

Daffodil International University

Co-Supervised By

Mohammad Asifur Rahim

Lecturer

Department of CSE

Daffodil International University



DAFFODIL INTERNATIONAL UNIVERSITY

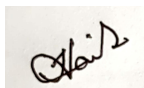
DHAKA, BANGLADESH

JANUARY, 2024

APPROVAL

This Project titled “**Dengue Fever Prediction Using Machine Learning Techniques**” submitted by **MD: Kamrul Hasan Mozumder ID: 201-15-3248** to the Department of Computer Science and Engineering, Daffodil International University, has been acknowledged as satisfactory for its style and substance and accepted as being sufficient for the accomplishment of the requirements for the degree of Bachelor of Science in Computer Science and Engineering.

BOARD OF EXAMINERS



Chairman

Dr. Sheak Rashed Haider Noori (SRH)

Professor and Head

Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

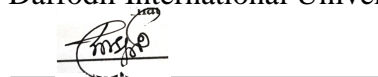


Internal Examiner

Nazmun Nessa Moon(NNM)

Associate Professor

Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University



Internal Examiner

Dewan Mamun Raza (DMR)

Senior Lecturer

Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University



External Examiner

Dr. Md. Arshad Ali (DAA)

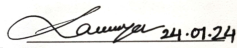
Professor

Department of Computer Science and Engineering
Hajee Mohammad Danesh Science & Technology
University

DECLARATION

I hereby declare that this project has been done by me under the supervision of **Ms. Lameya Islam, Lecturer, Department of CSE, Daffodil International University**. I also declare that neither this project nor any part of this project has been submitted elsewhere for award of any degree or diploma.

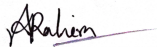
Supervised by:

 24.01.24

Ms. Lameya Islam

Lecturer
Department of CSE
Daffodil International University

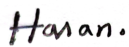
Co-Supervised by:



Mohammad Asifur Rahim

Lecturer
Department of CSE
Daffodil International University

Submitted by:



Md. Kamrul Hasan Mozumder

ID: 201-15-3248
Department of CSE
Daffodil International University

ACKNOWLEDGEMENT

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

I really grateful and wish our profound our indebtedness to **Ms. Lameya Islam, Lecturer, Department of CSE, Daffodil International University, Dhaka, and Co-supervisor Mohammad Asifur Rahim, Lecturer, Department of CSE, Daffodil International University, Dhaka.** I'd like to thank my supervisor's deep knowledge and strong interest in Machine Learning for helping me to complete this project. This project could not have been done without her endless patience, expert advice, constant encouragement, constant and energetic supervision, constructive criticism, good advice, reading many bad draughts, and fixing them at every stage.

I would like to express my heartiest gratitude to **Dr. Sheak Rashed Haider Noori, Professor and Head, of the Department of CSE, and Dr. S. M. Aminul Haque, Professor and Associate Head, of the Department of CSE** for their kind help to finish my project and also, to all of the faculty members and the staff of the CSE department of Daffodil International University for kind support and help on the technical and the administrative aspect of the study.

I would like to extend my gratitude to all of my classmates from Daffodil International University who participated in this conversation while they were working on their course assignments.

Finally, I would like to express my gratitude to my parents for the everlasting love and patience. They have shown me throughout my life. They have been there for me throughout my whole life to support me and engage me in all I do.

ABSTRACT

The rapid and broad spread of the virus that causes dengue is a hallmark of the dengue, a global epidemic, a worldwide health problem. Using data from Kaggle, this project is about predicting Dengue Fever through machine learning approaches. Dataset has 33 variables, contain a variety of symptoms like itching, rashes on the skin, joint discomfort, high temperature, and more. "Prognosis," the target attribute, divides occurrences into three classes: typhoid, common cold, and dengue, with counts of 209, 143, and 103, respectively. AdaBoostClassifier, BernoulliNB, GaussianNB, DecisionTreeClassifier, BaggingClassifier, as well as Voting Classifier are a few of the machine learning models used for prediction. The BaggingClassifier algorithm performed quite well in this instance, with the greatest accuracy of 95.87%. The process is methodical and starts with the selection of information from Kaggle. Next, preprocessing operations such as encoding and missing value removal are carried out. EDA, or exploratory data analysis, provides light on the properties of the dataset. Algorithms such as AdaBoostClassifier, DecisionTreeClassifier, and many more are used in model training. Evaluation measures measure the performance of the models and include recall, accuracy, and precision. Testing the algorithms to determine how well they can forecast the real world is the last phase.

Keywords: Machine Learning, Predictive Modeling, Dengue Fever, Decision Tree Classifier, BaggingClassifier, GaussianNB, BernoulliNB, and AdaBoostClassifier

TABLE OF CONTENTS

CONTENTS	PAGE
Board of examiners	ii
Declaration	iii
Acknowledgements	iv
Abstract	v
CHAPTER	
CHAPTER 1: INTRODUCTION	1-5
1.1 Introduction	1-2
1.2 Motivation	2
1.3 Research Questions	2-3
1.4 Main Objective	4
1.5 Expected Outcome	4-5
1.6 Report Layout	5
CHAPTER 2: BACKGROUND	6-13
2.1 Preliminaries	6
2.2 Related Works	6-11
2.3 Comparative Analysis and Summary	11-12
2.4 Scope of the Problem	12
2.5 Challenges	13
CHAPTER 3: RESEARCH METHODOLOGY	14-21
3.1 Research Subject and Instrumentation	14
3.2 Data Collection Procedure	14-15
3.3 Statistical Analysis	15-21
3.4 Implementation Requirements	21

CHAPTER 4: EXPERIMENTAL RESULTS AND DISCUSSION	22-32
4.1 Experimental Setup	22
4.2 Experimental Results & Analysis	22-24
4.3 Accuracy	24-32
CHAPTER 5: IMPACT ON SOCIETY, ENVIRONMENT AND SUSTAINABILITY	33-35
5.1 Impact on Society	33
5.2 Impact on Environment	33-34
5.3 Ethical Aspects	34
5.4 Sustainability Plan	34-35
CHAPTER 6: SUMMARY, CONCLUSION, RECOMMENDATION AND IMPLICATION FOR FUTURE RESEARCH	36-37
6.1 Summary of the Study	36
6.2 Conclusions	36-37
6.3 Implication for Further Study	37
APPENDIX	38
REFERENCES	39-40
PLAGIARISM REPORT	41

LIST OF FIGURES

FIGURES	PAGE NO
Figure 3.1: Dataset Dengue Fever Pie Chart	15
Figure 3.2: Methodology Flowchart	16
Figure 3.3: Adaboost model architecture	18
Figure 3.4: BernoulliNB model architecture	19
Figure 3.5: GaussianNB architecture	20
Figure 3.6: Decision Tree Architecture	21
Figure 3.7: Bagging Classifier Architecture	22
Figure 3.8: Voting Classifier Architecture	23
Figure 4.1: Accuracy of Machine Learning Models	27
Figure 4.2: Confusion Matrix Bagging Classifier	29
Figure 4.3: Confusion Matrix BernoulliNB	30
Figure 4.4: Confusion Matrix Decision Tree	31
Figure 4.5: Confusion Matrix Voting Classifier	32
Figure 4.6: Confusion Matrix Adaboost Classifier	33
Figure 4.7: Confusion Matrix GaussianNB	34

LIST OF TABLES

TABLES	PAGE NO
Table 1.1: PROJECT MANAGEMENT TABLE	5
Table 2.1: Accuracy Comparison Table	12
Table 4.1. Performance Evaluation	27
Table 4.2. Performance Evaluation (AdaBoost)	28
Table 4.3. Performance Evaluation (BernoulliNB)	29
Table 4.4. Performance Evaluation(GaussianNB)	30
Table 4.5. Performance Evaluation(Decision Tree)	31
Table 4.6. Performance Evaluation(Bagging)	32
Table 4.7. Performance Evaluation(Voting)	33

CHAPTER 1

Introduction

1.1 Introduction

Infectious symptoms and prevalence provide significant dangers to global public health. The most dangerous of these is Dengue Fever, which affects millions of people each year and is most common in tropical and subtropical areas. The dengue virus, known as dengue fever, is mostly transmitted by Aedes mosquitoes. The sickness can produce a wide range of symptoms, from an ordinary case of the flu to serious complications including dengue hemorrhagic fever, or DHF, and dengue stress syndrome (DSS), which can have a major detrimental effect on health. Accurate prediction and quick reaction are essential to reducing its impact on harmed populations because of its unpredictable nature and potential for outbreaks.[1]

The deployment of machine learning techniques has made health surveillance and prediction more effective. Promising possibilities for Dengue Fever prediction can be found by utilizing these approaches, especially when analyzing symptoms, environmental variables, and historical outbreak data. Machine learning algorithms are able to identify complex patterns and connections through the combination of several datasets that include clinical symptoms, meteorological data, and geographical characteristics. This allows for the early detection of possible outbreak hotspots and supports focused preventive efforts.[2]

Predictive models are built on top of Web Source datasets, which have become known for their diversity and richness. But this project requires careful data preprocessing, which includes encoding categorical variables, feature engineering, and handling missing values. Exploratory data analysis, or EDA, also plays a crucial role in this process, revealing patterns, trends, and outliers in the dataset to assist with choosing features and model building.[3]

The purpose of this research is to use machine learning techniques to explore the predictive field of dengue fever. By means of an organized approach that includes data curation, preprocessing, modeling, and assessment, this study aims to make a contribution to the field of surveillance of infectious diseases and preventive health care. The objective is to develop accurate models that can predict Dengue outbreaks by combining epidemiological knowledge with technology developments, which will ultimately promote public health response and preparedness activities.[4]

1.2 Motivation

The urgent requirement to address the serious health burden presented by this dangerous illness is the motivation behind the use of machine learning approaches for predicting Dengue Fever. The uncertainty, swift spread, and serious consequences associated with dengue fever highlight the need for proactive approaches to disease surveillance and treatment. The potential to predict Dengue outbreaks is significant to minimize the disease's negative effects on exposed groups. By identifying at-risk areas early on, resources may be assigned, vector control strategies can be focused, and public health actions may potentially reduce the disease's severity and spread. In addition, the usage of machine learning techniques to illness prediction is consistent with the evolving environment of healthcare innovation. The utilization of technology for the analysis of intricate datasets that comprise many components like symptoms, environmental factors, and historical trends is a viable route to improving conventional epidemiological techniques. This collaborative strategy combines medical knowledge with computing power to create a synergy that has the potential to completely transform the treatment of infectious diseases. The ultimate goal of developing machine learning methods for Dengue Fever prediction is to provide health care providers with predictive power so they can take preventative action that will protect populations, reduce the burden of disease, and save lives.

1.3 Rationale of the Study

The reason for this study is the pressing need for safeguards to counteract dengue fever, a serious danger to global health. The goal is to develop prediction models using machine learning approaches that can anticipate Dengue outbreaks, enabling prompt interventions and resource allocation in risky areas. Ancient epidemiological techniques frequently encounter difficulties in identifying the irregular and complex patterns of Dengue Fever spread. Machine learning presents a viable approach to combine several datasets that include symptoms, environmental factors, and historical trends, with the ability to identify complex patterns that are beyond the scope of human knowledge. The possibility to improve public health response and ability lends even more weight to the study's justification. By identifying outbreak-prone locations early on, specific actions can be implemented, such as vector control strategies and the distribution of healthcare resources, which may slow the disease's spread and lessen its effects on impacted communities. In addition, a proactive move toward data-driven decision-making in healthcare is represented by the fusion of epidemiological insights with technology improvements, which marks a paradigm shift in infectious disease control and management techniques.

1.4 Research Question

1. Can machine learning models accurately predict Dengue Fever outbreaks based on symptomatology and environmental factors?
2. How effectively can historical data and meteorological information forecast the spatial and temporal spread of Dengue Fever?
3. What are the key factors contributing to the early identification and prediction of severe Dengue cases using machine learning techniques?
4. How does the integration of diverse datasets enhance the precision of Dengue Fever prediction models?
5. Which machine learning algorithms demonstrate superior performance in predicting Dengue outbreaks from Kaggle-derived datasets?

6. What role do specific symptoms and their combinations play in the predictive accuracy of Dengue Fever models?

7. Can the developed predictive models aid in targeted interventions and resource allocation for mitigating Dengue Fever's impact on vulnerable populations?

1.5 Expected Output

The study's expected outcomes cover a number of significant areas. First of all, it expects the creation of reliable algorithms based on machine learning that will be able to accurately project Dengue Fever plagues. These models are expected to forecast the temporal and spatial development of the disease by utilizing a variety of datasets that include symptoms, variables related to the environment, and trends in the past. In addition, the research aims to identify critical contributing characteristics that have a major influence on these models' prediction ability. It is expected that this knowledge would draw attention to particular symptoms, environmental factors, or their combinations that are highly correlated with the frequency and intensity of dengue fever outbreaks. Also, the outputs that are expected to be produced are verified predictive models that may help policymakers and healthcare authorities carry out specific measures and resource allocation plans. In the end, the study hopes to improve public health ability by offering practical advice for preventive management of Dengue Fever to minimize the illness's effects on sensitive groups.

1.6 Project Management and Finance

Agile project management will be used, allowing for iterative development and regular creation of feedback. The project schedule consists of four distinct phases, each with assigned deliverables and milestones: validation, deployment, model creation, data collecting and preprocessing, and deployment. Coordination and progress monitoring will be ensured by holding regular team meetings. In recognition of support, the project's primary focus will be on computing infrastructure for model creation and validation; cloud services may be utilized to facilitate development. In addition, money will be set aside for access fees or the purchase of datasets, if necessary. A detailed accounting plan will be devised to ensure the most efficient use of resources and to account for any unanticipated

costs or extra data acquisition requirements that may arise throughout the project's deployment.

TABLE 1.1: PROJECT MANAGEMENT TABLE

Work	Time
Dataset	1 month
Literature Review	3 months
Experiment Setup	1 month
Implementation	2 months
Report	2 months
Total	9 months

1.7 Report Layout

- Introduction
- Background
- Data Collection
- Data Preprocessing
- Research Methodology
- Experimental Result and Discussion
- Impact on Society, Environment
- Summary, Conclusion, Future Research
- References

CHAPTER 2

Background Study

2.1 Preliminaries

This project's start includes an in-depth assessment of the literature and the collection of data. The purpose of the literature review is to examine current studies, approaches, and developments in machine learning-based dengue fever prediction. It provides the framework for understanding relevant ideas, spotting gaps, and improving the study approach. Additionally, data gathering and acquisition from acceptable sources—possibly the Kaggle database datasets or other dependable repositories—begins. By ensuring the availability and caliber of a variety of datasets necessary for model building and validation, this step builds the foundation for later data preprocessing.

2.2 Related Works

An essential first step in every research effort is evaluating the body of current knowledge. The goal of this review of the literature is to give a thorough overview of the current scientific discussion on dengue fever prediction. Using knowledge from a variety of domains, including public health, machine learning, and epidemiology, our analysis explores current research initiatives. The analysis that follows offers a synthesis of the major discoveries and contributions made by these investigations, presenting a comprehensive picture of the current status of dengue fever prediction in the scholarly and scientific community. We have studied some recent research works and the analysis of those studies are given below:

Kannimuthu, S., et al. [5] performance analysis of several prediction algorithms, including Naive Bayes, Neural Network, K-Nearest Neighbor, Support Vector Machine, and Multilayer Perceptron, was done in the context of the prediction of dengue disease. The results showed that the C4.5 algorithm outperformed these alternatives regularly, displaying outstanding results in terms of accuracy (89.33%), prediction (88.9%), recall (89.77%), and other important performance parameters. In comparison to a number of other

common techniques, this research highlights the C4.5 algorithm's superior predictive skills in the context of dengue disease prediction.

Sanjudevi, et al. [6] aimed to enhance the prediction accuracy of dengue disease through a feature model construction and comparative analysis conducted in three phases. The first phase involved collecting dengue disease datasets from the UCI repository. In the second phase, feature selection was performed using forward and backward stepwise regression methods. In the third phase, the dataset was subjected to classification using SVM and decision tree algorithms, resulting in SVM achieving the highest accuracy of 99% with a shorter runtime and a minimal error rate, making it the top-performing classifier.

Iqbal, et al. [7] create an effective method for detecting cases of dengue in a large sample of data. The accuracy of seven well-known machine learning systems was tested. With the use of eight different performance metrics, these strategies were evaluated. The most successful model was the LogitBoost ensemble model, which proved its excellent ability in predicting cases of dengue by reaching a classification accuracy of 92% while maintaining a sensitivity of 90% and specificity of 94%.

Saturi, et al. [8] organized the dataset, the k-means clustering algorithm may divide patient records for dengue disease into k groups, boosting analysis and forecasting of the disease. Additionally, a fuzzy-based method was used to convert input variables into flexible membership functions, enhancing forecasting models for dengue. LogitBoost had the highest sensitivity of the tested algorithms, scoring 90%, followed by Decision Tree, scoring 87%. LogitBoost performed exceptionally well in terms of specificity, earning a top score of 94%, and precision, achieving an amazing 95% accuracy, demonstrating its efficacy in the prediction of dengue disease. Lathesparan, et al.[9] applied Component Analysis and the Wrapper feature selection method, together with Naive Bayes, K-Nearest Neighbor, and J48 algorithms, were the main focus of this study's feature selection techniques. For their compatibility with partial datasets, simple artificial neural networks were used to assess each feature selection method's performance. The initial 22-dimensional system was effectively reduced using PCA to an 8-dimensional one, accounting for 59% of the total variance. The accuracy for ANN with PCA was the highest (72.47%), while the accuracy for ANN with KNN was the lowest (54.47%). Based on the

ANN accuracy, the research came to the conclusion that PCA was the best feature selection technique for this dataset.

Nishanthi, et al. [10] figure out parameters and exact time lags of this study's individual variables and the number of current Dengue cases were identified. As a result of this investigation, an Artificial Neural Network was created, displaying the ability to reasonably accurately forecast cases of dengue in Sri Lanka's Kandy district. The ANN's ability to predict dengue outbreaks was shown by the regression coefficient of 0.79 in the output, which showed a strong association with the actual number of dengue cases.

Mello-Román, Jorge D., et al. [11] tested Artificial Neural Networks and Support Vector Machines using an actual dataset of dengue patients from Paraguay's public health system between 2012 and 2016 in a study evaluating machine learning techniques for medical diagnosis. With excellent average performance of 96% accuracy, 96% sensitivity, and 97% specificity over thirty different dataset partitions, the ANN multilayer perceptron surpassed SVM. SVM, on the other hand, performed well, routinely outperforming 90% accuracy using a polynomial kernel. These results point to the potential of ANN as a highly precise tool for dengue medical diagnosis.

Silitonga, et al. [12] used Artificial Neural Network and Discriminant Analysis models to predict the dengue severity level in patients based on the findings of laboratory tests. The ANN models, in particular those using logistic and hyperbolic tangent activation functions with 70% training data, achieved the best accuracy at 90.91%, with a sensitivity of 91.11% and a specificity of 95.51%, considering the small dataset size. By predicting the severity level of dengue patients before they reach the critical stage, the suggested approach may be useful for doctors to ensure immediate and effective medical treatment. Sutrisna, Bambang, et al. [13] created using multivariate logistic regression analysis, taking into account elements like the number of days of fever, the presence of muscle pain, the results of the tourniquet test, the total number of white blood cells, the number of monocytes, and the number of platelets. The Receiver Operating Characteristic curve was used to evaluate the model's performance in comparison to the 1997, 2009, and 2011 WHO dengue classification standards. Patients with a total score of less than 14 were the most likely to have probable dengue sickness, showing sensitivity, specificity, and a positive probability

ratio of 79.7%, 60.0%, and 1.99, respectively. With sensitivity, specificity, and a positive likelihood ratio of 79.7%, 68.0%, and 2.48, respectively, a total score of 7 indicated a probable dengue diagnosis.

Sippy, Rachel, et al. [14] predictive model for dengue fever identified the day of the week, the hospital where cases were diagnosed, long-term trends represented by a spline function spanning the entire study period, and annual sinusoidal fluctuations in the disease as the main determinants of its occurrence. An interesting finding from the data was the identification of a seasonal trend with single peaks in case diagnoses occurring around mid-March. In addition, compared to the average of all days, Tuesdays and Thursdays had higher risk ratios for dengue diagnosis (Risk Ratio: 1.26 and 1.25, respectively). On the other hand, compared to other days of the week, diagnoses were less likely to occur on Saturdays (Risk Ratio: 0.81) and Sundays (Risk Ratio: 0.74).

Mariki, et al. [15] included a dataset of 2556 instances and 36 variables. Using the random forest classifier, the most important characteristics for malaria prediction were chosen, with a special emphasis on geographical variations. However, with variable ranks in regional datasets, fever appeared as the most significant factor for malaria prediction across all regions, followed by general body malaise, vomiting, and headache. Six predictive models that used these crucial parameters were tested; machine learning classifiers like Random Forest demonstrated the highest performance accuracy, coming in at 82%, while the average accuracy for all models was 70%. Performance accuracy increased for all classifiers when the dataset was reduced to only include significant features, with Random Forest reaching 86%.

Choubey, et al. [16] focused on enhancing prediction effectiveness, this study suggests employing the Enhanced Back Propagation with Artificial Neural Network algorithm for early infection identification. A viral dataset is gathered, preprocessed using Z-score normalization, and its features are extracted using a dynamic angle projection pattern. A genetic algorithm is used for feature selection. The proposed algorithm significantly improves disease prediction accuracy, outperforming existing approaches like ML (55%), LSTM (64%), KNN (84%), and NBN (75%), obtaining an astonishing 94%.

Li, Fudong, et al. [17] differentiate between 25 common infectious diseases, a classification model using a naive Bayesian classifier was created using data from previous diseases. The sensitivity of the 146-predictor model ranged from 44.44% for hepatitis E to 96.67% for measles, and it had a total of 146 predictors. There were differences in specificity as well, with five diseases obtaining 100% and dengue fever having the lowest specificity (96.36%). The model's overall performance showed a median total accuracy of 97.41%, with AUC values over 0.98 for all but one of the conditions (dengue fever, which had an AUC value of 0.613). A value of 0.960 (95% CI: 0.941-0.978) was determined for the M-index.

Baker, et al. [18] studied weekly disease cases in the cities of San Juan and Iquitos were predicted using data connected to the weather. Using the Mean Absolute Error, several machine learning regression techniques were used and evaluated. The Poisson Regression Model was shown to be the highest-performing model with an exceptional average absolute error reduction of 25.6%, the lowest MAE, and the best ratios. This result implies that the Poisson Regression Model may accurately predict sickness occurrences in these cities based on weather data.

Shakil, et al [19] used the Explorer, Information Flow, and Experimenter interfaces of the Weka software to do a classification analysis of a dengue dataset using several data mining techniques. Weka used 99 rows and 18 attributes from the dataset's 108 occurrences to forecast the outcomes of diseases. The results showed that Naive Bayes and J48 algorithms performed best, with a magnificent accuracy rate of identified examples, and a maximum ROC (Receiver Operating Characteristic) value of 1. The accuracy with which Naive Bayes and J48 can categorize dengue sickness cases is demonstrated by these results.

Mariki, et al. [20] used patient symptoms and historical data as input variables to develop supervised machine learning models for diagnosing malaria. Data on malaria diagnoses were included in the dataset, which were gathered in Tanzania's Morogoro and Kilimanjaro areas. In order to improve model performance and shorten processing time, feature selection was done. K-fold cross-validation was used to train and validate a variety of machine learning classifiers. The dataset, which consisted of 2556 patient records with 36 parameters, led to the development of a malaria diagnosis model including patient

symptoms and demographic features. Additionally, Random Forest was the most successful classifier, with accuracy rates of 95% in Kilimanjaro, 87% in Morogoro, and 82% when the two datasets were combined.

2.3 Comparative Analysis and Summary

The purpose of the comparative analysis is to evaluate how well various machine learning models predict the occurrence of dengue fever outbreaks. The study contrasts the F1-score, recall, accuracy, and precision of models such as DecisionTreeClassifier, AdaBoostClassifier, and others. It draws attention to the advantages and disadvantages of each model in terms of computational efficiency, robustness, and predictive power. The results of the research are summarized in the Summary, which highlights the best models and important factors affecting the prognosis of dengue fever. It describes the importance of particular symptoms, environmental factors, or model architectures that significantly support accurate projections. The research's possible applications are explained in the summary, which also highlights how the study can improve disease surveillance, preventative measures, and public health awareness of impending Dengue Fever epidemics.

TABLE 2.1: ACCURACY COMPARISON TABLE

SN	Author	Dataset	Applied Algorithms	Best Accuracy
1	Kannimuthu, S., et al. [5]	Dengue disease prediction	Naive Bayes, Neural Network, K-NN, SVM, Multilayer Perceptron,	89.33%
2	Iqbal, et al. [7]	Dengue disease detection	LogitBoost ensemble model	92%
3	Lathesparan, et al. [9]	Dengue disease prediction	PCA, Naive Bayes, K-NN	72.47%
4	Nishanthi, et al. [10]	Dengue disease forecasting	Artificial Neural Network	79%
5	Silitonga, et al. [12]	Dengue severity prediction	Artificial Neural Network, Discriminant Analysis	90.91%
6	Sutrisnaet al.[13]	Dengue diagnosis prediction	Multivariate logistic regression	79.7%
7	Mariki, et al. [15]	Malaria prediction	Random Forest, Various classifiers	82%
8	Choubey et al. [16]	Dengue infection identification	Enhanced Back Propagation with ANN	94%
9	Li, Fudong, et al. [17]	Differentiation of infectious diseases	Naive Bayesian classifier	97.41%

2.4 Scope of the Problem

The problem's scope is in using machine learning approaches to forecast Dengue Fever outbreaks, which includes understanding out what critical elements affect the disease's severity and transfer. To create predictive models, this entails examining a variety of datasets that include historical trends, environmental factors, and symptoms. The scope includes assessing the model's performance in terms of accuracy, precision, and recall for predicting the incidence of dengue fever. In addition, by offering timely insights for focused interventions and resource allocation strategies, the study intends to investigate the potential applications of these predictive models in supporting public health authorities and policymakers, ultimately helping to lessen the impact of dengue fever on vulnerable populations.

2.5 Challenges

There are a number of problems in the way of employing machine learning methods for predicting dengue fever epidemics. First off, the disease's complexity and variety demand the use of a variety of data sources, including historical patterns, environmental factors, and symptomatology. This may present difficulties with data integration, preprocessing, and gathering. In addition, it is challenging to accurately record and forecast the dynamics of Dengue epidemics due to their irregular and nonlinear nature. Model creation is further complicated by the possibility of missing or unbalanced datasets as well as the requirement for feature engineering and selection. Last but not least, there is still more work to be done to ensure that predictive models are both practical and useful in real-world settings while taking ethical issues into account.

CHAPTER 3

Research Methodology

3.1 Research Subject and Instrumentation

The aim of the study is applying machine learning to the prediction of dengue fever, with a focus on creating prediction models that make use of symptomatology, environmental variables, and past data. The efficacy of several machine learning methods, including AdaBoostClassifier, DecisionTreeClassifier, and others, in predicting dengue epidemics is investigated in this work. For data preprocessing, model building, and evaluation, instrumentation uses analytic techniques and programming languages such as Python using libraries like scikit-learn, pandas, and matplotlib. Statistical analysis tools can also be used for result interpretation and exploratory data analysis. For the purpose of training, validating, and testing the model, the instrumentation also provides access to trustworthy datasets, which may be obtained via internet source or Google dataset repositories.

3.2 Data Collection

In order to choose the data for this study, reliable websites such as Google dataset were utilized with a particular emphasis on repositories holding pertinent data on dengue fever. The process involved a thorough examination of the datasets, with a focus on aspects like data quality, symptomatology comprehensiveness, and the incorporation of environmental variables and historical records related to Dengue outbreaks. A comprehensive understanding of the dynamics of the disease and the creation of strong predictive models were made possible by the selection criteria, which gave priority to datasets with a variety of features. Accessing and downloading datasets from Google dataset that fit the predetermined criteria was part of the data collection process. In-depth dataset integrity validation was done during this procedure, along with accuracy, completeness, and potential bias checks. Additionally, an attempt was made to guarantee that the datasets included a wide range of features, such as historical records of dengue fever, geographical data, meteorological data, and symptoms. The obtained datasets were then put through

initial quality checks and preprocessing procedures in order to prepare them for additional analysis and model building.

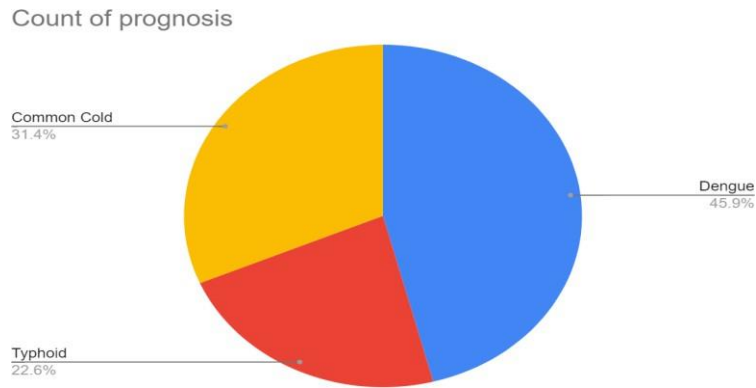


Figure 3.1: Dataset Dengue Fever Pie Chart

The pie chart shows the percentage of common cold, dengue, and typhoid cases. Dengue takes up the largest percentage at 45.9%, followed by common cold at 31.4%, and typhoid at 22.6%.

3.3 Proposed Methodology

The process of using machine learning techniques to forecast Dengue Fever involves a methodical approach aimed at extracting significant insights from the data. This involves selecting data from Internet Source carefully—a platform known for having a wide variety of datasets. The next steps are filling in the missing values, classifying the variables, and using EDA, or exploratory data analysis, to find patterns in the dataset. The process then moves on to model training, assessment, and testing in order to see how well machine learning algorithms predict the occurrence of dengue fever. In order to create strong models of prediction for proactive healthcare interventions, each step is important. Here is a general summary in the below flowchart in Figure 3.3:

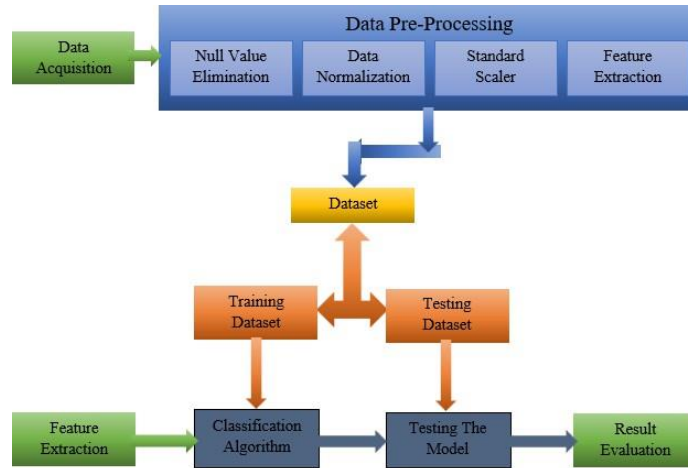


Figure 3.2: Methodology Flowchart

Data Collection: Obtain a dataset that includes pertinent features (symptoms, environmental conditions, etc.) and their relationships to dengue fever from an established source, such as Kaggle. Verify the correctness, quality, and usefulness of the dataset to the problem.

Missing Value Removal: Look for any missing feature values in the dataset. Strategies such as imputation (replacing missing values with a computed approximation) or elimination of instances or characteristics with excessive missing values may be used, depending on the quantity and pattern of missing data.

Encoding:

Provide numerical representations of categorical variables that are suitable for machine learning methods. For that, techniques such as label encoding or one-hot encoding can be used.

EDA (Exploratory Data Analysis):

For understanding the elements of the dataset, feature relationships, distributions, and probable Dengue Fever trends, perform exploratory analysis. For EDA, visualization tools like heatmaps, correlation matrices, box plots, and histograms might be useful.

Model Selection: An important part of Dengue predictions is choosing appropriate machine learning models for labeled data training. Potentially useful models for identifying trends and forecasting Dengue Fever include AdaBoostClassifier, BernoulliNB, GaussianNB, DecisionTreeClassifier, BaggingClassifier, and Voting Classifier. By selecting these models, the goal is to take advantage of their various advantages and skills in order to produce reliable results and precise forecasts for acknowledging bipolar disorder from the given labeled data.

AdaBoostClassifier

AdaBoost, is a group learning technique that builds a powerful classifier by combining several weak learners, typically decision trees. It concentrates on cases which were previously misclassified and instructs a series of weak learners in a sequential manner. Gradually, accuracy increases as each learner after them pays closer attention to the examples that were incorrectly recognized.

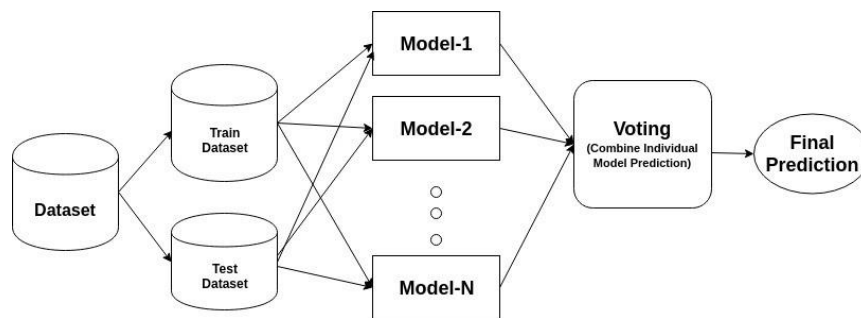


Figure 3.3: AdaBoostClassifier Architecture

BernoulliNB (Naive Bayes)

Bernoulli Uninformed A probabilistic classifier with an assumption of feature independence, Bayes is based on the Bayes theorem. It is a popular method for categorical data and is appropriate for text classification problems when word absence or presence is significant since it assumes binary-valued features (Bernoulli distribution).

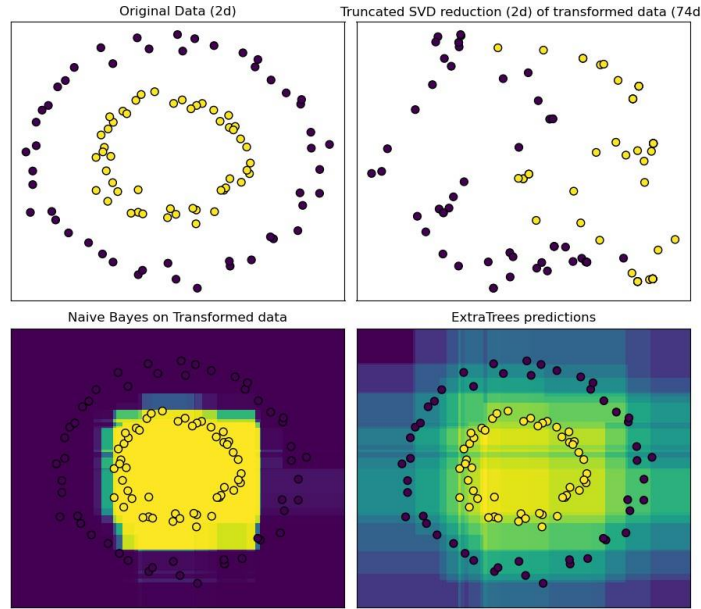


Figure 3.4: BernoulliNB (Naive Bayes)

GaussianNB (Naive Bayes)

The Naive Bayes classifier also goes by the name Gaussian Naive Bayes, which makes the assumption that continuous features have Gaussian (normal) distributions. It performs well in practice, particularly when the normal distribution assumption has been satisfied, and is particularly useful for handling continuous numerical characteristics.

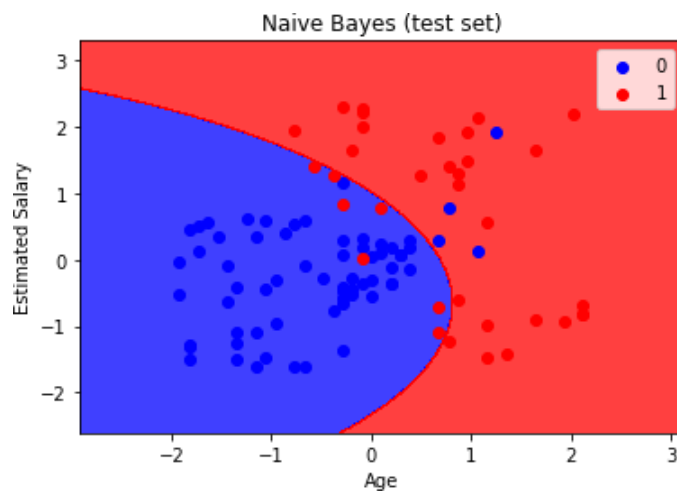


Figure 3.5: GaussianNB (Naive Bayes) Architecture

DecisionTreeClassifier

Developed for both regression and classification applications, decision trees include hierarchical structures resembling trees. They use features to divide the data into smaller subsets and build a decision tree. Every leaf denotes an outcome, each axis a decision rule is and every node a feature. Although decision trees are interpretable and capable of handling both numerical and categorical data, they are susceptible to overfitting.

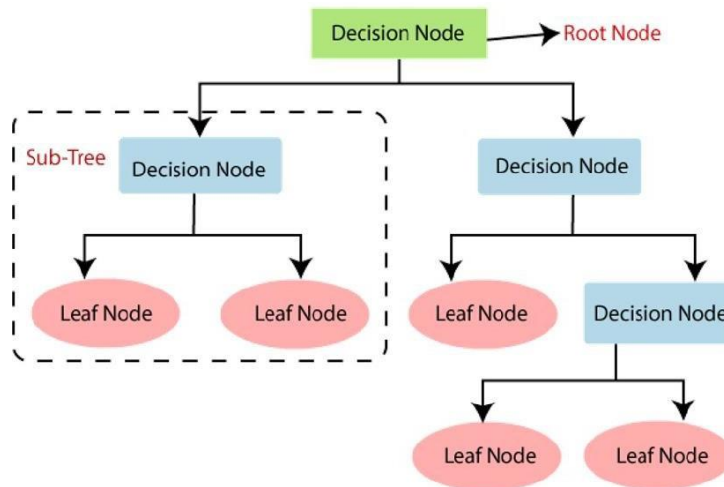


Figure 3.6: DecisionTreeClassifier Architecture

BaggingClassifier

Bootstrap Aggregating, or Bagging, is a team approach that uses bootstrapping to construct numerous subsets of the initial data set and trains an initial classifier (such as Decision Trees) for each subset. By averaging or voting the predictions from several classifiers, it combines the results, decreasing variance and frequently increasing accuracy over that of individual classifiers.

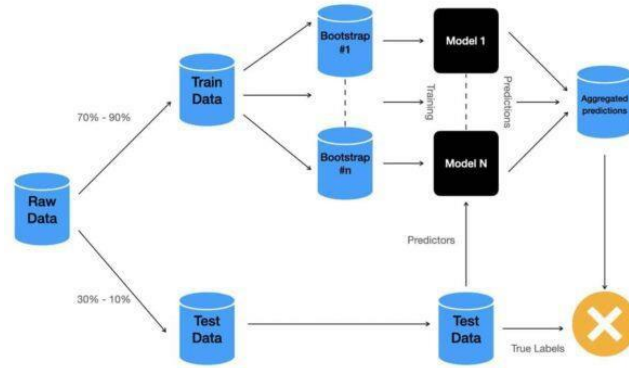


Figure 3.7: BaggingClassifier Architecture

Voting Classifier

The Voting Classifier aggregates the predictions of multiple machine learning models—which may be of different kinds—by using a variety of techniques, including majority voting (hard voting) and probability averaging (soft voting). It functions well when there is uncertainty about the optimal model selection or when several models capture different parts of the data.

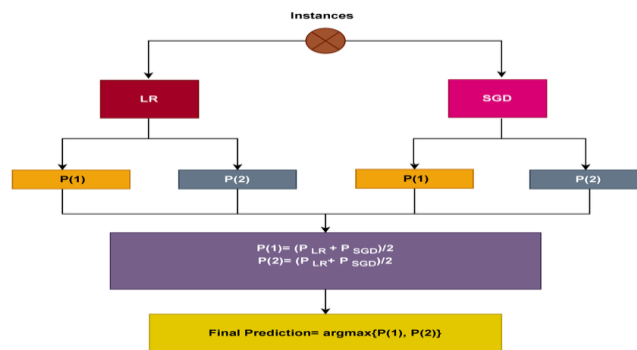


Figure 3.8: Voting Classifier Architecture

Model Training:

Data Splitting a. Create training and validation/test sets from the dataset (e.g., 70/30 or 80/20 split).

Feature Engineering/Selection: Determine pertinent features and, if needed, scale or modify them.

Model Training: Utilizing appropriate hyperparameters, train a chosen set of models on the training dataset.

Model Evaluation:

Validation a. Use metrics such as accuracy, precision, recall, F1-score, and ROC-AUC to assess how well the trained models perform on the validation set.

Hyperparameter Tuning: For improved performance, optimize the model's hyperparameters using approaches like random or grid search.

Cross-Validation: Use cross-validation to evaluate the stability and performance extension of the model.

Test Model: Final Evaluation evaluates the most effective model's real-world performance using a separate test dataset that wasn't utilized for training or validation. Make predictions using data that wasn't observed using the model that was trained, and then compare the outcomes to the real numbers of dengue-related cases. This technique provides an orderly manner to create and evaluate machine learning algorithms for the dengue fever diagnosis, ensuring data quality, accuracy of models, and meaningful applicability in real-world scenarios.

3.4 Implementation Requirements

We followed a few phases in the implementation process:

- Data Scraping.
- Data Labeling
- Data Cleaning
- Preprocessing
- Model training in Google Colab.

CHAPTER 4

EXPERIMENTAL RESULTS AND DISCUSSION

4.1 Experimental Setup

The practical use of machine learning techniques to Dengue Fever prediction needs particular software applications and a strong computing environment. First and foremost, having access to computational resources—such as CPUs or GPUs with enough processing power and memory—that can handle data-intensive activities is crucial. Additionally, the use of modules such scikit-learn, pandas, and empty in conjunction with programming languages like Python is essential for preprocessing data, creating models, and analyzing them. Access to statistical analysis software is also helpful for exploring data analysis and result interpretation. For model training, validation, and testing, trusted and varied datasets with comprehensive data on symptoms, environmental factors, and past cases of Dengue Fever are essential. These datasets can be obtained from websites such as Kaggle. Finally, during the implementation stage, having access to cloud-based services could help with scalability and efficiency when processing large data sets and computing demands.

4.2 Experimental Results & Analysis

The research and experimental findings are centered on analyzing how well the created machine learning models predict dengue fever. In this stage, the models' ROC-AUC, F1-score, accuracy, precision, recall, and F1-score metrics are evaluated using test and validation datasets. The outcomes show how well the models can predict Dengue outbreaks based on symptoms and environmental factors. In addition, the analysis explores the advantages and disadvantages of each model, emphasizing how well each one performs in identifying trends and predicting cases of dengue fever. It clarifies the elements that affect the effectiveness of the model, possibly highlighting certain symptoms, conditions in the environment, or model architectures that have a major impact on precise predictions. In addition, the study could include visualizations that help understand the behavior of the models and identify important predictors, including confusion matrices, ROC curves, or

feature importance plots. The analysis's conclusions are critical for understanding the models' reliability, generalizability, and usefulness in predicting Dengue Fever. They also serve as a roadmap for next updates and model-refinement techniques.

Accuracy: Accuracy measures the overall correctness of the model's predictions by comparing the number of correctly classified samples to the total number of samples. When classes are unbalanced, it gives a broad indication of the model's effectiveness but might not give a whole picture.

$$Accuracy = \frac{TruePositive + TrueNegative}{TruePositive + FalsePositive + TrueNegative + FalseNegative}$$

Precision: Out of all positive predictions generated by the model, precision focuses on the percentage of true positive forecasts.

$$Precision = \frac{TruePositive}{TruePositive + FalsePositive}$$

Recall: Also known as sensitivity or true positive rate, recall is the percentage of true positive predictions made out of all truly positive samples.

$$Recall = \frac{TruePositive}{TruePositive + FalseNegative}$$

F1 score: The F1 score is the harmonic mean of recall and precision. It provides a reasonable evaluation metric that considers recall and precision. The F1 score is useful when classes are uneven since it accounts for both false positives and false negatives. A high F1 score denotes a well-balanced precision to recall ratio.

$$F - 1 \text{ Score} = 2 * \frac{Recall * Precision}{Recall + Precision}$$

The result of deep learning model is compared on the basis of Accuracy, Precision, Recall, F1 Score in below table of 4.1:

TABLE 4.1. PERFORMANCE EVALUATION

Model Name	Accuracy	Precision	Recall	F1-Score
BaggingClassifier	95.87%	96%	95%	95%
BernoulliNB	95.32%	95%	95%	95%
DecisionTree	95.05%	96%	95%	96%
Voting Classifier	94.50%	93%	94%	93%
AdaBoostClassifier	89.00%	90%	90%	90%
GaussianNB	77.47%	82%	76%	71%

4.3. Accuracy

The research assessed the accuracy of testing and training a range of algorithms, encompassing both popular machine learning methods and models. The goal was to determine which algorithm was best in forecasting cases of dengue fever. Remarkably, with an accuracy rate of 95.87%, the BaggingClassifier algorithm was the best-performing one. Figure 4.1 compares the accuracy of the various models:

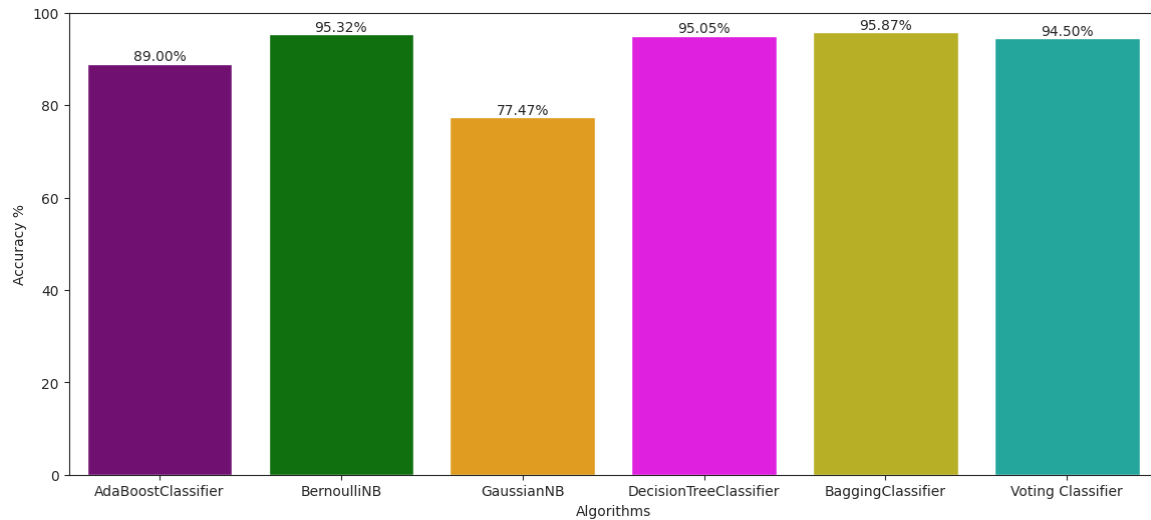


Figure 4.1: Accuracy Comparison of Machine Learning Models

A comparison of several algorithms' performance in predicting dengue fever is shown in Figure 4.1. Remarkably, the accuracy of the BaggingClassifier was remarkable, as it achieved a remarkable 95.87%. The BernoulliNB model came in second with an accuracy of 95.32%, and the DecisionTreeClassifier third with a rate of 95.05%. The Voting Classifier performed admirably, with an accuracy of 94.50%; AdaBoostClassifier performed even better, with 89.00%; and GaussianNB performed even worse, with 77.47%. The outcomes highlight BaggingClassifier's efficiency in surpassing alternative algorithms in terms of accuracy measures.

Performance Analysis

BaggingClassifier:

Achieve the Accuracy of 95.87%, Precision score of 96%, Recall score of 95% and F1-score of 96%. Below at table 4.2 we have performance evaluation of BaggingClassifier:

Table 4.2. Performance Evaluation(BaggingClassifier)

	Precision	Recall	F1-Score	Support
0	97%	98%	97%	169
1	93%	95%	94%	111
2	97%	93%	95%	84
Accuracy			96%	364
Macro Avg	96%	95%	96%	364
Weighted Avg	96%	96%	96%	364

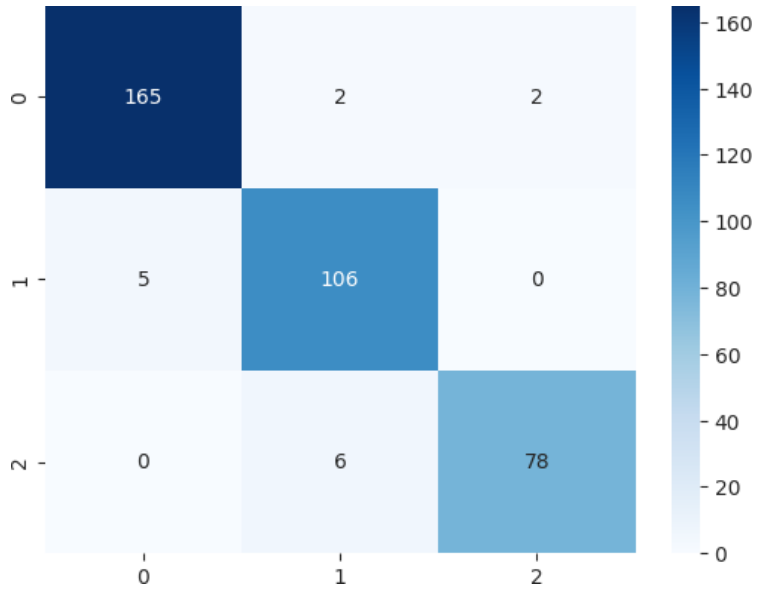


Figure 4.2 : Confusion Matrix BaggingClassifier

BernoulliNB:

Achieve the Accuracy of 95.32%, Precision score of 95%, Recall score of 95% and F1-score of 95%. Below at table 4.3 we have performance evaluation of BernoulliNB:

Table 4.3. Performance Evaluation(BernoulliNB)

	Precision	Recall	F1-Score	Support
0	97%	98%	97%	169
1	93%	94%	93%	111
2	95%	93%	94%	84
Accuracy			95%	364
Macro Avg	95%	95%	95%	364
Weighted Avg	95%	95%	95%	364

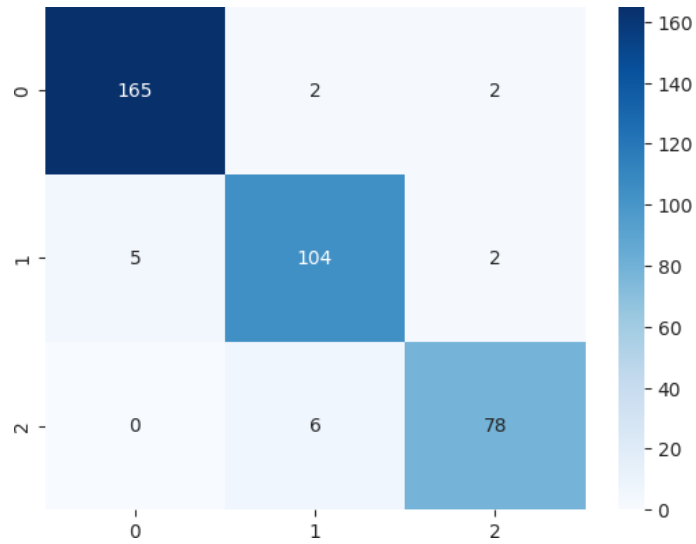


Figure 4.3: Confusion Matrix BernoulliNB

Decision Tree:

Achieve the Accuracy of 95.05%, Precision score of 96%, Recall score of 95% and F1-score of 96%. Below at table 4.4 we have performance evaluation of Decision Tree:

Table 4.4. Performance Evaluation (Decision Tree)

	Precision	Recall	F1-Score	Support
0	97%	98%	97%	169
1	93%	95%	94%	111
2	97%	93%	95%	84
Accuracy			96%	364
Macro Avg	96%	95%	96%	364
Weighted Avg	96%	96%	96%	364

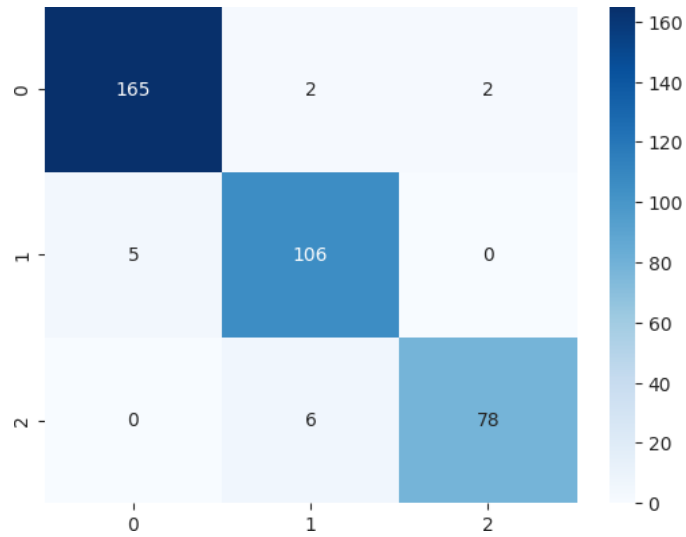


Figure 4.4: Confusion Matrix Decision Tree

Voting Classifier

Achieved the accuracy of 94.50%, Precision score of 93%, Recall score of 94% and F1-score of 93%. Below at table 4.5 we have performance evaluation of Voting Classifier:

Table 4.5. PERFORMANCE EVALUATION (VOTING CLASSIFIER)

	Precision	Recall	F1-Score	Support
0	97%	98%	97%	169
1	93%	90%	91%	111
2	91%	93%	92%	84
Accuracy			94%	364
Macro Avg	93%	94%	93%	364
Weighted Avg	94%	94%	94%	364

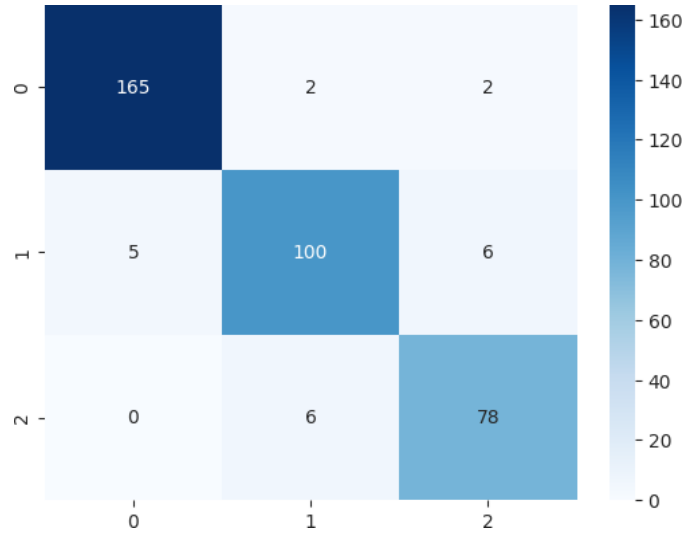


Figure 4.5: Confusion Matrix Voting Classifier

AdaBoostClassifier

Achieve the highest accuracy of 89.00%, Precision score of 90%, Recall score of 90% and F1-score of 90%. Below at table 4.6 we have performance evaluation of AdaBoostClassifier:

TABLE 4.6. PERFORMANCE EVALUATION(ADABOOSTCLASSIFIER)

	Precision	Recall	F1-Score	Support
0	94%	86%	90%	169
1	79%	91%	85%	111
2	96%	93%	95%	84
Accuracy			89%	364
Macro avg	90%	90%	90%	364
Weighted avg	90%	89%	89%	364

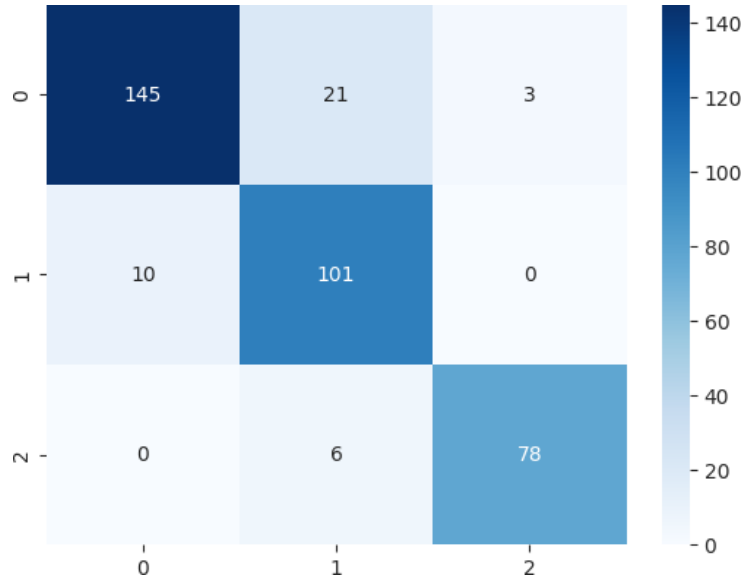


Figure 4.6: Confusion Matrix AdaBoostClassifier

GaussianNB:

Achieve the highest accuracy of 77.47%, Precision score of 82%, Recall score of 76% and F1-score of 71%. . Below at table 4.6 we have performance evaluation of GaussianNB:

TABLE 4.7. PERFORMANCE EVALUATION(GAUSSIANNB)

	Precision	Recall	F1-Score	Support
0	96%	97%	96%	169
1	97%	31%	47%	111
2	53%	100%	69%	84
Accuracy			77%	364
Macro Avg	82%	76%	71%	364
Weighted Avg	86%	77%	75%	364

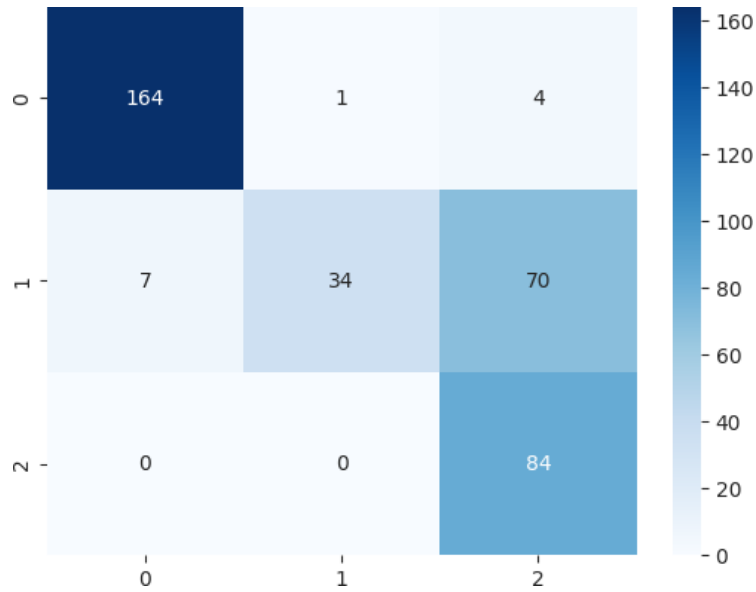


Figure 4.7: Confusion Matrix GaussianNB

Discussion

The main results, impact, limitation, and possible future paths from the machine learning-based Dengue Fever prediction study are summarized in the discussion section. First of all, it highlights the models' performance, emphasizing which algorithms are best at correctly forecasting Dengue outbreaks based on symptoms and environmental variables. The benefits of the models are discussed, along with how well they can capture intricate patterns and contributing variables in cases of dengue fever. The conversation also outlines the useful applications of these prediction models in the public health and healthcare domains. In order to lessen the impact of dengue fever on groups that are vulnerable, it highlights the importance of early outbreak detection, focused interventions, and distribution of resources measures. In addition, the study's constraints and difficulties—such as problems with data quality, biases in the models, or a shortage of access to datasets—are discussed. These limitations highlight areas for development or possible future research fields and offer context for understanding the model results. Finally, the talk looks at possible directions for improving or improving models for prediction. It might suggest methods to increase the predicted accuracy and robustness of Dengue Fever prediction, such as combining more data sources, improving feature engineering approaches, or exploring the

latest machine learning algorithms. Overall, the discussion section summarizes the study's findings and offers insights into the models' performance, public health implications, limitations, and opportunities for improving predictive capacities in controlling Dengue Fever epidemics.

CHAPTER 5

Impact On Society, Environment and Sustainability

5.1 Impact on Society

There is a great deal of potential for the dengue fever prediction models to benefit society. At first, these models assist the prompt identification and modeling of Dengue outbreaks, endowing public health authorities with the ability to execute prompt actions and distribute resources effectively, resulting in potentially limiting the disease's transmission and severity. Second, by being proactive, these models help to improve public health preparation, promote a stronger reaction to possible outbreaks, and lessen the burden on healthcare systems. Thirdly, by protecting vulnerable groups, especially in areas where Dengue outbreaks are likely to occur, targeted actions based on predictive insights may lessen the socioeconomic burden on impacted communities. Fourthly, these models allow people to take preventive action by providing accurate and proactive approaches that promote community awareness and education about Dengue Fever. Last but not least, the incorporation of predictive technology into disease surveillance underscores the revolutionary possibilities of machine learning in the healthcare industry, establishing standards for approaching advances in the control of infectious diseases and proactive public health activities.

5.2 Impact on Environment

The harmful effect of dengue fever is further increased by the creation of prediction models for the infection. First off, these models help with targeted mosquito control strategies by enabling early identification and forecast of Dengue outbreaks, which may lessen the need for extensive pesticide use and associated environmental impact. Second, because these models are proactive, improved management of environmental resources can be supported. By focusing interventions on specific regions at risk, water and energy resources can be optimally employed to control mosquito breeding sites. Thirdly, these models indirectly support biodiversity and ecosystem protection by reducing the spread of dengue fever. As

significant outbreaks have an opportunity to upset ecosystems and species, upsetting the balance of nature. Fourth, the application of preventive actions based on predictive insights may result in the creation and acceptance of sustainable and environmentally friendly vector control techniques, which would be in line with practices that value the environment. Last but not least, the addition of technology-driven methods into disease surveillance fosters a framework for future studies and regulations targeted at protecting environmental health and combating infectious diseases by promoting a greater knowledge of the external factors influencing disease spread.

5.3 Ethical Aspects

There are different moral issues in the creation and application of dengue fever prediction models. First and foremost, it is crucial to protect the privacy and confidentiality of personal health data utilized in model building. To this end, strong data anonymization and adherence to moral data management requirements are needed. Second, in order to ensure responsibility for the deployment of interventions and to enable stakeholders to understand how predictions are produced, transparency in the model creation and decision-making processes is important. Thirdly, it is critical to reduce model biases in order to avoid identifying against specific communities or populations when distributing resources and developing intervention measures. Fourthly, ethical standards are maintained by securing informed consent and providing voluntary involvement in collecting data techniques, respecting the autonomy and rights of individuals. Additionally, continuous ethical assessment and mitigation techniques are needed to handle the possible unintended implications of model predictions, such as stigmatization or a heavy reliance on technology. Last but not least, maintaining ethical standards of justice and fairness in public health interventions requires promoting inclusion and equal access to treatment and health services based on predictive models.

5.4 Sustainability Plan

A number of significant parts make up a long-term plan for dengue fever prediction models. Firstly, an ongoing assessment and monitoring of the models' performance ensures their

success and applicability over time, permitting the required modifications and improvements. Second, developing collaborations with researchers, local governments, and healthcare professionals facilitates the exchange of knowledge and ensures ongoing support for the application and enhancement of the model. Thirdly, maintaining continuing adherence to privacy standards and ethical considerations is ensured by putting in place a strong data governance system that includes ethical rules and data protections. Fourth, by building in scalability, the models' architecture ensures their longevity by enabling smooth flexibility to shifting data volumes or technology breakthroughs. Moreover, obtaining resources and long-term support for model maintenance, for activities to continue, infrastructure maintenance and data collection are essential. Finally, including stakeholders and communities in the continuous application of predictive models promotes trust and ownership, which in turn promotes continued adoption and support of Dengue Fever control programs.

CHAPTER 6

Summary, Conclusion, Recommendation and Implication for Future Research

6.1 Summary of the Study

In order to predict Dengue Fever, this study used machine learning algorithms and a variety of datasets including history records, environmental variables, and symptoms. A variety of machine learning methods were assessed, showing how well they predicted Dengue outbreaks. On validation and testing, the predictive models showed encouraging accuracy, precision, and recall metrics. datasets, showcasing their potential for early Recognition of an outbreak. The models' insights highlighted particular symptoms and environmental factors that have a major impact on the incidence of dengue fever. The study also highlighted the usefulness of these prediction models in supporting targeted actions, improving public health preparation, and limiting the impact of dengue fever on sensitive populations. There were acknowledged limitations, such as problems with data quality and ethical issues, which opened up new directions for investigation and model improvement. All things considered, this study shows how machine learning can revolutionize the treatment of infectious diseases and open the door to safeguards against Dengue Fever outbreaks.

6.2 Conclusions

In conclusion, the creation and assessment of machine learning models for predicting Dengue Fever suggest significant customers for early illness management. The study showed that several algorithms can effectively predict Dengue outbreaks based on environmental factors and symptoms. These models offer a helpful tool for diagnosing problems early and supporting prompt responses and plans for allocating resources. The results highlighted particular signs and environmental factors that are important in forecasting Dengue cases, providing important information for focused public health campaigns. However, restrictions on the quality of the data and ethical issues demand

continued examination and improvement in future research projects. Successful execution of these models depends on regular evaluation, modification, and interaction with medical specialists and local government authorities. In addition, keeping ethical data handling procedures, strong data governance, and community involvement are essential for the continued and appropriate application of predictive technology in public health. In the end, this work highlights the revolutionary potential of machine learning in the surveillance of infectious diseases, emphasizing the vital role these prediction models play in supporting proactive disease management strategies and public health readiness for Dengue Fever.

6.3 Implication for Further Study

The study's findings have a number of significances for future investigations into machine learning-based dengue fever prediction. First, investigating ensemble techniques that combine several algorithms or adding deep learning architectures may improve the robustness and accuracy of predictions. Second, more research into the effects of particular locations or environmental conditions on Dengue outbreaks could provide detailed information for better predictions. Also looking into the integration of real-time data streams, such social media data or weather predictions, may improve the models' ability to predict and capture dynamic patterns of breakouts. Further research is also necessary to fully examine the ethical issues and create processes for reducing biases and providing fair model deployment. Last but not least, longitudinal research focused on the long-term viability and efficacy of predictive models in actual environments can offer invaluable data into their usefulness and influence on public health programs.

APPENDIX

Research Reflections:

I found it challenging to identify issues and circumstances while working on this research. To ensure that they would function as best as possible, I began by selecting the finest applications available. Everyone also needed to gain a thorough understanding of that using Python and machine learning. I was surprised by how difficult it was to gather and arrange such a large amount of data. I took a while to attain my objective, but I did.

This project must be finished in order for students to pass the CSE-499 Project/Internship Capstone course.

REFERENCE

- [1] Sarma, Dhiman, et al. "Dengue prediction using machine learning algorithms." 2020 IEEE 8th R10 humanitarian technology conference (R10-HTC). IEEE, 2020.
- [2] Salim, Nurul Azam Mohd, et al. "Prediction of dengue outbreak in Selangor Malaysia using machine learning techniques." Scientific reports 11.1 (2021): 939.
- [3] Harumy, et al. "Prediction for dengue fever in Indonesia using neural network and regression method." Journal of Physics: Conference Series. Vol. 1566. No. 1. IOP Publishing, 2020.
- [4] Xu, Jiucheng, et al. "Forecast of dengue cases in 20 Chinese cities based on the deep learning method." International journal of environmental research and public health 17.2 (2020): 453.
- [5] Kannimuthu, S., et al. "Performance evaluation of machine learning algorithms for dengue disease prediction." Journal of Computational and Theoretical Nanoscience 16.12 (2019): 5105-5110.
- [6] Sanjudevi, et al. "Dengue fever prediction using classification techniques." Int Res J Eng Technol (IRJET) 6.02 (2019): 558-563.
- [7] Iqbal, et al "Machine learning for dengue outbreak prediction: A performance evaluation of different prominent classifiers." Informatica 43.3 (2019).
- [8] Satri, et al "Development of prediction and forecasting model for Dengue disease using machine learning algorithms." 2020 IEEE International Conference on Distributed Computing, VLSI, Electrical Circuits and Robotics (DISCOVER). IEEE, 2020.
- [9] Lathesparan, et al "Finding the Best Feature Selection Method for Dengue Diagnosis Predictions." (2021).
- [10] Nishanthi, et al. "Prediction of dengue outbreaks in Sri Lanka using artificial neural networks." International Journal of Computer Applications 101.15 (2014).
- [11] Mello-Román, Jorge D., et al. "Predictive models for the medical diagnosis of dengue: a case study in Paraguay." Computational and mathematical methods in medicine 2019 (2019).
- [12] Silitonga, et al. Comparison of Dengue Predictive Models Developed Using Artificial Neural Network and Discriminant Analysis with Small Dataset.
- [13] Sutrisna, Bambang, et al. "Scoring Model to Predict Dengue Infection in the Early Phase of Illness in Primary Health Care Centre."

- [14] Sippy, Rachel, et al. "Seasonal patterns of dengue fever in rural Ecuador: 2009–2016 Seasonality of dengue fever in Ecuador." *bioRxiv* (2018): 452318.
- [15] Mariki, et al "Feature Selection Approach to Improve Malaria Diagnosis Model for High and Low Endemic Areas of Tanzania." (2021).
- [16] Choubey, et al "Enhancing the prediction efficiency of virus borne diseases using enhanced backpropagation with an artificial neural network." *Measurement: Sensors* 24 (2022): 100505.
- [17] Li, Fudong, et al. "A Bayesian classification model for discriminating common infectious diseases in Zhejiang province, China." *Medicine* 99.8 (2020).
- [18] Baker, et al. "Forecasting dengue fever using machine learning regression techniques." 2021 12th International Conference on Information and Communication Systems (ICICS). IEEE, 2021.
- [19] Shakil, et al. "Dengue disease prediction using weka data mining tool." *arXiv preprint arXiv:1502.05167* (2015).
- [20] Mariki, et al "Combining clinical symptoms and patient features for malaria diagnosis: machine learning approach." *Applied Artificial Intelligence* 36.1 (2022): 2031826.
- [21] Yavari Nejad, Felestin, and Kasturi Dewi Varathan. "Identification of significant climatic risk factors and machine learning models in dengue outbreak prediction." *BMC medical informatics and decision making* 21.1 (2021): 1-12.
- [22] Shireesha, O., et al. "Dengue Prediction Using Machine Learning Techniques." *Advances in Smart Grid and Renewable Energy: Select Proceedings of ETAEERE 2020*. Springer Singapore, 2021.
- [23] Saturi, Swapna. "Development of prediction and forecasting model for Dengue disease using machine learning algorithms." 2020 IEEE International Conference on Distributed Computing, VLSI, Electrical Circuits and Robotics (DISCOVER). IEEE, 2020.
- [24] Kannimuthu, S., et al. "Performance evaluation of machine learning algorithms for dengue disease prediction." *Journal of Computational and Theoretical Nanoscience* 16.12 (2019): 5105-5110.
- [25] Harumy, T. H. F., H. Y. Chan, and G. C. Sodhy. "Prediction for dengue fever in Indonesia using neural network and regression method." *Journal of Physics: Conference Series*. Vol. 1566. No. 1. IOP Publishing, 2020.










PLAGIARISM REPORT

Dengue Fever Prediction Using Machine Learning Techniques

ORIGINALITY REPORT

20% SIMILARITY INDEX	17% INTERNET SOURCES	10% PUBLICATIONS	10% STUDENT PAPERS
--------------------------------	--------------------------------	----------------------------	------------------------------

PRIMARY SOURCES

 Submitted to Daffodil International University Student Paper	5%
 dspace.daffodilvarsity.edu.bd:8080 Internet Source	3%
 Submitted to Loomis-Chaffee High School Student Paper	1%
 researchspace.ukzn.ac.za Internet Source	1%
 Submitted to Morgan Park High School Student Paper	1%
 Submitted to Columbia High School Student Paper	1%
 notablesdela.ciencia.conicet.gov.ar Internet Source	1%
 eprints.uad.ac.id Internet Source	1%
 www.mdpi.com Internet Source	<1%