iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 11 **Issue:** X **Month of publication:** October 2023

DOI: https://doi.org/10.22214/ijraset.2023.56378

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue X Oct 2023- Available at www.ijraset.com

Malaria Disease Prediction and Grading System: A Performance Model of Multinomial Naïve Bayes (MNB) Machine Learning in Nigerian Hospitals

Temitope Olufunmi Atoyebi[1], Rashidah Funke Olanrewaju[2], N.V. Blamah[3], Morufu Olalere[4]
Department of Computer Science Nasarawa State University Keffi Nasarawa, Nigeria

Abstract: ***Malaria disease is the number one cause of death all over the Sub-Sahara world. Data mining can help extract valuable knowledge from available data in the healthcare sector. This allows training a patient health prediction model faster than in a clinical trial. Various implementation of machine learning algorithms such as Bayesian Theorem, Logistic Regression, K-Nearest Neighbor, Support Vector Machine and Multinomial Naïve Bayes (MNB), etc. have been applied on Public Hospital Malaria Disease datasets but there has been a limit to modeling using Multinomial Naïve Bayes Algorithm. This research applied MNB modeling to discover the relationship between 15 relevant attributes of the Public Hospitals data collected from Bwari General Hospital in Bwari Area Council and Maitama Hospital in Abuja Municipal Area Council, Abuja, FCT, and Nigeria. The goal is to examine how dependencies between attributes affect the performance of the classifier. The MNB produces a reliable and transparent graphical representation between the attributes with the ability to predict new scenarios. The model has an accuracy of 97%. It was concluded that the model outperformed the GNB classifier which has an accuracy of 100% and RF which also has an accuracy of 100%.***

Keywords: ***Multinomial Naïve Bayes (MNB) Classifier, Malaria Disease, Machine learning models, Artificial intelligence, Hospitals***

I. INTRODUCTION

Malaria is a major health problem in Nigeria, causing a significant number of deaths and illnesses. Statistics show that mortality and morbidity rates are at alarming levels due to malaria in Nigeria. Shockingly, five out of every 10 people with chronic malaria die from the disease (World Health Organization [19], underscoring the urgency of addressing this problem.

Additionally, the rapid incidence of malaria and verified records show that most households in Nigeria are at risk of contracting the disease [24].

In terms of inequality, urban areas are relatively equitable due to their positive attitude towards preventing transmission through mosquito bites, unlike rural areas which have rapid infection rates [21]. The predominant cases in the majority of public hospitals are mainly malaria treatment and deaths (National Center for Disease Control [23].

Malaria is a disease spread through the bite of mosquitoes carrying the plasmodium parasite. This is a global pandemic affecting more than three-quarters of the world's population. According to a 2018 report by the World Health Organization [25] sub-Saharan Africa has the highest number of malaria cases, with more than half of all deaths in the region attributed to the disease. The report also highlights that children under 5 years old and the elderly are at higher risk of dying from malaria. According to a report by [23] rural people are at high risk of malaria due to environmental factors, poor health behaviors, inadequate public health centers, and limited accessibility, with high health care costs and cultural beliefs.

Malaria detection using machine learning has made significant progress in recent years and has become a powerful tool in a number of medical applications, such as disease identification and diagnosis. It is because malaria is such a common and potentially fatal disease, scholars have been interested in finding ways to use machine learning algorithms to increase the efficiency and accuracy of diagnosis. One of the key areas where machine learning has a significant impact is the use of image analysis to detect malaria. Historically, the gold standard for diagnosing malaria was microscopic analysis of blood stains. However, this procedure is time-consuming and requires competent specialists. Researchers used machine learning methods for automated image analysis to overcome these difficulties. They were able to create models that could accurately detect and classify malaria parasites in blood samples using algorithms trained on large blood smear datasets or simply by analyzing risk factors related [14].

©IJRASET: All Rights are Reserved

It is a fact that most malaria cases and deaths occur in sub-Saharan Africa. This region has some of the poorest countries in the world, accounting for 90% of deaths (about 3,000 deaths per day) [19]. Malaria remains one of the leading causes of illness and death in the tropics. This is the most common deadly tropical disease. It causes large amounts of illness and death in children and pregnant women [23].

In 2018, there were 247 million cases of malaria and nearly one million deaths – mainly in children living in sub-Saharan Africa [20, 21]. Every 45 seconds, a child dies from malaria, which causes 20% of all child deaths [19, 20].Malaria kills 3,000 children every day in sub-Saharan Africa, or one million children every year. In sub-Saharan Africa, many households as well as children suffer from malaria, which often causes chills, tremors and fever.

Malaria infection during pregnancy is notable in sub-Saharan Africa, where mortality from virulent P.falciparum is a much greater problem than in most of the world. Anemia is another complication of malaria that can lead to death. This occurs when P.falciparum breaks down red blood cells and thereby reduces red blood cell production. Pathology associated with P.falciparium malaria is notably due to the adhesion of infected red blood cells in the brain causing metabolic disorders and organ dysfunction [3].

Malaria is the cause of 6.5% of abortions, 15% of premature births and 0.7% of intrauterine deaths (WHO, 2016).Malaria infection leads to increased morbidity, mortality, and premature birth and low birth weight due to intrauterine growth restriction (IUGR), may be the result of placental parasitism [19]. Malaria infection during pregnancy is notable in sub-Saharan Africa, where mortality from virulent P.falciparum is a much greater problem than in most of the world. Anemia is another complication of malaria that can lead to death. This occurs when P.falciparum breaks down red blood cells and thereby reduces red blood cell production. Pathology associated with P.falciparium malaria is notably due to the adhesion of infected red blood cells in the brain causing metabolic disorders and organ dysfunction [3].

The process of training and growth can apply to physical, emotional, or intellectual development. Malaria parasites may present symptomatically or asymptomatically. Malaria can have different levels of severity: mild, acute, and chronic. People with malaria may experience chills, severe headaches, seizures, joint pain, and high fever, loss of appetite, red eyes and difficulty breathing. However, carriers can also be asymptomatic, meaning they may not show any signs of the disease even when infected with the plasmodium parasite [19, 23].

The global malaria community and WHO aim to eliminate malaria by eliminating malaria in countries and implementing prevention measures. The pace of progress in malaria-endemic countries is influenced by a number of factors, including the strength of national health systems, the level of investment in malaria prevention, as well as other factors. Biological determinants, environmental factors, and social, demographic, political, and economic factors. This information was reported by WHO in 2023. Malaria can present as mild or acute. Mild malaria presents with mild symptoms such as headache and temporary loss of appetite and can be treated with self-medication and herbal medicine.

Severe symptoms of acute malaria require medical attention. If you experience acute symptoms of malaria, please see your doctor or nurse as soon as possible. Chronic malaria is a medical emergency that requires immediate attention from health care professionals [12]. The disease can be diagnosed through appropriate testing and an effective treatment plan should be put in place to manage it. Chronic malaria is a potentially fatal disease that affects many people around the world. Chronic malaria is the leading cause of death, especially in children under 5 years' old, pregnant women and the elderly. Studies have shown that more than half of malaria cases result in severe illness and can lead to death. It is essential to take necessary precautions and receive prompt medical attention if you experience symptoms. This evidence contradicts the health of the population as there is no technology to predict the occurrence of this phenomenon.

Malaria is a potentially fatal disease caused by a parasite that lives in the saliva of infected mosquitoes. Malaria is transmitted through mosquito bites and affects red blood cells and liver cells, changing the body's biochemistry. Sub-Saharan Africa is home to four common malaria parasite species: 4,444 P. falciparum causes the most severe cases of malaria in sub-Saharan Africa, followed by P. vivax and P.ovale. It is important to note that cerebral malaria is a serious complication of malaria that can lead to abnormal mental status and high mortality rates, ranging from 15% to 50% due to brain infection.

Malaria poses a major threat to nearly half of the world's population. In 2021, about 247 million people in 85 countries have the disease. Sadly, about 619,000 people also died from malaria that same year. Certain groups of people are at higher risk of severe malaria, such as infants, children under 5 years old, pregnant women, and people with HIV/AIDS. People not taking chemoprevention therapy, such as migrants, mobile populations and tourists, or people entering areas of intense malaria transmission without acquired immunity a part after prolonged exposure to this disease, is also vulnerable.

In areas where malaria is common, some people develop partial immunity, which reduces the risk of serious illness but does not provide complete protection.

In Africa, young children are at the highest risk of malaria-related death, while other age groups living in areas with low transmission rates are also vulnerable due to lack of immunity. Projections of malaria incidence highlight the need for sound prevention methods and life-saving interventions. Clinicians and their personal experiences can help predict malaria symptoms.

People who carry the malaria parasite face difficult decisions. There are two options for treating malaria. The first is to consult a clinician, who can diagnose the parasite and recommend treatment based on symptoms. The other option is to self-treat if malaria is suspected. Clinicians play an important role in the treatment of acute and chronic malaria by diagnosing patients, documenting them, admitting them to hospital if necessary, providing treatment, and monitoring their recovery, also assess the risk of death. This disappointment was due to suspected professional negligence and poor medical care [21, 23]. [8] used machine learning techniques to accurately predict malaria symptoms and treatment by rapidly analyzing data from non-governmental hospitals. However, it is important to note that the data used in this study came from public hospitals, which is slightly different from existing data. This is the weakness of this study and can be summarized as follows:

- Lack of structured patient information in public hospitals due to staff turnover.
- Health service shortages, professional negligence and poor health care are widespread, especially in rural areas.
- Socio-economic characteristics have not been taken into account in a comprehensive manner.
- Manually creating and maintaining records through traditional methods can make it difficult to collect data sets.

II. RELATED CONCEPTS

This section describes some basic concepts and terminologies such as machine learning, classification and Naïve Bayes (Multinomial).

A. Machine learning

The field of Machine Learning (ML) has been in existence since 1959. Arthur Samuel while working for IBM defined ML as a field of study that enables the computer to learn without being explicitly programmed. A formal definition of ML was proposed by Tom Mitchell using a well-posed learning problem, stating that A is said to learn from experience E with respect to some task T and some performance measure P, if its performance on T as measured by P, improves with experience E. To relate this definition to this study, we aimed to develop a malaria prediction system. The task T of this system is to predict the presence of Malaria disease. The performance measure P is the prediction accuracy of our model. The system will find out if we have additional clinical data about the patient's malaria status. Here the experience E refers to the set of already processed clinical datasets. Hence as more records of data are added to the system, we achieve a higher precision as regards its accuracy. ML is a recurring field of artificial intelligence (AI) with active research and applications over the past several decades. Artificial Intelligence (AI) has been identified as one of the major role players in this current industrial revolution and there has been a lot of evolution in different machine learning algorithms (Abraham, 2020). Self-driven cars, speech recognition, robotic controls, effective web search, face detection are only a few of the areas where ML is being used.

B. Random Forest

This random forest is one of the most popular and commonly used algorithms by data scientists.

Random forest (RF) is a supervised machine learning algorithm widely used in regression and classification challenges. It builds decision trees on different samples and takes majority votes for classification and averaging in case of regression. Random Forest is a general-purpose machine learning algorithm developed by [9]. It leverages a set of multiple decision trees to generate predictions or classifiers. By combining the output of these trees, the random forest algorithm yields more accurate and aggregated results. Its widespread popularity comes from its user-friendly nature and adaptability, which allows it to effectively solve classification and regression problems. The algorithm's strength lies in its ability to handle complex datasets and minimize over fitting, making it a valuable tool for various predictive machine learning tasks.

One of the most important features of the Random Forest algorithm is that it can handle data sets containing continuous variables, as in the case of regression, and categorical variables, as in the case of classification. It works best for classification and regression tasks. In this tutorial, we will understand how random forests work and implement random forests in a classification task.

C. Naïve Bayes

Naïve Bayes classifier provides an analytical engine that defines a set of model rules to classify data into different classes using a probabilistic approach [15].

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

Initially, it builds a model for each class attribute based on the other remaining attributes in the dataset. It correlates the class of each record using a previously designed model on new and even previously unseen data sets. This analysis helps to fully understand the data set and predict future trends [4].

This probabilistic classifier used to predict outcomes based on maximum likelihood. The model assumes that all variables in the data set used to predict the target value are independent. Classification models are based on the assumption that the value of an entity in a data set does not depend on the values of other entities in the data set. It focuses on the dependent variable and then thinks about the probability of the given value that the independent variables have, determining which value has the highest probability. The dependent variable will fall into this classification [7].

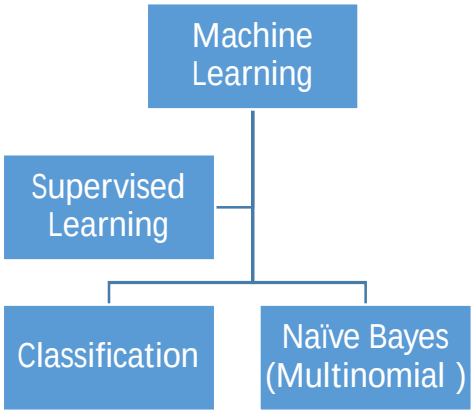


Figure 1.0 Naïve Baye Classification

1) Malaria Detection Systems Using Naïve Bayes Technique

Predictive models for data mining tasks include regression, classification, prediction, and time series analysis [16]. In this study, Naïve Bayes classifier is proposed to predict malaria subtypes and associated complications.

Typically, a prediction system is obtained by building a model by extracting rules and patterns in the training data set and using these extracted rules to predict whether the class of records has a class label or not. In the past, many researchers in the field of data mining have demonstrated how to provide a system to improve accuracy and precision by developing automated disease diagnosis and prediction systems using Use different classification techniques

The approach uses natural language processing and machine learning (ML) Naïve Bayes techniques and the diseases considered in the study include: Malaria, typhoid, dengue fever, tuberculosis and hepatitis B. However, this system focuses on Medline text as the target parameter and does not consider accuracy verification as a means to validate the model.

Similar research by [18] was also performed on a prediction system for liver diseases such as hepatitis, cirrhosis, bile duct and liver cancer using SVM and Naïve Bayes classification algorithm.

The classifier used a large dataset of liver function tests (LFTs) from the University Hospital database and compared their results based on classification, accuracy and real-time shows two classifiers.

[10] Conducted research on triage and focused on hospital management issues such as a model to predict admission and discharge of patients from the emergency department based on initial medical records using text mining. [5] Presented a technique for predicting the presence of Malaria disease using detailed medical records provided by the patient.

The techniques used are Laplace smoothing and Naïve Bayes classifiers, Random Forests and Support Vector Machines.

The system helps avoid redundant diagnostic tests performed on patients and some of the delays caused in initiating appropriate treatment by rapidly diagnosing Malaria disease in patients. The system provides alternative opinions regarding the patient's condition, just like an experienced doctor, as it predicts from a historical database containing a large number of Malaria patient records.

The system can provide quality services at an affordable cost and patients can also use it if they go for a health check-up.

2) Hybrid Model of Malaria Prediction System

Machine learning has made significant progress in recent years and has become a powerful tool in a number of medical applications, such as disease identification and diagnosis. It is because malaria is such a common and potentially fatal disease, scholars have been interested in finding ways to use machine learning algorithms to increase the efficiency and accuracy of diagnosis.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

One of the key areas where machine learning has a significant impact is the use of image analysis to detect malaria. The gold standard for diagnosing malaria has always been microscopic analysis of blood stains. However, this procedure is time-consuming and requires competent specialists.

Researchers used machine learning methods for automated image analysis to overcome these difficulties. They were able to create models that could accurately detect and classify malaria parasites in blood samples using algorithms trained on large blood smear datasets or simply by analyzing risk factors related [14].

These machine learning-based analysis algorithms have the potential to provide rapid and accurate malaria diagnosis, especially in areas with limited resources and limited access to trained experts.

Many of the prediction systems developed have used combined preprocessing, clustering, and classification techniques known as hybrid models to achieve the common goal of prediction.

Indeed, research has shown the weakness of a single model in providing effective prediction modeling [6]. Recently, [9] conducted an evaluation of the application of Naïve Bayes Network (NBN), Random Forest and Support Vector Machine to predict and improve disease diagnosis by doctors. The results show that NBN performs better than other classifiers in terms of accuracy, specificity, sensitivity, receiver operating characteristic (ROC), and area under the curve (AUC).

III. METHODOLOGY

This section discusses the various materials and methods in this study. The research design, Malaria disease dataset, data preprocessing and the tools used for the study are discussed in detail.

A. Network Design

This depicts the actual flow of the entire network building. The flow diagram presented in Figure 1.0 explains the sequence involved in network design

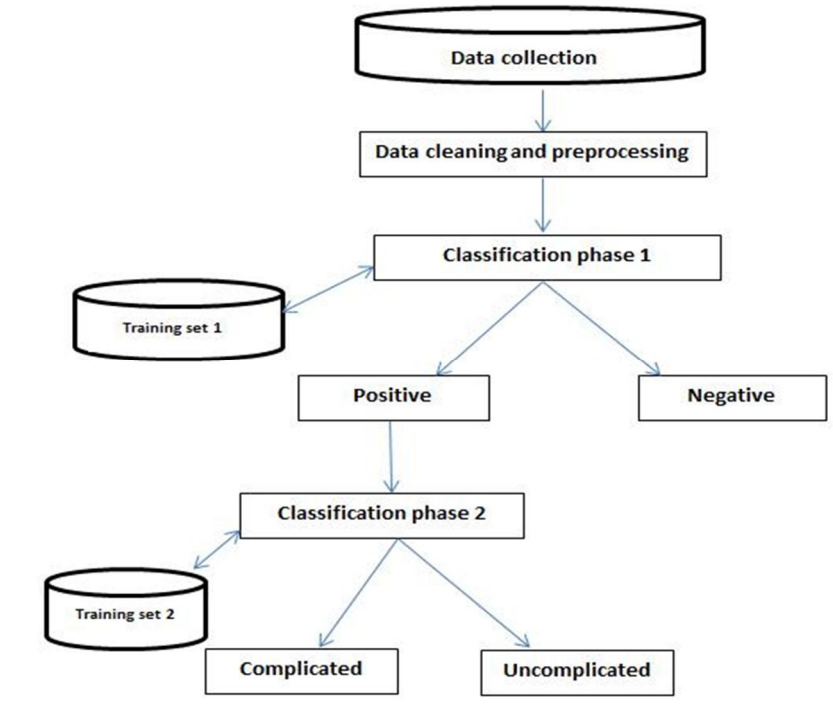


Figure 2.0: Sequence in Network Flow

B. Malaria Disease Data Collection

Malaria incidence data set is obtained from Public hospitals in Bwari Area Council and Maitama Hospital from Abuja Municipal Area Council in FCT, Abuja from 2017 to 2021. These are the data used for modeling and analysis. Also, putting in mind the geographical location and socio-economic factors inclusive which are available for patients inhabiting those areas. Naive Bayes (Multinomial) is the model used to analyze the collected data for malaria disease prediction and grading accordingly.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue X Oct 2023- Available at www.ijraset.com

C. Preprocessing Data

Data preprocessing is also known as cleaning data. It is one of the most important steps to achieve the best from the dataset. This is a process whereby data inconsistencies such as missing values, out of range values, unformatted data, and noise are removed from the data. The process is usually time-consuming because it involves a lot of experimentation trying out various data analysis tools. Our preprocessing involves data retrieval, handling missing values, target class transformation and data discretization.

1) Data Preparation

a) Data Preparation: This is a process of cleaning data, encoding data and having the data in a right format for either analysis or modelling.

b) Cleaning: The data was cleaned by removing duplicates and handling missing value, deleting the columns with missing values, replacing missing values with zeros, mean of the column and the most frequent values.

c) Data Encoding: This was achieved by changing Boolean, ordinal, categorical data to numerical data before feeding to a model, since models accept only numbers and not string.

2) Data Partitioning

The data which shall be collected will be divided into two portions; one portion of the data shall be extracted as a training set, while the other portion will be used for testing. The training portion shall be taken from a set stored in a database and will be called data which is *training set1*, while the training portion taking from another set store in a database is shall be called data which is *training set2*.

The dataset was split into two parts: a sample containing 70% of the training data and 30% for the purpose of this research. Then, using MNB classification algorithms implemented in Python, the models were trained on the training sample. On the 30% remaining data, the resulting models were tested, and the results were compared with the other Machine Learning models using the standard metrics.

3) Handling Missing Values

Missing data values is a common problem faced by analysts. This occurs due to different reasons such as incomplete extraction, corrupt data, failure to load the information, etc. This is a great challenge that must be fixed because good models are generated when you make the right decisions on how to fix it [11].

These are 5 ways of handling missing data: i. Deleting Rows ii. Replacing with mean/median/mode iii. Assigning a unique category iv. Predicting the missing values v. using algorithms which supports missing values. We adopted Deleting rows since we have few missing values.

4) Classification and Prediction

Base on the nature of variable in the dataset, this study will use Naïve Bayes (Multinomial) classification techniques; *Classification phase 1* and *Classification phase 2*. Working of the framework is illustrated as follows:

a) Data collection and preprocessing shall be done.

b) Preprocess data shall be stored in a training set 1 and training set 2. These datasets shall be used during classification.

c) Test data set is shall be stored in database test data set.

d) Part of test data set shall be compared for classification using classifier 1 and other part shall be classified using classifier 2 as follows:

- *Classifier stage 1:* It classify into positive or negative classes. If the patient is having malaria, then the patient is classified as positive (P), while a patient is classified as negative (N) if the patient does not have malaria.
- *Classifier stage 2:* It classify only data set that has been classified as positive by classifier 1, and then further classify them into complicated and uncomplicated class label.

The classifier will also capture data on environmental factors, genetics, gender and age, cultural and socio-economic variables. The system will be designed such that the core parameters as a determining factor should supply their value.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

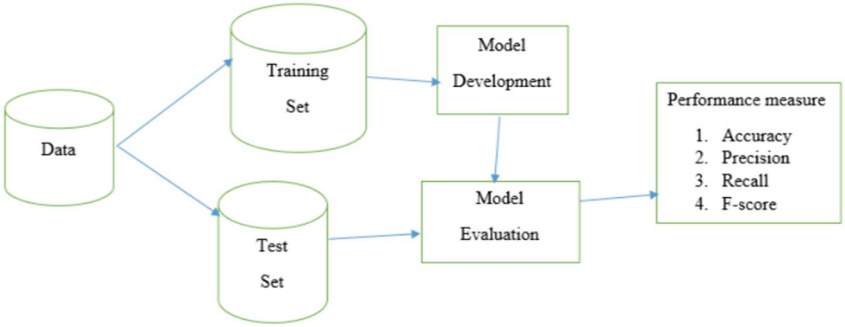


Figure 3. 0 Flow Diagram of Network Design

D. Performance Metrics

Performance metrics are used to evaluate how different algorithms perform based on various criteria such as accuracy, precision, recall etc. They are discussed below.

1) Confusion Matrix

The confusion matrix shows the performance of the algorithm in Table 1.0. It depicts how the classifier is confused while predicting. The rows indicate the actual instance of the class label while the columns indicate the predicted class instances. Table 2.0 below shows a confusion matrix for binary classification. True positive value signifies that the positive value is correctly predicted, false positive means the positive value is falsely classified, false negative means the negative value is falsely predicted while true negative value means the negative value is classified correctly. The Confusion matrix Figure 8.0 is used to calculate different performance metrics as discussed below.

2) Accuracy

Accuracy is the ratio between the number of correctly classified cases and all cases. It is the sum of TP and TN divided by the total number of instances.

3) Precision

Precision is the proportion of true positive instances that are classified as positive. It reflects the closeness of predicted values is to one another [20].

4) Recall

Recall is the proportion of positive instances that correctly classified as positive. Recall is known as sensitivity. F1 Score F1 score combines both precision and recall and finds a balance between both.

IV. RESULTS AND DISCUSSIONS

This section discusses the results of the research and compares the Multinomial Naïve Bayes model with the Gaussian Naive Bayes and Random Forest Classifiers.

A. Algorithmic Flow

1) Load the dataset
2) Train the model
3) Evaluate the model
4) Tune the model
5) Test (Make predictions) the model
6) Compare predictions

B. Algorithm Development

1) Step 1: Pandas was used to load data into data frame from an excel file (.xlsx)
2) Step 2: Scikit Learn test train split was used to split the data into training, validation and test set

3) Step 3: MNB were used to train, validate and test the model
4) Step 4: The following hyper parameter were fine-tuned; fitting model, outliers, over fitting
5) Step 5: The performance of the model was determined based on these metrics; which are accuracy score, confusion matrix, recall, precision, f-measure and error rate
6) Step 6: The trained model called predictive model is saved and can be used to make predictions

C. Performance Evaluation of the MNB Model

This is to show how well our model performed. The model was able to predict 51 out of the 60 test samples correctly thereby achieving an accuracy of 97%. The confusion matrix is as shown in Figure4.0 below. Performance metrics are also as shown in the Figure 5.0 below.

D. Performance Evaluation of the Gaussian Naïve Bayes (GNB) and Random Forest (RF) Algorithm

The performance evaluation of the Gaussian Naïve Bayes and Random Forest has an accuracy of 100% which is not too good for our accuracy due to the presence of noise and outliers. The confusion matrix is as shown in Figure 4.0 below. Performance metrics are also displayed as shown in Figure 5.0 below.

Table 1.0 Classification Report Table

SN	Model	Accuracy Score	Precision (Negative, Positive)	Recall (Negative, Positive)	F1-Score (Forged, Real)
1	MNB	97%	(98%)	(98%)	(97%)
2	GNB	100%	100%	100%	100%
3	RF	100%	100%	100%	100%

Table 2.0 Confusion Matrix Report

```
df.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2121 entries, 0 to 2120
Data columns (total 16 columns):
 #   Column                                                  Non-Null Count  Dtype
---  ------                                                  --------------  -----
 0   Patient Serial Number                                   2121 non-null   int64
 1   Pregnancies                                             2121 non-null   int64
 2   Availability of Treated Net                             2121 non-null   int64
 3   Season: Level of Rainfall-Stagnant water breeding       2121 non-null   int64
 4   Rate of Malaria                          Infection (Lab Diagnosis)  2121 non-null   float64
 5   Malaria Parasite Density Fever- Rapid Diagnostic Test(Strip)  2121 non-null   object
 6   Complaints/Symptoms                                     2109 non-null   object
 7   Age                                                     2121 non-null   int64
 8   Outcome                                                 2121 non-null   int64
 9   Electricity                                             2065 non-null   float64
 10  Laboratory Equipments                                   2065 non-null   object
 11  Doctor to Patient                                       2065 non-null   object
 12  Environment - Sanitised or not                          2065 non-null   float64
 13  Complicated/ Uncomplicated Malaria Diagnosis            2121 non-null   object
 14  Location (Urban/ Rural Area)                            2121 non-null   object
 15  Malaria Outcome Interpretation                          2121 non-null   object
dtypes: float64(3), int64(6), object(7)
memory usage: 265.2+ KB
```

Figure 8.0: Attributes of Malaria Disease dataset from Public Hospitals in Abuja, FCT

E. Discussion

It is obvious from the performance metrics of MNB, GNB and RF that the MNB outperformed the GNB and RF. The MNB model was able to classify 97% of the test dataset correctly compared to the 100% achieved by GNB and RF due to outliers and noise.
Interpreting the confusion matrix report, the model reported 43 false positives (FP) and 73 false negatives (FN) on the test set. This suffices to say the model correctly predicted 134 negative samples, while it wrongly predicted 43 negative samples as positive. The calculation below is used to determine the error rate which is calculated by taken the addition of the numbers not at the diagonal of the confusion matrix (misclassified samples) and dividing it by the total test data. Therefore, to calculate the total error generated by the model the calculation is done below:

		Actual Class	
		Negative	Positive
Predicted Class	Negative	*TN* 134	*FP* 43
	Positive	*FN* 73	*TP* 175

$$\text{Error Rate(Full Set)} = \frac{73+43}{425}$$

$$= 0.2729$$

F. Findings

The Figure 4.0 below and Table 1.0 above show the evaluation metrics for the multinomial Naïve Bayes model (MNB). The model architecture was trained and tested with the data collected. Based on the data samples, the model classifies the inputs into positive or negative. All the evaluation techniques used show optimized performance results according to Figure 5.0 below, the model (MNB) generated outputs with an error less than (0.3), and the model misclassified none having an accuracy score of 97%. In summary, the study performed a total of three (3) different evaluation metrics as stated and calculated above. This model is therefore valid.

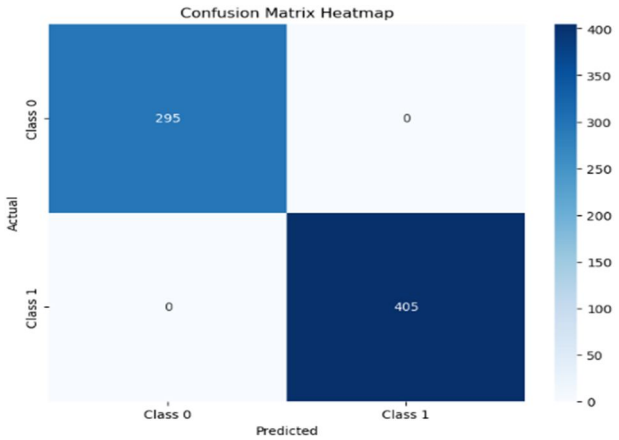


Figure 4.0 Confusion Matrix Heatup Map

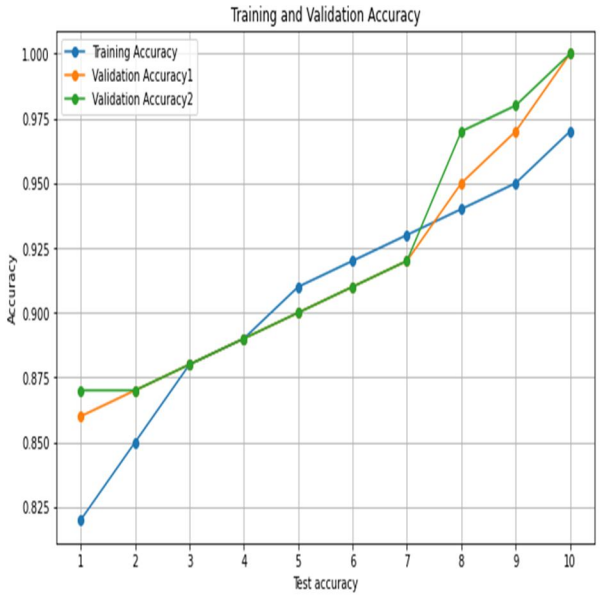


Figure 5.0 Training and Validity Check

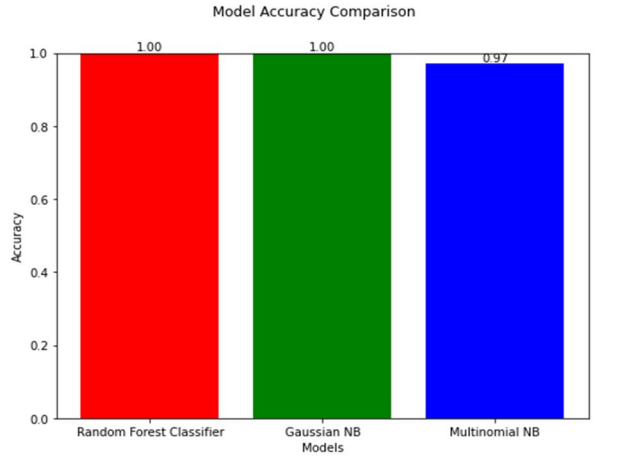


Figure 6.0 Model Accuracy Comparison

G. Derived Insight

What is the rate of patient with early malaria disease detection and grading with the chosen parameters? This model made us to realize that Malaria disease can be treated with a clean environment, with or without visiting doctors and equally eradicated with the availability of treated net, basic infrastructures based on the environment/houses patient they live, and detection of symptoms to reduce the severalty of the disease.

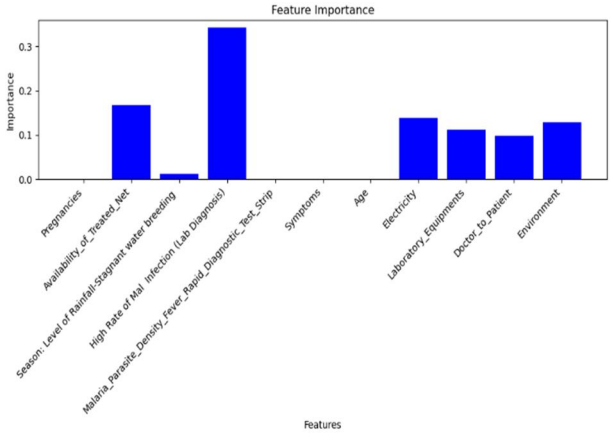


Figure 7.0 Derived Insight

V. CONCLUSION

The advancements in machine learning and deep learning have made it possible to create models for accurate predictions in various domains such as weather forecasting systems, speech recognition systems, and recommender systems. The accuracy of these models depends on the domain, problem statement, and accuracy requirements, and can be further improved after evaluating all relevant metrics. This study focuses on designing a multinomial Naïve Bayes (MNB) model. To address the problem of the prevalent malaria disease in Sub-Saharan Africa, a dataset was formulated using existing features and novel features (Hospital infrastructures). The data were collected secondarily from the General Teaching Hospitals in Abuja, Nigeria. The dataset was used to make analysis and prediction. During analysis, it was found out that early detection of Malaria disease can be achieved and solution prescribed before it will go out of hand in becoming a complicated issue.

The research work developed a Multinomial Naïve Bayes (MNB) model for Malaria disease prediction and grading in human being. This model was built using the Python package. The goal of this research is to compare the effectiveness of the Multinomial Naïve Bayes (MNB) classifiers in predicting Malaria diseases. We used three different implementations of Multinomial Naïve Bayes (MNB) classifier: the Multinomial Naïve Bayes (MNB), Gaussian Naïve Bayes and Random Forest. The Multinomial Naïve Bayes (MNB) produced a graphical representation of the dependencies between attributes.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

The obtained model helps us to identify the causal dependencies and conditional independencies between attributes. Dataset was collected from the public hospitals in Abuja, FCT. We used 2121 instances with 15 attributes; 13 numeric input attributes and two output. The Multinomial Naïve Bayes (MNB) outperformed the other Machine Learning algorithm in the prediction of Malaria diseases. This research will assist in making inferences about Malaria diseases, thereby serving as a diagnostic tool to support medical practitioners. As a future work, our model parameters can be increased and equally compared to other classifiers.

A. List of Figures

Figure 1.0: Naïve Baye Classification
Figure 2.0: Sequence in Network Flow
Figure 3.0: Flow Diagram of Network Design
Figure 4.0: Confusion Matrix Heatup Map
Figure 5.0: Training and Validity Check
Figure 6.0: Model Accuracy Comparison
Figure7.0: Derived Insight
Figure 8.0: Attributes of Malaria Disease dataset from Public Hospitals in Abuja, FCT

B. List of Tables

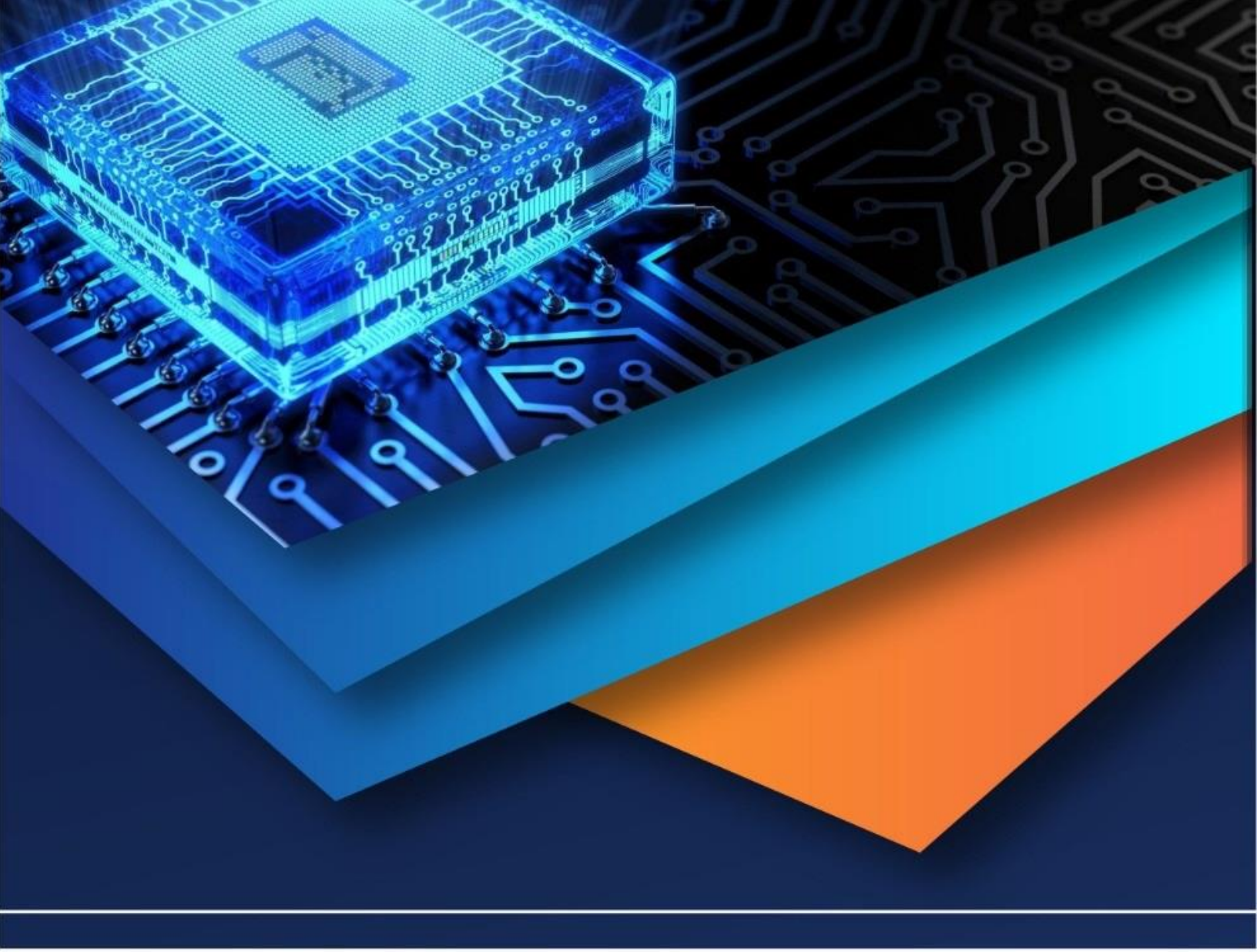
Table 1.0: Classification Report Table
Table 2.0: Confusion Matrix Report

REFERENCES

[1] Abraham and P. Siarry (2020) Engineering Applications of Artificial Intelligence https://www.journals.elsevier.com/engineeringapplications-of-artificial-intelligence.
[2] Adele. Cutler, D. R. Cutler, and J. R. Stevens(2012) , "Ensemble Machine Learning," Ensemble Mach. Learn., no. February 2014, 2012, doi: 10.1007/978-1-4419-9326-7.
[3] Amajoh, N, Odukoko, J.B. and Monsanya, M.E. (2017). Preliminary investigations on malaria and anemia in the Atlantic coastal margin of Ibeju Lekki, Lagos Nigeria. Journal of Malaria in Africa and the Tropics, 1, 17-22.
[4] Ameta, M. A. and Jain, M. K. (2017). Data mining techniques for the prediction of kidney diseases and treatment. Journal of Medical Systems Review, 6(2), 20–27.
[5] Cherian, V., & Bindu, M. S. (2017). Disease prediction using Naïve Bayes algorithm and laplace smoothing technique, 5(2), 68–73.
[6] Hakizimana, L., Cheruiyot, W. K., Kimani, S. and Nyararai, M. (2017). A hybrid based classification and regression model for predicting diseases outbreak in datasets. International Journal of Computer (IJC), 27(1), 69-83.
[7] Hijmans Abel & Bhulai S. (2017). Journal of Computer Science for Dutch Prediction of games using machine learning classifiers research paper business analytics. Corpus ID: 46691270
[8] Hillary. I. Okagbue, Pelumi E. Oguntunde, Emmanuella C.M. Obasi, Patience I. Adamu(Jan, 2021). Diagnosing Malaria from some Symptoms: A Machine Learning Approach and Public Health Implications. Health Technology- Springer Berlin Heidelberg, Vol. 11, Iss:, pp 23-37.
[9] Langarizadeh, M. and Moghbeli, F. (2016). Applying Naive Bayesian networks to disease prediction: a systematic review. Acta Informatica Medica, 24(5), 364.
[10] Lucini, F. R., S. Fogliatto, F., C. da Silveira, G. J., L. Neyeloff, J., Anzanello, M. J., de S. Kuchenbecker, R.D. and Schaan, B. (2017). Text mining approach to predict hospital admissions using early medical records from the emergency department. International Journal of Medical Informatics, 100, 1–8.
[11] M. Kishan (2019) Analytics India," 9 Feb 2019. [Online]. Available: https://www.analyticsindiamag.com/5-ways
[12] Nicholas J. White (2022) Severe Malaria Disease, White Malaria Journal No. 21:284 https://doi.org/10.1186/s12936-022-04301-8
[13] Patil, R., Chopade, P., Mishra, A., Sane, B. and Sargar, Y. (2016). Disease prediction system using data mining hybrid approach. Communications on Applied Electronics Published by Foundation of Computer Science (FCS), NY, USA, 4(9), 48–51.
[14] Poostchi, M., Silamut, K., Maude, R. J., Jaeger, S., & Thoma, G. (2018). Image analysis and machine learning for detecting malaria. Translational Research, 194, 36-55.
[15] Sajana. T., Narasingarao .M.R. (2018) Classification of Imbalanced Malaria Disease Using Naïve Bayesian Algorithm. International Journal of Engineering & Technology, 7 (2.7) (2018) 786-790 .https://: www.sciencepubco.com/index.php/IJET
[16] Tribhuvan, A. P., Tribhuvan, P. P. and Gade, J. G. (2015). Applying Naive Bayesian classifier for predicting performance of a student using Weka. Advances in Computational Research, 7(1), 239.
[17] Vijayarani, S. and Deepa, S. (2014). Naïve Bayes Classification for Predicting Diseases in Haemoglobin Protein Sequences. International Journal of Computational Intelligence and Informatics, 3(4).
[18] Vijayarani, S. and Dhayanand, S. (2015). Liver disease prediction using SVM and Naïve Bayes algorithms. International Journal of Science, Engineering and Technology Research (IJSETR), 4(4), 816-820.
[19] World Health Organisation (2016): Malaria fact sheet No 94, April. http://www.who.int/mediacentre/factsheets/fs094/en/
[20] World Health Organisation (2018): Malaria fact sheet No 98, October. http://www.who.int/mediacentre/factsheets/fs094/en/
[21] W.H.O, W. H. O. (2021). WHO guidelines for malaria, 13 July 2021.

[22] W.H.O, W. H. O. (2023). WHO guidelines for malaria, 27 March 2023.
[23] National Centre for Disease and Control, NCDC, 2018
[24] Federal Ministry of Health (FMoH), 2018.
[25] W.H.O, W. H. O. (2018). WHO guidelines for malaria, 27 March 2018.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)