CONTENTS

MALARIA PREDICTION RECOGNITION

1.INTRODUCTION

1.1 INTRODUCTION:

Artificial Intelligence is a way of making a computer, a computer-controlled robot, or a software think intelligently, in the similar manner the intelligent humans think. AI is accomplished by studying how human brain thinks and how humans learn, decide, and work while trying to solve a problem, and then using the outcomes of this study as a basis of developing intelligent software and systems.

Machine Learning may be defined as the field of computer science, more specifically an application of artificial intelligence, which provides computer systems the ability to learn with data and improve from experience without being explicitly programmed.

Basically, the main focus of machine learning is to allow the computers learn automatically without human intervention. It can be started with the observations of data. The data can be some examples, instruction or some direct experiences too. Then on the basis of this input, machine makes better decision by looking for some patterns in data.

1.2 OBJECTIVES OF RESEARCH:

- To promote the implementation of evidence based strategies for malaria control through sustained technical support and partnerships.
- To facilitate the access of populations at risk to effective treatment of malaria.
- To support the application of effective preventive measures against malaria for population at risk through integrated vector management.
- To strengthen capacity building for malaria control in the member countries.
- To assist in the strengthening of malaria surveillance systems and the monitoring and evaluation of malaria control at all levels.

1.3 PROBLEM STATEMET:

Malaria is a deadly infectious disease that is transmitted by protozoan parasites and poses huge public health crises in poor countries. The disease is the most common cause of outpatient visits in Nigeria. Most rural Nigerian towns experience high malaria transmission,

with an estimated 990,000 people dying from malaria in 1995; that is more than 2,700 deaths per day, or 2 deaths per minute. Among children under 5, malaria is responsible for the deaths of 1.5 to 2 million people every year. An initial literature review revealed that the relationship between culture and health care seeking behavior is unclear. Furthermore, the nature of the relationship between the local religions practiced in Anambra and malaria treatment seeking behavior needs to be clarified. The problem is that, while scientists know that malaria is one of the most complex diseases facing mankind, and is responsible for 25% infant mortality and 30% childhood mortality in Nigeria. Researchers do not know how local cultures and religions influence how people see malaria treatment.

2. Review of Literature:

On doing the literature survey of various methods for malaria value prediction, we come to the conclusion that to predict the resale value there are multiple approaches like

- Decision Tree
- Random Forest

Decision Tree Classifier:

Decision Tree is a supervised machine learning algorithm used to solve classification problems. The main objective of using Decision Tree in this research work is the prediction of target class using decision rule taken from prior data. It uses nodes and inter nodes for the prediction and classification. Root nodes classify the instances with different features. Root nodes can have two or more branches while the leaf nodes represent classification. In every stage, Decision tree chooses each node by evaluating the highest information gain among all the attributes.

Random Forest:

Random Forest is a flexible, easy to use machine learning algorithm that produces, even without hyper-parameter tuning, a great result most of the time. It is also one of the most used algorithms, because it’s simplicity and the fact that it can be used for both classification and regression tasks. In this post, you are going to learn, how the random forest algorithm works and several other important things about it.

3. DATA COLLECTION:

We have collected information about appropriate variables, scale & facts regarding our study. Some information was collected from internet, official website. the dataset that is collected is malaria prediction.csv. The independent variables are the symptoms of malaria (i.e.,fever, bitter_tongue, rigour, fatigue, vomittings etc.,) and the dependent variable shows that the person has a severe malaria or not .

4. METHODOLOGY :

4.1 EXPLORATORY DATA ANALYSIS:

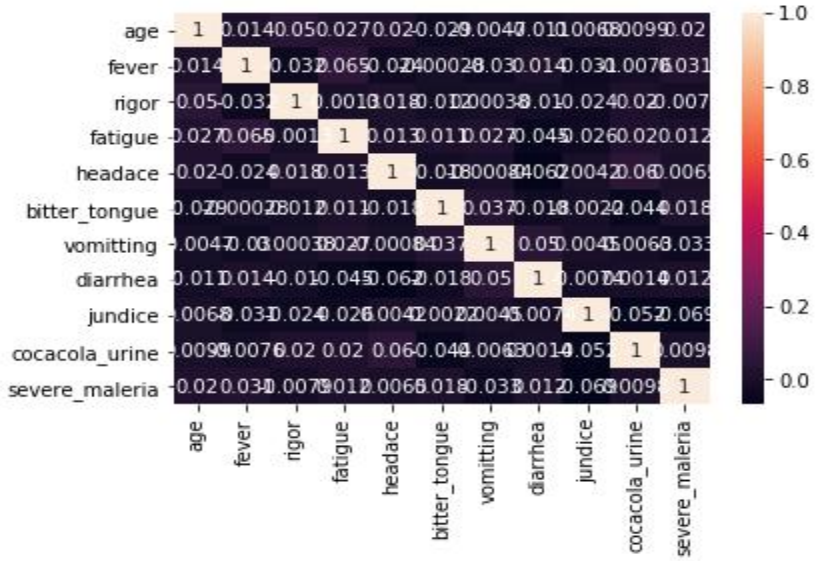
Procedures for analyzing data, techniques for interpreting the results of such procedures, ways of planning the gathering of data to make its analysis easier, more precise or more accurate, and all the machinery and results of (mathematical) statistics which apply to analyzing data.

Exploratory data analysis, robust statistics, non parametric statistics, and the development of statistical programming languages facilitated statisticians' work on scientific and engineering problems. Such problems included the fabrication of semiconductors and the understanding of communications networks, which concerned Bell Labs. These statistical developments, all championed by Tukey, were designed to complement the analytic theory of testing statistical hypotheses, particularly the Laplacian tradition's emphasis on exponential families.

4.1.1Figures and Tables

```
import seaborn as sns
sns.heatmap(malaria_dataset.corr(),annot=True)
```

```
<matplotlib.axes._subplots.AxesSubplot at 0xb6e42b0>
```



Dataset for Malaria Prediction:

	fever	rigor	fatigue	headace	bitter_tongue	vomitting	diarrhea	jundice	cocacola_urine	sex_Female	sex_Male
0	1	0	0	0	1	0	0	1	0	0	1
1	0	1	1	1	0	0	0	1	0	0	1
2	1	0	1	0	1	0	1	0	1	1	0
3	1	1	0	1	0	0	0	1	1	1	0
4	1	1	0	1	0	0	1	1	1	1	0
5	1	0	1	1	0	1	1	1	1	1	0
6	1	1	1	1	0	0	0	1	1	0	1
7	1	0	0	1	1	0	0	1	0	0	1
8	1	0	1	1	1	0	0	0	1	1	0
9	1	1	0	1	1	0	0	0	0	0	1
10	0	0	0	0	1	0	1	0	0	1	0
11	1	0	0	1	0	0	1	1	0	1	0
12	1	1	0	1	0	0	0	1	1	0	1
13	1	0	1	1	1	0	0	0	0	0	1
14	1	0	1	1	1	0	0	0	1	0	1
15	1	0	1	0	1	0	1	1	0	1	0

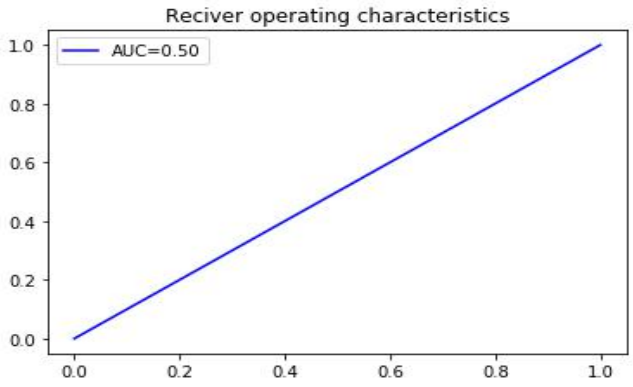
AUC and ROC Curves:

For Decision Tree:

```
import sklearn.metrics as metrics
fpr,tpr,threshold=metrics.roc_curve(y_test,y_predict)
roc_auc=metrics.auc(fpr,tpr)
roc_auc
```

```
0.5
```

```
plt.title("Reciver operating characteristics")
plt.plot(fpr,tpr,label="AUC=%0.2f"%roc_auc,color="blue")
plt.legend()
plt.show()
```

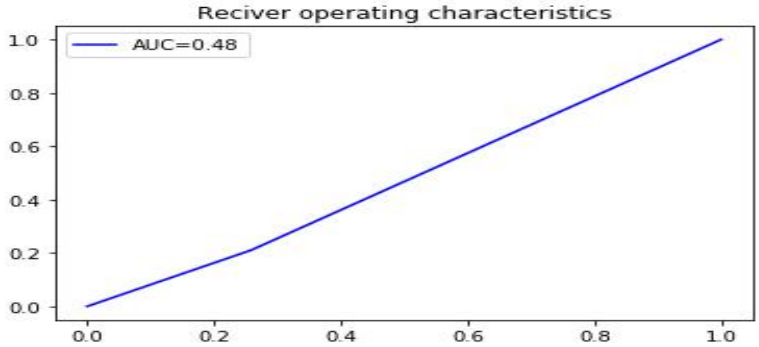


For Random Forest:

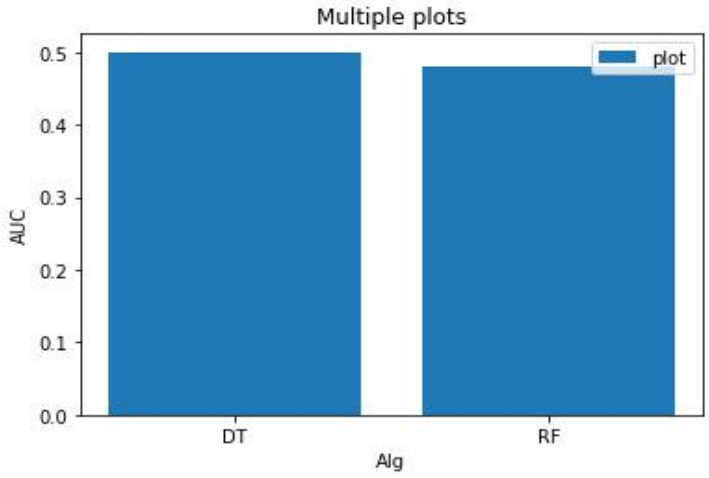
```
import sklearn.metrics as metrics
fpr,tpr,threshold=metrics.roc_curve(y_test,y_predict)
roc_auc=metrics.auc(fpr,tpr)
roc_auc
```

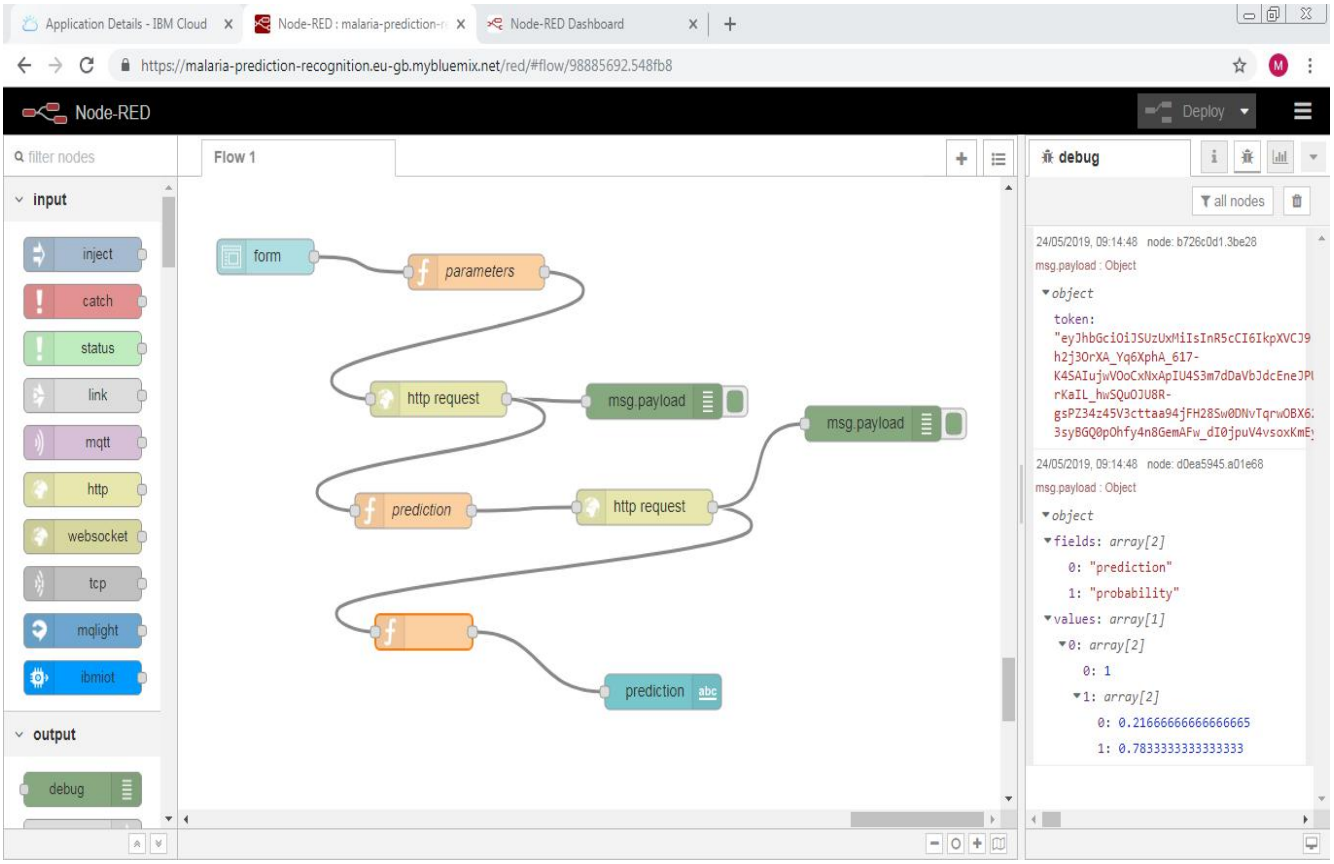
```
0.47599486521181
```

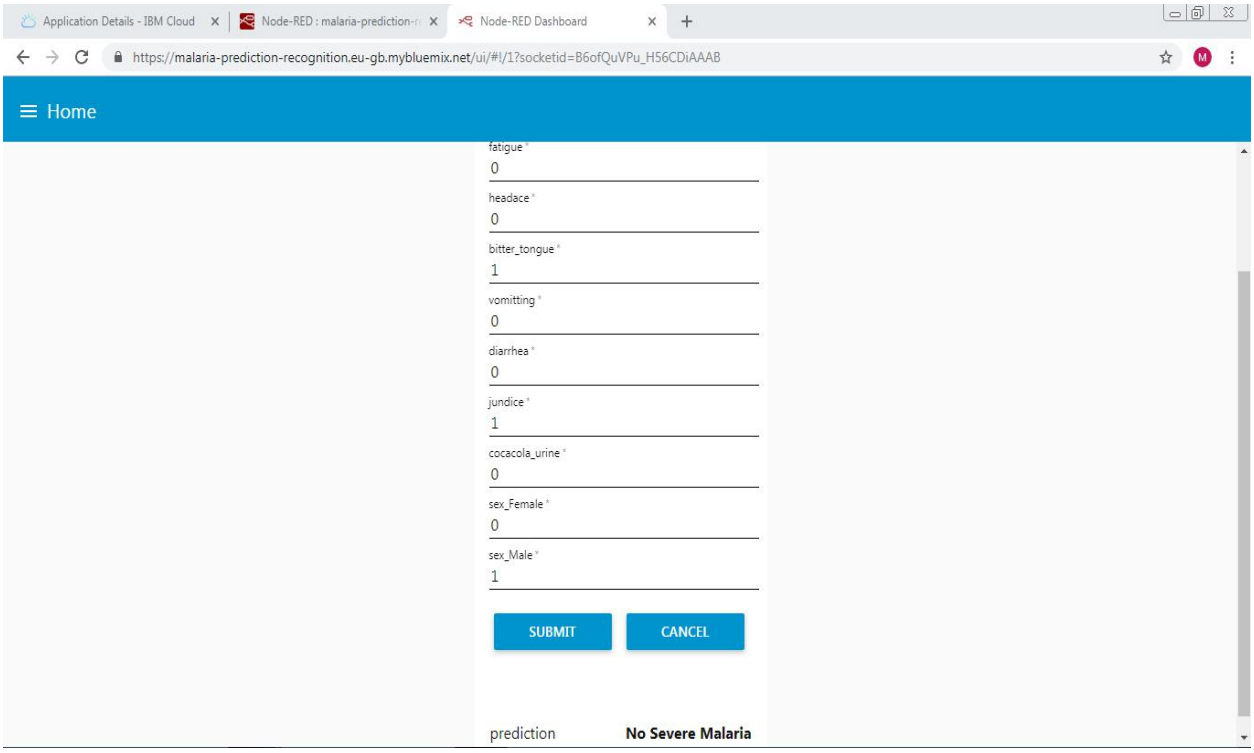
```
plt.title("Reciver operating characteristics")
plt.plot(fpr,tpr,label="AUC=%0.2f"%roc_auc,color="blue")
plt.legend()
plt.show()
```

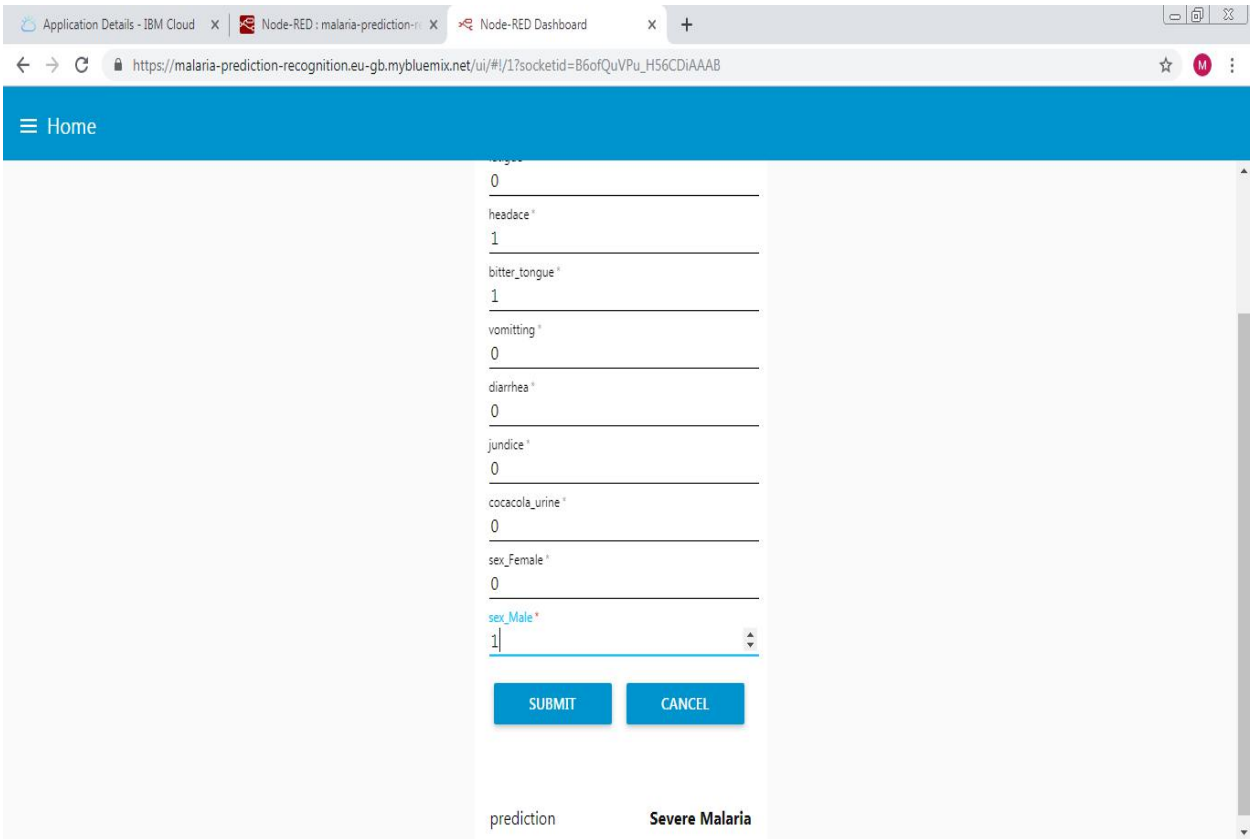


```
]: x=['DT','RF']
   y=[0.5,0.48]
   plt.bar(x,y,label="plot")
   plt.xlabel("Alg")
   plt.ylabel("AUC")
   plt.title("Multiple plots")
   plt.legend()
   plt.show()
```






Application Details - IBM Cloud
Node-RED : malaria-prediction-r
Node-RED Dashboard
https://malaria-prediction-recognition.eu-gb.mybluemix.net/ui/#!/1?socketid=B6ofQuVPu_H56CDiAAAB
Home
fatigue *
0
headace *
0
bitter_tongue *
1
vomitting *
0
diarrhea *
0
jundice *
1
cocacola_urine *
0
sex_Female *
0
sex_Male *
1
SUBMIT
CANCEL
prediction
No Severe Malaria


Application Details - IBM Cloud
Node-RED : malaria-prediction-r
Node-RED Dashboard
https://malaria-prediction-recognition.eu-gb.mybluemix.net/ui/#!/1?socketid=B6ofQuVPu_H56CDiAAAB
Home
0
headace *
1
bitter_tongue *
1
vomitting *
0
diarrhea *
0
jundice *
0
cocacola_urine *
0
sex_Female *
0
sex_Male *
1
SUBMIT
CANCEL
prediction
Severe Malaria

4.2 STATISTICAL TECHNIQUES AND DATA VISUALIZATION

Data visualization is an important skill in applied statistics and machine learning. Statistics does indeed focus on quantitative descriptions and estimations of data. Data visualization provides an important suite of tools for gaining a qualitative understanding. This can be helpful when exploring and getting to know a dataset and can help with identifying patterns, corrupt data, outliers, and much more. With a little domain knowledge, data visualizations can be used to express and demonstrate key relationships in plots and charts that are more visceral to yourself and stakeholders than measures of association or significance.

4.3 DATA MODELLING USING SUPERVISED ML TECHNIQUES

We consider Decision Tree here. Because the accuracy we get is more compared to all the models.Decision Tree is a classification algorithm. It is used to predict a binary outcome (0/1, Yes/No, True/False) from the set of independent variables. It is a special case of decision tree where the outcome variable is also a binary outcome .

5.FINDINGS AND SUGGESTIONS:

Findings:

- Malaria is common disease which causes by the mosquito bytes and if it is not cured within the time,It will be dangerous and very severe that it can kill humans.
- Malaria is found with particular health issues and it can be discovered with the blood cells.Malaria persons will get infected blood cells that can infect total body of that person.
- Symptoms that show a person has malaria are high fever,rigor,vomitting,body pains,headache,jaundice,fatigue etc.,
- If it becomes severe then there is no proper cure for that.
- So,we have to prevent this disease before it gonna be severe.

Suggestions:

- Don’t go to the places where mosquitoes are high numbered and don’t go to the very dirty and garbage areas.
- Use mosquito killing methods(All out,Jet coins,Nets) to prevent mosquitoes.

- Be careful at the time of receiving the blood from blood donors.
- Be alert from dangerous mosquitoes to prevent mosquito bite.
- Take preventive medicines and vaccines.

6.CONCLUSION:

One of the important medical problems is the detection of malaria at its early stage.In this study, systematic efforts are made in designing a system which results in the prediction of disease like malaria. During this work, five machine learning classification algorithms are studied and evaluated on various measures. Experimental results determine the adequacy of the designed system with an achieved accuracy of 0.70 using the Decision Tree algorithm. In future, the designed system with the used machine learning classification algorithms can be used to predict or diagnose other diseases. The work can be extended and improved for the automation of diabetes analysis including some other machine learning algorithms.