iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 10 **Issue:** XII **Month of publication:** December 2022

DOI: https://doi.org/10.22214/ijraset.2022.48002

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 10 Issue XII Dec 2022- Available at www.ijraset.com

Disease Prediction System using naïve bayes

Sameer Meshram[1], Shital Dongre[2], Triveni Fole[3]
[1, 2, 3]*Department of Information Technology Vishwakarma Institute of Technology*

Abstract: ***Accurate and on-time analysis of any health-re- lated problem is vital for the prevention and treatment of the illness. The standard way of diagnosis might not be suf-ficient. Developing a diagnosis system with machine learn- ing (ML) algorithms for prediction of any disease can helpin a very more accurate diagnosis than the traditional method. The proposed model is an Disease Prediction System with the help of machine learning algorithm Naive Bayes which takes the symptoms as the input and it gives the output as predicted disease. It results in saving time and also makes it easy to induce a warning about your health before it's too late. By using this model anyone can get the result as pre- dicted disease by simply given the symptoms as input. The accuracy of this model is more than existing models.***
Keywords: Disease prediction, more accuracy, symptoms,naïve bayes.

I. INTRODUCTION

It is observed that near about 67-70 % of the population of India were affected by some common diseases like flu, cold cough, diarrhoea etc in each couple of months .Numberof peoples even don't realize that some common symptoms can give rise to major diseases. It is estimated that 20-25 % of population results to very bad situation and deaths just because of avoiding some common symptoms .Due to such concern catching or predicting such small diseases at very early stage is able to resist unwanted casualties. The existingsystem has the capability to just predict any specific diseasewhich will bound its limits.

The main objective of the system is to predict such dis- eases that if we avoid or when unchecked can results into fatal situations. By embedding machine learning algorithmslike naive bays part of supervised learning algorithm .This model can generate a prediction of the most possible diseasematching with symptoms which is trained using machine learning mechanism. The proposed model is provide quick medical diagnosis to users. It is very easy to an user to get to know the predicted chronic disease by simply giving thesymptoms as input.

II. LITERATURE SURVEY

In the proposed model "Disease Prediction Using MachineLearning" the author has discussed that the Disease predic-tion based on history of patient treatment and health data by applying machine learning and data processing tech- niques with the ongoing struggle for the past decades. Alsoin the number of cases they are applying the data pro- cessing onto the medical profiles for prediction of specificdiseases. These approaches tried to predict the occurrenceof disease. **[1]**

In the proposed model "A Smart Health Prediction Using Data Mining" the information about mining techniques like association rule mining and classification, clustering and also proposed a model which makes use of various tech- niques like machine learning ,artificial Intelligence, man- agement techniques for the extraction of new patterns from large datasets and therefore the knowledge related to these patterns. By using this technique data can be extracted auto-matically or semi automatically. the various parameters in- cluded indata processing are clustering, forecasting, path analysis and predictive analysis.**[2]**

In the paper "General Disease Prediction System" the au- thor has discussed about data processing techniques and ID3 decision tree algorithms. within the project, the dis- ease prediction system will perform data processing in its preliminary stages, also the system are trained using ma- chine learning and data processing to form the prediction for the final and more commonly occurring disease that when uncheckedcan transform fatal disease.**[3]**

In the proposed mode "Smart E-Health Prediction System Using Data Mining" it discusses the Bayesian statistics and posterior distribution the puzzles that are solved highly and also told us about data processing techniques like as- sociation rule mining, classification, clustering. Also Bayesian statistics successfully can be applied on different fields like science, economic . In the fields like medical fields, the foreign students have solved some medical is- sues that are laborious to be settled in classic statistics by classification of Bayesian.**[4]**

In the paper "Medi-Insight: A Smart Health Prediction Sys- tem" the author proposes an answerfor identifying diseases supported symptoms Information technologies which are being increasingly implemented in healthcare organizations.

Their system can able to predict a person's lifetime diseases are going to be able to warn the person to possess early med- ical servicesor to manage and stop exceptional situations inlife with the assistance of Multinomial Naïve Bayes Algo- rithm. The Author used Multinomial Naive Bayes which may be a variant of Naive Bayes algorithm accustomed im- plement our disease prediction model. Multinomial naive Bayes assumes to ownfeature vector where each element represents theamount of times it appears (or, very often, its frequency). this method is additionally very efficient in lin- guistic communication processingor whenever the samples are composed ranging from a standard dictionary.**[5]**

In the proposed model, K means clustering for clustering al-gorithm and later cluster the data an approach of KNN is used here the dataset of diabetes and liver disorder are usedfor performing test operation onto the defined methods.the concept of fuzzy logic is used for improving the accuracy ofclassification .**[6]**

In the proposed model , combination Bayesian and KNN(K- Nearest Neighbour) algorithm are used .here a dataset of di- abetes patient used that gives an analysis of diabetes malady**[7]**

In the proposed model, diagnosis of hearth disease based onthe previous information and related data. Here the con- cerned risk factor of hearth disease predicted using the naïvebayes. The parameters considered are sex, gender, choles- terol, bp, blood sugar etc. They are act as input for the pre- diction.[**8]**

In the proposed model, machined learning algorithms for the prediction of chronic diseases are used. Here the model is tried as experiment over the data which is collected from thereal world hospital in 2013-2015.they used here the convo- lutional neural network approach for multiple disease pre-dictions reached the prediction accuracy of 90%.**[9]**

In the proposed model, for the prediction of multiple dis- eases like diabetes analysis, heart disease, diabetes etc an API named as Flask is used. The parameters here taken intoconsideration are bmi, sex, glucose, diabetes pedigree func-tion, heart rate/pulse rate, serum creatinine, potassium Glas-gowComascale .**[10]**

III. PROPOSED MODEL

As we were thinking on how and what we can do in thedisease prediction field, So we Developed Web-app which predicts disease on basis of user's symptoms. So in this pro-ject we are predicting a disease which an individual is af- fected by depending upon the symptoms patient is suffering .In the proposed model , system is asking for symptoms and evaluating those symptoms by using the Naïve Bayes algo- rithm which works robust and faster with datasets having small in size also it gives the nice accuracy.

Naive Bayes algorithm which learns the probability of an ob- ject with certain features belonging to a specific group/class. for example, if we are trying to identify a fruit based on its taste ,shape and color then an orange colored, tangy fruit andspherical would presumably be an orange. Such properties can be used individually contribute to the probability that thisfruit is an orange. Due to these reason it is referred to as "Na-ive". As for the "Bayes" part , it refers to statistician and phi-losopher Thomas Bayes and also the theorem named after him , Bayes' theorem , which is that the base for Naïve Bayes.

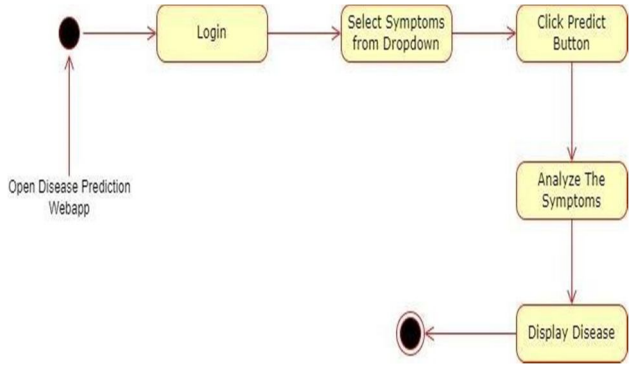


Fig: State Diagram

IV. METHODOLOGY

A. Algorithm

Step 1: Let Say D can be represented as training dataset every record can be denoted as n-dimensional attribute vector, thus we can write this as X=(x1,x2,x3,x4…..Xn) this is responsibleto predict n attributes of n measurements(Bi to Bn).

Step 2: Lets take m number of class for operation of predictionit can written as(C1,C2,C3…….Cm)

By applying the Naive Baye's theorem:

$$P(Ci \mid X) = \frac{P(X \mid Ci) * P(Ci)}{P(X)}$$

Step 3: Here for every class P(X) is constant, therefore P(X|C)*P(Ci) compulsorily be maximized.
Step 4: After, P(X| Ci) = P(x1| Ci)* P(x2| Ci) P(xm| Ci)
Step 5: for X class prediction, P(X|Ci)P(Ci) is need to find outfor each Ci class
Here classifier i.e naïve bayes is able to predict the class labelof X=Ci class
If,
P(X|Ci) P (Cj) > P(X|Cj)P(Cj)

B. Steps Of Model Building

1) *Collecting Data:* Data from excel, access, text filesetc., this step (gathering past data) can be act as a inspiration of the longer term learning. The higher the variability, volume and density of relevant data,better the aspects of learning becomes.
2) *Pre-processing the Data:* This steps used to re- move outliers and missing data to form the clean data. Exploratory analysis is maybe one method to review the nuances of the information in details thereby burgeoning the nutritional content.
3) *Training a Model:* This step involves choosing theacceptable algorithm and representation of knowledge within the type of the model. The data that we get after removing outliers and cleaned datais split into two parts – train and test (proportion reckoning on the prerequisites); the primary part (training data) is employed for developing the model. Another step(test data) is employed.

❖ Following are the Parameters that taken into consideration for training purpose are shown below:
Itching,skin_rash,nodal_skin_eruptions, continuous_sneez- ing,shivering, chills, joint_pain, stomach_pain, acidity, ul- cers_on_tongue, muscle_wasting, vomiting, burning_micturi- tion, spotting_urination, fatigue,weight_gain,anxiety, cold_hands_and_feets, mood_swings, weight_loss, restless-ness, lethargy,patches_in_throat, irregular_sugar_level, cough,high_fever, sunken_eyes, breathlessness, sweating, de- hydration, indigestion, headache, yellowish_skin, dark_urine, nausea,loss_of_appetite,pain_behind_the_eyes, back_pain,con- stipation, abdominal_pain, diarrhoea, mild_fever,yellow_urine, yellowing_of_eyes, acute_liver_failure, fluid_overload,well- ing_of_stomach,swelled_lymph_nodes,malaise,redness in eyes, ness_of_eyes,sinus_pressurerunny_nose, congestion, chest_pain, weakness_in_limbs, fast_heart_rate, pain_dur- ing_bowel_movements,pain_in_anal_region,bloody_stool,irri- tation_in_anus,neck_pain,dizziness,cramps,bruising,obe- sity,swollen_legs,swollen_blood_vessels, severe pain_in_chest puffy_face_and_eyes, enlarged_thyroid, brittle_nails, swol- len_extremeties, excessive_hunger,extra_marital_contacts,dry- ing_and_tingling_lips,slurred_speech,knee_pain,hip_joint_pain, muscle_weakness, stiff_neck,swelling_joints, movement_stiffness, spinning_movements, loss_of_bal- ance,unsteadiness,weakness_of_one_body_side,loss_of_smell, bladder_discomfort, foul_smell_of, urine,con-tinuous_feel_of_urine, passage_of_gases, internal_itching, toxic_look_(typhos), depression, irritability muscle_pain,al- tered_sensorium, red_spots_over_body, belly_pain, abnor-mal_menstruation, dischromic_patches, watering_from_eyes,increased_appetite, polyuria, family_history, mucoid_sputum,rusty_sputum, lack_of_concentration, visual_disturbances, re-ceiving_blood_transfusion, receiving_unsterile_injections,coma, stomach_bleeding, distention_of_abdomen, his- tory_of_alcohol_consumption, fluid_overload, blood_in_spu- tum, prominent_veins_on_calf, palpitations, painful_walking, pus_filled_pimples, blackheads, scurring, skin_peeling,sil-ver_like_dusting, small_dents_in_nails, inflamma- tory_nails,blister, red_sore_around_nose, yel- low_crust_ooze,prognosis.

❖ Following are the disease our model is predicting: Fungal, infection, Allergy, GERD, Chronic, cholestasis, Drug, Reaction, Peptic, ulcer, diseae, AIDS, Diabetes, Gas- troenteritis, Bronchial, Asthma, Hypertension, Mi- graine,Cervical,spondylosis, aralysis(brainhemorrhage), Jaundice,Malaria, Chicken pox, Dengue,Typhoid,hepatitis, A,Hepatitis, B,Hepatitis C, Hepatitis, D, Hepatitis, E, Alco- holichepatitis, Tuberculosis, Common, Cold, Pneumonia, Dimorphic, hemmorhoids(piles), Heart, attack, Varico- seveins, Hypothyroidism, Hyperthyroidism, Hypoglycemia, Osteoarthristis, Arthritis,(vertigo) Paroymsal, Positional, Vertigo, Acne, Urinary, tractinfection, Psoriasis, Impetigo Fungal infection

- *Evaluating the Model:* For determining the per- formance we have to denote the TP,FP,TN and FN considered as true positive i.e the instances are cor-rectly predicted. False negative i.e instances are in-correctly predicted as not required,false positive i.e instanced predicted are incorrect as required, true positive i.e instances predicted are correct as re- quired.

  Precision= $\frac{TP + TN}{TP+TN+FP+FN}$

  Accuracy = $\frac{TP}{TP+FP}$

- *Improving the Performance:* In this step it needsto choose different model altogether or increasing the number of variables to improve efficiency. Dueto these reason most of time need to spent in the datapreparation and data collection.

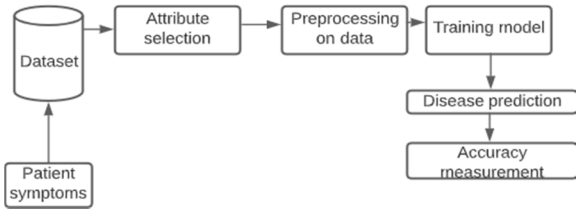


Fig: flow diagram

V. TOOLS AND TECHNOLOGY USED

A. *Front-end*

We have used HTML and CSS for front-end development withdjango framework which is a Python-based free and open- source web framework that follows the model template viewsarchitectural pattern and give us great advantage for integrat- ing with python.

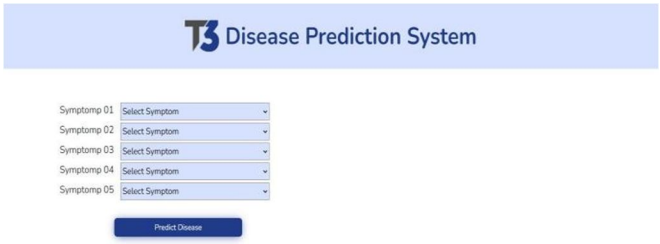
B. *Back -end*

We have used Python in back-end which providedus various libraries such as numpy, pandas and skylearn which helped us to implement this processin an easy manner.

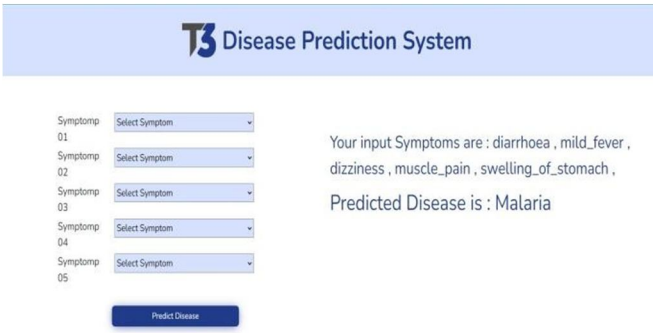
C. *Results*

1) *Homescreen*

2) *Dashboard of symptoms*



3) Predicted Result



D. Advantages

- High Accuracy compared to existing models.
- Computational Speed is High.
- Multiple disease prediction is possible.

Comparison of existing models with proposed model

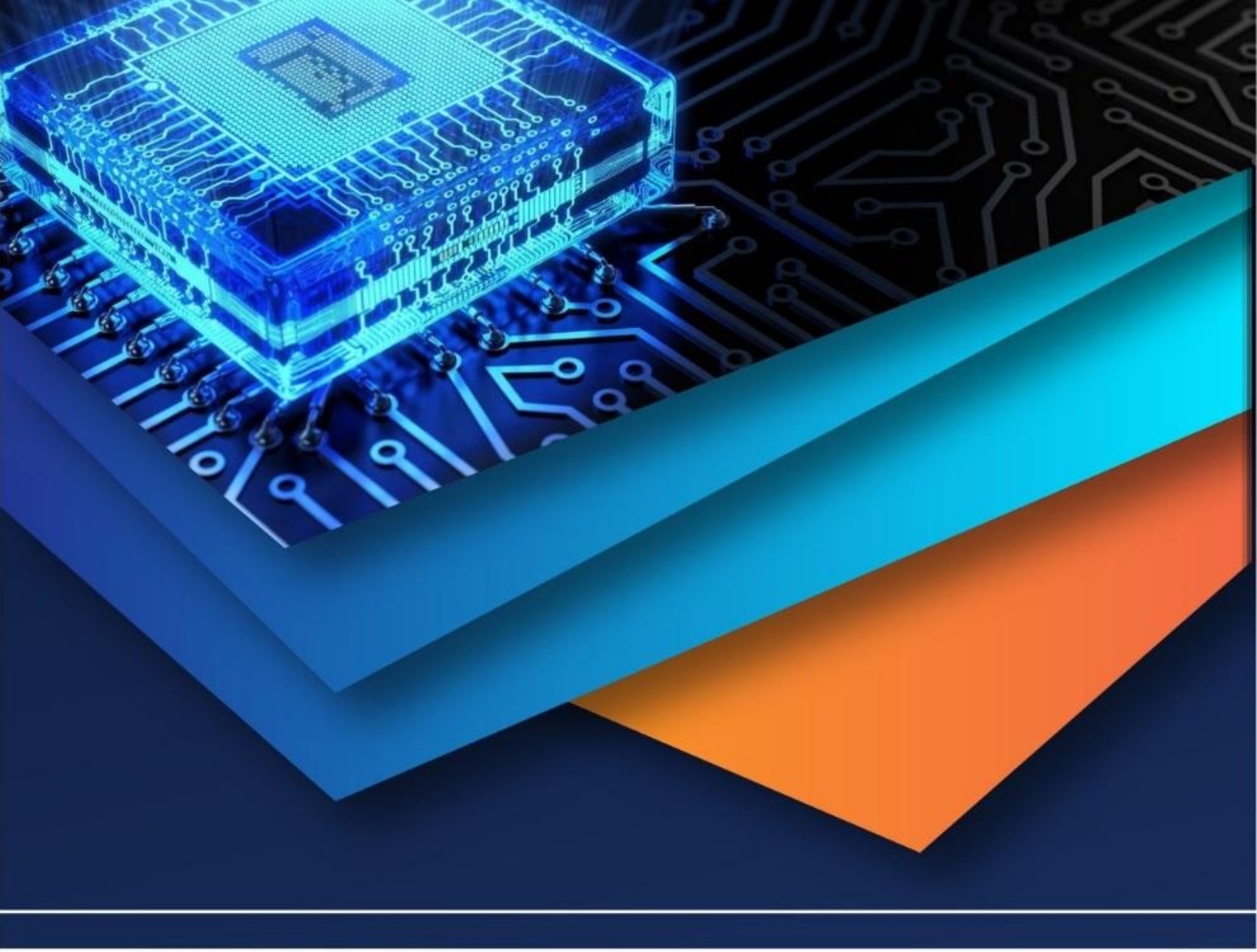
Related work	Accuracy	Diseases able to predict
Multi Disease Prediction Model by using Machine Learning and Flask API	91 %	30-40
Diabetes Disease Prediction Using Data Mining	70%	1
Disease Prediction by Machine Learning over Big Data from Healthcare Communities	94.8%	40-50
Disease Prediction System using Naïve Bayes(our proposed model)	98.2%	100-150

VI. CONCLUSION

The main motivation is to facilitate coordinated systemsand well-informed health care systems.The proposed model help any user to provide consultancy of his concerned symp- toms.Depending on the symptoms gives as input or in case nothing is given the model is capable of predicting the matched disease. The model having the capability of predict- ing the probability of diseases which can occurs results in re- duction of mortality ratio.

REFERENCES

[1] Akash C. Jamgade, Prof. S. D. Zade, International Re- search Journal of Engineering and Technology,Maaaay2019
[2] Prof. Krishna Kumar Tripathi,Shubham Jawad- war,Siddhesh Murudkar,Prince Mishra, “A Smart Health Prediction Using Data Mining” International Re- search Journal of Engineering and Technology (IR- JET),April 2018.
[3] Shratik J. Mishra 1, Albar M. Vasi 2, Vinay S. Menon3, Prof. K. Jayamalini4, “ General disease prediction sys- tem” , International Research Journal of Engineering and Technology (IRJET),March 2018
[4] G.Pooja reddy, M.Trinath basu, K.Vasanthi, K.Bala Sita Ramireddy, Ravi Kumar Tenali, “Smart E-Health Prediction System Using Data Mining”, International Journal of Innovative Technology and Exploring Engi- neering(IJITEE),April_2019.
[5] Yashaswi G Sagar1, Sahana Gajanana Acharya2, Vishal S Chincholi3, Riyal Vivek A4, Swetha P M5,” Medi-Insight: ASmart Health Prediction System”, International Research Journal of Engineering and Technology (IRJET),June 2021.
[6] [4] Naganna Chetty, Kunwar Singh Vaisla, Nagamma Patil,” An Improved Method for Disease Prediction us- ing Fuzzy Approach”,ICACCE,2015.
[7] Deeraj Shetty, Kishor Rit, Sohail Shaikh, Nikita Patil,” Diabetes Disease Prediction Using Data Min- ing”,ICIIECS,2017.
[8] Akkem Yaganteeswarudu,” Multi Disease Prediction Model by using Machine Learning and Flask API”, IEEE, July 2022.
[9] Min Chen, Yixue Hao, Kai Hwang, Fellow, Lu Wang, and Lin Wang,” Disease Prediction by Machine Learn- ing over Big Data from Healthcare Communi- ties”,IEEE, April 2016.
[10] Anjan Nikhil Repaka, Sai Deepak Ravikanti, Ramya G Franklin,” Design And Implementing Heart Disease Prediction Using Naives Bayesian”, IEEE,June 2019.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY