IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 5, June 2022
Impact Factor: 6.252

Disease Prediction using Machine Learning

Onkar Pawar[1], Suraj Gund[2], Dnyaneshwar Goykar[3]
Students, Department of Computer Science and Engineering[1,2,3]
SVERI College of Engineering, Pandharpur, India

Abstract: *"Disease Prediction" system based on predictive modelling predicts the disease of the user based on the symptoms that user provides as an input to the system. The system analyses the symptoms provided by the user as input and gives the probability of the disease as an output Disease Prediction is done by implementing the Naïve Bayes Classifier. Naïve Bayes Classifier calculates the probability of the disease. Therefore, average prediction accuracy probability 60% is obtained.*

Keywords: Predictive Modelling, Naïve Bayes Classifier

I. INTRODUCTION

At present, when one suffers from disease, then the person must visit to doctor which is time consuming and costly too. Also, if the user is out of reach of doctor and hospitals it may be difficult for the user as the disease cannot be identified. So, if the above process can be completed using a automated program which can save time as well as money, it could be easier to the patient which can make the process easier. There are other Heart related Disease Prediction System using data mining techniques that analysis the risk level of the patient.

Disease Predictor is a web-based application that predicts the disease of the user with respect to the symptoms given by the user. Disease Prediction system has data sets collected from different health related sites. With the help of Disease Predictor, the user will be able to know the probability of the disease with the given symptoms.

II. REQUIREMENT

For this, 15 real-world medical problems from the UCI machine learning repository (Asuncion and Newman, 2007) were selected for evaluating the performance of all algorithms. In the experiment it was found that NB outperforms the other algorithms in 8 out of 15 data sets so it was concluded that the predictive accuracy results in Naïve Baeyes is better than other techniques.

- Logistic Regression (LR),
- KStar (K*), Decision Tree (DT), Neural Network (NN)
- Simple rule-based algorithm (ZeroR).

2.1 Functional Requirement

- Predict disease with the given symptoms.
- Compare the given symptoms with the input datasets.

III. INPUT DATA SET

Table 2- Sample Data Sets

Symptoms	Disease
Runny nose ,Sore throat ,Cough ,Congestion, body aches, headache ,Sneezing , fever	Common cold
Fever ,profuse sweating ,headache ,nausea ,vomiting ,diarrhea ,anemia ,muscle pain ,convulsions ,coma bloody stools ,shaking chill	Malaria
poor appetite ,abdominal pain ,headaches ,generalized aches and pains ,fever ,lethargy ,intestinal bleeding or perforation ,diarrhea , constipation	Typhoid

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 5, June 2022
Impact Factor: 6.252

The algorithm implemented in this project is Naïve Bayes Classifier.

Naïve Bayes classifier depends on Bayes Theorem

Equation 1:

$$P(Y|X1, \dots\dots\dots, X_n) = \frac{P(Y)P(X1, \dots\dots, X_n|Y)}{P(X1, \dots\dots X_n)}$$

Where,

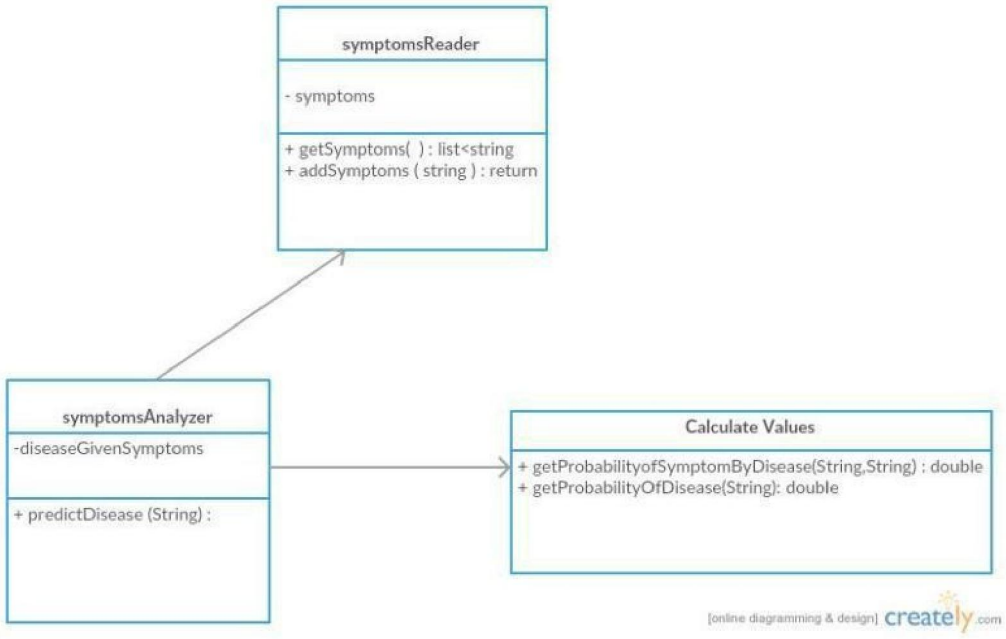
Y is the class variable

$X_1, X_2, \dots\dots, X_n$ are the dependent features

IV. WORKING

4.1 Class Diagram

It explains the classes used in the Disease Predictor classes used in total, Symptoms Reader: Reads the user input and creates the list of symptoms Symptoms Analyzer: According to symptoms parameter displays the subjective result. Calculate Values: Calculates the probabilistic model of the diseases.



4.2 State Diagram

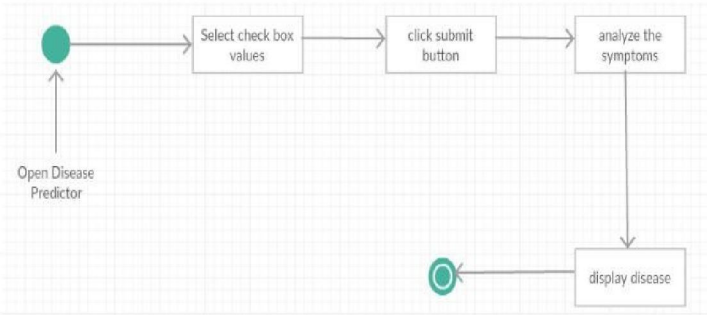


Figure 2: State Diagram

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 5, June 2022
Impact Factor: 6.252

4.3 Sequence Diagram

It explains the sequence of the Disease Predictor. Initially system shows the symptoms to be selected. The user selects the symptoms and submits to the system .The Disease Predictor predicts and display the result

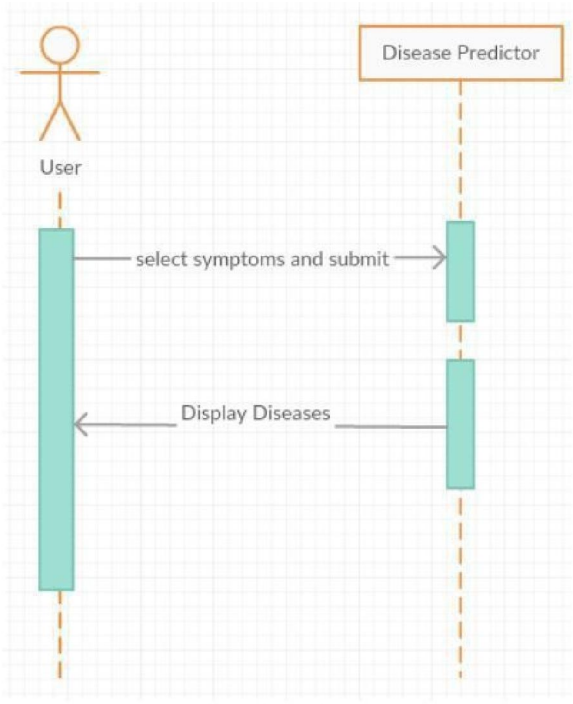


Figure 3: Sequence Diagram

V. OUTPUT

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 5, June 2022
Impact Factor: 6.252

Predictive Accuracy of Bayes and other Technique

Medical Problems	NB	LR	K*	DT	NN	ZeroR
Breast Cancer wise	97.3	92.98	95.72	94.57	95.57	65.52
Breast Cancer	72.7	67.77	73.73	74.28	66.95	70.3
Dermatology	97.43	96.89	94.51	94.1	96.45	30.6
Echoeardiogram	95.77	94.59	89.38	96.41	93.64	67.86
Liver Disorders	54.89	68.72	66.82	65.84	68.73	57.98
Pima Diabetes	75.75	77.47	70.19	74.49	74.75	65.11
Haeberman	75.36	74.41	73.73	72.16	70.32	73.53
Heart-c	83.34	83.7	75.18	77.13	80.99	54.45
Heart-statlog	84.85	84.04	73.89	75.59	81.78	55.56
Heart-b	83.95	84.23	77.83	80.22	80.07	63.95
Hepatitis	83.81	83.89	80.17	79.22	80.78	79.38
Lung Cancer	53.25	47.25	41.67	40.83	44.08	40
Lymphpgraphy	84.97	78.45	83.18	78.21	81.81	54.76
Postooerative Patient	68.11	61.11	61.67	69.78	58.54	71.11
Primary tumor	49.71	41.62	38.02	41.39	40.38	24.78
Wins	8\15	5\15	0\15	2\15	1\15	1\15

(Al-Aidaroos, Bakar, & Othman, 2012)

VI. CONCLUSION

This project aims to predict the disease on the basis of the symptoms. The project is designed in such a way that the system takes symptoms from the user as input and produces output i.e., predict disease. Average prediction accuracy probability of 55% is obtained. Disease Predictor was successfully implemented using grails framework

VII. ACKNOWLEDGMENT

We are pleased to acknowledge Dr. S.B.THIGALE (HOD CSE) sir for his valuable guidance during the course of this project work. We extend our sincere thanks to MR.V.MSALE who continuously helped us throughout the project and without his guidance, this project would have been an uphill task. We are also grateful to other members of the CSE faculty members and technical staff who cooperated with us regarding some issues.

REFERENCES

[1]. A. Davis, D., V. Chawla, N., Blumm, N., Christakis, N., & Barbasi, A. L. (2008). Predicting Individual Disease Risk Based on Medical History.

[2]. Adam, S., & Parveen, A. (2012). Prediction System for Heart Disease Using Naive Bayes.

[3]. Al-Aidaroos, K., Bakar, A., & Othman, Z. (2012). Medical Data Classification with Naive Bayes Approach. Information Technology Journal.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 5, June 2022
Impact Factor: 6.252

[4]. Darcy A. Davis, N. V.-L. (2008). Predicting Individual Disease Risk Based on Medical History. JyotiSoni, Ansari, U., Sharma, D., & Soni, S. (2011). Predictive Data Mining for Medical Diagnosis: An Overview of Heart Disease Prediction.

[5]. K.M. Al-Aidaroos, A. B. (n.d.). K.M. Al-Aidaroos, A. B. (n.d.). 2012. Medical Data Classification with Naive Bayes Approach.

[6]. NisharBanu, MA; Gomathy, B. (2013). Disease Predicting System Using Data Mining Techniques

BIOGRAPHY

Name: Suraj Balkrushna Gund
College_Name:Sveri's College Of Engineering Pandharpur Branch: Computer Science And Engineering Designation :Student

Name: Dnyaneshwar Annasaheb Goykar College_Name:Sveri's College Of Engineering Pandharpur Branch: Computer Science And Engineering
Designation: Student

Name: Dnyaneshwar Annasaheb Goykar College_Name:Sveri's College Of Engineering Pandharpur Branch: Computer Science And Engineering
Designation : Student