Applying Naïve Bayesian Classifier for Getting Probability Based Result for E-Knowledge Services In Healthcare

Vaishali V Kaneria
Research Scholar, KSV University, Gandhinagar,Gujarat, India
Asst. Prof., Department of MCA, AITS, Rajkot, Gujarat, India

Dr. N. N. Jani
Dean CS, KSV University,
Director, SKPIMCS, Gandhinagar

Ms. Richa Mehta
Research Scholar-Singhania University, Rajasthan, India

Gautam Kamani
Research Scholar, KSV University, Gandhinagar, Gujarat, India

Abstract

In the existing health system various systems of medicine are available. Here a special EHR has been prepared to integrate multiple SOM. By applying some data mining technique result can be derived intelligently from the existing available data in EHR. Naïve Bayesian classification technique is selected for getting the result because it can give probability based evaluation. Weka tool is used for the selected algorithm. Different test options are used in Weka to dissect the algorithm at better extent.

Index Terms— *SOM (System of Medicine), EHR (Electronic Health Record), Naïve Bayesian*

I. INTRODUCTION

There are many systems of medicine available to get cured. Here Allopathic and Ayurvedic system of medicine is taken as consideration. This research is focused on building a novel architecture integrating two SOM and a special care is taken to develop an EHR database involving medical experts in both the system in order to transform the architecture into a prototype and its testing to justify the research objectives.

The EHR database can have any number of attributes but here selected 5 attributes are shown in the below sample EHR data set. Further the data needs a classification before the processing. The classification is based on set of condition evaluation. The classifier which can take care of this conditional classification is supplemented with the use of Weka tool [1].

This work is targeted to obtain condition based probability results. To satisfy this purpose the Naïve bayes classification algorithm supported in Weka is selected for the purpose of computation of results.

II. SAMPLE EHR DATASET

Age	Gender	SOM	Disease	Cured
<=20	M	AYU	Allergy	NO
<=20	M	ALO	Allergy	NO
21-40	M	AYU	Allergy	YES
>40	M	AYU	RA	YES
>40	F	AYU	fever	NO
>40	F	ALO	fever	YES
21-40	F	ALO	fever	YES
<=20	M	AYU	RA	NO
<=20	F	AYU	fever	YES
>40	F	AYU	RA	YES
<=20	F	ALO	RA	YES
21-40	M	ALO	RA	YES
21-40	F	AYU	Allergy	YES
>40	M	ALO	RA	NO

III. SELECTION OF SUITABLE DM METHODOLOGY:

For getting the result some of the Data mining algorithms were studied. From that Naïve Bayesian is selected because in this algorithm probability based conditions can be evaluated and result can be get based on that.

Naïve Bayesian Classifier computations [2]:

- $P(C_i)$: P(Disease Cured = "yes") = 9/14 = 0.643
  P(Disease Cured = "no") = 5/14= 0.357

- Compute $P(V|C_i)$ for each class
  P(age = "<=20" | Disease Cured = "yes") = 2/9 = 0.222
  P(Diagnosis = "RA" | Disease Cured = "yes") = 4/9 = 0.444
  P(Gender = "F" | Disease Cured = "yes) = 6/9 = 0.667
  P(SOM = "AYU" | Disease Cured = "yes") = 6/9 = 0.667

  P(age = "<= 20" | Disease Cured = "no") = 3/5 = 0.6
  P(Diagnosis = "RA" | Disease Cured = "no") = 2/5 =0.4
  P(Gender = "F" | Disease Cured = "no") = 1/5 = 0.2
  P(SOM = "AYU" | Disease Cured = "no") = 2/5 = 0.4

Test Case V = (age <= 30 , Diagnosis = RA, Gender = F, SOM = AYU)

Calculating the probability of disease being cured based on the parameter given in V:

$P(V|C_i)$: P(V|Disease Cured = "yes") = 0.222 x 0.444 x 0.667 x 0.667 = 0.044

Calculating the probability of disease not being cured based on the parameter given in V:

$P(V|C_i)$: P(V|Disease Cured = "no") = 0.6 x 0.4 x 0.2 x 0.4 = 0.019

$P(V|C_i)*P(C_i)$: P(V|Disease Cured = "yes") * P(Disease Cured = "yes") = 0.028……..(a)

P(V|Disease Cured = "no") * P(Disease Cured = "no") = 0.007……..(b)

Decision:
Case V belongs to class ("Disease Cured = yes")
as the probability=0.028…….(a) >
probability=0.007…….(b)

IV. NAÏVE BAYESIAN ALGORITHM IN WEKA:

The implementation of naïve bayes algorithm is implemented in Weka environment with initially sample data and later on the prototype is to be implemented with EHR having huge datasets.

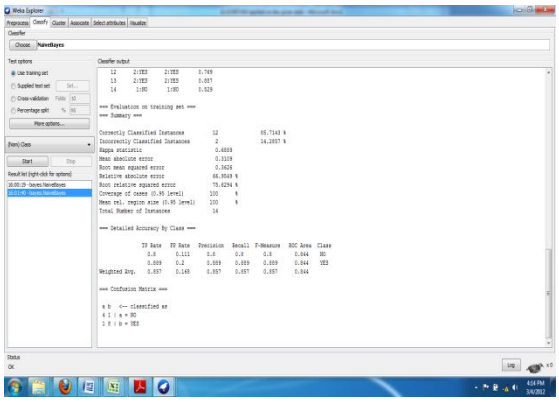
Weka supports four Naive Bayesian Algorithm Testing Options are:

1) Use training set
2) Supplied test set
3) Cross validation Folds
4) Percentage Split

V. APPLYING NAÏVE BAYESIAN

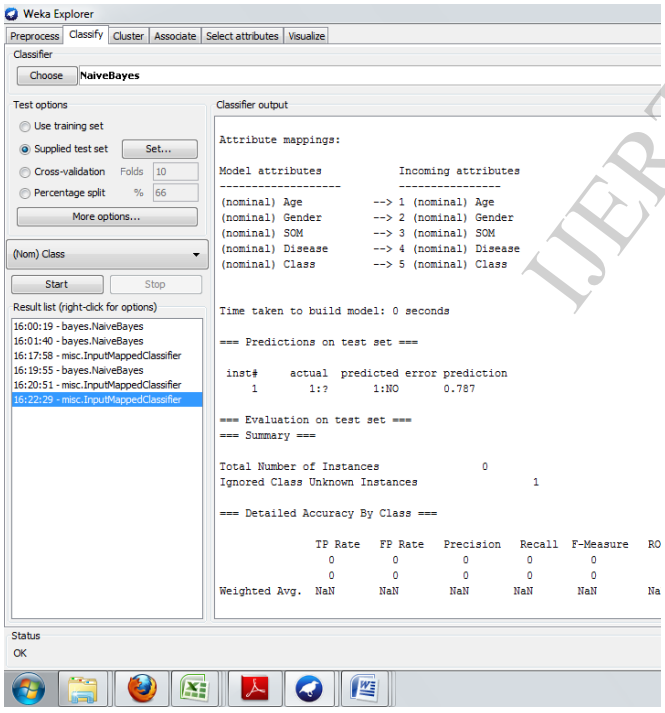
The EHR dataset is subjected to all four options of Naïve Bayesian algorithm.

1) Testing using Training set



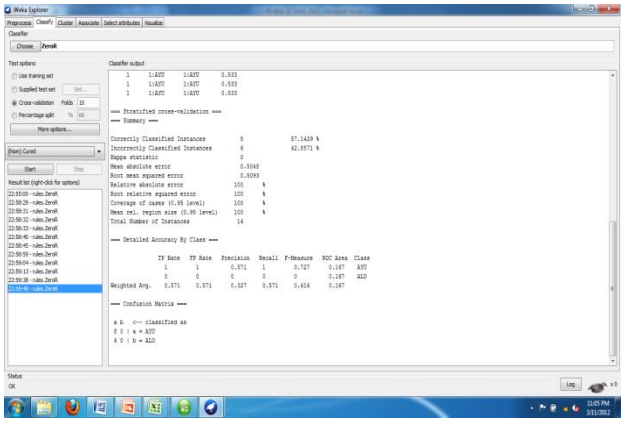
Correctly Classified Instances: 85.7%
Incorrectly classified instances: 14.3%

2) Testing using supplied test set



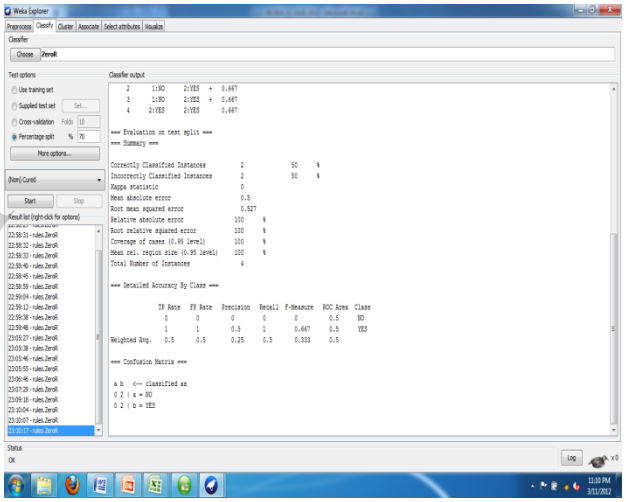
Correctly Classified Instances: 60%
Incorrectly classified instances: 40%

3) Testing using Cross validation Folds



Correctly Classified Instances: 57.14%
Incorrectly classified instances: 42.9%

4) Testing data using Percentage split



Correctly Classified Instances: 50%
Incorrectly classified instances: 50%

VI. CONCLUSION:

Here the results are shown which are taken using Weka tool. Using all the types of Naïve Bayesian options results has been taken. The Results from Naïve Bayesian training set, Naïve Bayesian supplied test set, Cross validation folds and Percentage split are not consistent. Improved algorithms are needed to be build for more consistent and improved result. This can be further utilized in terms of services for the users.

REFERENCES

[1] http://www.cs.waikato.ac.nz/ml/weka/
[2] http://www.inf.u-szeged.hu/~ormandi/ai2/06-naiveBayes-example.pdf
[3] http://www.ijetae.com/Volume2Issue2.html