Available Online at www.ijcsmc.com

International Journal of Computer Science and Mobile Computing

A Monthly Journal of Computer Science and Information Technology

ISSN 2320–088X

IJCSMC, Vol. 4, Issue. 5, May 2015, pg.341 – 348

RESEARCH ARTICLE

A Research on Expert System using Decision Tree and Naive Bays Classifier

Jatin Patel[1], Nikita D. Patel[2], Nikita S. Patel[3]

[1]PG *student*, Department of Computer Engineering, Kalol Institute of Technology and Research Centre, India

[2,3]Asst. Prof., Department of Computer Engineering, Kalol Institute of Technology and Research Centre, India

[1] jatince@yahoo.com

Abstract — Expert system means the system which makes decision based on provided knowledge by human expert and displays the result. But, knowledge provided by the human expert may varies from other expert. So, sometimes this procedure becomes subjective to the human expert. To eliminate this, we use decision tree classifier for initial phase of the expert system. From the decision tree we select the minimum list of attributes from the initial dataset and create new minimal dataset for use in Inference engine. In second phase we use naive bays classifier to create the prediction model. Which improve the accuracy of prediction of class on testing dataset.

Keywords— Expert System, classification, Decision Tree classifier, Naïve Bays classifier

I. INTRODUCTION

Expert system is the system which is take input from the user about specific domain and gives the probabilistic result based on the knowledge store in the system in terms of rules. This type of system is known as the Rule-based expert system. Rules are defined by knowledge engineer. Expert system is use for taking decision from the facts in real world.

In Rule-based expert system, system takes some facts from the real world and makes decision based on those facts. But the efficiency of the system is decrease when size of problem domain is large. To improve the efficiency and accuracy we use classification technique in expert system. As we know classification is technique to classify the large data into small classes. Data classification process contains two steps. 1Learning step: In this classifier model is constructed using training dataset. 2. Classification step: In this step constructed model is use to classify the testing data. So, in expert system we try to make small size database using classification.

II. INTRODUCTION TO EXPERT SYSTEM

In this section first of all we understand what are expert system and its working. In expert system knowledge represent is in terms of if-then rules. Rule based expert system means the system which declare rules using if-then statements.

As shown in figure1 these are the components of expert system. Knowledge base is a main component of this system. It is also called Rule base. Knowledge base is a collection of rules which are produce by the knowledge engineer through knowledge acquisition. Knowledge is acquired from the knowledge expert. User interface provide the input for the system. After getting the input from the user the Inference engine is come into the work. Inference engine takes input from the user interface and then perform the matching with the predefined rules which is store in the knowledge base. When it finds the perfect match then it takes decision as describe in the system. For this process of rule matching temporary working memory is used. At last the decision made by the system is transfer to the user or client using user interface.

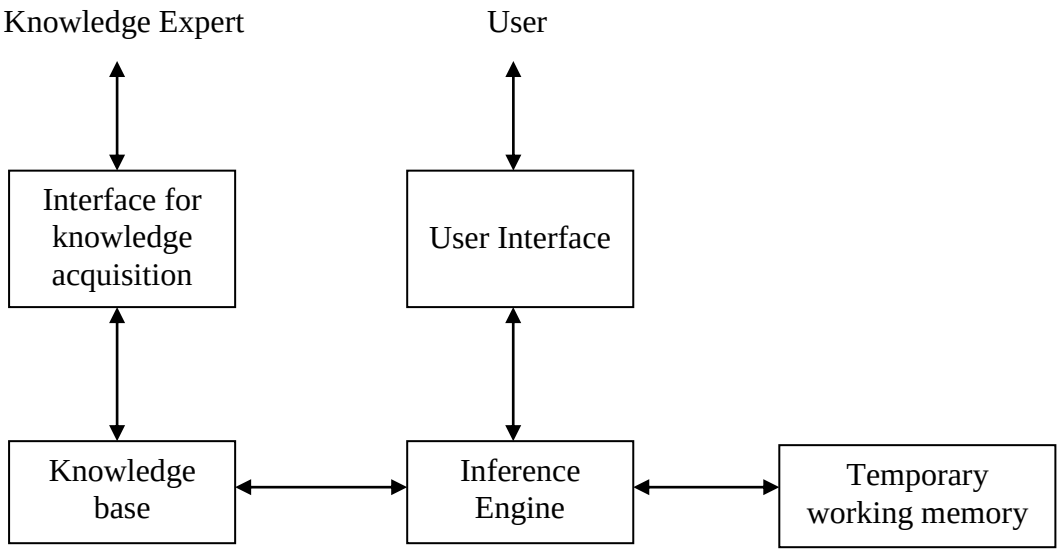

Figure 1: Components of Rule-Based Expert System[8]

III. DIFFERENT CLASSIFICATION TECHNIQUES

Decision tree algorithm: Decision tree [3] is a structure which is look like a flowchart structure, in this algorithm we perform specific test on an attribute which represented as a internal node of tree and leaf node represents the class label. Tree contains only one root node which indicates the starting of tree. This algorithm is simple and fast. Rules are easily converted to this algorithm. The accuracy of the decision tree algorithm is good. C4.5 is the most efficient and useful algorithm used for decision tree-based approach. In this algorithm initially dataset is sorted according to attribute value. This procedure is continuing until all the attributes are classified. When node of a tree contains same attribute records then that node is consider as a leaf node as represent the class label. Decision tree is performing well when size of a dataset is large.

This algorithm is deal with the different types of attributes 1) Discrete valued: in this type, possible value of node N is known and the procedure of splitting is performing based on these values. 2) Continuous valued: in this splitting condition is define and based on this condition node N have only two possible outputs.3) Discrete valued but a binary tree must be produced: for this we check where the value of attribute is match with the condition or not. It generates only two branches yes or no.

The pseudo code of this algorithm is as below:
Input: Data partition D, attribute list, Attribute selection method.
Output: a decision tree
Method:
1) create a node N;
2) if tuples in D are all of the same class C then
3) return N as a leaf node labeled with the class C;
4) if attribute_list is empty then
5) return N as a leaf node labeled with the majority class in D; //majority voting
6) apply Attributr_selection_method(D,attributr_list) to find the best splitting_criterion;
7) label node N with splitting_criterion;
8) if splitting_attribute is discrete value and multiple splits allowed then
9) attribute_list ← attribute_list-splitting_attribute;
10) for each outcome j of splitting_criterion
11) let Dj be the set of data tuples in D satisfying outcome j;
12) if Dj is empty then
13) attach a leaf labeled with the majority class in D to node N;
14) else attach the node return by generate_decision_tree(Dj, attribute_list) to node N;
15) endfor
16) return N;

Naïve- Bays algorithm: This algorithm is created based on Bay's theorem. It is based on the posterior probability. It is a statistical classifier. They can predict class membership probabilities such as the probability that a given tuple belongs to a particular class. Naïve Bayesian classifiers assume that the effect of an attribute value on a given class is independent of the values of the other attributes. This algorithm is perform accurately when the size of the dataset is medium or large. The results of both algorithm decision tree and naïve bays are very comparative in different domains.

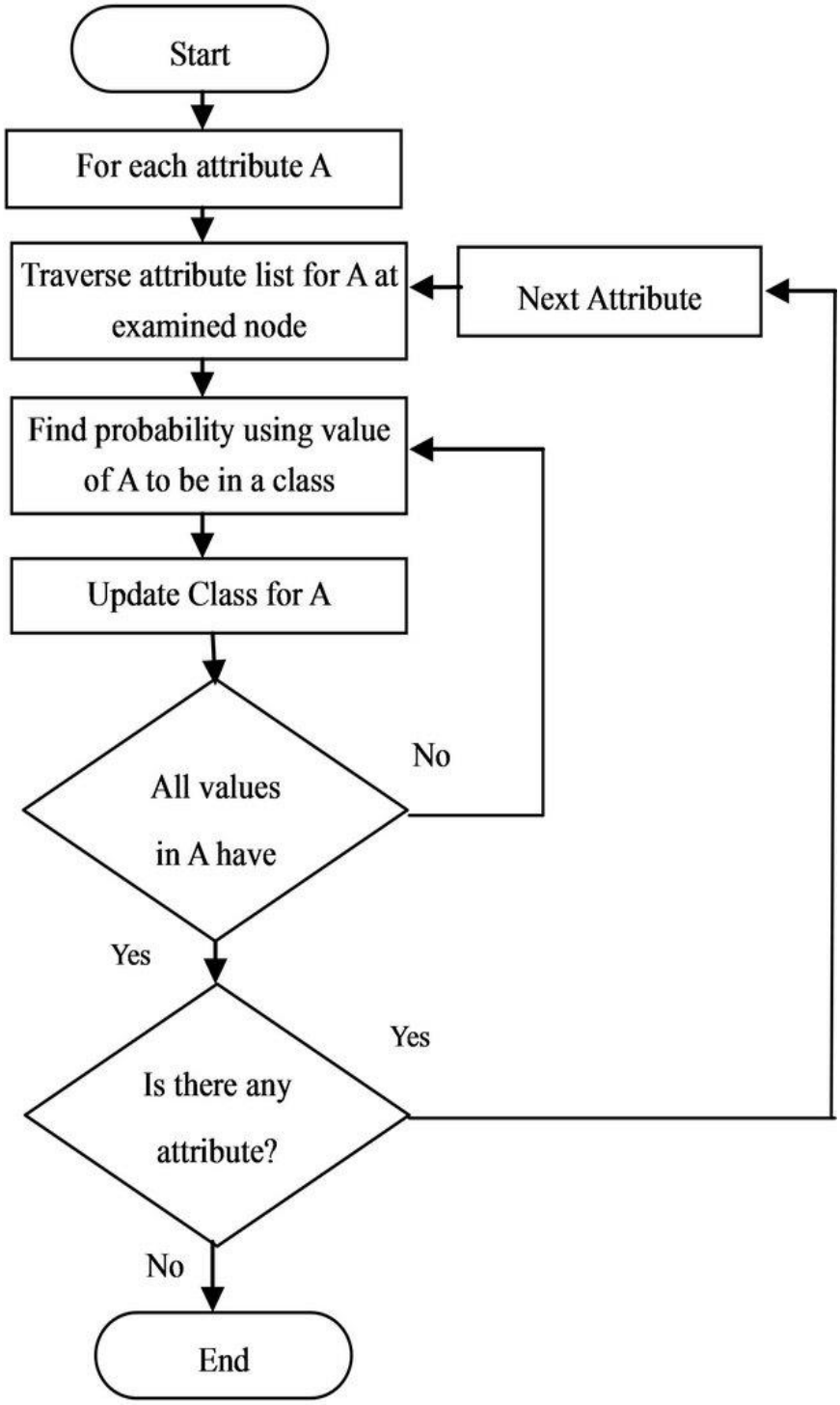

Figure 2: Flowchart of Naive Bayes algorithm[3]

IV. COMPARISON OF CLASSIFICATION TECHNIQUES

In this section, we use Cleveland heart disease dataset for comparing the performance of different classification technique in weka. For performing this comparison we divide dataset into two subsets.

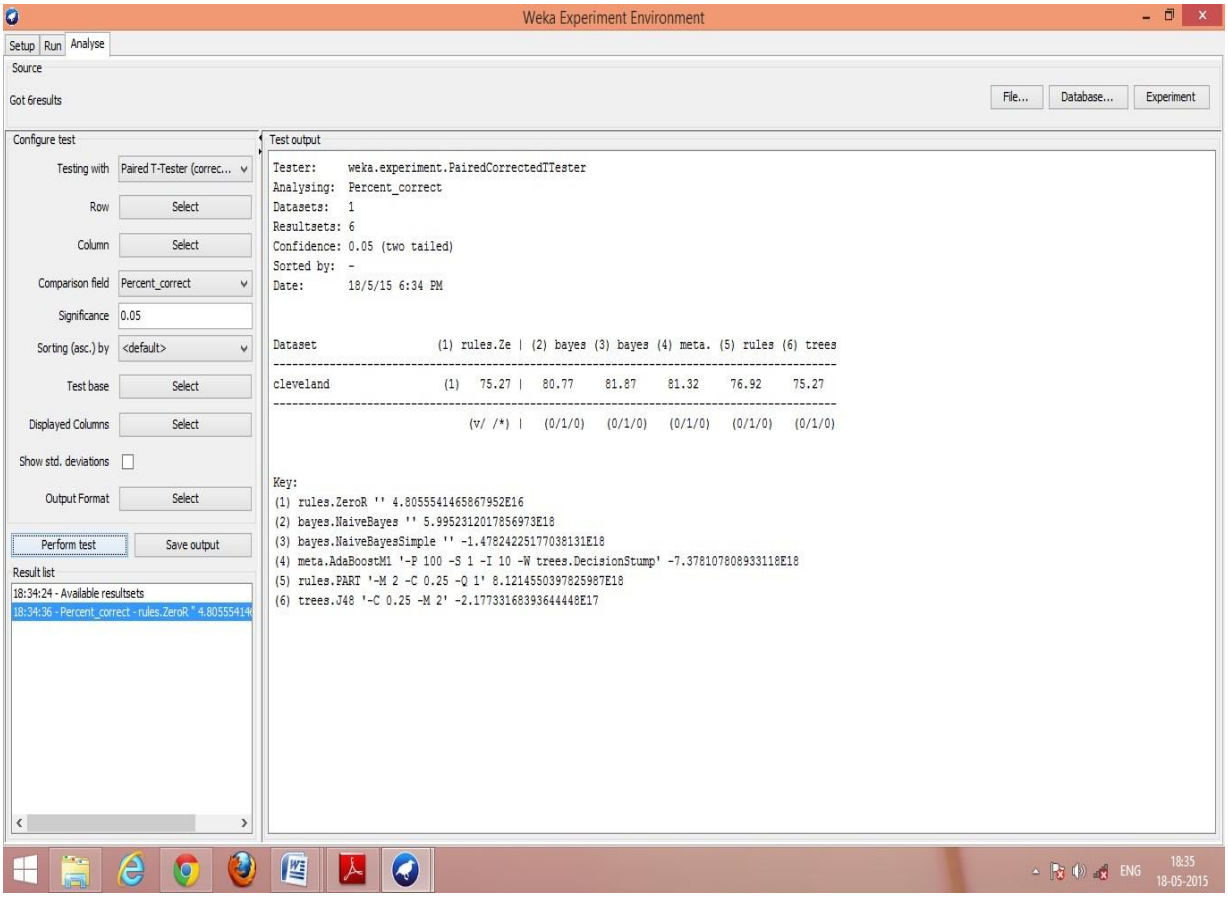

Figure 3: Comparison of classification techniques in WEKA

V. PROPOSED WORK

In this work, main aim is improve the accuracy of the medical prediction system. Till now in expert system knowledge base is created from the knowledge of human expert, which is subjective to the expert. So, in this work we find out the useful knowledge from the real dataset related to the particular domain.

Phase 1: In this phase we use whole dataset to generate the decision tree. After generating the decision tree we choose only those attributes which is used to split the nodes or internal nodes of the decision tree.

Phase 2: In this phase we use naive bays classifier to train the model. By using naive bays and internal nodes we create inference engine and then test the data using trained model.

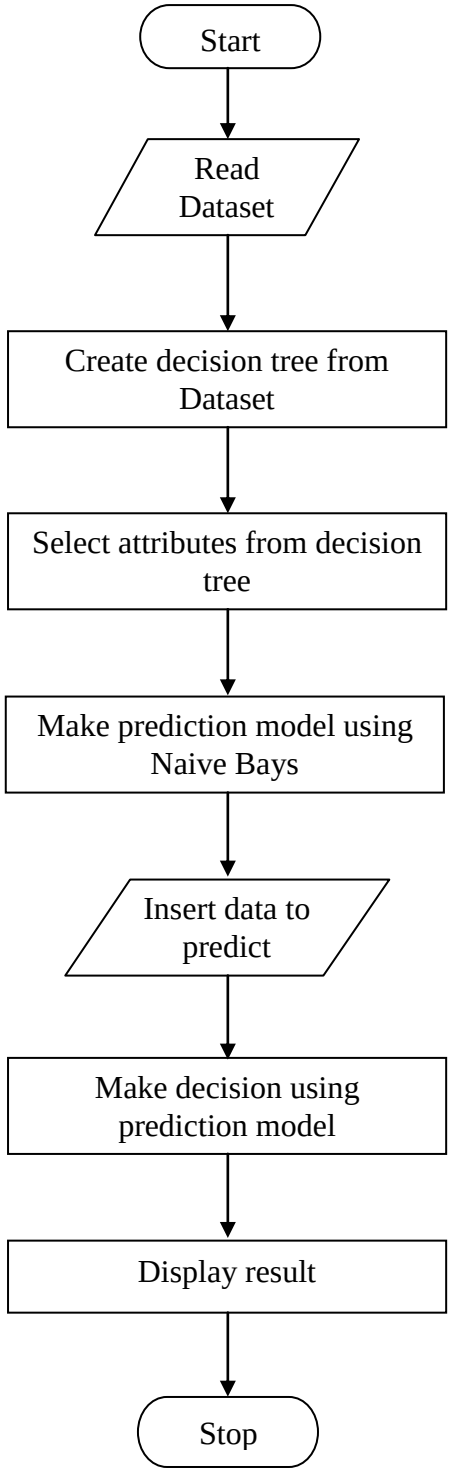

Figure 4: Flow of Proposed work

VI. EXPERIMENTAL WORK

After performing the steps included in proposed work we find out the following results and compare that result to the fuzzy heart prediction system.

Dataset	Weighted fuzzy rule Based system		Proposed system	
	Training	Testing	Training	Testing
Cleveland	0.509901	0.623529	0.89109	0.82178

TABLE 1: COMPARISON OF EXPERT SYSTEM

VII. CONCLUSION

In this paper our aim is to develop clinical prediction expert system. For that we initial use examples or raw data of heart patients and extract attributes from that data using decision tree classifier. In next step we generate knowledge base from extracted rules and pass it to inference engine. For working of inference engine we train our model using naïve bays classifier. Then prediction model predict class label for user input.

ACKNOWLEDGEMENT

The authors would like to thank the reviewers for their precious comments and suggestions that contributed to the expansion of this work.

REFERENCES

1. P.K. Anooj "Clinical decision support system: Risk level prediction of heart disease using weighted fuzzy rules" of King Saud University – Computer and Information Sciences (2012) production and hosting by Elsevier.Journal
2. Gökhan Engin, Burak Aksoyer, Melike Avdagic, Damla Bozanlı, Umutcan Hanay, Deniz Maden, Gurdal Ertek "Rule-based expert systems for supporting university students" ITQM 2014 ScienceDirect
3. Masud karim, Rashedur M. Rahman "Decision tree and Naïve Byes algorithm for classification and generation of actionable knowledge for Direct Marketing." Journal of software engineering and applications,2013,6,196-206
4. Zhi-Hua Zhou, Yuan Jiang "Medical Diagnosis with C4.5 rule preceded by artificial neural network ensemble" IEEE transactions on information technology in Biomedicine
5. Dahee chung, Kun change Lee, Seung chang Seong "General Bayesian network approach to Health informatics prediction: emphasis on performance comparison" Procedia- Social and Behavioral Science 81,2013, Science Direct
6. A.Ambika, Satyanarayana Gandi, Amarendra Kothalanka "An Efficient Expert system for Diabetes By Naïve Bayesian classifier " IJETT- Volume 4 Issue 10-Oct 2013
7. M.S Prasad Babu, Venkatesh Achanta, N.V. Ramana Murty, Swapna K. "Development of Maize Expert System using Ada-Boost algorithm and Naïve Bayesian Classifier" IJCATR Volume 1-Issue 3, 89-93, 2012
8. Maitri patel, Atul patel, Paresh virparia "Rule based Expert System for Viral Infection Diagnosis" IJARCSSE Volume 3, Issue 5, May 2013

9. V. Krishnaiah, Dr.G.Narsimha,Dr.N.Subhash Chandra "Diagnosis of Lung Cancer Prediction System using Data Mining Classification Techniques" IJCSIT Volume 4(1),2013,39-45
10. J. Huang, J. Lu and C. X. Ling, "Comparing Naïve Bayes, Decision Trees, and SVM with AUC and Accuracy," *Proceedings of Third IEEE International Conference on Data Mining* (*ICDM* 2003), 19-22 November 2003, pp. 553-556
11. G. Dimitoglou, J. A. Adams and C. M. Jim, "Comparison of the C4.5 and a Naïve Bayes Classifier for the Predic- tion of Lung Cancer Survivability," *Journal of Comput-ing*, Vol. 4, No. 2, 2012, pp. 1-9.
12. P. S. Vadivu and V. K. David, "Enhancing and Deriving Actionable Knowledge from Decision Trees," *Interna- tional Journal of Computer Science and Information Se-curity* (*IJCSIS*), Vol. 8, No. 9, 2010, pp. 230-236.
13. Prof. M.S. Prasad Babu, N.V. Ramana Murty, S.V.N.L.Narayana, "A Web Based Tomato Crop Expert Information System Based on Artificial Intelligence and Machine learning algorithms", *International Journal of Computer Science and Information Technologies, Vol. 1 (1), (ISSN: 0975-9646).,* 2010, pp6-15.
14. Dr. B D C N Prasad, P E S N Krishna Prasad and Y Sagar, "A Comparative Study of Machine Learning Algorithms as Expert Systems in Medical Diagnosis (Asthma)" – *CCSIT 2011, Part I, CCIS 131,* pp 570– 576.
15. Tulips Angel Thankachan, Dr. Kumudha Raimond, " A Survey on Classification and Rule Extraction Techniques for Datamining," IOSR Journal of Computer Engineering (IOSR-JCE),Volume 8, Issue 5 (Jan. - Feb. 2013), PP 75-78.
16. Setiawan, N.A., Venkatachalam, P.A., Hani, A.F.M., 2009. Diagnosis of coronary artery disease using artificial intelligence based decision support system. In: Proceedings of the International Conference on Man-Machine Systems, Batu Ferringhi, Penang, 11–13 October,1, 2009.

