International Journal for Rapid Research in Engineering Technology & Applied Science
ISSN (Online): 2455-4723
Vol 3 Issue 5 May 2017

Improve the Accuracy of ID3 Classifier for Heart Disease Prediction

Ms. Shikha Sharma* Mr. Chetan Chauhan**
M. Tech Student, *Shri Vaishnav institute Of Technology &Science*
Assistant Professor Department Of Computer Science Engg, SVITS, Indore
*S*harmashikha198@gmail.com*, chetan.chauhan1982@gmail.com**

Abstract:-

Data Classification may be a very fashionable and computationally overpriced task. Most of those information classification techniques square measure supported the conception of call trees. several researchers have worked on the malady prediction systems exploitation the information mining techniques. a number of the systems square measure for predicting one malady and a few for the predicting the multiple diseases. Still there's scope to enhance the potency of the malady prediction. during this paper, we tend to square measure presenting associate degree updated ID3 rule. a brand new attribute choice rule has projected during this paper. The accuracy of the projected methodology is best than the present rule.

Key Words -: data processing, classification rule, ID3 rule, call tree

INTRODUCTION:-

Data mining is that the non-trivial method of distinguishing valid, novel, doubtless helpful and ultimately intelligible pattern in knowledge. several governmental organization, businesses etc square measure finding some way to gather, store, analyze and report knowledge regarding people ,households or businesses, so as to support (short and long term) coming up with activities. System contains non-public or lead like their Social Security variety, financial gain of staff, getting of client etc.

Decision tree learning, utilized in statistics, data processing and machine learning, uses a call tree as a prophetical model that maps observations regarding associate degree item to conclusions regarding the item's target price. A lot of descriptive names for such tree models square measure classification trees or regression trees. In these tree structures, leaves represent category labels and branches represent conjunctions of options that cause those category labels. In call analysis, a choice tree may be accustomed visually and expressly represent choices and higher cognitive process. In data processing, a choice tree describes information however not decisions; rather the ensuing classification tree may be associate input for higher cognitive process. This

page deals with call trees in data processing. Call tree learning could be a methodology usually employed in data processing. The goal is to make a model that predicts the worth of a target variable supported many input variables. Every interior node corresponds to at least one of the input variables.

There are edges to kids for every of the doable values of that input variable. every leaf represents a worth of the target variable given the values of the input variables painted by the trail from the foundation to the leaf. A tree may be "learned" by cacophonous the supply set into subsets supported Associate in Nursing attribute worth take a look at. This method is continual on every derived set in a very algorithmic manner known as algorithmic partitioning. The rule is completed once the set at a node has all constant worth of the target variable, or once cacophonous now not adds worth to the predictions. This method of top-down induction of call trees (TDIDT) is Associate in Nursing example of a greedy formula, and it's out and away the foremost common strategy for learning call trees from information, however it's not the sole strategy. In fact, some approaches are developed recently permitting tree induction to be performed in a very bottom-up fashion.[2]

Literature Survey:-

Early methods of identifying patterns in data include Bays' theorem (1700s) and regression analysis (1800s). The proliferation, ubiquity and increasing power of computer technology has increased data collection and storage. As data sets have grown in size and complexity, direct hands-on data analysis has increasingly

been augmented with indirect, automatic data processing. This has been aided by other discoveries in computer science, such as neural networks, clustering [1], genetic algorithms (1950s), decision trees (1960s) and support vector machines (1980s).

In data processing and machine learning, a choice tree could be a prophetical model; that is, a mapping from observations concerning associate item to conclusions concerning its target worth. a lot of descriptive names for such tree models ar classification tree (discrete outcome) or regression tree (continuous outcome). In these tree structures, leaves represent classifications and branches represent conjunctions of options that lead those classifications. The machine learning technique for inducement call a tree from knowledge is termed decision tree learning.

Proposed Approach:-

In call tree classifiers, the factors used for the attribute choice is as follows: initial info gain of every attribute is computed then the attribute having most info gain is chosen.

This means that AN attribute with most values is chosen for cacophonic the tree. However in most of the cases, it's not necessary that AN attribute with most values are the simplest. Additionally ID3 algorithmic rule uses the idea of data gain for choosing an attribute. The data gain relies on the idea

of the chance. Chance primarily based technique is appropriate for random issues. However it can't be the common criteria for attribute choice.

For resolution this drawback, we tend to propose a additional correct call tree primarily based classifier. Our proposed solution will use new attribute selection criteria. It will give more weight to attributes with less value but more importance. Also it will reduce the weight of attribute with more values and less importance

ATTRIBUTE SELECTION: Our projected methodology uses a changed gain primarily based greedy approach to select the best attribute, which will be used for partitioning the training data set into smaller partitions. Similar to ID3, our proposed algorithm also chooses the attribute with highest information gain. But we have modified the formulae of information gain. The modified formulae contain utility value of each attribute. In this the choice criteria has improved, that ultimately can result's a lot of classification and prediction..Entropy measures the amount of information in an attribute [5].

ID3 (Examples, Target Attribute, Attributes)

Create a root node for the tree

If all examples square measure positive, come back the single-node tree Root, with label = +. If all examples square measure negative, come back the single-node tree Root, with label = -.

if vary of predicting attributes is empty, then come the only real node tree Root, with label = commonest value of the target attribute inside the examples.

Otherwise Begin
A ← The Attribute that best classifies examples.

Decision Tree attribute for Root = A. for every doable worth of A,
Add a replacement limb below Root, equivalent to the check A = .

Let Examples ( ) be the set of examples that have the worth for A

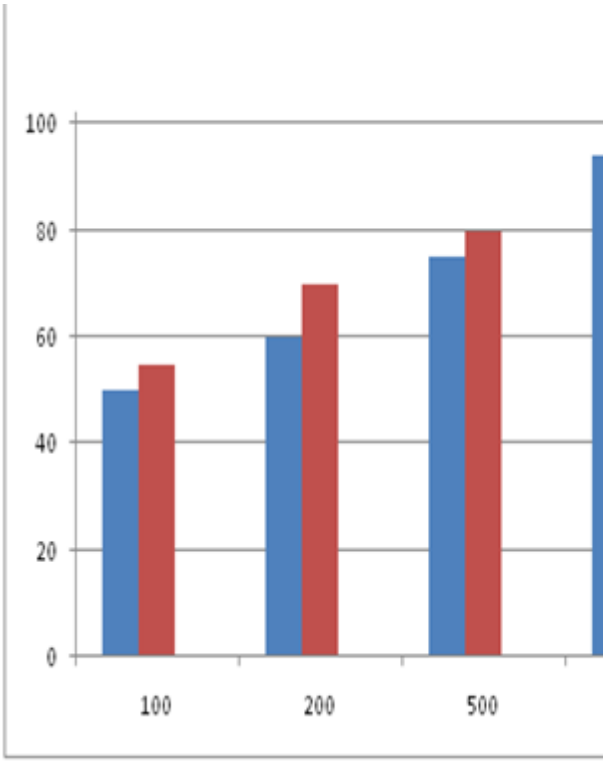
If Examples ( ) is empty
Then below this new branch add a leaf node with label = commonest target worth within the examples

Else below this new branch add the sub tree ID3 (Examples ( ),Target

- Attribute, Attribute–)
End
Return Root

Result Analysis:-

	Accuracy	**Error**
Previous Algorithm	**94.987%**	.1152
Proposed Algorithm	**99.99%**	.048



Conclusion:-

In this paper, we have presented a more accurate algorithm for classification.
Our proposed methodology uses greedy approach to select the best attribute.
To do so the information gain is used. The attribute with highest information
gain is selected. In this way accuracy has improved. We also surveyed the
existing data classification techniques. We restricted ourselves to the classic classification problem.

Future Work:-

Our research work has shown that the decision tree based data mining techniques can be applied to predict multiple diseases. The outcome of this work is very important for the medical field. It may help the patients and doctors up to a great extent. To improve the accuracy of the classifier further research should be conducted using different data mining classification algorithms like- rule based induction, expectation maximization, etc.

References:-

[1] Singh Vijendra. Efficient Clustering For High Dimensional Data: Subspace Based Clustering and Density Based Clustering, *Information Technology Journal; 2011*, 10(6), pp. 1092-1105.

[2] D Bremen, L., Friedman, J. H., Olsen, R. A., and Stone, C. J. "Classification and Regression Trees". Wadsworth International Group. Belmont, CA: The Wadsworth Statistics/Probability Series1984.

[3] Quinlan, J. R. "Induction of Decision Trees". *Machine Learning;* 1986, pp. 81- 106.

[4] Quinlan, J. R. Simplifying "Decision Trees. International Journal of Man-Machine Studies";

1987, 27:pp. 221-234.

[5] Gama, J. and Brazil, P. "*Linear Tree. Intelligent Data*Analysis",1999,.3(1): pp. 1-22.

[6] Langley, P. "*Induction of Recursive Bayesian Classifiers*". In Brazil P.B.(ed.), Machine Learning: ECML-93;1993,pp.153-164Springer,Berlin/Heidelberg~lew York/Tokyo.

[7] Witten, I. & Frank, E,"*Data Mining: Practical machine learning tools and* techniques", 2nd Edition, Morgan Kaufmann, San Francisco, 2005.ch. 3,4, pp 45-100.

[8] Yang., Webb, G. "*On Why Discretization Works for Naive-Bays* Classifiers*"*, Lecture Notes in Computer Science, vol. 2003, pp. 440 –452.

[9] H. Zantema and H. L. Oleander, "Finding Small Equivalent Decision Trees is Hard", *International Journal of Foundations of Computer Science*; 2000, 11(2):343-354.

[10]Huang Ming, Niue Waning and Liang Xu , "An improved Decision Tree classification algorithm based on ID3 and the application in score analysis", *Software Technol. Inst*., Dalian Jiao Tong Univ., Dalian, China, June 2009.