Decision Tree Data Mining Algorithms for Diagnosis of Angioplasty and Stents for Heart Disease Treatment

Amarjeet Kaur [1] (PHD Scholar), Dr.Ashok Jetwat[2] (Professor of Pacific University),
Dr.Gurpreet Singh[3] (Professor of St.Soldier Institute of Engg &Tech),
Department Of Computer Science & Engg)

Abstract:

Remote resources such as computers, databases, files etc along with people like analysts, professionals, end users are often involved in the complex process of analysis of data. This analysis is in an omnipresent manner and is very important for applications which deal in finance, process control, defense and many more domains. The ability to analyse large data amount is the demand of these applications. Decision tree a data mining technique which are CART, ID3 and C4.5 as are scalable and fast and are for data streams monitoring from omnipresent devices such as computers, palmtops etc.

Keyword: CART, ID3, C4.5

I. INTRODUCTION:

There are varieties of algorithms being used in classification technique. One if these are the decision tree approach. To represent both the regression models and classifiers decision tree in the state of predicative model is used. Decision trees basically us the hierarchal model of decisions and their consequences. The structure of decision tree includes branch, root node and leaf node. Attributes test is denoted on each interval node, the test outcome is denoted by branch and class labels are shown by leaf node. The topmost node is the root node of the tree. The tree learning is done by dividing the source into set which are generally based on a test of attribute value. The top down approach of decision tree sets an example of greedy algorithm. Apart from this bottom-up approach is also common these days.

There are mainly two types of data trees used in data mining.

1. Classification tree analysis-It is done when the class to which data depends in the predicted outcome.

2. Regression tree analysis-It is done when are all number can be taken as the predicted outcome

example (The cost of a building) To refer both of these procedures the term classification and regression tree CART analysis is used. Trees used for both regression and classification are same at some perspective but along with this they have differences too such as procedures which are used to determine the split point. There are techniques which construct more than one decision tree namely Bagging Decision Trees, Random Forest Classifier, Boosted Trees and Rotation Forest.

What are Heart Disease Treatment with Angioplasty and Stents:

First, you'll have what's called a cardiac catheterization. Medication will be given to relax you, then the doctor will numb where the catheter will go with anesthesia. Next, a thin plastic Tube called a sheath is inserted into an artery—some times in your groin, sometimes in your arm. Along, narrow, hollow tube called a catheter is passed through the sheath and guided up a blood vessel to the arteries surrounding the heart. A small amount of contrast liquid is put into your blood vessel through the catheter. It's photographed with an X-ray as it moves through your heart's chambers, valves, and major vessels. From these pictures, doctors can tell if your coron aryarteries are narrowed and, in some cases, whether

the heart valves are working correctly. If the doctor decide stopper form angioplasty, he will move the catheter into the artery that's blocked. He'll then do one of the procedures described below. The whole thin glasts from 1to3 hours, but the preparation and recovery can add much more time. You may stay in the hospital overnight for observation.

What Types of Procedures Are Used in Angioplasty?

There are several your doctor will choose from. They include:

Balloon:

A catheter with a small balloon tip is guided to then arrowing in your artery. Once in place, the balloon is in flated top used the plaque and stretch the artery open to boost blood flow to the heart.

Stent:

This is a small tube that acts as a scaffold to support the inside your coronary artery. A balloon catheter, placed over a guide wire, puts the stents into your narrowed coronary artery. Once the balloon is flated and the stent expands the size of the artery and holds it open.

(CHAID). Performs multi-level splits when computing classification trees.

- MARS: extends decision trees to better handle numerical data.

ID3 and CART are invented in dependently of one another at around same time, yet follow a similar approach for learning decision tree from training tuples.

The balloon is then deflated and removed while the stent stays in place. Over several weeks, your artery heals around the stent. These are ten placed during angioplasty to help keep the coronary artery open. The stent is usually made of metal and is permanent. It can also be made of a material that the body absorbs overtime. I will collect in this research how many people's are suffering with this disease. **CHAID** stands for *Chi*-squared Automatic Interaction Detector. The CHAID is a kind of analysis that find show variables are best combined to explain the effect of a given dependent variable. The model can be used in situation of market dispersion, predicting and interpreting responses or a multitude of other research problems.

Table.1.Dataset used

age	chest_pai	rest_bpres	blood_sug	rest_electro	max_heart_ra	exercice_angi	disease
43	asympt	140	F	normal	135	yes	positive
39	atyp_angin	120	F	normal	160	yes	negative
39	non_angin	160	T	normal	160	no	negative
42	non_angin	160	F	normal	146	no	negative
49	asympt	140	f	normal	130	no	negative
50	asympt	140	f	normal	135	no	negative
59	asympt	140	t	left_vent_hyp	119	yes	positive

There are many specific decision-tree algorithms. Not able ones include:

- ID3(IterativeDichotomiser3)
- C4.5algorithm,successorofID3
- CART(Classification And Regression Tree)
- CHi-squared Automatic Interaction Detector

CHAID analysis is mainly useful for data expressing categorized values instead of continuous values. For this kind of data some common statistical tools such as regression are not applicable and CHAID analysis is a perfect tool to discover the relationship between variables. One of the outstanding advantages of

CHAID analysis is that it can visualize the relationship between the target (dependent) variable and the related factors with a tree image.
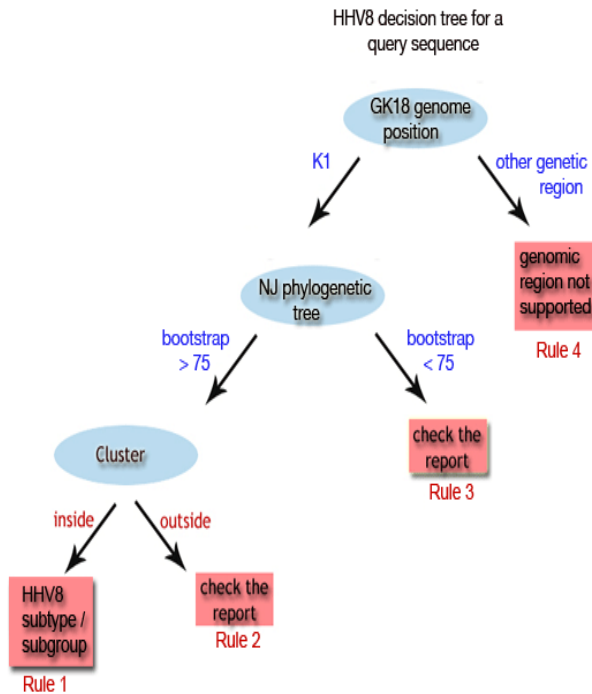
Figure.1 .Training Tuples

ID3 and C4.5 are developed by Quinlan for inducing *Classification Models*, also called *Decision Trees*, from data. We are given a set of accounts. Each record has the same construction, consisting of a number of quality/value pairs. One of these attributes represents the group of the record. The problem is to decide a decision tree that on the basis of answers to questions about the non-category attributes predicts correctly the value of the category attribute. Usually the category attribute takes only the values {true, false}, or {success, failure}, or something equivalent. In any case, one of its values will mean failure.

The basic ideas behind ID3 are that:

•   In the decision tree each node corresponds to an on-categorical attribute and each arc to a possible value of that attribute. A leaf of the tree specifies the expected value of the definite attribute for the records described by the path from the origin to that leaf. [This defines what a Decision Tree is.]

•   In the decision tree at every node must be related the non-categorical attribute which is most useful among the attributes not so far measured In the path from the root. [This establishes what a

"Good" decision tree is.]

•   Entropy is used to measure how in formative is a node. [This defines what we mean by "Good". By the way, this notion was introduced by Claude Shannon in Information Theory.]
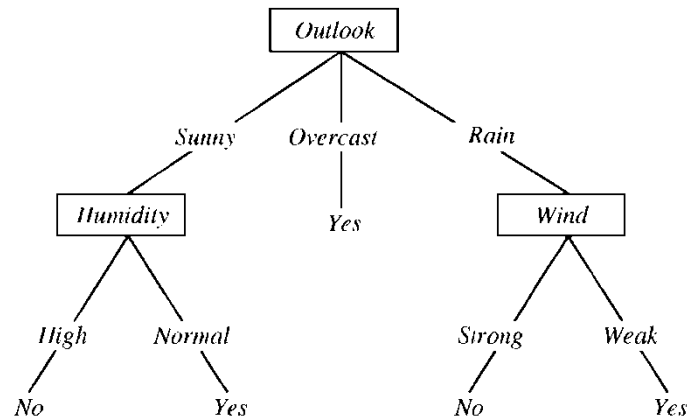

Figure.2.variable and the related factors with a tree image

II. CONCLUSION:

In this work, I propose a technology based on data mining algorithms for the induction of decision trees. It is well suited in our context for various reasons. To collect the data set from different hospitals of Angioplasty and Stents for Heart Disease Treatment and propose enhanced decision tree algorithm which will work on Angioplasty and Stents for Heart Disease Treatment dataset. Increase the efficiency of correct classified instances with a new classifier that combines the k- Nearest Neighbor (CART) distance based algorithm with the classification tree paradigm based on C45 algorithm and improve accuracy or reduces the error to the same dimensions as the quantity being predicted by using sum of square error as compare to the CART and C4.5 classification algorithm with new algorithm.

III.REFERENCES:

[1]. Yang, Q., Yin, J., Ling, C., & Pan, R. (2007). Extracting actionable knowledge from decision trees. Knowledge and Data Engineering, IEEE Transactions on,19(1), 43-56.

[2]. Ye, M., Wu, X., Hu, X., & Hu, D. (2013). Multi-level rough set reduction for decision rule mining. Applied intelligence, 39(3), 642-658.

[3]. Katz, G., Shabtai, A., Rokach, L., & Ofek, N. (2014). ConfDTree: A Statistical Method for Improving Decision Trees. Journal of Computer Science and Technology, 29(3), 392-407.

[4]. Soldic-Aleksic, J. (2012). COMBINED APPROACH OF KOHONEN SOM AND CHAID DECISION TREE MODEL TO CLUSTERING PROBLEM: A MARKET SEGMENT ATION EXAMPLE. Journal of Economics & Engineering, 3(1).

[5].Dan, J., Jianlin, Q., Xiang, G., Li, C., & Peng, H. (2010, June). A Synthesized Data Mining Algorithm Based on Clustering and Decision Tree. In Computer and Information Technology(CIT), 2010 IEEE 10th International Conference on(pp. 2722-2728). IEEE.

[6]. Karaolis, M. A., Moutiris, J. A., Hadjipanayi, D., & Pattichis, C. S. (2010). Assessment ofthe risk factors of coronary heart events based on data mining with decision trees. Information Technology in Biomedicine, IEEE Transactions on, 14(3), 559-566.

[7]. Kargupta, Hillol, and Byung-Hoon Park. "Mining decision trees from data streams in a mobile environment." Data Mining, 2001. ICDM 2001, Proceedings IEEE International Conference on. IEEE, 2001.

[8]. Van Hieu, Duong, Nawaporn Wisitpongphan, and Phayung Meesad. "Analysis of factors which impact Facebook users' attitudes and behaviours using decision tree techniques." Computer Science and Software Engineering (JCSSE), 2014 11th International Joint Conference on. IEEE, 2014.

Decision Tree Data Mining Algorithms for Diagnosis of Angioplasty and Stents for Heart Disease Treatment

Amarjeet Kaur [1] (PHD Scholar), Dr.Ashok Jetwat[2] (Professor of Pacific University),
Dr.Gurpreet Singh[3] (Professor of St.Soldier Institute of Engg &Tech),
Department Of Computer Science & Engg)