

A Review on Classification of Various Types of Decision Trees with Merits and Demerits

O.Yamini¹, Dr G.V.Ramesh Babu²

¹ Research Scholar, Department of Computer Science, S V University, Tirupati

² Associate Professor, Department of Computer Science, S V University, Tirupati

Abstract

Data mining includes several important technologies such as classification, clustering, regression, etc. Categorical mining technology, among data mining technologies, is becoming the most mature and active research direction enabling successful applications. Categorical mining can be applied to uncover useful information from large amounts of data stored in a large number of fields such as hospitals, inventory, banking, etc. Decision tree methods, neural network methods and statistics exploratory methods. In this paper a large number of decision tree algorithms such as Iterative Dichotomizer 3, CART (Classification and Regression Tree), C4.5, SLIQ and SPRINT are used in different fields. These algorithms have been explained clearly with merits and demerits along with some applications. Most of the decision tree methods are developed from the ID3 method.

Keywords: Data mining, Classification algorithms, ID3, CART, C4.5, SLIQ, SPRINT.

1. Introduction

Data mining is a step within the information discovery method in a database that consists of the application of information discovery and analysis algorithms that, within perfect limits of computational performance, generate a table that lists specific patterns on the records [1].

Data Mining is a sequence of tactics to discover the delivered price of a facts set of know-how that has now not been known manually. Remember that the word mining itself way the attempt to get a few valuables from a massive variety of basic substances. Consequently statistics Mining truly has long roots from fields together with artificial intelligent, gadget gaining knowledge of, statistics and databases. Data mining is the process of applying this approach to information which will uncovering hidden patterns. With other meanings information mining is the process for extracting patterns from the statistics. Statistics mining will become an increasingly critical device for converting statistics into data. It is often utilized in diverse profile practices, such as advertising, surveillance, fraud detection and

scientific discovery. It has been used for years by means of businesses, scientists and governments to filter the extent of statistics including flight Passenger travel records, census data and supermarket information scanners to generate market research reports. The main cause for the usage of information mining is to assist inside the analysis of behavioral statement collections. The facts is vulnerable to co linearity due to known association together with described into figure 1.

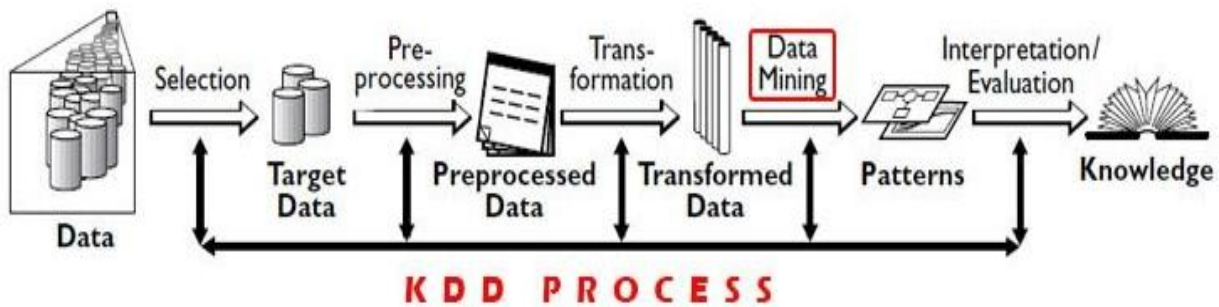


Fig1. Process Stages of KDD (knowledge discovery in databases)

2. Data Mining Functions

Data mining features or obligations may be used to specify the types of styles or information to be discovered for the duration of records mining. Some of the main records mining features are classification, clustering, category, outlier analysis, regression and prediction, and so forth. [1]

a) Classification

A classification algorithm is used to predict the data classes [6]. To date, a large collection of classification algorithms (or classifiers) have been proposed by researchers [2, 7]. Some popular classification algorithms are summarized.

Category	Name of algorithm	Based on the concept
Decision tree classifiers (machine learning based)	ID3 (iterative dichotomiser) CART (classification and regression trees) C4.5 (a descendant of ID3)	Information gain Gini index Gain ratio
Bayesian classifiers (statistics based)	Naive Bayesian (or Simple Bayesian) classifier Bayesian belief networks	Bayes' theorem Bayes' theorem and probabilistic graphical model
Rule based classifiers (machine learning based)	IF-THEN rule using decision tree Sequential covering algorithms: AQ, CN2	Decision tree Entropy, information gain

	and RIPPER	
Support vector machine classifier Classification using Backpropagation (neural network based)	Support vector machine Backpropagation	Linear optimal separating hyperplane Multilayer FF ANN
Classification using frequent patterns (associative classification based)	CBA (classification based on association) Frequent Itemset Mining CMAR (classification based on predictive association rules)	Frequent itemset mining with rule pruning strategy. Frequent itemset mining with foil (first order inductive learner)
Lazy learners (machine learning based)	kNN (K-nearest neighbour) CBR (case-based reasoning)	Learning by analogy and euclidean distance Database of problem solutions and background knowledge

b) Outlier analysis

The outlier’s location unit typically discarded by way of maximum of understanding mining strategies as noise or exceptions. Every so often, outliers would possibly have additional facts compared to opportunity data objects. So outlier analysis is very crucial for some application regions like intrusion detection, fraud detection, anomaly detection, etc.

Numerous records processing techniques generally use bunch to hit upon the outliers as a noise. The outlier detection techniques is assessed as classification-based strategies, applied arithmetic techniques, clustering-primarily based techniques, supervised, semi-supervised and unsupervised strategies, deviation-primarily based strategies and proximity-primarily based strategies [2]

c) Regression

Regression predicts the really worth of characteristic supported regression technique(s) over time. The long run values of variables area unit foretold with the help of historical statistic plot. Analytic wondering (also called as evolution analysis) discovers captivating patterns in the evolution history of the items. Identification of patterns in accomplice degree item’s evolution and matching of the items’ dynamical tendencies place unit the 2 predominant aspects of analytic wondering. A trend of the items, whose behavior evolves over the years, is delineated victimization analytic wondering and regression

fashions. Analytic wondering exposes time-varying developments of the statistics items at intervals the dataset. The association analysis can even be used for evolution evaluation [5].

d) Prediction

Regression evaluation is accustomed model the hyperlink among one or extra freelance or predictor variables and a mounted or reaction variable (that's non-stop-valued). Inside the context of information mining, the predictor variables region unit the attributes of interest describing the tuple (i.E., developing up the function vector). In trendy, the values of the predictor variables location unit brilliant. (techniques exist for managing instances anywhere such values can also be missing.) The reaction variable is what we would really like to expect—it's far what we have a propensity to spoken in segment half-dozen.1 because the foretold attribute. Given a tuple delineated with the aid of manner of predictor variables, we would really like to expect the associated worth of the response variable.

3. Classification by Decision Trees

Decision trees are the most typically used due to its ease of implementation and simplicity in expertise compared to different class algorithms. Decision Tree type algorithm may be applied in a serial or parallel fashion based totally on the quantity of statistics, reminiscence area to be had at the pc resource and scalability of the algorithm. [8]

Selection tree set of rules is a information mining induction strategies that recursively walls a facts set of data using intensity first grasping method or breadth-first approach till all the records gadgets belong to a particular class. [9] A decision tree shape is manufactured from root, internal and leaf nodes. The tree shape is utilized in classifying unknown records statistics. At every inner node of the tree, a choice of great break up is made using impurity measures. The tree leaves is made up of the magnificence labels in which the statistics items had been grouped. Choice tree may be a non-parametric supervised studying formula, that is used for each type and regression responsibilities. It's a gradable, tree shape that consists of a root node, branches, internal nodes and leaf nodes. As we are now capable of see from the subsequent diagram.

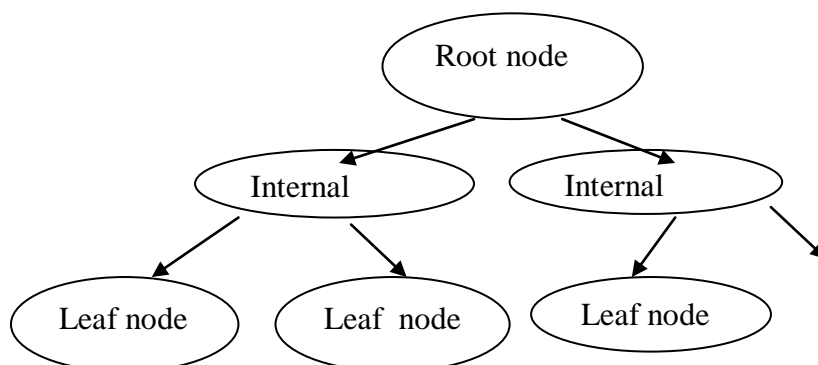


Fig 2: Decision Tree

Types of decision trees

Hunt's set of rules, developed within the Sixties to version human studying in psychology, is the basis of many famous choice tree algorithms, which include the following:

ID3: Ross Quinlan is credited with growing ID3, which stands for "Iterative Dichotomiser 3". This algorithm exploits entropy and gain information as metrics to assess candidate splits. You can locate a number of Quinlan's research in this algorithm from 1986

C4.5: This algorithm is taken into consideration a later version of ID3, also advanced with the aid of Quinlan. It may use gain or gain fee records to evaluate cut up factors inside the decision tree.

CART: The term CART is an acronym for "classification and regression tree" and was brought by Leo Breiman. This algorithm commonly uses Gini impurities to decide the suitable assets for separation. The Gini impurity measures how regularly a randomly decided on attribute is misclassified. Whilst comparing the use of Gini impurities, a lower fee could be more best. The subsequent table suggests the contrast of parameters between distinct selection tree algorithms. Those algorithms are some of the most influential data mining algorithms in the studies community [4].

SLIQ:

SLIQ is a selection tree classifier that can manage both numeric and categorical attributes. It uses a singular pre-sorting approach inside the tree-increase section. This sort- ing method is included with a breadth-fist tree growing approach to allow classification of disk-resident datasets.

Sprint:

Dash algorithm is a classical set of rules for building a selection tree that is a widely used approach of information class. But, the sprint algorithm has high computational cost within the calculation of characteristic segmentation.

The dash set of rules has many advantages. This set of rules is unrestricted by means of reminiscence, and it is a type of scalable and parallel method of building choice trees. But there are also a few shortcomings. For instance, finding the nice segmentation factor of discrete attributes needs a big quantity of calculation, and the partition of non-stop attributes is unreasonable.

Table: Comparison of decision tree algorithm parameters

ALGORITHMS	ID3	CART	C4.5	SLIQ	SPRINT
Measure	Entropy info-gain	Gini diversity index	Entropy info-gain	Gini index	Gini index
Procedure	Top-down	Constructs	Top-down	Decision	Decision

	decision tree construction	binary decision tree	decision tree construction	tree construction in a breadth first manner	tree construction in a breadth first manner
Pruning	Pre pruning using a single pass algorithm	Post pruning based on cost complexity measure	Pre pruning using a single pass algorithm	Post pruning based on MDL principle	Post pruning based on MDL principle

a) ID3 Algorithm:

Iterative Dichotomiser 3 is an algorithm used to generate choice timber. This set of rules is based on Occam's razor: it prefers small selection timber over big selection trees. However, it does not constantly produce the smallest tree, and is consequently a heuristic. ID3 set of rules is one critical method in the era of decision tree type and so is broadly implemented. ID3 set of rules searches thru attributes of the education instances and extracts the attribute that best separates the given examples. If the characteristic perfectly classifies the training units then ID3 stops; in any other case it recursively operates on the n (in which n = wide variety of feasible values of an attribute) partitioned subsets to get their "first-rate" attribute. The set of rules makes use of a grasping search, that is, it alternatives the first-class attribute and in no way seems lower back to reconsider earlier choices.

The significant precept of ID3 algorithm is based on data concept.

ALGORITHM

Step 1: Begin with calculate class entropy.

Step 2: select the attributes and for every characteristic, calculate information gain

Step 3: maximum statistics benefit attributes are found out.

Step 4: put off node attribute, for future calculation. Repeat steps 2-four until all attribute were used.

ID3 (E, T_A, A)

E- Examples are the training examples.

T_A- target attribute is the attribute whose fee is to be expected with the aid of the tree. [5]

A- Attributes are the listing of attributes which can be examined by using the discovered selection tree.

- to start with create a root node.
- return the single-node tree Root, as + If all examples are positive,
- return the single-node tree Root, as - If all examples are negative,
- If wide variety of predicting attributes is empty, then return the single node tree Root, with label = most commonplace price of the goal attribute within the examples
- otherwise begin

- A — The attribute that nice classifies examples
- choice Tree characteristic for Root — A
- For each fine value, v_i , of A, f
- add a new tree branch to Root, similar to the test $A = v_i$
- If $v_i = \text{empty}$
- Then upload a leaf node with label as maximum not unusual target value in examples.
- Else upload the sub tree ID3 ($E(v_i), T_A, A$)
- end
- return Root

Because of this capability, ID3 can be implemented in several domains to generate decision tress

Merits

1. It produces the excessive accuracy result than the C4.Five set of rules.
 2. ID3 algorithm generally makes use of nominal attributes for classification with out a lacking values.
 3. It produces fake alarm rate and omission charge reduced, increasing the detection charge
- And decreasing the gap intake

Demerits

- 1 It has long looking time.
2. It takes the greater reminiscence than the C4.Five to large software execution.[12]

Applications of ID3:

- ✓ Application of ID3 in Diabetes
- ✓ Efficiency of ID3 in monitoring Heart Attack
- ✓ ID3 in Identifying Cancer
- ✓ Utility of ID3 in Computer Forensics
- ✓ Utility of ID3 in Knowledge Acquisition for Tolerances Design
- ✓ Use of ID3 for Breast Tumor Diagnosis
- ✓ Application of ID3 to reduce Cost sensitive Decision Tree

✓ Application of ID3 in Educational Field[16]

b) C4.5

C4.5 is an set of rules used to generate a choice tree evolved by using Ross Quinlan.C4.Five is an extension of Quinlan's in advance ID3 algorithm. The decision trees generated via C4.5 may be used for category, and because of this, C4.5 is frequently known as a statistical classifier.It's far developed to address Noisy statistics better, missing data, Pre and put up pruning of selection timber, Attributes with continuous values and Rule Derivation..

C4.5 builds choice trees from a fixed of training facts in the same way as ID3, the use of the concept of records gain and entropy further to advantage ratio. The notion of advantage ratio delivered earlier favors attributes that have a big range of values. If we've an characteristic D that has a wonderful fee for every record, then entropy (D, T) is zero, thus records gain (D, T) is maximal. To compensate for this Quinlan suggests the usage of the subsequent ratio in place of statistics gain as[11]

$$GainRatio(D, T) = \frac{Gain(D, T)}{SplitInfo(D, T)}$$

So benefit ratio is a modification of the facts benefit that reduces its bias on high branch attributes. Break up information (D, T) is the information because of the split of T on the idea of cost of express attribute D as [15]:

$$Split Info(D, T) = - \sum_{i=1}^k \frac{D_i}{T} \log_2 \frac{D_i}{T}$$

Advantages

1. It produces the appropriate result.
2. It takes the much less reminiscence to huge program execution.
3. It takes much less version build time.
4. It has short searching time.

Disadvantages

1. Empty branches.
2. Insignificant branches.
3. Over fitting.[9]

c) CART

CART (classification And Regression Tree) is a version of the decision tree algorithm. It could take care of both category and regression duties. Scikit-analyze uses the type And Regression Tree (CART) set of rules to train decision bushes (additionally called “developing” bushes). CART become first produced through Leo Breiman, Jerome Friedman, Richard Olshen, and Charles Stone in 1984.

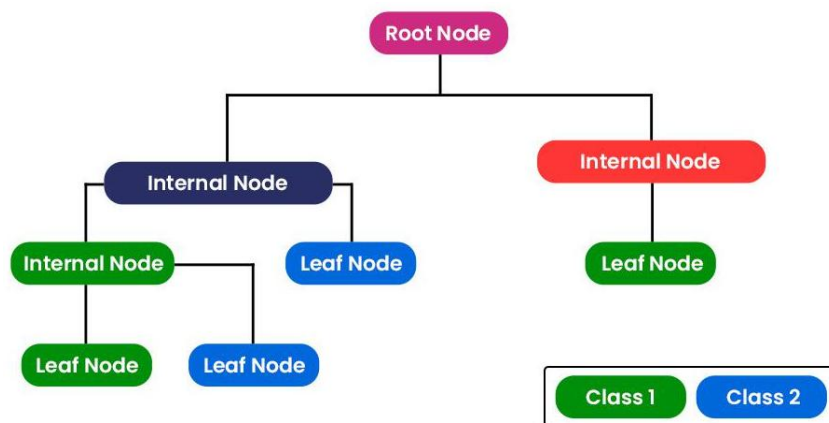
CART Algorithm

CART is a predictive set of rules utilized in gadget mastering and it explains how the goal variable’s values may be anticipated based totally on different subjects. It is a decision tree in which every fork is split into a predictor variable and each node has a prediction for the goal variable on the give up.

In the selection tree, nodes are break up into sub-nodes on the basis of a threshold price of an attribute. The foundation node is taken because the schooling set and is break up into by thinking about the first-class attribute and threshold fee. Similarly, the subsets also are split using the identical good judgment. This keeps until the ultimate natural sub-set is discovered within the tree or the maximum variety of leaves feasible in that developing tree.

The CART set of rules works via the subsequent manner:

- The exceptional split factor of each enter is obtained.
- Based totally on the exceptional cut up points of each input in Step 1, the brand new “excellent” break up point is identified.
- Split the chosen input according to the “pleasant” cut up point.
- Preserve splitting until a stopping rule is glad or no similarly appropriate splitting is available.



CART algorithm uses Gini Impurity to split the dataset into a decision tree. It does that by searching for the best homogeneity for the sub nodes, with the help of the Gini index criterion.

Gini index/Gini impurity

The Gini index is a metric for the type responsibilities in CART. It shows the sum of squared chances of every magnificence. It computes the diploma of probability of a particular variable this is wrongly being categorized while selected randomly and a variation of the Gini coefficient. It really works on specific variables, gives effects both “a success” or “failure” and therefore conducts binary splitting only.

The degree of the Gini index varies from zero to at least one,

- wherein zero depicts that each one the factors are allied to a certain elegance, or handiest one magnificence exists there.
- The Gini index of fee 1 signifies that all the factors are randomly dispensed throughout diverse lessons, and
- A fee of zero.Five denotes the factors are uniformly disbursed into a few lessons.

Mathematically, we will write Gini Impurity as follows::

$$\text{Gini} = 1 - \sum_{i=1}^n (p_i)^2$$

where p_i is the probability of an object being classified to a particular class.

CART model illustration

CART models are shaped by choosing input variables and evaluating cut up points on the ones variables until the right tree is produced.

Steps to create a decision Tree the usage of the CART set of rules:

- **Greedy Algorithm:** on this The enter area is divided using the grasping method which is referred to as a recursive binary spitting. This is a numerical method inside which all the values are aligned and numerous other split points are attempted and assessed using a price characteristic.
- **Preventing Criterion:** because it works its manner down the tree with the schooling statistics, the recursive binary splitting method defined above should know when to stop splitting. The

maximum common halting method is to make use of a minimum amount of schooling statistics allotted to each leaf node. If the depend is smaller than the specified threshold, the cut up is rejected and additionally the node is considered the remaining leaf node.

- **Tree pruning:** decision tree's complexity is described as the variety of splits within the tree. Trees with fewer branches are recommended as they're easy to grasp and less susceptible to cluster the statistics. Operating through each leaf node inside the tree and evaluating the effect of deleting it the usage of a keep-out test set is the fastest and only pruning approach.
- **Records instruction for the CART:** No unique information coaching is needed for the CART algorithm [16].

Merits of CART

- Results are simplistic.
- Category and regression timber are nonparametric and nonlinear.
- Category and regression trees implicitly perform feature selection.
- Outliers have no meaningful effect on cart.
- It requires minimal supervision and produces clean-to-apprehend fashions.

Demerits of CART

- Overfitting.
- Excessive Variance.
- Low bias.
- The tree shape can be risky.

Applications of the CART algorithm

- ✓ For brief records insights.
- ✓ In Blood Donors class.
- ✓ For environmental and ecological information.
- ✓ Within the monetary sectors.

d) SLIQ

Supervised learning in quest (SLIQ) algorithm

SLIQ is a classifier of decision tree, which could take each numerical and express attributes it builds compact and correct bushes. Pre-sorting technique is used in the tree growing section and an less expensive pruning set of rules. It is suitable for classification of big disk-resident datasets, separately of the wide variety of instructions, attributes and information [21].

Tree Building

Make tree (education facts t)

Partition (t)

Partition (facts s)

If (all points in s are within the identical class)

Then go back;

Compare splits for every characteristic a;

Use first-class break up to separation s into s1 and s2;

Partition (s1);

Partition (s2);

The gini-index is used to evaluate the “goodness” of the opportunity splits for an characteristic If a information set t includes examples from n training, gini(t) is given as

$$gini(T) = 1 - \sum P_j^2$$

Where p_j is the relative frequency of sophistication j in t. After splitting t into subset t1 and t2 the gin index of the break up statistics is defined

$$gini(T)_{split} = \frac{|T1|}{|T|} gini(T1) + \frac{|T2|}{|T|} gini(T2)$$

The primary technique applied by using SLIQ is a scheme that removes the need to sort statistics at each node it creates a separate list for each attribute of the training records. A separate list, called class listing, is produced for the class labels attached to the examples. SLIQ calls for that the Class list and (best) one characteristic list can be stored within the memory at any time.

SLIQ – algorithm

1. Start **Pre-sorting** of the samples.
2. As long as the forestall criterion has not been reached
 1. For each characteristic
 1. Area all nodes into a category histogram.
 2. Begin **evaluation** of the splits.
 2. Pick out a split.
3. Replace the decision tree; for every new node

Replace its magnificence listing (nodes).

Merits

- It can handle Large disk-resident data
- Requires some data per record that stay in memory required.
- There is no need to rewrite the lists during split.
- Reassignment of records is easy.

Demerits

- It must stay in memory all the time which limits the amount of data that can be classified by SLIQ

e) SPRINT

A choice tree based classifier known as sprint (acronym for Scalable Parallelizable Induction of choice bushes) is provided, which, as the call indicates, is both scalable w.R.T. Length of dataset in addition to parallelizable. Allow us to study some of the principle ideas involved in this

Algorithm[17]. Choice-tree is one of the crucial branches of facts mining algorithms. Maximum of the choice-tree algorithms, which include ID3, C4.Five, CART, and etc, require that the

Training pattern datasets live in reminiscence, which is impractical for information mining regarding with thousands and tens of millions datasets. To address the trouble of restricted main reminiscence, John Shafer proposed dash to apply to very massive scale training units and create compact correct selection-tree. Sprint has true expansibility and parallelizability, does now not restricted through the dimensions of memory, runs rapid, and lets in multiple processors create a decision-tree version at collaboratively. On this we take sprint as an example and discuss the design of Hadoop. Primarily based selection tree algorithms[20].

A selection tree classifier is built in two levels [3] [2]: a boom segment and a prune segment. Within the boom phase, the tree is built by way of recursively partitioning the facts till each partition is both “pure” (all participants belong to the same class) or small enough (a parameter set through the user). This procedure is shown in discern. The shape of the break up used to partition the statistics relies upon on the form of the characteristic used within the split. Splits for a continuous attribute A are of the form $\text{cost}(A) < c$ in which t is a cost in the area of A. Splits for a specific characteristic A are of the form $\text{price}(A) \in X$ where $X \subset \text{domain}(A)$. We don't forget only binary splits because they typically cause greater correct timber; but, our strategies may be prolonged to deal with multi-manner splits. Once the tree has been completely grown, it is pruned inside the 2nd segment to generalize the tree by casting off dependence on statistical noise or variation that can be precise best to the training set. The tree growth section is computationally lots extra high-priced than pruning, for the reason that facts is scanned more than one instances on this part of the computation. Pruning requires get right of entry to best to the fully grown selection tree. Our experience based on our preceding paintings on SLIQ has been that the pruning segment usually takes much less than 1% of the total time needed to construct a classifier. We consequently awareness most effective at the tree-boom section. For pruning, we use the set of rules utilized in SLIQ, that's based totally on the minimum Description length principle.

Remember, as an instance, the credit score trouble, wherein a credit score employer desires to classify clients based on a schooling database containing information approximately them. The classification tree is generated in a top down fashion as follows: The facts is recursively Partitioned until either every partition is sufficiently 'natural' (parameterized by way of a person designated self assurance), or is too small to yield statistically sizeable effects. If neither of The above two standards keep, the quality viable cut up is chosen (as an example, education degree (e-stage) at root node in and facts is partitioned in keeping with that split. We shall see in Segment 2.2.4 how the 'goodness' of a split is evaluated. As shown in determine, only binary splits are accomplished. For a non-stop attribute A, it's miles of the shape: $\text{cost}(A) < a$, in which A is a price in $\text{domain}(A)$, while for a specific attribute A, splits are of the shape: $\text{cost}(A) \geq S$, wherein S $\text{domain}(A)$ how splits are completed.

Algorithm :-Flow of SPRINT algorithm

```
Partition(Data D)
begin
if more than % (=confidence) of the elements in D
belong to the same class OR size
of D < minsize then
return;
end if
for each attribute A do
evaluate splits on attribute A (Section 2.2.4)
end for
Use the best split to partition D into D 1 and D2
Partition(D 1)
Partition(D 2)
end
```

The famous CART [18] and C4.5 [17] classifiers, for example, develop timber depth-first and repeatedly kind the records at every node of the tree to reach on the first-rate splits for numeric attributes. SLIQ, however, replaces this repeated sorting with one-time sort by the usage of separate lists for every characteristic (see [19] for info). SLIQ uses a statistics structure called a class listing which have to stay reminiscence resident always. The dimensions of this structure is proportional to the variety of parent three: example of attribute lists enter information, and this is what limits the

Wide variety of input facts that SLIQ can deal with. Sprint addresses the above two issues in another way from preceding algorithms; it has no restriction on the dimensions of input and but is a

quick set of rules. It stocks with SLIQ the benefit of a one-time kind, but uses specific- facts structures. Particularly, there is no shape just like the class list that grows with the size of enter and wishes to be reminiscence resident.

Merits

- Removes all of the memory restrictions
- Fast and scalable and it can easily parallelized.

4. Conclusions

Decision trees are truly responding to a problem of discrimination is one of the few methods that can be supplied fast sufficient to a non-specialist audience data processing with out getting misplaced in hard to understand mathematical formulations. Here in this paper, we discussed about the various types of decision trees and the importance of its algorithms along with the merits and demerits. Some of the applications are also given for the decision trees.

5. Authors' Biography



O.Yamini received her M.Phil degree from Sri Venkateswara University in year 2020. She is currently a research scholar in department of computer science in S V University, Tirupati. Her research interests includes Data Mining, Data Structures, Software Engineering, Cryptography and Network Security, Artificial Intelligence, IoT and Machine Learning.



Dr.G.V.RameshBabu is working as Assistant Professor in the Department of Computer Science, Sri Venkateswara University, Tirupati. He awarded his Ph.D in the year 2015 from S.V.University. He pursued M.Tech from Vinayaka Mission University, Selam in the year 2005 and MCA from Osmania University, Hyderabad. He has 15 years of teaching experience in P.G Course and 2 years in UG course. He published 25 Research articles in different reputed journals. He visited and presented 30 research papers at various National and International Conferences like China, Malaysia, and Singapore etc. He has specialisations in Privacy Preserving Data Mining, Cryptography & Network Security, Software Engineering and E-Commerce.

6. References

1. Heikki M (1996) Data mining: machine learning, statistics, and databases. In: SSDBM '96: proceedings of the eighth international conference on scientific and statistical database management, June 1996, pp 2–9

2. Han J, Kamber M, Pei J (2012) Data mining concepts and techniques, 3rd edn. Elsevier, Netherlands
3. Jain AK, Murty MN, Flynn PJ (1999) Data clustering: a review. *ACM Comput. Surv* 31(3):1–60
4. Fan Chin-Yuan, Fan Pei-Shu, Chan Te-Yi, Chang Shu-Hao (2012) Using hybrid data mining and machine learning clustering analysis to predict the turnover rate for technology professionals. *Expert Syst Appl* 39:8844–8851
5. Tan KC, Teoh EJ, Yua Q, Goh KC (2009) A hybrid evolutionary algorithm for attribute selection in data mining. *Expert Syst Appl* 36(4):8616–8630
6. Hea Z, Xua X, Huangb JZ, Denga S (2004) Mining class outliers: concepts, algorithms and applications in CRM. *Expert Syst Appl* 27(4):681e97
7. Liao TW, Triantaphyllou E (2007) Recent advances in data mining of enterprise data: algorithms and applications. World Scientific Publishing, Singapore, pp 111–145
8. Matthew N. Anyanwu&Sajjan G. Shiva “Comparative Analysis of Serial Decision Tree Classification” in *Algorithms International Journal of Computer Science and Security, (IJCSS) Volume (3): Issue (3) 230*
9. Diana Gorea, Sabin CorneliuBuraga “Towards Integrating Decision Tree with XML Technologies” in 8th International Conference on Development and Application Systems, Suceava, Romania, May 25 -27, 2006
10. Jiawei Han, Micheline Kamber, Data mining: concepts and techniques Morgan Kaufmann. 2006,58-61.
11. Badr HSSINA, Abdelkarim MERBOUHA,Hanane EZZIKOURI,Mohammed ERRITALI, A comparative study of decision tree ID3 and C4.5, (*IJACSA) International Journal of Advanced Computer Science and Applications,Special Issue on Advances in Vehicular Ad Hoc Networking and Applications.*
12. N. Chandra Sekhar Reddy, K. Sai Prasad and A. Mounika *International Journal of Computational Intelligence Research* ISSN 0973-1873 Volume 13, Number 8 (2017), pp. 2135-2142 © Research India Publications
13. L.Surya Prasanthi et al, ID3 and Its Applications in Generation of Decision Trees across Various Domains- Survey / (*IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 6 (6) , 2015, 5353-5357, ISSN: 0975-9646.*
14. D. Joița, "Unsupervised static discretization methods in data mining," *Titu Maiorescu University, Bucharest, Romania, 2010*
15. Radhwan H. A. Alsagheer *et al*, Popular Decision Tree Algorithms of Data Mining Techniques: A Review, *International Journal of Computer Science and Mobile Computing, Vol.6 Issue.6, June-2017, pg. 133-142, SSN 2320–088X IMPACT FACTOR: 6.017*
16. <https://www.geeksforgeeks.org/cart-classification-and-regression-tree-in-machine-learning>.
17. Yang Chen-zhu. The Research of Data Mining Based on HADOOP [D]. Chong Qing. Chongqing University 2010.11:pp42~43
18. Zhu Min,Wan Jian-yi,Wang Ming-wen. Design and implementation of parallel decision tree classification algorithm based on MR[J]. *Journal of Guangxi Normal University Natural Science.Vol.29 No.1 2011.3 : pp84*
19. <http://archive.ics.uci.edu/ml/datasets.html>

20. K. Veerapathap Reddy¹, T. Chandra Sekhara Reddy, Data Classifying Larger Datasets By Using Sprints , International Journal Of Innovative Research In Technology, 2014 IJIRT | Volume 1 Issue 5 | ISSN : 2349-6002
21. Kamlesh Kumar Joshi, Pawan Patidar, Joshi, An Approach To Construct Decision Tree Using Sliq And Knn For Land Grading System, 5(1): January-March, 2015, ISSN: 2277-5528 Impact Factor: 2.745 (SIJF)