IJMTST is UGC Approved Journal with Journal Id: 43137

International Journal for Modern Trends in Science and Technology, 8(S08):24-27, 2022

ISSN: 2455-3778 online
DOI: https://doi.org/10.46501/IJMTST08S0804
Available online at: https://www.ijmtst.com/vol8si08.html

Decision Tree Induction for Classification of Objects in Machine Learning

G.Sudha Rani[1] | P.Surekha[2]

[1]Computer Science and Information Technology, Chalapathi Institute of Engineering and Technology, Guntur, Andhra Pradesh, India.
[2]Computer Science and Engineering , St.Ann's College of Engineering and Technology., Chirala, Vetapalem, Andhra Pradesh, India.

To Cite this Article
G.Sudha Rani and P.Surekha. Decision Tree Induction for Classification of Objects in Machine Learning. International Journal for Modern Trends in Science and Technology 2022, 8(S08), pp. 24-27. https://doi.org/10.46501/IJMTST08S0804

Article Info
Received: 26 May 2022; Accepted: 24 June 2022; Published: 28 June 2022.

ABSTRACT

Decision Tree is a regulated learning technique utilized in information digging for grouping and relapse strategies. A tree helps us in dynamic purposes. The choice tree makes grouping or relapse models as a tree structure. It isolates an informational index into more modest subsets, and simultaneously, the choice tree is consistently evolved. The last tree is a tree with the choice hubs and leaf hubs. A choice hub has no less than two branches. The leaf hubs show a characterization or choice. We can't achieve more split on leaf hubs The highest choice hub in a tree that connects with the best indicator called the root hub. Choice trees can manage both clear cut and mathematical information. This paper is presents induction of decision tree for various real time applications in machine learning.

KEYWORDS: Decision Tree, Machine Learning, Real time objects, Classification

1. INTRODUCTION

In statistics, data mining, and machine learning, Decision Tree Learning is a supervised learning approach. To derive conclusions about a set of observations, a classification or regression decision tree is utilised as a predictive model in this approach. Classification trees are tree models in which the goal variable can take a discrete set of values; in these tree structures, leaves indicate class labels and branches represent feature combinations that lead to those class labels.

Regression trees are decision trees in which the target variable can take continuous values (usually real numbers). A tree is constructed by dividing the source set, which is the tree's root node, into subsets, which are the tree's successor offspring.

The splitting is based on a series of categorization feature-based splitting rules. Recursive partitioning is the process of repeating this method on each derived subset. Decision trees are a type of data mining technique that combines mathematical and computational techniques to aid in the description, categorization, and generalisation of a set of data.

(x,Y)=(x1,x2,x3,.....,xk,Y)

There are two types of decision trees used in data mining: When the expected outcome is the class (discrete) to

which the data belongs, classification tree analysis is used. When the projected outcome is a real number (for example, the price of a property or the length of a patient's hospital stay), regression tree analysis is used.

Breiman et al. first presented classification and regression tree (CART) analysis in 1984, and it is an umbrella word that refers to either of the aforementioned approaches. There are some parallels between regression and classification trees, but there are also some distinctions, such as the technique for determining where to divide.

Some strategies, known as ensemble methods, create many decision trees: Trees that have been boostedBuilding an ensemble gradually by training each new instance to emphasise previously mis-modelled training instances. AdaBoost is a good example.

These can be used to solve problems of the regression and classification types. An early ensemble method called bootstrap aggregated (or bagged) decision trees create several decision trees by resampling training data with replacement and voting the trees for a consensus forecast.

The majority of decision tree algorithm's function top-down, selecting a variable at each step that optimally separates the set of objects. Different algorithms employ various metrics to determine what is "best. "These are used to assess the homogeneity of the target variable across subsets.

The following are some examples.

These criteria are applied to each candidate subgroup, and the resulting values are averaged to offer a measure of the split's quality.

2. LITERATURE REVIEW

[M Kamber] Efficiency and scalability are fundamental issues concerning data mining in large databases. Although classification has been studied extensively, few of the known methods take serious consideration of efficient induction in large databases and the analysis of data at multiple abstraction levels.

[JR Quinlan] The technology for building knowledge-based systems by inductive inference from examples has been demonstrated successfully in several practical applications. This paper summarizes an approach to synthesizing decision trees that has been used in a variety of systems, and it describes one such system, ID3, in detail.

[John R, Koza] Genetic algorithms are highly parallel mathematical algorithms that transform populations of individual mathematical objects (typically fixed-length binary character strings) into new populations using operations patterned after (1) natural genetic operations such as sexual recombination (crossover) and (2) fitness proportionate reproduction (Darwinian survival of the fittest).

[Jong Woo Kim, Byung Hun Lee, Michael J. Shaw, Hsin-Lu Chang & Matthew Nelson]
Customization and personalization services are a critical success factor for Internet stores and Web service providers. This paper studies personalized recommendation techniques that suggest products or services to the customers of Internet storefronts based on their demographics or past purchasing behaviour. The underlining theories of recommendation techniques are statistics, data mining, artificial intelligence, and rule-based matching.

[Pat Langley, Herbert A. Simon] Machine learning is the study of computational methods for improving performance by mechanizing the acquisition of knowledge from experience. Expert performance requires much domain-specific knowledge, and knowledge engineering has produced hundreds of AI expert systems that are now used regularly in industry.

[D. Lavanya, KU Rani] Classification is one of the fundamental tasks in data miningand has also been studied extensively in statistics, machinelearning, neural networks and expert systems over decades[1,2]. The input for classification is a set of training records (training instances), where each record has several attributes.

[Arie Ben-David] Decision trees that are based on information-theory are useful paradigms for learning from examples. However, in some real-world applications, known information-theoretic methods frequently generate nonmonotonic decision trees, in which objects with better attribute values are sometimes classified to lower classes than objects with inferior values.

[Marlon Nunez] At present, algorithms of the ID3 family are not based on background knowledge. For that reason, most of the time they are neither logical nor

understandable to experts. These algorithms cannot perform different types of generalization as others can do (Michalski, 1983; Kodratoff, 1983), nor can they can reduce the cost of classifications. The algorithm presented in this paper tries to generate more logical and understandable decision trees than those generated by ID3-like algorithms; it executes various types of generalization and at the same time reduces the classification cost by means of background knowledge.

3. METHODS

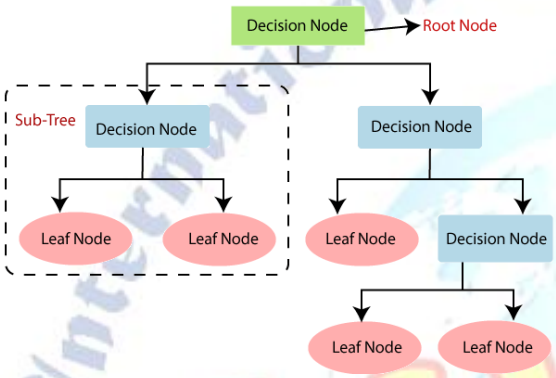


Figure-1 Decision Classification

There are various algorithms in Machine learning, so choosing the best algorithm for the given dataset and problem is the main point to remember while creating a machine learning model. Below are the two reasons for using the Decision tree:

- Decision Trees usually mimic human thinking ability while making a decision, so it is easy to understand.
- The logic behind the decision tree can be easily understood because it shows a tree-like structure.
- **Decision Tree Terminologies**
- **Root Node:**
- Root node is from where the decision tree starts. It represents the entire dataset, which further gets divided into two or more homogeneous sets.
- **Leaf Node**: Leaf nodes are the final output node, and the tree cannot be segregated further after getting a leaf node.
- **Splitting:** Splitting is the process of dividing the decision node/root node into sub-nodes according to the given conditions.
- **Branch/Sub Tree:** A tree formed by splitting the tree.
- **Pruning:** Pruning is the process of removing the unwanted branches from the tree.
- **Parent/Child node:** The root node of the tree is called the parent node, and other nodes are called the child nodes.

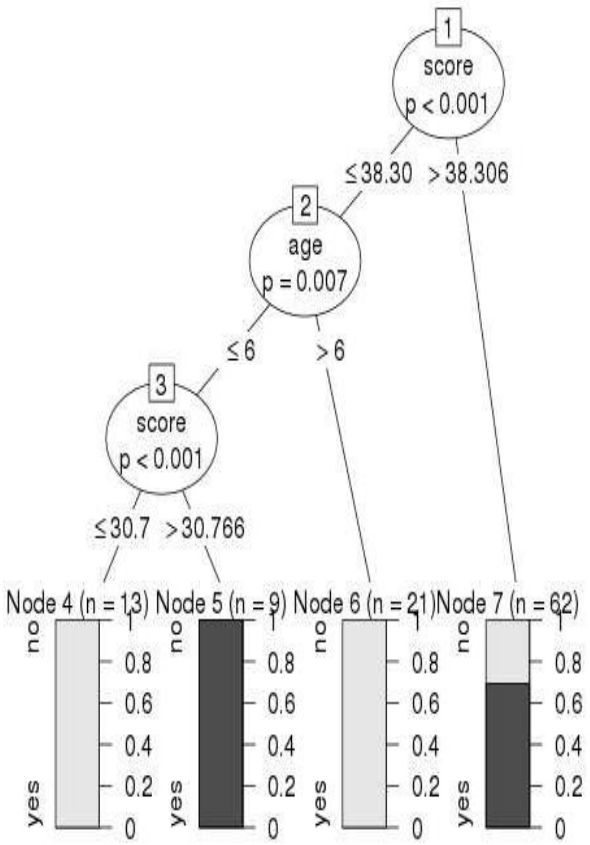
4. RESULTS

```
# Load the party package. It will automatically load other
# dependent packages.
library(party)

# Print some records from data set readingSkills.
print(head(readingSkills))
```

When we execute the above code, it produces the following result and chart –

```
nativeSpeaker age shoeSize score
1 yes 5 24.83189 32.29385
2 yes 6 25.95238 36.63105
3 no 11 30.42170 49.60593
4 yes 7 28.66450 40.28456
5 yes 11 31.88207 55.46085
6 yes 10 30.07843 52.83124
Loading required package: methods
Loading required package: grid
.....
```

```
# Load the party package. It will automatically load other
# dependent packages.
library(party)

# Create the input data frame.
input.dat <-readingSkills[c(1:105),]

# Give the chart file a name.
png(file ="decision_tree.png")

# Create the tree.
output.tree<-ctree(
nativeSpeaker~ age +shoeSize+ score,
  data = input.dat)

# Plot the tree.
plot(output.tree)

# Save the file.
dev.off()
```



(Provide comparative analysis with detail result discussion)

4. CONCLUSION AND FUTURE WORK

Decision trees are used to solve classification problems. Decision trees are one of the few methods that can be presented quickly to people who are not experts in data processing or machine learning without getting lost in difficult-to-understand mathematical formulations. In this chapter, we discussed the key elements required to build a decision tree from a dataset, as well as the pruning methods, pre-pruning and post-pruning. As a possible solution to the variance problem, we also mentioned ensemble meta-algorithms.

Conflict of interest statement

Authors declare that they do not have any conflict of interest.

REFERENCES

[1] Grąbczewski, K. (2014). Meta-learning in decision tree induction (Vol. 1). Cham: Springer International Publishing.

[2] rętowski, M. (2004, June). An evolutionary algorithm for oblique decision tree induction. In International Conference on Artificial Intelligence and Soft Computing (pp. 432-437). Springer, Berlin, Heidelberg.

[3] ernández, V. A. S., Monroy, R., Medina-Pérez, M. A., Loyola-González, O., & Herrera, F. (2021). A practical tutorial for decision tree induction: Evaluation measures for candidate splits and opportunities. ACM Computing Surveys (CSUR), 54(1), 1-38.

[4] Buntine, W. L. (2013). Decision tree induction systems: a Bayesian analysis. arXiv preprint arXiv:1304.2732.

[5] Bhukya, D. P., &Ramachandram, S. (2010). Decision tree induction: an approach for data classification using AVL-tree. International Journal of Computer and Electrical Engineering, 2(4), 660.