



RAMA UNIVERSITY

www.ramauniversity.ac.in

FACULTY OF ENGINEERING

DATA MINING & WAREHOUSEING LECTURE-22

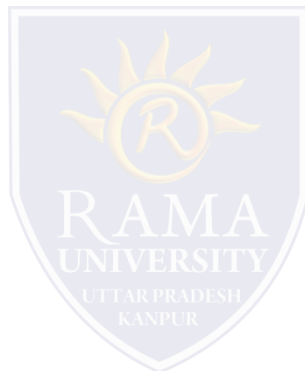
MR. DHIRENDRA

ASSISTANT PROFESSOR

RAMA UNIVERSITY

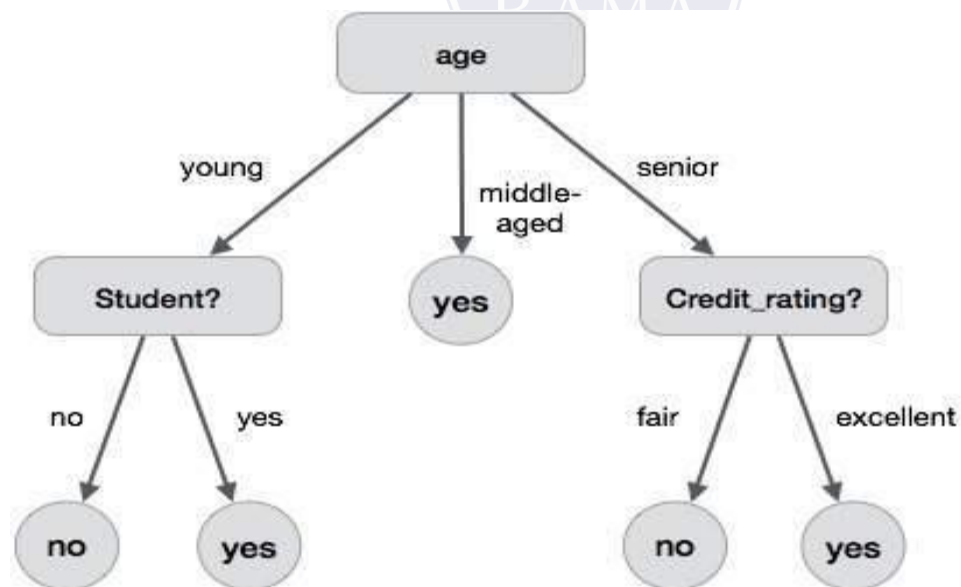
OUTLINE

- ❖ **DECISION TREE INDUCTION**
- ❖ **BENEFITS OF DECISION TREE**
- ❖ **TREE PRUNING**
- ❖ **COST COMPLEXITY**
- ❖ **MCQ**
- ❖ **REFERENCES**



Decision Tree Induction

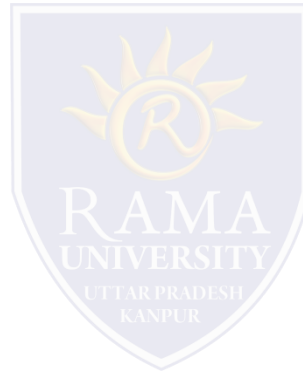
- A decision tree is a structure that includes a root node, branches, and leaf nodes. Each internal node denotes a test on an attribute, each branch denotes the outcome of a test, and each leaf node holds a class label. The topmost node in the tree is the root node.
- The following decision tree is for the concept `buy_computer` that indicates whether a customer at a company is likely to buy a computer or not. Each internal node represents a test on an attribute. Each leaf node represents a class.



Benefits of Decision Tree

The benefits of having a decision tree are as follows –

- It does not require any domain knowledge.
- It is easy to comprehend.
- The learning and classification steps of a decision tree are simple and fast.



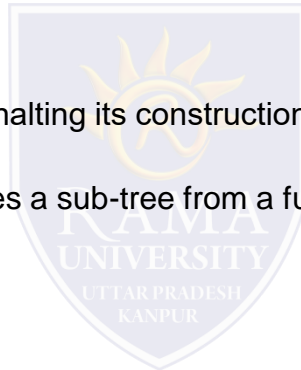
Tree Pruning

Tree pruning is performed in order to remove anomalies in the training data due to noise or outliers. The pruned trees are smaller and less complex.

Tree Pruning Approaches

There are two approaches to prune a tree –

- Pre-pruning – The tree is pruned by halting its construction early.
- Post-pruning - This approach removes a sub-tree from a fully grown tree.



Cost Complexity

The cost complexity is measured by the following two parameters –

- Number of leaves in the tree, and
- Error rate of the tree.



Multiple Choice Question

1. Which of the following is the other name of Data mining?

- a) Exploratory data analysis.
- b) Data driven discovery.
- c) Deductive learning.
- d) All of the above.

2.. Which of the following is a predictive model?

- a) Clustering
- b) Regression
- c) Summarization
- d) Association rules.

3. Which of the following is a descriptive model?

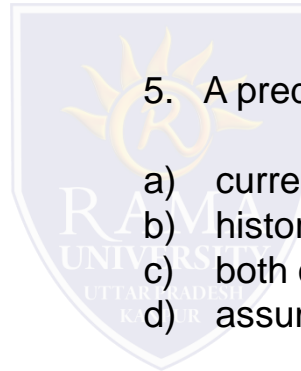
- a) Classification
- b) Regression
- c) Sequence discovery.
- d) Association rules.

4. A _____ model identifies patterns or relationships.

- a) Descriptive
- b) Predictive
- c) Regression
- d) Time series analysis.

5. A predictive model makes use of _____.

- a) current data.
- b) historical data.
- c) both current and historical data.
- d) assumptions



REFERENCES

- https://www.tutorialspoint.com/dwh/dwh_overview.htm
- <http://myweb.sabanciuniv.edu/rdehkharghani/files/2016/02/The-Morgan-Kaufmann-Series-in-Data-Management-Systems-Jiawei-Han-Micheline-Kamber-Jian-Pei-Data-Mining.-Concepts-and-Techniques-3rd-Edition-Morgan-Kaufmann-2011.pdf> DATA MINING BOOK WRITTEN BY Micheline Kamber
- <https://www.javatpoint.com/three-tier-data-warehouse-architecture>
- M.H. Dunham, “ Data Mining: Introductory & Advanced Topics” Pearson Education
- Jiawei Han, Micheline Kamber, “ Data Mining Concepts & Techniques” Elsevier
- Sam Anahory, Dennis Murray,” data warehousing in the Real World: A Practical Guide for Building Decision Support Systems, “ Pearson Education
- Mallach,” Data Warehousing System”, TMH
- R. Agrawal, A. Gupta, and S. Sarawagi. Modeling multidimensional databases. ICDE’97 S. Chaudhuri and U. Dayal. An overview of data warehousing and OLAP technology. ACM SIGMOD Record, 26:65-74, 1997
- S. Agarwal, R. Agrawal, P. M. Deshpande, A. Gupta, J. F. Naughton, R. Ramakrishnan, and S. Sarawagi. On the computation of multidimensional aggregates. VLDB’96 D. Agrawal, A. E. Abbadi, A. Singh, and T. Yurek. Efficient view maintenance in data warehouses. SIGMOD’97
- E. F. Codd, S. B. Codd, and C. T. Salley. Beyond decision support. Computer World, 27, July 1993.
- J. Gray, et al. Data cube: A relational aggregation operator generalizing group-by, cross-tab and sub-totals. Data Mining and Knowledge Discovery, 1:29-54, 1997.