

# Rule Based Classification

- Rule-based classification in data mining is a technique in which the records are ***classified / categorized*** by using a set of “**IF...THEN...**” rules.
- Rules are a good way of representing information or bits of knowledge.
- A ***rule-based classifier*** uses a set of **IF-THEN** rules for ***classification***.
- An ***IF-THEN rule*** is an expression of the form  
**IF *condition* THEN *conclusion***.

- An example is rule **R1**,  
**R1:**    **IF** *age = youth* AND *student = yes*  
          **THEN** *buys\_computer = yes*.
- The “**IF**” part (or left side) of a rule is known as the rule ***antecedent*** or ***precondition***. The “**THEN**” part (or right side) is the rule ***consequent***.
- In the rule ***antecedent***, consists of *one or more* attribute tests that are **logically ANDed**. The rule’s ***consequent*** contains a **class prediction**.
- **R1** can also be written as  
**R1:**  $(age = youth) \wedge (student = yes) \Rightarrow (buys\ computer = yes)$

- If the ***condition*** (i.e., all the attribute tests) in a ***rule antecedent*** holds ***true*** for a given tuple, we say that the rule antecedent is ***satisfied*** (or simply, that the rule is satisfied) and that the rule **covers** the tuple.
- A rule **R** can be assessed by its ***coverage*** and ***accuracy***.
- **Coverage**: A rule's coverage is the percentage of tuples that are **covered** by the rule.
- **Accuracy**: A rule's coverage is the percentage of tuples that are **correctly classified** by the rule.

- Given a tuple, **X**, from a class labelled data set, **D**,
- let  $n_{covers}$  be the number of tuples covered by a rule **R**,  $n_{correct}$  be the number of tuples correctly classified by the rule **R**; and  $|D|$  be the number of tuples in D.
- We can define the **coverage** and **accuracy** of rule R as

$$\mathbf{Coverage(R)} = \frac{n_{covers}}{|D|}$$

$$\mathbf{Accuracy(R)} = \frac{n_{correct}}{n_{covers}}$$

- **Example:** Rule **accuracy** and **coverage**. Let's go back to our customer data set. Our task is to predict whether a customer will buy a computer.

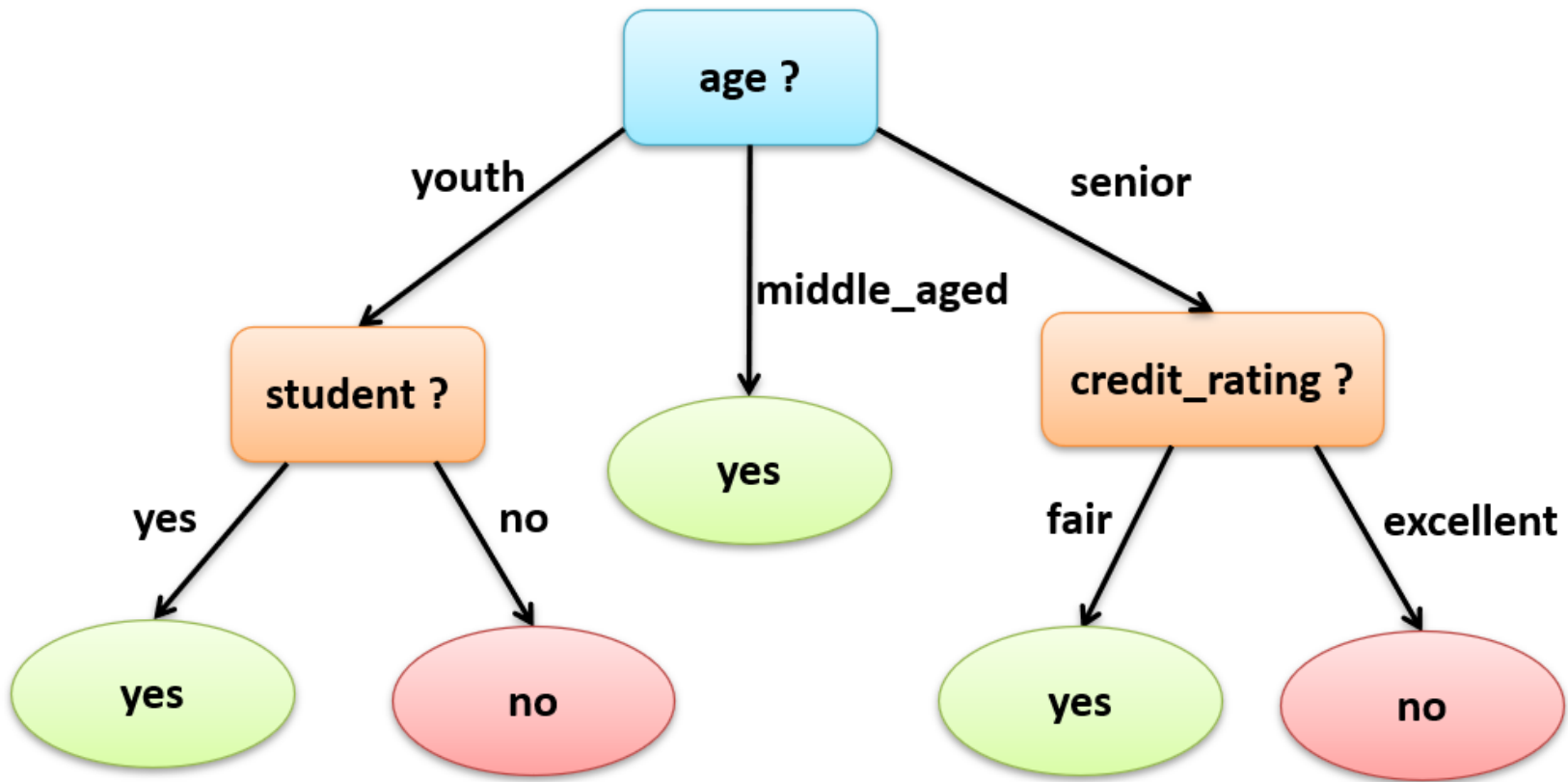
**R1:**    **IF** *age = youth* **AND** *student = yes*  
          **THEN** *buys\_computer = yes*.

- Consider rule **R1**, which covers **2** of the **14** tuples. It can correctly classify **both** tuples. Therefore,  
          *coverage* R1 (D) =  $2/14 = 14.28\%$  and  
          *accuracy* R1 (D) =  $2/2 = 100\%$ .

## Rule Extraction from a Decision Tree:

- *Decision tree classifiers* are a popular method of classification.
- In comparison with a decision tree, the ***IF-THEN*** rules may be easier for humans to understand, particularly if the decision tree is ***very large***.
- In this, we look at how to build a ***rule based classifier*** by extracting ***IF-THEN*** rules from a ***decision tree***.

- To extract rules from a decision tree, one rule is created for each path from the **root** to a **leaf** node.
- Each splitting criterion along a given path is logically **AND**ed to form the rule antecedent (“IF” part). The leaf node holds the ***class prediction***, forming the rule consequent (“THEN” part).
- **Example:** Extracting ***classification rules*** from the following decision tree.



The above decision tree can be converted into ***IF-THEN classification rules*** by tracing the path from the **root** node to each **leaf node** in the tree.

- The rules extracted from the above Figure are as follows:
- **R1:** IF *age = youth* AND *student = yes* THEN *buys\_computer = yes*
- **R2:** IF *age = youth* AND *student = no* THEN *buys\_computer = no*
- **R3:** IF *age = middle\_aged* THEN *buys\_computer = yes*
- **R4:** IF *age = senior* AND *credit\_rating = excellent* THEN *buys\_computer = yes*
- **R5:** IF *age = senior* AND *credit\_rating = fair* THEN *buys\_computer = no*