41256

Jan 2, 2022

Assignment No. 2

* Problem Statement:
  To find the decision based on a given scenario from a dataset using Decision Tree classifier.

* Objective:
  To understand how the decision tree classifier algorithm works on a given dataset.

* Outcome:
  To find the decision based on the given scenario of people with income, gender or marital status information from a dataset using Decision Tree classifier.

* Theory:

A decision Tree is a simple representation for classifing examples. It is a Supervised Machine Learning where the data is continously split according to a certain Parameter.

Decision Tree consists of →

1. Node : Test for the value of certain attribute
2. Edge/Branch : correspond to the outcome the test on connect to the next node or leaf.
3. Leaf nodes : Terminal Nodes that Predict the outcome.

⊕ There are two main type of Decision Tree

1. Classification Trees
2. Regression Trees

A) Classification Trees (Yes / No types)
→ Decision variable is catagorical / discrete
→ Built through a process known as binary recursive Partitioning.

B) Regression Tree (continous data types)
→ Target variables can take continous values.

* In Decision Tree classification, a new example is classified by submitting it to a series of tests that determine the class label of example. These tests are organised in hierarchical structure called a decision tree. They follow Divide-and-conquer Algorithm.
↳ Select a test for root node. create branch for each possible outcome of test.
↳ Split instance into subset. One for each

branch extending from the node.
↳ Repeat recursively for each branch, using only instance that reach the branch.
↳ Stop recursion for a branch if all its instance have the same class.

⊕ Advantages of classification with Decision Tree

1. Inexpensive to construct.
2. Extreamly fast at classifing unknown records
3. Easy to interpret for Small-Size trees
4. Accuracy comparable to other classification techniques for many simple data sets.
5. Excludes unimportant features.

⊕ Disadvantage of classification with Decision Tr

1. Easy to overfit
2. Decision Boundary restricted to being Parallel to attribute axis.
3. Decision tree models are often biased towards splits on features having a large number of levels.
4. Small changes in the training Data can result in large changes in decision logic.
5. Large tree can be difficult to interpret and the decision they make seem counter intuitive.

⊕ Application of Decision trees in real life:

1. Biomedical Engineering
2. Financial analysis
3. Astronomy
4. System control
5. Manufacturing & Production
6. Medicines
7. Physics

* Test Cases:

Split the dataset into training & testing dataset.

* Conclusion:

Thus Decision tree classifier was used to Predict & made a decision on the given Dataset.