Decision Tree Algorithm – Regression Problem:

[1]. Read Data, Identify Target Variable and Input Features (Attributes)

[2]. Calculate SD, CV, Mean using below equations

Mean = $\frac{\sum x}{n}$

SD=$Square\ Root\ of\ \frac{\sum(x-mean)\text{^}2}{n}$

CV=$\frac{SD}{mean}$

[3]. The dataset is then split on the different attributes. The standard deviation for each branch is calculated.

SD(Attribute)=$\sum_{Brach\ \epsilon\ \text{Attribute}} W(branch)SD(Branch)$

[4]. The resulting standard deviation is subtracted from the standard deviation before the split. The result is the standard deviation reduction.

SDR=SD-SD(Attribute)

[5]. The attribute with the largest standard deviation reduction is chosen for the decision node.

The dataset is divided based on the values of the selected attribute. This process is run recursively on the non-leaf branches, until all data is processed. (Termination: coefficient of deviation (CV) for a branch becomes smaller than a certain threshold (e.g., 10%) and/or when too few instances (n) remain in the branch)

Decision Tree Algorithm – Classification Problem – Information Gain Index:

[1]. Read Data, Identify Target Variable and Input Features (Attributes)

[2]. Calculate entropy of Target Variable using below equation

$$Entropy(Target) = -\sum_{Calsse Target} p(Class)log2(p(class)$$

[3]. Calculate entropy of each attribute using below equation

$$E(Target, Attribute) = \sum_{Branch \epsilon Attribute} p(branch)E(branch)$$

$$where\ E(branch) = E(a1, a2, \dots, an) = -\sum_{i \epsilon n} \frac{ai}{\sum_{i=1}^{n} ai} log2(\frac{ai}{\sum_{i=1}^{n} ai})$$

[6]. Calculate Information Gain using below equation

$$IG(Target, Attribute) = E(Target) - E(Target, Attribute)$$

[7]. Choose attribute with highest information gain as decision node

[8]. Repeat step 2 to 7 unit complete tree formed with leaf node

Decision Tree Algorithm – Classification Problem – Gini Index:

[1]. Read Data, Identify Target Variable and Input Features (Attributes)

[2]. Calculate Gini Index of each attribute using below equation

$$Gini(Target, Attribute) = \sum_{Branch \epsilon Attribut} p(branch)Gini(branch)$$

$$where\ Gini(branch) = Gini(a1, a2, \dots, an) = 1 - \sum_{i \epsilon n} [\frac{ai}{\sum_{i=1}^{n} ai}]\text{^}2$$

3. Choose attribute with lowest Gini as decision node,

Repeat step 2 and 3 unit complete tree formed with leaf nodes