Gini Impurity

- Gini Impurity is a metric utilized in constructing Decision Trees to ascertain the optimal way of dividing nodes in the tree based on the features of the dataset.
- It is used in predicting the likelihood that a randomly selected example would be incorrectly classified by a specific node.
- A feature with the smallest Gini Impurity is selected for splitting the node.
- For a dataset '**D**' with samples from '**k**' classes, probability of a sample belonging to class '**i**' is denoted as $\mathbf{p_i}$. The gini impurity is given by the formula,

$$Gini(D) = 1 - \sum_{i=1}^{k} p_i^2$$