Find the majority of group labels,

use distance

K too small or too big, maybe cause unfit or overfit.

K-nearest neighbor:

Simple to understand why a given unseen record was given a particular class.

Cons:

Expensive to classify new points.

KNN can be problematic in high dimensions.

K close to N, it is always underfit.(N is the number of the datapoints in dataset.)

K close to 1, it is overfit, highly flexible and prone to learning the noise data.

Hunter algorithm

Split, based on the attribute,

Binary:

Income, divorce

Use income

The max of gini: is 0.5 , scope(0-0.5)