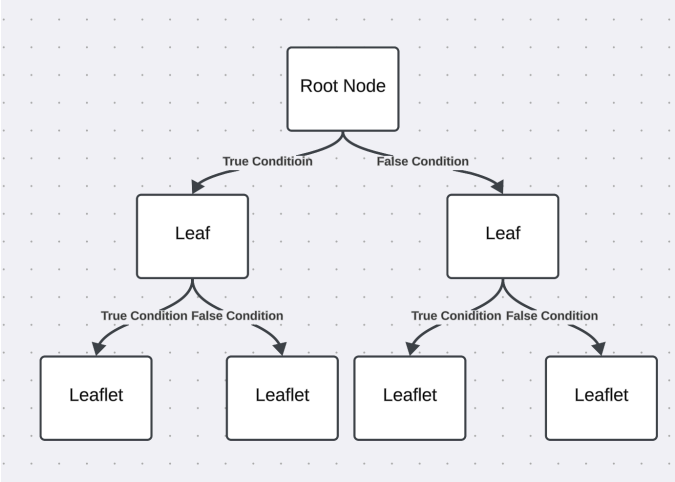
**Predicting Discrete Values**

**Classification**
Predictions based on categorical data, instead of numerical data.

**Use of Decision Tree Classification**
Sensitive to small variations in training features.




The difference between decision trees classification and decision tree regression is that the classification model predicts a category, on the other hand, decision tree regression is used to predict continuous numerical data.

**Naive Bayes**
Naive in the name is for the model's ability to consider individual words in a message, not the whole message. Naive Bayes excels in predicting data with words.

To featurize is to extract a message and featurize it into counts of individual words.

Find the probability of the label occurring in that dataset by taking the amount of times that the label occurs divided by the total number of entries in that dataset. That alone gives you the probability that a new entry will be that specific label, but Naive Bayes can hone in that probability by taking the probability of words occurring in that specific label and plugging it into an equation. You would take the probability of each word occurring in the specific label and divide it by the total number of the occurrences of other words in that label.

Data with words is best suited for Naive Bayes.