## Splitting on Impurity Metrics

In order to compare the performance of F-measure splitting to some baseline, we evaluated the same datasets with two more commonly used split-criteria, Information Gain and Gini Index.

### Information Gain

Information Gain is a measure of the entropy of a given split compared to the entropy of the un-split data. Each side of the split data is given an entropy according to the equation

Where is the number of labels in the split, and is the probability of a random example in the split having the label . The entropies of the two sides are then combined and Information Gain (IG) calculated according to[[1]](#footnote-1):

Where is the dataset of the parent node in the tree, f is the split function, is the number of examples in the parent dataset, , are the datasets resulting on each side of the split, and , are the number of examples on each side of the split.

### Gini Index

The Gini Index is an alternative measure Impurity which is similar to Information Gain but with some notable differences. Each side of the split data is given a Gini Index according to the equation

The impurities of the two sides are then combined and compared to the un-split data using the same method used in Information Gain.

Because Gini Index calculations don’t involve logarithms, they are slightly faster to compute, and a Gini Index is will always be in the range 0-1, regardless of how many labels exist in the dataset.

A more detailed discussion of the differences between Information Gain and Gini Index is described in [1][[2]](#footnote-2). For the purposes of this study, both Information Gain and Gini Index are used as comparators to splitting based on F Measure, such that the average performance of impurity metrics can be compared to the more results-based F Measure.

1. https://sebastianraschka.com/faq/docs/decision-tree-binary.html [↑](#footnote-ref-1)
2. L.E. Raileanu, K. Stoffel / Gini Index and Information Gain criteria (2004) https://www.unine.ch/files/live/sites/imi/files/shared/documents/papers/Gini\_index\_fulltext.pdf [↑](#footnote-ref-2)