Within a Gini score gives an idea of how good a split is by how mixed the classes are in the two groups created by the split. A perfect separation results in a Gini score of 0, whereas the worst case split that results in 50/50 classes. Gini index or Gini impurity measures the degree or probability of a particular variable being wrongly classified when it is randomly chosen.

**But what is actually meant by ‘impurity’?**

If all the elements belong to a single class, then it can be called pure. The degree of Gini index varies between 0 and 1, where 0 denotes that all elements belong to a certain class or if there exists only one class, and 1 denotes that the elements are randomly distributed across various classes. A Gini Index of 0.5 denotes equally distributed elements into some classes.

**After running the model, with a depth of 200 trees, the following observations can be made:**  
  
**(1)** From our sample it is hard to find delinquency at the early Tree stages. We had to go from a max depth of 3 to 5 to start seeing some delinquency nodes.  
  
**(2)** If the borrower credit score at origination is less than or equal to 657, and if the original interest rate is greater than 4.312%, there is a higher risk of delinquency.  
  
**(3)** If a coborrower credit score at origination is higher than 701.5 and the origin debt to income ratio is more than 15.5% while the borrower credit score is less than or equal to 733.5 then there is a higher risk of delinquency.

Under confusion matrix

The confusion matrix shows that non-delinquent loans are correctly identified nearly 100% of the time (these are true negatives), and incorrectly labels them as non-delinquent only 1% of the time (these are false positives).

For all of the delinquent loans in our dataset, we are able to correctly identify them 90% of the time (these are true positives), while our algorithm incorrectly misidentifies them 3% of the time (these are false negatives).

**Why is this important?** In terms of profitability to Fannie Mae, false negatives are the most important metric, because Fannie Mae loses money when the model incorrectly labels a delinquent loan as being a non-delinquent.