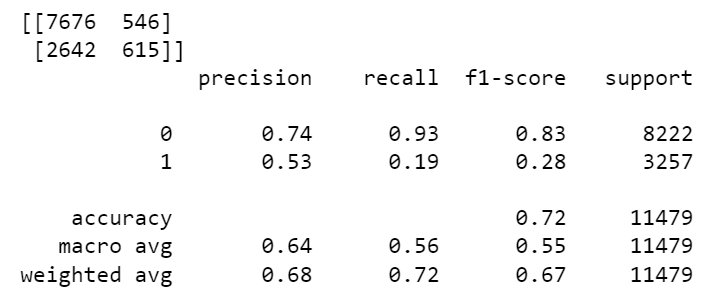
Notes:

* Decision trees can also be used to predict continuous values, though mostly binary
* Where to split a decision tress is most important
* Overfitting is the toughest problem here
* Solution:
  + Set constraints: max depth, min depth, features etc.
  + Pruning: remove subtrees etc. (minimum error)
* Gini index: used to split the data. Higher the value better the split. Good for categorical data.
* Chi square: used to split data. Good for categorical data. Works for multiple splits.
* Information entropy: how much variance is in the data
  + Used to understand information gain
* Lower entropy is better
* Root node: algorithm select best attribute to split record first
* Child node
* Decision trees are prone to overfitting
* Hyperparameter tuning
* This model clearly performs better than logistic regression



* Using GridSearch tells us the best parameters to use
* Bagging and boosting are ensemble methods to combine several decision trees
* Decision trees suffer from high variance
* Bagging solves for high variance
* Random forests are an extension of bagging
* Boosting fits consecutive decision trees, reduces error with each