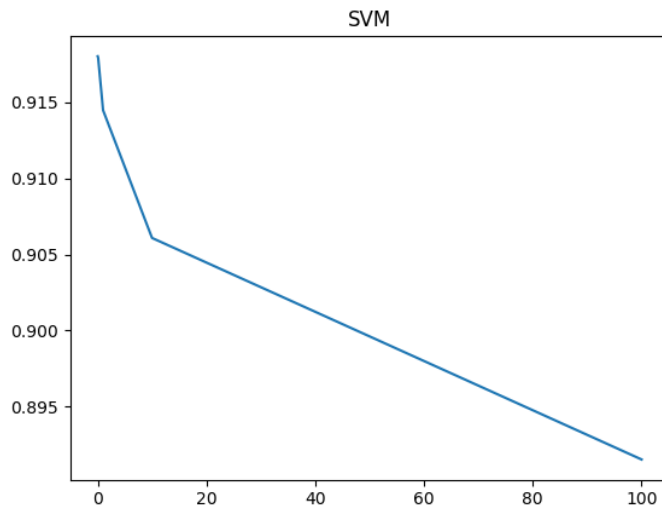**Q1) Discuss your observation. Is smaller or larger margin better for this dataset (need to explain which C values are likely to produce smaller v.s. larger values and then which end up being better in cross validation**
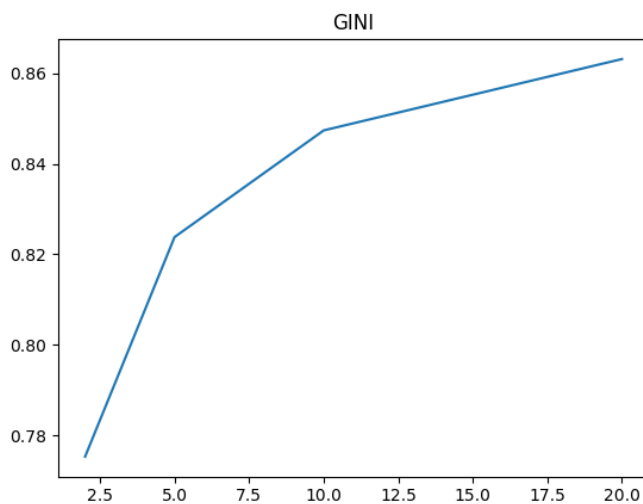
Answer-1

F-Measure is maximum for C=0.01 so therefore, the smaller the C the better the Margin (in our case).
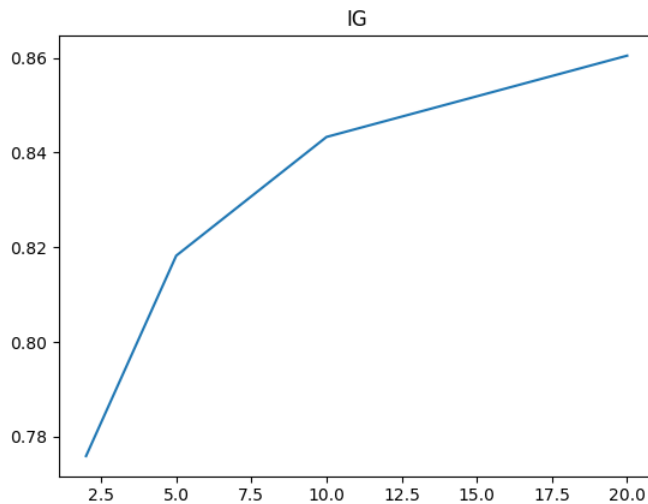
As the C values decrease the margin increases and for our case f measure also increases.

Q-2) Discuss your observations from the figures. Does larger tree mean better F-measure? Which criterion is better?

For Decision Tree with Criterion Gini maximum F-Measure is for K=20.

Therefore, Larger the tree better the F-Measure for GINI.

For Decision Tree with Criterion Information Gain maximum F-Measure is for K=20.

Therefore, Larger the tree better the F-Measure for Information GAIN.

Criterion Gini is better since it has higher F-measure.

Q3) Discuss your findings. Which are the best classifiers when you consider the different metrics? Is there a single winner for this dataset.

Classifiers V/S Precision


Classifiers V/S Recall
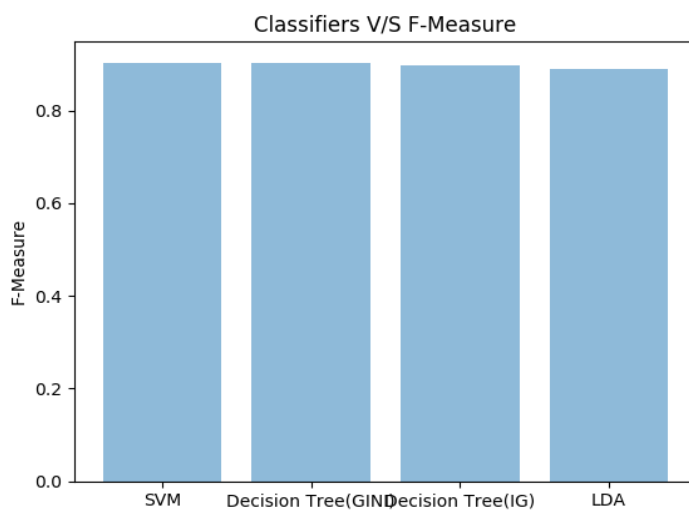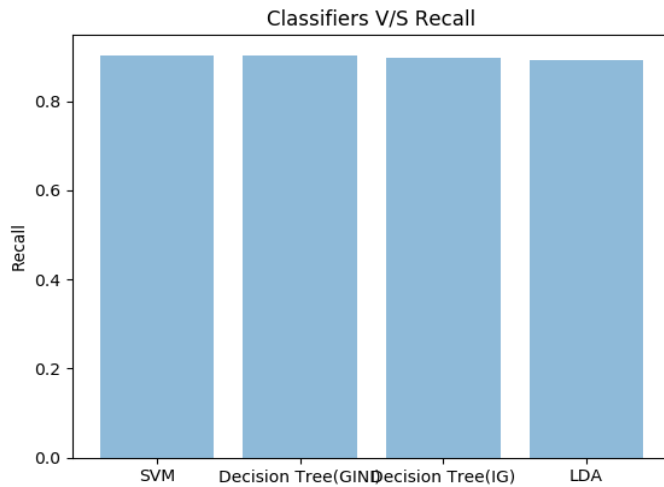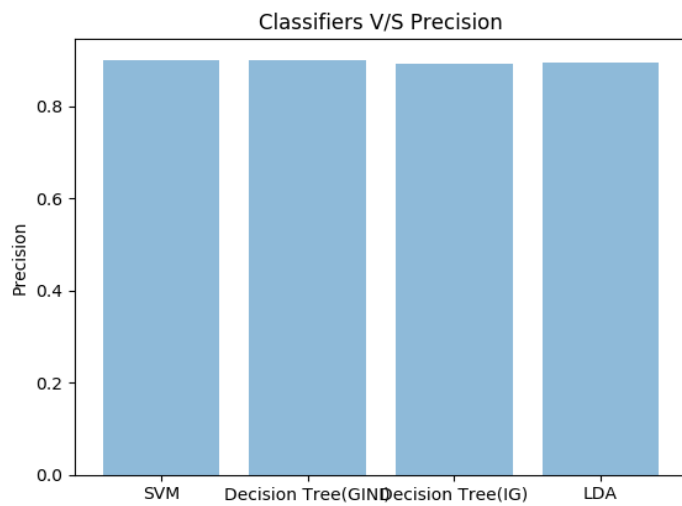
For F-Measure: SVM is proved to be the best classifier.

For Precision: SVM is proved to be best classifier.

For Recall: SVM is proved to be the best classifier.

Therefore, SVM is the single winner.

Q-4) Consider a classifier that is implemented in scikit-learn but we have not discussed in class: RandomForestClassifier. Learn how to create and train it in scikit-learn and add it to the comparison from the previous part (c). Explain the results.

Classifiers V/S Precision

Classifiers V/S F-Measure

Classifiers V/S Recall
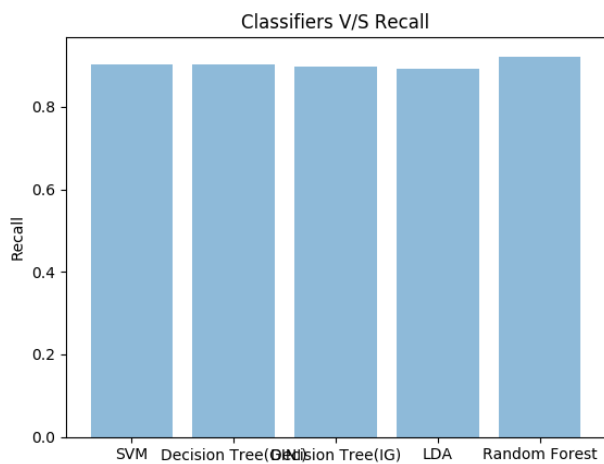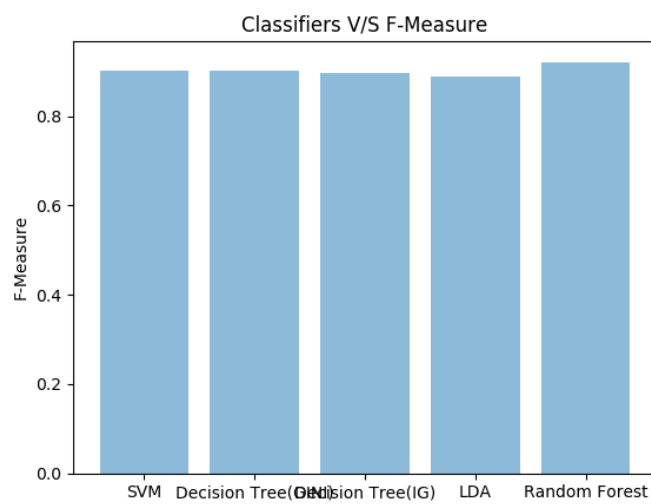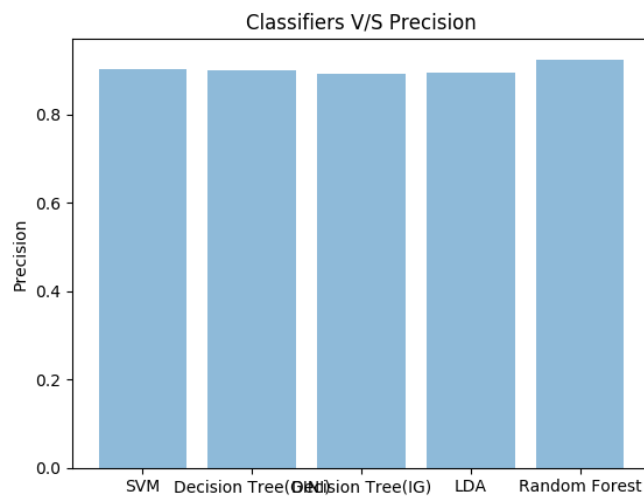
For all the measures RandomForestClassifier is the winner.

Therefore, RandomForestCLassifier is the best classifier for our dataset.