F1 score with test set from Driven Data
for different methods and data