F1 score evaluation

