Tfidf based similarity SVM + linear

