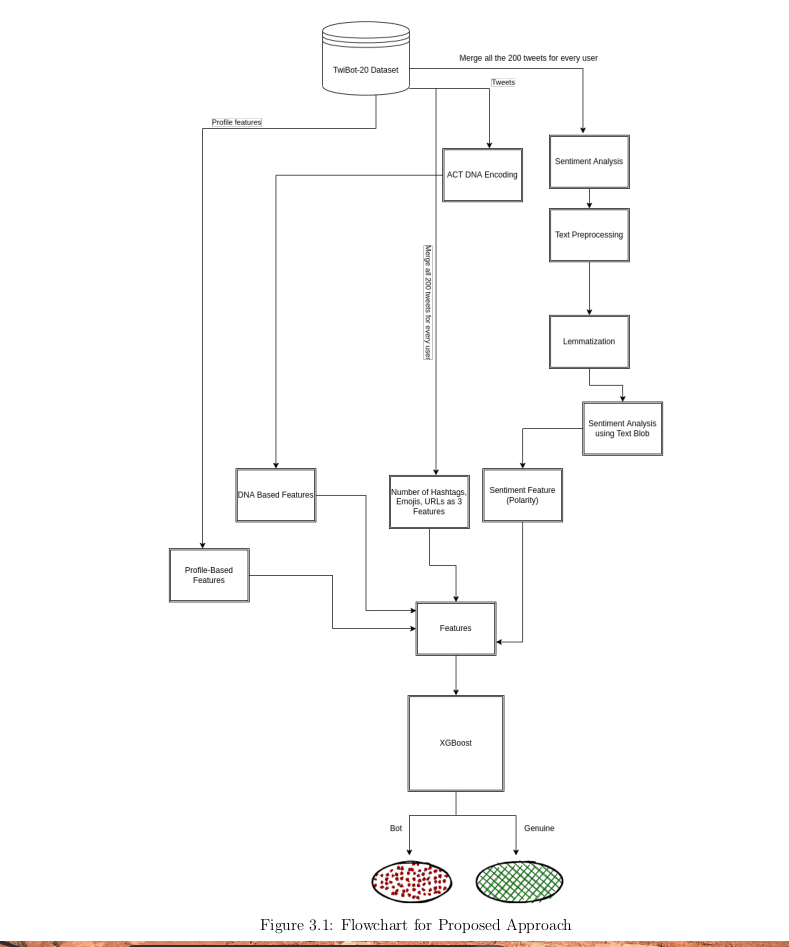
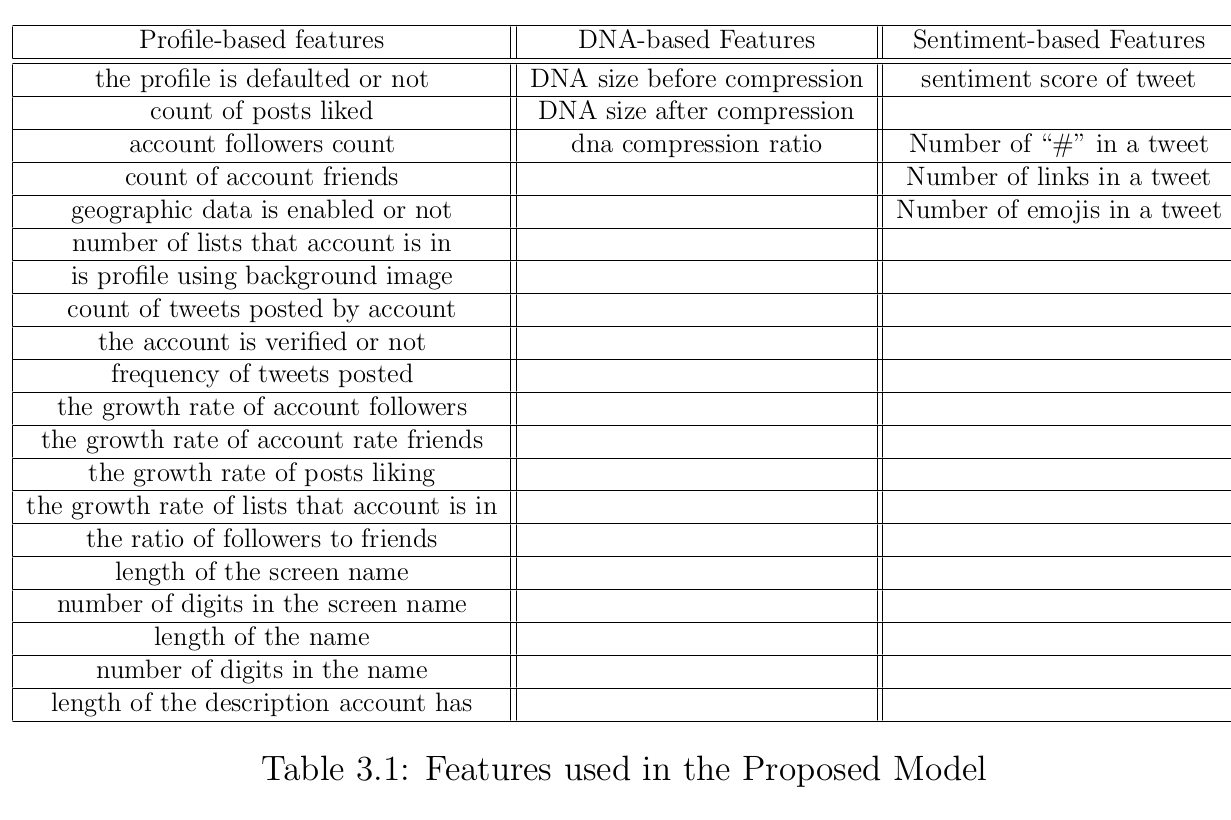
Detection of Social Bots in Twitter Network

Problem Statement: Given a set of Twitter user accounts, classify each user account into a gen-uine(human) or bot account.

Approach: Take profile-based feature, along with tweet sentiment feature, and DNA-based features, to feed into XGBoos for classifying each user account into genuine or bot. TwiBot-20 is the dataset used for model training and inference which contains 22K bot user accounts and 17K genuine user accounts. Tweets are converted into dna sequence A,C,T where T stands for a tweet that starts with “RT”, C- Tweet that has “@” at the start and A is another normal tweet. This sequence is then compressed using lossless compression algorithm zlib to get uncompressed/compressed sizes to be used as an input to the model, along with sentiment score of tweet.



Features:



Top 3 features: Verified profile?, geographic data available?, defaulted?