Two Stage competition: What that means: we will release a new, unseen dataset 7 days before the final submission deadline and your scores will be evaluated with that data. For this competition, you will NOT be required to upload your model before Stage 2.

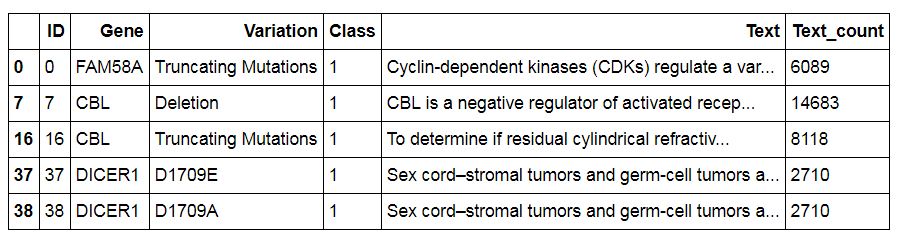
tagging genes:

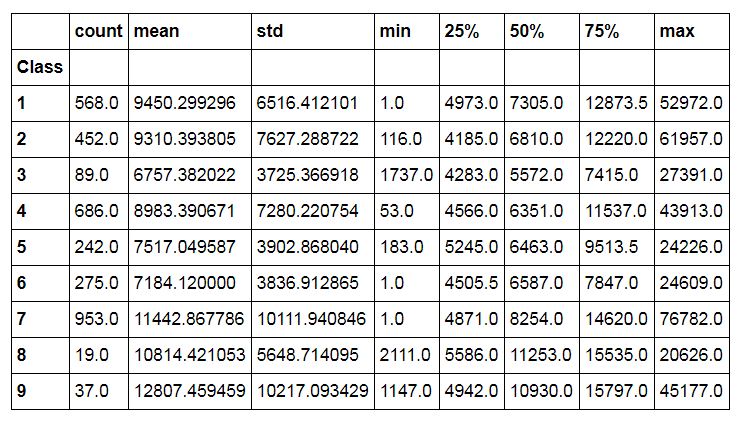
pre-trained Bio-NER taggers:

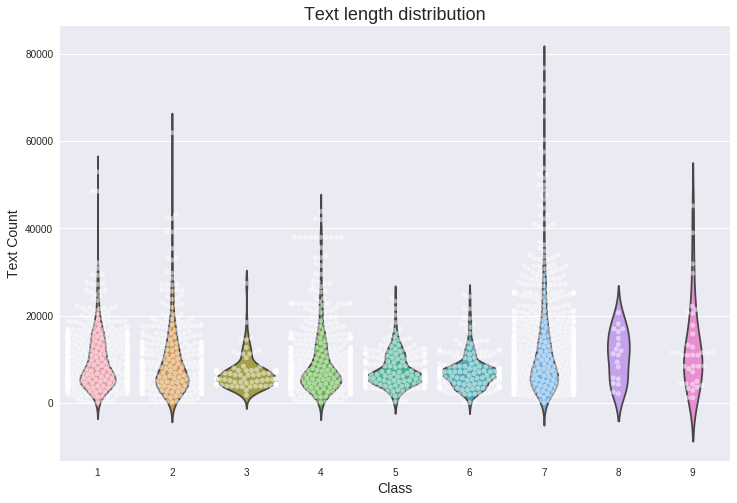
tmVar https://www.ncbi.nlm.nih.gov/research/bionlp/Tools/tmvar/

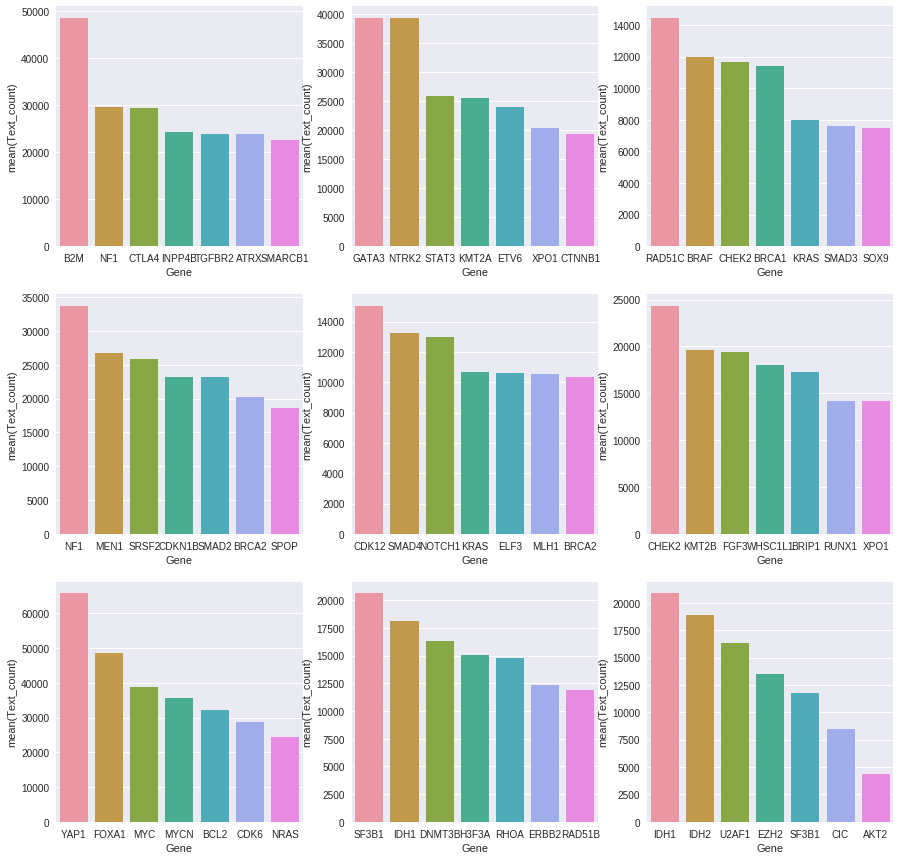
TaggerOne https://www.ncbi.nlm.nih.gov/research/bionlp/Tools/taggerone/

GNormPlus https://www.ncbi.nlm.nih.gov/research/bionlp/Tools/gnormplus/





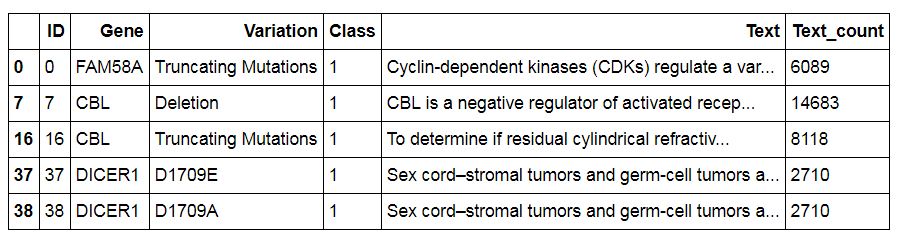




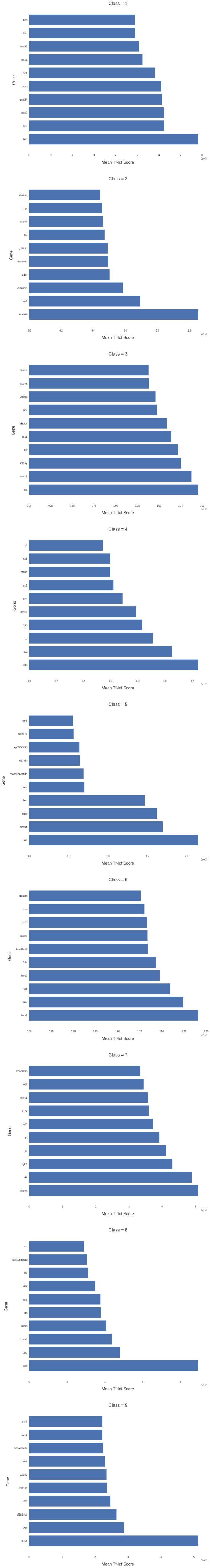
BRCA1 is highly dominating Class 5

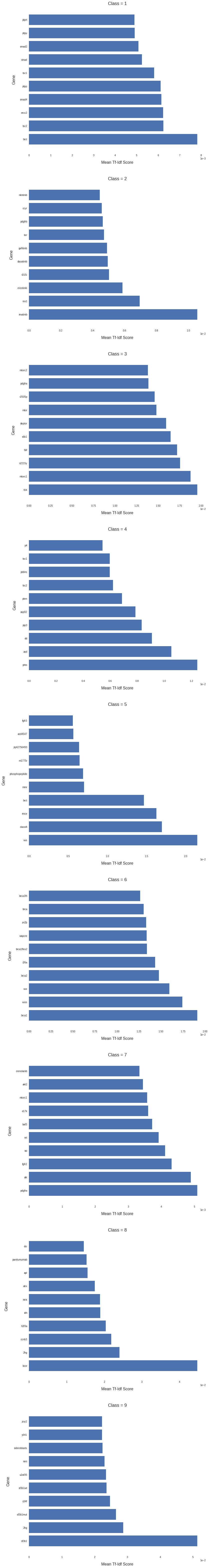
SF3B1 is highly dominating Class 9

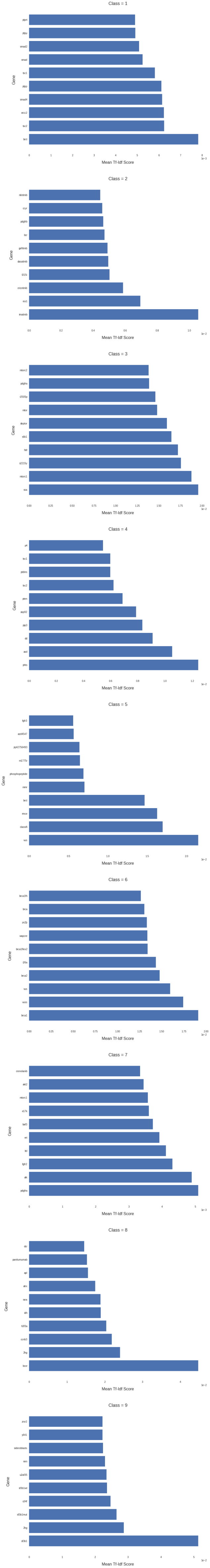
BRCA1 and BRCA2 are dominating Class 6



In information retrieval, tf–idf, short for term frequency–inverse document frequency, is a numerical statistic that is intended to reflect how important a word is to a document in a collection or corpus. It is often used as a weighting factor in information retrieval, text mining, and user modeling.







linear SVM