

Table 1: Cross-validation accuracy. Ibsen corpus

Features	Dimension	SVC	Naïve Bayes	Decision Tree	Logistic Regression
bigrams	78246	0.6467	0.8000	0.8170	0.9656
bigrams pos	231	0.7873	0.6170	0.6770	0.8172
bigrams pos punct	730	0.9746	0.9958	0.9917	0.9958
bigrams pos punct tf	730	0.9746	0.9958	0.9746	0.5446
bigrams pos punct tfidf	730	0.9746	0.9958	0.9746	0.5446
bigrams pos tf	231	0.8045	0.7870	0.7489	0.5446
bigrams pos tfidf	231	0.8045	0.7784	0.7489	0.5446
bigrams punct	69168	0.8380	0.9571	0.9957	0.9913
bigrams punct tf	69168	0.8422	0.9571	0.9913	0.5446
bigrams punct tfidf	69168	0.8422	0.9871	0.9913	0.5446
bigrams tf	78246	0.7150	0.8000	0.8078	0.5446
bigrams tfidf	78246	0.7150	0.8889	0.8078	0.5446
cohesive	124	0.6944	0.5408	0.6636	0.7404
cohesive punctuation	897	0.8380	0.8723	0.8810	0.9105
cohesive punctuation tf	897	0.8380	0.8723	0.8900	0.5446
cohesive punctuation tfidf	897	0.8380	0.9025	0.8900	0.5446
cohesive tf	124	0.6775	0.5366	0.7393	0.5446
cohesive tfidf	124	0.6775	0.5493	0.7393	0.5446
syntactic n2	89117	0.8034	0.9567	0.9400	0.9913
syntactic n2 tf	89117	0.7092	0.9567	0.9616	0.5399
syntactic n2 tfidf	89117	0.7092	0.9784	0.9616	0.5399
syntactic n3	270419	0.4851	0.8929	0.9272	0.9274
syntactic n3 tf	270419	0.4851	0.8929	0.8928	0.5399
syntactic n3 tfidf	270419	0.4851	0.9574	0.8928	0.5399
trigrams	157088	0.5404	0.8000	0.7531	0.7612
trigrams pos	2526	0.7649	0.5625	0.6083	0.7958
trigrams pos punct	6609	0.9612	0.9958	0.9578	0.9957
trigrams pos punct tf	6609	0.9697	0.9958	0.9708	0.5446
trigrams pos punct tfidf	6609	0.9697	0.9958	0.9708	0.5446
trigrams pos tf	2526	0.8083	0.6051	0.8214	0.5446
trigrams pos tfidf	2526	0.8083	0.6982	0.8214	0.5446
trigrams punct	156913	0.5616	0.8895	0.9576	0.9364
trigrams punct tf	156913	0.5616	0.8895	0.9234	0.5446
trigrams punct tfidf	156913	0.5616	0.9701	0.9234	0.5446
trigrams tf	157088	0.5446	0.8000	0.8134	0.5446
trigrams tfidf	157088	0.5446	0.8337	0.8134	0.5446
unigrams	10025	0.8466	0.8897	0.9234	0.9828
unigrams punct	10065	0.8636	0.9489	0.9915	0.9957
unigrams punct tf	10065	0.8931	0.9489	0.9915	0.5446
unigrams punct tfidf	10065	0.8931	0.9788	0.9915	0.5446
unigrams tf	10025	0.8679	0.8897	0.8511	0.5446
unigrams tfidf	10025	0.8679	0.9404	0.8511	0.5446

Table 2: Cross-validation accuracy. Quixote corpus

Features	Dimension	SVC	Naïve Bayes	Decision Tree	Logistic Regression
bigrams	296924	0.6296	0.3967	0.8649	0.9763
bigrams pos	226	0.8281	0.4439	0.4812	0.8464
bigrams pos punct	763	0.9868	0.8680	0.9683	0.9920
bigrams pos punct tf	763	0.9868	0.8680	0.9789	0.2699
bigrams pos punct tfidf	763	0.9868	0.8680	0.9789	0.2699
bigrams pos tf	226	0.8202	0.4385	0.5342	0.2699
bigrams pos tfidf	226	0.8202	0.4332	0.5342	0.2699
bigrams punct	271925	0.8121	0.6401	0.9419	0.9841
bigrams punct tf	271925	0.8121	0.6401	0.9576	0.2699
bigrams punct tfidf	271925	0.8121	0.9393	0.9576	0.2699
bigrams tf	296924	0.6323	0.3967	0.8622	0.2699
bigrams tfidf	296924	0.6323	0.8441	0.8622	0.2699
cohesive	142	0.8622	0.7488	0.7907	0.9179
cohesive punctuation	1210	0.9658	0.8679	0.8227	0.9735
cohesive punctuation tf	1210	0.9632	0.8679	0.8464	0.2699
cohesive punctuation tfidf	1210	0.9632	0.9023	0.8464	0.2699
cohesive tf	142	0.8701	0.7488	0.8810	0.2699
cohesive tfidf	142	0.8701	0.7488	0.8810	0.2699
syntactic n2	420400	0.6747	0.4496	0.8732	0.9683
syntactic n2 tf	420400	0.6668	0.4496	0.8732	0.2699
syntactic n2 tfidf	420400	0.6668	0.8943	0.8732	0.2699
syntactic n3	1870070	0.2169	0.1269	0.8521	0.9102
syntactic n3 tf	1870070	0.2115	0.1269	0.8548	0.2699
syntactic n3 tfidf	1870070	0.2115	0.6481	0.8548	0.2699
trigrams	748772	0.1243	0.0501	0.8703	0.8783
trigrams pos	2838	0.8413	0.4369	0.5343	0.8491
trigrams pos punct	8416	0.9947	0.9473	0.9656	0.9868
trigrams pos punct tf	8416	0.9947	0.9473	0.9656	0.2699
trigrams pos punct tfidf	8416	0.9947	0.9736	0.9656	0.2699
trigrams pos tf	2838	0.8492	0.4395	0.5660	0.2699
trigrams pos tfidf	2838	0.8492	0.5111	0.5660	0.2699
trigrams punct	736288	0.4713	0.5269	0.9524	0.9763
trigrams punct tf	736288	0.4765	0.5269	0.9524	0.2699
trigrams punct tfidf	736288	0.4765	0.8995	0.9524	0.2699
trigrams tf	748772	0.1270	0.0501	0.8703	0.2699
trigrams tfidf	748772	0.1270	0.3889	0.8703	0.2699
unigrams	22401	0.9604	0.8072	0.9550	0.9815
unigrams punct	22442	0.9604	0.8126	0.9603	0.9867
unigrams punct tf	22442	0.9604	0.8126	0.9603	0.2699
unigrams punct tfidf	22442	0.9604	0.9208	0.9603	0.2699
unigrams tf	22401	0.9604	0.8072	0.9523	0.2699
unigrams tfidf	22401	0.9604	0.9208	0.9523	0.2699