

	Naive Bayes		Decision Tree
Feature Set	Accuracy	# of features	Accuracy
(Unigrams + Bigrams + POS-uni + POS-bi) with binning + liwc_scores no binning + len of review	0.76	32768	0.49
(Unigrams + Bigrams) with binning + liwc_scores no binning + len of review	0.79	16384	0.48
Unigrams with binning + len of review	0.76	128	0.51
(Unigrams + Bigrams) with binning + len of review	0.78	4096	0.48
(Unigrams + Bigrams) with binning	0.80	32768	0.78
(Unigrams + Bigrams) with binning +LIWC set 1 no binning	0.81	16384	0.69
(Unigrams + Bigrams) with binning + LIWC set 1 with binning	0.81	256	0.66
(Unigrams + Bigrams) with binning + LIWC set 1 + 2. any LIWC feature is 1 when liwc_scores > 0 and 0 otherwise	0.80	32	0.71
(Unigrams + Bigrams) with binning + LIWC set 1 +2 with binning	0.81	128	0.72
<b>(Unigrams + Bigrams) as count of the words + LIWC set 1 and 2 with binning</b>	<b>0.83</b>	<b>32</b>	<b>0.78</b>
(Unigrams + Bigrams) as count of the words + LIWC set 1 +2 with binning + positive_score*	0.83	32	0.72
(Unigrams + Bigrams) as count of the words + LIWC set 1 +2 as liwc_scores no binning	0.83	128	0.58

positive\_score\*

if a unigram is in ["great", "best", "perfect", "friendly", "nice", "comfortable"]

positive\_score += 1

if a unigram is in ["wonderful", "excellent"]

positive\_score += 2

feature\_vector[fname] = positive\_score

I have chose the combination

(Unigrams + Bigrams) as count of the words + LIWC set 1 and 2 with binning

as the best since it gives best results for NB and DT and takes only 32 features.