Results for Sentiment Analysis

The following table depicts the performance of
1. Naive bayes classifier.
2. Naive bayes classifier with stopwords removal.
3. Binary naive bayes classifier.

over a movie review test data set containing 12,500 positive movie reviews and 12,500 negative movie reviews, trained using 12,500 positive movie reviews and 12,500 negative movie review examples.

	Accuracy (%)	Avg Exec Time
Naive Bayes	81.444	6.54 s
Binary NB	81.644	5.16 s
NB/stopwords removal	82.25	5.43 s

	Positive Recall (%)	Positive Precision(%)	Positive fMeasure (%)
Naive Bayes	74.976	86.115	80.16
Binary NB	73.576	87.73	80.03
NB/stopwords removal	76.224	86.67	81.11

	Negative Recall (%)	Negative Precision(%)	Negative fMeasure (%)
Naive Bayes	87.912	77.84	82.56
Binary NB	88.28	78.78	83.25
NB/stopwords removal	89.71	77.24	83.00

Comparision of naive bayes Learning with its enhancements
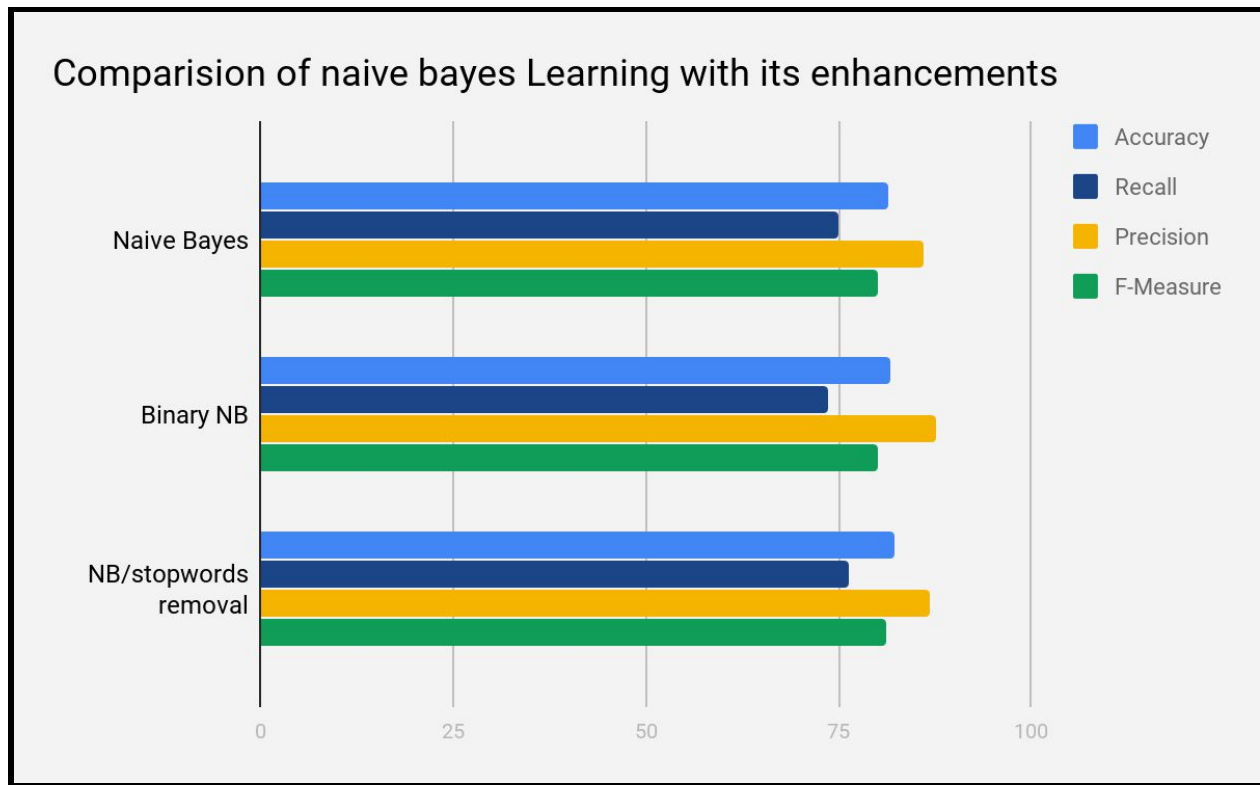
Effect of removing stopwords

- 1% improvement in accuracy in comparison to simple naive bayes learning.

- Common english terms like 'They', 'we', 'I' etc. are neutral words and do not contribute in assigning a sentiment to a query.

- Some class (here Positive or negative) might end up having having a lot more of the common English terms than the other class.This can potentially draw the bayesian classifier into selecting the class with a higher occurrence of the common English words.

- Removing stopwords reduces memory overhead since our vocabulary size becomes smaller.

- Prediction thus become faster and more accurate since we are eliminating noise and distracting features, in terms of stop words.

Prediction performance using binary naive bayes learning

- 0.24% improvement in accuracy in comparison to simple naive bayes learning.

- Rare words in a movie review primarily decide the sentiment of the review. Thus, if a particular rare word is used quite often in a movie review, the sentiment can sway more strongly depending on the quality of the rare word.

  However, a particular rare word is generally not used twice in a product or a movie review. Thus binary naive bayes learning fails to have a significant effect on our movie review data. Consequently, the improvement in prediction with respect to performance of simple naive bayes learning, is low.