Evaluation

To evaluate our sentiment analysis model, we have manually annotated each tweet as positive, negative or neutral. These 2000 annotated tweets are then classified on the topics given by the LDA algorithm. Now we compare the sentiment of each tweet to the sentiment predicted by Hindi wordNet baseline implementation and random forest algorithm.

The baseline wordNet classification gave us an accuracy of 62.37%. The system performs poorly because of spelling errors in the tweets as well as the spelling errors introduced post transliteration. Many tweets had words which were not present in the wordnet or the words with neutral polarity. On using stemming on the tweets the accuracy of this classifier increases by 2.3% i.e. 64.67%. This classifier was then applied to the complete tweet dataset.

The second classifier that we built was using random forests by using 1500 tweets to train and 500 tweets as test data, improved the accuracy of our sentiment classifier to 70.69%; and using stemming further improved our accuracy by 1.73%.

To evaluate our LDA topic model and the changing sentiment on topics over time, we cross verified the topics/keywords generated by the model with the trends published on twitter and also on google news. We saw overlap of our topics and some of the hashtags trending on Twitter. We also verified that topics like "Sonu Nigam's comment on Azan", "Gaurakhsaks of UP", "Mahaveer Jayanti", "Gurunanak Jayanti", "Hanuman Jayanti", and others were among the most discussed topics on Indian news channels/websites and were included in stories on Inshorts news application which provides a 60 words summary of each news item trending during the day.