Discussion

Noise: The tweets contain a mixture of hinglish and hindi words and also some english words written in hindi, this deteriorates the performance of our model. Also tweets also contain a lot of slang words which are not present in the word net.

Morphological Variations- Handling morphological variations is a big challenge for Hindi language. Hindi language is morphologically rich i.e. a lot of information is fused in the words as compared to the English language where we add another word for the extra information. With same root there can be many words in a language with varying information i.e multiple variations of same words can have the same root with respect to the sense of tense, gender, person and other information

Spelling Variations- In Hindi the same word with same meaning can have multiple spellings, thus it's not quite difficult to have all the occurrences of such words in a corpus and even while training a model it's quite complex to handle all the spelling variants.

Lack of resources-Lack of sufficient resources, tools and annotated corpora adds to the challenges while addressing the problem of sentiment analysis when dealing with indian languages.

This experiment also shows the limitation of the Hindi Senti-Wordnet specifically for twitter data. This is largely due to limited words present in the word net and poor overlap of the words present in the tweet and the wordnet due to spelling differences, mixture of english - hindi words i.e. very few tweets use pure hindi, and slang word usage.

The above factors combined make the task of sentiment analysis on hindi very difficult, but as more and more resources become available, especially hand annotated tweets we believe this will lead to better sentiment prediction models as shown by the random forest classifier. Thus with the available of more hand annotated data the machine learning models will be able to learn better by relying of the inherent properties of the tweets rather than word nets. We hope that our dataset will be a valuable contribution in this regards.