SENTIMENTAL ANALYSIS

As We Know

- Sentiment analysis or opinion mining is the automated extraction of writer’s attitude from the text
- It is one of the major challenges in natural language processing

What we have done

- We took a Hotel review dataset with three levels of emotions tagged(Positive , Negative, Neutral) for our model.
- Got started with pre-processing of reviews like removing stop words.
- Applied Machine learning Techniques to solve natural text processing problems.
- Models included Bag of words methods like SVM, Naive Bayes and alterations of that.
- We also used some pre-defined database of Tagged words for model.

What we have achieved

- Were able to distinguish positive , negative and neutral reviews
- Overall accuracy of SVM model was higher with around 70%
- Model correctly distinguished positive and negative reviews to good extent.
- Accuracy to determine neutral reviews was less.

Challenges

- On real we have dozens of Sentiments to be tagged with texts.
- Complexity of Texts , identifying different figure of speeches like sarcasm , metaphor etc.
- Finalizing model which solves our project problem.
- Improving overall accuracy of model.
- Model should be fast as well.

What we can do more

- Better pre-processing of input texts like for modeling for increasing accuracy and speed. Better pre-processing include –

1. Classification of Input texts into various category to which models can be applied. For example categorizing the data to different domains “ Market”,” Economy”,” Industry”, ” customers” , ”Employee”.
2. Better Feature Selection like wisely removing stopwords , removing punctuations and stemmimg the document such that it don’t effect .

- Accuracy can be increased by using different model apart from bag of words like part of speech tagging which itself is used by Google Search Engine.
- Various other Sentiment extraction methods like Logistic regression , K-neighbour Clustering,Trees can be applied according to their effects.
- On-whole We have to come up with a final model which may be combinations of many specified which is more accurate and fast.
- Since Machine learning is now towards Deep-learning .We can think of that too also!!