

## Sentiment Classification of Documents

**Authors :** Swarnadip Ghosh

**Abstract :** Sentiment Analysis is the process of detecting the contextual polarity of text. In other words, it determines whether a piece of writing is positive, negative or neutral. Sentiment analysis of documents holds great importance in today's world, when numerous information is stored in databases and in the world wide web. An efficient algorithm to illicit such information, would be beneficial for social, economic as well as medical purposes. In this project, we have developed an algorithm to classify a document into positive or negative. Using our algorithm, we obtained a feature set from the data, and classified the documents based on this feature set. It is important to note that, in the classification, we have not used the independence assumption, which is considered by many procedures like the Naive Bayes. This makes the algorithm more general in scope. Moreover, because of the sparsity and high dimensionality of such data, we did not use empirical distribution for estimation, but developed a method by finding degree of close clustering of the data points. We have applied our algorithm on a movie review data set obtained from IMDb and obtained satisfactory results.

**Keywords :** sentiment, Run's Test, cross validation, higher dimensional pmf estimation

**Conference Title :** ICSEM 2015 : International Conference on Statistics, Econometrics and Mathematics

**Conference Location :** Paris, France

**Conference Dates :** January 23-24, 2015