**Text Mining**

**Extracting IMDB Movie Reviews**

**Movie Name-The Godfather**

**Latent Dirichlet Allocation**

Topic 1 Topic 2 Topic 3 Topic 4 Topic 5 Topic 6

[1,] "movie" "film" "movie" "movie" "film" "film"

[2,] "the" "the" "the" "the" "the" "movie"

[3,] "film" "godfather" "godfather" "film" "godfather" "the"

[4,] "godfather" "movie" "good" "godfather" "family" "good"

[5,] "films" "films" "movies" "this" "films" "godfather"

Topic 7 Topic 8 Topic 9 Topic 10

[1,] "film" "film" "the" "film"

[2,] "family" "the" "film" "movie"

[3,] "movie" "godfather" "godfather" "the"

[4,] "the" "family" "movie" "godfather"

[5,] "godfather" "movie" "michael" "great"

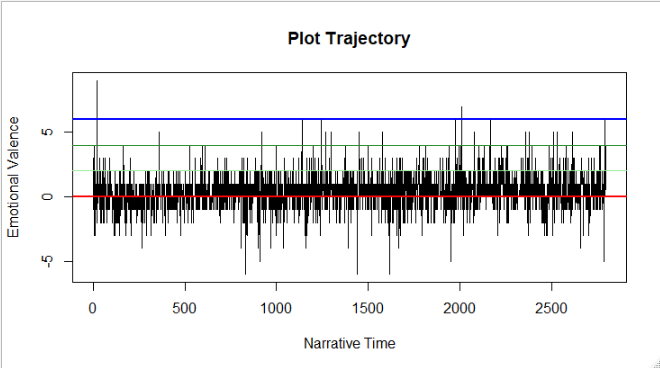
**10 topics with 5 terms each**

**Sentiment Analysis**

**Six methods for sentiment analysis**

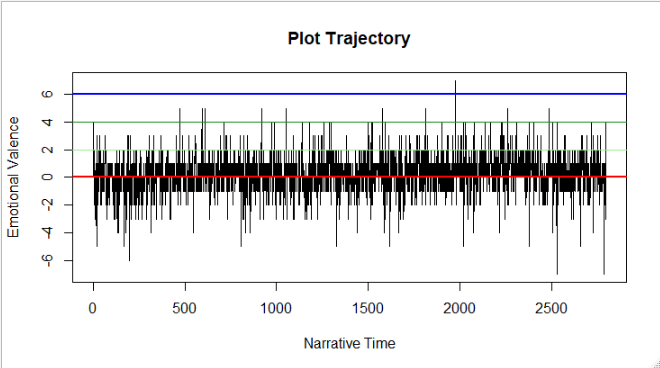
**syuzhet , afinn, bing, nrc, Stanford, custom**

**Analyzing using “nrc”**

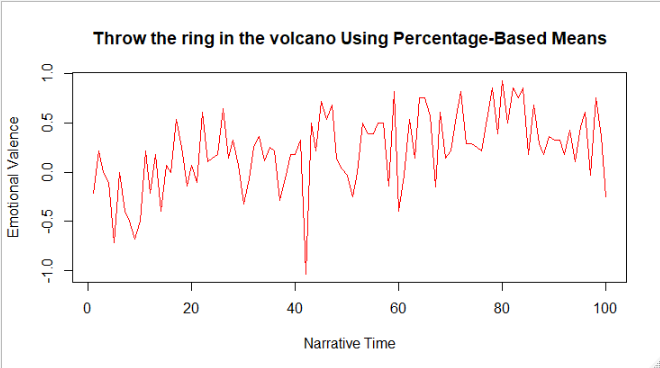


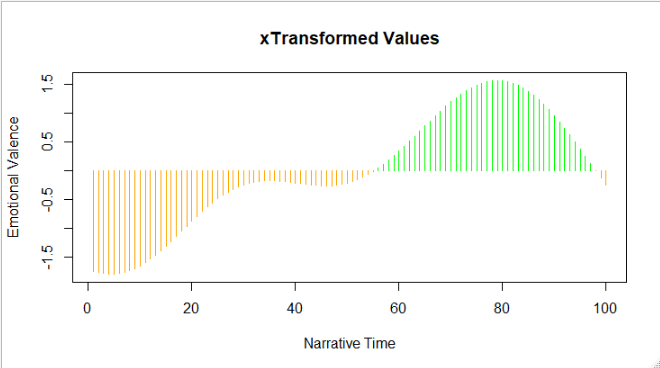
**From the above plot we can infer that,extreme positive reviews cross the limit of +7 ,where as extreme negative reviews cross -5.**

**Analyzing using “bing”**

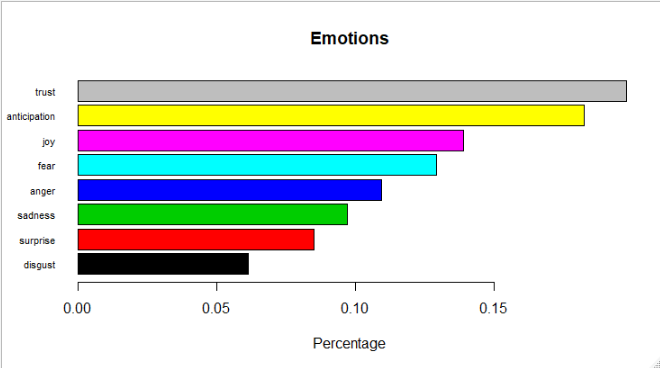


Based on above plot, positive and negative reviews are equally likely classified.



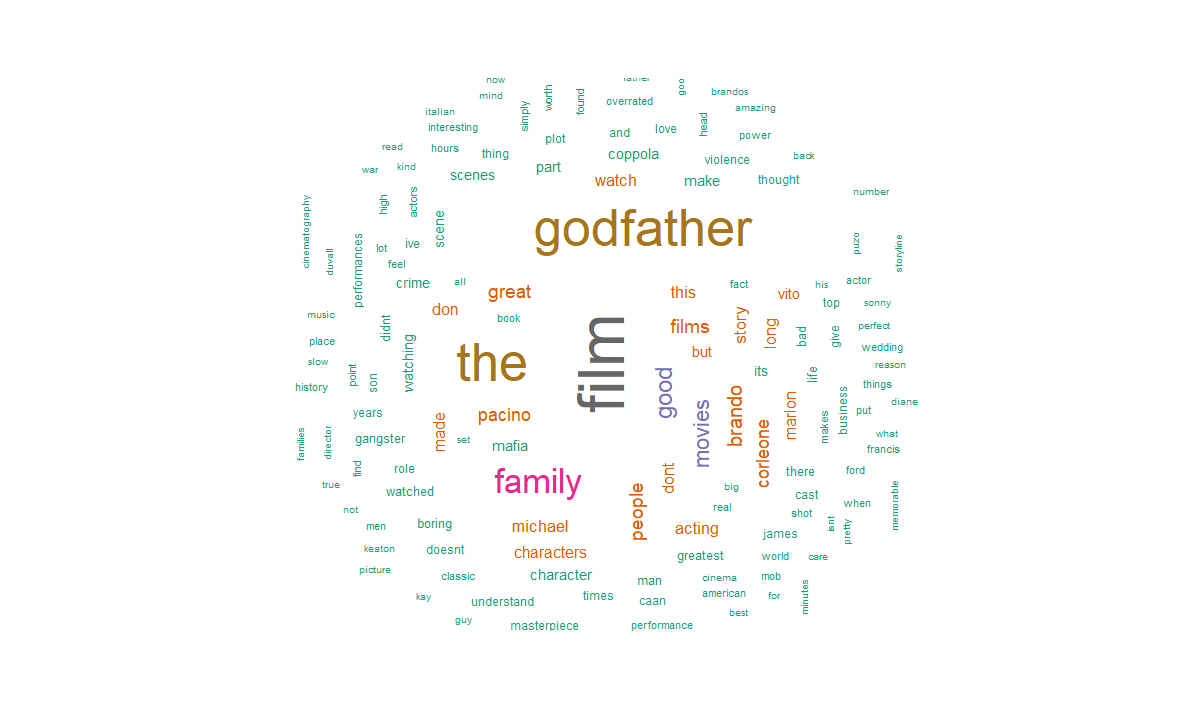


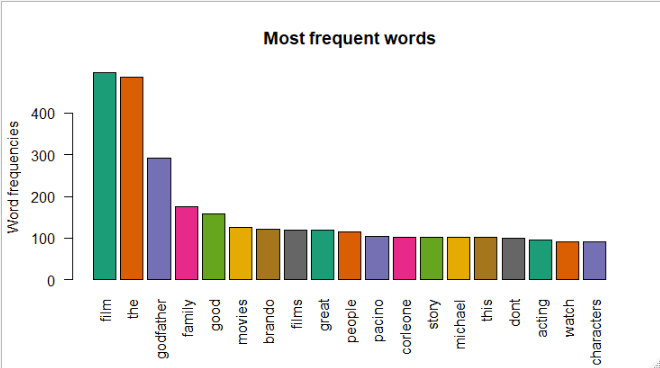
**Emotion Analysis**



**From the 8 emotions trust and anticipation are more in the reviews.**

**Word Cloud**





From the word cloud and bar plot, apart from movie, godfather, family, good movie is repeated most of the time. Also, we can comment that the godfather is good family movie.