Predicting Movie Ratings Based on Reviews

**Literature Survey**:

<http://aclweb.org/anthology/Y07-1050> : In this paper they have conducted various experiments on different classifiers like SVM, Maximum Entropy and Scoring which resulted in poor performance. Then they have collaborated all the three models which significantly increased the performance.

**Data Extraction:**

* Download the IMDB dataset from Kaggle (<https://www.kaggle.com/orgesleka/imdbmovies/version/1>)
* Perform Data Cleaning (missing attributes/ misplaced data)
* From the CSV file we can extract the movie title id for each given movie.
* Frame a url [https://www.imdb.com/title/**tt0032976**/reviews](https://www.imdb.com/title/tt0032976/reviews)

For example, title id of movie Rebecca is **tt0032976**.

* Using web scrapping, we can retrieve the User Reviews section from the above mentioned url.

**Problem Categorization:** It is a classification problem for review classification and regression to predict movie ratings.

**Success Metric:** Accuracy, precision, recall, f1 score

**Feature Extraction:** Review to TFIDF vector.  
**Model Selection/ Choice:** Based on empirical experiments

**Validation and Testing:** We can tune the parameters based on the validation set and calculate accuracy on the test data