**Project Title**

MovieBuzz

**Team information: member names.**

Sumesh Nellemakkal Balan

Ismail Wandeliwala

**Objective and overview of the project.**

IMDB – isolated movie community

Twitter – a general community

Analyze movies based on both the specialized IMDB data and general twitter data.

The project can be used to derive a relation between public opinion about a movie and its Box-office collection, Budget, rating etc. We will be using both IMDB data and twitter APIs to achieve the desired visualization of cross plotting of various details like trending of Movie, Director, Actor etc. against the movies rating in IMDB, Box office collection, Budget etc. Our aim is to mine both the data sources and find out some difference in the way movies are perceived by the general public and imdb users.

**What datasets you will use and how/where to acquire the datasets?**

IMDB data can be obtained from the ftp sites

http://www.imdb.com/interfaces/

<ftp://ftp.fu-berlin.de/pub/misc/movies/database/>

We will use the open source project – “ IMDBPy (GPL-2 License)” – to access IMDB data.

<http://imdbpy.sourceforge.net/>

Twitter APIs

**What do you plan to deliver at the end of the semester? How would your demo and system look like?**

We will deliver a web application, which can be used to compare two movies between its public opinion and IMDB rating, box-office collection, budget etc. The application will mine the data from twitter and show the analysis of the publics opinion in favor/against IMDB rating.

Given a movie name, the application will mine the data returned from twitter APIs and IMDB database and plot visualizations describing attributes such as IMDB rating, Box-office collection, Budget and Twitter Data. The visualization will also show the relation between public’s opinion (based on a twitter data) and movie rating, budget, collection etc.

**What are the challenges in this project? Why is it significant?**

The main challenge is in identifying the mood of the twitter data , ie Predicting positive or negative opinion by analysis of twitter data

Often the specific community ratings, such as IMDB, may not reflect the real opinion of the general public as relatively very small classes of people use such websites. But twitter is a very popular website, which does not distinguish itself for any particular community like movies, politics, literature, sports etc. It would be interesting to derive interesting information between IMDB and Twitter – so that it will give an opinion of how a specialized community thinks about a movie and what’s the opinion of general public.

**How do you plan to address the challenges? How would you design and implement the solution?**

Based on the analysis of HashTags in twitter we are planning to identify the opinion of the tweets. We need to find out the top trending HashTags for the movies and then classify them as positive or negative. We have to calculate an effective score based on the twitter data and visualize this score against the data from IMDB.

**How would you evaluate your project? How would you call it a success?**

If we can plot realistic visualizations of public opinion about movies based on twitter data against the data from IMDB, we believe the project will be a success.

**How would you partition the tasks and coordinate among group members?**

Fetching of data from IMDB/IMDBPy, Visualization – Sumesh

Twitter API, Interface design/Ismail

Classification of Twitter Data – Sumesh/Ismail