**Project Title**

MovieBuzz

**Team information: member names.**

Sumesh N Balan

Ismail Vandeliwala

**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 rating etc. We will be using both IMDB data and twitter APIs to achieve the desired visualization of cross plotting of various details like Movie, Director, and Actor etc. against the movies rating in IMDB and number of positive and negative tweets. 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.

**Challenges of the project**

Getting twitter data: Streaming data works only on new movies and it does not give data for old movies. Also it takes a lot of time to fetch the data. It takes around 10 minutes to fetch 100 tweets from one movie. So in order to get at least 50000 tweets, it takes 3 days to complete the process.

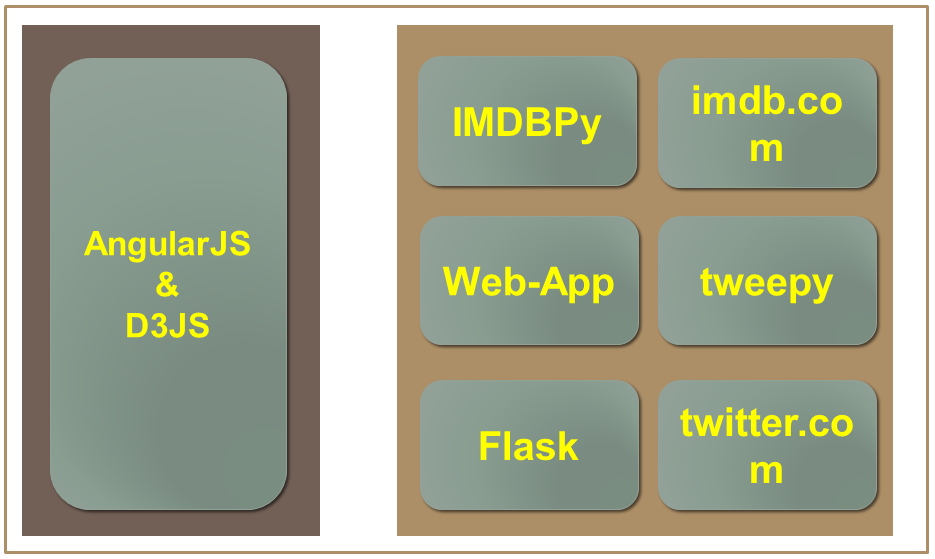
**Initial Implementation**

We implemented a web page in which user has an option to type movie name and press go button.

We collected around 500 tweets from 2 of the movies: Insurgent and Chappie

We have used flask as a web framework between front end (angular js) and python for middle layer.

**Architecture/Components**

****

**Change of initial plan**

Initial: We were planning to let user enter movie name and fetch the tweets and show the comparisons between imdb data and twitter data on the fly.

Now: As it takes lot of time to fetch tweets, we are going to gather tweets for 20-25 movies and store it in our project. Then user will be asked to select only those movie names and would be displayed different comparison graph between imdb and twitter data.

If it has to be visualized in real time, then the comparison could be just done for 100 tweets or less for the movie given.

**Evaluation**

We plan to use nltk – classification to classify the tweets for the movie into positive and negative. We will use 2 percent or 400 tweets (whichever is less) from the output given by ntlk to classify them manually and compare the manual results against the nltk classification results.

**Algorithm**

* User selects 2 movies from two different dropdowns and presses submit.
* Program fetches imdb data using imdb API for both movies and displays it on the page.
* Program reads the stored tweets for both movies and sends it to nltk.
* Program will run nltk classifier on those tweets and return the number of positive and negative tweets for both movies which will be displayed on web page.
* Program will display different visualization charts for different comparisons between number of tweets and imdb information.

**Tasks to be accomplished**

1) Have to implement nltk classification for classifying tweets into positive and negative.

2) Have to evaluate the classified tweets manually.

3) Have to implement different relations between imdb and twitter data using graph which will answer following questions:

3.1 How the public opinions differ from the critic’s opinion based on the critics rating.

3.2 Does the runtime of the movies affect the success of the movie?

3.3 Does the actor or director contribute to the success of the movie?

**Completion Dates**

* Task 1 and 2 – 04/12/2015
* Task 3 – 04/30/2015