JOB OPPORTUNITIES USING SOCIAL MEDIA

Big Data Analytics

ITE2013

PROJECT REPORT

SUBMITTED BY

Vinit Shahdeo	15BIT0335
Rishabh Agrawal	15BIT0252
Ratnesh Kanungo	15BIT0257

Problem Statement

We have collected the tweets related to jobs using python language and applied machine learning algorithms and sentiment analysis (using php) to find the job opportunities. Some specific hashtags played major role in finding job opportunities.

Abstract

The project is about searching the twitter for job opportunities. If you haven't been taking advantage of Twitter as a job search tool, it's time to jump in. When used intelligently, Twitter can have a profound impact on your job search success – or lack thereof. Small steps can help you turn Twitter into your own personal job search platform. We have collected the tweets related to jobs using python language and applied machine learning algorithms and sentiment analysis to find the job opportunities. Some specific hashtags played major role in finding job opportunities.

Introduction

Compared with social media job search heavyweights like LinkedIn, Twitter often flies under the radar as a tool for job search success. Don't let that underdog status fool you, though.

Twitter is all about enabling users to send out brief messages to large audiences. As a job seeker, you can use this to your advantage on multiple fronts including:

- Finding hidden job leads
- Networking with industry insiders
- Researching companies and interviewers
- Building your personal brand within the industry

In order to get the maximum job search benefit from Twitter, you need to develop a strategy for achieving your specific goals.

Popular hashtags used –

- #Hiring or #NowHiring

- #Jobs
- #Careers
- #TweetMyJobs
- #JobOpening
- #JobListing
- #JobPosting
- #JobOpportunities
- #HR
- #Graduate Jobs

Career and Industry specific hashtags

- #ITJobs
- #TechJobs
- #Marketing
- #Freelance (in case you're looking for quick or temporary jobs while the main job search continues)

Self-promoting hashtags to consider include:

- #HireMe
- #MBA
- #Candidate
- #JobSearching
- #Hire

Beyond the scope of networking, though, Twitter is an extremely useful tool for enhancing your efforts to research various companies as well as interviewers you may come across in your efforts to work for certain organizations.

Twitter is a great tool for getting to know people and businesses. Learn about the business, from an employee's point of view, by following the people who work for the organization. You can generally tell if they are shouting out the virtues of their employer or view work as essentially another four-letter word.

You can also learn a great deal about the company culture by observing company-posted tweets and responses to the tweets of others. Use Twitter to learn the following things and more:

- Have there been recent changes within the organization or structure of the company, new products, or industry-shaking news?
- How does the company define the position you're applying for? Your idea and the company's idea of a position or title are not always identical.
- What is the background of the people you'll be working with (specializations, fields, education, etc.)?
- Who are the company's major competitors and how does your future employer (let's be optimistic) stand out?

The more you know about the company before going into an interview, the better prepared you are to impress them with your knowledge.

Motivation

We have chosen to work with twitter since we feel it is a better approximation of public sentiment as opposed to conventional internet articles and web blogs. The reason is that the amount of relevant data is much larger for twitter, as compared to traditional blogging sites. Moreover the response on twitter is more prompt and also more general (since the number of users who tweet is substantially more than those who write web blogs on a daily basis). Sentiment analysis of public is highly critical in macro-scale socioeconomic phenomena like predicting the stock market rate of a particular firm. This could be done by analysing overall public sentiment towards that firm with respect to time and using economics tools for finding the correlation between public sentiment and the firm's stock market value. Firms can also estimate how well their product is responding in the market, which areas of the market is it having a favourable response and in which a negative response (since

twitter allows us to download stream of geo-tagged tweets for particular locations. If firms can get this information they can analyze the reasons behind geographically differentiated response, and so they can market their product in a more optimized manner by looking for appropriate solutions like creating suitable market segments. Predicting the results of popular political elections and polls is also an emerging application to sentiment analysis. One such study was conducted by Tumasjan et al. in Germany for predicting the outcome of federal elections in which concluded that twitter is a good reflection of offline sentiment.

Libraries & Languages Used –

1. **Python**
2. **PHP**
3. **Tweepy** – Python library for accessing Twitter API.
4. **TextBlob** – Python library for processing textual data.
5. **Mlab** – MongoDB Cloud Database.

Python Code–

Libraries used – Tweepy , TextBlob

Code for Sentiment Analysis

```
import re
import tweepy
from tweepy import OAuthHandler
from textblob import TextBlob

class TwitterClient(object):
    '''
    Generic Twitter Class for sentiment analysis.
    '''
    def __init__(self):
        '''
        Class constructor or initialization method.
        '''

        consumer_key = '18QHFMz0zvycM2KLrTMfrafI1'
        consumer_secret = 'WNwYGKBXmbfsY7ysZXxPJ4Voa7rgtLxGocuDHbIJ1TZLShtBVF'
        access_token = '843094924299976704-GNJyLjovEGFAiOWLswFBagKxlebRQUq'
        access_token_secret = 'L39Wz6lXKSavutPqhopmNwK7egJiSrwRxVohjbqVqbQvM'
```

```
        try:

            self.auth = OAuthHandler(consumer_key, consumer_secret)

            self.auth.set_access_token(access_token, access_token_secret)

            self.api = tweepy.API(self.auth)
        except:
            print("Error: Authentication Failed")

    def clean_tweet(self, tweet):
        '''
        Utility function to clean tweet text by removing links, special characters
        using simple regex statements.
        '''
        return ' '.join(re.sub("(@[A-Za-z0-9]+)|([^0-9A-Za-z \t])|(\w+:\/\/\S+)", " ",
tweet).split())

    def get_tweet_sentiment(self, tweet):
        '''
        Utility function to classify sentiment of passed tweet
        using textblob's sentiment method
        '''

        analysis = TextBlob(self.clean_tweet(tweet))

        if analysis.sentiment.polarity > 0:
            return 'positive'
        elif analysis.sentiment.polarity == 0:
            return 'neutral'
        else:
            return 'negative'

    def get_tweets(self, query, count = 10):
        '''
        Main function to fetch tweets and parse them.
        '''

        tweets = []

        try:
```

```
            fetched_tweets = self.api.search(q = query, count = count)


            for tweet in fetched_tweets:

                parsed_tweet = {}


                parsed_tweet['text'] = tweet.text

                parsed_tweet['sentiment'] = self.get_tweet_sentiment(tweet.text)

                if tweet.retweet_count > 0:

                    if parsed_tweet not in tweets:
                        tweets.append(parsed_tweet)
                else:
                    tweets.append(parsed_tweet)


            return tweets

        except tweepy.TweepError as e:
            print("Error : " + str(e))

def main():
    api = TwitterClient()
    tweets = api.get_tweets(query = 'Job Opportunities', count = 500)
    ptweets = [tweet for tweet in tweets if tweet['sentiment'] == 'positive']

    print("Positive tweets percentage: {} %".format(100*len(ptweets)/len(tweets)))

    ntweets = [tweet for tweet in tweets if tweet['sentiment'] == 'negative']

    print("Negative tweets percentage: {} %".format(100*len(ntweets)/len(tweets)))

    print("Neutral tweets percentage: {} % ".format(100*(len(tweets) - len(ntweets) -
len(ptweets))/len(tweets)))

    print("\n\nPositive tweets:")
    for tweet in ptweets[:10]:
        print(tweet['text'])

    print("\n\nNegative tweets:")
```

```
    for tweet in ntweets[:10]:
        print(tweet['text'])

if __name__ == "__main__":

    main()
```

Code for finding VARIANCE

```
import math
def find_variance(data):
  if len(data) == 0:
    return 0
  K = data[0]
  n = 0
  sum_ = 0
  sum_sqr = 0
  for x in data:
    n = n + 1
    sum_ += x - K
    sum_sqr += (x - K) * (x - K)
  variance = (sum_sqr - (sum_ * sum_)/n)/(n - 1)
  return variance

data=[52, 45, 45, 52, 26, 38, 15, 13]
variance=find_variance(data)
sd=variance**0.5
print(sd)
print('\n')
print(variance)
```

Sentiment analysis using PHP

```
<?php
if (PHP_SAPI != 'cli') {
        echo "<pre>";
}

$strings = array(
        1 => 'RT @Buzz_dont_tweet: Five amazing job opportunities with our new
#BackFromTheBrink projects! https://t.co/zmjWoQZGSn #ConservationJobs',
```

```
    2 => 'So fucking amped right now. Im going to have a meet up on friday with someone who has numerous job opportunities. its almost time',
    3 => 'RT @FirstEmploy: The latest job opportunities have just been uploaded! https://t.co/yWYuiKJ9v3 https://t.co/h4ESVRy2Wg',
    4 => 'Interested in a #job in #USA? This could be a great fit: https://t.co/Xg0lgz9Y6J #Engineering #Hiring #CareerArc https://t.co/U7RFaZPXNv',
    5 => '2 exciting job opportunities in Lab4Living. Children's prosthetics https://t.co/2hlV70ZHyR Healthcare Improvements https://t.co/etSHHYVBmf',
    6 => 'This Job Fair has many opportunities for multi-cultural and bilingual candidates. Look for the Prospanica',
    7 => 'We arere #hiring! Read about our latest #job opening here: Java Application Architect - https://t.co/NhKuDm0T7l… https://t.co/ki5bqBuPGg',
);




require_once __DIR__ . '/../autoload.php';
$sentiment = new \PHPInsight\Sentiment();
foreach ($strings as $string) {

    // calculations:
    $scores = $sentiment->score($string);
    $class = $sentiment->categorise($string);

    // output:
    echo "String: $string\n";
    echo "Dominant: $class, scores: ";
    print_r($scores);
    echo "\n";
}
```

Screenshots

Python Code & It's Ouput

jobtweets.py - C:\Users\Sony vaio\Desktop\BIG DATA PROJECT\jobtweets.py (3.4.2)

File Edit Format Run Options Windows Help

```
import re
import tweepy
from tweepy import OAuthHandler
from textblob import TextBlob

class TwitterClient(object):
    '''
    Generic Twitter Class for sentiment analysis.
    '''
    def __init__(self):
        '''
        Class constructor or initialization method.
        '''

        consumer_key = '18QHFMz0zvycM2KLrTMfrafI1'
        consumer_secret = 'WNwYGKBXmbfsY7ysZXxPJ4Voa7rgtLxGocuDHbIJ1TZLShtBVF'
        access_token = '843094924299976704-GNJyLjovEGFAiOWLswFBagKxlebRQUq'
        access_token_secret = 'L39Wz6lXKSavutPqhopmNwK7egJiSrwRxVohjbqVqbQvM'


        try:

            self.auth = OAuthHandler(consumer_key, consumer_secret)

            self.auth.set_access_token(access_token, access_token_secret)

            self.api = tweepy.API(self.auth)
        except:
            print("Error: Authentication Failed")
```

Ln: 8 Col: 45

8:40 AM 5/4/2017

Python 3.4.2 Shell

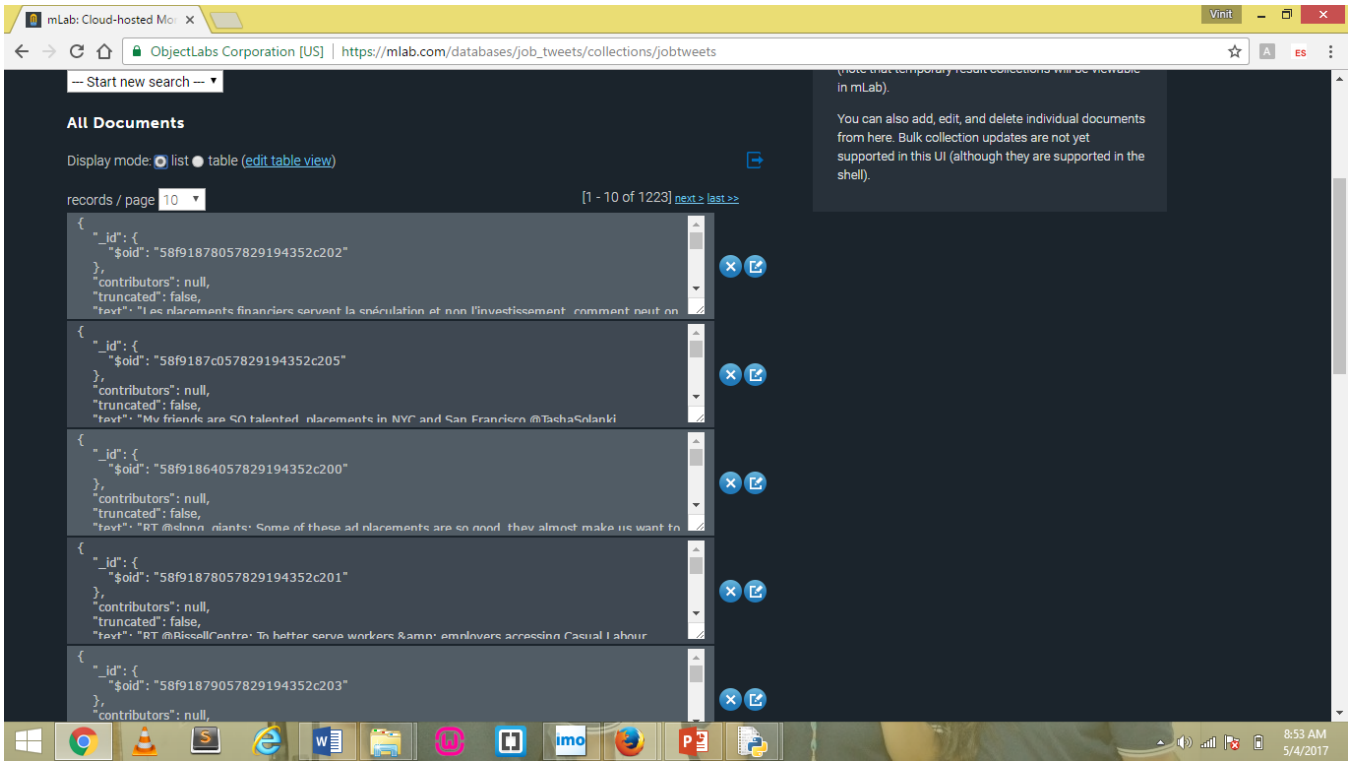
File Edit Shell Debug Options Windows Help

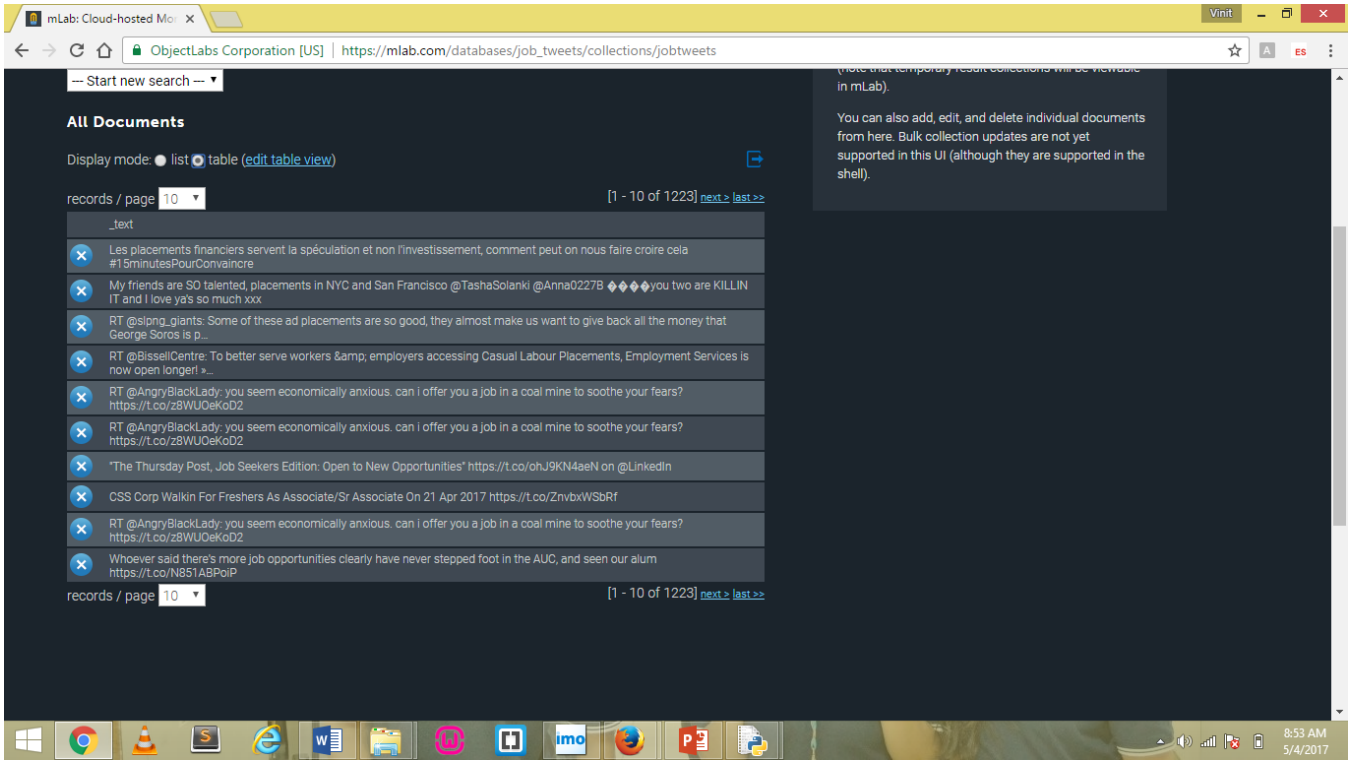
```
Indigenous Australia News Top story: Mental Health Nurse - Kununurra « Job oppo… https://t.co/kkiOPGQUOv, see more https://t.co/WNSmFmp7fl
RT @RealPeerReview: Gender integrated rugby has the potential to create lots of new job opportunities for hospital workers https://t.co/55q…
RT @ATLTechVillage: Great news- We have a Job Board and startups are hiring! 50+ job opportunities at the Village. Apply NOW: https://t.co/…
Qatar Large Job Opportunities - Attractive salary... https://t.co/79shs0n2iR
RT @EconomicTimes: Threat of #automation: Robotics and #ArtificialIntelligence to reduce #job opportunities at top banks https://t.co/8YG6…
RT @PrincesTrust_SW: Get into Retail with @marksandspencer will be back in #Bristol in July. Free training and job opportunities for 16-25…
Sexy boyfriend, great friends, good grades, job opportunities and oh yea, moving into a BOUJEE APARTMENT WITH THE LOML. Life is great. #mob
Santa Ana #CA #USA - Retail Merchandiser - We offer many opportunities wi... https://t.co/SIrxbrXqsH #LOS-ANGELES #SOUTHERN-CALIFORNIA #JOB
JOB LISTING: Second Mate /Educator (licensed)

Sound Experience is seeking applications for a licensed 100-ton... https://t.co/YUlNm96B6g


Negative tweets:
Joining the game has its perks! If you think you have what it takes apply for your dream job today:… https://t.co/WLFc1PqT5O
Threat of automation: Robotics and artificial intelligence to reduce job opportunities at ... https://t.co/rIJ9Fp8ipn via @economictimes
RT @Matthewcogdeill: #RETWEET "Life offers you endless Opportunities NEVER settle for dead end JOB, negative friends or abusive significant…
#RETWEET "Life offers you endless Opportunities NEVER settle for dead end JOB, negative friends or abusive signific… https://t.co/YyJRoWHuDO
CDOC CRIMINAL INVESTIGATOR I - State of Colorado Job Opportunities - Colorado https://t.co/5TNma7RkQT
>>>
```

Ln: 32 Col: 4

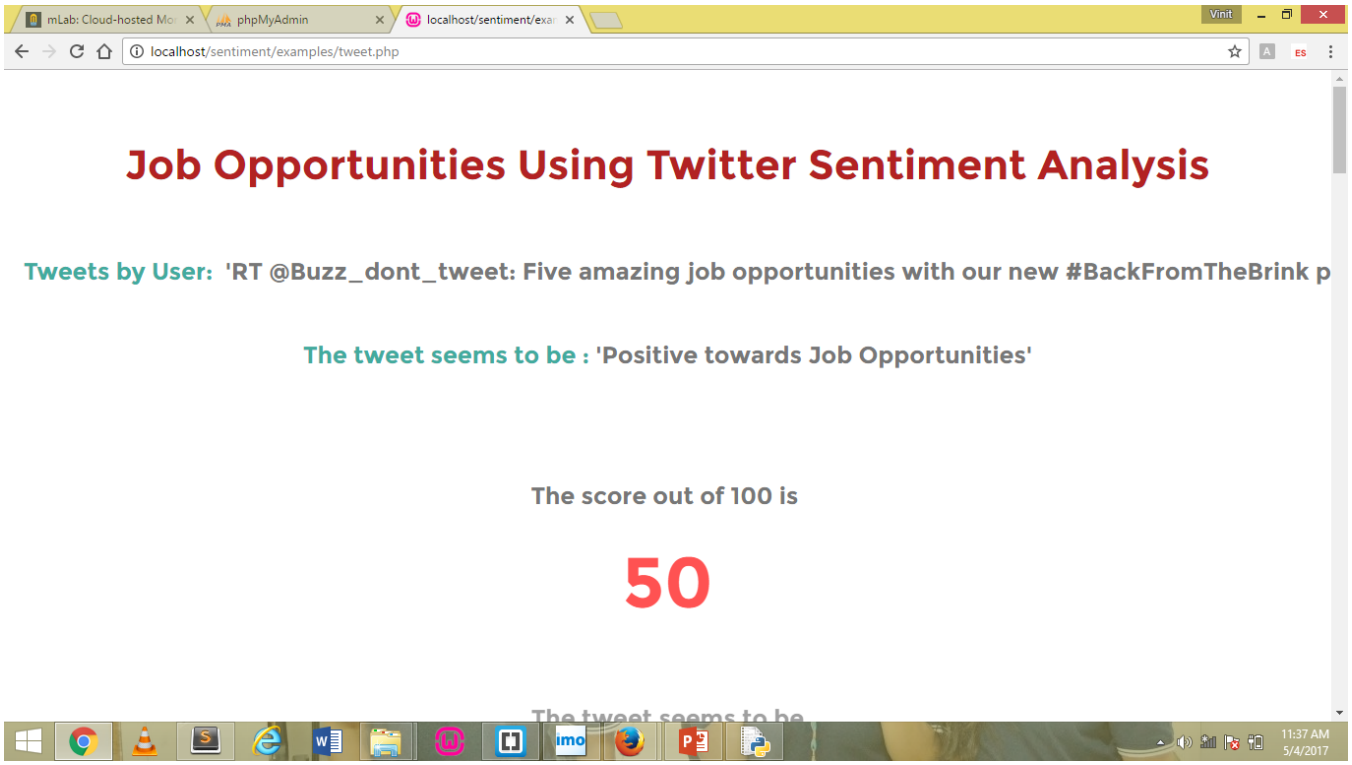
8:47 AM 5/4/2017

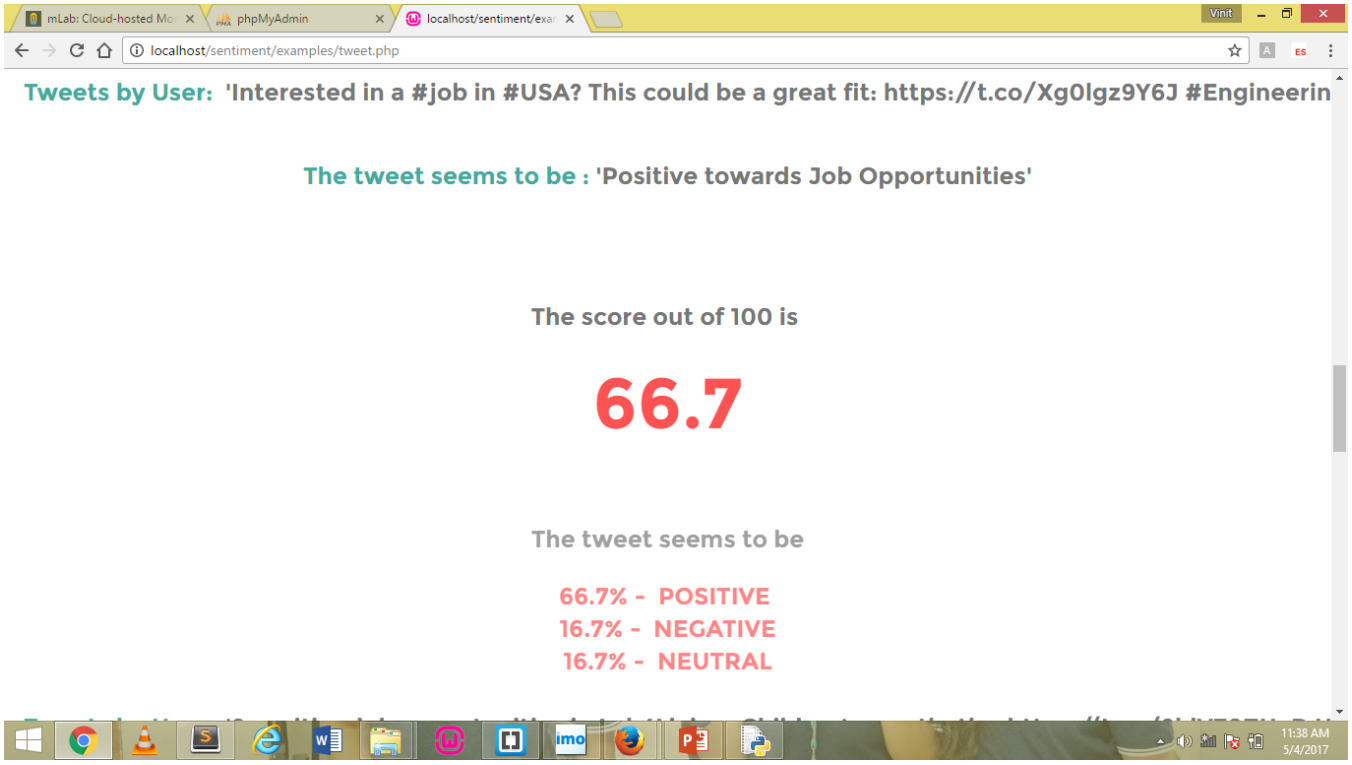
Mlab – MongoDB Cloud Database





Sentiment Analysis using PHP





Outcomes

The outcomes of doing sentiment analysis on popular hashtags using python are following –

#Hiring

Positive tweets percentage: 52.0 %

Negative tweets percentage: 1.0 %

Neutral tweets percentage: 47.0 %

#JobOpportunities

Positive tweets percentage: 45.333333333333336 %

Negative tweets percentage: 5.333333333333333 %

Neutral tweets percentage: 49.333333333333336 %

#Jobs

Positive tweets percentage: 45.833333333333336 %

Negative tweets percentage: 6.944444444444445 %

Neutral tweets percentage: 47.22222222222222 %

#Career

Positive tweets percentage: 52.11267605633803 %

Negative tweets percentage: 12.67605633802817 %

Neutral tweets percentage: 35.2112676056338 %

#Freelance

Positive tweets percentage: 26.136363636363637 %

Negative tweets percentage: 1.1363636363636365 %

Neutral tweets percentage: 72.72727272727273 %

#Marketing

Positive tweets percentage: 38.095238095238095 %

Negative tweets percentage: 10.714285714285714 %

Neutral tweets percentage: 51.19047619047619 %

#ITJobs

Positive tweets percentage: 15.0 %

Negative tweets percentage: 2.0 %

Neutral tweets percentage: 83.0 %

#jobopenings

Positive tweets percentage: 13.953488372093023 %

Negative tweets percentage: 39.53488372093023 %

Neutral tweets percentage: 46.51162790697674 %

Conclusion

Twitter is all about enabling users to send out brief messages to large audiences. If you haven't been taking advantage of Twitter as a job search tool, it's time to jump in. When used intelligently, Twitter can have a profound impact on your job search success – or lack thereof. Small steps can help you turn Twitter into your own personal job search platform. Try them today and see what a difference they make in your overall job search success.

Possible Improvements

Accuracy of the system can be improved by training on hand labeled tweets. This is because the jargons in tweets maybe different from the ones used in web. Thus, a system trained on basis of hashtags might not cover the words expressing sentiment on twitter. It can also be improved by using bigram counts in addition to unigram counts. As far as performance is considered, majority of the time is consumed to search for tweets. As only 100 results are allowed per query, twitter has to be queried multiple times to gather enough data. Thus, if twitter searches are made faster, the system will perform better in terms of response time.