

IMPROVED APPROACH FOR CLASSIFICATION AND PREDICTION IN SENTIMENT DATA ANALYSIS

Rohit Kumar Nonia

CSE Department

Nalanda College of Engineering, Gokhulpur

Bihar, India

Abhijit Bandyopadhyay

Computer Application Department

Raniganj Institute of Computer and Information Sciences, Raniganj

West Bengal, India

ABSTRACT

Sentiment Data Analysis is one of the prominent and widely used domains by the research scholars and practitioners. In this approach, there are number of tools and technologies available for fetching the live datasets, tweets, emotional attributes. Using these tools, the real time tweets and messages can be extracted from Twitter, Facebook, WhatAapp and many other social medial portals. In this article, the implementation of fetching live tweets from Twitter using python programming is presented. By this approach, the emotional attributes of the Internet users on social medial portals can be analyzed and prediction is done. Suppose we want to evaluate the overall cumulative score of a celebrity. For this, the python or PHP based programming scripts can fetch the live tweets about that celebrity or renowned person from Twitter. After that, using natural language processing toolkits, the fetched data in form of tweets or messages can be analyzed and prediction is done whether that particular person or movie or celebrity is getting fame or not.

Keywords - Sentiment Data Analysis, Social Mining, Social Media Mining, Opinion Mining

PREAMBLE

Big data analytics is the process of examining large data sets containing a variety of data types -- i.e., big data -- to uncover hidden patterns, unknown correlations, market trends, customer preferences and other useful business information. The analytical findings can lead to more effective marketing, new revenue opportunities, better customer service, improved operational efficiency, competitive advantages over rival organizations and other business benefits.

The primary goal of big data analytics is to help companies make more informed business decisions by enabling data scientists, predictive modelers and other analytics professionals to analyze large volumes of transaction data, as well as other forms of data that may be untapped by conventional business intelligence(BI) programs. That could include Web server logs and Internet clickstream data, social media content and social network activity reports, text from customer emails and survey responses, mobile-phone call detail records and machine data captured by sensors connected to the Internet of Things. Some people exclusively associate big data with semi-structured and unstructured data of that sort, but consulting firms like Gartner Inc. and Forrester Research Inc. also consider transactions and other structured data to be valid components of big data analytics applications.

Big data can be analyzed with the software tools commonly used as part of advanced analytics disciplines such as predictive analytics, data mining, text analytics and statistical analysis. Mainstream BI software and data visualization tools can also play a role in the analysis process. But the semi-structured and unstructured data may not fit well in traditional data warehouses based on relational databases. Furthermore, data warehouses may not be able to handle the processing demands posed by sets of big data that need to be updated frequently or even continually -- for example, real-time data on the performance of mobile applications or of oil and gas pipelines. As a result, many organizations looking to collect, process and analyze big data have turned to a newer class of technologies that includes Hadoop and related

tools such as YARN,MapReduce, Spark, Hive and Pig as well as NoSQL databases. Those technologies form the core of an open source software framework that supports the processing of large and diverse data sets across clustered systems.

Following is the statistical reports from InternetLiveStats.com and Statista.com about the real time data on social media and related web portals.

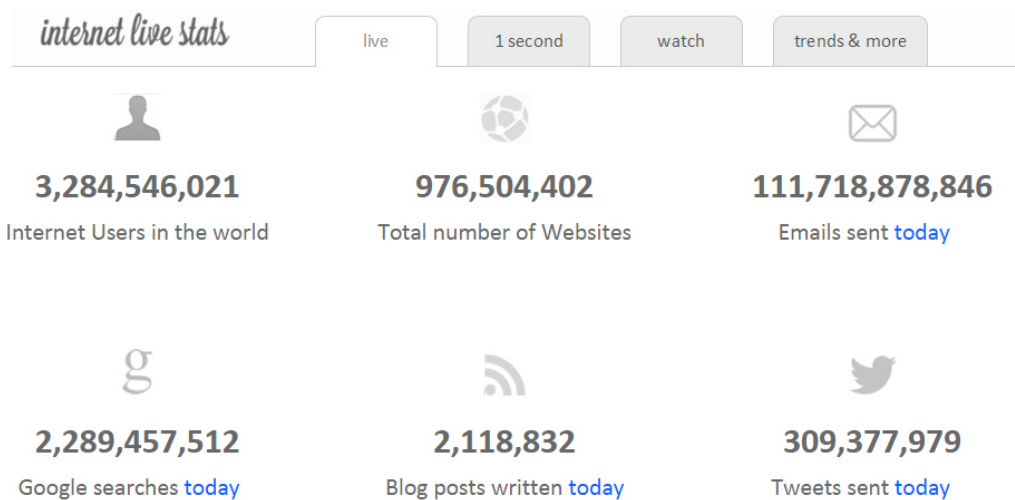


Figure 1 - Screenshot from InternetLiveStats.com

On Twitter, 350 Million Tweets are flowing daily by more than 500 Million Accounts. Around 571 New Websites are hosted every minute on World Wide Web. There are more than 5 billion users on mobile phones concurrently.

On WhatAapp, there are 700 Million Active Users. There are more than 1 Million New Users Registration. Around 30 Billion Messages Sent and 34 Billion Messages Received everyday on WhatAapp. On Facebook, 5 New Profiles are created every second. There are around 83 Billion fake Profiles. Around 300 Billion Photos are Uploaded everyday by 890 Billion Daily Active Users. The data of 320 TB is processed daily with 21 Minutes per User Average Time spent. All these statistical reports are amazing and growing in huge volume every time instance.

Now, the question is how to do research work on these datasets. By which technologies the real time datasets can be fetched. The live streaming data can be fetched using Python, PHP, PERL, Java and many other used for network programming.

FETCHING LIVE STREAMING DATA FROM TWITTER USING PYTHON CODE

For fetching live tweets from Twitter, the specific packages named tweepy and Twitter is required with python. After installation of these packages, the python code will be able to fetch live data from Twitter.

These can be installed using pip command as follows -

```
$ python -m pip install tweepy
```

```
$ python -m pip install Twitter
```

CODE TO FETCH LIVE TWEETS FROM TWITTER

```
from tweepy import Stream  
from tweepy import OAuthHandler  
from tweepy.streaming import StreamListener  
my_app_consumerkey = 'XXXXXXXXXXXXXXXXXXXXXXXXXXXX'  
my_app_consumersecret = 'XXXXXXXXXXXXXXXXXXXXXXXXXXXX'  
my_app_accesstoken = 'XXXXXXXXXXXXXXXXXXXXXXXXXXXX'  
my_app_accesssecret = 'XXXXXXXXXXXXXXXXXXXXXXXXXXXX'  
class TweetListener(StreamListener):  
    def on_data(self, mydata):  
        print mydata  
        return True  
    def on_error(self, status):  
        print status
```

```

auth = OAuthHandler(my_app_consumerkey, my_app_consumersecret)
auth.set_my_app_access_token(my_app_access_token, my_app_access_secret)
stream = Stream(auth, TweetListener())
stream.filter(track=[Name of the Celebrity or Movie or Person'])

```

```

{"created_at": "Wed Aug 26 05:42:38 +0000
2015", "id": 636413421097889792, "id_str": "636413421097889792", "text": "15mins from FourSeasons to cross WorliNaka
signal.. Fucking retarded city.. #mumbai", "source": "\u003ca href=\\"http://twitter.com/download/iphone\\"
rel=\\"nofollow\\" \u003eTwitter for
iPhone\u003c/a\u003e", "truncated": false, "in_reply_to_status_id": null, "in_reply_to_status_id_str": null, "in_reply_to_
user_id": null, "in_reply_to_user_id_str": null, "in_reply_to_screen_name": null, "user": {"id": 47670394, "id_str": "47670394
", "name": "PoojaS", "screen_name": "PoojaSolanki", "location": "Mumbai", "url": "http://www.poojas.wordpress.com", "descri
ption": "#INTENT: Anything & Everything...It only matters when you make it matter, when you let it
matter...\n\nalso @formspaceorder ... the creative
ME", "protected": false, "verified": false, "followers_count": 743, "friends_count": 507, "listed_count": 20, "favourites_count
": 167, "statuses_count": 13479, "created_at": "Tue Jun 16 16:40:00 +0000 2009", "utc_offset": 19800, "time_zone": "New
Delhi", "geo_enabled": false, "lang": "en", "contributors_enabled": false, "is_translator": false, "profile_background_color":
"9E888B", "profile_background_image_url": "http://pbs.twimg.com/profile_background_images/126747951/xcdab80bfe31
01214aeea33742408f46.png", "profile_background_image_url_https": "https://pbs.twimg.com/profile_background_images/
126747951/xcdab80bfe3101214aeea33742408f46.png", "profile_background_tile": true, "profile_link_color": "2600FF", "profi
le_sidebar_border_color": "A1A35D", "profile_sidebar_fill_color": "A3945D", "profile_text_color": "0B0002", "profile_use_b
ackground_image": true, "profile_image_url": "http://pbs.twimg.com/profile_images/569409479704207360/irhEBGYA_norm
al.jpeg", "profile_image_url_https": "https://pbs.twimg.com/profile_images/569409479704207360/irhEBGYA_normal.jpe
g", "profile_banner_url": "https://pbs.twimg.com/profile_banners/47670394/1375557451", "default_profile": false, "de
fault_profile_image": false, "following": null, "follow_request_sent": null, "notifications": null, "geo": null, "coordinates
": null, "place": null, "contributors": null, "retweet_count": 0, "favorite_count": 0, "entities": {"hashtags": [{"text": "mumbai
", "indices": [76, 83]}], "trends": [], "urls": [], "user_mentions": [], "symbols": []}, "favorited": false, "retweeted": false, "po
ssibly_sensitive": false, "filter_level": "low", "lang": "en", "timestamp_ms": "1440567758236"}

```

Figure 2 - Live Tweets fetched from Twitter in JSON Format

After execution of this script, the output dataset is fetched in JSON file format. The JSON file can be parsed using OpenRefine tool in the XML, CSV or any other readable format by the data mining and machine learning tools.

Openrefine is one of the powerful and effective tools that is used for processing the BigData and JSON file format.

In the similar way, the timeline of any person or Twitter id can be fetched using the following code -

```

import tweepy

import time

my_app_consumerkey = 'XXXXXXXXXXXXXXXXX'
my_app_consumersecret = 'XXXXXXXXXXXXXXXXX '
my_app_access_token = 'XXXXXXXXXXXXXXXXX '

```

```
my_app_accesssecret = 'XXXXXXXXXXXXXXX '  
auth = tweepy.auth.OAuthHandler(my_app_consumerkey,  
my_app_consumersecret)  
auth.set_my_app_accesstoken(my_app_accesstoken, my_app_accesssecret)  
api = tweepy.API(auth)  
list= open('Twitter.txt','w')  
if(api.verify_credentials):  
    print 'Connected to Twitter Server'  
currentuser = tweepy.Cursor(api.followers,  
screen_name="gauravkumarin").item()  
while True:  
    try:  
        u = next(currentuser)  
        list.write(u.screen_name + '\n')  
    except:  
        time.sleep(15*60)  
        u = next(currentuser)  
        list.write(u.screen_name + '\n')  
list.close()
```

The following script of Python can be used to parse the JSON to CSV format

```
JSON - CSV Parser  
import fileinput  
import json  
import csv  
import sys  
l = []  
for currentline in fileinput.input():  
    l.append(currentline)  
currentjson = json.loads("".join(l))  
keys = {}
```

```
for i in currentjson:
    for k in i.keys():
        keys[k] = 1
mycsv = csv.DictWriter(sys.stdout, fieldnames=keys.keys(),
                       quoting=csv.QUOTE_MINIMAL)
mycsv.writeheader()
for row in currentjson:
    mycsv.writerow(row)
```

FETCHING DATA FROM TWITTER USING PHP CODE

For fetching the live tweets using PHP code, the API *TwitterAPIExchange* is required. After including this API in this PHP code, the script will directly interact with the Twitter Servers and live streaming data.

```
<?php
error_reporting(0);
define('CURRENTDBHOST','localhost');
define('CURRENTDBUSERNAME','root');
define('CURRENTCURRENTDBPASSWORD','');
define(' CURRENTDBPASSWORD ','Twitter');
define('CURRENTTWEETTABLE','Tweets');
require_once('TwitterAPIExchange.php');
$settings = array(
    'oauth_my_app_accesstoken' => "XXXXXXXXXXXXXXXXXXXX",
    'oauth_my_app_accesstoken_secret' => "XXXXXXXXXXXXXXXXXXXX",
    'my_app_consumerkey' => "XXXXXXXXXXXXXXXXXXXX",
    'my_app_consumersecret' => "XXXXXXXXXXXXXXXXXXXX"
);
$url = "https://api.Twitter.com/1.1/statuses/user_timeline.json";
$myRequestMethod = "GET";
$getfield = '?screen_name=gauravkumarin&count=20';
```

```

$Twitter = new TwitterAPIExchange($settings);
$string = json_decode($Twitter->setGetfield($getfield)
->buildOauth($url, $requestMethod)
->performRequest(), $assoc = TRUE);
if($string["errors"][0]["message"] != "") {echo "<h3>Sorry, there
was a          problem.</h3><p>Twitter returned the following error
message:</p><p>
<em>".$string[errors][0]["message"]."</em></p>";exit();}
foreach($string as $items)
{
    echo "Tweeted by: ". $items['currentuser']['name']."<br />";
    echo "Screen name: ". $items['currentuser']['screen_name']."<br
/>";
    echo "Tweet: ". $items['text']."<br />";
    echo "Time and Date of Tweet: ".$items['timestamp']."<br />";
    echo "Tweet ID: ".$items['id_str']."<br />";
    echo "Followers: ". $items['currentuser']['followers']."<br /><hr
/>";
    echo
insertTweetsDB($items['currentuser']['name'],$items['currentuser']['scre
en_name'],$items['text'],$items['timestamp'],$items['id_str'],$items['current
user']['followers']);
}
function
insertTweetsDB($name,$screen_name,$text,$timestamp,$id_str,$followers){
    $mysqli = new mysqli(CURRENTDBHOST,
CURRENTDBUSERNAME,          CURRENTCURRENTDBPASSWORD,
MYDBNAME);
    if ($mysqli->connect_errno) {
        return 'Failed to connect to Database: (' . $mysqli-
>connect_errno . ')'. $mysqli->connect_error;

```



```

}
$querystmt='INSERT INTO
'.MYDBNAME.'.CURRENTTWEETTABLE.' (name, screen_name, text,
timestamp, id_str, followers) VALUES (?, ?, ?, ?, ?, ?);';
if ($insert_stmt = $mysqli->prepare($querystmt)){
    $insert_stmt->bind_param('sssid',
$name, $screen_name, $text, $timestamp, $id_str, $followers);
    if (!$insert_stmt->execute()) {
        $insert_stmt->close();
        return 'Tweet Creation cannot be done at this moment.';
    }elseif($insert_stmt->affected_rows>0){
        $insert_stmt->close();
        return 'Tweet Added.';
    }else{
        $insert_stmt->close();
        return 'No Tweet were Added.';
    }
}
}
}
}

```

Using these technologies, the parsing, processing and prediction on the real time tweets and their association with the particular event can be mapped. The news channels adopt this type of implementation for exit polls which predict the probability of winning by any political party or candidate. In similar aspect, the prediction on the business of any movie can be done after careful analysis on the live streaming data.

LIVE DATA FETCHING USING PYTHON

Using LinkChecker library in Python, the recursive and deep checking of the server pages can be done. Using LinkChecker, the site crawling is made easy with features

of integrating the regular expressions and filtering. The output can be generated in multiple formats including HTML, XML, CSV, SQL or simply the sitemap graph.

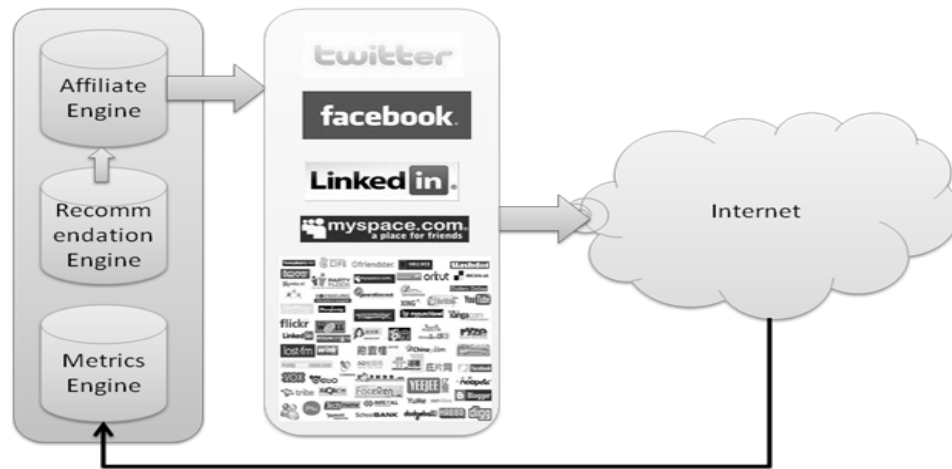


Figure 3 – Recommender System and Related Engines in Social Media

WEBSCRAPING

Python is used by the researchers and practitioners for collecting the live data for research and development. For example, we can fetch the live records of stock market, price of any product from E-Commerce websites. Such data collected is the foundation of BigData Analytics. If a researcher is doing research on big data analysis, the live data can be fetched using Python Script and then it can be processed based on the research objectives.

Here is the code snippet to fetch the live stock exchange data from the website timesofindia.com using Python

```
from bs4 import BeautifulSoup  
import urllib.request  
from time import sleep  
from datetime import datetime  
def getnews():
```

```

url = "http://timesofindia.indiatimes.com/business"
scraping = BeautifulSoup(page)
price = scraping.findAll("span",attrs={"class":"red14px"})[0].text
return price
with open("bseindex.out","w") as f:
    for x in range(2,100):
        sNow = datetime.now().strftime("%I:%M:%S%p")
        f.write("{0}, {1}\n".format(sNow, getnews()))
        sleep(1)

```

FETCHING LIVE DATA FROM SOCIAL MEDIA

In the same way, the twitter live feeds can be fetched using Python APIs. Using twitter developer account, the new app can be created and then the Python Script is mapped with the Twitter App



```

twitter_analytics — python2.7 — 80x24
{
  "statuses_count":24096,"created_at":"Wed Nov 23 23:34:19 +0000 2011","utc_offset":
  :-14400,"time_zone":"Eastern Time (US & Canada)","geo_enabled":true,"lang":"en",
  "contributors_enabled":false,"is_translator":false,"profile_background_color":"1
  91919","profile_background_image_url":"http://pbs.twimg.com/profile_backgroun
  d_images/738070196/237c3eb2ccb14a8df528a80986a8676.jpeg","profile_background_
  image_url_https":"https://pbs.twimg.com/profile_background_images/738070196/
  237c3eb2ccb14a8df528a80986a8676.jpeg","profile_background_tile":false,"profile
  _link_color":"009999","profile_sidebar_border_color":"FFFFFF","profile_sidebar_f
  ill_color":"DDEEF6","profile_text_color":"333333","profile_use_background_image
  :true,"profile_image_url":"http://pbs.twimg.com/profile_images/4566158900253
  40928/Qo_JEm96_normal.jpeg","profile_image_url_https":"https://pbs.twimg.com/
  profile_images/456615890025340928/Qo_JEm96_normal.jpeg","profile_banner_url":
  "https://pbs.twimg.com/profile_banners/419918470/1404682685","default_profi
  le":false,"default_profile_image":false,"following":null,"follow_request_sent":n
  ull,"notifications":null},"geo":null,"coordinates":null,"place":null,"contributo
  rs":null,"retweet_count":34,"favorite_count":30,"entities":{"hashtags":[],"trend
  s":[]},"url":[],"user_mentions":[],"symbols":[]},"favorited":false,"retweeted":f

```

Figure 4 – Fetching Live Tweets from Twitter in JSON Format

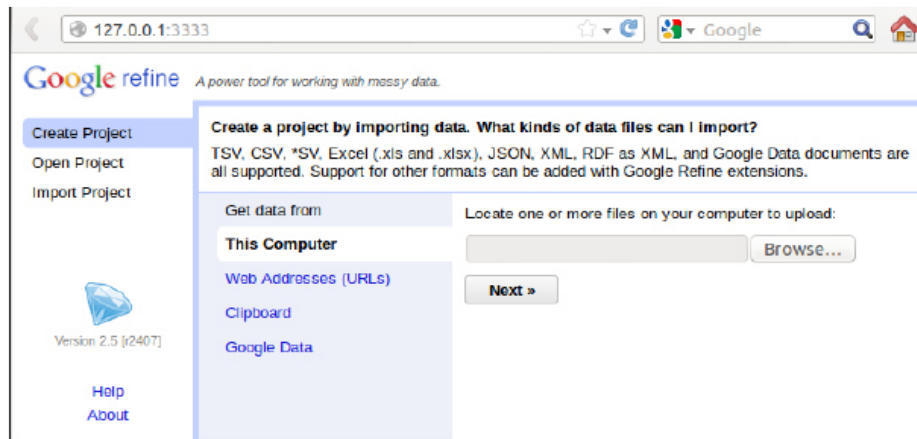


Figure 5 – Parsing of JSON using Google Refine

Sentiment classification

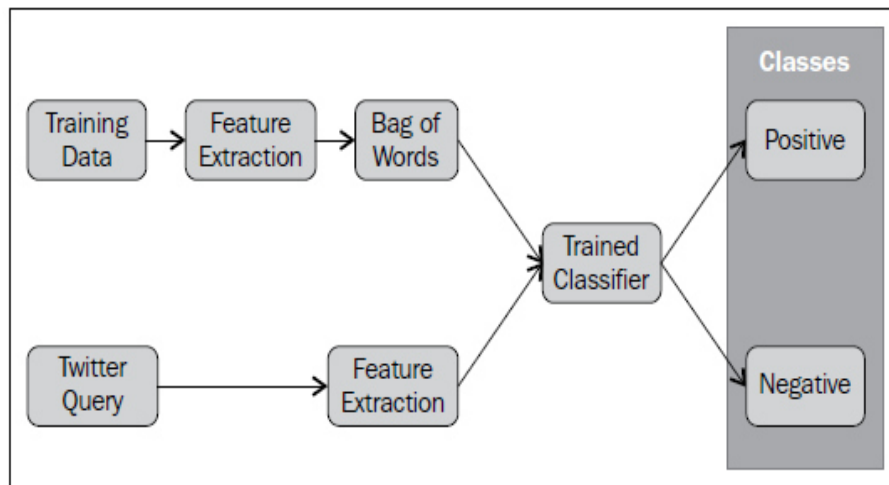


Figure 6 – Generation of Sentiments in Different Classes

REAL WORLD APPLICATIONS OF SOCIAL MINING

- Banking and Finance
- Marketing and Business Policies
- Consumer Behavior
- Market Basket Analysis
- Military and Defense

- Log Files Analysis
- Forensic Investigation
- Bio-Statistics and Bio-Informatics
- Web Usage Mining
- Medical Data Mining
- Expert Systems
- Knowledge Discovery

REVIEW OF EXISTING LITERATURE

Bollen (2010) [1] - Behavioral economics tells us that emotions can profoundly affect individual behavior and decision-making. Does this also apply to societies at large, i.e. can societies experience mood states that affect their collective decision making? By extension is the public mood correlated or even predictive of economic indicators? Here we investigate whether measurements of collective mood states derived from large-scale Twitter feeds are correlated to the value of the Dow Jones Industrial Average (DJIA) over time. This work analyze the text content of daily Twitter feeds by two mood tracking tools, namely OpinionFinder that measures positive vs. negative mood and Google-Profile of Mood States (GPOMS) that measures mood in terms of 6 dimensions (Calm, Alert, Sure, Vital, Kind, and Happy). This paper cross-validate the resulting mood time series by comparing their ability to detect the public's response to the presidential election and Thanksgiving day in 2008. A Granger causality analysis and a Self-Organizing Fuzzy Neural Network are then used to investigate the hypothesis that public mood states, as measured by the OpinionFinder and GPOMS mood time series, are predictive of changes in DJIA closing values. The results in this paper indicate that the accuracy of DJIA predictions can be significantly improved by the inclusion of specific public mood dimensions but not others. The authors find an accuracy of 87.6% in predicting the daily up and down changes in the closing values of the DJIA and a reduction of the Mean Average Percentage Error by more than 6%.

Bollen (2009) [2] - Microblogging is a form of online communication by which users broadcast brief text updates, also known as tweets, to the public or a selected circle of contacts. A variegated mosaic of microblogging uses has emerged since the launch of Twitter in 2006: daily chatter, conversation, information sharing, and news commentary, among others. Regardless of their content and intended use, tweets often convey pertinent information about their author's mood status. As such, tweets can be regarded as temporally-authentic microscopic instantiations of public mood state. In this article, we perform a sentiment analysis of all public tweets broadcasted by Twitter users between August 1 and December 20, 2008. For every day in the timeline, we extract six dimensions of mood (tension, depression, anger, vigor, fatigue, confusion) using an extended version of the POMS, a well-established psychometric instrument. The authors in this paper compare the results to the values recorded by stock market and crude oil price indices and major events in media and popular culture, such as the U.S. Presidential Election of November 4, 2008 and Thanksgiving Day. This work finds that events in the social, political, cultural and economic sphere do have a significant, immediate on the various dimensions of public mood. The authors speculate that large scale analyses of mood can provide a solid platform to model collective emotive trends in terms of their predictive value with regards to existing social as well as economic indicators.

Asur (2010) [3] – Sentiment Analysis is important part for social networking and content sharing. And yet, the content that is generated from these websites remains largely untapped. In this paper, the authors demonstrate how social media content can be used to predict real-world outcomes. In particular, this work uses the chatter from Twitter.com to forecast box-office revenues for movies. This paper shows that a simple model built from the rate at which tweets are created about particular topics can outperform market-based predictors. This work further demonstrates how sentiments extracted from Twitter can be further utilized to improve the forecasting power of social media.

Tan (2011) [4] – The authors show that information about social relationships can be used to improve user-level sentiment analysis. The main motivation behind the approach is that users that are somehow “connected” may be more likely to hold similar opinions; therefore, relationship information can complement what we can extract about a user’s viewpoints from their utterances. Employing Twitter as a source for our experimental data, and working within a semi-supervised framework, we propose models that are induced either from the Twitter follower/followee network or from the network in Twitter formed by users referring to each other using “@” mentions. The proposed transductive learning results reveal that incorporating social-network information can indeed lead to statistically significant sentiment classification improvements over the performance of an approach based on Support Vector Machines having access only to textual features.

Saif (2012) [5] - Sentiment analysis over Twitter offer organisations a fast and effective way to monitor the publics’ feelings towards their brand, business, directors, etc. A wide range of features and methods for training sentiment classifiers for Twitter datasets have been researched in recent years with varying results. In this paper, we introduce a novel approach of adding semantics as additional features into the training set for sentiment analysis. For each extracted entity (e.g. iPhone) from tweets, we add its semantic concept (e.g. “Apple product”) as an additional feature, and measure the correlation of the representative concept with negative/positive sentiment. The authors apply this approach to predict sentiment for three different Twitter datasets. The results show an average increase of F harmonic accuracy score for identifying both negative and positive sentiment of around 6.5% and 4.8% over the baselines of unigrams and part-of-speech features respectively. We also compare against an approach based on sentiment-bearing topic analysis, and find that semantic features produce better Recall and F score when classifying negative sentiment, and better Precision with lower Recall and F score in positive sentiment classification.

Davidov (2010) [6] - Automated identification of diverse sentiment types can be beneficial for many NLP systems such as review summarization and public media

analysis. In some of these systems there is an option of assigning a sentiment value to a single sentence or a very short text. In this paper the authors propose a supervised sentiment classification framework which is based on data from Twitter, a popular microblogging service. By utilizing 50 Twitter tags and 15 smileys as sentiment labels, this framework avoids the need for labor intensive manual annotation, allowing identification and classification of diverse sentiment types of short texts. The authors evaluate the contribution of different feature types for sentiment classification and show that the proposed framework successfully identifies sentiment types of untagged sentences. The quality of the sentiment identification was also confirmed by human judges. This paper also explores dependencies and overlap between different sentiment types represented by smileys and Twitter hashtags.

Saif (2012) [7] - Twitter has brought much attention recently as a hot research topic in the domain of sentiment analysis. Training sentiment classifiers from tweets data often faces the data sparsity problem partly due to the large variety of short and irregular forms introduced to tweets because of the 140-character limit. In this work the authors propose using two different sets of features to alleviate the data sparseness problem. One is the semantic feature set where this work extracts semantically hidden concepts from tweets and then incorporate them into classifier training through interpolation. Another is the sentiment-topic feature set where we extract latent topics and the associated topic sentiment from tweets, then augment the original feature space with these sentiment-topics. Experimental results on the Stanford Twitter Sentiment Dataset show that both feature sets outperform the baseline model using unigrams only. Moreover, using semantic features rivals the previously reported best result. Using sentiment topic features achieves 86.3% sentiment classification accuracy, which outperforms existing approaches.

Bifet (2009) [8] - Micro-blogs are a challenging new source of information for data mining techniques. Twitter is a micro-blogging service built to discover what is happening at any moment in time, anywhere in the world. Twitter messages are short, and generated constantly, and well suited for knowledge discovery using data stream

mining. The authors briefly discuss the challenges that Twitter data streams pose, focusing on classification problems, and then consider these streams for opinion mining and sentiment analysis. To deal with streaming unbalanced classes, this work propose a sliding window Kappa statistic for evaluation in time-changing data streams. Using this statistic this work performs a study on Twitter data using learning algorithms for data streams.

CONCLUSION AND SCOPE OF FUTURE WORK

There is huge scope of research and development using Python scripts and specialized APIs for assorted applications including cyber security, data mining, Internet of Things, cloud simulation, grid implementation and many others. Python is one of the effective programming languages that can process and handle any type of data stream. In the proposed work, the live fetching of social media messages shall be integrated with assorted advance data mining tools for effective prediction and modeling.

REFERENCES

- [1] Bollen, J., Mao, H., & Zeng, X. (2011). Twitter mood predicts the stock market. *Journal of Computational Science*, 2(1), 1-8.
- [2] Bollen, J., Pepe, A., & Mao, H. (2009). Modeling public mood and emotion: Twitter sentiment and socio-economic phenomena. *arXiv preprint arXiv:0911.1583*.
- [3] Asur, S., & Huberman, B. (2010, August). Predicting the future with social media. In *Web Intelligence and Intelligent Agent Technology (WI-IAT), 2010 IEEE/WIC/ACM International Conference on* (Vol. 1, pp. 492-499). IEEE.
- [4] Tan, C., Lee, L., Tang, J., Jiang, L., Zhou, M., & Li, P. (2011, August). User-level sentiment analysis incorporating social networks. In *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 1397-1405). ACM.
- [5] Saif, H., He, Y., & Alani, H. (2012). Semantic sentiment analysis of twitter. In *The Semantic Web–ISWC 2012* (pp. 508-524). Springer Berlin Heidelberg.

- [6] Davidov, D., Tsur, O., & Rappoport, A. (2010, August). Enhanced sentiment learning using twitter hashtags and smileys. In *Proceedings of the 23rd International Conference on Computational Linguistics: Posters* (pp. 241-249). Association for Computational Linguistics.
- [7] Saif, H., He, Y., & Alani, H. (2012). Alleviating data sparsity for twitter sentiment analysis. CEUR Workshop Proceedings (CEUR-WS. org).
- [8] Bifet, A., & Frank, E. (2010, January). Sentiment knowledge discovery in twitter streaming data. In *Discovery Science* (pp. 1-15). Springer Berlin Heidelberg.