International Journal of Scientific Research in Computer Science, Engineering and Information Technology
© 2018 IJSRCSEIT | Volume 3 | Issue 6 | ISSN : 2456-3307

Sentiment Analysis Using Parallel Computing Through GPU

Harshita Mandloi, Shraddha Masih
School of Computer Science and IT, DAVV, Indore, Madhya Pradesh, India

ABSTRACT

Parallel Computing is becoming important in the field of computer science and is proven as a high-performance solution. Over the couple of years, GPU has gained an important place in the field of high-performance computing. Social media is expanding at present and becoming important in society. Social network sites allow users to communicate with people in the network by sharing posts, images, videos, status. The proposed system gathers the information from the social media websites and performs the sentiment analysis on the social media data using GPU. The work concentrates on recognizing the sentiment information from the text reviews and using that to identify the items. The aim of this paper is to do analytics on social media data. Analysis is done on the data using K Nearest Neighbor algorithms and Support Vector Machine algorithm on the GPU.

Keywords: Parallel Computing, GPU, Social media, K nearest neighbor, SVM

I. INTRODUCTION

Nowadays social media, blogs, and other media produce a huge amount of data on the world wide web. This huge amount of data contain a critical opinion related to the information that can be used to benefit business and other aspects of commercial and scientific industries. Analysis of user sentiments can help in a business decisions. Sentiments can be categorized into 3 types:- Positive, Negative and Neutral.

Graphics Processing Unit has shown better computational performance as compared to the current main-stream multicore Central Processing Unit. Apart from graphics and multimedia processing, high-performance GPUs are also mapped to take care of General Purpose Computing.

Machine learning is the current trend in computing that is focused on designing and developing algorithms which allows computers to learn. It can capture characteristics of interest in order to make a prediction for a new data query. The gathered data is considered as training data which illustrate the relationships among the observed variables. Many important patterns can be recognized after applying the learning procedure. Supervised learning is the data mining task of classifying data into labeled data. This work is mainly focused on the classification tools- Support Vector Machine and K Nearest Neighbor. To improve the performance of this algorithm GPU is used with CUDA. Parallel SVM algorithm is implemented on GPU using CUDA framework. The implementation of parallel SVM achieves a great performance using GPU. This software utilizes the parallelism in both data level and task level to maximize the performance of the GPU.

II. BACKGROUND

A. Parallel Computing- Parallel computing is becoming important day by day. Dividing large task into sub task and assigning these tasks to multiprocessor and executing them concurrently is called parallel processing. Parallel computing is used for high performance computing in the field of data analytics. For the high performance computing the

CSEIT183687 | Received : 20 July 2018 | Accepted : 05 August 2018 | July-August-2018 [3 (6) : 428-434]

Graphic Processing Unit (GPU) has played an important role because of the low cost and parallel processing power.

B. Social Media- Social media is the changing the way of the people find information, share knowledge and communicate with each other. The most popular social networking sites are Facebook, Twitter, yahoo etc. The term social media refers to as:

Social: It refers to connect with other people by sharing information with them and receiving information from them.

Media: It refers to an instrument of communication like internet, TV, newspaper etc.

C. Sentiment Analysis- Sentiment Analysis is also called opinion mining. It is an analysis of the feeling, emotions, attitude. Sentiment analysis is the process of determining whether a piece of script is positive, negative or neutral. It is generally used in social media to check as it allows us to gain an overview of the public opinion behind certain topics. Its application is larges and powerful, the ability to extract insight from social data is a practice that is being widely adopted by organizations across the world.

Different Classes of Sentiment Analysis

i. Positive Sentiment: It refers to the positive attitude of the speaker about the text.

ii. Negative Sentiment: It refers to the negative attitude of the speaker about the text.

iii. Neutral Sentiment: In this no emotion are reflected about the text.

D. Graphic Processing Unit- Graphic Processing Unit is a solitary chip processor mainly used to manage and boost the performance of videos and graphics. A CPU consist of few cores while GPU consists of massively parallel architecture consisting of thousands of smaller and efficient cores designed to handle multiple tasks at the same time. It is not only used in a computer on a video card or motherboard, apart it is also used in mobile phones, display adapters, workstations and game consoles. The GPU is also known as Visual Processing Unit (VPU).

GPU programming Models

i. CUDA (Compute Unified Device Architecture)
ii. OpenCL (Open Computing Language)
iii. C++ AMP (Accelerated Massive Parallelism)
iv. OpenACC (Open Accelerator)

E. Support Vector Machine

It is a supervised machine learning algorithm which can be used for both classification and regression challenges. However, it is mostly used in classification problem. An SVM yields a guide of the sorted data with the margins between the two as far apart as possible. SVMs are utilized in text categorization, image classification, handwriting recognition and in the sciences. It is also known as Support Vector Network (SVN). An SVM creates parallel partitions by generating two parallel lines. Divide two categories by a clear gap that should be as wide as possible, do this partition by a known as hyper plane.

SVM Algorithm:

```
candidateSV = {closest pair from opposite classes }
while there are violating points do
        Find a violator
        candidateSV = candidateSV S
violator
if any αp < 0 due to addition of c to S then
        candidateSV = candidateSV \ p
repeat till all such points are pruned
end if
end while
```

F. K-Nearest Neighbor

K-Nearest Neighbor is one of the most basic yet essential classification algorithm in machine learning. It belongs to supervised machine learning domain and finds intense application in pattern recognition, data mining and intrusion detection.

KNN Algorithm:

In this algorithm we are assuming n as number of training data set and p as un unknown point.

```
I.   In this algorithm, store the training data set in an array of data points arr []. This each element of this array represents a tuple (x,y).
II.  for i=0 to m:
III. calculate Euclidean distance d (arr[i],      p)
IV.  Make set S of K smallest distances obtained. Each of these distances corresponds to an already classified data point.
V.   Return the majority label among S.
```

III. RELATED WORK

Usually CPU based problem are not suitable for parallel computing due their cost and scalability issues. The GPU is one of the most cost effective solution in the era of big data to solve this problem with massively parallel computing technology[1].

The CPU based algorithm are not fast enough to give a solution in a reasonable amount of time. These problems can become even larger to the point that not even a multi-core CPU based algorithm is fast enough. Today, these problems can be solved faster using massive parallel processing. These problem give a birth to an important hardware for parallel computing known as GPU [2]. The study uncovers the basic and the advanced topics that are required to understand parallel computing and especially GPU computing. The main contribution is in solving computational physics problems that requires massive parallel architecture. In general, data parallel issues are well suited for massively parallel computing in the GPU, that is the problem which divided into several independent sub-problem. This divided problem can be arbitrary distributed or a recursive divide and conquer strategy. Time dependent problems with small work can lead to inefficient performance and not well suited for GPU computing [3]. In this paper, it is revealed that the performance of two popular GPU integration tools developed in python. The GPU is becoming more popular when the running the machine learning applications including deep learning and computer vision. Cython and PyCUDA are the two most popular python-based framework that supports the high performance computing on the GPU. Cython is the static compiler that wraps C/C++/GPU code in the python. It introduced the Unified Memory feature. PyCUDA allows the user to directly access NVIDIA CUDA driver API and compile the kernel in a just in time fashion and also move data freely between python data objects and GPU memory. As the result Cython achieve comparable performance than PyCUDA.

Though the GPU is becoming the major component of high performance computing, computer architecture has already taken another steps towards more effective utilization of GPUs, in the form of integrated chips [4]. In this chip multicore-CPU and GPU both are integrated in silicon. This both device shares the same physical memory which makes it possible to copy data between device at high speed. The GPU does not have its own memory instead of it, it accesses the system memory with the CPU. Unlike the GPU, which is connected to the host through the PCI bus, the data transfer between memory and both processing units are through the same high speed bus which are controlled by a unified memory controller. The system memory is divided into two parts: the device memory and host memory. These two memories are designed for the GPU and CPU respectively, though both of them can be accessed by either device.

Sentiment Analysis is growing in the field of research with significant application in both industry and academic[5]. The most of the proposed solutions are centered around supervised, machine learning approaches and review oriented datasets. In this, I am going to focus on the informal communication such as online discussion, tweets and social network comments, unsupervised approaches that estimate the level of emotional intensity contained in the text in order to make predictions. In this experiment were made on the real world datasets, extracted from online social websites, the result will be very robust and reliable solutions for sentiment analysis of informal communication. The sentimental classification is the most useful application of sentiment analysis to classify review using sentiment analysis and classification are technically challenging [6]. The challenging task is to extract the information from the social networking sites and perform sentiment analysis. Using popular social media such as Facebook and Twitter, i have present new perspective to bring out more meaningful information about the social networks.

K-Nearest Neighbors (*k*-NNs) search is a popular and powerful distance based tool for solving classification problem [7]. It is the prerequisite for training local model based classifiers. Fast distance calculation can significantly improve the speed performance of these classifiers and GPUs can be very handy for their accelerations. Meanwhile, several GPU based sorting algorithms are also included to sort the distance matrix and seek for the k nearest neighbors. The speed performances of the sorting algorithms vary depending upon the input sequences. Another research deals with developing parallel processing algorithms for Graphic Processing Unit (GPU) in order to solve machine learning problems for large datasets [8]. In particular, it contributes to the development of fast GPU based algorithms for calculating distance matrix. It also presents the algorithm and implementation of a fast parallel Support Vector Machine (SVM) using the GPU. These application tools are developed using Computing Unified Device Architecture (CUDA), which is a popular software framework for General Purpose Computing using GPU (GPGPU). The GPU version of parallel SVM based on parallel Sequential Minimal Optimization (SMO) implemented in this dissertation is proposed to reduce the time cost in both training and predicting phases. This implementation of GPUSVM is original. It utilizes many parallel processing techniques to accelerate and minimize the computations of kernel evaluation, which are considered as the most time consuming operations in SVM. Although the many core architecture of GPU performs the best in data level parallelism, multi-task processing is also integrated into the application to improve the speed performance of tasks such as multiclass classification and cross-validation.

IV. METHODOLOGY

The methodology for software development will be as follows:

A. Acquire Social Media Data- To gather the social media data through the different websites of the social network sites using the API of the social media. To acquire the twitter data TwitterAPI is used. Tweepy is the python client for the official TwitterAPI.

B. Preprocess Data

In preprocessing, remove the unnecessary data like null values, hashtags, repeated letters, and URLs.

a. **Lower Case:** convert the tweets to lower case.
b. **URLs:** replace with generic world URLs
c. **@users:** I can eliminate "@users" via regex matching or replace it with generic core AT_user.
d. **Hashtag:** Hashtag can give us some useful information, so it is useful to replace them with the exact same words without the hash.

e. **e. Punctuations and Addition whitespace:** remove punctuation at the start and end of the tweets. It is also helpful to replace multiple whitespaces with the single whitespace.

C. Train and Test model using Algorithms-

a. Support Vector Machine Algorithm

b. K Nearest Neighbor Algorithm

D. Visualized Sentiment Analysis- Analyze the sentiment of the predicted values.

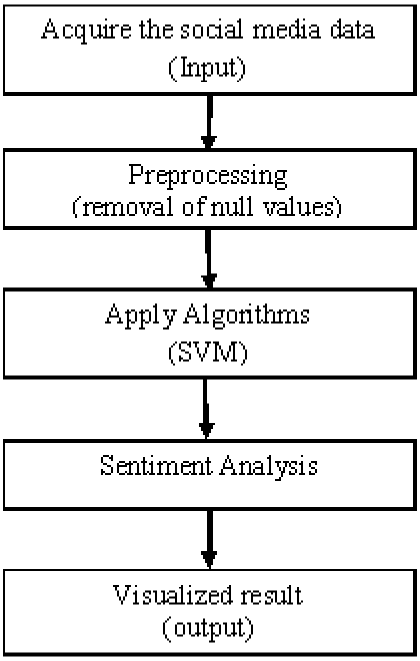
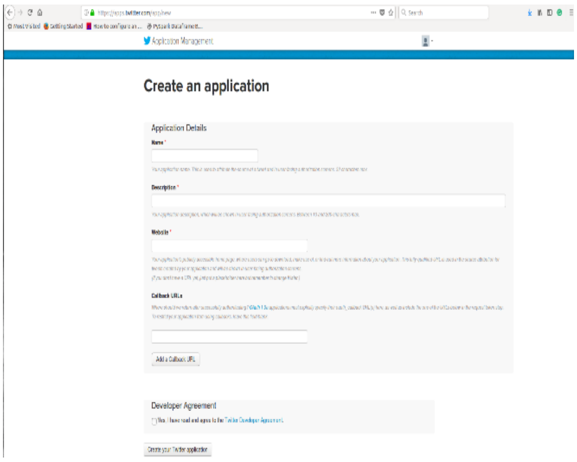


Figure 1. Flow of Methodology

V. IMPLEMENTATION

To acquire data, first an application is created to extract data from twitter. To start with the twitter you will need to have a twitter account and obtain credential information that is consumer keys, consumer secret, access token, access secret on the Twitter Developer site to access the Twitter API. If you already have a Twitter account then just go to https//apps.twitter.com/ and login with your twitter user account. This progression gives you a twitter development account under the same name as your user account. Then click to "Create New App" fill the form and click "Create your Twitter Application". Now in the next page, click on "keys and access tokens" and copy your API keys and "API secret", scroll down and click "create my access tokens", and your "access tokens" and "access secret".


https://apps.twitter.com/app/new

Application Management

Create an application

Application Details

Name *

Description *

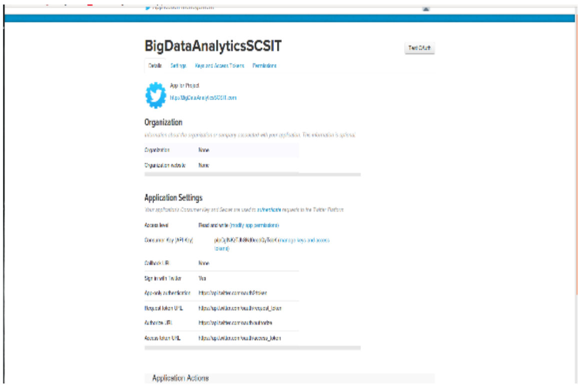
Website *

Callback URLs

Developer Agreement

□ Yes, I have read and agree to the Twitter Developer Agreement.

Figure 2. Create an Application


BigDataAnalyticsSCSIT

App for Project
https://BigDataAnalyticsSCSIT.com

Organization

Organization	None
Organization website	None

Application Settings

Access level	Read and write (modify app permissions)
Consumer Key (API Key)	[illegible] (manage keys and access tokens)
Callback URL	None
Sign in with Twitter	Yes
App-only authentication	https://api.twitter.com/oauth2/token
Request token URL	https://api.twitter.com/oauth/request_token
Authorize URL	https://api.twitter.com/oauth/authorize
Access token URL	https://api.twitter.com/oauth/access_token

Application Actions

Figure 3. Twitter Application

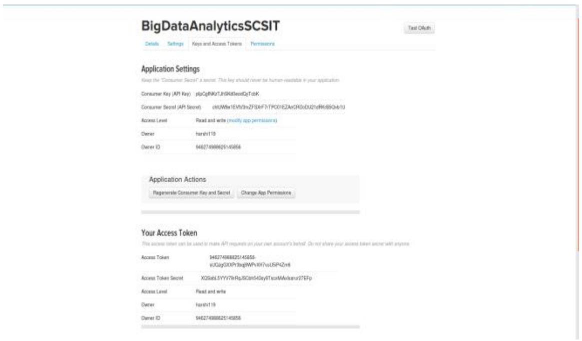
A. Data Access: To gather twitter data using the Twitter API, authentication keys are required to access the Twitter API. The Keys are:

a. Consumer Key

b. Consumer Secret Key

c. Access Token

d. Access Secret Token



BigDataAnalyticsSCSIT

Application Settings

Keep the "Consumer Secret" a secret. This key should never be human-readable in your application.

Consumer Key (API Key)	plpCgfNKrTJhSKd0eodQyTcbK
Consumer Secret (API Secret)	cWUW8e1EVfV9mZF8XrF7rTPC01EZAeCROeDU21dRK6B6Qvb1U
Access Level	Read and write (modify app permissions)
Owner	harshi119
Owner ID	946274988625145856

Application Actions

Your Access Token

This access token can be used to make API requests on your own account's behalf. Do not share your access token secret with anyone.

Access Token	946274988625145856- xUQJgGXXPr9bqjWPvXH7vuU5iP4Zm8
Access Token Secret	XQSabL5YYV7frRqJSCbh54Gsy9TscaMAvlsaruz27EFp
Access Level	Read and write
Owner	harshi119
Owner ID	946274988625145856

Figure 4. Twitter Authentication

B. Twitter Gathered Data:

Different Packages are used to extract the data:

a. Stream Listener: Create a class inheriting from Stream Listener
b. Using that class creates a stream object
c. Connect the TwitterAPI using the Stream.

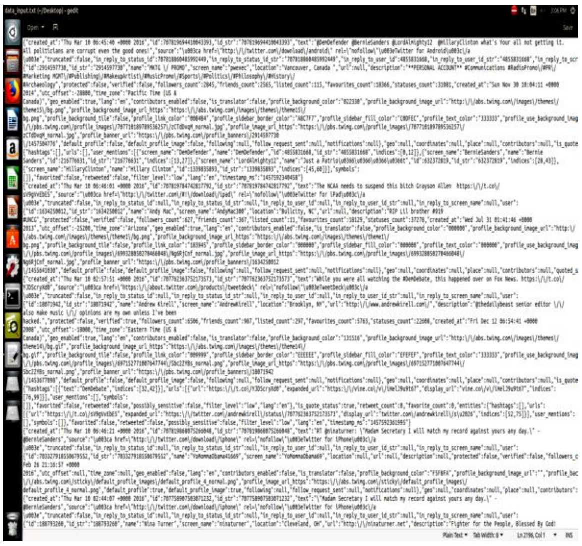


Figure 5. Twitter Data

C. Preprocess Data:

In the preprocessing, remove the un-necessary data like null values, hash tags, repeated letters and URLs.

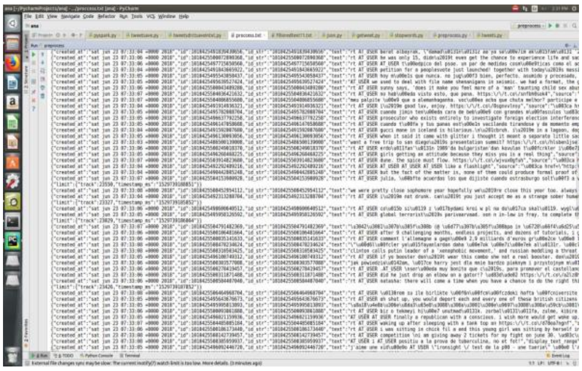
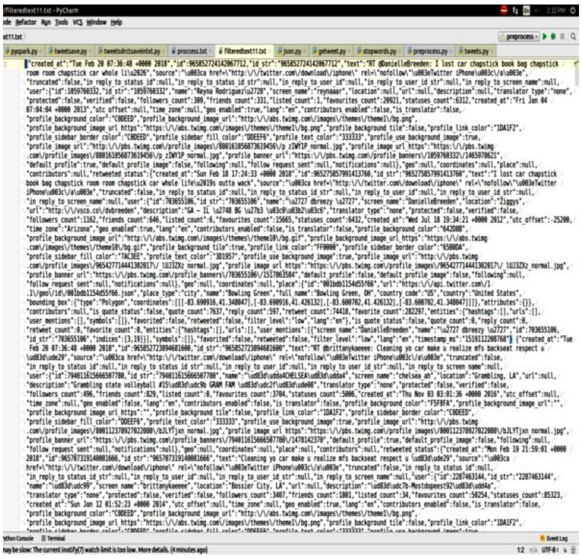


Figure 6. Preprocessing of Data

D. Removal of Stops Words:

'a', 'is', 'the', etc. are considered as stop words. These words don't indicate any sentiment and can be removed. The full list of stop words can be found at stop word list.



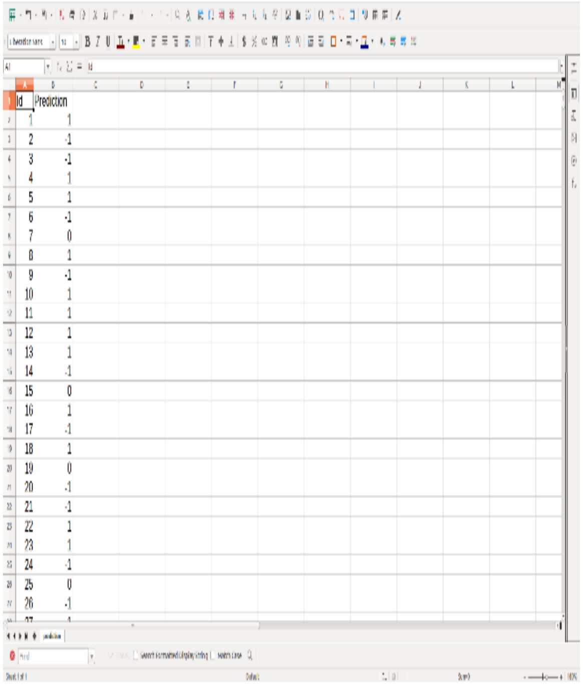
```
{"created_at":"Tue Feb 20 07:36:48 +0000 2018","id":965852724142067712,"id_str":"965852724142067712","text":"RT @DanielleBreeden: I lost car chapstick book bag chapstick
room room chapstick car whole li\u2026","source":"\u003ca href=\"http:\/\/twitter.com\/download\/iphone\" rel=\"nofollow\"\u003eTwitter 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":1059760332,"id_str":"1059760332","name":"Reyna Rodriguez\u2728","screen_name":"reynaaar","location":null,"url":null,"description":null,"translator_type":"none",
"protected":false,"verified":false,"followers_count":309,"friends_count":331,"listed_count":1,"favourites_count":20921,"statuses_count":6312,"created_at":"Fri Jan 04
07:04:04 +0000 2013","utc_offset":null,"time_zone":null,"geo_enabled":true,"lang":"en","contributors_enabled":false,"is_translator":false,
"profile_background_color":"C0DEED","profile_background_image_url":"http:\/\/abs.twimg.com\/images\/themes\/theme1\/bg.png",
"profile_background_image_url_https":"https:\/\/abs.twimg.com\/images\/themes\/theme1\/bg.png","profile_background_tile":false,"profile_link_color":"1DA1F2",
"profile_sidebar_border_color":"C0DEED","profile_sidebar_fill_color":"DDEEF6","profile_text_color":"333333","profile_use_background_image":true,
"profile_image_url":"http:\/\/pbs.twimg.com\/profile_images\/880161056873619456\/p_zIWY1P_normal.jpg","profile_image_url_https":"https:\/\/pbs.twimg
.com\/profile_images\/880161056873619456\/p_zIWY1P_normal.jpg","profile_banner_url":"https:\/\/pbs.twimg.com\/profile_banners\/1059760332\/1465978621",
"default_profile":true,"default_profile_image":false,"following":null,"follow_request_sent":null,"notifications":null},"geo":null,"coordinates":null,"place":null,
"contributors":null,"retweeted_status":{"created_at":"Sun Feb 18 17:24:33 +0000 2018","id":965275857991413760,"id_str":"965275857991413760","text":"I lost car chapstick
book bag chapstick room room chapstick car whole life\u2019s outta wack","source":"\u003ca href=\"http:\/\/twitter.com\/download\/iphone\" rel=\"nofollow\"\u003eTwitter
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":703655106,"id_str":"703655106","name":"\u2727 dbreezy \u2727","screen_name":"DanielleBreeden","location":"Ziggys",
"url":"http:\/\/vsco.co\/dvbreeden","description":"GA - IL \u2748 BG \u27b3 \u03c8\u03b2\u03c6","translator_type":"none","protected":false,"verified":false,
"followers_count":1162,"friends_count":646,"listed_count":6,"favourites_count":15665,"statuses_count":6432,"created_at":"Wed Jul 18 19:34:21 +0000 2012","utc_offset":-25200,
"time_zone":"Arizona","geo_enabled":true,"lang":"en","contributors_enabled":false,"is_translator":false,"profile_background_color":"642D8B",
"profile_background_image_url":"http:\/\/abs.twimg.com\/images\/themes\/theme10\/bg.gif","profile_background_image_url_https":"https:\/\/abs.twimg
.com\/images\/themes\/theme10\/bg.gif","profile_background_tile":true,"profile_link_color":"FF0000","profile_sidebar_border_color":"65B0DA",
"profile_sidebar_fill_color":"7AC3EE","profile_text_color":"3D1957","profile_use_background_image":true,"profile_image_url":"http:\/\/pbs.twimg
.com\/profile_images\/965427714441302017\/_lUJ3ZXz_normal.jpg","profile_image_url_https":"https:\/\/pbs.twimg.com\/profile_images\/965427714441302017\/_lUJ3ZXz_normal.jpg",
"profile_banner_url":"https:\/\/pbs.twimg.com\/profile_banners\/703655106\/1517863584","default_profile":false,"default_profile_image":false,"following":null,
"follow_request_sent":null,"notifications":null},"geo":null,"coordinates":null,"place":{"id":"001bdb1154d55f66","url":"https:\/\/api.twitter.com\/1
.1\/geo\/id\/001bdb1154d55f66.json","place_type":"city","name":"Bowling Green","full_name":"Bowling Green, OH","country_code":"US","country":"United States",
"bounding_box":{"type":"Polygon","coordinates":[[[-83.690916,41.340047],[-83.690916,41.426132],[-83.600702,41.426132],[-83.600702,41.340047]]]},"attributes":{}},
"contributors":null,"is_quote_status":false,"quote_count":7637,"reply_count":597,"retweet_count":74418,"favorite_count":282297,"entities":{"hashtags":[],"urls":[],
"user_mentions":[],"symbols":[]},"favorited":false,"retweeted":false,"filter_level":"low","lang":"en"},"is_quote_status":false,"quote_count":0,"reply_count":0,
"retweet_count":0,"favorite_count":0,"entities":{"hashtags":[],"urls":[],"user_mentions":[{"screen_name":"DanielleBreeden","name":"\u2727 dbreezy \u2727","id":703655106,
"id_str":"703655106","indices":[3,19]}],"symbols":[]},"favorited":false,"retweeted":false,"filter_level":"low","lang":"en","timestamp_ms":"1519112208768"} {"created_at":"Tue
 Feb 20 07:36:48 +0000 2018","id":965852723894681600,"id_str":"965852723894681600","text":"RT @brittanykaeeee: Cleaning yo car make u realize mfs backseat respect u
\ud83d\ude29","source":"\u003ca href=\"http:\/\/twitter.com\/download\/iphone\" rel=\"nofollow\"\u003eTwitter 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":794011615666507780,"id_str":"794011615666507780","name":"\ud83d\udda4CHELSEA\ud83d\udda4","screen_name":"chelsea_ah","location":"Grambling, LA","url":null,
"description":"Grambling state volleyball #15\ud83d\udc9b GRAM FAM \ud83d\udc2f\ud83d\ude08","translator_type":"none","protected":false,"verified":false,
"followers_count":496,"friends_count":829,"listed_count":0,"favourites_count":3704,"statuses_count":5006,"created_at":"Thu Nov 03 03:01:36 +0000 2016","utc_offset":null,
"time_zone":null,"geo_enabled":false,"lang":"en","contributors_enabled":false,"is_translator":false,"profile_background_color":"F5F8FA","profile_background_image_url":"",
"profile_background_image_url_https":"","profile_background_tile":false,"profile_link_color":"1DA1F2","profile_sidebar_border_color":"C0DEED",
"profile_sidebar_fill_color":"DDEEF6","profile_text_color":"333333","profile_use_background_image":true,"profile_image_url":"http:\/\/pbs.twimg
.com\/profile_images\/800112378927022080\/bJLYTjxn_normal.jpg","profile_image_url_https":"https:\/\/pbs.twimg.com\/profile_images\/800112378927022080\/bJLYTjxn_normal.jpg",
"profile_banner_url":"https:\/\/pbs.twimg.com\/profile_banners\/794011615666507780\/1478142370","default_profile":true,"default_profile_image":false,"following":null,
"follow_request_sent":null,"notifications":null},"geo":null,"coordinates":null,"place":null,"contributors":null,"retweeted_status":{"created_at":"Mon Feb 19 21:59:01 +0000
2018","id":965707319140081666,"id_str":"965707319140081666","text":"Cleaning yo car make u realize mfs backseat respect u \ud83d\ude29","source":"\u003ca
href=\"http:\/\/twitter.com\/download\/iphone\" rel=\"nofollow\"\u003eTwitter 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":2287463144,"id_str":"2287463144",
"name":"\ud83d\udc99","screen_name":"brittanykaeeee","location":"Bossier City, LA","url":null,"description":"\ud83d\udc7b-Mostdopeest92\ud83d\udd4e",
"translator_type":"none","protected":false,"verified":false,"followers_count":3407,"friends_count":1001,"listed_count":34,"favourites_count":50254,"statuses_count":85323,
"created_at":"Sun Jan 12 01:52:23 +0000 2014","utc_offset":null,"time_zone":null,"geo_enabled":true,"lang":"en","contributors_enabled":false,"is_translator":false,
"profile_background_color":"C0DEED","profile_background_image_url":"http:\/\/abs.twimg.com\/images\/themes\/theme1\/bg.png",
"profile_background_image_url_https":"https:\/\/abs.twimg.com\/images\/themes\/theme1\/bg.png","profile_background_tile":false,"profile_link_color":"1DA1F2",
```

may be slow: The current inotify(7) watch limit is too low. More details. (4 minutes ago)

Figure 7. Data without Stopwords

E. Applying SVM Algorithm

SVM Algorithm is applied on the dataset using machine learning libraries.



Id	Prediction
1	1
2	-1
3	-1
4	1
5	1
6	-1
7	0
8	1
9	-1
10	1
11	1
12	1
13	1
14	-1
15	0
16	1
17	-1
18	1
19	0
20	-1
21	-1
22	1
23	1
24	-1
25	0
26	-1

Figure 8. Predicted Values

F. Visualized Output

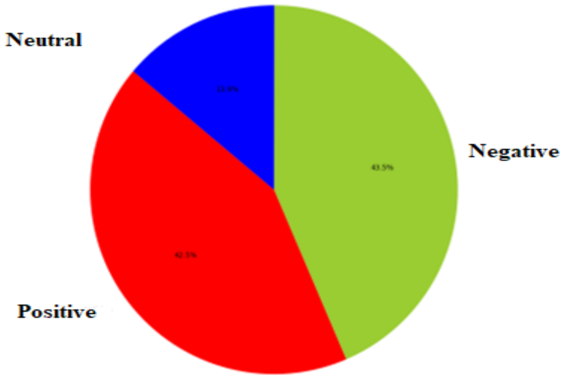


Figure 9. Visualization

Results:

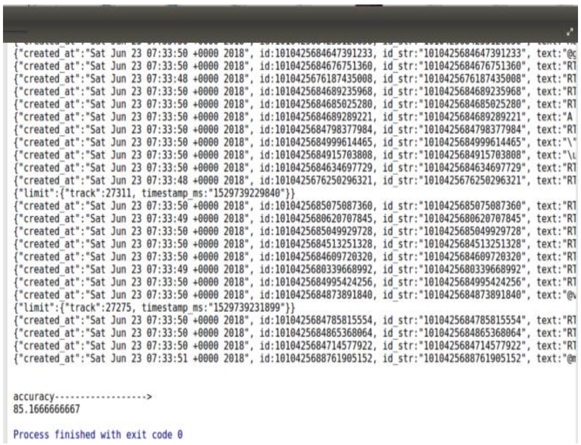
42.5% Positive Sentiments

43.3% Negative Sentiments

13.9% Neutral Sentiments

G. Applying k-NN Algorithm

K-Nearest Search Algorithm is applied on the dataset by importing the machine learning libraries as KNeighborsClassifier. After that, training data is fitted in as a parameter for K-NN. It learns about the data and give the accuracy of the data.



```
{"created_at":"Sat Jun 23 07:33:50 +0000 2018", id:1010425684647391233, id_str:"1010425684647391233", text:"@g
{"created_at":"Sat Jun 23 07:33:50 +0000 2018", id:1010425684676751360, id_str:"1010425684676751360", text:"RT
{"created_at":"Sat Jun 23 07:33:48 +0000 2018", id:1010425676187435008, id_str:"1010425676187435008", text:"RT
{"created_at":"Sat Jun 23 07:33:50 +0000 2018", id:1010425684689235968, id_str:"1010425684689235968", text:"RT
{"created_at":"Sat Jun 23 07:33:50 +0000 2018", id:1010425684685025280, id_str:"1010425684685025280", text:"RT
{"created_at":"Sat Jun 23 07:33:50 +0000 2018", id:1010425684689289221, id_str:"1010425684689289221", text:"A
{"created_at":"Sat Jun 23 07:33:50 +0000 2018", id:1010425684798377984, id_str:"1010425684798377984", text:"RT
{"created_at":"Sat Jun 23 07:33:50 +0000 2018", id:1010425684999614465, id_str:"1010425684999614465", text:"\"
{"created_at":"Sat Jun 23 07:33:50 +0000 2018", id:1010425684915703808, id_str:"1010425684915703808", text:"\u
{"created_at":"Sat Jun 23 07:33:50 +0000 2018", id:1010425684634697729, id_str:"1010425684634697729", text:"RT
{"created_at":"Sat Jun 23 07:33:48 +0000 2018", id:1010425676250296321, id_str:"1010425676250296321", text:"RT
{"limit":{"track":27311, timestamp_ms:"1529739229840"}}
{"created_at":"Sat Jun 23 07:33:50 +0000 2018", id:1010425685075087360, id_str:"1010425685075087360", text:"RT
{"created_at":"Sat Jun 23 07:33:49 +0000 2018", id:1010425680620707845, id_str:"1010425680620707845", text:"RT
{"created_at":"Sat Jun 23 07:33:50 +0000 2018", id:1010425685049929728, id_str:"1010425685049929728", text:"RT
{"created_at":"Sat Jun 23 07:33:50 +0000 2018", id:1010425684513251328, id_str:"1010425684513251328", text:"RT
{"created_at":"Sat Jun 23 07:33:50 +0000 2018", id:1010425684609720320, id_str:"1010425684609720320", text:"RT
{"created_at":"Sat Jun 23 07:33:49 +0000 2018", id:1010425680339668992, id_str:"1010425680339668992", text:"RT
{"created_at":"Sat Jun 23 07:33:50 +0000 2018", id:1010425684995424256, id_str:"1010425684995424256", text:"RT
{"created_at":"Sat Jun 23 07:33:50 +0000 2018", id:1010425684873891840, id_str:"1010425684873891840", text:"@v
{"limit":{"track":27275, timestamp_ms:"1529739231899"}}
{"created_at":"Sat Jun 23 07:33:50 +0000 2018", id:1010425684785815554, id_str:"1010425684785815554", text:"RT
{"created_at":"Sat Jun 23 07:33:50 +0000 2018", id:1010425684865368064, id_str:"1010425684865368064", text:"RT
{"created_at":"Sat Jun 23 07:33:50 +0000 2018", id:1010425684714577922, id_str:"1010425684714577922", text:"RT
{"created_at":"Sat Jun 23 07:33:51 +0000 2018", id:1010425688761905152, id_str:"1010425688761905152", text:"@m


accuracy------------------>
85.1666666667

Process finished with exit code 0
```

Figure 10. Accuracy of KNN

Result:

Accuracy= 85.1666666667

VI. CONCLUSION

Sentiment Analysis based on social media data is carried out and Positive, negative and neutral sentiment of the Twitter data were found out. Machine learning algorithm is used to classify the data and to analyze the Twitter data. Support Vector Machine is used to analyze the sentiments of the data using GPU. KNN algorithm is used to calculate the accuracy of the data.

VII. FUTURE WORK

GPU has brought an opportunity of accelerating many applications to various problems. In the future, we can add some extra features like analyzing the data with images or emotions tweets. We can also try to focus on the other sentiments of the data like very positive or very negative.

VIII. REFERENCES

[1]. Li, Shengren, and Nina Amenta. "Brute-force k-nearest neighbors search on the GPU." International Conference on Similarity Search and Applications. Springer, Cham, 2015.

[2]. Navarro, Cristobal A., Nancy Hitschfeld-Kahler, and Luis Mateu. "A survey on parallel computing and its applications in data-parallel problems using GPU architectures." Communications in Computational Physics 15.2 (2014): 285-329.

[3]. Accelerating Machine Learning Algorithms in Python Patrick Reilly, Leiming Yu and David Kaeli, Department of Electrical and Computer Engineering Northeastern University, Boston, MA

reilly.pa@husky.neu.edu,fylm,kaelig@ece.neu.edu.

[4]. Chen, Linchuan, Xin Huo, and Gagan Agrawal. "Accelerating MapReduce on a coupled CPU GPU architecture." Proceedings of the International Conference on High Performance Computing, Networking, Storage and Analysis. IEEE Computer Society Press, 2012.

[5]. Li, Peilong, et al. "Heterospark: A heterogeneous cpu/gpu spark platform for machine learning algorithms." Networking, Architecture and Storage (NAS), 2015 IEEE International Conference on. IEEE, 2015.

[6]. Bello-Orgaz, Gema, Jason J. Jung, and David Camacho. "Social big data: Recent achievements and new challenges." Information Fusion 28 (2016): 45-59.

[7]. Huq, Mohammad Rezwanul, Ahmad Ali, and Anika Rahman. "Sentiment analysis on Twitter data using KNN and SVM." Int J Adv Comput Sci Appl 8.6 (2017): 19-25.

[8]. Li, Qi, et al. "GPUSVM: a comprehensive CUDA based support vector machine package." Open Computer Science1.4 (2011): 387-405.

[9]. Thambawita, D. R. V. L. B., Roshan Ragel, and Dhammika Elkaduwe. "To use or not to use: Graphics processing units (GPUs) for pattern matching algorithms." Information and Automation for Sustainability (ICIAfS), 2014 7th International Conference on. IEEE, 2014.