Sentiment Analysis on Facebook Comments

Aijith S Poriyath[1], Amitha Joseph[2]

[1]UG Scholar, Department of Computer Science,
Santhigiri College of Computer Sciences, Vazhithala
aijith0407[at]gmail.com

[2]Assistant Professor, Department of Computer Science,
Santhigiri College of Computer Sciences, Vazhithala
amithajoseph[at]santhigiricollege.com

Abstract: *Sentiment Analysis has a big role in computer science. Because in current world the peoples usually buy items or even see a movie based on their review. Sentiment analysis is a process of opinion mining. That is mining useful information from an opinion. Opinion mining is basically web content mining. At the same time sentiment analysis is application of Natural Language processing. The computer can process the digital information and covert to the feeling in the real world. In other words, computer gives an opportunity for the peoples to share their opinion in correct feeling. Facebook is one of social media that is widely used in the entire world. Today about 3.5 billions of peoples are active Facebook users. They put status, post and comments in Facebook. Facebook allows the user to post short messages called as comments. These comments can be140 characters in length. So that there can be trillions of Facebook posts are posted by different users in each day. Here the question is the machine able to detect the sentiment in the data's. The answer is completely 'Yes'. This paper has an objective to implement sentiment analysis on the data from Facebook using python. Python is one of the popular languages that have rich set of libraries. The dataset is collected using Graph API and it is refined and stored using Open Refine. Each comment in Facebook will have a polarity of negative, neutral, and positive. The paper gives an idea about how sentiment analysis works using VADER. Life is digitalizing day by day. Many people's that express their views now through Facebook. Facebook give a virtual platform to the people to express their views. Today many origination based on Facebook or social media. Through the response of posts about their organization they can evaluate public view of the society through sentiment analysis. We can apply sentiment analysis on Facebook comments of the post. Through that a person gets a better understanding how the public will treat his post that will help to make their activities better. This paper also describes usefulness of sentiment analysis in Facebook.*

Keywords: VADER, Graph API, Open Refine, SentimentIntensity Analyzer, Pandas, nltk, Facebook

1. Introduction

Social media is a website or application which having digitally-mediated technologies that facilitate the creation or sharing/exchange of contents, thoughts, interests, and other forms of idea via virtual space and networks. This is the highly interactive platforms through which many people can share, create, discuss, participate, and modify the content or self created content posted online. Some of the example for the social media websites, with over 100 million registered users, includes Facebook, TikTok, Instagram, QZone, Weibo, Twitter, Tumblr, Baidu Tieba, and LinkedIn. This also consist of social media services include YouTube, Quora, Telegram, WhatsApp, Snapchat etc. This also includes bookmarking sites and news websites. With the innovations in the field of technology and availability internet connections low rate increase the popularity of social media. We are connected in all the time, when we texting each other on Whatsapp or checking out video and pictures of cute animals on Facebook etc. All the status, videos, pictures, that posted by public on their social media account contains data — data about their style, mentality and how other treat them. This information can be analysed. its applications are numerous. Apart from all the social media Facebook have some popularity over others. Facebook connect all people in this world. Facebook has helped to building a brand position in businesses. Moreover, it's a major source of entertainment for all people in this world. Since Facebook is a global social networking site available in almost all the continents, so location is not consider as a barrier. It has many advantages. There are thousands of Facebook posts. It is true that there is no end for the comments. But is it positive or negative? No one knows. In this world one of the most important thing the corporate valued is the data. Social media sentiment is the way in which people treat your brand and your organisations

The creations of large number databases for compensating the real world need arise the need for data mining. Sentiment analysis in Facebook comments is a process of opinion mining. Mining the use full information from opinion. At the same time sentiment analysis is application of Natural Language processing [1]. The computer can process the digital information and covert to the feeling in the real world. In other words, computer gives an opportunity for the peoples to share their opinion in correct feeling. Data mining is a process of extraction of useful information from huge amount of raw data. For example if you keep track of all your expenditure for a month and analyzing this result will help you to find out reduce the unnecessary expenditure for the next month. This process in computer science is called Data mining. It is also called a knowledge discovery process or may it be called knowledge mining from data. One of Data Mining is the Web Mining. In Web Mining also you there is Web usage mining, Web content mining and Web Structured mining. Web mining is mainly related to internet data [1].It is one of big field. In this paper we focus on opinion mining. One of the sub fields of web content mining. Data is precious. Mine in this data will helpful for the people to analyze some important information which useful. It may be a solution or some answer for the questions. Many IT today invest billions in the field of data mining. This will show how important the data mining. Sentiment analysis on Facebook comments which we perform here is considered as a process of opinion mining. The broadest classification of data mining is show below.
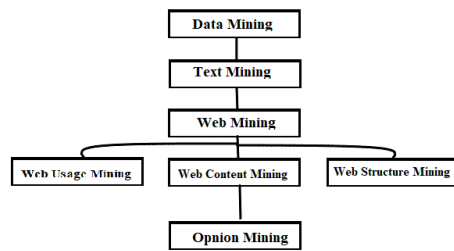
Figure 1: Classification of Data Mining

Sentiment analysis that we performed on Facebook comments will be useful for many companies. This will help companies, government and individual to get a awareness about how the people think about them and take about their posts, topic and organzation. Currently Facebook didn't provide this feature for the page users. Sentiment Analysis refers to the process which extract the opinions from a specific piece of text written in human language. In other words ii can be defined as process to finds out whether the particular piece of text is positive, negative, or neutral. It is an automated process capable of understanding the feelings or opinions that underlie a text. It is also one of the most interesting sub fields of NLP, a branch of Artificial Intelligence (AI) that focuses on how machines process human language. People make judgments about the all things around them when they are living in this world [2]. They make positive, negative and neutral attitudes about people, products, topics and events. These types of attitudes can be considered as sentiments That tree is ugly, That flower smells nice. Here we have to focusing on the keywords: ugly and nice. Here we know the first sentence is negative because it uses 'ugly.' The same get repeated in second sentence. Comments expressed on the social media are the most powerful forms of feedback for businesses because they are given facts. But the problem is that a person can't categories this whole comments. That must be a time consume task. The only solution is to use computer for this. Sentimental analysis to analyse social media posts, product reviews, and news websites, and capture the emotions behind what people are talking about. Businesses can also use it to analyse all sorts of internal data, like NPS surveys, employee surveys, and customer support chats. Whether positive or negative, opinions are powerful. In the digital world, they can grow or cause trouble to a brand's reputation. Stats show that 40% of buyers form an opinion of a business after reading 1-3 online reviews, and that 64% software buyers read at least 6 online reviews before making a purchase. We live in a world where 3.5 billion people are active social media users; that's 45% of the world's population! Every single minute of the day, people write more than 500, 000 Tweets and 510, 000 Face book comments [5]. Sentiment analysis allow us to mine this data and extract the feelings that underlie social media conversation, to understand how people are talking about a given product or topic. Sentiment analysis can be performed in two ways first one is the lexicon based method and machine learning approach. In Lexicon based approach we will use the dictionary with which contain a collection of words and its sentiment intensity. In machine learning approach we train the system to find the sentiment a statement [5]. Many people's that express their views now through Facebook.Facebook give a virtual platform to the people to express their views. Today many origination based on Facebook or social media. We can apply sentiment analysis on Facebook comments of the post. Through that a person gets a better understanding how the public will treat his post that will help to make their activities better. This paper 'Sentiment Analysis on Facebook Comment' discuss the importance of sentiment analysis on Facebook comment and also try to discuss how to done sentiment analysis by using python. Python is one of the most popular programming language that will provide many features. We uses python special featured library called 'nltk' to categories our comments. And also VADER analyser to find intensity of the comments. Throughout the document we discuss the how to done sentiment analysis using python and importance of sentiment analysis in this information age.

2. Working Procedure

Sentiment analysis is a simple process to find out whether a piece of text is positive, negative or neutral. The sentiment analysis approaches take one of two forms. One is polarity-based i.e, the pieces of texts are classified as either positive or negative and the other is valence-based, which the intensity of the sentiment is taken into account. For an example, the words 'good', 'excellent' has the same in the polarity-based approach, and 'excellent' would has more positivity than 'good' in the other approach. It is process of find an emotion of a pieace of text. Sentiment analysis has applications across a scope of space - it's extraordinary for anything where you can hear unstructured point of view information about an assistance or item. One application of sentiment analysis is for companies that have social media accounts to receive feedback. Clearly it's terrible business for these organizations to leave negative input unanswered excessively long, and sentiment analysis can give them a brisk method to discover and needs these miserable clients. There need to follow a set of steps to perform sentiment analysis on Facebook comments. As we know Facebook is one of the large company that keep the user data private. The user even can't download his own data from the Facebook. The user can request for a report for all account activity from Facebook. If Facebook accept the request and then we are able to download our feeds and posts data. But then also we are not able to access the comments. It is necessary to use some developing tools. In order to perform the sentimental analysis we are using python programming language. The syntax and build of Python applications resemble object-oriented languages like C, C++ & Java. It can be efficiently used for data analysis tasks that are synced with a web application. Python is a object oriented programming language and a great tool to execute algorithms. Python exclusive feature called VADER (Valence Aware Dictionary and sentiment Reasoner) offers the great advantages in sentiment analysis. But before that need to collect the data. Sentiment analysis on Facebook comment is tedious task with technology. we need to use different software to perform this. So that we need a set of steps need to pass before classifying the number of positives and negative and neutral statements. This steps are shown below
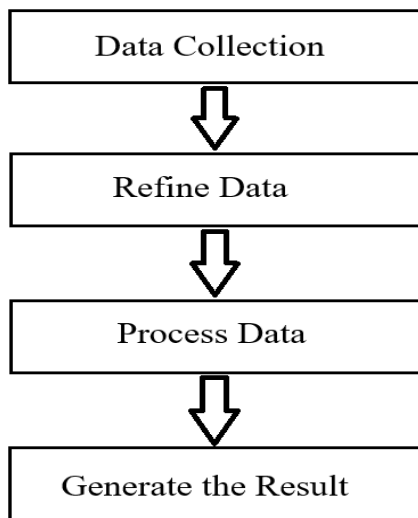
Figure 2: Steps for sentiment analysis using python VADER

2.1 Data Collection

Data collection is one of the important step to perform the sentiment analysis on Facebook comment. Facebook is a social media that keep all information to be private. Even the original users can't download there chats and comments of the posts directly from the Facebook. It was also one reason for lessening research in the field of Facebook. So that we will using Graph API. Graph API is a web application developed by Facebook with tab line Facebook for developers to get data into and out of the Facebook. It's an HTTP based API that can be use to manage profiles and upload photos, and perform variety of tasks programmatically. You can use Graph API to get your own Facebook account data. But if you need to get other users data then for this you need to take several permissions from users. You need to use OAuth protocol to achieve this. Through that anyone can authenticate and grant your permissions, after this you will get a Access Token. Access tokens allow you to access the Graph API. They typically perform two functions. They allow you to access a User's information without requiring the User's password and they allow us to identify your app, the User who is using your app, and the type of data that the User has permitted your to access. Access is token is nothing but a mechanism to make us authentic our self with the Graph API.
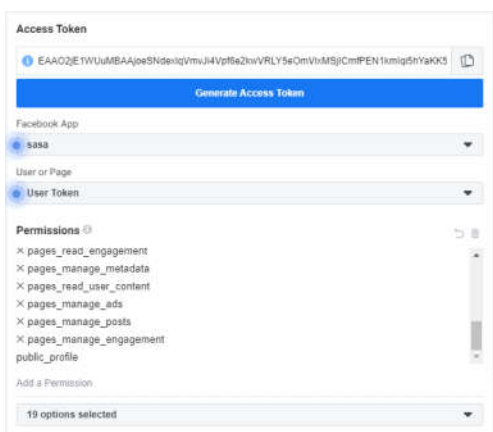

Figure 3: Access token in Graph API

Your are not only required the access token but also you should provide the Facebook page id with the post id separated by underscore to access the post. And also you need to specify the access is to comment by adding /comments at the end of the request. Page id can mostly find out by using findmyfbid.com. And the post id always given inside the url.
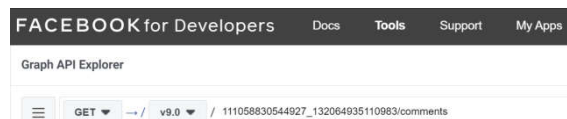

Figure 4: Example request to access the comments for a post

The data which is it produce which contain lot of information and not in suitable form which we can read. To read the data we need to convert into excel form.

Figure 5: Retrived comments using Graph API

The data which we received is stored inside a file with extension.JSON.

2.2 Refine Data

Next step in this process is refine the collected data. The data which we collected is in the form which shown in Figure 5. We need to convert the data into a form that we able to read to our program. The appropriate form is in the excel format. For that purpose we uses a software called open refine. Open Refine is a powerful tool for working with the data which is mess and clean it. It also help to transform one form to another. It also help to convert into another form and extending it with web services and external data. It always keeps the data private on your own computer until you want to share/Collaborate. Your private data never deletes from your computer unless you don't want it to. It works by running a small server on the computer and use web browser to interact with. The software open refine is available in more than 15 languages. This will operated on rows of information which have cells under columns. Which is very similar to the relational database tables. An Open Refine project consists of one table. Here users can clean the rows using the facets that define filtrations criteria. That is for an example, showing rows where a given column is not empty. Not like spreadsheets, many operations in Open Refine are done on visible rows: transformation of all cells in all rows under one column, creation of a new column based on existing column information etc. All actions that were

performed on the datasets are stored in the projects and can be replayed on another datasets.
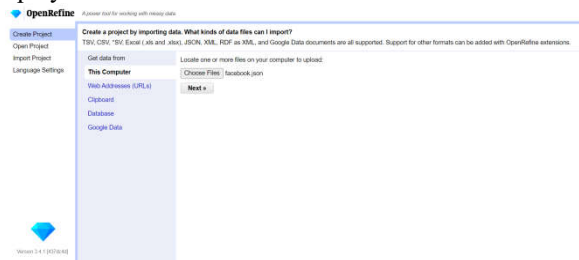

Figure 6: Interface of Open Refine

After you launch the open refine you got an interface that like in the above window. In that you were able to upload the document which you saved in JSON format. After upload the file you can see the data inside the file. Then you need to specify the delimiter of data which need to convert by free selection and that much of data got converted to excel preview.
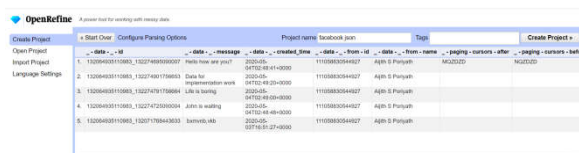

Figure 7: Data in row and column format inside open refine

After you got an excel preview you were able to create a project to click create project and export the data to excel file and able to stored it to the computer.

2.3 Process Data

After you refine the data you need to process the data. That means classify the comments based on positive, negative and neutral. We are using a lexicon based approach, so we uses the VADER lexicon. Lexicon act as a dictionary here. For processing our data first we need to read our excel file to our program. After that for manage it efficiently we convert the stored data in the excel file in to data frame in python. DataFrame is data structure available in python. It is available in the pandas library. Pandas library support data frame. DataFrame is two-dimensional data dtructure where size is mutable. Mainly they are heterogeneous tabular data structure with labelled axes. It is a two-dimensional data structure. That means data is aligned in a tabular form in rows and columns. DataFrame available in pandas consists of three principal components they are data, rows, and columns. Normally a Pandas DataFrame is used for storing data which is from existing other storage on computer. It can be SQL database, CSV or Excel files. It can be also created from the lists, dictionary, and from a list of dictionary etc.

Next we need to find the polarity score of each comment. After read the file and stored it in the data frame you can easily read each comment and find polarity of each comment. This can be achieved by using SentimentIntensityAnalyzer which we import from nltk.sentiment.vader.VADER - Valence Aware Dictionary and sEntiment Reasoner is a lexicon and rule-based sentiment analysis tool that is specifically focused to sentiments expressed in social media like facebook, instagram etc. This is used for sentiment analysis of text

which has both the polarities, positive or negative. VADER is used to know how much of positive or negative emotion or feeling the text has and also the intensity of the emotion. VADER's SentimentIntensityAnalyzer which is a class takes a string and returns a dictionary of scores for each of the four categories Positive, Negative, Neutral and Compound. VADER sentimental analysis relies on a dictionary that maps word to emotion intensities known as sentiment scores. This sentiment score for a text can be obtained by summing up the intensity of each word in the text. VADER belongs to a type of sentiment analysis that is based on lexicon. In this approach, each of the words in the lexicon is rated as to determine whether it is positive or negative. Below you can see an sample from VADER's lexicon where the positive words have positive ratings and negative words have negative ratings.

Table 1: Word and it's sentiment rating

Word	Sentiment Rating
tragedy	-3.4
rejoiced	2.0
insane	-1.7
disaster	-3.7
great	3.1

The lexicon needs to have a huge collection of the words in your text, otherwise it is not very accurate. On the other side, when there is a good fit between the lexicon dictionary and the text, then this approach is accurate well, and additionally quickly returns the results even on a large amounts of text. Incidentally the developers of Valence Aware Dictionary and Sentiment Reasoner used Amazon's Mechanical Turk to get most of the ratings, which is a very quick and cheap way to get their ratings. when Valence Aware Dictionary and Sentiment Reasoner analyses a piece of text it checks to see if any of the words in the text are present in the lexicon. For example, consider a sentence like "The food is good and the atmosphere is nice" this has two words in the lexicon say good and nice with rating of 1.2 and 1.3 respectively. VADER produces four category of values called sentiment metrics from these word rating, which will shown below. The first three, positive, neutral and negative, represent the ratio of the text that falls into those categories. As you can see, our example sentence was rated as 45% of it is positive, 55% of its is neutral and 0% of it is negative. The final metric is the compound score, which is the sum of all of the lexicon ratings which have been standardised to range between -1 and 1. In this case, our example sentence has a rating of 0.69, which means the sentence is strongly positive.

Table 2: Same statement intensities

Sentiment Metric	Value
Positive	0.45
Neutral	0.55
Negative	0.00
Compound	0.69

Consider a Statement "Life is boring" of course it is negative statement. Figure 6 show the result produced after processing the statement.

```
Life is boring
('neg', 0.535)
('neu', 0.465)
('pos', 0.0)
('compound', -0.3182)
```

Figure 8: Sentiment Intensities of a statement

2.4 Generate Result

This is the last stage of sentimental analysis. Find out the negativity, positivity and neutrality of a sentence doesn't provide meaning full information to the user. A user didn't get any meaning full information after find out how much percent each comment is positive, negative and neutral. It is necessary to count the number of positive and number of negative and number of positive comment in the data set. This will helpful for the user to analyse the data and help them to generate many decisions from the result. For this we will look to the compound. We will check if the compound is greater than 0 then we count to positive or if it is less than zero we count to negative else we count as neutral comment. Figure 9 that show the sample output that is produces after passing 5 comments.

```
Posetive Comments
0
Negative Comments
1
Neutral Comments
4
Total Comments
5
```

Figure 9: Sample result obtained

3. Advantage of this Approach

- For many purposes, such as evaluating public opinions or enhancing customer experiences, this approach is easy to understand.
- This lexical approach is very quick to implement, requiring just readily available libraries and code have less number of lines.
- Easy to capture the dataset for this analysis. We uses Graph API.
- It's efficient at analysing large datasets as because all data analysed in same way.
- It does not require any training data as in the case of machine learning approach.
- It can analyse the sentiment of a text containing emotions, slang, emojis, punctuation and much more.
- It works perfectly on social media text.
- It can work with multiple domains.
- It is fast enough to be used online with streaming data as because each time we download the Vader lexicon.
- It will perform the operation very fastly.

4. Disadvantage of this Approach

- Misspellings and grammatical mistakes may cause the final result may be in accurate
- Sarcasm and irony may be misinterpreted and result may not be accurate.
- Analysis is done as in language-specific. It take care of only English.
- jargon, nomenclature, memes, or turns of phrase may cause the result to change.

5. Application of Sentiment Analysis on Facebook Comments

Sentiment analysis is highly useful in social media monitoring as it allows us to gain an overview of the public opinion behind certain topics that the organization or company focus. The applications of sentiment analysis in business are broad and powerful. The ability to extract useful data from social data is a practice that is being widely used by many organizations across the world. Given below the area in which our approach is applicable.

5.1 In Business

In business field companies use it to develop their own strategies, to understand customers reaction towards products or brand, how they respond to their campaigns or product launches and why customers don't buy some products. we can also look how it help on business [8].

5.1.1 Understand your audience.

Marketers do the best works to raise their business when they understand their audience. That means you need to understand how your audience feels about your brand, your Facebook posts and their comments. Ongoing social media sentiment analysis help you to quickly know trends at the time customer preferences and desires change [4].

5.1.2 Under where you stand on your niche.

Brands cannot be all things to all people. Monitoring social sentiment that can help the organization understand where you stand in your business niche [8]. If number of positive comment increase they know they are growing.

5.1.3 Spot brand crises on early stages.

You never like your brand to fall into a crisis. Of course the business men's work spot brand crises early stage. Monitoring social sentiment can help you early to spot the problem. You can implement your own crisis response plan to minimizing negative sentiment or avoid it entirely

5.2 In Politics

In the political field, it is necessary to keep track of political view for detect consistency and inconsistency between statements and action performed by the government. Today all things which we evaluate through the social media mindset. Because social media is the platform that help us to understand carefully about all

things. Today politics is a field that need lot of predictions. The predictions are based on the number of followers, likes and comments. From this comments are more important evaluating data. But it is not possible for a human to classify all this or predict the result after reading some of the comments. Here the approach of sentiment analysis on this Facebook comment that will help to take a better decision and got a good analysis about the party. It can even may be used to predict election results as well!. The Obama administration that uses the sentiment analysis to gather public opinions to policy announcements and campaign messages ahead of 2012 election. Being able to quickly see the sentiment behind everything from forum posts to news articles means being better able to strategies and plan for the future..

5.3 For Individuals

Social media a wide platform that now a huge number of people find their income. Today being a social media influencer is consider one of the biggest provision that can attain. Many people that support social media for their day to day life. Facebook is also one of the biggest platform for earn. So they need to maintain a good relationship between all their followers and other users of social media. Sentiment analysis on the comment of Facebook posts that help them to evaluate the posts and at the same time it will help them maintain or improve their approaches.

6. Conclusion

In this study, the NLTK and the VADER analyser were applied to conduct a sentiment analysis. This approach is easy to use and time consuming method. The results indicated that the VADER Sentiment Analyser was an effective choice for sentiment analysis classification using Facebook Comments. VADER is quickly and easily classified huge amounts of data. However, the present approach has the following limitations. First a general lexicon was used to categorise specific data. Third, the data were not trained. The world now accumulated with huge amount of data. Each day there is tremendous amount of data that is accumulated in the social media sites. The methods' to find useful information from this data that help to speedy decision making higher productivity in various fields. Facebook is Multilingual social media platform. This is one of biggest challenge to perform the sentiment analysis on Facebook comment. Sentiment analysis on Facebook data provide variety opportunities in business, politics and for the individuals. Sentiment Analysis can apply in any form of data collected from the social media not only the comments. You can apply in your status set, chat, etc. Apply sentimental analysis on the chat will also help to understand the mentality of each chat group. This will be helpful when social media that will uses as a e-learning platform [5]. VADER classifies the sentiments very well. It is always available and easy to use, ready-made model which can be used across multiple sectors such as, social-media texts, analysing reviews etc. Advantage of using VADER is that it does not require any training data. Well, we can see that the results obtained are very much accurate.

References

[1] Rupinder Kaur, Dr. Harmandeep Singh, Dr. Gaurav Gupta, 2019, Sentimental Analysis on Facebook comments using Data Mining Technique, INTERNATIONAL JOURNAL OF ENGINEERING RESEARCH & TECHNOLOGY (IJERT) Volume 08, Issue 08 (August 2019),

[2] Sharef, Nurfadhlina & Mat Zin, Harnani & Nadali, Samaneh. (2016). Overview and Future Opportunities of Sentiment Analysis Approaches for Big Data. Journal of Computer Science. 12. 153-168. 10.3844/jcssp.2016.153.168.

[3] Taboada, Maite. (2016). Sentiment Analysis: An Overview from Linguistics. Annual Review of Linguistics. 2. 10.1146/annurev-linguistics-011415-040518.

[4] Anton Borg, Martin Boldt, Using VADER sentiment and SVM for predicting customer response sentiment, Expert Systems with Applications, volume 162, 2020, 113746, ISSN 0957-4174,

[5] Ortigosa, Alvaro & Martin, Jose & Carro, Rosa. (2014). Sentiment analysis in Facebook and its application to e-learning. Computers in Human Behavior. 31. 527-541. 10.1016/j.chb.2013.05.024.

[6] Ashwini Patil, Sanket Thakare, Analyzing Public Sentiment Variations on Twitter and Facebook, INTERNATIONAL JOURNAL OF ADVANCED RESEARCH IN COMPUTER SCIENCE & TECHNOLOGY (IJARCST 2017) Volume 05, Issue 02 (Apr. - Jun. 2017)

[7] Troussas, Christos & Virvou, Maria & Espinosa, Kurt & Llaguno, Kevin & Caro, Jaime. (2013). Sentiment analysis of Facebook statuses using Naive Bayes Classifier for language learning. 4. 1-6. 10.1109/IISA.2013.6623713.

[8] Umman Tugba Gürsoy, Diren Bulut, Cemil Yiğit. Social Media Mining and Sentiment Analysis for Brand Management. Global Journal of Emerging Trends in e-Business, Marketing and Consumer Psychology (GJETeMCP) An Online International Research Journal (ISSN: 2311-3170) 2017 Vol: 3 Issue: 1

Author Profile

Aijith S Poriyath Pursuing Bachelor of Computer Application from Santhigiri college of Computer Sciences, Vazhithala in 2018-2021.Presented a paper in National IT Conference about "Quantum Computing and Quantum Cryptography" held at Santhigiri College of Computer Sciences, Vazhithala.

Amitha Joseph received the MCA professional degree and MPhil in Computer Science. Data Mining is the area of specialization. Currently working as Asst. Professor in Santhigiri College of Computer Sciences, Vazhithala.