Journal of Engineering Sciences

Vol 13, Issue 05, MAY/2022

ISSN NO:0377-9254

SAAS PRODUCT COMPARISON AND REVIEWS USING NLP

Syed Shehriyar Ali, BE, Department of CSE

syedshehriyarali89@gmail.com

Mohammed Sarfaraz Shaikh, BE, Department of CSE

sarfaraz.mohsin21@gmail.com

Syed Safi Uddin, BE, Department of CSE

syed85150@gmail.com

Dr. Mohammed Abdul Bari, BE, M.Tech, M.Sc (UK), Ph.D, Associate Professor, Hod CSE Department

drabdulbari@islec.edu.in

ABSTRACT: Start-up companies use a variety of SaaS products to automate things and get the job done which is required by the consumers. SaaS products types of software that are hosted by a central provider and offered to customers through the internet. Rather than installing or downloading a copy of the application, users can access the product from a web or mobile browser. The SaaS company then manages and updates the software based on user needs. Often, they come across options for softwares which offers the same services yet have many differences. It takes alot of effort to decide which service suits the purpose well and is accustomed to the particular needs of the projects, thus it is a hard job to decide which product is best according to the company's various needs and fits best accordingly. As it is a time consuming task for the companies to filter out the existing services we are going to minimize the time complexity of this task by bringing all the Saas Products to a single platform where the users can compare the services, products etc and use the services that serve their purpose optimally. In our work, we use twitter data to analyze public views towards a product. Firstly, we have developed a natural language processing (NLP) based pre-processed data framework to filter tweets. Secondly, we incorporate Bag of Words (BoW) and Term Frequency-Inverse Document Frequency (TF-IDF) model concept to analyze sentiment. This is an initiative to use BoW and TFIDF are used together to precisely classify positive and negative tweets. We have found that by exploiting TF-IDF vectorizer, the accuracy of sentiment analysis can be substantially improved and simulation results show the efficiency of our proposed system. We achieved 85.25% accuracy in sentiment analysis using NLP technique.

Keywords- *Natural language processing (NLP), Twitter, data mining, Sentiment analysis.*

1. INTRODUCTION

The idea to bring SaaS Products to a single platform of different categories would help start-up companies decide by comparing those products. The project would be built on the concepts of web development and NLp modelling which would be integrated to showcase the products on the platform. It would compare the different products, their services and read the human input given as reviews which would be filtered by the model into positive and negative categories This platform will guide the companies and consumers towards gaining an idea and understanding for choosing optimal SaaS products and other services accordingly and would save their time in the long run.Every social networking sites like facebook, twitter,instagram etc become one of the key sources of information. It is found that by extracting and analyzing data from social networking sites, a business entity can be benefited in their product marketing. Twitter is one of the most popular sites where people used to express their feelings and reviews for a particular product.

Fig.1: Example figure

Now-a-days, internet services are generating a large amount of data which is increasing significantly day by day. Social networking sites are being used for microblogging where it has become a tremendous tool among Internet users for communication . Every big and small company are joining the social networking site to share their product and try to know the reviews of the products from the consumer. The company will use sentiment analysis to grasp the opinion of shoppers concerning their merchandise, so they will analyze client satisfaction and as per that they will improve their product. Particularly, the developed method to sentiment analysis using, by and large, to look at between any device, public figure, Sports team and so on. Twitter is the second biggest social networking platform after Facebook which generates 347,222 tweets every each minute and 21 million tweets per hour [1]. So it creates an opportunity for data mining and sentiment analysis based on users tweet. Since sentiment analyses are part of the data mining that can observe public educe about various topics and products. It is also the stem of natural language processing, text analysis, computational linguistics, bio-metrics, machine learning methods. We are choosing Twitter for sentiment analysis because it offers opportunities for the tenderness of enunciated disposition.

2. LITERATURE REVIEW

2.1 Impact of online consumer reviews on sales: The moderating role of product and consumer characteristics:

This article examines how product and consumer characteristics moderate the influence of online consumer reviews on product sales using data from the video game industry. The findings indicate that online reviews are more influential for less popular games and games whose players have greater Internet experience. The article shows differential impact of consumer reviews across products in the same product category and suggests that firms' online marketing strategies should be contingent on product and consumer characteristics. The authors discuss the implications of these results in light of the increased share of niche products in recent years.

2.2 The impact of online user reviews on hotel room sales:

Despite hospitality and tourism researchers' recent attempts on examining different aspects of online word-of-mouth [WOM], its impact on hotel sales remains largely unknown in the existing literature. To fill this void, we conduct a study to empirically investigate the impact of online consumer-generated reviews on hotel room sales. Utilizing data collected from the largest travel website in China, we develop a fixed effect log-linear regression model to assess the influence of online reviews on the number of hotel room bookings. Our results indicate a significant relationship between online consumer reviews and business performance of hotels.

2.3 Classification of opinion mining techniques:

The important part to gather the information is always seems as what the people think. The growing availability of opinion rich resources like online review sites and blogs arises as people can easily seek out and understand the opinions of others. Users express their views and opinions regarding products and services. These opinions are subjective information which represents user's sentiments, feelings or appraisal related to the same. The concept of opinion is very broad. In this paper we focus on the Classification of opinion mining techniques that conveys user's opinion i.e. positive or negative at various levels. The precise method for predicting opinions enable us, to extract sentiments from the web and foretell online customer's preferences, which could prove valuable for marketing research. Much of the research work had been done on the processing of opinions or sentiments recently because opinions are so important that whenever we need to make a decision we want to know others ' opinions. This opinion is not only important for a user but is also useful for an organization.

2.4 Opinion mining and sentiment analysis,:

An important part of our information-gathering behavior has always been to find out what other people think. With the growing availability and popularity of opinion-rich resources such as online review sites and personal blogs, new opportunities

Journal of Engineering Sciences
Vol 13, Issue 05, MAY/2022
ISSN NO:0377-9254

and challenges arise as people now can, and do, actively use information technologies to seek out and understand the opinions of others. The sudden eruption of activity in the area of opinion mining and sentiment analysis, which deals with the computational treatment of opinion, sentiment, and subjectivity in text, has thus occurred at least in part as a direct response to the surge of interest in new systems that deal directly with opinions as a first-class object. This survey covers techniques and approaches that promise to directly enable opinion-oriented information seeking systems. Our focus is on methods that seek to address the new challenges raised by sentiment aware applications, as compared to those that are already present in more traditional fact-based analysis. We include material on summarization of evaluative text and on broader issues regarding privacy, manipulation, and economic impact that the development of opinion-oriented information-access services gives rise to. To facilitate future work, a discussion of available resources, benchmark datasets, and evaluation campaigns is also provided.

2.5 Opinion mining in e-learning system:

As an education pattern, e-learning systems are becoming more and more popular. For developing of e-learning systems, it is important to know users' opinions and evaluation about them. It is involved in applying the automatic text analysis to extract the opinions and adopting automatic sentiment analysis to identify the sentiment of opinions from the Web pages on which users are discussing or describing their personal opinions and evaluation of the services. Conditional random fields is used for identifying and extracting the opinions. The negative sentences and degree adverbs are specially considered in sentiment processing. Then, we present the strength calculation for opinion sentiment orientation. The experiment shows that it is with high analysis precision on opinions extraction and sentiment analysis and helpful to e-learning system.

3. IMPLEMENTATION

Twitter is limited to 140 characters of text that's why users can explain their brief ideas via a short message. We have developed an NLP based preprocessed data framework to filter tweets where we incorporate Bag of Word (BoW) model and TF-IDF (Term Frequency - Inverse Document Frequency) model concept to sentiment analysis. In NLP Technique, it has done tokenization, stemming, lemmatization, removal of stop words, POS tagging, named entity recognition, co reference resolution, and text modeling as Bag of Word and TF IDF Model . The main aim is to identify the sentiment of the tweet by defining positive and negative polarity where tweets are collected by using Twitter streaming API from Twitter. We utilize these tweets as crude data. At that point, we use the proposed technique that gives the assessment of tweet. From the sentiment analysis, the client will understand the feedback of the services before creating a buying deal.

Our idea is to build a website which not only compares the existing Saas products on the website but also brings down the reviews of all the Saas products on one platform and categorize them into the best ones accordingly. With the use of NlP modelling integration we are gonna categorize the reviews into positive and negative segments which portray a valuable feedback of the products and consumers would gain a wide range of ideas on which product to which and the ones to disregard from their consideration. The categorization of human input given through the NLP modelling has not been implemented on any preexisting platform and this makes our project unique. This product would make choosing Saas products as easy as buying a mobile phone. As one individual just goes to Youtube or platforms like 91mobiles or gsmarena to check the user reviews on the mobile phone, our project would simplify the Saas product world just like that. Rather than going all over the internet looking for reviews for all the products and wasting time on finding the appropriate feedbacks All a consumer would need is to come on our platform search for the product, read the genuine reviews integrated by NLP and decide for themselves.

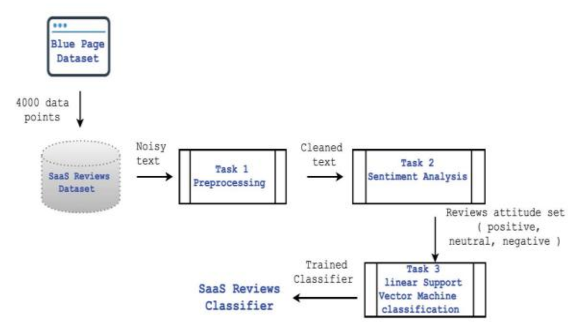

Fig.2: System architecture

4. METHODOLOGIES

Software as a Service (SaaS):

SaaS is also known as "On-Demand Software". It is a software distribution model in which services are hosted by a cloud service provider. These services are available to end-users over the internet so, the end-users do not need to install any software on their devices to access these services.

There are the following services provided by SaaS providers

Business Services - SaaS Provider provides various business services to start-up the business. The SaaS business services include ERP (Enterprise Resource Planning), CRM (Customer Relationship Management), billing, and sales.

Fig.3: services provided by SaaS providers

Document Management - SaaS document management is a software application offered by a third party (SaaS providers) to create, manage, and track electronic documents.

Example: Slack, Samepage, Box, and Zoho Forms.

Social Networks - As we all know, social networking sites are used by the general public, so social networking service providers use SaaS for their convenience and handle the general public's information.

Mail Services - To handle the unpredictable number of users and load on e-mail services, many e-mail providers offering their services using SaaS.

Advantages of SaaS cloud computing layer

1 SaaS is easy to buy:

SaaS pricing is based on a monthly fee or annual fee subscription, so it allows organizations to access business functionality at a low cost, which is less than licensed applications. Unlike traditional software, which is sold as a licensed based with an up-front cost (and often an optional ongoing support fee), SaaS providers are generally pricing the applications using a subscription fee, most commonly a monthly or annually fee.

2. One to Many:

SaaS services are offered as a one-to-many model means a single instance of the application is shared by multiple users.

3. Less hardware required for SaaS:

The software is hosted remotely, so organizations do not need to invest in additional hardware.

4. Low maintenance required for SaaS:

Software as a service removes the need for installation, set-up, and daily maintenance for the organizations. The initial set-up cost for SaaS is typically less than the enterprise software. SaaS vendors are pricing their applications based on some usage parameters, such as a number of users using the application. So SaaS does easy to monitor and automatic updates.

5. No special software or hardware versions required:

All users will have the same version of the software and typically access it through the web browser. SaaS reduces IT support costs by outsourcing hardware and software maintenance and support to the IaaS provider.

6. Multidevice support:

SaaS services can be accessed from any device such as desktops, laptops, tablets, phones, and thin clients.

7. API Integration:

SaaS services easily integrate with other software or services through standard APIs.

8. No client-side installation:

SaaS services are accessed directly from the service provider using the internet connection, so do not need to require any software installation.

NLP:

JES Journal of Engineering Sciences

Vol 13, Issue 05, MAY/2022

ISSN NO:0377-9254

Natural language processing (NLP) refers to the branch of computer science and more specifically, the branch of artificial intelligence or AI concerned with giving computers the ability to understand text and spoken words in much the same way human beings can. NLP combines computational linguistics rule-based modeling of human language with statistical, machine learning, and deep learning models. Together, these technologies enable computers to process human language in the form of text or voice data and to 'understand' its full meaning, complete with the speaker or writer's intent and sentiment.

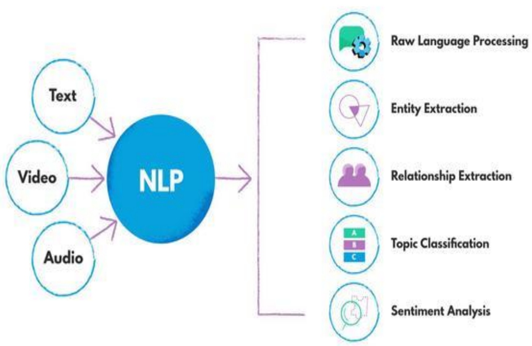

Fig.4: NLP processing

Several NLP tasks break down human text and voice data in ways that help the computer make sense of what it's ingesting. Some of these tasks include the following:

Speech recognition, also called speech-to-text, is the task of reliably converting voice data into text data. Speech recognition is required for any application that follows voice commands or answers spoken questions. What makes speech recognition especially challenging is the way people talk—quickly, slurring words together, with varying emphasis and intonation, in different accents, and often using incorrect grammar.

Part of speech tagging, also called grammatical tagging, is the process of determining the part of speech of a particular word or piece of text based on its use and context. Part of speech identifies 'make' as a verb in 'I can make a paper plane,' and as a noun in 'What make of car do you own?'

Word sense disambiguation is the selection of the meaning of a word with multiple meanings through a process of semantic analysis that determine the word that makes the most sense in the given context. For example, word sense disambiguation helps distinguish the meaning of the verb 'make' in 'make the grade' (achieve) vs. 'make a bet' (place).

Named entity recognition, or NEM, identifies words or phrases as useful entities. NEM identifies 'Kentucky' as a location or 'Fred' as a man's name.

Co-reference resolution is the task of identifying if and when two words refer to the same entity. The most common example is determining the person or object to which a certain pronoun refers (e.g., 'she' = 'Mary'), but it can also involve identifying a metaphor or an idiom in the text (e.g., an instance in which 'bear' isn't an animal but a large hairy person).

Sentiment analysis attempts to extract subjective qualities—attitudes, emotions, sarcasm, confusion, suspicion—from text.

Natural language generation is sometimes described as the opposite of speech recognition or speech-to-text; it's the task of putting structured information into human language.

5. EXPERIMENTAL RESULTS

Reviews	polarity_confidence	polarity
My company has b	0.24248223	positive
SiSense has consis	0.457113117	positive
Handles huge DB c	0.441015244	positive
We spent about 6	0.413340598	positive
SiSense helps us vi	0.94254607	positive
Zoho Reports succ	0.267500013	positive
The integration wi	-0.227008581	negative
Zoho Reports succ	0.212381914	neutral
We compared a lot	-0.102500021	negative
After setting up Da	0.080556497	neutral
Our use case is cen	0.167504683	neutral
You can have a gre	0	neutral
LogicMonitor as a	0.319127381	positive
Both LogicMonitor	-0.245000005	negative
LogicMonitor is a g	0.582140028	positive
If you really want	0.282369673	positive
I wanted this app	0.452391326	positive
Mention is a tool th	0.339855999	positive
we are a niche adv	0.895999968	positive

Fig.5: Shortened version of sentiment analysis of SaaS reviews

The aim of this task is to measure the emotional tone of each SaaS post/review in our dataset. We perform this task using Semantria sentiment analysis that provides a sentiment scoring for each review. This analysis can be applied at the document level or entity level. In our scenario, we apply the analysis at the document level, as each review is a document. The emotion phrases are identified in each document and are then scored to identify positive, negative, or neutral tones. Finally, the scores of the phrases are combined to obtain the overall score for the document.

Polarity of Reviews	Number of Reviewers	Polarity Threshold
Positive	2487	greater than 0.22
Neutral	1312	between 0.21 and 0.04
Negative	201	less that 0.05

Fig.6: Summery of sentiment analysis

	Actual negative	Actual negative	class precision
predict negative	6	30	16.67%
predict positive	195	2457	92.65%
class recall	2.99%	98.79%	

Fig.7: Term occurrences

	Actual negative	Actual negative	class precision
predict negative	7	20	25.93%
predict positive	194	2467	92.71%
class recall	3.48%	99.2%	

Fig.8: TF-IDF

	Actual negative	Actual negative	class precision
predict negative	3	9	25.00%
predict positive	198	2478	92.60%
class recall	1.49%	99.64%	

Fig.9: Binary term occurrences

Approach	Accuracy	classification error
TFIDF	91.63% +/- 0.24%	8.37% +/- 0.24%
Binary Term Occurrences	92.30% +/- 0.18%	7.70% +/- 0.18%
Term Occurrences	92.04% +/- 0.21%	7.96% +/- 0.21%

Fig.10: Compression of 3 Word Vector Approaches With 3 Folds

6. CONCLUSION

In our work, we use twitter data to analyze public views towards a product. Firstly, we have developed a natural language processing (NLP) based pre-processed data framework to filter tweets. Secondly, we incorporate Bag of Words (BoW) and Term Frequency-Inverse Document Frequency (TF-IDF) model concept to analyze sentiment. This is an initiative to use BoW and TFIDF are used together to precisely classify positive and negative tweets. We have found that by exploiting TF-IDF vectorizer, the accuracy of sentiment analysis can be substantially improved and simulation results show the efficiency of our proposed system.

7. FUTURE SCOPE

In future work, we plan to analyze more cloud consumers' reviews for different cloud services, such as Platform as a Service, and different supervised classification models (Navie Bayse, Decision Tree) and compare the results using different weight schema such as the N-gram model.

REFERENCES

[1] F. Zhu and X. Zhang, "Impact of online consumer reviews on sales: The moderating role of product and consumer characteristics," Journal of marketing, vol. 74, no. 2, pp. 133–148, 2010.

[2] Q. Ye, R. Law, and B. Gu, "The impact of online user reviews on hotel room sales," International Journal of Hospitality Management, vol. 28, no. 1, pp. 180–182, 2009.

[3] Q. Roller, "Quote roller reviews, pricing comparison of alternatives," GetApp, vol. 02-, 2015. [Online]. Available: http://www.getapp.com/ sales-software/a/quote-roller/

[4] N. Mishra and C. Jha, "Classification of opinion mining techniques," International Journal of Computer Applications, vol. 56, no. 13, pp. 1–6, 2012.

[5] S. Model, "Software as a service (saas) from the gartner IT glossary," Gartner IT Glossary, vol. 04-, 2012. [Online]. Available: http://www.gartner.com/it-glossary/software-as-a-service-saas/

[6] B. Pang and L. Lee, "Opinion mining and sentiment analysis," Foundations and trends in information retrieval, vol. 2, no. 1-2, pp. 1–135, 2008.

[7] D. Song, H. Lin, and Z. Yang, "Opinion mining in e-learning system," in Network and Parallel Computing Workshops, 2007. NPC Workshops. IFIP

International Conference on. IEEE, 2007, pp. 788–792.

[8] E. Marrese-Taylor, J. D. Velasquez, F. Bravo-Marquez, and Y. Matsuo, ´ "Identifying customer preferences about tourism products using an aspect-based opinion mining approach," Procedia Computer Science, vol. 22, pp. 182–191, 2013.

[9] S. F. Sayeedunnissa, A. R. Hussain, and M. A. Hameed, "Supervised opinion mining of social network data using a bag-of-words approach on the cloud," in Proceedings of Seventh International Conference on Bio-Inspired Computing: Theories and Applications (BIC-TA 2012). Springer, 2013, pp. 299–309.

[10] N. Bharathi, G. Brindha, B. Santhi, and P. Neelamegam, "Integration of fpga, cloud and opinion mining," Journal of Applied Sciences, vol. 14, no. 14, p. 1653, 2014.

[11] M. Cooke and N. Buckley, "Web 2.0, social networks and the future of market research," International Journal of Market Research, vol. 50, no. 2, p. 267, 2008.

[12] Y. Chen and J. Xie, "Online consumer review: Word-of-mouth as a new element of marketing communication mix," Management Science, vol. 54, no. 3, pp. 477–491, 2008. [Online]. Available: http://pubsonline.informs.org/doi/abs/10.1287/mnsc.1070.0810

[13] B. Liu, "Sentiment analysis and opinion mining," Synthesis Lectures on Human Language Technologies, vol. 5, no. 1, pp. 1–167, 2012.

[14] L. Zhuang, F. Jing, and X.-Y. Zhu, "Movie review mining and summarization," in Proceedings of the 15th ACM international conference on Information and knowledge management. ACM, 2006, pp. 43–50.

[15] A. Minanovic, H. Gabelica, and Z. Krstic, "Big data and sentiment analysis using knime: Online reviews vs. social media," in Information and Communication Technology, Electronics and Microelectronics (MIPRO), 2014 37th International Convention on. IEEE, 2014, pp. 1464–1468.