# Users' Sentiment Analysis in Social Media Context using Natural Language Processing

Nitish Bhardwaj[1], Anupam Shukla[2], Pradip Swarnakar[3]
[1, 2, 3] (Department of Information Technology, ABV-Indian Institute of Information Technology and Management, Gwalior,MP, India)
Prof. Anupam Shukla, E-Block, ABV-IIITM, Gwalior, Madhya Pradesh, India, PIN-474010
nitish.bhardwaj11@gmail.com

***Abstract: The main goal of this research work is to model a natural language processing toolkit to infer users' sentiment analysis for study of feedback about a product, person or thing using the means of social media interactions. A complete discussion on the mechanism behind the data gathering using social networking channels has been made in this research as well as a small account on the design of natural language processing toolkit has been also given. To demonstrate the validation and applicability of the end application, the proposed methodology will be implemented with input sets obtained from live users' interactions on social media and feeding them to a natural language model for observation. In this research, we will also check the validity of the result obtained comparing the results obtained by polls survey and by natural language model.***

**Keywords: Social Media Channels, Natural Language Processing, Natural Language Toolkit, Sentiment Polarity Score, Contextual Dictionary**

## I. INTRODUCTION

Consumers are day by day relying more on feedbacks posted on the social network channels to make a variety of decisions ranging from what movies to watch to what business to invest in [17]. Various researches in the past has suggested that consumers follow these reviews and consider them more unbiased and transparent than the traditional sources [3]. However, it is very unclear to use various mechanisms for manipulation of online reviews and influencing consumers' purchase decisions.

Since social network web sites have become popular media for people to share their opinions, enterprises have sought the opportunities to leverage this data for business intelligence applications such as enterprise marketing services and customer relationship management. It has become critical for enterprises to unlock customer sentiment embedded in the huge amount of social media data so that they can quickly respond to complaints and improve their product quality. Sentiment analysis is the study of using a machine to determine the polarity of an opinion - whether it is positive, negative, or neutral. However, while sentiment analysis (Pang, Lee, and Vaithyanathan 2002; Hu and Liu 2004) [4] [5] is no short-term hot research topic, few work has focused on social media data. Rather, research has been on more structured language such as product and movie reviews, due to the low accuracy resulting from much more informal writing, short sentences, sarcasm, and abbreviations seen in social media data (Wilson,Wiebe, and Hoffmann 2005a) [2].

In this paper, we will use the adjectives and polarity from lexicon-based sentiment analysis on human-labeled social comments using the generated sentiment dictionary (Wei Peng and Dae Hoon Park 2011) [1] [19] [20]. Then we will use NLTK- Natural Language Toolkit of Python for Machine Learning Approach to remove ambiguous data, sarcastic comments and to correlate canonical forms. This model will enable to infer the users' real sentiments for a product, policy or person. The information extracted from the model will be useful for various applications in multiple domains by enabling significance of social media channels for the welfare of living society.

## II. LITERATURE REVIEW

Adjectives are always important to impart inference from social media networks. For this purpose, the paper entitled as "**Generate Adjective Sentiment Dictionary for Social Media Sentiment Analysis Using Constrained Nonnegative Matrix Factorization**" [1] is a source of adjectives and polarity. The paper proposed to automatically generate an adjective sentiment dictionary from social media data with the following steps: (1) obtain a set of seed positive and negative adjective words and expand it using synonym and antonym relations from the WordNet (Fellbaum 1998) [6]; (2) extract all the adjectives linked to the adjective set by 'and' and 'but' using Part-Of-Speech (POS) technique on social media corpus; (3) construct a graph matrix (or a nonnegative symmetric matrix) where each entry is the edge weight between two adjectives calculated from the synonym relations from WordNet and the 'and' conjunction relations; (4) construct a constraint matrix (a nonnegative symmetric matrix) where each non-zero entry value denotes a Cannot-link weight between two adjectives calculated from the antonym relation from WordNet and the 'but' conjunction relation; (5) use our proposed Constrained Symmetric Nonnegative Matrix Factorization (CSNMF) algorithm to iteratively cut this adjective graph into positive and negative sets, where each adjective is assigned a positive score and a negative score.

**Table 1: Example words with sentiment scores [1]**

| Adjective | Positive Score | Negative Score |
|---|---|---|
| | | |
| 'cocking' | 0.902 | 0 |
| 'new' | 0.1511 | 0.0116 |
| 'sassy' | 0.8836 | 0 |
| 'yucky' | 0 | 0.9095 |
| 'irksome' | 0 | 0.8994 |
| 'long-winded' | 0 | 0.8895 |
| 'dark' | 0.0228 | 0.0297 |

This paper shows that combining links from both WordNet and the corpus to generate sentiment dictionaries does outperform using only one of them. Comparison between our method and some existing approaches show that our performance improvement is statistically significant. Our proposed method can also assign the sentiment strength score to each word in the dictionary, in which the top ranked words yield better precision. Finally, our dictionary shows comparable performance in determining the sentiment score for social network comments, compared to the human labeled ground-truth dictionaries.

The polarity defined in "Fuzzy Logic Models for the Meaning of Emotion Words" [7] proposed that words and natural language play a central role in how we describe and understand emotions. One can learn about emotions first-hand by observing physiological or behavioural data, but to communicate emotional information to others who are not first-hand observers, one must use natural language descriptions to communicate the emotional information. The field of affective computing deals with creating computer systems that can recognize and understand human emotions. To realize the goals of affective computing, it is necessary not only to recognize and model emotional behaviour, but also to understand the language that is used to describe such emotional behaviour.

**TABLE 1 Similarity between words of the Blog Moods vocabulary and the Emotion Category Word vocabulary.**

| | ANGRY | DISGUSTED | FEARFUL | HAPPY | NEUTRAL | SAD | SURPRISED |
|---|---|---|---|---|---|---|---|
| AMUSED | 0.004 | 0.003 | 0.005 | **0.060** | 0.004 | 0.005 | 0.053 |
| TIRED | 0.006 | 0.003 | 0.034 | 0.001 | 0.038 | **0.196** | 0.001 |
| CHEERFUL | 0.003 | 0.003 | 0.003 | **0.109** | 0.001 | 0.002 | 0.088 |
| BORED | 0.015 | 0.012 | 0.075 | 0.004 | 0.064 | **0.335** | 0.004 |
| ACCOMPLISHED | 0.015 | 0.013 | 0.008 | **0.151** | 0.006 | 0.008 | 0.139 |
| SLEEPY | 0.007 | 0.005 | 0.018 | 0.009 | **0.172** | 0.128 | 0.010 |
| CONTENT | 0.005 | 0.004 | 0.007 | **0.044** | 0.015 | 0.012 | 0.040 |
| EXCITED | 0.015 | 0.017 | 0.006 | **0.255** | 0.002 | 0.002 | 0.213 |
| CONTEMPLATIVE | 0.006 | 0.004 | 0.012 | 0.006 | **0.161** | 0.075 | 0.007 |
| BLAH | 0.014 | 0.010 | 0.049 | 0.005 | 0.166 | **0.359** | 0.007 |
| AWAKE | 0.020 | 0.017 | 0.016 | 0.061 | 0.015 | 0.014 | **0.068** |
| CALM | 0.003 | 0.002 | 0.011 | 0.007 | **0.137** | 0.069 | 0.008 |
| BOUNCY | 0.009 | 0.012 | 0.002 | **0.361** | 0.000 | 0.001 | 0.311 |
| CHIPPER | 0.002 | 0.002 | 0.001 | **0.066** | 0.002 | 0.003 | 0.059 |
| ANNOYED | **0.393** | 0.380 | 0.080 | 0.041 | 0.002 | 0.023 | 0.076 |
| CONFUSED | 0.026 | 0.020 | 0.064 | 0.014 | 0.046 | **0.170** | 0.017 |
| BUSY | 0.068 | 0.079 | 0.049 | 0.111 | 0.013 | 0.012 | **0.116** |
| SICK | 0.008 | 0.004 | 0.032 | 0.001 | 0.023 | **0.204** | 0.001 |
| ANXIOUS | **0.207** | 0.181 | 0.091 | 0.028 | 0.003 | 0.025 | 0.038 |
| EXHAUSTED | 0.015 | 0.011 | 0.048 | 0.003 | 0.046 | **0.298** | 0.004 |
| DEPRESSED | 0.008 | 0.005 | 0.050 | 0.001 | 0.015 | **0.218** | 0.001 |
| CURIOUS | 0.038 | 0.042 | 0.014 | **0.203** | 0.011 | 0.006 | 0.176 |
| DRAINED | 0.009 | 0.007 | 0.039 | 0.002 | 0.061 | **0.280** | 0.003 |
| AGGRAVATED | 0.578 | **0.618** | 0.114 | 0.047 | 0.002 | 0.020 | 0.087 |
| ECSTATIC | 0.000 | 0.000 | 0.000 | 0.108 | 0.000 | 0.000 | **0.117** |
| BLANK | 0.006 | 0.004 | 0.017 | 0.005 | 0.133 | **0.137** | 0.006 |
| OKAY | 0.016 | 0.013 | 0.035 | 0.017 | **0.076** | 0.057 | 0.020 |
| HUNGRY | **0.084** | 0.062 | 0.029 | 0.045 | 0.013 | 0.034 | 0.052 |
| HOPEFUL | 0.009 | 0.007 | 0.007 | 0.047 | 0.010 | 0.009 | **0.050** |
| COLD | 0.005 | 0.003 | 0.026 | 0.001 | 0.047 | **0.123** | 0.002 |
| CREATIVE | 0.027 | 0.037 | 0.007 | **0.524** | 0.001 | 0.002 | 0.462 |
| PISSED_OFF | **0.383** | 0.363 | 0.052 | 0.016 | 0.000 | 0.008 | 0.035 |
| GOOD | 0.004 | 0.003 | 0.004 | **0.067** | 0.005 | 0.006 | 0.060 |
| THOUGHTFUL | 0.005 | 0.003 | 0.004 | 0.011 | **0.079** | 0.029 | 0.012 |
| FRUSTRATED | 0.186 | **0.233** | 0.068 | 0.022 | 0.001 | 0.012 | 0.030 |
| CRANKY | 0.325 | **0.351** | 0.099 | 0.045 | 0.002 | 0.022 | 0.060 |
| STRESSED | 0.288 | **0.304** | 0.158 | 0.044 | 0.003 | 0.026 | 0.053 |

**Fig.1 Similarity between words of the Blog Moods vocabulary and the Emotion Category Word vocabulary [7]**

The psycholinguistic theory of communication accommodation [8] accounts for the general observation that users who post comments on social media networks tend to converge to each other's opinion. They resemble in a lot ways such as the selection of words, the trending words, gestures and expressions. This theory has been supported by many small scale and big data analysts.

Social interactions of users do not depend only on the static parameters but on dynamic interactions of users on social media networks. The length constraints on comments put a restriction on the users. So, this is a limiting factor which weakens robustness. For investigation of users' interaction on social media networks, the fuzzy model proposes a probabilistic framework for detection of users' sentiments of social media networks.

Polarity classification of words or adjectives [9] is important for applications such as Opinion Mining and Sentiment Analysis. A number of sentiment word/sense dictionaries have been manually or (semi) automatically constructed. The dictionaries have substantial inaccuracies. Besides obvious instances, where the same word appears with different polarities in different dictionaries, the dictionaries exhibit complex cases, which cannot be detected by mere manual inspection. We introduce the concept of polarity consistency of words/senses in sentiment dictionaries in this paper.

There are numerous works that, given a sentiment lexicon, analyze the structure of a sentence/document to infer its orientation, the holder of an opinion, the sentiment of the opinion, etc. (Breck et al., 2007; Ding and Liu, 2010) [9] [10]. Several domain independent sentiment dictionaries have been manually or(semi)-automatically created, e.g., General Inquirer(GI) (Stone et al., 1996), Opinion Finder (OF) (Wilson et al., 2005) [13], Appraisal Lexicon, SentiWordNet (Baccianella et al., 2010) [11] and Q-WordNet (Agerri and Garc´ıa-Serrano,2010) [12]. Q-WordNet and SentiWordNet are lexical resources which classify the synsets(senses) in Word-Net according to their polarities. We call them sentiment sense dictionaries (SSD). OF, GI and AL are called sentiment word dictionaries (SWD). They consist of words manually annotated with their corresponding polarities. The sentiment dictionaries have the following problems:

• They exhibit substantial (intra-dictionary) inaccuracies. For example, the synset{Indo-European, Indo-Aryan, Aryan} (of or relating to the former Indo-European people),has a negative polarity in Q-WordNet, while most people would agree that this synset has a neutral polarity instead.
• They have (inter-dictionary) inconsistencies. For example, the adjective cheap is positive in AL and negative in OF.
• These dictionaries do not address the concept of polarity (in) consistency of words/synsets.

We concentrate on the concept of (in) consistency in this paper. We define consistency among the polarities of words/synsets in a dictionary and give methods to check it.

The great acceptation of the Social Web [14] has converted social networks, blogs and wikis in almost perfect advertising mediums. However, many of the current social publicity strategies do not exploit all the potential of these mediums, since they obviate users' online life: the social contexts in which they are involved. Our proposal to reverse this situation is a model to infer users' social contexts by the application of several Natural Language Processing (NLP) and data mining techniques over users' interaction data on Facebook. We take advantage of both Facebook and *Groupon APIs* to provide a deployment scenario in which knowing users' social life allows ads to target the most potential customers, which is beneficial for both companies and possible customers.

Facebook has its own social publicity tool (Facebook Ads) to disseminate ads among its users. However, the fact that (i) the recommendation algorithm, (ii) the users' recommendation profiles and (iii) the set of all available ads are not publicly available prevents it to be a reference with which comparing any other social publicity strategy that considers Facebook users' data. Still, some recent works in the literature have inspected the use of Facebook API for item's recommendation [16] [18].

The manipulation of online reviews [15] plays an important role in analysis of ratings, readability, and sentiments. As consumers are relying on all the social media networks for decision making. The feedback method has been attached to the word of mouth (WOM) that the users put on social media networks. There is also a controversial issue that the companies tend to make social media networks to misrepresent their product but how much do they succeed is a debatable issue. This paper illustrates the various methods of online reviews, the pros and cons of various data gathering methods and the consumer's response to these reviews. This analysis also examines textual information in online reviews by combining sentiment mining algorithms with pointer/grading assessment.

It has been discovered that around 10.3% of the products are subject to manipulations by online reviews. Even sometimes the sentiments are deliberately used, the ratings used by natural language processing methods are useful for decision making. The findings from this research encourage to develop deep into the processing technique and filtering of noisy data.

## III. PROPOSED METHODOLOGY

**Brief explanation of the terminologies used**

*Social Media Channels* are referred as social networking websites such as Facebook, Twitter, and LinkedIn where users interact without the influence of pre-assumptions and with full-freedom.
*Natural Language Processing* is a mechanism to extract useful information from the conversational data from the social media channels. *Natural Language Toolkit* in Python is a toolkit equipped with advanced methods and in-built functions for better decision making using natural language processing. *Machine Learning* enables a smart system to remove ambiguous data and detection of sarcastic comments for better decision making. It is assisted with a storage system, automation of complex computation techniques and human analytical abilities based on the latest research in the field of artificial intelligence.

**Proposed methodology:**

The proposed methodology is divided into dependent subsystems each performing as isolated units which are explained as follows:

1. Data Collection:
    - Define Dictionary of positive and negative adjectives
    - Define polarity of positive and negative adjectives
    - Define canonical tagging
2. Data Processing
    - Extraction of useful data from social networking channels
    - Removal of ambiguous data; e.g., sarcasm, interrogative comments
    - Integration of multiple social media channels for accuracy
    - Extraction of sentiments from comments
3. Decision Making
    - Design of natural language processing model to draw inference from comments on channels
    - Use of polarity database and comments for decision making for sentiment analysis

**Data collection methodology:**

In this paper, we have used database extracted from users' interactions on social media networks. The data will be taken from multiple social channels for better decision making.
For design of natural language processing model, there are multiple dictionary sources as mentioned in the literature review. For our purpose, we will use dictionary defined by the WordNet (Fellbaum 1998) [6] and Fuzzy Logic Models for the Meaning of Emotion Words [7]. The dictionary will be updated using the tools of machine learning to enable dynamic updating.

**Data processing:**

The data extracted from users will be processed to filter the useful information. For this purpose, we use various techniques such as word-sense disambiguation, sarcasm detection, context level analysis of sentences, negative and interrogative sentences, trending words.

The other data, i.e., dictionary needs to be updated with the trending words of social media. The data used for dictionary are useful to detect adjectives for Agreement analysis and determine the positive and negative polarity.

The major parameters of Sentiment Analysis in context of Social Media Channel like Facebook are:

- Users
- Comments
- Likes

The users put multiple comments and then different comments have different number of likes. Using the Facebook API, the data can be extracted in JSON format which can be further represented in the desired format. The number of likes work as the filtering parameter to choose the most valid comments for analysis. These parameters are useful for removal of useless and noisy data of social networks.

**Processing of Comments:**

The pre-processing of data follows a hybrid approach; i.e.; combination of top-down and bottom-up analysis. This model extracts scores of multiple sentences from multiple comments of users and the score is relatively added based on the parameters of number of likes and polarity score. The scores are added and final polarity score of a user is used for calculation of agreement analysis.

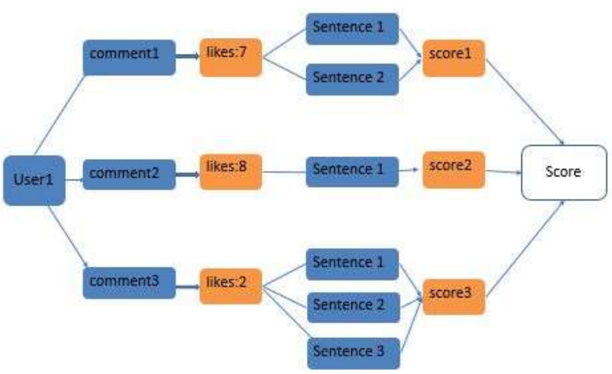


**Fig.2 Data Pre-processing Diagram**

The data extracted from social network are tested with the database of dictionary of adjectives. The sentiment score is calculated using following steps:

- Sentence Level Extraction from comments
- Discarding ambiguous sentences
- Detection of Adjectives used in sentences
- Counting the total polarity score.

$$total_{score} = \sum_{i} (adjective_{polarity} \times adective_{count_i})$$

- Comparison of positive and negative score

The positive score and the negative score are calculated using the same method but using different data files for positive adjectives and negative adjectives.

**Analysis of Score:**

The data analyzed from users’ interaction on social media will be processed to filter the useful information. The positive and negative scores reflect the opinion of users’ interaction. The polarity deviation from each comments can be used to draw inference regarding trending words and sarcastic comments.

For the purpose of more accurate results, we need to integrate the machine learning techniques to the Facebook API for analysis of dynamic data.

**Proposed model of Machine Learning Technique:**

In this paper, Machine Learning Technique integrated with NLTK (Natural Language Toolkit) will be used to model an approach to infer users’ sentiments analysis about a product, policy, person or thing.

This technique will be useful in decision making by various organizations to improve their quality of services and better customer services.

The results of this model will be useful in various ways:

- Providing customer satisfaction level about any product
- Agreement Analysis on any social issue
- Opinion mining about any person or thing
- Users’ Sentiments for any person

- Unbiased Exit Polls

The machine learning approach for sentiment analysis [4] is based on training the data set to classify the adjectives in the sets of different degrees of positive and negative adjectives. This approach enables classification of statements based on some predefined sets of sentences. The sentiment analysis used for data set with single attribute can be handled with this approach, but if we want to get the analysis of a product having multiple attributes, we need to consider contextual analysis, dynamic dictionary and the words introduced in some context. Therefore, we are now going to discuss the significance of contextual analysis and the approach to implement contextual analysis.

**Significance of Contextual Analysis:**

The sentiment analysis is incomplete without considering the context of the subject and the attributes. The context level analysis is also significant in extracting information from a subject based on multi-attributes. For example: the review of a camera depends on multiple attributes like picture quality, weight, portability, motion detection, red-eye removal, degree of focus, finer details etc. If we consider the attribute e.g. 'weight' for which the adjective 'lighter' is positive for subject like 'camera' but the same adjective for the same attribute is negative for subject like 'paper weight'. So, the adjective has various significance based on the context, i.e., the subject. Therefore, the contextual analysis is very significant in this case.

**Approach of Contextual Analysis:**

To model the context level analysis, we will use a topic modelling approach discussed in "Hidden Factors and Hidden Topics: Understanding Rating Dimensions with Review Text" [24]. This approach is used to model attributes and adjectives relation based on the context, i.e., the subject. The data set for analysis is taken from Stanford Network Analysis Project [24].

**Time Variant Analysis:**

The model also uses Machine learning to draw inference from patterns of users' sentiments over a given period of time. Facebook API plays an important in extraction of time dependent data for analysis. Users' sentiment disparity and fluctuating behavior play an important role in drawing conclusions.

There are some challenges yet to be explored with some new approaches. We try to deal with these in some naïve way. The different comments by the same user also poses a serious issue of concern for processing. This can be tackled with the parameter "Num_of_likes". The other challenge is "Sarcasm detection" which can be dealt by ruling out the variations or higher degree of disparity. The results analyzed using this natural language processing toolkit has been discussed below. The data taken has been put with reference.

## IV. RESULTS AND DISCUSSION

The result and analysis of data processed using the methodology explained above has given following observations.

**Analysis: 1**

*Social networking is a useful tool for present day business communication* [9]

Total no. of comments: 196

Positive Score : 90.997730

Negative Score : 9.002270

Conclusion:

Most of the students support significance of social media networks in business communication. Their statements are inclined to extreme positive sides to show the high degree of support.

**Analysis: 2**

*How do leadership can change organizational performance?* [9]

Total no. of comments: 173

Positive Score: 87.329186

Negative Score: 12.670814

Conclusion:

Majority of the students support importance of a good leadership in an organization. Students believe that leadership can play a significant role in changing organizational performance.

**Analysis: 3**

*Dynamic Data Extraction from Social Networking sites using APIs[10]*

*– CNN Poll Agreement Analysis*

*– Do you agree?*

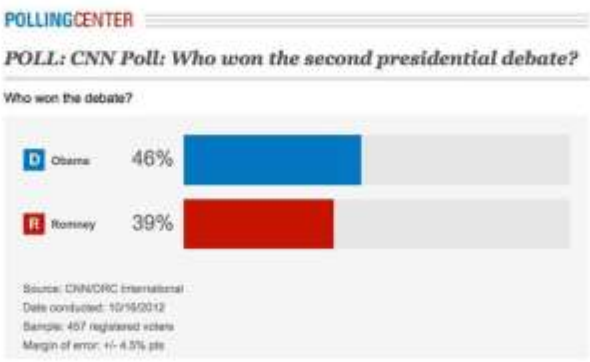


**Fig.3 CNN-Poll [23]**

Total no. of comments: 8350

Positive Score: 63.423695

Negative Score: 36.576305

Conclusion: Majority of the people support Obama for a good leadership in the country. They support Obama and it is evident from the polls result as well.

Similar analysis can be done on any other comments from social media networks to draw inference from users' interaction. The analysis can be useful in opinion mining and decision making.

## V. CONCLUSION

The wide application of soft computing techniques and its integration with latest API technology is becoming vociferously popular as the research and development among all the leading software and hardware companies. Particularly in decision making system where decision are based on complex system of analysis and processing of a large set of information, we basically need to deal with natural language analysis with following features:

- Context level analysis of Sentences
- Word-sense disambiguation, i.e., *analysis of Negative and interrogative sentences* [4]
- Trending words
- Sarcasm Detection [3]
- Detailed Analysis with time variant [10][14]

In the study of sentiment analysis, we discussed the significance and methodology of contextual analysis of statements discussed in social media network. The topic model approach discussed in "Hidden Factors and Hidden Topics: Understanding Rating Dimensions with Review Text" [24] deals with only pre-defined attributes or topics. We not only consider the topic modelling for predefined attributes but also consider inclusion of dynamic attributes. For example: in the analysis: 3, we must consider some attributes as the words like 'Obama' as positive and 'Romney' as negative sentiments. Similar contextual analysis attributes play an important in reconstructing the decision making model.

The recommendation system discussed in *"Facebook single and cross domain data for recommendation systems*" [16] utilizes comments information from social media channel Facebook for extraction of information. We not only use comments from Facebook but also use 'Number of likes' parameter which is very useful in pre-processing and removal of noisy comments. We also extract the user who has commented in order to deal with multiple comments of the same user. The multiple comments by the same user can cause data disparity and skewedness. This information also

ISBN:978-0-9891305-6-1 ©2014 SDIWC

helps to filter out the information of a user with ambiguous data.

The decision support systems discussed in "*Manipulation of online reviews: An analysis of ratings, readability, and sentiments*" [15] brings new dimensions in the opinion mining from the social media interactions. We take inspiration from their model to use social media networks for extraction of useful information. This model extract information form twitter for drawing inference for a decision making. We use the social channel Facebook for our study. Although Facebook has more noisy data but there are also means of filtering those noisy data like number of likes. Facebook unlike Twitter also covers a wide variety of population. Twitter is restricted to some classes of people. Facebook has its reach to the masses. The information taken from social media channels should not be skewed.

The machine learning technique discussed in "*Thumbs up? Sentiment classification using machine learning techniques*" [4] has been used by us with integration of Python's Natural Language Toolkit for better extraction of sentiments from the given comments of social media channels.

We also bring a new concept of time-variant analysis which plays an important role is decision support system. The future work is focused making the decision making more accurate and dynamic. The work will focus on basically two things:

- Detailed Analysis with time variant [10] [14]
  - Use of Facebook Graph API
  - Time variant dynamic data collection

The use of dynamic time-variant data will focus on bringing more transparency and validation of the results obtained by natural language processing.

## REFERENCES


[1] W. Peng and D. H. Park, "*Generate Adjective Sentiment Dictionary for Social Media Sentiment Analysis Using Constrained Nonnegative Matrix Factorization",* Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media, 2011

[2] T. Wilson, J. Wiebe, and P. Hoffmann, "*Recognizing contextual polarity in phrase-level sentiment analysis*" In HLT '05, 347–354.

[3] B. Bickart and R.M. Schindler, *"Internet forums as influential sources of consumer information"*, Journal of Interactive Marketing, pp. 31–40, March 2001.

[4] B. Pang, L. Lee, and S. Vaithyanathan, *"Thumbs up? Sentiment classification using machine learning techniques"* In EMNLP, pp.79–86, 2002.

[5] M. Hu and B. Liu, *"Mining and summarizing customer reviews"* In KDD '04, pp.168–177, 2004.

[6] C. Fellbaum, ed. 1998. *"WordNet: An Electronic Lexical Database."* Cambridge, MA: MIT Press, 1998.

[7] K. Abe, L. Sungbok, and N. Shrikanth, University of Southern California, USA, *"Fuzzy Logic Models for the Meaning of Emotion Words"*, IEEE Computational intelligence magazine, pp. 34-49, May 2013

[8] D. Cristian, G. Michael, Susan D., "*Mark my words! Linguistic style accommodation in social media*", International World Wide Web Conference Committee (IW3C2). WWW 2011, ACM 978-1-4503-0632-4/11/03, Hyderabad, India, March, 2011.

[9] B. Eric, C. Yejin, and C. Claire, *"Identifying expressions of opinion in context*" In IJCAI, 2007

[10] L. Bing, *"Sentiment analysis and subjectivity"* In Nitin Indurkhya and Fred J. Damerau, editors, Handbook of Natural Language Processing, Second Edition. CRC Press, Taylor and Francis Group, Boca Raton, FL. ISBN 978-1420085921

[11] B. Stefano, E. Andrea, and S. Fabrizio, *"SentiWordNet 3.0: An Enhanced Lexical Resource for Sentiment Analysis and Opinion Mining"*. In Proceedings of the

Seventh conference on International Language Resources and Evaluation (LREC'10), Valletta, Malta, May, 2010

[12] R. Agerri and A. Garc´ıa-Serrano, *"Q-wordnet: Extracting polarity from wordnet senses"* In LREC, 2010

[13] T. Wilson, J. Wiebe and P. Hoffmann, *"Recognizing contextual polarity in phrase-level sentiment analysis"* In HLT/EMNLP, 2005.

[14] S. Servia-Rodr´ıguez, A. Fern´andez-Vilas, P. Rebeca, R. D´ıaz-Redondo and J. Jos´e; "*Inferring Contexts from Facebook Interactions: a Social Publicity Scenario"*, IEEE Transactions on Multimedia, vol:5, pp.13, April 2013.

[15] N. Hu, I. Bose, N. Koh and L. Liu, "*Manipulation of online reviews: An analysis of ratings, readability, and sentiments*", Decision Support Systems, vol.52, pp.674–684, 3 November 2011.

[16] B. Shapira, L. Rokach and S. Freilikhman, *"Facebook single and cross domain data for recommendation systems,*" User Modeling and User-Adapted Interaction, pp.1–37, 2012.

[17] L. Guernsey, *"Suddenly, everybody's an expert on everything"* The New York Times, February 3 2000.

[18] J. Duan, S. Prasad, and J. Huang, *"Discovering unknown but interesting items on personal social network,*" Advances in Knowledge Discovery and Data Mining, pp.145–156, 2012.

[19] http://users.cis.fiu.edu/∼wpeng002/sentiDict/

[20] http://www.digg.com

[21] http://obcomments.netii.net/

[22]https://developers.facebook.com/docs/reference/apis/

[23]https://www.facebook.com/photo.php?fbid=10151231723796509

[24] J. McAuley, and J. Leskovec, "*Hidden factors and hidden topics: understanding rating dimensions with review text*", RecSys'13, October, 2013