ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 4, Issue 12, December 2017

Addressing the Issues of Text Analytics

[1] S.N.Sithi Shamila, [2] Dr.D.S.Mahendran, [3] Dr.Mohamed Sathik
[1] Asst. Prof,Dept of Computer Science, Wavoo Wajeeha Women's college, Kayalpatnam
[2] Assoc. Prof. Dept of Computer Science, Adithanar College of Arts and Science, Thiruchendur
[3] Principal, Sadakathullah Appa College, Palyamkottai

Abstract: - **The revolution that has swept through information Processing arena has placed the modern world in a unique vantage position. Huge quantum of data has been generated through the normal operations of day-to-day business activities in various fields such as public services, Research, Education, Industries and Institutes. The multiple level transactions captured in these fields have generated such an enormous amount of data which are so challenging to handle as it swings from megabytes to petabytes, and in a swift, to zettabytes as well. The inevitability of the dependence on the Internet in this technologically revolutionized context has paved way for the increased use of data, and this situation has obviously generated considerable research interest to study user view analysis in order to determine the several characteristics of the users which might influence the development of software systems and several other technological products. It has been observed that more than 90 percent of today's data is either unstructured or semi-structured, and this has complicated and challenged the study of decoding knowledge and information embedded in different recognizable patterns and analyze the text documents from the huge volume of accumulated data. It is in this context, the study of text mining, which is essentially a process of extracting interesting and non-trivial patterns from a large amount of text documents, assumes significance. Although there are innumerable techniques and tools to mine the text in order to discover valuable information which with not only future predictions in several areas of business, social, political and economic interest can be made, also it will influence the process of decision making in several spheres. Thus, Sentiment Analysis or Opinion Mining has attained an enormous importance in the modern world today, and it is in this context, an attempt has been made in this paper to briefly discuss and analyze text mining techniques and the issues connected to text mining which affect accuracy and relevance of the results obtained ..**

***Keywords*: Knowledge Discovery; Text Mining; Sentiment Analysis.**

I. INTRODUCTION

In this world of science and technology, human lives have been inevitably influenced by the forces of changes effected through the generation of huge amount of data which have been the sources of various interpretations and decisions in the matters connected to research and technology. The quantum of the 'Big Data' work that has been done in today's world has the potential to reverse the classically defined relationship between data and technological advancement. The world of business, the field of science, the operation of technology and functioning of government apparatus do indispensably depend In the modern world today, most social networks engage in the act of bringing people and connecting them on various similar interests and issues of social, political , religious, ideological and cultural significance. In times to come, it is expected that such networks might connect with other entitles such as software components, web-based services, data resources, workflows . These interactions which are expected in near future between people and nonhuman artifacts would explode us much more interesting data , and would engage data scientists' productivity too. In this context it is worth-mentioning that Big data analytics can accumulate the words of wisdom of crowds, reveal patterns of behavior and yield best practices. For better understanding of the role of analytics, they have been listed here as follows. Prescriptive analytics is really valuable, but largely not used. Where big data analytics in general sheds light on a subject, prescriptive analytics gives you a laser-like focus to answer specific questions. For example, in the health care industry, you can better manage the patient population by using prescriptive analytics to measure the number of patients who are clinically obese, then add filters for factors like diabetes and LDL cholesterol levels to determine where to focus treatment. Predictive analytics use big data to identify past patterns to predict the future. For example, some companies are using predictive analytics for sales lead scoring, entire sales process, analyzing lead source, CRM data, etc. Properly tuned predictive analytics can be used to support sales, marketing, or for other types of complex forecasts.

Diagnostic analytics are used for discovery or to determine why something happened. For example, for a social media marketing campaign, you can use descriptive analytics to assess the number of posts, mentions, followers, fans, page views, reviews, pins, etc. There can be thousands of online mentions that can be distilled into a single view to see what worked in your past campaigns and what didn't. Descriptive analytics or data mining are at the bottom of the big data value chain, but they can be valuable for uncovering patterns that offer insight. A simple example of descriptive analytics would be assessing credit risk; using past financial performance to predict a customer's likely financial performance. Descriptive analytics can be useful in the sales cycle, for example, to categorize customers by their likely product preferences and sales cycle. In this paper, we will see the analytics types, techniques used for Sentiment Classification based on Text and Challenges of Text mining.

II.ANALYTICS

A.Text Analytics.

Most of all information or data is available in textual form in databases. From these contexts, manual Analytics or effective extraction of important information are not possible. Text analytics or text mining refers process of deriving important information from text data. It will use to extract meaningful data from the text. It use many ways like associations among entities, predictive rules, patterns, concepts, events etc. based on rules. Text analytics widely use in government, research, and business needs. Data simply tells you what people did but text analytics tell you why. Social network feeds, emails, blogs, online forums, survey responses, corporate documents, news, and call center logs are examples of textual data held by organizations. Text analytics involve statistical analysis, computational linguistics, and machine learning. Text analytics enable businesses to convert large volumes of human generated text into meaningful summaries, which support evidence-based decision-making. For instance, text analytics can be used to predict stock market based on information extracted from financial news.We present a brief review of text analytics methods below.

Information extraction (IE) techniques extract structured data from unstructured text. For example, IE algorithms can extract structured information such as drug name, dosage, and frequency from medical prescriptions. Two sub-tasks in IE are Entity Recognition (ER) and Relation Extraction (RE) (Jiang, 2012). ER finds names in text and classifies them into predefined categories such as person, date, location, and organization. RE finds and extracts semantic relationships between entities (e.g., persons, organizations, drugs, genes, etc.) in the text. For example, given the sentence "Steve Jobs co-founded Apple Inc. in 1976", an RE system can extract relations such as Founder Of [Steve Jobs, Apple Inc.] or Founded In [Apple Inc., 1976].

Text summarization techniques automatically produce a summary of a single or multiple documents. The resulting summary conveys the key information in the original text(s). Applications include scientific and news articles, advertisements, emails, and blogs. Broadly speaking, summarization follows two approaches: the extractive approach and the abstractive approach. In extractive summarization, a summary is created from the original text units (usually sentences). The resulting summary is a subset of the original document. Based on the extractive approach, formulating a summary involves determining the salient units of a text and stringing them together. The importance of the text units is evaluated by analyzing their location and frequency in the text. Extractive summarization techniques do not require an 'understanding' of the text. In contrast, abstractive summarization techniques involve extracting semantic information from the text. The summaries contain text units that are not necessarily present in the original text. In order to parse the original text and generate the summary, abstractive summarization incorporates advanced Natural Language Processing (NLP) techniques. As a result, abstractive systems tend to generate more coherent summaries than the extractive systems do (Hahn & Mani, 2000). However, extractive systems are easier to adopt, especially for big data.

Question answering (QA) techniques provide answers to questions posed in natural language. Apple's Siri and IBM's Watson are examples of commercial QA systems. These systems have been implemented in healthcare, finance, marketing, and education. Similar to abstractive summarization, QA systems rely on complex NLP techniques. QA techniques are further classified into three categories: the information retrieval (IR)-based approach, the knowledge-based approach, and the hybrid approach. IR-based QA systems often have three sub-components. First is the question processing, used to determine details, such as the question type, question focus, and the answer type, which are used to create a query. Second is document processing which is used to retrieve relevant pre-written passages from a set of existing documents using the query formulated in question processing. Third is answer processing, used to extract candidate answers from the output of the previous component, rank them, and return the highest-ranked candidate as the output of the QA system. Knowledge-based QA systems generate a semantic description of the question, which is then used to query structured resources. The Knowledge-based QA systems are particularly useful for restricted domains, such as tourism,

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 4, Issue 12, December 2017

medicine, and transportation, where large volumes of pre-written documents do not exist. Such domains lack data redundancy, which is required for IR-based QA systems. Apple's Siri is an example of a QA system that exploits the knowledge-based approach. In hybrid QA systems, like IBM's Watson, while the question is semantically analyzed, candidate answers are generated using the IR methods.

B. Audio Analytics.

Audio analytics analyze and extract information from unstructured audio data. Currently, customer call centers and healthcare are the primary application areas of audio analytics. Call centers use audio analytics for efficient analysis of thousands or even millions of hours of recorded calls. These techniques help improve customer experience, evaluate agent performance, enhance sales turnover rates, monitor compliance with different policies (e.g., privacy and security policies), gain insight into customer behavior, and identify product or service issues, among many other tasks. Audio analytics systems can be designed to analyze a live call, formulate cross/up-selling recommendations based on the customer's past and present interactions, and provide feedback to agents in real time.

In healthcare, audio analytics support diagnosis and treatment of certain medical conditions that affect the patient's communication patterns (e.g., depression, schizophrenia, and cancer) (Hirschberg, Hjalmarsson, & Elhadad, 2010). Also, audio analytics can help analyze an infant's cries, which contain information about the infant's health and emotional status (Patil, 2010). Speech analytics follows two common technological approaches: the transcript-based approach (widely known as large-vocabulary continuous speech recognition, LVCSR) and the phonetic-based approach. These are explained below.

LVCSR systems follow a two-phase process: indexing and searching. In the first phase, they attempt to transcribe the speech content of the audio. This is performed using automatic speech recognition (ASR) algorithms that match sounds to words. The words are identified based on a predefined dictionary. If the system fails to find the exact word in the dictionary, it returns the most similar one. The output of the system is a searchable index file that contains information about the sequence of the words spoken in the speech. In the second phase, standard text-based methods are used to find the search term in the index file. Phonetic-based systems work with sounds or phonemes. Phonemes are the perceptually distinct units of sound in a specified language that distinguish one word from another (e.g., the phonemes/k/and/b/differentiate the meanings of “cat” and “bat”). Phonetic-based systems also consist of two phases: phonetic indexing and searching. In the first phase, the system translates the input speech into a sequence of phonemes. This is in contrast to LVCSR systems where the speech is converted into a sequence of words. In the second phase, the system searches the output of the first phase for the phonetic representation of the search terms

C.Video Analytics.

Video is a major issue when considering big data. Videos and images contribute to 80 % of unstructured data. For example YouTube has innumerable videos being uploaded every minute containing massive information. Apart from videos, surveillance cameras generate a lot of information in seconds. Video analytics is still in its infancy compared to other types of data mining (Panigrahi, Abraham, & Das, 2010), various techniques have already been developed for processing real-time as well as pre-recorded videos. The increasing prevalence of closed-circuit television (CCTV) cameras and the booming popularity of video-sharing websites are the two leading contributors to the growth of computerized video analysis. A key challenge, however, is the sheer size of video data. To put this into perspective, one second of a high-definition video, in terms of size, is equivalent to over 2000 pages of text (Manyika et al., 2011).

The primary application of video analytics in recent years has been in automated security and surveillance systems. In addition to their high cost, labor-based surveillance systems tend to be less effective than automatic systems (e.g., Hakeem et al., 2012 report that security personnel cannot remain focused on surveillance tasks for more than 20 minutes). Video analytics can efficiently and effectively perform surveillance functions such as detecting breaches of restricted zones, identifying objects removed or left unattended, detecting loitering in a specific area, recognizing suspicious activities, and detecting camera tampering, to name a few. Upon detection of a threat, the surveillance system may notify security personnel in real time or trigger an automatic action (e.g., sound alarm, lock doors, or turn on lights).

Another potential application of video analytics in retail lies in the study of buying behavior of groups. Among family members who shop together, only one interacts with the store at the cash register, causing the traditional systems to miss data on buying patterns of other members. Video analytics can help retailers address this missed opportunity by providing information about the size of the group, the group's demographics, and the individual members’ buying behavior. Automatic video indexing and retrieval constitutes another domain of video analytics applications. The widespread emergence of online and offline videos has highlighted the need to index multimedia content for easy search and retrieval. The indexing of a video can be performed based on different levels of information available in a video including the metadata, the soundtrack, the transcripts, and the visual content of the video. In the metadata-based approach, relational database In terms of the system architecture, there exist two approaches to video analytics, namely server-based and edge-based.

Server-based architecture. In this configuration, the video captured through each camera is routed back to a centralized and dedicated server that performs the video analytics. Due to bandwidth limits, the video generated by the source is usually compressed by reducing the frame rates and/or the image resolution. The resulting loss of information can affect the accuracy of the analysis. However, the server-based approach provides economies of scale and facilitates easier maintenance.

Edge-based architecture. In this approach, analytics are applied at the 'edge' of the system. That is, the video analytics is performed locally and on the raw data captured by the camera. As a result, the entire content of the video stream is available for the analysis, enabling a more effective content analysis. Edge-based systems, however, are more costly to maintain and have a lower processing power compared to the server-based systems.

D.Social Media Analytics & Research

Many websites dedicated to social media are among the most popular—Wikipedia (collective knowledge generation), MySpace and Facebook (social networking), YouTube (social networking and multimedia content sharing), Digg and Delicious (social browsing, news ranking, and bookmarking), Second Life (virtual reality), and Twitter (social networking and microblogging), to name just a few. Because social media is already a critical part of the information ecosystem and as social media platforms and applications gain widespread adoption with unprecedented reach to users, consumers, voters, businesses, governments, and nonprofit organizations alike, interest in social media from all walks of life has been skyrocketing from both application and research perspectives. For-profit businesses are tapping into social media as both a rich source of information and a business execution platform for product design and innovation, consumer and stakeholder relations management, and marketing. For them, social media is an essential component of the next-generation business intelligence platform. For politicians, political parties, and governments, social media represents the ideal vehicle and information base to gauge public opinion on policies and political positions as well as to build community support for candidates running for public offices. Public-health officials could potentially use social media as valuable, early clues about disease outbreaks and to provide feedback on public-health policies and response measures. For homeland security and intelligence analysis communities, social media presents Immense opportunities to study terrorist group behavior, including their recruiting and public relation schemes and the grounding social and cultural contexts. Even think tanks and social science and business researchers are conceptually using social media as an unbiased sensor network and a laboratory for natural experimentation, providing valuable indicators and helping test hypotheses about social production and interactions as well as their economic, political, and societal implications. For many individuals, social media has become a unique information source to deal with information- and cognitive-overload problems, find answers to specific questions, and discover more valuable opportunities for social and economic exchange. In addition, it has become a platform for them to network and contribute to all kinds of dynamic dialogues by sharing their expertise and opinions. It is safe to claim that social media has already penetrated a spectrum of applications with remarkable impact. Given the continued interest and the ever-growing information and meta-information generated through social media, it is expected to continue enabling new exciting applications and revolutionizing many existing ones.

III. SENTIMENT ANALYSIS/OPINION MINING

Sentiments can be described as emotions, judgements, opinions or ideas prompted or coloured by emotions. Sentiment analysis (also sentiment mining, sentiment classification, opinion mining, subjectivity analysis, review mining or appraisal extraction and in some cases polarity classification) deals with computational treatment of opinion, sentiment and subjectivity in text.Humans are subjective creatures and opinions are important. Being able to interact with people on that level has many advantages for information systems. Comparatively few categories (positive/negative, 3 stars, etc) compared to text categorization Crosses domains, topics, and users Categories not independent (opposing or regression-like) Characteristics of answers to opinion- based questions are different from fact- based questions, so opinion-based information extraction differs from trad information extraction. Some of the challenges in Sentiment Analysis are: People express opinions in complex ways, in opinion texts, lexical content alone can be misleading. Another challenge can be in the form of Intra-textual and sub-sentential reversals, negation, topic change common. Humans tend to express a lot of remarks in the form of sarcasm, irony, implication etc. which is very difficult to interpret. For Example- "How can someone sit through the

IFERP
connecting engineers... developing research

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 4, Issue 12, December 2017

movie" is extremely negative sentiment yet contains no negative lexographic word. Even if a opinion word is present in the text ,their can be cases where a opinion word that is considered to be positive in one situation may be considered negative in another situation.People can be contradictory in their statements. Most reviews will have both positive and negative comments, which is somewhat manageable by analyzing sentences one at a time. However, in the more informal medium like twitter or blogs(Social media), the more likely people are to combine different opinions in the same sentence which is easy for a human to understand, but more difficult for a computer to parse. Sometimes even other people have difficulty understanding what someone thought based on a short piece of text because it lacks context. A good example would be "The laptop is good but I would prefer, the operating system which I was using". There is a huge demand of sentiment analysis. Before buy-ing any product it is a practice now, to review its rating as rated by other persons who are using it. Online advice and Recommendations the data reveals is not the only reason behind the buzz in this area. There are other reasons like the company wants to know "How Successful was their last campaign or product launch" based upon the sentiments of the customers on social media. Sentiment analysis concentrates on attitudes, whereas traditional text mining focuses on the analysis of facts. There are few main fields of research Predominate in Sentiment analysis: sentiment classification, feature based Sentiment classification and opinion summarization. The task of Sentimant Analysis can be broadly classified into three levels: document level, sentence level, and feature based approaches (aspect level).[Bing Liu]

3.1 Document level Sentiment Analysis

Classification of the overall sentiment of a document based on the overall sentiment of the opinion holder. This problem is basically a text classification problem. Here in general it is assumed that the document is written by a single person and expresses opinion about a single entity. One of the major challenge in the document level classification is that all the sentence in a document may not be relevant in expressing the opinion about an entity. Therefore subjectivity/objectivity classification is very important in this type of classification. The irrelevant sentences must be eliminated from the processing works. Both supervised and unsupervised learning methods can be used for the document level classification. Any supervised learning algorithm like nave Bayes classifier, Support Vector Machine, or Maximum Entropy etc can be used to train the system. For training and testing data, the reviewer rating (in the form of 1-5 stars), can be used. The features that can be used for the machine learning are term frequency, adjectives from Part of speech tagging, Opinion words and phrases, negations, dependencies etc. Labeling the polarities of the document manually is time consuming and hence the user rating available can be made use of. The unsupervised learning can be done by extracting the opinion words inside a document. The point-wise mutual information can be made use of to find the semantics of the extracted words. Thus the document level sentiment classification has its own advantages and disadvantages. Advantage is that we get an overall polarity of opinion text about a particular entity from a document. Disadvantage is that the different emotions about different features of an entity could not be extracted separately.

3.2 Sentence level Sentiment Analysis

Here polarity of each contributing senetence is derived. Again, here the assumption is that each sentence is written by a single person and expresses a single positive or negative opinion/sentiment. Sometimes Document-level sentiment classification is too coarse for our pupose. One of the reason can be, the size of the document is too large or a more granular level of sentiments needs to be derived. A lot of early work in the region of sentence level analysis focuses on identifying subjective sentences. Most techniques use supervised learning. This can be divided into two task: first identify which sentence old opinion (subjective sentences) and then classify each sentence as positive/negative or the star rating. But there will be complex sentences also in the opinionated text. In such cases, sentence level sentiment classification is not desirable. Knowing that a sentence is positive or negative is of lesser use than knowing the polarity of a particular feature of a product. The advantage of sentence level analysis lies in the subjectivity/ objectivity classification. Some challenges in this approach could be: many objective sentences can imply sentiments or Many subjective sentences do not express positive or negative sentiments/opinions. A single sentence may contain multiple opinions and subjective and factual clauses.

3.3 Feature based or Aspect level Sentiment Analysis

A more granular approach that gives some extra information. For example" Sentiment classification at both the document and sentence (or clause) levels are useful, but they do not find what people liked and disliked. The product or the review.They do not identify the targets of opinions. Much of the research is based on online reviews and blog related data. In the cae of reviews, where the entity (product or service) is known. Its a easier problem. But for blogs, forum discussions, etc., it much harder because the entity is unknown there may also be many comparisons, and there is also a lot of irrelevant information. This problem is somewhat similar to the problem of Named Entity Resolution.

IV. CHALLENGES OF TEXT MINING

Sentiment Analysis approaches seek to study the semantic properties of the words from a text and classify them as positive or negative , and in the contexts wherein the words don't fall under either of these two categories, then, they shall be classified as objective (read as non-sentiment bearing word). The general challenges are summarized as follows:

1. Anaphora Resolution - The word 'anaphora' implies a backward looking reference. The problem involved in solving what a pronoun or a noun phrase refers back is called Anaphora Resolution . For Example : We had a great party and we enjoyed it ourselves; 'it' was a nice moment. What does 'it' refer to ?

2. It is observed that Social Networking Platforms are replete with a kind of language which is marked by poor spelling, poor punctuation, poor grammar, hash tags, emoticons in Twitter and Facebook etc.

3. Implicit Sentiment –A sentence might indicate an implicit sentiments without the morphological units suggesting the presence of any such sentiment words like good , better , best , worst ,bad etc. but the sentences may have its positive or negative feedback about the product , services and policies. Consider the following examples.

It is hard to think that the writer of this book has not lost his balance Going by car takes much less time than going by train. These above sentences bear examples of showing the negative sentiment but without using any negative sentiment word .

4. Sarcasm and Ridicule –Sentences suggesting sarcasm and ridicule try to turn the thoughts upside down by exploiting the semantic and syntactic arrangements. Such sentences demand undivided attention on the part of reader to fathom out the meaning in its true spirit. These sentences not only mislead us but also mean the just opposite of what is apparently stated.

Example : Who will listen to this great 'pragaspathi'? He can't hold of weight of his own head. Who will trust him for guidance ?

5. Thwarted Expectations- In this context, the author creates an initial impression of building up glorifying adjectives ,but towards the end , he thwarts the readers expectation by the sudden swift of turn around.

Example :He speaks French so beautifully , but French of France is still unknown to him.

The above sentence carries a feature called 'anti-climax'. The author builds up the tempo of positive only to drag it to the crucial negative conclusion at the end.

6. World Knowledge- Sometimes we need to have some information about the historical background of the word to detect and determine sentiments .
For example : It is a catch -22 situation.

The sentiment of this sentence can be understood only when the reader has prior information about the Joseph Keller's book Catch-22. Also the sentiment of the sentence ' He is a Frankenstein' can be detected only when the reader knows about the damaging influence of the grotesque creation of the scientist which has become a force of destruction itself for the scientist) Hence, in order to decipher the sentiments, world knowledge has to be incorporated into the system .

7. Subjectivity Detection – There is a difference between opinionated and non-opinionated text. Subjectivity detection helps to filter out objective facts from opinionated statement. But, this is often difficult because of the complex nature of the operative functions of grammar.
Example :
I enjoy writing novels.
I do not like the book, ' I enjoy writing novels' .

The first statement bears an objective fact about the likes of the speaker whereas the second statement presents an opinion about a particular book. It is to be noted that in the second sentence the object of the sentence is 'book' and it is apposition is the clause that follows it.

8. Entity Recognition – A text or sentence may have several distinctive entities. Hence, it is important to know the entity towards which an opinion is generated.

Example :
Nizar is more intelligent than Nizam Football is more entertaining than Cricket

In the above examples, positive opinions are generated for Nizar and Football and negative ones are generated for Nizam and Cricket

9. In text analytics, we find sometimes the same word will serve as positive for one entity in a particular domain and the negative in another domain .

For example :
It is a cheap price and cheap product too.

A positive or negative sentiment world may have their opposite meaning in a particular domain so it is hard to predict by its keyword meaning. In the above sentence the word 'cheap' in one context means 'not very costly' (which is a positive sentiment) and 'not of good quality' in another context (which is negative sentiment).

10. Interrogative Sentence: An interrogative sentence is one which either seeks information through asking questions or seeks confirmation through questions. These sentences neither reveal positive or negative sentiments. But the key words used in the Interrogation may suggest positive or negative sentiments.

Example : What are advantages does Samsung enjoy over Apple ?

Can you tell me if Indian gooseberry is better than Kiwis. The above sentences don't suggest positive or negative sentiments , but they present valuable information with which they can be interpreted .

11. Conditional sentences – This sentence suggest performance of an action based on a condition stated . It neither suggests positive or negative sentiments .
Example :If you are good , I will be good.
The above example does not state if the speaker is good or bad , but it does suggest that the condition of being good depends on the condition of something else being met.

12. Author and Reader understanding point (person to person varying): Sometimes, a sentence conveys a sentiment which may be positive for one person and negative for the other.

For example : The prices of tomatoes have been slashed . This sentence carries positive sentiment for the buyers , but for the farmers , it suggests negative sentiment.

13. Spam Reviews –These are the kinds of sentiments posted by the competitor or opposite parties or organization in order to increase the value of their products in contrast with their counter parts products . These are otherwise called Spam Reviews .

14. Rhetorical questions: These are the kinds of questions which are raised for the sake of theatrics or for the sake of seeking confirmation from the audience, but not in expectation of any answers from the readers or listeners. Hence, they should be treated as simply as queries, but they could suggest positive sentiment or negative depending upon the context. Hence, they should be treated only as neutral .
Example :What is the meaning of this life ?
The above question suggests either the speaker wants to know the meaning of life or the speaker is so frustrated and feels empty and in that state of utter helplessness, he says that life has no meaning at all . Hence, this utterance cant be taken in face value unless the context is clear.

V. CONCLUSION

To address these issues of Text mining, we will propose a novel methodology for multimodal sentiment analysis, which consists in harvesting sentiments from Web videos by demonstrating a model that uses audio, visual and textual modalities as sources of information. We used both feature- and decision-level fusion methods to merge affective information extracted from multiple modalities. Deep learning architectures like Convolutional Neural Network and Recurrent Neural Networks will be helpful to explore the research in Multimodal Sentiment Analysis.

REFERENCES

1. Diana Maynard, Adam Funk. Automatic detection of political opinions in tweets. In: Proceedings of the 8th international conference on the semantic web, ESWC'11; 2011

2. Hastie, T., Tibshirani, R., & Friedman, J. (2009). The Elements of Statistical Learning: Data Mining, Inference, and Prediction.Springer.

3. Hu Minging, Liu Bing. Mining and summarizing customerreviews. In: Proceedings of ACM SIGKDD international conference on Knowledge Discovery and Data Mining (KDD'04);2004.

4. Liu B. Sentiment analysis and opinion mining. Synth Lecture Human Language Technology 2012.

5. Machine Learning with Naïve Bayes Classifier [Online].Available:http://blog.datumbox.com/machine-learningtutorial-the naive-bayes-text-classifier/

6. TsytsarauMikalai, Palpanas Themis. Survey on mining subjective data on the web. Data Mining Knowledge Discovery.

7. Wilson T, Wiebe J, Hoffman P. Recognizing contextual polarity in phrase-level sentiment analysis. In: Proceedings of HLT/EMNLP; 2005

8. Yelena Mejova, Padmini Srinivasan. Exploring feature definitionand selection for sentiment classifiers. In: Proceedings of the fifthinternational AAAI conference on weblogs and social media;2011.