Sentiment Analysis and Opinion Mining within Social Networks using Konstanz Information Miner

Banan Awrahman[1], Bilal Alatas[2]
[1]*Department of Field Crops, Halabja Technical Agriculture College, Sulaimani Polytechnic University, Sulaimani, Iraq*.
[2]*Department of Software Engineering, Firat University, Elazig, Turkey*.
balatas@firat.edu.tr

***Abstract*—Evaluations, opinions, and sentiments have become very obvious due to rapid emerging interest in e-commerce which is also a significant source of expression of opinions and analysis of sentiment. In this study, a general introduction on sentiment analysis, steps of sentiment analysis, sentiments analysis applications, sentiment analysis research challenges, techniques used for sentiment analysis, etc., were discussed in detail. With these details given, it is hoped that researchers will engage in opinion mining and sentiment analysis research to attain more successes correlated to these issues. The research is based on data input from web services and social networks, including an application that performs such actions. The main aspects of this study are to statistically test and evaluate the major social network websites: In this case Twitter, because it is has rich data source and easy within social networks tools. In this study, firstly a good understanding of sentiment analysis and opinion mining research based on recent trends in the field is provided. Secondly, various aspects of sentiment analysis are explained. Thirdly, various steps of sentiment analysis are introduced. Fourthly, various sentiment analysis, research challenges are discussed. Finally, various techniques used for sentiment analysis are explained and Konstanz Information Miner (KNIME) that can be used as sentiment analysis tool is introduced. For future work, recent machine learning techniques including big data platforms may be proposed for efficient solutions for opinion mining and sentiment analysis.**

***Index Terms*—Konstanz Information Miner (KNIME); Opinion Mining; Sentiment Analysis.**

I. INTRODUCTION

The rapid utilization of the internet and interactive activities such as ticket booking, blogging, and chatting has led to the extraction, transformation loading and analysis of large amounts of data at a rapid rate, which is called big data. Such data may be analyzed by utilizing and combining extracted data from web and text mining techniques in various actual applications. The vast amount of information related to customers' feedback / reviews is quite cumbersome for analysis and needs a specific technique to have a common summary of the opinion. Various e-commerce sites, social network, news reports, and blog forums are used as platforms for expressions which could be used to understand the public opinion and consumers' preferences, political movements, social gathering events, product preferences, marketing campaigns, reputation monitoring [1]. To achieve these tasks, the scientific community and academia are strictly working on the analysis of the attitudes in the last one and a half decade. A study of computational analysis of feelings, emotions, opinions, and attitudes expressed in the texts in relation to the nature is called sentiment analysis [2]. Analysis of attitudes also known as opinion mining or extraction of the review assessment analysis ratio is the job of discovering, retrieval, classification opinions, attitudes, and views on various themes expressed in the text entry.

Evaluations, opinions, and sentiments have become apparent due to emerging interest to e-commerce which is also a significant source of expression of opinions and sentiment analysis [3]. Currently, e-commerce sites' customers depend mainly on the existing customer service providers and manufacturers' reviews posted online in order to make a clear analysis posted by customers to their service and product standards and quality improvements. For example, the reviews given in the e-commerce websites like eBay, Asos, and Amazon may affect customers' decisions in purchasing products and subscribe to services [4]. Social media and the internet are generally replacing offline media rapidly in developed nations because they encourage ordinary people to engage in political debate and give them the enablement to bring one-sided thinking on international issues in an interactive mode. Broad exchange of idea platforms are provided by social media sites that encourage group conversations with views that are open for the public. The best means to achieve fast feedbacks and responses on various international issues and organizations in the shape of video text messages and new images are provided by online media. Therefore, opinions provided by people can be analyzed for the study of consumer behavior market models and trends in the societies [5].

Two hundred and fifty five million people login on Twitter on a monthly basis and it manages five hundred million tweets per day [6]. Therefore, it can be utilized to extract diverse opinions of groups of people from different backgrounds for various purposes such as improving services and products. Social media sites and online media platforms are utilized for public expressions and experience sharing in discussions, blogs, and reviewing of products publicly. The obtained information contains data that are highly unstructured such as combined texts, image animations, and videos which are very useful in public decision making on various issues [7].Pang et al. [8] conducted an in-depth research of more than 300 published scholarly articles by covering applications of common problems for analysis of sentiment and opinion mining major tasks, namely mining sentiment classification opinion and the definition of polarity and synthesis. At that point, Tang et al. [9] discussed four major issues related to mining of opinions that are the views of classification of word sediment, classification of subjectivity, extracting opinions and classification of sentiment documents. For the classification of subjectivity, they highlighted some approaches such as scaling dependent multiple (Naïve

Bayes) NB classifier cutting through the classifier and NB classifier.

O'Leary et al. [10] exhibited an overview on web journal mining, which incorporates presentations on online journal pursuit and a mining sort of web journal and the kind of sentiments that can be analyzed and their application to be separated from the sites.

Montoyo et al. [11] recorded some open problems alongside the accomplishments acquired up to this point in the field of analysis of sentiment and analysis of subjectivity.

Tsytsarau and Palpanas [12] introduced a study on sentiment analysis by concentrating on feeling mining assessment accumulation, including spam identification and disagreement analysis. They analyzed feeling mining techniques which were utilized on some basic data set.

Cambria et al. [13] highlighted the complexities required in sentiment analysis as for ebb and flow demand alongside plausible future exploration possibilities. Feldman [14] concentrated on five specific issues in the area of sentences: level document, level sentiment analysis, perspective based sentiment analysis, relative sentiment analysis and acquisition of lexicon sentiment. They additionally recorded few clear problems like statement writing, sentiment analysis, programmed element acknowledgment, dialogue on a multi-element in the same audit, mockery discovery, and subjectivity order at a better level.

Liu [15] displayed diverse published undertakings and works in sentiment analysis and sentiment mining. Significant undertakings records are generated by sentiment lexicon, subjectivity, summarization of opinion, sentiment analysis aspect based, analysis of similar assessments, and detection of opinions spam, sentiment search, recovery and nature of audits.

Medhat et al. [16] introduced a review that highlighted choices and opinions order techniques. An exceptionally short portrayal about element choice strategies has been displayed and a detailed talk on classification techniques of sentiment analysis has been introduced. They outlined 54 articles stating the assignments achieved by utilizing algorithm space, extremity information scope of language type and source of data. The researchers' significant aim is to discuss the methods connected in the reviewed papers. Alongside these reviewed papers, a lot of studies have been cited in this field and the lexica of various types have been made by scientists to assess new areas of analysis sentiment algorithm. Particularly, in the last four years noteworthy worries of scientists are drawn on the miniaturized scale of web journals that have been effectively connected for business sector forecasts [17], social publicizing and film industry expectations [18].

This study will follow the mechanisms of data mining to identify the benefits of its algorithms and methods. The research will be based on data input from web services and social networks, including an application that performs such actions. The main aspects of this study are to statistically test and evaluate the major social network websites: in this case, the Twitter.

II. SENTIMENT ANALYSIS

Analysis of sentiment, also known as mining of opinion is an area of research that analyses sentiments, opinions of people, assessments, examinations, mentalities and feelings towards substances, for example the item administrations, organizing people's issues, occasion themes, and their characteristics. There are many marginally diverse tasks and names involved in the analysis of sentiment, such as the extraction of opinions, analysis of sentiment, mining of opinion and sentiments, analysis of subjectivity, analysis of influence, analysis of feeling, and mining of survey and so forth. Nevertheless, currently they fall within the group of analysis of sentiment or mining of opinion. Although the phrase 'analysis of sentiment' is commonly used in the industry, yet both the analyses of sentiment and the meaning of opinion are utilized most of the time in the community of education. This is because in principle they fundamentally represent the same area of research. The phrase 'analysis of sentiment' has been initially used [19], followed by the phrase 'mining of opinion' [20].

Despite the fact that philology and Natural Language Processing (NLP) has been studied and researched for quite some time, it was not until the year 2000 that very few studies have been conducted on sentiments and opinions. From that point forward, sentiment analysis and opinion mining have turned into an exceptionally active study field. There are a few explanations behind this phenomenon. To start with, it possesses a broad range of uses in many areas. The business side of analysis of sentiment has also developed because of the multiplication of business applications it carries. This gives a sound reasoning to dig into and invest in the area more. Second, it provides opportunities for new arrays of research subjects in many aspects that have never been focused before. Third, the social networking sites provide a vast amount of opinion-oriented information for the very first time in human history. It must be mentioned that the analysis of sentiment currently focuses on the research on social networking sites. This new trend in the analysis of sentiment highly affects NLP as profoundly as it impacts the social sciences, management sciences, economics and political sciences, and they are all influenced by individuals' opinions. Despite the fact that the analysis of sentiment study essentially began from mid-2000, there existed some prior works on the clarification of allegories sentiment, descriptive word subjectivity that view on focuses and influences [20, 21].

Mining of opinion, analysis of subjectivity and analysis of sentiment are interrelated fields of study which use different strategies taken from NLP retrieved from organized information and mining of unstructured data. A real piece of information generated around the world is unstructured by itself, such as content discourse sound, video and pose as a focal navigation challenges [22]. To manage such unstructured content, information customary techniques for NLP, such as data recovery and data extraction must be utilized [23]. Keeping in mind the objective is to achieve a feeling out of the extracted content that various exploration efforts have been seen lately, this leads to computerized analysis of sentiment, which is an augmented NLP territory of examination [24]. Analysis of sentiment is just not a solitary issue that can be managed alone, but rather it is a diverse issue [25]. Different strides are expected to perform mining of opinion from given writings since the writings for mining of opinion are originated from a few assets in assorted organization. Information procurement and pre-processing are mainly the normal sub-tasks needed for content mining and analysis of sentiment, which are

discussed in this segment. The steps of analysis of sentiment is shown in Figure 1.

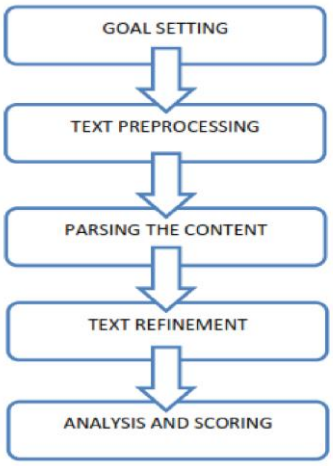

Figure 1: General steps of analysis of sentiment [26]

Figure 1 shows how text words are being scored. They are driven from a set of web pages. The first step is to set a clear goal. Afterwards, the texts undergo the pre-processing stage, in which it is forwarded to be read and organized as text blocks in order to make it easy for compilation. In parsing the content step, the text is being cut into token pieces. In the following step, the tokens are filtered and refined according to the regulation, which has been predefined. In the final step, the filtered tokens are scored and analyzed to the measure according to the prompted goal.

Data Acquisition

Because of the different online presence and wide availability, data retention by reinforcing media types are expected to perform the type of analysis and the dataset very subjectively. Some chatting sites like Facebook, Twitter gave access to its Application Programming Interface (API) to publicly collected information from their sites. Statistical information like user profiles' data were drawn from Twitter REST API and information like tweets were taken from Streaming API2 [27]. Twitter4J API3 has been misused to concentrate on a gushing number of tweets [28, 29]. Likewise, SnapChat and Facebook have made Tancent API5 and Facebook Graph API4 accessible. These APIs helped us to remove posts and other data from their site too, and they have been studied [30, 31]. Xu et al. [32] gathered data of around 23,507 companionships and 5,012 individuals from an item audit social media "UrCosme.com" and attempted to scale the peer effect of each person against one another to other Technorati. Also, Google Scholar is generally the most go-to site for internet researchers and it is also well studied [33].

The data set used in this research was taken from the internet in real-time. The model was based on the current date and time of data receiving and followed in an up-to-date schema. It gathered text fields as input vectors from the most frequently written phrases from Twitter. This data set has been followed by a test set in the same type using cross-validation method. A random percentage of training set was tested and scored against the error rate and performance.

Twitter has been used as a main source of data input in the project. There has been a collection of about 1k to 5k tweets from the USA, data received from (tender or eHarmony) set. Although, the tweets expired in one or two days, the main focus was on testing and evaluating the project. As its main case, there will be data about global warming and major worldwide topic throughout Twitter, the timing, will be captured but it is less important due to short life of tweets.

Pre-processing

Fully fledged raw data acquired from different sources must be pre-processed before they were analyzed. Some of the preprocessing steps are well-known, such as separating the tokens, word stop removal of parts of speech which include tagging, extraction, and representation. Some other technics were used to break a sentence into images or other important markers, such as highlighting marks, separating the tokens, uprooting words, and expressions. Stop words were not added to the analysis, and subsequently they were removed amid pre-processing steps. The procedure was caused by the need of different administrators to bring parts of speech and words to the root structure. The parts of speech in tagging content are crucial to use in the dialect discourse to detect different components. Due to their poor condition, printing data was a big challenge. They were often added to the pre-processing and it was one of the steps required to remove the highlights, and this required a compelling level. Besides removing the highlights, highlight identification is fundamental in achieving any analysis. [34] Reports feeling examinations investigate the Point Mutual Information (PMI), chi-square, and highlight the distinctive component selection methods such as sleep semantic indexing. Despite this, a little bit more is now written to offer measurable component removal strategies. It is summarized below.

III. SENTIMENT ANALYSIS APPLICATIONS

Aside from the actual practice, many applied research articles have also been published on this subject. For example, [34] predicted performance of sales with sentiment models. In [35], researches on the correlation between public opinions and betting line of NLP on Twitter and blogs were done. In [36], opinion polls of the public were linked with a sentiment from Twitter. In [37], views of politics were studied and in [38], a technique for political blogs comment related to prediction was investigated.

A. Applications in Business and Government Intelligence

Sentiment adds crucial connection to social discussions. Thus, it would be difficult and misleading to use estimation of notice only. It would be likely that when you were quantifying notice for your company's new items, you could expect an increase in the notice, and this meant that your product was welcomed. Everything taken into consideration, more says indicates that more people are talking about the product. However, it should be highlighted that some of the notices could be negative. Measuring sentiment assists one to comprehend the general feeling surrounding a specific subject, empowering you to make a more comprehensive and more complete picture of the social discussions that are important to you. This allows your organizations to track new discovery, new item observation, brand recognition, and notoriety management to name a few. The subjective information must be known as "the outline is affected",

"client administration was immediate" [39].

B. Applications in Sociology, Psychology and Political Sentiment Analysis

It permits people to have an opinion on something from a global view like motion picture surveys, political perspectives, and opinion over a worldwide issue, recognizing reasonableness of recordings in light of remarks, and so forth. Sentiment is the sense of feeling attached to a social media post. This way it helps the tone of the discussion to be put in a quantification level that can show whether the individual is satisfied, annoyed, or angry? For legislative issues, it can be used to analyze patterns, recognize an ideological tendency, target publicizing/messages as needed, and assessment of open/voters' opinions. In human sciences, thought engendering through gatherings is an important idea, the opinions that individuals have and their responses to thoughts are relevant to the selection of new thoughts. In brain researches, sentiment analysis gives a stage to enlarge mental examinations and tries different things with information and removes from regular language content [40].

C. Applications to Review-Related Websites

It is obvious the same abilities that an audit-arranged web user would have, could likewise serve remarkably well as the principle for the creation and mechanized running of survey and opinion collection sites. That is, as another option to locals like opinions that request input and audits, one could envision destinations that proactively accumulate such data. It's fundamental that points need not be limited to item surveys, but rather incorporate elements such as opinions about candidates in an election, political issues. There are also the utilizations of the technological advances we talk about two more conventional survey requesting locals, too. Abridging user surveys is an important issue. One could envision that mistakes in user appraisals could be settled: there are situations where users have unmistakably and unintentionally chosen a low evaluating when their survey shows a positive assessment [41].

D. Applications as a Sub-Component Technology

Mining of opinions and analysis of sentiment frameworks have an important potential part as an empowering advance for different frameworks. One probability is that an increase in the work suggested by the frameworks since it may profit from such a framework not to recommend things that get a great deal of negative criticism [42]. Recognition of "blazes" (excessively warmed or antagonistic language) in an email or different sorts of correspondence [43] is another possible use of subjectivity discovery and characterization. In online frameworks that show promotions as sidebars, it is useful to recognize website pages that contain touchy and substantially inappropriate items for advertisements arrangement [44]. For more advanced frameworks, it could be useful to raise item advertisements when relevant positive sentiments are identified and maybe all more important, cancel the advertisements when relevant negative articulations are found. Likewise, it has been opposed that data extraction can be improved by arranging data found in subjective sentences [45]. Sentiment analysis can also be useful in question answering. For instance, opinion-situated inquiries may require distinctive treatment and this will help to get that specific treatment [46].

E. Applications across Different Domains

One energizing and unexpected development has been the conversion of enthusiasm for opinions and sentiment in the field of software engineering field with the enthusiasm for opinions and sentiment in other types of fields. As it is well understood, opinions matter in an extraordinary amount in governmental issues. Some works have focused on understanding what voters are speculating [47], though different undertakings have been conducted as a long haul objective in the illumination of politicians' positions, for example, what public figures are backing or restricting to enhance the nature of data that voters have entered [48]. Sentiment analysis has particularly been proposed as a key empowering technology in e-Rulemaking, permitting the programmed analysis of the opinions that individuals submit about pending approach or government-control recommendations [49]. On a related note, there has been examination concerning opinion mining in weblogs which gave it to legitimate matters, in some cases known as "Blawgs" [50]. Connections with human sciences guarantee to be to a great degree of productivity. For instance, the issue of how thoughts and developments diffuse [51] includes the topic of who is emphatically or contrarily arranged towards whom, and consequently who might be pretty much open to new data transmission from a given source.

IV. Sentiment Analysis Research Challenges

As mentioned above, pervasive genuine applications are just part of the motivation behind why sentiment analysis is a prevalent examination issue. Moreover, it is exceptionally difficult as a natural language processing research subject, and covers many other new issues as well [52].

A. Various Analysis of Sentiment Levels

In general, there are three fundamental levels of studying sentiment analysis:

The first level is the Sentence: The assignment at this level reaches to the sentences and figures out if a sentence contains a positive, negative, or an impartial opinion. Impartial normally means the sentence does not contain any opinion. Subjectivity grouping can be seen at this level. The sentences that have been identified are referred as targeted sentences. They can separate the true data from sentences, known as subjective sentences which express subjective viewpoints and opinions. In any case, it should be clear that subjectivity is not the same thing as sentiment as the same number of target sentences can infer opinions, e.g., "We purchased the auto a month ago and the windshield wiper has tumbled off" [53].

The second level is the Document: The main work at this level is to group and show the communicated opinion's as positive or negative sentiment. For instance, if we look at a poll item, the grouping should show if the communication in general is in favor or against the item. This undertaking is normally known as report level sentiment order. It is assumed at this level that every report puts forth opinions on a single element (e.g., a solitary item). In this way, it would not be a proper way to record which one assesses or analyzes different elements [54].

The third level is the Aspect and Entity: Analysis of both

the above report level and the sentence level does not find what precisely individuals enjoyed or disliked. Perspective level performs better grained analysis. Perspective level was previously called highlight level. Rather than taking a look at language structure, this level takes a look at the opinion itself. This depends on the possibility that an opinion includes a sentiment (favored or disliked) and an objective (of opinion). Opinion that is not recognized but its objective has a limited usage. Furthermore, understanding the significance of opinion targets assists us to better comprehend the issues of sentiment analysis [55].

B. Sentiment Lexicon and Its Issues

Sentiment words, which are also known as opinion words, are the most important indicators of sentiments. Positive or negative sentiments are usually derived from those words. For example, wonderful, awesome, and spectacular are certain sentiment words that have positive sentiment. In the opposite side, awful, bad, and unpleasant are used as words that carry negative sentiment. Apart from the separate words, there are expressions and sayings, such as "out of the blue". These words and expressions are crucial in sentiment analysis for evident reasons. Putting the words and expressions together is a sentiment dictionary which is also called opinion vocabulary. Over the years, analysts have put together many calculations to create those vocabularies.

C. Opinion Spam Detection

A key component of social media is that it empowers anyone from anywhere on the planet to completely express his/her perspectives and opinions without unveiling his/her actual identity and without the fear of undesirable outcomes. These opinions are therefore profoundly profitable. This anonymity additionally accompanies a cost. It permits individuals with concealed plans or harmful expectations to effortlessly amuse the framework to give individuals the feeling that they are free individuals from people. In general, they post fake opinions to elevate or to dishonor target items, administrations, organizations, or people without revealing their actual aims, or the individual or organization that they are subtly working for. Such people are called opinion spammers and their exercises are called opinion spamming [56].

D. NLP Issues

Finally, we should not overlook that sentiment analysis is an NLP issue. It touches each parts of NLP; for example, nullification handlings, and word sense disambiguation, which include more troubles subsequent to these, have not been tackled issues in NLP. In any case, it is similarly useful to understand that sentiment analysis is a profoundly confined natural language processing issue because the framework does not have to completely understand the semantics of every sentence or report, however it just needs to understand a few parts of it, i.e., positive or negative sentiments and their objective substances or subjects. In this sense, sentiment analysis offers an incredible stage for natural language processing scientists to make tangible advances on all fronts of natural language processing with the capability of having a tremendous useful effect [57].

V. TECHNIQUES USED FOR SENTIMENT ANALYSIS

Comprehensively, there exist two sorts of techniques for analysis of sentiment: lexical-based and machine-learning-based.

Machine learning techniques often depend on directed order tactics, where sentiment discovery is enclosed as pair (i.e., positive or negative). This methodology prepares classifications based on marked information it obtains [58]. One of the advantages of learning-based strategies is their capacity to adjust and make prepared models for specific aims and settings, their disadvantages are the availability of named information and subsequently the low relevance of the technique for the new information. This is because marking information may be unreasonable or even restrictive for some undertakings. Lexical-based strategies make use of a specified use of words, in which each word is connected with a specific sentiment, and the strategies differ per the setting in which they were made [59].

Other techniques for analysis of sentiments are also discussed below.

A. Linguistic Inquiry and Word Count

Linguistic analysis and counting words are analysis of content apparatus that assesses enthusiastic, subjective, and auxiliary segments of a given content in view of the use of a lexicon that have words in them and their arranged classifications. Nevertheless, recognizing positive and negative effects in a given content, linguistic inquiry and word count give different arrangements of sentiment classifications. For instance, "concur" has a place with the accompanying word classes: consent, full of feeling, positive feeling, and intellectual procedure [60].

B. Emoticons

The least difficulty in recognizing extremity is the positive and negative impact of a message depends on the emoticons it contains. There has been a spread of emoticons lately, to the degree that a few (e.g. <3) are currently incorporated into English Oxford Dictionary. They are essentially confronting based and speak to gloat or dismal sentiments, despite the fact that an extensive types of them exist that are non-facial: for example, <3 is a symbol of a heart and communicates adoration or friendship. You might assume that there are more messages that have more than one emoticon and their collective number express feeling [61].

C. SentiStrength

Machine learning based strategies is appropriate for applications that need content-driven or versatile extremity recognizable proof models. One of the successful works by M. Thelwall, "Absolute entirety: Sentiment quality location in the social web with Sentistrength", looked at an extensive variety of managed and unsupervised order techniques, including straightforward logistic relapse, (Support Vector Machines) SVM, J48 grouping tree, JRip standard based Happiness Index which is used as a scale for sentiment that utilizes known effective English words [62].

D. SenticNet

SenticNet is a strategy that includes counterfeit consciousness and semantic Web procedures. The objective of SenticNet is to give the extremity of judgment skills, ideas from the regular language content at a level of semantic, not a syntactic level. The strategy uses (NLP) methods to make an extremity for almost 14,000 ideas [63].

E. Konstanz Information Miner (KNIME)

Undertakings of word reference making and sentiment analysis procedure are finished by the means of Konstanz Information Miner (KNIME) which is a user-accommodating graphical workbench capacities of the whole analysis process. KNIME uses six distinctive strides to process writings: perusing and parsing archives, named substance acknowledgment, shifting and manipulation, word numbering and catchphrase extraction, and transformation and perception. Taking after work processes and assignments are created and executed utilizing KNIME [64]:

- Retrieving data from a database
- Dictionary improvements and use
- Review scores

VI. Sentiment Analysis using Konstanz Information Miner

We have already used Twitter as a dataset because it has a rich data source and easy to work within KNIME tool. For collecting data from Twitter, we need to use Twitter API. Data were loaded from Twitter by an API connector that is shown in Figure 2. After tweet is chosen, it is converted to document.

After that, you need to remove punctuation marks such as (. , ; - .) and other commands (@, $, £, ….). Then, (http|com|rt|via|ht) were removed using "rexex filter" node. And then the document was tagged with part of speech to understand the language of your topic (noun, verb, adv….). And then they were added to a word bag with separation (noun, verb, adv., adjective,…) That is shown in Figure 3.

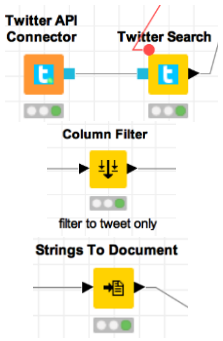

Figure 2: Parsing document

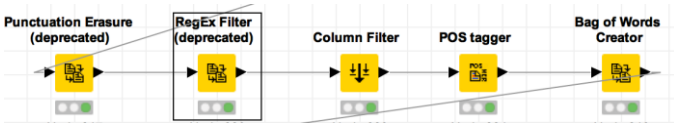

Figure 3: Steps of preprocessing

TF-IDF (Term Frequency-Inverse Document Frequency) is a significant number that one would need to create in KNIME if he needs to take out junk words and inspect only the crucial words in his analysis that is shown in Figure 4.

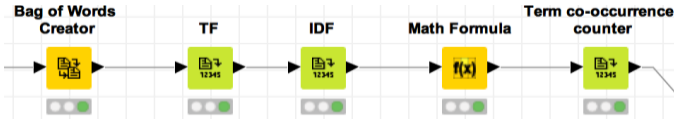

Figure 4: Calculation of TF-IDF

It calculates co-occurrence data from Twitter in an easy way. This is because of the node that is built in it and takes care of the whole analysis. This goes from a group of words node, one only needs to add-on the Term Co-Occurrence Counter node. This will help the production of two main columns which is from the co-occurring words. It will then create statistics regarding how frequently those terms show up besides each other. That is shown in Figure 5.

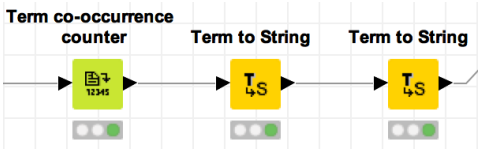

Figure 5: Calculation of frequency

Parallel Latent Dirichlet Allocation (LDA) nodes determine how many topics are identified for this model. And colors can be applied to segment the keywords beyond frequency or weight. Using the "Color Manager" node and feeding into the "Tag Cloud" node (the default word cloud visualization node built into KNIME), different colors to different topic can be applied as shown in Figure 6.

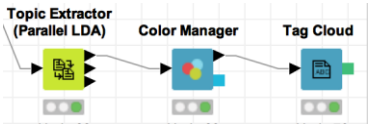

Figure 6: Visualizations

Figure 7 shows tag cloud visualizations result.

Figure 7: Tag cloud visualizations

As you can see many words with different colors. And you can see words like (eplileptic), (long), (prime), and (term) stand out they were used more frequently in the document.

VII. Conclusions

Sentiment analysis and opinion mining researches are indeed a good advancement in sciences and engineering fields. In this study, a general introduction on sentiment analysis, steps of sentiment analysis, sentiments analysis applications, research challenges, and techniques used for sentiment analysis were discussed in detail. With these details given, it is hoped that researchers will engage in sentiment analysis and opinion mining researches to attain more successes correlated to this issue.

In this study, firstly a good understanding of sentiment analysis and opinion mining researches that are based on recent trends in the field was provided. Secondly, various aspects of sentiment analysis were explained. Thirdly, various steps of sentiment analysis were introduced. Fourthly, various sentiment analysis, research challenges were discussed. Finally, various techniques used for

sentiment analysis were explained and KNIME that can be used as sentiment analysis tool was briefly introduced.

For future work, recent machine learning techniques or computational intelligence optimization techniques including big data platforms, such as Hadoop MapReduce programming are proposed for efficient solutions for opinion mining and sentiment analysis.

REFERENCES

[1] M. R. Saleh, M.T. Martín-Valdivia, A. Montejo-Ráez, L.A. Ureña-López, Experiments with SVM to classify opinions in different domains, Expert Systems with Applications 38 (2011) 14799–14804.

[2] W. Medhat et al. Sentiment analysis algorithms and applications: A survey, Ain Shams Eng J (2014), http:// dx.doi.org/10.1016/j.asej.2014.04.011

[3] A. Montoyo, P. Martínez-Barco, A. Balahur, Subjectivity and sentiment analysis: An overview of the current state of the area and envisaged developments, Decision Support Systems 53 (2012) 675–679.

[4] Y.M. Li, T.-Y. Li, Deriving market intelligence from microblogs, Decision Support Systems 55 (2013) 206–217

[5] D. Kang, Y. Park, Review-based measurement of customer satisfaction in mobile service: Sentiment analysis and VIKOR approach. Expert Systems with Applications (2013), http://dx.doi.org/10.1016/j.eswa.2013.07.101

[6] J. Bollen, H. Mao, X. Zeng, Twitter mood predicts the stock market, Journal of Computational Science 2 (2012) 1-8

[7] O. Popescu, and C. Strapparava, Time corpora: Epochs, opinions and changes, Knowledge-Based Systems 69 (2014): 3-13.

[8] B. Pang, L. Lee, Opinion mining and sentiment analysis, Foundations and Trends in Information Retrieval 2 (2008) 1–135

[9] H. Tang, S. Tan, X. Cheng, A survey on sentiment detection of reviews, Expert Systems with Applications 36 (2009) 10760–10773.

[10] D.E. O'Leary, Blog mining-review and extensions: "From each according to his opinion", Decision Support Systems 51 (2011) 821–830.

[11] A. Montoyo, P. Martínez-Barco, A. Balahur, Subjectivity and sentiment analysis: An overview of the current state of the area and envisaged developments, Decision Support Systems 53 (2012) 675–679.

[12] M. Tsytsarau, T. Palpanas, Survey on mining subjective data on the web, Data Min Knowl Disc (2012) 24:478–514, DOI 10.1007/s10618-011-0238-6

[13] E. Cambria, B. Schuller, Y. Xia, C. Havasi, New Avenues in Opinion Mining and Sentiment Analysis, Knowledge-Based Approaches to Concept-Level Sentiment Analysis, IEEE Intelligent Systems, 2013.

[14] R. Feldman, Techniques and Applications for Sentiment Analysis, Review Articles, Communications of the ACM, Vol. 56 No. 4, Pages 82-89, April 2013.

[15] B. Liu, Sentiment analysis and opinion mining, Morgan and Claypool publishers, May 2012.

[16] W. Medhat et al. Sentiment analysis algorithms and applications: A survey, Ain Shams Eng J (2014), http:// dx.doi.org/10.1016/j.asej.2014.04.011

[17] J. Bollen, H. Mao, X. Zeng, Twitter mood predicts the stock market, Journal of Computational Science 2 (2012) 1-8

[18] Y.-M. Li, Y.-L. Shiu, A diffusion mechanism for social advertising over microblogs, Decision Support Systems 54 (2012) 9–22

[19] J. Du et al., Box office prediction based on microblog. Expert Systems with Applications (2013), http://dx.doi.org/10.1016/j.eswa.2013.08.065

[20] T. Nasukawa, J. Yi, Sentiment analysis: Capturing favorability using natural language processing. In Proceedings of the KCAP-03, 2nd Intl. Conf. on Knowledge Capture (2003).

[21] K. Dave, S. Lawrence, D. M. Pennock, Mining the peanut gallery: Opinion extraction and semantic classification of product reviews. In Proceedings of International Conference on World Wide Web (WWW- 2003) (2003).

[22] V. Hatzivassiloglou, J. L. Klavans, M. L. Holcombe, R. Barzilay, M.-Y. Kan, K. R. McKeown, Simfinder: A flexible clustering tool for summarization. In Proceedings of the Workshop on Summarization in NAACL-01. (2001).

[23] J. Wiebe, R. F. Bruce, T. P. O'Hara, Development and use of a gold-standard data set for subjectivity classifications. In Proceedings of the Association for Computational Linguistics (ACL-1999). (1999).

[24] A. Montoyo, P. Martínez-Barco, A. Balahur, Subjectivity and sentiment analysis: An overview of the current state of the area and envisaged developments, Decision Support Systems 53 (2012) 675–679

[25] S. Poria, A. Gelbukh, A. Hussain, N. Howard, D. Das, S. Bandyopadhyay, Enhanced SenticNet with Affective Labels for ConceptBased Opinion Mining, Knowledge-Based Approaches to Concept-Level Sentiment Analysis, IEEE Intelligent Systems, (2013) 1.

[26] M. Godsay, The process of sentiment analysis: a study. International Journal of Computer Applications 127, 7 (2015), 26-30

[27] S. Kumar, F. Morstatter, H. Liu, Twitter Data Analytics, August 19, 2013, Springer.

[28] F.H. Khan et al., TOM: Twitter opinion mining framework using hybrid classification scheme, Decision Support Systems (2013), http://dx.doi.org/10.1016/j.dss.2013.09.004.

[29] E. Kontopoulos, C. Berberidis, T. Dergiades, N. Bassiliades, Ontology-based sentiment analysis of twitter posts, Expert Systems with Applications 40 (2013) 4065–4074.

[30] H. Bao, Q. Li, S. S. Liao, S. Song, H. Gao, A new temporal and social PMF-based method to predict users' interests in micro-blogging, Decision Support Systems 55 (2013) 698–709.

[31] W. Li, H. Xu, Text-based emotion classification using emotion cause extraction. Expert Systems with Applications (2013), http://dx.doi.org/10.1016/j.eswa.2013.08.073

[32] W. Medhat et al. Sentiment analysis algorithms and applications: A survey, Ain Shams Eng J (2014), http:// dx.doi.org/10.1016/j.asej.2014.04.011

[33] S. Wang, D. Li, X. Song, Y. Wei, H. Li, A feature selection method based on improved Fisher's discriminant ratio for text sentiment classification, Expert Systems with Applications 38 (2011) 8696–8702.

[34] G. Vinodhini, R. M. Chandrasekaran, Opinion mining using principal component analysis based ensemble model for e-commerce application, CSI Transactions on ICT (2014): 1-11.

[35] A. Abbasi, H. Chen, A. Salem, Sentiment Analysis in Multiple Languages: Feature Selection for Opinion Classification in Web Forums, ACM Transactions on Information Systems, Vol. 26, No. 3, Article 12, Publication date: June 2008.

[36] B. Pang, L. Lee, S. Vaithyanathan, Thumbs up? Sentiment classification using machine learning techniques, Proceedings of the ACL-02 conference on empirical methods in natural language processing (Vol. 10, pp. 79–86). Association for Computational Linguistics, 2002.

[37] B. Pang, L. Lee, A sentiment education: Sentiment analysis using subjectivity summarization based on minimum cuts, in: Proceedings of the 42nd annual meeting on Association for Computational Linguistics (p. 271), 2004, July.

[38] J. Liu, Yunbo Cao, Chin-Yew Lin, Yalou Huang, and Ming Zhou. Low-quality product review detection in opinion summarization. In Proceedings of the Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL-2007). 2007.

[39] B. O'Connor, Ramnath Balasubramanyan, Bryan R. Routledge, and Noah A. Smith. From Tweets to Polls: Linking Text Sentiment to Public Opinion Time Series. In Proceedings of the International AAAI Conference on Weblogs and Social Media (ICWSM 2010). 2010.

[40] A. Tumasjan, Timm O. Sprenger, Philipp G. Sandner, and Isabell M. Welpe. Predicting elections with twitter: What 140 characters reveal about political sentiment. In Proceedings of the International Conference on Weblogs and Social Media (ICWSM-2010). 2010.

[41] Lu´ıs Cabral and Ali Hortac¸su. The dynamics of seller reputation: Theory and evidence from eBay. Working paper, downloaded version revised in March, 2006. URL http://pages.stern.nyu.edu/˜lcabral/workingpapers/CabralHortacsu_Mar06.pdf.

[42] J. Tatemura, Virtual reviewers for collaborative exploration of movie reviews. In Proceedings of Intelligent User Interfaces (IUI), pages 272–275, 2000.

[43] Ellen Spertus. Smokey: Automatic recognition of hostile messages. In Proceedings of Innovative Applications of Artificial Intelligence (IAAI), pages 1058–1065, 1997.

[44] X. Jin, Y. Li, T. Mah, J. Tong, Sensitive webpage classification for content advertising. In Proceedings of the International Workshop on Data Mining and Audience Intelligence for Advertising, 2007.

[45] Ellen Riloff, Janyce Wiebe, and William Phillips. Exploiting subjectivity classification to improve information extraction. In Proceedings of AAAI, pages 1106–1111, 2005

[46] L. V. Lita, A. H. Schlaikjer, W. Hong, E. Nyberg, Qualitative dimensions in question answering: Extending the definitional QA

task. In Proceedings of AAAI, pages 1616–1617, 2005. Student abstract.
[47] M. Laver, K. Benoit, J. Garry, Extracting policy positions from political texts using words as data. American Political Science Review, 97(2):311–331, 2003.
[48] T. Mullen, R. Malouf, A preliminary investigation into sentiment analysis of informal political discourse. In AAAI Symposium on Computational Approaches to Analyzing Weblogs (AAAICAAW), pages 159–162, 2006.
[49] N. Kwon, S. Shulman, E. Hovy, Multidimensional text analysis for eRulemaking. In Proceedings of Digital Government Research (dg.o), 2006.
[50] J. G. Conrad, F. Schilder, Opinion mining in legal blogs. In Proceedings of the International Conference on Artificial Intelligence and Law (ICAIL), pages 231–236, New York, NY, USA, 2007. ACM
[51] E. Rogers, Diffusion of Innovations. Free Press, New York, 1962. ISBN 0743222091. Fifth edition dated 2003.
[52] C. Yubo, J. Xie, Online consumer review: Word-of-mouth as a new element of marketing communication mix. Management Science, 2008. 54(3): p. 477-491.
[53] W. Theresa, J. Wiebe, R. Hwa, Just how mad are you? Finding strong and weak opinion clauses. In Proceedings of National Conference on Artificial Intelligence (AAAI-2004). 2004.
[54] M. Hu, B. Liu, Mining and summarizing customer reviews. In Proceedings of ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD-2004). 2004.
[55] Z. Lei, B. Liu, Aspect and entity extraction for opinion mining, Data mining and knowledge discovery for big data. Springer Berlin Heidelberg, 1-40, 2014.
[56] N. Jindal, B. Liu, Opinion spam and analysis. In Proceedings of the Conference on Web Search and Web Data Mining (WSDM-2008). 2008
[57] N. Preslav, et al., SemEval-2016 task 4: Sentiment analysis in Twitter, In Proceedings of the 10th international workshop on semantic evaluation (SemEval 2016), San Diego, US (forthcoming). 2016.
[58] B. Pang, L. Lee, and S. Vaithyanathan. Thumbs up?: sentiment classification using machine learning techniques. In ACL Conference on Empirical Methods in Natural Language Processing, pages 79–86, 2002.
[59] Y. R. Tausczik and J. W. Pennebaker. The psychological meaning of words: Liwc and computerized text analysis methods. Journal of Language and Social Psychology, 29(1):24–54, 2010.
[60] J. Golbeck, Negativity and anti-social attention seeking among narcissists on Twitter: A linguistic analysis, First Monday (2016).
[61]] J. Park, V. Barash, C. Fink, and M. Cha. Emoticon style: Interpreting differences in emoticons across cultures. In International AAAI Conference on Weblogs and Social Media (ICWSM), 2013.
[62] P. S. Dodds and C. M. Danforth. Measuring the happiness of large-scale written expression: songs, blogs, and presidents. Journal of Happiness Studies, 11(4):441– 456, 2009.
[63] E. Cambria, R. Speer, C. Havasi, and A. Hussain. Senticnet: A publicly available semantic resource for opinion mining. In AAAI Fall Symposium Series, 2010.
[64] A. Mihanović, H. Gabelica, Ž. Krstić, Big Data and Sentiment Analysis using KNIME: Online Reviews vs. Social Media, In 37th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO), 1464–1468, 2014.