

Sentiment Analysis or Opinion Mining: A Review

Bilal Saberi[#], Saidah Saad[#]

*Centre for Artificial Intelligence Technology (CAIT), Faculty of Information Science and Technology,
Universiti Kebangsaan Malaysia (UKM), 43600 Bangi, Selangor, Malaysia
E-mail: bilal.sabri79@yahoo.com, saidah@ukm.edu.my*

Abstract— Opinion Mining (OM) or Sentiment Analysis (SA) can be defined as the task of detecting, extracting and classifying opinions on something. It is a type of the processing of the natural language (NLP) to track the public mood to a certain law, policy, or marketing, etc. It involves a way that development for the collection and examination of comments and opinions about legislation, laws, policies, etc., which are posted on the social media. The process of information extraction is very important because it is a very useful technique but also a challenging task. That mean, to extract sentiment from an object in the web-wide, need to automate opinion-mining systems to do it. The existing techniques for sentiment analysis include machine learning (supervised and unsupervised), and lexical-based approaches. Hence, the main aim of this paper presents a survey of sentiment analysis (SA) and opinion mining (OM) approaches, various techniques used that related in this field. As well, it discusses the application areas and challenges for sentiment analysis with insight into the past researcher's works.

Keywords— opinion mining; sentiment analysis; NLP; machine learning; lexical-based

I. INTRODUCTION

These days, sentiment analysis is gaining importance in the research study of text mining and natural language processing (NLP). There has been a rise in accessibility of online applications and a surge in social platforms for opinion sharing, online review websites, and personal blogs, which have captured the attention of stakeholders such as customers, organizations, and governments to analyze and explore these opinions. Therefore, the major role of sentiment classification is to analyze an online document such as a blog, comment, review and new items as a comprehensive sentiment and categories it as positive, negative, or neutral [1], [2].

Lately, the study of sentimental analysis has become popular among researcher scholars, and a number of research studies are being conducted on the subject. It is also known as opinion mining and sentiment classification. The sentimental analysis constitutes text classification and segregates sentiments for subjective texts, which are mainly related to consumer's reviews on products and services. Sentiments are classified into two: positive and negative sentiments. In a few cases, there may not be any sentiments, which are termed as neutral. The sentimental analysis is an intricate process, which consists of several tasks such as sentiment analysis (SA) subjectivity analysis, opinion mining (OM) and sentiment orientation. It is considered a novel, evolving new research field in machine learning (ML),

natural language processing (NLP) and computational linguistics. The sentiment analysis involves three major levels – word level, sentence level, and document level. The level of the analysis determines the task required for the process. The word level is the most complex one owing to the difficulty in carrying out the analysis, whereas the analysis is simpler at the sentence and document levels [3].

Semantic-based analysis and machine learning are the two major techniques used for the review of sentimental analysis. Also, a method is used to combine both the techniques. There have been many studies that have used machine-learning technique [4]-[10]. A Semantic-based analysis is a renowned technique of sentiment analysis [11], [12].

The remaining of this paper is arranged as the followings: Next section describes the sentiment analysis and opinion mining. After that, various levels of classifying sentiments are presented. The description of the previous work that was done on sentiment analysis techniques is present in Section IV. As well, Section V presents the resources of the sentiment analysis. Section VI describes the challenges in sentiment analysis. Finally, the conclusion of this review is declared in section VII.

A. Sentiment Analysis (SA)

In the natural language processing (NLP), Sentiment Analysis (SA) has become one of the many fields of computational studies [13], [14]. In general, SA deals with the mining of information related to sentiments or opinions

from a group for a specified topic. In addition, sentiments from some applications are received at the document level. Opinion-based summarization, emotion or mood extraction, and genre distinctions are taken into account for other research studies concerning sentiment or opinion-based studies.

SA has gained much popularity among various fields such as politics (to forecast election outcomes from different political forums) [15], business (to analyze online sentiments in the social media for stock market prediction) [16] and marketing (to estimate sales of specific products) [7]. When carrying out SA, it is assumed that the documents would contain opinions. However, in most cases, only objective information and facts are stated in these documents (one such example is the news document). Also, at times, documents assumed to contain sentiments (opinions) may include factual sentences as a part. Therefore, identifying the type and nature of sentences form the most fundamental part of SA. Therefore, based on the subjective or objective, sentences are extracted, categorised and used in the given analysis. Subjectivity classification forms the main task in SA, which involves classifying sentences as objective or subjective.

In general, SA involves a set of complex processes. The analysis is surrounded by a series of tasks, namely sentiment classification, subjective analysis, opinion holder extraction, and aspect or object-based extraction. The subjective analysis involves evaluating a text document or a sentence to label the same as subjective or objective. After this step, these documents or sentences tagged as objective are immediately discarded, as they are not much of use for the SA process. The sentiment classification involves the task of examining the sentiment polarity of the filtered sentences. These sentences are categorised as neutral, negative or positive sentiment depending on the case. One of the most important tasks is the aspect or object-based extraction that is the prime focus element in SA. In some cases, the opinion holder extraction task becomes highly vital because, at times, it is essential to know the author of the opinion.

1) Subjectivity Classification

One of the main tasks involved in sentiment analysis (SA) is subjectivity classification. The main aim is to classify documents or sentences into one of the two classes; objective or subjective [17]. Usually, when dealing with SA, the sentences in the text are labeled as subjective and objective. The subjective sentences undergo the process of sentiment analysis. The objective sentences identified are discarded. This is because objective sentences represent factual information. Hence, they are no benefits to the overall process of SA.

In this process of SA, subjective sentences are important to the process since they contain sentiments or opinions. The views perspectives, assessments, values, claim thoughts, observations, personal beliefs, and opinions of individuals are expressed in subjective sentences [17]. These types of sentences are classified into different classes, which may be positive, negative, or neutral. They are classified as neutral if the opinion is neither positive nor negative. For example, “*I think this Canon camera has a very large lens.*” In this example, the opinion holder is giving an observation, or

opinion, about the feature, *lens* of the entity, *Canon camera*. The opinion holder claims that the camera has very large lens. However, when analyzed in terms of sentiments, it is not clear whether this sentence provides a negative sentiment (e.g., *therefore the camera cannot fit in my bag*) or a positive sentiment (e.g., *therefore this means that it takes better quality photos*). Consequently, in the analysis, this sentence would be classified as a neutral sentence.

There are numerous research studies on subjectivity classification as an individual issue [17]-[20]. In this way, objective sentences may be discarded, leaving only subjective sentences for analysis in terms of sentiments.

Several researchers that work with sentiment analysis (SA) have focused on a model that carries out the task of subjectivity classification [21]. They used semi-supervised machine learning approach (Naïve Bayes classifier and several binary options). Later, a model that used unsupervised machine learning approach being created for the task of subjectivity classification [1].

A naïve Bayes classifier also being used as a supervised machine learning approach, along with sentence similarity, for subjectivity classification [22]. One weakness in the use of supervised machine learning techniques is the annotation of a lot of training samples. Therefore, a bootstrapping technique is used to overcome this problem [21]. This technique can categorize training samples automatically.

Besides the use of English language in the research studies of subjectivity classification, there are several research works in the Arabic language [23] and the Urdu language [24]. [23] used support vector machine (SVM) as supervised machine learning for the subjectivity and sentiment analysis. As well, [24] used techniques such as bootstrap learning and resource sharing from a syntactically similar language.

B. Classification Levels

Generally, there are three different levels of the sentiment analysis (SA). Document-based is the first level of the SA. Sentence based is the second level, while word or phrase based is the third level.

1) *Document Level*: The first level of SA is document level based. In this level, the document is taken into consideration as a whole. So it is classified based on a comprehensive sentiment of the whole document. In this case, the opinion holder is often assumed a single individual or source [12]. Another issue of SA at the document level is sentiment regression [6], [25]-[28]. Some had used supervised learning to predict the rating scores for a document, in terms of the extent of positive or negative it was [6]. There is an approach to extract a linear based combination method of the polarities in a text document [3], [12]. A major issue in focusing only at the document level is at this level not all the sentences in the document that express opinions are subjective sentences. In order to obtain results that are more accurate after SA, it is more precise to look at each sentence individually. In this way, only the objective sentences may be discarded, while the subjective sentences are extracted for the analysis of sentiments. Hence, SA at the sentence level has become a major focus for research in SA.

2) *Sentence Level*: There is numerous research works in the classification of every sentence in a document or text as objective or subjective sentences. After this is done, the subjective sentences are analyzed. They attempted the classification of only subjective sentences for the analysis of sentiments [30]. Generally, machine learning is used to detect subjective sentences. Meanwhile, there is researcher proposed a model that uses a log probability rate and numerous root terms to create a score for every subjective sentence [7]. Subsequently, there is a model that integrates the sentiments of the terms for each sentence, to obtain the overall sentiment for that particular sentence being taken into consideration [26]. However, there is a limitation with sentiment analysis at the sentence level. Some objective sentences that actually contain sentiments may be missed the cut. Study the following example: "I purchased this mug last week, and it has formed cracks on its sides". The previous sentence is an example of an objective sentence, as it presents facts. Upon careful examination, it is found that the sentence has expressed an indirect sentiment. The opinion holder of the sentence has expressed a negative sentiment about the cracks on the sides of the mug. To overcome this problem, it is needed to take into consideration the sentiment analysis at a word or phrase level. The following section clarifies further about sentiment analysis at the word level.

3) *Word/Phrase Level*: It is imperative to take into consideration words or phrases when dealing with sentiment analysis (SA) at the word or phrase level. This is essential because a word is the smallest unit with a meaning that is available in a text. Therefore, this is the most detailed type of SA. As such, it has caught the attention of many researchers so that a number of research studies for SA at word level is simply countless. Earlier research studies take into consideration the polarity of both phrases and words, for implementation the sentiment classification at the sentence and document levels. As a result, it is common to generate lists of word lexicons manually and automatically. Generally, the word lexicons of SA include adjectives (lovely, nice, beautiful, amazing, old, bad, horrible), adverbs (quickly, slowly, poorly, horribly) and some verbs (like, hate, love, despise) [31], [32]. Sometimes, nouns such as (trash, junk) are also regarded as sentiments.

II. MATERIAL AND METHOD

Of the various approaches available for sentiment analysis (SA), only two main groups are widespread. The first group solves the problems of SA by implementing the machine learning approach. In this group, multiple techniques are employed in a bid to extract salient features that more accurately give information about the polarity of sentiments. The technique used is constantly monitored, as the process requires manually annotated corpus.

The second group uses a linguistically inclined method called the lexicon-based approach. According to reference [8], the investigation is initiated with words or sentences showing characteristics of semantic polarity.

As well, there is one more group, which is used to combine machine learning with the Lexicon group. This group named combination method or semi-supervisor.

A. Machine Learning Approach

It is necessary for two different collections of documents in the use of machine learning for the purpose of classification. These two collections of documents include training collection (used by the classifier to differentiate between text features) and test collection (employed in estimating the precision of the classifier). There are numerous machine learning (ML) approaches. They are developed for the purpose of classifying texts into negative or positive classes. The performances of the approaches; Support Vector Machines (SVM), Naive Bayes (NB), and Maximum Entropy (ME) with SA and classification are highly successful. The other approaches include ID3, Centroid Classifier, Winnow Classifier, K-Nearest Neighbour, and Association Rules mining approach.

Naïve Bayes (NB) classification method is commonly utilize for classifying text documents [10], [27], [33], [34]. This technique is based on a probabilistic model and employs cooperative probabilities of certain terms and their respective group for the estimation of the probability of a certain group, with a text document as input.

As well, the Support Vector Machine (SVM) being proposed as a classifier to solve the problems for recognizing pattern between two groups [18]. Support Vector Machine (SVM) aims to identify the best margin separation of the hyperactive plane between two groups of data. It was originally intended for solving separable cases, but can be extended to solve the linearly non-separable case by mapping the original data vector to spaces of higher dimensions. Many research workers consider the SVM classifier as the best method to use for the purpose of text classification [9], [10], [21], [34]-[37] have employed this method.

Besides using the individual machine learning approaches that have been mentioned earlier for sentiment classifying, many researchers have been conducted to make comparisons of the machine learning approaches and to select the best algorithm for sentiment classification. Agarwal et al. revealed that the Naïve Bayes (NB) classifier achieved exceptional results compared to Support Vector Machine (SVM) on reviews in Cantonese [34]. He focused his attention on Cantonese writing, which is a type of Chinese language. Zhang et al. also found the support vector machine (SVM) outperformed naive Bayes (NB) and n-gram model on reviews of mining purpose [33].

Chen et al. proposed a model that uses the k-means clustering method [38]. The documents are classified into two groups, a positive group, and a negative group. Then, the TF-IDF weighting feature method is implemented on the text. Then, a voting technique is conducted to retrieve the more stable results.

The Association Rules mining has attracted much of the researcher's attention to the various methods of data mining. The main reason for it is becoming a well-researched area of study. The methods of Association Rules mining involved an initial data exploration approaches. They are often applied to tremendously large data sets. An example is the market basket data in grocery stores [39].

The classification that uses Association Rules is also known as associative classification. It is a field of study in Data Mining (DM), which integrates association rule

discovery with classification. In fact, AR applies discovery association rule methods to discover the knowledge and then choose a group of rules to construct a classifier [40]. The idea behind the design of AR is building a model (classifier) that consists of some rules (extracted knowledge) from labelled input (training dataset utilised in the classification of prediction) for an unclassified precise test data [41].

One of the most common association rules mining algorithms is called Apriori algorithm. This algorithm is utilised in discovering rules (knowledge) [42]. An Apriori algorithm process is based on two important definitions: Support and Confidence, where the support of the itemset is some transactions in D that contain the itemset. An association rule is expressed as $X \Rightarrow Y$, where $X, Y \subseteq I$ are two sets of items and $X \cap Y = \emptyset$, D is the dataset, and I is a set of items. On the other hand, confidence refers to the probability that a transaction contains Y given that it contains X , so the confidence is provided as $\text{support}(XY)/\text{support}(X)$.

Firstly, the frequent rule items are discovered effectively. This is necessary to generate new candidate item sets, which is performed by mapping all the training data in each iteration from the frequent itemsets that have been discovered. After the discovery of all frequent itemsets, the Apriori algorithm assigns class labels and generates all the class association rules (CARs). Before assigning a class, however, the confidence value for each CAR must be computed. The rule item that meets the minimum value of Minimum confidence (min.conf.) becomes a rule. Min. conf. means constraint is applied to these frequent itemsets to form rules-sets. After this, the rules are ranking then ordered by the confidence value and support value. Pruning is also carried out to eliminate misleading and (repeated) rules. The technique has the benefit of accelerating the classification method and produces more accurate classifier (model) [43].

The final phase of this algorithm is the prediction phase. In this phase, attempts are made to predict the accuracy of the classifier through testing the known label of test data with the classified results generated by the classifier. The average precision, therefore, refers to the percentage of the test set that is rightly classified by the classifier. This manner of construction is known as supervised learning, whereby the training data are assigned by labels that refer to the group of the new data that has been classified based on the training dataset.

The design of AR used for sentiment classification is made up of a text classifier approach, which mines association rules. This classifier associates the terms of a document and its categories. The text documents are demonstrated as a collection of transactions whereby each document is noted as a transaction, and the terms in the document, that are made up of terms are noted as the items of the transaction. The system then discovers associations between the words in documents and the labels assigned to them. Each category is viewed as a separate text collection, and the AR mining is applied to each of the groups. The resultant rules from all the separate categories are then merged to form the classifier. Then, the training set is used for the evaluation of classification quality, classification of the text documents for tests, the number of rules covered, and the use of attributive probability. Furthermore, many

researchers have adopted sentiment analysis using association rules in their research works [44]-[48].

B. Lexicon-Based Approach

The semantic orientation (SO) approach is yet another technique employed in sentiment analysis (SA). This approach uses a learning technique that is unsupervised as initial training is not required. To a certain extent, this technique determines how far away or close a term is from being positive or negative.

Unsupervised learning uses lexical rules in sentiment classification tasks. [16] and a semi-supervised learning approach uses WordNet as the prime lexical resource [49]. A seed set extracted from the WordNet is used in this model. This concept stated that words with close orientation usually have similar gloss. They also presented a statistical-based approach to determine the semantic orientation of the seed terms with the help of gloss classification. They suggested a model that utilizes k-means clustering method [38]. The documents are categorized into two groups – positive and negative. After this, the term frequency-inverse document frequency (TF-IDF), which is a weighting feature method, is executed on the text, followed by a voting technique to obtain results that are more accurate. Some researcher prepared a sentimental lexicon [50]. They argued that the preparation of a manual sentiment lexicon is more effective compared to the preparation of an automatic sentiment lexicon that produces more precise lexicons.

In addition, the literature review and conclusion of previous research studies state that machine learning technique is efficient than the sentiment orientation technique. This was attributed to many features that determine the polarity of documents. In contrast, the sentiment orientation technique performs efficiently with various domains, as it is not dependent on other specific domain. In addition, the sentiment orientation technique has to face a difficult task while constructing a lexicon with the right polarity, just like the selection of correct features with the machine learning technique.

C. Combination Method

Besides using the individual machine learning approaches or Lexicon approach that have been mentioned earlier for sentiment classifying, there are very few research techniques that utilize both the machine learning and the lexicon-based approaches.

The improved Naïve Bayes and SVM algorithms as classifiers are also being proposed [8]. Unigrams and bigrams are used as a feature selection to close the gap of accuracy between positive and negative. The researchers concluded that the combination of machine-learning methods and dictionary-based methods could substantially improve sentiment classification.

D. Resources of Sentiment Analysis or Opinion Mining

Sentiment analysis (SA) always begins with the procedure of the collection of data, such as Amazon or from a social media website, such as Twitter, or by leveraging pre-existing resources, such as publicly available sentiment analysis datasets.

Sentiment analysis can be classified based on the input domains such as Blogs, Review Sites, News articles, or Social Networks.

1) *Blogs and Forums*: In addition, researchers have used Web forum postings [49], [51] and blogs, [14] as a source for their sentiment analysis research. Users who use forums or message boards have to be registered before they are allowed to registered users send documents for publications sites. Generally, forums are dedicated to only one subject; hence, the use of forums as a database ensures the operation of sentiment analysis in a single domain [51]. As well, bloggers document day after day activities in and around their areas, counties, or around the globe, conveying their reviews in blogs [37]. A high percentage of these blogs include testimonies various products, issues and event. Most research in the area of sentiment analysis has expressed the importance of blogs as a resource pool for expression of personal opinions [9], [27].

2) *Reviews*: In sentiment analysis, many studies focused on reviews because of their availability and richness with the sentiment. Movie and product reviews, in particular, were among the most studied [28], [14]. The purpose of the opinions (reviews) is to demonstrate the effectiveness of a certain item, so it is a matter of a single domain. Meanwhile, the sentiment analysis of reviews benefits both the organizations and prospective customers of the products. It enables companies to predict the sales of a product. In addition, the attributes preferred and resented by the reviewers can be discovered. Amazon product reviews (www.amazon.com) or professional review sites such as www.dpreview.com, www.imdb.com, and www.cent.com (product reviews) are excellent data sources for sentiment analysis researchers [28]. Review text is usually moderate in length (about 50 words), and people tend to use formal language in review writing.

3) *News Articles*: News articles, more especially financial news articles are a popular source of sentiment analysis [52]. News articles texts are usually structured and formal. One issue that arises from the text extraction in this domain is the use of graphics in news articles. Sometimes, the graphs and figures may include information that is not found in the text of the article. Thus, the use of the existing methods, such information will be ignored.

4) *Social Networks*: News articles, more especially financial news articles are a popular source of sentiment analysis [52]. News articles texts are usually structured and formal. One issue that arises from the text extraction in this domain is the use of graphics in news articles. Sometimes, the graphs and figures may include information that is not found in the text of the article. Thus, the use of the existing methods, such information will be ignored.

- **Twitter:**

Twitter is a well-known microblogging service or social networking. It allows its users, to post and read posts that are restricted to 140 characters, referred to as "tweets". The sentiment analysis on Twitter reveals an impending pattern in the estimation of poll results [20], [21].

- **Facebook:**

Facebook was introduced as a social networking platform in 2004. Since then, it has become a trending social networking service, allowing users to post their personal profiles, share videos, photos and other information. Others can view the profiles, videos, and images on the owner's friend list. As such, social media is the fastest and most preferred source of information on the Web linking users throughout the world. Therefore, people can now make an impact on one another effortlessly and conveniently. Due to the exceptional rise in the amount of information available, the demands for a computerized strategy that reacts to changes in sentiment and to increase tendencies is inevitable [53], [29].

III. RESULT AND DISCUSSION

Generally, in natural language processing (NLP), the classification or sentiment analysis (SA) is regarded as a specific case of text classification. Despite the number of classes in sentiment analysis being small, the process of sentiment classification is complex than the traditional topic text classification [14]. The classification in topic text classification depends on the use of keywords; however, it does not work efficiently in case of sentiment analysis. The nature of the problem defines other difficulties in sentiment analysis. The negative sentiment may sometimes be represented in a sentence without using any notable negative words. In addition, there is a fine line between whether a sentence should be considered subjective or objective. Identifying the opinion holder, the person who voices sentiments in the text is the most complex task in sentiment analysis. The sentiment analysis greatly relies on the subject or field of the data. The words may sometimes have a positive sentiment in a particular field, and the same words may have another polarity sentiment in a different field [14].

IV. CONCLUSIONS

To conclude, the use of an opinion mining or sentiment analysis to mine a numerous of unstructured data has become an essential study problem. Development of better products, services, and good business management are the products of sentiment analysis. The review paper presented the related work done until the time in Sentiment Analysis field. The research work in the articles indicates that the improvement of the Sentiment Classification algorithms or opinion mining is still an uncovered research area. As well, Naive Bayes (NB) and Support Vector Machines (SVM) are the common frequently used supervised Machine learning Approaches for Opinion mining or Sentiment classification. Amongst the surveyed approaches, the Analyzers of the sentiment are dependent language. No existing method found which it is more general and suitable to be language dependent. Nevertheless, the interest in languages other than English in the field of sentiment analysis or opinion mining is getting attending, as there is still the lack of researches and available resources in other languages. The sources of the social media such as microblogs, forums, news sources, and blogs present a tremendous amount of information about people's opinion and feelings on a particular issue or product. Nevertheless, by using these social media in micro-blogging

sites or networking sites as a source of data for OM or SA task is still needs significant deeper analysis and research. The main challenges involve the difficulty of classification, tackling language generalization and negations. As well, natural language processing tools have attracted researchers recently in a field of sentiment analysis and still needs more improvement, while certain algorithms that have been used in OM or SA gives good results. In addition, all algorithms that exist still no one complete method so far to solve all the challenges.

ACKNOWLEDGMENT

This research has been supported by the UKM university research Grant GUP-2015-003.

REFERENCES

- [1] M. M. Altawaier & S. Tiun "Comparison of Machine Learning Approaches on Arabic Twitter Sentiment Analysis". *International Journal on Advanced Science, Engineering and Information Technology*, Vol. 6 (2016) No. 6, pages: 1067-1073.
- [2] K. Dave, S. Lawrence and D. M. Pennock. "Mining the peanut gallery: Opinion extraction and semantic classification of product reviews". In *Proceedings of the 12th international conference on World Wide Web*, pp. 519-528.
- [3] A. Balahur, R. Mihalcea and A. Montoyo "Computational approaches to subjectivity and sentiment analysis": Present and envisaged methods and applications. *Computer Speech & Language*, 28(1): 1-6, 2014.
- [4] E. Boiy and M. F. Moens "A machine learning approach to sentiment analysis in multilingual Web texts". *Information retrieval*, 12(5): 526-558, 2009.
- [5] M. Ghiassi, J. Skinner and D. Zimbra. "Twitter brand sentiment analysis: A hybrid system using n-gram analysis and dynamic artificial neural network". *Expert Systems with applications*, 40(16): 6266-6282, 2013.
- [6] B. Liu, M. Hu and J. Cheng. "Opinion observer: analyzing and comparing opinions on the web". In *Proceedings of the 14th international conference on World Wide Web*, pp. 342-351.
- [7] J. Brooke, M. Tofiloski and M. Taboada "Cross-Linguistic Sentiment Analysis: From English to Spanish". In *RANLP*, pp. 50-54.
- [8] P. Arora, A. Bakliwal and V. Varma "Hindi subjective lexicon generation using WordNet graph traversal". *International Journal of Computational Linguistics and Applications*, 3(1): 25-39, 2012.
- [9] R. Feldman. "Techniques and applications for sentiment analysis". *Communications of the ACM*, 56(4): 82-89, 2013.
- [10] M. Taboada, J. Brooke, M. Tofiloski, K. Voll and M. Stede Lexicon-based methods for sentiment analysis. *Computational linguistics*, 37(2): 267-307, 2011.
- [11] R. Xia, C. Zong and S. Li. "Ensemble of feature sets and classification algorithms for sentiment classification". *Information Sciences*, 181(6): 1138-1152, 2011.
- [12] J. Kamps, M. Marx, R. J. Mokken and M. d. Rijke. "Using WordNet to measure semantic orientations of adjectives". 2004.
- [13] M. Bautin, L. Vijayarenu and S. Skiena "International Sentiment Analysis for News and Blogs". In *ICWSM*.
- [14] R. Prabowo and M. Thelwall. "Sentiment analysis: A combined approach". *Journal of Informetrics*, 3(2): 143-157, 2009.
- [15] J. Bollen, H. Mao and X. Zeng "Twitter mood predicts the stock market". *Journal of Computational Science*, 2(1): 1-8 2011.
- [16] T. Xu, Q. Peng and Y. Cheng. "Identifying the semantic orientation of terms using S-HAL for sentiment analysis". *Knowledge-Based Systems*, 35: 279-289, 2012.
- [17] T.T. Dang, N. T. X. Huong, A.C. Le and V.N. Huynh. "Automatically Learning Patterns in Subjectivity Classification for Vietnamese". *Knowledge and Systems Engineering*. Springer, pp. 629-640, 2015.
- [18] P. D. Turney. "Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews". In *Proceedings of the 40th annual meeting on association for computational linguistics*, pp. 417-424.
- [19] T. Wilson, P. Hoffmann, S. Somasundaran, J. Kessler, J. Wiebe, Y. Choi, C. Cardie, E. Riloff and S. Patwardhan. "Opinion Finder: A system for subjectivity analysis". In *Proceedings of hlt/emnlp on interactive demonstrations*, pp. 34-35.
- [20] E. Riloff, J. Wiebe and W. Phillips. "Exploiting subjectivity classification to improve information extraction". In *Proceedings of the National Conference On Artificial Intelligence*, pp. 1106.
- [21] S. Yu and S. Kak. "A survey of prediction using social media". *arXiv preprint arXiv:1203.1647*, 2012.
- [22] M. Abdul-Mageed, M. Diab, & M. Korayem, 2011. Subjectivity and Sentiment Analysis of Modern Standard Arabic. *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, hlm. 587-591.
- [23] S. Mukund, R. Srihari, & E. Peterson, 2010. An Information-Extraction System for Urdu---A Resource-Poor Language. *ACM Transactions on Asian Language Information Processing (TALIP)*, 9(4), 15.
- [24] G. Li and F. Liu. "A clustering-based approach on sentiment analysis". In *Intelligent Systems and Knowledge Engineering (ISKE)*, International Conference on, pp. 331-337, 2010.
- [25] L. Qu, G. Ifrim and G. Weikum. "The bag-of-opinions method for review rating prediction from sparse text patterns". In *Proceedings of the 23rd International Conference on Computational Linguistics*, pp. 913-921.
- [26] M. D. Molina-González, E. Martínez-Cámara, M-T Martín-Valdivia and J. M. Perea-Ortega. "Semantic orientation for polarity classification in Spanish reviews". *Expert Systems with Applications*, 40(18): 7250-7257, 2013.
- [27] P. Chaovalit and L. Zhou "Movie review mining : A comparison between supervised and unsupervised classification approaches". In *System Sciences. HICSS'05. Proceedings of the 38th Annual Hawaii International Conference on*, pp. 112c-112c, 2005.
- [28] Y. Chen and J. Xie "Online consumer review: Word-of-mouth as a new element of marketing communication mix". *Management Science*, 54(3): 477-491, 2008.
- [29] L. Pan. "Sentiment Analysis in Chinese". Brandeis University, 2012.
- [30] A. Aue and M. Gamon, "Customizing sentiment classifiers to new domains: A case study". In *Proceedings of recent advances in natural language processing (RANLP)*, pp. 2.1.
- [31] A. Esuli and F. Sebastiani. "Determining the semantic orientation of terms through gloss classification". In *Proceedings of the 14th ACM international conference on Information and knowledge management*, pp. 617-624.
- [32] V. Vapnik "The nature of statistical learning theory". 1995.
- [33] Z. Zhang, Q. Ye, Z. Zhang and Y. Li. "Sentiment classification of Internet restaurant reviews written in Cantonese". *Expert Systems with Applications*, 38(6): 7674-7682, 2011.
- [34] A. Agarwal, B. Xie, I. Vovsha, O. Rambow and R. Passonneau, Sentiment analysis of twitter data. In *Proceedings of the workshop on languages in social media*, pp. 30-38.
- [35] A. Go, R. Bhayani and L. Huang "Twitter sentiment classification using distant supervision". *CS224N Project Report*, Stanford, 1: 122009.
- [36] P. Melville, W. Gryc and R. D. Lawrence. "Sentiment analysis of blogs by combining lexical knowledge with text classification". In *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 1275-1284.
- [37] J. Liu and S. Seneff. "Review sentiment scoring via a parse-and-paraphrase paradigm". In *Proceedings of the Conference on Empirical Methods in Natural Language Processing: Volume 1-Volume 1*, pp. 161-169, 2009.
- [38] Y.-L. Chen, K. Tang, R.-J. Shen, & Y.-H. Hu., 2005. Market Basket Analysis in a Multiple Store Environment. *Decision support systems* 40(2): 339-354.
- [39] F. Thabtah, Q. Mahmood, L. McCluskey, and H. Abdel-Jaber. (2010a). "A new Classification based on Association Algorithm". *Journal of Information and Knowledge Management*, Vol 9, No. 1, pp. 55-64. World Scientific.

- [40] E.Baralis, S.Chiusano and P. Garza. (2008)." A Lazy Approach to Associative Classification". IEEE Trans. Knowledge Data Engineering. 20(2): 156-171.
- [41] R. Agrawal and R. Srikant, (1994) "Fast algorithms for mining association rules". In Proc. of the Int. Conf. on Very Large Databases, pages 487-499, San Diego, Chile,
- [42] D. Patil, V. Wadhai, and J. Gokhale, (2010)," Evaluation of Decision Tree Pruning Algorithms for Complexity and Classification Accuracy", International Journal of Computer Applications (0975 – 8887) Volume 11– No.2
- [43] Z. Hai, , K. Chang, & J.-J. Kim, 2011. Implicit Feature Identification via Co-Occurrence Association Rule Mining. Dlm. (pnyt.). Computational Linguistics and Intelligent Text Processing, hlm. 393-404. Springer.
- [44] W. Y. Kim, J. S.Ryu, , K. I. Kim, & U. M. Kim, 2009. A Method for Opinion Mining of Product Reviews Using Association Rules. Proceedings of the 2nd International Conference on Interaction Sciences: Information Technology, Culture and Human, hlm. 270-274.
- [45] Y. Man, O. Yuanxin, & S. Hao,. 2014. Investigating Association Rules for Sentiment Classification of Web Reviews. Journal of Intelligent & Fuzzy Systems: Applications in Engineering and Technology 27(4): 2055-2065.
- [46] Q. Su, X. Xu, , H. Guo, Z. Guo, X. Wu, X. Zhang , B. Swen, & Z. Su, 2008. Hidden Sentiment Association in Chinese Web Opinion Mining. Proceedings of the 17th international conference on World Wide Web, hlm. 959-968.
- [47] W. Wang, H. Xu, , & W. Wan, (2013). Implicit feature identification via hybrid association rule mining. Expert Systems with Applications, 40(9), 3518-3531.
- [48] B. Pang, L. Lee and S. Vaithyanathan. "Thumbs up?: sentiment classification using machine learning techniques". In Proceedings of the ACL-02 conference on Empirical methods in natural language processing-Volume 10, pp. 79-86.
- [49] Q. Ye, Z. Zhang and R. Law. "Sentiment classification of online reviews to travel destinations by supervised machine learning approaches". Expert Systems with Applications, 36(3): 6527-6535, 2009.
- [50] A. Abbasi, H. Chen and A. Salem Sentiment analysis in multiple languages: Feature selection for opinion classification in Web forums. ACM Transactions on Information Systems (TOIS), 26(3): 12, 2008.
- [51] A. B. Goldberg and X. Zhu," Seeing stars when there aren't many stars: graph-based semi-supervised learning for sentiment categorization". In Proceedings of the First Workshop on Graph Based Methods for Natural Language Processing, pp. 45-52.
- [52] M. Abdul-Mageed, M. T. Diab and M. Korayem Subjectivity and sentiment analysis of modern standard Arabic. In Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies: short papers-Volume 2, pp. 587-591.