



## A SURVEY ON SENTIMENT ANALYSIS IN TOURISM

Sarah Anis

Faculty of Computer and  
information sciences,  
Ain shams University  
Cairo, Egypt  
[Sarrahaniss@gmail.com](mailto:Sarrahaniss@gmail.com)

Sally Saad

Faculty of Computer and  
information sciences,  
Ain shams University  
Cairo, Egypt  
[Sallysaad@cis.asu.edu.eg](mailto:Sallysaad@cis.asu.edu.eg)

Mostafa Aref

Faculty of Computer and  
information sciences,  
Ain shams University  
Cairo, Egypt  
[mostafa.m.aref@gmail.com](mailto:mostafa.m.aref@gmail.com)

*Abstract: Tourism-related websites have turned into an incredible data source that impacts the tourism industry from many points of view. Tourists express their opinions regarding products and services online daily. The interest in understanding and analyzing customer opinions has increased significantly over the past few years as it is vital for the decision making of both customers and companies. Sentiment analysis is the practice of applying natural language processing, statistics and machine learning methods to extract and identify the common opinion behind the text in a review, blog discussion, news, comments or any other document. Sentiment analysis has great potential to directly understand tourists' opinions. This paper tackles a comprehensive overview of the latest update in this field. The main target of this survey is to give a nearly full image of sentiment analysis approaches, techniques, and challenges in analyzing the correct meaning of sentiments and detecting the suitable sentiment polarity in the field of tourism.*

*Keywords—Sentiment Analysis; Natural Language Processing; Machine Learning; Sentiment Classification; Tourism*

### I. INTRODUCTION

Nowadays, people generally prefer to communicate and socialize on the web. With the widespread usage of social media in our daily lives, social media websites became a vital and major source of data about user reviews in various fields. Sentiment classification is the task of classifying the data into some categories mainly to positive or negative opinions that have been expressed on a certain product, organization, or event [1]. This survey can be useful for newcomer researchers as it generally gives a refined summary to different sentiment analysis (SA) methods and techniques with brief details of their strengths and weaknesses in a single research paper. Moreover, the fundamental prospects lying ahead for sentiment analysis are identified and discussed. This work will help researchers in this field choose the right technique for a certain application and will likewise be useful for beginners aiming at sentiment analysis to have an all-encompassing perspective on the whole research field. The rest of this paper is organized as follows: section 2 includes a general overview of the sentiment analysis process, while section 3 tackles the sentiment detection techniques specifically and their related articles; section 4

discusses pieces of literature on various sentiment classification (SC) approaches and techniques. Section 5, highlights possible trends and challenges in sentiment analysis. Finally, the conclusion is presented in section 6.

## II. SENTIMENT ANALYSIS OVERVIEW

This paper focuses on the classification of sentiments of text reviews in tourism-related websites. Users lean towards internet booking because it is less tedious, less expensive, and they can get point by point information about facilities and location of the required destination. Studies about customers' online behavior uncovered that the decision of acquiring a product is very much influenced by other buyers' opinions [2]. Sentiment analysis enables us to gain the common opinion behind user reviews describing their hotel, restaurant or any tourist destination experience. The sentiment analysis is a complex process that involves five different steps [3] described in Fig. 1.

1. Data collection: gathers review data from user-generated content online and then store them in a database. These data are unstructured and disorganized making manual analysis almost impossible. In this way, sentiment analysis is used to classify.
2. Text preparation (Preprocessing): prepares data by eliminating non-textual and irrelevant content from text.
3. Sentiment detection: sentences are classified into subjective or objective categories. Subjective sentences include opinions, personal beliefs and views while objective expressions include facts, evidence, and measurable observations [4]. After this step, subjective sentences are held and sentences with objective expressions are discarded.
4. Sentiment classification: in this step, the polarity of a review is determined on which subjective sentences are classified into positive, negative or neutral.
5. Presentation of output: the main objective of sentiment analysis is to be further used in decision-making. Thus, the display of the classification output is very important. The results are mainly presented on graphs and charts.



Fig. 1. Sentiment analysis process

Customer reviews on social media often reflect joy, dissatisfaction, frustration, happiness, and different sentiments. Taking advantage of these huge volumes of subjective information is of great value to tourism associations and organizations which aim to increase profitability and enhance or maintain customer satisfaction. There are three types of sentiment polarity classification which are binary, ternary or ordinal [5]. In binary classification, the polarity of a given review is classified as positive or negative assuming that the text is subjective in the first place. The assumption made that reviews are subjective is not necessarily true, customer reviews provided through the text is considered either subjective or objective which means a ternary classification that requires the third category is needed. Nowadays, most of the sentiment analysis approaches apply sentiment detection first which differentiate between objective and subjective reviews using a binary classifier then determines the sentiment polarity of subjective reviews using a binary polarity classifier. In addition to the binary and ternary classifiers,

ordinal classification can be employed to automatically rate reviews based on an ordinal regression model by the means of a numbered ranking scale [6, 7]. Sentiment analysis can be performed at word, sentence, paragraph or document levels. In document-level, the whole document is assigned a single sentiment while in paragraph-level sentiment analysis, the sentiment of each paragraph is determined separately. Other approaches consider applying sentiment analysis on each sentence or even each word individually. Sentence-level classification is more challenging than paragraph-level and document-level classification since sentences have fewer words than paragraphs and documents.

### III. SENTIMENT DETECTION

Sentiment detection known as subjectivity detection could be viewed as a pre-processing step to sentiment classification that can prevent a sentiment classifier from considering the deceptive text in online reviews and eventually lead to better classification results. Documents containing only subjective text are ideal but many web reviews are a mixture between subjective and objective texts making the process of finding subjective reviews a challenging problem. Lexicon-based and machine-learning are the two main approaches for subjectivity detection [8]. In the lexicon-based approach, a list of subjective words is built and used as an opinion lexicon, and the frequency of lexicon words determine the subjectivity score of the document. On the other hand, the machine-learning or classification-based approach uses term occurrences or linguistic features for building the classifier. Chaturvedi et al. [9] proposed a framework for subjectivity detection that integrates convolutional neural networks (CNN) and an extended version of extreme learning machine (ELM) paradigm that uses Bayesian networks and fuzzy recurrent neural networks. The role of Bayesian networks was to detect dependencies in high dimensional data by building a network of connections between the hidden neurons of ELM then the overall structure generated by the Bayesian networks is inherited by a fuzzy recurrent neural network which models the temporal features in the predictor. They showed that their proposed framework can work well with standard subjectivity detection problems. Karimi et al. [8] used two main approaches for subjectivity detection. The first approach namely lexicon-based used a list of subjective words as an opinion lexicon to differentiate between subjective documents and objective ones based on the frequency of the lexicon words in documents to determine the overall subjectivity score of the document. The second approach was based on machine learning or classification methods. In this approach, the classifier is trained then used to identify the category of test documents. The proposed supervised method is called Language Model-based Supervised Opinion Retrieval (LMSOR) in which they used a collection of subjective and objective documents, and in the semi-supervised method that is called Language Model-based Semi-Supervised Opinion Retrieval (LMSSOR), a set of lexicons along with their polarity weights was utilized. In both methods, a measure was computed by comparing each document to determine its subjectivity degree then this measure was employed for documents ranking. Their experimental results showed that both of the proposed methods achieved high accuracy for both the English and Persian languages.

While most of the previous work had focused on document-level subjectivity analysis, Rilof et al. [10] had focused on classifying sentences as objective or subjective. They used the extraction pattern to represent subjective expressions then extracted features were employed to train machine learning classifiers such as Support Vector Machine (SVM) [11] and ELM [12]. Sentence-level subjective classification methods are mainly based on supervised machine learning. However, it is not always applicable to obtain labeled sentences since manual annotation is costly and time-consuming. On the other side, the use of fuzzy logic techniques has proved to be powerful in sentence-level classification as Opinions are fuzzy [13]. Rustamov et al. [14] presented two supervised machine learning approaches for sentence-level subjectivity detection in a movie review database, which are Fuzzy Control System (FCS), and Adaptive Neuro-Fuzzy Inference System (ANFIS). Their feature extraction algorithm was

based on statistical occurrences of words in the corpus. They showed their approach improved the accuracy of subjectivity detection and the fact that it didn't depend on any linguistic knowledge made it applicable to any language.

Other recent papers attempt to perform sentiment detection at the aspect level, creating what is known as Aspect-Based SA (ABSA) [15]. An aspect is an attribute or a feature of a certain service or product. In ABSA a relation between various aspects of an item and its polarity is established. There're two types of aspects, implicit and explicit. The main difference between implicit and explicit aspects is that an implicit aspect is not clearly stated in the text while an explicit aspect is directly stated and well classified which therefore makes extracting implicit aspects more complex. Ray et al. [16] proposed a deep learning approach for aspect extraction from text and analysis of the user's sentiment according to the aspect. They used a deep convolutional neural network (CNN) for tagging each aspect in the subject sentences and combined it with a set of rule-based approaches to improving the performance of aspect extraction method. Svitlana et al. in [17] suggested that incorporating various author demographics like age, gender, etc. can effectively enhance sentiment analysis. They conducted an empirical study on the differences in subjective language between male and female users on Twitter and their results showed a statistical improvement in terms of F-measure compared to the gender-independent baseline for subjectivity classification and thus enhancing the overall polarity sentiment classification.

#### IV. SENTIMENT CLASSIFICATION

Sentiment classification is the main task of sentiment analysis, which determines the overall polarity of opinion whether it's positive or negative. The key to sentiment classification task is selecting effective features [18]. In sentiment classification, feature selection is the process of extracting relevant features which should be sentiment words that indicate a certain polarity.

The approaches for automated sentiment classification can be generally divided into a machine learning approach, lexicon-based approach and hybrid approach [19]. The machine learning (ML) approach applies the famous ML algorithms and uses linguistic features for predicting the polarity of sentiments based on trained as well as testing data sets. The text classification methods using the ML approach can be mainly categorized into supervised and unsupervised learning methods. The supervised methods make use of a large number of labeled training data. The unsupervised methods are used when there is a limited or no availability of these labeled training data. Different machine learning classification approaches can be employed to the classification task, but the Support Vector Machines and Naïve Bayes are the most commonly used [20]. On the other side, the lexical approach starts with a predefined set of sentiment words and it does not need any prior training for data classification. It is divided into a dictionary-based approach and corpus-based approach which use statistical or semantic methods to find sentiment polarity. While the hybrid approach is the combination of both the machine learning and the lexicon-based approaches [21].

##### A. Machine Learning Approach

Machine learning is one of the most important approaches gaining the enthusiasm of researchers due to its flexibility and accuracy. Mostly the supervised learning consists of two stages: Training data and Classification. In the training stage, a set of labeled corpora is prepared and will be used by the algorithm for learning purpose. After completing the training phase, the classifier is deployed for analysis as a set of feature vectors from the trained data. A model is constructed based on the training data set which is employed over the new testing data for classification. Other techniques that also used to classify sentiments based on machine learning are semi-supervised and unsupervised learning methods.

Supervised sentiment classification approaches consist of two main types: Probabilistic Classification and Non-probabilistic Classification [22]. Some of the popular algorithms for probabilistic classification category in the field of sentiment analysis are NB, Bayesian network and maximum entropy. On the other hand, the most commonly used methods for non-probabilistic classification are neural networks, SVM, the nearest neighbor, decision tree and rule-based methods.

Polarity classification is a very complex process, because of the uncertainty Natural language and the informal text that is found in the online blogs and reviews. In addition, new topics, expressions, and trends made it more challenging to detect the sentiment polarity of a text. Recent work of sentiment analysis mainly focuses on a specific domain since it is hard to adapt SA dynamically, which improved the accuracy of the results compared to generalized systems [23].

Online travel-related information is very important for both travelers and travel agencies [24, 25]. Ye et al. [24] indicated that support vector machines and the character-based N-gram model achieved high accuracy for sentiment classification of reviews in the field of tourism compared to Naïve Bayes classifier. Moreover, they stated that the three methods reached accuracies of at least 80% when trained on large datasets.

Semi-supervised machine learning approaches work on both unlabeled data and some labeled data together. Semi-supervised learning has caught the attention of many researchers recently in the field of sentiment analysis for its high accuracy when labeled data are scarce. Nádia Félix et al. [26] proposed a semi-supervised framework to perform tweet sentiment classification in twitter. Moreover, they used an iterative self-training approach that enhanced tweet classification. A type of Semi-supervised machine learning approach is Graph-based methods. Many researchers tend to use Graph-based learning in the most recent decade as it has been affirmed to be effective in the field of sentiment analysis [27]. Jinpeng Wang et al. [28] utilized a graph-based semi-supervised method for classifying tweets into six distinct classes. Their experiments showed that using graph-based learning allowed them to achieve desirable results.

In unsupervised learning methods also known as the clustering technique, labeled data is not required instead it identifies commonalities in the unlabeled input dataset. Partition clustering and hierarchical clustering are the two main approaches of unsupervised learning. The main difference between them is the set of clusters whether it's nested or not. In the partitioning clustering approach, clusters simply do not overlap which means that each data object will be added to just one cluster. The main goal of partition clustering is to separate clusters so that data within each cluster are very similar to each other and very different from data in other clusters. The similarity measure is based on the Euclidean distance and the main drawback of this clustering technique is its sensitivity to noise. A widely used partitioning clustering method is K-means.

Conversely, Hierarchical clustering is a collection of overlapped clusters sorted out as a hierarchy. In this approach, clusters can have sub-clusters. Hierarchical clustering methods generally fall into two types based on their structure: Divisive methods (top-down) and agglomerative (Bottom-up). In the divisive method, all observed data start in one cluster initially, and data that are less similar to each other will split to individual clusters and move down the hierarchy. This process is done recursively until there's a single cluster for each data point. The agglomerative method is the opposite of divisive method in which each data is assigned to a separate cluster and iteratively data that are more similar to each other are merged in one cluster until there is only a single cluster or a specified number of clusters left [22]. In both methods, the merges and splits are achieved in a greedy manner.

### *B. Lexicon-Based Approach*

Lexicon-based Approach makes use of dictionaries only without the need for the training phase of data to train the classifier. There are many lexical databases (dictionaries) that can be used to perform sentiment analysis. WordNet [29] is a widely used lexical knowledge base. Lexical databases such as WordNet group words together with the relations between them such as synonymy, hypernymy, and meronymy [30]. Approaches based on such resources map each word with word senses in the lexical database and utilize the latter as concepts. However, lexical resources are generic which lack proper names, neologisms, slang, and terms that relates to a specific domain. BabelNet is a lexical database that has been developed as an improvement to WordNet. BabelNet combines WordNet with the largest multilingual online encyclopedia known as Wikipedia; this combination is performed via an automatic mapping between WordNet and Wikipedia in order to fill lexical gaps in other resources [31]. The usage of Wikipedia made a significant enhancement in WordNet because it is currently the biggest knowledge repository on the internet [32]. Late work in the area of automatic text analysis includes BabelNet as it supports many languages and with rich semantic relations.

In the Dictionary-based lexicon approach, the input text is changed over to tokens. Then, each token is matched with the lexicon in the dictionary. If there is a positive match to the token in the dictionary, the total score of text is incremented. The score is decremented or the word is tagged as negative otherwise. The classification of a text depends on the total score it accomplishes. On the other hand, Corpus-based approach takes advantage of the syntactic pattern of text that it focuses on the whole text rather than words to determine new sentiment words and their polarity by searching for the lexicons that match certain part of another sentence in a large corpus [33]. The problem with the corpus-based approach is that its performance differs in various domains, because a word could be interpreted as positive in one domain and gives a negative meaning in another domain [22], e.g. “his voice is low and gentle” is considered a positive opinion in music domain while “he was feeling low” should have a negative score in another domain [34]. It has been demonstrated that the dictionary-based technique is increasingly useful since it is hard to prepare a very large dataset can cover every single English word [46]. In General, the limitation of lexical-analysis is that its performance drastically decreases with the huge growth of the number of words in the dictionary [22].

### *C. Hybrid Approach*

Various researchers have investigated the utilization of hybrid approaches by combining different techniques with the point of accomplishing better outcomes than a standard methodology dependent on just a single tool. Maurel et al. [35] on the use of multiple techniques and tools in SA, showed how different resources and methodologies can contribute extra features to each other and exploit the advantages of each method. The approach that they have used for the classification of subjective texts was a hybrid between a symbolic rule-based technology and a statistical technology based on machine learning.

Another hybrid method for sentiment classification was proposed by Keith et al. [36] based on the combination between an unsupervised classifier and a supervised classifier which are scoring algorithm and support vector machines respectively. This proposed method takes advantage of the information provided by the scoring algorithm and its associated components and the flexibility of SVM. However, their hybrid method had a higher computational cost compared to using each method individually since it required the usage of the scoring algorithm and training the SVM classifier. But they claimed that this drawback has no significant effect as long as the dataset used in the application is considerably small.

The main goal of using lexicon-based or machine learning methods in a hybrid manner is for when one classifier fails, the following one takes place to classify and so on until the whole document is

classified. Dhaoui et al. [37] empirically compared the performance of lexicon-based approach, machine learning approach, and the hybrid approach. Results of the two approaches independently were relatively similar in terms of accuracy. However, the hybrid approach demonstrates significantly improved performance in positive polarity classification.

Based on our study, supervised machine learning techniques have shown preferable performance over unsupervised lexicon-based methods for sentiment classification. However, supervised machine learning techniques require large labeled training data which is relatively expensive and not necessarily available, whereas obtaining unlabeled data is easy. Thus, unsupervised strategies are more useful in case the application domain lack labeled training data. Besides, the hybrid approach can attain the benefits of both approaches, the high accuracy of machine learning algorithms and the stability of the lexicon-based approach, but it requires higher computational complexity and can be more sensitive to noisy reviews.

A brief comparison of the most well-known methods and techniques used for machine learning and lexicon-based approaches is shown in Tables 1 and 2 which includes the advantages and disadvantages of each method.

**Table 1: Comparison between different machine learning Classification Methods.**

<b>Method</b>	<b>Type</b>	<b>Advantages</b>	<b>Disadvantages</b>
Naive Bayesian	Probabilistic - Supervised machine learning algorithm	<ul style="list-style-type: none"> <li>• Simple.</li> <li>• Requires a small amount of training data to estimate the parameters.</li> <li>• Obtains good results in subjective classification.</li> </ul>	<ul style="list-style-type: none"> <li>• Class conditional independence assumption is violated by real-world data which may cause loss of accuracy.</li> </ul>
Maximum entropy	Probabilistic - Supervised machine learning algorithm	<ul style="list-style-type: none"> <li>• Incorporates various sources of information in a unified model.</li> <li>• Model constructed can satisfy all empirical constraints.</li> </ul>	<ul style="list-style-type: none"> <li>• Time consuming when dataset is large.</li> <li>• Over-fitting problem</li> <li>• High computational complexity of model parameter estimation procedure.</li> </ul>
Decision tree	Non-Probabilistic - Supervised machine learning	<ul style="list-style-type: none"> <li>• Easy to understand and interpret.</li> <li>• Inexpensive to construct.</li> <li>• Good performance with large datasets.</li> </ul>	<ul style="list-style-type: none"> <li>• May suffer from over-fitting.</li> <li>• Once a mistake is made at higher level, any sub tree is wrong.</li> </ul>

Support vector machine	Non-Probabilistic - Supervised machine learning algorithm	<ul style="list-style-type: none"> <li>• High accuracy</li> <li>• High dimensional input space.</li> <li>• Work well even with complex non-linear data points.</li> </ul>	<ul style="list-style-type: none"> <li>• Kernel selection.</li> <li>• Parameter tuning is critical to get good results.</li> </ul>
K-nearest neighbour	Non-Probabilistic - Supervised machine learning algorithm	<ul style="list-style-type: none"> <li>• Simple</li> <li>• Powerful</li> <li>• Fast training time</li> </ul>	<ul style="list-style-type: none"> <li>• Relatively slow in case of large training set.</li> <li>• Sensitive to noise.</li> </ul>
Self-training	Semi-Supervised machine learning algorithm	<ul style="list-style-type: none"> <li>• Simple.</li> <li>• Doesn't depend on the classification model.</li> </ul>	<ul style="list-style-type: none"> <li>• Poor performance of Traditional self-training.</li> <li>• Errors can re-enforce themselves.</li> </ul>
Co-training	Semi-Supervised machine learning algorithm	<ul style="list-style-type: none"> <li>• High accuracy.</li> <li>• Works well when conditional independence holds.</li> </ul>	<ul style="list-style-type: none"> <li>• In order to achieve high performance many features should be available.</li> </ul>
Graph-based	Semi-Supervised machine learning algorithm	<ul style="list-style-type: none"> <li>• Good performance with simple and complex domains. <ul style="list-style-type: none"> <li>• Can handle binary and multi-class classification.</li> </ul> </li> </ul>	<ul style="list-style-type: none"> <li>• Performance is vulnerable to graph structure.</li> <li>• Sensitive to noise.</li> </ul>
K-means	Partitioning Clustering - unsupervised machine learning algorithm	<ul style="list-style-type: none"> <li>• High speed performance</li> <li>• Efficient when dataset is large.</li> <li>• Easy to implement.</li> </ul>	<ul style="list-style-type: none"> <li>• Selection of optimal number of clusters is high.</li> <li>• It can converge to local optimum</li> <li>• Sensitive to noise.</li> <li>• Random selection of centroids.</li> </ul>



Fuzzy c-means	Partitioning Clustering - unsupervised machine learning algorithm	<ul style="list-style-type: none"> <li>• always converges</li> </ul>	<ul style="list-style-type: none"> <li>• High computational time</li> <li>• Sensitive to noise</li> <li>• Sensitive to initial selections (speed, local minima)</li> </ul>
Agglomerative algorithm	Hierarchical Clustering- unsupervised machine learning algorithm	<ul style="list-style-type: none"> <li>• Doesn't depend on the number of clusters and center of gravity.</li> <li>• Robust against noise.</li> </ul>	<ul style="list-style-type: none"> <li>• High time complexity in large datasets</li> </ul>
Divisive algorithm	Hierarchical Clustering- unsupervised machine learning algorithm	<ul style="list-style-type: none"> <li>• Robust</li> <li>• Doesn't depend on the number of clusters and center of gravity.</li> </ul>	<ul style="list-style-type: none"> <li>• Inefficient use of memory.</li> <li>• Nonlinear time complexity.</li> </ul>

*Table 2: Comparison between different lexicon-based Classification Methods.*

Method	Advantages	Disadvantages
Dictionary-based	<ul style="list-style-type: none"> <li>• Does not require any labeled training samples</li> <li>• Efficient</li> <li>• Easy access to words lexicon</li> </ul>	<ul style="list-style-type: none"> <li>• Relatively low accuracy when dealing with different domains.</li> </ul>
Corpus-based	<ul style="list-style-type: none"> <li>• Good for capturing semantic relations between single words</li> <li>• Efficient</li> <li>• Provides better results when domains are different.</li> </ul>	<ul style="list-style-type: none"> <li>• Can't be used alone</li> <li>• Can't handle logical relations between words.</li> </ul>

## V. CHALLENGES IN SENTIMENT ANALYSIS

There are several challenges in sentiment analysis. In this section, we discuss some of the difficulties that still remain and have not been explored adequately. Additionally, we talk about recommendations to deal with these difficulties. The goal of this section is to encourage researchers and specialists more to work particularly in the areas that are relatively less investigated.

Lexicons can be annotated manually or by automatic means according to their polarity. Building resources is not a sentiment analysis task, but it could help enhance it. The main challenges that face the

work in building resources are words ambiguity, multilingual words, text granularity [5, 38] and the differences in the meaning of opinions among multiple domains. Moreover, the manual construction of sentiment dictionaries needs broad head work with relatively restricted coverage on different forms of informal words [39]. At word-level sentiment analysis, most work focuses on determining the semantic orientation (SO) of words, in which a prior polarity is assigned to words in the lexicon out of context. However, the polarity of some words may differ emphatically with context, making the task of assigning each word to a specific sentiment category in the lexicon very hard. Words like big, small, strong, and weak are examples of Polarity-ambiguous words [40]. In the work of ambiguous words disambiguation, a mutual bootstrapping algorithm is adopted in [41] to automatically build the aspect and polarity-ambiguous lexicon. Firstly, they start with a small set of words which are ambiguous to extract some features then the sentiment orientation of the features is determined using the relations between features and polarity-ambiguous words in the labeled text. Secondly, they keep adding more polarity-ambiguous words. Other several iterations, the polarity-ambiguous word lexicon is built. Finally, the sentiment of polarity-ambiguous words in context can be decided due to the prior polarity retrieved of ambiguous words. Experiments showed that their method had effective results in sentence-level analysis. Xia et al. [42] focused on resolving the ambiguity of words using opinion-level features that includes opinion target, modifying word and indicative words. They used the Bayesian model to calculate the probability that a given sentiment word has a specific sentiment polarity, and they made an independent assumption amongst the features that influence the model. They claim that their model performed well in two different domains.

Text reviews may also contain multilingual words or words of different domains as various users write reviews in their own languages even on English forums thus making it hard to apply opinion mining techniques for efficient sentiment analysis [43]. On the other hand, many challenges are facing the machine learning technique of sentiment classification [22]. It could be in designing the classifier, availability of training dataset, or the correct interpretation of an unexpected expression. In the field of tourism, several reviews problems make them hard to evaluate [44]. Those challenges include that reviews are not clear and brief, also Scalar reviews are not useful when hotels offer different services. Furthermore, some features are more vital than others so the general rating is not objective however more impacted on those viewpoints. Bucur et al. [44] claimed that an aspect-oriented analysis would improve the performance of sentiment analysis from reviews in the domain of tourism. They also mentioned that a future heading for enhancing the performance could be the utilization of ontology, oriented to tourism domain.

Moreover, online reviews often contain slang or irony expressions. An irony expression normally signifies the opposite meaning of the word, typically used for humor or sarcasm. Reyes et al. [45] discussed the identification of irony reviews. They aimed to extract a set of discriminating features that represent irony. A model to represent irony reviews was proposed in terms of some features like n-grams, POS-grams, and funny profiling. They constructed their own dataset with ironic reviews collected from Amazon. They used three classifiers in classification, namely NB, SVM, and Decision Tree. Their evaluation showed relevant results of the model with the three classifiers for retrieving representative content.

Negation expressions are also considered one of the important challenges in sentiment analysis. Negation expressions will generally make positive words negative and make negative words more positive [46]. Several aspects in the area of negation handling are still not explored like how people use negations differently according to their community and cultures [47].

People generally write text on the web in a very expressive way that often contains many emotions and new techniques such as repeating characters to stress on a specific word. Thus, making content generated by users utterly different from standard text since it contains a lot of misspellings, abbreviations, slang words, and grammatical errors which downgrades the performance of automatic analysis because tools are primarily trained on the standard text [48]. Van Hee et al. [49] investigated a complex normalization system as a prior step to the classification task to enhance the performance of a sentiment classifier by automatically translating irrelevant text into standard text. Results showed the relevance of their normalization system for improving the overall process of sentiment analysis. Schouten et al. [50] claimed that the evolution of constructing a knowledge base that includes stereotypes and user-generated words that is rich in ironic and sarcastic words along with negation expressions when combined with machine learning approaches would offer better results for handling user-generated content.

Thwarting has been identified as a common challenge in different forms of texts in sentiment analysis [51, 52]. Thwarted expressions fundamentally refer to the phenomenon wherein the writer of the text first develops certain expectations for the review content, just to deliver a planned contrast to the earlier discussion [53]. In thwarting expressions only some part of text determines the overall polarity of the sentence or document e.g. “The actors are popular. The supporting cast is talented. It sounds like a great plot, but somehow this movie disappoints me” while the overall polarity is negative simple sentiment analysis approaches would classify this example as positive. A thwarting detection approach was proposed by Ramteke et al. [54] that used a rule-based method and a machine learning approach namely SVM for thwarting detection. Their results showed that the ML approach outperformed the rule-based approach.

Discourse Structure analysis is critical for Sentiment Analysis/Opinion Mining. It is one of the challenges that can majorly affect the overall opinion behind a text. In recent work, the importance of discourse relations in the field of sentiment analysis has been widely recognized. Discourse structure is how an entire text is organized and how words are ordered. In conventional lexicon-based methods, all words and sentences are dealt with similarly, and the structural aspects of a text are discarded. However, to correctly detect the polarity of text, Discourse structure and words order in the text should be taken into consideration. For instance, “A is better than B”, is the exact inverse opinion from, “B is better than A”. In this work, Wang et al. [55] utilized discourse relations by using explicit connectives for Chinese sentiment classification. They accomplished preferable outcomes over the state-of-the-art method.

Another emerging direction in sentiment analysis is the development of a machine learning algorithm that doesn't depend on a specific domain on which it was trained but also can perform well on other domains. The problem is that a word or sentence can have different meanings in different domains which mean it could be interpreted as positive sentiment in a domain and negative in another. Andreevskaja et al. [56] proposed a novel approach that combined a corpus-based classifier trained on a set of labeled in-domain data with a lexicon-based system trained on WordNet, to overcome the problem of system portability across different domains. They expect that the ensemble of more classifiers would take advantage of more approaches which would be able to accomplish better accuracy on in-domain, just as on other domains.

## VI. CONCLUSION

In this paper, we have discussed the process of sentiment analysis and its different approaches. The aim of this study is to provide an adequate survey by investigating the popular and most recent methods along with their challenges. A wide range of strategies have been developed and tested but still, a lot of work is yet to be done. Some algorithms that have been utilized in the field of sentiment analysis

achieved great outcomes, yet at the same time no method can overcome all the presented challenges. The commonly used approach for sentiment classification is supervised machine learning, it is well-known for its high accuracy, but they require a large amount of annotated training data which is costly in terms of human effort and time. Studies considered using semi-supervised learning methods for many sentiment analysis applications as they provide a good alternative to supervised learning when there's a small amount of labeled data available. One of the improvements made for the lexicon-based methods is expanding the dictionary to a larger one to make it able to deal with more words [57].

Recent work aims to improve the accuracy of sentiment classification by using a hybrid approach that combines lexicon-based approaches along with an ensemble of machine learning methods which exploit the advantages of multiple classifiers, although hybrid methods have higher computational complexities [20]. ML algorithms are commonly used for their simplicity and domain adaptability since it has the ability to learn from the training data. On the other hand, Lexicon-based algorithms are also frequently used to tackle general SA issues because they are scalable and computationally efficient [20, 39, 58]. Moreover, the choice of the subjectivity detection algorithm could effectively improve the overall performance of sentiment classification. Studies demonstrated that still, the English language is the most commonly used language in applications of sentiment analysis [59] due to the availability of its resources, which opens a new challenge to recent researchers specially in opinion mining of text reviews from social media to construct resources for other non-English languages including Latin, German, Far East, and Middle East Languages. However, we've concluded that the solutions for sentiment analysis still have far to go before achieving the certainty level demanded by practical applications.

#### REFERENCES

- [1] Md Shoeb, Jawed Ahmed, "Sentiment Analysis and Classification of Tweets Using Data Mining", *International Research Journal of Engineering and Technology (IRJET)*, vol. 04, no. 12, 2017.
- [2] F. Krebs, B. Lubascher, T. Moers, P. Schaap and G. Spanakis, "Social Emotion Mining Techniques for Facebook Posts Reaction Prediction", *arXiv.org*, 2017. [Online]. Available: <https://arxiv.org/abs/1712.03249>.
- [3] A. D'Andrea, F. Ferri, P. Grifoni and T. Guzzo, "Approaches, Tools and Applications for Sentiment Analysis Implementation", *International Journal of Computer Applications*, vol. 125, no. 3, pp. 26-33, 2015. Available: 10.5120/ijca2015905866.
- [4] R. Feldman, "Techniques and Applications for Sentiment Analysis", *Communications of the ACM*, 2013. [online]. Available: <https://cacm.acm.org/magazines/2013/4/162501-techniques-and-applications-for-sentiment-analysis/fulltext>.
- [5] A. Alaei, S. Becken and B. Stantic, "Sentiment Analysis in Tourism: Capitalizing on Big Data", *Journal of Travel Research*, vol. 58, no. 2, pp. 175-191, 2017. Available: 10.1177/0047287517747753.
- [6] J. Broß, *Aspect-Oriented Sentiment Analysis of Customer Reviews Using Distant Supervision Techniques*. Berlin: Freie Universität Berlin, 2013.
- [7] Xu B., Zhao TJ., Wu JW., Zhu CH., "Automatically ranking reviews based on the ordinal regression model", In: Deng H., Miao D., Lei J., Wang F.L. (eds) *Artificial Intelligence and Computational Intelligence. AICI 2011. Lecture Notes in Computer Science*, vol 7004. Springer, Berlin, Heidelberg, 2011. [Online]. Available: <https://dl.acm.org/citation.cfm?id=2045938>.
- [8] S. Karimi and A. Shakery, "A language-model-based approach for subjectivity detection", *Journal of Information Science*, vol. 43, no. 3, pp. 356-377, 2016. Available: 10.1177/0165551516641818.

- [9] I. Chaturvedi, E. Ragusa, P. Gastaldo, R. Zunino and E. Cambria, "Bayesian network based extreme learning machine for subjectivity detection", *Journal of the Franklin Institute*, vol. 355, no. 4, pp. 1780-1797, 2018. Available: 10.1016/j.jfranklin.2017.06.007.
- [10] E. Rilof and J. Wieb, "Learning Extraction Patterns for Subjective Expressions," In *Proceedings of the 2003 Conference on Empirical Methods in Natural Language Processing (EMNLP-03)*, pp. 105–112, 2003.
- [11] A. Tripathy, A. Agrawal and S. Rath, "Classification of Sentimental Reviews Using Machine Learning Techniques", *Procedia Computer Science*, vol. 57, pp. 821-829, 2015. Available: 10.1016/j.procs.2015.07.523.
- [12] L. Oneto, F. Bisio, E. Cambria and D. Anguita, "Statistical Learning Theory and ELM for Big Social Data Analysis", *IEEE Computational Intelligence Magazine*, vol. 11, no. 3, pp. 45-55, 2016. Available: 10.1109/mci.2016.2572540.
- [13] M. AL-Maimani, N. Salim and A. M. Al-Naamany, "Semantic and Fuzzy aspects of opinion mining", *Journal of Theoretical and Applied Information Technology*, vol. 63, no. 2, 2014. Available: <http://www.jatit.org>.
- [14] S. Rustamov, M. A. Clements, "Sentence-Level Subjectivity Detection Using Neuro-Fuzzy Models", In *Proceedings of the 4th Workshop on Computational Approaches to Subjectivity Sentiment and Social Media Analysis*, pp. 108–114, 2013.
- [15] M. Al-Ayyoub, A. Khamaiseh, Y. Jararweh and M. Al-Kabi, "A comprehensive survey of arabic sentiment analysis", *Information Processing & Management*, vol. 56, no. 2, pp. 320-342, 2019. Available: 10.1016/j.ipm.2018.07.006.
- [16] P. Ray and A. Chakrabarti, "A Mixed approach of Deep Learning method and Rule-Based method to improve Aspect Level Sentiment Analysis", *Applied Computing and Informatics*, in press.
- [17] Svitlana Volkova, Theresa Wilson, and David Yarowsky, "Exploring demographic language variations to improve multilingual sentiment analysis in social media." In *Proceedings of the 2013 Conference on Empirical Methods on Natural Language Processing*, pages 1815–1827, 2013.
- [18] P. Nakov et al., "Developing a successful SemEval task in sentiment analysis of Twitter and other social media texts", *Language Resources and Evaluation*, vol. 50, no. 1, pp. 35-65, 2016. Available: 10.1007/s10579-015-9328-1.
- [19] Diana Maynard, Adam Funk, "Automatic detection of political opinions in tweets", In: *Proceedings of the 8th international conference on the semantic web, ESWC'11*; p. 88–99, 2011.
- [20] W. Medhat, A. Hassan and H. Korashy, "Sentiment analysis algorithms and applications: A survey", *Ain Shams Engineering Journal*, vol. 5, no. 4, pp. 1093-1113, 2014. Available: 10.1016/j.asej.2014.04.011.
- [21] J. Wehrmann, W. E. Becker and Rodrigo C. Barros. "A Multi-Task Neural Network for Multilingual Sentiment Classification and Language Detection on Twitter", *The 33rd ACM/SIGAPP Symposium On Applied Computing*, DOI:10.1145/3167132.3167325, France, 2018.
- [22] F. Hemmatian and M. Karim Sohrab, "A survey on classification techniques for opinion mining and sentiment analysis", *Springer Netherlands*, 2017.
- [23] H.G. Yoon, H. Kim, C.O. Kim and M. Song, "Opinion polarity detection in Twitter data combining shrinkage regression and topic modeling", *Journal of Informetrics*, vol.10, no. 2, pp. 634-644. 2016.
- [24] Q. Ye, Z. Zhang and R. Law, "Sentiment classification of online reviews to travel destinations by supervised machine learning approaches," *Expert systems with applications*, vol.36, no. 3, pp. 6527–6535. 2009.
- [25] Xiomarah M., Edward Rolando, Jordan P. E., Ruben G, and Vicente G, "A proposal for sentiment analysis on Twitter for tourism-based applications", *The 17th international conference on intelligent software methodologies, tools and techniques*, Spain, 2018.

- [26] Nádia Félix Felipeda Silva, Luiz F.S.Coletta, Eduardo R.Hruschka, Estevam R.Hruschka Jr., "Using unsupervised information to improve semi-supervised tweet sentiment classification", *Information Sciences*, Volumes 355–356, Pages 348-365, 2016.
- [27] F. Xing, E. Cambria and R. Welsch, "Natural language based financial forecasting: a survey", *Artificial Intelligence Review*, vol. 50, no. 1, pp. 49-73, 2017. Available: 10.1007/s10462-017-9588-9.
- [28] Jinpeng Wang, Gao Cong, Wayne Xin Zhao, Xiaoming Li," Mining user intents in twitter: a semi-supervised approach to inferring intent categories for tweets ", In *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence*, 2015.
- [29] G. Miller, "WordNet: a lexical database for English", *Communications of the ACM*, vol. 38, no. 11, pp. 39-41, 1995. Available: 10.1145/219717.219748.
- [30] E. Gabrilovich and S. Markovitch, "Wikipedia-based Semantic Interpretation for Natural Language Processing", *Journal of Artificial Intelligence Research*, vol. 34, pp. 443-498, 2009. Available: 10.1613/jair.2669.
- [31] R. Navigli and S. Ponzetto, "BabelNet: The automatic construction, evaluation and application of a wide-coverage multilingual semantic network", *Artificial Intelligence*, vol. 193, pp. 217-250, 2012. Available: 10.1016/j.artint.2012.07.001.
- [32] Gabrilovich, E., S. Markovitch, "Computing semantic relatedness using Wikipedia-based explicit semantic analysis", *Proceedings of IJCAI*, 1606-1611, 2007.
- [33] E. Pamungkas, R. Sarno and A. Munif, "B-BabelNet: Business-Specific Lexical Database for Improving Semantic Analysis of Business Process Models", *TELKOMNIKA (Telecommunication Computing Electronics and Control)*, vol. 15, no. 1, p. 407, 2017. Available: 10.12928/telkomnika.v15i1.3176.
- [34] Neha S. Joshi, Suhasini A. Itkat, "A Survey on Feature Level Sentiment Analysis", (*IJCSIT International Journal of Computer Science and Information Technologies*, Vol. 5 (4), 5422-5425, 2014.
- [35] S. Maurel, P. Curtoni, L. Dini, "A hybrid method for sentiment analysis", *INFORSID*, 2008.
- [36] B. Keith, E. Fuentes, and C. Meneses, "A Hybrid Approach for Sentiment Analysis Applied to Paper" In *Proceedings of ACM SIGKDD Conference*, Halifax, Nova Scotia, Canada, 2017.
- [37] C. Dhaoui, C. Webster and L. Tan, "Social media sentiment analysis: lexicon versus machine learning", *Journal of Consumer Marketing*, vol. 34, no. 6, pp. 480-488, 2017. Available: 10.1108/jcm-03-2017-2141.
- [38] G. Vinodhini and R. Chandrasekaran, "Sentiment analysis and opinion mining: A survey," *International Journal*, vol. 2, no. 6, 2012.
- [39] Gwanghoon Yoo, Jeesun Nam,"A Hybrid Approach to Sentiment Analysis Enhanced by Sentiment Lexicons and Polarity Shifting Devices", *The 13th Workshop on Asian Language Resources*, Kiyooki Shirai, Miyazaki, Japan. pp.21-28, 2018.
- [40] Y. Wu and P. Jin, "SemEval-2010 task 18: disambiguating sentiment ambiguous adjectives", *Language Resources and Evaluation*, vol. 47, no. 3, pp. 743-755, 2012. Available: 10.1007/s10579-012-9206-z.
- [41] Y. Cao, P. Zhang, A. Xiong, "Sentiment Analysis Based on Expanded Aspect and Polarity-Ambiguous Word Lexicon", *International Journal of Advanced Computer Science and Applications*, vol. 6, no. 2, 2015. Available: 10.14569/ijacsa.2015.060215.
- [42] Y. Xia, E. Cambria, A. Hussain and H. Zhao, "Word Polarity Disambiguation Using Bayesian Model and Opinion-Level Features", *Cognitive Computation*, vol. 7, no. 3, pp. 369-380, 2014. Available: 10.1007/s12559-014-9298-4.
- [43] Khan, K., Baharudin, B. B., Khan, A.," Mining opinion from text documents: A survey", In *Digital Ecosystems and Technologies, DEST'09. 3rd IEEE International Conference on* (pp. 217-222), 2009.

- [44] C. Bucur, "Using Opinion Mining Techniques in Tourism", *Procedia Economics and Finance*, vol. 23, pp. 1666-167, 2015. Available: 10.1016/s2212-5671(15)00471-2
- [45] Reyes Pérez, A.; Rosso, P., "Making objective decisions from subjective data: Detecting irony in customers reviews", *Decision Support Systems*. 53(4):754-760. doi:10.1016/j.dss.2012.05.027, 2012.
- [46] Zhu, X., Guo, H., Mohammad S., Kiritchenko S., "An empirical study on the effect of negation words on sentiment", In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 304–313, Baltimore, Maryland, 2014.
- [47] S. M. Mohammad, "Challenges in sentiment analysis," in *A Practical Guide to Sentiment Analysis*, D. Das, E. Cambria, and S. Bandyopadhyay, Eds. Springer, 2016.
- [48] Orphee De Clercq, "The many aspects of fine-grained sentiment analysis An overview of the task and its main challenges", *LT3, Language and Translation Technology Team Ghent University Ghent, Belgium*, 2016.
- [49] C. Van Hee, M. Van de Kauter, O. De Clercq, E. Lefever, B. Desmet, V. Hoste, "Noise or Music? Investigating the Usefulness of Normalisation for Robust Sentiment Analysis on Social Media Data," *Expert Systems with Applications*, 2017.
- [50] K. Schouten, F. Frasinca, "Survey on aspect-level sentiment analysis", *IEEE Transactions on Knowledge and Data Engineering*, vol. 28, no. 3, pp. 813–830, 2016.
- [51] Pang, B., Lee L., Vaithyanathan S., "Thumbs up? Sentiment Classification using Machine Learning Techniques", In *Proceedings of EMNLP*, pages 79-86), 2002.
- [52] Ohana, B. and Tierney, B., "Sentiment classification of reviews using SentiWordNet", In *9th. IT & T Conference*, page 13, 2009.
- [53] B. Pang and L. Lee, "Opinion Mining and Sentiment Analysis", *Foundations and Trends® in Information Retrieval*, vol. 2, no. 12, pp. 1-135, 2008. Available: 10.1561/15000000011.
- [54] Ramteke, Ankit, Pushpak Bhattacharyya, J. Saketha Nath. "Detecting Turnarounds in sentiment analysis: Thwarting", In *Conference for association of computational linguistics*. 2013.
- [55] Wang F, Wu Y, Qiu L., "Exploiting discourse relations for sentiment analysis", In *Proceedings of the 24th International Conference on Computational Linguistics (COLING 2012)*, pp. 1311–20. Stroudsburg, PA: Assoc. Comput. Linguist. 2012.
- [56] Andreevskaia, A., & Bergler, S., "When specialists and generalists work together: Overcoming domain dependence in sentiment tagging." In *Proceedings of ACL-08: HLT*, Vol. Columbus, OH, pp. 290-298, 2008.
- [57] Bing Liu, "Sentiment Analysis and Subjectivity", *Handbook of Natural Language Processing*, Second Edition, (editors: N. Indurkha and F. J. Damerau), 2010.
- [58] Lu, B. and Tsou, B.K, "Combining a large sentiment lexicon and machine learning for subjectivity classification", In *Proceedings of the 9th International Conference on Machine Learning and Cybernetics*, pages 3311–3316, 2010.
- [59] K. Dashtipour, S. Poria, A. Hussain, E. Cambria, A. Y. A. Hawalah, A. Gelbukh, and Q. Zhou, "Multilingual sentiment analysis: State of the art and independent comparison of techniques", *Cognitive Computation*, 2016.