International Journal on Emerging Technologies **11**(3): 267-274(2020)

ISSN No. (Print): 0975-8364
ISSN No. (Online): 2249-3255

A Framework for Sentiment Analysis of Online News Articles

Jeelani Ahmed and Muqeem Ahmed
Department of Computer Science and Information Technology,
Maulana Azad National Urdu University, Hyderabad (Telangana), India.

(Corresponding author: Jeelani Ahmed)
(Received 31 January 2020, Revised 02 April 2020, Accepted 06 April 2020)
(Published by Research Trend, Website: www.researchtrend.net)

ABSTRACT: The current innovation era has changed the conventional way of life in many areas. Information Technology has been overflowed with an enormous quantity of data, which is being distributed each moment of consistently, by a large number of users, in the form of online journals, feedbacks, reviews, news blogs, comments, micro-blogging websites, social media and some more. Due to this absolute amount of assessment of precious web assets, a significant part of the research is concentrating on the sentiment analysis. Opinion analysis is the act of implementing text analysis and natural language processing techniques to recognize and extricate abstract data from content. News analysis can be utilized to plot the company's conduct after some time and in this manner, yield significant vital bits of knowledge about companies. Sentiment analysis is additionally helpful in web-based life checking to consequently describe the general inclination or state of mind of purchasers as reflected in web-based life toward a particular brand or organization and decide if they see positively or negatively. Today, it is an everyday exercise for some users to read the news on the internet. Persons' viewpoints, in general, will encounter a variation as per the news they go through. News reports occasions that involve feelings – good, bad, neutral. The human feelings available in the textual information can be studies by using sentiment analysis. Finding the sentiments news articles involves many challenges as it requires going through each and every word of the news text articles to find the particular sentiments. This paper presents a brief study of sentiment analysis and opinion mining and an experimental approach of sentiment analysis of news articles obtained from various news websites.

Keywords: Natural Language Processing, News Analysis, Opining Mining, Sentiment Analysis, Text Classification, Text Mining.

Abbreviations: BBC, British Broadcasting Corporation; BOW, Bag of Words; DA, Dependency Analysis; DJIA, Dow Jones Industrial Index; FWL, Fixed Window Length;LB, Lexical based; MLB, Machine Learning Based; MPQA, Multi-Perspective Question Answering; OSN, Online Social Networks; POS, Parts of Speech; RoS, Rest of Sentence; SVM, Support Vector Machine; TSAM, Tweets Sentiment Analysis Model; TF-IDF, Term Frequency- Inverse Document Frequency; MALLET, Machine Learning for Language Toolkit; HTML, Hyper Text Markup Language.

I. INTRODUCTION

With the appearance of the web, there has been an extraordinary change in the public activity, way of life and choices of ordinary citizens. Today it is ordinary action and standard practice for every individual to peruse news on the web and watch promotions in regards to a film, an item or a book before really placing the cash into it. As it has changed their way of life, it likewise has sway on the public activity of a person. The presentation of the new online web-based world, for instance, websites, news channels, message sheets, and news articles are impacting their public activity and the method in which persons take a glimpse at various stuff around them. Persons' viewpoints, in general, will encounter a variation as per the news they go through.

The web-based life has now consumed the signature space on the web. The new client-driven web has an enormous volume of information consistently. Clients are currently co-makers of substance on the web, as opposed to merely latent buyers. The client is currently adding to web-based life going from articles, blog entries, news, tweets, audits and photograph/video transfer. This addition is making a lot of the information on the web as unstructured content.

Access to mainstream correspondence stages has given a path for people, in general, to create feelings, assessments, suppositions, assessments, evaluations, and frames of mind towards stuff, for instance, things, occasions, issues, associations, people, administrations and their properties.

The capacity of such data made today is enormous. By using this Online Social Networks (OSN), we divide this data as a component of our regular day to day existence for enhanced learning and coordinating into our encompassing world. Due to the impact of these networks every day, we start pulling back our meaningful choices and activities with specific existing ideas made by others by their audits. Hence, the change of patterns in this kind of OSNs every once in a while is fascinating and testing. Thorough examinations can prompt suitable expectations. It provides an advantage to Text Mining and explores the importance of Sentiment Analysis, an application written to distinguish opinions.

The way a person reading the news has changed due to the sudden rise of the web, portable innovations, and the internet. Online news and weblogs have replaced Conventional magazines and physical newspapers. Intuitiveness and promptness are the two reasons due

to which readers are more interested in migrating towards online news [1]. These days, individuals need to devour as much news, from the several numbers of sources, as they can, on issues that are essential to them or matters that grab their eye. Intelligence alludes to the natural inclination portrayed by the majority that causes them to expend updates on their advantage. Today, Quickness is a factor that people are more concerned to know the news with deferral in time [2]. Due to the innovation, we acclimate, which enables individuals to profit by giving them every news on the occasions as they needed every day. Online news sites have created robust methodologies to pull people groups' consideration [3]. Online news communicates feelings in regards to news elements, which may involve individuals, puts or even things while writing about occasions that have as of late happened [4]. Thus intuitive feeling rating facilities are provided by different channels of a few news sites, i.e., news can be of any opinion neutral, positive or negative [5].

Discovering the extremity or quality of an opinion which communicated in a composed content can be known as Opinion Mining or Sentiment Analysis [6, 7]. It will be a tedious procedure if we go for the manual marking of feeling words. For sentiment analysis, there are two well-known methodologies are available. The primary procedure utilizes a dictionary of weighted words, and the subsequent procedure depends on the tactics of machine learning. The second methodology utilizes a word lexicon method in which sentiments words are compared with a provided lexicon dictionary to discover the polarity of words. Instead of AI techniques, this methodology does not have to preprocess information, and it does not involve model training [8].

Opinion mining is one kind of natural language processing tool which follows the mind-set of people about a particular theme or item. It is likewise known as Sentiment Analysis; it consists of creating an application for gathering and observing various feelings about the text in various tweets, surveys, remarks or blog entries. Sentiment Analysis can be used in various fields, for example, for launching a new item in the market. It also helps in the decision-making process of an advertisement campaign. It discovers the types of items which are prevalent in the market and even distinguish which socioeconomics like or aversion specific highlights. Sentiment Analysis faces some difficulties as well. A particular feeling word may be positive or negative at one and other circumstances, respectively. A person does not express his opinion in an equivalent mood in every case. In Opinion Mining, in any case, "the picture was great" is altogether different from "the picture was not great". Individuals may be conflicting while declarations. While analyzing, we may find that most of the reviews have both remarks positive and negative.

Facts are investigated in usual text mining, while Sentiment analysis focuses on attitudes. Sentiment analysis covers many fields of research which include opinion summarization, sentiment classification, and feature-based sentiment classification. Characterizing whole reports can be managed by Sentiment classification as indicated by sentiments towards specific articles. Feature-based Sentiment classification thinks about the sentiments on properties of specific articles features task is not quite the same as conventional text summarization because solitary the features of the item are mined on which the clients have communicated their sentiments.

In the majority of the detailed research of the Sentiment analysis, it is noticed that two sorts of content reports, for example, Subjective and Objective are utilized for preparing. Subjective text is characterized as "Linguistic expression of individuals' feelings, sentiments, feelings, and assessments." The Subjective text contains online journals, item or film reviews. Objective text characterized as the text communicates as the substances, occasions, and properties. The Objective text contains news stories or news stories reviews. The objective text archive of the work speaks to make a starter evaluation of the techniques to characterize the discovery and classification of opinion in announced discourse. In Online Social Networks, the individuals express their feelings openly, which means yet this is not over with news stories. One of the benefits of using proposed approach is that their capability of adapting and creating trained models for precise contexts and purposes. In contrast, a predefined list of words is use by the lexical-based techniques, where each word is assigned a particular sentiment. These techniques fluctuate, as indicated by the context where they were made.

The rest of the article is composed as pursues: Related work of sentiment analysis is present in section II. The next section contains the proposed framework and its implementation while section IV comprises of results of the implementation. At long last, Section V exhibits the conclusion of this work.

II. RELATED WORK

Today, Opinion Mining is an extensive and emergent area of study. Subjective and Objective sentiments both may be available in text. The presence of someone's various linguistic expressions such as beliefs, evaluations, emotions, sentiments, speculations, and opinions referred to as Subjective text [2, 5]. Sentiment Analysis is a contemporary area of Information Retrieval that is more related to the opinion that it express instead of the topic of the document [9]. By sentiment analysis, a wide range of human emotions is capture, and most of the studies of sentiment analysis are focused on recognizing the polarity of a provided text. Which means a particular message related to a topic belongs to positive or negative sentiments is automatically calculated [10]. Polarity analysis can offer several applications, especially when dealing with the news articles. According to Pang and Lee (2012), Sentiment Analysis methods are of two types lexical-based and machine learning-based. Methods of machine learning use supervised algorithms of classification, where sentiments recognition has treated as positive or negative. In this method, it trains the classifier using labeled data [11].

The following are the discussed cases of work done by different authors using Lexical based (LB) or machine learning-based (MLB) methods.

Dos Rieis *et al.,* presented a framework to determine the comparison between news popularity and emotion polarity [9]. An experiment using various sentiment analysis techniques is carryout that used 69907

headlines of various news websites- Daily mail, Reuters, BBC, and The New York Times. They investigated the opinions of the polarity of these headlines by extricating the properties from news headlines text. Their research concludes that the fame of news articles impacted by the polarity of the headline. The result of this research is that a neutral news headline has less interest as compared to the positive and negative news headlines.

Godbole *et al.,* (2007) developed a lexicon-based algorithm that is use in discovering the feeling words and objects assigned in the news text and blogs by observing the entity's appearance in the same sentence [7]. They selected news articles and blogs from seven different domains politics, business, sports, crimes, health, and general. Two main tasks of Subjectivity and Polarity has conducted in the experiment, sentiment related to each entity positive and negatives found in the Polarity task and how much sentiment an entity hold found in the Subjectivity. The calculation score of both Polarity and Subjectivity has performed. They use a static corpus that has been crawled online and not uses dynamic corpus.

Islam *et al.,* (2017) proposed a methodology for the classification of online news articles. Sentence level sentiment analysis has carried out and to find sentiment polarity a dynamic dictionary has used, which contains a set of positive and negative words [8]. In this experiment, the following tasks have performed. Selection and Extraction of news articles sentences from the news website, Searches for the positive and negative words in those sentences and describe the words polarities, find the final news article polarity by calculating polarity of all sentences. 91% accuracy has achieved using this experiment.

Meyer *et al.,* (2017) use machine learning and a lexicon-based approach to perform sentiment analysis of financial news articles. Accurate conclusion and results have found by performing eight experiments [12]. Sentiment polarities have found using a lexicon-based approach, which uses the General Inquirer Lexicon (H4N) alongside the Bag of Words (BOW) model. Machine learning approach was used for Parts of Speech (POS) syntactic model and using this approach better accurate result was obtained.

Shirsat *et al.,* (2017) performed document-level sentiment analysis and determine the polarity of a news article, for which a dataset of 2225 documents is used [13]. Preprocessing and post-processing gave done on the entire news dataset. Positive, negative or neutral are the sentiments in which these articles have classified.

Agarwal *et al.,* (2016) use python packages to perform sentiment analysis to classify words, and for identification of positive and negative words, SentiWordNet 3.0 is used [14]. The total influence of the news sentiments has calculated. High accuracy was achieved using this approach; however, this approach is tedious.

Lei *et al.,* (2014) have developed a model for identifying human feelings actuated by news stories, tweets [15]. Different modules such as a selection of documents, lexicon generation and parts of speech tagging were part of the developed model. Initially, the model creates a training set at that point applies methods of POS tagging and feature extraction. Then, based on the document, probabilities of the feelings are calculated for the generation of social emotion lexicons. For testing the model accuracy, a dataset of 40,897 news articles has gathered.

Fong *et al.,* (2013) present work on different machine learning methods and comparisons of various algorithms for efficient sentiment analysis [16]. Positive, negative and neutral classes based on these three classes the text is classified. The work recommended that better results can obtain using Naïve Bayes classifier as it provides high accuracy when compared with c4.5, decision tree, maximum entropy, and winnow classifiers.

Zhou *et al.,* (2013) present a work that developed an approach of the Tweets Sentiment Analysis Model (TSAM) using which people's opinions and social interests for a particular social event were gathered [17]. Australian federal elections 2010 event is use as an example of a dataset. The study of emotions expressed and opinions sentiments of text were perform. This work suggests that only use those words which pose some sentiments as opposed to utilizing all words for sentiment analysis. The work presents a lexicon-based sentiment analysis system which provides various techniques for enhancing the accuracy of classifier like Naïve Bayes.

Li and Liu (2010) used a k-means clustering algorithm to develop an approach of sentiment analysis. For applying weighting on the raw data, the TF-IDF technique is used [18]. A progressively steady clustering result has attained using a voting mechanism. By applying multiple implementations of the clustering process, the result has achieved, and for improving the clustering result, the term score has used. Cauterizations of documents into positive and negative groups have performed.

Popescu and Etzioni (2007) proposed an approach OPINE, an unsupervised information extraction system that extricated opinions and product properties from reviews [19]. In the first step OPINE extricated noun phrases from reviews and returns phrases which have a high frequency than the threshold and for extricating precise properties, these phrases can be assessed by OPINEs feature assessor. In order to find the opinion words, OPINE applies manual extraction rules.

Somprasertsri and Lalitrojwong (2010) developed an approach to recognize the semantic relationships between opinion and product features [20]. The approach of his work is to mine the opinion and item features following the semantic information and syntactic information by implementing ontological knowledge and dependency relations with a probabilistic based model.

Theussl *et al.,* (2009) uses R package *tm* to propose a system for large scale sentiment analysis [21]. They calculate sentiment scores using the polarity of annotating terms and the New York Times static text corpus. The Authors also explain the utilization of dispersed text mining methods along with the Map-Reduce concept. They also reviewed some related authors worked and presented the idea of generating sentiment scores from various news websites articles and also relates to economic situations. They focus on the whole text and do not concentrate on the filtration of corpus to related text and lines.

Balahur *et al.,* (2011) perform a comparative analysis of multilingual sentiment analysis by machine translation and supervised learning [22]. The work explains methods to gather multilingual data using machine translation and classify sentiment using supervised learning. The work uses algorithms, various features, three different machine translation systems, and meta-classifiers. This study obtains the multilingual data for sentiment analysis that shows that Machine translation systems are a better choice for implementation.

Goonatilake and Herath (2007) proposed an approach that shows that there is a secure connection between oil prices, economic indices and news articles [23]. They calculated the influence of news articles on significant indices like NASDAQ, S&P 500 and Dow Jones Industrial Index (DJIA). They develop a regression model for measuring the everyday fluctuations of stock prices and categorize the news articles into four classes.

Breen (2011) proposed an approach of airline satisfaction mechanism, in which the system calculates the sentiment and polarity scores by using keyword tagging on twitter feeds [24]. Using this approach, the authors explained well the idea of customer sentiments and satisfaction ratings about airlines.

Yu *et al.,* (2007) proposed a methodology of text mining in which the system discover the opinions of news items and explain its influence on energy requirement [25]. In this approach, the news sentiment has compared with the market up-downs of energy requirements and prices by quantifying the news sentiment and represented them as time series.

Agić *et al.,* (2010) presented an approach that correlates between online news websites reporting and stock exchange trade figures [26]. The authors explained that news articles sentiments that have taken from a period of a trend are correlated overall market trends. They found that polarity phrases in the news articles control the general sentiment of the news articles. They also research the financial domain for classification and automatic detection of polarity in unseen texts using a rule-based polarity term detection module [26].

Zhai *et al.,* (2009) presented a methodology of collecting the sentiment scores of financial news articles from online news websites and uses it for the market predictor [27]. Automatic and Manual, two approaches have used to discover the sentiments. Human-based marking of the sentiment tag for each news article positive, negative, or neutral is a manual method of sentiment classification. They used the Stanford Classifier and coded a Java data processing module for analyzing and predicting the sentiments in an automatic method of sentiment classification. For performance analysis, they allocate positive article +1, negative article -1, and neutral article 0 scores, and overall sentiment value has calculated using these scores [27].

Devitt and Ahmad (2007) have performed extensive research Financial News for Sentiment Polarity Identification [28]. For the detection of sentiment and polarity in news articles, the authors used a lexical based text representation. News articles have used for Part-of-speech tagging and using topological features available in WordNet cohesive structures are acquired. Text and concepts which have a connection between relations in WordNet have collected as nodes in a graph structure. The efficiency of the cohesion graph has accessed by executing Polarity metrics.

Palanisamy *et al.,* (2013) discovered the sentiments of Twitter's tweet, which comprises hashtags, emoticons, and word variations by using a lexicon-based method [29]. They used Serendio taxonomy for building their lexicon which comprises stop words, negation, and phrases, positive and negative. Words sentiment of the tweet is calculated depending upon the guesstimate created on the sentiment orientation. Sentiment entity's average sum such as hashtags, emoticons, and word variation can be named as the sentiment of a tweet. In light of the contextual sentiment orientation of text, they categorized the tweet's polarity and obtained an F-score of 0.8004 [29].

Wilson *et al.,* proposed an approach of subjectivity analysis of news articles [30, 31]. They use different syntactic and lexicon features, along with machine learning methods like SVM [32] and BoosTexter [33]. Multi-Perspective Question Answering (MPQA) corpus [31] has utilized to conduct their experiments. For sentiment analysis, they recommend a two-step method in which first step polar or neutral texts have classified from a piece of text and second step recognize the polar text polarity. Using the BoosTexter algorithm for the one-step method, they get better outcomes of accuracy 74.3%.

Cambria *et al.,* (2012) presented an initial idea of social media sentiment analysis [34]. They applied their technique on the text sentences using common-sense concepts for performing sentiment analysis of the text sentences. They used comments obtained from a YouTube video for a product review and tried to predict ratings by performing sentiment analysis of the user's comments. They obtained great results of 86% recall, 97% precision, and 91% F-measure.

Cambria *et al.,* (2013) proposed a methodology of sentiment-mining framework motivated biologically [35]. The Authors performed big social data analysis using an ensemble approach of multidimensional scaling, artificial neural networks applications, and a construction-based concept-level system [36].

Padmaja *et al.*, (2014) presented three techniques of news articles sentiment analysis [37]. The author presented an approach of recognizing the scope of negation in news articles by utilizing three already available techniques, namely Dependency Analysis (DA), Fixed Window Length (FWL), and Rest of Sentence (RoS). They obtained the news articles from two news website The Hindu and NDTV, which later they differentiated as good and bad news articles. The author utilized Support Vector Machine and Naïve Byes for the machine learning approach and SentiWordNet for a lexical based approach.

Padmaja *et al.*, (2014) has presented a methodology in which comparison is carryout between the negation scopes of various news articles [37]. The author uses three already existing methodologies for discovering the news article's scope of negation for two different political parties' viz. Congress and BJP. Dependency Analysis (DA), Fixed Window Length (FWL), and Rest of Sentence (RoS) have used with 0.75, 0.69, and 0.58 F measures, respectively. Dependency Analysis performs better than two other techniques.

Fong *et al.*, says that several challenges associated with sentiment analysis, out of which automatically detection of positive and negative text is an essential challenge [16]. Here the author proposed a novel method of sentiment analysis MALLET (Machine Learning for Language Toolkit). The author implements several algorithms using MALLET. This paper focuses on checking the neutral vocabulary in news articles as compared to the emotional vocabulary of opinion.

Raina (2013) presented an approach of sentic computing in which sentiment analysis of news articles has performed using common sense knowledge [38]. The ability to understate the human sentiment of computers can be known as sentic computing. SentiNet and ConceptNet have used for getting the common sense of knowledge. Using this approach, the author obtained some excellent results as 71% accuracy and 91% precision.

III. PROPOSED WORK

Online news article's sentiment analysis has done using the Lexicon-based technique. Both supervised and unsupervised approaches may utilize for sentiment analysis. In a supervised approach of sentiment analysis, a set of labels have used as training data, which has utilized to fabricate the classifier model, which, in turn, used for testing with unclassified data. In the Unsupervised approach of sentiment analysis, training data is not involved in any model training. In this approach, the polarity of words has used for performing the sentiments of words.

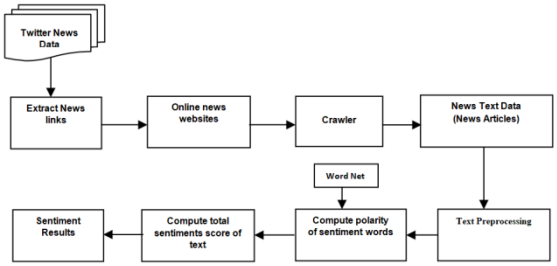

Fig.1. Sentimental Analysis of Online News Articles.

At phrase level, word level, sentence level, or document level, sentiment analysis of the text can be implemented. This work focuses on the document level, which recognizes various opinions such as positive, negative, or neutral in the news articles documents. This approach of sentiment analysis uses WordNet lexical dictionary.

This approach consists of 5 steps. In the first step, data collection has performed from news websites. Next, data preprocessing have performed to minimize the inconsistencies. The WordNet lexical dictionary has used for calculation of the polarity of the word. These steps have elaborated below

A. Data Collection

In the first module, Twitter data has collected using tweepy, and news links have extracted, and crawling has performed using a web crawler; the news articles have downloaded from various news websites. These articles are downloaded in the HTML format, and using an HTML parser; the text is extracted and stored in a temporary text file. In this paper, we used news articles from various online websites such as Indian Express, The Hindu, The Wire, The Quint, The Print, Telegraph India, and Scroll in for performing sentiment analysis. The news articles comprise seven categorize, such as Crime, Business, Political, Entertainment, Media, Tech, and World News.

B. Data Pre-processing

News article's text has preprocessed for minimizing the inconsistencies, and the dataset can be used more effectively in sentiment analysis. In first "Tokenize" operator is used for performing tokenizing. It is a process of breaking a group of sentences into individual symbols, phrases, or words, which are known as tokens. Punctuation marks are removed during the tokenization. In the next step, the "filter stop word (English)" operator is used to remove stop words. In the last step, Stemming has performed using a "stem (WordNet)" operator. In this process, the inflected or derived words have minimized.

C. Polarity Computation of Words

Once the preprocessing has carried out, TF-IDF is used to know the importance of certain words in a document. Term Frequency-Inverse Document Frequency technique have used to identify the words that often appear in a manuscript are marked essential, and weight has assigned to them. Once the identification of essential words has made, sentiment scores are assigned to discovered words using a dictionary. In this experiment, the lexical database for the English language known as the WordNet dictionary has used. More than 90k different word senses and 118k different word forms are available in the WordNet dictionary [17]. The reason behind using the WordNet dictionary is that it helps to determine the opinion words and assign sentiment scores.

D. Calculate Total Sentiment Score

After the polarity computation of each word, phrases, or sentences, each document has assigned with a polarity. Here each news article is treated as a document. The polarity assigned to document is the sum of all polarities of all words/phrases and sentences in new articles. After the calculation of polarity sentiments of news articles is calculate. Text having +1 sentiment score is considered as positive, and -1 sentiment score has considered as the negative. The 0 sentiment score represents neutral sentiments. The sentiment score has calculated using the SentiWordNet 3.0.0 dictionary. An extension of the WordNet dictionary is the SentiWordNet 3.0.0 dictionary. Synset IDs are used to connect WordNet and SentiWordNet. The total sentiments of news articles have calculated using Score sentiment function based on WordNet and SentiWordNet dictionary.

E. Sentiment Results

By observing the total sentiment score of news articles, these have classified as positive, negative, and neutral sentiments. Then by taking the average of total word sentiments, news articles sentiments were calculated.

IV. RESULT DISCUSSION

This experiment uses news articles from the seven news websites, as described in the earlier section. The experiment crawled one-month news articles from

various news websites. After the implementation of news articles having a sentiment score, +1 have treated as positive, and score -1 have considered negative, whereas score 0 treated as neutral. The following table shows the experimental results.

It has noticed that most of the articles are belongs to either positive or negative sentiments, and only some articles belong to neutral sentiments. Most of the news articles in the crime and politics categories are belong to negative sentiments; however, the majority of the news articles in entertainment, business, and sport belong to positive news. The politics, tech, and world news share an equal part of positive and negative sentiments. Graphical results have shown in Table 1.

Figs. 3 to 5 provides a graphic comparison of sentiments types with its corresponding news articles.

In the Positive and negative sentiments graphs, some categories like politics, world news and tech news almost share the same percentage. While the Crime category shares the large space in negative sentiment and less space in the positive sentiment.

Entertainment and Tech categories nearly share the equal percentage in every sentiment of the news articles. If we closely observe each category of neutral sentiments we come to a conclusion that every category shares nearly equal percentage.

Table 1: Results of sentiment analysis of news articles.

News Category	Positive	Negative	Neutral	Total Articles
Politics	510	505	145	1160
Crime	123	877	200	1200
Entertainment	345	253	134	732
Business	423	302	183	908
Tech	263	220	110	593
World News	293	273	161	727
Sport	517	300	163	980

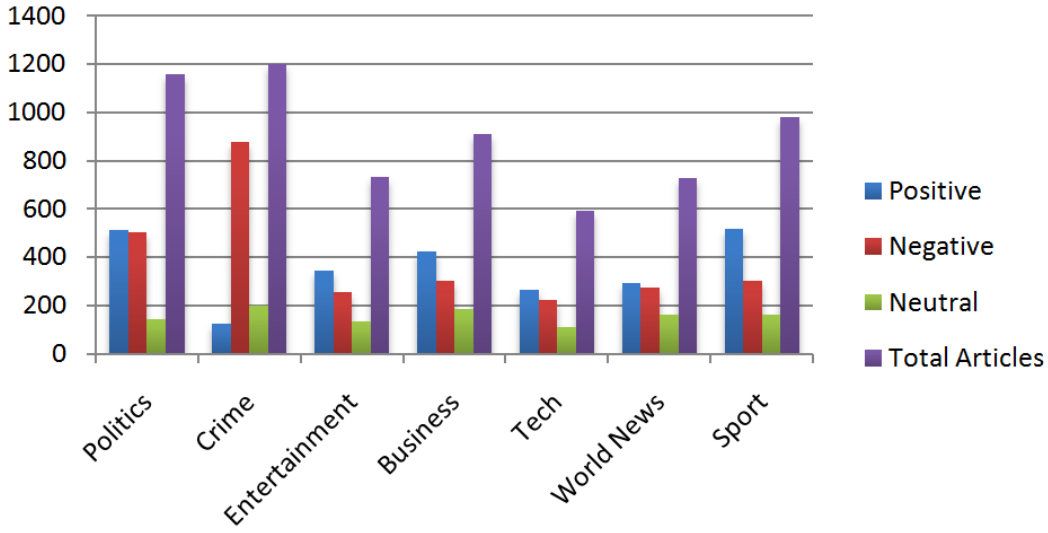

Fig. 2. Comparison of categorized News articles with different moods.

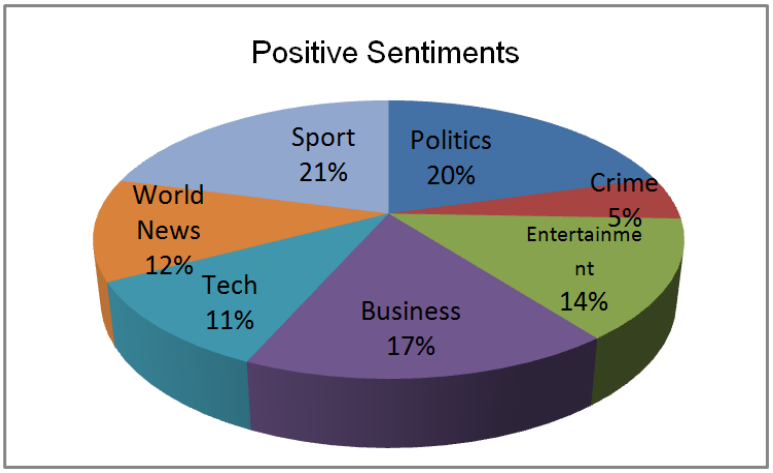

Fig. 3. Percentage of every category under positive sentiments news articles.

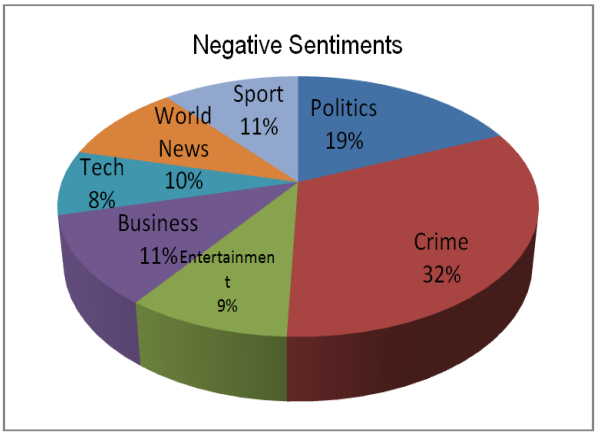

Fig. 4. Percentage of every category under negative sentiments news articles.

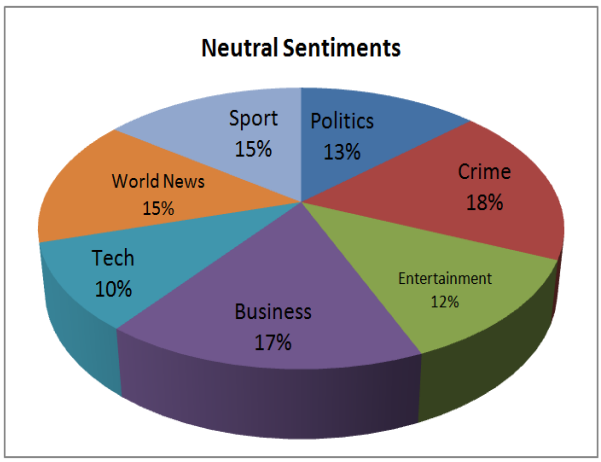

Fig. 5. Percentage of every category under neutral sentiments news articles.

V. CONCLUSION

There are many applications of the information systems in which Sentiment Analysis have used, like summarizing review, opinion mining, classifying reviews, and various real-time applications. This paper applies sentiment analysis on news articles crawled from various online news websites from September 2019 to October 2019. One of the limitations of this approach is that it has noticed that sentiment analysis focused on English words and work on other languages like Arabic, Italian, and other languages need to perform. Many challenges may arise in dealing with other languages like implicit product features handling, sentence/ document complexity, negation expressions handling, and a summary of opinions of product features/attributes. Future work of this research could be implementing this sentiment analysis technique for other languages as well by overcoming the above mentioned challenges.

ACKNOWLEDGEMENTS

The authors send gratitude to the university for the facilities and support provided during the research.

Conflict of Interest. The authors declare no conflict of interest.

REFERENCES

[1]. [online] 5G Radio Access, Research and Vision: http://www.ericsson.com/res/docs/whitepapers/wp-5g.pdf [Last Accessed: 08-February-2014].

[2]. Kanter, T., Forsström, S., Kardeby, V., Walters, J., Jennehag, U., & Österberg, P. (2012). MediaSense – an Internet of Things Platform for Scalable and Decentralized Context Sharing and Control. *The Seventh International Conference on Digital Telecommunications,* 27-32.

[3]. [online] MediaSense | The Internet of Things Platform,http://www.mediasense.se/ [Last Accessed: 08-February-2014]

[4]. Cugola, G., & Jacobsen, H. (2002). Using Publish/Subscribe Middleware for Mobile Systems. *ACM SIGMOBILE Mobile Computing and Communications Review*, *6*, 25–33.

[5]. Zarko, I. P., Antonic, A., & Pripužic, K. (2013). Publish/subscribe middleware for energy-efficient mobile crowdsensing. *ACM conference on Pervasive and ubiquitous computing adjunct publication (UbiComp '13 Adjunct). Zurich*. , 1099-1110.

[6]. Dos Rieis, J. C. S., de Souza, F. B., de Melo, P. O. S. V., Prates, R. O., Kwak, H., & An, J. (2015). Breaking the news: First impressions matter on online news. In *Ninth International AAAI conference on web and social media,* 357-366.

[7]. Godbole, N., Srinivasaiah, M., & Skiena, S. (2007). Large-Scale Sentiment Analysis for News and Blogs. *Icwsm*, *7*(21), 219-222.

[8]. Islam, M. U., Ashraf, F. B., Abir, A. I., & Mottalib, M. A. (2017). Polarity detection of online news articles based on sentence structure and dynamic dictionary. In *2017 20th International Conference of Computer and Information Technology (ICCIT)*, 1-5.

[9]. Pereira, J., Fabret, F., Llirbat, F., &Shasha, D. (2000). Efficient matching for web-based publish/subscribe systems. In *International Conference on Cooperative Information Systems*, Eilat, Israel, September 6-8, 2000, 162-173.

[10]. Fabret, F., Jacobsen, H. A., Llirbat, F., Pereira, J., Ross, K. A., & Shasha, D. (2001). Filtering algorithms and implementation for very fast publish/subscribe systems. In *Proceedings of the 2001 ACM SIGMOD international conference on Management of data*, 115-126.

[11]. Pang, B., Lee, L., Vaithyanathan, S.: Thumbs up Sentiment classi_cation using machine learning techniques. In *Proceedings of the 2002 Conference on Empirical Methods in Natural Language Processing(EMNLP).* (2002) 79-86.

[12]. Meyer, B., Bikdash, M., & Dai, X. (2017). Fine-grained financial news sentiment analysis. In *Southeast* Conference, 1-8.

[13]. Shirsat, V. S., Jagdale, R. S., & Deshmukh, S. N. (2017). Document level sentiment analysis from news articles. In *2017 International Conference on Computing, Communication, Control and Automation (ICCUBEA),* 1-4.

[14]. Agarwal, A., Sharma, V., Sikka, G., & Dhir, R. (2016). Opinion mining of news headlines using SentiWordNet. In *2016 Symposium on Colossal Data Analysis and Networking (CDAN)*, 1-5.

[15]. Lei, J., Rao, Y., Li, Q., Quan, X., & Wenyin, L. (2014). Towards building a social emotion detection system for online news. *Future Generation Computer Systems*, *37*, 438-448.
[16]. Fong, S., Zhuang, Y., Li, J., & Khoury, R. (2013). Sentiment analysis of online news using mallet. In *2013 International Symposium on Computational and Business Intelligence*, 301-304.
[17]. Zhou, X., Tao, X., Yong, J., & Yang, Z. (2013). Sentiment analysis on tweets for social events. In *Proceedings of the 2013 IEEE 17th International Conference on Computer Supported Cooperative Work in Design (CSCWD),* 557-562.
[18]. Li, G., & Liu, F. (2010). A clustering-based approach on sentiment analysis. *2010 IEEE International Conference on Intelligent Systems and Knowledge Engineering*, Hangzhou, 331-337.
[19]. Popescu, A. M., & Etzioni, O. (2007). Extracting product features and opinions from reviews. In *Natural language processing and text mining*, 9-28. Springer, London.
[20]. Somprasertsri, G., & Lalitrojwong, P. (2010). Mining Feature-Opinion in Online Customer Reviews for Opinion Summarization. *J. UCS*, *16*(6), 938-955.
[21]. Theussl, S., Feinerer, I., & Hornik, K. (2009). Distributed text mining with tm. In *The R User Conference*.
[22]. Hofmarcher, P., Theußl, S., & Hornik, K. (2011). Do Media Sentiments Reflect Economic Indices?. *Chinese Business Review*, *10*(7), 487-492.
[23]. Goonatilake, R., & Herath, S. (2007). The volatility of the stock market and news, *International Research Journal of Finance and Economics*, *11,* 53-65.
[24]. Breen, J. (2011). R by example: Mining Twitter for consumer attitudes towards airlines. *Boston Predictive Analytics Meetup Presentation*.
[25]. Yu, W. B., Lea, B. R., & Guruswamy, B. (2007). A Theoretic Framework Integrating Text Mining and Energy Demand Forecasting. *IJEBM*, *5*(3), 211-224.
[26]. Agić, Ž., Ljubešić, N., & Tadić, M. (2010). Towards sentiment analysis of financial texts in croatian. In *Proceedings of the Seventh International Conference on Language Resources and Evaluation*.
[27]. Zhai, J. J., Cohen, N., & Atreya, A. (2009). CS224N Final Project: Sentiment analysis of news articles for financial signal prediction, 1-8.
[28]. Devitt, A., & Ahmad, K. (2007). Sentiment polarity identification in financial news: A cohesion-based approach. In *Proceedings of the 45th annual meeting of the association of computational linguistics,* 984-991.
[29]. Palanisamy, P., Yadav, V., & Elchuri, H. (2013). Serendio: Simple and Practical lexicon based approach to Sentiment Analysis. In *Second Joint Conference on Lexical and Computational Semantics (* SEM), 2: Proceedings of the Seventh International Workshop on Semantic Evaluation (SemEval 2013),* 543-548.
[30]. Wilson, T., Wiebe, J., & Hoffmann, P. (2009). Recognizing contextual polarity: An exploration of features for phrase-level sentiment analysis. *Computational linguistics*, *35*(3), 399-433.
[31]. Wilson, T. A. (2008). *Fine-grained subjectivity and sentiment analysis: recognizing the intensity, polarity, and attitudes of private states* (Doctoral dissertation, University of Pittsburgh).
[32]. Joachims, T. (1998). *Making large-scale SVM learning practical* (No. 1998, 28). Technical Report.
[33]. Schapire, R. E., & Singer, Y. (2000). BoosTexter: A boosting-based system for text categorization. *Machine learning*, *39*(2-3), 135-168.
[34]. Cambria, E., Grassi, M., Hussain, A., & Havasi, C. (2012). Sentic computing for social media marketing. *Multimedia tools and applications*, *59*(2), 557-577.
[35]. Cambria, E., Mazzocco, T., & Hussain, A. (2013). Application of multi-dimensional scaling and artificial neural networks for biologically inspired opinion mining. *Biologically Inspired Cognitive Architectures*, *4*, 41-53.
[36]. Cambria, E., Rajagopal, D., Olsher, D., & Das, D. (2013). Big social data analysis. *Big data computing*, *13*, 401-414.
[37]. Padmaja, S., Fatima, S. S., & Bandu, S. (2014). Evaluating sentiment analysis methods and identifying scope of negation in newspaper articles. *Int. J. Adv. Res. Artif. Intell*, *3*(11), 2014, 1-6.
[38]. Raina, P. (2013). Sentiment analysis in news articles using sentic computing. In *2013 IEEE 13th International Conference on Data Mining Workshops*, 959-962.

How to cite this article: Ahmed, J. and Ahmed, M. (2020). A Framework for Sentiment Analysis of Online News Articles. *International Journal on Emerging Technologies*, *11*(3): 267–274.