

Determination of news biasedness using content sentiment analysis algorithm

SV. Shri Bharathi¹, Angelina Geetha²

¹Department of Computer Science and Engineering, Saveetha Institute of Medical and Technical Sciences, India

²Department of Computer Science and Engineering, Hindustan Institute of Technology and Science, India

Article Info

Article history:

Received Jan 29, 2019

Revised Apr 30, 2019

Accepted May 21, 2019

Keywords:

Data and text mining

Machine learning

News bias

News values

Sentiment analysis

ABSTRACT

Nowadays, identifying news biases in the social media is one of the most fundamental problems. News bias is a complex process that comprises several dimensions to be taken into account and it is interlinked with social, political and economic problems. In general, news bias has the ability to reflect opinion of people about a topic or government policies and actions. The proposed algorithm develops a system which can detect the biasedness of news topics from different news Websites. This approach automatically collects the news contents from various online news media portals and then consolidates them for the determination of news biasedness. In the experimental study, the news topics are gathered from various Websites of U.S., U.K., and India. For training dataset 3265 news sentences were collected under various news topics from 20 different news Websites. The effectiveness of classification of algorithm is proved by the extensive experimental study. The proposed algorithm provides a method improves the determination of news biasedness, which in turn may help in providing impartial, unbiased and reliable information.

*Copyright © 2019 Institute of Advanced Engineering and Science.
All rights reserved.*

Corresponding Author:

SV. Shri Bharathi,

Department of Computer Science and Engineering,

Saveetha Institute of Medical and Technical Sciences,

Saveetha School of Engineering, Thandalam, Chennai 602 105, Tamilnadu, India.

Email: shribharathi01@gmail.com

1. INTRODUCTION

Newspaper reading plays a vital role in our daily lives. Nowadays, increasing popularity of online social networks has gathered hundreds of millions of users. Recently, with widespread access to online news websites, users can browse and retrieve news more easily. In general, websites are expected to give fair and impartial news comprising of facts. In order to develop a successful approach in the prediction of news bias, the background knowledge of news websites tendencies is necessary for determining whether news articles are fair and credible.

In general, bias is defined as prejudice against or in favor of one thing, person, or group compared with another, usually in a way considered to be unfair. Media bias is when journalists, news producers, and news outlets show bias in the selection of events and stories as well as the ways they are reported. News bias cannot be understood without understanding of the context of media industry as a whole.

Research within the field of automatic identification of news bias additionally has been shifting its attention to opinion mining and sentiment analysis in the news. Most opinion analysis has been done on very subjective texts like product launch, reviews or blogs, where the opinion of the author is expressed freely in a very subjective and biased way. Recently, sentiment analysis has been drawing attention to news articles.

This research paper investigates the public news sentiment, as expressed in large scale collections of daily news posts collected from related Websites. This can be used to predict the biasedness of news contents.

This paper is organized as follows: Section 2 gives a survey on related work. Section 3 describes system for Sentiment Analysis of News Biasedness in sentence level documents by using the proposed algorithm. Section 4 presents the experimental results. Finally, Section 5 concludes our work with a scope for future enhancements.

2. RELATED WORK

Comparison between the classifier and lightweight neural network architecture with Logistic Regression, Gradient Tree Boosting, SVM and Naive Bayes aims to identify the articles published in news media bias [1]. Machine learning and crowdsourcing techniques are combined and supervised learning algorithms are used to identify the articles pertaining to political events. Online human judges are used to classify the political articles [1-2].

In order to measure the news bias a new algorithm is combined with the theoretical and methodological part [3]. Also political bias content in large daily newspapers are concentrated to identify the source of bias [4]. A new opinion mining system is developed using Support Vector Machine and NLP tools on newspaper headlines. A classification model was built and news headlines are fetched and processed in core NLP techniques [5].

The prediction of opinion mining on stock market by carried out by combing the mathematical model with RSS news feeds which results into higher accuracy [6]. Some popular classifiers such as decision tree, random forests, and support vector machine are compared with each other and final result will be combined with the popular majority vote classifier technique [7].

A new fuzzy situational analysis model (situational model) is introduced to assesss the stock market current state [8]. The stock market prediction accuracy is explored by combining one or more stock level indicators with twitter and RSS feeds [9].

The unfairness of the American media is demonstrated and from a series of interrelated factors it identifies the partiality stems [10]. Ideological online news media was analysed by providing citizens with easier access to relevant information, for the better understanding of biased news sites several times. Also bias can be estimated based on the language used to describe an issue, as when an outlet selects more negative terminology for the opposition [11]. A novel hybrid approach was introduced which combines SVM and random forest method to detect the sentiments of text documents [12].

Identification and extraction of factual information was focused automatically based on the interest of events. This work learns the informative clues of subjectivity and also builds a subjective-objective sentence classifier that does not require annotated data as input [13].

A technique called opinion question answering and text summarization was presented and provided the users with the correct information containing answer of a question [14]. The combination of LDA with Antelope which is another NLP tool was analysed and the measurement of semantic framework is incorporated with sentiment bias with various corpus of political news [15].

A sentiment classification framework is presented and incorporated with sarcasm detection. The framework was evaluated using a non-linear Support Vector Machine and Malay social media data for the detection of sarcasm [16]. Three machine learning algorithms such as K-Nearest Neighbors (KNN), Naïve Bayes, and Support Vector Machine (SVM) are used to determine the sentiment of customer satisfaction of Traveloka, Ticket and Agoda by analyzing Facebook posts and comments data from their fan [17].

From daily newspapers lots of political bias content were analysed. This work identified the source news bias. This work needed to give attention to preferences of publishers or customers [18-19]. From online news comments the method analysed the polarity of news comments and achieved high order precision [20]. The performance of the model was achieved by combining the two classes namely anger and disgust [21]. In order to know the revenue of the website two sources were monitored [22].

Efficient prediction of model that scores emotions from all relevant real time stock news available in public domain are analysed [23]. For the prediction of sentiments, a feature based vector model was used, based on novel weighting algorithm [24]. To carry out the sentiment analysis, the social media contents of twitter political data were analysed [25].

3. PROPOSED METHOD

3.1. News Bias

In general, journalists are supposed to provide the readers with impartial, objective, unbiased and reliable information because the reality is somehow different. Without understanding of the context of media industry as a whole news bias cannot be understood. News bias is also regarded as a problem of ideology.

3.2. Several Level Dimensions of News Bias

News bias can be used to understand groups and situations which are not a regular part of people lives. The objective of news bias is to provide balanced reporting and to avoid misleading definitions, Imbalanced reporting, and selective omission, distortion of facts and lack of transparency. Various lists of news values and selection criteria are given in Table 1.

Table 1. Several Levels of Dimensions of News Bias

S.No	Dimension Level	Selection Criteria
1	Frequency	The time distance, time-span of an event and how often it is in the news
2	Threshold/relevance	The size, impact and/or the intensity of an event
3	Unambiguity	Clarity of the meaning of the event to the public
4	Meaningfulness/	The event is of great value and meaningful to the audience if it culturally and geo-politically close to the location of the audience
5	Proximity	The event should match conventional expectations of the people and be harmonious.
6	Consonance	The event has to happen unexpectedly and unplanned
7	Unexpectedness	The event should be continuous and connected over a period of time
8	Continuity	The event should be balanced and complemented by other pieces of information, citations etc. to form a unified news event
9	Personalization	The event is seen as actions of individuals, it should be personalized affecting people, and it should have a human interest
10	Negativity	The event should report bad news; when it bleeds, it leads.

3.3. Different Metrics of News Bias Identification

Numerous techniques that are used for detecting biases are given in Table 2. In this proposed work, the bias-related sentiment news articles are analysed. This paper attempts to design and implement a predictive system for measuring the news biasedness from websites. Based on the polarity of the news sentiments with a score ranging from +1 to -1, the sentiment score is calculated and then based on the positivity and negativity of the score the biasedness is identified for the specific news topic.

Table 2. Different Metrics of News Bias Identification

S.No	Dimension Level	Selection Criteria
1	Keyword Analysis	Keyword frequency analysis is performed using Term Frequency -Inverse Document Frequency (TF-IDF) weights for every word in the text.
2	Analysis of Grammatical differences	News bias is done with regards to the usage of various parts of speech, like adjectives, adverbs and nouns and how these properties differ when reporting about articles from different categories.
3	Readability differences	News bias occurred in readability analysis and it is an indicator of how understandable a text is to a particular group of readers. Readability measures have been used extensively to help to evaluate and develop textbooks, business publications, medical literature and nowadays also news articles.
4	Geographical bias	Analysing geographical bias is to uncover unequal geographical distribution and intensity of the coverage of media attention to certain events in certain countries. News outlets have a certain geographical focus and have been traditionally divided into local, national and international.
5	Topic coverage bias	Topic similarity across various sources are analysed in topic coverage bias. Similarity between the topic or category is analysed in this news bias
6	Speed of reporting bias	News bias depends on the speed of the news reporting among publishers. It measures roughly how quickly a publisher produces an article about an event in comparison to other publishers.
7	News wire citation bias	Automatic detection of news bias is associated with newswire citations. In this bias detection how frequently the selected news publishers cite different news agencies or how newswire hyperlinks are useful to predict bias of websites using their linking pattern.
8	Similarity in the coverage events	The detection news bias is based on the most popular articles across different news sources. Bias is to compute how much the events covered by the two publisher overlap do.
9	Gender bias	Gender bias means studying the mentions of men and women and their relations across different topics in various media channels.
10	Sentiment Analysis	

The architecture of the proposed system is given in Figure 1. News topics are collected from the relevant news web sites. From the collection of news topic, search the news that is to be checked for biasedness. Now the entire queried news topics are stored in a separate document. Now the pre-processing steps are applied and it performs the cleaning process. Unwanted contents such as comma, semi colon, numerals, symbols, date and time etc. are termed as noises and are removed in this step. After pre-processing the Sentence parsing module will separates the document sentences into an individual sentence and stored it in a file [5].

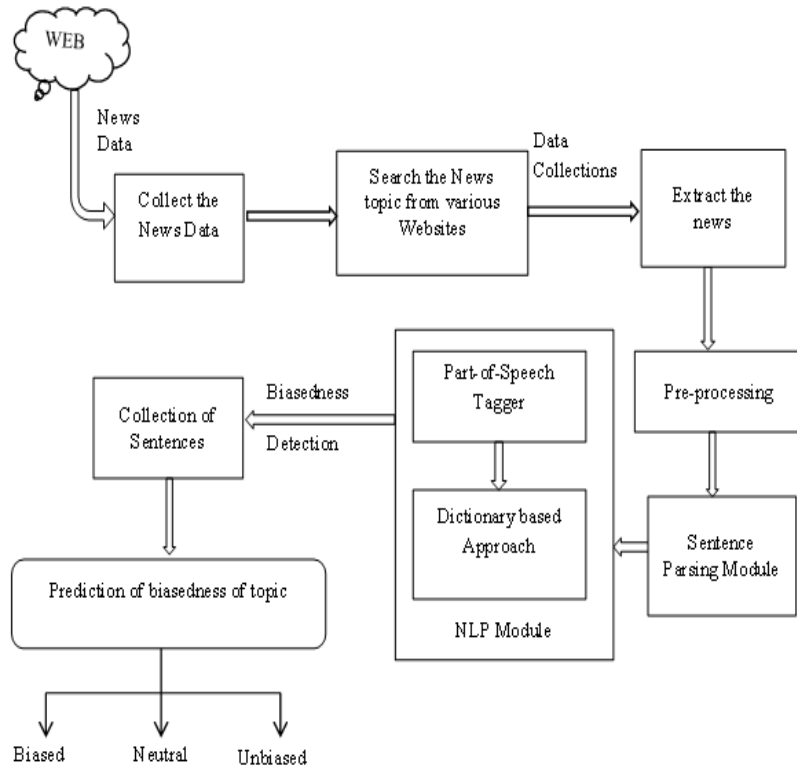


Figure 1. Sentiment analysis of news biasedness in sentence level documents

3.4. Sentiment Analysis of News Biasedness in Sentence Level Documents

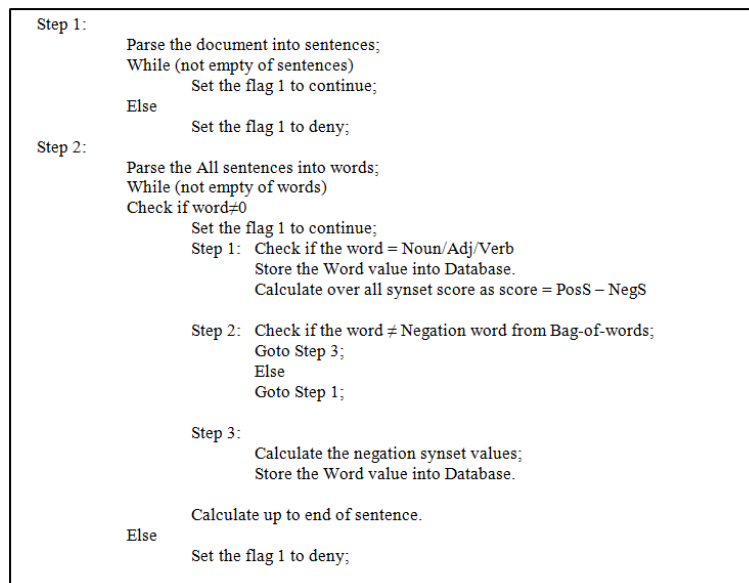
Now all the collected individual sentences are passed to the Natural Language Processing (NLP) module. NLP places a vital role, in order to identify extract sentiment for the words having a positive, negative and neutral value. To find the polarity of the sentence, the part-of-speech tagger, dictionary based approach are used. In Part-of-speech (POS) tagger, all separated sentences are passed and it is a piece of software that reads text in some language and assigns parts of speech to each word, such as verb, noun, adjective, etc. Now each word carries noun, verb or adjective etc. and is passed to the dictionary based approach. The dictionary is used to find the opinion words and their polarities and also used to determine the opinion of words and their polarities. The three types of classified opinions are positive, negative or neutral. The overall result of each sentence is calculated using contents sentiment algorithm [6].

The calculation of sequence of words is as follows (1).

$$Sequence\ of\ words(W) = W_1 + W_2 + \dots + W_n \tag{1}$$

Where n = Number of words

The overall result of each sentence is calculated using sentence score sentiment algorithm. The algorithm for calculating the biasedness of news topic is as follows [9].



Total score of summarization of opinions will explore the result of the biasedness of the news topic. In general, the score ranges from -1.0 to 0.0 and 0.0 to 1.0. For each word the value is assigned and their sum is calculated for finding the calculation of total score value of each sentence. Synset consists of a set of one or more synonyms. If the score value ranges between 0.0 to 1.0 then the sentence is said to be positive. If the score value ranges between -1.0 to 0.0 then the sentence is said to be negative. A neutral value is assigned if its value is 0.0 [13].

4. RESULTS AND DISCUSSION

In the experimental study the biasedness prediction of news is collected from major news portals from U.S. and the U.K. Sentiment Analysis of News Biasedness in sentence level documents like CNN, New York Times, ABC news, BBC etc. Without filtering or manual discarding the web pages were crawled and indexed from the web.

In our experimental study, the dataset is collected from online news web pages for various news topics. Various distinct news topics like “Obama’s farewell speech”, “us election 2016 results”, “Indian Currency demonetization” etc. is queried for analysing the biasedness of topic [11]. Using the classification technique, for a particular topic the absolute frequency of topics is measured. Various list of news Websites and data set are given in Tables 3 and 4.

Table 3. List of Various News Websites

Country Name	News Websites	URL Name
United States	1. New York Times	http://www.nytimes.com/
	2. CNN	http://www.cnn.com/
	3. USA Today	http://www.usatoday.com/
	4. The Guardian	http://www.theguardian.com/
	5. Washington Post	https://www.washingtonpost.com
	6. StudentNewsDaily	https://www.studentnewsdaily.com
	7. News Busters	https://www.newsbusters.org
	8. Denver post	http://www.denverpost.com
	9. The Age	https://www.theage.com.au
	10. Thomson Reuters	https://in.reuters.com
United Kingdom	1. BBC	http://www.bbc.co.uk/
India	1. The Hindu	http://www.thehindu.com/
	2. India Times	http://www.indiatimes.com/
	3. Indian Express	http://www.indianexpress.com/
	4. NDTV	http://www.ndtv.com
	5. Economic Times	https://economictimes.indiatimes.com
	6. Yahoo	https://www.yahoo.com
	7. Business Standard	http://www.business-standard.com
	8. Zee News India	http://zeenews.india.com/
	9. TimesnowNews	http://www.timesnownews.com

Table 4. Dataset for Training and Testing Samples

S.No	News Topic	Training Set	Testing Set
1.	Obama's farewell speech	150	70(USA Today, New York Times)
2.	US election 2016 results	180	100 (CNN, News, The Guardian)
3.	Hillary Clinton lost in US election 2016	140	60 (BBC)
4.	Indian Currency Demonetization	280	130(The Hindu, India Times, Indian Express)
5.	TamilNadu Assembly Election in 2016 in India	130	70 (The Hindu, NDTV)
6.	Is the media biased toward Clinton or Trump? Here is some actual hard data	145	30 (Washington Post)
7.	Broadcast Networks Skip Weak Economic Growth?	150	40(StudentNewsDaily, News Busters)
8.	Maruti's dominance, small car bias & global issues make India a tough market for big auto MNCs	175	35(India Times)
9.	The pathetic neediness of Trump	250	30 (Denver Post, The Age)
10.	Airlines scramble to minimise losses as Bali volcano costs grow	230	20(Reuters, Yahoo, Business standard)
11.	US charges three Chinese for hacking American corporations	140	40 (Zee News India, India Times)
12.	PM Modi Launches Hyderabad Metro, Takes First Ride: 10 Points	150	50 (NDTV)
13.	Reliance Communications enters deal to sell Reliance BIG TV	120	40(Business standard, TimesnowNews)
14.	Obama reports for jury duty in Chicago and is dismissed	170	50(The Guardian, BBC)
15.	Obama's speech at Gates Foundation	140	50(CNN, New York Times)

Dataset for Training and Testing samples are given in Table 3. There are two phases in the proposed system. First one is training phase and second one is testing phase. In first phase, punctuations, tokenization process are processed before the selection of news contents. After that the news contents are classified according to the predicted category like biased, unbiased and neutral. Second in testing phase, the selected news topics are compared with already available training set of the news topics. The previous work made use of the WiSARD libraries and NLTK library. The Text was represented as bag-of-words, where order is not considered [1].

The accuracy measure of the topic classification is shown in Table 5. The accuracy classification of the measure is calculated by precision and recall. Precision is the ratio of total number of correctly identified positive instances and total number of instances that are identified as positive (2).

$$\text{Precision} = \frac{tp}{tp+fp} \quad (2)$$

Table 5. Accuracy Classification Measures of Biased Topic

S.No	News Topic	Total	Positive	Negative	Neutral	Precision
1.	Obama's farewell speech	220	180	30	10	81.8%
2.	US election 2016 results	280	100	120	60	89.25%
3.	Hillary Clinton lost in US election 2016	200	120	70	10	80%
4.	Indian Currency Demonetization	310	80	220	10	93.54%
5.	TamilNadu Assembly Election in 2016 in India	200	100	95	5	85%
6.	Is the media biased toward Clinton or Trump? Here is some actual hard data	175	135	25	15	77.14%
7.	Broadcast Networks Skip Weak Economic Growth?	190	150	30	10	78.94%
8.	Maruti's dominance, small car bias & global issues make India a tough market for big auto MNCs	210	160	33	17	76.19%
9.	The pathetic neediness of Trump	280	220	60	0	78.57%
10.	Airlines scramble to minimise losses as Bali volcano costs grow	250	210	40	10	84%
11.	US charges three Chinese for hacking American corporations	180	150	25	5	83.33%
12.	PM Modi Launches Hyderabad Metro, Takes First Ride: 10 Points	200	160	30	10	80%
13.	Reliance Communications enters deal to sell Reliance BIG TV	160	120	35	5	75%
14.	Obama reports for jury duty in Chicago and is dismissed	220	180	40	0	81.81%
15.	Obama's speech at Gates Foundation					

The accuracy percentage of 15 news topics between the WiSARD algorithm (Previous) and our proposed algorithm is shown in the below Figure 2. The first line of each topic describes the accuracy level of WiSARD algorithm and second line of each topic shows the accuracy level of our proposed algorithm. For each news topic the accuracy level of our approach is more than 7 % improvement than the previous one.

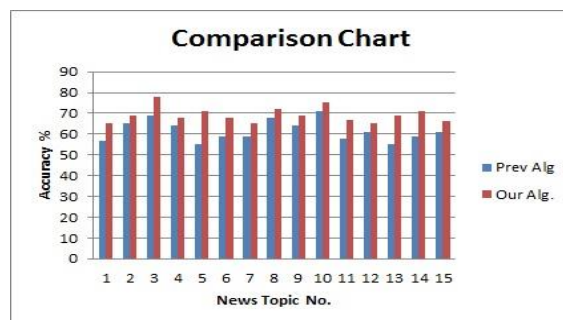


Figure 2. Accuracy level of the news topics between the two algorithms

5. CONCLUSION

Media has tremendous power in setting the mood and behaviour of common people in day to day activities. It is essential that news media should be fair and accurate. Various applications such as marketing, financial dealings, stock exchange values, international trade depends on the biased of news. In the proposed work we have designed an algorithm for estimating the biasedness of news. The novelty of our work is we have collected the news from various online news media portals and then consolidated for the determination of news biasedness. Our approach focuses on the detection of biasedness of news topics and measures the accuracy level in comparison with previous algorithm. Our further work focuses on considering the temporal components related to publishing of news and study its impact on stock exchange market fluctuations.

REFERENCES

- [1] R. Dutra Cavalcanti, *et al.*, "Evaluating Weightless Neural Networks for Bias Identification on News," 14th IEEE International Conference on Networking, Sensing and Control, ICNSC'17, pp. 257-262, 2017.
- [2] C. Budak, *et al.*, "Fair and Balanced? Quantifying Media Bias through crowdsourced content analysis," *Journal on Public Opinion Quarterly*, vol. 80, pp. 250-271, 2016.
- [3] K. S. Watanabe, *et al.*, "Measuring news bias: Russia's official news agency ITAR-TASS' coverage of the Ukraine crisis," *European Journal of Communication*, vol. 32, pp. 224-241, 2017.
- [4] R. Wayne, *et al.*, "Framing the Right Suspects: Measuring Media Bias," *Journal of Media Economics*, vol. 26, pp. 122-147, 2013.
- [5] C. Jashubhai Rameshbhai *et al.*, "Opinion mining on newspaper headlines using SVM and NLP," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 9, pp. 2152 -2163, 2019.
- [6] SV. Shri Bharathi, *et al.*, "Sentiment Analysis for Effective Stock Market Prediction," *International Journal of Intelligent Engineering and Systems*, vol. 10, pp. 146-153, 2017.
- [7] Zamahsyari, *et al.*, "Sentiment Analysis of Economic News in Bahasa Indonesia Using Majority Vote Classifier," Proceedings of the International Conference on Data and Software Engineering, ICoDSE'16, pp. 1-6, 2016.
- [8] Siham Abdulmalik Mohammed *et al.*, "The Decision-making Model for the Stock Market under Uncertainty," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 7, pp. 2782-2790, 2017.
- [9] SV. Shri Bharathi, *et al.*, "Sentiment Analysis of Twitter and RSS News Feeds and Its Impact on Stock Market Prediction," *International Journal of Intelligent Engineering and Systems*, vol. 10, pp. 146-153, 2017.
- [10] D. Quackenbush, *et al.*, "Public Perceptions of Media Bias: A Meta-Analysis of American Media Outlets during the 2012 Presidential Election," *The Elon Journal of Undergraduate Research in Communications*, vol. 4, pp. 51-65, 2013.
- [11] R. K. Garrett, *et al.*, "Driving a Wedge between Evidence and Beliefs: How Online Ideological News Exposure Promotes Political Misperceptions," *Journal of Computer-Mediated Communication*, vol.1, pp. 331-348, 2016.
- [12] Y. Al Amrani *et al.*, "A Novel Hybrid Classification Approach for Sentiment Analysis of Text Document," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 8, pp. 4554-4567, 2018.
- [13] Chaudhary Jashubhai Rameshbhai and Joy Paulose E. Lloret, *et al.*, "Towards a unified framework for opinion retrieval, mining and summarization," *Journal of intelligent Information Systems*, vol. 39, pp.711-747, 2012.
- [14] S. Doumit, *et al.*, "Online News Media Bias Analysis using LDA-NLP Approach," Proceedings of the International Conference on Complex Systems, pp. 249-263, 2011.
- [15] Md. Mohd Suhairi, *et al.*, "Modified framework for sarcasm detection and classification in sentiment analysis," *Indonesian Journal of Electrical Engineering and Computer Science (IJECS)*, vol. 13, pp. 1175-1183, 2019.
- [16] D. Abimanyu, *et al.*, "Indonesian online travel agent sentiment analysis using machine learning methods," *Indonesian Journal of Electrical Engineering and Computer Science (IJECS)*, vol. 14, pp. 117-122, 2019.
- [17] Kohei Watanabe, "Measuring News Bias: Russia's Official News Agency ITAR-TASS' Coverage of the Ukraine Crisis", *European Journal of Communication*, vol. 32, pp. 224– 241, 2017.

- [18] Wayne R. Dunham, "Framing the Right Suspects: Measuring Media Bias", *Journal of Media Economics*, vol. 26, pp. 122-147, 2013.
- [19] Jianwei Zhang *et al.*, "Sentiment Bias Detection in Support of News Credibility Judgment", Hawaii International Conference on System Sciences, pp. 1-10, Jan 2011.
- [20] Suhuan Sun *et al.*, "Polarity Words Distance-Weight Count for Opinion Analysis of Online News Comments", *Procedia Engineering*, pp. 1916-1920, 2011.
- [21] Plaban Kumar Bhowmick *et al.*, "Classifying Emotion in News Sentences: When Machine Classification Meets Human Classification", *International Journal on Computer Science and Engineering*, vol.2, pp.98-108, 2010.
- [22] Yoonjung Choi *et al.*, "Coarse-Grained +/-Effect Word Sense Disambiguation for Implicit Sentiment Analysis", *IEEE Transactions on Affective Computing*, vol. 8, pp.471-479, 2017.
- [23] SV. Shri Bharathi *et al.*, "Sentiment Analysis for online stock market news using RSS feeds", *International Journal of Current Engineering and Scientific Search*, vol.4, No.4, 2017.
- [24] Hanshi Wang *et al.*, "Feature-based Sentiment Analysis Approach for Product Reviews", *Journal of Software*, vol. 9, pp. 274-279, 2014.
- [25] Monireh Ebrahimi *et al.*, "Challenges of Sentiment Analysis for Dynamic Events", *IEEE Intelligent Systems*, vol. 32, pp. 70-75, 2017.

BIOGRAPHIES OF AUTHORS



Mrs. SV. Shri Bharathi is a research scholar in the Department of Computer Science and Engineering at B S Abdur Rahman Crescent Institute of Science and Technology, Chennai, TamilNadu, India. Her research interest includes Data Mining, Knowledge Discovery, Information Extraction, and Machine Learning. She has published more than 10 International / National Journals which are indexed by Scopus and Web of Science. Currently she is working as Assistant Professor in the Department of Computer Science and Engineering, Saveetha School of Engineering at Saveetha Institute of Medical and Technical Sciences, Chennai, Tamil Nadu, India.



Dr. Angelina Geetha is working as a Professor in Department of Computer Science and Engineering at Hindustan Institute of Technology and Science, Chennai, TamilNadu, India. Her research interest includes Data Analytics, Data Mining, Information Retrieval, Knowledge Management, Web mining. She has authored/ co-authored more than 20 journals and conference papers, most of which are indexed by Scopus and ISI Web of Science. She also served on the program and organizing committees of numerous national and international conferences.