



## Uncovering flaming events on news media in social media

Praboda Rajapaksha, Reza Farahbakhsh, Noel Crespi, Bruno Defude

### ► To cite this version:

Praboda Rajapaksha, Reza Farahbakhsh, Noel Crespi, Bruno Defude. Uncovering flaming events on news media in social media. IPCCC 2019: 38th International Performance Computing and Communications Conference, Oct 2019, Londres, United Kingdom. pp.1-9, 10.1109/IPCCC47392.2019.8958759 . hal-02363443

**HAL Id: hal-02363443**

**<https://hal.science/hal-02363443>**

Submitted on 14 Nov 2019

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Uncovering Flaming Events on News Media in Social Media

Praboda Rajapaksha, Reza Farahbakhsh, Noel Crespi, Bruno Defude

► **To cite this version:**

Praboda Rajapaksha, Reza Farahbakhsh, Noel Crespi, Bruno Defude. Uncovering Flaming Events on News Media in Social Media. International Performance Computing and Communications Conference, Oct 2019, London, United Kingdom. hal-02363443

**HAL Id: hal-02363443**

**<https://hal.archives-ouvertes.fr/hal-02363443>**

Submitted on 14 Nov 2019

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Uncovering Flaming Events on News Media in Social Media

Praboda Rajapaksha<sup>\*†</sup>, Reza Farahbakhsh<sup>\*</sup>, Noël Crespi<sup>\*</sup>, Bruno Defude<sup>\*</sup>

<sup>\*</sup>*Institut Polytechnique de Paris, Telecom SudParis, CNRS Lab UMR5157, Evry, France.*

<sup>†</sup>*Uva Wellassa University, 90000, Badulla, Sri Lanka.*

{praboda.rajapaksha, reza.farahbakhsh, noel.crespi, bruno.defude}@telecom-sudparis.eu

**Abstract**—Social networking sites (SNSs) facilitate the sharing of ideas and information through different types of feedback including publishing posts, leaving comments and other type of reactions. However, some comments or feedback on SNSs are inconsiderate and offensive, and sometimes this type of feedback has a very negative effect on a target user. The phenomenon known as flaming goes hand-in-hand with this type of posting that can trigger almost instantly on SNSs. Most popular users such as celebrities, politicians and news media are the major victims of the flaming behaviors and so detecting these types of events will be useful and appreciated. Flaming event can be monitored and identified by analyzing negative comments received on a post. Thus, our main objective of this study is to identify a way to detect flaming events in SNS using a sentiment prediction method. We use a deep Neural Network (NN) model that can identify sentiments of variable length sentences and classifies the sentiment of SNSs content (both comments and posts) to discover flaming events. Our deep NN model uses Word2Vec and FastText word embedding methods as its training to explore which method is the most appropriate. The labeled dataset for training the deep NN is generated using an enhanced lexicon based approach. Our deep NN model classifies the sentiment of a sentence into five classes: Very Positive, Positive, Neutral, Negative and Very Negative. To detect flaming incidents, we focus only on the comments classified into the Negative and Very Negative classes. As a use-case, we try to explore the flaming phenomena in the news media domain and therefore we focused on news items posted by three popular news media on Facebook (BBCNews, CNN and FoxNews) to train and test the model. The experimental results show that flaming events can be detected with our proposed approach, and we explored main characteristics that trigger a flaming event and topics discussed in the flaming posts.

**Index Terms**—Flaming detection, Sentiment analysis, deep neural networks, social media, Facebook, News media, FastText, Word2Vec.

## I. INTRODUCTION

In modern Internet parlance, cyberbullying has become increasingly common, especially in the Internet communities and SNSs such as Facebook, Twitter, Instagram and YouTube. In this era, several new phenomena have appeared, leading to important debates and discussions and among them one hot topic is ‘Flamings’. The Flaming can be considered as a serious issue on SNSs where many users express disagreements, insults or offensive words in the form of comments on a forum, blog or chat room intended to inflame emotions and sensibilities of others. These comments do not contribute any useful content to the discussion groups and instead attempt to wound another person socially and psychologically.

These comments might be posted by genuine users and fraud or spam generated content [1]. Two examples of flaming include 1) ‘Delete your account’: a Clinton-Trump Twitter flame war<sup>1</sup> and 2) flame war between Donald Trump and Pope Francis on the popes calling Trump ‘disgraceful’ for his immigration recommendations<sup>2</sup>. As a result of this high-level visibility, flaming has become an interesting topic among social researchers as they seek to understand the phenomena and explore the impact of these types of activities on targets.

One use-case we target in this study is news media in Facebook as many American adults consume news on social media and majority of them are commonly use Facebook [2]. In addition we can observe that the number of fans in news media Facebook pages<sup>3</sup> are considerably higher similar to other categories. News media tend to publish news items of interest to more diversified and varied readers, and therefore many news readers interact with news items daily via commenting, sharing and reacting. A set of news media is relying on the content of most popular news media as content consumers [3], [4] and so, it is important to understand these flaming types of events in the news media domain. Thus, our main objective is to explore news items’ flaming events in terms of negative feedback with insults and other offensive words. The existence of flamings on news items may reduce number of followers, or sometimes these types of posts can go viral and increase the number of followers. Therefore detecting flamings and identifying the topics that the community most strongly disagree with is a useful conception for news media.

Sentiment polarity prediction is one of the main ways of detecting flaming events in SNS [5], [6]. Many advancements have been made to the sentiment classification methods to date. However, these methods are domain-specific and therefore their results are strongly biased on the words used in that domain. As a result, we build a word embedding-based multiclass (5-classes) sentiment classification deep NN-based approach by focusing on Facebook news items and variable length user feedbacks that can be modified and applied in any other social media category other than news. In addition, we use an improved lexicon-based sentiment classification method to generate a true labeling list with which to train the deep NN

<sup>1</sup><https://www.sbs.com.au/news/delete-your-account-clinton-trump-in-twitter-flame-war>

<sup>2</sup><https://www.lifewire.com/what-is-flaming-2483253>

<sup>3</sup><https://fanpagelist.com/category/news/>

model. The flaming effect analyses will be done based on the comments classified as Negative and Very Negative. A flaming event takes place when many users give negative feedback, and so a post with a large number of negative comments received within a short time will possibly a flaming event. We will also explore what types of topics were mainly affected by flamings that are published by BBCNews, CNN and FoxNews.

This paper offers the following contributions:

- 1) a word embedding-based sentiment prediction deep NN model focused on Facebook comments.
- 2) an exploration of which word embedding method (Word2Vec or FastText) works better on Facebook comments.
- 3) identification of flaming posts on BBCNews, CNN and FoxNews Facebook pages published in February 2018.
- 4) identification of flaming posts' associated topics.

## II. LITERATURE SURVEY

Researchers have investigated flamings on YouTube [7], email threads [8] and comments received on Twitter and Facebook [5], [6], news sites and news channels [9]. These works show that flaming events can be appeared in any online platforms especially in the SNSs as any user can comment on a public content. Sentiment polarity prediction is one of the main ways of detecting flaming events in SNSs [5], [6].

Sentiment analysis of social media content has become more and more popular, as it can be used for mining opinions on services, products, companies, etc. and these models can be implemented as supervised, unsupervised or semi-supervised [10] methods. However, with the increase of user-generated content in SNSs it is not effective to apply lexicon-based unsupervised methods, and therefore supervised methods can be automated to detect polarity. Apart from that, analyses on sentiment classification are aspect-based and domain-specific. Therefore, model needs to train with domain specific data to achieve better accuracy when building a sentiment prediction model. Recently, it has been widely acknowledged that deep learning-based representation models have achieved great success in text sentiment classification tasks compared to traditional machine learning models [11]. Furthermore, in recent works, word embedding-based method are applied for sentiment classification. A few have used Word2Vec embeddings [12], [13]. Deep learning has emerged as a powerful machine learning technique and is also popularly used in sentiment analysis in recent years [14], [15]. Wang et al. proposed a CNN-RNN architecture [16] to analyze the sentiment of short text while some other studies tried apply methods based on CNN [17] and RNN [18]. Their experimental results shown that the proposed method outperforms lexicon-based, regression-based, and obtained an obvious improvement upon the state-of-the-art. In addition, many research works considered only the binary sentiment classification and few studies have used multi-class sentiment classification producing promising results [19].

Hence, in this work we try to build a deep NN model combining CNN and RNN to classify sentiment of variable length text using word embeddings that can be generalized as a semi-supervised model to be adapted to any domain specific

TABLE I  
STATISTICS OF PUBLIC POSTS OBTAINED FROM BBC, CNN, AND FOXNEWS FACEBOOK PAGES DURING FEBRUARY 2018. (PP REPRESENTS THE ABBREVIATION FOR PRE-PROCESSING)

| News media | #Fans | Number of comments of |                    |                      |                        |
|------------|-------|-----------------------|--------------------|----------------------|------------------------|
|            |       | 300 posts Before PP   | 300 posts After PP | 100 PPposts Training | 200 PPposts Experiment |
| BBCNews    | 46.2M | 398,453               | 312,881            | 107,874              | 205,007                |
| CNN        | 29.9M | 595,268               | 312,881            | 217,386              | 288,606                |
| FoxNews    | 16.3M | 1,162,734             | 312,881            | 280,342              | 773,439                |

dataset. Also, the model will consider multi-class labels to train and experiment with a real datasets.

## III. EXPERIMENTS

News media flaming events can be detected by identifying negative comments received on shared news items in SNSs. Therefore, sentiment prediction models can be adapted to cluster senses of the user feedbacks in order to explore the existence of flaming events in news media in SNS using a rule based technique. Many previous sentiment analysis works have used Twitter as the SNS to analyze and build their sentiment prediction models. As Facebook employ different content properties such as permitting to share variable length posts and comments, unlike in Twitter, we build our methodology by proposing a sentiment prediction model focusing on news items shared on Facebook. Most efficient sentiment analysis algorithms are supervised learning which requires sufficient amount of training data. Hence, we introduce a deep-NN based supervised sentiment prediction model that require considerable number of true labels to train. The labeled dataset is generated using an unsupervised model which is an improved unsupervised sentiment classification model. In the following sections we explain our sentiment classification deep NN-based approach in detail.

### A. Dataset description and pre-processing

In this study, we try to explore existence of the flaming events in news media in Facebook as news media plays a major role to promote and distribute their news items in Facebook [2]. Therefore, this study targets on top three news media (BBCNews, CNN and FoxNews) in Facebook having the highest number of fans as a usecase to explore flamings.

We implemented a crawler to collect random 300 public posts and related comments of BBCNews, CNN and FoxNews in February 2018 using Facebook Graph API. To respect the ethical aspects, we collected only the texts of public posts, comments and timestamps from their Facebook accounts and do not collected neither sensitive data nor personal data. The analyses are mainly based on this dataset and a brief description of the dataset is presented in Table I. The training and experimental datasets are separated after pre-processing the original news items and user feedback.

We applied basic preprocessing techniques such as replace URL with a space, remove user mentions, hashtags, retweets, special characters and multiple spaces, removed multiple letters from a word ('haaappy' will be replaced by 'happy')

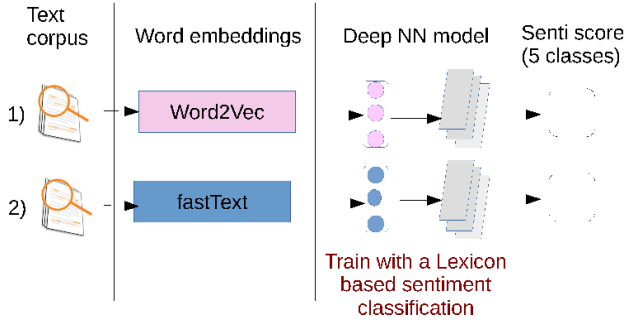


Fig. 1. The architecture of the multiclass sentiment prediction using word embeddings

,merge characters of a word written with spaces or dots in between each character ('h a p p y', 'h.a.p.p.y' will be reformatted to happy). As the objective of this work is to identify the sentiment of SNSs content, emojis impact a lot on the overall sentiment value. Therefore, we included all the positive and negative emojis in the text without ignoring. We also performed stemming on each word in the sentence to generate the stem. Lemmatization is not applied as lemmatizers have to search through a corpus while stemmers do basic string operations and thus is faster than lemmatizers. Also, stemming can work with unknown words while lemmatizer do not recognize them. We feed to the neural network separately the datasets after being lemmatized and stemmed, and the accuracy of the model was very low with the lemmatized dataset. We do not remove stopwords in sentences while pre-processing.

Considerable amount of comments have been removed after preprocessing and the main reason was that the majority of those comments were URLs. We can observe from the Table I that news items published by FoxNews has received a higher number of comments, while having less number of fans, than BBCNews and CNN. We use comments of 100 pre-processed posts from each news media as a training dataset to train the NN model and comments of the remaining 200 news posts from each news media as the experimental dataset to explore flaming events.

### B. Proposed sentiment analysis architecture

The architecture of our variable length text sentiment prediction model is shown in Figure 1. At first, as depicted in Figure 1, we build a lexicon-based approach to generate the training dataset for our deep NN model. The technique we used in the lexicon-based approach is explained in detail in Section III-C. Many previous studies have tried to employ word-embedding methods for machine learning or deep learning-based sentiment classification [15], [12], [13]. In our prediction model we experiment with two different word embedding methods: Word2Vec and FastText. Word2Vec is widely used for both shallow and deep neural networks and was developed by Google in 2013 [20]. In the Word2Vec model, words with similar meaning are mapped to a nearby vector space, thereby making it simple to explore semantically similar words. As the Word2Vec model is based on a predefined dictionary, one

shortcoming of this representation is that rare words may not be mapped with other vectors. However, FastText was developed by Facebook [21] in 2016 using the n-gram representation of each word in a sentence and therefore helps to detect misspelled words and slang across different languages. In the experiments we will explore which word embedding method is more suitable for sentiment classification of the Facebook content.

The deep NN model shown in Figure 1 uses the best performed word embedding method to predict the sentiment of a sentence based on a multiclass (5 classes) classification technique, where five different classes can be defined as; Very Positive(4), Positive(3), Neutral(2), Negative(1) and Very negative(0). The accuracy of the model can be improved by re-training the model with experimental datasets by increasing the labeled dataset occasionally as a semi-supervised approach.

### C. Methodology for generating true labels

The deep NN sentiment prediction model need to use a set of true labels to train the model. We introduce an improved lexicon-based approach to identify the sentiment of Facebook content and then use this model to build the training corpus for a deep NN model to automatically classify the sentiments. In this approach we can customize sentiment labels to be 2-class (positive, negative), 3-class (positive, neutral, negative), or 5-class(very positive, positive, neutral, negative, very negative) depending on the requirement. This section explains this unsupervised lexicons-based sentiment classification approach.

The consideration of negators, adverbs and model verbs associated with words is important for sentiment classification. Therefore it is worth to identify important sentiment of words and word phrases separately including various negators, models, and degree adverbs, as well as their combinations. Kiritchenko et al. [22] created a lexicon dataset, Sentiment Composition Lexicon for Negators, Models, and Degree Adverbs (SCL-NMA) and it includes 3207 phrases with related sentiment scores where each phrase contains at most 4 words. Sentiment value of a phrase ranges from -1 to 1 where phrases with negative meanings have assigned negative values and phrases with positive meanings have assigned positive values. The higher the meaning of a phrase is positive the more it is closer to 1 and similarly, sentences having very negative meanings have assigned score very closer to -1.

We used an statistical based approach to identify the polarity of a sentence in English. A sentence may have one or many lexicons found in the SCL-NMA dataset and also both negative and positive lexicons. Hence, we explored all the lexicons in the sentence to classify its overall sentiment. This approach is very useful to identify the sentiment of variable length sentences and following equation shows the prediction of sentiment of a new sentence. Assume that SCL-NMA dataset consists of lexicons  $L_1, L_2, \dots, L_x$ .

$$SentiScore = \frac{[\sum_{n=1}^N Ln] + [C + S + E]}{[N + C + S + E]} \quad (1)$$

TABLE II  
CONFUSION MATRIX OF OUR LEXICON-BASED SENTIMENT  
CLASSIFICATION METHOD (MACRO PRECISION-60.66%, MACRO  
RECALL-62.01%, F1 SCORE - 61.31%)

|        |     | Predicted |        |        | Precision |
|--------|-----|-----------|--------|--------|-----------|
|        |     | Pos       | Neg    | Neu    |           |
| Actual | Pos | 128       | 19     | 35     | 70.33%    |
|        | Neg | 57        | 83     | 37     | 46.89%    |
|        | Neu | 37        | 12     | 90     | 64.75%    |
| Recall |     | 57.66%    | 72.81% | 55.56% |           |

TABLE III  
CONFUSION MATRIX OF VADAR SENTIMENT CLASSIFICATION (MACRO  
PRECISION-37.85%, MACRO RECALL-51.41%, F1 SCORE - 43.59%)

|        |     | Predicted |        |        | Precision |
|--------|-----|-----------|--------|--------|-----------|
|        |     | Pos       | Neg    | Neu    |           |
| Actual | Pos | 24        | 12     | 156    | 12.5%     |
|        | Neg | 0         | 15     | 162    | 8.46%     |
|        | Neu | 1         | 1      | 137    | 92.57%    |
| Recall |     | 70.56%    | 53.57% | 30.11% |           |

Where N is the number of lexicons in the sentence. Letters C, S, and E are constant and set to be 1, -1 or 0 depending on the lexicon polarity. These constant values are added to the SentiScore formula based on the existence of different properties of the lexicons in a sentence as explained below.

The overall sentiment of a sentence is increased when it has lexicons (found in the SCL-NMA dataset) with capital letters. Therefore, we identify those lexicons in capital letters and use these details in Equation 1, the letter C, to evaluate the overall sentiment score.

$$C = \sum capital\_lexicons \quad (2)$$

where  $capital\_lexicons=1$ , if the lexicons' in capital letters has a positive sentiment value,  $capital\_lexicons=-1$ , if the lexicons' in capital letters has a negative sentiment value and  $capital\_lexicons=0$ , if no capitalized lexicons in the sentence.

Emojis in a sentence are affecting more on the sentiment score. Thus, E in Equation 1 refers to all emojis in the sentence those that helps to increase the strength of the sentiment.

$$E = \sum emojis \quad (3)$$

where  $emojis=1$  for a positive emoji,  $emojis=-1$  for a negative emoji and  $emojis=0$  if no emojis presence in a sentence.

People use exclamation marks to stress words and this is an important property when evaluating sentiments. Thus, we observe the presence of exclamation marks in a sentence and this feature is presented in Equation 1 as S.

$$S = exclamation\_mark \quad (4)$$

where  $exclamation\_mark=1$  if a positive lexicon is attached to exclamation mark,  $exclamation\_mark=-1$  if a negative lexicon is attached to exclamation mark, and  $exclamation\_mark=0$  if no any exclamation mark is attached to lexicons.

The algorithm consider n-grams as the features, where n ranges from 1 to 3 and generate features separately for lexicon dataset and sentences that need to identify its sentiments. In summary, Equation 1 helps to identify the sentiment of

any sentence written in English regardless of the length of the sentence. This approach helps to generate a labeling dataset from user feedback received on news items shared on Facebook in order to train our deep NN model .

#### D. Comparing unsupervised true lable generation method with a baseline method

VADAR lexicon and rule-based sentiment analysis tool [23], which is widely used for classifying social media content, is considered as the baseline method to compare our unsupervised true label generation model. For a given sentence, VADAR returns its sense from 3 classes: positive, negative or neutral. Therefore, in order to compare our algorithm with VADAR, our model is customized to classify sentence sentiment in to the same 3 classes.

We use a manually annotated dataset having 498 sentences with its sentiment values: 177 negative tweets, 182 positive tweets and 139 neutral tweets published by the Stanford University [24]. This dataset is applied on both VADAR and our unsupervised sentiment classification technique and their confusion matrices are shown in Table III and Table II, respectively. The precision, recall and F1 score parameters of our approach is much higher than VADAR for the classification of positive and negative sentiments as detailed in Table III and Table II. The results is proven that our lexicon based approach can classify the sentiment label of a sentence much better than VADAR. However, we can still improve the accuracy of our model by increasing the lexicon dataset with additional lexicons.

This approach is unique as our training dataset for the deep NN can be generated automatically, as opposed to having manual annotate Facebook posts and comments. Moreover, this model is introduced as a multiclass sentiment classification approach rather than classifying only into 2 or 3 classes. In this model, we assigned a sentiment label for each sentence from five classes as follows.

$$Prediction = \begin{cases} VeryPositive, IF(sentiScore \geq 0.5) \\ Positive, IF(0.5 > sentiScore > 0) \\ Neutral, IF(sentiScore = 0) \\ Negative, IF(0 > sentiScore > -0.5) \\ VeryNegative, IF(sentiScore \leq -0.5) \end{cases} \quad (5)$$

The baseline value for distinguishing Very Positive from Positive and Very Negative from Negative classes is set to be 0.5 and -0.5 respectively.

#### E. Methodology for an automatic sentiment prediction model

Since deep learning-based sentiment classification models are better than the traditional machine learning models [11] we try to adapt one deep learning-based model in this work to predict sentiments of Facebook content. These deep models can learn text representation from original data that can capture relations between contexts words in a scalable manner.

Recent research on text sentiment classification used convolution operations using n-gram features. However, the convolution neural network (CNN) completely ignores the sequence

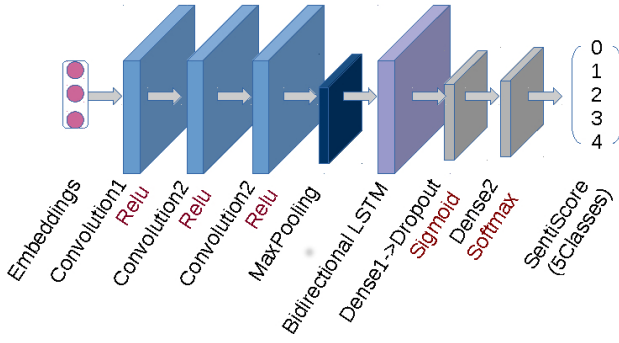


Fig. 2. The deep model for extracting sentiment scores

information of the text while paying attention to the local features of a sentence. On the other hand, long short-term memory (LSTM) networks are good at learning sequential correlation in the text by using an adaptive gating mechanism but lost the ability to learn local features of the text. Therefore, to involve both sequence correlation information and local features, it is important to explore an effective combination strategy that takes advantage of the CNN and the LSTM network. Hence, we experiment with several different architectural methods such as: i) CNN-LSTM-Word2Vec : pre-trained Word2Vec vectors, ii) CNN-LSTM-Word2Vec : custom Word2Vec model train with our own data, iii) CNN-LSTM-Fasttext : pre-trained Fasttext vectors, iv) CNN-LSTM-Fasttext : custom Fasttext model tran with our own data, and v) Adjusting the number of CNN layers and dropout layers, and maxpooling layer.

Models are trained separately with the custom Word2Vec and FastText embedding methods using 315K total number of comments (experimental dataset in Table I) with their sentiment labels generated from our lexicon-based method. The training dataset consist of 42.3%-neutral, 20.5%-very positive, 16.6%-positive, 12.9%-very negative and 7.7%-negative labels. And also these models are trained with both pre-trained word embedding. Most promising results with high accuracy is given to the architecture shown in Figure 2 and achieved high accuracy with our own domain specific word embedding model rather than using pre-trained embedding.

As shown in Figure 2, model was implemented with three CNN layers and a Bi-LSTM layer. The inputs of the NN model uses 1D convolutions. One dropout layer was added within two dense layers (hidden layers) close to the output layer, and another dropout layer on the Bi-LSTM layer. The model compilation is done by adjusting three parameters: loss, optimizer and metrics. The Adam [25] optimizer was used in our model to adjust the learning rate throughout the training and the learning rate was fixed to be very small (0.0001) leading to obtain more accurate results. The categorical\_crossentropy loss function was used in our implementation. This model uses three activation functions: Relu- for CNN layers, Softmax- for the output layer and Sigmoid- for the dense layer. The method of dealing with variable length sentences is that, at first we need to set a maximum number of token values (30 in this research work) and then if a comment has more than 30 tokens,

we divided it into sub-sentences and evaluate the sentiment of each sub-sentence separately. Finally, the sentiment score for the entire sentence is the average sentiment scores of sub-sentences.

At first, deep NN model is trained for 40 epochs and evaluated the accuracy and validation loss for each iteration separately for Word2Vec and FastText. In the Word2Vec model, we analyzed that at the 13th epoch, the performance of the training dataset continue to decrease than the validation dataset, indicating an over-fitting. Therefore, the best-fit for this model is to training with 12 epoch as the model has good skill on both the training dataset and unseen test dataset. The accuracy of the model is identified as 85%. Similarly, when the FastText model used in the deep NN model, it shows the best fit after 15th epoch, in which training and validation loss is almost equal and with 78% accuracy. In both scenarios, total number of parameters trained by the NN model is 3,050,053.

The results manifested that, Word2Vec model out performs the FastText model when applying for multiclass sentiment classification of Facebook content. As a result, we will use our generated models with Word2Vec embedding to explore the existence of any flaming events in news media in Facebook.

#### IV. ANALYSIS OF FLAMING EVENTS IN BBCNEWS, CNN AND FOXNEWS IN FACEBOOK

In social media, relatively large number of critical comments can be directed at individuals, companies, brands and etc. and this behavior is called as flaming. One way of exploring flaming behavior is to identify whether a post received higher number of very negative comments within a short time. This section explains, how the implemented deep NN sentiment prediction model can be used to identify the existence of flaming or similar kind of event in news media in Facebook.

##### A. Experimental dataset

In order to analyze flaming events and their existence in news media in Facebook, we randomly selected comments from 200 posts (Table I and Figure 3) shared by BBCNews, CNN and FoxNews in February 2018 as our experimental dataset. First, we try to explore sentiment labels for each comment in the experimental dataset using our deep NN model which is trained with the Word2Vec model as Word2Vec model given a high accuracy than the FastText model (Section III ).

Figure 3 shows the statistics of the sentiment prediction of the comments belongs to 600 posts in our experimental dataset (200 posts from each BBCNews, CNN and FoxNews). As indicated in Figure 3, number of Neutral comments received on these posts are much higher than the other classes. The second most number of sentiment predictions of comments are belongs to the Positive class, followed by Very Positive and Very Negative. The least number of predicted sentiment values are from the Negative class. In general, number of Neutral comments received on the posts are always higher than the total number of positive comments (Very Positive + Positive) and total number of negative comments (Very Negative + Negative). We observed that the text classified into a Neutral



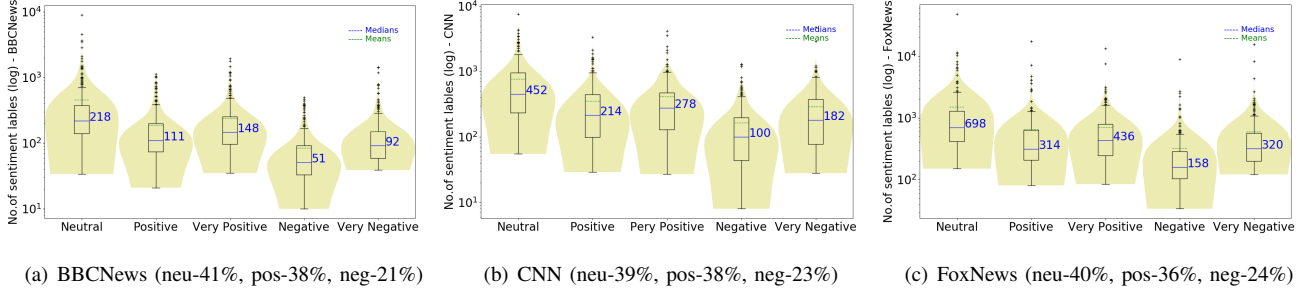


Fig. 3. Five class sentiment value distribution of 200 random posts shared by BBCNews, CNN and FoxNews during February 2018.

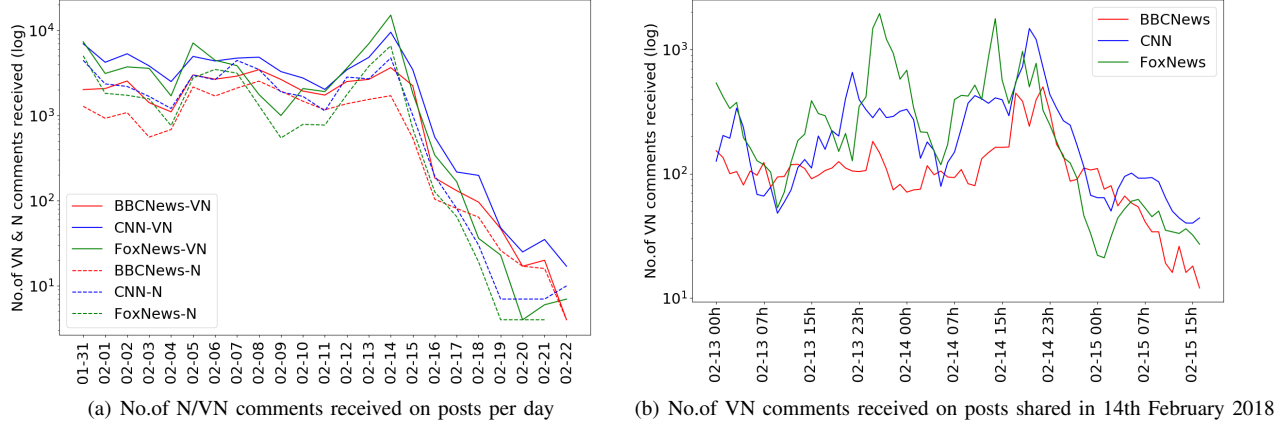


Fig. 4. The distribution of Negative (N) and Very Negative (VN) comments received on posts shared by BBCNews, CNN and FoxNews in February 2018.

class have a large number of stop-words, names, URLs, and single words without indication of any feelings.

### B. Flaming events in news media - time-series approach

Receiving avalanches of negative emotions or flood of low happiness comments from users within short period of time tend to generate flaming events and a time-series approach can be one of the optimal ways to explore it. Figure 4 exhibits the time-series distribution of received Negative and Very Negative comments in our experimental dataset. Figure 4(a) exhibits negative comments received per day during February 2018. As shown in Figure 4(a), number of Very Negative comments is always higher than the number of Negative comments and therefore we focus only on Very Negative comments for further analysis. We can observe that, on 14th February 2018, a flood of negative comments have received on posts shared by all three news media compared to the other days. Apart from that, 31st January, 2nd and 5th February exhibited another flaming type of distributions showing few other spikes of number of Very Negative comments.

As shown in Figure 4(a), the main flaming type of event taken place in 14th February. Therefore, to identify more information about this event the consideration of the posts shared during one day before and after 14th February is also important. Figure 4(b) shows this distribution in terms of the Very Negative comments received during these 3 days. We can observe that all three news media have shared the flaming posts

in 14th February and received a large number of Very Negative comments but, FoxNews has shown a flaming type of event on 13th February where some of the comments on these posts have received on 14th February as well. One main inspection from these figures indicate our hypothesis that the flaming type of posts received Very Negative comments within short duration of time (approximately 2-3hrs) after posting news.

Figure 4(a) and 4(b) indicate varied values of number of comments received on 14th February. We explored that on 14th February, BBCNews, CNN, FoxNews have published only 14.5% posts, 19% posts and 16.5% posts respectively out of 200 posts. Next we identified how many posts have received a higher number of negative comments than the other types of comments. The statistics shown that only 26.6% posts from BBCNews, 34.2% from CNN and 30.3% from FoxNews have received more negative comments than the other types of comments on 14th February. Hence, as per this analysis, only a set of posts contributed to the flaming event on 14th February 2018 while other comments have received on the posts shared by the other days. Next section will explain more about flaming event posts and their discussed topics based on one statistical approach.

### C. Flaming events in news media - statistical approach

The analysis of comments based on the time-series data has proven an existence of flaming event in BBCNews, CNN and FoxNews. Identification of types of the posts of these flaming



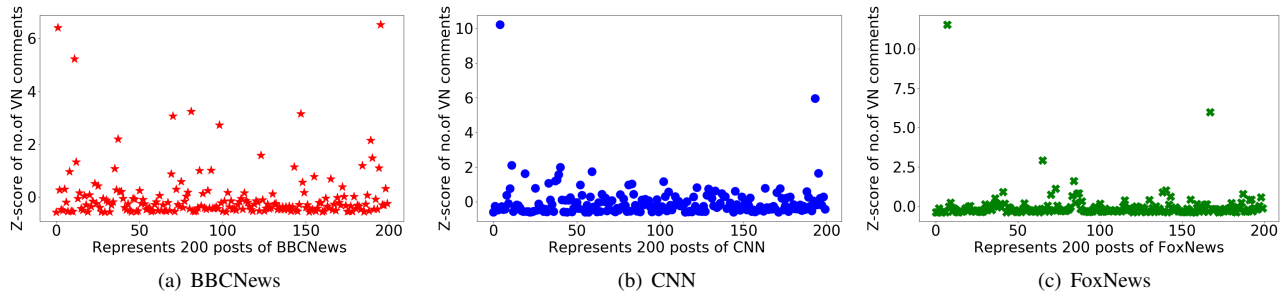


Fig. 5. The distribution of z-score values of number of Very Negative (VN) comments received per post (200 posts in total).

TABLE IV  
STATISTICS OF THE FLAMING TYPES OF POSTS SHARED BY BBCNews, CNN AND FoxNews (VN STANDS FOR VERY NEGATIVE).

| News media | Date       | #Reactions | #Comments(VN)   | Topics   |
|------------|------------|------------|-----------------|--|
| BBCNews    | 2018-02-01 | 13590      | 4110 (44.28%)   | Palestinian teenager slapped an Israeli soldier.   |
|            | 2018-02-02 | 33110      | 3946 (37.56%)   | A father has tried to attack a doctor who abused his three daughters.  |
| CNN        | 2018-02-14 | 28718      | 2614 (63.54%)   | Shooting at a Florida high school.   |
|            | 2018-02-02 | 113632     | 19598 (30.99%)  | A father has tried to attack a doctor who abused his three daughters.  |
| FoxNews    | 2018-02-14 | 39611      | 4670 (73.17%)   | Shooting at a Florida high school.   |
|            | 2018-01-31 | 86717      | 130185 (20.61%) | House Democratic Leader Nancy Pelosi and other House Democratic leaders hold a news conference.  |
|            | 2018-02-13 | 72765      | 35619 (29.41%)  | The women of 'The View' took a shot at Vice President Mike Pence's Christian faith on Tuesday, mocking the former governor of Indiana for talking to Jesus and even calling it a 'mental illness'. |

events are useful and therefore in this section we will explore more details about the posts that attracts flaming comments.

At first, we try to identify these flaming posts with the use of one statistical approach that can predict these types of behaviors considering an outliers prediction method where outliers in this scenario are the posts those that received many insults or offensive comments. The standard score (z-score) is one outlier prediction method which allows to identify whether a particular value is equal to the mean, below the mean or above the mean. We calculate z-score for a distribution of number of Very Negative comments received on each posts within February 2018 using following equation where  $\mu$  and  $\sigma$  represents mean and standard deviation, respectively. For each post we assign 'x', the number of Very Negative comments.

$$z = \frac{x - \mu}{\sigma}$$

Figure 5 exhibits the distribution of z-score values of Very Negative comments received on posts in our experimental dataset. As shown in Figure 5(a), BBCNews has three outliers having z-score value greater than 5. Figure 5(b) exhibits two outliers for CNN with the z-score above 6 and as shown in Figure 5(c), FoxNews has two outliers having z-score greater than 6. Next, we will analyze the popularity of these 7 outliers posts and their discussed topics.

Table IV contains the information on above identified 7 outliers including published date, total number of reactions received, number of comments received, number of Very Negative comments (VN) and the discussed topics. We observed that, these posts are widely popular and have received enormous amount of reader reactions (#reactions) compared to

the other posts in the dataset. We also observed that these posts have received huge amount of comments and more of them are Very Negative (#comments). In addition, all 7 posts have received more than 20% of the comments as Very Negative.

An interesting observation of this analysis is to explore the types of topics discussed by each news media in these outlier posts. Table IV exhibits that BBCNews and CNN have posted similar topics on i) 2nd February - about a father was trying to kill a doctor who abused his daughters and ii) 14th February - about an incident of shooting at a Florida high school. In addition, BBCNews has an extra flaming event on iii) 1st February - about one Palestinian teenager slapping an Israeli soldier. On the other hand, two flaming posts published by FoxNews are different to what BBCNews and CNN have shared; on iv) 31st January - type of a political discussion and v) 13th February - related to a religious discussion. In summary, we identified only 5 posts associated to the 7 outliers detected in the previous section. Therefore, we can conclude that, as exhibited in Figure 4 31st January and 2nd, 13th, 14th February exhibited flaming types of news posts with the analysis of a statistical method. The topics discussed on these flaming posts are mainly very sensitive information, political related details or related to religious matters. We can deduce from these information that, people try to send more aggressive and negative comments on sensitive news items.

These results conclude that we can detect flaming events in news media in Facebook using a sentiment prediction approach, and we can apply this analysis to other types of categories in Facebook such as celebrities, politicians, etc. Moreover, if the SCL-NMA dataset can be improved by adding more hateful and insult wordings with their sentiment scores,

we can detect flaming events with higher accuracy. Our deep NN model can also be used to adapt to other use-cases, as it is a general model to detect the sentiments of textual content from five different classes. The proposed model can be enhanced to identify flaming wars in SNSs as the model we developed can be automated and behave as a semi-supervised approach. The comments received on flaming posts might be shared by real users or might be Spam generated content [1]. Therefore, it is important to identify these patterns of posting flaming events based on commentator's behavioral properties.

## V. CONCLUSION

This paper is mainly focused on the detection of flaming types of events in news media in Facebook. We proposed a deep Neural Network (NN) model to predict sentiment polarity of SNSs comments based on word embedding. We also proposed an improved lexicon-based multiclass sentiment classification method to generate true labels for our NN model. We experimented with two word embedding models: Word2Vec and FastText and found that Word2Vec model performed better (85% accuracy) than FastText model (78%). The model returns five sentiment classes; very negative, negative, neutral, positive and very positive. As a use-case, we explored some existing flaming types of events in BBCNews, CNN and FoxNews Facebook pages. A flaming event can be present in SNSs when a post receives many negative comments withing a short period of time. However, we identified that, these news media received a higher number of neutral comments than other categories. First, we explored flaming behavior based on a time-series analysis of the comments received on posts during February 2018 and a major flaming event is identified on 14th February 2018. Then, using one statistical based approach we justified the existence of flaming types of incidents in news items that were identified from the time-series analysis. The results shown that, in news media, flaming types of events are more common when they published very sensitive information (murders, rape etc.), political based contents and information related to religious belief.

## REFERENCES

- [1] S. Shehnepoor, M. Salehi, R. Farahbakhsh, and N. Crespi, "Netspam: A network-based spam detection framework for reviews in online social media," *IEEE Transactions on Information Forensics and Security*, vol. 12, no. 7, pp. 1585–1595, 2017.
- [2] Pwe, "News use across social media platforms 2018," <http://www.journalism.org/2018/09/10/news-use-across-social-media-platforms-2018>, 2018, [Online; accessed 20-March-2019].
- [3] P. Rajapaksha, R. Farahbakhsh, and N. Crespi, "Scrutinizing news media cooperation in facebook and twitter," *IEEE Access*, 2019.
- [4] P. Rajapaksha, R. Farahbakhsh, N. Crespi, and B. Defude, "Inspecting interactions: Online news media synergies in social media," in *2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*. IEEE, 2018, pp. 535–539.
- [5] S. Ozawa, S. Yoshida, J. Kitazono, T. Sugawara, and T. Haga, "A sentiment polarity prediction model using transfer learning and its application to sns flaming event detection," in *2016 IEEE Symposium Series on Computational Intelligence (SSCI)*. IEEE, 2016, pp. 1–7.
- [6] S. Yoshida, J. Kitazono, S. Ozawa, T. Sugawara, T. Haga, and S. Nakamura, "Sentiment analysis for various sns media using naïve bayes classifier and its application to flaming detection," in *2014 IEEE Symposium on Computational Intelligence in Big Data (CIBD)*. IEEE, 2014, pp. 1–6.
- [7] P. J. Moor, A. Heuvelman, and R. Verleur, "Flaming on youtube," *Computers in human behavior*, vol. 26, no. 6, pp. 1536–1546, 2010.
- [8] A. K. Turnage, "Email flaming behaviors and organizational conflict," *Journal of Computer-Mediated Communication*, vol. 13, no. 1, pp. 43–59, 2007.
- [9] F. Mungeam and H. Crandall, "Commenting on the news: How the degree of anonymity affects flaming online," *Unpublished thesis, MA Program in Communication and Leadership Studies, Gonzaga University*, pp. 1–38, 2011.
- [10] F. H. Khan, U. Qamar, and S. Bashir, "A semi-supervised approach to sentiment analysis using revised sentiment strength based on sentiwordnet," *Knowledge and information Systems*, vol. 51, no. 3, pp. 851–872, 2017.
- [11] A. Gatt and E. Krahmer, "Survey of the state of the art in natural language generation: Core tasks, applications and evaluation," *Journal of Artificial Intelligence Research*, vol. 61, pp. 65–170, 2018.
- [12] A. H. Ombabi, O. Lazzez, W. Ouarda, and A. M. Alimi, "Deep learning framework based on word2vec and cnn for users interests classification," in *2017 Sudan Conference on Computer Science and Information Technology (SCCSIT)*. IEEE, 2017, pp. 1–7.
- [13] X. Ge, X. Jin, and Y. Xu, "Research on sentiment analysis of multiple classifiers based on word2vec," in *2018 10th International Conference on Intelligent Human-Machine Systems and Cybernetics (IHMSC)*, vol. 2. IEEE, 2018, pp. 230–234.
- [14] L. Zhang, S. Wang, and B. Liu, "Deep learning for sentiment analysis: A survey," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 8, no. 4, p. e1253, 2018.
- [15] M. U. Salur and İ. Aydin, "Sentiment classification based on deep learning," in *2018 26th Signal Processing and Communications Applications Conference (SIU)*. IEEE, 2018, pp. 1–4.
- [16] X. Wang, W. Jiang, and Z. Luo, "Combination of convolutional and recurrent neural network for sentiment analysis of short texts," in *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*, 2016, pp. 2428–2437.
- [17] C. Dos Santos and M. Gatti, "Deep convolutional neural networks for sentiment analysis of short texts," in *Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers*, 2014, pp. 69–78.
- [18] X. Wang, Y. Liu, S. Chengjie, B. Wang, and X. Wang, "Predicting polarities of tweets by composing word embeddings with long short-term memory," in *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, vol. 1, 2015, pp. 1343–1353.
- [19] Y. Liu, J.-W. Bi, and Z.-P. Fan, "Multi-class sentiment classification: The experimental comparisons of feature selection and machine learning algorithms," *Expert Systems with Applications*, vol. 80, pp. 323–339, 2017.
- [20] T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado, and J. Dean, "Distributed representations of words and phrases and their compositionality," in *Advances in neural information processing systems*, 2013, pp. 3111–3119.
- [21] A. Joulin, E. Grave, P. Bojanowski, M. Douze, H. Jégou, and T. Mikolov, "Fasttext. zip: Compressing text classification models," *arXiv preprint arXiv:1612.03651*, 2016.
- [22] S. Kiritchenko and S. M. Mohammad, "The effect of negators, modals, and degree adverbs on sentiment composition," *arXiv preprint arXiv:1712.01794*, 2017.
- [23] C. J. Hutto and E. Gilbert, "Vader: A parsimonious rule-based model for sentiment analysis of social media text," in *Eighth international AAAI conference on weblogs and social media*, 2014.
- [24] A. Go, R. Bhayani, and L. Huang, "Twitter sentiment classification using distant supervision," *CS224N Project Report, Stanford*, vol. 1, no. 12, 2009.
- [25] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.