5-2019

Identifying Fake News using Emotion Analysis

Brady Gilleran

Identifying Fake News using Emotion Analysis

An Undergraduate Honors College Thesis
in the
Department of Computer Science
College of Engineering
University of Arkansas
Fayetteville, AR

By

Brady Gilleran

bjgiller@uark.edu

May 2019

University of Arkansas

Abstract

This paper presents research applying Emotion Analysis to “Fake News” and “Real News” articles to investigate whether or not there is a difference in the emotion used in these two types of news articles. The paper reports on a dataset for Fake and Real News that we created and the natural language processing techniques employed to process the collected text. We use a lexicon that includes predefined words for eight emotions (anger, anticipation, disgust, fear, surprise, sadness, joy, trust) to measure the emotional impact in each of these eight dimensions. The results of the emotion analysis are used as features for machine learning algorithms contained in the Weka package to train a classifier. This classifier is then used to analyze a new document to predict/classify it to be “Fake” or “Real” News.

Keywords – emotion analysis, machine learning, classifier, fake news

Table of Contents

1. Introduction

1.1 Background

Fake news is a type of yellow journalism or propaganda that consists of deliberate disinformation or hoaxes spread via traditional print and broadcast new media or via social media [1]. It is usually written to mislead people or persuade them into believing something is true or having them act a certain way. This has caused people to have total confusion and little understanding about important social and political issues, caused students to propagate misinformation in research assignments, and it can even cause issues with physical health due to fake information about treating diseases or illnesses [2].

Because of the serious consequences of inaccurate information, it is so important to be able to classify this misinformation and isolate it to prevent all the harmful things that Fake news can cause. Sites like Snopes.com [3] and FakeCheck.org [4] have done a great job to help this issue. However, people need a faster way of checking articles since it takes time before the articles are submitted to either site, then a human must verify its validity. Our goal is to address this problem by exploring the feasibility of automatically identifying misinformation, specifically fake news, via its emotional content. We will address this as a classification problem, developing a system capable of labeling news articles as either "real" or "fake".

For us to be able to begin classification, the first step is to have features on which to base our classification. For our work, we begin by performing emotional analysis on a labelled text collection. Measuring emotion and sentiment is a good way for groups to provide feedback on a topic or new idea. Many companies and interest groups have started investing in this to see how their products and services are performing so they can better market towards their customers and therefore increase their profits. We base our approach on information gained from Dr. Susan Gauch as well as surveys [5] of research using different Natural Language Processing (NLP) strategies to analyze the emotion of text.

1.1 Goals

Our research has three main goals. First, we will create a document collection of both real and fake news sources and perform emotional analysis on the two classes of documents in the collection. We will then perform statistical analysis on the average of the two corpus's and see if there are any differences between the two types of news.

Second, we will use the results of the emotion analysis as features in a Weka classifier [6] for machine learning purposes. We will perform 5-fold validation to train the classifier and test on the other folds to see whether we can classify a news article is fake or real based on its emotion vector.

Finally, we will assess the successfulness of the classification and discuss possible reasons for the results, hoping to provide on how to improve the features or the data to have a more accurate machine classifier in the future.

2 Related Work

2.1 Emotion Analysis

Emotion analysis refers to a recent extension of work in Sentiment Analysis of text used in natural language processing. Lexicon-based approaches consist of using a pre-defined classifier to identify words within a text that are associated with a given emotion. These words, and their associated emotion weights (if any), are accumulated to calculate the prevalence of each emotion within the text. Based on work by [7], we analyzed the presence of eight commonly used emotions, i.e., Anger, Anticipation, Disgust, Fear, Joy Sadness, Surprise, and Trust. In the lexicon we used [7], each word's emotion value for each emotion is binary not scalar. For instance, the word "reward" has a 1 in anticipation, joy, surprise, and trust. So, reward will contribute to those emotion scores whenever it is encountered in a snippet of text.

The emotion analysis process is typically done by retrieving the emotion score of every word in the document and adding that to the emotional score for the document or document snippet. This creates a raw Emotion Score that is then normalized based on length of the snippet to create an 8-dimension emotion vector for each document.

Emotion analysis has most frequently been used to rate tweets [8]. More recently, research has moved forward to include Emoticons (Emoji's) in the calculation of the Emotion Score [9]. There are also efforts to include analysis tone of voice as a feature contributing to the emotion score for audio recordings [10].

2.2 Machine Learning Classifiers

Machine Learning Classifiers use "features" to classify items into different categories based on the values of the data in relation to the feature. In a two-dimensional space, the classifier would be to draw a line and hopefully have all the data from one class be on one side of the line and all data from the other class fall on the other side. Then given a point, you could classify it based on its relation to the line/value on the graph. Obviously, as classifiers work with more features, the topography necessary to separate the class instances becomes more complex and it becomes increasingly difficult to cleanly separate the classes. There are many classification techniques available based on vector space models, e.g., k-Nearest Neighbors and

Support Vector Machines, probability theory, e.g., Naïve Bayes, and logistic regression, e.g., Ada Boost, among many others.

Weka (Waikato Environment for Knowledge Analysis) is a powerful machine learning toolkit developed by the University of Waikato in New Zealand [6]. Its popularity has increased due to it being free under GNU General Public License and its ease of use on multiple operating systems due to using Java as a back end for the algorithms. It includes many of the most popular classification algorithms. This broad collection of classifiers allows users to try different built-in algorithms to find some of the patterns that humans cannot see and choose the best classifier for a given problem.

2.3 Fake News

Many people have researched classification of Fake News using other features as a means of training the classifier. Anthony Chan on Kaggle.com [11] reported on their attempt to classify news as fake or real based on the keywords that they contained. They used three different inputs for their classifier: title, content, and publication. They then classifiers trained on the extracted keywords to train their classifier to predict whether the article was fake or real.

Aaron Edell wrote on Medium.com about another Fake News Classifier.[8] Their initial research employed sentiment analysis, however that produced poor results. Then, they built a model based instead of detecting fake news based on keywords and achieved much higher accuracy. They built a program called FakeBox, that had a machine learning model that analyzed the way the article was written and told you if it was like another article written. Therefore, if it was like Real News. The keys to identifying Real News was “little to no biased words, strong adjectives, opinion, or colorful language.” Their success seemed to point the way to using emotion analysis as classification features.

3 Implementation and Methods

3.1 Emotion Score System

3.1.1 Preprocessing

The first step to creating the Emotion Score of a document, it to perform Natural Language Processing and prepare it to be processed by the Emotion Score Algorithm. This includes removing all punctuation, down casing all letters, and separating all the words into individual tokens so we that can look up the emotional of each individual word in our lexicon.

This word processing turns this article from the in text written article to a linked list of strings, each node containing a word to be scored.

Real News Article	Linked List of Words
Alex Jones Apologizes for Promoting 'Pizzagate' Hoax Alex Jones a prominent conspiracy theorist and the host of a popular right-wing radio show has apologized for helping to spread and promote the hoax known as Pizzagate. The admission on Friday by Mr. Jones the host of "The Alex Jones Show" and the operator of the website Infowars was striking. In addition to promoting the Pizzagate conspiracy theory he has contended that the Sept. 11 attacks were inside jobs carried out by the United States government and that the 2012 shooting at Sandy Hook Elementary School in Newtown Conn. was a hoax concocted by those hostile to the Second Amendment. The Pizzagate theory which posited with no evidence that top Democratic officials were involved with a satanic child pornography ring centered around Comet Ping Pong a pizza restaurant in Washington D.C. grew in online forums before making its way to more visible venues including Mr. Jones's show. And its prominence after the election drew attention to the proliferation of false and misleading news much of it politically charged that circulated on platforms like Facebook Twitter and YouTube.	*[alex, jones, apologizes, for, promoting, pizzagate, hoax, alex, jones, a, prominent, conspiracy, theorist, and, the, host, of, a, popular, rightwing, radio, show, has, apologized, for, helping, to, spread, and, promote, the, hoax, known, as, pizzagate, the, admission, on, friday, by, mr, jones, the, host, of, the, alex, jones, show, and, the, operator, of, the, website, infowars, was, striking, in, addition, to, promoting, the, pizzagate, conspiracy, theory, he, has, contended, that, the, sept, attacks, were, inside, jobs, carried, out, by, the, united, states, government, and, that, the, shooting, at, sandy, hook, elementary, school, in, newtown, conn, was, a, hoax, concocted, by, those, hostile, to, the, second, amendment, the, pizzagate, theory, which, posited, with, no, evidence, that, top, democratic, officials, were, involved, with, a, satanic, child, pornography, ring, centered, around, comet, ping, pong, a, pizza, restaurant, in, washington, dc, grew, in, online, forums, before, making, its, way, to, more, visible, venues, including, mr, joness, show, and, its, prominence, after, the, election, drew, attention, to, the, proliferation, of, false, and, misleading, news, much, of, it, politically, charged, that, circulated, on, platforms, like, facebook, twitter, and, youtube]*

Table 3.1: Alex Jones article with Corresponding Individual Tokens

The article on the right is now ready to be scored for each of the eight emotions based on their appearance in the lexicon.

3.1.2 Emotion Scoring

Once we have the linked list of words, we use a lexicon of emotion words to identify the emotions associated with each word. For this paper, we used a lexicon built by the National Research Council of Canada.[7]. It contains 14,182 unique words, each labeled with 8 emotion values (either 0 or 1).

Step one is to create an Emotion object. This object is an 8-dimensional array with a string attribute. The string contains the word from the lexicon and the 8-dimensional array stores the emotion score (either 0 or 1). Each object as it is created is put in a Linked List of Emotions to be used to cross-reference in the document.

Step two is to is to check our Linked List of tokens from the previous step. We do this by looping over every token with every word in our lexicon. Then if the token matches the word in the lexicon, the 8-dimensional array is incremented by one for each emotion that is displayed. The algorithm for every document is as follows.

```
Loop over Linked List of Tokens {
        Loop over all Emotion Object words in Lexicon {
                        If Token = LexiconObject.Word{
                                Increment value of array location of Emotion that is displayed
        }       }       }
```

Figure 3.1 Emotion Scoring Algorithm

[alex, jones, apologizes, for, promoting, pizzagate, hoax(anger, disgust, sadness, surprise), alex, jones, a, prominent, conspiracy(fear), theorist, and, the, host, of, a, popular, rightwing, radio, show(trust), has, apologized, for, helping, to, spread, and, promote, the, hoax(anger, disgust, sadness, surprise), known, as, pizzagate, the, admission, on, friday, by, mr, jones, the, host, of, the, alex, jones, show(trust), and, the, operator, of, the, website, infowars, was, striking, in, addition, to, promoting, the, pizzagate, conspiracy(fear), theory(anticipation, trust), he, has, contended, that, the, sept, attacks, were, inside, jobs, carried, out, by, the, united(trust), states, government(fear), and, that, the, shooting(anger, fear), at, sandy, hook, elementary, school(trust), in, newtown, conn, was, a, hoax(anger, disgust, sadness, surprise), concocted, by, those, hostile(anger,disgust,fear), to, the, second, amendment, the, pizzagate, theory(anticipation, trust), which, posited, with, no, evidence, that, top(anticipation, trust), democratic, officials, were, involved, with, a, satanic(anger), child(anticipation, joy), pornography(disgust), ring, centered, around, comet, ping, pong, a, pizza, restaurant, in, washington, dc, grew, in, online, forums, before, making, its, way, to, more, visible, venues, including, mr, joness, show(trust), and, its, prominence, after, the, election, drew, attention, to, the, proliferation, of, false, and, misleading(anger, disgust), news, much, of, it, politically, charged, that, circulated, on, platforms, like, facebook, twitter, and, youtube]

Figure 3.2: Alex Jones article with Emotional Words Highlighted with Corresponding Emotion

After looping through the document, we now have our 8-dimensonal Array with the Raw Emotion Score for the document. Below is the raw score of the Alex Jones article. If you look up at Table 3.2, you can see that the highlighted words correspond with the scores below.

Anger	Anticipation	Disgust	Fear	Joy	Sadness	Surprise	Trust
7	4.0	6.0	5.0	1.0	3.0	3.0	8.0

Table 3.3: Raw Alex Jones Article Emotion Score Vector

Once that is done, we divide each element by the size of the sum of the values in the Linked List to normalize them by document length. We then multiplied the results by 1,000 for readability. Below is a resulting example of the final of score of the Alex Jones article listed above. Based on this data, a quick observation would reveal that the highest score was Trust and the lowest score was Joy.

Anger	Anticipation	Disgust	Fear	Joy	Sadness	Surprise	Trust
38.25	21.86	32.79	27.32	5.46	16.39	16.39	43.72

Table 3.4: Normalized Alex Jones Article Emotion Score Vector

3.1.3 Fake News Classification

Now we need to use the features calculated from the Emotional Score Module to train a machine learning classifier to predict which articles are real and which articles are fake. For this, we used a classifier package called Weka or the Waikato Environment for Knowledge Analysis developed at The University of Waikato in Hamilton New Zealand [4].

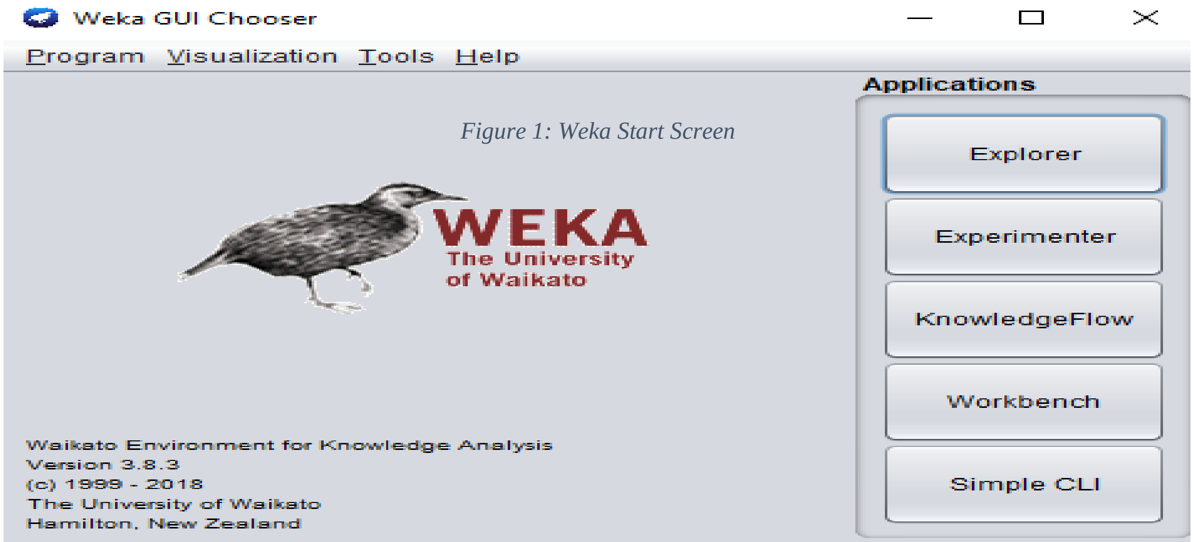

Table 3.5: Weka Classifier Program

Weka requires the data to be trained to be in a certain format (called arff) that contains the information for all the features as well as the classification for every piece of data in the dataset. An example is show below.

```
@RELATION realVsFakeNews

@ATTRIBUTE anger	REAL
@ATTRIBUTE anticipation REAL
@ATTRIBUTE disgust REAL
@ATTRIBUTE fear	REAL
@ATTRIBUTE joy	REAL
@ATTRIBUTE sadnessREAL
@ATTRIBUTE surprise		REAL
@ATTRIBUTE trust	REAL
@ATTRIBUTE class	{FakeNews,RealNews}

@DATA
2.8436018957,3.3175355450,1.8957345972,2.8436018957,1.4218009479,0.9478672986,0.9478672986,4
.7393364929,FakeNews
0.0000000000,0.0000000000,0.0000000000,5.9405940594,0.0000000000,0.0000000000,0.0000000000,1
.9801980198,FakeNews
```

Figure 3.6: Arff file format for features

After we import the file, you can select on which features you would like to train the classifier, and you can see some information about the distribution of each feature in your dataset for each of the classes you have. This interface is called the Weka Explorer.

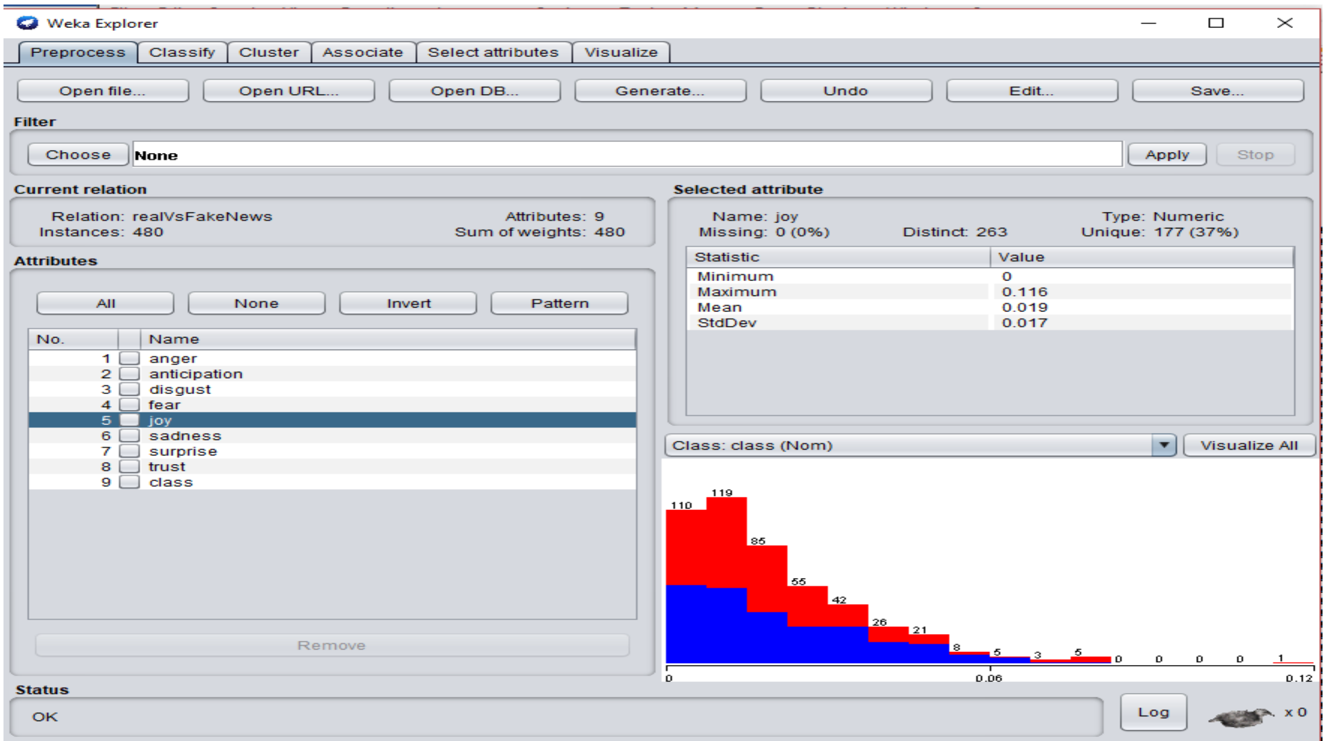

Figure 3.7: Weka Explorer

After exploring the data, we can select the Classify tab in order to configure the classifier for our experiment. Weka provides a wide selection of classifiers. We decided to test my module using two different classifiers: Adaboost and Support Vector Machine.

AdaBoost, short for “Adaptive Boosting”, is the first boosting algorithm proposed in the literature[12]. It focuses on classification problems and aims to convert a set of weak classifiers into a strong one[12]. The final equation for classification can be represented as

$$F(x) = sign(\sum_{m=1}^{M} \theta_m f_m(x)),$$

where f_m stands for the m_{th} weak classifier and Θ_m is the corresponding weight.[7] It is exactly the weighted combination of M weak classifiers. [12]

A Support Vector Machine (SVM) is a discriminative classifier formally defined by a separating hyperplane[13]. In other words, given labeled training data (supervised learning), the algorithm outputs an optimal hyperplane which categorizes new examples[13]. In a two-dimensional space, this hyperplane is a line dividing a plane in two parts where in each class lay in either side[13].

4 Results and Analysis

4.1 Emotional Distribution in Dataset

The Fake News dataset was collected from the Michigan Institute for Data Science [14]. The dataset was comprised of two different directories each with 240 text files each containing a single Fake and Real News article. In order to process them all, we collected all the Fake Articles and Real articles into their own respective text files and performed the text processing detailed in 3.1.1. We then performed the emotion scoring described in Chapter 3 for each of the two subclasses in our data set to determine whether there was any difference. Table 4.1 presents the mean values for the emotion scores for each of the news article classes, based on the 240 articles in the class, and the percent difference between the two classes.

Emotion Feature	Real News	Fake News	(Fake vs Real)
Anger	126.068	138.161	8.75%
Anticipation	265.32	285.503	7.07%
Disgust	69.0142	78.9964	12.64%
Fear	164.842	171.111	3.66%
Joy	178.338	201.642	11.56%
Sadness	125.061	135.133	7.45%
Surprise	142.205	163.584	13.07%
Trust	422.341	424.358	0.48%

Table 4.1: Mean Feature Difference between Real and Fake News

As you can see, for every single emotion, there was an observable increase in the Emotion Score for the Fake News versus Real News. The biggest difference observed was in the Disgust, Joy, and Surprise scores. This supports our belief that Fake News uses more Emotion than Real News does. This supports our hypothesis of the very nature of Fake News. While Real News usually tries to logically inform you about current issues, Fake News attempts to misinform you by tying into your emotions to make you react in a certain way. This makes people bypass the many inaccuracies that are usually in these articles. The reason Trust values are so similar in each is due to the similar nature of both types of news. News articles try to show themselves as a credible news source; Otherwise, there would be no reason for you to believe it and be sucked into the article, whether true or false. Since there is a difference in the means, when train our model using the classifier, we hope to see some accurate classification results.

4.2 Machine Learning Classifier

Using 240 examples for each class, we trained Adaboost and Support Vector Machine classifiers using a 5-fold validation. In other words, we used 384 articles for training the classifier (80%), evenly split between Fake and Real News, and 96 for testing, also evenly split. Table 4.1 summarizes our initial results, which were frankly disappointing.

	Correctly Classified	Percent Correct
Adaboost	246/ 480	51.25 %
Support Vector Machine	233/480	48.54 %

Table 4.2: Initial Classification Comparison: Adaboost vs. SVM

From this, we can set that Adaboost worked better by around 3 percent, although this is still close to random. To see if we could improve the results, we investigated what happens when we removed the least important feature, as identified by Weka’s Explore module, Trust. The lack of importance of Trust to the classifier is also supported our data from 4.1 which showed that Trust had a less than 0.5% difference between Fake and Real News. Table 4.2 presents the results after Trust is removed.

	Correctly Classified	Percent Correct
Adaboost	253/480	52.71 %
Support Vector Machine	243/480	50.63 %

Table 4.3: Classification Comparison: Adaboost vs. SVM without Trust Feature

As you can see, by removing the poorest feature, we can create a more accurate classifier that can better predict data be removing bad distributions. However, it is till not where we need it to be. It is our suspicion that, although we have a clear difference in the means between the two classes for all features, this difference is small (generally less than 10%). So, it may be necessary to have much more training data or to differentially weight the features so that the stronger features carry more weight in the classification.

5 Conclusion

5.1 Conclusions

The Emotional Score program we developed can determine the Emotional Score of the articles of Fake and Real News from the dataset. There is a clear difference in means between the emotion features of Anger, Anticipation, Disgust, Fear, Joy, Sadness, and Surprise, although there is little difference in Trust. When the features are used to train a classifier in Weka, if there is not enough data and/or the differences in the means are drowned out by the standard deviation, so that we achieve classification results that are barely better than random. However, it does improve slightly when Trust is removed.

5.2 Future Work

Future work should focus on increasing the dataset to provide more examples on which to train the classifier. This should provide more accurate classification. Future work should also pursue combining the Emotion Score feature into current and new ways of classifying Fake and Real News. There is already current work using other features for classification such as keywords, number of capitalized words, word order, etc. Combining those features with the emotion analysis presented in this paper should help those classifiers become better as well targeting Fake News before it spreads and does further harm to society.

6.0 References

[1] LibGuides: Fake News and Alternative Facts: Finding Accurate News: Why is Fake News Harmful? https://researchguides.austincc.edu/c.php?g=612891&p=4258046

[2] Tufekci, Zeynep (January 16, 2018). "It's the (Democracy-Poisoning) Golden Age of Free Speech". Wired.

[3] Snopes.com Fact checking site for Online websites through users
https://www.snopes.com/https://www.snopes.com/

[4] FactCheck.org Fact checking site for Online websites through the Government
https://www.snopes.com/https://www.snopes.com/

[5] Manzoor Hakak, Nida & Mohd, Mohsin & Kirmani, Mahira & mohd, Mudasir. (2017). Emotion analysis: A survey. 397-402. 10.1109/COMPTELIX.2017.8004002.

[6] Weka 3: Machine Learning Software in Java
https://www.cs.waikato.ac.nz/~ml/weka/

[7] Crowdsourcing a Word-Emotion Association Lexicon, Saif Mohammad and Peter Turney, Computational Intelligence, 39(3), 555-590, 2013.

[8] I trained fake news detection AI with >95% accuracy, and almost went crazy. (2019). – Aaron Edell
https://towardsdatascience.com/i-trained-fake-news-detection-ai-with-95-accuracy-and-almost-went-crazy-d10589aa57c

[9] A global Analysis of Emoji Usage – Ljubvsic, Nikola – 2016, Berlin Germany
https://www.aclweb.org/anthology/W16-2610

[10] Hybrid Approach for Emotion Classification of Audio Conversation Based on Text and Speech Mining – Jasmine Bhaskar, K. Sruthi, Prema Nedgungadi – 2015

[11] Fake News Classifier – Anthony Chan via Kaggle.com – 2018
https://www.kaggle.com/anthonyc1/fake-news-classifier-final-project

[12] SauceCat Boosting algorithm: Adaboost – Medium user SauceCat - 2017
https://towardsdatascience.com/boosting-algorithm-adaboost-b6737a9ee60c

[13] Chapter 2: SVM (Support Vector Machine) – Theory – Savan Patel – 2017
https://medium.com/machine-learning-101/chapter-2-svm-support-vector-machine-theory-f0812effc72

[14] Automatic Detection of Fake News: Veronica Perez-Rosas, Bennett Kleinberg, Alexandra Lefevre, Rada Mihalcea - International Conference on Computational Linguistics (COLING) – 2018