Factors that Affect Emotion Elicited from News Readers

Peigen Xu and **Charibeth Cheng**
De La Salle University, Manila, Philippines
peigen_xu@dlsu.edu.ph and chari.cheng@delasalle.ph

Abstract

Content providers today are no longer just in the business of providing information. They now also want to know what their audience think and feel. This is true for news agencies which usually provide several avenues for its readers to express their opinions and sentiments. These reader-generated content contains information that may be useful to content providers. This research uses data from the sentiment solicitation tools of Rappler [1], a social news network based in the Philippines. We explored different document representations, feature combinations, feature selection, and use of affective information in building a multi-class classification model, that identifies emotions elicited from readers when they read a news article. The performance of the classifiers were examined using SVM, NB, and k-NN learning algorithms. We also analyzed the relevant word features of each emotion, and the relevance between the emotion classes. We discovered that there is a relationship between the emotion elicited from readers and the theme of the article.

1 Introduction

Sentiment analysis is the computational study of opinions, sentiments and emotions in text. While traditional sentiment analysis focuses on the sentiment or emotion in the text, our work is focused on the emotions that is being elicited by a news article from readers [Medhat *et al.*, 2014].

Rappler is a Philippine-based online news that actively uses social media platforms for news distribution. It allows readers to share their "mood" towards an article through its Mood Meter feature.

To answering the question, "How does the story make you feel?", readers choose from a pre-defined set of emotions, namely: *happy*, *sad*, *angry*, *don't care*, *inspired*, *afraid*, *amused*, and *annoyed*. Each article in Rappler would then have multiple labelled emotions crowd-sourced from their readers. We used this data in creating a multi-label emotion classifier. Figure 1 shows a sample Mood Meter report of an article.

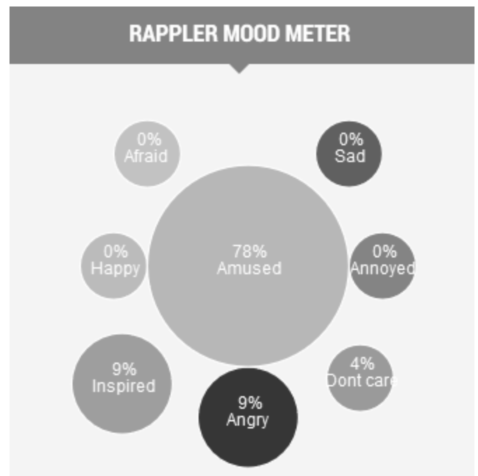

Figure 1: Rappler Mood Meter

Multi-label classification has been applied to different domains. The multi-label k Nearest Neighbor (ML-kNN) algorithm was used in three different real-world multi-label learning problems, namely: yeast gene functional analysis, natural scene classification, and automatic web page categorization [Zhang and Zhou, 2007]. The same algorithm was also used by Bhowmick *et al.* to perform sentence level emotion classification.

This paper presents a computational model that identifies the emotions elicited from readers given an English news article. The paper is organized as follows: Section 2 describes related work on emotion classification using different language resource, features, and algorithm; Section 3 describes our dataset; Section 4 discusses the methodology; Section 5 discusses the experiment results; lastly, Section 6 discusses the conclusions.

2 Related Work

Emotion detection researches have been done in the past decades. The work of Shivhare *et al.* used emotion ontology built by Parrott, together with a weighting algorithm that deals with the depth of ontology and parent-child relationship

[1] http://www.rappler.com

of the document to calculate the emotion class. The ontology is built with the definition of schemas and aspects such as entities and attributes, the relationship between the entities, and the domain vocabulary. The result of their research achieved an average accuracy of 79.57% from six emotion classes. The emotion detection model was tested on 135 blog posts.

The work of Yang *et al.* investigated a criteria for document level classification and presented observations in classifying the author's emotion from the content of the document. They used 5,410,933 blog posts from Taiwan Yahoo! Kimo that were tagged with emoticon representing the corresponding emotion. Three classifiers namely: Support Vector Machine (SVM), Conditional Random Field (CRF), and Bayesian Classifier with varying size of features were evaluated. They obtained the highest precision, recall and F-Score of 59.23%, 58.58%, and 58.90% respectively, classifying the document level emotion based only to the emotion gathered of the last sentence. This implies authors tend to summarize emotion of the overall document in the last sentence of the document.

Related to previous work of Yang *et al.*, Kumar and Suresh proposed a methodology for emotion detection using lexical chains. They used WordNet as the lexical database and manually collected an emotion keyword list as the knowledge base to construct the lexical chain. A lexical chain may denote semantic relationships such as synonym, hyponym, meronym, and holonym. 1660 news articles were tested, the precision, recall and F-measure of 96%, 80%, and 86% respectively. These results outperformed the work of Yang *et al.*. However, 340 news articles were not included in the test due to missing emotion centric words in the lexical database, and some of the news article didn't contain sufficient content for the algorithm to form strong lexical chains.

To further examine the performance of a reader's perspective emotion classifier, Lin *et al.* performed several experiments to show that certain feature combinations can achieve relatively good accuracy. They examined five features from the news articles namely: Chinese character bigram, segmented Chinese word, metadata of the news articles (news category, agency, hour of publication, reporter, event location), affix similarity (similarity between a news article and an emotion class), and emotional word. They collected Chinese news articles from Yahoo! Kimo News, where 25,975 for training, and 11,441 were used for testing. Results of using SVM to the feature combinations showed that the combination of all five features performed the highest accuracy of 76.88%. They found out that affix similarity and bigram are the two notable features that capture some important emotion details that cannot be expressed by other features.

A news article may trigger several, different emotions from each reader, so multi-label classification algorithms were investigated to model this fact. The work of Bhowmick *et al.*, used multi-label kNN(ML-kNN) to classify reader's emotion at the sentence level, using word-based features such as word polarity and semantic frame. The semantic frame is generated from the Berkeley FrameNet project [2] that groups similar lexical units. Using 1305 sentences annotated with 4 emotion classes, they also performed x^2 feature selection to reduce the feature dimensions. Different evaluation metrics were used such as hamming loss, partial match accuracy, one-error, coverage, etc. to evaluate the result of multi-label classification. Their result suggests that the semantic frame feature improves the overall performance.

[2]http://framenet.icsi.berkeley.edu/

3 Dataset

We collected 24,229 news articles from Rappler, which were all written in English. Rappler allows readers to share the emotion they felt after reading an article as either *afraid*, *amused*, *angry*, *annoyed*, *don't care*, *happy*, *inspired* or *sad*. News articles included information such as: title, news body, topic tags, and the emotion value from the readers. For instance, the news article entitled *Duterte takes oath as 16th President of the Philippines* had the topic tags of *Duterte administration*, *Philippine president* and *Rodrigo Duterte* with readers' emotion of 80% *happy*, 10% *inspired*, and 10% *don't care*.

The collected news were written from January 2015 to September 2016. Initially, out of the 24,229 total news article crawled 10,941 were labeled *happy*, which was almost half the size of the corpus. To balance the dataset, 1,008 articles were randomly selected for each emotion labels, thus a total of 8,064 news articles were used as the main corpus in this research. Label selection methods are discussed in the next section.

4 Methodology

4.1 Document Representation

Similar to the work of [Yang *et al.*, 2007] in document-level classification, each news article is expressed in six baseline document representations:

- Title - title of the news article
- First sentence - first sentence of the news article
- Last sentence - last sentence of the news article
- Longest sentence - longest sentence of the news article
- Topic tag - tags assigned to the news article
- Whole article - content of the news article, using the Bag-of-Words technique

The document representations and their combinations were later analyzed based on their effectiveness in classifying the emotion elicited from the reader.

4.2 Feature Extraction

Case-insensitive Bag-of-Words technique was used to determine the relevant terms among the document representations. Each document representation is vectored into uni-grams and bi-grams. The term frequency-inverse document frequency (TF-IDF) value of each word feature of each document is used as the feature value. A minimum word frequency of 2 among all documents is also set to reduce the feature size. Also, stop word removal is performed before feature extraction.

4.3 Label Selection

Single Label

Each news article from Rappler has a set of eight emotion values in percentage that correspond to the percentage of readers who felt that emotion after reading the article. For single label classification, the emotion label with the highest user-selection is the label of the article. In case of a tie among multiple emotion values, one is randomly selected among the set.

Multi-label

Each news article can possibly have values for each of the 8 emotion labels. To determine the appropriate emotion labels associated with each article, we considered the following scenarios: first, by setting a specific threshold value. For instance, when the threshold is set to 30% and the emotion values of an article are: 35% *happy*, 30% *inspired*, 10% *angry*, and 25% *afraid*, then the emotion labels associated with the article would only be *happy* and *inspired*.

Second, by getting the *top N* emotion values as class labels including the emotions with the same rank. For example, getting *top 3* labels with 25% *happy*, 20% *inspired*, 20% *angry*, 20% *afraid*, 10% *sad*, 5% *don't care* will result to: *happy*, *inspired*, *angry*, and *afraid* being the labels associated with the article.

Last, by getting the *top N* emotion values excluding the emotions with the same rank. For example, getting *top 3* labels with 25% *happy*, 20% *inspired*, 20% *angry*, 20% *afraid*, 10% *sad*, 5% *don't care* will result to only one label: *happy*.

4.4 Model Building

The algorithms considered in the experiment include Naive Bayes, SVM, and k-NN. To be able to analyze the performance more accurately, each set of experiment was conducted 10 times with their average values as the metric. Before building the model, the training and testing data are randomly split into 75% and 25% respectively. Scikit-learn [Pedregosa *et al.*, 2011], a machine learning library for the Python programming language, was used to build the models. The three classification methods used in this experiment are multi-class, binary-class, and multi-label classification. The difference among the three is discussed in following sections:

Multi-Class Classification

Multi-class classification is a classification task using more than two classes; in case of this study are the eight emotion labels. Multi-class classification assumes that each instance is assigned to only one label i.e., a news article is classified to one emotion label. This is also the method used to identify the best performing document representation. The best performing document representation is used for the other two classification solutions.

Binary Classification

Binary classification is the traditional classification technique that predict the true value between two options. In the case of this study, each emotion label is used to compare against the other labels. The corpus used for this solution is rearranged by collecting 1,000 articles of the target label and another 1,000 of mix and randomly selected articles from the other labels. The resulting corpus for this solution is eight sets of documents for the eight labels.

Multi-Label Classification

Multi-label classification is a classification task that classifies each instance to a set of target labels. This can be interpreted as predicting properties of data that are not mutually exclusive. In case of this study, a news article can classified as *happy*, *inspired*, and *afraid*.

5 Discussion

5.1 Baseline Experiment

In order to verify the best document representation for news article, the multi-class classification model is used. The accuracy results in Table 1 show that among the six identified document representations, *whole article* has the highest performance followed by the *topic tag*. In contrast to the work of Yang *et al.*, the *last sentence* document representation has the lowest performance in this experiment.

Combinations of different document representations are also examined and the results are shown in Table 2. The general result is similar to the work of Lin *et al.*, with no improvement in performance.

Combining the best document representations, i.e, *whole article* and *topic tag*, yielded a slightly better performance. Combining low performing document representations (*last* + *longest sentence*) improved a little in performance but still cannot outperform any top baseline document representation. The second best performing document representation, the *topic tag*, has around 3% to 6% performance difference as compare to the *whole article*; while *topic tag* contains only around 10% of the *whole article's* feature size. In terms of algorithm, SVM outperforms NB and k-NN, and NB outperforms k-NN in majority of the experiments.

5.2 Binary Classification

Using different document representations and learning algorithms, binary classification aims to examine the classification difficulty of each class label. Figure 2 shows the average classification accuracy of each emotion label using algorithms. It is notable that *afraid* and *angry* are consistently the easiest to classify by having the highest average accuracy. *Happy* and *sad* have almost the same average accuracy, regardless in different document representation. *Don't care*, *amused*, and *annoyed* are the most difficult emotions to classify with less than 1% difference in average accuracy. In fact, the highest performing label *afraid* has almost 10% difference in accuracy versus the worst performing label *annoyed*

5.3 Multi-Label Classification

In multi-label classification, each news article can be classified to more than one emotion class. Since each new article includes the percentage value of each emotion, section 4.3.2 described the three strategies used for the label selection in this study. The evaluation metrics for the multi-label classification are compared according to the label cardinality.

Figure 3 shows the results based on coverage, ranking loss, average precision, and hamming loss from the average score

Table 1: Results of Single Document Representation

Document Representation	Feature Size	Accuracy SVM	NB	k-NN
WholeArticle	180431	45.41%	42.32%	39.97%
TopicTag	17195	39.60%	39.33%	35.46%
Title	6861	36.54%	37.72%	29.24%
FirstSentence	16681	35.92%	36.23%	31.53%
LongestSentence	21269	32.40%	33.64%	30.23%
Last Sentence	10633	26.66%	28.04%	23.77%

Table 2: Results of Combination Document Representation

Document Representation	Feature Size	Accuracy SVM	NB	k-NN
Whole + Topic + Title + First	192319	46.07%	43.01%	39.62%
Whole + Topic	187434	45.93%	42.69%	39.65%
Topic + Title + First	30778	43.40%	42.68%	38.27%
Topic + Title + Longest + First	43420	43.63%	42.45%	38.15%
Topic + Title + Longest + Last	42329	42.60%	41.91%	37.89%
Topic + Title	18750	41.50%	41.63%	37.79%
Title + First + Longest	34225	41.50%	41.42%	36.91%
Topic + Last	23185	40.52%	40.73%	37.20%
Last + Longest	28515	35.62%	36.53%	33.43%

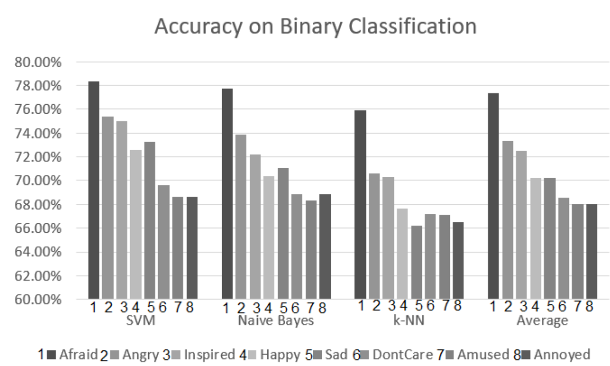

Figure 2: Binary Classification

of different algorithms with respect to the label cardinality. For threshold label selection method, as the threshold decreases the label cardinality increases, because it accepts the emotion as the class label for even minority opinions. It was unexpected that the two *top N label* selection methods yield a big difference in the resulting label cardinality. The *include tie* method yields almost double the value of *exclude tie* method; *top 3*, for instance, 4.90 and 2.30 of label cardinality were outputted from the two methods. This phenomenon suggests that certain emotion labels coexist, or are more related to each other, as can be shown by these emotion labels having the same emotion values in an article.

It is noticeable that label cardinality is one factor that affects the classifier's performance for all evaluation metrics. When the label cardinality increases, coverage also increases. Coverage measures the scope of possible labels needed to be investigated in order to get the right labels. This also applies to the average precision, whose value increases as the label cardinality increases. However, ranking loss and hamming loss were not directly proportional to label cardinality. Performance of the hamming loss seem to peak as the label cardinality reached half of the actual number of labels used. While the ranking loss also seem to peak as the label cardinality reached a quarter of the actual number of labels, it starts to drop as the label cardinality continue increases. For hamming loss, when *topic tags* were used as document representation, there was an increase from 0.1465 with 1.27 label cardinality, to 0.3348 with a 4.08 label cardinality. But when the label cardinality goes up to 4.9, the hamming loss goes down to 0.3086. In addition, as the trend line from different metrics show no perfect label cardinality that performs best among all the evaluation metrics, an ideal label cardinality would depend on the metric aimed for. For this experiment, a 4.9 label cardinality or more than half of the actual number of labels performs best in average precision, which is the metric that measures the overall performance of the multi-label classifier.

5.4 Emotion Correlation

Two analysis on emotion labels were done to examine their correlations. These include the co-occurrence between the emotion labels and word feature analysis of different emotions.

Emotion Co-occurrence Distribution

Based on the three multi-label label selection strategies discussed in Section 4.3.2; the first strategy uses 10%, 20%, and 30% threshold value, and for the second and the third strategies we get the *top 2* and *top 3* emotions. With these

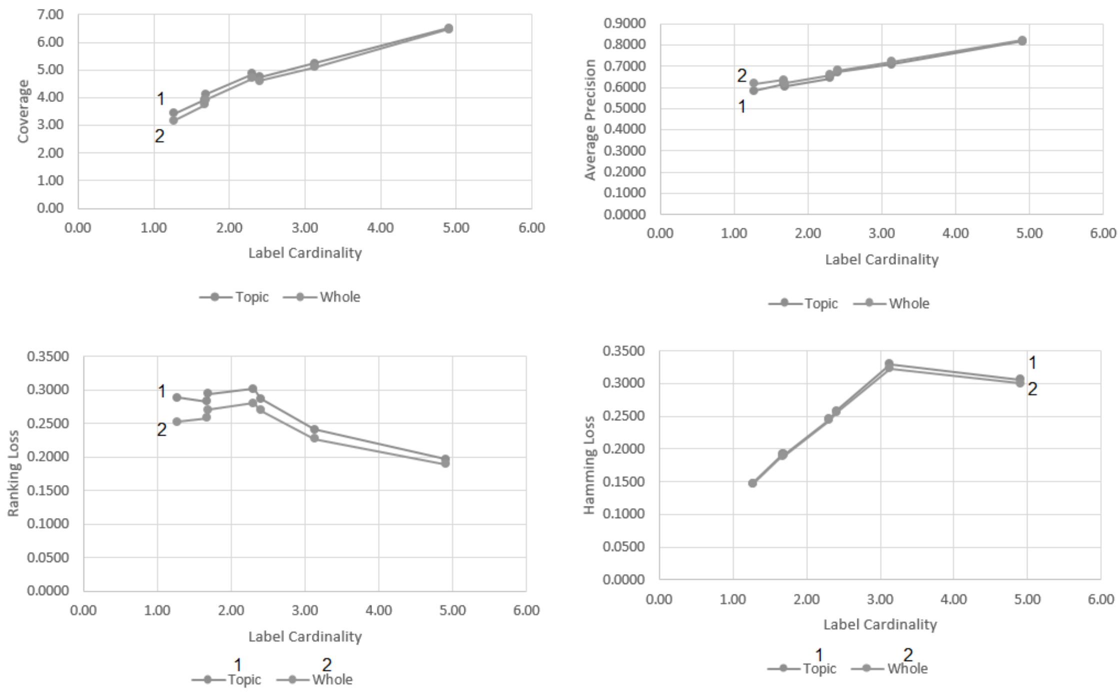

Figure 3: Multi-Label Classification Evaluation (Metrics vs Label Cardinality)

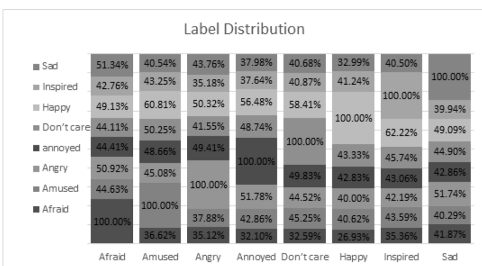

Figure 4: Emotion Distribution

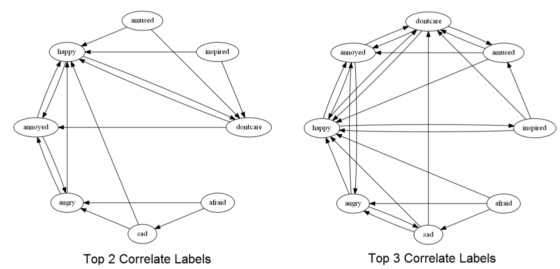

Figure 5: Balanced Corpus: Top 2 and 3 Correlate Labels

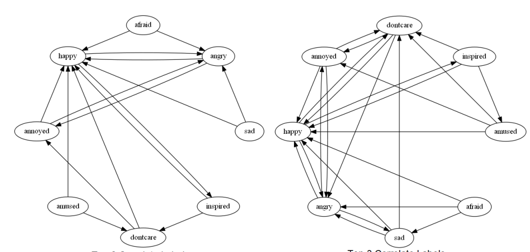

Figure 6: Unbalanced Corpus: Top 2 and 3 Correlate Labels

seven multi-label label selection strategies, Figure 4 shows the percentage of coexistence between each emotion label. The value shown in the graph is the average result from all label selection strategies. By getting the top two and three relevant labels from each label, a visualization is presented in Figure 5. From the original unbalanced dataset of 24,229 articles, the label correlation, as shown in Figure 6, is seen to be relatively close to the resulting graph using the balanced data. The result shows *happy* to be related to almost all other labels. One reason could be that *happy* has relatively high occurrence among all documents, regardless of the methods for multi-label label selection. Conversely, there are only a few of emotions related to *inspired* and *afraid*.

Word Feature Analysis

To further understand each emotion label, the word features that have high tf-idf score were examined from each emotion. The top 100 word features were analyzed and each feature was manually categorized to a theme. The theme names are mainly influenced by the news categories (e.g., entertainment, sports, science/nature, etc.) used by Rappler. For example, *Zika virus*, was manually categorized to *health/disease*. The list below describes the themes that emerged in data:

- **world politics**: country names, world leader names, ISIS related issue
- **sports**: basketball, football, boxing, sports celebrity, team names
- **politics**: local politician names, political parties
- **entertainment**: T.V shows, events, movies
- **security**: human right, security issues, terrorism
- **health/disease**: disease names, health care issues
- **science/nature**: natural disaster, weather report

Results show that themes about *election* and *politics* are associated with all emotions. It is surprisingly to see that the *religious* theme triggers not only *inspired* emotion, but also *angry*. The positive emotions of *happy* and *amused* are commonly triggered by the themes of *sports* and *entertainment*. Also, the themes about *health/disease*, *science/nature*, and *security* trigger the negative emotions of *afraid* and *sad*. In fact, there are distinct themes that only occur in one emotion; such as *health/disease* in *afraid*, *celebrity* in *don't care*, *development* in *happy*, as well as *immigration* and *deceased celebrities* in *sad*. *Annoyed* and *angry* emotion share a relatively similar themes, which are *politics*, *world politics*, *PH election*, and *local government*. Also *don't care* and *happy* share the themes: *sports*, *entertainment*, *politics*, *PH election*.

5.5 Model Application to Other News Sources

We further tested the classification models on three local events namely: the gruesome murder of a Korean in Manila, the Trump inauguration, and the national university athletics(UAAP 79) basketball championship. Comparable news articles were collected from four news sources including Rappler, GMA News (a news and public affairs program on TV with online content), Inquirer (a newspaper/broadsheet), and Tempo (a tabloid). The predicted emotions were compared to the elicited emotions from the related news articles from Rappler. There were cases that the model predicted more than half of the possible emotion set, in those cases we took the top 3 predicted results based on the confidence values.

The prediction results for the murder incident from different news sources generated the emotions: *angry*, *annoyed*, and *sad*, with *angry* as the true emotion based on the Rappler's Mood Meter. This case, we selected 3 news articles from each of the news sources that discuss the same issue. For the Trump inauguration, with the majority's true emotion of *happy*, *angry*, and *annoyed*, the classifier predicted *annoyed* and *amused*. For the UAAP 79 basketball championship where *happy* was the true emotion, the prediction were *happy* and *dont care*.

Based on the prediction results, there is a high probability of the same emotion predictions for articles of the same topic, regardless of source, i.e., whether the source is a broadsheet, TV news or tabloid.. This strengthens our hypothesis that the theme of the news article is a major factor in the emotion elicited from readers. In the case of Trump inauguration, the model was consistent in coming up with the same prediction regardless of the news source; however, the prediction did not perfectly match the actual emotion indicated by the readers of Rappler. It is possible that the model was trained on the topics that were very different from the Trump inauguration, thus the prediction failed. This implies that the model is required to be continuously trained with recent articles for its knowledge to be relevant.

6 Conclusions

Among the document representations, articles are still best represented by *whole article*, with the disadvantage of large feature size. The next best performing document representation is *topic tag* while results to only 10% of the feature size when using the *whole article*. In terms of classification difficulty among the emotion classes, it was observed that *afraid* is the easiest class to classify, while *annoyed* is the most difficult.

Further examination on the word features from each classes was performed. The emotions of *afraid* and *inspired* were seen to be the most distinguishable emotion of all. It is noticeable that most of emotion have a set of unique themes and some emotions such as *angry* and *annoyed* share relatively close themes.

Out of 8 emotion labels, *afraid* and *inspired* are the ones with word distinct characteristics. It would be interesting to investigate further other characteristics of these emotion labels, by considering other sources. The most dominant emotion label is *happy*. Does this fact say something about the readers, or about the articles themselves?

References

[Bhowmick *et al.*, 2009] Plaban Kr Bhowmick, Anupam Basu, Pabitra Mitra, and Abhishek Prasad. Multi-label text classification approach for sentence level news emotion analysis. In *International Conference on Pattern Recognition and Machine Intelligence*, pages 261–266. Springer, 2009.

[Kumar and Suresh, 2012] M. Naveen Kumar and R. Suresh. Emotion detection using lexical chains. *International Journal of Computer Applications*, 57(4), 2012.

[Lin *et al.*, 2008] Kevin Hsin-Yih Lin, Changhua Yang, and Hsin-Hsi Chen. Emotion classification of online news articles from the reader's perspective. pages 220–226. IEEE, December 2008.

[Medhat *et al.*, 2014] Walaa Medhat, Ahmed Hassan, and Hoda Korashy. Sentiment analysis algorithms and applications: a survey. *Ain Shams Engineering Journal*, 5(4):1093–1113, December 2014.

[Parrott, 2001] W Gerrod Parrott. *Emotions in social psychology: essential readings*. Psychology Press, 2001.

[Pedregosa *et al.*, 2011] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas,

A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

[Shivhare *et al.*, 2015] Shiv Naresh Shivhare, Shakun Garg, and Anitesh Mishra. *2015 International Conference on Computing, Communication & Automation (ICCCA 2015): Greater Noida, [UP], India, 15 - 16 May 2015*. IEEE, Piscataway, NJ, 2015.

[Yang *et al.*, 2007] Changhua Yang, Kevin Hsin-Yih Lin, and Hsin-Hsi Chen. Emotion classification using web blog corpora. pages 275–278. IEEE, November 2007.

[Zhang and Zhou, 2007] Min-Ling Zhang and Zhi-Hua Zhou. ML-KNN: A lazy learning approach to multi-label learning. *Pattern recognition*, 40(7):2038–2048, 2007.