Text Characteristics Vector: Rethinking Human-Centered Sentiment Analysis with Emotion-Related Text Characteristics

Cheng Guo[1] [0009−0004−8972−2444], Yanfu Liu[2], Yue Yuan[2], Wuhao Zhang[3], and Sourojit Ghosh[2] [0000−0001−5143−6187]

[1] University of California, San Diego
[2] University of Washington, Seattle
[3] University of Minnesota, Twin Cities
c5guo@ucsd.edu

Abstract. To evaluate how emotional expression is related to valence, we built a sentimental classifier based only on quantitative information of the count of eight text characteristics on a dataset of fanfiction reviews. We also integrated the adjusted character count into the counting. To evaluate our text characteristics vector (TCV) model, we compared it with a model built based on the same data through a TF-IDF vectorizer. Our TCV model performed equally to the TF-IDF model with both F1 scores around 0.90. Our research is critical since it shows how emotional expression can be deterministic on the valence of text. Our result reveals the importance of a human-centered approach to NLP. The result is generalizable to all reviews on any social media. We release data and code in our research for reproducing purposes.

Keywords: human-centered natural language processing · natural language processing · sentiment analysis · qualitative coding.

1 Introduction

Due to the expedited growth of social media, people have increased their involvement in sharing their opinions and checking out reviews on social media regarding buying certain products [20], planning travels [4], and choosing restaurants [36]. As researchers study those reviews, they concluded that most reviews have emotional input, which is often correlated with *valence* i.e. the 'goodness' or 'badness' of text [11]. While the correlation is asserted ([21], [31]), it is still unclear how strong the correlation between emotion and valence is or whether we can approach sentiment analysis with emotions in text. Moreover, there is still more research required on identifying emotions from texts.

In this paper, we demonstrate a novel method of identifying the valence of a given text based on the emotional expression within it, which we establish to be a function of the following eight characteristics: emoticons/emojis, exclamation marks, capitalizations, repetitions, action verbs, intentional misspellings, keyboard smashing, and text length. We adopt as our dataset a publicly-available trove of online fanfiction reviews [34], which are known to be rife with emotional expression as reviewers tend to express strong emotions in reaction to stories about their favorite characters [10]. Operating on a ground-truth dataset which we manually code with a taxonomy of emotions from [14], we design two machine learning models: one that uses a standard TF-IDF vectorizer used for this purpose by [13] and another one that predicts the valence based on the aforementioned characteristics (hereafter referred to as 'the TCV model', abbreviated from Text Characteristics Vector Model). We examine whether the TCV model can match the performance of the previous one, as we compare results produced by the two both manually and through computational metrics. Through our work, we make two novel contributions:

(1) We present novel insights into the process of detecting the nature and degrees of emotional expression within short texts. We adopt a slate of eight characteristics of texts – emoticons/emojis, exclamation marks, capitalizations, repetitions, action verbs, intentional misspellings, keyboard smashing, – initially proposed by [13], to which we add an eighth: text length, as a metric of identifying the valence of a short form text. Based on these characteristics, we design a machine learning model for the task of predicting the valence of short-form text (e.g. tweets, social media comments, etc.). We demonstrate that this approach produces comparable results to models that employ TF-IDF vectorizers, the commonly-adopted procedure for this purpose. Our approach proceeds by processing an input dataset of short texts into the counts and frequencies of the aforementioned characteristics, and then applying the counts to classify the texts into Positive, Neutral or Negative valences. This approach demonstrates a significant speedup at the classification stage over TF-IDF vectorizers, as it only operates on the counts and frequencies rather than the words and contents within the text. Upon comparing results from TF-IDF vectorizers and our model over 100,000 texts, we observed that our model produces 91% similar results to the former. Upon examination of differences, we observe that only 1844 valid disagreements, since the remaining 7606 are not in English and our training data was only in English texts. Manual examination of a random sample of 50 English texts for which the models disagree reveals that although we do agree more overall with the TF-IDF vectorizer's classification of reviews over the TCV model's, we do agree more with the TCV model's classification of Positive Valences. While this approach does not yet represent an improvement upon TF-IDF vectorizers, we believe that it shows promising results and is encouraging as a direction of future work in the field of sentiment analysis.

(2) We contribute towards the growing field of human-centered machine learning [8] by adopting a human-centered approach to designing a sentiment classifier. We extend the work of [13], which covers degrees of positive emotions present within short texts but produces statedly underwhelming results, by establishing how our human-centered approach of building our ground-truth dataset from manually coding data ourselves led to the aforementioned observation and the emergence of the pattern between texts and the eight characteristics. We hope that our work adds to the literature in the field which seeks to motivate researchers to adopt more human-centered methods in their ML/NLP tasks, by presenting a successful case study.

2 Related Work

2.1 Emotion Classification in Text

The study of emotions and valence within texts has long since been the subject of attention of scholars from a wide range of fields [11]. That people express emotions in text is unsurprising and, over the past few decades, there has been substantial work in analyzing emotions within texts.

Within the field of NLP, researchers have attempted to classify emotions in speech ([5], [7]), tweets [23], and movie reviews [30], to name a few. Caschera et al. [7] and Turney [30] used as their method of classification part-of-speech (POS) tagging, which builds vectors based on the modality and syntax of each word in a sentence and joining them together to identify an overall emotion within the sentence [32]. When classifying emotion in tweets, Mohammad [23] focused on emotion-related hashtags like #joy and developed binary classifiers, like classifying whether a tweet has joy emotion or not, which ends up with a better classifier for positive emotions since more positive tweets were collected. Researchers have also applied Pointwise Mutual Information (PMI) ([23], [30]), a metric that determines the possibility of two different words occurring in the same sentence, to calculate the semantic orientation and the emotion in a sentence. Researchers have also attempted the machine learning approach for emotion classification, such as KNN or neural networks [27].

Recently, deep-learning approaches have been widely studied and applied to solving NLP tasks. Regarding emotion classification, researchers have used a convolutional neural network (CNN) on short texts and achieved better accuracy than SVM and other deep-learning methods, including LSTM and RNN [33]. Deep-learning approaches embed words into feature vectors that can collect information like semantic relevance that is hard to capture from traditional techniques. When classifying emotions of text on the Microblog of China, researchers have found that CNN could obtain an accuracy of about 7.0% better than other approaches [33]. Beyond emotion classification, deep-learning techniques like CNN have proved to perform better in classifying tobacco and health-related datasets [18] since they can extract more information from texts and study them in sequential or hierarchal structures.

2.2 Sentiment Analysis

Since Nasukawa and Yi [25] first coined the term ‘sentiment analysis’, it has risen to become synonymous with the process of computationally determining emotions within the text and is one of the most recognized NLP tasks [22]. Nasukawa and Yi [25] defined it as the identification of sentiment expressions, including their polarity and strengths, and their relations to the subject of the text. It generally proceeds by training ML models on a ground-truth dataset of sentiment-labeled texts, and then executing the model on a dataset of unlabeled texts to produce sentiment labels. This type of modeling is called supervised learning and is the most common, though semi-supervised and unsupervised versions are also in common usage for sentiment analysis tasks.

Over the years, researchers have applied various techniques to extract sentiment from various texts, mostly on Twitter ([16], [17], [26]) and movie reviews ([3], [28], [29]) texts. The researchers started with preprocessing text data, including lemmatization ([17], [28]), removing special characters ([16], [17], [28]), and correcting misspelled words ([17], [26]). Some of them removed slang words [16], but others replaced them with words in the dictionary to maintain the emotional contribution of those words [26]. For the processing stage, they created vectors by feature extraction, including part-of-speech (POS) tagging ([17], [26]) and assigning weights to words ([16], [26]). Then, they would apply ML models, mostly Naïve Bayes or SVM or both ([3], [16], [17], [26], [28], [29]). For Twitter text, they ended with that while Naïve Bayes and SVM result in models with similar accuracy ([17], [26]), the decision trees method performs better [17]. For movie reviews text, some of them concluded Naïve Bayes is better ([3], [28]), while others claim SVM is better [29].

2.3 Human-Centered Sentiment analysis

Over the past few years, the emerging field of human-centered machine learning [8] has led to the birth of a human-centered approach towards sentiment analysis (e.g. [2], [13], [24]). This approach towards sentiment analysis asks to go beyond simply treating texts as data to classify, but rather to consider the emotions and thoughts of the humans who produced such texts. Ghosh et al. [14] demonstrate how, in classifying the same dataset of online fanfiction reviews as we did, the importance of understanding the subjective differences between different degrees of positive emotions as a function of the ways in which emotional content is expressed, arguing that if a reviewer made a conscious choice of using capitalization (e.g. 'I LOVE THIS' over 'I love this'), then that choice must be respected in understanding the degree of positive emotion expressed. In this vein, they later [13] designed a human-centered machine learning model that attempted to distinguish between different degrees of emotions. We extend their work, which produced underwhelming results, by examining how patterns within the contents of reviews can be used to determine their valence.

2.4 Fanfiction and NLP

As a field with an abundant amount of text, fanfiction attracted lots of NLP researchers to analyze and classify texts. Researchers from Carnegie Mellon University formulated a fully structured pipeline for analyzing fanfiction texts [35]. They introduced SpanBERT-based language models and built a dataset with features like character coherence and quote attribution [35]. They evaluated their pipeline and compared it with pipelines like BookNLP and CoreNLP, hoping that their structured analysis could be generalized to domains other than fanfiction [35]. Researchers have also evaluated various traditional and novel approaches to NLP tasks on fanfiction to see which performs the best. The researchers tested traditional approaches by combining TF-IDF with Naïve Bayes and SVM, and they compared them with deep-learning approaches combined with Word2Vec [9]. In conclusion, they discovered that SVM achieved better accuracy than deep learning approaches and TF-IDF with Naïve Bayes is the best approach, specifically for classification tasks related to fanfiction [9].

3 Methods

3.1 Online Fanfiction Reviews

We use for this study a publicly-available dataset of online fanfiction reviews collected by [34]. The full dataset[4] contains metadata from 6,807,100 stories by 1,516,335 unique authors in 44 different languages, as well as over 176 million reviews across all the stories. In the world of online fanfiction, a 'review' is a comment left on a story by a user. Prior research into these reviews (e.g. [6], [10], [14]) has demonstrated that such reviews are rife with emotional expression, especially expressing extremely positive or extremely negative sentiments clearly and often. Therefore, for our stated goals of detecting valence of short-form texts, this dataset is appropriate.

We began with a random sample of 10000 reviews to qualitatively code (explained in Section 3.2) and used the labeled data to train our model. For testing, we extracted another random sample of 100,000 reviews from [34]'s dataset.

3.2 The Eight Text Characteristics

The purpose of this study was to examine whether the text characteristics of Positive emotions proposed by [13] could indeed be used to predict the valence of short form texts. We began this study by qualitatively coding the aforementioned dataset of 10000 fanfiction reviews with a taxonomy of 11 emotions proposed by [14], which are further divided into three categories of Positive, Negative, and Unclassified emotions (full slate of emotions shown in Table 2). Four researchers individually coded 9364 reviews (dataset shortened after removing non-English reviews to avoid errors in translation similarly as [14]), with a specific attention towards the three Positive emotions: Like, Joy/Happiness, and Anticipation/Hope. During the coding process, we had detailed discussion on *why* each researcher coded a review with a Positive emotion, paying close attention to the role played by one of [13]'s characteristics of Positive reviews. In doing so, we examined whether we could find patterns between the emotion chosen and the text characteristics, specifically whether the presence of one of the characteristics was a stronger indication of a particular emotion over the other. Through tabulations of agreements/disagreements and observations of the presence of the

[4] http://research.fru1t.me./

characteristics as determining factors, we adopt the slate of seven characteristics proposed by [13] – emoticons/emojis, exclamation marks, capitalizations, repetitions, action verbs, intentional misspellings, and keyboard smashing– for our work. We further identify an eighth characteristic that played a role in our determining different emotions: text length. We reasoned that longer reviews are indicative of a stronger degree of user investment and might contain a clearer expression of emotions, and therefore adopted it into our slate of characteristics.

3.3 Qualitative Coding

To build the model, we first need qualitatively coded data. To determine the ground truth of the sentiments in fanfiction reviews, a team of four independently coded a random sample of 1000 reviews (different than the ones referred to above) each over five weeks between February and March 2023. For each review, each coder determined its valence by selecting one of the four: positive, neutral, negative, or unknown. Since there are many more positive reviews than all the other three categories combined [10], we decided to put all three other categories into one, named "not positive," coded with 0 in the coded dataset, and positive reviews are coded with 1. By combining four persons' codes, we determined the final valence for each of the 1000 reviews on a majority vote scheme, while if two people coded 0 and two coded 1, we invited a fifth person as a tie-breaker. Now we have 1000 reviews, each with its coded valence in 1 or 0.

3.4 Quantitative Modeling

After we have the qualitative data, we started building the model. We first extracted the count of the following text characteristics: emoticons/emojis, exclamation marks, capitalizations, repetitions, action verbs, intentional misspellings, and keyboard smashing on each review as quantitative measures, then we also take the length of each review as another measure. These eight text characteristics are proved to have correlated with the emotion expressed in the review, especially positive emotions like joy, hope, and like [13]. For counting the capitalization and total length of the text, we incorporated Adjusted Character Count (ACC) since capitalization can embed strong emotions in the text [14]. If in the text, more than three characters except for whitespace are next to each other on the QWERTY keyboard, we would count that as one keyboard smashing. For repetition, we take the number of the character on the keyboard that is typed the most in the sentence. For intentional misspelling, currently we counted every word in the review that is misspelled. These eight quantitative measures are combined as one text characteristics vector (TCV). Based on the TCV, we build a quantitative model to predict a review's valence in the above two categories: positive or not positive. We chose Linear SVC since it tends to result in a higher accuracy compared to other machine-learning models ([1], [15], [19]). We also build a second model based on TF-IDF statistics and Linear SVC with the 1000 reviews and our code, since TF-IDF is a well-known algorithm in NLP [12]. We compared the two models through their F1 scores because we want to minimize both false positives and false negatives.

3.5 Manual Examination

Beyond building models on 1000 reviews and comparing their performance, we applied the two models to 100,000 fanfiction reviews and retrieved the predictions through both models. We extracted all the English reviews that are classified differently. Then we randomly selected a sample of 50 reviews and compared the results to see which model's result we agree with.

4 Findings

4.1 Model Performances

We have built two ML models, one is a Linear SVC model on the vector of eight text characteristics for each review (TCV model), and the other one is also a Linear SVC model built on tokenizing reviews with the TF-IDF vectorizer. Both models have an F1 score of around 0.90 when classifying reviews as positive or not positive. Specifically, the TCV model has an F1 score of about 0.9130, and the model with TF-IDF vectorizer reached an F1 score of 0.9043. Based solely on the F1 scores, we can say that the model on text characteristics has achieved an acceptable F1 score as with the model with the TF-IDF vectorizer. However, further consideration is required to examine specific reviews and their predicted results from the two models to validate the models' performance.

4.2 Manual Examination

Among the 100,000 reviews, 9450 are classified differently. Among the 9450 reviews, there are 1834 in English, 7606 non-English reviews, and 10 reviews consisting of only emoticons. We randomly selected 50 reviews from the 1834 English reviews that are classified differently and compared the results to see which model's result we agree with. Among the 50 reviews we selected, there are 32 reviews in which we agree with the model built on TF-IDF vectorizers, and 18 we agree with the model from the text characteristics method. Among the 32 reviews we agree with the TF-IDF model, three of them we both agree to be positive reviews and 29 of them we both agree to be non-positive. Among the 18 reviews we agree with the model from the text characteristics method, 16 of them we both agree to be positive, and two of them we both agree to be non-positive.

There are five reviews where the TF-IDF model characterizes it to be positive, while the text characteristics model predicts it is not positive. Among these five, four of them contain almost exclusively capital letters. Three of these four we agree with the TF-IDF model as positive reviews. Those three reviews contain words like "funny" or "groovy" and encouragement to the author to update soon.

Table 1. Examples from Manual Examination

Text	**TF-IDF Model**	**TCV Model**	**Manual Examination**
HEY, I DIDN'T REVIEW TIL THE END, IT'S GROOVY! I LOVE HIGH SCHOOL FICS...EVEN GOT ONE OF MY OWN...WELL HURRY UP WITH THE NEXT CHAPTER PLEASE, THIS STORY IS GREAT!	Positive	Not Positive	Positive
Well my prediction was right. And I'm SO SAD! I almost cried! AW WHY VICTORIA? Poor Albus! Why could'nt a[nother] bad guy die? *weep* Oh, and isn't the Griffin supposed to be a Hippogriff? :S I hate u for that death chapter.	Not Positive	Positive	Not Positive
I'm happy to see that this story is being updated again - I love it! Poor Relena, though. :(	Not Positive	Positive	Positive
Anna and kokoro nonoko and yuu sumire and monchu otonashi and kitsuneme those are what I think the parings should be... with Natsume and Mikan, Ruka and Hotaru of course	Positive	Not Positive	Not Positive

The TF-IDF model characterizes the other 45 reviews as not positive, and we agree with it on 29 reviews. Although we agree that these 29 reviews are not positive, they are neither negative. Those reviews discuss the plot of the fanfiction while the reviewer did not express like or dislike for the fanfiction itself. Those reviews generally used more capitalization, exclamation marks, and emoticons compared to the other 16 reviews, where we agree with the text characteristics model regarding their positivity. Those 16 reviews did not contain many special text characteristics, but there is more encouragement for updating from the author.

5 Discussion

5.1 New Framework for Sentiment Analysis

In the previous approach to sentiment analysis, researchers first preprocess the data through steps like lemmatization ([17], [28]) and removing special characters ([16], [17], [28]). For the process stage, they would extract vectors from text, like part-of-speech (POS) ([17], [26]) or TF-IDF vector [12], then apply machine learning algorithms like Naïve Bayes or SVM ([3], [16], [17], [26], [28], [29]). However, in our approach, we skipped the preprocessing stage since we want to analyze text characteristics beyond their stem form, so we directly extract a vector composed of the count of eight text characteristics and apply Linear SVC to it. Thus, we eliminated the preprocessing stage.

5.2 Manual Examination

During the manual examination, we observed three shortcomings that may hinder model performance.

Counting Emoticons Since we only considered the count of the eight text characteristics, the model did not include what those text characteristics entail, especially for emoticons. For emoticons, :) implies happy emotions, and :(entails sad emotions. In our code, both of them are counted as one emoticon, so did not consider the different emotions of those emoticons.

Counting Repetitions We take the count of the most-typed character in the sentence. If in a sentence, there is the usage of a lot of certain English characters, but the reviewer did not repeat it on purpose, that count in the vector represents wrong information of the text.

Identifying Keyboard Smashing If more than three characters except for whitespace in the text are next to each other on the QWERTY keyboard, we would count that as one occurrence of keyboard smashing. If the reviewer is using keyboards other than the QWERTY keyboard, like T9 or Dvorak keyboards, we cannot identify if they used keyboard smashing.

5.3 Run Time of Building Models

When building a model, it is crucial to consider the Run Time. While we expect the two models we built to cost roughly the same amount of time, this is not true. There are two stages in building the TCV model, the first one is the processing stage, in which we extract the count of those eight features, and the second stage is the classification stage, which is building the model based solely on the eight counts. If we only consider building the model, then the text characteristics method has less time cost since there is no need to vectorize all reviews. However, this model takes a significant amount of time in the processing stage when extracting the count of the text characteristics. In specific, when counting the misspelled words of a review, our code takes the most amount of time, which is about 6 minutes on 1000 reviews, and the total time required for all features is about 10 minutes on 1000 reviews. When classifying the 100,000 reviews, since we did not remove non-English reviews, the time usage varies due to our code recognizing every non-English word as misspelled. The total time for building the model with the TF-IDF vectorizer with 1000 reviews is around 5 minutes and building the model of text characteristics vector takes at least 10 minutes.

5.4 Implications

The new Text Characteristics Vector model we proposed has the following implications.

Capitalization People usually express strong emotions through capitalization, but the following review is a counterexample: *I AM GONNA KILL U! HOW COULD U JUST LEAVE IT THERE AND THEN PICK UP THE SEQUEL 7 YEARS LATER! AND Y'D U HAV 2 MAKE THEM BREAK UP! U BETTER FIX IT IN UR SEQUEL CUZ THEY'RE SUPPOSED TO B 2GETHER AND IF THE SEQUELS SEVEN YEARS LATER THAT WOULD BE LIKE NEITHER OF THEM FELTR GUILTY! POST THE SEQUL SOON PLZ CUZ THAT WAS SUCH A HORIBLE ENDING!* Even though the reviewer used most exclusively capitalized letters, we think the reviewer did not express a positive attitude since they asked the author to fix the relationships of two characters in the fanfiction. We agree on the TCV model in this review is not positive. The TF-IDF method would turn words into lower cases and remove all the exclamation marks, and we do not agree with the results from the TF-IDF model.

Emoticon in Text While we discussed emoticons can entail different emotions, sometimes people use emoticons not with the emotion it implies. Consider the following review: *:(why did ya stop posting? i'm heartbroken...post more please! its my birthday!* Although the reviewer used a sad emoticon :(at the beginning of the review, they encourage the author to post more of this fanfiction as they are reading it in their birthday, and we agree with the TCV model to categorize this review as positive. The TF-IDF method will remove the emoticons before processing, and the TF-IDF model categorizes it as not a positive review.

Generalization of the TCV Model The results from our model could be applied to all short-formed text in social media like Twitter, Reddit, or movie reviews since people express strong emotions in texts through the eight characteristics. Examples can be this tweet from Donald Trump on January 8, 2021: *The 75,000,000 great American Patriots who voted for me, AMERICA FIRST, and MAKE AMERICA GREAT AGAIN, will have a GIANT VOICE long into the future. They will not be disrespected or treated unfairly in any way, shape or form!!!* There are some capitalizations and exclamation marks, so it can be categorized as a positive tweet through the text characteristics vector, but people who did not vote for Trump may categorize this tweet not as a positive one.

Human-Centered Machine Learning Researchers who attempted Human-Centered Machine Learning have encouraged future designers to manually code the data [13] and in our research, four coders have coded 1000 reviews independently. We also have a fifth coder for tie-breaking. Our process proves that the human-centered machine-learning approach can be applied to sentiment analysis and result in a model with an F1 score over 0.90. We focused on text characteristics that would normally be removed during the preprocessing stage and created new possibilities for future NLP researchers to extract information from texts beyond the stem form.

6 Future Plan

We would like to improve the Text Characteristics Vector in the following ways.

Emoticon For the future version of the vector, we would like to categorize emoticons based on the emotion it entails, for example :) and :> for happy, :(and :< for sad. Then we would count different categories separately so that the model will consider emoticons based on the emotion it implies.

Repetition In the current algorithm, we just counted the English character that appeared the most times as the number for repetition. However, in the future version, we hope to count the repetition that is intended, like e in yeees, or h in ahhhh. Thus, we need a more precise way to count repetitions in the future version.

Intentional Misspelling Currently we are counting all the words that are not in English, but sometimes people have unintentional typos. We hope to examine whether the misspelling is intentional or not based on the content of the text in a future version of the vector.

Keyboard Smashing While the current algorithm for counting keyboard smashing is acceptable, we would like to develop a more rigorous version of it in the future version of the vector so that it can identify a keyboard smash in a convincing way.

Deep Learning As discussed in Section 2.1, researchers have incorporated deep-learning approaches to classify emotions in text resulting in higher accuracy. We hope to explore the potential of applying the text characteristics vector in deep learning as a neuron in the network to improve performance and make the research more comprehensive and human-centered.

7 Conclusion

We have approached sentiment analysis in a human-centered way by building an ML model based solely on the count of eight text characteristics, including emoticons/emojis, exclamation marks, capitalizations, repetitions, action verbs, intentional misspellings, keyboard smashing, and text length, and compared it with a model with TF-IDF vectorizers. Both models achieved equally well results. Through analyzing the reviews that are categorized differently between the two models, we found that we are more inclined to agree with the TCV model when it predicts a review to be positive, while more likely to agree with the TF-IDF model when it predicts a review to be negative. We provided evidence for encouraging the human-centered approach in sentiment analysis on short texts from social media in that it can substitute the TF-IDF vectorizing process and result in an equally-better model.

References

1. Ahmed, H., Awan, M., Khan, N., Yasin, A., Faisal Shehzad, H.: Sentiment analysis of online food reviews using big data analytics. İlköğretim Online **20**, 827–836 (04 2021). https://doi.org/10.17051/ilkonline.2021.02.93
2. Ali, K., Dong, H., Bouguettaya, A., Erradi, A., Hadjidj, R.: Sentiment analysis as a service: a social media based sentiment analysis framework. In: 2017 IEEE international conference on web services (ICWS). pp. 660–667. IEEE (2017)

3. Baid, P., Gupta, A., Chaplot, N.: Sentiment analysis of movie reviews using machine learning techniques. International Journal of Computer Applications **179**, 45–49 (12 2017). https://doi.org/10.5120/ijca2017916005
4. Baka, V.: The becoming of user-generated reviews: Looking at the past to understand the future of managing reputation in the travel sector. Tourism Management **53**, 148–162 (2016). https://doi.org/https://doi.org/10.1016/j.tourman.2015.09.004, https://www.sciencedirect.com/science/article/pii/S0261517715300091
5. Borchert, M., Dusterhoft, A.: Emotions in speech - experiments with prosody and quality features in speech for use in categorical and dimensional emotion recognition environments. In: 2005 International Conference on Natural Language Processing and Knowledge Engineering. pp. 147–151 (2005). https://doi.org/10.1109/NLPKE.2005.1598724
6. Campbell, J., Aragon, C., Davis, K., Evans, S., Evans, A., Randall, D.: Thousands of positive reviews: Distributed mentoring in online fan communities. In: Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing. p. 691–704. CSCW '16, Association for Computing Machinery, New York, NY, USA (2016). https://doi.org/10.1145/2818048.2819934, https://doi.org/10.1145/2818048.2819934
7. Caschera, M.C., Grifoni, P., Ferri, F.: Emotion classification from speech and text in videos using a multimodal approach. Multimodal Technologies and Interaction **6**(4) (2022). https://doi.org/10.3390/mti6040028, https://www.mdpi.com/2414-4088/6/4/28
8. Chancellor, S.: Toward practices for human-centered machine learning. Communications of the ACM **66**(3), 78–85 (2023)
9. Donaldson, C., Pope, J.: Data collection and analysis of print and fan fiction classification. In: International Conference on Pattern Recognition Applications and Methods (2022), https://api.semanticscholar.org/CorpusID:246948382
10. Frens, J., Davis, R., Lee, J., Zhang, D., Aragon, C.: Reviews matter: how distributed mentoring predicts lexical diversity on fanfiction. net. arXiv preprint arXiv:1809.10268 (2018)
11. Frijda, N.H., et al.: The emotions. Cambridge University Press (1986)
12. Gebre, B.G., Zampieri, M., Wittenburg, P., Heskes, T.: Improving native language identification with TF-IDF weighting. In: Proceedings of the Eighth Workshop on Innovative Use of NLP for Building Educational Applications. pp. 216–223. Association for Computational Linguistics, Atlanta, Georgia (Jun 2013), https://aclanthology.org/W13-1728
13. Ghosh, S., Ali, M., Batra, A., Guo, C., Jain, M., Kang, J., Kharchenko, J., Suravajhela, V., Zhou, V., Aragon, C.: "do we like this, or do we like like this?" : Reflections on a human-centered machine learning approach to sentiment analysis. In: Proceedings of the 4th International Conference on Artificial Intelligence in HCI in the context of the 25th International Conference on Human-Computer Interaction (HCI International). Springer (2023)
14. Ghosh, S., Froelich, N., Aragon, C.: "i love you, my dear friend": Analyzing the role of emotions in the building of friendships in online fanfiction communities. In: Proceedings of the 15th International Conference on Social Computing and Social Media in the context of the 25th International Conference on Human-Computer Interaction (HCI International). Springer (2023)
15. Gulati, K., Saravana Kumar, S., Sarath Kumar Boddu, R., Sarvakar, K., Kumar Sharma, D., Nomani, M.: Comparative analysis of machine learning-based classification models using sentiment classification of

tweets related to covid-19 pandemic. Materials Today: Proceedings **51**, 38–41 (2022). https://doi.org/https://doi.org/10.1016/j.matpr.2021.04.364, https://www.sciencedirect.com/science/article/pii/S2214785321032843, cMAE'21
16. Hasan, A., Moin, S., Karim, A., Shamshirband, S.: Machine learning-based sentiment analysis for twitter accounts. Mathematical and Computational Applications **23**(1) (2018). https://doi.org/10.3390/mca23010011, https://www.mdpi.com/2297-8747/23/1/11
17. Jain, A.P., Dandannavar, P.: Application of machine learning techniques to sentiment analysis. In: 2016 2nd International Conference on Applied and Theoretical Computing and Communication Technology (iCATccT). pp. 628–632 (2016). https://doi.org/10.1109/ICATCCT.2016.7912076
18. Kamath, C.N., Bukhari, S.S., Dengel, A.: Comparative study between traditional machine learning and deep learning approaches for text classification. In: Proceedings of the ACM Symposium on Document Engineering 2018. DocEng '18, Association for Computing Machinery, New York, NY, USA (2018). https://doi.org/10.1145/3209280.3209526, https://doi.org/10.1145/3209280.3209526
19. Kamath, C.N., Bukhari, S.S., Dengel, A.: Comparative study between traditional machine learning and deep learning approaches for text classification. In: Proceedings of the ACM Symposium on Document Engineering 2018. DocEng '18, Association for Computing Machinery, New York, NY, USA (2018). https://doi.org/10.1145/3209280.3209526, https://doi.org/10.1145/3209280.3209526
20. Kwahk, K.Y., Ge, X.: The effects of social media on e-commerce: A perspective of social impact theory. In: 2012 45th Hawaii International Conference on System Sciences. pp. 1814–1823 (2012). https://doi.org/10.1109/HICSS.2012.564
21. Lelieveld, G.J., Hendriks, H.: The interpersonal effects of distinct emotions in online reviews. Cognition and Emotion **35**(7), 1257–1280 (2021). https://doi.org/10.1080/02699931.2021.1947199, https://doi.org/10.1080/02699931.2021.1947199, pMID: 34187323
22. Liu, B.: Introduction, p. 1–17. Studies in Natural Language Processing, Cambridge University Press, 2 edn. (2020). https://doi.org/10.1017/9781108639286.002
23. Mohammad, S.: #emotional tweets. In: *SEM 2012: The First Joint Conference on Lexical and Computational Semantics – Volume 1: Proceedings of the main conference and the shared task, and Volume 2: Proceedings of the Sixth International Workshop on Semantic Evaluation (SemEval 2012). pp. 246–255. Association for Computational Linguistics, Montréal, Canada (7-8 Jun 2012), https://aclanthology.org/S12-1033
24. Nalis, I., Neidhardt, J.: Not facial expression, nor fingerprint–acknowledging complexity and context in emotion research for human-centered personalization and adaptation. In: Adjunct Proceedings of the 31st ACM Conference on User Modeling, Adaptation and Personalization. pp. 325–330 (2023)
25. Nasukawa, T., Yi, J.: Sentiment analysis: Capturing favorability using natural language processing. In: Proceedings of the 2nd International Conference on Knowledge Capture. p. 70–77. K-CAP '03, Association for Computing Machinery, New York, NY, USA (2003). https://doi.org/10.1145/945645.945658, https://doi.org/10.1145/945645.945658
26. Neethu, M.S., Rajasree, R.: Sentiment analysis in twitter using machine learning techniques. In: 2013 Fourth International Conference on Computing, Communications and Networking Technologies (ICCCNT). pp. 1–5 (2013). https://doi.org/10.1109/ICCCNT.2013.6726818

27. Petrushin, V.: Emotion in speech: Recognition and application to call centers. Proceedings of Artificial Neural Networks in Engineering (01 2000)
28. Rahman, A., Hossen, M.S.: Sentiment analysis on movie review data using machine learning approach. In: 2019 International Conference on Bangla Speech and Language Processing (ICBSLP). pp. 1–4 (2019). https://doi.org/10.1109/ICBSLP47725.2019.201470
29. Samal, B., Behera, A.K., Panda, M.: Performance analysis of supervised machine learning techniques for sentiment analysis. In: 2017 Third International Conference on Sensing, Signal Processing and Security (ICSSS). pp. 128–133 (2017). https://doi.org/10.1109/SSPS.2017.8071579
30. Turney, P.D.: Thumbs up or thumbs down? semantic orientation applied to unsupervised classification of reviews. In: Proceedings of the 40th Annual Meeting on Association for Computational Linguistics. p. 417–424. ACL '02, Association for Computational Linguistics, USA (2002). https://doi.org/10.3115/1073083.1073153, https://doi.org/10.3115/1073083.1073153
31. Ullah, R., Amblee, N., Kim, W., Lee, H.: From valence to emotions: Exploring the distribution of emotions in online product reviews. Decision Support Systems **81**, 41–53 (2016). https://doi.org/https://doi.org/10.1016/j.dss.2015.10.007, https://www.sciencedirect.com/science/article/pii/S0167923615002018
32. Voutilainen, A.: Part-of-speech tagging. The Oxford handbook of computational linguistics pp. 219–232 (2003)
33. Xu, D., Tian, Z., Lai, R., Kong, X., Tan, Z., Shi, W.: Deep learning based emotion analysis of microblog texts. Information Fusion **64**, 1–11 (2020). https://doi.org/https://doi.org/10.1016/j.inffus.2020.06.002, https://www.sciencedirect.com/science/article/pii/S156625352030302X
34. Yin, K., Aragon, C., Evans, S., Davis, K.: Where no one has gone before: A metadataset of the world's largest fanfiction repository. In: Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems. pp. 6106–6110 (2017)
35. Yoder, M., Khosla, S., Shen, Q., Naik, A., Jin, H., Muralidharan, H., Rosé, C.: FanfictionNLP: A text processing pipeline for fanfiction. In: Akoury, N., Brahman, F., Chaturvedi, S., Clark, E., Iyyer, M., Martin, L.J. (eds.) Proceedings of the Third Workshop on Narrative Understanding. pp. 13–23. Association for Computational Linguistics, Virtual (Jun 2021). https://doi.org/10.18653/v1/2021.nuse-1.2, https://aclanthology.org/2021.nuse-1.2
36. Zhang, Z., Ye, Q., Law, R., Li, Y.: The impact of e-word-of-mouth on the online popularity of restaurants: A comparison of consumer reviews and editor reviews. International Journal of Hospitality Management **29**(4), 694–700 (2010). https://doi.org/https://doi.org/10.1016/j.ijhm.2010.02.002, https://www.sciencedirect.com/science/article/pii/S0278431910000198

Table 2. The taxonomy of emotion codes, along with definitions, examples and type of emotion (Positive, Negative, Unclassified) from [14]

Emotion Code	**Definition**	**Example**	**Emotion Type**
Like	The reviewer expresses generic or slightly positive emotions, without going into too much depth.	Wow, I really like this chapter	Positive
Joy/Happiness	The reviewer has more than just a slightly positive reaction to the story and has taken time to adequately express this.	I LOVE this story! Excellent work!	Positive
Anticipation/Hope	The reviewer is expressing their hope of seeing upcoming work.	Good job I'll be waiting for more	Positive
Surprise	The reviewer is surprised, either pleasantly or otherwise.	Whoa I did not see that coming	Positive
Dislike	The reviewer expresses generic or slightly negative emotions, without going into too much depth.	I was a little disappointed	Negative
Disturbed/Disgust	The reviewer expresses discomfort with the content of the story, either with some specific parts or the general tone.	Ugh Snape makes me want to crawl out of my skin	Negative
Anger/Frustration	The reviewer expresses an extreme negative reaction either to the story or the lack of updates.	This is absolutely garbage	Negative
Sadness	The reviewer expresses sadness, either mildly or through tears	Broke my heart :,(I cried a bit	Negative
Confused	The reviewer expresses confusion, as most often indicated by one or more questions.	Why would Harry do that??	Negative
Unknown	The text is either indecipherable or is in a language other than English.	me encanta!	Unclassified
No emotion	Any emotion cannot be reliably assigned to the text.	I'm a Boy	Unclassified