2023 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)

# Monitoring Dynamics of Emotional Sentiment in Social Network Commentaries

Ismail Hossain*, Sai Puppala†, Md Jahangir Alam‡ and Sajedul Talukder§
School of Computing, Southern Illinois University Carbondale, IL, USA
Email: *ismail.hossain@siu.edu, †saimaniteja.puppala@siu.edu, ‡mdjahangir.alam@siu.edu, §sajedul.talukder@siu.edu

***Abstract*—The proliferation of social media offers a real-time reflection of public sentiments. Sentiment analysis on such platforms yields crucial insights for sectors like market research, politics, business strategy, and public health. In this study, we introduce an innovative framework to examine evolving sentiments in social media comments and understand their wider implications. Utilizing a pre-trained BERT base uncase model, we estimate emotional values from comments and align them with various sentiment trends such as Approval, Toxicity, and Neutral, among others. By leveraging machine learning, we train on a distinctive dataset, correlating emotional values with sentiment trends to generate trend likelihood scores. Through a bottom-up methodology, we compile emotional ratings across comment threads to forecast overarching sentiment scores. Our results reveal that the BERT base uncase model excels in emotional prediction, achieving an AUC of 0.91. Meanwhile, Decision Tree models stand out, registering an F1 score above 0.40 on a macro average basis.**

***Index Terms*—sentiment analysis, NLP, BERT, social network**

## I. Introduction

In the digital era, social media platforms have transcended their initial role as communication channels, evolving into powerful barometers of public sentiment. Their omnipresence in contemporary society amplifies their influence on shaping worldviews and opinions [1]. Platforms like Twitter, Reddit, and Facebook empower individuals to vocalize their perspectives, catalyzing ripple effects that foster diverse dialogues across global networks. Whether between close acquaintances or distant users, these platforms nurture conversations that weave intricate webs of discourse. Harnessing the power of sentiment analysis emerges as a pivotal tool, particularly when gauging the dynamic ebbs and flows of emotions within comment sections [2]. These spaces, pulsating with candid exchanges and emotive discussions, serve as invaluable reservoirs for distilling collective sentiment. By meticulously tracking these sentiment oscillations, businesses can tailor marketing tactics, navigate crises, and fuel product innovations. Simultaneously, for policymakers and academic researchers, such insights become instrumental in gauging public pulse on pivotal societal issues, policy enactments, or public health scenarios.

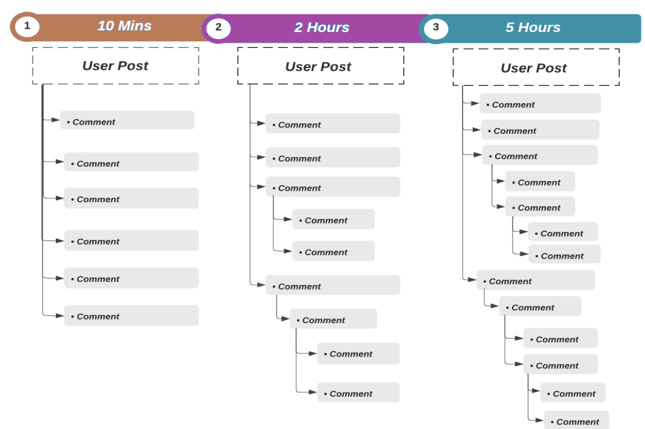


Fig. 1: A sample of a hierarchy thread representing the possible increase in comments over the period of time for a user post.

In this paper, we present an in-depth methodology for tracing sentiment evolutions within social media comment sections. This technique offers actionable insights and enhances our comprehension of public sentiment dynamics in today's social media landscape. Our focus centers on delving into comment threads, aiming to project overarching trends from individual sentiment expressions. In doing so, we introduce a system that identifies sentiments like Approval, Toxicity, Obscenity, Threat, Hate, Offensive, and Neutral, deriving from users' emotional responses. This begins with utilizing a pre-trained BERT base uncase model to gauge sentiments from comments and their subsequent replies. The BERT output maps these comments to one of the predefined trends. Leveraging a range of machine learning models, we



*ASONAM '23*, November 6-9, 2023, Kusadasi, Turkey

ACM ISBN 979-8-4007-0409-3/23/11…$15.00
https://doi.org/10.1145/3625007.3627730

curate a dataset juxtaposing emotion scores against these trends, yielding trend probability scores. After comprehensive training, the methodology is tested on hierarchical comment data. As we navigate each tier of this comment structure, sentiment scores are consistently extracted and consolidated from top to bottom, culminating in collective sentiment metrics for the complete comment hierarchy. These metrics then guide the prediction of overarching sentiment trends. Conclusively, our BERT-based model attains an AUC of 0.91, while the Decision Tree models notably surpass their counterparts, securing an F1 score exceeding 0.40 on a macro scale.

Figure 1 depicts the chronological expansion of the comment section. The visual representation uses varying rectangular box sizes to mirror the lengths of user comments. By harnessing the power of data science, we've meticulously charted, scrutinized, and visualized collective sentiments, granting a tangible insight into online interaction patterns. This research harnesses cutting-edge sentiment analysis tools and innovative visualization methods, shedding light on the evolving contours of public sentiment. The importance of monitoring these sentiment transitions resonates across diverse arenas, from market analysis and crisis management to political reviews, public health surveillance, and sociological studies.

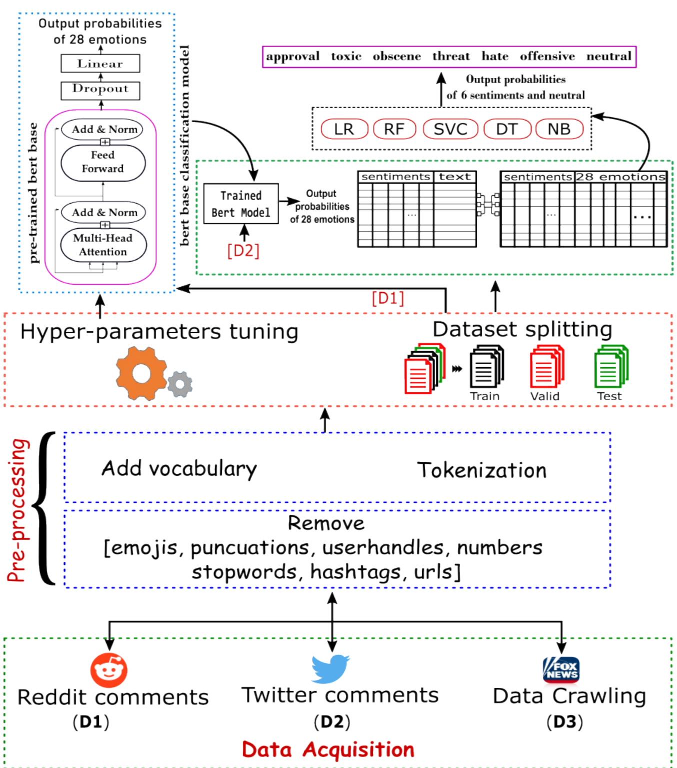


Fig. 2: System flow architecture starting at data acquisitions from different sources to the machine learning model tuning and different algorithms.

## II. Motivation

Analyzing collective sentiment from comment sections is imperative due to its multifaceted implications. First and foremost, these sections serve as a barometer for public opinion, allowing businesses and policymakers to tap into the general mood and views surrounding a particular topic or event. This understanding can subsequently guide marketing strategies, product innovations, and policy decisions. Additionally, shifts in sentiment can function as an early warning system, highlighting emerging concerns or issues that might necessitate prompt attention. This proactive approach can prevent minor grievances from escalating into significant challenges. Furthermore, by acknowledging and addressing the sentiments expressed, organizations can enhance user engagement and foster deeper trust. For content creators, gauging sentiment becomes instrumental in tailoring content strategies to resonate better with audiences. In sectors like finance, sentiment analysis can even serve as a predictor for stock market movements, reflecting public sentiment about companies or industries. Moreover, in a world rife with rapid information dissemination, continuous sentiment monitoring aids in effective crisis management, helping institutions identify and manage potential PR crises swiftly. Beyond the immediate, sentiment analysis of comment sections can reveal deeper, evolving trends in public opinion, offering a nuanced understanding of shifts in societal values and attitudes. In sum, comment sections are more than just feedback mechanisms; they are a gold mine of insights that, when properly analyzed, can greatly influence strategies and decision-making across a spectrum of industries.

## III. Related Works

In aspect-based opinion mining [3], sentiment aggregation is crucial. The traditional arithmetic mean falls short of capturing the majority opinion effectively. To address this, Abbasimehr et al. [4] propose a sentiment aggregation system, incorporating WSAM-OWA, a variation of the OWS operator. WSAM-OWA considers both the majority opinion and information importance in aggregation.

On the lines of trend prediction Matsumoto et al. [5] conducts a study that introduces a model that predicts the popularity of tweets on Twitter by predicting the number of "likes" and "RTs." Also, Jay et al. [6] and Gloor et al. [7] in their papers talk about the different trend prediction strategies.

Previous research has mainly concentrated on identifying and characterizing different forms of negative behaviors, such as hate speech [8], [9], harassment [10], cyberbullying [11], and general toxic behavior [12], [13]. These studies typically analyze the content in isolation and retrospectively [14]. Although they are useful for monitoring and reducing exposure to toxic content, they have limited potential in predicting and preventing toxic behaviors beforehand [15]. To forecast toxicity, it is necessary to consider the social and conversational context in which such behaviors occur. Previous studies on conversational analysis have examined various outcomes, including the growth [16], [17], or re-entry of conversations [16], [18], their productivity [19], controversy [20], [21], and the likelihood of leading to disagreement [22]. Recently, there has been a focus on predicting toxicity by considering pragmatic cues [23] and learned representations [24] of the language used in the initial exchanges of a conversation.

## IV. Methodology

### A. Text classification:

Our task was structured into four primary phases. Initially, we gathered data from three distinct datasets, as detailed in Section V-A. In the second phase, we conducted preprocessing, a critical step in text classification. This process encompassed handling userhandles, punctuation, emojis, hashtags, numbers, and URLs, while also eliminating stop words. Depending on the specific problem, we considered whether to retain or remove special characters and numbers. We also incorporated domain-specific vocabulary into the text. Furthermore, we employed a tokenizer based on a transformer model to prepare the inputs for our model.

Moving on to the third phase, we carried out modeling and analysis. This entailed partitioning the dataset into training, validation, and test sets, as well as fine-tuning hyperparameters for model training and evaluation. Given that we utilized a pre-trained language model, we fine-tuned it using our dataset. Finally, during the experimentation and evaluation phase, we assessed the model's performance and computed emotion scores for the textual data.

### B. Scores Aggregation:

Sentiment analysis is a natural language processing task employed to identify and categorize reviews conveyed in online texts. It aims to classify them into various classes. The choice of an aggregation algorithm significantly influences the effectiveness of sentiment analysis methods. A technique involves a straightforward process of summing up and averaging a total of $n$ replies, including the comment itself. The resulting average scores, represented by a 1x28 shape, are then normalized to obtain the final aggregated scores for the sub-reply tree. This process is repeated for all sub-reply trees using the Bottom-Up approach.

By utilizing methods, score aggregation can be performed for any number of categories. The ultimate score that we aim for, obtained by estimating the emotional score for each comment, is determined by integrating the aggregation method across all comment threads associated with a specific social media post. This combined calculated score is then utilized to predict the current trend of the user's post. In Figure 3, a section of a reply tree is depicted, where each node represents a reply comment. Our machine-learning model has generated scores for each emotion category. Initially, all leaf nodes are assigned emotional scores, and by applying equation (1), we obtain the aggregated score for a particular parent node, including the node itself. This calculation is carried out for every node in the reply tree.

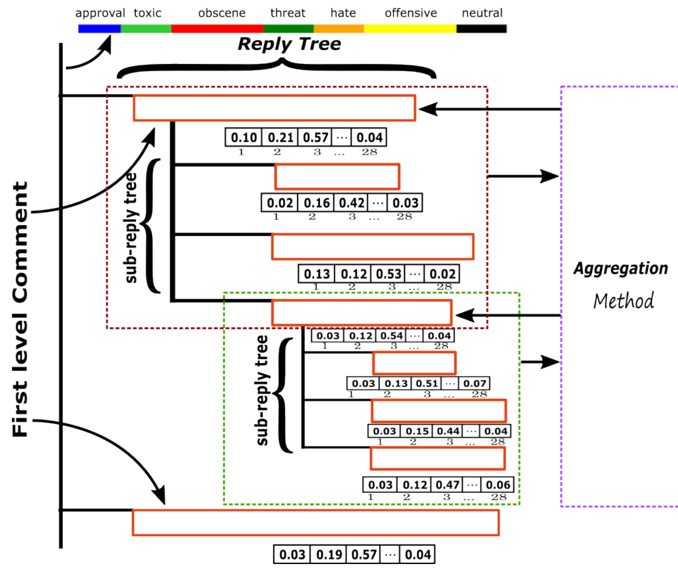


Fig. 3: Comment threads for user post scores aggregation sample representation using different approaches.

### C. Trend Prediction:

After obtaining the results of score aggregation (as described in Section IV-B), our focus shifts to predicting trends based on these derived emotion scores. To accomplish this, we require a model capable of forecasting trend scores using the provided aggregated scores. In Section V-A, we leverage the mentioned dataset to generate emotion scores for each comment by applying our trained Bert-based uncased classification model. By replacing each comment with the 27 emotion scores and one neutral score obtained from the text classification model, we now have a total of 28 features. Among these features, one serves as the target variable representing the desired trend to be predicted.

Subsequently, we proceed to train a machine learning model using Logistic Regression (LR), Support Vector Machine (SVM), Decision Tree (DT), and Random Forest (RF) as experimental approaches to predict the multi-level probability scores based on these features. The trained machine learning model is then saved for future trend score prediction after calculating the emotion scores for each comment or reply and aggregating them into a single set of 27 emotions and one neutral score, as described in Section V-C.

## V. Experiment

### A. Dataset

The ability to comprehend emotions conveyed through language holds numerous practical uses, ranging from the identification of detrimental online conduct to the prediction of social media post trends. Progress in this field can be enhanced by utilizing extensive datasets of significant scale, encompassing a detailed typology that can be easily adapted to multiple subsequent tasks. As a dataset, we have considered three different corpora, (i) GoEmotion: the largest manually annotated dataset of 58k English Reddit comments [25], this

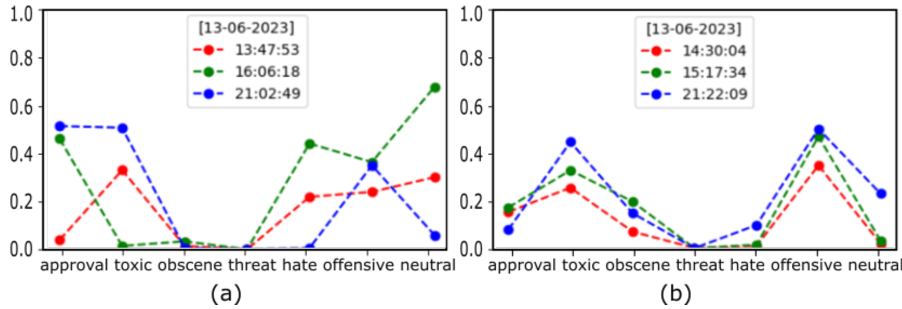


Fig. 4: Trend observations for possible comment threads for user posts based on different timestamps for Fox News articles.

dataset is used to fine-tune the BERT model where this dataset is labeled with 27 emotions and neutral (denoted as D1) (ii) Twitter comments from Kaggle (denoted as D2 in Figure 2) consists of two different datasets one (more than 24k tweets) [26] is labeled by hate, offensive and neither (we renamed 'neither' column to 'neutral') and another (more than 159k Twitter comments) [27] is taken where the labels are toxic, severe_toxic, obscene, threat, insult, identity_attack (renamed the column name to 'hate') and we combined the 'toxic' and 'severe_toxic' column into 'toxic' if any of both contains binary 1. We then added one more trend and its 'approval' which we get by combing the positive [25] emotions 'admiration', 'approval', 'gratitude', and 'realization'. One question could arise Why do these three make one and how? So, the answer is that representing the people's support, agreement, recognition, or acceptance of the user post. (iii) Crawled users' comments from Fox News articles (a. https://tinyurl.com/3tyvp4j7, b. https://tinyurl.com/kbr7vscj) comment section which is used to evaluate the model and generate the comments 27 emotions and neutral scores and then we performed aggregation and applied this dataset in order to get the trend scores (as shown the results in table I and graph in Figure 4). To retrieve data from the official Fox News website, we follow a manual procedure, which involves copying the HTML file that contains the extended comment section. This comment section is organized as an unordered list, with the presence of a reply comment triggering the appearance of another unordered list. This hierarchical pattern is consistent throughout the entire comment section. Once we've identified this hierarchical structure of comments and replies within the HTML file, we employ a recursive approach with the assistance of a Python package. Additionally, we create a Python class that is structured in a tree-like manner. This method enables us to collect and assign all the extracted data to a variable, streamlining subsequent processing. The dataset is then forwarded to the next stage of data preprocessing, which includes tasks like removing punctuation, stopwords, and tokenization.

### B. Model Finetuning

We employ a neural network architecture that comprises a BERT model, a Dropout layer, and a Linear layer to facilitate both regularization and classification. During the forward pass, the BertModel layer generates two outputs. The second output, referred to as the pooled output, is transmitted to the Dropout layer, while the subsequent output is directed to the Linear layer. The final layer possesses a dimension of 28, corresponding to the number of classes for prediction. The output from this final layer is employed to assess the model's performance. Our experiment encompasses the exploration of various configurations, including batch sizes of 8, 16, and 32, learning rates of 1e-5, 2e-5, and 5e-5, 10 epochs of training, and dropout rates of 0.3, 0.5, and 0.9. During the training phase, the model is trained with multiple combinations of hyperparameters, and the best-performing model is saved based on achieving the minimum validation loss.

### C. Result

The Bert base uncased model demonstrates outstanding performance when evaluated with the dataset, achieving a maximum AUC score of 0.91. Leveraging this high-performing model, we conducted an assessment of the dataset described in the Dataset section, which includes annotations for seven distinct trends. The trend scores for various machine learning models are presented in Table I.

To predict multi-level scores, we employed various multi-level classification techniques, such as MultiOutputClassifier (MOC), LabelPowerset (LP), and ClassifierChain (CC), in combination with five classification machine learning models. When analyzing user comments in relation to the post or reply, we utilized the initial state of the comment section from the most recent reference URL (a), as depicted in Figure 4, to generate the trend scores (as illustrated in Table I).

In our pursuit of predicting trends in comment sections using periodic data, we present the trend scores for each timestamp (as shown in Figure 4). In this figure, the color red signifies the initial state when the data was first collected and trend scores were generated. Subsequently, we observe two subsequent states marked as green and blue, representing two other comment section states.

## VI. Ethical Considerations

We have developed our protocols to collect and analyze our data in an ethical, IRB-approved manner. For analysis purposes, we only stored anonymized data that we collected

TABLE I: Performance of ML models

| Models | Trend | | | | | | |
|---|---|---|---|---|---|---|---|
| | Approval | Toxic | Obscene | Threat | Hate | Offensive | Neutral |
| DT | 1.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| LR-CC | 0.5914 | 0.1523 | 0.0536 | 0.0064 | 0.0526 | 0.2424 | 0.0875 |
| LR-LP | 0.6487 | 0.0435 | 0.021 | 0.0037 | 0.0266 | 0.0639 | 0.0281 |
| NB | 0.6598 | 0.1125 | 0.0544 | 0.0035 | 0.0422 | 0.2182 | 0.0468 |
| RF | 0.667 | 0.1216 | 0.0602 | 0.0041 | 0.0471 | 0.2261 | 0.0524 |
| SVC | 0.6418 | 0.1682 | 0.0662 | 0.0152 | 0.0377 | 0.1722 | 0.0566 |

from the different corpora whose handling does not fall within the PII definition of NIST SP 800-122 [28]. Under GDPR [29], the use of the information without context, e.g., name or personal identification number, is not considered to be "personal information".

## VII. Conclusion and Future Directions

In this research, we've delved deep into the intricacies of comment section dynamics, leveraging state-of-the-art techniques to distill overarching sentiments and opinions without the need for exhaustive manual review. Our comprehensive approach encompasses three pivotal stages of trend prediction within these comment sections.

We initiated our process by harnessing a BERT model to yield emotional scores from user comments, setting the stage for the subsequent training of trend prediction models. This layered approach of first extracting emotional context and then leveraging it for trend prediction showcases the sophistication of our methodology.

Representing the comment data as a tree, our methodology systematically evaluates every leaf comment using the trained BERT model, capturing a wide range of 27 emotions alongside neutral scores. The data is then synthesized through an aggregation method, working from the finer details up, ensuring a thorough and comprehensive sentiment analysis.

Moving forward, we're enthusiastic about diversifying our toolkit: we aim to incorporate a variety of machine learning models and neural networks to ascertain the most adept model for our needs. Moreover, we're also in the process of designing an intuitive interface, providing users with a hands-on opportunity to explore the system's capabilities and assess its real-world utility.

## VIII. Acknowledgment

This research was supported by NSF grant CNS-2153482.

## References

[1] Sajedul Talukder and Bogdan Carbunar. A study of friend abuse perception in facebook. *ACM transactions on social computing*, 3(4):1–34, 2020.

[2] Ronen Feldman. Techniques and applications for sentiment analysis. *Communications of the ACM*, 56(4):82–89, 2013.

[3] Somayyeh Aghababaei and Masoud Makrehchi. Mining social media content for crime prediction. In *2016 IEEE/WIC/ACM International Conference on Web Intelligence (WI)*, pages 526–531. IEEE, 2016.

[4] Hossein Abbasimehr and Mostafa Shabani. A sentiment aggregation system based on an owa operator. In *2019 5th international conference on web research (ICWR)*, pages 1–5. IEEE, 2019.

[5] Kazuyuki Matsumoto, Reishi Amitani, Minoru Yoshida, and Kenji Kita. Trend prediction based on multi-modal affective analysis from social networking posts. *Electronics*, 11(21):3431, 2022.

[6] Yaniv Altshuler, Wei Pan, and Alex Pentland. Trends prediction using social diffusion models. In *Social Computing, Behavioral-Cultural Modeling and Prediction: 5th International Conference, SBP 2012, College Park, MD, USA, April 3-5, 2012. Proceedings 5*, pages 97–104. Springer, 2012.

[7] Peter A Gloor, Jonas Krauss, Stefan Nann, Kai Fischbach, and Detlef Schoder. Web science 2.0: Identifying trends through semantic social network analysis. In *2009 international conference on computational science and engineering*, volume 4, pages 215–222. IEEE, 2009.

[8] Thomas Davidson, Dana Warmsley, Michael W. Macy, and Ingmar Weber. Automated hate speech detection and the problem of offensive language, 2017.

[9] Mainack Mondal, Leandro Araújo Silva, and Fabrício Benevenuto. *A Measurement Study of Hate Speech in Social Media*. 2017.

[10] Antigoni-Maria Founta, Constantinos Djouvas, Despoina Chatzakou, Ilias Leontiadis, Jeremy Blackburn, Gianluca Stringhini, Athena Vakali, Michael Sirivianos, and Nicolas Kourtellis. Large scale crowdsourcing and characterization of twitter abusive behavior., 2018.

[11] Mengfan Yao, C. Chelmis, and Daphney-Stavroula Zois. *Cyberbullying Ends Here: Towards Robust Detection of Cyberbullying in Social Media*. 2019.

[12] Jeremy Blackburn and Haewoon Kwak. Stfu noob! predicting crowdsourced decisions on toxic behavior in online games, 2014.

[13] Ellery Wulczyn, Nithum Thain, and Lucas Dixon. Ex machina: Personal attacks seen at scale, 2017.

[14] Paula Fortuna and Sérgio Nunes. A survey on automatic detection of hate speech in text, 2018.

[15] David Jurgens, Libby Hemphill, and Eshwar Chandrasekharan. A just and comprehensive strategy for using nlp to address online abuse, 2019.

[16] Lars Backstrom, Jon Kleinberg, Lillian Lee, and Cristian Danescu-Niculescu-Mizil. Characterizing and curating conversation threads: Expansion, focus, volume, re-entry, 2013.

[17] Justine Zhang, Jonathan Chang, Cristian Danescu-Niculescu-Mizil, Lucas Dixon, Yiqing Hua, Nithum Thain, and Dario Taraborelli. Conversations gone awry: Detecting early signs of conversational failure, 2018.

[18] Sarah Shugars and Nicholas Beauchamp. Why keep arguing? predicting engagement in political conversations online:. *SAGE Open*, 2019.

[19] Vlad Niculae and Cristian Danescu-Niculescu-Mizil. Conversational markers of constructive discussions, 2016.

[20] Mauro Coletto, Kiran Garimella, Aristides Gionis, and Claudio Lucchese. Automatic controversy detection in social media: A content-independent motif-based approach. *Online Social Networks and Media*, 2017.

[21] Jack Hessel and Lillian Lee. Something's brewing! early prediction of controversy-causing posts from discussion features., 2019.

[22] Lu Wang and Claire Cardie. A piece of my mind: A sentiment analysis approach for online dispute detection, 2016.

[23] Justine Zhang, Cristian Danescu-Niculescu-Mizil, Christina Sauper, and Sean J. Taylor. Characterizing online public discussions through patterns of participant interactions, 2018.

[24] Jonathan S. Chang and Cristian Danescu-Niculescu-Mizil. Trouble on the horizon: Forecasting the derailment of online conversations as they develop, 2019.

[25] Dorottya Demszky, Dana Movshovitz-Attias, Jeongwoo Ko, Alan Cowen, Gaurav Nemade, and Sujith Ravi. Goemotions: A dataset of fine-grained emotions. *arXiv preprint arXiv:2005.00547*, 2020.

[26] ANDRII SAMOSHYN. Dataset: Hate speech and offensive language dataset, kaggle, 2019.

[27] Jigsaw/Conversation AI. Dataset: Jigsaw unintended bias in toxicity classification, kaggle, 2019.

[28] Guide to protecting the confidentiality of personally identifiable information (pii). https://tinyurl.com/ylyjst5y, 2021. Accessed: 2021-02-12.

[29] General data protection regulation (gdpr). https://gdpr-info.eu/, 2021. Accessed: 2021-02-12.