Automatic classification of humor using Covid-19 related tweets

Anonymous ACL submission

Abstract

With the increased usage of social media platforms owing to the pandemic and lockdown situations all around the world, we have observed an exponential increase in the production of content revolving around the pandemic. With a huge chunk of this content falling under the genre of humor, especially from content creators, government institutions, entertainment companies, etc., this study aims to create a dataset and refine a preexisting hierarchical annotation schema in an attempt to accomplish universality. Using the dataset of 5118 tweets, containing both humorous and non-humorous data, scraped using pandemic-related keywords from Twitter, we develop a refined hierarchical annotation schema and attempt to use multi-label multi-class machine learning models to classify the humorous tweets based on four levels of annotation viz., type, technique, situation, and relevance. The dataset is also analysed to find a correlation between the tweets of a certain type of humor and the situation of the pandemic. Further, this study discusses the importance of contextual knowledge in order to achieve the above-mentioned objective. The use case for the findings of this study includes but is not limited to individuals or organizations who are involved in discussions regarding the crisis, brands for building digital marketing strategies, and behavioral researchers. The classification models can also be used to aid further research in the field of humor.

1 Introduction

Humor is a complex phenomenon, and its classification and analysis are challenging tasks. The ability to recognize and produce humor is an essential aspect of social intelligence, and it is often used as a measure of mental agility and creativity. Thus, the study of conversational humor is of significant interest to researchers in various fields, including linguistics, psychology, and artificial intelligence. Linguists have long been interested in the structure and function of humor, focusing on the linguistic devices used to create and express humor. They have been concerned with finding patterns and defining rules about what we find funny. More specifically, many linguists find interest in studying how grammatical manipulations affect humor generation. Sociologists are interested in how it influences society at large. Psychologists have studied humor from a cognitive and affective perspective, examining how people perceive, process, and respond to humorous stimuli. Meanwhile, researchers in artificial intelligence and natural language processing have explored the potential of computational methods for analyzing and generating humor.

Researchers have previously tackled humor detection tasks by framing them as binary classification problems and training BERT-based models to achieve satisfactory accuracies. However, classifying humor based on its characteristics, such as type and technique, presents an even more daunting challenge, particularly given the absence of a formal taxonomy of humor characteristics (Yang et al., 2015). It is also important to note that most humor detection tasks have been performed on canned jokes, one-liners, and similar sources, resulting in more homogeneous data sets.

Refining this binary classification approach to not only detect humor but also classify humorous utterances based on their type and technique and further understanding the context and relevance of these utterances is crucial for gaining a comprehensive understanding of humor. Such classification systems can also provide insights into topics that may be considered sensitive or potentially offensive to specific sections of society.

In this study, we go beyond the binary classification of humor and attempt multi-class, multi-label classification on a manually annotated dataset of Covid-19-related tweets.

2 Objective

Conversational agents and chatbots are becoming increasingly common in various domains, including customer service, education, and entertainment. These systems rely on NLP and machine learning techniques to understand and generate human-like responses. However, they often struggle with understanding and generating humor, which can make interactions with users less engaging and enjoyable. A possible solution to developing a feeling of connection is to add an element of humor. By developing machine learning-based approaches to conversational humor classification, we can improve the performance of conversational agents and chatbots and make them more user friendly.

Humor is an important coping mechanism that can help individuals reduce stress, improve mood, and enhance social connections. By developing conversational agents and chatbots that can generate and understand humor, we can potentially improve the mental health and well-being of users.

Given these needs, comprehensive data annotation methods, which are neither constrained by domain nor by language, are a necessity. Furthermore, given the extensive use of ML models for executing the above-mentioned advancements, we can foresee the need for a massive amount of data. Annotation of this data manually is a herculean task as well as a futile undertaking. The use of state-of-the-art models to automatically tag data is hence, a prerequisite in order to smoothly carry out this task.

3 Literature Survey

Numerous studies have been conducted on the types of humor in both canned jokes as well as conversational humor. There are innumerable classification schemes and a wide variety in the set of types given in each study. Hay (1995), Martin et al. (2003), Dynel (2009) and many others have each come up with a new set of types in their studies. While Hay (1995) agrees for the need of a unified theory and a thorough, well defined, taxonomy, they also argue that most taxonomies are designed for a specific context and fail to provide adequate coverage of different datasets. While this is the case with most taxonomies, for this paper we have chosen to follow Pamulapati et al. (2020)'s annotation schema containing type and technique tags for conversational humor data. This schema seemed to comprehensively be able to tag all occurrences of humor in our dataset while taking into account the context, cultural significance, benignity, etc. of each utterance. The hierarchical annotation schema, which, when further refined as part of this study, facilitated our goal of achieving a multi-class, multi-label classification of humorous data.

Humor as a coping strategy against the adverse psychological outcomes of a worldwide pandemic is advocated by numerous research scholars. Reizer et al. (2022) for example, studies the direct and indirect associations between humor and optimism in predicting well-being during Israel's lockdown period. Their findings suggest that individuals who maintained a sense of humor during this challenging period experienced better psychological well-being and were better equipped to handle the stress and uncertainty associated with the lockdown. Alarcon et al. (2013), Martin and Ford (2018), Kuiper (2012) are some of the few others who strongly favor arguments emphasizing the vital role of humor in managing stress and fostering resilience during crises. The prevalence of humor-related content on social media platforms during the pandemic suggests that individuals were actively seeking out and engaging with humorous content to cope with the stressful circumstances brought on by the pandemic.

Computational humor analysis and recognition is a slightly more complex task as the perception of humor varies largely from person to person. However, some of the earliest attempts for automatic humor recognition were contributions of Taylor and Mazlack (2004) ; Mihalcea and Strapparava (2006). For any work on humor recognition to be fruitful, it needs to be paired with an accurate analysis of why something is funny. Ahuja et al. (2018) address this question of what makes us laugh in their work. Similarly, Mihalcea and Strapparava (2006) study humor specific stylistic features like alliteration, antonymy and slang for charecterizing humorous texts.

In addition to rule-based and machine learning-based systems, researchers have also explored the use of deep learning techniques for humor recognition and generation (Hossain and Muhammad, 2020). Deep learning models, such as recurrent neural networks (RNNs), long short-term memory (LSTM) networks, and transformers, have shown promising results in various natural language processing tasks, including humor analysis. Moreover, researchers have attempted to address the challenges associated with the subjectivity and context-dependency of humor by incorporating additional contextual information into their models. Cultural differences and variations in humor perception have been acknowledged in computational humor studies. To address this challenge, researchers have experimented with cross-cultural and multilingual humor datasets. One such example is the work by Potash et al. (2017), who developed a cross-cultural humor dataset containing jokes from various countries and languages and employed machine learning algorithms to classify humor across different cultures.

Finally, a study similar to ours had been conducted by de Haas (2016) during the Ebola crisis. While they discuss the change in humor styles during the Ebola crisis, we aim to do a similar study during the Covid-19 pandemic and compare and contrast results. Our study additionally considers factors like contextual knowledge and tries to find a correlation not just between the stage of development of the crisis and the type of humor but also between the type of humor and the topic being discussed.

4 Dataset and annotation schema

Given that Covid-19 is a relatively recent phenomenon, to the best of our knowledge, this dataset of jokes revolving around the pandemic is the first and only publicly available dataset. However, creating such a dataset comes with its own challenges. Creating a humorous tweets dataset in the domain of Covid-19 requires careful consideration of these challenges to ensure that the content is appropriate, ethical, and relevant to the current situation. Covid-19 is a serious and sensitive issue that has affected people's lives worldwide. Therefore, creating humorous content around such a topic can be perceived as insensitive, inappropriate, and even offensive. Creating funny tweets requires a deep understanding of the context, audience, and cultural nuances. In the case of Covid-19, it can be challenging to create humorous content that resonates with people across different regions, cultures, and languages. Humorous content about Covid-19 can easily cross the line and become offensive. Therefore, ensuring that the content is ethical, responsible, and respectful is also essential. This was ensured by both annotators manually verifying each utterance during the process of annotation. However, given the varying degree of tolerance of each individual, it is impossible to claim that any humor dataset can be completely appropriate across all ages, genders, cultures etc.

This Covid-19 tweets dataset contains 1395 humorous tweets and 3723 non-humorous tweets comprising news items, country-wise statistics, etc. This paucity in the number of humorous tweets collected can be attributed to the sensitive nature of the topic. However, it is noteworthy that this dataset gives a comprehensive overview of all the tags mentioned in the annotation schema with a considerable number of utterances obained for each tag. We collected the humorous data using Twitter scrapers to extract tweets using keywords related to the pandemic, like "corona jokes," "pandemic humor," and other similar words. The output was then manually filtered to remove non-humorous tweets like news articles, sub-tweets without context, external links, etc. These keywords were selected based on the understanding that if they were found as hashtags in the tweet, then the tweet would most likely be humorous. This further reduced the task of manually filtering the non-humorous tweets from the scrape results. The results contained tweets from the general public from all over the world, as no location filter was implemented. For the non-humorous data, we used tweets from various news channels and official accounts reporting statistics or country-wise updates regarding the pandemic.

The time period over which tweets were collected ranged from February 2020, around which time the virus was first detected, to December 2021. The tweets collected during this span captured various stages of the pandemic, including the three major waves, vaccination drives, lockdown situations, travel bans, etc.

4.1 Hierarchical Annotation Schema

For the purpose of this study, we refine the annotation schema provided by Pamulapati et al. (2020). For the task of refining the schema, we have taken into consideration the "type" and "technique" tags and added the tags of "situation" and "relevance" to the same level. These tags give contextual information, which, as discussed earlier, is crucial in understanding why and to whom an utterance is humorous. Accordingly, the data from the Covid-19 tweets dataset was annotated using six tags for each tweet. The tags were first to identify if the tweet was humorous or not. Next, we identify the type and if the technique used to make the tweet funny is direct or indirect. Further, we also specify the exact direct/ indirect technique used. The fifth tag identifies the situation on which the tweet is based. This gives us an idea about the status of the pandemic. And the last relevance tag identified whether the tweet is universally relevant or is only specific to a particular culture. Fig 1 shows the refined schema containing three levels.

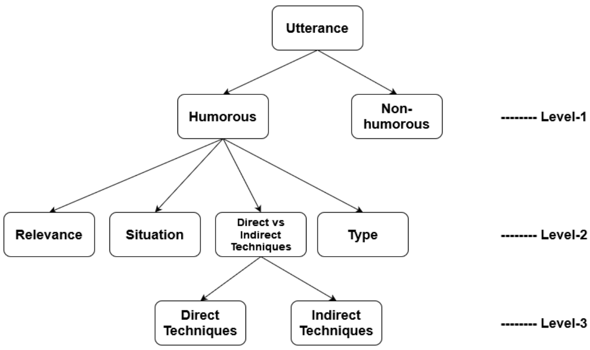

Figure 1: Modified schema containing three levels

4.1.1 Level-1 Humorous vs Non-humorous

The Covid-19 tweets dataset contains 1395 humorous tweets and 3723 non-humorous tweets. The first level of annotation was to classify the tweets as humorous or not. However, this level itself is not as seemingly straightforward. Apart from humor's subjective nature, the chosen domain stands as a problem in this case. The pandemic was a grave time for a majority of the population all around the world. Millions of people fell sick, and a significant number of people lost their lives. Though light-hearted, creating humor in such cases may sometimes come across as insensitive. Hence, tagging such data even manually comes with a lot of challenges, and automating the process further is undeniably arduous.

4.1.2 Level-2 Type

This study focuses on four types of humor generated by the utterances: Teasing, Retort, Banter, and Schadenfreude. Understanding these humor types is essential for analyzing the social dynamics and psychological effects of humor in various contexts. This set of tags comes from a hierarchical annotation schema for conversational humor given by Pamulapati et al. (2020). The list of types includes Teasing, Retort, Banter, and Schadenfreude.

4.1.3 Level-2 Direct vs Indirect techniques

Humor techniques can be broadly classified into two categories: direct and indirect. These approaches play a vital role in shaping how the audience conveys and interprets humor. The classification is based on the presence of pretense or any hidden meaning in the utterance. This is further classified into more specific techniques in Level-3 of the annotation schema.

Direct humor techniques are characterized by their straightforward and unambiguous nature, with no hidden meaning or pretense. In these cases, the meaning of the utterance is taken at face value, without any need for interpretation or inference.

On the other hand, indirect humor techniques involve an element of pretense or subtlety, with the intended meaning of the utterance differing from its literal interpretation (Muller, 2011). In these instances, the speaker tries to convey something more than the apparent meaning of the utterance.

4.1.4 Level-2 Situation

The situation tag can be explained as the topic that is being discussed as part of the conversation/ joke. This tag gives us an idea about the stage of development in the pandemic. For example, if the topic is about the lockdown, we can understand that the pandemic is just developing, and the virus is spreading rapidly. However, if the topic is about reopening schools and offices, we know that it is only towards the end of a wave when cases have started to decline.

We have tagged our dataset using the following topic tags. Health/ Hygiene, Lockdown, Testing/ Vaccination, Quarantine, Government Orders, Eating Habits, Online School, Work From Home (WFH), Hoarding Groceries, Social Media, Social Distance, and Miscellaneous.

To be able to make any assumptions from a trending topic or even to be able to find the tweet funny and deduce a trending topic from it, contextual knowledge and an awareness of what is happening around the world are fundamental. This trending topic can be universally significant or only be confined to a specific culture. It is found in this study that there is a high interdependence between the type of humor and the trending topic.

4.1.5 Level-2 Relevance

The importance of context in humor cannot be understated, as it significantly impacts how jokes are processed and appreciated. This final tag provides information about the significance of a trending topic across the globe, shedding light on the contextual and cultural aspects that influence humor perception. According to Jiang et al. (2019), understanding the contextual factors that shape humor appreciation is essential for both researchers and practitioners, as it helps identify culture-specific topics that may be humorous within a particular cultural context but not necessarily to those outside that culture.

This is what gives importance to this tag as this helps us identify culture specific topics that may not even be funny to people outside that specific culture. It is also to be noted again that relatability and contextual knowledge are important to understand the humor in the tweet. This understanding has practical implications for content creators and marketing agencies, who can leverage the power of humor more effectively by considering the cultural and contextual factors that shape humor appreciation.

4.1.6 Level-3 Techniques

Direct humor techniques include exaggeration, fallacious reasoning, allusion, profanity, and stylistic figures. Whereas, since indirect humor techniques often require the audience to engage in cognitive processes, such as inference and interpretation, in order to appreciate the humor (Suls, 1972), utterances employing indirect humor techniques can be further classified as either satire or sarcasm.

Note : Examples for each tag are given in appendix A

5 Inter Annotator Agreement

This dataset was tagged manually by two annotators, A1 and A2. Cohen's Kappa values for each class were calculated separately. For level-1, i.e., identifying humorous tweets, the Kappa value was calculated to be 0.92 (almost perfect agreement). This can be attributed to the fact that determining the non-humorous data came with no ambiguity. Any instances of news, statistics, etc., could be directly tagged as non-humorous. Further, since the humorous tweets were scraped using specific keywords, the instances where either one of the annotators found such a tweet unfunny though it existed were significantly low.

At the second level, the Kappa values for Relevance and Situation were relatively higher, with 0.9 and 0.85, respectively. Since both the annotators have similar cultural backgrounds in terms of age, educational background, country, social background, etc., identifying cultural nuances in the humorous tweets posed no ambiguity. The tags for the situation were also straightforward. Since there could be more than one tag situation tag associated with each tweet, it could be possible that either annotator failed to recognise the second situation resulting in a slightly lesser score.

The Kappa value for identifying the type of humor was calculated to be 0.52 (Moderate agreement). The issue discussed by Pamulapati et al. (2020), which resulted in the low Kappa value for the Type tag, is seen in this case as well. Namely, the failure to identify a particular type and giving the NULL tag, and an overlap between retort and teasing tags stand as the main concerns. Finally, the identification of direct and indirect techniques resulted in a Kappa value of 0.62 (Substantial agreement). And the tagging of the specific techniques resulted in a Kappa value of 0.58(Moderate agreement). This can be attributed to issues like a failure to recognise allusion, misunderstanding wordplay, differences in humor perception in cases like exaggeration, etc.

6 Analysis

Based on the analysis of our tagged dataset, it has been observed that satire and sarcasm are the most prevalent humor techniques, accounting for 27.1% and 27.3% of the total data, respectively. Schadenfreude, defined as the pleasure derived from others' misfortune, emerged as the most common humor type, with 29.5% of the utterances being tagged as such. This observation can be justified considering the context and timing of these jokes, which occurred amidst panic and chaos caused by the global crisis. The presence of schadenfreude aligns with the benign violation theory discussed by McGraw and Warren (2010), which posits that humor arises when a situation is perceived as both a violation of norms and benign simultaneously. The prevalence of satire and sarcasm can also be explained by previous studies that have analyzed humor during challenging times, revealing similar trends. (de Haas, 2016), (Cai et al., 2014).

Our data supports the idea that humor development follows a pattern of an initial increase after the onset of a crisis, followed by a decrease. This observation can be substantiated by examining the number of occurrences of each trending topic. For instance, the beginning of the crisis was marked by a high number of jokes related to the lockdown situation. However, the frequency of such jokes gradually decreased as vaccines became available and the situation evolved.

Furthermore, our analysis revealed a clear correlation between the type of humor and the topic being discussed. Conversations about the lockdown situation have a high occurrence of schadenfreude, possibly reflecting individuals' attempts to cope with the challenges and frustrations brought about by the restrictions. In contrast, conversations about the government and its policies are often characterized by satirical humor, which serves as a means of critiquing and questioning the decisions made by those in power.

In conclusion, examining our tagged dataset provides valuable insights into the humor techniques and types employed during times of crisis, highlighting the prevalence of satire, sarcasm, and schadenfreude as ways of coping with the challenges and uncertainties faced by individuals and societies. The observed correlations between humor types and the topics being discussed underscore the role of humor as a versatile tool for communication, critique, and emotional regulation in the face of adversity.

7 Methodology

The process of analyzing and understanding humor in the context of Covid-19 related tweets begins with the preprocessing of the dataset. The dataset includes both humorous and non-humorous tweets, many of which contain elements that may interfere with the model's ability to recognize and process the humor present. These elements include external links, hashtags, emojis, and other non-textual components that are not directly relevant to the humor analysis task at hand. Consequently, these elements must be carefully removed during the data preprocessing step, ensuring the dataset is cleaned and ready for annotation.

Once the data has been cleaned, it is then annotated to provide valuable information about the humor present in each tweet. The annotation process involves categorizing tweets according to different dimensions of humor, such as type, technique, and situation. However, given the low number of humorous tweets available in the Covid-19 domain, some adjustments had to be made to the annotated tags to ensure the model could be effectively trained. Low representation of specific tags can lead to imbalanced data, which may adversely affect the model's ability to generalize and accurately predict the corresponding labels.

Tag	Labels	No. of occurrences
Type	Schadenfreude	449
	No value	943
Technique	Sarcasm	387
	Satire	376
	Stylistic Figures	189
	Allusion	164
	Exaggeration	50
	No value	226
Situation	Lockdown	562
	Health/ Hygiene	368
	Misc	295
	Work from home	100
	Government	67
Relevance	Universal	1237
	Cultural	155

Table 1: Frequencies of each tag after data manipulation

All tags with fewer than 50 instances were recategorized during the annotation process to address this issue. In the case of situation tags, they were labeled as miscellaneous, while for type and technique tags, they were marked as unknown or null. This approach allowed for more balanced training data and helped improve the model's ability to accurately predict humor-related categories across a broader range of tweets. The edited dataset's statistics containing the labels and the number of occurrences of each label in each tag can be found in the table 1. Though this reduces the scope of the study, it was done due to the absence of a sufficient number of jokes based on the pandemic to train the models. Further, to contain the scope of this paper to multi-class, multi-label classification we refrain from discussing the binary classification task of humorous vs non-humorous given that such a classification had already been explored. (Ahuja et al., 2018), (Kumar et al., 2022).

The dataset is split into training and test sets in the ratio of 80:20 and this ratio strikes a balance between having enough data for training and ensuring a reasonable-sized test set for evaluation. It helps in evaluating the model's performance, avoiding overfitting, enabling fair comparisons, utilizing data efficiently, and facilitating iterative model improvement. Two different types of machine learning models were employed to approach the multi-label multi-class classification for type, technique, and situation and cultural relevance. The first type consisted of basic models using K-Nearest Neighbors (KNN) and decision trees. Using these basic models, the study aimed to establish a foundation for more advanced approaches to humor classification. All the tweets are vectorized by using Tf-idf vectorization and the tf-idf vectors are used as the input features for KNN and Decision trees. This is beneficial for multiple reasons such as feeding numerical inputs to the models, dimentionality reduction, feature importance as tf-idf assigns higher weights to words that are more informative etc. This can contribute to improved performance and better interpretability in humor classification tasks.

The second type of models employed in the study included three BERT-based models: BERT, RoBERTa, BERTweet to achieve the said tasks. These advanced models leverage pre-trained language representations and fine-tuning techniques to achieve state-of-the-art performance on a wide range of natural language processing tasks. By utilizing these BERT-based models, the study aimed to capture more nuanced aspects of humor and gain a deeper understanding of the various dimensions of humor present in Covid-19 related tweets.

7.1 Finetuning the models

The pre-trained BERT models are not task-specific, meaning they do not have any knowledge about the specific task that they are being used for. Fine-tuning is the process of taking the pre-trained BERT model and adapting it to a particular task by training it on a smaller dataset that is specific to the task. This process allows the model to learn the nuances of the task and improve its performance on that task. We fine-tune the three BERT-based models we have by taking the following steps.

- **Attention pooling:** We employ attention pooling technique to aggregate information from different tokens in the sequence. This is done to better capture the humor-related aspects of the text by focusing on important tokens.

- **Focal loss:** The data we have is highly imbalanced for the labels “Relevance" and “Situation" as can be seen in table 1. Hence, it becomes extremely important to tackle the imbalance carefully to prevent the model from being skewed to a particular class. By employing focal loss, we assign higher weights to the misclassified or hard-to-classify examples. By increasing the weight for these difficult examples, focal loss encourages the model to focus more on learning from them. This effectively reduces the dominance of well-classified examples and helps in dealing with class imbalance. It also down-weighs easy examples so that their contribution to the overall loss is reduced.

- **Optimizer:** We set an optimizer with learning rate and warmup scheduler rather than using the default optimizer which helps in adapting the learning rate during training.

8 Results and analysis

The accuracies obtained can be seen in table 2 and table 3.

The results obtained predictably show that the BERT-based models outperformed the classic ML models, such as KNN and Decision trees. This is because the BERT-based models use BiDirectional Encoder Representations, as mentioned earlier, to understand the context of the tweet and nuances of the language, which is essential for understanding humor. They capture long-range dependencies between words, making them better suited for understanding the context of the tweet and nuances of language in humor classification.

Traditional machine learning models like KNN and decision trees performed well in binary classification tasks because they can learn to distinguish between two classes based on a set of input features. As there are only two possible outcomes in binary classification, it becomes easier for these models to learn the decision boundary that separates the two classes. However, they struggle in the multi-class multi-label classification as it requires more complex decision boundaries.

Fine-tuning the pre-trained BERT-based models for our specific task improved the models' accuracy as we added a few layers on top of the existing ones specific to our task. This process allowed the models to adjust their weights and parameters to understand the context better and identify patterns in the data. We also dealt with issues such as imbalances in the data, capturing humor-related aspects in the tweet, etc., by using fine-tuning techniques.

Model	Type	Technique	Situation	Relevance
KNN	0.68	0.29	0.29	0.88
Decision tree	0.57	0.29	0.21	0.86
BERT	0.68	0.46	0.45	0.85
RoBERTa	0.67	0.44	0.48	0.88
BERTweet	0.67	0.41	0.37	0.88

Table 2: Accuracy for multi-class multi-label classification

Model	Type	Technique	Situation	Relevance
BERT	0.70	0.49	0.46	0.89
RoBERTa	0.65	0.48	0.55	0.93
BERTweet	0.69	0.43	0.41	0.89

Table 3: Accuracy for multi-class multi-label classification after fine-tuning

As a result, the accuracy of all the models had increased for multi-class labels by around 5%, which is significant considering the size of the dataset at hand.

9 Conclusion and future work

To summarise, the study of humor detection and classification is an important and growing research area with significant potential to expand our understanding of humor and its various manifestations in society. By refining classification models, incorporating context-aware approaches, and exploring humor across diverse cultural settings, researchers can contribute to developing more sophisticated systems that can better capture and appreciate the complexities of humor.

The enhancement of the annotation schema for conversational humor through the addition of 'situation' and 'relevance' tags, and its subsequent validation on the Covid-19 tweets dataset, has resulted in a more comprehensive schema that accounts for the influence of culture on humor perception. By successfully validating this schema on a unique dataset of humorous tweets related to Covid-19, we have not only demonstrated its wide applicability but also created a valuable resource for future research. This updated annotation schema captures the nuances of humor across different contexts and cultures, thus paving the way for deeper insights into the nature of humor in a global context.

Incorporating both basic and advanced machine learning models in the humor classification process allowed the researchers to leverage the strengths of each approach, ultimately leading to a more accurate and comprehensive understanding of humor in the context of Covid-19 related tweets.

While the study has made significant progress in understanding and classifying humor during the pandemic, there remains ample room for further exploration. Future work could focus on expanding the dataset by collecting and annotating more tweets, ensuring a diverse and representative sample. Future research could investigate the performance of models such as GPT-3, T5, and XLNet in the context of humor classification, potentially leading to further improvements in classification accuracy and generalizability.

Moreover, the insights from this research can also inform the design of artificial intelligence systems and conversational agents capable of engaging in humorous interactions with humans. to become more relatable, engaging, and effective in various applications, ranging from customer service to mental health support.

Overall, the multi-class, multi-label classification of humorous tweets related to Covid-19 represents a significant step forward in understanding humor's complex and dynamic nature in the context of crises. By pursuing the future research directions outlined in this section, we can continue to advance our knowledge in this area and improve the classification models and their applications.

References

Vikram Ahuja, Taradheesh Bali, and Navjyoti Singh. 2018. What makes us laugh? investigations into automatic humor classification. In *Proceedings of the Second Workshop on Computational Modeling of People's Opinions, Personality, and Emotions in Social Media*, pages 1–9.

Gene M Alarcon, Nathan A Bowling, and Steven Khazon. 2013. Great expectations: A meta-analytic examination of optimism and hope. *Personality and individual differences*, 54(7):821–827.

Huajian Cai, Yu Luo, and Vincent Ng. 2014. Humor during challenging times: A study of chinese people with cancer. *Health Communication*, 29(9):908–918.

MM de Haas. 2016. Change in humor styles on twitter during development of ebola crisis.

Marta Dynel. 2009. Beyond a joke: Types of conversational humour. *Language and Linguistics Compass*, 3(5):1284–1299.

Jennifer Hay. 1995. Gender and humour: Beyond a joke. *Unpublished Master's thesis, Victoria University of Wellington, Wellington, New Zealand*.

Md Moinul Hossain and Golam Kibria Muhammad. 2020. Deep learning techniques for humor recognition and generation: A survey. *Artificial Intelligence Review*, 54(3):2389–2421.

Tonglin Jiang, Hao Li, and Yubo Hou. 2019. Cultural differences in humor perception, usage, and implications. *Frontiers in psychology*, 10:123.

Nicholas A Kuiper. 2012. Humor and resiliency: Towards a process model of coping and growth. *Europe's Journal of Psychology*, 8(3).

Vijay Kumar, Ranjeet Walia, and Shivam Sharma. 2022. Deephumor: A novel deep learning framework for humor detection. *Multimedia Tools and Applications*, 81(12):16797–16812.

Rod A Martin and Thomas Ford. 2018. *The psychology of humor: An integrative approach*. Academic press.

Rod A Martin, Patricia Puhlik-Doris, Gwen Larsen, Jeanette Gray, and Kelly Weir. 2003. Individual differences in uses of humor and their relation to psychological well-being: Development of the humor styles questionnaire. *Journal of research in personality*, 37(1):48–75.

A Peter McGraw and Caleb Warren. 2010. Benign violations: Making immoral behavior funny. *Psychological science*, 21(8):1141–1149.

Rada Mihalcea and Carlo Strapparava. 2006. Learning to laugh (automatically): Computational models for humor recognition. *Computational Intelligence*, 22(2):126–142.

Eve Muller. 2011. Pretense. In *The Oxford Handbook of Philosophy of Language*, pages 601–617. Oxford University Press.

Vaishnavi Pamulapati, Gayatri Purigilla, and Radhika Mamidi. 2020. A novel annotation schema for conversational humor: Capturing the cultural nuances in kanyasulkam. In *Proceedings of the 14th Linguistic Annotation Workshop*, pages 34–47.

Peter Potash, Alexey Romanov, and Anna Rumshisky. 2017. Cross-cultural humor prediction. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 2419–2429.

Abira Reizer, Yifat Munk, and Lotem Katz Frankfurter. 2022. Laughing all the way to the lockdown: On humor, optimism, and well-being during covid-19. *Personality and Individual Differences*, 184:111164.

Jerry M Suls. 1972. A two-stage model for the appreciation of jokes and cartoons: An information-processing analysis. *The psychology of humor: Theoretical perspectives and empirical issues*, 1:81–100.

Julia M Taylor and Lawrence J Mazlack. 2004. Computationally recognizing wordplay in jokes. In *Proceedings of the Annual Meeting of the Cognitive Science Society*, volume 26.

Diyi Yang, Alon Lavie, Chris Dyer, and Eduard Hovy. 2015. Humor recognition and humor anchor extraction. In *Proceedings of the 2015 conference on empirical methods in natural language processing*, pages 2367–2376.

A Appendix

Model architechture for BERT-based models is represented in figure 2

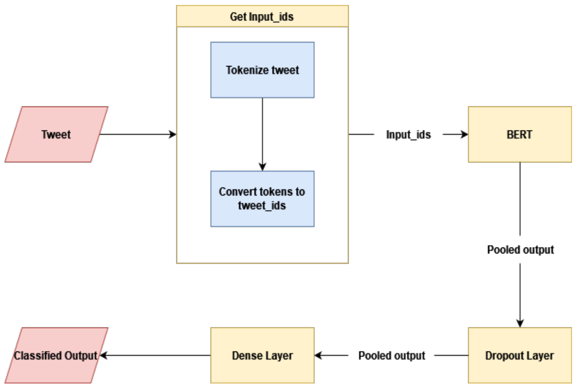

Figure 2: Flowchart depicting the model architecture

Examples tweets for each tag can be found in tables 4, 5 and 6

Type	Example
Teasing	Feb : Ok, Boomer. April : Are you OK, Boomer?
Retort	I'll try being nicer if you'll try being smarter
Banter	Exchange of teases and retorts
Schadenfreude	"We became like the money in our banks. Existing but untouchable!"

Table 4: Examples for each 'type' tag

Situation	Example
Lockdown	It's time for an exciting adventure, move to the other side of the couch.
Health/ Hygiene	1665 - learnt about the existence of microorganisms 1892 - discovered viruses 1928 - discovered Penicillin (the first antibiotic) 2020 - learnt how to wash hands properly
Work From Home	Somebody stole my lunch out of the fridge at work today. The worst part about it... I'm working from home.
Social distance	Practice social distancing? Buddy, I've doing that my whole life!
Government	2019: Netflix and chill 2020: Government coronavirus press conference and cry
Eating habits	Nothing says pandemic desperation like eating a pint of Cinnabon frosting with your bare hands...
Testing/ Vaccine	I got my COVID test today, it says 50. What does that mean? Also, my IQ test came back positive.
Hoarding Groceries	Kinda wish I hadn't rid myself of those yellow and white pages. They would have come in handy for toilet paper..... and reading material.
Social Media	It's time we extended the curfew to Whatsapp Groups.
Quarantine	Chuck Norris got the Coronavirus. The virus is now in quarantine for a month.
Online School	There will be no homeschooling for the rest of the month due to technical difficulties. Technically I'm finding it difficult to be good at it.
Misc	Every time I see my email inbox, I feel flattered that so many brands that I don't even use are so sincerely standing by me during this crisis.

Table 5: Examples for each 'situation' tag

Technique	Example
Exaggeration	Weird year this is. January ended in a couple of weeks, February finished in just a few days, while March dragged on for about a year.
Fallacious Reasoning	Cancelled all my weekend plans because of coronavirus. Staying home. Once this dumbass idiot coronavirus goes away, then I'll think of other reasons to cancel plans.
Allusion	An Englishman, an Irishman and a Scotsman walk into a bar. Ah, those were the days...
Profanity	Covid-19 is like my dick. It suddenly shows up and ruins peoples' lives.
Stylistic Figures	From "Lockdown may get extended" to "Lockdown extended to May".
Sarcasm	Really looking forward to being bad at making eye contact again.
Satire	If you want people in Mumbai to stay at home, announce elections instead of lockdown.

Table 6: Examples for each 'technique' tag