

# Guilt Detection in Text

Ruojia Tao  
Cornell tech master student  
[rt456@cornell.edu](mailto:rt456@cornell.edu)

Yaxuan Huang  
Cornell tech master student  
[yh798@cornell.edu](mailto:yh798@cornell.edu)

Anqi Dong  
Cornell tech master student  
[ad829@cornell.edu](mailto:ad829@cornell.edu)

Hongjiao Zhang  
Cornell tech master student  
[hj452@cornell.edu](mailto:hj452@cornell.edu)

## Abstract

*We are trying to use the deep learning model and natural language processing model to implement a model that tries to classify specific emotions, such as guilt.*

## Keywords

Deep learning, Natural language processing, Sentiment analysis, Emotion recognition

## 1. Introduction

Current focus of deep learning and the natural language processing model is on helping artificial intelligence understand human emotion. Nowadays, artificial intelligence can easily understand the positive and negative emotions in humans' words. However, it is still hard for them to understand complex emotions, for example, guilt. By classifying complex emotions, we can expand the usage of artificial intelligence into more areas. Guilt emotion arises when people feel they did something wrong. This emotion has a high correlation with mental illness. With the background that people are likely to real their emotions more in online social media, if we can classify more emotions on the posts from online social media, we may able to detect posts from people that have suicidal tendencies and help them.

## 2. Related Work

Our main experiment is based on this paper published in March 2023:

Guilt Detection in Text: A Step Towards Understanding Complex Emotions [1].

This paper has information about the literature review, available datasets, baseline models, and possible future directions of the model developments. This paper provides us with a good starting point. We can use the background information to give it to us to find the resource we need and also use the paper included in the literature review as the reference.

To have a whole picture of text-based emotion detection, a review called Text-based emotion detection: Advances, challenges, and opportunities[2] is considered.

This paper introduces the main approaches adopted by researchers in the design of text-based emotion detection systems. It further discusses some recent state-of-the-art proposals in the field. Also, it presents the comparison between different emotion-labeled data sources, which is a great reference for choosing our dataset.

Another paper relevant to specific negative motion detective:

ReDDIT: Regret Detection and Domain Identification from Text [3]

This paper implements a deep learning model mainly focused on the regret emotion which may be different but similar to our goal and we can use similar ways to train the deep learning model.

Apart from specific motion detection, this paper discusses the overall comparison among models:

Comparative Analyses of Bert, Roberta, Distilbert, and Xlnet for Text-Based Emotion Recognition[4]

This paper analyzes the efficacy of BERT, RoBERTa, DistilBERT, and XLNet pre-trained transformer models in recognizing emotions from texts. The implemented models are fine-tuned on the ISEAR data to distinguish emotions into anger, disgust, sadness, fear, joy, shame, and guilt. This paper finds that the recorded model accuracies in decreasing order are 0.7431, 0.7299, 0.7009, 0.6693 for RoBERTa, XLNet, BERT, and DistilBERT, respectively.

There are also some papers talking about guilt in a physiological way. That is related to our project.

## 3. Methodology

The feature engineering methods they used are TF-IDF and bag of words, which are basic NLP methods. We will try encoding features with TF-IDF and S-Bert as feature engineering.

For the deep learning models, in the paper above they implemented a model of CNN with BiLST and we will start

with a model and try different deep learning models to see the accuracy of each model.

3.1. Datasets

Currently, the datasets we have are:

Vent [5]: a large annotated dataset of texts (texts are from 33M posts), emotions, and social connections from the Vent social media platform. In Vent, each post is associated with emotion, self-annotated with emotion by the post’s author at the posting time. The problem with this dataset is it needs permission to download the data and we will try to apply for it.

Tweet\_eval[6] from hugging face: this is a multi-classification task from hugging face the tasks include different emotions. If we cannot get permission of using the Vent dataset we will use this dataset.

4. Risks

Since the vent dataset is not a public dataset, it is possible that we cannot get permission immediately after we applied for it, so we probably have to change our topics and plan if we cannot get the dataset on time. Currently, we get the dataset so it is no more a risk, we will delete this risk in future versions.

Also, the current deep learning models do not work well on the dataset and their output accuracy is only about 20 percent, which makes it hard to tell how well we can do.

5. Timeline

Table 1. Timeline for project

week 1	Finalize the dataset we need to use and finalize project scope
week 3	Finish data analysis and feature engineering
week 5	Build up CNN model and evaluate the perform tuning the parameters and model to improve performance
week 7	Try another model and replicate the evaluation : and tuning steps
Future Weeks	If we have more time we will try to improve t performance of the current model we have or we build other models, depend on current mode performance

6. Ethical Considerations

When conducting research on detecting guilt in text, there are several ethical considerations that must be taken into account.

The first problem is Privacy and Confidentiality. Since the dataset is collected from the Vent social media platform, It is essential to ensure that the data collected for this project

is obtained with the consent of the participants and that their privacy is protected. The collected data should be anonymized to prevent the identification of individual participants. Additionally, the data should be stored securely to prevent unauthorized access. Therefore, if we have the chance to get data from Vent, we need to keep it in good use.

In addition to that, we need to pay more attention to the analysis more than the result. We might not be able to offer instant help as soon as we predict the tendency out of our model, along with the fact that the model result might not be accurate. Therefore, we should view the data more as the research data rather than the practical data.

References

[1] Meque, Abdul Gafar Manuel, et al. "Guilt Detection in Text: A Step Towards Understanding Complex Emotions." arXiv preprint arXiv:2303.03510 (2023).

[2] Acheampong, Francisca Adoma, Chen Wenyu, and Henry Nunoo-Mensah. "Text-based emotion detection: Advances, challenges, and opportunities." Engineering Reports 2.7 (2020): e12189.

[3] Balouchzahi, Fazlourrahman et al. "ReDDIT: Regret Detection and Domain Identification from Text." ArXiv abs/2212.07549 (2022): n. pag.

[4] Adoma, Acheampong Francisca, Nunoo-Mensah Henry, and Wenyu Chen. "Comparative analyses of bert, roberta, distilbert, and xlnet for text-based emotion recognition." 2020 17th International Computer Conference on Wavelet Active Media Technology and Information Processing (ICCWAMTIP). IEEE, 2020.

[5] Lykousas, N., Patsakis, C., Kaltenbrunner, A. & Gómez, V. Sharing Emotions at Scale: The Vent Dataset. Proc. Int. AAAI Conf. on Web Soc. Media 13, 611–619, DOI: 10.1609/icwsm.v13i01.3361 (2019).

[6] "Tweet\_eval · Datasets at Hugging Face." Huggingface.co, huggingface.co/datasets/tweet\_eval.

[7] Balahur, A., Hermida, J. M. & Montoyo, A. Detecting implicit expressions of sentiment in text based on commonsense knowledge. In Proceedings of the 2nd Workshop on Computational Approaches to Subjectivity and Sentiment Analysis (WASSA 2.011), 53–60 (Association for Computational Linguistics, Portland, Oregon, 2011).

[8] Haile, Sileshi Bogale et al. "Emotion Classification for Amharic Social Media Text Comments Using Deep Learning." SSRN Electronic Journal (2022): n. pag.

[9] Mubeen, S. Mohammad Malik et al. "Linguistic Based Emotion Detection from Live Social Media Data Classification Using Metaheuristic Deep Learning Techniques." International Journal of Communication Networks and Information Security (IJCNIS) (2022): n. pag.

[10] Mubeen, S. Mohammad Malik et al. "Linguistic Based Emotion Detection from Live Social Media Data Classification Using Metaheuristic Deep Learning Techniques." International Journal of Communication Networks and Information Security (IJCNIS) (2022): n. pag.

[11] Abbaschian, Babak Joze et al. "Deep Learning Techniques for Speech Emotion Recognition, from Databases to Models." Sensors (Basel, Switzerland) 21 (2021): n. pag.