Stress Detection using Social Media Posts

11.12.2025

Abstract

This project investigates whether stress expressed in social media posts can be automatically detected using machine learning models. Using Reddit and Twitter datasets introduced by Rastogi et al., the task is to classify social media post text as either expressing stress or not. This project compares two machine learning models: BLSTM and BERT. The models are trained separately on Reddit data, Twitter data, and a combined dataset to evaluate both the predictive accuracy and also the social media platform specific influence on the models. The results show that BERT performs the best across all datasets, achieving the highest validation accuracy and loss. Surprisingly, the Reddit trained models perform the best even though they have less training data than the combined dataset trained models. Additional tests on custom sentence inputs reveal interesting platform specific behavioral differences between the models.

Contents

1 Introduction

Nobody can escape stress. It's a prevalent thorn in everyday life for many and has a negative effect on peoples' mental well-being. With the rising popularity of social media, it has become a way for many to share their stress and find comfort from others. This creates a potential avenue for others to detect those in stress and in need of help. This raises an interesting question: can we harness the superior computing power of computers to help with this problem?

In recent years, stress detection from social media has become an active research area within natural language processing (NLP). Earlier work has shown that machine learning models can capture cues from text that correlate with different psychological states with some studies using posts from social media platforms like Reddit and Twitter. However, different platforms show different writing styles and word choices. This raises an interesting question: how well do models trained on one platform generalize to another, and do platform-specific patterns influence the models predictions?

In this project, I explore these questions by training and comparing two types of machine learning models for text-based stress detection using the datasets introduced by Rastogi et al.[1]. The goal is to classify posts as either "stress" or "no stress," evaluate how well the models perform, and investigate how their behavior changes when trained on Reddit, Twitter, or a combined dataset.

The project uses two widely studied model types: a Bidirectional Long Short-Term Memory network (BLSTM) and BERT (Bidirectional Encoder Representations from Transformers). BLSTM serves as a traditional neural network model baseline that learns language representation directly from the training data, while BERT represents a more modern and powerful transformer-based approach with strong contextual understanding learned through pretraining.

In addition to evaluating model accuracy and loss, this project also examines model behavior using custom test sentences to highlight potential biases or platform-specific tendencies. This analysis helps answer a broader question: do social media platforms differ in how users express stress using them?

2 Related work

A number of studies have explored machine learning based automatic stress detection and mental health analysis using social media data. I have chosen 4 of these articles that I deemed to be the most relevant to borrow ideas for this project.

The work most closely related to this project is done by Rastogi et al. [1], who are the authors of a research paper titled “Stress detection from social media articles: New dataset benchmark and analytical study”. In this paper, they introduce the Reddit and Twitter datasets used here. They also compare several machine learning models and show that BERT variants perform really well on this data.

A second relevant article is by Suhail et al[2], and is titled “A Comparative Study of Sentiment Analysis for Mental Health Related Posts at Reddit & Twitter Using Machine Learning and Pre-Trained Models”. This paper investigates how effectively social media posts on Reddit and Twitter can be used to detect anxiety and depression. The paper compares different pre-trained machine learning models for automatic detection of these mental health issues.

A third relevant paper is by Jadhav et al. [3], and is titled “Text Based Stress Detection Techniques Analysis Using Social Media”. This paper reviews different machine learning methods for detecting stress by analyzing textual social media data such as tweets, comments and chat messages. This study compared three different machine learning models and concluded that the BLSTM model performed the best. The findings from the paper motivate the use BLSTM in this project.

Finally, the fourth paper is by Febriansyah et al. [4] and is titled “Stress detection system for social media users”. This study focuses on detecting stress using different machine learning models by using a Reddit post dataset for training classical machine learning models. Their work show that even simple machine learning models, such as SVMs (Support Vector Machines), can reach a good performance for this kind of task.

3 Problem formulation

The main objective of this project is to train two machine learning models on social media post text data to classify text as either expressing stress or not. We will train the two machine learning models on this task, evaluate their performance, and compare them with each other. However, the goal of this project is not only to build accurate classifiers, but also to analyze the models' behavior when trained on different datasets. This brings us to the two central questions of this project:

1. How well do the BLSTM and BERT models perform on the combined dataset compared to the individual Reddit and Twitter datasets and how much better does BERT perform overall?

2. Do the models learn platform specific patterns, and how does the behavior change when trained on the individual Reddit and Twitter datasets?

To explore these questions, this project compares two models from different machine learning model families: RNN (Recurrent Neural Network) based BLSTM and transformers based BERT.

In addition to quantitative performance metrics such as validation loss and accuracy, this project also examines the qualitative performance of the models by using custom test sentences tailored to highlight different preferences of the models. This uncovers potential social media platform induced biases in the models, and it also shows how well the models generalize to normal stress-related sentences.

4 Dataset description

The dataset used for this project consists of stress related social media posts from two large social media platforms, Reddit and X, built by Rastogi et al.[1]. These posts are labeled with either 0 or 1, indicating whether the post expressed stress or not. The Reddit dataset has a total of 5556 observations, and the Twitter dataset a total of 2051 observations. Similarly, the number of words in each dataset are 98 953 with a mean word post length of 17.81, and 52 117 with mean of 25.41 respectively.

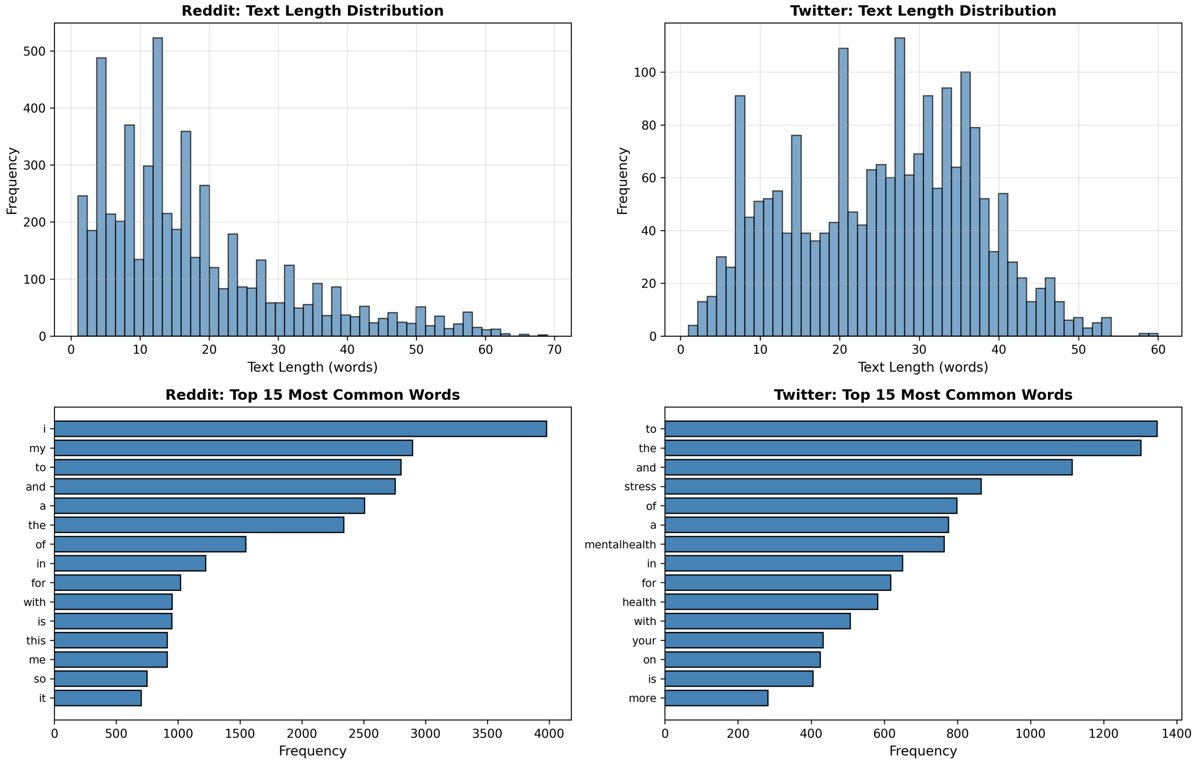

Figure 1: Raw Dataset exploration

Observing the raw data plots in Figure 1, we can see that the datasets are slightly different. For example, Reddit posts seem to be quite short compared to Twitter posts. Also, when observing the top 15 words plots, the Reddit plot consists of only short meaningless words while the Twitter plot includes more substantial words, like ”stress” and ”mentalhealth” for example. After preprocessing the data and plotting again in Figure 2, we can see that the trends remain. Reddit words are still mostly neutral, while Twitter words carry more information.

Figure 3 displays the label distributions in each dataset. The Reddit data is very balanced, but the Twitter data has a large label discrepancy, having many more ”stress” labels than ”no stress” labels. However, after combining the two datasets into one, the percentage difference between the labels comes down to only 47.3% to 52.7%, which makes the final dataset almost balanced.

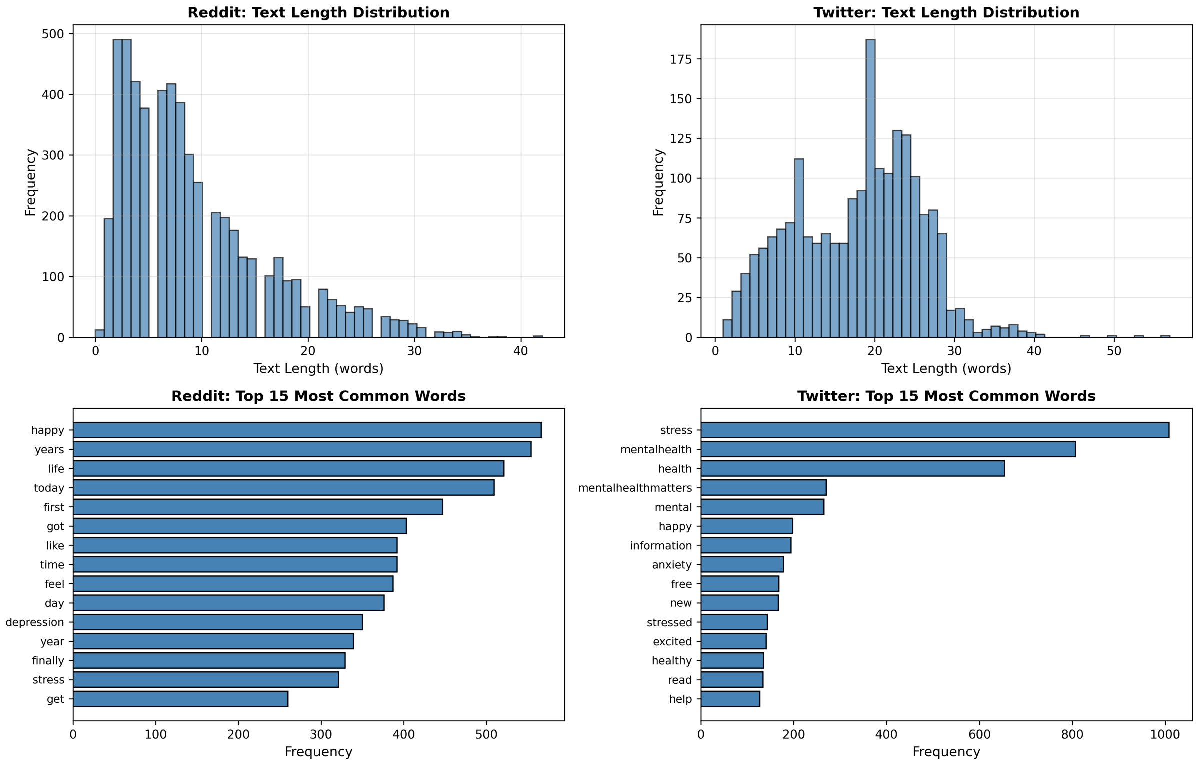

Figure 2: Preprocessed Dataset exploration

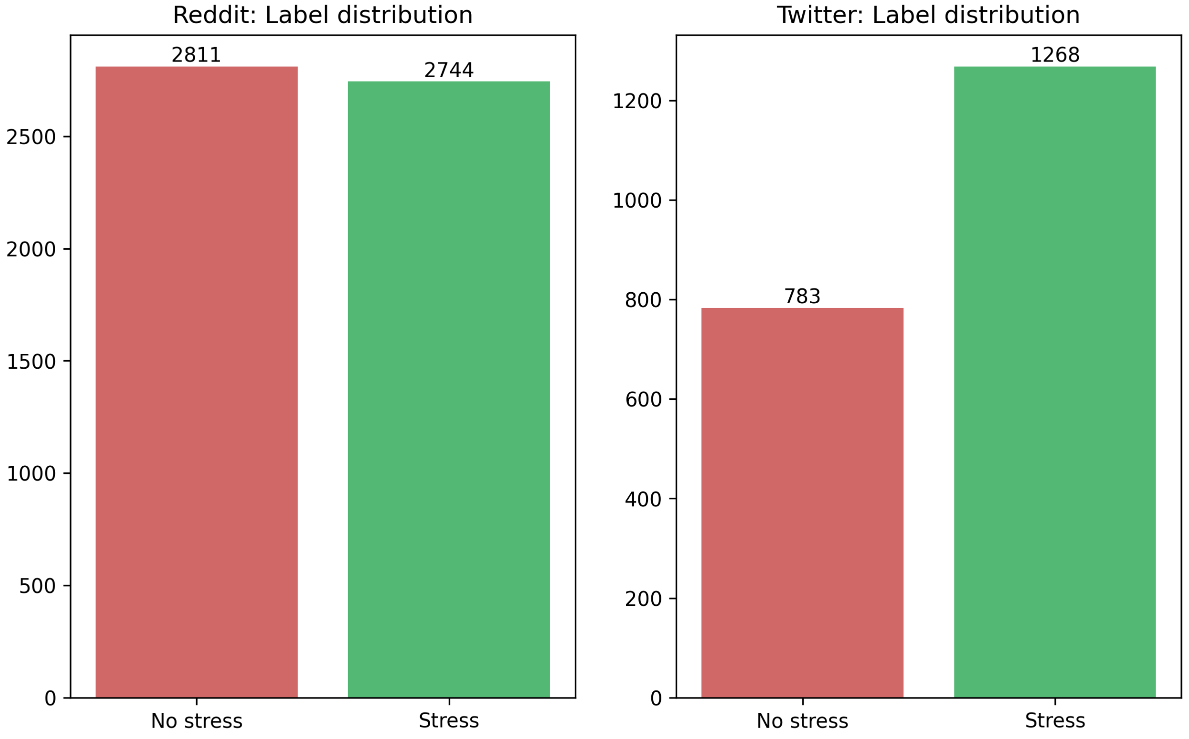

Figure 3: Dataset label distributions

5 Methods

We train two models: BLSTM and BERT. A key difference between the models is that BERT comes with pre-trained contextual embeddings trained on huge amounts of textual data so it already has a deep understanding on word meanings and contextual differences. On the other hand, our BLSTM model learns its embedding values only from our training data, meaning that it becomes specialised only for the task that we train it for.

5.1 BLSTM

LSTM is a supervised machine learning method, which is an RNN, meaning that it has a hidden state acting as a memory, that allows the network to use past computations in the current step. The main idea behind LSTM is the ability to maintain a memory of past observations while also taking in new information when making decisions. In our case this means that the model remembers what words have previously occured in the input sequence while progressing through the words (tokens). BLSTM is a modified version of LSTM that input bidirectionality. This means that the input fed into the neural network contains the text two times: once normally and once with the words reversed. This allows the model to understand the context of each word better, which is important for a classification task like ours.

5.2 BERT

BERT is also a supervised machine learning method, but unlike BLSTM, it is not an RNN. BERT is a transformer based model that maintains context using self-attention instead of a hidden state like LSTMs. Transformers process all words in the input sequence simultaneously rather than reading them from left to right, like RNNs. This allows the BERT model to understand the full context of each word.

Another important difference is that BERT produces contextual embeddings. This means that the vector representation of a word changes depending on the sentence it appears in. In contrast, typical LSTM models use static embeddings, where each word has the same vector regardless of its context. Contextual embeddings allow BERT to capture different meanings of the same word more effectively.

BERT also differs from LSTM models in how it learns its language representation. It does not learn from scratch. BERT is pretrained on a massive text corpus and already has an excellent general language understanding. The model training just finetunes it to work with our classification task.

6 Results

6.1 Model comparison

Model	**Reddit**	**Twitter**	**Combined**
BLSTM	0.902	0.786	0.874
BERT	0.955	0.903	0.936

Table 1: Validation accuracies for BLSTM and BERT across datasets.

Model	**Reddit**	**Twitter**	**Combined**
BLSTM	0.248	0.503	0.303
BERT	0.146	0.233	0.170

Table 2: Validation losses for BLSTM and BERT across datasets.

The training validation losses are a great way to estimate how well the model performs. Table 2 shows the validation losses of each trained model. We can see that the BERT validation losses are lower on each of the datasets, which was expected. Quite surprisingly, the models trained on only the Reddit dataset have the smallest validation loss, even though that dataset is roughly 27% smaller compared to the combined dataset.

To test, whether some interesting insights can be gained from the models, we test the Reddit and Twitter BERT models on custom text inputs focusing on different topics and compare them.

Even though neither models statistics show no clear bias for either of the categories, there seems to be clear pattern. Twitter BERT seems to predict stress for events that should be stress-free in nature, while Reddit BERT correctly gives these a no-stress label. Both models predict the test sentences correctly.

Sentence	Topic	BERT Reddit	BERT Twitter
I am so happy and fun today	no stress test	no stress	no stress
The sun is shining and the weather looks nice today	no stress test	no stress	no stress
My friend died in a car crash	stress test	stress	stress
I got married today to my wife	marriage	no stress	stress
My husband surprised me with a breakfast this morning	marriage	no stress	stress
I saw my friend today and we hanged out	friends	no stress	stress
I went for a coffee with my best friend today	friends	no stress	stress
I have a lot of work piling up	work stress	stress	stress
They are firing a lot of people at my work	work stress	stress	stress

Table 3: Predictions of BERT Reddit and BERT Twitter models on custom test sentences.

7 Conclusion and Discussion

7.1 Model limitations

The biggest limiting factor for all of our models is the size of our dataset. Especially the LSTM models, that must learn their language representation solely from the data they are given, struggle with smaller amounts of data. The Twitter models' performance reflects this issue quite clearly, but this is not their only problem.

An interesting result seen in Tables 1 and 2 is that the amount of training data is not the sole factor driving the performance of the models. The Reddit models have the best performance with both of our model types. This means that the Twitter dataset has some characteristics that make learning difficult for the models. One reason could be that because the Twitter dataset has, on average, longer texts, the model struggles to find the components of those texts that contribute to the right classification.

The class imbalance in the Twitter dataset probably also negatively affects the model performance. Even though the models don't become imbalanced themselves, the imbalance of having many more stress labels could also participate in the frequent stress predictions observed in Table 3.

7.2 Model insights

The custom input test shown in Table 3 provides some potentially interesting insights into the differences between the two social media platforms. Based on the results, it appears that Twitter users seem to share more negative aspects of their lives than positives. This is evidenced by the fact that the BERT model, which was trained on Twitter data, classified sentences about marriage or friends as stress related, whereas the model trained on Reddit data classified the same sentences as no stress. It must be acknowledged that this test was very small, so no certain conclusions can be drawn from the results. However, the results do suggest a trend, which could certainly be plausible.

7.3 Potential improvements

As discussed in the model limitations section, the main bottleneck for the models is dataset size. The clearest improvement idea is therefore to use a larger dataset. This data does not have to be social media post data; classified stress-related text data would also benefit the models in their main task.

If one wants to get more out of BERT, using stronger and more modern BERT variants would most likely improve the model performance. The research paper by Rastogi et al.[1] showed that DistilBERT performed the best out of their models on this same task.

Finally, the behavior difference between the Reddit-trained and Twitter-trained model found in this project suggests that this area could be interesting to continue studying in future work.

References

[1] A. Rastogi, Q. Liu, and E. Cambria, "Stress detection from social media articles: New dataset benchmark and analytical study," in *Proceedings of the 2022 International Joint Conference on Neural Networks (IJCNN)*, IEEE, Jul. 2022, pp. 1–8. DOI: `10.1109/IJCNN55064.2022.9892889`.

[2] L. Suhail, S. Masood, and A. Haider, "A comparative study of sentiment analysis for mental health related posts at reddit and twitter using machine learning and pre-trained models," *Journal of Innovative Computing and Emerging Technologies*, vol. 4, no. 2, Oct. 2024. DOI: `10.56536/jicet.v4i2.128`.

[3] S. Jadhav, A. Machale, P. Mharnur, P. Munot, and S. Math, "Text-based stress detection techniques analysis using social media," in *2019 5th International Conference on Computing, Communication, Control and Automation (ICCUBEA)*, IEEE, Aug. 2019, pp. 1–5. DOI: `10.1109/ICCUBEA47591.2019.9129201`.

[4] M. Febriansyah, Nicholas, R. Yunanda, and D. Suhartono, "Stress detection system for social media users," *Procedia Computer Science*, vol. 216, pp. 672–681, 2023. DOI: `10.1016/j.procs.2022.12.183`.