

# Identifying Depression on Social Media

Kali Cornn (kcornn)  
Department of Statistics

## Background

- Millennials may turn to social media for help or discussion of mental health issues. Previous work has found Facebook statuses can reveal depressive symptoms.
- Reddit is unique in that users are able to post on “throwaway” accounts not associated with their main account.
- Goal: develop models taking in a Reddit text comment and outputting classification (“depressed”, “non-depressed”)

## Data

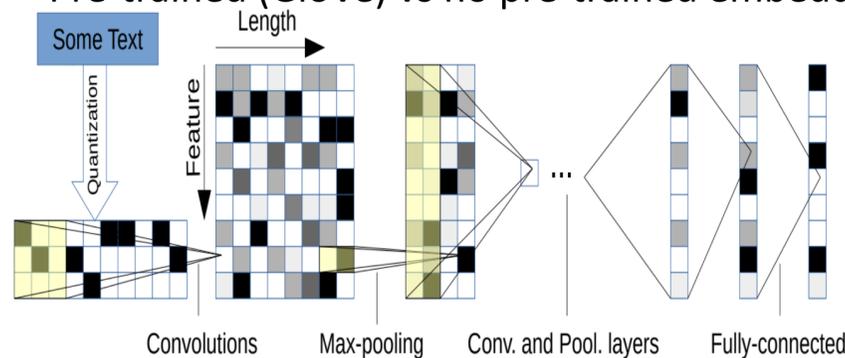
- Scraped comments from two subreddits: /r/depression and /r/AskReddit; all initially unlabeled.
- Labeled comment 1 if “depressed” (/r/depression), 0 if “non-depressed” (/r/Askreddit)
- Shuffled and split data into 80%-10%-10% ratio for training, dev, and test sets, which had 143,712, 47,904, and 47,905 comments, respectively.

## Baseline Models

- Converted comments to a matrix of tf-idf features with L2 norm.
- Logistic regression** utilized inverse of regularization strength  $C = 1$ .
- SVM with linear kernel** utilized penalty parameter of the error term  $C = 0.25$ .

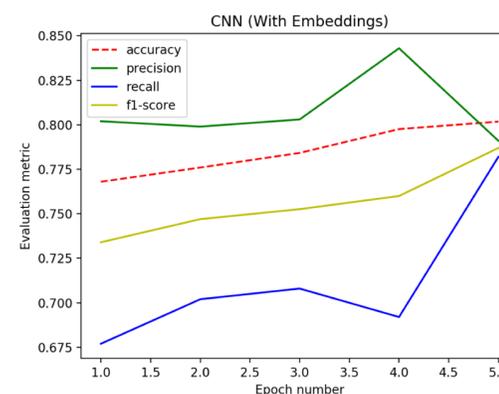
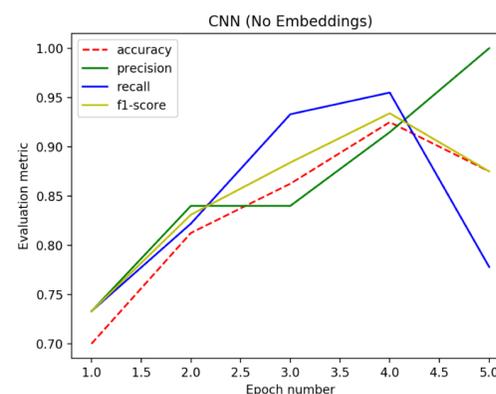
## Models

- BERT** model is pre-trained language representation model that can be fine-tuned with a single output layer that performs binary classification [1].
- Character-level CNN** consisted of 6 convolutional layers and 3 fully-connected layers based on [2].
  - Pre-trained (GloVe) vs no pre-trained embeddings



## Test Results

Model	Accuracy	Precision	Recall	F1-Score
LogReg	0.848	0.856	0.813	0.834
SVM	0.850	0.854	0.821	0.837
BERT	0.857	0.849	0.844	0.847
<b><u>CNN (no embed)</u></b>	<b><u>0.925</u></b>	<b><u>0.915</u></b>	<b><u>0.956</u></b>	<b><u>0.935</u></b>
CNN (w/ embed)	0.776	0.799	0.702	0.747



## Discussion

- Advantage of using BERT: is bidirectional – can access context from “past” and “future.”
- CNN model with GloVe embeddings performed worst, but plots of evaluation metrics were smoother over the 5 epochs.
- CNN model without GloVe embeddings performed well, at the expense of high training time.

## Future Work

- Create dataset by using human evaluation rather than automatically assigning comment to a category (depressed/non-depressed) based on subreddit.
- Run CNN models for a longer time.
- Use SGD as an optimizer and decrease the learning rate over time.
- While the CNN performed well, explore other neural network architectures such as CNN-LSTM encoder-decoder model, which has been effective in learning Tweet embeddings.

## References

- [1] Devlin, J., Chang, M., Lee, K. and Toutanova, K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding.
- [2] Zhang, X., Zhao, J. and LeCun, Y. (2019). Character-level Convolutional Networks for Text Classification.