# Using Deep Learning with Unsupervised Clustering for Suicide and Depression Identification

Texas A&M
Department of Computer Science and Engineering

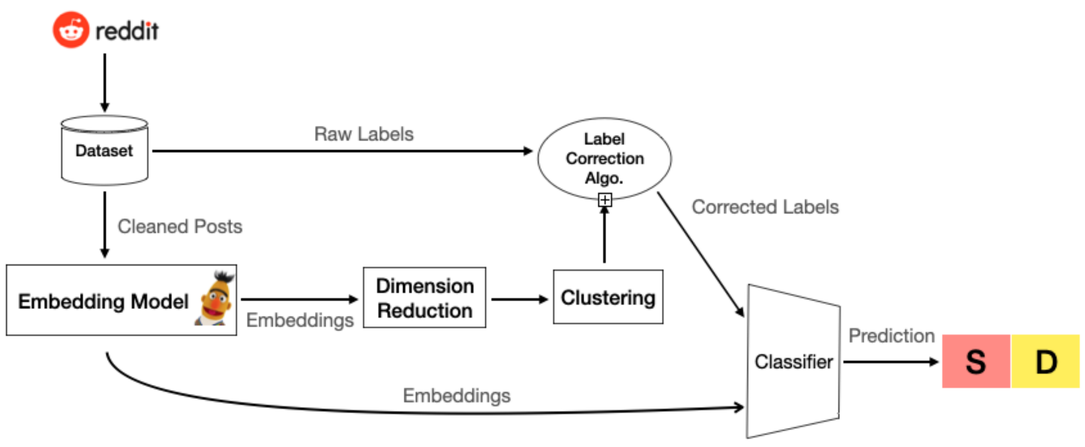
Professor: Dr. James Caverlee

Aaryan Kothapalli, Jia-Hua Cheng, Chien-Peng Huang

## Overview

- **Objective**: to differentiate between suicide and depression given a sentence in a post/tweet/social media outlet.
- **Solution**: utilize unsupervised clustering-based label correction process and neural network sentiment analysis to classify texts between depression and suicidal.

## Methodology

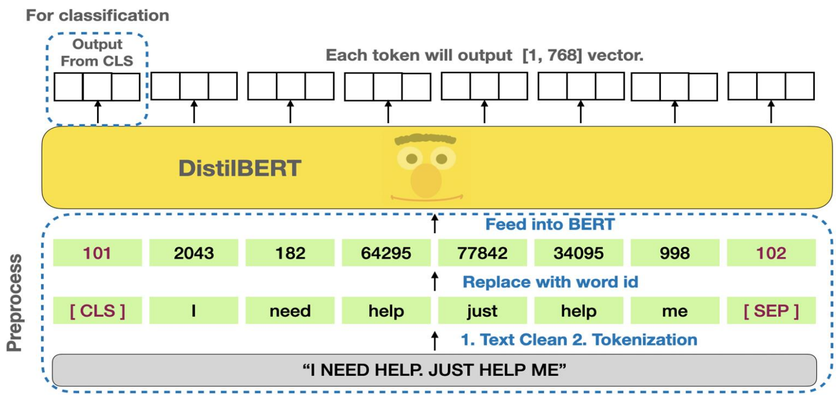



- **Libraries Used:** NumPy, Pandas, PyTorch, TF, Matplotlib, and Scikit-learn

- **About the Dataset:**
  - Name: *Suicide and Depression Detection*
  - Source: Dataset on Kaggle, sourced directly from Reddit's r/depression
  - Content: Suicide or depression sentences, differentiated by 1 for depression, or 2 for suicide.
  - Date period: January 2009 – January 2021
- **Preprocessing:**
  - Filter content length < 100 words for simplicity (~150,000 data)
  - Cleaning and removing noise:
    - URLs: since URLs are not valuable for this problem, we remove them
    - Emoji: for simplicity, we remove emoji. There is an alternative way for emoji, which is to replace emoji with meaningfully related word.
    - Contraction: we unpack contraction words (e.g., I'll -> I will, you're -> You are)
    - Punctuation: it is removed most of the time. But if we think it relates to emotion, we keep it.
    - Case normalization: we make every word lowercase

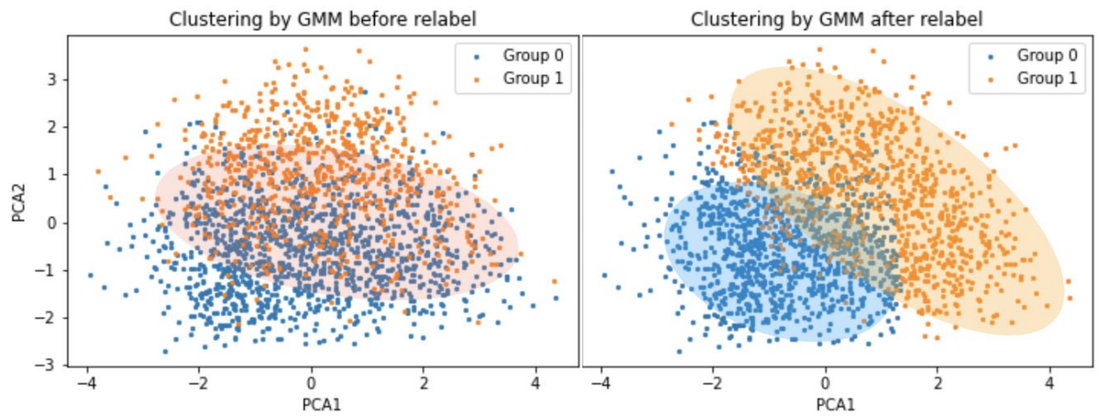
| Cleaned Sentence | label |
|---|---|
| i need help just help me I am crying so hard | 1 |
| I am fucked assignment is due tomorrow and I haven not … | 0 |
| I accidently cut my dick with scissors help | 0 |
| it ends tonight. I can not do it anymore. I quit. | 1 |
| today is our project presentation…What if my professor does not like it… | 0 |

- **Embedding Models:**
  - BERT (Bidirectional Encoder Representations from Transformers) is a popular transfer-based word embedding model. Instead of proceeding word by word sequentially like RNN/LSTM, it totally avoids recursion, by processing sentences as a whole and by learning relationships between words.
  - We feed 150,000 posts into distilBERT to extract only the dimension with [CLS] for our classification task. The final output after BERT is a (150000, 768) vector. We then merge with one label column from the raw dataset. This (150000, 768+1) dataset is the cleaned, sentence embedded representation.




## Methodology, cont.

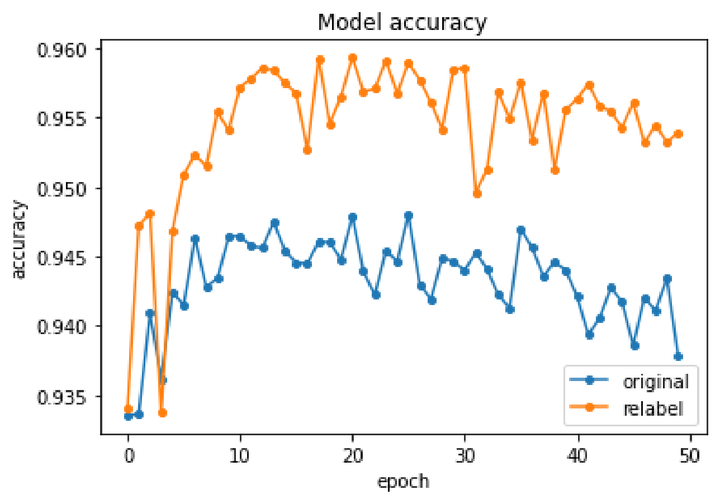
- **Confidence Correction with GMM clustering:**
  - Steps:
    - Decompose dataset to 2D by PCA
    - Cluster the data into two groups by Gaussian Mixture Model
    - Relabel the data which has different label with its initial and GMM.
  - Details:
    - Input dataset is (n, 768). n = number of posts, 768 = number of features extracted from BERT. Output = (n, 2) as we use PCA to decompose the dataset into 2D.
    - Cluster the data into 2 groups through GMM.
    - Relabel the data with confidence threshold 95%.




- **Classifier:**
  - With the corrected label set, we train the data in a deep neural network to determine whether the sentences portray depressive or suicidal sentiment. To achieve this, we used a Fully Dense Neural Network (DNN) given its historical performance with similar datasets.
  - The neural network takes in an input of shape ( ,768). 768 = number of features. And there are about 140xxx rows, each with 768 features.
  - The input is split into Train, Valid, and Test split to account for bias and accuracy.
  - The network is composed of an input layer of size 768, a hidden ReLu layer of size 128, and another hidden ReLu layer of size 64. Output layer is a sigmoid activation layer.
  - Adam optimizer is used with binary cross entropy for loss.

## Results

| Method(s) used | Testing Accuracy (over 50 epochs) |
|---|---|
| DNN classification without GMM clustering confidence correction | 94%, loss=0.244 |
| DNN classification with GMM clustering confidence correction | 96%, loss=0.199 |




Our Fully Dense Network with unsupervised GMM clustering confidence correction achieved a **96%** testing accuracy in successfully determining whether input sentences portray depressive or suicidal sentiment.

## Acknowledgment

- It is important to note that there may be false positives or false negatives, given the way we gathered data (Reddit is an informal social site).
- This project should not be used as a primary classification tool to differentiate between depressive and suicidal sentiment analysis given that all input data was sourced from a single site.
- Our team did not infringe on any copyright in the making of this project. We used open-source libraries for all code that was implemented. Everything was obtained legally and ethically.