Depression Among Users of Social Media

By: Torrence Barbour, Satwik Nijampudi, Shanthan Gunti, Ryan Edwards

Research Question:

Given a social media post that was posted, can we detect if the user suffers from depression?

Overview

- Our project goal is to find out if a trained AI can pick out any signs of depression from a social media post
- Depression has been one of the major causes of death and major sources of stress and anxiety among teenagers in recent years.
- According to the National Institution of Mental Health, approximately 21 million people in 2020 experienced at least one two-week long depressive episode in the US alone.
- With the advancement of technology in the field of AI and NLP, we believe that we can implement filters in social media platforms like Reddit to identify signs of depression and try to offer them the resources required to improve their condition.

Machine Learning Algorithm to be Used:

- Bayesian Network or Support Vector Machine
- Both will be implemented under Supervised Learning
- Bayesian Networks essentially take different aspects of the text and filter them through weights and probabilities in order to determine the likelihood of a given result. In this case, the result is if the user has depression or not.
- A Support Vector Machine works like regression does, but since this is classification it is a bit different. It will create a line/plane/dividing boundary to classify the results with the most generally accurate results.

AI Impact and Ethical Implications

Impact

- This kind of AI can be used to streamline medical treatment/advising to historically underserved communities in this area, all from the comfort of their own homes.
- Having depression can lower general quality of life which can in turn affect every aspect of life itself. Even without a medical diagnosis, knowing you may be at risk can let you take the first steps to better your life.

Implications

- While AI can warn of the chance of depression, this does not mean that it is an accurate medical diagnosis. Results should not be used as a diagnosis.
- With the current stigma of mental health, users will not be very open to the idea of being profiled. This could cause issues.

Sources of Bias

- The only social media platform used was Reddit which may not be representative of all social media platforms
- The average age of Reddit users may not be fully representative of all demographics (age, gender, ideology, socioeconomic status, etc.)
- Only 7,650 posts were gathered in the dataset which may not train an AI with the level of precision required for applicable use in the field

Dataset

- “Depression: Reddit Dataset (Cleaned)”
- “This data is collected through web scraping subreddits and is cleaned using multiple NLP (Natural Language processing) techniques.” (The dataset description)

Sources:

- Machine Learning Natural Language Processing
- Undiagnosed Depression: A Community Diagnosis
- Major Depression
- Introduction to Bayesian Networks
- Bayesian Networks in Python Tutorial
- Natural Language Processing (NLP) Zero to Hero
- Bayesian Network | Machine Learning-Python
- Support Vector Machine Explained
- Understanding the Basics of SVM With Example and Python Implementation
- A guide to Text Classification(NLP) using SVM and Naive Bayes with Python
- Text Classification: What it is And Why it Matters
- Natural Language Processing (NLP): 7 Key Techniques