



UnMask

Jacob Long, Matt Bailey, Manan Shah and David Cuellar

Project Objective

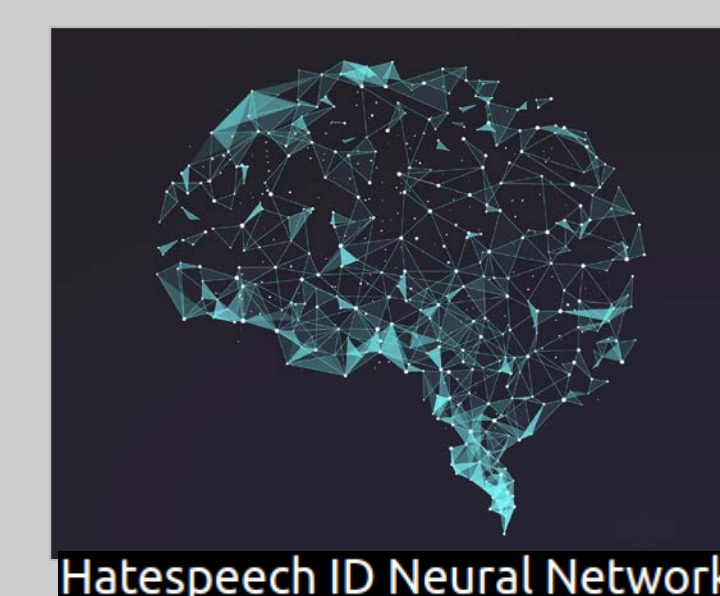
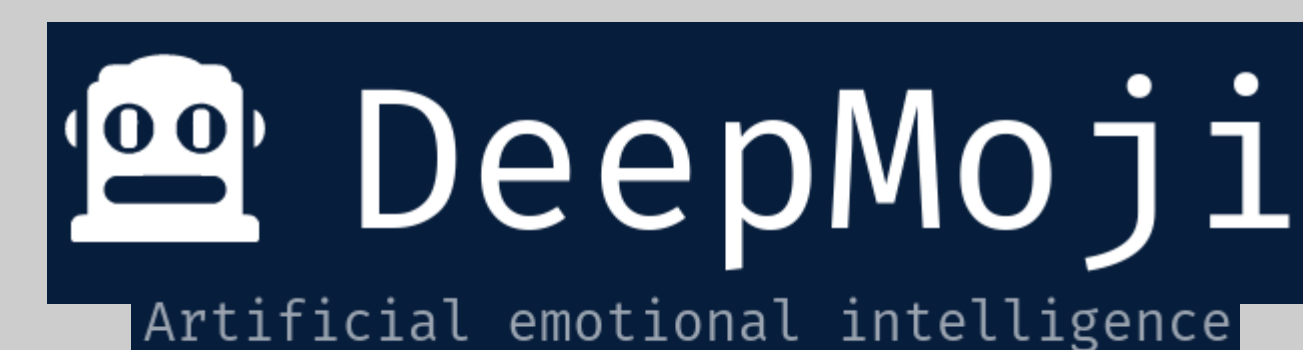
- Predictive Behavioral Modelling software that will integrate deep learning neural networks (MIT's DeepMoji and IIT-Hyderabad's Deep Learning for Hate speech identification network)
- Analyze a given user's relationship networks from twitter to predict likelihood that user may engage in discriminatory behaviors and risk adverse outcomes to their employer
- That likelihood will be represented by a "behavioral profile", which includes a risk coefficient and most common sentiment found in a sample of tweets in the form of an emoji, returned to the customer, who provides input
- The end result is a web-based application intended for use by HR departments and hiring managers to screen the social media presences of prospective employees for indicators of hateful / risky sentiment.
- By providing an automated, quantitative means for analyzing the social media of prospective hires, the customer would insulate themselves from legal risks of discovering protected information via social media screening, as would occur if such screenings were performed manually.

User Interface

The diagram illustrates the user interface flow. At the top, a navigation bar includes 'Search', 'About', 'Contact', and 'EECS 582'. Below this is a search form with the label 'Twitter username:' and a 'Submit' button. Two arrows point from the search form to two example result screens. The left screen shows 'Here are your results!' with a 'Risk coefficient: 0% (No risk)' and 'Most common sentiment:' followed by a 😄 emoji. The right screen shows 'Here are your results!' with a 'Risk coefficient: 74% (High risk)' and 'Most common sentiment:' followed by a 😡 emoji.

Design

- MIT's Deepmoji neural network and IIT Hyderabad's Hate Speech ID neural network are both given the tweets as input.
- The output of both nets are then combined and mathematically analyzed in a python wrapper function developed by the UnMask team.
- The output is returned to the front end - a web app ran on the Django framework - in the form of a percentage and an emoji.



Ideas for Expansion

The UnMask team met the specifications that were set out for them, but have ideas for future expansion of the suite.

- More robust profile returned
 - Sample tweets
 - Network analysis
 - Confidence coefficient
- Scrape tweets from account based off:
 - User location
 - Specified time period of tweets
 - Specified type of hate speech
- Extended form submission options
- Search History
- Efficiency improvements