

# Mental Health Risk Text Classifier

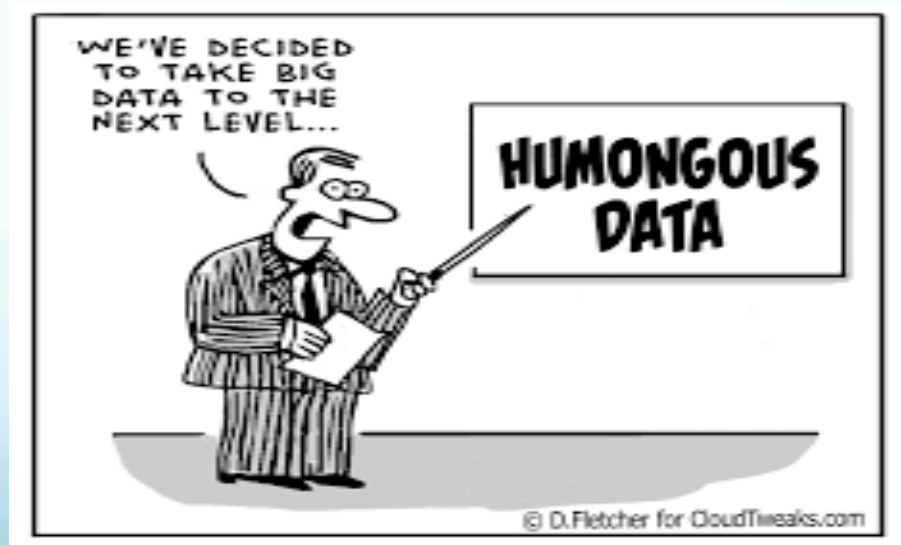
**Can Text Data  
Identify Risk?**

Leah Pilossoph, RN  
BSN, MSs, CCRN, OCN



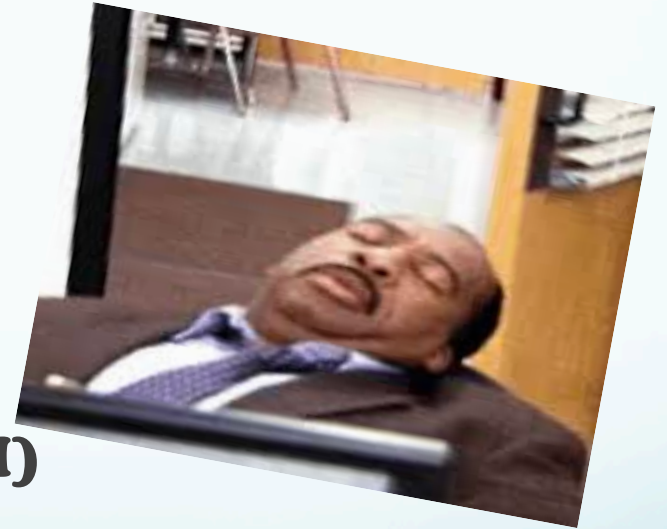
# The Problem

- **Diagnosing is limited**
- **Based on sporadic conversations between clinicians and patients**
- **Some questionnaires; many feel they are outdated**
- **Need More Data!!!**



# Methods

- **PRAW: Python Reddit API Wrapper for web-scraping**
- **Hypothesis testing – One Tailed Two Sample T-Tests**
- **Term Frequency-Inverse Document Frequency Vectorization**
- **Count Vectorization**
- **Multinomial Naïve Bayes**
- **Extreme Gradient Boosting**
- **Recurrent neural networks**
  - **Long short-term memory (LSTM)**
  - **Gated recurrent units (GRU)**



# In an Absolute State: Elevated Use of Absolutist Words Is a Marker Specific to Anxiety, Depression, and Suicidal Ideation



Mohammed Al-Mosaiwi and Tom Johnstone

Department of Psychology, School of Psychology and Clinical Languages, University of Reading

Clinical Psychological Science  
2018, Vol. 6(4) 529–542  
© The Author(s) 2018



Reprints and permissions:  
[sagepub.com/journalsPermissions.nav](http://sagepub.com/journalsPermissions.nav)  
DOI: 10.1177/2167702617747074  
[www.psychologicalscience.org/CPS](http://www.psychologicalscience.org/CPS)



- **Absolutely**
- **All**
- **Always**
- **Complete**
- **Completely**
- **Constant**
- **Constantly**
- **Definitely**
- **Entire**
- **Ever**
- **Every**
- **Everyone**
- **Everything**
- **Full**
- **Must**
- **Never**
- **Nothing**
- **Totally**
- **Whole**

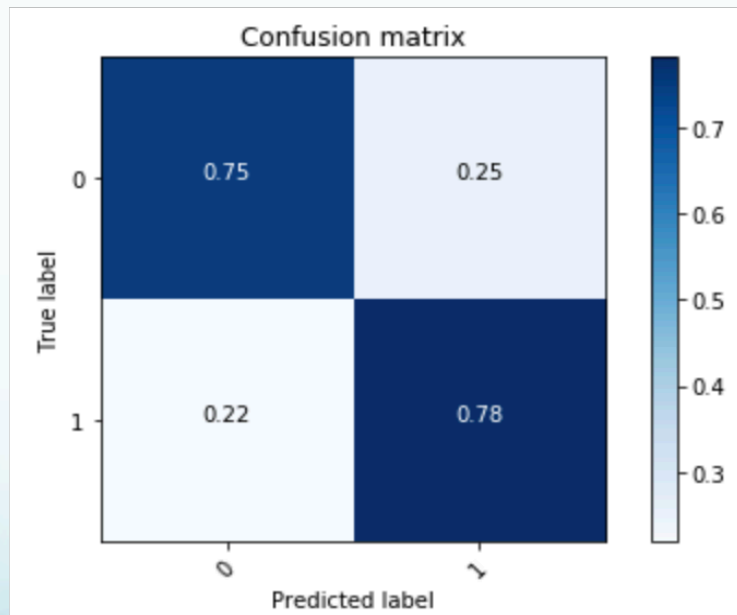
# Hypothesis Test:

## Confirm Study's Results

- **Absolutist Words are used significantly more frequently in mental-health forums vs. hobby forums**
  - **P-value =  $4.98367479 \times 10^{-18}$**
- **Absolutist Words are used significantly more frequently in mental-health forums vs. chronic disease forums**
  - **P-value =  $1.75108500 \times 10^{-26}$**
- **Absolutist Words are used significantly more frequently in suicide-related forums vs. other mental-health forums**
  - **P-value = 0.00491004**

# Text Classification with Reddit Posts

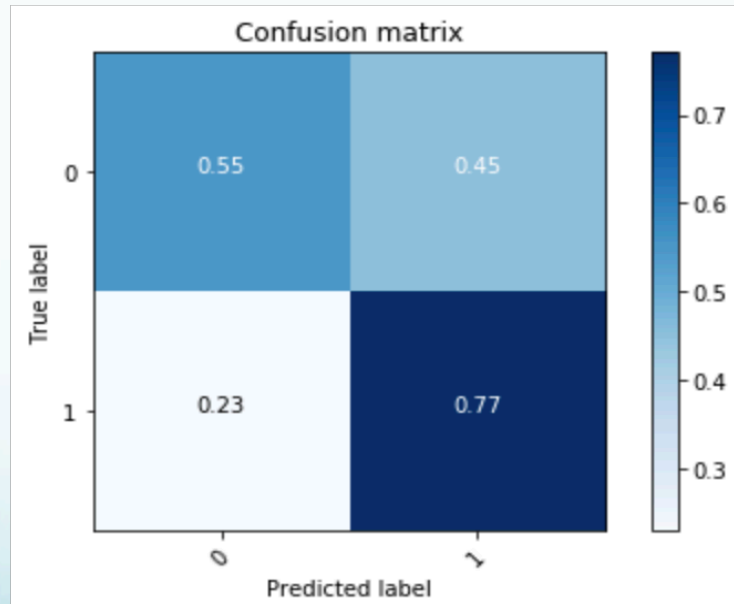
**Mental Health Risk Forums  
vs. Hobby and Leisure  
Forums:**



- Accuracy: 77%
- Precision: 76%
- Recall: 78%
- F1 Score: 77%

# Text Classification with Reddit Posts

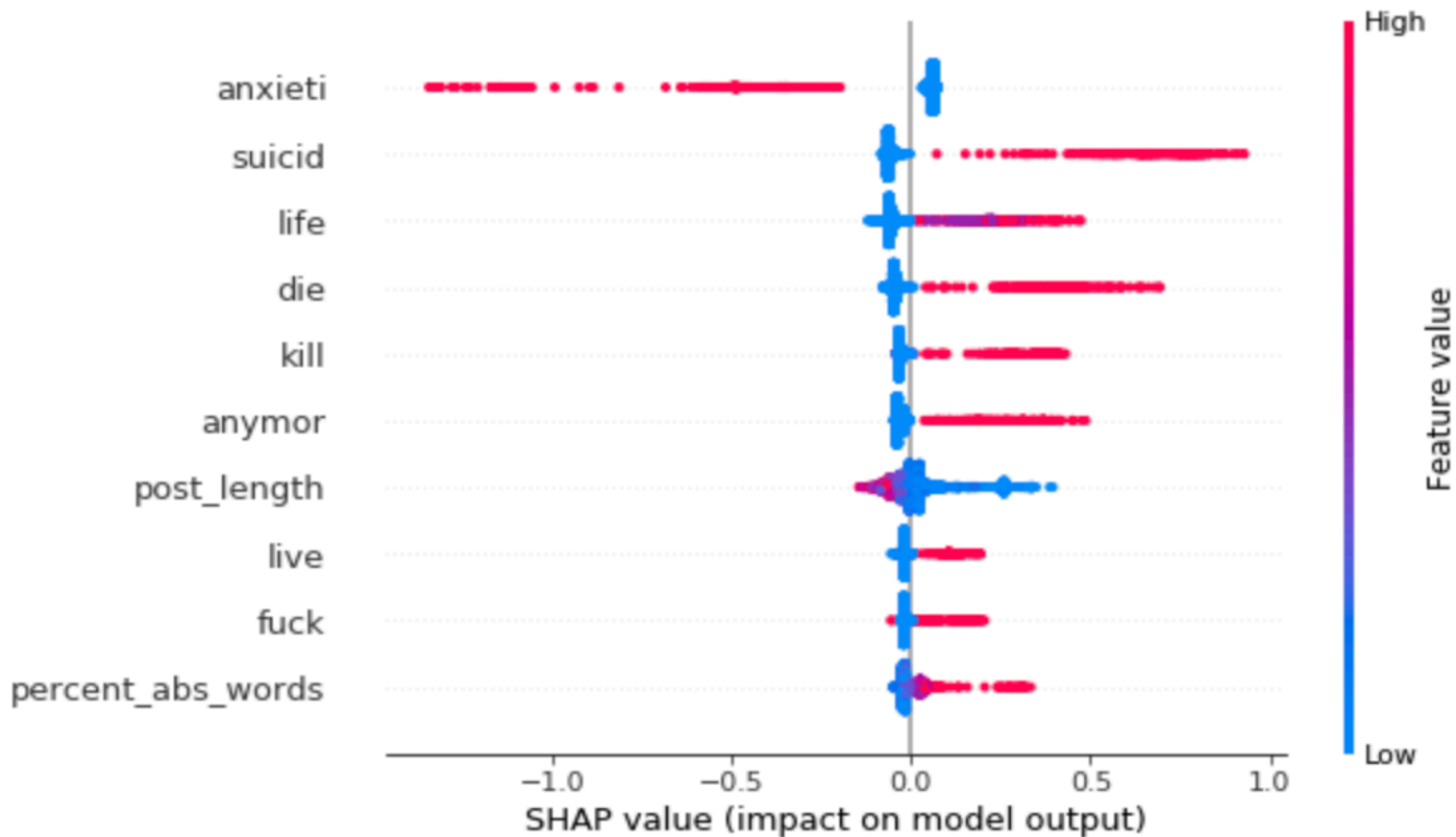
**Suicide Risk Forums vs.  
Other Mental-Health  
Forums:**



- Accuracy: 66%
- Precision: 64%
- Recall: 77%
- F1 Score: 70%



# Feature Importance





# Conclusions

## Features:

- **Several key words are predictors:**
  - “anymore”, “life”, “live”, “die”, etc...
- **Percentage of Absolutist Words in post is a predictor:**
  - Higher percentage → likely to classify as suicidal
- **Post Length is a predictor:**
  - Shorter post → likely to classify as suicidal

## Classifier:

- **Can improve due to imperfect data source**

# Use Cases

- **Parents/Family members:**
  - **Monitor loved ones on social media**
- **Impactful targeted ads on social media:**
  - **Identify high risk, provide access to help**
- **Tool to aid in Clinical Diagnosis**
- **Clinical Dashboard for monitoring:**
  - **Therapy Impact**
  - **New Medication Impact**
  - **See next slide for prototype...**

# Mental Health Tracking Prototype

<https://suicidal-ideation-tracker-tool.herokuapp.com/>

## Mental Health Time Series Tracking

Developed by Leah Pilossoph, RN

Filter by Patient:

→

Select a social media username

Select a date range

Percent of Absolutist Words in Post (Scaled 0-5)



Length of Post (Scaled 0-5)



# Future Work

- **Further Analysis:**
  - **Phrase Analysis**
  - **Latent Dirichlet Allocation (LDA) → Topic Analysis**
- **More Data!**
- **Better Data!**
- **Address Potential Privacy Issues**

# Questions?



A cartoon drawing of a hand holding a sign that says "ROAST ME". The hand is simple, with two fingers visible. The sign is white with a black outline and the words "ROAST ME" are written in large, bold, red, sans-serif capital letters. The sign is tilted slightly to the right.

# Thank you!