


Project 3: Subreddit NLP Classification

Brandon Griffin
10/23/2020



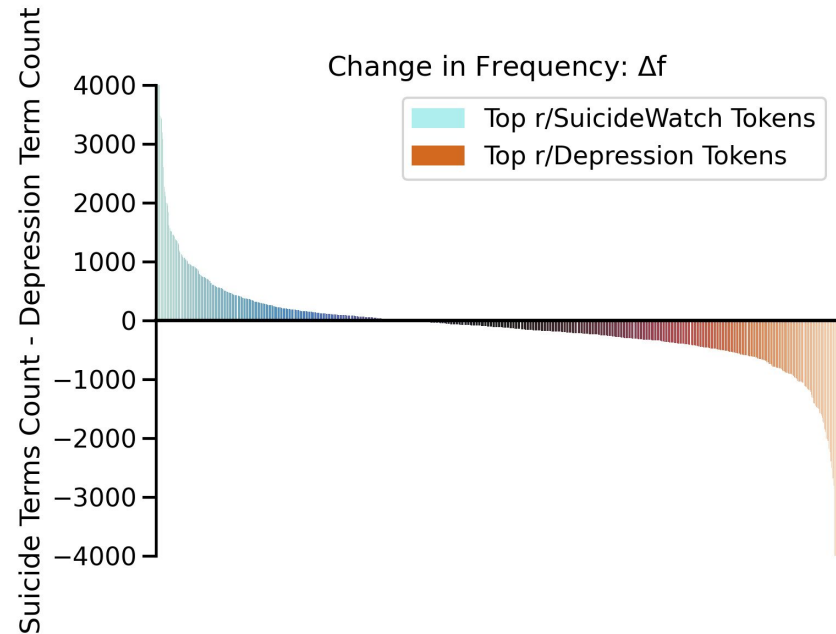
Introduction

Problem Statement: Can Natural Language Processing be used to tell suicidal ideation apart from depression in a collection of electronic records?

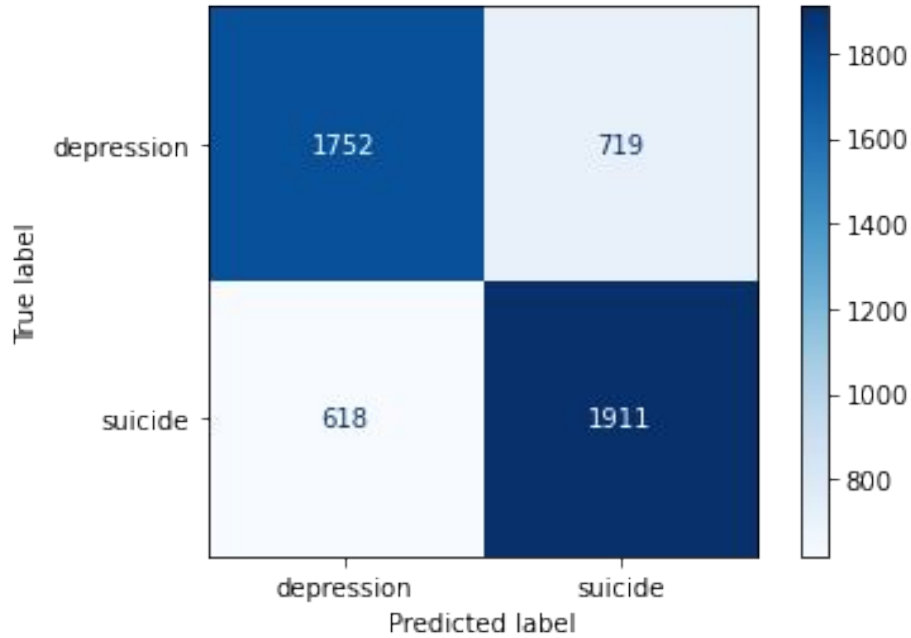
Target Audience: A psychiatric treatment center is investigating the association of suicidal ideation and depression. There are a large number of incomplete health records available for clinical research on, but first they need to be encoded correctly. I have been hired to quantify the distinction between help them determine if they can automate the process.

Methodology/Procedure

die	8213
kill	7960
suicide	7308
end	4524
want	4429
...	...
feel like	-4300
depressed	-5861
like	-6066
feel	-6736
depression	-11046

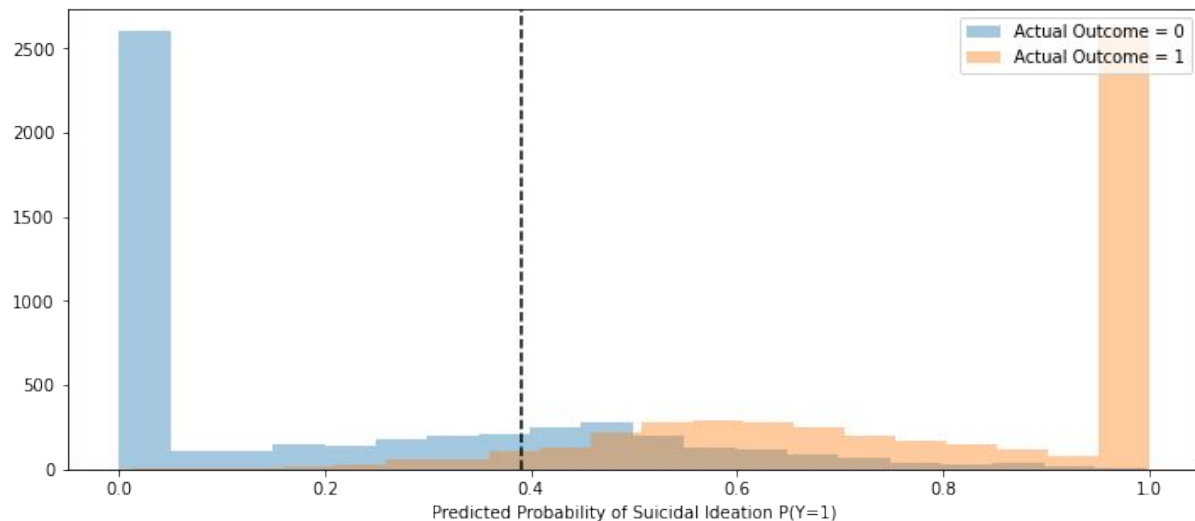


Data Considered



Metric	Score
Sensitivity	76%
Accuracy	73%
Precision	71%

Primary Findings



Metric	Score
Sensitivity	90%
Threshold	$P(Y=1) \leq 0.39$
Precision	64%

Recommendations to client

1. Suicidal ideation can be distinguished from depression with a sensitivity of 90% when a threshold of $P(Y=1) \leq 0.39$ is enforced
2. A hybrid Machine Learning and Rule based approach may be necessary

Sources

Pompili M. Critical appraisal of major depression with suicidal ideation. *Ann Gen Psychiatry*. 2019;18:7. Published 2019 May 31. doi:10.1186/s12991-019-0232-8