



Predicting Mental Health Based on Social Media Posts

STAT 571 Modern Data Mining

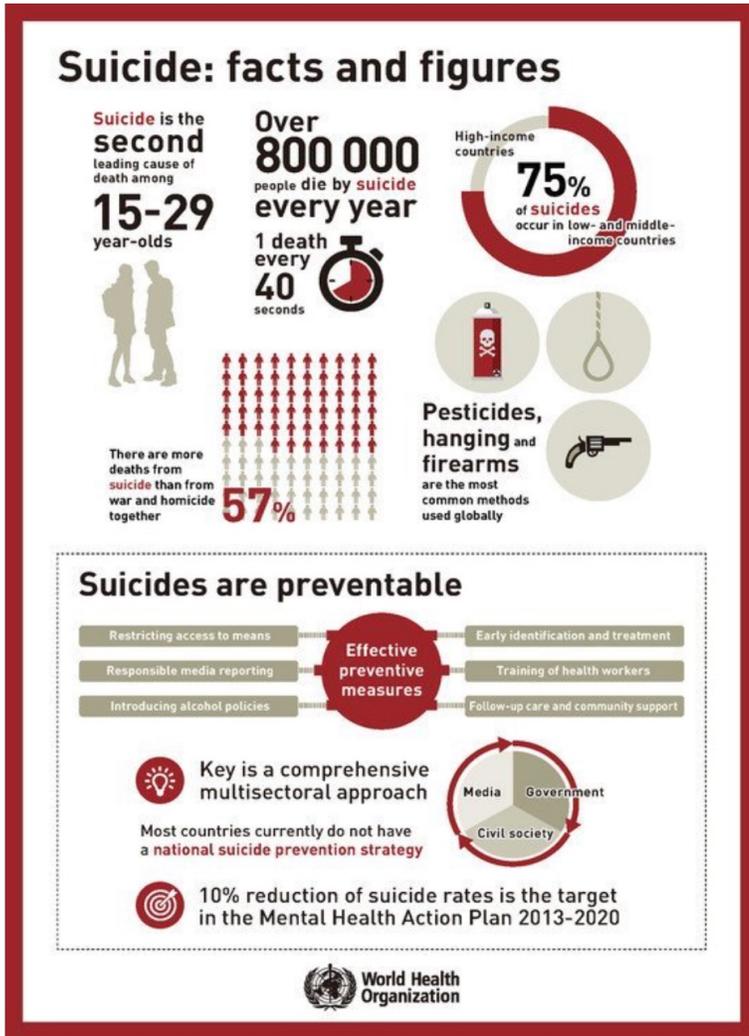
Final project

Group 35

Xueqi Sun & Xu Han

Research Background:

Can we predict depression and suicidal tendencies?



WHO | Facts and figures on suicide.

Depression is one of the **most prevalent** mental health conditions in the US and it is the **leading cause** of disability worldwide.

In 2020, an estimated **14.8 million U.S. adults** aged 18 or older and **4.1 million U.S. adolescents** aged 12 to 17 had at least one major depressive episode with severe impairment in the past year.

ADAA

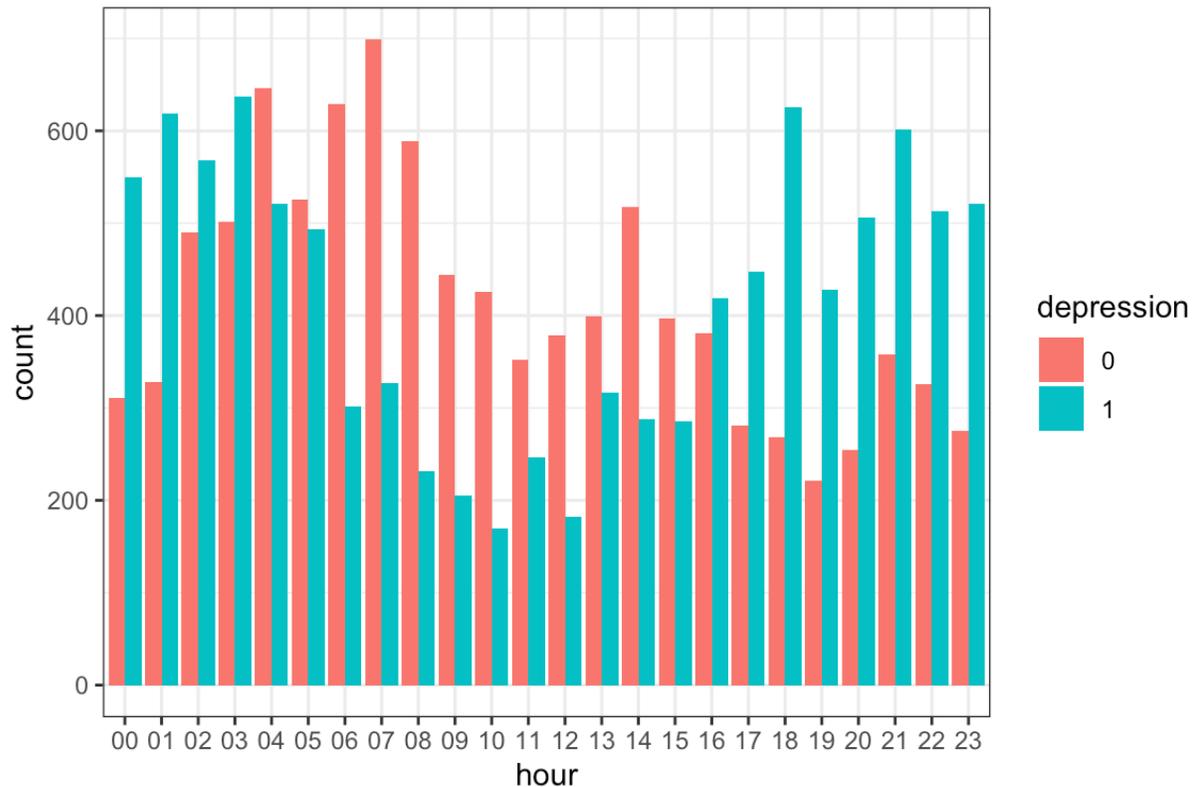
Datasets:

Data1: Depression: Twitter Dataset + Feature Extraction

Data2: Suicide and Depression Detection

Data3: Depression 2: Twitter Dataset

Post created time vs. Depression



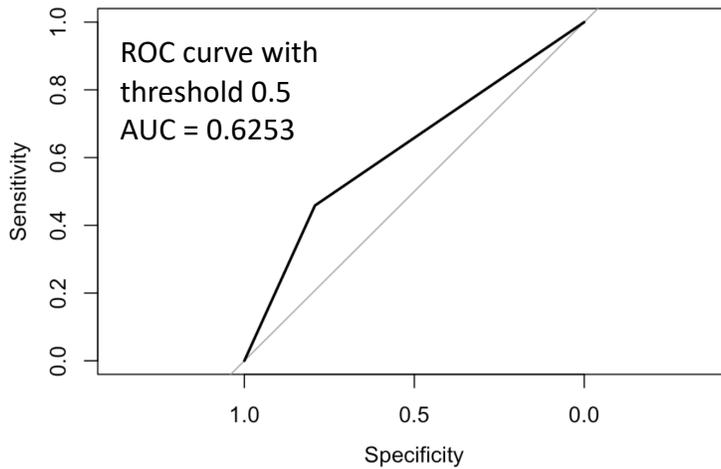
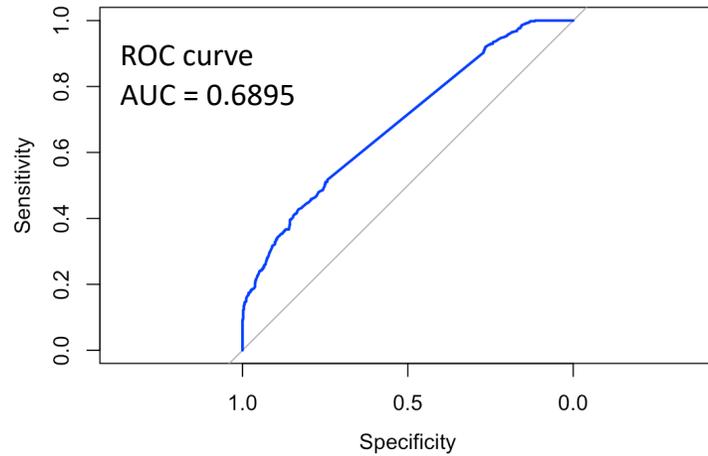
```
## Coefficients:
##           Estimate Std. Error z value Pr(>|z|)
## hour00  0.57013    0.07095   8.036 9.30e-16 ***
## hour01  0.63509    0.06830   9.299 < 2e-16 ***
## hour02  0.14772    0.06166   2.396 0.01658 *
## hour03  0.23817    0.05968   3.991 6.59e-05 ***
## hour04 -0.21505    0.05888  -3.652 0.00026 ***
## hour05 -0.06479    0.06269  -1.034 0.30132
## hour06 -0.73370    0.07001 -10.480 < 2e-16 ***
## hour07 -0.75969    0.06700 -11.339 < 2e-16 ***
## hour08 -0.93169    0.07751 -12.020 < 2e-16 ***
## hour09 -0.77281    0.08444  -9.152 < 2e-16 ***
## hour10 -0.92454    0.09091 -10.170 < 2e-16 ***
## hour11 -0.35830    0.08310  -4.312 1.62e-05 ***
## hour12 -0.73353    0.09018  -8.134 4.16e-16 ***
## hour13 -0.23322    0.07530  -3.097 0.00195 **
## hour14 -0.58701    0.07350  -7.986 1.39e-15 ***
## hour15 -0.33145    0.07764  -4.269 1.96e-05 ***
## hour16  0.09507    0.07079   1.343 0.17927
## hour17  0.46420    0.07613   6.097 1.08e-09 ***
## hour18  0.84676    0.07302  11.597 < 2e-16 ***
## hour19  0.66096    0.08283   7.979 1.47e-15 ***
## hour20  0.68527    0.07680   8.923 < 2e-16 ***
## hour21  0.51806    0.06676   7.760 8.51e-15 ***
## hour22  0.45338    0.07083   6.401 1.54e-10 ***
## hour23  0.63898    0.07454   8.573 < 2e-16 ***
```


Building Models

Model	MCE
Logistic regression	0.37225
Neural Network	0.38175
Random Forest	0.36525
Random Forest with PCA	0.3775

Predicted Depression \ True Depression	0	1
	0	1504
1	541	986

Checking the Final Model with Validation Data



Predicted Depression \ True Depression	0	1
	0	1
0	774	554
1	203	469

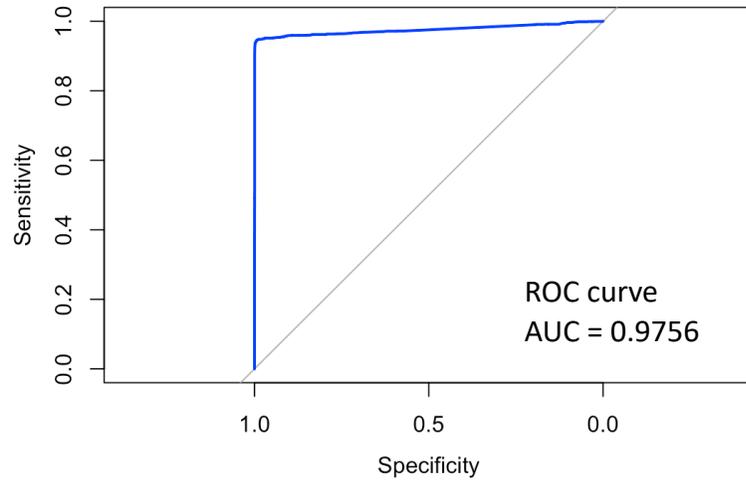
- MCE = 0.3785
- phi coefficient = 0.2652866

Predicting Depression for the Suicidal Data

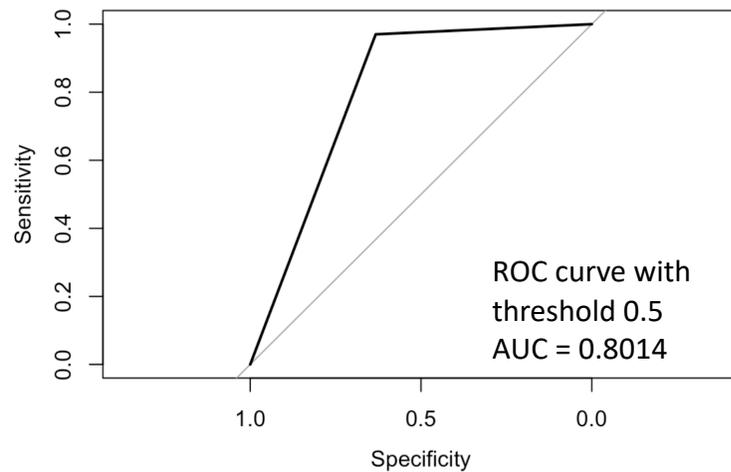
Predicted Depression \ True Suicide	0	1
	0	1
0	3617	795
1	7927	10869

- The proportion that predicted depression does not match suicide is 0.3758187.
- phi coefficient = 0.3123986

Testing the Model on a Different Set of Depression Data



		True Depression	
		0	1
Predicted Depression	0	5061	69
	1	2939	2245



- MCE = 0.2916424
- The phi coefficient of the contingency table is 0.5029368, indicating a well agreement.



Summary

