machine learning in

# identifying suicide attempt predictors

using demographic data

Bengielyn Mae Danao

# existing study

**HHS Public Access**

Author manuscript

*J Affect Disord*. Author manuscript; available in PMC 2017 March 15.

Published in final edited form as:

*J Affect Disord*. 2016 March 15; 193: 109–116. doi:10.1016/j.jad.2015.12.066.

## Identifying a clinical signature of suicidality among patients with mood disorders: a pilot study using a machine learning approach

**Ives Cavalcante Passos, MD**[1,2,*], **Benson Mwangi, PhD**[1,*], **Bo Cao, PhD**[1], **Jane E Hamilton, PhD**[1], **Mon-Ju Wu, PhD**[1], **Xiang Yang Zhang, MD, PhD**[1,3], **Giovana B. Zunta-Soares, MD**[1], **Joao Quevedo, MD, PhD**[1], **Marcia Kauer-Sant'Anna, MD, PhD**[2], **Flávio Kapczinski, MD, PhD**[2], and **Jair C. Soares, MD, PhD**[1]

**72%**
accuracy

# suicide statistics

**800k**
**deaths** per year
around the world

**one** death
every 40 secs

**2nd** leading
cause of death
for ages **15-24**

reduce suicide rates by predicting which individuals are at high risk by screening risk factors

# data source

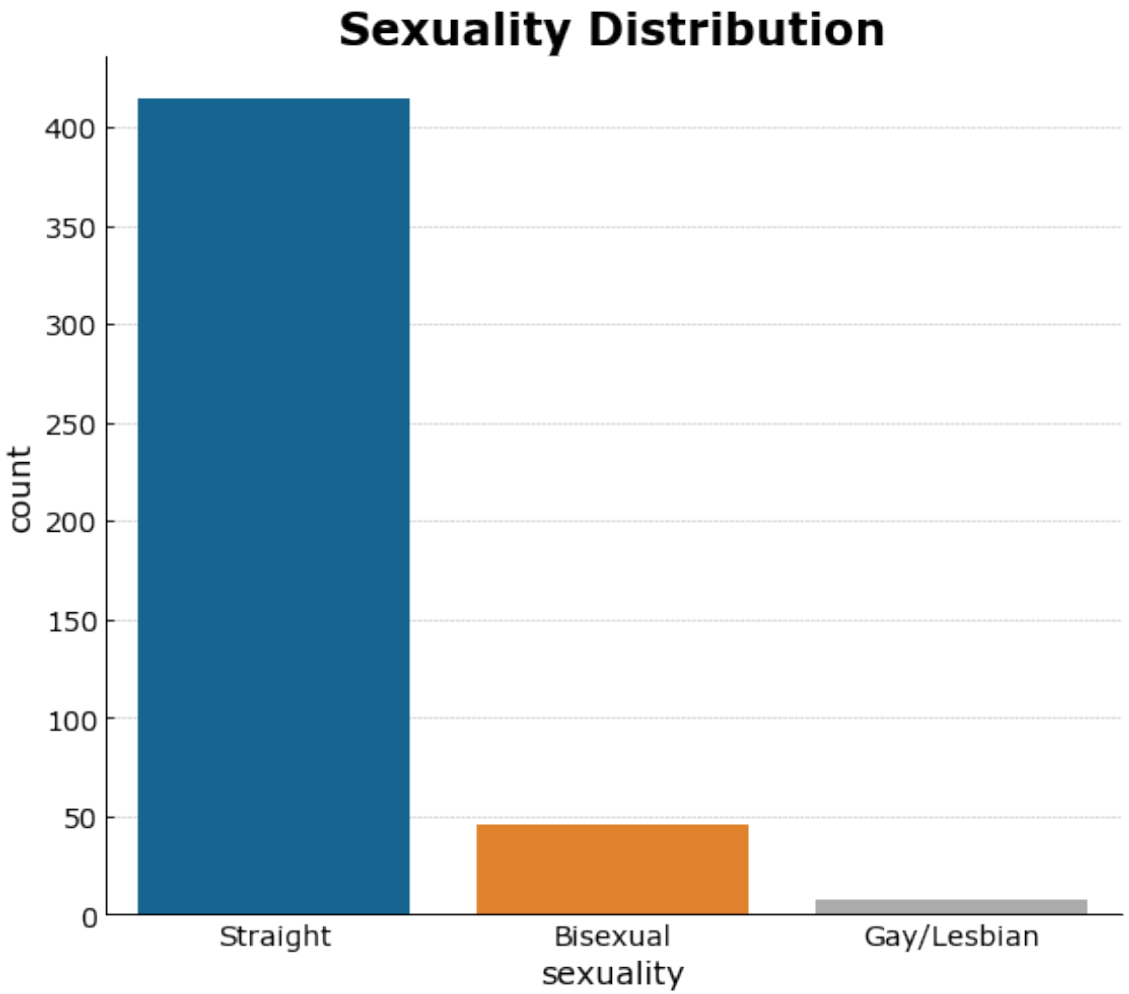
demographic data from a survey collected from the subreddit r/ForeverAlone

1. Gender
2. Sexual orientation
3. Age
4. Income level
5. Race / ethnicity
6. Bodyweight description
7. Virginity
8. Legality of prostitution in the area
9. No. of friends in real life
10. Social fear / anxiety
11. Are you depressed?
12. What kind of help do you want from others?
13. Have you attempted suicide?
14. Employment status
15. Job title
16. Education level
17. What have you done to improve yourself?

# more on the data

**86**% male

**14**% female



# more on the data

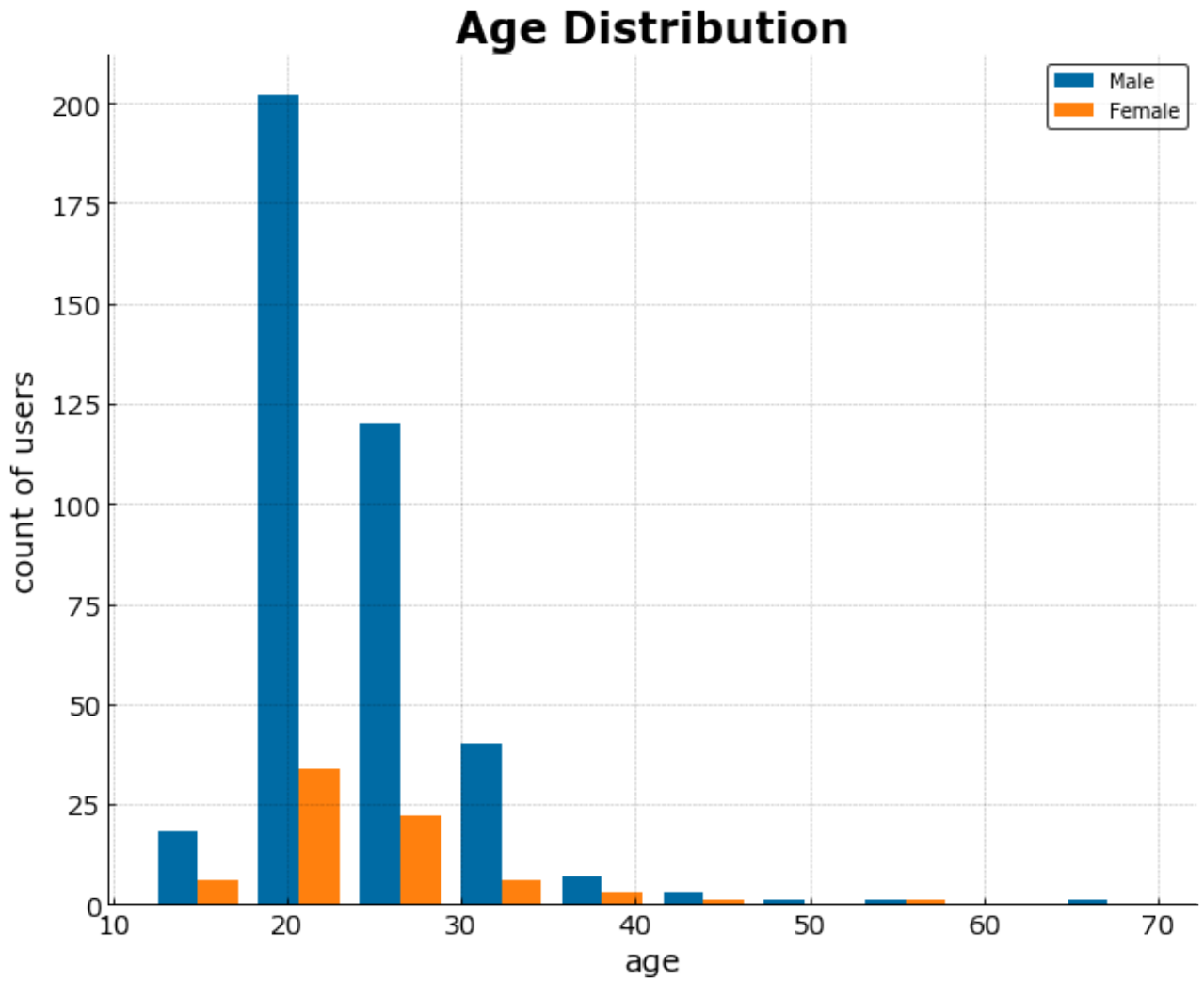
**74%** mixed race

**12%** asian

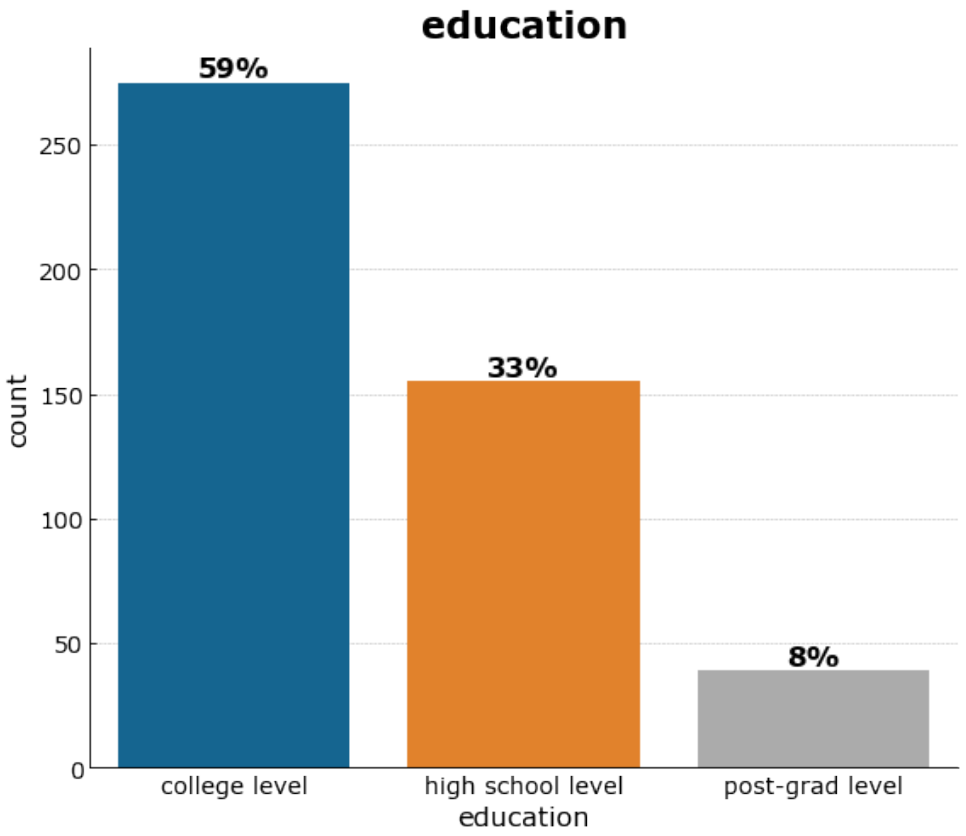
**7%** hispanic

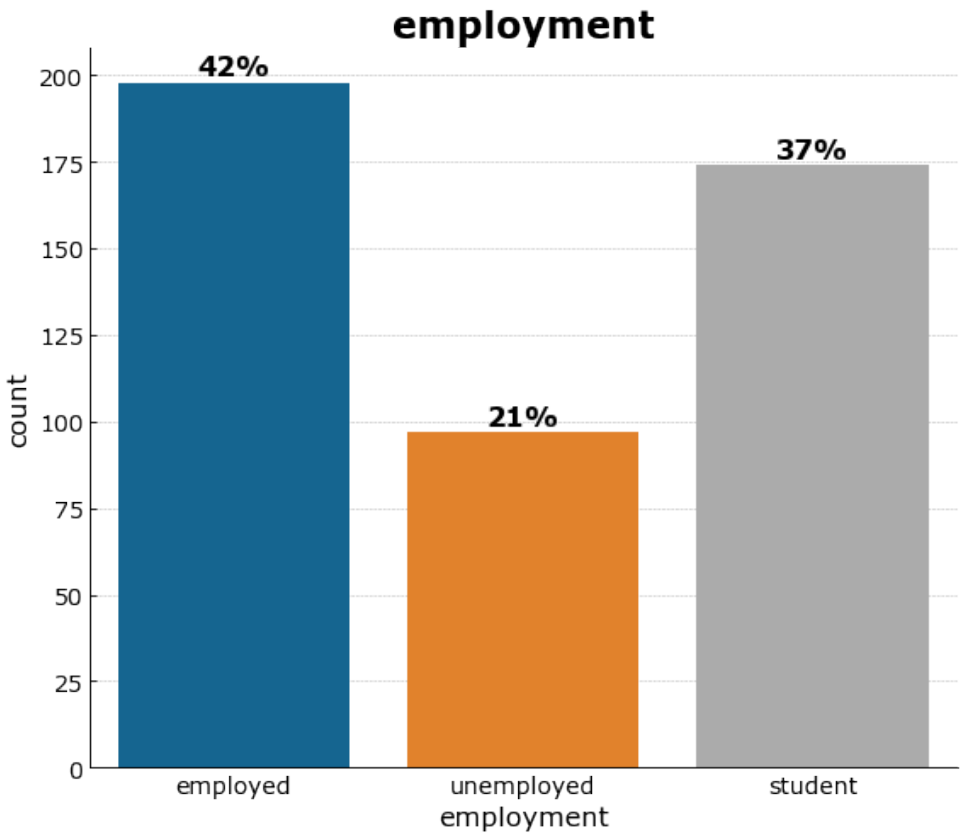
**5%** african

**2%** others



# more on the data



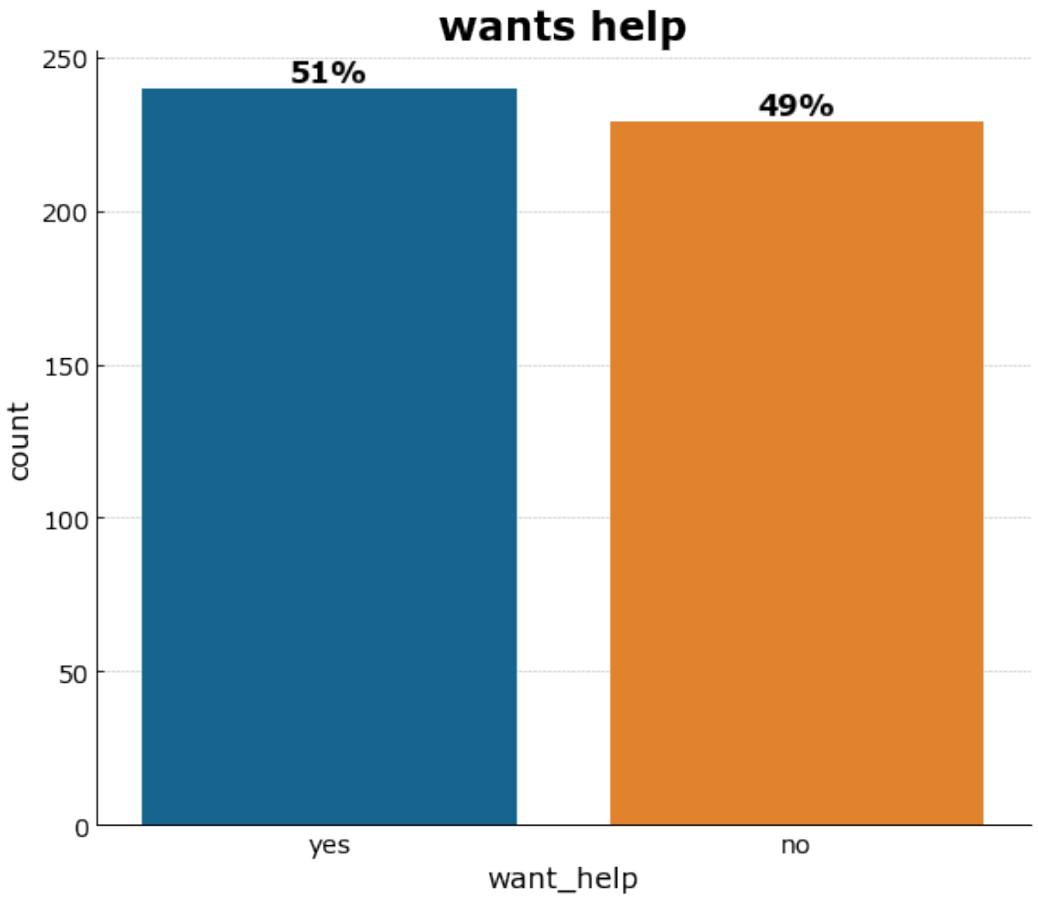

**86%**
low income

**11%**
mid income

**3%**
high income

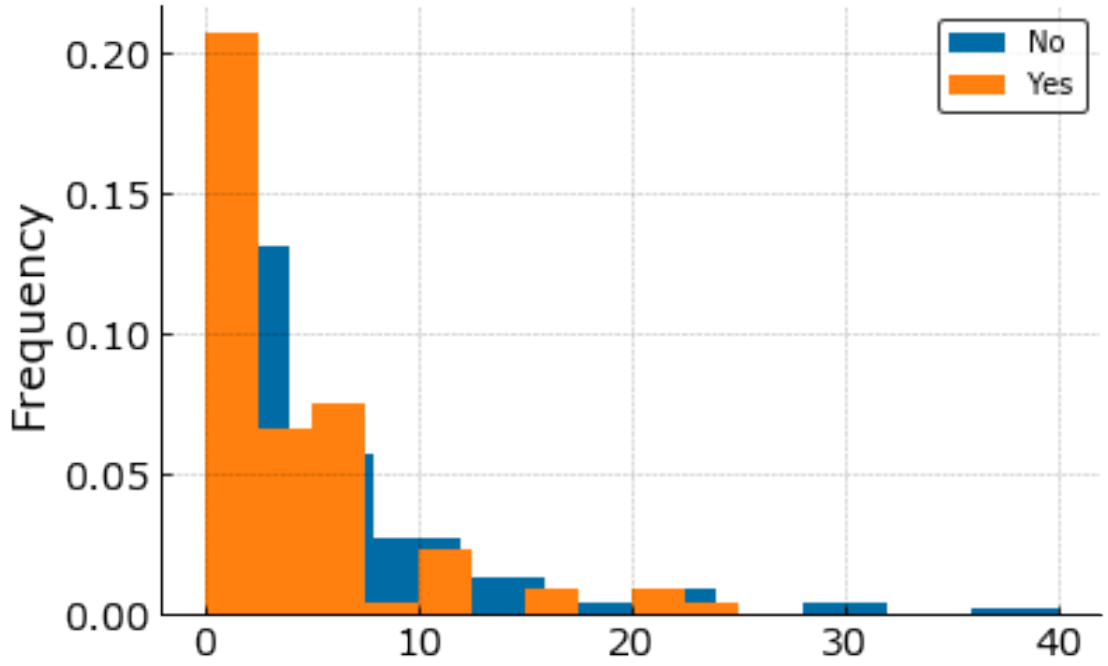
# more on the data

**77**%
willing to improve



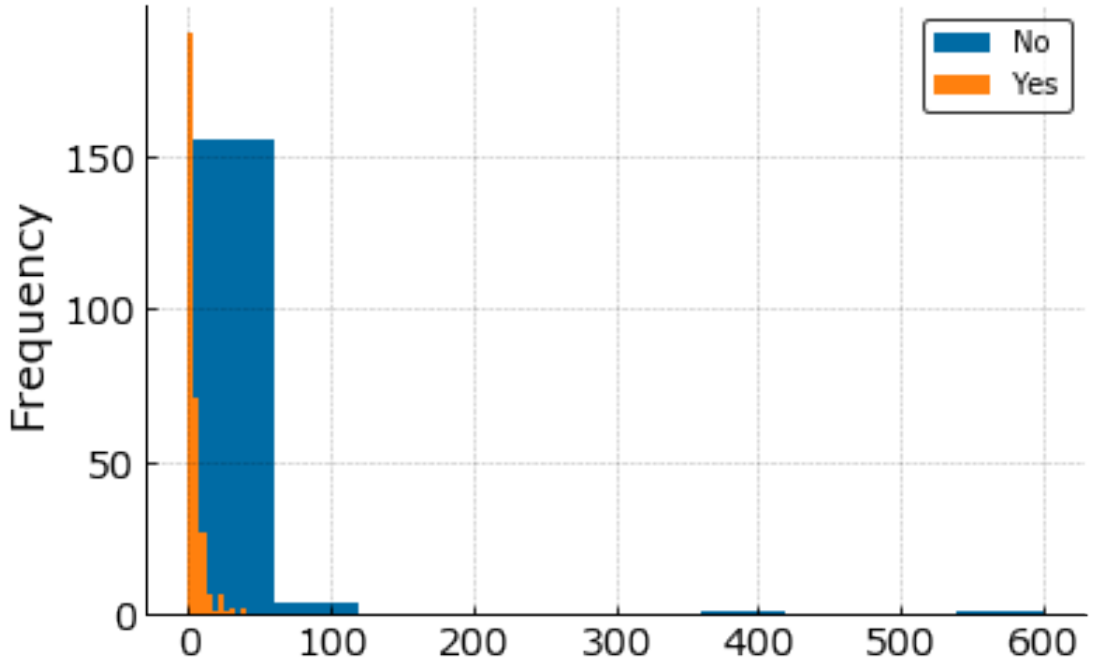
# more on the data

most people who have attempted suicide have less number of friends

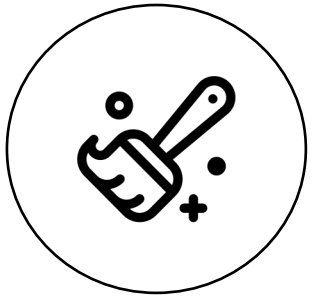


# more on the data

people with social fear have less friends



# methodology



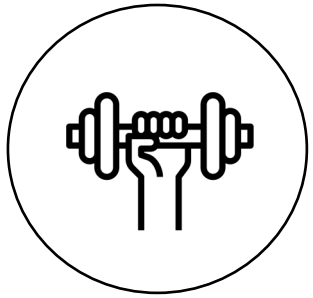
**data cleaning**
preprocessing

label: attempted suicide

15 features

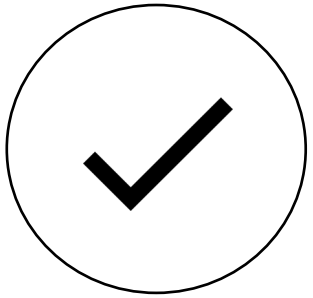
**oversampling**
balance the data

since the data is not balanced at 20-80, SMOTE was used for oversampling



**modelling**
highest accuracy

used grid search to optimize the hyperparameters



**validation**
evaluate model

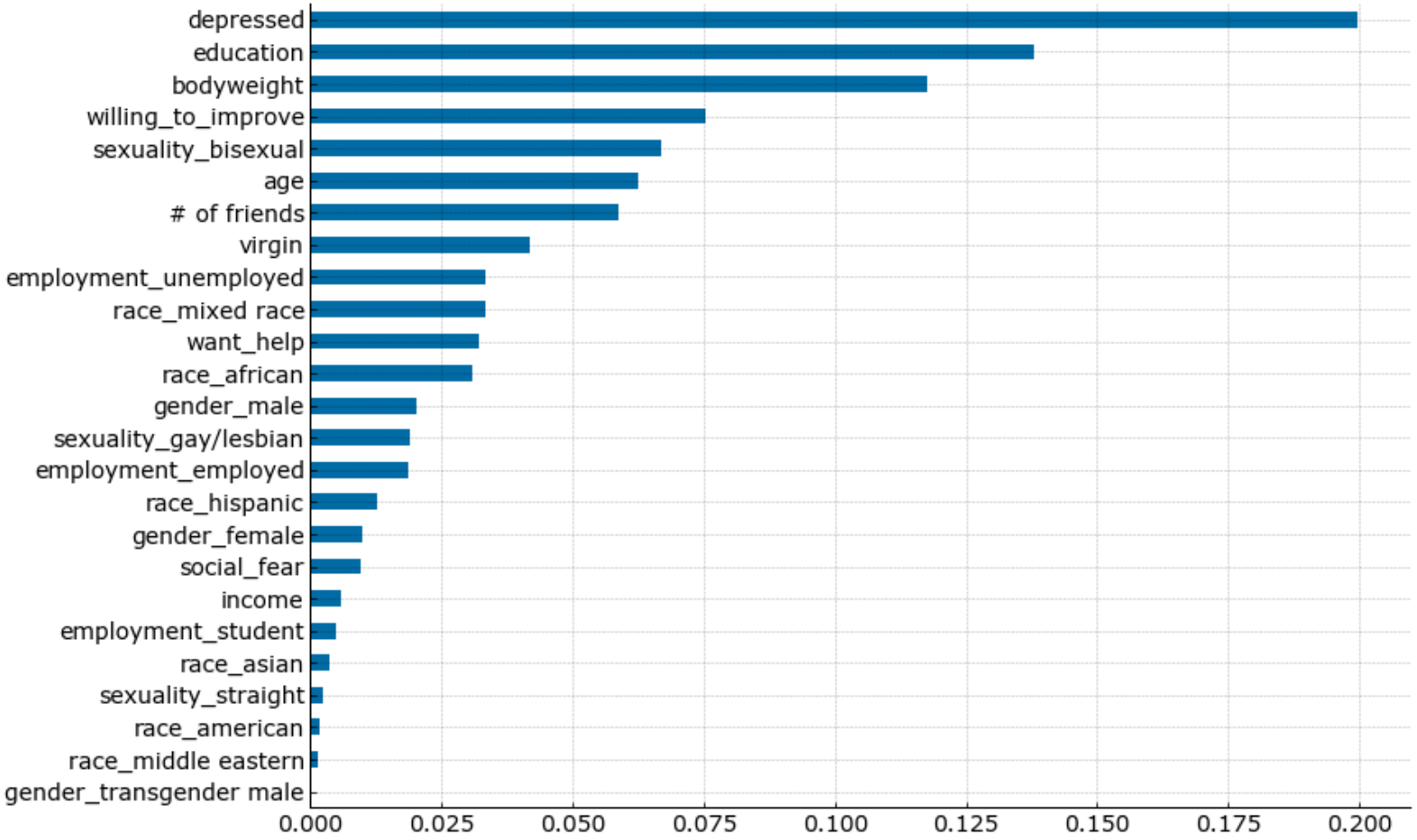
run a cross-validation and classification report

# random forest classifier

**80.5%**

accuracy

max depth: 7
n estimators: 150



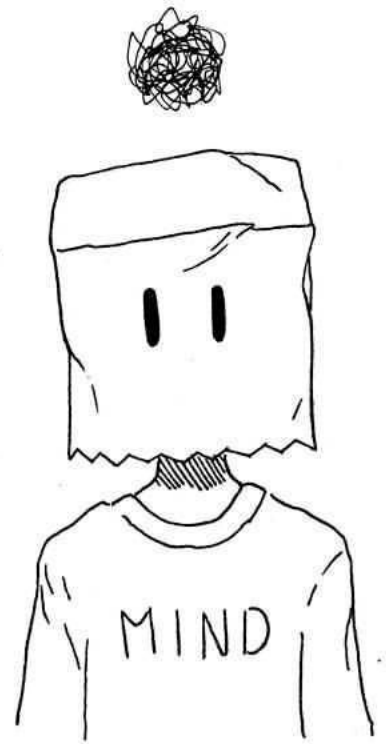
# classification report

| | precision | recall | f1-score | support |
|---|---|---|---|---|
| 0 | 0.81 | 0.93 | 0.87 | 94 |
| 1 | 0.36 | 0.17 | 0.23 | 24 |
| micro avg | 0.77 | 0.77 | 0.77 | 118 |
| macro avg | 0.59 | 0.55 | 0.55 | 118 |
| weighted avg | 0.72 | 0.77 | 0.74 | 118 |

```
=== Mean AUC Score ===
Mean AUC Score - Random Forest:  0.9365041617122474
```

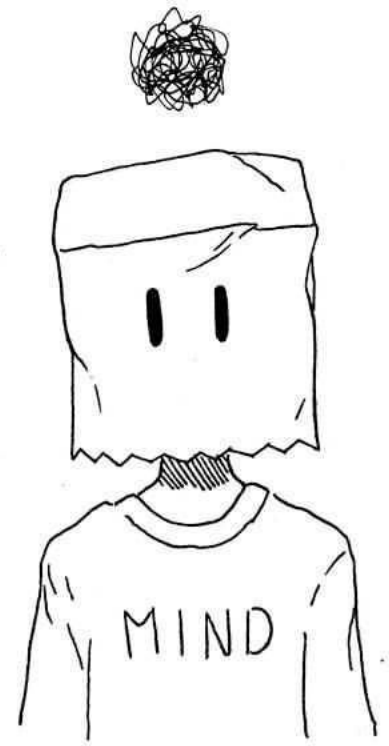
# conclusion

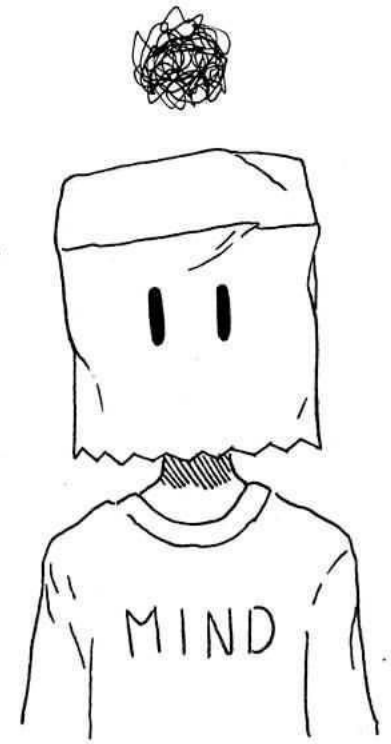
in the prediction model, **depression** is the highest predictor of suicide attempts followed by **education**. This is consistent with previous studies saying majority of suicide tendencies are found to experience anxiety disorders. Several studies also suggest that academic work and stresses are associated with suicide attempts as well.



# conclusion

this study is subject to **methodological limitations** – data from the reddit survey included information about suicide attempts, demographics, and psychological status that was examined using **simple questions and scales** only. This might have affected the performance of the prediction models.



# conclusion

the study showed that a machine learning model on mental health data (if made available) is **capable of identifying predictors** of suicide attempts and other critical suicide risk such as self-harm