

# Adolescent Mind Kenya

A Path to Well-being

Project by:  
Psytech



Understanding and supporting mental health in adolescents.

# Project Team:



Franklin Kuiria  
**Scrum master**

Rosemary Wanjiru  
**Machine learning  
Engineer**

Dorcas Kabutie  
**Data Engineer**

Samuel Njogu  
**Data Analyst**

# Overview

According to The World Health Organisation(WHO, 2023), 1 in 7 teenagers aged 10-19 experience some kind of mental issues.

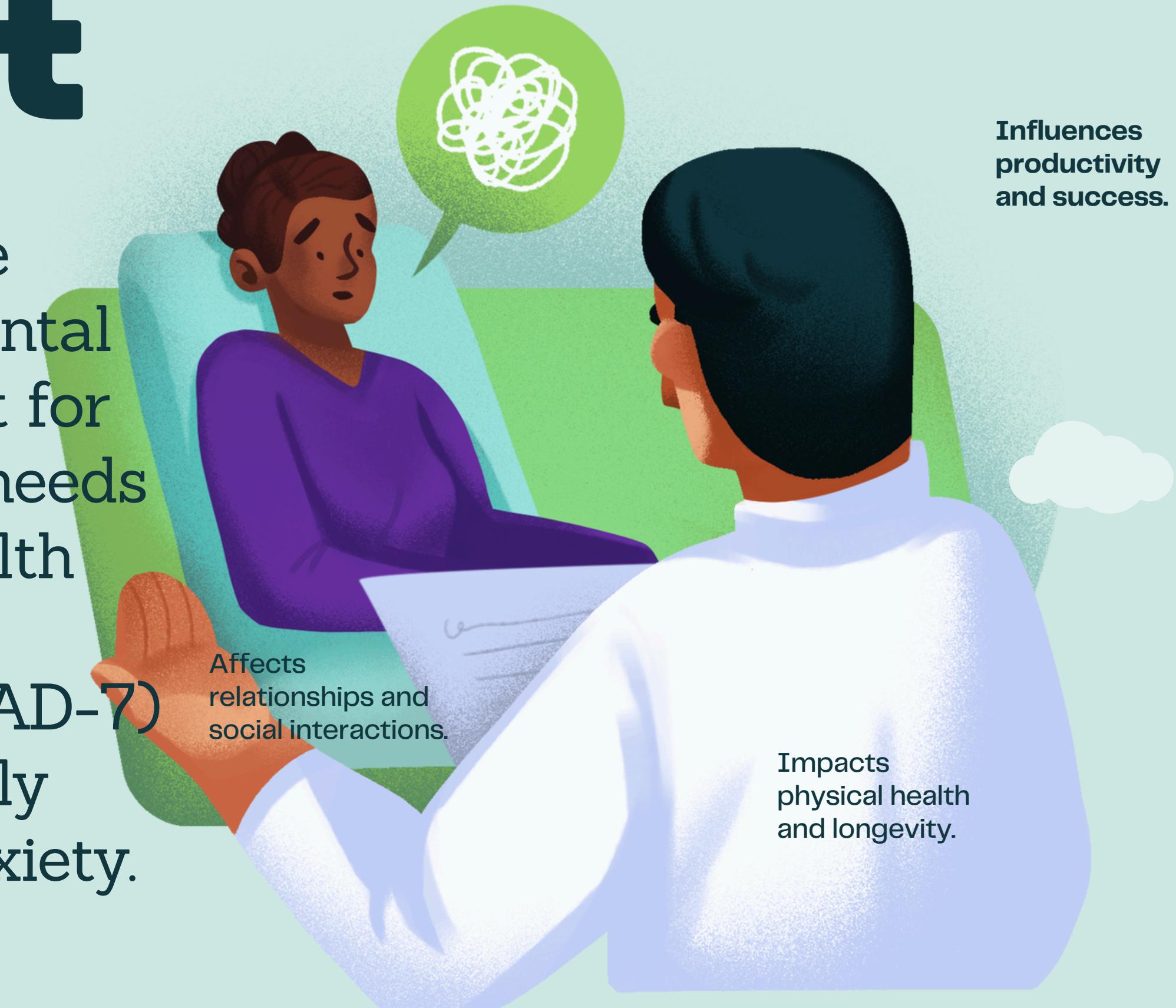
AdolescentMind' s goal is to provide additional support to teenagers by analyzing their demographics, psychological factors, family background and school related features.



It affects how we think, feel, and behave in daily life.

# Problem Statement

Adolescents frequently face challenges that affect their mental and general wellbeing. Support for addressing their mental health needs remains minimal. Patient Health Questionnaire(PHQ-9) and Generalized Anxiety Disorder(GAD-7) are scales for identifying early symptoms of depression and anxiety.



Contributes to overall life satisfaction and happiness.

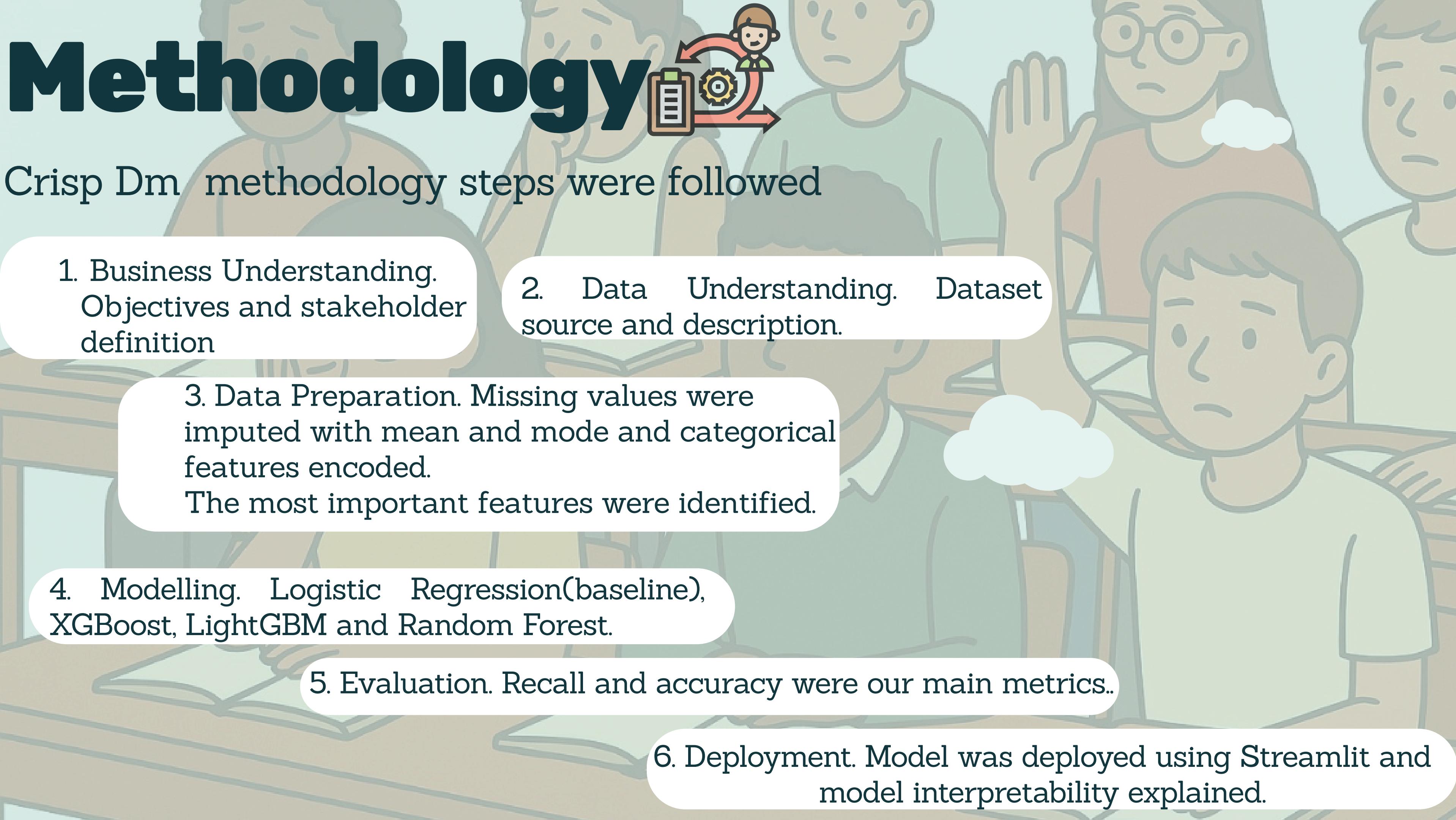
Influences productivity and success.

# Objectives

**Main :** To build a machine learning model capable of identifying depression and anxiety levels among Kenyan adolescents



- To find out the major demographic factors that contribute to depression and anxiety.
- To identify the major factors contributing to anxiety and/or depression.
- To identify how different counties contribute to students level of anxiety and depression.
- To interpret model outputs and find out the most important factors that contribute to depression and anxiety prediction.



# Methodology

Crisp Dm methodology steps were followed

1. Business Understanding.  
Objectives and stakeholder definition
2. Data Understanding. Dataset source and description.
3. Data Preparation. Missing values were imputed with mean and mode and categorical features encoded.  
The most important features were identified.
4. Modelling. Logistic Regression(baseline), XGBoost, LightGBM and Random Forest.
5. Evaluation. Recall and accuracy were our main metrics..
6. Deployment. Model was deployed using Streamlit and model interpretability explained.

# Data Understanding

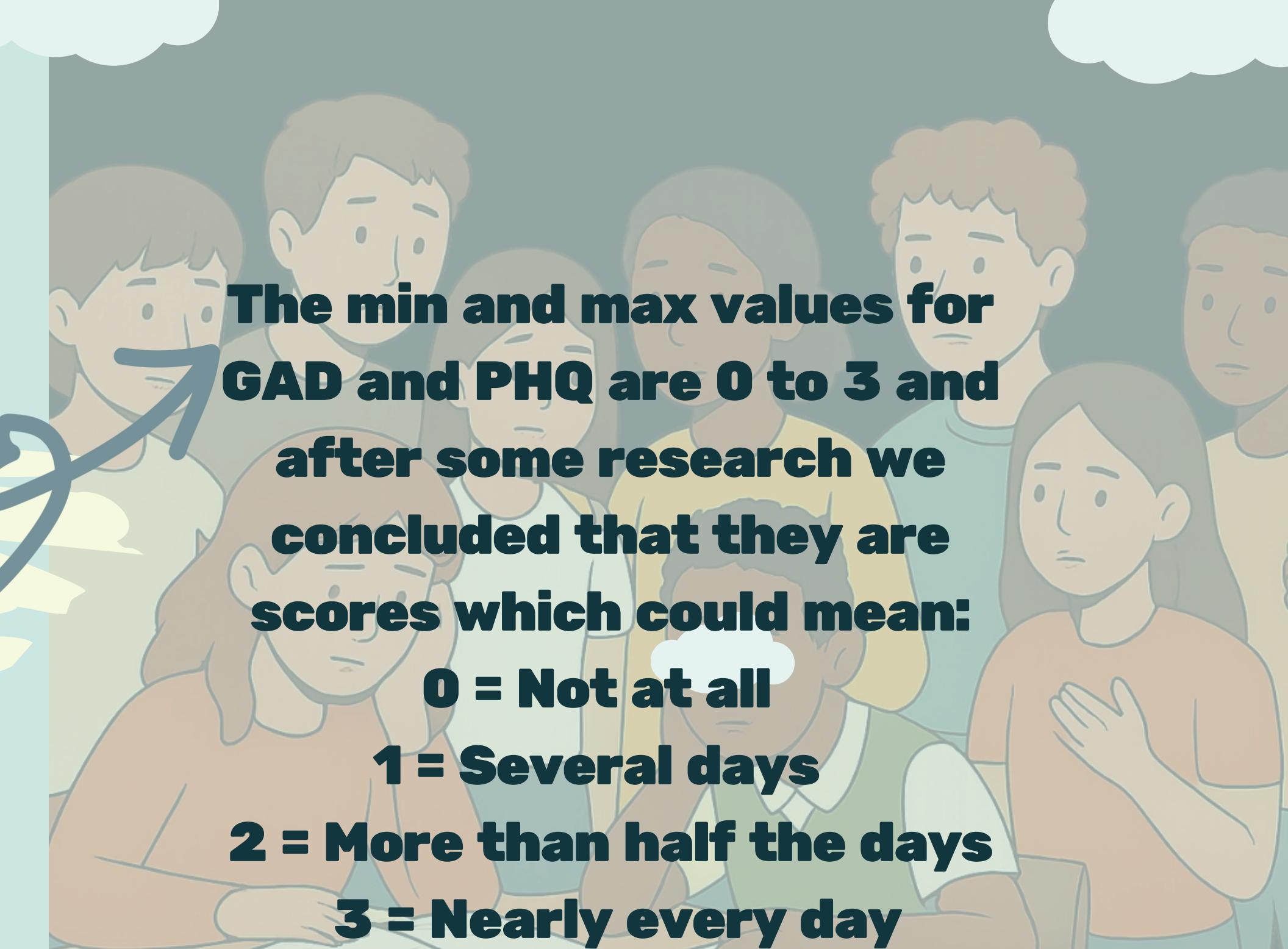
## Source:

[https://osf.io/preprints/osf/yvdsc\\_v2](https://osf.io/preprints/osf/yvdsc_v2)

**Shape:** 17,089 records,  
191 features

## Target:

GAD(Generalized Anxiety Disorder)  
PHQ(Patient Health Questionnaire)

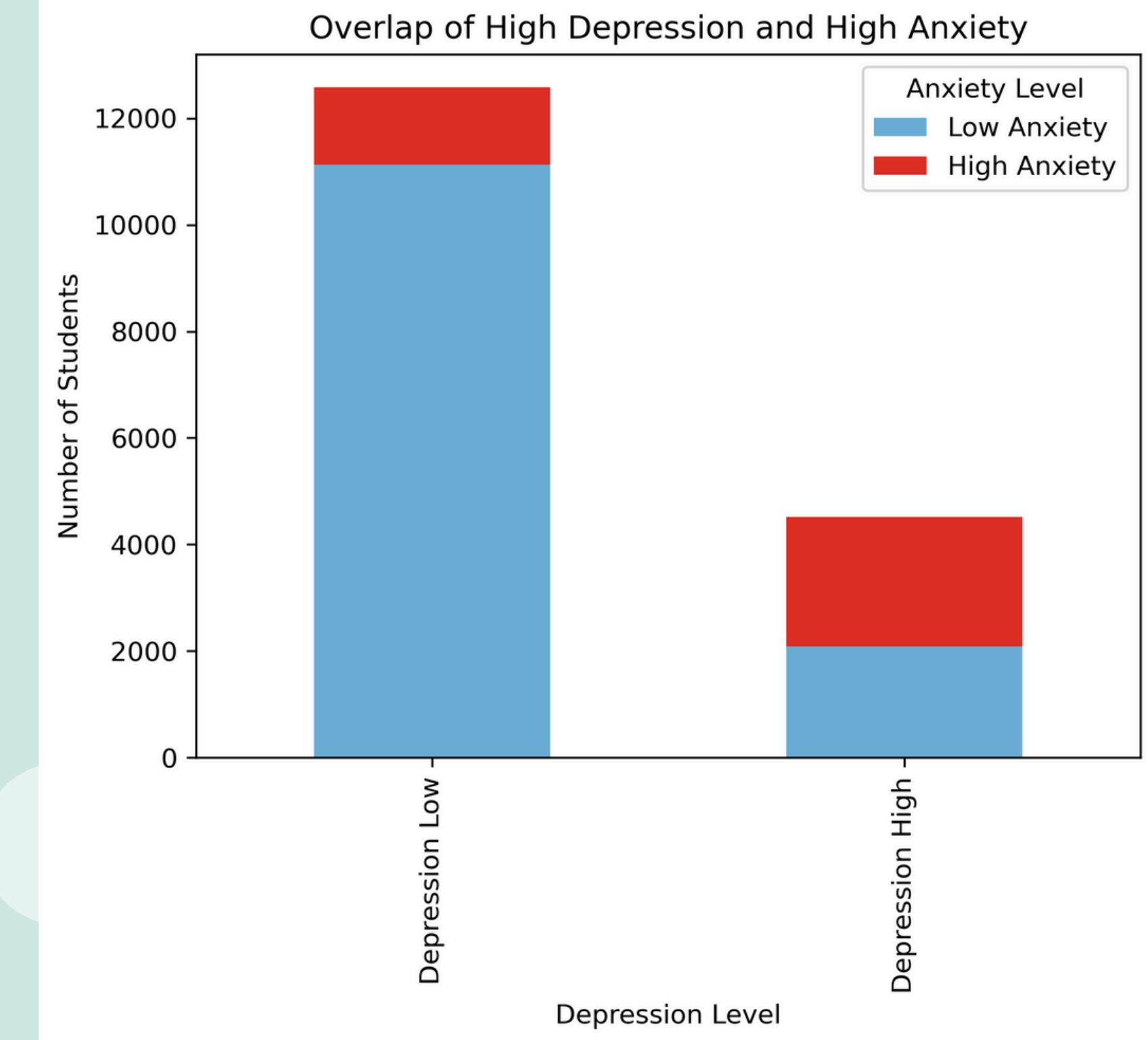


The min and max values for GAD and PHQ are 0 to 3 and after some research we concluded that they are scores which could mean:

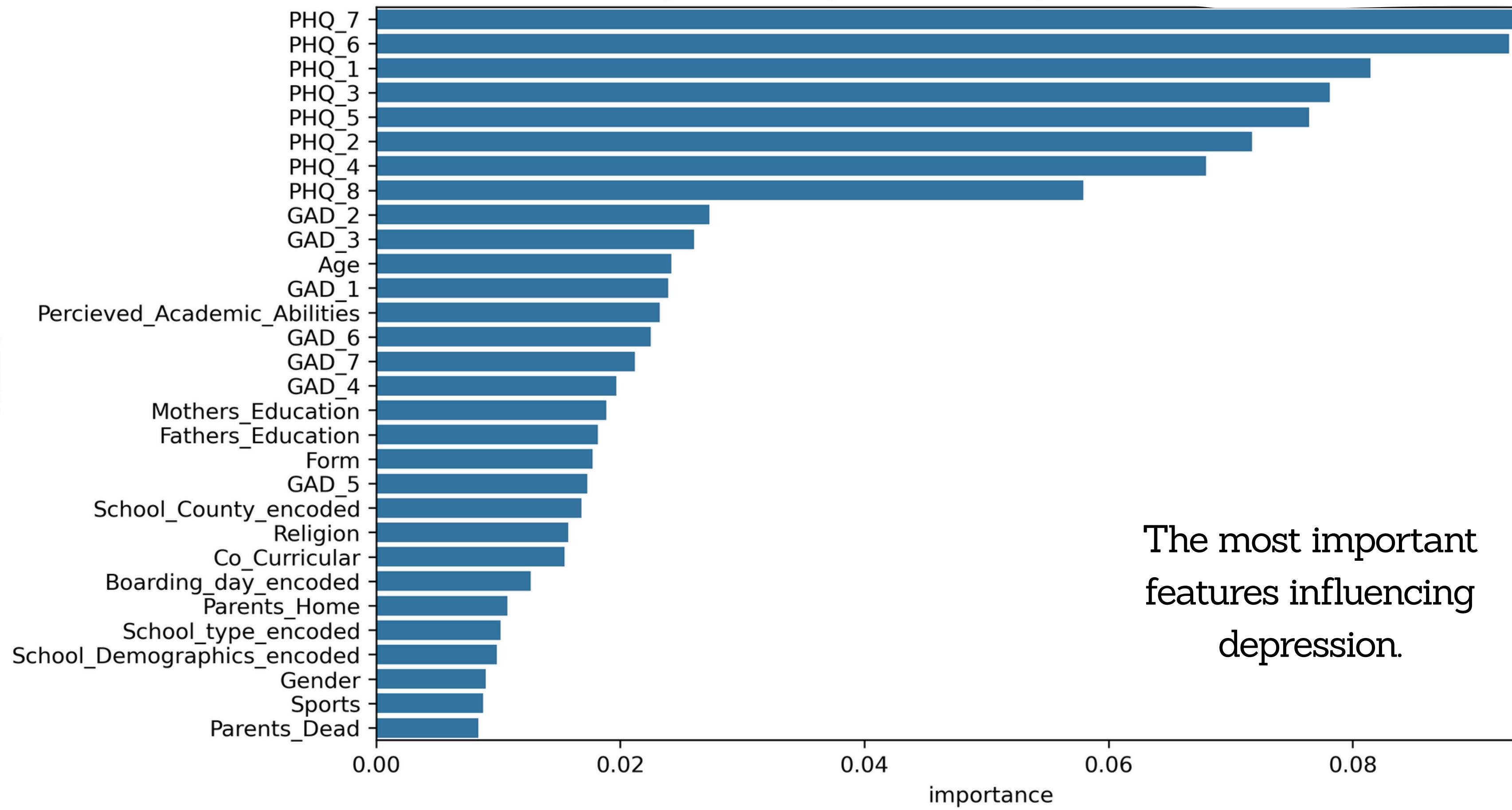
- 0 = Not at all**
- 1 = Several days**
- 2 = More than half the days**
- 3 = Nearly every day**

# Data Analysis

- This shows how many students fall into each group.
  - The red portion represents students with high anxiety, and within that, you can clearly see the large segment overlapping with high depression.

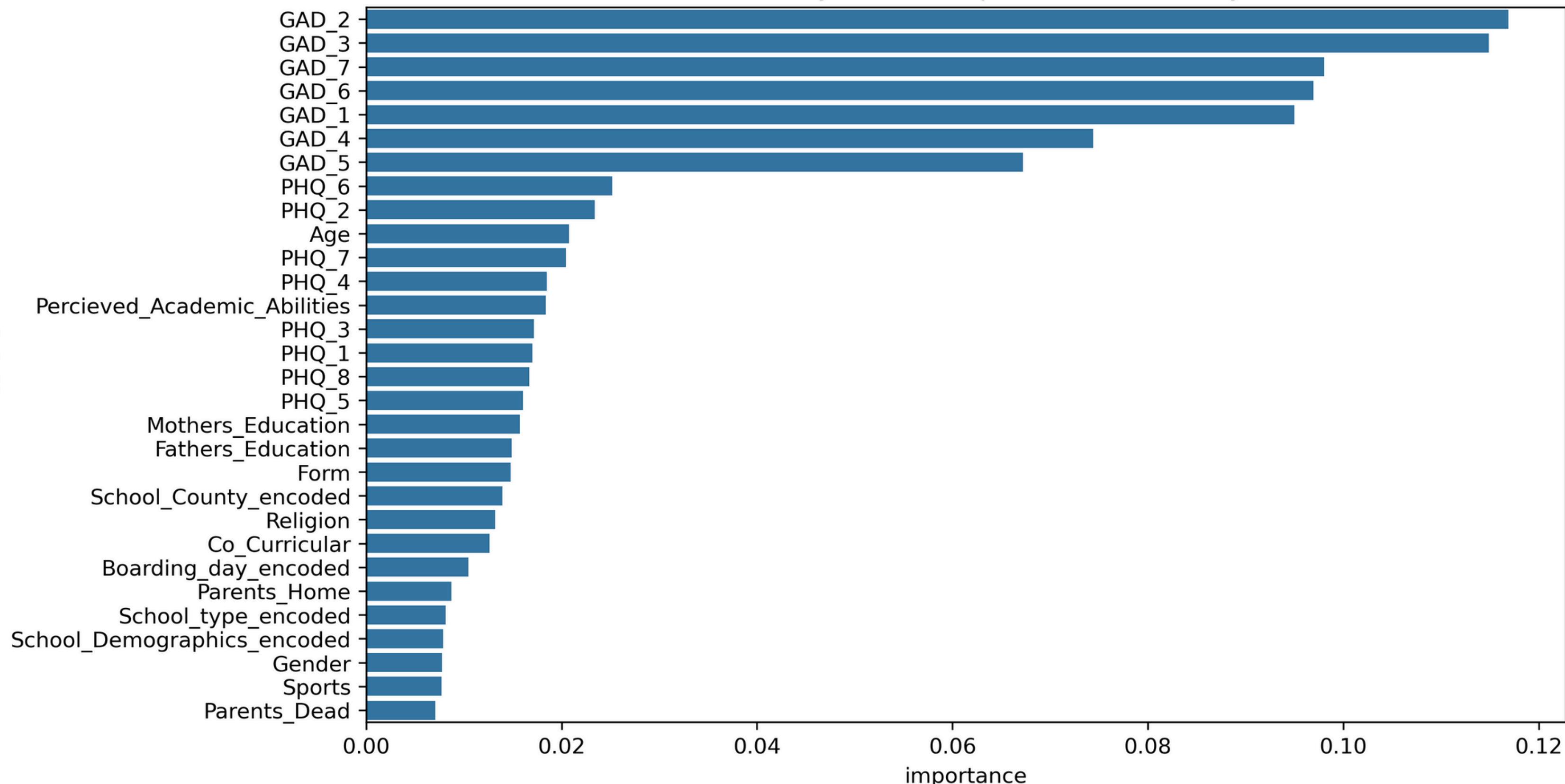


## Preliminary Feature Importance for Depression



The most important  
features influencing  
depression.

## Preliminary Feature Importance for Anxiety



These are the features that mostly influence anxiety.

# Modelling

## Is\_depressed

Model	Accuracy	Recall
Logistic Regression	0.809	0.617
Random Forest	0.806	0.597
XGboost	0.86	0.76
LightGBM	0.84	0.74

## Has Anxiety

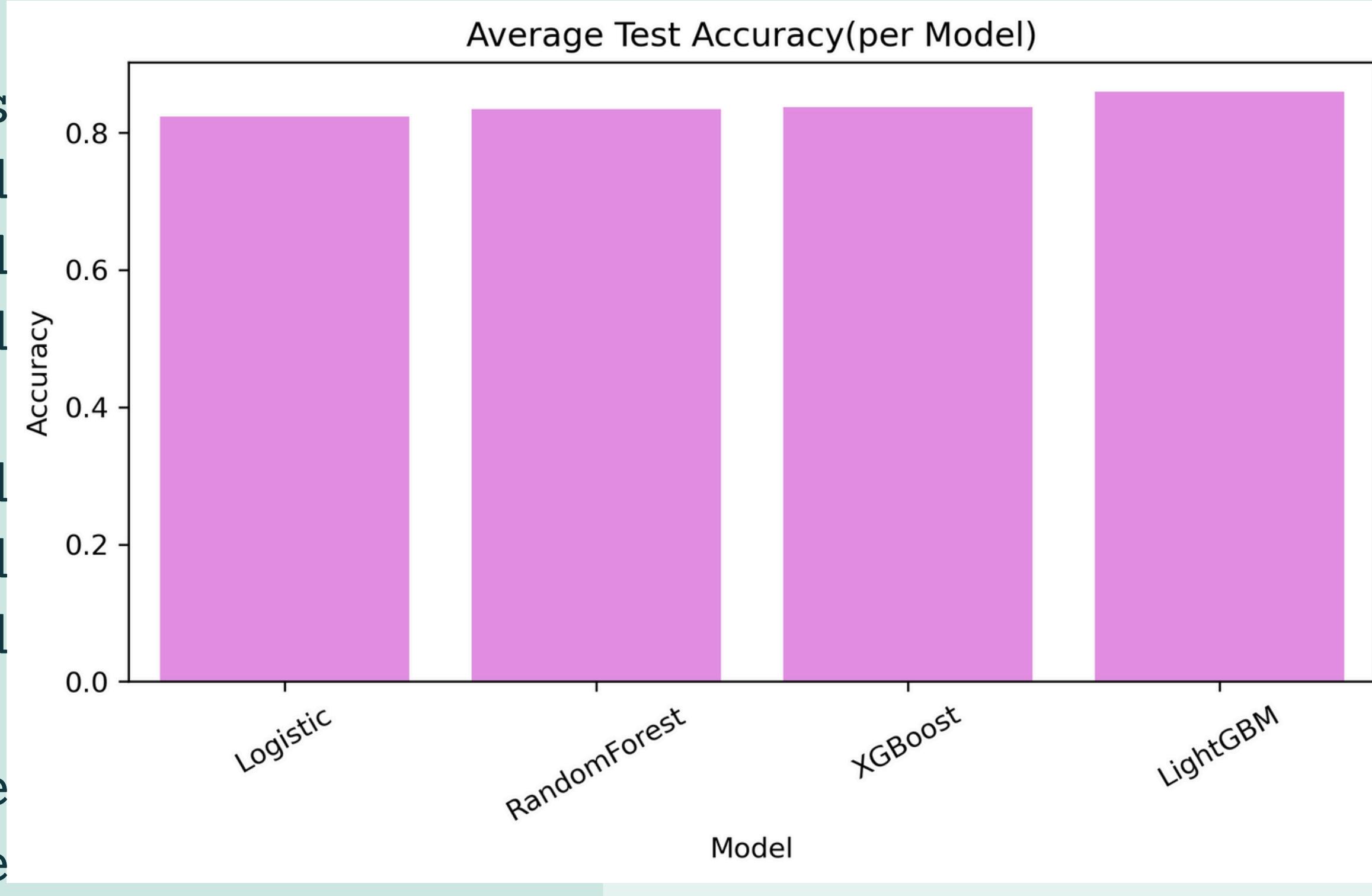
Model	Accuracy	Recall
Logistic Regression	0.87	0.817
Random Forest	0.862	0.794
XGboost	0.89	0.834
LightGBM	0.878	0.83

### Evaluation Metrics

- Recall: Ability to correctly identify true cases.
- Accuracy: Overall correctness of predictions

# Evaluation

- Overall accuracy of the models generally was high, detected severe medical and mental conditions were often missed especially for depression.
- XGBoost and LightGBM provided the best results for levels of mild and moderate anxiety and depression.
- Anxiety detection was more reliable overall, but the identification of the very severe cases continues to be a challenge across models.



The test accuracy across all models are high indicating that the models generalize well on unseen data.

# Recommendations

- Give priority to improving recall.
- Deploy the model as an early mental health screening tool.
- Collect data from all geographical locations in Kenya.
- Deploy the models simultaneously as performance varies by severity level and type of symptoms.
- Expand data collection efforts to improve detection of severe cases.



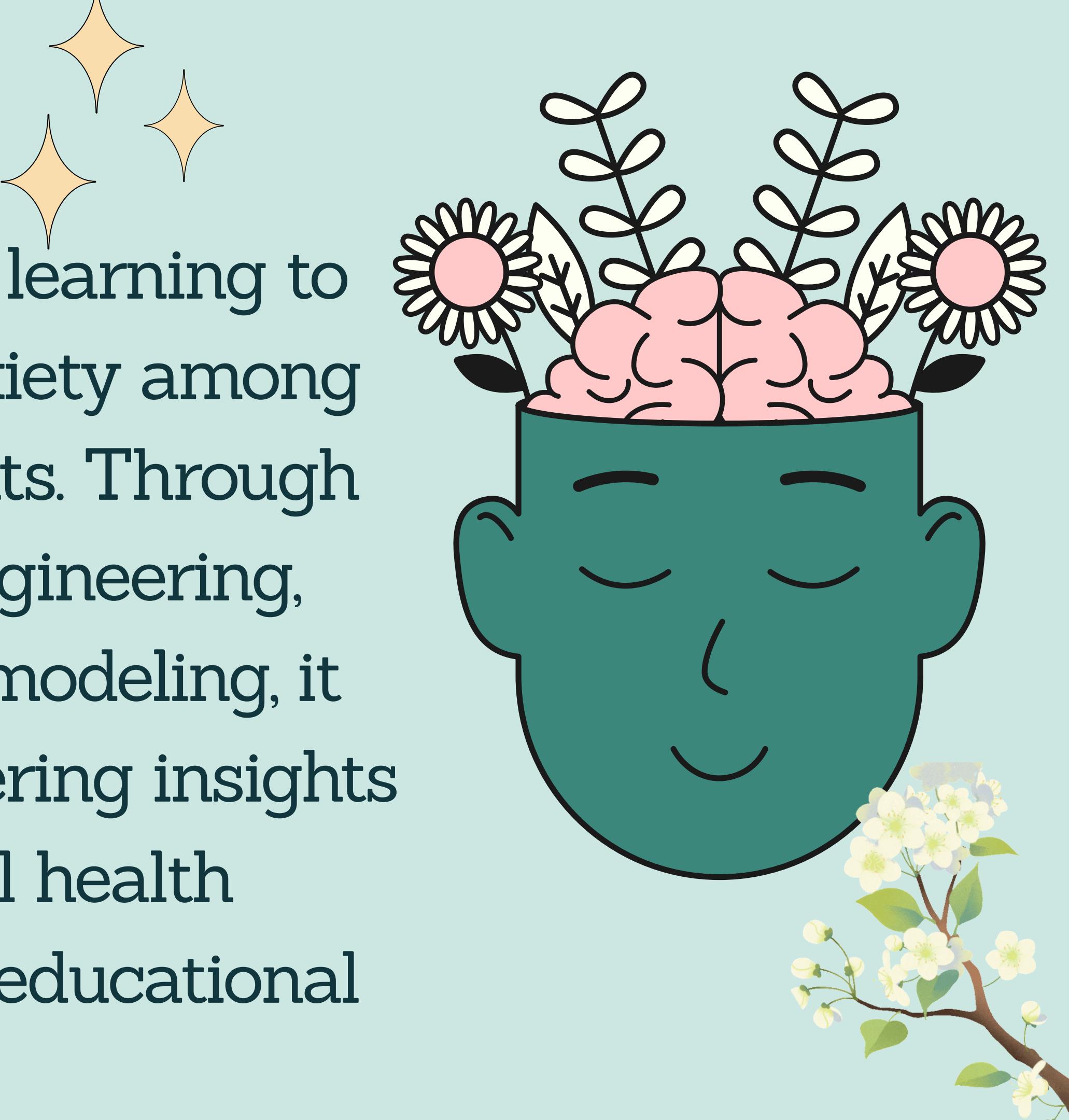
# Findings



- Top Predictors: PHQ-9 and GAD-7 responses, academic self-perception, and gender.
- Best Model: LightGBM (strongest recall(minimized false negatives) and interpretability)
- Emotional and academic self-view indicators strongly influence risk classification.

# Conclusion

This project uses machine learning to predict depression and anxiety among Kenyan high school students. Through preprocessing, feature engineering, clustering, and predictive modeling, it identifies at-risk groups, offering insights to support early mental health intervention and informed educational policies.



# Next Steps

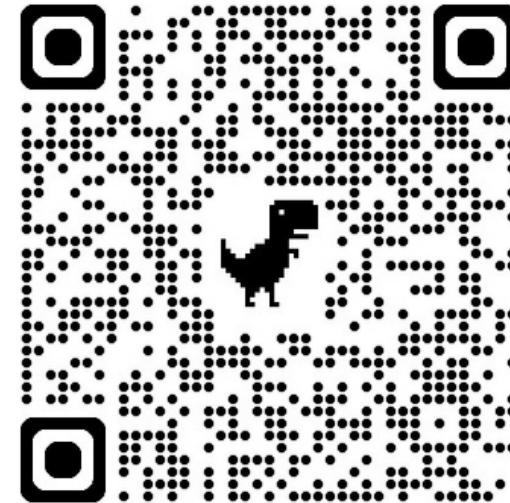
- 1.Launch and monitor.
- 2.Ensure that the system is ethically used.
- 3.Increase data collection. Sample more severe levels of depression and anxiety to enhance model generalization.



# Deployment

## AdolescentMind

Depression & Anxiety Screening for Kenyan High School Students



**Disclaimer :**

**The app gives insight and awareness and not diagnosis.  
If you score high, reach out to the provided resources  
for support.**



# Emergency

## Hotlines:

Kenya Red Cross: 1199

Befrienders Kenya: +254 722 178177

Lifeline Kenya: +254 722 178177



# Thank you

We got you

