

Mental Health and Adverse Childhood Experiences: A Relational Investigation

By: Randa Ampah, Isabel Delgado, Aysha Hussen, Aniyah McWilliams, and Jessica Oseghale

DS 6021: Introduction to Predictive Modeling
Professor Afriyie
12/12/2025



Table of Contents

01

Research Question

Motivations

02

Data

Data Cleaning

03

Methods

Lasso, HAC, Logistic, KNN, & MLR

04

Results

Accuracy, R-squared, & More

05

Conclusion

Findings & Next Steps

Research Questions

Adverse Childhood Experiences

- Adverse Childhood Experiences (ACEs) are traumatic events that occur when an individual is aged 0-17 years old

Research Questions:

- How do distinct clusters or patterns of ACEs predict mental health outcomes?
- Which specific ACEs are the strongest predictors of adult depressive disorders?
- To what extent can mental health disorders be accurately predicted from ACEs?

Data

Data Overview

- Data originates from the CDC Behavioral Risk Factor Surveillance System
- Combined data from the years 2019, 2021, and 2023 via telephone interviews

Data Cleaning

- Narrowed data down to 1,264,973 rows and 18 Columns
- Used replace map to convert numeric to a more informative numeric type
- Converted 'yes' and 'no' to 0 and 1
- Dropped N/A data



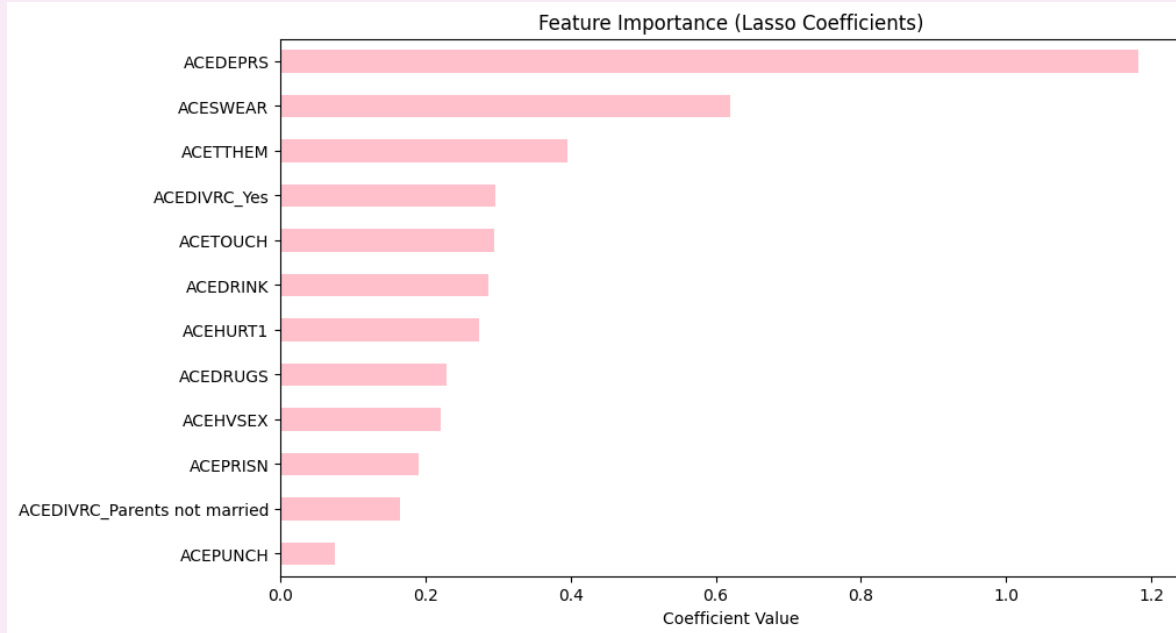
Investigation



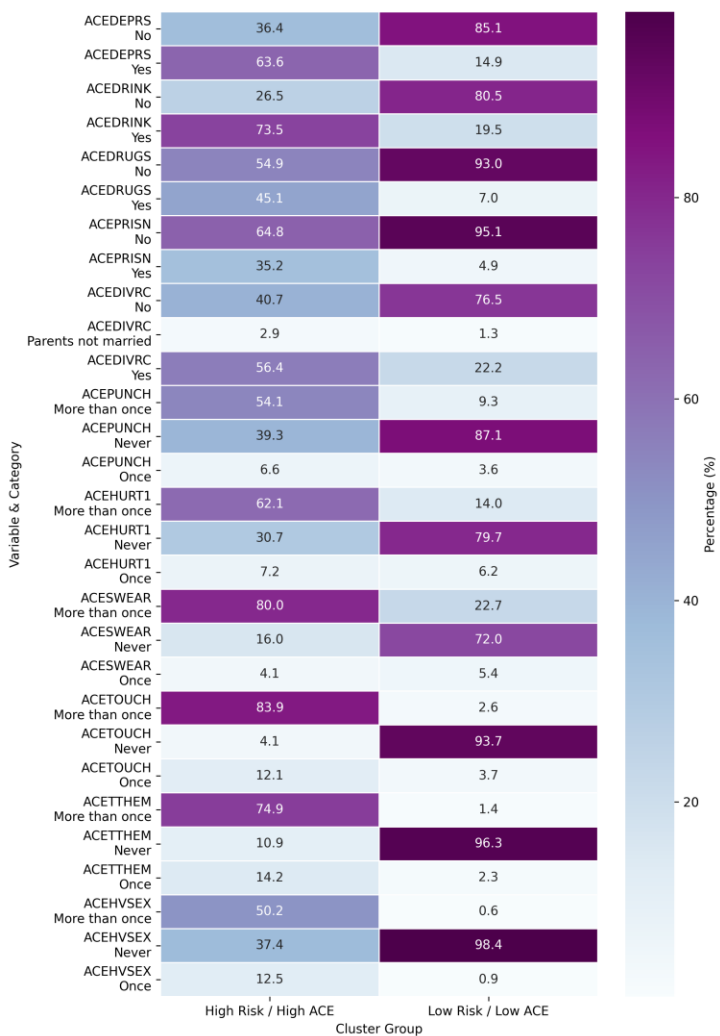
Lasso Regression

Objective: predict the number of days someone was depressed

Target Variable: MENTHLTH – how many days in the past 30 days was your mental health not good?



Proportion of ACE Categories by Cluster



Hierarchical Cluster Analysis

Method: Hierarchical Clustering (Gower Distance, average Linkage)

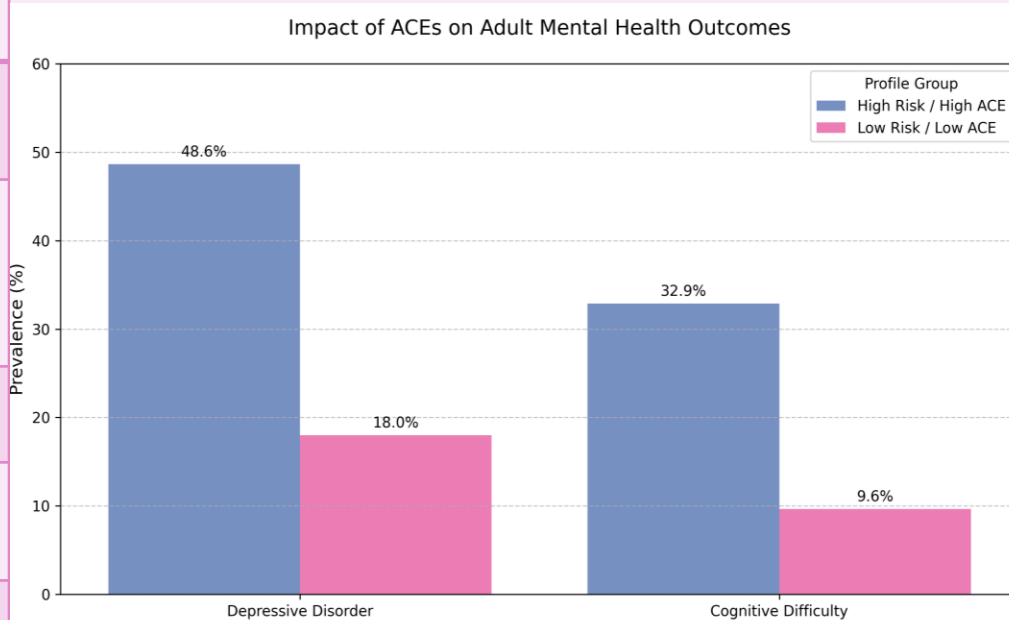
Sample: N: 10,000

Result: Two distinct profiles identified.

- **Cluster 1 (5%):** High/Multi-Type Exposure
- **Cluster 0 (95%):** Low/Baseline Exposure.

Hierarchical Cluster Analysis

Key Metric	High ACE	Low ACE	Relative Impact
Sexual Abuse Prevalence	84%	3%	28x
Witnessed Domestic Violence more than once	54%	9%	6x
Depression Diagnosis	49%	18%	2.7x
Activity Limitation days/month	10.8	3.5	3.1x
Poor Mental Health days/month	9.3	4.9	1.9x



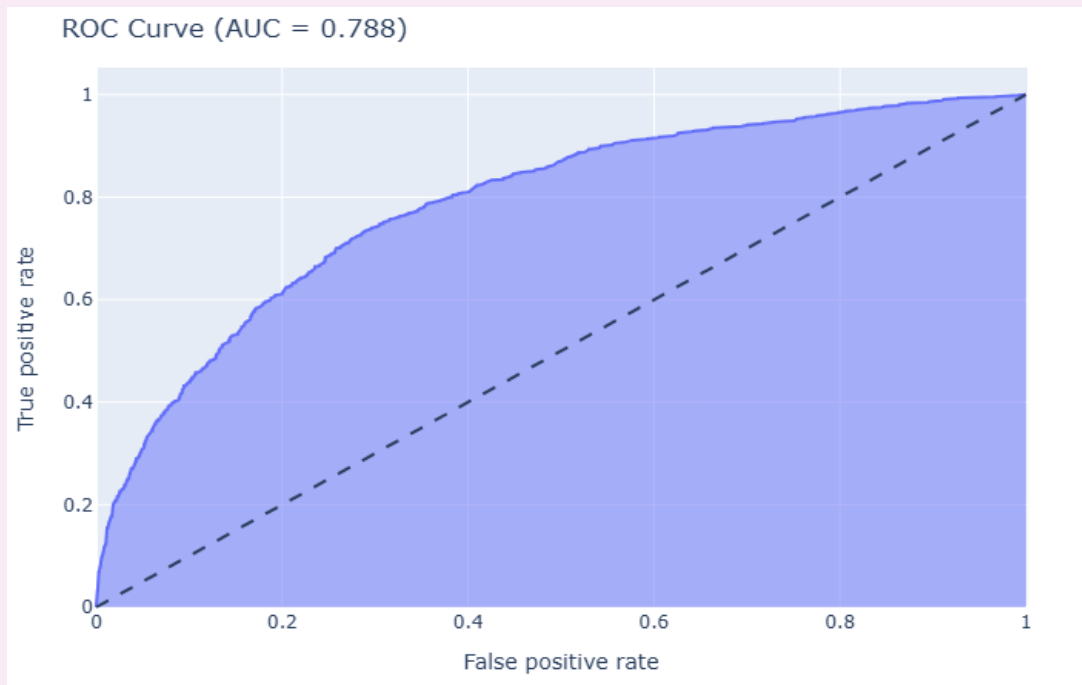


Clinical Prediction

Logistic Regression

Objective: Predict likelihood that a respondent has ever been told they have a depressive disorder.

Target Variable: ADDEPEV3: have you ever been told you have a depressive disorder?



Outcome: 0 = No diagnosis ,
1 = Has been diagnosed

Threshold: 0.70

Accuracy: 0.788

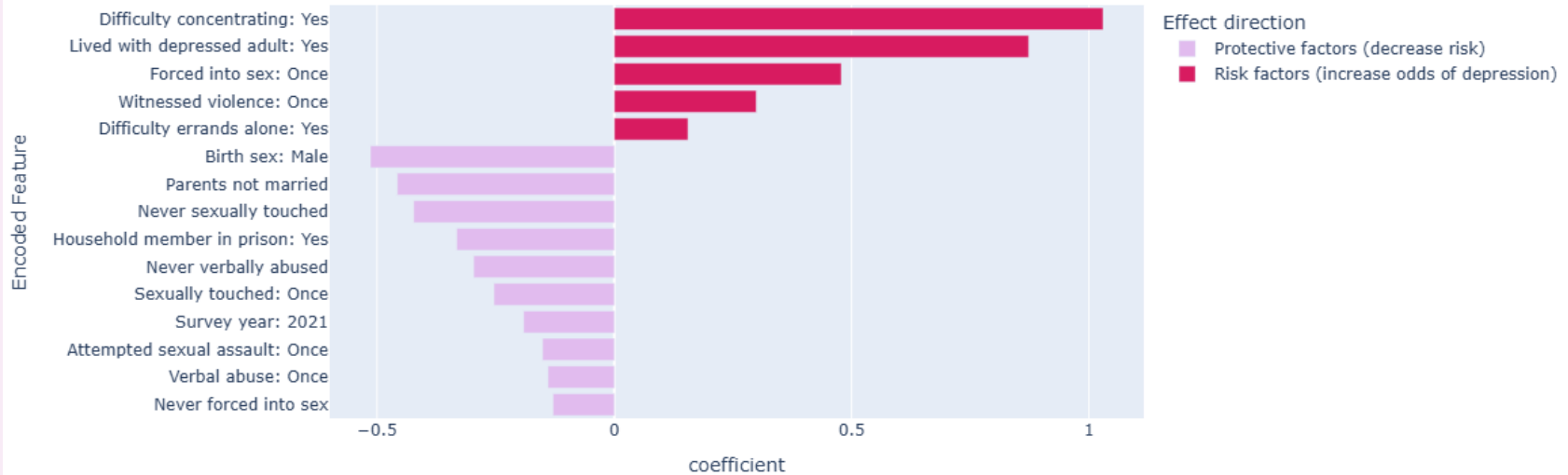
Log Loss: 0.491

Logistic Regression

Objective: Predict whether someone has been told they have a depressive disorder

Target Variable: ADDEPEV3, **Outcome:** 0 never told, 1 has been told, **Threshold:** 0.70

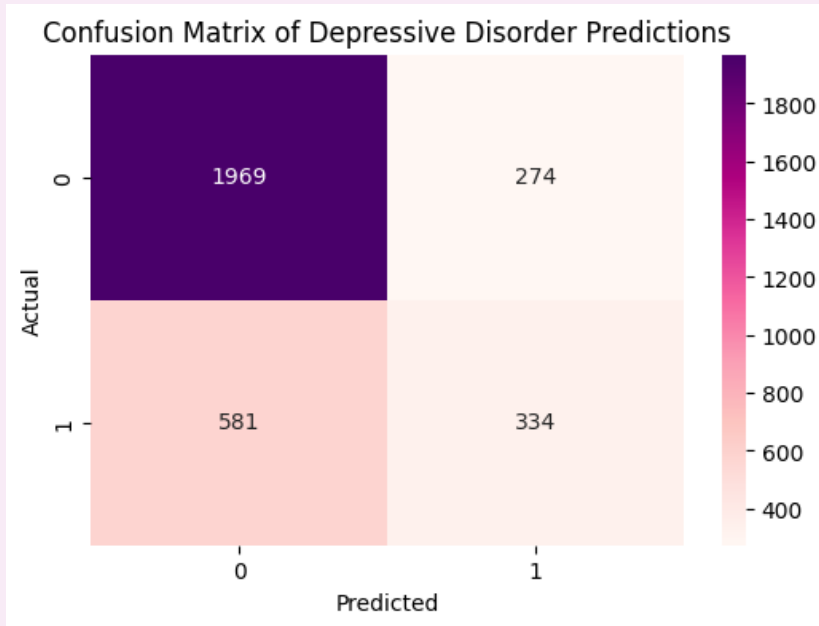
Top Logistic Regression Coefficients



K-Nearest Neighbor

Objective: predict whether someone has been told they have a depressive disorder

Target Variable: ADDEPEV3: have you ever been told you have a depressive disorder?



K = 11

Outcome:

- 0 (never had depression)
- 1 (has had depression)

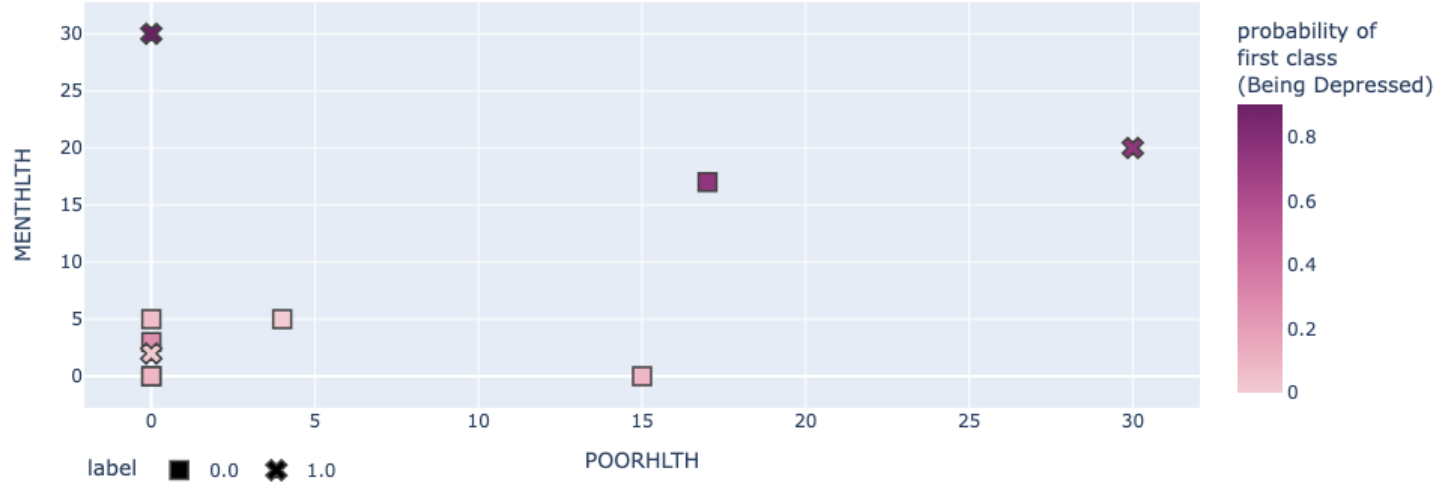
Accuracy: 0.729

Balanced Accuracy: 0.621

AUC: 0.708

K-Nearest Neighbor

KNN Depression Classification Displayed Across Days Depressed and Days Unmotivated

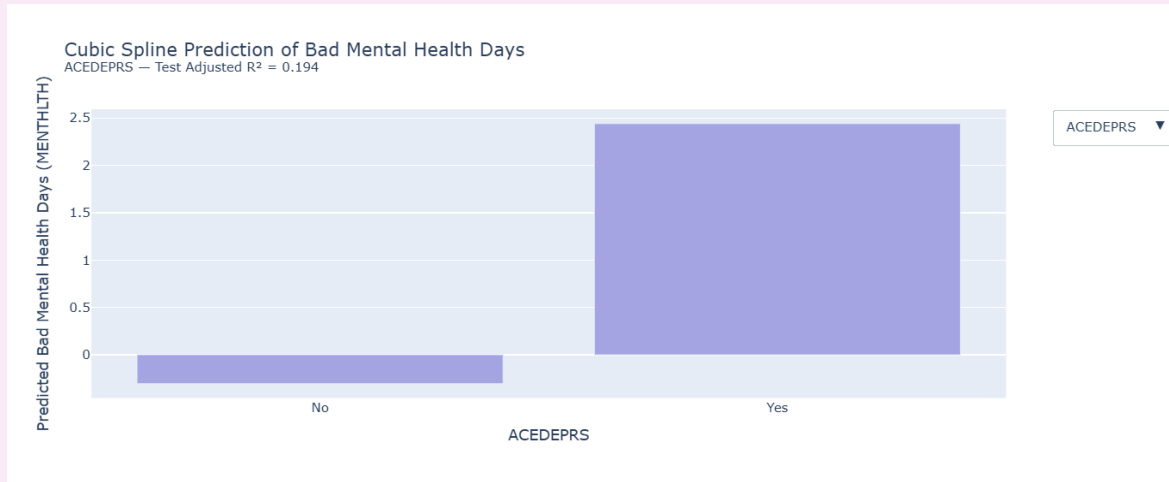


Multiple Linear Regression

Objective: predict the number of bad mental health days

Target Variable: MENTHLTH – how many days in the past 30 days was your mental health not good?

➤ Only data from 2019 & 2023



ACEDEPRS

Predicted Values:

No = -0.0304 (0 days)

Yes = 2.4398 (~2.4 days)

Conclusion

Main Takeaways

- Able to predict mental health with ~75% accuracy
- Strongest Predictors:
 - Living with someone depressed
 - Living with someone who swore at you
 - Divorced Parents

Next Steps

- Explore different K clusters for HAC to explore combination clusters within ACEs
- Incorporate more lifestyle variables into the model
 - Chronic Health Conditions
 - Age Groupings
 - Physical Activity

Thank you!

Any Questions?

We want to specifically thank Professor Johnson and Professor Afriyie!

