

Predicting Mental Illness

Ayush Ranjan



Data Science Solution

AXA Data Science Task

Overview

- Introduction
- Modelling Approach
- Results
- Conclusion

Introduction

- ▶ **Objective:** Predict mental illness based on demographic and lifestyle data.
- ▶ **Key Features:** Age, income, marital status, smoking, physical activity.
- ▶ **Challenge:** Class imbalance – fewer individuals with a history of mental illness.

Key Points:

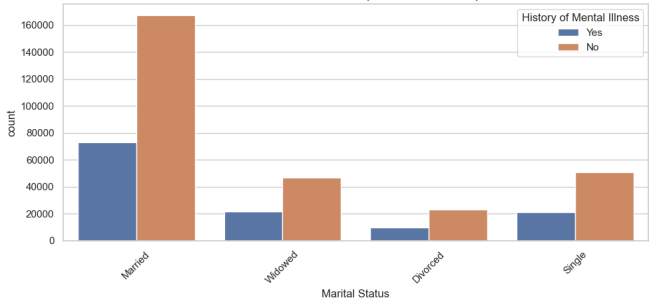
- ▶ Mental health prediction for improving public health outcomes.
- ▶ Early identification of at-risk individuals.

Dataset Overview and Insights

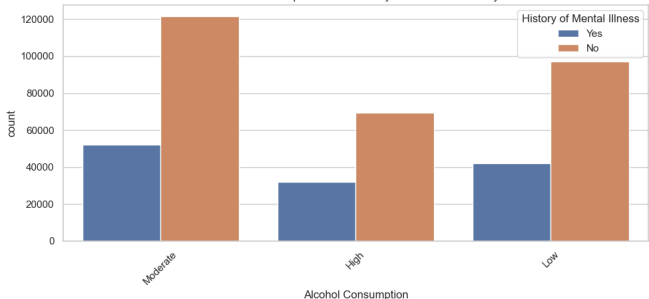
- ▶ **Size of Dataset:** 400,000+ rows.
- ▶ **Key Features:**
 - ▶ Age: Skewed towards younger individuals.
 - ▶ Income: Variation exists, which may influence mental health.
 - ▶ Lifestyle Factors: Smoking, physical activity strongly correlated with mental health.
- ▶ **Class Imbalance:** 70% no history of mental illness, 30% with a history.

Dataset Overview and Insights

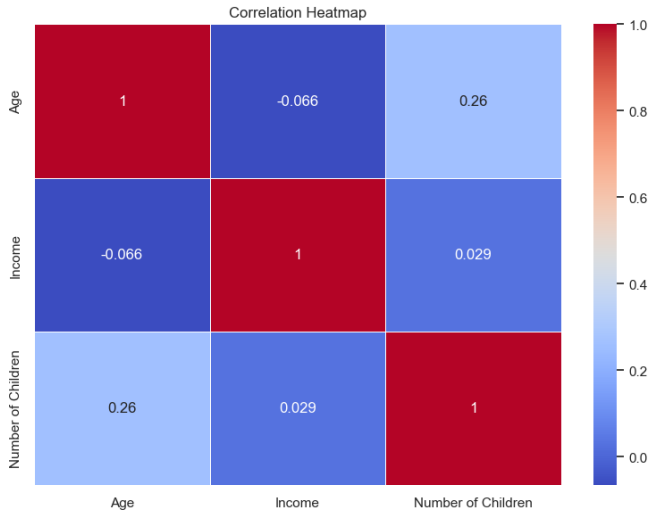
Marital Status Distribution by Mental Illness History



Alcohol Consumption Distribution by Mental Illness History



Dataset Overview and Insights



Modeling Approach

- ▶ **Models Used:**

- ▶ Random Forest
- ▶ XGBoost
- ▶ LightGBM

- ▶ **Stacking Classifier:**

- ▶ Logistic Regression as meta-learner.
- ▶ Combines predictions from base models.

- ▶ **Handling Class Imbalance:**

- ▶ SMOTE (Synthetic Minority Over-sampling Technique) to generate synthetic data for minority class.
- ▶ Class weighting to penalize misclassifications of minority class.

Results and Model Performance

- ▶ **Overall Accuracy:** 63%
- ▶ **ROC-AUC Score:** 0.523

Class 0 (No History of Mental Illness):

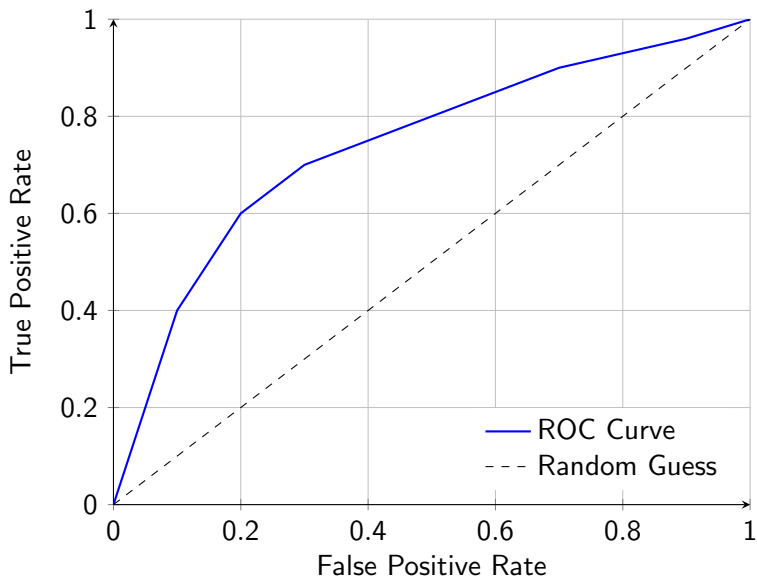
- ▶ Precision: 0.71
- ▶ Recall: 0.79
- ▶ F1-Score: 0.75

Class 1 (History of Mental Illness):

- ▶ Precision: 0.35
- ▶ Recall: 0.26
- ▶ F1-Score: 0.30

Confusion Matrix

	Predicted No (Class 0)	Predicted Yes (Class 1)
Actual No (Class 0)	45,678	11,793
Actual Yes (Class 1)	18,456	7,827



Limitations of the Current Approach

- ▶ Class imbalance remains an issue despite SMOTE and class weighting.
- ▶ Recall for Class 1 is low (26%) – many true cases of mental illness were missed.
- ▶ Bias toward Class 0 persists.

Suggested Improvements

- ▶ **Advanced Feature Engineering:** Introduce new features or interaction terms.
- ▶ **Hyperparameter Tuning:** Use GridSearchCV or RandomizedSearchCV to optimize model parameters.
- ▶ **Cost-Sensitive Learning:** Penalize misclassifications of minority class more heavily.

Conclusion

- ▶ Stacking classifier achieved reasonable performance, but class imbalance remains a challenge.
- ▶ Next steps: Focus on improving recall for Class 1 through advanced techniques.

Takeaway: Early intervention can be enabled by improving mental illness prediction models.

The End