

# Depression Prediction and Transcript Sentiment Analysis

Henry Musto

# Task Overview

## 1. Depression Prediction:

- Analyse tabular data to predict mental illness using demographic, lifestyle, and family history features.
- Utilise Machine Learning to prediction mental illness occurrence.

## 2. Transcript Sentiment Analysis:

- Analyse customer support transcripts to determine sentiment trends and call outcomes.
- Evaluate multiple sentiment models for performance and consistency.

# Depression Prediction Methodology

## 1. Data Exploration:

- Visualised categorical and numerical features (e.g., smoking, alcohol consumption).
- Analysed relationships between predictors and mental illness.

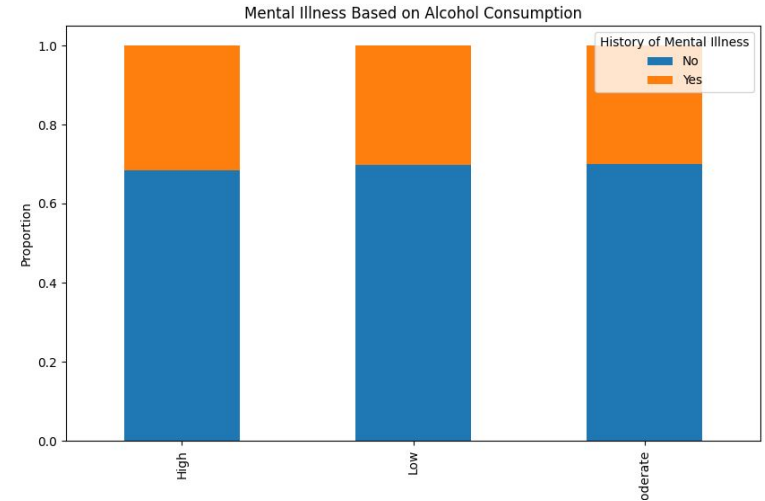
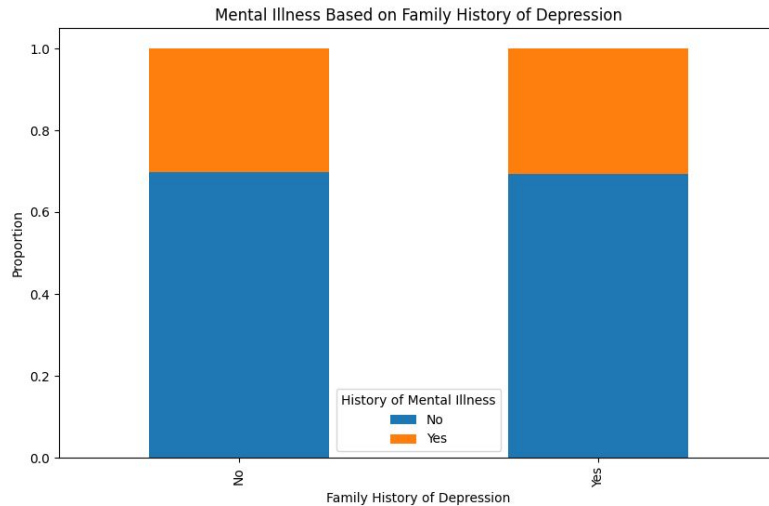
## 2. Preprocessing:

- Label encoding for categorical variables.
- Applied SMOTE to handle class imbalance.
- Scaled the data

## 3. Modeling:

- Trained Random Forest and Logistic Regression models.
- Tuned models using Bayes Optimisation.
- Evaluated models using ROC-AUC scores.

# Depression Prediction Visualisations



The relationship between predictors and the onset of mental illness is not clear and so a model may struggle to predict based on these variables.

# Transcript Analysis Methodology

## 1. Data Preparation:

- Processed transcript data to extract customer messages.
- Defined keywords for call outcomes (e.g., resolved, follow-up).

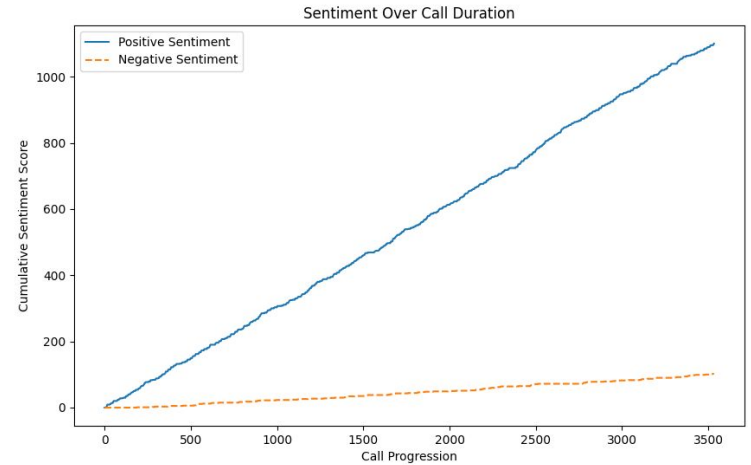
## 2. Sentiment Analysis:

- Used transformer-based models to analyze customer sentiment.

## 3. Evaluation:

- Measured agreement rate between models by running two models over the same data.
- Conducted consistency testing across model by repeatedly running the same model over the data and measuring how often it made the same prediction.

# Transcript Sentiment Analysis Visualisations



The words that are expressed most often are positive and the majority of interactions get more positive as the call progresses, suggesting they may be working towards a positive outcome.

# Insights and Next Steps

## 1. Insights:

- *Mental Health* - Relationship between variables and outcome was unclear, although the models did pick up some predictive ability with LR doing better than RF.
- *Transcripts* - I was able to consistently predict sentiment and outcome with an agreement of 1.0 between two separate models.

## 2. Next Steps:

- Enhance feature engineering for better model performance.
- Better define good performance for sentiment modelling, possibly using manually labelled data.