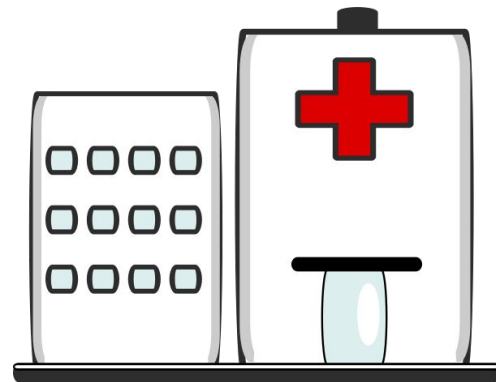# Obesity Prediction

Lucas Wilkerson

Summary

Data from the 2022 BRFSS was analyzed

Predictive modeling utilized:

- **Goal → Build a model to predict Obesity risk/ presence**

Insight generated to guide public health professionals focus and resources to tackling obesity

Outline

- Business Problem/ Stakeholder
- Data/Methods
- Modeling
- Conclusions/Recommendations

Business Problem/ Stakeholder

A team of public health professions wants to improve obesity rates by:

- Early risk factor detection
- Decreasing risk through public health education/awareness
- Broad Health Interventions/Initiatives

Data/Methods

EDA:

- 209405 Rows
- 39 Columns

Highest Correlated Features:

- Overweight
- General health
- Exercise
- Physical health
- Mental health
- Alcohol consumption

Prevalence of Obese Individuals

139429 (66.6%)

69976 (33.4%)

Normal Weight vs Obesity

Modeling

Models:

- **Logistic Regression Model**
- **Random Forest Classifier**
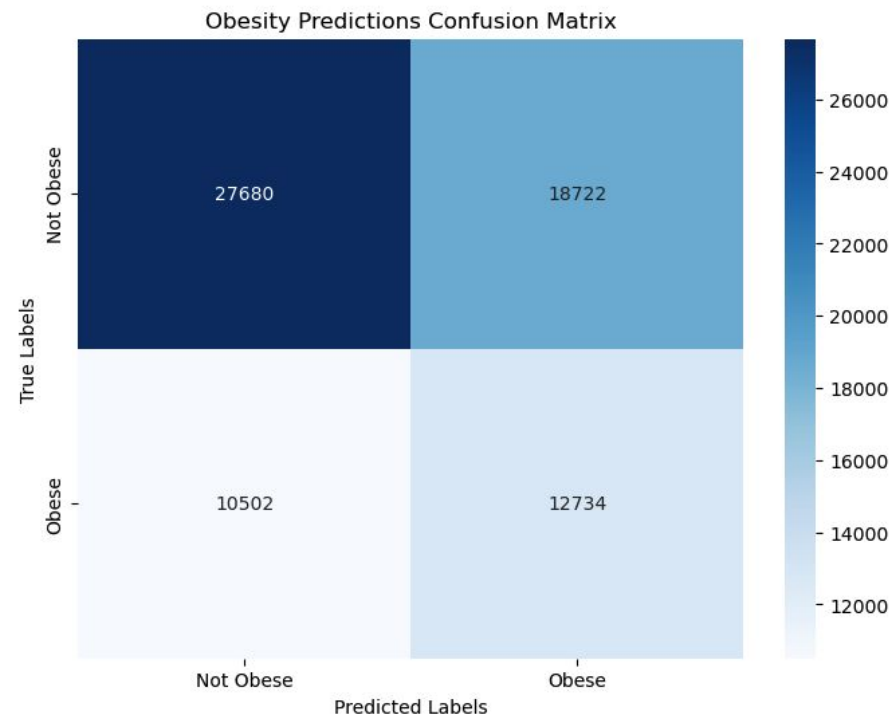- **Neural Network**

Best performing model:

- **Tuned Logistic Regression**
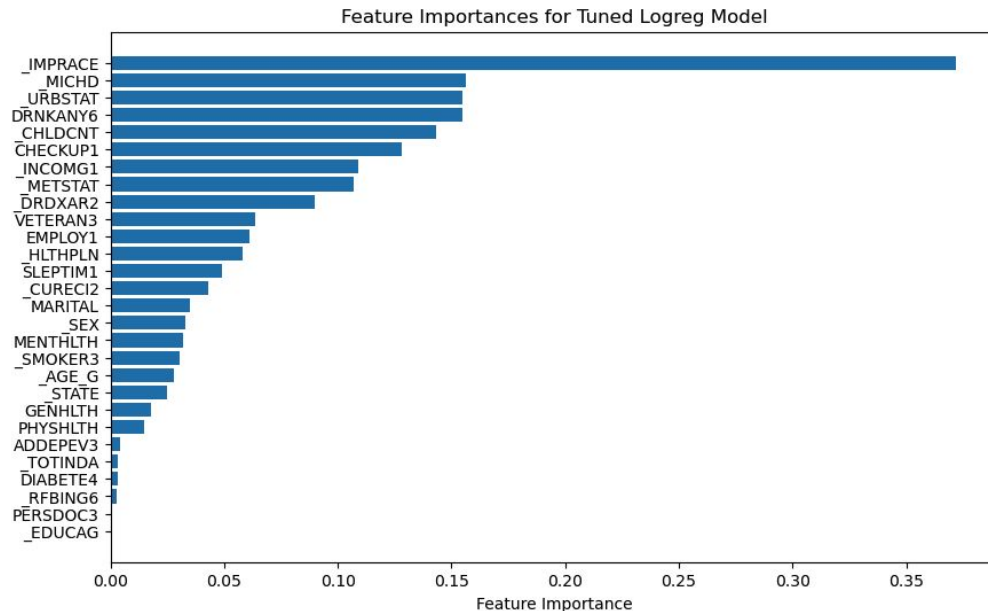
Model Performance: Best Model

Evaluation Metrics:

- Accuracy Score: **0.6912**

- F1 Score: **0.6702**

- Recall Score: **0.9402**

Obesity Predictions Confusion Matrix

Feature Importance:

Most Important Features:

- Race

- History of Heart attack or
  heart disease

- Urban/Rural Location

- Alcohol Consumption

Feature Importances for Tuned Logreg Model

Conclusions/Recommendations

Early Risk Factor Detection:

- Early detection/ screening tool

Public Education/ Awareness:

- Public health education initiatives

Health Interventions:

- Addressing health areas such as general and mental health along with exercise

Limitations

- Higher chance of false positives with higher recall score
- Extensive dataset to analyze with limited computation resources
- Dataset was filtered down and had a high amount of missing values

Thank You!

Repository Link: predicting_obesity
Email: ldwilker10@gmail.com
GitHub: @ldwilker10
LinkedIn: https://www.linkedin.com/in/lucasdukewilkerson/