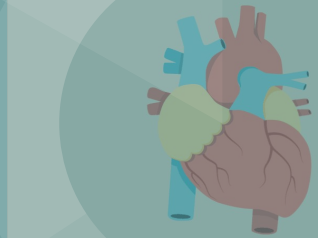
A Quest for Heart Attack Prediction



AI Bootcamp Project 2

Manali Patel, Matthew Dlugosz, Katia Monteros, Madeline Jankowski

Background

Every 34 seconds, someone in the United States dies from heart disease and CVD remains the number one cause of mortality worldwide.

With an increase in the aging population, this burden becomes overwhelming to health care systems

Many people have risk factors for heart disease that can lead to heart attacks, some of which are modifiable.

Risk predictions can help lead preventative medicine and lifestyle modifications

Project Overview

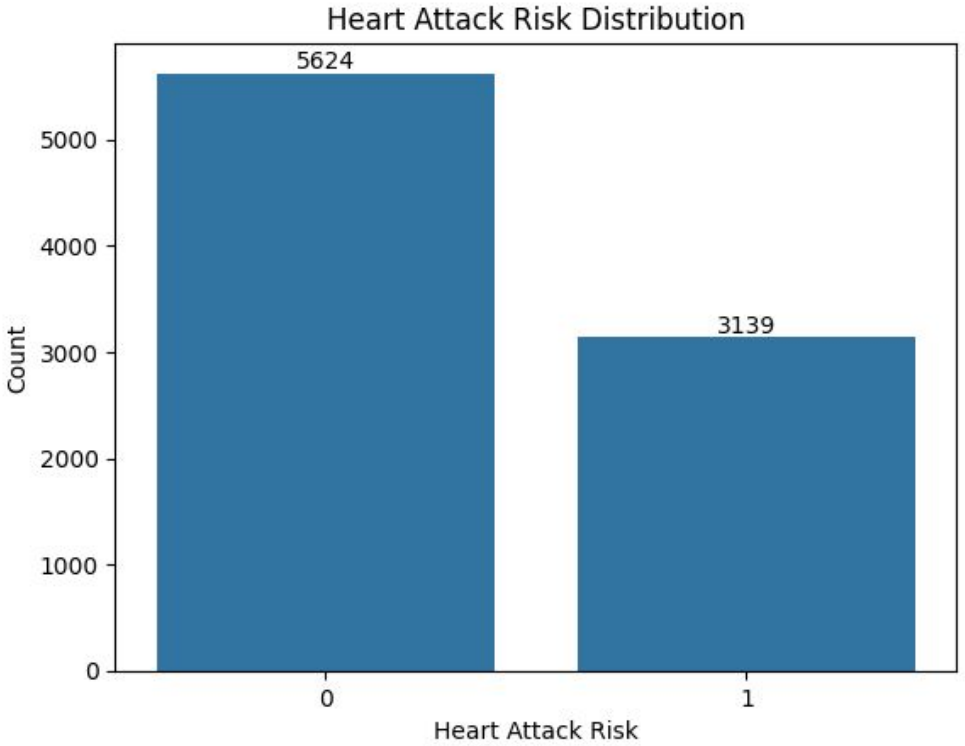
Hypothesis: We predicted that the studied traditional and risk-enhancing features would be the leading predictors in our machine learning model and could keening assess one's risk for having a heart attack.

Aim: We aimed to train a machine learning model on the data to predict the risk of a heart attack

https://www.heart.org/en/health-topics/heart-attack/understand-your-risks-to-prevent-a-heart-attack

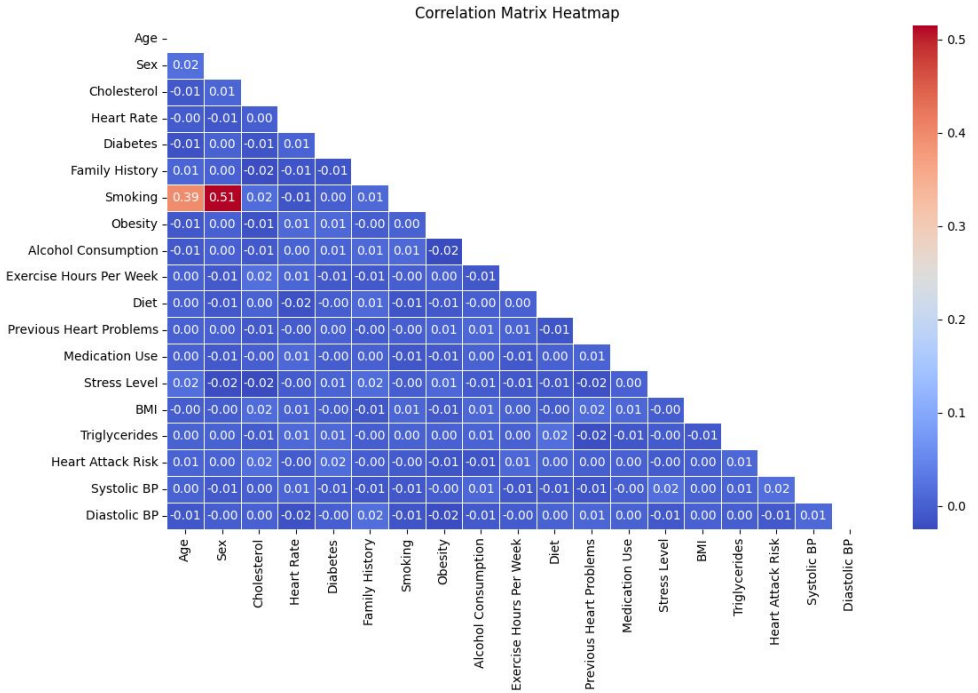
Dataset Analysis

- 8,763 patients
- 26 features including:
  - Demographics
  - Medical history
  - Lifestyle factors
  - Socioeconomic factors



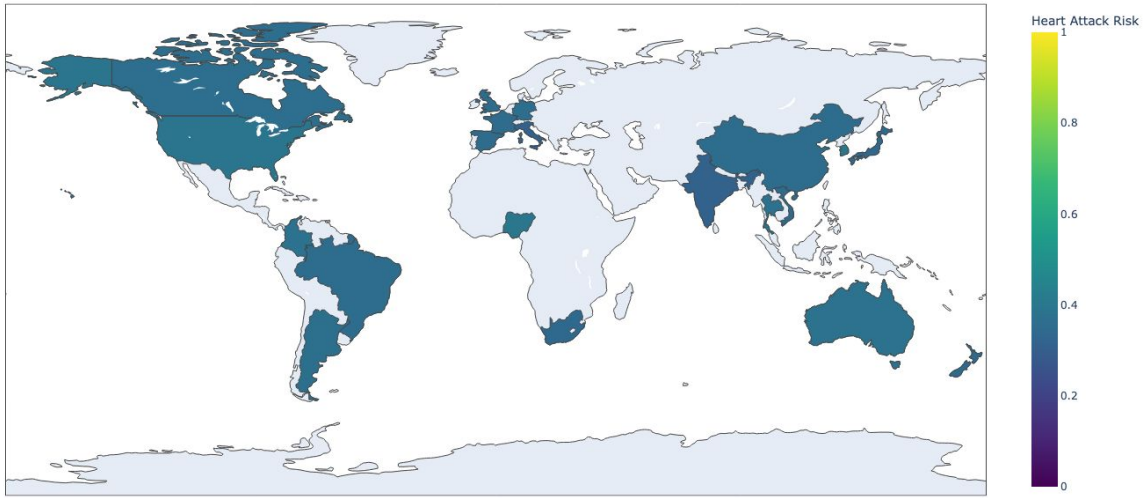
Dataset Analysis

- Smoking vs Age: 0.39
- Smoking vs Sex: 0.51



Dataset Analysis

Heart Attack Risk by Country from dataset

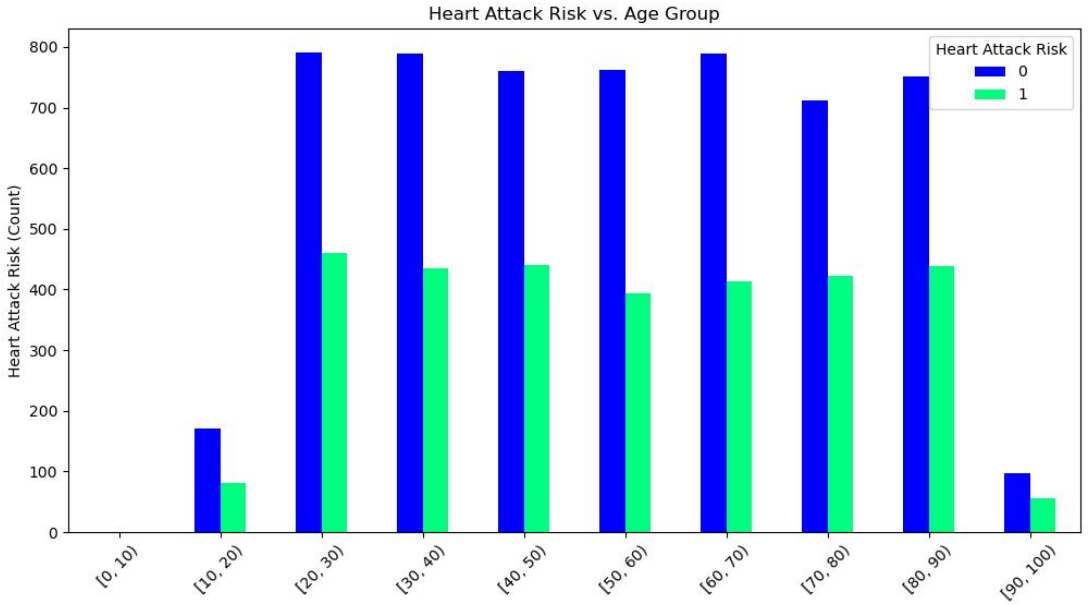


Known geographical areas with increased burden of CVD:

- Central Europe
- Eastern Europe
- Central Asia

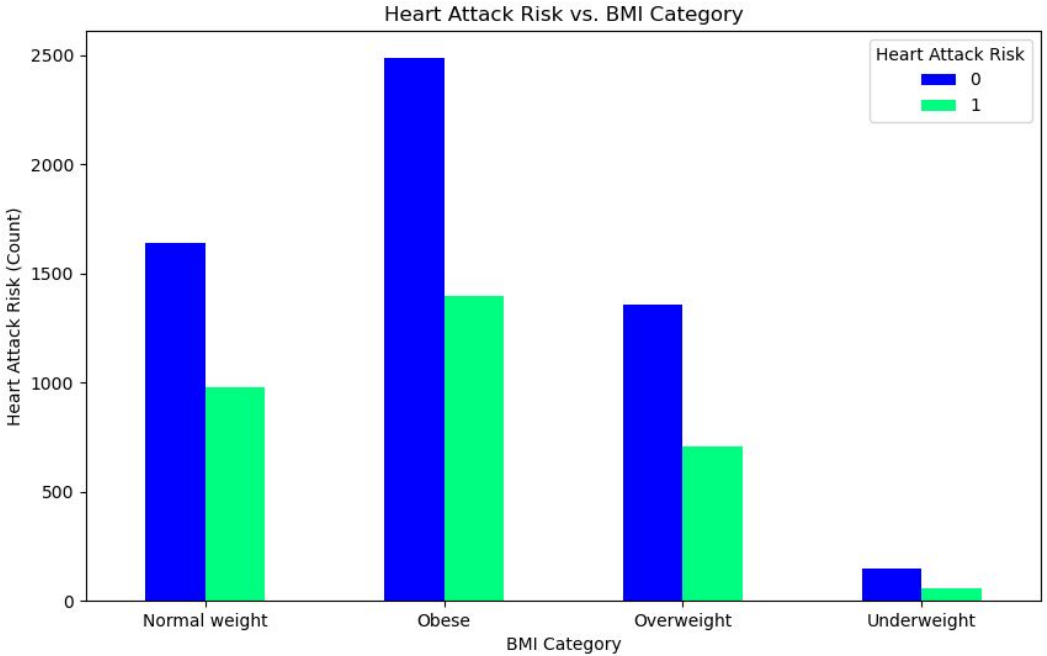
https://world-heart-federation.org/wp-content/uploads/World-Heart-Report-2023.pdf

Traditional Risk: Heart Attack Risk vs Age



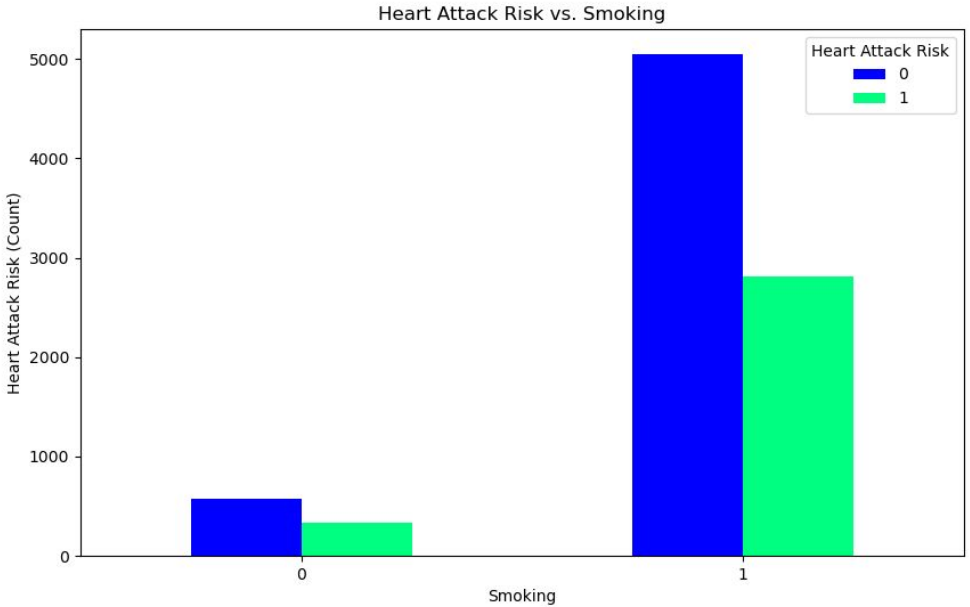
Studies show ages 55-75 are at highest risk for heart attack

Traditional Risk: Heart Attack vs BMI

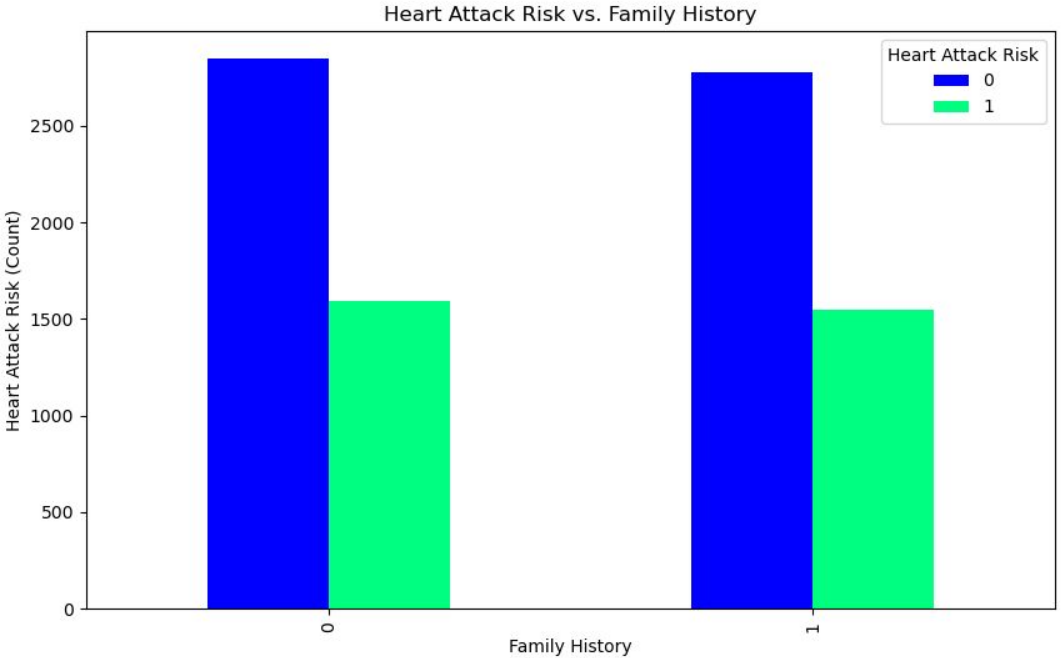


BMI Category (kg/m^2)
Underweight: 18.5 >
Normal Weight: 18.5 - 25
Overweight: 25 - 30
Obese: < 30

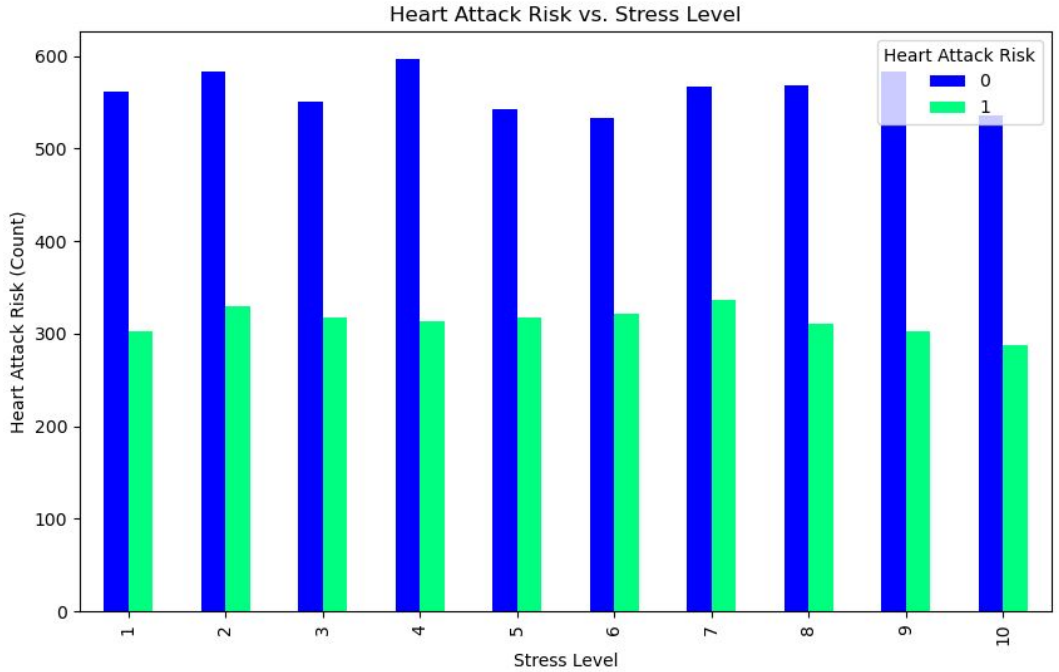
Traditional Risk: Heart Attack Risk vs Smoking



Risk Enhancing Factors: Heart Attack Risk vs Family History



Other Contributors: Heart Attack Risk vs Stress Level

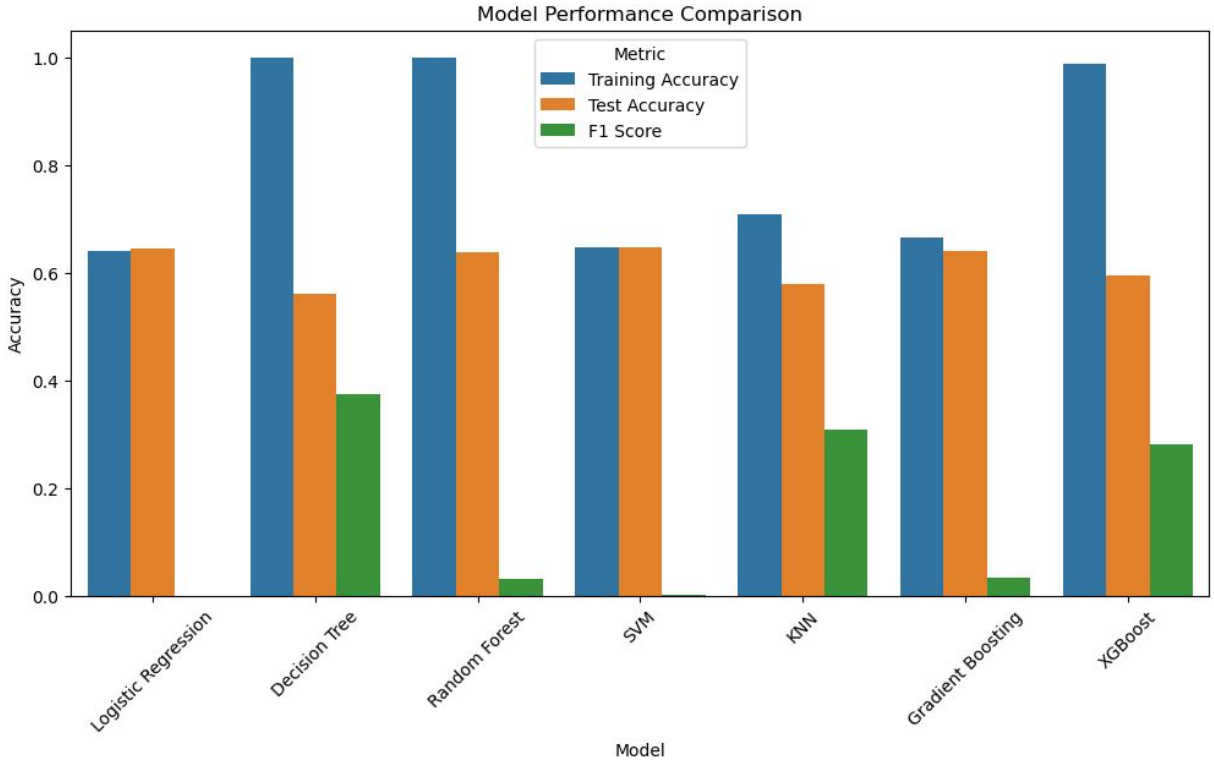


Approach to project goals

- Machine Learning Models (with optimization)
  - Logistic Regression
  - Random Forest Classifier
  - Decision Tree Classifier
  - XG Boost
  - Gradient Boost
  - Support Vector Machine (SVM)
  - K-Nearest Neighbor (KNN)
- Model Evaluation Metrics
  - Accuracy
  - Precision
  - Recall
  - F1 Score

Results

- LR was the top performing model and had most similar scores between train and test sets
- No model performed better than a 65% on test set
- Other metrics were also considered



Results

- F1 score are low
- Potential Imbalance issues
- Accuracy is misleading
- Minority is the positive class

Model Comparison:

Model	Training Accuracy	Test Accuracy	F1 Score
Logistic Regression	0.640	0.646	0.000
Decision Tree	1.000	0.564	0.374
Random Forest	1.000	0.643	0.055
SVM	0.647	0.646	0.003
KNN	0.709	0.578	0.308
Gradient Boosting	0.665	0.639	0.034
XGBoost	0.988	0.594	0.281

Conclusions

- Based on our hypothesis, we predicted that traditional and risk-enhancing features would be key predictors for our model
- But when really diving into the data, we see that the algorithm is constantly predicting no heart attack (0).
- When looking at individual traditional and risk-enhancing features, you can see that there is a slight correlation to heart attack risk.
    - Underweight had the highest percent difference at an 84%

Barriers

Future Development

- Using a different dataset with actual patient data with included outcomes
  - Hospital databases could be combined, but hard to access
- Predict the risk of overall cardiovascular disease instead of just heart attack
- Establish connections with programs like Program for Cardiovascular Health Metrics at University of Washington that are focusing on areas such as:
  - Global burden of disease
  - Disparities in CV health outcomes
  - Registry and outcomes science
  - Interventions and scenarios