

Final Project Executive Summary

Problem Statement

This project aims to use machine learning techniques to predict heart attack risk using a comprehensive dataset that includes a wide range of health and lifestyle parameters. By identifying key predictors of heart attack risk, this study will provide insights that can inform preventive measures and healthcare strategies, ultimately aiming to reduce the incidence and impact of heart disease.

Modeling And Results

Both models were trained using the preprocessed data. The Random Forest model ended up being the slightly better model. This model demonstrated an okay performance: 62% accuracy, 25% precision, 1% sensitivity, and 98% specificity. The ROC curve and AUC analyses confirmed the model's risk differentiation. Using more datasets and having more frequent training can increase these values, so it can be a better machine model.

Evaluation

The models were evaluated using several metrics, including accuracy, precision, sensitivity (recall), and specificity. Additionally, ROC curves and AUC values were used to assess the models' ability to differentiate between high and low-risk individuals. Feature importance analysis using the Random Forest model identified several significant predictors of heart attack risk. The top features included BMI, Triglycerides, Exercise Frequency, Heart Rate, Age, and Cholesterol.

Conclusion

This project provided valuable insights into the application of machine learning techniques for predicting heart attack risk. Key learnings include: Importance of Data Preprocessing, Model Selection, Feature Importance.