**CS-UY 4563: Machine Learning**

**Final Project Written Report**

**Heart Attack Prediction**

**April 29, 2024**

**Professor Linda N. Sellie**

**Thu Vu, Ariel Wang**

Introduction

This project focuses on the development of predictive models to identify the risk factors associated with heart attacks using both supervised and unsupervised machine learning techniques. For the supervised analysis, we implemented models including Logistic Regression, Support Vector Machines (SVM), and Neural Networks. These models were trained to predict heart attack occurrences based on a set of labeled data points. In the realm of unsupervised learning, we employed the k-means clustering algorithm to uncover inherent groupings and patterns within the data that could signify hidden risk factors. Additionally, we incorporated various feature transformation techniques and regularization methods to enhance the performance and generalization ability of our models. The specifics of the feature transformations and regularization strategies employed will be detailed further in subsequent sections of this report.

Dataset

The dataset utilized for our heart attack prediction project comprises 303 individual records, each annotated with a range of clinical features pertinent to cardiovascular health assessment. The features include 'Age', representing the patient's age, and 'Sex', indicating the patient’s gender. 'exang' captures whether exercise induces angina, with values 1 for yes and 0 for no. 'ca' denotes the number of major vessels detected through fluoroscopy, ranging from 0 to 3. 'cp' relates to the type of chest pain experienced by the patient, classified into four categories: typical angina, atypical angina, non-anginal pain, and asymptomatic. 'trtbps' and 'chol' record the resting blood pressure in mm Hg and cholesterol level in mg/dl, respectively, the latter obtained via a BMI sensor. 'fbs' indicates if fasting blood sugar exceeds 120 mg/dl, with 1 for true and 0 for false. 'rest\_ecg' describes the resting electrocardiographic results, categorized into normal, having ST-T wave abnormality, or showing signs of left ventricular hypertrophy by Estes' criteria. 'thalach' measures the maximum heart rate achieved. The 'target' feature labels each example with 0 or 1, where 0 suggests a lower chance of a heart attack and 1 indicates a higher risk. This rich dataset forms the backbone of our analysis, allowing us to apply and evaluate various machine learning models effectively.

Preprocessing

Logistic Regression

Support Vector Machine (SVM)

Neural Network

Result

Conclusion

Work Cited

Heart Attack Analysis & Prediction Dataset

<https://www.kaggle.com/datasets/rashikrahmanpritom/heart-attack-analysis-prediction-dataset>