Joseph Krinke

IDS 702

Dr. Akande

**Introduction**

A perfusion defect is when an area of the heart has reduced blood flow under increased stress. The decreased amount of blood flow can lead to damage to the heart muscles, which can lead to heart failure and other medical problems. Categories of this disease include reversible defects and irreversible defects. The current diagnosis technique is a cardiac perfusion scan, an expensive test which requires an individual to exercise while connected to electrodes. However, this test is time-consuming, expensive, and may cause harm to patients. On top of this, symptoms for this ailment are vague, which can lead to late diagnosis and unneeded damage to the heart. The goal of this project is to build a logistic regression model to predict an individual’s perfusion status using physiological data. Being able to diagnose the disease earlier could allow for early intervention and produce better outcomes for patients. Alternatively, one could use the results from this model to determine whether to test someone, which would reduce testing costs and expose patients to less harm.

**Data Source**

The data used was obtained from the HCI Machine Learning Repository. This dataset was collected from four hospitals: the Hungarian Institute of Cardiology, University Hospital, Zurich, University Hospital, Basil, and the V.A. Medical Center. This data included 76 attributes related to a person’s demographics and heart health. However, the vast majority of research has been focused on the V.A. dataset, so I chose to use that subset for my analysis. The subset of data I chose had 14 attributes and 270 observations; no missing data was observed. A list of all variables is below.

|  |  |  |
| --- | --- | --- |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |

**Exploratory Data Analysis**

**Model Selection**

**Results**

**Conclusion**