Accessible Approach to Blood Pressure Prediction and Anomaly Detection

Introduction:

High blood pressure, or hypertension is one of the most important risk factors for morbidity and mortality. Unfortunately, the measurement of blood pressure is not as straightforward as one might expect due to its inherent variability and inconsistencies in measurement techniques. New approaches have emerged to meet this problem. However, many of them are largely tailored towards privileged groups, ignoring the frequently cited observations that those in rural communities, low-income communities and of minority status have some of the highest risk of hypertension. Thus, the primary goal of this study is to investigate how machine learning might be used in a straightforward and accessible manner to assist healthcare workers in measuring blood pressure. The study accomplishes this by comparing newer modeling techniques against traditional OLS regression and using these techniques to design a simple tool that can determine if an observed blood pressure is anomalous or not.

Methodology:

Blood pressures vary unequally across variables, so the stratification of individual’s into different blood pressure groups has been shown to improve model quality (Stamler, Jeremiah, et al., 1975). Typically, this is done with a few demographic variables such as age and gender. However, the ability of K-modes and K-prototype to cluster on categorical and continuous data provides an attractive alternative to automate and potentially improve upon the standard method of stratification.

In terms of predictions methods, the study calls for approaches whose predictions are easily interpreted. Thus, methods such as linear regression, LASSO regression, decision trees and, to a lesser extent, K-nearest neighbors are prime candidates. While many studies opt to predict blood pressure labels such as “pre-hypertension” and “hypertension”, this study aims to predict continuous systolic and diastolic blood pressure. The prediction of continuous values is necessary for the goal of anomaly detection. However, it does introduce additional complexities to the study.

Once the best predictive technique is chosen, it will be used for the purpose of anomaly detection. This can be accomplished by comparing the blood pressure predicted by the model to an observed blood pressure. If the observed is over some predetermined cut-off distance from the predicted, it will be considered erroneous. If it is under, it will be considered valid. The distance between predicted and observed will be considered in terms of both systolic and diastolic using mahalanobis distance.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Cluster** | **Gender** | **Diabetes** | **Race** | **Age** | **Hypertension Meds** | **Hypertension** |
| 1 | Male | No | White | 65 | Yes | Yes |
| 2 | Female | No | Black | 55 | Yes | Yes |
| 3 | Male | No | White | 39 | No | No |
| 4 | Female | No | White | 39 | No | No |