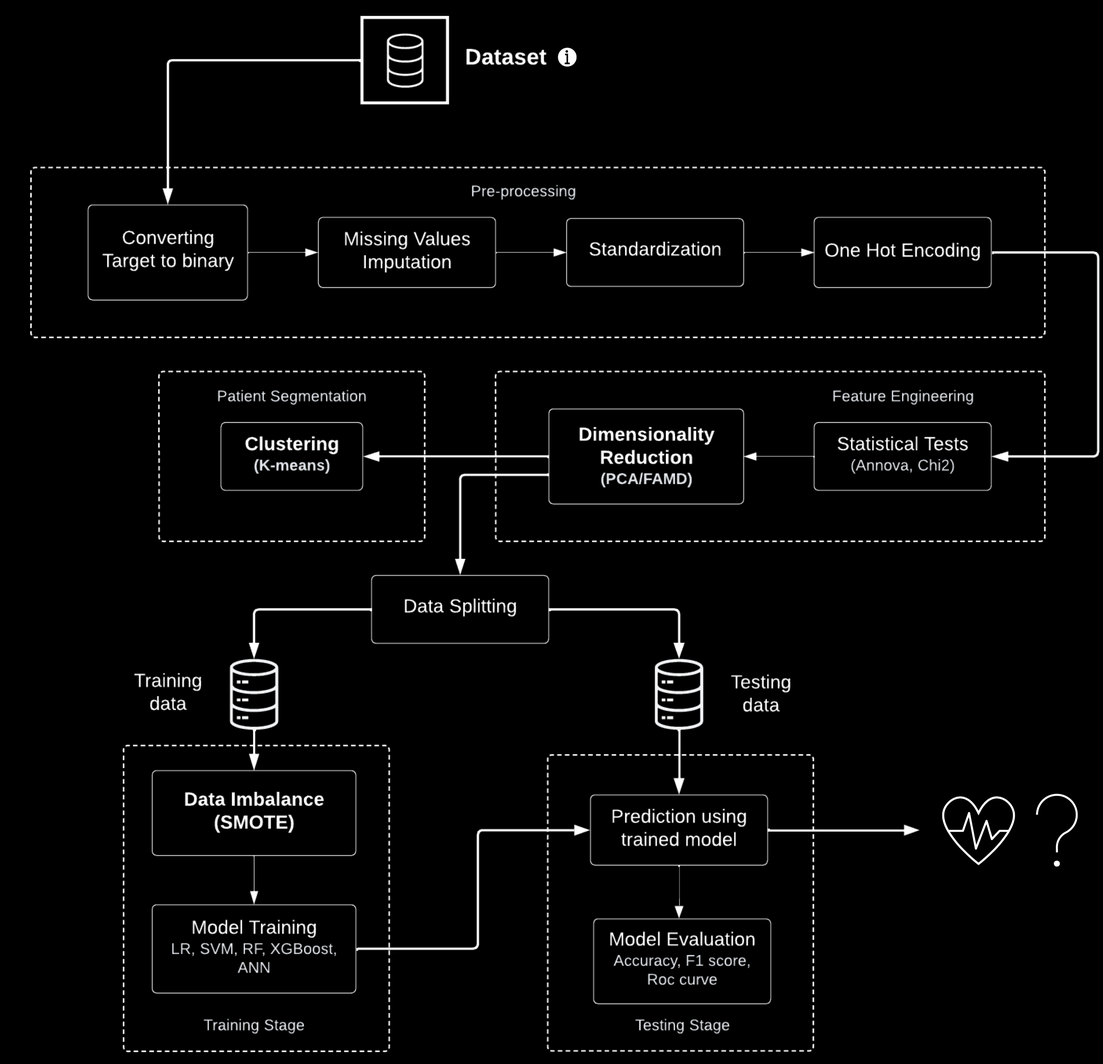
HEART FAILURE ANALYSIS AND PREDICTION

Prajwal Amin, 201685598, MSc Data Science and Analytics
Leeds University, 2022/23

UNIVERSITY OF LEEDS

Introduction

Heart disease, a leading global health concern, is at the center of this research, which aims to enhance early prediction and understanding of heart failure. The approach involves exploring the impact of dimensionality reduction and clustering techniques, which are relatively underutilized in this context, to boost predictive model performance. The ultimate goal is to equip clinicians with robust, data-driven tools, paving the way for quicker and more effective early interventions in heart disease treatment.



Objectives

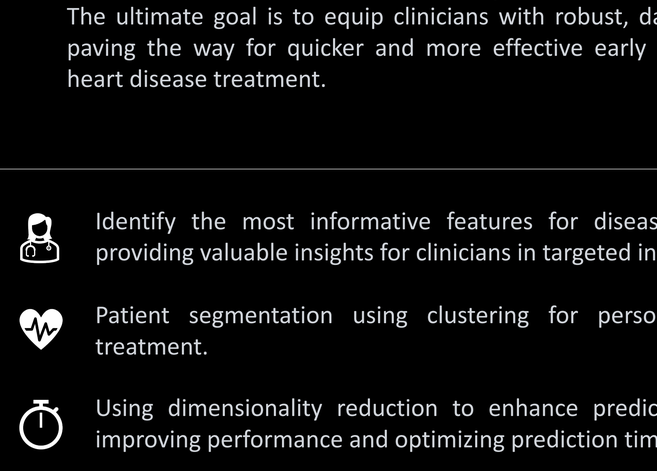
- Identify the most informative features for disease classification, providing valuable insights for clinicians in targeted interventions.
- Patient segmentation using clustering for personalized disease treatment.
- Using dimensionality reduction to enhance predictive models by improving performance and optimizing prediction time.

Dataset — *Cleveland Clinic Foundation* — *303 Samples* — *25 Attributes* *(After encoding)*

Results

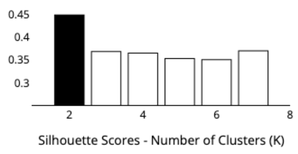
Feature Selection

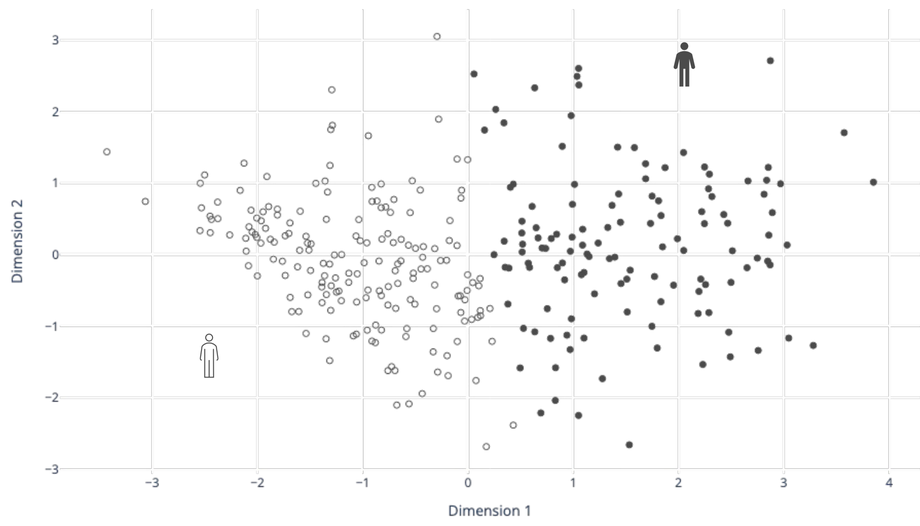
The bar chart showcases the top 15 features of higher importance, identified through SelectKBest using ANOVA (numerical) and Chi-square (categorical) test. Subsequently, the models have been trained using the data consisting of these selected features.



Patient Segmentation

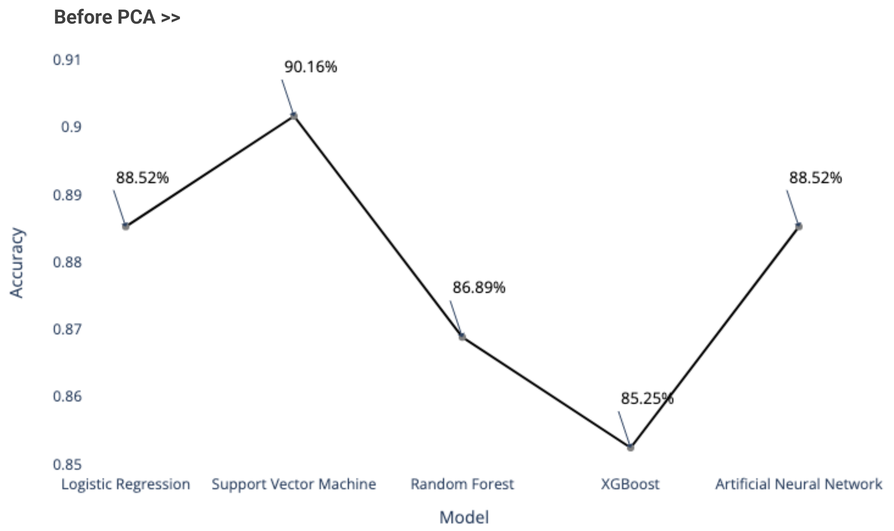
Clustering after PCA reveals distinct patient segments, guided by the optimal number of clusters using silhouette method. The clusters provide insights into patient details, enabling personalized approaches for treatment and intervention strategies.

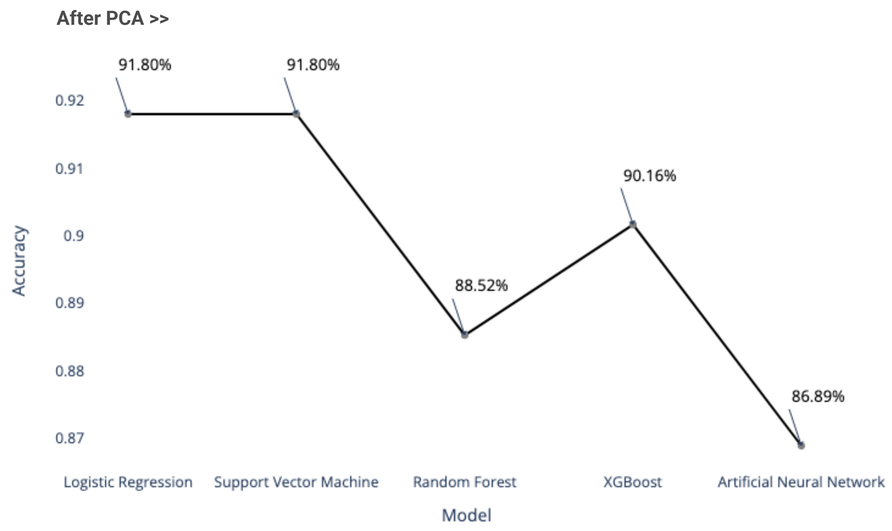




Model Comparison

The line graph visually depicts the performance of the predictive models before and after implementing PCA dimensionality reduction. This comparison offers valuable insights into the effect of dimensionality reduction on the accuracy of the models.





Findings

- Identified features through feature selection that have strong influence in disease outcomes.
- Clustering reveals distinct patient subgroups based on medical profiles. However, further improvements can be made to enhance the clustering results and achieve more precise subgroup differentiation.
- While dimensionality reduction techniques led to improved performance for some models, it is important to note that not all models benefitted equally from PCA.

Further Information ✉ mm22pla@leeds.ac.uk

References

- Hassan, D., Hussein, H.I. & Hassan, M.M. 2023. *Heart disease prediction based on pre-trained deep neural networks combined with principal component analysis*. Biomedical Signal Processing and Control, vol. 79.
- Géron, A. 2019. *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow. O'Reilly Media. Chapter 8: Dimensionality Reduction.* O'Reilly Media, Inc.
- Senan, E.M., Abunadi, I., Jadhav, M.E. and Fati, S.M., 2021. *Score and Correlation Coefficient-Based Feature Selection for Predicting Heart Failure Diagnosis by Using Machine Learning Algorithms*. Computational and Mathematical Methods in Medicine, 2021.