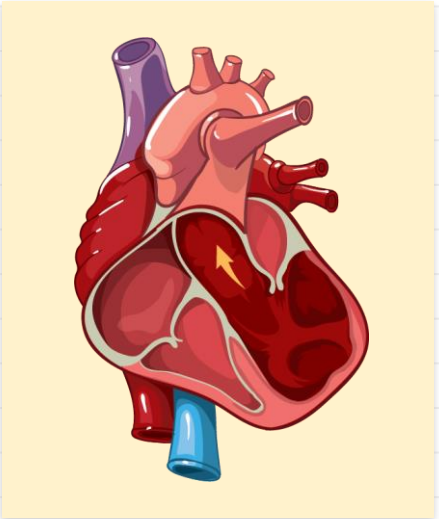
Promoting Cardiovascular Health for All



Sprint 2

Yaire Souffront

01

Overview

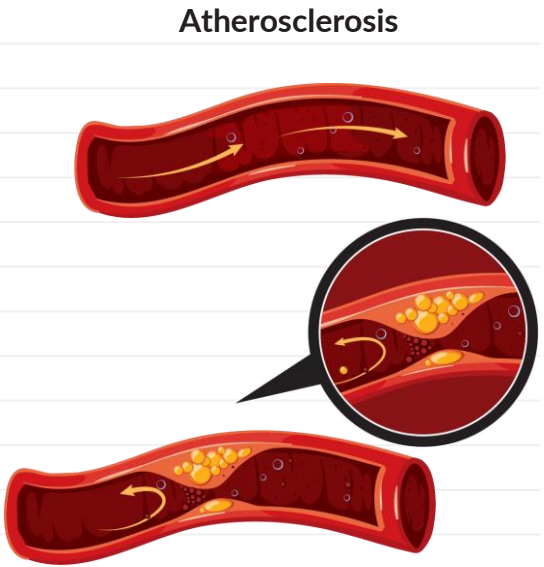
What is Cardiovascular Disease (CVD)?

Causes

Can Vary
Plaque Buildup in Arteries
Infections, Aging

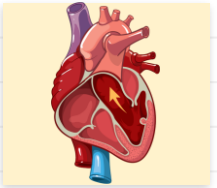
Risk Factors

High Blood Pressure
High Cholesterol
Lack of Physical Activity



Sources: Cleveland Clinic , WHO

NHANES Datasets

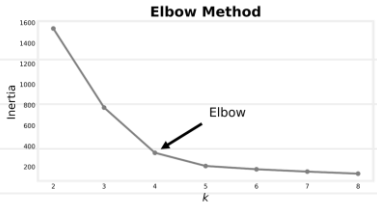


Demographic	Holds demographic data of participants, such as US citizenship, household income, gender, and age
Questionnaire	Contains basic level information of participant's medical history, physical activity, family-level information, and dietary behavior
Labs	includes cholesterol levels, blood glucose, and other metabolic indicators
Diet	Provides detailed information on participants' dietary intake
Examination	Contains physical examination of participants such as bmi, height, waist circumference, and blood pressure readings.

Here's the Data

02

Workflow



Optimize Clustering:

Using Elbow Method

Logistic Reg

Cluster Labels = Pseudo-Target Variable

Validation of Clusters:

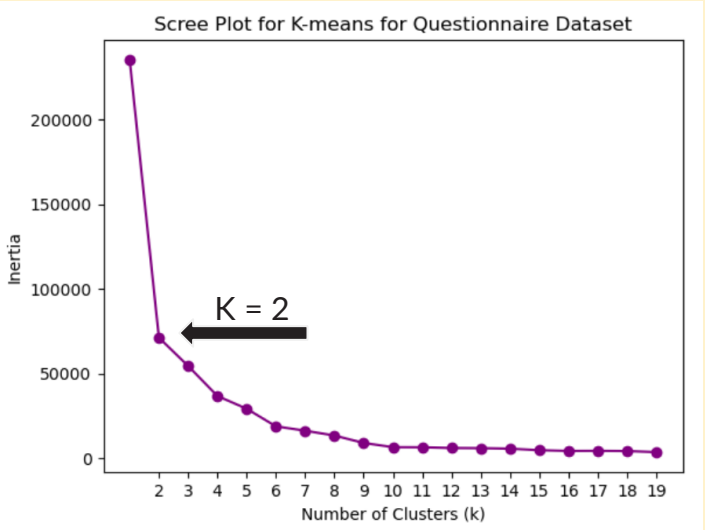
Using Silhouette Analysis

Optimize Log Reg

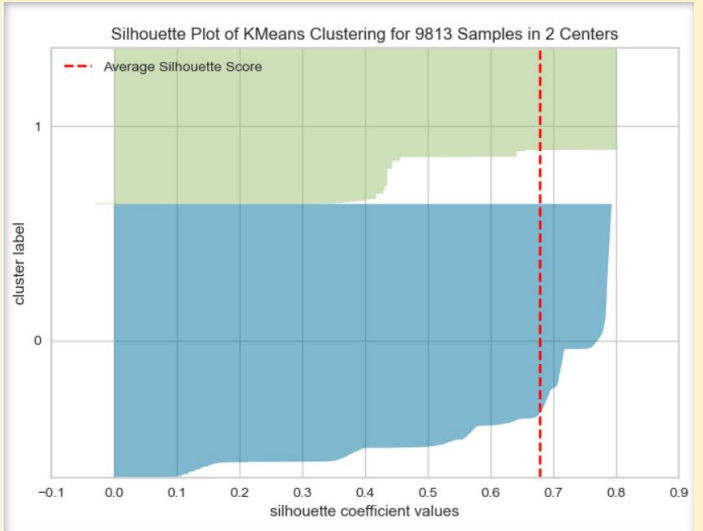
Regularization

Finding an Ideal ‘K’ Value

Elbow Method

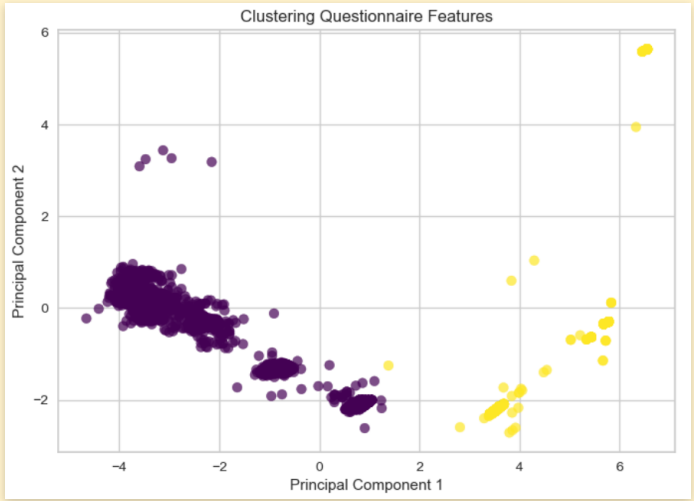


Silhouette Score



Cluster Labels for Log Reg

Questionnaire Clustering

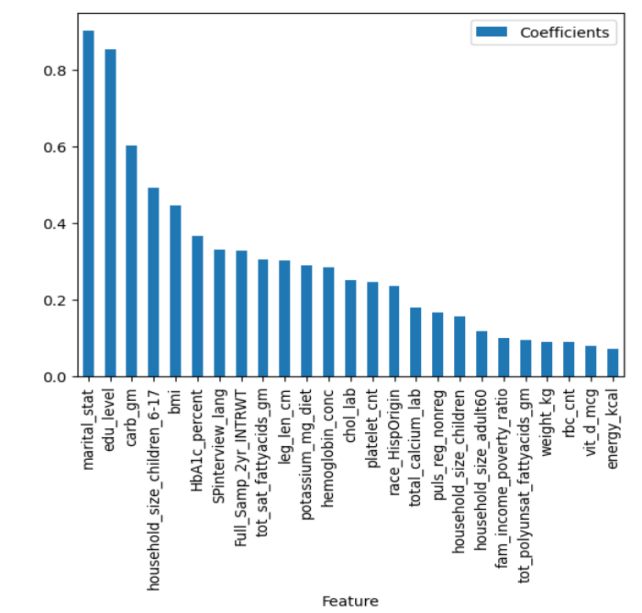


Cluster Labels

	0	1
high_bp	-0.747	1.319
closerel_heart attack	-0.631	1.115
insured	0.124	-0.219

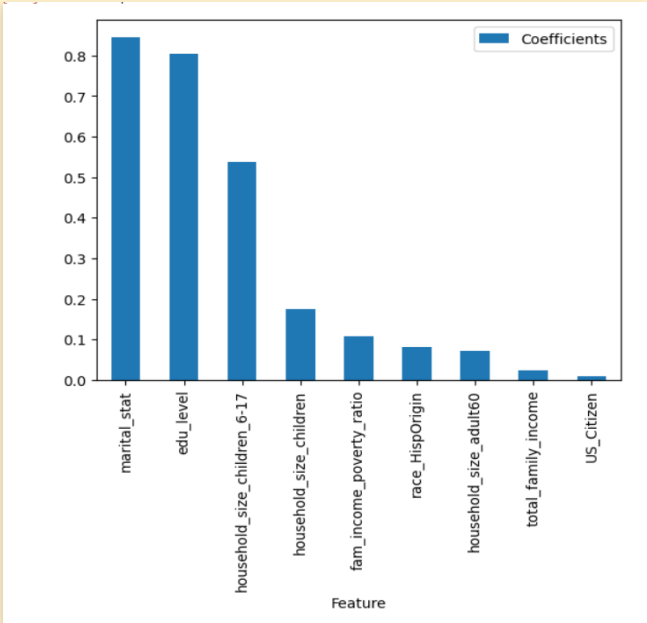
Feature Selection & Log Reg Optimization

Before



Baseline Accuracy: 98.13%

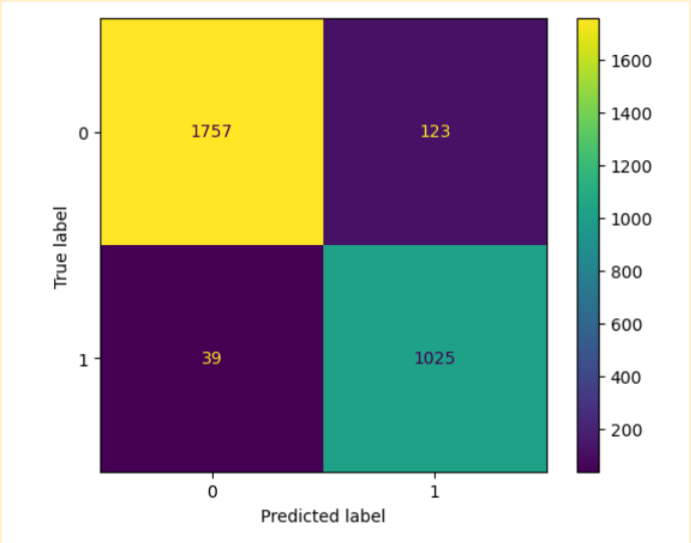
After



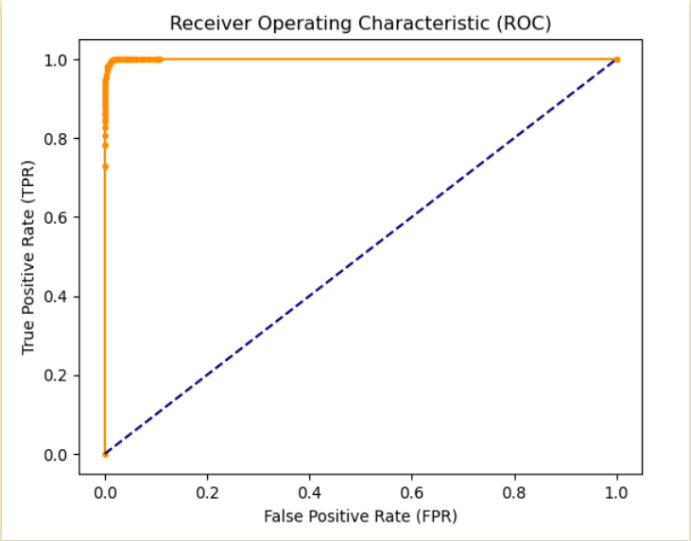
Accuracy: 98.17%

Model Evaluation

Confusion Matrix



ROC Curve



03

Next Steps

Random Forest

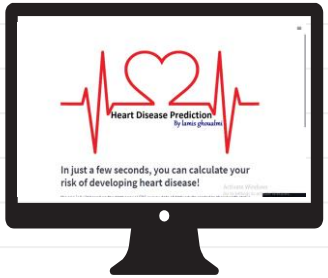
Regularization

Neural Networks

Pytorch

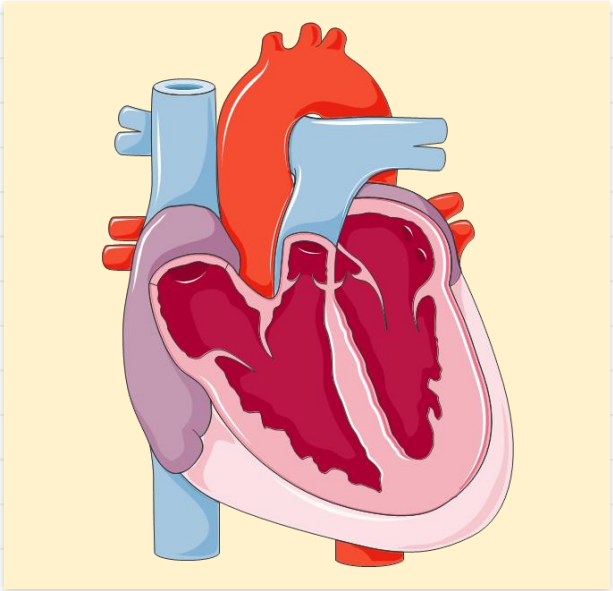
Streamlit

App Creation



Deployment

Thanks!



CREDITS: This presentation template was created by **Slidesgo**, and it includes icons by **Flaticon**, infographics & images by **Freepik**

Illustration by Smart-Servier Medical Art