

PREDICTING HEART ATTACK RISKS WITH MACHINE LEARNING

BUSINESS CONTEXT

Reduction of revenue from health insurance schemes.

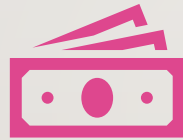
High number of claims related to heart health.

Identify high risk clients.

WHY ACCURATE PREDICTION MATTERS



Precision
Underwriting



Financial
Stability



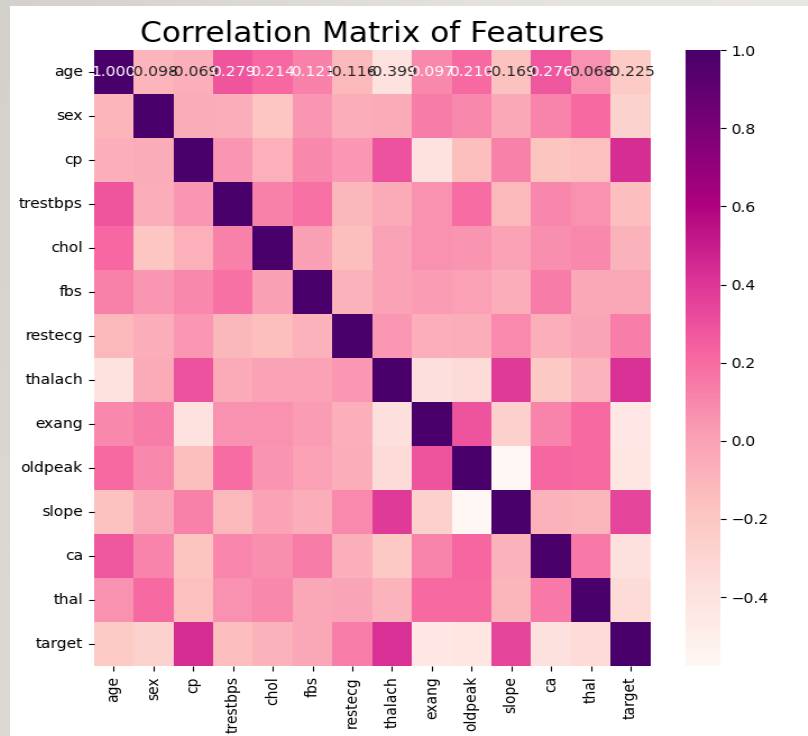
Customer
Satisfaction



Reduced Loss
Ratios

DATA SOURCE : HEART DISEASE DATA SET - UCIRVINE MACHINE LEARNING REPOSITORY

TARGET : PRESENCE OF HEART DISEASE IN THE PATIENT (0= NO PRESENCE, 1 = PRESENCE)



13 Features :

- Age
- Sex
- Chest pain type (4 values)
- Resting blood pressure
- Serum cholesterol in m/dl
- Fasting blood sugar > 120 mg/dl
- Resting electrocardiographic results (values 0,1,2)
- Maximum heart rate achieved
- Exercise induced angina
- Oldpeak = ST depression induced by exercise relative to rest
- The slope of the peak exercise ST segment
- Number of major vessels (0-3) colored by flourosopy
- Thal: 0 = normal; 1 = fixed defect; 2 = reversable defect

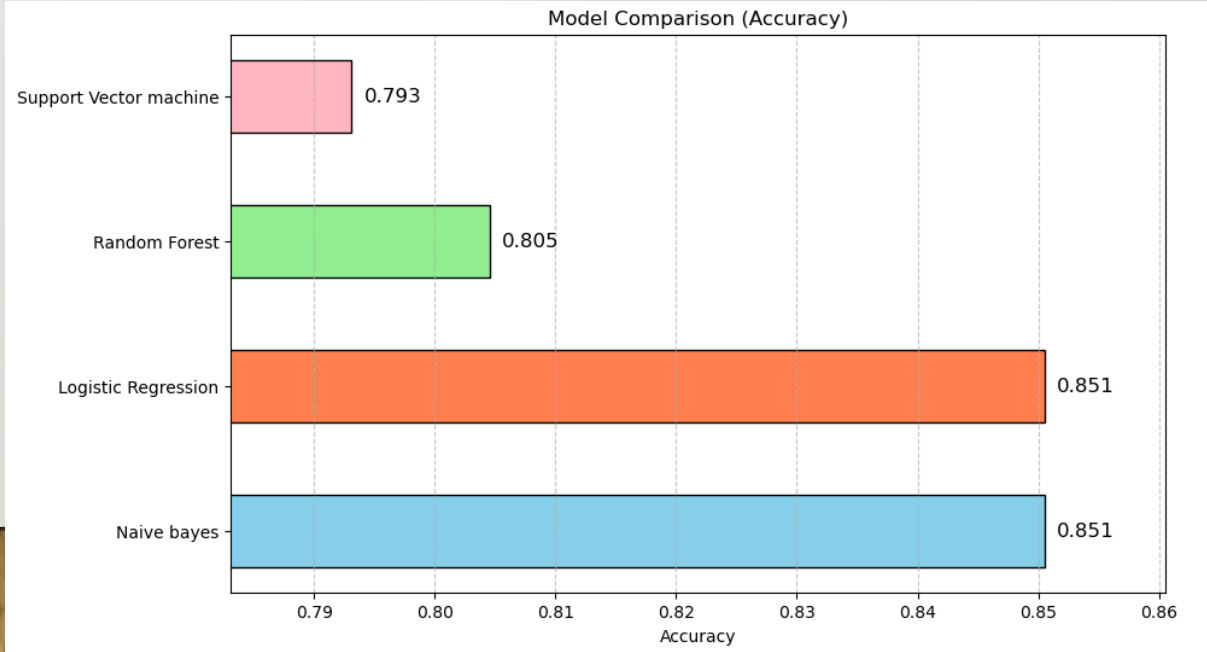
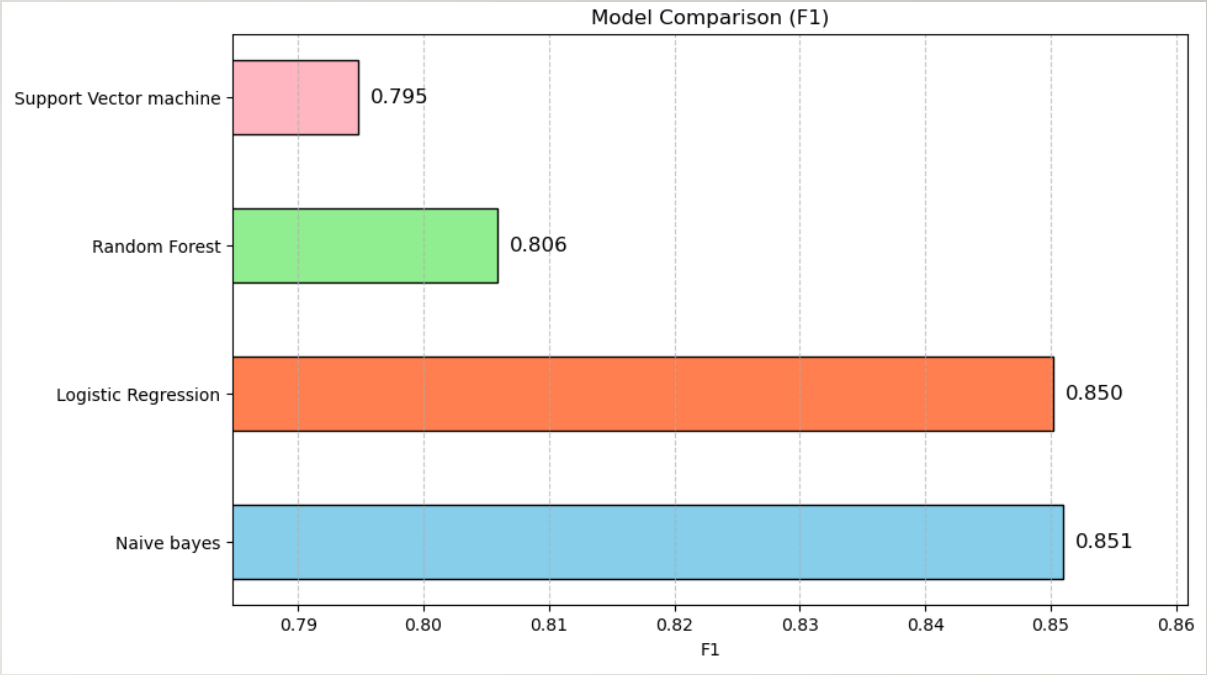
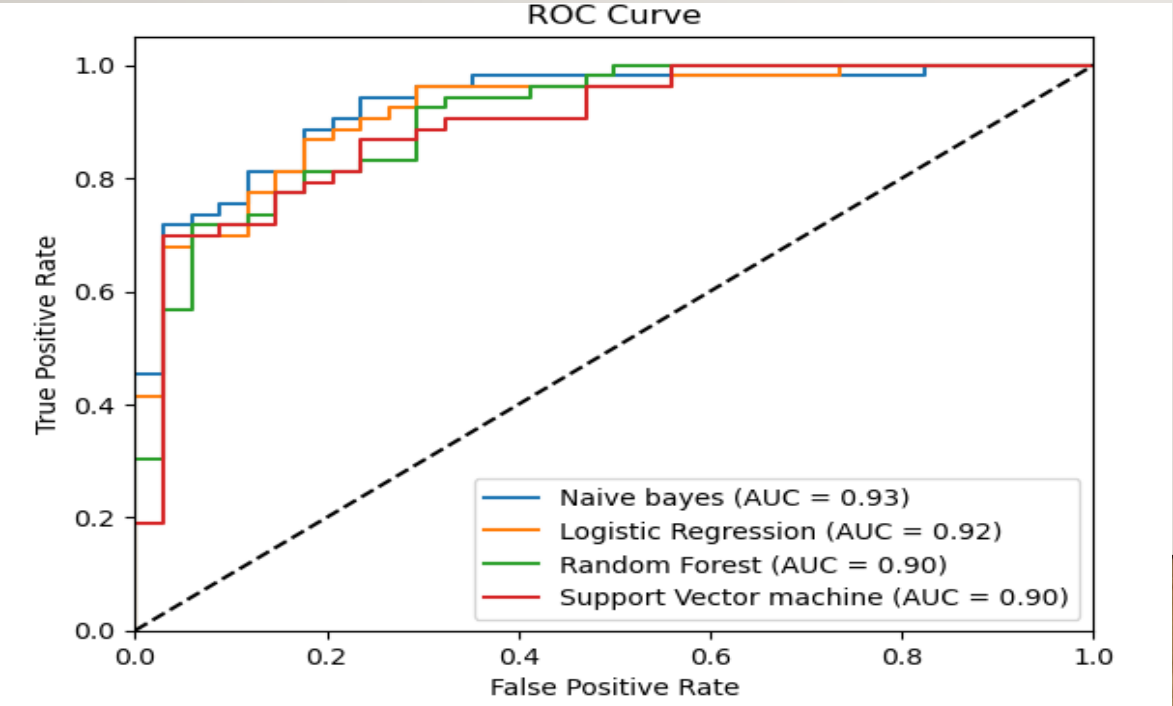
MODEL BUILDING



- Random Forest
- Support Vector Machine
- Logistic Regression
- Naïve Bayes

MODEL EVALUATION

Matrix	Random Forest	SVM	Logistic Regression	Naïve Bayes
Train accuracy	0.905	0.860	0.850	0.845
Test accuracy	0.805	0.793	0.851	0.851
ROC-AUC train	0.981	0.944	0.914	0.908
ROC AUC test	0.905	0.895	0.918	0.929
Mean cross validation score-Train	0.835	0.810	0.815	0.830
Mean cross validation score-Test	0.918	0.883	0.884	0.884
F1 for 0	0.760	0.750	0.810	0.810
F1 for 1	0.830	0.820	0.880	0.880





Q & A



THANK YOU
