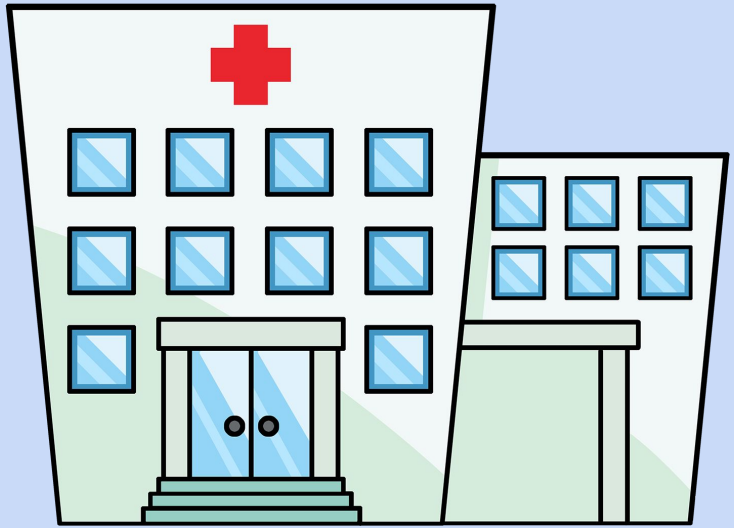
HOSPITALIZED MACHINE LEARNING

PREDICTING HEART DISEASE BASED ON PATIENT DATA



Elias Olcina

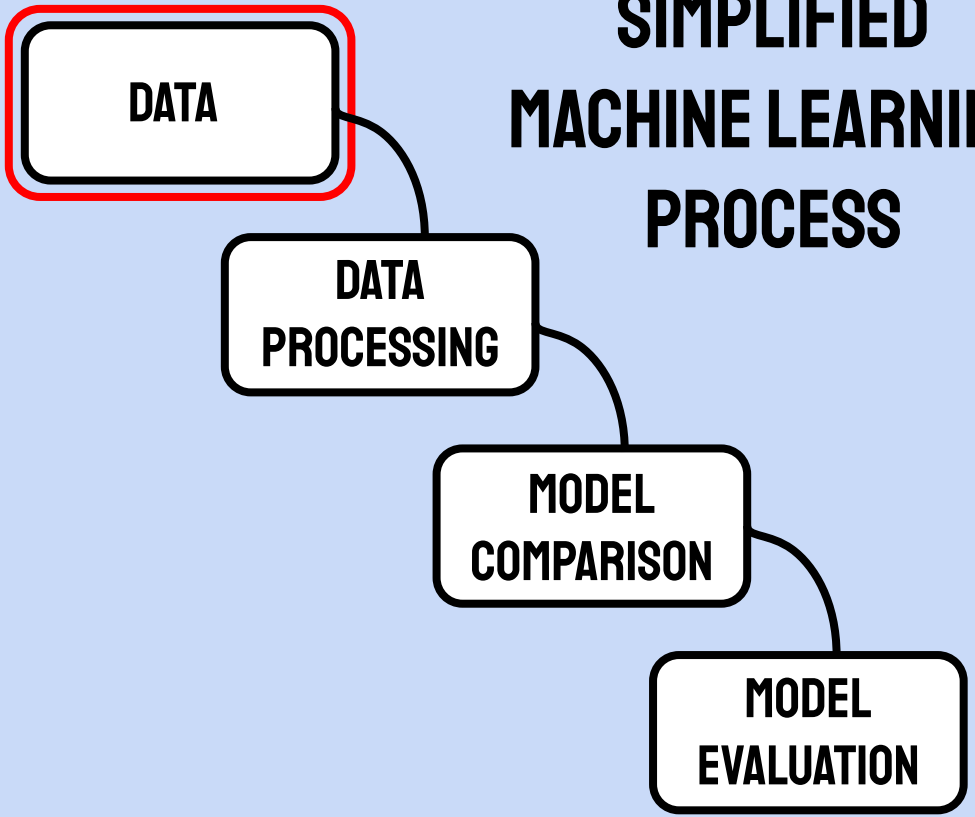
DATA SET

- 918 Patients
- 11 Features
- 1 Heart Disease

- About a 50/50 split of HD

- Train/Test-Split (80/20)

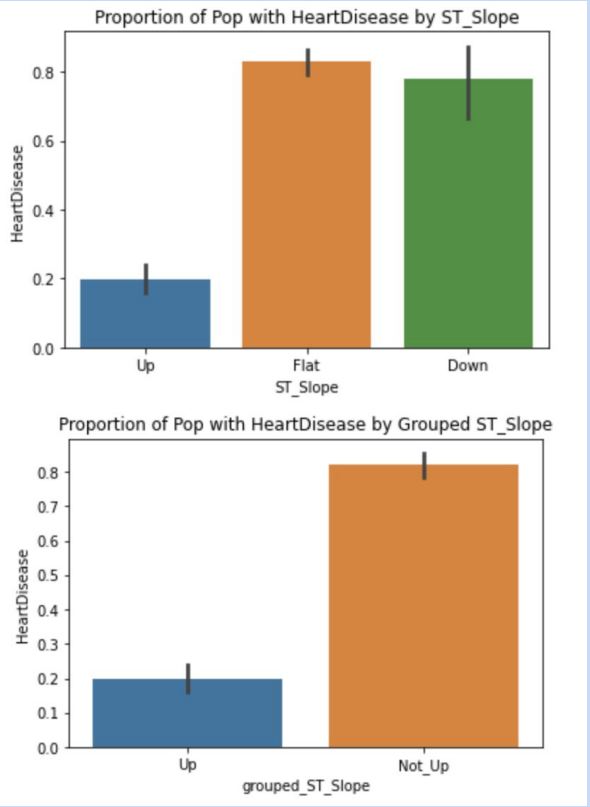
SIMPLIFIED MACHINE LEARNING PROCESS

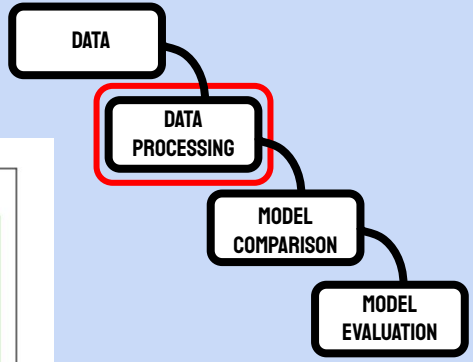


PROCESSING DATA

Model Variables:

- Sex
- Cholesterol
- Fasting Blood Sugar
- Max Heart Rate
- Exercise Angina
- Oldpeak
- Chest Pain Type - Grouped
- ST Slope - Grouped

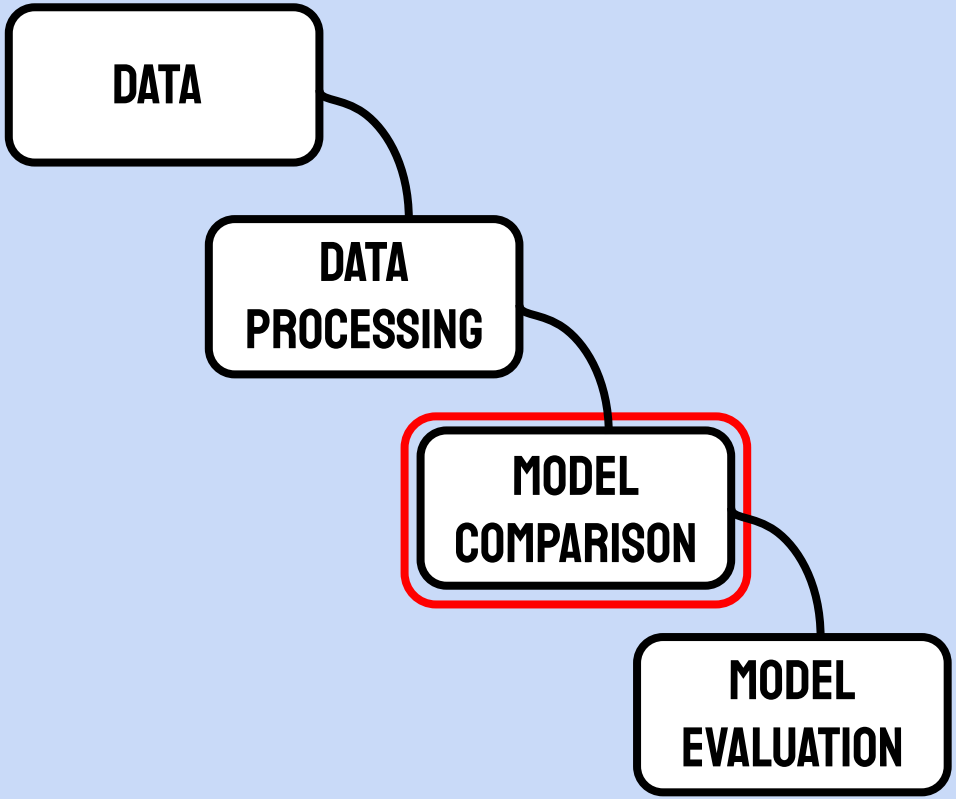




MODEL COMPARISON

7 Different Models:
RandomForestClassifier()
SVC()
LogisticRegression()
DecisionTreeClassifier()
KNeighborsClassifier()
LinearDiscriminantAnalysis()
GaussianNB()

How do we compare?
- Scoring

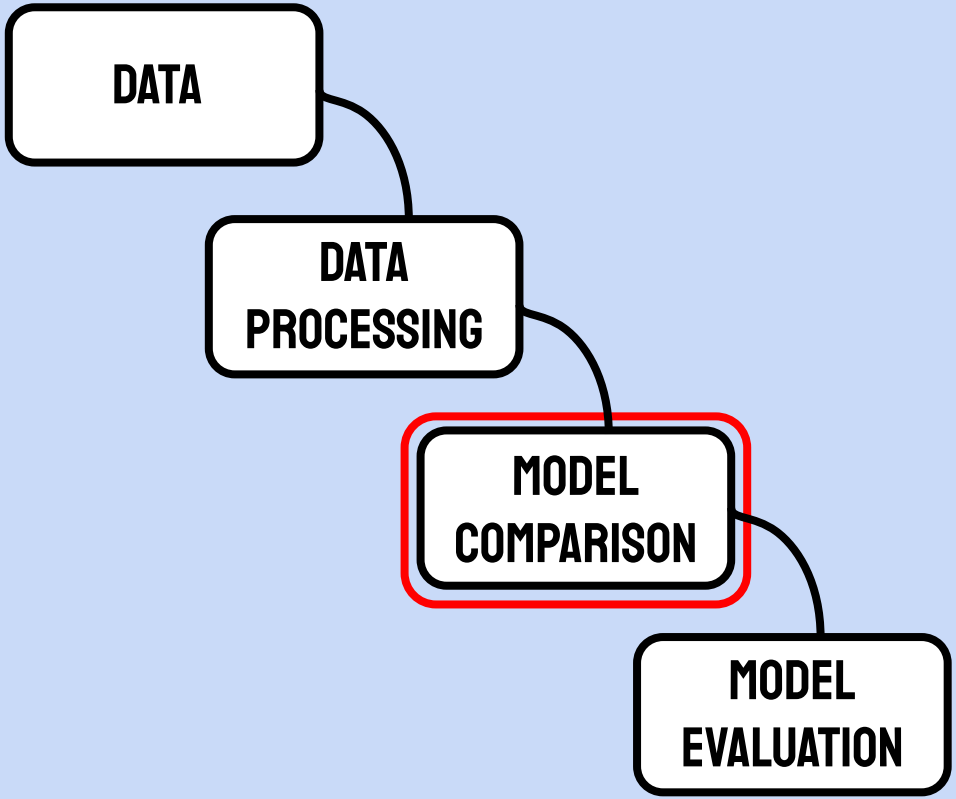


SCORING

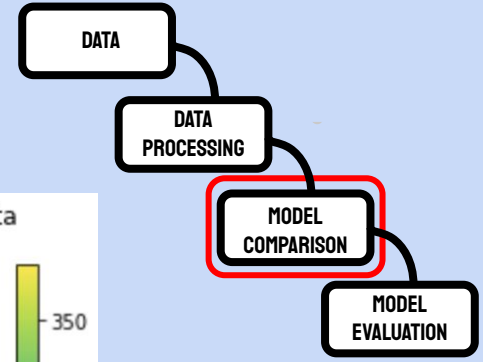
Recall:
How many out of the patients with heart disease did we correctly classify as sick?
Precision:
How many out of the patients we classified as sick, did actually have a heart disease?

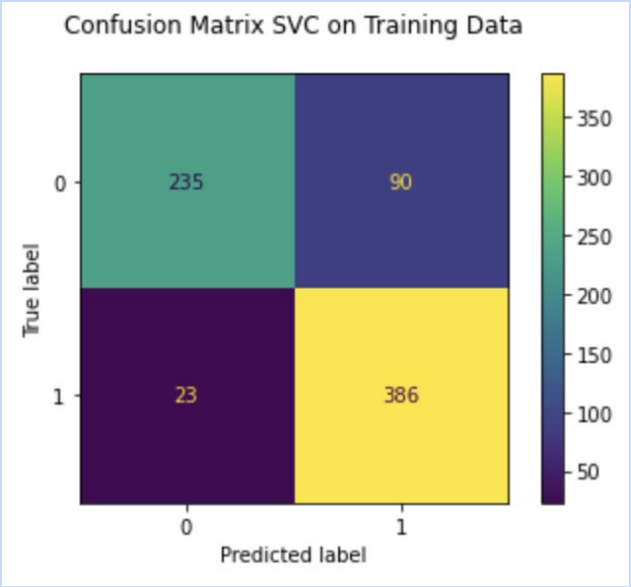
Assumption:
Telling a patient with a HD that they are healthy is worse than telling a healthy patient they have a HD

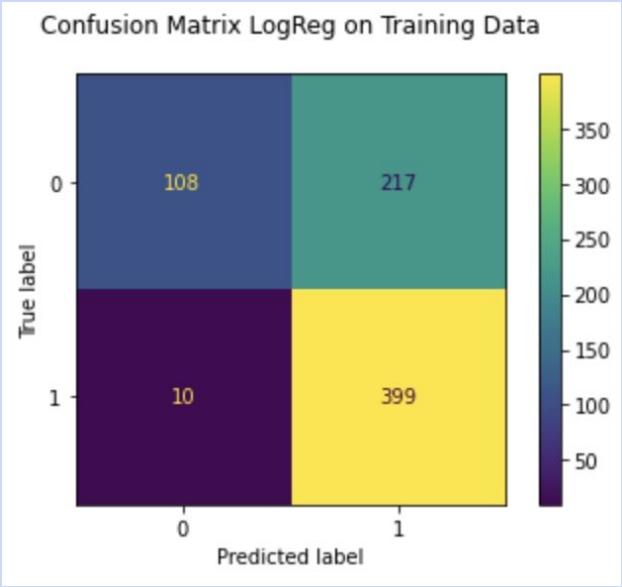
Based on this logic, Best Models:
Support Vector Classifier
Logistic Regression



SVC VERSUS LOGREG







Recall: 94.3 % Precision: 81.1%

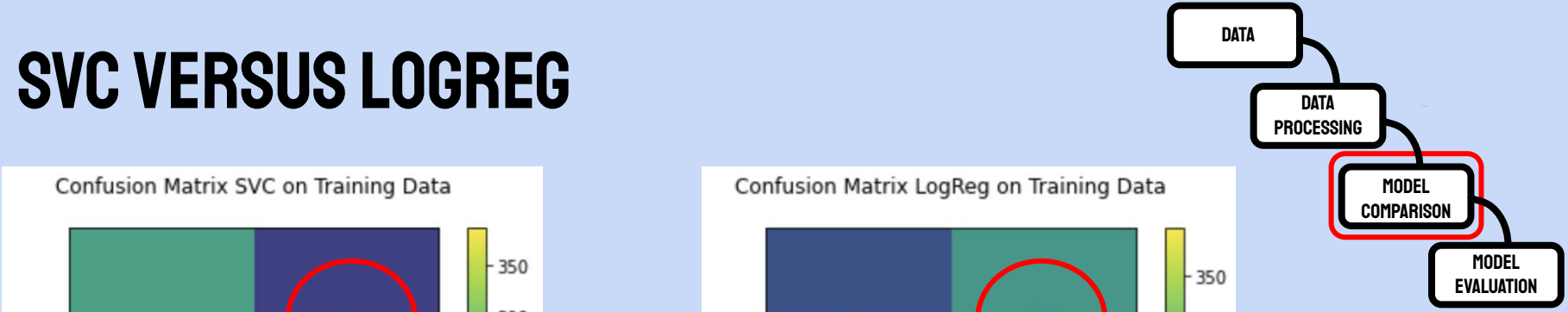
23: Have Heart Disease, Classified as Healthy
90: Are Healthy, Classified with Heart Disease

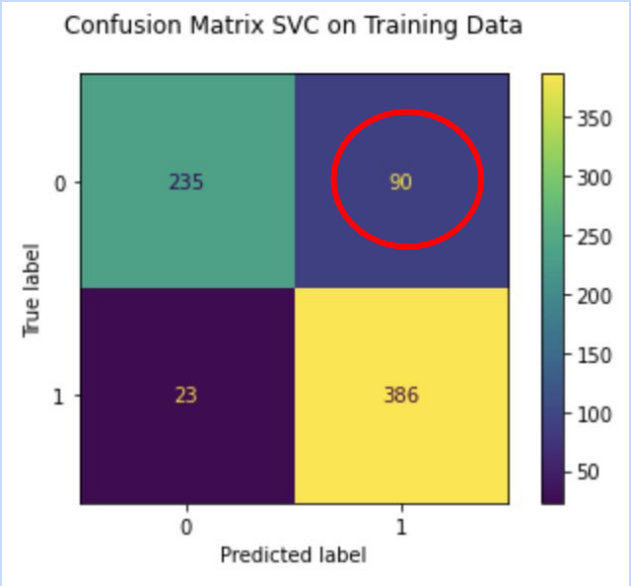
Recall: 97.5 % Precision: 64.8%

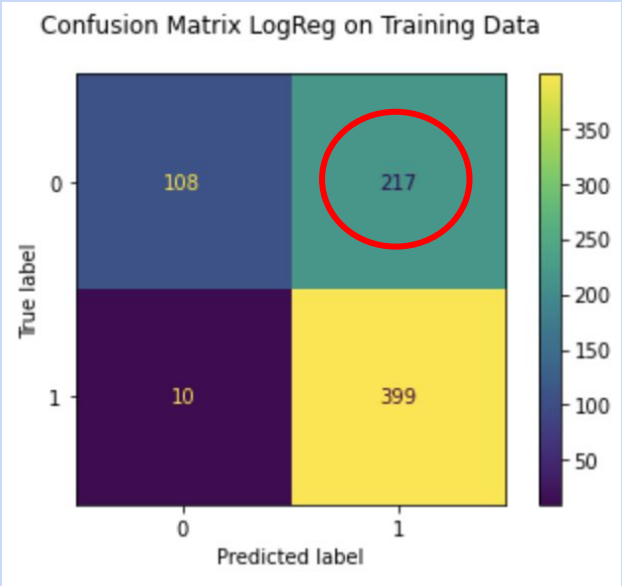
10: Have Heart Disease, Classified as Healthy
217: Are Healthy, Classified with Heart Disease

MONEY

SVC VERSUS LOGREG





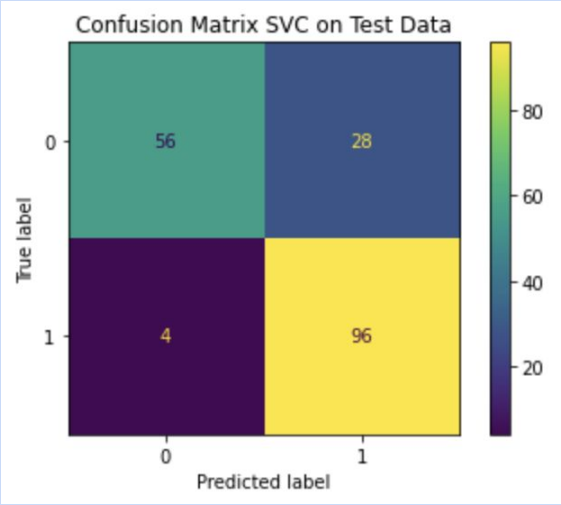


Cost for Treatment on Avg: 18,900$

Unnecessary cost of treatment: 18,900 * (217-90) = 2,400,300$

MODEL EVALUATION

Model performance on Test Data:
Recall: 0.96 Precision: 0.77



Total: 184 patients

4: Have Heart Disease but Classified as Healthy

28: Are Healthy, but Classified with Heart Disease

DATA

DATA PROCESSING

MODEL COMPARISON

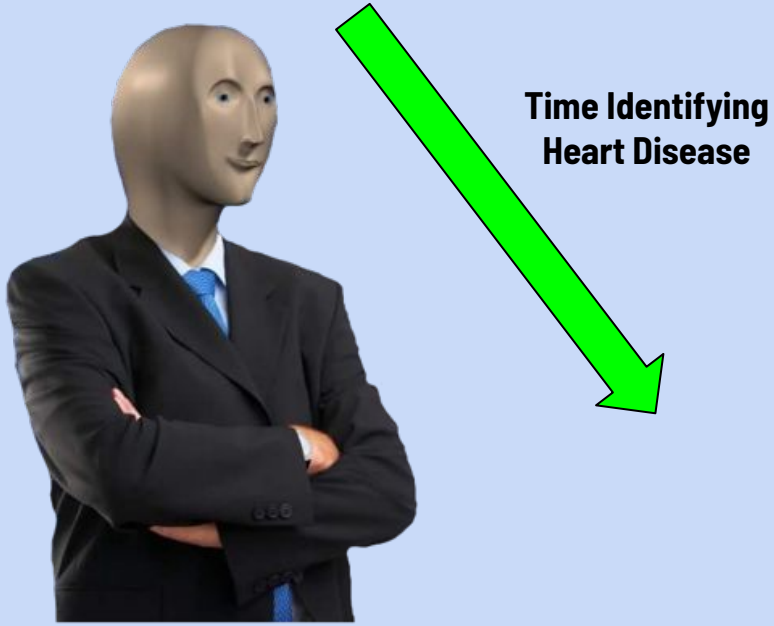
MODEL EVALUATION

PROPOSITION:

Model as a tool to identify patients with heart disease, very quickly

With your help, we can create a new model with a better scoring:

- How much worse is it to classify a person with heart disease as healthy, compared to the opposite?



QUESTIONS?