

PREDICTING CARDIOVASCULAR DISEASE

SULEYMAN QAYUM

OUTLINE



BUSINESS
UNDERSTANDING



DATA
UNDERSTANDING



DATA
PREPARATION



RESULTS



CONCLUSION



LIMITATIONS &
NEXT STEPS

BUSINESS UNDERSTANDING

- ❖ The *U.S. Department of Health and Human Services (HHS)* have allocated funding for the development of a machine learning model that can accurately predict cardiovascular disease
 - *Final goal is to integrate it into the major EHR systems as a preliminary screening tool for cardiovascular disease.*
 - *They also request that the most important features be identified*

DATA UNDERSTANDING

- ❖ Dataset contains almost 70,000 samples pertaining to patient medical information
- ❖ Dataset was chosen because:
 - ❖ *contains features that can be easily and quickly obtained in a medical setting*
 - ❖ *large number of samples*
 - ❖ *well-balanced*
 - ❖ *no missing values*

DATA PREPARATION

❖ Data Cleaning

- *Over 6,000 samples removed due to presence of outliers or absurd values*
- *Highlighted some issues with the data*

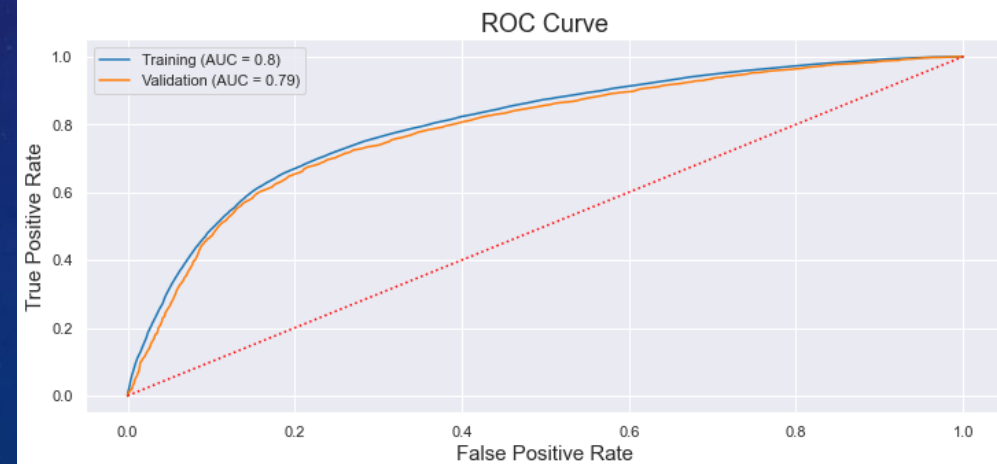
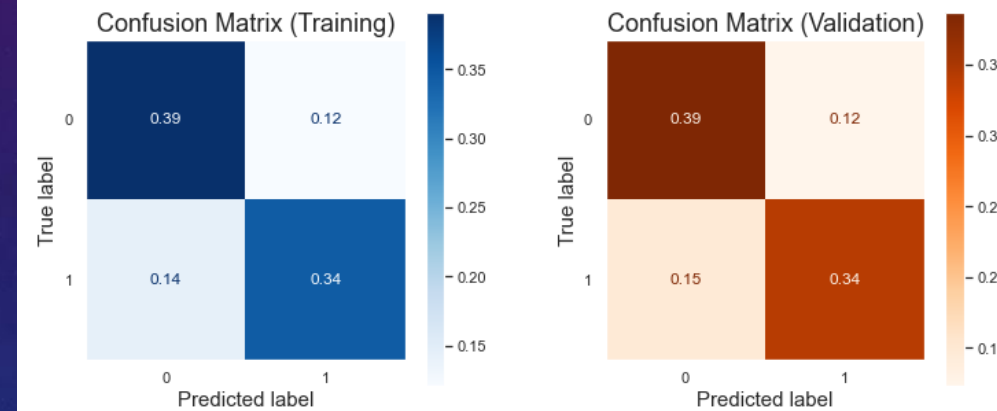
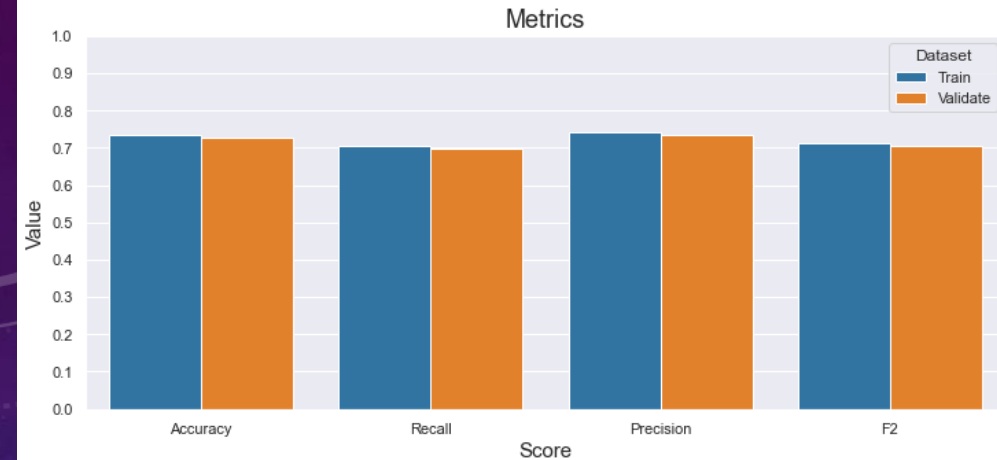
❖ Feature Engineering

- *BMI*
- *BMI Category*
- *BP Category*

RESULTS

❖ The information shown summarizes the final model:

- *Too many false negatives*
- *Accuracy needs improvement*
- *Generalizes very well*

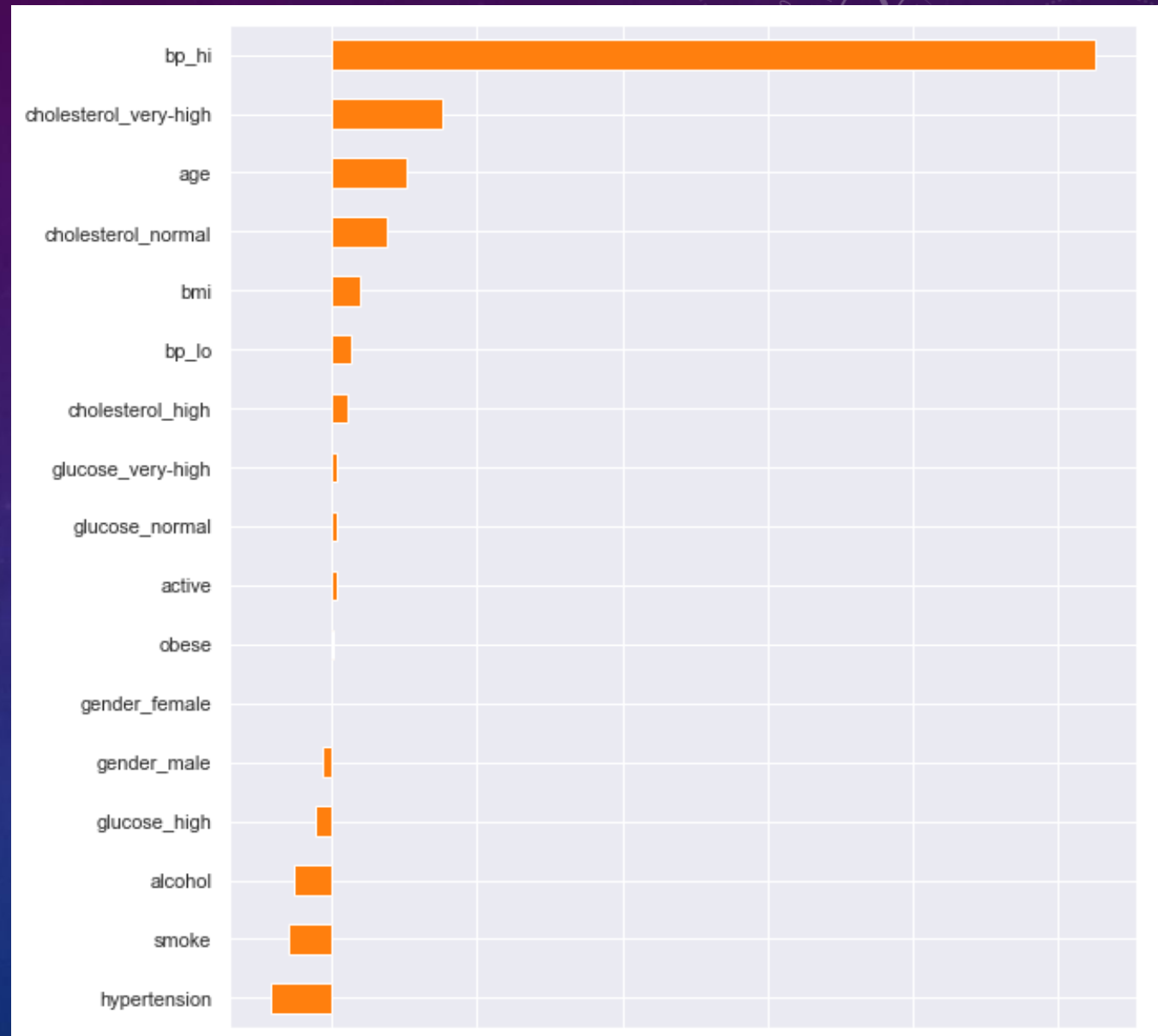


❖ Features that *matter*:

- *Systolic blood pressure (far the most important)*
- *Cholesterol level*
- *Age*
- *Body Mass Index (BMI)*
- *Diastolic blood pressure*

❖ Features that *don't matter*:

- *Whether or not patient smokes*
- *Whether or not patient drinks alcohol*
- *Gender*



LIMITATIONS & NEXT STEPS

- ❖ Main problems are related to data
 - ❖ *Some features too vague or non-informative*
 - *i.e. whether or not patient smokes, drinks*
 - ❖ *Difficult to evaluate quality of features*
 - *Requires domain-specific knowledge*
- ❖ More data needs to be gathered
- ❖ Exploit EHR Systems for online training
 - *Model shows promise – would be a great starting point*

THANK YOU

Email: sqayum33@gmail.com

Github: @sqayum

