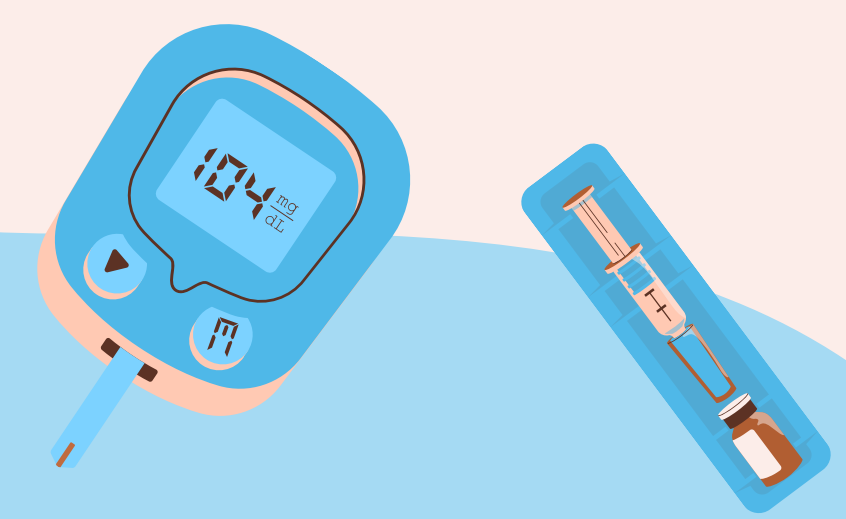
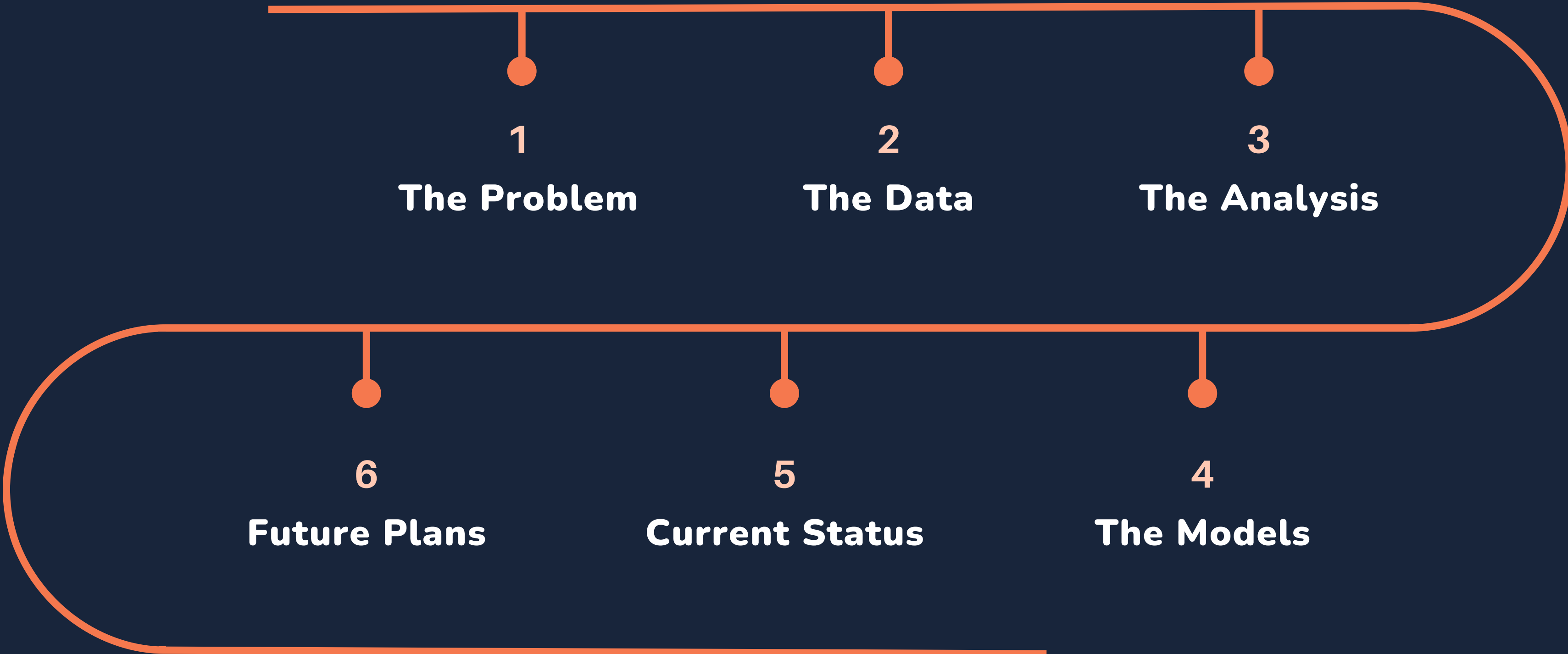


## An illustration of various medical supplies for insulin administration. In the center is an orange insulin bottle with a white cap and a label that reads "INSULIN Injection 10ml". To the left of the bottle is a white insulin pump with a blue syringe attached. To the right of the bottle is a blue syringe with a scale from 0 to 100 units. Further to the right is an orange container holding several white insulin strips. The background is a solid blue color.



# TODAY'S AGENDA



# THE PROBLEM

- There are 5.7 Million Canadians living with Diabetes Mellitus in 2022<sup>(1)</sup>
- Diabetic patients have **complex medical needs**, especially in the ER<sup>(2)</sup>
- The prevalence of diabetes is **2.1 times** higher among adults living in the lowest-income group<sup>(3)</sup>



# THE DATA

## What does it look like?

After filtering for only patients with diabetes, there were  
110K rows and 487 columns

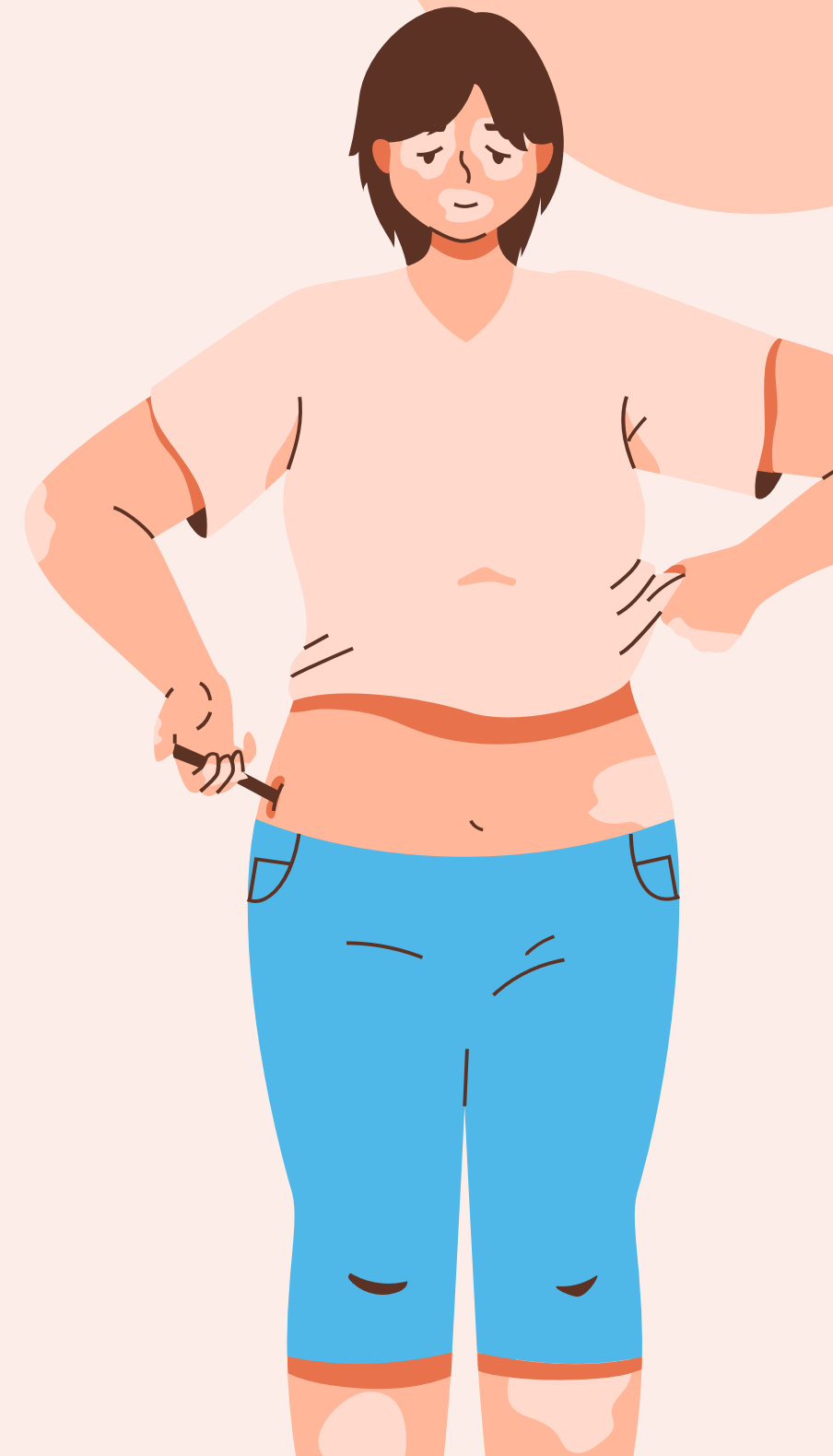
## What changes were made?

Changing all numeric columns to categorical  
columns.

Dummying the columns  
Feature Engineering/Elimination

## Motivation

To aid in the pre-processing of data for modeling

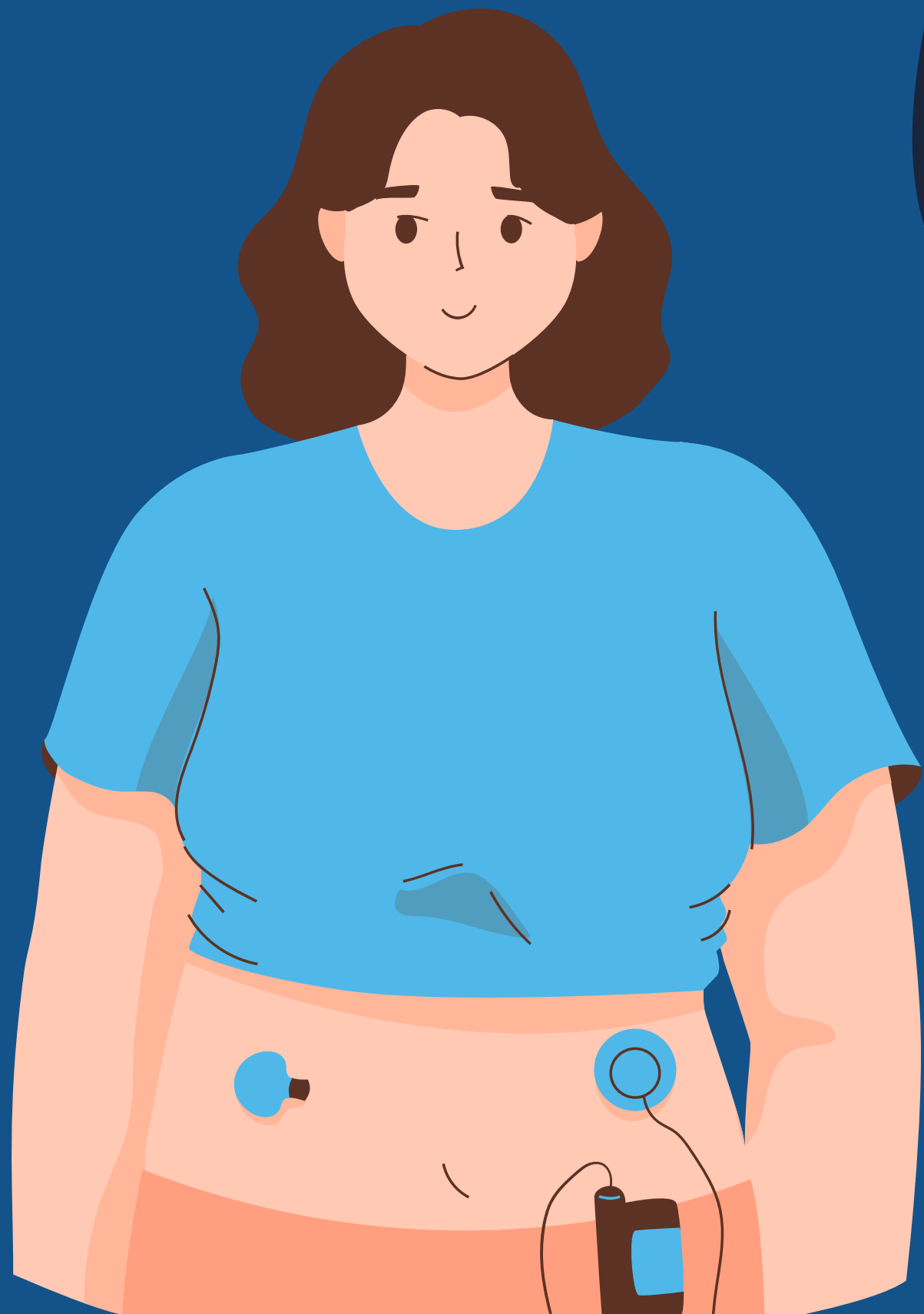


# THE ANALYSIS

- The majority of patients are over 40
- About 76% of the patients have hypertension (high blood pressure)
- About 53% of the patients have hyperlipidemia (high cholesterol)



# MODEL METRICS



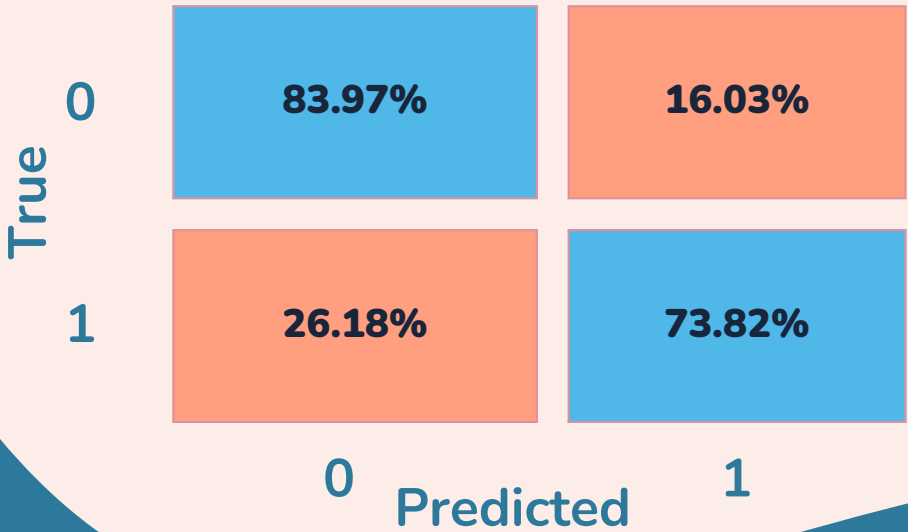
- Train Accuracy
- Test Accuracy
- 5 Fold Cross Validation (CV)
- Mean CV Accuracy
- Classification Report
  - Precision, Recall, F-1 Score
- Confusion Matrix
- ROC - AUC Curve

# Logistic Regression

## Base

Train Accuracy: **0.792**  
Test Accuracy: **0.793**  
Mean CV Accuracy: **0.789**

CLASSIFICATION REPORT	PRECISION	RECALL	F-1 SCORE
0	0.79	0.84	0.81
1	0.80	0.74	0.77

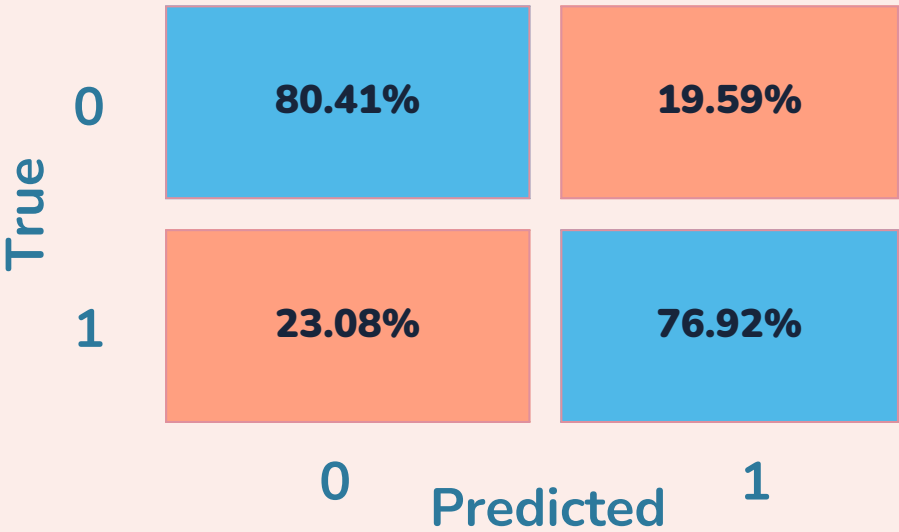


ROC AUC: 0.87

## Optimized

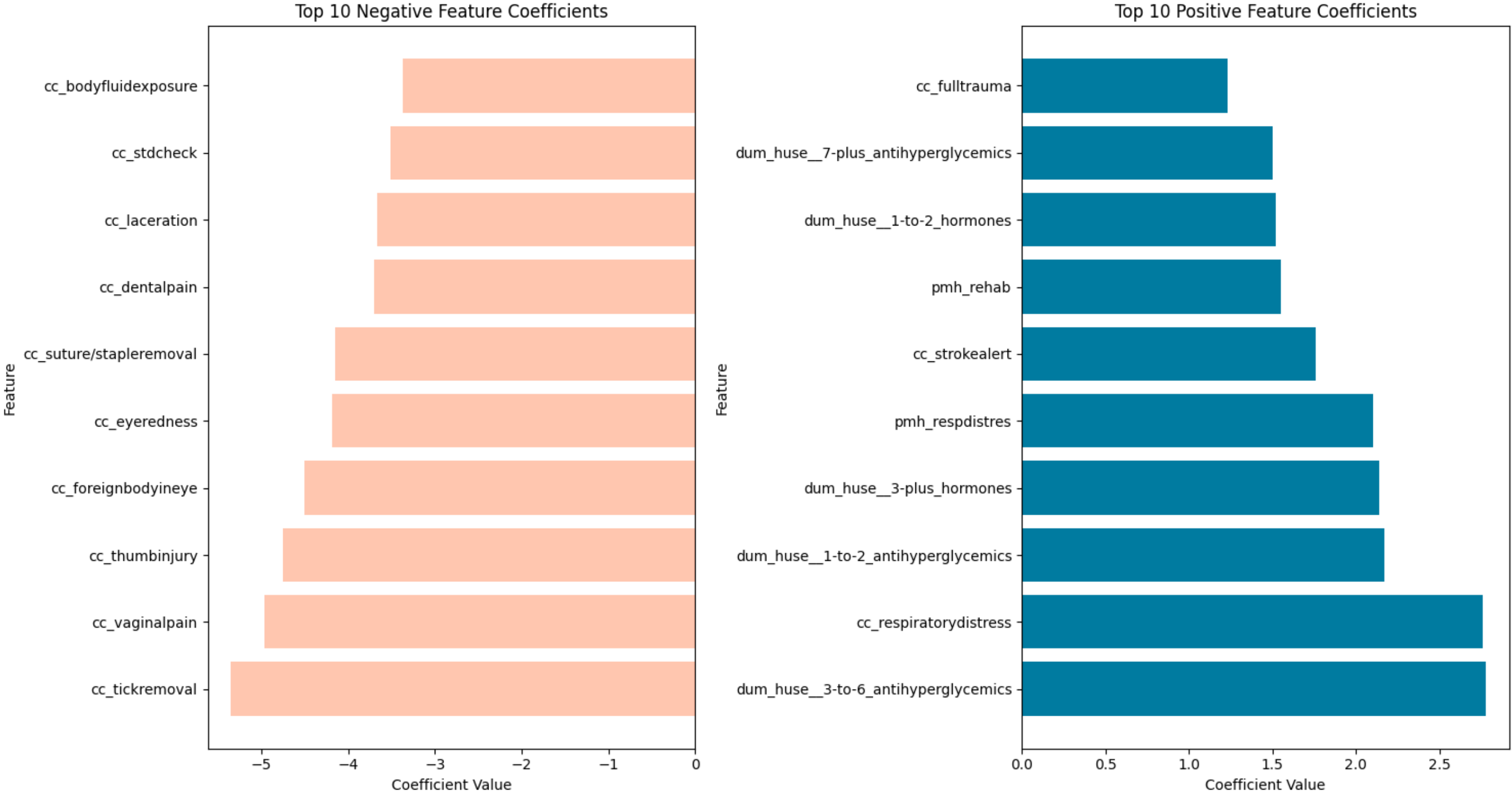
Train Accuracy: **0.787**  
Test Accuracy: **0.788**  
Mean CV Accuracy: **0.789**

PRECISION	RECALL	F-1 SCORE	CLASSIFICATION REPORT
0.80	0.80	0.80	0
0.77	0.77	0.77	1



ROC AUC: 0.87

# Feature Importance





# DECISION TREE

## Base

Train Accuracy: **0.995**  
Test Accuracy: **0.724**  
Mean CV Accuracy: **0.700**

CLASSIFICATION REPORT	PRECISION	RECALL	F-1 SCORE
0	0.74	0.75	0.74
1	0.70	0.69	0.70

True	0	74.99%	25.01%
	1	30.68%	69.32%
		0 Predicted	1

ROC AUC: 0.72

## Optimized

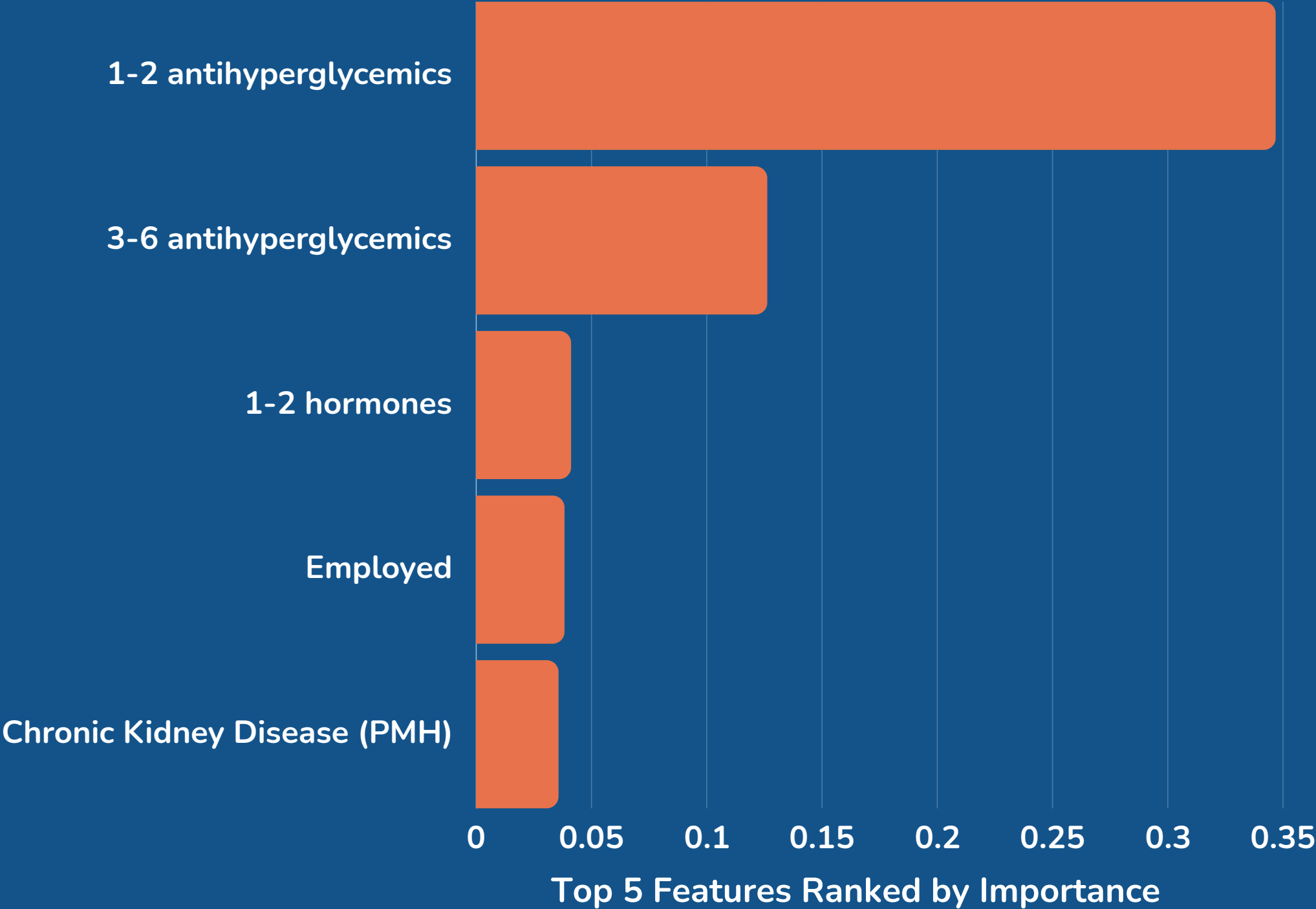
Train Accuracy: **0.8000**  
Test Accuracy: **0.762**  
Mean CV Accuracy: **0.747**

PRECISION	RECALL	F-1 SCORE	CLASSIFICATION REPORT
0.76	0.81	0.79	0
0.76	0.70	0.73	1

True	0	81.42%	18.58%
	1	23.08%	76.92%
		0 Predicted	1

ROC AUC: 0.80

# DECISION TREE



# Random Forest

## Base

Train Accuracy: **0.995**  
Test Accuracy: **0.785**  
Mean CV Accuracy: **0.776**

CLASSIFICATION REPORT	PRECISION	RECALL	F-1 SCORE
0	0.78	0.84	0.81
1	0.79	0.72	0.76

True	0	83.65%	16.35%
	1	27.50%	72.50%
		0 Predicted	1

ROC AUC: 0.86

## Optimized

Train Accuracy: **0.849**  
Test Accuracy: **0.779**  
Mean CV Accuracy: **0.787**

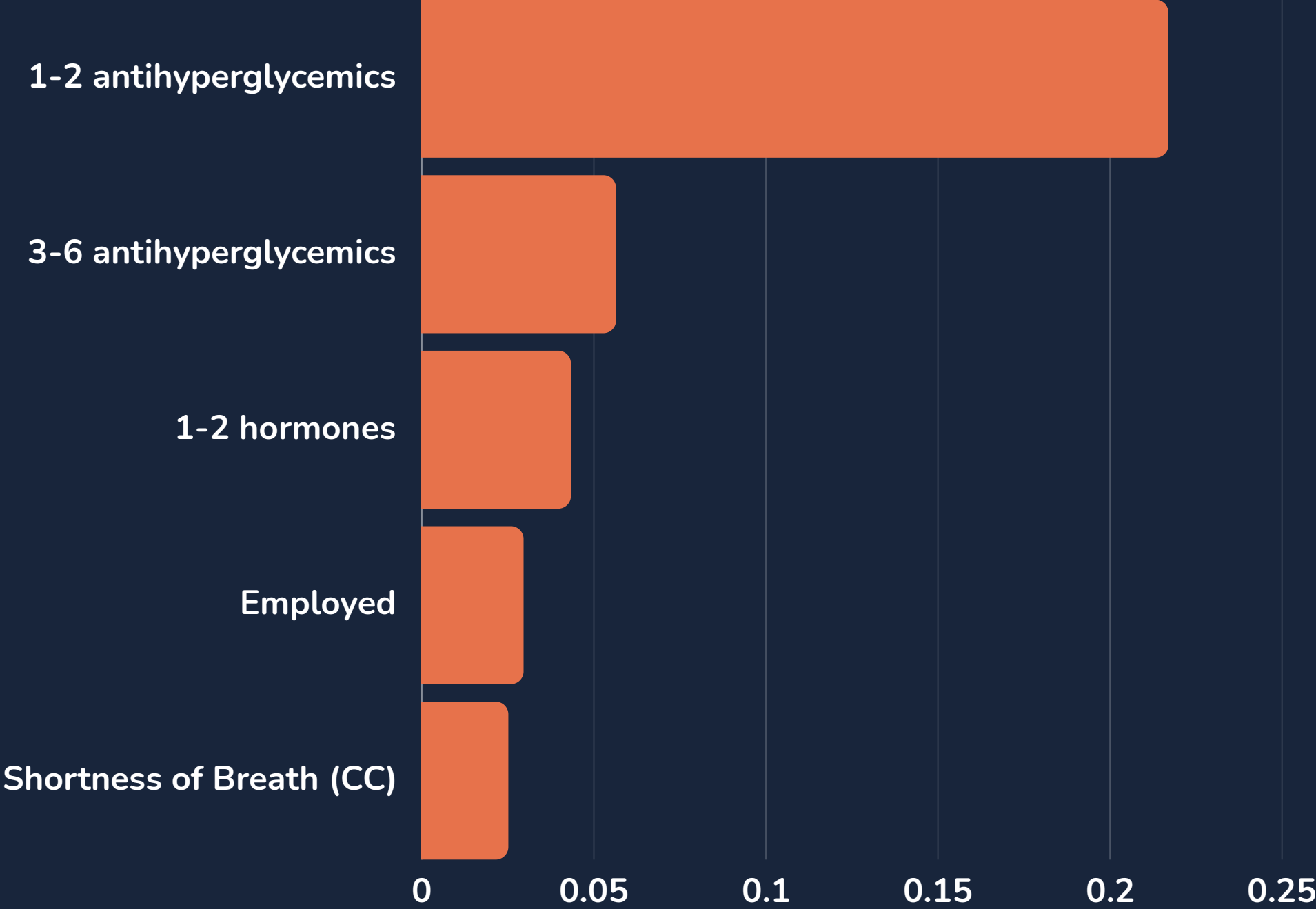
PRECISION	RECALL	F-1 SCORE	CLASSIFICATION REPORT
0.77	0.86	0.81	0
0.81	0.70	0.75	1

True	0	85.53%	14.47%
	1	29.07%	70.93%
		0 Predicted	1

ROC AUC: 0.87

# Random Forest

Top 5 Features Ranked by Importance

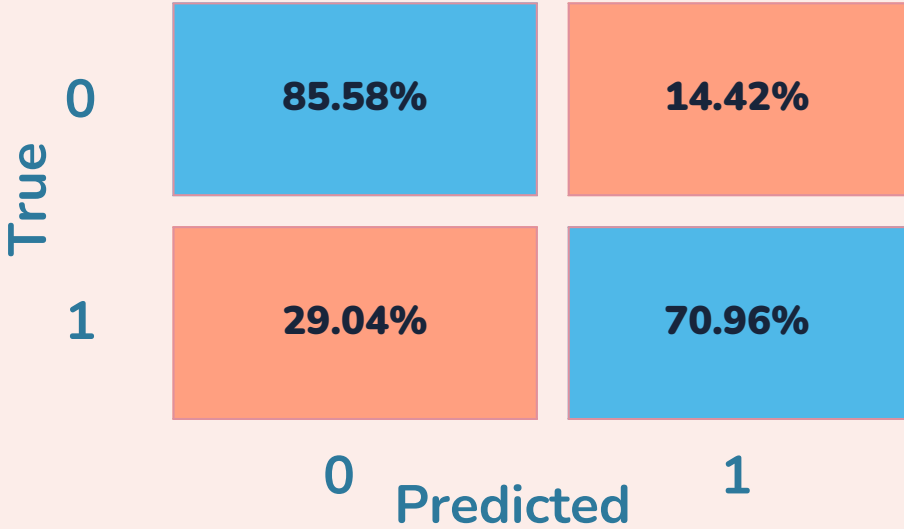


# XGBOOST

## Base

Train Accuracy: **0.812**  
Test Accuracy: **0.788**  
Mean CV Accuracy: **0.789**

CLASSIFICATION REPORT	PRECISION	RECALL	F-1 SCORE
0	0.77	0.86	0.81
1	0.81	0.71	0.76

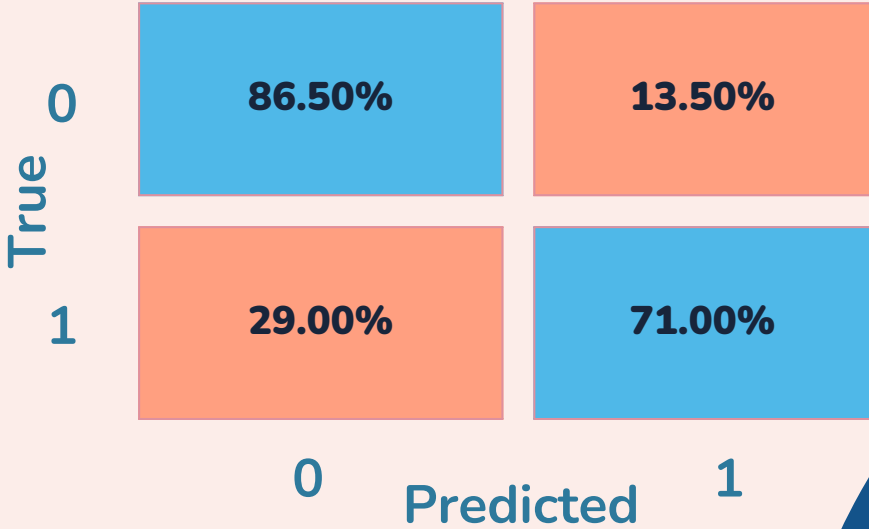


ROC AUC: 0.88

## Optimized

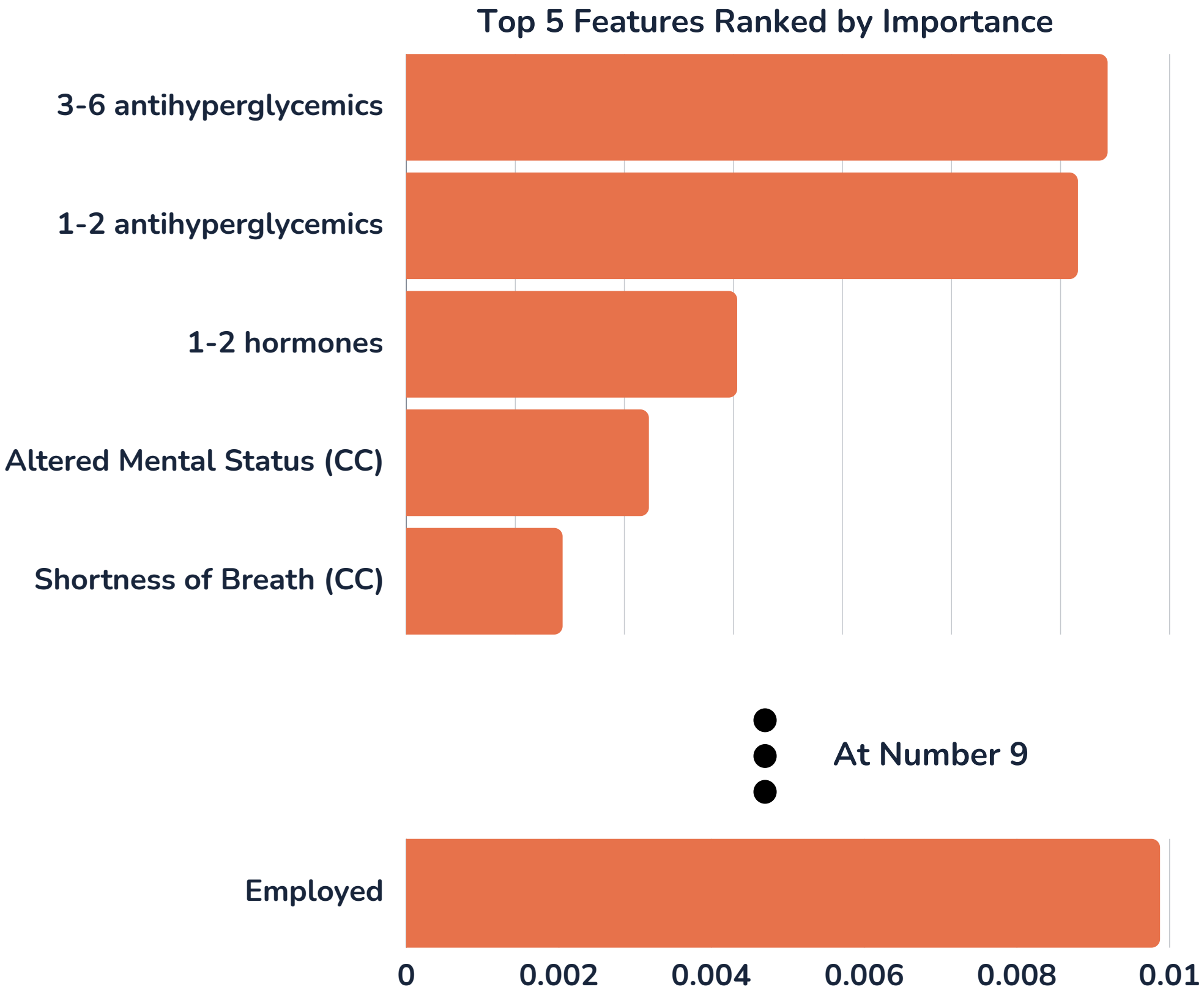
Train Accuracy: **0.817**  
Test Accuracy: **0.795**  
Mean CV Accuracy: **0.789**

PRECISION	RECALL	F-1 SCORE	CLASSIFICATION REPORT
0.78	0.87	0.82	0
0.81	0.71	0.76	1



ROC AUC: 0.88

# XGBOOST



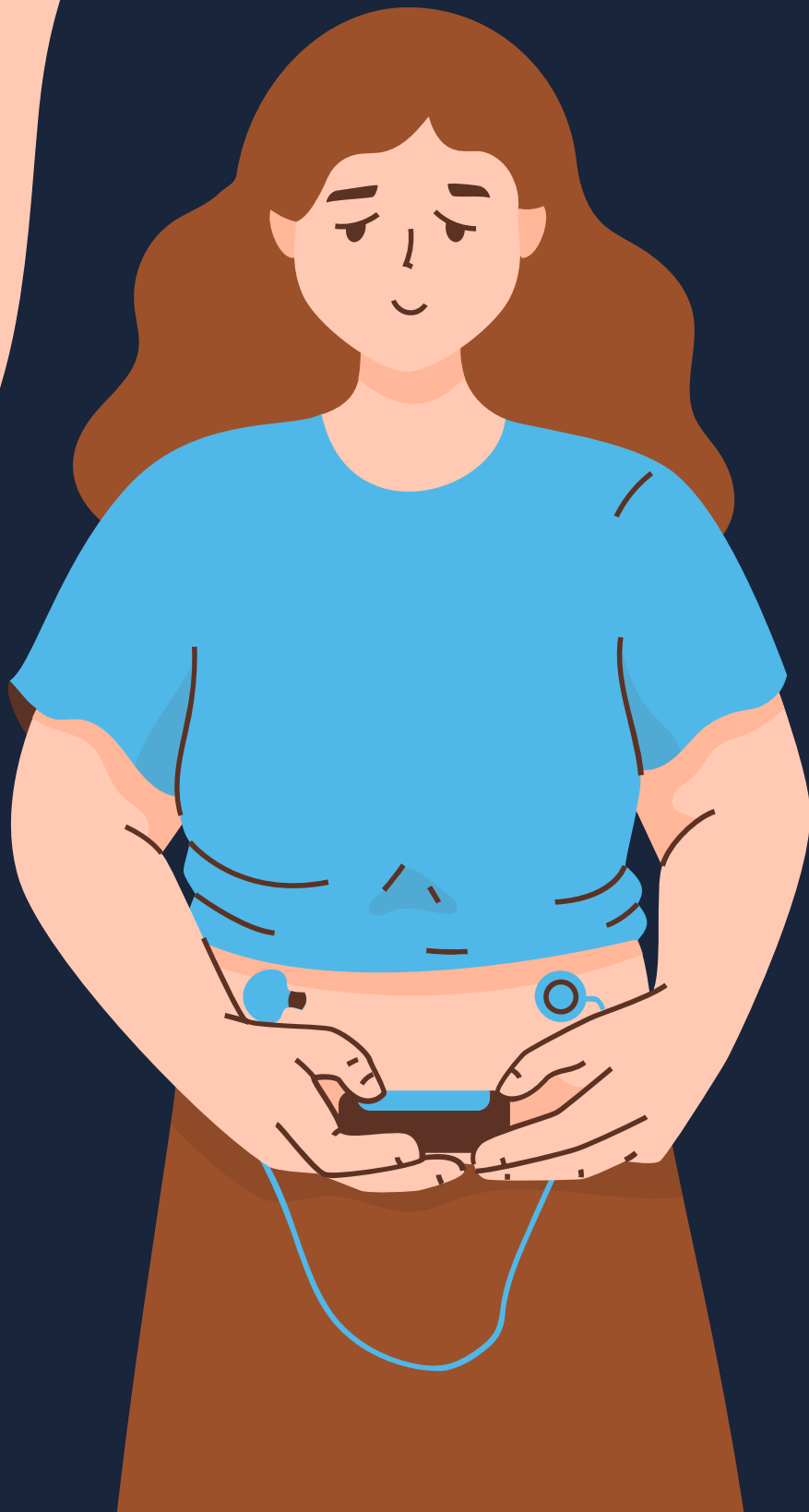
# CURRENT STATUS

- Increasing Explainability



# ML IN PRODUCTION

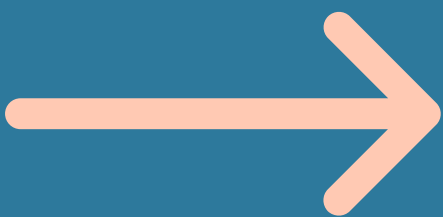
- Use top 2 models to push to production
- Utilize **Streamlit** to create a **web app** to **interact** with the **models**





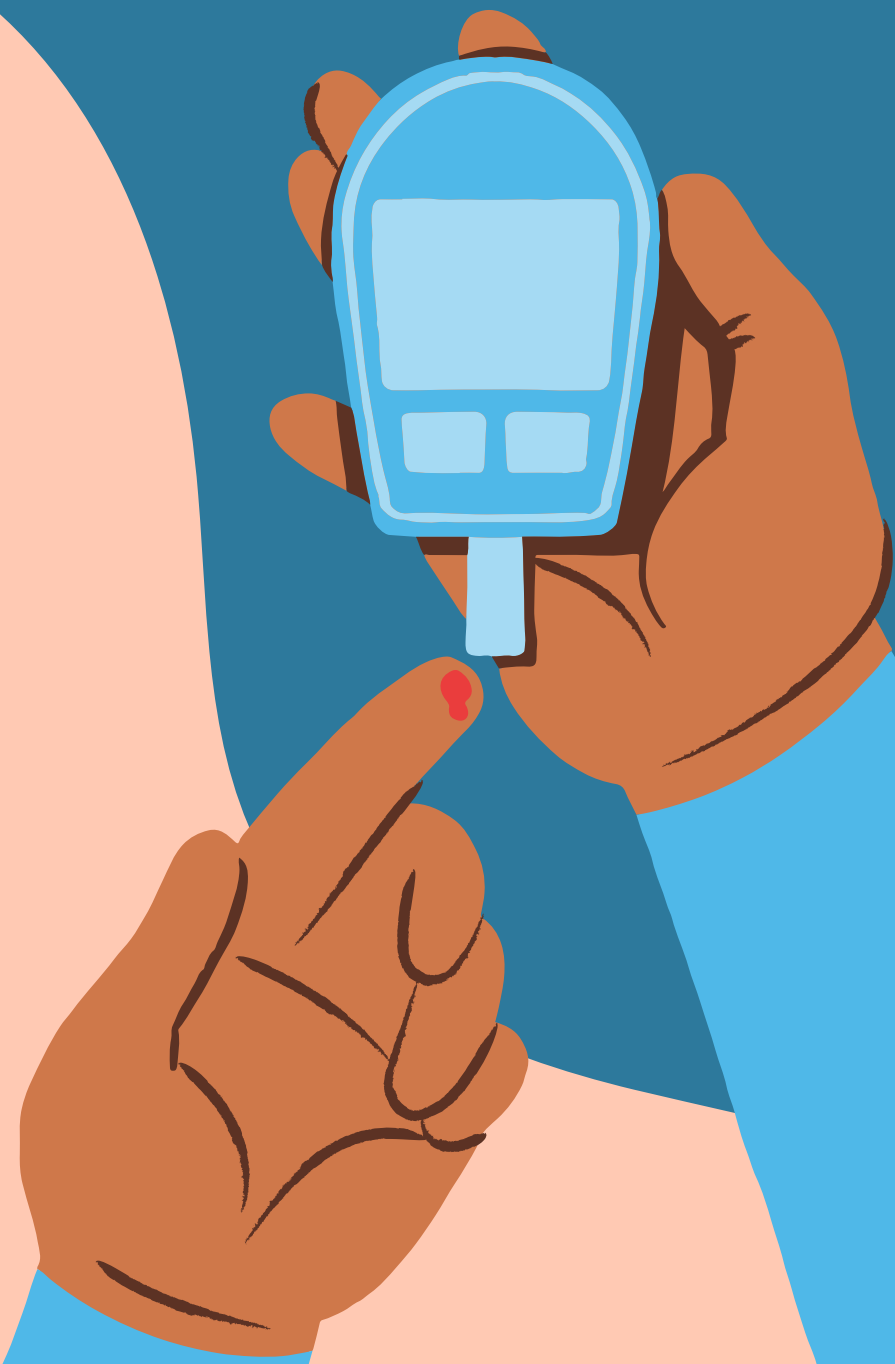


**BUT WHERE  
DOES  
BIAS/HEALTH  
DISPARITY FIT  
IN?**



# FUTURE PLANS:

- More research into IBM's AIF360 Toolkit
- Find a way to successfully implement it into model pre-processing
- Assess model performance and compare to baseline





**THANK YOU!  
SO LONG  
AND  
THANKS FOR  
ALL THE FISH!**