Mitigating Health Disparity in Hospital Admission for Diabetic Patients

Kristen Lo - BrainStation Data Science Capstone

TODAY'S AGENDA

1
The Problem

2
The Data

3
The Analysis

6
Future Plans

5
Current Status

4
The Models

THE PROBLEM

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

(1)Diabetes Canada (2)Grant, R et al. (3) Gov of Canada

THE DATA

What does it look like?

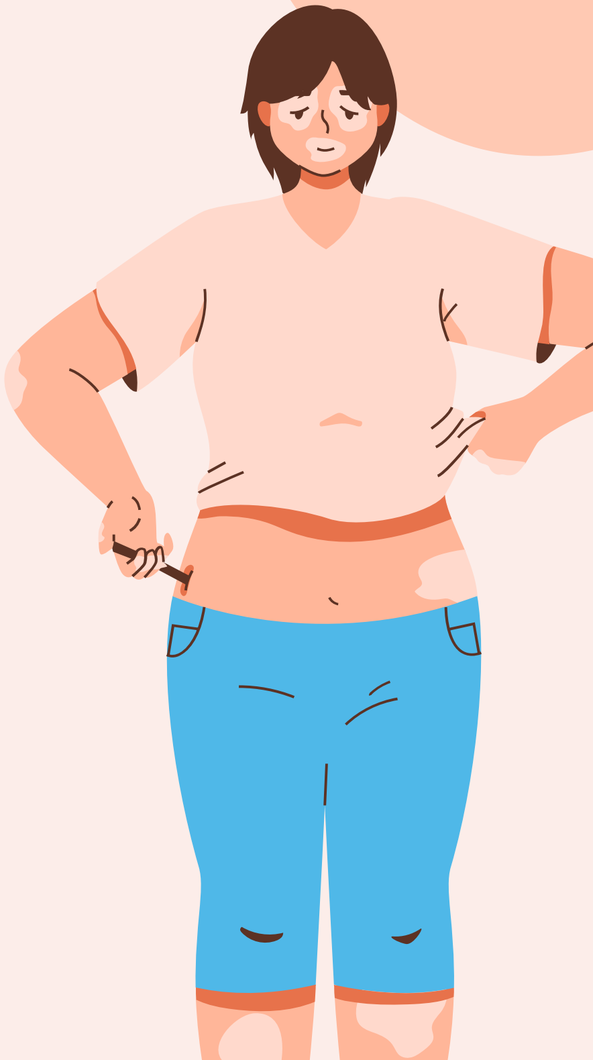
After filtering for **only patients with diabetes**, there were **110K rows and 487 columns**

What changes were made?

Dummying the columns
Feature Engineering/Elimination

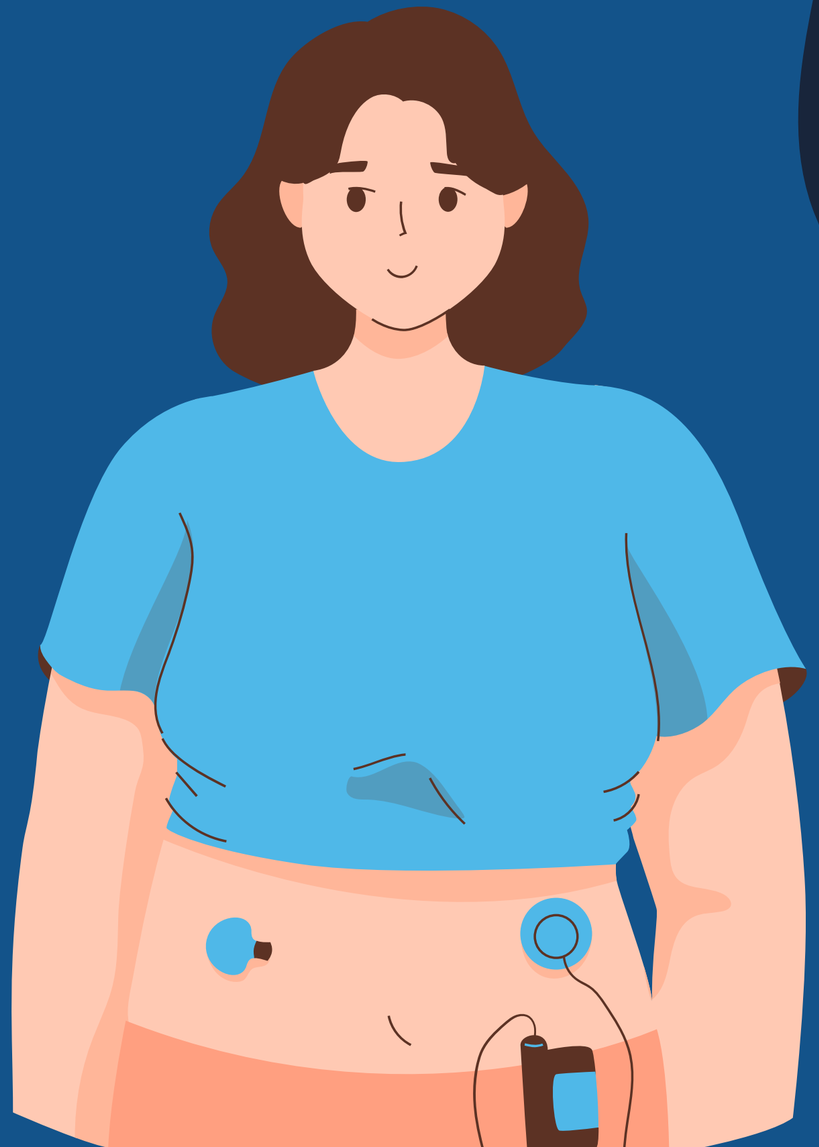
What does it look like after?

After **preprocessing** and **Feature Elimination**, there were **105K rows** and **419 columns**



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

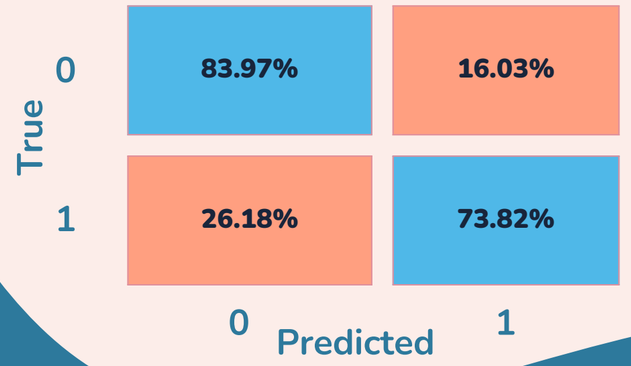
- Train Accuracy
- Test Accuracy
- 5 Fold Cross Validation (CV)
- Mean CV Accuracy
- Classification Report
- Confusion Matrix
- ROC - AUC Curve

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



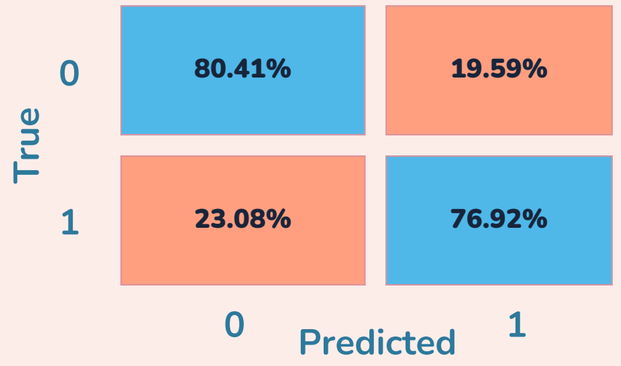
ROC AUC: **0.87**

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

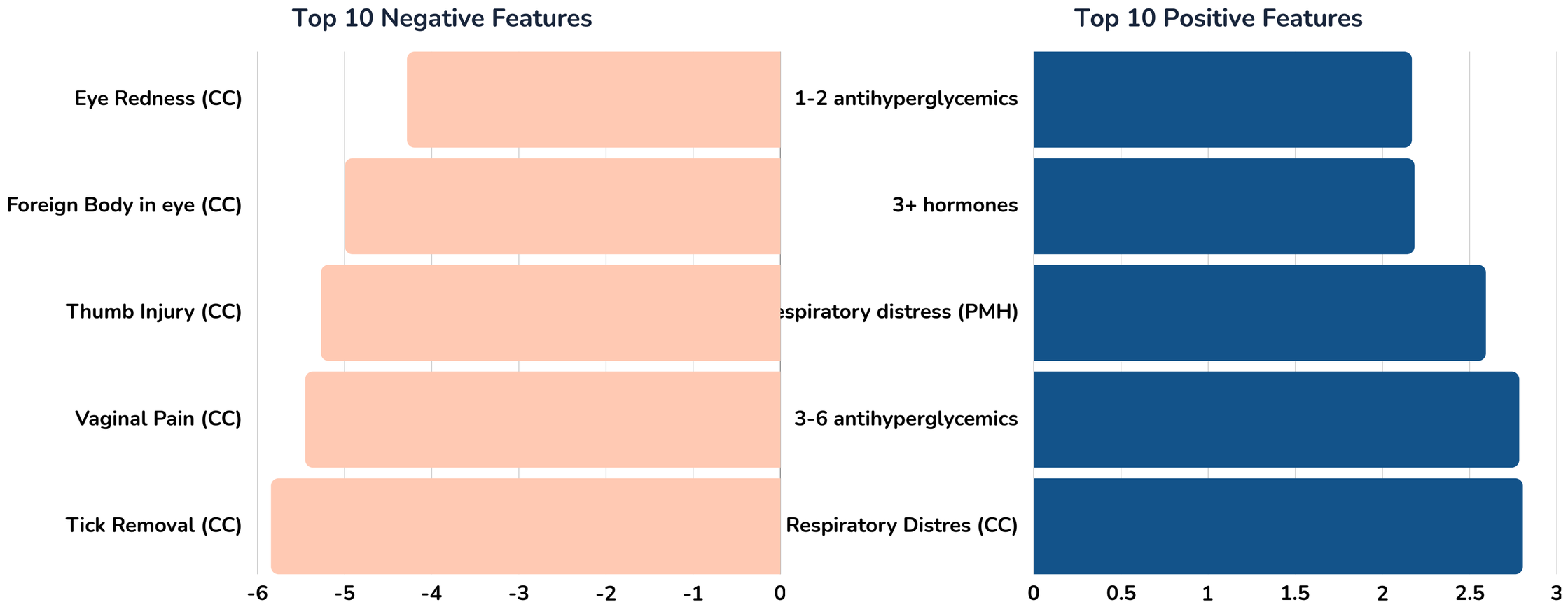
ROC AUC: **0.87**



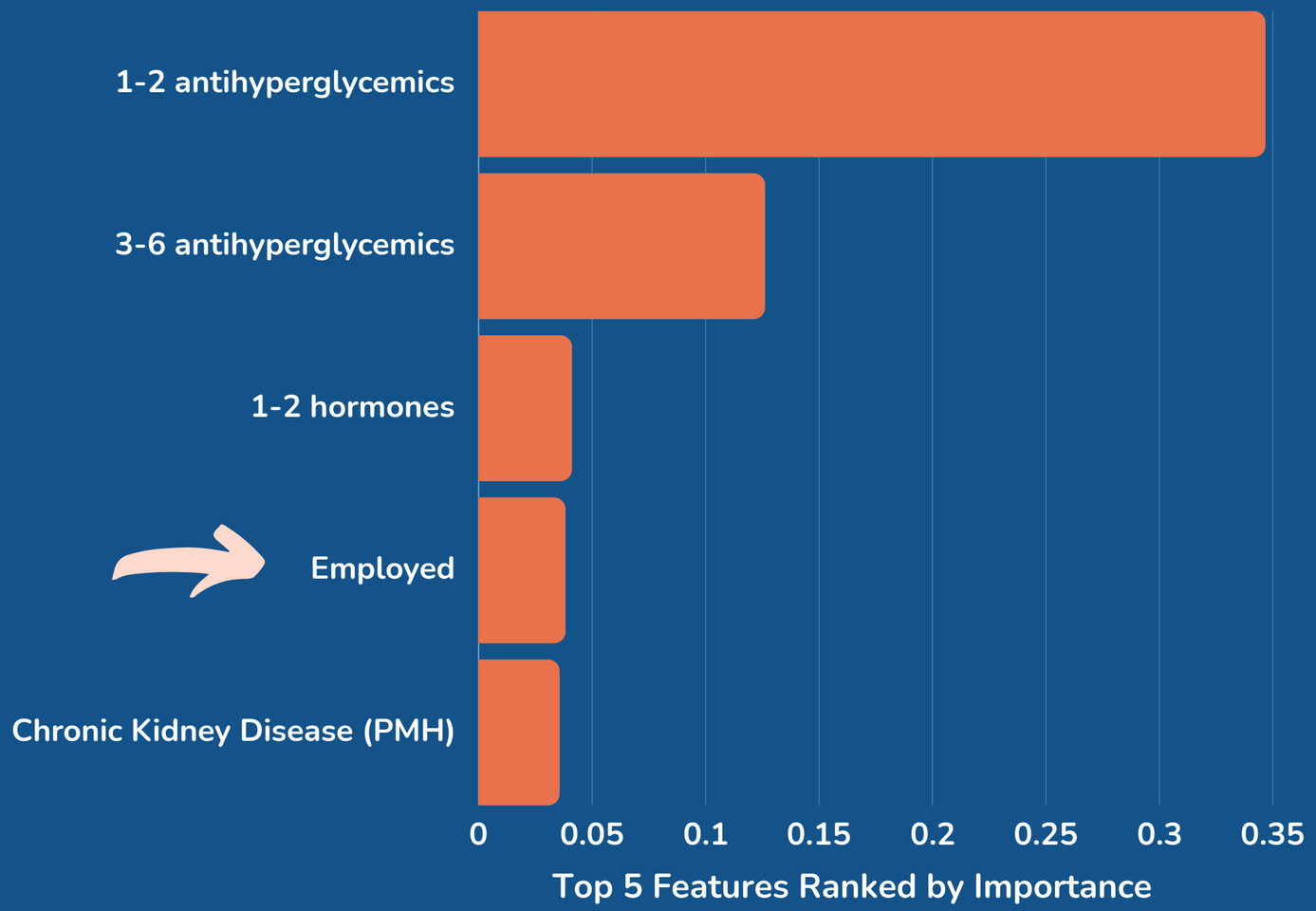
Summary of Model Performance

MODEL TYPE	TRAIN ACC	TEST ACC	MEAN CV	ROC-AUC
LOGISTIC REGRESSION (BASE)	0.792	0.793	0.789	0.87
DECISION TREE (OPTIMIZED)	0.800	0.782	0.747	0.80
RANDOM FOREST (OPTIMIZED)	0.849	0.779	0.787	0.87
XGBOOST (OPTIMIZED)	0.817	0.795	0.789	0.88

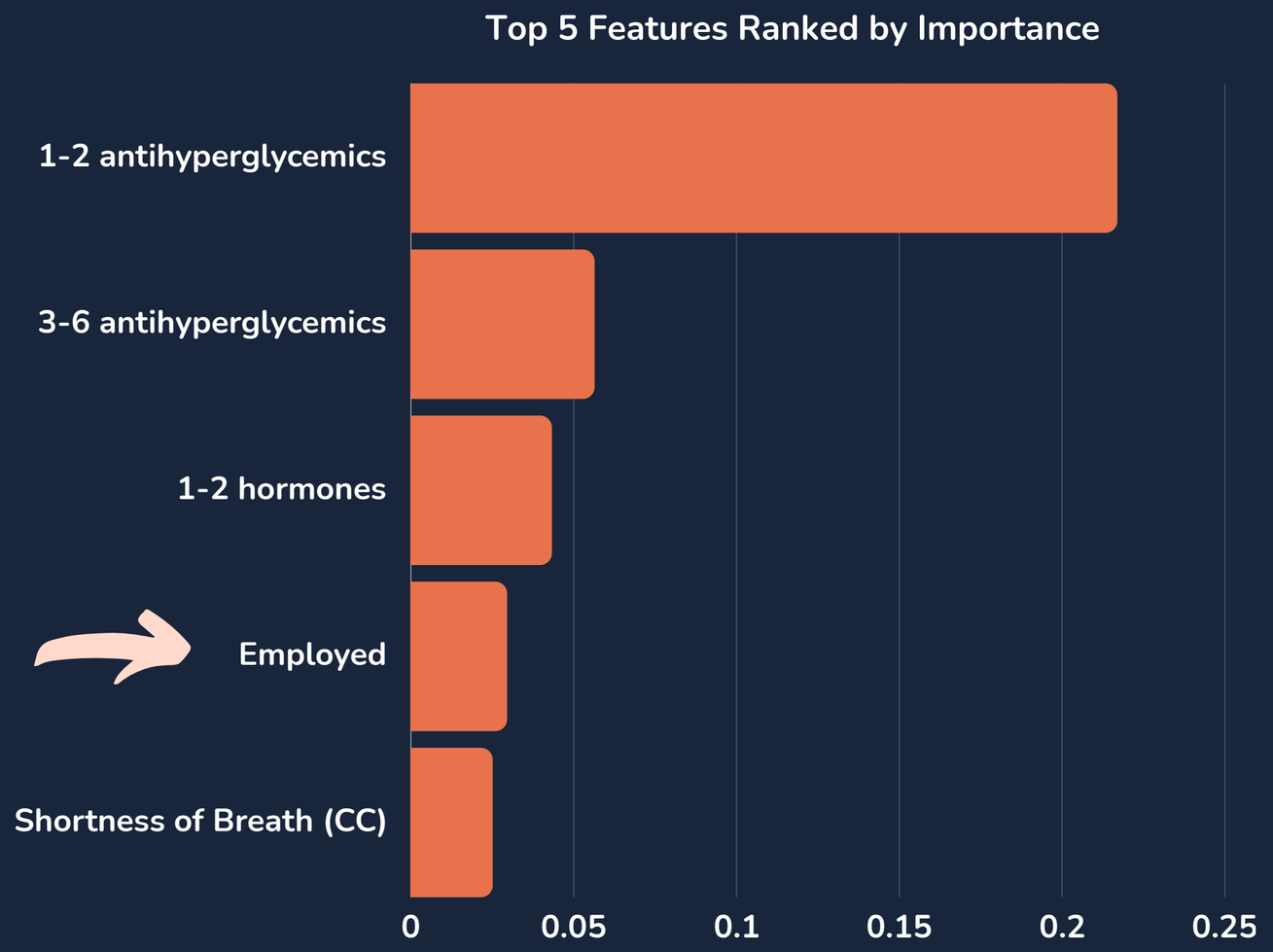
Feature Importance



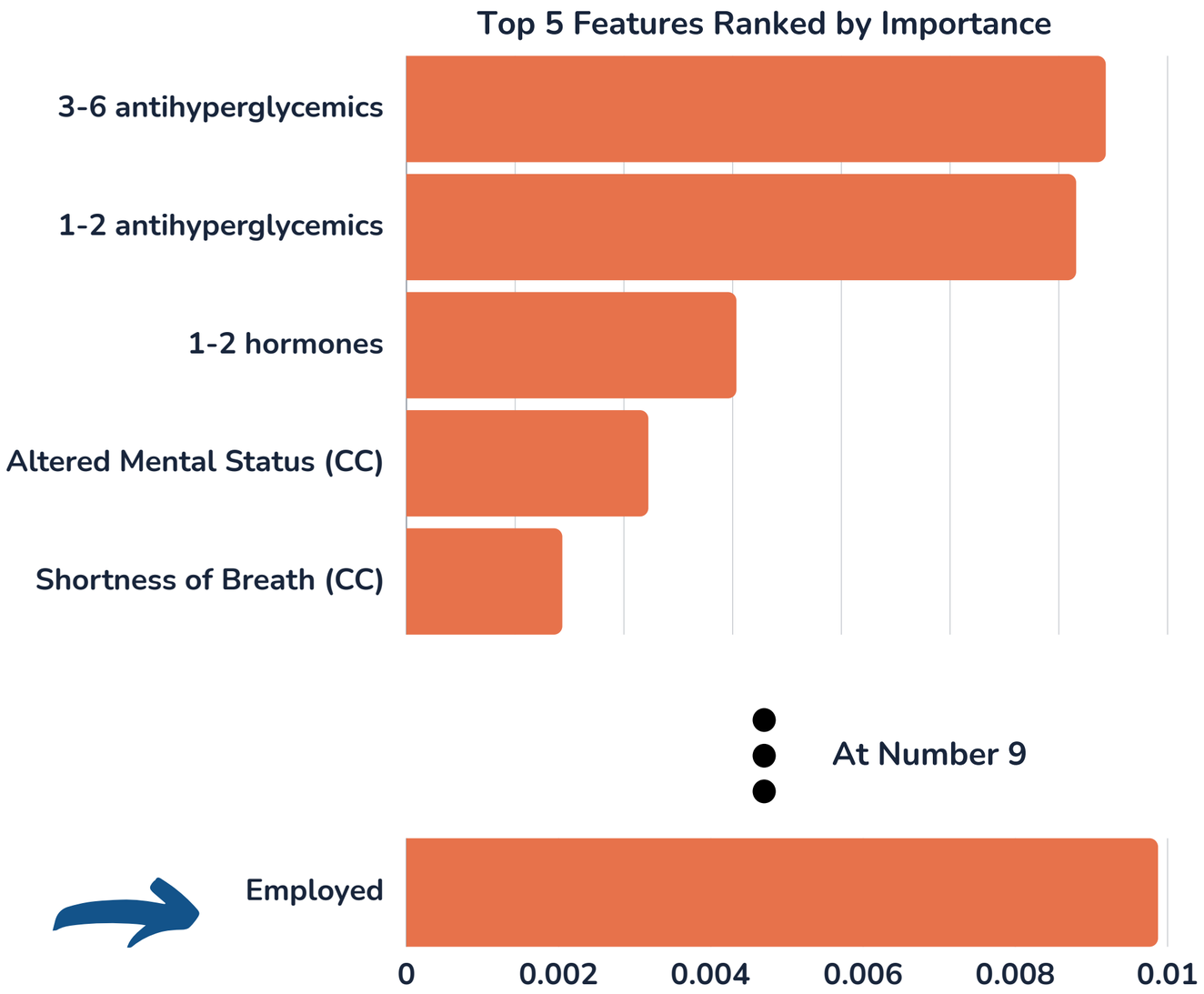
DECISION TREE



Random Forest

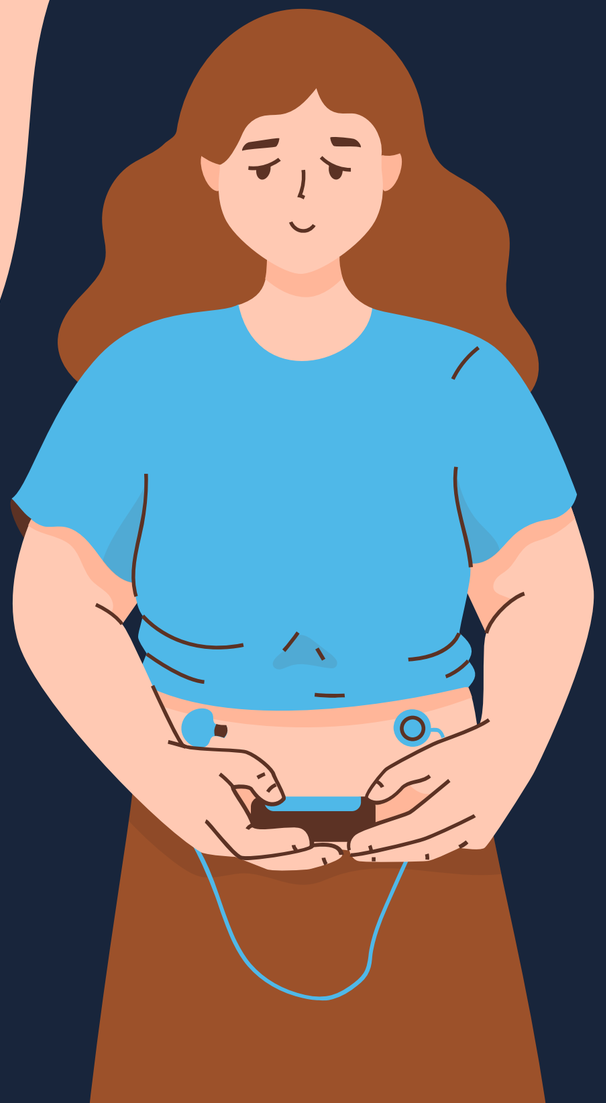


XGBOOST


Top 5 Features Ranked by Importance
3-6 antihyperglycemics
1-2 antihyperglycemics
1-2 hormones
Altered Mental Status (CC)
Shortness of Breath (CC)
At Number 9
Employed
0
0.002
0.004
0.006
0.008
0.01

CURRENT STATUS

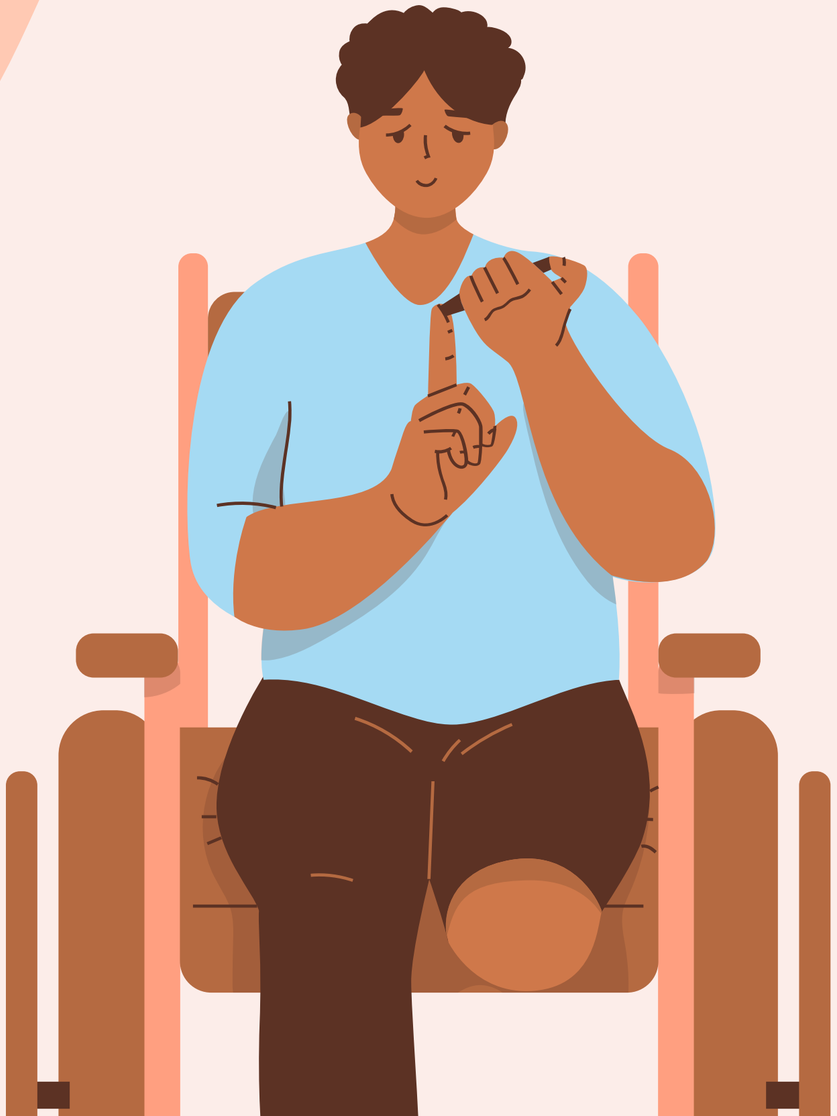
- Increasing **Explainability**



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 HEALTH DISPARITY FIT IN?

FUTURE PLANS:

- More research into **IBM's AIF360 Toolkit**
- Find a way to **successfully implement** it into **model pre-processing**
- Perhaps predict the **Emergency Severity Index(ESI)** next

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