

# Detecting Diabetes

---

## Health App Features



Mekdes Wassie

# Problem Statement

---

Project is aimed at **identifying features** which can be added to a health app that does a preliminary diabetes screening and recommends interventions to individuals based on the outcome.

# What is Diabetes?

---

**High Blood Sugar**

**Insulin**

Low or can't be used

**Health Problems**

kidney disease, heart disease

**Lifestyle Changes**

# Methodology

---

- We utilized machine learning models to identify individuals with diabetes.
- Using the National Health and Nutrition Examination Survey (NHANES) dataset of 2017 to 2020, we searched for all available feature variables within the data to develop models for diabetes detection.

# Exploratory Data Analysis

---

## Positive Correlation

- age
- waist-hip ratio
- BMI
- blood pressure

# Data Distribution



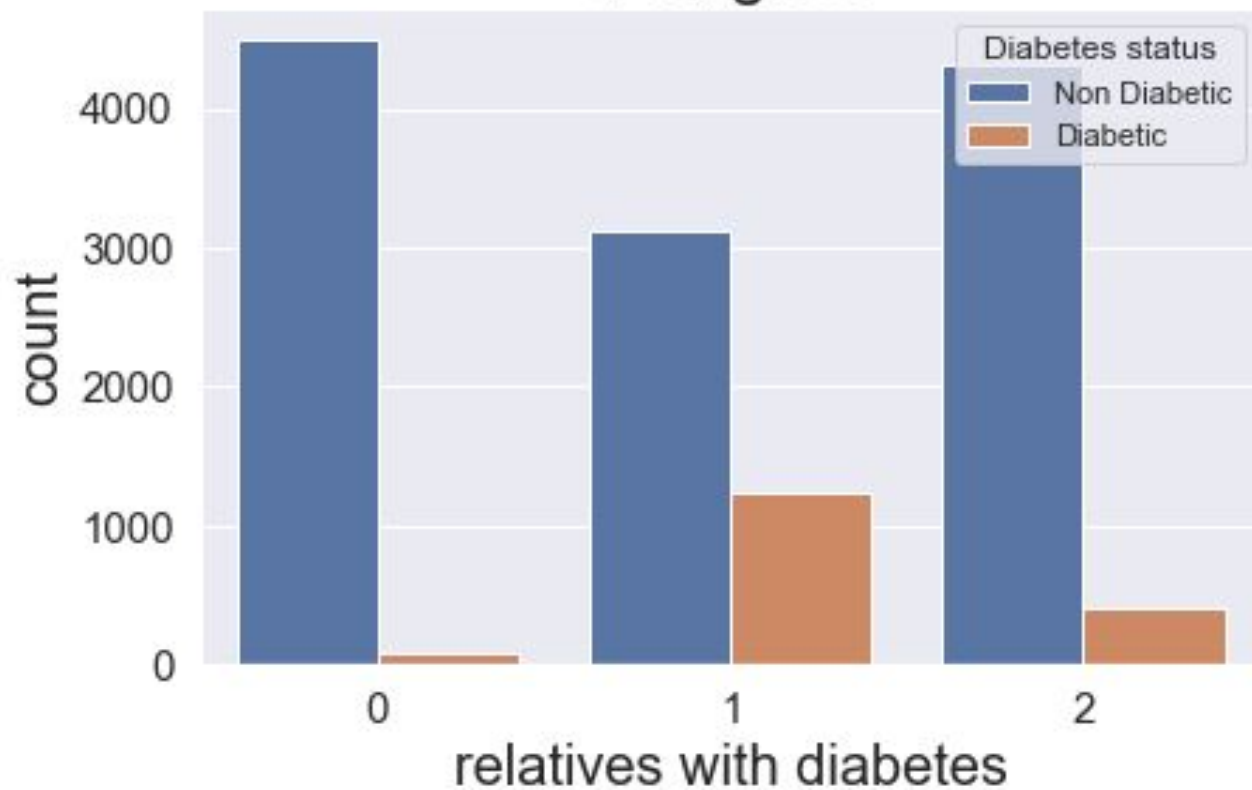
## Outcome Variable

a yes or no response to the question 'have your dr told you have diabetes?'.

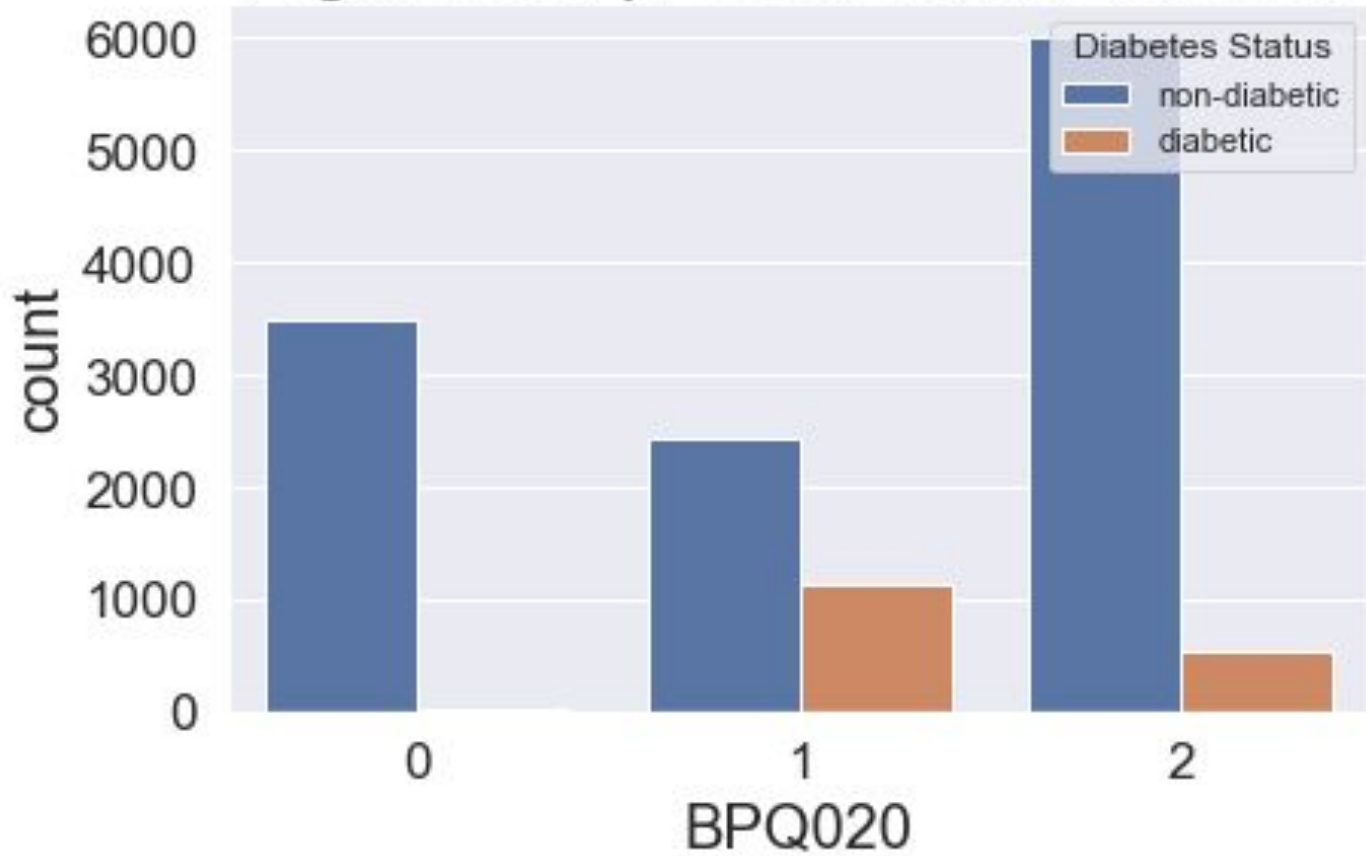
## Class Imbalance

There is a class imbalance; the positive class makes up only 12%

## Pedigree

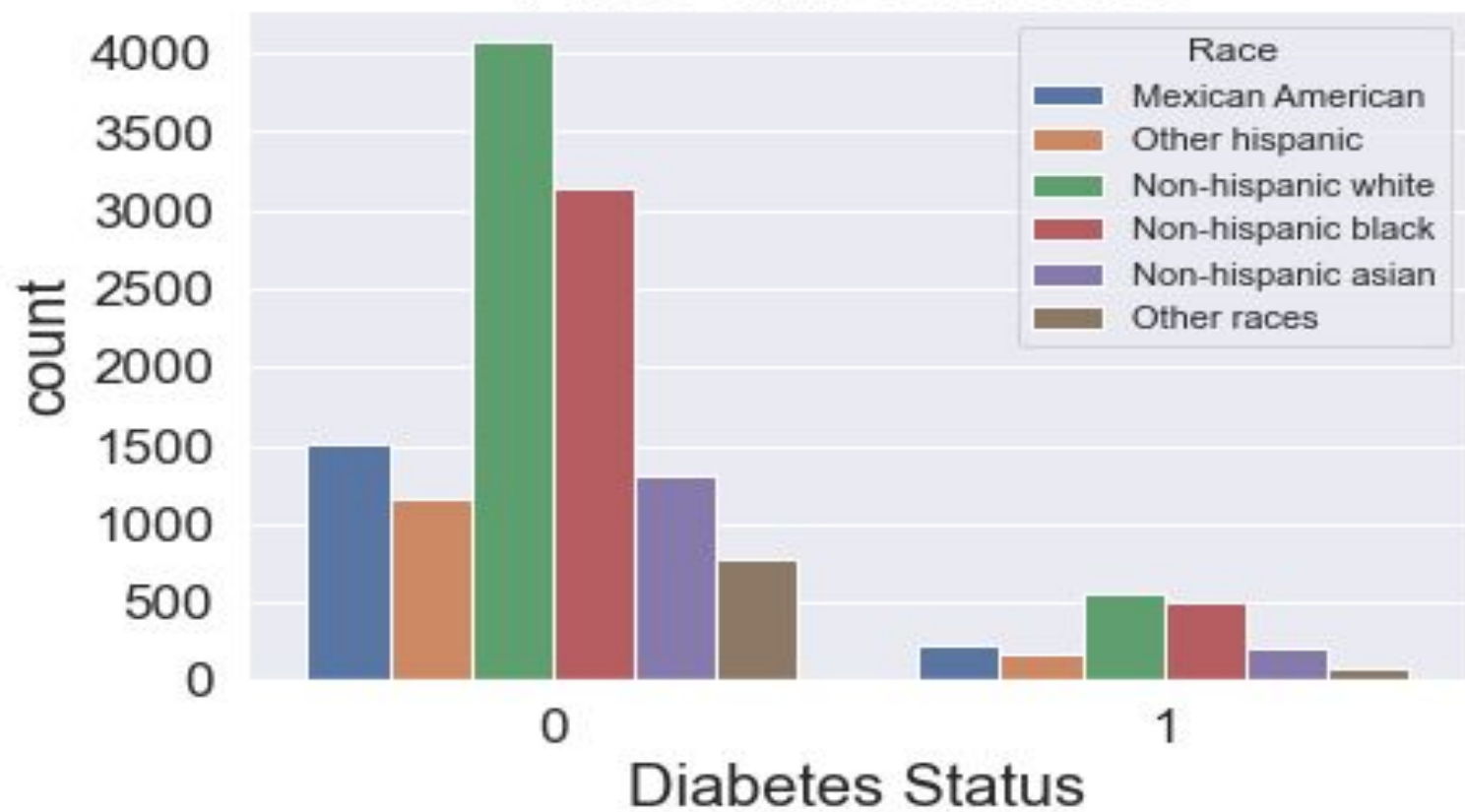


## High Blood pressure and Diabetes

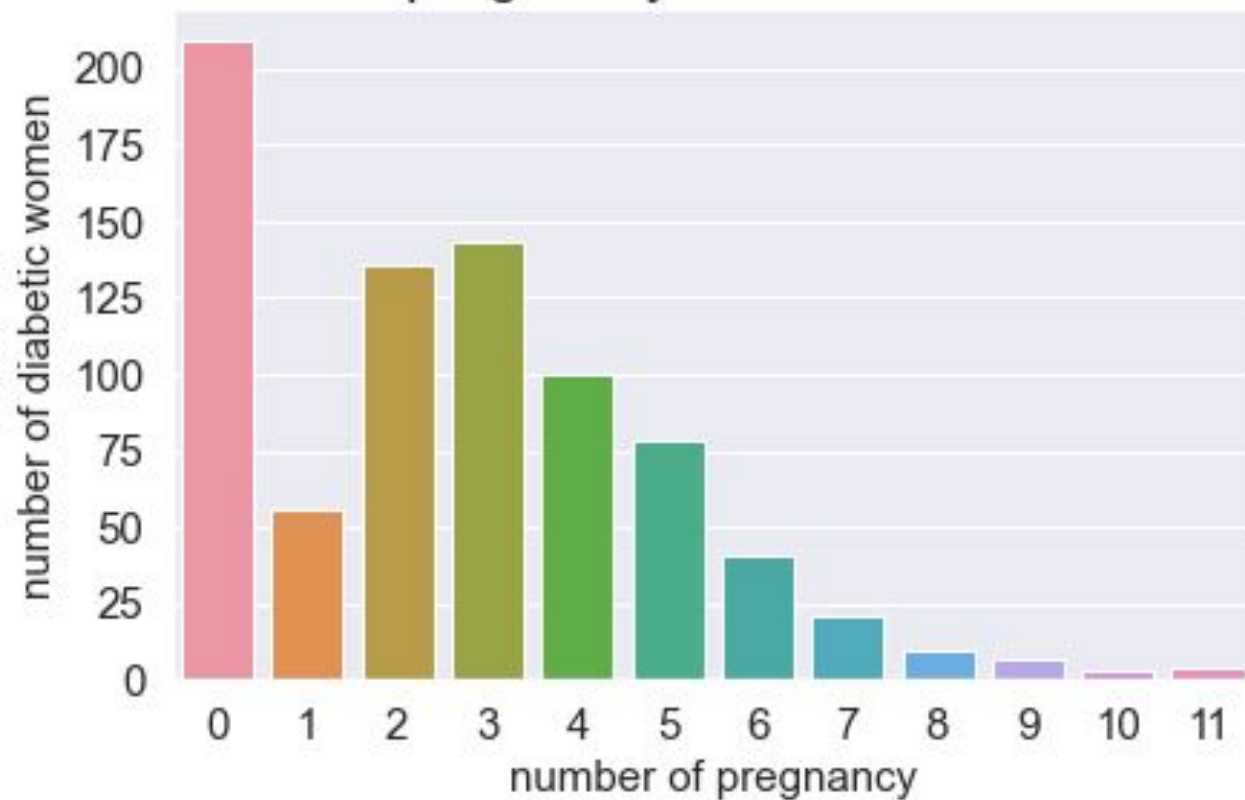




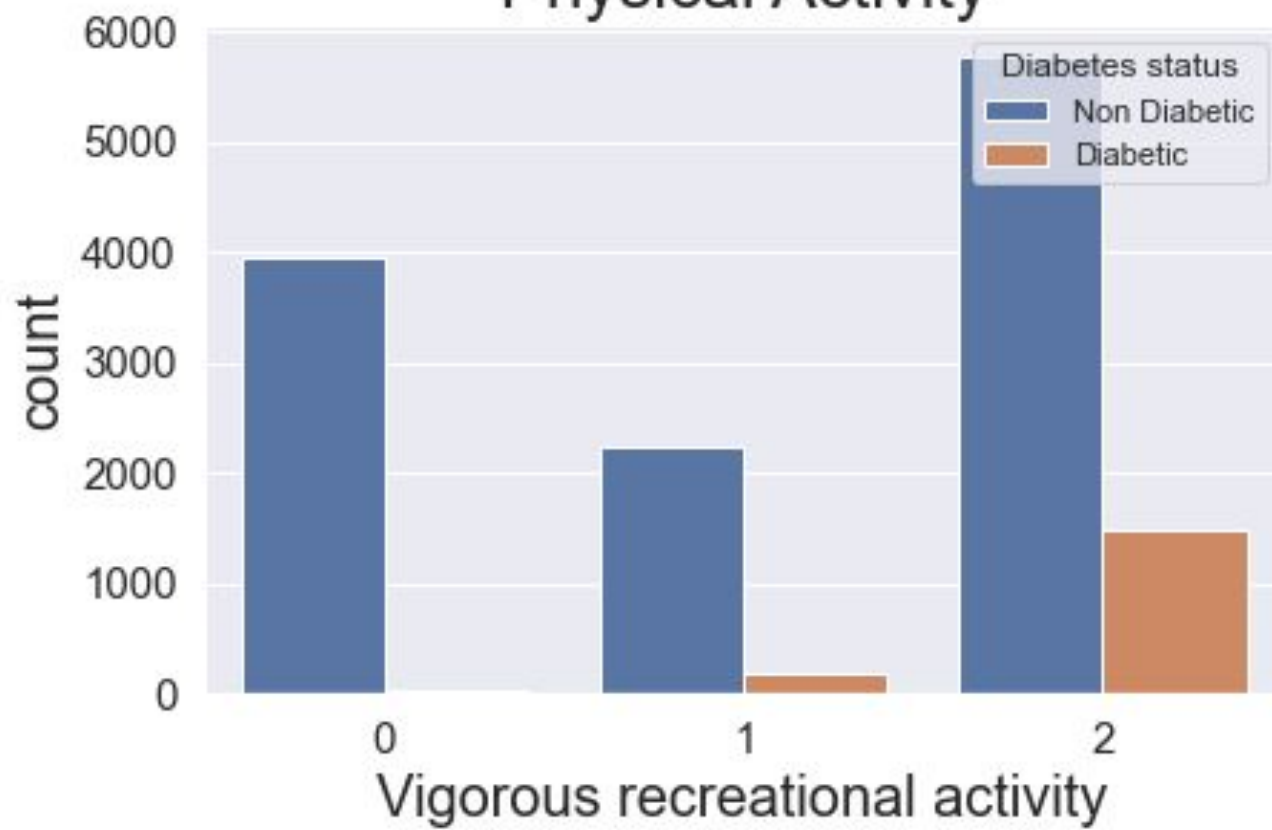
## Race and Diabetes



pregnancy and diabetes



## Physical Activity



# Modeling

## Considerations

**Interpretability**

Which features have strong predictive power

**False negatives**

Optimize for false negatives

**Imbalanced data**

techniques



# Modeling

---

## Features

- age
- waist-hip ratio
- body mass index
- physical activity features
- race
- High blood pressure
- pedigree
- smoking status
- number of pregnancies

# Modeling

---

## Binary Classifier

- Logistic Regression
- Random Forest
- Gradient Boosting
- Multinomial Naive Bayes
- Knn

# Modeling

---

Model of Choice

## Logistic Regression (with undersampling)

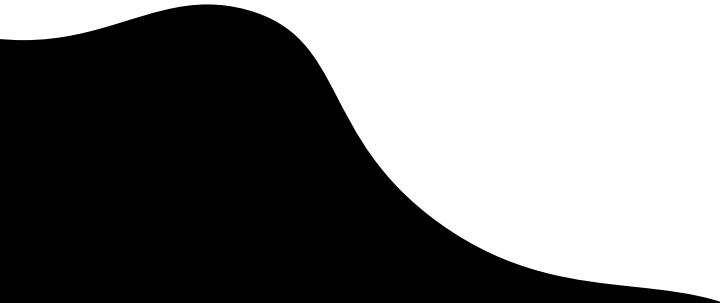
- Recall score of 85%
- Accuracy of 88%
- 66 **False negatives** and 684 False positives

## Logistic Regression

- Recall score of 24%
- Accuracy of 78%
- 327 **False negatives** and 95 False positives

# Findings

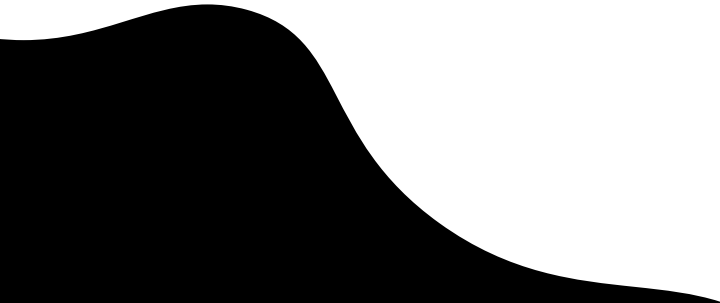
---

- If one has a high blood pressure, the odds of suffering from diabetes increases by 67%
  - If one has close relatives that are diabetic, the odds of being diabetic increases by 41%
  - If one has no close relatives that are diabetic, the odds of being diabetic decreases by 55%
  - As age increases by one year, the odds of being diabetic increases by 6%
- 



# Findings

---

- As BMI increases by one unit, the odds of being diabetic increases by 6%
  - As number of pregnancy increases by one, the odds of being diabetic decreases by 5%
  - If you are non-hispanic white, the odds of getting diabetic decreases by 17%.
- 

# Recommendations

- High blood pressure and family history of diabetes are stronger predictors of diabetes. Age, BMI and number of pregnancies also have predictive power.
- Therefore, these features may be added to the app.
- But this model is not perfect. It was trained on imbalance data so getting better data with equal representation of the two classes will help the model to better classify and minimize false positives and false negatives.