



SUPERVISED LEARNING DIABETES DATASET

Akash Kadia

A decorative graphic on the left side of the slide, composed of several overlapping geometric shapes and patterns. It includes a blue triangle with white concentric circles, a purple triangle with white concentric circles, a pink square with white concentric circles, and a grey square with white concentric circles. The overall design is modern and abstract.

FLOW STRUCTURE

- ☐ Project Goal
- ☐ EDA - Exploratory Data Analysis
- ☐ Preprocessing & Feature Engineering
- ☐ Training ML Model
- ☐ Conclusion



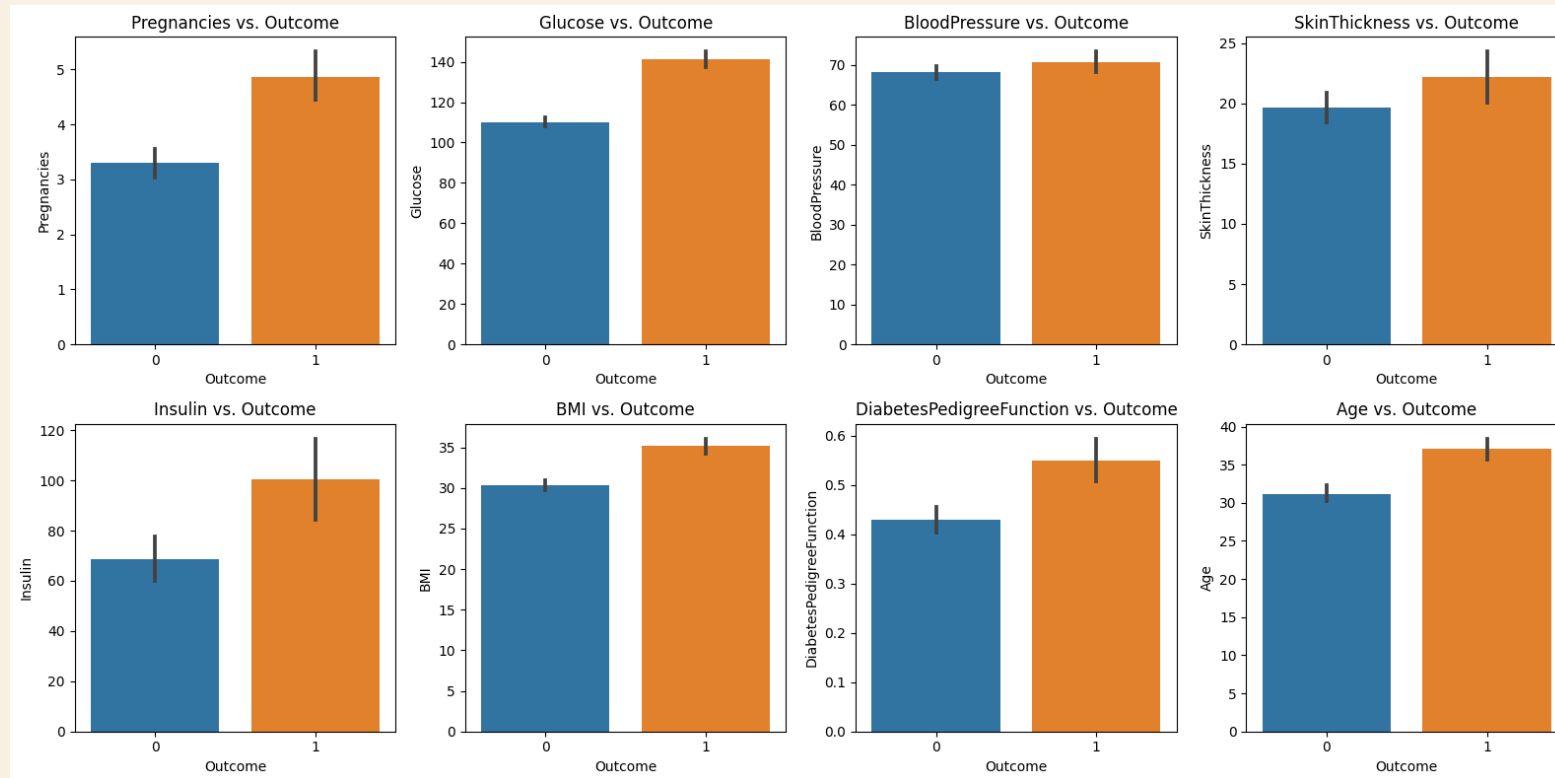
PROJECT GOAL

- ❑ Supervised Learning Techniques
- ❑ Build a Machine Learning model
- ❑ predict whether a patient has diabetes or not

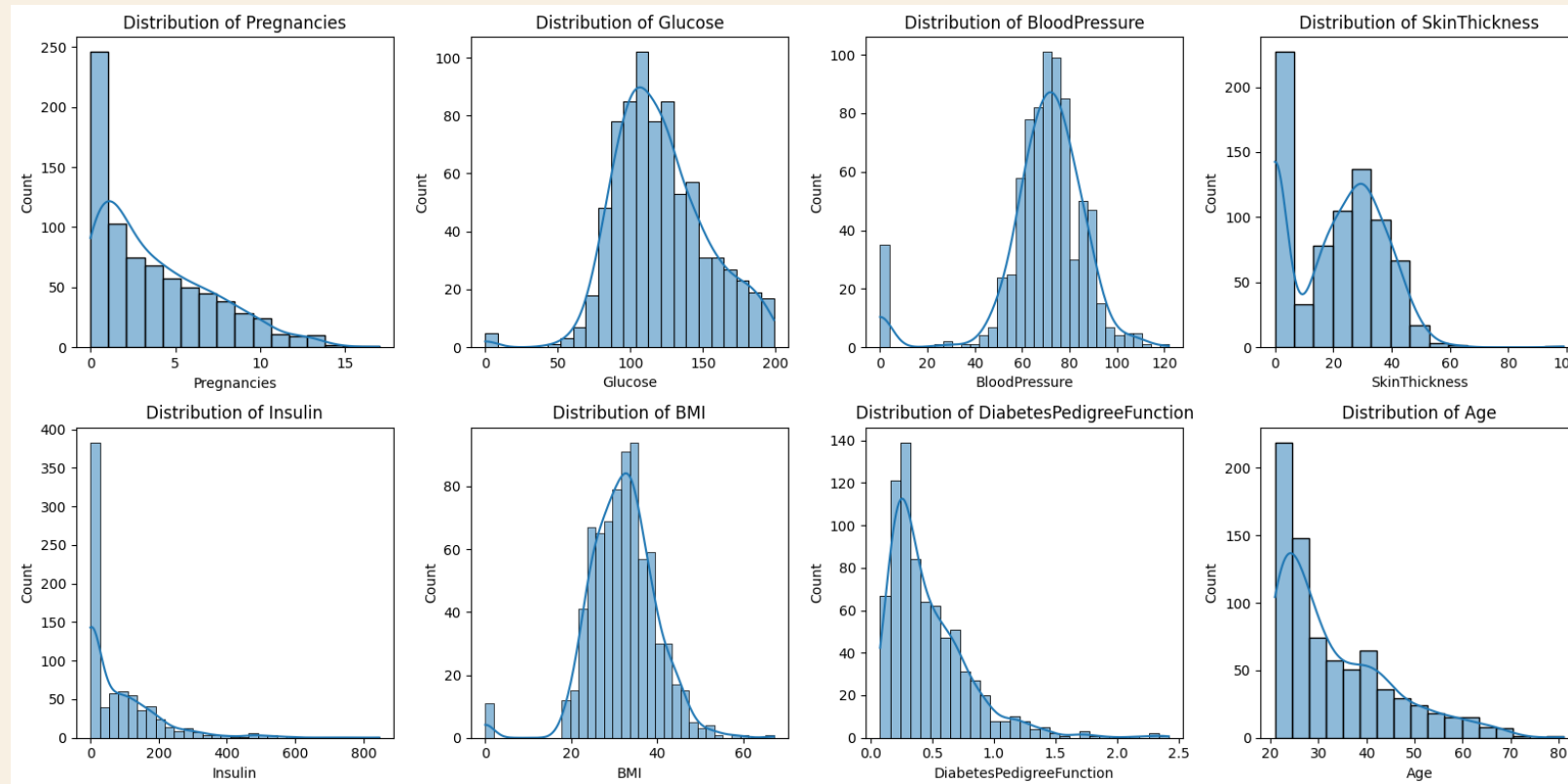


EXPLORATORY DATA ANALYSIS

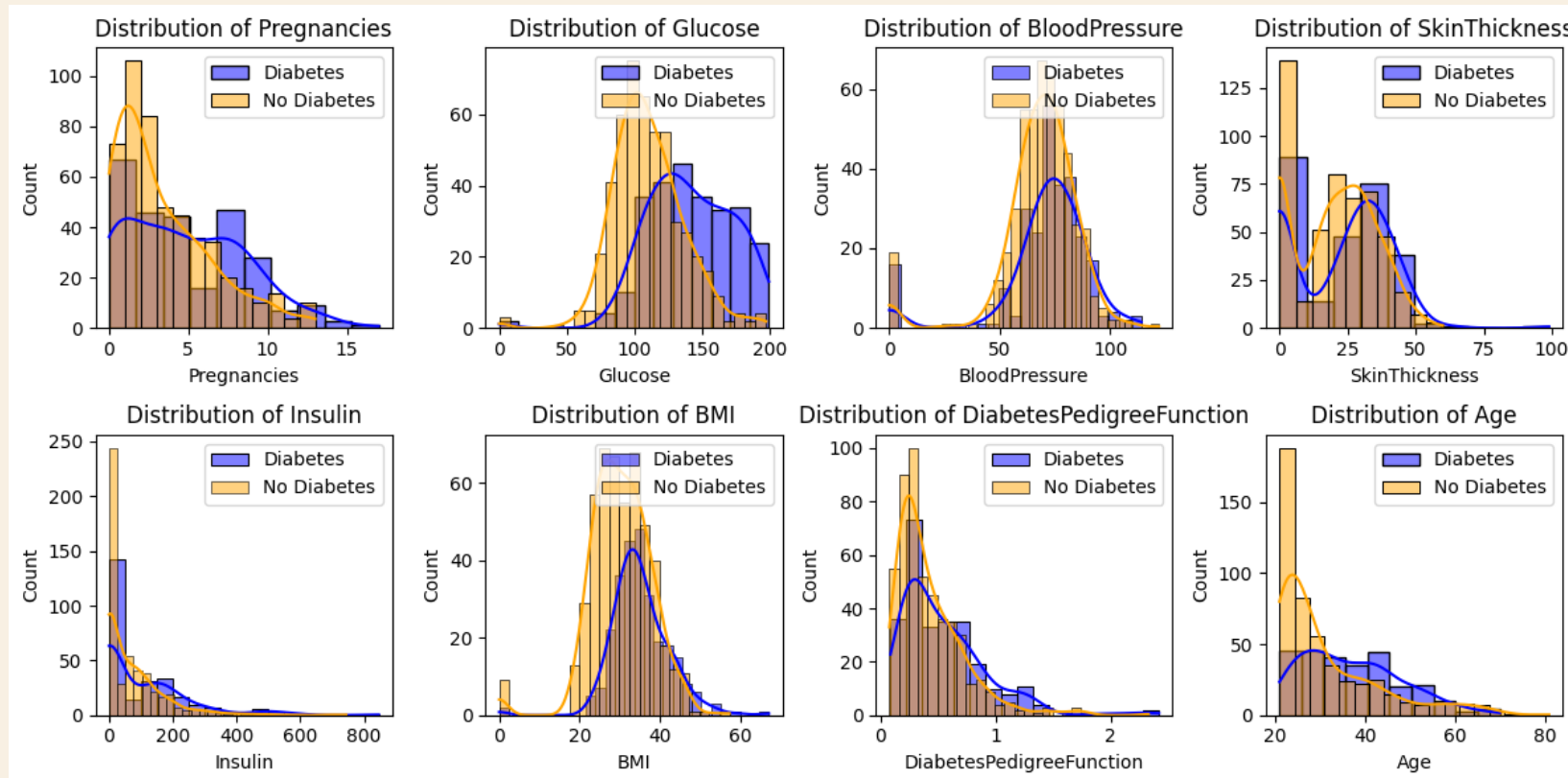
RELATIONSHIP BETWEEN PREDICTOR VARIABLES AND THE OUTCOME VARIABLE



DISTRIBUTION OF EACH PREDICTOR VARIABLES



DISTRIBUTION OF PREDICTOR VARIABLES FOR INDIVIDUALS WITH AND WITHOUT DIABETES





PROCESSING & FEATURE ENGINEERING



FEATURE ENGINEERING

With the help of Subject Expert

AGE GROUP BINNING

- Grouping age in ranges
- Figuring out Type-1 diabetes

BMI CATEGORIES

- Categories
 - Underweight
 - Normal
 - Overweight
 - Obese
- Obese are more diabetes positive

RATIO

Insulin to Glucose ratio

The top-left corner of the slide features a complex arrangement of geometric shapes and patterns. It includes a dark purple square, a blue square with concentric circles, a magenta triangle, a light purple circle, a grey rectangle, a magenta square, a dark purple square with a white semi-circular pattern, and a magenta square. A white line with a dot extends from the top-left towards the center of the slide.

MODEL BUILDING

MODEL COMPARISON

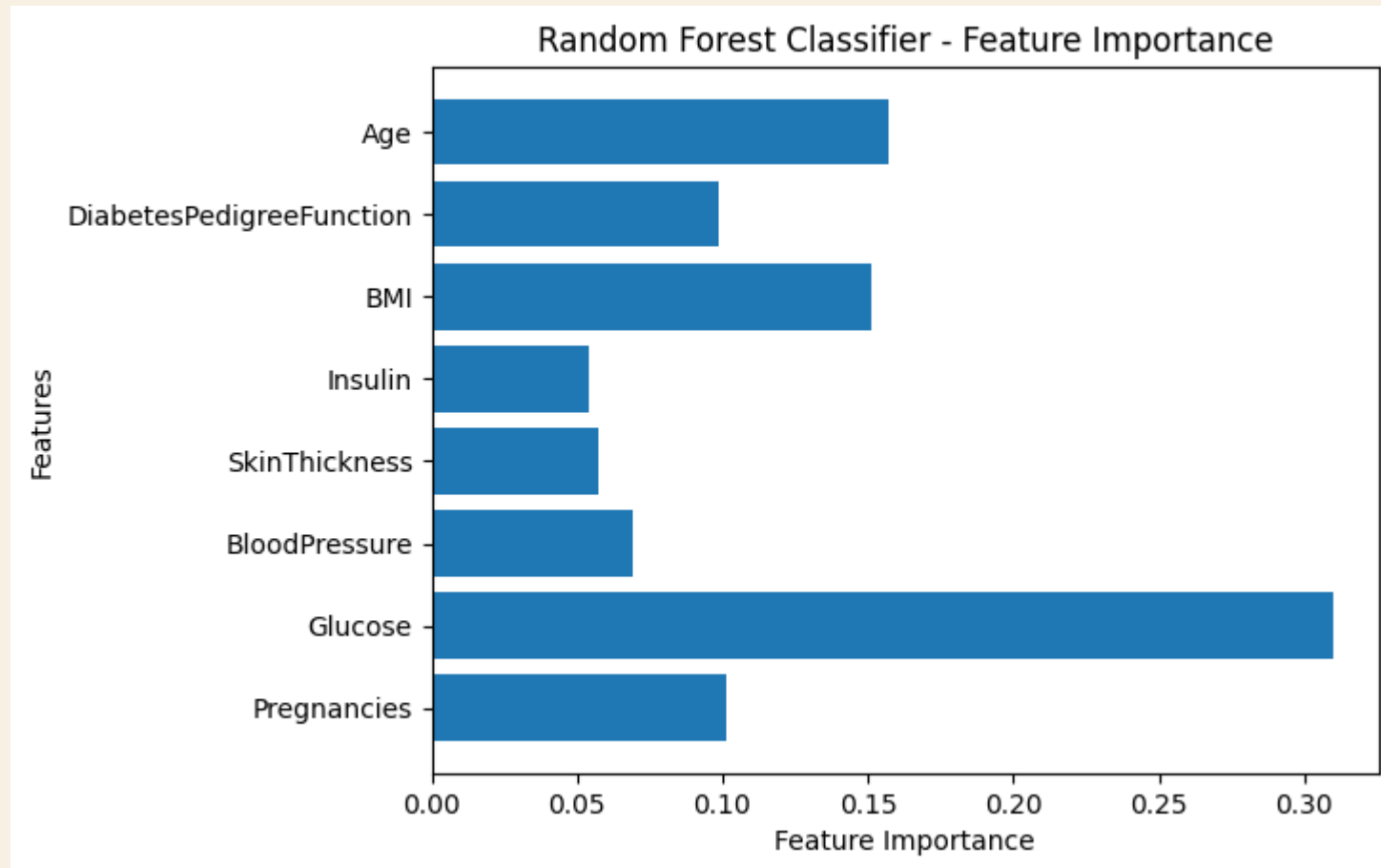
K NEIGHBORS CLASSIFIER

```
Model 2 (Random Forest) Evaluation:  
Accuracy: 0.78  
Precision: 0.6  
Recall: 0.46153846153846156  
F1-Score: 0.5217391304347826  
ROC-AUC: 0.7988565488565489
```

RANDOM FOREST CLASSIFIER

```
Model 1 (KNeighborsClassifier) Evaluation:  
Accuracy: 0.76  
Precision: 0.55  
Recall: 0.4230769230769231  
F1-Score: 0.47826086956521735  
ROC-AUC: 0.7697505197505197
```

FEATURE IMPORTANCE



An abstract geometric design on the left side of the slide. It features a dark blue background with various geometric shapes and patterns. A white circle is positioned near the top left. Below it, a light blue semi-circle is visible. To the right of the semi-circle, there is a pink triangle with diagonal lines. Further down, there is a pink square with a pattern of concentric lines. At the bottom, there is a pink triangle with a pattern of concentric lines. The overall design is modern and minimalist.

THANK YOU