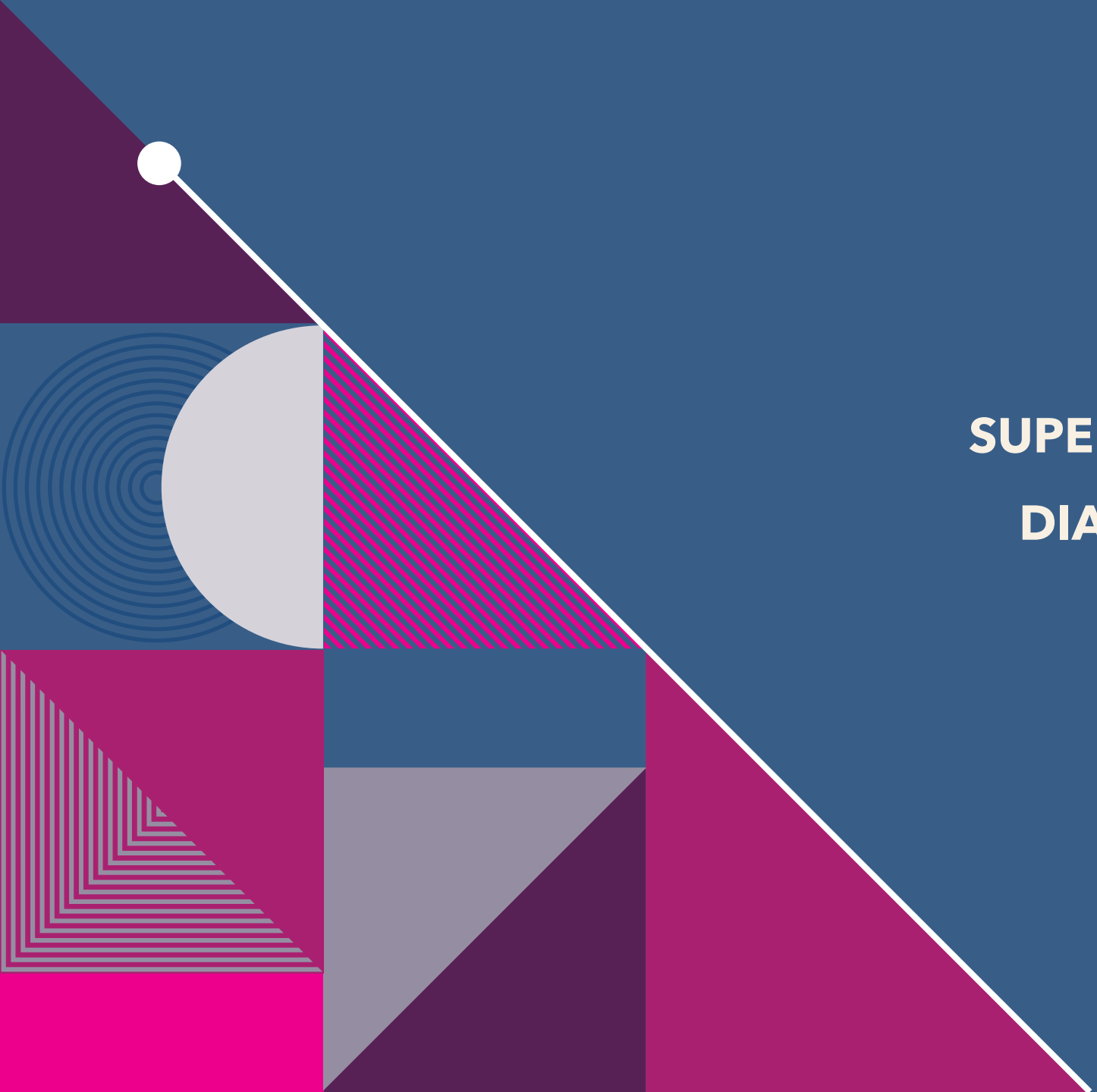


SUPERVISED LEARNING

DIABETES DATASET

Akash Kadia

FLOW STRUCTURE

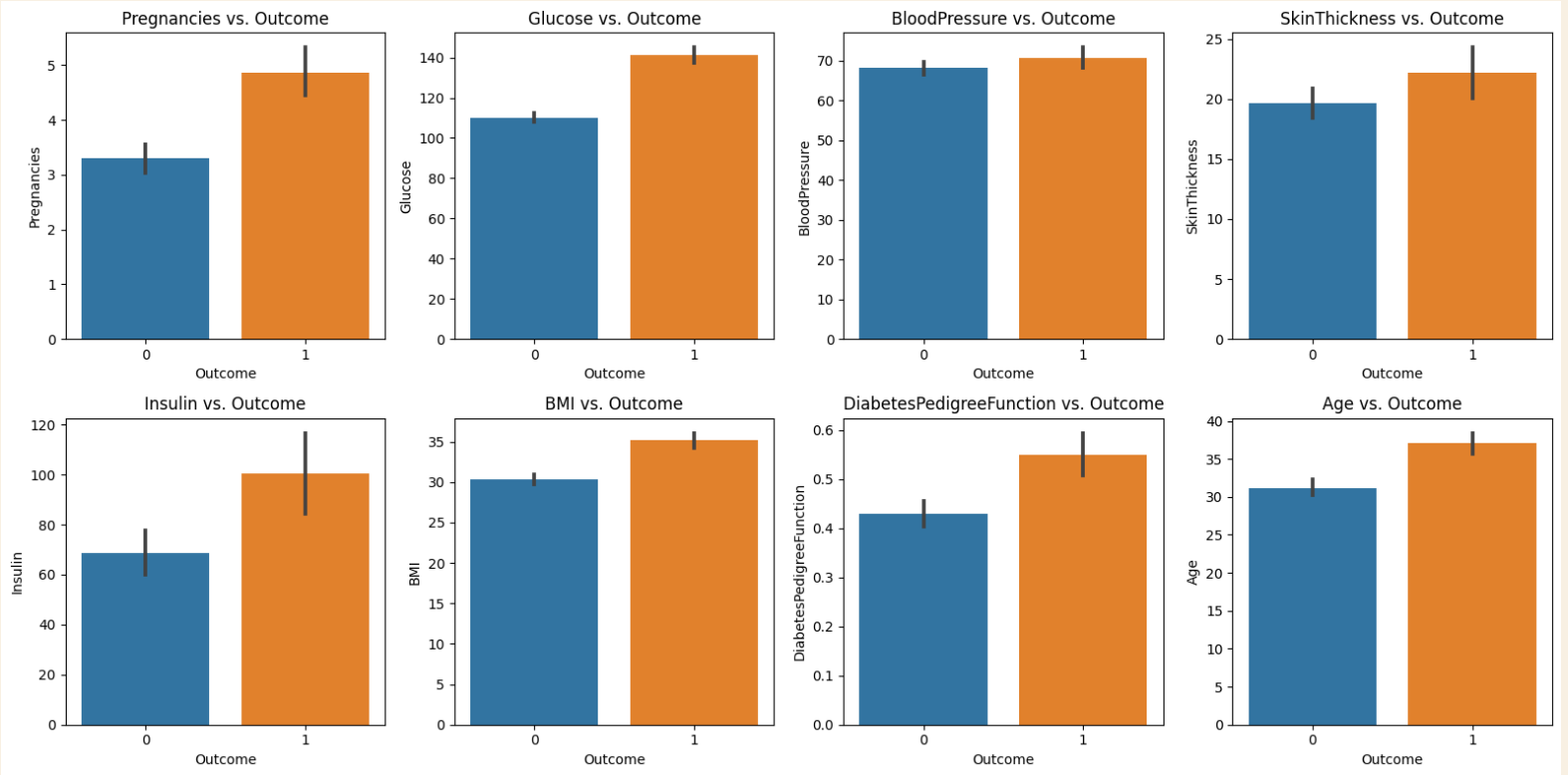
- ❑ Project Goal
- ❑ EDA – Exploratory Data Analysis
- ❑ Preprocessing & Feature Engineering
- ❑ Training ML Model
- ❑ Conclusion

PROJECT GOAL

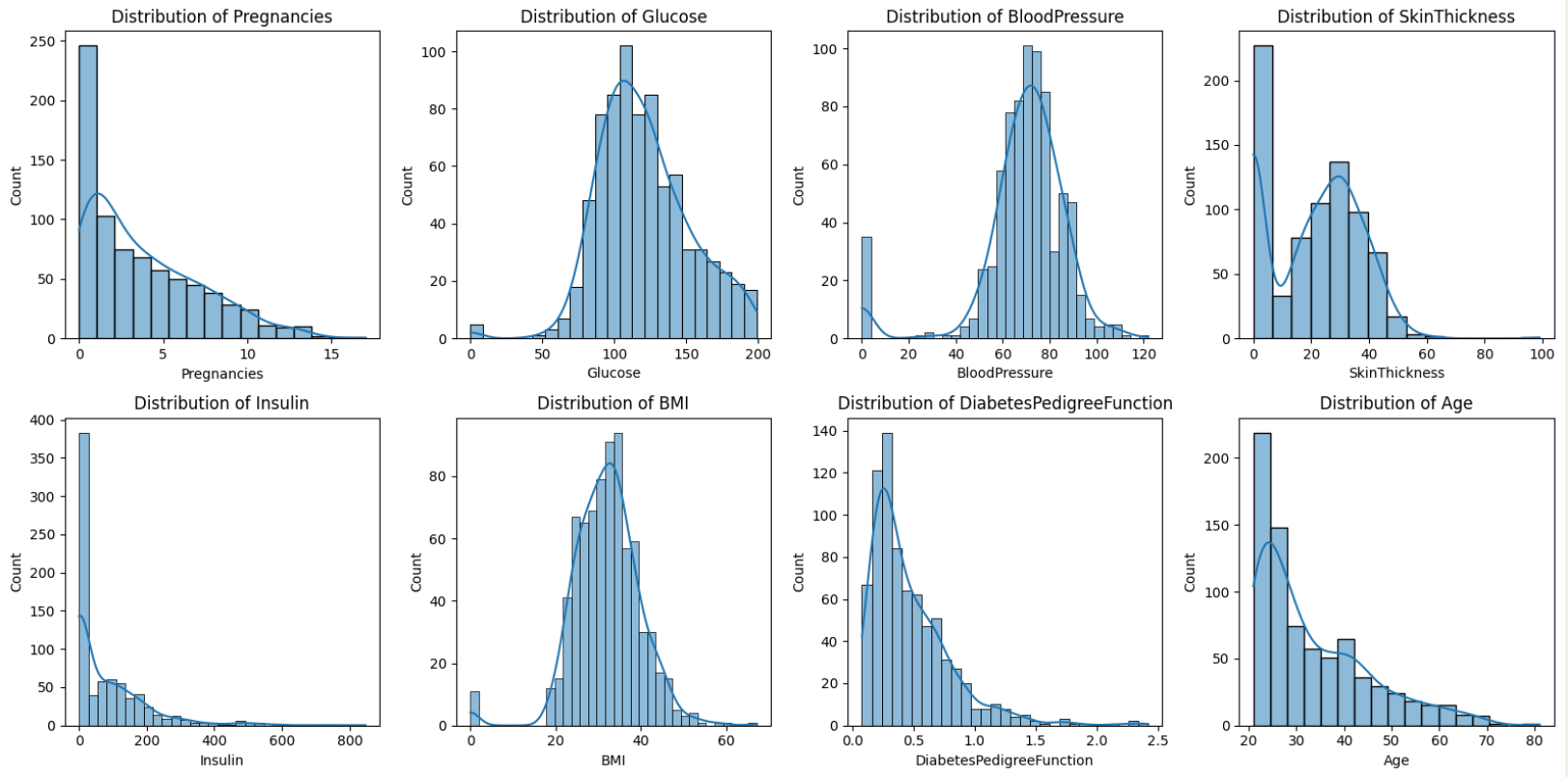
- ❑ Supervised Learning Techniques
- ❑ Build a Machine Learning model
- ❑ predict whether a patient has diabetes or not

EXPLORATORY DATA ANALYSIS

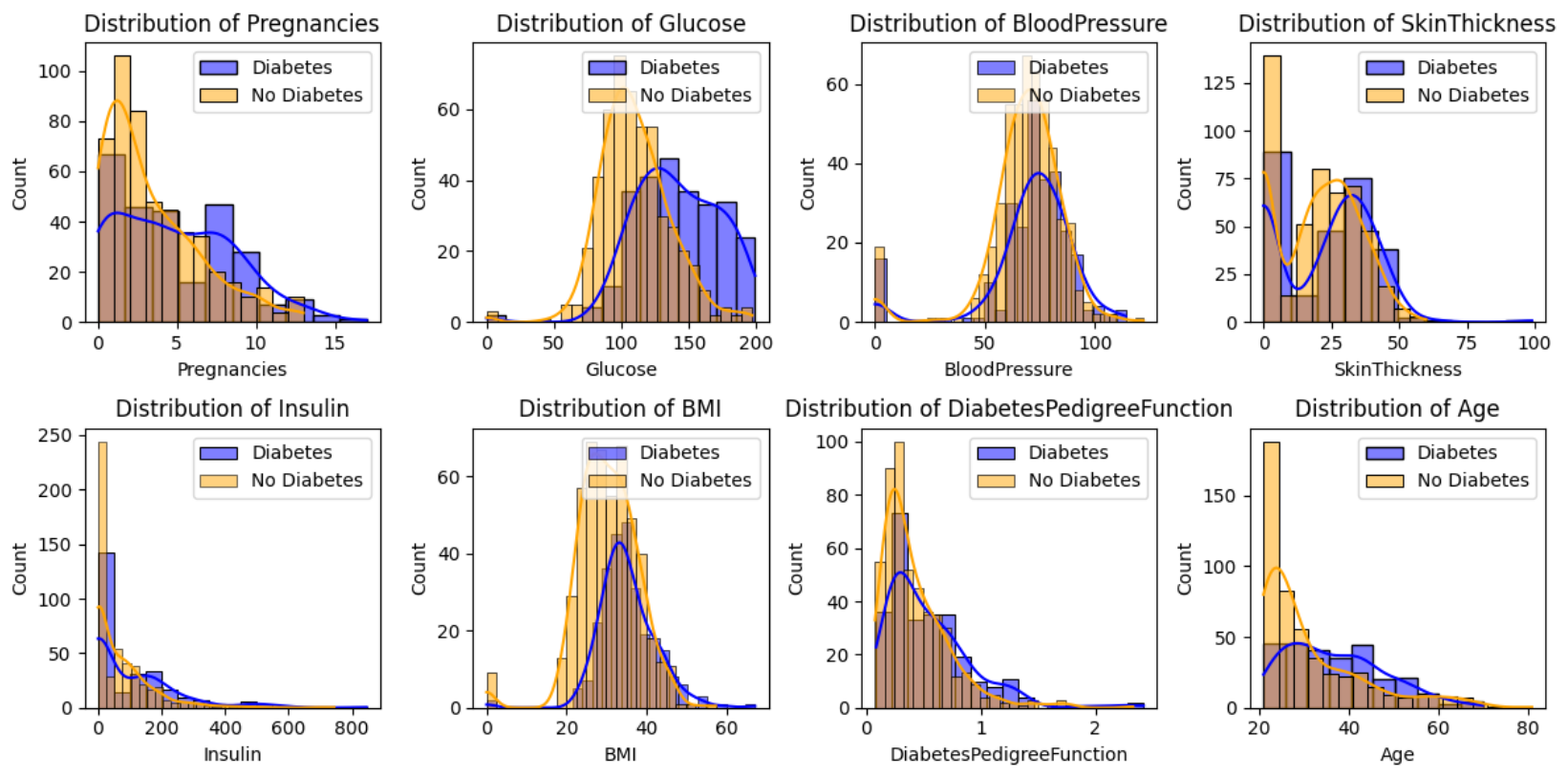
RELATIONSHIP BETWEEN PREDICTOR VARIABLES AND THE OUTCOME VARIABLE



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

MODEL BUILDING

MODEL COMPARISON

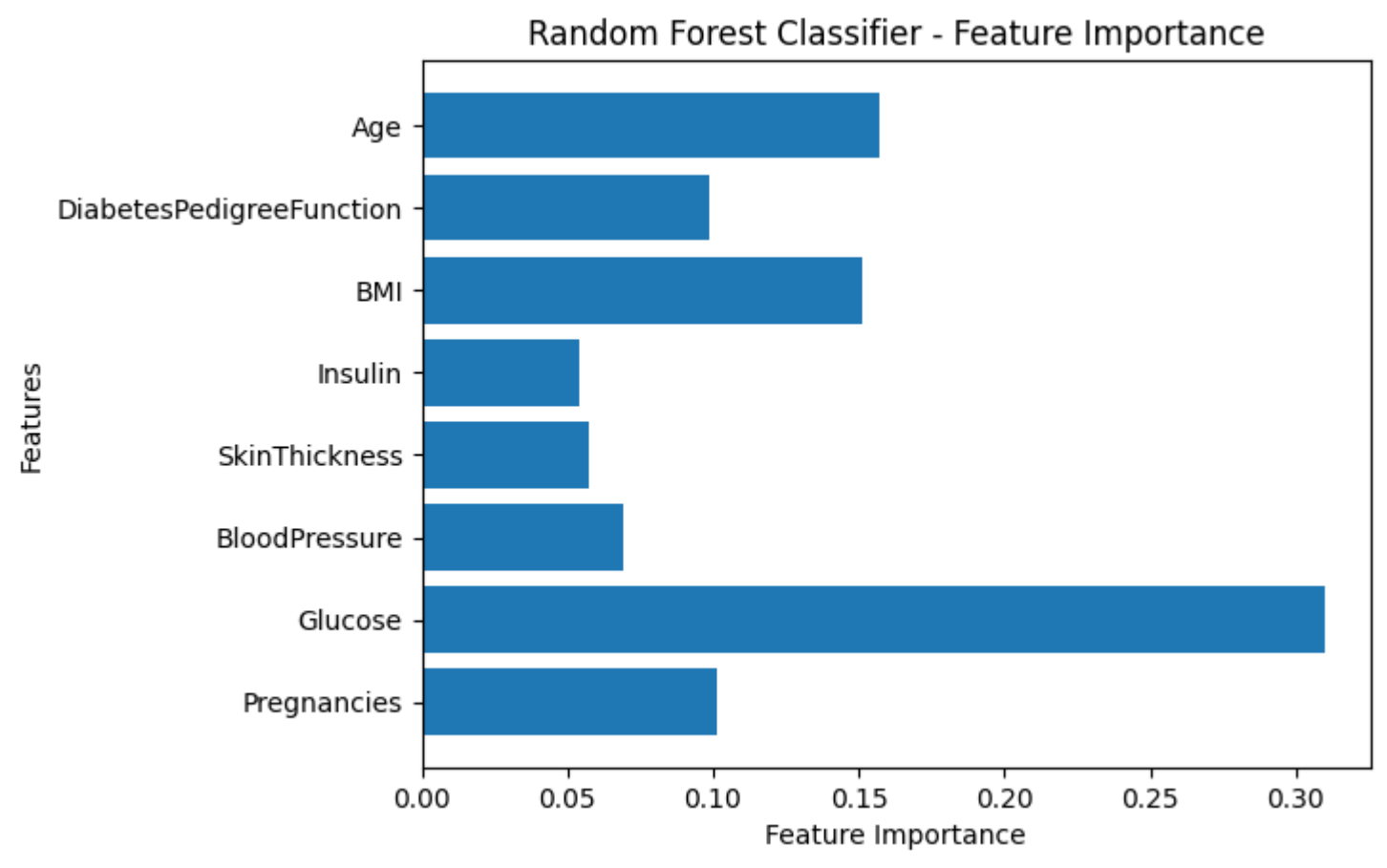
K NEIGHBORS CLASSIFIER

```
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



THANK YOU