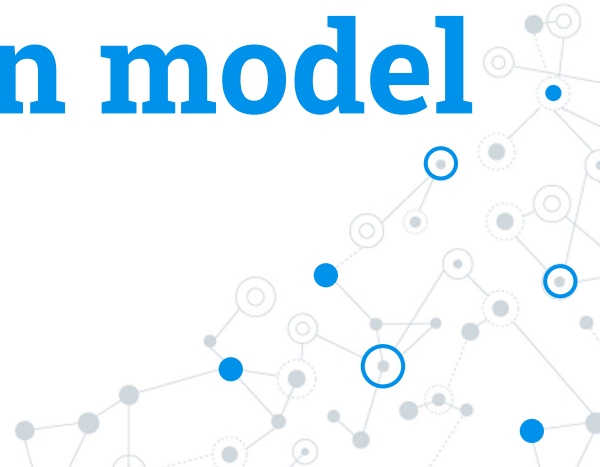
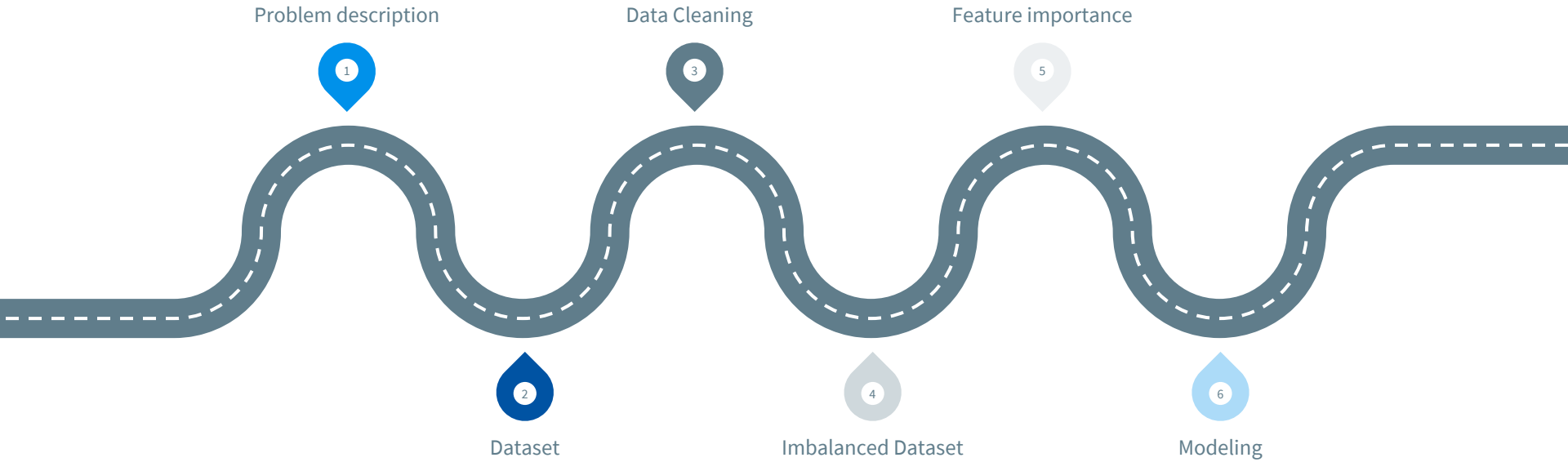


**Predict customer chur  
using classification model**



# Roadmap





# Problem description

Customer churn is one of the most vital data points for businesses to track.

Customer churn analysis helps you identify key stages in the customer journey where people are falling off, allowing you to pinpoint specific strategies to improve their interactions with your brand and improve brand loyalty.



# Dataset kaggle

7043

Rows

21

Columns





# Dataset

**Train set: 55%**

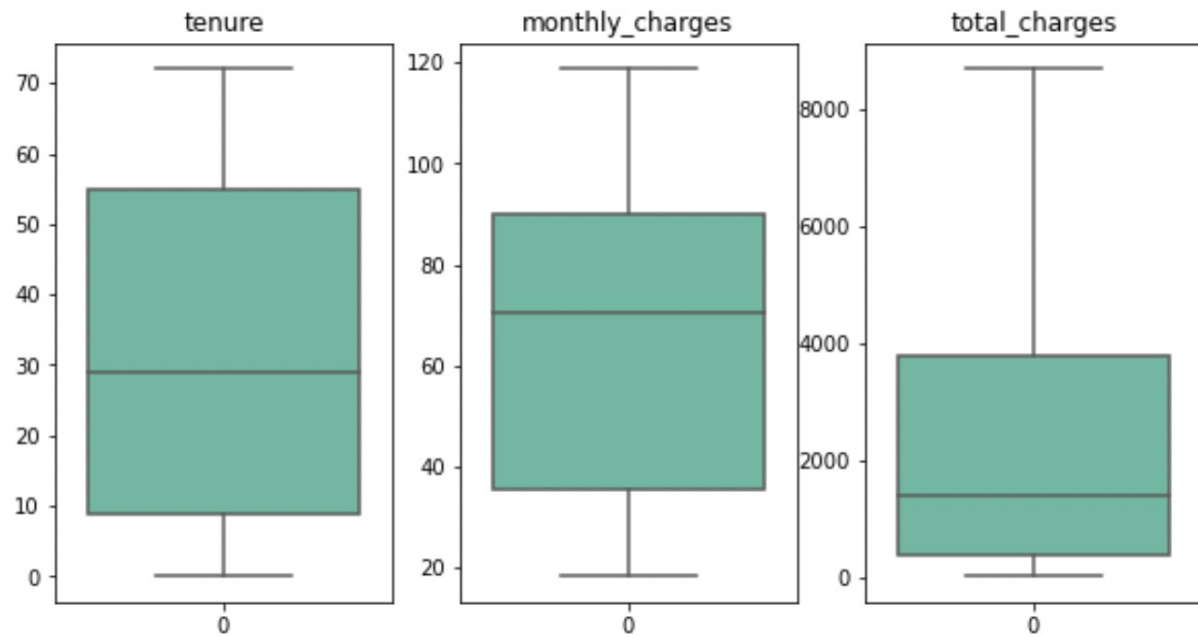
**Validation set: 25%**

**Test set: 20%**

A decorative network diagram in the top-left corner, featuring a complex web of interconnected nodes and lines. The nodes are represented by small circles, some of which are larger and have concentric circles, suggesting different levels of connectivity or importance. The lines are thin and gray, creating a mesh-like structure.

# 1. Data Cleaning

## 1. Outliers



## 2. Null Values

1. Split the dataframe to train and validation
2. Assign mean to null values in the train

## 3. Drop duplicate rows

## 4. Change columns data type

## 5. Drop unique columns



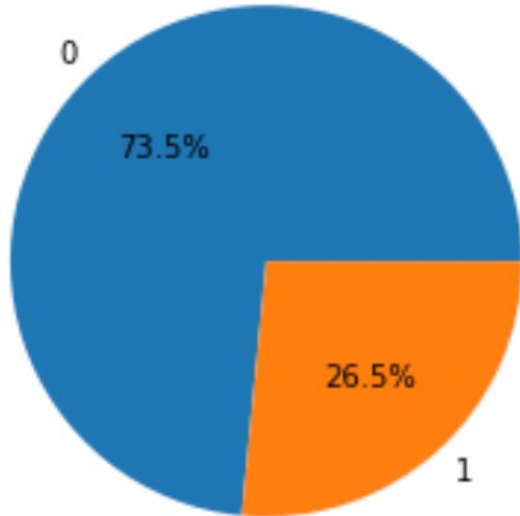


A decorative network diagram in the top-left corner, featuring a complex web of interconnected nodes and lines. The nodes are represented by circles of varying sizes, some with concentric rings, and the lines are thin and gray. The diagram is partially cut off by the top and left edges of the slide.

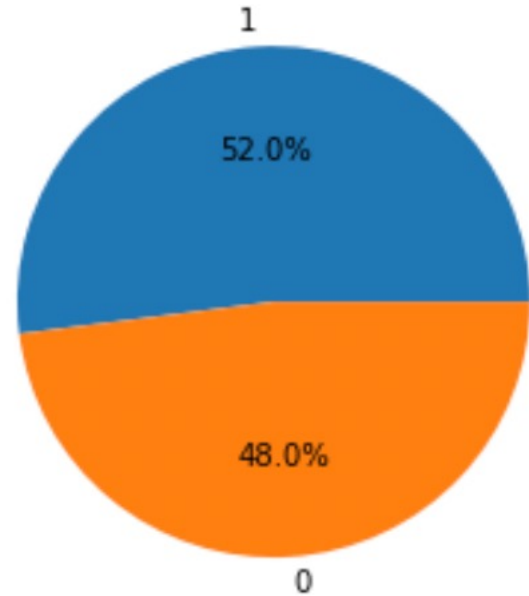
## 2. Imbalanced Dataset

## Over-sampling (by Random over sampling)

Distribution of Churning Rate



Distribution of Churning Rate

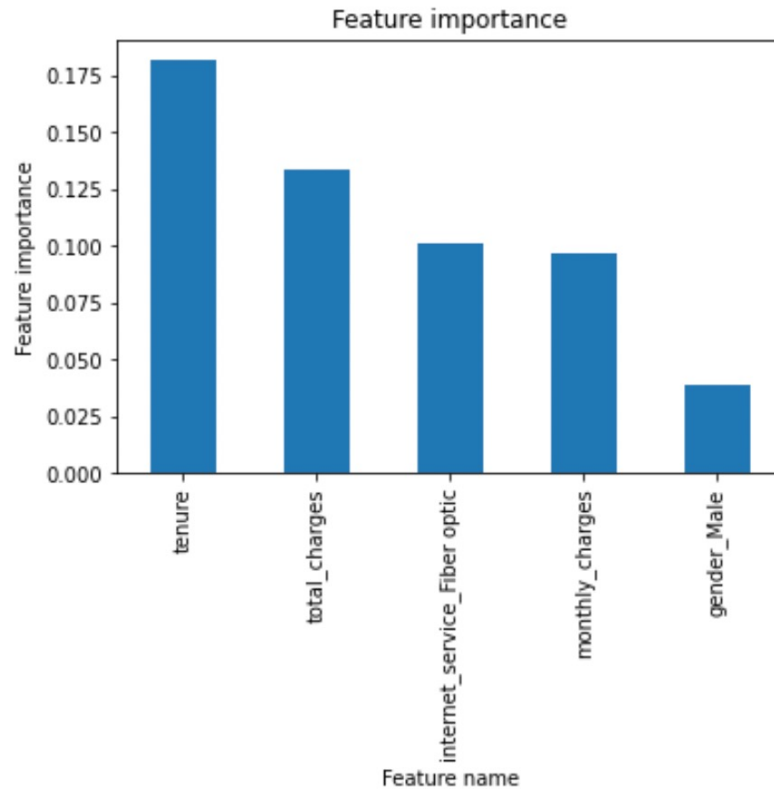




# 3. **Feature Importance**



# Feature Importance





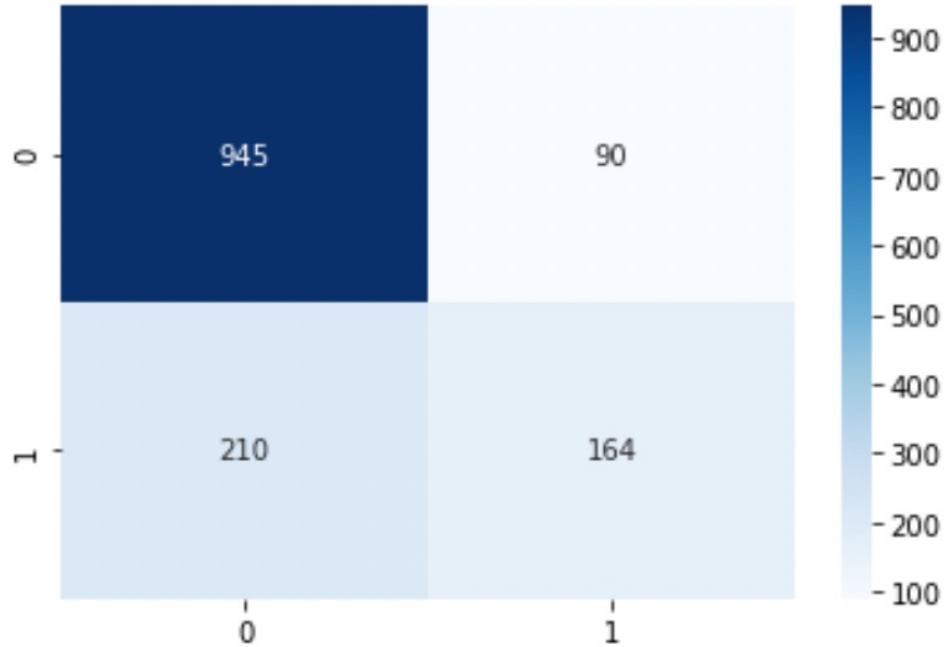
# 4. Modeling

## 1. Logistic Regression (baseline)

Accuracy: 79%

Recall: 43%

Precision: 64%

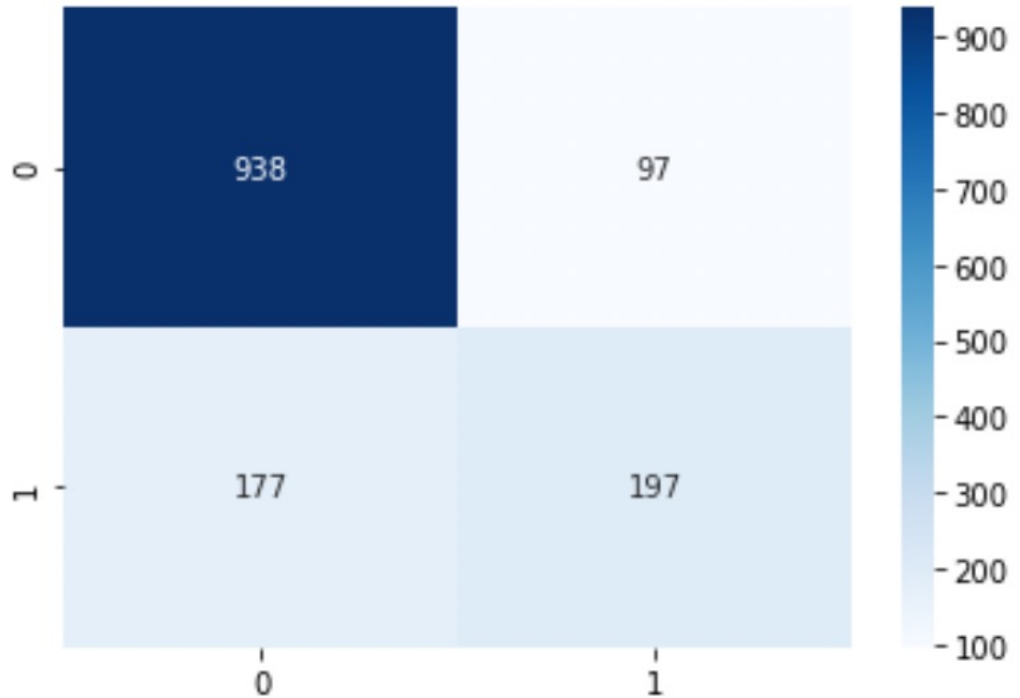


## 1. Logistic Regression (After Feature engineering)

Accuracy: 81%

Recall: 52%

Precision: 67%

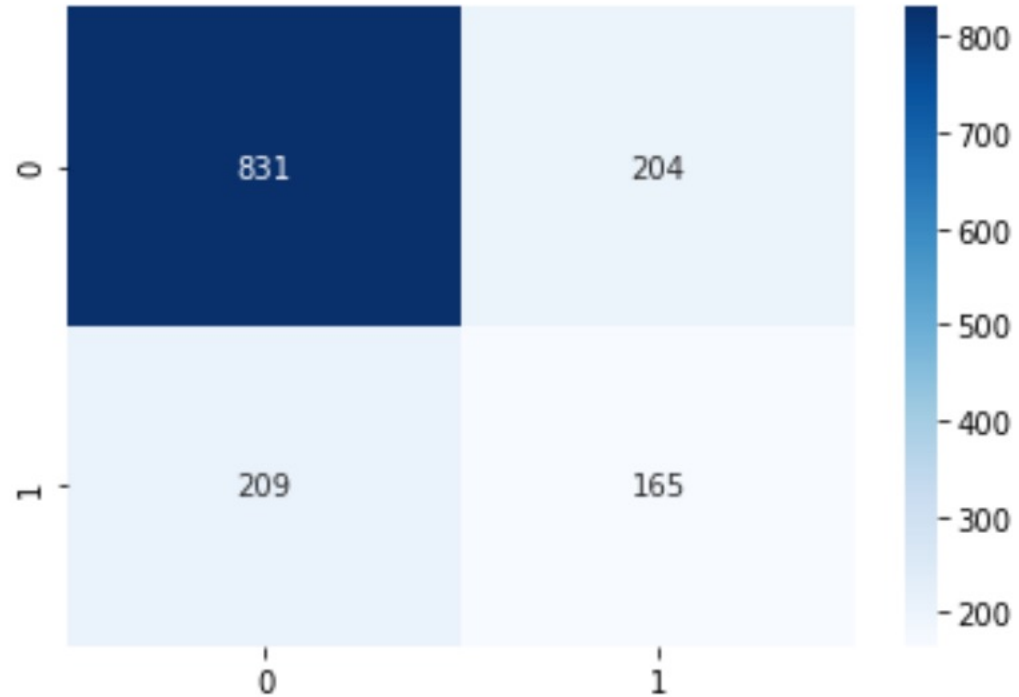


## 2. k Nearest Neighbors

Accuracy: 71%

Recall: 44%

Precision: 44%



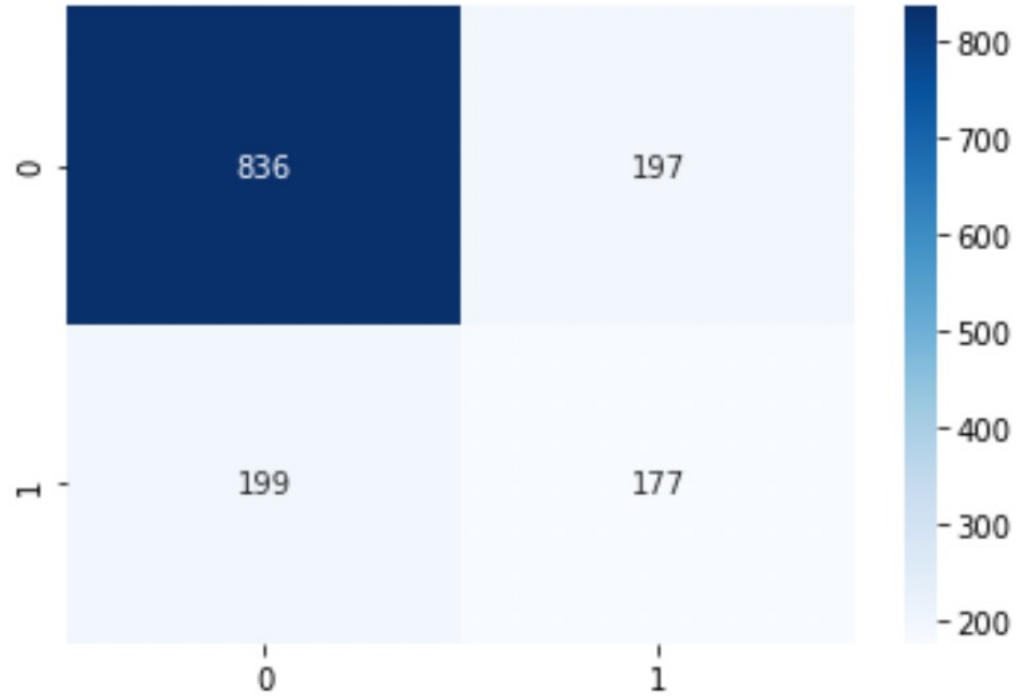


### 3. Decision Tree

Accuracy: 71%

Recall: 49%

Precision: 48%

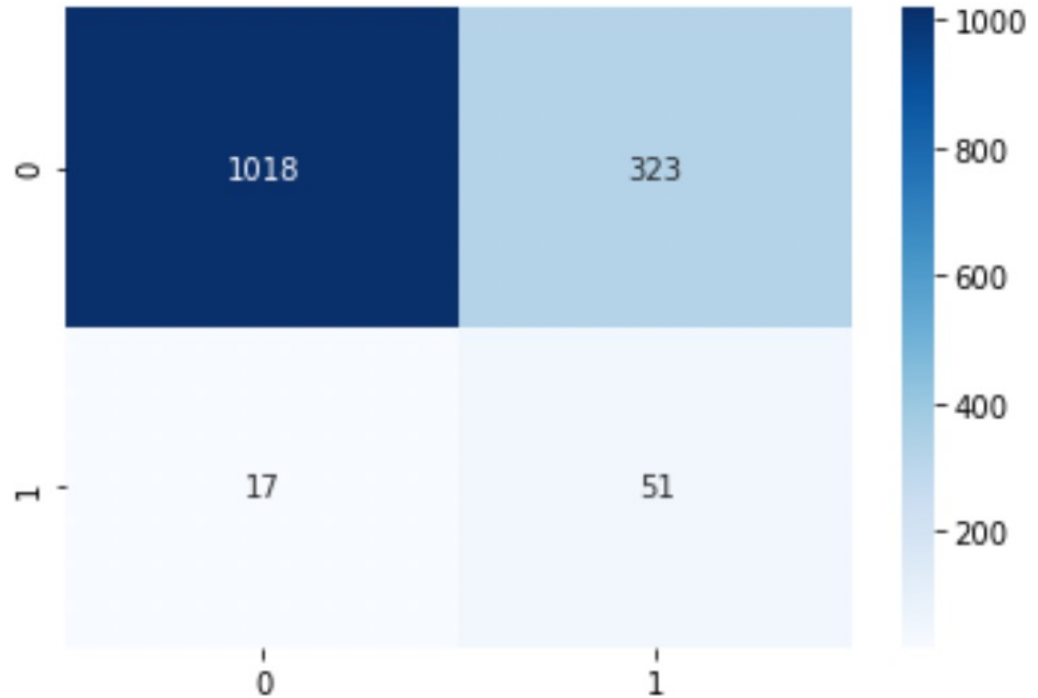


## 4. Random Forest

Accuracy: 81%

Recall: 13%

Precision: 75%

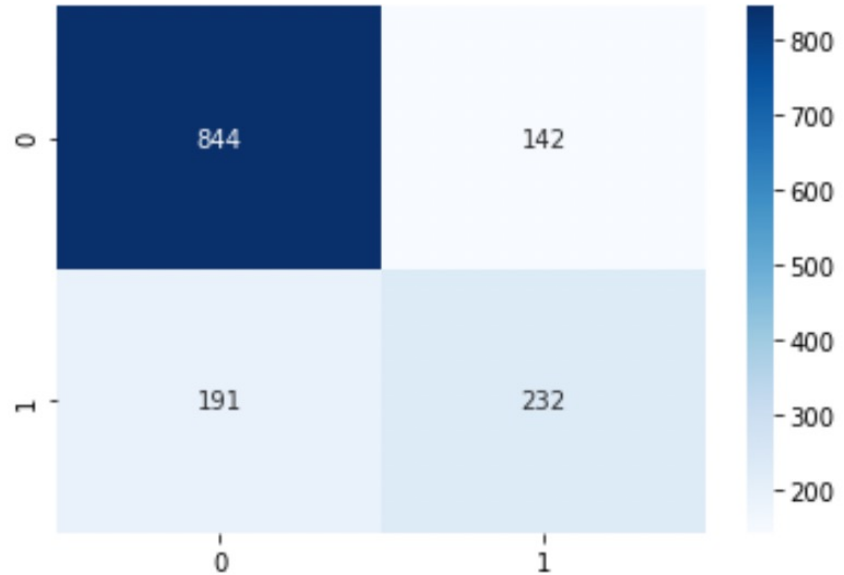


## 5. XGBoost (The best)

Accuracy: 81%

Recall: 62%

Precision: 54%



The background of the slide is a light gray network pattern. It consists of numerous small circles, some of which are solid gray and others are hollow with a gray outline. These circles are interconnected by a web of thin, light gray lines, creating a complex, organic-looking structure that fills the entire frame.

# **Thanks!**

## **Any questions?**