

Identifying Customer Churn

Objective

This project will create a binary classification model to determine which customers are likely to leave the telecom company.

Once customers are identified, the customer service team will contact them proactively to determine why they might leave. Resulting strategies to lower churn could include addressing individual customer concerns by customer service or general engineering improvements or sales policy changes.

To identify general improvements, we do not need to identify every churn instance, only enough to get sufficient customer feedback.

The Dataset

State

Account Length

Area Code

Phone Number

International Plan

Voicemail Plan

Number VM Messages

Total Day Minutes

Total Day Calls

Total Day Charge

Total Eve Minutes

Total Eve Calls

Total Eve Charge

Total Night Minutes

Total Night Calls

Total Night Charge

Total Intl Minutes

Total Intl Calls

Total Intl Charge

Customer Service Calls

Final Feature Set Importance

Total Day Minutes	0.18	Total Eve Calls	0.05
Customer Service Calls	0.14	Total Day Calls	0.05
Total Eve Minutes	0.07	Total Night Calls	0.05
Total Intl Calls	0.07	Number VM Messages	0.03
Total Night Minutes	0.06	Area Code	0.05
Total Intl Minutes	0.06	Intl Plan	0.02
Account Length	0.05	VM Plan	0.02

Evaluation Criteria

F1 Score – a measure of both:

precision (true positives/predicted positives)

recall (predicted true positives/actual true positives)

Accuracy – the total number of predictions the model gets correct

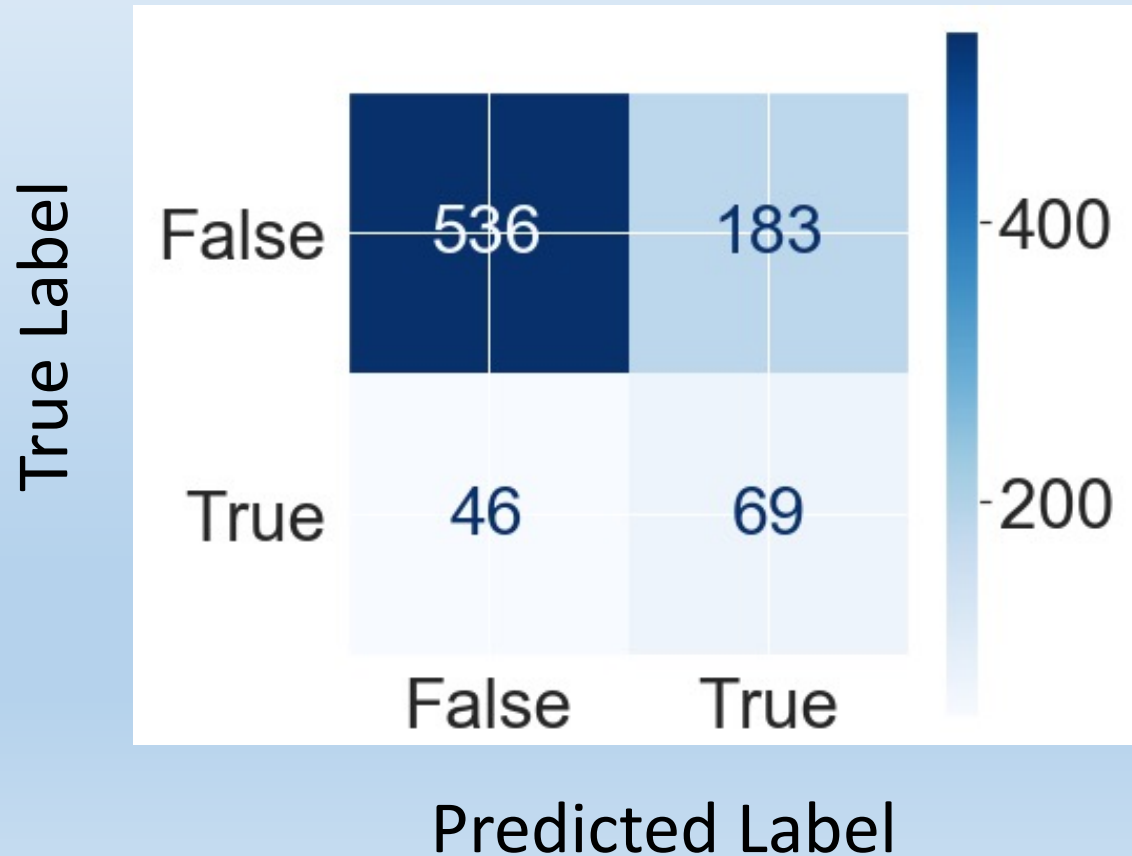
Confusion Matrix – shows True Negatives, False Positives

False Negatives, True Positives

Logistic Regression

F1 Score 0.37

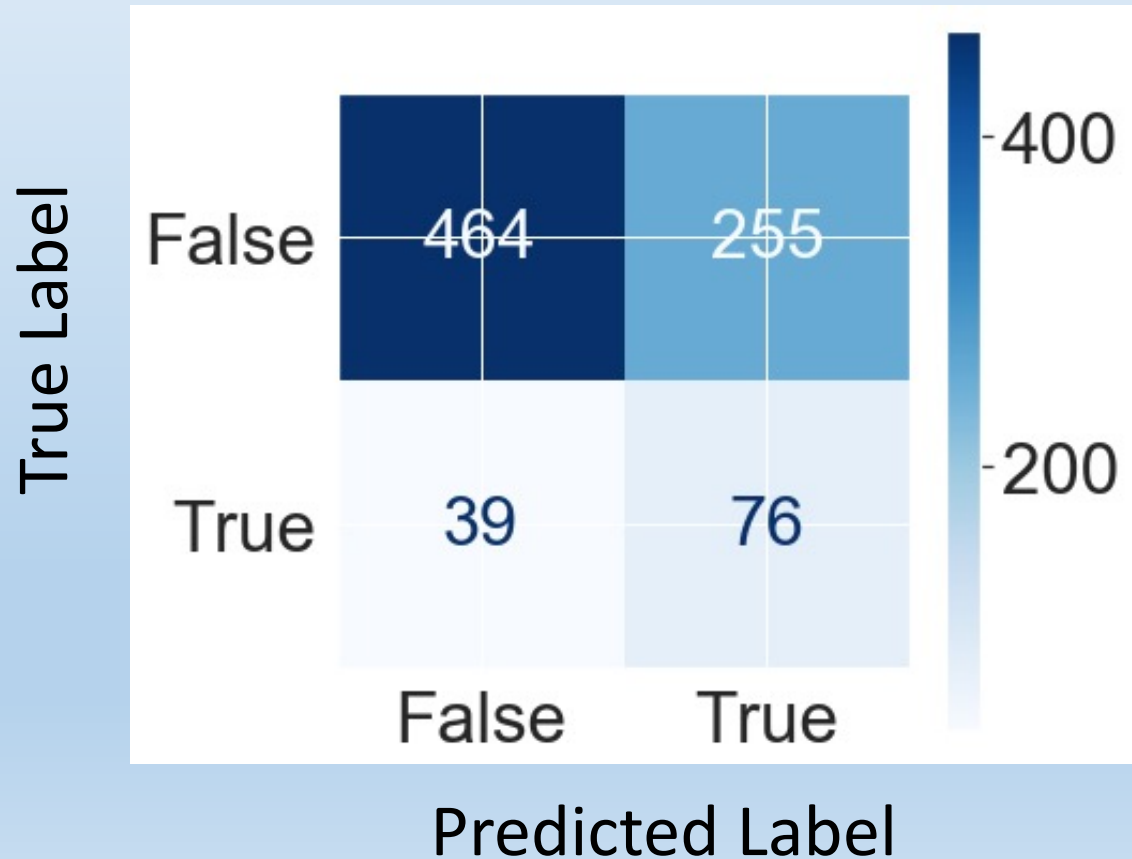
Accuracy 0.71



GaussianNB

F1 Score 0.35

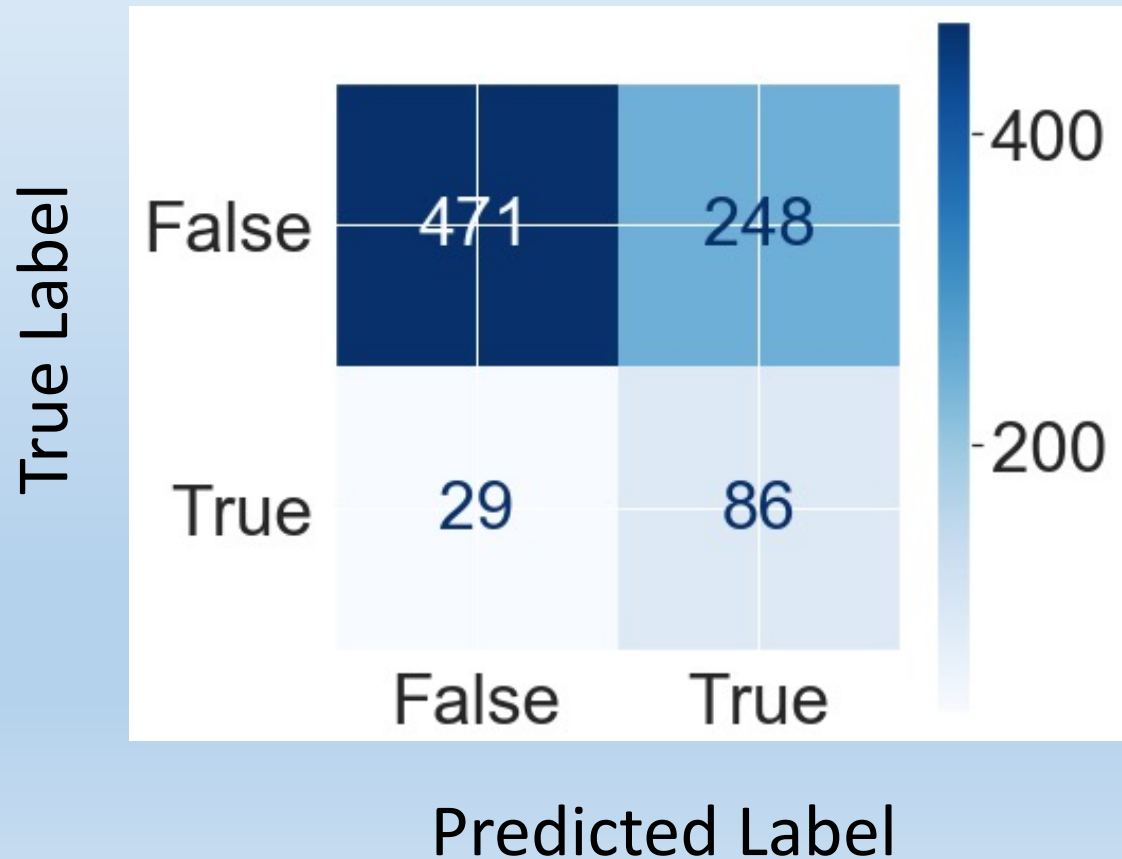
Accuracy 0.66



K Nearest Neighbors

F1 Score 0.40

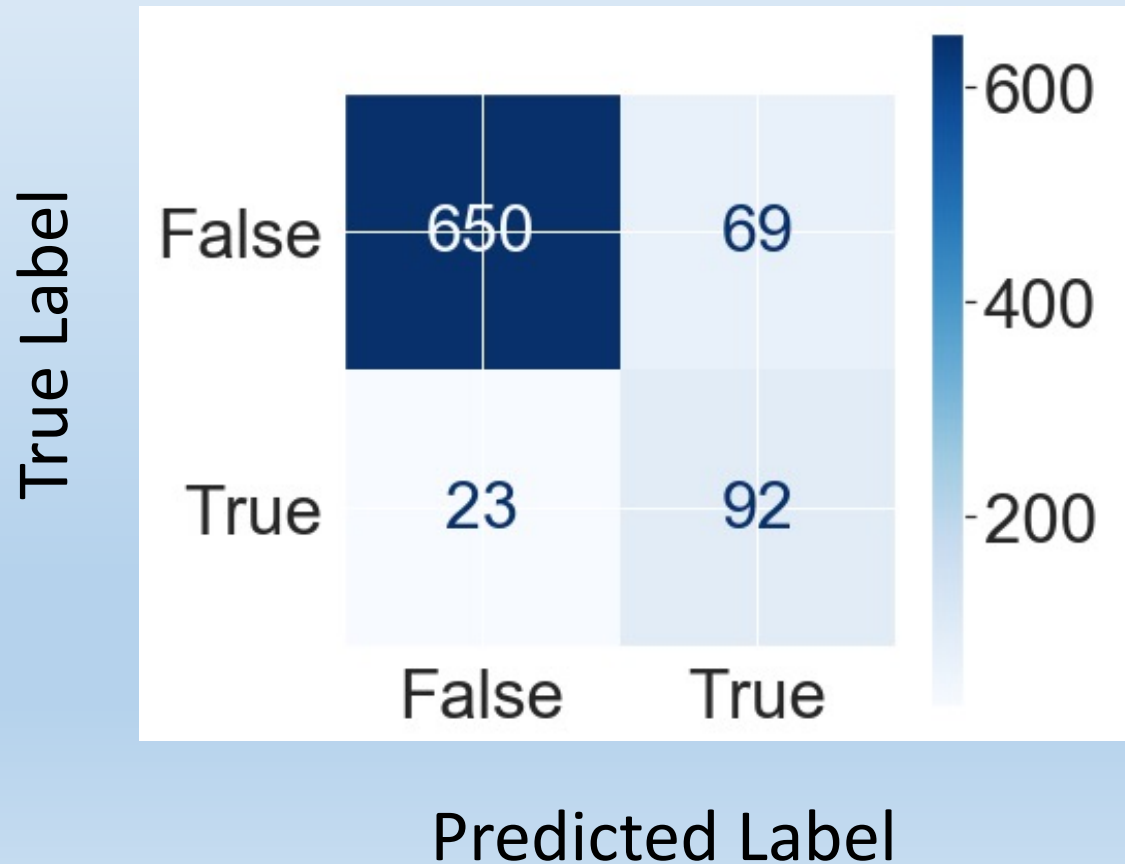
Accuracy 0.70



Gradient Boost

F1 Score 0.63

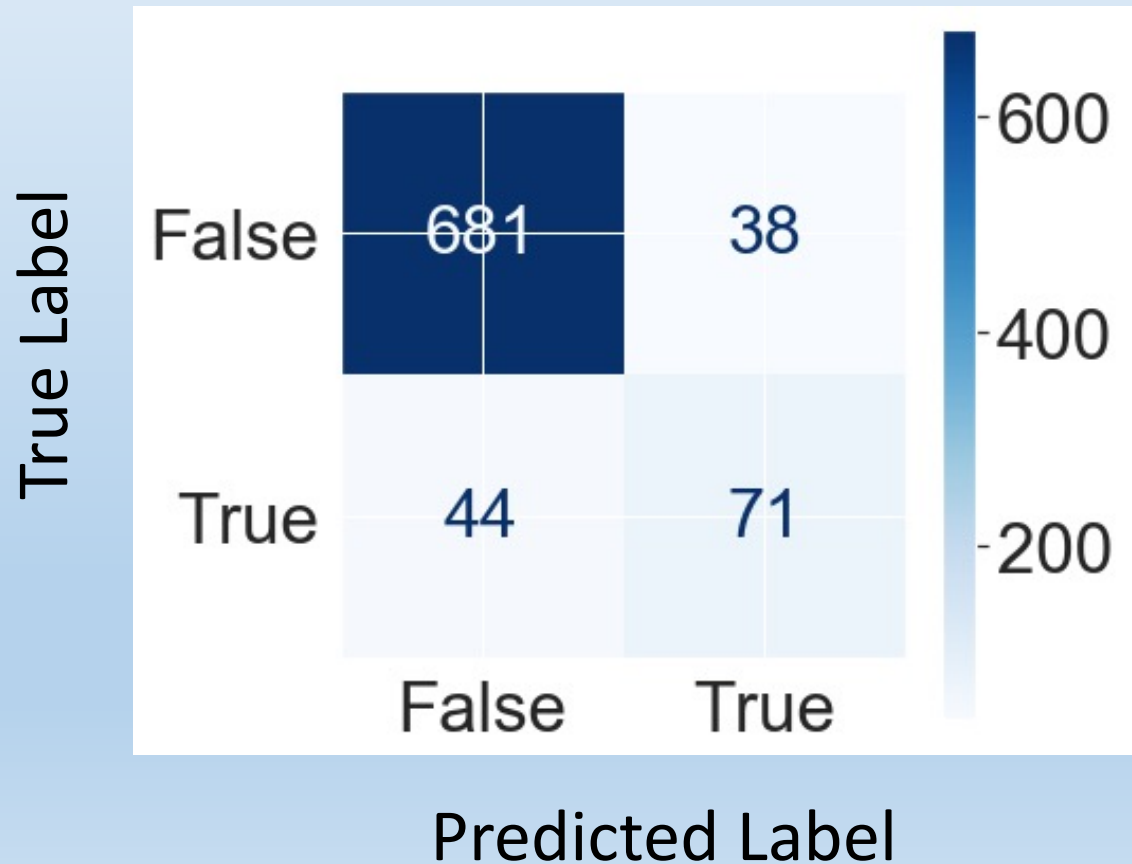
Accuracy 0.88



Random Forest

F1 Score 0.61

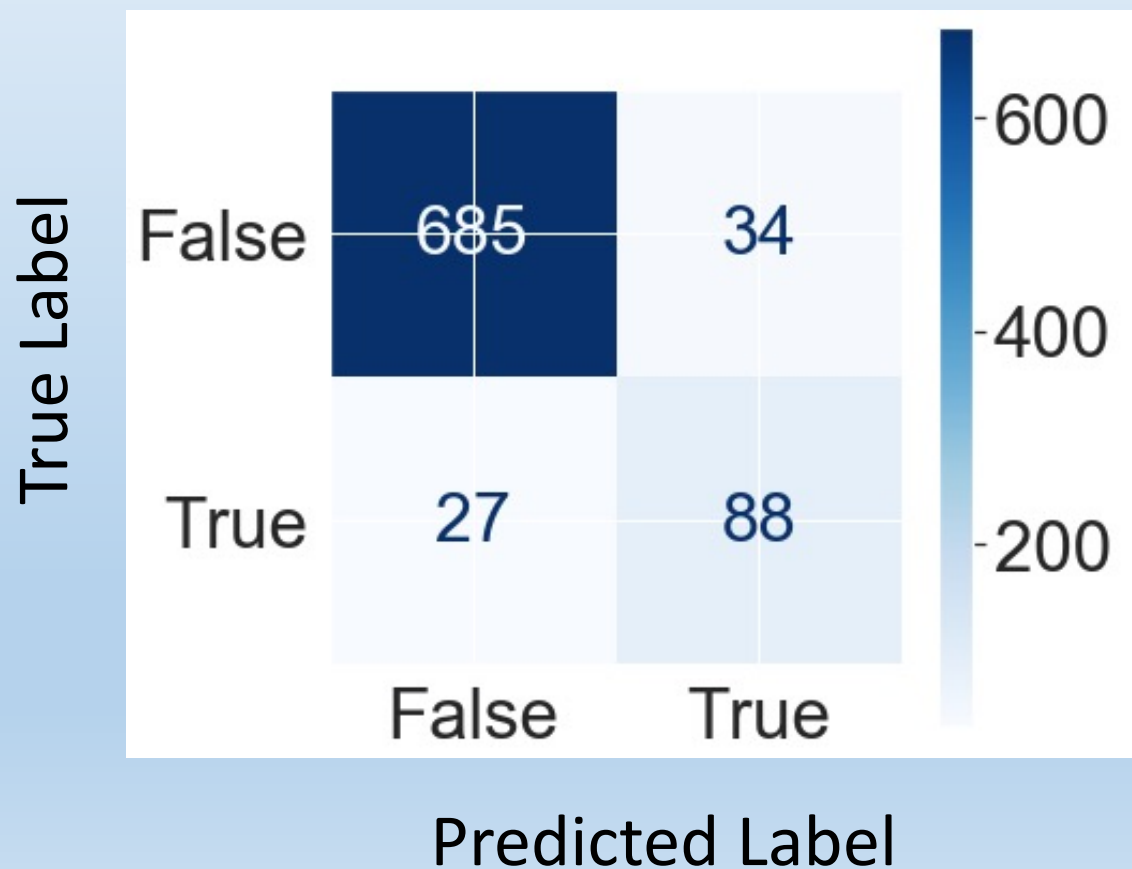
Accuracy 0.90



XGBoost

F1 Score 0.75

Accuracy 0.93



Conclusion

The XGBoost model gives the best predictive model, with an F1 Score of 0.75 and an accuracy score of 0.93.

The model gives more false positives than negatives, which is preferred as it emphasizes capturing as much churn as possible.

The two most important features for predicting churn are the total number of day minutes used and the number of customer service calls.