PREDICTING CUSTOMER CHURN

12 AUGUST 2024

VORTEX

CALIFORNIA MARKET

VINAY BEESA GNANESHWAR

“A 5% INCREASE IN CUSTOMER LOYALTY CAN PRODUCE PROFIT INCREASES FROM 25% TO 85%”

REICHHELD & SASSER (H.B.R)

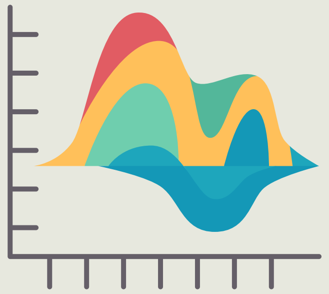
PROBLEM

26% Churn Rate

VORTEX

VORTEX

DATASET



- Monthly Charges
- Age
- Tenure in Months
- Total Charges.

- Dependents
- Referred a Friend
- Contract type
- Payment Method.

7043	60	4	CHURN
CUSTOMERS	UNIQUE VARIABLES	MODELS	TARGET

VORTEX

DATA CLEANING

Column names in each DataFrame:

	df1	df2	df3	df4	df5
0	Customer_ID	LoyaltyID_df2	Customer_ID	Service ID_df4	Status ID_df5
1	Count_df1	Customer_ID	Count_df3	Customer_ID	Customer_ID
2	Country_df1	Senior Citizen_df2	Gender_df3	Count_df4	Count_df5
3	State_df1	Partner_df2	Age_df3	Quarter_df4	Quarter_df5
4	City_df1	Dependents_df2	Under 30_df3	Referred a Friend_df4	Satisfaction Score_df5
5	Zip Code_df1	Tenure_df2	Senior Citizen_df3	Number of Referrals_df4	Customer Status_df5
6	Lat Long_df1	Phone Service_df2	Married_df3	Tenure in Months_df4	Churn Label_df5
7	Latitude_df1	Multiple Lines_df2	Dependents_df3	Offer_df4	Churn Value_df5
8	Longitude_df1	Internet Service_df2	Number of Dependents_df3	Phone Service_df4	Churn Score_df5
9	Gender_df1	Online Security_df2	nan	Avg Monthly Long Distance Charges_df4	CLTV_df5
10	Senior Citizen_df1	Online Backup_df2	nan	Multiple Lines_df4	Churn Category_df5
11	Partner_df1	Device Protection_df2	nan	Internet Service_df4	Churn Reason_df5
12	Dependents_df1	Tech Support_df2	nan	Internet Type_df4	nan
13	Tenure Months_df1	Streaming TV_df2	nan	Avg Monthly GB Download_df4	nan
14	Phone Service_df1	Streaming Movies_df2	nan	Online Security_df4	nan
15	Multiple Lines_df1	Contract_df2	nan	Online Backup_df4	nan
16	Internet Service_df1	Paperless Billing_df2	nan	Device Protection Plan_df4	nan
17	Online Security_df1	Payment Method_df2	nan	Premium Tech Support_df4	nan
18	Online Backup_df1	Monthly Charges_df2	nan	Streaming TV_df4	nan
19	Device Protection_df1	Total Charges_df2	nan	Streaming Movies_df4	nan
20	Tech Support_df1	Churn_df2	nan	Streaming Music_df4	nan
21	Streaming TV_df1	nan	nan	Unlimited Data_df4	nan
22	Streaming Movies_df1	nan	nan	Contract_df4	nan
23	Contract_df1	nan	nan	Paperless Billing_df4	nan
24	Paperless Billing_df1	nan	nan	Payment Method_df4	nan
25	Payment Method_df1	nan	nan	Monthly Charge_df4	nan
26	Monthly Charges_df1	nan	nan	Total Charges_df4	nan
27	Total Charges_df1	nan	nan	Total Refunds_df4	nan
28	Churn Label_df1	nan	nan	Total Extra Data Charges_df4	nan
29	Churn Value_df1	nan	nan	Total Long Distance Charges_df4	nan
30	Churn Score_df1	nan	nan	Total Revenue_df4	nan
31	CLTV_df1	nan	nan	nan	nan
32	Churn Reason_df1	nan	nan	nan	nan

Column names in pre_dummy_df:

	pre_dummy_df
0	Monthly_Charges
1	Dependents
2	Churn
3	Age
4	Under_30
5	Number_of_Dependents
6	Referred_a_Friend
7	Number_of_Referrals
8	Tenure_in_Months
9	Offer
10	Avg_Monthly_Long_Distance_Charges
11	Multiple_Lines
12	Internet_Service
13	Internet_Type
14	Avg_Monthly_GB_Download
15	Online_Security
16	Online_Backup
17	Device_Protection_Plan
18	Premium_Tech_Support
19	Streaming_TV
20	Streaming_Movies
21	Streaming_Music
22	Unlimited_Data
23	Contract
24	Payment_Method
25	Total_Charges
26	Total_Refunds
27	Total_Extra_Data_Charges
28	Total_Long_Distance_Charges
29	Total_Revenue

VORTEX

DATA PREPROCESSING

→

→

DROPPING VARIABLES

- 102 to 30 variables
- Handled similar columns.
- Dropping all IDs.

STANDARDIZATION

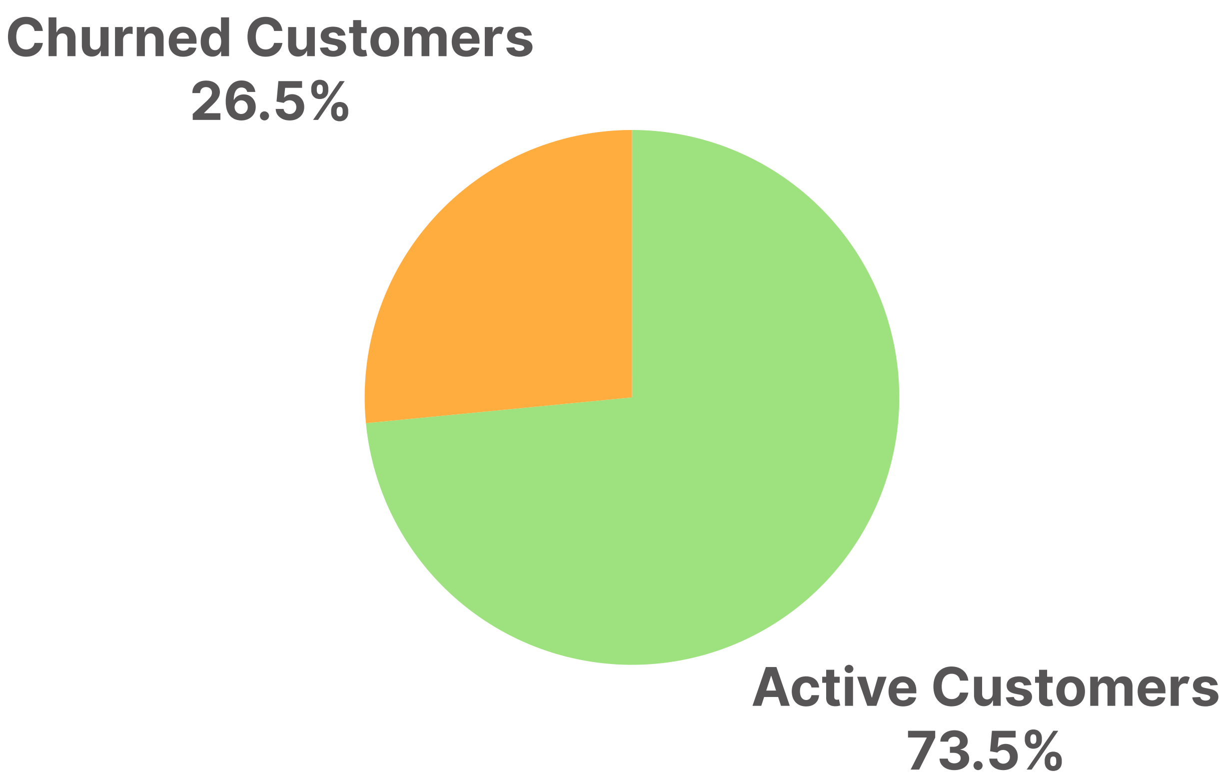
- Homogenization of nomenclature.
- Imputing missing values.
- Association of accurate data type.
- Outlier Detection.

PRE-MODELLING

- Dummy variables - binary.
- Creation of Train & Test dataset.

WHAT DOES THE DATA TELL US?

VORTEX



Churn vs Active

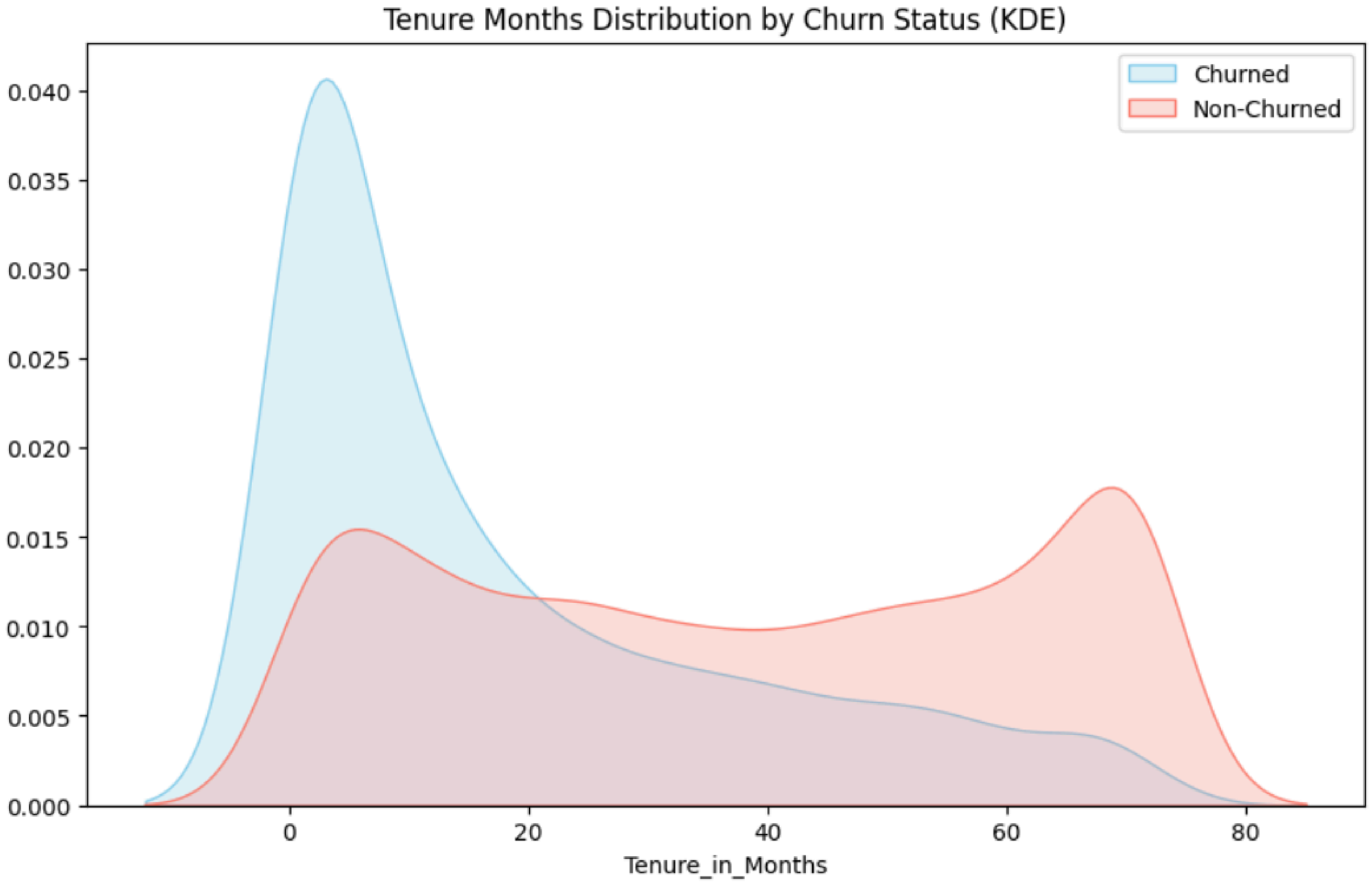
Active - 5174

Churned - 1869

Contract

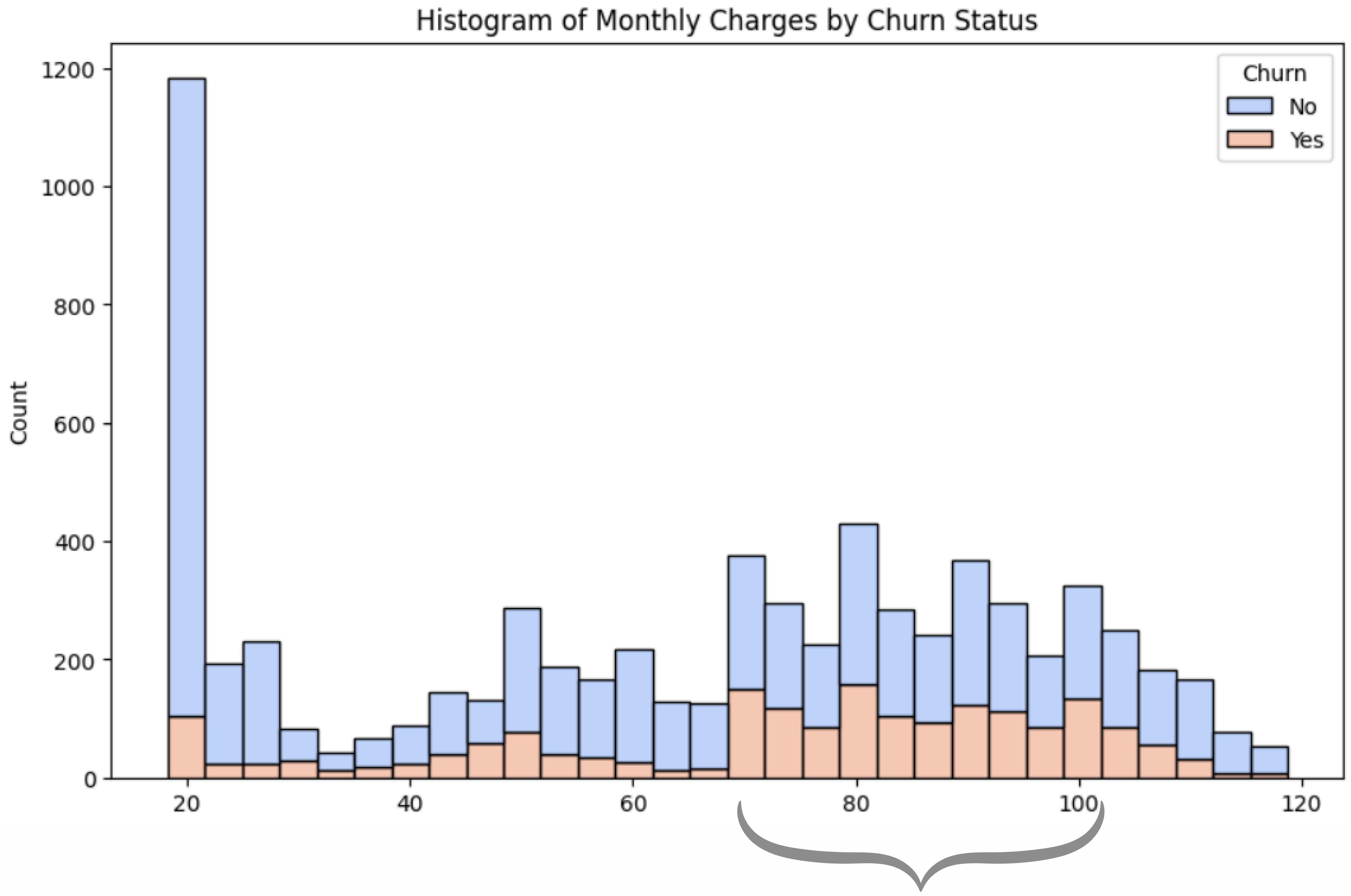
High Churns - 0 to 12 months

Low Churns - 12 onwards



VORTEX

CHURN BY CHARGES



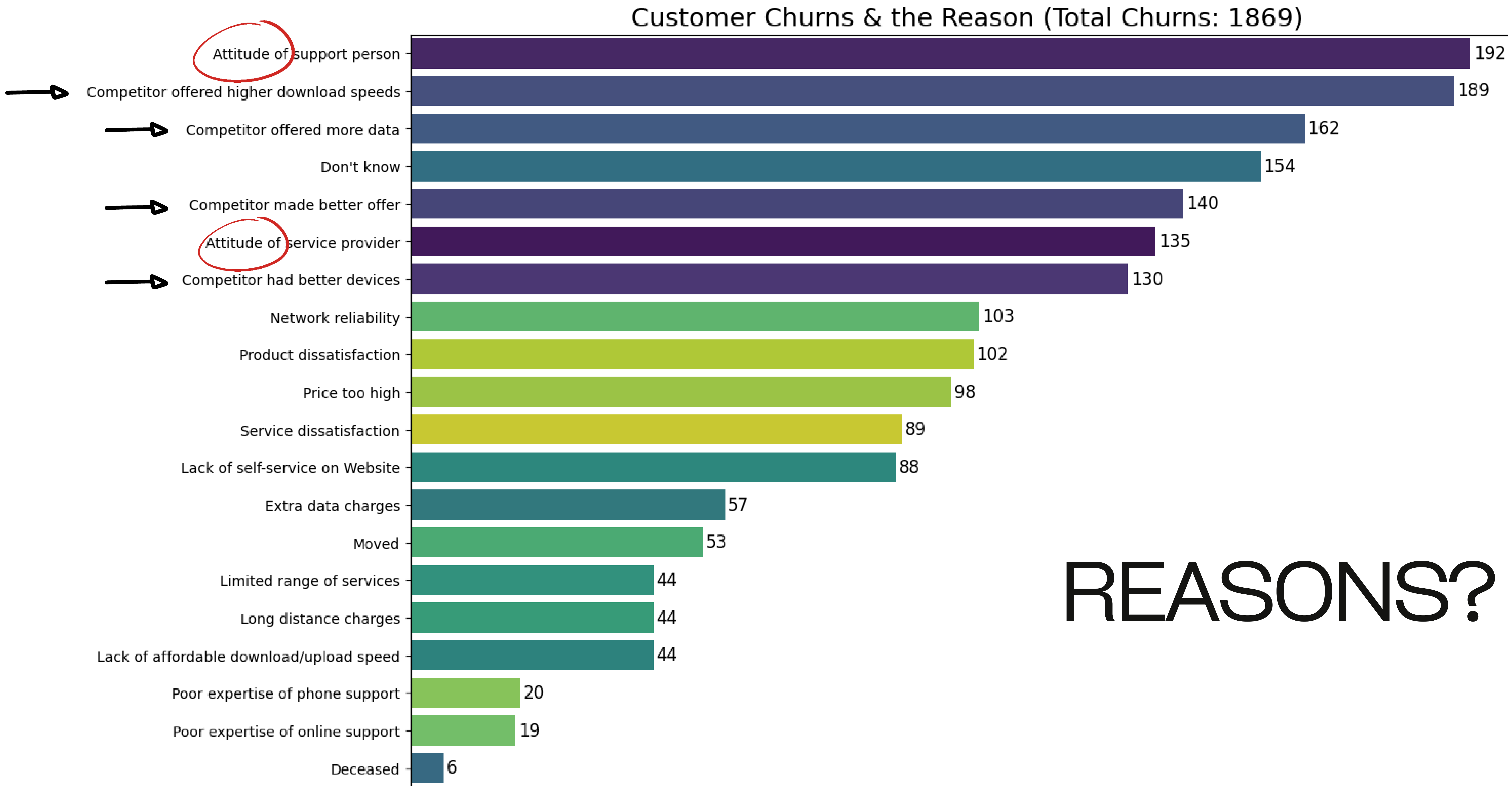
$20

Majority of 'No' Churns

$70 - $100

Majority of Churns

VORTEX



REASONS?

VORTEX

MODELS

LOGISTIC REGRESSION

Binary Predictor

K NEIGHBOUR CLASSIFIER

Neighbor Voting

NEURAL NETWORK

Deep Learning

RANDOM FOREST

Tree Ensemble

RANDOM FOREST

01

Ensemble Learning Method

02

Prediction & Accuracy

03

Robustness to Noise

VORTEX

WHAT DOES IT DO?

Bootstrapping

Growing Decision Trees

Predictions

Aggregating Results

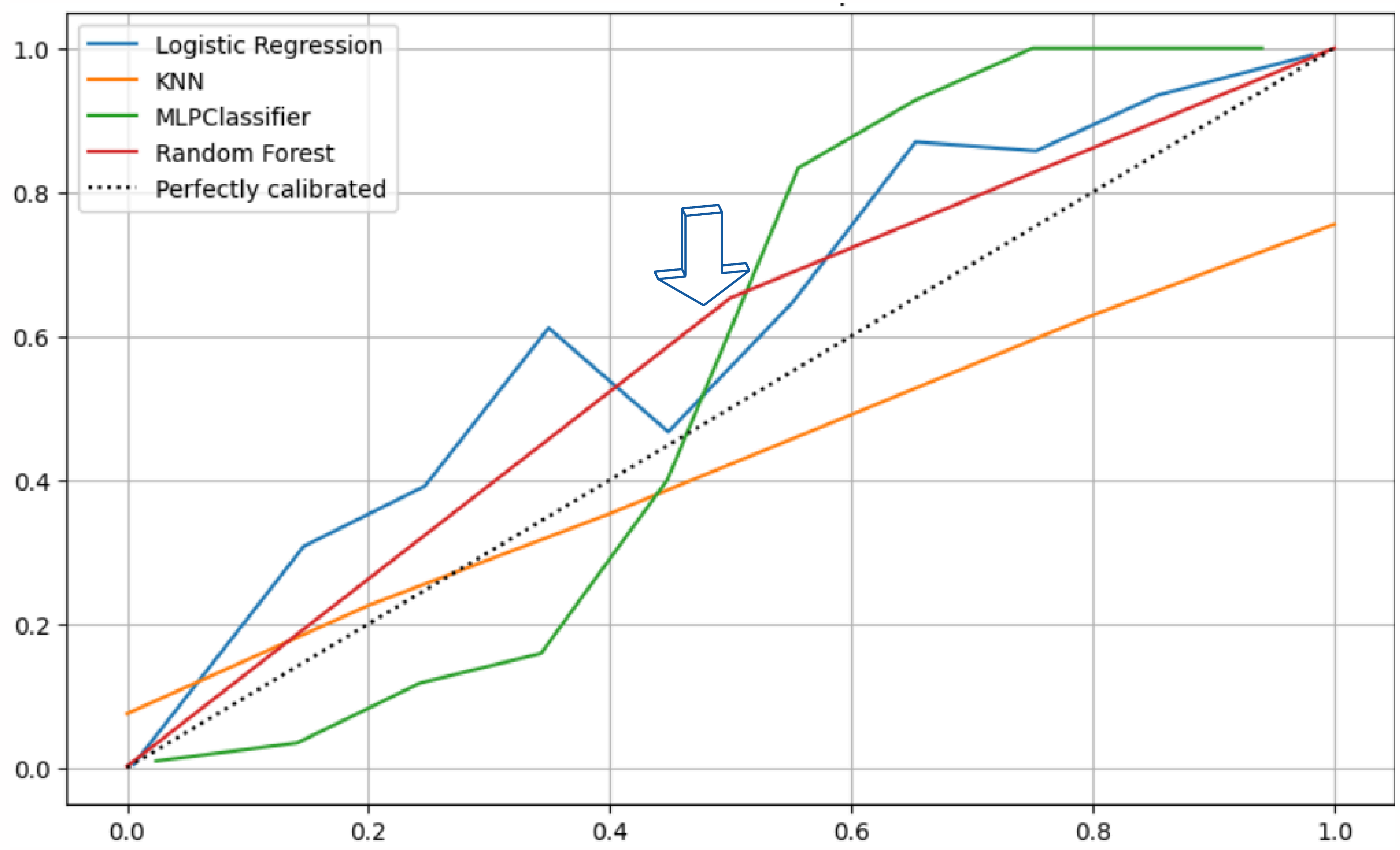
97%

Predicting Churn

98%

Accuracy

VORTEX



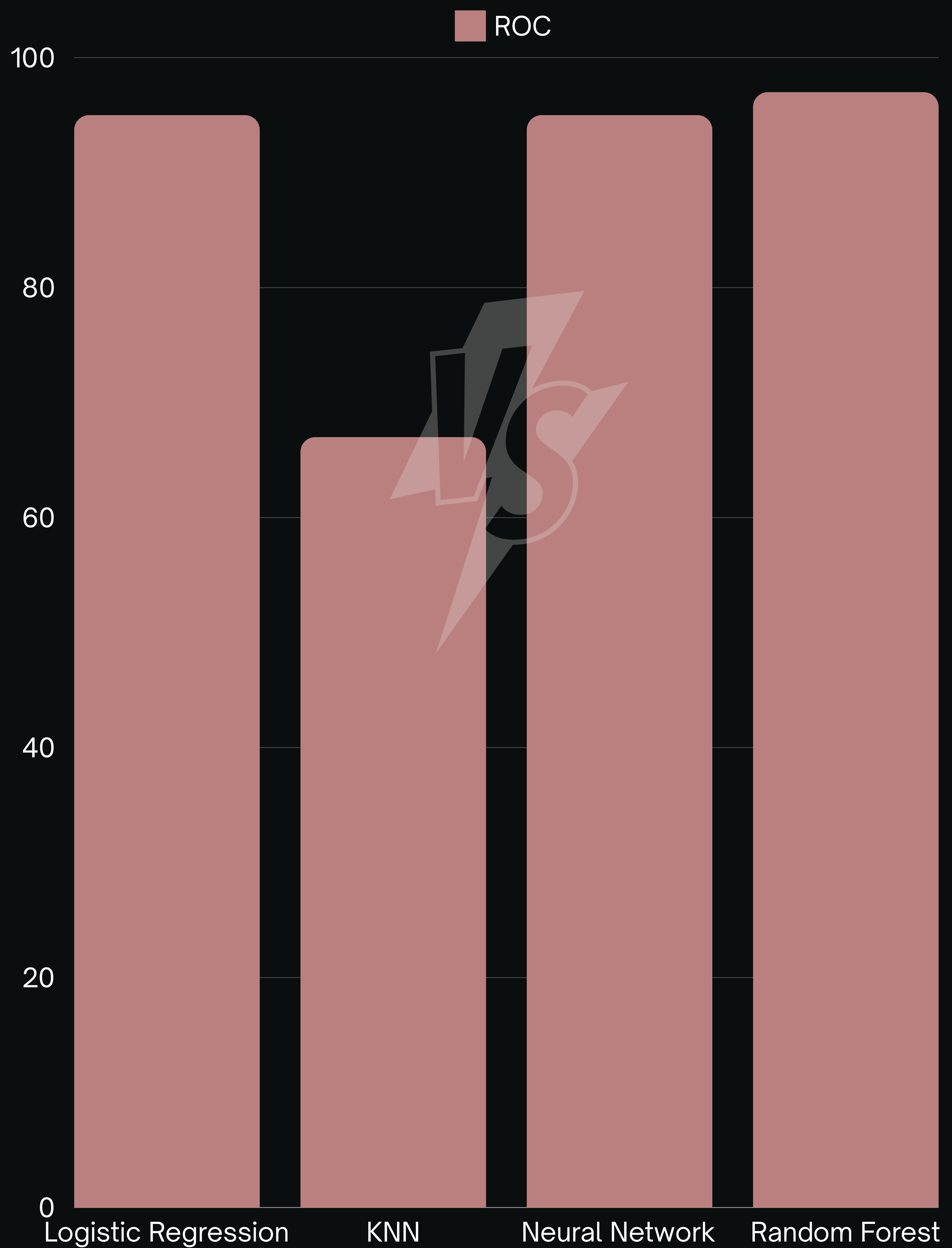
WHY?

Confidence & Accuracy

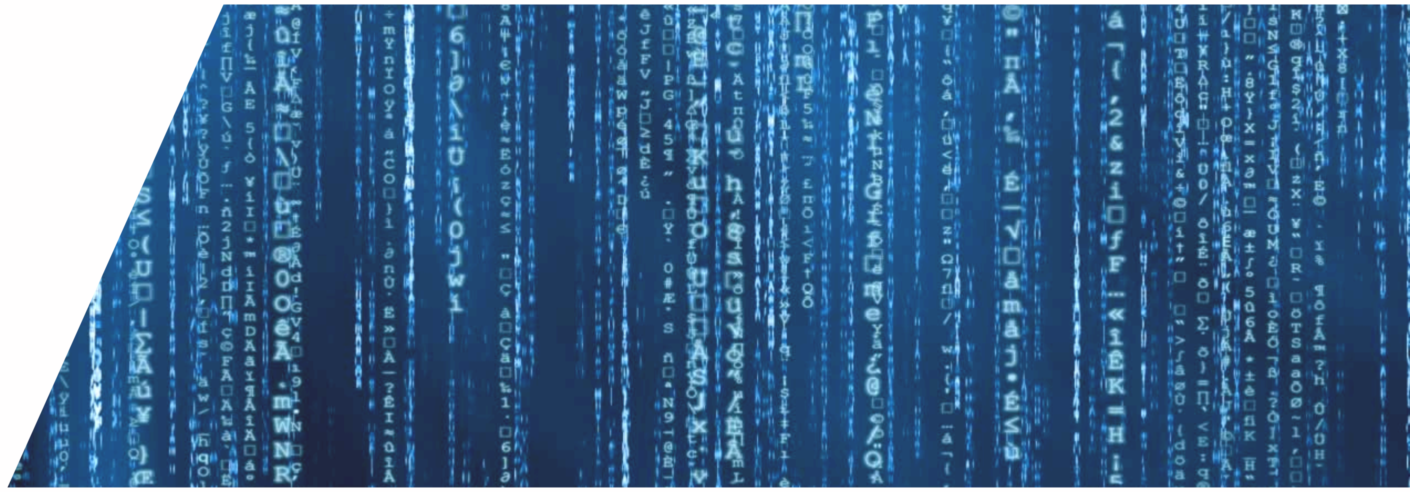
Handles Overfitting

VORTEX

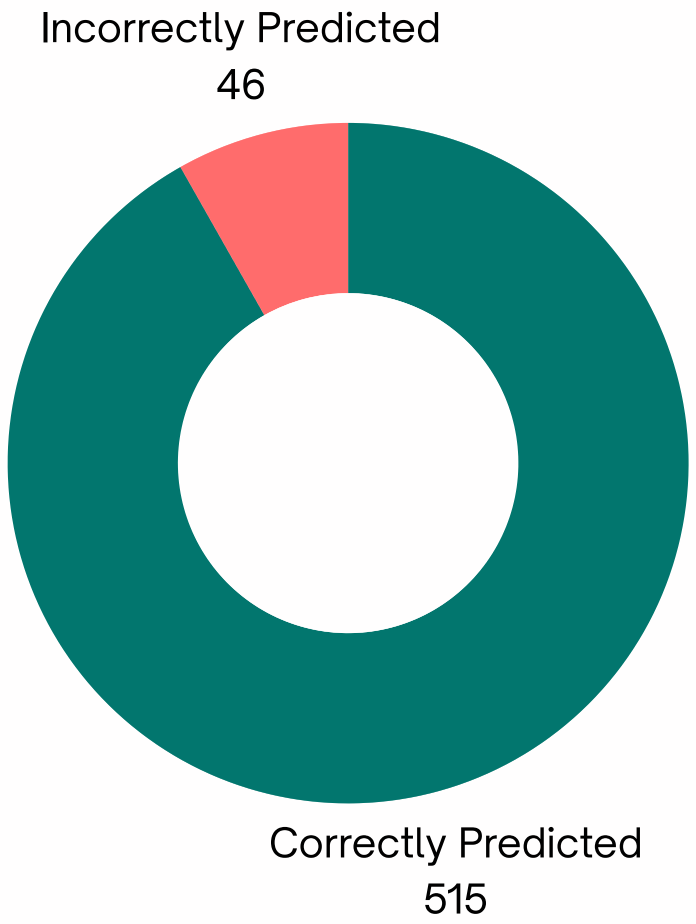
HOW DO THEY COMPARE?



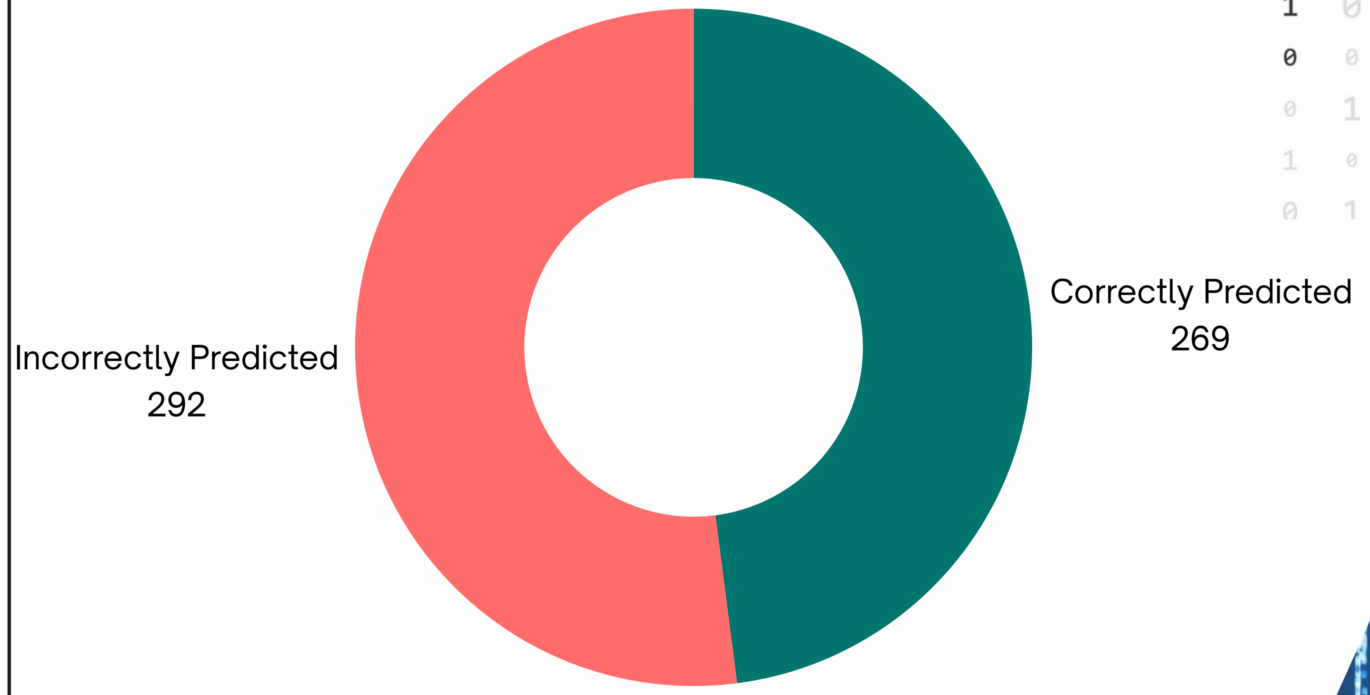
CONFUSION MATRIX



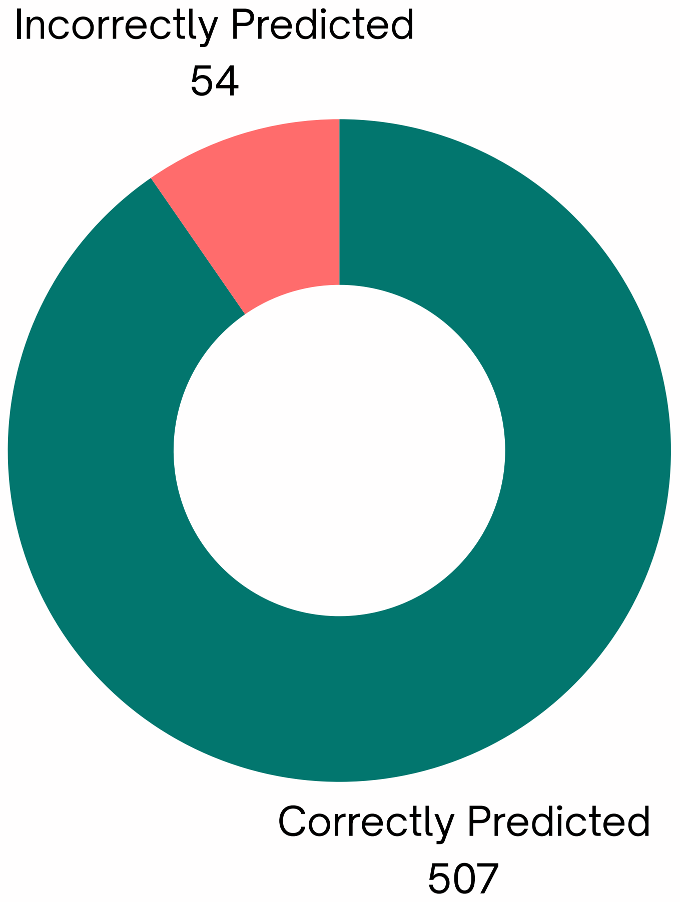
Logistic Regression



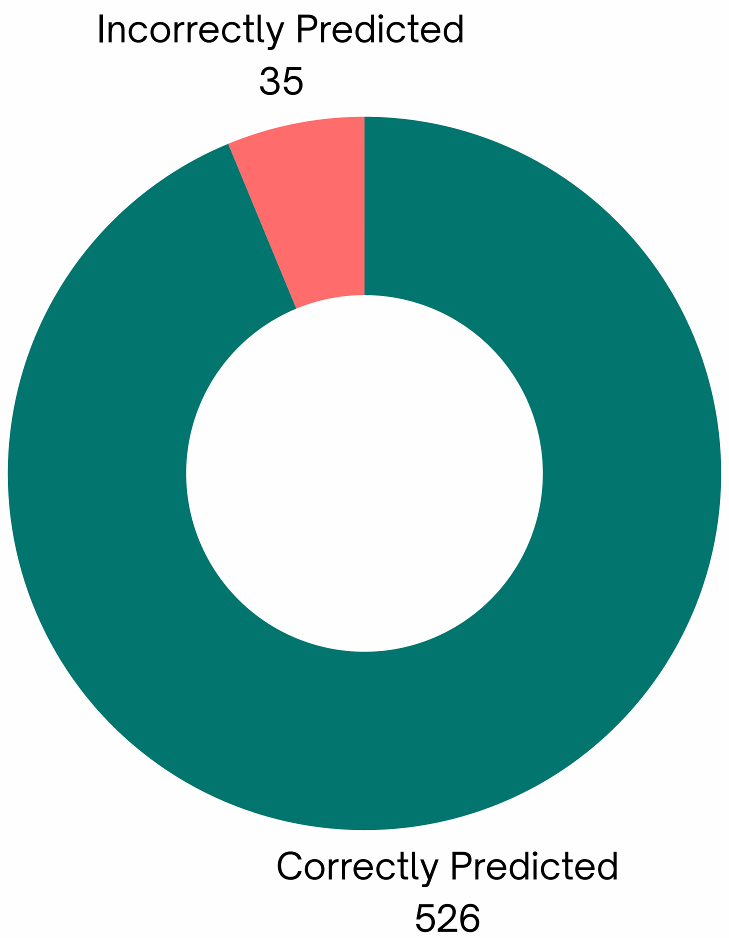
K-Nearest Neighbour



Neual Network



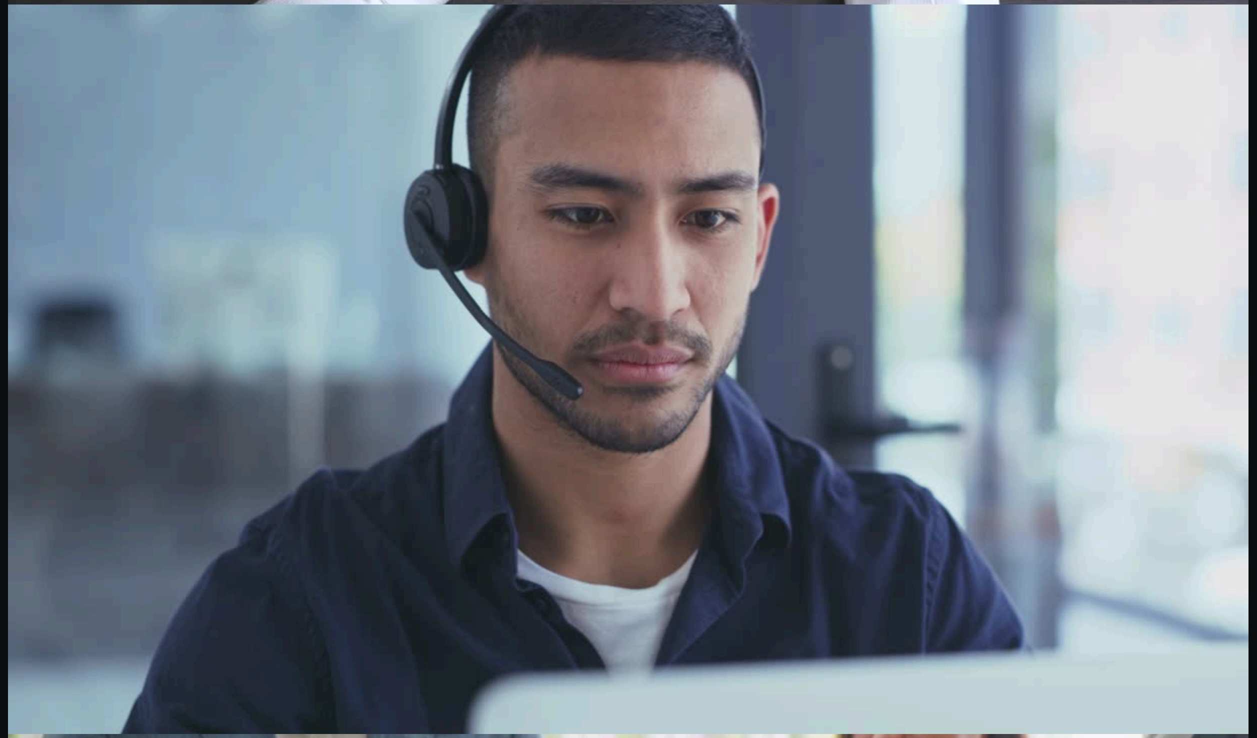
Random Forest



VORTEX

WHAT DO WE FOCUS ON?

VORTEX



CONTRACT LENGTH

BETTER SERVICE

REFERRALS

IMPLEMENTATION-
CONTRACT LENGTH

Incentives

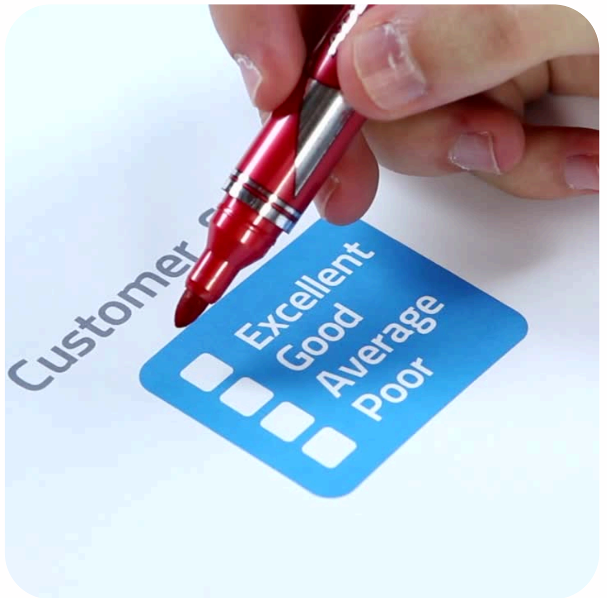
Early Renewal Options

Engage Customers

VORTEX

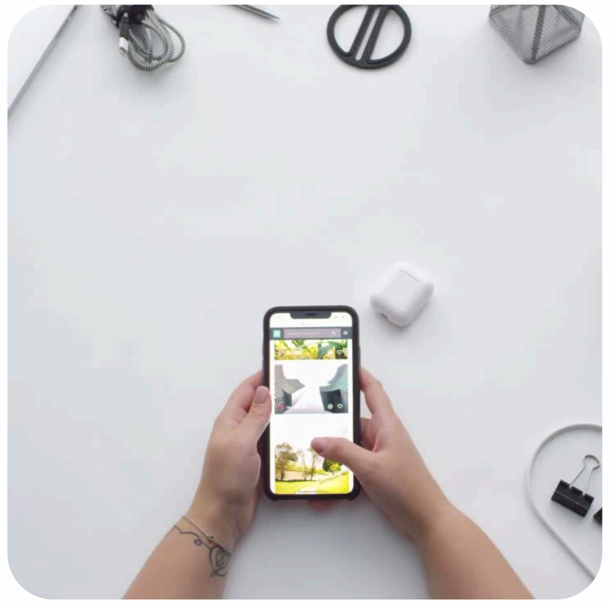
IMPLEMENTATION-

BETTER SERVICE



Enhance Customer Support

Proactive



Personalized Communication

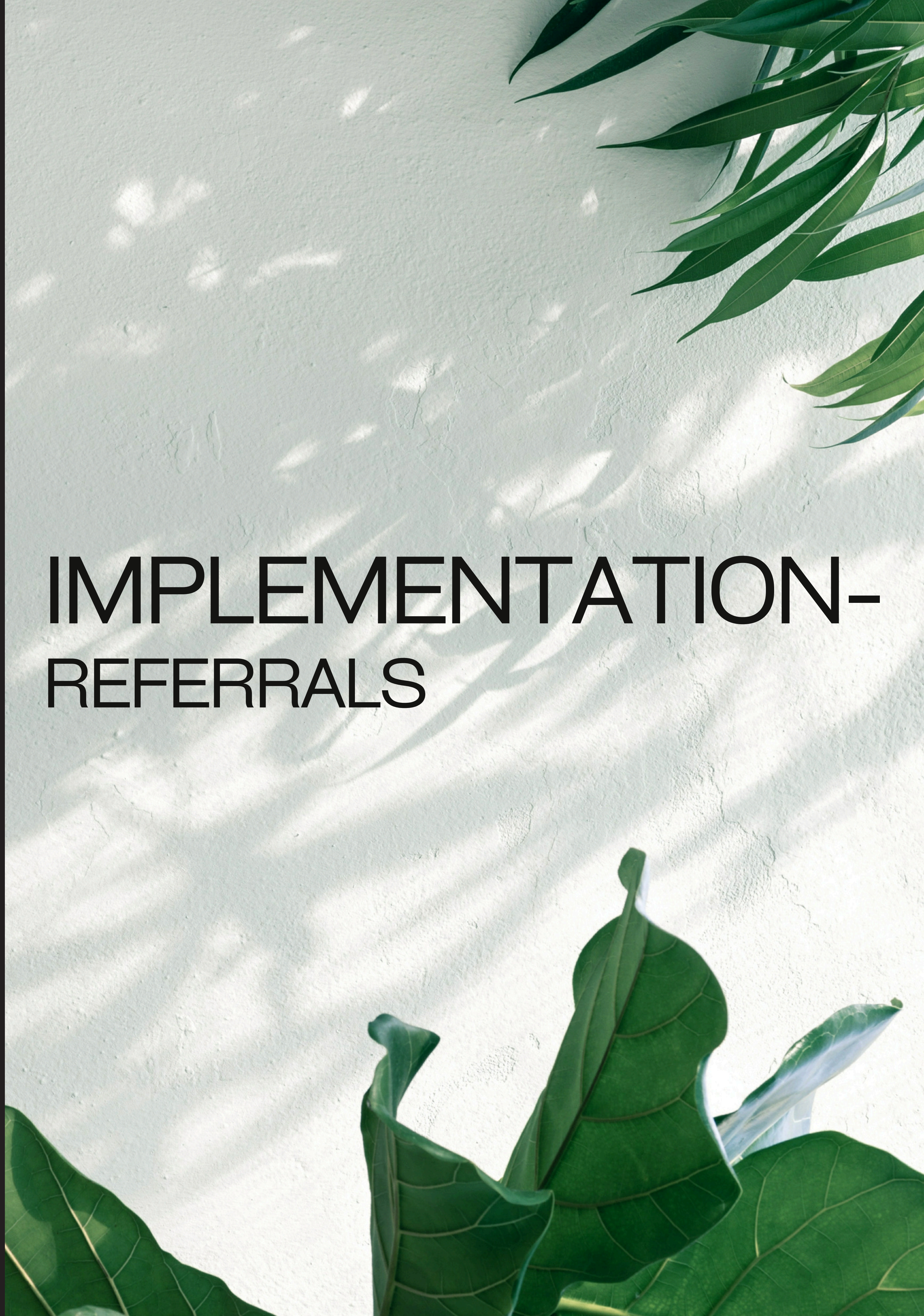
VORTEX

IMPLEMENTATION- REFERRALS

Targeted Referral

Reward Retention

Optimize Effectiveness



VORTEX

FINAL WORDS

-Working on these implementations.

-Work on better service thereafter.

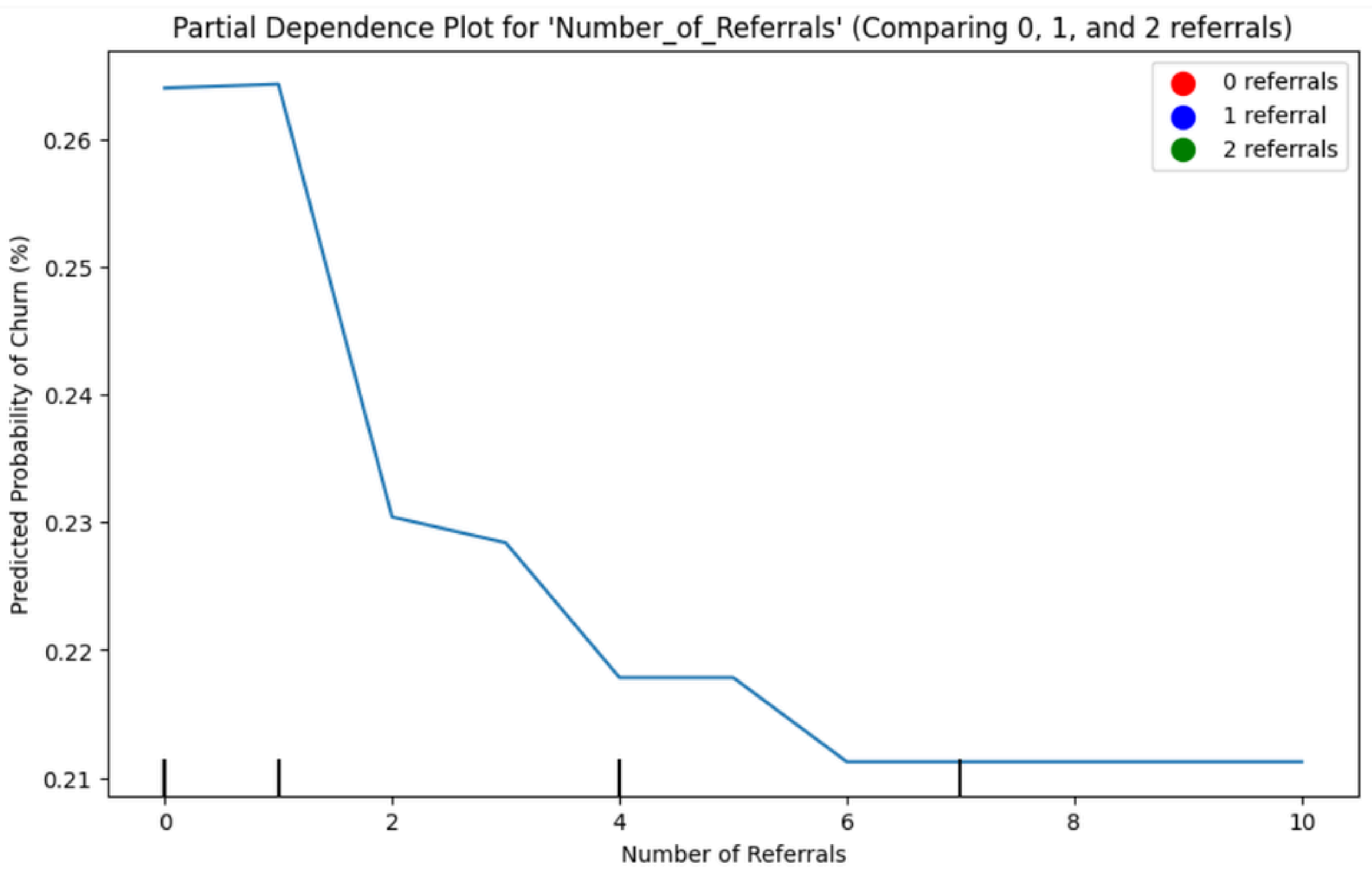
-Further fine-tune prediction model.

-Continuous Customer Feedback Integration.

VORTEX

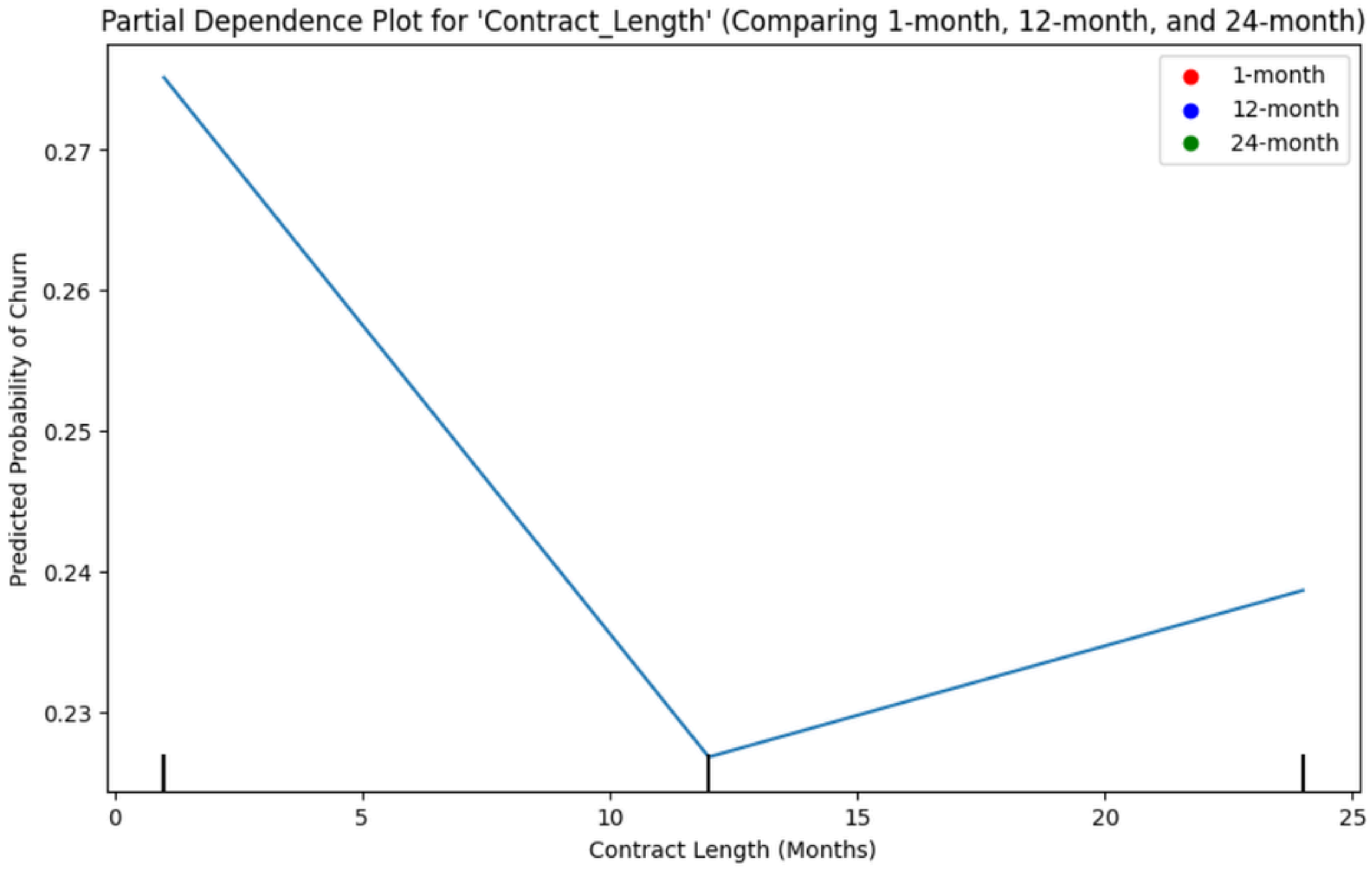
Thank you.

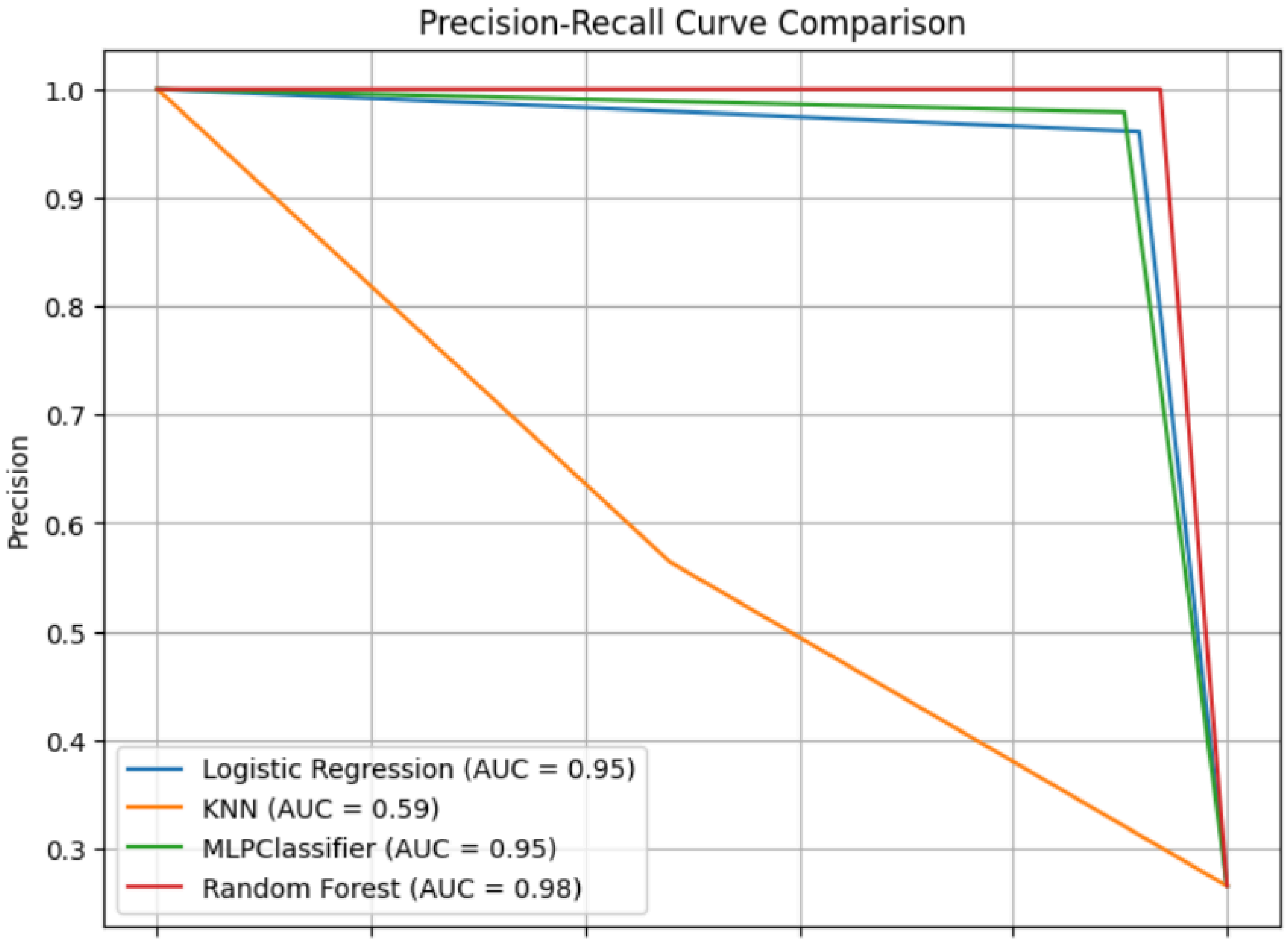
VORTEX

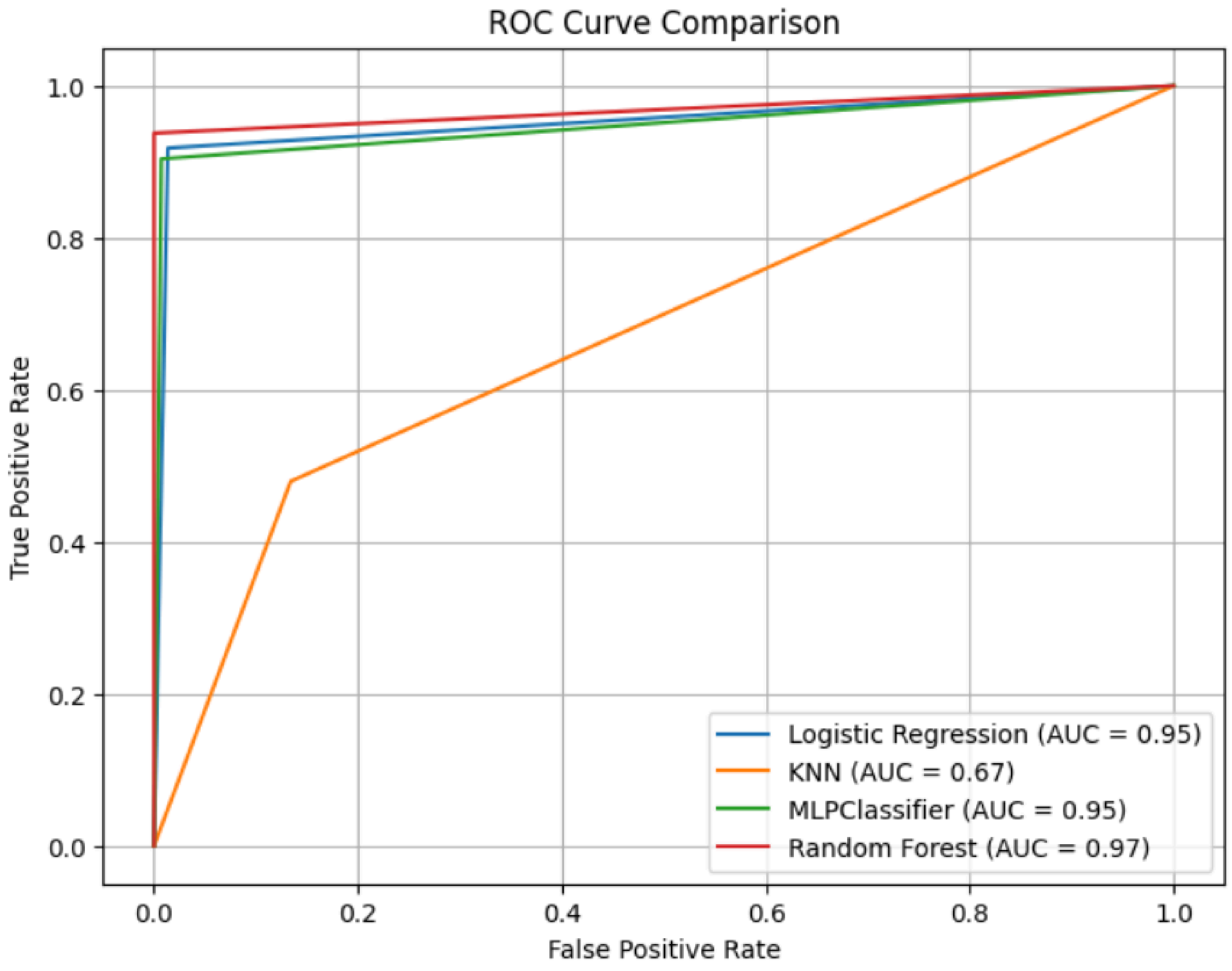


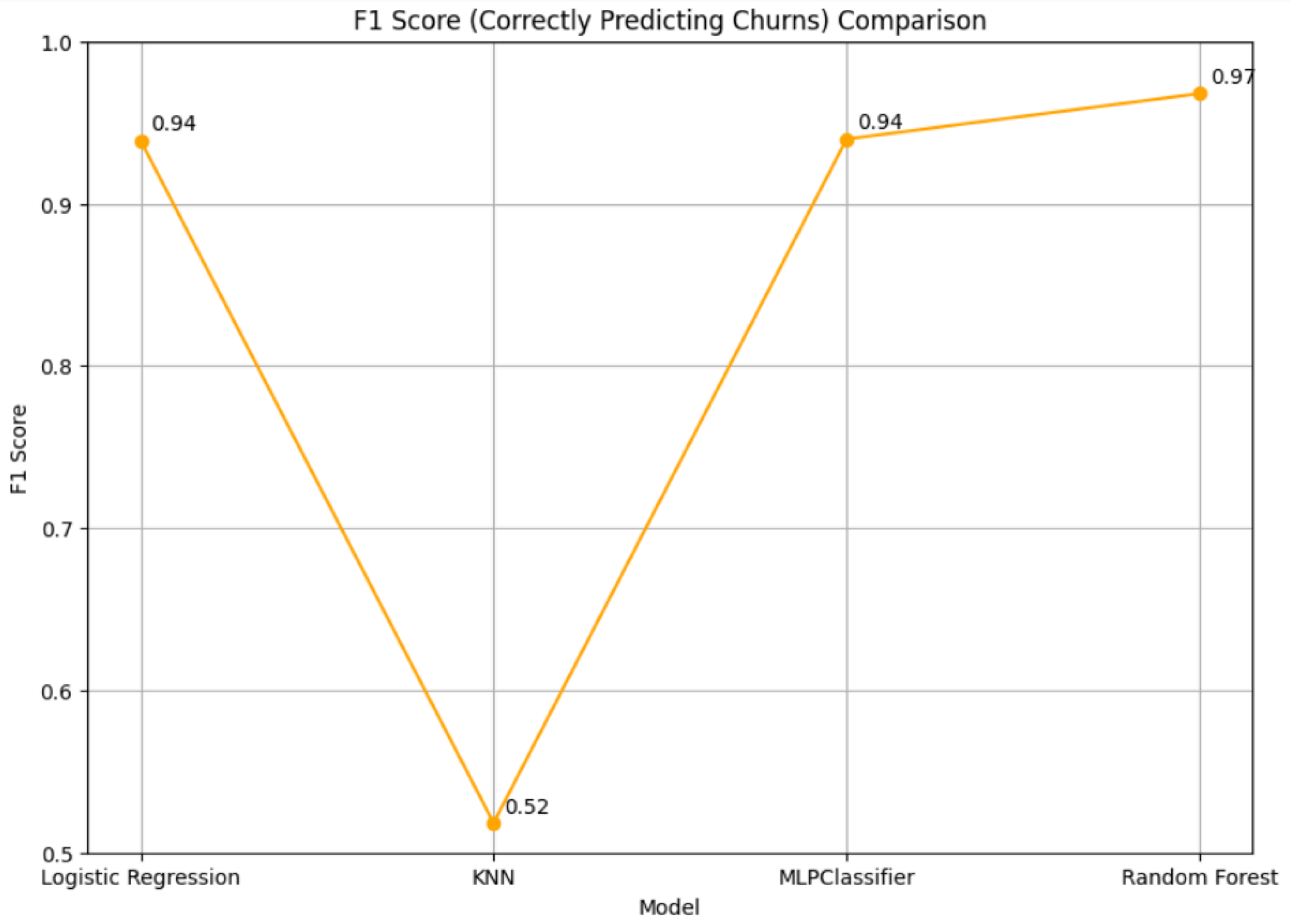
Impact on Churn with increase in Referrals

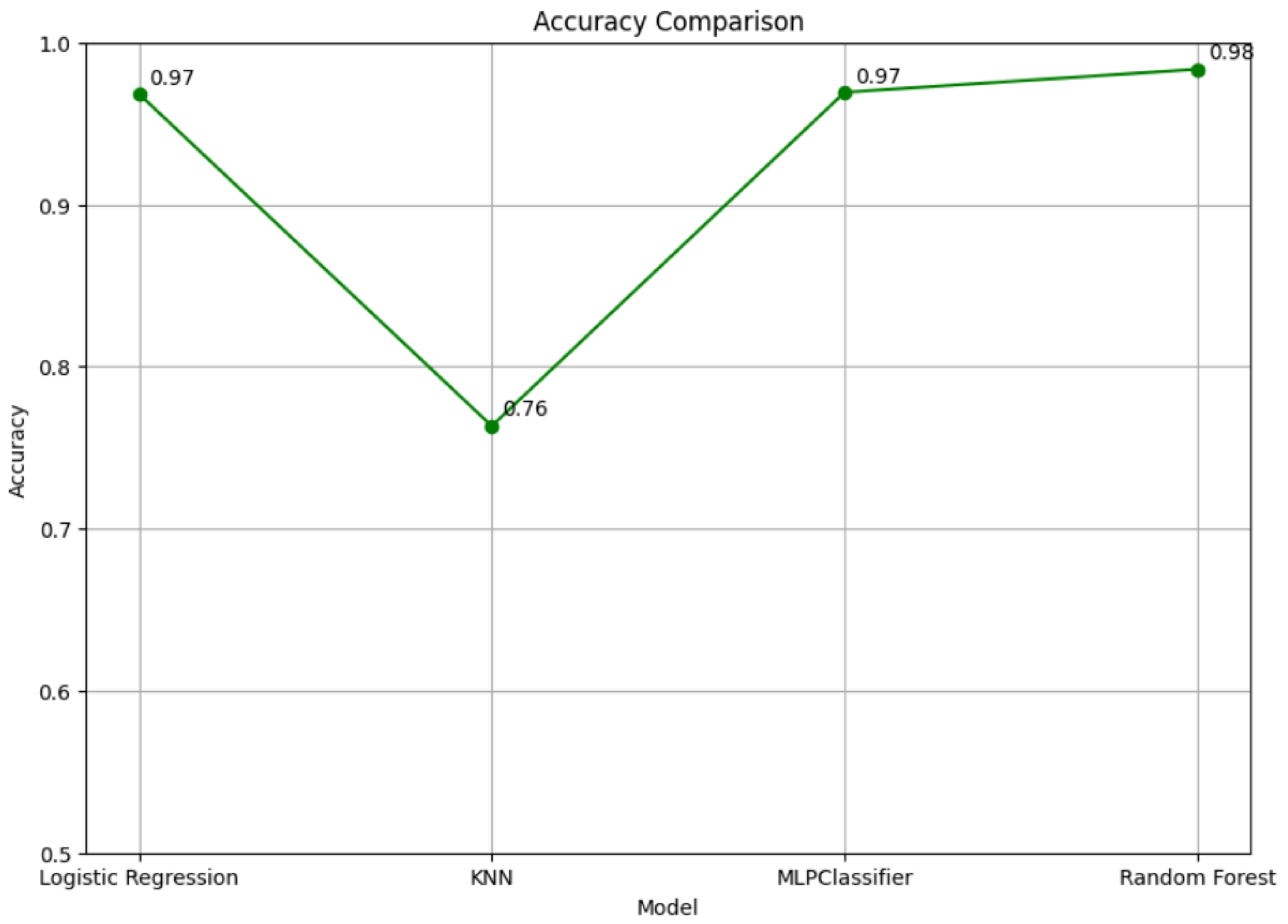
Impact on Churn with change in Contract Length

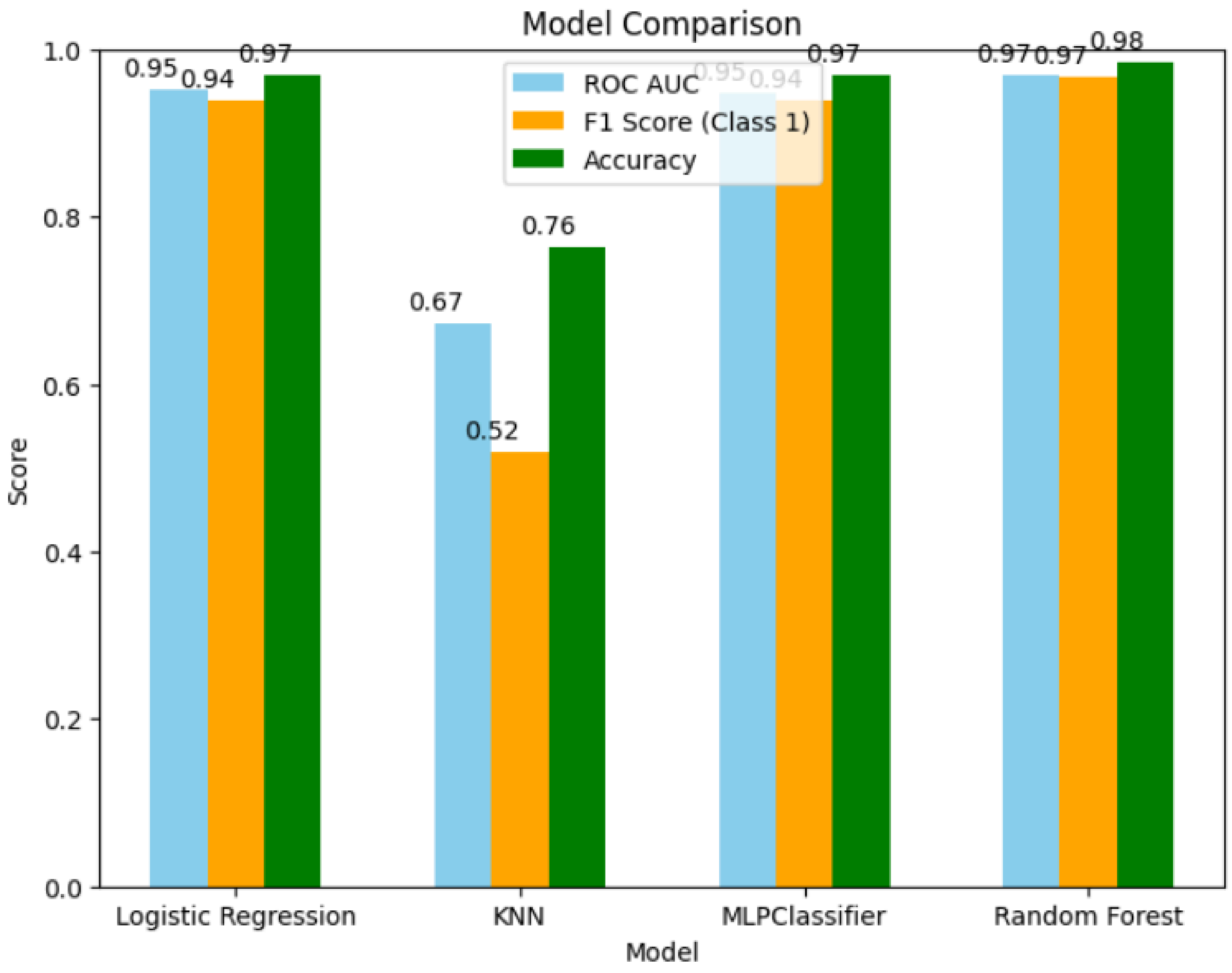



Precision-Recall Curve Comparison
Precision
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
Logistic Regression (AUC = 0.95)
KNN (AUC = 0.59)
MLPClassifier (AUC = 0.95)
Random Forest (AUC = 0.98)


ROC Curve Comparison
True Positive Rate
False Positive Rate
0.0
0.2
0.4
0.6
0.8
1.0
Logistic Regression (AUC = 0.95)
KNN (AUC = 0.67)
MLPClassifier (AUC = 0.95)
Random Forest (AUC = 0.97)


F1 Score (Correctly Predicting Churns) Comparison
F1 Score
Model
Logistic Regression
KNN
MLPClassifier
Random Forest
0.94
0.52
0.94
0.97
1.0
0.9
0.8
0.7
0.6
0.5


Accuracy Comparison
Accuracy
Model
Logistic Regression
KNN
MLPClassifier
Random Forest
0.97
0.76
0.97
0.98
1.0
0.9
0.8
0.7
0.6
0.5


Model Comparison
ROC AUC
F1 Score (Class 1)
Accuracy
0.95
0.94
0.97
0.67
0.52
0.76
0.95
0.94
0.97
0.97
0.97
0.98
1.0
0.8
0.6
0.4
0.2
0.0
Score
Logistic Regression
KNN
MLPClassifier
Random Forest
Model