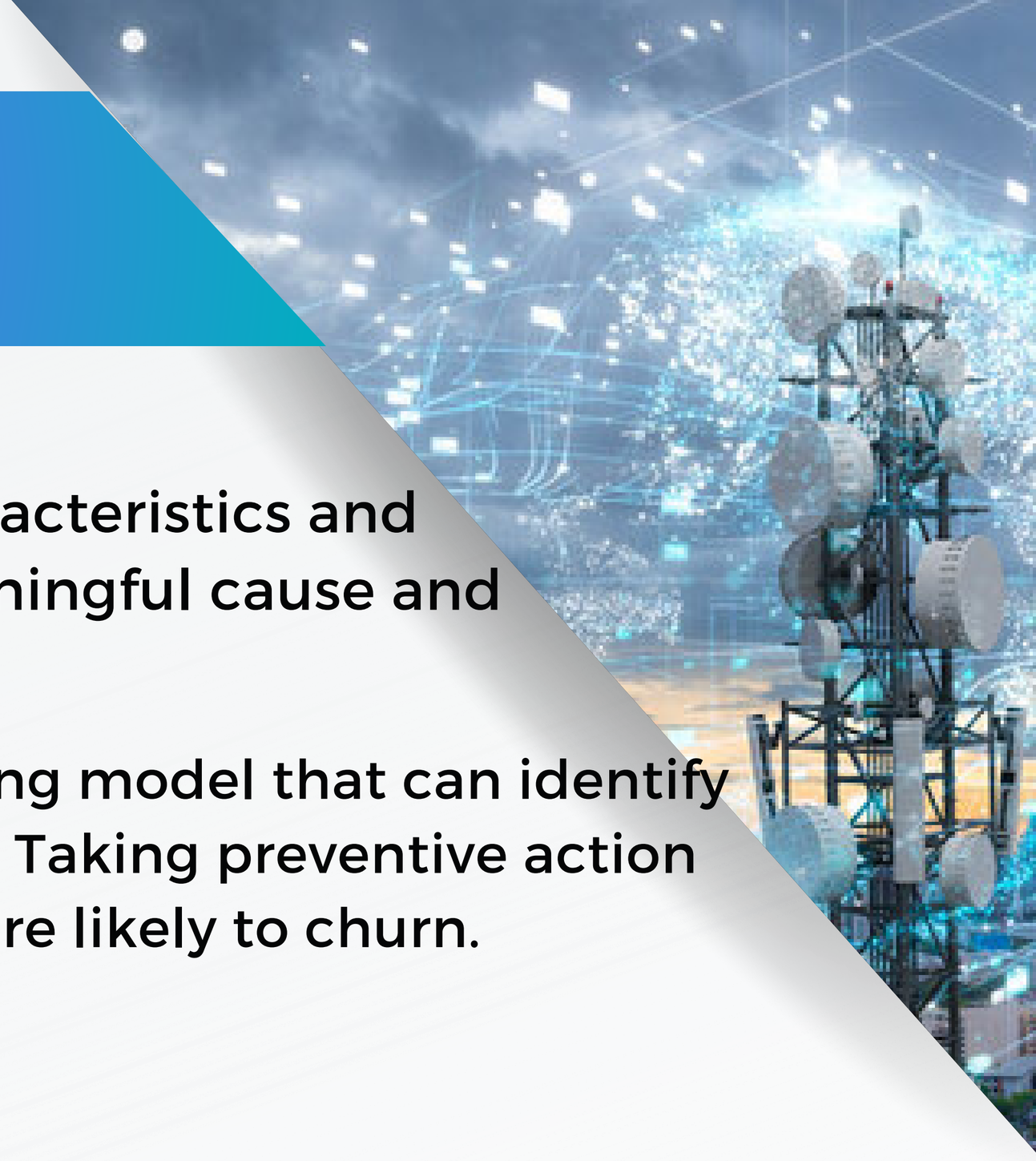
CUSTOMER CHURN ANALYSIS & MACHINE LEARNING MODEL FOR TELECOM COMPANY

BERFIN OKÇU

TARGETS



Analyzing customer characteristics and behaviours. Finding meaningful cause and effect relationships.

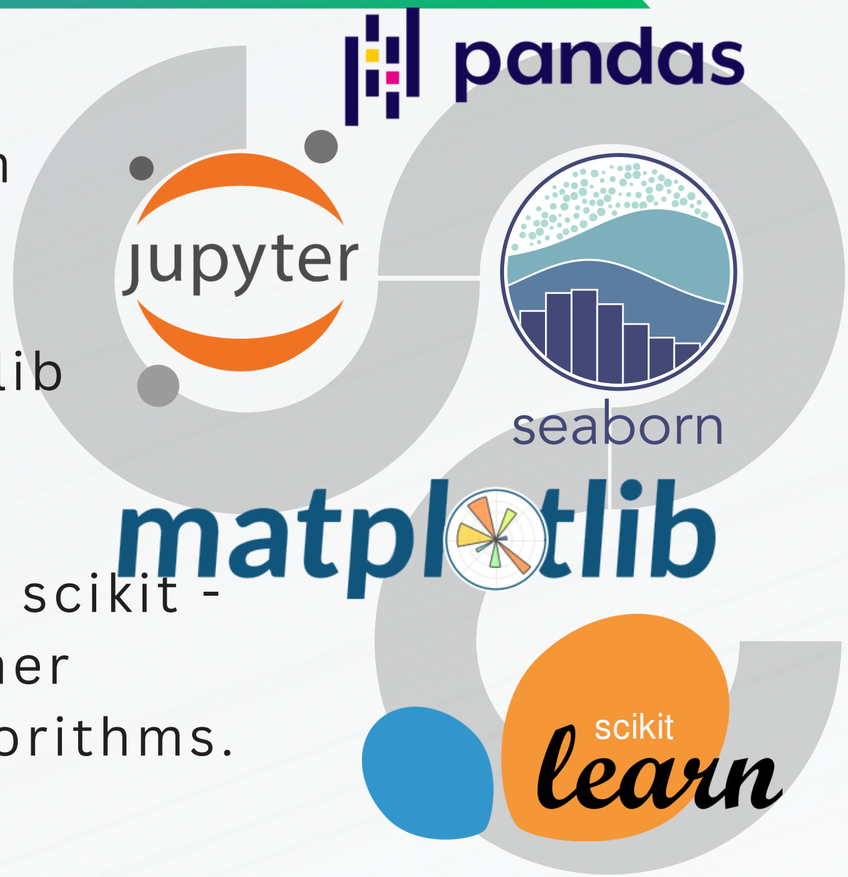
Building a machine learning model that can identify potential customer losses. Taking preventive action steps for customers who are likely to churn.

METHODOLOGY

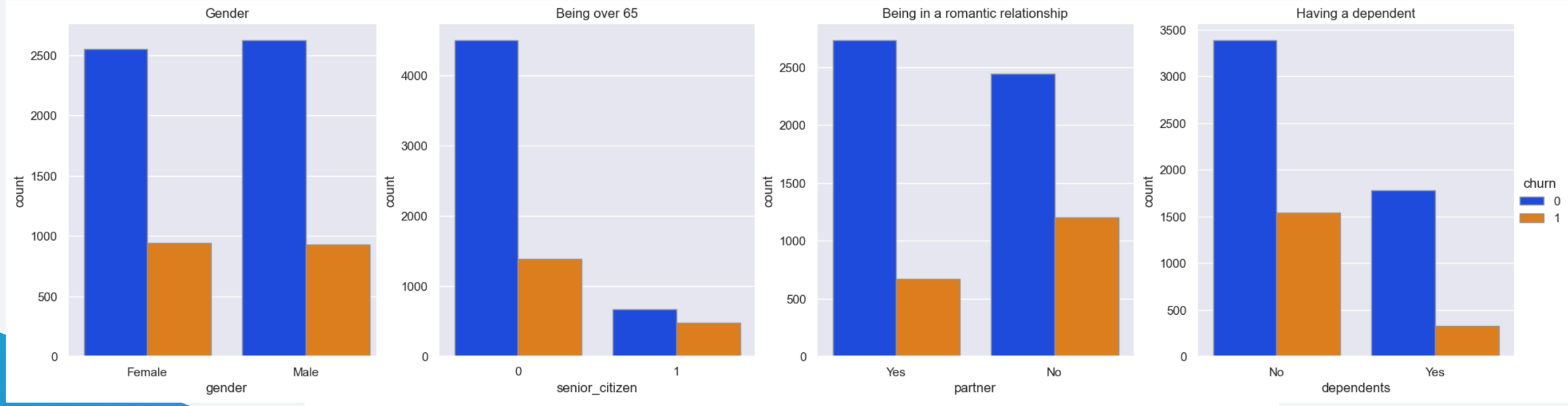
01 Analyzing customer data with pandas in Jupyter notebook environment

02 Visualizing these analysis with matplotlib and seaborn libraries

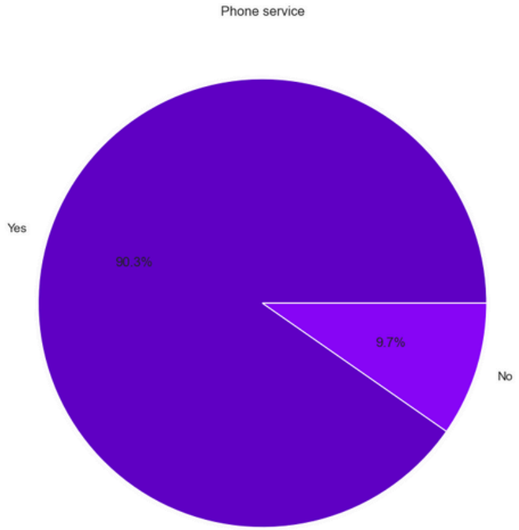
03 Building a machine learning model with scikit - learn that can predict potential customer churn using different classification algorithms.

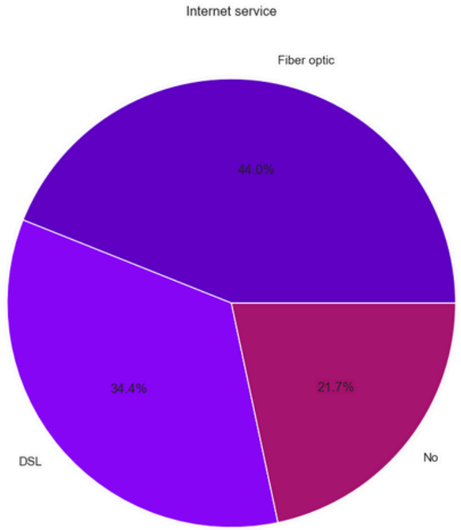


ANALYSIS

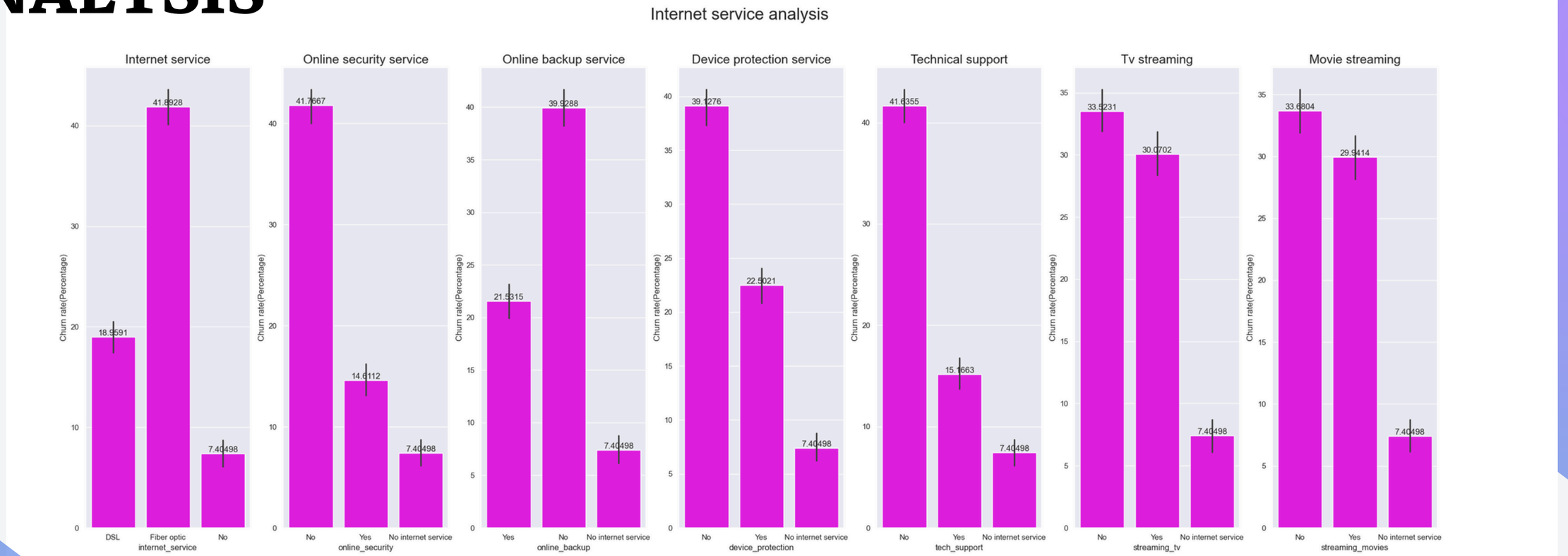


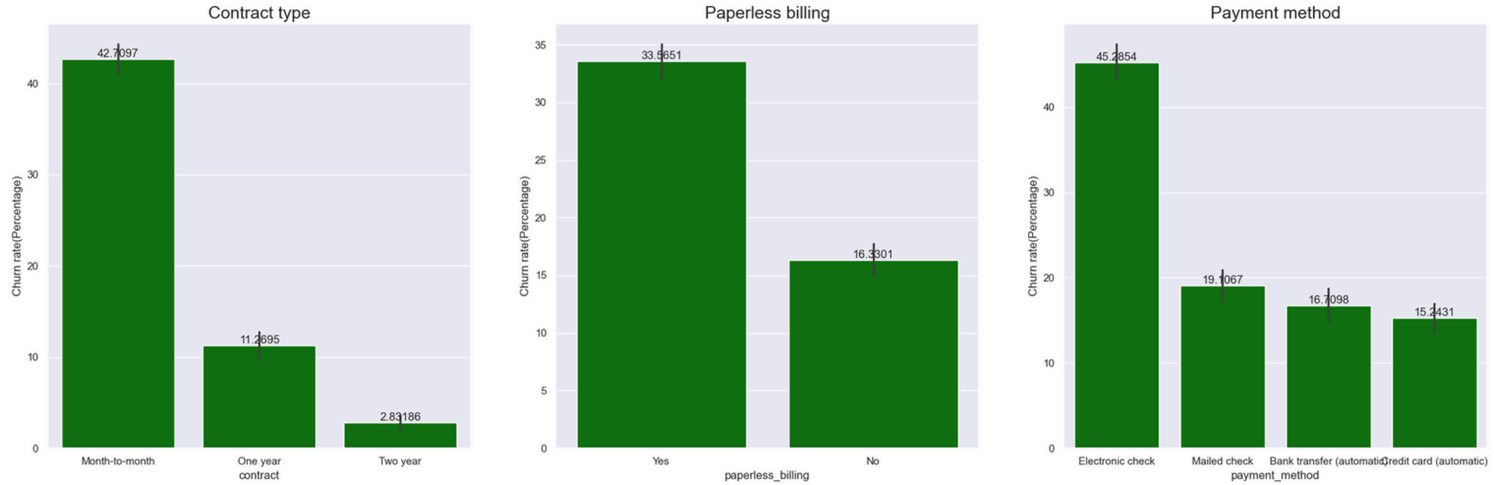
Services





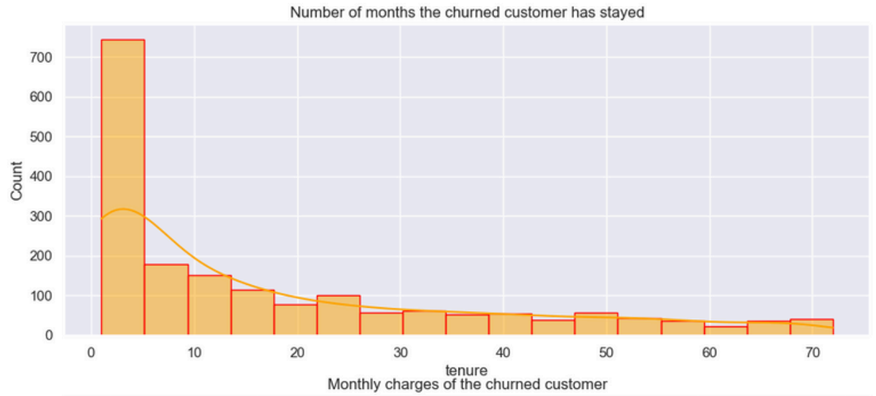
ANALYSIS

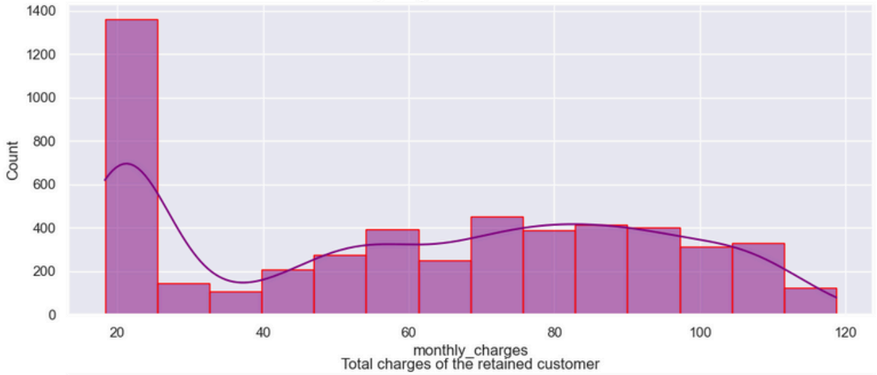

Internet service analysis
Internet service
Online security service
Online backup service
Device protection service
Technical support
Tv streaming
Movie streaming
Churn rate(Percentage)
18.9591
41.8928
7.40498
41.7667
14.6112
7.40498
21.5315
39.9288
7.40498
39.1276
22.5021
7.40498
41.0355
15.1663
7.40498
33.5231
30.0702
7.40498
33.6804
29.9414
7.40498
DSL
Fiber optic
No
internet_service
No
Yes
No internet service
online_security
Yes
No
No internet service
online_backup
No
Yes
No internet service
device_protection
No
Yes
No internet service
tech_support
No
Yes
No internet service
streaming_tv
No
Yes
No internet service
streaming_movies

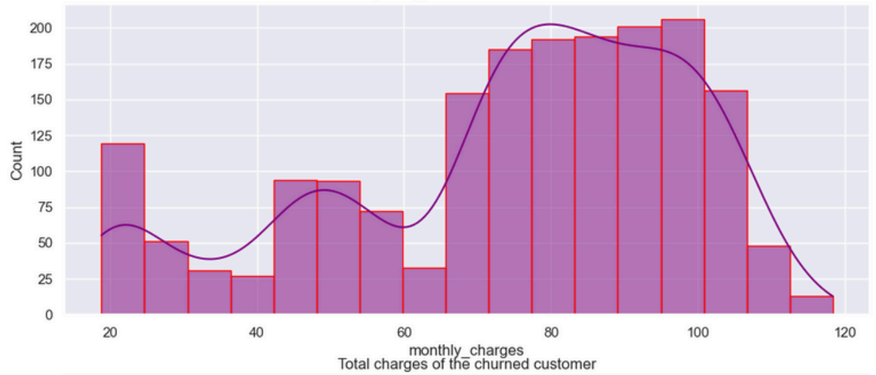

Payment analysis
Contract type
Paperless billing
Payment method
Churn rate(Percentage)
42.7097
11.2695
2.83186
33.5651
16.3301
45.2854
19.1067
16.7098
15.2431
Month-to-month
One year
Two year
contract
Yes
No
paperless_billing
Electronic check
Mailed check
Bank transfer (automatic)
Credit card (automatic)
payment_method

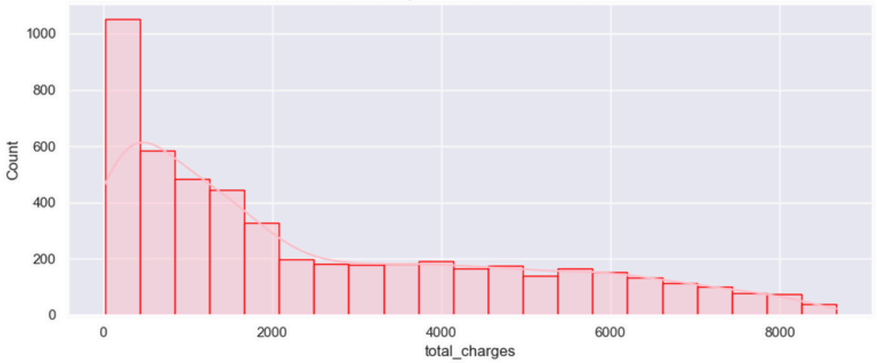
ANALYSIS

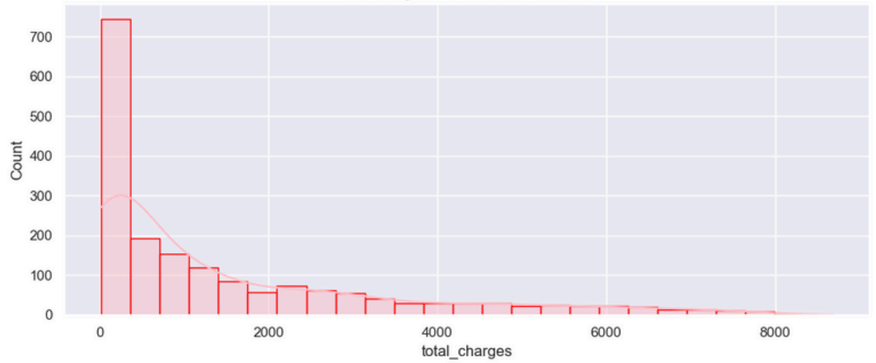

Number of months the retained customer has stayed
Count
tenure


Number of months the churned customer has stayed
Count
tenure

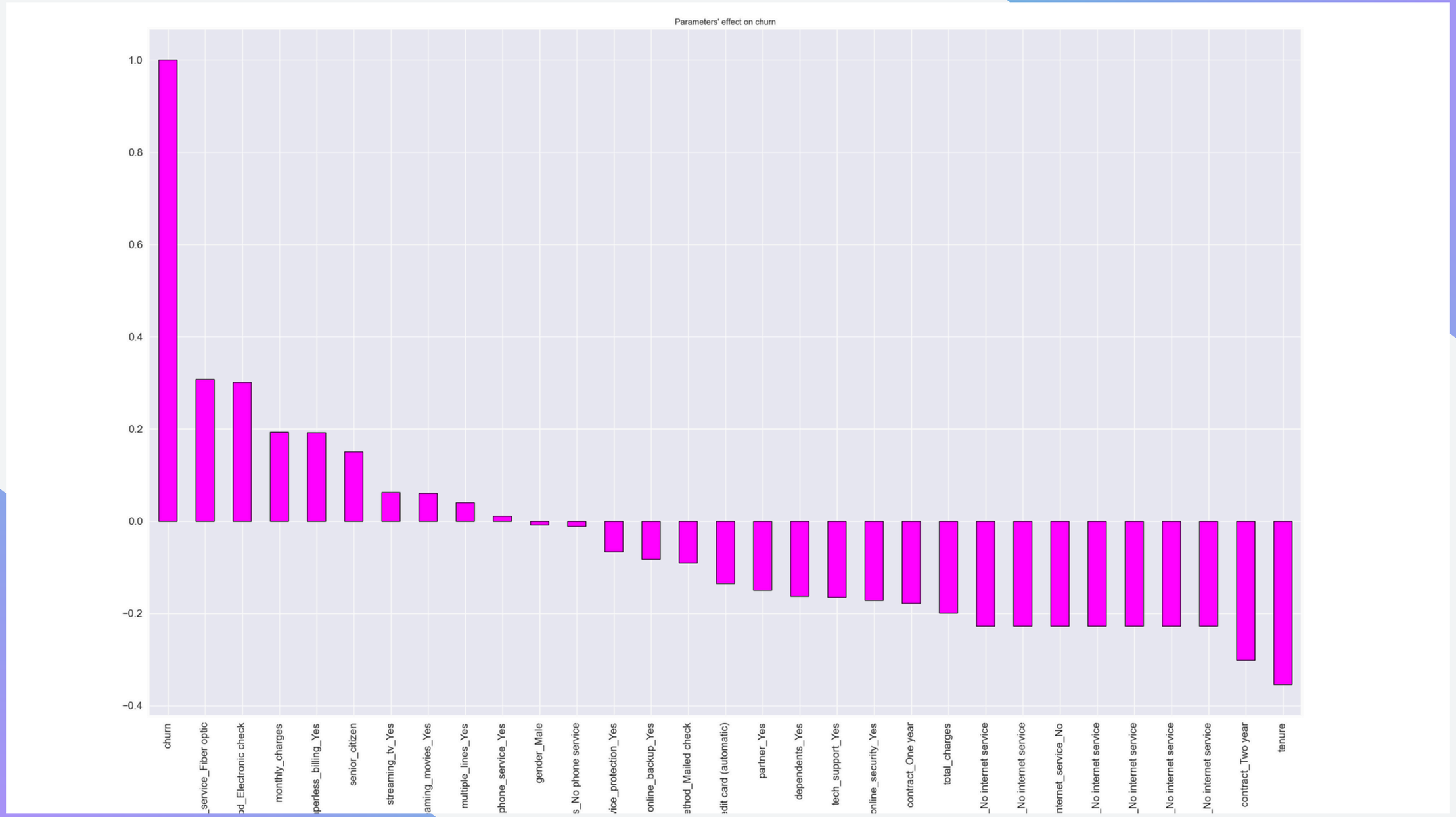

Monthly charges of the retained customer
Count
monthly_charges


Monthly charges of the churned customer
Count
monthly_charges


Total charges of the retained customer
Count
total_charges


Total charges of the churned customer
Count
total_charges

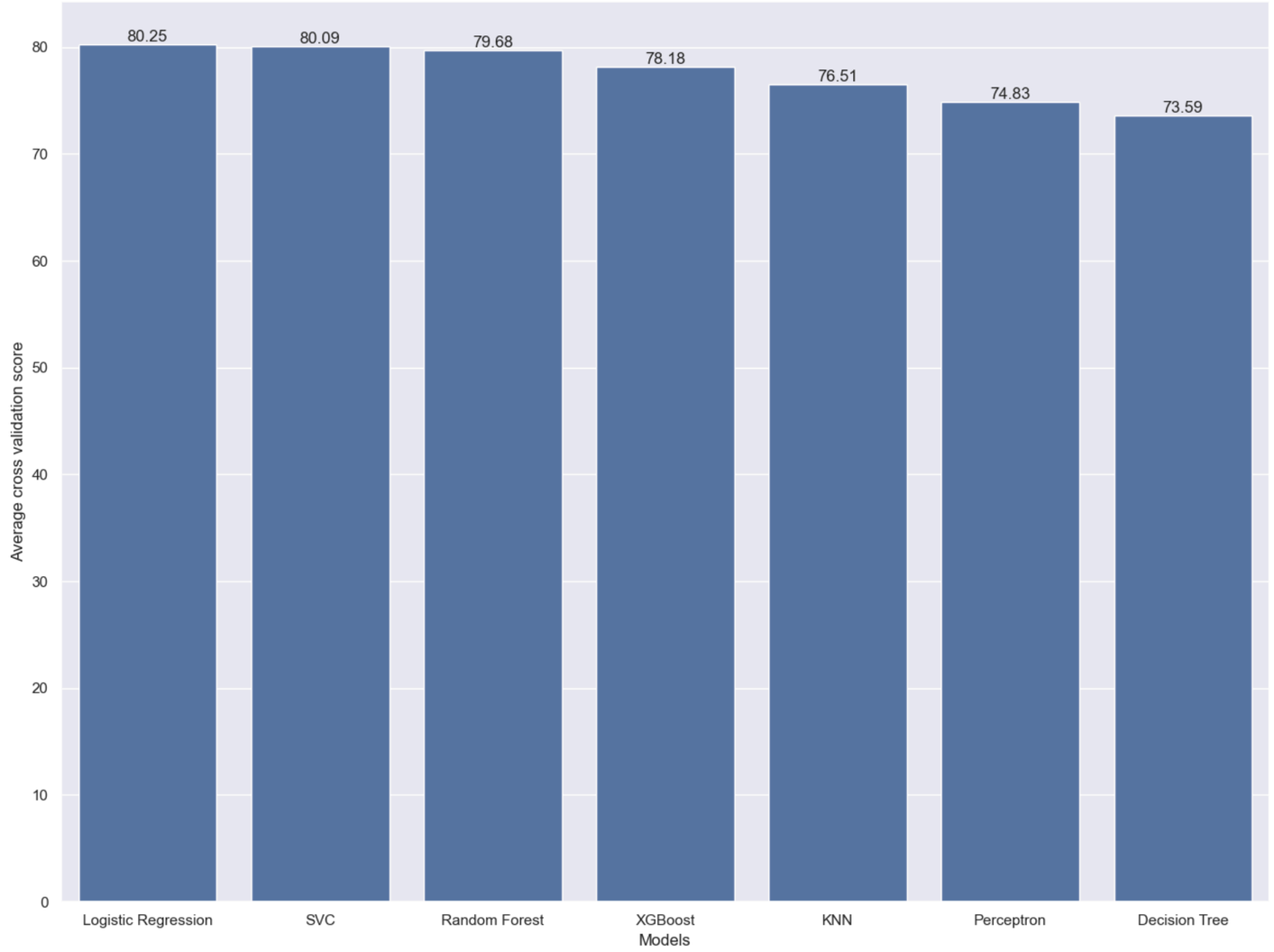
Parameters' Effects



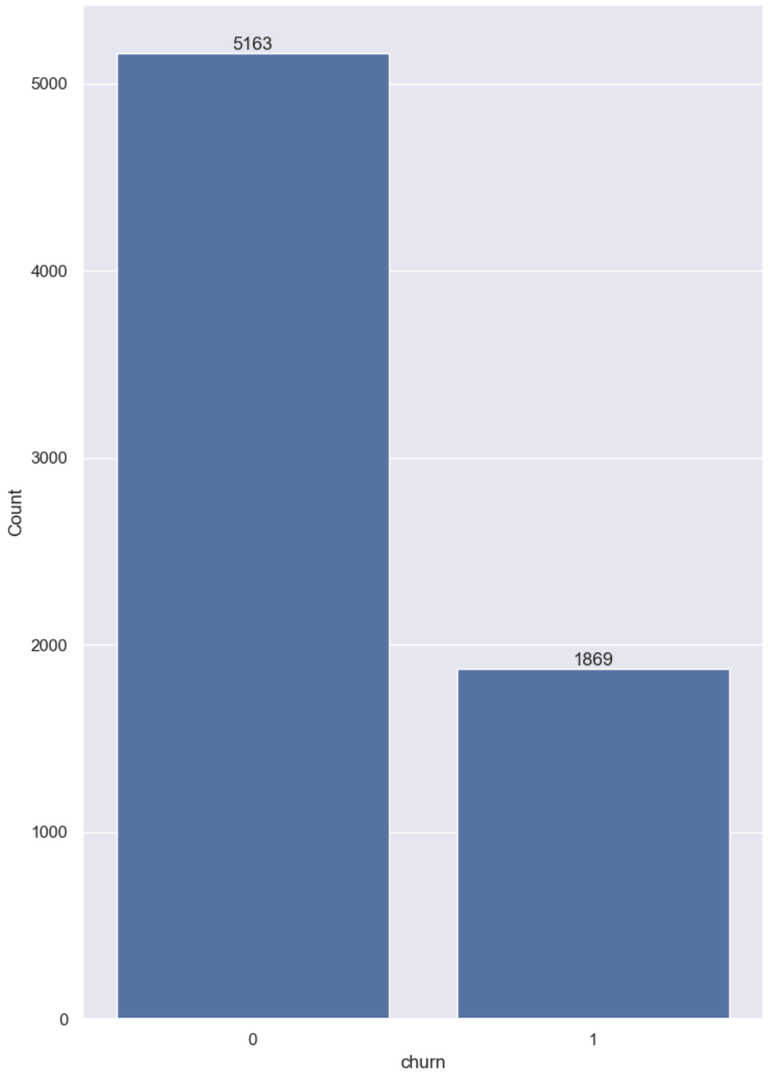
Cross Validation for Machine Learning Model

Train Test Split
-70% Train
-30% Test

-Logistic Regression
-Support Vector Classifier
-Random Forest
-XGBoost

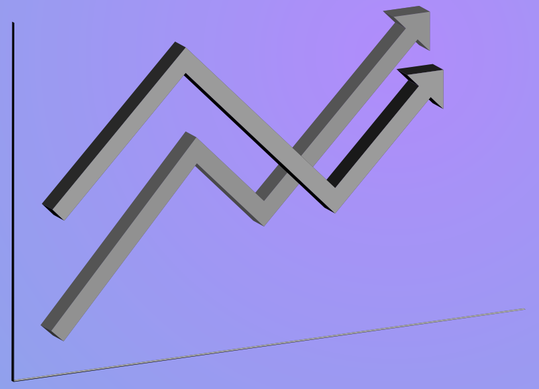


Imbalanced Dataset



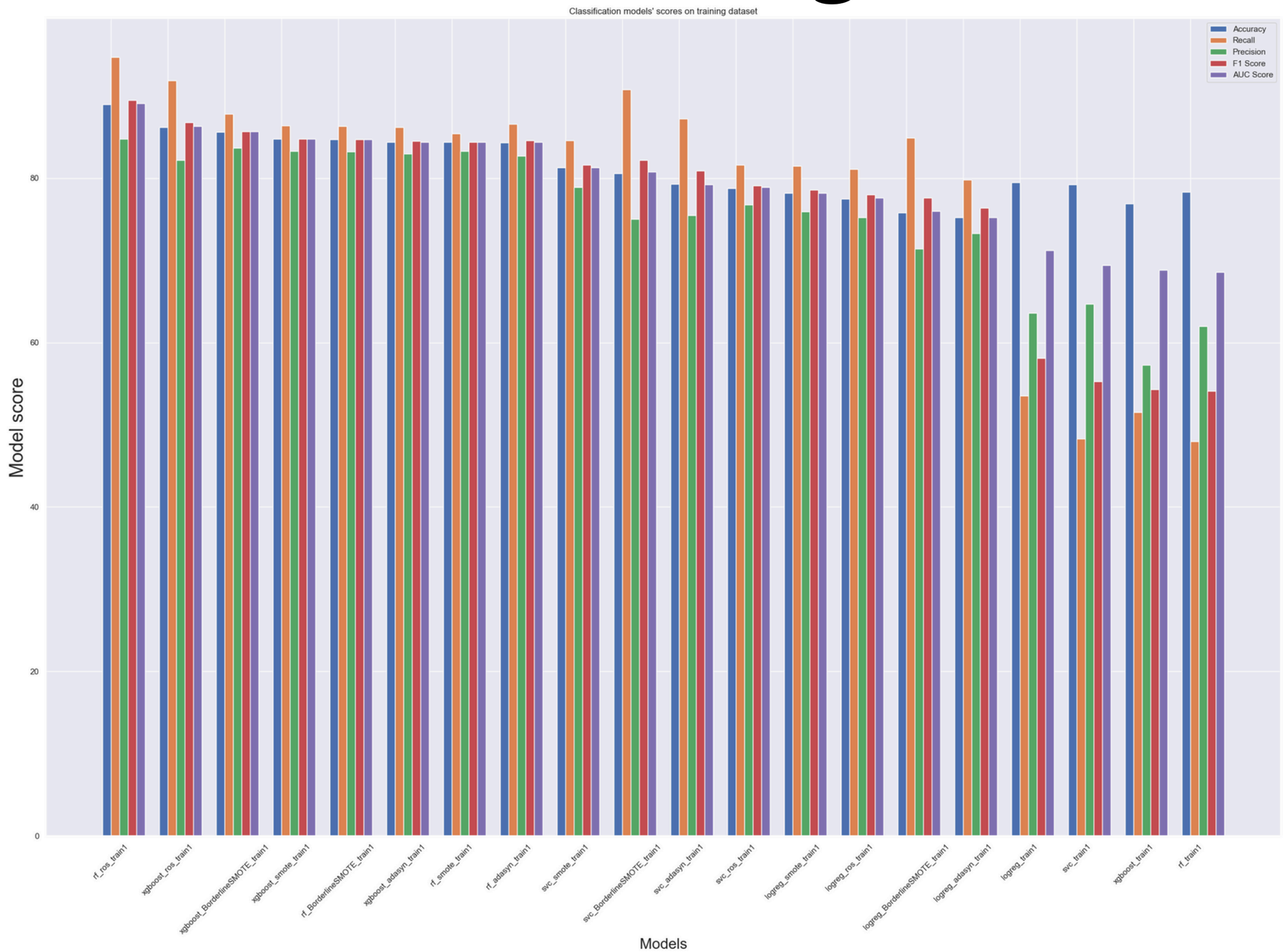
74.5% **24.5%**

OVERSAMPLING

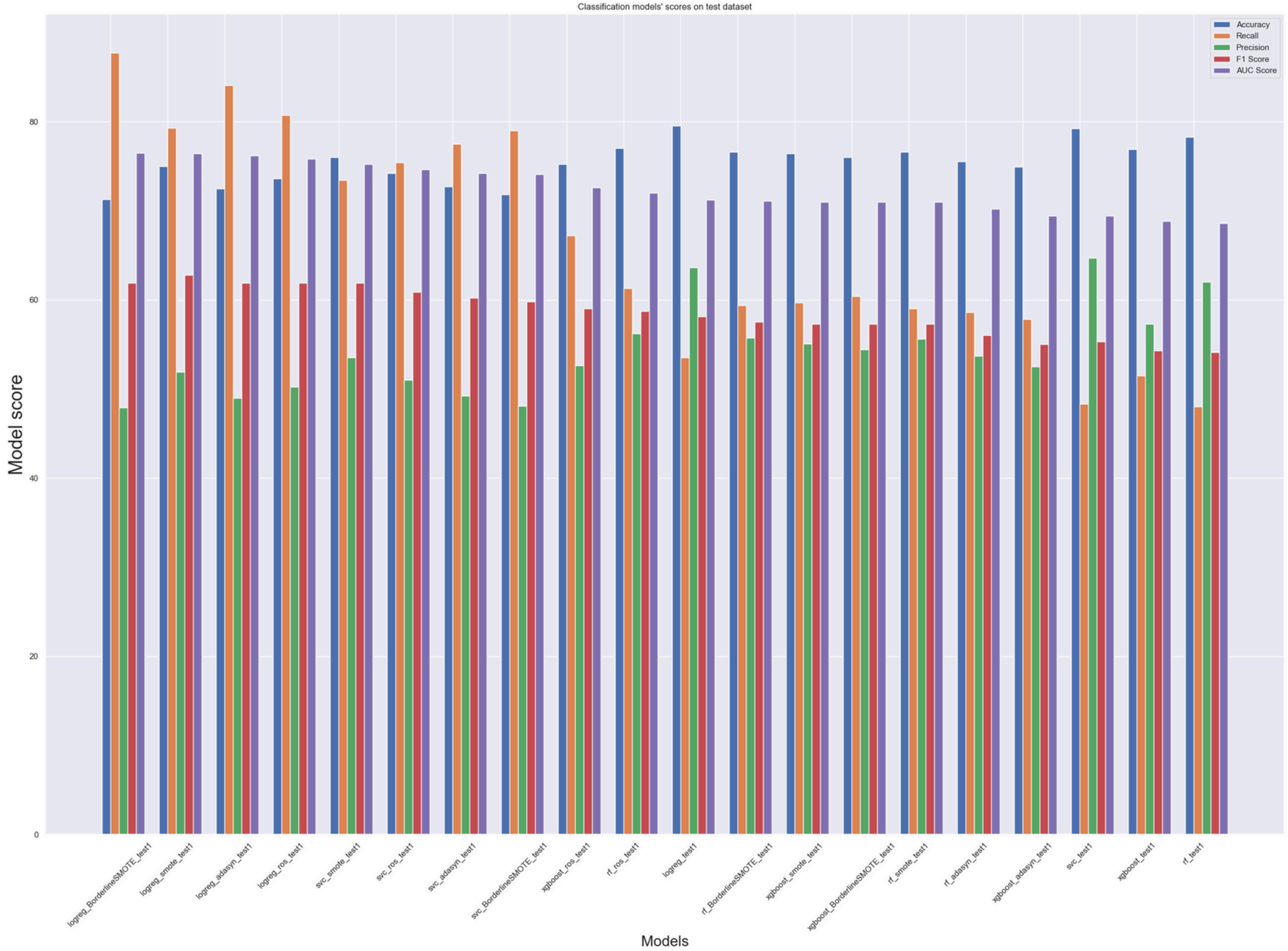


-Without oversampling
-Random Over Sampler
-SMOTE
-Adasyn
-BorderlineSMOTE

Model Scores on Training Dataset

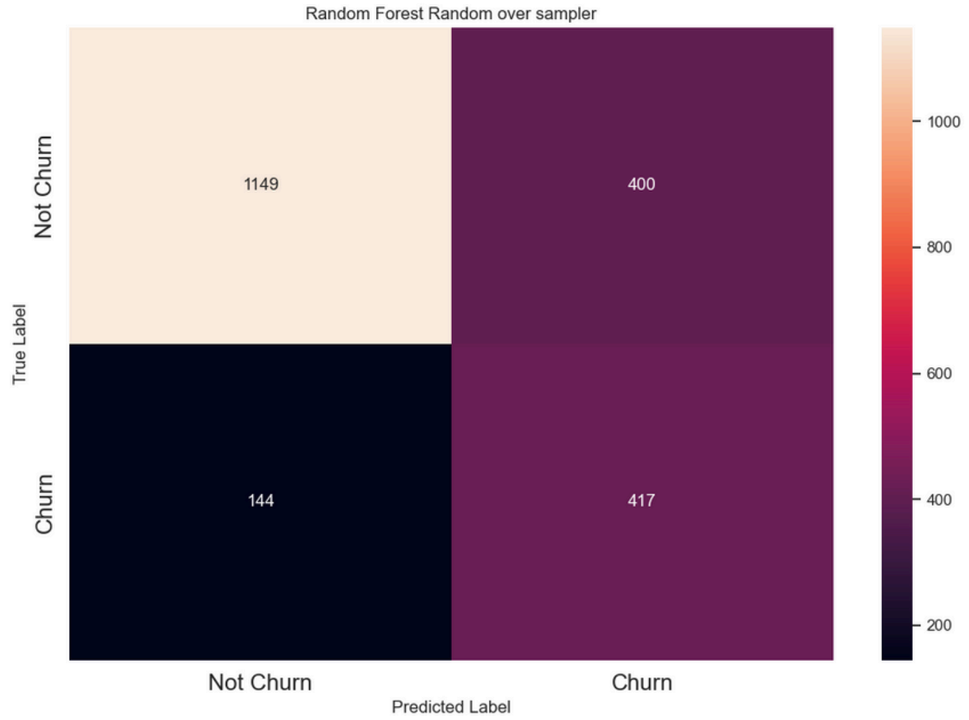


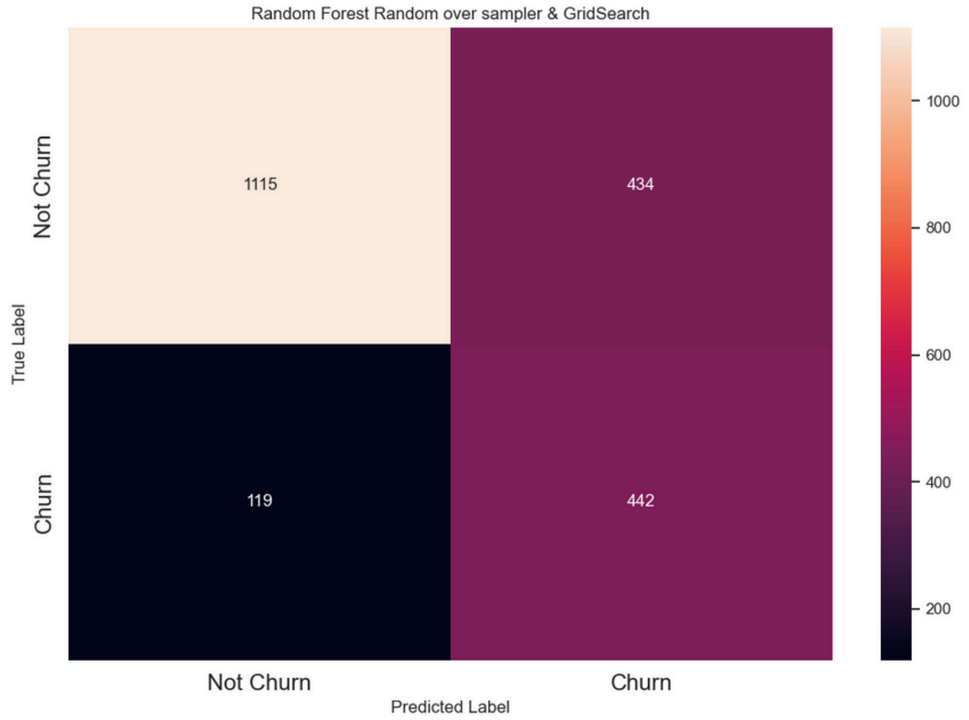
Model Scores on Test Dataset



Model Optimization Hyperparameter Tuning GridSearchCV

```
RandomForestClassifier(max_depth=6, max_features=None, max_leaf_nodes=9,
                       min_samples_leaf=2, min_samples_split=5,
                       n_estimators=25)
```





Final Model

Logistic Regression borderlineSMOTE

	Accuracy	Recall	Precision	F1 Score	AUC Score
LogisticRegression(solver='liblinear')Class 1	0.717	0.87	0.482	0.62	0.765