

TELCO CUSTOMER CHURN ANALYSIS



AGENDA



01 BACKGROUND

02 DATA
PREPROCESSING

03 MODELING

04 RESULTS

05 CHALLENGES

06 RECOMMENDATIO
N

01 BACKGROUND

- Telecom carriers have a high penetration in the US population already
- The competition among telecom companies is stiff
- The cost of acquiring a new customer is higher than retaining an old customer
- Our goal is to analyze key factors determining churn and develop a strategy to prevent churn
- Predict the customers who are likely to churn and take proactive actions

DATA EDA

Predictors:

Churn	gender	SeniorCitizen	Partner	Dependents	PhoneService	MultipleLines	InternetService
No	1.507263	1.128995	1.472400	1.343405	1.901220	2.311253	1.893473
Yes	1.497592	1.254682	1.642055	1.174425	1.909042	2.363831	1.814874

Churn	OnlineSecurity	OnlineBackup	DeviceProtection	TechSupport	StreamingTV	StreamingMovies
No	1.878172	1.904126	1.907806	1.880108	1.910905	1.915553
Yes	1.278759	1.780631	1.412520	1.286784	1.556447	1.558587

Churn	Contract	PaperlessBilling	PaymentMethod	Tenure_Category	Duration	tenure	MonthlyCharges	TotalCharges
No	1.887081	1.463878	2.497385	2.843696	2.291497	37.650010	61.307408	2555.344141
Yes	1.140182	1.250936	1.813269	1.894596	1.636704	17.979133	74.441332	1531.796094

factorize() - No:1, Yes:2, Others:3

Shape: (7043,21)

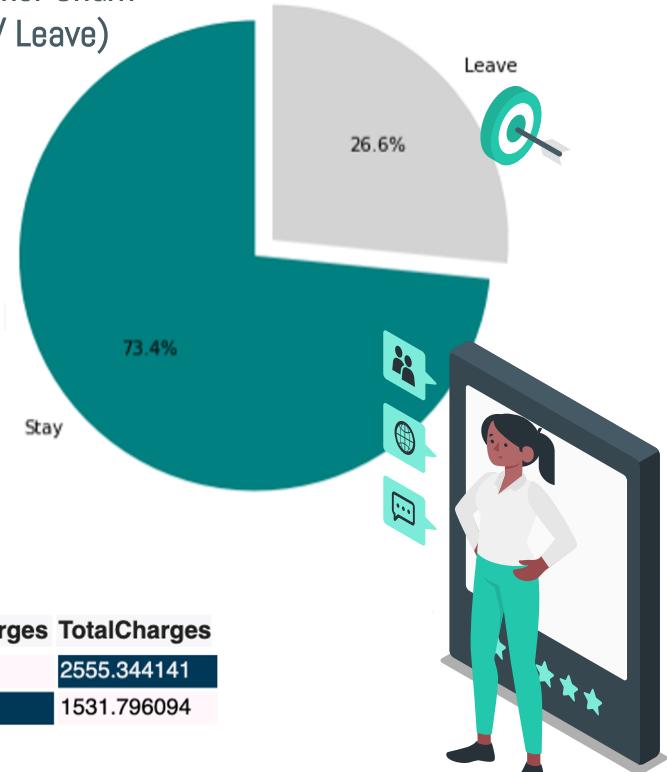
Missing value : 0

Change data types

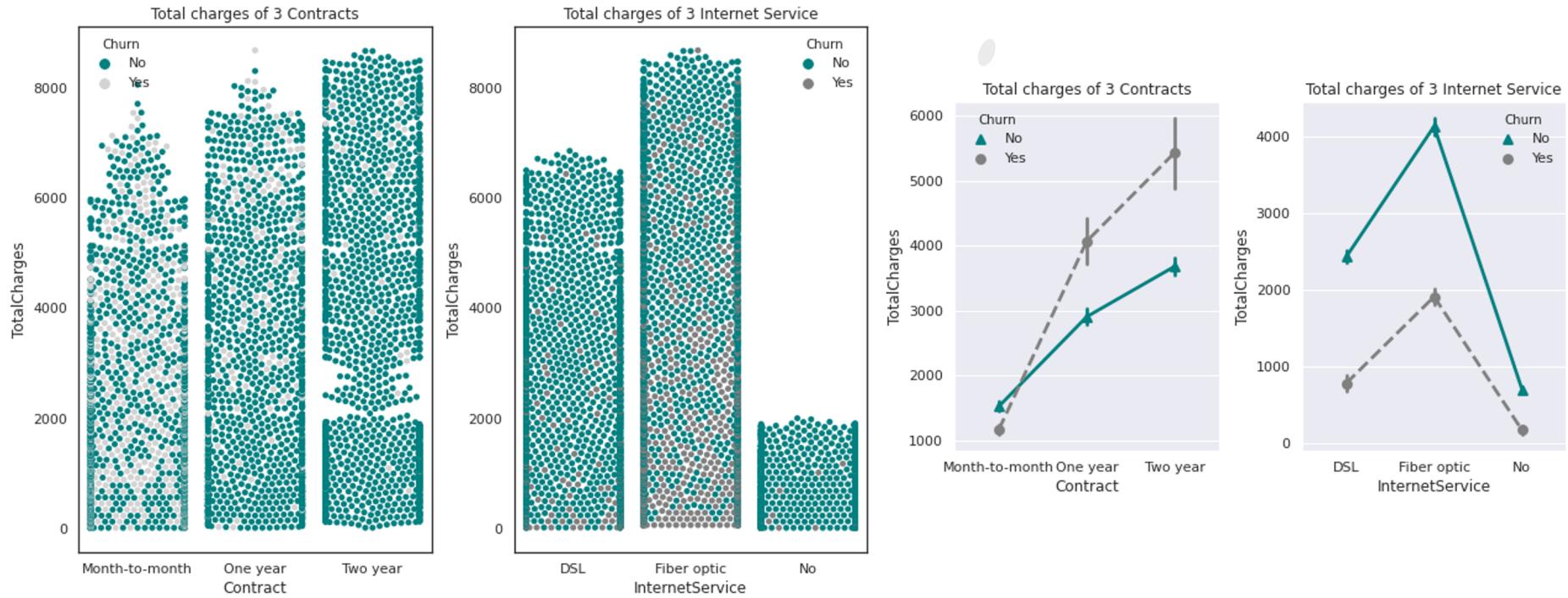
Duplicate rows: 11

Target

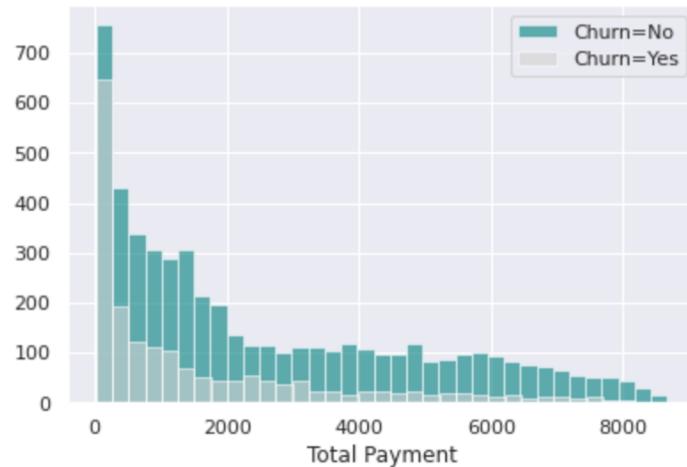
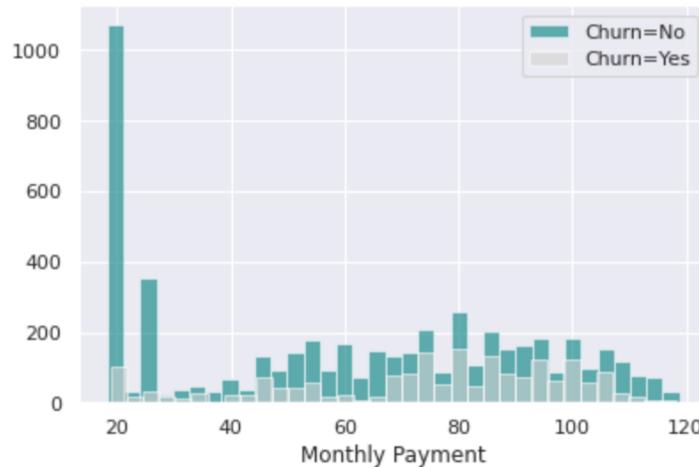
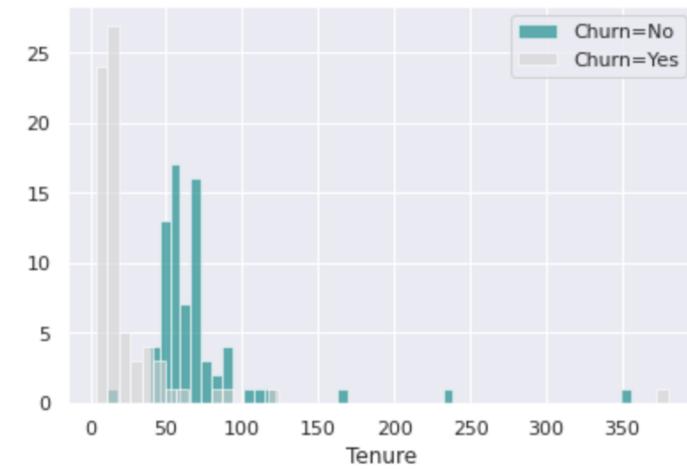
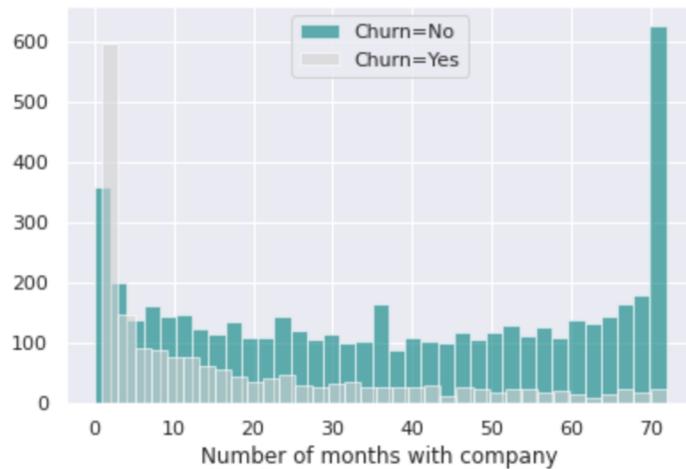
Customer Churn
(Stay / Leave)



DATA EDA

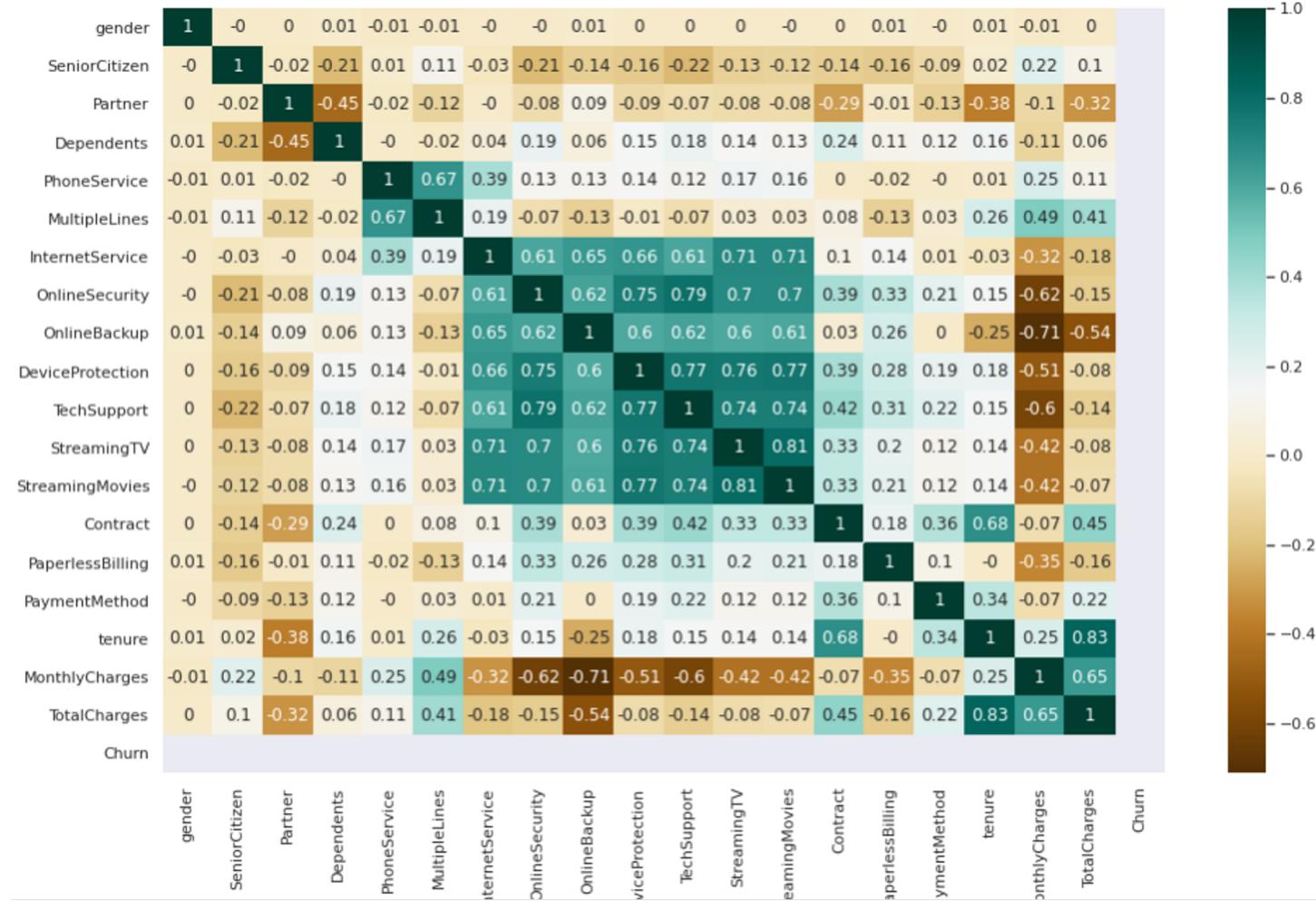


DATA EDA



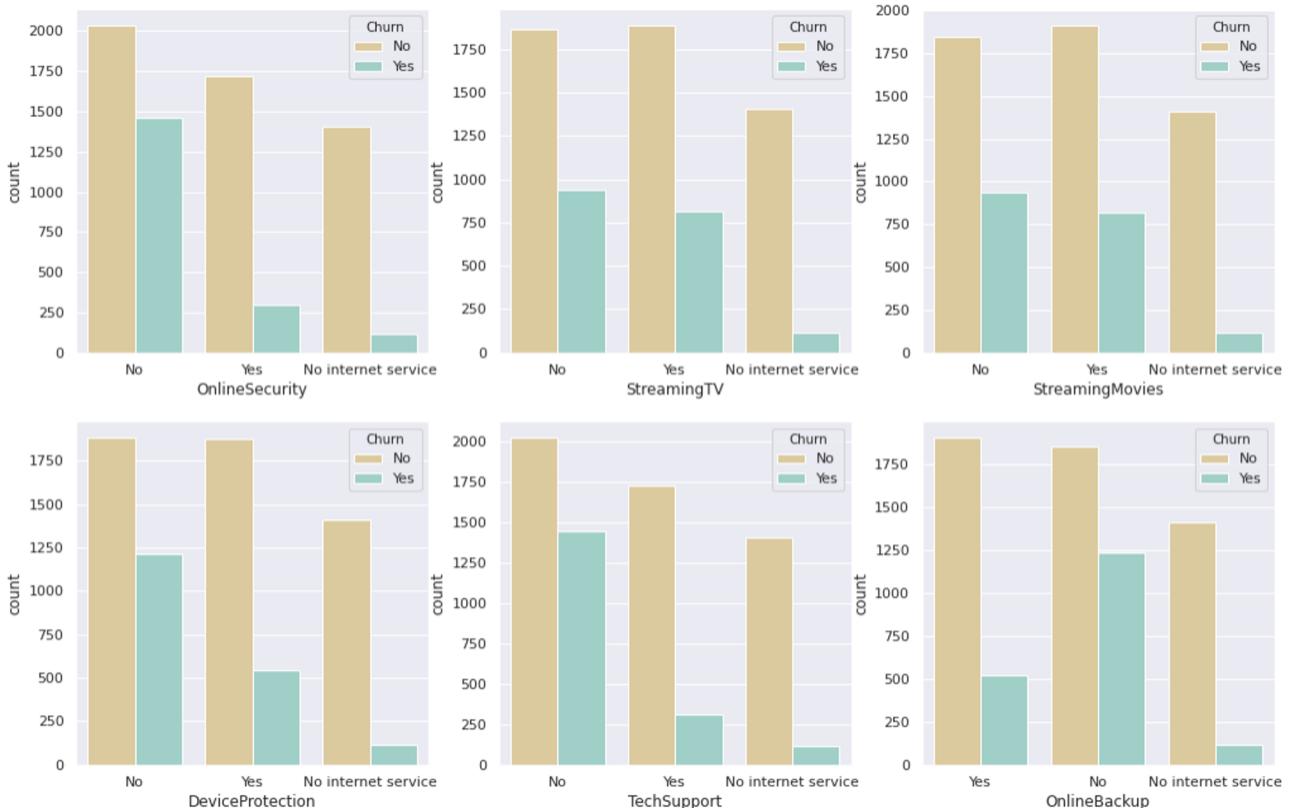
DATA EDA

**Look!
High correlations
between predictors!**



DATA EDA

Their distributions look similar!



02 DATA PREPROCESSING

FEATURE ENGINEERING

Numeric: Polynomials

Interaction Terms

Log features

Categorical: One-hot recoding

STANDARDIZATION

Technique: Min_Max_Scaler

84)

UNDERSAMPLING



Shape: (7043,21) -> (7032,84)

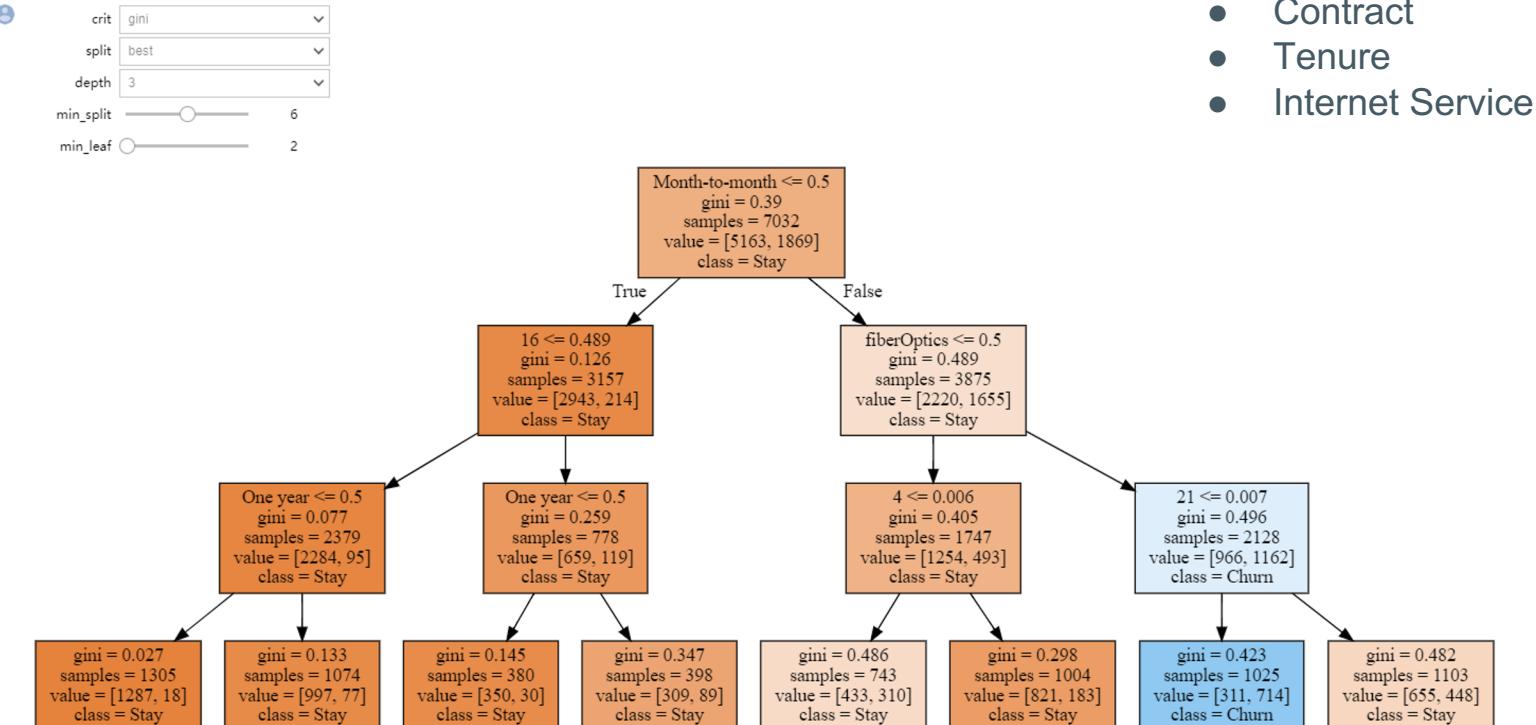
-> X: (3738,

Y: (3738, 1)

03 MODELING: INTERACTIVE TREE



Any insight?



MODELING: NEURAL NETWORK

Train

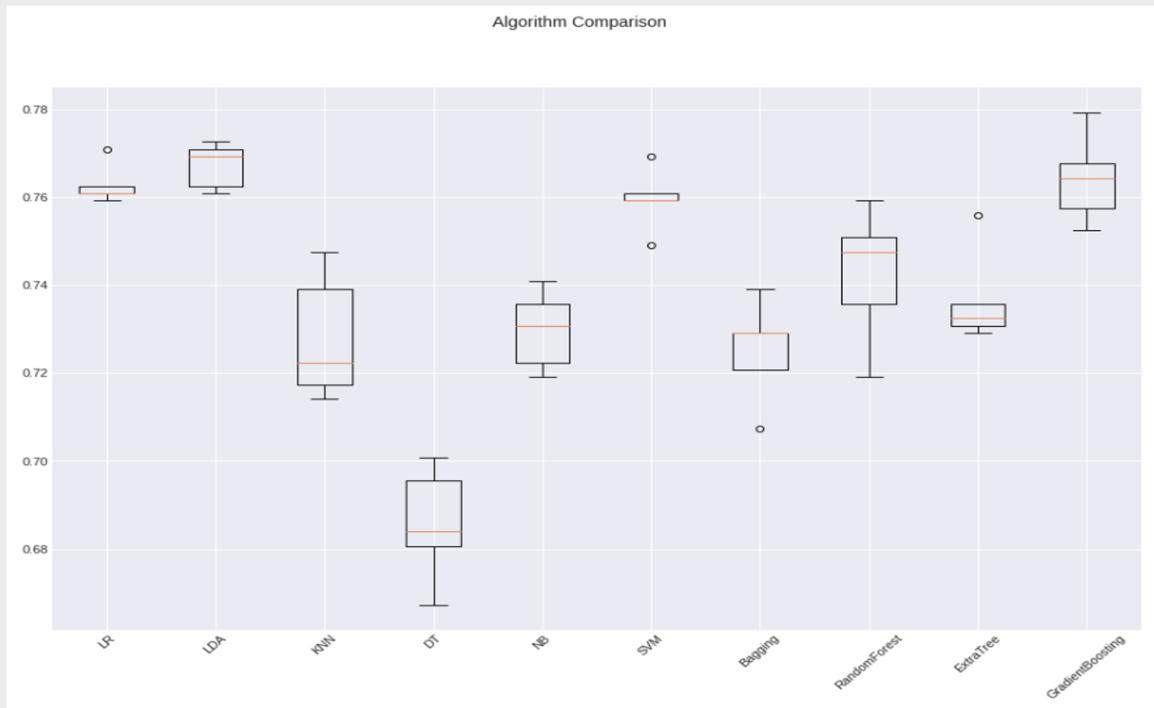
[[1211 284] [151 1344]]		precision	recall	f1-score	support
0	0.89	0.81	0.85	1495	
1	0.83	0.90	0.86	1495	
				0.85	2990
accuracy					2990
macro avg	0.86	0.85	0.85	2990	
weighted avg	0.86	0.85	0.85	2990	

Test

[[275 99] [79 295]]		precision	recall	f1-score	support
0	0.78	0.74	0.76	374	
1	0.75	0.79	0.77	374	
				0.76	748
accuracy					748
macro avg	0.76	0.76	0.76	748	
weighted avg	0.76	0.76	0.76	748	

Kind of
Overfitting ?

MODELING: SPOT CHECKING



Spot Check

- LR: 0.763
- LDA: 0.767
- KNN: 0.728
- DT: 0.686
- NB: 0.730
- SVM: 0.760
- BAGGING: 0.725
- RF: 0.742
- ET: 0.737
- GB: 0.764

MODELING

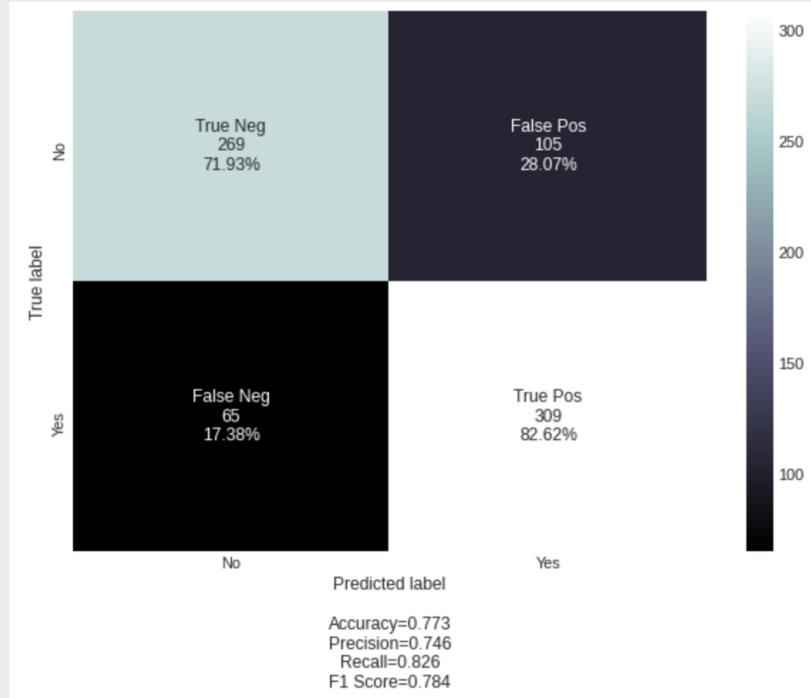


Hyperparameter tuning

Grid Search Model	Test set accuracy score for best params
LR	0.774
LDA	0.773

BEFORE

	precision	recall
0	0.79	0.76
1	0.77	0.80

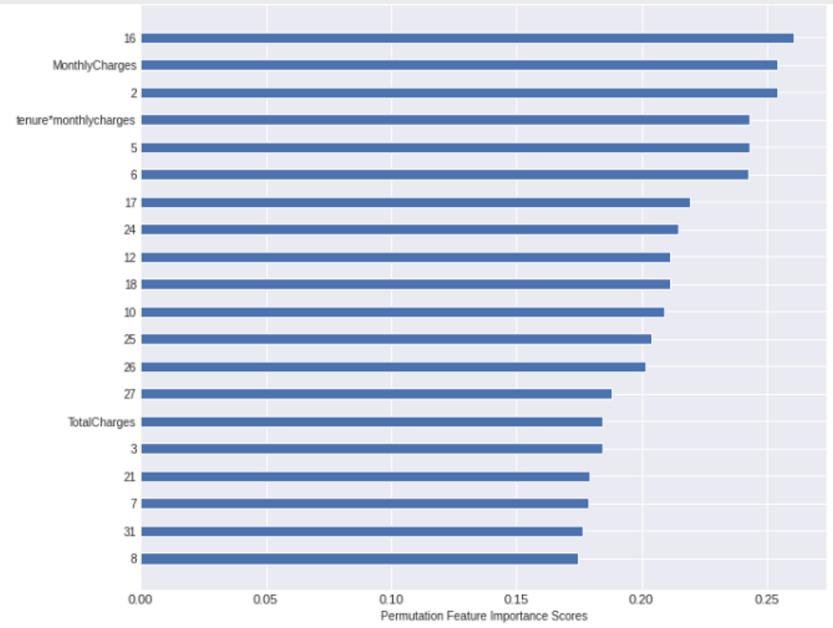


MODELING

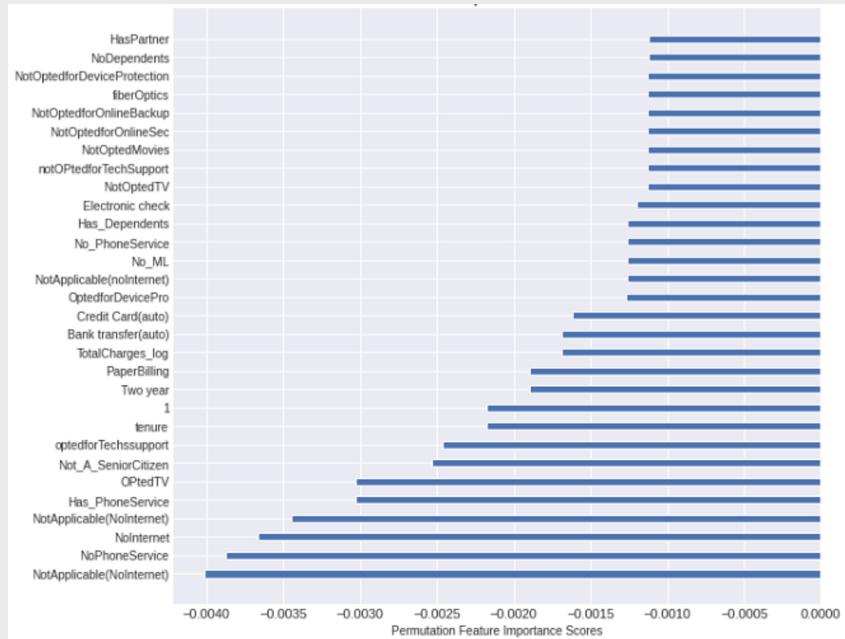


Permutation Feature Importance

TOP 20



LAST 30

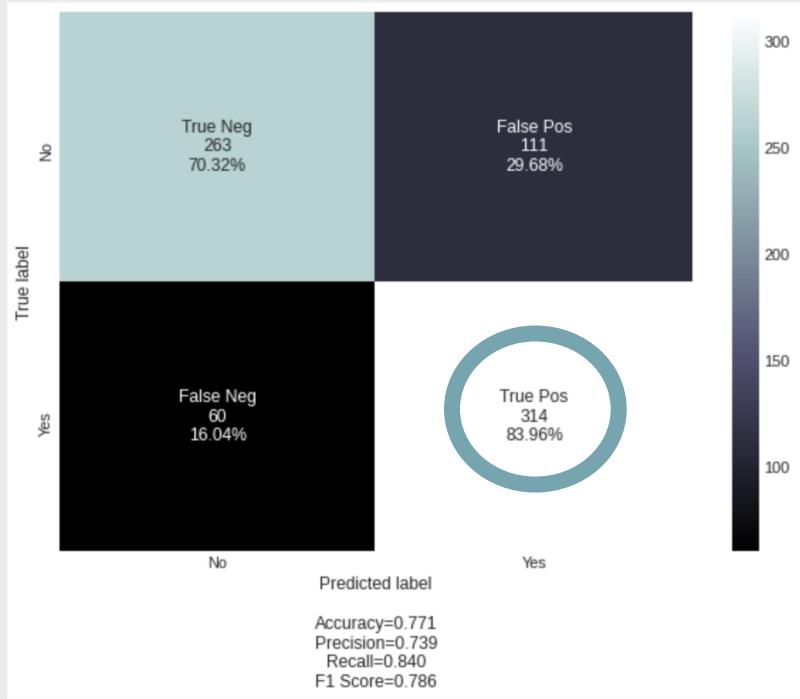


MODELING

Refining the model

Two ways:

1. Only including the most significant variables into model
Doesn't work :(
1. Delete 30 most insignificant variables
improves recall (YES-prediction): increases 1%!



04 RESULT

Model Comparison



Logistic Regression Model	Total Accuracy (Test)	Recall of YES (Test)	Recall of NO (Test)
Before Grid Search	0.78	0.80	0.76
After Grid Search	0.77	0.82	0.73
Feature Importance (top20)	0.74	0.75	0.71
Feature Importance (drop last 30)	0.78	0.83	0.72

Model	Neural Network	Logistic Regression
Train: Accuracy	.85	.77
Test: Accuracy	.74	.78
Train: Recall of YES	.90	.82
Test: Recall of YES	.80	.83

DECISION TREE SHAP EXPLANATION IMPORTANCE?

Permutation Important Test Might not do a good job
on:

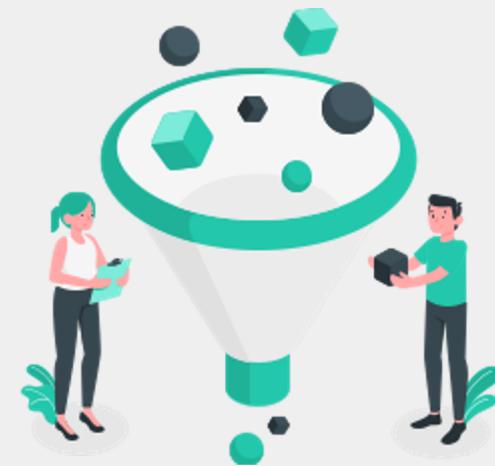
imbalance features:

An extreme example:

0
0
0
0
0
0
0
0
0
1

→

0
0
1
0
0
0
0
0
0
0



RECOMMENDATION

- Keep tracking the prediction of models and **identify risk** in advance
- Retain our clients:
 1. For **new clients**, try to establish long-term relationships (Contract,Tenure) to increase customer retention
 2. For **existing customers**, provide rewards to increase customer loyalty
 3. For **high-value customers**, provide customized plans and personalized services



THANKS

