

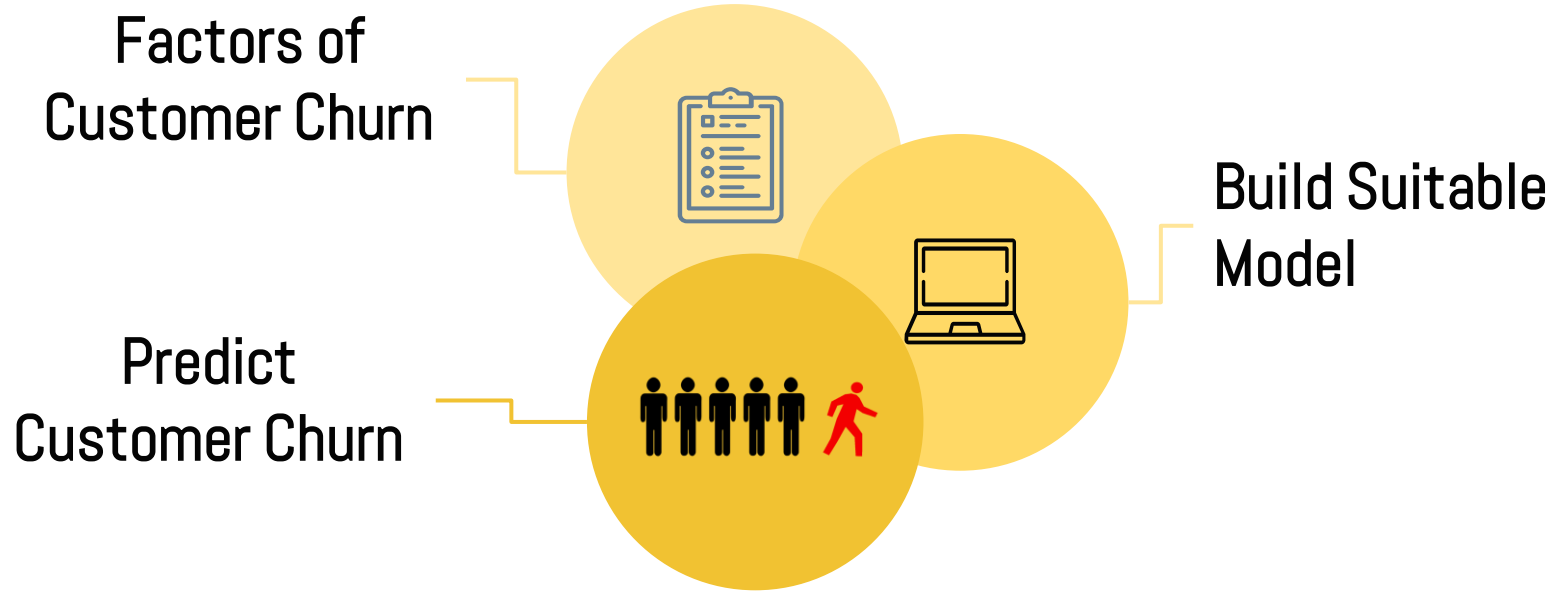
Predicting Customer Churn in the Telecommunications Industry

Why Predict Customer Churn?

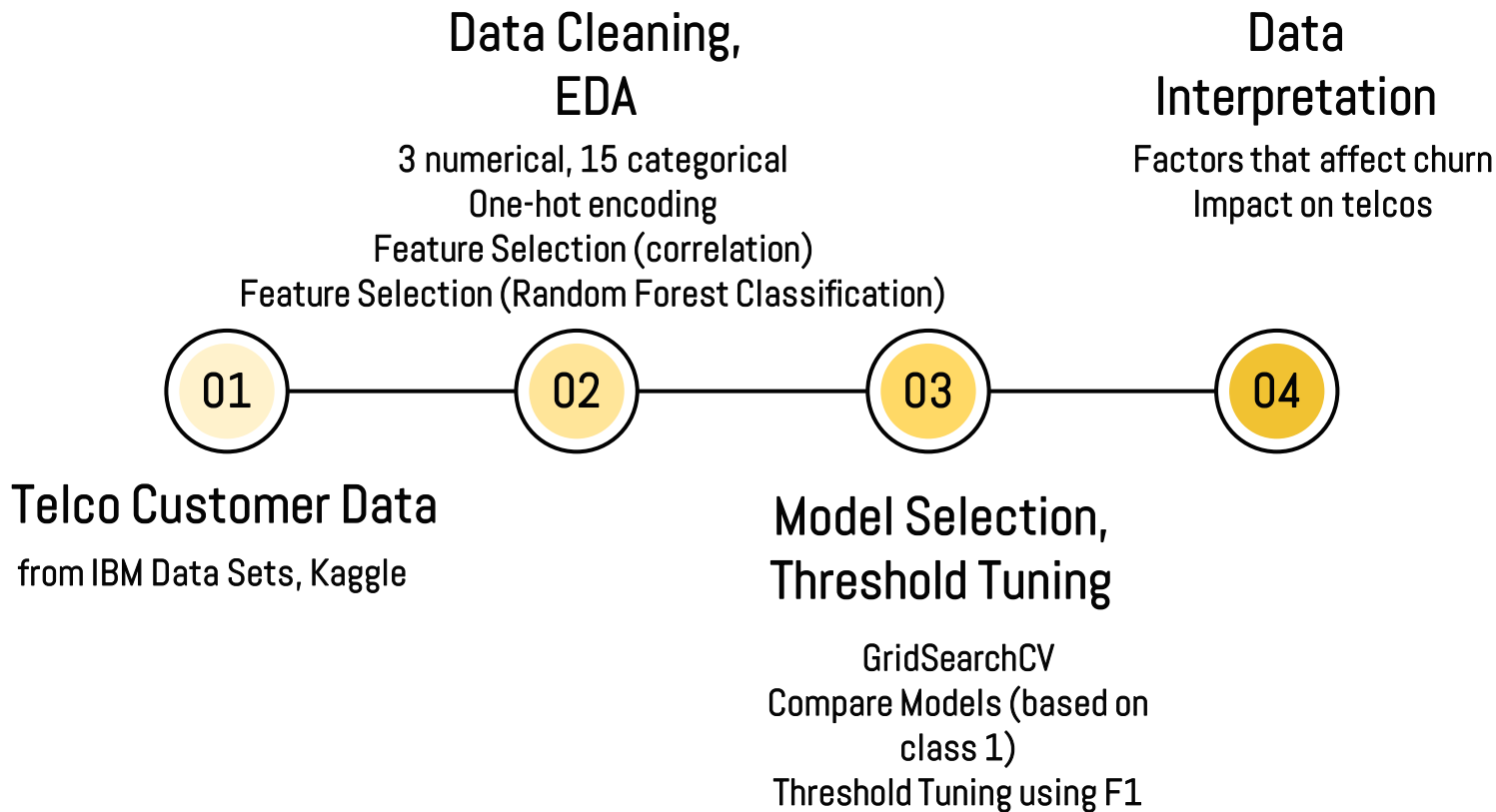
- Getting new customers is much more expensive than retaining existing ones. Some studies have shown that it costs six to seven times more to acquire a new customer than to keep an existing one.
- Not only do loyal customers help sales, but they are also more likely to buy high-margin supplemental goods and services.
- According to *BeyondPhilosophy.com*:
“Loyal customers reduce costs associated with consumer education and marketing, especially when they become Net Promoters for your organization.”
- Hence it is important to be able to proactively determine the customers most at risk of leaving and take preventative measures against this.

Project Goals

3



Project Stages





Tools Used

 pandas

matplotlib 

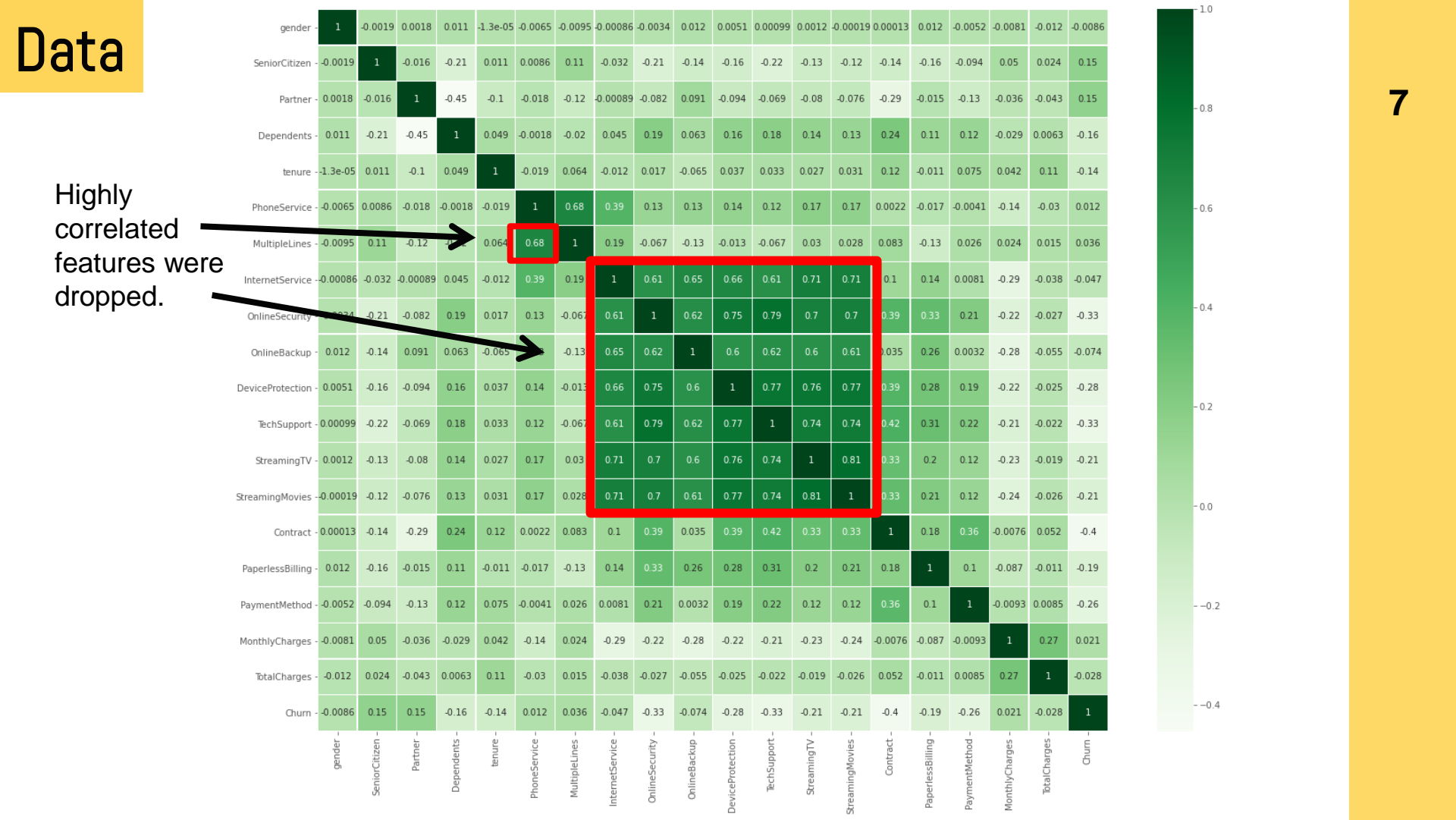
 scikit
learn  statsmodels

 Seaborn

 python™

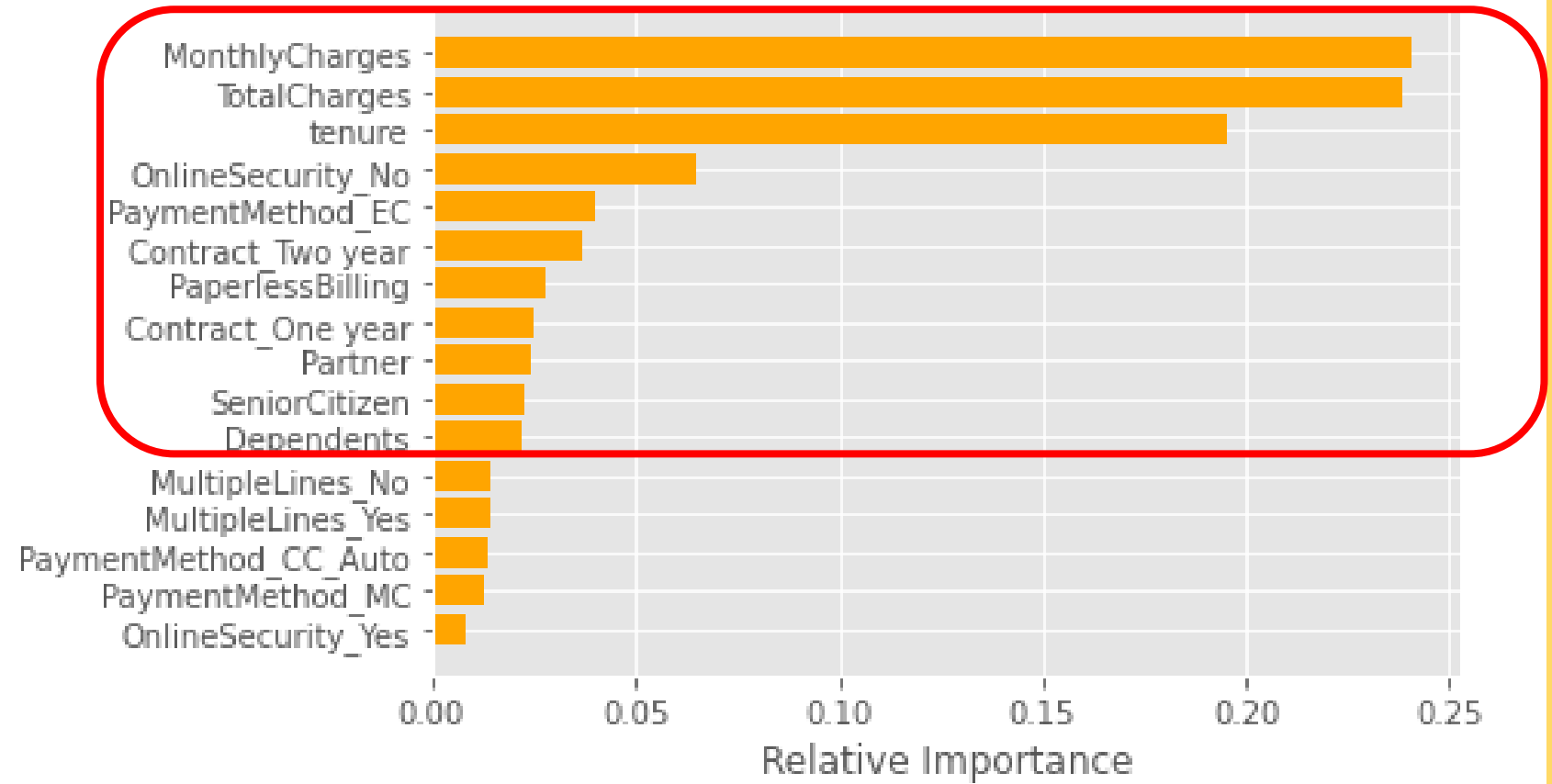
Data Clean and Exploratory Data Analysis

- Data set has 26.5% churn.
- Check that each customer id is unique.
- Check for data set imbalance.
- Replace blank spaces in total charges column, convert to numerical data type.
- Dropped 7 features that have high correlation values >0.6 , and 1 feature of little effect on Churn.
- 9 Categorical features (One-hot encoding), 3 Numerical features
- 16 feature selection down to 11 features through Random Forest Classification.

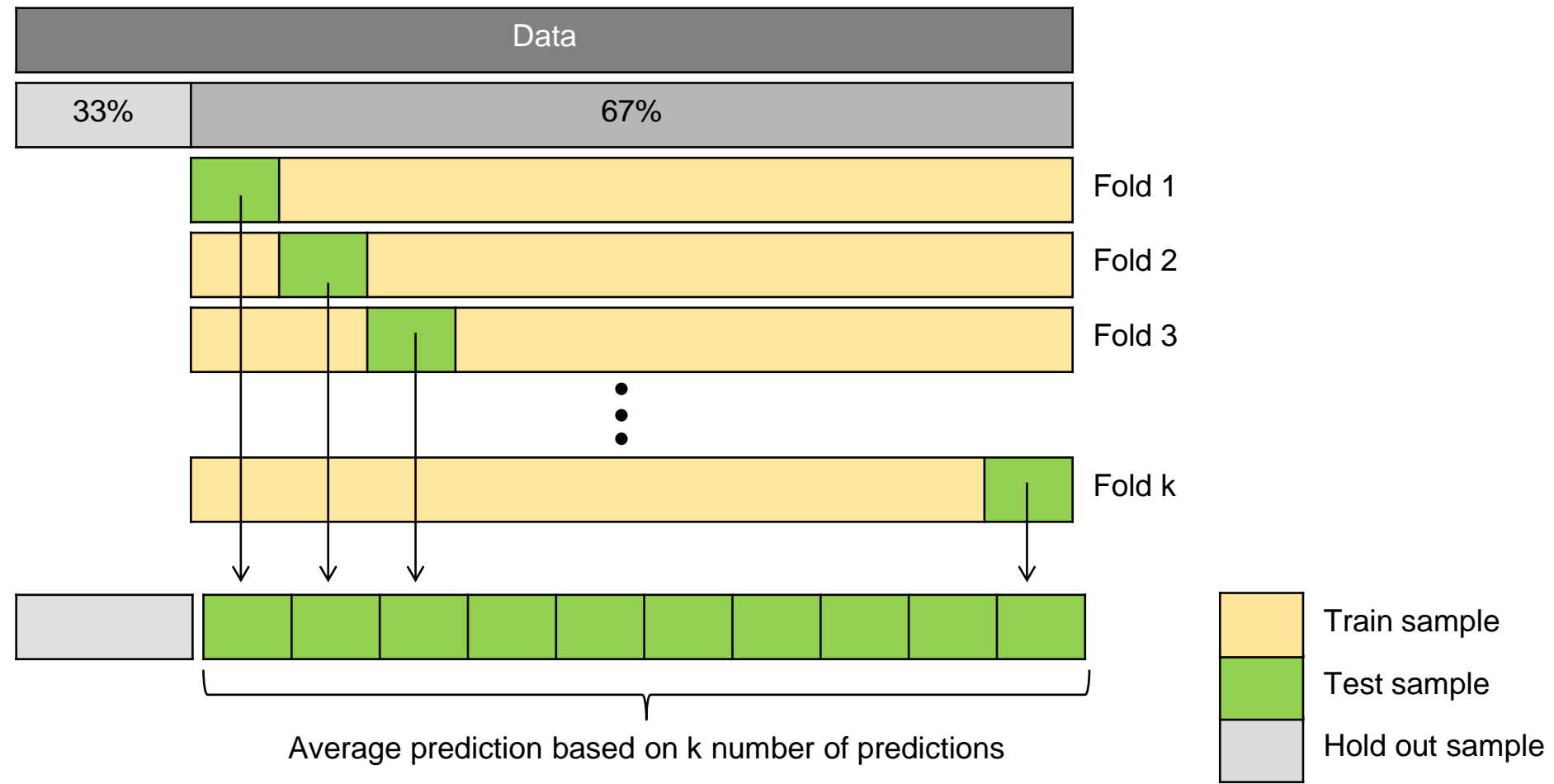


Features selected by
Random Forest Classification

Feature Importances

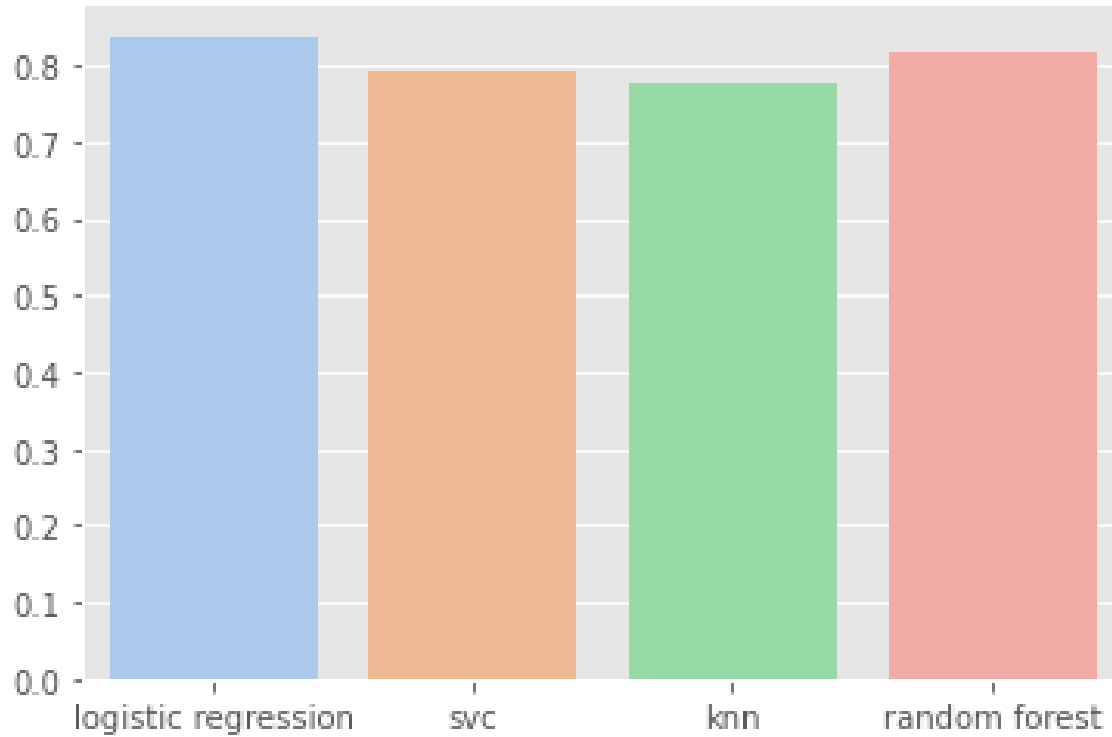


Each Model undergoes GridSearchCV using k-fold Cross Validation Procedure



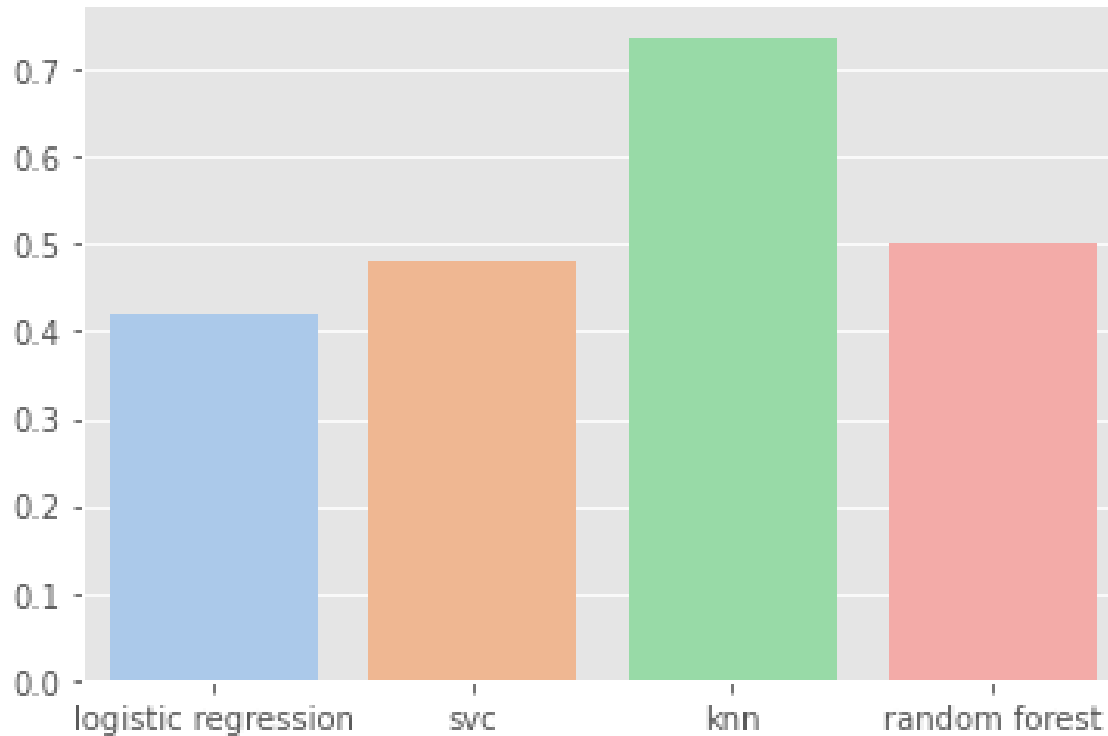
Data

AUC ROC

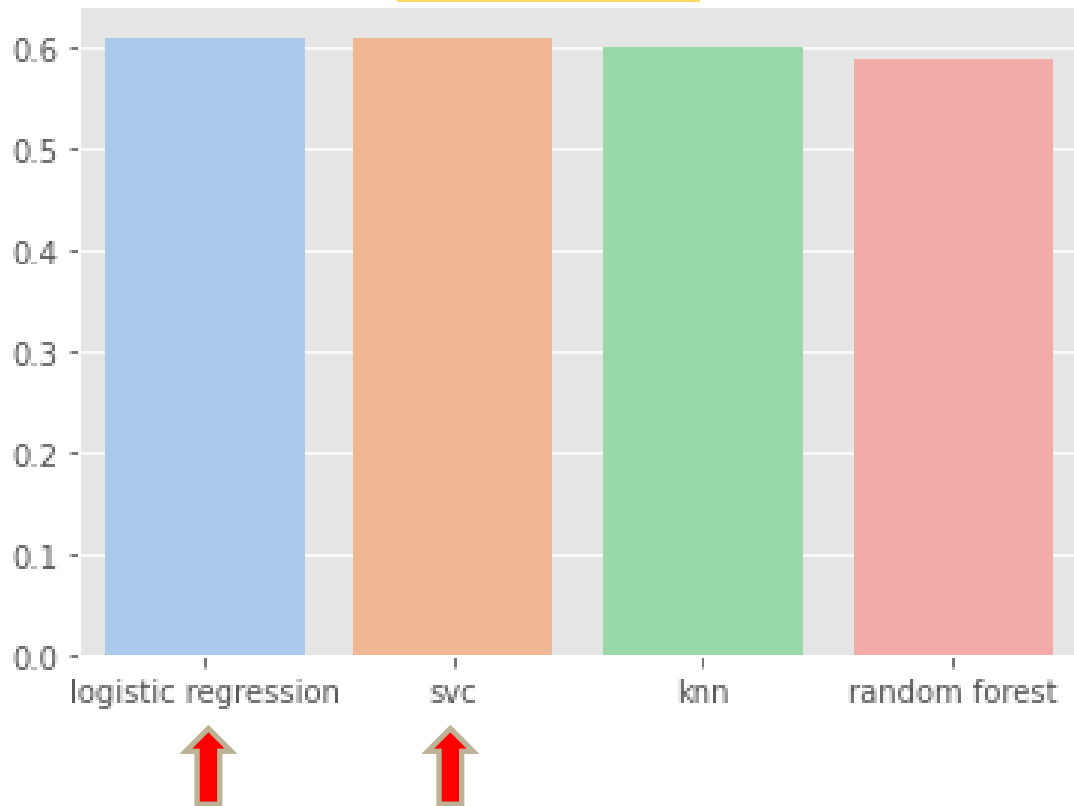


Data

Log Loss

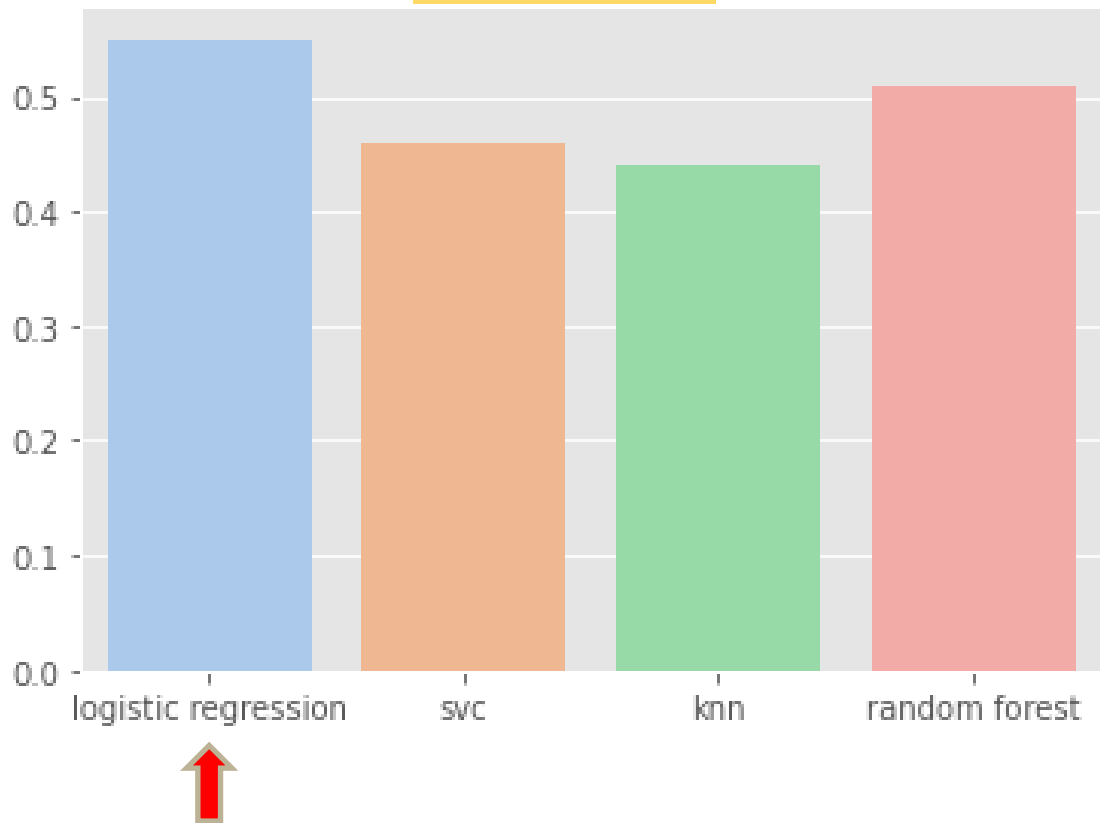


Precision

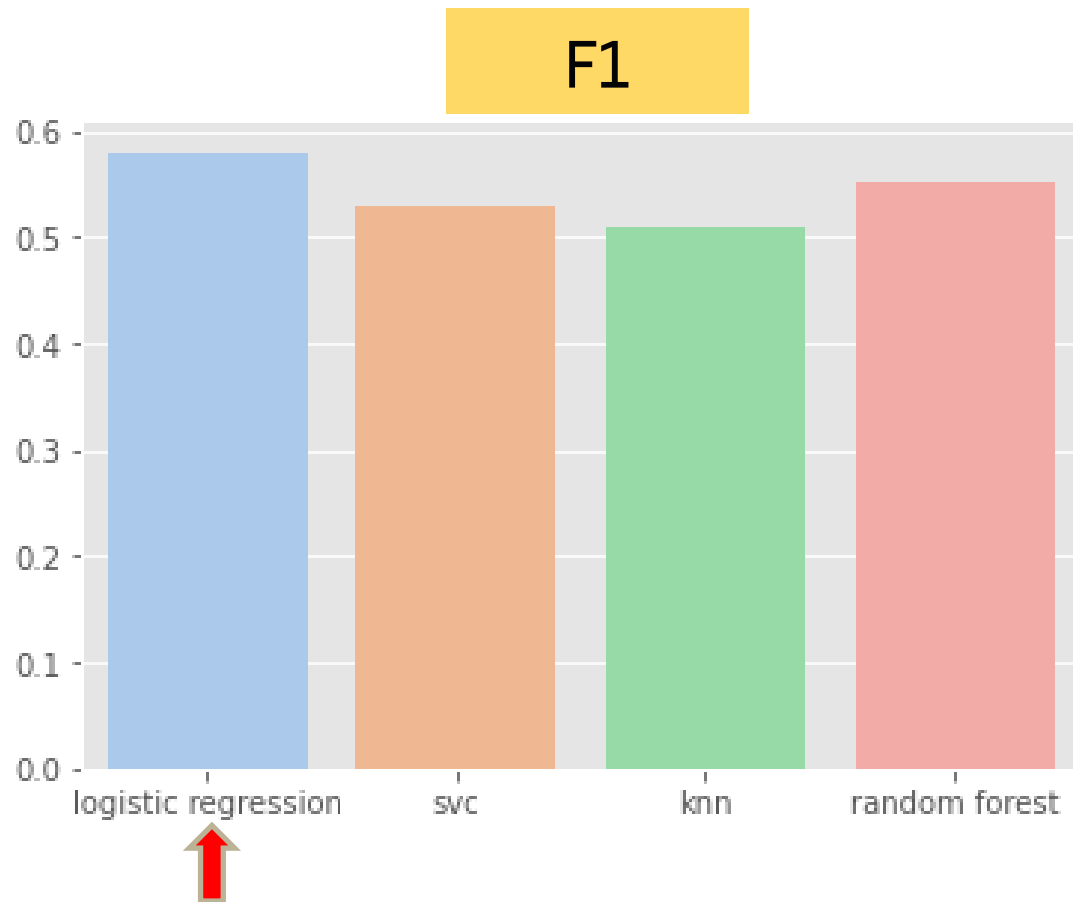


Data

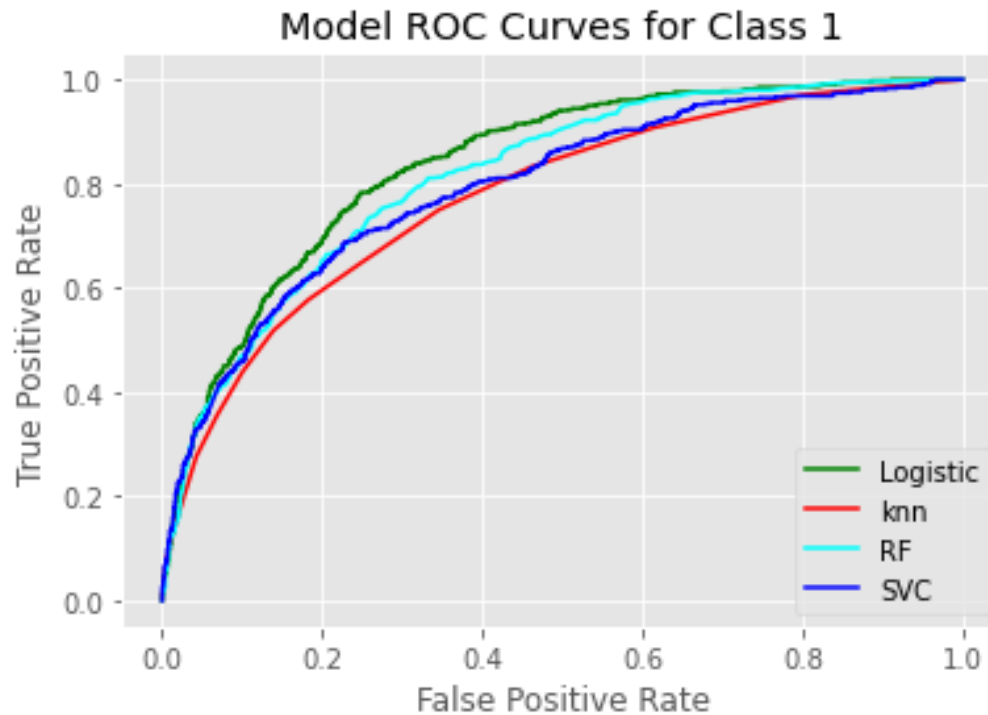
Recall



Data



Tuning Hyperparameters from 10-fold GridSearchCV



Model	Parameters	AUC ROC	Log Loss	F1
Logistic Regression	C = 5.0941, Penalty = L2, Solver = lbfgs	0.838	0.42	0.58
KNN	n neighbours = 15	0.775	0.7371	0.51
Random Forest	n estimators = 300, criterion: 'gini'	0.815	0.5027	0.55
SVC	C = 1, Kernel = Linear , Gamma =1	0.793	0.4811	0.53

Tuning threshold in Logistic Regression Model

$$\text{F-Measure} = (2 * \text{Precision} * \text{Recall}) / (\text{Precision} + \text{Recall})$$

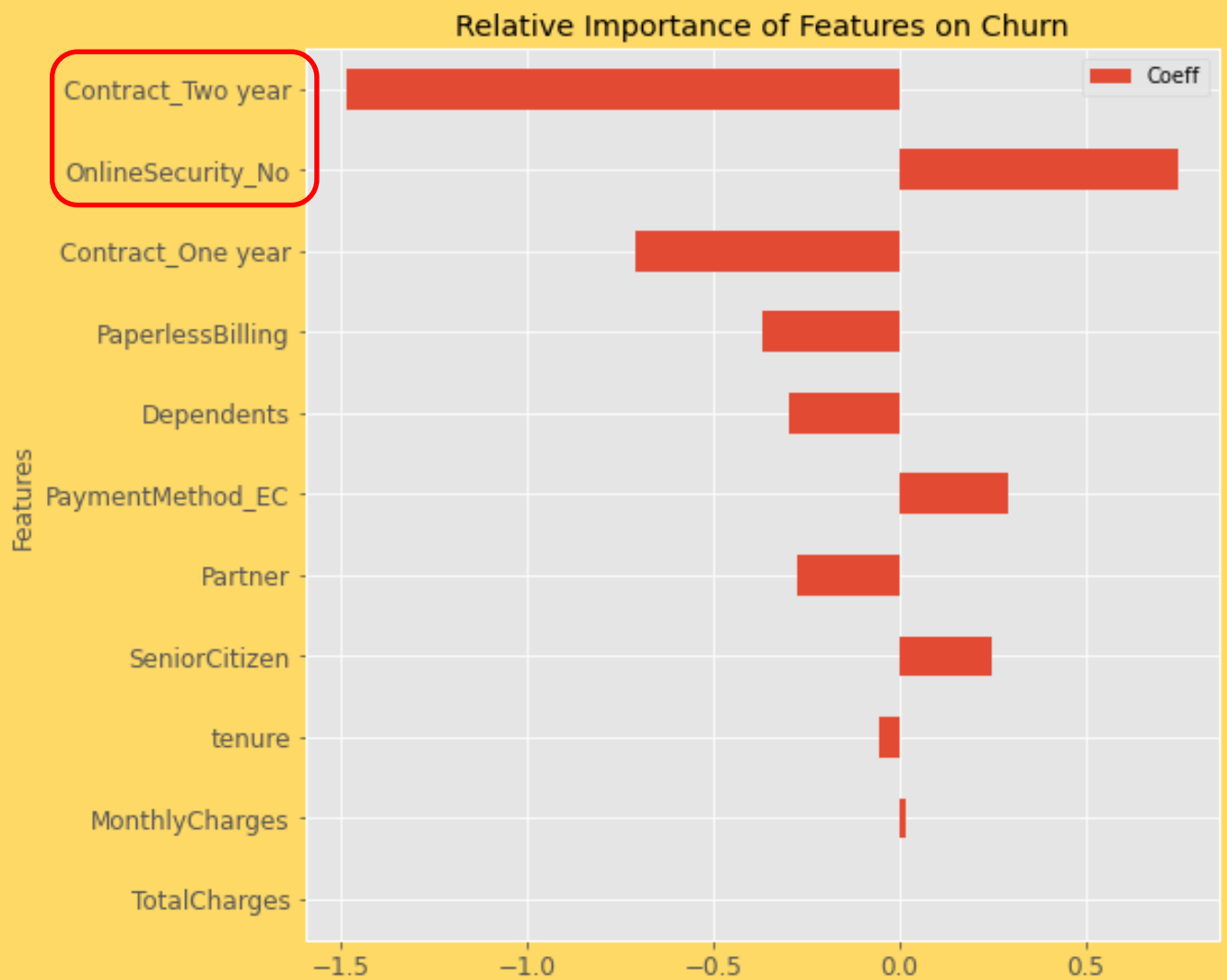
Threshold = 0.5			
Class	Precision	Recall	f1
1	0.61	0.55	0.58

Based
on
fmax



Threshold = 0.361258			
Class	Precision	Recall	f1
1	0.52	0.78	0.62

Top Features



To reduce customer churn, telcos need to focus on:

- 1) Longer term contract plans for customers.
- 2) Provide online security for customers.

THANKS!

Do you have any questions?

syleo22@gmail.com

https://github.com/syleo22/SiuYin_Projects

<https://www.linkedin.com/in/syleo/>



CREDITS: This presentation template was created
by **Slidesgo**, including icons by **Flaticon**, and
infographics & images by **Freepik**

Please keep this slide for attribution