

# データ分析課題

Telco Customer Churn

# 目次

- ▶ 分析概要と目的
- ▶ データ概要
- ▶ 分析タスク01-モデル学習・評価
- ▶ 分析タスク02-施策提示
- ▶ 今後の課題
- ▶ Appendix



# 01

## 分析概要と目的

# ▼ 分析概要と目的

## 概要と目的

あなたは、電話会社に勤務するデータ分析担当者です。カスタマーサクセス部から契約顧客の解約率の高さが課題になっていると相談を受けており、何かしらの施策を打つ必要があります。

社内で収集したデータを活用し、下記のタスクを実行してください。

タスク実行後は、事業部に対して報告を行ってください。

1. 解約する可能性の高そうな顧客を判別するモデルを作る
2. 解約率を下げるための施策を提示する

データセットはkaggleのTelco Customer Churnを使用する。

02

データ概要

# ▼ データ概要

【図表1】 データセット(一部)

customerID	gender	SeniorCitizen	Partner	Dependents	tenure	PhoneService	MultipleLines	InternetService	OnlineSecurity	OnlineBackup	DeviceProtection
7590-VHVEG	Female	0	Yes	No	1	No	No phone service	DSL	No	Yes	No
5575-GNVDE	Male	0	No	No	34	Yes	No	DSL	Yes	No	Yes
3668-QPYBK	Male	0	No	No	2	Yes	No	DSL	Yes	Yes	No
7795-CFOCW	Male	0	No	No	45	No	No phone service	DSL	Yes	No	Yes
9237-HQITU	Female	0	No	No	2	Yes	No	Fiber optic	No	No	No

TechSupport	StreamingTV	StreamingMovies	Contract	PaperlessBilling	PaymentMethod	MonthlyCharges	TotalCharges	Churn
No	No	No	Month-to-month	Yes	Electronic check	29.85	29.85	No
No	No	No	One year	No	Mailed check	56.95	1889.5	No
No	No	No	Month-to-month	Yes	Mailed check	53.85	108.15	Yes
Yes	No	No	One year	No	Bank transfer (automatic)	42.3	1840.75	No
No	No	No	Month-to-month	Yes	Electronic check	70.7	151.65	Yes

使用データ：WA\_Fn-UseC\_-Telco-Customer-Churn.csv

データ件数：7043件

特徴量(カラム)の数：21カラム

解約者数：1869人(約0.27)

非解約者数：5174人(約0.73)

# 特 徴 量 ( カ ラ ム ) 概 要

【図表2】 特徴量概要

特徴量	説明
customerID	顧客固有のID
gender	性別（男性、女性）
SeniorCitizen	高齢者かどうか (1=はい、 0=いいえ)
Partner	パートナーの有無 (はい、 いいえ)
Dependents	扶養家族の有無 (はい、 いいえ)
tenure	在籍期間(月)
PhoneService	電話サービスの有無 (はい、 いいえ)
MultipleLines	複数回線の有無 (はい、 いいえ、 電話サービスなし)
InternetService	インターネット・サービス・プロバイダー (DSL、 光ファイバー、 いいえ)
OnlineSecurity	オンライン・セキュリティの有無 (はい、 いいえ、 インターネット サービスなし)
OnlineBackup	オンライン・バックアップの有無 (はい、 いいえ、 インターネット サービスなし)
DeviceProtection	デバイス保護機能の有無 (はい、 いいえ、 インターネット サービスなし)
TechSupport	テクニカル・サポートの有無 (はい、 いいえ、 インターネット サービスなし)
StreamingTV	ストリーミングTVサービスの有無 (はい、 いいえ、 インターネット サービスなし)
StreamingMovies	ストリーミング映画サービスの有無 (はい、 いいえ、 インターネット サービスなし)
Contract	契約形態 (月単位、1年単位、2年単位)
PaperlessBilling	ペーパーレス請求の有無 (はい、 いいえ)
PaymentMethod	お支払い方法 (電子小切手、 郵送小切手、 銀行振込(自動)、 クレジットカード(自動))
MonthlyCharges	月額料金(ドル)
TotalCharges	総支払額(ドル)
Churn	解約の有無 (はい、 いいえ)

03

# 分析タスク01

- モデル学習・評価



# ▼ 分析タスク01

## 分析タスク01

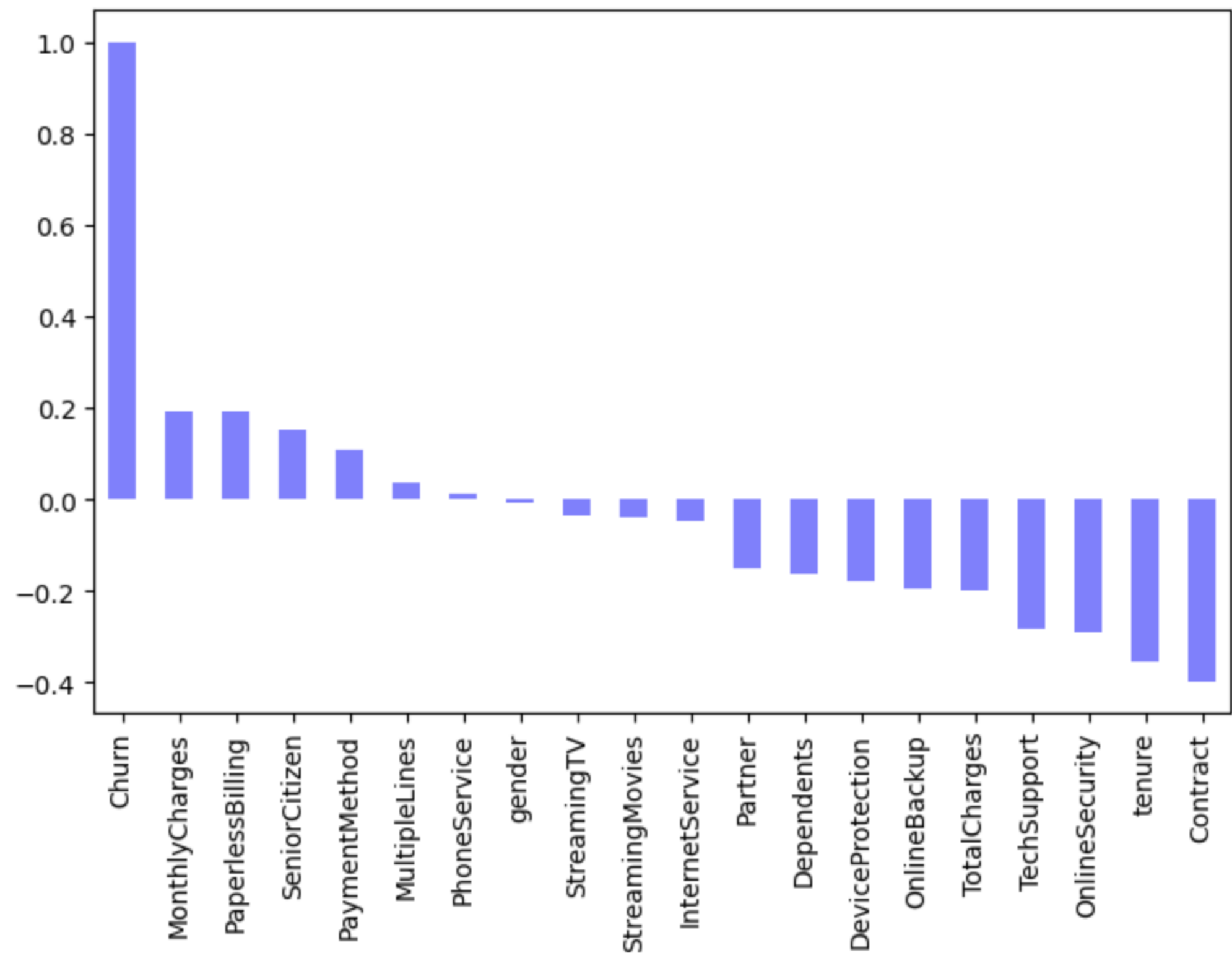
解約する可能性の高そうな顧客を判別するモデルを作る

### 前提

- ✓ データ分析：データの特性や構造を理解するために、探索的データ分析(EDA)を実施する。
- ✓ モデル：データセットに最も適したアルゴリズムを見つけるため、複数の分類モデル(ロジスティック回帰、サポートベクターマシン、決定木、ランダムフォレスト、XGBoost、LightGBM)を試行し、精度が最も高いモデルを選択する方針とした。
- ✓ 評価指標：モデルの精度を決める評価指標は再現率(Recall)を使用する。解約したユーザーのうち、モデルが正確に特定できたユーザーの数を示す指標である。解約するユーザーを見逃さないようにしたいため、重視すべき指標とした。
- ✓ 目標精度：再現率80%以上を目標とする。再現率80%の場合、解約するユーザーの80%は予測できるモデルということである。

# 探索的データ分析(EDA)

【図表3】 「Churn(解約)」と特徴量との相関

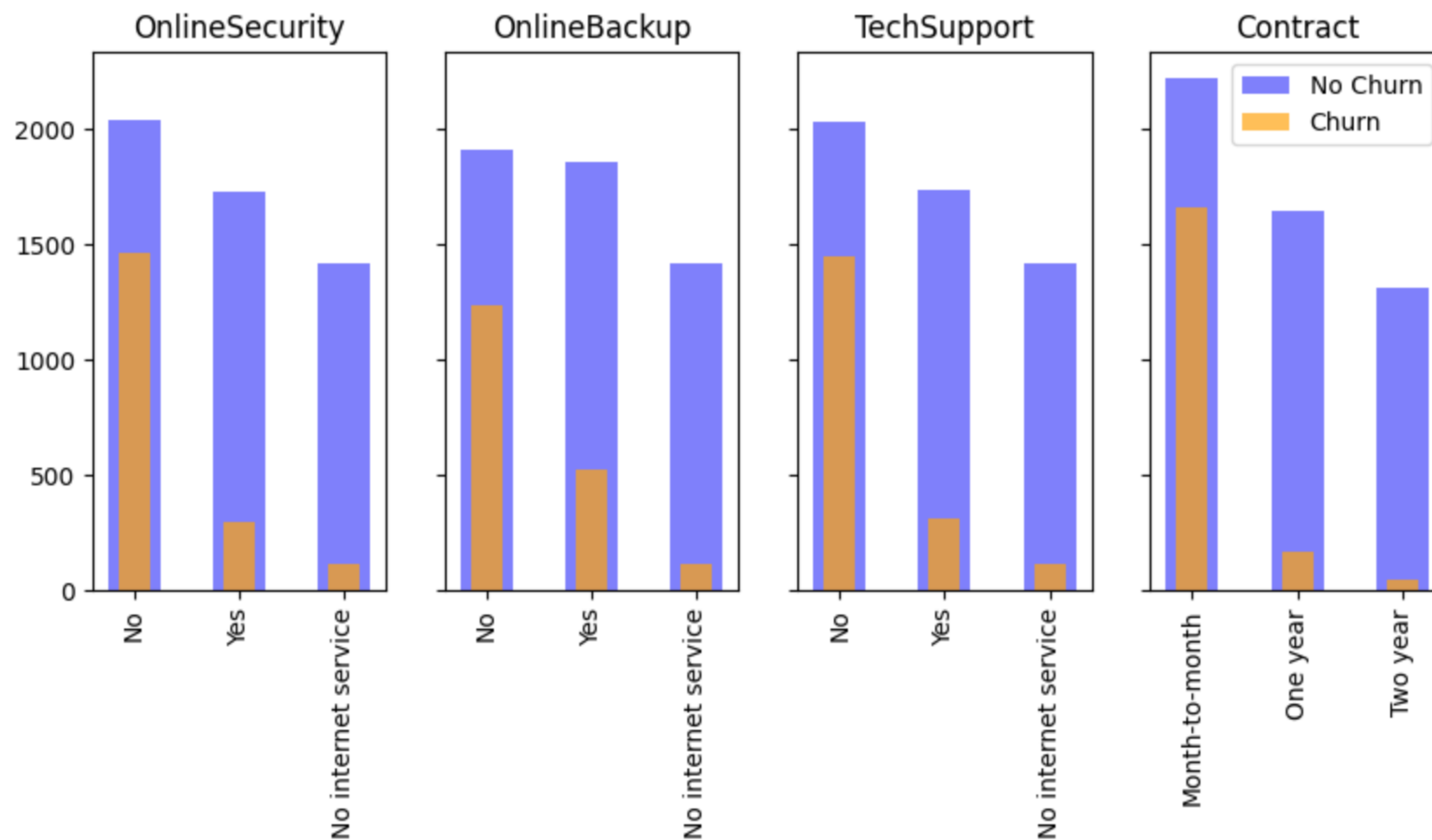


相関分析を実施しChurn(解約)と特徴量との相関を確認した。(図表3)  
解約と相関が強かったのは以下の特徴量であった。

- Contract(契約形態)
- tenure(在籍期間)
- OnlineSecuriy(オンライン・セキュリティの有無)
- TechSupport(テクニカル・サポートの有無)
- TotalCharges(総支払額)
- OnlineBackup(オンライン・バックアップの有無)
- MonthlyCharges(月額料金)

# 探索的データ分析(EDA)

【図表4】 「Churn(解約)」 とカテゴリ特徴量



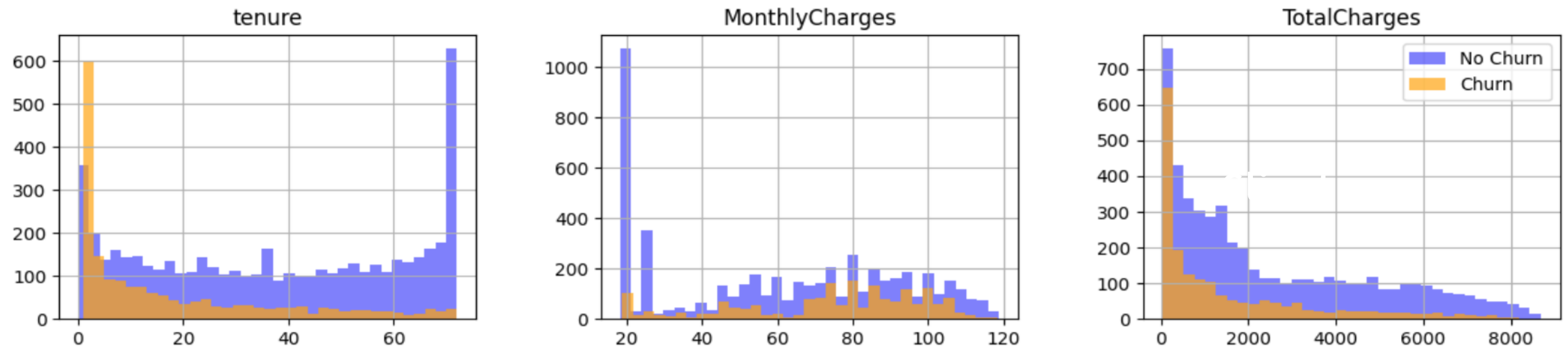
図表3でChurn(解約)と相関が強かったカテゴリ特徴量をグラフにした。

- ・インターネット・サービスに加入しているユーザーでは、オンライン・セキュリティやオンライン・バックアップ、テクニカル・サポートに加入していないユーザーは、解約する可能性が高い。

- ・Contract(契約形態)では月単位のユーザーは、解約する可能性が高い。

# 探索的データ分析(EDA)

【図表5】 「Churn(解約)」 と数値特徴量



図表3でChurn(解約)と相関が強かった数値特徴量をグラフにした。

- tenure(在籍期間)では、1ヶ月～2ヶ月の解約者が最も多い。5ヶ月までにユーザーの多くが解約している。
- MonthlyCharges(月額料金)では、70ドル～105ドルのユーザーの多くが解約している。
- TotalCharges(総支払額)では、500ドル以下のユーザーの多くが解約している。

# ▼ モデル学習・評価

【図表6】 モデル比較

Model	Accuracy	Precision	Recall	F1_Score	AUC
ロジスティック回帰	0.74	0.51	0.80	0.62	0.84
XGBoost	0.77	0.56	0.70	0.62	0.84
ランダムフォレスト	0.77	0.55	0.70	0.62	0.84
サポートベクターマシン	0.77	0.55	0.70	0.61	0.81
決定木	0.75	0.52	0.65	0.58	0.78
LightGBM	0.77	0.56	0.65	0.60	0.83

探索的データ分析結果を基にモデルの学習・評価を行った。  
モデルの学習結果は図表6のようになった。

- ・ロジスティック回帰が、再現率(Recall)80%となり最も評価の高いモデルとなった。目標精度の再現率80%以上は達成できた。

分析タスク01結果：  
評価指標である再現率(Recall)が最も高く、目標精度の再現率80%以上に達しているため、解約予測を行うモデルにロジスティック回帰を採用する。