

「機械学習プロセス徹底理解」 Step5 – アイデア4

河野 慎司

特徴量エンジニアリング 単独乗船

単独で乗船している人と、
家族など複数名で乗船している人を区別するクラスを、
特徴量として追加する。

特徴量エンジニアリング 単独乗船

```
# 単独乗船（単独==1，複数名==0）
# 'Total Family'が0人のとき1，0人でないとき0 を 'Alone'カラムに代入
train_set['Alone'] = np.where(train_set['Total Family'] == 0, 1, 0)

# 単独乗船とそれ以外のデータをgroupbyでまとめて，それぞれ生存の平均をとる
AlonePlot = train_set['Survived'].groupby(train_set['Alone']).mean()
AlonePlot
```

実行結果

```
↳ Alone  
0 0.505650  
1 0.303538  
Name: Survived, dtype: float64
```

単独(1)での乗船者よりも複数人(0)で乗船している人の方が、生存率は高くなる。