

Random Forest Classifier

Model1

- **Features:** 'TotalCharges', 'MonthlyCharges', 'tenure', 'contract_MtoM'
- **Hiperparameters:** n_estimators=1000, random_state=42

Confusion Matrix:

```
[[1385 167]
```

```
 [ 283 278]]
```

	precision	recall	f1-score	support
0	0.83	0.89	0.86	1552
1	0.62	0.50	0.55	561
accuracy			0.79	2113
macro avg	0.73	0.69	0.71	2113
weighted avg	0.78	0.79	0.78	2113

Train Score: 0.9980

Test Score: 0.7870

Difference (Train - Test): 0.2109

ROC-AUC: 0.6940

Model2

- **Features:** All
- **Hiperparameters:** n_estimators=50, random_state=42

Confusion Matrix:

```
[[1387 165]
```

```
 [ 297 264]]
```

	precision	recall	f1-score	support
0	0.82	0.89	0.86	1552
1	0.62	0.47	0.53	561
accuracy			0.78	2113
macro avg	0.72	0.68	0.70	2113
weighted avg	0.77	0.78	0.77	2113

Train Score: 0.9978

Test Score: 0.7814

Difference (Train - Test): 0.2164

ROC-AUC: 0.6821

Model3 (RandomSearchCV)

best_model

- **Features:** All
- **Hiperparameters:** 'n_estimators': 500, 'min_samples_split': 10, 'min_samples_leaf': 4, 'max_features': 'log2', 'max_depth': 2, 'class_weight': None

Confusion Matrix:

```
[[1542  10]
```

```
[ 495  66]]
```

	precision	recall	f1-score	support
0	0.76	0.99	0.86	1552
1	0.87	0.12	0.21	561
accuracy			0.76	2113
macro avg	0.81	0.56	0.53	2113
weighted avg	0.79	0.76	0.69	2113

Train Score: 0.7554

Test Score: 0.7610

Difference (Train - Test): -0.0056

ROC-AUC: 0.5556

Model4 (GridSearchCV)

- **Features:** All
- **Hiperparameters:** 'class_weight': None, 'max_depth': 3, 'max_features': 'sqrt', 'min_samples_leaf': 5, 'min_samples_split': 5, 'n_estimators': 500

Confusion Matrix:

```
[[1492  60]
```

```
[ 407 154]]
```

	precision	recall	f1-score	support
0	0.79	0.96	0.86	1552
1	0.72	0.27	0.40	561
accuracy			0.78	2113
macro avg	0.75	0.62	0.63	2113
weighted avg	0.77	0.78	0.74	2113

Train Score: 0.7870

Test Score: 0.7790

Difference (Train - Test): 0.0080

ROC-AUC: 0.6179

Model5

- **Features:** All
- **Hiperparameters:** n_estimators= 500, min_samples_split= 2, min_samples_leaf= 4, max_features= 'sqrt', max_depth= 5, class_weight= 'balanced', random_state=42

Confusion Matrix:

```
[[1099 453]
 [ 103 458]]
```

	precision	recall	f1-score	support
0	0.91	0.71	0.80	1552
1	0.50	0.82	0.62	561
accuracy			0.74	2113
macro avg	0.71	0.76	0.71	2113
weighted avg	0.81	0.74	0.75	2113

Train Score: 0.7462

Test Score: 0.7369

Difference (Train - Test): 0.0094

ROC-AUC: 0.7623

Model6

- **Features:** All
- **Hiperparameters:** n_estimators=600, max_depth=5, max_features='sqrt', min_samples_leaf=3, min_samples_split=2, random_state=42
- Oversampling and undersampling: SMOTEEN

Confusion Matrix:

```
[[1107 445]
 [ 107 454]]
```

	precision	recall	f1-score	support
0	0.91	0.71	0.80	1552
1	0.51	0.81	0.62	561
accuracy			0.74	2113
macro avg	0.71	0.76	0.71	2113
weighted avg	0.80	0.74	0.75	2113

Train Score: 0.9327

Test Score: 0.7388

Difference (Train - Test): 0.1939

ROC-AUC: 0.7613

Model7

- **Features:** All

- **Hiperparameters:** n_estimators=500, max_depth=None, max_features='sqrt', min_samples_leaf=5, min_samples_split=5, random_state=42, class_weight = 'balanced'

Confusion Matrix:

[[1219 333]

[153 408]]

	precision	recall	f1-score	support
0	0.89	0.79	0.83	1552
1	0.55	0.73	0.63	561
accuracy			0.77	2113
macro avg	0.72	0.76	0.73	2113
weighted avg	0.80	0.77	0.78	2113

Train Score: 0.8499

Test Score: 0.7700

Difference (Train - Test): 0.0799

ROC-AUC: 0.7564

Logistic Regression

Model1

- **Features:** All
- **Hiperparameters:** random_state=42
- StandarScaler

Confusion Matrix:

[[1394 158]

[245 316]]

	precision	recall	f1-score	support
0	0.85	0.90	0.87	1552
1	0.67	0.56	0.61	561
accuracy			0.81	2113
macro avg	0.76	0.73	0.74	2113
weighted avg	0.80	0.81	0.80	2113

Train Score: 0.8045

Test Score: 0.8093

Difference (Train - Test): -0.0048

ROC-AUC: 0.7307

Model2 (SMOTE)

- **Features:** All
- **Hiperparameters:** random_state=42
- StandarScaler
- SMOTE

Confusion Matrix:

```
[[1130  422]
```

```
 [ 116  445]]
```

	precision	recall	f1-score	support
0	0.91	0.73	0.81	1552
1	0.51	0.79	0.62	561
accuracy			0.75	2113
macro avg	0.71	0.76	0.72	2113
weighted avg	0.80	0.75	0.76	2113

Train Score: 0.7782

Test Score: 0.7454

Difference (Train - Test): 0.0328

ROC-AUC: 0.7607

Model3 (SMOTEENN)

- **Features:** All
- Logarithmic transformation on TotalCharges and 'MonthlyCharges'
- **Hiperparameters:** random_state=42
- StandarScaler

- without SMOTEEN

Confusion Matrix:

```
[[1408  144]
```

```
 [ 261  300]]
```

	precision	recall	f1-score	support
0	0.84	0.91	0.87	1552
1	0.68	0.53	0.60	561
accuracy			0.81	2113
macro avg	0.76	0.72	0.74	2113
weighted avg	0.80	0.81	0.80	2113

Train Score: 0.8095

Test Score: 0.8083

Difference (Train - Test): 0.0012

ROC-AUC: 0.7210

- with SMOTEENN

```

Confusion Matrix:
[[1009  543]
 [  79 482]]

```

	precision	recall	f1-score	support
0	0.93	0.65	0.76	1552
1	0.47	0.86	0.61	561
accuracy			0.71	2113
macro avg	0.70	0.75	0.69	2113
weighted avg	0.81	0.71	0.72	2113

```

Train Score: 0.9442
Test Score: 0.7056
Difference (Train - Test): 0.2385
ROC-AUC: 0.7547

```

Model4 (RandomSearchCV)

best_model

- **Features:** All
- **Scoring:** Precision
- Logarithmic transformation on TotalCharges and 'MonthlyCharges'
- **Hiperparameters:** 'C': np.float64(1.1105192452767676), 'class_weight': None, 'max_iter': 500, 'penalty': 'l2', 'solver': 'liblinear'
- StandarScaler

```

Confusion Matrix:
[[1429  123]
 [ 356 205]]

```

	precision	recall	f1-score	support
0	0.80	0.92	0.86	1552
1	0.62	0.37	0.46	561
accuracy			0.77	2113
macro avg	0.71	0.64	0.66	2113
weighted avg	0.75	0.77	0.75	2113

```

Train Score: 0.8103
Test Score: 0.8083
Difference (Train - Test): 0.0020
ROC-AUC: 0.7216

```

Model5 (scoring= F1)

- **Features:** All
- **Scoring:** F1
- Logarithmic transformation on TotalCharges and 'MonthlyCharges'

- **Hiperparameters:** `C= np.float64(0.5908361216819946),`
`class_weight= 'balanced',max_iter= 100,penalty= 'l2',solver=`
`'newton-cg'`
- StandarScaler

Confusion Matrix:

[[1192 360]

[179 382]]

	precision	recall	f1-score	support
0	0.87	0.77	0.82	1552
1	0.51	0.68	0.59	561
accuracy			0.74	2113
macro avg	0.69	0.72	0.70	2113
weighted avg	0.78	0.74	0.75	2113

Train Score: 0.7584

Test Score: 0.7440

Difference (Train - Test): 0.0145

ROC-AUC: 0.7620

- **After threshold optimization:**

[[1010 542]

[70 491]]

	precision	recall	f1-score	support
0	0.94	0.65	0.77	1552
1	0.48	0.88	0.62	561
accuracy			0.71	2113
macro avg	0.71	0.76	0.69	2113
weighted avg	0.81	0.71	0.73	2113

Train Score: 0.7428

Test Score: 0.7104

Difference (Train - Test): 0.0324

ROC-AUC: 0.7630

Model6

pipeline

- **Features:** All
- Logarithmic transformation on TotalCharges and
'MonthlyCharges'
- **Hiperparameters:** `C= 0.01, class_weight= 'balanced',max_iter=`
`100,penalty= 'l1',solver= 'newton-cg'`
- StandarScaler

```

Confusion Matrix:
[[1428 124]
 [ 289 272]]

```

	precision	recall	f1-score	support
0	0.83	0.92	0.87	1552
1	0.69	0.48	0.57	561
accuracy			0.80	2113
macro avg	0.76	0.70	0.72	2113
weighted avg	0.79	0.80	0.79	2113

```

Train Score: 0.8014
Test Score: 0.8045
Difference (Train - Test): -0.0031
ROC-AUC: 0.7025

```

XGBoost Classifier

Model1

- **Features:** All
- **Hiperparameters:** random_state=42, use_label_encoder=False, eval_metric='logloss'

```

Confusion Matrix:
[[904 131]
 [177 197]]

```

	precision	recall	f1-score	support
0	0.84	0.87	0.85	1035
1	0.60	0.53	0.56	374
accuracy			0.78	1409
macro avg	0.72	0.70	0.71	1409
weighted avg	0.77	0.78	0.78	1409

```

Train Score: 0.9340
Test Score: 0.7814
Difference (Train - Test): 0.1526
ROC-AUC: 0.7001

```

Model2 (RandomSearchCV)

best_xgb_model

- **Features:** All
- **Hiperparameters:** 'subsample': 0.5, 'n_estimators': 100, 'min_child_weight': 7, 'max_depth': 3, 'learning_rate': 0.05, 'gamma': 0.3, 'colsample_bytree': 1.0

Confusion Matrix:

```
[[937 98]
```

```
[178 196]]
```

	precision	recall	f1-score	support
0	0.84	0.91	0.87	1035
1	0.67	0.52	0.59	374
accuracy			0.80	1409
macro avg	0.75	0.71	0.73	1409
weighted avg	0.79	0.80	0.80	1409

Train Score: 0.8166

Test Score: 0.8041

Difference (Train - Test): 0.0125

ROC-AUC: 0.7147

Ensamble - VotingClassifier

- **Features:** All
- **Hiperparameters:** Model5 of Logistic Regression and Model2 of XGBoost
- Standard Scaler
- Logarithmic transformation on TotalCharges and 'MonthlyCharges'

Train Score: 0.8078

Test Score: 0.7899

Difference (Train - Test): 0.0179

Ensemble Confusion Matrix

```
[[852 183]
```

```
[113 261]]
```

Ensemble Classification Report:

	precision	recall	f1-score	support
0	0.88	0.82	0.85	1035
1	0.59	0.70	0.64	374
accuracy			0.79	1409
macro avg	0.74	0.76	0.75	1409
weighted avg	0.80	0.79	0.80	1409

Ensemble ROC-AUC: 0.850727221059702