# TELCO CHURN PREDICTION MODEL

## A Model for Predicting Customer Retention in Telecom

**Peerapat.t**

For project's material please visit : github.com/peerapat-t

# TELCO CHURN PREDICTION MODEL

A Model for Predicting Customer Retention in Telecom

## Problem context

- Retaining a customer is generally more cost-effective than acquiring a new one, as demonstrated by a Bain & Company study.
- The study revealed that acquiring a new customer can be 5 to 7 times more expensive than retaining an existing one.
- Existing customers, being familiar with your product or service, tend to exhibit greater loyalty and are more inclined to maintain their business relationship with you.

## Challenges

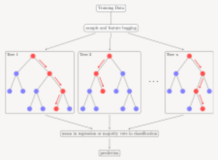
- Imbalanced dataset
- Metrics for measurement
- Threshold optimization

## Tools

python

pandas

## Methods



Random forest

LightGBM

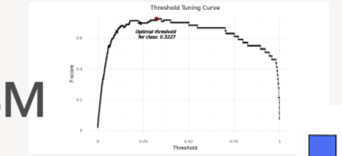
Random forest

XGBoost

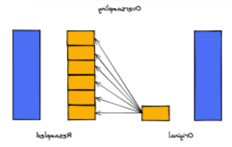
Random forest

SHAP

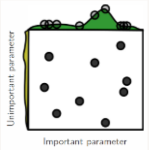
Shapley



Threshold tuning



Resampling (SMOTE, ADASYN, Random undersampling)



Randomized search

## Business impact


The model improved revenue by 164.24% compared to the 'do nothing' program and by 44.51% compared to the 'retain all' program.

# TABLE OF CONTENT

# 1. PROBLEM STATEMENT

- The cost of retaining a customer is typically much lower than the cost of acquiring a new customer.
- For example, a study by Bain & Company found that it costs 5 to 7 times more to acquire a new customer than it does to retain an existing customer.
- This is because existing customers are already familiar with your product or service and they are more likely to continue doing business with you.

# 2. BUSINESS VALUE

- A churn prediction model in a telco company provides substantial business value by forecasting which customers are likely to leave the service.
- This predictive capability allows the company to take proactive measures to retain customers, such as offering personalized incentives or addressing their concerns.
- By reducing churn, the company can achieve cost savings associated with customer acquisition, preserve existing revenue streams, and enhance overall customer satisfaction.

# 3. METHODOLOGY

### Do nothing

- Business not take any action to retain customers.

### Retain all customers

- Business take action to retain every customer.

### Use a churn prediction model

- Business use churn prediction model to predict the probability of churning and use this score to decide whether a customer will churn.
- Churn prediction can help businesses identify customers who are at risk of leaving.
- This allows businesses to take action to retain those customers, such as offering them discounts or special promotions.

| Option | Cost | Effectiveness |
| --- | --- | --- |
| Do nothing | Least expensive | Least effective |
| Retain all customers | Most expensive | Most effective |
| Use a churn prediction model | More cost-effective than retaining all customers | More effective than doing nothing |

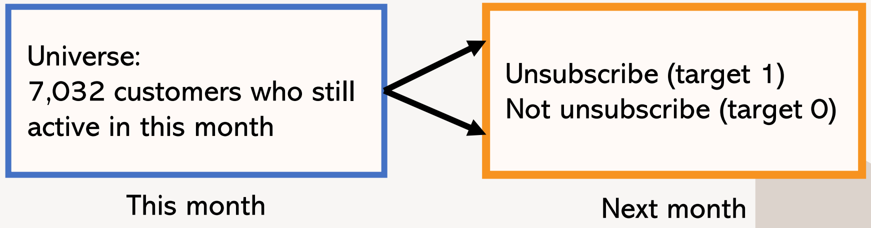
# 3. METHODOLOGY

**Target**

- There are 7,043 customers in the dataset, including churn and not-churn customers.
- Churn customers are those who have unsubscribed within the last month.

**Features**

- The dataset contains 22 features, which can be categorized into 3 groups:
  1. Services that each customer has signed up for: This includes phone, multiple lines, internet, online security, online backup, device protection, tech support, and streaming TV and movies.
  2. Customer account information: This includes how long they've been a customer, contract, payment method, paperless billing, monthly charges, and total charges.
  3. Demographic info about customers: This includes gender, age range, and if they have partners and dependents.



# 4. RESULT

## Model result

| Method | Total (Test) | Actual churn | Overspend (FP) | Save (TP) | Gains (or Loss) |
|---|---|---|---|---|---|
| Do nothing | 1,409 | 373 | 0 | 0 | -186,500 |
| Retain all | 1,409 | 373 | 1,036 | 373 | 82,900 |
| Churn prediction model | 1,409 | 373 | 253 | 302 | 119,800 |

* These calculations are based on the assumption that the cost of promotion is 100, the cost of loss is 500, and the savings per customer is 500.

- Utilizing data from the test set involving approximately 1,409 customers, the churn prediction model yielded a gain of 164.24% when compared to the 'do nothing' program, and 44.51% when compared to the 'retain all' program.

# 4. RESULT

## Feature importance

| Rank | Features | Sign | Meaning |
|---|---|---|---|
| 1 | Contract | + | Customers on contract or top-up plans are more likely to churn, possibly due to the ease of changing numbers for top-up customers. |
| 2 | Tenure | + | Long-time customers have a higher churn probability, potentially due to elderly individuals reducing phone usage to cut expenses. |
| 3 | Monthly charges | - | Customers paying lower charges are more likely to churn, influenced by industry trends towards appealing, low-priced packages from other operators. |
| 4 | Internet service | - | Customers without internet in their package are at a higher churn risk, as competitors offer inclusive internet services, reflecting its growing importance. |
| 5 | Payment method (electronic check) | - | Non-users of electronic payment for bills are more likely to churn, possibly due to the inconvenience of payment, prompting a switch to other providers. |

# 5. CONCLUSIONS/RECOMMENDATIONS

## Conclusion

- The churn prediction model, based on the analysis of a test set comprising around 1,409 customers, has demonstrated its effectiveness in reducing customer churn significantly.
- When compared to the 'do nothing' program, which involves no proactive retention efforts, the model achieved a remarkable gain of 164.24%. Additionally, in comparison to the 'retain all' program, which indiscriminately attempts to retain all customers, the model still outperformed with a gain of 44.51%. These results underscore the value of predictive modeling in identifying and mitigating customer churn.

## Recommendation

- Implement the churn prediction model as a core part of your customer retention strategy to identify at-risk customers.
- Develop personalized retention strategies based on the model's insights to optimize resource allocation.
- Regularly monitor and improve the churn prediction model to ensure its accuracy and effectiveness over time.

# 6. FUTURE WORK

- Improve model performance by creating more features and performing feature engineering.
- Experiment with different machine learning models, such as SVC and deep learning.
- Tune hyperparameters using sequential search techniques, such as Optuna.
- Segment customers using clustering and retain them with personalized promotions.

# END OF PRESENTATION

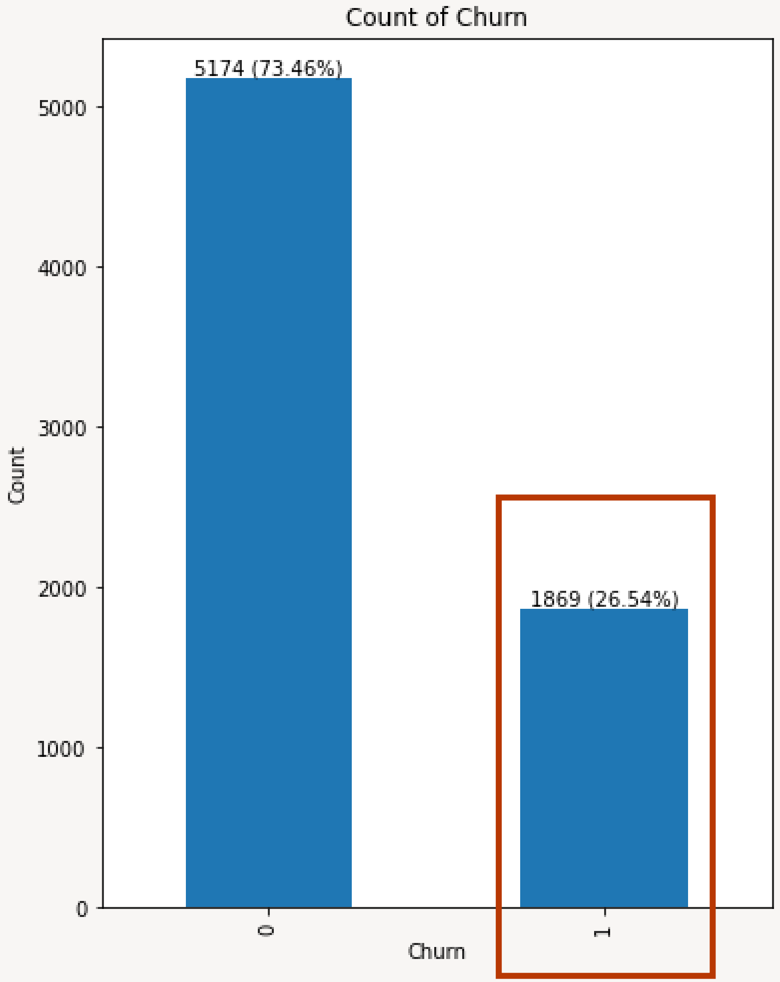
# 7. APPENDIX

## Technique

| Step | Topic | Cost |
|---|---|---|
| 1 | Dataset | • Split the data into train, test, and validation sets.<br>• Use the train set to train the model, the validation set to tune the threshold, and the test set to evaluate the model. |
| 2 | Normalized | • Use Min-Max scaling to normalize the data. |
| 3 | Resampling | • Try oversampling (SMOTE, ANASYN) and undersampling (Random undersampling). |
| 4 | Model | • Try random forest, LightGBM, and XGBoost. |
| 5 | Hyperparameter tuning | • Use RandomizedCV to find the best hyperparameters for each model. |
| 6 | Threshold tuning | • Use cost-sensitive learning to tune the threshold.<br>• Assign a cost of 5 times more to acquiring new customers than to retaining existing customers. |
| 7 | Interpreting | • Use SHAP values. |

# 7. APPENDIX

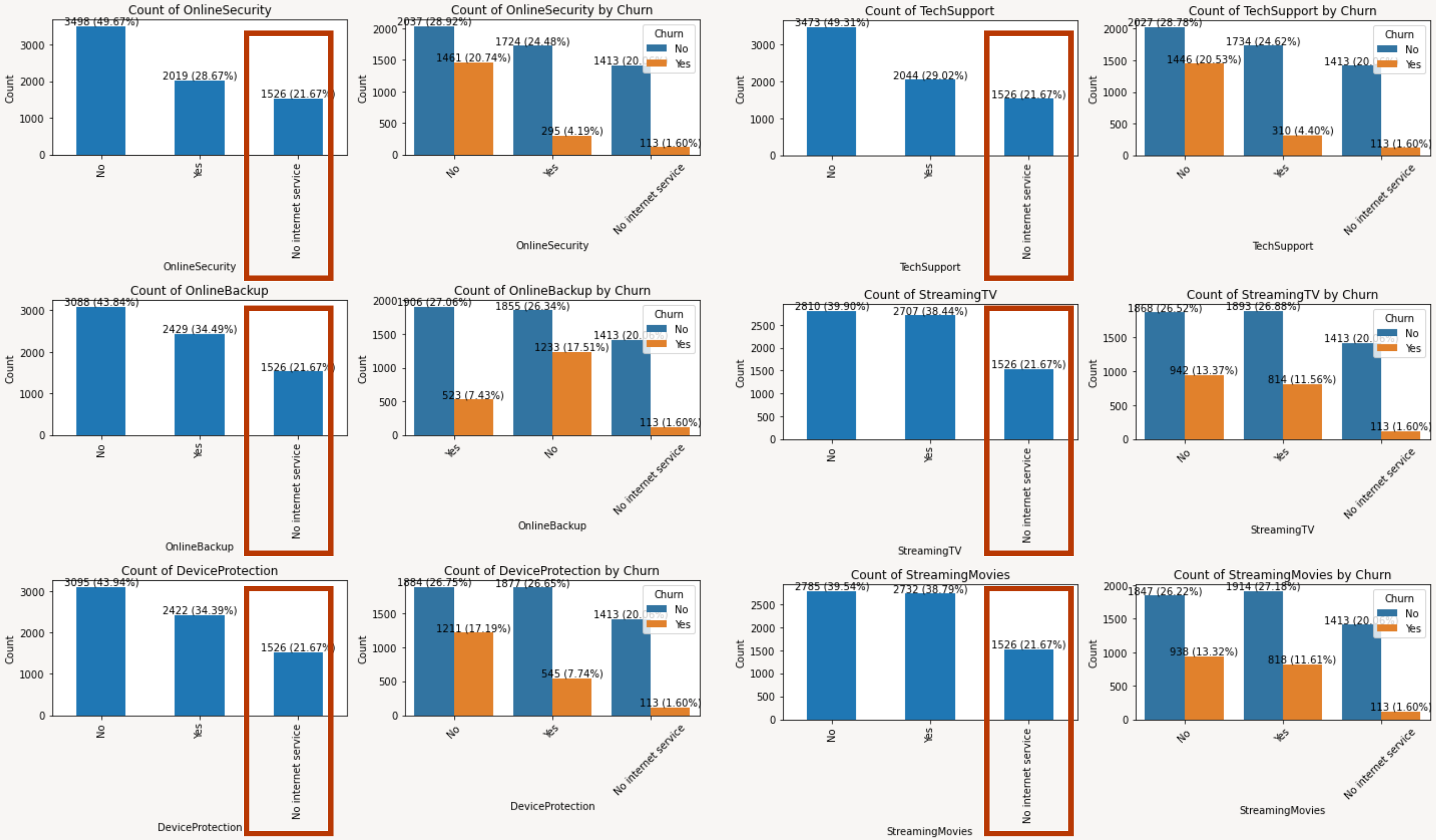
## Exploratory data analysis (EDA)




- If we look at the target distribution, There are only 27% of churn customers that the data is imbalanced.
- Resampling method or threshold tuning should be applied to handle this problem.
- Metrics like accuracy can be misleading and ineffective. Instead, it is essential to explore alternative evaluation measures, such as precision, recall, F1-score, or AUC-ROC.

# 7. APPENDIX

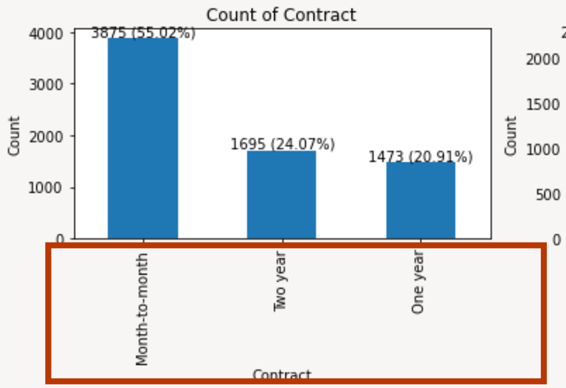
## Exploratory data analysis (EDA)

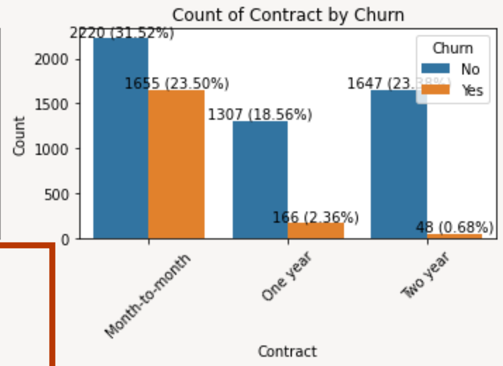



- The phrase "No internet service" appears in 6 variables, all with the same frequency of 1,526.

# 7. APPENDIX

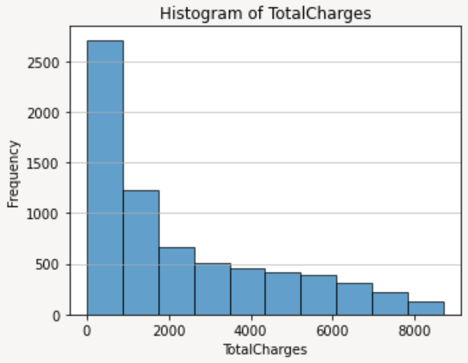
## Exploratory data analysis (EDA)

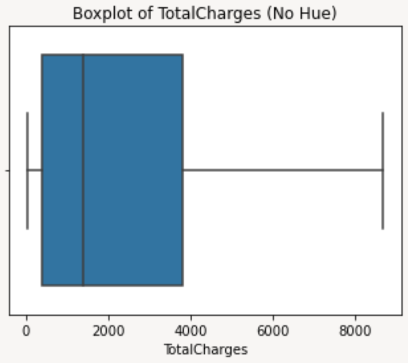


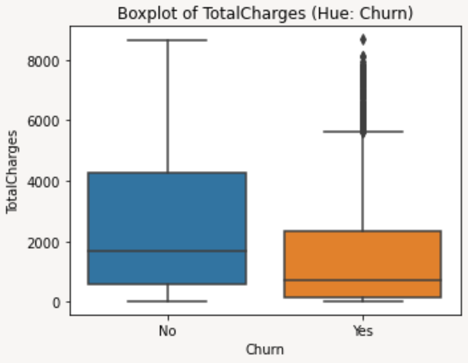




0, 1, 2

- Can represent the "Contract" variable as an ordinal variable with values 0, 1, and 2.



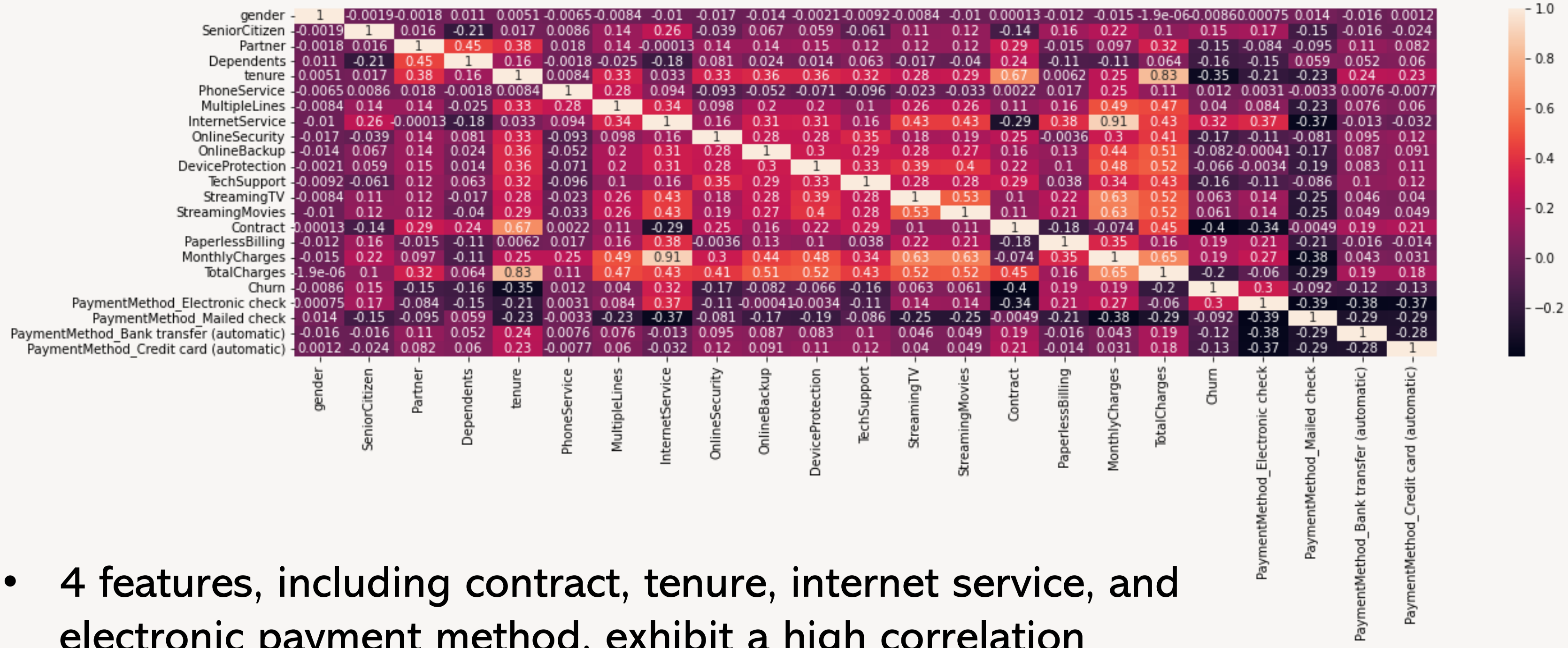




- The distribution of 'TotalCharges' is right-skewed. Use the median to replace missing values instead of the mean.

# 7. APPENDIX

## Exploratory data analysis (EDA)




```
Churn                                      1.000000
Contract                                   0.396713
tenure                                     0.352229
InternetService                            0.316846
PaymentMethod_Electronic check             0.301919
TotalCharges                               0.199037
MonthlyCharges                             0.193356
PaperlessBilling                           0.191825
OnlineSecurity                             0.171226
TechSupport                                0.164674
Dependents                                 0.164221
SeniorCitizen                              0.150889
Partner                                    0.150448
PaymentMethod_Credit card (automatic)      0.134302
PaymentMethod_Bank transfer (automatic)    0.117937
PaymentMethod_Mailed check                 0.091683
OnlineBackup                               0.082255
DeviceProtection                           0.066160
StreamingTV                                0.063228
StreamingMovies                            0.061382
MultipleLines                              0.040102
PhoneService                               0.011942
gender                                     0.008612
Name: Churn, dtype: float64
```

- 4 features, including contract, tenure, internet service, and electronic payment method, exhibit a high correlation (greater than 0.3 in absolute terms) with the target variable.

# 7. APPENDIX

## Model evaluation

| | Model | Observation | TP | TN | FP | FN | Precision | Recall | AUCROC |
|---|---|---|---|---|---|---|---|---|---|
| 0 | Dummy Model (All Churn) | 1409 | 373 | 0 | 1036 | 0 | 0.264727 | 1.000000 | 0.500000 |
| 1 | Dummy Model (All Not Churn) | 1409 | 0 | 1036 | 0 | 373 | NaN | 0.000000 | 0.500000 |
| 2 | Random Forest (SMOTE) | 1409 | 212 | 901 | 135 | 161 | 0.610951 | 0.568365 | 0.839842 |
| 3 | LightGBM (SMOTE) | 1409 | 216 | 888 | 148 | 157 | 0.593407 | 0.579088 | 0.838736 |
| 4 | XGBoost (SMOTE) | 1409 | 253 | 885 | 151 | 120 | 0.626238 | 0.678284 | 0.857387 |
| 5 | Random Forest (ADASYN) | 1409 | 208 | 894 | 142 | 165 | 0.594286 | 0.557641 | 0.835892 |
| 6 | LightGBM (ADASYN) | 1409 | 217 | 895 | 141 | 156 | 0.606145 | 0.581769 | 0.840977 |
| 7 | XGBoost (ADASYN) | 1409 | 245 | 880 | 156 | 128 | 0.610973 | 0.656836 | 0.855058 |
| 8 | Random Forest (RUS) | 1409 | 302 | 783 | 253 | 71 | 0.544144 | 0.809651 | 0.861091 |
| 9 | LightGBM (RUS) | 1409 | 306 | 764 | 272 | 67 | 0.529412 | 0.820375 | 0.857970 |
| 10 | XGBoost (RUS) | 1409 | 313 | 749 | 287 | 60 | 0.521667 | 0.839142 | 0.862201 |

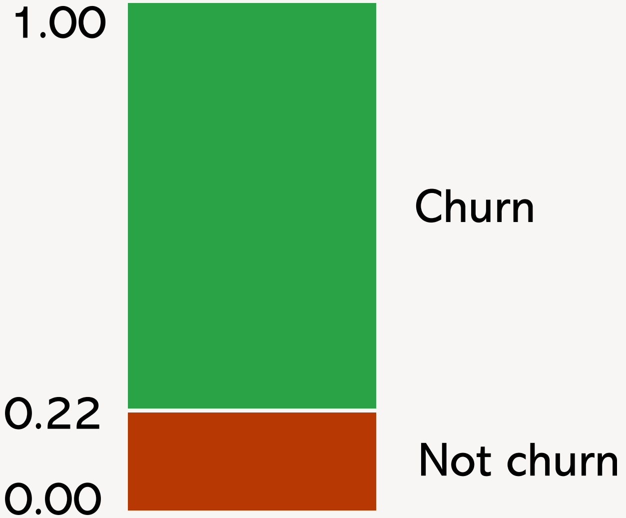
| | Model | Observation | TP | TN | FP | FN | Precision | Recall | AUCROC | Best Threshold | Validation Gain | Test Gain |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | Dummy Model (All Churn) | 1409 | 373 | 0 | 1036 | 0 | 0.264727 | 1.000000 | 0.500000 | 0.00 | 70300 | 82900 |
| 1 | Dummy Model (All Not Churn) | 1409 | 0 | 1036 | 0 | 373 | NaN | 0.000000 | 0.500000 | 0.00 | -152500 | -186500 |
| 2 | Random Forest (SMOTE) | 1409 | 212 | 901 | 135 | 161 | 0.610951 | 0.568365 | 0.839842 | 0.05 | 88000 | 109700 |
| 3 | LightGBM (SMOTE) | 1409 | 216 | 888 | 148 | 157 | 0.593407 | 0.579088 | 0.838736 | 0.01 | 92700 | 115700 |
| 4 | XGBoost (SMOTE) | 1409 | 253 | 885 | 151 | 120 | 0.626238 | 0.678284 | 0.857387 | 0.07 | 96300 | 116000 |
| 5 | Random Forest (ADASYN) | 1409 | 208 | 894 | 142 | 165 | 0.594286 | 0.557641 | 0.835892 | 0.08 | 91200 | 110400 |
| 6 | LightGBM (ADASYN) | 1409 | 217 | 895 | 141 | 156 | 0.606145 | 0.581769 | 0.840977 | 0.02 | 88100 | 111500 |
| 7 | XGBoost (ADASYN) | 1409 | 245 | 880 | 156 | 128 | 0.610973 | 0.656836 | 0.855058 | 0.08 | 97700 | 117700 |
| 8 | Random Forest (RUS) | 1409 | 302 | 783 | 253 | 71 | 0.544144 | 0.809651 | 0.861091 | 0.22 | 96400 | 119800 |
| 9 | LightGBM (RUS) | 1409 | 306 | 764 | 272 | 67 | 0.529412 | 0.820375 | 0.857970 | 0.16 | 95800 | 111700 |
| 10 | XGBoost (RUS) | 1409 | 313 | 749 | 287 | 60 | 0.521667 | 0.839142 | 0.862201 | 0.19 | 97300 | 114900 |

- Before performing threshold tuning, XGBoost (RUS) seems good because it has a highest AUCROC.
- However, prior to that, Random forest (RUS) is the better model as it achieves maximum gains.
- Ultimately, for the final decision, we choose Random forest (RUS) to deploy.

# 7. APPENDIX

## Threshold

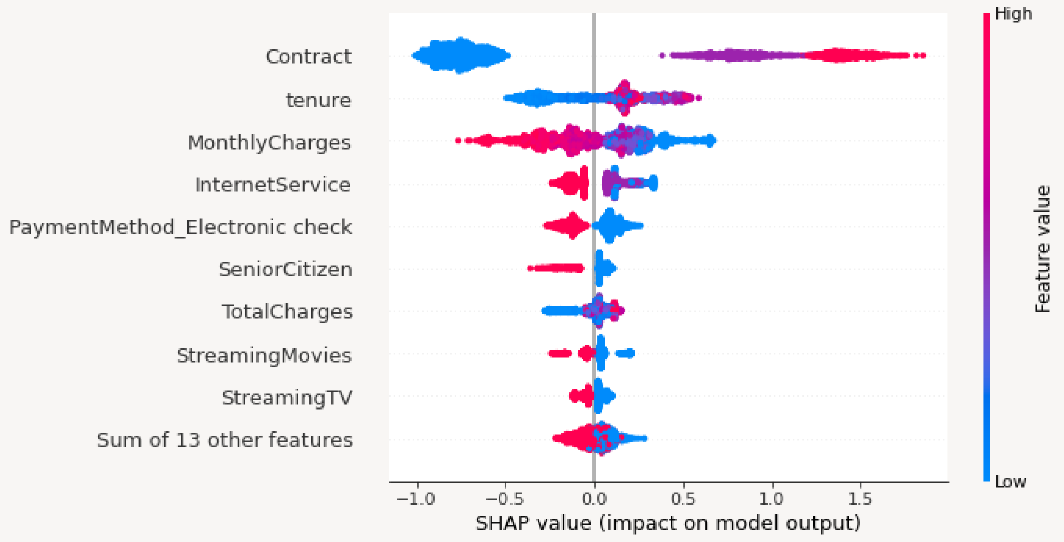
| | Model | Observation | TP | TN | FP | FN | Precision | Recall | AUCROC | Best Threshold | Validation Gain | Test Gain |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | Dummy Model (All Churn) | 1409 | 373 | 0 | 1036 | 0 | 0.264727 | 1.000000 | 0.500000 | 0.00 | 70300 | 82900 |
| 1 | Dummy Model (All Not Churn) | 1409 | 0 | 1036 | 0 | 373 | NaN | 0.000000 | 0.500000 | 0.00 | -152500 | -186500 |
| 2 | Random Forest (SMOTE) | 1409 | 212 | 901 | 135 | 161 | 0.610951 | 0.568365 | 0.839842 | 0.05 | 88000 | 109700 |
| 3 | LightGBM (SMOTE) | 1409 | 216 | 888 | 148 | 157 | 0.593407 | 0.579088 | 0.838736 | 0.01 | 92700 | 115700 |
| 4 | XGBoost (SMOTE) | 1409 | 253 | 885 | 151 | 120 | 0.626238 | 0.678284 | 0.857387 | 0.07 | 96300 | 116000 |
| 5 | Random Forest (ADASYN) | 1409 | 208 | 894 | 142 | 165 | 0.594286 | 0.557641 | 0.835892 | 0.08 | 91200 | 110400 |
| 6 | LightGBM (ADASYN) | 1409 | 217 | 895 | 141 | 156 | 0.606145 | 0.581769 | 0.840977 | 0.02 | 88100 | 111500 |
| 7 | XGBoost (ADASYN) | 1409 | 245 | 880 | 156 | 128 | 0.610973 | 0.656836 | 0.855058 | 0.08 | 97700 | 117700 |
| 8 | Random Forest (RUS) | 1409 | 302 | 783 | 253 | 71 | 0.544144 | 0.809651 | 0.861091 | 0.22 | 96400 | 119800 |
| 9 | LightGBM (RUS) | 1409 | 306 | 764 | 272 | 67 | 0.529412 | 0.820375 | 0.857970 | 0.16 | 95800 | 111700 |
| 10 | XGBoost (RUS) | 1409 | 313 | 749 | 287 | 60 | 0.521667 | 0.839142 | 0.862201 | 0.19 | 97300 | 114900 |



# 7. APPENDIX

## SHAP

| Features | Sign | Meaning |
|---|---|---|
| Contract | + | Customers on contract or top-up plans are more likely to churn, possibly due to the ease of changing numbers for top-up customers. |
| Tenure | + | Long-time customers have a higher churn probability, potentially due to elderly individuals reducing phone usage to cut expenses. |
| Monthly charges | - | Customers paying lower charges are more likely to churn, influenced by industry trends towards appealing, low-priced packages from other operators. |
| Internet service | - | Customers without internet in their package are at a higher churn risk, as competitors offer inclusive internet services, reflecting its growing importance. |
| Payment method (electronic check) | - | Non-users of electronic payment for bills are more likely to churn, possibly due to the inconvenience of payment, prompting a switch to other providers. |



# 7. APPENDIX

**Reference**


- https://medium.com/@stephen.blount99/putting-a-price-on-customer-churn-38a184e530b8

- https://www.kaggle.com/datasets/blastchar/telco-customer-churn/data

- https://www.kaggle.com/code/bandiatindra/telecom-churn-prediction

- https://machinelearningmastery.com/smote-oversampling-for-imbalanced-classification/

# 8. ABOUT ME

- Peerapat Tancharoen holds a Bachelor's degree in Economics from Srinakharinwirot University, graduating with first honors and a GPA of 3.67. He also earned a Master's degree in Economics from Thammasat University, achieving a GPA of 3.98.


**Contact me on**

- linkedin.com/in/peerapat-tancharoen-664759220/
- peerapat.tcr@gmail.com