SECURITY AS A SERVICE

Katherine Salazar
June 1, 2021

PIPELINE

1
Plan

2
Acquire

3
Prepare

4
Explore

5
Model

6
Deliver

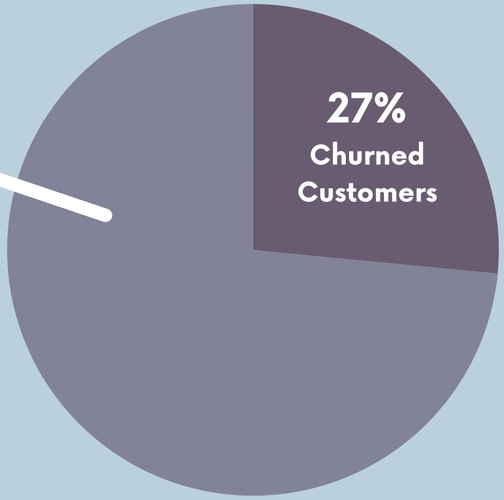
EXECUTIVE SUMMARY

KT&T

73%

Retained Customers

Internet Service
Security Focused



Goals

- Understand customer churn cause
- Explain Security Features to churn

Model 2

DecisionTreeClassifier() has the best performance of all 4 models

Baseline `0.7423`

Recommend

Security as a Service customer focus

UNIVARIATE STATS TAKEAWAYS

is_churn

Only 27% of security & risk focused customers churn

Features

Security & risk focused customers are niche, but less likely to churn

Fiber Optic

35% high rate of internet service churn

Tenure & Monthly Charges

Running more test for visible trends

KT&T

BIVARIATE STATS TAKEAWAYS

Features

Security & risk focused customers are niche, but less likely to churn

Tenure

Most customers churn at 1 year

Fiber Optic

High rate of internet service 725 out of 1008 customers churn

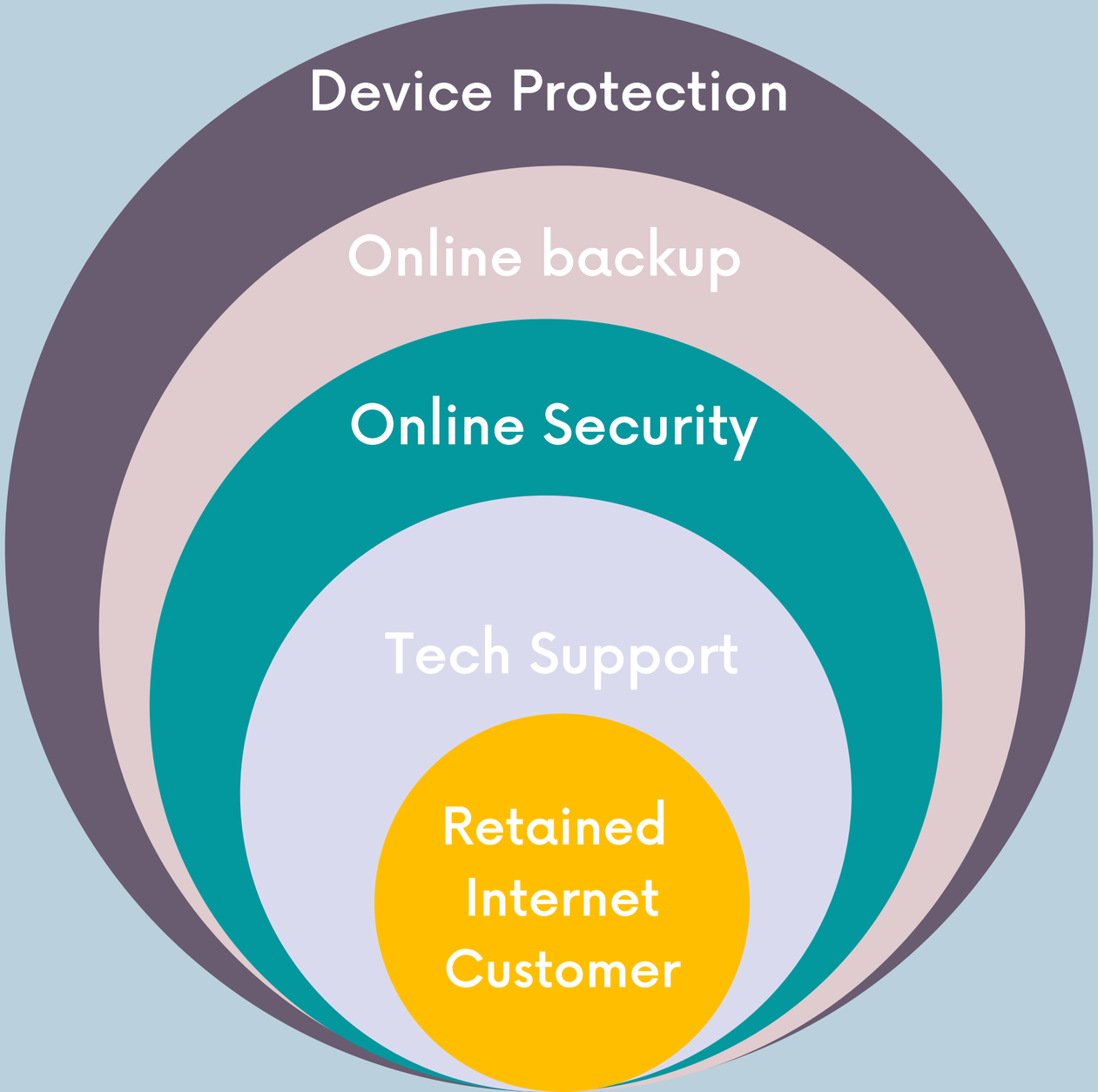
Monthly Charges

Customers who churn are likely to pay $74 a month

KT&T

MULTIVARIATE STATS & ANALYSIS TAKEAWAYS

When a customer chooses security features they are less likely to churn



MODEL COMPARISON

Baseline 0.7423

```
The accuracy for our model is 0.8694
The True Positive Rate is 0.644, The False Positive Rate is 0.0523,
The True Negative Rate is 0.948, and the False Negative Rate is 0.356
```

	precision	recall	f1-score	support
0	0.884566	0.947728	0.915059	2927.000000
1	0.810409	0.643701	0.717499	1016.000000
accuracy	0.869389	0.869389	0.869389	0.869389
macro avg	0.847488	0.795714	0.816279	3943.000000
weighted avg	0.865458	0.869389	0.864153	3943.000000

Random Forest

Model 1

```
clf2 = DecisionTreeClassifier()
```

```
The accuracy for our model is 0.9906
The True Positive Rate is 0.971, The False Positive Rate is 0.00273,
The True Negative Rate is 0.997, and the False Negative Rate is 0.0285
```

	precision	recall	f1-score	support
0	0.990163	0.997267	0.993702	2927.000000
1	0.991960	0.971457	0.981601	1016.000000
accuracy	0.990616	0.990616	0.990616	0.990616
macro avg	0.991061	0.984362	0.987652	3943.000000
weighted avg	0.990626	0.990616	0.990584	3943.000000

Decision Tree

Model 2

```
clf3 = DecisionTreeClassifier(max_depth=3)
```

```
The accuracy for our model is 0.7918
The True Positive Rate is 0.344, The False Positive Rate is 0.0526,
The True Negative Rate is 0.947, and the False Negative Rate is 0.656
```

	precision	recall	f1-score	support
0	0.806105	0.947386	0.871054	2927.000000
1	0.693837	0.343504	0.459513	1016.000000
accuracy	0.791783	0.791783	0.791783	0.791783
macro avg	0.749971	0.645445	0.665283	3943.000000
weighted avg	0.777176	0.791783	0.765011	3943.000000

Decision Tree

Model 3

```
The accuracy for our model is 0.828
The True Positive Rate is 0.429, The False Positive Rate is 0.0335,
The True Negative Rate is 0.967, and the False Negative Rate is 0.571
```

	precision	recall	f1-score	support
0	0.829862	0.966519	0.892992	2927.00000
1	0.816479	0.429134	0.562581	1016.00000
accuracy	0.828050	0.828050	0.828050	0.82805
macro avg	0.823171	0.697826	0.727787	3943.00000
weighted avg	0.826414	0.828050	0.807855	3943.00000

KNN

Model 4

FOCUSED MODEL COMPARISON

KT&T

Random Forest

Model 1

	predict_churn	predict_no_churn
actual_churn	true negative: 2774	false positive: 153
actual_no_churn	false negative: 362	true positive: 654

Decision Tree

Model 2

```
clf2 = DecisionTreeClassifier()
```

	predict_churn	predict_no_churn
actual_churn	true negative: 2919	false positive: 8
actual_no_churn	false negative: 29	true positive: 987

Model 2
DecisionTreeClassifier()
has the best performance of all 4 models

Baseline `0.7423`

Decision Tree

Model 3

```
clf3 = DecisionTreeClassifier(max_depth=3)
```

	predict_churn	predict_no_churn
actual_churn	true negative: 2773	false positive: 154
actual_no_churn	false negative: 667	true positive: 349

CONCLUSIONS & NEXT STEPS

Security as a Service

Rest assure,you are secure!

Explore HIPAA, GDPR and CPRA certifications

Aditional Data Security

Explore corporate security focused customers

APPENDIX

PROJECT DESCRIPTION AND GOALS

KT&T

Description

Introducing Security as a Service for telco_churn data based on data modeling

Project Planning

- trello kanban board
- agile program management

Goals

Security focused customers are less likely to churn, customers who don't have security features are more likely to churn based on model 2

Hypothesis

- H0: There is no relationship between internet customer churn and Security Features
- Ha: There is a relationship between internet customer churn and Security Features

DATA DEFINITIONS

Feature	Definition	Data Type
internet_service_type_id	DSL, fiber optic, or no service	int(0-3)
customer_id	ID of Customer	object
monthly_charges	in USD $	float
total_charges	in USD $	float
tenure_years	Customer tenure in years	int
is_churn	Customer has left the company or stayed	int - boolean
security_features	online security or not	int - boolean
backup_features	backup or not	int - boolean
device_protection_features	device protection or not	int - boolean