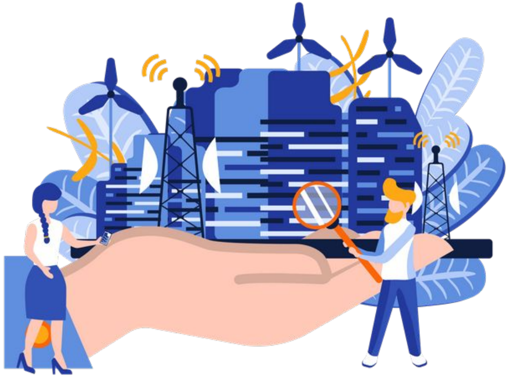
Telco Customer Churn Analysis, Mitigation, and Strategic Action

An in-depth data-driven analysis to ***improve business performance*** *and achieve* ***long-term strategic alignment***



Meet our research analyst, GroundZero team member

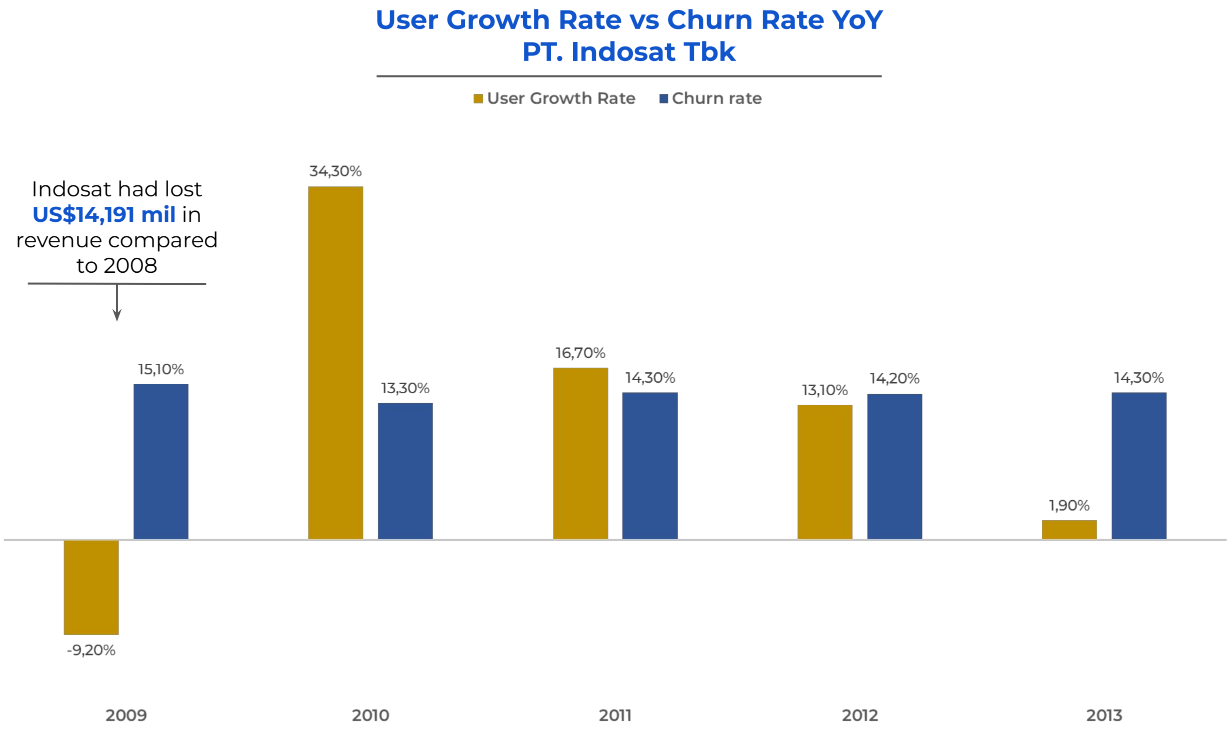
Ruben Daniel Hosea

Andreas Lukita

Ronaldo WIjaya

Phase 1: Introduction

Telecommunication companies in Indonesia are experiencing challenges in maintaining customer loyalty to operator services



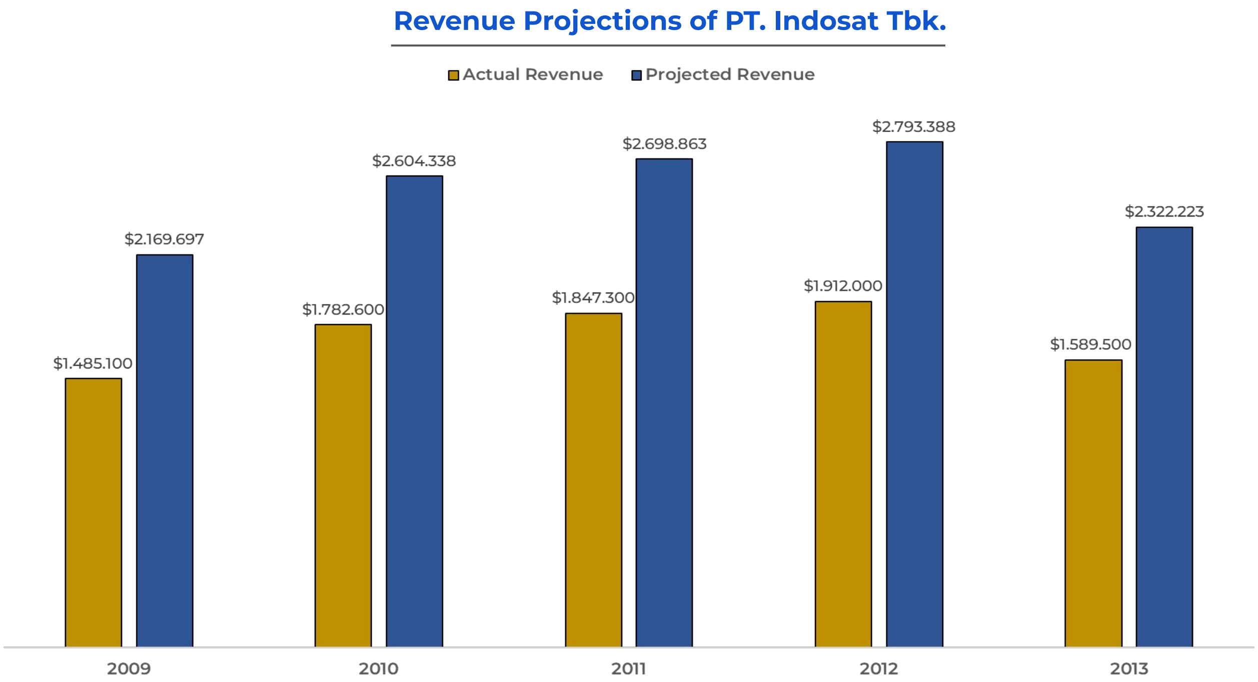
Sources: (Annual Report of PT Indosat Tbk)

As shown from the graph, the company's case is having **challenges maintaining their total subscribers** in the telecommunications industry competition.

Based on research from (Abdulsalam et al., 2022), **customer churn results in significant profit loss** if it exceeds the threshold set by the company.

The research was proven on the case of PT Indosat Tbk from 2009 - 2013.

Losing customers makes a tangible impact on the company, especially on potential revenue in annualized terms



According to a study by (Zeng et al., 2023), when the **customer churn rate drops by 5%**, then **companies can make a profit of 25% - 85%**.

We projected using the **geometric mean** (square root of multiplication from 25% and 85%) **of research above at 46%** of PT. Indosat Tbk. revenue.

As a result, it turns out that the company is **losing significantly higher potential revenue from year to year** as a result of the churn rate.

Sources: (Annual Report of PT Indosat Tbk & Study Research)

The issue of customer loyalty and churn has been researched in detail, companies need to increase awareness based on this research to scale up the growth in the market

Popular Studies on Telecommunication Customer Churn

1. Research from (Somosi et al., 2021) reveals that factors such as **tenure, usage intensity, and mature age are indicators of loyal customers**.

2. The study from (Jain et al., 2020) reinforces **customer churn reasons** such as **lack of engagement, high call rates or SMS charges, and lack of promotions** using a machine learning algorithm approach.

3. From the study of (Ã et al., 2008), customer churn is caused by factors such as **low customer satisfaction, high percentage of customer switching costs and the influence of customer demographics**.

Recommended Action From the Studies

1. Focusing on several variables in terms of **length of use, service, price, usage, and demographics of customers**.

2. **Make special offers with plans that match market** characteristics to retain customers effectively.

3. Applying appropriate pricing to offer programs to encourage customers to accept better alternatives.

4. Companies need to **involve AI technology** approaches to **accelerate effectiveness, reduce unnecessary costs, and know the characteristics of their respective customers**.

Based on these issues, we found several key issues that need to be resolved by the company's management

1. Which customer segmentation has the highest potential revenue loss due to churn?

2. What segmentation should companies focus on to generate optimal results?

3. What are the factor affecting customer decision to churn?

Phase 2: Data and Methodology

We get the exact same issue in the data we receive, here are the steps we take to analyze and resolve the issue

01
Dataset given by committee

Read and understand the available data structures and perform data transformation as necessary

02
Identify the issue and problem

Exploring and identifying data to find business problems that occur, which is customer churn.

03
EDA using Excel and PowerBI

Perform visualization by doing some quantitative approaches to get in-depth customer analysis.

04
Building ML model with Python

Combining machine learning models to create a ‘score’ to predict the likelihood of customer to churn

05
Result posted on Powerpoint

Present the customer segmentation and possible mitigation on customer with high risk of churn

We grouped tenure to enhance the comprehensiveness of the analysis by referring to scientific research on telecommunication customer characteristics.

Tenure Month Grouping

1. **Low Tenure Group :**
   Tenure Month Range from 0 - 3 Months

2. **Mid Tenure Group :**
   Tenure Month Range from 4 - 11 Months

3. **High Tenure Group :**
   Tenure Month above 11 Months

Description

Customer of Low tenure has several key characteristic which require them to be grouped:

a. Little information about customer Needs
b. No History of Customer Buying capability

Customer In Mid tenure are more mature in the sense that company has more information about customer needs based on purchase history as well as customer buying power.

Customer in high tenure are now mature and less likely to churn.

Source:
Customer Segmentation Model Research Based on Organizational Customer Life Cycle in Telecom Operator
Author: Chengrong Hu1,a, Huaying Shu2,b , Xinchun Qiao3,c

Also, we do calculations to create a projection of the company's revenue and the projected user expenses that occurs in the company.

Potential Revenue

Potential Revenue : CLTV - Monthly Purchase * Tenure Months

Potential revenue is used instead of CLTV to take into account the financial impact each customer had

This metric will be used often to evaluate customer segmentation that is worth the effort and investment

Total Spending

Total Spending : Monthly Purchase * Tenure Months

Total Spending shows the amount of money that has been spent on the company.

Then, we do product segmentation to make it easier to categorize the characteristics of product usage by customers over several periods.

Product Segmentation

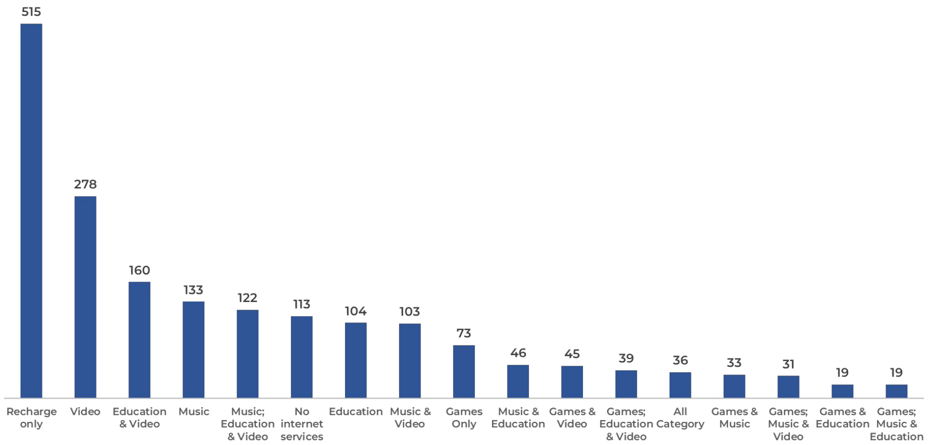
Column 'Games Product', 'Music Product',Education Product, and 'video product' is concatenated when the value is 'YES'

This will result in Several product Combination, The following are special mentioned category :

1. **Recharge only**: Customer did not use any of the product
2. **All Category**: use All product
3. **No Internet Service**: customer phone did not have internet service

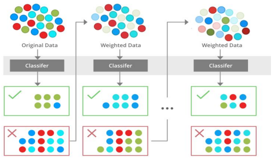
This Product segmentation is made to understand the demand for certain product combination within the customer.

Product Distribution

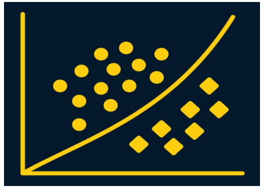


Finally, we apply the selected algorithm by performing the necessary preprocessing steps according to the characteristics of the dataset.

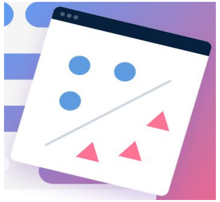
Algorithm Choices



XGBoost



Logistic Regression

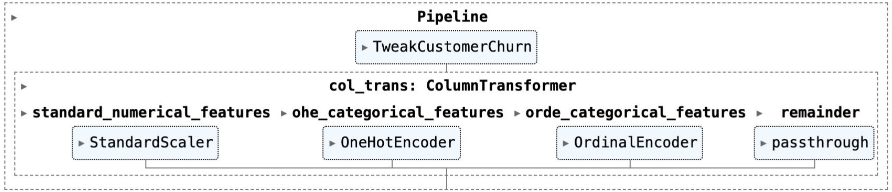


Support Vector Machines

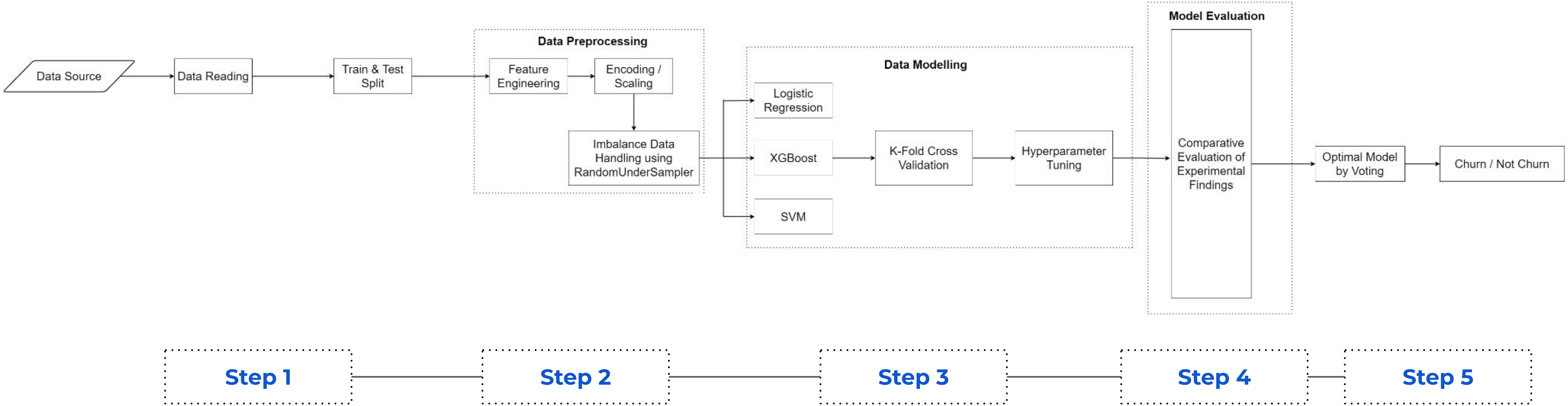
Preprocessing steps

Here are the idea behind our preprocessing steps:

1. Avoid data leakage by leveraging **Pipeline**.
2. Write custom transformers to **introduce new features** such as net_cltv_total_spending, tenure_segmentation, etc.
3. **Encode** categorical features accordingly using OneHotEncoder, OrdinalEncoder. **Scale** numerical features for Logistic Regression and SVM using Standard Scaler.
4. Address **data imbalance** (No Churn:Churn 73%:27%) using RandomUnderSampler.



Here is the entire process of training our ML model from scratch to deployment

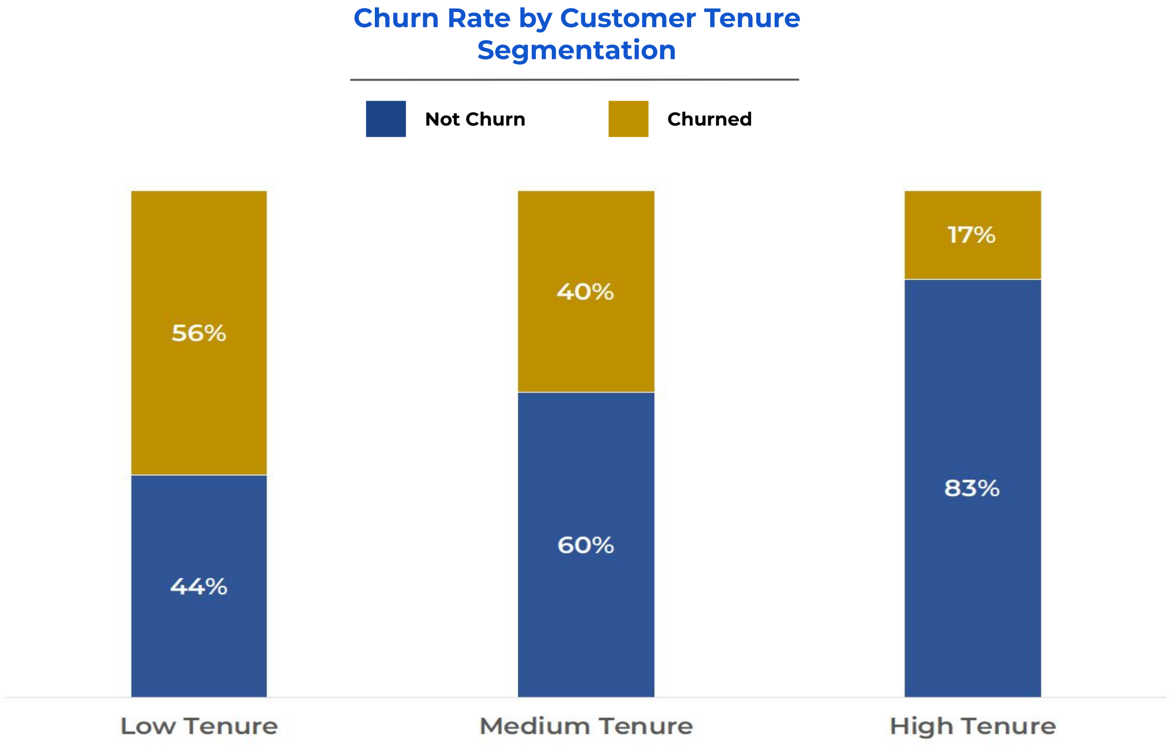


Phase 3:
Results

Part 1:

"What customer segmentation has the highest customer churned rate and potential revenue loss?"

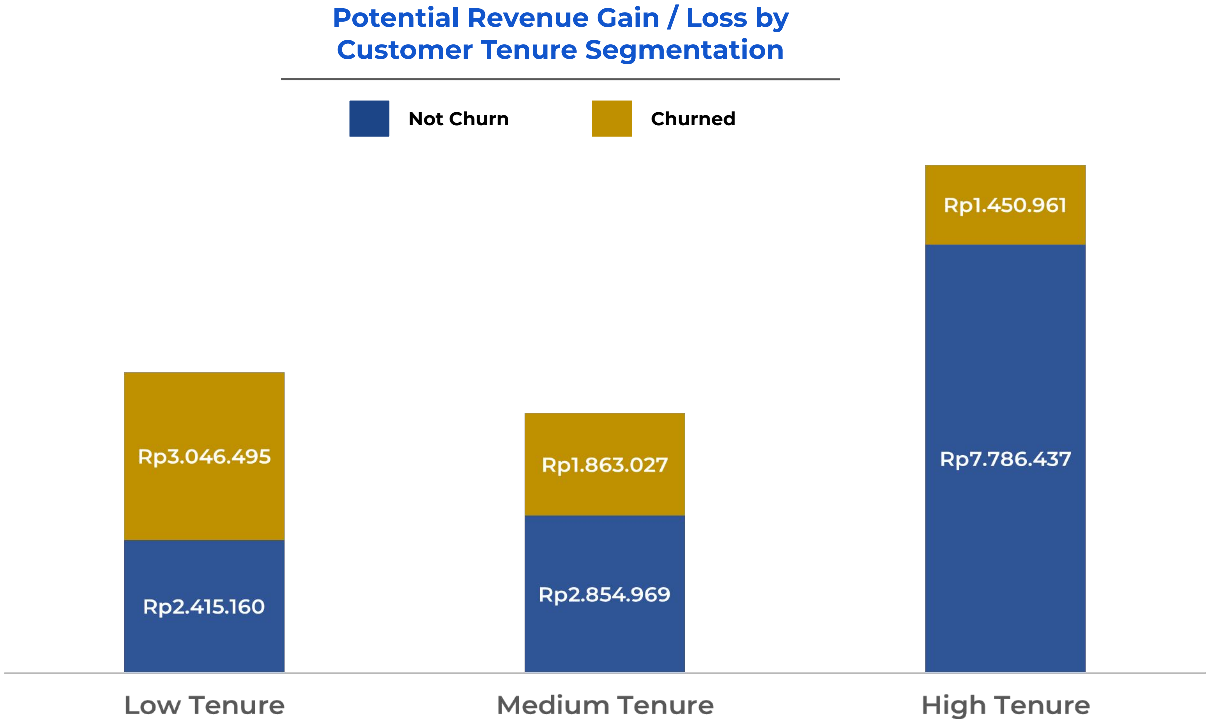
Based on customer tenure distribution, customer churn rate decreases from 56% to 17% when their tenure segmentation increases



The decrease in churn rate is driven by an increase in the tenure for each customer. **The higher the tenure value, the lower the possibility of the customer being churned.**

From this analysis, a point can be made that **tenure is one of the influential variables** on customer churn prediction.

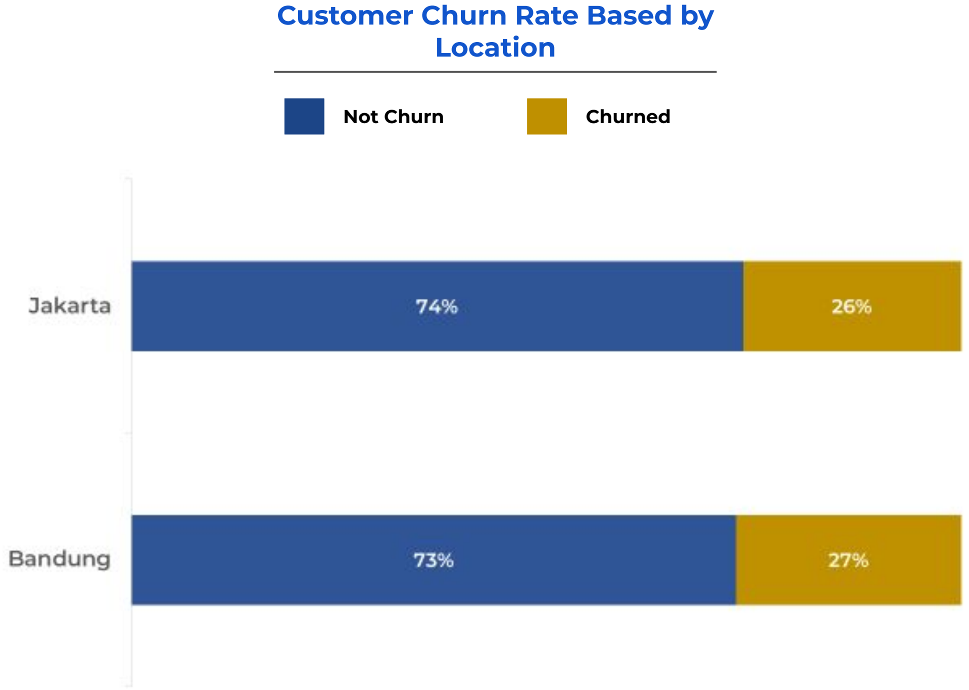
Furthermore, tenure segmentation also reflects how much the company loses potential revenue from customer churn, which is 6 million IDR in the current period.



From the analysis on the left, tenure segmentation also reflects the extent of potential revenue loss. **The higher the tenure value, the lower the company loses its potential revenue**.

At this point it can be concluded that tenure segmentation provides an **overview of the potential revenue loss of customers**.

Based on customer location distribution, Jakarta and Bandung have similar proportion of customer churn frequency.



We can see that the ratio of churned and not churned in both locations have very similar value.

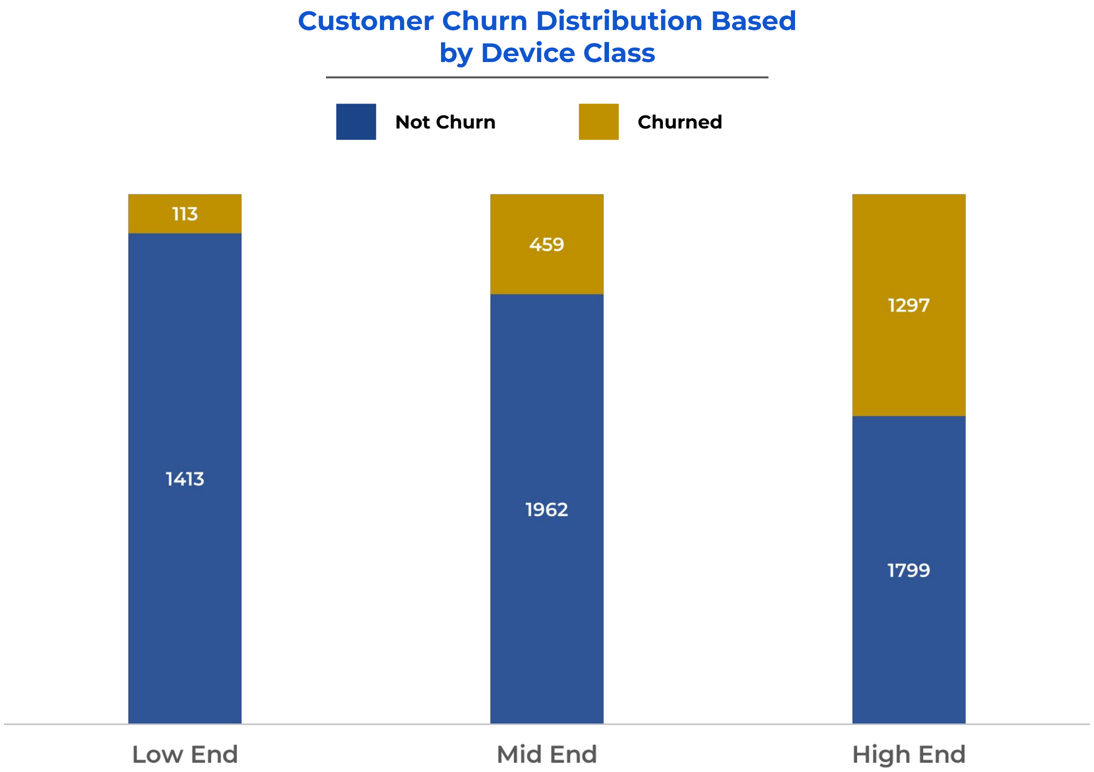
Thus, we can derive a point that **location has minimal influence** on customer decision to churn.

Upon looking at the potential revenue loss, the company experienced a bigger financial loss in Jakarta compared to Bandung.

Based on its location, it turns out that **Jakarta experienced the highest loss of potential revenue** compared to Bandung.

At this point, it can be concluded that **location affects the potential revenue loss** experienced by the company.

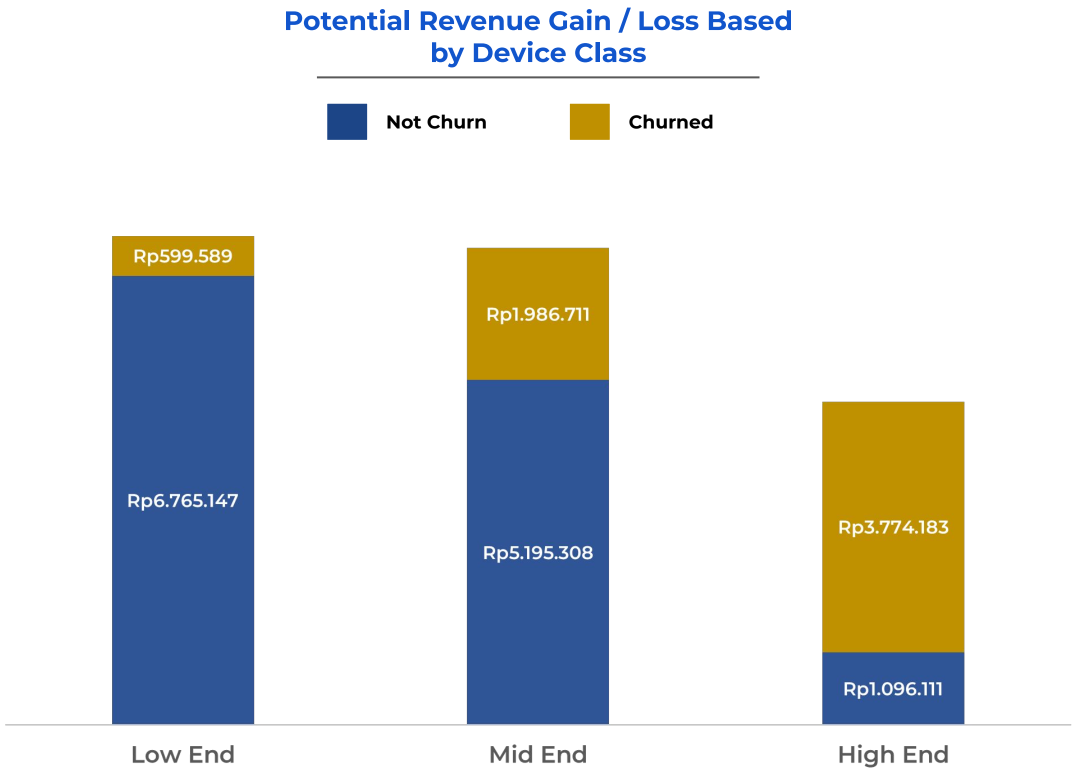
Based on the device class distribution, the highest churned customers are in the high end category, with a churn rate percentage reaching 42% in its sample class.



We can see that there is a significant increase in the churned population of each device class.

While analyzing, we can reach a point that **device class has a contribution to customer decisions to churn.**

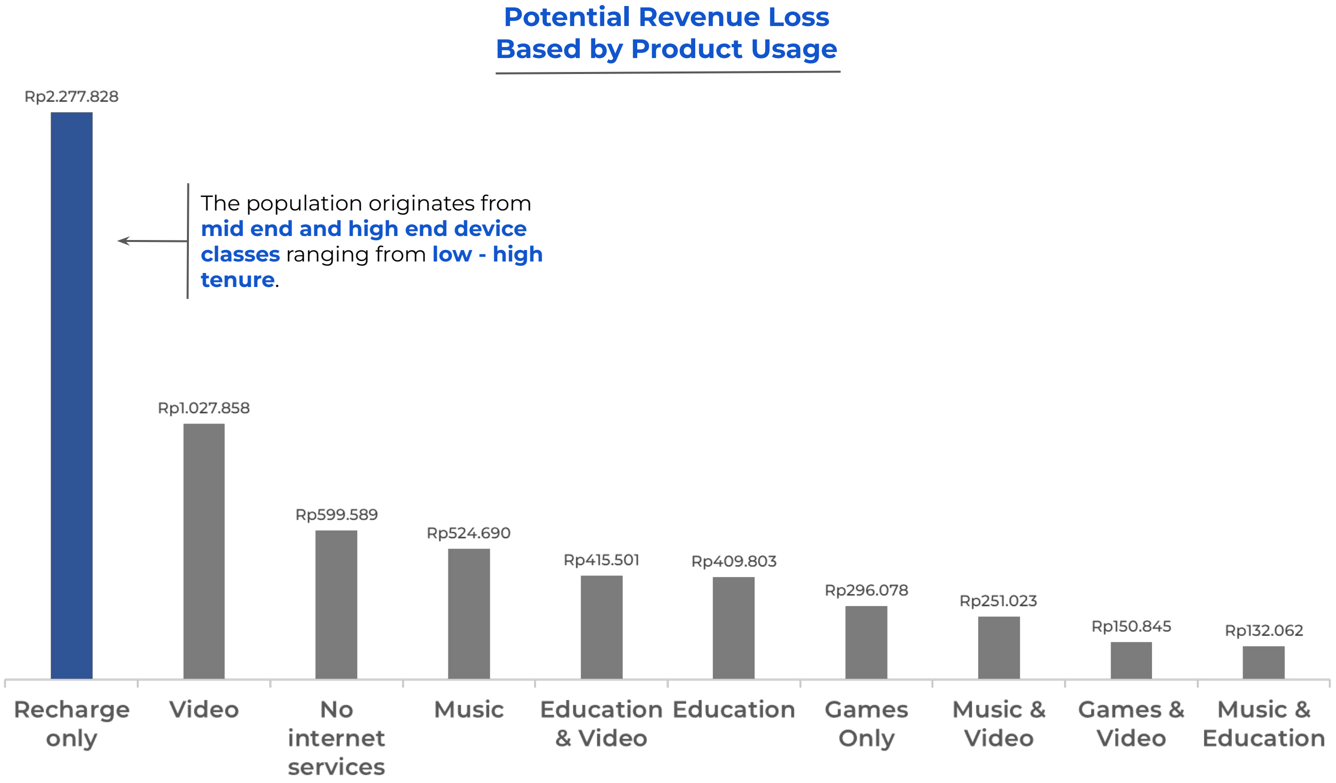
The impact of customer churn on the high end device class is illustrated through the potential revenue loss experienced by the company of 3.7 million IDR



It is projected through potential revenue that the **high end device class experiences the highest loss** compared to other categories.

Provisional analysis, device class **has a contribution to potential revenue loss** to the company

Related to the use of their products, here are the top 10 based on the potential revenue loss amongst the churned customers. The company has lost around 6 million IDR due to this problem.



From the graph beside, the biggest loss occurred from the **Recharge only category around 2M IDR**, followed by the **Video category about 1M IDR**, and **No internet service around 500K IDR**.

The following observations show a **relationship between product usage to both potential revenue and churned customers**.

Part 2:

“What segmentation should companies focus on to generate optimal results?”

Based on the analysis above, we focus on customer segmentation that has a direct impact on the company, in terms of potential revenue loss and high churn rate.

Customer Characteristic

Low Tenure | Jakarta | Mid - High End Device Class

1. Based on Slide 17, We saw that **low tenure customer has the highest potential revenue loss (3 Million IDR)** compared to other tenure grouping

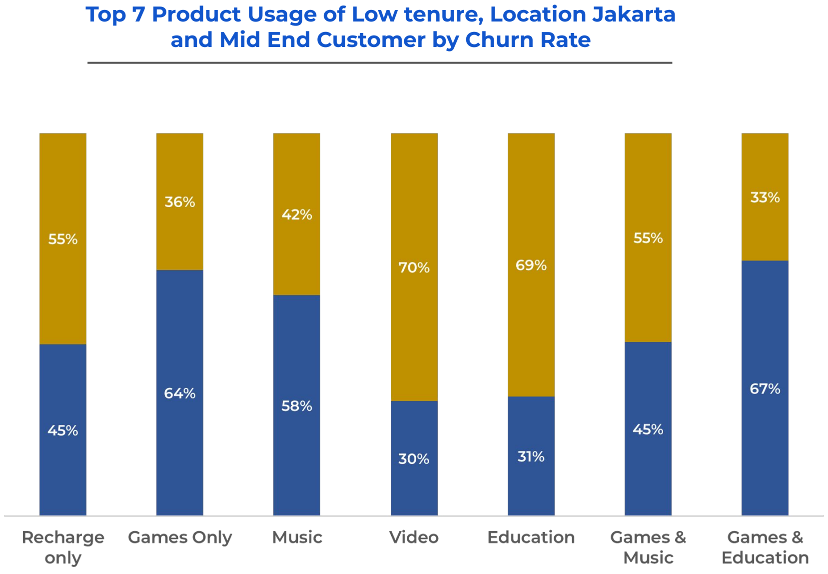
2. We also saw in slide 19 that **Jakarta has the highest potential revenue loss compared to Bandung**. However we also see in slide 16 that location did not affect decision to churn.

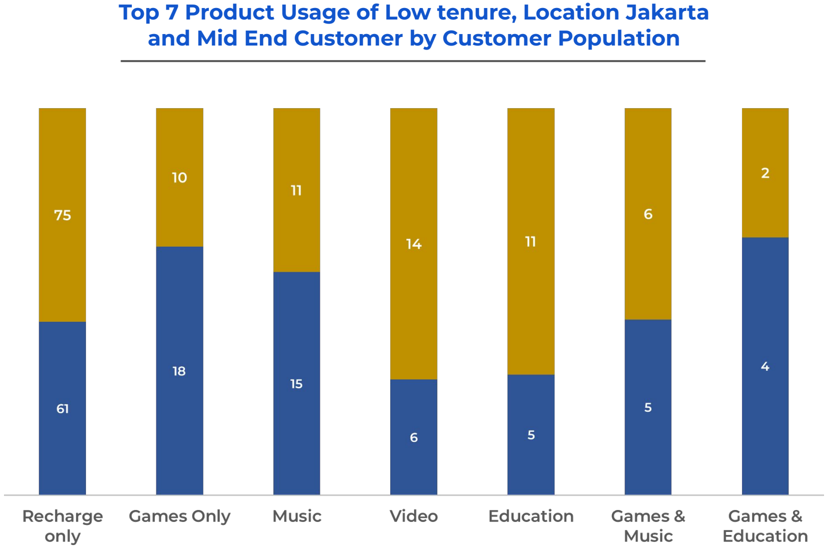
3. Lastly, Based on slide 20 & 21, we can see that churn mostly happen in **mid end to high end device class**.

Therefore, we focus on exploring usage products based on customer segmentation of **low tenure**, **located in Jakarta**, and having **mid end - high end device class**.

For Mid End low tenure customer, Recharge only has the highest demand, accounting about 52% of total demand in every product segment.
In addition, We also see that product segment Games Only and Music only and Video Only are among top 4 high demand product with churn rate lower than 45%

Not Churn Churned

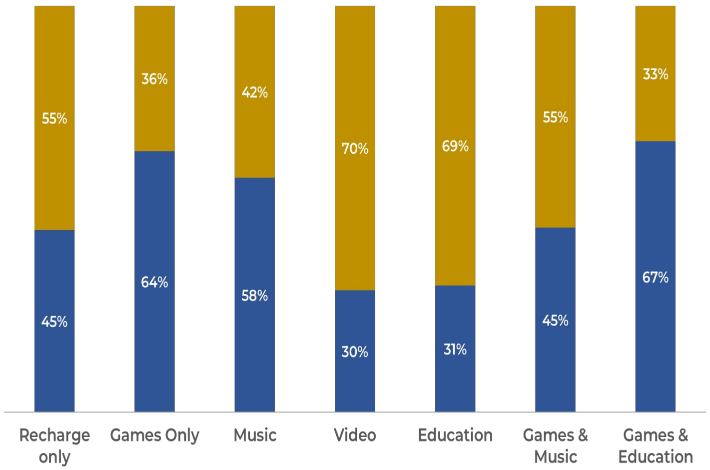




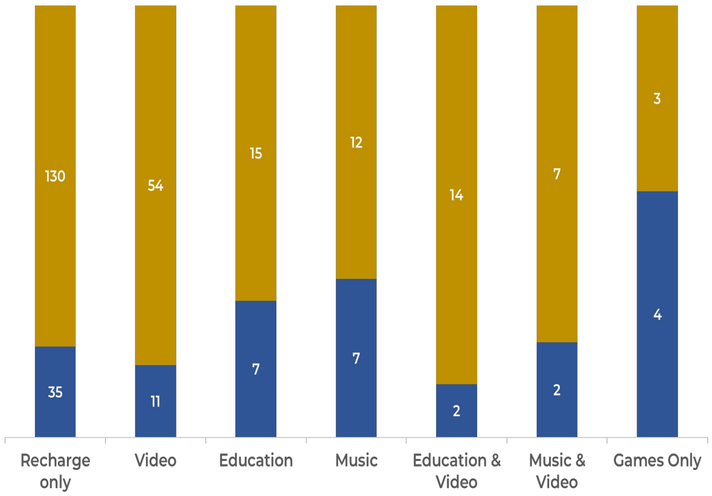
For High End low tenure customer, Recharge only is the product with the highest demand accounting for 51% of total demand followed by video only product accounting for 20% of total product demand.
In addition, We also see that product segment Education only and Music only are among top 4 high demand product with churn rate lower than 70%. While Games only has the lowest churn rate but with the lowest product demand

Not Churn Churned

Top 7 Product Usage of Low tenure, Location Jakarta and High End Customer by Churn Rate



Top 7 Product Usage of Low tenure, Location Jakarta and High End Customer by Customer Population

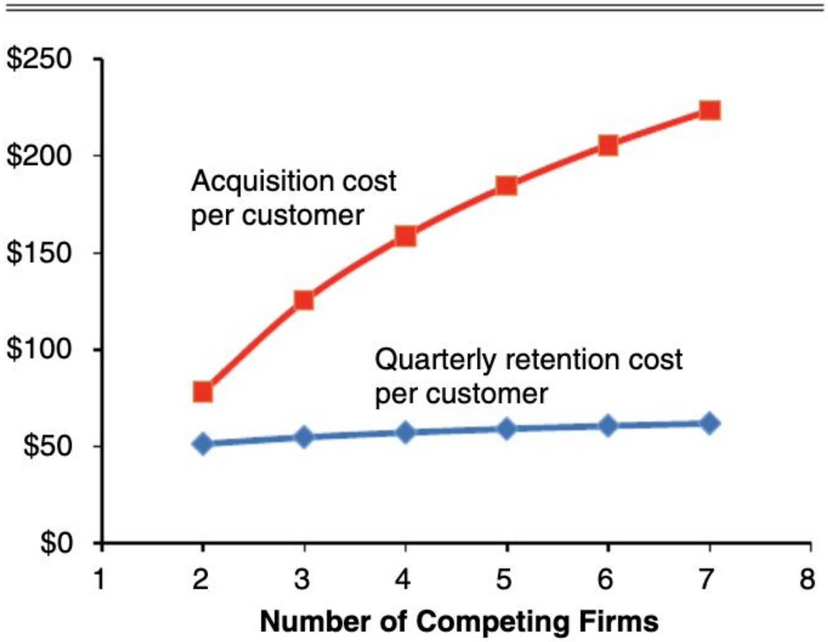


Customer retention is preferable to customer acquisition based on the research we have explored, saving more resources to be allocated to other marketing programs.

Secondary Research

In the paper ***Customer Acquisition and Retention Spending: An Analytical Model and Empirical Investigation in Wireless Telecommunications Markets***, Min, Zhang, Srivastava, and Kim posit that the stability of customer retention expenses remains consistent regardless of the degree of market competition.

However, as the competitive landscape expands with more contenders, the **costs associated with acquiring new customers rise sharply, at times reaching multiple fold the costs of retaining customers**.



Sources: (S. Zhang et al., 2016)

The modeling of our customer churn prediction refers to the above research ideas. Focusing on retaining customers rather than acquiring new customers.

Impact on Model Building

We want our model to be able to **capture more False Negative (FN) cases relative to False Positive (FP) cases**.

Why?

Missing out on a customer who churns **hurts the bottom line more** than spending resources attending to the wrong customer who is unlikely to churn. In essence,

Marginal cost of failing to catch incident of churning

>

Marginal cost of spending resources to prevent churn

AI Model Performance

Based on the research above, our AI Model is train to optimise Recall over Precision

Success Rate

(Soft Voting Classifier Model)

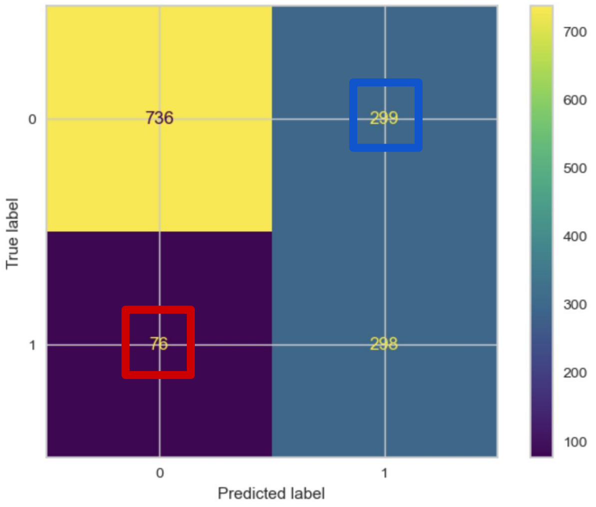
Out of 10 customers who churned, we manage to identify

8/10 *customer*

Please refer to the Appendix for details.

Understanding False Positive and False Negative

Confusion Matrix Comparison



Soft Voting Classifier Model

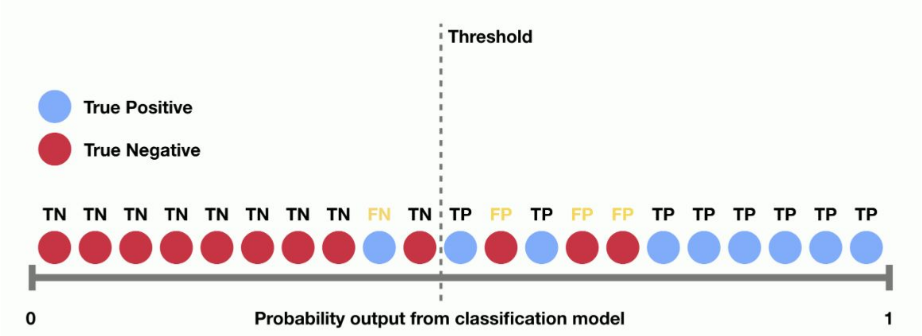
FP 299

FN 76 *(to minimize)*

***FP**: Our model predicts churn, but otherwise*
***FN:** Our model predicts not churn, but otherwise*

Adjustment of Probability Threshold

Default probability threshold = 0.50



For illustration purpose

Please refer to the Appendix for details.

So what?

1. Due to our **business needs**, we prefer to optimise recall over precision as the cost of customer acquisition is higher than that of customer retention.
2. Different costs associated with false positives and false negatives. In essence, false negative is **costlier** than false positive.
3. Adjusting the threshold allows us to **minimize the cost or risk** associated with wrong predictions.
4. Granted, **precision-recall tradeoff** still applies in that lowering our threshold to catch more false negative results in increase recall but decrease precision.
5. As such, we adjust our probability threshold to **0.404**

AI Model Value Proposition

Cost Saving

From the dataset, potential revenue loss (CLTV - tenure months x monthly purchase) per customer churned is

IDR 3,400,000

*With 80% recall, **we can save you***

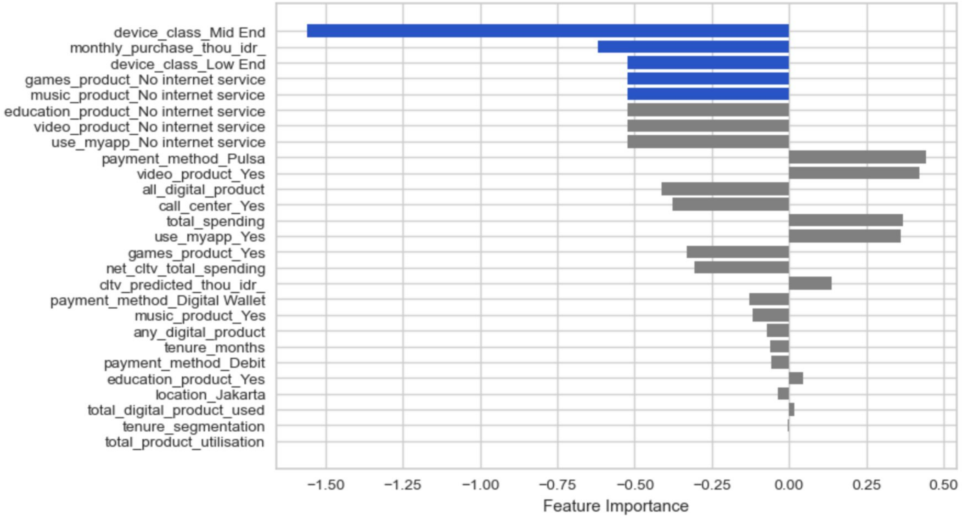
IDR 5,083,000,000

Part 3:

“What are the factor affecting customer decision to churn?”

Grokking Logistic Regression Result

Feature Importance

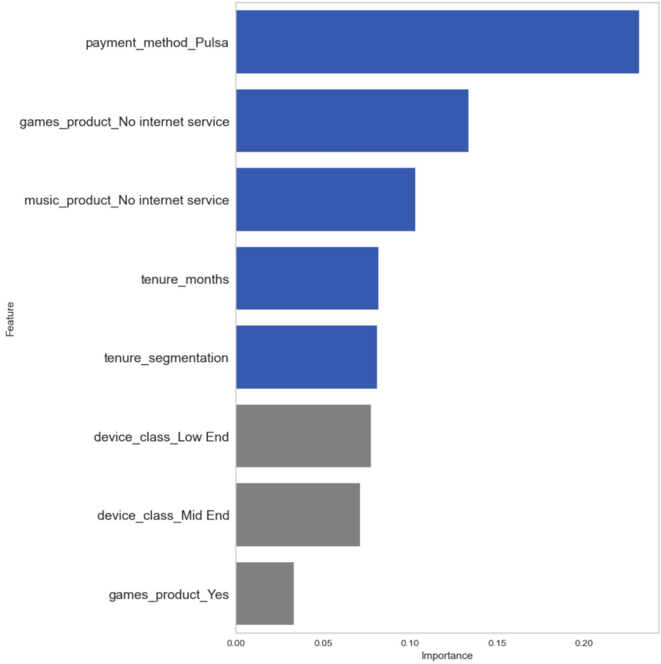


Key Insights

1. Relative to the High End device users, users with **Mid End** and **Low End** devices are less likely to churn.
2. As the value of **monthly_purchase_thou_idr_** increases, this decreases the log odds of churning suggesting probability of churning decreases.
3. The **top 5** most influential features towards churning prediction are highlighted in blue.

Grokking XGBoost Result

Feature Importance

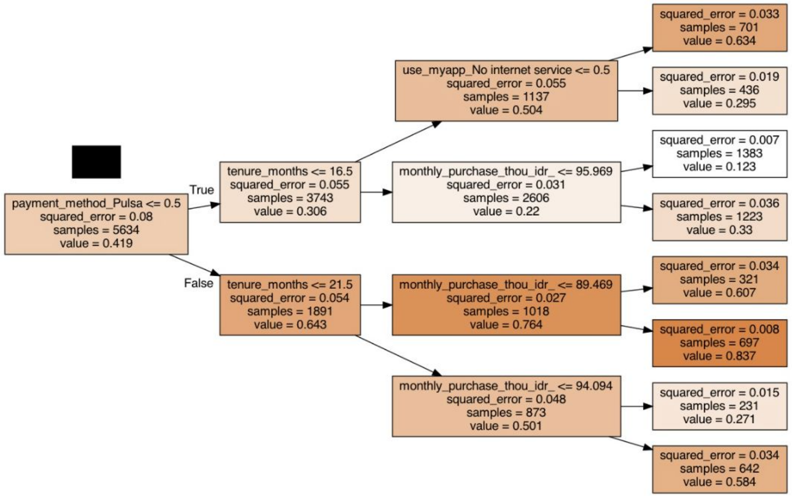


Key Insights

1. XGBoost grants several **advantages** compared to a simple Logistic Regression. These include handling of non-linear relationships, automatically learn complex feature interactions, and performance over imbalance dataset.
2. From the feature importance plot, we observe the **top 5** features contributing most to customer decisions to churn include payment_method_Pulsa, games_product_No internet service, tenure_months, etc.

Let's be honest, XGBoost is a complex model! We want a model easy to understand, but is able to approximate XGBoost prediction power. There is beauty in simplicity.

Surrogate Model



(Decision Tree depth = 3)

Please refer to the Appendix for details.

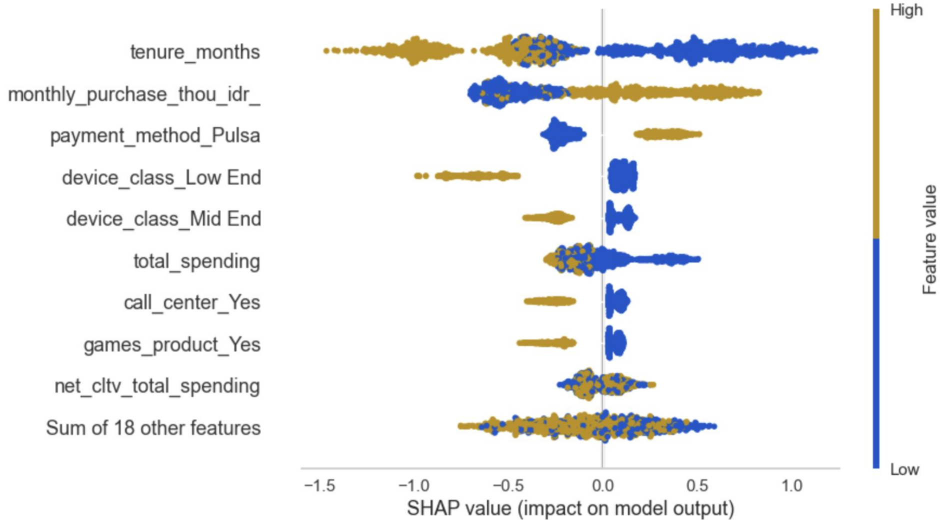
*With XGBoost, accuracy score is **75%** while recall score is **78.1%**, but with a simpler surrogate model, we manage*

71.5% *accuracy*

77.3% *recall*

To understand the XGBoost Results, we use Beeswarm plots to visualize the contribution of each feature to the model prediction.

Beeswarm Plot using SHAP Framework



Key Insights

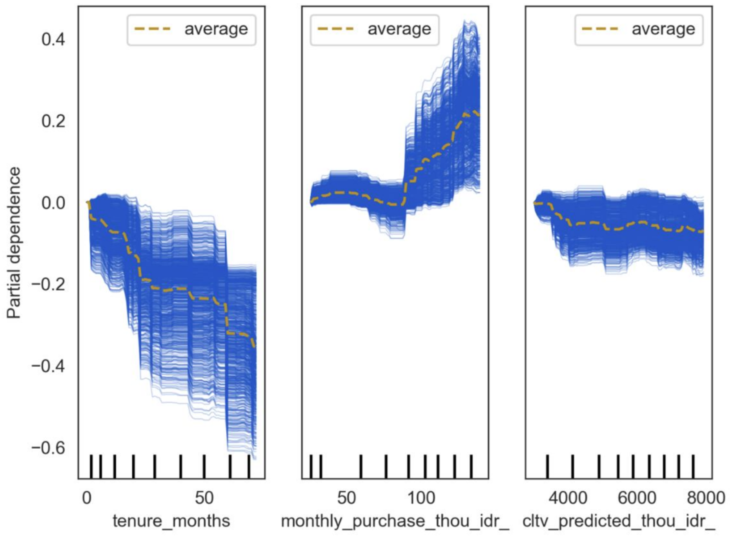
1. The y-axis lists features with decreasing order of impact on the model's output—**tenure_months** might be the **most influential feature**, followed by **monthly_purchase_thou_idr**. This is in line with finding in slide 17

2. **monthly_purchase_thou_idr_** has points predominantly to the right, suggesting that higher monthly purchases in thousands of IDR are associated with a higher model output (**more likely to churn**).

3. For **total_spending**, higher spending are associated with a lower model output (**less likely to churn**).

In addition, we also use PDP and ICE visualization to cross check the findings from Beeswarm Plot In slide 41

Partial Dependence Plot (PDP) and Individual Conditional Expectation (ICE)

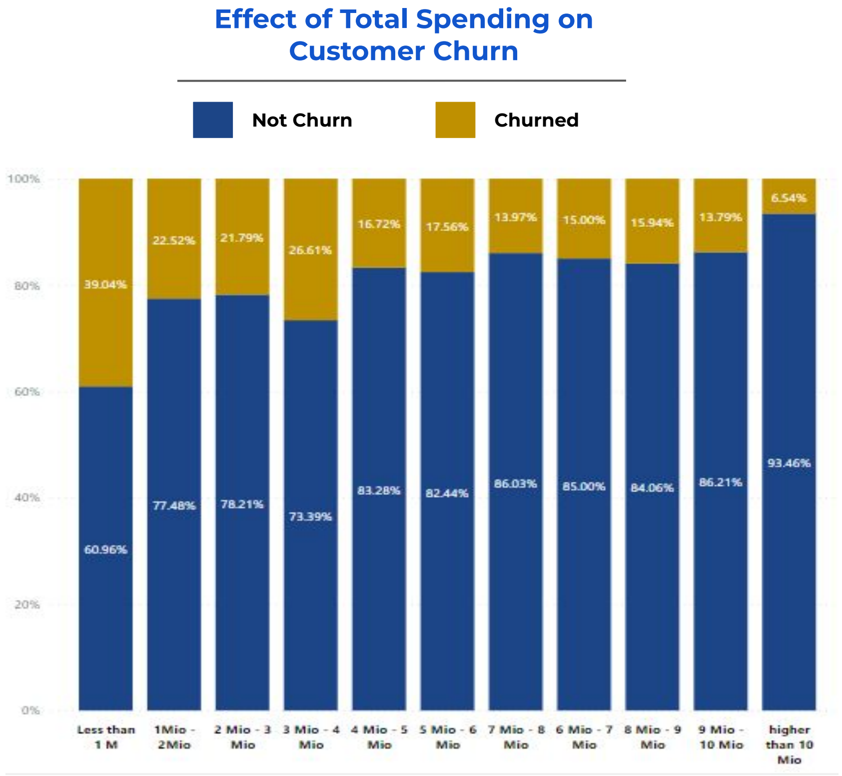


Key Insights

1. **tenure_months**: generally slope downwards as tenure increases. This indicates that, for individual customers, **as the tenure increases, the predicted outcome decreases (less likely to churn)**.

2. **monthly_purchase_thou_idr_**: generally shows an increasing trend, dips initially below monthly purchase 100, then starts to increase. This indicates that, for individual customers, **as the monthly purchase increases, the predicted outcome increase (more likely to churn).**

It is also found using EDA that total spending could reliably predict customer likeliness to churn.



Key Insights

1. Customer with total spending **below IDR 1.000.000 are more likely to churn (39.04%)** compared to customer who has spend above IDR 4.000.000 which **has average churn rate of 15%.**

Visualization of the data as a whole shows that customer churn do increases with monthly purchases in accordance to machine learning findings

Key Insights

When monthly purchase is **below IDR 40,000**, Non churn customer are consistently higher than the churn customer.

However, when monthly purchase is above **IDR 90,000,** there is an increase in the number of churn customer.

We perform the deployment model through streamlit which documentation can be seen through Github with the link provided below.

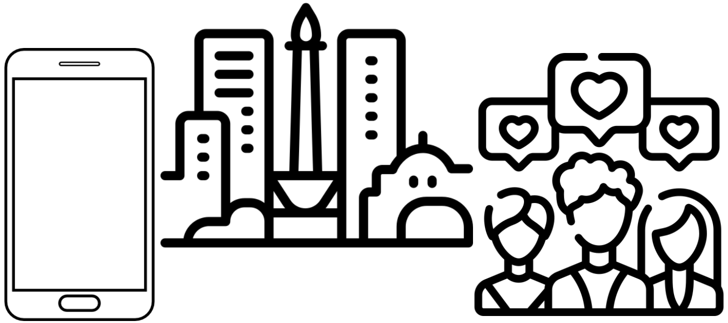
https://groundzero-telkomchurn.streamlit.app/

https://github.com/AndreasL7/telkom_customer_churn

Phase 4: Conclusion

Strategic recommendation from current dataset problem

Mid End - Low Tenure- Living in Jakarta

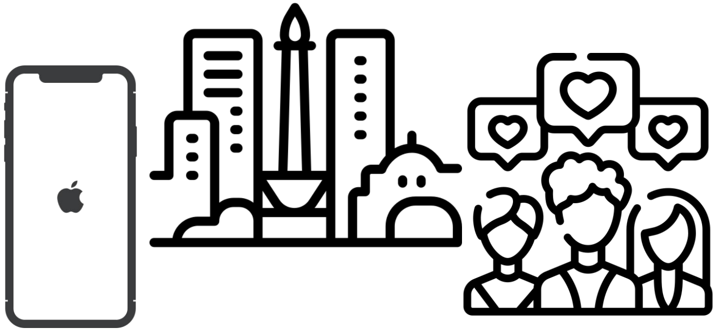


Suggestion:

1. For new user, recommend product video only, or music only, or games only. (High demand in general and low churn probability).
2. Introduce Customer Loyalty Program to increase tenure month such as level with rewarding program, etc
3. Introduce promotion for the first three month to reduce Monthly purchase for new customer

Strategic recommendation from current dataset problem

High end - low tenure - living in Jakarta



Suggestion:

1. Increase marketing on **games product** which has lowest churn rate but low customer base
2. For new user, recommend **education only and music only** product which has **high demand** among high end user and low churn rate
3. Introduce Customer loyalty program to increase tenure month such as level with rewarding program, etc
4. Introduce promotion for the first three month to reduce Monthly purchase for new customer

Appendix

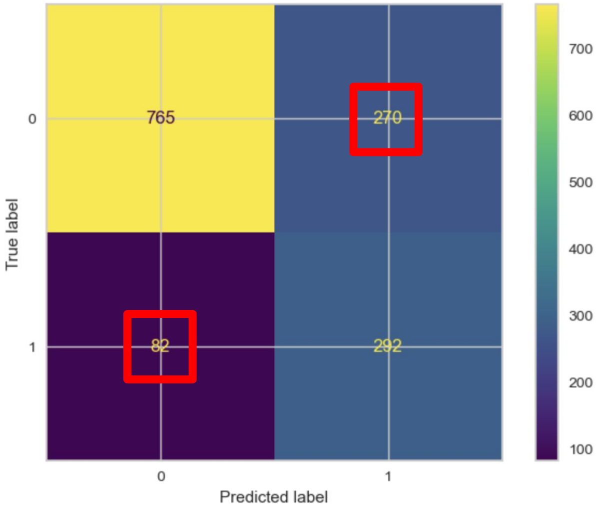
Model Performance

Train to optimise Recall

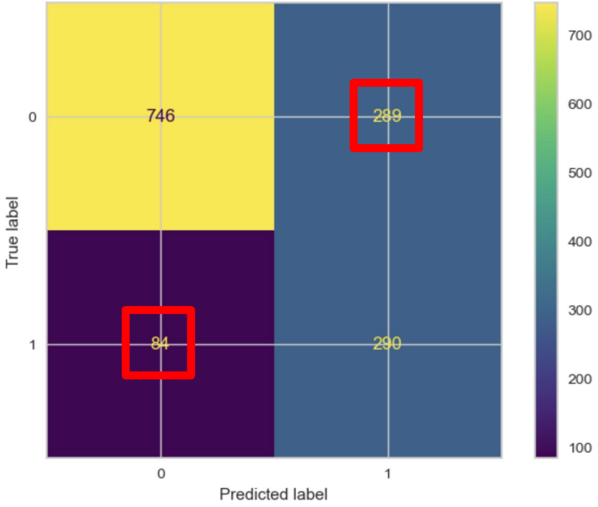
Metrics	XGBoost		Logistic Regression		Support Vector Classifier		Voting Classifier—Hard Voting	Voting Classifier—Soft Voting
	Based	Tuned	Based	Tuned	Based	Tuned	Based	Based
Accuracy	0.737	0.750	0.737	0.735	0.720	0.713	0.737	0.734
Precision	0.503	0.520	0.503	0.501	0.483	0.476	0.503	0.499
Recall	**0.797**	**0.781**	**0.775**	**0.775**	**0.783**	**0.821**	**0.794**	**0.797**
F1	0.617	0.624	0.610	0.609	0.597	0.603	0.616	0.614
ROC-AUC	0.824	0.838	0.833	0.833	0.825	0.826	-	0.837
Average Precision Score	0.633	0.650	0.642	0.642	0.639	0.588	-	0.647

Model Performance

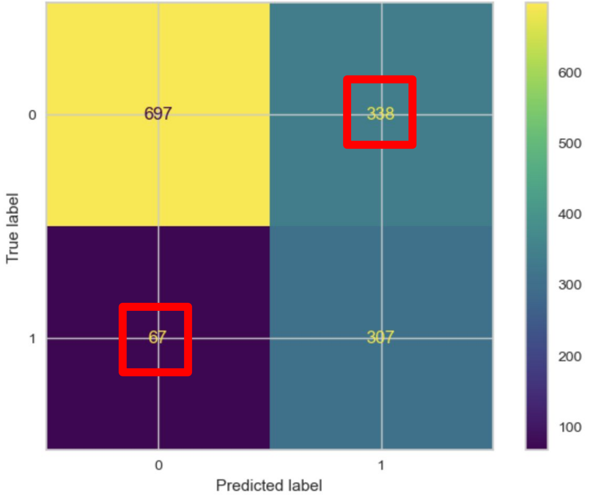
Confusion Matrix Comparison



XGBoost



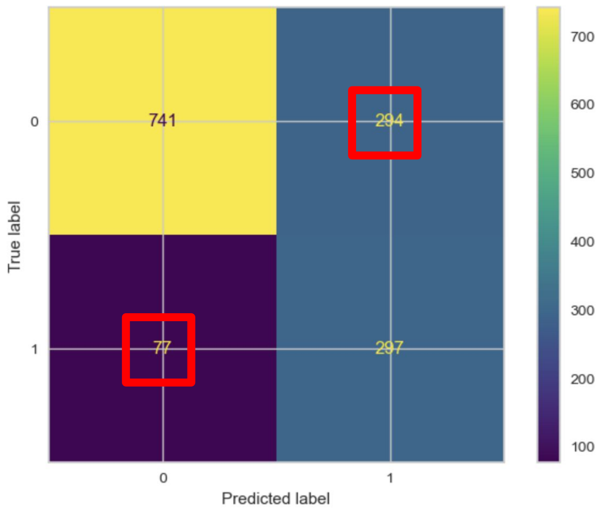
Logistic Regression



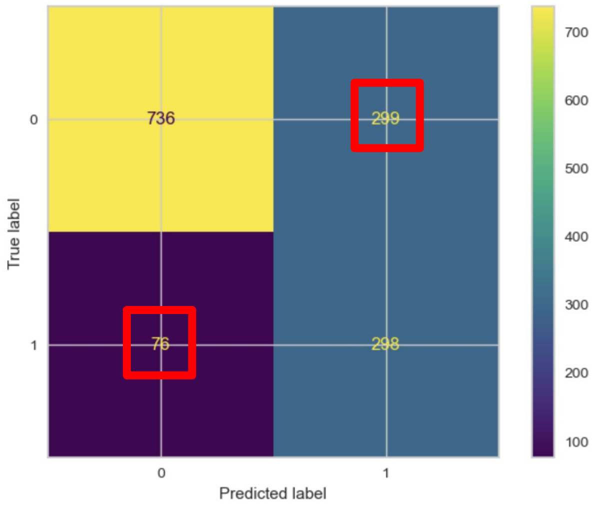
Support Vector Machine

Model Performance

Confusion Matrix Comparison



Hard Voting



Soft Voting

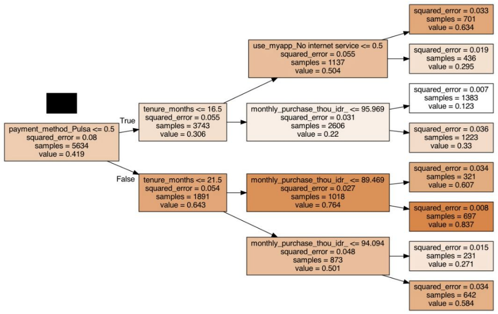
Adjustment of Probability Threshold

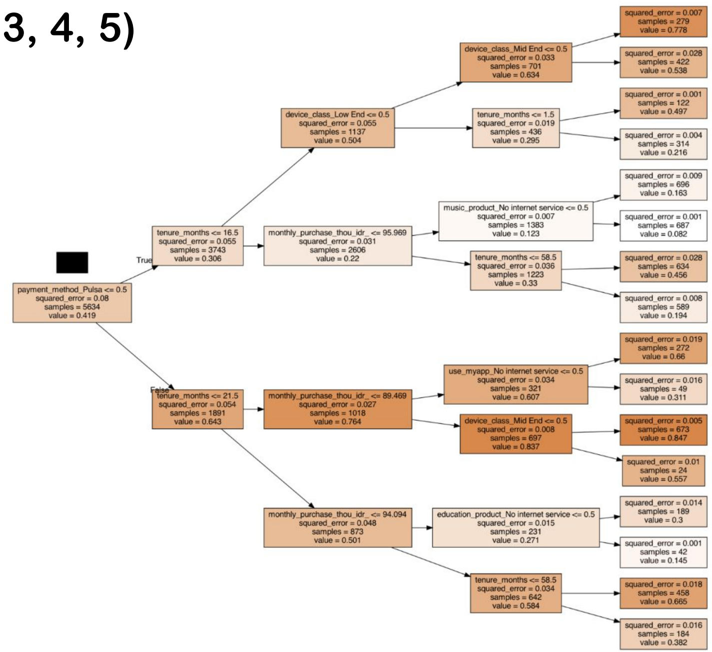
Default probability threshold = 0.50

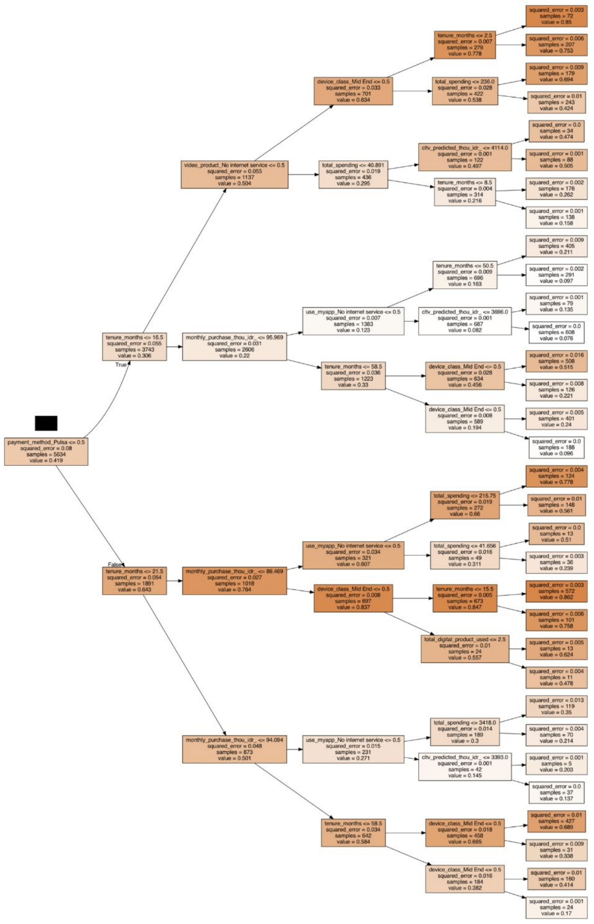
Metrics	XGBoost			Logistic Regression			Support Vector Classifier			Voting Classifier—Soft Voting		
	f1 (0.505)	fβ=1.5 (0.384)	fβ=1.8 (0.303)	f1 (0.596)	fβ=1.5 (0.364)	fβ=1.8 (0.343)	f1 (0.707)	fβ=1.5 (0.455)	fβ=1.8 (0.212)	f1 (0.586)	fβ=1.5 (0.404)	fβ=1.8 (0.242)
Accuracy	0.755	0.705	0.657	0.776	0.681	0.661	0.790	0.712	0.600	0.769	0.705	0.615
Precision	0.526	0.470	0.432	0.565	0.449	0.434	0.591	0.476	0.394	0.549	0.469	0.404
Recall	**0.775**	**0.874**	**0.922**	**0.687**	**0.885**	**0.909**	**0.679**	**0.832**	**0.944**	**0.723**	**0.845**	**0.952**
F1	0.627	0.611	0.588	0.620	0.595	0.587	0.632	0.605	0.556	0.624	0.603	0.567
ROC-AUC	0.838	0.838	0.838	0.833	0.833	0.833	0.826	0.826	0.826	0.837	0.837	0.837
Average Precision Score	0.650	0.650	0.650	0.642	0.642	0.642	0.588	0.588	0.588	0.647	0.647	0.647

Grokking XGBoost Result

Surrogate Models
(Decision Tree depth = 3, 4, 5)







Grokking XGBoost Result

Surrogate Models

Metrics	XGBoost (Tuned)	Decision Tree (Depth=3)	Decision Tree (Depth=4)	Decision Tree (Depth=5)
Accuracy	0.750	0.715	0.743	0.720
Precision	0.520	0.478	0.511	0.483
Recall	**0.781**	**0.773**	**0.733**	**0.794**
F1	0.624	0.590	0.602	0.601
ROC-AUC	0.838	0.734	0.740	0.743
Average Precision Score	0.650	0.429	0.445	0.438

End of Presentation