TELECOM CHURN PREDICTION

ENHANCED RANDOM FOREST APPROACH USING K-MEANS CLUSTERING

CONTENT

TEAM INTRODUCTION

David Kurniawan

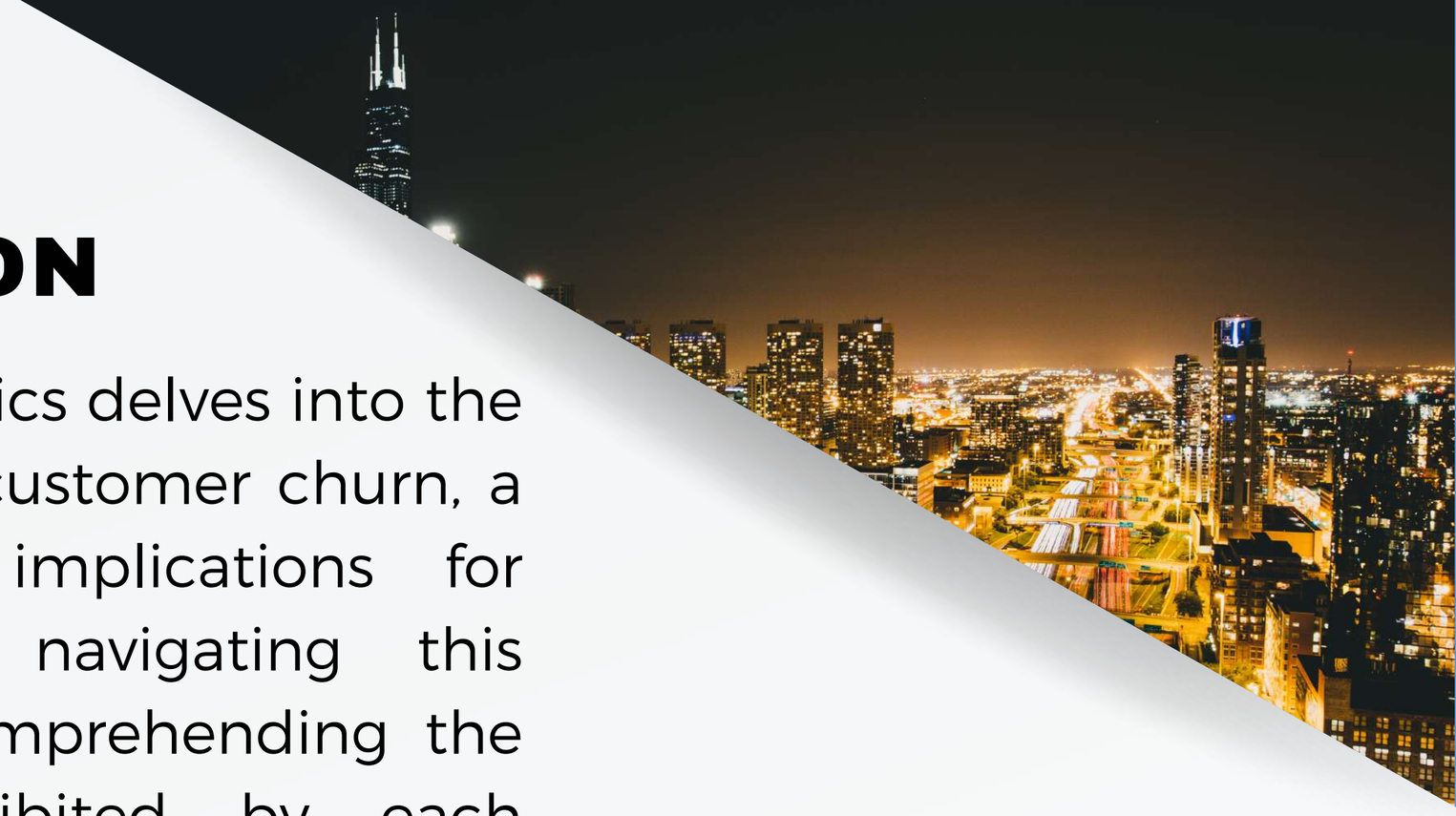
Maranatha Christian University
Majoring in Information System

Sherly Santiadi

Maranatha Christian University
Majoring in Informatics Engineering

Marselius Agus Dhion

Maranatha Christian University
Majoring in Information Systems



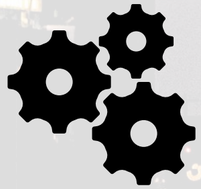
INTRODUCTION

The proposal is presented topics delves into the critical matter of predicting customer churn, a challenge with profound implications for business sustainability. In navigating this concern, our focus is on comprehending the distinct characteristics exhibited by each customer, thereby facilitating the development of a targeted strategy for those identified as potential churners within the company's service model.

DATA AND METHODOLOGY

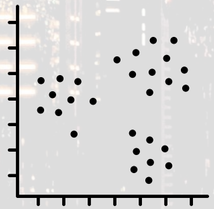
Random Forest

Using random forest approach with raw data as an input without any feature engineering



Feature Engineering

Using random forest approach with categorize data as an input



K-Means Clustering

Using k-means clustering to identify incorrect predictions in customer segmentation to enhance the random forest model

Conditional

Using conditional statement based on k-means clustering to enhance the random forest model

RESULT

https://github.com/sntdshrly/dsw2023/

https://bit.ly/telkomsel-dashboard

Categorize Regression Features

- Tenure months
  - New customer: 0 -12
  - Existing customer: 13 - 48
  - Loyal customer: > 48
- Monthly purchase
  - Low: < 50
  - Medium: 50 -100
  - High: > 100
- CLTV
  - Low: < 4000
  - Medium: 4000 - 6000
  - High: > 6000

Pearson Correlation

3 Top Features that Correlate Strongly with Churn Label

- Tenure months
- Payment months
- Device class

K-Means Clustering

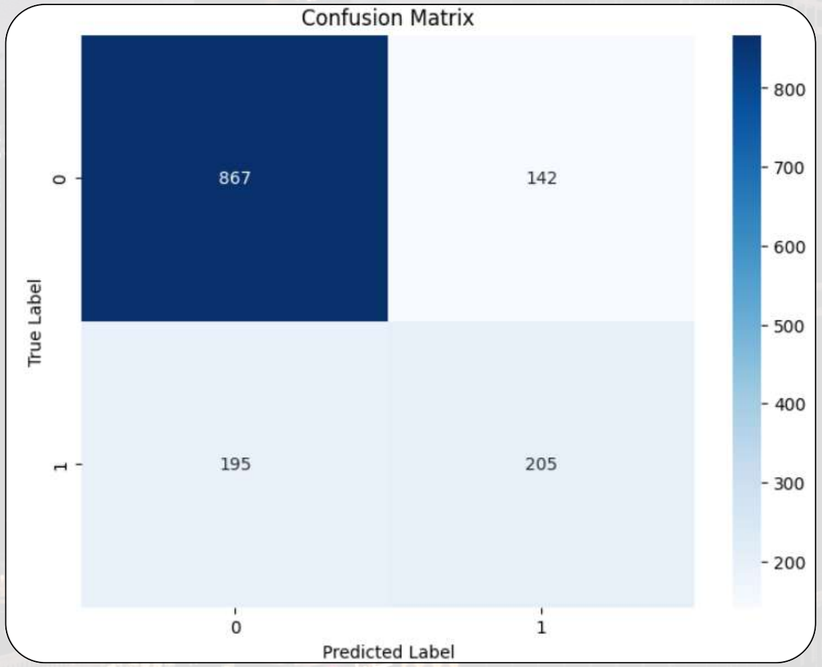
After performing the k-means clustering approach, we found that the top 2 prediction errors involve

- Existing or loyal customers who use pulsa as their payment method and have high-end device classes. In these cases, the model tends to predict non-churn, whereas the actual prediction is churn.

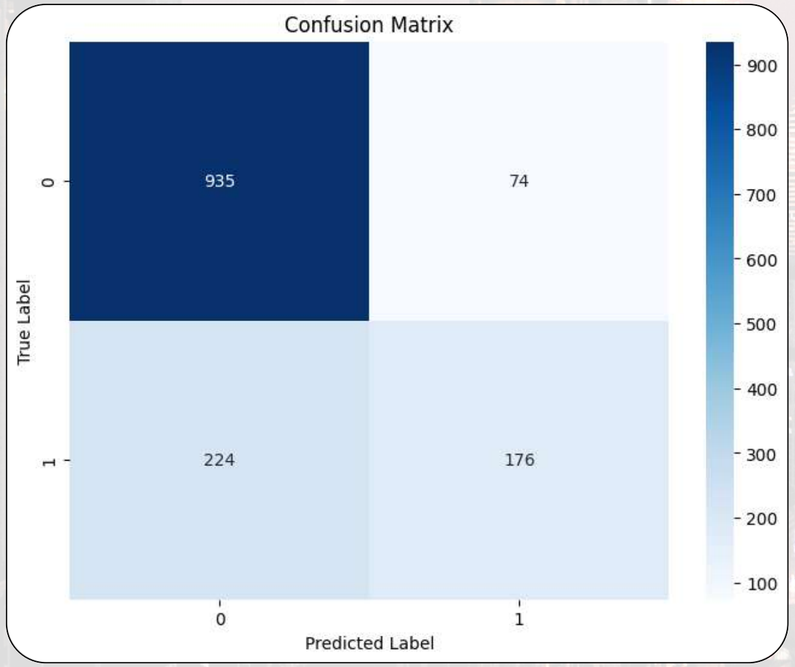
K-Means Clustering

- New customers who use pulsa as their payment method and have high-end device classes. In these cases, the model tends to predict churn, whereas the actual prediction is non-churn.

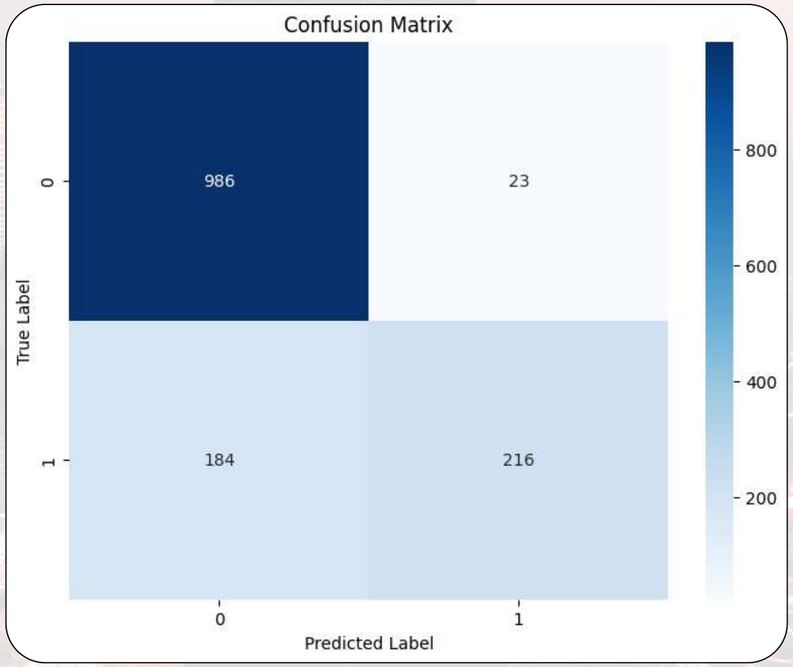
RANDOM FOREST



RANDOM FOREST WITH FEATURE ENGINEERING



ENHANCED RANDOM FOREST



CONCLUSION

The issue addressed in this proposal is how to predict customer churn. By understanding the characteristics of each customer, we aim to develop a tailored approach for those predicted to churn from the company's service model. One noteworthy observation is the tendency of customers who make pulsa payments and use high-end devices to churn over time. Consequently, a specialized approach is required for this customer segment—such as offering points that can be exchanged for vouchers.