



PROJECT PRESENTATION

Title: VIP_Churn_Predictor

By Krishna Priya

INTRODUCTION



Good morning, I am Krishna Priya.
Losing any customer is bad, but losing a 'VIP' customer is a disaster. Standard churn models are too simple—they often can't tell the difference between a one-time buyer and your best customer who is silently walking away. That's an expensive, preventable mistake, and my project is built to solve it."

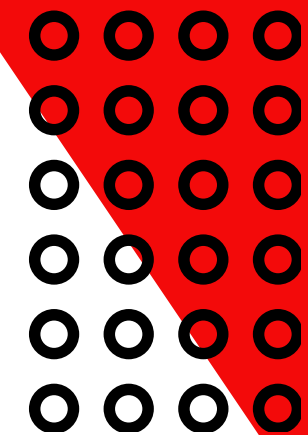
BACKGROUND

GENERAL OVERVIEW

This project is a two-stage "smart" churn predictor. A regression model predicts a customer's value, which is then fed into a classification model to predict their

HISTORY

This project is based on a classic marketing concept called RFM analysis, where businesses have long manually tracked a customer's Recency, Frequency, and Monetary value. My project simply automates and enhances this proven idea, using machine learning to turn a static, historical report into a live, predictive model.



GOALS

GOAL 01

- Build a regression model to predict a customer's future financial value, creating a "VIP Score."

GOAL 02

- Build a classification model that uses this "VIP Score" to predict which high-value customers are at risk of churning.

PROBLEMS

- The core problem is that losing a "VIP" customer is far more costly than losing a regular one. Standard churn models often fail by treating all customers equally, sending the same "at-risk" alert for a low-value buyer as for a high-value VIP. This is inefficient and wastes the marketing budget. We needed a "smarter" system that could specifically identify which of our most valuable customers were at the highest risk of leaving, so we could focus our efforts on them.

HYPOTHESIS

- Our central hypothesis is that a standard churn model, using only historical RFM features, is insufficient for this task. We propose that a two-stage model will be significantly more effective. We hypothesize that by first using a regression model to predict a customer's future financial value (a "VIP Score"), and then feeding that score as a new, synthesized feature into our final classification model, we can build a "smarter" predictor that is far more accurate at identifying the specific high-value customers who are at risk of leaving.

METHODOLOGY

METHOD 01

- The "Value" Model First, we used a Regression model (trained on RFM data) to predict a customer's future spending. This created a single "VIP Score" for each customer.

METHOD 02

- The "Risk" Model Second, we used a Classification model to predict churn. We fed it the original RFM data plus the new "VIP Score" to make the prediction "smarter."

DATA ANALYSIS

Our analysis began with the raw UCI Online Retail dataset, which contained over 500,000 transactions. The first critical step was data cleaning, where we removed all unusable rows, such as those with missing customer IDs or negative quantities (representing cancellations). We then transformed this raw transaction log into a powerful, customer-centric dataset by engineering Recency, Frequency, and Monetary (RFM) features for every unique customer, summarizing their entire purchase history into three key numbers. Finally, we created our two target variables, CLV_3_Month for regression and is_churned for classification, allowing us to build our predictive models.

CONCLUSION

In conclusion, this project successfully built and deployed a "smart" VIP churn predictor. We proved our hypothesis by first using a regression model to quantify customer value, which achieved an R-squared of 0.625. We then fed this "VIP Score" into a classification model, which successfully predicted churn with a realistic 67.6% accuracy. The entire, end-to-end pipeline was packaged into a live Streamlit web app, turning a complex analysis into a practical, real-time tool for making smarter business decisions.



THANK YOU

By Krishna Priya