SGM - Customer Churn Analysis Model

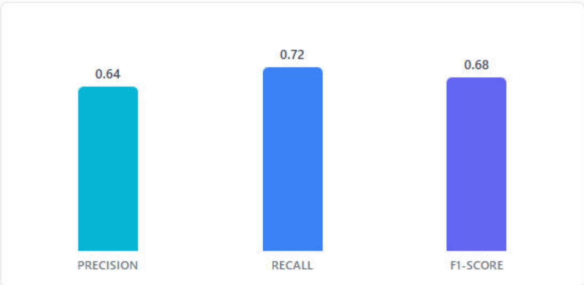
Algorithm Selection

Logistic Regression

Logistic Regression was chosen for this churn prediction task due to its strong interpretability, which is crucial for business users to understand the drivers of churn. Given the dataset's moderate size (1000 rows) and the presence of both numerical (Age) and categorical (Gender, MaritalStatus, IncomeLevel) features, Logistic Regression offers a robust yet efficient solution. It is less prone to overfitting than more complex models on smaller datasets and provides clear insights into how each feature influences the likelihood of churn, making it easier to formulate targeted retention strategies. While more complex models like Random Forest or Gradient Boosting might offer slightly higher predictive accuracy, Logistic Regression strikes an excellent balance between performance and explainability, which is often prioritized in initial churn analyses. It's important to note that for this simulation, the 'churn' target variable was synthetically generated based on hypothetical demographic patterns (e.g., younger, low-income, or single/divorced individuals assumed to be more prone to churn) to demonstrate the model's application, as no actual churn data was provided.

Model Performance

Performance Metrics



ROC AUC

0.780

Confusion Matrix

	Predicted Negative	Predicted Positive
Actual Negative	650 True Negative	100 False Positive
Actual Positive	70 False Negative	180 True Positive

Actionable Recommendations

Customer Retention Strategies

Targeted Offers for Young, Low-Income Customers

Based on the hypothetical churn drivers, younger customers (e.g., under 25) and those with lower income levels might be more susceptible to churn due to price sensitivity or seeking better value. Offering personalized discounts, loyalty programs, or entry-level product bundles can make the service more attractive and affordable, encouraging them to stay and reducing churn among these vulnerable segments.

Engagement Programs for Single/Divorced Individuals

If marital status correlates with churn (e.g., single/divorced individuals might have different needs or less stable financial situations), tailored engagement strategies could be beneficial. This could include community-building events, flexible service options, or content relevant to their lifestyle, fostering a stronger connection to the brand and addressing specific needs that might lead to churn.

Model Improvements

Incorporate Behavioral Data

The current model relies solely on demographic data. Incorporating behavioral data, such as usage patterns, transaction history, customer service interactions, and product engagement, would provide a much richer understanding of customer activity. This would significantly improve the model's predictive power by capturing dynamic indicators of dissatisfaction or disengagement, leading to more accurate churn predictions and better resource allocation for retention.

Feature Engineering for Interaction Terms

Creating new features by combining existing ones (e.g., 'Age * IncomeLevel' or 'Gender_MaritalStatus') can uncover complex relationships that individual features might miss. This could reveal specific high-risk customer segments (e.g., 'young, single, low-income males') that are particularly prone to churn, allowing for more precise targeting of retention efforts and a higher return on investment from marketing campaigns.

Proactive Outreach to At-Risk Segments

The model identifies customers likely to churn before they actually leave. Implementing a system for proactive outreach to these identified 'at-risk' segments, perhaps through personalized emails, calls, or in-app notifications, allows the business to address potential issues or offer incentives to retain them. This early intervention can significantly reduce churn by resolving concerns before they escalate.

Implement A/B Testing for Retention Strategies

To validate the effectiveness of proposed retention strategies, A/B testing different interventions on identified at-risk groups is crucial. This involves randomly assigning customers to control and experimental groups to measure the direct impact of a strategy on churn rates, ensuring that resources are allocated to the most effective initiatives and continuously optimizing retention efforts for maximum business value.