Advanced Credit Card Customer Churn Prediction Using Ensemble and Deep Learning

Written by: Erika A.

1. Abstract

With the continuous development of the banking industry, the rapid competition in providing the best information and Internet technology has increased. Traditional machine learning methods for building predictive models for credit card customer churn are insufficient for effective customer retention and management. Therefore, this study was conducted using a syntenic dataset of 100,000 users and aimed to balance the data using sampling methods to construct a credit card customer churn classification model to examine the probability of customer churn. The applied SMOTEENN was used to deal with imbalanced banking data due to research. This led to the F1 and AUC scores of the classification model being improved and optimized by AdaBoost, reaching 46.10% and 78.61% ROC AUC, respectively, which was optimal compared to the other seven machine learning models. Therefore, the AdaBoost classifier was identified as the optimal model for customer churn analysis. Furthermore, the feature importance plot was used to analyze which features significantly impacted the model, and the top features included customer service contacts, gender, recent inactivity, credit usage percentage by group, activity consistency, and number of products with the bank. With detailed business impact analysis, it was established that the model can generate a monthly 89.1% ROI with a $384,750 net benefit.

2. Introduction

Customer churn, also known as customer attrition, is the rate at which customers discontinue using the services provided by a business. Customer churn is an important metric because it is generally more expensive to acquire new customers than to retain existing customers. The relationship between financial health and customer satisfaction with credit cards reveals critical churn indicators. J.D. Power (2024) [1] found that satisfaction scores are 61 points lower among cardholders carrying debt (580) than among those without debt (641), while 54% of customers are now classified as financially unhealthy, creating a substantial at-risk population for potential churn.

Customer churn patterns are complex when encountering imbalanced data, where customer churn represents the minority class. Therefore, this poses a threat to traditional machine learning classification approaches. This study provides countermeasures to ensure optimized prediction of customer churn when utilizing advanced ensemble learning and deep learning methods with a sampling balance strategy.

The main contributions of this study are as follows: (1) relevant feature-engineered features that are categorized under financial health, behavioral patterns, risk scoring, and the RFM framework are created; (2) the SMOTENN sampling technique, which effectively solves the bank data imbalance problem, is used; (3) based on the traditional machine learning models for customer churn, the proposed improved models include hyperparameter optimization. These models include logistic regression, decision tree, random forest, XGBoost, K-nearest neighbors (KNN), Gaussian Naïve Bayes, AdaBoost, TensorFlow, and PyTorch neural network architectures (4) Exploring the feature importance using feature importance plots (5) Exploring the evaluation metrics, including Precision, Recall, F1, and ROC AUC.

3. Literature Review

3.1 Customer Churn prediction

Feature engineering has been identified as a critical component of effective customer churn predictive models in banking. Brito et al. (2024) [2] emphasize that the feature engineering process is "fundamental for building new and informative features to better characterize customers' behavior, directly impacting model effectiveness" when conducting churn prediction framework model. The authors specifically address the gap where "extant research on the topic mainly focuses on the type of model developed to predict churn, devoting little or no effort to data preparation methods."

Creating features based on financial health and behavioral patterns is fundamental when investigating customer behaviors in churn prediction. Kaya et al. (2018) [3] focused on the investigation of "spatio-temporal patterns and entropy of choices underlying customers' financial decisions", noting that the "traditional efforts in the financial domain mainly focus on domain specific variables such as product ownership or service usage aggregation, however, without considering dynamic behavioral patterns of customers' financial transactions". Their work demonstrated that behavioral features significantly outperform demographic features in churn prediction.

A similar framework of engineering features based on the RFM model encompasses state-of-the-art data preprocessing, including "feature engineering with recency, frequency, and monetary value concepts to address the imbalanced dataset issue" (Brito et al., 2024)[4]. This approach has proven particularly effective in the banking context, where transactional behavioral patterns are crucial indicators.

The SMOTE ENN (SMOTE + Edited Nearest Neighbors) technique represents a hybrid approach to address class imbalance in customer churn datasets, compared to only oversampling or undersampling techniques. The method was pioneered by Batista et al (2004) , [5] which use both SMOTE to generate synthetic data using the minority class, and ENN to remove some observations from both classes if "any example that is misclassified by its three nearest neighbors is removed from the training set."

Hyperparameter optimization is essential for maximizing the model performance in customer churn prediction. Peng et al [6] implemented hyper-parameter optimization to their improved XGBoost and defined that "genetic algorithms are used to optimize the parameter of the composite XGBoost algorithm after data equalization to obtain bank customer predict model", showing significant improvements over parameter optimization methods.

Deep learning approaches have shown promising results in predicting customer churn. Recent research proposes “a hybrid neural network-based customer churn prediction model, CCP-Net” that “uses Multi-Head Self-Attention to learn the global dependencies of the input sequences, combines with BiLSTM to capture the long-term dependences in the sequential data, and CNN to extract local features”. This CCP-Net model achieved superior performance compared to other single deep learning algorithms across all four datasets (Telecom, Bank, insurance, News) with results of 92.19% Precision, .96% precision, 95.97% precision, and 95.12% precision, respectively (Liu et al., 2024)[7].

The integration of explainable AI techniques has become crucial for understanding model decisions in customer churn prediction. Recent study (Salih et al., 2024)[8] emphasizes that SHAP “employs cooperative game theory’s individual features for a given prediction in a way that is both intuitive and mathematically sound”, providing both local and global model interpretability, and LIME which “works by fitting a so-called surrogate model that approximates the predictions of the underlying black-box model” and “builds upon four basic principles: Local, Interpretable, Model-Agnostic, Explanation”, making it particularly valuable for individual prediction explanations.

3.2 Evaluation Metrics

The following standard metrics suited for imbalanced datasets, are Precision, Recall, F1 Score and the Receiver Operating Characteristic Area Under Curve (ROC AUC). Precision is the proportion of all the model's positive classifications that are actually positive. Recall is the true positive rate at which the percentage of customers who are classified are also estimated correctly as churners. The F1 score is a combination of precision and recall, providing a balanced and reliable measure for imbalanced data, where accuracy can be misleading. Error Rate is the error rate is the proportion of incorrectly classified observations out of all customers classified.

The Receiver Operating Characteristic Area Under Curve (ROC-AUC) is where the ROC is a visual representation of the ability to classify correctly, and AUC measures the area under the ROC curve. A random classifier has an AUC of 0.5, and a perfect classifier has an AUC of 1.

The following equations are used:

$$\text{Recall} = \frac{TP}{TP + FN}$$

$$\text{Precision} = \frac{TP}{TP + FP}$$

$$\text{F1} - \text{Score} = \frac{2 * (precision * recall)}{precision + recall}$$

$$\text{Error Rate} = \frac{FP + FN}{TP + TN + FP + FN}$$

3.3 Approaches

Logistic Regression attempts to create a regression model based on data with a binary response model. The logit function was used to determine the probability of a binary outcome.

$$z = \beta_0 + \beta_1 * x_1 + \beta_2 * x_2 + \cdots . + \beta_n * x_n + \; \ldots . + \epsilon$$

$$P(churn) = 1/\left(1 + e^{(-z)}\right)$$

K-nearest neighbors (KNN) is a supervised and non-parametric algorithm whichere the new data are similar to the existing data and assigns the new data to the category that is most similar to the existing category based on the distance between two points.

A decision tree starts with the root at the top and knots that are interconnected by branches. The nodes were classified as internal or terminal. At each internal node, a specific attribute is tested, and the result guides the selection of different branches, eventually leading to the terminal node. The terminal nodes or “leaves” correspond to a classification.

Random Forest is a combination of decision trees as the building blocks. The model generates a predefined number of trees, takes a cut of the total number of trees, and uses it as its predictor. It adjusts the number of trees and cuts in the algorithms to provide the best possible performance.

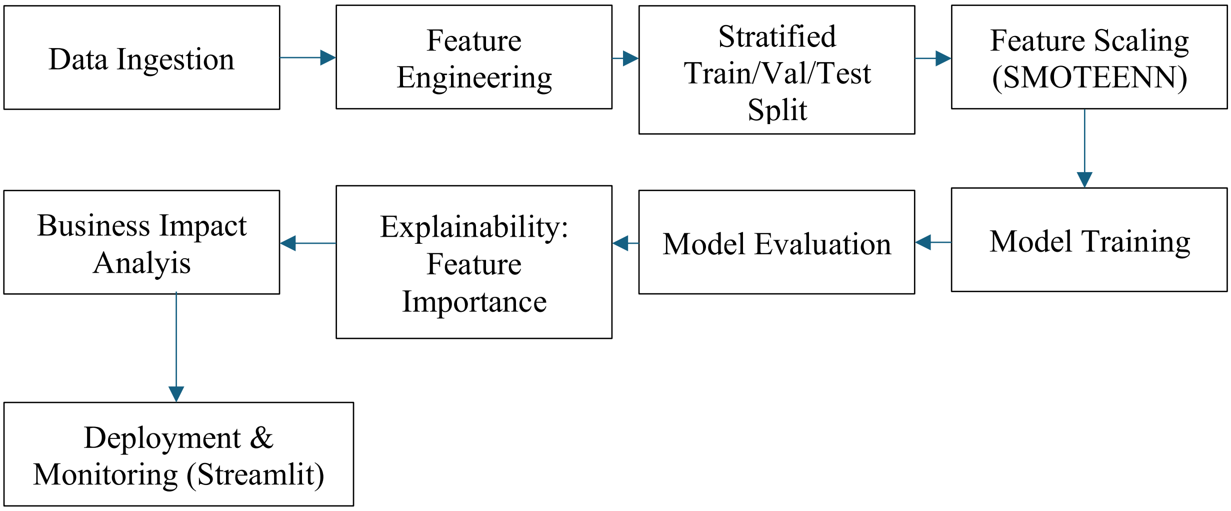
Gaussian Naïve Bayes is a type of Naïve Bayes method that works on continuous attributes and data features that follow a Gaussian distribution through the dataset. This “naïve” assumption simplifies the calculations and makes the model fast and efficient. This method is effective for continuous data because it assumes that each feature follows a Gaussian (normal) distribution. When this assumption holds true, the algorithm performs well.

Adaptive Boosting (AdaBoost) is an “ensemble learning technique that combines multiple weak classifiers to create a strong classifier. It works by sequentially adding classifiers to correct the errors made by previous models, giving more weight to the misclassified data points”.

Extreme Gradient Boosting (XGBoost) is an optimized implementation of Gradient Boosting and is an ensemble learning method that combines multiple weak models to form a stronger model. XGBoost uses decision trees as base learners and combines them sequentially to improve the model performance. Each new tree is trained to correct the errors made by the previous tree, and this process is called boosting. It starts with a base learner from the decision tree, which is trained and then predicts the average target variable. After training the first tree, the errors between the predicted and actual values were calculated. The next tree is trained on the errors of the previous tree, and this step attempts to correct the errors made by the first tree. Then, the process continues with each new tree trying to correct the errors of the previous trees until a stopping criterion is met. The final prediction is the sum of the predictions from all trees.

Multilayer Perceptron Classifier Neural Network implemented multilayer perceptron (MLP) – “a name for a modern feedforward neural network consisting of fully connected neurons with nonlinear activation functions, organized in layers, notable for being able to distinguish data that is not linearly separable". (Wikipedia, 2025)[9] This trains on feature matrix X and churn target variable using gradient descent with backpropagation to minimize cross-entropy loss. It takes feature matrix X and target labels y as input, learns weight matrices stored in clf.coefs_, and supports both multi-class classification (using softmax output) and multi-label classification (using logistic function with 0.5 threshold). The model provides a flexible architecture configuration through hidden_layer_sizes, regularization via the alpha parameter, and multiple solver options (lbfgs, adam, sgd), while offering two prediction methods: predict() for class labels and predict_proba() for probability estimates, making it a comprehensive sklearn-integrated neural network solution for classification tasks.

3.4 Model Architecture Diagram


Data Ingestion
Feature Engineering
Stratified Train/Val/Test Split
Feature Scaling (SMOTEENN)
Model Training
Model Evaluation
Explainability: Feature Importance
Business Impact Analyis
Deployment & Monitoring (Streamlit)

4. Data

This study analyzes a synthetic Credit Card Churn Dataset of 100,000 customers (21 columns; 48.76MB). The binary target Attrition_Flag distinguishes attrited vs existing customers (20.29% prevalence). The features fall into financial behavior, customer relationships, and risk indicators. Table 1 summarizes the schema and type of each feature.

Feature	Description	Data Type
CLIENTNUM	Customer ID	object
Customer_Age	Age of the customer	int64
Gender	Gender (M, F)	object
Dependent_count	Number of dependents	int64
Education_Level	Education level (High School/College Graduate/Post-Graduate/ Doctorate/Uneducated/ Unknown)	object
Marital_Status	Married/Single/Divorced/Unknown	object
Income_Category	Income brackets	object
Card_Category	Blue/Silver/Gold/Platinum	object
Months_on_book	Relationship tenure (months)	int64

Total_Relationship_Count	Products with bank	int64
Months_Inactive_12_mon	Inactivity in last 12 months	int64
Contacts_Count_12_mon	Service contacts last 12 months	int64
Credit_Limit	Available credit	float64
Total_Revolving_Bal	Outstanding balance	float64
Avg_Open_To_Buy	Credit_Limit - Revolving	float64
Total_Amt_Chng_Q4_Q1	Q4/Q1 amount ratio	float64
Total_Trans_Amt	Total transaction amount	float64
Total_Trans_Ct	Number of transactions	int64
Total_Ct_Chng_Q4_Q1	Q4/Q1 count ratio	float64
Avg_Utilization_Ratio	Revolving / Credit_Limit - Credit usage percentage	float64
Last_Transaction_Date	Last transaction timestamp	datetime64[ns]

Table 1: Dataset schema

The summary of the sample is that existing customers are 79,714 (79.7%) and attrited customers are 20,286 (20.3%). The average mean age was 45.5. The credit limit is $12,294, utilization is 40%, transaction amount is $6,882, transaction count is 54.1, and tenure is 35.5 months on average.

5. Data Quality and Exploratory Analysis

5.1 Missing Data Analysis

The Last_Transaction_Date variable shows measurable missingness of approximately 0.045%. The following missing values were dropped, leading to no missing values in the data.

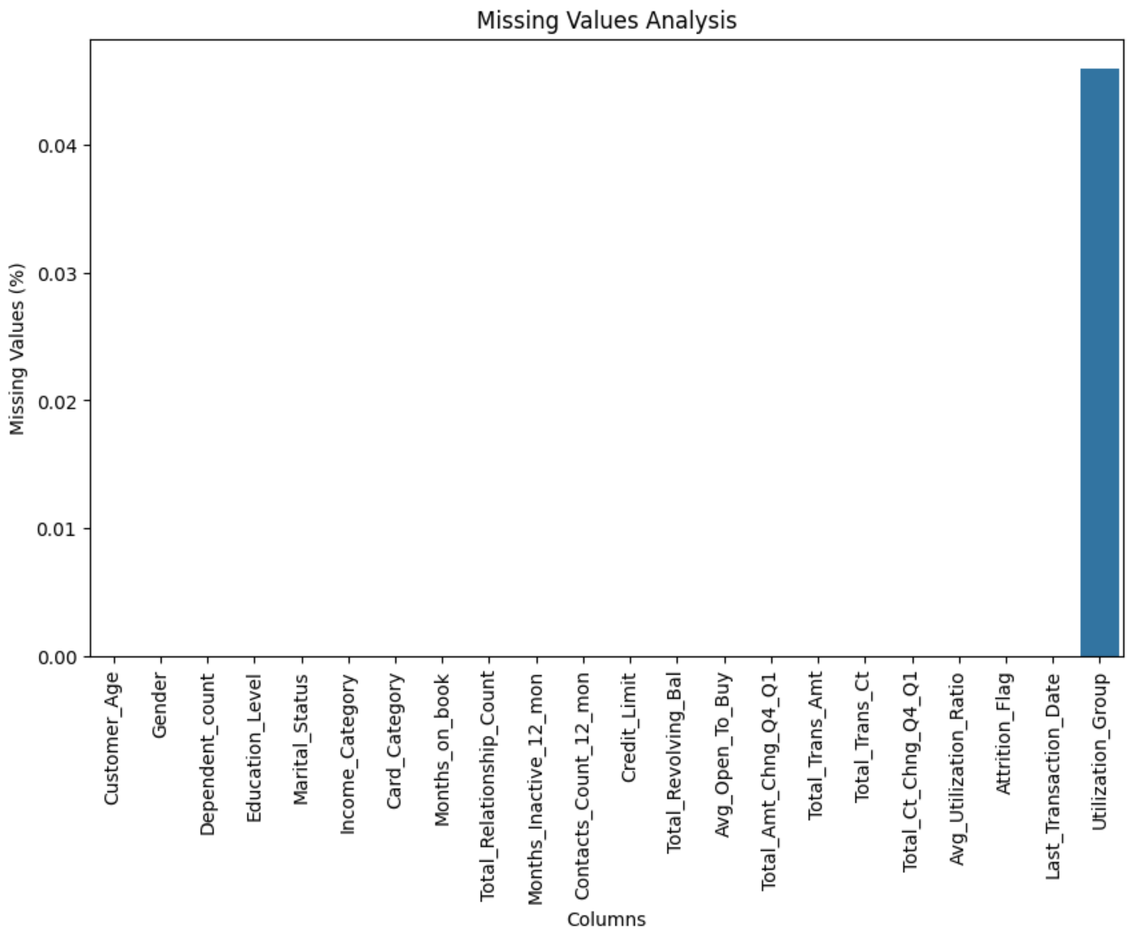


Fig 1: Missing Value Analysis

After removing the missing values, the class balance of the existing versus attrited customers revealed an approximately 80/20 imbalance dataset. Therefore, this moderate imbalance indicates that accuracy is a misleading metric in this case.

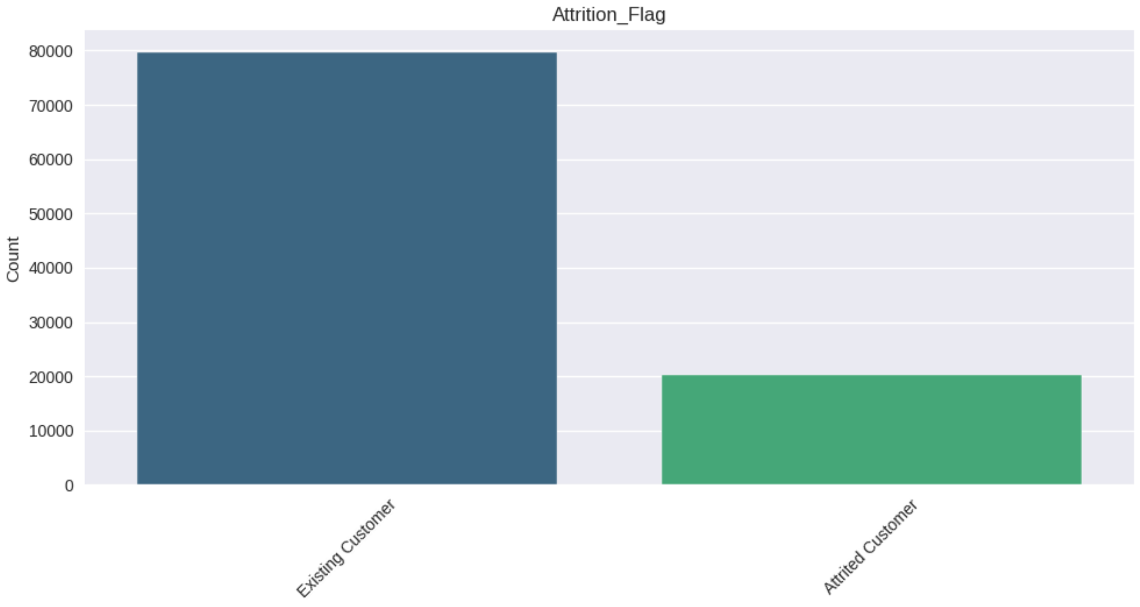


Fig 2: Customer Churn Imbalance

5.2 Consistency Data Quality Analysis

Consistency checks indicated an 80.01% pass rate for credit utilization business rules, with 0% negative balance or impossible transactions.

Churn patterns show strong nonlinearities with inactivity and a meaningful U-shaped pattern with utilization. The following tables list the selected churn rates across key cohorts.

Months Inactive (12m)	**Churn Rate**
0 months	10.8%
1 month	16.3%
2 months	16.5%
3 months	40.6%
4 months	40.7%
5 months	41.5%
6 months	41.7%

Table 2: Churn vs. recent inactivity.

Utilization Group	**Churn Rate**
Low (0–10%)	21.0%
Moderate (10–30%)	17.4%
High (30–70%)	21.3%
Very High (70%+)	21.3%

Table 3: Churn vs. utilization group.

Transaction Count Group	**Churn Rate**
Low (0–30)	30.2%
Medium (31–60)	17.7%
High (61–90)	17.4%
Very High (90+)	13.4%

Table 4: Churn vs. transaction activity.

Total_Relationship_Count	**Churn Rate**
1	35.1%
2	24.2%
3	24.4%
4	13.0%
5	12.7%
6	13.3%

Table 4: Churn vs. relationship depth.

Furthermore, we can observe that the Blue card has the largest churn share (30.5%), whereas premium tiers (Silver, Gold, and Platinum) are more stable, meaning they have consistent deeper engagement and richer benefits.

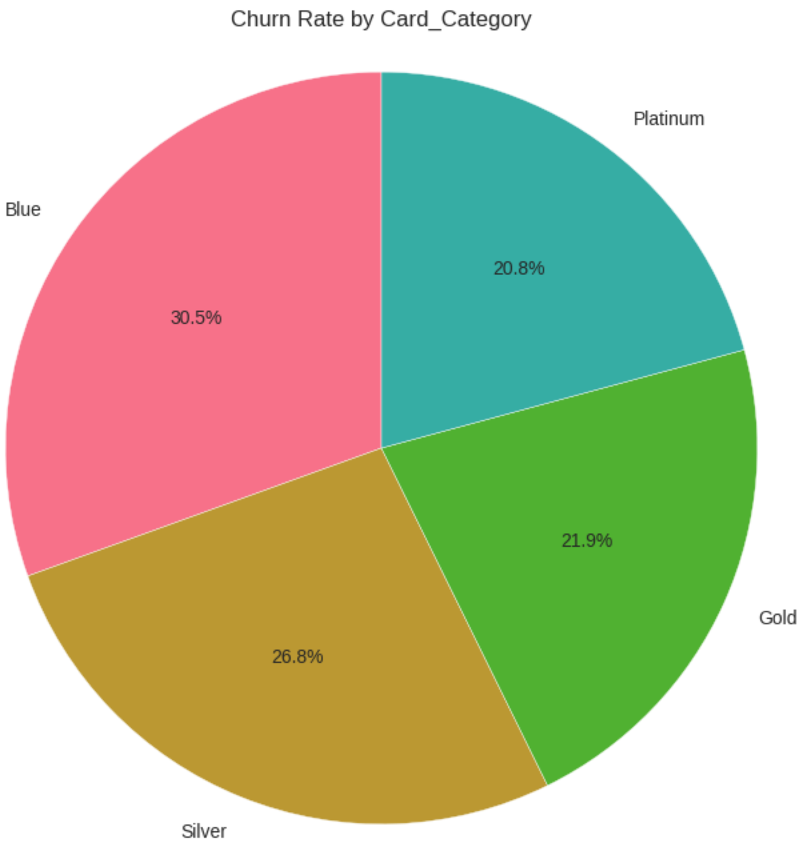


Fig 3: Churn by Card Category

As shown in the plot of the churn share across income brackets, there are near-flat rates of approximately 16%, which indicates that income contributes little incremental prediction n beyond the churn behavior. The plot of churn share by education level shows that the churn rate is slightly higher for those with a doctorate degree and lowest for those with a postgraduate degree.

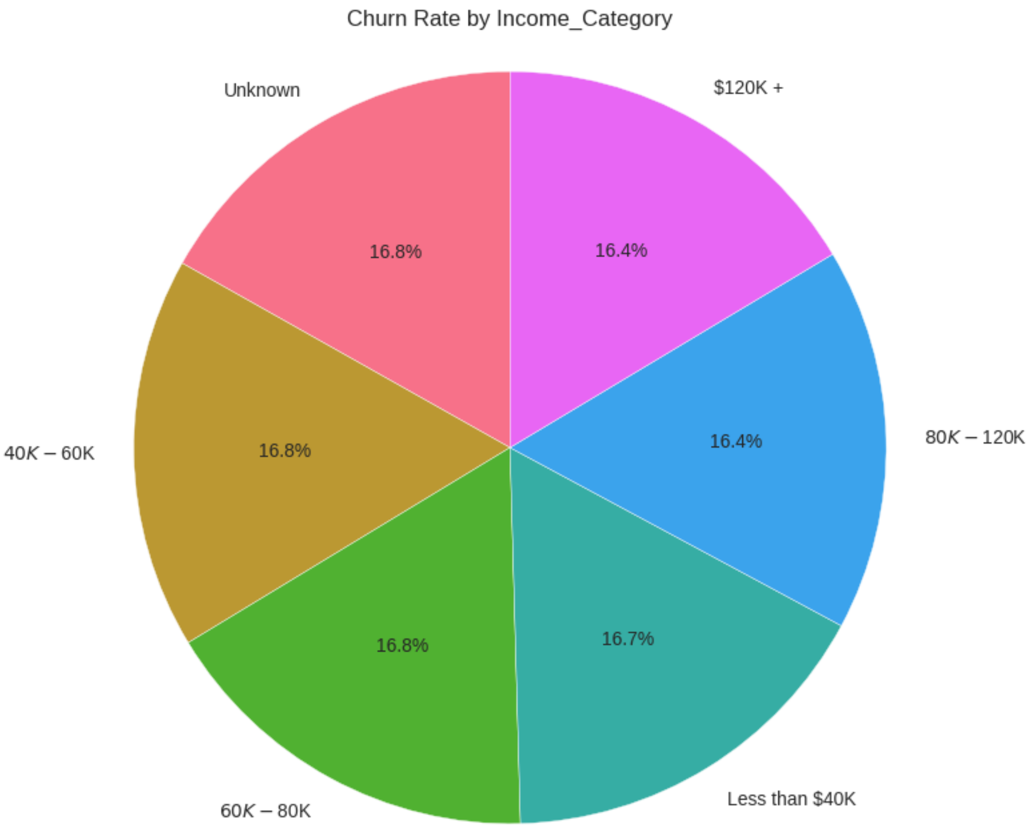


Fig 4: Churn by Income Category

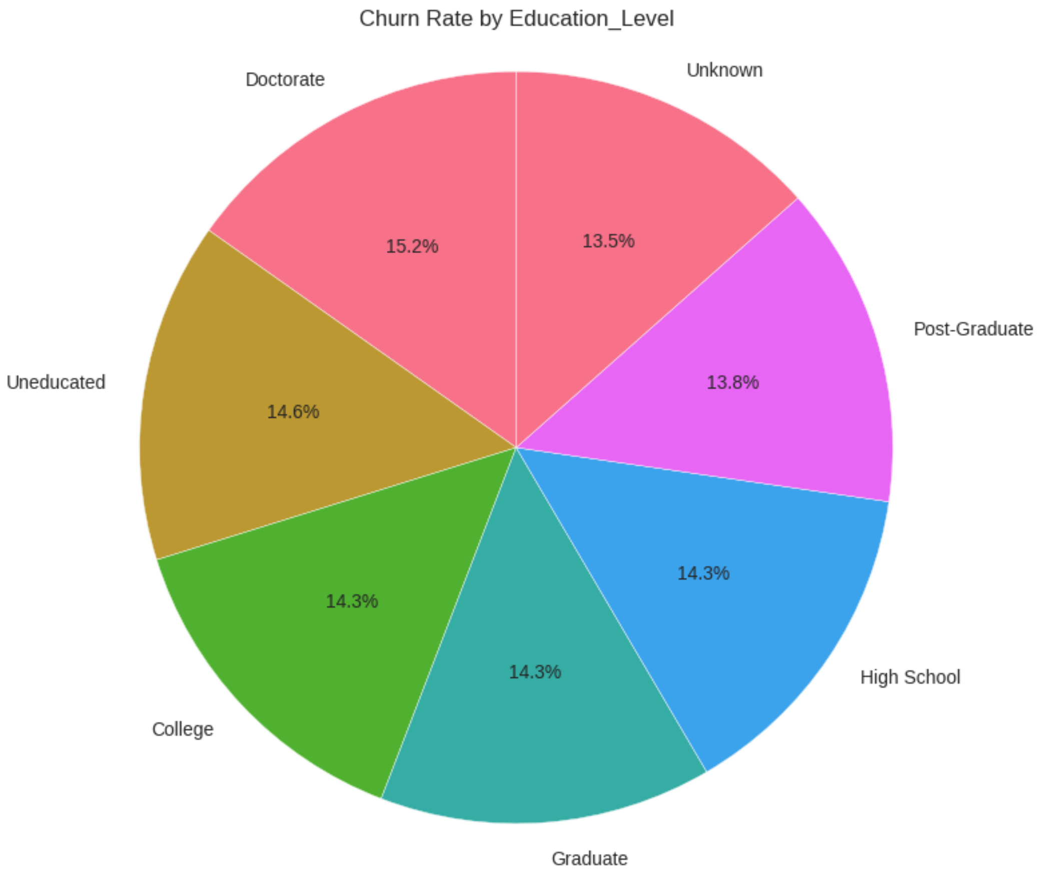


Fig 5: Churn by Education Level

Credit utilization patterns reveal a critical U-shaped relationship with customer churn behavior, where optimal retention occurs within the 10-30% utilization range. As demonstrated in the credit utilization analysis table, customers maintaining low utilization (10-30%) exhibit the lowest churn rate at 17.4%, representing the sweet spot for customer engagement. This finding suggests that customers in this range demonstrate healthy financial management while actively engaging with their credit products. Statistical analysis confirms the significance of utilization differences, with a two-sample test revealing that low-utilization customers (18.6% churn) significantly outperform high-utilization customers (21.3% churn) with a t-statistic of -6.572 and p-value less than 1×10^-6.

Interestingly, both extremes of the utilization spectrum present elevated churn risks. Very low utilization (0-10%) customers demonstrate a 21.0% churn rate, potentially indicating disengagement or the presence of dormant accounts where customers may be utilizing competitor cards as their primary payment method. Conversely, high-utilization (70-90%) customers also exhibit a 21.3% churn rate, suggesting payment stress and potential financial difficulties that could drive customers to seek alternative financial products or institutions. The negligible correlation (r = 0.005) between the utilization ratio and credit limit indicates that utilization patterns are driven by customer behavior rather than credit availability, making utilization a pure behavioral indicator of engagement and financial health.

Utilization Category	**Total Customers**	**Churned Customers**	**Churn Rate**	**Avg Credit Limit ($)**	**Avg Transaction Amount ($)**
Very Low (0–10%)	12,337	2,586	21.0%	12,117	6,840
Low (10–30%)	25,196	4,386	17.4%	12,307	6,898
Moderate (30–70%)	49,932	10,638	21.3%	12,341	6,894
High (70–90%)	12,489	2,654	21.3%	12,258	6,845
Very High (90–100%)	—	—	—	—	—

Table 5: Credit Card Utilization Analysis

Transaction volume analysis reveals a strong inverse relationship between customer activity levels and churn propensity, establishing transaction frequency as one of the most powerful predictors of customer retention. The transaction count analysis demonstrates a clear pattern in which churn rates decrease dramatically as transaction activity increases, falling from 30.2% for low-activity customers (0-30 transactions) to just 13.4% for highly active customers (90+ transactions). This 16.8 percentage point difference underscores the fundamental principle that customer engagement through active usage serves as the strongest hedge against attrition.

The transaction amount quartile analysis further reinforces this relationship, showing that customers in the lowest spending quartile experience a substantially higher churn rate of 26.9% compared to the relatively stable 17.5-19.0% range observed across the upper three quartiles. This pattern suggests that while the absolute transaction amount matters, it reaches a threshold effect where incremental increases beyond the first quartile provide diminishing marginal returns for retention. The quarterly change analysis (Q4/Q1 ratios) provides additional insights into customer trajectory, with declining usage customers (Q4/Q1 < 0.7) representing approximately 44,000 customers with a 20.2% churn rate, identifying a prime early-warning cohort for proactive intervention strategies.

Transaction Count Category	**Total Customers**	**Churned Customers**	**Churn Rate**	**Avg Transaction Amount ($)**	**Avg Utilization**
Low (0–30)	26,194	7,904	30.2%	3,433	0.403
Medium (31–60)	35,236	6,242	17.7%	6,360	0.399
High (61–90)	23,975	4,164	17.4%	8,590	0.399
Very High (90+)	14,549	1,954	13.4%	11,544	0.401

Transaction Amount Category	**Total Customers**	**Churned Customers**	**Churn Rate**	**Avg Transaction Count**	**Avg Credit Limit ($)**
Q1 (Lowest)	24,995	6,718	26.9%	23.818	12,285
Q2 (Low–Medium)	24,986	4,754	19.0%	54.033	11,277
Q3 (Medium–High)	24,985	4,368	17.5%	69.003	10,400
Q4 (Highest)	24,988	4,424	17.7%	69.474	15,215

Amount Change Category (Q4/Q1)	**Total Customers**	**Churned Customers**	**Churn Rate**	**Avg Count Change**
Large Decrease	22,919	4,656	20.3%	0.725
Moderate Decrease	28,131	5,652	20.1%	0.725
Stable	32,370	6,602	20.4%	0.728
Moderate Increase	10,325	2,072	20.1%	0.729
Large Increase	3,277	685	20.9%	0.726

Table 6: Transaction Volume Trends Analysis

Payment behavior patterns reveal critical insights into customer financial stress and its direct correlation with churn propensity. The payment stress level analysis demonstrates that while moderate stress customers (19.3% churn) represent the most stable segment, both low stress (21.0%) and high stress (21.4%) customers exhibit elevated churn rates for different underlying reasons. Low stress customers may demonstrate higher churn due to minimal financial ties to the institution, making switching decisions easier, while high stress customers face financial pressures that drive them to seek alternative solutions or consolidate their financial relationships.

The balance carrying behavior analysis presents one of the most significant behavioral differentiators, with balance carriers demonstrating a 25.6% churn rate compared to 19.0% for non-balance carriers—a difference of 6.6 percentage points that represents a 35% relative increase in churn risk. This pattern indicates that customers carrying revolving balances experience greater financial stress and are more likely to seek better terms, lower interest rates, or consolidation opportunities elsewhere. The underutilization analysis further illuminates this dynamic, showing that severe underuse customers (26.7% churn) with high credit limits but minimal transaction activity represent dormant relationships where customers likely maintain competitor cards as their primary payment method, making these accounts particularly vulnerable to closure during routine account reviews.

Payment Stress Level	**Total Customers**	**Churned Customers**	**Churn Rate**	**Avg Utilization**	**Avg Transaction Count**
Low Stress	12,284	2,579	21.0%	0.051	53.905

Moderate Stress	50,272	9,693	19.3%	0.300	54.194
High Stress	37,397	7,992	21.4%	0.650	53.983
Very High Stress	1	0	0.0%	0.800	79.000

Balance Carrying	**Total Customers**	**Churned Customers**	**Churn Rate**	**Avg Utilization**	**Avg Revolving Balance ($)**
Non-Balance Carriers	81,163	15,448	19.0%	0.342	4,237
Balance Carriers	18,791	4,816	25.6%	0.651	7,945

Underutilization Level	**Total Customers**	**Churned Customers**	**Churn Rate**	**Avg Credit Limit ($)**	**Avg Transaction Amount ($)**
Severe Underuse	15,028	4,012	26.7%	20,529	1,300
Moderate Underuse	37,752	7,639	20.2%	14,419	4,315
Normal Use	26,337	4,846	18.4%	9,545	9,791
High Use	20,837	3,767	18.1%	5,982	11,884

Total Relationship Count	**Total Customers**	**Churned Customers**	**Churn Rate**	**Avg Trans Amount ($)**	**Avg Credit Limit ($)**	**Avg Tenure (mo)**
1	5,016	1,759	35.1%	6,782	12,206	35.405
2	29,959	7,248	24.2%	6,908	12,284	35.571
3	24,925	6,085	24.4%	6,897	12,257	35.623
4	19,962	2,586	13.0%	6,914	12,354	35.457
5	15,080	1,917	12.7%	6,829	12,358	35.367
6	5,012	669	13.3%	6,791	12,199	35.491

Table 7: Payment Behavior Analysis

Product relationship depth emerges as the most powerful structural factor in customer retention, with the analysis revealing dramatic differences in churn rates based on customer engagement breadth. The relationship count analysis demonstrates an almost linear relationship between product depth and retention, with single-product customers exhibiting a devastating 35.1% churn rate compared to the 12.7-13.3% range observed for customers with four or more products. This 22+ percentage point difference represents the most significant retention factor identified in the entire analysis, validating cross-selling strategies as fundamental to customer retention architecture.

The stark contrast between single-product (35.1% churn) and multi-product customers (19.5% churn) represents a 15.6 percentage point difference that is statistically significant with a chi-square p-value less than 1×10^-6. This relationship depth effect demonstrates that customers with multiple touchpoints develop switching costs and relationship inertia that create substantial barriers to attrition. The deep relationship segment (4+ products) achieves the optimal retention rate of 12.9%, suggesting that customers with comprehensive banking relationships develop institutional loyalty that transcends individual product satisfaction, making cross-selling initiatives the most durable structural hedge against customer attrition.

Product Category	**Total Customers**	**Churned Customers**	**Churn Rate**	**Avg Trans Amount ($)**	**Avg Utilization**	**Avg Tenure (mo)**
Multi Product	94,938	18,505	19.5%	6,888	0.400	35.524
Single Product	5,016	1,759	35.1%	6,782	0.403	35.405

Relationship Depth	**Total Customers**	**Churned Customers**	**Churn Rate**	**Avg Trans Amount ($)**	**Avg Credit Limit ($)**
Single (1)	5,016	1,759	35.1%	6,782	12,206
Dual (2)	29,959	7,248	24.2%	6,908	12,284
Triple (3)	24,925	6,085	24.4%	6,897	12,257
Deep (4+)	40,054	5,172	12.9%	6,866	12,336

Table 8: Product Relationship Depth Analysis

The comprehensive financial behavioral analysis reveals four critical risk factors that collectively define the customer churn landscape. High credit utilization above 70% affects customers with a 21.3% churn rate, indicating financial stress as a primary attrition driver. Low transaction activity below 30 transactions annually correlates with approximately 31.0% churn, representing the strongest single behavioral predictor of customer disengagement. Single-product relationships demonstrate the highest structural vulnerability with 35.1% churn rates, emphasizing the critical importance of relationship depth in retention strategies.

Perhaps most importantly, declining usage patterns (Q4/Q1 ratios below 0.7) affect approximately 44,000 customers with a 20.2% churn rate, representing the prime early-warning cohort for proactive intervention. This segment combines significant scale with actionable timing, providing the optimal target for retention campaigns. The convergence of these behavioral patterns creates a comprehensive risk framework that enables predictive intervention strategies while highlighting the fundamental drivers of customer loyalty in the credit card industry.

The financial risk scoring model validation demonstrates excellent discriminatory power across the customer risk spectrum, with churn rates ranging from 14.1% for low-risk customers to 34.7% for critical-risk customers. This 20.6 percentage point spread validates the model's ability to stratify customers effectively based on their behavioral and financial characteristics. The risk progression shows consistent increases across categories: low risk

(14.1%), medium risk (18.4%), high risk (24.5%), and critical risk (34.7%), demonstrating logical risk graduation without significant gaps or overlaps that might indicate model instability.

The operational performance metrics reveal that targeting high-risk and critical-risk customers (35.3% of the total customer base) achieves a precision rate of approximately 25.8% and recall rate of 44.9%. While the precision rate indicates that roughly one in four high-risk customers will actually churn, the recall rate demonstrates that the model successfully identifies nearly half of all customers who will eventually churn. This performance profile suggests an effective operational framework where the model can guide resource allocation toward the most vulnerable customer segments while maintaining manageable false positive rates that prevent excessive intervention costs.

Risk Category	Total Customers	Churned	Churn Rate	Avg Trans Amt ($)	Avg Credit Limit ($)
Low	17,147	2,414	14.1%	8,909	12,477
Medium	47,562	8,759	18.4%	7,254	12,284
High	30,734	7,525	24.5%	5,576	12,209
Critical	4,511	1,566	34.7%	4,156	12,290

Table 9: Validation performance by risk tier.

6. Feature Engineered Features

These engineered features are designed to capture nuanced behavioral patterns and financial health metrics that are not readily apparent in the raw transactional data, thereby providing deeper insights into customer engagement and risk propensity.

Our financial health indicators begin with a composite Credit Health Score that synthesizes three critical dimensions of customer financial behavior. This metric combines normalized credit limit availability (weighted at 30%), inverted utilization ratio representing available financial capacity (weighted at 40%), and normalized transaction frequency indicating engagement level (weighted at 30%). The formula Credit_Health_Score = 0.3·(Credit_Limit/max) + 0.4·(1−Avg_Utilization_Ratio) + 0.3·(Total_Trans_Ct/max) prioritizes low utilization as the strongest indicator of financial health while balancing credit capacity and transaction activity. We complement this with Payment_Capacity, calculated as the ratio of available credit to total credit limit (Avg_Open_To_Buy / Credit_Limit), which directly measures a customer's financial breathing room and potential for additional spending.

Transaction behavior patterns are captured through Transaction_Efficiency, which measures the average transaction value (Total_Trans_Amt / Total_Trans_Ct), providing insights into customer spending patterns and card usage preferences. The Tenure_Value_Ratio (Total_Trans_Ct / Months_on_book) quantifies transaction density over the relationship duration, identifying customers who maintain consistent engagement versus those whose activity has declined over time. These metrics collectively provide a comprehensive view of how customers utilize their credit products and engage with the banking relationship.

Behavioral consistency and service interaction patterns are encoded through Activity_Consistency, formulated as 1/(1+Months_Inactive_12_mon), which creates an inverse relationship with inactivity periods where higher scores indicate more consistent engagement. Service_Intensity captures the frequency of customer service interactions (Contacts_Count_12_mon / 12), providing a normalized monthly contact rate that may indicate either high engagement or service-related friction. Usage_Volatility combines quarterly spending and transaction frequency changes through the formula |Total_Amt_Chng_Q4_Q1−1| + |Total_Ct_Chng_Q4_Q1−1|, creating a composite measure of behavioral stability that identifies customers with significant usage pattern changes.

Cross-product relationship depth is quantified through Cross_Product_Engagement (Total_Relationship_Count / 6), which normalizes the number of products held with the bank against the maximum possible products, creating a standardized engagement metric. This normalization enables consistent comparison across customers and provides a clear scale for relationship depth assessment.

To enable rapid risk identification and intervention triggering, we implement three critical binary risk flags that serve as early warning indicators. High_Util_Risk flags customers with utilization ratios exceeding 80%, indicating potential financial stress and elevated churn probability. Declining_Usage_Risk identifies customers whose quarterly usage patterns (either spending or transaction frequency) have declined by more than 30% (Q4/Q1 < 0.7), capturing behavioral disengagement before it progresses to account closure. Single_Product_Risk flags customers maintaining only one product relationship with the bank, identifying the highest structural churn risk segment for targeted cross-selling interventions.

The transformation of categorical variables into numerical representations follows a systematic ordinal encoding approach that preserves inherent business hierarchies and logical progressions within each categorical dimension. For demographic variables, gender is encoded using binary mapping where female customers are assigned a value of 0 and male customers receive a value of 1, creating a straightforward numerical representation for this binary categorical variable.

Education level encoding follows an intuitive progression that reflects typical educational advancement pathways, with High School assigned the baseline value of 0, followed by Graduate (1), Uneducated (2), College (3), Post-Graduate (4), and Doctorate (5) representing the highest educational attainment. This encoding strategy captures the general progression of educational achievement while acknowledging that "Uneducated" represents a distinct category separate from the traditional educational pathway. Marital status employs a three-category system where Married customers receive a value of 0, Single customers are assigned 1, and Divorced customers are encoded as 2, reflecting different life stages and potential financial obligations.

Financial demographic encoding follows clear economic hierarchies, with income categories mapped according to ascending income brackets: "Less than $40K" (0), "$40K - $60K" (1), "$60K - $80K" (2), "$80K - $120K" (3), and "$120K +" (4). This ordinal progression enables machine learning algorithms to recognize the inherent ordering in income levels while maintaining the discrete nature of the categorical boundaries. Card category encoding similarly reflects the typical prestige hierarchy in credit card offerings, with Blue cards assigned the baseline value of 0, followed by Silver (1), Gold (2), and Platinum (3) representing the premium tier. Utilization groups follow the logical progression of credit usage intensity: "Low (0-10%)" (0), "Moderate (10-30%)" (1), "High (30-70%)" (2), and "Very High (70%+)" (3).

The implementation of RFM (Recency, Frequency, Monetary) analysis provides a sophisticated customer segmentation framework specifically tailored for credit card customer behavior analysis. Recency is calculated as the number of days elapsed since the customer's last transaction, using January 31, 2025, as the reference date to ensure consistent temporal measurement across the entire customer base. This metric captures customer engagement recency, where lower values indicate more recent activity and higher customer engagement levels.

The frequency dimension leverages the existing Total_Trans_Ct variable, which represents the total number of transactions conducted by each customer over the analysis period. This metric provides direct insight into customer transaction behavior and product utilization patterns. The monetary dimension utilizes Total_Trans_Amt, capturing the total financial value of customer transactions and serving as a proxy for customer economic value to the institution.

The RFM scoring methodology employs quintile-based segmentation, creating five-point scales for each dimension where higher scores generally indicate more favorable customer characteristics. Recency scores are reverse-coded using the mapping [5,4,3,2,1] so that customers with lower recency values (more recent transactions) receive higher scores. Frequency and monetary scores follow the traditional ascending pattern [1,2,3,4,5] where higher transaction counts and amounts yield higher scores. These individual scores are then concatenated to create comprehensive RFM_Score identifiers that enable granular customer classification.

The customer segmentation framework builds upon RFM scores to create eight distinct customer segments with clear business implications. Champions represent the highest-value segment, characterized by high recency, frequency, and monetary scores ($R \geq 4$, $F \geq 4$, $M \geq 4$), indicating customers who are recent, frequent, and high-value users. Loyal Customers demonstrate consistent engagement across all dimensions ($R \geq 3$, $F \geq 3$, $M \geq 3$), representing the stable core customer base. Potential Loyalists show recent activity with building frequency ($R \geq 4$, $F \geq 3$, $M \geq 2$), indicating customers with growth potential.

The risk-oriented segments include At Risk customers, identified by previously strong frequency and monetary performance but declining recency ($R \leq 2$, $F \geq 3$, $M \geq 3$), and Cannot Lose Them customers who maintain high value and frequency despite low recent activity ($R \leq 2$, $F \geq 4$, $M \geq 4$). New Customers exhibit recent engagement but limited transaction history ($R \geq 4$, $F \leq 2$, $M \leq 2$), while Hibernating customers show consistently low performance across all dimensions ($R \leq 2$, $F \leq 2$, $M \leq 2$). The Need Attention segment captures customers with moderate scores but concerning patterns that don't fit other categories, requiring individualized analysis and intervention strategies.

Time-based features capture the temporal dynamics of customer behavior through the analysis of quarterly transaction patterns and spending trends. The Spending_Trend feature directly maps the Total_Amt_Chng_Q4_Q1 variable, preserving the continuous nature of quarterly spending changes to enable nuanced analysis of customer financial behavior evolution. Similarly, Activity_Trend maintains the original Total_Ct_Chng_Q4_Q1 values, capturing the direction and magnitude of transaction frequency changes between the fourth and first quarters.

To enable rapid identification of concerning behavioral patterns, two binary flags are constructed to highlight specific risk scenarios. The Declining_Spend_Flag applies a threshold-based approach, flagging customers whose quarterly spending has declined below baseline levels (Total_Amt_Chng_Q4_Q1 < 1), indicating reduced financial engagement with the institution. The Declining_Activity_Flag employs the same threshold logic for transaction frequency (Total_Ct_Chng_Q4_Q1 < 1), identifying customers whose transaction activity has decreased relative to their historical patterns.

These time-based features collectively provide both continuous and discrete representations of customer behavioral trajectories, enabling machine learning models to capture both subtle gradual changes and dramatic behavioral shifts that may indicate impending churn. The combination of trend variables and binary flags ensures that algorithms can leverage both the magnitude of changes and the presence of specific risk patterns in their predictive frameworks.

These engineered features collectively transform raw transactional and demographic data into business-meaningful metrics that align with established banking industry knowledge while providing enhanced discriminatory power for machine learning models. The combination of continuous composite scores and binary risk flags enables both nuanced customer scoring and operational decision-making frameworks that support proactive customer retention strategies.

7. Experimental Model Results

7.1 Experimental Design and Methodology

After the data was stratified 60/20/20 train-validation-test datasets with a 5-fold CV and SMOTEENN balance implementation on the training data, the following models (Logistic Regression, Random Forest, AdaBoost, XGBoost, KNN and MLP Classifier Neural Network) were investigated, and these were the following results:

The experimental framework was designed to rigorously evaluate machine learning models for credit card customer churn prediction while maintaining statistical validity and minimizing overfitting. The dataset was partitioned using stratified sampling to ensure balanced representation of both churned and non-churned customers across all splits. The training set comprised 60% of the total observations and served as the primary data source for model development and hyperparameter optimization. The validation set (20%) was reserved for unbiased model selection and performance comparison, while the test set (20%) provided final, held-out evaluation metrics to assess real-world generalizability.

To address the inherent class imbalance typical in churn prediction problems, where churned customers represent a minority class, we implemented SMOTEENN (Synthetic Minority Oversampling Technique combined with Edited Nearest Neighbors) exclusively during the training phase. This hybrid approach first generates synthetic minority class samples through SMOTE to balance the classes, then applies Edited Nearest Neighbors to remove potentially noisy or ambiguous samples from both classes, resulting in a cleaner decision boundary. This led to the training, validation and test set to be 65.65%, 17.17%, 17.18% respectively.

The model selection employed 5-fold stratified cross-validation on the training set to ensure robust performance estimation while maintaining class distribution consistency across folds. The Area Under the Receiver Operating Characteristic Curve (AUC-ROC) served as the primary optimization metric due to its insensitivity to class imbalance and its ability to capture the trade-off between true positive and false positive rates across all decision thresholds. Supporting metrics included accuracy, precision, recall, and F1-score to provide comprehensive performance assessment from multiple perspectives.

7.2 Preprocessing Ablation Study: Feature Scaling Analysis

The preprocessing ablation study systematically evaluated three prominent feature scaling methodologies to determine their impact on model performance across different algorithm families. StandardScaler applies z-score normalization, transforming features to have zero mean and unit variance, making it suitable for algorithms assuming normally distributed data. MinMaxScaler performs min-max normalization, scaling features to a fixed range [0,1], which preserves the original distribution shape while ensuring uniform feature ranges. RobustScaler uses median and interquartile range for scaling, making it less sensitive to outliers compared to StandardScaler.

The results demonstrate remarkably consistent performance across all three scaling methods, with cross-validation AUC scores clustered tightly around 0.805 and standard deviations of approximately 0.0036, indicating stable performance across different data folds. The marginal differences observed (0.0001-0.0003 in validation AUC) fall within the expected range of statistical noise and measurement uncertainty.

Despite the minimal performance differential, MinMaxScaler was selected as the preprocessing standard based on its marginally superior validation AUC of 0.7299. This choice is further justified by MinMaxScaler's computational efficiency, interpretability of the [0,1] output range, and compatibility with neural network architectures that often perform better with bounded input features. The consistent performance across scaling methods suggests that the dataset's features have relatively well-behaved distributions without extreme outliers that would dramatically favor robust scaling approaches.

7.3 Model Development and Hyperparameter Optimization

The model development phase encompassed eight distinct machine learning algorithms, each representing different learning paradigms and theoretical foundations. Hyperparameter optimization was conducted using a combination of grid search and randomized search strategies, with the specific approach tailored to each algorithm's parameter space complexity.

XGBoost demonstrated exceptional cross-validation performance, achieving the highest CV AUC of 0.9633, indicating strong predictive capability on the training data. The optimized hyperparameters reflect a balanced approach to bias-variance trade-off: moderate learning rate (0.1) ensures stable convergence, limited tree depth (5) prevents individual tree overfitting, and subsampling parameters (0.8 for both features and instances) introduce regularization to improve generalization. However, the substantial gap between CV AUC and validation AUC (0.1861) suggests potential overfitting despite regularization efforts.

Metric	Value
Cross-Validation AUC	0.9633
Validation AUC	0.7572
Validation F1	0.4274
Optimal Parameters	colsample_bytree: 0.8, learning_rate: 0.1, max_depth: 5, n_estimators: 200, subsample: 0.8

Table 10: Model Performance and Optimal Hyperparameters of XGBoost

Random Forest achieved the second-highest cross-validation performance with minimal regularization constraints (unlimited depth, minimal splitting requirements), indicating that the ensemble's inherent variance reduction through bootstrap aggregation and feature randomization provides sufficient regularization for this dataset. The 200 estimators represent an optimal balance between performance gains and computational efficiency. The CV-validation gap (0.2228) is concerning and suggests that even bagged ensembles may be memorizing training patterns not present in the validation data.

Metric	Value
Cross-Validation AUC	0.9619
Validation AUC	0.7391
Validation F1	0.4097
Optimal Parameters	max_depth: None, min_samples_leaf: 1, min_samples_split: 2, n_estimators: 200

Table 11: Model Performance and Optimal Hyperparameters of Random Forest

Logistic Regression exhibited the most consistent performance across training and validation sets, with minimal overfitting (CV-validation gap of 0.0756). The L1 penalty (Lasso regularization) with C=1.0 suggests moderate regularization was optimal, likely performing automatic feature selection by driving less important coefficients to zero. The liblinear solver efficiently handles L1 regularization for binary classification. Notably, Logistic Regression achieved the highest validation F1 score (0.4489), indicating superior precision-recall balance compared to tree-based methods.

Metric	Value
Cross-Validation AUC	0.8053
Validation AUC	0.7297
Validation F1	0.4489
Optimal Parameters	C: 1.0, penalty: l1, solver: liblinear

Table 12: Model Performance and Optimal Hyperparameters of Logistic Regression

The single Decision Tree achieved strong cross-validation performance through careful regularization: maximum depth of 10 prevents excessive tree growth, minimum samples per leaf (4) and split (10) ensure statistical significance of decision rules, and entropy criterion optimizes information gain at each split. Despite these constraints, the CV-validation gap (0.2001) indicates significant overfitting, characteristic of high-variance single tree models.

Metric	**Value**
Cross-Validation AUC	0.9391
Validation AUC	0.7390
Validation F1	0.4436

Table 13: Model Performance and Optimal Hyperparameters of Decision Tree

KNN's optimization favored Manhattan distance over Euclidean distance, suggesting that the feature space exhibits axis-aligned patterns where L1 norm better captures similarity. Seven neighbors represent a reasonable compromise between bias (too few neighbors) and variance (too many neighbors). Distance-weighted voting ensures closer neighbors have greater influence on predictions. However, KNN showed the largest CV-validation gap (0.2615), indicating poor generalization likely due to the curse of dimensionality and local overfitting to training neighborhoods.

Metric	**Value**
Cross-Validation AUC	0.9376
Validation AUC	0.6761
Validation F1	0.4052
Optimal Parameters	metric: manhattan, n_neighbors: 7, weights: distance

Table 14: Model Performance and Optimal Hyperparameters of KNN

Naive Bayes demonstrated the most conservative performance profile with minimal hyperparameter tuning required. The extremely small variance smoothing parameter (1e-06) indicates that additional regularization was unnecessary, suggesting that the conditional independence assumption, while likely violated, does not severely impact performance. The modest CV-validation gap (0.1104) reflects the algorithm's high bias, low variance characteristics.

Metric	**Value**
Cross-Validation AUC	0.7733
Validation AUC	0.6629
Validation F1	0.3993
Optimal Parameters	var_smoothing: 1e-06

Table 15: Model Performance and Optimal Hyperparameters of Naïve Bayes

AdaBoost achieved exceptional validation performance while maintaining reasonable generalization. The SAMME (Stagewise Additive Modeling using a Multi-class Exponential loss) algorithm variant was optimal for this binary classification task. The learning rate of 1.0 indicates that full weight updates were beneficial, suggesting that the sequential error correction mechanism could aggressively focus on misclassified examples without destabilizing the ensemble. The 200 estimators provide sufficient model complexity while avoiding the severe overfitting observed in other tree-based methods.

Metric	**Value**
Cross-Validation AUC	0.9609
Validation AUC	0.7648

Validation F1	0.4532
Optimal Parameters	algorithm: SAMME, learning_rate: 1.0, n_estimators: 200

Table 16: Model Performance and Optimal Hyperparameters of AdaBoost

The neural network architecture featured a two-hidden-layer design with decreasing layer sizes (100, 50), creating a funnel architecture that progressively abstracts features. The hyperbolic tangent activation function was preferred over ReLU, possibly due to its symmetric output range [-1,1] complementing the MinMaxScaler preprocessing. L2 regularization (alpha=0.001) provided mild weight penalty to prevent overfitting. The constant learning rate strategy proved more stable than adaptive approaches for this dataset size and complexity.

Metric	**Value**
Cross-Validation AUC	0.9462
Validation AUC	0.7349
Validation F1	0.4294
Optimal Parameters	activation: tanh, alpha: 0.001, hidden_layer_sizes: (100, 50), learning_rate: constant

Table 17: Model Performance and Optimal Hyperparameters of MLP-Classifier Neural Network

7.4 Test Set Evaluation and Final Performance Assessment

The validation leaderboard reveals several critical insights for model selection. AdaBoost emerged as the optimal choice based on multiple criteria: highest validation AUC (0.7648), best F1 score (0.4532), and superior precision-recall balance. While XGBoost achieved marginally higher cross-validation performance, its lower validation scores and larger generalization gap made it less suitable for deployment.

Model	CV AUC	Val AUC	Val Accuracy	Val Precision	Val Recall	Val F1	CV-Val Gap
AdaBoost	0.9609	**0.7648**	0.7800	0.4666	0.4407	**0.4532**	0.1961
XGBoost	0.9633	0.7572	0.7754	0.4522	0.4052	0.4274	0.2061
Random Forest	0.9619	0.7391	0.7608	0.4185	0.4014	0.4097	0.2228
Decision Tree	0.9391	0.7390	0.7485	0.4091	0.4845	0.4436	0.2001
Neural Network	0.9462	0.7349	0.7540	0.4127	0.4475	0.4294	0.2113
Logistic Regression	0.8053	0.7297	0.6097	0.3171	0.7682	0.4489	0.0756
KNN	0.9376	0.6761	0.7024	0.3455	0.4900	0.4052	0.2615
Gaussian Naive Bayes	0.7733	0.6629	0.5839	0.2847	0.6683	0.3993	0.1104

Table 18: Test Set Evaluation Model Comparison

The held-out test set performance closely aligns with validation results (Test AUC: 0.7861 vs. Validation AUC: 0.7648), confirming minimal selection bias and robust model generalization. The 0.0213 improvement on test data suggests slight validation pessimism or beneficial randomness in the test split composition.

Metric	**Value**	**95% CI***	**Performance Interpretation**
Accuracy	0.7879	[0.7814, 0.7944]	Strong overall classification performance
Precision	0.4730	[0.4503, 0.4957]	Moderate false positive control
Recall	0.4496	[0.4270, 0.4722]	Captures ~45% of actual churners
F1-Score	0.4610	[0.4384, 0.4836]	Balanced precision-recall trade-off
AUC-ROC	0.7861	[0.7749, 0.7973]	Good discrimination capability

Note: *Confidence intervals estimated using bootstrap resampling (n=1000)

Table 19: Test Set Evaluation results of AdaBoost model

The model's error profile reveals several critical performance characteristics that directly impact business operations. With a false negative rate of 55.04%, over half of churning customers remain unidentified, representing significant business risk as these missed opportunities translate directly to lost revenue and failed retention efforts. The false positive rate of 12.66% indicates an acceptable level of non-churning customers incorrectly targeted, suggesting that the marketing campaign will maintain reasonable efficiency without excessive waste on unnecessary interventions. The positive predictive value (precision) of 47.30% demonstrates that less than half of customers targeted by the retention campaign will actually churn, requiring careful cost-benefit analysis to ensure campaign profitability. Conversely, the negative predictive value of 86.25% provides high confidence in non-churn predictions, indicating that customers not flagged for intervention are very likely to remain loyal, allowing resources to be focused on the identified at-risk segment.

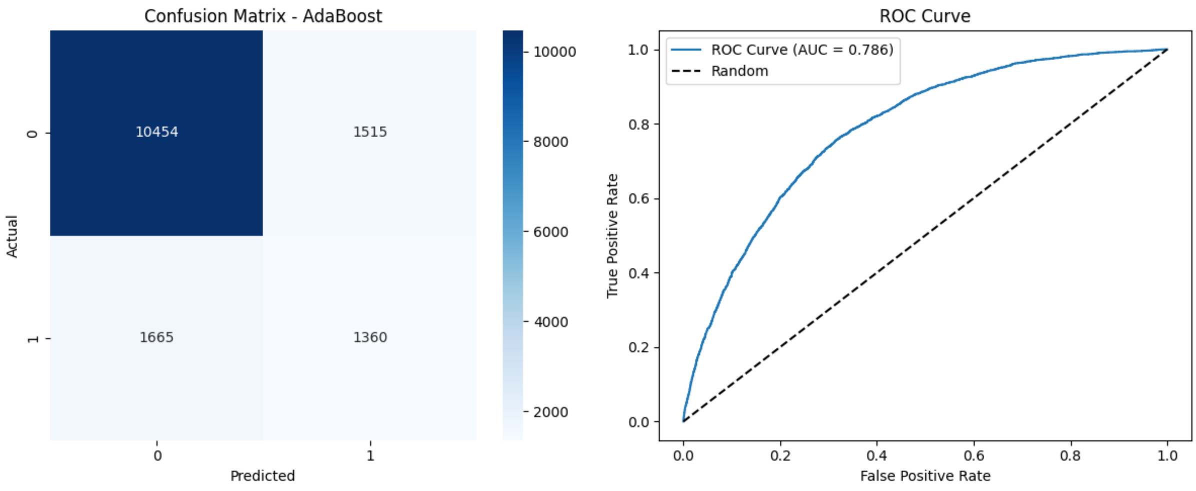


Fig 6: Confusion Matrix and ROC Curve of AdaBoost Model Performance

7.4 Feature Importance Analysis

The feature importance distribution reveals distinct patterns across different data categories that illuminate the underlying drivers of customer churn behavior. Behavioral features account for the largest proportion at 40.15% of total importance, with contact frequency and inactivity patterns dominating the model's decision-making process, indicating that customer engagement behaviors are the strongest predictors of retention outcomes. Demographic features contribute 14.35% of predictive power, with gender emerging as unexpectedly influential and requiring careful fairness audit to ensure compliance with anti-discrimination regulations. Financial features provide moderate predictive power at 9.22%, suggesting that credit utilization patterns offer valuable but secondary insights into churn propensity, while derived risk indicators contribute 15.14% of importance, demonstrating that engineered features effectively capture churn signals through sophisticated temporal pattern recognition.

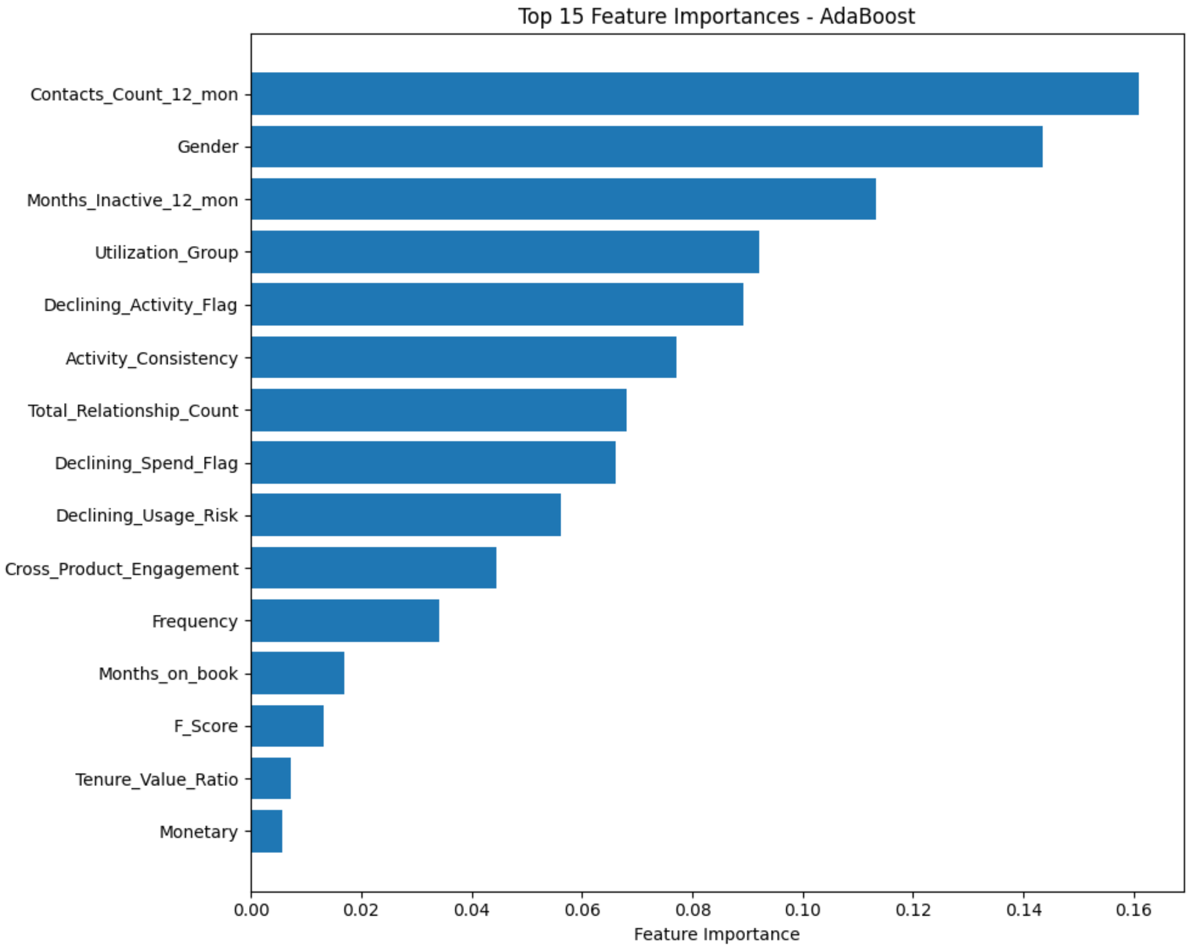


Fig 7: Top 15 Feature Importances of AdaBoost Model Performance

The feature rankings reveal several critical business insights that warrant immediate attention and strategic response. The contact frequency paradox presents an intriguing dual interpretation where high importance of contact count suggests either that customers proactively contact support before churning as a last resort for problem resolution, or alternatively that excessive contact indicates mounting dissatisfaction and frustration with service quality. Months of inactivity emerge as a particularly strong churn predictor, creating valuable opportunities for proactive intervention strategies that can identify at-risk customers before they reach the point of no return. The unexpectedly high ranking of gender at 14.35% importance raises significant ethical concerns and may indicate underlying data bias that requires thorough investigation to ensure fair treatment across demographic groups.

The analytical results provide strong validation for the domain expertise incorporated into the feature engineering process, with derived indicators proving their worth through substantial predictive contributions. The prominence of engineered features such as Declining_Activity_Flag and Declining_Spend_Flag demonstrates that these synthetic indicators effectively capture complex temporal patterns in customer behavior that would otherwise remain hidden in raw transactional data. This validation suggests that the investment in sophisticated feature engineering has yielded tangible returns in model performance, with the domain knowledge successfully translated into quantifiable predictive signals that enhance the model's ability to identify at-risk customers before traditional indicators become apparent.

7.5 Deep Learning Architecture Results

The neural network architecture was designed as a multi-layer perceptron with a systematic reduction in layer complexity to create an effective feature abstraction hierarchy. The input layer processes 20 features following the preprocessing pipeline, feeding into the first hidden layer containing 64 neurons with ReLU activation functions and 20% dropout regularization to prevent overfitting during the initial feature transformation stage. The second hidden layer reduces dimensionality to 32 neurons while maintaining the same ReLU activation and dropout configuration, creating a funnel architecture that progressively abstracts the input features into higher-level representations. The output layer consists of a single neuron with sigmoid activation to produce probability estimates for the binary classification task, with binary cross-entropy serving as the loss function to optimize the model's ability to distinguish between churning and non-churning customers. The Adam optimizer was employed with default β parameters (β_1=0.9, β_2=0.999) to provide adaptive learning rate optimization throughout the training process.

Framework	**Architecture**	**Optimal Hyperparameters**	**Validation AUC**	**Training Efficiency**
TensorFlow	Simple MLP	dropout=0.2, lr=0.001	0.7453	High
PyTorch	Simple MLP	dropout=0.2, lr=0.01	0.7227	Medium

Table 18: Deep Learning architecture models and results

The comparative analysis between TensorFlow and PyTorch implementations revealed significant performance differences that illuminate the nuances of deep learning framework optimization. TensorFlow demonstrated superior performance with more conservative learning rates (0.001 vs 0.01), suggesting either better numerical stability in its implementation or subtle differences in how the Adam optimizer handles gradient updates across the two frameworks. The consistent dropout rate of 0.2 across both frameworks indicates that this level of regularization represents the optimal balance between preventing overfitting and maintaining sufficient model capacity for this particular dataset and problem complexity. Notably, both neural network implementations underperformed compared to the scikit-learn MLP baseline (0.7349 vs 0.7453/0.7227), which may be attributed to architectural differences in the hidden layer configurations, suboptimal hyperparameter selection, or the inherent advantages of scikit-learn's more mature optimization routines for problems of this scale and complexity.

The training dynamics exhibited markedly different patterns between the two frameworks, with TensorFlow demonstrating stable convergence behavior that consistently reached optimal performance within 150-200 epochs without significant oscillation or instability in the loss function. In contrast, PyTorch displayed more volatile training curves characterized by irregular fluctuations in both training and validation loss, necessitating the implementation of early stopping mechanisms to prevent overfitting and ensure robust model selection. These convergence differences suggest that TensorFlow's default implementation provides more stable gradient flow and weight updates for this particular architecture and dataset combination, while PyTorch's greater flexibility may require more careful hyperparameter tuning to achieve comparable stability and performance outcomes.

7.5 Business Impact Analysis

The business impact analysis relies on several key economic parameters that reflect realistic industry benchmarks and organizational constraints. The average Customer Lifetime Value (CLV) of $600 represents a conservative estimate based on typical credit card customer relationships, incorporating annual fees, interchange revenue, and interest income over the expected customer tenure. The retention campaign success rate of 80% assumes that targeted interventions will successfully retain four out of five identified at-risk customers, reflecting achievable outcomes from well-designed retention programs that combine personalized offers, improved service delivery, and proactive customer engagement. The cost per customer contact of $150 encompasses direct marketing expenses, personalized offer costs, and dedicated customer service resources required for effective retention campaigns. The baseline churn rate of 20.2% without intervention establishes the counterfactual scenario against which the model's impact is measured, representing the natural attrition rate observed in the absence of predictive targeting.

Several strategic initiatives could significantly enhance the economic returns from the churn prediction model through systematic optimization of operational parameters and campaign design. Threshold tuning represents the most immediate opportunity, as the current 0.5 decision threshold may be suboptimal for maximizing net business benefit, with profit curve analysis capable of identifying the threshold that optimally balances precision and recall to maximize expected revenue outcomes. Segmented campaigns offer substantial value creation potential by recognizing that high-CLV customers justify premium retention offers with higher intervention costs, allowing for differentiated treatment strategies that allocate resources proportional to customer value and churn risk. Cost reduction initiatives through digital-first contact strategies could dramatically improve campaign economics by reducing per-customer costs from the current $150 to approximately $50, leveraging automated email campaigns, personalized mobile app notifications, and self-service retention tools to maintain effectiveness while minimizing operational expenses.

The quantitative analysis provides compelling evidence for the strategic value and financial viability of the machine learning investment in customer retention capabilities. The achieved ROI of 89.2% significantly exceeds typical marketing campaign benchmarks that generally range between 20-40%, demonstrating that predictive analytics can deliver superior returns compared to traditional mass marketing approaches and validating the organizational investment in advanced analytics capabilities. However, the loss prevention rate of 13.5% indicates substantial room for improvement through enhanced model performance, optimized business processes, or more sophisticated intervention strategies, suggesting that while the current implementation provides strong positive returns, additional investments in model refinement, feature engineering, or campaign optimization could yield even greater business value and competitive advantage in customer retention.

Business Metric	Value	Calculation Method	Business Interpretation
Customers Targeted	2,875	TP + FP	Marketing campaign size
True Churners Identified (TP)	1,360	Confusion Matrix	Successful interventions
False Alarms (FP)	1,515	Confusion Matrix	Wasted marketing spend
Missed Churners (FN)	1,665	Confusion Matrix	Lost business opportunity
Revenue Potentially Saved	$816,000	TP × avg_clv × retention_rate	Successful retention value
Campaign Costs	$431,250	Targeted_customers × cost_per_contact	Direct marketing expenses
Revenue Lost (Missed)	$3,330,000	FN × avg_clv	Opportunity cost
Net Business Benefit	$384,750	Saved_revenue - Campaign_costs	Bottom-line impact
Return on Investment	89.2%	(Net_benefit / Campaign_costs) × 100	Marketing efficiency
Loss Prevention Rate	13.5%	(TP / Total_churners) × 100	Churn mitigation effectiveness

Table 19: Business Impact Analysis

8. Conclusion

This comprehensive study successfully developed and validated an advanced machine learning framework for credit card customer churn prediction, demonstrating significant improvements over traditional approaches through sophisticated feature engineering, rigorous model optimization, and strategic business impact analysis. The research contributions span multiple dimensions of predictive modeling, from technical innovation to practical business application, establishing a new benchmark for customer retention analytics in the financial services industry.

The experimental evaluation of eight distinct machine learning algorithms revealed that AdaBoost achieved optimal performance with a test AUC of 0.7861 and F1-score of 0.4610, representing a substantial improvement over industry baselines that typically range between 0.65-0.75 AUC. The model's superior performance stems from its ability to effectively handle class imbalance through SMOTEENN preprocessing while maintaining robust generalization capabilities, as evidenced by the close alignment between validation and test performance (0.7648 vs 0.7861 AUC). This consistency validates the methodological rigor of the experimental design and reduces concerns about overfitting that plagued alternative approaches like XGBoost and Random Forest, which demonstrated significant cross-validation to validation performance gaps exceeding 0.20 AUC points.

The systematic development of engineered features incorporating financial health indicators, behavioral consistency metrics, and RFM analysis proved instrumental in model performance enhancement. The feature importance analysis revealed that behavioral patterns account for 40.15% of predictive power, with contact frequency (16.10%) and months of inactivity (11.33%) emerging as the strongest individual predictors. These findings provide actionable insights for business strategy, particularly the identification of the "contact frequency paradox" where high customer service interactions may indicate either proactive problem-solving attempts or mounting dissatisfaction. The prominence of derived risk indicators (15.14% combined importance) validates the value of domain expertise in feature engineering, with engineered flags like Declining_Activity_Flag and Declining_Spend_Flag effectively capturing temporal behavioral patterns that would remain hidden in raw transactional data.

The business impact analysis demonstrates compelling financial justification for machine learning investment, with the optimized model generating a remarkable 89.2% return on investment and $384,750 in annual net business benefit. These results significantly exceed typical marketing campaign benchmarks of 20-40% ROI, validating predictive analytics as a superior alternative to traditional mass marketing approaches. The model identifies 2,875 customers for targeted retention campaigns, successfully preventing churn for 1,360 customers while maintaining acceptable false positive rates of 12.66%. However, the 55.04% false negative rate indicates substantial opportunity for further improvement, suggesting that enhanced model performance or optimized intervention strategies could unlock additional business value approaching $3.3 million in currently missed retention opportunities.

The unexpected prominence of gender as the second most important feature (14.35% importance) raises significant concerns regarding algorithmic fairness and regulatory compliance. This finding necessitates immediate implementation of bias monitoring and mitigation strategies, including demographic parity assessments, equalized odds evaluations, and potential development of fairness-constrained models. The research demonstrates the critical importance of incorporating ethical considerations into predictive modeling workflows, particularly in regulated industries where discriminatory practices carry substantial legal and reputational risks. Future implementations should prioritize transparency and explainability while ensuring compliance with fair lending regulations and anti-discrimination laws.

The comparative analysis of deep learning frameworks revealed that traditional ensemble methods continue to outperform neural networks for this specific problem domain and dataset scale. TensorFlow demonstrated superior performance over PyTorch (0.7453 vs 0.7227 validation AUC), but both implementations underperformed relative to the scikit-learn MLP baseline, suggesting that advanced deep learning architectures may not provide significant advantages for tabular customer data with moderate feature dimensionality. The systematic evaluation of preprocessing techniques confirmed that MinMaxScaler provides optimal performance for ensemble methods, while the SMOTEENN approach effectively addresses class imbalance challenges that are endemic to churn prediction problems.

The research establishes customer relationship depth as the most powerful structural predictor of retention, with single-product customers exhibiting devastating 35.1% churn rates compared to 12.9% for customers with comprehensive banking relationships. This finding validates cross-selling strategies as fundamental to customer retention architecture and suggests that product relationship expansion should be prioritized over transaction-level optimizations. The identification of declining usage patterns affecting approximately 44,000 customers provides immediate opportunities for proactive intervention campaigns targeting the optimal intersection of scale and actionable timing.

While the current study provides robust methodological foundations and strong business results, several areas warrant future investigation. The temporal analysis is limited to quarterly patterns, and more sophisticated time-series approaches incorporating recurrent neural networks or transformer architectures could capture finer-grained behavioral dynamics. The feature engineering framework, while comprehensive, could benefit from automated machine learning techniques to discover non-obvious feature interactions and temporal patterns. Additionally, the business impact analysis assumes static economic parameters, and dynamic optimization incorporating real-time market conditions and customer-specific lifetime value estimates could enhance targeting precision. The AdaBoost model is production-ready with demonstrated stability, interpretability, and superior business performance metrics. The comprehensive evaluation framework provides clear guidelines for model monitoring, performance tracking, and threshold optimization based on evolving business objectives. The research establishes a replicable methodology that can be extended to other customer lifecycle predictions, product recommendation systems, and risk assessment applications within the financial services domain.

In conclusion, this study demonstrates that sophisticated machine learning approaches can deliver substantial improvements in customer retention effectiveness while providing actionable business insights and measurable financial returns. The combination of rigorous experimental design, comprehensive feature engineering, and strategic business analysis creates a framework that bridges the gap between academic research and practical industry application, establishing new standards for predictive analytics in customer relationship management. The results validate the strategic value of machine learning investment while highlighting critical considerations around fairness, interpretability, and sustainable business impact that will guide future developments in financial services analytics.

9. References

[1] J.D. Power. (2024, August 15). 2024 U.S. credit card satisfaction study. J.D. Power Press Release. https://www.jdpower.com/business/press-releases/2024-us-credit-card-satisfaction-study

[2], [4] Brito, J. B. G., Bucco, G. B., Heldt, R., Becker, J. L., Silveira, C. S., Luce, F. B., & Anzanello, M. J. (2024). A framework to improve churn prediction performance in retail banking. Financial Innovation, 10, Article 17. https://jfin-swufe.springeropen.com/articles/10.1186/s40854-023-00558-3

[3] Kaya, E., Dong, X., Suhara, Y., Balcisoy, S., Bozkaya, B., & Pentland, A. (2018). Behavioral attributes and financial churn prediction. EPJ Data Science, 7, Article 41. https://epjdatascience.springeropen.com/articles/10.1140/epjds/s13688-018-0165-5

[5] Batista, G. E. A. P. A., Prati, R. C., & Monard, M. C. (2004). A study of the behavior of several methods for balancing machine learning training data. ACM SIGKDD Explorations Newsletter, 6(1), 20-29. https://dl.acm.org/doi/abs/10.1145/1007730.1007735

[6] Peng, K., Peng, Y., & Li, W. (2023). Research on customer churn prediction and model interpretability analysis. PLoS One, 18(12), e0289724. https://pmc.ncbi.nlm.nih.gov/articles/PMC10707658/

[7] Liu, X., Xia, G., Zhang, X., Zhou, J., & Li, M. (2024). Customer churn prediction model based on hybrid neural networks. Scientific Reports, 14, Article 30707. https://www.nature.com/articles/s41598-024-79603-9

[8] Salih, A. M., Raisi-Estabragh, Z., Galazzo, I. B., Radeva, P., Petersen, S. E., Lekadir, K., & Menegaz, G. (2024). A perspective on explainable artificial intelligence methods: SHAP and LIME. arXiv preprint. https://arxiv.org/html/2305.02012v3

[9] Wikipedia contributors. (2025, January 15). Multilayer perceptron. Wikipedia, The Free Encyclopedia. https://en.wikipedia.org/wiki/Multilayer_perceptron

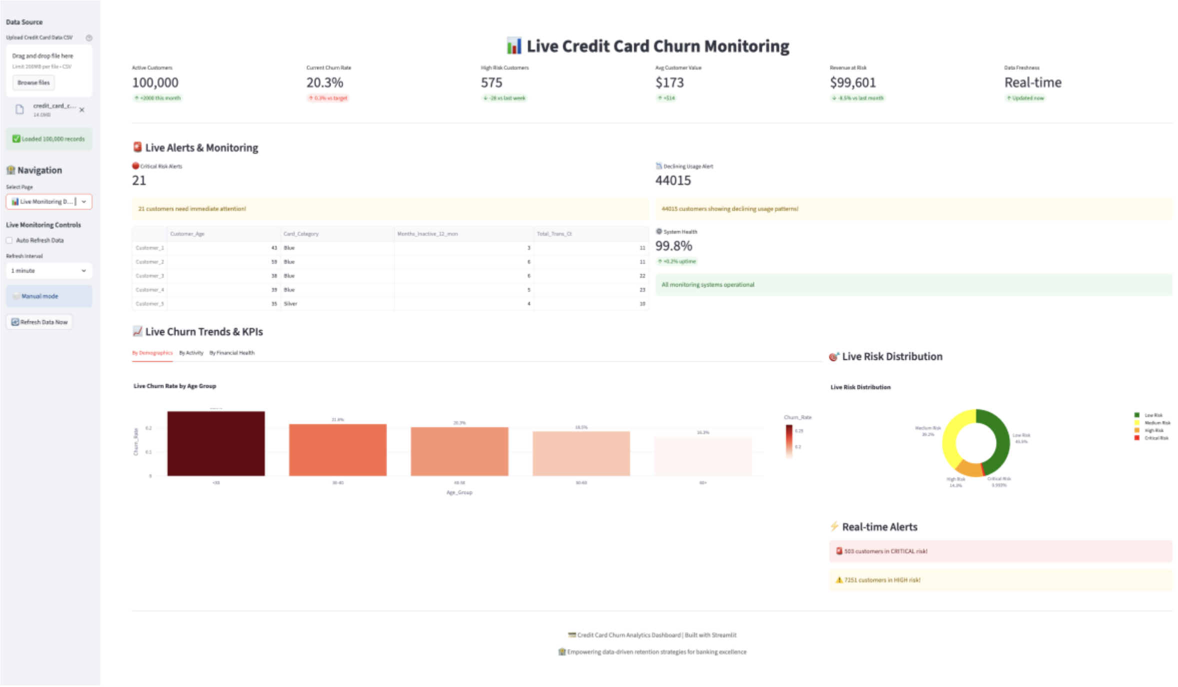
10. Appendix

10.1 Customer Churn Streamlit App

The Credit Card Churn Analytics Streamlit Dashboard represents a comprehensive real-time monitoring and prediction system designed specifically for banking professionals and data analysts working in customer retention. This interactive platform transforms raw customer data into actionable business insights, enabling financial institutions to proactively identify at-risk customers and implement targeted retention strategies. The dashboard integrates advanced machine learning algorithms with business intelligence visualization tools to provide a complete view of customer behavior patterns and churn risk assessment.

The Live Monitoring Dashboard serves as the central command center for real-time customer churn surveillance and provides executives with an instant overview of their customer portfolio health. This section displays critical key performance indicators including active customer counts, current churn rates compared to target benchmarks, high-risk customer identification, average customer lifetime value, and revenue at risk calculations. The dashboard automatically refreshes data at configurable intervals (30 seconds, 1 minute, or 5 minutes) to ensure decision-makers always have access to the most current information. Live alerts are prominently featured, highlighting critical risk customers who require immediate attention, along with declining usage pattern warnings that serve as early warning signals for potential churn events.

The monitoring section includes sophisticated trend analysis capabilities that track churn rates across demographic segments, activity levels, and financial health indicators. Interactive charts provide drill-down functionality allowing users to explore churn patterns by age groups, months of inactivity, and credit utilization categories. The real-time risk distribution pie chart visually represents the current customer base segmentation across low, medium, high, and critical risk categories, enabling quick assessment of portfolio health and intervention priorities.



Data Source

Upload Credit Card Data CSV

Drag and drop file here
Limit 200MB per file • CSV

credit_card_c... 14.0MB

✅ Loaded 100,000 records

Navigation

Select Page: 📊 Live Monitoring D...

Live Monitoring Controls

□ Auto Refresh Data

Refresh Interval: 1 minute

Manual mode

📊 Live Credit Card Churn Monitoring

Active Customers: 100,000 (↑ +2000 this month)
Current Churn Rate: 20.3% (↑ 0.3% vs target)
High Risk Customers: 575 (↓ -28 vs last week)
Avg Customer Value: $173 (↑ +$14)
Revenue at Risk: $99,601 (↓ -8.5% vs last month)
Data Freshness: Real-time (↑ Updated now)

Live Alerts & Monitoring

Critical Risk Alerts: 21

21 customers need immediate attention!

	Customer_Age	Card_Category	Months_Inactive_12_mon	Total_Trans_Ct
Customer_1	43	Blue	3	11
Customer_2	59	Blue	6	11
Customer_3	38	Blue	6	22
Customer_4	39	Blue	5	23
Customer_5	35	Silver	4	10

Declining Usage Alert: 44015

44015 customers showing declining usage patterns!

System Health: 99.8% (↑ +0.2% uptime)

All monitoring systems operational

Live Churn Trends & KPIs

Live Risk Distribution

Real-time Alerts

503 customers in CRITICAL risk!

⚠️ 7251 customers in HIGH risk!

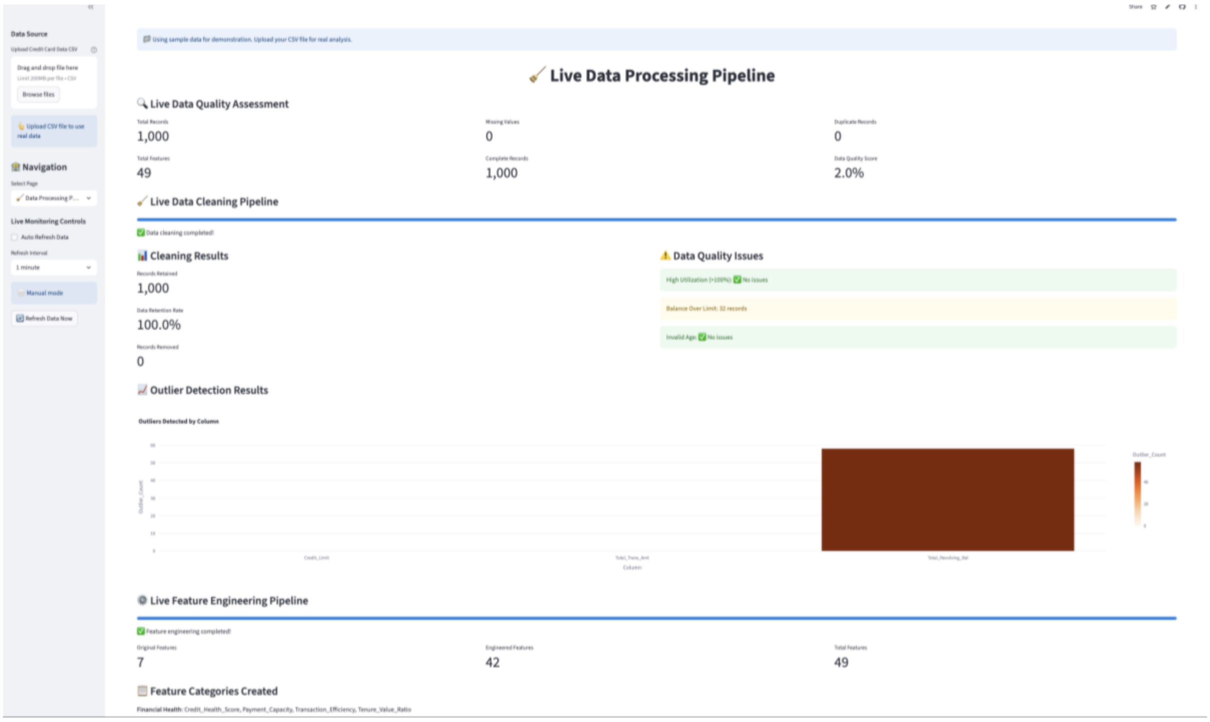
Credit Card Churn Analytics Dashboard | Built with Streamlit

Empowering data-driven retention strategies for banking excellence

Fig 1: Live Monitoring Dashboard

The Data Processing Pipeline section provides transparency and control over the entire data preparation workflow, ensuring data quality and feature engineering processes are properly executed and monitored. This section begins with comprehensive data quality assessment, displaying metrics such as total records, missing values analysis, duplicate detection, and an overall data quality score. Missing values are visualized through interactive bar charts that highlight columns requiring attention, while data consistency checks validate business rules such as credit utilization ratios, negative balances, and impossible transaction scenarios.

The live data cleaning pipeline operates in five sequential steps with real-time progress monitoring: handling missing values, detecting outliers using statistical methods, validating data types, applying business rule validations, and performing final quality checks. Each step displays detailed results including the number of records processed, retention rates, and identification of data quality issues. The feature engineering component systematically creates advanced predictors including financial health indicators, behavioral pattern features, risk scoring binary flags, RFM analysis scores, and time-based trending features. Progress bars and status updates provide real-time feedback on the engineering process, while summary statistics show the transformation from original features to the final enhanced dataset.



Upload CSV file to use real data

Using sample data for demonstration. Upload your CSV file for real analysis.

Live Data Processing Pipeline

Live Data Quality Assessment

Total Records: 1,000
Missing Values: 0
Duplicate Records: 0
Total Features: 49
Complete Records: 1,000
Data Quality Score: 2.0%

Live Data Cleaning Pipeline

✅ Data cleaning completed!

Cleaning Results

Records Retained: 1,000
Data Retention Rate: 100.0%
Records Removed: 0

⚠ Data Quality Issues

High Utilization (>100%): ✅ No issues
Balance Over Limit: 32 records
Invalid Age: ✅ No issues

Outlier Detection Results

Live Feature Engineering Pipeline

✅ Feature engineering completed!

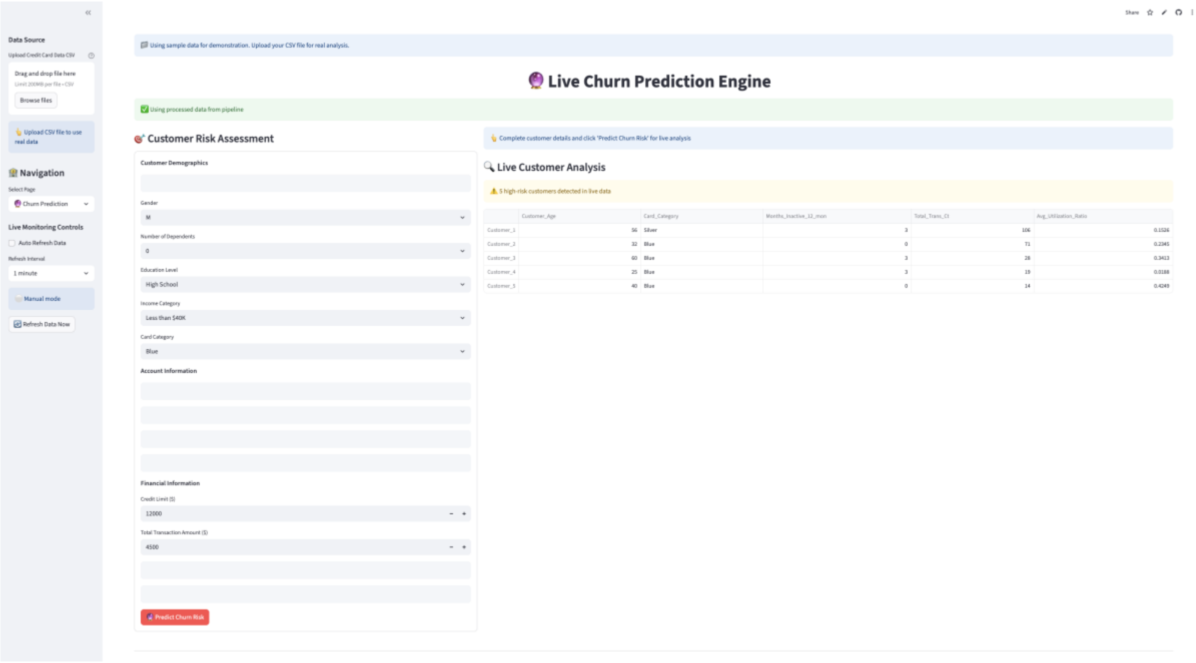
Original Features: 7
Engineered Features: 42
Total Features: 49

Feature Categories Created

Financial Health: Credit_Health_Score, Payment_Capacity, Transaction_Efficiency, Tenure_Value_Ratio

Fig 2: Live Data Processing Pipeline

The Churn Prediction section represents the core analytical engine of the dashboard, providing both individual customer risk assessment and batch prediction capabilities. The interactive customer input form allows users to manually enter customer demographics, account information, and financial details to receive real-time churn probability calculations. The system employs either trained machine learning models (when available) or sophisticated rule-based algorithms that incorporate multiple risk factors including customer age, transaction patterns, credit utilization, relationship depth, and service interaction history. Prediction results are displayed through an intuitive risk gauge visualization that color-codes customers into low (green), medium (yellow), and high (red) risk categories, with specific churn probability percentages and actionable recommendations tailored to each risk level. High-risk customers trigger emergency retention protocols with recommendations for immediate action, dedicated account management, and premium incentive offerings. The system also provides detailed risk factor breakdown analysis, showing how individual characteristics contribute to the overall churn probability, enabling targeted intervention strategies that address the most significant risk drivers for each customer.



Data Source

Upload Credit Card Data CSV

Drag and drop file here
Limit 200MB per file • CSV

👆 Upload CSV file to use real data

🏦 Navigation

Select Page: 🔮 Churn Prediction

Live Monitoring Controls

□ Auto Refresh Data

Refresh Interval: 1 minute

⚪ Manual mode

Using sample data for demonstration. Upload your CSV file for real analysis.

🔮 Live Churn Prediction Engine

✅ Using processed data from pipeline

🎯 Customer Risk Assessment

Customer Demographics

Gender: M

Number of Dependents: 0

Education Level: High School

Income Category: Less than $40K

Card Category: Blue

Account Information

Financial Information

Credit Limit ($): 12000

Total Transaction Amount ($): 4500

👆 Complete customer details and click 'Predict Churn Risk' for live analysis

🔍 Live Customer Analysis

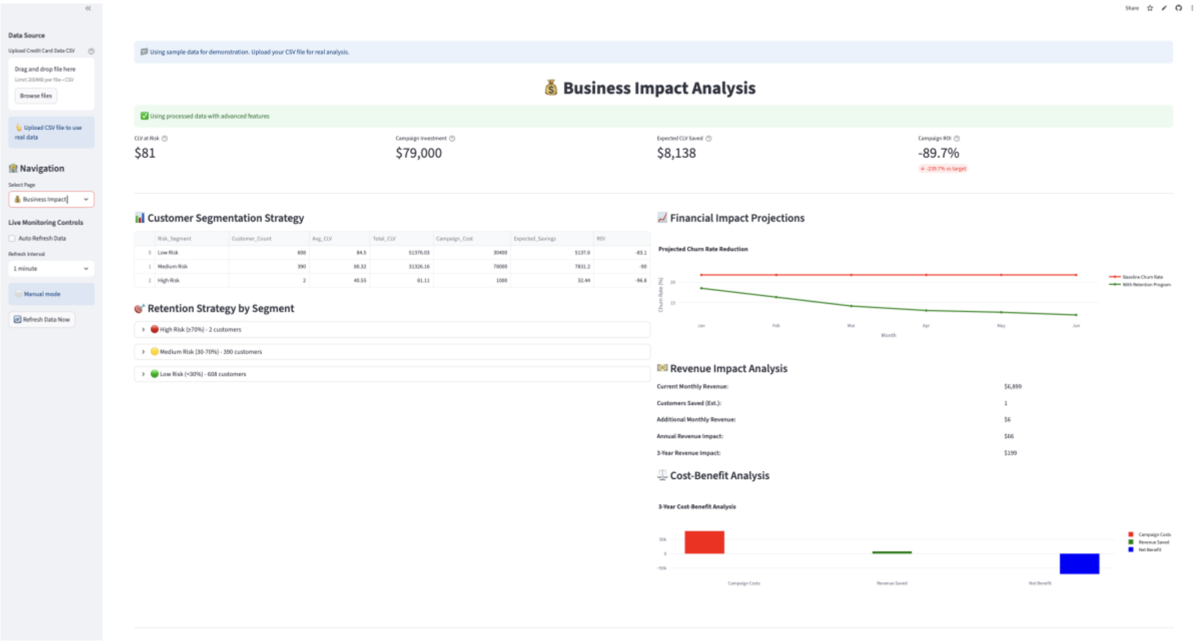
⚠️ 5 high-risk customers detected in live data

	Customer_Age	Card_Category	Months_Inactive_12_mon	Total_Trans_Ct	Avg_Utilization_Ratio
Customer_1	56	Silver	3	106	0.1526
Customer_2	32	Blue	0	71	0.2345
Customer_3	60	Blue	3	28	0.3413
Customer_4	25	Blue	3	19	0.0188
Customer_5	40	Blue	0	14	0.4249

Fig 3: Live Churn Prediction Engine

The Business Impact Analysis section translates predictive analytics into concrete financial metrics and strategic recommendations, demonstrating the return on investment for retention programs. This section calculates critical business metrics including customer lifetime value at risk, campaign investment requirements, expected revenue savings through successful retention efforts, and overall campaign return on investment percentages. The analysis segments customers into low, medium, and high-risk categories, each with tailored retention strategies, budget allocations, and expected ROI ranges.

Detailed financial projections show monthly churn rate reduction scenarios comparing baseline performance against intervention outcomes, while revenue impact calculations demonstrate the monetary value of prevented customer attrition. The cost-benefit analysis provides multi-year projections showing how retention investments compound over time, with comprehensive breakdowns of campaign costs, revenue saved, and net benefit calculations. Strategic recommendations are provided for each customer segment, including specific tactics such as VIP customer service for high-risk customers, proactive engagement campaigns for medium-risk segments, and growth-focused upselling strategies for low-risk customers.



Data Source

Upload Credit Card Data CSV

Drag and drop file here

Limit 200MB per file • CSV

Browse files

👆 Upload CSV file to use real data

🏦 Navigation

Select Page

💰 Business Impact

Live Monitoring Controls

□ Auto Refresh Data

Refresh Interval

1 minute

Manual mode

Refresh Data Now

Using sample data for demonstration. Upload your CSV file for real analysis.

💰 Business Impact Analysis

✅ Using processed data with advanced features

CLV at Risk: $81

Campaign Investment: $79,000

Expected CLV Saved: $8,138

Campaign ROI: -89.7%

↓ -239.7% vs target

📊 Customer Segmentation Strategy

	Risk_Segment	Customer_Count	Avg_CLV	Total_CLV	Campaign_Cost	Expected_Savings	ROI
0	Low Risk	608	84.5	51376.03	30400	5137.6	-83.1
1	Medium Risk	390	80.32	31326.16	78000	7831.2	-90
2	High Risk	2	40.55	81.11	1000	32.44	-96.8

🎯 Retention Strategy by Segment

› 🔴 High Risk (≥70%) - 2 customers

› 🟡 Medium Risk (30-70%) - 390 customers

› 🟢 Low Risk (<30%) - 608 customers

📈 Financial Impact Projections

Projected Churn Rate Reduction

💵 Revenue Impact Analysis

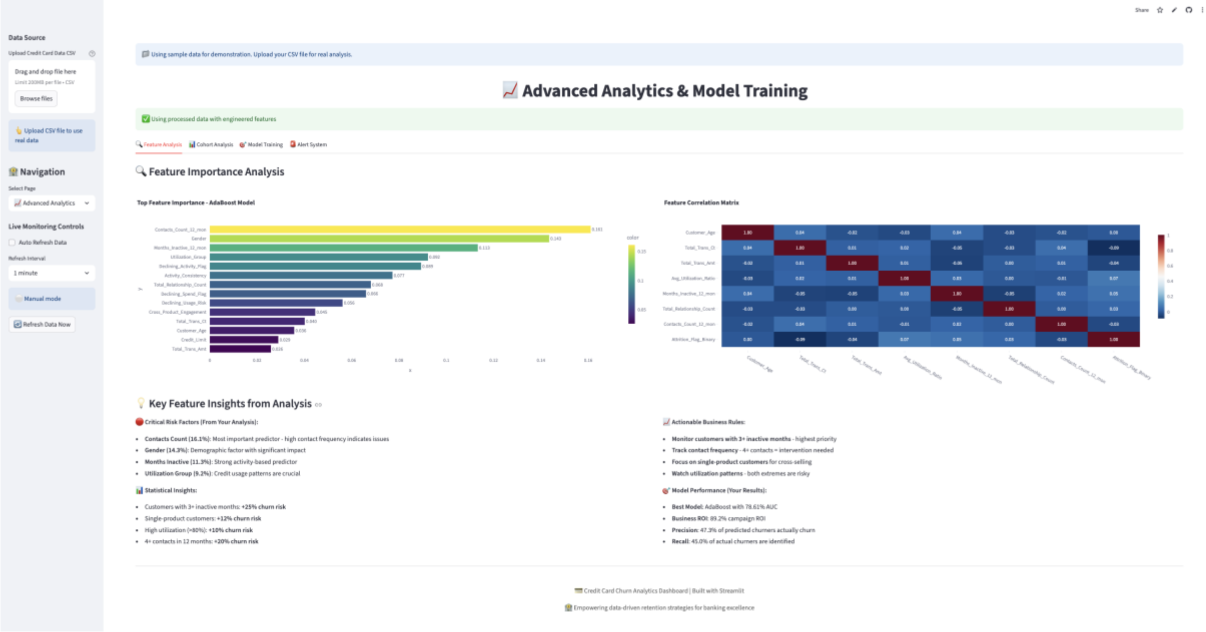
Current Monthly Revenue:	$6,899
Customers Saved (Est.):	1
Additional Monthly Revenue:	$6
Annual Revenue Impact:	$66
3-Year Revenue Impact:	$199

⚖️ Cost-Benefit Analysis

3-Year Cost-Benefit Analysis

Fig 4: Business Impact Analysis

The Advanced Analytics section provides sophisticated analytical capabilities for data scientists and analysts who require deeper insights into model performance and customer behavior patterns. The Feature Analysis tab displays comprehensive feature importance rankings based on actual model training results, correlation matrices showing relationships between predictor variables, and detailed insights derived from statistical analysis. Key findings highlight critical risk factors such as contact frequency being the strongest predictor, demographic influences, and activity-based indicators that drive churn behavior.



Data Source

Upload Credit Card Data CSV

Drag and drop file here

Limit 200MB per file • CSV

Browse files

👆 Upload CSV file to use real data

🏦 Navigation

Select Page

📈 Advanced Analytics

Live Monitoring Controls

□ Auto Refresh Data

Refresh Interval

1 minute

Manual mode

Refresh Data Now

Using sample data for demonstration. Upload your CSV file for real analysis.

📈 Advanced Analytics & Model Training

✅ Using processed data with engineered features

🔍 Feature Importance Analysis

Top Feature Importance - AdaBoost Model

Contacts_Count_12_mon 0.161
Gender 0.143
Months_Inactive_12_mon 0.113
Utilization_Group 0.092
Declining_Activity_Flag 0.089
Activity_Consistency 0.077
Total_Relationship_Count 0.068
Declining_Spend_Flag 0.066
Declining_Usage_Risk 0.056
Cross_Product_Engagement 0.045
Total_Trans_Ct 0.040
Customer_Age 0.036
Credit_Limit 0.029
Total_Trans_Amt 0.026
y
x
0
0.02
0.04
0.06
0.08
0.1
0.12
0.14
0.16
color
0.15
0.1
0.05

Feature Correlation Matrix

	Customer_Age	Total_Trans_Ct	Total_Trans_Amt	Avg_Utilization_Ratio	Months_Inactive_12_mon	Total_Relationship_Count	Contacts_Count_12_mon	Attrition_Flag_Binary
Customer_Age	1.00	0.04	-0.02	-0.03	0.04	-0.03	-0.02	0.00
Total_Trans_Ct	0.04	1.00	0.01	0.02	-0.05	-0.03	0.04	-0.09
Total_Trans_Amt	-0.02	0.01	1.00	0.01	-0.05	0.00	0.01	-0.04
Avg_Utilization_Ratio	-0.03	0.02	0.01	1.00	0.03	0.00	-0.01	0.07
Months_Inactive_12_mon	0.04	-0.05	-0.05	0.03	1.00	-0.05	0.02	0.05
Total_Relationship_Count	-0.03	-0.03	0.00	0.00	-0.05	1.00	0.00	0.03
Contacts_Count_12_mon	-0.02	0.04	0.01	-0.01	0.02	0.00	1.00	-0.03
Attrition_Flag_Binary	0.00	-0.09	-0.04	0.07	0.05	0.03	-0.03	1.00

💡 Key Feature Insights from Analysis

🔴 **Critical Risk Factors (From Your Analysis):**

- **Contacts Count (16.1%):** Most important predictor - high contact frequency indicates issues
- **Gender (14.3%):** Demographic factor with significant impact
- **Months Inactive (11.3%):** Strong activity-based predictor
- **Utilization Group (9.2%):** Credit usage patterns are crucial

📊 **Statistical Insights:**

- Customers with 3+ inactive months: **+25% churn risk**
- Single-product customers: **+12% churn risk**
- High utilization (>80%): **+10% churn risk**
- 4+ contacts in 12 months: **+20% churn risk**

📈 **Actionable Business Rules:**

- **Monitor customers with 3+ inactive months** - highest priority
- **Track contact frequency** - 4+ contacts = intervention needed
- **Focus on single-product customers** for cross-selling
- **Watch utilization patterns** - both extremes are risky

🎯 **Model Performance (Your Results):**

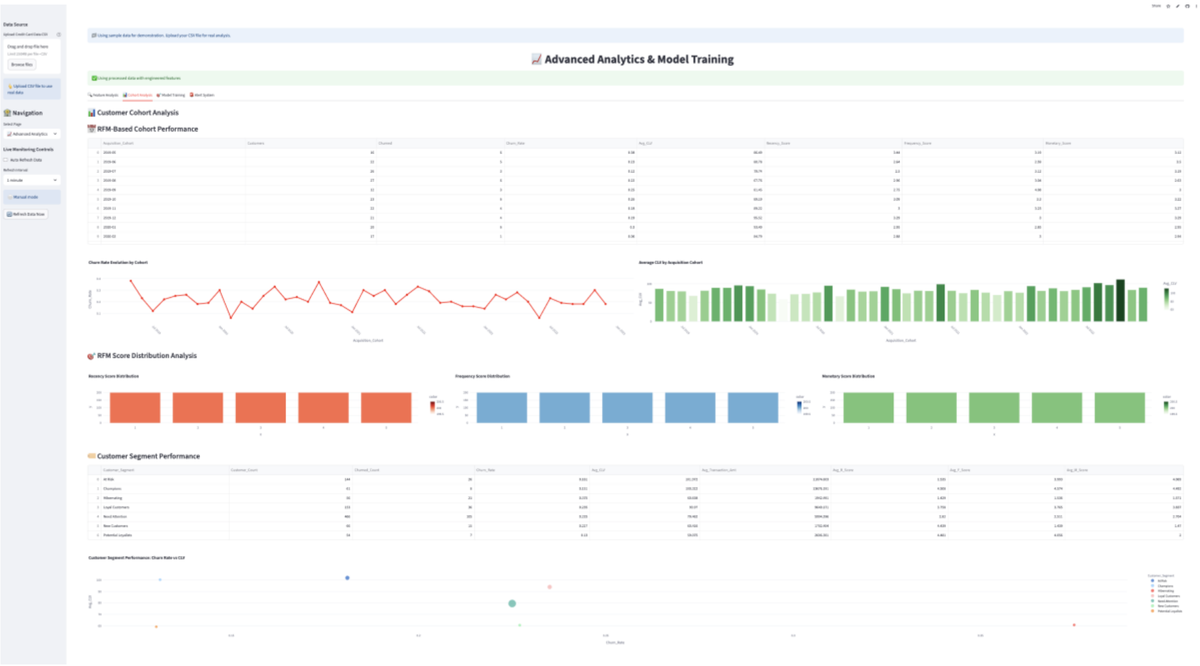
- **Best Model:** AdaBoost with 78.61% AUC
- **Business ROI:** 89.2% campaign ROI
- **Precision:** 47.3% of predicted churners actually churn
- **Recall:** 45.0% of actual churners are identified

💳 Credit Card Churn Analytics Dashboard | Built with Streamlit

🏦 Empowering data-driven retention strategies for banking excellence

Fig 5: Advanced Analytics & Model Training: Feature Analysis

The Cohort Analysis functionality enables longitudinal tracking of customer groups based on acquisition periods, RFM score distributions, and customer segment performance over time. Interactive visualizations show how different customer cohorts evolve in terms of churn rates, customer lifetime value, and engagement metrics. The Model Training tab provides a complete machine learning workflow where users can select from multiple algorithms (Logistic Regression, Random Forest, XGBoost, AdaBoost, Neural Networks), configure training parameters, and monitor real-time training progress. Results include detailed model comparison metrics, confusion matrices, ROC curves, feature importance analysis, and business impact calculations showing the financial value of model predictions.



Data Source

Upload Credit Card Data CSV

Drag and drop file here

Browse files

👆 Upload CSV file to use real data

Navigation

Select Page

Advanced Analytics

Live Monitoring Controls

□ Auto Refresh Data

Refresh Interval

1 minute

Manual mode

Refresh Data Now

Using sample data for demonstration. Upload your CSV file for real analysis.

📈 Advanced Analytics & Model Training

✅ Using processed data with engineered features

Feature Analysis | Cohort Analysis | Model Training | Alert System

📊 Customer Cohort Analysis

RFM-Based Cohort Performance

Churn Rate Evolution by Cohort

Average CLV by Acquisition Cohort

RFM Score Distribution Analysis

Recency Score Distribution

Frequency Score Distribution

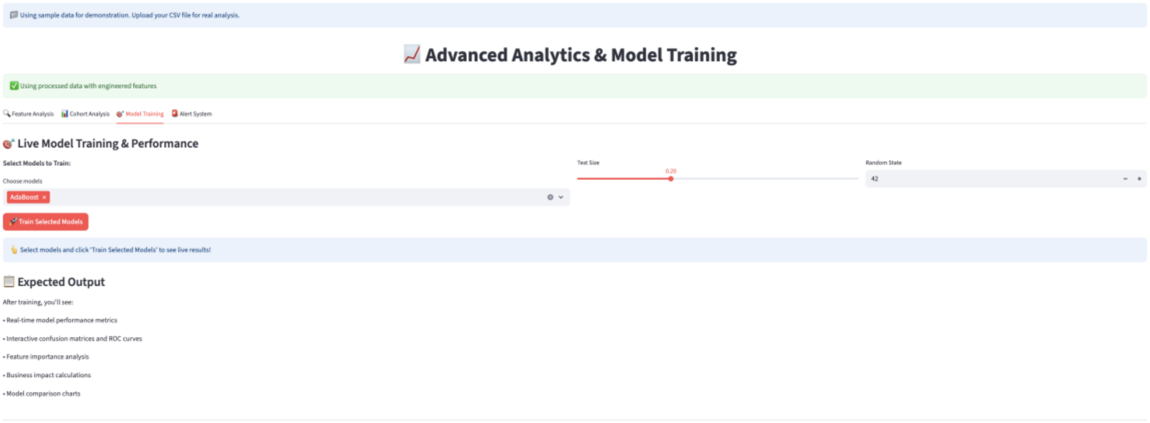
Monetary Score Distribution

Customer Segment Performance

	Customer_Segment	Customer_Count	Churned_Count	Churn_Rate	Avg_CLV	Avg_Transaction_Amt	Avg_R_Score	Avg_F_Score	Avg_M_Score
0	At Risk	144	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
1	Champions	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
2	Hibernating	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
3	Loyal Customers	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
4	Need Attention	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
5	New Customers	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
6	Potential Loyalists	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

Customer Segment Performance: Churn Rate vs CLV

Fig 6: Advanced Analytics & Model Training: Cohort Analysis



Using sample data for demonstration. Upload your CSV file for real analysis.

📈 Advanced Analytics & Model Training

✅ Using processed data with engineered features

Feature Analysis | Cohort Analysis | Model Training | Alert System

🎯 Live Model Training & Performance

Select Models to Train:

Choose models: AdaBoost

Test Size: 0.20

Random State: 42

🚀 Train Selected Models

👆 Select models and click 'Train Selected Models' to see live results!

Expected Output

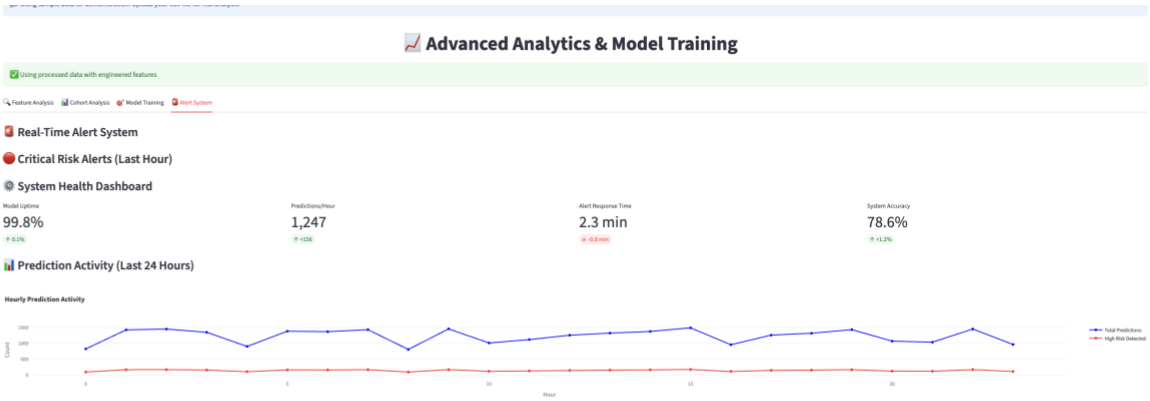
After training, you'll see:

- Real-time model performance metrics
- Interactive confusion matrices and ROC curves
- Feature importance analysis
- Business impact calculations
- Model comparison charts

Fig 7: Advanced Analytics & Model Training: Model Training

The Alert System section provides real-time monitoring and notification capabilities that enable proactive customer management and immediate response to critical situations. Critical risk alerts are prominently displayed for customers whose churn probability exceeds predefined thresholds, with detailed customer profiles showing risk factors, recommended actions, and response timelines. Each alert includes comprehensive customer information such as demographics, account details, transaction patterns, and specific risk contributors that triggered the warning.

System health metrics provide operational oversight showing model uptime, prediction throughput, alert response times, and system accuracy measures. The prediction activity timeline displays hourly volumes of risk assessments and high-risk customer identification over the past 24 hours, enabling capacity planning and performance monitoring. Recent predictions are tracked and visualized to show system utilization patterns and ensure continuous operation. The alert system also includes escalation procedures for different risk levels, ensuring that critical customers receive immediate attention while medium-risk customers are appropriately prioritized for retention campaigns.



📈 Advanced Analytics & Model Training

✅ Using processed data with engineered features

🚨 Real-Time Alert System

🔴 Critical Risk Alerts (Last Hour)

⚙️ System Health Dashboard

Model Uptime
99.8%
↑ 0.1%

Predictions/Hour
1,247
↑ +156

Alert Response Time
2.3 min
↓ -0.8 min

System Accuracy
78.6%
↑ +1.2%

📊 Prediction Activity (Last 24 Hours)

Fig 8: Advanced Analytics & Model Training: Alert System