Reduce Customer Churn for TruSource

Executive Summary

Saren Chatham, Fran Jimenez, Shamsa Khoja, Willow Noltemeyer, Arnold Songa

Context:

Customer churn directly threatens TruSource's revenue stability and long-term growth. Acquiring new customers is significantly more expensive than retaining existing ones, so the business challenge is not only predicting who will churn, but identifying early warning signals that allow the company to intervene before customers leave.

EDA and Data Processing:

We analyzed 5,636 customer records with 35 predictors covering account characteristics, service usage, billing behavior, and plan details. Early exploratory analysis revealed several patterns that shaped our modeling approach. Churn was concentrated in the early customer lifecycle, especially within the first year. Automatic payment methods were associated with lower churn, suggesting that payment friction may influence retention. We also observed strong relationships among tenure, contract type, add-ons, internet technology, and monthly fees, indicating meaningful behavioral interactions rather than isolated drivers.

These findings informed our preprocessing decisions. We removed identifier variables and a zero-variance fiscal quarter feature. Missing categorical values were imputed as "Unknown" or "No offer" to preserve behavioral signal. Customers with zero total billed and zero tenure were treated as newly onboarded and assigned total billed equal to zero. Categorical variables were one-hot encoded for linear models, and we engineered several features to better capture engagement and billing friction. These included a combined count of all add-on services, an aggregated binary streaming indicator, and indicators for extra data charges and refunds. Feature engineering focused on simplifying sparse behavioral signals into interpretable and operationally useful variables.

Modeling Approach:

To predict customer churn with confidence, we evaluated a range of statistical and machine-learning methods, progressing from simple and explainable baselines to more advanced ensemble techniques.

We began with Logistic Regression to establish a transparent benchmark and understand directional drivers of churn. We then expanded to tree-based models to capture non-linear relationships and interactions common in customer behavior data. Finally, we tested boosting algorithms designed to maximize predictive accuracy.

Model selection was based on performance on unseen data, using cross-validation and two complementary metrics:

- ROC–AUC → ability to separate churners from non-churners
- PR–AUC → effectiveness at correctly identifying true churners

While Logistic Regression delivered a strong baseline, boosting methods provided better real-world detection power. CatBoost achieved the best balance of accuracy and precision, particularly excelling in identifying high-risk customers.

CatBoost's advantage comes from native handling of categorical variables (plans, payment types, offers), reduced information loss versus one-hot encoding and a strong ability to model complex, non-linear behavior patterns. This makes it especially well-suited for our churn environment.

Recommendations & Trade-offs:

While tenure and month-to-month contracts were strong predictors of churn, these are well-known drivers. Our analysis focused on deeper interactions and meaningful customer subgroups to generate non-obvious, operational recommendations.

Using K-Means clustering on high-impact features, we identified three distinct customer segments that require differentiated retention strategies.

Segment 1: Loyal Premium Bundled Users (33%)
These customers have long tenure, high monthly fees, multiple add-ons, and high unlimited data adoption. Churn in this group is less about contract structure and more likely driven by service failures or unexpected billing shocks.

Recommendation: Implement a VIP Retention Shield. Provide priority support, proactive service monitoring, and targeted long-distance plan optimization to reduce overpayment friction. Avoid aggressive upselling that risks eroding trust.

Tradeoff: Premium support requires dedicated resources, and over-investing in already loyal customers may produce diminishing returns.

Segment 2: Split Wallet Customers (29%)
These mid-tenure customers have low monthly fees, low data usage, and limited add-ons. Many appear to split services with competitors. They are not deeply engaged and represent unrealized wallet share rather than immediate high value.

Recommendation: Offer structured data or fiber trials combined with a diagnostic outreach strategy to understand service splitting behavior. Focus on increasing engagement and converting them into primary-provider customers rather than offering blanket discounts.

Tradeoff: Competing for wallet share may require promotional investment, and some customers may never become high-value accounts.

Segment 3: Early-Life Month-to-Month Majority (38%)
This is the largest and most strategically important segment. These customers have short tenure, high monthly fees, high unlimited data adoption, and a very high prevalence of month-to-month contracts and electronic check payments. They show strong revenue potential but are highly unstable.

Recommendation: Intervene early. Promote contract conversion before month 18, bundle add-ons to increase stickiness, and actively push auto-pay adoption to reduce payment friction. Prioritize this group in predictive churn campaigns due to its size and revenue impact.

Tradeoff: Incentivizing contract lock-in can compress margins, and repeated discounting risks creating dependency that leads to churn after promotions expire.