MACHINE LEARNING CASE STUDY

Customer Churn Model

Identifying At-Risk Customers with a Decision Tree

Roger Estrada · Machine Learning in Business · AI Strategy · 2026

A pruned decision tree flagged at-risk customers with 78.5% test accuracy and an AUC of 0.815, and showed that tenure, monthly charges, and fiber-optic service are the strongest churn signals. A retention team can act on that ranking.

Business problem

Losing a customer costs more than keeping one, and the loss is usually visible only after it happens. A churn model scores each customer before they leave. It lets a retention team put limited outreach where it matters and shows which customer attributes to act on.

The technical question: which model separates leavers from stayers, and which features signal the highest risk?

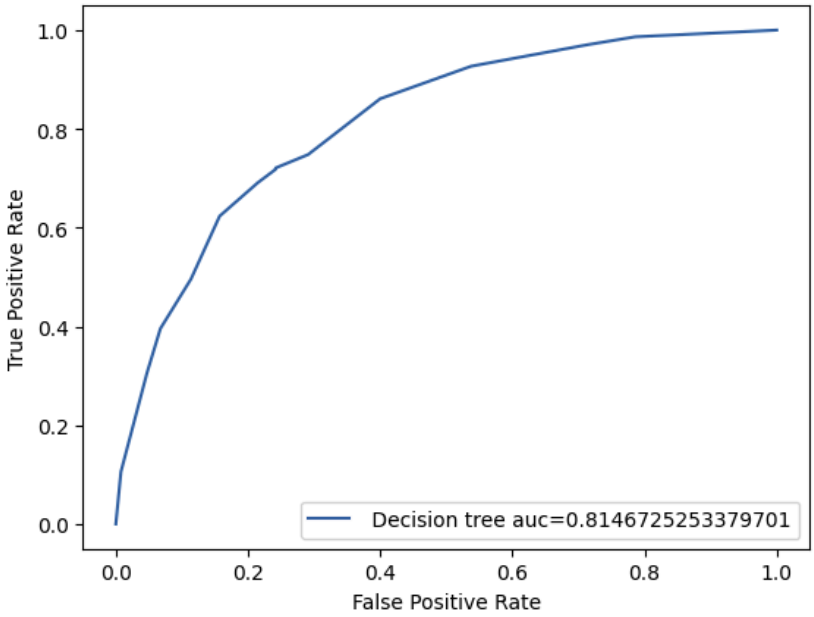
Approach and results

Data. 7,032 telecom customers described by 19 attributes covering tenure, contract, services, billing, and payment method. 1,869 customers churned (26.6%). The data had no missing values after cleaning. Categories were converted to numeric indicators, and the data was split 80/20 into training and test sets.

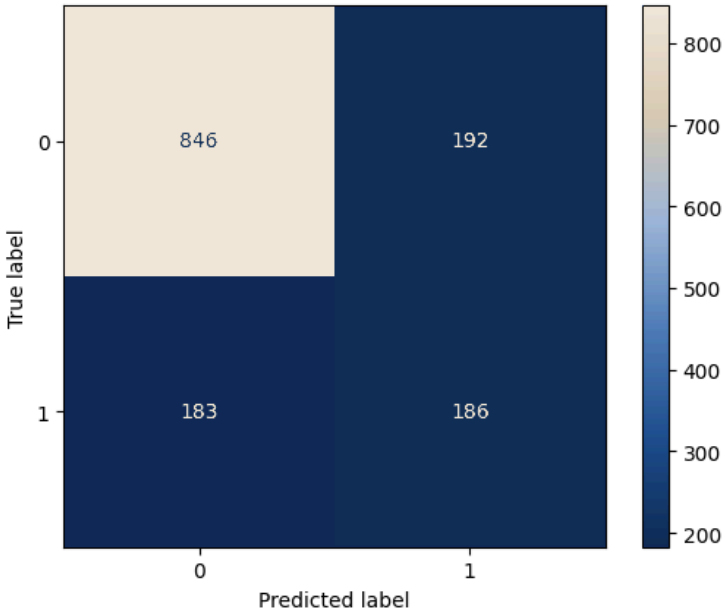
Baseline. In the test set, 73.8% of customers stayed, so a model that predicts "stays" for everyone is already 73.8% accurate. That is why accuracy alone is a weak measure here and the ROC curve and AUC carry the evaluation.

Model	Train accuracy	Test accuracy	Test AUC
Unpruned decision tree	0.998	0.733	not reported
Pruned decision tree (max depth 4)	0.791	0.785	0.815

The unpruned tree memorized the training data (99.8% accuracy) and scored 73.3% on new customers, below the "predict stays" baseline. Limiting the tree to four levels removed the overfitting. Test accuracy rose to 78.5%, close to training accuracy, so the model generalizes.



The confusion matrix for the unpruned tree shows what the errors look like in practice: of 369 customers who left, 186 were caught, and 192 loyal customers were flagged in error.



What drives churn. Feature importance from the tree points to four attributes.

Feature	Importance
Tenure	0.202
Monthly charges	0.199
Total charges	0.197
Fiber-optic internet service	0.106

Tenure, monthly charges, and total charges are closely related, so together they describe how long and how much a customer has paid. Fiber-optic service stands out as a separate signal worth investigating for price or service-quality problems.

Enterprise application

For a credit union or bank, the same workflow applies to member attrition, dormant-account reactivation, and product-depth growth. Only the features and the target change. A scored list lets outreach, pricing offers, and service interventions go to the members where they change the outcome, which ties the model to retention rate and products per member.

The interpretable tree also gives service and marketing teams plain rules to act on, such as which tenure and billing profiles carry the highest risk.

Governance and limits

- **Public data.** The model was built on public telecom data, not member data. A production version needs member-level features and privacy controls.
- **Single model.** This stage shows one interpretable model. Ensemble methods and tuned variants are the next step for lifting AUC.
- **Measuring impact.** A prediction does not prove lift. Production use needs a defined intervention and a holdout group to measure whether outreach changes the outcome.
- **Drift.** Customer behavior changes, so the model needs performance monitoring and a retraining trigger.