Customer Churn Prediction - Related Research Summaries

This document presents summaries of recent research papers on customer churn prediction. Each summary outlines the study's domain, dataset, methodology, and performance metrics in approximately 75 words. These summaries provide insights into various approaches for understanding and predicting customer churn across sectors.

GA-XGBoost for Banking Churn Prediction

This study focused on churn prediction in the banking sector, using an optimized XGBoost model with genetic algorithm tuning. The approach involved resampling techniques like SMOTEENN to address class imbalance, achieving an impressive AUC of 99.02%. SHAP analysis was used for model interpretability, identifying key features influencing churn, such as transaction volume and customer age range.

Telecom Sector Churn Prediction

In a telecom-specific study, researchers applied machine learning techniques to predict churn by analyzing customer activity patterns. The model utilized features like call duration, internet usage, and monthly charges to gauge customer loyalty, achieving high predictive accuracy. Data preprocessing, including feature scaling and dimensionality reduction, enhanced model performance, which was further fine-tuned using cross-validation techniques.

Comparative Study Using Ensemble Models

A comparative analysis of machine learning models, including Random Forest, XGBoost, and Logistic Regression, was conducted across telecom datasets. Ensemble methods, especially Random Forest, demonstrated the highest accuracy in predicting churn, while logistic regression, despite lower accuracy, offered better interpretability. Models were evaluated using accuracy, precision, and recall, with Random Forest performing consistently across datasets.

Deep Learning in Customer Churn Prediction

This paper investigated deep learning models, specifically CNN and LSTM, to predict customer churn in e-commerce. The dataset comprised of over 200,000 transaction records. By capturing temporal patterns, LSTM yielded higher accuracy than CNN, reaching an AUC of 0.92. Data augmentation was used to improve training balance, and dropout regularization prevented overfitting.

Retail Churn Prediction Using Decision Trees

A study on retail churn prediction applied decision tree models on transactional data to classify churn probabilities. Key features included purchase frequency, basket size, and

discount usage. The decision tree model achieved a recall of 0.81, making it effective in identifying high-risk churn customers. Pruning was employed to avoid model complexity and overfitting.

Boosted Models for Financial Services Churn

In the financial sector, a study used boosted models like AdaBoost and CatBoost to predict churn. Datasets included customer account details and transaction histories. The CatBoost model achieved the highest F1-score of 0.87, while SHAP values indicated features like credit card activity and loan payments were most impactful. Feature selection reduced model complexity without affecting accuracy.

Telecommunications Churn Prediction with SVM

Support Vector Machines (SVM) were used in a telecommunications churn study, leveraging customer service records and monthly spending as inputs. The model reached an accuracy of 83%, with optimized hyperparameters enhancing performance. Data imbalance was handled using undersampling, and the final model was effective at isolating high-risk churners. Precision and recall were both notably high.

Logistic Regression for Predicting Churn in SaaS

In this SaaS-focused churn study, logistic regression was utilized due to its interpretability. The dataset contained user engagement metrics like login frequency and feature usage. The model achieved 76% accuracy and high interpretability, with key predictors being usage frequency and time since last login. It's especially suited for scenarios needing straightforward churn predictions with easy-to-understand outcomes.

Comparing Neural Networks and Gradient Boosting in E-commerce

This study compared neural networks and gradient boosting for churn prediction in e-commerce. Neural networks excelled in capturing complex customer behavior patterns, achieving 90% accuracy, but gradient boosting provided faster results with slightly lower accuracy (88%). Feature engineering focused on purchase history and browsing time to optimize model predictions.