

Churn Prediction with Advanced Feature Engineering and Ensemble Methods

Resumo

Este estudo explora a aplicação prática de técnicas avançadas de machine learning ao problema de predição de churn em um conjunto de dados tabulares de clientes. O foco está na extração de insights acionáveis por meio da combinação de pré-processamento, engenharia de atributos e modelagem em ensemble.

Metodologia

O conjunto de dados foi submetido a uma análise exploratória detalhada para identificar os atributos mais preditivos e padrões subjacentes associados ao churn. A integridade dos dados foi assegurada com o tratamento de valores ausentes, outliers e inconsistências categóricas. A engenharia de atributos desempenhou papel central, incluindo a criação de termos de interação, variáveis de razão, binning direcionado de variáveis contínuas e mean encoding para categorias selecionadas, enriquecendo a capacidade representacional dos dados.

As etapas de pré-processamento incluíram a codificação one-hot de variáveis categóricas e a normalização de variáveis numéricas. Diversos modelos de machine learning foram implementados e comparados, incluindo máquinas de gradient boosting (LightGBM, CatBoost), regressão logística regularizada e uma rede neural MLP. Para capturar relações complexas e reduzir a variância, foi empregado um ensemble stacking, aproveitando as forças preditivas dos modelos individuais em um framework de meta-aprendizagem. O desempenho foi avaliado via validação cruzada estratificada, utilizando ROC AUC como métrica principal.

Resultados e Impacto Prático

O modelo final de ensemble atingiu um ROC AUC de 75,07%, demonstrando alta capacidade discriminativa para identificar clientes com risco de churn. A integração de atributos engenheirados e ensemble learning não só elevou a acurácia preditiva, como também resultou em uma solução robusta a overfitting. A análise de importância de atributos evidenciou que as novas features de interação e razão contribuíram significativamente para o desempenho.

No contexto de negócios, tal abordagem possibilita a identificação antecipada de clientes propensos ao cancelamento, viabilizando estratégias de retenção direcionadas. A metodologia é escalável e pode ser adaptada a problemas similares em setores de assinatura, onde a redução de churn impacta diretamente a estabilidade de receita e o valor do ciclo de vida do cliente.

Conclusão

O trabalho demonstra os benefícios práticos da combinação de engenharia avançada de atributos com técnicas de ensemble para predição de churn. O modelo resultante oferece previsões confiáveis, interpretáveis e acionáveis, agregando valor imediato à pipeline analítica de negócios.