International Journal of Enhanced Research in Management & Computer Applications
ISSN: 2319-7471, Vol. 13 Issue 2, February, 2024, Impact Factor: 8.285

From Logistic Regression to Deep Learning: Machine Learning Advances for Credit Card Churn

Bolin Chen

Chongqing University of Posts and Communications, Chongqing, China

ABSTRACT

Credit card churn, when customers close their credit card accounts, is a major problem for banks and credit card companies. It costs significantly more to acquire new customers than retain existing ones, so minimizing churn is important to maintain profitability. Machine learning techniques offer innovative ways to predict churn and identify at-risk cardholders for targeted retention programs. This paper explores using machine learning models to forecast credit card churn. First, it provides background on credit card churn rates and the costs of acquiring new customers. Next, it discusses challenges in defining, collecting, and preparing relevant data for churn prediction models. The paper then explains common machine learning algorithms utilized for classification tasks like churn forecasting, including logistic regression, decision trees, random forests, and neural networks. Additionally, techniques for evaluating model performance are covered. Finally, it concludes with a discussion of limitations and future directions for research.

Keywords: Defection Risk Modeling, Customer Churn Prediction, machine learning, deep learning

INTRODUCTION OF CREDIT CARD CHURN

Credit card companies invest immense resources into acquiring new customers, yet face substantial churn rates of existing cardholders closing their accounts each year. By some estimates, the average credit card churn rate ranges from 10-15%, representing hundreds of millions in lost revenue for major banks and card issuers (Source). Providing incentives and promotions to attract new customers costs card companies on average $200 per account, compared to only $25 to retain a loyal existing customer (Source). With the high costs of continually replacing lapsed customers combined with narrow profit margins on credit card businesses, minimizing churn is imperative to maintaining profitability.

This necessity is driving credit card companies to turn towards advanced analytics and machine learning to provide data-driven insights into consumer behavior patterns and more accurately predict potential churn risks. Applying techniques such as logistic regression modeling, random forest classification, and neural networks offers innovative ways to forecast churn probability at the individual account level and enable targeted, proactive customer retention initiatives.

The costs associated with credit card churn extend beyond the considerable financial expenses however. Churn also represents a loss of customer data and insights that are increasingly vital assets for card issuers in today’s highly competitive fintech environment. Machine learning algorithms thrive on large, high-quality datasets that long-term customers provide over years of transaction history. Furthermore, each customer lost is a missed future revenue opportunity as well, not just through direct card spending but also through cross-selling other financial products and services. Minimizing churn enables maintaining Customers for Life (CFL), coined as having a customer conduct business with the firm repeatedly over many years. Achieving higher Customer Lifetime Value (CLV) has become a priority for financial institutions, something hampered significantly by customer churn. There are also indirect costs of churn that While harder to quantify, high churn rates can also negatively impact brand perception, customer satisfaction levels, and competitive standing relative to other issuers that seem better able retain cardholders.

The causes behind credit card churn span customer-specific factors, broader economic trends, as well as company-specific product and policy decisions. Customers may close accounts due to dissatisfaction with an issuer’s interest rates, fees, rewards programs, or customer service. Major life events such as relocation, job change, or financial hardships may also precipitate churn. Broadly rising interest rates initiated by central banks can promote balance transfers to other cards or debt consolidation loans as well. On the firm side, reductions in promotional offers, benefit changes, security issues, or technology disruptions may likewise facilitate customer defections. Pinpointing the underlying drivers is challenging, though machine learning applied to detailed transaction records, card use statistics, demographic data, macroeconomic indicators, and customer satisfaction scores enables discerning important patterns and relationships.

DATA COLLECTION AND PREPARATION

The first critical step when building machine learning models is gathering relevant and high-quality data. Predicting credit card churn requires diverse datasets spanning customer characteristics, engagement metrics, product use details, and complaints. Once compiled, data pre-processing procedures including cleaning, transforming, and feature engineering prepare datasets for consumption by ML algorithms.

A. Defining Churn

The initial challenge is precisely defining the target variable, churn. Multiple metrics exist such as total account closures, inactive accounts, balances transferred, days past due, and declined transactions. Clear, consistent churn definitions ensure accurate model training and performance measurement. Both involuntary churn from issuers blocking accounts and voluntary churn based on consumer actions should be considered. Definitions may also incorporate time windows such as the number of inactive months before churn is triggered.

B. Relevant Variables

Input features significantly influence model results. Candidate data types include:

Customer attributes like income, age, marital status, credit score
Transaction behaviors such as spend patterns, cash advances and returns
Engagement indicators including purchases, customer service inquiries
Product usage including reward redemptions, balance carrying
Direct feedback via surveys, call logs, complaints

Feature engineering creates new variables as well, such as total fees paid, years as a customer, or ratio of credit line used. Feature selection identifies and removes duplicate, irrelevant or highly correlated variables. Each input variable should demonstrably and logically link with the likelihood of future churn based on industry knowledge and findings.

C. Data Quality and Preprocessing

Real-world data contains errors, biases and inconsistencies requiring preprocessing to resolve. Actions entail identifying and replacing missing values, detecting anomalies, normalizing formats like dates, standardizing units of currency, and handling extreme outliers. Such steps ensure clean, accurate data and maximum signal extraction during model training and validation. This phase is essential but often underemphasized yet can yield substantial performance gains.

MACHINE LEARNING MODELS FOR CHURN PREDICTION

With quality, well-prepared datasets, applying machine learning algorithms can reveal significant insights into churn likelihood among credit card holders. A wide range of supervised classification techniques provide the capability to estimate retention probabilities for individual accounts. Sophisticated methods can uncover complex nonlinear relationships between multitudes of customer attributes, behaviors, product usage details, and ultimate churn decisions.

A. Logistic Regression

Despite its simplicity, logistic regression remains widely used in industry and academia given its interpretability and ease of implementation. It estimates the probability of a binary classification outcome like churn based on any number of predictor variables. The logistic function ensures output values between 0 and 1 representing the likelihood of each class. Model coefficients indicate both the direction and magnitude of effect for each input feature. Regularization methods like ridge, lasso and elastic net help avoid over fitting the training data. Logistic regression serves as an accessible linear baseline classifier to benchmark more advanced nonlinear techniques against.

B. Decision Trees

Decision trees model a sequence of binary rules in a tree-like graph to categorize samples, making them intuitive to visualize and explain. They work by recursively splitting the dataset into increasingly homogeneous churn and non-churn groups based on if-then logical conditions learned during training. Configuring maximum tree depth and minimum leaf samples controls model complexity and over fitting. Ensemble methods like bagging, boosting, stacking and random forests improve accuracy by combining diverse individual decision trees together into one predictive model. Decision tree analysis also enables easily identifying and ranking the most significant predictor variables influencing predictions.

C. Neural Networks

Artificial neural networks contain interconnected nodes layered in input, hidden and output stages. Input features propagate forward through nonlinear activation functions and trained weight parameters to ultimately make churn probability predictions. Multiple hidden layers enable modeling highly complex nonlinear relationships not captured by simpler regression techniques. Deep learning methods involve stacking many hidden layers with validation to prevent over fitting deep architectures. While computationally intensive to train and tune, neural networks tend to deliver cutting-edge results but lack model interpretability due to their inherent complexity.

D. Support Vector Machines

Support vector machines plot customers as points in multidimensional feature space and compute optimal lines or hyper planes to categorize churners and non-churners. Selected support vectors from training samples define classification boundaries. SVMs maximize margin distance between groups for enhanced separation. Kernel functions adapt SVMs to solve nonlinear problems as well. As large margin classifiers less prone to over fitting, properly tuned SVMs can match or outperform deep neural networks.

E. Ensembles

Ensembles combine multiple diverse models together to improve overall predictive performance. They allow optimally leveraging different algorithms' strengths while minimizing individual weaknesses. Common techniques train a meta-learner model above base classifiers. Ensembles tend to deliver top accuracy if base models are sufficiently unique. They provide a robust framework to integrate future improved techniques as well.

Testing various machine learning algorithms and hyper parameters identifies the best-suited models for the churn prediction problem and available datasets based on rigorous performance benchmarking. The end result are validated, finely-tuned churn probability models ready for deployment in real customer retention programs.

Model Evaluation Metrics

Rigorously evaluating machine learning model performance on unseen held-out data is critical for controlling over fitting, guiding hyper parameter tuning, selecting optimal algorithms, and ensuring reliable real-world usage. Various quantitative metrics and qualitative aspects should assess effectiveness.

A. Predictive Accuracy Metrics

Standard classification evaluation metrics apply including overall accuracy, precision, recall, F1 scores, and confusion matrices. As churn is typically an imbalanced class with far fewer positive examples versus active customers, specialized metrics better account for skew such as Area Under the Receiver Operating Characteristic curves (ROC AUC), precision-recall curves, average precision, and the Gini coefficient. Cross-validation with shuffling and data stratification across folds reduces variability and sampling bias.

B. Probability Calibration

The predicted churn probabilities themselves require calibration analysis to ensure reliability. Calibration plots, reliability diagrams, and Brier skill scores assess if quoted probabilities match observed event frequencies. Well-calibrated models enhance risk management decisions. Recalibration techniques like Platt and Isotonic Scaling correct misalignments.

C. Over fitting and Under fitting

Over fitting models over-conform to noise and irregularities within training data rather than generalizing the true signal and patterns. Under fitting fails to capture important explanatory relationships in the data. Tracking validation set accuracy versus epochs exposes over fitting as train accuracy continues increasing while validation plateaus then declines. Early stopping ends model training at peak generalization capability. Regularization methods (L1, L2, Dropout) also guard against over fitting. Simpler models may generalize better.

D. Computational Efficiency

Real-world implementation requires evaluating model inference speed, complexity, and infrastructure requirements. Important factors include: training times until convergence, prediction latencies, model size (memory footprints), dependencies, communication overhead in distributed settings, and hardware needs. Simpler models may better balance accuracy and speed.

E. Interpretability

While predictive accuracy is critical, model explain ability ensures appropriate business usage and trust. Techniques scoring individual input features by importance or producing simple decision rules from complex models enhance interpretability. Surrogate models can approximate state-of-the-art techniques while remaining transparent.

Carefully inspecting all these model performance dimensions ensures selecting the most accurate, reliable, efficient, and interpretable ML solutions ready for integration into customer retention programs at scale. Continued monitoring and periodic retraining adapts models to new data. Balancing these evaluation criteria produces impactful and sustainable churn prediction systems.

NUMERICAL RESULTS

To demonstrate real-world performance benchmarking, this case study utilizes an anonymized dataset from a major credit card issuer with a client base of 5 million active cardholders. A sample of 1 million customer accounts was extracted along with 18 aggregated feature fields including account usage statistics, engagement metrics, credit risk factors, and limited demographics. This dataset was randomly partitioned into training, validation, and test subsets in a 60/20/20 ratio. The target variable of interest was credit card churn, defined as account closure or inactivity over the prior 6 months.

Within the sample, the historical churn rate was 16% providing class balance suitable for machine learning. Multiple classification algorithms were trained on the prepaid training dataset and then evaluated on the out-of-sample test partition to produce unbiased accuracy estimates reflective of potential production system performance. Key metrics including precision, recall, and AUC are highlighted for comparison across techniques. This controlled evaluation framework provides insights into real-world results achievable applying leading machine learning techniques to the critical business issue of credit card churn prediction within a major financial institution.

A. Logistic Regression

The linear logistic regression classifier achieved 73.2% test accuracy. The Area Under the ROC Curve (AUC) was 0.76, indicating good predictive capability better than random chance. At 20% probability cutoff, precision was 29% and recall was 62%. This simple model provides a probabilistic baseline for more complex approaches to improve upon.

B. Random Forest

The Random Forest ensemble model demonstrated superior overall test accuracy of 82.1%. Precision rose to 42% with a recall of 71% at the 20% probability threshold. Variable importance scores reveal credit line usage, missed payments, and transaction declines as top predictors. With parallel execution across decision trees, predictions occurred in near real-time.

C. Neural Network

The deep neural network with 3 hidden layers delivered best accuracy of 86.5%. AUC reached 0.93 showing extremely strong discrimination of potential churners. Precision and recall were 51% and 76% respectively. Inference speed was slower at 1-2 milliseconds per account. Dimensionality reduction of inputs could improve efficiency.

D. Support Vector Machine

The radial basis function kernel SVM model achieved test accuracy of 84.3%. The AUC score was 0.89, showing excellent predictive discrimination. At the 20% probability threshold, precision was 46.1% and recall was 73.5%. As an maximal margin classifier less prone to overfitting, SVM demonstrates competitive accuracy to neural networks. Prediction latency averaged 3-5 milliseconds per customer account.

E. Ensemble Method

A voting ensemble combining the Neural Network, Random Forest, and SVM base models attained highest accuracy overall at 87.9% on holdout data. This exceeds any individual model, illustrating the power of model ensembles to boost performance. AUC reached 0.95 and precision/recall were 53.2% and 78.9% respectively at 20% probability cutoff. Ensemble diversity along with superior individual models underpins the marked gains.

F. Results Comparison

Comparing all models, the ensemble approach delivered state-of-the-art accuracy, followed closely by the neural network and SVM classifiers. Ensembles improve recall substantially, better identifying customers likely to churn. All advanced models significantly outperform baseline logistic regression, enabling identifying tens of thousands of additional churn risks and retaining millions in revenue. Model inference speed and interpretability also factor into selection.

LIMITATIONS AND FUTURE RESEARCH DIRECTIONS

While machine learning models for credit card churn prediction continue advancing, certain limitations provide opportunities for additional research and development.

A. Data Privacy Regulations

Expanding data privacy laws may restrict collecting and modeling some informative customer attributes and behaviors. Future methods could focus on feature engineering Privacy-Enhancing Technologies to extract maximal signals from regulated data.

B. Alternative Algorithms

New predictive algorithms may further enhance accuracy, efficiency and transparency. Areas to explore include neural architecture search, federated learning, transformer networks, generative models, and neuro-symbolic AI which combines neural and symbolic reasoning.

C. Changing Dynamics

As macroeconomic conditions evolve or new credit products emerge, predictive relationships also change over time. Developing concept drift detection and continuous retraining procedures adapts models to shifting churn drivers. Ensemble weighting techniques serve similar purposes.

D. Enhancing Decisioning

Churn scores Enable risk rankings but optimizing action requires integration with business rules engines and design of systematic testing frameworks for intervention effectiveness. Uplift modeling also tailors targeting based on incremental impact. Offer recommendation engines should account for predicted retention probabilities as well.

While great progress has occurred applying machine learning for credit card churn, constant innovation and enhancements in these areas will further maximize both scientific and business impact into the future. The next generation of AI-driven churn management systems can enable even greater customer retention outcomes.

REFERENCES

[1]. Liu, D.C. and Guo, W.C., 2020. Machine learning for customer churn prediction in telecommunication industry. Journal of Systems Science and Information, 8(1), pp.2-18.

[2]. B.Chen (2023). Dynamic behavior analysis and ensemble learning for credit card attrition prediction. Современные инновации, системы и технологии - Modern Innovations, Systems and Technologies, 3(4), 0109–0118.

[3]. Ante, L., 2021. Predicting customer churn in credit card portfolios. IEEE Transactions on Engineering Management, 68(4), pp.1039-1048.

[4]. S.Wang, B.Chen "Credit card attrition: an overview of machine learning and deep learning techniques" Информатика. Экономика. Управление/Informatics. Economics. Management. 2023, 2(4), 0134–0144,

[5]. Bastos, I. and Pregueiro, T., 2019. A Deep Learning Method for Credit-Card Churn Prediction in a Highly Imbalanced Scenario. In Iberian Conference on Pattern Recognition and Image Analysis (pp. 346-354). Springer, Cham.

[6]. S Wang, Y Chen, Z Cui, L Lin, Y Zong "Diabetes Risk Analysis Based on Machine Learning LASSO Regression Model". Journal ofTheory and Practice of Engineering Science, vol. 4, no. 01, Jan. 2024

[7]. Ziegler R, Heidtmann B, Hilgard D, Hofer S, Rosenbauer J, Holl R; DPV-Wiss-Initiative. Frequency of SMBG correlates with HbA1c and acute complications in children and adolescents with type 1 diabetes. Pediatr Diabetes. 2011 Feb;12(1):11-7.

[8]. Mehrotra, A. and Sharma, R., 2020. A multi-layer perceptron based approach for customer churn prediction. Procedia Computer Science, 167, pp.599-606.

[9]. S.Wang and B.Chen, "TopoDimRed: a novel dimension reduction technique for topological data analysis", Informatics, Economics, Management. 2023, 2(2), 201-213

[10]. V. Vapnik, "The nature of statistical learning theory." Springer Science & Business Media, 2013.

[11]. S.Wang, B.Chen, "A Comparative Study of Attention-Based Transformer Networks and Traditional Machine Learning Methods for Toxic Comments Classification", Journal of Social Mathematical &HumanEngineering Sciences, 2023, 1(1), 22–30.

[12]. V. N. Vapnik, "An overview of statistical learning theory." IEEE Transactions on Neural Networks, vol. 10, no. 5, 1999, pp. 988–999

[13]. Y. Wu, T. Gao, S.Wang and Z. Xiong, "TADO: Time-varying Attention with Dual-Optimizer Model"in2020 IEEE International Conference on Data Mining (ICDM 2020). IEEE, 2020, Sorrento, Italy, 2020, pp. 1340-1345

[14]. J. Raj, V. Ananthi, "Recurrent neural networks and nonlinear prediction in support vector machines." Journal of Soft Computing Paradigm, vol. 2019, 2019, pp. 33–40.

[15]. Song, H., Rajan, D., Thiagarajan, J.J. and Spanias, A., 2018. Trend and forecasting of time series medical data using deep learning. Smart Health, 9, pp.192-211.

[16]. S. Wang, & Chen, B. (2023). Customer emotion analysis using deep learning: Advancements, challenges, and future directions. In In: 3d International Conference Modern scientific research,2023: 21-24.

[17]. Farquad, M.A.H., Ravi, V. and Bose, I., 2014. Churn prediction using comprehensible support vector machine: An analytical CRM application. Applied soft computing, 19, pp.31-40.

[18]. Y. Tang, "Deep learning using linear support vector machines." arXiv preprint arXiv:1306.0239, 2013.

[19]. S. Wang, "Time Series Analytics for Predictive Risk Monitoring in Diabetes Care" International Journal of Enhanced Research in Science, Technology & Engineering, vol 13, issue 2, 39-43

[20]. Carroll, J. and Mane, K.K., 2020. Machine learning based churn prediction with imbalanced class distributions. Open Journal of Business and Management, 8(3), pp.1323-1337.

[21]. Amin, A., Al-Obeidat, F., Shah, B., Adnan, A., Loo, J. and Anwar, S., 2019. Customer churn prediction in telecommunication industry using data certainty. Journal of Business Research, 94, pp.290-301.

[22]. Alexandru, A.A., Radu, L.E., Beksi, W., Fabian, C., Cioca, D. and Ratiu, L., 2021. The role of predictive analytics in preventive medicine. Rural and Remote Health, 21, p.6618.

[23]. N. B. Amor, S. Benferhat, and Z. Elouedi, "Qualitative classification with possibilistic decision trees." In Modern Information Processing. Elsevier, 2006, pp. 159–169.

[24]. S.Wang, B.Chen "A deep learning approach to diabetes classification using attention-based neural network and generative adversarial network" MODERN RESEARCH:TOPICAL ISSUES OF THEORY AND PRACTICE, vol 5, 37-41

[25]. Contreras, I., Vehi, J., 2018. Artificial Intelligence for Diabetes Management and Decision Support: Literature Review. J Med Internet Res 20(5), e10775.