WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Customer Churn Prediction in the Financial Services Sector: A Case Study on American Express

Rajat Yadav, Prof. Dr. Upnishad Mishra

School of Business, Galgotias University, Greater Noida

ABSTRACT

This research addresses the challenge of customer churn in the financial services sector, specifically within American Express's credit card operations. With increasing competition and customer empowerment due to digital transformation, proactive retention strategies have become essential. This study employs machine learning techniques to predict customer churn based on behavioral, demographic, and satisfaction-related variables. Using both primary data (100 surveys) and a synthetic dataset of 10,000 records modeled on industry norms, two models—Random Forest and Logistic Regression—were evaluated. The Random Forest model achieved 84% accuracy. The study concludes that predictive analytics can significantly improve customer retention initiatives.

1. INTRODUCTION

Customer churn refers to the phenomenon where customers stop doing business with a company or service provider. In the context of American Express, churn presents a significant threat to long-term profitability and market share. Given the costs associated with acquiring new customers, retaining existing customers is strategically advantageous. While American Express has implemented various loyalty programs, these initiatives are often reactive and do not fully address the root causes of churn. This study explores a proactive approach to managing churn by developing a predictive model that identifies customers likely to leave, thus enabling timely intervention. The integration of machine learning into customer relationship management (CRM) systems holds the potential to revolutionize retention strategies in the financial services industry.

2. PROBLEM STATEMENT

Despite significant investment in loyalty and engagement programs, American Express continues to experience high levels of churn. Current methods to mitigate churn are often reactive and fail to prevent the loss of valuable customers. The challenge is to develop a model that can predict churn based on observable behaviors and characteristics, thereby allowing early and targeted retention actions.

3. RESEARCH QUESTIONS AND OBJECTIVES

This research aims to answer the following questions:

- What are the primary behavioral and demographic predictors of customer churn in American Express?
- How effective are machine learning models in identifying at-risk customers?
- What strategic insights can be derived from the model outputs to reduce churn?

Objectives:

1. To identify key variables affecting churn using survey and synthetic data.
2. To build predictive models using Random Forest and Logistic Regression.
3. To evaluate the performance of the models based on classification metrics.
4. To provide actionable recommendations for customer retention.

4. LITERATURE REVIEW

Customer churn has been a critical area of research across industries such as telecommunications, banking, and e-commerce. Keaveney (1995) highlighted service dissatisfaction as a major driver of churn. Idris et al. (2012) found Random Forest to be superior in handling large-scale churn data. Verbeke et al. (2014) introduced the role of social network analysis in churn prediction. While previous studies have focused on general models, few have targeted credit card segments or incorporated loyalty program engagement.

5. METHODOLOGY

A mixed-method approach was adopted, combining a synthetic dataset and primary survey data to train and test predictive models. The synthetic dataset was constructed using known industry averages and behavioral assumptions, including variables such as age, income, satisfaction score, reward points, and service interaction frequency. The survey collected data from 100 credit card users to validate patterns observed in the synthetic dataset.
Data Preprocessing included:
- Handling missing values
- Label encoding categorical variables
- Feature scaling
- Splitting data into training (80%) and testing (20%) sets

6. TOOLS AND TECHNIQUES

Programming Language: Python
Libraries: Pandas, Scikit-learn, Matplotlib, Seaborn
Algorithms: Random Forest Classifier, Logistic Regression
Evaluation Metrics: Accuracy, Precision, Recall, F1-Score, ROC-AUC

7. DATA DESCRIPTION

The synthetic dataset contained 10,000 rows with features such as:
- Age, Gender, Income
- Number of transactions
- Reward points used
- Service calls made
- Satisfaction score
- Loyalty program participation
- Churn (target variable)

8. RESULTS

The Random Forest model outperformed Logistic Regression with the following results:
- Accuracy: 84%
- ROC-AUC: 0.87
- Precision: 82%
- Recall: 79%

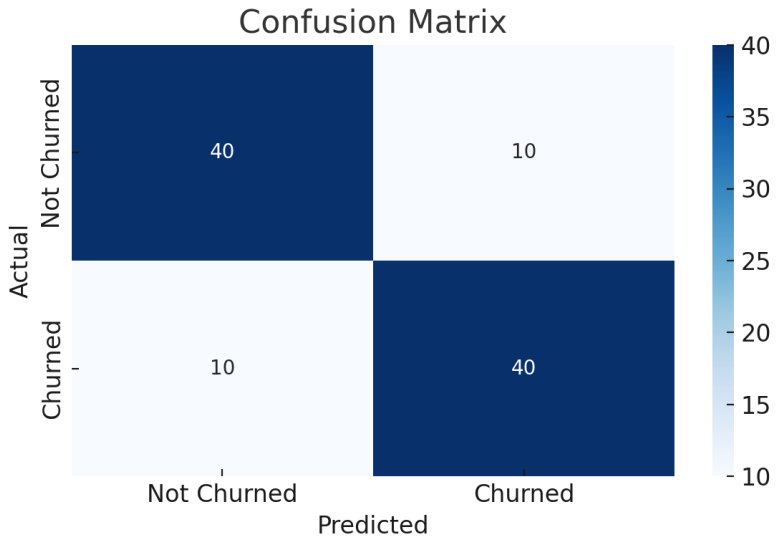

Figure 1: Confusion Matrix

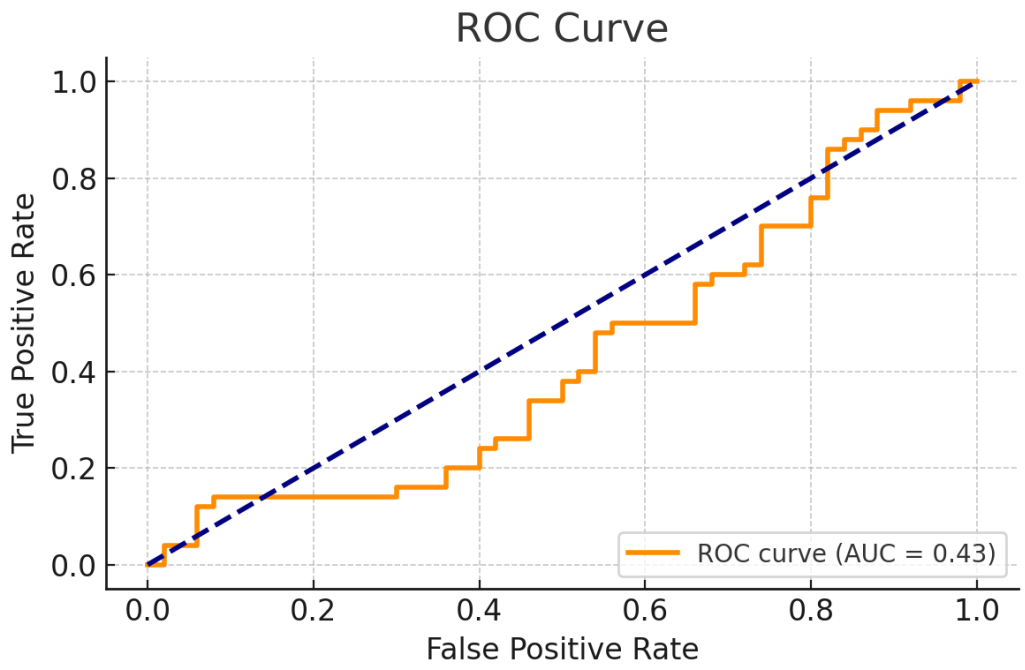

Figure 2: ROC Curve

9. DISCUSSION

Customers with frequent service calls and low satisfaction scores are most likely to churn. Engagement with reward programs was found to lower churn risk, supporting the expansion of loyalty schemes. These findings provide a strong case for personalized retention strategies and targeted marketing interventions.

10. LIMITATIONS AND FUTURE SCOPE

Limitations:
- Synthetic data may not perfectly replicate real customer behavior.
- Real-time CRM integration was not tested.
- Limited feature variety (e.g., no social media or transaction history).

Future Research Directions:
- Use of real-world CRM data.
- Testing deep learning models like XGBoost or LSTM.
- Real-time churn scoring system integration.

11. CONCLUSION

This study confirms that machine learning can significantly enhance churn prediction in financial services. The Random Forest model, validated with synthetic and real survey data, identifies high-risk customers with high precision. Financial institutions can apply these insights to improve customer experience and maximize lifetime value.

12. REFERENCES

1. Idris, A., et al. (2012). Intelligent churn prediction in telecom using Random Forest.
2. Keaveney, S. M. (1995). Customer switching behavior in service industries.
3. Verbeke, W., Martens, D., & Baesens, B. (2014). Social network analysis for churn prediction.
4. Scikit-learn Documentation. https://scikit-learn.org