CRM-Based Customer Churn Prediction Using Machine Learning: A Customer 360 Analytics Approach

Author: Marisha Gavkar | Golden Gate University

Course Number: MSBA 326 | Instructor: Bahman Zohuri

Abstract

Customer churn is a critical challenge for subscription-based and recurring-revenue business models, directly impacting profitability, customer lifetime value, and strategic growth. This MBA capstone project presents an end-to-end, CRM-oriented analytics framework for predicting customer churn using machine learning. The study restructures raw operational data into CRM-style entities, constructs a unified Customer 360 analytical layer, and applies an interpretable predictive model to identify customers at high risk of churn. The project architecture mirrors real-world enterprise analytics environments such as Microsoft Dynamics 365 and Microsoft Fabric. Results indicate that contract type, customer tenure, pricing, service adoption, and support availability are significant predictors of churn. The findings demonstrate how machine learning can be operationalized within CRM systems to support proactive retention strategies and managerial decision-making.

Keywords: Customer Churn, CRM Analytics, Customer 360, Business Analytics, Machine Learning

---

1. Introduction

In competitive subscription-based industries, retaining existing customers is more cost-effective than acquiring new ones. Customer churn, defined as the termination or non-renewal of a service contract, poses a significant risk to revenue stability and long-term profitability. As organizations accumulate large volumes of customer data through Customer Relationship Management (CRM) systems, advanced analytics and machine learning have become essential tools for predicting churn and enabling proactive intervention.

The purpose of this MBA final project is to design and implement an industry-aligned churn prediction framework that integrates CRM data modeling with machine learning techniques. Rather than analyzing the dataset as a single flat file, the project decomposes operational data into CRM-style entities and reconstructs them into a Customer 360 analytical view. This approach reflects how modern organizations operationalize analytics using platforms such as Microsoft Dynamics 365, Dataverse, and Microsoft Fabric.

The objectives of this study are: - To model raw customer data into CRM-style entities - To construct a unified Customer 360 dataset for analytics - To develop and evaluate a machine learning model for churn prediction - To translate analytical results into actionable CRM strategies

---

2. Business Context and Problem Statement

Telecommunications and subscription-based service providers rely heavily on recurring revenue models, where profitability depends on customer retention and engagement over time. High churn rates increase acquisition costs, reduce lifetime value, and negatively affect operational planning. Traditional business intelligence reports typically identify churn after it has occurred, limiting their strategic value.

Problem Statement: How can a CRM-driven analytics framework leverage machine learning to predict customer churn in advance and support data-driven retention strategies?

From a managerial perspective, the challenge is not solely predictive accuracy but also interpretability. Decision-makers require insights that align with CRM levers such as pricing, contracts, feature adoption, and customer support in order to design effective interventions.

---

3. Dataset Description

This project uses the publicly available *Telco Customer Churn* dataset provided by IBM. The dataset contains customer-level information related to demographics, subscriptions, services, billing, and churn outcomes.

3.1 Variables

Key variables include: - **Customer Attributes:** Customer ID, tenure - **Subscription Details:** Contract type, monthly charges, total charges - **Service Adoption:** Internet services, streaming services, security features - **Billing Information:** Payment method, paperless billing - **Support Indicators:** Technical support availability - **Target Variable:** Churn (Yes/No)

Although originally designed for telecommunications analysis, the dataset closely resembles a generic subscription-based CRM system, making it suitable for broader business analytics applications.

---

4. Methodology

The project follows a modular analytics pipeline designed to mirror enterprise CRM and analytics architectures.

4.1 Data Understanding and Preparation

Initial data exploration focused on understanding variable definitions, identifying missing values, and resolving data quality issues. The *TotalCharges* variable required transformation from a text field to a numeric measure. Exploratory analysis confirmed expected churn patterns across tenure, contract type, and pricing variables.

4.2 CRM Entity Modeling

To align with real-world CRM design principles, the dataset was decomposed into the following entities:

- **Account Entity:** Customer tenure and lifetime value metrics
- **Subscription Entity:** Contract type, monthly recurring revenue, and billing preferences
- **Usage / Feature Adoption Entity:** Enabled services and feature breadth indicators
- **Support Entity:** Technical support availability and support risk flags
- **Churn Label Entity:** Customer renewal outcome

This decomposition demonstrates an understanding of data grain, entity relationships, and operational CRM modeling.

4.3 Customer 360 Construction

The individual CRM entities were joined at the account level to create a unified Customer 360 analytical dataset. This dataset represents the analytics-ready layer commonly used in enterprise platforms such as Microsoft Fabric and Customer Insights, separating operational data from downstream analytics and modeling.

4.4 Feature Engineering

Business logic was applied to translate CRM attributes into model-ready features, including: - Feature adoption counts - Contract and pricing risk indicators - Support dependency flags

This step illustrates how domain knowledge is transformed into predictive signals.

4.5 Machine Learning Model

A Logistic Regression model was selected as the baseline classifier due to its interpretability and suitability for binary classification problems. The dataset was split into training and validation sets, and categorical variables were encoded using standard techniques. Model performance was evaluated using accuracy, precision, recall, and ROC-AUC metrics.

---

5. Results and Analysis

The churn prediction model demonstrated strong performance, particularly in identifying customers likely to churn. High recall for churned customers is critical in retention-focused use cases, where missing at-risk customers can lead to revenue loss.

5.1 Key Drivers of Churn

The most influential predictors of churn included: - **Contract Type:** Month-to-month contracts exhibited significantly higher churn risk than long-term contracts - **Customer Tenure:** Newer customers were more likely to churn than long-tenured customers - **Monthly Charges:** Higher monthly charges were associated with increased churn probability - **Service Adoption:** Customers with fewer enabled services showed higher churn risk - **Technical Support:** Lack of technical support availability increased churn likelihood

These findings are consistent with established customer behavior and CRM theory.

---

6. Managerial Implications

The results of this study provide several actionable insights for CRM and business leaders:

- **Contract Strategy:** Encourage migration from month-to-month contracts to longer-term agreements through targeted incentives
- **Onboarding Programs:** Focus retention efforts on early-tenure customers through education and engagement
- **Product Bundling:** Increase feature adoption to raise switching costs and perceived value
- **Proactive Support:** Offer technical support outreach to customers identified as high risk

By embedding churn predictions into CRM workflows, organizations can shift from reactive reporting to proactive customer management.

---

7. Limitations and Future Research

Despite its strengths, the study has several limitations: - The dataset represents a single industry context - Only a baseline machine learning model was implemented - Behavioral and interaction data were not available

Future research could incorporate advanced models, real-time data integration, and deployment of the model as an API within CRM platforms.

---

8. Conclusion

This MBA capstone project demonstrates how CRM principles, data engineering, and machine learning can be integrated into a practical churn prediction framework. By restructuring raw operational data into CRM-style entities and constructing a Customer 360 analytical layer, the project reflects real-world enterprise analytics practices. The findings provide both predictive insights and clear managerial actions, fulfilling the academic and practical objectives of a graduate-level business analytics project.

References (APA 7th Edition)

IBM. (2019). *Telco customer churn dataset*. IBM Developer. https://www.ibm.com/

Microsoft. (2023). *Dynamics 365 customer insights documentation*. Microsoft Learn. https://learn.microsoft.com/

Provost, F., & Fawcett, T. (2013). *Data science for business: What you need to know about data mining and data-analytic thinking*. O'Reilly Media.