

Are You Still There?

Predicting Churn of Fictitious Telecommunication Customers

Jett Palmer

Introduction

Churn is the **number of customer a company loses** over a given period of time. Understanding churn allows businesses to **more accurately forecast revenue** and identify factors that predict customer **retention**. This study analyzes customer service records from a fictitious telecommunications company to investigate which customer profiles are most strongly associated with churn and **how retention resources could be allocated** to most effectively reduce it. Online connectivity is deeply embedded in contemporary social and cultural life, where owning a cell phone and having internet access has become the prevailing norm, making telecommunications services an increasingly essential part of participation in modern society.

Methods

- Minimal cleaning was performed (i.e., type casting, level re-coding).
- All available **customer account, billing, and service subscription variables** were included to maximize predictive performance while allowing analysis of customer profiles that were at risk.
- Logistic regression and decision trees** were compared using cross-validation. Logistic regression demonstrated **marginally superior performance** (ROC-AUC = 0.845 vs. 0.83).
- ROC-AUC was used to assess performance as the data are **moderately imbalanced** (26.5% churn), and the cost of false positives and false negatives is similar in this context.
- Logistic coefficients were used to identify the **strongest drivers of churn** and to construct **high-risk and low-risk customer profiles**.

Results

- Logistic regression achieved **strong predictive performance**. The model assigns a **higher churn probability** to a randomly chosen churned customer **84.5% of the time**, compared to a randomly chosen retained customer (*Figure 1*).
- Customers were most likely to churn when they had **higher total charges, month-to-month contracts, no tech support plan, and electronic check payments**.
- Customers were most likely to be retained when they had **long tenure, two-year contracts, DSL internet service, phone service bundles, and paper billing**.
- Our risk classification identified approximately **4.8% of customers** as both **high risk** (by prob. of churning) and **high value** (by billings) (*Figure 2*).
- If **outreach costs \$50 per customer** and **reduces churn probability by 20%**, retention efforts (e.g., offering **phone service bundles**) could generate a positive expected net benefit for **35% of customers**.

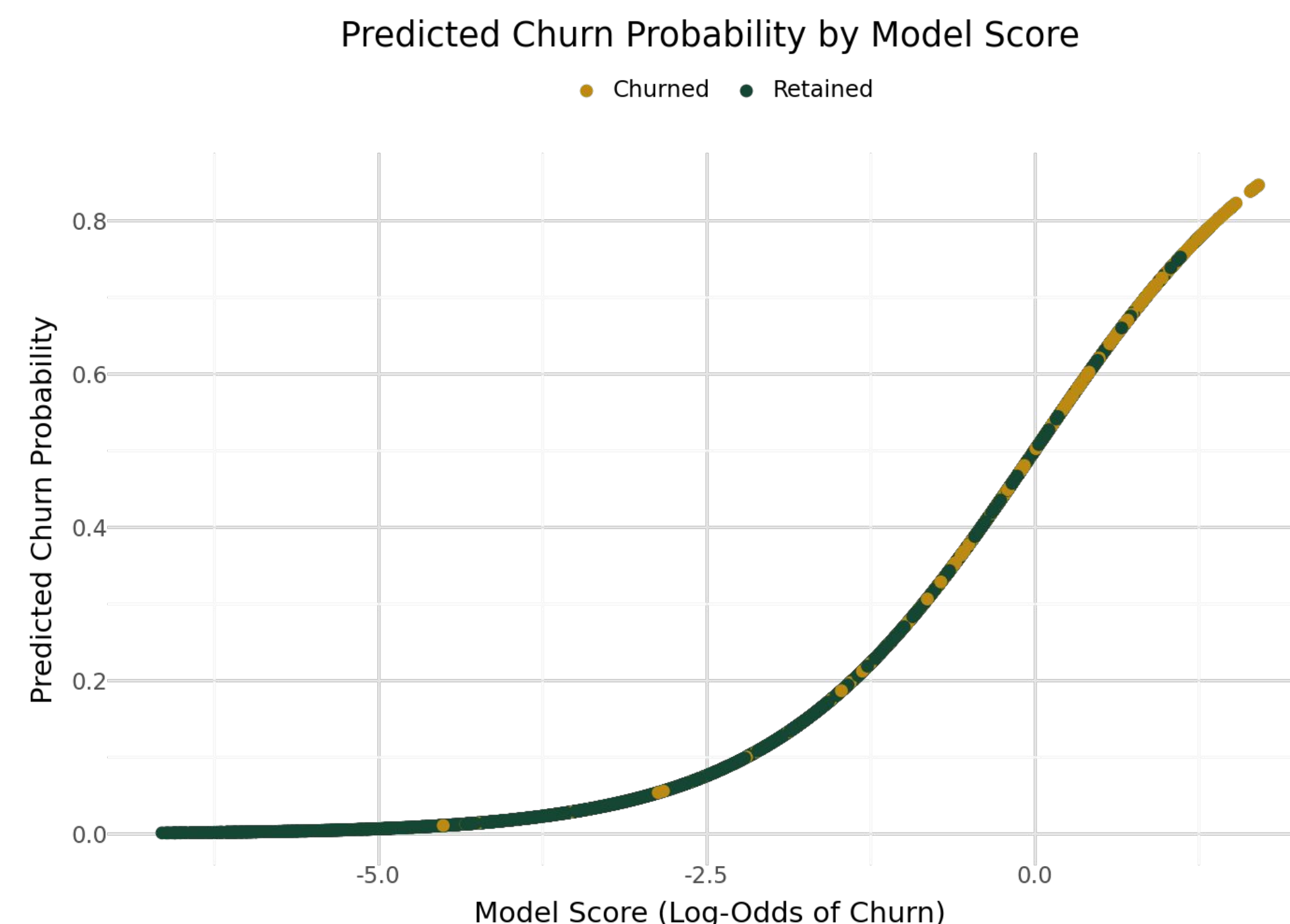


Figure 1: Logistic Curve of Predicted Probabilities vs. Log Odds of Churn

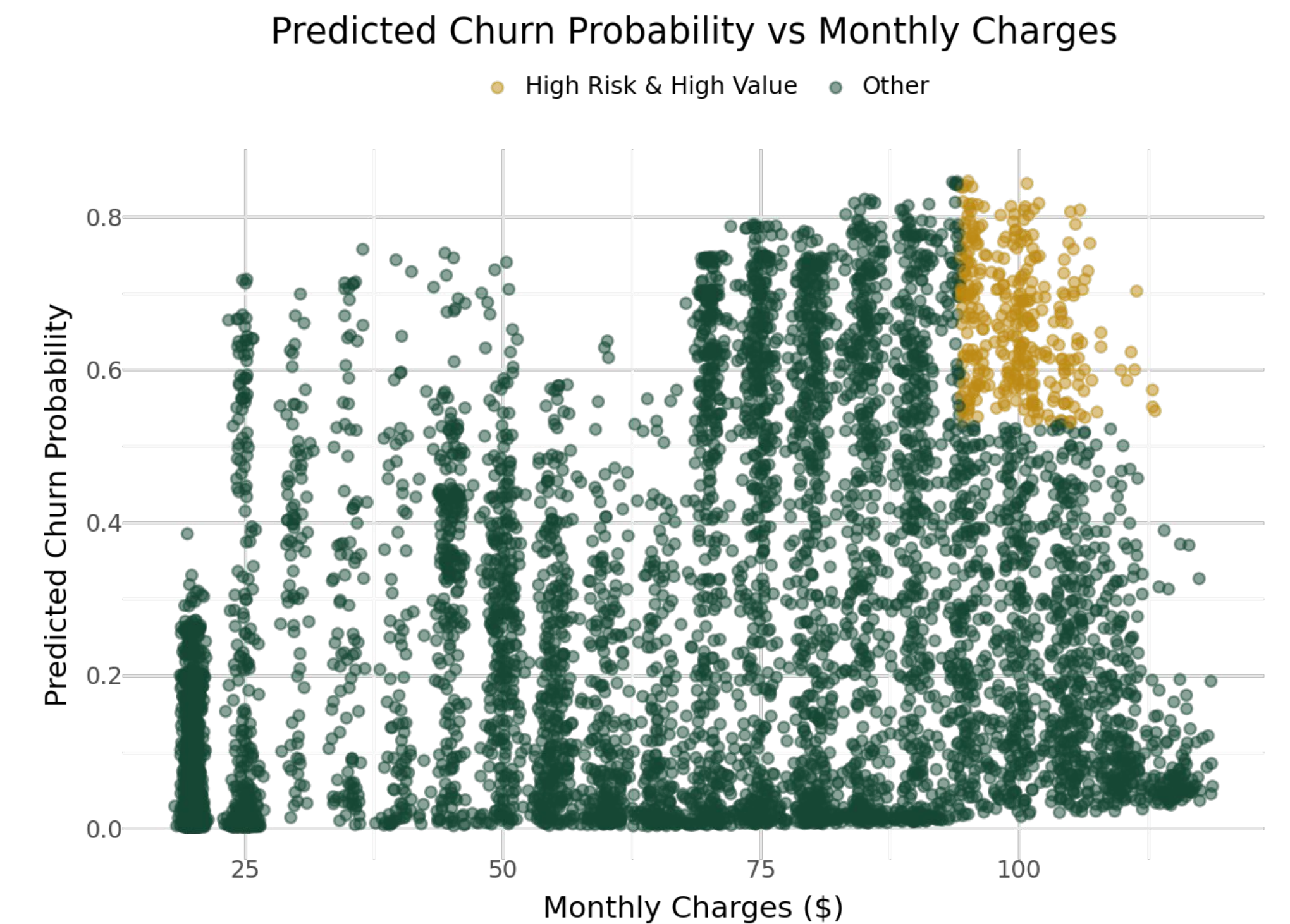


Figure 2: Gold Represents the 80th Percentile of Billings & Churn Probability

Conclusion & Limitations

- The dataset used in this study is **synthetic**, created as a sample dataset for IBM's Cognos Analytics software, and we were unable to confirm whether it reflects patterns from a real telecommunications customer base.
- The primary contribution of this analysis is **methodological** rather than empirical. Specifically, we demonstrate how **churn-risk feature engineering** can help organizations **allocate retention resources** more strategically.
- Telecommunications access is closely tied to participation in modern social and economic life. Even **without explicit demographic variables**, institutional and socioeconomic patterns may be indirectly captured in **service usage or billing variables**, meaning retention strategies informed by these models could unintentionally and **disproportionately** prioritize or neglect certain groups of customers.