

Unveiling Customer Churn: Harnessing the Power of Supervised Machine Learning in Telecom Services

***Mr. Chinmaya Mishra**

****Prof. (Dr.) Manoranjan Dash**

*****Dr Alaka Samantaray**

ABSTRACT

Customer retention is a vital requirement for organizations, to recognize the significance of customers as their most valuable assets. This principle holds true for telecom companies operating in a competitive environment. The advent of machine learning, leveraging existing information technologies and extensive data, presents a powerful tool for extracting valuable insights. In this study, we employed the binary logistic regression technique to develop a predictive model that incorporates this knowledge. The outcomes reveal distinctive characteristics of churned customers, enabling retail managers to proactively identify potential churners. Armed with this classification, managers can devise targeted strategies for customers displaying similar features, aiming to prevent churn.

Keywords: Binary Logistics, Churn Prediction, Supervised Machine Learning

Introduction

In the fiercely competitive telecommunications sector, customer churn, where customers terminate their association with a company, presents a formidable obstacle for businesses. Predicting and comprehending customer churn can yield valuable insights for telecom companies, empowering them to devise effective retention strategies and enhance customer satisfaction. The telecom industry, characterized by cutthroat competition, rapid technological advancements, and a diverse range of services, constantly challenges telecom companies to retain their

**Mr. Chinmaya Mishra, Faculty of Management Sciences, Siksha 'O' Anusandhan (Deemed to be University), Bhubaneswar, Email: chinmaya_mishra123@yahoo.co.in*

***Prof. (Dr.) Manoranjan Dash, Faculty of Management Sciences, Siksha 'O' Anusandhan (Deemed to be University), Bhubaneswar, Email: manoranjandash@soa.ac.in*

****Dr. Alaka Samantaray, Faculty of Management Sciences, Siksha 'O' Anusandhan (Deemed to be University), Bhubaneswar, Email: alakasamantaray@soa.ac.in*

customer base and mitigate churn. Customer churn not only results in revenue loss but also entails substantial costs associated with customer acquisition and retention endeavors. In the telecommunications sector, conventional approaches to analyzing customer churn have typically involved descriptive statistics and basic metrics. However, the landscape has evolved with the advent of data analytics and machine learning. These advancements have given rise to more sophisticated techniques, such as binary logistic regression, which serve as potent tools for predicting churn and unveiling the underlying factors driving it. Customer churn refers to the phenomenon where customers switch from one company to another. Apart from causing a loss in income, churn also has various negative effects on a company's operations like churn management is the process of identifying customers who intend to switch to a competing service provider and now this is an integral part of customer relationship management, considering its importance in establishing long-term relationships and maximizing the value of the customer base. This study aims to investigate customer churn in the telecommunications sector by employing a binary logistic regression model. Logistic regression is a widely used statistical technique for modeling binary outcomes, making it suitable for predicting whether a customer will churn or not. Through the utilization of this model, the study seeks to uncover the key factors that influence customer churn. By providing actionable insights, telecom companies can effectively mitigate churn rates and enhance customer retention strategies. In a binary logistic regression, the relationship between a set of independent variables (such as customer demographics, service usage, and contract details) and a binary outcome variable (churned or not churned) is modeled. By estimating the probabilities of churn based on these variables, telecom companies can gain valuable insights into the factors that significantly influence customer attrition. With this information, they can tailor their strategies to effectively address churn. The outcomes of this study will contribute to the existing knowledge of customer churn analysis within the telecommunications sector. The findings will assist telecom companies in making data-driven decisions aimed at reducing churn rates, enhancing customer loyalty, and improving overall business performance. By employing a robust analytical approach and harnessing the power

of binary logistic regression, this study aims to provide a comprehensive understanding of customer churn dynamics in the telecommunications industry. Ultimately, this research endeavor aims to enable telecom companies to develop improved customer retention strategies, giving them a competitive edge in the market.

Literature Review

Keiningham et al.,(2007), in their study, explored the relationship between customer satisfaction, loyalty (measured by Net Promoter Score), and revenue growth in the telecommunications sector. It highlights the importance of customer loyalty in driving business growth and emphasizes the need to focus on customer retention to minimize churn rates. Verhoef et al ., (2002) examined the impact of customer relationships on referrals and service adoption, this study provides insights into the role of customer satisfaction, trust, and commitment in reducing churn rates. It emphasizes the importance of building strong customer relationships to foster loyalty and reduce the likelihood of churn. Van et al.,(2004) et al., .While focusing on the financial services industry, this study demonstrates the application of proportional hazard models in predicting customer attrition. The research highlights the relevance of various factors such as customer demographics, transactional history, and service usage patterns in identifying customers at high risk of churn. Liu et al.,(2014) employed data mining techniques to identify the key factors contributing to customer churn in the telecommunications industry. The study emphasizes the importance of customer demographics, contact details, customer service interactions, and usage patterns in predicting churn, providing valuable insights for customer retention strategies. Wang et al.,(2016) study focused on applying survival analysis modeling, specifically the Cox proportional hazards model, to predict customer churn in the telecommunications sector. The research emphasizes the significance of contract-related variables, customer behavior, and service usage patterns in predicting churn probabilities. Kumar et al., (2010) study focused on the concept of customer engagement and its impact on customer churn. It highlights the importance of measuring and understanding the overall customer

engagement value to identify customers who are undervalued or overvalued. By recognizing the different levels of engagement, telecom companies can design appropriate retention strategies for each customer segment. Reichheld (2003), This influential article introduces the concept of the Net Promoter Score (NPS) as a metric for customer loyalty. It discusses the correlation between NPS, customer satisfaction, and business growth. By analyzing NPS scores, telecom companies can identify potential churners and implement strategies to improve customer loyalty. Verhoef et al.,(2007) explored the multichannel nature of customer interactions and its impact on customer churn. It emphasizes the need to understand the research-shopper phenomenon, where customers gather information from multiple channels before making a purchase decision. Telecom companies can leverage this understanding to optimize their multichannel strategies and reduce churn rates. Hwang et al.,(2004) focused on customer churn prediction in the Korean mobile telecommunications industry. It examines the effectiveness of data mining techniques, including decision trees, neural networks, and logistic regression, in predicting churns. The research identifies key factors such as call patterns, service usage, and customer demographics that significantly impact churn rates. Li et al.,(2017) explored the application of ensemble learning techniques, including random forests and gradient boosting, for customer churn prediction in the telecommunications industry. The study highlights the advantages of combining multiple models to improve predictive accuracy and identifies relevant features for churn prediction, such as customer behavior and service usage patterns.

Model Building

Customer churn analysis in the telecommunications sector using a binary logistic regression model can be an effective approach to predict and understand customer churn.

- ❖ **Data Collection:** Gather relevant data about your customers, including demographic information, usage patterns, customer service interactions, and historical churn data. Ensure that you have a binary churn variable indicating whether each customer has churned or not.

- ❖ **Data Pre-processing:** Clean and pre-process the data. Handle missing values, remove duplicates, and transform categorical variables into numeric representations (e.g., one-hot encoding).
- ❖ **Feature Selection:** Identify the most relevant features that are likely to impact churn. This can be done through various techniques such as statistical tests, correlation analysis, or domain knowledge. Consider including variables like customer demographics, service usage, contact details, payment history, and customer interactions.
- ❖ **Train-Test Split:** Split your dataset into a training set and a testing set. The training set will be used to train the logistic regression model, while the testing set will be used to evaluate its performance.
- ❖ **Model Training:** Fit a binary logistic regression model to the training data. In logistic regression, the dependent variable is binary (churned or not churned), and the independent variables are the selected features. This model will estimate the probability of churn based on the input features.
- ❖ **Model Evaluation:** Evaluate the performance of your model using the testing set. Common evaluation metrics for binary classification models include accuracy, precision, recall, F1 score, and the receiver operating characteristic (ROC) curve.
- ❖ **Interpretation of Results:** Analyze the coefficients of the logistic regression model to understand the impact of each feature on the likelihood of churn. Positive coefficients indicate a positive relationship with churn, while negative coefficients indicate a negative relationship.
- ❖ **Predictions:** Once the model is trained and evaluated, you can use it to make predictions on new, unseen data. Provide the input features of a customer to the trained model, and it will output the probability of churn.
- ❖ **Customer Segmentation:** Segment your customers based on their predicted churn probabilities. This will help you prioritize your retention efforts by focusing on customers at high risk of churn.
- ❖ **Retention Strategies:** Based on the insights gained from the model and customer segmentation, develop targeted retention strategies to reduce

churn. These strategies could include personalized offers, proactive customer support, loyalty programs, or service improvements.

Key Factors for Customer Churn

Subscribers churn analysis varies across countries and cultures, leading to potential variations in the pivotal factors that play an important role in this analysis. Therefore, this paper aims to compile factors identified by scholars that impact customer relationships and supplement them with insights from experts within the same country. One key factor that has a direct impact on customers is service quality. Service providers are keenly aware of its significance as a critical determinant of customer ecstasy, as highlighted by Mannan et al. (2017). “Service quality was positively and significantly related to customer loyalty”. Bhuian et al. (2018) found a direct and significant relationship between service quality and customer loyalty, while Adebiyi et al. (2015) noted that call quality is an important factor influencing customer retention decisions. Promotion is one of the factors emphasized by researchers. Effective promotional strategies have been shown to increase mind awareness about the brand, customer retention, and company profitability. When customers are satisfied with the promotions offered by their service providers, the likelihood of churn decreases. “Factors that significantly contribute to the prediction of customer churning intention are; tariffs, transparency level, promotions, technical assistance, privacy, and response to complaints” (Hamelin et al, 2009). Therefore, this study will incorporate the following factors to analyze customer churn in the telecommunications sector, considering their relevance and impact: service quality, Usage and revenue, technical assistance, and response to complaints.

- ❖ Customer Data
- ❖ Usage and Revenue Data
- ❖ Customer Care Data as per SLA
- ❖ Network Data

Like the other business companies, Telecommunication also has millions of customers. This implies the necessity of maintaining the database that contains the information about these customers. This database contains personal information like name, address, age, gender, income, and service-oriented information like plan, contact information, credit score, payment history, etc, Usage and Revenue Data ensure whenever a call is made on the telecommunication network, detailed record of a call is stored with certain descriptive information. Due to the increase in customers, detailed data of the call has a huge number of records. Each record in the call detail data defines the characteristics of the call (Qining Lin., 2007). It contains the basic information like-:

- I. Out Going Call Minutes
- II. Incoming Call Minutes
- III. Data Downloading Quantity in MBS
- IV. Prepaid recurring revenue

- ❖ **Customer Data:** Similar to other businesses, the telecommunications industry serves millions of customers, necessitating the maintenance of a comprehensive customer database. This database contains personal information such as name, address, age, gender, income, and service-oriented details including plans, contact information, credit scores, payment history, and more.
- ❖ **Usage and Revenue Data:** For each call made within the telecommunications network, a call detail record is generated, resulting in a substantial number of records due to the large customer base. These call detail records capture various call characteristics, including outgoing call minutes, incoming call minutes, data downloading quantities in megabytes (MBs), prepaid recurring revenue, postpaid billing amounts, and the utilization of value-added services.
- ❖ **Customer Care Data:** To meet service level agreements and ensure customer satisfaction, telecom companies maintain call center databases that gather valuable information. This includes the total number of calls made by

a customer during a specific period, the percentage of calls categorized as queries or requests for services, the average query handling time by customer care executives, and the number of repeat calls made for the same issue.

- ❖ **Network Data:** Telecommunication networks are complex systems composed of interconnected components, generating a vast amount of error and status messages. Recording and analyzing network data is crucial for efficient network management. Network data encompasses percentages of dropped calls, percentages of calls successfully connected on the first attempt, average data downloading and uploading speeds in megabits per second (Mbps), location-specific busy and non-busy hours, as well as average network uptime and downtime.

The data analysis process in this study is divided into several key stages. It begins with data collection, where information is gathered from the telecom company's database. The collected data comprises various types of variables to find as many factors as possible that might influence customer churn.

Next is the data display stage, where the factors considered in the prediction of customer churn in the telecommunications sector are summarized and presented using tables. This visual representation helps in understanding and interpreting the data effectively. The focus of this paper is customer churn prediction utilizing the binary logistic regression model.

Data Analysis and Results

The following table shows the Cases included in the analysis, we have a total of 7043 cases, and the dependent variable in this is those who will churn are classified as 0, and those who will not churn are classified as 1.

Table-1

Case Processing Summary			
Unweighted Cases ^a		N	Percent
Selected Cases	Included in Analysis	7043	100.0
	Missing Cases	0	.0
	Total	7043	100.0

Unselected Cases	0	.0
Total	7043	100.0
a. If weight is in effect, see classification table for the total number of cases.		

Table-1.1

Dependent Variable Encoding	
Original Value	Internal Value
Yes	0
No	1

Table-2

Omnibus Tests of Model Coefficients				
		Chi-square	df	Sig.
Step 1	Step	8150.146	9	.000
	Block	8150.146	9	.000
	Model	8150.146	9	.000

Table-3

Hosmer and Lemeshow Test			
Step	Chi-square	df	Sig.
1	1.341	8	.967

Omnibus Tests of Model Coefficients were used to test the model fit. The model shows a good fit. The Hosmer and Lemeshow test indicates a poor fit if the sig value is less than 0.05. Here the model adequately fits the data.

Table-4

Contingency Table for Hosmer and Lemeshow Test						
		Churn = Yes		Churn = No		Total
		Observed	Expected	Observed	Expected	
Step 1	1	704	704.000	0	.000	704
	2	704	704.000	0	.000	704
	3	461	461.000	243	243.000	704
	4	0	.000	704	704.000	704
	5	0	.000	705	705.000	705
	6	0	.000	704	704.000	704
	7	0	.000	704	704.000	704
	8	0	.000	704	704.000	704
	9	0	.000	704	704.000	704
	10	0	.000	706	706.000	706

Table-5

Classification Table ^a					
Observed			Predicted		
			Churn		Percentage Correct
			Yes	No	
Step 1	Churn	Yes	1869	0	100.0
		No	0	5174	100.0
	Overall Percentage				100.0
a. The cut value is .500					

Table-6

Model Summary			
Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	.000 ^a	.686	1.000
a. Estimation terminated at iteration number 20 because maximum iterations have been reached. Final solution cannot be found.			

Nagelkerke's R² is an adjusted version of the Cox & Snell R-Square that adjusts scale of the statistics to cover the full range from 0 to 1. The model correctly classified 100% of the cases overall referred to as the percentage accuracy in classification. It is the rate of correct classification. it indicated the degree to which the observed outcomes are predicted by our model.

Table-7**Variables in the Equation**

		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I. for EXP(B)	
								Lower	Upper
Step 1 ^a	TOM	.000	3.292	.000	1	1.000	1.000	.002	634.610
	TOI	.001	6.614	.000	1	1.000	1.001	.000	426410.691
	DaM	.000	1.936	.000	1	1.000	1.000	.022	44.435
	DaG	.231	1981.078	.000	1	1.000	1.260	.000	.
	Noc	-.011	200.674	.000	1	1.000	.989	.000	6.449E+170
	CC	-.278	2843.129	.000	1	1.000	.757	.000	.
	ADD	12.523	610.170	.000	1	.984	274698.927	.000	.

CD	1536.538	68040.541	.001	1	.982	.000	.000	.
CM	4767.566	258178.335	.000	1	.985	.000	.000	.
Constant	4255.092	229653.297	.000	1	.985	.		
a. Variable(s) entered on step 1: TOM, TOI, DaM, DaG, Noc, CC, ADD, CD, CM.								

The percentage in the first two row provided the information regarding specificity and sensitivity of the model in terms of predicting the churn. Specificity also called as True Negative Rate is 100%. Sensitivity also called True Positive Rate was 100%

Conclusion

Managers in service companies place significant emphasis on building strong customer relationships. When the relationship between customers and the service provider is positive, it fosters loyalty, which directly impacts the company's profitability. To ensure sustained success, companies must try to find out the reasons behind customer churn and proactively work to reduce churn rates. In this paper, we proposed a model for customer churn analysis utilizing the Binary Logistic Regression Model, a well-known statistical methodology. The purpose of this model is to provide companies with a clear idea of how they can analyze their data, identify customers with a higher probability of churning, and determine the key factors influencing this analysis. Armed with this knowledge, companies can closely watch these factors and develop strategies to decrease churn probability effectively by employing this model, we can delve into the daily dynamics of customer behavior. This enables the service provider to proactively take appropriate actions to retain customers before their dissatisfaction reaches a level where they voluntarily decide to churn. Through the application of this algorithm, companies can gain valuable insights and implement targeted measures to improve customer retention and satisfaction.

References

- Keiningham, T. L., Cool, B., Aksoy, L., Andreassen, T. W., & Williams, L. "A longitudinal examination of net promoter and firm revenue growth". *Journal of Marketing*, (2007),71(3), 39-51.
- Verhoef, P. C., Franses, P. H., & Hoekstra, J. C. "The effect of relational constructs on customer referrals and number of services purchased from a multiservice provider: Does age of relationship matter?" *Journal of the Academy of Marketing Science*, (2002), 30(3), 202-216.
- Van den Poel, D., & Larivière, B. (2004). Customer attrition analysis for financial services using proportional hazard models. *European Journal of Operational Research*, 157(1), 196-217
- Liu, D. R., Chen, K. Y., & Huang, M. C. "Applying data mining techniques to analyze factors influencing customer churn in the telecommunications industry. *International Journal of Electronic Business Management*", (2014),12(4), 311-322.
- Wang, W., Zhang, Y., & Ouyang, M. "Predicting customer churn in the telecommunications industry—An application of survival analysis modeling". *Expert Systems with Applications*, (2016),46, 22-33.
- Kumar, V., Aksoy, L., Donkers, B., Venkatesan, R., Wiesel, T., & Tillmanns, S. "Undervalued or Overvalued Customers: Capturing Total Customer Engagement Value". *Journal of Service Research*, (2010) 13(3), 297-310.
- Reichheld, F. F. "The one number you need to grow". *Harvard Business Review*, (2003),81(12), 46-54.
- Verhoef, P. C., Neslin, S. A., & Vroomen, B. "Multi-channel customer management: Understanding the research-shopper phenomenon". *International Journal of Research in Marketing*, (2007),24(2), 129-148.
- Hwang, H., Jung, T., & Suh, E. "Analysis of customer churn prediction in the Korean mobile telecommunication service industry". *Expert Systems with Applications*, (2004),26(2), 241-250.
- Li, X., Wang, W., & Sun, P. "Customer churn prediction in the telecommunications industry using ensemble learning. *Neurocomputing*", (2017),228, 254-261.
- Verbeke W, Martens D, Mues C, Baesens B "Building comprehensible customer churn prediction models with advanced rule induction techniques". *Expert Syst Appl* (2011) 38:2354–2364
- Tsai C-F, Lu Y-H "Customer churn prediction by hybrid neural networks". *Expert Syst Appl* (2009) ,36:12547–12553.

- Sharma A, Panigrahi PK “A neural network based approach for predicting customer churn in cellular network services”. *Int J Comput Appl* (2011) ,27(11):26–31
- Chen K, Hu Y-H, Hsieh Y-C “Predicting customer churn from valuable B2B customers in the logistics industry: a case study”. (2014) *IseB* 13:475–494. doi:10.1007/s10257-014-0264-1
- Hadden J, Tiwaria A, Roy R, Ruta D “Computer assisted customer churn management: State-of-the-art and future trends”. *Comput Oper Res* (2005),34:2902–2917
- Risselada H, Verhoef PC, Bijmolt THA,” Staying power of churn prediction models. *J Interact Mark* (2010), 24:198–208
- Mannan, M., Mohiuddin, M.F., Chowdhury, N. and Sarker, P. “Customer satisfaction, switching intentions, perceived switching costs, and perceived alternative attractiveness in Bangladesh mobile telecommunications market”. *South Asian Journal of Business Studies*, .(2017) 6(2), 142-160
- Bhuiyan, S.N., Al-Balushi, M. and Butt, I. “Antecedents and consequences of customer loyalty in Qatar”. *Journal for Global Business Advancement*, (2018) 11(1), pp.41-63.
- Adebisi, S.O., Oyatoye, E.O. and Kuye, O.L. “An Analytic Hierarchy Process Analysis: Application to Subscriber Retention Decisions in the Nigerian Mobile Telecommunications”. *International Journal of Management and Economics*, .. (2015),48(1), 63-83.
- Hamelin, N., Nassali, A., & Hancar, T. “Determining churn drivers in moroccan telecom sector. *Research Yearbook*”, (2009). 573.
- Qining Lin, “Mobile Customer Clustering Analysis Based on Call Detail Records, *Communications of the IIMA*”, (2007). 7 (4), 95-100.
- Darji, H.H “Data Mining in Telecommunication Industry”, *International Journal for Scientific Research & Development*”,(2014). 2(8),7-9.