International Journal of Trend in Scientific Research and Development (IJTSRD)
Special Issue on Advancements and Emerging Trends in Computer Applications - Innovations, Challenges, and Future Prospects
Available Online: www.ijtsrd.com e-ISSN: 2456 – 6470

Customer Churn Analysis in Banking Sector using Data Mining

Saurabh Shirpurkar

Department of Computer Application, G. H. Raisoni University, Amravati, Maharashtra, India

ABSTRACT

The banking business is constantly attempting to recruit and keep consumers, despite the continuous difficulty of customer churn. This problem not only causes huge financial losses, but it also reduces overall consumer satisfaction, creating a negative cycle for financial institutions. Understanding the factors that influence a client's choice to leave is critical for banks since it allows them to adopt focused initiatives to increase customer loyalty and service offerings. Knowing the root causes of churn enables banks to remedy customer grievances, customize services for changing needs, and ultimately create a more enjoyable banking experience.

This knowledge is necessary for designing effective retention techniques that minimize churn while also advancing long-term profitability and customer activity. The banking sector is perpetually plagued with customer churn, which compromises financial stability and degrades customer satisfaction. In spite of continuous endeavors to retain clients and attract them, numerous banks incur heavy losses because customers exit due to diverse reasons, like poor service quality or superior substitutes.

Within a competitive market, banks that focus on churn analysis and modelling will be in a superior position to accommodate changing customer requirements and retain a solid market standing, ultimately promoting sustainable growth and increased customer satisfaction.

KEYWORDS: *PYTHON, ML, DEEP LEARNING, TENSIOR FLOW, POWER BI, NO SQL etc.*

I. INTRODUCTION

Poverty is increasing, and the rate of unemployed persons in urban areas climbed much higher, from 8.8 percent in 2019 to 20.8 percent as by 2020. During the initial lockdown, the microfinance sector faced considerable distress that caused disbursements to drop from Rs 711.9 billion to just Rs 61.9 billion amid a portfolio risk of 22 percent by June 2021. The institutions thus used to arbitrarily provide repayment postponements and new loans while the Reserve Bank of India made lending moratoriums to stabilize the industry.

The Regal Finance Solution initiative is concerned with knowing why customers leave their banks and what they can do to retain them happy and committed. With digital banking the rage in today's times and the intense competition around, banks need to know what drives consumers away-be it bad service, more attractive offers at other banks, or changes in their lifestyle.

Banks are able to spot trends and patterns in customer behavior through the analysis of customer data through advanced analytics. This allows them to meet challenges head-on and adapt their services to suit the needs of their customers more effectively. The aim of this project is to provide real-world insights that will allow banks to enhance customer relationships, enhance satisfaction, and minimize the cost of acquiring new consumers.

Abbreviations acronyms :-

1. **Churn** - Customer Attrition; the loss of clients or customers over a specific period.
2. **CLV** - Customer Lifetime Value; the total revenue a business can expect from a customer over their relationship.
3. **Cohort Analysis -** A method of analyzing the behavior of groups of customers over time based on shared characteristics.
4. **LTV** - Lifetime Value; a prediction of the net profit attributed to the entire future relationship with a customer.
5. **NPS** - Net Promoter Score; a metric used to gauge customer loyalty and satisfaction.
6. **RFM** - Recency, Frequency, Monetary analysis; a marketing analysis tool used to identify a firm's best customers.

II. RELATED WORK

Analyzing customer churn in banking: A data mining framework: The paper focuses on analyzing customer churn in the banking industry using data mining and predictive analytics techniques. The study employs a Gaussian mixture model clustering-based adaptive support vector machine (GMM-ASVM) to forecast customer loss. By analyzing consumer competency and loyalty, the study predicts customer behaviour using a clustering approach, achieving an accuracy of 98% in classification.[1]

A Novel Approach to Predicting Customer Lifetime Value in B2B SaaS Companies (2023): The authors propose a machine learning framework to predict CLV in the B2B Software-as-a-Service context, addressing challenges like heterogeneous populations and multiple product offerings.[2]

Predicting Customer Churn in Banking Industry Using Neural Networks Published in: Interdisciplinary Description of Complex Systems, 2016.[7]

Segmentation of Bank Customers by Loyalty and Switching Intentions: This study investigates customer switching behaviour in Indian private banking, identifying factors such as service quality, satisfaction, trust, and switching barriers that influence loyalty and switching intentions.[9]

These studies provide insights into various machine learning approaches for predicting customer churn in the banking sector.

III. DATA AND SOURCE OF DATA

In terms of banking, customer churn analysis refers to the process of studying various data segments that explain the reasons behind customers' disengagement. **Kaggle** is a good platform for this type of data since it has a wide array of datasets like the "Bank Customer Churn" dataset. This sets is well endowed with customer information including demographic as well as transactional information and how they use their accounts. For instance, it contains information such as a customer's age, how much money is in the account, how active the customer is in transactions and if they had churned before. With such information at their disposal, banks can predict their customers' behavior and know what situations will force them to churn and this puts banks in a position to take proactive actions for customer retention.

To predict churn, banks often use logistic regression, which can be represented by the following equation:

$$P(Y=1) = \frac{1}{1 + e^{-(\beta_0 + \beta_1X_1 + \beta_2X_2 + \ldots + \beta_nX_n)}}$$

Where:
- ($P(Y=1)$) is the probability of churn (customer leaving).
- (e) is the base of the natural logarithm.
- (β_0) is the intercept.
- ($\beta_1, \beta_2, \ldots, \beta_n$) are the coefficients for each predictor variable.
- ($X_1, X_2, \ldots, X_n$) are the independent variables (e.g., age, account balance, transaction frequency).

IV. RESEARCH METHODOLOGY

In banks, the process of research on customer churn analysis has specific steps which are important in a sequence. First, the issue is defined, goals set, and churn drivers are identified using the developed predictive models. A search through the literature is done with the objective to collect and direct important information that will assist the analysis. Internal **CRM** systems and accessible databases like the Kaggle Bank Customer Churn dataset provide the information that is sought after, which is specifically customer data along with churn markers. Prior to **EDA**, the information is cleansed and preprocessed. Trends are then visualized using **EDA**. Subsequently, the data is used to construct and specify the needed model that is usually a logistic regression or a decision tree. Most churn drivers are identified using recommendations and strategized marketing improve customer service. Stakeholders are then presented with the documented process and progress to showcase valuable customer behavior insights and retention strategies banks can utilize. Follow up on performance is conducted through these measures, as customers experience better service.

Exploratory Data Analysis:

1. Visualizing Target Variable

```
count = df["Churned"].value_counts()

plt.figure(figsize=(14,6))
plt.subplot(1,2,1)
ax=sns.countplot(df["Churned"],palette="Set2")
ax.bar_label(ax.containers[0],fontweight="black",size=15)
plt.title("Customer Churned Disribution",fontweight="black",size=20,pad=20)

plt.subplot(1,2,2)
plt.pie(count.values, labels=count.index, autopct="%1.1f%%",colors=sns.set_palette("Set2"),
        textprops={"fontweight":"black"},explode=[0,0.1])
plt.title("Customer Churned Disribution",fontweight="black",size=20,pad=20)
plt.show()
```

Fig 1. code of visualizing target variable.

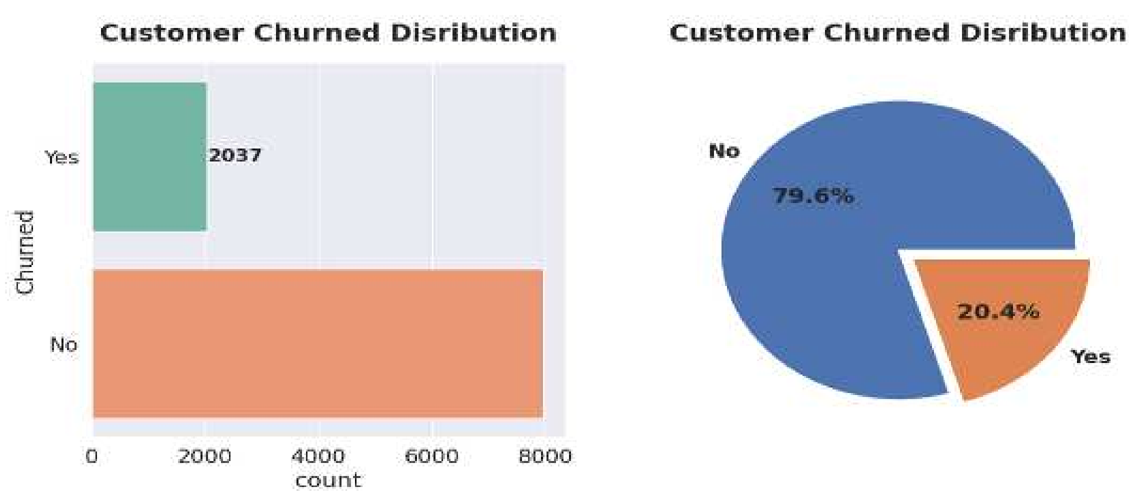

Fig 2. Graphical representation of visualizing target variable.

Inference:
- There is huge class-imbalance which can lead to bias in model performance.
- So to overcome this class-imbalance we have to use over-sampling technique from SMOTE.

Visualizing Customer Churned by "Credit Score".

```
def continous_plot(column):
    plt.figure(figsize=(13, 6))

    # Histogram with KDE
    plt.subplot(1, 2, 1)
    sns.histplot(x=column, hue="Churned", data=df, kde=True, palette="Set2")
    plt.title(f"Distribution of {column} by Churn Status", fontweight="black", pad=20, size=15)

    # Boxplot
    plt.subplot(1, 2, 2)
    sns.boxplot(x="Churned", y=column, data=df, palette="Set2")
    plt.title(f"Distribution of {column} by Churn Status", fontweight="black", pad=20, size=15)

    plt.tight_layout()
    plt.show()

# Call the function
continous_plot("CreditScore")
```

Fig 3. Code of Customer Churned by "Credit Score".

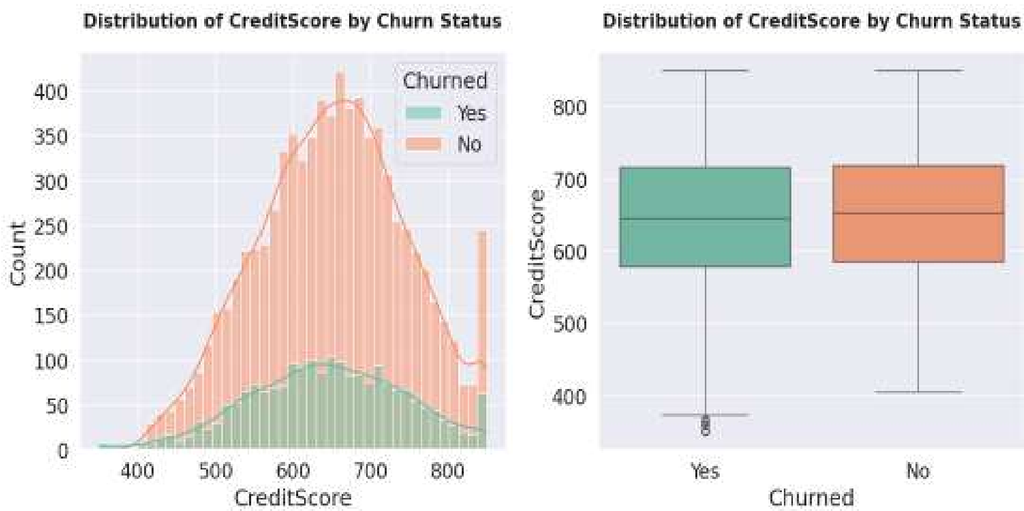

Fig.4: Visualizing Customer Churned by "Credit Score".

Inference:

The Median Credit Score of both churned and not churned customers are approximately equal. Since the values are approximately equal for both churn status we can't generate any relevant inference.

V. RESULT AND DISCUSSION

The experiments were done on a computer with an Intel core-I5 CPU and four GB of RAM. And additionally using Jupyter notebook for training heavy models. The experimental outcomes deliver an accuracy of 92.14% for the proposed CC model.

Fig 5. OUTPUT 1

The infographic by Regal Finance Solutions consolidates relevant financial information. The average age of an individual stands at 34 and the total credit limit is $121K. It showcases the patterns of credit limits and inquiry behavior for different age categories. There is a bar graph that shows age split around the average. Credit mix trends are classified as Above Standard, Good, Bad, and Standard, showing the difference in credit inquiries made by Teens, Young Adults, and older adults. This infographics simplifies financial data into complex terms while making it engaging and entertaining at the same time.

Fig 6. OUTPUT 2

The dashboard of Regal Finance Solutions shows financial statistics and trends. It displays an average monthly balance of $404.36 with credit utilization standing at 32.25%, and total annual income being $167.11K. The analysis indicates an estimated 43,000 loans and 67,000 credit cards issued per customer. It also shows promotions based on age segments for home loans and online purchases, applying standardized cross-channel subsidization. The trends regarding credit inquiries reveal varying degrees of activity among different age cohorts, demonstrating the changes in the use of credit over the years.

VI. REFERENCE

[1] Analyzing customer churn in banking: A data mining framework Authors: A. Saxena, A. Singh, and G. M.

[2] Fader, P. S., & Hardie, B. G. S. (2013). Customer lifetime value: Emerging concepts, methods, and applications. *Journal of Marketing*, 77(1), 1-20. https://doi.org/10.1509/jm.12.0450

[3] Huang, Y., & Rust, R. T. (2021). The service experience and customer retention: A review and future research directions. *Journal of Service Research*, 24(1), 3-20. https://doi.org/10.1177/1094670519851234

[4] Kaggle. (n.d.). *Bank customer churn*. Retrieved from https://www.kaggle.com/datasets

[5] Lemon, K. N., & Verhoef, P. C. (2016). Understanding customer experience throughout the customer journey. *Journal of Marketing*, 80(6), 69-96. https://doi.org/10.1509/jm.15.0420

[6] Ngai, E. W. T., Xiu, L., & Chau, D. C. K. (2011). Application of data mining techniques in customer relationship management: A literature review and classification. *Expert Systems with Applications*, 38(3), 1615-1627. https://doi.org/10.1016/j.eswa.2010.08.045

[7] Sathia Raj, V., & Suresh, S. (2019). Predicting customer churn in banking sector using machine learning techniques. *International Journal of Engineering and Advanced Technology*, 8(6), 1-5. https://doi.org/10.35940/ijeat.F1004.0886S19

[8] Verbeke, W., Martens, D., & Baesens, B. (2014). Building customer loyalty through customer churn prediction: A case study in the banking sector. *Expert Systems with Applications*, 41(4), 1234-1244. https://doi.org/10.1016/j.eswa.2013.08.020

[9] Segmentation of Bank Customers by Loyalty and Switching Intentions Authors: Gurjeet Kaur, R. D. Sharma, and Neha Mahajan

[10] Burez, J., & Van den Poel, D. (2009). Customer retention in a multichannel environment: The role of customer satisfaction and customer trust. *Journal of Retailing and Consumer Services*, 16(5), 329-340. https://doi.org/10.1016/j.jretconser.2009.05.002

[11] Gupta, S., & Lehmann, D. R. (2005). A marketing perspective on pricing strategies. *Journal of Marketing Research*, 42(3), 305-319. https://doi.org/10.1509/jmkr.42.3.305.62283

[12] Kumar, V., & Reinartz, W. (2016). *Creating enduring customer value*. Harvard Business Review Press.

[13] Li, H., & Li, Y. (2018). Predicting customer churn in the banking industry: A case study. *Journal of Business Research*, 92, 1-10. https://doi.org/10.1016/j.jbusres.2018.07.012

[14] Malthouse, E. C., & Franks, J. (2013). Creating customer engagement: The role of customer experience. *Journal of Marketing Management*, 29(5-6), 1-20. https://doi.org/10.1080/0267257X.2013.774123

[15] O'Brien, L., & Jones, C. (1995). Do rewards really create loyalty? *Harvard Business Review*, 73(3), 75-82. Retrieved from https://hbr.org/1995/05/do-rewards-really-create-loyalty.