Using A Predictive Approach to Customer Management: From Segmentation to Retention

Padmanaban.S[1], Mahalakshmi.G[2]
[1]*Student, Master of Computer Applications, Dr. M.G.R. Educational and Research Institute, Chennai*
[2]*Assistant Professor, Master of Computer Applications, Dr. M.G.R. Educational and Research Institute, Chennai*

Abstract - Customer segmentation groups people based on their buying habits, demographics, and preferences to improve marketing strategies and keep customers. Traditional static segmentation models have a hard time capturing dynamic behaviors, which makes marketing less effective and people less interested. This project uses advanced clustering methods based on Regency, Frequency, and Monetary (RFM) metrics to combine dynamic segmentation and churn prediction. It puts customers into three groups: Gold, Silver, and Normal. Then, Logistic Regression uses RFM scores, behavioral patterns, and demographic data to figure out how likely someone is to leave. Businesses can use segmentation and predictive analytics together to get useful information that helps them create targeted retention strategies, personalized offers, and loyalty programs. This two-step method fills the gap between traditional segmentation and predictive modelling of customer behaviour. It makes better use of resources and improves customer relationship management

INTRODUCTION

The choices and actions people or households make when choosing, acquiring, utilizing, and discarding goods or services are referred to as consumer behavior. These decisions are influenced by numerous sociological, psychological, and cultural elements. Problem identification, information collection, assessment, purchase, and post-purchase evaluation are all steps in the process. Companies use consumer behavior analysis to develop marketing strategies that meet the needs of their target audience and prevent financial losses. Purchase decisions are influenced by social influences, product attributes, marketing campaigns, and environmental factors. Forecasting demand and improving product design, pricing, promotion, and distribution are made easier with an understanding of consumer behavior. Complex buying, dissonance-reducing buying, habitual buying, and variety-seeking buying are the four categories of consumer behavior. Businesses can increase customer retention, provide tailored marketing campaigns, and spur business expansion by analyzing consumer trends.

OBJECTIVE

The project's goal is to use data analytics and machine learning to create and deploy a dynamic customer segmentation and churn prediction system for improved customer management and retention tactics. To dynamically divide up the client base according to demographics, frequency, and spending patterns.

- To use RFM analysis to find high-value and vulnerable clients.
- To use clustering algorithms for robust segmentation, like K-Means.
- To use historical data and logistic regression to forecast customer attrition.

SCOPE

Data Collection: Compile demographic, behavioral, and transactional data for analysis.

Data preprocessing includes cleaning, handling missing values, normalizing, and encoding categorical data. Customer Segmentation: Divide customers into Gold, Silver, and Normal groups using RFM analysis and K-Means clustering. Create features such as Recency, Frequency, Monetary values, and Product Preferences through feature engineering. Churn Prediction: Using historical data, train a logistic regression model to determine which customers are most likely to leave.

- Model Evaluation: Evaluate performance with metrics such as Inertia, Silhouette Score, ROC-AUC, Precision, and Recall.

REVIEW OF LITERATURE

Jingyuan Li (2024): This study predicts e-commerce churn using Random Forest, XG Boost, LightGBM, and Logistic Regression. SMOTE is used to address class imbalance, and SHAP and LIME are used to enhance model interpretability.

Tayab Khan and colleagues (2024): This study uses statistical modelling to identify important factors like contract length, service usage patterns, and customer interactions in order to predict churn in the telecom industry.

Waghol Prajwal et al. (2024): This study investigates a number of machine learning methods for churn prediction, including SVM, Random Forest, and XG Boost. It compares various models for accuracy and emphasizes how crucial churn reduction is to company profitability

SYSTEM CONFIGURATION

Processors: Intel Core TM i5 with 8 GB of DRAM; 320 GB of disk space

- Windows 10 is the operating system.
- RAM: 8GB at the very least (16GB or more for managing big datasets).
- Storage: 500GB SSD (1TB SSD or cloud storage for scalability).
- Internet: Fast internet for data retrieval and cloud-based model training Python for programming; Pandas and NumPy for library analysis; Scikit-Learn for clustering techniques (K-Means, DBSCAN, Hierarchical Clustering); and Scikit-Learn for predictive analytics (Logistic Regression)
- Data Visualization: Seaborn, Matplotlib Flask is the model deployment, and MySQL is the database

PYTHON

Guido van Rossum developed Python, a high-level, general-purpose programming language, between 1985 and 1990. Web development, data science, machine learning, GUI applications, and other fields make extensive use of it. Python is notable for its readability because it is accessible and uses English keywords rather than punctuation. Both the procedural and object-oriented programming paradigms are supported. Python's ease of use enables programmers to create code that is clear and easy to maintain. Python is used by big businesses like Google, Amazon, and Facebook because of its robust standard library, which contains frameworks like Selenium for web scraping, OpenCV for image processing, and Django for web development. Its syntax, which is based on indentation, improves readability. Because of its versatility and broad industry adoption, Python is a crucial skill for aspiring software engineers, especially in web development and data science.

PANDAS

Built on Python, Pandas is an open-source data analysis and manipulation tool that is strong, adaptable, and simple to use. It offers quick and simple data structures for handling relational and labelled data. Pandas, which enables users to import data from multiple formats like CSV, JSON, Parquet, SQL, and Excel, is crucial for useful data analysis. It facilitates effective data wrangling, cleaning, selection, reshaping, and merging. With its Data Frame structure, Pandas, a popular tool in data science, streamlines intricate procedures in tabular data manipulation. Pandas is used by researchers and analysts to turn unstructured data into insightful knowledge. Because of its strong functionality, it is a high-level building block for efficiently managing real-world datasets, guaranteeing smooth data processing in Python applications for a variety of industries

NUMPY

Numerical Python, or NumPy, is a library that includes multidimensional array objects and a number of processing routines. NumPy can be used to perform logical and mathematical operations on arrays. A general-purpose array processing package is called NumPy. It offers tools for working with multidimensional arrays as well as a high-performance array object.

MATPLOTLIB

Matplotlib is a complete Python visualization toolkit for making static, animated, and interactive visualizations. Matplotlib makes difficult things possible and easy things easy. Matplotlib is a plotting library for NumPy, a numerical mathematics extension for the Python programming language. Using general-

purpose GUI toolkits such as the interface, Qt, or GTK, it offers an object-oriented API for integrating plots into applications.

SCIKIT-LEARN

The 3-Clause BSD license governs the distribution of scikit-learn, a Python machine learning module built on top of SciPy. Scikit-learn, formerly known as sci kit learn and sometimes referred to as learn, is a free machine learning library for Python. Support-vector machines, random forests, gradient boosting, k-means, and DBSCAN are just a few of the classification, regression, and clustering algorithms it offers. It is also made to work with the Python scientific and numerical libraries NumPy and SciPy.

MYSQL

The Structured Query Language, the most widely used language for managing and accessing database records, is the foundation of the relational database management system MySQL. Under the GNU license, MySQL is a free and open-source program. The Oracle Company supports it. The MySQL database offers guidance on database administration and data manipulation through a variety of SQL queries. These queries include: create tables, drop tables, select records, update records, delete records, and insert records. To aid in your comprehension of the MySQL database, there are also provided MySQL interview questions. The most widely used database management system software for relational database administration at the moment is MySQL. Oracle Company supports this open-source database software.

WAMPSERVER

A Windows web development environment is called WampServer. With Apache2, PHP, and a MySQL database, it enables you to build web applications. Additionally, PhpMyAdmin makes it simple for you to manage your database. Using PHP Apache2 and a MySQL database, WAMP Server is a dependable web development tool that enables you to create web applications. The program has many features and an easy-to-use interface, which makes it the go-to option for developers worldwide. No payment or subscription is needed to use the software, which is available for free.

CONCLUSION

By employing logistic regression for churn prediction and RFM-based segmentation, the suggested system improves marketing tactics and customer retention. It helps businesses target high-value users and identify at-risk customers by classifying customers into Gold, Silver, and Normal groups. Businesses can tailor engagement through customized offers and proactive retention strategies by anticipating churn. To guarantee accuracy in customer insights, the system combines modules for segmentation, prediction and recommendation

REFERENCE

[1] Nikhil Patel and Sandeep Trivedi(2020)- Using AI Chatbots, Predictive Modelling, Machine Learning, Personalization, NLP, and Customer Service to Boost Customer Loyalty.

[2] Sakshi Koli; Rajesh Singh; Rashmi Mishra; Preeti Badhani(2023)- Using machine learning techniques to retain customers requires the use of customer segmentation techniques.

[3] Zhang Wei;Chen Lihua;Xu Lijuan(2024)- Data-Mining-Based Prediction Model for E-Commerce User Purchase Behavior

[4] Websites Referred

[5] www.python.org

[6] www.numpy.org

[7] www.scikit-learn.org

[8] www.matplotlib.org