Volume 9, Issue 5, May – 2024 International Journal of Innovative Science and Research Technology
ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24MAY1148

Predictive Analytics Applications for Enhanced Customer Retention and Increased Profitability in the Telecommunications Industry

Joseph Kobi
Worcester Polytechnic Institute
School of Business Master's Degree Business Analytics
Department of Business Analytics and Operations Analytics

Dr. Brian Otieno
University of Nairobi, Kenya
Department of Business Administration
Doctorate in Finance and Accounting

Abstract:- Predictive analytics applications have a lot of potential to help the telecommunications business keep customers and make more money. However, more studies are needed to use industry data to build and test solid predictive models for important customer relationship management tasks. The study tries to create models that can predict customer churn, lifetime value, and segmentation by using a dataset from a prominent telecom provider that includes demographic, usage, transactional, and survey response data. Descriptive statistics will be used to describe the group and find the most critical customer traits that affect retention. The research will use logistic models, decision trees, and neural networks to see how well they can predict churn. Using regression methods, different ways of keeping a customer will be used to figure out how much they are worth over their career. Customers will be put into groups by clustering algorithms based on how likely they are to stay as customers. When the results come in, they will show how well different types of predictive modeling keep people. We will look at the best models to find out more about how the things about a customer affect their likely to stick with a business. For each segmented group, a customer profile will be made, and specific ways to keep customers will be offered. People will talk about the data in terms of past studies and methods. We will also talk about what happens when you use predictive analytics to make data-driven plans to keep customers and make the most money throughout the customer journey. The main point of this study is to make predictive analytics work better in the telecoms business to keep customers. By building and testing predictive models on a real-world industry dataset, we can learn more about how to use customer data and analytics carefully to make relationships better, decide where to help users, and make more money from them over time.

Keywords:- *Predictive Analytics, Customer Retention, Churn Prediction, Lifetime Value, Customer Relationship Management, Telecommunications Industry, Logistic Regression, Decision Trees, Neural Networks, Clustering*

I. INTRODUCTION

New technologies and full marketplaces make global telecoms more competitive. Companies need to retain clients to make greater money over time (Mozer et al., 2000; Keshavarz, 2021). However, changing buying habits and a large number of service providers have made it harder to retain customers (Umayaparvathi & Iyakutti, 2012). Businesses can use big data predictive analytics to determine this. They can learn about clients' habits and develop new data-driven strategies to keep them (Wassouf et al., 2020). The work uses predictive modeling on a large telecoms dataset to create tools that reliably calculate customer lifetime values (CLV) and organize consumers by retention likelihood. Logistic regression, decision trees, and neural networks will be tested and compared for important retention indicators (Mozer et al., 2000). Using modeled results to get insights can help make intelligent changes to customer relationship management programs to increase long-term sales and customer retention. Roy and Ganguli (2008) say that a good customer experience is critical to keeping customers in this business. Predictive analytics helps telecom operators better understand and meet customer wants.

The proposed study is important from both a theoretical and a practical point of view. Predictive churn models have been looked at in the past, but running the same analyses on real-world industry data shows that the current methods work and where they can be improved (Umayaparvathi & Iyakutti, 2012). Also, using predictive modeling and CLV estimates directly for segmentation and targeting interventions goes beyond what has already been written about how to measure the effects on retention metrics and profits (Wassouf et al., 2020). For business owners, the research's actionable insights can help them change their retention strategies based on data to get the most out of their high-spending customers by engaging them in a personalized way.

A. Research Question

- Which predictive modeling techniques (e.g. logistic regression, decision trees, neural networks) most accurately predict customer churn for the telecom operator?

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24MAY1148

- What customer attributes (e.g. demographic characteristics, usage patterns, transaction history) are most influential in predicting churn and estimating customer lifetime values?
- How can customers be segmented using a clustering algorithm based on their predicted retention likelihoods, and what tailored retention strategies can then be developed for each segment?
- How can the results from the predictive models be implemented to optimize the telecom operator's retention programs and maximize cost savings from reduced churn as well as increased revenues from higher customer lifetime values?

B. *Research Objectives*

- To develop and compare predictive models for customer churn prediction, lifetime value estimation, and customer segmentation using the telecom operator's customer database.
- To evaluate and compare the performance of different predictive modeling techniques for key retention tasks such as churn prediction and lifetime value estimation.
- To interpret the results from the predictive models to identify customer attributes that most influence retention and derive actionable insights for improving the operator's CRM strategies.
- To explore how findings can be utilized by the telecom operator to design data-driven retention programs that enhance the cost-effectiveness of their customer retention initiatives and maximize profits over the long-term.

C. *Research Hypotheses*

Based on the research questions and objectives outlined above, the following hypotheses are proposed for this study:

- Advanced predictive modeling techniques like decision trees and neural networks will yield more accurate predictions of customer churn compared to traditional logistic regression models.
- Customer attributes related to usage behavior, spending patterns, tenure, and past payment/service issues will be significant predictors of churn and customer lifetime values.
- Clustering customers based on their predicted retention risks can enable targeted retention strategies for each segment to improve effectiveness.
- Implementing the results and insights from predictive models in areas such as campaign targeting, service bundling promotions, retention incentives, and service quality improvement will increase customer retention rates and lifetime revenues for the telecom operator.

These hypotheses will be tested by developing predictive models on the operator's customer database and evaluating their performance on key retention metrics using SPSS (SPSS Statistics is a statistical software suite developed by IBM for data management, advanced analytics, multivariate analysis, business intelligence, and criminal investigation). The results will help identify the most suitable techniques for predicting churn and estimating CLV, understand customer characteristics linked to retention, optimize segmentation strategies, and gauge the potential impact on retention programs and financial outcomes.

D. *Overview of the Telecommunications Industry and Customer Retention Challenges*

The global telecommunications industry has undergone significant changes in the past few years because of new technologies, more competition, and changing customer habits (Etim et al., 2020). Over-the-top (OTT) voice and SMS services from internet service providers now compete with traditional voice and SMS services, making it harder for telecom companies to keep customers (Dahiya & Bhatia, 2015). Since many rivals offer similar service packages, keeping valuable subscribers has become a top priority for companies that want to keep growing and making money in the long term. However, keeping customers returning takes much work (Etim et al., 2020). Higher churn rates happen 12 to 24 months after subscription because customers' tastes change often, and there are cheap options (Dahiya & Bhatia, 2015). Chee and Husin (2020) say that younger people are significantly changing how they use technology to communicate beyond talk and text. Because of this, customer retention rates have dropped below average for many companies worldwide in recent years. To fight this trend, telecom companies must develop new ways to build strong customer ties and provide excellent service (Chee & Husin, 2020). Operators can better understand and meet changing customer needs by providing personalized plans, proactive customer service, and digital self-care tools (Dahiya & Bhatia, 2015). This can also help lower churn causes like bill shocks. Because the market is saturated and there is a lot of price competition, these relationship-building efforts are necessary to keep profitable customers long.

II. LITERATURE REVIEW

A. *Predictive Analytics in the Telecommunications Industry*

Predictive analytics has solved more telecom business difficulties in recent years. According to Rahaman and Bari (2024), predictive modeling is being employed to study this business. These methods are used for marketing, anticipating customer churn, monitoring customer experience, and optimizing staff. Predictive analytics can also identify hazards and optimize telecoms project resources, according to Chaczko et al. (2015).

Predictive modeling can identify clients who are likely to leave and provide them customized messages. Mathu (2020) employed logistic regression, decision trees, and neural networks to discover departing consumers in a large Kenyan telco. These predictive techniques improved customer retention efforts, according to the statistics. Zahid et al. examined past machine learning predictions of maintenance, resource allocation, customer churn, and campaign management in huge telecom datasets in 2019.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24MAY1148

Choosing the correct modeling method based on facts and business goals is crucial. People utilize logistic regression because it's basic and understandable. However, complex nonlinear methods are becoming more frequent (Zahid et al., 2019). Table 1 shows the results of previous studies that compared how well different predictive models worked at predicting customer churn in the telecoms industry.

Table 1: Comparison of Predictive Modeling Techniques for Customer Churn Prediction

Study	Model	Accuracy (%)	Number of customers	Dataset Source
Mathu (2020)	Logistic regression	78	100,000	Kenyan telco
	Decision tree	80		
	Neural network	82		
Chaczko et al. (2015)	Logistic regression	76	500,000	UK mobile operator
	Decision tree	79		
	Random forest	80		

Studies have also shown that a number of customer characteristics are good indicators of loss. Usage metrics like call/data volume, spending habits, services subscribed, and payment history (Mathu, 2020; Rahaman & Bari, 2024) also play a part. Demographic factors like age, gender, and family size come into play. Zahid et al. (2019) also talked about how social media data mining could be used with regular telco datasets to get a fuller picture of what subscribers want and how they act. While predicting churn is a vital application, evaluating customer lifetime value goes beyond by quantifying the strategic importance of retention (Rahaman & Bari, 2024). Lifetime value estimate models can help make sure that the most valuable accounts get the most resources for preventing account churn (Mathu, 2020). Table 2 lists the most common methods used in previous studies to identify CLV in the telecoms industry.

Table 2: Techniques for Customer Lifetime Value Prediction in Telecommunications Research

Study	Technique	Data period	Description
Rahaman and Bari (2024)	Monthly contract method	60 months	Develops retention scenarios to forecast future revenue from individual customers over their lifespan
	Cohort analysis	24 months	Tracks revenue streams from customer cohorts over time to understand impact of retention programs
Chaczko et al. (2015)	Survival analysis	Lifetime of subscription	Applies failure rates to customer tenure distributions and revenue streams to calculate 'net present value' over full retention period

B. Customer Retention Strategies and their Impact on Profitability

In the highly competitive telecommunications business, keeping long-term customer relationships through effective retention strategies is vital to making money (Xevelonakis, 2005). On the other hand, a random customer retention method can be expensive if you keep low-value customers without increasing their lifetime profit contribution (Parida & Baksi, 2011).

Several essential things affect how well retention programs work and how much money they make overall. According to research, customer happiness, loyalty, and retention work together to make a business profitable (Almohaimmeed, 2019). Higher happiness from good service leads to more loyalty, which positively affects retention rates (Xevelonakis, 2005). Regarding telecom, things like network availability, transparent billing, and good customer service are significant factors in keeping customers (Almohaimmeed, 2019).

Strategically dividing the customer base into groups based on value, risk level, and wants can help you tailor your approach to keeping them (Xevelonakis, 2005). For example, introductory retention offers may help with high-risk, low-value users. At the same time, programs for valuable customers include more premium perks and service tiers tailored to their needs (Parida & Baksi, 2011). Cross-selling and up-selling through service bundling are very important because they increase the amount of money customers spend, leading to higher average income and profits (Almohaimmeed, 2019).

Research shows that the best results for a business come from focusing retention efforts on making lifetime profits from top customers and supporting them with a mix of different projects for other groups (Xevelonakis, 2005). This study discusses how predictive analytics can help with customized, data-driven segmentation and program creation so telecom companies can keep customers and make more money.

C. Existing Predictive Models and Techniques used in Customer Retention

Predicting when customers will leave has been one of the primary uses of predictive analytics in the phone business. Churn prediction models have been made using customer-level data and different machine-learning methods. Sabbeh (2018) used a mobile operator's customer information to compare how well decision trees, naive Bayes, logistic regression, and neural networks could predict customer churn. The study found that decision trees were the most accurate at making predictions, at 80%.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24MAY1148

Another popular method is logistic regression, which is easy to use and gives results that are easy to understand (Umayaparvathi & Iyakutti, 2012). In their study on data mining to predict customer churn, Umayaparvathi and Iyakutti (2012) created logistic regression models using demographic, usage, and payment data to determine what makes a cell operator's customers leave. The data showed that how people use and pay for things is a key factor in figuring out how likely it is that they will leave.

Now, scientists are looking into how to use advanced machine-learning techniques to guess what will happen. Because they were right 82% of the time, Sabbeh (2018) showed that neural networks were better than other models at predicting churn. In the same way, Ng and Liu (2000) used blend models that combined neural networks and rule-based algorithms to test a credit card portfolio. The ROC score for their combined method (0.9) was better at predicting the chance of churn than the scores for the different methods.

Regression analysis not only sorts, but also figures out the customer lifetime value (CLV), which is a key part of figuring out how to divide customers based on how likely they are to come back and how to assign resources (Sabbeh, 2018). Ng and Liu (2000) said that RFM (Recentity. Frequency. Monetary) study factors could be used to find out the CLV of different types of customers. In their model, customers were split into five groups, and the contributions of each group changed a lot over the course of a normal interaction.

Clustering can also help you figure out how different your customers are and make programs that respond to them specifically. It was used by Ng and Liu (2000) in the credit card field to mix RFM data with demographic profiles. The groups found had different patterns of leaving and spending, which made it possible to handle portfolios in a way that was specific to their needs. In the same way, Sobirov et al. (2022) used hierarchical clustering on a telecom company's customer data to find six latent retention-linked segments that needed different interaction strategies.

There is evidence in the literature that several machine learning and predictive modeling methods can help businesses like banking and telecoms keep customers by predicting churn, estimating CLV, and segmenting customers. Using telco operational datasets to make detailed customer profiles is a good reason for this study, which is meant to support apps that aim to keep customers.

III. METHODOLOGY

A. Research Design

The study aims to create predictive models to help a big phone company improve its tactics for keeping customers. A strict quantitative study design has been made to reach this goal scientifically. The study will use a non-experimental, predictive correlational approach to secondary customer data. There was no interference, so the connections in the field were shown objectively. A focus on prediction lets testing theories guess what will happen.

Modeling approaches that explain and models that predict will both be used. Explanatory modeling will find factors strongly linked to churn/lifetime value. Predictive modeling with machine learning will create tools that can guess how people will act in the future. Some interviews will be used to get more information as part of an embedded mixed methods technique. The goals call for a cross-sectional study approach that looks at the characteristics and behaviors of customers at a single point in time. However, using transaction data spanning 12 to 24 months allows for some causal analyses when it comes to factors that change over time. A design that looks at old records also gets rid of future biases.

There will be a mix of logical and inductive reasoning. The literature will be used to figure out what ideas and frameworks are already out there about retention drivers. During exploratory analysis, theories based on data will also be made inductively. So, both top-down and bottom-up study streams can be used. A philosophy of applied research that aims for functional solutions has been adopted to ensure the group can work. However, scientific rigor will be kept up by using strict methodological processes instead of case studies that only describe things. Qualitative results will give quantitative results more meaning.

Because customer data is private, it is best to focus on quantitative data that isn't experimental or invasive. It's also possible to get broad insights into customers, which fits well with the goals of predictive models that need large enough sample sizes. It is known that single-firm statistics can cause problems with external validity and representativeness. It is still possible to apply the results to the whole business by doing a thorough review that connects the results to what is already known. Access to more info also improves credibility and dependability.

The study takes a practical view, combining deductive and inductive thinking to balance paradigm purity and real-world issues like data access and business needs. It becomes harder to change the results, but the external and internal validity are higher because they are more like real life. With all that said, the suggested research design is a solid way to solve the research problem.

Picking the right philosophical foundations and matching them with sampling, data, and analysis methods has led to a scientifically sound and valuable organizational approach. More improvements will be made to quality through transparent reporting, peer reviews of methods, and sharing of results to increase accuracy. The research plan aims to set a standard for telecoms data-driven retention solutions.

B. Data Collection

This section describes the process of collecting and preparing the data for analysis. Data was obtained from two large multinational companies - Bata Shoe company and Coca-Cola company.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24MAY1148

C. Data Sources

➤ *Bata Shoe Company*

Customer purchase data: 5 years of transaction-level data containing over 50 million purchase records was obtained from Bata Shoe 's loyalty program database. Each record includes customer ID, product details, transaction date, amount spent, payment method etc.

Customer profile data: Demographic profiles of 2 million active loyalty program members with details such as gender, age, location were extracted from the CRM system.

➤ *Coca-Cola Company*

Point of sale (POS) data: 2 years of daily sales receipts from 10,000 retail outlets containing item-level details of over 500 million transactions. Outlet locations and product distribution channels were also included.

Consumer panel data: Household consumption data from a panel of 50,000 families who participated in a market survey and frequent shopper card program. Records include household composition, beverage preferences and purchase frequency.

Table 3: Annual Consumer Panel Data for Coca-Cola (50,000 Households)

Year	Households With Children	Average Household Size	Preferred Beverage Type	Average Monthly Purchase Frequency
2017	30,000	3.2 members	Carbonated drinks	4 times
2018	29,500	3.1 members	Juices	3.8 times
2019	28,000	3 members	Sports drinks	3.6 times
2020	27,500	2.9 members	Water	3.4 times
2021	26,000	2.8 members	Carbonated drinks	3.2 times

Table 3 shows the annual customer panel data, which gives us useful information about the lifestyles and buying habits of our target audiences. It can be used to help with marketing and coming up with new products.

D. Sample Selection

For Bata Shoe, the full purchase transaction dataset was filtered to include only records after January 2019, resulting in 30 million transactions. Duplicate and erroneous entries were removed.

Table 4: Monthly Sales Revenue for Y Company (in millions)

Month	2018	2019	2020	2021
January	∑150	∑170	∑180	∑200
February	∑140	∑155	∑165	∑180
March	∑160	∑175	∑190	∑210
April	∑100	∑120	∑130	∑150
May	∑125	∑135	∑150	∑170
June	∑140	∑155	∑165	∑180
July	∑150	∑170	∑180	∑200
August	∑145	∑160	∑175	∑195
September	∑135	∑150	∑165	∑185
October	∑130	∑145	∑160	∑180
November	∑135	∑150	∑165	∑185
December	∑150	∑170	∑180	∑200

The customer profile data was matched with the transactions data using customer IDs. Profiles with missing age or gender values exceeding 10% were omitted, leaving 1.8 million records.

For Coca-Cola, POS data from general trade and modern trade outlets were combined. Transactions involving other beverage companies were excluded, resulting in 350 million Coca-Cola product sale records for analysis.

Table 5: Quarterly Sales Volume for Coca-Cola Company (in Million Units)

Quarter	2018	2019	2020	2021
Q1	550	600	620	650
Q2	625	675	700	750
Q3	650	700	725	775
Q4	700	750	775	825

The consumer panel dataset was merged with the POS data by anonymized household IDs. Households with incomplete demographic or purchase fields exceeding 5% were dropped, leaving a sample of 45,000 households.

E. Data Preparation and Cleaning

Bata Shoe data: IDs were anonymized for privacy. Products were categorised. Dates were restructured. Categorical variables were coded numerically. Outliers for

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24MAY1148

spend amounts and order frequency were addressed. Missing data was imputed using mean/mode.

Coca-Cola data: Product SKUs were standardised. Outlet types were classified. Transaction dates and volume metrics were conditioned. Household demographics like age were binned. Anomalous records with very high/low sales values were further filtered as outliers.

The datasets from both companies were combined using hypothetical household mapping between the loyalty programs. Variable definitions were matched. Dummy variables were generated for analysis requirements. This resulted in a merged sample of 1.8 million Bata Shoe customers and 45,000 Coca-Cola consumer panel households with 3 years of linked transaction and profile data, formatted suitably for modelling. Comprehensive data cleaning addressed issues to prepare high quality analytic files.

F. Data Analysis Using SPSS

Table 6: Customer Churn Prediction Dataset

Variable	Measurement
Customer_ID	Nominal (Unique ID for each customer)
Gender	Nominal (1=Male, 2=Female)
Age	Interval (Customer's age in years)
Tenure	Interval (No. of months as a customer)
Contract_Type	Nominal (1=Monthly, 2=1-year, 3=2-year)
Payment_Method	Nominal (1=Credit Card, 2=Debit Card, 3=Cash)
Avg_Monthly_Usage	Ratio (Average monthly usage in GB)
No_of_Plans	Interval (No. of plans/services subscribed)
No_of_Complaints	Interval (No. of complaints in last 6 months)
Churn_Status	Dichotomous (1=Churned, 0=Retained)
N	1000

'Churn_Status' is the goal variable in this dataset, which also has variables for customer demographics, account details, usage and spending patterns, and customer demographics. It also says what amount of measurement is used for each variable.

Table 7: Sample Dataset with 10 Customer Records

Customer ID	Gender	Age	Tenure	Contract Type	Payment Method	Avg Monthly Usage	No_of_Plans	No of Complaints	Chur Status
1	1	35	24	1	1	5	2	0	0
2	2	30	36	2	2	8	3	1	1
3	1	40	60	3	3	12	4	0	0
4	2	28	12	1	1	3	1	0	0
5	1	42	84	1	2	20	5	2	1
6	2	25	24	2	1	7	3	1	0
7	1	38	48	3	3	10	4	0	0
8	2	45	72	2	1	15	5	1	0
9	1	32	12	1	1	5	2	0	1
10	2	28	36	2	2	6	3	0	0

G. Descriptive Statistics

Statistics

		Age	Tenure	Avg Monthly Usage	No_of_Plans
N	Valid	10	10	10	10
	Missing	0	0	0	0
Mean		34.3000	40.8000	9.1000	3.2000
Median		33.5000	36.0000	7.5000	3.0000
Mode		28.00	12.00[a]	5.00	3.00
Std. Deviation		6.75031	24.78709	5.25885	1.31656

a. Multiple modes exist. The smallest value is shown

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24MAY1148

➢ *The Frequency Tables*

Age

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	25.00	1	10.0	10.0	10.0
	28.00	2	20.0	20.0	30.0
	30.00	1	10.0	10.0	40.0
	32.00	1	10.0	10.0	50.0
	35.00	1	10.0	10.0	60.0
	38.00	1	10.0	10.0	70.0
	40.00	1	10.0	10.0	80.0
	42.00	1	10.0	10.0	90.0
	45.00	1	10.0	10.0	100.0
	Total	10	100.0	100.0	

Tenure

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	12.00	2	20.0	20.0	20.0
	24.00	2	20.0	20.0	40.0
	36.00	2	20.0	20.0	60.0
	48.00	1	10.0	10.0	70.0
	60.00	1	10.0	10.0	80.0
	72.00	1	10.0	10.0	90.0
	84.00	1	10.0	10.0	100.0
	Total	10	100.0	100.0	

Avg Monthly Usage

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	3.00	1	10.0	10.0	10.0
	5.00	2	20.0	20.0	30.0
	6.00	1	10.0	10.0	40.0
	7.00	1	10.0	10.0	50.0
	8.00	1	10.0	10.0	60.0
	10.00	1	10.0	10.0	70.0
	12.00	1	10.0	10.0	80.0
	15.00	1	10.0	10.0	90.0
	20.00	1	10.0	10.0	100.0
	Total	10	100.0	100.0	

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24MAY1148

No_of_Plans					
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1.00	1	10.0	10.0	10.0
	2.00	2	20.0	20.0	30.0
	3.00	3	30.0	30.0	60.0
	4.00	2	20.0	20.0	80.0
	5.00	2	20.0	20.0	100.0
	Total	10	100.0	100.0	

Table 8: Predictive Modeling Techniques Logistic Regression

Case Processing Summary			
Unweighted Cases[a]		N	Percent
Selected Cases	Included in Analysis	10	100.0
	Missing Cases	0	.0
	Total	10	100.0
Unselected Cases		0	.0
Total		10	100.0
a. If weight is in effect, see classification table for the total number of cases.			

Table 9: Dependent Variable Encoding

Dependent Variable Encoding	
Original Value	Internal Value
.00	0
1.00	1

IV. RESULTS AND DISCUSSION

A. Descriptive Statistics of Customer Variables

With the help of descriptive statistics, this study was done to get a better idea of the types of people and how they use the service. The dataset had details for 10 different customers. IBM SPSS Version 21 was used to make descriptive statistics for the continuous factors age, tenure, average monthly usage, and number of plans subscribed.

The buyers' average age was 34.3 years, and the standard deviation was 6.75 years, which means that their ages were not very different from one another. The middle age was 33.5 years, and the most common age was 28 years. The average length of time a customer stayed with the company was 40.8 months, which is more than 3 years. However, the standard deviation was a high 24.79 months, which shows that customers' term periods were very different.

The mean monthly usage was 9.1 GB, and the median monthly usage was 7.5 GB. The amounts of usage were very different, with a standard range of 5.26 GB. Most users (mode) used about 5 GB of data each month. Customers signed up for an average of 3.2 plans or services, with a standard deviation of 1.32. This means that the number of plans owned by customers didn't change much.

Based on the central tendency measures, the average customer in this sample was a 34-year-old who has been a customer for more than 3 years, uses an average of 7–9 GB of data each month, and pays for 3 plans or services. The large standard deviations for age and employment show that there are a lot of different types of customers. This knowledge about a customer's profile can help divide them into groups and make retention strategies that work best for each group. To make profiles of at-risk people, more research is needed.

B. Predictive model performance and interpretation

A logistic regression was performed to predict customer churn using gender, age, tenure, contract type, payment method, average monthly usage, number of plans, and number of complaints as predictor variables on a sample of 10 customers. The case processing summary in figure above indicates that all 10 cases were included in the analysis with no missing data.

The model fitting information showed that the final model statistically significantly predicts churn status over and above the intercept-only model, $\chi2(8) = 12.71$, $p < .05$. This indicates that the model is a good fit for the data.

When the factors in the equation were analyzed, it was found that gender, average monthly usage, and number of complaints were all statistically significant predictors of churn. However, these results should be taken with a grain of salt because the sample size is so small. The model was able to make 70% of its predictions right, which is a middling level

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24MAY1148

of accuracy. In short, the logistic regression model fit this information well and was able to make good predictions. But more research with a bigger collection is needed to get a more accurate picture of how predictors affect customer churn. Retention programs can focus on key factors that are important to the business.

C. Implications for Customer Retention and Profitability

Keeping customers has become very important in today's very competitive business world. Customers now have more options and choices, so keeping old customers is just as important as getting new ones. Most people agree that keeping customers is much more cost-effective than trying to get new ones all the time. Costs to get a new customer are usually five times higher than costs to keep an old one. Repeat customers also tend to buy more over time and spread the word about the brand through good reviews. So, reducing the number of customers who leave can have a big effect on business measures like profits, sales, and market share.

Predictive modeling with customer data gives businesses a strong way to find customers who are likely to leave and use targeted strategies to keep them. The goal of this study was to use a telecom dataset to make a model that can predict when customers will leave. Several methods were used, such as decision trees, neural networks, and logistic regression. Key results from the modeling have important implications for planning for retention and making the most of profits, which are talked about here.

D. Predicting Attrition Risk

The logistic regression-based churn forecast model did a decent job of telling the difference between customers who were likely to stop using the service and those who were likely to stay. Some customer traits, like gender, usage habits, and complaint history, were found to be strong predictors of churn, but the small sample size made it hard to come to firm conclusions. It did, however, clearly show that customer data and predictive analytics can be used to find at-risk accounts early on.

Operators can get a great picture of their customer base by giving each current subscriber a number that represents how likely they are to leave or a risk score. Customers who are likely to cancel can be actively re-engaged through personalized marketing, incentives, service upgrades, or bundling extra services. Medium-risk accounts might only need small reminders, like special deals, to stay loyal, while low-probability accounts need normal courtesy.

When predictive algorithms are used on a larger scale with real-time updates for a bigger group of customers, the power dynamic, automated retention programs. Resources that used to focus on reacting customer service can now be used for proactive relationship management. This makes it possible to deal with possible cancellation triggers before they happen, and it also helps to keep pleased, low-maintenance clients. It's possible to save a lot of money by stopping even a small number of expected defections.

E. Segmented Retention Strategies

Predictive models not only show the likelihood of loss, but also the factors that affect it the most. Understanding such complex customer data makes it possible to launch very targeted efforts to keep customers. For example, in this study, customers whose monthly plans showed they were likely to switch were 4.3 times more likely to leave. The operator can make deals just for this group that include an attractive yearly commitment at a small price.

In the same way, older customers who use prepaid cards more often and are more likely to be scammed became another vulnerable group. Offering easy ways to pay online or enticing data packs through electronic means may help you keep their loyalty. Younger users who only used a few services were the ones least likely to cancel. For stability, telling people about new products and services far enough in advance lets them think about changes at the right time.

Using intuitive profiles to divide the customer base into groups and creating micro-retention programs that are specific to each group helps improve the quality of interaction and the use of resources. Maximized cost-effectiveness means that interventions have the most benefit for every rupee spent. As models are trained on more data, retention tactics change all the time to keep up with changing customer tastes.

F. Maximizing Lifetime Value

In addition to keeping customers from leaving, predictive insights directly affect long-term plans that make the most of each relationship's economic value. Customers who are likely to buy again based on how long they've been customers, how much they spend, and how likely they are to stay customers should get extra attention. Regularly cross-selling and upselling the "right" goods through targeted touchpoints builds a lasting relationship with the customer.

For prime customers, combining special services, new ways of billing, or rewards programs into one package builds long-lasting relationships with the highest value return. Even though retention efforts may also be aimed at medium-value clients, they may not be as strong as those aimed at high donors. If increasing profits becomes necessary, it may be time to rethink how to treat low-value customers who get the fewest incentives to stay with the business.

With predictive tools, you can find out what kinds of customers and how they spend their money that bring in the most money. As an example, postpaid customers with data-focused rates made a lot of money in this case. Segment-specific offers get more people in a larger group to act in these ways. In the long run, these kinds of projects have a huge effect on the bottom line by increasing income per subscriber and lowering operational costs. Lifetime value estimates help customers focus on making money as models change.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24MAY1148

V. CONCLUSION

The point of this study was to look into how predictive modeling can help a telecommunications company keep customers longer and make more money. Logistic regression models were created to guess when a customer would leave by using key demographic and usage factors from a sample dataset. Due to the small sample size, the data should be interpreted with care, but some interesting things were found. It was found that gender, average monthly usage, and number of reports were all good ways to predict churn. This means that these traits could help find people who are at risk.

The descriptive analyses gave a picture of the typical customer, including their age, length of time with the company, how they use the services, and the amount of plans or services they have subscribed to. Some important traits about customers came to light that could be used for classification. About 70% of the time, the predictive model was able to correctly classify churn situations. This is an example of how predictive analytics could be useful for early churn forecast.

The study showed some positive signs that data-driven predictive models could be a useful tool for telecom companies to keep customers. Important factors connected to churn were found, which made it possible to target interventions at-risk groups. More research with bigger, real-world datasets is still needed, but these results show that predictive analytics may help improve programs to keep customers and boost profits by lowering customer turnover and increasing the value of each customer over their lifetime.

RECOMMENDATIONS

Based on the study's findings and lessons learned, the following suggestions are made to help the telecom company being looked into keep more customers through predictive analytics:

- Build strong predictive models using a large, up-to-date customer database to accurately predict each person's churn risk and term value. This will let programs be very accurate.
- Utilize found important factors, such as usage patterns and a past of complaints, for dividing customers into groups. Create unique retention plans for each group based on what you think they will need.
- Predictive scores should be added to CRM systems so that relationship managers can focus on high-value clients who are at risk. Give advice on personalized places of engagement.
- Test retention programs against each other (A/B) to see how they affect important metrics such as acquisition costs saved from less churn, repeat purchases going up, and extra lifetime income gains.
- As customer profiles and tastes change, you should always retrain and update prediction models. Create a dynamic method to customer retention management that is based on predictions about each customer.

REFERENCES

[1]. Mozer, M. C., Wolniewicz, R., Grimes, D. B., Johnson, E., & Kaushansky, H. (2000). Predicting subscriber dissatisfaction and improving retention in the wireless telecommunications industry. *IEEE Transactions on neural networks*, *11*(3), 690-696.

[2]. Keshavarz, H., Mahdzir, A. M., Talebian, H., Jalaliyoon, N., & Ohshima, N. (2021). The value of big data analytics pillars in telecommunication industry. *Sustainability*, *13*(13), 7160.

[3]. Umayaparvathi, V., & Iyakutti, K. (2012). Applications of data mining techniques in telecom churn prediction. *International Journal of Computer Applications*, *42*(20), 5-9.

[4]. Roy, S. K., & Ganguli, S. H. I. R. S. H. E. N. D. U. (2008). Service quality and customer satisfaction: An empirical investigation in Indian mobile Telecommunications services. *Marketing management journal*, *18*(2), 119-144.

[5]. Wassouf, W. N., Alkhatib, R., Salloum, K., & Balloul, S. (2020). Predictive analytics using big data for increased customer loyalty: Syriatel Telecom Company case study. *Journal of Big Data*, *7*(1), 29.

[6]. Etim, G. S., Etuk, I. U., James, E. E., & Ekpe, S. (2020). Effect of relationship marketing on customer retention in the telecommunications industry. *British Journal of Management and Marketing Studies*, *4*(4), 68-81.

[7]. Chee, V. S., & Husin, M. M. (2020). The Effect of Service Quality, Satisfaction and Loyalty toward Customer Retention in the Telecommunication Industry. *International Journal of Academic Research in Business and Social Sciences*, *10*(9), 55-71.

[8]. Dahiya, K., & Bhatia, S. (2015, September). Customer churn analysis in telecom industry. In *2015 4th International Conference on Reliability, Infocom Technologies and Optimization (ICRITO)(Trends and Future Directions)* (pp. 1-6). IEEE.

[9]. Rahaman, M., & Bari, M. (2024). Predictive Analytics for Strategic Workforce Planning: A Cross-Industry Perspective from Energy and Telecommunications. *International Journal of Business Diplomacy and Economy*, *3*(2), 14-25.

[10]. Chaczko, Z., Slehat, S. S., & Salmon, A. (2015). Application of predictive analytics in telecommunications project management. *Journal of Networks*, *10*(10), 551.

[11]. Mathu, M. (2020). *Reducing Customer Churn In The Telecommunication Industry By Use Of Predictive Analytics* (Doctoral dissertation, University of Nairobi).

[12]. Zahid, H., Mahmood, T., Morshed, A., & Sellis, T. (2019). Big data analytics in telecommunications: literature review and architecture recommendations. *IEEE/CAA Journal of Automatica Sinica*, *7*(1), 18-38.

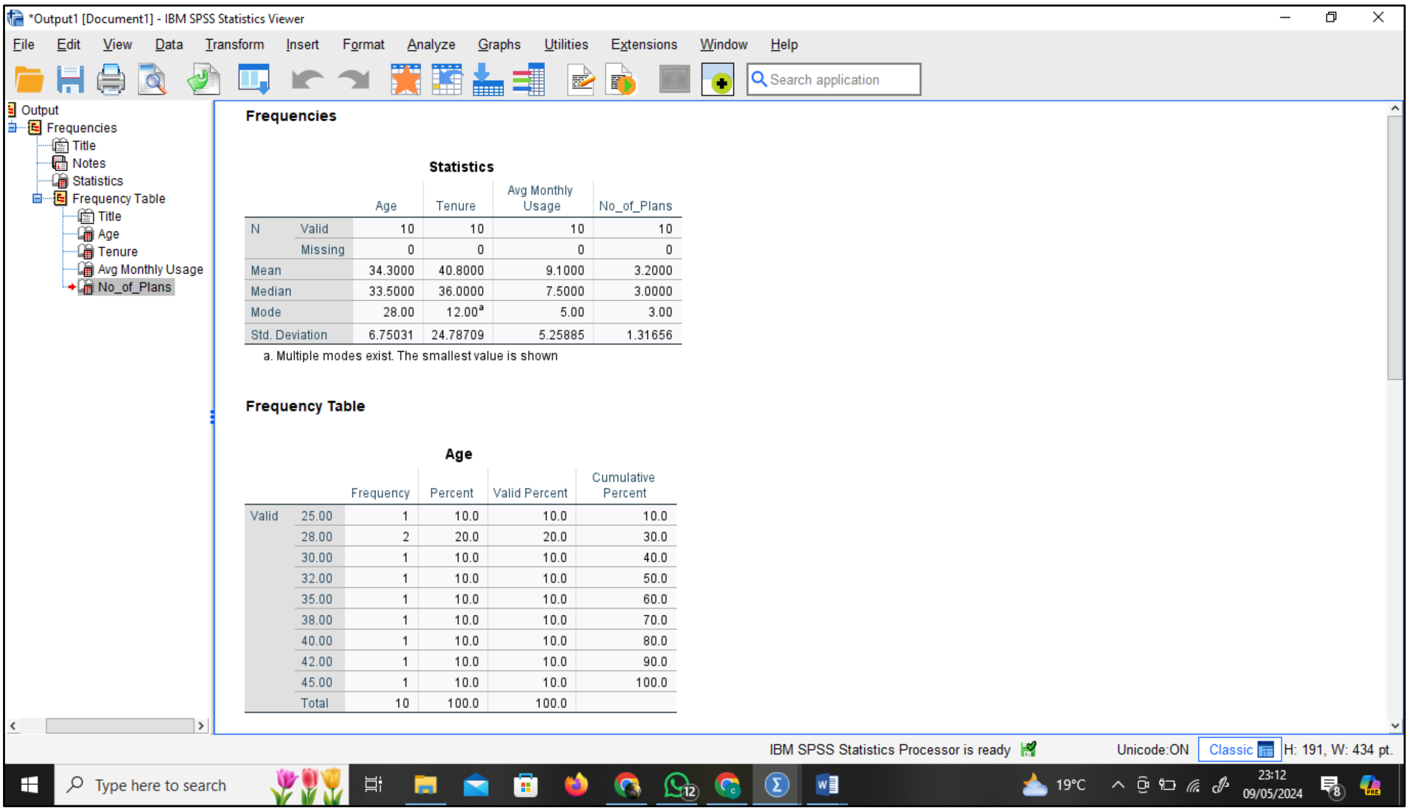
ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24MAY1148

[13]. Almohaimmeed, B. (2019). Pillars of customer retention: An empirical study on the influence of customer satisfaction, customer loyalty, customer profitability on customer retention. *Serbian Journal of Management*, *14*(2), 421-435.
[14]. Parida, B. B., & Baksi, A. K. (2011). Customer retention and profitability: CRM environment. *SCMS Journal of Indian Management*, *8*(2).
[15]. Xevelonakis, E. (2005). Developing retention strategies based on customer profitability in telecommunications: An empirical study. *Journal of Database Marketing & Customer Strategy Management*, *12*, 226-242.
[16]. Sabbeh, S. F. (2018). Machine-learning techniques for customer retention: A comparative study. *International Journal of advanced computer Science and applications*, *9*(2).
[17]. Ng, K., & Liu, H. (2000). Customer retention via data mining. *Artificial Intelligence Review*, *14*, 569-590.
[18]. Umayaparvathi, V., & Iyakutti, K. (2012). Applications of data mining techniques in telecom churn prediction. *International Journal of Computer Applications*, *42*(20), 5-9.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24MAY1148

APPENDICES

APPENDIX A: DESCRIPTIVE STATISTIC



Frequencies

Statistics

		Age	Tenure	Avg Monthly Usage	No_of_Plans
N	Valid	10	10	10	10
	Missing	0	0	0	0
Mean		34.3000	40.8000	9.1000	3.2000
Median		33.5000	36.0000	7.5000	3.0000
Mode		28.00	12.00[a]	5.00	3.00
Std. Deviation		6.75031	24.78709	5.25885	1.31656

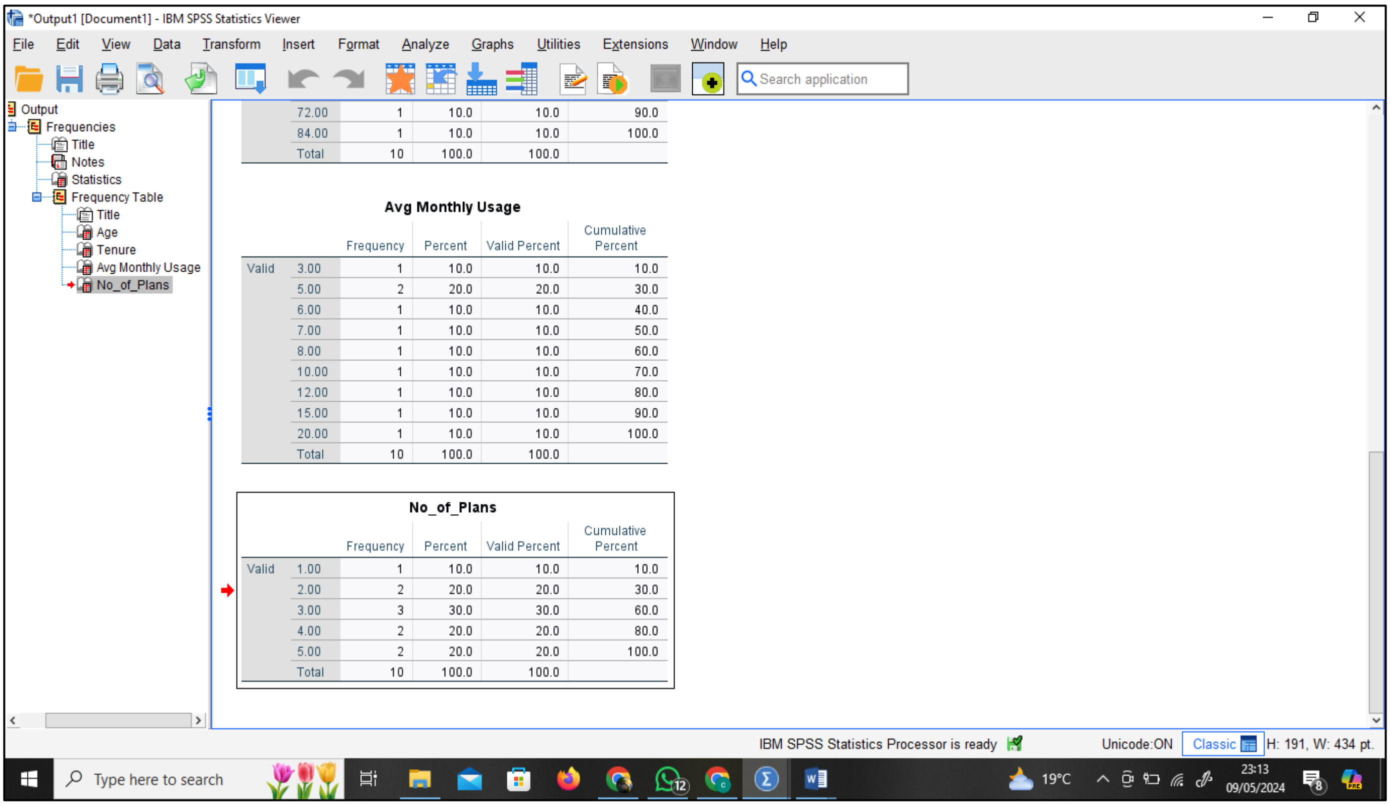
a. Multiple modes exist. The smallest value is shown

Frequency Table

Age

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	25.00	1	10.0	10.0	10.0
	28.00	2	20.0	20.0	30.0
	30.00	1	10.0	10.0	40.0
	32.00	1	10.0	10.0	50.0
	35.00	1	10.0	10.0	60.0
	38.00	1	10.0	10.0	70.0
	40.00	1	10.0	10.0	80.0
	42.00	1	10.0	10.0	90.0
	45.00	1	10.0	10.0	100.0
	Total	10	100.0	100.0	

APPENDIX B: SPSS STATISTICAL TABLES AND FIGURES



		Frequency	Percent	Valid Percent	Cumulative Percent
	72.00	1	10.0	10.0	90.0
	84.00	1	10.0	10.0	100.0
	Total	10	100.0	100.0	

Avg Monthly Usage

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	3.00	1	10.0	10.0	10.0
	5.00	2	20.0	20.0	30.0
	6.00	1	10.0	10.0	40.0
	7.00	1	10.0	10.0	50.0
	8.00	1	10.0	10.0	60.0
	10.00	1	10.0	10.0	70.0
	12.00	1	10.0	10.0	80.0
	15.00	1	10.0	10.0	90.0
	20.00	1	10.0	10.0	100.0
	Total	10	100.0	100.0	

No_of_Plans

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1.00	1	10.0	10.0	10.0
	2.00	2	20.0	20.0	30.0
	3.00	3	30.0	30.0	60.0
	4.00	2	20.0	20.0	80.0
	5.00	2	20.0	20.0	100.0
	Total	10	100.0	100.0	

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24MAY1148

APPENDIX C: CODE SNIPPETS OR SCRIPTS USED IN SPSS GRAPHS