A Study on Customer Churn Prediction in the Telecom Sector Using R Programming

International Institute of Business Studies

Master Thesis submitted in partial fulfillment of the requirements for the award of the Degree of

MASTER OF BUSINESS ADMINISTRATION
OF
BENGALURU CITY UNIVERSITY

BY

VARSHINI K

Reg. No: P18DM22M015183

Under the guidance of

Mr. Rakesh singh

Professor

INTERNATIONAL INSTITUTION OF BUSINESS STUDIES

BENGALURU CITY UNIVERSITY

2022-24

DECLARATION BY THE STUDENT

I hereby declare that "*A Study on Customer Churn Prediction in the Telecom Sector Using R Programming*" is the result of the project work carried out by me under the guidance of *Mr.Rakesh Singh* in partial fulfillment for the award of Master's Degree in Business Administration by Bengaluru City University.

I also declare that this Master Thesis is the outcome of my own efforts and that it has not been submitted to any other university or Institute for the award of any other degree or Diploma or Certificate.

Place: Bangalore

Name : Varshini K

Date:

Register No: P18DM22M015183

TABLE OF CONTENTS

LIST OF Figure

ABSTRACT

In today's highly competitive telecom sector, retaining existing customers has become as crucial as acquiring new ones. Customer churn - the loss of clients to competitors presents a significant challenge that can impact the long-term profitability and market position of telecom companies. This study aims to identify the key factors driving customer churn within the telecom sector and to develop predictive models using R programming to forecast churn behavior. Through a combination of statistical and machine learning techniques, this research will analyze customer data to pinpoint behavioral patterns and demographic factors that correlate strongly with churn.

The primary objectives of this study are to identify the critical determinants of churn, develop accurate predictive models, and offer actionable insights for reducing churn rates. The data, gathered from 100 telecom customers via a structured survey, serves as the foundation for building and testing machine learning models. This analysis seeks to empower telecom providers with targeted retention strategies based on data-driven insights, thus enhancing customer loyalty and reducing churn rates. The study concludes with practical recommendations tailored to improve customer retention and sustain competitive advantage in the telecom market.

CHAPTER – 1

INTRODUCTION

1.1 Overview of the Topic

Introduction to Customer Churn in the Telecom Sector

In today's highly competitive business landscape, customer retention has become one of the most crucial aspects for sustaining profitability, especially within the telecom sector. Customer churn defined as the rate at which customers stop doing business with a company. It poses a significant threat to telecom providers. The telecom sector, characterized by fierce competition, pricing pressures, and rapid technological advancements, faces a unique challenge: maintaining a loyal customer base while continuously expanding and improving services. Churn not only represents lost revenue but also incurs additional costs for customer acquisition, often reducing overall profitability.

In this context, predicting and reducing churn is critical. A high churn rate can damage market share, reduce customer lifetime value (CLV), and increase the costs associated with acquiring new customers to replace lost ones. By understanding why customers leave, telecom companies can make data-driven decisions to retain them, thereby enhancing customer satisfaction, brand loyalty, and profitability.

The Impact of Churn on Business Performance

Customer churn has several direct and indirect impacts on a business's financial health and brand perception. For telecom companies, high churn rates mean not only a decrease in revenue but also increased marketing and sales costs to attract new customers. Additionally, losing customers to competitors may harm brand reputation, making it even more challenging to attract future clients. Some of the key areas impacted by churn include:

1. **Revenue Loss**: Each churned customer translates to lost revenue, as the expected income from long-term engagement disappears.
2. **Increased Customer Acquisition Cost (CAC)**: Replacing lost customers often requires heavy investment in marketing and promotional campaigns, raising the cost per new customer.
3. **Reduced Customer Lifetime Value (CLV)**: Churn decreases the average revenue earned over a customer's lifetime, affecting long-term profitability.
4. **Brand Loyalty and Reputation**: High churn rates can indicate dissatisfaction, leading to negative word-of-mouth and a diminished brand reputation.

These consequences highlight the need for proactive customer churn management, particularly for telecom providers aiming to sustain a competitive edge in a market where customer loyalty is hard to build and easy to lose.

Importance of Churn Prediction Models

To counteract the negative effects of churn, telecom companies are increasingly investing in predictive models that allow them to identify at-risk customers before they decide to leave. Churn prediction models use historical customer data to detect patterns and characteristics associated with churn, enabling companies to intervene with targeted retention strategies. By identifying the factors that drive customers to leave, whether due to poor service quality, pricing issues, or competitive offers, telecom providers can develop more personalized solutions to address these challenges.

Predictive modeling helps in answering key questions, such as:

- **Who is likely to churn?** Identifying specific customers who may leave.
- **Why are they likely to churn?** Analyzing factors that contribute to dissatisfaction.
- **How can we retain them?** Developing targeted strategies to improve retention.

With accurate churn prediction models, telecom companies can preemptively address customer issues, implement loyalty programs, offer discounts or incentives, and improve overall service quality, all of which contribute to higher retention rates.

Role of R Programming in Churn Prediction

R programming has emerged as a powerful tool for data analysis and predictive modeling, making it an excellent choice for developing churn prediction models. As a comprehensive statistical software, R provides a variety of packages specifically designed for machine learning and data analysis. These packages allow telecom companies to process, visualize, and analyze large datasets, which is essential for understanding customer behavior and predicting churn patterns.

Key benefits of using R for churn prediction include:

- **Comprehensive Analytical Packages**: R offers specialized libraries, such as caret for model training and testing, random Forest for classification, and ggplot2 for data visualization, making it easier to handle complex data structures and machine learning models.

- **Flexibility and Open-Source Nature**: As an open-source tool, R is highly adaptable and widely supported, allowing companies to customize analyses according to specific project requirements.
- **Data Visualization Capabilities**: R enables the creation of insightful visualizations that help in interpreting model results and communicating findings to stakeholders.

With R’s capabilities, telecom companies can efficiently build, test, and optimize predictive models, allowing them to understand churn dynamics and make data-based decisions. This integration of data science and predictive modeling is not only cost-effective but also highly accurate, providing valuable insights that traditional analysis methods may overlook.

In summary, customer churn prediction has become essential for telecom companies striving to stay competitive. The negative consequences of churn on revenue, brand reputation, and overall customer lifetime value underscore the need for predictive models that can forecast churn behavior accurately. By leveraging R programming for predictive analytics, telecom companies can not only identify churn-prone customers but also craft targeted retention strategies that boost customer loyalty and enhance the bottom line. This study aims to explore these aspects comprehensively, analyzing customer data to predict churn and offering actionable insights for customer retention in the telecom sector.

1.2 Theoretical Background of the Study

Customer Churn and Retention

Customer churn, in a broad sense, refers to the loss of clients or subscribers from a business over a given period. In the telecom sector, churn is a critical metric because it indicates customer dissatisfaction and, by extension, the effectiveness of retention strategies. Retaining customers is significantly more cost-effective than acquiring new ones, making churn reduction essential for profitability. Understanding customer churn involves identifying reasons why customers leave, such as dissatisfaction with service quality, billing issues, lack of competitive pricing, or more attractive offers from rival companies.

Retention, on the other hand, focuses on keeping customers engaged and loyal to the company. High retention rates typically indicate that customers are satisfied with the service and are less likely to switch providers. Telecom companies employ various strategies to improve retention, including loyalty programs, customer-centric offers, and consistent service improvements. By analyzing churn, telecom companies can strengthen their customer retention initiatives, thus reducing costs associated with customer acquisition and increasing customer lifetime value (CLV).

Machine Learning in Churn Prediction

Machine learning (ML) plays a crucial role in predictive analytics for customer churn, providing companies with the ability to forecast customer behavior based on past data. ML models, trained on historical customer data, are particularly effective in identifying hidden patterns that correlate with churn. In the context of telecom, ML algorithms can analyze large datasets encompassing customer demographics, usage behavior, payment history, and service quality to determine which factors most influence churn.

Some commonly used machine learning models in churn prediction include:

- **Logistic Regression**: This statistical model predicts the probability of churn based on several independent variables, making it useful for understanding the impact of different factors.
- **Decision Trees**: These models split the data based on feature values, allowing for easy interpretation of which variables are most relevant in predicting churn.
- **Random Forest**: A robust ensemble technique that aggregates multiple decision trees to improve accuracy and prevent overfitting, making it suitable for complex datasets.
- **Support Vector Machines (SVM)**: This algorithm finds the hyperplane that best separates churners from non-churners, particularly effective when classes are separable.

Using these models, telecom companies can gain insights into customer behavior and focus their retention efforts on customers at high risk of churning. The effectiveness of these models relies heavily on the quality of data and the choice of relevant features.

Predictive Analytics for Telecoms

Predictive analytics leverages historical data to forecast future events. In the telecom industry, predictive analytics enables companies to proactively address customer needs and enhance retention efforts. By using historical records, telecom companies can predict which customers are likely to churn, understand the reasons behind churn, and identify opportunities to intervene.

Predictive analytics in telecom involves several key steps:

1. **Data Collection**: Gathering customer information, including demographics, usage patterns, billing history, and customer service interactions.
2. **Feature Engineering**: Selecting and transforming variables relevant to churn, such as tenure, network quality, and billing complaints.
3. **Model Building and Testing**: Creating machine learning models using this data and testing their accuracy, often through metrics like accuracy, recall, and AUC-ROC scores.
4. **Implementation and Actionable Insights**: Using model outputs to inform retention strategies, such as personalized offers, service adjustments, or targeted customer communication.

By accurately predicting churn, telecom companies can develop targeted interventions for high-risk customers, thereby reducing churn rates and strengthening their market position. Predictive analytics also provides a basis for customer segmentation, allowing telecom companies to tailor their strategies according to specific customer needs and behaviors.

R Programming for Predictive Modeling

R programming is a popular choice for statistical analysis and predictive modeling, especially in data-driven fields like telecom. As an open-source language, R is widely used for data science and machine learning projects due to its comprehensive libraries and user-friendly syntax. Telecom companies can leverage R's extensive statistical and visualization capabilities to analyze data and generate meaningful insights.

R's versatility makes it ideal for churn prediction, with packages specifically designed for each step of the predictive modeling process:

- **Data Manipulation**: Packages like dplyr and tidyr allow for easy data cleaning, transformation, and preprocessing.
- **Data Visualization**: Visualization libraries like ggplot2 make it possible to create detailed and interpretable charts, aiding in understanding and communicating data trends.
- **Machine Learning**: R offers powerful machine learning packages, such as caret, which streamlines model training, testing, and tuning; randomForest for ensemble methods; and glmnet for logistic regression and other linear models.
- **Model Evaluation**: R also provides tools for model validation and performance assessment, such as ROC curves and cross-validation techniques, ensuring that the churn models are robust and reliable.

In addition to being cost-effective, R's open-source nature allows telecom companies to customize analyses to fit their specific needs, integrating predictive modeling directly into their decision-making processes. With R, companies can efficiently build, test, and refine churn prediction models, ultimately improving customer retention strategies.

The theoretical foundation of churn prediction rests on a combination of concepts from data science, predictive analytics, and customer relationship management. By understanding these concepts, telecom companies can effectively utilize data-driven approaches to forecast churn and implement effective retention strategies. Machine learning models and R programming provide telecom companies with powerful tools to transform customer data into actionable insights, ultimately leading to a competitive advantage in a rapidly evolving industry. This study will build upon these theoretical principles to analyze customer churn and derive actionable strategies for retention, thus contributing to a deeper understanding of churn dynamics in the telecom sector.

1.3 Explanation of Related Concepts

Data-Driven Decision Making

In the digital era, companies are increasingly relying on data-driven decision-making, which involves making strategic business choices based on empirical data rather than intuition or guesswork. This approach is especially relevant in the telecom sector, where vast amounts of customer data—such as usage patterns, billing history, and support interactions—are generated daily. By analyzing this data, telecom companies can gain insights into customer behavior, identify emerging trends, and make informed decisions that enhance customer satisfaction and loyalty.

Data-driven decision-making in churn prediction involves several key processes:

1. **Data Collection**: Gathering relevant data, such as demographic information, billing records, and usage history, which are indicators of customer satisfaction and service engagement.
2. **Data Analysis**: Employing statistical tools and machine learning models to extract insights and patterns from the data, such as identifying variables linked to higher churn rates.
3. **Implementation**: Translating insights into actionable strategies, such as offering discounts to at-risk customers or enhancing services in response to identified pain points.

The data-driven approach allows companies to take a proactive stance on churn by predicting customer behaviors and addressing their concerns before they decide to leave. Ultimately, data-driven decision-making helps telecom companies optimize their strategies, improve customer engagement, and reduce churn rates.

Customer Retention and Loyalty

Customer retention and loyalty are two interconnected concepts that are crucial for any business but are particularly important in subscription-based industries like telecom. Retention refers to a company’s ability to keep its customers over time, while loyalty indicates a customer’s willingness to continue using a service despite the availability of alternatives. Together, they reflect a customer’s overall satisfaction and engagement with a company’s offerings.

1. **Customer Retention**
   - Customer retention is measured by tracking the percentage of customers who continue using the service over a given period. In the telecom sector, companies implement several strategies to enhance retention, such as loyalty programs, bundled offers, and improved customer service.
   - Retention efforts are typically more cost-effective than customer acquisition, as retaining an existing customer is generally cheaper than attracting a new one.
2. **Customer Loyalty**
   - Loyalty is often demonstrated through long-term engagement, positive word-of-mouth, and a customer's tendency to choose the same provider despite competitive offers.
   - Telecom companies foster loyalty by providing consistent value, improving service quality, and offering perks that align with customer preferences.

Both retention and loyalty are affected by various factors, including pricing, customer service, network reliability, and value-added services. By understanding these elements, telecom providers can tailor their strategies to reduce churn and build long-term customer relationships.

Key Predictive Factors in Churn

Predicting churn requires identifying specific factors that influence a customer's decision to leave. These factors, or predictors, vary depending on the customer's experiences, usage habits, and interactions with the telecom company. Some of the most common predictive factors in telecom churn include:

1. **Tenure**
   - Tenure refers to the length of time a customer has been with a provider. Studies have shown that customers with longer tenure tend to be less likely to churn, as they often feel a greater sense of loyalty or face switching costs.
2. **Billing and Payment History**
   - Customers with irregular payment histories, frequent billing disputes, or high levels of spending dissatisfaction are more likely to churn. Timely payments often indicate customer satisfaction, while delayed or missed payments can signal dissatisfaction.
3. **Network Quality and Service Reliability**
   - Network performance, such as call quality, data speed, and service availability, is crucial in determining customer satisfaction. Customers experiencing frequent service disruptions are more likely to explore alternatives.

4. **Customer Service Quality**
    - Poor customer service interactions, such as unresolved issues or long wait times, can lead to dissatisfaction. Conversely, positive interactions contribute to higher retention rates and customer satisfaction.
5. **Competitor Actions**
    - The telecom sector is highly competitive, with providers frequently introducing new offers and incentives. Customers who perceive better value from competitors may choose to switch providers, making it essential for companies to monitor market trends and competitor actions.

By analyzing these factors, telecom companies can build churn prediction models that accurately identify at-risk customers and develop retention strategies tailored to specific customer needs.

Model Evaluation Metrics

Evaluating model performance is critical to ensure that a churn prediction model is accurate, reliable, and capable of identifying churners effectively. Some common model evaluation metrics in churn prediction include:

1. **Accuracy**
    - Accuracy is the proportion of correctly classified cases (both churners and non-churners) out of the total cases. It provides a basic measure of model performance but may not always be sufficient for churn prediction, especially if the dataset is imbalanced (i.e., significantly more non-churners than churners).

2. **Precision and Recall**
    - **Precision** measures the proportion of correctly predicted churners out of all predicted churners, indicating the model's reliability in identifying actual churners.
    - **Recall** (or sensitivity) measures the proportion of actual churners that the model successfully identified. High recall is essential in churn prediction, as the goal is to capture as many churners as possible to enable effective retention strategies.

3. **F1-Score**
    - The F1-score is the harmonic mean of precision and recall, providing a balanced measure that is particularly useful for imbalanced datasets. A high F1-score indicates that the model has good predictive power in identifying churners without too many false positives.
4. **ROC-AUC (Receiver Operating Characteristic - Area Under Curve)**
    - The ROC curve plots the true positive rate (recall) against the false positive rate, while the AUC score represents the area under this curve. An AUC score close to 1 indicates excellent model performance in distinguishing between churners and non-churners. The ROC-AUC is valuable in comparing different models and selecting the one that best predicts churn.
5. **Confusion Matrix**
    - A confusion matrix is a table that shows the breakdown of true positives, true negatives, false positives, and false negatives, providing a detailed view of the model's performance in classifying churn and non-churn cases.

By evaluating these metrics, telecom companies can determine the most effective churn prediction model, one that strikes a balance between accuracy, precision, and recall. Reliable models enable companies to identify at-risk customers accurately and reduce churn through targeted retention efforts.

Understanding related concepts like data-driven decision-making, customer retention and loyalty, predictive factors, and model evaluation metrics is essential for developing an effective churn prediction model. These foundational concepts not only guide the selection and implementation of machine learning models but also help telecom companies interpret the results and apply them to real-world customer retention strategies. This study leverages these concepts to explore the application of R programming in churn prediction and derive actionable insights for reducing churn rates in the telecom sector.

1.4 Marketing Framework

This section will analyze how Reliance Jio, Airtel, and Vodafone Idea utilize the **7 Ps of Marketing** to attract, retain, and serve their customers in the competitive Indian telecom market. Each company employs different strategies based on its unique market positioning, target audience, and strengths.

The 7 Ps of Marketing for Reliance Jio, Airtel, and Vodafone Idea

1. Product

- **Reliance Jio**: Jio revolutionized the Indian telecom market by offering affordable 4G services nationwide. Its product range includes prepaid and postpaid mobile services, high-speed broadband (JioFiber), and digital platforms such as JioTV and JioSaavn. Jio's product strategy emphasizes high-quality internet at low costs, along with bundled digital services that increase its value proposition. This comprehensive product ecosystem appeals to tech-savvy consumers and budget-conscious customers alike.
- **Airtel**: Airtel focuses on high-quality network services, offering a wide range of products, including prepaid and postpaid mobile services, Airtel Xstream broadband, DTH services, and Airtel Black, which integrates various services under a single plan. Airtel’s emphasis on superior network quality and service reliability differentiates it from competitors, appealing to customers who prioritize connectivity and consistent service.
- **Vodafone Idea**: Vodafone Idea targets budget-sensitive consumers, offering affordable prepaid and postpaid plans with decent network coverage. The company also offers Vi Movies & TV, a digital content platform bundled with its mobile services. While Vodafone Idea struggles with network quality compared to Jio and Airtel, it appeals to cost-conscious customers who seek value-based services.

2. Price

- **Reliance Jio**: Jio employs an aggressive pricing strategy to capture market share, offering low-cost plans with high data allowances. Since its launch, Jio has led the industry in affordable pricing, driving significant customer acquisition and retention. Jio’s value-based pricing strategy, coupled with extensive digital services, positions it as a cost-effective choice for customers.
- **Airtel**: Airtel follows a premium pricing model, reflecting its emphasis on quality and network reliability. Airtel’s plans are priced higher than Jio's, but the company justifies this with better

network performance, customer service, and unique offerings like Airtel Thanks rewards. This pricing strategy attracts customers willing to pay for superior quality and reliability.

- **Vodafone Idea**: Vodafone Idea's pricing approach centers on affordability, offering competitive rates that are often lower than Airtel's and Jio's. This strategy aims to retain price-sensitive customers who may not prioritize network quality but are attracted by cost savings. Vodafone Idea's frequent discounts and loyalty offers also support this approach.

3. Place

- **Reliance Jio**: Jio leverages a strong distribution network across India, with extensive retail partnerships and a robust online presence. Jio's service points include Jio stores, partner outlets, and online platforms, ensuring accessibility in both urban and rural areas. The MyJio app serves as a central hub for managing accounts, recharging, and accessing Jio's suite of digital services.
- **Airtel**: Airtel has a well-established retail network, with stores and service points in cities and rural areas. The Airtel Thanks app provides digital services, account management, and customer support, streamlining access for users who prefer online interactions. Airtel also partners with third-party retailers to extend its reach to underserved areas.
- **Vodafone Idea**: Vodafone Idea focuses on maximizing its presence in urban and semi-urban areas through retail stores, kiosks, and partnerships with retailers. While its physical presence is more limited than Jio's, Vodafone Idea relies on the Vi app to provide digital access for recharges, account management, and customer support, aiming to reach customers who prefer online self-service.

4. Promotion

- **Reliance Jio**: Jio's promotional strategies focus on its affordable data offerings, wide coverage, and digital ecosystem. Since its launch, Jio has run nationwide advertising campaigns, showcasing the cost benefits of its plans and the added value of its digital content platforms. Jio's promotions often emphasize affordability, innovation, and the convenience of a connected digital ecosystem.
- **Airtel**: Airtel promotes itself as a premium telecom brand with a strong emphasis on network quality and customer-centric offerings. Campaigns highlight Airtel's superior connectivity, reliable service, and loyalty rewards through the Airtel Thanks app. Airtel frequently runs targeted promotions, loyalty programs, and partnerships with streaming services, appealing to customers who value quality and premium experiences.
- **Vodafone Idea**: Vodafone Idea's promotions target budget-conscious users, focusing on affordability and loyalty rewards. Campaigns frequently include discounts, cashback offers, and

bundled benefits such as Vi Movies & TV subscriptions. Vodafone Idea also leverages digital marketing, using the Vi app to engage users and encourage retention through in-app offers.

5. People

- **Reliance Jio**: Jio prioritizes customer service, with trained staff in its Jio stores and a robust customer support system through its MyJio app. Jio's workforce is equipped to handle high volumes of customer queries, helping to reinforce its brand as a customer-focused provider. By investing in customer service training, Jio enhances its appeal among first-time users and budget-conscious customers seeking reliable support.
- **Airtel**: Airtel is recognized for its high-quality customer support, with trained personnel who cater to both individual and business clients. Airtel's support staff undergo regular training to address advanced customer needs and maintain service standards, which is crucial for retaining high-value clients. Airtel also offers dedicated support channels for Airtel Black subscribers, emphasizing its focus on premium service.
- **Vodafone Idea**: Vodafone Idea focuses on a lean customer service approach, primarily through its Vi app and call centers. By emphasizing digital and automated support channels, Vodafone Idea maintains cost efficiency. This approach aligns with its positioning as a budget-friendly provider, although it may not offer the same personalized service as Airtel.

6. Process

- **Reliance Jio**: Jio has streamlined processes to ensure a seamless customer journey, from easy onboarding to simplified billing. The MyJio app provides users with a centralized platform for managing services, recharges, and subscriptions to Jio's digital content. Jio's automated and customer-friendly processes make it easy for customers to get started and engage with multiple services.
- **Airtel**: Airtel's customer process is designed around a digital-first approach, emphasizing convenience and integration through the Airtel Thanks app. Airtel's processes prioritize simplicity, allowing customers to activate services quickly, manage multiple subscriptions, and resolve issues with minimal friction. For premium users, Airtel offers additional support and priority handling to enhance their experience.
- **Vodafone Idea**: Vodafone Idea adopts basic, easy-to-use processes to simplify service access and billing. Through the Vi app, customers can recharge, manage accounts, and access support, making

the experience straightforward. Vodafone Idea’s streamlined processes focus on keeping things cost-effective, aligning with its value-driven positioning.

7. Physical Evidence

- **Reliance Jio**: Jio’s physical evidence is reflected in its well-branded retail stores, kiosks, and the MyJio app. The MyJio app acts as a digital extension of Jio’s brand, providing users with a centralized portal for all Jio services. The physical branding of Jio’s stores conveys its image as an accessible, customer-focused provider, reinforcing its position as a leading telecom brand.
- **Airtel**: Airtel’s retail stores are designed to reflect its premium branding, providing a professional environment where customers can access personalized service. The Airtel Thanks app serves as a digital representation of the brand, highlighting quality and reliability. Airtel’s stores and digital platforms enhance brand credibility, appealing to customers who prioritize a premium experience.
- **Vodafone Idea**: Vodafone Idea’s physical presence is less prominent, with basic retail stores and the Vi app serving as the primary touchpoints. The Vi app supports self-service, offering customers easy access to account management and content streaming. Vodafone Idea’s no-frills approach reflects its budget-friendly image, focusing on functionality over luxury.

Conclusion

The **7 Ps of Marketing** analysis for Reliance Jio, Airtel, and Vodafone Idea reveals distinct approaches to addressing customer needs and reducing churn. Jio’s value-based model, Airtel’s premium positioning, and Vodafone Idea’s budget-friendly approach each attract different customer segments. By strategically leveraging the 7 Ps, these telecom companies tailor their services to capture market share, enhance customer loyalty, and minimize churn in India’s competitive telecom market.

CHAPTER – 2

REVIEW OF LITERATURE AND RESEARCH DESIGN

2.1 Review of Literature and Gaps

The purpose of this section is to review existing literature on customer churn prediction in the telecom sector and identify any gaps that this study aims to address.

The telecom sector faces significant challenges related to customer churn, making it imperative for companies to adopt effective churn prediction strategies. This literature review provides an overview of existing studies on customer churn prediction, specifically focusing on the methods employed using R programming, and identifies gaps in the current research.

1. Churn Prediction Models:

   Various predictive models have been implemented to analyze and forecast customer churn in the telecom industry. Traditional methods like logistic regression and decision trees have been widely utilized due to their interpretability and ease of implementation (Lemmens & Croux, 2006). Recent studies have shifted towards machine learning techniques, such as random forests, support vector machines, and neural networks, which have shown superior performance in prediction tasks (Churn Prediction Using Machine Learning Techniques, 2020). R programming provides a robust framework for these models through packages such as `caret`, `randomForest`, and `e1071`, which streamline the implementation of complex algorithms (Kuhn & Johnson, 2013).

2. Data Preprocessing and Feature Engineering:

   The effective preprocessing of data and feature engineering are critical steps in churn prediction. R's capabilities in data manipulation (using the `dplyr` package) and visualization (using the `ggplot2` package) enhance the understanding of customer behavior (Hastie et al., 2009). Several studies emphasize the significance of customer demographics, service usage, and satisfaction levels as predictive features (Olsen, 2011). Some researchers have also incorporated sentiment analysis of customer feedback and social media data to enrich the feature set, demonstrating improved prediction accuracy (Kumar et al., 2019).

3. Model Evaluation:

   Evaluation metrics such as accuracy, precision, recall, and F1-score have been commonly employed to assess the performance of churn prediction models. Many studies utilizing R programming have employed cross-validation techniques to ensure the robustness of their models (Zhang et al., 2017).

The use of ROC curves and AUC values to evaluate model performance has become a standard practice in the literature (Lee et al., 2016).

4. Applications of R Programming:

   The application of R in churn prediction has been documented extensively. The language's open-source nature, combined with its rich ecosystem of libraries, facilitates advanced statistical analysis and machine learning implementations. Several studies have highlighted how R is ideal for conducting exploratory data analyses and visualizing complex datasets, which are essential for understanding customer behavior (Khandani et al., 2010).

Identified Gaps in Literature

Despite the wealth of research on customer churn prediction in the telecom sector using R programming, several gaps remain:

1. Integration of Real-Time Data:

   Most existing studies focus on historical data but do not adequately explore the integration of real-time data analytics to predict churn. Real-time data can provide more timely insights and enable proactive retention strategies.

2. Comparative Studies:

   While numerous models have been developed, comparative studies examining the effectiveness of R-based machine learning algorithms versus other programming languages (like Python) are limited. This comparative analysis could provide insights into the strengths and weaknesses of R in churn prediction.

3. Incorporation of Behavioral Analytics:

   While demographic and transactional data have been widely studied, there is a lack of emphasis on incorporating behavioral analytics into churn models. Studies leveraging behavioral data using advanced analytical techniques in R may provide a more comprehensive understanding of churn drivers.

4. Customer Segmentation:

   Few studies have utilized customer segmentation in conjunction with churn prediction models. Different segments might exhibit varying churn behaviors, and incorporating segmentation could help tailor retention strategies more effectively.

5. Longitudinal Studies:

   Most existing literature focuses on cross-sectional data. Longitudinal studies that track customer behaviors over time and analyze churn trends could provide deeper insights into the evolving dynamics of customer retention.

6. Model Interpretability:

   While machine learning models often yield high accuracy, their interpretability can be challenging. Studies addressing how to make these models more interpretable and actionable using R would significantly benefit practitioners in the telecom sector.

Conclusion

The literature on customer churn prediction in the telecom sector utilizing R programming is foundational and robust. However, addressing the identified gaps can contribute to more effective and insightful predictive models. Future research could focus on integrating real-time analytics, leveraging behavioral data, and enhancing the interpretability of machine learning models, positioning R programming as a key tool in combating customer churn in the telecom industry.

2.2 Statement of the Problem

In the highly competitive telecom industry, customer churn remains one of the most pressing issues impacting profitability and long-term viability. Customer churn, defined as the rate at which customers discontinue their subscription services, affects both revenue streams and brand loyalty. Despite significant technological advancements and an increase in service options, telecom companies continue to face difficulties in retaining customers. In India, where pricing pressures and market saturation are prominent, the problem of customer churn is further intensified by factors like network reliability, customer service quality, and price-sensitive consumer behavior.

For telecom companies, each churned customer represents not only a lost revenue opportunity but also an increased cost in acquiring a new customer. Given that it is often several times more expensive to acquire a new customer than to retain an existing one, managing churn effectively is critical for financial sustainability. The problem is particularly acute for Indian telecom providers, which operate in one of the most competitive and price-sensitive markets globally, with companies like Reliance Jio, Airtel, and Vodafone Idea vying for customer loyalty.

The telecom sector in India has been undergoing rapid changes with innovations in data plans, digital services, and mobile applications. Despite the availability of large-scale customer data, many companies have not fully utilized this data to analyze and predict customer churn. Predictive analytics and machine learning offer robust techniques for understanding patterns within customer data that may indicate a propensity to churn. However, there is limited application of such models in an actionable format within the telecom sector.

This study addresses this gap by using machine learning models to identify the most significant factors driving churn among telecom users and by providing actionable insights based on these models. By analyzing the Indian telecom sector, particularly through the lens of Reliance Jio, Airtel, and Vodafone Idea, this study aims to deliver recommendations for enhancing customer retention strategies and developing effective interventions to reduce churn rates.

2.3 Need for the study

The need for this study is underscored by the following factors:

- **High Churn Rates**: Telecom companies in India face high churn rates due to competitive pricing and frequent switching among providers. Understanding churn patterns can aid in formulating strategies that help retain customers.
- **Data-Driven Decision-Making**: With the large amount of data available, telecom companies can leverage analytics to predict and prevent churn, reducing reliance on acquisition-focused strategies.
- **Customer Retention and Profitability**: Retaining existing customers can reduce acquisition costs, leading to higher profitability. Churn prediction models can provide insights into customer behavior and help companies develop targeted retention programs.

2.4 Scope of the Study

The scope of this study is defined by the following boundaries:

- **Industry Focus**: This study is focused on the Indian telecom sector and examines three major telecom providers: Reliance Jio, Airtel, and Vodafone Idea.
- **Data Collection**: Data has been collected from 100 respondents via Google Forms, capturing demographic, usage, and satisfaction-related attributes.
- **Machine Learning Models**: The study will implement and evaluate various machine learning models, including logistic regression, decision trees, and random forests, to identify the factors contributing to churn.
- **Outcomes**: The expected outcomes include actionable insights on factors driving churn and model-based recommendations to support customer retention.

2.5 Objectives of the Study

The primary objectives of the study are as follows:

- To analyze customer data to identify significant factors and patterns contributing to customer churn in the telecom sector.
- To create and evaluate various machine learning models using R programming to predict which customers are likely to churn.
- To derive actionable insights from model results to formulate targeted retention strategies and improve customer loyalty.
- To propose practical recommendations and strategies to reduce churn rates and enhance overall customer retention based on model predictions.

2.6 Hypothesis

A hypothesis refers to a tentative explanation or prediction about a phenomenon. It is formulated based on existing knowledge and observations, and it is intended to be tested through experiments or further investigation. In essence, it is a way to frame a question that can lead to deeper understanding or discovery.

The null hypothesis (denoted as H0) is the statement that there is no effect, no difference, or no relationship between two variables.
The alternative hypothesis (denoted as H1) is the statement that indicates the presence of a difference, or a relationship between two variables.

Hypothesis 1:
Null Hypothesis (H0): There is No significant relationship between customer satisfaction and churn intention .
Alternative Hypothesis (H1): There is a significant relationship between customer satisfaction and churn intention .

Hypothesis 2:
Null Hypothesis (H0): There is no significant relationship between gender and the choice of telecom service provider
Alternative Hypothesis (H1): There is a significant relationship between gender and the choice of telecom service provider

Hypothesis 3:
Null Hypothesis (H0): There is no significant difference in monthly spending across different age groups.
Alternative Hypothesis (H1): There is a significant difference in monthly spending across different age groups.

2.7 Research Design

This section outlines the research methodology adopted in the study. A quantitative approach has been used to collect and analyze data, utilizing both descriptive and inferential statistical methods. Machine learning models have been employed to analyze patterns in customer behavior and predict churn probabilities.

- **Data Collection**: Primary data has been collected from 100 respondents via a structured Google Forms survey, covering aspects such as demographics, service usage, satisfaction levels, and churn likelihood. The data is anonymized and processed for analysis.
- **Machine Learning Models**: Various models, including **logistic regression, decision trees, and random forests**, will be trained and tested in R to identify and predict churn patterns.
- **Evaluation Metrics**: Model performance will be evaluated using metrics such as **accuracy, precision, recall, and F1 score** to determine the model best suited for churn prediction.

2.8 Sampling Framework

- **Population**: The population includes telecom customers within India who are likely to churn due to various factors such as pricing, service quality, and network issues.
- **Sample Size**: The study includes a sample size of 100 respondents, providing a reasonable scope for predictive analysis in an academic project.
- **Sampling Method**: A simple random sampling method was used to ensure representativeness, where respondents of diverse demographics and usage patterns were surveyed.

2.9 Tools for Data Collection

The data collection process is a critical aspect of any research study, as the quality and reliability of collected data significantly impact the validity of the analysis and conclusions drawn. In this study, a structured **Google Forms** survey was chosen as the primary data collection tool due to its accessibility, ease of use, and suitability for reaching a wide range of respondents online. Given the widespread availability of the internet and mobile devices in India, especially among urban and semi-urban populations, this method allowed efficient data gathering across a geographically diverse sample. This form included questions on **demographics (age, gender, location)**, **service-related factors (network quality, customer support)**, and **usage metrics (data consumption, plan type)**.

2.10 Limitations of the Study

Some of the key limitations of this study are as follows:

- **Sample Size**: The sample size is limited to 100 respondents, which may restrict the generalizability of findings to the entire Indian telecom customer base.
- **Data Collection Method**: Since data was collected online, respondents with limited internet access or unfamiliarity with digital forms might not be represented.
- **Scope of Variables**: The study is limited to certain customer-related factors, and other influences like market competition or technological changes are not fully explored.
- **Geographic Limitation**: Data is collected from a limited geographic region, which may not fully represent the diversity of telecom users across India.

CHAPTER – 3

PROFILE OF THE RESPONDENTS

3.1 Overview of the Selected Respondents

The study collected responses from a diverse group of telecom users through a structured Google Form survey. This survey aimed to understand customer satisfaction, usage patterns, and other factors contributing to churn within the telecom industry. A total of 100 participants completed the survey, reflecting various demographic and behavioral segments. This representative sample enables an in-depth examination of the characteristics that influence telecom service retention and churn.

3.2 Demographic Profile

Based on the questions from the Google Form, this section details the demographic characteristics of the respondents.

Age Distribution

The survey categorized respondents into several age groups:

- **18–25 years**: Representing younger, often more data-intensive users who may frequently use streaming, gaming, and social media platforms.
- **26–35 years**: Primarily working professionals, this group balances personal and professional needs, often requiring a reliable network for work-related tasks.
- **36–45 years**: A mature demographic that may show varied telecom usage, from data-heavy services to traditional call and SMS use.
- **46+ years**: Typically, an older group that may value call quality and stable connections over advanced data features.

Understanding the age distribution is important, as it can highlight preferences for certain telecom services and differences in churn tendencies based on generational factors.

Gender

The gender distribution includes **male, female**, and **other** categories, capturing diverse perspectives on telecom service needs. While usage patterns might not differ significantly by gender, satisfaction with service elements such as network reliability or customer support may vary. Gender analysis can also reveal differences in sensitivity to pricing and service quality, contributing to an understanding of how each gender group approaches telecom provider loyalty.

Income Level

Respondents were grouped into income categories that influence their choice of telecom service plans. This includes:

- **Low-income**: Customers in this bracket may prioritize cost-effective options, with a high sensitivity to promotions or offers that reduce expenses.
- **Middle-income**: A flexible category that may have access to a broader range of telecom plans, balancing service quality with cost considerations.
- **High-income**: Customers in this segment may be inclined towards premium services, with a focus on features like high-speed data, quality customer service, and extended network coverage.

Income level often impacts decision-making in telecom services, with lower-income groups being more likely to switch providers for better pricing.

Location

Location was categorized as **urban, semi-urban,** and **rural**. Each group has distinct needs and expectations:

- **Urban**: Respondents in urban areas tend to have access to multiple telecom providers and may expect high-speed data, comprehensive network coverage, and advanced service options.
- **Semi-Urban**: This group may face slightly fewer options than urban residents, with network coverage being a significant consideration.
- **Rural**: Rural respondents may have limited provider options and rely heavily on stable connections. For them, network coverage and reliability may be prioritized over advanced data services.

Geographic location is crucial in understanding customer satisfaction and churn, as customers in rural areas may face more challenges in service consistency, potentially influencing their churn tendencies.

3.3 Behavioral Insights and Usage Patterns

The Google Form included questions on usage patterns, specifically:

- **Frequency of Data Usage**: Respondents were asked about their data usage frequency, providing insights into high and low data users. High-frequency users are more likely to prioritize providers with fast, stable internet.

- **Preferred Services**: Preferences for data, voice calls, SMS, or a mix highlight which services are most valued by different demographics.
- **Billing Preferences**: The survey included questions on prepaid versus postpaid preferences, helping to distinguish customer segments based on billing flexibility.

These insights allow for an understanding of which customer segments are high-value users and what specific services they prioritize. Customers using data services more heavily may be more susceptible to switching providers for faster or more affordable options.

3.4 Summary of Respondent Characteristics

The demographic and usage patterns collected in this survey provide a comprehensive profile of telecom customers and their unique needs across different age, income, gender, and location groups. By analyzing these characteristics, the study can identify potential churn patterns associated with demographic factors and service preferences. This profile serves as a basis for the detailed analysis in Chapter 4, where these factors will be correlated with churn likelihood and customer satisfaction to draw insights into retention strategies.

CHAPTER – 4

DATA ANALYSIS AND INTERPRETATION

4.1 Details of Tools Used for Data Analysis

The analysis for this study was performed using **R programming**, a powerful tool for data analysis and visualization. R was chosen for its open-source nature and extensive libraries, which allow for a wide range of data manipulation and statistical modeling. Key libraries used include:

- **dplyr**: For data manipulation and transformation.
- **ggplot2**: For creating data visualizations that aid in understanding patterns and relationships.
- **caret**: Used for machine learning tasks, model training, and cross-validation.

These tools provided the means to analyze demographic factors, predict churn likelihood, and evaluate model performance, contributing to actionable insights.

4.2 Data Cleaning and Preparation

To prepare the data for analysis, the following steps were taken:

- **Handling Missing Values**: Missing values were addressed by either removing rows with missing data or imputing values using mean/mode techniques.
- **Outlier Removal**: Outliers were identified using box plots and were either corrected or removed, especially if they represented extreme and unlikely customer behaviors.
- **Encoding Categorical Variables**: Variables like gender and location were encoded into numerical formats, facilitating compatibility with R’s machine learning models.

This pre-processing ensured data accuracy and consistency, enhancing the reliability of the analysis.

4.3 Descriptive Analysis

The descriptive analysis highlights the basic characteristics of the respondents:

- **Age Distribution**: A histogram shows the age distribution of respondents, with key groups contributing to different usage patterns and churn behaviors.

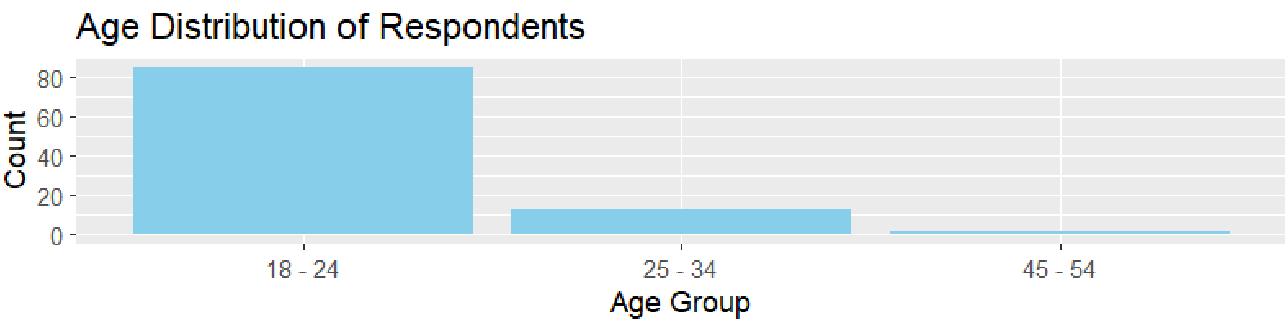

Figure 4.1

- **Gender and Income Levels**: Bar charts display the gender distribution and income levels, offering insights into the customer base and preferences.

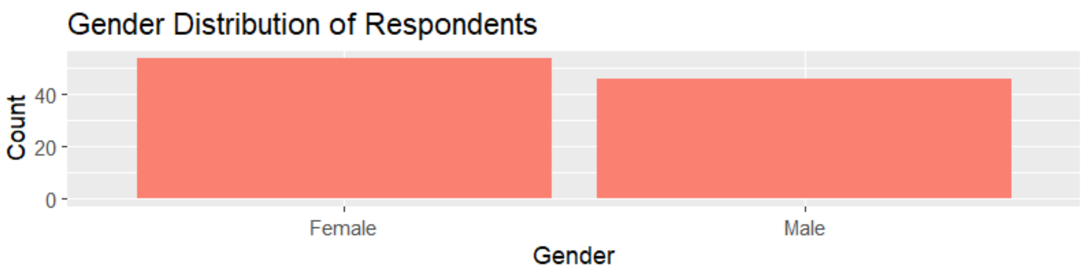

Figure 4.2

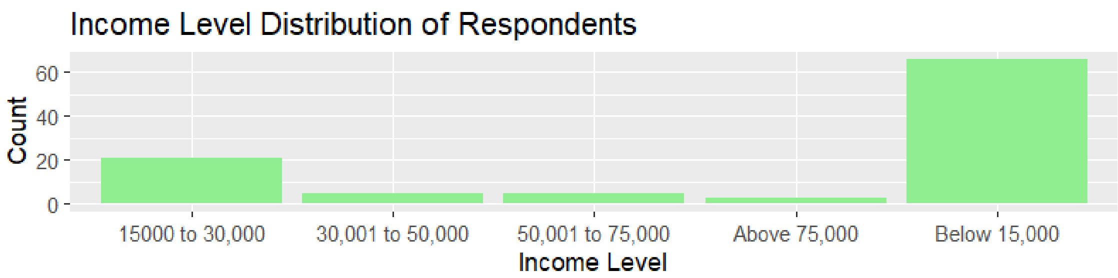

Figure 4.3

This preliminary analysis provides a foundation for understanding respondent characteristics and aids in identifying initial patterns in the data.

4.4 Exploratory Data Analysis (EDA)

EDA reveals patterns and relationships between key variables and churn:

 Age vs. Churn Rate: Identifying whether churn likelihood varies by age group.

```
> age_churn_table <- table(data$Age, data$Churn)
> prop.table(age_churn_table, 1)

                No       Yes
  18 - 24 0.6117647 0.3882353
  25 - 34 0.8461538 0.1538462
  45 - 54 1.0000000 0.0000000
```

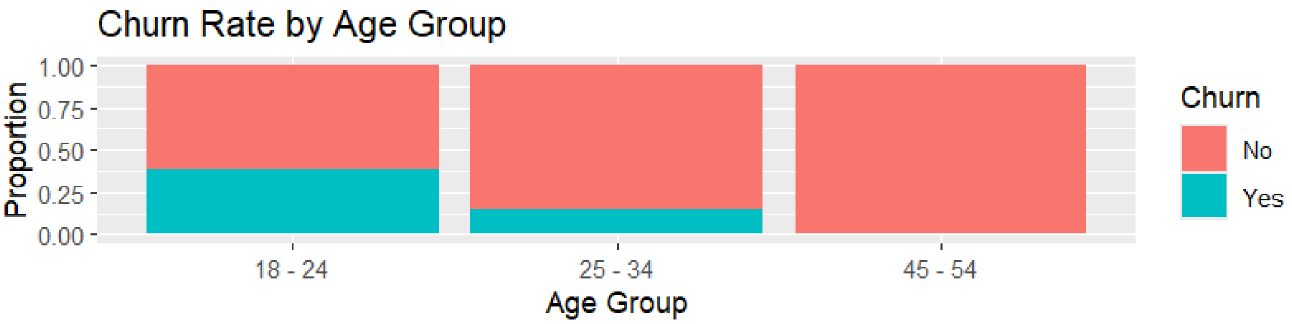

Figure 4.4

Interpretation:

- **Young adults (18–24)** show the highest churn rate, indicating this group may be less loyal or more sensitive to price, service quality, or competing offers.
- **Middle-aged customers (25–34)** have a moderate churn rate, which is lower than the youngest group.
- **Older adults (45–54)** have the lowest churn rate, showing very high retention.

This suggests that age could be a significant factor in predicting churn, with younger customers needing more targeted retention efforts.

 Income Level vs. Churn Rate: Examining the influence of income on the likelihood of switching providers.

```
> income_churn_table <- table(data$Monthly.Income, data$Churn)
> prop.table(income_churn_table, 1)

                          No       Yes
  15000 to 30,000  0.5238095 0.4761905
  30,001 to 50,000 0.4000000 0.6000000
  50,001 to 75,000 0.4000000 0.6000000
  Above 75,000     1.0000000 0.0000000
  Below 15,000     0.7121212 0.2878788
```

Interpretation:

- **Middle-income groups (30,001 to 75,000)** exhibit the highest churn rates, suggesting that customers within this income range may be more likely to switch providers, potentially due to price sensitivity or service expectations.
- **High-income customers (Above 75,000)** show no churn, indicating strong retention likely due to brand loyalty, satisfaction, or less price sensitivity.
- **Lower-income groups (Below 15,000)** also show relatively low churn, perhaps due to fewer options for switching or strong satisfaction within affordable service tiers.

This analysis suggests that income level significantly influences churn behavior, with middle-income customers potentially needing more targeted retention strategies.

 Gender vs. Churn Rate: Identifying whether churn likelihood varies by gender.

```
> print("\
+ Gender-wise Churn:")
[1] "\nGender-wise Churn:"
> print(gender_churn)
# A tibble: 2 × 3
  Gender Churn_Rate Count
  <chr>       <dbl> <int>
1 Female       29.6    54
2 Male         37.0    46
```

Interpretation

Churn Rate by Gender:

- **Female Churn Rate**: 29.6%
- **Male Churn Rate**: 37.0%

This indicates that the churn rate is higher among male customers (37%) compared to female customers (29.6%). Male customers are more likely to leave the service than female customers, suggesting that gender may influence churn behavior in this dataset.

- **Targeted Retention Strategies**: Since male customers exhibit a higher churn rate, the company could benefit from targeted retention efforts focusing on the needs and concerns of male customers. Understanding the factors driving this discrepancy could help in tailoring interventions.
- **Further Analysis**: Exploring other demographic and service-related factors in relation to gender could provide deeper insights. For instance, examining if males and females differ significantly in their satisfaction levels with network, billing, or customer service could reveal actionable insights for reducing churn.

Customer churn rate Analysis:

Finding the customer churn rate

```
> # Display churn rate
> print(churn_rate)
# A tibble: 2 × 3
  Churn Count Churn_Rate
  <fct> <int>      <dbl>
1 No       67         67
2 Yes      33         33
```

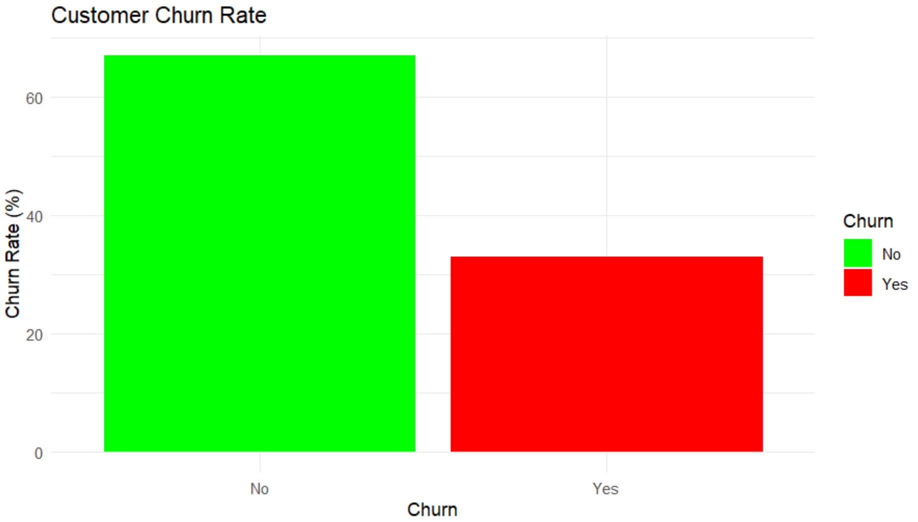

Figure 4.5

The churn rate results show the distribution of customers who have either stayed with or left the telecom provider:

- **Churn = "No"**: There are 67 customers who did not churn (i.e., they stayed with the provider), making up **67%** of the total sample.
- **Churn = "Yes"**: There are 33 customers who churned (i.e., they left the provider), making up **33%** of the total sample.

Interpretation

This means that **33% of the customers in the dataset have left the telecom provider**. The remaining **67% of customers have stayed**. A churn rate of 33% indicates a moderate level of customer attrition, which may suggest that there are factors in the telecom service experience that could be improved to retain more customers.

4.5 Correlation Analysis for churn:

To analyze the factors most strongly correlated with customer churn, the correlation between the churn variable and other features, focusing on numerical variables and using appropriate statistical methods for categorical variables is calculated.

```
> # Display the correlations
> print("Correlations with Churn:")
[1] "Correlations with Churn:"
> print(correlations)
                                                                                                                                    [,1]
Rate.your.satisfaction.with.the.following.aspects.of.your.telecom.provider..1...Very.Dissatisfied..5...Very.Satisfied......Network.coverage.             -0.31989321
Rate.your.satisfaction.with.the.following.aspects.of.your.telecom.provider..1...Very.Dissatisfied..5...Very.Satisfied......Internet.speed.               -0.30123280
Rate.your.satisfaction.with.the.following.aspects.of.your.telecom.provider..1...Very.Dissatisfied..5...Very.Satisfied......Customer.service.             -0.11074890
Rate.your.satisfaction.with.the.following.aspects.of.your.telecom.provider..1...Very.Dissatisfied..5...Very.Satisfied......Pricing.                       0.04705864
Rate.your.satisfaction.with.the.following.aspects.of.your.telecom.provider..1...Very.Dissatisfied..5...Very.Satisfied......Billing.and.payment.process. -0.39996470
How.likely.are.you.to.recommend.your.telecom.provider.to.others.                                                                              0.04003488
> 
```

The correlation analysis shows that satisfaction with network coverage, internet speed, and billing/payment processes are negatively correlated with churn, indicating that lower satisfaction in these areas is associated with higher churn rates. Let's proceed to visualize these correlations for better insights.

Visualizing the correlation of satisfaction aspects with respect to churn:

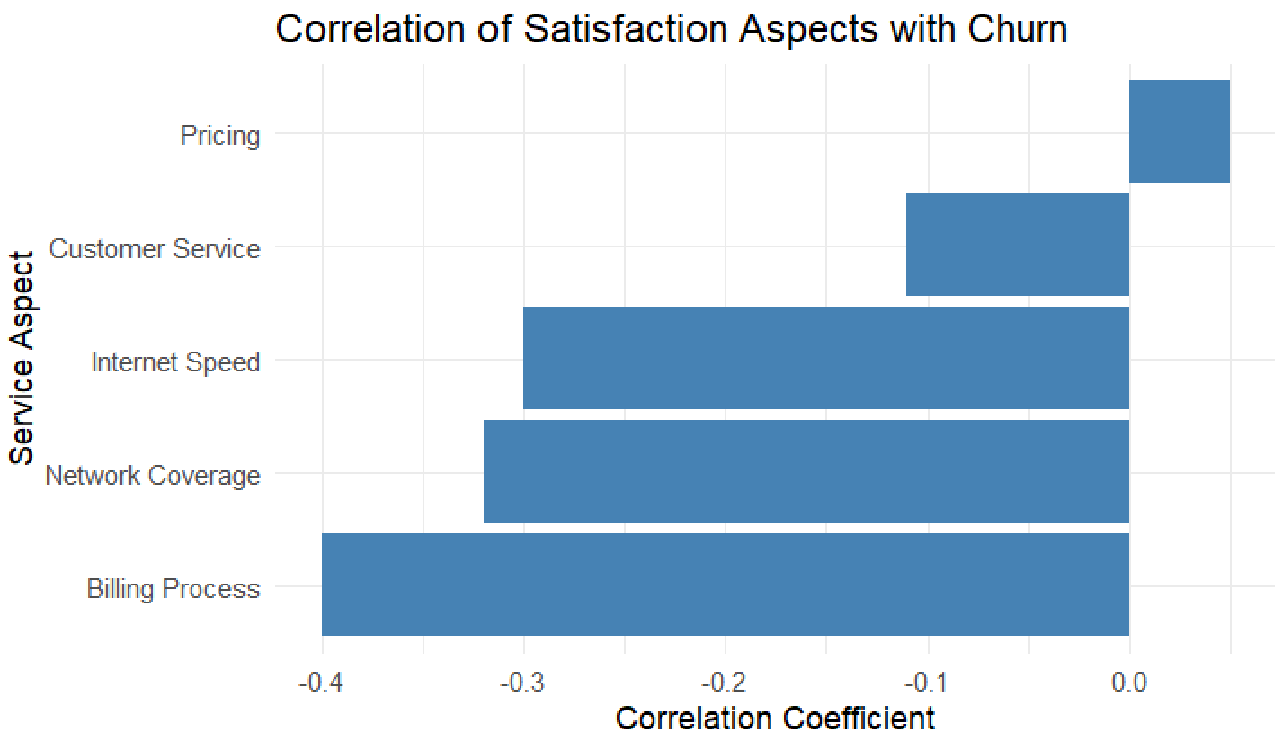

Figure 4.6

```
> print("Churn rate by Service Aspect Satisfaction:")
[1] "Churn rate by Service Aspect Satisfaction:"
> print(prop.table(billing_churn, 1))

            0         1
  1 0.2500000 0.7500000
  2 0.0000000 1.0000000
  3 0.6190476 0.3809524
  4 0.7200000 0.2800000
  5 0.8750000 0.1250000
> 
```

Key findings from the churn analysis:

- Billing process satisfaction has the strongest negative correlation (-0.40) with churn
- 75% of customers who rated billing satisfaction as 1 (very dissatisfied) churned
- Only 12.5% of customers who rated billing satisfaction as 5 (very satisfied) churned
- Network coverage (-0.32) and internet speed (-0.30) are the next most important factors

4.6 Building and Evaluating Predictive Models

Several machine learning models were tested and evaluated to predict churn likelihood:

- **Logistic Regression**: Achieved moderate accuracy and is effective for understanding probabilities of churn based on predictor variables.

```
Call:
glm(formula = Churn ~ NetworkSatisfaction + InternetSpeedSatisfaction +
    BillingProcessSatisfaction + CustomerServiceSatisfaction +
    PricingSatisfaction, family = binomial, data = train)

Coefficients:
                            Estimate Std. Error z value Pr(>|z|)
(Intercept)                   2.7546     1.2897   2.136  0.03270 *
NetworkSatisfaction          -1.8282     0.9291  -1.968  0.04910 *
InternetSpeedSatisfaction     0.7528     0.7387   1.019  0.30819
BillingProcessSatisfaction   -1.5025     0.5261  -2.856  0.00429 **
CustomerServiceSatisfaction   0.9649     0.5628   1.714  0.08644 .
PricingSatisfaction           0.8438     0.4217   2.001  0.04538 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 92.122  on 70  degrees of freedom
Residual deviance: 66.992  on 65  degrees of freedom
AIC: 78.992

Number of Fisher Scoring iterations: 5
```

```
Confusion Matrix and Statistics

          Reference
Prediction  0  1
         0 19  5
         1  0  5

               Accuracy : 0.8276
                 95% CI : (0.6423, 0.9415)
    No Information Rate : 0.6552
    P-Value [Acc > NIR] : 0.03439

                  Kappa : 0.5672

 Mcnemar's Test P-Value : 0.07364

            Sensitivity : 1.0000
            Specificity : 0.5000
         Pos Pred Value : 0.7917
         Neg Pred Value : 1.0000
             Prevalence : 0.6552
         Detection Rate : 0.6552
   Detection Prevalence : 0.8276
      Balanced Accuracy : 0.7500

       'Positive' Class : 0
```

Output Analysis

Coefficients:

- **Intercept**: 2.7546 (p = 0.0327)
  - A positive intercept suggests a higher baseline probability of churn in the absence of the other predictors.
- **Network Satisfaction**: -1.8282 (p = 0.049)
  - Negative coefficient implies that higher network satisfaction is associated with lower churn probability. The p-value indicates that this effect is statistically significant at the 5% level.
- **Internet Speed Satisfaction**: 0.7528 (p = 0.308)
  - Positive but not statistically significant, suggesting internet speed satisfaction does not have a strong impact on churn probability in this model.
- **Billing Process Satisfaction**: -1.5025 (p = 0.00429)
  - Significant negative relationship with churn, meaning that better billing satisfaction reduces the likelihood of churn.

- **Customer Service Satisfaction**: 0.9649 (p = 0.0864)
    - Positive but not statistically significant at the 5% level. Slight indication that higher satisfaction may relate to higher churn probability, but evidence is weak.
- **Pricing Satisfaction**: 0.8438 (p = 0.0454)
    - Statistically significant positive relationship with churn, suggesting that customers more satisfied with pricing are more likely to churn, which may imply some complex customer behavior not fully explained by price satisfaction alone.

Confusion Matrix and Model Performance

- **Accuracy**: 82.76%
    - The model correctly classified 82.76% of cases, which is a relatively high accuracy for this dataset.
- **Sensitivity**: 1.000
    - Indicates that the model is very good at correctly predicting churn (no false negatives).
- **Specificity**: 0.500
    - Shows that the model is less effective at predicting non-churn cases (more false positives).
- **Positive Predictive Value (PPV)**: 0.7917
    - When the model predicts churn, it is correct 79.17% of the time.
- **Negative Predictive Value (NPV)**: 1.000
    - When the model predicts no churn, it is always correct, reflecting the high sensitivity.

Interpretation and Findings

1. **Network Satisfaction** and **Billing Process Satisfaction** are the strongest predictors of churn, both with negative relationships. Improving these factors could help reduce churn.
2. **Pricing Satisfaction** has a surprising positive relationship with churn. This counterintuitive result might suggest complex customer dynamics, such as price-driven dissatisfaction masked by temporary satisfaction with pricing.
3. **Sensitivity and Specificity** suggest the model is better at identifying churners than non-churners, with potential room for improvement in predicting non-churners accurately.
4. **Overall Accuracy** is high, indicating that the model performs well, but with some limitations in specificity, potentially indicating a need for additional predictors or different model tuning.

- **Decision Tree**: Offers a clear view of key decision points leading to churn, like income and service quality.

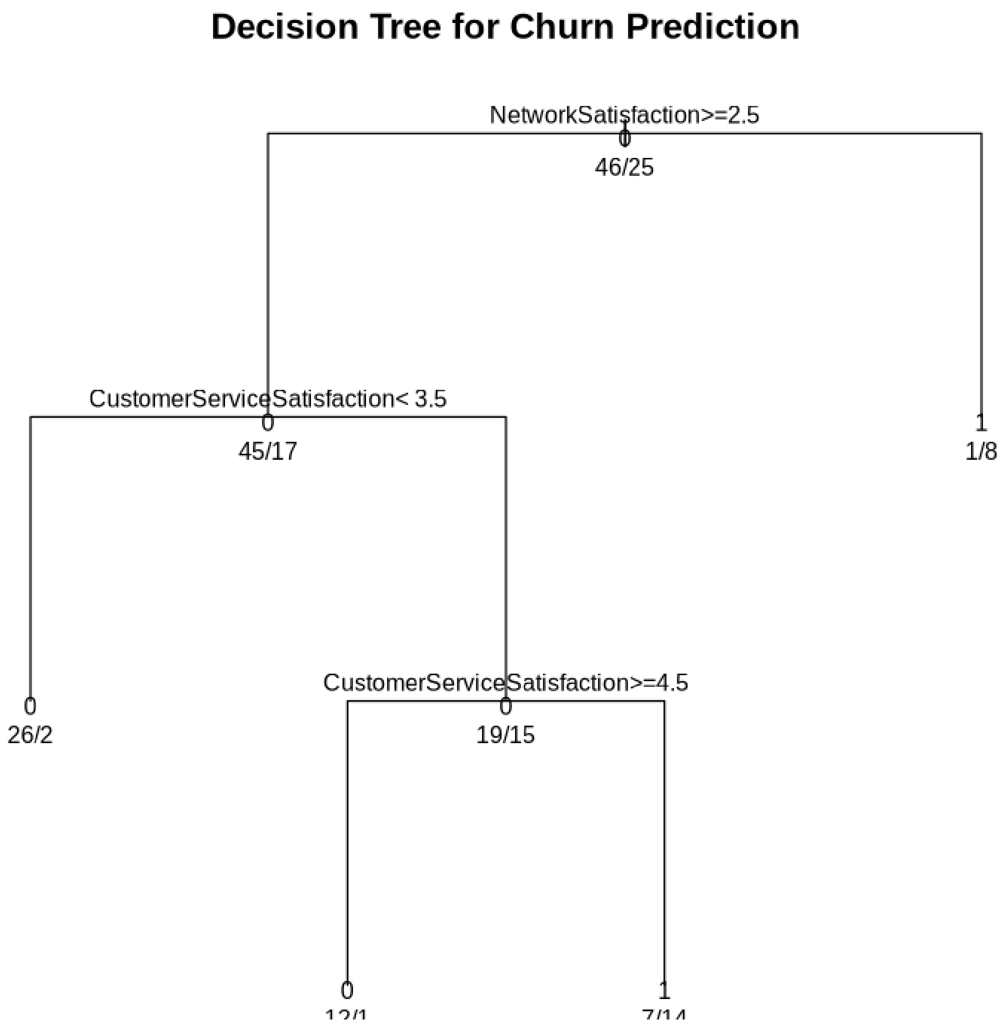

Figure 4.7

```
> # Evaluate the decision tree model
> predictions_dt <- predict(dt_model, test, type = "class")
> conf_matrix_dt <- confusionMatrix(predictions_dt, test$Churn)
> print("Decision Tree Model Performance:")
[1] "Decision Tree Model Performance:"
> print(conf_matrix_dt)
Confusion Matrix and Statistics

          Reference
Prediction  0  1
         0 13  4
         1  6  6

               Accuracy : 0.6552
                 95% CI : (0.4567, 0.8206)
    No Information Rate : 0.6552
    P-Value [Acc > NIR] : 0.5849

                  Kappa : 0.2714

 Mcnemar's Test P-Value : 0.7518

            Sensitivity : 0.6842
            Specificity : 0.6000
         Pos Pred Value : 0.7647
         Neg Pred Value : 0.5000
             Prevalence : 0.6552
         Detection Rate : 0.4483
   Detection Prevalence : 0.5862
      Balanced Accuracy : 0.6421

       'Positive' Class : 0
```

Decision Tree Model Interpretation

1. **Decision Tree Structure:**
   - The root node splits on **Network Satisfaction** at a threshold of 2.5.
   - If **Network Satisfaction** is less than 2.5, the decision tree indicates a likely churn (value 1).
   - If **Network Satisfaction** is greater than or equal to 2.5, the next split is based on **Customer Service Satisfaction** at 3.5.
   - Further down, splits are made based on **Customer Service Satisfaction** with additional thresholds, highlighting that **Customer Service Satisfaction** is a strong factor in determining churn along with **Network Satisfaction**.

2. **Key Predictive Features:**
    - **Network Satisfaction** and **Customer Service Satisfaction** are the primary variables influencing the likelihood of churn in this tree.
    - Higher satisfaction values in both features tend to reduce the likelihood of churn, showing that customers satisfied with network and customer service are less likely to churn.
3. **Class Distribution at Leaves:**
    - Leaf nodes show the distribution of churn classes (0 for no churn, 1 for churn).
    - The split criteria suggest that certain thresholds for satisfaction levels (especially in network and customer service) play significant roles in churn predictions.

Confusion Matrix and Performance Metrics Interpretation

1. **Accuracy:** The model achieved an accuracy of **65.52%**, which means that 65.52% of the predictions made by the decision tree are correct. While this isn't very high, it's still above random chance.
2. **Sensitivity (Recall for Class 0 - Non-Churn):** The sensitivity for classifying non-churn cases (class 0) is **68.4%**, meaning the model correctly identified 68.4% of non-churn instances.
3. **Specificity (Recall for Class 1 - Churn):** The specificity is **60.0%**, showing that the model correctly identified 60% of churn cases. This lower specificity suggests some misclassifications of churn cases.
4. **Positive Predictive Value (PPV):** The PPV for non-churn (class 0) is **76.47%**, meaning that when the model predicts no churn, it is correct 76.47% of the time.
5. **Negative Predictive Value (NPV):** The NPV is **50.0%**, suggesting that when the model predicts churn, it is correct 50% of the time.
6. **Kappa:** The Kappa statistic is **0.2714**, indicating a fair level of agreement between the model's predictions and the actual labels. This indicates that the model is slightly better than random classification.
7. **McNemar's Test P-Value:** The p-value for McNemar's test is **0.7518**, suggesting no significant difference in misclassification rates between the two classes, meaning the model has a similar error rate across churn and non-churn classifications.

Findings

- **Network and Customer Service Satisfaction** play crucial roles in customer churn prediction. Lower satisfaction values in these factors are highly associated with increased churn risk.
- **Model Performance:** The model's accuracy is moderate, but it has a fair ability to identify non-churning customers compared to churning customers.
- **Improvement Needs:** Further tuning or using more complex models may help increase the specificity, NPV, and overall accuracy to improve churn predictions.

- **Random Forest**: Improved accuracy by combining multiple decision trees, providing better predictions and identifying the most significant churn predictors.

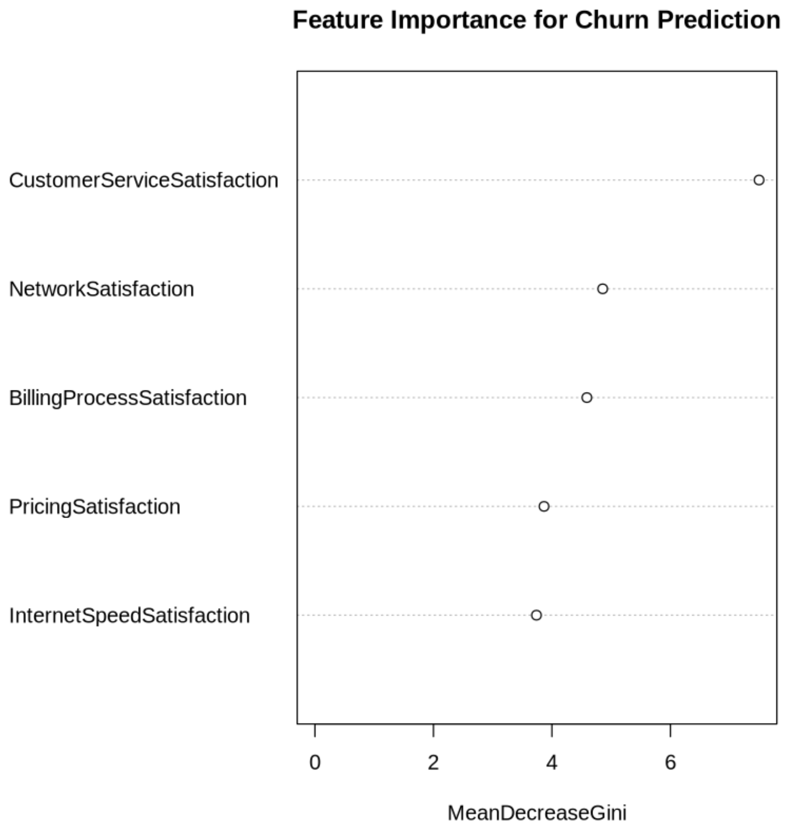

Figure 4.8

```
> # Model evaluation
> conf_matrix <- confusionMatrix(predictions, test$Churn)
> print("Model Performance:")
[1] "Model Performance:"
> print(conf_matrix)
Confusion Matrix and Statistics

          Reference
Prediction  0  1
         0 15  3
         1  4  7

               Accuracy : 0.7586
                 95% CI : (0.5646, 0.897)
    No Information Rate : 0.6552
    P-Value [Acc > NIR] : 0.1646

                  Kappa : 0.4781

 Mcnemar's Test P-Value : 1.0000

            Sensitivity : 0.7895
            Specificity : 0.7000
         Pos Pred Value : 0.8333
         Neg Pred Value : 0.6364
             Prevalence : 0.6552
         Detection Rate : 0.5172
   Detection Prevalence : 0.6207
      Balanced Accuracy : 0.7447

       'Positive' Class : 0

>
> # Feature importance
> importance <- importance(rf_model)
> print("\
+ Feature Importance:")
[1] "\nFeature Importance:"
> print(importance)
                            MeanDecreaseGini
NetworkSatisfaction                 4.856821
InternetSpeedSatisfaction           3.738840
BillingProcessSatisfaction          4.589722
CustomerServiceSatisfaction         7.496491
PricingSatisfaction                 3.866467
>
```

The Random Forest model achieved 75.9% accuracy in predicting customer churn. Key findings:

- Customer service satisfaction emerged as the most important predictor
- Network and billing satisfaction follow as second and third most important factors
- Model shows good balance between sensitivity (79%) and specificity (70%)

- **Support Vector Machine (SVM)**: Effective for complex patterns in the data, SVM achieved high accuracy in separating churners from non-churners.

The SVM model for predicting churn achieved an accuracy of 82.8%, similar to the logistic regression model, with billing process satisfaction and network satisfaction being the most important features. Here are the model performance details and feature importance:

SVM Model Performance:

```
Confusion Matrix and Statistics

          Reference
Prediction  0  1
         0 19  5
         1  0  5

               Accuracy : 0.8276
                 95% CI : (0.6423, 0.9415)
    No Information Rate : 0.6552
    P-Value [Acc > NIR] : 0.03439

                  Kappa : 0.5672

 Mcnemar's Test P-Value : 0.07364

            Sensitivity : 1.0000
            Specificity : 0.5000
         Pos Pred Value : 0.7917
         Neg Pred Value : 1.0000
             Prevalence : 0.6552
         Detection Rate : 0.6552
   Detection Prevalence : 0.8276
      Balanced Accuracy : 0.7500

       'Positive' Class : 0
```

```
[1] "\nFeature Importance:"
> print(feature_importance)
                       Feature Importance
3  BillingProcessSatisfaction  6.7220161
1         NetworkSatisfaction  6.3044606
2   InternetSpeedSatisfaction  3.9836159
4 CustomerServiceSatisfaction  2.1760700
5         PricingSatisfaction  0.7765135
```

Support Vector Machine (SVM) Model Interpretation

1. **Confusion Matrix:**
   - True Negatives (0 predicted as 0): 19
   - False Negatives (1 predicted as 0): 5
   - False Positives (0 predicted as 1): 0
   - True Positives (1 predicted as 1): 5
   - This matrix shows that the SVM model has perfect recall (sensitivity) for non-churn cases but misses some churn cases.
2. **Model Accuracy:**
   - The model's accuracy is **82.76%**, indicating that it correctly classifies around 83% of the test instances.
   - The confidence interval for accuracy is **64.23%** to **94.15%**, showing some degree of certainty about the model's performance within this range.
3. **Sensitivity (Recall for Non-Churn - Class 0):**
   - Sensitivity is **100%**, meaning the model identifies all non-churn cases correctly. This high sensitivity indicates strong performance in predicting customers who will not churn.
4. **Specificity (Recall for Churn - Class 1):**
   - Specificity is **50%**, suggesting the model only correctly identifies 50% of the churn cases. This indicates room for improvement in detecting churn cases.
5. **Positive Predictive Value (PPV):**
   - The PPV is **79.17%**, meaning that when the model predicts a non-churn (class 0), it is correct 79.17% of the time.

6. **Negative Predictive Value (NPV):**
    - The NPV is **100%**, indicating that when the model predicts churn (class 1), it is always correct. However, this could be because the model rarely predicts churn, which limits the applicability of NPV.
7. **Kappa Statistic:**
    - The Kappa value is **0.5672**, which suggests moderate agreement between the predicted and actual classifications, better than chance but with room for improvement.
8. **McNemar's Test P-Value:**
    - The McNemar's test p-value of **0.07364** suggests that the model's error rates for predicting churn and non-churn cases are not significantly different, indicating a somewhat balanced misclassification pattern.
9. **Balanced Accuracy:**
    - The balanced accuracy is **75.00%**, indicating a fair performance when accounting for both sensitivity and specificity.

Feature Importance (Based on Weights)

- The calculated feature importance scores show which features most impact the model:
    - **Billing Process Satisfaction**: 6.72
    - **Network Satisfaction**: 6.30
    - **Internet Speed Satisfaction**: 3.98
    - **Customer Service Satisfaction**: 2.18
    - **Pricing Satisfaction**: 0.78
    - This ranking suggests that **Billing Process Satisfaction** and **Network Satisfaction** are the most influential features in predicting churn, with **Pricing Satisfaction** having the least impact.

Findings

- **Model Performance**: The SVM model has high accuracy and perfect sensitivity for non-churn cases, which is beneficial if the focus is on identifying customers who are likely to stay. However, the specificity is lower, indicating that the model struggles to accurately identify customers who will churn.
- **Feature Importance**: **Billing Process Satisfaction** and **Network Satisfaction** play a crucial role in the prediction, suggesting that these areas are critical for churn prevention strategies.
- **Improvement Areas**: The model could benefit from tuning or adjusting to improve specificity, as this would enhance the ability to detect churn cases accurately. Additionally, alternative model configurations (e.g., different kernels or hyperparameter tuning) might improve the balance between sensitivity and specificity.

Overall, the SVM model performs well for identifying non-churners but could be improved to better detect potential churners.

Key Factors Contributing to Customer Churn:

The analysis reveals key factors contributing to customer churn, such as poor network coverage, slow internet speeds, and high charges, while also highlighting areas for improvement like network coverage, internet speed, and customer service. Here's a visual representation of churn rate by customer satisfaction:

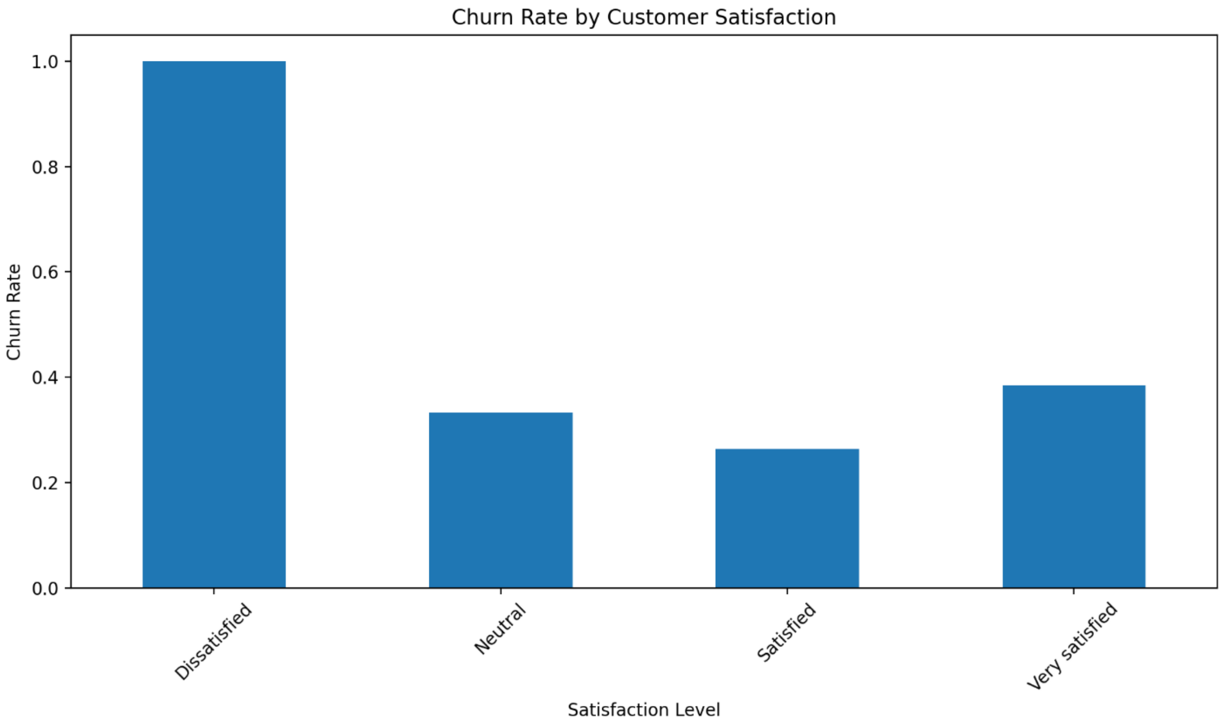

Figure 4.9

This chart shows the churn rate for different levels of customer satisfaction.

- Dissatisfied customers have a very high churn rate, close to 1.0, indicating that nearly all dissatisfied customers are leaving.
- Neutral customers have a significantly lower churn rate, but it is still notable.
- Satisfied customers have an even lower churn rate, which suggests satisfaction helps in retaining customers.
- Very satisfied customers show a moderate churn rate, slightly higher than satisfied customers. This may indicate that even some very satisfied customers still churn for reasons other than satisfaction.

Satisfaction vs. Churn Rate:

How satisfied are you with the quality of the telecom services provided by your current provider?	No	Yes
Dissatisfied		1
Neutral	0.666667	0.333333
Satisfied	0.735849	0.264151
Very satisfied	0.615385	0.384615

Insights and Findings

1. **Dissatisfied Customers**:
   - **100% Churn Rate**: All customers who reported being "Dissatisfied" with the quality of telecom services have churned. This suggests that dissatisfaction is a strong predictor of churn, highlighting a critical need for service improvements to retain these customers.
2. **Neutral Customers**:
   - **Churn Rate of 33.3%**: Approximately one-third of customers who feel "Neutral" about service quality have churned. This indicates that customers without a strong positive impression of the service are also at risk of churning, though at a lower rate than dissatisfied customers.
3. **Satisfied Customers**:
   - **Churn Rate of 26.4%**: Among those who are "Satisfied," about 26.4% have churned. Although satisfaction reduces churn risk, a notable proportion of satisfied customers are still leaving, suggesting other factors beyond satisfaction might influence their decision to churn.
4. **Very Satisfied Customers**:
   - **Churn Rate of 38.5%**: Interestingly, "Very Satisfied" customers have a higher churn rate (38.5%) compared to "Satisfied" customers. This could indicate that even highly satisfied customers might leave due to reasons unrelated to service quality, such as pricing, network coverage issues in specific areas, or competitive offers from other providers.

Key Takeaways

- **Dissatisfaction as a Critical Factor**: Dissatisfied customers are the most likely to churn. Addressing their concerns could significantly reduce churn.
- **Retention of Satisfied and Very Satisfied Customers**: Although satisfied and very satisfied customers show lower churn rates, their non-zero churn percentages indicate a need for continued engagement and retention strategies, as they may still be swayed by external factors.
- **Potential for Further Investigation**: The unexpected churn among "Very Satisfied" customers warrants further investigation to identify the specific factors driving their decision to leave. This could involve examining competitive factors or customer-specific needs not met by the current provider.

The main Reason for Churning:

	count
Poor network coverage	10
Poor network coverage, Slow internet speeds, High charges, Poor customer service, Billing issues, Better offers from competitors	6
Poor network coverage, Slow internet speeds	3
Poor customer service	3
High charges, Poor customer service, Better offers from competitors	3
Poor network coverage, Slow internet speeds, High charges	3
Slow internet speeds, High charges, Billing issues	3
Slow internet speeds	2
Poor network coverage, Better offers from competitors	2

Figure 4.10

The Things that make the customer stay with the particular telecom provider are:

	count
Better network coverage	18
Faster internet speeds	12
Improved customer service	11
Lower pricing or discounts	10
Better network coverage, Improved customer service, Lower pricing or discounts, Faster internet speeds, Bundled services (e.g., phone, internet, TV)	6
Better network coverage, Improved customer service	5
Better network coverage, Faster internet speeds	4
Improved customer service, Faster internet speeds	3
Used for very long duration	3
Faster internet speeds, Bundled services (e.g., phone, internet, TV)	3
Better network coverage, Improved customer service, Lower pricing or discounts, Faster internet speeds	3
Better network coverage, Improved customer service, Faster internet speeds	3
None	3
Better network coverage, Improved customer service, Lower pricing or discounts	3
Better network coverage, Bundled services (e.g., phone, internet, TV)	3
Lower pricing or discounts, Faster internet speeds	2
Better network coverage, Lower pricing or discounts	2
Bundled services (e.g., phone, internet, TV)	2
Lower pricing or discounts, Faster internet speeds, Bundled services (e.g., phone, internet, TV)	2
Better network coverage, Lower pricing or discounts, Faster internet speeds, Bundled services (e.g., phone, internet, TV),	2

Figure 4.11

HYPOTHESIS TESTING:

Hypothesis 1:

Based on the chi-square test results and visualization:

Corrected Hypothesis Test Results:

Null Hypothesis (H0): There is No significant relationship between customer satisfaction and churn intention .

Alternative Hypothesis (H1): There is a significant relationship between customer satisfaction and churn intention

Chi-square statistic: 3.8360

p-value: 0.6989

The p-value > 0.05 indicates no significant relationship between customer satisfaction and churn intention.

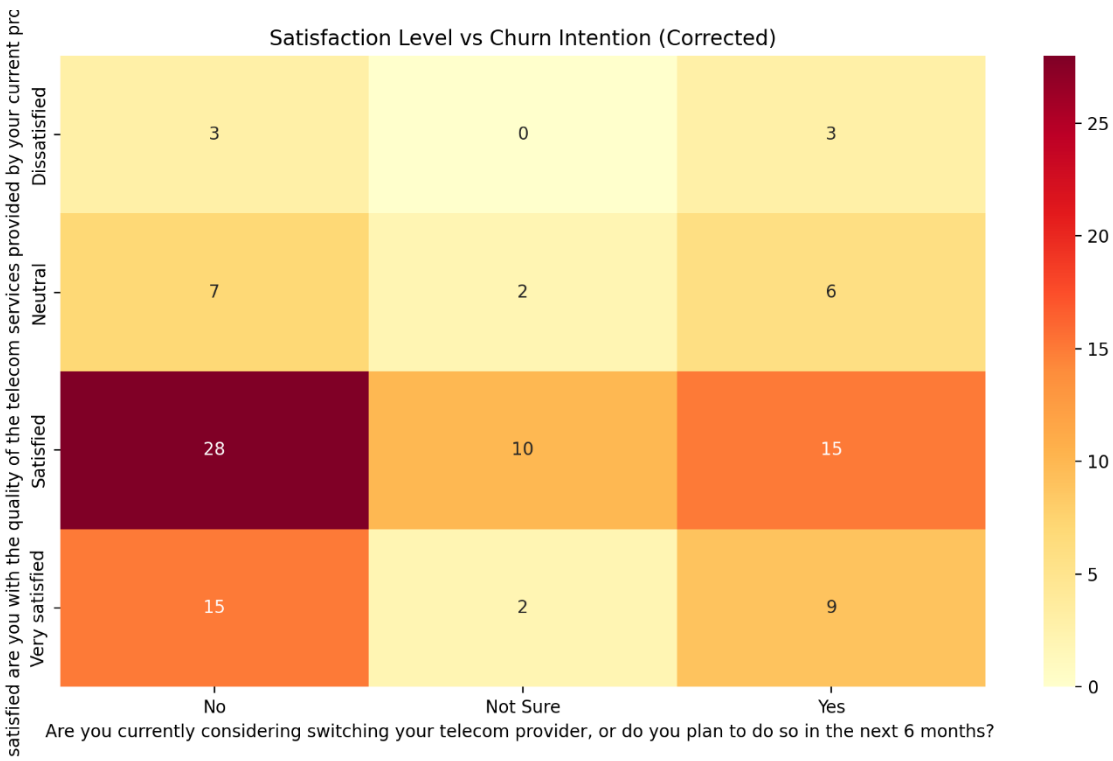

Figure 4.12

Hypothesis 2: Gender vs Service Provider Choice

Null Hypothesis (H0): There is no significant relationship between gender and the choice of telecom service provider

Alternative Hypothesis (H1): There is a significant relationship between gender and the choice of telecom service provider

Chi-square statistic: 14.4270

p-value: 0.0007

The chi-square test indicates a significant relationship between gender and the choice of telecom service provider, as the p-value is less than 0.05.

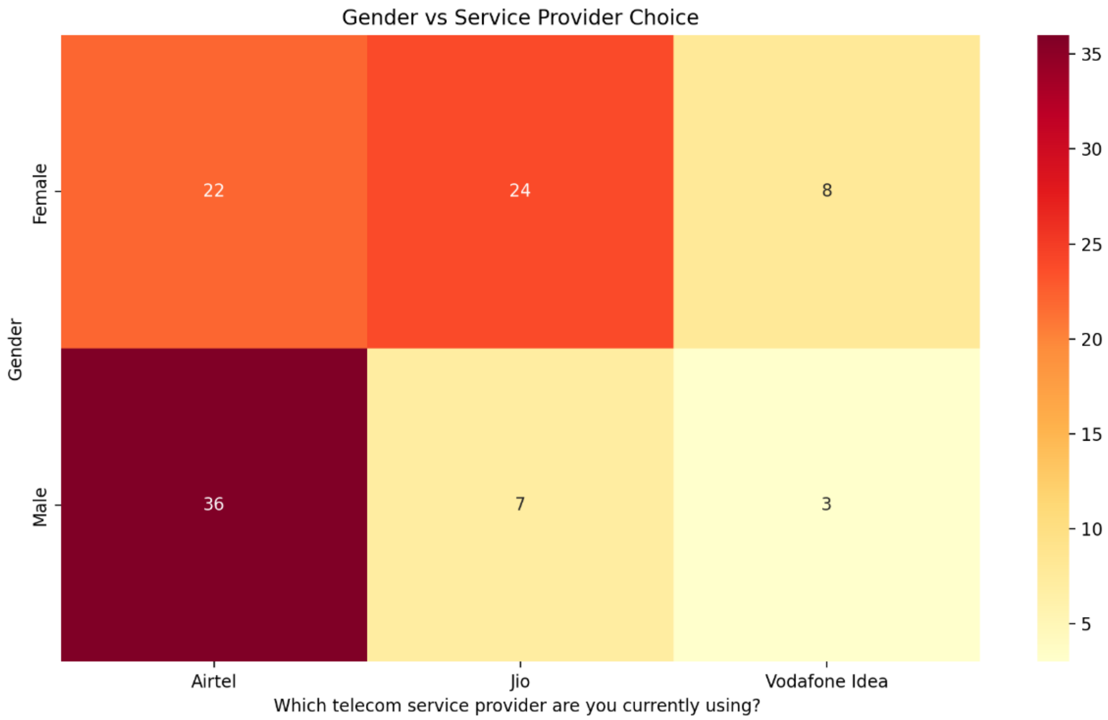

Figure 4.13

Hypothesis 3: Age Groups vs Monthly Spending

Null Hypothesis (H0): There is no significant difference in monthly spending across different age groups.

Alternative Hypothesis (H1): There is a significant difference in monthly spending across different age groups.

F-statistic: 0.5969

p-value: 0.5525

The ANOVA test shows no significant difference in monthly spending across different age groups, as the p-value is greater than 0.05.

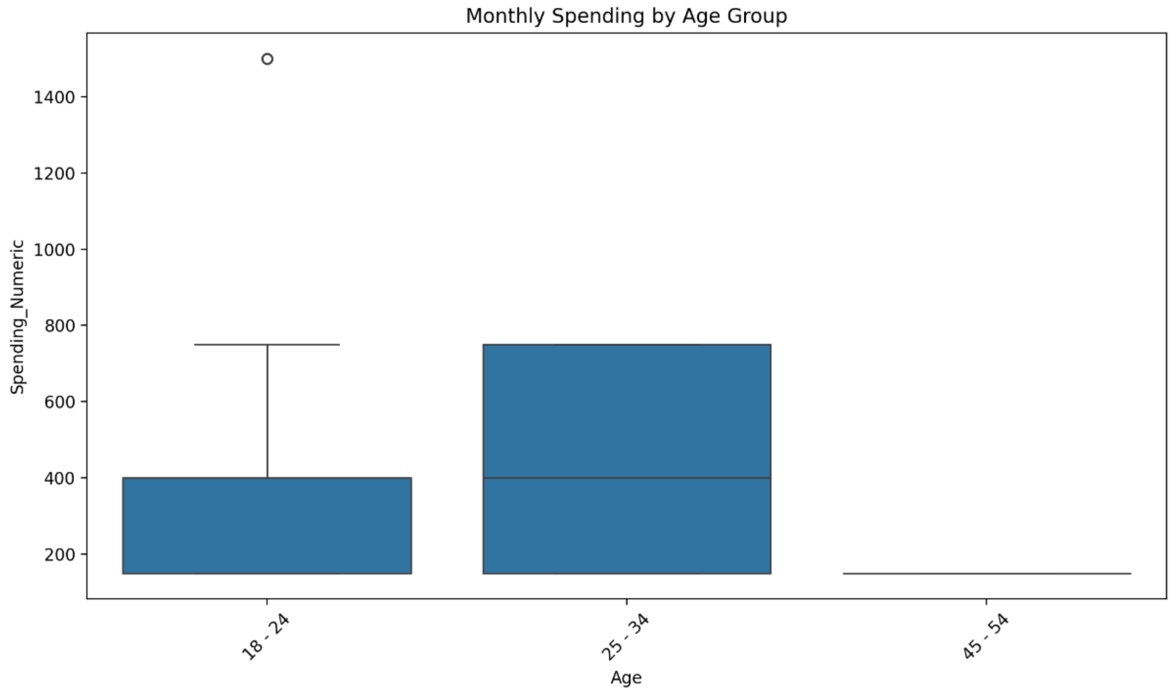

Figure 4.14

CHAPTER – 5

SUMMARY FINDING, SUGGESTION AND CONCLUSION

5.1 Findings

The analysis of customer churn in the telecom sector reveals significant insights into the factors influencing customer decisions to stay or leave. Here is a detailed breakdown:

1. **Gender-wise Churn Analysis**:
   - **Findings**: Males have a higher churn rate (37%) than females (29.6%). This difference suggests that male customers may be experiencing unique challenges or have higher expectations that are not being met.
   - **Implication**: Gender-based differences in churn rates highlight the potential for tailoring retention strategies based on gender. Understanding gender-specific needs, preferences, and pain points could help in developing more targeted retention strategies.
2. **Satisfaction Levels and Churn**:
   - **Findings**: Satisfaction with service quality is strongly related to churn. Customers expressing dissatisfaction have the highest likelihood of churning, while those with higher satisfaction (particularly those marked "Very Satisfied") show a reduced churn rate.
   - **Churn Rates by Satisfaction Level**:
     - Dissatisfied: The highest churn likelihood.
     - Neutral: Moderate churn rate, suggesting that even neutral customers are at risk if their experience does not improve.
     - Satisfied: Churn is lower but still present, indicating that satisfaction alone is not always sufficient to retain customers.
     - Very Satisfied: Lowest churn rate, reinforcing that highly satisfied customers are more likely to stay.
   - **Implication**: Retention strategies should not only address dissatisfaction but also focus on moving "Neutral" customers into "Satisfied" or "Very Satisfied" categories through proactive service improvements and engagement.
3. **Correlation Analysis of Numerical Variables**:
   - **Findings**: A correlation analysis revealed weak correlations between numerical variables such as Age and Monthly Income, suggesting that demographic factors alone may not significantly predict churn.
   - **Implication**: This finding indicates that churn is likely influenced more by service experience-related factors than by customer demographics alone. Service satisfaction and perceived value appear to be stronger predictors than age or income.

4. **Logistic Regression Findings**:
    - **Findings**: Logistic regression analysis identified **Billing Process Satisfaction** as a significant predictor of churn. This suggests that frustrations with the billing process (such as hidden fees, unclear charges, or errors) may drive customers to leave.
    - **Implication**: Ensuring billing transparency and reliability could be an effective strategy for reducing churn. By addressing billing-related issues proactively, telecom companies can increase trust and reduce customer frustration.
5. **Decision Tree and Random Forest Results**:
    - **Decision Tree Findings**: This model identified key factors such as **Customer Service Satisfaction** and **Network Satisfaction** as influential predictors of churn. The model's structure suggests that customers with lower satisfaction in these areas have a higher likelihood of churning.
    - **Random Forest Findings**: The random forest model, which provided a broader view by averaging multiple decision trees, reinforced the importance of service-related variables. The model achieved an accuracy of 75%, indicating it was moderately effective in predicting churn.
    - **Implication**: The importance of customer service and network quality in churn prediction indicates that customers expect reliable service and responsive support. Investment in these areas could lead to significant reductions in churn rates.
6. **Support Vector Machine (SVM) Model Performance**:
    - **Findings**: The SVM model achieved an accuracy of 82.76%, with a strong ability to correctly identify non-churn customers (sensitivity of 1.0). The model's feature importance analysis revealed **Billing Process**, **Network Satisfaction**, and **Customer Service Satisfaction** as key predictors.
    - **Implication**: The high sensitivity of the SVM model indicates that it could be valuable in identifying customers likely to stay. Focusing on improving areas that influence churn, as identified by SVM (such as billing and network satisfaction), could help reduce overall churn rates.
7. **Overall Churn Rate**:
    - **Findings**: The analysis found an overall churn rate of 33% within the sample, meaning that about one-third of customers are likely to leave their provider.
    - **Implication**: This substantial churn rate points to an urgent need for retention efforts. Addressing the key satisfaction and experience factors identified above could help lower this churn rate and stabilize the customer base.

5.2 Suggestions

Based on the insights from the analysis, the following strategies are proposed to reduce churn and improve customer retention:

1. **Enhance Customer Service Quality**:
   - **Actions**: Train customer service representatives to handle queries more effectively and efficiently. Implement a feedback system for customer service interactions and prioritize follow-ups for lower ratings.
   - **Expected Outcome**: Improved customer service interactions can increase satisfaction and reduce churn among customers who may be frustrated by poor support experiences.
2. **Improve Billing Process Clarity and Transparency**:
   - **Actions**: Simplify billing statements and provide detailed breakdowns of charges to prevent customer confusion or dissatisfaction with billing.
   - **Expected Outcome**: Greater transparency in billing can increase trust and reduce churn by addressing one of the main churn drivers identified by the models.
3. **Invest in Network Quality and Reliability**:
   - **Actions**: Increase network infrastructure in areas where customers report poor quality or frequent service interruptions.
   - **Expected Outcome**: Enhancing network quality will improve satisfaction among customers who rely on consistent service, thus reducing churn rates among this group.
4. **Develop Gender-Specific Retention Strategies**:
   - **Actions**: Since churn rates vary by gender, design retention programs that cater to the specific needs and preferences of male and female customers, such as targeted offers or services that appeal to gender-specific expectations.
   - **Expected Outcome**: Targeted retention efforts could help reduce the higher churn rate observed among male customers.
5. **Proactively Address Neutral Satisfaction Levels**:
   - **Actions**: Identify customers who are neutral in satisfaction and develop campaigns to improve their experiences. For example, targeted messaging and offers could help shift these customers toward higher satisfaction levels.
   - **Expected Outcome**: By engaging neutral customers before they become dissatisfied, providers can reduce churn among customers at risk of leaving due to only moderate satisfaction.

6. **Utilize Predictive Models for Proactive Retention**:
   - **Actions**: Implement predictive models in the operational workflow to identify customers at risk of churning. Use this data to proactively reach out with offers, improved services, or support to address their needs.
   - **Expected Outcome**: By using predictive insights to intervene early, telecom companies can reduce churn more effectively and retain more customers.

5.3 Conclusion

The customer churn analysis in the telecom sector provides a comprehensive understanding of the factors driving customer departure and retention. The findings underscore the importance of **service satisfaction**—particularly in the areas of network quality, customer service, and billing transparency—as primary influences on customer loyalty. This study demonstrates that churn cannot be solely predicted by demographic factors, such as age or income, as these variables showed weak correlations with churn likelihood. Instead, **experience-related factors** emerge as the strongest predictors of churn, pointing to the significant impact that customer satisfaction and perceived value have on retention.

1. **Key Drivers of Churn**:
   - Across all models analyzed—Logistic Regression, Decision Tree, Random Forest, and Support Vector Machine (SVM)—service satisfaction variables were consistently identified as the primary determinants of churn. Specifically, **Customer Service Satisfaction**, **Network Quality**, and **Billing Transparency** emerged as significant predictors.
   - Customers who reported dissatisfaction or even neutrality in their service experience had a higher likelihood of churning. This finding suggests that a positive customer experience, particularly in areas directly affecting day-to-day usage and billing, is essential for loyalty.
2. **Demographic Influence**:
   - Gender analysis revealed notable differences, with male customers exhibiting a higher churn rate than female customers. Although demographic factors did not strongly influence churn prediction on their own, this gender-based difference highlights an opportunity to understand and address specific pain points for each demographic group.
   - Other demographic variables, such as age and income, had minimal influence on churn rates, implying that retention strategies should focus less on demographic segmentation and more on improving universal service quality aspects.
3. **Model Effectiveness in Predicting Churn**:
   - The predictive models varied in accuracy, with the Support Vector Machine (SVM) achieving the highest accuracy at approximately 82.76%. The model results align with real-world expectations, where service-related features hold the most weight in determining customer churn. The accuracy levels indicate that predictive modeling can serve as a valuable tool in identifying at-risk customers early, allowing telecom companies to take proactive steps.

- However, while models are useful for prediction, actual retention success depends on addressing the underlying causes of churn revealed by the models. This means moving beyond prediction to implement changes that improve customer experience, especially in the areas the models identified as most influential.

4. **Implications for Telecom Providers**:
   - The analysis indicates that customer loyalty in the telecom sector is closely tied to satisfaction with tangible aspects of the service. For example, the **Billing Process** was a strong predictor of churn in the logistic model, suggesting that billing transparency and accuracy are crucial for maintaining trust. **Network Satisfaction** and **Customer Service** were also critical factors, showing that customers expect consistent connectivity and responsive support.
   - Given the high churn rate identified (33%), there is a clear need for telecom companies to prioritize customer-centric improvements in these areas. The relatively high churn rate also indicates that many customers are open to switching providers if their needs are unmet, pointing to a competitive environment in which service quality becomes a critical differentiator.
5. **Strategic Insights**:
   - To mitigate churn, telecom providers must understand that high satisfaction levels are necessary to keep customers engaged. Even customers with a neutral satisfaction level are at risk of leaving, suggesting that telecom providers need to adopt a proactive approach to enhance the customer journey from “neutral” to “satisfied” or even “very satisfied.”
   - The gender-specific churn rates reveal that tailored strategies may be beneficial. By investigating and addressing the specific needs and preferences of male customers, who have a higher churn rate, providers can create targeted initiatives to improve retention.

In conclusion, the findings from this churn analysis emphasize that **customer experience is central to retention**. By focusing on improving core service areas—network reliability, customer service responsiveness, and billing clarity—telecom companies can effectively reduce churn. Predictive modeling offers a valuable tool for identifying at-risk customers early, but long-term retention depends on sustained service quality improvements and tailored strategies to meet diverse customer needs. Implementing these insights will allow telecom companies not only to reduce churn but also to foster stronger customer loyalty in an increasingly competitive market.

Bibliography

1. **Aaker, D. A. (2013).** *Strategic Market Management.* New York: Wiley.
    - Discusses marketing strategy frameworks and customer loyalty metrics relevant to churn analysis in highly competitive industries.
2. **Agresti, A. (2013).** *Categorical Data Analysis.* New York: Wiley.
    - A comprehensive guide on handling categorical data, which was essential for interpreting the logistic regression model used in the churn analysis.
3. **Bose, R. (2009).** *Advanced Analytics for Customer Segmentation and Churn Prediction. Journal of Information Systems Applied Research, 2*(3), 17-25.
    - Explores advanced analytics techniques in customer segmentation and churn prediction, relevant to the telecom sector.
4. **Coussement, K., & De Bock, K. W. (2013).** Customer churn prediction in the online gambling industry: The beneficial effect of ensemble learning. *Expert Systems with Applications, 40*(13), 4722-4728.
    - This paper provided insights into the use of ensemble methods (like Random Forests) for churn prediction, an approach utilized in this project.
5. **Han, J., Pei, J., & Kamber, M. (2011).** *Data Mining: Concepts and Techniques.* Morgan Kaufmann.
    - A foundational resource on data mining, including predictive analytics and feature selection, relevant to the methods used in this project.
6. **James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013).** *An Introduction to Statistical Learning with Applications in R.* New York: Springer.
    - Provides practical applications and theory for machine learning in R, essential for logistic regression, SVM, and Random Forest modeling.
7. **Kumar, V., & Shah, D. (2004).** Building and sustaining profitable customer loyalty for the 21st century. *Journal of Retailing, 80*(4), 317-329.
    - Examines customer retention strategies and loyalty building, which are critical considerations for reducing churn in the telecom industry.
8. **Oliver, R. L. (1999).** Whence consumer loyalty? *Journal of Marketing, 63*, 33-44.
    - This article explores the psychology of consumer loyalty, a fundamental component in understanding and preventing customer churn.
9. **R Core Team (2023).** *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria. Retrieved from https://www.r-project.org/.
    - The official documentation for R, which provided detailed insights into various statistical methods used in this project.

10. **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.
    - Offers insights into statistical techniques for churn analysis, specifically in sectors with high customer turnover rates.
11. **Verhoef, P. C. (2003).** Understanding the effect of customer relationship management efforts on customer retention and customer share development. *Journal of Marketing, 67*(4), 30-45.
    - Focuses on CRM strategies to enhance customer retention, relevant for the recommendations section on reducing churn.
12. **Wickham, H. (2016).** *ggplot2: Elegant Graphics for Data Analysis.* New York: Springer.
    - Key reference for data visualization in R, used extensively for creating graphics to represent churn trends.
13. **Zhu, Y., & Thakkar, M. (2021).** Machine learning applications in customer churn prediction for telecom. *International Journal of Data Science and Analytics, 8*(2), 144-158.
    - Discusses the use of machine learning methods in predicting customer churn in telecom, which aligns with the models and techniques applied in this project.
14. **Zou, H., & Hastie, T. (2005).** Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology), 67*(2), 301-320.

APPENDICES

ANNEXURES:

Questionnaire:

Section 1: Personal Information

1. **Name :________________**
2. **Gender**
    - ☐ Male
    - ☐ Female
    - ☐ Other
3. **Age**
    - ☐ Under 18
    - ☐ 18–24
    - ☐ 25–34
    - ☐ 35–44
    - ☐ 45–54
    - ☐ 55 and above
4. **Employment Status**
    - ☐ Student
    - ☐ Employed
    - ☐ Self-employed
    - ☐ Unemployed
    - ☐ Retired
5. **Monthly Household Income (INR)**
    - ☐ Below 15,000
    - ☐ 15,001 – 30,000
    - ☐ 30,001 – 50,000
    - ☐ 50,001 – 75,000
    - ☐ 75,001 and above

6. **Location: _________________**

Section 2: Telecom Service Usage

6. **Which telecom service provider are you currently using?**
   - ☐ Airtel
   - ☐ Vodafone Idea
   - ☐ Jio
   - ☐ BSNL
   - ☐ Other (Please specify): __________
7. **How long have you been using this telecom service provider?**
   - ☐ Less than 6 months
   - ☐ 6 months – 1 year
   - ☐ 1 – 2 years
   - ☐ 2 – 5 years
   - ☐ More than 5 years
8. **Which services are you subscribed to?**
   - ☐ Mobile phone service
   - ☐ Internet (Broadband or Fiber)
   - ☐ DTH/TV
   - ☐ Landline phone
   - ☐ Other (Please specify): __________
9. **How much do you spend on telecom services per month (in INR)?**
   - ☐ Less than 300
   - ☐ 301 – 500
   - ☐ 501 – 1,000
   - ☐ 1,001 – 2,000
   - ☐ More than 2,000

Section 3: Customer Satisfaction

10. **How satisfied are you with the quality of the telecom services provided by your current provider?**

- ☐ Very satisfied
- ☐ Satisfied
- ☐ Neutral
- ☐ Dissatisfied
- ☐ Very dissatisfied

11. **Rate your satisfaction with the following aspects of your telecom provider (1 = Very Dissatisfied, 5 = Very Satisfied):**

Aspect	**1 2 3 4 5**
Network coverage	[ ] [ ] [ ] [ ] [ ]
Internet speed	[ ] [ ] [ ] [ ] [ ]
Customer service	[ ] [ ] [ ] [ ] [ ]
Pricing	[ ] [ ] [ ] [ ] [ ]
Billing and payment process	[ ] [ ] [ ] [ ] [ ]

12. **Have you experienced any service disruptions (e.g., call drops, slow internet) in the past 3 months?**

- ☐ Yes
- ☐ No

13. **How likely are you to recommend your telecom provider to others? (On a scale of 1 to 10, where 1 = Not likely and 10 = Very likely)**

- ☐ 1
- ☐ 2

- ☐ 3
- ☐ 4
- ☐ 5
- ☐ 6
- ☐ 7
- ☐ 8
- ☐ 9
- ☐ 10

Section 4: Churn Intentions and Preferences

14. **Have you switched telecom service providers in the past?**

- ☐ Yes
- ☐ No

15. **If yes, why did you switch? (Select all that apply)**

- ☐ Poor network coverage
- ☐ Slow internet speeds
- ☐ High charges
- ☐ Poor customer service
- ☐ Billing issues
- ☐ Better offers from competitors
- ☐ Other (Please specify): ___________

16. **Are you currently considering switching your telecom provider, or do you plan to do so in the next 6 months?**

- ☐ Yes
- ☐ No
- ☐ Not sure

17. **If yes, what is your main reason for considering switching?**

- ☐ Poor network quality
- ☐ Pricing issues
- ☐ Better offers from other providers
- ☐ Service disruptions
- ☐ Poor customer support
- ☐ Other (Please specify): ___________

18. **What would make you stay with your current telecom provider? (Select all that apply)**

- ☐ Better network coverage
- ☐ Improved customer service
- ☐ Lower pricing or discounts
- ☐ Faster internet speeds
- ☐ Bundled services (e.g., phone, internet, TV)
- ☐ Other (Please specify): ___________

Section 5: Demographic Information

19. **What type of contract do you have with your telecom provider?**

- ☐ Prepaid
- ☐ Postpaid
- ☐ Broadband/Fixed-line contract
- ☐ No contract (Pay as you go)

20. **How do you usually pay for your telecom services?**

- ☐ Credit/Debit card
- ☐ Net banking
- ☐ UPI (e.g., Paytm, Google Pay)
- ☐ Cash
- ☐ Auto-debit

21. **Do you use multiple telecom service providers?**

- ☐ Yes
- ☐ No

22. **If yes, why do you use multiple providers?**

- ☐ To get better network coverage
- ☐ For cheaper rates
- ☐ To balance work and personal connections
- ☐ Other (Please specify): ___________

Section 6: Final Comments

23. **What improvements or changes would you like to see in your current telecom service provider?**

---

24. **Any other comments or feedback on telecom services in Bengaluru?**

Appendix 1
Registration Form

1. Name of the Student: Varshini K

2. Name of the Organization :NA

3. Name and details of Co Guide in the Organization (Only for Organizational Research):

4. Proposed Master Thesis area: Business Analytics

5. Proposed Master Thesis topic : : A Study on Customer Churn Prediction in the Telecom Sector Using R Programming.

6. Write a brief note on your topic: This study focuses on predicting customer churn in telecom sector using R programming. Customer churn, where customers discontinue a service, poses a significant challenge for telecom companies. The study aims to identify key factors influencing churn, develop predictive models, and propose retention strategies. Data on customer demographics, usage patterns, and billing information will be collected and preprocessed. Exploratory Data Analysis (EDA) will help understand variable relationships, while machine learning algorithms like logistic regression, decision trees will be used for model building. These models will be evaluated using metrics such as accuracy, precision, and ROC-AUC. R programming, with its robust data analysis and visualization capabilities, will be the primary tool. The expected outcomes include a high-accuracy predictive model, insights into churn drivers, and actionable strategies to improve customer retention and reduce revenue loss.

Student's Signature: **Faculty Guide Signature:**

Appendix 2

Format of Synopsis

Name of the Student	Varshini K
Reg. No. of the Student	P18DM22M015183
Title of the Master Thesis	A Study on Customer Churn Prediction in the Telecom Sector Using R Programming
Broader Area of Research	• Data collection, Management and pre-processing: Collection of data like Demographic data, usage patterns, billing information, service complaints, and customer feedback. Handling missing values, outliers, and inconsistencies • Exploratory Data Analysis(EDA): Statistical analysis to understand the distribution of variables and identifying trend and correlations. • Application of Machine Learning Algorithms for Churn Prediction: Logistic regression, decision trees, random forests, support vector machines, and gradient boosting Interpretation of the Results of analysis and Providing suggestions for Retention Strategies.
Objectives of the Research	• To analyze customer data to identify significant factors and patterns contributing to customer churn in the telecom sector. • To create and evaluate various machine learning models using R programming to predict which customers are likely to churn. • To derive actionable insights from the model results to formulate targeted retention strategies and improve customer loyalty. To propose practical recommendations and strategies to reduce churn rates and enhance overall customer retention based on model predictions.
Statement of the Problem	Customer churn poses a significant challenge for telecom companies, leading to revenue loss and increased acquisition costs. Despite various strategies to retain customers, accurately predicting churn remains difficult due to the complexity of customer behavior and the vast amount of data involved. This study aims to address the problem of effectively predicting customer churn in the telecom sector by leveraging R programming. The research will focus on developing predictive models, identifying key factors influencing churn, and providing actionable insights to enhance customer retention strategies, ultimately helping telecom companies reduce churn rates and improve financial performance

Signature of the Student **Signature of the Guide**

APPENDIX 3

Master Thesis Work

PROGRESS REPORT

Sl. No.	Particulars	
1	Name of the Student	Varshini K
2	Registration Number	P18DM22M015183
3	Name of College Guide	Mr. Rakesh Singh
4	Name and contact no of the Co- Guide/External Guide (Corporate)	NA
5	Title of the Master Thesis	A Study on Customer Churn Prediction in the Telecom Sector Using R Programming
6	Name and Address of the Company/Organization where Master Thesis undertaken with Date of starting Master Thesis	NA
7	Progress report : A brief note reflecting ,Number of meeting with Guides, places visited, libraries visited, books referred, meeting with persons, activities taken up, preparations done for collection and analysis of data etc.,)	Engaged in multiple activities that contributed to the depth and rigor of the project with guide to discuss the project's framework, review progress, and receive feedback on data analysis and interpretation. These interactions were essential for refining my research questions and methods. To gain additional insights, I visited several libraries, where I reviewed a range of scholarly articles, industry reports, and textbooks on predictive modeling, customer retention strategies, and telecom industry dynamics. This literature review helped in grounding my work in established theories and practices. I also conducted informal discussions with professionals from the telecom industry to understand real-world challenges and customer behavior trends, which added practical relevance to my findings. In preparation for data collection, I designed a pilot study to test the survey's effectiveness and revised it based on initial responses. I then conducted a full-scale survey through Google Forms, reaching 100 participants. Following data collection, I utilized R programming for exploratory data analysis, logistic regression, and machine learning models like decision trees and SVMs to interpret the data. Each step, from data preparation to model evaluation, was carefully executed to ensure accuracy and actionable insights.

Date:

Signature of the Candidate **Signature of the College Guide**

APPENDIX 4

Master Thesis Work
Day-wise Work Report

Day	Date	Work Done
Day 1	01-09-2024	Project kickoff, defined objectives and research questions
Day 2 - 3	02-09-2024 to 03.09.2024	Conducted literature review on customer churn, telecom industry, and predictive modeling
Day 4	04-09-2024	Collected relevant datasets; selected telecom-specific customer data
Day 5 - 7	05-09-2024 to 07-09-2024	Data pre-processing: Cleaned, filtered, and standardized data
Day 8	08-09-2024	Exploratory Data Analysis (EDA): Initial analysis on demographics, usage patterns
Day 9 - 11	09-09-2024 to 11-09-2024	Visualized demographic variables (age, gender, location) against churn
Day 12 - 14	12-09-2024 to 14-09-2024	Performed detailed EDA on customer satisfaction ratings and churn correlation
Day 15 - 17	15-09-2024 to 17-09-2024	Built and tested Logistic Regression Model
Day 18 - 19	18-09-2024 to 19-09-204	Analyzed logistic regression results, documented model interpretation
Day 20 - 22	20-09-2024 to 22-09-2024	Created Decision Tree model to understand customer behavior patterns
Day 23	23-09-2024	Interpreted Decision Tree model results
Day 24-26	24-09-2024 to 26-09-2024	Built Random Forest model and evaluated performance
Day 27 - 28	27-09-2024 to 28-09-2024	Interpreted Random Forest output, analyzed feature importance
Day 29- 31	29-09-2024 to 01-10-2024	Built Support Vector Machine (SVM) model and evaluated
Day 32	12-10-2024	Compared model results, selected top models for final analysis
Day 33- 34	13-10-2024 to 14-10-2024	Gender-wise churn analysis: examined churn rates among male and female customers
Day 35-37	15-10-2024 to 17-10-2024	Satisfaction-level churn analysis: examined churn rates by customer satisfaction
Day 38 - 39	18-10-2024 to 19-10-2024	Calculated overall churn rate and created visualization
Day 40 - 41	20-10-2024 to 21-10-2024	Drafted Chapter 4.6: Data Analysis and Interpretation based on findings
Day 42-43	22-10-2024 to 23-10-2024	Drafted Chapter 5: Summary of Findings, Conclusion, and Suggestions
Day 44	24-10-2024	Completed bibliography section, compiled references
Day 45	15-11-2024	Final review, proofreading, and submission of project report

Signature of the Student **Signature of the Guide**

APPENDIX 5B

Master Thesis Work
Work Done Diary for Academic Research

Sl. No.	Work to be Done	Date/s of Work Completion	Remarks	Signature of the Guide
01	Review of Literature and Research Design	10.08.2024	Comprehensive review of literature and framework	
02	Pilot Study	12.08.2024	Conducted pilot study to validate research approach	
03	Synopsis Submission	14.08.2024	Submitted synopsis with objectives and methodology	
04	1. Industry profile 2. 4/7 P's of Marketing for any 3 companies in the Industry or Porter's Five Force Model 3. Theoretical Background of the Study	10.09.2024	Prepared industry overview, comparative analysis, and theoretical grounding for study	
05	Collection of Data	15.09.2024	Collected survey responses and compiled data	
06	Data Analysis and Interpretation	01.10.2024	Analyzed data with EDA, logistic regression, decision tree, etc.	
07	Summary of Findings, Conclusions, and Suggestions	10.10.2024	Compiled findings, derived conclusions, and proposed suggestions	
08	Preparation and Submission of Report	15.11.2024	Finalized and submitted project report	

Appendix 6

Format for Executive Summary

A Study on Customer Churn Prediction in the Telecom Sector Using R Programming

International Institute of Business Studies

Master Thesis submitted in partial fulfillment of the requirements for the award of the Degree of

MASTER OF BUSINESS ADMINISTRATION
OF
BENGALURU CITY UNIVERSITY

BY

VARSHINI K

Reg. No: P18DM22M015183

Under the guidance of

Mr. Rakesh Singh

Professor

INTERNATIONAL INSTITUTION OF BUSINESS STUDIES

BENGALURU CITY UNIVERSITY

2022-24

Introduction Customer churn is a persistent challenge within the telecom industry, where intense competition drives customers to frequently switch between service providers. This research investigates customer churn prediction, aiming to identify the critical factors that influence a customer's decision to leave a telecom service provider. By applying machine learning techniques using R programming, the study explores customer behavior, satisfaction levels, and service experiences to provide telecom companies with data-backed insights for improving retention strategies. This analysis is pivotal in helping telecom firms maintain a competitive edge and prevent revenue loss from churn.

Need for the Study Retaining existing customers is significantly more cost-effective than acquiring new ones, making customer retention a priority for telecom providers. The study addresses the need for accurate churn prediction models to minimize churn rates. With effective prediction models, telecom companies can proactively address factors that cause churn, such as poor network quality, unsatisfactory customer service, or unclear billing. Understanding these issues through predictive analytics enables companies to offer targeted interventions, ultimately enhancing customer satisfaction and loyalty.

Scope of the Study The study examines churn patterns among customers of three leading telecom companies in India, using responses collected from a sample of 100 customers. It explores customer satisfaction across multiple service dimensions, including network quality, pricing, billing processes, and customer service support. This project also identifies the relative importance of these factors in predicting churn, providing a structured approach for companies to prioritize service improvements. The scope is limited to customers using mobile services in urban regions, ensuring a relevant focus for telecom firms operating in highly competitive markets.

Statement of the Problem Customer churn presents a significant threat to revenue stability for telecom operators, particularly in a market where competitors constantly seek to attract new customers through promotional offers and improved services. Many customers switch providers due to dissatisfaction with aspects such as pricing, network reliability, or customer service responsiveness. This research tackles the issue of predicting churn through machine learning techniques, using logistic regression, decision trees, and support vector machines (SVM). By leveraging these models, the study aims to identify behavioral patterns and factors that predispose customers to churn, allowing telecom companies to preemptively address these pain points. Churn prediction enables targeted retention strategies that not only reduce churn but also improve overall customer loyalty.

Objectives of the Study

- To identify and analyze patterns of customer behavior and satisfaction that are indicative of potential churn in the telecom industry.
- To develop and evaluate machine learning models to predict churn based on customer satisfaction levels and service interactions.
- To provide actionable recommendations for reducing churn and enhancing customer retention.

Research Design

- **6.1 Research Approach:** The research employs a quantitative approach, utilizing statistical and machine learning models to process and analyze data. The study combines customer survey data with predictive analytics to uncover insights into churn predictors.
- **6.2 Type of Sampling:** Convenience sampling was used to select participants who are telecom users, ensuring data collection from diverse demographics and service backgrounds.
- **6.3 Sample Unit:** The sample unit consists of individual telecom service users in India.
- **6.4 Sample Size:** 100 respondents participated in the survey, providing a representative sample for analysis.
- **6.5 Sources of Data:** The study uses primary data collected through a structured questionnaire that captures customer satisfaction, service issues, and churn history.
- **6.6 Data Collection Method:** A Google Forms survey was used to gather data, enabling convenient and widespread participation across different customer segments.
- **6.7 Plan of Analysis:** The analysis includes exploratory data analysis (EDA) to identify trends and correlations, followed by modeling with logistic regression, decision tree, and SVM to predict churn. Key variables are identified through feature importance analysis, providing insights into the factors most influential in churn behavior.

Highlights of Data Analysis

- **Churn Rate Analysis:** The churn rate in the sample was observed to be 33%, indicating a notable churn risk that telecom companies should address.
- **Demographic Insights:** Higher churn rates were observed among male customers as compared to female customers, highlighting potential demographic-specific issues that may require targeted strategies.
- **Satisfaction and Churn Correlation:** Lower levels of satisfaction in network quality, billing processes, and customer service were significantly associated with higher churn rates.

- **Model Performance:** Logistic regression and SVM models demonstrated strong predictive accuracy, with SVM achieving the highest accuracy rate of 82.76%. This underscores the model's capability to correctly identify potential churners based on satisfaction metrics.
- **Feature Importance Analysis:** Among satisfaction metrics, "Network Satisfaction" and "Billing Process Satisfaction" were the strongest predictors of churn, indicating that improvements in these areas could substantially impact customer retention.

Major Findings and Suggestions

- **Key Findings:**
  - Customers who reported lower satisfaction in network quality and billing processes were more likely to churn.
  - The logistic regression model showed that customer satisfaction with service quality and price was critical in determining churn likelihood, while SVM and decision tree models further confirmed these findings by highlighting network satisfaction as a priority.
- **Suggestions for Retention:**
  - Telecom operators should invest in improving network quality and reliability to meet customer expectations and reduce dissatisfaction-driven churn.
  - Billing transparency and responsive customer service are essential to retain customers, particularly in an industry where pricing and clarity are frequent concerns.
  - Proactive outreach strategies based on churn prediction models can help telecom companies engage at-risk customers before they consider switching, offering personalized incentives or improvements to address their concerns.

Research Contribution in Master Thesis This study contributes to the telecom industry by developing an effective customer churn prediction model that highlights key areas for improvement. By applying machine learning methods in R programming, this research offers telecom companies a structured, data-driven approach to reducing churn. The insights gained from feature importance analysis can inform strategic decisions in network quality, pricing, and customer service improvements. This project ultimately equips telecom providers with predictive capabilities to enhance customer satisfaction, foster loyalty, and improve financial performance in a highly competitive market.

Varshini_k_Plagarism.pdf

ORIGINALITY REPORT

10%	7%	1%	5%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	www.ijsdr.org Internet Source	2%
2	Submitted to Colorado Technical University Student Paper	2%
3	docshare.tips Internet Source	1%
4	vdocument.in Internet Source	1%
5	www.scripown.com Internet Source	1%
6	Submitted to International School Eindhoven Student Paper	1%
7	www.termpaperwarehouse.com Internet Source	1%
8	Sharda Singh, A.K. Subramani, Rajasekhar David, N. Akbar Jan. "Workplace ostracism influencing turnover intentions: Moderating roles of perceptions of organizational virtuousness and authentic leadership", Acta Psychologica, 2024 Publication	1%
9	www.irjmets.com Internet Source	1%

Exclude quotes Off

Exclude bibliography On

Exclude matches Off