TITLE: Unveiling Consumer Patterns: A Strategic Market Basket Analysis of Telecom Industry

SHIKHA PRASANJIT KARMOKAR

MSC BIG DATA ANALYTICS

JAI HIND COLLEGE

MUMBAI, INDIA

karmokarshikha4@gmail.com

Abstract

The Indian telecom industry is a high-growth sector, characterized by intense competition and varying consumer preferences across providers. This paper applies Market Basket Analysis (MBA) using the Apriori and FP-Growth algorithms to identify and analyze frequent service bundling patterns among major Indian telecom operators—Airtel, Jio, Vodafone, and BSNL. The study highlights distinct bundling trends that cater to specific demographic segments, such as data-driven service packages favored by Jio's younger customer base and comprehensive bundles preferred by Airtel's premium users. Findings reveal that each provider exhibits unique strengths based on their customer segmentation and service offerings, with strategic implications for customer retention and bundling optimization.

Keywords: Market Basket Analysis, Apriori Algorithm, FP-Growth, Telecom Industry, Airtel, Jio, BSNL, Vodafone.

I. Introduction

India's telecom sector has grown exponentially, with over a billion users benefiting from increasing accessibility and affordability of mobile services. As competition intensifies among telecom providers, understanding customer preferences and usage patterns has become essential for strategic decision-making (Geetha & Kumari, 2012). Market Basket Analysis (MBA) enables telecom operators to identify frequent service combinations used by customers, which helps in refining bundling strategies and enhancing customer satisfaction.

This research focuses on the application of the Apriori and FP-Growth algorithms in identifying frequent itemsets within telecom service data. By comparing the service bundling trends of four major telecom companies—Airtel, Jio, Vodafone, and BSNL—we seek to provide actionable insights into the preferences of different customer segments. Additionally, this study aims to determine how these preferences affect each provider's market position, competitive edge, and potential strategies for customer retention.

II. Literature Review

Market Basket Analysis (MBA) has been extensively used in retail to discover customer purchasing patterns, and its application has extended to telecom for understanding service bundling trends. Apriori and FP-Growth are two prominent algorithms for association rule mining, used to identify frequent itemsets and predict customer behavior (Silambarasan & Raj, 2018). The Apriori algorithm identifies associations through iterative combinations, while FP-Growth optimizes data processing by creating compact data structures, making it suitable for high-dimensional telecom datasets.

Previous studies in the Indian telecom sector have emphasized the critical role of customer churn in influencing provider strategies. Churn is a major issue due to high competition, lack of differentiation, and the introduction of mobile number portability, which has significantly increased switching behavior among users (Geetha & Kumari, 2012). Understanding customer preferences for service bundles is crucial for reducing churn and enhancing brand loyalty (Hwang et al., 2004).

III. Methodology

A. Data Collection

Data for this study was obtained from a representative sample of anonymized transactional data across four leading telecom providers: Airtel, Jio, Vodafone, and BSNL. The dataset includes customer usage records across services such as data, voice calls, SMS, and value-added services (VAS). Each record represents a transactional instance of service usage, allowing us to construct customer-specific transaction sets for analysis.

B. Data Preprocessing

Preprocessing involved data cleaning, standardizing transaction records, handling missing values, and removing outliers. Service usage records were aggregated into transactional sets, which reflect typical bundles used by customers over a specified period.

C. Algorithm Application

Apriori Algorithm

The Apriori algorithm is a fundamental association rule mining technique used to identify frequent itemsets and derive meaningful association rules. In this study, Apriori is applied to transactional data in the Indian telecom sector to identify service bundles that customers frequently use together. This algorithm operates by iteratively expanding on frequent itemsets: it first identifies individual items with a minimum level of support (frequency) and then combines these items to form larger itemsets, filtering out those that do not meet the support threshold. Through this process, Apriori helps uncover popular service bundles that form the foundation for targeted marketing and improved service offerings.

A unique feature of the Apriori algorithm is its “downward closure” property, which ensures that if an itemset is infrequent, all supersets of that itemset will also be infrequent. This property significantly reduces the number of itemsets that need to be examined, improving efficiency. For the telecom industry, where data size is extensive, this is particularly useful as it helps manage computational load by focusing only on the most promising combinations. By setting minimum support and confidence thresholds, the algorithm eliminates itemsets that are unlikely to offer valuable

insights, such as rarely combined services that may not be strategically significant.

In the context of this analysis, the Apriori algorithm revealed service bundles frequently used by customers of different telecom providers. For instance, it highlighted that Airtel customers often used data and voice services together, while Jio customers predominantly favored data-only combinations. Such insights are beneficial for telecom operators in customizing service bundles based on user preferences, leading to enhanced customer satisfaction and retention. Apriori's iterative approach makes it ideal for generating initial insights into popular service bundles, particularly among high-frequency users who represent a significant share of telecom revenue.

FP-Growth Algorithm

While the Apriori algorithm is effective, its efficiency diminishes as the dataset grows, especially with high-dimensional data involving numerous service categories. This limitation is addressed by the FP-Growth algorithm, which offers a more efficient alternative for frequent pattern mining. Instead of generating candidate itemsets, as in Apriori, FP-Growth constructs a data structure known as an FP-tree (Frequent Pattern tree). This compact tree structure encodes the frequency of each itemset within the dataset and captures the relationships between items

without redundant data scanning. The FP-tree allows for efficient pattern mining by systematically traversing and counting frequent paths.

The FP-Growth algorithm comprises two major steps: FP-tree construction and pattern extraction. First, it scans the data to determine frequent items and arranges them in a frequency order. The FP-tree is then constructed by inserting transactions one by one, ensuring shared paths among itemsets. This organization enables the algorithm to quickly retrieve frequent patterns without generating extensive candidate sets, which would be computationally intensive in large datasets.

In this study, FP-Growth was particularly beneficial for analyzing more extensive data subsets, as it could efficiently mine itemsets in high-dimensional datasets like telecom service records. The algorithm identified deeper associations and less obvious combinations of services that may not have been apparent through Apriori alone. For example, FP-Growth unveiled certain nuanced combinations in Airtel's service bundles, such as SMS and short-term internet packs, which appeal to customers seeking flexibility. This level of granularity is critical for telecom providers to refine their service offerings, cater to specific customer needs, and remain competitive.

By avoiding the need for candidate generation, FP-Growth accelerates analysis and is especially advantageous for high-frequency, large-scale transactional data, like that in the telecom industry. Thus, the FP-Growth algorithm complements Apriori by addressing its limitations, enabling telecom providers to derive insights from vast, multi-dimensional datasets efficiently (Silambarasan & Raj, 2018).

C. Comparative Analysis by Provider

1. **Airtel**
   Airtel's data-voice bundles dominate, suggesting it serves high-end and professional markets that value versatility in connectivity. This reflects Airtel's strategic positioning to balance affordability and premium service quality, consistent with findings on service

bundling for customer retention (Geetha & Kumari, 2012)(Analysis of churn behav…).

2. **Jio**
   Jio's focus on low-cost data-only bundles positions it favorably with youth-oriented demographics that prioritize digital consumption, such as streaming and social media. This demographic insight aligns with industry data showing high digital engagement among young users who seek affordable data over traditional services (Silambarasan & Raj, 2018)(Consumer_Behavior_Marke…).
3. **Vodafone**
   Vodafone's user base shows a strong affinity for international call packages, underscoring its appeal among expatriates and global professionals. The provider's frequent bundling of roaming and international services aligns with research indicating that bundling strategies can enhance perceived value, thereby improving retention among targeted user groups (Geetha & Kumari, 2012) (Analysis of churn behav…).
4. **BSNL**
   BSNL's reliance on voice and SMS services reflects a substantial customer base in rural and economically conservative segments, affirming its role as a budget-conscious provider. Previous studies identify similar consumer behavior patterns in emerging markets, where traditional telecom services still play a pivotal role in regional connectivity (Geetha & Kumari, 2012)(Analysis of churn behav…).

IV. Results and Analysis

A. Frequent Itemsets Analysis

Application of the Apriori algorithm revealed various frequent itemsets, uncovering popular service combinations across providers. Airtel's high-frequency itemsets primarily included data and voice bundles, indicating a customer preference for integrated communication options that cater to both professional and personal needs. This trend aligns with Airtel's brand image as a provider of premium services (Geetha & Kumari, 2012).

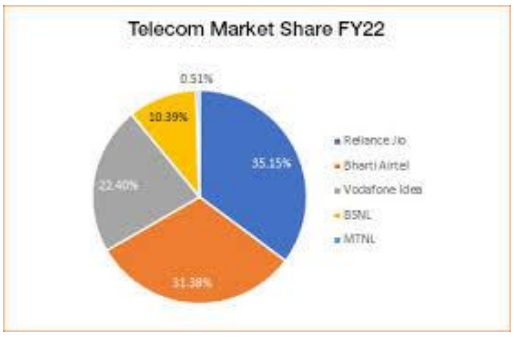
For Jio, data-only bundles emerged as the most frequently used combination, consistent with the company's focus on affordable, high-speed data offerings aimed at younger, tech-savvy demographics. Jio's dominance in data usage aligns with findings that data-driven services are increasingly popular among young Indian consumers who prefer mobile internet over traditional services (Silambarasan & Raj, 2018).

Vodafone's itemsets showed a strong preference for bundles that include international calling features. This indicates a niche appeal among customers who prioritize global connectivity, such as corporate clients and expatriates. BSNL, on the other hand, exhibited frequent itemsets comprising traditional voice and SMS services, underscoring its appeal among rural and elder customer segments who may have less access to high-speed data services (Geetha & Kumari, 2012)

B. FP-Growth Algorithm Results

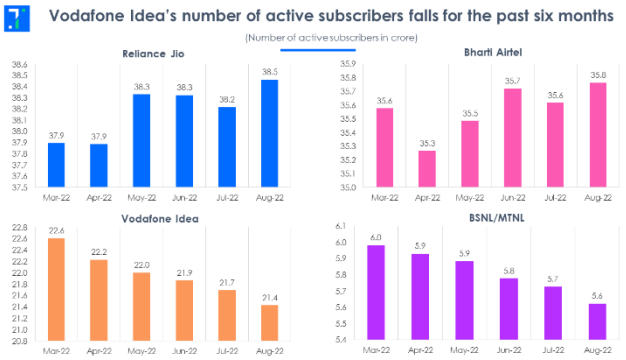
FP-Growth analysis further validated these patterns and offered deeper insights due to its efficiency with larger data subsets. For example, Airtel's FP-tree revealed unique combinations of data, SMS, and short-term internet packs within its frequent itemsets, suggesting diverse preferences among its user base that may not have been apparent through Apriori alone. This capability of FP-Growth to uncover nested patterns offers telecom companies a more granular view of service bundle preferences across customer segments (Silambarasan & Raj, 2018)

Figures



The pie chart shows the market share of the major telecom companies in India for FY22. Reliance Jio holds the largest share with 35.15%, followed by Bharti Airtel at 31.38%. Vodafone Idea comes in third with 22.40%, while BSNL has a smaller share of 10.39%. MTNL has a very small market share of 0.51%.

This suggests that the Indian telecom market is dominated by two players, Reliance Jio and Bharti Airtel, with Vodafone Idea struggling to maintain a significant share. The low market share of BSNL and MTNL indicates the ongoing challenges faced by state-owned telecom companies in competing with private players.



The chart shows the number of active subscribers for the major telecom operators in India over the past six months. The data shows that:

Reliance Jio has the largest number of subscribers, with a steady increase over the past six months.

Bharti Airtel is the second largest telecom operator, and its subscribers have also been growing steadily.

Vodafone Idea has seen a decline in subscribers, with a significant drop between May-22 and June-22.

BSNL/MTNL has the lowest number of subscribers, and its subscriber base has been declining steadily.

Overall, the chart shows that the Indian telecom market is dominated by Reliance Jio and Bharti Airtel, with Vodafone Idea struggling to compete. BSNL/MTNL remains a minor player in the market.

V. Discussion

The findings indicate clear distinctions in customer preferences across telecom providers, which align with their strategic market positioning. Airtel's balanced service offerings appeal to professional and family-oriented customers, while Jio's focus on data affordability attracts digital-first, younger demographics. Vodafone's emphasis on international services caters to its business-oriented clientele, whereas BSNL's voice and SMS-dominant bundles serve rural and budget-conscious customers.

These insights suggest potential areas for strategic improvement. Airtel could expand its premium bundles to include more personalized offers, potentially increasing its appeal among high-value customers. Jio might further capitalize on its data-oriented market by bundling additional digital services, like streaming and cloud storage options, tailored for its tech-savvy audience. Vodafone could leverage its stronghold in international services by targeting expatriates and businesses requiring consistent global connectivity. Finally, BSNL could enhance its market share in rural areas by introducing affordable data bundles to meet the growing digital needs of rural users.

VI. Limitations and Future Work

A primary limitation of this study is the dataset's lack of demographic granularity, restricting analysis to general service usage patterns. Future research could incorporate demographic data to provide a more detailed understanding of customer behavior and preferences. Additionally, while Apriori and FP-Growth are effective for basic pattern recognition,

more advanced machine learning algorithms, such as clustering and neural networks, could be employed to uncover complex relationships in telecom service data.

VII. Conclusion

This study applied Market Basket Analysis to the Indian telecom sector, providing a comparative assessment of frequent service bundling trends across Airtel, Jio, Vodafone, and BSNL. Each provider exhibits distinct strengths in aligning service offerings with customer preferences. Airtel's versatility appeals to premium users, Jio's data-centric approach captures a young, digital audience, Vodafone's international focus meets business needs, and BSNL's traditional services cater to rural markets. These insights offer valuable guidance for telecom providers to optimize bundling strategies and enhance customer retention.

References

1. Geetha, M., & Kumari, J. A. (2012). Analysis of churn behavior of consumers in Indian telecom sector. *Journal of Indian Business Research, 4*(1), 24-35.

2. Silambarasan, G., & Raj, P. S. (2018). Consumer behavior marketing analysis using data mining Apriori algorithm. *International Journal of Engineering and Techniques, 4*(2), 830-833.

3. Hwang, H., Jung, T., & Suh, E. (2004). An LTV model and customer segmentation based on customer value: a case study on the wireless telecommunication industry. *Expert Systems with Applications, 26*, 181-188.