**"PREDICTIVE MODELLING OF CONSUMER LOYALTY: ANALYZING RFM (RECENCY, FREQUENCY, AND MONETARY) AND SATISFACTION FACTORS"

***Dr S.B. Muley, Kartavi Naik, Lajari Murkar, Nikita Pakhale, Riya Gidwani**
*Assistant Professor * Department of Statistics Kishinchand Chellaram College*
Vidyasagar Principal K.M. Kundnani Chowk, 124, Dinshaw Vacha Road, Churchgate, Mumbai 400020
**sakharam.muley@kccollege.edu.in*

Abstract
The study is designed to develop a predictive model for a consumer to be Loyal online or Loyal offline. The loyalty of the customer is decided based on Recency, Frequency, and Monetary, or RFM, which are the three main variables that are used to assess consumer behavior and independent variables for which responses are collected as satisfaction levels on product quality, discounts, and offers, availability of a range of products, etc. In this study, a detailed survey was conducted and responses were collected. These are further analyzed using different statistical methods like Logistic Regression, Naïve Bayes Classification, and Support Vector Machine.
***Keywords:** Online shopping, retailers mechanism, E-commerce, Loyal online, Loyal offline*

****This study is part of a Guided Research Project at M.Sc. Students at Department of Statistics, K.C. College, Mumbai.**

Introduction

The online purchase has become a routine shopping way for people living in the cities. Online shopping became the most convenient and practical option for them as it provides a facility of 'Buy any time anywhere'. The advancement in internet infrastructure and online payment systems using mobile has significantly given a boost the consumers to shop online.

According to IBEF, In FY23, the Gross Merchandise Value (GMV) of e-commerce reached US$ 60 billion, increasing 22% over the previous year. In FY22, the GMV of e-commerce stood at US$ 49 billion. India's Business-to-Business (B2B) online marketplace will be a US$ 200 billion opportunity by 2030. With over 800 million users, India was the second-largest internet market in the world with 125.94 lakh crore UPI transactions in 2022.

According to a Deloitte India Report, as India is moving towards becoming the third-largest consumer market, the country's online retail market size is expected to reach US$ 325 billion by 2030, up from US$ 70 billion in 2022, largely due to the rapid expansion of e-commerce in tier-2 and tier-3 cities.

The e-commerce market's share of Tier-3 cities grew from 34.2% in 2021 to 41.5% in 2022, shows data. See the chart below which shows the increasing trend over the years in the future. Website

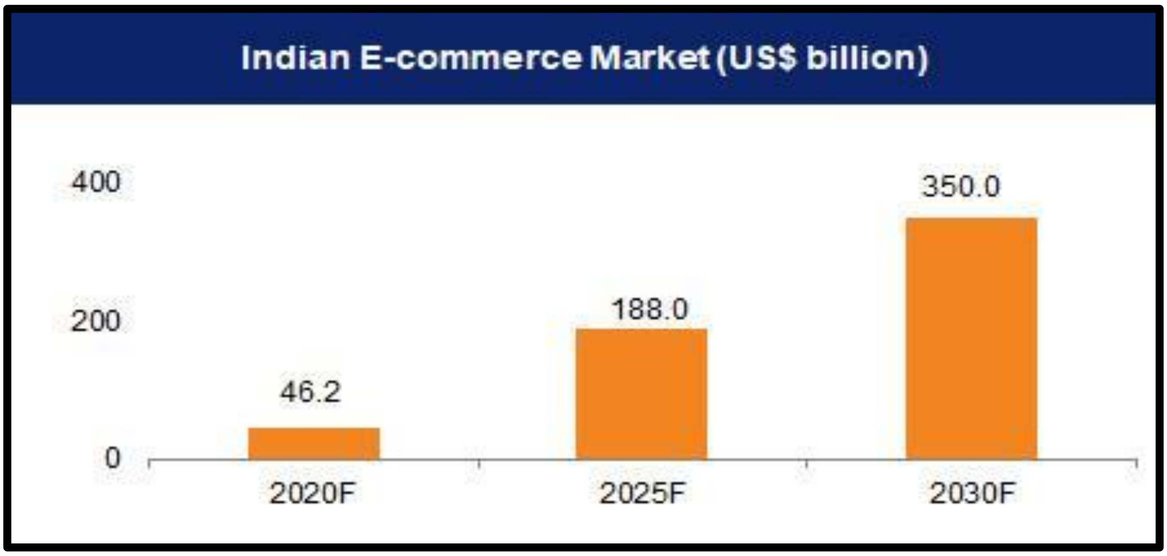

Chart 1: The chart is taken from the IBEF website showing Indian E-commerce Market for 2020 and projected for 2025, and 2030.

But still, many studies have recognized in-store purchases as a preference for some specific types of products and online for some specific types of products. Customers who choose to shop online choose to purchase a product from a website, App, etc. This data provides a leading edge for an opportunity to be a part of such a preferred chain.

Therefore, understanding consumer behavior in online purchasing and in-store purchasing is treated as an applied discipline in which some factors have a more significant impact on customers' behavior and expected actions. If e-market entrepreneurs are aware of the factors that influence online behavior and the relationships between these factors, they can build new marketing strategies based on these Critical Customer aspects to convert potential customers into active customers. Similarly, this will also help offline stores to stand in this storm of online stores.

Literature Review

Multiple studies have been conducted to examine various aspects of online shopping. For instance, Agarwal and Joseph (2021) conducted a study in India to identify key variables that influence trust and risk perception among online customers. Similarly, Barbosa (2018) investigated how digital platforms are being used for buying goods while inside a physical store. Dubey and Balaji (2021) explored the challenges faced by e-commerce businesses in managing customer expectations. Mr. G. Bhongade et al. (2018) presented a semi-supervised approach for opinion mining using online product reviews. Prashar et al. (2015) identified and ranked the factors that influence the selection of web portals among online shoppers in India, while their 2019 study identified four types of online shoppers and suggested strategies for targeting them effectively. Rao and Patro (2017) identified variables that influence online shopping behavior and proposed measures to make the web shopping experience more effective and trustworthy. Finally, Reddy and Prasad (2012) analyzed consumer perceptions of e-portals' characteristics and the factors influencing trust and privacy. Shaheen et al. (2019) studied the role of online reviews in driving customer engagement in e-commerce through online devices. These studies provide valuable insights into the evolving needs and preferences of consumers in the digital age, helping businesses to develop effective strategies to stay ahead of the competition.

Research Methodology

The study design for the current research is of descriptive and analytical type. It is descriptive in the sense it exists at present and it includes facts and findings and analytical as it involves analysis to test the hypothesis. The relevant data were collected through a questionnaire for which the Cronbach alpha value for the questionnaire is 0.89 which shows acceptable reliability for the instrument.

Using the questionnaire so designed the data is collected from 324 respondents in either personal or online circulation. To determine which customers are most likely to respond to a marketing effort or develop into loyal customers, Recency, Frequency, and Monetary or RFM analysis serves as a method for customer segmentation. Recency, Frequency, and Monetary, or RFM, are the three main variables that are used to assess consumer behavior.

Recency refers to how recently the customer made a purchase, Frequency refers to how often a customer makes a purchase, and Monetary refers to how much money a customer spends on purchases.

By examining these criteria, organizations may identify their most valued consumers, target specific customer groups with targeted marketing campaigns, and establish strategies to boost customer loyalty and retention.

RFM Score calculation

- Recency was measured on a scale of 1 to 5, where 1 is the purchase a month ago and 5 being the purchase a year ago. Customers with high recency are the ones with scales 1, 2, and 3 and customers with low frequency are the ones with scales 4 and 5
- Frequency was measured on a scale of 1 to 4, high-frequency customers fall under the scale of 1 and 2, and low-frequency customers fall under the scale of 3 and 4.
- Spending was measured on a scale of 1 to 4, where high-spending customers are the one who spends 2000- more than 5000 and falls under a scale of 3 and 4, and low-spending customers are the one who spends less than 1000-2000 and falls under the scale of 1 and 2

E.g. RFM Score of 1,1,1 means, high recency, high frequency, and low spending.

Based on the RFM Score customers were segmented into 2 Classes, loyal online consumers, and loyal offline consumers.

- Loyal online consumers are consumers with high recency, high frequency, and high spending.
- Loyal offline consumers are consumers with low recency, low frequency, and low spending.

Here in this case the 2 classes segmented based on RFM as mentioned above are the dependent variable and factors like Product Quality, Discounts and Offers, product prices, Return Policy, Exchange Policy, Payment options, Availability of a wide range of brands, Varieties of items, and Shipping time as independent variable.

Respondents were asked to rate their purchase satisfaction from 1 to 5 and 1 being least satisfied and 5 being highly satisfied. The binary logistic regression, Naïve Bayes Classification, and Support vector machine (SVM) techniques are used to analyze whether the selected respondent can be considered as Loyal online or Loyal offline and what factors make respondents loyal online or offline. This will help industries, online stores, or offline stores could include them in their critical quality parameters. The data was analyzed using Python software.

Result and Discussion

Binary Logistic Regression Analysis

Logistic regression is a powerful statistical tool that can help us achieve our objective of identifying the factors that significantly influence customer loyalty to an online purchasing platform.

Results

Variables	Coefficient value	p-value	Interpretations
Product Quality	-0.62821	0.006622	Significant Negative
Discounts and Offers	0.122438	0.568365	Non-Significant Positive
product prices	-0.02794	0.895578	Non-Significant Negative
Return Policy	-0.17065	0.494906	Non-Significant Negative
Exchange Policy	0.029756	0.884263	Non-Significant Positive
Payment options	-0.20205	0.392011	Non-Significant Negative
Availability of a wide range of brands	0.57463i13	0.011167	Significant Positive
Varieties of items	0.245755	0.341218	Non-Significant Positive
Shipping time	0.064378	0.769521	Non-Significant Positive

Model Accuracy: 88%

Findings

- 'Product Quality' (p-value<0.05), 'Products Prices', 'Return Policy', 'Exchange Policy', 'Payment options', and 'Shipping time' have negative coefficients, indicating that they have a negative impact on customer loyalty and significant in case of 'Product Quality' shows that more the satisfaction towards product quality, less likely the consumer loyal to online.
- On the other hand, 'Discounts and Offers', 'Availability of a wide range of brands' (p-value<0.05), and 'Varieties of items' have positive coefficients, positively impacting customer loyalty and significant in the case of 'Availability of a wide range of brands' indicating more the satisfaction more likely the person is loyal to online.
- On passing the unseen data for 2 metrics the model predicted that they belong to the loyal online consumer segment.

Naïve Bayes Classification

Objective

To develop a model to predict customer loyalty segment based on satisfaction level and to identify the factors that have a significant impact on customer loyalty segment.

Naive Bayes is a kind of classifier that uses the Bayes Theorem. It predicts membership probabilities for each class such as the probability that a given record or data point belongs to a particular class. The class with the highest probability is considered the most likely class.

Mathematical Model

We want to predict the Loyalty Segment of a customer based on their features such as Product Quality, Discounts, and Offers, Products Prices, Return Policy, Exchange Policy, Payment options, Availability of a wide range of brands, Varieties of items, and Shipping time.

To do this, Naive Bayes calculates the probability of each possible Loyalty Segment given the values of these features. It assumes that the features are conditionally independent given the Loyalty Segment, meaning that each feature does not influence the other features.

Based on this assumption, the conditional probability of a Loyalty Segment y given the feature vector X can be calculated as follows:

P(y | X) = P(y) * P(Product Quality | y) * P(Discounts and Offers | y) * P(Products Prices | y) * P(Return Policy | y) * P(Exchange Policy | y) * P(Payment options | y) * P(Availability of wide range of brands | y) * P(Varieties of items | y) * P(Shipping time | y)

P(y) is the prior probability of the Loyalty Segment y, which can be estimated using the training data.

Finally, the Naive Bayes algorithm chooses the Loyalty Segment, y that maximizes the posterior probability P(y | X). In other words, it predicts the

Loyalty Segment with the highest probability based on the given feature values.

Naive Bayes classification can help us identify the factors that have the greatest impact on customer loyalty, by calculating the conditional probabilities of each independent variable given each loyalty segment. These conditional probabilities can then be used to rank the independent variables in terms of their influence on customer loyalty.

Results of Naïve Bayes Classification

Results

Variables	Coefficient value	p-value	Interpretations
Constant	1.579117	0.038151	Significant Positive
Product Quality	-0.671663	0.006622	Significant Negative
Discounts and Offers	0.1114085	0.568365	Non-Significant Positive
product prices	-0.030929	0.895578	Non-Significant Negative
Return Policy	-0.189976	0.494906	Non-Significant Negative
Exchange Policy	0.037524	0.884263	Non-Significant Positive
Payment options	-0.211573	0.392011	Non-Significant Negative
Availability of a wide range of brands	0.610514	0.011167	Significant Positive
Varieties of items	0.244380	0.341218	Non-Significant Positive
Shipping time	0.068323	0.769521	Non-Significant Positive

Model Accuracy: 84%

Findings

- The 'Product Quality' input feature has a negative coefficient of -0.671663 and a low p-value of 0.006622, indicating that it is significant and negatively affects the target variable and significant in the case of 'Availability of a wide range of brands' indicating more the satisfaction less likely the person is loyal to online.
- The 'Availability of a wide range of brands' feature has a positive coefficient of 0.610514 and a low p-value of 0.011167, indicating that it is significant and positively affects the target variable and the significant in case of 'Availability of a wide range of brands' indicating more the satisfaction more likely the person is loyal to online.
- The other input features do not have significant p-values, suggesting that they are not important for predicting the target variable in this model.
- The prediction on unseen data through naïve Bayes predicted that the customer will be classified as a loyal online consumer.
- For a new test case the predicted probability of the target variable being in class 1 (or any specific class, depending on the model setup) is $\exp(1.579117) / (1 + \exp(1.579117)) = 0.829$.

Support Vector Machine

Objective

To identify the factors that significantly influence customer loyalty to an online purchasing platform.

Support Vector Machine (SVM) is a Supervised Machine Learning algorithm that is used for regression and/or classification. Although it is occasionally quite helpful for regression, classification is where it is most often used. In short, SVM identifies a hyper-plane that establishes a distinction between the different kinds of information.

SVM works by mapping data to a high-dimensional feature space so that data points can be categorized, even when the data are not otherwise linearly separable. A separator between the categories is found, then the data are transformed in such a way that the separator could be drawn as a hyperplane.

SVM is used because it helps us to determine the most crucial attributes that help forecast customer loyalty in terms of our objective of examining the variables that influence customers' loyalty to your online shopping platform.

Based on the values of the input attributes, the SVM algorithm finds the hyperplane that best divides the data into distinct classes (in your example, loyal vs. non-loyal consumers). The weights allocated to each feature in the hyperplane equation are represented by the coefficients you acquire from SVM, which show how significant each feature is in predicting customer loyalty.

Hence, you may identify which factors have the most effects on client loyalty and change your marketing approach accordingly by studying the coefficients and p-values of the SVM model.

Results

Variables	Coefficient value	p-value	Interpretations
Product Quality	-0.0001769	0.005393	Significant Negative
Discounts and Offers	-0.0000110	0.583045	Non-Significant Negative
product prices	0.0001036	0.990823	Non-Significant Positive
Return Policy	0.0000399	0.580631	Non-Significant Positive
Exchange Policy	-0.0000323	0.888167	Non-Significant Negative
Payment options	0.0000427	0.271933	Non-Significant Positive
Availability of a wide range of brands	0.0000173	0.011011	Significant Positive
Varieties of items	0.0000883	0.306320	Non-Significant Positive
Shipping time	0.0000011	0.886314	Non-Significant Positive

Model Accuracy: 84%

Findings

- The p-values of the coefficients indicate the statistical significance of the relationship between each feature and the Loyalty Segment.
- A p-value less than 0.05 suggests that the relationship is statistically significant, while a p-value greater than 0.05 suggests that the relationship is not statistically significant.
- In this case, only the features Product Quality, and Availability of a wide range of brands have p-values less than 0.05, indicating that they are statistically significant. This indicates that the satisfaction in the case of 'Availability of a wide range of brands' indicates more likely the person is loyal online. But also the results of ‘Product Quality’ indicate that more the satisfaction in case of 'Product Quality' indicates less likely the person is loyal to online.

Conclusion

The results of the Logistic Regression, Naive Bayes, and SVM models all suggest that the quality of the products and the availability of a wide range of brands have a significant impact on customer loyalty. Other factors such as product prices, return policy, exchange policy, payment options, and shipping time did not show a significant impact on customer loyalty in any of the models. Based on these results, we can suggest some marketing strategies to be adopted by the stores. The results of the product quality indicate that if the consumer is highly satisfied in online shopping for the product, the customer may be more loyal offline for that product and this shows that the loyal online consumer gets transformed into a loyal offline one.

Marketing Strategies

1. *Improve product quality*: As the logistic regression, Naïve Bayes, and SVM models all indicate that product quality has a significant impact on customer loyalty, focusing on improving product quality can be a viable strategy. This could involve investing in better raw materials, improving production processes, and introducing quality control measures to ensure consistent quality.
2. *Offer discounts and promotions*: The logistic regression model indicates that discounts and offers have a positive impact on customer loyalty. Offering discounts and promotions can help attract new customers and retain existing ones. These discounts could be targeted at specific products, time-bound offers, or loyalty-based discounts.
3. *Expand brand and product range*: The logistic regression and Naive Bayes models indicate that the availability of a wider range of brands and product varieties has a positive impact on customer loyalty. Offering more brands and products can attract a wider customer base and increase loyalty among existing customers.
4. *Improve customer service*: Although not explicitly mentioned in the models, good customer service is a key factor in customer loyalty. Improving customer service can involve investing in training for customer service personnel, implementing customer feedback systems, and addressing customer complaints promptly.
5. *Streamline return and exchange policies*: The logistic regression model indicates that return and exchange policies hurt customer loyalty. Streamlining these policies can help reduce customer frustration and improve loyalty.
6. *Improve shipping time*: The logistic regression model indicates that shipping time has a negative impact on customer loyalty. Improving shipping time can involve investing in faster shipping methods, better logistics management, and improving supply chain efficiency.

References

1. Agarwal, Anuraag, and Sanskrity Joseph. "A review on Effect of e trust and e risk on Consumers of retail e-markets in India: A Comparative Study Based on Sociodemographic Variables." Journal of Ravishankar University(PART-A) 26, no.

1(February 12, 2021):49–62. http://dx.doi.org/10.52228/jrua.2020-26-1-8.
2. Barbosa, Diogo. "Digital In-Store Shopping." U.Porto Journal of Engineering 2, no. 1(March 19, 2018):44–48. http://dx.doi.org/10.24840/2183-6493_002.001_0005
3. Dubey, Anil Roy, and B. Balaji. "A Study on E-Commerce Aspects Impacting Consumer's Online Shopping Behavior in United Arab Emirates." Journal of Computational and Theoretical Nanoscience 18, no. 4 (April 1, 2021):1331–35. http://dx.doi.org/10.1166/jctn.2021.9393
4. G. Bhongade, Bhagyashree, and Ashwini V.Z. "A Semi-supervised Approach for Opinion Mining using Online Product Review." International Journal of Engineering & Technology 7, no. 4.5 (September 22, 2018):143. http://dx.doi.org/10.14419/ijet.v7i4.5.20031
5. Prashar, Sanjeev, T. Sai Vijay, and Chandan Parsad. "Antecedents to Online Shopping." International Journal of E-Business Research 11, no. 1(January 2015):35–55. http://dx.doi.org/10.4018/ijebr.2015010103.
6. Prashar, Sanjeev, Sai Vijay Tata, Chandan Parsad, Abhishek Banerjee, Nikhil Sahakari, and Subham Chatterjee. "Clustering E-Shoppers based on Shopping Values and Web Characteristics." Journal of Global Information Management 27, no. 2 (April 2019): 24–38. http://dx.doi.org/10.4018/jgim.2019040102.
7. Rao, Katta Rama Mohana, and Chandra Sekhar Patro. "Shopper's Stance Towards Web Shopping." International Journal of Online Marketing 7no. 3(July 2017):42–54. http://dx.doi.org/10.4018/ijom.2017070104
8. Reddy, Anuradha, and G. V. Bhavani Prasad. "Consumer Perceptions on Security, Privacy, and Trust on E-Portals." International Journal of Online Marketing 2, no. 2(April 2012):10–24. http://dx.doi.org/10.4018/ijom.2012040102.
9. Srivastava, Gautam, Neeraj Anand, and Arvind Kumar Jain. "Impact of supply chain management on the digital marketing of perishable goods (fruits and vegetables)." Journal of Global Economy 14, no. 4 (November 8, 2018): 51–60. http://dx.doi.org/10.1956/jge.v14i4.497.
10. Shaheen, Musarrat, Farrah Zeba, Namrata Chatterjee, and Raveesh Krishnankutty. "Engaging customers through credible and useful reviews: the role of online trust." Young Consumers 21, no. 2 (September 19, 2019): 137–53. http://dx.doi.org/10.1108/yc-01-2019-0943.