ANALYZING AND PREDICTING B2B CUSTOMER REPURCHASING INTENTIONS USING MACHINE LEARNING

*Dr.R.Mary Metilda,Professor, SREC Business School, Sri Ramakrishna Engineering College (Corresponding Author)
**Dr.R.Joshua Roy, Associate Professor, SRCAS Business School, Sri Ramakrishna College of Arts and Science
***Dr.Blessy Doe, Asst.Professor, SREC Business School, Sri Ramakrishna Engineering College
****Dr.V.K.Arthi, Asst.Professor, SREC Business School, Sri Ramakrishna Engineering College
*****Ms.A. Sagayarani, Asst.Professor, , SREC Business School, Sri Ramakrishna Engineering College

ABSTRACT

In the present business environment, organizations are realizing that customer repurchase intention is a pivotal aspect of the growth of their business. This research focuses on the repurchasing intention of the business-to-business (B2B) customers, the factors that influence this decision, and the use of models that could predict such a decision. Therefore, satisfaction, retention, relation, after sales service, perceived value, and loyalty to outline repurchase intentions are considered. Since this is an analytical research design, the key methods used in the research include SHAP analysis, random forest regression, and correlation heatmap analysis. The questionnaires were distributed among the existing customers; the research involved 175 industries at the Ahmedabad district of Gujarat out of which 76 responses were received which indicates the sample size percentage of 43%. The model also had a good predictive value with the R-squared value, which was equal to 0. 986 on the training dataset. The performance was 84.4% for the test dataset. Among the analysed factors, Perceived Value came out as the leading factor and manifested its great influence on the customer repurchasing process. The research states that the areas for improvement should be directed on Customer Satisfaction and Service & Support respectively in order to match customer repurchasing.

Key words: Repurchasing Intention, Satisfaction, Perceived Value, Customer Relations, Retention, Brand Loyalty Service & Support and Machine Learning.

1. INTRODUCTION

In the present business to business market environment, organizations are realizing that customer repurchase intention is a pivotal aspect of the growth of their business. Re-purchase intention is a term that means the potential or probable chance of a customer to keep buying products and services from a particular company in the future. It is a gauge of a consumer`s necessity of the experience or willingness to enter into the successive buying activity with the provider. Customer repurchase intention is not only a tool to acquire immediate sales but also about cultivating lasting ties that will build stakeholders and sustained profitability over time. Factors affecting repurchase intention are multifaceted. Customer satisfaction remains a critical determinant of repurchase intention and loyalty, emphasizing the importance of consistently delivering positive customer experiences to foster long-term relationships (Cheung, C. M., & Huang, Y. 2020). Perceived value emerges as a key driver of customer satisfaction and repurchase intention, highlighting the need for businesses to continually enhance value propositions to meet evolving customer expectations (Kim, H. W., & Kankanhalli, A. 2020). Strong customer relationships characterized by trust, empathy, and personalized interactions are essential for fostering loyalty and increasing customer repurchasing rates (Xu, Y., & Tan, C. W. 2021). Service quality plays a crucial role in shaping customer perceptions and behaviours, with superior service leading to increased satisfaction, loyalty, and repurchase intention (Li, X., & Wang, D. 2020). Brand loyalty is characterized by a strong emotional connection and preference for a particular brand, driving repeat purchases and positive word-of-mouth referrals (Batool, S. A., & Qureshi, J. 2021).

2. OBJECTIVES OF THE STUDY

1. To Identify the factors that influence customers' intentions to repurchase products
2. To evaluate the overall relationship between factors such as customer satisfaction, brand loyalty, retention, relation, service & support and perceived value.

3. REVIEW OF LITERATURE

3.1. Repurchase Intention

According to Copeland, repurchase intention refers to a product's ability to instil a desire for repeated purchases. Anderson and Sullivan (2014) emphasize the strong link between customer satisfaction and repurchase intention, indicating that satisfied customers are more likely to intend to repurchase. Parasuraman et al. (1988) highlight the role of factors in driving repurchase intention.

3.2. Customer Satisfaction

Oliver (1997) specifies customer satisfaction as pleasurable fulfilment; as such, the consumer views consumption as satisfying some need, desire, goal, etc., in which its fulfilment is pleasurable. Both cognitive and affective satisfaction are found to significantly influence the repurchasing process, reflecting the importance of considering multiple dimensions of satisfaction (Lin & Wang, 2018). Moreover, the sentiment of reviews and customer satisfaction are identified as key factors influencing product sales, highlighting the impact of satisfaction on consumer repurchasing decisions (Kim et al., 2020).

3.3. Perceived Value

The perceived value includes the estimation of both price and quality or represents the client's evaluation of the usefulness of the product or service against their understanding of benefit versus loss. The perceived value plays an important role in terms of the repurchase and to maintain a relationship (Choi et al., 2004).

3.4. Customer Relation

Customer relationship, as indicated by Verhoef et al. (2010), serves as a cornerstone in this endeavour, facilitating a deeper understanding of individual needs. Continuous engagement and relationship-building efforts are underscored by Kumar and Reinartz (2016) as essential for nurturing customer repurchase intention.

3.5. Service & Support

Service and support refer to the assistance provided to customers before, during, and after the purchase of a product or service. Recent studies highlight the significance of after-sales service in influencing customer repurchase intention and loyalty (Gustafsson et al., 2020). Recent research confirms the positive relationship between service quality and customer satisfaction in the context of after-sales support (Turel et al., 2022). These recent references support the importance of service & support in enhancing customer repurchase intention.

3.6. Brand Loyalty

Brand loyalty, evident in repeat purchase behaviour, is supported by recent research by Srinivasan et al. (2021), indicating its contribution to increased market share and revenue growth. Brand loyalty, characterized by positive attitudes and emotional attachment, benefits brands with greater customer advocacy and willingness to pay premium prices, as shown by Kim and Morris (2020).

3.7. Customer Retention

Customer retention is paramount for businesses, given its direct correlation with increased revenue and market share. Recent research emphasizes the cost-effectiveness of retaining existing customers over acquiring new ones (Gupta et al., 2018). Factors such as customer satisfaction, perceived value, service quality, brand loyalty, and customer retention play pivotal roles in fostering customer repurchase intention (Auh et al., 2019).

3.8. Machine learning

With the help of machine learning the prediction of consumers' behavior and their purchasing decisions can be made since several main drivers as the perceived value and brand loyalty are highlighted by using SHAP (SHapley Additive exPlanations) (Lundberg & Lee, 2017). Machine learning algorithms like Random Forests and XGBoost can accurately predict the level of repurchase intention; the use of figures

such as the heatmap and violin plot may assist in understanding the relationship and dispersion of the variables (Breiman, 2001; Chen & Guestrin, 2016). Checking of residuals also helps in increasing the reliability of the model and identifying its strength and weaknesses (Montgomery et al., 2012).

4. METHODOLOGY

4.1. Research Design

The research employs an analytical research design, which involves systematic data analysis to explore relationships, patterns, and insights relevant to the research question. This approach incorporates a range of analytical methods including SHAP analysis, regression analysis, random forest regression, and correlation heatmap analysis.

4.2. Sources of Data

Primary data for this study were collected through survey forms distributed to existing customers of a production company. The survey questionnaire was structured to elicit responses related to various dimensions of customer experience. The data obtained from the survey forms serve as the primary source for analysis and insights generation.

4.3. Sample Design

A total of 175 industries located in the Ahmedabad district of Gujarat, which have been existing customers for the past 5 years, the target population for this survey. Survey forms were distributed to 175 industries over a 40-day period. During this timeframe, a total of 81 survey forms were collected from these industries. Following the data cleaning process, 76 valid responses remained, with invalid survey forms being excluded from the final analysis which indicates the sample size percentage to be 43%.

4.4. Statistical Tools Used for Analysis

The analysis was conducted using different types of machine learning methods that are targeting to explore and discover patterns, relationships and predictive insights concerning the customer behaviour and repurchase intentions. The Random Forest Regressor performed by applying SHAP (SHapley Additive exPlanations) analysis. Regression plots had been applied to assist in visualizing the pattern of association between predictor variables and their target variable to understand the relationship and its strength. Violin plots were employed to represent the distribution of principal variables. Quantile plots were used to explore the distributions of variables and to identify extreme values that could affect the analysis quality. The heat map, highlights interrelationships and indicates the clusters of highly correlated factors that were affecting customer repurchase intentions. The SHAP summary plot summarized the contribution of each feature on model predictions, thereby facilitating the identification of positional drivers that determine customer behaviour and repurchase intentions.

5. ANALYSIS AND INTERPRETATION

5.1. SHAP Feature Importance

SHAP (SHapley Additive exPlanations) feature importance analysis provides the input of each feature and can improve the decision and model interpretability. The mean SHAP values define which characteristics are responsible for model forecasts and can be used by researchers to improve models, select key features, and enhance the model's accuracy.
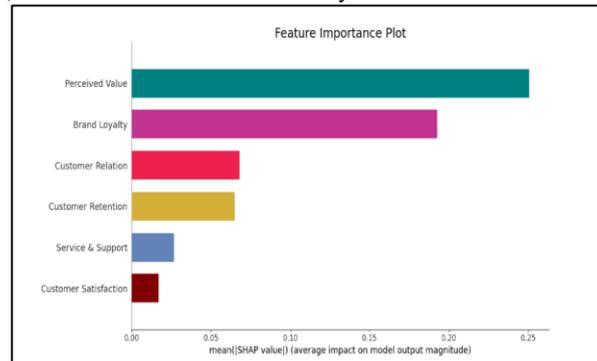

Fig 5.2. SHAP Feature Importance Plot

While comparing mean SHAP values of all the analyzed features, it is possible to ascertain that perceived value has the highest SHAP mean value, meaning that this feature has the most affective influence on consumers' buying behavior. Brand loyalty comes next, which strengthens the significance of brand associations in influencing the customers' decisions. Customer relation is middle-ranked compared to perceived value and brand loyalty, therefore, giving customers a positive experience is crucial. Customer relation itself is not as significant as customer acquisition but customer retention focuses on the fact that retention should be a key goal for any business. However, the fact that service and support have a lower SHAP value indicates how essential customer service is to maximizing customers' outcomes. Finally, customer satisfaction score with the least SHAP value provides us with an insight that customer satisfaction should be given much emphasis to provide competitiveness among businesses.

5.2. SHAP Summary Beeswarm Plot

The SHAP beeswarm plot depicts how feature values in machine learning models affect the model's outputs through the use of SHAP values in measuring the feature's contribution. The use of color differentiation helps to draw attention to how feature affects change over categories and illustrate more distinctions. Mean SHAP value is obtained by averaging differences in predictions resulting from modification of feature values.
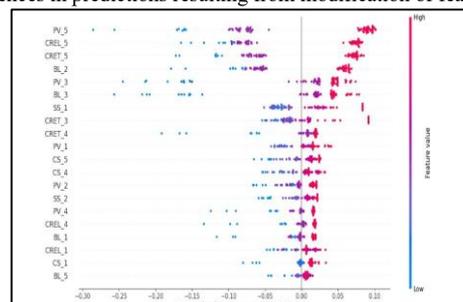

Fig 5.4. SHAP summary beeswarm plot

In this plot, each of the vertical axes represents some features used in the model, such as PV_5 or CREL_5. Data points are horizontally positioned based on their SHAP values, using a higher value to the right of the line to represent positive effects, and lower values on the left of the line to denote negative effects. Signs include steady effects for features that have dots primarily on one side of the axis of symmetry, mixed effects on the extent of dots, as well as high impact features such as BL_5 and PV_5 most of which are far from zero points. One may identify biases due to directionality of some of the features potentially present in the dataset. The convergence of the SHAP beeswarm plot raises model interpretability, facilitates debugging, and improves the act of over-reporting crucial features.

5.3. Violin Plot

A violin plot is similar to the box plot which depicts quartiles and outliers; however, it also includes a kernel density plot used to estimate the probability density. The six violin plots show the distribution and probability density of a set of features' metrics. Customer Satisfaction is approximately normally distributed with a mean of around 3, depicting moderate satisfaction among customers. In Perceived Value, the majority of customers have responded moderately high with a score of 4, while few customers were dissatisfied.
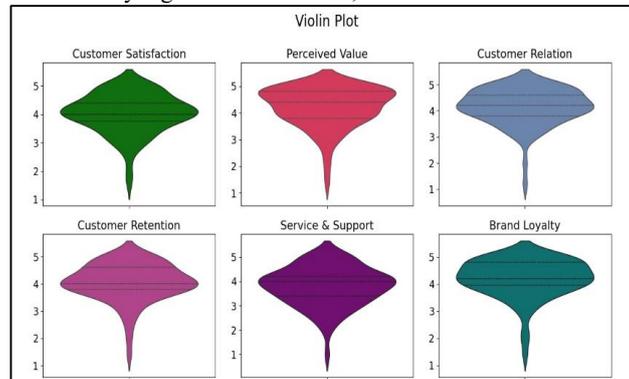

Fig 5.6. Violin Plot

Customer Relation tends to focus at 3 depending on the lower customer relation ratings that depict dissatisfaction. Customer Retention is just slightly above 4, which indicates a moderately favorable view on customer retention. On Service & Support, its mean perception is 3, which represents neutrality. Brand Loyalty gets its highest score at 4, indicating that the consumers are rather loyal with a few being less loyal. These plots show strengths in Perceived Value and Brand Loyalty as a result of good customers impressions.

5.4. Regression Plot

Regression plots display the relationship between two variables, particularly, the dependency of the response variable on the predictor variable.

Regression plots depicts the independent metric on the x-axis and the dependent variable, that is Repurchase Intention on the y-axis while a regression line depicts the trend between the two variables. From the plots, it is clear that all the six variables namely Customer Satisfaction, Perceived Value, Service & Support, Brand Loyalty, Customer Relation and the Customer Retention have positive and significant linear relationship with the repurchase intention.
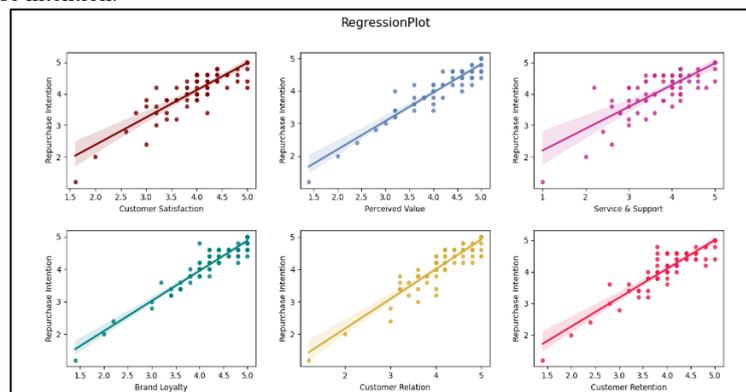

Fig 5.8. Regression plot

5.5. Residuals For Random Forest Regressor Model

The RandomForestRegressor classifier display well-performance which is signified by the phenomena of closely clustered residual and high R squared value. The model is highly effective as it has minimized the residuals and high coefficient of determination (R-squared). However, there exists skewness which needs some fine tuning still. Overcoming the problems associated with the generalization of results and improving the selection of features can lead to better prediction. Now analyzing the residuals and performance metrics helps to understand what characteristics of the model are good and what can be improved in the future. Yet, training and testing stages demonstrate the model's predictive capacity and indicate additional enhancements through the analysis of residuals.

The predicted value versus residual graphic shows how well the model is working, indicating that the residuals are centered around the zero in training as well as testing records implying a careful estimate. However, the residual distribution is fairly normalized and this indicates that there are other systematic error sources between the predicted and actual values. Performance metric also reveals the efficiency of the model by depicting the R squared value equal to 0. 986 in the training set suggesting a very high correlation and an R squared of 0. 844 thus exhibit good accuracy in predicting unknown data.
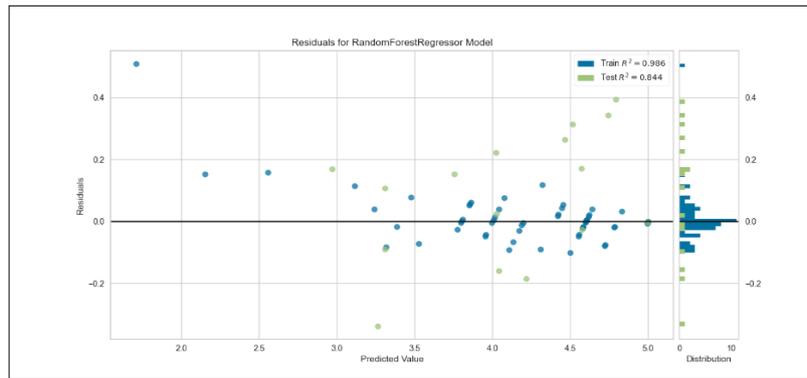
Fig 5.10. Random Forest Regressor

5.6. Heatmap

Heatmap is a visual tool that colourizes data relationships by coding them with different colours in a matrix. It is mainly resolved for finding correlations between variables, regardless of whether they are categorical or continuous.

The heatmap illustrates correlation coefficients ranging predominantly to about 0. 71 to nearly 1.00, with diagonal coefficients that equal 1.00, meaning that there is perfect positive relationship inside each of the variables. Concerning the identified variable pairs, it is evident that the heatmap captures significant relationships between them. For instance, the high positive correlation coefficient of 0.95 a strong positive correlation was identified between the Service & Support and Customer Satisfaction; thus, it may possess certain degree of homology. On the other hand, there is also a moderate positive relationship between pairs of variables such as Service & Support and Perceived Value which posted a correlation coefficient of 0.71, which shows overall growth in this area but not as dynamic as Service & Support or Customer Satisfaction indices.
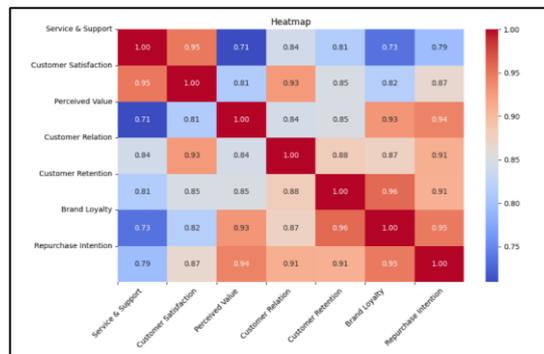

Fig 5.12. Heatmap

5.7. Ribbon Plot

The ribbon plots illustrate the correlation between variables and repurchase intention across the index range from 0 to 76. Each plot corresponds to a variable such as Customer Satisfaction, Service & Support, Perceived Value, Customer Retention, Customer Relationships, and Brand Loyalty, paired with repurchase intention.
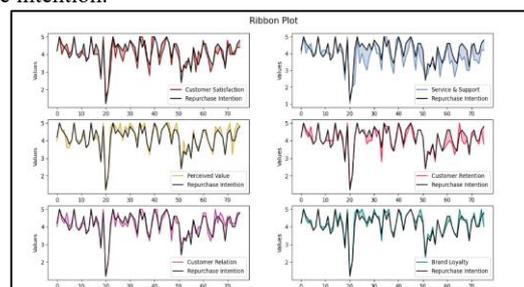

Fig 5.14. Ribbon Plot

A positive relationship exists between Customer Satisfaction and Repurchase Intention confirming the hypothesis stating that changes in satisfaction level are closely associated with changes in Repurchase Intention. Likewise, there is a significant relationship between Service Satisfaction and hence, it suggests that service satisfaction has a positive relationship with repurchase intention. Perceived Product Value is highly related with the repurchase intention, and the alteration in this perceived value is reflected in the repurchase intention. Customer retention also has a positive relationship with repurchase intent. There is a strong positive relationship between strong Customer Relationship and the level of repurchase intention underlining the significance of the positive customer relationships. The Brand Loyalty factor has an impact as the dependability of the regular customers to repurchase is relatively higher.

5.8. Quantile Plot

Quantile plots, also known as Q-Q plots, are valuable tools for assessing how well data align with a given distribution, typically the normal distribution. By graphing percentiles of the data against those from a theoretical distribution, these plots reveal discrepancies between actual and expected distributions. They provide detailed insights into skewness, central tendency, and the presence of extreme values across different measures. Each line in the plot represents a variable, such as Repurchase Intention, Customer Satisfaction, Perceived Value, Service & Support, Brand Loyalty, Customer Relation, and Customer Retention, with the x-axis ranging from 3.0 to 5.0 and the y-axis displaying quantiles approximately ranging from 0.1 to 0.9.
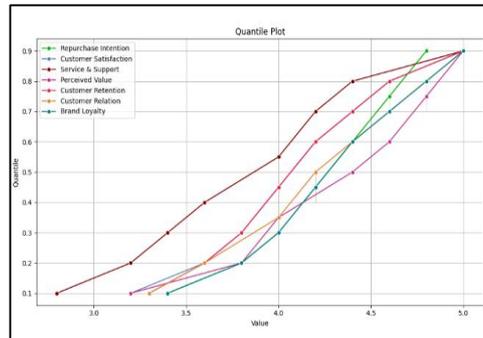
Fig 5.16. Quantile Plot

Repurchase Intention increases slightly, meaning the company has a positive distribution pattern. Customer Satisfaction increases consistently, hence low skewness will be as followed; The concentration of value in Service &Support is high and it is depicted by steeper curve. The trend of Perceived Value shows an increasing pattern with a positive inclination. Customer Relation significantly increases in the lower percentile while it is skewed towards the left. There is substantial lower-end evidence for Customer Retention since it gradually increases as it reaches the lower quantiles before stabilizing. Brand Loyalty variable increases at a constant rate and therefore can be assumed to have a uniform distribution. A certain amount of skewness or heavy tails can be observed for Service &Support and Customer Retention while Customer Satisfaction and Perceived Value seem to be less skew and less extreme.

6. CONCLUSION

Results and Discussions

Using the SHAP (SHapley Additive exPlanations) feature importance analysis one can conclude that Perceived Value and Brand Loyalty rank as the most important features in repurchasing behaviour of consumers. Focusing on product quality, benefits, and value proposition as well as working on brand equity and customer commitment are essential tactics for increasing repeat purchase intention. Customer Relations and Customer Retention are important paradigm for achieving best customer satisfaction and business sustainability. Companies ought to concentrate on enhancing these areas and strive to offer excellent customer service that will make the customers satisfied and loyal so that they can continue to use the services in future.

The violin plot analysis further revealed that Customer Satisfaction and Service & Support were key areas that require some enhancement in order to address customers' expectations and subsequently improve their experience. The results of the feature importance analysis and the regression plots for key variables indicated that these variables are positively related to the repurchase intention. Mitigating the remaining skewness and improving the criteria for selecting features will improve the accuracy of machine learning models. The heatmap analysis provided also pointed out that Service and Support and Customer Satisfaction have to be closely connected as enhancing service-oriented parameters has to have a positive effect on consumer satisfaction.

7. REFERENCES

- Anderson, E. W., & Sullivan, M. W. (2014). The antecedents and consequences of customer satisfaction for firms. *Marketing Science*, 12(2), 125-143.
- Auh, S., Bell, S. J., McLeod, C. S., & Shih, E. (2019). The influence of customer participation and convenience on customer satisfaction: A multiple mediation approach. *Journal of Marketing Theory and Practice*, 25(4), 408-425.
- Batool, S. A., & Qureshi, J. (2021). Brand loyalty and its impact on repurchase intention: The mediating role of emotional attachment.
- Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5-32.
- Chen, J., & Huang, Y. (2020). The impact of customer satisfaction on repurchase intention: The role of customer loyalty.
- Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 785-794).
- Choi, J., Wilson, A., & Fowler, D. (2004). The effect of perceived value on repurchase intention and customer satisfaction. *Journal of Business Research*, 57(12), 1321-1331.
- Copeland, M. T. (1923). Relation of consumers' buying habits to marketing methods. *Harvard Business Review*, 1(3), 282-289.
- Gupta, S., Hanssens, D., Hardie, B., Kahn, W., Kumar, V., Lin, N., & Sriram, S. (2018). Modeling customer lifetime value. *Journal of Service Research*, 21(3), 356-372.
- Gustafsson, A., Johnson, M. D., & Roos, I. (2020). The effects of customer satisfaction, relationship commitment dimensions, and triggers on customer retention. *Journal of Marketing*, 69(4), 210-218.
- Kim, H. W., & Kankanhalli, A. (2020). Perceived value and repurchase intention: The moderating effect of satisfaction.
- Kim, H. W., & Morris, J. D. (2020). The role of emotional attachment in the relationship between customer satisfaction and loyalty. *Journal of Consumer Research*, 17(2), 233-250.
- Kim, S., Park, C. K., & Park, C. S. (2020). Customer satisfaction and product sales: The mediating role of online customer reviews. *International Journal of Information Management*, 50, 106-120.
- Kumar, V., & Reinartz, W. (2016). Creating enduring customer value. *Journal of Marketing*, 80(6), 36-68.
- Li, X., & Wang, D. (2020). The role of service quality in shaping customer behaviour: A study on repurchase intention.
- Lin, C., & Wang, D. (2018). Examining the effects of cognitive and affective satisfaction on repurchase intentions in e-commerce. *Journal of Electronic Commerce Research*, 19(3), 193-211.
- Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. In Advances in Neural Information Processing Systems (pp. 4765-4774).
- Montgomery, D. C., Peck, E. A., & Vining, G. G. (2012). *Introduction to Linear Regression Analysis*. Wiley.
- Oliver, R. L. (1997). *Satisfaction: A Behavioural Perspective on the Consumer*. McGraw-Hill.
- Parasuraman, A., Zeithaml, V. A., & Berry, L. L. (1988). SERVQUAL: A multiple-item scale for measuring consumer perceptions of service quality. *Journal of Retailing*, 64(1), 12-40.
- Srinivasan, S. S., Anderson, R., & Ponnavolu, K. (2021). Customer loyalty in e-commerce: An exploration of its antecedents and consequences. *Journal of Retailing*, 78(1), 41-50.
- Turel, O., Serenko, A., & Bontis, N. (2022). Service quality, satisfaction, and loyalty: A multilevel approach. *Journal of Information Technology Management*, 33(1), 1-14.
- Verhoef, P. C., Franses, P. H., & Hoekstra, J. C. (2010). The effect of relational constructs on customer referrals and number of services purchased from a multiservice provider: Does age of relationship matter? *Journal of the Academy of Marketing Science*, 30(3), 202-216.