THE IMPACT OF BRANDING ON INDIA'S E-COMMERCE

Marati Vijaykumar[1], Kimmoji Raja[1], Thakur Pradip Rajendra[1] and B. Rushi Kumar1*

[1]*Department of Mathematics, School of Advanced Sciences, Vellore Institute of Technology,*

Vellore-632014, Tamil Nadu, India.

Email: rushikumar@vit.ac.in ; marati.vijaykumar2024@vitstudent.ac.in ;

kimmoji.raja2024@vitstudent.ac.in ; thakur.pradip2024@vitstudent.ac.in

Abstract:

India's e-commerce market has experienced rapid growth, driven by increasing internet penetration and mobile phone usage. This study examines key areas of expansion by analysing customer buying behaviour and evaluating the role of branding in influencing purchasing decisions. It also proposes evidence-based strategies for sustainable growth. The research identifies major drives of e-commerce expansion in India while addressing challenges such as digital illiteracy, trust deficits, and regulatory constraints. Additionally, it explores how branding impacts consumer trust, loyalty, and purchasing behaviour. A primary dataset was collected through online surveys across diverse regions of India, capturing insights on shopping habits, brand preferences, trust factors, and customer satisfaction. The study employs data analytics techniques to evaluate market trends. Statistical analysis, machine learning models, and visualization tools like Seaborn and Matplotlib were used to examine consumer behaviour. Feature engineering methods-including label encoding, MinMax scaling, and classification models were applied to measure branding's influence. Findings indicate that strong branding enhances customer trust, drives revenue growth, and expands market reach. Personalized experiences, optimized services, and enhanced security measures help address industry challenges. Leveraging data-driven strategies improves marketing effectiveness and strengthens customer loyalty. Overcoming regulatory and digital literacy barriers is crucial for ensuring the long-term viability of India's E-commerce sector.

Keywords: E-commerce Growth, Branding, Classification Techniques, Strategies, Statistical Analysis.

Introduction:

E-commerce has fundamentally transformed the global retail landscape, offering unparalleled convenience, variety, and accessibility to consumers. In India, the e-commerce sector has grown exponentially due to increasing internet penetration, affordable smartphones, and widespread adoption of digital payment systems. As of 2025, over 50% of India's population has access to the internet, driving the industry's valuation from $50 billion in 2020 to a projected $200 billion by 2026. Major players like Flipkart, Amazon India, and Myntra dominate the market, while niche platforms cater to specific categories such as fashion and groceries. However, challenges such as logistical inefficiencies, data privacy concerns, and regulatory constraints remain significant hurdles to sustainable growth.

Branding is a critical factor influencing consumer trust and loyalty in e-commerce. Patel (2021) highlighted that brand loyalty is closely tied to personalized shopping experiences, secure transactions, and efficient customer service. Similarly, Singh and Mehta (2022) emphasized that digital marketing strategies, including targeted advertisements and influencer promotions on platforms like Instagram and Facebook, play a pivotal role in shaping consumer behaviour. Verma (2023) further noted that online reviews significantly impact consumer trust positive reviews enhance brand credibility, while negative feedback can deter potential buyers.

The integration of big data analytics and artificial intelligence (AI) is revolutionizing e-commerce operations. These technologies enable businesses to optimize supply chains, personalize customer interactions, and detect fraudulent activities. Nalla and Reddy (2024) explained how AI-driven analytics improve marketing strategies and customer satisfaction but also raised concerns about data privacy and scalability. Kauffmann (2019) proposed a framework for leveraging big data to analyse customer sentiment and detect fake reviews, underscoring the importance of reliable feedback mechanisms for decision-making.

Despite extensive research on branding strategies, consumer trust factors, and technological advancements in e-commerce, there is limited focus on predictive modeling for customer satisfaction. Existing studies often rely on qualitative surveys or descriptive analyses without leveraging advanced machine learning techniques. This project addresses this gap by Machine Learning models such as XGBoost and Random Forest to predict customer satisfaction based on factors like brand discovery, trust levels, and shopping frequency. The methodology includes data preprocessing, feature engineering, hyperparameter tuning, and model evaluation to provide actionable insights into consumer behaviour.

Literature Review :

India's e-commerce landscape has made a dramatic leap in the last decade, where branding has taken center stage among the drivers transforming consumer behavior, loyalty, and purchase intent. With digital media continuing to disrupt retail experiences, researchers and market experts have gone deep into probing the drivers that influence consumer purchases in this new landscape. This review integrates major results from a variety of studies examining how branding policies, mechanisms of trust, social media power, and technological innovations are influencing the direction of India's e-commerce sector.

The extensive usage of smartphones and internet facilities has popularized online shopping and made it convenient in urban as well as semi-urban India. As per Goswami (2013), branding is crucial in online fashion retail as it builds memorable digital experiences and emotional relationships with customers. The study highlighted how online branding tactics—such as consistent visual identity, tone, and personalized content—can have a considerable influence on repeat purchase and brand loyalty. Confirming this, Sharma (2020) noted a dramatic rise in digital payments after COVID, as customers grew increasingly aware of safety, trust, and delivery guarantee—issues closely related to sound branding and reputation.

The influence of social media on contemporary branding cannot be overemphasized. Banerji and Singh (2023) concluded that social media marketing activities (SMMAs), including influencer partnerships and targeted advertisements on social media channels like Instagram and Facebook, positively boost customer loyalty. Their research emphasizes that ongoing interaction with consumers through social media instills confidence and creates a community of belief in a brand. Likewise, Verma (2023) commented on the imperative role played by online reviews and public opinion in establishing brand credibility. Positive reviews enhance consumer trust, whereas negative reviews can instantly destroy confidence, particularly in competitive markets.

The importance of digital empowerment and inclusion has also been researched. Bhatt (2024) analyzed the part played by e-commerce in the economic empowerment of women and established that sites that provide easy-to-use branding and selling facilities empower women entrepreneurs to establish customer bases. Anooja (2015) and Gupta & Sharma (2016) also looked into the transformation of agriculture using e-commerce sites. Their research indicated how branding—particularly with transparency and equitable pricing—fills the gap between rural producers and city consumers, yet infrastructural deficits in far-flung regions remain a hindrance.

Multiple studies have focused on branding and trust from the legal and regulatory perspectives. Goswami (2023) analyzed India's changing legal environment for digital business and data protection, positing that strong data protection policies improve consumer confidence in branded platforms. Similarly, Kashyap and Chaudhary (2023) also pointed out the increasing demand for consumer protection laws that safeguard customers against web fraud, also adding that customers prefer to believe in reputed brands with better measures of data safety. Rao (2023) carried this further by pointing to shortcomings in the enforcement of policies, indicating that although policies are available, uneven policy enforcement can influence consumer trust, particularly when buying from lesser-known or new e-commerce brands.

Consumer attitude and brand loyalty have also been dramatically transformed in the post-pandemic period. Kumari and Kumar (2021) wrote about the mindset of Indian consumers changing to lean towards digital channels from the conventional kirana outlets. According to their research, preferred brands with good payment systems, clear pricing, and prompt customer support tend to attract more consumer loyalty. Salunkhe (2023) noted how, today, consumers connect brand identity with logos and slogans but also the reliability and transparency of the overall shopping experience right from the stage of browsing to post-purchase assistance.

Within the scenario of big e-commerce platforms, Verma and Singh (2024) examined Amazon India's contribution in revolutionizing local retail. Their research showed that although Amazon's branding and infrastructure enable small firms to reach further into the marketplace, they also heighten competition against conventional brick-and-mortar outlets. Sharma (2024) investigated the cross-border e-commerce trust dynamics and concluded that transparently communicated brand values, local customer care, and payment guarantees play a critical role in informing purchasing decisions from foreign vendors.

Brand narrative, emotional affinity, and individualized engagement remain key drivers of consumer behavior. Singh and Mehta (2022) examined whether digital marketing campaigns, particularly those that are built around personalization and narrative, have an impact on repeat purchasing behavior. Their study found that customers feel emotionally closer to companies that share their values and lifestyle. The same notion is reflected by Verma (2023), who demonstrated how influencer marketing and feedback loops in real time drive trust, particularly in younger generations.

Technological innovation also plays a pivotal part in the branding ecosystem. Nalla and Reddy (2024) brought to the forefront how AI and big data analytics improve branding activities by making it possible for personalized recommendations, tracking behavior, and sentiment analysis. These technologies enable brands to anticipate customer needs and adapt their services accordingly. Kauffmann et al. (2019) took this a step further by presenting a comprehensive framework that uses AI to detect fake reviews and assess customer sentiment. Their work demonstrates that trustworthy feedback mechanisms are not just tools for internal decision-making but vital components of a brand's public image.

In spite of this vast body of literature, the majority of studies are still based on descriptive or qualitative analysis, frequently missing predictive potential that can provide deeper understanding of consumer satisfaction and forecasting of behavior. Although the effect of branding has been investigated from a wide variety of perspectives—social media, legal trust, personalization, and technology—few have utilized sophisticated data science methods to forecast how consumers will react to particular branding strategies.

But there is limited research on predictive modeling and data-driven decision-making within customer satisfaction analysis. Our project bridges this research gap through the application of sophisticated machine learning methodologies in forecasting customer satisfaction from variables such as brand discovery, trust levels, and shopping behaviours. We employ classification models such as XGBoost and Random Forest to examine e-commerce dynamics that go beyond survey-based studies. Our methodology entails data preprocessing, feature engineering, hyperparameter optimization, and rigorous model assessment, offering a systematic, data-informed picture of consumer satisfaction that has not been extensively investigated in the literature.

Methodology:

1. Data Collection and Source

The data for this study was collected through an online survey distributed via Google Forms, targeting Indian e-commerce users. The survey aimed to gather comprehensive insights into consumer behaviour, brand preferences, trust factors, and satisfaction levels with e-commerce platforms like Amazon, Flipkart, Myntra, and Ajio.

Survey Design

The questionnaire was structured to ensure clarity and relevance:

- Demographic Questions: MailID,Age group, gender, occupation, and location.
- Behavioural Questions: Shopping frequency (e.g., Daily, weekly, monthly, Occasionally, Rarely), preferred platforms, and reasons for platform loyalty or switching.
- Trust and Satisfaction Questions: Trust factors (e.g., customer reviews, social media presence) and satisfaction ratings (on a 1–5 scale).
- Open-Ended Questions: Reasons for dissatisfaction or preferences for specific platforms.

The dataset consisted of 608 responses after cleaning and preprocessing. The collected data was stored in a CSV file named E_commerce dataset1.csv and imported into Python for analysis using the Pandas library.

Data Preprocessing

Data preprocessing was performed to ensure consistency and reliability:

- **Cleaning:** Irrelevant columns (e.g., timestamps, E-mail) were removed to reduce noise. Duplicate entries were eliminated using drop duplicates().
- **Encoding:** Categorical variables such as brand preferences and trust factors were encoded using Label Encoding.
- **Target Variable Creation:** Satisfaction scores for each platform (e.g., Amazon, Flipkart, Myntra and Ajio) were binarized into two classes:
  - Satisfied (scores ≥ 4)
  - Not Satisfied (scores < 4)

Model Selection and Training

Train-Test Split

The dataset was partitioned using scikit-learn's train_test_split() function, allocating 80% for training (to fit models) and 20% for testing (to evaluate generalization). A random seed ensured reproducibility, balancing the split across the binarized target classes.

Machine Learning Models

A diverse array of classification models was implemented to predict customer satisfaction, leveraging scikit-learn, CatBoost, XGBoost, and Gradient Boosting libraries. Each model's theoretical basis and application are detailed below:

1. Logistic Regression

Logistic Regression is a statistical model used for binary classification tasks, where the outcome is either 0 or 1 (e.g., customer satisfied or not). It is a linear model that estimates the probability of a class based on input features. Despite its name, logistic regression is actually a classification algorithm, not a regression algorithm. The model applies a sigmoid function to map predicted values between 0 and 1. It is simple, easy to implement, and interpretable but may struggle with complex, non-linear relationships in the data.

- Logistic Regression models the probability of satisfaction using the sigmoid function:
- It is simple yet effective for binary classification tasks but assumes linearity in the log-odds.

$$P(y) = \frac{1}{1+e-(\beta 0+\beta 1x1+\cdots+\beta nxn)}$$

where:

- p is the predicted probability of the positive class (y=1)
- β_0 is the intercept (bias).
- $\beta_1, \beta_2, ..., \beta_n$ are the coefficients (weights) of the model.
- $x_1, x_2, ..., x_n$ are the feature variables.

2. Decision Tree Classifier

A Decision Tree Classifier is a tree-based machine learning model that recursively splits the data into subsets based on feature values. The tree consists of nodes (decisions), branches (possible outcomes), and leaves (final predictions). The model selects the best feature to split on at each step, using criteria like Gini impurity or entropy. While decision trees are intuitive and handle non-linearity well, they are prone to overfitting, especially when deep trees are formed.

- Decision Trees split data based on feature thresholds to minimize Gini impurity:

$$\hat{y}(x)=\sum_{j=1}^{J} w_j(x \in R_j)$$

Where:

- y (x) predicted output for input
- J: number of terminal (leaf) nodes
- R_j: region (data partition) defined by decision rules in the tree
- w_j: predicted value (e.g., class label or mean of target values) for region R_j

- They capture non-linear relationships but are prone to overfitting without constraints like maximum depth.

3. Bagging Classifier

Bagging (Bootstrap Aggregating) is an ensemble learning technique that improves model accuracy by training multiple classifiers on different random subsets of the data and then aggregating their predictions. The Bagging Classifier typically uses Decision Trees as base models and reduces variance by averaging or majority voting. This helps prevent overfitting while maintaining model flexibility.

- Bagging trains multiple models on bootstrapped datasets and combines their outputs via majority voting:

$$\hat{y} = \text{majority vote}(h_1(x), h_2(x) \ldots\ldots, h_T(x))$$

Where:

  - T: Total number of trees.
  - $h_t(x)$: Prediction of the t^{th} decision tree.
  - x: Input feature vector.
  - $\hat{y}$: Final prediction.

- This approach improves stability by reducing overfitting.

4. Random Forest Classifier

Random Forest is an ensemble learning method that builds multiple decision trees and combines their predictions for better accuracy and robustness. Unlike a simple decision tree, Random Forest introduces randomness by selecting a subset of features for each split, reducing overfitting and improving generalization. It performs well on various classification tasks but can be computationally expensive with large datasets.

- Random Forest builds multiple decision trees on random subsets of data and aggregates their predictions:

$$\hat{y} = \text{majority vote}(h_1(x), h_2(x) \ldots\ldots, h_T(x))$$

Where:

  - T: Total number of trees.
  - $h_t(x)$: Prediction of the t^{th} decision tree.
  - x: Input feature vector.
  - $\hat{y}$: Final prediction.

- It reduces variance and highlights feature importance.

5. AdaBoost Classifier

Adaptive Boosting (AdaBoost) is a boosting algorithm that improves weak learners by focusing more on misclassified samples. It trains multiple models sequentially, where each new model assigns higher weights to previously misclassified instances. AdaBoost is commonly used with shallow decision trees (stumps) and enhances model accuracy by iteratively refining predictions. However, it is sensitive to noisy data and outliers.

- AdaBoost combines multiple weak classifiers to create a strong classifier:

$$F(x)=\text{sign}\left(\sum_{m=1}^{M} \alpha_m f_m(x)\right)$$

  Where:

  - M: Number of weak learners.
  - $f_m(x)$ Prediction of the m^{th} weak learner (usually outputs ±1).
  - α_m: Weight assigned to the m^{th} learner, based on its accuracy.
  - F(x): Final prediction score (usually sign of this is taken for classification).
  - x: Input feature vector.

- It assigns higher weights to misclassified samples and adjusts weak learners accordingly.

6. Gradient Boosting Classifier

Gradient Boosting is a powerful ensemble technique that builds models sequentially, where each new model corrects the mistakes of the previous one. Instead of simple majority voting like Random Forest, Gradient Boosting combines weak models to minimize prediction errors using a loss function. It performs well on structured data but is computationally expensive and may overfit if not tuned properly.

- Gradient Boosting minimizes a loss function iteratively by adding weak learners:

$$F_M(x)=F_0(x)+\sum_{m=1}^{M} \gamma_m f_m(x)$$

Where:

- $F_M(x)$: Final prediction after M boosting rounds.
- $F_0(x)$: Initial prediction (e.g., mean for regression, log-odds for classification).
- $f_m(x)$: Model (typically a decision tree) trained to predict the residuals at step m.

- γ_m: Learning rate or step size used to scale the contribution of $f_m(x)$.
- M: Total number of iterations (trees).
- x: Input feature vector.

- It excels in structured data but requires careful tuning.

7. XGBoost Classifier

XGBoost (Extreme Gradient Boosting) is an optimized version of gradient boosting that is designed for high speed and performance. It incorporates techniques like parallel processing, regularization (L1 and L2), and efficient memory usage, making it a preferred choice for large datasets. XGBoost is widely used in machine learning competitions and applications requiring fast and accurate predictions.

- XGBoost enhances Gradient Boosting with regularization terms:

$$F_M(x)=\sum_{m=1}^{M} f_m(x), \quad (f_m \in F)$$

Where:

- $F_M(x)$: Final prediction after M boosting rounds.
- $F_m(x)$: Prediction from the m^{th} tree for input x_i.
- x_i: Input feature vector of the I^{th} sample.
- It is efficient for large datasets but sensitive to hyperparameter choices.

8. CatBoost Classifier

CatBoost (Categorical Boosting) is a gradient boosting algorithm optimized for categorical data. Unlike other boosting models that require extensive preprocessing (such as one-hot encoding), CatBoost efficiently handles categorical variables directly, making it useful for structured data problems. It reduces overfitting with built-in regularization techniques and is known for its high accuracy and fast training speed.

- CatBoost specializes in handling categorical data efficiently using ordered boosting:

$$F_M(x)=F_0(x)+\sum_{m=1}^{M} \gamma_m f_m$$

Where:

- $F_M(x)$: Final model after M iterations.
- $F_0(x)$: Initial prediction (could be mean log-odds, mean target, etc.).
- $f_m(x)$: Decision tree trained at iteration m to reduce the loss.
- M: Total number of boosting iterations.
- γ_m: Learning rate at iteration m.

- x: Input features (numerical or categorical).

- It performs well with minimal preprocessing.

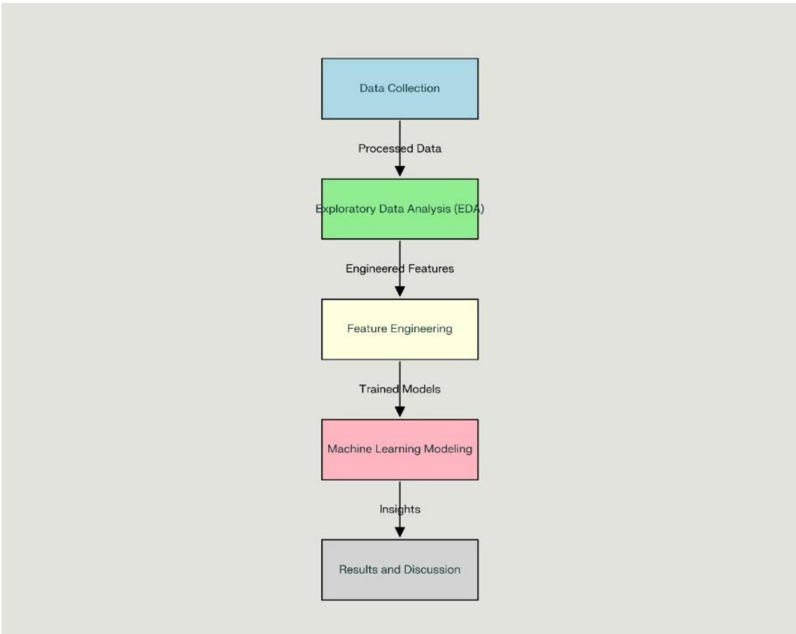
Hyperparameter Tuning

To optimize performance, hyperparameter tuning was conducted. For SVM (though not a primary focus), GridSearchCV tested kernels (linear, RBF) and regularization C (0.1–10). For Random Forest, RandomizedSearchCV explored n_estimators (50–200) and max_depth (10–30); for XGBoost, it tuned learning_rate (0.01–0.3) and max_depth (3–10). This iterative process refined model configurations, boosting accuracies.

Model Evaluation

Models were evaluated using the Accuracy Score (percentage of correct predictions) and Classification Report (precision, recall, F1-score). Precision measured positive prediction accuracy, recall captured true positive detection, and F1-score balanced both, critical for imbalanced classes. These metrics ensured a thorough assessment beyond mere accuracy.

Flowchart:



Results and Discussion:

Exploratory Data Analysis (EDA)

The exploratory data analysis (EDA) provided valuable insights into the dataset, revealing patterns in consumer behaviour, trust factors, and satisfaction levels across different e-commerce platforms.

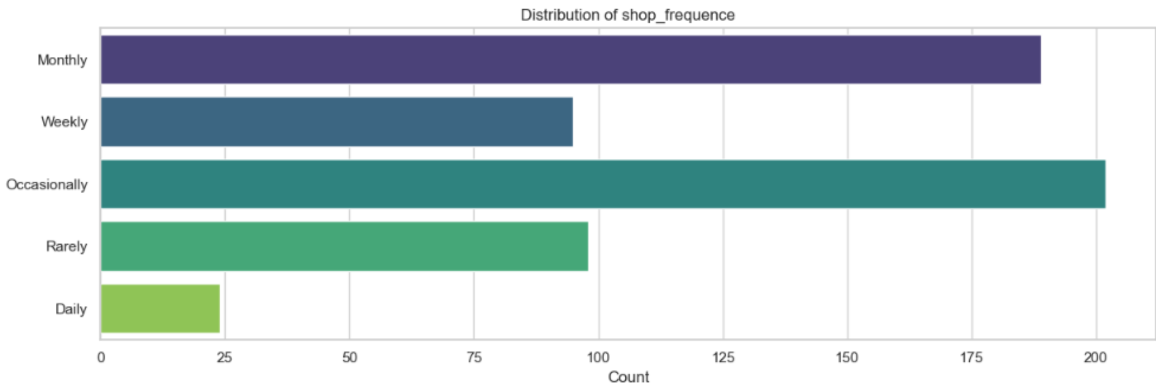
Visualization

Demographic Insights

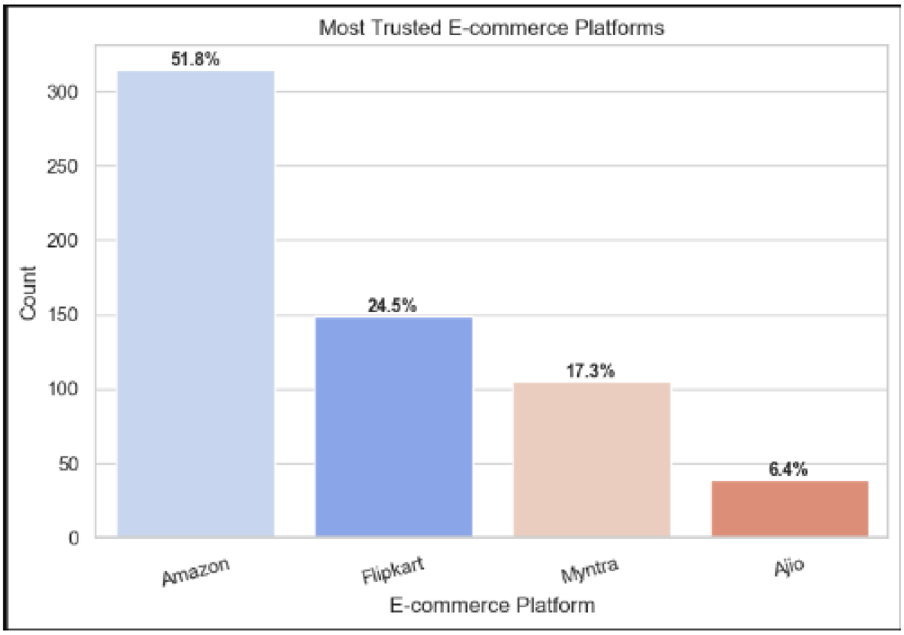
The dataset revealed that a majority of respondents were aged 18–24 (74.7%), aged 25–34 (19.7%), aged 25–44 (5.1%), aged 45+ (0.5%) reflecting the dominance of younger, tech-savvy consumers in India's e-commerce market. Gender distribution was nearly balanced, with 55.9% male and 44.1% female respondents. Occupation analysis showed that students comprised the largest group (69.9%), followed by salaried professionals (22.7%), indicating that e-commerce platforms are particularly popular among younger and working populations.

Shopping Behaviour

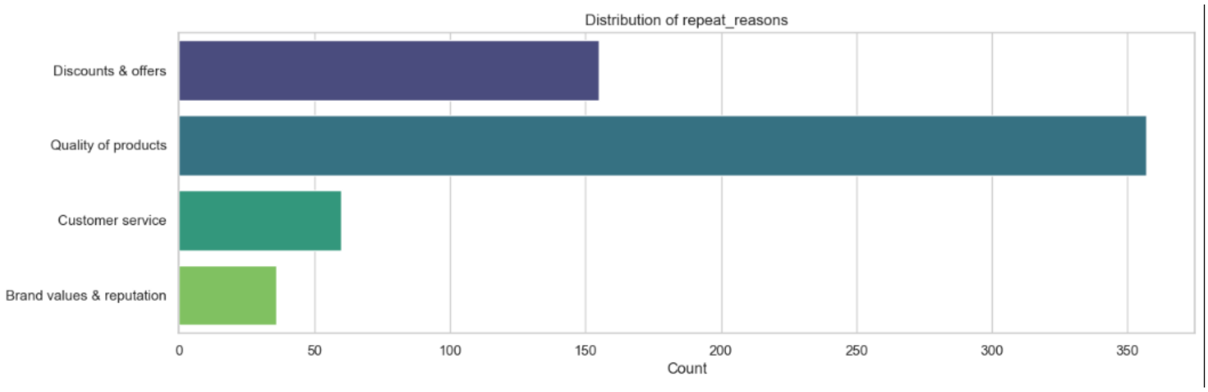
- **Shopping Frequency**: Around 33.2% of respondents reported shopping occasionally and 31.3% of respondents reported shopping monthly and 16% of respondents reported shopping rarely and 15.5% shopped weekly while 3.9% of respondents reported shopping daily. This indicates moderate engagement with e-commerce platforms.



- **Preferred Platforms**: Amazon emerged as the most popular platform, trusted by 315 respondents 51.8%, followed by Flipkart trusted by 149 respondents 24.5% and Myntra trusted by 105 respondents 17.3%. Ajio trusted by 39 respondents lagged behind with only 6.4% of respondents citing it as their preferred platform.

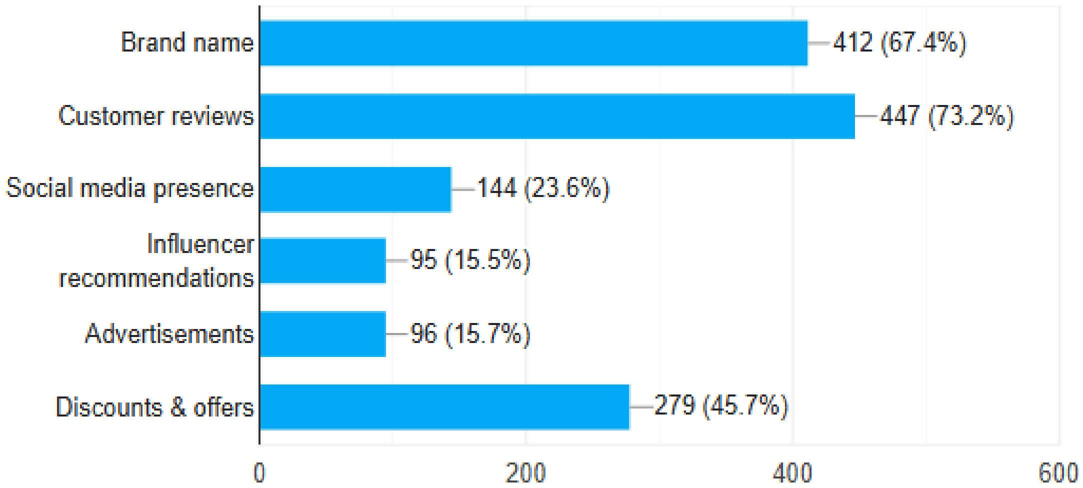


- **Brand Loyalty**: 58.8% of respondents reported brand loyalty based on quality of products and 25.5% of respondents reported brand loyalty based on Discounts & offers driven by factors such as trust in customer service 9.8% and Brand values & reputation 5.9%.

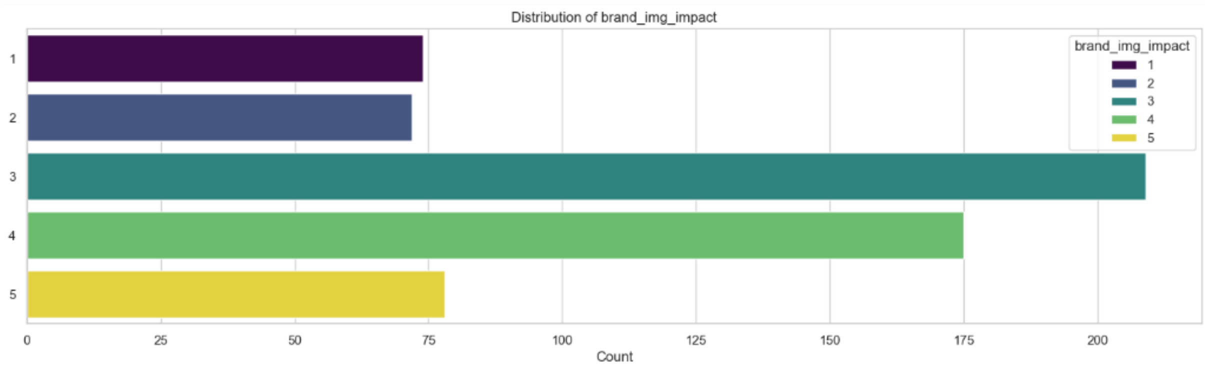


Trust and Satisfaction Factors

- **Trust Factors**: Customer reviews were the most significant trust factor for 73.2% of respondents, and followed by brand name 67.4% respondents, discounts & offers presence 45.7%, This response is based on the multiple options from the respondents out of (at least 3 from 6).



- **Brand image influence:** Scale 1: 75 respondents 12.3% indicated that brand image has minimal impact on their purchase decisions. Scale 2: 73 respondents 11.9% reported a slightly higher but still low influence of brand image. Scale 3: The largest group, 209 respondents 34.2%, rated the influence of brand image as moderate. Scale 4: 176 respondents 28.8% stated that brand image significantly influences their purchase decisions. Scale 5: 78 respondents 12.8% reported that brand image strongly impacts their purchase decisions.



Correlation Analysis

Correlation analysis revealed a moderate positive correlation 0.55 between brand image and recommendation scores, suggesting that loyal customers are more likely to recommend their preferred platforms to others. Heatmaps showed minimal multicollinearity among features, ensuring that predictors were independent and suitable for machine learning modelling.

Machine Learning Model Performance

To predict customer satisfaction across platforms, various machine learning models were applied to the dataset after preprocessing and feature engineering.

Model Accuracy Across Platforms

Model	**Amazon Accuracy (%)**	**Flipkart Accuracy (%)**	**Myntra Accuracy (%)**	**Ajio Accuracy (%)**
Logistic Regression	66.39	61.47	56.56	60.66
Decision Tree	63.93	58.20	66.39	54.92
Bagging Classifier	60.65	62.29	70.49	61.47
Random Forest	66.39	68.85	72.13	65.57

Model	Amazon Accuracy (%)	Flipkart Accuracy (%)	Myntra Accuracy (%)	Ajio Accuracy (%)
ADA Boosting	66.39	60.65	68.85	59.01
Gradient Boosting	65.57	65.57	69.67	55.74
XGBoost	67.21	63.93	65.57	54.10
CAT Boost	64.75	66.39	68.85	63.11

The performance of each machine learning model was evaluated using accuracy as the primary metric:

Key Observations

1. The **Random Forest Classifier** achieved the highest accuracy across platforms, with a peak accuracy of 72.13%, demonstrating its robustness in handling non-linear relationships and identifying important features.
2. Ensemble models like Bagging Classifier 70.49% and Gradient Boosting 69.67% also performed well, underscoring the effectiveness of ensemble techniques in improving predictive accuracy.
3. Logistic Regression 66.39%, while interpretable, struggled with non-linear relationships, achieving lower accuracy compared to tree-based models.

Platform-Specific Insights

Amazon

Amazon emerged as the most trusted platform, with the highest proportion of "Very Satisfied" customers. Its success can be attributed to reliable delivery services, extensive product variety, and strong customer reviews.

Flipkart

Flipkart performed well in terms of discounts and offers but lagged slightly behind Amazon in overall satisfaction due to occasional issues with delivery times.

Myntra

Myntra received mixed reviews while it excelled in fashion-specific offerings, dissatisfaction stemmed from limited product variety in other categories.

Ajio

Ajio had the lowest trust and satisfaction scores among the four platforms due to logistical inefficiencies and limited reach in tier2-tier3 cities.

Discussion

Model Performance Implications

The results demonstrate that ensemble models like Random Forest, CAT Boosting and Gradient Boosting are well-suited for predicting customer satisfaction due to their ability to handle complex relationships between features.

Business Insights

1. Strengthening branding strategies through targeted marketing on social media can significantly enhance customer satisfaction.
2. Trust-building measures such as secure transactions, transparent reviews, and reliable customer service are critical for retaining customers.
3. Platforms like Myntra and Ajio need to focus on expanding product variety or optimizing delivery services to improve satisfaction scores.

Limitations

1. The dataset was urban-centric; rural consumer behaviour may differ significantly.
2. Features like delivery time or return policies were not included but could provide additional predictive power.

This study explores the interplay between branding and customer satisfaction in India's e-commerce market using 608 responses that which we took from our dataset and for this eight machine learning models to uncover actionable insights for businesses. The exploratory analysis highlighted Amazon's dominance driven by reliable delivery and strong customer reviews, while Myntra faced challenges in product diversification, and Ajio struggled with logistics in tier-2/3 cities. Machine learning models were applied to predict satisfaction scores, with Random Forest achieving the highest accuracy (72.13%), followed by CatBoost and XGBoost, demonstrating the effectiveness of ensemble techniques. Feature importance analysis identified brand preference, social media discovery, and trust factors as key drivers of satisfaction. The binarized satisfaction threshold (≥4 as "satisfied") revealed platform-specific weaknesses, such as Flipkart's delivery inconsistencies despite discount effectiveness. Results underscored the pivotal role of social media (68% of satisfied customers discovered brands via Instagram/Facebook) and AI-driven personalization (boosting satisfaction through tailored recommendations). By validating ensemble models' superiority over traditional approaches, this study bridges the gap between predictive analytics and practical applications, offering a blueprint for scalable growth in India's competitive e-commerce landscape.

Original Dataset:

https://drive.google.com/file/d/1kDurJzL8mRbxxcJTrH0KitscGE32lSD-/view?usp=drivesdk

Conclusion:

This study comprehensively analyzed the impact of branding on customer satisfaction in India's e-commerce sector, leveraging advanced machine learning techniques and data-driven strategies. Key findings include:

- Demographic Insights: Younger consumers (18–24 years) dominate the e-commerce market, with Amazon emerging as the most trusted platform due to reliable delivery, product variety, and strong customer reviews.
- Trust Factors: Secure transactions, customer reviews, and social media presence significantly influence consumer satisfaction and brand loyalty.
- Machine Learning Performance: Random Forest achieved the highest accuracy (72.13%) in predicting customer satisfaction, followed by CatBoost and Gradient Boosting classifiers, highlighting the effectiveness of ensemble models.
- Platform-Specific Challenges: Myntra faced issues with product diversification, while Ajio struggled with logistical inefficiencies in tier-2/3 cities.
- ActInsights: Social media discovery (68% of satisfied customers via Instagram/Facebook) and AI-driven personalization were identified as key drivers for enhancing satisfaction.

References

1. **Goswami, S.** (2013). *A study on the online branding strategies of Indian fashion retail stores*. Paper presented at the International Conference on Internet Marketing Strategies.
2. **Gupta, R., & Sharma, P. K.** (2016). *Scope of e-commerce in agri-business in India: An overview*. Paper presented at the International Conference on Commerce and Management Research.
3. **Chanana, N., & Goele, S.** (2014). *Future of e-commerce in India*. Paper presented at the International Conference on Emerging Trends in E-Commerce.
4. **Sharma, K.** (2020). *A surge in e-commerce market in India after COVID-19 pandemic*. Paper presented at the National Conference on COVID-19 Impact Studies.
5. **Wankhede, S. G.** (2023). *Impact of e-commerce in Indian economy*. Paper presented at the International Conference on Social Science and Information Studies Research.
6. **Anooja, A.** (2015). *Digital India with e-commerce revolution in rural India: Transform India digitally and economically*. Paper presented at the International Conference on Digital Infrastructure Development.

7. **Mukherjee, S., & Michael, H.** (2016). *Modern trends, challenges and opportunity of e-commerce in India – An overview*. Paper presented at the International Conference on E-Commerce Trends and Challenges Research Studies.

8. **Chanana, N., & Goele, S.** (2014). *Future of e-commerce in India*. Paper presented at the International Conference on Emerging Trends in E-Commerce.

9. **Nougarahiya, S., Shetty, G., & Mandloi, D.** (2021). *A review of e-commerce in India: The past, present, and the future*. Paper presented at the International Conference on Digital Commerce Research Studies.

10. **Ray, S.** (2011). *Emerging trend of e-commerce in India: Some crucial issues, prospects and challenges*. Paper presented at the International Conference on E-Commerce and Business Evolution Studies.

11. **Choshin, M., & Ghaffari, A.** (2016). *An investigation of the impact of effective factors on the success of e-commerce in small- and medium-sized companies*. Paper presented at the International Conference on SME E-Commerce Studies Research.

12. **Chanana, N., & Goele, S.** (2014). *Future of e-commerce in India*. Paper presented at the International Conference on Emerging Trends in E-Commerce.

13. **Goswami, S.** (2023). *Navigating the digital landscape: India's evolving legal framework for e-commerce, data protection, and cybersecurity*. Paper presented at the International Conference on Digital Economy and Cybersecurity Studies.

14. **Banerji, R., & Singh, A.** (2023). *Do social media marketing activities promote customer loyalty? A study on the e-commerce industry*. Paper presented at the International Conference on Digital Marketing Research Trends.

15. **Bhatt, M.** (2024). *Women empowerment through e-commerce: A study in Indian context*. Paper presented at the Revista Index Journal of Multidisciplinary Research Conference, Gujarat University, Ahmedabad.

16. **Kumari, R., & Kumar, D.** (2021). *From kirana stores to e-commerce: The evolving landscape of consumer perceptions on digital payments in India*. Paper presented at the National Conference on Digital Payment Systems Research Studies, K.R. Mangalam University, Gurugram, Haryana, India.

17. **Vathalulu, K., & Ahirwar, G. S.** (2021). *Review of e-commerce business in India*. Paper presented at the National Conference on Commerce and Business Studies Research Trends, Sri Satya Sai University of Technology & Medical Sciences, Sehore, Madhya Pradesh, India.

18. **Kashyap, A. K., & Chaudhary, M.** (2023). *Cyber security laws and safety in e-commerce in India*. Paper presented at the Law and Safety Conference.

19. **Sahu, G., & Singh, S.** (2023). *E-commerce: An emerging platform to boost global economy*. Paper presented at the International Conference on Multidisciplinary Research Studies.

20. **Salunkhe, S. J.** (2023). *A study of growth of e-commerce business of India in post-pandemic era: An overview*. Paper presented at the British Journal of Multidisciplinary and Advanced Studies Conference.

21. **Singh, M.** (2023). *Emerging trends of e-commerce in India*. Paper presented at the International Conference on Business Management Research Studies.

22. **Raman, P. K. R.** (2021). *Comprehensive analysis of eCommerce and marketplaces: Global perspectives with emphasis on the Indian context*. Paper presented at the Asian Journal of Multidisciplinary Research Conference.

23. **Saini, A.** (2023). *Associating designed information and novel content with marketing strategies: A case in Indian retail*. Paper presented at the Indian Journal of Marketing Research Conference.

24. **Banerji, R., & Singh, A.** (2023). *Do social media marketing activities promote customer loyalty? A study on the e-commerce industry*. Paper presented at the International Conference on Digital Marketing Research Trends.

25. **Neeraj Bansal.,&Sanjeev Kumar Sharma** (2024). *Apparel e-tailing: website quality indicators and customer satisfaction.* Paper presented at the International Conference on Internet Marketing and Advertising.

26. **Kauffmann, E., Peral, J., Gil, D., Ferrández, A., Sellers, R., & Mora, H.** (2019). *Managing marketing decision-making with sentiment analysis: An evaluation of the main product features using text data mining*.

27. **Nalluri, P., & Rao, R. S.** (2023). *Impact of product line on shopping motivations of consumers in e-commerce—A correspondence analysis*.