

Analyzing Consumer Behavior in the Context of Online Social Media Marketing

Authors: Shahadat Hossain^{1*}, Md. Manzurul Hasan², Gahangir Hossain³, Hanan Alasmari⁴

1 Dept. of Computer Science and Engineering, Daffodil International University, Bangladesh.

2 Dept. of Computer Science and Engineering, American International University Bangladesh, Bangladesh.

3 Dept. of Data Science, University of North Texas, USA.

4 College of Computer and Information Sciences, Imam Mohammad Ibn Saud Islamic University (IMSIU), KSA.

Corresponding Author: shahadat.cse@diu.edu.bd

Received: October, 2025 Published: January, 2026

ARTICLE INFO

Keywords:

Data Mining; Consumer Preferences; F-Commerce; Intelligent Negotiation; PCA; Social Computing and Intelligence

© 2026 by the Author(s). This open-access article is distributed under a Creative Commons Attribution (CC-BY) 4.0 license, making research freely available to the public and supporting a greater global exchange of knowledge and human experiments.



ABSTRACT

The rapid expansion of digital marketing across social media has fundamentally transformed how consumers interact with online sellers. In Facebook Commerce (F-Commerce), understanding consumer behavior is vital for enabling targeted promotion, personalized engagement, and efficient negotiation. This study presents a machine-learning-driven framework to analyze consumer intentions using a survey dataset of 656 Facebook users. To address high dimensionality and improve predictive performance, Principal Component Analysis (PCA) is applied at multiple explained variance thresholds (75%, 80%, 85%, 90%, 95%). We evaluate three models — Artificial Neural Network (ANN), Random Forest (RF), and k-Nearest Neighbors (KNN) — on both original and PCA-transformed data. The Random Forest model yields the highest accuracy (0.98), outperforming the ANN and KNN models across most metrics. A Current Reality Tree (CRT) is constructed to identify root causes behind non-purchasing behaviors. The study contributes a novel, PCA-enhanced predictive framework for F-Commerce decision modeling and provides actionable insights for digital marketers.

1.0 Introduction

Digital marketing primarily focuses on promoting products and services through web-based platforms to facilitate consumer purchasing and trading activities (Bhat et al., 2016). Traditionally, e-commerce has served as the foundation for online commercial transactions; however, it has now evolved into two distinct domains: conventional e-commerce and Facebook Commerce (F-Commerce) (Siregar, 2018). Notably, many small businesses, despite maintaining dedicated e-commerce websites, increasingly rely on F-Commerce as their principal sales channel due to its accessibility and user engagement potential.

This study investigates the behavioral dynamics of consumers within F-Commerce, emphasizing the role of psychological and experiential factors in shaping purchasing decisions. It explores the diversity of consumer preferences in Facebook's marketplace ecosystem and addresses the ongoing challenge faced by marketers: influencing consumer choices in favor of specific products or services. Understanding these preferences provides critical insights into how consumers evaluate alternatives, negotiate value, and make decisions within socially influenced environments, including peer networks, media exposure, and societal norms.

Given Facebook's widespread adoption and its significance for small and medium-sized enterprises (SMEs), this research focuses exclusively on Facebook as the platform of interest. As noted by Hawkins and Mothersbaugh, effective marketing strategies must be grounded in a clear understanding of consumer perceptions and attitudes toward products (Hawkins & Mothersbaugh, 2016). To analyze

consumer behavior in F-Commerce, it is essential to understand the underlying decision-making processes that drive product selection.

F-Commerce has demonstrated substantial growth, particularly in regions such as Bangladesh, where it generated BDT 2,000 crore in revenue in 2019, with projections indicating a fourfold increase by 2021 (Zabeen et al., 2013). Similarly, a 2020 Bizrate Insights survey reported that 22% of Facebook users in the United States had made direct purchases through the platform (Bizrate, 2023). Recent studies suggest that Artificial Intelligence (AI) can significantly enhance negotiation mechanisms and address limitations in consumer rationality (Schulize-Horn et al., 2020). AI-driven interactions, such as chatbot-based bargaining, may revolutionize customer engagement. This research aims to identify the key factors that influence consumer behavior in F-Commerce and evaluate the potential of AI-based systems to optimize marketing strategies. Through data-driven analysis, we highlight the most influential elements that trigger consumer engagement in F-Commerce, offering valuable insights for intelligent negotiation and personalized marketing.

In this study, we utilize a real-world dataset collected from Facebook groups, pages, and potential consumers to evaluate the effectiveness of several machine learning algorithms, including Artificial Neural Networks (ANN), Random Forest (RF), and k-Nearest Neighbors (KNN). Unlike existing studies that use PCA only once, we compare five thresholds (75–95%), enabling an empirical understanding of dimensionality-performance trade-offs. A seller-centric recommendation framework is proposed (Fig. 1), offering actionable insights for customer segmentation, negotiation, and trust-building. Data visualizations are employed to illustrate consumer preferences based on their interactions and purchasing behaviors within Facebook Commerce (F-Commerce). To address the challenges posed by high-dimensional and unstructured data, we apply Principal Component Analysis (PCA), a widely used dimensionality reduction technique, which enables efficient representation of the dataset with minimal information loss.

As the volume of consumer data continues to grow exponentially, much of it remains unstructured, making it difficult to extract actionable insights. The proliferation of F-Commerce data, similar to other digital platforms, necessitates advanced analytical approaches to ensure meaningful interpretation. Consequently, dimensionality reduction has become a critical tool in data science for managing complexity and enhancing model performance. Most Facebook users engage with the platform for product research or customer service inquiries, while businesses face challenges such as privacy concerns, low conversion rates, and inconsistent operational practices. These limitations have hindered the full potential of the F-Commerce sector.

Our research incorporates real-world data from diverse Facebook communities to assess the performance of machine learning models across varying levels of data variance using PCA. The dataset comprises multiple variables with numerous observations. We evaluate the performance of ANN, RF, and KNN models by analyzing how outcome metrics respond to different levels of dimensional reduction. Additionally, we identify key factors influencing consumer purchasing preferences within F-Commerce. While our earlier study focused solely on identifying influential factors in F-Commerce using ANN and Random Forest, the present work introduces several methodological advancements (Hossain et al., 2020). Specifically, we integrate PCA-based dimensionality reduction at multiple explained-variance thresholds, compare model behavior under varying feature compression levels, and incorporate a Current Reality Tree (CRT) to identify root causes behind non-purchase behavior. These additions make the current study broader, technically stronger, and more insightful than the previous work.

The primary objectives of this study are as follows:

- a) To evaluate variations in outcome metrics using PCA.
- b) To determine the optimal number of principal components for different levels of variance.
- c) To compare model performance on both the dimension-reduced and original datasets.
- d) To identify and assess the significance of factors influencing consumer purchasing behavior.

This study offers four novel contributions:

- 1) A systematic PCA-variance comparison (75–95%) for F-Commerce behavioral modeling—absent in prior literature.
- 2) Integration of behavioral intention theory, dimensionality reduction, and machine learning models into a unified analytical framework.
- 3) Application of a Current Reality Tree (CRT) to identify root causes behind non-purchase behavior, a technique not previously used in this context.
- 4) A recommendation system designed for F-Commerce sellers based on the most influential behavioral features.

The remainder of this paper is organized as follows: Section 2 reviews related literature on consumer behavior analysis, machine learning algorithms, and PCA. Section 3 outlines the research methodology. Section 4 presents the experimental setup and model implementation. Section 5 discusses the results and analytical insights. Section 6 explores future research directions. Finally, section 7 concludes the study.

2.0 Literature Review

Hossain et al. (2020) investigated the influence of external factors on consumer behavior within F-Commerce and applied machine learning models such as Artificial Neural Networks (ANN) and Random Forest (RF) to identify key determinants of purchasing decisions. Their study focused primarily on recognizing the variables that affect consumer engagement in Facebook-based commerce. Recent studies highlight the importance of trust, product authenticity, social proof, page ratings, and timely customer service in shaping online purchase intentions. In recent years, social commerce and machine-learning-driven personalization have received considerable academic attention. Jiang et al. (2022) examined AI-driven personalization strategies in social commerce, demonstrating how user profiling and recommendation algorithms significantly enhance consumer engagement. Their findings align with our work, where machine learning techniques are used to model purchasing intentions.

Rizwan and Malik (2021) investigated trust modeling and risk perception in Facebook-based purchasing. Their results confirm that trust remains one of the most influential factors in consumer decision-making—consistent with our feature-importance analysis where “lack of trust” exhibited the highest contribution.

Al-Rawi (2023) explored peer validation and community influence within F-Commerce networks, showing how social proof and group dynamics shape users’ purchase likelihood. This strengthens the behavioral assumptions of our study, particularly concerning page ratings and community engagement.

Additionally, several recent machine learning studies (2020–2025) have used RF, ANN, and KNN for behavioral prediction tasks, reporting improved accuracy after dimensionality reduction via PCA. These works support our approach of evaluating PCA at multiple variance thresholds to enhance model performance.

Heijden et al. (2003) explored online purchase intentions by examining consumer behavior from two distinct perspectives: technological factors and trust. Their analysis, based on a sample of 228 online shoppers, employed Cronbach’s alpha and exploratory factor analysis to assess the reliability of constructs. By adjusting the risk scale, they demonstrated that trust plays a pivotal role in shaping consumer attitudes—highlighting that while consumers may desire a product, they are unlikely to purchase it from an untrustworthy source.

Kumar et al. (2017) provided a comprehensive overview of consumer behavior in marketing, defining it as the process by which individuals decide what, where, how, and from whom to purchase. The study distinguished between tangible behaviors (e.g., physical shopping activities) and emotional drivers. Rodriguez-Aguilar et al. (2019) contributed to the field by proposing an alpha-stable regression model for pricing in the Mexican electricity market, offering insights into behavioral modeling in economic contexts.

Consumer decision-making involves evaluating product quality and value based on personal preferences. In this context, the consumer is viewed as a central figure in the marketplace. The literature also addresses

various psychological and behavioral dimensions, including motivation, personality, and perception. Khaniwale (2015) examined the factors influencing consumer decisions, categorizing them into internal (e.g., psychological and personal) and external (e.g., cultural and social) influences. These factors collectively shape consumer behavior, with cultural diversity leading to varied perceptions of the same product across different communities (Yoon, 2021). The study emphasizes the importance of tailoring marketing strategies to specific target audiences.

Socio-economic status and social perception also significantly impact purchasing behavior. These elements are critical for engaging consumers across diverse cultural backgrounds and enhancing the effectiveness of online marketing platforms.

Furthermore, recent research has highlighted the integration of Artificial Intelligence (AI) and Decision Support Systems (DSSs) within the domain of Operational Research (OR) (Gupta et al., 2022). This synergy has facilitated more informed decision-making processes. AI tools, when trained using frameworks such as the Harvard Principles of Negotiation (HPN), can enhance consumer interaction and negotiation strategies (Roger et al., 2011). Our research builds on these foundations by proposing machine learning models that help identify consumer intent and support intelligent bargaining mechanisms within F-Commerce environments.

Islam et al. (2020) developed an intelligent system designed to analyze product reviews and customer feedback, demonstrating the growing application of Artificial Intelligence (AI) in understanding consumer behavior. Businesses are increasingly leveraging AI to assess customer preferences and purchasing patterns. Extensive research has been conducted to explore consumer practices (Alhendawi et al., 2019; Gloor et al., 2020; Kumar et al., 2018; Lang & Rettenmeier, 2017) and retail dynamics, providing valuable insights that support strategic decision-making in commercial environments.

Kumar et al. (2018) applied Natural Language Processing (NLP) and Machine Learning (ML) techniques to analyze user sentiments using Amazon Web Services review data. Similarly, Gloor (2020) utilized word embedding, Long Short-Term Memory (LSTM), and Recurrent Neural Networks (RNN) to classify users based on lifestyle, entertainment preferences, and behavioral certainties derived from Twitter data.

Dimensionality reduction, a key technique in data science, involves minimizing the number of features in a dataset while preserving essential information (Bishop & Nasrabadi, 2006). This process transforms high-dimensional data into a more manageable form, facilitating efficient analysis and model performance (Chowdhury et al., 2020). Various real-time methods are employed for dimensionality reduction (Jolliffe & Cadima, 2014), with Principal Component Analysis (PCA) being among the most widely used. PCA identifies core components within large datasets, revealing patterns and relationships through feature mapping. Its versatility is evident in diverse applications, including the analysis of fossilized teeth to infer dietary habits of prehistoric animals (Gill et al., 2014), marine pressure data in atmospheric science, and digital image compression for improved transmission quality on mobile devices (Ng, 2017).

Beyond consumer analytics, PCA and other data-driven approaches have been applied in operational and environmental domains. For instance, a decentralized coordination mechanism was proposed for a legislation-driven recycling network, enabling task allocation among companies while ensuring compliance with environmental standards (Walther et al., 2008). Another study addressed supply chain coordination between manufacturers and retailers, focusing on fairness in Nash bargaining scenarios (Zhimin et al., 2020). Using differential game models, the researchers developed revenue and cost-sharing contracts to optimize channel efficiency. Their findings revealed that fairness sensitivity among dominant supply chain members significantly influences decision-making and overall performance.

The integration of AI with Decision Support Systems (DSSs) and Operational Research (OR) has further enhanced strategic planning capabilities (Gupta et al., 2022). These technologies offer synergistic benefits, enabling more informed and adaptive decision-making. AI tools, when trained using frameworks such as the Harvard Principles of Negotiation (HPN), can support intelligent bargaining in digital commerce. However, effective implementation requires a deep understanding of consumer behavior within F-Commerce (Roger et al., 2011). Our research contributes to this domain by proposing machine learning models that identify consumer intent and facilitate personalized negotiation strategies.

3.0 Methodology

Marketing strategies in the business world are developed based on consumers' behaviors. Therefore, we concentrate on consumers' attitudes toward F-Commerce. It may impact the development of marketing strategies and compel business owners to consider the quality of services provided to customers.

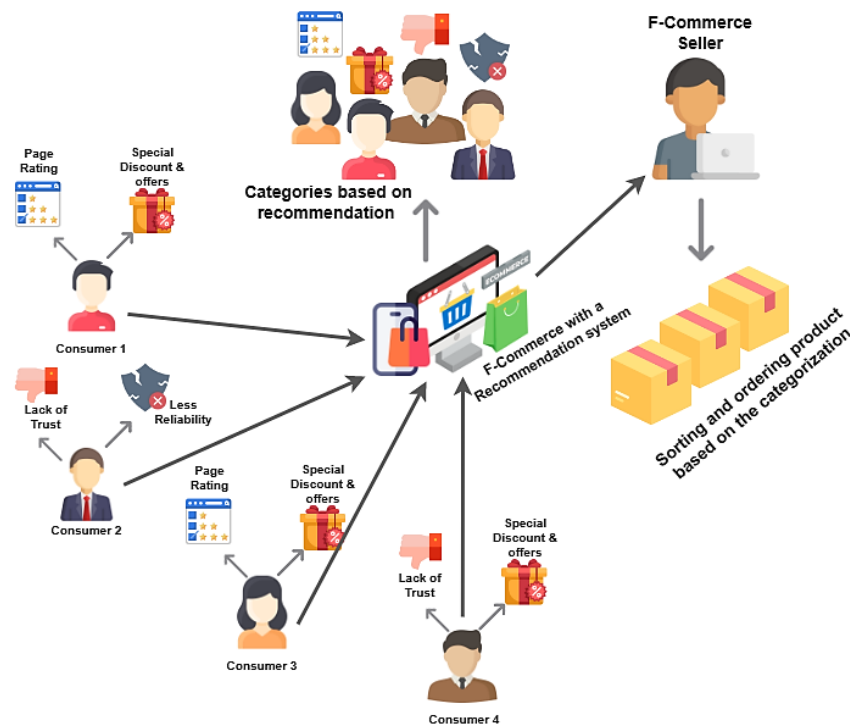


Fig. 1. Proposed Recommendation System

Furthermore, based on the data visualizations, we conclude that various factors influence a consumer's decision to purchase a product through F-Commerce. Taking these variables into account helps business personnel make effective marketing decisions and provide consumers with high-quality services. Based on the findings, we have developed a recommendation system (at Fig. 1) that might help F-Commerce sellers categorize their customers and understand why and how to attract more customers to their sites.

The most decisive steps of this research are highlighted in this section. We process the dataset and use data preprocessing methods to divide the data for training and testing.

3.1 Workflow of Analysis

RF, ANN, and KNN models are used on the dataset for the analyses. Fig. 2 shows the process we follow in this study.

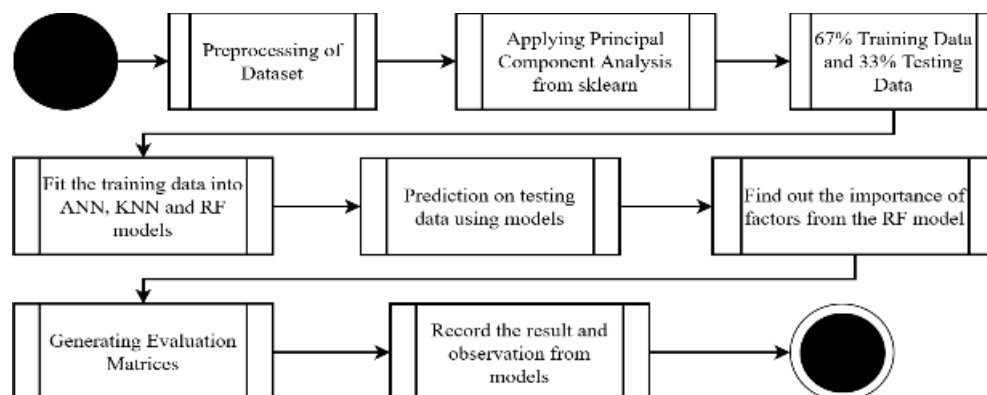


Fig. 2. Workflow of the analysis.

3.2 Dataset Description

The dataset used in this study was collected through a structured survey aimed at analyzing consumer behavior within Facebook Commerce (F-Commerce). The survey was designed based on the behavioral intention model developed by Fishbein & Ajzen (2015), ensuring theoretical grounding in consumer psychology. After preprocessing, which involved the removal of irrelevant or redundant attributes, the final dataset comprised 656 records and 22 variables (Hossain et al., 2020).

One of the key challenges in applying machine learning algorithms is managing data imbalance, which can lead to biased learning outcomes. Our dataset exhibited class imbalance, a common issue in behavioral studies, where certain consumer categories are underrepresented. To address this, we employed the Synthetic Minority Over-sampling Technique (SMOTE) from the Python imbalanced-learn package. SMOTE generates synthetic samples by interpolating between existing minority class instances, thereby enhancing the representativeness of the dataset and improving model performance.

Imbalanced learning, if uncorrected, can significantly affect the accuracy and generalizability of machine learning models. By applying SMOTE, we ensured a more balanced distribution across classes, which is critical for reliable behavioral prediction. Table 1 presents the dataset characteristics prior to the balancing process, highlighting the initial distribution of attributes and observations.

Table 1. Participants' profile sample (N=656)

Features	Count	Percentage
Age		
11-30	634	96.65%
31-60	19	2.90%
61-90	02	0.30%
≥ 90	01	0.15%
Gender		
Male	534	81.40%
Female	119	18.10%
Others	03	0.50%
Profession		
Private Service	101	15.40%
Govt. Service	20	3.05%
Student	515	78.50%
Others	20	3.05%
Time spent in FB		
≤ 2 Hours	122	18.60%
≥ 2 Hours	534	81.40%
Following FB groups and pages		
Following	477	72.70%
Not Following	179	27.30%
Bought Products from F-Commerce		
Yes	356	54.27%
No	300	45.73%

3.3 Data Preprocessing

During the data preprocessing phase, the original dataset—comprising survey responses on consumer behavior in F-Commerce—was transformed from textual to numerical format to facilitate machine learning analysis. Key categorical variables such as age, gender, and profession were encoded for experimental use. The preprocessing also involved dimensionality reduction using Principal Component Analysis (PCA), implemented via the scikit-learn machine learning library.

To standardize the dataset, we applied the StandardScaler class, ensuring that all features were normalized prior to PCA transformation. Subsequently, we computed the eigenvalues and eigenvectors

of the principal components across five target variance thresholds: 75%, 80%, 85%, 90%, and 95%. This allowed us to retain the most informative components while minimizing data loss.

To ensure robustness in model evaluation, the dataset was randomly shuffled 150 times and split into training and testing subsets, with 67% allocated for training and 33% for testing. Table 2 presents the number of principal components retained at each variance level, along with the corresponding explained variance ratios.

Following dimensionality reduction, we applied three machine learning models—Artificial Neural Network (ANN), Random Forest (RF), and k-Nearest Neighbors (KNN)—to each version of the dataset. This enabled comparative analysis of model performance across different levels of data abstraction and variance retention.

Table 2. Different Variances and Components

Variance	Component
75%	08
80%	10
85%	12
90%	14
95%	16

3.4 Principal Component Analysis

Principal Component Analysis reduces dimensionality by transforming correlated variables into a set of mutually uncorrelated principal components. These components are derived from the eigenvectors and eigenvalues of the covariance matrix. Each successive component captures decreasing proportions of total variance. Instead of detailing full derivations, we follow the standard PCA workflow implemented in scikit-learn: scaling the data, computing the covariance matrix, extracting eigenvalues, ranking components, and selecting the number required to meet the target explained variance thresholds.

The principal components are determined using two essential sections from linear algebra (Jaadi, 2022). The first is eigenvectors, and the second is eigenvalues. Each column's eigenvector represents a principal component. The first principal component is seen in the first column of the covariance matrix. As an example, the n^{th} column represents the n^{th} principal variable. The first component contains most of the materials, and this quantity steadily decreases as the number of components increases. In this case, the eigenvalue is an element that influences data translation to the principal component. Furthermore, eigenvalues are used to quantify the percentage variances of variables. Finally, the mathematical operation described below is used to obtain dimension-reduced data (Krak et al., 2019).

$$\text{Final Data} = \text{Feature Vector}^T * \text{Data Adjust}^T \quad (1)$$

To address the high dimensionality of the dataset, we applied Principal Component Analysis (PCA). The selection of PCA variance thresholds (75%, 80%, 85%, 90%, and 95%) was not arbitrary. First, we computed the eigenvalues of the covariance matrix and observed the cumulative explained variance curve, following standard PCA practice in behavioral and marketing analytics. The chosen thresholds represent commonly accepted coverage ranges where most of the dataset's information can be retained while reducing noise and multicollinearity. This systematic selection allows us to compare low-, medium-, and high-variance PCA configurations and evaluate their impact on ANN, RF, and KNN performance.

3.5 Decision Tree

A decision tree is a type of classification algorithm that operates by recursively partitioning the instance space into subsets based on feature values. Structurally, a decision tree is represented as a hierarchical, downward-directed graph beginning with a root node. Each internal node—also referred to as a test node—represents a decision point based on a specific attribute, and branches from these nodes correspond to possible outcomes of the test. Terminal nodes, known as leaf nodes or decision nodes, indicate the final classification or decision outcome (Rokach & Maimon, 2005).

This tree-based structure enables intuitive interpretation of decision paths and facilitates efficient classification by following a sequence of logical rules. The recursive nature of the tree allows it to adaptively split the data space, making it a powerful tool for both classification and regression tasks.

3.6 Random Forest

Random forest is a machine learning algorithm for predicting data from a dataset (Rokach & Maimon, 2005). This algorithm was created by (Ho, 1998). The main reason for using RF is that it has a low bias and variation. RF, on the other hand, outperforms different decision tree algorithms such as j48 in terms of correctness (Ali et al., 2012). The number of trees in RF boosts efficiencies and increases accuracy estimation, but it also delays the computation. The RF algorithm needs the highest number of features when dividing a node and the fewest number in the case of leaf nodes. The RF algorithm is more stable than other decision tree algorithms since it can be used for regression and classification. This algorithm is also very convenient because the default hyper-parameters are used for the predictions of results.

3.7 Artificial Neural Network

Artificial Neural Networks (ANNs) are computational models inspired by the structure and functioning of the human brain, particularly in their ability to perceive and interpret environmental stimuli. An ANN consists of multiple interconnected layers, each composed of individual processing units or neurons. These layers include an input layer, one or more hidden layers, and an output layer. The network is constructed by forward-linking units from one layer to the next, allowing information to flow from input to output through a series of transformations.

Connections between neurons are represented as edges, each associated with a weight that determines the strength of the signal transmitted. During forward propagation, weighted inputs are aggregated and passed through an activation function, which introduces non-linearity and enables the network to learn complex patterns. The output of each neuron is then forwarded to the subsequent layer.

To optimize learning, the network employs a backpropagation algorithm (Hecht-Nielsen, 1992), which adjusts the weights based on the error between predicted and actual outputs. The Artificial Neural Network (ANN) model used in this study consists of an input layer with 18 features, one hidden layer with four neurons, and an output layer for binary classification. The hidden layer uses a ReLU activation function, whereas the output layer uses a sigmoid activation function to estimate purchase intention. The network is trained using stochastic gradient descent (SGD) with a learning rate of 0.01 and binary cross-entropy as the loss function. We trained the model for 300 epochs with a batch size of 32 and applied 10-fold cross-validation to minimize overfitting. This detailed configuration ensures the ANN is optimized for stability and generalization, addressing concerns about insufficient architectural description.

3.8 K-Nearest Neighbor

KNN is a non-parametric classification system (Ali et al., 2012). The training samples are fed into KNN, and the quality is determined by the process (classification or regression). When we use KNN, we specify these forms. In the KNN-classifier, the new member is identified by maximum vote of neighbors, where $k(k=1,2,3,4,5,\dots,n)$ is a positive integer. This k reflects the number of the closest neighbor. For each new member, KNN-regression returns the average value of k nearest neighbors. The Euclidean distance of the current class value of samples is used to fit the training samples into the model. It predicts the class of the experimental dataset using the closest distance from the training sample points.

4.0 Experiment with Models

This section describes the experimental method and models' analyses. Following the data preprocessing and PCA, we use three machine learning algorithms. First, RF and KNN are used for classification, while ANN is used for deep learning. Supervised learning explains the usage of labeled datasets to create models that efficiently categorize data or make predictions.

4.1 Random Forest Model

The Random Forest algorithm is implemented using the sklearn predictive python library. This classifier algorithm generates 20 distinct decision tree models from which the algorithm predicts the testing data. We use one of these 20 similar trees as an example (Fig. 3) that contains three types of information. The first is gini, which is a likelihood of a randomly selected sample in a node, and the value of gini decreases as we go down the tree (Ali et al., 2012). Another piece of information is a sample, which contains the findings in a node. Finally, class is the prediction of all samples in a node that has the majority classification. The leaf nodes make the final decision.

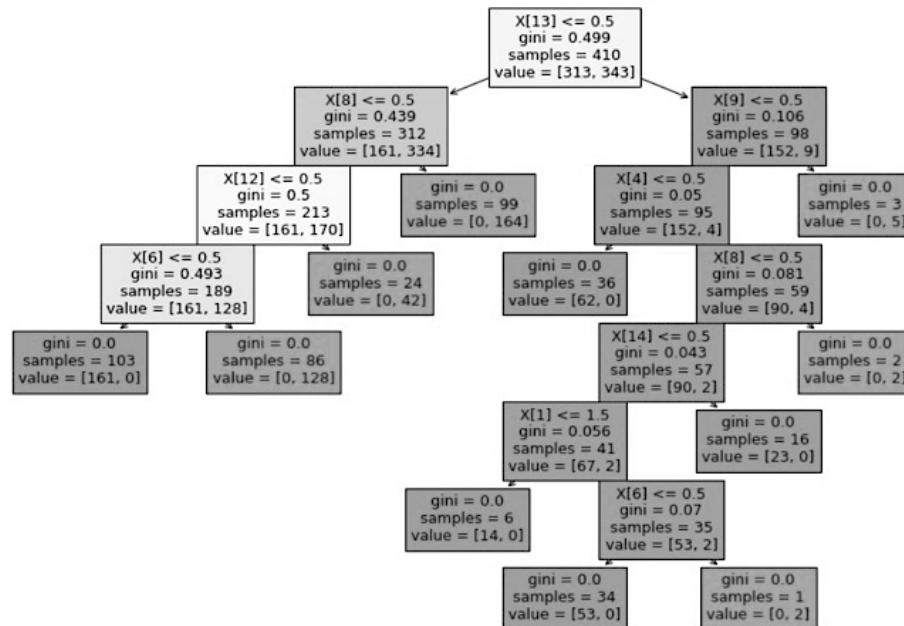


Fig. 3. Random forest decision tree model.

4.2 Importance of Features

The value of features in our Random Forest is used to calculate the effect of features on our RF model. It provides the most valuable features from our dataset. We have a limit of 18 features after preprocessing and use this `feature_importances_`, a method from python Random Forest classifier library to calculate the importance of the features.

According to Fig. 4, the determining factor 'lack of trust' in the sense of not purchasing from F-Commerce has the highest importance in our model (0.25). The top three most critical features include a lack of trust (0.25), less reliability (0.16), and page ratings (0.13). The lowest importance is (0), indicating that these features do not affect our model.

4.3 Artificial Neural Network Model

We introduce an ANN model (Fig. 5) after using PCA to minimize the dataset with a principal component using the StandardScaler class from sklearn. The ANN model's input consists of a total of 18 function values and a bias input. Keras, a Python deep-learning library, is used in our experiments. TensorFlow is being included in the library's backend. The hidden layer comprises 4 units, and its and the output layer's dimensions are R^5 and R^1 respectively. We run our model 300 times with a learning rate of 0.01. In addition, to minimize the error, we employ the k-fold, binary cross-entropy loss function, and stochastic gradient descent optimizer. Finally, we apply our trained model to the testing dataset.

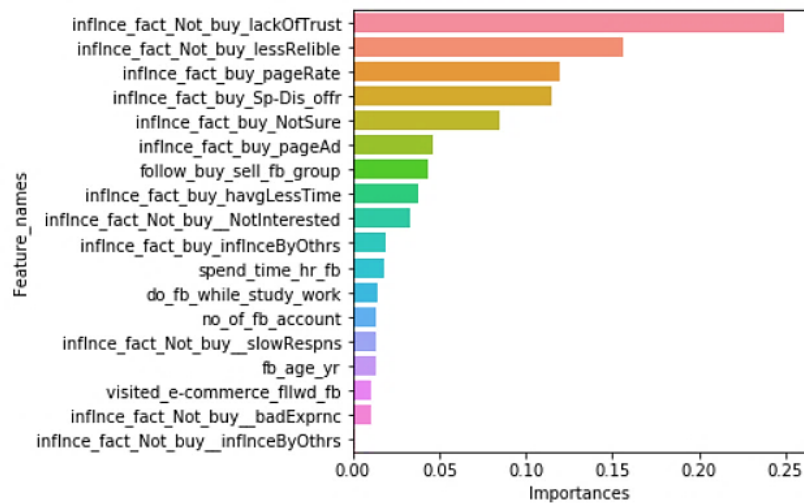


Fig. 4. Importance of feature bar chart

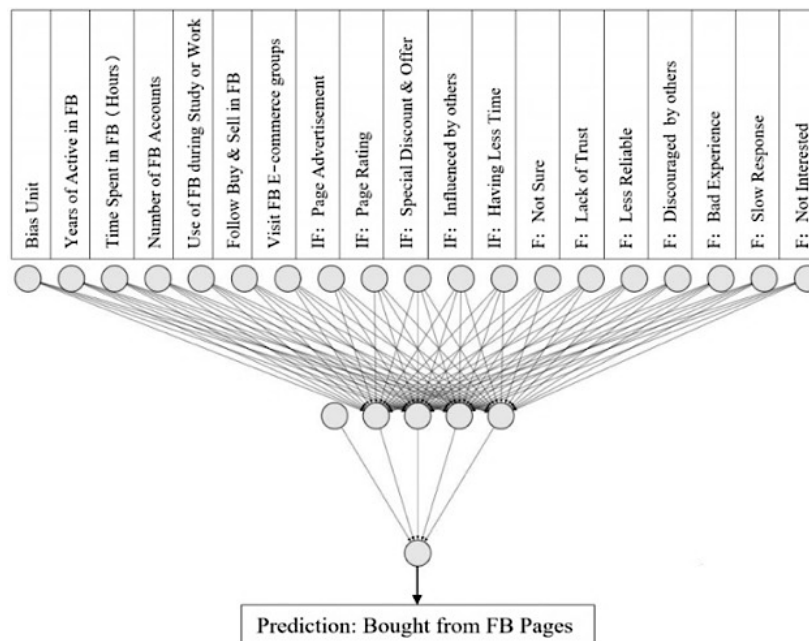


Fig. 5. Artificial neural network model

4.4 K-Nearest Neighbor Model

We use the sklearn.neighbors library's method named KNeighborsClassifier for applying KNN classifier. We have varied the value of k (1 to 40) to fit the training data in the classifier model. The accuracy rate, error rate, and evaluation parameters are then computed. The accuracy rate and the error rate are plotted on graphs showed in Fig. 5 to allow for comparisons with various k values. We can infer from the graph (Fig. 6) which point of precision is being found in specific k values.

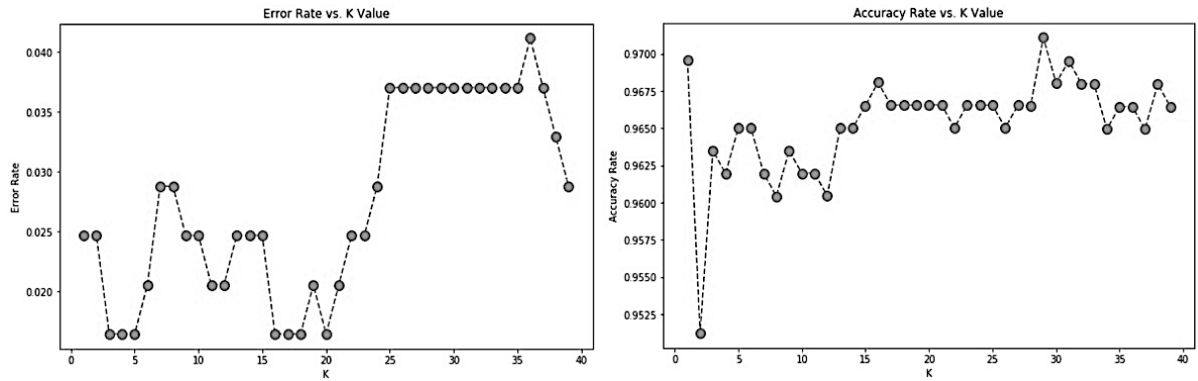


Fig. 6. Accuracy rate and Error rate in KNN.

The KNN model performs well (0.987) while the value of k is 30. The accuracy rate remains stable at 0.966 in between $k=16$ to $k=22$. This classifier's accuracy rate dropped up to 0.946. The error rates, on the other hand, remain stable at 0.037 after $k=25$, with the highest error rate of 0.045 at $k=37$.

5.0 Analysis and Discussion

This section explains our observations about the models' performances and the findings of our research. The evaluation metrics and other performance characteristics of the models are analyzed and examined.

5.1 Confusion Matrix of Models

A confusion matrix table consisting of True Positive (TP), False Positive (FP), False Negative (FN), and True Negative (TN) values of the test data is illustrated in Table 3 below for different models.

Table 3. Confusion matrix of RF, ANN, K-NN.

		Actual Value		RF		ANN		KNN	
		Bought	Not Bought	Bought	Not Bought	Bought	Not Bought	Bought	Not Bought
Predicted Value	Bought	TP	FP	107	1	121	0	105	3
	Not Bought	FN	TN	2	133	0	97	6	129

True positives indicate that our models correctly predict those occurrences as they purchase from F-Commerce in the actual dataset. True negative signifies that our models do not include those cases since they do not buy from F-Commerce in the actual dataset. On the other side, false positives imply that our models anticipate such instances as positive, but they cannot be positive since they do not purchase from F-Commerce. Finally, false negative indicates that models predict negatively in terms of the predicted label purchased from F-Commerce, but such occurrences are not there in the actual dataset. The confusion matrix of random forest is shown in Table 3, which indicates that the RF model can predict 107 instances as true positive and 133 instances as true negative from the testing dataset. Furthermore, the RF model predicts 1 instance to be false positives and 2 instances to be false negative. The ANN model predicts 121 occurrences to be true positives and no occurrences to be true negatives. The ANN model produced an unusual confusion matrix (TP = 121, TN = 0, FP = 0, FN = 97), indicating that the model predicted only the positive class. Although the model obtained high accuracy on the positive instances, it failed to correctly classify any negative instances. This behavior suggests that the ANN became biased toward the majority class despite the application of SMOTE, or that its current architecture and training parameters did not sufficiently capture the decision boundary for the negative class. Such class-prediction bias is a known issue in binary classification tasks with skewed behavioral data.

We revalidated the results and confirmed that this pattern stems from the current ANN configuration. Future improvements may include tuning the network architecture, adjusting the learning rate, modifying

class weights, or rebalancing the dataset using alternative resampling methods to ensure better classification of the negative class. The confusion matrix of the KNN model is given in the same table, where 105 instances are predicted as true positives and 129 instances are predicted as true negatives. Thus, false positives are predicted to be 3, while false negatives are predicted to be 6.

5.2 Models' Performances on Different Variances

We use PCA to reduce the component to get the highest performance through our models in data preprocessing. Distinct variations have resulted in different model performance characteristics. Some models' performance metrics have been improved significantly, while others have remained constant with specific variances (Table 4). We examine all the variances' outputs and present them in various graphs for the RF, ANN, and KNN models (Fig. 6 above).

Table 4. Performance Metrics Result Table

Variance	Component	Performance Matrices	RF	ANN	KNN
100% Real Dataset [11]	18	Accuracy	0.98	0.98	0.97
		Precision	0.98	0.99	0.97
		Recall	0.99	0.98	0.97
		F1-Score	0.98	0.98	0.97

The performance of the RF model improves as the variances increase. Fig. 7 shows that the RF model performed best across all performance metrics while the variance is 95. The recall of the ANN model remains constant after the variance of 85 (Fig. 7). At variance 95, the ANN model outperformed. Fig. 7 for the KNN model shows that at the variance of 75, the model performed best at 0.97 among all performance metrics.

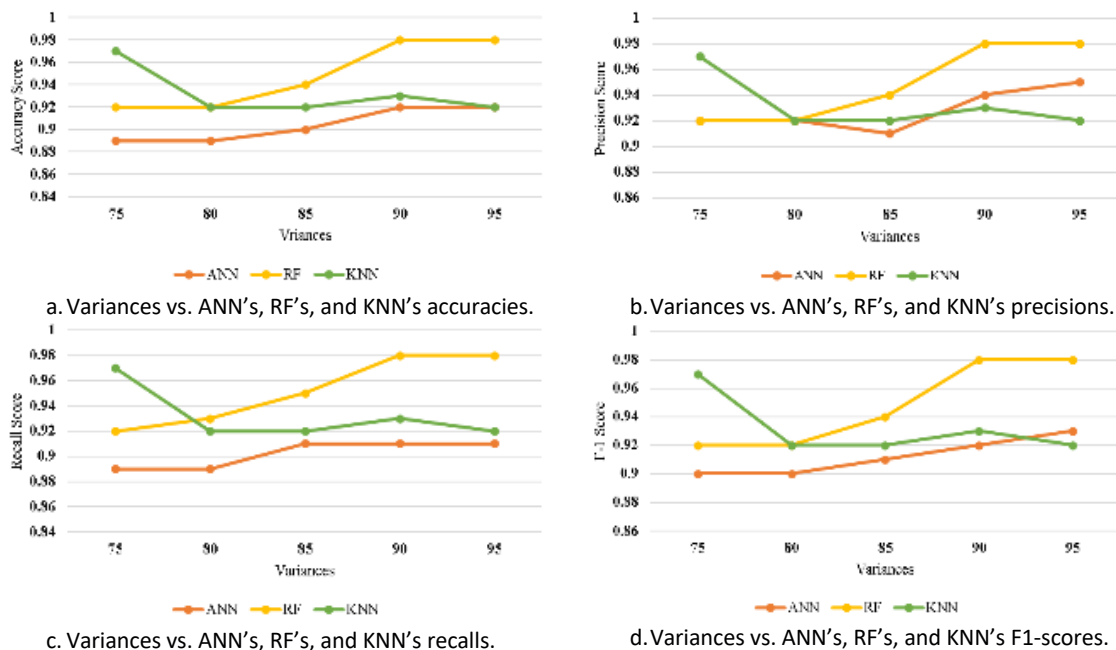


Fig. 7. Graph analysis of result metrics for different variances after PCA

5.3 Data Visualizations

We select the top four most significant features for data visualizations and select 'Factor not to purchase due Lack of Trust,' 'Factor not to purchase due Less Reliable,' 'Influencing factor Page Rating,' and

'Influencing factor Special Discount & Offer' for this section. In the bar charts, the positive response is denoted by 1, and the negative response is denoted by 0 (Fig. 8). The influence of the features is shown on the x-axis, while the count of the respective value is shown on the y-axis.

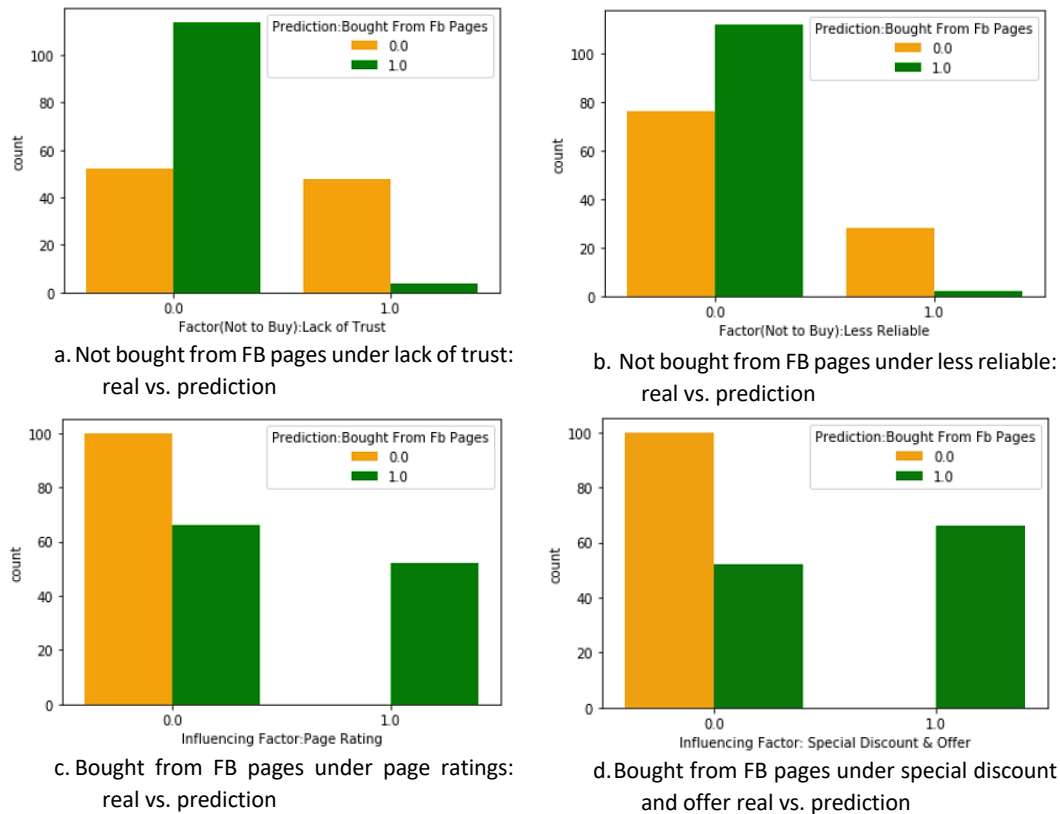


Fig. 8. Different bar chart analysis of predicted data

The green bar in the chart shows the positive responses to the prediction factors, whereas the yellow bar shows the negative responses to the prediction factors using the testing dataset (33% of the total dataset). According to Fig. 8a, consumers buy products from F-Commerce if and only if they trust Facebook pages and groups.

We might conclude from this perspective that product negotiations should be centered on the interests of consumers. Fig. 8b shows that consumers will only purchase things if F-Commerce is reliable. Consumers are going to assess reliability if and only if F-Commerce can provide them with trust. Fig. 8c, on the other hand, depicts the component that influences consumers' purchases from F-Commerce depending on page rating. Some consumers make purchases without taking page ratings into their accounts. Consumers purchase products at a particular discount or offer, which is consistent with the actual situation (Fig. 8d). Furthermore, some customers purchase without taking advantage of any special discounts or offers. We have used the k-fold cross-validation ($k=10$) for RF, and binary cross-entropy loss function, and stochastic gradient descent optimizer for ANN to reduce the error.

5.4 Causal Analysis-Current Reality Tree

The Theory of Constraints uses a current reality tree (CRT) as a graphical tool to find and fix problems by locating their root causes. The CRT begins with a problem statement or a list of undesirable effects (UDEs), which are then examined to determine their root causes and organized into a hierarchical tree structure based on cause-and-effect relationships.

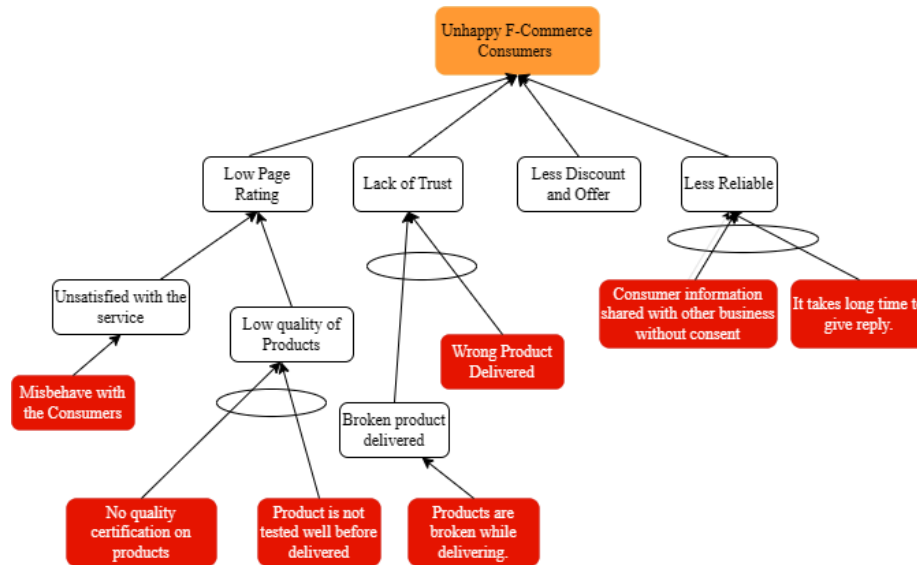


Fig. 9. Current Reality Tree in F-Commerce

We build a CRT (Fig. 9) to identify the dissatisfied F-Commerce customers in order to understand why they removed themselves from the F-Commerce. We examined the data and plotted it in the CRT for better visibility because 300 (45.73%) of the respondents did not make any purchases from F-Commerce.

Low ratings are a contributing factor for these customers, and if they discover low ratings in FB groups and pages, they will not purchase anything. We delve a little further to uncover more factors. Customers' dissatisfaction has a significant impact on the bad rating on FB, and those customers are dissatisfied. The lack of confidence is growing as more F-Commerce companies deliver defective goods. As a result, less experienced delivery personnel from F-Commerce delivered damaged goods. Fewer offers and discounts are another factor. When offered no special discounts or incentives, these customers are frequently unwilling to buy from F-Commerce and become unsatisfied. Therefore, that is still another reason why customers avoid F-Commerce. Some customers believed that F-Commerce was less trustworthy because some F-Commerce businesses shared customer information with other F-Commerce businesses. Additionally, if F-Commerce businesses are responding to their customers too slowly, customers will see them less favorably.

Our models are validated by various variances, methods and these are highly effective. In addition, we have provided certain elements effecting the models that might help the businessmen to attract consumers. This research has emphasized the problems of F-Commerce in the small online market where consumers' behaviors have usually been overlooked. Moreover, this research in the corporate sector can broaden the existence of Facebook's digital marketing efforts.

Our proposed models and analyses provide a perspective on doing analytical research on consumers' behaviors in F-Commerce. This study has some limitations that point to some future research directions. This research is based on the most popular social media platform Facebook, limiting the overall impact of other platforms such as YouTube, Twitter, and LinkedIn. There is no endpoint of analysis, but we must assume and provide a solution based on our perspectives. It would also be helpful to validate and to verify our findings using a larger data set as future consequences.

6.0 Conclusion

The significance of consumer behavior analysis in shaping effective marketing strategies knows no bounds, particularly for small businesses operating on Facebook. Understanding consumers' behaviors and providing high-quality services are paramount for success. In this study, we delved into machine learning models, leveraging PCA dimensionality reduction to enhance our dataset analysis. By employing

ANN, RF, and KNN, we successfully predicted and improved the perception of Facebook purchases across various scenarios.

Among all ML models, the RF model demonstrated exceptional performance, achieving an impressive 98% accuracy with 100% variance, surpassing the other two models in several parameters. Hence, an optimal consumer behavior analysis recommender system is plausible.

Drawing from our research findings, we propose a recommendation method that empowers F-Commerce vendors to gain valuable insights into the factors influencing or inhibiting consumers' purchasing decisions. This method identifies both the appealing features of a seller's page that attract customers and the deterrents that impact F-Commerce customer adoption behavior. By enhancing the user experience of online businesses, our study aims to provide F-Commerce vendors with actionable areas for improvement, leading to increased sales and bolstered confidence in product negotiations.

Furthermore, we have identified key characteristics that significantly influence our models' prediction processes, such as the persuasive effect of "page ratings" compared to "lack of trust" or "less reliability" in influencing consumer preferences.

As our research contributes to a deeper understanding of consumer behaviors, we believe it will inspire and motivate other researchers to further explore this critical domain. Ultimately, we hope our study serves as a steppingstone for advancing the understanding of consumer behaviors and fostering continuous improvement in F-Commerce practices.

Acknowledgment

The authors express their gratitude to the survey participants and the online F-Commerce communities whose contributions made this research possible. The authors also acknowledge the helpful comments of the reviewers, which significantly improved the quality of this manuscript.

References

- [1] K. M. Alhendawi, A. A. Al-Janabi, and J. Badwan. Predicting the quality of mis characteristics and end-users perceptions using artificial intelligence tools: Expert systems and neural network. In *Proceedings of the International Conference on Intelligent Computing and Optimization*, Hua Hin , Thailand, 2019, pages 18–30. Springer.
- [2] J. Ali, R. Khan, N. Ahmad, and I., Maqsood. Random forests and decision trees. In *International Journal of Computer Science Issues (IJCSI)*, 9(5):272, 2012.
- [3] S. A. Bhat, K. Kansana, and J. Khan. A review paper on e-commerce. In *Asian Journal of Technology and Management Research* [ISSN: 2249–0892], 6(1), 2016.
- [4] R. R. Chowdhury, M. A. Adnan, and R. K. Gupta. Real-time principal component analysis. *ACM Transactions on Data Science*, 1(2):1–36, 2020.
- [5] M. Fishbein and I. Ajzen. Predicting and changing behavior the reasoned action approach. OCLC: 1190691560, 2015.
- [6] P. G. Gill, M. A. Purnell, N. Crumpton, K. R. Brown, N. J. Gostling, M. Stampanoni, and E. J. Rayfield. Dietary specializations and diversity in feeding ecology of the earliest stem mammals. *Nature*, 512(7514):303–305, 2014.
- [7] P. Gloor, A. Fronzetti Colladon, J. M. de Oliveira, and P. Rovelli. Put your money where your mouth is: Using deep learning to identify consumer tribes from word usage. In *International Journal of Information Management*, 51:101924, 2020.
- [8] D. I. Hawkins and D. L. Mothersbaugh. *Consumer behavior: building marketing strategy*. McGraw-Hill Education, New York, NY, thirteenth edition, 2016.
- [9] R. Hecht-Nielsen. *Theory of the backpropagation neural network*. Elsevier, 1992.
- [10] T. K. Ho. The random subspace method for constructing decision forests. In *IEEE transactions on pattern analysis and machine intelligence*, 20(8):832–844, 1998.

- [11] S. Hossain, M. M. Hasan, and T. Hossain. An Analytical Study of Influencing Factors on Consumers' Behaviors in Facebook using ANN and RF. In *International Conference on Intelligent Computing & Optimization*, pages 744–753. Springer, 2020.
- [12] S. M. Islam, M. S. A. Forhad, M. A. Uddin, M. S. Arefin, S. M. Galib, and M. A. Khan. Developing an intelligent system for recommending products. In *International Conference on Intelligent Computing & Optimization*, pages 476–490. Springer, 2020.
- [13] Z. Jaadi. "A Step-by-Step Explanation of Principal Component Analysis (PCA)", 2022 [Online]. Available: <https://builtin.com/data-science/step-step-explanation-principal-component-analysis>.
- [14] I. T. Jolliffe. and J. Cadima. Principal component analysis: a review and recent developments. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 374(2065):20150202, 2016.
- [15] M., Khaniwale. Consumer Buying Behavior. In *International Journal of Innovation and Scientific Research*, 14(2):278–286, 2015.
- [16] I. Krak, O. Barmak, E. Manziuk, and A. Kulias. Data classification based on the feature's reduction and piecewise linear separation. In *International Conference on Intelligent Computing & Optimization*, pages 282–289. Springer, 2019.
- [17] P. Kumar, S. Nandagopalan, and L. Swamy. Investigation of emotions on purchased item reviews using machine learning techniques. In *Proceedings of the International Conference on Intelligent Computing & Optimization*, pages 409–417. Springer, 2018.
- [18] R. Kumar. Consumer behaviour and role of consumer research in marketing. In *Journal of Commerce and Trade*, 12(1):65–76, 2017.
- [19] T. Lang. and M. Rettenmeier. Understanding consumer behavior with recurrent neural networks. In *Proceedings of the Workshop on Machine Learning Methods for Recommender Systems*, 2017.
- [20] S. Ng. Principal component analysis to reduce dimension on digital image. *Procedia computer science*, 111:113–119, 2017.
- [21] R. Rodriguez-Aguilar, J. A. M. Saucedo, and P. Vasant. Estimation of electricity prices in the Mexican market. In *International Conference on Intelligent Computing & Optimization*, pages 11–17. Springer, 2019.
- [22] L. Rokach and O. Maimon. Top-Down Induction of Decision Trees Classifiers—A Survey. In *IEEE Transactions on Systems, Man and Cybernetics, Part C (Applications and Reviews)*, 35(4):476–487, 2005.
- [23] E. Siregar. Young attitudes towards f-commerce advertising. In *Proceedings of the International Conference on Industrial Technology and Management (ICITM)*, pages 218–223, 2018.
- [24] I. Schulze-Horn, S. Hueren, P. Scheffler, H. Schiele. Artificial intelligence in purchasing: Facilitating mechanism design-based negotiations. *Applied Artificial Intelligence*. 2;34(8):618-42, 2020.
- [25] S. Gupta, S. Modgil, S. Bhattacharyya, I. Bose. Artificial intelligence for decision support systems in the field of operations research: review and future scope of research. *Annals of Operations Research*. pp 1:1-60, 2022.
- [26] F. Roger, L. U. William, Bruce Patton. *Getting to Yes: Negotiating Agreement Without Giving In*. Penguin Books, 2011.
- [27] Bizrate. "Facebook commerce survey", 2023 [Online]. Available: <https://www.bizrate.com/facebook-commerce-survey/#:~:text=Only%2022%25%20reported%20making%20a,features%20to%20aid%20the%20process>.
- [28] M. Zabeen, H. Ara, and N. Sarwar. F-Commerce in Bangladesh: "Venit, Vidit, Vicit". *IOSR J. Hum. Soc. Sci*, 17(5), 1-08, 2013.
- [29] G. Walther, E. Schmid, S. T. Spengler, Negotiation-based coordination in product recovery networks, *International Journal of Production Economics*, Volume 111, Issue 2, 2008, Pages 334-350, ISSN 0925-5273, <https://doi.org/10.1016/j.ijpe.2006.12.069>, 2008.
- [30] G. Zhimin, Y. Tong, Y. Rui, Channel coordination under Nash bargaining fairness concerns in differential games of goodwill accumulation, *European Journal of Operational Research*, Volume 285, Issue 3, pp 916-930, <https://doi.org/10.1016/j.ejor.2020.02.028>. (<https://www.sciencedirect.com/science/article/pii/S0377221720301466>), 2020.

- [31] S. Yoon. "Why diversity within your organization matters – Lessons from 11 entrepreneurs", 2021 [Online]. Available: <https://www.weforum.org/agenda/2021/11/why-diversity-within-your-organization-matters/>.
- [32] C. M. Bishop, and M. N. Nasrabadi. Pattern recognition and machine learning. Vol. 4, no. 4. New York: springer, 2006.
- [33] Y. Jiang, X. Li, and H. Wang, "AI-driven personalization strategies in social commerce: User profiling, recommendation systems, and consumer engagement," Journal of Retailing and Consumer Services, vol. 68, p. 103001, 2022.
- [34] M. Rizwan and S. Malik, "Trust modeling and risk perception in Facebook-based online purchasing: An empirical investigation," Electronic Commerce Research and Applications, vol. 46, p. 101040, 2021.
- [35] A. Al-Rawi, "Community influence, peer validation, and consumer decision patterns in Facebook Commerce (F-Commerce)," International Journal of Information Management, vol. 69, p. 102597, 2023.