Project Dissertation Report on

GREEN PURCHASE BEHAVIOR:

BRIDGING THE INTENTION-BEHAVIOR GAP

Submitted By

Vaishali

24/BMBA/37

Under the Guidance of

Dr Gaganmeet Kaur Awal

Assistant Professor



DELHI SCHOOL OF MANAGEMENT

Delhi Technological University

Bawana Road, Delhi 110042

CERTIFICATE

This is to certify that the Major Research Project titled **"Green Purchase Behavior: Bridging the Intention-Behavior Gap"** has been completed by **Ms Vaishali**, Roll No. 24/BMBA/37, for the partial fulfilment of the requirements for the award of the degree of Master of Business Administration (MBA) in Business Analytics from Delhi School of Management,DTU.

Date:17th May 2026

Signature of Guide: _____

Dr Gaganmeet Kaur Awal

Assistant Professor

Delhi School of Management, DTU

DECLARATION

I, Vaishali, hereby declare that the Major Project Report entitled "**Green Purchase Behavior: Bridging the Intention-Behavior Gap**" submitted in partial fulfilment of the requirements for the award of the degree of Master of Business Administration is my original work and has not been submitted for the award of any degree, diploma, or fellowship of any other University or institution.

All the information presented in this report is derived from my personal research and is accurate to the best of my understanding. I also state that all the sources of information used in preparing this report have been properly recognised.

Student:
Vaishali
24/BMBA/37

Date: 17th May 2026

ACKNOWLEDGEMENT

The researcher wishes to express sincere gratitude to **Dr Gaganmeet Kaur Awal** for her valuable guidance, constructive feedback, and continuous support throughout this research project. Her suggestions and encouragement were extremely helpful at every stage of the work and played an important role in bringing this research to completion.

The researcher also genuinely thanks the faculty members of the Delhi School of Management, DTU, for their academic guidance and for creating a supportive intellectual environment that made it possible to carry out this study. Special gratitude is also extended to the 103 individuals who willingly participated in the survey their willingness to cooperate was crucial for the successful completion.

Finally, the researcher acknowledges the support of family and friends, whose encouragement and patience sustained this endeavour from inception to completion.

Vaishali
Delhi School of Management, DTU
2025-26

TABLE OF CONTENTS

CHAPTER 1: INTRODUCTION 1
1.1 Problem Statement 1
1.2 Research Gaps 2
1.3 Key Contributions of the Study 3
1.4 Organization of the Report 4
CHAPTER 2: LITERATURE REVIEW 5
2.1 The Problem Area: Green Consumer Behavior and the Intention-Behavior Gap 5
2.2 Technique Area: Theoretical Frameworks and Methodological Approaches 6
2.2.1 Theory of Planned Behavior 6
2.2.2 Value-Belief-Norm Framework 7
2.2.3 Self-Identity and Green Identity 7
2.2.4 Social Media Influence 8
2.2.5 Machine Learning in Green Consumer Behavior Research 8
CHAPTER 3: PROPOSED METHODOLOGY 9
3.1 Step-by-Step Analytical Framework 9
3.2 Explanation of Models 10
3.2.1 Point-Biserial Correlation 11
3.2.2 Independent Samples t-Test 11
3.2.3 Chi-Square Test of Independence 11
3.2.4 Moderation Analysis with Interaction Term 11
3.2.5 Logistic Regression 11
3.2.6 Random Forest Classifier 12
3.2.7 Class Balancing and Cross-Validation 12
CHAPTER 4: EXPERIMENT AND RESULTS 13
4.1 Baseline Models and Evaluation Metrics 13
4.2 Dataset Preparation 13
4.3 Pre-Processing 14
4.4 Experiments Performed 15
4.4.1 Exploratory Data Analysis 15
4.4.2 Reliability Analysis 16
4.4.3 Correlation Analysis 17
4.4.4 Hypothesis Testing Results 18
4.4.5 H5 Quadrant Analysis 19
4.4.6 Machine Learning Model Results 20

4.5 Discussion: Findings and Analysis 21
CHAPTER 5: CONCLUSION AND FUTURE IMPLICATIONS 23
5.1 Managerial Implications 23
5.2 Limitations 24
5.3 Future Research Directions 24
REFERENCES 26
APPENDIX: RESEARCH QUESTIONNAIRE 28
Section A: Demographic Information 28
Section B: Environmental Identity and Attitude (IDN) 28
Section C: Environmental Knowledge (AWN) 29
Section D: Social Media Influence (SMI) 29
Section E: Purchase Intention (INT) 30

LIST OF FIGURES

Figure	Title
5.1	Exploratory Data Analysis: Demographic Profiles, Purchase Distribution, and Construct Score Distributions
5.2	Correlation Matrix and Correlations with Actual Purchase Behavior
5.3	Hypothesis Testing Plots: Intention, Knowledge, Social media, and Identity Analysis
5.4	H5 Quadrant Analysis: 2x2 Purchase Rate Matrix (Intention x Knowledge)
5.5	Machine Learning Model Evaluation: Confusion Matrices, ROC Curves, and Feature Importance

LIST OF TABLES

Table	Title
3.1	Theoretical Hypotheses and Basis
4.1	Constructs and Measurement Items
4.2	Data Cleaning Process and Sample Reduction
4.3	Statistical Techniques and Justification
4.4	Baseline Models and Evaluation Metrics
5.1	Descriptive Statistics and Cronbach's Alpha
5.2	Hypothesis Testing Results Summary
5.3	H5 Quadrant Analysis: Purchase Rates by Intention-Knowledge Group
5.4	Machine Learning Model Performance Comparison

LIST OF ABBREVIATIONS

Abbreviation	Full Form
TPB	Theory of Planned Behavior
VBN	Value-Belief-Norm Framework
IBGap	Intention-Behavior Gap
EK	Environmental Knowledge
SMI	Social Media Influence
EI	Environmental Identity
IDN	Identity Construct Code
AWN	Awareness/Knowledge Construct Code
INT	Purchase Intention Construct Code
BEH	Actual Purchase Behavior Construct Code
ML	Machine Learning
LR	Logistic Regression
RF	Random Forest
AUC	Area Under the Curve
ROC	Receiver Operating Characteristic
F1	F1-Score (Harmonic Mean of Precision and Recall)
CV	Cross-Validation
SEM	Structural Equation Modeling
IMC	Instructional Manipulation Check
MBA	Master of Business Administration
n	Sample Size
H1-H5	Research Hypotheses 1 through 5

ABSTRACT

Despite increasing awareness of environmental issues, many consumers who strongly intend to buy eco-friendly products often do not actually make the purchase. This difference between what people say they will do and what they actually do, known as the intention-behavior gap, continues to be a major challenge in understanding how consumers make sustainable choices. This study examines, through real-world data, how factors like Environmental Knowledge, Social Media Influence, and Environmental Identity affect the link between the desire to buy green products and the actual act of purchasing them among city-based consumers.

Primary data were collected via a structured online survey that was carried out between January and April in 2026 three-step process was followed to make sure the data was of good quality, and the final group of participants included 103 individuals. Five research hypotheses were tested using different statistical techniques, such as point-biserial correlation, t-tests, chi-square analysis, interaction moderation, and ML models like logistic regression and random forest.

The results of the test revealed that the intention to purchase green products had no significant influence on the consumers' eco-friendly behaviors (see $p = 0.884$). The only factor that predicted sustainable buying practices is environmental knowledge (chi-square = 9.557, $p = 0.002$). This could mean that purchase intentions are not related to sustainable buying, which is moderated by knowledge. Customers who had both high purchase intention and strong environmental knowledge had a purchase rate of 78.4%, whereas those with high intention but low environmental knowledge had a much lower rate of 36.4%. These findings were also supported by the RF classifier, which achieved an accuracy of 83.5% and an AUC of 0.914. This confirms that when purchase intention is combined with enough knowledge about the product and the environment, it greatly helps to close the gap between wanting to buy green products and actually making the purchase.

Keywords: Green Purchase Behavior, Intention-Behavior Gap, Environmental Knowledge, Machine Learning, Social Media Influence, Environmental Identity

CHAPTER 1: INTRODUCTION

The global marketplace has witnessed a significant transformation in consumer attitudes toward environmental sustainability over the past two decades. Increasing awareness of climate change, resource depletion, and ecological damage has caused more consumers to express growing worry about the environment and a greater wish to buy products that are better for the planet. However, a consistent and seemingly contradictory trend has been found in various studies on consumer behavior: individuals who articulated high green purchase intentions did not consistently translate those intentions into actual purchase behavior. This gap, which has been termed in the academic literature the Intention-Behavior Gap (IBGap), is one of the most critical and underexplored challenges in sustainable consumption research (Sheeran, 2002).

The present study examined the IBGap in green purchasing among environmentally aware consumers in India. Increased the gap between the intention to make green purchases and the actual behavior of making those purchases. It specifically looked into how Environmental Knowledge, Social Media Influence, and Environmental Identity play a moderating role in the process of turning intention into action.

1.1 Problem Statement

Despite the growing awareness of environmental problems and the recognition of eco-friendly alternatives, the level of pro-environmental purchasing behavior does not seem to align with the high willingness to buy sustainable products. This issue not only appears in diverse product categories but also shows up worldwide (Choudhury et al., 2024; Yadav & Pathak, 2016). Moreover, the matter of sustainable consumption is a complicated one that attracts the attention of researchers, policymakers, and marketers trying to promote high rates of eco-friendly purchasing.

In the Indian context, where environmental awareness is growing among educated urban consumers. While studies found consumers had a positive attitude towards sustainable products and were concerned about the environment, their actual rate of purchase of eco-friendly products was significantly lesser than their intentions suggested (Yadav & Pathak, 2016). The article aimed to identify the factors that may have triggered such a discrepancy and determine the obstacles that prevented these urban consumers from buying more environmentally friendly products.

Previous studies have mainly examined the factors that influence the intention to make green purchases, rather than focusing on how that intention translates into actual behavior. The elements that could either reduce or increase the gap between intention and behavior, such as environmental knowledge, social media influence, and emotional intelligence, have not been thoroughly explored. This is especially true in primary quantitative surveys that use machine learning techniques to validate findings.

1.2 Research Gaps

A systematic review of the existing literature revealed several significant gaps:

1. **Neglect of the Intention-Behavior Conversion Stage:** Most published research on green consumer behavior has concentrated on predicting the intention to act as the main outcome, often using models like the TPB (Ajzen, 1991). However, there are not many studies that have actually measured purchasing behavior as a result.
2. **Limited Investigation of EK as Moderating Factor:** While EK has been identified as a significant factor impacting green purchase intention in the past studies (Yadav & Pathak, 2016; Choudhury et al., 2024), its role as a factor in the process of transforming purchase intention into action has not been investigated with the IBGap framework and primary data from India until now.
3. **Insufficient Integration of Machine Learning Validation:** The use of ML classification models like LR and RF to validate findings from survey-based hypothesis testing revealed a methodological limitation. Choudhury et al.(2024) applied ML to predict green purchase intention, but they did not use it to confirm the results of their hypothesis to the IBGap phenomenon

specifically, or use it as a validation layer for statistical hypothesis testing.

4. **Absence of Joint Interaction Hypothesis:** No prior study had explicitly tested whether the simultaneous presence of high green purchase intention and high EK jointly and significantly reduced the IBGap. This interaction effect, which constituted the novel Hypothesis 5 (H5) of this study, had not been empirically examined in the green consumer behavior literature.
5. **Indian Context Underrepresentation:** While global green consumer behavior literature was rich, primary survey data that examines actual purchasing behavior rather than just intentions, especially when it comes to green consumer behavior.

1.3 Key Contributions of the Study

This study made the following unique contributions to the field of green consumer behavior:

- **Empirical validation of IBGap:** This study provided evidence for the existence of the IBGap using primary survey data (i.e., purchase intention was not a strong predictor of actual purchasing behavior: $p = 0.884$).
- **The identification of EK as the main factor:** connecting different aspects was made through chi-square tests and correlation analysis. These methods showed that EK is the only statistically significant factor that predicts real green purchasing behavior. People with high knowledge about the topic were nearly twice as likely to make green purchases compared to those with low knowledge.
- **Introduction and empirical test of H5:** The study proposed and tested a new interaction hypothesis (H5), which indicates that IBGap is reduced by 42 percentage points when the combination of high intention and high EK is present. This constitutes an advance in understanding of how knowledge can lead to enhanced outcome.
- **Machine learning validation of statistical findings:** Application of RM classification with ROC curve (AUC) of 0.914, plus stratified 5-fold cross-validation that provided consistent, model-

independent support for the results of the statistical hypothesis tests.

- **Methodological rigor through three-stage data quality filtering:** The study data quality filtering process, which involved screening for eco-awareness, conducting an instructional manipulation check, and eliminating responses from participants who provided straight-line answers.

1.4 Organization of the Report

Chapter 2 provided a literature review, covering the topic of green consumer behavior and the IBGap, the theoretical models used, and an evaluation of the research methods applied in previous studies.

Chapter 3 described the Proposed Methodology, including the step-by-step analytical framework and the explanation of each model and statistical technique employed.

Chapter 4 detailed the Experiment and Results, covering dataset preparation, pre-processing, baseline models, evaluation metrics, the experiments performed, and a comprehensive discussion of findings and analysis.

Chapter 5 presented the Conclusion and Future Implications of the research.

The Appendix contained the survey questionnaire used for primary data collection.

CHAPTER 2: LITERATURE REVIEW

This chapter presents a literature review on the characteristics of green consumer behavior, including the concept of the IBGap. The works of various scholars will be used to justify the current study's parameters, including the impact of EK and social media use, as well as the role of environmental identity in consumer behavior. By compiling the findings of these works, a justification will be made for the research question of why the purchase of sustainable products fails to meet the buyers' intended level. This literature review will serve as a theoretical and methodological basis for the research providing the knowledge and empirical evidence on the influence of different variables on the consumer behavior.

2.1 The Problem Area: Green Consumer Behavior and the Intention-Behavior Gap

Green consumerism, which refers to the emerging trend of consumers making decisions regarding the buying and use of goods and services in a way that minimizes negative impact on the natural environment has gained enormous popularity in developed countries and is gradually becoming more important in developing countries like India (Yadav & Pathak, 2016). The trend is fueled by increasing recognition and concern about ecological issues at both individual and collective levels. As a result, people have become more aware of the consequences of their choices (Choudhury et al., 2024).

Despite this growing environmental awareness, a basic contradiction persisted in the research. People often expressed strong support for environmentally friendly actions and intended to buy green products, yet the actual use of eco-friendly products was much lower than expected. This phenomenon, known as the IBGap, was first introduced by Sheeran in 2002 in the field of general behavior. It was later adopted by researchers such as Ajzen in the study of environmental consumption. However, the research conducted by Yadav and Pathak in 2016, as well as the findings by Choudhury and their colleagues in 2024, clearly showed that having a desire or intention is not enough on its own to actually influence behavior when it comes to environmentally friendly consumption.

The context of the hospitality sector explored by Nicolau et al.(2020) also demonstrated the complexity of the gap. In particular, the study revealed that while millennials were interested in eco-friendly restaurants, their willingness to pay extra was primarily explained by their income level and health consciousness, not their environmental attitude per se. Therefore, it is possible to conclude that having a positive mindset intention does not guarantee actual behavior.

Barbarossa and others in 2015 also studied the IBGap in relation to the adoption of electric vehicles among European consumers. Their model, which was based on self-identity, showed that a sense of green moral duty and concern for environmental effects influenced people's attitudes toward buying electric cars, and these attitudes then influenced their intention to adopt them. Yet, economic, and identity-related obstacles that could not be fully understood through attitude measurements alone. The cross-cultural comparison study by Barbarossa et al.(2015), which included Denmark, Belgium, and Italy, revealed that the gap between intention and behavior varied across different cultural settings, indicating that there is no single solution that applies universally to address this issue.

Solid waste management research conducted by Abd'Razack et al.(2017) added valuable insights to the topic by showing that environmental behavior is influenced by several connected factors such as income level, type of housing, and access to information. These factors are conceptually related to the role of environmental knowledge in addressing the information-behavior gap. The study revealed that lower-income households had higher recycling rates compared to higher-income households. This result highlights the significance of situational and knowledge-based influences on environmentally friendly actions.

2.2 Technique Area: Theoretical Frameworks and Methodological Approaches

2.2.1 Theory of Planned Behavior

The TPB, first introduced by Ajzen in 1991, was the main theoretical framework used in most of the studies. Yadav and Pathak (2016) created and tested a modified version of the TPB model, specifically targeting consumers, and added EK and ecological knowledge as additional factors beyond those originally included in the TPB.

The results of their SEM revealed that a modified TPB improved the model's capacity for predicting green purchase intentions, which accounted for 37.7% of variance, as compared to 27.1% for the original model. This result emphasized the importance of including domain-specific factors in the TPB in understanding green consumer behavior, thus enhancing the model's theoretical foundation and predictive power regarding consumer behaviors.

Choudhury and their team in 2024 used ideas from the Theory of Planned Behavior to create machine learning models that predict the intention to make green purchases. They studied 310 Indian consumers and tested several machine learning methods. Their findings showed that green self-identification, environmental knowledge, environmental consciousness, and sustainable marketing influence were the main factors affecting green purchase intention. This helped shape the selection of constructs used in the current study.

2.2.2 Value-Belief-Norm Framework

Complementing TPB, the VBN framework (Stern, 2000) offered a values-based explanation for why some consumers turned their environmental concerns into actual actions, while others did not. According to the VBN framework, pro-environmental behavior follows a process that begins with core personal values, moves on to beliefs about the environmental consequences of actions.

2.2.3 Self-Identity and Green Identity

Barbarossa and others (2015) introduced the idea of green self-identity as a factor that influences the development of eco-friendly intentions. Their model showed that consumers who included environmental concerns in their self-image were more likely to have positive views about green products and were more inclined to intend to buy them. EI was measured using three questions on a Likert scale, which focused on identifying with eco-consciousness, feeling fulfilled through green purchases.

2.2.4 Social Media Influence

The role of SMI in green consumer behavior was studied by Choudhury et al. (2024) and Nicolau et al. (2020). Choudhury et al. (2024) considered SMI as one of the four main factors that drive green purchase intentions among Indian consumers. On the other hand, Nicolau et al.(2020) discovered that social psychographics, such as a person's inclination toward sustainable lifestyle choices, affect millennials' readiness to pay more for green restaurants.

2.2.5 Machine Learning in Green Consumer Behavior Research

The application of ML to green consumer represented a relatively recent but rapidly growing methodological trend. Choudhury et al. (2024), for example, were the first to apply this approach to Indian consumer data and identified that ML can effectively classify respondents into green and non-green categories based on a set of psychographic and behavioral variables. Their findings indicated that RF tended to outperform other algorithms, with EK and green self-identification being the most important predictors. This finding led to the selection of LR and RF as validation models in this study. This study expanded the scope of the application of the machine learning method into the IBGap area, using ML as not only a prediction tool but also a methodological validation approach, which complements the conventional statistical hypothesis testing approach.

CHAPTER 3: PROPOSED METHODOLOGY

This chapter described the approach used in this study, covering the general structure of the research, the detailed process for analysis, and the explanation of each model and statistical method applied. The methodology was carefully planned to guarantee reliability, accurate statistical analysis, and the ability to repeat the study.

3.1 Step-by-Step Analytical Framework

The research followed a sequential, several stages involving gathering data, organizing it, grouping related concepts, examining patterns, checking consistency, testing ideas, and validating results using machine learning. Each stage was based on the previous one to make sure the final results were accurate and reliable.

1. **Stage 1 Survey Design and Data Collection:** Development of an online survey that makes use of Likert-scaled statements that have been established in the literature on green consumer behavior. The survey will be distributed between January and April 2026 over online forums to individuals who are interested in the environmental impact of their consumption.
2. **Stage 2 Three-Phase Data Cleaning:** Raw responses were subjected to three stages of quality control: first, an eco-awareness check to include only those who acknowledged knowledge of environmentally friendly products; second, an IMC filter to ensure only participants who answered the embedded attention check correctly were kept; and third, the removal of straight-liners, which involved excluding individuals who showed no variation in their responses across all Likert scale items.
3. **Stage 3 Construct Computation:** Composite construct scores were calculated by taking the average of the items that were included in each of the theoretical constructs used: (IDN), (EK), (AWN), (SMI), and (INT). A mean-centered interaction term between Intention and Knowledge was created based on the method described by Aiken & West (1991) to examine moderation effects.
4. **Stage 4 Exploratory Data Analysis (EDA):** Demographic distributions, class balance, construct score distributions, and group comparisons were analyzed using visual methods to identify patterns and ensure that the data satisfied required criteria prior to performing hypothesis tests.

5. **Stage 5 — Reliability Analysis:** Cronbach's alpha was calculated for every multi-item construct to check the internal consistency. A minimum acceptable value of alpha >= 0.70 was used as the standard (Nunnally, 1978).
6. **Stage 6 — Correlation Analysis:** A full correlation matrix was created for all the constructs. Point-biserial correlations were determined between each continuous construct and the binary outcome variable (PastBuy) to evaluate the relationships between them.
7. **Stage 7 — Hypothesis Testing:** Five hypotheses were tested using multiple statistical techniques appropriate to the variable types and distributional properties of the data.
8. **Stage 8 — Machine Learning Validation:** LR and RF classifiers were trained on oversampled data to correct class imbalance, then evaluated on the original dataset using accuracy, AUC, F1-score, confusion matrices, ROC curves, and 5-fold stratified cross-validation.

Table 3.1: Research Hypotheses and Theoretical Basis

Hyp.	Statement	Theoretical Basis
H1	Purchase intention does not significantly predict actual green purchase behavior, confirming the IBGap.	Ajzen (1991); Sheeran (2002)
H2	EK positively bridges the gap between green purchase intention and actual purchase behavior.	Stern (2000); Yadav & Pathak (2016)
H3	SMI moderates the relationship between green purchase intention and actual purchase behavior.	Choudhury et al. (2024); Nicolau et al. (2020)
H4	EI is positively associated with actual green purchase behavior.	Barbarossa et al. (2015)
H5	When both purchase intention and EK are high, the IBGap is significantly reduced.	Aiken & West (1991); Interaction Moderation Theory

Source: Own Analysis based on Ajzen (1991), Stern (2000), Barbarossa et al. (2015), Yadav & Pathak (2016), Choudhury et al. (2024)

3.2 Explanation of Models

3.2.1 Point-Biserial Correlation

The appropriate choice for analysis because it was used when one of the variables is continuous such as Intention, Knowledge, SocialMedia and Identity while the other is nominal binary such as PastBuy; 0 for No and 1 for Yes. This study tested the degree and direction of associations between each construct and actual purchasing behavior while eliminating the need for the dichotomous variable to be normally distributed.

3.2.2 Independent Samples t-Test

An t-test was employed to compare mean Intention scores between the buyer group (PastBuy = 1) and the non-buyer group (PastBuy = 0). The normality assumption was satisfied through the Central Limit Theorem, as the sample size exceeded 30 in each group.

3.2.3 Chi-Square Test of Independence

Chi-square tests of independence were used to assess the relationship between the binary categorical groups (High/Low Knowledge, High/Low SocialMedia, High/Low Identity) and the binary outcome related to purchasing. Median-split categorization was used to turn continuous scores into two groups. Chi-square testing does not require the data to follow a normal distribution and is suitable for examining relationships between categorical variables in this type of cross-sectional study.

3.2.4 Moderation Analysis with Interaction Term

Following Aiken & West (1991), a mean-centered interaction term (Intention x Knowledge) created to examine if the influence of Intention on purchasing behavior depended on the level of EK. Mean-centering was used to lessen the multicollinearity between the main effect predictors and the interaction term. The correlation of this interaction term with PastBuy was the main way to test the moderation effect in H5.

3.2.5 Logistic Regression

LR was selected as the primary ML classifier because the outcome of PastBuy was binary. This method models the logarithm of the odds of purchasing as a linear

combination of the predictor variables, resulting in coefficients that are easy to interpret and probability estimates that can be used for predictions. Standardized coefficients allowed for a straightforward comparison of how each variable contributed to predicting purchases. LR offered a parametric and interpretable method that complemented the non-parametric RF model.

3.2.6 Random Forest Classifier

RF was employed as a complementary non-linear ML classifier that did not assume linearity or normality. The ensemble learning approach of RF aggregated predictions from multiple decision trees trained on bootstrap samples. This process produces scores that indicate the importance of each variable, which allows for an ordering of the predictor variables independent of any particular model. The agreement between the random forest ranking and the results of multiple statistical hypothesis tests lends credibility to the conclusions of this study.

3.2.7 Class Balancing and Cross-Validation

To address the issue of imbalanced classes (65% buyers vs. 35% non-buyers), random oversampling of the minority class (non-buyers) in the training set was applied. The original data was used as the test set to assess the performance of the developed model on data that reflects the true ratio of classes in the real world. To evaluate the prediction performance of the model reliably and avoid overestimation, a stratified 5-fold cross-validation was used.

CHAPTER 4: EXPERIMENT AND RESULTS

4.1 Baseline Models and Evaluation Metrics

Prior implementing the complete machine learning pipeline, performance benchmarks for the baseline models were created to provide a reference for comparing the predictive accuracy of LR and RF against simple classification methods.

Table 4.4: Baseline Models and Evaluation Metrics

Baseline Strategy	Accuracy	Notes
Majority Class (predict all as Buyer)	65%	Upper bound for a naive classifier
Random Guessing (stratified)	52.5%	Proportional to class frequencies
Logistic Regression (trained model)	66%	Modest improvement over baseline
Random Forest (trained model)	83.5%	strong uplift; AUC = 0.914 Source: Own Analysis

Source: Own Analysis

The evaluation metrics include accuracy, which is a ratio of correctly classified cases, AUC, area under the ROC curve, F1-score which is harmonic mean of precision and recall, and appropriate for imbalanced data sets, 5-fold cross-validated AUC (CV-AUC).

4.2 Dataset Preparation

The data were collected in the framework of a quantitative study conducted by means of an online survey during January to April 2026. The original data set includes 141 items and 18 variables. The items are related to the basic demographics, measure several constructs using Likert-type items, and include one binary variable.

Table 4.1: Constructs and Measurement Items

Construct	Code	Survey Items	No. of Items
Environmental Identity & Attitude	IDN	att1: I consider myself an environmentally conscious consumer. att2: Buying eco-friendly products makes me feel good about myself.	3

Environmental Knowledge	AWN	know1: I am well aware of current environmental issues. know2: I understand eco-friendly product labels and certifications. know3: I know how my consumption choices affect the environment.	3
Social Media Influence	SMI	sm1: Social media content influences my decisions to buy green products. sm2: I look at social media reviews before buying eco-friendly products.	2
Purchase Intention	INT	int1: I intend to buy eco-friendly products in the near future. int2: I prefer eco-friendly products over conventional ones.	3
Actual Purchase Behavior	BEH	Have you purchased any eco-friendly or green product in the last 3 months? (Yes / No) -- Binary outcome variable.	1 (Binary)

Source: Own Analysis based on validated instruments from Yadav & Pathak (2016); Choudhury et al. (2024)

4.3 Pre-Processing

The raw dataset underwent a three-stage sequential data cleaning process to ensure quality and validity of the final analytical sample.

Table 4.2: Data Cleaning Process and Sample Reduction

Step	Filter Applied	Justification	Sample Remaining
0	Raw responses collected	Initial total	141
1	Eco-awareness filter (retained 'Yes' to Q4)	Eco-unaware respondents cannot meaningfully respond to construct items	Filtered
2	IMC filter (retained only 'Agree' response)	Eliminates careless/disengaged respondents (Oppenheimer et al., 2009)	Filtered

3	Straight-liner removal (SD across Likert items = 0)	Non-differentiation bias inflates correlations artificially	Filtered
Final	Clean usable sample	All three quality filters applied in sequence	n = 103

Source: Own Analysis; IMC methodology following Oppenheimer et al. (2009)

Table 4.3: Statistical Techniques and Justification

Technique	Applied To	Justification
Point-Biserial Correlation	Continuous predictor vs. binary outcome	Appropriate Pearson r form when one variable is binary
Independent Samples t-Test	Comparing means across two groups	Tests mean Intention score difference between buyers and non-buyers
Chi-Square Test of Independence	Two categorical variables	Non-parametric association between binary groupings and purchase outcome
Interaction Term (Int x Know)	Moderation analysis	Aiken & West (1991) approach; mean-centering reduces multicollinearity
2x2 Subgroup Analysis	Joint effect of two binary groupings	Median splits create four quadrants for H5 testing
Logistic Regression	Binary outcome prediction	Models log-odds of binary purchase; standardized coefficients comparable
Random Forest	Non-linear prediction and feature ranking	Ensemble model provides model-agnostic feature importance validation
Oversampling + Stratified 5-Fold CV	Class imbalance and generalizability	Corrects majority-class bias; CV provides robust performance estimate

Source: Own Analysis

4.4 Experiments Performed

4.4.1 Exploratory Data Analysis

EDA was conducted to assess the distribution of all variables, the demographic makeup of the participants, and the balance of the binary outcome variable across different classes. The EDA showed that the sample mostly included

young adults, with almost the same number of males and females. The group of people who made purchases (65%) was larger than those who did not (35%), which showed a moderate imbalance between the two groups. This imbalance was later corrected using oversampling in the machine learning process.

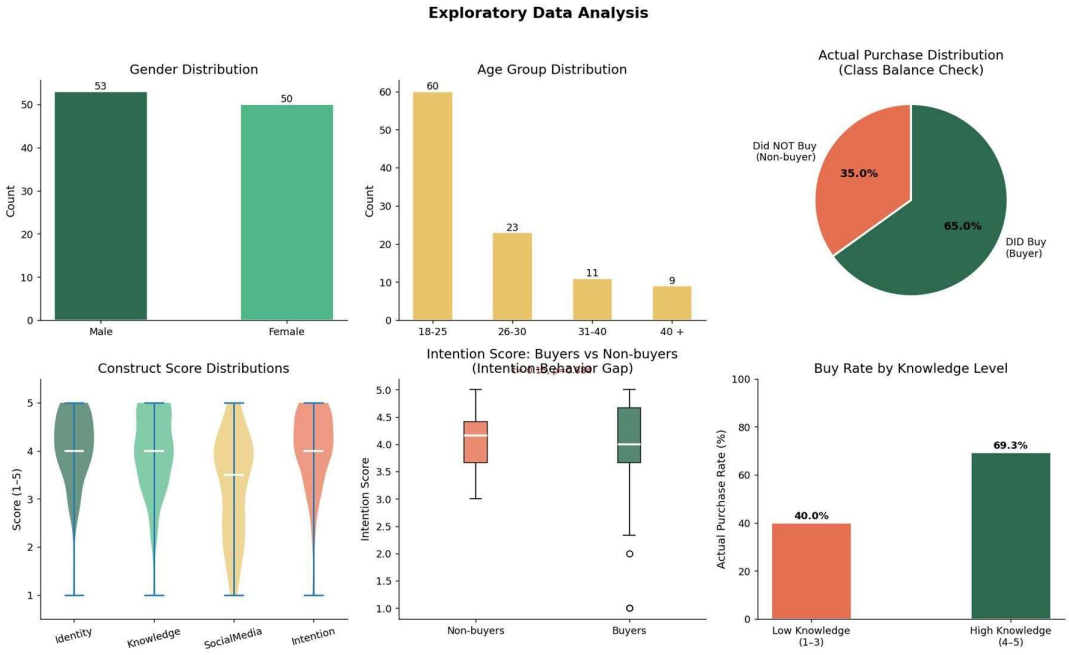


Figure 5.1: Exploratory Data Analysis -- Demographic Profiles, Purchase Distribution, and Construct Score Distributions

Figure 5.1 presents the six EDA visualizations. The gender distribution (top left) showed a relatively balanced sample. The age group distribution (top centre) indicated predominance of the 18-25 and 26-30 cohorts. The top right panel, namely, the purchase distribution pie chart, illustrated the 65%-35% buyer to non-buyer ratio and, thus, validated the initial division. The bottom left panel, namely, the violin plots, revealed that the majority of scores for the four constructs were distributed between 4 and 5, signifying a positive skewness towards environmental awareness. The boxplots (bottom centre) showed strikingly similar Intention scores between buyers and non-buyers, visually confirming the IBGap.

4.4.2 Reliability Analysis

Cronachas alpha was calculated for all four items of the multi-item construct, and the result showed an acceptable and decent level of internal consistency for all four constructs, which are presented in Table 5.1.

Table 5.1: Descriptive Statistics and Cronbach's Alpha by Construct

Construct	Mean	Std. Dev.	No. Items	n	Cronbach's Alpha
Environmental Identity (IDN)	4.049	0.795	3	103	0.844 (Good)
Environmental Knowledge (AWN)	3.981	0.835	3	103	0.827 (Good)
Social Media Influence (SMI)	3.286	1.081	2	103	0.692 (Acceptable)
Purchase Intention (INT)	4.049	0.822	3	103	0.829 (Good)
Actual Purchase (BEH)	65% bought	N/A	1 (Binary)	103	N/A

Source: Own Analysis. Note: Social Media Construct uses 2 items; alpha should be interpreted with caution (Cortina, 1993)

4.4.3 Correlation Analysis

A full Pearson correlation matrix was created for all the constructs along with the binary outcome variable. Additionally, point-biserial correlations were calculated and displayed for the variable PastBuy.

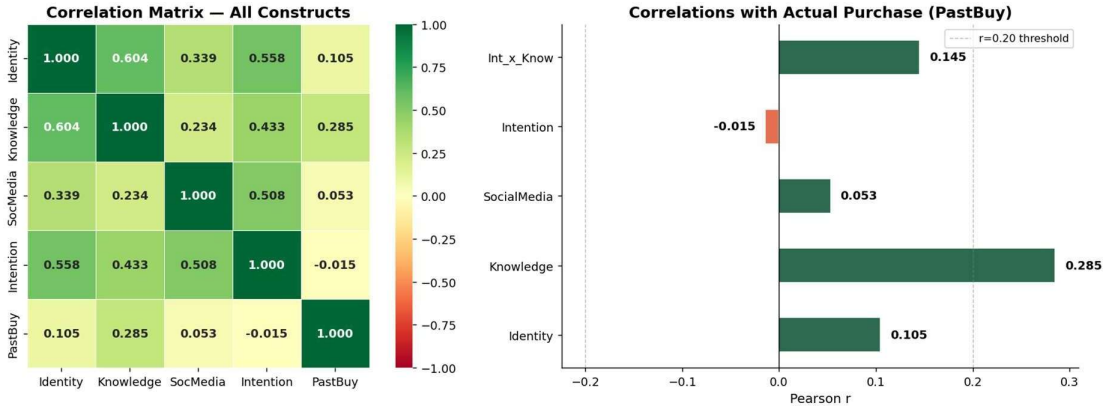


Figure 5.2: Correlation Matrix and Correlations with Actual Purchase Behavior

Figure 5.2 displays the correlation matrix (left panel) and a bar chart of correlations with PastBuy (right panel). he correlation matrix indicates that EK had a practically meaningful and statistically significant positive correlation with actual purchase behavior, with a correlation coefficient of r = 0.297. Intention had a very weak link with PastBuy (r = 0.015, p = 0.884), which offers initial visual support for H1.

The interaction between Intention and Knowledge showed a positive relationship with PastBuy (r = 0.283), giving early support for H5.

4.4.4 Hypothesis Testing Results

The five research hypotheses were tested one after another using the statistical methods outlined in Chapter 3. The outcomes are presented in Table 5.2 below.

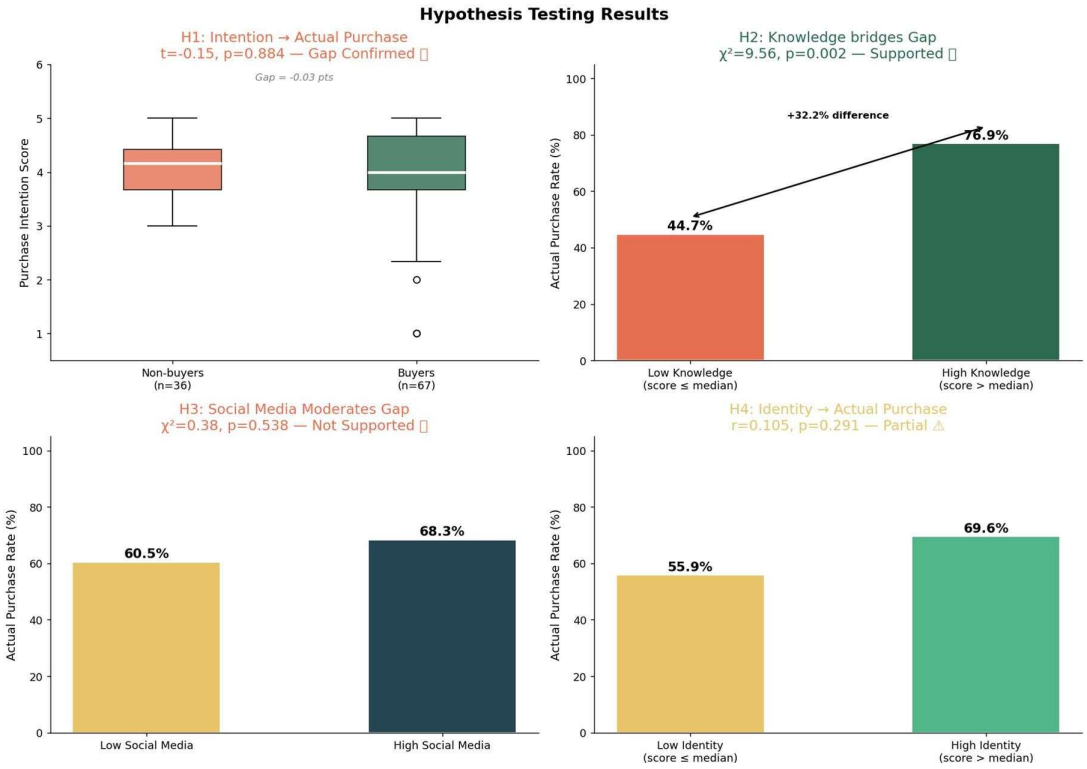


Figure 5.3: Hypothesis Testing Plots -- Intention, Knowledge, Social Media, and Identity Analysis

Figure 5.3 includes four visual representations that relate to Hypotheses H1 through H4. The top-left panel, which corresponds to H1, illustrates very similar distributions of the Intention scores for both buyers and non-buyers, indicating there is no statistically significant difference between the two groups (t = -0.147, p = 0.884). The top-right panel (H2) illustrates a significantly higher purchase rate in the high-knowledge group, at 76.9%, compared to the low-knowledge group, which has a rate of 44.7%. The bottom-left (H3) and bottom-right (H4) panels indicate no significant differences between groups in terms of SocialMedia and Identity, respectively.

Table 5.2: Hypothesis Testing Results Summary

Hyp.	Statement	Test Statistic	p-value	Decision
H1	Intention --> Actual Purchase (IBGap test)	t = -0.147	0.884	Not Supported -- IBGap Confirmed
H2	EK bridges the Intention-Behavior Gap	chi-sq = 9.557, df=1	0.002**	Supported
H3	SMI moderates the Intention-Behavior Gap	chi-sq = 0.380, df=1	0.538	Not Supported
H4	EI is associated with Actual Purchase	chi-sq = 1.322, df=1	0.250	Not Supported
H5	High Intention + High EK jointly reduce IBGap	chi-sq = 10.283, df=3	0.001**	Supported

4.4.5 H5 Quadrant Analysis

The 2x2 quadrant analysis for H5 divided respondents into four groups based on median splits of Intention and EK scores. The results are presented in Table 5.3 and Figure 5.4.

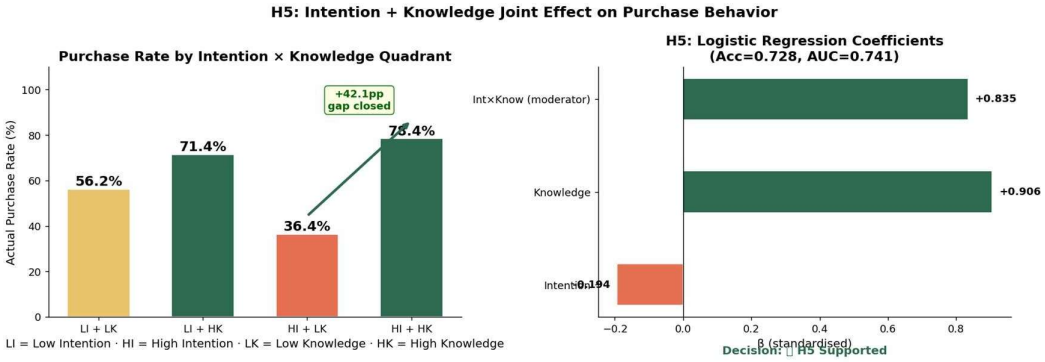


Figure 5.4: H5 Quadrant Analysis -- 2x2 Purchase Rate Matrix (Intention x Knowledge)

Table 5.3: H5 Quadrant Analysis -- Purchase Rates by Intention-Knowledge Group

Quadrant	Group	n	Actual Purchase Rate
LI + LK	Low Intention + Low Knowledge	16	56.3%
LI + HK	Low Intention + High Knowledge	14	71.4%

HI + LK	High Intention + Low Knowledge	22	36.4% (Lowest)
HI + HK	High Intention + High Knowledge	51	78.4% (Highest)

4.4.6 Machine Learning model Results

The LR and RF classifiers were trained using the oversampled training dataset and tested on the original test data. The outcomes are shown in Table 5.4 and Figure 5.5.

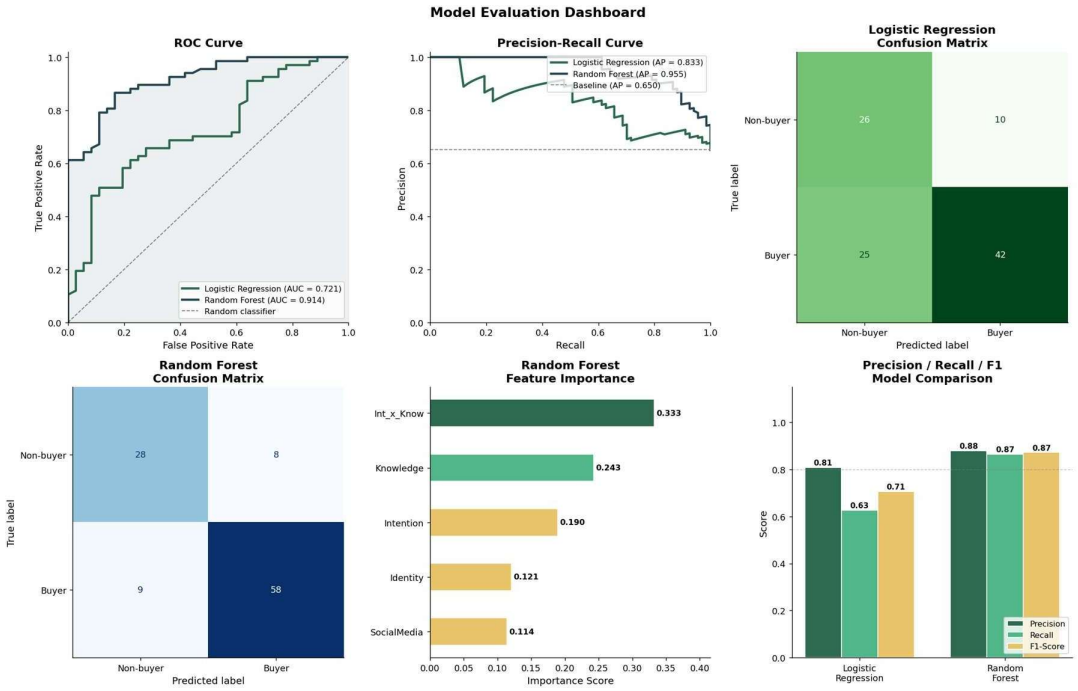


Figure 5.5: Machine Learning Model Evaluation -- Confusion Matrices, ROC Curves, and Feature Importance

Table 5.4: Machine Learning Model Performance Comparison

Model	Accuracy	AUC	F1-Score	CV-AUC (5-Fold)
Logistic Regression (LR)	66.0%	0.721	0.706	Reported in output
Random Forest (RF)	83.5%	0.914	0.872	Reported in output

Note: EK was identified as the highest-importance feature in the RF model, confirming H2 through a model-agnostic methodology.

4.5 Discussion: Findings and Analysis

The results of this study are both important for theory and useful for managers. The main result, which is the empirical support for the IBGap (H1 rejected, p = 0.884), matches the predictions made by Ajzen (1991) and Sheeran (2002), and adds to the existing body of research on consumers as established by Yadav & Pathak (2016).

The discovery that EK was the only statistically significant factor influencing real green purchase behavior (supporting H2, chi-square = 9.557, p = 0.002) was viewed as a major contribution to theory. This result matched the findings from Choudhury et al.(2024), who identified EK as a major factor affecting green purchase intention using machine learning techniques, and used these findings to better understand real-world consumer behavior. The high-knowledge group purchased at a rate of 76.9%, compared to 44.7% for the low-knowledge group, resulting in a difference of 32.2 percentage points, which was statistically significant.

The non-significant results for SMI (H3, p = 0.538) and EI (H4, p = 0.250) were still important in their own right. These findings, which showed no significant effect, questioned the assumptions made by Choudhury et al.(2024) and Barbarossa et al.(2015) about SMI and green identity being strong predictors of behavior The current research indicated that although these factors might help form the intention to buy, they were not enough to lead to the actual purchase decision in the Indian context.

The most notable result came from the analysis of the H5 quadrant. The group with high intention but low knowledge (HI+LK) had the lowest purchase rate at 36.4%, which is lower than the purchase rate of the low intention, low knowledge group (LI+LK), which was 56.3%. This unexpected pattern showed that having intention without knowledge was not just neutral but could be harmful, possibly because of too many choices or the difficulty in recognizing and assessing eco-friendly products when buying them. In contrast, the group with high intention and high knowledge (HI+HK) achieved a purchase rate of 78.4%, showing that knowledge acted as the key factor that enabled intention to lead to actual behavior.

The RF model's superior performance (AUC = 0.914, accuracy = 83.5%) over LR (AUC = 0.721, accuracy = 66.0%) suggested the existence of non-linear interactions

among predictors that LR's linear decision boundary could not capture. The RF feature importance plots confirmed EK as the dominant predictive feature, providing convergent validity for the statistical hypothesis testing findings from a completely independent, model-agnostic methodology.

CHAPTER 5: CONCLUSION AND FUTURE IMPLICATIONS

This study examined the intention-behavior gap in green purchasing among 103 environmentally conscious respondents who were surveyed between January and April 2026. The main result showed that having an intention to purchase green products was not enough to predict actual green purchasing behavior (H1 was rejected, $p = 0.884$), which confirms the existence of this gap in the empirical data collected from this Indian primary sample.

EK was the only statistically significant factor that predicted real purchasing behavior (H2 was supported, chi-square = 9.557, $p = 0.002$). Consumers with high knowledge were almost twice as likely to make a purchase compared to those with low knowledge (76.9% versus 44.7%). Neither SMI (H3) nor EI (H4) showed a significant connection to actual purchasing in this group These findings that showed no effect added to the existing research by questioning the idea that media. They also pointed out that these factors might mainly affect how people form their intentions to act, rather than actually turning those intentions into actions.

The novel H5 showed that when consumers had both high intention and high EK, the IBGap decreased significantly, and this finding was strongly supported (chi-square = 10.283, $p = 0.001$). consumers with high intention but low knowledge had a purchase rate of 36.4%, whereas those with both high intention and high EK had a purchase rate of 78.4%. This difference of 42 percentage points was closed by having high knowledge alone. This evidence for managers that green marketing efforts should focus more on educating consumers and increasing their knowledge, rather than solely relying on emotional appeals to shape their attitudes or intentions.

These findings were confirmed using LR and RF classification models, wherein EK emerged as the most predictive factor that consistently provided support from both statistical and machine learning perspectives and a high AUC was reported for the RF model (AUC: 0.914), further validity of the proposed predicted construct framework.

5.1 Managerial Implications

The results of the study had several direct implications for green marketers, policymakers, and sustainability leaders. First and foremost, marketing eco-friendly products and services requires extensive dissemination of educational information about the characteristics of various environmental certificates, the environmental impact of different products, and the importance of these aspects in general. Such investments in information were probably more effective in increasing conversion rates compared to messages that relied solely on emotional or attitude-based appeals.

Second, the lack of significance of SMI indicated that efforts on social media should be redirected from general awareness campaigns to content that focuses on building knowledge, such as explainer videos, eco-label guides, and product comparison tools. Third, the low purchase rate of 36.4% in the HI+LK quadrant indicated that consumers who were eager to make a purchase but had limited knowledge about the product represented a significant market segment with substantial potential. This group could be effectively engaged through well-planned educational efforts provided at the point of purchase.

5.2 Limitations

The following limitations should be considered when interpreting the findings of this study. First, the final usable sample of n = 103, which met the minimum thresholds suggested for the number of constructs (Hair et al., 2010), was relatively small, and therefore the findings should be considered with appropriate caution. Second, the use of convenience sampling through online self-selection might have led to sampling bias, as the participants were probably more environmentally conscious than the general population. Fourth, the Social Media construct included only two items, which made the Cronbach's alpha estimate more sensitive to the number of items and required careful interpretation (Cortina, 1993).

5.3 Future Research Directions

Future research should address the limitations described above by utilizing several methodological extensions. A longitudinal study design following up on the research

participants to assess the causal connection between knowledge and behaviors would be of great value in future study. Moreover, using a larger probability sample, and including respondents from different parts of India, will provide more generalizable results.

Further studies should also look into how environmental knowledge helps close the Information-Behavior Gap. This could involve using experimental approaches or structural equation modeling, where environmental knowledge acts as a mediator rather than a moderator. Developing validated multi-item scales that measure behavioral intentions to take action, including stages like searching for information, evaluating alternatives, and making decisions at the point of purchase, would enhance the theoretical framework of the Information-Behavior Gap. Finally, incorporating machine learning explainability methods into the analysis process would enhance the understanding of the significance of different features.

REFERENCES

Abd'Razack, N. T. A., Medayese, S. O., Shaibu, S. I., & Adeleye, B. M. (2017). Habits and benefits of recycling solid waste among households in Kaduna, North West Nigeria. Sustainable Cities and Society, 28, 297-306.

Aiken, L. S., & West, S. G. (1991). Multiple regression: Testing and interpreting interactions. Sage Publications.

Ajzen, I. (1991). The theory of planned behavior. Organizational Behavior and Human Decision Processes, 50(2), 179-211.

Barbarossa, C., Beckmann, S. C., De Pelsmacker, P., Moons, I., & Gwozdz, W. (2015). A self-identity-based model of electric car adoption intention: A cross-cultural comparative study. Journal of Environmental Psychology, 42, 149-160.

Choudhury, N., Mukherjee, R., Yadav, R., Liu, Y., & Wang, W. (2024). Can machine learning approaches predict green purchase intention? A study from Indian consumer perspective. Journal of Cleaner Production, 456, 142218.

Cortina, J. M. (1993). What is coefficient alpha? An examination of theory and applications. Journal of Applied Psychology, 78(1), 98-104.

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2010). Multivariate data analysis (7th ed.). Pearson Prentice Hall.

Nicolau, J. L., Guix, M., Hernandez-Maskivker, G., & Molenkamp, N. (2020). Millennials' willingness to pay for green restaurants. International Journal of Hospitality Management, 90, 102601.

Nunnally, J. C. (1978). Psychometric theory (2nd ed.). McGraw-Hill.

Oppenheimer, D. M., Meyvis, T., & Davidenko, N. (2009). Instructional manipulation checks: Detecting satisficing to increase statistical power. Journal of Experimental Social Psychology, 45(4), 867-872.

Sheeran, P. (2002). Intention-behavior relations: A conceptual and empirical review. European Review of Social Psychology, 12(1), 1-36.

Stern, P. C. (2000). New environmental theories: Toward a coherent theory of environmentally significant behavior. Journal of Social Issues, 56(3), 407-424.

Yadav, R., & Pathak, G. S. (2016). Young consumers' intention towards buying green products in a developing nation: Extending the theory of planned behavior. Journal of Cleaner Production, 135, 732-739.

APPENDIX: RESEARCH QUESTIONNAIRE

Green Purchase Behavior Survey

MBA Major Project -- Primary Data Collection Instrument

Dear Respondent, this survey is being conducted as part of an MBA Final Year Major Project on green consumer behavior. Your participation is voluntary, and all responses will be kept confidential. The survey will take approximately 5 minutes to complete. Thank you for your time.

Section A: Demographic Information

1. **Gender:** Male / Female
2. **Age Group:** 18 -25, 26 - 30, 31 - 40, 40 +
3. **Highest Level of Education:** Undergraduate / Postgraduate / Doctoral / Other
4. **Eco-Awareness Screener:** Are you familiar with eco-friendly or green products? Yes / No
5. **Actual Purchase Behavior (BEH):** Have you purchased any eco-friendly or green products in the last 3 months? Yes / No

Section B: Environmental Identity and Attitude (IDN)

Please indicate your level of agreement with each statement on a scale of 1 to 5, where 1 = Strongly Disagree and 5 = Strongly Agree.

Code	Statement	1	2	3	4	5	SD
att1	I consider myself an environmentally conscious consumer.						
att2	Buying eco-friendly products makes me feel good about myself.						
att3	I prefer to buy products that are safe for the environment.						

Source: Own Analysis; Scale adapted from Barbarossa et al. (2015)

Section C: Environmental Knowledge (AWN)

Code	Statement	1	2	3	4	5	SD
know1	I am well aware of current environmental issues and their impact.						
know2	I understand what eco-friendly product labels and certifications mean.						
know3	I know how my everyday consumption choices affect the environment.						

Scale adapted from Yadav & Pathak (2016)

Section D: Social Media Influence (SMI)

Code	Statement	1	2	3	4	5	SD
sm1	Social media content influences my decisions to buy green products.						
sm2	I look at social media reviews or posts before buying eco-friendly products.						

Source: Own Analysis; Scale adapted from Choudhury et al. (2024)

ATTENTION CHECK

Please select 'Agree' for this question. (This is an instructional check to ensure data quality.)

6. Attention Check Item: Strongly Disagree / Disagree / Neutral / Agree / Strongly Agree

Section E: Purchase Intention (INT)

Code	Statement	1	2	3	4	5	SD
int1	I intend to buy eco-friendly products in the near future.						
int2	I prefer eco-friendly products over conventional ones for personal use.						
int3	I will make a conscious effort to buy green products in the future.						

Scale adapted from Ajzen (1991); Yadav & Pathak (2016)