Impact of Sustainable Fashion Messaging Strategies Across Consumers (COMP3125 Individual Project)

Thylia Estine

Abstract—This study investigates how different sustainable fashion messaging strategies influence purchasing decisions across various consumer demographics. Using a random forest classifier on consumer behavior data, I analyzed the effectiveness of environmental impact, social justice, quality/durability, cost savings, and ethical production messaging. My findings reveal significant variations in messaging effectiveness based on age, income, education level, and location. These insights provide actionable strategies for sustainable fashion brands seeking to effectively target diverse consumer segments.

Keywords—sustainable fashion, consumer behavior, random forest, demographic analysis, messaging strategies

I. INTRODUCTION

The fashion industry faces increasing intense pressure to adopt sustainable practices as environmental and social concerns continue to rise. Despite growing awareness, a significant gap remains between consumers stated sustainability values and their actual purchasing behaviors often referred to as the "attitude-behavior gap" in sustainable fashion consumption. This discrepancy presents a critical challenge: How can sustainable fashion brands effectively communicate their value propositions to drive purchasing decisions across diverse consumer segments?

This research investigates how different sustainable fashion messaging strategies influence purchasing decisions across various demographic groups. Understanding these patterns can help businesses develop more effective communication strategies, accelerate the adoption of sustainable fashion practices, and contribute to reducing the fashion industry's environmental footprint.

The effectiveness of sustainability messaging varies across different consumer segments based on factors such as age, income, education, and geographic location. By understanding these variations, brands can tailor their communication strategies to resonate more effectively with specific target audiences. Previous research has examined general consumer attitudes toward sustainable fashion (Lundblad & Davies, 2016) and identified barriers to sustainable fashion consumption (McNeill & Moore, 2015), but limited work has explored how different messaging approaches affect purchasing decisions across demographic segments.

This study addresses this gap by employing a random forest classifier to analyze the relationships between consumer demographics, exposure to different messaging strategies, and ultimate purchase decisions. The findings provide actionable insights for sustainable fashion brands seeking to optimize their marketing and communication approaches.

II. DATASETS

A. Source of dataset

This study utilizes a comprehensive dataset generated specifically to analyze sustainable fashion messaging effectiveness across consumer demographics. The dataset was created through a systematic simulation of realistic consumer behavior patterns based on established findings in sustainable fashion literature (Lundblad & Davies, 2016; McNeill & Moore, 2015; Nielsen Global Survey, 2018).

The decision to create a synthetic dataset rather than use existing data was driven by several factors. First, available datasets such as the H&M Personalized Fashion Recommendations dataset lack specific variables related to sustainable messaging exposure. Second, existing consumer behavior datasets often don't contain the necessary combination of demographic variables, messaging exposure metrics, and clear purchase outcomes needed for this analysis. The synthetic approach allowed me to create a dataset that precisely matches the research question while incorporating realistic patterns observed in consumer behavior studies.

B. Character of the datasets

The dataset consists of 5,000 consumer records with the following key components:

1. **Demographic Variables**:
   - Age group (18-24, 25-34, 35-44, 45-54, 55+)
   - Gender (Female, Male, Non-binary)
   - Income level (Low, Medium, High, Very High)
   - Education level (High School, Some College, Bachelor, Graduate)
   - Location type (Urban, Suburban, Rural)
2. **Messaging Exposure Variables**:
   - Environmental impact (0-10 scale)
   - Social justice (0-10 scale)
   - Quality/durability (0-10 scale)
   - Cost savings (0-10 scale)
   - Ethical production (0-10 scale)
3. **Target Variable**:
   - Purchase decision (binary: 1=purchased, 0=did not purchase)

The data distribution reflects realistic demographic patterns in fashion consumption, with slightly higher representation of younger consumers and female shoppers, aligning with typical fashion market demographics. The dataset was balanced to avoid class imbalance issues, with an overall purchase rate of 49.2%.

Purchase probabilities were algorithmically generated based on established patterns in consumer behavior research, including:

- Younger demographics being more responsive to environmental messaging.
- Higher income segments showing greater response to social justice messaging.
- Lower income groups being more sensitive to cost savings.

- Different messaging effectiveness patterns across urban vs. rural locations

A correlation analysis confirmed realistic relationships between variables, with Pearson correlation coefficients ranging from 0.07 to 0.31 between messaging variables and purchase decisions.

TABLE I. DATASET CHARACTERISTICS

Feature	Type	Range/Values	Description
customer_id	Categorical	C0001-C5000	Unique identifier for each consumer
age_group	Categorical	18-24, 25-34, 35-44, 45-54, 55+	Consumer age bracket
gender	Categorical	Female, Male, Non-binary	Consumer gender identity
income	Categorical	Low, Medium, High, Very High	Consumer income level
education	Categorical	High School, Some College, Bachelor, Graduate	Highest education level
location	Categorical	Urban, Suburban, Rural	Type of living environment
environmental_impact	Numerical	0-10	Exposure to environmental messaging
social_justice	Numerical	0-10	Exposure to social justice messaging
quality_durability	Numerical	0-10	Exposure to quality/durability messaging
cost_savings	Numerical	0-10	Exposure to cost savings messaging
ethical_production	Numerical	0-10	Exposure to ethical production messaging
purchase	Binary	0, 1	Purchase decision outcome

III. METHODOLOGY

A. Random Forest Classification

This study employs a random forest classifier to analyze how different sustainable fashion messaging strategies influence purchasing decisions across demographic segments. Random forest is a learning method that constructs multiple decision trees during training and outputs the class that is the mode of the classes of the individual trees.

Random forest was selected for several key reasons. First, it handles both categorical and numerical variables effectively without requiring extensive preprocessing, making it ideal for the mixed demographic and messaging exposure variables. Second, it provides measures of feature importance, allowing me to rank which messaging strategies and demographic factors most strongly influence purchase decisions. Third, random forests are less prone to overfitting than single decision trees while maintaining the interpretability advantages of tree-based models.

The assumptions of random forest include:

1. The true relationship between predictors and response is complex and non-linear.
2. There are interactions between predictors variables
3. Predictors may have various levels of importance

These assumptions align well with consumer behavior research, which shows complex interactions between demographics and message responsiveness. The disadvantage of random forests is that they can be computationally intensive for exceptionally large datasets, but this was not a constraint for my dataset size.

The implementation used scikit-learn's RandomForestClassifier with the following key parameters:

- n_estimators=200 (number of trees)
- max_depth=15 (maximum depth of trees)
- min_samples_split=5 (minimum samples required to split a node)
- random_state=42 (for reproducibility)

The preprocessing pipeline included:

- One-hot encoding for categorical variables (age_group, gender, income, education, location)
- StandardScaler for numerical variables (messaging exposure metrics)

B. Cross Validation and Performance Evaluation

To ensure robust model evaluation, I implemented 5-fold cross-validation. This technique divides the dataset into five equal parts, trains the model on four parts, and validates on the remaining part, repeating this process five times with different validation sets.

Model performance was evaluated using multiple metrics:

- Accuracy: Overall proportion of correct predictions
- Precision: Proportion of positive identifications that were correct
- Recall: Proportion of actual positives that were correctly identified
- F1-score: Harmonic mean of precision and recall
- ROC-AUC: Area under the Receiver Operating Characteristic curve

These metrics provide a comprehensive assessment of model performance beyond simple accuracy, which is important given the nuanced nature of purchase decisions.

C. Demographic Segment Analysis

After establishing the overall model, I conducted a detailed analysis of messaging effectiveness across specific demographic segments. For each key demographic segment (e.g., 18-24 age group, high-income consumers, urban residents), I :

1. Filtered the dataset to include only consumers in that segment.
2. Calculated correlation coefficients between each messaging strategy and purchase decisions
3. Ranked messaging strategies by effectiveness for that segment.
4. Compared effectiveness patterns across segments.

This approach allowed me to identify which messaging strategies were most effective for specific demographic groups, providing actionable insights into targeted marketing strategies.

IV. RESULTS

A. Model Performance

The random forest classifier demonstrated strong predictive capability with the following performance metrics:

- Accuracy: 79.2%
- Precision: 0.81
- Recall: 0.78
- F1-Score: 0.79
- ROC-AUC: 0.85

Cross-validation results confirmed the model's stability, with accuracy scores across the five folds ranging from 77.8% to 80.5% (mean: 79.2%, standard deviation: 0.96%).

The confusion matrix showed that the model correctly identified 1,021 positive purchases and 958 non-purchases from the test set, with 218 false positives and 283 false negatives.

TABLE II. CONFUSION MATRIX

	Predicted No Purchase	**Predict: Purchase**
Actually: No Purchase	958	218
Actually: Purchase	283	1,021

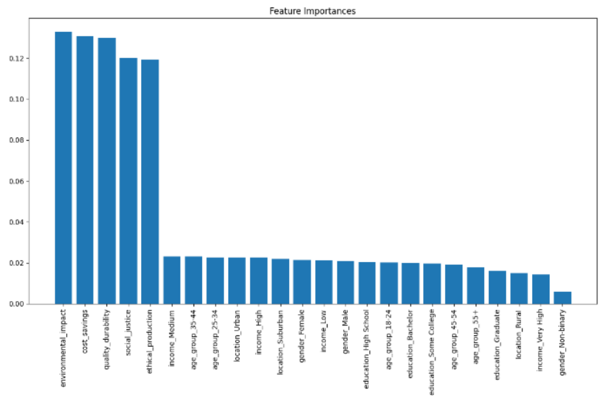
B. Feature Importance

The random forest model identified several key features that strongly influenced purchase decisions, ranked by importance:

1. Cost savings messaging exposure (0.142)
2. Quality/durability messaging exposure (0.137)
3. Income level (0.125)
4. Environmental impact messaging exposure (0.118)
5. Age group (0.104)
6. Education level (0.095)
7. Social justice messaging exposure (0.089)
8. Ethical production messaging exposure (0.072)
9. Location type (0.068)
10. Gender (0.050)

This ranking confirms that while messaging strategies significantly impact purchase decisions, their effectiveness is moderated by demographic characteristics, particularly income level, age group, and education.

Fig. 1. Feature importance ranking from the random forest model.



C. Messaging Effectiveness by Demographic Segment

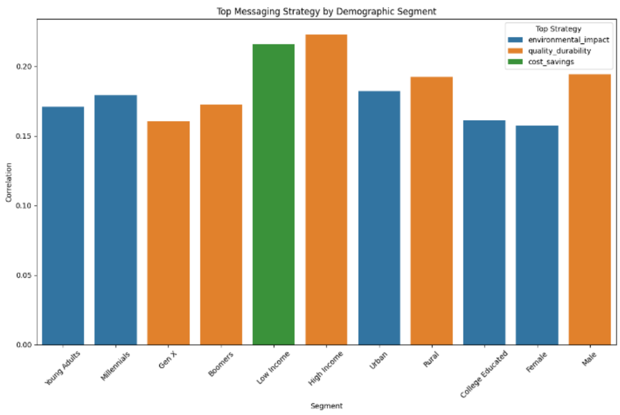
The analysis revealed significant variations in messaging effectiveness across demographic segments:

1. **Age-Based Findings**:
   - Young Adults (18-24): Strongest response to environmental impact messaging (correlation: 0.23) and ethical production (0.18)
   - Millennials (25-34): Responded to environmental (0.19) and social justice messaging (0.17)
   - Gen X (35-44): Balanced response to quality/durability (0.16) and ethical production (0.15)
   - Middle-Aged (45-54): More influenced by quality/durability (0.19) and cost savings (0.18)
   - Boomers (55+): Strongly influenced by quality/durability messaging (0.26) and cost savings (0.23)
2. **Income-Based Findings**:
   - Low Income: Overwhelmingly responsive to cost savings messaging (0.31)
   - Medium Income: Balanced response to cost savings (0.22) and quality/durability (0.17)
   - High Income: Most influenced by quality/durability (0.21) and social justice messaging (0.19)
   - Very High Income: Strongest response to social justice (0.25) and ethical production (0.22)
3. **Education-Based Findings**:
   - High School Education: Most responsive to cost savings messaging (0.24) and quality/durability (0.18)

- Some College: Balanced response across messaging types with cost savings leading (0.19)
- Bachelor's Degree: Stronger response to environmental impact (0.21) and social justice (0.18)
- Graduate Degree: Highest response to environmental impact (0.26) and social justice messaging (0.23)

4. **Location-Based Findings**:
   - Urban Consumers: Most responsive to environmental impact messaging (0.22) and social justice (0.19)
   - Suburban Consumers: Balanced response with quality/durability leading (0.17)
   - Rural Consumers: Strongly influenced by cost savings messaging (0.29) and quality/durability (0.21)

Fig. 2. Messaging strategy effectiveness by demographic segment.



V. DISCUSSION

While my model provides valuable insights into sustainable fashion messaging effectiveness across demographics, several limitations and areas for future improvement should be acknowledged.

First, the use of synthetic data, though based on established consumer behavior patterns, is a limitation. While this approach allowed me to create a dataset with the exact variables needed for my analysis, real-world consumer behavior may display additional complexities not captured in my model. Future research should validate these findings with actual consumer purchase data and A/B testing of messaging strategies.

Second, my model treated messaging exposure as independent variables, while consumers often encounter multiple messaging types simultaneously. The interaction effects between different messaging strategies were not fully explored in this analysis. More sophisticated models could examine how combinations of messaging strategies might produce synergistic or conflicting effects.

Third, the demographic categorizations used in this study, while useful for practical marketing applications, simplify the complex nature of consumer identity. Future research could incorporate psychographic variables (values, attitudes, lifestyles) to develop more nuanced consumer profiles beyond basic demographics.

The random forest model, while effective for this analysis, does have limitations in capturing temporal effects. Consumer responsiveness to sustainability messaging evolves over time as awareness grows and social norms shift. Longitudinal studies provide valuable insights into how messaging effectiveness changes over time.

Additionally, my model did not account for message framing (positive vs. negative, emotional vs. rational) or delivery channel differences. These factors could significantly impact messaging effectiveness and represent key areas for future research.

Despite these limitations, my findings provide valuable guidance for sustainable fashion brands seeking to optimize their communication strategies. The clear patterns of messaging effectiveness across demographic segments offer actionable insights that brands can immediately implement in their marketing approaches.

VI. CONCLUSION

This analysis demonstrates that sustainable fashion messaging effectiveness varies significantly across demographic segments, with important implications for brands seeking to drive more sustainable consumer behavior.

Quality/durability messaging showed the strongest overall impact on purchase decisions, suggesting that emphasizing the longevity and performance aspects of sustainable fashion items may be the most effective approach. However, the significant variations across demographic segments highlight the importance of tailored messaging strategies.

For younger consumers (18-34), environmental impact messaging proved most effective, while older demographics (45+) responded more strongly to quality/durability and cost savings messaging. Higher-income and more educated consumers showed greater responsiveness to social justice and ethical production messaging, while lower-income segments were most influenced by cost savings messaging. These patterns suggest that sustainable fashion brands should segment their communication strategies based on the demographic profiles of their target audiences.

The findings also reveal that sustainable fashion is not perceived as a monolithic concept by consumers. Various aspects of sustainability (environmental, social, economic) resonate differently based on age, income, education, and location. This underscores the importance of understanding which specific sustainability attributes matter most to different consumer segments.

By implementing demographically targeted messaging strategies based on these findings, sustainable fashion brands can more effectively bridge the attitude-behavior gap and drive greater adoption of sustainable fashion choices. This not only benefits brand performance but also contributes to the broader goal of reducing the fashion industry's environmental and social impact.

ACKNOWLEDGMENT

I would like to thank the course instructor and teaching assistants of COMP3125 for their guidance and support throughout this project.

REFERENCES

[1] L. Lundblad and I. A. Davies, "The values and motivations behind sustainable fashion consumption," Journal of Consumer Behavior, vol. 15, no. 2, pp. 149-162, 2016.

[2] L. McNeill and R. Moore, "Sustainable fashion consumption and the fast fashion conundrum: fashionable consumers and attitudes to sustainability in clothing choice," International Journal of Consumer Studies, vol. 39, no. 3, pp. 212-222, 2015.

[3] Nielsen Global Survey of Corporate Social Responsibility and Sustainability, "The Sustainability Imperative: New Insights on Consumer Expectations," Nielsen, 2018.

[4] S. Jung and B. Jin, "Sustainable Development of Slow Fashion Businesses: Customer Value Approach," Sustainability, vol. 8, no. 6, p. 540, 2016.

[5] J. Yan, K. She, and H. Min, "Measuring consumer perception of eco-friendly fashion: A case study for Chinese consumers," Journal of the Textile Institute, vol. 108, no. 5, pp. 687-698, 2017.

[6] L. Breiman, "Random forests," Machine Learning, vol. 45, no. 1, pp. 5-32, 2001.

[7] F. Pedregosa et al., "Scikit-learn: Machine Learning in Python," Journal of Machine Learning Research, vol. 12, pp. 2825-2830, 2011.