

# Predicting E-Commerce Greenwashing Using Random Forest Machine Learning: Legal Analysis and Marketing Management Strategies

Firman Syahputra<sup>1)\*</sup>, Juliya Maria<sup>2)</sup>, Hilda Elsera Br Sembiring<sup>3)</sup>

<sup>1,2)</sup> Universitas Battuta, Medan, Indonesia

<sup>3)</sup> Universitas Satya Terra Bhinneka, Medan, Indonesia

<sup>1)</sup>[24firman@gmail.com](mailto:24firman@gmail.com), <sup>2)</sup>[91juliyahasibuan@gmail.com](mailto:91juliyahasibuan@gmail.com), <sup>3)</sup>[hildaelsera@satyaterrabinneka.ac.id](mailto:hildaelsera@satyaterrabinneka.ac.id)

Submitted : Aug 20, 2026 | Accepted : Sep 14, 2026 | Published : Oct 1, 2026

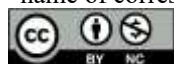
**Abstract:** The rapid growth of green marketing and e-commerce has increased the use of environmental claims in product descriptions, while potentially misleading claims may create consumer confusion and reduce trust. In Indonesia, such claims are also relevant to consumer protection because misleading information may require further legal examination. This study aims to develop and compare Natural Language Processing (NLP)-based machine learning models for the preliminary screening of indications of greenwashing in e-commerce product descriptions and to examine their implications for consumer protection and sustainable marketing management. A dataset of 610 e-commerce product descriptions was classified into *Non-Greenwashing* and *Indication of Greenwashing* categories. Text was represented using TF-IDF unigram and bigram features. Random Forest, Random Forest with Strict Feature Engineering, Support Vector Machine, and Multinomial Naive Bayes were evaluated using an independent stratified test set and leakage-safe five-fold cross-validation. Accuracy, Precision, Recall, F1-score, Macro-F1, Balanced Accuracy, ROC-AUC, and PR-AUC were used for evaluation. Random Forest with TF-IDF achieved the strongest independent-test performance, with 92.62% Accuracy and 95.81% F1-score. In five-fold cross-validation, SVM achieved the highest mean Accuracy ( $95.08 \pm 2.09\%$ ) and F1-score ( $97.13 \pm 1.23\%$ ), while Random Forest with Strict Feature Engineering achieved the highest ROC-AUC ( $96.07 \pm 1.46\%$ ) and PR-AUC ( $99.20 \pm 0.38\%$ ). Predictive-signal analysis identified textual features associated with model decisions. The findings indicate that NLP-based machine learning can support preliminary legal screening and sustainable marketing management. However, predictions should support, rather than replace, human legal and managerial assessment.

**Keywords:** Greenwashing; E-Commerce; Natural Language Processing; Random Forest; Machine learning

## INTRODUCTION

In recent decades, consumer environmental awareness has risen sharply, prompting companies to adopt green marketing to enhance business competitiveness. Numerous global studies and development policies underscore the importance of sustainable consumption (such as the UN's SDG agenda regarding sustainable consumption patterns). Companies have begun promoting eco-friendly practices in advertising materials and on labels; however, not all such claims are truthful. The term "greenwashing" emerged to describe instances where eco-friendly claims are used merely as a marketing tactic, without any genuine commitment. (Javed et al., 2023) found that perceived greenwashing can increase consumer skepticism and negatively affect brand credibility. From the consumer perspective, such perceptions may undermine consumers' confidence in environmental claims and the credibility of the brand. (Steenis et al., 2023) found that when product sustainability claims are inconsistent with actual product characteristics, consumers may perceive themselves as deceived, resulting in reduced purchase intention and brand trust.

\*name of corresponding author



This is anCreative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

The phenomenon of greenwashing, namely false or misleading claims regarding the environmental benefits of products, has attracted increasing attention in the sustainable marketing literature. In response to growing consumer environmental awareness and the rapid expansion of e-commerce, companies are increasingly adopting green marketing strategies. For example, Indonesia's digital economy reached approximately US\$59 billion in 2022, making the country the largest e-commerce market in Southeast Asia. However, this economic opportunity is accompanied by environmental challenges, as approximately 96% of e-commerce packaging consists of single-use plastic. This situation has increased public awareness of plastic waste issues and created potential opportunities for greenwashing practices. Longitudinal evidence also suggests that consumers are becoming increasingly skeptical. For example, (Fatchurrohman et al., 2026) reported that 60% of Indonesian consumers in 2025 did not trust corporate green claims.

The Indonesian context demonstrates similar dynamics. Growing interest in green products is accompanied by regulations that remain relatively general. A normative legal study by (Istanti & Samsul, 2023) indicates that consumers who feel "deceived" by false environmental claims may experience confusion and potential losses. Other legal research emphasizes that, although Indonesia does not yet have specific regulations governing greenwashing, Law No. 8 of 1999 on Consumer Protection, particularly Articles 4 and 7–10, can provide a legal framework for addressing misleading advertising and information practices. Similarly (Ramadhani & Hermawan, 2025) reported that greenwashing practices in Indonesia can increase consumer skepticism, highlighting the importance of applying consumer protection regulations to safeguard consumers. Despite these emerging concerns, recent empirical literature remains limited regarding the automation of greenwashing detection, particularly in e-commerce channels.

The development of Natural Language Processing (NLP) and machine learning provides an opportunity to automatically perform preliminary identification of environmental claim patterns. A systematic review of the relationship between greenwashing, sustainability reporting, and artificial intelligence indicates that the application of AI/NLP for greenwashing detection remains relatively limited and represents an emerging research area (Moodaley & Telukdarie, 2023). More recent research has also demonstrated the potential of AI-supported detection for analyzing ambiguous language in environmental claims (Janice et al., 2024). In the context of text classification, Random Forest is an applicable algorithm that can be combined with TF-IDF representations. Research on text classification has demonstrated that Random Forest can process textual features and provide feature ranking, while other studies have shown that combinations of text-weighting methods and machine learning can be applied to environmental text analysis (Chaudhary et al., 2016; Jalal et al., 2022).

From the perspective of Indonesian law, greenwashing is not merely a marketing issue but is also related to the information provided to consumers. Legal studies in Indonesia indicate that misleading environmental claims can be examined from the perspective of consumer protection and businesses' obligations to provide accurate and truthful information (Istanti & Samsul, 2023; Wongkar & Apsari, 2021). Therefore, an approach is needed that goes beyond text classification and can also provide information regarding the characteristics of potentially misleading environmental claims. Despite the growing attention to greenwashing, several research gaps remain. First, existing studies have largely examined greenwashing from consumer, marketing, sustainability reporting, or legal perspectives, while empirical studies that automatically classify environmental claims in e-commerce product descriptions remain limited. Second, existing AI-related research on greenwashing is still emerging and has not sufficiently connected automated text classification with preliminary consumer-protection and legal analysis in the Indonesian e-commerce context. Third, there is a need for more comprehensive evaluation of classification performance in imbalanced datasets, particularly through class-specific metrics and complementary measures such as Macro-F1, Balanced Accuracy, ROC-AUC, and PR-AUC. These gaps indicate the need for an approach that combines NLP-based classification, interpretable textual signals, and contextual interpretation of environmental claims.

To address these gaps, this study proposes an NLP-based machine learning framework for the preliminary screening of indications of greenwashing in e-commerce product descriptions. The contribution of this study lies in three aspects: (1) developing a classification approach using TF-IDF and machine learning models to distinguish Non-Greenwashing from Indication of Greenwashing; (2) evaluating model performance using an independent test set and leakage-safe five-fold cross-validation with multiple metrics to account for class imbalance; and (3) interpreting predictive textual signals and their implications for preliminary consumer-protection screening and sustainable marketing management. The proposed framework is intended as a decision-support mechanism rather than a substitute for legal judgment.

Based on these issues, this study aims to develop and compare machine learning models for predicting indications of greenwashing in e-commerce products using NLP, identify textual features associated with model predictions, and interpret the findings from the perspectives of consumer protection and sustainable marketing management strategies.

## LITERATURE REVIEW

Greenwashing can be understood as the use of environmental communication or claims that create a positive perception of the environmental attributes of a product or company but are not fully supported by adequate evidence or actual conditions. Greenwashing as a combination of problematic behavior and misleading communication, making the communication aspect an important component in identifying greenwashing (Pizzetti et al., 2021). In marketing, the problem does not arise only when claims are completely false, but also when claims are overly general, ambiguous, or disproportionate to the actual environmental benefits. Research on claim–fact discrepancy indicates that overly general sustainability claims about products that are only partially sustainable may be perceived as deceptive (Steenis et al., 2023; Vieira et al., 2020). Empirical studies have shown that greenwashing influences consumer perceptions by generating skepticism and cynicism, as well as affecting subsequent purchasing behavior (Nyilasy et al., 2013; Sajid et al., 2025).

The research data, consisting of product names and descriptions, are unstructured textual data. Natural Language Processing (NLP) is used to transform such text into a form that can be computationally analyzed through cleaning, tokenization, stopword removal, stemming, and numerical representation. NLP methods such as term frequency and TF-IDF are commonly used for text classification (Jalal et al., 2022). Nevertheless, research specifically focusing on greenwashing detection using AI remains limited (Moodaley & Telukdarie, 2023). Several studies addressing ESG-washing have applied automated analysis to sustainability reports, demonstrating the potential of NLP applications in the environmental domain (Janice et al., 2024; Lagasio, 2024).

TF-IDF assigns weights to terms according to their importance within documents. Text classification studies have demonstrated that TF-IDF can be combined with various machine learning algorithms, including Random Forest. Traditional approaches such as TF-IDF remain widely used because of their simplicity and interpretability (Chaudhary et al., 2016; Jalal et al., 2022). Random Forest was selected because it combines multiple decision trees and provides feature-importance information (Jalal et al., 2022) demonstrated that feature ranking and optimization of the number of trees can improve the performance of Random Forest for text classification. Such methods can highlight the specific "green" terms most indicative of misleading claims. Marketing strategies must take reputational risks into account. Consumers are becoming increasingly discerning (Steenis et al., 2023), 95% of products making green claims employ at least one greenwashing tactic. Companies need to implement authentic green marketing to maintain trust (Nyilasy et al., 2013). The study also suggests consumer literacy training as part of management strategy (Istanti & Samsul, 2023).

Table 1. Summary of Related Studies and Research Gaps

| Author(s) & Year              | Research Focus                                                                                     | Limitations                                                         |
|-------------------------------|----------------------------------------------------------------------------------------------------|---------------------------------------------------------------------|
| (Vieira et al., 2020)         | Systematic review of greenwashing concepts, forms, and typologies                                  | Literature-based; no automated product-level detection              |
| (Moodaley & Telukdarie, 2023) | Greenwashing, sustainability reporting, and AI potential                                           | Conceptual; no e-commerce product classification                    |
| (Nugroho, 2026)               | NLP-based detection of green promises and unverifiable claims in Indonesian sustainability reports | Six reports only; corporate-level data; no TF-IDF–RF classification |
| (Zeng et al., 2025)           | ML-based prediction and interpretation of ESG greenwashing                                         | Firm-level indicators; no product-level textual analysis            |
| (Lagasio, 2024)               | ML-based prediction and interpretation of ESG greenwashing                                         | ESG-oriented; no NLP-based product-claim detection                  |
| (Jalal et al., 2022)          | Random Forest with feature ranking for text classification                                         | Generic text domain; not greenwashing-specific                      |
| (Steenis et al., 2023)        | Consumer responses to sustainability claims and claim–fact discrepancies                           | Consumer-focused; no automated NLP detection                        |
| (Gorovaia, 2024)              | NLP-based greenwashing detection in CSR reports                                                    | CSR reports only; no e-commerce product data                        |

The reviewed studies demonstrate three main research directions. First, conceptual and systematic studies have established the forms and consequences of greenwashing but do not provide automated product-level detection (Vieira et al., 2020; Moodaley & Telukdarie, 2023). Second, recent machine learning and natural language processing studies have demonstrated the potential of automated greenwashing detection, but most applications focus on corporate sustainability, environmental, or corporate social responsibility reports rather than individual e-commerce products (Gorovaia, 2024; Lagasio, 2024; Nugroho, 2026; Zeng et al., 2025). Third, consumer-oriented studies have examined responses to sustainability claims and claim–fact discrepancies without developing

automated classification models (Steenis et al., 2023). These studies indicate that product-level automated screening of greenwashing-related claims in e-commerce remains insufficiently explored.

## METHOD

This study adopts a quantitative experimental approach using machine learning to predict potential greenwashing in e-commerce products. The research integrates Natural Language Processing (NLP), Term Frequency–Inverse Document Frequency (TF-IDF), feature engineering, and Random Forest as the primary classification model. Random Forest is selected for its ability to process combined textual and numerical features and provide feature-importance analysis, while SVM and Multinomial Naive Bayes are employed as benchmark models. The computational results are further interpreted from consumer protection and sustainable marketing perspectives. The overall research methodology is illustrated in Figure 1.

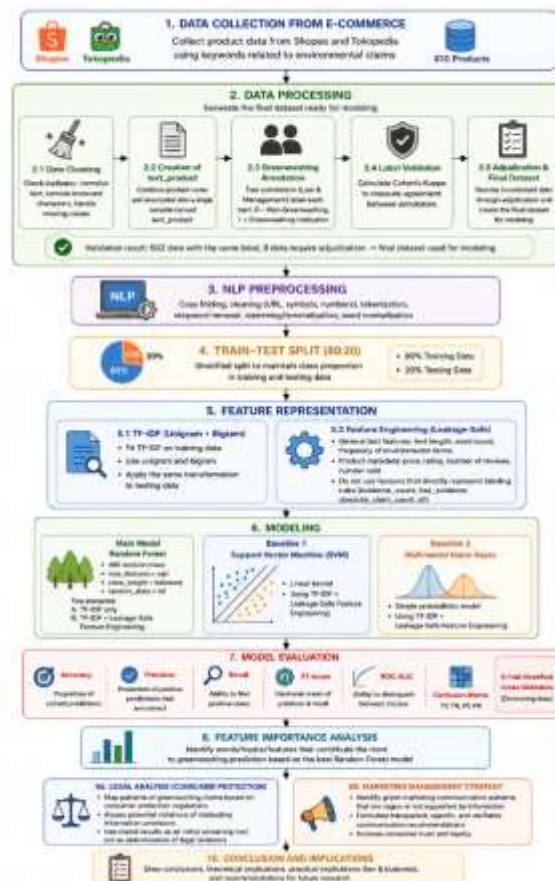


Fig. 1. Research Methodology Flowchart

### Data Collection

The dataset consisted of 610 e-commerce product records collected from Shopee and Tokopedia. Data collection focused on product information containing environmental-related claims. Three search keywords were used to identify relevant products: “ramah lingkungan,” “eco-friendly,” and “ecopack.” The collected records contained product-level information, including product name, product description, category, price, rating, number of reviews, number of products sold, seller information, location, platform, and product link.

The collected products were screened before annotation to ensure that the records were relevant to the research objective. Products were considered for inclusion when their product names or descriptions contained environmental-related information relevant to the study. Records that were duplicated or did not provide sufficient information for the annotation process were excluded during data preparation. After the data preparation and screening process, 610 product records were retained for annotation and subsequent analysis.

### Dataset and Annotation Procedure

The annotation process was conducted to classify product information according to the presence or absence of indications of greenwashing in environmental-related communication. Two annotators independently reviewed the collected product information using predefined annotation guidelines. The annotation focused on the

\*name of corresponding author



This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

environmental claims presented in the product name and description rather than on independently verified environmental performance.

A product was assigned to the Non-Greenwashing class (label 0) when no environmental claim was identified, or when the environmental claim was sufficiently specific and accompanied by identifiable supporting information, such as certification, product composition, measurable material content, or other verifiable information stated in the product information. In contrast, a product was assigned to the Indication of Greenwashing class (label 1) when its environmental claims were potentially misleading, exaggerated, overly general, absolute, ambiguous, or lacked a clear supporting basis in the available product information.

The annotation guidelines explicitly stated that the presence of environmental terms such as “eco,” “green,” “organic,” or “ramah lingkungan” alone was not sufficient to assign the greenwashing indication label. The classification considered the context of the claim, the specificity of the environmental statement, and the availability of supporting information within the product information.

### Inter-Annotator Agreement

To assess the reliability of the annotation process, inter-annotator agreement was measured using Cohen's Kappa (Li et al., 2023). Cohen's Kappa evaluates the agreement between two annotators while accounting for the possibility of agreement occurring by chance

$$\kappa = \frac{P_o - P_e}{1 - P_e} \quad (1)$$

where  $P_o$  represents the observed agreement and  $P_e$  represents the expected agreement by chance.

The annotation involved 610 product records. The two annotators produced 603 identical labels and 7 different labels, corresponding to an observed agreement of 0.98852 (98.85%). The expected agreement by chance was 0.747288 (74.73%). Based on these values, Cohen's Kappa was calculated as 0.954591, indicating a very high level of agreement between the two annotators.

Table 2. Inter-Annotator Agreement Matrix

| Annotator 1 / Annotator 2      | Non-Greenwashing<br>(0) | Indication of Greenwashing<br>(1) | Total |
|--------------------------------|-------------------------|-----------------------------------|-------|
| Non-Greenwashing (0)           | 87                      | 2                                 | 89    |
| Indication of Greenwashing (1) | 5                       | 516                               | 521   |
| Total                          | 92                      | 518                               | 610   |

The agreement matrix shows that 87 products were assigned label 0 by both annotators, while 516 products were assigned label 1 by both annotators. Seven records received different labels and were therefore identified as disagreement cases for further review.

### Disagreement and Adjudication

The seven disagreement cases were reviewed using the predefined annotation guidelines. The review considered the wording, context, specificity, and supporting information in the product description. The final labels were established after the disagreement cases were reviewed. The resulting labels represent indications based on product communication and do not constitute legal determinations of greenwashing.

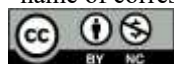
### Final Label Definition

The resulting target variable represented two classes: label 0 for Non-Greenwashing and label 1 for Indication of Greenwashing. The term Indication of Greenwashing was deliberately used to distinguish the machine learning classification from a legal or factual determination of greenwashing. The classification represents an indication derived from the characteristics of environmental communication available in the e-commerce product information. Consequently, the model predictions should be interpreted as preliminary screening signals rather than definitive evidence of deceptive environmental practices.

### Text Preprocessing

The textual information was prepared before feature extraction to reduce irrelevant variation and obtain a consistent representation of the product text. The product name and description were combined to form the textual

\*name of corresponding author



This is anCreative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

input for classification. The preprocessing procedure included text normalization and removal of irrelevant textual elements before the TF-IDF transformation(Wang et al., 2025)

### Train-Test Split

The dataset was divided into 80% training data and 20% testing data using a stratified split. Stratification preserved the class distribution between the two subsets. The testing set was kept completely separate from model training to provide an unbiased assessment of model generalization.

### TF-IDF Feature Representation

The processed text was converted into numerical features using Term Frequency–Inverse Document Frequency (TF-IDF). TF-IDF was used to represent the importance of terms within the documents. Both unigram and bigram features were used to capture individual terms and two-word expressions. TF-IDF remains a commonly used representation for text classification because of its simplicity and interpretability (Tang et al., 2022).

The TF-IDF representation was fitted only on the training data and then applied to the corresponding test data. During cross-validation, the vectorizer was fitted separately within each training fold to prevent data leakage. The TF-IDF formulation is:

$$TF - IDF_{(t,d)} = TF_{(t,d)} \times IDF_{(t)} \quad (2)$$

where:

$$IDF_{(t)} = \log \left( \frac{N}{df_t} \right) \quad (3)$$

TF(t,d) represents the frequency of term  $t$  in document  $d$ ,  $N$  is the number of documents, and  $df_t$  is the number of documents containing term  $t$ .

### Feature Representation

A leakage-safe feature engineering configuration was additionally evaluated to examine whether general textual and product-level characteristics could improve classification performance. The engineered features consisted of non-target-derived textual and product metadata characteristics. Features directly derived from the target label or annotation outcome were excluded from the predictive feature set to prevent target leakage. The feature engineering configuration was evaluated separately from the TF-IDF-only configuration to determine whether the inclusion of additional non-label-derived information improved the classification performance.

### Machine Learning Modeling

Random Forest was used as the primary classifier because it can combine multiple decision trees and provide feature-importance information. Its application to text classification has been demonstrated in recent research (Jalal et al., 2022). Support Vector Machine and Multinomial Naïve Bayes were used as benchmark models. Support Vector Machine is suitable for high-dimensional classification problems, while Multinomial Naïve Bayes is commonly applied to text classification using term-based representations. The dataset was divided into 80% training and 20% testing using stratified sampling. This resulted in 488 training records and 122 test records.

### Model Evaluation

Model performance was evaluated using Accuracy, Precision, Recall, F1-score, Macro-F1, Balanced Accuracy, ROC-AUC, PR-AUC, and confusion matrix. Multiple metrics were used because the dataset had an unequal distribution between the two classes. Recent research recommends considering the characteristics of individual evaluation metrics when assessing classification performance, particularly for imbalanced datasets (Mihelich et al., 2024; Richardson et al., 2024; Syahputra et al., 2026).

The evaluation metrics were calculated as follows:

$$Accuracy = \frac{(TP + TN)}{(TP + TN + FP + FN)} \quad (4)$$

$$Precision = \frac{TP}{(TP + FP)} \quad (5)$$

$$Recall = \frac{TP}{(TP + FN)} \quad (6)$$

\*name of corresponding author



This is anCreative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

$$F1\text{-score} = 2 * \frac{(\text{Presisi} \times \text{Recall})}{(\text{Presisi} + \text{Recall})} \quad (7)$$

where TP represents true positives, TN represents true negatives, FP represents false positives, and FN represents false negatives.

Macro-F1 calculates the average F1-score of all classes with equal weight:

$$\text{Macro} - F1 = \frac{1}{K} \sum_{i=1}^K F1_i \quad (8)$$

For two classes:

$$\text{Macro} - F1 = \frac{F1_0 + F1_1}{2} \quad (9)$$

This metric was used to provide a more balanced assessment of performance across the two classes (Takahashi et al., 2022)

Balanced Accuracy is the average recall of the two classes:

$$\text{Balanced Accuracy} = \frac{\text{Sensitivity} + \text{Specificity}}{2} \quad (10)$$

where:

$$\text{Sensitivity} = \frac{TP}{(TP + FN)} \quad (11)$$

and:

$$\text{Specificity} = \frac{TN}{(TN + FP)} \quad (12)$$

This metric gives equal importance to both classes and is therefore useful for imbalanced datasets (Dewi et al., 2025; Richardson et al., 2024).

ROC-AUC measures the ability of the model to distinguish between the two classes across classification thresholds (Mihelich et al., 2024; Richardson et al., 2024). The ROC curve uses True Positive Rate (TPR) and False Positive Rate (FPR):

$$TPR = \frac{TP}{(TP + FN)} \quad (13)$$

$$FPR = \frac{FP}{(FP + TN)} \quad (14)$$

The ROC-AUC can be expressed as:

$$ROC - AUC = \int_0^1 TPR(FPR) dFPR \quad (15)$$

A higher ROC-AUC indicates better discrimination between the classes.

PR-AUC summarizes the relationship between Precision and Recall across classification thresholds:

$$PR - AUC = \int_0^1 \text{Precision}(\text{Recall}) d\text{Recall} \quad (16)$$

PR-AUC was included as a complementary metric because it focuses on the relationship between precision and recall and can provide additional information when class distributions are unequal (Mihelich et al., 2024; Richardson et al., 2024)

### Feature Importance Analysis

Feature importance was analyzed for the best-performing Random Forest model to identify the textual and numerical features contributing most strongly to classification decisions. This analysis provides insight into the linguistic and product characteristics associated with potential greenwashing (Zeng et al., 2025). However, feature importance is interpreted as predictive contribution rather than causal evidence

### Legal and Sustainable Marketing Analysis

The predicted patterns were further interpreted from consumer protection and sustainable marketing perspectives. The legal analysis examined potentially misleading environmental claims in relation to transparency, truthful information, and consumer protection (Istanti & Samsul, 2023). The model was treated as an early screening tool, not as a legal mechanism for determining whether a seller or company had violated the law (Balqis et al., 2023).

\*name of corresponding author



This is anCreative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

The marketing analysis examined potentially exaggerated or insufficiently supported environmental communication and provided recommendations for more transparent, specific, and verifiable green marketing practices (Steenis et al., 2023; Ur Rahman & Nguyen-Viet, 2023).

## RESULT

### Dataset Characteristics

The dataset used in this study consisted of 610 e-commerce product records collected from Shopee and Tokopedia. The products were selected using environmental-related search keywords, particularly terms associated with ramah lingkungan, eco-friendly, and ecopack. Each product record contained textual product information and selected numerical attributes that were subsequently used for classification. The final dataset consisted of two target classes: Non-Greenwashing (label 0) and Indication of Greenwashing (label 1). Based on the final labeling, 87 products (14.26%) were categorized as Non-Greenwashing, while 523 products (85.74%) were categorized as Indication of Greenwashing. Thus, the dataset exhibited a substantial class imbalance, with the indication-of-greenwashing class representing the majority of observations.

Table 3. Final Class Distribution

| Label | Categori                   | Number | Percentage |
|-------|----------------------------|--------|------------|
| 0     | Non-Greenwashing           | 87     | 14.26%     |
| 1     | Indication of Greenwashing | 523    | 85.74%     |
| Total |                            | 610    | 100%       |

The dominance of the indication-of-greenwashing class is important for interpreting model performance. A model can achieve high overall accuracy by correctly predicting the majority class while performing poorly on the minority class. Therefore, this study evaluates the models using not only Accuracy, Precision, Recall, and F1-score, but also Macro-F1, Balanced Accuracy, ROC-AUC, and PR-AUC.

### Classification Performance

An independent stratified test set containing 20% of the dataset was used to evaluate the generalization performance of the models. The split was performed using a fixed random state of 42 to ensure reproducibility. The final modeling pipeline compared four configurations: Random Forest with TF-IDF, Random Forest with TF-IDF and leakage-safe feature engineering, Support Vector Machine with TF-IDF, and Multinomial Naive Bayes with TF-IDF. The results are presented in Table 4.

Table 4. Classification Performance on the Independent Test Set

| Model                   | Accuracy | Precision | Recall  | F1-score |
|-------------------------|----------|-----------|---------|----------|
| RF + TF-IDF             | 92.62%   | 93.64%    | 98.10%  | 95.81%   |
| RF + TF-IDF + Strict FE | 90.98%   | 93.52%    | 96.19%  | 94.84%   |
| SVM + TF-IDF            | 90.16%   | 94.29%    | 94.29%  | 94.29%   |
| NB + TF-IDF             | 88.52%   | 88.24%    | 100.00% | 93.75%   |

Table 4 shows that RF + TF-IDF achieved the highest accuracy (92.62%) and F1-score (95.81%) on the independent test set, indicating the strongest overall classification performance. SVM + TF-IDF obtained the highest precision (94.29%), whereas NB + TF-IDF achieved perfect recall (100.00%). However, the lower precision and accuracy of Naive Bayes suggest a greater tendency toward false-positive predictions. These results indicate that RF + TF-IDF provides the most balanced overall performance among the evaluated models.

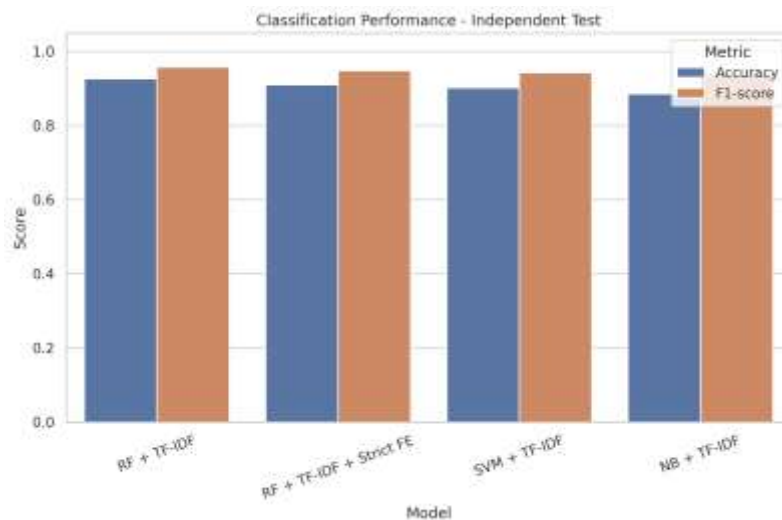


Fig. 2. Performance Comparison Chart

Figure 2 provides a visual comparison of accuracy and F1-score across the four models. RF + TF-IDF shows the highest values for both metrics, with an accuracy of 92.62% and an F1-score of 95.81%. RF + TF-IDF + Strict FE follows with an accuracy of 90.98% and an F1-score of 94.84%, while SVM + TF-IDF and NB + TF-IDF show lower values for both measures. The relatively close F1-scores across the models indicate that all approaches provide effective positive-class prediction, although their overall classification behavior differs, particularly in terms of precision and recall.

### Class-Specific Performance

To provide a more detailed evaluation, the performance of each model was assessed separately for the Non-Greenwashing and Indication of Greenwashing classes using five-fold cross-validation. The results are presented in Table 5.

Table 5. Class-Specific Performance on the Independent Test Set

| Model                   | Class                      | Precision Mean $\pm$ SD | Recall Mean $\pm$ SD | F1-score Mean $\pm$ SD |
|-------------------------|----------------------------|-------------------------|----------------------|------------------------|
| RF + TF-IDF             | Non-Greenwashing           | 92.53 $\pm$ 7.43%       | 64.38 $\pm$ 4.62%    | 75.69 $\pm$ 3.13%      |
| RF + TF-IDF             | Indication of Greenwashing | 94.36 $\pm$ 0.72%       | 99.04 $\pm$ 0.96%    | 96.64 $\pm$ 0.53%      |
| RF + TF-IDF + Strict FE | Non-Greenwashing           | 93.13 $\pm$ 6.68%       | 68.95 $\pm$ 5.21%    | 78.92 $\pm$ 1.41%      |
| RF + TF-IDF + Strict FE | Indication of Greenwashing | 95.05 $\pm$ 0.75%       | 99.04 $\pm$ 0.96%    | 97.00 $\pm$ 0.28%      |
| SVM + TF-IDF            | Non-Greenwashing           | 85.71 $\pm$ 11.01%      | 80.59 $\pm$ 9.16%    | 82.53 $\pm$ 7.06%      |
| SVM + TF-IDF            | Indication of Greenwashing | 96.79 $\pm$ 1.51%       | 97.51 $\pm$ 2.20%    | 97.13 $\pm$ 1.23%      |
| NB + TF-IDF             | Non-Greenwashing           | 100.00 $\pm$ 0.00%      | 35.75 $\pm$ 7.03%    | 52.37 $\pm$ 7.42%      |
| NB + TF-IDF             | Indication of Greenwashing | 90.34 $\pm$ 1.20%       | 100.00 $\pm$ 0.00%   | 94.92 $\pm$ 0.66%      |

As shown in Table 4, all models achieved higher performance for the Indication of Greenwashing class than for the Non-Greenwashing class. SVM + TF-IDF achieved the highest F1-score for Non-Greenwashing (82.53%), while RF + TF-IDF + Strict FE achieved the highest F1-score for Indication of Greenwashing (97.00%). NB + TF-IDF obtained perfect recall for Indication of Greenwashing (100.00%), but its recall for Non-Greenwashing was substantially lower (35.75%). These results indicate that identifying the *Non-Greenwashing* class remains more challenging, highlighting the importance of using class-balanced metrics such as Macro-F1 and Balanced Accuracy for further evaluation.

\*name of corresponding author



This is anCreative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

### Macro-F1 and Balanced Accuracy

Macro-F1 and balanced accuracy were used to evaluate performance more fairly across the two classes. SVM + TF-IDF achieved the highest mean Macro-F1 (89.83%) and balanced accuracy (89.05%). RF + TF-IDF + Strict FE ranked second, with Macro-F1 of 87.96% and balanced accuracy of 84.00%. Naive Bayes obtained the lowest values, 73.64% and 67.88%, respectively. These results confirm that high positive-class recall does not necessarily correspond to balanced classification performance.

Table 6. Macro-F1 and Balanced Accuracy

| Model                   | Macro-F1 Mean $\pm$ SD | Balanced Accuracy Mean $\pm$ SD |
|-------------------------|------------------------|---------------------------------|
| RF + TF-IDF             | 86.17 $\pm$ 1.80%      | 81.71 $\pm$ 2.15%               |
| RF + TF-IDF + Strict FE | 87.96 $\pm$ 0.78%      | 84.00 $\pm$ 2.17%               |
| SVM + TF-IDF            | 89.83 $\pm$ 4.13%      | 89.05 $\pm$ 4.45%               |
| NB + TF-IDF             | 73.64 $\pm$ 4.03%      | 67.88 $\pm$ 3.52%               |

These results indicate that SVM + TF-IDF provides the most balanced performance between the two classes, while the lower scores of NB + TF-IDF suggest a stronger bias toward the majority class. Therefore, Macro-F1 and Balanced Accuracy provide important complementary evidence to the conventional classification metrics.

### Confusion Matrix Analysis

The independent-test confusion matrices provide a direct view of classification errors. RF + TF-IDF produced 10 true negatives, seven false positives, two false negatives, and 103 true positives. RF + TF-IDF + Strict FE produced 10 true negatives, seven false positives, four false negatives, and 101 true positives. SVM produced the highest number of true negatives (11) and the lowest number of false positives (6), while recording six false negatives. Naive Bayes produced zero false negatives and 105 true positives, but only three true negatives and 14 false positives.

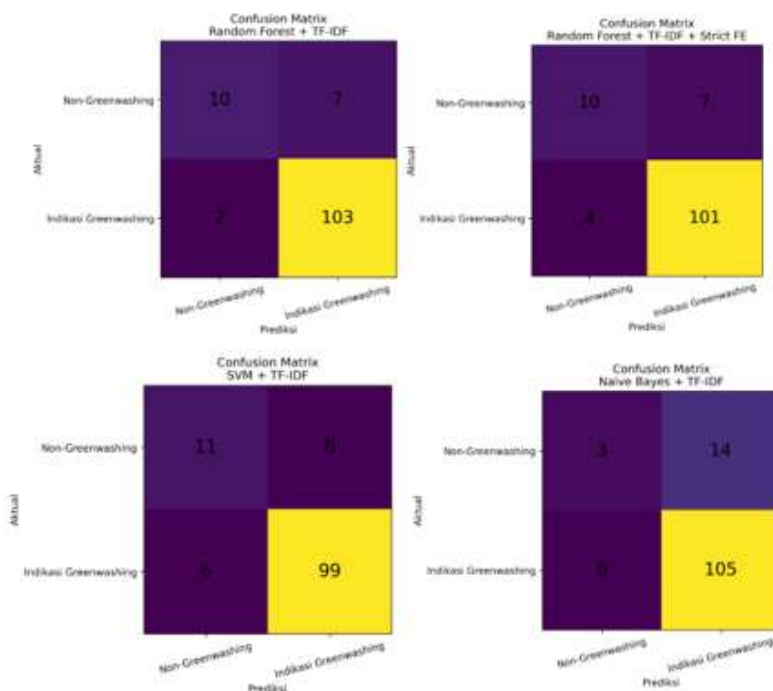


Fig. 3. Confusion Matrices of the Evaluated Machine Learning Models

The confusion matrices indicate that all models were highly effective in identifying Indication of Greenwashing, but their ability to correctly identify Non-Greenwashing varied. SVM + TF-IDF showed the strongest discrimination for the Non-Greenwashing class, whereas NB + TF-IDF showed a stronger tendency to classify observations as Indication of Greenwashing. These findings are consistent with the class-specific performance results and reflect the effect of the imbalanced class distribution.

\*name of corresponding author



### Cross-Validation Results

Five-fold stratified cross-validation was conducted with TF-IDF and scaling fitted separately within each training fold. This procedure prevents information from the validation fold from influencing the preprocessing stage. Table 7 summarizes the mean and standard deviation of the evaluation metrics across the five folds.

Table 7. 5-Fold Cross-Validation Result

| Model                   | Accuracy      | F1-score      | Macro-F1      | Balanced Accuracy |
|-------------------------|---------------|---------------|---------------|-------------------|
| RF + TF-IDF             | 94.10 ± 0.90% | 96.64 ± 0.53% | 86.17 ± 1.80% | 81.71 ± 2.15%     |
| RF + TF-IDF + Strict FE | 94.75 ± 0.45% | 97.00 ± 0.28% | 87.96 ± 0.78% | 84.00 ± 2.17%     |
| SVM + TF-IDF            | 95.08 ± 2.09% | 97.13 ± 1.23% | 89.83 ± 4.13% | 89.05 ± 4.45%     |
| NB + TF-IDF             | 90.82 ± 1.22% | 94.92 ± 0.66% | 73.64 ± 4.03% | 67.88 ± 3.52%     |

SVM + TF-IDF achieved the highest mean accuracy (95.08%), precision (96.79%), F1-score (97.13%), Macro-F1 (89.83%), and balanced accuracy (89.05%). RF + TF-IDF + Strict FE achieved a slightly lower accuracy of 94.75% and F1-score of 97.00%, but its F1 standard deviation was only 0.28%, indicating highly stable performance across folds. These complementary results suggest that SVM is strongest in class-balanced predictive performance, while the strict-feature Random Forest provides particularly stable discrimination.

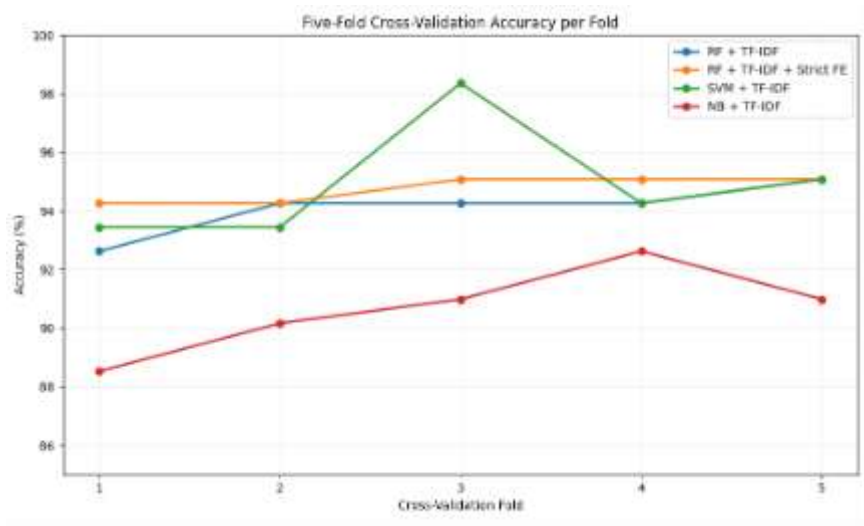


Fig. 4. Five-Fold Cross-Validation Accuracy per Fold

Figure 4 presents the accuracy of the four classification models across the five cross-validation folds. SVM + TF-IDF shows the highest accuracy in Fold 3 (98.36%), while its accuracy varies across the other folds. RF + TF-IDF and RF + TF-IDF + Strict FE show relatively stable accuracy across the five folds, ranging from 92.62% to 95.08%. NB + TF-IDF consistently records the lowest accuracy, ranging from 88.52% to 92.62%. Overall, the figure indicates that the Random Forest configurations provide relatively stable accuracy across folds, while SVM exhibits greater variation, particularly in Fold 3. The variation across folds is consistent with the standard deviation reported in the cross-validation summary, where SVM has a higher accuracy SD (2.09%) compared with RF + TF-IDF (0.90%) and RF + TF-IDF + Strict FE (0.45%).

### ROC-AUC Analysis

To further evaluate the discriminatory ability of the models, ROC-AUC analysis was conducted using five-fold cross-validation. ROC-AUC measures the ability of a classifier to distinguish between the Non-Greenwashing and Indication of Greenwashing classes across different classification thresholds. The results are presented in Table 8, while the corresponding ROC curves on the independent test set are shown in Figure 5.

Table 8. ROC-AUC performance across five cross-validation folds

| Model       | ROC-AUC Mean ± SD |
|-------------|-------------------|
| RF + TF-IDF | 95.99 ± 1.72%     |

\*name of corresponding author



This is anCreative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

|                         |                    |
|-------------------------|--------------------|
| RF + TF-IDF + Strict FE | $96.07 \pm 1.46\%$ |
| SVM + TF-IDF            | $95.02 \pm 4.06\%$ |
| NB + TF-IDF             | $92.28 \pm 3.24\%$ |

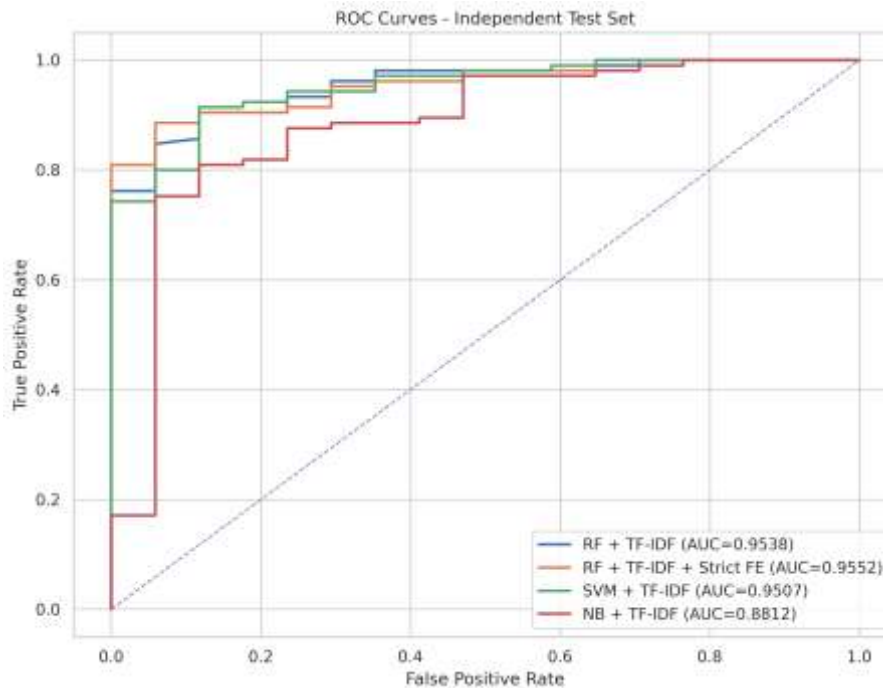


Fig. 5. ROC Curves of the Four Classification Models on the Independent Test Set

As shown in Table 7, RF + TF-IDF + Strict FE achieved the highest mean ROC-AUC of  $96.07\% \pm 1.46\%$ , followed by RF + TF-IDF at  $95.99\% \pm 1.72\%$ . SVM + TF-IDF obtained  $95.02\% \pm 4.06\%$ , while NB + TF-IDF achieved the lowest value of  $92.28\% \pm 3.24\%$ . The ROC curves in Figure 5 further illustrate the strong discriminatory performance of all models, with the curves generally positioned well above the random classification line. Overall, the results indicate that the Random Forest configurations provide strong discrimination between the two classes, with the Strict FE configuration achieving the highest cross-validation ROC-AUC.

### Precision-Recall AUC Analysis

To further evaluate model performance under the imbalanced class distribution, Precision-Recall AUC (PR-AUC) analysis was conducted using five-fold cross-validation. PR-AUC evaluates the trade-off between precision and recall across different classification thresholds. The results are presented in Table 9, while the corresponding Precision-Recall curves on the independent test set are shown in Figure 6.

Table 9. PR-AUC Performance Based on Five-Fold Cross-Validation

| Model                   | PR-AUC Mean $\pm$ SD |
|-------------------------|----------------------|
| RF + TF-IDF             | $99.18 \pm 0.41\%$   |
| RF + TF-IDF + Strict FE | $99.20 \pm 0.38\%$   |
| SVM + TF-IDF            | $98.45 \pm 1.99\%$   |
| NB + TF-IDF             | $98.25 \pm 1.00\%$   |

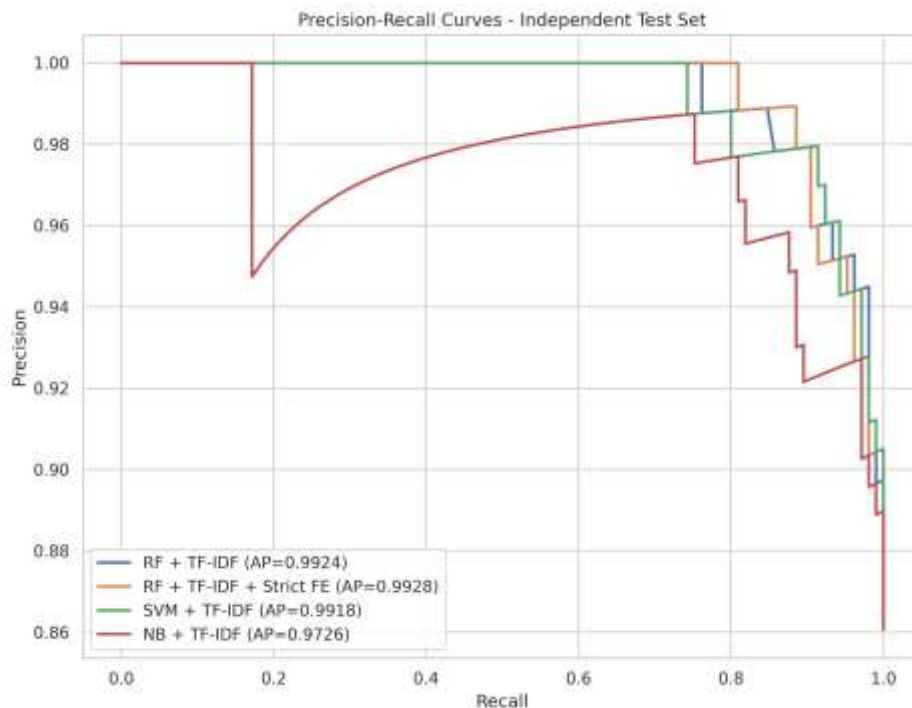


Fig. 6. Precision-Recall Curves of the Four Classification Models on the Independent Test Set

As shown in Table 9, RF + TF-IDF + Strict FE achieved the highest mean PR-AUC of  $99.20\% \pm 0.38\%$ , followed by RF + TF-IDF at  $99.18\% \pm 0.41\%$ . SVM + TF-IDF and NB + TF-IDF obtained mean PR-AUC values of  $98.45\% \pm 1.99\%$  and  $98.25\% \pm 1.00\%$ , respectively. The Precision-Recall curves in Figure 6 further illustrate the high precision and recall achieved by all models on the independent test set. Overall, the high PR-AUC values indicate that all models maintained strong precision-recall performance, with RF + TF-IDF + Strict FE achieving the highest and relatively stable result across the five folds.

#### Comparison Between Independent Test and Cross-Validation

The independent-test and cross-validation results show broadly consistent model behavior, although the absolute values differ because the two evaluation settings use different data partitions. Cross-validation produced higher accuracy and F1-score than the single independent test for all models, while ROC-AUC and PR-AUC remained high under both settings. RF + TF-IDF achieved 92.62% test accuracy versus 94.10% cross-validation accuracy, while SVM achieved 90.16% versus 95.08%. RF + TF-IDF + Strict FE showed the largest practical improvement in F1 from 94.84% on the independent test to 97.00% in cross-validation. These differences indicate that model performance is partition-dependent and should therefore be interpreted using both evaluation strategies rather than a single score.

Table 10. Comparison of Independent-Test and Cross-Validation Performance

| Model                   | Test Accuracy | CV Accuracy Mean | Test F1 | CV F1 Mean | Test ROC-AUC | CV ROC-AUC Mean | Test PR-AUC | CV PR-AUC Mean |
|-------------------------|---------------|------------------|---------|------------|--------------|-----------------|-------------|----------------|
| RF + TF-IDF             | 92.62%        | 94.10%           | 95.81%  | 96.64%     | 95.38%       | 95.99%          | 99.24%      | 99.18%         |
| RF + TF-IDF + Strict FE | 90.98%        | 94.75%           | 94.84%  | 97.00%     | 95.52%       | 96.07%          | 99.28%      | 99.20%         |
| SVM + TF-IDF            | 90.16%        | 95.08%           | 94.29%  | 97.13%     | 95.07%       | 95.02%          | 99.18%      | 98.45%         |
| NB + TF-IDF             | 88.52%        | 90.82%           | 93.75%  | 94.92%     | 88.12%       | 92.28%          | 97.26%      | 98.25%         |

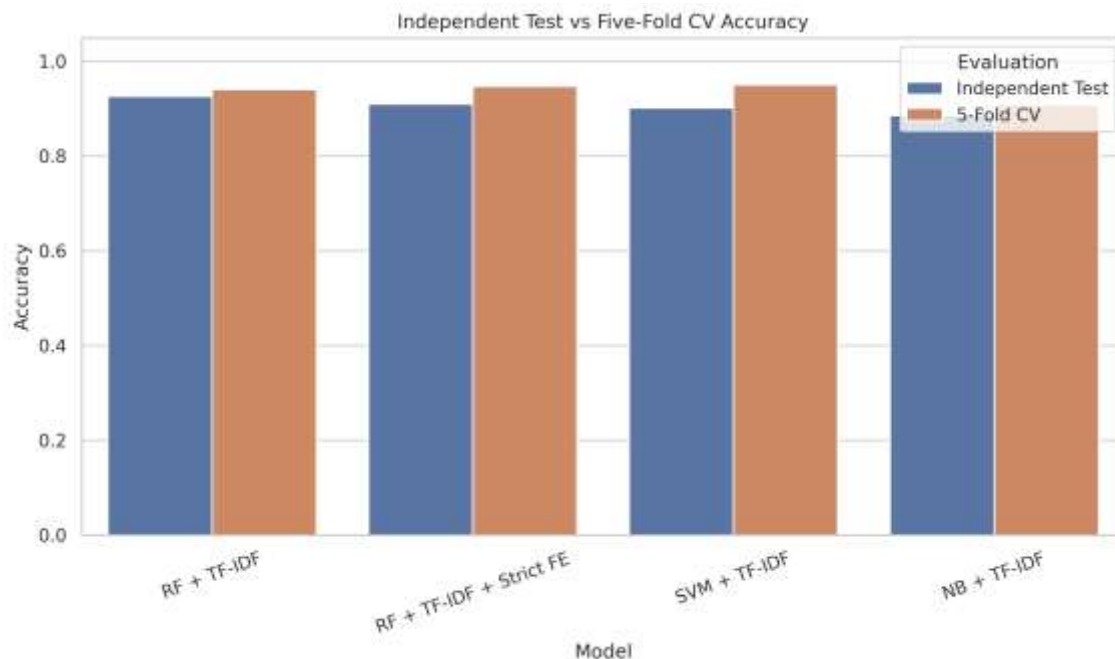


Fig. 7. Independent-test versus five-fold cross-validation accuracy

As shown in Table 10, cross-validation generally produced higher Accuracy and F1-score than the independent test. SVM + TF-IDF achieved the highest cross-validation Accuracy (95.08%) and F1-score (97.13%), while RF + TF-IDF + Strict FE achieved the highest cross-validation ROC-AUC (96.07%) and PR-AUC (99.20%). Figure 7 confirms the higher Accuracy obtained from cross-validation across all models. Overall, the results indicate consistent model performance across the two evaluation settings, with cross-validation providing slightly higher performance estimates and the independent test offering a more conservative evaluation on unseen data

#### Predictive Signals Associated with Model Decisions

To improve model interpretability, feature importance analysis was performed on the Random Forest model using TF-IDF and strict text-based features. The analysis identifies textual features that contributed relatively more to the model's prediction decisions. The top 10 predictive signals are presented in Figure 8.

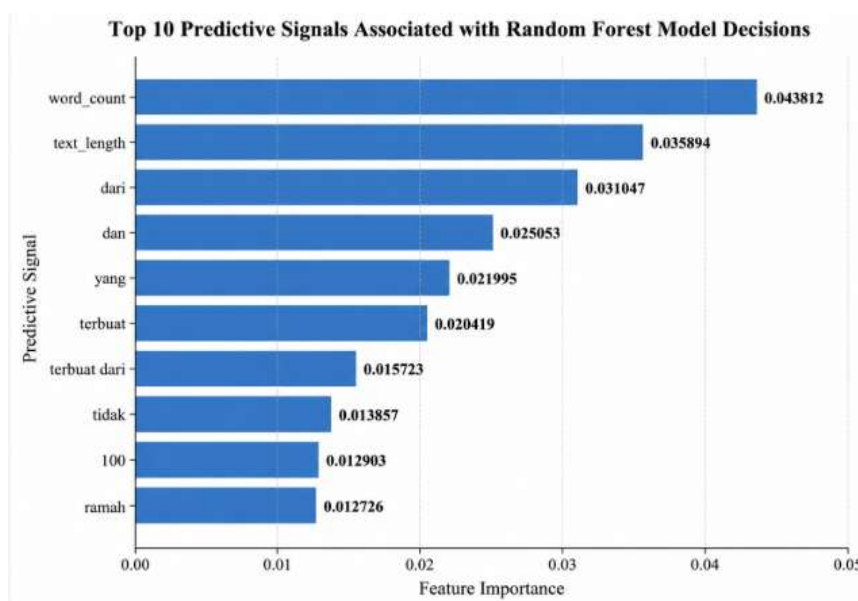


Fig. 8. Top 10 Predictive Signals Associated with Random Forest Model Decisions

\*name of corresponding author



This is anCreative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

As shown in Figure 8, *word\_count* and *text\_length* were the most prominent predictive signals, followed by textual features such as *dari*, *dan*, *yang*, *terbuat*, *terbuat dari*, *tidak*, *100*, and *ramah*. From a marketing management perspective, these signals can help identify product descriptions that may require closer review, particularly when environmental claims are expressed through material composition or environmental attributes. From a legal analysis perspective, the identified signals can support an initial screening process for potentially problematic environmental claims, but they cannot be regarded as legal evidence of greenwashing or regulatory violations. Legal assessment still requires examination of the meaning, substantiation, context, and accuracy of the specific environmental claims. Overall, the feature-importance analysis provides an interpretable link between textual patterns, marketing communication, and preliminary legal screening. The identified signals may assist marketers in reviewing environmental claims and help legal or compliance teams prioritize descriptions for further examination. However, the importance scores represent model behavior rather than causal determinants or definitive legal findings

## DISCUSSIONS

The experimental results demonstrate that Natural Language Processing combined with machine learning can provide an effective preliminary screening mechanism for identifying product descriptions containing indications of potentially misleading environmental claims. Random Forest with TF-IDF achieved the strongest performance on the independent test set, with an Accuracy of 92.62% and F1-score of 95.81%. However, the class-specific results show that identifying the Non-Greenwashing class remains more difficult, with a Recall of only 58.82% for Random Forest + TF-IDF. The five-fold cross-validation results further show that SVM achieved the highest mean Accuracy ( $95.08 \pm 2.09\%$ ), F1-score ( $97.13 \pm 1.23\%$ ), Macro-F1 ( $89.83 \pm 4.13\%$ ), and Balanced Accuracy ( $89.05 \pm 4.45\%$ ), while Random Forest with Strict Feature Engineering achieved the highest ROC-AUC ( $96.07 \pm 1.46\%$ ) and PR-AUC ( $99.20 \pm 0.38\%$ ). These findings indicate that model performance varies according to the evaluation criterion and that class-balanced metrics are important for this imbalanced dataset.

From a legal analysis perspective, the proposed model can support preliminary screening and prioritization of environmental claims that require further examination. Product descriptions predicted as Indication of Greenwashing can be prioritized for human review, potentially reducing the amount of information that must be examined manually. However, the model output should not be interpreted as definitive evidence of greenwashing or a legal violation. Legal assessment requires examination of the specific claim, supporting evidence, communication context, and applicable consumer-protection and advertising requirements. Therefore, the proposed approach is more appropriately positioned as a decision-support mechanism for preliminary legal and compliance screening.

From a marketing management perspective, the findings suggest that environmental claims should be communicated with greater specificity and adequate supporting information. General environmental expressions may require further clarification when they do not clearly describe the material, production process, certification, or scope of the claimed environmental benefit. The predictive-signal analysis identified features such as *word\_count*, *text\_length*, *terbuat*, *terbuat dari*, and *ramah* as relatively important signals in the Random Forest model. These signals can assist sellers and e-commerce platforms in identifying product descriptions that require closer review. Nevertheless, these features should not be interpreted as inherent indicators of greenwashing, because their importance reflects patterns learned by the model from the dataset.

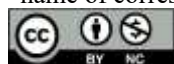
Overall, the findings address the identified research gap by demonstrating the potential of product-level NLP and machine learning for preliminary screening of greenwashing indications in e-commerce. The integration of comparative model evaluation and predictive-signal analysis also provides practical information for consumer-protection screening and sustainable marketing management. However, the proposed approach remains dependent on the characteristics of the dataset, including its size, platform coverage, selected keywords, and class distribution. Future research should expand the dataset and compare classical NLP methods with transformer-based or other contextual models to improve the identification of more complex environmental claims.

## CONCLUSION

The results demonstrate that Natural Language Processing combined with machine learning can effectively support the preliminary screening of e-commerce product descriptions containing indications of potentially misleading environmental claims. Random Forest with TF-IDF achieved the strongest performance on the independent test set, with an Accuracy of 92.62% and F1-score of 95.81%, while five-fold cross-validation showed that SVM achieved the highest mean Accuracy of  $95.08 \pm 2.09\%$  and F1-score of  $97.13 \pm 1.23\%$ . Random Forest with Strict Feature Engineering also achieved the highest ROC-AUC ( $96.07 \pm 1.46\%$ ) and PR-AUC ( $99.20 \pm 0.38\%$ ). These findings indicate that classical machine learning can capture relevant textual patterns for identifying indications of greenwashing, although the lower performance for the Non-Greenwashing class confirms the importance of considering class imbalance and multiple evaluation metrics.

From a legal perspective, the proposed model can support preliminary screening and prioritization of environmental claims for further human and evidentiary review, but its predictions should not be considered

\*name of corresponding author



This is anCreative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

definitive evidence of greenwashing or legal violations. From a marketing management perspective, the model can assist sellers and e-commerce platforms in reviewing environmental claims and improving the specificity and substantiation of environmental communication. Overall, the proposed approach provides complementary support for legal screening and marketing management; however, its application is limited by the dataset size, platform coverage, selected keywords and class distribution. Future research should expand the dataset and investigate transformer-based or other contextual NLP approaches to improve the identification of more complex environmental claims.

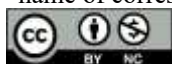
### ACKNOWLEDGMENT

We gratefully acknowledge that this work was funded by the DIKTI Grant of the Ministry of Higher Education, Science, and Technology (Kemdiktisaintek) through the Directorate General of Research and Development (Ditjen Risbang), Fiscal Year 2026. The authors also wish to extend their appreciation to Universitas Battuta and the Research and Community Service Institute (LPPM) Team of Universitas Battuta for their invaluable support in the implementation of the 2026 DIKTI Grant program.

### REFERENCES

- Balqis, T., Widiyanti, I. D., & Zulaika, E. (2023). Legal Protection for Eco-Friendly Product Consumer Losses Due to the Practice of Greenwashing. *Law Review*, 23(2), 140–154. <https://doi.org/10.19166/lr.v20i1.6915>
- Chaudhary, A., Kolhe, S., & Kamal, R. (2016). An improved random forest classifier for multi-class classification. *Information Processing in Agriculture*, 3(4), 215–222. <https://doi.org/10.1016/j.inpa.2016.08.002>
- Dewi, R., Sri, R., Saleh, A., Yani, D., Tanjung, H., & Jinan, A. (2025). Enhancing Machine Learning Algorithm Performance for PCOS Diagnosis Using Smotenc on Imbalanced Data. *JITK (Jurnal Ilmu Pengetahuan Dan Teknologi Komputer)*, 11(1), 55–63. <https://doi.org/10.33480/jitk.v11i1.6676>
- Fatchurrohman, M., Su, M., & Puspitasari, F. D. (2026). Asimetri Informasi dan Greenwashing dalam Era Pemasaran Digital : Peran Strategi Digital Marketing terhadap Persepsi Keberlanjutan Konsumen. *SENTRI: Jurnal Riset Ilmiah*, 5(2), 1714–1724. <https://doi.org/10.55681/sentri.v5i2.5714>
- Gorovaia, N. (2024). Identifying greenwashing in corporate - social responsibility reports using natural - language processing. *Wiley: European Financial Management*, 31(1), 427–462. <https://doi.org/10.1111/eufm.12509>
- Istanti, A. N., & Samsul, I. (2023). Analysis of Greenwashing Policy in Providing Consumer Protection in Indonesia in View of Act Number 8 of 1999 Concerning Consumer Protection. *De Lega Lata: Jurnal Ilmu Hukum*, 8(1), 8–16. <https://doi.org/10.30596/dll.v8i1.12477>
- Jalal, N., Mehmood, A., Sang, G., & Ashraf, I. (2022). A novel improved random forest for text classification using feature ranking and optimal number of trees. *Journal of King Saud University - Computer and Information Sciences*, 34(6), 2733–2742. <https://doi.org/10.1016/j.jksuci.2022.03.012>
- Janice, F., Boedijanto, O., & Delina, L. L. (2024). Energy Research & Social Science Potentials and challenges of artificial intelligence-supported greenwashing detection in the energy sector. *Elsevier*, 115. <https://doi.org/10.1016/j.erss.2024.103638>
- Javed, U., Rashid, M. A., Hussain, G., & Shafique, S. (2023). Consumers ' perception of green brand attributes and its outcomes : a sustainable perspective. *Journal of Environmental Planning and Management*, 1–23. <https://doi.org/10.1080/09640568.2023.2178882>
- Lagasio, V. (2024). International Review of Financial Analysis ESG-washing detection in corporate sustainability reports. *International Review of Financial Analysis*, 96(PB), 103742. <https://doi.org/10.1016/j.irfa.2024.103742>
- Li, M., Gao, Q., & Yu, T. (2023). Kappa statistic considerations in evaluating inter-rater reliability between two raters : which , when and context matters. *BMC Cancer*, 23(799), 1–5. <https://doi.org/10.1186/s12885-023-11325-z>
- Mihelich, M., Castagnos, F., & Dognin, C. (2024). Interplay of ROC and Precision-Recall AUCs : Theoretical Limits and Practical Implications in Binary Classification. *Proceedings of Machine Learning Research*, 235, 1–8.
- Moodaley, W., & Telukdarie, A. (2023). Greenwashing , Sustainability Reporting , and Artificial Intelligence : A Systematic Literature Review. *Sustainability*, 15(2), 1481. <https://doi.org/10.3390/su15021481>
- Nugroho, A. (2026). Green promises and unverifiable claims : Detecting corporate greenwashing in Indonesian sustainability reports through corpus linguistics and natural language processing. *Indonesian Journal of Language and Economic Discourse*, 1(2). <https://doi.org/10.67490/ijle.v1i2.904>
- Nyilasy, G., Gangadharbatla, H., & Paladino, A. (2013). Perceived Greenwashing: The Interactive Effects of Green Advertising and Corporate Environmental Performance on Consumer Reactions. *Journal Business Ethics*. <https://doi.org/10.1007/s10551-013-1944-3>
- Pizzetti, M., Gatti, L., & Seele, P. (2021). Green lies and their effect on intention to invest. *Journal of Business Research*, 127(3), 228–240. <https://doi.org/10.1016/j.jbusres.2021.01.028>

\*name of corresponding author



This is anCreative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

- Ramadhani, B., & Hermawan, A. (2025). E-Commerce dan Kelestarian Lingkungan : Analisis Dampak dan Upaya Berkelanjutan. *Jurnal Rimba*, 3(2), 229–240. <https://doi.org/10.61132/rimba.v3i2.1760>
- Richardson, E., Trevizani, R., Greenbaum, J. A., & Carter, H. (2024). Article The receiver operating characteristic curve accurately assesses imbalanced datasets The receiver operating characteristic curve accurately assesses imbalanced datasets. *Patterns*, 5(6), 100994. <https://doi.org/10.1016/j.patter.2024.100994>
- Sajid, M., Ertz, M., & Peethambaran, M. (2025). *Does greenwashing dim green actions ? Examining how greenwashing impacts consumers ' pro-environmental intentions*. 21(10), 2336–2355. <https://doi.org/10.1108/SRJ-03-2025-0211>
- Steenis, N. D., Herpen, E. Van, Lans, I. A. Van Der, & Trijp, H. C. M. Van. (2023). Partially Green , Wholly Deceptive ? How Consumers Respond to ( In ) Consistently Sustainable Packaged Products in the Presence of Sustainability Claims Partially Green , Wholly Deceptive ? How Consumers Respond to. *Journal of Advertising*, 52(2), 159–178. <https://doi.org/10.1080/00913367.2022.2047841>
- Syahputra, F., Dahri Yani Hakim, T., Dewi, R., & Hilda Elsera Br, S. (2026). Analisis Sentimen Masyarakat terhadap Isu Transformasi Digital Pemerintah di Twitter Menggunakan Naive Bayes dan Support Vector Machine. *Infosys (Information System) Journal*, 10(2), 73–85. <https://doi.org/10.22303/infosys.10.2.2026>
- Takahashi, K., Yamamoto, K., Kuchiba, A., & Koyama, T. (2022). Confidence interval for micro-averaged F 1 and macro-averaged F 1 scores. *Applied Intelligence*, 52, 4961–4972. <https://doi.org/10.1007/s10489-021-02635-5>
- Ur Rahman, S., & Nguyen-Viet, B. (2023). Towards sustainable development : Coupling green marketing strategies and consumer perceptions in addressing greenwashing. *Business Starategy and the Environment*, June 2022, 2420–2433. <https://doi.org/10.1002/bse.3256>
- Vieira, S., Netto, D. F., Felipe, M., Sobral, F., Regina, A., Ribeiro, B., & Robert, G. (2020). Concepts and forms of greenwashing : a systematic review. *Environmental Sciences Europe*, 32(19), 1–12. <https://doi.org/10.1186/s12302-020-0300-3>
- Wang, X., Gao, X., & Sun, M. (2025). Construction and analysis of corporate greenwashing index : a deep learning approach. *EPJ Data Sci.*, 14(44). <https://doi.org/10.1140/epjds/s13688-025-00562-w>
- Wongkar, E. E. L. T., & Apsari, P. K. (2021). Telaah Kebijakan Sustainable Consumption and Production ( SCP ) Era E-Commerce. *Indonesian Center for Environmental Law*.
- Zeng, F., Wang, J., & Zeng, C. (2025). An optimized machine learning framework for predicting and interpreting corporate ESG greenwashing behavior. *PLoS ONE*, 20(3), 1–25. <https://doi.org/10.1371/journal.pone.0316287>