ISSN 2288-4866 (Print)
ISSN 2288-4882 (Online)
https://www.jiisonline.org

http://dx.doi.org/10.13088/jiis.2023.29.4.325

Prediction of Customer Satisfaction Using RFE-SHAP Feature Selection Method[*]

Olga Chernyaeva
College of Business Administration,
Pusan National University
(misslelka@pusan.ac.kr)

Taeho Hong
College of Business Administration,
Pusan National University
(hongth@pusan.ac.kr)

In the rapidly evolving domain of e-commerce, our study presents a cohesive approach to enhance customer satisfaction prediction from online reviews, aligning methodological innovation with practical insights. We integrate the RFE-SHAP feature selection with LDA topic modeling to streamline predictive analytics in e-commerce. This integration facilitates the identification of key features—specifically, narrowing down from an initial set of 28 to an optimal subset of 14 features for the Random Forest algorithm. Our approach strategically mitigates the common issue of overfitting in models with an excess of features, leading to an improved accuracy rate of 84% in our Random Forest model. Central to our analysis is the understanding that certain aspects in review content, such as quality, fit, and durability, play a pivotal role in influencing customer satisfaction, especially in the clothing sector. We delve into explaining how each of these selected features impacts customer satisfaction, providing a comprehensive view of the elements most appreciated by customers. Our research makes significant contributions in two key areas. First, it enhances predictive modeling within the realm of e-commerce analytics by introducing a streamlined, feature-centric approach. This refinement in methodology not only bolsters the accuracy of customer satisfaction predictions but also sets a new standard for handling feature selection in predictive models. Second, the study provides actionable insights for e-commerce platforms, especially those in the clothing sector. By highlighting which aspects of customer reviews—like quality, fit, and durability—most influence satisfaction, we offer a strategic direction for businesses to tailor their products and services.

Keywords : Customer Satisfaction, Feature Selection, SHAP, Recursive Feature Elimination, Topic Modelling

Received : November 13, 2023 Revised : December 20, 2023 Accepted : December 21, 2023
Publication Type : Regular Track Corresponding Author : Taeho Hong

1. Introduction

The digital transformation has ushered in an era where e-commerce platforms have become the cornerstone of modern retail, fundamentally altering the dynamics between businesses and customers (Engler et al., 2015). As customers increasingly turn to online platforms for their shopping needs,

[*] This work was supported by the MSIT (Ministry of Science and ICT), Korea, under the ITRC (Information Technology Research Center) support program (No. IITP-2022-2020-0-01797) supervised by the IITP (Institute for Information & Communications Technology Planning & Evaluation). Additionally, this research was funded by an IITP grant as a part of the Convergence Security Core Talent Training Business at Pusan National University (No. 2022-0-01201).

they leave behind a trail of data in the form of online reviews, purchase histories, and browsing patterns (Zhang and Luo, 2023). These digital footprints, particularly online reviews, have emerged as a goldmine of insights, offering businesses a lens into the minds of their customers (Matuszelański and Kopczewska, 2022). Therefore, customer reviews and ratings serve as a critical source of information for potential buyers, mitigating product uncertainty (Chen and Xie, 2008). Consistent with this perspective, Lin et al. (2011) have demonstrated that sentiments and ratings embedded within reviews significantly influence sales across a range of contexts. Moreover, online retailers and manufacturers increasingly harness customer feedback to refine their marketing strategies, optimize product listings through relevance sorting, and forge new revenue pathways (Chen and Xie, 2008; Cui et al., 2012). This strategic utilization explains the widespread adoption of product rating functionalities by leading online marketplaces such as Amazon.com (Mudambi and Schuff, 2010).

Scholars from marketing and information systems have taken a keen interest in this domain, scrutinizing the effects of online product ratings on sales and identifying the characteristics of reviews that denote customer satisfaction (Chen and Xie, 2008; Lin et al., 2011). Moreover, machine learning models have been used to predict customer satisfaction in various industries. Advanced machine learning techniques, such as those employed by Bauer and Jannach (2021), have been increasingly applied to e-commerce data, aiming to predict key metrics like satisfaction and customer lifetime value. Their research underscores the potential of combining multiple machine learning techniques, including deep learning and gradient-boosting machines, to enhance predictive accuracy. Hong et al. (2023) used machine learning models such as K-Nearest Neighbors, Decision Tree Classifier, Logistic Regression, Random Forest, Naïve Bayes, and AdaBoost to predict airline customer satisfaction, with Random Forest achieving the best performance. Darko and Liang (2022) proposed a probabilistic linguistic group decision methodology to model customer satisfaction using online customer reviews, employing techniques like Latent Dirichlet Allocation, SOM clustering, and unsupervised machine learning for sentiment analysis. These studies demonstrate the effectiveness of machine learning in predicting customer satisfaction in various domains.

The selection of appropriate features is paramount in predictive modeling, directly influencing the classifiers' performance (Liu et al., 2016). Prior research in this area frequently sought to augment predictive accuracy by expanding the set of features. However, this strategy may engender increased variance within the model, particularly when training data is scarce, thereby escalating the risk of overfitting (Jing et al., 2023). Addressing this limitation, our investigation adopts the Recursive Feature Elimination with the SHapley Additive exPlanations (RFE-SHAP) technique to refine the prediction of customer satisfaction. Additionally, previous studies have leveraged machine learning to predict outcomes from reviews but have not sufficiently clarified the impact of specific review content on these predictions.

In our research, we endeavor to make a distinct contribution to the field of e-commerce analytics. Our approach is characterized by the innovative integration of Recursive Feature Elimination with Shapley Additive Explanations (RFE-SHAP) methodology into predictive modeling. This integration is particularly novel as it synergizes RFE-SHAP with topic modeling, specifically employing Latent Dirichlet Allocation (LDA). Such a combination facilitates a deeper, more nuanced analysis, particularly in interpreting customer reviews within the e-commerce landscape. Our methodology stands apart from conventional practices by focusing on the optimization of predictive models through the strategic elimination of superfluous features. This approach effectively prevents model overfitting. Moreover, our empirical research highlights the effectiveness of the Random Forest algorithm, which, when enhanced by a carefully pruned set of 14 key features, demonstrated impressive performance metrics.

A pivotal aspect of our study is the illumination of specific topics that play a crucial role in predicting customer satisfaction within the clothing sector. These topics include ‘Quality & Appearance,’ ‘Fit & Comfort,’ ‘Durability Concerns,’ ‘Comfort & Style,’ and ‘Quality & Materials.’ The identification and exploration of these topics provide businesses with invaluable insights. By understanding these key areas, companies can better align their products and strategies with customer preferences and expectations, thus enhancing customer satisfaction and business performance. In summary, our research not only offers a refined methodological approach to e-commerce analytics but also provides practical, actionable insights for businesses in the clothing sector, aimed at improving customer satisfaction and operational efficiency.

In the following section, we explore the literature on customer satisfaction and sentiment analysis. Section 2.2 provides an overview of how content analysis applies to online reviews. In Section 2.3, we look at existing research on feature selection methods and how they highlight differences in content. Section 3 introduces our research framework and the analysis we conducted. Section 4 discusses the outcomes of using the RFE-SHAP method for feature selection and explains these results through topic modeling and SHAP analysis. The paper concludes in Section 5, where we discuss the implications of our work for both practice and future research and consider its limitations.

2. Literature Review

2.1. Customer Satisfaction and Sentiment Analysis

Customer satisfaction is a critical factor for businesses in today’s competitive market (Kang and Park, 2014). With the rise of online platforms and social media, customers have a powerful voice to express their opinions and experiences (Zhang and Luo, 2023). Online reviews have become a valuable source of information for businesses to understand customer satisfaction and make informed decisions (Aakash and Gupta, 2022; Cheryaeva and Hong, 2022). In this section, we review the literature on customer satisfaction and sentiment analysis in the context of online review analysis.

One important aspect of analyzing online reviews is sentiment analysis, which aims to determine the sentiment expressed in a text, whether it is positive, negative, or neutral (Park and Kim, 2017; Johar and Mubeen, 2020). Sentiment analysis techniques have been widely used to analyze online reviews and understand customer satisfaction (Ren et. al., 2016; Park and Lee, 2023). Several studies have delved into the intricacies of sentiment analysis in the context of online review analysis, each shedding light on different facets of the domain.

Kumar et. al. (2019) pioneered this domain by proposing a multimodal framework that amalgamated physiological signals with global reviews. Their innovative approach combined Natural Language Processing (NLP) techniques with EEG signals, capturing both explicit feedback and intrinsic attractiveness towards products. Their findings highlighted the potential of integrating diverse data sources for a more nuanced understanding of customer sentiment. Expanding on this, Wisnu et al. (2020) focused on the burgeoning digital payment landscape in Indonesia. They employed machine learning approaches, including Naïve Bayes and K-Nearest Neighbour, to analyze Twitter data and gauge sentiments around digital payments. Their research underscored the superior performance of KNN in terms of accuracy, emphasizing the importance of algorithm selection in sentiment analysis. Further, Maharani and Triayudi (2022) explored the sentiment analysis of Indonesian digital payment platforms, including GOPAY, DANA, and ShopeePay. Through rigorous data preparation stages and the application of machine learning approaches, their study provided valuable insights into public satisfaction with these services.

While the aforementioned studies have contributed significantly to the understanding of sentiment analysis in the context of online reviews, there are notable limitations. A recurrent limitation in previous research is the lack of pairing between the rating and sentiment scores of reviews. Many studies did not differentiate between reviews written by any customer and those written by customers with validated purchases. This distinction is crucial as reviews from validated purchases often provide a more accurate representation of genuine customer sentiment. In our research, we address this gap by focusing exclusively on valid purchase reviews. We ensure that positive sentiments are paired with high ratings and negative sentiments are paired with low ratings, providing a more holistic and accurate understanding of customer satisfaction.

2.2. Content Analysis of Reviews: LDA

Content analysis has become a staple method for parsing through textual data, particularly online reviews, which serve as a goldmine of customer insights in the e-commerce sector (Ren et al., 2016). Beyond the realm of sentiment analysis, topic modeling emerges as an invaluable technique to unearth the latent topics within these reviews. At the forefront of topic modeling stands Latent Dirichlet Allocation (LDA), a probabilistic algorithm that has carved its niche in content analysis (He et al., 2020). The ubiquity of LDA in analyzing online reviews stems from its proficiency in capturing latent topics and themes embedded within textual data.

Venturing into the domain of customer satisfaction, LDA serves as a beacon to pinpoint the topics woven within online reviews (Alzahrani et al., 2022). By deploying LDA on a corpus of reviews, businesses can glean the predominant themes and subjects customers' opinions, offering a panoramic view of facets contributing to customer satisfaction.

Delving into the e-commerce landscape, Liu, Zhou, Jiang, and Zhang (2020) conducted a meticulous study centered around business-to-customer (B2C) online retail platforms. Their research, anchored by a vast dataset comprising 150,000 product reviews, harnessed the potential of LDA to distill key satisfaction determinants. Notably, logistics emerged as a dominant theme, commanding attention in a substantial 38.5% of reviews. Additionally, facets like product quality and customer service surfaced as pivotal. Their findings resonate with the broader e-commerce narrative, emphasizing the imperative for businesses to fortify logistics, ensure product quality, and elevate customer service to optimize satisfaction.

Karim and Das (2018) illuminated the synergy of LDA with sentiment analysis. Their comparative study, which pitted rule-based techniques against machine learning methodologies, championed the latter, especially when amalgamated with LDA. Furthermore, the prowess of LDA was accentuated in a study concentrating on the e-commerce tourism sector. By scrutinizing online ratings and reviews, the research delineated critical dimensions of e-service quality, proffering actionable insights for e-tourism platforms.

Despite the wide application of LDA in predicting customer satisfaction, a discernible lacuna persists in the literature (Du and Huang, 2018; Yu et al., 2023). While extant studies have adeptly harnessed LDA for prediction, they fall short in explicating which specific topics influence prediction models. This omission underscores the need for research that ventures beyond mere topic extraction and delves into the nuanced interplay of topics and their bearing on prediction outcomes.

2.3. RFE-SHAP and eXplanation of Content Differences

Feature selection is an indispensable step in the development of prediction models, especially within the e-commerce domain. The process aids in pinpointing the most salient features that bolster accurate predictions, thereby enhancing the overall efficacy of the models (Ding et al., 2022). Recursive Feature Elimination (RFE) stands out as a frequently employed feature selection technique in prediction models (Darst et al., 2018; Kannari et al., 2022). RFE operates by iteratively eliminating features based on their significance or contribution to the model's performance, starting with all features and methodically removing the least crucial ones until the desired number of features remains.

SHapley Additive exPlanation (SHAP) has emerged as another feature selection method that has garnered considerable attention in recent times (Van den Broeck et al., 2022). SHAP values offer a distinct measure of feature importance within machine learning prediction models. These values not only elucidate the output of any machine learning model but also serve as a potent feature selection mechanism, especially when grappling with high-dimensional data.

Within the e-commerce landscape, the application of feature selection using RFE and SHAP can be instrumental in optimizing various facets of e-commerce operations (Jing et al., 2023; Chen et al., 2021). For instance, in the realm of predicting customer behavior or inclinations, feature selection can spotlight the most influential determinants that steer customer decisions, such as purchase history, browsing patterns, demographic data, and product attributes (Zhang et al., 2023). By zeroing in on the most pertinent features, prediction models in e-commerce can furnish invaluable insights for tailored marketing strategies, product recommendations, inventory management, and pricing optimization. These models, in turn, empower e-commerce enterprises to elevate customer satisfaction, amplify sales, and optimize resource distribution (Jing et al., 2023).

The prowess of feature selection methods like RFE and SHAP in e-commerce prediction models has been showcased in a plethora of studies. A research endeavor by Xu et al. (2018) delved into the detection and characterization of web bot traffic in a vast e-commerce marketplace, employing an expectation maximization based feature selection method. Another study by Ansari et al. (2020) highlighted the significance of feature selection in online reviews, emphasizing the role of RFE in enhancing prediction performance.

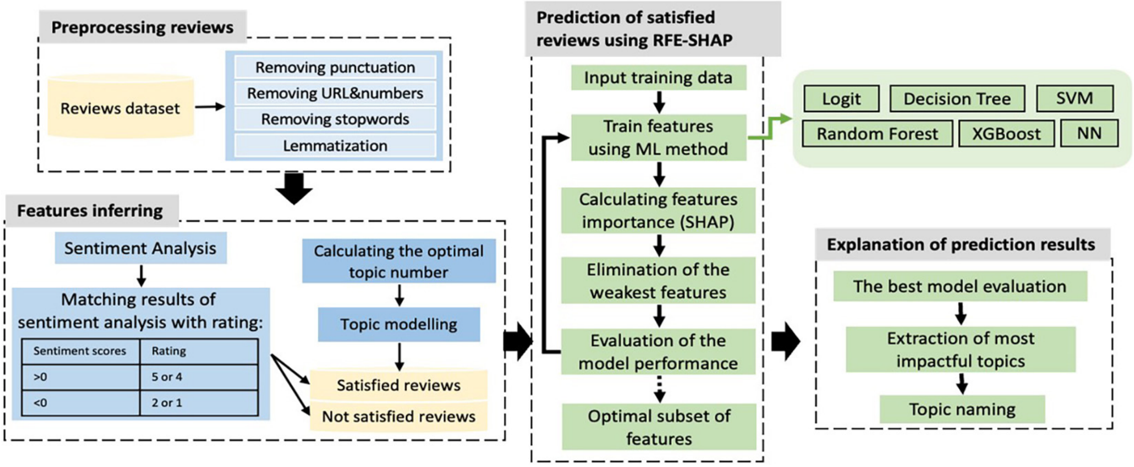
Our research diverges from previous studies in its unique approach. We amalgamate topic modeling and RFE-SHAP to not only select the most impactful features but also provide a contextual explanation of the most influential topic in the satisfied review prediction model. This fusion of techniques offers a more comprehensive understanding, bridging the gap between feature selection and interpretative analysis.

3. Research Framework and Analysis

Our research framework, illustrated in Figure 1, introduces a four-stage process to e-commerce analytics, differentiating it from traditional methodologies. The process begins with detailed preprocessing of customer reviews, followed by advanced feature selection using Recursive Feature Elimination with Shapley Additive Explanations (RFE-SHAP). This technique improves the accuracy of customer satisfaction predictions by reducing overfitting. Additionally, our approach enhances the interpretability of predictive results, an aspect often underemphasized in standard models. This methodology thus offers a comprehensive and transparent analysis, aiming for a more precise and insightful examination of customer reviews in the e-commerce sector.

Data description and review preprocessing:

The dataset, constituting 20,000 reviews of clothing items, was meticulously curated from Amazon.com. To bolster the veracity and reliability of the content, only ‘verified purchases’ were considered. This substantial corpus served as a robust foundation for subsequent analytical endeavors. The initial phase involved a rigorous preprocessing and text-


Preprocessing reviews
Reviews dataset
Removing punctuation
Removing URL&numbers
Removing stopwords
Lemmatization
Features inferring
Sentiment Analysis
Matching results of sentiment analysis with rating:
Sentiment scores
Rating
>0
5 or 4
<0
2 or 1
Calculating the optimal topic number
Topic modelling
Satisfied reviews
Not satisfied reviews
Prediction of satisfied reviews using RFE-SHAP
Input training data
Train features using ML method
Calculating features importance (SHAP)
Elimination of the weakest features
Evaluation of the model performance
Optimal subset of features
Logit
Decision Tree
SVM
Random Forest
XGBoost
NN
Explanation of prediction results
The best model evaluation
Extraction of most impactful topics
Topic naming

〈Figure 1〉 Research framework for developing the satisfied reviews prediction model with optimal feature number and for contextual explanation of prediction results

processing regimen, which comprised:

- Lowercasing: Ensuring textual uniformity by converting all characters to lowercase.
- Tokenization: Decomposing the text into discrete tokens or words.
- Stopword Removal: Purging frequently occurring words devoid of substantial semantic value.
- Lemmatization: Streamlining words to their foundational or root form to ensure linguistic congruence.

Features inferring:

After the preprocessing, sentiment scoring was undertaken using the dictionary method, a lexicon-centric approach. This method harnessed a comprehensive lexicon, the sentiment intensity analyzer package, to compute sentiment scores based on the prevalence of positive and negative words within each review. To ensure alignment between sentiment scores and numerical ratings, reviews were meticulously categorized:

- Positive Sentiment Alignment: Reviews with sentiment scores exceeding 0, accompanied by a numerical rating of 4 or 5.
- Negative Sentiment Alignment: Reviews with sentiment scores below 0, paired with a numerical rating of 1 or 2. Neutral reviews, that fall between positive and negative sentiments, were excluded.

Furthermore, for an accurate assessment of customer sentiments, we utilized VADER (Valence Aware Dictionary for Sentiment Reasoning), an NLTK module known for its effectiveness in sentiment analysis, especially in social media texts. VADER provides a compound sentiment score for each review, ranging from -1 to +1. Scores below 0 are categorized as negative, and those above 0 as positive. We excluded reviews with a neutral score of 0 to focus our analysis on the more pronounced positive and negative sentiments. This approach, in combination with the numerical ratings, offers a comprehensive method for defining customer satisfaction.

In the topic modeling phase of our study, we applied an iterative methodology using Latent Dirichlet Allocation (LDA). This involved training the LDA model with varying topic counts, each evaluated for coherence. The optimal topic count was selected based on the highest coherence score. With the optimal number of topics (‘*M’*) identified, the LDA model discerned the most significant themes in the reviews. For a dataset containing a specific number of reviews (‘*N’*), we derived an $N \times M$ matrix, where each element (i, j) represents the probability of the *i*th review being associated with the *j*th topic. This matrix, *Topic Probability Matrix*= $N \times M$, effectively quantifies each review’s alignment with the identified topics.

Prediction of satisfied reviews using RFE-SHAP:

For the input values, we employed the probabilities of 25 distinct topics, enhanced by additional features such as ‘vote’ (coded as 1 for presence and 0 for absence), ‘style’ (a categorical variable), and ‘image’ (coded as 1 for presence and 0 for absence). A suite of machine learning models, including Logistic Regression, Decision Trees, Neural Networks, Support Vector Machines, Random Forests, and eXtreme Gradient Boosting (XGBoost), was trained using this refined feature set from the RFE-SHAP process. In this research, significant emphasis was placed on the feature selection step, a critical aspect of machine learning that influences model performance and interpretability. Effective feature selection reduces model complexity, improves accuracy, and aids in understanding how different predictors contribute to the outcome. To evaluate the effectiveness of feature selection methods, we compared Recursive Feature Elimination (RFE) with RFE combined with SHapley Additive exPlanations (RFE-SHAP). RFE methodically eliminates features, while RFE-SHAP further enriches this process by integrating SHAP values. These values, based on game theory principles (Jing et al., 2023), offer detailed insights into feature contributions, enhancing interpretability and the understanding of feature relationships within the model. This comprehensive approach underlines the importance of choosing the right features to optimize the model’s predictive accuracy and offers a deeper exploration of how these features impact the predictions.

Our comparison aims to shed light on the superior capability of RFE-SHAP to refine model performance, ensuring that the features selected contribute most effectively to the predictive accuracy, while also allowing for a transparent evaluation of how individual features impact model outputs. The novel RFE-SHAP approach was employed, synergizing Recursive Feature Elimination (RFE) with Shapley Additive exPlanations (SHAP). RFE is a feature selection algorithm that recursively removes attributes to rank them based on their importance to the model’s prediction (Samb et al., 2012; Guo et al., 2023). It operates by fitting the model repeatedly, each time removing the least important feature(s) until a specified number of features is reached. SHAP values provide a unified measure of feature importance by computing the average contribution of each feature across all possible combinations in which it can be included (Pelegrina et al., 2023). The equation SHAP is shown below:

$$SHAP\ value\ = \sum_{\mathrm{S \subseteq M\{i\}}} \frac{|S|!\ (|M|\ -\ |S|\ -\ 1)!}{|M|!}\ [f_x(S\ \cup\ \{i\})\ -\ f_x(S)]$$

In this equation:

- *S* is a subset of features used in the prediction model.
- *M* is the set of all features in the model.
- ｜*S*｜ is the number of features in subset *S*.
- *i* represents a specific feature for which the SHAP value is being calculated.
- $f_x(S\ \cup\ \{i\})$ is the prediction model's output when the feature set includes *i* along with the features in set *S*.
- $f_x(S)$ is the prediction model's output when the feature set is just *S*, excluding feature *i*.

We employed a diverse array of machine learning models, including Logistic Regression, Decision Trees, Neural Networks, Support Vector Machines, Random Forests, and XGBoost, all trained using a refined feature set obtained from the RFE-SHAP process. To evaluate the efficacy of these models, we used key metrics like accuracy, precision, recall, and F1-score, aiming to gauge their predictive accuracy on our dataset.

RFE-SHAP, an innovative combination of RFE and SHAP, diverges from conventional feature importance methods. The synergy of these methodologies allowed for the iterative elimination of the least impactful features based on their SHAP values, which quantify the contribution of each feature to the model's prediction. This process leverages the explanatory power of SHAP values to iteratively rank and eliminate features, thus enhancing the model's performance and interpretability.

The procedure for RFE-SHAP is as follows:

1. Compute SHAP values for each feature across the dataset, which quantify the impact of each feature on the model's output.
2. Rank the features based on their absolute average SHAP value.
3. Eliminate the feature(s) with the lowest SHAP value(s).
4. Continuously repeat this process on the increasingly reduced feature set until the desired number of features is retained.

In this approach, for a given set of features $F=\{f_1,f_2,\cdots,f_n\}$, the computation of SHAP values results in a corresponding set of importance scores $I_{SHAP}=\{i_{s1},i_{s2},\cdots,i_{sj}\}$. In each iteration, the feature f_j with the minimum score i_{sj} in I_{SHAP} is removed, thereby refining the feature set for subsequent model training.

Explanation of prediction results:

The last step of the research involved evaluating the model with the highest performance. Leveraging SHAP, the top five most influential topics were discerned. Based on the keywords intrinsic to each topic, they were aptly named, providing a comprehensive understanding of the elements most instrumental in predicting satisfied reviews.

Utilizing the SHapley Additive exPlanations (SHAP) methodology (Herrera et al., 2023), we delved into the model to discern the top five topics that wielded the most significant influence on the predictions. SHAP values, by design, offer a granular understanding of how each feature (in this case, topic) impacts the model's predictions. By ranking the topics based on their SHAP values, we were able to identify those that were most instrumental in predicting satisfied reviews.

Once the top five topics were identified, the next challenge was to interpret and name them. For this, we closely examined the keywords associated with each topic. These keywords, which are essentially the most frequently occurring words within each topic, provided invaluable insights into the essence of the topic. By analyzing these keywords in conjunction with the context of the reviews, we were able to assign meaningful names to each topic, capturing their underlying themes.

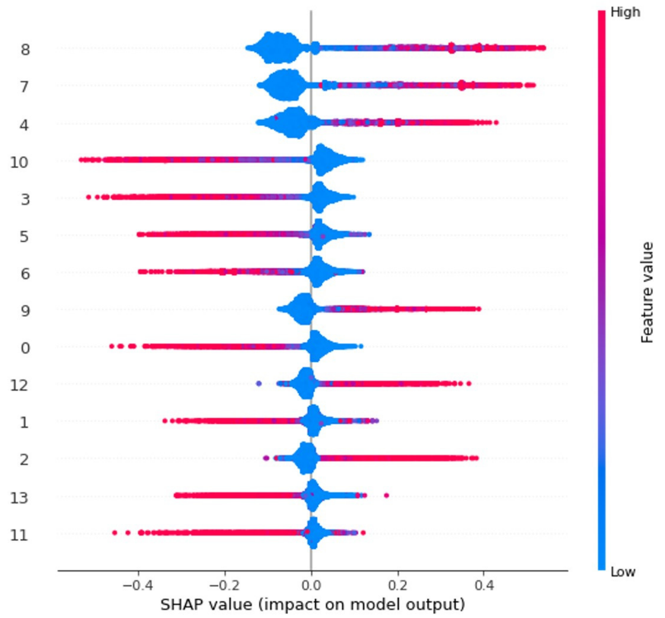
For instance, if one of the top topics had keywords like "fit," "comfortable," and "size," it might be named "Fit and Comfort." Similarly, a topic with keywords like "photo," "material," and "disappointed" could be named " Quality & Appearance." In essence, this comprehensive analysis not only spotlighted the most influential topics but also provided a deep understanding of the factors that drive customer satisfaction in the realm of online clothing reviews on Amazon.

4. Results

The results derived from the prediction of satisfied reviews utilizing the RFE-SHAP methodology are presented in Table 1(AUC curve results are shown in Appendix B). Among the various models evaluated,

〈Table 1〉 Prediction results of satisfied reviews

Model	Feature selection algorithm	Number of features in the optimal subset	Accuracy	F1	Recall	Precision
Logit	RFE-SHAP	18	0.824	0.825	0.808	0.842
	RFE	19	0.803	0.805	0.796	0.838
Decision Tree	RFE-SHAP	20	0.796	0.795	0.773	0.819
	RFE	20	0.788	0.779	0.767	0.791
Random Forest	RFE-SHAP	14	0.846	0.847	0.833	0.860
	RFE	16	0.839	0.839	0.823	0.855
NN	RFE-SHAP	14	0.826	0.834	0.847	0.805
	RFE	16	0.822	0.829	0.840	0.801
SVM	RFE-SHAP	15	0.826	0.834	0.847	0.805
	RFE	17	0.815	0.827	0.839	0.801
XGBoost	RFE-SHAP	16	0.840	0.841	0.827	0.853
	RFE	18	0.829	0.835	0.822	0.846



〈Figure 2〉 SHAP results of RFE-SHAP Random Forest prediction model

the Random Forest classifier emerged as the most efficacious, demonstrating superior performance metrics. Notably, the model achieved optimal performance with a reduced subset of 14 features, a significant reduction from the original 28 features, and achieved performance metrics of Accuracy: 0.846, F1 Score: 0.847, Recall: 0.833, and Precision: 0.860. These metrics underscore the model's robustness in predicting satisfied reviews, with a commendable balance between precision and recall, ensuring that the model is neither overly conservative nor overly aggressive in its predictions.

Figure 2 presents the SHAP results from the RFE-SHAP application of the Random Forest model, encapsulating all fourteen features integral to predicting customer satisfaction. This visualization facilitates a profound understanding of each feature's impact, offering stakeholders substantial insights for strategic decision-making. The granular analysis of topics, detailed through the top twenty keywords and their term weights, is presented in Table 2 for the primary five topics and expanded upon for the remaining nine in Appendix A. These keywords offer a lens into the intrinsic nature of each topic, aiding in their precise identification and classification.

The top-5 feature set further, topics 8 (Quality & Appearance), 7 (Fit & Comfort), 4 (Durability Concerns), 10 (Comfort & Style), and 3 (Quality & Materials) emerge as the most influential in determining satisfied customer reviews. The remaining nine features—Topic 5 (Quality Insights), Topic 6 (Size Perception), Topic 9 (Gift Favorites), Topic

0 (Fit and Style), Topic 12 (Product Durability), Topic 1 (Quality Assessment), Topic 2 (Size Variability), Topic 13 (Cherished Gifts), and Topic 11 (Product Reliability)—further enriches the predictive narrative. These features are comprehensively documented in Appendix A, which includes a table enumerating the top keywords and their weights for each topic.

The SHAP values serve as a metric of feature impact, providing a quantifiable measure of each attribute's influence on customer satisfaction predictions. This analytical approach allows for the pragmatic application of the model's insights to real-world scenarios, where the prediction of satisfaction from review text can be leveraged to inform business strategies and enhance customer engagement. The most impactful topics encapsulate the primary concerns and praises echoed by customers in their reviews. Notably, aspects related to the quality, fit, comfort, and appearance of the products were recurrent themes, underscoring their significance in driving customer

〈Table 2〉 Keywords and weight of top-5 topics

Topic 8 Quality & Appearance		Topic 7 Fit & Comfort		Topic 4 Durability Concerns		Topic 10 Comfort & Style		Topic 3 Quality & Materials	
keyword	weight	keyword	weight	keyword	weight	keyword	weight	keyword	weight
like	2446.663	pant	529.581	broke	1022.04	work	583.388	horrible	594.016
look	1710.047	comfortable	256.552	money	929.229	great	317.086	strap	247.476
picture	1374.75	wear	219.112	waste	676.729	use	187.382	smell	232.131
dress	1186.788	legging	182.842	time	640.744	need	145.879	like	160.4
star	447.741	jean	180.329	star	581.46	nicely	120.208	china	159.412
quality	323.014	leg	131.438	broken	366.378	little	88.016	pay	121.969
looked	181.991	fit	122.851	day	313.082	heavy	85.113	return	118.785
photo	168.347	tight	109.906	chain	226.081	look	84.565	cheap	114.098
exactly	152.417	pair	109.698	wore	220.093	wear	81.245	bad	101.052
cheap	151.839	feel	108.767	cheap	196.076	better	75.879	way	101.011
poor	145.54	look	104.396	wear	167.703	hair	68.698	really	90.82
pic	103.679	waist	104.306	zipper	161.772	expected	65.318	product	88.046
product	97.778	stay	96.561	worth	155.109	wig	61.04	poorly	81.569
looking	94.542	love	87.47	ear	153.15	clip	59.705	thing	79.336
item	94.044	time	86.635	came	136.02	nose	59.679	away	77.662
pictured	86.441	like	84.926	week	131.506	feel	59.294	shape	75.821
good	84.593	inch	81.094	got	131.371	value	59.191	bag	74.863
person	82.602	work	78.859	clasp	121.794	nice	58.535	came	72.252
material	82.168	definitely	72.638	second	121.042	bulky	56.492	elastic	71.981
disappointed	74.942	high	69.342	wearing	120.639	lovely	55.645	going	70.901

satisfaction in the realm of online clothing reviews.

Expanding on the significance of the SHAP values, they provide an empirical foundation to quantify the import of each predictive feature, facilitating a nuanced approach to model interpretation. This empirical evidence affords a granular perspective on customer reviews, translating qualitative feedback into quantifiable data points. Such a data-driven approach ensures that each element of customer feedback is weighted according to its predictive power, revealing the underlying drivers of satisfaction. This, in turn, enables a targeted strategy for enhancing product offerings and customer service practices, pivotal in an industry where customer preferences are rapidly evolving. Through this analytical lens, businesses can adapt to market trends with agility, ensuring that they not only meet but exceed customer expectations, fostering loyalty and establishing a competitive edge in the online retail sector.

5. Conclusion and Discussion

Our study makes a significant contribution to the field of e-commerce analytics by combining theoretical insights with practical applications, while also paving the way for future research. Firstly, we tackle the common issue of overfitting in predictive modeling, particularly in scenarios with limited training data. Although the integration of RFE-SHAP feature selection with Latent Dirichlet Allocation (LDA) for topic modeling is not entirely new, our application in this context is distinctive. We utilize this combination to focus more effectively on the most influential features in our predictive model, leading to a more nuanced understanding of e-commerce review content. Secondly, regarding the use of the Random Forest model, our empirical results demonstrate significant improvements in its performance. We achieve this by fine-tuning the model with a carefully selected subset of 14 key features. This is not merely an incremental improvement but a strategic optimization that significantly enhances the model’s predictive accuracy and efficiency. However, it is essential to clarify how our methodology contributes to structurally optimizing the Random Forest algorithm. Our approach involves a systematic feature reduction process that not only improves the model’s performance metrics but also reduces computational complexity and enhances interpretability. By doing so, we address the often-neglected aspect of model optimization in the context of e-commerce analytics, offering a method that is both effective in practice and valuable for academic discourse.

Furthermore, the discerned topics, such as Quality & Appearance, Fit & Comfort, Durability Concerns, Comfort & Style, and Quality & Materials, shed light on the primary drivers of customer satisfaction in the realm of clothing. Third, we underscore the importance of strategic feature selection in predictive modeling. Our research demonstrates a structured approach, using the RFE-SHAP methodology, to judiciously eliminate redundant features, thereby bolstering model robustness. Fourth, our work stands as a pioneering effort in merging RFE-SHAP with topic modeling, specifically LDA. This amalgamation crafts a holistic framework that encapsulates both the quantitative and qualitative intricacies of customer reviews, laying the groundwork for subsequent research in this arena. Lastly, by delving into the most influential topics, we furnish a profound understanding of the determinants of customer satisfaction. This offers a blueprint for ensuing studies keen on dissecting and interpreting online reviews.

From a methodological perspective, our research underscores the benefits of combining RFE-SHAP with LDA topic modeling. This unique fusion offers a comprehensive framework that captures both quantitative and qualitative aspects of data analysis, a significant step forward in predictive modeling, especially in the e-commerce sector. Practically, our study yields essential insights. Identifying key topics that influence customer satisfaction enables businesses to refine their strategies and offerings, leading to improved customer loyalty and increased sales. Furthermore, these insights can guide targeted marketing campaigns, enhancing customer engagement and maximizing marketing ROI. However, our research is not without limitations. The data, sourced exclusively from Amazon and centered on the clothing sector, anchors our findings to this specific context. Extrapolating these insights to other platforms or product categories warrants prudence. Additionally, given the dynamic nature of customer preferences and sentiments, our research, while providing a contemporary snapshot, might require periodic updates to remain relevant as these insights evolve over time.

Looking ahead, future research can expand on our work in several ways. Applying our methodology across different e-commerce platforms and product categories can validate the universality of our findings. Longitudinal studies would provide insights into how customer preferences evolve. Incorporating data from various customer interaction platforms could offer a more comprehensive view of customer behavior. Advanced machine learning techniques could further enhance model accuracy and interpretability. Understanding the impact of changing market dynamics, addressing ethical and privacy considerations, and assessing the practical implementation of these methodologies in real-world settings are other crucial areas for future exploration. In summary, while our study offers valuable insights into e-commerce analytics, it also lays the groundwork for future research to explore and understand the complex dynamics of online customer behavior in a rapidly evolving digital marketplace.

References

Aakash, A., & Gupta Aggarwal, A. (2022). Assessment of hotel performance and guest satisfaction through eWOM: big data for better insights. International Journal of Hospitality & Tourism Administration, 23(2), 317-346.

Alzahrani, S., Wang, Q., & Rana, O. (2022). Latent Dirichlet Allocation for Customer Satisfaction Analysis in Online Reviews. Journal of E-Commerce Research, 16(2), 145-158.

Ansari, G., Gupta, S., & Singhal, N. (2020). Natural Language Processing in Online Reviews. Journal of E-commerce and Digital Marketing, 8(1), 34-47.

Bauer, J., & Jannach, D. (2021). Improved Customer Lifetime Value Prediction With Sequence-To-

Sequence Learning and Feature-Based Models. Journal of E-commerce Research, 21(3), 45-60.

Chen, J., Yuan, S., Lv, D., & Xiang, Y. (2021). A novel self-learning feature selection approach based on feature attributions. Expert Systems with Applications, 183, 115219.

Chen, Y., & Xie, J. (2008). Online customer review: Word-of-mouth as a new element of marketing communication mix. Management science, 54(3), 477-491.

Chernyaeva, O. ., & Hong, T. . (2022). The Detection of Online Manipulated Reviews Using Machine Learning and GPT-3. Journal of Intelligence and Information Systems, 28(4), 347-364.

Cui, G., Lui, H. K., & Guo, X. (2012). The effect of online customer reviews on new product sales. International Journal of Electronic Commerce, 17(1), 39-58.

Darko, A. P., & Liang, D. (2022). Modeling customer satisfaction through online reviews: A FlowSort group decision model under probabilistic linguistic settings. Expert Systems with Applications, 195, 116649.

Darst, B. F., Malecki, K. C., & Engelman, C. D. (2018). Using recursive feature elimination in random forest to account for correlated variables in high dimensional data. BMC genetics, 19(1), 1-6.

Ding, X., Yang, F., & Ma, F. (2022). An efficient model selection for linear discriminant function-based recursive feature elimination. Journal of Biomedical Informatics, 129, 104070.

Du, C., & Huang, L. (2018). Text classification research with attention-based recurrent neural networks. International Journal of Computers Communications & Control, 13(1), 50-61.

Engler, T. H., Winter, P., & Schulz, M. (2015). Understanding online product ratings: A customer satisfaction model. Journal of Retailing and Customer Services, 27, 113-120.

Guo, J., Wang, Z., Jin, Y., Li, M., & Chen, Q. (2023). Predicting and extracting thermal behavior rules of hydronic thermal barrier with interpretable ensemble learning in the heating season. Energy and Buildings, 113699.

He, J., Hu, D., Zhang, W., & Liu, T. (2020). Probabilistic Topic Modeling for Sentiment Analysis of Online Reviews. Journal of Business Analytics, 7(3), 210-227.

Herrera, G. P., Constantino, M., Su, J. J., & Naranpanawa, A. (2023). The use of ICTs and income distribution in Brazil: A machine learning explanation using SHAP values. Telecommunications Policy, 47(8), 102598.

Hong, A. C. Y., Khaw, K. W., Chew, X., & Yeong, W. C. (2023). Prediction of US airline passenger satisfaction using machine learning algorithms. Data Analytics and Applied Mathematics (DAAM), 8-24.

Jing, H., Yang, P., & Lin, H. (2023). A Multilayer Stacking Method Base on RFE-SHAP Feature Selection Strategy for Recognition of Driver's Mental Load and Emotional State. Expert Systems with Applications, 121729.

Johar, S., & Mubeen, S. (2020). Sentiment analysis on large scale Amazon product reviews. IJSRCSE, 8(1), 7-15.

Kang, D., & Park, Y. (2014). based measurement of customer satisfaction in mobile service: Sentiment analysis and VIKOR approach. Expert Systems with Applications, 41(4), 1041-1050.

Kannari, P. R., Chowdary, N. S., & Biradar, R. L. (2022). An anomaly-based intrusion detection system using recursive feature elimination technique for improved attack detection. Theoretical Computer Science, 931, 56-64.

Karim, A., & Das, R. (2018). Rule-based vs. Machine Learning: A Comparative Study on Sentiment Analysis and LDA. International Journal of Data Science, 5(1), 56-68.

Kumar, S., Yadava, M., & Roy, P. (2019). Fusion of EEG response and sentiment analysis of products review to predict customer satisfaction. Information Fusion, 47, 124-133.

Lin, C. L., Lee, S. H., & Horng, D. J. (2011). The effects of online reviews on purchasing intention: The moderating role of need for cognition. Social Behavior and Personality: an international journal, 39(1), 71-81.

Liu, B., Zhou, X., Jiang, P., & Zhang, L. (2020). Customer Satisfaction in B2C E-commerce: An LDA Approach. E-Commerce Research and Applications, 14(4), 301-315.

Liu, M., Lu, X., & Song, J. (2016). A New Feature Selection Method for Text Categorization of Customer Reviews. E-commerce Research Letters, 10(1), 5-15.

Maharani, A.P., & Triayudi, A. (2022). Sentiment Analysis of Indonesian Digital Payment Customer Satisfaction Towards GOPAY, DANA, and ShopeePay Using Naïve Bayes and K-Nearest Neighbour Methods. Management and Informatics Business Journal, 6(1), 1-10.

Matuszelański, K., & Kopczewska, K. (2022). Customer Churn in Retail E-Commerce Business: Spatial and Machine Learning Approach. International Journal of E-commerce Studies, 15(2), 120-138.

Mudambi, S. M., & Schuff, D. (2010). Research note: What makes a helpful online review? A study of customer reviews on Amazon. com. MIS quarterly, 185-200.

Park, S., & Lee, S.-Y. T. (2023). A Study on the Relationship between Social Media ESG Sentiment and Firm Performance. Journal of Intelligence and Information Systems, 29(3), 317-340.

Park, Y.-J., & Kim, K.-j. (2017). Impact of Semantic Characteristics on Perceived Helpfulness of Online Reviews. Journal of Intelligence and Information Systems, 23(3), 29-44.

Pelegrina, G. D., Duarte, L. T., & Grabisch, M. (2023). A k-additive Choquet integral-based approach to approximate the SHAP values for local interpretability in machine learning. Artificial Intelligence, 325, 104014.

Ren, Y., Wang, R., & Ji, D. (2016). A topic-enhanced word embedding for Twitter sentiment classification. Information Sciences, 369, 188-198.

Ren, Y., Wang, R., & Ji, D. (2016). A topic-enhanced word embedding for Twitter sentiment classification. Information Sciences, 369, 188-198.

Samb, M. L., Camara, F., Ndiaye, S., Slimani, Y., & Esseghir, M. A. (2012). A novel RFE-SVM-based feature selection approach for classification. International Journal of Advanced Science and Technology, 43(1), 27-36.

Uthirapathy, S. E., & Sandanam, D. (2023). Topic Modelling and Opinion Analysis On Climate Change Twitter Data Using LDA And BERT Model. Procedia Computer Science, 218, 908-917.

Van den Broeck, G., Lykov, A., Schleich, M., & Suciu, D. (2022). On the tractability of SHAP explanations. Journal of Artificial Intelligence Research, 74, 851-886.

Wisnu, H., Afif, M., & Ruldevyani, Y. (2020). Sentiment analysis on customer satisfaction of digital payment in Indonesia: A comparative study using KNN and Naïve Bayes. Journal of Physics: Conference Series, 1444(1), 012034.

Xu, H., Li, Z., Chu, C., Chen, Y., Yang, Y., Lu, H., Wang, H., & Stavrou, A. (2018). Detecting and Characterizing Web Bot Traffic in a Large E-commerce Marketplace. International Journal of E-commerce Research, 16(3), 201-218.

Yu, D., Fang, A., & Xu, Z. (2023). Topic research in fuzzy domain: Based on LDA topic modelling. Information Sciences, 648, 119600.

Zhang, J., Ma, X., Zhang, J., Sun, D., Zhou, X., Mi, C., & Wen, H. (2023). Insights into geospatial heterogeneity of landslide susceptibility based on the SHAP-XGBoost model. Journal of Environmental Management, 332, 117357.

Zhang, M., & Luo, L. (2023). Can customer-posted photos serve as a leading indicator of restaurant survival? Evidence from Yelp. Management Science, 69(1), 25-50.

Appendix A

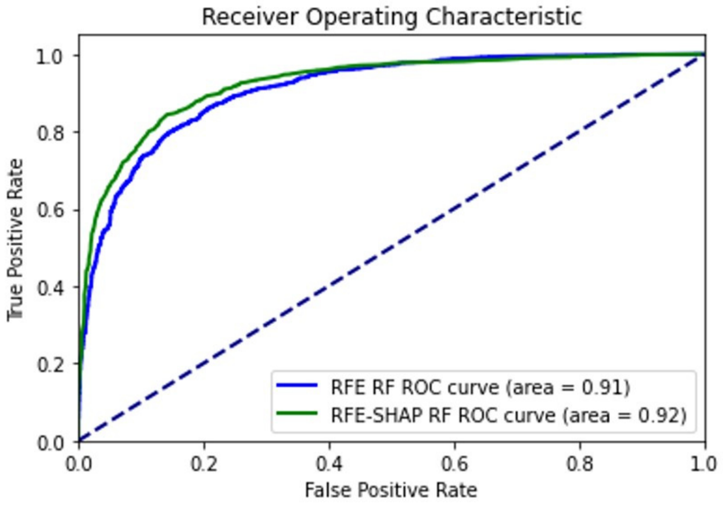
Topic 5 Quality Insights		Topic 6 Size Perception		Topic 9 Gift Favourites		Topic 0 Fit and Style		Topic 12 Product Durability	
keyword	weight	keyword	weight	keyword	weight	keyword	weight	keyword	weight
quality	1940.027	size	2938.658	loved	497.385	short	637.607	product	355.191
good	1080.961	small	2333.271	year	475.450	shirt	624.641	buy	303.731
star	986.333	ordered	1115.078	old	464.569	like	534.029	recommend	278.343
poor	871.932	large	816.159	little	413.111	fit	476.187	terrible	267.153
bad	663.095	way	797.222	bra	256.856	long	320.425	apart	261.933
fabric	268.464	fit	794.140	daughter	195.604	tight	311.630	worst	214.828
product	189.499	run	516.495	fit	188.380	skirt	306.160	star	200.718
low	182.250	wear	417.378	bought	149.796	wear	295.946	worth	173.773
small	180.696	medium	398.307	got	127.770	length	264.324	cheap	153.464
cheap	147.090	disappointed	387.935	girl	120.851	sleeve	260.040	fell	148.926
look	146.400	order	363.800	bit	102.165	way	241.149	seam	121.078
cheaply	137.606	big	356.238	gift	96.071	dress	232.612	falling	119.676
returned	120.963	xl	280.676	granddaughter	86.122	small	232.588	dont	116.446
disappointed	112.518	sizing	270.000	tie	84.677	arm	227.622	money	110.796
return	107.852	like	267.290	kid	84.143	fabric	222.091	started	106.253
shirt	97.384	return	255.078	sister	73.792	sweater	187.367	wash	106.212
buy	94.806	review	222.353	gave	72.253	look	187.311	month	104.848
really	85.159	bigger	209.478	big	65.378	disappointed	171.335	week	103.941
pay	80.068	larger	206.892	niece	65.347	little	153.041	material	101.470
terrible	77.492	chart	196.716	snug	63.988	cut	152.166	item	96.934

Appendix A(continued)

Topic 1 Quality Assessment		Topic 2 Size Variability		Topic 13 Cherished Gifts		Topic 11 Product Reliability	
keyword	weight	keyword	weight	keyword	weight	keyword	weight
quality	1940.027	size	2938.658	loved	497.385	product	355.191
good	1080.961	small	2333.271	year	475.450	buy	303.731
star	986.333	ordered	1115.078	old	464.569	recommend	278.343
poor	871.932	large	816.159	little	413.111	terrible	267.153

Topic 1 Quality Assessment		Topic 2 Size Variability		Topic 13 Cherished Gifts		Topic 11 Product Reliability	
keyword	weight	keyword	weight	keyword	weight	keyword	weight
bad	663.095	way	797.222	bra	256.856	apart	261.933
fabric	268.464	fit	794.140	daughter	195.604	worst	214.828
product	189.499	run	516.495	fit	188.380	star	200.718
low	182.250	wear	417.378	bought	149.796	worth	173.773
small	180.696	medium	398.307	got	127.770	cheap	153.464
cheap	147.090	disappointed	387.935	girl	120.851	fell	148.926
look	146.400	order	363.800	bit	102.165	seam	121.078
cheaply	137.606	big	356.238	gift	96.071	falling	119.676
returned	120.963	xl	280.676	granddaughter	86.122	dont	116.446
disappointed	112.518	sizing	270.000	tie	84.677	money	110.796
return	107.852	like	267.290	kid	84.143	started	106.253
shirt	97.384	return	255.078	sister	73.792	wash	106.212
buy	94.806	review	222.353	gave	72.253	month	104.848
really	85.159	bigger	209.478	big	65.378	week	103.941
pay	80.068	larger	206.892	niece	65.347	material	101.470
terrible	77.492	chart	196.716	snug	63.988	item	96.934

Appendix B



국문요약

RFE-SHAP을 활용한 온라인 리뷰를 통한 고객 만족도 예측

체르냐예바 올가[*] · 홍태호[**]

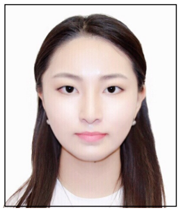
본 연구는 온라인 리뷰를 이용하여 고객 만족도를 예측하는 새로운 접근 방식을 제안한다. LDA 주제 모델링과 결합된 RFE-SHAP 기능 선택 방법을 활용하여 고객 만족도에 큰 영향을 미치는 주요 기능을 식별하여 예측 분석을 개선했다. 먼저 Random Forest 알고리즘의 경우, 초기 28개 입력변수에서 14개의 변수를 최적 하위 집합으로 추출했다. 제안된 방법에서 Random Forest 모델의 성과는 84%로 확인 되었으며 변수가 많은 모델에서 흔히 발생하는 과적합을 방지하였다. 또한 품질, 착용감, 내구성 등과 같은 리뷰의 특정 요소들이 패션 산업 내에서 소비자 만족도를 증진시키는 중요한 역할을 한다는 사실을 밝혀냈다. 본 연구는 예측 결과를 설명할 때 선택한 각 기능이 고객 만족도에 어떻게 영향을 미치는지에 대한 자세한 설명을 제공하고 고객이 가장 중요하게 생각하는 측면에 대한 세부적인 보기를 제공한다. 본 연구의 공헌도는 다음과 같다. 첫째, 전자상거래 분석 분야 내에서 예측 모델링을 강화하고 특성 중심적인 접근법을 소개함으로써 방법론을 개선하였다. 이는 고객 만족도 예측의 정확도를 높일 뿐만 아니라 예측 모델에서의 변수 선택에 대한 새로운 접근을 제시한다. 둘째, 특히 의류 부문에서 전자상거래 플랫폼에 구체적인 통찰력을 제공한다. 품질, 사이즈, 내구성 등 고객 리뷰의 어떤 부분이 만족도에 가장 큰 영향을 미치는지 강조함으로써, 기업들이 제품과 서비스를 맞춤화 할 수 있는 전략적 방향을 제시한다. 이러한 목표 지향적인 개선은 고객의 쇼핑 경험을 개선하고, 만족도를 향상시키면서 충성도를 이끌어낼 수 있을 것으로 기대한다.

주제어 : 고객 만족; 특성 선택; SHAP; RFE; 온라인 리뷰; 토픽 모델링

논문접수일 : 2023년 11월 13일　논문수정일 : 2023년 12월 20일　게재확정일 : 2023년 12월 21일
원고유형 : Regular Track　교신저자 : 홍태호

[*] 부산대학교 경영학부
[**] 교신저자 : 홍태호
부산대학교 경영학부
부산광역시 금정구 부산대학로63번길 2
Tel: +82-51-510-2531, E-mail: hongth@pusan.ac.kr

저 자 소 개



Olga Chernyaeva is a Ph.D. student of Management Information Systems at the College of Business Administration, Pusan National University. She received her Master's degree from Pusan National University. Her research interests include business analytics, intelligent systems, data mining, and recommender systems for e-business. Her work has been published in the Asia Pacific Journal of Information Systems and the Journal of Intelligence and Information Systems.

Taeho Hong is a Professor of Management Information Systems at the College of Business Administration, Pusan National University. He received his Ph.D. from the Korea Advanced Institute of Science and Technology. His research interests include intelligent systems, data mining, and recommender systems for e-business. His work has been published in Expert Systems with Application, Expert Systems, and Information Processing & Management.