TEKNOBUGA Vol. 13 No.1 (2025)

p-ISSN 2085 4056
e-ISSN 2528 7087

JURNAL TEKNOLOGI BUSANA DAN BOGA

https://journal.unnes.ac.id/journals/teknobuga/index

Making Sense of Fashion Feedback: Comparing Two Popular Text Analysis Tools

Muhammad Syafiq*, Wawan Saputra, Carlya Agmis Aimandiga, Cici Suhaeni, Bagus Sartono, and Gerry Alfa Dito

Statistics and Data Science Program Study, School of Data Science, Mathematics, and Informatics, IPB University

*Corresponding author: muhammadsyafiq@apps.ipb.ac.id

ABSTRACT - The rapid expansion of the fashion industry, propelled by digital technology and e-commerce, has resulted in a significant volume of customer-generated reviews. These reviews serve as a valuable source for understanding customer satisfaction and behavior. This study aims to (1) analyze customer sentiment, (2) predict product recommendations, and (3) examine the relationship between sentiment classification and recommendation decisions using text embeddings from Word2Vec and GloVe. The research utilized over 23,000 fashion product reviews sourced from Kaggle. Text data were preprocessed and vectorized using Word2Vec and GloVe, followed by classification and prediction tasks using six machine learning models: Random Forest, SVM, Naïve Bayes, LSTM, Logistic Regression, and Gradient Boosting. The results revealed that Word2Vec consistently outperformed GloVe across all models and tasks, with the Word2Vec-LSTM combination achieving the highest accuracy of 87.35% and F1 score of 92.35% in imbalanced data scenarios. Correlation analysis also confirmed a strong and statistically significant relationship between sentiment and recommendation labels, with Spearman's Rho of 0.8340 and Kendall's Tau of 0.8120. These findings suggest that high-quality sentiment representation can effectively support product recommendation systems. This study contributes to the understanding of embedding effectiveness in fashion-related text analysis and opens avenues for hybrid and transformer-based representations in future research.

Keywords: Fashion reviews, sentiment analysis, product recommendation, Word2Vec, GloVe.

INTRODUCTION

The fashion industry continues to grow rapidly with advancements in digital technology and the increasing use of e-commerce platforms. This growth facilitates customers in providing reviews of the products they have purchased, whether through comments, ratings, or recommendations. Customer reviews are highly valuable data sources as they offer insights into satisfaction levels, preferences, and experiences with specific products (Saragih *et al*., 2022). In the fashion industry, sentiment analysis of customer reviews is crucial as it helps companies understand market perceptions of their products. Additionally, product recommendation predictions can be used to optimize marketing strategies and enhance the shopping experience through more accurate product personalization (Gupta & Bhatnagar, 2023).

The application of sentiment analysis and product recommendation prediction techniques has become more effective with advancements in text representation methods such as Word2Vec and GloVe. Word2Vec is an embedding technique capable of capturing semantic relationships between words based on their context (Mikolov *et al*., 2013). On the other hand, GloVe offers a *embedding* approach based on global co-occurrence matrices, producing high-quality vector representations that identify semantic and syntactic relationships between words (Pennington *et al*., 2014). Several previous studies have utilized text representation techniques like Word2Vec and GloVe for customer sentiment analysis. Research by Adewopo *et al*. (2021) evaluated the effectiveness of Word2Vec and GloVe in sentiment classification of product reviews using machine learning algorithms such as SVM and Random Forest. The results showed that Word2Vec achieved slightly higher accuracy than GloVe in most models, although the difference was not significant. This indicates that the type of embedding can influence classification performance, depending on the model architecture used. Meanwhile, Patil & Atique (2020) compared Word2Vec and GloVe in the context of sentiment classification for fashion product reviews using a deep learning approach. They used BiLSTM as the base model and found that the combination of BiLSTM + Word2Vec achieved better accuracy than BiLSTM + GloVe, with a difference of up to 2% on the e-commerce review dataset. This study confirms that sequential architectures like BiLSTM can explore embeddings more optimally. On the other hand, research by Singh *et al*. (2022) explored the relationship between customer sentiment and their tendency to recommend products based on review texts. The study used Word2Vec as input features for a binary classification model to predict whether customers would recommend a product or not. *Logistic Regression* and *Gradient Boosting* were used as classification algorithms, and the results showed that sentiment features had a strong correlation with customer recommendation behavior.

However, although many studies have discussed the application of Word2Vec and GloVe for sentiment analysis, few have specifically explored and compared both embeddings comprehensively in the dual context of sentiment classification and product recommendation prediction, particularly within the fashion domain. Moreover, existing works often focus on either classical or deep learning models—but not both—and rarely evaluate model performance under different data balance conditions. Therefore, this study presents a novel contribution by simultaneously addressing these gaps: (1) evaluating the effectiveness of Word2Vec and GloVe representations across six machine learning models (including both classical and deep learning approaches), (2) analyzing performance in both imbalanced and balanced datasets, and (3) examining the statistical relationship between sentiment and recommendation labels. The main questions addressed in this study are: how do machine learning models perform in classifying customer sentiment toward fashion products based on Word2Vec and GloVe text representations; how do machine learning models perform in predicting whether customers will recommend fashion products based on review texts; and to what extent is there a significant relationship between sentiment classification and recommendation decisions in customer reviews.

METHOD

Dataset

This study employed secondary data from the publicly available "Women's E-Commerce Clothing Reviews" Customer Dataset obtained from Kaggle (https://www.kaggle.com/datasets/nicapotato/womens-ecommerce-clothing-reviews. The dataset comprises 23,486 customer reviews evaluating various women's fashion products. Three key features were utilized in the analysis:

1. Review Text: The primary textual feature containing customers' written evaluations of products.
2. Rating: Numeric sentiment labels on a 5-point scale, categorized as:
   Negative sentiment: Ratings 1-2
   Neutral sentiment: Rating 3
   Positive sentiment: Ratings 4-5
3. Recommended IND: A binary recommendation indicator where:
   0 = Customer does not recommend the product
   1 = Customer recommends the product

Research Method

The analytical steps undertaken to complete this research are as follows:

1. **Collecting the Customer Dataset,** specifically the *Women's E-Commerce Clothing Reviews* obtained from the Kaggle website.
2. **Pre-processing**
   The preprocessing stage aims to clean the text data from irrelevant elements and transform it into a format suitable

for machine learning models. The first step is expanding contractions (e.g., "don't" becomes "do not") to maintain consistency and clarity in word usage. URLs are then removed as they do not provide meaningful semantic value. Next, all text is converted to lowercase to avoid duplication due to case sensitivity (e.g., "Good" and "good" are treated as the same word). Punctuation marks and numerical digits are removed, followed by tokenization, which breaks down the text into individual words.

The next stage involves removing stopwords, including both default English stopwords from NLTK and fashion domain-specific stopwords, to focus on more informative content. Repeated characters (e.g., "niiiice") and duplicate words are also reduced to minimize redundancy. Stemming is applied to reduce words to their root forms, which helps in unifying word variants. The cleaned words are then reconstructed into a single string for consistency, and finally, the text is split back into tokens for subsequent analysis stages using embedding models like Word2Vec and GloVe.

3. **Text Representation**

After the text data has been cleaned through preprocessing, the next step is to convert the text into numerical form so it can be used by machine learning models. In this study, two main techniques are applied: Word2Vec and GloVe. Word2Vec is a neural network-based word representation method developed by Mikolov and colleagues (2013) at Google. This method represents each word in a corpus as a low-dimensional vector in a continuous space, using two main approaches: Continuous Bag of Words (CBOW) and Skip-Gram. The CBOW model predicts a target word based on its surrounding context, while Skip-Gram does the opposite, predicting the context from the target word. Word2Vec effectively captures both semantic and syntactic relationships between words. For example, word vectors can demonstrate linear relationships such as: vector("king") - vector("man") + vector("woman") ≈ vector("queen") (Mikolov et al., 2013). The use of Word2Vec in various studies has shown its effectiveness in improving the accuracy of NLP tasks such as text classification and sentiment analysis (Li & Xu, 2020; Dewi & Nugroho, 2021).

Unlike Word2Vec's predictive approach, GloVe is a count-based method developed by Pennington, Socher, and Manning (2014) at Stanford. GloVe combines the statistical approach of a co-occurrence matrix with the advantages of learning-based embeddings. GloVe's vector representations rely on the ratio of word co-occurrence probabilities across different contexts, allowing it to capture global information from the entire corpus. GloVe excels at capturing global semantic relationships and is efficient to train since it only requires matrix factorization. In practice, GloVe is widely used for NLP tasks such as opinion analysis, information extraction, and semantic search (Amalia & Suhartono, 2020). Studies comparing Word2Vec and GloVe indicate that both perform well, though their effectiveness may vary depending on the domain and dataset size (Rong, 2014; Budiman et al., 2021). Therefore, the choice of embedding method should be aligned with the context and representation needs.

To further illustrate the differences between Word2Vec and GloVe embeddings, a t-distributed Stochastic Neighbor Embedding (t-SNE) visualization was created based on key sentiment-related terms. The figure below shows how both models position similar words differently in vector space, highlighting the contrast in how semantic relationships are captured.

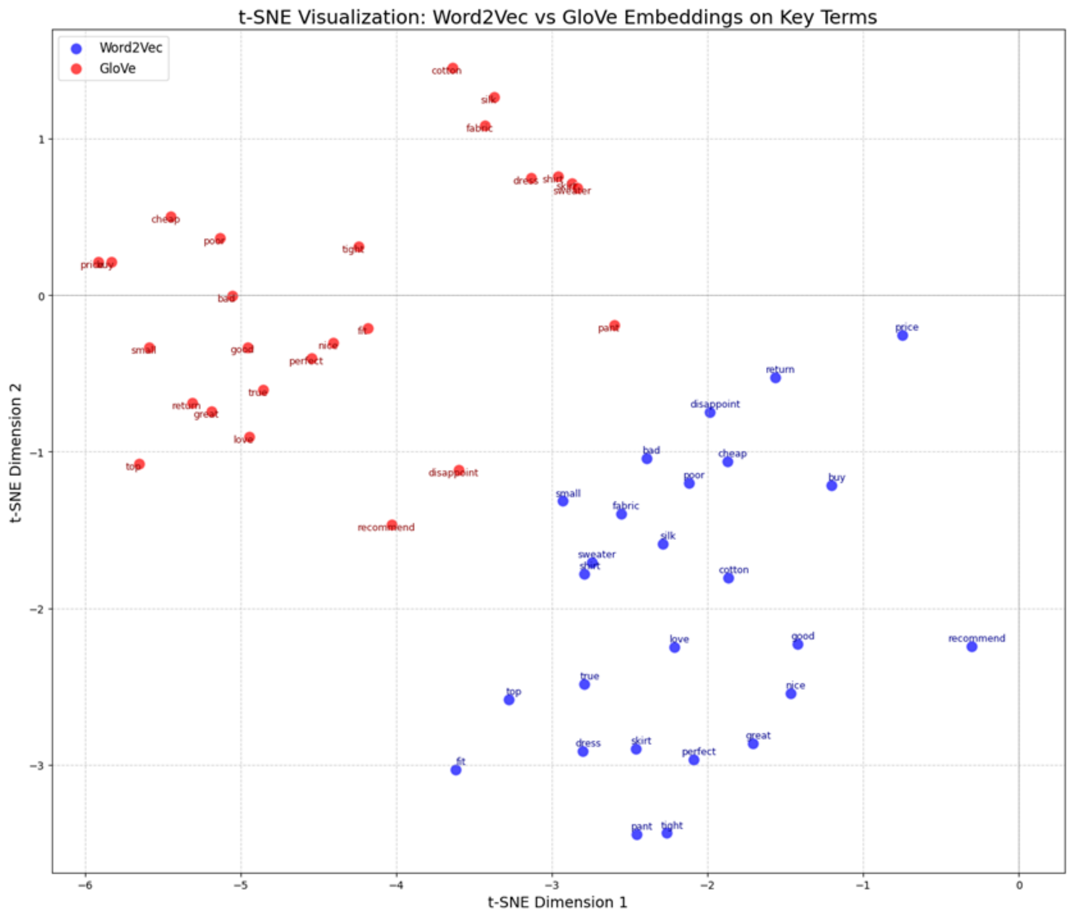

FIGURE 1. Word2Vec vs GloVe t-SNE visualization.

4. **Implementation of Machine Learning Models**

Once the text data is converted into numerical representations using Word2Vec and GloVe, the next step is to apply various machine learning algorithms to perform sentiment analysis and product recommendation prediction. Each model is trained using customer review data represented by these embedding techniques. Before training, the dataset is split into training and testing sets to evaluate model performance on unseen data. This study uses a stratified sampling approach to ensure that the class distribution remains consistent across both sets. Specifically, the data is divided with 70% allocated for training and 30% for testing, which preserves the proportion of sentiment classes. The machine learning models used include Random Forest, SVM, Naive Bayes, LSTM, Logistic Regression, and Gradient Boosting, applied to both balanced and imbalanced datasets.

5. **Model Evaluation**

To assess the performance of the machine learning models, this study employs two primary evaluation metrics: accuracy and F1-score. Accuracy measures the proportion of correctly predicted instances out of the total samples, providing a general sense of model effectiveness. However, in cases where the dataset is imbalanced, such as when one sentiment class dominates, accuracy alone may be misleading. Therefore, the F1-score, which is the harmonic mean of precision and recall, is also used to provide a more balanced evaluation. This metric is particularly valuable for understanding how well the model handles both false positives and false negatives. By evaluating models with both metrics across embeddings generated by Word2Vec and GloVe on balanced and imbalanced datasets, the study aims to identify the most robust and context-appropriate combination for sentiment analysis and product recommendation tasks.

The flowchart used to facilitate understanding of the steps is illustrated in **FIGURE 2**.

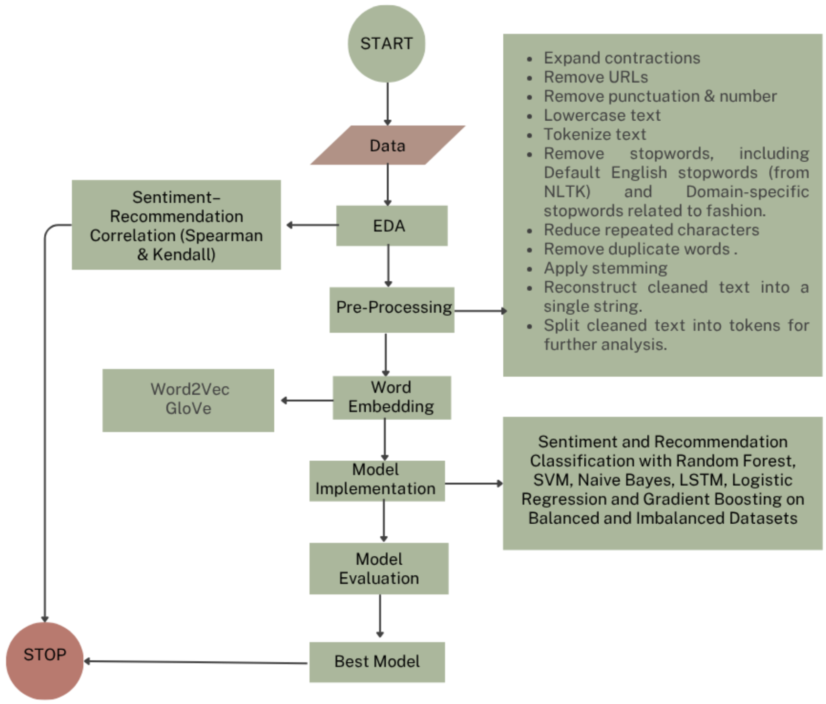

FIGURE 2. Research flowchart.

RESULT AND DISCUSSION

Sentiment And Recommendation Distribution (Target Variables)

The distribution of sentiment and recommendation labels in the dataset reveals a significant class imbalance. The figure below illustrates the proportion of each sentiment and recommendation category based on the women's e-commerce clothing reviews dataset.

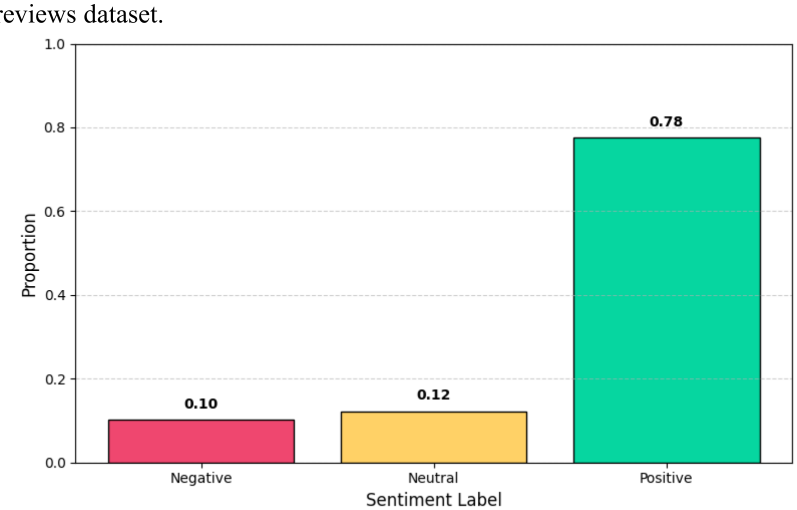

FIGURE 3. Sentiment distribution.

Most of the reviews in the dataset are positive, accounting for 78%, while neutral sentiments make up only 12%, and negative sentiments comprise 10%. This distribution indicates that the majority of customers who left reviews tend to be satisfied with the products they purchased. This imbalance in class proportions should be taken into account during model training, as it may lead to bias toward the majority class.

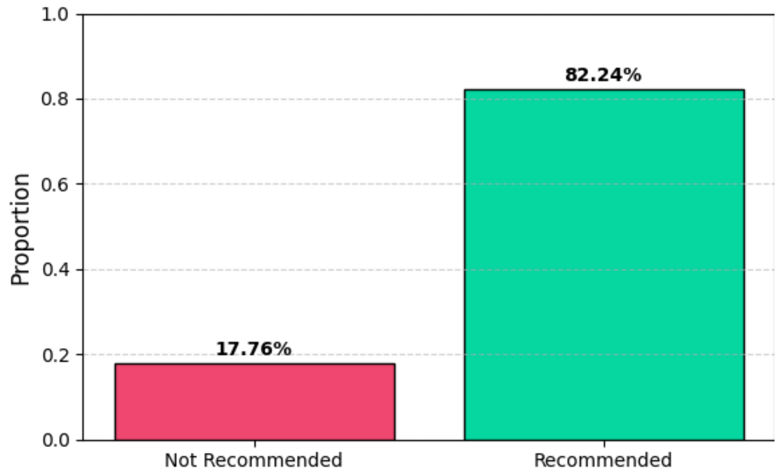

FIGURE 4. Recommendation distribution.

For the target variable of recommendation, 82.24% of the reviews indicate that customers recommend the product, while 17.76% do not. This pattern reinforces the findings from the sentiment distribution, where positive perceptions of the product are also reflected in customer recommendation behavior.

Model Performance

To analyze the effectiveness of Word2Vec and GloVe text representations in sentiment classification and product recommendation prediction tasks, this study implements various machine learning algorithms as well as a deep learning model based on Long Short-Term Memory (LSTM). The evaluation is conducted systematically under two data scenarios: Scenario 1: Imbalanced**,** using the original data without adjusting the class distribution. Scenario 2: Balanced, where data balancing techniques are applied. For classical models, SMOTE (Synthetic Minority Over-sampling Technique) is used, while for the LSTM model, class weighting is applied based on label proportions in the training data.

The Word2Vec and GloVe text representations are used as input for each model. For logistic regression, random forest, gradient boosting, support vector machine, and naïve Bayes, the embedding vectors from Word2Vec and GloVe are arranged into feature matrices. In contrast, for the LSTM model, the embedding representations are used as initial weights in the embedding layer, with the input consisting of token sequences obtained from tokenization and padding of the cleaned review texts.

Model Performance for Sentiment Classification

The sentiment classification task aims to categorize customer reviews into three classes: positive, neutral, and negative, based on their rating scores. The table below presents the evaluation results of various classification models trained using Word2Vec and GloVe text representations on both imbalanced and balanced datasets. The evaluation metrics used are Accuracy and F1 Score.

TABLE 1. Model performance for sentiment classification.

Embedding	Model	Data	Accuracy	F1
Word2Vec	Gradient Boosting	Imbalanced	0.796622	0.761979
		Balanced (SMOTE)	0.700823	0.735248
GloVe		Imbalanced	0.780443	0.704680
		Balanced (SMOTE)	0.623190	0.664559
Word2Vec	Logistic Regression	Imbalanced	0.797758	0.758022
		Balanced (SMOTE)	0.715441	0.747433
GloVe		Imbalanced	0.787539	0.726590
		Balanced (SMOTE)	0.634261	0.678854
Word2Vec	LSTM	Imbalanced	0.808686	0.783267
		Balanced (Class Weight)	0.739569	0.760696
GloVe		Imbalanced	0.802015	0.768601
		Balanced (Class Weight)	0.691456	0.729640
Word2Vec	Naive Bayes	Imbalanced	0.607011	0.662105
		Balanced (SMOTE)	0.579620	0.637655
GloVe		Imbalanced	0.359353	0.416018
		Balanced (SMOTE)	0.363611	0.426651
Word2Vec	Random Forest	Imbalanced	0.795061	0.749170
		Balanced (SMOTE)	0.752200	0.761163
GloVe		Imbalanced	0.775050	0.682443
		Balanced (SMOTE)	0.722254	0.710431
Word2Vec	SVM	Imbalanced	0.794209	0.728574
		Balanced (SMOTE)	0.704655	0.739755
GloVe		Imbalanced	0.775334	0.677216
		Balanced (SMOTE)	0.618791	0.667841

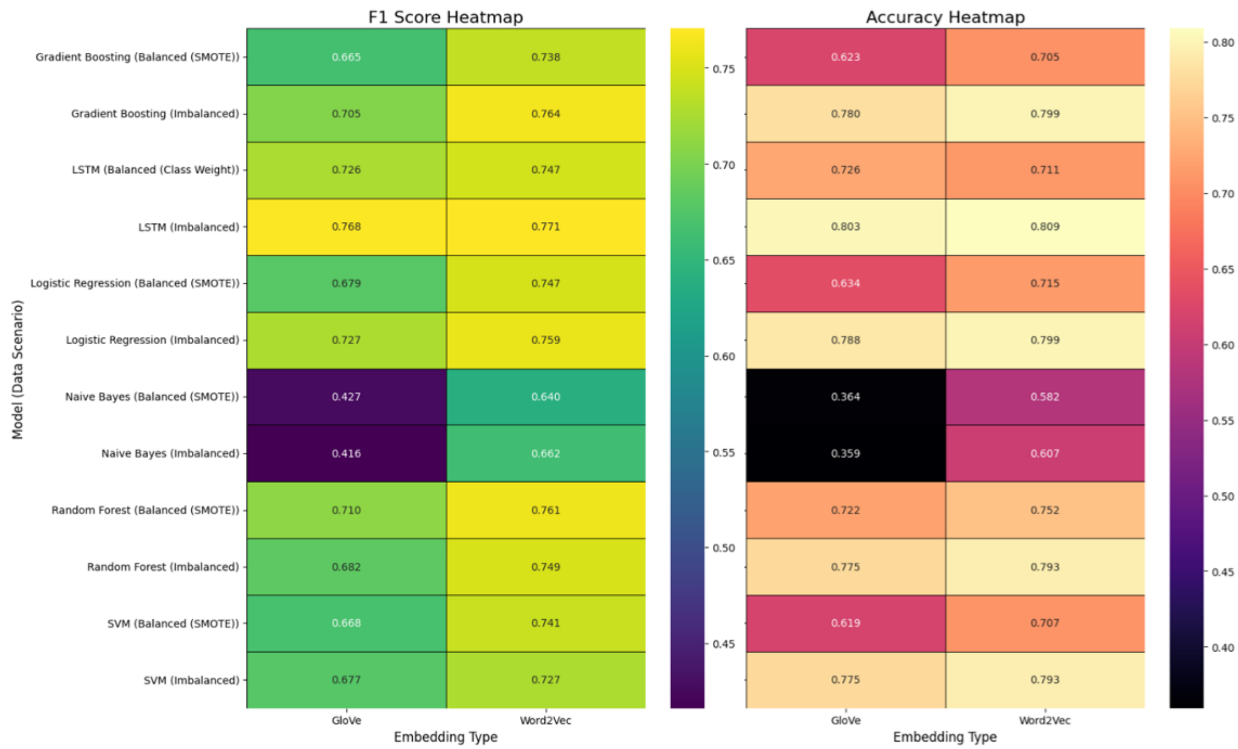

FIGURE 5. F1 score and accuracy heatmap for sentiment classification.

The evaluation results for the sentiment classification task indicate that the LSTM model using Word2Vec representation on imbalanced data delivers the best overall performance, achieving an accuracy of 80.87% and an F1 score of 78.33%. The LSTM model effectively captures the sequential context of words in reviews, and Word2Vec provides efficient word representations that enhance the model's performance. LSTM with GloVe also demonstrates competitive performance, although slightly lower compared to Word2Vec.

Other models such as Gradient Boosting, Random Forest, and Logistic Regression also yield fairly strong results when using Word2Vec, with F1 scores ranging from 75% to 76% on imbalanced data. However, most models experience a performance decline after balancing with SMOTE, as reflected in reduced accuracy and F1 scores for Logistic Regression, SVM, and Gradient Boosting. This suggests that synthetic data balancing is not always beneficial, especially when the characteristics of the minority class are not adequately representative.

Compared to GloVe, Word2Vec consistently produces higher performance across almost all models, both classical and deep learning. GloVe tends to result in lower F1 scores, particularly in models such as SVM and Naïve Bayes. Notably, Naïve Bayes with GloVe on balanced data achieves only 36.36% accuracy and an F1 score of 42.66%, significantly lower than other models. This indicates that Naïve Bayes is not suitable for embedding-based data, particularly when using global representations like GloVe.

Overall, these findings reinforce that the combination of Word2Vec and LSTM is the most effective approach for sentiment classification in fashion product reviews. Furthermore, the selection of balancing techniques and text representations should be tailored to the specific characteristics of the dataset and the model in use, to avoid a decline in classification performance.

Model performance for product recommendation classification

The table and heatmap below presents the evaluation results of various classification models in predicting product recommendations, based on a combination of word2vec and glove text representations, as well as two data scenarios: imbalanced and balanced. The evaluation was conducted using accuracy and f1 score metrics.

TABLE 2. Model Performance for recommendation classification.

Embedding	Model	Data	Accuracy	F1
Word2Vec	Gradient Boosting	Imbalanced	0.855379	0.841221
		Balanced (SMOTE)	0.793500	0.811949
GloVe		Imbalanced	0.836787	0.789681
		Balanced (SMOTE)	0.717286	0.745174
Word2Vec	Logistic Regression	Imbalanced	0.863185	0.847344
		Balanced (SMOTE)	0.809537	0.826110
GloVe		Imbalanced	0.842038	0.808092
		Balanced (SMOTE)	0.750071	0.775602
Word2Vec	LSTM	Imbalanced	0.873545	0.923513
		Balanced (Class Weight)	0.833239	0.892291
GloVe		Imbalanced	0.860772	0.918379
		Balanced (Class Weight)	0.754471	0.831088
Word2Vec	Naive Bayes	Imbalanced	0.693585	0.729158
		Balanced (SMOTE)	0.668606	0.707482
GloVe		Imbalanced	0.444366	0.485644
		Balanced (SMOTE)	0.419387	0.462014
Word2Vec	Random Forest	Imbalanced	0.855947	0.838308
		Balanced (SMOTE)	0.826994	0.833043
GloVe		Imbalanced	0.825149	0.758106
		Balanced (SMOTE)	0.784842	0.779834
Word2Vec	SVM	Imbalanced	0.862191	0.841917
		Balanced (SMOTE)	0.804428	0.822115
GloVe		Imbalanced	0.822311	0.742129
		Balanced (SMOTE)	0.736730	0.764798

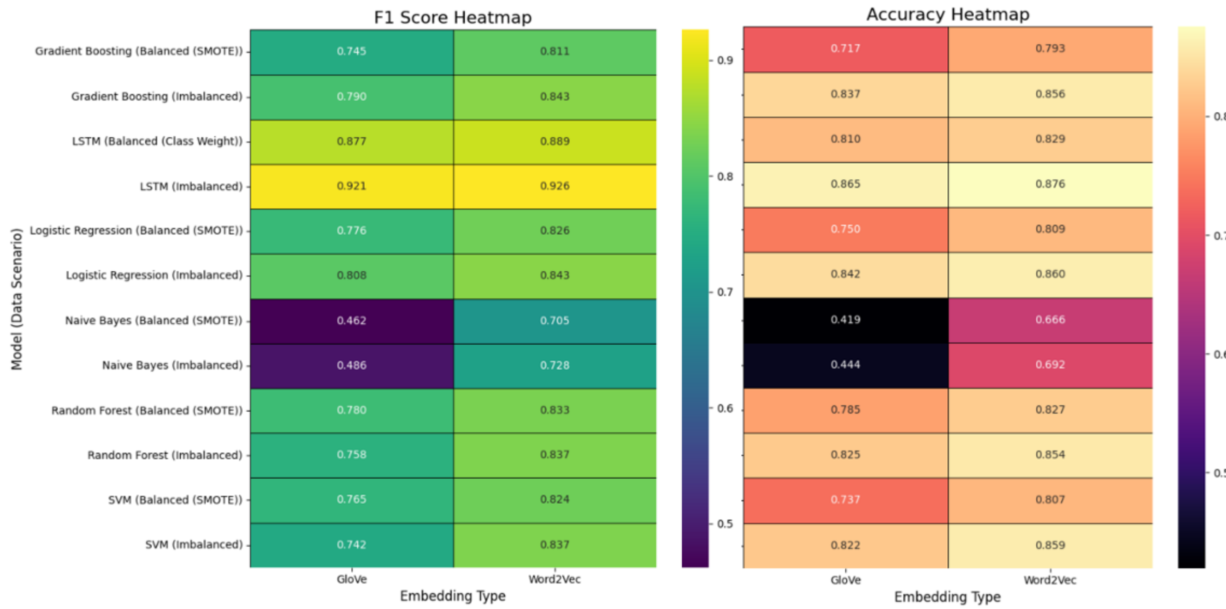

FIGURE 6. F1 score and accuracy heatmap for product recommendation.

Evaluation of various binary classification models for product recommendation prediction tasks indicates that the combination of Word2Vec text representation with the LSTM model delivers the best overall performance. On the imbalanced dataset, LSTM with Word2Vec achieved an accuracy of 87.35% and an F1 score of 92.35%. These results highlight LSTM's ability to effectively capture the sequential context of review texts, especially when paired with Word2Vec, which excels at capturing semantic relationships between words. In general, Word2Vec demonstrated more stable and superior performance compared to GloVe. Nearly all models, both classical and deep learning, achieved higher accuracy and F1 scores when using Word2Vec. In contrast, GloVe representations tended to result in lower performance, particularly for models like Naïve Bayes and SVM under the balanced scenario. This suggests that GloVe may be less suitable for handling short review texts commonly found in the e-commerce fashion domain.

The impact of data balancing techniques also varied. Some classical models such as Logistic Regression, SVM, and Random Forest experienced a slight performance drop after balancing using SMOTE. For instance, Logistic Regression with Word2Vec showed a decrease from 86.32% accuracy (F1: 84.73%) on the imbalanced data to 80.95% accuracy (F1: 82.61%) on the balanced data. This indicates that balancing does not always guarantee improved performance, particularly when the synthetic minority samples fail to adequately represent real-world characteristics. Other classical models like Random Forest and Gradient Boosting performed relatively competitively and consistently, although still underperforming compared to LSTM, especially in terms of F1 score, a crucial metric for evaluating performance on imbalanced data. Meanwhile, Naïve Bayes consistently showed the lowest performance, especially when using GloVe, with F1 scores as low as 48% on imbalanced data and even lower after balancing. Overall, these findings confirm that the combination of LSTM with Word2Vec is the most effective approach for modeling customer reviews in predicting fashion product recommendations. Meanwhile, the effectiveness of data balancing techniques should be carefully considered depending on the distribution and quality of minority class data.

One possible explanation for Word2Vec consistently outperforming GloVe is its predictive nature, which learns local contextual relationships between words. In short-text environments such as e-commerce reviews, where sentence structures are informal, fragmented, or include slang and emojis, Word2Vec can better capture semantic nuances through local co-occurrence windows. In contrast, GloVe relies on global co-occurrence statistics, which may be less effective when textual context is limited or when rare, domain-specific expressions are used. Fashion reviews often include colloquial terms like "super cute," "a bit tight," or expressive elongations (e.g., "loooove it!"), which Word2Vec is more capable of contextualizing effectively. While LSTM demonstrates superior performance, especially in terms of F1-score on imbalanced data, it also comes with higher computational cost and complexity in deployment. Classical models like Logistic Regression or Random Forest offer lower accuracy but are more interpretable, faster to train, and easier to deploy on lightweight systems or real-time applications. Therefore, practitioners must balance performance gain with resource constraints and interpretability requirements when choosing between deep learning and classical approaches.

Analysis of the Correlation Between Sentiment Classification and Product Recommendation

To examine the extent to which customer sentiment influences their tendency to recommend a product, an analysis was conducted on the relationship between sentiment labels and recommendation labels. The table below presents the distribution of recommendation proportions across different sentiment categories.

TABLE 3. Model sentiment-wise distribution of product recommendations.

	Not Recommended (%)	Recommended (%)
Negative	95.429996	4.570004
Neutral	58.585859	41.414141
Positive	1.059974	98.940026

The table shows that customers with positive sentiment almost entirely recommended the product (98.94%), while those with negative sentiment were predominantly inclined not to recommend it (95.43%). Meanwhile, customers with neutral sentiment exhibited a more balanced distribution, although the majority still tended not to recommend the product (58.59%). To visualize this relationship more intuitively, the following stacked bar chart displays the proportions of recommended and non-recommended products for each sentiment category.

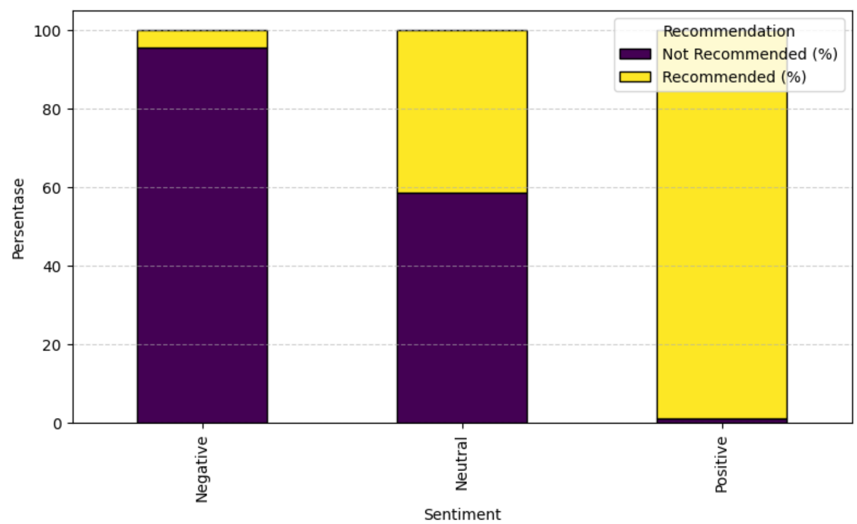

FIGURE 7. Bar chart showing recommendation proportions across sentiment categories.

This visualization further confirms that the more positive the sentiment expressed in a review, the higher the customer's tendency to recommend the product. Conversely, negative sentiment is consistently associated with the decision not to recommend. This clear pattern supports the assumption that sentiment expressed in reviews has a strong influence on customer recommendation behavior, and serves as a crucial foundation in the development of text-based recommendation systems. To statistically examine the relationship between customer sentiment and the tendency to recommend a product, an ordinal correlation analysis was conducted using two methods: Spearman’s Rho and Kendall’s Tau. The results of these tests are presented in the following table.

TABLE 4. Spearman’s Rho and Kendall’s Tau

Metric Correlation	Correlation Coefficient	p-value
Spearman's Rho	0.8340	0.0000
Kendall's Tau	0.8120	0.0000

The Spearman's Rho value of 0.8340 and Kendall's Tau value of 0.8120, both with p-values < 0.001, indicate a strong and statistically significant relationship between sentiment scores and customer recommendation behavior. The more positive the sentiment expressed in a review, the higher the likelihood that the customer will recommend the product. These findings reinforce the earlier exploratory analysis and confirm that customer sentiment can serve as a relevant indicator in text-based review recommendation systems.

CONCLUSION

This study evaluates the effectiveness of two popular text representation methods—Word2Vec and GloVe—in classifying customer sentiment and predicting product recommendations based on fashion product reviews using various machine learning models. The results demonstrate that Word2Vec consistently outperforms GloVe across classical and deep learning approaches, owing to its ability to capture nuanced local semantics commonly found in customer narratives. Word2Vec combined with LSTM achieved the highest performance, particularly in imbalanced data scenarios, reaching an accuracy of 87.35% and an F1 score of 92.35%. This suggests that models leveraging local semantic context are highly effective in interpreting sentiment and recommendation tendencies in fashion-related feedback. In contrast, GloVe generally exhibited lower performance, especially in models sensitive to word sequence or reliant on global semantic features. Furthermore, correlation analysis revealed a strong and statistically significant association between customer sentiment and recommendation behavior, evidenced by a Spearman's Rho of 0.8340 and a Kendall's Tau of 0.8120. These insights highlight that accurately modeling sentiment is a strong predictor of customer recommendation decisions in the fashion domain.

The findings of this study hold significant practical value for fashion retailers, UX designers, and e-commerce platform developers. By demonstrating that Word2Vec-based sentiment representations—especially when combined with LSTM—yield superior performance in both sentiment classification and recommendation prediction, the study provides actionable insights for enhancing customer experience systems. Retailers can leverage these insights to build more accurate and personalized recommendation engines, while UX teams may incorporate sentiment-aware modules to better interpret and visualize customer feedback. Furthermore, the integration of such models into review monitoring tools can help detect shifts in customer perception in real time. As a future direction, implementing a prototype system or interactive dashboard that showcases real-time sentiment-to-recommendation mapping would significantly increase the applied relevance of the study.

Future research may explore modern contextual embeddings such as BERT, FastText, or other transformer-based models, as well as hybrid representations that combine local and global semantic understanding. Incorporating fashion-specific linguistic patterns, informal tone, emojis, and varying review lengths could further improve model adaptability to real-world user-generated content. Nonetheless, this study has several limitations, including the use of a single domain-specific dataset, which may limit the generalizability of findings across different product categories or cultural contexts. Additionally, the absence of external validation using datasets from other platforms leaves room for further verification. Addressing these limitations through fairness-aware modeling, cross-domain evaluation, and the integration of human oversight will be essential to enhance the robustness and practical applicability of future recommendation systems.

REFERENCES

Amalia, F., & Suhartono, D. (2020). Sentiment analysis on Indonesian hotel reviews using word embedding and deep learning. *Journal of Information Systems Engineering and Business Intelligence, 6*(1), 1–10. https://doi.org/10.20473/jisebi.v6i1.2020.1-10

Adewopo, A., Oyelade, O. J., & Omidiora, E. O. (2021). A comparative study of Word2Vec and GloVe for sentiment classification using machine learning. *Journal of Intelligent Systems, 30*(3), 441–454.

Dewi, I. P., & Nugroho, A. S. (2021). Penerapan Word2Vec untuk klasifikasi sentimen ulasan produk e-commerce. *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIIK), 8*(2), 267–274. https://doi.org/10.25126/jtiik.2021821961

Gupta, R., & Bhatnagar, V. (2023). Improving product recommendation with sentiment-aware contextual embeddings. *International Journal of Data Science and Analytics, 15*(2), 85–98. https://doi.org/10.1007/s41060-023-00321-7

Li, J., & Xu, W. (2020). Comparative study of word embeddings in sentiment analysis. *Procedia Computer Science, 174*, 142–149. https://doi.org/10.1016/j.procs.2020.06.037

Patil, S., & Atique, M. (2020). Comparative performance analysis of word embeddings on BiLSTM model for sentiment classification. *Procedia Computer Science, 167*, 1101–1110.

Pennington, J., Socher, R., & Manning, C. D. (2014). GloVe: Global vectors for word representation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)* (pp. 1532–1543). Association for Computational Linguistics.

Rong, X. (2014). Word2Vec parameter learning explained. *arXiv preprint arXiv:1411.2738*. https://arxiv.org/abs/1411.2738

Saragih, S. N., Hartati, S., & Andriani, D. (2022). Analisis sentimen ulasan produk fashion menggunakan metode Naïve Bayes dan lexicon-based. *Jurnal Ilmu Komputer dan Aplikasi, 20*(3), 165–172.

Singh, R., Kaur, H., & Dhiman, G. (2022). Customer recommendation prediction based on sentiment-aware word embeddings. *International Journal of Advanced Computer Science and Applications, 13*(6), 543–551.