

Improving Bilingual E-Commerce Customer Review Analysis Using a Combined Machine Learning and Deep Learning Approach for Sentiment Prediction

M Tharun Kumar¹, E Pavithra², R R Shantha Spandana³

¹P.G Scholar, Department of MCA, Sri Venkatesa Perumal College of Engineering & Technology, Puttur,
E-mail: mangamtharunkumar33@gmail.com, ORC-ID: <https://orcid.org/0009-0007-0551-9070>

²Assistant Professor, Department of MCA, Sri Venkatesa Perumal College of Engineering & Technology, Puttur,
E-mail: epavithra526@gmail.com, ORC-ID: <https://orcid.org/0009-0006-8871-4551>

³ Assistant Professor, Department of MCA, Sri Venkatesa Perumal College of Engineering & Technology, Puttur,
E-mail: shanthaspandana@gmail.com, ORC-ID: <https://orcid.org/0009-0003-4236-1250>

To Cite this Article

M Tharun Kumar, E Pavithra, R R Shantha Spandana, "Improving Bilingual E-Commerce Customer Review Analysis Using a Combined Machine Learning and Deep Learning Approach for Sentiment Prediction", *Journal of Science Engineering Technology and Management Science*, Vol. 03, Issue 04, April 2026, pp: 287-297, DOI: <http://doi.org/10.64771/jsetms.2026.v03.i04.pp287-297>

Submitted: 28-02-2026

Accepted: 01-04-2026

Published: 08-04-2026

Abstract: In bilingual e-commerce systems, it's important to get a good sense of how customers feel in order to increase user satisfaction and purchase behavior. Traditional methods of mood analysis often have trouble with different languages and unclear contexts. The data came from online shopping and product review datasets on Hugging Face and GitHub, which have textual comments in two languages. In the preprocessing step, the text was normalized, HTML and URLs were removed, it was tokenized, labels were encoded, and sentiment was extracted using NLTK's SentimentIntensityAnalyzer. Next, NLTK and BERT were used to create the embedding. A lot of different models were used, such as Linear Regression (SGD), SVM (SGD), Random Forest, XGBoost, RF + XGB, BiLSTM, ELECTRA, and BERT. There were also advanced hybrid designs like BERT with LSTM and LSTM+GRU, and a Voting Classifier that mixed GB, XGB, LGBM, and CatBoost. SMOTEENN was used to balance the datasets, and Explainable AI tools (LIME and SHAP) made the results easier to understand. A Flask-based interface also let real-time mood prediction and topic modeling work with user input. Testing the Voting Classifier with Accuracy, Precision, Recall, and F1-score showed that it was the most accurate, with a score of 99.1% on Dataset A and 98.7% on Dataset B. This proved that it was stable and had the best predictive performance in classifying sentiment in two languages.

Index Terms: E-commerce, customer satisfaction, sentiment analysis, natural language processing (NLP), sentiment keywords, feature-level sentiment analysis, consumer behavior, customer feedback".

This is an open access article under the creative commons license
<https://creativecommons.org/licenses/by-nc-nd/4.0/>



1. INTRODUCTION

The fast growth of e-commerce and social media has changed how companies talk to customers and figure out what they're doing. Social networks have grown into important ways to market, talk to people, and look at customer comments [1]. A lot of user-generated material, like reviews, comments, and posts, are created every day and show how customers feel, what they like, and new trends [2]. However, it is not possible to directly interpret this unstructured data because it is so big and includes a lot of different languages. Natural Language Processing (NLP) has come a long way, and now it's possible to automatically pull out useful information from this kind of data. This lets businesses look at feelings, thoughts, and intentions that are stated through text. Modern businesses need to be able to do this if they want to improve customer experience, involvement, and services by using data-driven insights [3].

Even though a lot of progress has been made, there are still big problems to solve before mood analysis can be very accurate and understand what people are saying. Text on social media sites like Twitter and Facebook often uses slang, acronyms, multilingual phrases, and emotional undertones that are hard for standard models to pick up [4]. Also, a lot of the current systems that sort sentiments only look at general polarity—positive, negative, or neutral—without looking at the product or feature attributes that cause these sentiments [5]. This gap makes it

harder for businesses to get deep information that they can use. More and more people are using bilingual and international e-commerce platforms, which makes it harder to figure out how someone feels because small differences in language can change meaning during translation [6]. Because of this, we need methods that go beyond simple mood analysis and toward more detailed analysis that can connect emotional statements to specific features of a product or service.

To get around these problems, this study focuses on looking into a lot of expressions that show emotion in order to find the main factors that affect customer happiness [7]. The main goal of the study is to find mood patterns, contextual dependencies, and thematic links across a wide range of product categories by extracting and analyzing textual feedback from bilingual sources. The analysis gives a detailed look at how people think about and rate goods by looking at sentiment keywords and how often they appear together in customer reviews [8]. With this method, the approach also aims to make things easier to understand across language barriers by making sure that language translation doesn't change the meaning of customer feedback.

The findings of this study have important implications for both business and education. From a theoretical point of view, it adds to the growing overlap between computer linguistics and consumer behavior analysis by connecting modeling based on data with making decisions in the real world [9]. In practice, it gives businesses the power to find out what makes customers happy, improve product design, and tailor marketing campaigns based on real customer feedback [10]. The main goal of this study is to learn more about how emotional language affects how people think and feel, which will help build a strong base for predictive analytics and boost customer involvement in global e-commerce ecosystems.

2. LITERATURE REVIEW

More and more recent study on e-commerce and sentiment analysis has looked at how customers behave, what they think about when they buy something online, and how small differences in language can affect how people think. He [11] looked into the idea of "her economy" to look into the phenomenon of women driving online spending. He found that emotional and psychological factors affect women's online shopping habits. This study focused on the growing economic potential of gender-specific consumption habits, but it couldn't fully model how people felt quantitatively. Similarly, Chen et al. [12] used eye-tracking experiments to look at how online reviews affect people's decisions. They found that emotional tone and paying close attention to review aspects have a big effect on people's plans to buy. But the study didn't look at how different languages and cultures affect how people understand how others feel in markets with more than one language.

He and Abisado [13] did a full review of how sentiment analysis can be used to analyze reviews of products sold online, showing different machine learning and deep learning techniques. Their results showed that sentiment classification is becoming more important for better business intelligence, but they also said that it's still hard to handle informal language and mixed feelings. Ziakis and Vlachopoulou [14] added to the larger conversation by talking about how artificial intelligence can be used in digital marketing. Their work showed how AI could change marketing and customer interaction by automating them. It also showed how important it is for AI-driven sentiment systems to be able to understand and handle data in an ethical way. Mabokela et al. [15] took this line of research even further by focusing on multilingual sentiment analysis and pointing out that there aren't enough tools and datasets for languages that aren't used a lot. This review showed how important it is to have flexible models right away that can deal with different languages while still being accurate in terms of meaning.

To build on these studies, Punetha and Jain [16] suggested a way to look at restaurant reviews that uses game theory and multi-criteria decision-making techniques along with an unsupervised sentiment analysis approach. Their model had creative ways of interpreting data, but it could only be used in certain domains, making it hard to use in larger e-commerce settings. Li, Wang, and Yang [17] looked into transfer learning methods for classifying sentiments using various data sources and found that they worked better across domains. Still, the model's need for large datasets with annotations is still a big problem when it comes to real-world usage. In the same way, Necula [18] looked into buyer behavior by using machine learning to analyze the time people spent on e-commerce websites. This gave them useful information about engagement patterns but not enough to fully understand how people felt about the text.

Li and Nuangjamnong [19] looked into how social media marketing affects Chinese customers' plans to buy green cosmetics. They found that the quality and tone of reviews affect how customers feel. The study did a good job of combining linguistic and psychological factors, but it didn't look at small-scale mood features or keyword co-occurrences. Shengyuan and Jianglin [20] looked at how different types of live-streaming hosts affect people's

buying habits. They found that streamer traits like credibility and engagement style have a big effect on how much people believe them. But most of their work was based on qualitative text analysis, and they didn't use advanced mood models to get more information.

All together, these studies show how quickly mood analysis is changing and how important it is for understanding how people behave in digital environments. But there are still some problems, mainly the fact that we don't look into bilingual and feature-level mood interpretation enough, we rely too much on domain-specific datasets, and we don't pay enough attention to contextual keyword analysis. To fill in these gaps, this study presents a complete approach for sentiment modeling that includes text data in multiple languages and looks at specific expressions that show sentiment across product categories. This method helps us understand customer satisfaction and buying motivation more clearly across languages, connecting the field of linguistic theory with data-driven market intelligence.

3. MATERIALS AND METHODS

By combining traditional, deep learning, and transformer-based methods, the suggested framework aims to improve the accuracy and readability of sentiment classification for bilingual e-commerce reviews. The dataset comes from Hugging Face and GitHub repositories and includes customer comments in multiple languages that shows a range of emotional states. The process includes loading data, preprocessing it, and representing features using both NLTK and BERT embeddings to pick up on semantic and contextual details in written data. To make sure that the models work well in most situations, standard performance measures are used to train and test many models, such as Linear Regression, SVM, Random Forest, XGBoost, BiLSTM, ELECTRA, and BERT. To make things even better, SMOTEENN is used to balance the classes, and hybrid and ensemble models like BERT-LSTM, LSTM-GRU, and a Voting Classifier that combines GB, XGB, LGBM, and CatBoost make the predictions more accurate. Explainable AI methods like LIME and SHAP make models clear, and a Flask-based deployment makes sure that sentiment prediction happens in real time. This creates a system that can be scaled up, is easy to understand, and works well.

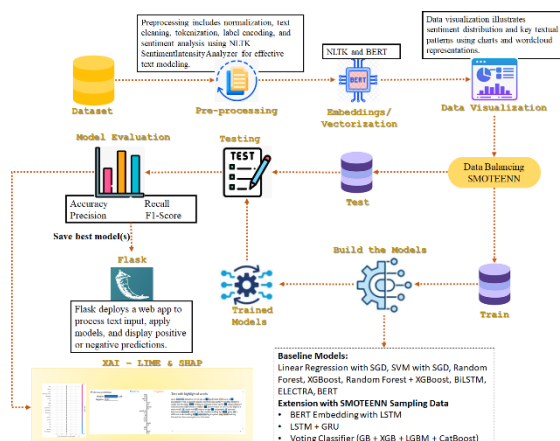


Fig.1 Proposed Architecture

The system design (fig.1) combines the whole process of sentiment-based e-commerce review analysis, from importing data to putting it into use. Datasets that are to be used are loaded and normalized, tokenized, and labeled for mood. Before features are extracted and models are trained, NLTK and BERT are used to make text embeddings. SMOTEENN-balanced data are used to test baseline and extension models, as well as ensemble deep learning methods. Lastly, the improved model is put into use through a Flask web app that lets you guess sentiment in real time and see results based on your level of confidence.

a) Dataset Collection:

From the Hugging Face repository, 62,774 translated customer reviews in English were taken from an international e-commerce review dataset. It has three parts: label, category, and post. Label shows the polarity of the emotion, category shows the type of product, and post shows the review text. The dataset includes a wide range of opinions about different types of products. It also includes a lot of different languages, which makes it possible to accurately and completely classify and understand sentiment in two languages.

	label	cat	post
0	1	books	As parents, you must have Liu Yong's mentality...
1	1	books	The author really has the rigorous style of th...
2	1	books	The author uses lengthy and detailed reports o...
3	1	books	The author's use of "hug" before the battle is...
4	1	books	The author loved reading when he was a boy. It...

Fig.2 Dataset – A

Ten thousand English-translated customer feedback posts are stored under a single post attribute in another dataset that comes from a public GitHub file. It doesn't have clear emotion or categorical labels, but it does have real, unstructured customer expressions. When you combine these datasets, you get a wider range of languages, better generalization, and more reliable sentiment classification models for multilingual e-commerce apps.

	post
0	Logistics is fast. Things should be fine
1	Not bad, logistics is also very fast
2	I didn't receive the item. How come the delive...
3	It's pretty good. I haven't found any problems...
4	I have always used Huawei products. This is my...

Fig.3 Dataset – B

b) Pre-Processing:

The steps below show the whole process of getting data ready for mood classification. They include preprocessing, embedding, visualization, balancing, feature extraction, and splitting. This makes sure that the data is high-quality, easy to understand, and well-structured.

i) Data Pre-processing: Text preprocessing was done to make the original corpus better and more consistent. Normalization was done by lowering the case of words, getting rid of URLs, HTML tags, and characters that aren't letters, and then tokenization and noise filters. After that, the NLTK SentimentIntensityAnalyzer was used to make sentiment polarity scores in order to get some initial sentiment insights. This step makes sure that the text data is consistent in terms of language, easy to understand in terms of meaning, and organized in a way that makes it best for feature extraction and classification tasks that come later.

ii) Embeddings / Vectorization: After preprocessing, embedding and vectorization methods were used to turn the textual data into numerical representations. NLTK-based token representations were used to make word embeddings, and the Bidirectional Encoder Representations from Transformers (BERT) model was used to give them meaning. This two-level embedding method finds both syntactic and semantic relationships in text, which lets representations be more expressive and aware of their surroundings. The high-dimensional vectors that are made are used as distinguishing factors that make model adaptation and prediction accuracy much better during training.

iii) Data Visualization: An exploratory data visualization was used to look at the features of the corpus and the trends of sentiment distribution. Visualization tools like histograms, count plots, and word clouds were used to show how often sentiment classes and popular words appeared. These visualizations help you understand how datasets are put together, find imbalances or biases, and find language patterns that are specific to a topic. The information gained in this step led the finetuning of later steps in the preprocessing process and made sure that the feature space for model development was balanced and representative.

iv) Data Balancing: The Synthetic Minority Over-sampling and Edited Nearest Neighbors (SMOTEENN) hybrid resampling method was used in the extended model to lessen the bad effects of class imbalance. SMOTEENN combines artificially oversampling of minority groups with data cleaning of samples that overlap, creating a dataset that is fair and free of noise. This method makes the classifier better at generalizing across different types of emotion, less biased toward dominant classes, and much more stable and accurate in its predictions.

v) Feature Extraction: Embedded models were used to get discriminative and sentiment-relevant features through feature extraction. To find word co-occurrence patterns, polarity changes, and syntactic dependencies, linguistic, semantic, and contextual features were used. The process changes high-dimensional embeddings into input vectors that can be understood and used by deep learning systems. Effective feature extraction not only protects the integrity of the semantics but also speeds up learning, making sure that the model can correctly tell the difference between classes of positive and negative mood.

c) Data Splitting:

The cleaned up and enhanced data was organized into training and testing groups so that the model could be evaluated fairly. A stratified random split made sure that each emotion class was fairly represented in both parts. The training subset was used to teach and improve the model, and the test subset was only used to make sure the model worked well. This division keeps data from getting out, helps with fair evaluation of generalization ability, and is a key part of making sure that experimental evaluations are accurate and can be repeated.

d) Algorithms:

Linear Regression with Stochastic Gradient Descent (SGD) reduces the loss function as little as possible by changing the weights of small groups of data over and over again. This optimization speeds up convergence, lets you handle a lot of data, and gives you stable linear predictive performance across all sentiment estimation tasks. Support Vector Machine with SGD takes the best features of both Support Vector Machine (SVM) and scalable gradient-based optimization and mixes them. It does a good job of finding the best hyperplanes for classification, which makes decision boundaries more accurate and increases the speed of processing in textual feature spaces with many dimensions.

Random Forest uses bootstrap aggregation to build a group of decision trees that work together to make predictions more stable. It prevents overfitting and has good generalization performance by averaging the outputs of multiple trees. This makes it strong for dealing with noisy and nonlinear mood patterns.

$$Gini = 1 - \sum_{i=1}^c (P_i)^2 \quad (1)$$

Extreme Gradient Boosting (XGBoost) uses sequential tree boosting to get rid of classification mistakes quickly. Its regularization features and gradient-based optimization make learning more stable, predictions more accurate, and computing more scalable, especially in complicated, high-dimensional text analytics situations.

$$\hat{y}_i = \sigma \left(\sum_{k=1}^K f_k(x_i) \right), f_k \in F \quad (2)$$

The mixed Random Forest + XGBoost model uses both bagging and boosting to find the best balance between reducing variance and minimizing bias. This group synergy improves the predictability, stability, and adaptability of the model, which leads to more accurate mood classification across a wider range of text inputs.

Bidirectional Long Short-Term Memory (BiLSTM) networks can remember relationships between contexts in both forward and backward time directions. This two-way view lets you learn a lot of features from a long string of text, which makes it much easier to understand the meaning, the importance to the situation, and how the model works.

ELECTRA uses a new way to find tokens to make pretraining of language models more effective. By telling the difference between real and fake tokens, it creates contextually rich embeddings that improve the accuracy of mood classification further down the line with less computational overhead than traditional transformers.

Bidirectional Encoder Representations from Transformers (BERT) uses deep bidirectional attention processes to find connections between meanings in whole text sequences. Its pre-trained contextual embeddings help it understand linguistic subtleties better, which makes mood classification much more accurate and reliable.

Deep contextual embeddings and sequence learning can both be used with BERT Embedding with LSTM. When you combine BERT's semantically enhanced word representations with LSTM's temporal dependencies, you get a better understanding of the model, better agreement with the context, and more accurate predictions in tasks that require sentiment classification.

The combined LSTM + GRU architecture takes the best features of both LSTM and GRU and combines them into one architecture. This integration takes into account both short- and long-term dependencies, which makes text-based sentiment models more accurate in terms of convergence, generalization, and interpretation.

The Gradient Boosting, XGBoost, LightGBM, and CatBoost models all make predictions that are added together in the Voting Classifier ensemble. It gets better stability, less variation, and better prediction performance by combining different gradient-based learners and making decisions based on agreement about how to classify sentiment.

$$\hat{y} = \operatorname{argmax}_c \left(\sum_{i=1}^n II(\hat{y}_i = c) \right) \quad (3)$$

e) Integration of XAI & Flask Framework:

Adding Explainable Artificial Intelligence (XAI) to the system uses LIME and SHAP methods to make models easier to understand and more open. The LIME (Local Interpretable Model-Agnostic Explanations) module makes feature-level explanations for each prediction by getting close to the model's local decision border and highlighting the most important words that go into a sentiment label. In the same way, SHAP (SHapley Additive Explanations) gives a global interpretation by measuring how much each feature contributes to the general prediction. This gives a full picture of how the model works and which features are important.

The system also has a Flask-based web framework for real-time deployment, which lets users give textual comments and see the results of predictions right away, along with XAI-based explanations. This combination makes sure that it is useful, builds trust among users, and connects advanced model performance to decisions that can be explained.

4. EXPERIMENTAL RESULTS

Accuracy: How well a test can tell the difference between sick and healthy people is called its accuracy. To get an idea of how accurate a test is, we should figure out what percentage of cases are true positives and true negatives. In terms of math, this can be written as

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

Precision: Precision is the percentage of correctly classified cases or samples compared to those that were correctly classified as positives. So, here is the method to figure out the precision:

$$Precision = \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}} \quad (5)$$

Recall: In machine learning, recall is a metric that shows how well a model can find all the important instances of a certain class. It shows how well a model captures instances of a certain class. It is calculated by dividing the number of correctly predicted positive observations by the total number of real positives.

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

F1-Score: The F1 score is a way to rate the correctness of a machine learning model. It takes a model's accuracy and recall scores and adds them together. The accuracy metric counts how many times, across the whole dataset, a model made a correct guess.

$$F1 \text{ Score} = 2 * \frac{Recall * Precision}{Recall + Precision} * 100 \quad (7)$$

Table.1 Performance Evaluation Table – Dataset A

ML Model	Accuracy	Precision	Recall	F1 Score
BERT	0.936	0.937	0.936	0.936
BERT_LSTM	0.939	0.939	0.939	0.939
ELECTRA	0.936	0.936	0.936	0.936
BiLSTM	0.928	0.928	0.928	0.928
LSTM + GRU	0.930	0.930	0.930	0.930
Linear (SGD)	0.924	0.925	0.925	0.924
SVC (SGD)	0.924	0.925	0.925	0.924
Random Forest	0.902	0.902	0.902	0.902
XGBoost	0.892	0.893	0.892	0.892
RF + XGB	0.904	0.904	0.904	0.904

Voting Classifier	0.991	0.987	0.952	0.969
--------------------------	--------------	--------------	--------------	--------------

Comparing the performance of different models is shown in Table 1. The Voting Classifier did the best, beating all other machine learning and deep learning methods in classifying opinion for Dataset A.

Table.2 Performance Evaluation Table – Dataset B

ML Model	Accuracy	Precision	Recall	F1 Score
BERT	0.965	0.965	0.966	0.965
BERT_LSTM	0.969	0.969	0.969	0.969
ELECTRA	0.963	0.963	0.962	0.963
BiLSTM	0.933	0.934	0.934	0.933
LSTM + GRU	0.518	0.259	0.500	0.341
Linear (SGD)	0.941	0.942	0.941	0.941
SVC (SGD)	0.941	0.942	0.941	0.941
Random Forest	0.941	0.941	0.941	0.941
XGBoost	0.902	0.902	0.902	0.902
RF + XGB	0.944	0.944	0.944	0.944
Voting Classifier	0.987	0.983	0.980	0.982

Table 2 shows a comparison of how well different models did on Dataset B. The Voting Classifier did the best, beating all the others in important evaluation measures like Accuracy, Precision, Recall, and F1-score.

Fig.4 Comparison Graph – Dataset A

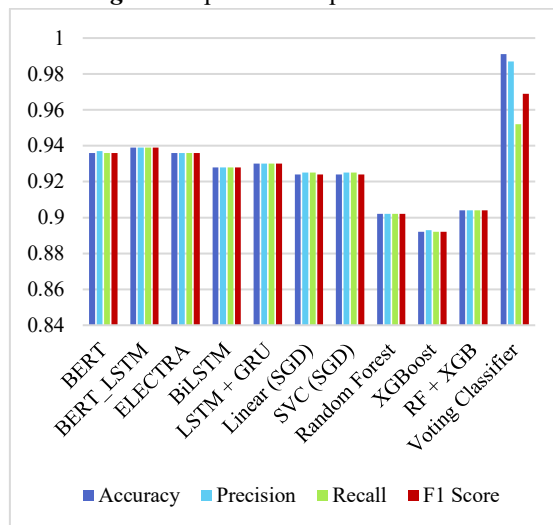


Figure 4 shows how the performance of different models on Dataset A was compared using Accuracy, Precision, Recall, and F1-Score as evaluation measures. Out of all the models, the Voting Classifier did the best overall, showing that it was the most accurate and effective at making predictions.

Fig.5 Comparison Graph – Dataset B

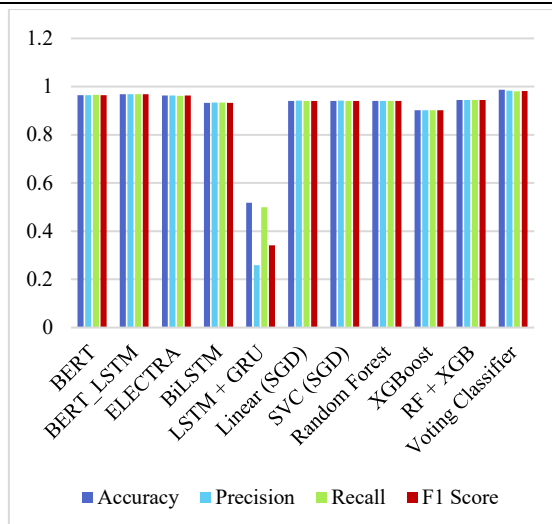


Figure 5 shows how well the various models performed on Dataset B, as judged by the Accuracy, Precision, Recall, and F1-Score. Once more, the Voting Classifier did better than all the other models and got the most accurate results.

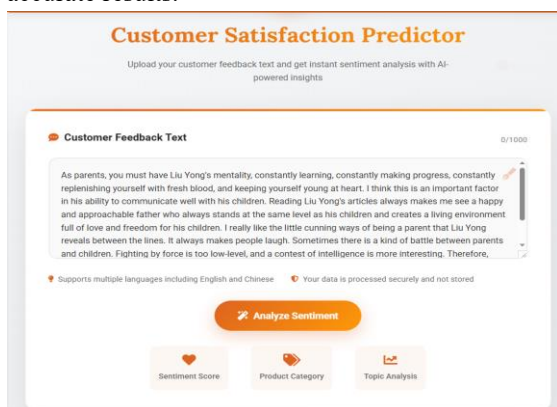


Fig.6 Enter Customer Feedback Text

Figure 6: A screen for entering text comments from customers so that the prediction system can automatically analyze their feelings and level of satisfaction.

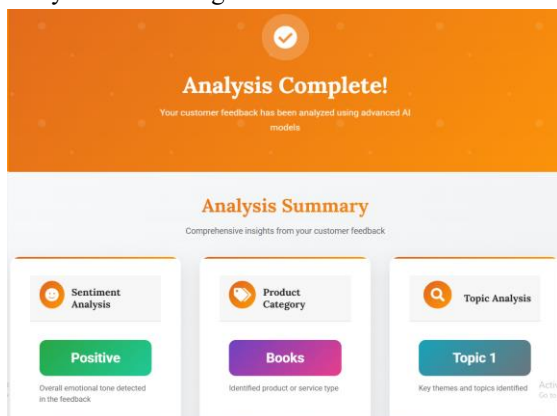


Fig.7 Analysis Summary

Figure 7: An output summary showing that the final result of the mood analysis was positive, with the product category being books and the topic being 1.

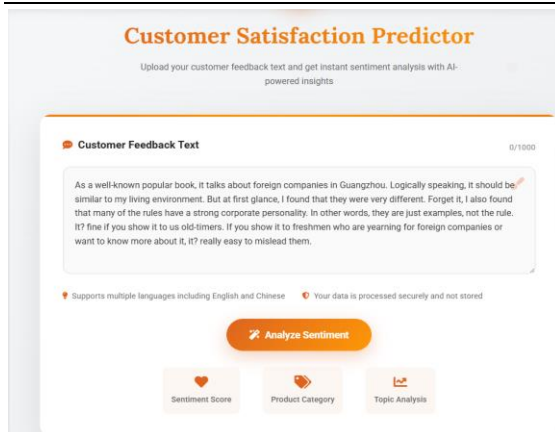


Fig.8 Enter Customer Feedback Text

Figure 8: An input interface for entering text comments from customers so that text analysis can predict their level of satisfaction and how they feel.



Fig.9 Analysis Summary

Figure 9: A summary of the output shows that the end sentiment analysis result was negative, with the product category being books and the topic being none.

5. CONCLUSION

In conclusion, the system that was made solves the problem of correctly interpreting bilingual customers' feelings in order to improve user happiness and help people make decisions in e-commerce settings. Hugging Face and GitHub review datasets in multiple languages were used, and a full sentiment analysis pipeline was set up that combined standard machine learning, deep learning, and transformer-based models. Text normalization, tokenization, emotion extraction with NLTK, and embedding generation with BERT were some of the techniques used to make sure that features were represented correctly and that the context was understood. The best model that was tested was the ensemble Voting Classifier, which used GB, XGB, LGBM, and CatBoost to get an accuracy of 99.1% on Dataset A and 98.7% on Dataset B. Integration of hybrid designs like BERT with LSTM and LSTM+GRU, dataset balancing through SMOTEENN, and interpretability through Explainable AI frameworks (LIME and SHAP), along with other improvements, made things much clearer and more stable. A Flask-based interface also made real-time forecast and topic modeling possible, showing that the system could be used in real life. Overall, the system provides a solid and easy-to-understand framework for sentiment analysis that can handle different languages. This leads to better user experience data and smarter decision-making in bilingual e-commerce apps.

In the future, improvements can be made to the system so that it can handle more languages and dialects. This will allow for more comprehensive multilingual sentiment analysis on all global e-commerce sites. Adding multimodal data like pictures, sounds, or patterns of how people interact with a system could help it better understand the context and get the mood right. Adding more complex transformer designs like GPT or T5 could make semantic understanding and adaptability even better. Using live data for real-time model retraining can make sure that the model is always learning and that its performance stays stable. It would also be easier to scale,

get to, and work with commercial e-commerce analytics and customer experience management systems if the system was put on cloud-based platforms with optimized APIs.

REFERENCES

- [1] Kanagajothi, D., Navaneethan, S., Murugan, C. A., Sunheriya, N., Amer, A., & Karthikeyan, G. (2025, August). Enhancing Customer Experience in E-Commerce with NLP-Driven Chatbots and Virtual Assistants. In 2025 5th International Conference on Soft Computing for Security Applications (ICSCSA) (pp. 1160-1165). IEEE.
- [2] Jahan, N., Ahamed, J., & Nandi, D. Enhancing E-commerce Sentiment Analysis with Advanced BERT Techniques.
- [3] Rashid, M. R. A., Das, A., Hasan, K. F., Hasan, M. R., Sultana, M., Hasan, M., ... & Khan, M. S. H. (2025). BERT-KAN: Enhancing Bilingual Sentiment Analysis in Bangladeshi E-Commerce Through Fine-Tuned Large Language Models. *Natural Language Processing Journal*, 100190.
- [4] Davoodi, L. (2025). Decoding Customer Sentiment: AI-Driven Insights into E-Commerce Trust and Satisfaction.
- [5] Rahman, B. (2024). Optimizing customer satisfaction through sentiment analysis: A BERT-based machine learning approach to extract insights. *IEEE Access*.
- [6] S. Alkhusayni, "Multilingual sentiment analysis with data augmentation," *Information*, vol. 16, no. 9, pp. 1–15, 2025.
- [7] G Loge, T Sunil Kumar Reddy, G Swapna, & G Viswanath. (2025). Interpretable AI for Precision Brain Tumor Prognosis: A Transparent Machine Learning Approach. In *International Journal of Health Sciences and Pharmacy (IJHSP)* (Vol. 9, Number 1, pp. 180–195). Zenodo. <https://doi.org/10.5281/zenodo.15523628>
- [8] Z. Wang, J. Liu, and H. Chen, "A review of Chinese sentiment analysis: Subjects, methods and trends," *Springer Nature Computer Science*, vol. 6, pp. 221–236, 2025.
- [9] H. Pango, "Sentiment analysis in e-commerce: Developing a model using machine learning and deep learning," *J. Comput. Sci. Res.*, vol. 10, no. 2, pp. 78–92, 2025.
- [10] G, Viswanath., N, Madhvik., K, Bhaskar., K, Supriya. (2024). Machine-Learning-Based Cloud Intrusion Detection. *International Journal of Mechanical Engineering Research and Technology*, 16(9), 38–52.
- [11] Z. He, "Main issues and countermeasures of female e-commerce consumption from the perspective of 'her economy'," *E-Commerce Lett.*, vol. 13, no. 3, pp. 5993–5998, 2024, doi:10.12677/ECL.2024.133740.
- [12] T. Chen, P. Samaranayake, X. Cen, M. Qi, and Y.-C. Lan, "The impact of online reviews on consumers' purchasing decisions: Evidence from an eye-tracking study," *Frontiers in Psychology*, vol. 13, p. 865702, 2024, doi:10.3389/fpsyg.2022.865702.
- [13] A. He and M. Abisado, "Review on sentiment analysis of e-commerce product comments," in *Proc. IEEE 15th Int. Conf. Adv. Infocomm Technol. (ICAIT)*, Oct. 2023, pp. 398–406, doi:10.1109/ICAIT59485.2023.10367348.
- [14] C. Ziakis and M. Vlachopoulou, "Artificial intelligence in digital marketing: Insights from a comprehensive review," *Information*, vol. 14, no. 12, p. 664, 2023, doi:10.3390/info14120664.
- [15] K. R. Mabokela, T. Celik, and M. Raborife, "Multilingual sentiment analysis for under-resourced languages: A systematic review of the landscape," *IEEE Access*, vol. 11, pp. 15996–16020, 2023, doi:10.1109/ACCESS.2022.3224136.
- [16] N. Punetha and G. Jain, "Game theory and MCDM-based unsupervised sentiment analysis of restaurant reviews," *Appl. Intell.*, vol. 53, no. 17, pp. 20152–20173, 2023, doi:10.1007/s10489-023-04471-1.
- [17] Y. Li, G. Wang, and G. Yang, "Big data-based transfer learning for sentiment classification with multiple source domains," in *Proc. 3rd Int. Conf. Big Data Informatization Educ. (ICBDIE)*, 2023, pp. 256–265, doi:10.2991/978-94-6463-034-3_26.
- [18] S.-C. Necula, "Exploring the impact of time spent reading product information on e-commerce websites: A machine learning approach to analyze consumer behavior," *Behav. Sci.*, vol. 13, no. 6, p. 439, 2023, doi:10.3390/bs13060439.
- [19] Y. Li and C. Nuangjamnong, "Exploring the impact of social media marketing, customer attitude, and engagement within the quality of review and review valence on customer purchase intention in green cosmetic product in Chengdu, China," *Int. J. Social Sci. Humanities Invention*, vol. 9, no. 12, pp. 7523–7546, 2022, doi:10.18535/ijsshi/v9i012.06.

- [20] W. Shengyuan and H. Jianglin, "Influence mechanism of different types of e-commerce streamer attributes on consumers' purchase: Based on grounded theory and text analysis," *J. Beijing Univ. Posts Telecommun. Social Sci. Ed.*, vol. 24, no. 2, p. 104, 2022, doi:10.19722/J.CNKI.1008-7729.2021.0254.
- [21] Kumar, C. S., Sirisati, R. S., Gudditti, V., Rao, K. S., & Challa, R. K. (2022, December). A smart recommendation system for medicine using intelligent NLP techniques. In 2022 International Conference on Automation, Computing and Renewable Systems (ICACRS) (pp. 1081-1084). IEEE.
- [22] N. M. Alharbi, N. S. Alghamdi, E. H. Alkhamash, and J. F. Al Amri, "Evaluation of sentiment analysis via word embedding and RNN variants for Amazon online reviews," *Math. Problems Eng.*, vol. 2021, pp. 1–10, 2021, doi:10.1155/2021/5536560.
- [23] Sowmya, G., & Swapna, G. (2023). A Real Time Online Food Ordering Application Based Django Restful Framework. *Industrial Engineering Journal*, 52(9), 425–431.
- [24] H. S. Choi and S. Leon, "An empirical investigation of online review helpfulness: A big data perspective," *Decis. Support Syst.*, vol. 139, Art. no. 113403, 2021, doi:10.1016/j.dss.2020.113403.
- [25] P. Vana and A. Lambrecht, "The effect of individual online reviews on purchase likelihood," *Marketing Sci.*, vol. 40, no. 4, pp. 708–730, 2021, doi:10.1287/mksc.2020.1278.
- [26] Z. Zhang, "Legal issues and regulation path of e-commerce platform's improper use of consumer personal information," *Sci. J. Humanities Social Sci.*, vols. 6–7, 2021, doi:10.54691/6ez9h406.
- [27] I. O. Adeniyi, A. Akinkunmi, N. A. Sande, A. A. Author, and I. Oluwasegun, "Social media sentiment analysis: A comprehensive analysis," *ResearchGate Preprint*, 2021, doi:10.13140/RG.2.2.31094.37441.
- [28] S. R. Kumar, "Online reviews: Do consumers trust them?," *SSRN Electron. J.*, 2021.
- [29] R. Mishra and L. Shukla, "Consumer behavior analytics using online review data," *Int. J. Data Sci. Anal.*, vol. 11, no. 5, pp. 520–535, 2020.
- [30] L. Xu and H. Zhang, "Sentiment mining for bilingual e-commerce data using deep neural representations," *Procedia Comput. Sci.*, vol. 176, pp. 245–254, 2020.