IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 16 February 2025, accepted 11 March 2025, date of publication 14 March 2025, date of current version 27 March 2025.

Digital Object Identifier 10.1109/ACCESS.2025.3551666

RESEARCH ARTICLE

Examining Customer Satisfaction Through Transformer-Based Sentiment Analysis for Improving Bilingual E-Commerce Experiences

SHIZHONG SHAN[1], (Member, IEEE), JIN SUN[2], AND REMSON MARK C. MACAWILE[3]

[1]China National Democratic Construction Association, Economic Research Institute, Yango University, Fujian 350015, China
[2]College of Business and Accountancy, National University, Manila 1008, Philippines
[3]College of Engineering, National University, Manila 1008, Philippines

Corresponding author: Shizhong Shan (hellomadisonshan@gmail.com)

ABSTRACT As e-commerce continues its rapid expansion, accurately gauging and interpreting customer sentiment has become critical for optimizing user satisfaction and predicting purchasing behavior. This study advances feature-level sentiment analysis by examining Chinese e-commerce product reviews, translated into English, to uncover the contextual impact of sentiment-bearing keywords across diverse product categories. The research intends to discover trends, reasons, and broad conclusions associated with the volume, co-occurrence, and contextual use of keywords across various product types. Using sentiment intensity analysis, N-gram modeling, and machine learning, this research attempted to evaluate the performance of these models: ELECTRA, BERT, BiLSTM, Random Forest with XGBoost, and SVC with SGD. Empirical results demonstrate that ELECTRA yields the highest accuracy (98.09%) and F1 score (0.9711), illustrating superior performance in capturing nuanced sentiment cues, with low loss of 0.062. BiLSTM followed with 96.09% accuracy and F1 score of 0.9583. Although BERT performed strongly with an accuracy of 95.28% and an F1 score of 0.9528, it is extremely resource demanding. Ensemble methods like Random Forest with XGBoost and SVC with SGD had moderate performance, finding 82.43%, and 88.84% accuracy, respectively, but had better runtimes. The results supports the advantage of using transformer based models in attempting more sophisticated contextual analysis by uncovering paramount sentiment patterns, product specific issues, and customer needs. Furthermore, by translating Chinese reviews into English, the study highlights the potential for cross-lingual frameworks to enrich global e-commerce strategies. Future research directions include integrating cultural factors, extending analysis to additional languages, and refining computational approaches for resource-limited environments.

INDEX TERMS E-commerce, customer satisfaction, sentiment analysis, natural language processing (NLP), machine learning models, sentiment keywords, feature-level sentiment analysis, consumer behavior, customer feedback, N-gram modeling, transformer-based models, ELECTRA, BERT, BiLSTM, random forest, XGBoost, purchase intentions, data-driven decision-making, product category prediction, sentiment intensity analysis.

I. INTRODUCTION

The rise of e-commerce has transformed business and shopping practices around the world. Businesses of all kinds utilize social media for marketing, customer engagement, and analyzing feedback [1]. Social media has become an important source of user-generated reviews, comments, and posts that reflect the sentiments, preferences, and trends of customers [2]. Unfortunately, interpreting such large volumes of information is tedious and time-consuming if done manually, and is incredibly error-prone. The development of Natural Language Processing (NLP) models provides a good alternative for the classification and analysis of social media posts content. NLP models can read and understand

The associate editor coordinating the review of this manuscript and approving it for publication was Barbara Guidi.

the text data which allows a company to classify, determine the sentiment, recognize patterns, and even forecast buying behavior.

With proper sentiment analysis, businesses can make more advanced decisions, engage with customers more effectively, and better tailor their products or services. Yet, there are still specific issues in constructing reliable and precise models when it comes to complex sociolinguistic phenomena such as slang and context in social media language [3]. For example, the application of e-commerce in China is changing dramatically with the incorporation of artificial intelligence and big data technology. China has been estimated to account for more than half of the revenue from e-commerce worldwide in the year 2023. While e-commerce platforms operate in a multilingual landscape, this study specifically focuses on bilingual sentiment analysis, examining customer feedback from Chinese e-commerce platforms that has been translated into English. NLP is also being used by firms such as Alibaba, Baidu, and Amazon to evaluate customer actions and develop marketing approaches. This enables them to identify not only improve customer relations through the use of chatbots but also provides comprehensive insights into consumer purchasing intentions and sentiments, thereby increasing sales and loyalty [4].

Sentiment analysis, a technique that involves the computational study of opinions, emotions, and attitudes within text data, has emerged as a valuable tool for companies to monitor and respond to customer feedback [5]. With the growing demand for effective customer feedback analysis in bilingual e-commerce environments, this research presents an improved approach by integrating transformer-based models with feature-level sentiment analysis to better interpret customer satisfaction. Traditional sentiment analysis approaches focus on the overall sentiment polarity classification, whereas this work studies the expression of sentiment-bearing keywords and analyzes their contextual co-occurrence by product categories. This micro-level perspective allows a closer analysis of the drivers of consumer attitudes and spending. One effective way to gauge customer sentiment is through the analysis of social media posts, where customers often express their opinions and experiences [6]. Companies are really putting this sentiment analysis to work through mining and analyzing customers' opinions, emotions and feedback through text and speeches. Hoar et al. demonstrates that sentiment analysis is not just limited to monitoring the reception of political communications during this coronavirus period or movies on IMDB, rather its applications are numerous and even more surprising. It is evident from the work of Jerripothula et al. that sentiment analysis has recently emerged as an essential tool through which e-commerce stores are analyzing customer reviews [7]. New studies have emerged demonstrating the rich use of sentiment analysis within various disciplines. For instance, during the COVID-19, political communications were measured while sentiments of users on IMDb were analyzed. This proves sentiment analysis can help understand people's opinions as well as consumer behavior. With regards to e-commerce, sentiment analysis has been utilized by companies to scrutinize customer reviews and gain more insights into the product features that are making people satisfied or not. Even if the progress in sentiment analysis has been remarkable, there remains a gap in feature level sentiment analysis in trying to find sentiment words that contribute to customer satisfaction. Most of the previous works focus on the general sentiment of a product or brand while missing out on the specific keywords and themes associated with the customers' experience [8]. This study breaks away traditional sentiment analysis methods by embedding a micro-level approach that starts to use a sentiment polarity categorization at a more granular level (sentiment-bearing keywords), and investigates their contextual co-occurrence in product categories as opposed to using traditional approaches where the focus is on product performance metrics. This leads to a more detailed understanding of consumer attitudes and purchasing behavior. Closing this gap would allow businesses to better identify the resonating attributes and refine their business strategy by providing how to gain actionable intelligence. This research study seeks to understand customer satisfaction based on sentiment keywords found in social media posts. By considering keyword volume, co-occurrence, and contextual use, the research hopes to discover patterns which will explain the key drivers of customer sentiment. It conclusively offers richer content analysis compared to deep content analysis on the factors which result to satisfaction or dissatisfaction, thereby giving certain direction for data-driven decisions and product improvements [9]. This paper adopts NLP techniques to build powerful classifiers that can perform sentiment analysis on large amounts of unstructured social media content. Based on the previous studies, it will examine the effectiveness of some machine learning and ensemble approaches, including word embedding and RNN algorithms in sentiment classification tasks [10]. These methods will also be applied to study the connections between sentiment words and customer satisfaction, which creates prospects for integrating sentiment metrics with business performance measures such as sales or customer profitability. Moreover, this study also compares the performance of various state-of-the-art machine learning and deep learning models, including ELECTRA, BERT, and BiLSTM for sentiment classification purposes. Using a dataset of Chinese e-commerce reviews translated to English, the research seeks to overcome linguistic and cultural challenges in bilingual sentiment analysis to ensure that consumer opinions are properly reflected in both languages. These contributions align this study at the interplay between computational linguistics and business analytics, thus linking more advanced natural language processing techniques and their associated real-world applications in e-commerce. It further underscores the importance of employing both qualitative and quantitative analysis to distill the sentiment and specific concerns of bilingual customers in order to fine-tune marketing strategies and foster greater engagement.

This study sought to address the challenges posed by predictive modeling of sentiment analysis. Determining product categories or features associated with the highest positive reviews is still an active challenge for businesses at large, helping them to improve customer satisfaction and sales [10].

II. RELATED STUDIES

The related studies reviewed in this section address a spectrum of factors influencing consumer behavior and satisfaction in e-commerce. These include demographic and behavioral aspects, platform-specific features, trust dynamics, and sentiment analysis as a methodological tool. Together, these components provide a comprehensive foundation for understanding the multifaceted drivers of consumer behavior in e-commerce.

Existing studies suggest that consumer purchasing behavior is affected by many factors. Personal factors can be summarized as gender and age; External factors include product services, product price, and product quality [11]. User behavior is comprehensively affected by personal factors, psychological factors, technical factors, and social factors, and put forward a series of optimization suggestions from the aspects of search engine optimization, social media marketing, content marketing, and email marketing [12].

A. INSIGHTS INTO E-COMMERCE CONSUMER BEHAVIOR

In terms of different genders, studies show that men's single shopping amount in online shopping will be higher than that of women, while women's shopping frequency is ahead of men [13]. He Zhoubin [14] focused on the unique "her economy" phenomenon formed by women's financial management and consumption, and found that product quality and safety issues, opaque prices and false promotions, inadequate after-sales services, personal privacy and information security issues in e-commerce are Problem focus. Ok analyzed the "other economy" marketing model in the e-commerce market and proposed that the female economy needs precise marketing strategies [15].

Among different age groups, the consumers' consumption behavior is also very different. Young people are more likely to accept new things, so consumption on e-commerce platforms is more popular among young people. Liu and Xiao believes that consumers with different education levels will express different online purchase intentions. The higher the education level and the better understanding of e-commerce, the stronger their purchase intention [16]. In terms of the language and attributes of e-commerce live broadcasts, analyzed the language characteristics and content of e-commerce anchors, and believed that the language characteristics and content of anchors play a very important role in e-commerce [17]. There are different types and attributes of e-commerce anchors based on grounded theory and text analysis, and believed that the type of e-commerce anchors has a great impact on consumer purchasing behavior [18]. Zhang conducted an empirical analysis based on the privacy policies of 12 e-commerce platforms in 6 categories and believed that the protection of personal information of e-commerce consumers urgently needs to be standardized and protected [19]. On the other hand, a study on personal income, found that high-income groups are more willing to try new shopping methods and have a stronger willingness to purchase online [20]. The difference in purchase intention between rich online shopping experience and not is not obvious, but rich online shopping experience can affect consumers' perceived risk. There are different influencing factors of e-commerce user loyalty based on the perspective of community shopping experience and believed that user satisfaction and trust positively affect user loyalty [21], which further affects the accumulation of loyalty and helps e-commerce companies improve their market competitiveness. and adapt to changing market conditions. The level of professional knowledge will affect the process of searching, obtaining and final decision-making for product information, so the reader's different knowledge level or experience will also affect the usefulness of online reviews [22]. Therefore, the individual gender, age, education, and cognitive level of the user, as well as the voice, language content, and type of the e-commerce platform anchor all have an impact on the purchase intention of e-commerce platform users.

B. EXTERNAL EVALUATION FACTORS OF E-COMMERCE PLATFORMS

When exploring external factors of consumers' purchase intention, most scholars focus on cross-border e-commerce websites or platforms, believing that website quality and platform services will affect purchase intention. Cross-border e-commerce platforms and the construction of e-commerce platforms has a significant impact on consumers' online purchase intentions [23]. In terms of online reviews. The usefulness of comments on online reviews means that online shoppers change their views on products through browsing online reviews, which in turn affects their self-perception of purchase desire [24]. The product type has a significant moderating effect on review length, as well as the number of historical reviews by the publisher and the usefulness of online reviews. In another study, the review length and star rating also matters in purchasing items [25]. The longer the length of the review, the richer information consumers can obtain, which can further enhance their knowledge and understanding of the product, so the perceived usefulness of the review is higher. When consumers are faced with a large amount of positive information, they will think that reviewers have more subjective emotional tendencies, and neutral or negative reviews can reveal more deficiencies of goods and services. Reflect more real and useful information. It has also been suggested that the number of days online reviews are published, the more helpful it is for consumers to make purchase decisions [1]. In addition, that extreme comments can get more usefulness votes than neutral comments [26]. Scholars such as Yin Guopeng and others [27] concluded through empirical analysis of movie review data on Douban.com that there is a certain correlation between the length of reviews and the usefulness of reviews. The relationship between them can be represented by an inverted U shape. When within a certain range, relative Longer comments are more attractive for readers to browse carefully, but if they exceed a certain threshold, they will increase the reader's information load, reduce the reader's interest, and then negatively inhibit the usefulness of the comments [28], [29], [30].

1) DISCLOSURE OF PERSONAL INFORMATION OF COMMENTERS

The authenticity of the review information has been linked to factors such as the disclosure of personal details, which fosters trust among consumers [31], [32]. Some studies suggest that the word-of-mouth is the core indicator that distinguishes different reviewers. Consumers are more likely to further read the comments written by reviewers with high word-of-mouth. Therefore, compared with those posted by reviewers with low word-of-mouth The usefulness of comments will be higher [33], [34]. Qin Guishen et al. targeted anchors and consumers in e-commerce live broadcasts and found that anchors can control the live broadcast through user interaction mechanisms, editing and guidance mechanisms, security and review mechanisms, value-added and cost mechanisms, infrastructure and rules, etc. Live broadcast users' consumption concepts and purchasing behaviors have an impact [2].

C. E-COMMERCE USER SHOPPING EXPERIENCE

Research shows that user experience is a decisive factor in users' repeat purchasing behavior. Search engine user experience from the dimensions of visual experience and logical rule experience and proposed optimization methods [15]. There is a correlation between brand experience, brand community and consumer loyalty and user experience has a significant positive impact on user loyalty. However, when consumers participate in online shopping activities, there are large differences in the impact and mechanism of different experience components on user loyalty [35], and customer loyalty varies in different models and types of e-commerce platforms. There are also differences in the direction, path and intensity of influencing factors. The influencing factors of user loyalty in the context of a single online community or traditional e-commerce platform cannot be applied to community-based e-commerce invariably [36]. The research by Kim et al. shows that high-quality service is a factor that affects consumer brand satisfaction. The shopping experience of e-commerce users is a complex and comprehensive experience, which is closely related to visual, brand, service and other variables [29].

Relying on the big data era represented by artificial intelligence, cloud computing, and the Internet of Things, the number of e-commerce users is growing rapidly. Therefore, studying e-commerce user evaluation and feedback is an important part of maintaining and expanding customers. User sentiments, comments, and reviews are predictive of the performance of products and buying behavior of consumers [37].

1) SENTIMENT ANALYSIS

Sentiment analysis has become an essential technique for understanding public opinions, preferences, and emotions, particularly in contexts such as e-commerce, social media, and user-generated content (UGC) [38]. The surge of user-generated material on social media sites like Twitter, Facebook, and Instagram has provided academics with an abundance of data to examine popular mood and opinion. Sentiment analysis (SA), or opinion mining, offers insights into public attitudes by classifying text into sentiment categories such as positive, negative, or neutral. The applications of SA encompass e-commerce decision-making and policy assessments in education [39]. Consumers' review search behaviors and purchase intentions differ as a function of exposure to products. In a study, Fu et.al identified the acceptance barriers to using recycled water reuse in businesses and e-commerce intelligence. The paper helps understand how different degrees of human contact influence consumer behavior and their review searching efforts [40]. The paper suggests that better design of product information and reviews on e-commerce sites enable consumers to make informed decisions and willingly accept new and sustainable products.

III. METHODS

A. DATASETS

Social media dataset is an invaluable tool for researching customer behavior and preferences in the retail and e-commerce industries since it encompasses a range of attitudes, from critical complaints to favorable recommendations [41]. Customer reviews are a pivotal aspect of e-commerce, offering insights into consumer satisfaction, product quality, and service effectiveness [42]. The dataset utilized in this study comprises consumer reviews sourced from various Chinese online marketplaces, reflecting diverse purchasing and product usage experiences. Both datasets were in Chinese and translated to English using the Google Translation API. We verified a random subset of the translated reviews manually (to ensure the consistency of translation) and text normalization (stopwords removal, tokenization, and stemming). This study utilizes the following datasets:

1) ONLINE SHOPPING - 10 CATEGORY DATASET

The online_shopping_10_cats dataset (Dataset A) is a rich resource widely used for sentiment classification tasks, particularly in Chinese-language contexts [43], [44]. It comprises over 62,000 product reviews spanning ten distinct categories, including books, tablets, mobile phones, fruits, shampoos, water heaters, dairy products (e.g., Mengniu), clothing, computers, and hotels. Each category contains approximately 6,000 reviews, balanced between positive and negative sentiments, with around 3,000 reviews for each sentiment class. The dataset includes several features: the original Chinese text of the review, the product category, and a binary label indicating sentiment, where '1' represents a positive review and '0' a negative one. To enable sentiment analysis in English, the dataset has been enhanced by translating the Chinese reviews into English using the Google Translate API, adding an English translation feature to its structure [45]. This dataset is instrumental for researchers and practitioners focusing on sentiment analysis, particularly in cross-lingual applications, offering a diverse and balanced collection of data that can support the development and evaluation of robust sentiment classification models. The dataset captures a wide range of reviews across multiple product types and dives into consumer behavior and preferences to understand sentiment patterns and purchase satisfaction.

2) PRODUCTS DATASET

The Products Dataset (Dataset B) is scraped from different Chinese e-commerce sites [46]. This dataset comprises 10,000 customer reviews spanning a diverse range of product categories, including electronics, clothing, dairy products, and personal care items. Primarily written in Chinese, the dataset offers a rich source of textual data for sentiment analysis, customer feedback interpretation, and natural language processing (NLP) research. It includes a spectrum of sentiments, from positive endorsements to critical complaints, making it a valuable resource for studying consumer behavior and preferences in retail and e-commerce sectors. Initially, the dataset was in Chinese and later translated into English using the Google Translate API to facilitate broader accessibility and analysis. It primarily contains a single column labeled "Reviews," housing textual feedback from consumers. The content spans numerous domains, such as electronics, personal care, and grocery items, providing a rich landscape of customer sentiments. These reviews vary from straightforward product assessments to detailed grievances about service quality and product defects. This textual diversity enhances the dataset's utility for sentiment classification tasks, allowing the capture of nuanced consumer opinions. This dataset serves as a comprehensive repository of product reviews collected from an online retail environment, making it well-suited for academic research in sentiment analysis, opinion mining, and natural language processing. Its bilingual nature, dominated by Chinese text, poses unique challenges and opportunities for NLP research in non-English settings. Despite having only one column, the dataset indicates a mix of positive, neutral, and negative sentiments, often accompanied by specific product details and contextual complaints. This variability makes it particularly suitable for natural language processing tasks, such as sentiment analysis, customer satisfaction evaluation, and market trend prediction. With this, it offers direct analysis into consumer behavior and attitudes. Furthermore, the dataset serves as a foundational resource for understanding customer sentiment in Chinese e-commerce while using English-translated content for global applicability. Future studies could expand on its utility by integrating categorical tags or numerical sentiment scores for further refinement in supervised machine-learning models.

B. DATASET ANALYSIS

1) DATASET A SENTIMENT-BASED DESCRIPTIVE STATISTICS

This study used the Sentiment Intensity Analyzer from the NLTK package to perform the analysis across different product categories using the dataset. This tool derived better sentiment scores (ranging between -1, most negative, and 1, most positive) from the more naïve dataset comprised of binary labels. To explore the relationship between the sentiment expressed and the verbosity of feedback, statistical measures like the mean, median, standard deviation, and count of sentiment scores were calculated for each category, in addition to the average length of the reviews.

Fig. 1 shows box plots of sentiment per category, including median and distribution by category. Average post length, categorized into 'positive' and 'negative,' is indicated on the y-axis, and the corresponding frequencies expressed in bars indicate y-axis values in Fig. 2, which indicate a positive correlation with the length of feedback posts. 2. Fig. 3 provides an overall perspective of the sentiment score distributions across all categories with histograms with density plots overplotted detailing the frequency and nature of sentiment scores.

The Books category demonstrated the highest average sentiment score of around 0.481, a median of 0.727, as shown in the standard deviation of 0.551 which reveals a good reception but great divergence from readers. Reviews in this category were also the longest, averaging about 479 words, suggesting that more thorough criticism is usually more favorable. Conversely, the average sentiment scores for the Clothes and Fruit Categories were considerably lower at 0.303 and 0.224 respectively, both with high standard deviations of more than 0.57, indicating mixed feelings amongst consumers. Notably, hotel reviews had the longest reviews on average, 441 words on average, which could represent more detailed critiques or praise specific to services provided. Both categories exhibited average sentiment scores with the Water Heater category at a relatively high average score of 0.493, potentially indicative of product and owner satisfaction due to the essential nature of the product even though the volume of reviews were ultimately less than those of the Water Softener category.

2) DATASET A N-GRAM ANALYSIS

N-grams allow for a more granular exploration of consumer sentiment where word order and context play a critical role in understanding sentiment nuances. The methodology for analyzing the Chinese grocery product reviews through N-gram modeling began with data preparation, where a.txt file containing the reviews was uploaded. The next step involved

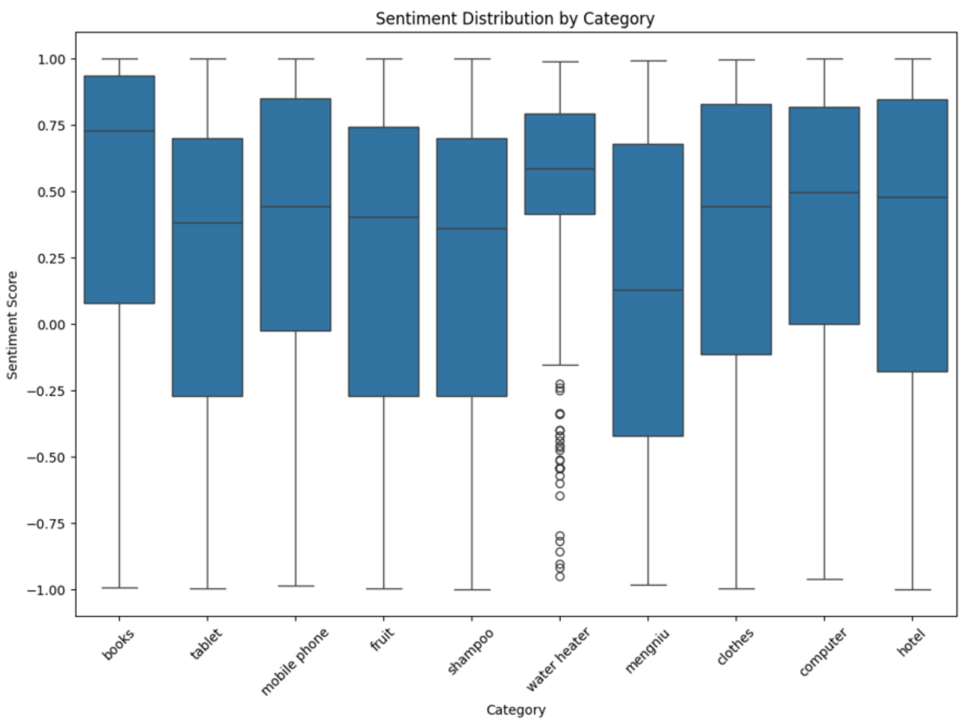

FIGURE 1. Sentiment distribution by category.

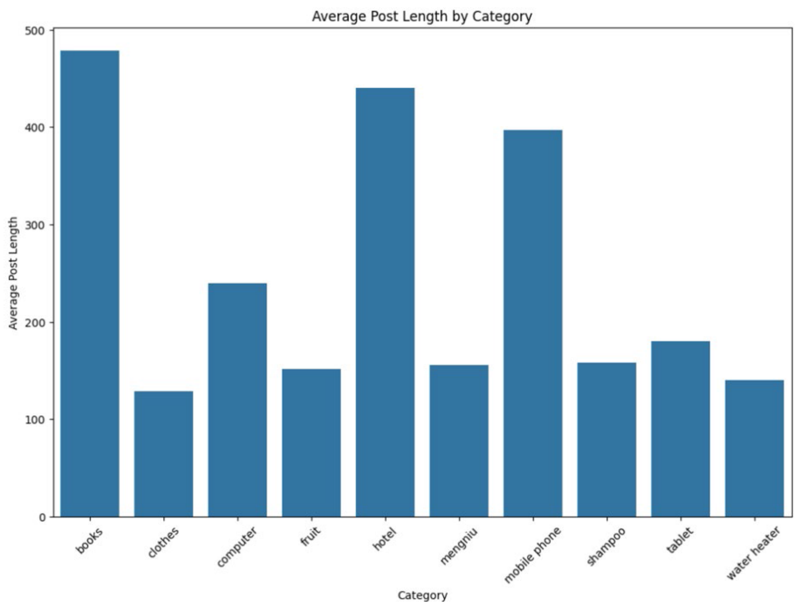

FIGURE 2. Average post length by category.

acquiring a comprehensive list of Chinese stopwords downloaded from a trusted online repository on GitHub [47]. These stopwords were integral to filtering out common, non-informative words during the preprocessing stage, thereby enhancing the quality of the analysis. Text preprocessing included several critical steps. Initially, non-Chinese characters, punctuation, and symbols were removed from the text using regular expressions. The clean text was then tokenized using Jieba, a widely recognized Python library for Chinese text segmentation, ensuring the accurate splitting of text into meaningful words or phrases. The tokenized words were subsequently filtered against the Chinese stopword list

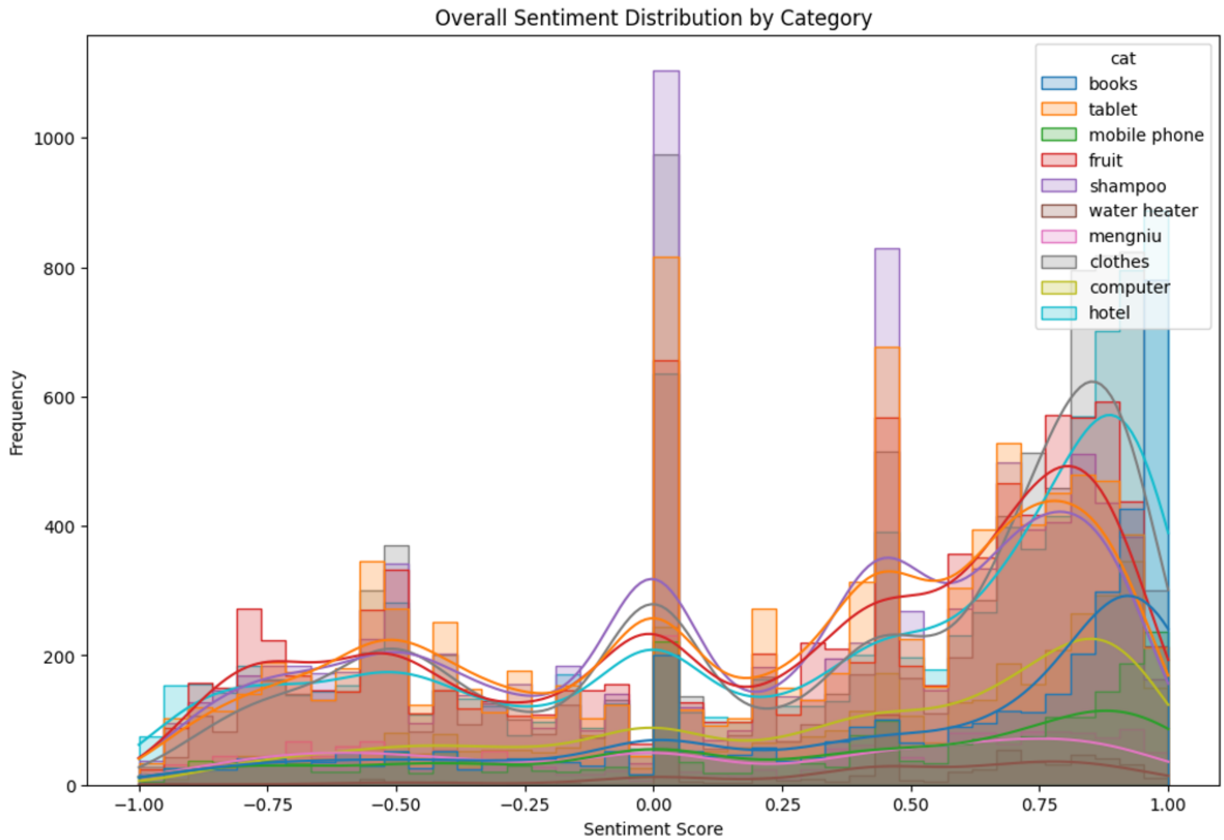

FIGURE 3. **Overall sentiment distribution.**

to remove irrelevant terms and retain only those essential for meaningful analysis.

N-grams facilitate the identification of sentiment-laden keywords and phrases that directly influence customer satisfaction and buying decisions. For example, recurring terms such as ‘great quality’ or ‘delayed delivery’ highlight product-specific drivers of satisfaction or dissatisfaction. By focusing on bi-grams and tri-grams, the method balances computational efficiency with interpretability, making it suitable for analyzing moderately sized datasets while uncovering meaningful patterns.

N-grams, which are sequences of consecutive words, were generated from the preprocessed text with an N-gram size (n) of 30. These N-grams were counted for frequency using the Python collections. Counter library, resulting in a frequency distribution that highlighted the most common word sequences. From this distribution, the top 25 most frequent N-grams were selected for visualization. To visualize the relationships between terms, a network graph was constructed. Each word in the selected N-grams was represented as a node, and consecutive words formed edges, with edge weights corresponding to the frequency of their co-occurrence. The network was structured using NetworkX, and node positions were computed using the spring layout algorithm to ensure an intuitive and evenly spaced arrangement. Figure 4 shows the top N-Gram generated using the dataset. The n-gram was rendered as an interactive visualization using Plotly. Nodes were sized and colored based on their degree of connectivity, and edges represented the strength of the relationships between terms. The resulting visualization allowed for an in-depth exploration of recurring themes and relationships within the reviews. Insights derived from this analysis were interpreted to identify key patterns, such as frequently mentioned product attributes, brands, and customer concerns. The n-gram data presents a comprehensive overview of prevalent themes, feelings, and focal points in a dataset presumably related to customer feedback, especially within a Chinese environment. Analyzing these patterns allows this research to extract information about the audience’s primary issues and priorities, as well as prospective avenues for enhancement or additional investigation. Predominant phrases in the dataset include “Consumer,” “Service,” “Good,” “No,” and “Problem,” which indicates a significant emphasis on customer experiences and views of quality and service. This infers that the consumers articulate their experiences with products and services, notably expressing both happiness and disappointment. Particular product categories such as “Milk,” “Dairy,” “Meat Products,” “Wine,” and “Baiju” underscore the significance of consumable products in the dataset, potentially pertaining to food safety,

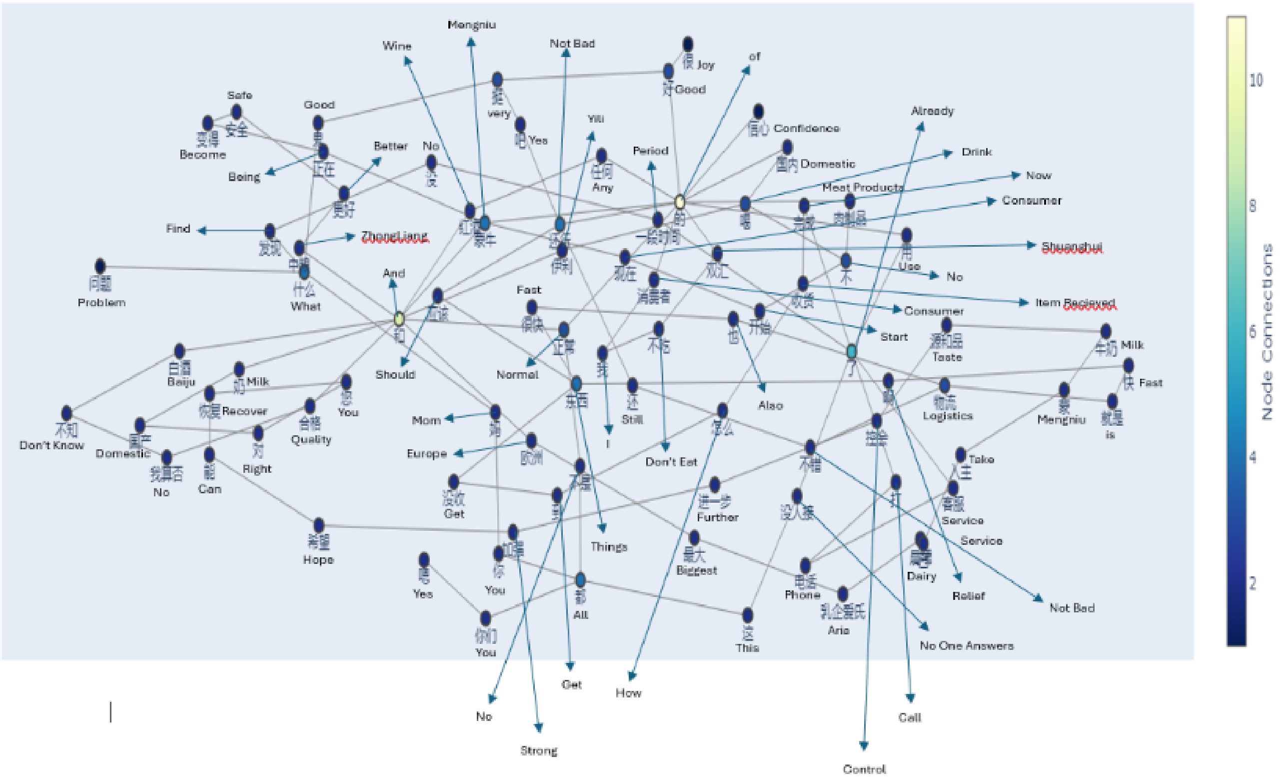

FIGURE 4. Dataset B NGRAM.

product quality, or brand reputation. The data also showed presence of expressions such as "Good," "Not Bad," "Hope," and "Confidence" signifies positive feeling, but phrases like "Don't Eat," "No One Answers," "Problem," and "No" imply discontent or anxiety. The contrast between these positive and negative attitudes highlights the diverse experiences of customers. The term "Don't Eat" serves as a straightforward warning against particular items, whereas "Confidence" and "Hope" signify confidence in specific companies or aspirations for enhancement. There is also the presence of the use of brand names such as "Shuanghui," "Mengniu," "Yili," and "ZhongLiang" indicates that these firms are prominent market participants and are often referenced by customers. The context in which these brands are used, such as in conjunction with phrases like "Quality" or "Safe," undoubtedly offers insights on public perception. Brands connected with favorable connotations may be perceived positively, but those related with negative attitudes could have reputational difficulties. The n-gram also showed terms like "Safe," "Quality," and "Problem" indicate a significant emphasis on safety and quality, possibly representing the audience's foremost objectives. The recurrence of these phrases indicates persistent difficulties or concerns that may necessitate specific measures, such as enhancing product standards or revising safety practices. Likewise, "Logistics" and "Service" may underscore operational difficulties, such as prompt delivery or client assistance. Words such as "Fast," "Normal," and "Good" offer insight into consumer inclinations, highlighting efficiency and dependability. The recurrence of "Milk" and "Dairy" indicates that these are fundamental goods in the dataset, mirroring cultural or market-specific consumption patterns. Expressions such as "No One Answers" and "Don't Know" suggest possible deficiencies in communication or knowledge, which may adversely impact customer satisfaction. The dataset's concentration on domestic items, shown by phrases such as "Domestic," highlights a preference for local goods. This may be associated with national pride, economic considerations, or perceptions of quality and confidence in local brands. The n-grams "Already" and "Recover" may pertain to consumer anticipations for timely assistance or product restitution following a problem. There is a nuanced relationship between pleasure and discontent, emphasizing consumer experiences about product safety, service quality, and brand trust. Comprehending these trends might assist firms or researchers in addressing consumer preferences and enhancing audience engagement.

Phrase frequency and sentiment patterns in this paper uncover substantial insights into the fundamental themes and emotional nuances present in the collection. An analysis of phrase frequency and corresponding sentiment ratings reveals numerous significant insights, enhancing comprehension of customer perception and dialog. The predominant phrases, including "Consumer," "No," "You," and "Service," suggest a strong emphasis on consumer-oriented dialogues, possibly pertaining to encounters with products or services. The prevalence of "Consumer" as the most frequent phrase (10 occurrences) highlights its significance in the examined discourse, indicating a dataset likely sourced from consumer feedback or contact logs. The frequent use of the phrase "No" (8 times) may indicate recurring themes of dissatisfaction, negativity, or denial, maybe associated with client complaints or service deficiencies. The use of phrases such as "Service" and "Logistics" underscores a significant focus on operational efficiency and customer experience, presumably shaped by the caliber of service delivery systems. The sentiment analysis indicates a varied emotional environment. Positive sentiment expressions, such as "Not Bad," "Good," and "Safe," indicate an inherent appreciation or contentment with specific features of items or services. The phrase "Not Bad" attains a moderately favorable emotion score of 0.35, indicating reserved acceptance rather than fervent endorsement. This restrained optimism indicates that although certain customer experiences are adequate, they are not outstanding. Conversely, phrases like "No," "Problem," and "No One Answers" signify negative attitude, reflecting displeasure or irritation. These phrases, together with their prevalence, indicate persistent pain points in consumer encounters, including unsolved issues or insufficient customer care. Neutral phrases such as "You," "Get," and "Consumer" highlight their utilitarian roles in speech, acting as links or references rather than emotionally laden statements. Their frequency underscores their structural significance in shaping the larger context of the dataset. Commonly referenced phrases generally exhibit neutral to mildly negative sentiment ratings, indicating that customer discussions frequently focus on concerns or requirements rather than solely happy experiences. The juxtaposition of phrases with elevated sentiment ratings, such as "Safe" and "Good," against their comparatively lower frequency suggests that unambiguous statements of satisfaction proof of focused efforts to enhance good customer experiences.

C. MODEL SELECTION AND TRAINING PARAMETERS

For the task of predicting product categories based on online review posts, a variety of machine learning models were employed to ensure a comprehensive evaluation of performance across different architectures and methodologies. The models used were Linear Regression with Stochastic Gradient Descent (SGD), Support Vector Classification (SVC) with SGD, Random Forest, XGBoost, BiLSTM, ELECTRA, and BERT. Each model was chosen for its unique strengths in handling text data and predictive tasks.

For regression of the model, Linear Regression with SGD, provides a baseline for comparison by modeling linear relationships between features. SVC with SGD builds on this by implementing classifier chains, effectively decomposing multi-label classification tasks into a sequence of binary classification problems, which enhances its robustness and adaptability for high-dimensional data. Random Forest and XGBoost are ensemble methods that leverage decision trees for which the former emphasizes simplicity and stability, while the latter enhances accuracy through gradient boosting. The BiLSTM model, a deep learning architecture, is particularly effective in capturing sequential dependencies in text data by processing inputs in both forward and backward directions. ELECTRA and BERT, state-of-the-art transformer-based models, excel in understanding contextual relationships within text, making them particularly suited for complex natural language processing tasks.

The computational complexity and capacity to identify patterns in the data of the models vary greatly. Compared to ensemble approaches or deep learning architectures, standard models such as SVC and Linear Regression may have trouble capturing deeper semantic correlations, despite their computational efficiency. Although Random Forest and XGBoost perform well on structured data, they are unable to effectively make use of text's sequential structure. While transformer-based models like ELECTRA and BERT offer unmatched performance by utilizing pre-trained language representations and attention processes, BiLSTM fills this gap by integrating temporal dependencies.

1) COMPUTING MACHINE SPECIFICATIONS

Furthermore, the models were trained using a Dell Precision 7770 workstation equipped with an Intel Core i9-12950HX Processor, 32GB DDR4 RAM (3200MHz), 1TB M.2 PCIe NVMe SSD storage, and an NVIDIA A3000 GPU with 12GB GDDR6 memory. This hardware configuration made it possible to handle computationally demanding jobs efficiently, especially for deep learning models like ELECTRA, BERT, and BiLSTM. By exploiting the NVIDIA A3000 GPU's parallel processing capabilities, TensorFlow models were accelerated with CUDA, greatly cutting down on training durations and increasing computational efficiency. Faster gradient calculations and matrix operations were made possible by CUDA, which is essential for deep learning frameworks, especially for models with big parameter spaces. A consistent framework for comparison analysis with regard to time limitations was provided by the 50 epochs that each model was trained for, which allowed for enough iterations for parameter optimization and convergence.

IV. RESULTS

A. MODEL PERFORMANCE

This section shows the result of the six prediction models on their performance to annotate the online reviews with their respective categories. As user-generated content continues to grow at an explosive rate, the performances of these models are vital as it helps to automate the organization and analysis of huge amounts of user-generated content, allowing businesses to understand consumer behavior and preferences more thoroughly.

TABLE 1. Model performance.

Model	*loss*	*accuracy*	*precision*	*recall*	*f1-score*	*Runtime (Hours)*
Linear Regression with SGD	2.4564	0.3320	0.3430	0.3169	0.2771	0.0417
SVC with SGD	0.2524	0.8884	0.8760	0.8627	0.8685	0.0933
Random Forest with XGBoost	0.6122	0.8243	0.8957	0.8012	0.8358	0.5817
BiLSTM	0.1326	0.9609	0.9613	0.9555	0.9583	2.9833
ELECTRA	0.0625	0.9809	0.9819	0.9623	0.9711	10.2883
BERT	0.1319	0.9528	0.9529	0.9528	0.9528	27.8917

The results highlight the effectiveness of each approach in predicting product categories based on online review posts that provide valuable insights into their strengths and limitations. The metrics used showed a detailed comparison of the different models' predictive accuracy, precision, recall, F1-score, and overall runtime. The results are displayed in Table 1.

The models' comparison yields unique strengths and weaknesses for each model in relation to their ability to classify product categories from online review posts. As seen through the accuracy, precision, recall, F1-score, and runtime metrics highlighted in Table 1.

V. DISCUSSION

To validate the consistency between datasets and model predictions, the cosine similarity analysis between translated text pairs, along with performance agreement rates between models, was examined.

A. DATASET TRANSLATION ACCURACY

A mean cosine similarity score of 0.6491 between the original Chinese product reviews and their English translations using Google Translate indicates moderate to strong alignment between the two text values. This indicates that the translation does contain the actual meaning of the input text but still could give some different structure and word choice. This means that, in some cases, google translate will translate the Chinese into interpreted language rather than a direct translation. However, as shown by the performance across the Dataset A and Dataset B, most of the translated English reviews can keep the meaning of their Chinese counterpart. The yielded cosine similarity score implies that much of the meaning is preserved, although there may be drift in the meaning of certain translated words due to the challenges in mapping those words through translation.

B. MODEL PERFORMANCE ANALYSIS

ELECTRA was the best performer among all the models by achieving the maximum accuracy of 0.9809 and F1-score of 0.9711 with an astonishingly low loss of 0.0625. The specialized transformer architecture is particularly good at understanding contextual relationships in text, and its ability to classify categories with pre-training on broad corpuses is what allows it to make blunders of this nature. Regardless of these positives, the most striking fact about ELECTRA is its considerable runtime of 10.29 hours, which speaks to the resources needed for most transformer models. With a tradeoff to computational resources, BiLSTM performed admirably with an accuracy of 0.9609 and an F1 score of 0.9583, boosted by a low loss of 0.1326. These results were possible because BiLSTM is particularly good at capturing sequential dependencies in the text because of the bidirectional nature of its architecture. In addition, BiLSTM has a favorable runtime of 2.98 hours, which together with the performance is greatly beneficial in environments where resources are limited unlike in those needed for transformer-based models.

With BERT, another transformer-based model, the accuracy score and F1 score were 0.9528 with a loss of 0.1319 BERT's results were only marginally lower than that of ELECTRA's score, however, it's still a strong contender for natural language processing. The runtime of 27.89 hours is the longest among all compared models, further proving the extensive nature of resources needed to process BERT. These actual computational requirements set boundaries of what can be achieved with this model and show deep consideration is needed when the resources versus performance trade-off is weighed.

Random Forest with XGBoost achieved an accuracy of 0.8243 with an F1 Score of 0.8358. Its ensemble approach performed well given the dataset's high complexity as evidenced by the high precision of 0.8957, and moderate recall of 0.8012. These numbers along with the 0.58 hours runtime shows that Random Forest with XGBoost gets the job done for tasks needing a good level of accuracy with low computational cost. On the other hand, SVC with SGD exhibited even better performance with accuracy of 0.8884 and F1 Score of 0.8685. Its optimization of hyperplanes for class separation shows effectiveness for multidimensional problems. Furthermore, SVC with SGD was able to achieve a precision of 0.8760 and a recall of 0.8627, which depicts a slight bias towards false positive minimization. Because its 0.09 hours runtime is so short, SVC with SGD is useful for accurate, albeit not perfectly accurate, classifications.

The acquisition of lower tier performance data was conducted by utilizing Linear Regression with SGD, which served as the baseline model. Along with this, an F1 score of 0.2771 was calculated to include deviations and misclassifications into consideration. These values are paired with an accuracy of 0.3320. The loss of the model is very high (2.4564) which indicates that the model is very poor at capturing the non-linear relationships that exist in the data. While it did have the fastest runtime of 0.0417 hours, its performance and accuracy is not sufficient for an attempt as complex in nature as this.

The results reveal that transformer-based models outperform others, with ELECTRA being the most useful for text-classifying tasks. Its capability to perform best is due to its unique transformer architecture that captures text's context at a deeper level compared to other models. But this extreme accuracy and F1-score comes with truly high costs in terms of resources, as seen by the long runtime. The high resource consumption of ELECTRA makes it suitable to problems where classification accuracy has to be maximized, for instance for large-scale text mining requiring accurate classification, but not for other less powerful systems that are resource-constrained. BiLSTM, on the other hand, is a strong performer because it achieves a superior balance between accuracy and resource consumption. With bidirectional capability, BiLSTM captures sequential dependencies within the text data allowing understanding of complex relationships within the data. ELECTRA begets higher performance and lower runtimes makes it more enticing for applications where resources are scarce, such as real-time systems and embedded systems. BiLSTM now seems attractive for mid-tier systems where performance and application operating sides are both important.

Having a lower threshold for accuracy largely opens up options for choosing models. One such example is the Random Forest with XGBoost and SVC with SGD implementation for classification tasks. Random Forest with XGBoost reporting accuracy and recall in the higher ends makes logging time lower helps it fit into the category of lesser accuracy tasks. Similarly, XGBoost performs well with a high level of dimensionality because it is blended with SVC that uses SGD. In addition, SVC with SGD is lower in weight and reports high classification accuracy with lower dimensions of data. This would only be possible because of his hyperplane optimization method which increases the speed of target variable prediction while reducing the accuracy thresholds. Still, the slightly lower recall compared to the precision dominant value suggests he misses more than he should. This makes the model more convenient for classification tasks rather than the other way around. A good instructor would not use anything less than Linear Regression with SGD. While this unsophisticated algorithm does sit in the lower end of the performance metric range, its simplicity does offer an unhelpful amount in terms of calculations. Unfortunately, like most other simpler algorithms, it struggles with anything non-linear, making the more complex tasks of text categorization close to impossible. Still, it does formulate strong bounds for more advanced algorithms.

C. MODEL VALIDATION USING DATASET A

In this study, Dataset A was the primary dataset for model training, and Dataset B was the verification dataset to evaluate model. By addressing the differences in structure and labeling, this technique effectively enables sentiment classification models trained on the more structured and labeled Dataset A to generalize and accurately predict sentiment in Dataset B, which is more diverse and lacks structured labeling. In comparison of the transformer models, ELECTRA and BERT, each model was used to predict Dataset A, which are the product reviews in the collected scrapped dataset from different Social Media sites of Chinese reviews in the total of 10,000 data. It resulted in 90.6% match in the predictions of ELECTRA and BERT, showing high agreement in the predicting Chinese reviews data from the two transformer models trained using Dataset B, with 62,774 rows of data. This comparison of two transformer models serves as a means to validate the trained models.

1) PREDICTION COMPARISON AND AGREEMENT OF TRANSFORMER MODELS

In the comparison of the transformer models, ELECTRA and BERT, each model was used to predict the product reviews in the collected scrapped dataset from different Social Media sites of Chinese reviews in a total of 10,000 unlabeled data. It resulted in 90.6% match in the predictions of ELECTRA and BERT in Table 2, showing high agreement in the predicting Chinese reviews data from the two transformer models trained with 62,774 rows of data.

TABLE 2. Prediction comparison of transformer models.

Transformer Model	Comparisons
Matched Predictions	9060
Not-Matched Predictions	940

The analysis of the ELECTRA and BERT models' predictions demonstrates a strong level of agreement, as evidenced by an agreement percentage of 90.6% and a Cohen's Kappa score of 0.7617. The agreement percentage reflects the models' consistent alignment in their classification of reviews across the 10 categories. However, this metric does not account for random agreement, which Cohen's Kappa addresses. The Kappa score of 0.76 indicates substantial agreement between the two models, emphasizing that their alignment goes beyond chance. Despite this strong performance, the remaining 23.8% disagreement highlights differences in how the models interpret specific categories or handle edge cases.

The confusion matrix provides deeper insights into the areas of agreement and disagreement. A notable observation is the overwhelming alignment in the "computer" category,

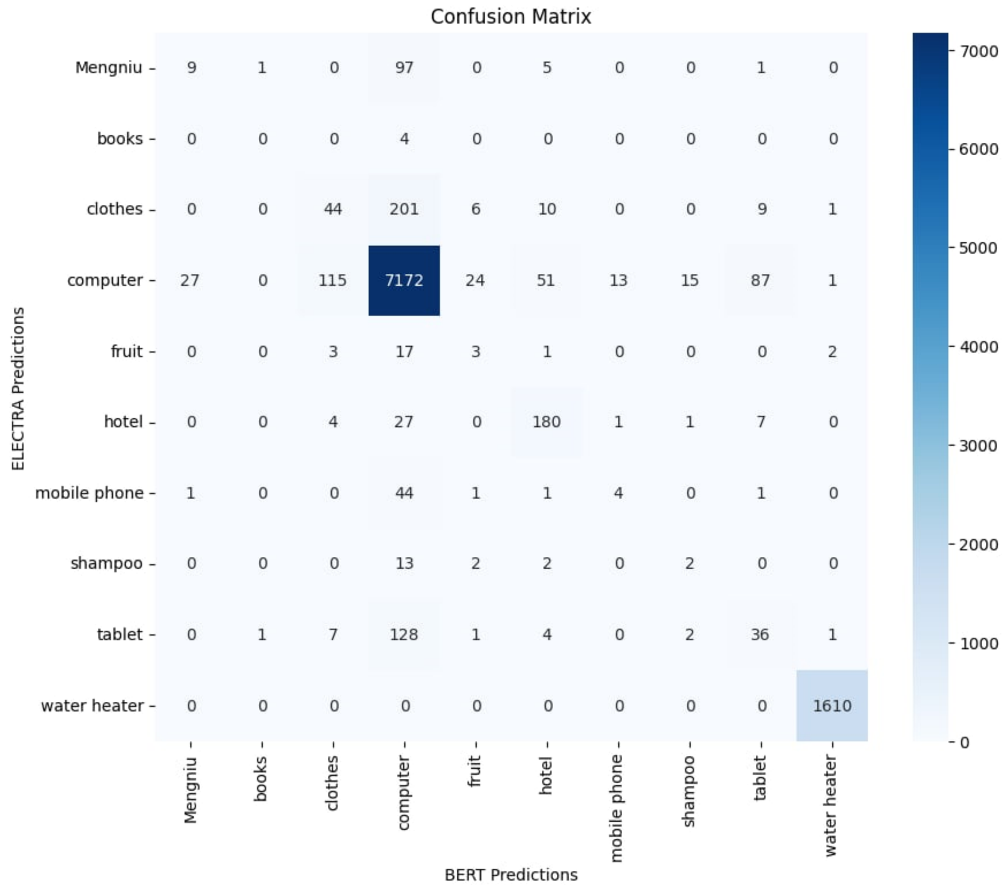

FIGURE 5. Confusion matrix.

with 7,172 matching predictions, making it the most agreed-upon class. However, this dominance suggests an imbalance in the dataset, where the "computer" category appears to be overrepresented. This overrepresentation could inflate the agreement metrics while overshadowing the models' performance on smaller categories. For instance, categories like "books," "fruit," "shampoo," and "water heater" show sparse diagonal values, indicating underrepresentation and limited robustness in these predictions. Additionally, off-diagonal values, such as 201 disagreements between "clothes" and "computer" or 128 mismatches between "tablet" and "computer," reveal significant confusion in semantically overlapping categories. These disagreements likely stem from the shared contexts or keywords in reviews, making it challenging for the models to distinguish between them.

These results highlight key implications for improvement. The imbalance in the dataset, particularly the dominance of the "computer" category, skews the evaluation, leading to inflated agreement percentages and overshadowing the models' ability to generalize across minority classes. Furthermore, the substantial disagreements in categories like "clothes" and "tablet" suggest insufficient training examples or unclear boundaries between these classes.

To enhance the performance of the classification system, several recommendations are proposed. First, balancing the dataset by augmenting underrepresented categories such as "books," "fruit," and "water heater" would improve the models' ability to classify these categories reliably. Second, a detailed error analysis should be conducted to investigate patterns in misclassification, particularly for off-diagonal disagreements such as those between "clothes" and "computer." This can help identify specific linguistic or contextual features leading to errors. Third, redefining or merging semantically overlapping categories, such as "tablet" and "computer," could reduce classification ambiguity. Finally, leveraging ensemble techniques or rule-based post-processing could refine the models' predictions, particularly in cases where their outputs differ.

D. ADVANTAGES OF THE SELECTED MODELS

This study examined several recent machine learning and deep learning models suitable for e-commerce sentiment analysis. And the relative strengths of each approach vary with the nature of the datasets: computational efficiency, accuracy, understanding of context, and resource limitations. Linear Regression using SGD provided a solid baseline, demonstrating that more complex techniques can gradually advance from simple, efficient approaches to accurately modeling the subtle patterns in language that characterize typical user comments. Due to its augmentable ability of non-linear decision boundary and high dimensional textual features, it is balanced with the classification performance and runtime efficiency with SVC and SGD. By combining a large number of simpler algorithms such as decision trees (e.g. the Random Forest with XGBoost), ensemble methods achieved a strong performance through reducing overfitting risks and increasing interpretability, but tree-based methods are unable to model deeper contextual dependencies. On the other hand, BiLSTM's recurrent framework learned sequential dependencies in both forward and backward directions, allowing it to identify nuanced sentiment signals in text with relatively low computational cost. ELECTRA proved to be the best-performing classifier among transformer-based networks as it utilises a novel pre-training objective that enhances its ability to discriminate contextual relationships better. However, its relatively high requirements for resource and time utilization may render it unsuitable for some real-time or resource-limited applications. Likewise, BERT showed superior contextual understanding and classification performance but came with the highest computational expense of the compared methods. This consolidation of studies highlights a core trade-off: transformer-based models provide unprecedented depth of semantic nuance but at the cost of massive computational resources, whereas BiLSTM and classical ensemble techniques are more pragmatic alternatives in scenarios where speed, interpretability, and hardware constraints are top concerns.

VI. CONCLUSION AND FUTURE WORK

This study demonstrates how sentiment analysis enhances customer behavior and satisfaction understanding in the context of bilingual e-commerce product reviews. With state-of-the-art machine learning models like ELECTRA, BERT, and BiLSTM, it was shown how transformer-based architectures outperform other models in achieving high-performance metrics, where ELECTRA achieved the highest accuracy of 98.09% and F1 score of 0.9711. This study highlights the advanced methods in natural language processing for recognizing complex sentiment expressions, product-centric concerns, and critical intelligence outliers, which are essential for sound decision-making in e-commerce. In spite of these contributions, the study does reveal gaps. For example, the use of machine-translated reviews presents certain biases on culture and language that influence sentiment detection. In addition, the use of transformer models such as ELECTRA and BERT consume vast amounts of resources making their scalability and deployment in low-resource settings extremely problematic. This gap presents an opportunity to study other less resourceful methods for analyzing and preserving contextual accuracy and depth. There should be an expansion of the dataset to encapsulate a wider range of cultural and linguistic information for robust models to be built in the future. Moreover, there should be attempts to improve the computational effectiveness of the sentiment analysis frameworks to enable their use in dynamic e-commerce systems in real-time. This milestone has far reaching effects on the global competitive landscape and transforms borderless e-commerce by improving customer experience, business strategies, and innovation. Furthermore, this study has provided a vital stepping stone for the development of bilingual sentiment analysis systems.

The results of this paper provided promising groundwork for future efforts stemming from the sentiment analysis practice area of e-commerce. Future research for this study will have us apply our methodology to other languages in addition to Chinese and English. A promising direction is the use of multilingual transformer models, e.g., XLM-R, mBERT, and mT5, which can handle several languages without relying on translation. These models utilize cross-lingual embeddings, enabling sentiment analysis directly in a variety of languages without losing linguistic nuances. Including reviews in Spanish, French, Arabic and other common e-commerce languages will enable the analysis of multilingual sentiment trends. Furthermore, studying the effect of low-resource Languages on the e-commerce sentiment analysis would help in understanding the existing limitations and area of improvements needed in these models. The authors will pursue ongoing research that incorporates social media, purchase history, and even demographic data as these elements can greatly enhance a model's ability to understand and predict consumer actions. The ideas of few-shot learning and zero-shot learning are particularly exciting because they reduce the requirement for highly annotated data and enable comprehensible models that generalize across many languages and domains with minimal data. This is especially important for e-commerce and social media, where the streams of data are large and volatile and require responsive and adaptable methods. All of these approaches aim to extend the findings of this particular study to advance sentiment analysis and other applications in the field of e-commerce.

ACKNOWLEDGMENT

The authors extend their sincere appreciation to all persons and organizations whose contributions facilitated this work. They also like to express their gratitude to their colleagues and peers for their valuable comments throughout the completion of this work.

REFERENCES

[1] P. Vana and A. Lambrecht, "The effect of individual online reviews on purchase likelihood," *Marketing Sci.*, vol. 40, no. 4, pp. 708–730, May 2021, doi: 10.1287/MKSC.2020.1278.

[2] Q. Guishen, "Analysis of the impact of e-commerce anchors on user consumption behavior–based on the perspective of gatekeeper theory," *E-Commerce Lett.*, vol. 13, no. 2, pp. 3733–3738, 2024, doi: 10.12677/ecl.2024.132456.

[3] I. O. Adeniyi, A. Akinkunmi, N. A. Sande, A. A. Author, and I. Oluwasegun, "Social media sentiment analysis: A comprehensive analysis," 2024, doi: 10.13140/RG.2.2.31094.37441.

[4] C. Ziakis and M. Vlachopoulou, "Artificial intelligence in digital marketing: Insights from a comprehensive review," *Information*, vol. 14, no. 12, p. 664, Dec. 2023, doi: 10.3390/info14120664.

[5] K. N. Manasa and M. C. Padma, "A study on sentiment analysis on social media data," in *Emerging Research in Electronics, Computer Science and Technology* (Lecture Notes in Electrical Engineering), vol. 545, 2019, pp. 661–667, doi: 10.1007/978-981-13-5802-9_58.

[6] B. Hoar, R. Ramachandran, M. Levis, E. Sparck, K. Wu, and C. Liu, "Summative Student course review tool based on machine learning sentiment analysis to enhance life science feedback efficacy," 2023, *arXiv:2301.06173*.

[7] K. R. Jerripothula, A. Rai, K. Garg, and Y. S. Rautela, "Feature-level rating system using customer reviews and review votes," *IEEE Trans. Computat. Social Syst.*, vol. 7, no. 5, pp. 1210–1219, Oct. 2020.

[8] 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.

[9] N. Punetha and G. Jain, "Game theory and MCDM-based unsupervised sentiment analysis of restaurant reviews," *Int. J. Speech Technol.*, vol. 53, no. 17, pp. 20152–20173, Mar. 2023, doi: 10.1007/s10489-023-04471-1.

[10] N. M. Alharbi, N. S. Alghamdi, E. H. Alkhammash, 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, May 2021, doi: 10.1155/2021/5536560.

[11] 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 Psychol.*, vol. 13, pp. 661–667, Jun. 2022, doi: 10.3389/fpsyg.2022.865702.

[12] A. Rosário and R. Raimundo, "Consumer marketing strategy and e-commerce in the last decade: A literature review," *J. Theor. Appl. Electron. Commerce Res.*, vol. 16, no. 7, pp. 3003–3024, Nov. 2021, doi: 10.3390/JTAER16070164.

[13] S. R. Kumar, "Online reviews: Do consumers trust them," *SSRN Electron. J.*, 2001.

[14] 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, Jun. 2024, doi: 10.12677/ECL.2024.133740.

[15] E. Ok. (2024). *Attaining Branding Excellence in the Digital Era Through Strategic Search Engine Optimization*. [Online]. Available: https://www.researchgate.net/publication/385597363

[16] R. Liu and J. Xiao, "Factors affecting users' satisfaction with urban parks through online comments data: Evidence from shenzhen, China," *Int. J. Environ. Res. Public Health*, vol. 18, no. 1, p. 253, Dec. 2020, doi: 10.3390/ijerph18010253.

[17] X. Tao and W. Hu, "The influence of characteristics of e-commerce anchors information source on college students' purchase intention," *Int. J. Bus. Manage.*, vol. 19, no. 5, p. 47, Aug. 2024, doi: 10.5539/ijbm.v19n5p47.

[18] 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, Apr. 2022, doi: 10.19722/J.CNKI.1008-7729.2021.0254.

[19] 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, 2024, doi: 10.54691/6ez9h406.

[20] 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, May 2023, doi: 10.3390/bs13060439.

[21] X. Tong, D. Tao, and R. Lifset, "Varieties of business models for post-consumer recycling in China," *J. Cleaner Prod.*, vol. 170, pp. 665–673, Jan. 2018, doi: 10.1016/j.jclepro.2017.09.032.

[22] T.-A. Pham and M. Yoo, "Nighttime vehicle detection and tracking with occlusion handling by pairing headlights and taillights," *Appl. Sci.*, vol. 10, no. 11, p. 3986, Jun. 2020, doi: 10.3390/app10113986.

[23] I. Ventre and D. Kolbe, "The impact of perceived usefulness of online reviews, trust and perceived risk on online purchase intention in emerging markets: A Mexican perspective," *J. Int. Consum. Marketing*, vol. 32, no. 4, pp. 287–299, Aug. 2020, doi: 10.1080/08961530.2020.1712293.

[24] A. Samanta and T. Guha, "On the role of head motion in affective expression," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Mar. 2017, pp. 2886–2890.

[25] H. S. Choi and S. Leon, "An empirical investigation of online review helpfulness: A big data perspective," *Decis. Support Syst.*, vol. 139, Dec. 2020, Art. no. 113403, doi: 10.1016/j.dss.2020.113403.

[26] J. W. Kim, A. Guess, B. Nyhan, and J. Reifler, "The distorting prism of social media: How self-selection and exposure to incivility fuel online comment toxicity," *J. Commun.*, vol. 71, no. 6, pp. 922–946, Dec. 2021, doi: 10.1093/joc/jqab034.

[27] Y. Guopeng, "What online reviews do consumers find more useful?–Effect of Social Factors," *Manage World*, no. 12, pp. 115–124, 2012.

[28] C. Grange and I. Benbasat, "Opinion seeking in a social network-enabled product review website: A study of word-of-mouth in the era of digital social networks," *Eur. J. Inf. Syst.*, vol. 27, no. 6, pp. 629–653, Nov. 2018, doi: 10.1080/0960085x.2018.1472196.

[29] S. Kim, S. Ham, H. Moon, B.-L. Chua, and H. Han, "Experience, brand prestige, perceived value (functional, hedonic, social, and financial), and loyalty among GROCERANT customers," *Int. J. Hospitality Manage.*, vol. 77, pp. 169–177, Jan. 2019, doi: 10.1016/j.ijhm.2018.06.026.

[30] Q. Zhang and M. Abisado, "A novel context-aware deep learning algorithm for enhanced movie recommendation systems," *Math. Model. Eng. Problems*, vol. 10, no. 6, pp. 2031–2038, Dec. 2023, doi: 10.18280/mmep.100613.

[31] B. Bickart and R. M. Schindler, "Internet forums as influential sources of consumer information," *J. Interact. Marketing*, vol. 15, no. 3, pp. 31–40, Aug. 2001, doi: 10.1002/DIR.1014.

[32] J. Lee, S. Rajtmajer, E. Srivatsavaya, and S. Wilson, "Online self-disclosure, social support, and user engagement during the COVID-19 pandemic," *ACM Trans. Social Comput.*, vol. 6, nos. 3–4, pp. 1–31, Dec. 2023, doi: 10.1145/3617654.

[33] M. Jasin, "The role of social media marketing and electronic word of mouth on brand image and purchase intention of SMEs product," *J. Inf. Syst. Manage.*, vol. 1, no. 4, pp. 54–62, Aug. 2022, doi: 10.4444/JISMA.V1I4.258.

[34] G. Jiang, F. Liu, W. Liu, S. Liu, Y. Chen, and D. Xu, "Effects of information quality on information adoption on social media review platforms: Moderating role of perceived risk," *Data Sci. Manage.*, vol. 1, no. 1, pp. 13–22, Mar. 2021, doi: 10.1016/j.dsm.2021.02.004.

[35] J. Chen, G. Kou, Y. Peng, X. Chao, F. Xiao, and F. E. Alsaadi, "Effect of marketing messages and consumer engagement on economic performance: Evidence from Weibo," *Internet Res.*, vol. 30, no. 5, pp. 1565–1581, Jun. 2020, doi: 10.1108/intr-07-2019-0296.

[36] A. Harmadi, A. Kuswanto, and G. Nuryanto, "THE effect of online service quality and offline service quality on trust and satisfaction and its impact on loyalty: An empirical study on online and offline transportation customers," *Int. J. Res. Publication*, vol. 142, no. 1, pp. 1–12, Jan. 2024, doi: 10.47119/ijrp1001421220246040.

[37] V. Gooljar, T. Issa, S. Hardin-Ramanan, and B. Abu-Salih, "Sentiment-based predictive models for online purchases in the era of marketing 5.0: A systematic review," *J. Big Data*, vol. 11, no. 1, p. 107, Aug. 2024, doi: 10.1186/s40537-024-00947-0.

[38] R. R. Tobias, R. E. Roxas, and M. Abisado, "Science mapping of social media analytics in health through artificial intelligence," in *Proc. IEEE Region Conf. (TENCON)*, Dec. 2021, pp. 750–755, doi: 10.1109/TENCON54134.2021.9707362.

[39] 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.

[40] H. Fu, G. Manogaran, K. Wu, M. Cao, S. Jiang, and A. Yang, "Intelligent decision-making of online shopping behavior based on Internet of Things," *Int. J. Inf. Manage.*, vol. 50, pp. 515–525, Feb. 2020, doi: 10.1016/j.ijinfomgt.2019.03.010.

[41] K. L. Tan, C. P. Lee, and K. M. Lim, "A survey of sentiment analysis: Approaches, datasets, and future research," *Appl. Sci.*, vol. 13, no. 7, p. 4550, Apr. 2023, doi: 10.3390/app13074550.

[42] 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, Dec. 2022, doi: 10.18535/ijsshi/v9i012.06.

[43] Q. 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.

[44] *Dirtycomputer/online_shopping_10_cats ? Datasets at Hugging Face*. Accessed: Jan. 20, 2025. [Online]. Available: https://huggingface.co/datasets/dirtycomputer/online_shopping_10_cats

[45] Y. Kong, Z. Xu, and M. Mei, "Cross-domain sentiment analysis based on feature projection and multi-source attention in IoT," *Sensors*, vol. 23, no. 16, p. 7282, Aug. 2023, doi: 10.3390/s23167282.

[46] *Hellomadisonshan/ProductsSocialMedia: This Dataset is Social Media Posts and Reviews Scraped in Chinese Language*. Accessed: Jan. 22, 2025. [Online]. Available: https://github.com/hellomadisonshan/ProductsSocialMedia

[47] Y. Hu, H. Wang, T. Ji, X. Xiao, X. Luo, P. Gao, and Y. Guo, "CHAMP: Characterizing undesired app behaviors from user comments based on market policies," in *Proc. IEEE/ACM 43rd Int. Conf. Softw. Eng. (ICSE)*, May 2021, pp. 933–945, doi: 10.1109/ICSE43902.2021.00089.

SHIZHONG SHAN (Member, IEEE) was born in Fuzhou, Fujian, China. He received the degree in business administration from José Rizal University, Philippines. He is currently the Executive Dean of China National Democratic Construction Association, Economic Research Institute, Yango University. Previously, he was the General Manager at Fujian Ruisheng Investment Company Ltd., leading key initiatives that drove organizational growth. He brings over a decade of experience in academic leadership, strategic policy development, and research. He is an accomplished academic and policy leader. His work focuses on digital transformation, sustainable development, and talent retention, with many of his recommendations adopted by government bodies. He has been recognized with numerous awards for excellence in social research and public opinion analysis. He remains dedicated to fostering critical thinking, innovation, and collaboration, aiming to make lasting contributions to academia.

JIN SUN received the bachelor's degree in electrical engineering and automation from Huaihai Institute of Technology and the master's degree in management from Asia Pacific College. He is currently pursuing the Ph.D. degree in business administration with National University. He is also the Director at Fuzhou Zhongmin Huixin Venture Capital Partnership, where he oversees initiatives in international education and research collaboration. He has been instrumental in establishing partnerships between Chinese and Philippine universities, including National University and Asia Pacific College, fostering collaborative research efforts aimed at publication in leading journals, such as SCI and SSCI. His professional background includes roles as a Service Engineer at Intervision Vision (Shanghai) Company Ltd. and the Investment Manager at Fujian Ruisheng Investment Company Ltd., reflecting a career that bridges technical expertise and strategic management. With a focus on innovation, education, and research, he continues to contribute to strengthening international academic and professional collaborations.

REMSON MARK C. MACAWILE is currently pursuing the Ph.D. degree in engineering management with the Polytechnic University of the Philippines (PUP). He is also an Assistant Vice President for Transaction Banking at one of the largest banks in the Philippines, where he leads product innovation and development, leveraging over a decade of experience in product management, operations, and financial services. He is an accomplished academic and banking executive, serving as an Adjunct Faculty Member at National University and Asia Pacific College. Since 2014, he has taught undergraduate and graduate courses in engineering and business, mentored research projects, and contributed to Asian Institute of Management (AIM's) curriculum development by integrating industry practices into academic frameworks. His executive roles complement his academic endeavors, fostering a practical and interdisciplinary approach to education and research. With a strong academic foundation and extensive professional expertise, he bridges the gap between theory and practice, advancing both scholarship and industry development. His research interests include Fintech innovation, MSME capital formation, and cash management.