



Decoding E-Sourcing Narratives on Reddit: A Comprehensive Text Mining Approach

Yuhui (Jessie) Liu, Ph.D. Candidate, University of Missouri

Dr. Li Zhao, Associate Professor, University of Missouri

Keywords: e-sourcing, online sourcing, text mining, Reddit, N-gram, sentiment, LDA

E-sourcing, also known as online sourcing, employs digital-enabled technology to streamline procurement processes (Chartered Institute of Procurement and Supply, 2019). Serving as online intermediaries, they create opportunities for buyers and sellers to discover new products/services and explore market possibilities through information sharing (Chang and Wong, 2010). In 2022, the global B2B eCommerce market, valued at US\$20.4 trillion, surpasses the B2C market by more than fivefold, with the Asia Pacific region commanding a significant market share of nearly 78%. Leading players in the global B2B eCommerce market include Amazon Business, Alibaba, Rakuten, Unite (formerly Mercateo), Global Sources, Walmart, and IndiaMART (Statista, n.d.). However, there is limited research delving into the needs and experiences of the e-sourcing community. This study aims to (1) identify the e-sourcing needs of SMEs through comments related to fashion online sourcing, and (2) explore SMEs' e-sourcing experiences.

This study guided by the conceptual framework combined the triangular alignment model (TAM) (Schwarz et al, 2009) and technology–organization–environment (TOE) theories (Tornatzky & Fleischer, 1990), encompassing five dimensions of e-sourcing experience (Li, Zhao & Lee, 2021): efficiency, competitiveness, IT capacity, friendliness, and reliability. This comprehensive framework aids in understanding SMEs' motivations for e-sourcing platform adoption, contributing to improved performance. Additionally, leveraging unsupervised machine learning aims to not only validate but also refine and extend the framework.

Table 1 High-frequency N-grams in each dimension

Dimension	Top words and the most frequently associated words
Efficiency	Make: term, compare, make, home Like: learn, always, probably, similar, still get: something, high, right, sewing, good time: question, worth, term, catch, something
Competitiveness	Product/people: expensive, price, depend, compare, cheap Good/Get: something, high, right, sewing, good Business/Brand: probably, absolutely, hobby, buy, board company: price, compare, catch, cheap, trench coat want: sewer, worth, always, sewing, would
Reliability	find: fabric; definitely, different, frustrate thing: question, worth, term, catch, something fashion: find, fabric; definitely, different, frustrate also: hard, learn, always, compare, make
Friendliness	lot: learn, always, probably, similar, still take: question, worth, term, catch, something

To achieve the research objective, we employed a comprehensive text-mining approach. Initially, we decided to choose Reddit as our primary data source because it has an active and relatively independent community outside of specific e-commerce platforms. By using keywords such as "fashion online

sourcing," "Apparel online sourcing," and "Clothing online sourcing," we extracted relevant comments from over 300 subreddits, after data cleaning, resulting in a dataset comprising 576 posts containing approximately 170k words. Examples of these subreddits include "flipping," "entrepreneur," "female fashion advice," "small businesses," and "dropshipping," among others.

Our data mining process comprised three main components: N-gram analysis (Srikanth & Srihari, 2002) to identify key terms within our theoretical framework, sentiment analysis (Nahar et al, 2019) to assess the sentiments of the sourcing community, and LDA topic modeling (Blei et al, 2003) to explore the predominant topics within the e-sourcing community.

As depicted in Table 1, based on the N-gram analysis results, we categorized the most frequently occurring words and frequently associated words under the themes outlined in our theoretical framework. For instance, under the theme of "efficiency," the word "time" was identified and frequently associated with terms were found, including "questions," "worth," "term," and "catch," Similarly, in the context of "competitiveness," words like "product" and "people" were commonly linked with "expensive," "price," "depend," "compare," and "cheap," whereas terms like "business," "brand," Within the theme of "reliability," However, for "friendliness" and "IT capacity," we did not identify significant findings among the top frequent words. These findings suggest a pattern highlighting competitiveness and efficiency by seeking relevant information.

Table 2 *Sentiment Analysis*

Sentiment	Percentage
Positive	32%
Trust	19%
Anticipation	15%
Negative	14%
Joy	10%
Fear	7%
Sadness	6%
Anger	6%
Surprise	6%
Disgust	4%

As shown in Table 2, based on sentiment analysis using the method developed by Nahar et al. (2019), the proportion of matched words in each sentiment category suggests a prevalence of positive sentiments within e-sourcing communities compared to negative ones. Trust, anticipation, and joy were commonly expressed sentiments. For instance, comments such as "I'm looking for an app/platform so that people can order customs easily. Suggestions?", "Help Identifying fabric...", reflecting positive sentiments and anticipation within the community. However, negative sentiments including fear, sadness, anger, and disgust were also present.

Examples include comments such as "P(online payment company) is disgusting... I rather pay a larger fee and go elsewhere..." and "This makes me so sad and so angry at the same time...S(company name) is way worse...". The sentiment analysis effectively helps us understand the needs of the e-sourcing community.

Table 3 *Topics of the fashion e-sourcing comments extracted from Reddit online communities*

Topic & Percentage	Key words	Description
Product source oriented 52.5%	Source/like/sell/buy/brand/time	Online sourcing, brand-supplier dynamics, and staying updated on fashion industry trends.
Human source oriented 21.9%	Like/people/think/try/work/know	Explore quality assurance and risk mitigation in e-sourcing, uncovering strategies for maintaining high standards.
Ethic sourcing Concern 7.9%	Say/region/cotton/include/Xinjiang	Ethical sourcing discussion, unveiling information on associated fashion brands.
Secondhand fashion 5%	Thrift/store/casual/size/people	Exploring flipping, thrift, and second-hand markets, emphasizing diverse sourcing methods.
Domestic sourcing 3.9%	Feb/kohl/mode/China/fox/hat fan	Leveraging information gaps across platforms for profit in domestic sourcing, with a focus on brand concerns
Technology concern 3.8%	Year/company/growth/share/technology	Sharing novel methods and sourcing-related technologies, with a focus on competitive considerations.

Lastly, as shown in Table 3, the LDA topic modeling provided us with an insight into the online sourcing community on the aspects of the topic proportions. More than half (52.5%) of the topics

were assigned to product source orientation, focusing on sourcing efficiency and competitiveness. Examples include discussions on seeking products or agencies, such as "Look online for actual makers...their process in a way showing transparency with their sourcing and production methodology," Human source orientation accounted for 21.9% of the categories and also addressed efficiency and competitiveness concerns. Additionally, our data revealed topics related to ethical concerns, domestic sourcing, and technology, indicating the prevalent interests within the community.

This study holds significance by shedding light on the e-sourcing online community's needs and experience, thereby contributing to the broader discourse on fashion supply chain dynamics. These findings carry implications for future research, indicating a need to test efficiency, competitiveness, and reliability within e-sourcing practices. Further quantitative studies on these aspects can enhance our understanding of the intricacies involved and pave the way for more informed decision-making processes within the fashion e-commerce supply chain context.

Reference

- Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet allocation. *Journal of Machine Learning Research*, 3(Jan), 993-1022.
- Chartered Institute of Procurement and Supply (2019), "What is procurement?", available at: <https://cips.org/en/cips-for-individuals/what-is-procurement/> (accessed 1 March 2024).
- Chang, H.H. and Wong, K.H. (2010), "Adoption of e-procurement and participation of e-marketplace on firm performance: trust as a moderator", *Information and Management*, Vol. 47 No. 5, pp. 262-270.
- Li, X., Zhao, L., & Lee, S. H. (2021). Measuring user experiences with e-sourcing platforms: the development of the e-sourcing experience index. *Journal of Fashion Marketing and Management: An International Journal*, 25(3), 430-447.
- Nahar, L., Sultana, Z., Iqbal, N., & Chowdhury, A. (2019, May). Sentiment analysis and emotion extraction: A review of research paradigm. In *2019 1st International Conference on Advances in Science, Engineering and Robotics Technology (ICASERT)* (pp. 1-8). IEEE.
- Schwarz, A., Jayatilaka, B., Hirschheim, R. and Goles, T.A. (2009), "Conjoint approach to understanding IT application services outsourcing", *Journal of the Association for Information Systems*, Vol. 10 No. 10, pp. 748-781.
- Srikanth, M., & Srihari, R. (2002, August). Biterm language models for document retrieval. In *Proceedings of the 25th annual international ACM SIGIR conference on Research and development in information retrieval* (pp. 425-426).
- Statista. (n.d.). B2B eCommerce - In-depth Market Insights & Data Analysis. Retrieved from <https://www.statista.com/study/44442/in-depth-report-b2b-e-commerce/>
- Tornatzky, L.G. and Fleischer, G. (1990), *The Process of Technological Innovation*, Lexington books, Lexington, MA.