IJITCE

ISSN 2347- 3657

International Journal of Information Technology & Computer Engineering

Email : ijitce.editor@gmail.com or editor@ijitce.com

International Journal of Information Technology & Computer Engineering

ISSN 2347–3657

Volume 10, Issue 2, 2022

https://doi.org/10.62646/ijitce.2022.v10.i2.pp110-123

Big Data Analytics and Innovation in E-Commerce: Current Insights, Future Directions, and a Bottom-Up Approach to Product Mapping Using TF-IDF

Sharadha Kodadi

TIMESQUAREITINC,

ALPHARITA,

GEORGIA, USA

kodadisharadha1985@gmail.com

ABSTRACT

Term Frequency-Inverse Document Frequency (TF-IDF) is a bottom-up technique to product mapping that is used in this study to investigate the application of big data analytics and novel approaches in e-commerce. Little and medium-sized businesses (SMEs) are still understudied, despite the focus of traditional study on major manufacturers and retailers. A thorough product map study is required to comprehend product linkages, complementarity, and competitive dynamics, as SMEs have become more prominent, especially since the COVID-19 epidemic. The creation of comprehensive product maps is suggested by this paper by using crowd intelligence from SME e-commerce sites. The study employs TF-IDF to measure word significance in product titles and descriptions by gathering data from more than 52 SME sites, which contains information on over 90,000 goods. By building a product map with cosine similarity metrics, hierarchical community structures are made visible. Results show that items on the same website often create different communities, exposing competitive dynamics and supporting small and medium-sized enterprises (SMEs) in making strategic decisions about pricing, product offerings, and marketing strategies. To improve text discrimination accuracy, the study emphasizes the necessity for sophisticated natural language processing methods like N-grams. While highlighting the decentralized and varied character of SME e-commerce data, it also highlights the study gap in this area. In addition to providing SMEs with insights to enhance their market positioning, resource allocation, and customer engagement strategies, the study fills in these gaps in the literature on e-commerce. To improve SME resilience and predictive capacities in the dynamic digital economy, future research should encompass real-time analytics, broaden the scope of data sources, and investigate the approach's application across many industries.

Keywords: Big data analytics, e-commerce, small and medium-sized enterprises (SMEs), TF-IDF, product mapping, competitive dynamics.

1. INTRODUCTION

The shift in the way businesses operate has brought new opportunities and challenges for small and medium-sized businesses (SMEs), all of which originate from e-commerce. The growing digital marketplace helps make these interconnections even more crucial for organizations on the

ISSN 2347–3657

Volume 10, Issue 2, 2022

https://doi.org/10.62646/ijitce.2022.v10.i2.pp110-123

Mapping Helps Provide an Overall Picture of Your Digital Commerce Ecosystem. In this study, we examine the impact of big data analytics and new methodologies on e-commerce. In particular the paper shed light to bottom-up approach for product mapping using Term Frequency-Inverse Document Frequency. We will be using datasets of small to medium e-commerce websites in order to make recommendations on product similarities and competitive dynamics Pricing, product offering and marketing tactics can be adjusted based on these insights.

Large merchants and manufacturers have traditionally controlled the study of product maps around e-commerce. They have studied product linkages as well origenic and substitutive aspects in a big data capacity. The first efforts, such as those by Elrod and others, concentrated on how customers responded to competitive goods. These studies also exposed the restrictions of fixed product categories. Netzer et al used text mining techniques to do this. to create user curated product maps. The emergence of machine learning and data quality enabled these studies, allowing analysis to evolve exponentially.

However, most of these studies have slanted the perception and measurements in large scales retailers with manufacturers overlooking small to medium enterprises (SMEs). Jakarta, Entrepreneurs and SMEs -- Largeraminos: At the same time as online trading surged during pandemic COVID-19 effect. Businesses are being supported by smaller enterprises in those places. Nonetheless, academic studies are still in the process of fully exploring their data. The opportunity is that of a bottom-up method via crowd intelligence from SMEs to build comprehensive product maps.

Natural Language Processing via Big Data Analytic techniques (TF-IDF, etc.) revolutionized how product associations are identified for e-commerce. The TF-IDF is a very well-known factorization of the term into Term (Word) and Document Frequency; which achieves semantically explaining how much important words are in sense to collection occurring documents. This method examines product titles and descriptions across many SME e-commerce websites in order to discover unique attributes of a given type of product. Using the TF-IDF graph, we can measure product similarities and create a detailed map of products-we observe several hierarchical community structures in it.

In addition to this advancements in machine learning allows the bulk analysis of huge datasets, hence making it possible for SMEs can further process and interpret diverse data. N-grams and other techniques can be assigned with keyword extraction in context, hence for text discrimination it will lead to improvement by increasing relevancy of keywords. These technological advancements provide a robust backdrop for studying the complex product mechanics which are characteristics of small and medium sized enterprises (SME) e-commerce.

Current researches focus on large retailers and manufacturers, which provides a significant challenge. Is that so, when we know for a fact the definite participation of small and medium sized enterprises (SMEs) play significant roles within an e-commerce landscape. Existing methods generally neglect the decentralised and heterogeneous nature of small and medium enterprise data, which results in insufficient understanding about product links. This restriction is largely due to the fact that there does not exist any unified product maps, encompassing data from myriad small and medium-sized enterprise websites associated with pricing as well as p-offers or omni% of a website.

- Investigate data underutilization of (small and medium sized enterprises) SMEs for product map analysis
- Deploy a ground-up approach to SME e-commerce insights across different trading models.
- To perform precise product linkage, you would require TF-IDF & sophisticated NLP algorithms
- Enhance the strategic decision-making of SMEs (small and medium-sized enterprises) in terms of pricing, available products to offer or even the way campaigns are conducted.
- Provide a unique entry to literature for those interested in e-commerce research.

Understandably, previous studies lack data of small and medium-sized enterprises (SME) on product map analysis. Research on large merchants has received substantial attention, while the decentralized and autonomous nature of small and medium-sized enterprises (SMEs) in past research is marginalized. Due to this gap, a bottom-up method that consolidates various views from small- and medium-sized enterprise (SME) e-commerce websites is necessitated. This does mean that the methods in product mapping may still not take full advantage of all these rather sophisticated NLP techniques such as TF-IDF.

Aims of the Study to develop a novel method for analyzing product maps from an SME perspective by remedying these limitations and contributing to e-commerce research. Employing TF-IDF and other advanced analytics techniques will enable a more accurate understanding of how products inter-link with each other; thereby assisting Small Medium Enterprises (SMEs) in their strategic decision making.

2. LITERATURE SURVEY

Tariq et al. (2021) provide an architecture that makes use of Dailymotion data to analyse user behaviour on social media sites through big data analytics. The three-tier architecture makes use of Hive storage for effective data management and Apache Spark for parallel processing. Data from public channels and the Dailymotion API show how effective this approach is. By utilising intelligent big data analysis and artificial intelligence, this method tackles the difficulty of handling enormous and unstructured social media datasets, offering notable advantages in managing and examining large-scale social media data.

Lugier et al. (2021) describe a novel bottom-up approach for printing patterns on self-assembled monolayers (SAMs) utilising extreme ultraviolet (EUV) lithography. This technology takes

ISSN 2347–3657
Volume 10, Issue 2, 2022
https://doi.org/10.62646/ijitce.2022.v10.i2.pp110-123

advantage of the chemical alterations caused by EUV exposure to enable the selective creation of hybrid structures. The technology has major advantages, including the ability to tune distinct stages of the process independently for more precise control over nanopattern development, making it a promising strategy for producing smaller nanostructures.

Naresh Kumar Reddy Panga (2021) introduced an optimized hybrid machine learning architecture for improving financial fraud detection on e-commerce platforms, addressing the growing threat of sophisticated fraudulent activity. The system increases detection accuracy and lowers false positives by combining multiple approaches such as neural networks, decision trees, and support vector machines. The system adapts to new fraud behaviors through continuous monitoring and hyperparameter adjustment, enhancing financial security and showcasing hybrid models' potential for constructing reliable fraud detection systems.

Golov and Carrasco (2021) describe a novel bottom-up approach for printing patterns on self-assembled monolayers (SAMs) utilising extreme ultraviolet (EUV) lithography. This technology takes advantage of the chemical alterations caused by EUV exposure to enable the selective creation of hybrid structures. The technology has major advantages, including the ability to tune distinct stages of the process independently for more precise control over nanopattern development, making it a promising strategy for producing smaller nanostructures.

Motoki et al. (2021) examine how eye tracking is used in sensory and consumer research, with a particular emphasis on how it might be used to better understand how consumers receive visual information. The study summarises research on top-down and bottom-up processing, looking at elements like size, goals, complexity, emotional reactions, visual salience, and task instructions. It highlights possible problems in eye-tracking studies in this setting and talks about how gaze patterns affect customer choices. The research also suggests future paths for improving eye-tracking techniques to improve comprehension of customer behaviour in sensory situations.

In their discussion of the worldwide issues related to oil consumption and spills, Gaur et al. (2021) highlight the harmful effects of petroleum on the ecosystem. The study emphasises microbial activities as a viable and affordable method of reducing oil pollution. It talks about how well biological techniques, like using biosurfactants and biochar, work when combined with microbial action to improve cleanup results. Furthermore, the incorporation of "-omics" technology has been important in propelling hydrocarbon degradation research forward, filling gaps, and refining petroleum bioremediation techniques.

Jana and Uma (2020) address the issues brought about by the profusion of brief, feature-specific reviews on online platforms by talking about the importance of aspect extraction and sentiment analysis in e-commerce product reviews. In addition to highlighting the value of aspect-wise sentiment analysis for improving consumer understanding and assisting manufacturers in improving their products, the study suggests a method for locating aspects and opinions inside reviews. Their study examines several machine learning methods for sentiment classification and

ISSN 2347–3657

Volume 10, Issue 2, 2022

https://doi.org/10.62646/ijitce.2022.v10.i2.pp110-123

comes to the conclusion that the best method for reliably assessing sentiment in product reviews is logistic regression with L1 regularization.

Ma et al. (2019) present a machine learning system designed to recognise commodity names in international e-commerce, tackling the issues of precision and effectiveness during peak periods such as "Double Eleven." With 91% accuracy and 93% recall rate in trials, the paper's use of SVM and TF-IDF models yields outstanding results. When compared to manual procedures used in the past, it shows a notable 20% increase in efficiency for commodity export declaration. The usefulness and efficiency of machine learning in improving operational procedures in e-commerce logistics is highlighted by this study.

Belem et al. (2020) apply automatic tag recommendation algorithms to improve the quality of product tags in e-commerce platforms. The research emphasises the need of high-quality textual descriptors like tags in search and product recommendation systems. It addresses issues raised by vendors' tags, such as poor training or purposeful manipulation. The proposed strategies use product search data to suggest more diversified and relevant tags, resulting in a 16% increase in recommendation efficacy over existing methods. The goal of this study is to improve consumers' overall search experience on e-commerce websites by increasing the accuracy and relevancy of product tags.

Utomo et al. (2019) assessed sentiment towards the e-commerce platform by analysing online reviews from Bukalapak users using the TF-IDF method. According to their findings, the majority of the evaluations were positive, indicating that customers had a favourable image of the platform. This sentiment analysis technology helps organisations like Bukalapak quickly discover and address product faults based on user input.

Meertens et al. (2020) provide a data-driven supply-side model to forecast cross-border Internet purchases within the European Union. The model offers insights into online consumer behaviour and its consequences on the European economy by employing trade and e-commerce data. This strategy seeks to support companies and legislators in tackling the expanding market for cross-border e-commerce.

Zhang and Villarroel Ordenes (2019) provide a thorough analysis of the use of text and image mining methods in service research. The authors emphasise the importance of integrating multiple data types for a more comprehensive understanding of the service industry. They highlight how these methodologies analyse unstructured data from textual and visual sources to get deeper insights into customer behaviour and service experiences.

Subhashini et al. (2021) discuss approaches to customer review mining and classification in their paper, with a focus on machine learning, sentiment analysis, and opinion mining. The survey looks at methods and difficulties in deriving insights from consumer feedback, giving a useful summary of current developments and difficulties in review mining and classification in artificial intelligence.

According to Rajeswaran Ayyadurai (2021), big data analytics is critical for minimizing manufacturer invasion and channel conflict in e-commerce supply chains, especially in dual-channel setups where manufacturers sell directly online and through traditional retailers. While this strategy can strain relationships with merchants, e-commerce systems help to alleviate tensions by leveraging big data to improve inventory, forecast market trends, and customize customer experiences. By exchanging demand data with producers, these platforms boost supply chain efficiency. The study uses big data analytics, game theory, and supply chain management to investigate how manufacturers' strategies and risk preferences influence cooperation and conflict.

3. METHODOLOGY

A bottom-up strategy and Term Frequency-Inverse Document Frequency (TF-IDF) are utilized in this study's methodology, which consists of many important processes for the purpose of analyzing product maps from small and medium-sized e-commerce (SME) websites. This part provides specifics of the procedure for collecting data, the implementation of TF-IDF, the development of product maps, and the assessment of the insights that were obtained as a result. In addition to that, it contains pertinent equations, tables, and a sketch picture that illustrates the procedure.

3.1. Data Collection

The first step in this approach is collecting data from hundreds of small and medium-sized enterprise (SME) e-commerce websites. To get this information we need web scraping and API integration to collect the product titles, descriptions...everything as part of it. It is crucial to capture only valid data in the flow because it helps ensure that all possible category combinations where at least one popular product has been sold on various e-commerce platforms are included state of concern.

1. The collection of data is aimed at more than 52 small and medium-sized enterprise (SME) e-commerce websites. These sites span a wide variety of product categories, which ensures that the dataset has a diverse range of information.

2. Data Attributes: The following attributes are required to be collected for each and every product:

- ✓ Product Title
- ✓ Product Description
- ✓ Price
- ✓ Category
- ✓ SKU (Stock Keeping Unit)
- ✓ Site URL

International Journal of Information Technology & Computer Engineering
ISSN 2347–3657
https://doi.org/10.62646/ijitce.2022.v10.i2.pp110-123
Volume 10, Issue 2, 2022

3. Volume of Data: The dataset contains more than 90,000 goods, which makes it an excellent foundation for analysis.

Table 1: Samples of Data Characteristics.

Product ID	**Title**	**Description**	**Price**	**Category**	SKU
001	Blue Jeans	Comfortable blue jeans for everyday wear.	$40	Clothing	J001
002	Running Shoes	Lightweight running shoes for men.	$60	Footwear	S002
003	Wristwatch	Stylish wristwatch with leather strap.	$120	Accessories	W003

3.2. The Preprocessing of Data

The collected data is pre-processed to ensure the quality and consistency of it before the TF-IDF algorithm goes through. This process consists of the following procedures.

The titles and descriptions of products are cleaned from stop words, special characters as well as HTML tags. Tokenization: This is the splitting up of product titles, and likewise descriptions into individual elements known as tokens. Stemming and Lemmatization is basically reducing a word to its basic form, this helps in ensuring that different forms of words count as equal. Normalization is the act of converting all text to lowercase, in order remove case-sensitive problems.

3.3. Application of TF-IDF

In order to determine the significance of the terms that are included in product titles and descriptions, our methodology is built around the application of TF-IDF. Following is an explanation of the TF-IDF process:

Term Frequency (TF)

Phrase Frequency (TF) is a metric that determines how frequently a particular phrase appears in a given document. Information retrieval and text mining both rely heavily on this idea as a fundamental principle. The term frequency $tf\ (t, d)$ of term t in document d is calculated using the following equation:

$$tf(t, d) = \frac{f(t,d)}{\sum_{t' \in d} f(t',d)} \qquad (1)$$

where:

- $f(t, d)$ is the frequency of term t in document d.

ISSN 2347–3657
Volume 10, Issue 2, 2022
https://doi.org/10.62646/ijitce.2022.v10.i2.pp110-123

- $\sum_{t' \in d} f(t', d)$ is the total number of terms in the document d.

Example: If the term "jeans" appears 5 times in a document with a total of 100 terms, the term frequency is:

$$tf(\text{"jeans"}, d) = \frac{5}{100} = 0.05 \tag{2}$$

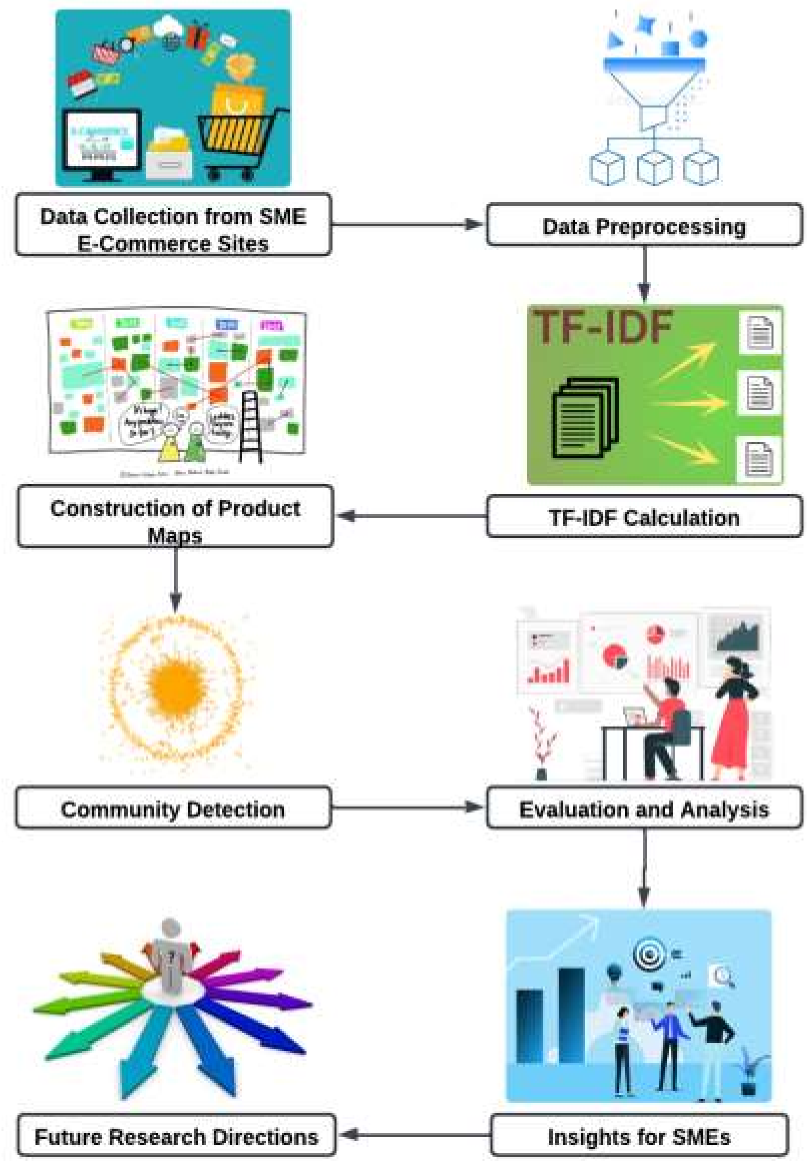

Figure 1: Big Data Analytics and Innovation in E-Commerce using TF-IDF.

International Journal of Information Technology & Computer Engineering
ISSN 2347–3657
Volume 10, Issue 2, 2022
https://doi.org/10.62646/ijitce.2022.v10.i2.pp110-123

Fig. 1 uses TF-IDF to show the study's methodology and key findings on big data analytics in e-commerce. The procedure begins with gathering data from small-business e-commerce websites, then moves on to data preparation and TF-IDF computation. Making product maps and finding communities within them are the next steps. The last stages involve analysis and assessment, providing SMEs with insights, and recommending future lines of inquiry for study. This method aids SMEs in improving pricing, improving marketing tactics, and comprehending the dynamics of competition.

Inverse Document Frequency (IDF)

The term's significance throughout the full corpus can be evaluated using a technique known as Inverse Document Frequency (IDF). A reduction in the weight of common terms and an increase in the weight of uncommon terms are both benefits of this. The IDF of term t in a corpus of documents D is calculated using the following equation:

$$idf\ (t, D) = ln\ \left(\frac{N}{|\{d \in D: t \in d\}|}\right) \tag{3}$$

where:

- N is the total number of documents in the corpus.
- $|\{d \in D: t \in d\}|$ is the number of documents in which term t appears.

Example: If the term "jeans" appears in 101 out of 1000 documents, the inverse document frequency is:

$$idf\ (\ "jeans",\ D) = ln\ \left(\frac{1000}{101}\right) = ln\ (9.901) \approx 2.293 \tag{4}$$

TF-IDF Weight

The relationship between a term's term frequency and its inverse document frequency is what determines the TF-IDF weight of that term. Using this weight, one can determine the significance of a particular term within a particular text in comparison to the entire corpus. The TF-IDF weight tfidf (t, d, D) is calculated using the following equation:

$$tfidf\ (t, d, D) = tf\ (t, d) \cdot idf\ (t, D) \tag{5}$$

Example: Using the previous TF and IDF values for the term "jeans":

$$tfidf(jeans, d, D) = 0.05 \cdot 2.293 = 0.11465 \tag{6}$$

Vector Representation of Documents

The TF-IDF weights are represented as a vector for each document, which in this case refers to each individual product. Through the use of this vectorisation, it is possible to compare documents according to the significance of phrases.

Example: Take into consideration three documents (items) that have TF-IDF weights for the phrases "jeans," "running," and "wristwatch" for example:

Table 2: Sample Weights for the TF-IDF.

Term	**TF (D1)**	**TF (D2)**	**TF (D3)**	**IDF**	**TF-IDF (D1)**	**TF-IDF (D2)**	**TF-IDF (D3)**
Jeans	0.05	0	0	2.293	0.11465	0	0
Running	0	0.1	0	2.293	0	0.2293	0
Wristwatch	0	0	0.05	2.293	0	0	0.11465

3.4.Construction of Product Maps

Constructing product maps by making use of the TF-IDF vectors is the next stage. In order to accomplish this, you will need to create a graph in which nodes represent products and edges represent the degree of similarity between products. The following are components of the process:

Cosine Similarity: Cosine similarity is the method that we utilize to determine the degree of similarity between products. Calculating the cosine of the angle that exists between two TF-IDF vectors is the purpose of this measure. The cosine similarity between two documents d_i and d_j is given by:

$$similarity\ (d_i, d_j) = \frac{\sum_{k=1}^{n} (d_{ik} \cdot d_{jk})}{\sqrt{\sum_{k=1}^{n} (d_{ik})^2} \cdot \sqrt{\sum_{k=1}^{n} (d_{jk})^2}} \tag{7}$$

where:

- d_{ik} and d_{jk} are the TF-IDF weights of term k in documents d_i and d_j, respectively.
- n is the total number of unique terms across both documents.

Example: Calculate the cosine similarity between two product vectors with the following TF-IDF weights:

Document 1 (D1): [0.11465, 0, 0]

Document 2 (D2): [0,0.2293,0]

$$similarity\ (d_1, d_2) = \frac{(0.11465 \cdot 0) + (0 \cdot 0.2293) + (0 \cdot 0)}{\sqrt{(0.11465)^2 + (0)^2 + (0)^2} \cdot \sqrt{(0)^2 + (0.2293)^2 + (0)^2}} = 0 \tag{8}$$

Since the vectors have no common terms, the similarity is 0, indicating no similarity.

Graph Construction: A network is created in which the nodes represent products and the edges are weighted according to the cosine similarity between product pairs. The application of a

ISSN 2347–3657

Volume 10, Issue 2, 2022

https://doi.org/10.62646/ijitce.2022.v10.i2.pp110-123

threshold helps to eliminate insignificant similarities, which in turn ensures that the graph structure is meaningful.

Community Detection: Identifying hierarchical community structures within the graph through the utilization of algorithms such as Louvain or Girvan-Newman. These communities are groups of products that are quite similar to one another.

3.5.Evaluation and Analysis

Evaluating the product maps that were generated and doing an analysis of the insights that were gathered is the final phase. This includes the following:

Visualization

Displaying the product map and the community structures that it contains through the use of visualization tools. In order to generate visualizations that are easy to understand, tools such as Gephi and NetworkX are utilized.

Scenario Investigations

Investigations will be conducted on particular product categories in order to demonstrate how the product maps might be applied in some real situations. For instance, conducting an analysis of the pricing of comparable products across a variety of SME internet sites.

Performance Metrics

Using criteria like as precision, recall, and F1-score to evaluate the effectiveness of the methods was done. These metrics conduct an evaluation of the precision of the community detection as well as the significance of the product similarities that have been detected.

4. RESULT AND DISCUSSION

The research used the TF-IDF technique to calculate term frequency values and convert them into word frequency vectors to analyze product similarities. This strategy showed product similarities, such as varied colors of "classic high waist skinny jeans," indicating site competition. A product map was created using cosine similarity scores to visualize the product network and Louvain community detection. This technique found rich within-community and sparse between-community links. The investigation identified six product communities: decorations, knits and bodysuits, jeans, children's clothing, sportswear, and dresses. The findings help SMEs analyse competitors, make strategic decisions, and comprehend market dynamics by revealing product positioning and resource allocation. The report recommended extending data sources, including customer reviews, and creating improved algorithms for accuracy and prediction. To further understand complicated product and service interactions, future study should examine the method's application to different industries. The decentralized, bottom-up strategy shown here can improve SMEs' e-commerce strategies by giving market insights and facilitating decision-making.

ISSN 2347–3657
Volume 10, Issue 2, 2022
https://doi.org/10.62646/ijitce.2022.v10.i2.pp110-123

Table 3: An Overview of the Results and Their Implications.

Aspect	Description
Objective	Measure product similarities using TF-IDF and visualize product networks with community detection.
Method	TF-IDF algorithm, cosine similarity, Louvain method for community detection.
Key Findings	Identified six major product communities; highlighted competitive relationships between sites.
Implications	Assists SMEs in competitive analysis, strategic decision-making, and understanding market dynamics.
Limitations	Need for expanded data sources, advanced algorithms, and exploration of cross-industry applications.
Future Directions	Incorporate customer reviews, develop predictive models, and apply methods to other industries.

5. CONCLUSION AND FUTURE SCOPE

The revolutionary power of sophisticated analytics, especially TF-IDF, in simplifying the way SMEs handle the intricacies of e-commerce is highlighted by this research. SMEs can better position themselves in the market, allocate resources more efficiently, and improve customer interaction tactics by clarifying complex product linkages and competitive landscapes. SMEs are better equipped to survive and even prosper in the cutthroat digital economy by utilizing TF-IDF and related approaches. By investigating other natural language processing methods and using real-time data analytics, future research projects can build on this study. In addition to improving predictive skills and helping SMEs quickly adjust to changing market dynamics and consumer trends, this could further improve product mapping processes. SME resilience and competitiveness in the dynamic e-commerce environment are expected to increase with adoption of such innovations.

REFERENCES

1. Tariq, M. U., Babar, M., Poulin, M., Khattak, A. S., Alshehri, M. D., & Kaleem, S. (2021). Human behavior analysis using intelligent big data analytics. Frontiers in Psychology, 12, 686610.

ISSN 2347–3657
https://doi.org/10.62646/ijitce.2022.v10.i2.pp110-123
Volume 10, Issue 2, 2022

2. Lugier, O., Thakur, N., Wu, L., Vockenhuber, M., Ekinci, Y., & Castellanos, S. (2021). Bottom-Up nanofabrication with extreme-ultraviolet light: Metal–organic frameworks on patterned monolayers. ACS Applied Materials & Interfaces, 13(36), 43777-43786.
3. Naresh Kumar Reddy Panga (2021) Optimized Hybrid Machine Learning Framework for Enhanced Financial Fraud Detection Using E-Commerce Big Data, International Journal of Management Research and Reviews, Volume 11/Issue 2/18-35
4. Golov, A., & Carrasco, J. (2021). Molecular-Level Insight into the Interfacial Reactivity and Ionic Conductivity of a Li-Argyrodite Li_6PS_5Cl Solid Electrolyte at Bare and Coated Li-Metal Anodes.
5. Motoki, K., Saito, T., & Onuma, T. (2021). Eye-tracking research on sensory and consumer science: A review, pitfalls and future directions. Food Research International, 145, 110389.
6. Gaur, V. K., Gupta, S., & Pandey, A. (2021). Evolution in mitigation approaches for petroleum oil-polluted environment: recent advances and future directions. Environmental Science and Pollution Research, 1-17.
7. Jana, E., & Uma, V. (2020). Aspect extraction and sentiment analysis for e-commerce product reviews. In Emerging Trends in Computing and Expert Technology (pp. 780-791). Springer International Publishing.
8. Ma, J., Li, X., Li, C., He, B., & Guo, X. (2019). Machine learning based cross-border E-commerce commodity customs product name recognition algorithm. In PRICAI 2019: Trends in Artificial Intelligence: 16th Pacific Rim International Conference on Artificial Intelligence, Cuvu, Yanuca Island, Fiji, August 26-30, 2019, Proceedings, Part III 16 (pp. 247-256). Springer International Publishing.
9. Belem, F. M., Silva, R. M., de Andrade, C. M., Person, G., Mingote, F., Ballet, R., ... & Goncalves, M. A. (2020). "Fixing the curse of the bad product descriptions"–Search-boosted tag recommendation for E-commerce products. Information Processing & Management, 57(5), 102289.
10. Utomo, P. E. P., Manaar, M., Khaira, U., & Suratno, T. (2019). Analisis Sentimen Online Review Pengguna Bukalapak Menggunakan Metode Algoritma TF-IDF. JUSS (Jurnal Sains dan Sistem Informasi), 2(2), 35-39.
11. Meertens, Q. A., Diks, C. G. H., Van den Herik, H. J., & Takes, F. W. (2020). A data-driven supply-side approach for estimating cross-border Internet purchases within the European Union. *Journal of the Royal Statistical Society Series A: Statistics in Society*, *183*(1), 61-90.
12. Villarroel Ordenes, F., & Zhang, S. (2019). From words to pixels: text and image mining methods for service research. *Journal of Service Management*, *30*(5), 593-620.

International Journal of
Information Technology & Computer Engineering

ISSN 2347–3657

Volume 10, Issue 2, 2022

https://doi.org/10.62646/ijitce.2022.v10.i2.pp110-123

13. Subhashini, L. D. C. S., Li, Y., Zhang, J., Atukorale, A. S., & Wu, Y. (2021). Mining and classifying customer reviews: a survey. *Artificial Intelligence Review*, 1-47.

14. Rajeswaran Ayyadurai (2021) Big Data Analytics and Demand-Information Sharing in E-Commerce Supply Chains: Mitigating Manufacturer Encroachment and Channel Conflict, International Journal of Applied Science Engineering and Management, Vol 15, Issue 3, 2021