

Unveiling Metaverse Social Trends: Analysing Big Data Regarding Online Sports News With LDA-Based Topic Modelling

Ju Chan Na¹, Eun Joung Kim^{2*}, Jung Yoon Kim^{3*}

Abstract

The rise of digitalization has led to a significant increase in the attention given to online sports news. In a similar vein, the Metaverse has gone through various stages of evaluation. At present, it is in a developmental phase where real-world socio-economic and cultural activities are seamlessly integrated into a 3D virtual environment. Understanding the impact and implications of the Metaverse 3.0 era is highly sought after. This study seeks to investigate social issues related to the Metaverse in South Korea. It will analyse a large dataset of online sports news articles from June 2021 to May 2022 using LDA-based topic modelling. As a result, four main topics were identified: a new platform, a new business model, pervasive online cultural and educational media, and a new frontier for business and investment. New Platform (Topic 1) showcased keywords indicating the potential of Metaverse as a cutting-edge platform, while New Business (Topic 2) highlighted keywords indicating digital media companies' pursuit of a new profit model within the virtual realm of Metaverse. The increasing prevalence of online culture and educational media (Topic 3) highlights the growing demand for a wide range of content that caters to the virtual world's cultural and educational experiences. "New Frontier for Business and Investment (Topic 4) highlights the significant involvement of digital media companies in acquiring investments, technologies, and intellectual property (IP) to gain a competitive edge in the global Metaverse market. This study solely examined online sports data, but future research could encompass a wider range of raw data sources, such as interviews with practitioners, user reviews, and research reports. By doing so, a more comprehensive understanding of the potential challenges that the Metaverse may encounter can be obtained.

Keywords: Metaverse, Sports News, Text Mining, Big Data, Latent Dirichlet Allocation (LDA), Topic Modelling.

1. Introduction

Due to the rise of digitalization, online sports news has become more prevalent. It is worth noting that digital platforms and social media have become increasingly dominant as the primary sources of sports news consumption (Morehead, O'Hallarn, & Shapiro, 2016). Users are increasingly relying on digital platforms like Instagram, Twitter, and sports-focused applications for real-time updates, news, and other highlights. In addition, the use of multimedia content, such as interactive graphs and videos, has significantly improved the user experience (Pather, 2021). In addition, there has been an increasing trend in the integration of sports and technology, particularly in the use of virtual or augmented reality for sports coverage (Viscione & D'Elia, 2019). This approach allows sports enthusiasts to have a more immersive experience. The expansion of sports-related content encompasses more than just match results and sports-

related information. It also includes behind-the-scenes stories, athlete profiles, and social justice issues. Thus, it demonstrates a comprehensive and all-encompassing perspective on sports journalism. In addition, the impact of sports was apparent during the pandemic (Pu, Kim, & Daprano, 2021). This is due to the growing emphasis on resilience and adaptability in sports, as well as changes in fan engagement caused by restrictions on in-person attendance.

The Metaverse has gained popularity as a virtual space for people to interact, share experiences, and consume content. Considering the dynamics of online sports news is highly important in this context (Pizzolante et al., 2023). As a result, the prevalence of online sports news has grown significantly. The LDA-based topic modelling allows for the extraction of significant themes and patterns from large datasets. Thus, it reveals the pattern of public interest in the digital realm. The metaverse is a dynamic digital environment where individuals participate in a wide range

¹ Department of IT Convergence Engineering, Gachon University, Seongnam-si, Republic of Korea.

Email: najuchan@naver.com

² Department of Game Content, College of Smart Content, Kyungil University, Gyeongsan-si, Gyeongbuk, 38428, Republic of Korea.

Email: ejkim@kiu.ac.kr

³ Department of Game Media, College of Future Industry, Gachon University, Seongnam-si, 13120, Gyeonggi-do, Republic of Korea.

*Corresponding author Email: kjyoon@gachon.ac.kr & ejkim@kiu.ac.kr

of activities. Additionally, it encompasses the consumption of sports content. Examining the impact of big data on online sports news in this context can provide valuable insights into how users interact with and consume sports-related information in the metaverse. LDA-based topic modelling is useful for identifying and analysing the main themes and topics in online sports news (Onan, Korukoglu, & Bulut, 2016). This information is valuable for content creators who can align their content with the preferences and interests of the audience in the metaverse.

By considering the pertinent and widely debated subjects, content can be crafted to optimise engagement and resonance within the online community. The concept of the "Metaverse" is increasingly becoming a reality for the public, thanks to advancements in Augmented Reality (AR) and Virtual Reality (VR) technology. Society's desire for virtual connection has also contributed to this growing trend. The term "Metaverse" is derived from "Meta," meaning transcendence, and "Universe," meaning universe. It describes a 3D virtual environment that exists independently of reality. Described as a digital realm closely reflecting and interweaving with the physical world, it is celebrated as a novel digital paradigm for future generations. It has the potential to revolutionise human activities by removing physical limitations, leading to a substantial increase in socialisation, productivity, shopping, and entertainment. There has been a continuous and ongoing industrial interest and investment in the Metaverse since its initial attraction in 2020. Roblox and Zepeto are widely regarded as leading platforms in the Metaverse. These platforms continue to dominate sales in both Google Play and the App Store. Notably, Zepeto has garnered a massive user base of 300 million, with a strong emphasis on fostering user communication.

Metaverse strives to offer immersive and interactive experiences focused on online sports news trends and insights, enhancing the user's overall experience. Developers can use the findings to integrate various features that align with community interests, such as interactive discussions, virtual events, or sports simulations that provide an immersive experience. Therefore, the user's experience is enhanced, and user satisfaction is also heightened. Regarding this matter, businesses operating in the metaverse and engaging in sports-related ventures can gain valuable insights into social trends. By utilising LDA-based topic modelling, it becomes possible to analyse online sports news data and identify market trends, assess sentiment, and estimate opportunities for marketing campaigns. This information is valuable for businesses looking to optimise their practices. It also connects their services or products with

the growing trends and attitudes in the virtual sports industry.

Extensive research has been conducted on the topic of big data analytics, with numerous studies exploring various aspects (Himeur et al., 2023; Maroufkhani, Iranmanesh, & Ghobakhloo, 2023; Talaoui et al., 2023). In addition, researchers have extensively studied LDA-based topic modelling (Herqutanto, Zafari, & Sutoyo, 2023; Uthirapathy & Sandanam, 2023; Xue et al., 2024). However, it is worth noting that there is a lack of scholarly attention dedicated to studying metaverse social trends. In addition, there is a scarcity of research that has examined large datasets on online sports news using LDA-based topic modelling. The objective of this study is to address the existing research gap by examining the social trends within the metaverse. The researcher has also aimed to focus on online sports news and big data through LDA-based topic modelling. This research is significant due to its relevance to the emerging trend of online sports news. It has the potential to provide valuable insights. These dynamics pertain to sports-related conversations in the Metaverse. Furthermore, it impacts the generation of content, user experiences, market strategies, and overall advancement of virtual spaces related to sports enthusiasts. In addition, this study offers valuable insights into fan behaviour and preferences by considering how sports news is consumed in the metaverse. Given the ongoing digital transformation in the sports industry, the metaverse is seen as the next frontier. Understanding the impact of social trends in the digital realm is crucial for sports organisations looking to adjust their operations, communication strategies, and marketing to meet the changing needs of fans.

2. Literature Review

The Metaverse is a virtual space that enables people to communicate and interact with others using avatars. It is currently gaining significant attention, particularly from the younger generation (Park & Kim, 2020). During the 2000s, the Internet-based virtual reality game Second Life gained popularity and was referred to as Metaverse 1.0. From 2016 to 2019, there was a significant development and commercialization of immersive technologies such as augmented reality (AR), virtual reality (VR), and extended reality (ER). This period marked a transition for Metaverse (Huang, Backman, & Backman, 2010; Kaplan & Haenlein, 2009). Once the foundational technologies supporting the 3D immersive digital world of Metaverse reached a state of advancement and stability, the virtual realm of digital games for young individuals merged with social network services (SNS). This resulted in the creation of a vast digital

platform where users can engage in socialising, gaming, shopping, and even conducting transactions for buying and selling goods. Research on the Metaverse has been conducted in various areas, including technical aspects (Chen, 2022), distribution of live performances (Baía Reis

& Ashmore, 2022), copyright issues (Kim & Jeon, 2021), and educational tools (Hwang & Chien, 2022). Figure 1 illustrates the various stages of Metaverse and the corresponding technologies that have been implemented over time.

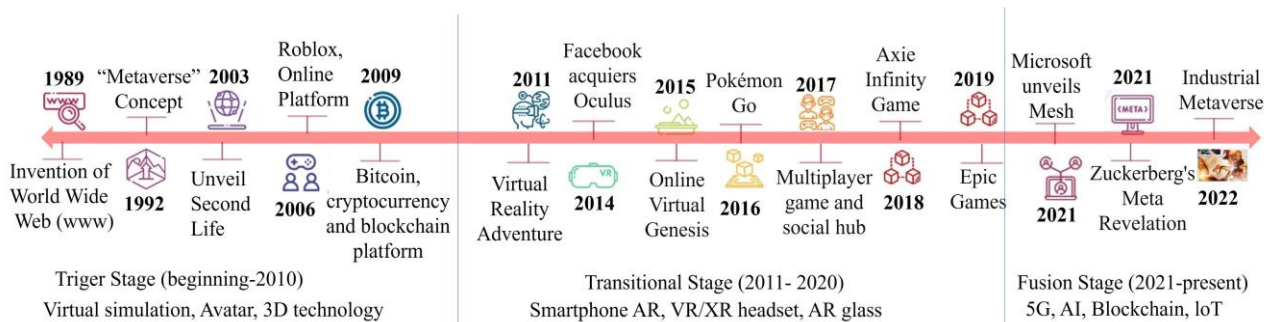


Figure 1. Different stages of Metaverse and Applied Technologies.

However, with the recent rebranding of Facebook to Meta, there has been a surge in discussions regarding the potential of the Metaverse as the next significant technological advancement and the optimal path for its development. The utilisation of text mining analytics primarily focused on the examination of extensive online news data (Lee, 2021). Additionally, social networking sites such as Instagram, Twitter, and blogs were employed to gain insights into the general public's perception of the Metaverse (Zhe et al., 2023). Topic modelling is a commonly employed technique in text mining that allows for the identification of latent semantic structures and abstract topics within a set of documents. Probabilistic topic models are utilised in machine learning algorithms to automatically extract latent keywords and topics from extensive text bodies. This approach proves valuable in identifying keywords and major trends within large amounts of text, including online news data (Asmussen & Møller, 2019; Kim, Eun, & Kim, 2016).

The field of big data analysis garners significant interest from both academic and industry circles due to its focus on extracting valuable insights from collected data (Kyu-Dong, Ju, & Hwang, 2022; Zhu et al., 2018). The sheer volume of text data generated by social media platforms, media companies, and online communities is too immense to be comprehended in real-time. Various text mining analytics are currently employed for extracting meaningful information (Kim, 2019).

By analysing review data, companies can gain insights into consumer behaviour and product preferences. Similarly, news data can shed light on the issues that are being shaped. In a study conducted by Wui, Na, and Park (2018), the researchers examined users' interest in VR games by analysing user reviews from an online game platform. Additionally, Youm and Kim (2022) proposed a text

mining approach to enhance the design of mobile RPGs based on an analysis of user reviews from the Google Play Store (Youm & Kim, 2022). In a recent study, Kim, Son, and Jeon (2021) analysed online news data to gain insights into public opinion regarding probability-based items sold in gaming (Kim et al., 2021). Additionally, Lee and Youk (2021) examined news coverage of digital games following the World Health Organization's classification of 'gaming disorder' as a behavioural addiction (Lee & Youk, 2021). This study aimed to investigate the social issues related to Metaverse in South Korea. It analysed big data from online news articles about sports during the period when the term Metaverse gained rapid popularity. The analysis was conducted using LDA-based topic modelling. This study specifically examines the time from June 1, 2021, to May 31, 2022, during which there was a significant surge in social interest and enough news reports for data analysis. The aim is to uncover valuable insights into the future direction and trends of the Metaverse.

2.1. LDA-Based Topic Modelling and Visualization

Topic modelling using LDA is a text mining technique that extracts underlying principles from large amounts of text data. In the beginning, the probabilistic latent semantic analysis (pLSA) technique was employed. However, after the publication of the Latent Dirichlet Allocation (LDA) algorithm by Blei, Ng, and Jordan (2003), LDA has become the predominant technique for topic modelling. Visualisation and LDA-based topic modelling are effective tools for data analysis. This topic specifically pertains to the management of extensive unstructured data sets. LDA is a significant probabilistic model used to identify latent topics within a collection of documents. It is important to grasp the underlying themes present in the vast amount of textual data.

However, it is especially important when working with lengthy news articles, research papers, or social media posts. This pertains to the specific topic, providing a succinct and organised overview. LDA can be used to create topic-based indice (Farkhod et al., 2021; Gupta et al., 2022; Uthirapathy & Sandanam, 2023). As a result, there are enhancements in information retrieval systems. Users can easily locate relevant articles or documents by searching for specific topics of interest. The enhanced accessibility and usability of data leads to overall improvements. The results obtained from complex topic modelling can be transformed into visual representations such as bar charts, word clouds, and network graphs. It helps to analyse and understand the connections and patterns in the data. The use of visualisation tools facilitates comprehension of the main text's key themes and associations for both technical and non-technical audiences.

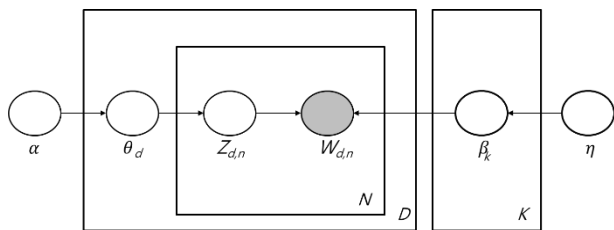


Figure 2. Simple Sketch of LDA-Based Topic Modeling.

Figure 2 presents a simplified outline of LDA-based topic modelling. In this context, K represents the number of topics, α is a parameter that influences the θ value, and η is a parameter that determines the β value. θ represents the ratio of topics per document, while β denotes the generation ratio of W, which is a word per topic. The topic of the n^{th} word in document d is denoted as $Z_{d,n}$. $W_{d,n}$ is a variable that represents the n^{th} word in document d. The θ and β values adhere to the Dirichlet distribution, while Z is derived from θ , representing the topic ratio for each article. A word (W) is formed based on Z, a value that reflects the topic of the word, and β , the proportion of word production per topic (Jelodar et al., 2019; Park & Song, 2013).

Topic modelling is an unsupervised machine learning method that can analyse a collection of documents and identify patterns of words and phrases in them. Several studies have been conducted to determine the research direction in

mathematics education (Jin & Ko, 2019) and its relationship to the fourth industrial revolution (Cho & Woo, 2019). Kim's research team analysed news data released during the COVID-19 pandemic to discover patterns in reporting (Kim, 2020). Cho and colleagues determined the characteristics that contribute to the quality of airline services by analysing internet reviews (Cho & Lee, 2021). LDAvis is a software tool designed to facilitate the visualisation of topic models in an interactive manner. Sievert and Shirley (2014) created it, and Park and Oh (2017) made further improvements. The visualisation of LDA findings allows for the observation of trends and patterns across themes. This is particularly significant in identifying movements in public mood, dominant issues, or changes in focus within a single topic throughout time. Consequently, the decision-makers are furnished with significant insights (Sharma, Sharma, & Sakshi, 2022). Utilising Latent Dirichlet Allocation (LDA) for text analysis allows for the examination of how a topic develops and transforms over time (Lee, Kim, & Han, 2022).

Nevertheless, it can also be crucial in monitoring the advancement of conversations. By employing this approach, one can discern pivotal occurrences that influence the transitions in subject matter. The LDA-based topic modelling and visualisation are crucial for revealing patterns, organising information, and providing interpretable insights from large datasets. These techniques are crucial for extracting valuable knowledge, facilitating decision-making, and enhancing comprehension of intricate textual data sets.

3. Research Methods

An analysis was conducted to extract meaningful topics from raw online news data. This process involved multiple steps, as shown in Figure 3. At first, online new data was collected using BIG KINDS. It was then pre-processed by filtering out unnecessary words through three procedures: tokenization, text cleansing, and moralization. The processed data were subsequently analysed using three methods: frequency, TF-IDF, and LDA. Ultimately, the results of topic models can be visualised to facilitate easy interpretation. This allows for the identification of current trends and the future direction of the Metaverse.

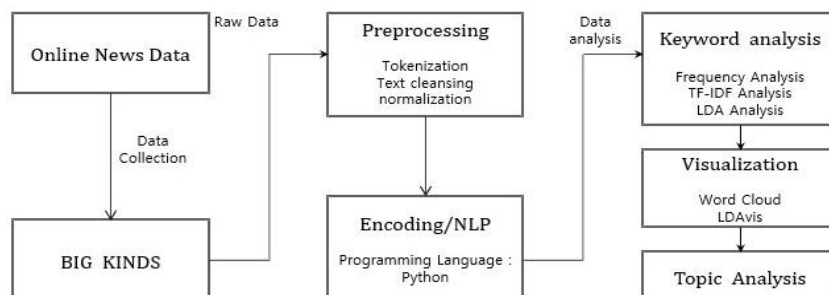
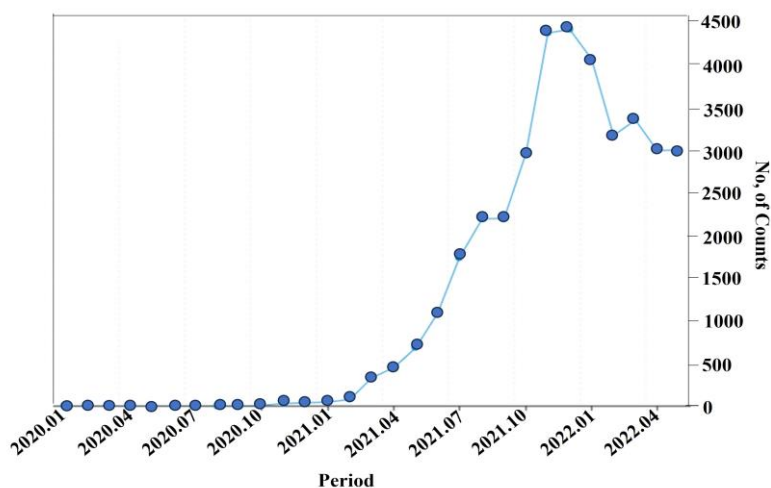


Figure 3. Procedure Followed for Online News Data Extraction.

Data collection and processing were conducted using BIG KINDS and Colab. BIG KINDS, a news big data solution operated by the Korea Press Foundation, allows users to search, view, download, and collect news articles from around 54 media companies dating back to 1990 (Chun, Kim, & Lee, 2023). Colab is a Python development environment that operates within the browser through Google Cloud.

3.1. Data Collection

Data was collected from a large dataset using the search term "Metaverse" during the period of June 1, 2021, to May 31, 2022. Throughout this timeline, there was a significant rise in the amount of data related to the Metaverse, as shown in Figure 4. A grand total of 36,388 data points were gathered.



Period	Counts
June 2021	1115
July 2021	1769
August 2021	2229
September 2021	2235
October 2021	3045
November 2021	4555
December 2021	4582
January 2022	4144
February 2022	3230
March 2022	3406
April 2022	3027
May 2022	3051
Total	36,388

Figure 4. Number of News Data Issued During January 2020 to April 2022.

3.2. Preprocessing Process

Preprocessing refers to the process of ensuring the integrity of data by removing special characters, digits, and other unnecessary keywords that may hinder analysis (Yang, 2021). Removed The names of specific entities, regions, countries, and other relevant keywords related to online sports news were condensed and included in Tables S1 and S2 for analysis purposes.

3.3. Keyword Analysis

An analysis was conducted to identify the main keywords of Metaverse, using frequency and TF-IDF analysis. Frequency analysis involves identifying the co-occurrence of words, while TF-IDF is a statistical measure that quantifies the significance of a keyword in a document relative to a collection of documents (Kim & Kim, 2021). The TF-IDF formula is expressed as follows.

$$TF-IDF(t, d, D) = TF(t, d) * IDF(t, D)$$

The codes used for frequency and TF-IDF analysis are shown in Table S3.

4. Results

4.1. Frequency and TF-IDF Analysis

The top 30 keywords were selected and ranked after frequency and TF-IDF analysis, as shown in Table 1.

Table 1

Frequency and TF-IDF Rank.

Rank	Keyword	Freq	Rank	Keyword	TF-IDF
1	Metaverse	26139	1	Metaverse	0.643032
2	Platform	8751	2	Education	0.473875
3	Investment	8222	3	Investment	0.45992
4	Business	8050	4	Business	0.421983
5	Technology	7609	5	Digital	0.402431
6	Education	7172	6	Platform	0.401022
7	Digital	7041	7	Game	0.401005
8	Game	5778	8	Technology	0.39175
9	Content	5028	9	Content	0.336051
10	Covid 19	4921	10	AI	0.328332
11	Service	4887	11	NFT	0.312159
12	AI	4704	12	Service	0.307731
13	NFT	4109	13	SK	0.301433
14	Global	3974	14	Covid-19	0.300811
15	Virtual	3693	15	Global	0.267154
16	SK	3528	16	Virtual	0.260609
17	Online	3267	17	KT	0.241063
18	KT	2625	18	Online	0.233456
19	Blockchain	2338	19	Kakao	0.213336
20	Virtual world	2303	20	Startups	0.205513
21	Startups	2246	21	Blockchain	0.200899
22	Kakao	2167	22	Naver	0.191299
23	Naver	1977	23	Virtual world	0.181494
24	Reality	1780	24	ICT	0.169953
25	ICT	1690	25	Tourism	0.162548
26	Contact-free	1669	26	Reality	0.156445
27	Convergence	1559	27	Zepeto	0.147354
28	Tourism	1490	28	Contact-free	0.146289
29	Virtual reality	1472	29	Convergence	0.145753
30	Zepeto	1464	30	Scientific Technology	0.139175

The frequency analysis revealed that the words "Metaverse" (26139), "Platform" (8751), "Investment" (8222), "Business" (8050), and "Technology" (7609) were among the top 30 keywords. The TF-IDF analysis revealed that the

following terms were among the top 30 ranks: Metaverse (0.643), Education (0.473), Investment (0.459), Business (0.421), Digital (0.402), and Platform (0.401). Figure 5 illustrates the word cloud visualisation based on frequency.

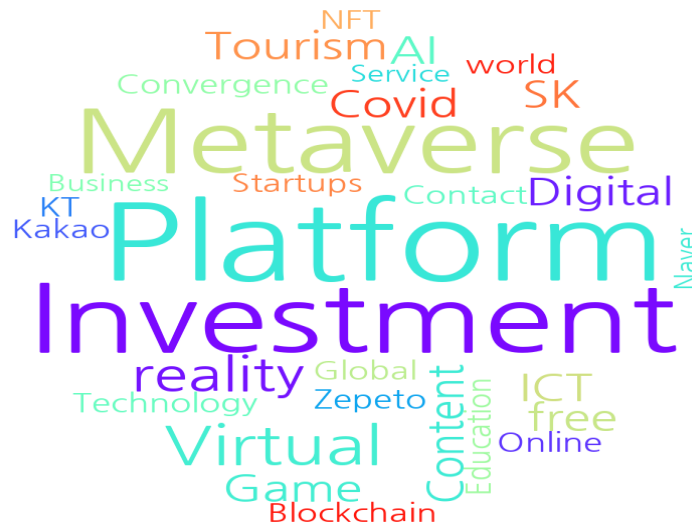


Figure 5. Word Clouds Visualization of Metaverse.

4.2. LDA Result

To identify the ideal number of topics based on K values, one can measure and compare both coherence scores and perplexity values. The study found that the highest ratings were achieved when there were four subjects considered, with a focus on easy interpretability.

Figure 6 and Table 2 illustrate that Topic 1 (New Platform) encompasses terms such as Metaverse, Platform, Business, Kakao, Virtual World, Service, Zepeto, Scientific Technology, Avatar, and Virtual Space. Hence, it can be construed as a fundamental framework that enables people to access or immerse themselves in the virtual realm of the Metaverse.

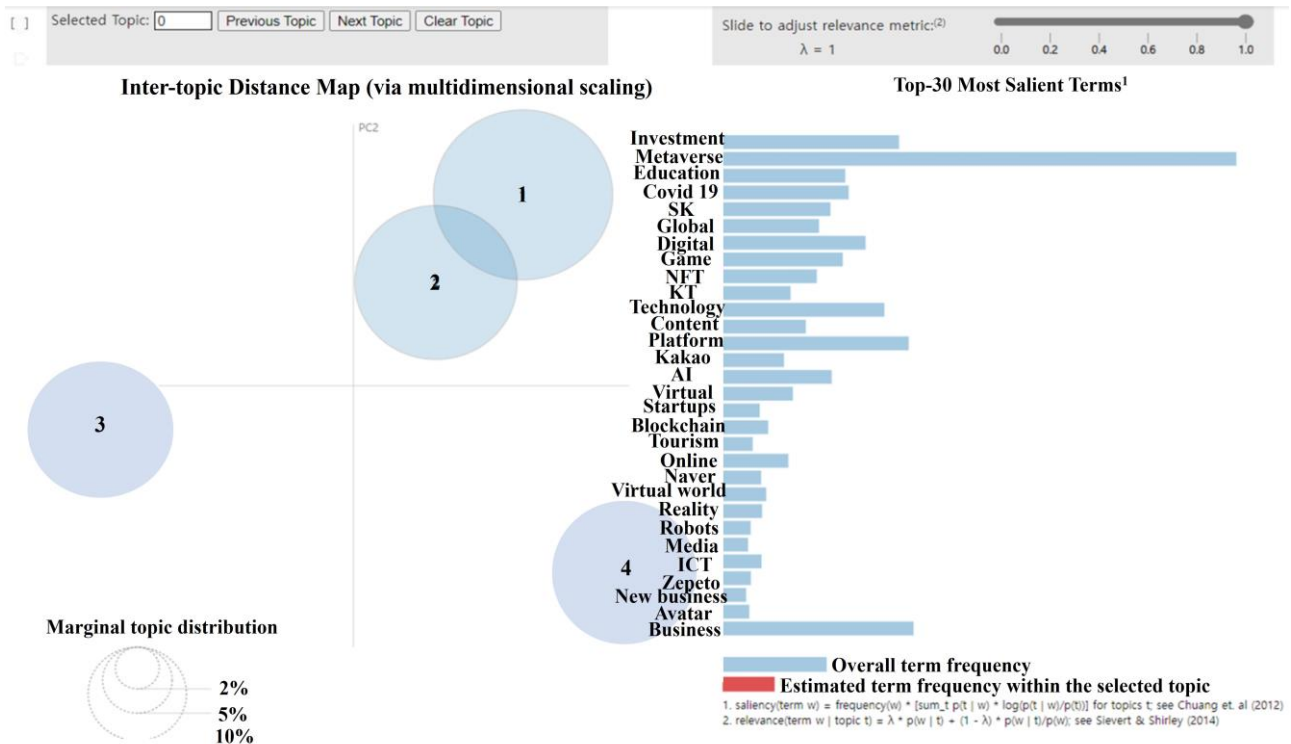


Figure 6. Visualization of LDA-Based Topic Modelling of Metaverse Including Four Topics (1) New Platform (2) New Business Model (3) 'Pervasive Online Cultural and Educational Media (4) 'New Frontier for Business and Investment.

Topic 2, 'New Business Model', encompasses several terms such as Technology, Digital, NFT, Content, Blockchain, Service, Metaverse, and Game. This signifies the sincere endeavours to integrate NFT and blockchain technologies into the content of the Metaverse.

Topic 3 (Pervasive Online Cultural and Educational Media) encompasses terms such as COVID-19, Contact-free, Education, Performance, Exhibition, Online, AI, and Tourism. It pertains to digital material that can be accessed and enjoyed from home, such as online educational

resources, virtual concerts, and tourism applications. This demonstrates that the influence of COVID-19 continues to affect the trend of contact-free living and its associated entertainment culture.

Topic 4 (New Frontier for Business and Investment) includes a range of relevant keywords such as investment, global, KT, AI, startups, Naver, and business. The intense competition among corporations in the Metaverse industry highlights their efforts to secure funding and gain a competitive edge in the market.

Table 2

Keywords According to the Topics.

Topic	Cluster	Keyword
Topic 1	New Platform	Metaverse, Platform, Business, Virtual, Kakao, Virtual World, Reality, Service, Zepeto, Scientific technology, Avatar, Virtual space, Online, MZ, Arts, Technology, Digital, Exhibition Hall, Videos, Virtual Reality, Contact-free, Science, Convergence, Naver, Internet, Contest, Mobile, Performance, Offline, R&D
Topic 2	New Business Model	Technology, Digital, NFT, Content, Blockchain, Service, Metaverse, Business, Game, ICT, Exhibition, Robots, Virtual asset, AI, Videos, Solution, U+, Facebook, YouTube, Information technology, Mobile, Virtual reality, Augmented reality, Ecology, Cloud, New technology, Convergence, Network, IP, Next generation COVID-19, Education, SK, Games, Online, AI, Tourism, Contact-free, Offline,
Topic 3	Pervasive Online Cultural and Educational Media	Metaverse, ifland, Design, Microsoft, Learning, Software, ICT, New employees, Convergence, Exhibition, Industrial revolution, Mobile, Performance, Business, Platform, YouTube, Information technology, Service, Infrastructure, PC, Video Investment, Global, KT, AI, Startups, Metaverse, Naver, Business,
Topic 4	New Frontier for Business and Investment	Service, Platform, Media, New business, Technology, Apple, Entertainment, Com2us, Big data, Commerce, Cryptocurrency, Movies, Solution, Live, music, Cloud, Ecology, Information technology, IP, Infrastructure, Mobile, Next generation

5. Discussion

The objective of this study was to evaluate the current social trends in the metaverse. The researcher conducted an analysis of extensive online sports news data using LDA-Based topic modelling. The results demonstrate the consequences that arise from the categorised keywords of four subjects. The New Platform cluster suggests that the Metaverse is being used by young athletes as an online communication space. This results in the development of numerous exhibitions or performances within the virtual realm under the Metaverse clouds. This highlights the rise of a new platform for public interaction and brand marketing. Therefore, it is essential to incorporate human-touched design or spatial storytelling within the Metaverse to attract a wide range of uses. Users can establish meaningful connections with their virtual avatars and engage in experiences beyond the confines of reality.

In addition, the New Business Model cluster suggests that digital companies are utilising technologies such as NFTs and Blockchain for virtual transactions and business purposes. At present, the process of obtaining virtual items typically involves users being granted usage rights by the service provider, rather than actual ownership. Thus, in the event of service termination, the items will cease to exist as well. When the NFT is utilised, users gain complete ownership of virtual items. From the user's perspective, the utilisation of NFT technology enhances the value and dependability of virtual items. The sports content industry is undergoing rapid changes, with new content being released daily. Therefore, digital companies, including those in the public sector, are actively exploring the potential business prospects offered by the Metaverse and its cutting-edge technologies.

In addition, the Pervasive Online Cultural and Educational Media cluster reflects the shift from traditional outdoor

activities to a more integrated and widespread way of incorporating cultural and educational content into daily life. Indeed, the digital content industry is currently receiving significant attention as a promising source of growth. However, it is worth noting that the availability of Metaverse content remains limited, with a concentration among a few major companies. Hence, it is imperative for the public sector to establish a consultative body to engage in comprehensive discussions with industry and academia. Additionally, a well-defined mid-to-long-term plan should be formulated to nurture the content industry.

In the realm of Business and Investment, it is evident that digital media companies are actively pursuing investments, technologies, and intellectual property (IP) to gain a competitive edge in the global Metaverse market. Efficiently navigating the unexplored digital territories of the Metaverse, and keeping up with its evolving trends, necessitates the rapid development of skilled individuals. To achieve this goal, it is imperative to have active investment at the national level, both from an economic and policy standpoint. The impact of digital media companies on the sports industry cannot be underestimated, as the developments in the metaverse for business and investment have shown. The future of sports can be shaped in transformative and innovative ways through the integration of virtual experiences, the emergence of virtual sports competition, and the cultivation of talent for the metaverse.

6. Conclusion

This study aimed to investigate significant societal issues in the Metaverse. An LDA-based topic modelling approach was utilised to analyse online sports news data collected over a 12-month period, spanning from June 2021 to May 2022. From the analysis, a total of four issues were identified: New Platform, New Business, Pervasive Online Culture, and Educational Media. Keywords associated with "New Platform" (Topic 1) indicate that the Metaverse has the potential to serve as a platform for future generations. On the other hand, keywords related to "New Business" (Topic 2) suggest that digital media companies are exploring new ways to generate profits within the virtual realm of the Metaverse.

To meet the growing demand for virtual world culture and education, it is necessary to provide additional material. This is particularly important in the context of Pervasive Online Culture and Education Media (Topic 3). Search terms related to "New Frontier for Business and Investment" (Topic 4) suggest that digital media companies are actively seeking investments, technology,

and intellectual property (IP) to gain a competitive advantage in the global Metaverse market. The Metaverse has gained significant attention during the COVID-19 pandemic, establishing itself as a popular platform for communication and entertainment among the younger generation. There appears to be a possible change in the main demographic of users in the digital market.

The Metaverse is revolutionising collaboration, commerce, and culture, particularly in the realm of sports, by introducing innovative ways to transform and transcend virtual experiences. To ensure competitiveness in the Metaverse market, government departments and private corporations collaborate to establish adaptable environments where companies and creators can explore solutions and innovations that pave the way for progress.

7. Research Implications

The Present Study Holds Various Theoretical and Practical Implications Which Are Stated Below:

Theoretical Implications

This study has important theoretical implications as it contributes to the existing literature on online sports news and LDA-based topic modelling. It helps to comprehend how individuals engage with sports information in virtual environments. Furthermore, this study provides guidance for content creators to tailor their work to specific interests. This study contributes to the development of virtual experiences that align with user preferences by unveiling prevalent themes. The findings of this study provide valuable information about the changing dynamics within the digital sports community. Additionally, it has an impact on marketing strategies. In addition, this study highlights the connection between sports and technology. This provides a theoretical foundation for incorporating new elements into the Metaverse. This study offers a theoretical basis for understanding and adjusting to the changing landscape of sports news in virtual environments.

Practical Implications

In addition to its theoretical implications, this study provides valuable practical insights. The findings of this study can be utilised by content creators and sports platforms to customise their content. By doing so, we can ensure that it is in line with the current and significant subjects in the virtual sports community. This optimisation has the potential to enhance user experience and satisfaction. The research findings can be utilised by developers to create immersive and user-friendly virtual environments. This can include virtual events and

interactive discussions that align with the prevalent interests in the metaverse.

Sports organisations and companies in the metaverse space can make well-informed strategic decisions based on identified trends. Countries that acknowledge the potential of the metaverse in sports can also facilitate collaborations between public and private sectors. By adopting this approach, we can encourage innovation and facilitate the growth of talent. In addition, nations that understand the potential of the metaverse can establish themselves as significant participants in the worldwide virtual sports landscape.

8. Research Limitations and Future Indications

Given the limitations of this study in terms of the volume and type of online news data, it would be beneficial to expand the research to incorporate a wider range of raw data sources. This could include conducting interviews with practitioners, analysing user reviews, and examining research reports. This additional research will offer in-depth analysis to effectively address the potential obstacles

that Metaverse may encounter. Further research opportunities lie in the examination of social dynamics and community structures within the metaverse sports environment. Additionally, one can analyse user interactions and identify influential individuals or content creators. In addition, it can encompass studying the development of virtual sports communities.

Conflict of Interest: The authors state that there are no conflicts of interest related to the publication of this paper.

Funding Statement: This research is funded by the Ministry of Culture, Sports and Tourism and the Korea Creative Content Agency (Project Number: R2020040243)

Acknowledgements

This research is funded by the Ministry of Culture, Sports and Tourism and the Korea Creative Content Agency (Project Number: R2020040243)

Data Availability

The datasets utilised in this study can be obtained from the corresponding author upon request.

References

- Asmussen, C. B., & Møller, C. (2019). Smart literature review: a practical topic modelling approach to exploratory literature review. *Journal of Big Data*, 6(1), 93. <https://doi.org/10.1186/s40537-019-0255-7>
- Baía Reis, A., & Ashmore, M. (2022). From video streaming to virtual reality worlds: an academic, reflective, and creative study on live theatre and performance in the metaverse. *International Journal of Performance Arts and Digital Media*, 18(1), 7-28. <https://doi.org/10.1080/14794713.2021.2024398>
- Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet Allocation. *Journal of Machine Learning Research*, 3(Jan), 993-1022. <https://www.jmlr.org/papers/volume3/blei03a/blei03a.pdf>
- Chen, Z. (2022). Exploring the application scenarios and issues facing Metaverse technology in education. *Interactive Learning Environments*, 1-13. <https://doi.org/10.1080/10494820.2022.2133148>
- Cho, K.-W., & Woo, Y. (2019). Topic Modeling on Research Trends of Industry 4.0 Using Text Mining. *Journal of the Korea Institute of Information and Communication Engineering*, 23(7), 764-770. <http://doi.org/10.6109/jkiice.2019.23.7.764>
- Cho, M.-K., & Lee, B.-J. (2021). Comparison of service quality of full service carriers in Korea using topic modeling: based on reviews from TripAdvisor. *Hotel and Tourism Research*, 23(1), 152-165. <https://doi.org/10.31667/jhts.2021.2.86.152>
- Chun, J., Kim, J., & Lee, S. (2023). A content analysis of South Korean newspaper coverage of adolescent cyberbullying. *Children & Society*, 37(6), 2011-2029. <https://doi.org/10.1111/chso.12695>
- Farkhod, A., Abdusalomov, A., Makhmudov, F., & Cho, Y. I. (2021). LDA-based topic modeling sentiment analysis using topic/document/sentence (TDS) model. *Applied Sciences*, 11(23), 11091. <https://doi.org/10.3390/app112311091>
- Gupta, R. K., Agarwalla, R., Naik, B. H., Evuri, J. R., Thapa, A., & Singh, T. D. (2022). Prediction of research trends using LDA based topic modeling. *Global Transitions Proceedings*, 3(1), 298-304. <https://doi.org/10.1016/j.gltp.2022.03.015>
- Herqutanto, M. F., Zatarí, R. P., & Sutoyo, R. (2023). Topic Modeling Using LDA-Based and Machine Learning for Aspect Sentiment Analysis. In *2023 International Conference on Informatics, Multimedia, Cyber and Informations System (ICIMCIS)* (pp. 142-148). IEEE. <https://doi.org/10.1109/ICIMCIS60089.2023.10349056>
- Himeur, Y., Elnour, M., Fadli, F., Meskin, N., Petri, I., Rezgui, Y., Bensaali, F., & Amira, A. (2023). AI-big data analytics for building automation and management systems: a survey, actual challenges and future perspectives. *Artificial Intelligence Review*, 56(6), 4929-5021. <https://doi.org/10.1007/s10462-022-10286-2>
- Huang, Y.-C., Backman, S. J., & Backman, K. F. (2010). Student attitude toward virtual learning in second life: A flow theory approach. *Journal of Teaching in Travel & Tourism*, 10(4), 312-334. <https://doi.org/10.1080/15313220.2010.525425>

- Hwang, G.-J., & Chien, S.-Y. (2022). Definition, roles, and potential research issues of the metaverse in education: An artificial intelligence perspective. *Computers and Education: Artificial Intelligence*, 3, 100082. <https://doi.org/10.1016/j.caeai.2022.100082>
- Jelodar, H., Wang, Y., Yuan, C., Feng, X., Jiang, X., Li, Y., & Zhao, L. (2019). Latent Dirichlet allocation (LDA) and topic modeling: models, applications, a survey. *Multimedia Tools and Applications*, 78, 15169-15211. <https://doi.org/10.1007/s11042-018-6894-4>
- Jin, M., & Ko, H. K. (2019). Analysis of trends in mathematics education research using text mining. *Communications of mathematical education*, 33(3), 275-294. <https://doi.org/10.7468/jksmee.2019.33.3.275>
- Kaplan, A. M., & Haenlein, M. (2009). The fairyland of Second Life: Virtual social worlds and how to use them. *Business Horizons*, 52(6), 563-572. <https://doi.org/10.1016/j.bushor.2009.07.002>
- Kim, G., & Jeon, J. H. (2021). A Study on the Copyright Survey for Design Protection in Metaverse Period. *International Journal of Advanced Smart Convergence*, 10(3), 181-186. <https://doi.org/10.7236/IJASC.2021.10.3.181>
- Kim, I.-H. (2019). Newspaper Big Data and Text Mining for Digital Humanities. *Academic Journal of Grammar and Literature*, 78, 41-62. <https://doi.org/10.15565/jll.2019.06.78.41>
- Kim, J.-H., & Kim, S.-S. (2021). A study on the analysis of agricultural R&D keywords using textmining method. *Journal of the Korea Academia-Industrial Cooperation Society*, 22(2), 721-732. <https://doi.org/10.5762/KAIS.2021.22.2.721>
- Kim, S.-J., Eun, H.-J., & Kim, Y.-S. (2016). Food recommendation system using big data based on scoring taste adjectives. *International Journal of u-and e-Service, Science and Technology*, 9(1), 39-52. <http://dx.doi.org/10.14257/ijunesst.2016.9.1.05>
- Kim, T.-J. (2020). COVID-19 news analysis using news big data: Focusing on topic modeling analysis. *The Journal of the Korea Contents Association*, 20(5), 457-466. <https://doi.org/10.5392/JKCA.2020.20.05.457>
- Kim, T., Son, W., & Jeon, S. (2021). Mining Loot Box News: Analysis of Keyword Similarities Using Word2Vec. *Journal of Information Technology Services*, 20(2), 77-90. <https://doi.org/10.9716/KITS.2021.20.2.077>
- Kyu-Dong, P., Ju, S.-D., & Hwang, M.-C. (2022). A Study on Ways to Improve Management of the Integrated Culture Voucher Program using Big Data Analysis. *Journal of Digital Art Engineering and Multimedia*, 9(2), 201-214. <http://dx.doi.org/10.29056/jdaem.2022.06.08>
- Lee, J. W., Kim, Y., & Han, D. H. (2022). LDA-based topic modeling for COVID-19-related sports research trends. *Frontiers in Psychology*, 13, 1033872. <https://doi.org/10.3389/fpsyg.2022.1033872>
- Lee, J. Y. (2021). A Study on Metaverse Hype for Sustainable Growth. *International Journal of Advanced Smart Convergence*, 10(3), 72-80. <https://doi.org/10.7236/IJASC.2021.10.3.72>
- Lee, S.-J., & Youk, E.-H. (2021). Content analysis of news coverage on games after the inclusion of gaming disorder in ICD-11. *Journal of Korea Game Society*, 21(3), 91-106. <https://doi.org/10.7583/JKGS.2021.21.3.91>
- Maroufkhani, P., Iranmanesh, M., & Ghobakhloo, M. (2023). Determinants of big data analytics adoption in small and medium-sized enterprises (SMEs). *Industrial Management & Data Systems*, 123(1), 278-301. <https://doi.org/10.1108/IMDS-11-2021-0695>
- Morehead, C. A., O'Hallarn, B., & Shapiro, S. L. (2016). Tell me how you really feel: Analyzing debate, desire, and disinhibition in online sports news stories. *International Journal of Sport Communication*, 9(1), 13-35. <https://doi.org/10.1123/ijsc.2015-0056>
- Onan, A., Korukoglu, S., & Bulut, H. (2016). LDA-based Topic Modelling in Text Sentiment Classification: An Empirical Analysis. *International Journal of Computational Linguistics and Applications*, 7(1), 101-119. <http://www.ijcla.org/2016-1/IJCLA-2016-1-pp-101-119-preprint.pdf>
- Park, J.-H., & Song, M. (2013). A Study on the Research Trends in Library & Information Science in Korea using Topic Modeling. *Journal of the Korean Society for information Management*, 30(1), 7-32. <https://doi.org/10.3743/KOSIM.2013.30.1.007>
- Park, J., & Oh, H.-J. (2017). Comparison of Topic Modeling Methods for Analyzing Research Trends of Archives Management in Korea: focused on LDA and HDP. *Journal of Korean Library and Information Science Society*, 48(4), 235-258. <https://doi.org/10.16981/kliss.48.201712.235>
- Park, Y. M., & Kim, Y.-w. (2020). User Perceptions of a Digital Archive for Content Creation: Focused on the Korea Creative Content Agency (KOCCA)'s 'Culturing'. *Journal of the Korean BIBLIA Society for library and Information Science*, 31(1), 285-310. <https://doi.org/10.14699/kbiblia.2020.31.1.285>
- Pather, S. (2021). *The impact of digital media platforms on sport reporting and audience engagement: A case study of Twitter in South Africa* (Doctoral dissertation, University of Johannesburg (South Africa)). <https://hdl.handle.net/10210/501590>
- Pizzolante, M., Borghesi, F., Sarcinella, E., Bartolotta, S., Salvi, C., Cipresso, P., Gaggioli, A., & Chirico, A. (2023). Awe in the metaverse: Designing and validating a novel online virtual-reality awe-inspiring training. *Computers in Human Behavior*, 148, 107876. <https://doi.org/10.1016/j.chb.2023.107876>

- Pu, H., Kim, J., & Daprano, C. (2021). Can esports substitute traditional sports? The convergence of sports and video gaming during the pandemic and beyond. *Societies*, 11(4), 129. <https://doi.org/10.3390/soc11040129>
- Sharma, C., Sharma, S., & Sakshi. (2022). Latent DIRICHLET allocation (LDA) based information modelling on BLOCKCHAIN technology: a review of trends and research patterns used in integration. *Multimedia Tools and Applications*, 81(25), 36805-36831. <https://doi.org/10.1007/s11042-022-13500-z>
- Sievert, C., & Shirley, K. (2014). LDAvis: A method for visualizing and interpreting topics. In *Proceedings of the Workshop on Interactive Language Learning, Visualization, and Interfaces* (pp. 63-70). Association for Computational Linguistics. <https://doi.org/10.3115/v1/w14-3110>
- Talaoui, Y., Kohtamäki, M., Ranta, M., & Paroutis, S. (2023). Recovering the divide: A review of the big data analytics—strategy relationship. *Long Range Planning*, 56(2), 102290. <https://doi.org/10.1016/j.lrp.2022.102290>
- Uthirapathy, S. E., & Sandanam, D. (2023). Topic Modelling and Opinion Analysis On Climate Change Twitter Data Using LDA And BERT Model. *Procedia Computer Science*, 218, 908-917. <https://doi.org/10.1016/j.procs.2023.01.071>
- Viscione, I., & D'Elia, F. (2019). Augmented reality for learning in distance education: the case of e-sports. *Journal of Physical Education and Sport*, 19, 2047-2050. <https://doi.org/10.7752/jpes.2019.s5304>
- Wui, M.-Y., Na, J. Y., & Park, Y. I. (2018). A study on the Elements of Interest for VR Game Users Using Text Mining and Text Network Analysis-Focused on STEAM User Review Data. *Journal of Korea Game Society*, 18(6), 69-82. <https://doi.org/10.7583/JKGS.2018.18.6.69>
- Xue, Y., Kambhampati, C., Cheng, Y., Mishra, N., Wulandhari, N., & Deutz, P. (2024). A LDA-Based Social Media Data Mining Framework for Plastic Circular Economy. *International Journal of Computational Intelligence Systems*, 17(1), 8. <https://doi.org/10.1007/s44196-023-00375-7>
- Yang, H.-C. (2021). Topic Modeling Analysis of Franchise Research Trends Using LDA Algorithm. *The Korean Journal of Franchise Management*, 12(4), 13-23. <https://doi.org/10.21871/KJFM.2021.12.5.13>
- Youm, D., & Kim, J. (2022). Text Mining Approach to Improve Mobile Role Playing Games Using Users' Reviews. *Applied Sciences*, 12(12), 6243. <https://doi.org/10.3390/app12126243>
- Zhe, C., Huilan, G., Jiang, W., & Zhongyi, H. (2023). The Ideal and Reality of Metaverse: User Perception of VR Products Based on Review Mining. *Data Analysis and Knowledge Discovery*, 7(1), 49-62. <https://doi.org/10.11925/infotech.2096-3467.2022.0371>
- Zhu, L., Yu, F. R., Wang, Y., Ning, B., & Tang, T. (2018). Big data analytics in intelligent transportation systems: A survey. *IEEE Transactions on Intelligent Transportation Systems*, 20(1), 383-398. <https://doi.org/10.1109/TITS.2018.2815678>