NAVIGATING DIGITAL MEDIA'S ROLE IN HOME-BASED SPORTS ACTIVITIES AMID THE PANDEMIC: A SENTIMENT AND TOPIC MODELING APPLICATION

Rolando R. Codilan,
Eastern Samar State University, Borongan City, Philippines
Correspondence: rcodilanessu@gmail.com

ABSTRACT: *The global outbreak of COVID-19 led to an abrupt crisis in sports, causing a shift from traditional to home-based sports activities. This study explores the sentiments expressed on digital media regarding home-based sports events during the pandemic, utilizing sentiment analysis and Latent Dirichlet Allocation (LDA) on the top ten online publications. The analysis reveals prevalent themes, including challenges in the transition to online competition, difficulties in digital training, mental and physical health issues, concerns about poor internet connection, and the risk of future injuries when athletes return to play. The sentiment analysis indicates a mix of positive and negative attitudes among athletes, with a leaning towards negativity. The study underscores the need for sports organizations to adapt to digitalization and emphasizes the role of eSports during the pandemic.*

Keywords: Content analysis, opinion mining, text mining, Digital Media, Home-Based Sports Events

1. INTRODUCTION

The abrupt and unexpected global outbreak of COVID-19 triggered a major crisis in sports [1]. Technology changes the lives of people, from the simple to the digital community [2]. The pandemic led to a near halting of global sports activities, compelling a paradigm shift towards home-based sports through the utilization of the Internet and virtual modalities. While sports will undoubtedly persist as a vital societal component, the pandemic is anticipated to leave an enduring imprint on the management of sports organizations. Digitalization is set to assume a pivotal role, necessitating sports organizations to embrace flexibility, adaptability, and resilience for recovery from this crisis [3]. This transformation imposes a shift to a home-based format, replacing traditional face-to-face competition. The absence of live traditional sports combined with COVID-19 strict stay-at-home policies has resulted in more children becoming engaged in eSports, offering alternative avenues for physical activity. Children with no access to mainstream sports lose the benefits of physical activity commonly associated with traditional sports. However, eSport has been found to have social, motivational, emotional, and cognitive benefits [4]. Moreover, during the halt of various sports activities, eSports has filled the void with broadcast programming on television and online, professional athletes partaking in eSports tournaments and activities, and even wagering on eSports events, which in turn has also drawn advertisers to promote their products and services through these activities [5]. This study addresses the lack of research on sentiments related to digital media's role in home-based sports events during the pandemic by examining the top ten published documents. Employing Topic Modeling, Latent Dirichlet Allocation, and content analysis, the research aims to extract and comprehend perspectives on home-based sports activities. Leveraging semantic mining and latent finding in documents, these methodologies prove valuable in natural language processing [6]. LDA, with its capacity to analyze extensive text and discern unseen attributes through specific weighting, becomes instrumental in summarizing, clustering, and linking vast datasets [7]. Each LDA-generated topic embodies a theme, offering a unique insight into the sentiments expressed across global internet users concerning digital media's role in home-based sporting activities during the pandemic

2. OBJECTIVES OF THE STUDY

In this study, the researcher examined the top ten online publications focusing on digital media as a tool for home-based sports events during the pandemic. The objective was to delve into various themes, associated discussions, and materials available on the internet, employing Latent Dirichlet Allocation (LDA) as a response mechanism. Specifically, the study aimed to determine the frequently occurring words in the collection of online published documents, utilize LDA to identify the overall sentiment of the documents and uncover the hidden predominant themes revealed by LDA in the published articles.

3. METHODS

3.1 Research Design

This study employs content analysis on online documents, which involves the analysis of existing evidence or information to provide a meaningful assessment of the outcomes. Content analysis involves a thorough evaluation of available data to clarify a complex phenomenon using topic modeling and latent Dirichlet Allocation (LDA). The data was gathered from the top 10 published articles online. The study focused on the point of view and sentiment of the sudden change in sports from face-to-face to digital media as a tool for home-based sports events during the pandemic. This study applied an investigative sequential mixed methods approach where the initial phase starts with data collection using qualitative information and analysis, followed by quantitative data generation to guide the research and determine the order of data collection and analysis with a concluding stage of data analysis. The collected data is then converted to digital format and compiled into a single data set that combines the two outcomes [8].

3.2 Data Source

Most text data are in unstructured or semi-structured forms, and since these words could not be classified or indexed manually, this study applied a data pre-processing procedure. Data pre-processing first extracts converts, and cleans up the text data. After parsing the content of the document, the document can be retrieved, consolidated, and converted into a "corpus," and then the noise is cleaned up for spaces, punctuation marks, numbers, English letters, and so on [9].

3.3 Data Pre-processing and Cleaning

Pre-processing plays a crucial role in any data-mining process as it directly influences the success of the overall data-mining project [10]. It simplifies the complexity by eliminating extraneous information from the original dataset. Many review datasets tend to have noisy data, including outliers and duplications in the outsourced data, adversely affecting the quality of the outcomes. This study employs Natural Language Processing (NLP) methods to effectively filter out noisy and unnecessary data. Specifically, for language-related tasks such as text summarization and word processing, a key pre-processing step involves removing stop words. In this approach, the elimination of unused or irrelevant data is achieved through stop-word removal [11].

3.4 Stop Words Removal

In natural language processing (NLP), Stop words are commonly used words in a language, such as "the" and "and," that are often removed in natural language processing and text mining. Their removal is essential to reduce computational complexity by lowering the dimensionality of the term space. Additionally, excluding stop words enhances readability, focusing analyses on more meaningful terms and improving the accuracy of tasks like text summarization and sentiment analysis. In essence, removing stop words streamlines text for more effective and accurate natural language processing [12].

3.5 Topic Modeling (Latent Dirichlet Allocation)

Latent Dirichlet Allocation (LDA) is a generative probabilistic model intricately designed for corpus analysis. In topic modeling, LDA is a key method that represents documents as combinations of hidden topics. Each topic is characterized by a distribution over words, with LDA representing these topics through word probabilities. The words with the highest probabilities within each topic serve as insightful indicators of the topic's essence [13, 14]. This advanced textual analysis technique delves into the statistical correlations among words within an extensive document set, unraveling and quantifying the underlying topics [15]. The fundamental concept of LDA involves treating documents as combinations of words, each document as a fusion of multiple topics, and each topic as a combination of several words. The advantages of LDA include its characteristics of supervised learning, flexible extension capabilities, significantly improved computational speed, and recognized effectiveness and value [16].

For each document w in a corpus D, LDA assumes the following generative process:

Choose N ~ Poisson(ξ)

- Choose θ ~ Dir(α)
- For each of the N words wn
- (a) Choose a topic zn ~ Multinomial(θ)
- (b) Choose a word wn from p(wn |zn,β), a multinomial probability conditioned on the topic zn.

3.5 Document

A document is usually a bunch of words, terms (which are just multiple words), symbols, diagrams, or tables all grouped under a title that represents it. Examples of documents include

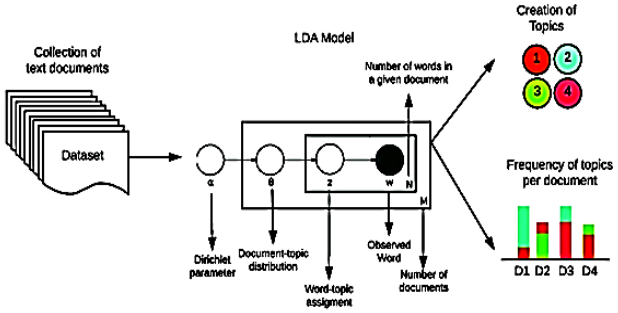

Figure 1. Latent Dirichlet Allocation Model [22]

research papers and news articles when we talk about topic modeling. Symbolically, if we have a document called "d," we can think of all its words as {w1 + w2 + w3 +….+ wn}, where each "w" represents a word in that document. The total number of documents is represented by D={w1,w2,…,wn}, with w1 being the words in a specific document [17]. The technique used here, called LDA, is applied to the text data. It works by breaking down the big matrix of all the words in all the documents into two smaller matrices: one for the document topics and one for the topic words. So, LDA, similar to PCA, is a method that breaks down matrices [18]. Let's consider a corpus of documents with words like "freekick," "dunk," "rebound," "foul," "shoot," "NBA," and "Liverpool.":

3.6 Tokenization

Tokenization serves as a method to break down a piece of text into smaller units known as tokens, which can encompass words, characters, or subwords. Aravindpai [19] categorizes tokenization into three main types: word, character, and subword (n-gram characters) tokenization. For instance, consider the sentence "This movie is good." The most common tokenization method involves using spaces as delimiters. Using space as a delimiter, the sentence is tokenized into four units: "This movie is good." In this case, where each token represents a word, it exemplifies Word tokenization.

3.7 Bag of words, sentences, and documents assumptions

The "Bag of Words" assumption signifies the disregard for the word order within documents, focusing instead on the frequency of word occurrences. The model prioritizes the number of times words appear in a document rather than their sequence. Analogously, the "Bag of Sentences" assumption dismisses the order of sentences within documents, while the "Bag of Documents" assumption extends this concept to ignore the order of documents in the corpus [17].

3.8 Topics

Distinct communities often offer varied interpretations of topics. In this paper, topics are characterized as concealed patterns or concise representations of documents within a text corpus. Ali Daud [17] technically defined topics as "semantically related probabilistic clusters of words," serving as a connecting link between words and entities (such as documents or authors) to unveil latent associations. Formally, a topic is described as "a probability distribution over words or terms in a vocabulary," while informally, it is conceptualized as "an underlying semantic theme; a document, comprising numerous words, can be succinctly modeled as stemming from a smaller number of topics," [20].

3.9 Topic models and modeling

Topic models operate on the premise that documents can be portrayed as a blend of topics, with each topic represented as a probability distribution over words. The generation of documents follows a straightforward probabilistic procedure within generative models [21]. The overarching method of unearthing latent topics from text corpora through these models is commonly termed topic modeling. Technically, this involves identifying a topic (denoted as 'z') in a document ('d') with a defined probability distribution of words in a vocabulary ('V') using topic models [17]. Latent Dirichlet Allocation (LDA) achieves a dual purpose: pinpointing topics within the corpus and assigning these identified subjects to documents in the same corpus. The stepwise process of LDA is succinctly depicted in the accompanying schematic image

3.10 Interpretation and Comprehensive Discussion

To comprehend the data, the researcher meticulously detailed the findings, providing a thorough explanation of the revealed occurrences, supported by relevant literature. The discussion aimed to address the gaps highlighted in the introduction. Recommendations for presenting the analysis in the discussion encompass a comprehensive articulation of findings, elucidating the how, why, and suggested actions. The flow of data in Orange Data Mining, as illustrated in Fig. 2 [23] was followed. The Import Documents widget facilitated the loading of raw data, and subsequent preprocessing and cleaning were conducted using the Preprocess Text widget. Following these steps, topic modeling, sentiment analysis, multidimensional scaling, and a word cloud were generated.

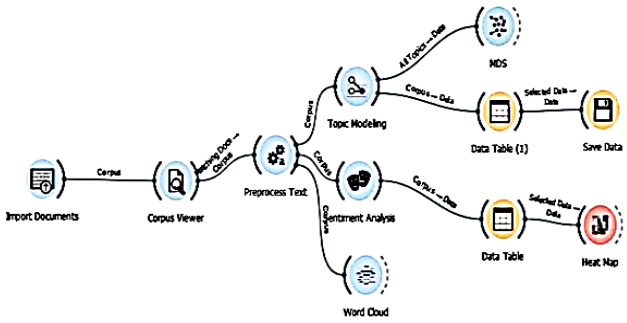

Figure 2. Research Framework

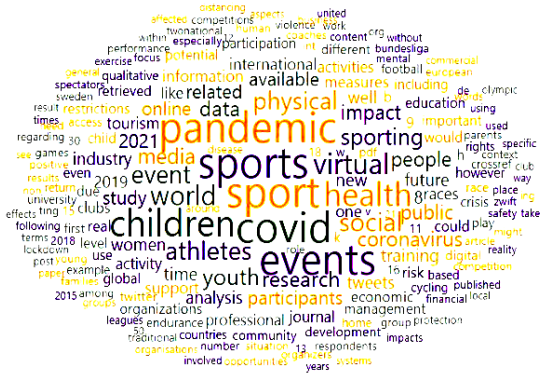

Figure 3. Word Cloud

4. RESULTS

Based on the results, some interesting influences have been established in this study, which are addressed below. With respect to work on sentiment analysis under the Latent Dirichlet Allocation approach, sentiment analysis receives different terms. Within these common terms, the researcher found the most commonly used terms in the top ten published documents were displayed in a word cloud produced from the corpus of papers (Shown in Figure 3). The terms pandemic, sports, children, physical, virtual, health, public, athletes, sporting, coronavirus, event, impact, training, COVID, online, and training were found to be the most frequently used in the corpus of papers, among others. This indicates that athletes have difficulties

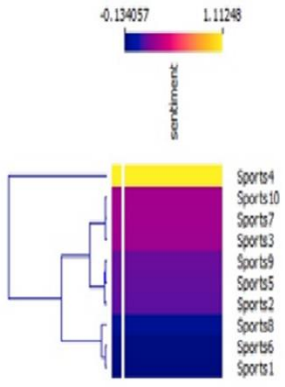

Figure 4. Heat Map

in home-based training and the use of digital media as a tool for home-based sports events during a pandemic.

The corpus was imported into a sentiment analysis widget that analyzes each document's sentiment using Latent Dirichlet Allocation techniques. Table 1 was found to have positive records for sports10, sports2, sports3, sports4, sports5, sports7, and sports9. This shows that during the pandemic, the athletes referenced in the documents had a positive attitude toward digital media as a medium for home-based sporting events. Based on the compound value of their attitudes, Sports1, Sports6, and Sports8 were classified as negative documents, as shown in Table 1. This would entail the athletes employing several derogatory terms, which may be explained by their struggles with digital media as a tool for home-based sporting activities during the epidemic. The results of the word cloud in Figure 3 agree with this information.

To visualize the sentiment analysis results, the sentiment analysis results were imported into a heat map widget. Each page's thoughts are represented by a set of colors on the heat map. The color blue represents negative feelings, whereas the color yellow represents positive emotions. Figure 4 illustrates that the documents lean toward the color blue, meaning that the majority of the athletes used negative phrases in their responses, resulting in the documents' compound value being negative.

Table 1. Documents with positive and negative sentiments analysis result

	name	path	content	sentiment
title	True			
1	Sports1	C:/Users/L2WS40User/Desktop/Top 10 Sports\Sports1.pdf	Article Theand ImpactofCOVID-19on...	-0.134057
2	Sports10	C:/Users/L2WS40User/Desktop/Top 10 Sports\Sports10.pdf	Journal of Global Sport Management...	0.268007
3	Sports2	C:/Users/L2WS40User/Desktop/Top 10 Sports\Sports2.pdf	The impact of COVID-19 on sport, p...	0.120919
4	Sports3	C:/Users/L2WS40User/Desktop/Top 10 Sports\Sports3.pdf	Elliottetal.BMCPublicHealthhttps://d...	0.257295
5	Sports4	C:/Users/L2WS40User/Desktop/Top 10 Sports\Sports4.pdf	Sporting events during the COVID-1...	1.11248
6	Sports5	C:/Users/L2WS40User/Desktop/Top 10 Sports\Sports5.pdf	International sporting events during ...	0.132979
7	Sports6	C:/Users/L2WS40User/Desktop/Top 10 Sports\Sports6.pdf	An Overview of the Sport-Related Im...	-0.128558
8	Sports7	C:/Users/L2WS40User/Desktop/Top 10 Sports\Sports7.pdf	European Journal of Information Sys...	0.262727
9	Sports8	C:/Users/L2WS40User/Desktop/Top 10 Sports\Sports8.pdf	Ann Appl Sport Sci 9(1): e964, 2021. ...	-0.073758
10	Sports9	C:/Users/L2WS40User/Desktop/Top 10 Sports\Sports9.pdf	GETTING BACK TO THE EVENT: COV...	0.142045

Latent Themes

The latent topics within the corpus of documents were loaded into a topic modeling widget that used latent semantic indexing to identify them. The model, as illustrated in Figure 10, was programmed to generate five (5) topics, each with ten (10) words. Latent themes for the created subjects were created based on the researcher's theoretical stance (shown in Table 2).

Latent Topic 1

It discusses the sudden transition from traditional sports to online competition has presented various challenges for both coaches and athletes. The coronavirus pandemic has prompted many institutions and athletes to switch to virtual training, which has been an adjustment for those who are used to in-person training. Numerous facets of life, particularly sports, have been altered by this catastrophe. Some of the problems encountered may be personal, such as nervousness related to technology use or stepping outside one's comfort zone.

Latent Topic 2

Difficulties and challenges in the conduct of digital training and Learning. The COVID-19 pandemic response resulted in an unprecedented situation within the sporting community. Resultantly, athletes were forced to abruptly modify training habits without adequate time for coaches, strength, and conditioning professionals, and athletes to develop a structured strategy moving forward. Previous commentaries [24] and consensus statements have been recently published regarding concerns for athletes returning to sport after a disruption in regular training or recommendations for in-home training [25, 26].

Latent Topic 3

Talks about the Mental and Physical Health issues related to the use of digital Media as a Tool for Home-Based Sports. In 2020, the sports world was halted by a global pandemic that resulted in a stoppage of most athletic competitions. The Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) responsible for the respiratory condition known more commonly as COVID-19 can result in pneumonia-like symptoms and may elicit an inflammatory cascade impacting the bronchioles, alveoli, cardiovascular system, and myocardial tissue [27]. Officials at the national and municipal levels in the Philippines put in place several limitations and safeguards to stop community transmission and "flatten the curve." These preventive measures included social withdrawal, bans on public gatherings, and stay-at-home directives for various Philippine cities and municipalities, which led to the closure of the majority of non-essential enterprises. The majority of sports training facilities and neighborhood fitness centers were closed as a result, denying athletes access to equipment for strength and conditioning. As a result, athletes were forced to drastically alter their training routines and surroundings without enough time to buy the right training gear.

Latent Topic 4

Concerns the poor internet connection concerning the use of digital Media as a Tool for Home-Based Sports. When faced with significant tasks, an athlete who is unable to complete training and exercises feels anxious and unmotivated. Additionally, it was found that their learning had been impeded by a lack of suitable home-based sports training materials and a sluggish or unstable internet connection. Athletes are prone to fear, which causes them to provide an incomplete variety of their assigned task. Home-based Sports environments can generate a feeling of anonymity to athletes which makes it easier for athletes to withdraw or participate minimally or completely disappear from the Sports training. These views suggested that athletes learning sports at home experienced anxiety, which prevented them from playing. Athletes were demotivated to attend practice because they were spending too much time preparing activities for home-based sports and not enough time coaching.

Latent topic 5

Risk of future injuries when athletes return to play. Many of the athletes were forced to finish their training on their own as a result of the COVID-19 shutdown procedures, probably in unusual settings with different resources. Few athletes in the study reported having access to sports facilities but nearly half of the athletes said they had access to resistance bands and dumbbell weights. Strength personnel were likely faced with multiple difficulties when attempting to create effective training programs with any level of coherence or systematic periodization scheme given that there was no degree of uniformity across athletes. Table 2. Latent Themes of the Topics generated through LDA

Table 2. Latent Themes of the Topics generated through LDA

1. The sudden transition from traditional sports to online competition has presented various challenges for both coaches and athletes. Abrupt change from conventional to digital Media as a Tool for Home-Based Sports
2. Difficulties and challenges in the conduct of digital training and Learning
3. Mental and Physical Health issues related to the use of digital Media as a Tool for Home-Based Sports
4. Poor internet connection concerning the use of digital Media as a Tool for Home-Based Sports.
5. Risk of future injuries when athletes return to play

An athlete's impression of and willingness to train may be influenced by the interruption of regular sporting events and the prohibition of in-person activities taken together. Another major concern stemming from disruptions in proper physical training for sports is the increased risk of future injuries when athletes return to play [28]. Furthermore, a lack of an adequate preparatory strength and conditioning period likely resulted in a detraining effect, thus predisposing athletes to the greater risk of soft-tissue injuries during explosive activities [28].

5. CONCLUSIONS

The global impact of the COVID-19 pandemic prompted an unprecedented shift in sports activities, necessitating a transition from traditional to home-based formats. The study focused on analyzing sentiments expressed in the top ten online publications related to digital media's role in home-based sports events during the pandemic. Through employing Latent Dirichlet Allocation (LDA), the researchers aimed to uncover prevalent themes, determine frequently occurring words, and identify the overall sentiment within the collection of documents. The findings revealed a spectrum of sentiments among athletes, with a leaning towards negativity, indicating the challenges associated with digital media as a tool for home-based sports activities. The latent themes identified encompassed issues such as the abrupt transition to online competition, challenges in digital training, mental and physical health concerns, poor internet connectivity, and the potential risk of injuries upon athletes' return to play. As sports organizations grapple with the aftermath of the pandemic, these insights underscore the importance of adaptability, resilience, and strategic planning in navigating the evolving landscape of digitalized sports. Moreover, the study sheds light on the pivotal role of eSports in filling the void left by traditional sports, signaling a broader shift in the sports industry.

REFERENCES

[1] Alam, M. M., & Abdurraheem, I. I. (2023). COVID-19 and the financial crisis in the sports sector around the world. Sport in Society, 26(1), 154-167.

[2] Fibrianto, A. S., & Yuniar, A. D. (2019, June). Technological development and its impact on community social behavior. In 1st International Conference on Social Knowledge Sciences and Education (ICSKSE 2018) (pp. 210-213). Atlantis Press.

[3] Sara Keshkar et.al (2021). The Effects of Coronavirus Pandemic on the Sports Industry: An Update. pp 4, 10. Available at: http://www.aassjournal.com

[4] Kovess Masfety et al. (2016). Is time spent playing video games associated with mental health, cognitive and social skills in young children? Social Psychiatry and Psychiatric Epidemiol 51, pp. 349–357.

[5] Adgate B. (2020) Esports is filling the programming void. Forbes. Retrieved from:https://www.forbes.com/sites/bradadgate/2020/04/21/esports-is-filling-the-programming-void/#29439bd3533b.

[6] Hamed Jelodar et.al (2017). Latent Dirichlet Allocation (LDA) and Topic modeling: models, applications, a survey

[7] Hindle, A., Campbell, J. C, Hindle, A., Stroulia, A. (2014). Latent Dirichlet Allocation: Extracting Topic. The Art and Science of Analyzing Software Data. 139-159.

[8] K. Isoaho, D. Gritsenko, and E. Mäkelä, (2021)"Topic modeling and text analysis for qualitative policy research," Policy Studies Journal, vol. 49, no. 1, pp. 300-324.

[9] I-Cheng Chang et al.,2021 Applying Text Mining, Clustering Analysis, and Latent Dirichlet Allocation Techniques for Topic Classification of Environmental Education Journals.

[10] Alnawas, Anwar, and Nursal ARICI (2021). "Effect of word embedding variable parameters on Arabic sentiment analysis performance

[11] Geler, Zoltan et.al(2021). "Sentiment prediction based on analysis of customers assessments in food serving businesses." Connection Science, pp: 1-19

[12] Fan, Y., Arora, C., & Treude, C. (2023). Stop Words for Processing Software Engineering Documents: Do they Matter?. arXiv preprint arXiv:2303.10439.

[13] Blei D M, McAuliffe J. Supervised topic models. In: Advances inNeural Information Processing Systems (NIPS) 21. Cambridge, MA, MIT Press, 2007, 121–128

[14] Casillano, N. F. B. (2022). Discovering Sentiments and Latent Themes in the Views of Faculty Members towards the Shift from Conventional to Online Teaching Using VADER and Latent Dirichlet Allocation. International Journal of Information and Education Technology, 12(4).

[15] Hamed Jelodar et.al (2017). Latent Dirichlet Allocation (LDA) and Topic modeling: models, applications, a survey.

[16] Fernandes, S.; Bala, J. (2013). Performance analysis of PCA-based and LDA-based algorithms for face recognition. Int. J. Signal. ProcessSyst. 1, 1–6.

[17] Ali Daud et.al (2020). Knowledge discovery through directed probabilistic topic models: a survey

[18] Vaibhav Jayaswal (2020). Text Vectorization: Bag of words (Bow) how to convert text features into vectors

[19] Aravindpai (2020) What is Tokenization in NLP? Here's All You Need To KnowBendle and Wang (2016). Uncovering the message from the mess of big data

[20] Blei D M, McAuliffe J. (2007).Supervised topic models. In: Advances in Neural Information Processing Systems 121–128

[21] Steyvers M, Griffiths T(2017). Probabilistic topic models.

[22] Seth, Neha (2021). https://www.analyticsvidhya.com/blog/2021/06/part-2-topic-modeling-and-latent-dirichlet-allocation-lda-using-gensim-and-sklearn/

[23] Demsar, T. Curk,,et.al (2013). "orange: data mining toolbox in python," Journal of Machine Learning Research, vol. 14, pp. 2349−2353

[24] Sarto, F., Impellizzeri, F. M., Sporri, J., Porcelli, S., Olmo, J., Requena, B., et al. (2020). Impact of potential physiological changes due to COVID-19 home confinement on athlete health protection in elite sports: a call for awareness in sports programming. Sports Med. 50, 1417–1419. doi: 10.1007/s40279-020-01297-6

[25] Hammami, A., Harrabi, B., Mohr, M., & Krustrup, P. (2020). Physical activity and coronavirus disease 2019 (COVID-19): specific recommendations for home-based physical training, Managing Sport and Leisure, 10. https://doi.org/10.1080/23750472.2020.1757494

[26] Jukic, I., Calleja-González, J., Cos, F., Cuzzolin, F., Olmo, J., Terrados, N., Njaradi, N., Sassi, R., Requena, B., Milanovic, L., Krakan, I., Chatzichristos, K., & Alcaraz, P. E. (2020). Strategies and Solutions for Team Sports Athletes in Isolation due to COVID-19. Sports (Basel, Switzerland), 8(4), 56. https://doi.org/10.3390/sports8040056

[27] Buscarini, E., Manfredi, G., Brambilla, G., Menozzi, F., Londoni, C., Alicante, S., et al. (2020). GI symptoms as early signs of COVID-19 in hospitalised Italian patients. Gut. 69, 1547–1548. doi: 10.1136/gutjnl-2020-321434

[28] Myer, G. D., Faigenbaum, A. D., Cherny, C. E., Heidt, R. S. Jr., and Hewett, T. E. (2011). Did the NFL Lockout expose the Achilles heel of competitive sports? J. Orthop. Sports Phys. Ther. 41, 702–705. doi: 10.2519/jospt.2011.0107