MAKÜ
BURDUR MEHMET AKİF ERSOY ÜNİVERSİTESİ

OPEN ACCESS

Burdur Mehmet Akif Ersoy University Journal of Sport Sciences, 2024, Volume :2 Number:1

Exploring the Evolving Perception of Sports Doping through Text Analytics: Figuring out Public Sentiments

Sukumaran C[1] *B-C,* , *Ahmet Yavuz KARAFİL*[2] *D* , *Marimuthu K*[3] *E* ,
Sugumar Chokkalingam[1] *A,*
[1]*Government Law College,* [2]*Burdur Mehmet Akif Ersoy University,* [3]*Tezpur University,* [4]*Gandhigram Rural Institute*

Author Contribution Rate:
A) Study Design
B) Data Collection
C) Data Analysis
D) Article Writing
E) Critical Reading

ABSTRACT

This study employs text analytics to investigate the evolving public perception of sports doping, utilizing natural language processing techniques such as sentiment analysis and Latent Dirichlet Allocation (LDA) topic modeling. The data, comprising 64,725 English tweets collected from January 2018 to December 2022 using the snscrape Python library, underwent refinement through NLP algorithms, including the removal of stop words, punctuation, digits, and hyperlinks. Stemming and lemmatization techniques were applied for text uniformity and structure enhancement. The sentiment analysis revealed dynamic shifts in neutral, pessimistic, and optimistic sentiments across different years, indicating changing public attitudes toward sports doping. Visual representations through figures and tables enhance comprehension of the sentiment distribution trends. LDA topic modeling identified critical themes in the sports doping discourse, encompassing anti-doping regulations, financial impacts on sports, marijuana use, women's participation, dope screening procedures, ethical considerations, anti-doping organizations, specific doping instances, and consequences of doping bans. The research illuminates a nuanced understanding of the intricate issue by presenting these findings visually. The study highlights a notable shift in public attitudes towards sports doping. It underscores the dynamic nature of opinions, emphasizing the importance of continuous observation and comprehension of factors influencing these fluctuations. The LDA analysis provides valuable insights into various dimensions of the sports doping debate, contributing to a more comprehensive understanding of this complex and evolving issue.

Keywords: Sports Doping, Sentiment Analysis, Text Analytics, Topic Modeling, Latent Dirichlet Allocation.

Sporda Dopinge İlişkin Gelişen Algının Metin Analitiği Yoluyla Keşfedilmesi: Kamuoyunun Düşüncelerini Anlamak

ÖZET

Bu çalışma, duygu analizi ve Latent Dirichlet Allocation (LDA) konu modellemesi gibi doğal dil işleme tekniklerini kullanarak sporda doping konusunda kamuoyunda gelişen algıyı araştırmak için metin analitiği kullanmaktadır. Ocak 2018'den Aralık 2022'ye kadar snscrape Python kütüphanesi kullanılarak toplanan 64.725 İngilizce tweetten oluşan veriler, durak kelimelerin, noktalama işaretlerinin, rakamların ve köprülerin kaldırılması da dahil olmak üzere NLP algoritmaları aracılığıyla iyileştirilmiştir. Metin bütünlüğü ve yapısının iyileştirilmesi için stemming ve lemmatization teknikleri uygulanmıştır. Duygu analizi, farklı yıllar arasında nötr, kötümser ve iyimser duygulardaki dinamik değişimleri ortaya çıkararak, sporda dopinge karşı değişen kamu tutumlarını göstermiştir. Şekiller ve tablolar aracılığıyla görsel temsiller, duygu dağılım eğilimlerinin anlaşılmasını artırmaktadır. LDA konu modellemesi, sporda doping söyleminde dopingle mücadele düzenlemeleri, spor üzerindeki mali etkiler, marihuana kullanımı, kadınların katılımı, doping tarama prosedürleri, etik hususlar, dopingle mücadele kuruluşları, belirli doping vakaları ve doping yasaklarının sonuçlarını kapsayan kritik temaları belirlemiştir. Araştırma, bu bulguları görsel olarak sunarak karmaşık bir konunun incelikli bir şekilde anlaşılmasına ışık tutmaktadır. Çalışma, sporda dopinge yönelik kamu tutumlarında kayda değer bir değişim olduğunu vurgulamaktadır. Görüşlerin dinamik doğasının altını çizerek, bu dalgalanmaları etkileyen faktörlerin sürekli gözlemlenmesinin ve anlaşılmasının önemini vurgulamaktadır. LDA analizi, sporda doping tartışmasının çeşitli boyutlarına ilişkin değerli içgörüler sunarak bu karmaşık ve gelişen konunun daha kapsamlı bir şekilde anlaşılmasına katkıda bulunmaktadır.

Anahtar Kelimeler: Sporda Doping, Duygu Analizi, Metin Analitiği, Konu Modelleme.

E-ISSN: 3023-5359

Doi Number:
doi.org/10.5281/zenodo.10899883

Introduction

The issue of doping in elite and professional sports remains a pressing concern as it erodes public trust and compromises the integrity of sporting outcomes **(**De Hon et al., 2015; Heuberger et al., 2022; Sefiha & Reichman, 2017**)**. When athletes misuse performance-enhancing drugs (PEDs), they risk their health by using prohibited substances such as anabolic beta-blockers, hormones, glucocorticoids, metabolic stimulants, modulators, anabolic agents, beta-2 agonists, unapproved substances, diuretics, narcotics, masking agents, cannabinoids, growth factors, mimetics, peptide hormones, and substances its related compounds (Reardon & Creado, 2014). These substances can cause severe harm to their cardiovascular and liver systems and increase their chances of developing cancer, diabetes, and stroke (Perishable, 2019). Over the past 15 years, the Anti-Doping Administrative Management System (ADAMS) has recorded a yearly average rise of 5.81% in the number of samples tested for anti-doping measures, encompassing Olympic and non-Olympic sports (Abeza & Sanderson, 2022). In response to doping issues in 2000, the World Anti-Doping Agency (WADA) was established to tackle the problem. The organization introduced the World Anti-Doping Code (WADC) in 2001 and has since revised it several times, including in 2020 (Pöppel, 2021).

It is vital to not only focus on the regulations set forth by WADA and the WADC regarding doping but also consider the impact of this issue on public discourse and attitudes. Social media, particularly Twitter, helps express opinions and uncover underlying sentiments regarding sports doping (Abeza & Sanderson, 2022). Analyzing sports doping-related tweets from 2018 to 2022 using natural language processing and machine learning techniques can offer a comprehensive understanding of the multifaceted impact of sports doping on the sporting realm, encompassing public reactions and media coverage.

The objective of this research is to answer two crucial questions: firstly, to investigate the changes in the general public's attitude towards sports doping, and secondly, to identify the key themes and topics that arise in conversations about sports doping. By addressing this research gap, our study aims to enhance the comprehension of the consequences of sports doping and offer helpful guidance to WADA and policymakers to implement effective measures against doping.

Method

Data collection and pre-processing

To extract tweets from Twitter, this study utilized the snscrape Python library, designed explicitly for this task (Subramanian, 2021). Our data collection primarily targeted Twitter, a popular social media platform boasting approximately 450 million active users in 2023 (Xu, 2023). The tweets were scraped, and the duplicate tweets were removed with 64725 unique English tweets using the keyword “Sports doping” from 1 January 2018 to 31 December 2022.

Various natural language processing algorithms were employed during the pre-processing stage to refine the data. The primary objective was to eliminate irrelevant elements, such as stop words, punctuation, digits, and hyperlinks. Stop words, commonly used English words that do not significantly affect the meaning of a sentence, were specifically targeted for removal. Specific terms, like "a," "an," and "the," were removed during data cleaning to improve the accuracy of the analysis. Once the erroneous bits were removed, two preliminary processing methods, stemming and lemmatization, were applied to standardize and organize the text for further research. These approaches aim to simplify the data, increase its consistency and quality, and generate

significant insights by combining various word combinations into their root words using lemmatization and removing word suffixes to reveal a word's core form using stemming. By performing these crucial procedures, the complexity of the data was reduced, making it easier to analyze and generate valuable insights (Praveen et al., 2021; SV & Ittamalla, 2022).

Research Methodology

Study 1 Sentiment analysis

When evaluating data, sentiment analysis is crucial in determining the sentiment score of a specific word, phrase, or section of a document. This automated approach enables accurate retrieval and examination of subjective evaluations regarding different aspects of an object or entity (Birjali et al., 2021; Wankhade et al., 2022).

The study aimed to gain insight into doping followers' intense and emotional reactions to sports. To do this, sentiment analysis was utilized to generate the sentiment rating of the data studied, exposing information about fans' attitudes concerning doping in sports.

To establish the general tone of the tweets, we utilized the TextBlob tool for sentiment analysis. This method assigned a sentiment score to each tweet based on the polarity of its message. Using the TextBlob library, a natural language processing tool, we analyzed the sentiment conveyed in the text, which helped us interpret the overall feeling (Singh et al., 2021).

The sentiment ratings were then allocated to categorize the tweets into three unique classes: positive, negative, or neutral (Vernikou et al., 2022). The study's main goal was to better understand the changing patterns in sports doping sentiments from 2012 to 2018 by carefully examining the temporal changes in the frequency of these categories.

Study 2 Latent Dirichlet Allocation

We employed Latent Dirichlet Allocation (LDA) topic modeling in the second part of our study to learn more about the critical arguments that ordinary people bring up while discussing doping in sports on internet forums. Topic modeling is a set of algorithms that condense large amounts of text by mechanically discovering hidden topics and subjects throughout a given dataset. LDA assumes that documents in the corpus consist of various latent themes, each of which is a multinomial distribution of words (Farkhod et al., 2021; Nanda et al., 2023; Praveen et al., 2021).

The pre-processed tweets were first transformed into a document-term matrix to execute LDA. This matrix quantifies how often each term occurs in tweets. The LDA model underwent training on the above matrix to detect latent themes in the text. The coherence score of the model was utilized to determine the appropriate number of topics to be identified (Blei et al., 2003).

The topics were established first, and each tweet was assigned a likelihood score for every case. This enabled them to identify the significant issues usually associated with doping in sports. LDA was a powerful tool for assessing enormous volumes of text data and extracting key themes and subjects. This study provided important information into the mindsets and views about doping in sports.

Findings

Study 1: Sentiment Analysis

In Table 1, you can see the results of the sentiment analysis. It shows the number of tweets collected each year and the corresponding percentages of neutral, negative, and optimistic sentiments. These sentiments were determined by rating specific tweets during the analysis process. We carefully studied the content of the text to form the sentiment categories, which are based on either

positive, negative, or neutral emotions (Praveen et al., 2021).

The provided information unveils intriguing insights into the evolution of attitudes throughout time. In 2018, out of 16,234 tweets, 43.70% were classified as neutral, 19.30% as unfavorable, and 37.03% as favorable. In 2019, 41.40% of the 13,514 tweets were neutral, 19.24% were negative, and 39.40% were positive. This trend continued in subsequent years, with varying proportions of attitudes observed.

After analyzing data over time, it is evident that attitudes toward doping in sports have shifted. The percentage of neutral opinions increased from 35.70% in 2022 to 43.70% in 2018, indicating changes in the public's perspective. Negative sentiments varied from 18.50% in 2020 to 23.10% in 2022, depending on the negativity or criticism expressed in tweets. Happy emotions also fluctuated from 37.03% in 2018 to 41.24% in 2022, demonstrating users' varying levels of optimism or support. The analysis findings are visually depicted in Figures 1 and 2.

Results

Table 1. Sentiment analysis

Year	Total Tweets	Neutral (%)	Negative (%)	Positive (%)
2018	16234	7090 (43.70)	3133 (19.30)	6011 (37.03)
2019	13514	5594 (41.40)	2600 (19.24)	5320 (39.40)
2020	10694	4509 (42.20)	1978 (18.50)	4207 (39.34)
2021	10109	3821 (37.80)	2231 (22.11)	4057 (40.13)
2022	14174	5060 (35.70)	3268 (23.10)	5846 (41.24)
Grand Total	**64725**	**26074**	**13210**	**25441**

Note: The primary column represents the total tweet count per year. Each value indicates the proportion and percentage of positive, negative, and neutral tweets relative to the overall count across all years. For example, there were 7,090 neutral tweets, accounting for 43.70% of the total neutral tweets (n = 26,074) collected in the entire dataset in 2018.

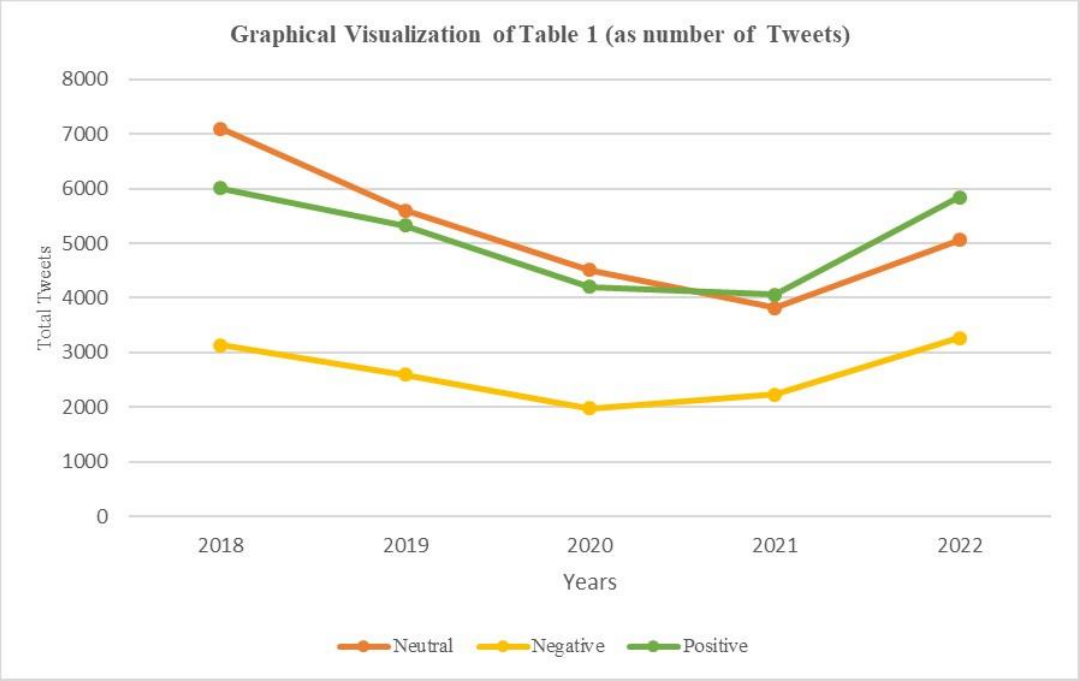

Figure 1 Displays the number of tweets with positive, negative, or neutral sentiments for each year over five years based on Table 1.

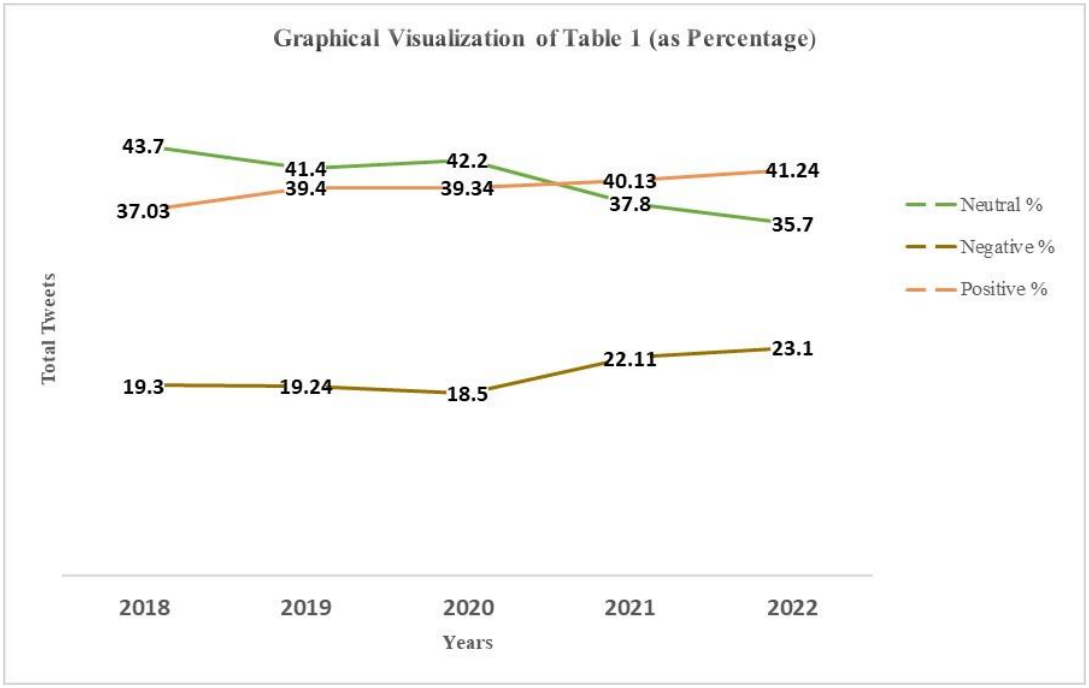

Figure 2 Exposing the percentage of positive, negative, and neutral tweets yearly over five years.

Table 2. Latent Dirichlet Allocation Topic Modeling

Topic	Top Words
Anti-doping policies	sport, clean, dope, athlete, amp (Amphetamine), anti-dope
The financial impact of sport	financial, league, club, wash, city, adult
Sharron's Context	Sharron, Ukraine, context, marijuana, fall, propaganda
Women and sports doping	sport, dope, woman, like, athlete, compete, don't
Dope testing in sport	sport, dope, test, athlete, anti-dope, drug
Gender and doping ethical issues	female, male, category, law, testosterone, person, illegal
Anti-doping agency	anti-dope, sport, WADA, agency, world, intern
Doping in a specific sport	skate, figure, race, cycle, Armstrong, Lance
Doping bans and consequences	dope, sport, ban, year, Olympic, medal
Olympic Doping and Legal Issues	Olympic, Russian, sport, dope, court, ban

(Note: The topic labels were assigned through manual analysis, while the LDA model generated the top words associated with each topic.)

Study 2: Topic modeling

Latent Dirichlet Allocation Topic Modelling

The study on Latent Dirichlet Allocation (LDA) has uncovered several topics related to sports doping. The central theme was anti-doping regulations, highlighting the importance of maintaining a fair and clean sporting environment. Another crucial aspect was the financial implications of sports, including leagues, clubs, and economic concerns. Discussions centered on marijuana and its use, specifically in the case of Sharron, which was treated as a separate issue. The topic of women's involvement in sports doping was also explored, examining the link between doping practices and female athletes' participation and competitiveness. Dope screening in sports was also identified as a distinct issue, focusing on the testing processes and protocols used to identify doping substances. Ethics surrounding gender and doping was another significant subject, which included discussions on male and female groups, legal consequences, and testosterone usage. The anti-doping agency topic emphasized the importance of organizations like the World Anti-Doping Agency (WADA) in fighting doping in sports. The prevalence of doping within specific sports, such as skating, figure skating, and cycling, with notable individuals like Lance Armstrong, was also addressed. Lastly, the consequences and prohibitions of doping, particularly in the Olympics, were highlighted, underscoring the significance of bans, legal proceedings, and doping practices.

Discussion and Conclusion

In Study 1: Evolution of Public Sentiments Towards Sports Doping over Time

The effectiveness of anti-doping initiatives and pinpointing areas that need more education and awareness can be assessed through sentiment analysis (Heuberger et al., 2022; Pöppel, 2021). Upon analyzing the sentiment tendencies towards doping in sports, as presented in Table 1, distinct patterns and shifts in public opinion have been identified. Our analysis of individual tweets' sentiment ratings showcased a range of neutral, negative, and positive sentiments that varied over time.

Based on the data, it is evident that there have been noticeable fluctuations in sentiment over the years. To illustrate, in 2018, out of the 16,234 tweets analyzed, 43.70% were deemed impartial, 19.30% were negative, and 37.03% were positive. Similarly, in 2019, 41.40% of the 13,514 tweets had a neutral stance, 19.24% were unfavorable, and 39.40% were positive. Over the years, the percentages have experienced significant alterations.

According to data gathered from public opinion, there has been a significant rise in doping in sports. For instance, the percentage of neutral tweets decreased from 43.70% in 2018 to 35.70% in 2022, indicating an 8% difference. The alterations demonstrate notable changes in the prevailing attitudes of the general public. It's essential to take note of the shift in public sentiment perception as negative posts have increased from 18.50% in 2020 to 23.10% in 2022. Similarly, the percentage of positive posts in the public sentiment has remained relatively stable, ranging from 37.03% in 2018 to 41.24% in 2022.

Various aspects can influence an individual's stance on doping in sports, encompassing personal and societal surroundings, ethical principles, the viewpoints of influential peers, the understanding of equitable competition,

monetary and tangible rewards, the aspiration for exceptional performance, notoriety, and athletic achievements (García-Grimau et al., 2021).

Study 2: Principal Themes and Topics in Discussions on Sports Doping

Sefiha and Reichman previously remarked through topic modeling that social media is vital in deterring doping among young athletes3. Our analysis of discussions surrounding sports doping employed the LDA model. We revealed various themes and subjects related to doping in sports, presented in Table 2, and shed light on its diverse aspects.

One of the prime themes of our discussions is to uphold genuineness in athletics to maintain fair competition by enforcing anti-doping rules and principles (Dougherty & Baron, 2022). Our exploration found the financial components, according to Barget and Chavinier, that contribute to the success of a sports organization. This encompassed various elements such as competition, club operations, and finances (Barget & Chavinier-Rela, 2017). We investigated the theme concerning Sharron Davies, a former Olympian from Plymouth, who has sparked controversy with her recent Twitter comments. In the past, she has likened trans athletes to doping cheats and suggested that their inclusion has led to discrimination against women. These remarks were made in response to an article published by BBC (*Olympian Sharron Davies Reveals Death Threats during Trans Debate*, 2022). Research has unequivocally demonstrated that female athletes may resort to diverse doping techniques owing to gender-specific physiological and psychological differences. This issue demands urgent attention and a comprehensive inquiry (Collomp et al., 2022).

Our thorough analysis jumped deep into the critical and sensitive topic of doping screening in sports that making decisions based on scientific evidence is crucial for accurate and practical outcomes, and anti-doping tests could become more reliable and enhance the effectiveness of anti-doping measures, thus providing a fair and just system for athletes (Nissen-Meyer et al., 2022). Our analysis has centered on the essential matter of ethical considerations in drug testing and gender-specific doping practices, as well as the rigorous adherence to gender classifications in the realm of athletics. Recent reports have highlighted pronounced gender disparities in sports news coverage, with male athletes receiving greater prominence from media outlets (Uyar et al., 2022). Our research also consisted of a transparent exploration of the role and operations of anti-doping organizations, with a particular focus on the globally recognized WADA.

During our discussions, we delved deeply into the topic of doping incidents and scandals that have plagued the sports industry. In the course of our discussion, we touched upon the issue of certain high-profile individuals who were found to have used performance-enhancing substances, Lance Armstrong being one of them. As a former American road racing cyclist, Armstrong was involved in doping, revoking his seven consecutive Tour de France titles and one Olympic medal (Davis & Ryall, 2017). Eventually, the WADA decided to impose a four-year suspension on Russia's participation in global sports events and the Olympics due to doping. While this would have prevented Russia from competing in the 2020 Olympics, some Russian athletes can participate under a neutral flag (Panja, 2019).

Promoting fair play in sports requires

eliminating the abuse of performance enhancing drugs. This will not only reduce negative sentiments associated with it but also ensure a level playing field. Our data indicate a significant increase in negative sentiments due to impact of sports doping occurrences in the competition world, emphasizing the need for its elimination. It is recommended that future studies thoroughly examine all aspects of topic modeling and its relation to evidence-based treatments. By conducting thorough investigations of all the topics, we can increase awareness of programs that prioritize safe and ethical practices in global sports.

References

Abeza, G., & Sanderson, J. (2022). Theory and social media in sport studies. *International Journal of Sport Communication*, *15*(4), 284–292.

Barget, E., & Chavinier-Rela, S. (2017). The analysis of amateur sports clubs funding: A European perspective. *Athens Journal of Sports*, *4*(1), 7–34.

Birjali, M., Kasri, M., & Beni-Hssane, A. (2021). A comprehensive survey on sentiment analysis: Approaches, challenges and trends. *Knowledge-Based Systems*, *226*, 107134.

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of Machine Learning Research*, *3*(Jan), 993–1022.

Collomp, K., Ericsson, M., Bernier, N., & Buisson, C. (2022). Prevalence of prohibited substance use and methods by female athletes: Evidence of gender-related differences. *Frontiers in Sports and Active Living*, *4*, 839976.

Davis, P., & Ryall, E. (2017). Evaluating violent conduct in sport: A hierarchy of vice. *Sport, Ethics and Philosophy*, *11*(2), 207–218.

De Hon, O., Kuipers, H., & Van Bottenburg, M. (2015). Prevalence of doping use in elite sports: A review of numbers and methods. *Sports Medicine*, *45*, 57–69.

Dougherty, J. W., & Baron, D. (2022). Substance Use and Addiction in Athletes: The Case for Neuromodulation and Beyond. *International Journal of Environmental Research and Public Health*, *19*(23), 16082. https://doi.org/10.3390/ijerph192316082

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.

García-Grimau, E., De la Vega, R., De Arce, R., & Casado, A. (2021). Attitudes toward and susceptibility to doping in Spanish elite and national-standard track and field athletes: An examination of the Sport Drug Control Model. *Frontiers in Psychology*, *12*, 679001.

Heuberger, J. A., Henning, A., Cohen, A. F., & Kayser, B. (2022). Dealing with doping. A plea for better science, governance and education. *British Journal of Clinical Pharmacology*, *88*(2), 566–578.

Nanda, G., Jaiswal, A., Castellanos, H., Zhou, Y., Choi, A., & Magana, A. J. (2023). Evaluating the Coverage and Depth of Latent Dirichlet Allocation Topic Model in Comparison with Human Coding of Qualitative Data: The Case of Education Research. *Machine Learning and Knowledge Extraction*, *5*(2), 473–490.

Olympian Sharron Davies reveals death threats during trans debate. (2022). https://nypost.com/2022/06/21/olympian-sharron-davies-reveals-death-threats-during-trans-debate/

Panja, T. (2019). Russia banned from Olympics and global sports for 4 years over doping. *The New York Times*, *9*.

Perishable. (2019, May 1). *Effects of Performance-Enhancing Drugs | USADA*. https://www.usada.org/athletes/substances/effects-of-performance-enhancing-drugs/

Pöppel, K. (2021). Efficient ways to combat doping in a sports education context!? A systematic review on doping prevention measures focusing on young age groups. *Frontiers in Sports and Active Living*, *3*, 673452.

Praveen, S., Ittamalla, R., & Deepak, G. (2021). Analyzing the attitude of Indian citizens towards COVID-19 vaccine–A text analytics study. *Diabetes & Metabolic Syndrome: Clinical Research & Reviews*, *15*(2), 595–599.

Reardon, C. L., & Creado, S. (2014). Drug abuse in athletes. *Substance Abuse and Rehabilitation*, *5*, 95–105. https://doi.org/10.2147/SAR.S53784

Sefiha, O., & Reichman, N. (2017). Social media and the doping of sport surveillance. *Sociology Compass*, *11*(10), e12509.

Singh, M., Jakhar, A. K., & Pandey, S. (2021). Sentiment analysis on the impact of coronavirus in

social life using the BERT model. *Social Network Analysis and Mining*, *11*(1), 33.

Subramanian, D. (2021, August 10). *Scrape Tweets using snscrape and Build a Sentiment Classifier*. Medium. https://pub.towardsai.net/scraping-tweets-using-snscrape-and-building-sentiment-classifier-13811dadd11d

SV, P., & Ittamalla, R. (2022). General public's attitude toward governments implementing digital contact tracing to curb COVID-19–a study based on natural language processing. *International Journal of Pervasive Computing and Communications*, *18*(5), 485–490.

Uyar, Y., Gentile, A., Uyar, H., Erdeveciler, Ö., Sunay, H., Mîndrescu, V., Mujkic, D., & Bianco, A. (2022). Competition, gender equality, and doping in sports in the red queen effect perspective. *Sustainability*, *14*(5), 2490.

Vernikou, S., Lyras, A., & Kanavos, A. (2022). Multiclass sentiment analysis on COVID-19-related tweets using deep learning models. *Neural Computing and Applications*, *34*(22), 19615–19627.

Wankhade, M., Rao, A. C. S., & Kulkarni, C. (2022). A survey on sentiment analysis methods, applications, and challenges. *Artificial Intelligence Review*, *55*(7), 5731–5780.

Xu, Q. (2023). Competing as the First Out Transgender Female Olympian: A Twitter Network Analysis of Laurel Hubbard During the 2020 Tokyo Games. *Communication & Sport*, *11*(5), 854–878. https://doi.org/10.1177/21674795221090422

Corresponding Author Information Author Name Surname: *Sukumaran, C*

Institution: *Government Law College*

Address: *Khajamalai, Thiruchirappalli, Tamil Nadu, India – 620 023*

Orcid Number: *0000-0002-6689-3927*