

# Using NLP to ascertain changes in the fast-fashion industry based on UN sustainable development goals

Jaanya Chadha<sup>1,\*</sup>, Srinidhi Gubba<sup>1,\*</sup>, Ameeta Agrawal<sup>2</sup>

<sup>1</sup> Westview High School, Portland, Oregon

<sup>2</sup> Department of Computer Science, Portland State University, Portland, Oregon

\* Authors contributed equally to this work

## SUMMARY

The fast fashion industry largely contributes to pollution and global emissions, which can be mitigated through strengthening policy. Policy change requires knowledge of the changing conditions of the fast fashion industry that can be evaluated through the sustainability reports published by fast fashion companies. However, due to the reports' variations in details and lengths, it is difficult to find trends that promote the United Nations Sustainable Development Goals (UN SDGs), such as "Responsible Consumption and Production" and "Climate Action". Therefore, we ask if fast fashion companies have changed their sustainability focuses in terms of the UN SDGs from 2018–2019 to 2020–2021. We hypothesized that the later years would showcase higher similarities towards the UN health-related goals due to COVID-19. After checking the reports' similarities against the keywords of each UN SDG, we found that the similarity percentage of the "Good Health and Well-Being" goal had more than doubled from 2018–2019 to 2020–2021. This is important because it shows how the industry reacts to global changes like a pandemic. We also noticed that the three most frequently addressed goals were the same for both time periods. Specifically, the most dominant topic was "Responsible Consumption and Production," with a similarity score of 15.54% for 2018–2019 and 14.1% for 2020–2021. This suggests that the industry may have some consistent focuses on sustainability over time. With these insights, policymakers can better address what changes the fast fashion industry needs to make.

## INTRODUCTION

Fast fashion, which is the unsustainable practice used by many fashion companies to create a large quantity of clothing at low prices, has taken its toll on the planet in many ways. Textile dyeing itself is the second largest cause of pollution to water (1). Fast fashion is expected to increase global carbon emissions by 50% over the next decade (2). As companies attempt to convert to more sustainable practices, substantial reporting is available to assess their changes and areas of impact (3). The United Nations (UN) is currently working on several initiatives dedicated to making the fashion industry more sustainable (4). One notable project is the UN Alliance for Sustainable Fashion. Sustainability encapsulates many different areas of improvement, including social,

environmental, and economic problems. Part of the objectives of the UN Alliance for Sustainable Fashion is to foster more knowledge and data sharing between the companies and organizations involved (4).

Sustainability reports can play an important role for policymakers, as this is the platform for corporate impacts to be communicated and viewed in a sustainable perspective (5). These reports ranging from 5 to 10 pages, contain information about environmental and human actions being taken by the companies to become more sustainable (6). Many companies have specific sections for goals more associated with the environment and sections for goals more associated with human affairs like wages and equality. They end off with discussing methodology of how they reported their measured actions and impacts. Oftentimes, companies and policymakers use the UN Sustainable Development Goals (SDGs) as a form of assessment for sustainable progress. These goals consist of 17 topics ranging from "Good Health and Well-Being" to "Reduced Inequalities" that are designed to promote worldwide prosperity and peace by 2030 and create a more sustainable future for all (5).

Natural language processing (NLP) is a form of machine learning that is used to understand human language and speech. There are many applications of NLP including analyzing sentiment, transforming speech into text, and translating between different languages. (7). Since NLP can quantify raw text, it can be used to identify common themes and topics, making text analysis more efficient. This is important for analyzing large amounts of textual input such as sustainability reports, government documents, and social media.

In this project, the main NLP technique used was cosine similarity, which quantifies text into vectors to compare them (7). We compared sustainability reports to 20 of the most important words for each UN SDG compiled by the UN Department of Economic and Social Affairs (1). This is important as it helped assess how much each UN SDG was addressed in individual company sustainability reports, which when combined can provide insight into the industry as a whole. Through interviews with people who work on sustainability, policy, and fashion, we concluded that sharing industry knowledge can help policymakers make more informed decisions (8–10). The research in this project examined if fast fashion companies changed their sustainability focuses in terms of the UN SDGs from 2018–2019 to 2020–2021. We utilized these UN SDGs as a form of assessment for this project. After learning about how sustainability reports could provide policy-makers with essential data, we started examining some of the reports and noticed how they were very tedious and long. Many organizations do not have the

time or resources to go through these hundreds of pages of reports. Taking this into account, we researched how machine learning (a form of artificial intelligence) could be used to analyze these reports in an efficient manner.

In this study, we examined if fast fashion companies changed their sustainability focuses in terms of the UN SDGs from 2018–2019 to 2020–2021 by utilizing sustainability reports. These two time periods were chosen to examine the effects before and during the pandemic, as there were changes made in health and work conditions. Our hypothesis was that the later years assessed would showcase higher similarities towards the UN health-related goals due to COVID-19 (11, 12). One generalized goal that the UN SDGs include is the “Good Health and Well-Being” goal, noted as Goal 3 in the UN SDG list (13). This tool can help identify goals that are and are not addressed frequently in the fashion industry with the hopes that these insights will be used to progress towards a more sustainable environment. By using natural language instead of numerical values, this project can provide alternative analysis techniques to identify trends that may not be found within other types of data. This can also support the UN Alliance for Sustainable Fashion’s mission of data sharing. The results noted that between the years the top 3 dominant UN SDGs were consistent, with them being “Responsible Consumption and Production” (Goal 12), “Clean Water and Sanitation” (Goal 6), and “Climate Action” (Goal 13). Yet when looking at the individual SDGs, some goals had significantly increased or decreased over the time period, which could be correlated to external factors.

### RESULTS

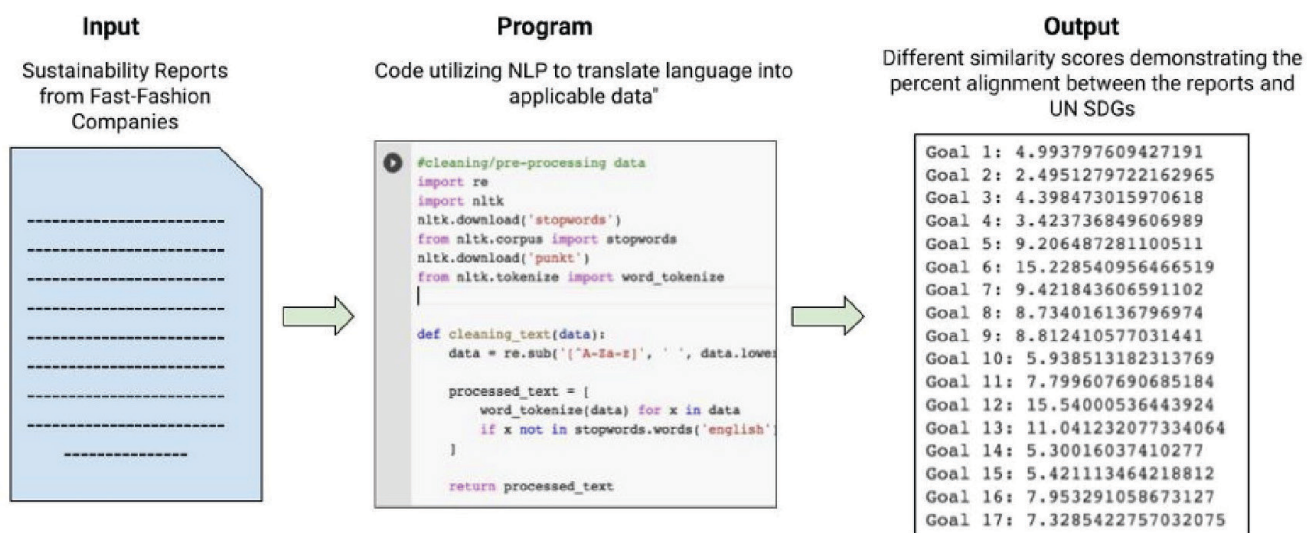
This project examined whether fast fashion companies have changed their sustainability focuses in terms of the UN SDGs from 2018–2019 to 2020–2021. The ten brands that were included in the study were American Eagle Outfitters, Gap, Guess, H&M, Nike, Prada, Primark, PVH, Uniqlo, and Zara. To answer this question, we collected sustainability

reports of the companies from the two time periods and inputted it into a cosine similarity algorithm (Figure 1). Based on the orientation of the vector, the cosine similarity algorithm will produce a value from -1 to 1 (or -100 to 100 in terms of percentage): the closer to 1, the more similar the vectors. The keywords and text extracted from the sustainability reports were vectorized and inputted into the algorithm. The output resulted in similarity scores between the reports and the keywords associated with each UN SDG and helped us identify our results (Figure 2).

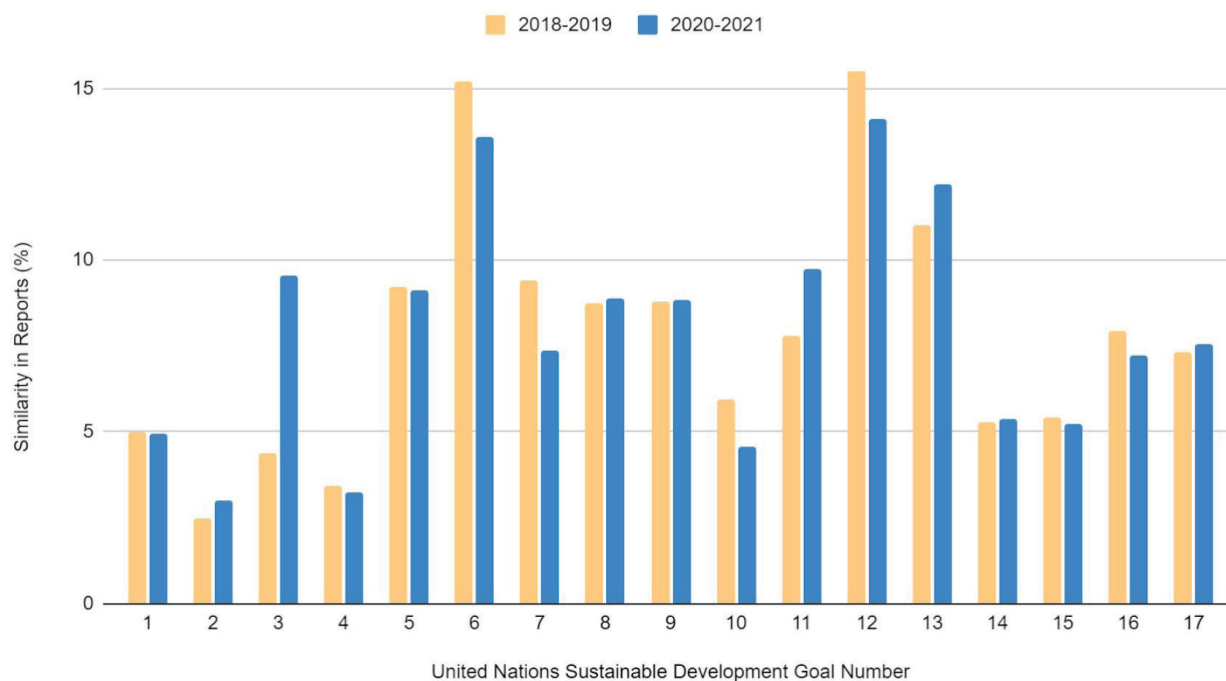
The top three most addressed UN SDGs were the same for both years. These three goals were “Responsible Consumption and Production” (Goal 12), “Clean Water and Sanitation” (Goal 6), and “Climate Action” (Goal 13) (Table 1). The most dominant goal, UN SDG 12, had the highest similarity percentages in both years with a score of 15.54% in 2018–2019 and 14.10% in 2020–2021 (Table 1). UN SDG 6 presented the second-highest similarity score with a score of 15.22% in 2018–2019 and 13.59% in 2020–2021 (Table 1). The third most dominant goal was UN SDG 13 with a score of 11.04% in 2018–2019 and 12.2% in 2020–2021 (Table 1).

When looking through the goals with the most change between the years, Goal 3: “Good Health and Well-Being” had the largest change: the reports had a similarity of 4.40% with the keywords of this goal in 2018–2019 and have more than doubled to 9.57% in 2020–2021 (Table 1). Some goals had noticeable decreases from 2018–2019 to 2020–2021 which included “Affordable and Clean Energy” (Goal 7) and “Reduced Inequalities” (Goal 10) (Figure 3, Figure 4). The goals with the least amount of observed change were UN SDGs 1, 4, 5, 8, 9, and 14 (Figure 4). Additionally, “Sustainable Cities and Communities” (Goal 11) had an increase from 2018–2021 going from 7.8% to 9.75% (Table 1). The higher percentages indicate more alignment to the keywords correlated with each UN SDG.

To test whether the overall similarity scores were statistically significant, we conducted a matched-pairs *t*-test



**Figure 1: Overview of algorithm process.** The diagram shows the inputs, the program, and the outputs of the technology used for this research. The left part of our diagram displays the inputs, which are the company sustainability reports. The middle portion of the diagram shows a section of our code and explains what is being done. The right portion of the diagram displays the outputs, which shows what the results (the similarity scores) look like.



**Figure 2: Report similarities against keywords from UN SDGs.** Graph showing percent similarities for each UN SDG for each time period tested, either 2018–2019 (yellow) or 2020–2021 (blue). The cosine similarity algorithm was used to measure the similarity between the keywords of each goal and the words in each year's reports.

with data from the two time periods. The summation of how much each goal changed over the two time periods was combined to calculate the overall *p*-value. Our calculations resulted in a larger *p*-value (approximately 0.74) compared to the significance level (0.05), which means that the data overall was not statistically significant as the companies did not significantly change their goals between the two reporting periods.

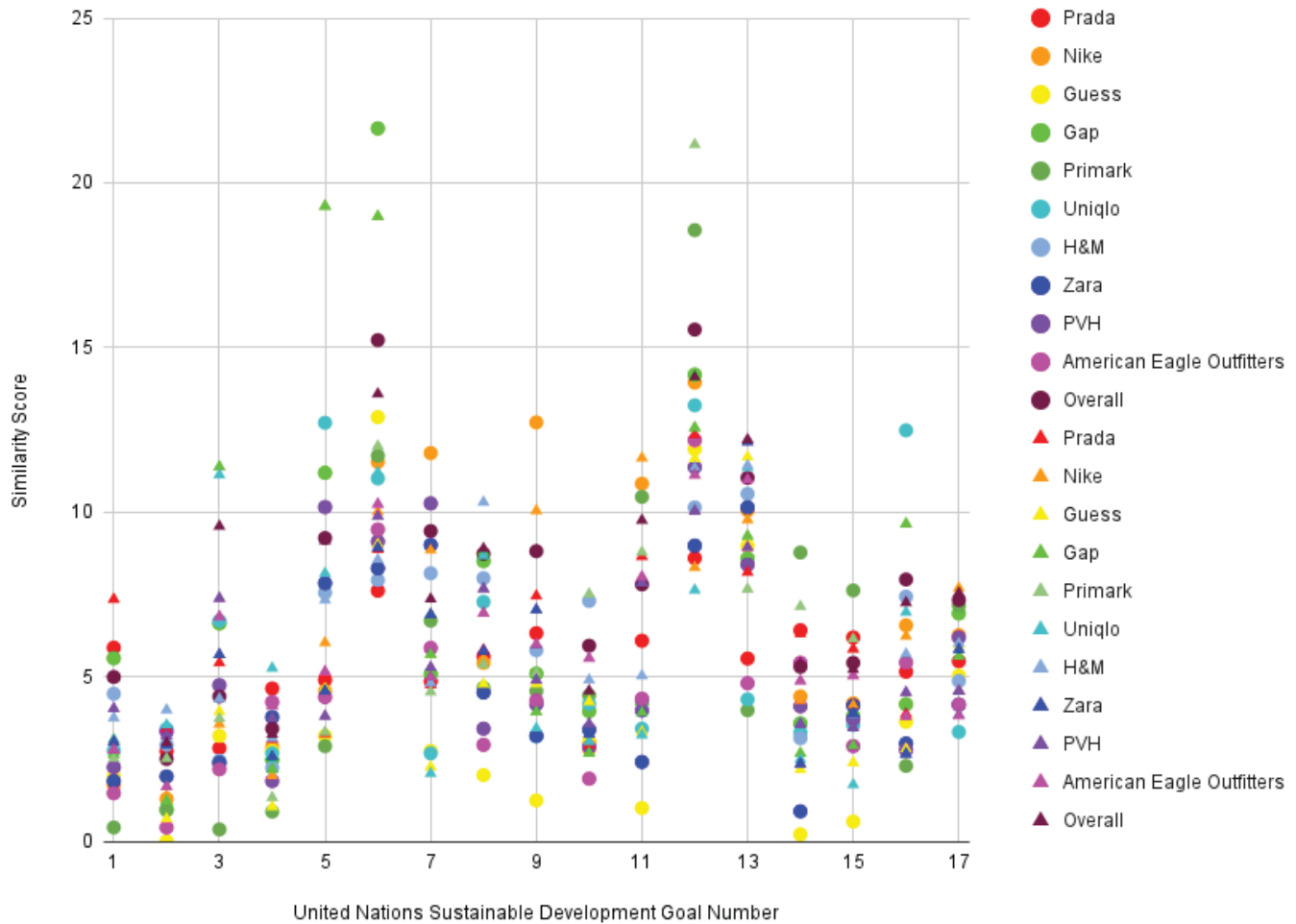
## DISCUSSION

The purpose of this study is to see whether fast fashion companies have changed their sustainability focuses in terms of the UN SDGs between two time periods (2018–2019 and 2020–2021). It was hypothesized that the reports from 2020–2021 would show more similarity with the UN health-related goals due to the pandemic caused by COVID-19. This hypothesis was tested by compiling company sustainability reports from both time periods and inputting it into a cosine similarity algorithm, which output the alignment between the company reports and each UN SDG. Our results showed that both time periods tested had the same top three addressed goals: “Responsible Consumption and Production” (SDG 12), “Clean Water and Sanitation” (SDG 6), and “Climate Action” (SDG 13). The matched-pairs t-testing validated that there was no statistical difference between the similarity scores for the main goals between the different time periods. This does not support our hypothesis as we predicted later years would show a higher similarity, yet there was a similar amount of focus on the main goals despite different years.

These top three goals can be explained through the actions and motivations reflected in the sustainability reports. For instance, to promote Goal 12: “Responsible Consumption

UN SDG Number	Similarity to SDG (%)	
	2018–2019	2020–2021
1: No Poverty	4.99	4.97
2: Zero Hunger	2.50	3.00
3: Good Health and Well-Being	4.40	9.57
4: Quality Education	3.42	3.25
5: Gender Equality	9.21	9.13
6: Clean Water and Sanitation	15.22	13.59
7: Affordable and Clean Energy	9.42	7.36
8: Decent Work and Economic Growth	8.73	8.89
9: Industry, Innovation, and Infrastructure	8.81	8.82
10: Reduced Inequalities	5.94	4.57
11: Sustainable Cities and Communities	7.80	9.75
12: Responsible Consumption and Production	15.54	14.10
13: Climate Action	11.04	12.20
14: Life Below Water	5.30	5.40
15: Life on Land	5.42	5.22
16: Peace, Justice, and Strong Institutions	7.95	7.25
17: Partnerships for the Goals	7.33	7.54

**Table 1: Report similarities against keywords from UN SDGs.** The table shows percent similarities for each UN SDG for each time period tested. The cosine similarity algorithm was used with sustainability reports and keywords as inputs.



**Figure 3: Similarity scores of fashion brands in 2018–2019 and 2020–2021.** The diagram shows the 10 companies and their individual alignment to the UN SDGs in 2018–2019 (represented by circles) and in 2020–2021 (represented by triangles). This shows how much each company contributes to the overall trend in the industry.

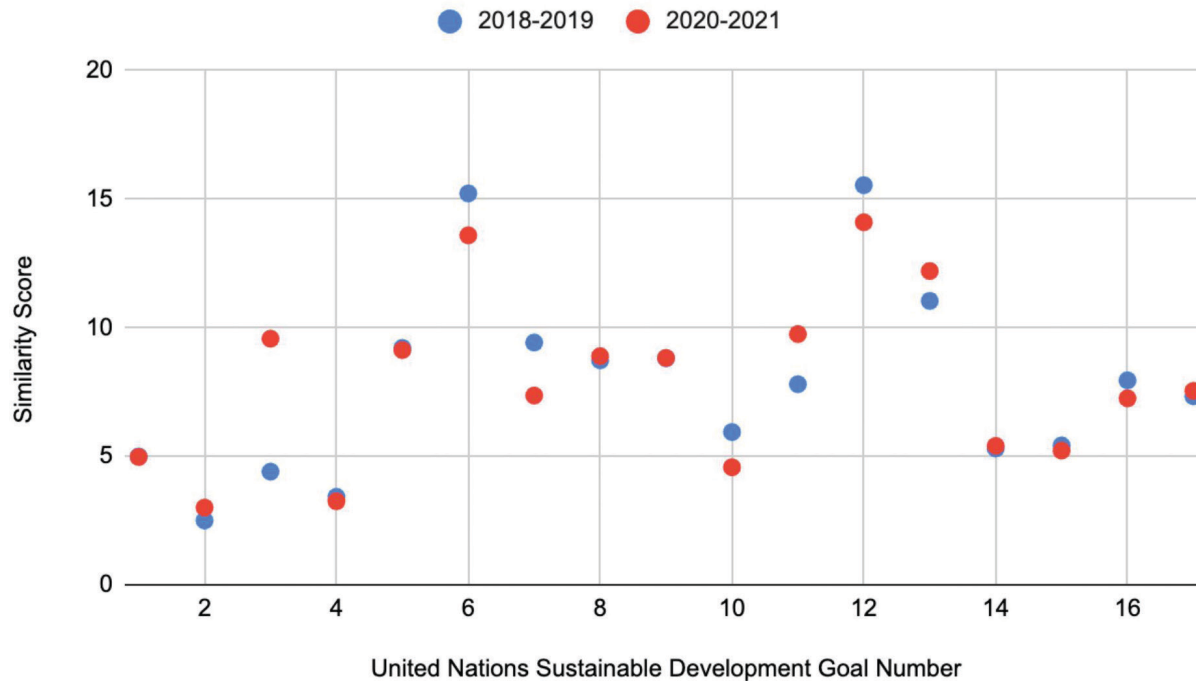
and Production,” one company educated over 150 employees on sustainable design and materials. Other companies are taking similar approaches by educating designers and associates about sustainable production (14). Companies are also addressing “Clean Water and Sanitation” through their management of wastewater (14). One company had about 6,500 cubic meters of wastewater in one of their locations. To reduce the impacts of this wastewater, the company purified the water prior to sending it to municipal water treatment facilities. Others are addressing water management through how it is disposed of, cleaned, and used (14). “Good Health and Well-Being” (SDG 3) had much higher similarity within the 2020–2021 sustainability reports. While we did not find a statistical difference between all SDGs over the two years, the increase in Goal 3 and change in similarity scores of other goals, were noted. Whether those specific differences are statistically different cannot be taken into account with the matched-pairs *t*-test. The test we conducted evaluated the change in similarity across all the goals and thus resulted in only one p-value; we had an insufficient amount of data to test each individual goal. The increase in SDG 3 similarity score could potentially be correlated to the increased concern about COVID-19 in this period (11, 12). Also, by

looking at each individual company, we can see that almost each individual company’s alignment with SDG 3 increased (Figure 3, Figure 4). This suggests that there was a trend throughout the industry rather than among a few companies.

Some goals showed slight declines in the similarity levels over time. For example, Goals 7 and 10 both had lower similarity levels in the post-pandemic time period. Factors like lower economic profit and distancing guidelines may have impacted the industry’s ability to support initiatives for clean energy and reducing inequalities. On the other hand, Goal 11 showed a slight increase, which could be due to how companies addressed this goal. For example, one company originally focused on the indirect influence on Goal 11 but in the later years had specific initiatives that were designed to directly impact this SDG.

This project can achieve more accurate results by expanding the datasets used, as currently only 10 of the largest fast fashion companies were reviewed and the top 20 keywords from each UN SDG were examined (1,15). A larger dataset would allow NLP techniques to analyze more language data and potentially find more concentrated alignment among the companies towards the UN goals.

For policymakers or civic organizations, it can be very



**Figure 4: Overall similarity scores in 2018–2019 and 2020–2021.** The diagram shows each time period’s similarity score alignment to the UN SDGs.

tedious and sometimes unreliable to read hundreds of pages of sustainability reports to find specific trends or topics that address how a retail company is working towards becoming more sustainable (14). This algorithm can make processing large sets of information more efficient, accessible, and fill in gaps where data is needed (16). It is important to acknowledge that many of these reports address past actions as well as future plans. The cosine similarity takes into consideration the whole report, which encapsulates both these topics. The results of this study may not directly translate to the current actions of these fast fashion companies since some proposed ideas that they hope or plan to implement.

We noticed that the most addressed areas were related to sustainable consumption/production. The results of this experiment showed how global changes impact sustainability and identified which topics are most and least addressed. With the tool created from this experiment, future research can be conducted to see how the fast fashion industry addresses sustainability over longer periods, which may show a more drastic change that can provide insight into the direction these companies are heading. If the NLP algorithm is more developed, it can also be provided to civic organizations to support their research endeavors (6,7). The algorithm can help collate data across different initiatives too. Another area of experimentation that can be explored is sustainability in different high-emission industries: how are other fields addressing sustainability, and how does that compare to fast fashion companies?

**MATERIALS AND METHODS**

This experiment was conducted by first collecting and compiling two different sets of reports: 10 of the largest brands’ sustainability reports from 2018–2019 and from 2020–2021. The companies were American Eagle Outfitters,

Gap, Guess, H&M, Nike, Prada, Primark, PVH, Uniqlo, and Zara (17). The reports used were downloaded from company websites. Each time period’s reports were analyzed together and not individually.

Each report’s text was extracted by using PyPDF, a Python library that can perform multiple tasks on pdf files. The text was then pre-processed and cleaned for the NLP model to understand the input. This was done by removing stopwords (such as “and”, “is”, and “the”), lowercasing, and lemmatizing all the extracted text for conformity, and finally tokenizing the text by breaking each word into its own token. The NLP cosine similarity algorithm was then applied to the processed language data. The outputs were the percent similarity to each UN SDGs’ keywords.

There were three main steps conducted: input, program, and output (Figure 1). A Google Collaboratory file was created to test the different algorithms and code. The library PyPDF was used to extract the text from the sustainability reports and was converted into a list of words and pre-processed. This is done by removing stopwords (such as “the” and “is”), numbers, and unnecessary punctuation. Additionally, all words were lowercased in the text to provide uniformity within the data. Finally, the text was tokenized, meaning that all of the data was split into singular words or “tokens”. The list was inputted into the cosine similarity algorithm along with the lists of each UN SDG’s 20 most important words. The keywords used for the comparison were compiled by the UN Department of Economic and Social Affairs (1).

To create the figures and tables, the Google suite services was utilized, which included Google Sheets and Google Drawings. A matched-pairs *t*-test was conducted to evaluate the statistical significance of the data, using the GraphPad *t*-test tool by *Dotmatics* (23). The following link is our code repository, which shows the techniques utilized for this project:

Github Link, <https://github.com/jaanyac/NLP-Model.git>

**Received:** August 27, 2022

**Accepted:** May 15, 2023

**Published:** September 11, 2023

### ACKNOWLEDGEMENTS

We sincerely appreciate the input and insights given by Mr. Minal Mistry, Ms. Sarah Gray, and Mr. Dave Wortman during the ideation stage of this research. We would also like to thank the creators of the open-source libraries, coding tools, and sustainability reports used for this project.

### REFERENCES

1. LaFleur, Macelo T. "Art Is Long, Life Is Short: An SDG Classification System for DESA Publications". *UN Department of Social and Economic Affairs*, 2019 [www.un.org/esa/desa/papers/2019/wp159\\_2019.pdf](http://www.un.org/esa/desa/papers/2019/wp159_2019.pdf). Accessed 21 January 2022.
2. World Bank Group. "How Much Do Our Wardrobes Cost to the Environment?" *World Bank*, World Bank Group, 23 Sept. 2019, [www.worldbank.org/en/news/feature/2019/09/23/costo-moda-medio-ambiente](http://www.worldbank.org/en/news/feature/2019/09/23/costo-moda-medio-ambiente). Accessed 10 March 2021.
3. Seth, Neha. "Topic Modeling and Latent Dirichlet Allocation (LDA) Using Gensim." *Analytics Vidhya*, 26 Aug. 2021, [www.analyticsvidhya.com/blog/2021/06/part-2-topic-modeling-and-latent-dirichlet-allocation-lda-using-gensim-and-sklearn/](http://www.analyticsvidhya.com/blog/2021/06/part-2-topic-modeling-and-latent-dirichlet-allocation-lda-using-gensim-and-sklearn/). Accessed 18 April 2021.
4. "Putting the Brakes on Fast Fashion." *UNEP*, [www.unep.org/news-and-stories/story/putting-brakes-fast-fashion](http://www.unep.org/news-and-stories/story/putting-brakes-fast-fashion). Accessed 10 March 2021.
5. "Take Action for the Sustainable Development Goals - United Nations Sustainable Development." *United Nations*, United Nations, [www.un.org/sustainabledevelopment/sustainable-development-goals/](http://www.un.org/sustainabledevelopment/sustainable-development-goals/). Accessed 18 April 2021.
6. Nguyen, Hoang Tien, et al. "Enhancing Sustainability in the Contemporary Model of CSR: A Case of Fast Fashion Industry in Developing Countries." *Social Responsibility Journal*, Emerald Publishing Limited, 7 May 2020. DOI: <https://doi.org/10.1108/SRJ-03-2019-0108>. Accessed 18 June 2021.
7. Dutta, Bhumika. "What Is Topic Modeling in NLP?" *Analytics Steps*, [www.analyticssteps.com/blogs/what-topic-modelling-nlp](http://www.analyticssteps.com/blogs/what-topic-modelling-nlp). Accessed 18 June 2021.
8. Mistry, Minal. "Insights for Sustainability Project." 7 Apr. 2022.
9. Gray, Sarah. *Insights from Senior Research Analyst for WRAP*, 4 Apr. 2022.
10. Wortman, David. *Insights from Oregon's Statewide Sustainability Officer*, 6 Apr. 2022.
11. Taylor Brydges, Monique Retamal & Mary Hanlon (2020) Will COVID-19 support the transition to a more sustainable fashion industry?, *Sustainability: Science, Practice and Policy*, 16:1, 298-308, DOI: <https://doi.org/10.1080/15487733.2020.1829848>. Accessed 15 June 2021.
12. "The Pandemic Is Changing Employee Benefits." *Harvard Business Review*, 21 Dec. 2021, [hbr.org/2021/04/the-pandemic-is-changing-employee-benefits](http://hbr.org/2021/04/the-pandemic-is-changing-employee-benefits). Accessed 22 May 2021.
13. "Health." *United Nations*, United Nations, [www.un.org/en/global-issues/health](http://www.un.org/en/global-issues/health). Accessed 20 May 2021
14. Moorhouse, Debbie. "Making Fashion Sustainable: Waste and Collective Responsibility." *One Earth* (Cambridge, Mass.), U.S. National Library of Medicine, 24 July 2020, [www.ncbi.nlm.nih.gov/pmc/articles/PMC7380204/](http://www.ncbi.nlm.nih.gov/pmc/articles/PMC7380204/). Accessed 14 March 2021.
15. "Home." *The UN Alliance for Sustainable Fashion*, 4 Apr. 2022, <https://unfashionalliance.org/>. Accessed 10 July 2022.
16. DeepAI. "Cosine Similarity." *DeepAI*, DeepAI, 17 May 2019, <https://deepai.org/machine-learning-glossary-and-terms/cosine-similarity>. Accessed 15 May 2021.
17. "The Top 10 Fashion Retailers in 2022." *Upp*, 14 Nov. 2022, [upp.ai/resources/the-top-fashion-retailers/](https://upp.ai/resources/the-top-fashion-retailers/). Accessed 17 December 2022.
18. Smith, P. "Sales of Major Apparel Manufacturers/Retailers Worldwide 2021." *Statista*, 2 June 2022, [www.statista.com/statistics/242114/sales-of-the-leading-10-apparel-retailers-worldwide/](http://www.statista.com/statistics/242114/sales-of-the-leading-10-apparel-retailers-worldwide/). Accessed 19 July 2022.
19. Gunawan1, D, et al. "IOPscience.The Implementation of Cosine Similarity to Calculate Text Relevance between Two Documents" *Journal of Physics: Conference Series*, IOP Publishing, 1 Mar. 2018, DOI: <https://doi.org/10.1088/1742-6596/978/1/012120>. Accessed 23 May 2022.
20. "Topic Modeling: An Introduction." *MonkeyLearn Blog*, 26 Sept. 2019, [monkeylearn.com/blog/introduction-to-topic-modeling/](http://monkeylearn.com/blog/introduction-to-topic-modeling/). Accessed 18 June 2021.
21. Sharma, Arun. "Understanding Latent Dirichlet Allocation (LDA)." *GreatLearning Blog: Free Resources What Matters to Shape Your Career!*, 3 Mar. 2022, [www.mygreatlearning.com/blog/understanding-latent-dirichlet-allocation/](http://www.mygreatlearning.com/blog/understanding-latent-dirichlet-allocation/). Accessed 10 June 2021.
22. Camacho-Collados, Jose, and Mohammad Taher Pilehvar. "On the Role of Text Preprocessing in Neural Network Architectures: An Evaluation Study on Text Categorization and Sentiment Analysis." *ACL Anthology*, [aclanthology.org/W18-5406/](http://aclanthology.org/W18-5406/). [www.emerald.com/insight/content/doi/10.1108/SRJ-03-2019-0108/full/html](http://www.emerald.com/insight/content/doi/10.1108/SRJ-03-2019-0108/full/html). Accessed 05 February 2022.
23. Dotmatics. GraphPad. GraphPad Software, [www.graphpad.com/quickcalcs/ttest1/](http://www.graphpad.com/quickcalcs/ttest1/). Accessed 22 October 2022.

**Copyright:** © 2023 Chadha, Gubba, and Agrawal. All JEI articles are distributed under the attribution non-commercial, no derivative license (<http://creativecommons.org/licenses/by-nc-nd/3.0/>). This means that anyone is free to share, copy and distribute an unaltered article for non-commercial purposes provided the original author and source is credited.