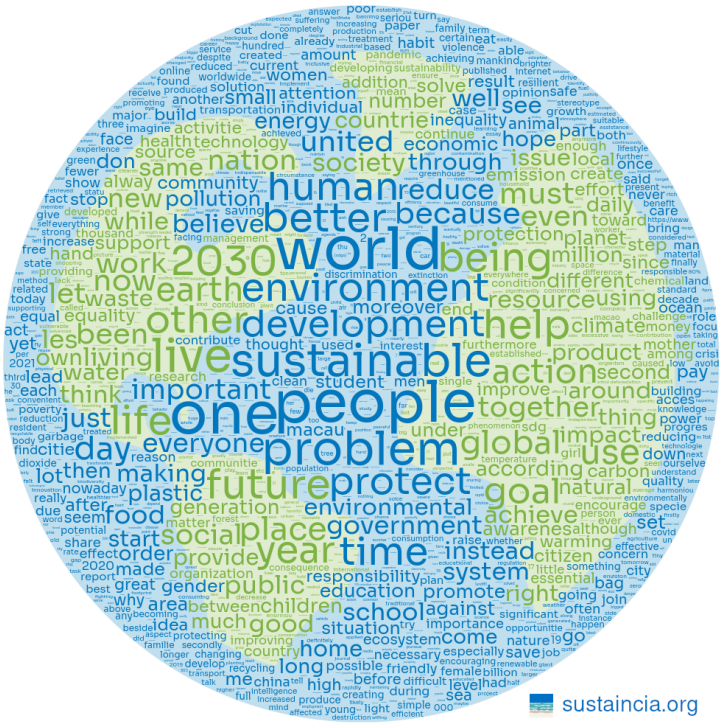


All word tokens used in the analysis

Sustainability Essay Macao Contest Analysis

Introduction

Sustainability is important for the continuation of life on our planet, especially for our future generations. Our students are our future. Unfortunately, they will also be the bearer of problems we cause today; therefore their participation in and awareness of sustainability is essential. They need to be part of the solution and we need to know about their concerns.

A writing competition[1] for Macao students on sustainability (a first that we're aware of) titled "The Change I Want To Make" was held in Oct 2021. Results of the contest were announced Dec 2021 and this is an analysis of these winning entries.

Objectives

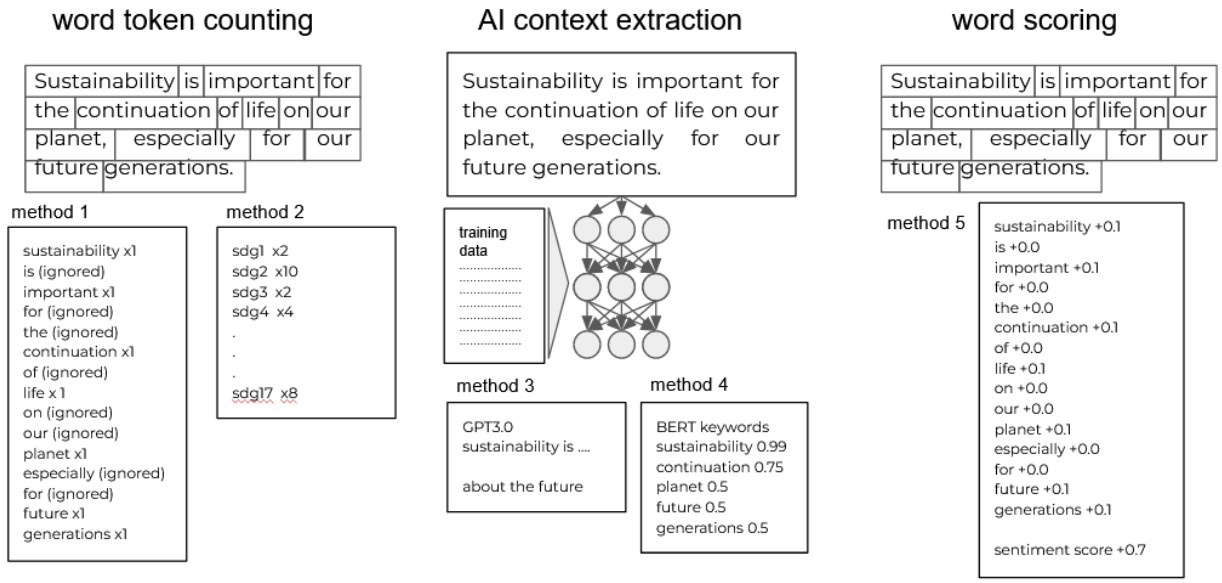
Sustaincia.org[2], as co-organizer of this contest, wanted to see if using machine analysis could offer some insights into the kind of sustainable topics which were receiving attention in aggregate. The analysis would take an in-depth look at the winning work of the students; but since these were judged works, the insights cover judges as well. Much work is put into maintaining objectivity as much as possible using machine analysis, to avoid writing styles and other subjective attachments. Key insights were found beyond the original goal and discussed in the last "Conclusions and Recommendations" section. The next section will describe the multi methods used, followed by the actual outputs of these methods.

[1] see https://essaymacao.com for more details

[2] https://sustaincia.org is a locally registered association encouraging the use of technology and social efforts for sustainable development

Methodologies

Five methodologies were used which can be categorized into three general approaches: 1) word token counting, 2) AI context extraction and 3) word scoring.



But first here is some background about the data source:

- Contest Background
  - 1,644 students registered for the contest
  - From Upper Primary age 8 to University less than age 26
  - Winning essays 17 x 4 categories = total 68 winning essays
- Human Judging Criterias
  - Content 35%, Organization 30%, Mechanics 20%, Originality 15%, with a total of 10 scores[3]
  - done by hand with randomized assignment
- Dataset
  - 68 winning articles
  - total words 37,297
  - unique terms 4,720

Word token counting approaches:

- Method 1 - terms count
  - a numerical count of all the terms
  - common terms were ignored e.g. he, she , it ....
  - as terms usually appear multiple times in an article especially when the writing revolves around the term, therefore we count each term only once per article
  - combine all the counts from all articles to get aggregated data

Output	[term , number of articles mentioning the term at least once], ...

[3] https://essaymacao.com/wp-content/uploads/2021/10/20211011JudgingRubric.pdf

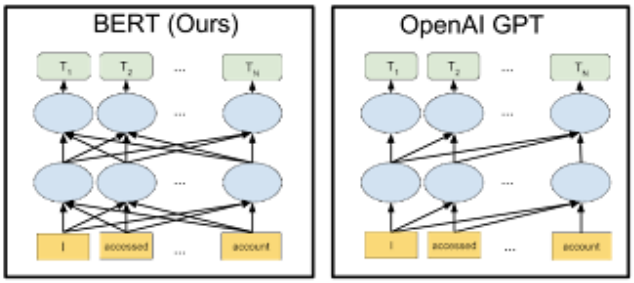
- Method 2 - sdg terms count
    - assign terms to each of the 17 sustainable development goals from UN
    - for each goal count number of articles mentioning the SDG terms once per goal
    - keywords were first extracted from UN's SDG website, further keywords were referenced from the following:
        - https://leicester.figshare.com/articles/dataset/SDG_Research_Publication_Keywords/12839519/1
        - https://data.utoronto.ca/sustainable-development-goals-sdg-report/sdg-report-appendix/
        - https://github.com/Aurora-Network-Global/sdg-queries
        - https://www.its.ac.id/drpm/wp-content/uploads/sites/71/2021/04/Daftar-keywords-Sustainable-Development-Goals.pdf
        - https://www.sdgmapping.auckland.ac.nz

Output	#SDG goal number of articles counted towards this goal (1) the term number of articles mention the term at least once (2) * (1) is not the sum of (2)s as multiple terms could be matched to the same article, and that would be counted only once for (1)

AI context extraction approaches:

- Method 3 - Gpt-3 generative pre-trained transformer
    - word counts disregards context, so try another automated technique call natural language processing, which uses neural networks created using large training sets, then using these neural networks to model the articles and then use model to output information like keyword extraction
    - problem with this method was found when the length of the writing caused missing output, perhaps the token size were too large[4]
    - turns out Gpt-3 (openai) is good for predictive conversation, but not so good for keyword extraction, as sometimes there is no output, therefore this method was discarded
- Method 4 - BERT Bidirectional Encoder Representations from Transformers
    - is a technique developed by Google for Natural Language Processing (NLP) pre-trained using a large text corpus and which looks at both word and its sentence position to map context - https://ai.googleblog.com/2018/11/open-sourcing-bert-state-of-art-pre.html

[4] see test.sh and test-s.sh



- On SQuAD v1.1 (Stanford Question Answering dataset for reading comprehension), BERT achieves 93.2% F1 score (a measure of accuracy), surpassing the previous state-of-the-art score of 91.6% and human-level score of 91.2%
- Using the model, we map the whole document and then extract 5 ngram3 words which match most with the document's representation

Output	each row is an output from a single article [term , the angle of the term to the article (1 means close)], ... variance is then calculated for the angle for each article $Mean(M) = (\sum_{i=0}^{n-1} arr[i])$ / n And variance (V) = $(\sum_{i=0}^{n-1} (arr[i] - M)^2)$ / n

Word scoring approach:

- Method 5 - Sentiment Analysis
  - assign sentimental score to each word and then summing the score
  - then the sum is normalize by the number of words
  - words sentimental values are declare with afinn-165 http://www2.imm.dtu.dk/pubdb/pubs/6010-full.html

Findings/results

The five methods described above were applied to the students overall and then again separately for each of the 4 age groups. Note that output for method 3 is skipped due to the situation explained in the above methodology section.

Results from each method is then presented as:
1) actual output from the program
2) summary observation

For method 2 related to 17 SDG goals, a visualization of the output is given as a pie chart and each goals' percentage.

For method 5, each writing is analyzed individually and an average is calculated for each age group and for overall.

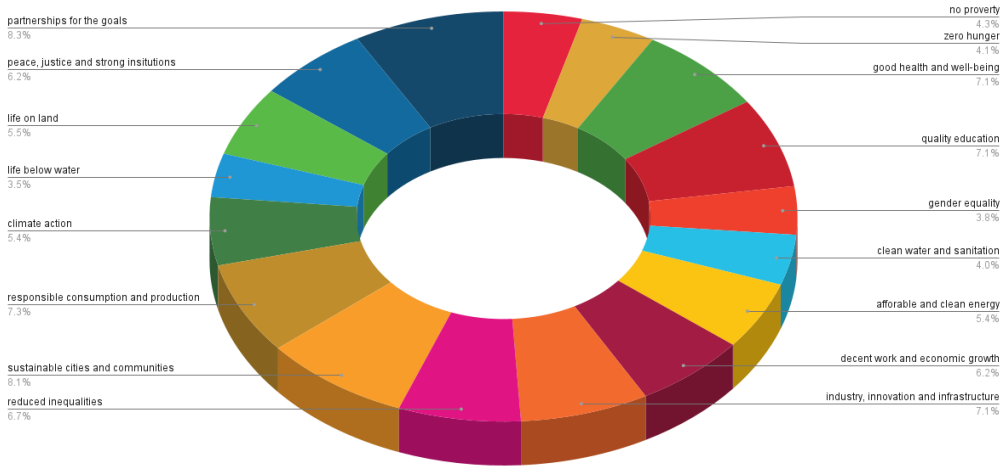
Methods 1/2/4 Results for Overall:

Output Method 1 (top 10 keywords)	Findings
['one', 17], ['sustainable', 16], ['people', 16], ['development', 15], ['world', 14],['time', 13], ['better', 13], ['goal', 13],['human', 13], ['2030', 12]	Highest frequency 3 words are 'one', 'sustainable', 'people' found in ~90% of all articles; the next 6 highest frequency words are 'development', 'world', 'time', 'better', 'goal', 'human' found in ~75% of article; and '2030' is in ~70% of articles.

Method 2 (17 sdg goals)		
#1	25	poverty : 10 income : 4 distribution : 3 wealth : 1 wealthy : 2 socio : 1 poor : 11 poorest : 1 money : 12 homeles : 3 prosperity : 4 microfinance : 1
#2	24	hunger : 5 hungry : 4 agriculture : 5 agricultural : 2 nutrition : 1 nutrient : 1 famine : 4 food : 23 farm : 6 farming : 1 starving : 2 starvation : 5 malnutrition : 1 malnourish : X meal : 4 crop : 4
#3	41	health : 15 healthy : 7 healthcare : 1 healthier : 5 medicine : 2 medical : 6 death : 7 die : 11 pandemic : 10 disease : 6 doctor : 4 physician : 1 nursing : 1 nurse : 2 sick : 1 sicknesse : 1 ill : 1 illnesse : 1 treatment : 8
#4	41	education : 17 educate : 6 school : 22 teacher : 6 student : 16 kid : 4 young : 9 skill : 6 teaching : 4 children : 18 learn : 9 learning : 5 vocational : 1
#5	22	gender : 16 woman : 6 women : 15 girl : 11 empower : 4 female : 11 male : 6 sex : 4 sexism : 3 genital : 1 genitalia : 1 men : 13 man : 11 marriage : 3 marry : 2 married : 3 feminine : 1 femininity : 1 feminism : 1 feminist : 1 trafficking : 2
#6	23	water : 21 sanitation : 2 sewage : 2 drain : 1 pond : 1 lake : 1 river : 2 toilet : 4 rainwater : 2 bathing : 1 drinking : 3 aquifer : 2 desalination : 2 diarrheal : 1 hygienic : 2 wastewater : 3
#7	31	energy : 20 power : 12 solar : 3 hydrogen : 1 renewable : 7 wind : 4 hydro : 1 electricity : 5 wave : 2 light : 10
#8	36	job : 10 economic : 17 economy : 4 economie : 3 growth : 10 employment : 3 employee : 5 employed : 2 unemployment : 1 welfare : 1 gdp : 2 labor : 2 wage : 1 migrant : 1 worker : 7 slavery : 2
#9	41	industry : 6 industrie : 4 industrial : 5 innovate : 1 innovation : 5 infrastructure : 2 sector : 5 region : 6 regional : 2 domestic : 6 international : 5 digital : 2 internet : 8 mobile : 1 factory : 2 factorie : 4 scientific : 1 technology : 15 technologie : 6 technological : 2
#10	39	inequality : 14 inequalitie : 4 unequal : 3 equal : 13 discriminate : 2 discrimination : 9 inclusion : 1 inclusive : 5 country : 13 countrie : 20 clas : 2 disability : 1 disabilitie : 1 disabled : 2 ethnic : 3 ethnicity : 1 racism : 1 opportunitie : 8 opportunity : 2 assistance : 5
#11	47	city : 12 citie : 15 settlement : 2 planning : 2 space : 7 local : 16 building : 11 overcrowded : 1 slum : 1 community : 16 communitie : 10 neighbour : 1 house : 3 housing : 2 disaster : 4 transport : 7
#12	42	consumer : 2 consumption : 7 consume : 7 consuming : 5 recycle : 5 recycling : 9 retailer : 1 buy : 6 purchase : 3 production : 7 produce : 9 exploit : 1 waste : 21 shop : 3 shopping : 5 reuse : 4 reusable : 7 reused : 1 supply : 3 fashion : 3
#13	31	climate : 18 weather : 5 temperature : 8 celsiu : 2 warm : 4 warming : 15 hot : 2 atmosphere : 4 carbon : 16 dioxide : 9 emit : 2 emission : 16 emitted : 1 emitter : 1 greenhouse : 7 ice : 2 methane : 4
#14	20	fishing : 1 fish : 3 overfishing : 2 sea : 14 ocean : 13 marine : 7 island : 1 whale : 3 turtle : 2 coral : 3 acidification : 2 aquatic : 1 aquaculture : 1
#15	32	land : 10 animal : 14 mountain : 3 forest : 8 deforestation : 4 biodiversity : 4 biodiverse : 1 ecosystem : 13 ecology : 2 drought : 1 flood : 5 desert : 1 hunting : 1 wildlife : 2 specie : 10 soil : 4
#16	36	peace : 8 peaceful : 7 justice : 2 fair : 4 accountable : 3 bribery : 1 corruption : 2 institute : 2 institution : 3 security : 4 police : 2 law : 9 legal : 1 un : 7 conflict : 2 crime : 5 violence : 9 transparent : 1
#17	48	cooperation : 4 cooperative : 1 together : 25 forum : 2 platform : 3 discus : 2 international : 5 national : 6 nation : 25 partnership : 1 volunteering : 2 coordinated : 1 organisation : 2 organization : 12 society : 23

Findings

Out of 17 SDG goals, the goal with the highest count is #17 Partnership for the Goals, while the lowest counted goal is #14 Life Below Water.



Output Method 4 (BERT) first refer to outputs for each group on next pages	average variance
primary school	0.02373074913
middle school	0.01920732786
highschool	0.01661393478
university	0.01629507769

Findings

Each age group's AI keywords extraction returns top five 3-ngram for each document, these have an angle which we calculate variance with, then these variances are then averaged, we find that the average variance decreases as age increases, in another word, keywords get closer to model representation of the document as age increase. Keep in mind, word count and word usage skill also increase with age.

Methods 1/2/4 Results for Primary school:

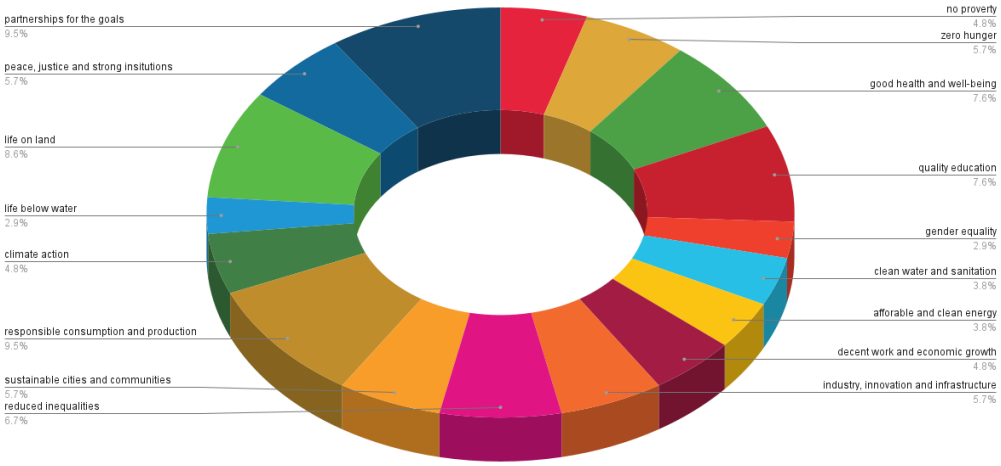
Output Method 1 (top 10 keywords)	Findings
['world', 12], ['one', 10], ['people', 10], ['live', 10], ['2030', 9], ['problem', 9], ['better', 8], ['help', 7], ['think', 7], ['future', 7]	Highest frequency word is “world” found in ~71% of articles; next highest 6 words are ‘one’, ‘people’, ‘live’, ‘2030’, ‘problem’, ‘better’ found in ~59% of articles and next 3 words ‘help’, ‘think’, ‘future’ are found in ~41% of articles.

Output Method 2 (17 sdg goals)

#1	5	poverty : 3 income : X distribution : X wealth : X wealthy : X socio : X poor : 1 poorest : X money : 2 homeles : 2 prosperity : 1 microfinance : X
#2	6	hunger : 1 hungry : 2 agriculture : 3 agricultural : X nutrition : 1 nutrient : X famine : 1 food : 5 farm : 3 farming : X starving : X starvation : 1 malnutrition : 1 malnourish : X meal : X crop : 3
#3	8	health : 5 healthy : 1 healthcare : X healthier : 1 medicine : X medical : X death : 1 die : 3 pandemic : X disease : 2 doctor : X physician : X nursing : X nurse : 1 sick : X sickness : 1 ill : X illnesse : X treatment : X
#4	8	education : 3 educate : 1 school : 5 teacher : 1 student : 3 kid : X young : 1 skill : 1 teaching : X children : 5 learn : 1 learning : X vocational : X
#5	3	gender : 3 woman : 1 women : 3 girl : 1 empower : 1 female : 1 male : 1 sex : X sexism : X genital : X genitalia : X men : 2 man : 2 marriage : X marry : 2 married : X feminine : X femininity : X feminism : X feminist : X trafficking : X
#6	4	water : 4 sanitation : X sewage : X drain : X pond : X lake : X river : X toilet : X rainwater : X bathing : X drinking : X aquifer : X desalination : X diarrheal : X hygienic : X wastewater : X
#7	4	energy : 3 power : 2 solar : 1 hydrogen : X renewable : 1 wind : 1 hydro : X electricity : 1 wave : X light : 1
#8	5	job : 4 economic : 1 economy : X economie : 1 growth : X employment : X employee : X employed : X unemployment : X welfare : X gdp : X labor : X wage : X migrant : X worker : X slavery : X
#9	6	industry : 1 industrie : X industrial : X innovate : X innovation : X infrastructure : X sector : 1 region : X regional : X domestic : 1 international : X digital : X internet : 1 mobile : X factory : X factorie : X scientific : X technology : 2 technologie : X technological : 1
#10	7	inequality : 1 inequalitie : X unequal : X equal : 1 discriminate : 1 discrimination : 2 inclusion : X inclusive : X country : X countrie : 4 clas : X disability : 1 disabilitie : X disabled : 1 ethnic : X ethnicity : X racism : X opportunitie : 1 opportunity : 2 assistance : 2
#11	6	city : 1 citie : 2 settlement : X planning : X space : X local : 1 building : 3 overcrowded : X slum : X community : X communitie : 1 neighbour : X house : X housing : 1 disaster : 1 transport : 2
#12	10	consumer : 1 consumption : 1 consume : X consuming : X recycle : 1 recycling : 4 retailer : X buy : 1 purchase : X production : 2 produce : 2 exploit : X waste : 4 shop : X shopping : X reuse : X reusable : 2 reused : X supply : X fashion : X
#13	5	climate : 4 weather : 2 temperature : X celsiu : X warm : X warming : 3 hot : X atmosphere : X carbon : 3 dioxide : 2 emit : 1 emission : 1 emitted : X emitter : X greenhouse : 1 ice : X methane : 2
#14	3	fishing : X fish : 1 overfishing : 1 sea : 1 ocean : 3 marine : 1 island : X whale : 1 turtle : 1 coral : 1 acidification : X aquatic : X aquaculture : X
#15	9	land : 2 animal : 5 mountain : X forest : 3 deforestation : X biodiversity : 2 biodiverse : 1 ecosystem : 3 ecology : X drought : 1 flood : 2 desert : X hunting : X wildlife : 1 specie : 2 soil : X
#16	6	peace : 2 peaceful : X justice : X fair : X accountable : X bribery : X corruption : X institute : X institution : X security : 2 police : 1 law : X legal : 1 un : X conflict : X crime : X violence : 1 transparent : X
#17	10	cooperation : X cooperative : X together : 6 forum : X platform : X discus : X international : X national : 1 nation : 5 partnership : X volunteering : X coordinated : X organisation : X organization : 1 society : 5

Findings

Out of 17 SDG goals, the goals with the highest count are #12 Responsible consumption and production , #17 Partnerships for the goals (tie); whereas the lowest count are #5 Gender Equality and #14 Life Below Water.



Output Method 4 (BERT)	variance
[['growing trees worldwide', 0.1747], ['expanding wooded areas', 0.0243], ['live coming rebuild', 0.1801], ['planet benefit improving', 0.0584], ['world goal 2030', 0.2241]],	0.005927010
[['need protect ecosystem', -0.0055], ['water beautiful beaches', 0.2902], ['overfishing causes decrease', 0.2551], ['people polluting water', 0.3309], ['trashing products ocean', 0.0452]],	0.018617118
[['people need government', 0.1953], ['funding housing start', -0.319], ['families reduce rent', -0.1823], ['children growing orphanages', 0.121], ['150 million homeless', 0.2829]],	0.053178582
[['super typhoon', -0.1639], ['super typhoon heavy', 0.1042], ['warming greatly affects', 0.1392], ['concerns climate change', -0.1992], ['climate change', -0.0869]],	0.019158606
[['huge cleaning cost', 0.0493], ['ocean encourage governments', 0.1347], ['keeping ocean clean', -0.1853], ['fish 2030', 0.1008], ['waste double 2030', 0.2076]],	0.017861398
[['want trees flowers', -0.0137], ['adding stronger facilities', -0.1495], ['help homeless people', -0.0085], ['protect forest disasters', 0.2667], ['dream better planet', 0.0954]],	0.019117962
[['build life parks', -0.1473], ['dropped dramatically 2017', 0.032], ['making world biodiverse', 0.2806], ['rhinos eating grass', 0.1319], ['250 white rhinos', 0.1375]],	0.019993738
[['women increased significantly', 0.2186], ['women spiritual health', -0.089], ['gender equality majority', 0.0334], ['women need work', -0.1076], ['girls forced marry', -0.1286]],	0.016773318
[['fertilizer harmful health', 0.1615], ['waste turn fertilizer', 0.0217], ['wasting food encouraging', 0.1762], ['4317 recycle food', -0.1942], ['food waste high', -0.1524]],	0.023701242
[['main cause famine', 0.1067], ['000 people die', 0.3367], ['million children suffering', -0.1648], ['hunger achieving food', 0.2155], ['climate change food', -0.0564]],	0.032643154
[['building urban farms', 0.2482], ['rewilding farmland increase', 0.1188], ['want centuries human', 0.1074], ['products kill animals', 0.0746], ['modern humans selfishly', 0.257]],	0.005787680
[['protect women violence', -0.0627], ['think females weak', 0.1222], ['violence forced marry', 0.0924], ['illustrated inequality gender', -0.1408], ['job people discriminate', -0.1989]],	0.015946794
[['heavy meat eater', 0.0945], ['meat burger yummy', 0.2103], ['cow emits methane', 0.1861], ['climate change statistic', 0.185], ['farmers use antibiotics', -0.273]],	0.032815574
[['warming makes frequent', -0.1121], ['crops causes starvation', 0.0627], ['year increase carbon', 0.2476], ['storms tornadoes', 0.3419], ['extreme drought kills', -0.2919]],	0.053516806
[['world genders opportunities', -0.134], ['opportunity gender stereotype', 0.1132], ['gender stereotype want', 0.1862], ['policed women bodies', -0.0056], ['equality empower women', 0.0763]],	0.012029018
[['china destroying', 0.0139], ['2019 australian bushfires', -0.1395], ['scenes hurricanes hail', -0.1086], ['causing millions lives', -0.1331], ['bushfires burned months', 0.4098]],	0.043347564
[['schools live bad', -0.1416], ['2014 awarded kailash', -0.0995], ['government build libraries', -0.2226], ['nobel peace prize', 0.1035], ['teachers deserve pay', -0.1853]]	0.013007172

Findings

For AI keyword extraction for each document in the group, the resulting 3-ngram keywords angles return an average variance of 0.02373074913 (highest overall).

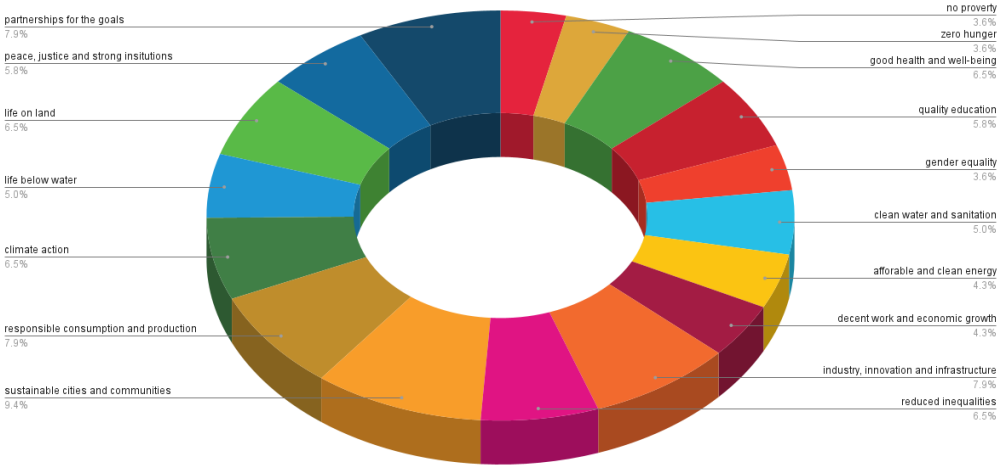
Methods 1/2/4 Results for Middle school:

Output Method 1 (top 10 keywords)	Findings
['world', 12], ['live', 12], ['2030', 10], ['people', 10], ['other', 10], ['help', 10], ['environment', 10], ['life', 9], ['one', 9], ['day', 9]	Highest frequency words are 'world' and 'live found in ~71% of articles; next 5 highest frequency words are '2030', 'people', 'other', 'help', 'environment' found in ~59% of articles; next 3 words 'life', 'one', 'day' are found in ~53% of articles.

Output Method 2 (17 sdg goals)		
#1	5	poverty : 2 income : 1 distribution : X wealth : X wealthy : 1 socio : 1 poor : 3 poorest : 1 money : 4 homeles : X prosperity : 1 microfinance : 1
#2	5	hunger : 2 hungry : X agriculture : 1 agricultural : 1 nutrition : X nutrient : 1 famine : 1 food : 6 farm : 1 farming : 1 starving : 1 starvation : 1 malnutrition : X malnourish : X meal : 1 crop : X
#3	9	health : 2 healthy : 2 healthcare : X healthier : 2 medicine : X medical : X death : 2 die : 3 pandemic : 1 disease : X doctor : X physician : X nursing : X nurse : X sick : X sickness : X ill : X illnesse : X treatment : 1
#4	8	education : 3 educate : 1 school : 5 teacher : 1 student : 2 kid : 1 young : X skill : 2 teaching : 1 children : 2 learn : X learning : 1 vocational : X
#5	5	gender : 3 woman : 2 women : 4 girl : 2 empower : 1 female : 3 male : 1 sex : 1 sexism : X genital : X genitalia : X men : 3 man : 1 marriage : 1 marry : X married : X feminine : X femininity : X feminism : X feminist : X trafficking : 1
#6	7	water : 7 sanitation : 1 sewage : 2 drain : X pond : 1 lake : 1 river : 1 toilet : 1 rainwater : 1 bathing : X drinking : 1 aquifer : 1 desalination : X diarrheal : X hygienic : 1 wastewater : 2
#7	6	energy : 4 power : X solar : X hydrogen : X renewable : 1 wind : X hydro : X electricity : X wave : 1 light : X
#8	6	job : 2 economic : 3 economy : X economie : X growth : 1 employment : X employee : 1 employed : 1 unemployment : X welfare : X gdp : 1 labor : 1 wage : X migrant : 1 worker : 1 slavery : 1
#9	11	industry : 2 industrie : 2 industrial : 1 innovate : X innovation : 1 infrastructure : X sector : X region : 1 regional : 1 domestic : 2 international : X digital : X internet : 2 mobile : X factory : 1 factorie : 2 scientific : X technology : 1 technologie : 1 technological : X
#10	9	inequality : 4 inequalitie : X unequal : X equal : 4 discriminate : 1 discrimination : 2 inclusion : X inclusive : 2 country : 4 countrie : 4 clas : X disability : X disabilitie : X disabled : X ethnic : X ethnicity : 1 racism : 1 opportunitie : 1 opportunity : X assistance : X
#11	13	city : 2 citie : 4 settlement : X planning : X space : 1 local : 4 building : X overcrowded : X slum : X community : 2 communitie : 3 neighbour : X house : X housing : X disaster : 2 transport : 2
#12	11	consumer : X consumption : X consume : 3 consuming : 1 recycle : 1 recycling : 2 retailer : X buy : 3 purchase : X production : 1 produce : 2 exploit : X waste : 6 shop : 2 shopping : 2 reuse : 1 reusable : 2 reused : 1 supply : 1 fashion : X
#13	9	climate : 5 weather : 1 temperature : 4 celsiu : 1 warm : 1 warming : 5 hot : 1 atmosphere : 2 carbon : 5 dioxide : 4 emit : X emission : 4 emitted : X emitter : 1 greenhouse : 1 ice : X methane : 1
#14	7	fishing : 1 fish : 2 overfishing : 1 sea : 6 ocean : 4 marine : 2 island : X whale : X turtle : X coral : 1 acidification : 1 aquatic : X aquaculture : X
#15	9	land : 2 animal : 4 mountain : 1 forest : 1 deforestation : 1 biodiversity : X biodiverse : X ecosystem : 5 ecology : X drought : X flood : 1 desert : X hunting : X wildlife : X specie : 1 soil : 2
#16	8	peace : X peaceful : 2 justice : X fair : X accountable : 2 bribery : X corruption : 1 institute : X institution : 1 security : X police : X law : 5 legal : X un : X conflict : X crime : 1 violence : 2 transparent : X
#17	11	cooperation : 1 cooperative : X together : 7 forum : 1 platform : X discus : 1 international : X national : 1 nation : 5 partnership : X volunteering : X coordinated : 1 organisation : 1 organization : X society : 5

Findings

Out of 17 SDG goals, the goal with the highest count is #11 Sustainable Cities and Communities; while the goal with the lowest counts are #1 No Poverty , #2 No Hunger , and #5 Gender Equality.



Output Method 4 (BERT)	variance
[['promoting urban development', 0.2654], ['ideal 2030 people', 0.2465], ['enforcement green ecological', 0.0979], ['remediation eco tourism', 0.4227], ['future ideal 2030', 0.2106]],	0.0109503896
[['sea birds losing', 0.2711], ['pollution gets worse', -0.0514], ['oil spill accidents', 0.02], ['water pollution efficiently', 0.1202], ['100 million marine', 0.3574]],	0.0231632224
[['business enrichment expense', -0.0386], ['entrepreneurs observe code', 0.0242], ['caring sustainable future', 0.4011], ['2030 agenda sustainable', -0.0276], ['environmental pollution entrepreneurs', -0.0197]],	0.0282136376
[['sea photosynthesis factories', -0.1253], ['fishing small fish', 0.1892], ['moratorium develop tourism', -0.2641], ['fishermen lose income', -0.2567], ['sea pollution 2030', -0.2252]],	
[['blessing mankind destructive', -0.0538], ['biggest tax contributors', 0.063], ['laws need politicians', -0.0897], ['climate change change', -0.1231], ['climate change success', 0.0113]],	0.0289606776
[['crimes targeting asian', 0.0094], ['street racism stopped', 0.2825], ['racism world peaceful', -0.1232], ['causing corruption society', 0.2713], ['classmates severe racism', -0.0948]],	0.0045589944
[['poorest countries world', 0.1756], ['poverty education good', 0.1165], ['want make civilization', 0.1766], ['people starving everyday', 0.1669], ['warming global warming', 0.3018]],	0.0307661944
[['gender equality families', 0.2932], ['harvard university 28th', -0.0003], ['suffering gender inequality', 0.2972], ['president harvard president', -0.2583], ['future woman president', 0.2369]],	0.0037580536
[['pollution getting year', -0.0271], ['enter immerse oceans', 0.0987], ['organisms oceans healthier', 0.0709], ['century ocean pollution', 0.2596], ['oceans healthier public', -0.1302]],	0.0464901464
[['controversial long gender', 0.2545], ['change bias women', 0.0424], ['inequality classmates encouraging', 0.0924], ['news gender inequality', 0.4028], ['powerful female heroes', 0.5157]],	0.0170122376
[['casino industry forced', -0.0629], ['female citizens immigrants', 0.2122], ['trafficking victims deceived', 0.2414], ['slavery hearing females', 0.2411], ['heartbreaking macau government', -0.2649]],	0.0322463864
[['great scientists', -0.0467], ['activities destroying marine', 0.011], ['2030 oceans', 0.0526], ['scientists think save', 0.2301], ['make ocean healthier', -0.0591]],	0.0417274616
[['fancy hotels modern', -0.1081], ['save hundreds trees', -0.0532], ['city low poverty', 0.1173], ['textbooks reused efficiently', 0.0237], ['richest cities world', -0.0605]],	0.0108892376
[['water improving quality', 0.2356], ['countries abundant water', 0.3622], ['goal make goal', 0.3602], ['2030 think sdgs', 0.2358], ['2015 united nations', 0.0465]],	0.0062382704
[['enormous storm terminated', 0.1836], ['change better 2030', 0.0851], ['frequent disasters wildfires', 0.104], ['hottest country annual', -0.0409], ['mali world hottest', 0.2899]],	0.0133070624
[['massive plastic products', -0.1222], ['family turn garbage', 0.0959], ['food lunch restaurants', -0.1328], ['public environmental awareness', 0.0113], ['global warming undeniably', -0.1416]],	0.0120359224
[['remember eating delicious', -0.0759], ['farming lager problem', 0.1326], ['billion tons food', 0.0853], ['better future today', 0.1735], ['fish feed day', 0.0809]]	0.0090386536
	0.0071680256

Findings

For AI keyword extraction for each document in the group, the resulting 3-ngram keywords angles return an average variance of 0.01920732786 (second highest overall)

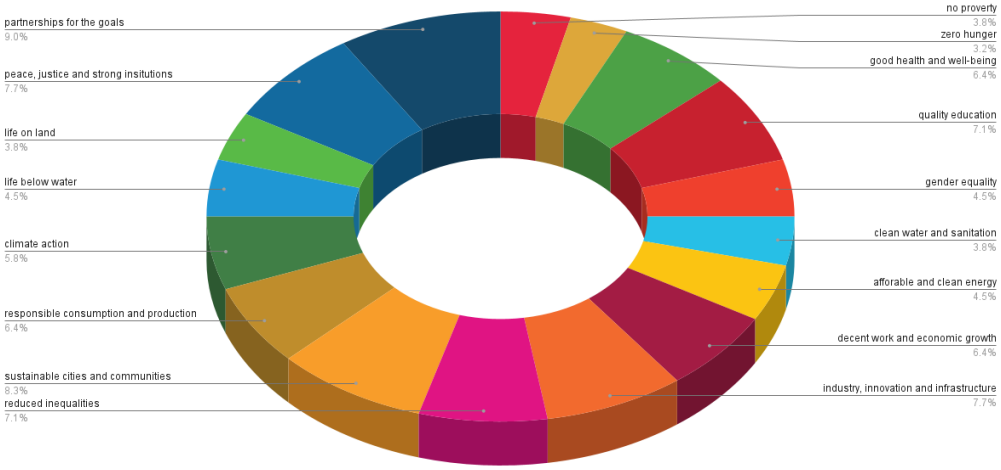
Methods 1/2/4 Results for High school:

Output Method 1 (top 10 keywords)	Findings
['one', 16], ['people', 16], ['world', 15], ['sustainable', 12], ['problem', 12], ['future', 12], ['time', 12], ['environment', 11], ['being', 10], ['life', 10]	Highest frequency words are 'one', 'people', 'world', found in ~94% of articles; next 5 highest frequency words are ' sustainable', 'problem', 'future', 'time', 'environment' found in ~71% of articles; next 2 highest frequency words are 'being' and 'life' found in in ~59% of articles.

Output Method 2 (17 sdg goals)		
#1	6	poverty : 2 income : 1 distribution : 2 wealth : X wealthy : X socio : X poor : 1 poorest : X money : 2 homeles : X prosperity : 2 microfinance : X
#2	5	hunger : 1 hungry : X agriculture : X agricultural : X nutrition : X nutrient : X famine : X food : 6 farm : 1 farming : X starving : 1 starvation : 2 malnutrition : X malnourish : X meal : 1 crop : X
#3	10	health : 4 healthy : X healthcare : 1 healthier : 1 medicine : X medical : X death : 2 die : 3 pandemic : 1 disease : 2 doctor : X physician : X nursing : X nurse : X sick : X sicknesse : X ill : X illnesse : X treatment : 1
#4	11	education : 6 educate : 3 school : 6 teacher : 1 student : 5 kid : 1 young : 5 skill : 1 teaching : 1 children : 3 learn : 3 learning : 1 vocational : X
#5	7	gender : 5 woman : 2 women : 4 girl : 3 empower : 2 female : 4 male : 2 sex : 2 sexism : 2 genital : 1 genitalia : X men : 5 man : 4 marriage : 1 marry : X married : 2 feminine : X femininity : X feminism : 1 feminist : X trafficking : X
#6	6	water : 6 sanitation : 1 sewage : X drain : X pond : X lake : X river : 1 toilet : 1 rainwater : 1 bathing : 1 drinking : 1 aquifer : X desalination : 2 diarrheal : 1 hygienic : X wastewater : 1
#7	7	energy : 6 power : 4 solar : X hydrogen : 1 renewable : 2 wind : X hydro : X electricity : 1 wave : X light : 2
#8	10	job : 2 economic : 5 economy : 2 economie : 2 growth : 2 employment : 1 employee : 2 employed : 1 unemployment : X welfare : X gdp : 1 labor : X wage : 1 migrant : X worker : 3 slavery : X
#9	12	industry : X industrie : 2 industrial : 2 innovate : X innovation : 2 infrastructure : 1 sector : 4 region : 2 regional : X domestic : 3 international : 3 digital : X internet : 3 mobile : 1 factory : 1 factorie : 1 scientific : X technology : 6 technologie : 3 technological : X
#10	11	inequality : 5 inequalitie : 2 unequal : 1 equal : 3 discriminate : X discrimination : 2 inclusion : X inclusive : X country : 7 countrie : 8 clas : 1 disability : X disabilitie : X disabled : X ethnic : 1 ethnicity : X racism : X opportunitie : 1 opportunity : X assistance : 1
#11	13	city : 3 citie : 3 settlement : X planning : 1 space : 1 local : 8 building : 3 overcrowded : X slum : X community : 8 communitie : 2 neighbour : 1 house : X housing : X disaster : X transport : 2
#12	10	consumer : X consumption : 2 consume : 1 consuming : 2 recycle : 2 recycling : 2 retailer : X buy : X purchase : X production : 1 produce : 3 exploit : X waste : 5 shop : 1 shopping : 1 reuse : 1 reusable : 2 reused : X supply : 1 fashion : X
#13	9	climate : 5 weather : 1 temperature : 2 celsiu : 1 warm : 1 warming : 3 hot : 1 atmosphere : 2 carbon : 3 dioxide : 1 emit : X emission : 5 emitted : X emitter : X greenhouse : 3 ice : 1 methane : 1
#14	7	fishing : X fish : X overfishing : X sea : 5 ocean : 5 marine : 3 island : 1 whale : 2 turtle : X coral : 1 acidification : 1 aquatic : 1 aquaculture : 1
#15	6	land : 1 animal : 1 mountain : 1 forest : 1 deforestation : 1 biodiversity : 1 biodiverse : X ecosystem : 3 ecology : X drought : X flood : X desert : X hunting : X wildlife : X specie : 5 soil : 1
#16	12	peace : 2 peaceful : 2 justice : X fair : 3 accountable : X bribery : X corruption : X institute : 1 institution : 1 security : 1 police : X law : 2 legal : X un : 3 conflict : X crime : 2 violence : 2 transparent : X
#17	14	cooperation : 1 cooperative : 1 together : 5 forum : 1 platform : 2 discus : 1 international : 3 national : 2 nation : 8 partnership : 1 volunteering : X coordinated : X organisation : 1 organization : 6 society : 7

Findings

Out of 17 SDG goals, the goal with the highest count is #17 Partnership for the Goals; while goal with the lowest count is #2 No Hunger.



Output Method 4 (BERT)	variance
[['biodegradable plastics seawater', -0.1027], ['products better ocean', 0.0514], ['worldwide ocean totally', 0.3029], ['2030 pouring trash', 0.0605], ['big truck trash', -0.1474]],	0.024979890
[['healthcare myriad', 0.3829], ['future looks unsurprisingly', 0.1515], ['environmentally friendly macau', -0.024], ['goals better brighter', 0.143], ['nonentity poverty potable', 0.2368]],	0.017634538
[['travel demands advanced', -0.0643], ['home smarter mobility', 0.0557], ['pooling cheaper reduces', -0.0048], ['traffic accidents caused', 0.0745], ['air pollution caused', 0.0144]],	0.002375516
[['outnumbered females 34', -0.184], ['feminism women recently', -0.1555], ['sexism big', 0.0233], ['family members sexism', 0.0502], ['learn gender equality', -0.0044]],	0.009299062
[['inequality important goals', 0.0511], ['fighting women right', 0.3196], ['education courageous women', 0.0044], ['world gender inequality', 0.099], ['women recent research', 0.2861]],	0.016169626
[['energy reactors powerful', -0.0256], ['global warming', 0.1616], ['factors building green', -0.0635], ['save money environmental', -0.0518], ['reduced global warming', -0.0221]],	0.006794686
[['men gang raped', -0.1606], ['old girl florida', 0.0663], ['public schools transphobia', 0.116], ['homemakers instead breadwinners', -0.1263], ['200 million girls', -0.0736]],	0.011732330
[['2021 https blog', -0.2726], ['hunger facts statistics', -0.1278], ['food waste restaurants', -0.1896], ['aggravates global warming', -0.0102], ['millions tourists month', 0.0139]],	0.011627894
[['politicians actors discredit', 0.0026], ['misinformation effectively finally', 0.1851], ['news travel faster', -0.0282], ['social media underestimate', 0.0777], ['influential shaping world', 0.0449]],	0.005440246
[['billion year scientists', 0.0928], ['rampant deforestation soil', 0.1078], ['increasing number vegetarians', -0.0661], ['climate change efficient', -0.0141], ['needs man greed', 0.048]],	0.004286598
[['debris floating ocean', -0.1792], ['ocean organisms seriously', -0.0234], ['trash flowing ocean', -0.1612], ['bleaching chemical pollution', -0.2258], ['pollution dire needed', -0.0083]],	0.007640658
[['sex education finally', 0.2658], ['sadly women lost', 0.4582], ['afghanistan august 2021', 0.2729], ['public breastfeeding unjustifiable', -0.2228], ['female employees afghanistan', -0.1228]],	0.066990526
[['hope injured war', 0.1082], ['world sustainable green', -0.0869], ['future brighter objectives', 0.246], ['degradation caused anthropogenic', 0.3874], ['environmental city innovations', -0.0591]],	0.032479710
[['disgusting floating mom', 0.1839], ['whale notebook mother', -0.057], ['play whales dolphins', -0.0612], ['dumping unused whales', -0.035], ['jellyfish tasted disgusting', -0.3202]],	0.025560928
[['better conditions planet', -0.0983], ['brothers sisters world', 0.0528], ['projects benefit community', 0.1651], ['partnership brothers sisters', 0.177], ['environmentally kind future', 0.2126]],	0.012875914
[['new water resources', 0.1087], ['bathing bathtub day', -0.0301], ['water continues increase', 0.0636], ['safe drinking water', -0.0991], ['water rationing necessary', -0.0494]],	0.005803978
[['global warming simple', 0.1595], ['year dump massive', 0.2929], ['undeniable ocean heat', -0.0053], ['weathers caused global', 0.1756], ['increasing ocean acidification', -0.1149]]	0.020744790

Findings

For AI keyword extraction for each document in the group, the resulting 3-ngram keywords' angle returns an average variance of 0.01661393478 (third highest overall).

Methods 1/2/4 Results for University:

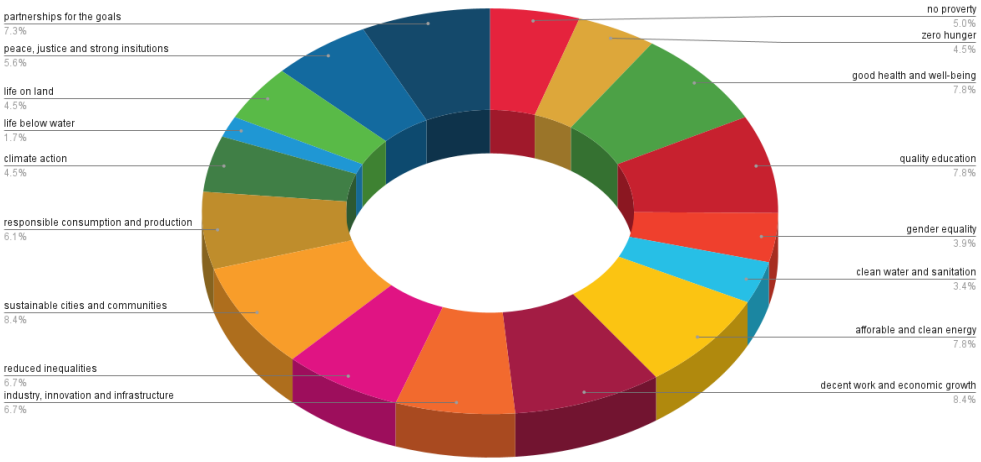
Output Method 1 (top 10 keywords)	Findings
['one', 17], ['sustainable', 16], ['people', 16], ['development', 15], ['world', 14], ['time', 13], ['better', 13], ['goal', 13], ['human', 13], ['2030', 12]	Highest frequency words are 'one', 'sustainable', 'people' found in ~100% of articles; next 6 highest frequency words are 'development', 'world', 'time', 'better', 'goal', 'human' found in ~88% of articles; next 1 frequency word '2030' is found in ~71% of articles.

Output Method 2 (17 sdg goals)

#1	9	poverty : 3 income : 2 distribution : 1 wealth : 1 wealthy : 1 socio : X poor : 6 poorest : X money : 4 homeles : 1 prosperity : X microfinance : X
#2	8	hunger : 1 hungry : 2 agriculture : 1 agricultural : 1 nutrition : X nutrient : X famine : 2 food : 6 farm : 1 farming : X starving : X starvation : 1 malnutrition : X malnourish : X meal : 2 crop : 1
#3	14	health : 4 healthy : 4 healthcare : X healthier : 1 medicine : 2 medical : 6 death : 2 die : 2 pandemic : 8 disease : 2 doctor : 4 physician : 1 nursing : 1 nurse : 1 sick : 1 sicknesse : X ill : 1 illnesse : 1 treatment : 6
#4	14	education : 5 educate : 1 school : 6 teacher : 3 student : 6 kid : 2 young : 3 skill : 2 teaching : 2 children : 8 learn : 5 learning : 3 vocational : 1
#5	7	gender : 5 woman : 1 women : 4 girl : 5 empower : X female : 3 male : 2 sex : 1 sexism : 1 genital : X genitalia : 1 men : 3 man : 4 marriage : 1 marry : X married : 1 feminine : 1 femininity : 1 feminism : X feminist : 1 trafficking : 1
#6	6	water : 4 sanitation : X sewage : X drain : 1 pond : X lake : X river : X toilet : 2 rainwater : X bathing : X drinking : 1 aquifer : 1 desalination : X diarrheal : X hygienic : 1 wastewater : X
#7	14	energy : 7 power : 6 solar : 2 hydrogen : X renewable : 3 wind : 3 hydro : 1 electricity : 3 wave : 1 light : 7
#8	15	job : 2 economic : 8 economy : 2 economie : X growth : 7 employment : 2 employee : 2 employed : X unemployment : 1 welfare : 1 gdp : X labor : 1 wage : X migrant : X worker : 3 slavery : 1
#9	12	industry : 3 industrie : X industrial : 2 innovate : 1 innovation : 2 infrastructure : 1 sector : X region : 3 regional : 1 domestic : X international : 2 digital : 2 internet : 2 mobile : X factory : X factorie : 1 scientific : 1 technology : 6 technologie : 2 technological : 1
#10	12	inequality : 4 inequalitie : 2 unequal : 2 equal : 5 discriminate : X discrimination : 3 inclusion : 1 inclusive : 3 country : 2 countrie : 4 clas : 1 disability : X disabilitie : 1 disabled : 1 ethnic : 2 ethnicity : X racism : X opportunitie : 5 opportunity : X assistance : 2
#11	15	city : 6 citie : 6 settlement : 2 planning : 1 space : 5 local : 3 building : 5 overcrowded : 1 slum : 1 community : 6 communitie : 4 neighbour : X house : 3 housing : 1 disaster : 1 transport : 1
#12	11	consumer : 1 consumption : 4 consume : 3 consuming : 2 recycle : 1 recycling : 1 retailer : 1 buy : 2 purchase : 3 production : 3 produce : 2 exploit : 1 waste : 6 shop : X shopping : 2 reuse : 2 reusable : 1 reused : X supply : 1 fashion : 3
#13	8	climate : 4 weather : 1 temperature : 2 celsiu : X warm : 2 warming : 4 hot : X atmosphere : X carbon : 5 dioxide : 2 emit : 1 emission : 6 emitted : 1 emitter : X greenhouse : 2 ice : 1 methane : X
#14	3	fishing : X fish : X overfishing : X sea : 2 ocean : 1 marine : 1 island : X whale : X turtle : 1 coral : X acidification : X aquatic : X aquaculture : X
#15	8	land : 5 animal : 4 mountain : 1 forest : 3 deforestation : 2 biodiversity : 1 biodiverse : X ecosystem : 2 ecology : 2 drought : X flood : 2 desert : 1 hunting : 1 wildlife : 1 specie : 2 soil : 1
#16	10	peace : 4 peaceful : 3 justice : 2 fair : 1 accountable : 1 bribery : 1 corruption : 1 institute : 1 institution : 1 security : 1 police : 1 law : 2 legal : X un : 4 conflict : 2 crime : 2 violence : 4 transparent : 1
#17	13	cooperation : 2 cooperative : X together : 7 forum : X platform : 1 discus : X international : 2 national : 2 nation : 7 partnership : X volunteering : 2 partnership : X coordinated : X organisation : X organization : 5 society : 6

Findings

Out of 17 SDG goals, the goals with the highest counts are #8 Decent Work and #11 Sustainable Cities and Communities; while goals with the lowest count is #14 Life Below Water.



Output Method 4 (BERT)	variance
[['unfair treatment women', 0.0726], ['coronavirus outbreak exacerbates', 0.103], ['men persecution women', 0.0787], ['women laidbare coronavirus', 0.1822], ['girl suffered circumcision', -0.0966]], [['2012 tropical deforestation', 0.0093], ['global warming climate', -0.1051], ['cut trees faster', -0.0279], ['deforestation people cut', 0.4548], ['destroying entire ecosystem', -0.034]], [['father hybrid rice', -0.0489], ['robot nanny', 0.1642], ['sniper battle new', 0.4423], ['science fiction essays', -0.089], ['2030 epidemic completely', 0.0796]], [['hope better change', 0.0806], ['coming tour buses', 0.1099], ['bicycle sharing prioritized', -0.1005], ['macau reasons bicycle', 0.1529], ['goal building renewable', 0.2016]], [['world better better', 0.0207], ['health global', 0.0835], ['health problems important', -0.035], ['war fair world', 0.3388], ['development 2030 agenda', 0.078]], [['want change 2030', 0.1802], ['precarious environmental pollution', 0.2527], ['hunting killing wildlife', 0.2912], ['food crazy hunting', 0.2957], ['faster city representatives', 0.0933]], [['future unpredictable magic', -0.0736], ['work climate change', 0.0307], ['earth perspective 2030s', -0.0207], ['moviethe day tomorrow', 0.367], ['global warming global', 0.3926]], [['unfortunate female employee', -0.0583], ['international murder violence', 0.0188], ['leaders discussing company', -0.2372], ['starvation want threats', -0.0122], ['blueprint winning businesses', -0.0858]], [['science technology tourism', 0.2352], ['challenge tourist cities', 0.3302], ['threatens sustainability tourism', 0.3455], ['world 2030', 0.3839], ['2030 live better', 0.3045]], [['million refugees food', -0.1648], ['experience volunteering community', -0.224], ['year old sister', 0.1226], ['com amazon amazon_destruction', -0.1906], ['boy wished police', -0.1007]], [['sustainable reasons influencing', 0.356], ['tasting bitter fruits', -0.0295], ['chemicals global warming', 0.0404], ['pandemic necessary replace', 0.1754], ['students environmental awareness', -0.1122]], [['construct new home', 0.0247], ['nurse constantly considering', 0.2131], ['2030 jobs perspectives', 0.0842], ['growth diseases improving', 0.0111], ['world better sustainable', 0.132]], [['grandmas went shopping', 0.1203], ['week recent research', 0.018], ['monster soap dispenser', -0.0265], ['apartment beijing', -0.0447], ['landfills takes 200', 0.2558]], [['life tougher game', -0.1318], ['murder mystery games', 0.1812], ['vaccines online teaching', 0.1774], ['children graduate finally', 0.0366], ['mother family handwork', 0.1112]], [['gender injustice longer', -0.0357], ['eliminate gender injustice', -0.0452], ['men women unfair', 0.1817], ['frequently brought sexism', -0.0981], ['gender inequality essential', 0.0286]], [['blue planet 2020', -0.0053], ['danger contamination rains', -0.0245], ['post epidemic reflection', 0.1081], ['destroying space human', 0.0953], ['city zootopia deerwandered', 0.1664]], [['library shared hundreds', -0.0131], ['massive unscheduled school', -0.0185], ['outstood long pandemic', 0.273], ['classes online counseling', 0.1642], ['bullying severely disadvantaged', -0.1341]]	0.008299090 0.040451254 0.035826930 0.010635908 0.016451196 0.005894366 0.039745700 0.007889326 0.002453738 0.015334680 0.027119042 0.005478990 0.012415438 0.013463082 0.009315610 0.005193000 0.021048972

Findings

For AI keyword extraction for each document in the group, the resulting 3-ngram keywords' angle returns an average variance of 0.01629507769 (lowest overall).

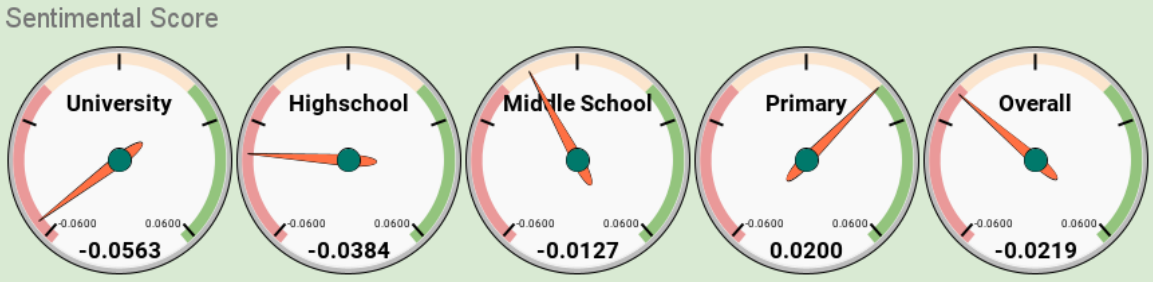
Method 5 Sentiment Analysis Results

Negative scores reflects negative sentiment
Positive scores reflects positive sentiment
The greater the magnitude the more negative / positive sentiment

Groups	University	Highschool	Middle School	Primary
Individual Sentiment values (sorted)	-0.1311688312 -0.1146907216 -0.1086387435 -0.0912162162 -0.0811724915 -0.0811724915 -0.0796645702 -0.0612500000 -0.0596707819 -0.0466392318 -0.0425764192 -0.0369897959 -0.0363636364 -0.0145208132 -0.0025940337 0.0106951872 0.0207522698	-0.1367127496 -0.1009036145 -0.0971014493 -0.0935143288 -0.0892586989 -0.0864779874 -0.0512422360 -0.0426229508 -0.0376068376 -0.0323325635 -0.0318559557 -0.0125698324 -0.0105263158 0.0099626401 0.0451388889 0.0451612903 0.0688622754	-0.144495413 -0.082969432 -0.070663812 -0.063180828 -0.049356223 -0.048611111 -0.035545024 -0.027363184 -0.026666667 -0.002159827 0.005025126 0.013544018 0.023668639 0.028953229 0.060267857 0.062770563 0.141176471	-0.151750973 -0.106719368 -0.063291139 -0.046875000 -0.045977011 -0.031872510 0.000000000 0.004444444 0.004608295 0.008968610 0.034749035 0.046875000 0.049808429 0.078740157 0.132075472 0.170731707 0.254901961
Average	-0.0562871365	-0.0384470839	-0.012682683	0.019965712

Findings

Overall the average sentimental value is -0.0218627979 which is generally negative sentiment.



Interestingly an increase in positive sentiment corresponds to a drop in age, in other words primary students' writings have much more positive outlook than university students, something we should really reflect on.

Discussion

These machine analyses are making quantitative measures on writing, in another word, **scoring words.** Even the complex AI analyses are just assigning values using multiple cascading techniques and levels of value adjustments.

Although we are analyzing using code and mathematical models, there are still decisions between being analytical and being justifiable, inevitably bias seeps in and it is important to be careful of overfitting and directed bias. Whenever that happens, it is good to step back and try a different method to counterbalance, we also try to clarify such situations, so that insight can still be obtained while providing an objective view.

Conclusions and Recommendations

With these caveats in mind, there is utility in **using machine analysis to summarize insights from large corpus of texts,** especially written by different ages with numerous concepts. First, it is a strain for anyone to manually review or to do so consistently. Even if multiple judges divide the review with a scoring matrix, each judge has their own bias and expertise affecting how they score, there is also a need to consolidate different scores. In fact, this analysis also includes the combined view of the contest judges, as these articles were picked over others.

Another advantage machine analysis has over human review is the possibility to reproduce results. This is not a trivial point, it means that work can be checked, re-execute on other datasets, and more importantly allow others to make their own adjustments and make improvements.

A good place to start is the simpler method like word count, it is easy to pick up and the drawbacks will appear quickly. More advanced methods require understanding of some Natural Language Processing (NLP) jargons, as this field is extensively studied and many combinations of techniques exist. Also, we found that it is useful to separately analyze the age groups.

We try to maintain objectivity and provide data so the reader can draw their own conclusions, and encourage others to clone the code, decide how they want to change it to achieve results while avoiding bias susceptibility; **our experience is that this kind of exercise will teach you about your own biases about sustainability**.

Written communication is not done with text scoring, be it with humans or machines, it is about conveying emotions and reasoning against prejudices, beliefs and preconceptions.

> *The effectiveness of a writer to speak to a reader; depends on the readers' reception of the language used, the emotional state and what they value or deem important, at that moment.*

In fact, we will conclude with a plea for more sustainability writing from all ages and backgrounds, regardless of score, because writing is inherently subjective. Let's reach more readers and launch more sustainability writers.

Contributors to this paper received one edition of the globe NFT on the first page