Educational and Information Technologies (2023) 28:10965–10994
https://doi.org/10.1007/s10639-022-11493-8

Analysing the sentiments about the education system trough Twitter

Mary Luz Mouronte-López[1] · **Juana Savall Ceres**[2] · **Aina Mora Columbrans**[1]

Received: 18 July 2022 / Accepted: 28 November 2022 / Published online: 10 February 2023

Abstract
This paper applies Information and Communication Technologies (ICT) as well as data analysis to gain a better understanding of the existing perception on the education system. 45,278 tweets were downloaded and processed. Using a lexicon-based approach, examining the most frequently used words, and estimating similarities between terms, we detected that a predominantly negative perception of the education system exists in most of the analysed countries. A positive perception is identified in certain low-income nations. Men exhibit a more positive sentiment than women as well as a higher subjectivity in some countries. The countries that exhibit the most positive perceptions India, Canada, Pakistan, Australia, South Africa and Kenya are also those that manifest the highest subjectivity.

Keywords Education system · Twitter · Sentiment analysis · Data science · Social networks · Perception of education

1 Introduction

Social networks have become one of the main forms of social media where citizens can express opinions on different matters (Ayala, 2014). Twitter, Facebook and Insta

✉ Mary Luz Mouronte-López
maryluz.mouronte@ufv.es

Juana Savall Ceres
j.savall@ufv.es

Aina Mora Columbrans
ainamora@hotmail.com

1 Higher Polytechnic School of Universidad Francisco de Vitoria, Ctra. Pozuelo a Majadahonda Km 1.800, Pozuelo de Alarcón, 28223, Madrid, Spain

2 Faculty of Education and Psychology of Universidad Francisco de Vitoria, Ctra. Pozuelo a Majadahonda Km 1.800, Pozuelo de Alarcón, 28223, Madrid, Spain

gram are among the most popular social networks. Twitter had 217 million daily active users by February 2022 Twitter by the numbers: stats (2022). Facebook had 2.934 billion monthly active users by July 2022. However, the total number of persons that make use of this network each month have gone down by roughly 2 million (– 0.07%) during the three months prior to July (Facebook Statistics and Trends, 2022). Instagram had over 1.44 billion monthly active users by August 2022 The Small Business Blog (n.d.).

On Facebook, individuals can stay in touch with each other, receiving e-mails or alerts on their cell phones whenever someone posts a status update (Amidon, 2011). Instagram is utilised to share both photos and videos. Additionally, it also provides tools that enable the modification of pictures (Jackson & Luchner, 2018). Twitter users have been able to post messages of up to 280 characters each since the end of September 2017 (previously the size was 140 characters). This limit establishes many of the attributes of the language utilised in the tweets (Farzindar & Inkpen, 2015). Twitter users send messages about any topic within of the aforementioned limit and lead others to receive their tweets. In addition to posting, Twitter users can reply, retweet and quote[1]. Neither Facebook nor Instagram allow for a similar textual exchange of messages, but instead allow for annotations to be made on the content. Twitter, which requires its users to be 13 years of age or older, has a tool that is not only used for disseminating information, but also for raising awareness about issues affecting society (Quadri et al., 2018).

Compared to the access to information handled on Instagram or Facebook, Twitter data is more readily available through robust Application Programming Interface (API) (Macy et al., 2015; Weller et al., 2014). This allows researchers to obtain a significant volume of tweets with extremely useful metadata, such as type of tweet, longitude and latitude (Weber et al., 2021), or other textual description related to the location from which the tweet was posted. The used hashtags in the tweets as well as the relations between tweets, in addition to other data, can also be retrieved (Weber et al., 2021).

The study of social networks makes it possible to explore social trends and behaviors, as well as to help to create a collective image about different realities. Social networks are nowadays a highly disseminated loudspeaker for all kinds of events and news (Beykikhoshk et al., 2015; Karami et al., 2018). These platforms have a great deal of potential in facilitating guided decision-making by governments and institutions.

In this paper we use the social network Twitter to examine the public perception of education systems in different countries.

1.1 Sentiment analysis

Classifying and quantifying the sentiments expressed in the social networks allows researchers to analyse opinions that are often not expressed in more formal settings.

[1]To reply is answered to another users's tweet, to Retweet is the way one sends another user's tweet to their followers, and to quote is to retweet writing an additional comment.

This is especially apparent when the points of view are related to sensitive issues that are affected by the social desirability. The emotional dimension of a controversial issue is often developed more dynamically and unpredictably in a social network than in an informal face-to-face discussion (Kušen et al., 2017). The sentiment analysis thus becomes a relevant resource both for capturing the pulse of society on certain matters and for acting as a tool to both raise awareness and for predicting future behaviour (Kalyanam et al., 2016).

Sentiment analysis is a widely studied topic (Mohammad et al., 2018; Kharde & Sonawane, 2016), being used in areas as diverse as the stock market, the healthcare system, and the field of business, among others. In the area of the stock market, (Bharathi & Geetha, 2017), the sentiments of common people are united through their news feeds and combined with data related to the Index of the Bombay Stock Exchange in order to anticipate the behavior of the current stock market. A model is suggested to forecast the stock prices based on the combining of a Self-Organizing Map (SOM) and fuzzy – Support Vector Machines (f-SVM) (Naseem et al., 2018). Uhr et al. (2014) applies the sentiment analysis on financial market news implementing a software prototype for computing results on different levels of exploration, splitting text into sentences and subsets of tokens from sentences. SentimentWortschatz (SentiWS) software (Remus et al., 2010) was used. Kolasani and Assaf (2020) predicts the future movement of the United States' stock market by analysing the sentiment of Twitter posts related to the stock market. The average sentiment value of the tweets is obtained using a Support Vector Machines Model (SVM). In the healthcare field, (Balakrishnan et al., 2021) examines the effectiveness of deep learning methods to identify the sentiment modification in breast cancer patients' narratives. A sentiment change study was carried out to estimate the modification in the satisfaction level of the patients a deep learning model bidirectional Long Short-Term memory (LSTM) with sentiment embedding features, which provided the best result. Hopper and Uriyo (2015) utilised the sentiment analysis technique to review patient feedback for a group of gynecologists, where time-to-next-complaint procedures were used. In the business area (Suerdem & Kaya, 2015) emulates conventional attitude and market research tools by using Naïve Bayes and Support Vector Machines models in the Turkish language context. Rajkumar et al. (2019) utilizes Naïve Bayes and Support Vector Machines models to categorize product reviews as positive or negative.

Specifically, Twitter's sentiment analysis has been used in a wide range of areas related to public trust, ranging from predicting election outcomes to examining resentment against government policies (Arcila-Calderón et al., 2017; Calderón et al., 2015; Blasco & Coenders, 2020). It has also been utilised in the study of migration (Arguedas et al., 2020; Arcila-Calderón et al., 2020), disability (Gómez-Marí et al., 2022) or diseases caused by the Ebola virus (Percastre-Mendizábal et al., 2019) or SARS-CoV-2 (Sued & Cebral, 2020).

Sentiment analysis is a procedure that mechanically detects attitudes, opinions, or emotions from text, in which several approaches can be utilised to conduct the analysis: Vohra and Teraiya (2013):

- A knowledge-based procedure, in which a words lexicon showing positive or negative perceptions is built. Some heuristics can also be used to add more words to the initial lexicon (Musto et al., 2014; Bonta et al., 2019; Mehmood & Balakrishnan, 2020).
- Relationships based approach, in which the relationships between certain parts of the text, whose sentiment were previously analysed, are used in order to identify the existing opinions in the principal unit. Additionally, all those connections that can exist between elements in relation to their meaning can be examined (Alrasheed, 2021; Janda Janda et al., 2019).
- Language models implementation. The majority of existing lexicons contain only unigrams along with their sentiment scores. It has been observed that sentiment n-grams formed by combining unigrams with intensifiers or negations sometimes show improved results in the sentiment classification task. This has led to the development of new methods that process frequent n-grams as features in addition to single words (Dey et al., 2018; Walaa et al., 2014; Kaur et al., 2021).

1.2 Motivation and objectives of the study

Our purpose is to answer the following research questions: What is the predominant perception towards the education system?. Are there differences in these sentiments between sexes? To answer them we use Twitter. This is due both to the unique characteristics of this social network and to the evidence suggesting that this site has been one of the most studied social networks in the academic research (Weller et al., 2014; Agarwal et al., 2011; Kharde & Sonawane, 2016). The goals of this research are:

1. Regarding education systems, to identify the sentiments expressed globally, on Twitter, as well as by gender. The perceptions of education are globally examined on Twitter, considering those countries from which more than 50 tweets were posted and where the gender of the author as well as the country of origin of the tweet were identified during the analysis period.
2. To detect commonalities and differences between countries with the highest number of tweets.
3. To compare the results of points 1 and 2 with reports that may have been carried out at an international level on the perception of the education system by educators and users.

Concerning perceptions on the education system, various survey-based studies have been performed with the aim of gaining in-depth knowledge of students' perceptions connected to the educational system. The World Innovation Summit for Education (WISE 2020) published a comprehensive survey of young people's views[2] on their education and future. Also, in the PISA 2018 (OECD, 2019a) Report, for the first time, 15-year-old students were asked how they normally feel about their lives. While this study focuses on the perception of the students themselves, our research covers a

[2]The WISE global survey was conducted to young people between the ages of 16 and 25.

much wider age range (the younger and the adult populations expressing their opinion on Twitter).

It should be noted that the determining perceptions of the education system of students and teachers has been the purpose of several pieces of research, these have aimed to solve problems related to the quality of teaching and learning, as well as unemployment-related and job dissatisfaction issues among university graduates. WISE (2020) and Ipsos Group S.A (2021) examined the perception of the students themselves towards the education system. Existing research often uses structured empirical surveys collecting qualitative and quantitative information (Safari & Barigye, Sydney et al., 2021; Kartal et al., 2015).

Public opinions on education have also been analysed theoretically (Benjamin, 1993), as well as the influence of the media on such opinions. Greater and more effective publicity about the role played by educational centers in society has been required (Benjamin, 1993). An analysis recently published (Cheeti, 2021), utilised Twitter sentiments analysis to explore the impact of the COVID-19 pandemic on education. However, the geographical dimension was not considered. An analysis by origin country of the posted tweet, as well as by gender, is carried out in our investigation.

With respect to gender inequalities, (UNESCO, n.d.) states that significant gender differences exist in access, school performance and continuity in education. Girls are at a disadvantage compared to boys in many countries. UNESCO points out that 16 million girls will never receive an education (UNESCO, n.d.). Geographic isolation, poverty, minority status, disability, early marriage and pregnancy, gender-based violence and traditional attitudes about the status and role of women are all among the main causes of inequality (UNESCO, n.d.). Due to the above, it is relevant to know if gender differences exist in perceptions on the education system. Dissimilarities between the opinions of men and women have already been detected in various social issues such as the most important environmental problems (Dreiling & Belkhir, 1997; Mouronte-López & Subirán, 2022), the epidemic of COVID-19 (Galasso et al., 2020), corruption (Bauhr & Charron, 2020), or leisure(Fontenelle & Zinkhan, 1993).

2 Materials and methods

2.1 Overview used resources

2.1.1 Software programs

T-Hoarder tool T-Hoarder tool (Congosto et al., 2017) was used for the downloading of data from Twitter. This tool is a piece of software that carries out tweet crawling, data filtering, and shows a synopsis of information about Twitter activity on a certain topic. The tool is available for the UNIX operating system, having been developed in Python language (Van Rossum & Drake, 2009).

T-Hoarder uses two API provided by Twitter to retrieve data: REST (REpresentational State Transfer) API and the Streaming API. The REST API makes it possible to perform all kinds of queries on Twitter data synchronously. Utilising this API,

Table 1 Information provided by T-Hoarder for each tweet (Congosto et al., 2017)

Field	Description
`idtweet`	Unique number that is given sequentially to each message by Twitter, allowing its characterization.
`timestamp`	Date and GMT hour in which the tweet is posted.
`@author`	A screen name symbolising the author of the tweet.
`text`	Text content of each tweet.
`app`	Application from which a tweet was posted.
`id_author`	Sequential number which is allocated by Twitter to each user, when he or she first registers on the social network.
`followers_author`	Number of followers that the user has at the moment that he/she sent the tweet.
following_author	Number of users followed by the user at the moment in which he/she sends the tweet.
statuses_author	Number of tweets posted by the user.
`location`	Location that the user registered on his/her profile.
`url`	If an URL is included in the tweet, this field contains its value, otherwise it is empty.
`geolocation`	If the tweet is geolocated, this field includes the longitude and latitude corresponding to the place where the user is located. Otherwise, the field is empty.
`name`	Name provided by the author's tweet.
`bio`	Short description related to the author's tweet.
`url_media`	If the tweet contains multimedia information, its url is included in this field, otherwise this field is empty.
`type_media`	If the tweet includes multimedia information, this field describes the type of the multimedia information, for instance photo or video, otherwise this field is empty.
`lang`	Language in which the tweet was written.

only tweets posted in the week prior to the time in which the API is utilised can be obtained (Congosto et al., 2017).

The Streaming API establishes a socket[3] between Twitter and a server through which information is received asynchronously. This API allows us to download tweets in real time. T-Hoarder utilises the `Tweepy` library (Twitter, n.d.), which resolves low level access to the aforementioned API (Congosto et al., 2017).

Using T-Hoarder we download tweets both synchronously and asynchronously, through the REST and the Streamming API. Utilising T-Hoarder the information shown in Table 1 can be obtained by each tweet.

[3] A bidirectional communication link.

In-house software

Several programs were developed in R (n.d.) and Python programming languages, where the following features were implemented:

Data processing which includes all that is indicated below. More detail on the utilised procedures can be found in the Overview of used methods Section.

- Elimination of all duplicate tweets, exclusion of replied, retweeted and quoted category interactions since they did not contain personal opinions. Only tweeting type messages were considered.

  In order to optimise the processing, similarly to Mouronte-López and Subirán (2022), certain substitutions were made: all URLs were replaced by the word "LINK", the word `"USER"` took the place of all users' mentions, while hashtags were replaced by the word "HASHTAG". Emojis were also translated into text since they contain information about the sentiment of the tweet. The punctuation marks, articles, conjunctions, and words with a length lower than 3 letters were removed since they do not provide any such information. A reduction of words to their root term (stemming) was performed. Additionally, all inflected forms of the same word were replaced by its lemma (lemmatization). The correction of misspelt words was also carried out. Finally, the text was converted into lowercase.

  The program that carried out the mentioned functionalities was implemented in Python language. `NLTK` and `pandas` libraries were used.
- Analogously to Mouronte-López and Subirán (2022), for each processed tweet, the gender of its author was obtained. Longitude and latitude, as well as the textual description on the localization indicated in each tweet, were also used to gain knowledge of the country from which each tweet was posted.

  The program that implemented this feature was developed in Python language. The `Nominatim` geolocation service and the online service `genderize.io` were used.
- For each tweet, the sentiment corresponding to the processed text was obtained. This program was also implemented in Python language. The `TextBlob`[4] library was utilised.

Performing statistical calculations the following statistical calculations were carried out:

- Percentage of tweets with positive/negative sentiment in each country.
- Mean, Standard Deviation, Median and Mode of both sentiment and subjective exhibiting tweet text in each country.
- Construction of a graphical representation of the information in two-dimension (heat-map). Several colors were used to show different levels of analogy between countries, considering a specific similarity metric.
- Obtaining the most frequently used words.

This program was implemented in R language. The `R.utils` library was used.

[4]It provides a simple API for performing natural language processing tasks.

2.2 Overview of used methods

2.2.1 Downloading and processing of tweets

We downloaded only tweets in the English language from March 03^{rd} until August 31^{st} in 2021. This period was chosen because the project was granted in the first quarter of 2021 and results had to be obtained during the last quarter of that year.

For the analysis of sentiments about the education system, tweets including the sentences: `"education system"` and `"educational system"` were chosen. As previously stated, Twitter users can perform various kinds of interactions. However, because we aimed to detect individuals' sentiment, tweets exclusively related to `tweeting` interaction are analysed.

As previously indicated, similarly to (Mouronte-López & Subirán, 2022), a processing of the tweets is executed in which emoticons, usernames, punctuations, and links are eliminated. The text is tokenized, stemmed, lower-cased and only words with a length longer than 2 letters are utilised for the analysis. Duplicate tweets, if any, were also removed in the procedure. After this process, 45,278 tweets were obtained for the analysis of the education system.

2.2.2 Detecting the geographic location from which each tweet is sent and identifying the gender of its author

Analogously to (Mouronte-López & Subirán, 2022), to mechanically obtain the gender for each tweet, the online service `genderize.io` was utilised. It is an API, which anticipates the gender of a person from their name, which is registered in the `name` of each tweet.

To identify, based on geolocation data, the country from which each tweet was produced, the Open Street Map (OSM) `Nominatim` geolocation service was utilised. Similarly to (Mouronte-López & Subirán, 2022), we assess the tweeting location not only by inspecting the *geolocation* field but also checking the *location* field in each tweet. It allowed us to maximise the number of tweets whose home country was identified.

Of the 45,278 tweets downloaded referring to the education system, 43,050 included the author's country of origin and gender. 19,726 were produced by women and 23,324 by men. The tweets corresponded to 16,030 unique users.

It can be noted that, for this investigation, only those countries from which more than 50 tweets, where it was possible to identify the author's gender and the home country, were selected for the analysis by country. We utilize all tweets in the study at a global level.

2.2.3 Evaluating similarities between countries

Regarding the messages exchanged on Twitter, the analogy between two countries was estimated by assessing the similarity between frequencies of used words. It is assumed that if tweets from two countries are written in a similar way, they will tend to contain the same words with analogous frequencies. In order

Table 2 Frequency of utilised words for $Country1$ and $Country2$

Word	Country1	Country2
$Word_1$	FreqCountry1Word$_1$	FreqCountry2Word$_1$
$Word_2$	FreqCountry1Word$_2$	FreqCountry2Word$_2$
...	...	...
$Word_m$	FreqCountry1Word$_m$	FreqCountry2Word$_m$

to evaluate this similitude, we use the $Sim2Count$ metric, which can be defined as:

$$Sim2Count = 1 - S_C \tag{1}$$

Where

$$S_C = \frac{\sum_{i=1}^{i=m} Country1_i Country2_i}{\sqrt{\sum_{i=1}^{i=m} Country1_i^2}\sqrt{\sum_{i=1}^{i=m} Country2_i^2}} \tag{2}$$

$Country1_i$ and $Country2_i$ symbolise the i component of the vectors $Country1$ and $Country2$, respectively.

$Country1_i$ represents the frequency of the word i in $Country1$ ($FreqCountry1$Word$_i$) and $Country2_i$ is the frequency of the word i in $Country2$ ($FreqCountry2$Word$_i$), this nomenclature is shown in Table 2.

The cosine similarity metric, S_C, has been used frequently in order to examine similarities between texts (Lahitani et al., 2016; Ahmed et al., 2013).

2.2.4 Identifying the polarity of tweets

As previously mentioned, to check the polarity of each tweet the `TextBlob` library in Python was used. From the processed text obtained for each tweet this library provides two different values: polarity and subjectivity. The polarity is in the [–1,1] interval and supplies the actual sentiment, and whether it is negative or positive. The subjectivity is in the [0,1] interval, and it refers to the amount of personal opinion existing in the sentence. The higher the value the more subjectivity there is in the text (Mouronte-López & Subirán, 2022).

3 Results

Tables 3 and 4 depict, for the tweets which referred to the educational system, the number of unique users and the total amount of tweets as well as by sex. A total of

Table 3 For interactions related to the educational system, number of unique users

Total	Women	Men
16,070	6,297	9,773

Table 4 For interactions related to the educational system, number of tweets

Total	Women	Men
43,050	19,726	23,324

43,050 tweets were downloaded, coming from 16,070 unique users. Of these, 60.82% were male and 39.18% female (9,773 men versus 6,297 women). Tables 5 and 6 show analogous information by each analysed country.

Tables 7, 8 and 9 show the average, the standard deviation, the median and the mode of expressed sentiments by sex and country. Figures 1 and 2 show the percentage of tweets with positive and negative sentiments by country. Figures 3 and 4 display, as bar diagrams, the median of the sentiment by country, and desegregated by sex in each nation. It can be observed that negative perceptions of education systems are prevalent according to the obtained median. This happens for both sexes as well as for all analysed countries, except in India, Canada, Pakistan, Australia, South Africa and Kenya.

Tables 10, 11, 12 and Fig. 5 depict identical information referring to the subjectivity of tweets.

In Table 9 can be observed that, according to median and mode values, the men's perception of the education system is better than that of women in all analysed

Table 5 For interactions related to the educational system, number of unique users by country and gender

Country	Number of unique users	Number of unique male users	Number of unique female users
United States	4,029	2,218	1,811
Myanmar	1,429	785	644
Singapore	963	492	471
United Kingdom	896	555	341
India	597	490	107
Canada	229	143	86
Pakistan	206	163	43
Japan	143	77	66
Australia	125	82	43
Germany	118	61	57
France	114	39	75
Thailand	81	52	29
South Africa	78	42	17
Kenya	57	47	10

Table 6 For interactions related to the educational system, number of tweets by country

Country	Number of Tweets	Country	Number of Tweets
United States	6,311	Germany	181
Myanmar	2,589	France	170
Singapore	1,705	Australia	154
United Kingdom	1,450	Thailand	138
India	674	South Africa	94
Canada	284	Korea, Republic of	69
Japan	244	Kenya	60
Pakistan	237		

Table 7 For interactions related to the educational system, statistical parameters corresponding to expressed sentiments by sex

Sex	Average sentiment	Standard deviation	Median	Mode
Women	– 0.04927	0.16496	– 0.05000	– 0.10000
Men	– 0.03352	0.18818	– 0.05000	– 0.10000

Table 8 For interactions related to the educational system, statistical parameters corresponding to expressed sentiments by country

Country	Average sentiment	Standard deviation	Median	Mode
United States	– 0.01950	0.14797	– 0.05000	– 0.10000
Myanmar	– 0.03097	0.10467	– 0.05000	– 0.10000
Singapore	– 0.02947	0.11305	– 0.05000	– 0.10000
United Kingdom	– 0.00966	0.14864	0	0
India	0.05277	0.28004	0	0
Canada	0.01554	0.19240	0	0
Pakistan	0.02566	0.25279	0	0
Japan	– 0.03431	0.10115	– 0.05000	– 0.10000
Australia	0.01804	0.19191	0	– 0.10000
Germany	– 0.00562	0.13037	– 0.05000	– 0.10000
France	– 0.02881	0.11951	– 0.05000	– 0.10000
Thailand	– 0.01235	0.14067	– 0.05000	– 0.10000
South Africa	0.00233	0.24754	0	0
Kenya	0.00314	0.21817	0	0

Table 9 For interactions related to the educational system, statistical parameters corresponding to expressed sentiments by sex and country

Country	Women				Men			
	Average sentiment	Standard deviation	Median	Mode	Average sentiment	Standard deviation	Median	Mode
United States	− 0.02240	0.13600	− 0.05000	− 0.10000	0.21542	0.23325	0.10000	0.10000
Myanmar	− 0.03049	0.10856	− 0.05000	− 0.10000	0.15298	0.17365	0.10000	0.10000
Singapore	− 0.03472	0.11203	− 0.05000	− 0.10000	0.16603	0.19251	0.10000	0.10000
United Kingdom	− 0.01683	0.14906	− 0.05000	− 0.10000	0.22900	0.23571	0.10000	0.10000
India	0.06201	0.25427	0	0	0.37450	0.28789	0.36667	0
Canada	− 0.00050	0.14407	0	− 0.10000	0.34471	0.28785	0.32875	0
Pakistan	0.05687	0.25899	0.00952	0	0.35759	0.26160	0.40000	0
Japan	− 0.04524	0.09409	− 0.05000	− 0.10000	0.16170	0.17913	0.10000	0.10000
Australia	0.01836	0.18390	0	0	0.30406	0.25790	0.25000	0.10000
Germany	0.00554	0.11997	0	− 0.10000	0.20250	0.23656	0.10000	0.10000
France	− 0.03358	0.10214	− 0.10000	− 0.10000	0.19082	0.23330	0.10000	0.10000
Thailand	− 0.01693	0.10716	− 0.05000	− 0.10000	0.20814	0.25918	0.10000	0.10000
South Africa	− 0.0017	0.2265	0	0	0.37519	0.27903	0.36875	0
Kenya	0.01548	0.14288	0	0	0.33188	0.26742	0.31786	0

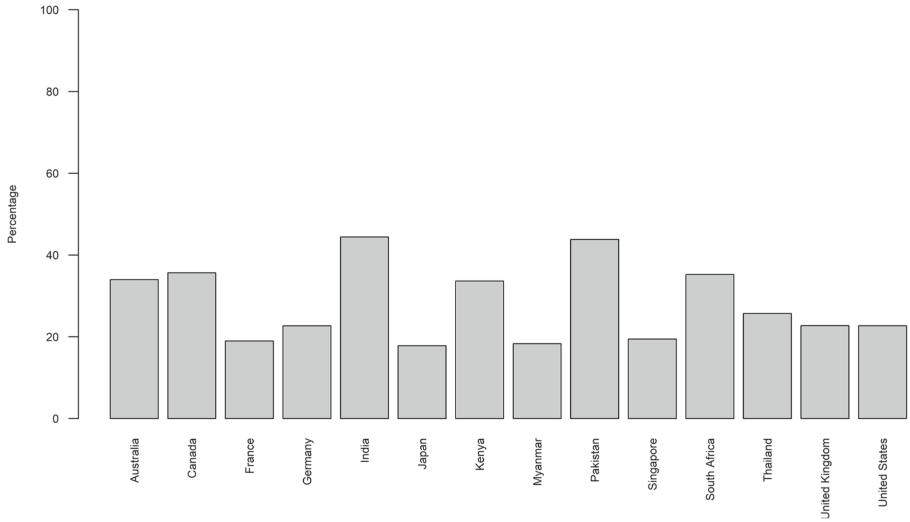

Fig. 1 For interactions related to educational system, percentage of tweets with a positive sentiment

countries. Because the standard deviation is significant for the obtained mean value, we consider the median value more appropriate for the statistical analysis.

In Table 10 can be seen that globally, medians and modes exhibit similar values for both sexes. Again, the standard deviation is relevant considering the average subjectivity. Due to this, the median is a more suitable value for the statistical study.

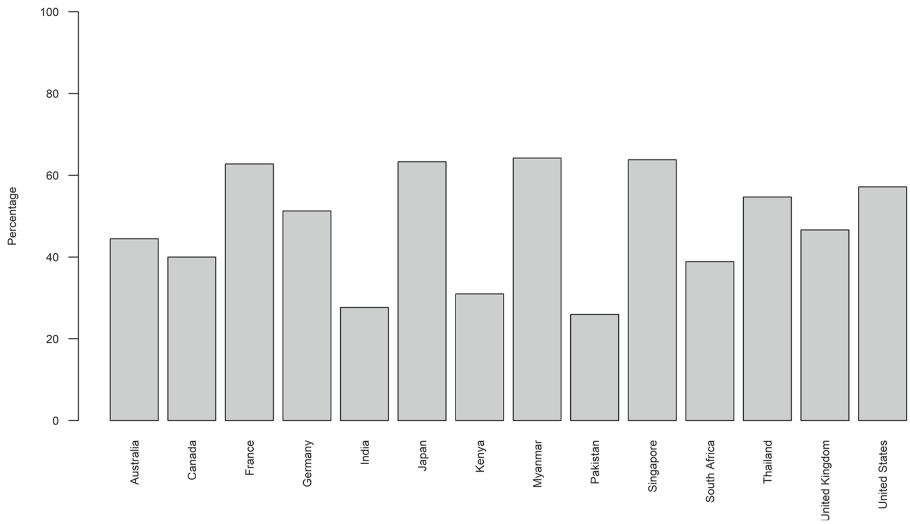

Fig. 2 For interactions related to educational system, percentage of tweets with a negative sentiment

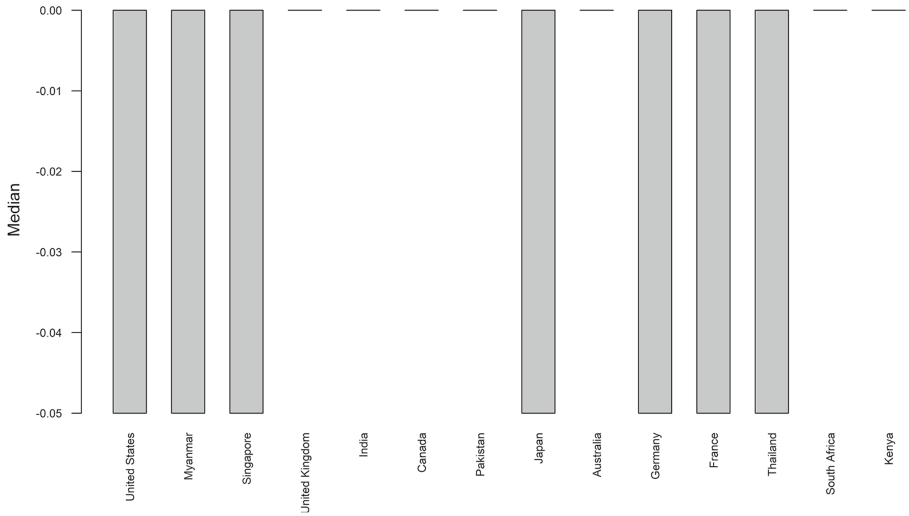

Fig. 3 For interactions related to educational system, median of expressed sentiments by country

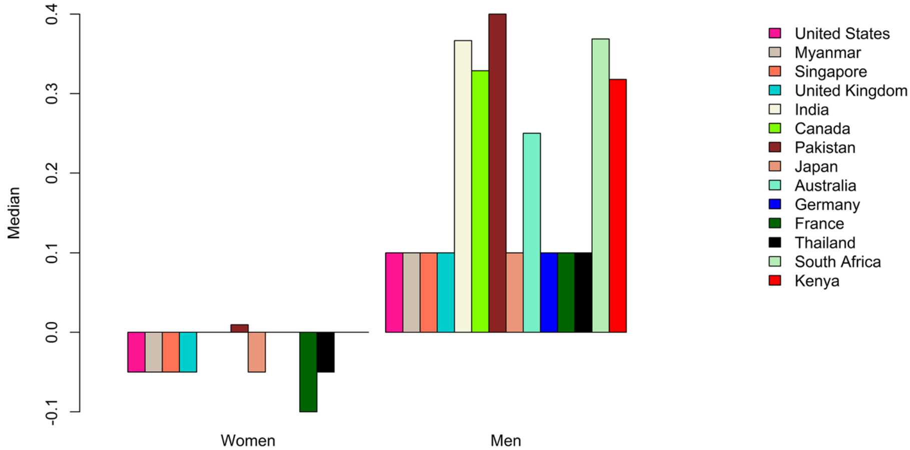

Fig. 4 For interactions related to educational system, median of expressed sentiments by country and sex

Table 10 For interactions related to the educational system, statistical parameters corresponding to expressed subjectivity by sex

Sex	Average subjectivity	Standard deviation	Median	Mode
Women	0.18225	0.20276	0.1	0.10000
Men	0.21785	0.22831	0.1	0.10000

Table 11 For interactions related to the educational system, statistical parameters corresponding to expressed subjectivity by country

Country	Average subjectivity	Standard deviation	Median	Mode
United States	0.19861	0.22381	0.10000	0.10000
Myanmar	0.15575	0.17640	0.10000	0.10000
Singapore	0.16529	0.19154	0.10000	0.10000
United Kingdom	0.18351	0.22528	0.10000	0.10000
India	0.38002	0.28711	0.38333	0
Canada	0.27650	0.25859	0.20000	0.10000
Pakistan	0.36339	0.27254	0.40000	0
Japan	0.14293	0.15663	0.10000	0.10000
Australia	0.24824	0.22716	0.16667	0.10000
Germany	0.19859	0.24392	0.10000	0.10000
France	0.15855	0.17980	0.10000	0.10000
Thailand	0.20280	0.24958	0.10000	0.10000
South Africa	0.37142	0.29880	0.34444	0
Kenya	0.30207	0.25374	0.30000	0

In Table 11 it can be observed that, considering the median values, four countries present the highest subjectivity. Again, the standard deviation is relevant considering the average subjectivity, the median is considered a more appropriate statistical metric for the study.

In Table 12 it can be noted that, considering the median values, some countries show differences in the subjectivity by sex but others do not. Once more, the standard deviation is significant considering the average subjectivity so we take the median as a more suitable statistical metric for the study (Fig. 6).

Figure 7 depicts the similarity heat-map for the studied countries utilising the $Sim2Count$ metric. It can be observed that the most similar are Myanmar and Japan and the most divergent in the perception of the education system are the United States and Pakistan, as well as India and Singapore.

Additionally, Fig. 8 displays the most frequently used words globally in tweets, and Table 13 shows analogous information by nation, the frequency of occurrence of words in the tweets is indicated in parentheses.

As an example of a country with an average positive and negative sentiment, Figs. 9 and 10 show the cloud of the most frequently used words in Japan and India. It can be noted that the graph shows words related to a lack of possessions as well as wishes in India. While in Japan the diagram displays words with the meaning related to protest, complaint or disagreement. All the above will be examined in more detail in the Discussion Section (Fig. 11).

Table 12 For interactions related to the educational system, statistical parameters corresponding to expressed subjectivity by sex and country

Country	Women				Men			
	Average sentiment	Standard deviation	Median	Mode	Average sentiment	Standard deviation	Median	Mode
United States	0.18220	0.21367	0.10000	0.1	0.21542	0.23325	0.1	0.10000
Myanmar	0.16174	0.18129	0.10000	0.10000	0.15298	0.17365	0.10000	0.10000
Singapore	0.15860	0.18119	0.10000	0.10000	0.16603	0.19251	0.10000	0.10000
United Kingdom	0.19822	0.21683	0.10000	0.1	0.22900	0.23571	0.10000	0.10000
India	0.37917	0.29018	0.39375	0	0.37450	0.28789	0.36667	0
Canada	0.22203	0.22754	0.10000	0.10000	0.34471	0.28785	0.32875	0
Pakistan	0.38526	0.28146	0.40000	0	0.35759	0.26160	0.40000	0
Japan	0.11774	0.12159	0.10000	0.10000	0.16170	0.17913	0.10000	0.10000
Australia	0.25929	0.21609	0.21000	0.10000	0.30406	0.25790	0.25000	0.10000
Germany	0.18744	0.24945	0.10000	0.10000	0.20250	0.23656	0.10000	0.10000
France	0.14843	0.16148	0.10000	0.10000	0.19082	0.23330	0.10000	0.10000
Thailand	0.17009	0.22041	0.10000	0.10000	0.20814	0.25918	0.10000	0.10000
South Africa	0.36894	0.34516	0.24107	0.10000	0.37519	0.27903	0.36875	0
Kenya	0.29415	0.19692	0.35000	0	0.33188	0.26742	0.31786	0

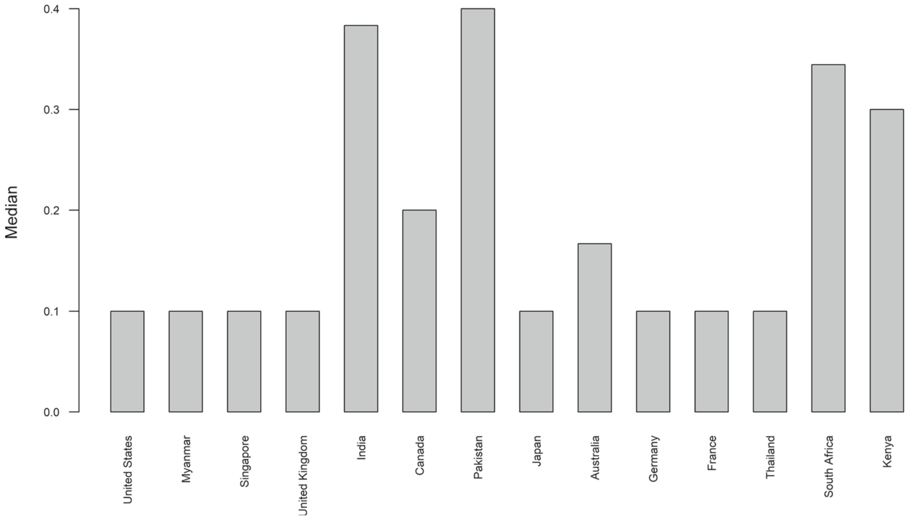

Fig. 5 For interactions related to educational system, median of expressed subjectivity by country

4 Discussion

The study of the perception of the educational system is a matter of relevant interest. The public perception of the quality of education programs has been explored in specific countries (Shrestha, 2013) as well as the public opinion on the Higher Education (University of Scranton, 2004).

The World Innovation Summit for Education (WISE 2020) published a comprehensive survey of young people's views on their education and future. This survey, conducted by Ipsos Group S.A (2021), was carried out in 20 countries around the world, aiming to understand how young people perceive their education and how

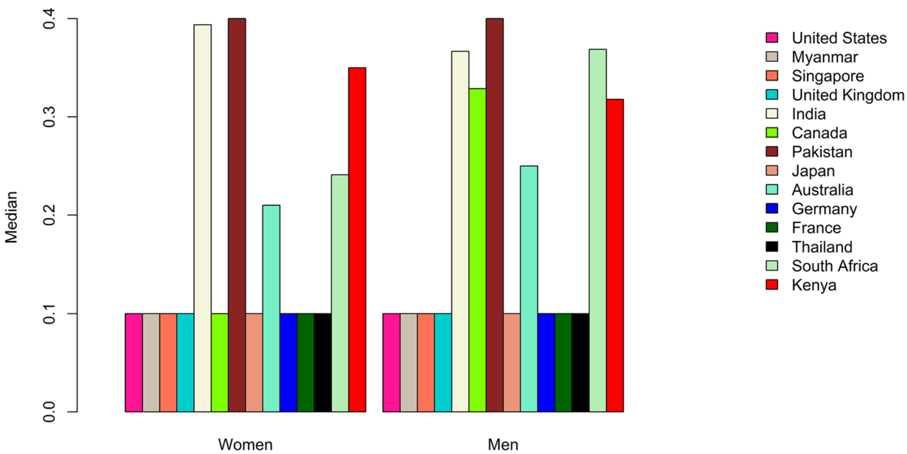

Fig. 6 For interactions related to educational system, median of expressed subjectivity by country and sex

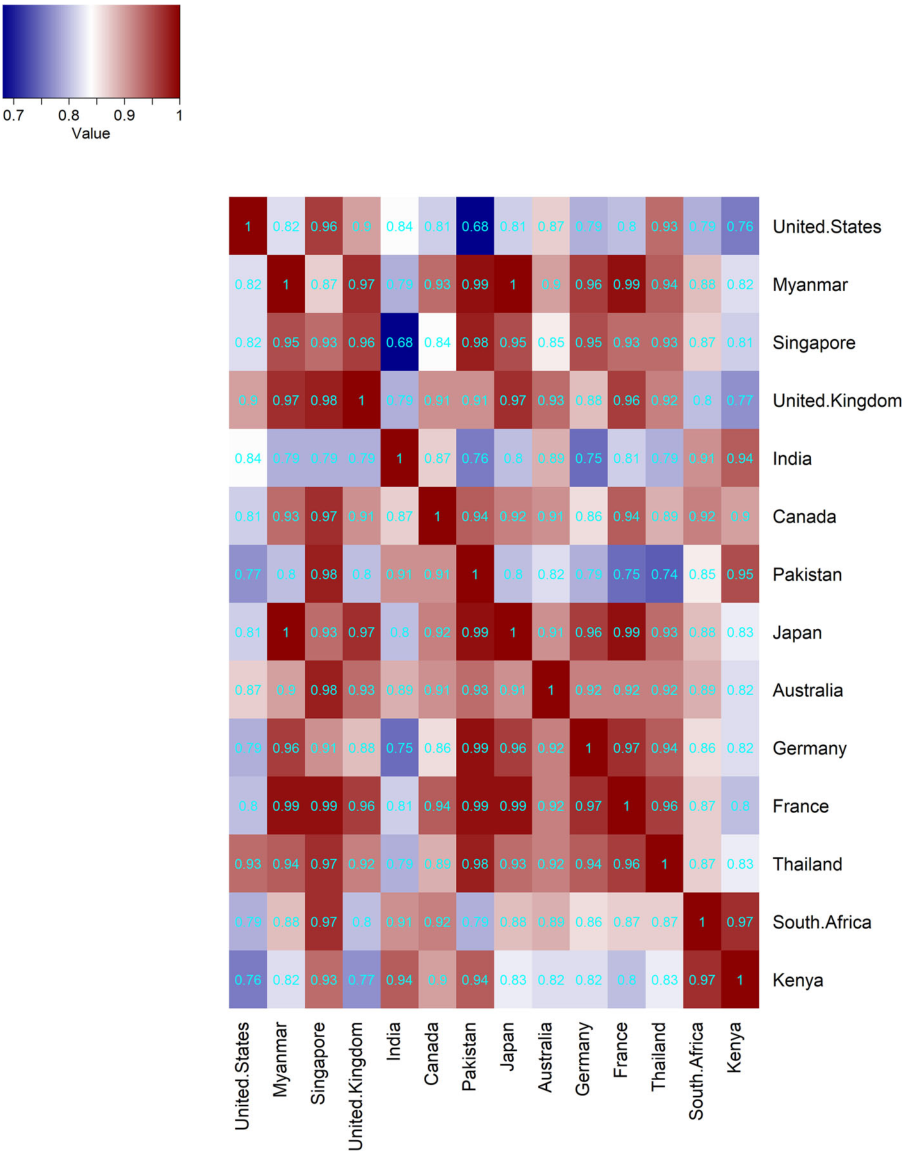

Fig. 7 For interactions related to the education system, a heat-map showing the calculated similarity between countries using the *Sim2Count* metric

prepared and confident they feel about their personal destiny. The study revealed that students really value their training. Around 90% agreed that education was not just to learn a career or profession, but it has a value in itself. Additionally, nearly 84% of young people thought that learning outside school was as important as learning in the classroom. The analysis also highlights that 80% of young people were "satisfied" or "somewhat satisfied" with their education, although only 27% was "totally satisfied".

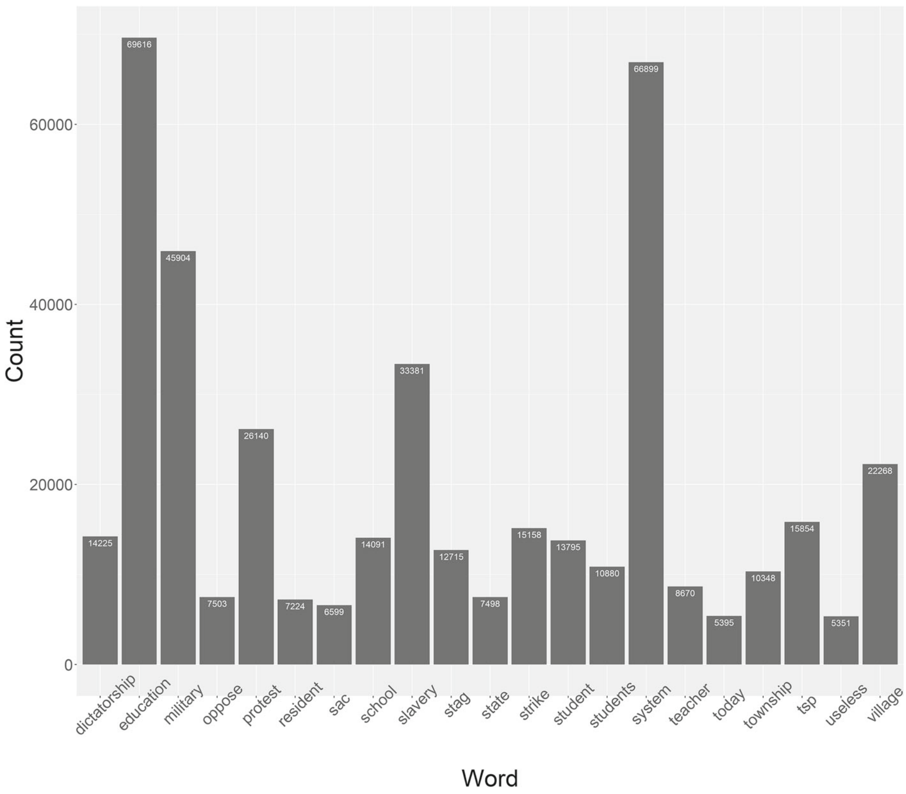

Fig. 8 For interactions related to the educational system, most frequently used words (articles and auxiliary verbs have been excluded)

The survey also showed two main areas of improvement that young people would like to implement in their education. Firstly, students demanded a more individualised educational approach (more than 60% said that they would like their teachers to provide them with personal advice on career guidance and on ways of learning as well as studying what is tailored to their needs). Secondly, students demanded more time to learn 21^{st} century skills. In this respect, about half of the young people (41%) thought that the current education lacked training in new technologies (artificial intelligence, coding, etc.) as well as appropriate instruction in communication and organisational skills. Around 44% believed that education did not foster creativity and curiosity.

Also, the Organisation for Economic Co-operation and Development's (OECD's) in its Programme for International Student Assessment (PISA) (OECD, 2019a) provided a comprehensive and rigorous international assessment of the learning outcomes of 15-year old students. The results obtained provide important insights into the quality and equity of education systems, allowing educators and policy makers to learn from practices in different countries. The third of six volumes of this survey focuses on the physical and emotional health of students, as well as the role of teachers and parents in shaping the environment and social life at school. In the

Table 13 For interactions related to the educational system, the most frequently used words by country[1]

Country	Words (times)
United States	education(12619); system(12044); military(8958); slavery(6420); protest(4888); village(4494); strike(3266); tsp(3013); dictatorship(2975); school(2576); student(2512); stag(2440); students(2094); township(2025); teacher(1736); state(1437); may(1428); sac (1414); resident(1389); oppose(1355); useless(1200); asean(1194); include(1087)
Myanmar	education(4777); system(4523); military (4083); slavery (3014); protest(2414); village (1932); tsp (1473); strike (1317); dictatorship (1184); school(1170); stag (1166); student (1109); students (973); township(923); may (804); teacher (724); oppose (684); resident (662); state (660); sac (560); asean (483); useless(483); district (483)
Singapore	education (3327); system (3157); military(2811); slavery (2013); protest(1549); village(1362); tsp(934); dictatorship(865);strike(851); stag (834); school(807); student(767); students(673); township(621); oppose(527); teacher(515); may (482); resident(476); state(450); sac(397); include(339);local(317);district(310)
United Kingdom	education(3191);system(3052); military (1856); slavery(1365); protest(1075); user (988); village(913); signature (793); tsp(668); strike(636); dictatorship(602); school (561); student(527); stag(499); support(496); june (473); government(438); students (435); reform (410); township(409); please (407); grow (407)
India	education(1402);system(1297); user(817); student(228);exam(190); the (172);indian(168); dont(139);india (139); change(112);make (110); need(104); students(104); school(103); get (84); know(83);govt(80); like(78);wrong (72); always(71); teach(70); want (68); our(65)
Canada	system(526); education (515); military(162); user(152); slavery(121); school(102); protest(91); the (89); student(73); village(71); strike(64); tsp(62); stag(60); teacher(52); students(50);need(47); public(47); dictatorship(45); take(38); may (37); state (35);oppose (34); township(33)
Pakistan	system(472); education (415); user(343); educational(82); the(65); pakistan(58); corruption(49); student (46); nation(45); eliminate(44); may(43); department(41); upgrade(40); request(40);exam (39); prosper(38); traditional(38); government (36); want(34); country(28); dont(26); need (25); one(25)
Japan	education(536); system(498); military(425); slavery (330); protest(250); village(208); tsp(147); strike(144); dictatorship(139); school(113); student(112); stag(106); township(102); students(98); oppose(81); teacher(76);may(75); resident(69); state(64); region(54); sac(53); demonstrate(53); boycott(51)
Australia	education(364);system (350); military(188);slavery(140); protest(98); village(83);school(78); student(63); user (63); dictatorship(62); tsp(58); stag(58); strike(51); teacher(50); township(41); students(38); may(37); state(36); oppose(35); need(35); resident(30); the(28); today(24)
Germany	education(442); system (420); military(310); slavery(224); protest(180); village(156); strike(138); tsp (114);dictatorship(111);stag(105); school (103); sac(83); student(80); students(79); teacher (67); boycott(64); township(64); domestic(54); asean(53); useless(53); oppose (52); state(49); demonstrate(43)
France	education(336); system(323); military(266); slavery(199); protest(149); village(133); dictatorship(96); tsp(91); strike(84); student(71); township(69); school(69); stag(64); students(58); may(52); teacher(50); oppose(46); resident(43); state(35); include(32); sac(31); district(31); local(30)
Thailand	education(266); system(253); military(217); slavery(158); protest (118); village(118); tsp(89); dictatorship(82); strike(76); school (60); students(56); student(55); stag(55); teacher(52); township(46); sac(40); include(34); useless(33);asean (33); may(32);boycott(30); state(30); resident(28)

Table 13 (continued)

Country	Words (times)
South Africa	education(175); system(160); user(35); military(26); need(23); south(21); the(20); slavery(18); school (18); strike(17); protest(17); people(17); get (15); useless(13); africa(12); village(12); student(12); dictatorship(12); tsp(11); year(11); change(11); country(11); our(11)
Kenya	education(129); system(126); user(65); the(45); student(19); school (17); need(14); kenya(14); kenyan(13); our(12); learn(11); fail(11); people(10); make(9); dont(9); cbc(8); government(8); job(8); future(7); one(7); public(7); want(7); get(7)

[1]asean: Association of SouthEast Asian Nations

stated volume, for the first time, 15-year-old students were asked how they frequently felt about their lives. Students reported positive affective states: “happy”, “joyful”, “proud”, “content” and “cheerful”, but also negative affective conditions: “scared”, “miserable”, “fearful” and “sad”.

In the aforementioned survey, more than 85% of students of the OECD countries reported that they sometimes or always felt “happy”, “cheerful” or “joyful”. In contrast, around 6% of students reported that they “always felt sad”. It must be noted that across countries and economies, girls were more likely than boys to report that they “sometimes” or “always” felt sad. The investigation also highlighted that the more time students spent on the Internet outside of school, the more likely they were to report feeling sad or miserable. Across virtually all countries and economies, students were more likely to report positive sentiments when they had a high sense of belonging to the school and greater cooperation between students existed. In contrast, if students had experienced bullying, they were more likely to manifest sadness. The

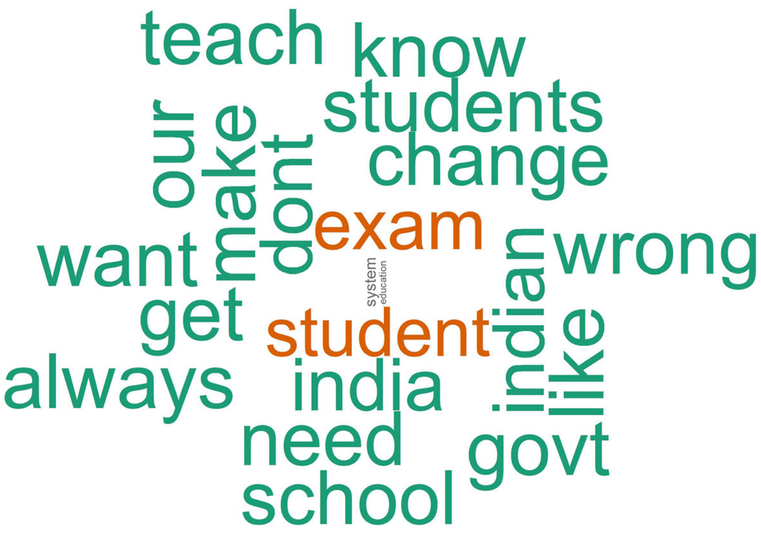

Fig. 9 Most frequently used words in India (articles and auxiliary verbs have been excluded)

Fig. 10 Cloud of the most frequently used words in Japan (articles and auxiliary verbs have been excluded)

survey also noted that the difference between those students that manifested "low use" and "frequent use" of Internet outside the classroom and that indicated feeling miserable "sometimes" or "always" was 7 percentage points among boys and 13 percentage points among girls, on average across OECD countries. It can be noted that, among the five school climate indices analysed in the survey, those that best predicted positive student feelings in OECD countries were the index of sense of belonging to school, followed by the indices of student co-operation and exposure to bullying.

The study also described life at school as a key aspect of pupils' existence. School was not only the place where children acquired knowledge, but also where they made friends, established trusting relationships with their teachers and developed an attachment to their place of learning. However, negative reactions to education could also be generated at school. There are also other factors and events outside the establishment, such as economic crises and natural disasters, as well as family-related problems, which can influence pupils' sentiments.

Research exists that explains that students who feel attached to their school, persevere and love learning report positive affective states, such as enthusiasm, inspiration and happiness. The causes of negative affective states, such as sadness, fear, despair or shame, are often more difficult to detect than positive affective states (Anderman, 1999; Weber et al., 2016). Certain student behaviours, such as teamwork, self-regulation, general positive attitudes towards school and life, as well as a sense of belonging to school, possessing hope and receiving love, seem to protect students from experiencing negative emotions (Anderman, 1999; Weber et al., 2016).

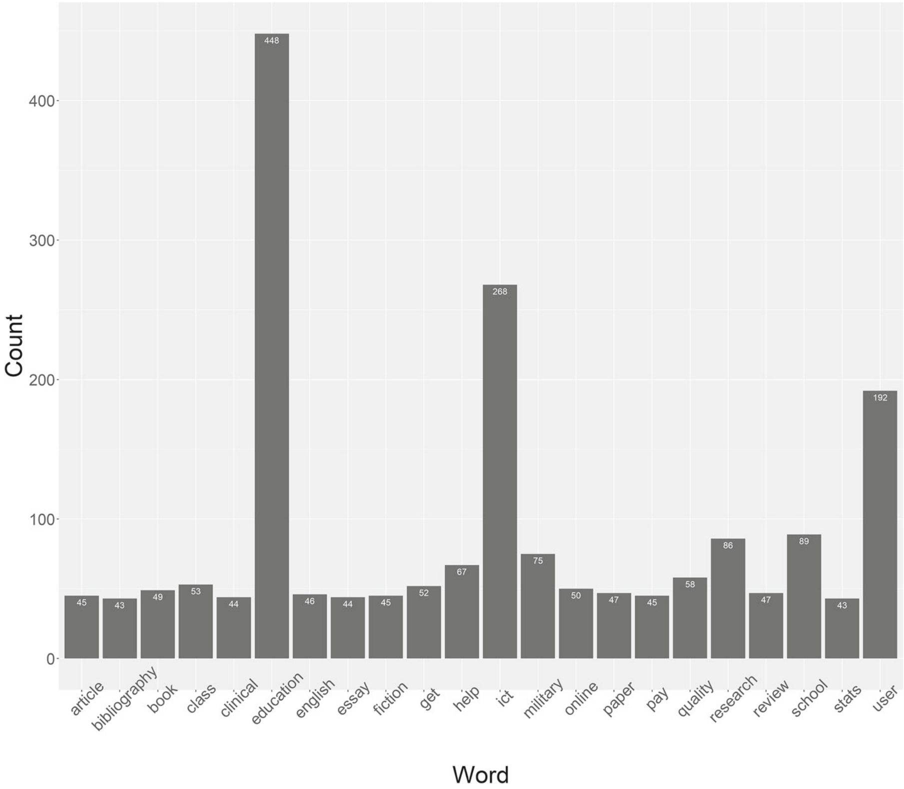

Fig. 11 For interactions related to the educational system, most frequently used words (articles and auxiliary verbs have been excluded)

Regarding our investigation, which covers a wider age range and type of person, Tables 7 and 8, show some statistical results such as average sentiment, standard deviation, median and mode[5]. It can be observed that negative perceptions of education systems are prevalent according to the obtained median. This happens for both sexes as well as for all analysed countries, except in India, Canada, Pakistan, Australia, South Africa and Kenya. These nations exhibit a higher percentage of positive tweets, as shown in Figs. 1 and 2, which can be put in relation to their income level.

Based on their incomes by 2022, the aforementioned nations can be classified, as follows (World Bank, n.d.):

- Lower-Middle Income Economies ($1,046 to $4,095): India, Pakistan, Kenya.
- Upper-Middle-Income Economies ($4,096 to $12,695): South Africa.
- High-Income Economies ($12,696 OR MORE): Australia and Canada.

[5]The median must be considered as the most appropriate metric, since the mean exhibits a relevant standard deviation

Globally, as Fig. 8 shows, the words “dictatorship” and “slavery” are the most used. The vision by country is displayed in Table 13. It can be observed that in all countries except in three with lower-middle (India, Pakistan and Kenya) or upper-middle (South Africa) incomes, the term “slavery” is one of the most frequently utilised words. In all nations except India, Pakistan and Kenya, the vocabulary “dictatorship” is one of the most posted words. In India and South Africa, the word “change” is also highly used, as well as “job” and “future” in Kenya. In the aforementioned six countries in which the sentiment towards the education system is positive, this is India, Canada, Pakistan, Australia, South Africa and Kenya, the term “need” has a relevant use. The above suggests that in those nations where education is perceived as an opportunity to improve, change and progress, the sentiment towards the education system is more favourable.

It must be noted that in all countries but three of them with lower-middle income (India, Pakistan and Kenya) or upper-middle income (South Africa) the term “slavery” is one of the most frequently used words in the tweets.

Japan, which is one of the countries showing negative sentiment, as Fig. 3 shows, traditionally has occupied a good position in most international educational performance surveys. Thus, it is among the best-placed countries in the Program for International Student Assessment (PISA), especially in Science and Mathematics (with 529 and 527 points, respectively, in the 2018 edition) (OECD, 2019b). Japan’s results also show great equity, as the impact of socioeconomic status on student performance is below the OECD average.

The Japanese education system is characterized by emphasising effort over ability or intelligence. The Japanese belief in the idea of achievement through hard work has led to the proliferation of so-called “juku”, after-school classes to reinforce learning. Students work from an early age in this meritocracy system to achieve good results, especially in the exams that give access to the best higher secondary schools and the most prestigious universities (National Center on Education and the Economy, n.d.). This competitiveness could be one of the factors implicated in the negative perception of the educational system. Figure 10 depicts a word cloud diagram, in which the terms “oppose”, “dictatorship”, “boycott” and “demonstrate” are among the most frequently used words. Figure 7 shows that very high similarities exist regarding the used vocabulary between Japan, Pakistan, France, and Myanmar. High similitude $S_C > 0.97$ has also been found between several countries. It should be noted that even if there are nations that are similar in the words used, they may not be analogous in terms of frequency of use, and may therefore show very different sentiment and subjectivity.

In India, one of the countries with the most positive sentiments towards education, as Fig. 3 shows, the access to education is far from universal. According to UNESCO, there are 252,863,750 illiterate Indians over the age of 15 (162,780,856 of whom are women) and the enrolment rate in Higher Education in 2020 was 29.4% (31.3 for women and 27.8 for men). It could be established that education is a precious asset and therefore the assessment of the education system is more positive. Figure 9 shows a word cloud diagram in which “need”, “teach”, “change”, “want” are the most recurrent terms.

Table 14 More relevant indicators related to this research

Country	Human development index	Rank	Gender Development Index (GDI)	Internet users, total (% of population)	Expected years of schooling (years)
India	0.645	131	0.820	34.5%	12.2
Canada	0.929	16	0.986	91%	16.2
Pakistan	0.557	154	0.745	15.5%	8.3
Australia	0.944	8	0.976	86.5%	22
South Africa	0.709	114	0.986	56.2%	13.8
Kenya	0.601	143	0.937	17,8%	11.3

Additionally, it must also be noted that, if the sentiments in each country are desegregated by sex, high differences are detected between men and women, as is shown in Table 9. This is especially noticeable in nations such as India, Canada, Pakistan, Australia, South Africa and Kenya, where the difference between the medians of the sentiments between the two sexes is higher than 0.24 [6].

Regarding the detected subjectivity in the tweets, as is shown in Tables 10, 11 and 12 the interactions do not exhibit very high subjectivity, nor do they show very relevant differences between sexes. However, India, Canada, Pakistan, Australia, South Africa and Kenya are the ones that show the highest subjectivity. The largest values correspond, in order, to Pakistan, India, South Africa and Kenya. These four countries have low overall human development indicators (UNDP, n.d.) (particularly India, Pakistan and Kenya), far below Australia or Canada, as it can be observed in Table 14.

The results of the social perception of national education systems indicate that there is large room for improvement. Greater involvement of families and the local community in the education could contribute to such enhancement. Research shows that family involvement in children's learning is strongly associated not only with the academic success, but also with their socio-emotional development. The OECD, in the aforementioned PISA reports, introduces parental involvement as a key indicator of educational effectiveness.

In addition, raising awareness of the social effects of education beyond employability could also favour a good social perception. As our study shows, in low-income countries education is perceived as a valuable asset, as a hope for a better quality of life. The OECD worked on the development of indicators on the potential social outcomes of education since the 2009 edition of *Education at a glance report* (OECD, 2021). The 2021 edition focuses on the impacts of education on health statuses and subjective well-being.

Health is an important policy area in OECD countries, as evidenced by the rapid increase in life expectancy over recent decades and in the context of the current COVID-19 pandemic. While education does not impact on health in isolation from

[6] We consider the median as the most appropriate metric, since the mean has a significant standard deviation.

other factors or in an unidirectional way, the latest OECD report (2021) shows that, on average, there is a gap by level of education. This gap is exemplified through of a difference of 5 years of life expectancy between the population with tertiary education and those with only primary or first stage of secondary education. Life expectancy is a consequence of the economic circumstances in a country, health conditions and other mortality risks that affect individuals throughout their life trajectory. Higher educational attainment provides means to improve socioeconomic status and is associated with better living habits and healthcare (OECD, 2019c).

In some countries, as previously mentioned, self-demand and competitiveness are strongly encouraged in academic training, both by families and teachers. We think that special attention should be paid to the psychological factors that can affect students in those countries.

Sentiment analysis in social networks can be a valuable tool for taking the pulse of certain social issues. In the case of educational systems, studies such as this one can help to guide educational policies, as well as to increase the transparency and efficiency of the education system.

5 Conclusions

Data science is an interdisciplinary field of scientific methods, processes and systems for extracting knowledge from data. Such science encompasses statistics, data analysis and its related methods, in order to understand and analyse current phenomena/problems (Deloitte, 2018). We apply data science to the study of the sentiments that citizens have about a problem as relevant as the education system.

The analysis of public opinion expressed in social networks makes it possible to visualise dominant opinions without the bias of social desirability. This type of analysis, as we have mentioned, has great potential both to capture the pulse of society on sensitive issues.

According to the results of our research, a predominantly negative perception of the education system has been found in most of the analysed countries. A positive perception is detected in certain low-income nations. Men exhibit a more positive sentiment than women as well as a higher subjectivity. The countries that exhibit the most positive perceptions are also those that manifest the highest subjectivity.

This research can be continued by extending the analysis to tweets written in some of the most widely spoken languages in the world today. The five most extensively spoken languages in the world are (not only mother tongues are considered): English (1,121 million speakers), Mandarin Chinese (1,107 million speakers), Hindi (698 million speakers), Spanish (512 million speakers), and French (284 million speakers) (LINGUA, 2022). We could use `TextBlob` to perform the sentiment analysis but we first translated the text of each tweet into English. Using `TextBlob` translation and language identification can be done through the Google Translate API. Other software tools could also be utilised to directly perform sentiment analysis in languages other than English.

An examination through focus group interviews can also complement the obtained results with qualitative information. Therefore, a phenomenological approach could be adopted, one focused on the collection of individual experiences of the participants in relation to a collective phenomenon (perception of the national education system), establishing different age cohorts.

Additionally, the semantic similarity between tweets can be can be determined generating graph-based spatial relationships between their texts. In this approach, the texts are represented as a graph using synonymous relationships. Community detection as well as other network analysis metrics can be utilised.

The methodology used in this research can also be used to examine citizens' perceptions of other current issues.

Acknowledgements This research was performed as a result of the Project: Hopper: Women, Society, Technology and Education which was granted in an internal call in 2021 at the Universidad Francisco de Vitoria. This work was also partially supported by Telefonica Chair at Francisco de Vitoria University. The authors acknowledge Mari Luz Congosto Martínez for her assistant on the use of the T-Hoarder tool.

Data Availability The data utilised to support the findings of this research are available from the corresponding author upon request.

Declarations

The research complied with all the relevant national regulations. We only used public tweets.

Conflict of Interests The authors declare that they have no conflict of interests.

References

Agarwal, A., Xie, B., Vovsha, I., Rambow, O., & Passonneau, R. (2011). Sentiment analysis of Twitter data. In *Proceedings of the workshop on languages in social media*.

Ahmed, H., Razzaq, M., & Qamar, A. (2013). Prediction of popular tweets using similarity learning. In *ICET 2013–2013 IEEE 9th international conference on emerging technologies*. https://doi.org/10.1109/ICET.2013.6743524.

Anderman, L. H. (1999). Classroom goal orientation, school belonging and social goals as predictors of students' positive and negative affect following the transition to middle school. *Journal of Research & Development in Education*, *32*(2), 89–103.

Alrasheed, H. (2021). Word synonym relationships for text analysis: a graph-based approach. *Public Library of Science (PLoS) ONE*, *16*(7), e0255127. https://doi.org/10.1371/journal.pone.0255127.

Amidon, M. (2011). *Social networking: MySpace, Facebook & Twitter*. USA: Technology Pionners.

Arcila-Calderón, C., Ortega-Mohedano, F., Jiménez-Amores, J., & Trullenque, S. (2017). Análisis supervisado de sentimientos políticos en español: clasificación en tiempo real de tweets basada en aprendizaje automático. *El Profesional de la Información*, *26*(5), 973–982. https://doi.org/10.3145/epi.2017.sep.18.

Arcila-Calderón, C., Blanco-Herrero, D., & Valdez, M.B. (2020). Rechazo y discurso de odio en Twitter: análisis de contenido de los tuits sobre migrantes y refugiados en español. *Revista Española de Investigaciones Sociológicas*, *172*, 21–40. https://doi.org/10.5477/cis/reis.172.21.

Arguedas, M., Beita, J., Rodríguez, F., Umaña, J., & Vaca, M. (2020). Crisis migratoria en Colombia y Costa Rica: una visión desde el análisis de sentimientos. *Revista Humanidades*, *10*(2). https://doi.org/10.15517/h.v10i2.42238. Universidad de Costa Rica.

Ayala, T. (2014). *Redes sociales, poder y participación ciudadana Revista Austral de Ciencias Sociales,* vol. 26, pp. 23–48. Chile: Universidad Austral de Chile Valdivia.

Beykikhoshk, A., Arandjelović, O., Phung, D., Venkatesh, S., & Caelli, T. (2015). Using Twitter to learn about the autism community. *Social Network Analysis and Mining*, *5*(22), 1–17. https://doi.org/10.1007/s13278-015-0261-5.

Bharathi, S. S., & Geetha, A. (2017). Sentiment analysis for effective stock market prediction. *International Journal of Intelligent Engineering and Systems*, *10*, 146–154. https://doi.org/10.22266/ijies2017.0630.16.

Blasco, M., & Coenders, G. (2020). Análisis de sentimiento de la agenda de los partidos políticos españoles en Twitter durante la Moción de Censura de 2018. Un enfoque de datos composicionales. *Revista Mediterránea de Comunicación*, *11*(2), 185–198.

Bonta, V., Kumaresh, N., & Janardhan, N. (2019). A Comprehensive study on lexicon based approaches for sentiment analysis. *Asian journal of Computer Science and Technology*, *8*, 1–6. https://doi.org/10.51983/ajcst-2019.8.S2.2037.

Balakrishnan, A., Idicula, S. M., & Jones, J. (2021). Deep learning based analysis of sentiment dynamics in online cancer community forums: an experience. *Health Informatics Journal*, 1–18. https://doi.org/10.1177/14604582211007537.

Bauhr, M., & Charron, N. (2020). Do men and women perceive corruption differently? Gender differences in perception of need and greed corruption. *Politics and Governance*, *8*, 92–102. https://doi.org/10.17645/pag.v8i2.2701.

Benjamin, T. B. (1993). Public perceptions of higher education. *Oxford Review of Education*, *19*(1), 47–63. https://doi.org/10.1080/0305498930190104.

Calderón, N. A., Fisher, B., Hemsley, J., Ceskavich, B., Jansen, G., Marciano, R., & Lemieux, V.L. (2015). Mixed-initiative social media analytics at the world bank: observations of citizen sentiment on Twitter data to explore "trust" of political actors and state institutions and its relationship to social protest. In *2015 IEEE international conference on big data* (pp. 1678–1687). https://doi.org/10.1109/BigData.2015.7363939.

Cheeti, S. S. (2021). Twitter based sentiment analysis of impact of Covid-19 on education globaly (May 2021). *International Journal of Artificial Intelligence and Applications (IJAIA)*, *12*(3). Retrieved 07 Sept 2022, from https://ssrn.com/abstract=3973030.

Congosto, M. L., Basanta-Val, P., & Sanchez-Fernandez, L. (2017). T-hoarder: a framework to process Twitter data streams. *Journal of Network and Computer Applications*, *83*, 28–39. https://doi.org/10.1016/j.jnca.2017.01.029.

Dey, A., Jenamani, M., & Thakkar, J. J. (2018). Senti-N-Gram: An n -gram lexicon for sentiment analysis. *Expert Systems With Applications*, *103*, 92–105. https://doi.org/10.1016/j.eswa.2018.03.004.

Deloitte (2018). Data science transforming the ICT domain. Retrieved 22 May 2022, from https://www2.deloitte.com/content/dam/Deloitte/in/Documents/technology-media-telecommunications/in-tmt-data-science-transforming-noexp.pdf.

Dreiling, M., & Belkhir, J. A. (1997). Race, gender & class. *Environmentalism and Race, Gender, Class Issues*, *5*(1), 153–169.

Facebook Statistics and Trends (2022). Retrieved 07 Sept 2022, from https://datareportal.com/essential-facebook-stats#:~:text=Facebook.

Farzindar, A., & Inkpen, D. (2015). *Natural language processing for social media*. USA: Morgan & Claypool Publishers.

Fontenelle, S. M., & Zinkhan, G. M. (1993). Gender differences in the perception of leisure: a conceptual model. In L. McAlister, & M. L. Rothschild (Eds.) *NA–Advances in Consumer Research,* (vol. 20 pp. 534–540). Provo: Association for Consumer Research.

Galasso, V., Pons, V., Profeta, P., Becher, M., Brouard, S., & Foucault, M. (2020). Gender differences in COVID-19 attitudes and behavior: panel evidence from eight countries. In *Proceedings of the national academy of sciences of the United States of America,* (vol. 117(44)). https://doi.org/10.1073/pnas.2012520117.

Gómez-Marí, I., Tijeras-Iborra, A., Manzanera-Hernández, S. M., & Tárraga-Mínguez, R. (2022). La discapacidad es trending topic:, análisis del impacto en Twitter de diez campañas de sensibilización. *Prisma Social: Revista de Investigación Social*, *36*, 142–165.

Hopper, A., & Uriyo, M. (2015). Using sentiment analysis to review patient satisfaction data located on the internet. *Journal of Health Organization and Management*, *29*, 221–33. https://doi.org/10.1108/JHOM-12-2011-0129.

Ipsos Group S.A. (n.d) (2021). https://www.ipsos.com/en.

Janda Janda, H., Pawar, A., Du, S., & Mago, V. (2019). Syntactic, semantic and sentiment analysis: the joint effect on automated essay evaluation. *IEEE Access*, 1–1. https://doi.org/10.1109/ACCESS.2019.2933354.

Jackson, C. A., & Luchner, A. F. (2018). Self-presentation mediates the relationship between self-criticism and emotional response to Instagram feedback. *Personality and Individual Differences*, *133*, 1–6.

Kalyanam, J., Quezada, M., Poblete, B., & Lanckriet, G. (2016). Prediction and characterization of high-activity events in social media triggered by real-world news. *Public Library of Science (PloS) one*, *11*(12), e0166694. https://doi.org/10.1371/journal.pone.0166694.

Karami, A., Dahl, A., Turner-McGrievy, G., Kharrazi, H., & Shaw, G. (2018). Characterizing diabetes, diet, exercise, and obesity comments on Twitter. *International Journal of Information Management*, *38*(1), 1–6. https://doi.org/10.1016/j.ijinfomgt.2017.08.002.

Kharde, V., & Sonawane, S. (2016). Sentiment analysis of Twitter data: a survey of techniques. *International Journal of Computer Applications*, *139*, 5–15. https://doi.org/10.5120/ijca2016908625.

Kartal, S., Yirci, R., & Özdemir, T. (2015). Teachers', parents' and students' perceptions regarding the problems in Turkey's national education system: a qualitative study. *7*.

Kaur, H., Ahsaan, S. U., Alankar, B., & Chang, V. (2021). A proposed sentiment analysis deep learning algorithm for analysing COVID-19 tweets. *Information Systems Frontiers*, *23*, 1417–1429. https://doi.org/10.1007/s10796-021-10135-7.

Kolasani, S., & Assaf, R. (2020). Predicting stock movement using sentiment analysis of Twitter feed with neural networks. *Journal of Data Analysis and Information Processing*, *8*, 309–319. https://doi.org/10.4236/jdaip.2020.84018.

Kušen, E., Strembeck, M., Cascavilla, G., & Conti, M. (2017). On the influence of emotional valence shifts on the spread of information in social networks. In *Proceedings of the 2017 IEEE/ACM international conference on advances in social net-works analysis and mining 2017,* (pp. 321–324).

Lahitani, A. R., Permanasari, A. E., & Setiawan, N.A (2016). Cosine similarity to determine similarity measure: study case in online essay assessment. In *2016 4th international conference on cyber and it service management, 2016,* pp. 1–6. https://doi.org/10.1109/CITSM.2016.7577578.

LINGUA (2022). The 20 most spoken languages in the world in 2022. https://lingua.edu/the-20-most-spoken-languages-in-the-world-in-2022/.

Macy, M., Mejova, Y., & Weber, I. (2015). *Twitter: a digital socioscope*. New York: Cambridge University Press.

Mouronte-López, M. L., & Subirán, M. (2022). What do Twitter users think about climate change? Characterization of Twitter interactions considering geographical, gender and account typologies perspectives. *Weather, Climate, and Society*, (published online ahead of print 2022). Retrieved 12 Sept 2022, from https://journals.ametsoc.org/view/journals/wcas/aop/WCAS-D-21-0163.1/WCAS-D-21-0163.1.xml.

Musto, C., Semeraro, G., & Polignano, M. (2014). A comparison of lexicon-based approaches for sentiment analysis of microblog. *CEUR Workshop Proceedings*, *1314*, 59–68.

Mehmood, Y., & Balakrishnan, V. (2020). An enhanced lexicon-based approach for sentiment analysis: a case study on illegal immigration. Online Information Review. ahead-of-print. http://doi.org/10.1108/OIR-10-2018-0295.

Mohammad, N., Syed, Z., Sana, F., Mala, I., & Bhagwan, D. (2018). A survey on sentiment analysis, classification and applications. *International Journal of Pure and Applied Mathematics*, *119*.

Naseem, M., Zia, S., Fatima, S., Mala, I., & Bhagwan, D. (2018). A survey on sentiment analysis, classification and applications. *International Journal of Pure and Applied Mathematics*, *119*(10), 1203–1211.

National Center on Education and the Economy (n.d.) National center on education and the economy (NCEE). https://ncee.org/country/japan/.

OECD. (2019). *PISA 2018 results (volume III): what school life means for students' lives PISA*. Paris: OECD Publishing. https://doi.org/10.1787/acd78851-en.

OECD. (2019). *Snapshot of trends in reading, mathematics and science performance, in PISA 2018 results (Volume I): what students know and can do*. Paris: OECD Publishing.

OECD. (2019c). *Health at a glance 2019 OECD indicators*. Paris: OECD Publishing. https://doi.org/10.1787/4dd50c09-en.

OECD. (2021). *Education at a glance 2019: OECD indicators*. Paris: OECD Publishing. https://doi.org/10.1787/19991487.

Twitter by the numbers: stats, d.emographics.&.f.un.f.acts. (2022). Retrieved 07 Sept 2022, from https://www.omnicoreagency.com/twitter-statistics/.

Percastre-Mendizábal, S., Pont-Sorribes, C., & Suau-Gomila, G. (2019). La gestión comunicativa en redes sociales digitales de la emergencia del Ébola en españa. The communicative management in social media of the emergence of Ebola in Spain, 80–90. https://doi.org/10.20318/recs.2019.4437.

Quadri, S., Karusala, N., & Arriaga, R. (2018). #AutismAwareness: a longitudinal study to characterize tuiting patterns for Indian and US users. *ICPS Proceedings*, 11–19. https://doi.org/10.1145/3297121.3297128.

R. (n.d.) The R project for statistical computing. https://www.r-project.org/.
Rajkumar, J., Vishal, S., & Sachin, D. (2019). Sentiment analysis on product reviews using machine learning techniques. *Proceeding of CISC 2017*. https://doi.org/10.1007/978-981-13-0617-4_61.
Remus, R., Quasthoff, U., & Heyer, G. (2010). SentiWS – a puplicly available German-language resource for sentiment analysis. In *Proceedings of the 7th international language ressources and evaluation (LREC'10),* 1168–1171.
Safari, E., & Barigye, S. (w. d.). Students perceptions regarding the quality high learning system in Rwanda case study of Mount Kenya University Rwanda. Retrieved 07 Sept 2022, from http://erepository.mkuit.ac.rw/bitstream/handle/123456789/5756/paper.
Shrestha, B. K. (2013). Public perception of the quality of academic education program. *Journal of Education and Research*, *3*, 52–64. https://doi.org/10.3126/jer.v3i0.7852.
The Small Business Blog (n.d.) How many people use Instagram in 2022? (Instagram Statistics) (2022). Retrieved September 07, 2022, from https://thesmallbusinessblog.net/instagram-statistics/.
Sydney, M., Dulosa, S., Inocian, R., Bokeron, M. R. A., & Moneva, J. (2021). Perception of students towards education. *International Journal of Novel Research in Education and Learning*, *6*(1), 35–45.
Sued, G. E., & Cebral, M. (2020). Voces autorizadas en Twitter durante la pandemia de COVID-19: actores, léxico y sentimientos como marco interpretativo para usuarios ordinarios. *Revista de Comunicación y Salud*, *10*(2), 549–568. https://doi.org/10.35669/rcys.2020.10(2).549-568.
Suerdem, A., & Kaya, E. (2015). Using sentiment analysis to detect customer Attitudes in social media comments. *Research in Computing Science*, *90*, 207–215. https://doi.org/10.13053/rcs-90-1-16.
Twitter (n.d.) Tweepy. An easy-to-use python library for accessing the Twitter API. Retrieved 07 Sept 740 2022, from https://www.tweepy.org/.
UNDP (n.d.) United Nations Development Programme. Human development reports. Global human development indicators. https://hdr.undp.org/en/countries.
UNESCO (n.d.) Gender equality and education. Retrieved 07 Sept 2022, from https://www.unesco.org/en/gender-equality/education.
Uhr, P., Zenkert, J., & Fathi, M. (2014). Sentiment analysis in financial markets - a framework to utilize the human ability of word association for analysing stock market news reports. In *IEEE international conference on systems, man and cybernetics*. https://doi.org/10.1109/SMC.2014.6974028.
Van Rossum, G., & Drake, F. L. (2009). *Python 3 reference manual scotts valley*. CA: CreateSpace.
University of Scranton (2004). University of Scranton University of Scranton (2004). Environmental scanning report: public perceptions of higher education. Retrieved 07 Sept 2022, from https://www.scranton.edu/pir/institutional-research/kates-from-ir/Public-Perceptions-of-Higher-Education.pdf.
Vohra, S., & Teraiya, J. (2013). Applications and challenges for sentiment analysis: a survey. *International Journal of Engineering Research & Technology*, *2*(2).
Walaa, M., Ahmed, H., & Hoda, K. (2014). Walaa sentiment analysis algorithms and applications: a survey. *Ain Shams Engineering Journal*, *5*(4). https://doi.org/10.1016/j.asej.2014.04.011.
Weber, M., Wagner, L., & Ruch, W. (2016). Positive feelings at school: on the relationships between students character strengths, school related affect, and school functioning. *Journal of Happiness Studies*, *17*(1), 341–355. https://doi.org/10.1007/s10902-014-9597-1.
Weber, D., Nasim, M., Mitchell, L., & Falzon, L. (2021). Exploring the effect of streamed social media data variations on social network analysis. *Social Network Analysis and Mining*, *11*(1), 1–45. https://doi.org/10.1007/s13278-021-00770-y.
Weller, K., Bruns, A., Burgess, J., Mahrt, M., & Puschmann, C. (2014). *Twitter and society*. New York: Peter Lang.
World Bank (n.d.) World Bank country and lending groups (2022). Retrieved May 22, 2022, from https://datahelpdesk.worldbank.org/knowledgebase/articles/906519.
WISE (2020). World innovation summit for education. https://www.wise-qatar.org/.