THE ROLE OF SOCIAL MEDIA IN SUSTAINABLE CONSUMPTION: A CLASS-WISE ANALYSIS

CEREN CUBUKCU CERASI, YAVUZ SELIM BALCIOGLU, ASLI KILIC, FARID HUSEYNOV
Gebze Technical University, Gebze, Kocaeli, Turkey
cerencubukcu@gtu.edu.tr, ysbalcioglu@gtu.edu.tr, aslikilic@gtu.edu.tr, fhuseynov@gtu.edu.tr

Abstract Public opinion must change if green consumerism is to be promoted and making this change will be easier with an understanding of society's movement toward green consumption. Businesses can benefit from this adjustment and make their processes and products more sustainable and eliminate any kind of waste. This study uses sentiment analysis to determine social media's contribution to the promotion of sustainability. It intends to quantify public attitudes toward green consumption and identify significant points of attitude shift by applying text mining and deep learning algorithms on structural and semi-structured data from YouTube. Long Short-Term Memory, Support Vector Machines and Random Forest algorithms were used in this study. In addition, class-wise performance was measured. The keywords related to eco-friendly consumption were selected and then a class-wise performance analysis was performed related to each word structure. Encouraging sustainability and sustainable consumption by businesses can be very beneficial for them. In this way, companies will be able to strengthen both their own and their country's resilience with a sustainable economy.

Keywords:
sustainability, green consumption, sentiment analysis, class-wise analysis, social media

JEL:
Q56, C38

University of Maribor Press

DOI https://doi.org/10.18690/um.epf.3.2023.9
ISBN 978-961-286-736-2

1 Introduction

In recent years, the public has become increasingly aware of the negative environmental effects of modern civilization. As a result, there has been a growing interest in sustainable practices, with an emphasis on environmentally friendly consumption. Green or sustainable consumption means using goods and services that are produced in an eco-friendly and sustainable way (Kong et al., 2002). This includes using products made from recycled or renewable materials, which are made ethically and have a low carbon footprint. Sustainable consumption also involves reducing waste, preserving resources, and choosing options that have minimal impact on the environment. This shift towards green consumerism requires a change in public opinion. A transition to a sustainable civilization can be aided by promoting attitudes and values that support sustainability (Akenji, 2014). This would involve education and awareness-raising campaigns, as well as training programs for leaders in various sectors. Moving towards sustainable lifestyles is a gradual process that requires changes in the way we live our lives, including the way we shop and behave (Gilg et al., 2015).

In order to contribute to the relevant literature, this research conducted sentiment analysis on social media data, specifically YouTube, to provide an example of how to explore the current status and popularity of sustainability related topics with YouTube commentators. The study's focus on text mining and deep learning algorithms to understand public attitudes and identify significant points of attitude shift presents a valuable resource for understanding consumer behavior and the public's view on sustainability. This study tries to show how deep learning methods and social media data can play a valuable role in promoting sustainable consumption among the general population. The results of this study provide significant insights for both scholars and businesses working towards a more sustainable economy.

2 Literature Review

Most of the research in literature related to sustainability have used questionnaires on various factors and a group of participants. For example, Ahamad and Ariffin (2018), investigated the knowledge, attitudes and practices of students towards sustainable consumption with the data obtained through a questionnaire. As a result of the research, it was revealed that although the students' knowledge about sustainable consumption was at a high level, they were at a moderate level in attitudes and practices.

Jalali and Khalid (2019) conducted a study on the green product purchasing behaviors, values and precautions of influencers with the survey data they conducted with Instagram users. The Theory of Reasoned Action (TRA) and the Uses and Gratifications Theory (UGT) were used in their studies.

Sajeewanie et al. (2019) proposed a model for predicting consumer purchase intentions and behavior towards green products. Their model took perceived consumer behavior, habit, subjective norms, environmental attitude, environmental knowledge and awareness, price, brand image, and marketing information into account.

Wang (2021) collected data with a questionnaire survey, to investigate the effect of consumers' green cognition on green consumption behavior. The outcome of the research shows that, green consumption knowledge of consumers and their views on environmental problems can primarily enable the development of green consumption awareness of consumers. The methodology will be described in the section after this. Results and an analysis of that data will then be presented, followed by conclusions.

3 Methodology

The 6489 comments that were obtained from Youtube are notably relevant to the three major green keywords, which we will refer to as T0, T1, and T2. You may find the relevant comments by doing a search on Youtube for usernames, their acronyms, and the names of the most common keywords that each user used, such as #Balance of trade, #Sustainability, and #Green items. For the sake of this research and analysis, the analysis has been carried out in a collective fashion. As a result, the hypothesis that the distribution of emotions that are represented in comments that are relevant toward a single remark is indicative of the emotions associated with that term is not accepted. In order to accommodate this, the class of attitudes that is taken into consideration includes those that also reflect T0 and T1, T0 and T2, T1 and T2, none, and unrelated. It is important to note that the mining of comments from Youtube took place from January 2020 all the way through December 2020, as this year was the year of the Covid-19 pandemic, the most data was obtained.

The purpose of using classification methods such as Long Short Term Memory (LSTM), Support Vector Machines (SVM), and Random Forest was to analyze and quantify public attitudes towards sustainable consumption as expressed through YouTube comments. By adhering to these selection criteria, we aimed to create a diverse and comprehensive dataset of YouTube videos and their corresponding comments, allowing for an in-depth analysis of public attitudes towards sustainable consumption.

In the end, the following fields were chosen; firstly, ID: An individual ID will be assigned to each remark. In order to maintain the users' anonymity while yet identifying them, it is recommended that a numerical ID be used as an identifier rather than the username. Secondly, text: Displays the user's first remark that was uploaded to the site. Thirdly, class: There are three columns, with two of them being designated for each of the three keywords. The binary data that may be entered into these columns is determined by whether the comment is oriented towards any one, two, or all of them, or even none of them at all. After combining the three columns, the resulting binary number is then transformed to decimal so that it may be used as an easy-to-understand indicator. It might be anywhere between 0 and 6, inclusive.

3.1 Preprocessing

As compared to other online social media platforms for text and video, YouTube has its own unique traditions, which is one of the reasons why extra caution is used during the cleaning and pre-processing of unprocessed comments. The following procedures are included in the pre-processing stage: first, Non-standard lexical tokens, such as, emotions, and anomaly punctuation, are filtered out just after the tokenization process; second, duplicate comments are excluded to maintain the uniqueness of each comment; third, standard stop words are removed; and finally, case folding is performed in order to convert all of the tokens into lower case letters. There are a considerable number of terms that are out of vocabulary, however these words are not removed from the data.

3.2 Annotation

There are three annotators who are possibly aware of the green keywords and intelligently aware of the keywords that are utilized either in favor of or against a user for annotation. Instead of assigning a definitive positive or negative feeling to each comment, annotators were tasked with determining if the statement in question was an opinion or hypothesis. Annotation categories included the following six subcategories: T0, T1, T2, (T0 and T1), (T1 and T2), and none. T0 = green products, T1 = sustainability, T2 = balance of trade, shown in Table 1. A suitable class was given to the collection of superfluous annotations so that the distribution of comments could be identified, rather than the group being thrown out as unnecessary.

Table 1: Classes and Corresponding keywords

Classes	Corresponding keyword
None	None
T0	Green products
T1	Sustainability
T2	Balance of trade
T0 and T1	"Green products" and "Sustainability"
T0 and T2	"Green products" and "Balance of trade"
T1 and T2	"Sustainability" and "Balance of trade"

3.3 Feature Extraction

There are three distinct kinds of feature vectors that are based on term-frequency, and they are as follows: Term Frequency – contrary document density for unigrams, bigrams, and trigrams. The Term density – contrary document density is a major statistical method that is used to transform a collection of unprocessed documents into a matrix in the form of term density vs. contrary document density. In other means, it gives a term more weight if it appears less often in the material being analyzed.

3.4 Long Short-Term Memory

Recurrent neural network (RNN) units that can learn long-term dependencies are referred to as long short-term memory (LSTM) units. A recurrent neural network (RNN) made of LSTM units is essentially what an LSTM network is. A cell, an input gate, an output gate, and a forget gate are the component parts that make up a typical LSTM unit (Gao et al., 2020). The cell can remember values for arbitrarily long periods of time, and its three gates are responsible for controlling the flow of information into and out of the cell. The analysis of sequential data is where the LSTM model (Saini, 2021), which belongs to the category of deep learning, shines the brightest. It is used in the processing of jobs like language translation, music production, and voice recognition, amongst others.

3.5 Support Vector Machines (SVM)

A method for machine learning that is drawn from analytical learning theory and is based on the notion of structural uncertainty reduction is referred to as a discriminative classifier. This approach is also known as the Support Vector Machine (SVM) (Pisner & Schnyer, 2020). A separation hyper-plane that has optimality between the two classes of a training dataset is found using the support vector machine (SVM). The hyperplane with the greatest distance from the training dataset that is located closest to it is used to calculate this optimal separation. Conventionally, supervised learning approaches like support vector machines (SVM) are used for classification, regression, and the identification of outliers (Abdullah & Abdulazeez, 2021).

3.6 Random Forest

A random forest (RF) is a model that is made up of a collection of tree-structured models with the formula h(x, k), where k = 1,..., where the k are independent random vectors that are distributed equally and where each tree casts a unit vote for the class that has the most users given the input x. The bagging algorithm serves as the foundation for the RF, which also makes use of the Ensemble learning approach. Using the given subset of data (Breiman, 2001), it generates many random trees and then aggregates the results of all of those trees. One benefit of using a RF is that, in comparison to using basic Decision Tree models, it can produce more accurate predictions because of its reduction of overfitting and variance (Schonlau & Zou, 2020).

4 Results and Discussion

The data that are shown in Table 2 are the result of the study of annotations. The bulk of the 6489 comments that were gathered exhibit a feeling toward the term T0. This represents 30.1% of the total comments. In contrast to this, the proportion of comments that reflect attitudes toward T1 is just 19.05%, while the percentage of comments that show sentiments for T2 is even a much lower percentage, at only 19.09%. This demonstrates that there are a significant number of people on Youtube that prefer the green keyword T1. It's interesting to note that out of all the comments, there are 1097 that indicate a preference for a combination of the terms T 0 and T1. This accounts for 16.7% of the total comments. Yet, the favor seen for T0 and T2 is 8.9%, and the favor observed for T1 and T2 is 9.4%, both of which are very low.

In addition to LSTM, several other machine learning techniques, such as Support Vector Machines and Random Forest, are used in order to do a comparative analysis of the comment data which can be seen in Table 3. Cross-validation using stratified 6-folds is performed on the whole data set for the purpose of validation. The LSTM and the SVM had the greatest precision out of all the classifiers, coming in at 0.76 and 0.73 respectively. The LSTM model has the greatest recall at 0.77, which is equivalent to the model's term-frequency representations. The LSTM and SVM models exhibit the greatest F1-Scores, which are respectively 0.74 and 0.7. By looking at everything together, the performance of RF is the worst of all of them. The LSTM classifier has the capability of achieving maximum accuracy.

Table 2: Comments statistics

Sentiment Classes	No. of comments in each class	% of comments in each class
None	126	1.95
T0	1947	30.1
T1	1249	19.05
T2	894	13.9
T0 and T1	1097	16.7
T0 and T2	573	8.9
T1 and T2	603	9.4
Total	6489	100

Table 3: Evaluation of classifiers regarding a variety of characteristics

Classification method	Precision	Recall	F1-Score
LSTM	0.76	0.74	0.7498
SVM	0.73	0.71	0.7198
RF	0.68	0.72	0.6994

In Table 4, each of the seven classes' precision, recall, and F1-Score are shown according to their respective class. The relevant classes with the greatest precision, 0.81, is the class T0, whereas the relevant classes with the lowest precision are T0 and T1 together. Surprisingly, the class T0 results in the greatest recall, and the class T0 results also provide the highest F1-Score. The fact that the "None" class obtained substantial values of achievement measures may be attributed to the fact that it is clearly differentiated from all of the other classes since all of the other classes have feelings, however the "None" class does not contain any emotion at all.

Table 4: Class-wise performance of LSTM

Classes	Precision	Recall	F1- Score
None	0.81	0.72	0.758
T0	0.79	0.81	0.793
T1	0.75	0.74	0.744
T2	0.76	0.73	0.738
T0 and T1	0.65	0.69	0.664
T0 and T2	0.71	0.74	0.724
T1 and T2	0.73	0.75	0.736

5 Conclusion

In order to investigate the many different learning models, some key characteristics are retrieved from the comments based on the inverse document frequency of phrase frequency. The examination of the gathered data reveals the existence of a green behavior on Youtube, which accounts for more than thirty percent of all the green keywords-driven comments that were retrieved. The assessment of the annotated dataset itself was performed using LSTM in addition to numerous machine learning models, and encouraging results were found using LSTM and SVM.

This study only focuses on sentiment analysis of social media data from YouTube, which may not be representative of the entire population's attitudes towards green consumption. Also, this study only uses three different algorithms, and there may be other algorithms that could provide better results for sentiment analysis. The practical value of this article is that it provides insights into public attitudes towards green consumption and sustainability, which can be helpful for businesses and policymakers in developing strategies to promote sustainable practices.

In conclusion, the findings of this study have the potential to provide important findings to raise awareness of sustainable consumption in businesses and encourage end consumers. Encouraging sustainability and sustainable consumption by businesses can be very beneficial for them, as academic studies in this field have shown that companies trying to protect the environment and prevent pollution can benefit from improved operational efficiency, lower costs, better brand image and better profits (Chen, 2015; Stål and Jansson, 2017). In this way, companies will be able to strengthen both their own and their countries' resilience with a sustainable economy.

Acknowledgment

This study is one of the preliminary outputs of the project numbered 2022-A-113-07 funded by Research Fund of Gebze Technical University.

References

Abdullah, D. M., & Abdulazeez, A. M. (2021). Machine learning applications based on SVM classification a review. *Qubahan Academic Journal*, 1(2), 81-90. doi:10.48161/qaj.v1n2a50

Ahamad, N. R., & Ariffin, M. (2018). Assessment of knowledge, attitude and practice towards sustainable consumption among university students in Selangor, Malaysia. *Sustainable Production and Consumption, 16,* 88-98. doi:10.1016/j.spc.2018.06.006

Akenji, L. (2014). Consumer scapegoatism and limits to green consumerism. *Journal of Cleaner Production*, *63*, 13-23. doi:10.1016/j.jclepro.2013.05.022

Breiman, L. (2001). Random forests. *Machine learning*, *45*, 5-32. doi:10.1023/A:1010933404324

Chen, L. (2015). *Sustainability and company performance: Evidence from the manufacturing industry* (Vol. 1698). Linköping University Electronic Press.

Gao, Z., Yuan, T., Zhou, X., Ma, C., Ma, K., & Hui, P. (2020). A deep learning method for improving the classification accuracy of SSMVEP-based BCI. *IEEE Transactions on Circuits and Systems II: Express Briefs, 67*(12), 3447-3451. doi:10.1109/TCSII.2020.2983389

Gilg, A., Barr, S., & Ford, N. (2005). Green consumption or sustainable lifestyles? Identifying the sustainable consumer. *Futures*, *37*(6), 481-504. doi:10.1016/j.futures.2004.10.016

Jalali, S. S., & Khalid, H. (2019). Understanding Instagram Influencers Values in Green Consumption Behaviour: A Review Paper. *Open International Journal of Informatics*, *7*(Special Issue 1), 47-58.

Kong, N., Salzmann, O., Steger, U., & Ionescu-Somers, A. (2002). Moving Business/Industry towards sustainable consumption: The role of NGOs. *European Management Journal*, *20*(2), 109-127. doi:10.1016/S0263-2373(02)00022-1

Pisner, D. A., & Schnyer, D. M. (2020). Support vector machine. In *Machine learning* (pp. 101-121). Academic Press. doi:10.1016/B978-0-12-815739-8.00006-7

Saini, K. (2021). A study on stock price prediction using LSTM model. *International Journal for Research in Applied Science and Engineering Technology*, *9*(VI), 2618–2622. doi:10.22214/ijraset.2021.35568

Sajeewanie, L. A. C., Ab Yajid, M. S., Khatibi, A., Azam, S. F., & Tham, J. (2019). Integrated model for green purchasing intention and green adoption: Future research direction. *Journal of Sociological Research*, *10*(2), 23-66. doi:10.5296/jsr.v10i2.14996

Schonlau, M., & Zou, R. Y. (2020). The random forest algorithm for statistical learning. *The Stata Journal*, *20*(1), 3-29. doi: 10.1177/1536867X20909688

Stål, H. I., & Jansson, J. (2017). Sustainable consumption and value propositions: Exploring product–service system practices among Swedish fashion firms. *Sustainable Development*, *25*(6), 546-558. doi:10.1002/sd.1677

Wang, Y. (2021, July). Research on the Influence Mechanism of Green Cognition Level on Consumers' Green Consumption Behavior: An Empirical Study Based on SPSS. In *2021 International Conference on Management Science and Software Engineering (ICMSSE)* (pp. 175-178). IEEE. doi:10.1109/ICMSSE53595.2021.00044