sustainability

MDPI

Article

Analysing Sustainability and Green Energy with Artificial Intelligence: A Turkish English Social Media Perspective

Fahrettin Kayan [1], Yasemin Bilişli [2,*], Mehmet Kayakuş [3,*], Fatma Yiğit Açıkgöz [1], Agah Başdeğirmen [4] and Meltem Güler [5]

[1] Department of Marketing and Advertising, Social Sciences Vocational School, Akdeniz University, Antalya 07058, Türkiye; kayan@akdeniz.edu.tr (F.K.); fatmayigit@akdeniz.edu.tr (F.Y.A.)
[2] Department of Office Services and Secretariat, Social Sciences Vocational School, Akdeniz University, Antalya 07058, Türkiye
[3] Department of Management Information Systems, Faculty of Social and Human Sciences, Akdeniz University, Antalya 07800, Türkiye
[4] Department of Management and Organisation, Isparta Vocational School, Isparta University of Applied Sciences, Isparta 32200, Türkiye; agahbasdegirmen@isparta.edu.tr
[5] Department of Communication (External), Akdeniz University, Antalya 07070, Türkiye; meltemguler1970@gmail.com
* Correspondence: yasemin@akdeniz.edu.tr (Y.B.); mehmetkayakus@akdeniz.edu.tr (M.K.); Tel.: +90-5327280787 (Y.B.); +90-5062654830 (M.K.)

Abstract: This study explores how linguistic and cultural differences shape social media discourses on green energy and sustainability by analysing English and Turkish tweets. Leveraging artificial intelligence-based text mining methods, the research examines users' perceptions, emotions, and concerns about green energy on social media platforms. The findings reveal that in both languages, negative sentiments outweigh positive ones, with users frequently expressing their criticisms and apprehensions. However, significant thematic differences emerge based on language and culture. English tweets generally adopt a global and industrial perspective, while Turkish tweets are more focused on local, technical, and operational issues. By integrating sustainability into the analysis, this study highlights the interconnectedness of green energy discussions with broader environmental and societal goals. Social media platforms are shown to play a critical role in raising environmental awareness and influencing consumer perceptions. The results underline the importance of developing sustainability policies that consider regional dynamics, cultural contexts, and user expectations. Additionally, this study provides valuable insights for advancing climate research, media strategies, and digital marketing efforts. Ultimately, it emphasises the need for inclusive, informed, and innovative approaches to foster greener and more sustainable futures globally.

Keywords: green energy; sustainability; climate; digital marketing; marketing communication; machine learning; social media

Academic Editors: Paulo Brito, Sergio Correia and Mohamed Lahby

Received: 16 January 2025
Revised: 6 February 2025
Accepted: 21 February 2025
Published: 22 February 2025

Citation: Kayan, F.; Bilişli, Y.; Kayakuş, M.; Açıkgöz, F.Y.; Başdeğirmen, A.; Güler, M. Analysing Sustainability and Green Energy with Artificial Intelligence: A Turkish English Social Media Perspective. *Sustainability* **2025**, *17*, 1882. https://doi.org/10.3390/su17051882

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Green energy has become one of the most important issues in the modern world with the goals of reducing the effects of global warming, ensuring sustainable development, and protecting environmental balances. The use of renewable energy sources is seen as an important strategy that not only provides environmental benefits but also supports economic growth. Today, it has become a global priority both to ensure environmental

sustainability and to provide long-term solutions to the energy crisis [1,2]. Social media platforms provide an effective environment for users to share their thoughts on this issue, express their support, or voice their concerns. This study aims to reveal how discussions on this concept are shaped in different cultural contexts by analysing consumer perceptions of green energy through Turkish and English tweets.

The aim of this study is to reveal consumers' perspectives on green energy in a comparative manner by analysing the emotional tendencies, perceptions, and emphasised themes in Turkish and English tweets. By using artificial intelligence methods such as sentiment analysis, text mining, and topic modelling, positive and negative emotions of social media users about green energy were determined. In addition, understanding the differences in the reactions to this concept in different languages and cultures contributes to the effective design of communication strategies related to green energy. This study also highlights how social media data can be used to understand consumer behaviour.

Social media analyses on green energy provide critical clues for digital marketing campaigns. The positive and negative emotions identified in this study allow brands to understand consumer perception and develop sustainability-orientated strategies. In this context, marketing communication strategies can be enriched with the data obtained from social media to develop messages in line with the environmental sensitivities of consumers. Storytelling techniques that emphasise green energy projects can be used as an effective tool to create brand loyalty and strengthen consumers' positive emotions. Messages of support for renewable energy projects can be used as a powerful communication element in the marketing of environmentally friendly products and services. This use of social media data offers a new perspective in the design of consumer-orientated digital campaigns. The findings of the article offer opportunities in various sectors.

In this study, modern methods for social media analysis are used. Analysing English and Turkish tweets with sentiment analysis, word frequency, and topic modelling (LDA) methods provides a comprehensive evaluation of the text data from both quantitative and qualitative perspectives. Visualisation tools such as PyLDAvis help to better understand and represent the topics obtained. Furthermore, the comparison of data in two different languages shows how linguistic and cultural differences can be managed in multilingual analyses. The combination of these methods allowed for deeper conclusions that can be drawn from social media data.

This study makes a unique contribution by analysing the similarities and differences between global and regional perceptions of green energy by comparing Turkish and English tweets. While previous studies have generally been conducted in a single language or in a limited geography, this analysis addresses consumer perceptions in two different cultural contexts within the same methodological framework. The importance of this study stems from the fact that it provides insights for organisations and decision-makers aiming to raise awareness on green energy to develop effective strategies in different cultural contexts. In addition, the fact that it provides an example of how artificial intelligence-supported analysis methods can be used in the field of social sciences makes this study unique in terms of both methodology and content.

Based on the comparative sentiment and topic modelling analysis of Turkish and English tweets, this study hypothesises that linguistic and cultural differences influence consumer perceptions of green energy. Specifically, it is expected that English tweets will focus more on global and industrial themes, while Turkish tweets will emphasise local and technical concerns. The findings of this study aim to test this hypothesis by analysing the sentiment distribution and thematic structures in both datasets.

2. Literature

Green energy is critical for sustainable development and the prevention of environmental degradation. The use of renewable energy sources is recognised as a key strategy for reducing the carbon footprint and addressing the negative impacts of fossil fuels on the environment [3]. Green energy has become a priority area for policymakers and industry leaders, especially in combating climate change and increasing energy security [4]. As a result of this priority, studies to understand the awareness levels of societies towards green energy have gained momentum. These studies play an important role in understanding individuals' responses to sustainable energy policies and the factors affecting their energy choices. For example, Sovacool (2009) stated that individuals' perceptions of green energy projects are closely related to environmental benefits and economic costs [5]. Similarly, Wolsink (2007) found that attitudes towards renewable energy projects often vary depending on the local context in which the projects will be implemented and the perceived benefits [6]. Ellis et al. (2007) argue that perceptions of renewable energy projects are influenced not only by 'objective' policy barriers but also by a range of factors such as governance, participation processes, aesthetic concerns, and power inequalities. Recent studies show that community participation and local economic benefits play a critical role in the formation of these perceptions [7]. In addition, Gross's (2007) research shows that individuals' beliefs about whether they are fairly represented in decision-making processes have an impact on their perceptions of projects [8].

In addition to environmental concerns, perceptions of economic and technological success play an important role in the acceptance of renewable energy technologies. Wüstenhagen and Menichetti (2012) reveal that innovative technologies receive more public support and that this support is often associated with economic benefits rather than environmental awareness. In addition, the influence of media discourses on perceptions of projects has become increasingly evident [9]. Foust and Murphy (2009) showed that the media's dramatic and confrontational coverage of projects can negatively affect perceptions [10].

While increased awareness of climate change positively affects individuals' general attitudes towards renewable energy projects, the reflections of these attitudes at the local level exhibit a complex structure. Whitmarsh et al. (2020) found that environmental sensitivity has increased among younger generations, but individual interests may create opposition to local projects. Economic factors also play an important role in the formation of these perceptions [11]. Johnson et. all (2015) emphasise that fluctuations in energy prices and investment costs can affect public confidence in renewable energy, while political stability and regulations are critical for long-term support [12].

Douglas' (1970) cultural theory suggests that individuals' perceptions of renewable energy projects are influenced by their relationships with wider social groups [13]. Hulme (2009) emphasises how environmental issues are associated with individual experiences and how these perceptions should be taken into account in policy design [14]. Furthermore, Pasqualetti (2011) suggests that the perception of wind turbines as an aesthetic attraction by some communities may increase social acceptance [15].

Current research indicates that perceptions towards renewable energy should be analysed in a more dynamic way. While social media analyses reveal the rapidly changing nature of social perceptions, sentiment analyses provide a more in-depth examination of these perceptions through large datasets [16]. For example, Castillo-Manzano et all (2017) found that both information sharing and social participation were high in social media users' discussions on green energy [17].

In this context, there is a need to utilise more postpositivist approaches to understand how individuals' perspectives on renewable energy are shaped through a combination of their presuppositions and sources of information. Devine-Wright (2006) emphasises that

perceptions of renewable energy projects need to be considered not only at the individual level but also in broader social and cultural contexts. These developments offer important insights that can contribute to the more effective design of renewable energy policies and projects [18].

This study is unique in that it analyses social media discourses on green energy through both English and Turkish tweets and examines how language and cultural differences are reflected in these discourses. While the existing literature usually focuses on analyses conducted in the context of a single language or culture, this study examined discourses in two languages from a comparative perspective. The findings show that global and industrial themes are prominent in English tweets, while local and technical issues are prominent in Turkish tweets. This indicates that green energy debates are shaped by regional dynamics and user perceptions, suggesting that sustainability policies should be developed in accordance with local contexts. Furthermore, this study emphasises the power of social media platforms to raise environmental awareness by conducting both sentiment analysis and thematic analysis with artificial intelligence-based text mining methods.

3. Materials and Methods

In this study, a total of 14,916 English and 1358 Turkish tweets collected from the X platform were analysed to analyse user views on green energy. The dataset was analysed comprehensively to reveal the thoughts, perceptions, and emotional aspects of social media users about green energy. The data were analysed using advanced data processing methods such as text mining, sentiment analysis, and topic modelling.

Text mining is a technique for extracting meaningful and useful information from a dataset and was applied in this study to analyse users' comments on green energy in more depth. This method was used for systematic processing of text data and the extraction of specific features. In addition, sentiment analysis was used to identify users' feelings and attitudes towards green energy. At this stage, a machine learning-based approach was adopted for sentiment analysis classification, and the support vector machines (SVM) algorithm was used. SVM is a powerful machine learning method used to find the hyperplane that best classifies the data points, and in this study, it was preferred to ensure the accurate classification of emotional loads in tweets.

Topic modelling, on the other hand, makes it possible to group users' opinions about green energy on a topic basis and to make an analysis based on certain themes and keywords. This method enables the data to be analysed in a broader perspective and provides useful information to understand the basic views and concerns of users about green energy.

3.1. Dataset

There are several important reasons why the X platform was preferred in this study. Firstly, the X platform has a large user base around the world and offers an environment where the public quickly expresses their opinions and interacts, especially through social media. The content shared on this platform clearly reveals users' thoughts and emotional reactions to current events, social issues, and various topics. Especially on a topic of social importance such as green energy, comments on the X platform provide access to the views of individuals from different geographies and demographics [19].

While this study primarily focuses on data from the X platform, other social media platforms such as Facebook, Instagram, and LinkedIn also provide valuable insights into public perceptions of green energy. However, these platforms differ in user demographics, content-sharing mechanisms, and accessibility of data. Unlike X, which allows easy access to real-time discussions and a wide range of public opinions, platforms

such as Instagram and LinkedIn focus more on visual content and professional discourse, respectively. Future research could explore these platforms to gain a more comprehensive view of consumer sentiment on green energy.

The X platform stands out as a space where users interact in real time and share their opinions freely. This provides researchers with a large and diverse dataset, allowing for a more accurate and comprehensive analysis of public opinions on green energy. Moreover, the diversity and breadth of the content on the platform make it possible to extract meaningful data through methods such as text mining and sentiment analysis used in the research.

In this study, a total of 14,916 English and 1358 Turkish tweets collected over a period of 2 weeks were selected as the dataset. There are several important reasons for choosing this specific time. Firstly, the rapid change and updating of content on social media platforms can lead to the problem of timeliness and authenticity, which are lost over time in longer-term datasets. Therefore, data collected over a short period of time allow for a more accurate reflection of the current views and perceptions of the public on the topic at the centre of the research, namely, green energy. Especially considering the rapid spread of discussions on environmental issues and energy policies on social media, a short period of time provides more meaningful data to understand the current feelings and thoughts of the public. Moreover, this 2-week period increases the applicability of methods such as text mining and sentiment analysis used in this study. Short-term datasets allow the model to run faster and more efficiently, thus speeding up the analysis process and allowing researchers to examine social sentiments and trends in a given period of time. A dynamic and rapidly changing topic such as green energy is directly linked to intense interactions monitored through social media; therefore, by using up-to-date data, public perceptions were measured more accurately and effectively.

The fact that the dataset consists of tweets in English and Turkish allows for the analysis of global and local perceptions by comparing public opinions in two different languages. This diversity enables a deeper exploration of public perspectives and concerns about green energy in different cultural contexts. While the data in English reflect universal views and debates on green energy on a global scale, the Turkish tweets reflect the feelings and thoughts of the Turkish public in a local context.

Finally, the size of the dataset, limited to 14,916 English and 1358 Turkish tweets, was sufficient to provide the depth of analysis and sample size required for this study, given the intensity of social media interaction in the selected time. These data allow for a comprehensive analysis of social media discussions on green energy.

Although the Turkish dataset (679 tweets after preprocessing) is smaller than the English dataset (7950 tweets), it still provides meaningful insights into consumer perceptions. Previous studies have shown that even smaller datasets can yield valuable results in sentiment analysis, particularly when combined with robust preprocessing and feature selection. Additionally, the findings obtained from Turkish tweets were thematically consistent with English tweets, further supporting the reliability of the analysis. (See Table 1).

Table 1. Sample dataset.

Tweet	Language
Incentives for the production of green technologies such as batteries for electric vehicles and renewable energy systems stand out as a new area of transatlantic competition as well as trade protectionism between the US and the European Union (EU).	Tr
While the production cost of green hydrogen produced using renewable energy sources is calculated to be USD 1 per kilogram in 2050, global investments in the hydrogen sector reached USD 1.1 billion last year.	Tr

Until 2050, the sector that will never fall off the agenda of the world and the country will be green energy. You will lose a lot if you are not in a sector with such infinite potential and development for long-term investment.	Tr
Let's protect our future by providing all the energy it needs from renewable energy sources.	Tr
Greenhouse gas emissions can be reduced by transition to a green economy. Investments can be made in renewable energy sources. Work can be done for the protection and restoration of natural habitats.	Tr
A clean environment and a healthy future is possible with renewable energy! In order to fulfil our duty for a greener future, we should prioritise increasing our product and service diversity for renewable energy sources.	En
Wind and solar are the primary sources of electricity in the EU in 2022. Electricity generated from these 2 sources in the EU reached 22.3% of the total, surpassing other sources for the first time, including coal, nuclear and natural gas.	En
It is on a mission to build homes in Wales that run on green energy and can reduce the amount of CO_2 in our atmosphere.	En
Denmark produced a record amount of green energy in 2022	En
The Chinese group plans to build a giant green hydrogen project in Egypt with investments ranging from USD 5–8 billion, which will produce more than 140,000 tonnes of green hydrogen per year.	En

3.2. Text Mining

Text mining is a process for extracting meaningful information from unstructured text data. This process utilises computer-based techniques to uncover patterns and relationships hidden in large amounts of text data. Drawing from areas such as natural language processing (NLP), machine learning, and information retrieval, text mining uses various algorithms to make text data analysable and usable [20]. In recent years, text mining has gained importance with the increase in data from digital sources such as social media, customer reviews, newsletters, and blogs [21].

The first step of the text mining process is text preprocessing. This step ensures that the text is cleaned and made suitable for analysis. Since raw text often contains redundant information and noise, the preprocessing process aims to remove such elements [22]. Preprocessing steps include tokenisation, stop-word extraction (removing words that do not carry meaning), and normalisation. Tokenisation is the process of dividing text into words or sentences. Stop-word extraction removes important words of the language as well as words that do not carry meaning. Normalisation is the stemming of words [23].

Another concept frequently encountered in text mining is n-grams. N-gram refers to consecutive sequences between words in the text. These sequences are used to understand the relationships between words. While unigram refers to the analysis of a single word, longer sequences such as bigrams and trigrams can be used to analyse more complex relationships between words. N-grams are an important tool for understanding the linguistic structure of text and making more meaningful inferences [24].

Text mining also works with methods such as classification and clustering. Classification is the process of categorising texts into certain categories. For example, sentiment analysis can be used to determine whether a text carries a positive, negative, or neutral emotion. Machine learning algorithms can be used for this classification process, learning from large datasets and classifying texts accurately.

In conclusion, text mining is a powerful technique for extracting valuable information from text data. Methods such as text preprocessing, n-gram analysis, and classification are the cornerstones of this process. This process enables more in-depth analyses of text data and allows meaningful inferences to be made from large datasets.

3.3. Sentiment Analysis

Sentiment analysis is a natural language processing (NLP) method for identifying the emotional tone in text data. This analysis aims to recognise positive, negative, or neutral expressions of emotion in texts. Sentiment analysis is widely used, especially in analysing large amounts of unstructured data such as social media, customer feedback, product reviews, and news comments [25]. This process is important in various fields, such as companies measuring customer satisfaction, conducting opinion polls, and evaluating the impact of policies on the public [26].

Sentiment analysis usually tries to determine whether the text expresses a certain emotional tone. In this analysis, words and phrases in the text are analysed to see whether they convey a positive, negative, or neutral emotion. There are three main approaches used in sentiment analysis: rule-based, machine learning-based, and hybrid approaches.

Rule-based sentiment analysis identifies sentiments using predefined rules and lexicons. This approach allows words and sentence structures to be analysed based on certain rules. For example, using a sentiment lexicon, positive or negative words in the text are identified, and the sentiment is determined according to the context of these words in the sentence. This method, although simple and interpretable, may have difficulties in fully understanding the context of the language. It may also fail to recognise complex language features such as irony, nuance, and polysemy [27,28].

Machine learning-based sentiment analysis uses algorithms to classify text. In this approach, models that learn from large datasets use labelled data to determine the emotional tone of the text. Algorithms include support vector machines (SVM), decision trees, naive Bayes, and deep learning techniques [29]. Machine learning-based approaches can better grasp the context and are more robust to the complexity of language. However, the training process of these methods can be time-consuming and requires large, labelled datasets.

Hybrid approaches are a combination of rule-based and machine learning-based methods. These methods aim to achieve a more accurate result by combining the strengths of both approaches. For example, while rule-based methods analyse sentiment for specific words and sentence structures, machine learning algorithms can examine these results more comprehensively. Hybrid approaches utilise the advantages of both methods to address language complexity more effectively [30].

3.4. Machine Learning

Machine learning is a branch of artificial intelligence that enables computers to learn from data and perform certain tasks as they gain experience with the data, without being explicitly programmed [31]. Machine learning methods can generally be classified into three main types: regression, classification, and clustering. Regression aims to predict a continuous target variable. Classification aims to categorise data into certain categories [32]. Clustering, on the other hand, divides data into groups with similar characteristics, is a type of analysis performed with unlabelled data, and is used in applications such as customer segmentation. These three methods are selected according to different data types and analysis needs and are used to solve various problems.

This study not only applied sentiment analysis and topic modelling to social media data but also introduced methodological adaptations to facilitate a cross-linguistic comparison. For instance, preprocessing steps were tailored to accommodate linguistic differences, ensuring that stop-word removal and tokenisation were optimised for both English and Turkish. Additionally, topic modelling was fine-tuned to maintain a balanced thematic distribution despite the disparity in dataset sizes.

In this study, we preferred traditional machine learning methods such as support vector machines (SVM) instead of deep learning models. One of the key reasons was the

relatively small size of the Turkish dataset, which limits the effectiveness of deep learning models that require large amounts of labelled data for training. Furthermore, SVM has been widely applied in sentiment analysis with high accuracy and computational efficiency, particularly in cases where interpretability is important. While transformer-based models (e.g., BERT) were considered, they were not utilised due to computational constraints and the need for extensive labelled data for fine-tuning.

Support Vector Regression

Support vector machines (SVM) is a powerful machine learning algorithm that is widely used, especially for classification problems [33]. SVM aims to separate data with the widest margin between two classes. This algorithm can work effectively even on non-linear datasets because it makes it possible to perform linear classification by transforming the data into a high-dimensional space using kernel functions [34].

SVM allows the data to be separated by the widest possible margin to find the linear separation between two classes. This margin is the distance between the closest points (support vectors) from both classes. The aim of SVM is to maximise this margin, i.e., to find the hyperplane that separates the classes the farthest apart. For a linear classification problem, this hyperplane corresponds to the partitions of the data in a plane [35].

Mathematically, SVM tries to find a linear hyperplane so that the data points are placed on two different sides of this plane. The mathematical formula used for this linear classification is as follows:

$$w * x + b = 0 \quad (1)$$

where w is the normal vector of the linear hyperplane, x is the data point, and b is the bias term. The goal is to find the parameters w and b that will give the best discrimination to the data points.

The main feature of SVM is that it uses data called 'support vectors' in classification. Support vectors are data points in the immediate vicinity of the boundary separating classes, and these points are most critical to the accuracy of the classification model. These support vectors determine the margin and play an important role in training the model [36]. The selection of support vectors is very important for classification accuracy.

The goal of SVM is to maximise the margin (distance) between two classes. This can be expressed mathematically as follows:

$$Maximize\ \frac{2}{\|w\|} \quad (2)$$

In Equation (2), $\|w\|$ is the norm of the linear hyperplane and represents the distance between classes. While the margin distance can theoretically be zero in special cases, the norm $\|w\|$ itself cannot be zero because it represents the vector's magnitude. A zero norm would imply the absence of a separating hyperplane, which contradicts the definition of support vector machines. This distinction is crucial in ensuring the mathematical validity of the model.

SVM can also be used for non-linear classification problems. For this, kernel functions are used that linearise the data by projecting them into a higher-dimensional space. These kernel functions enable linear discrimination by transforming data points into a higher-dimensional space [37]. Some commonly used kernel functions are the following:

Linear Kernel: Performs simple linear classification.

$$K(x, x') = x * x' \quad (3)$$

Polynomial Kernel: Transforms data points into a polynomial of higher order.

$$K(x, x') = (x * x' + c)^d \quad (4)$$

Radial Basis Function (RBF) Kernel: Transforms data into a very high-dimensional space and is particularly suitable for nonlinear classification problems.

$$K(x, x') = \exp\left(-\gamma \|x - x'\|^2\right) \quad (5)$$

These kernel functions are effective tools for finding nonlinear boundaries and enable SVM to deal with nonlinear datasets. SVM can solve nonlinear problems by transforming them into a high-dimensional feature space. SVM can work quite effectively even with small datasets and generally has good generalisation ability. By maximising the margin, the classification accuracy can be high [38].

3.5. Topic Modelling

Topic modelling is a text mining technique for discovering hidden themes (topics) or topical structures from large text datasets. This method aims to extract meaningful and coherent information, especially from unstructured texts [39]. Topic modelling is used to understand how to classify texts into specific topics and what themes each text contains. This technique often finds application in the fields of natural language processing (NLP), text mining, information retrieval, and big data analysis.

Topic modelling essentially represents each document with the probabilities of a set of topics. Each topic is expressed as a distribution of words, and each document is considered as a mixture of these topics. In this way, the main themes in texts can be extracted, and texts can be made more meaningful [40].

The most widely used method of topic modelling is the Latent Dirichlet Allocation (LDA) algorithm. LDA calculates the probability that each document and each word belong to a particular topic. Latent Dirichlet Allocation (LDA) is a topic modelling method that extracts hidden themes (topics) from text data. This method works on the assumption that each document can belong to more than one topic, especially in large text datasets, and extracts words that represent these topics. LDA assumes that each document is composed of one or more topics, and each topic is represented by specific words. The goal is to accurately model each document and the words associated with each topic [41,42].

The basic working principle of LDA assumes that in each text collection (e.g., a set of documents), each document may be generated by several topics, and each topic is represented by certain words. In this case, the themes (topics) contained in each document and the words contained in each topic are latent variables [43,44].

The main purpose of LDA is to determine what topics each document consists of and what words each topic consists of. Mathematically, it represents each document as a mixture of topics and each topic as a mixture of words [45]. To learn the parameters of this model, LDA explores the text data, and the latent variables (topic distributions and word distributions) associated with these data.

The basic components of the LDA model are as follows:

- Number of topics (K): This specifies the number of topics that the model will predict.
- Document topics (θ_d): Each document is represented as a mixture of K topics. Here, θ_d is the topic distribution of the dth document.
- Topic word distributions (ϕ_k): Each topic represents a distribution of words, that is, ϕ_k is the word distribution of the kth topic.
- Word vector (w_d, n): The words that make up each document. Each word is generated by a topic.

We can express the mathematical structure of the LDA model as follows:

$$p(w, z\, I\alpha, \beta) = \prod_{d=1}^{D} \left(\prod_{n=1}^{N_d} p\left(\left(w_{d,n} | z_{d,n}, \beta \right) p\left(z_{d,n} | \theta_d \right) \right) \right) p(\theta_d | \alpha) p(\beta | \eta) \quad (6)$$

where

- $w_{d,n}$ is the nth word in the dth document, and $z_{d,n}$ is the predicted topic label for this word.
- θ_d represents the topic distribution of the dth document.
- β and α are the hyperparameters of the Dirichlet distributions.

This formula shows how LDA models the data and tries to learn topics. Basically, it allows the model to learn the correct topic distributions from given word and document data.

LDA works using a method called "Bayesian learning". This process involves making statistical predictions to understand which topics each document and word belong to. Based on the initialised topic distributions, LDA learns the relationship between documents and words in an iterative process.

This process usually works in the following steps:

At first, it starts with assigning each document and each word to a random topic. This is the beginning of the learning process. At each step of the model, the topic distributions for each word are updated. This update makes a prediction about which topics the words are more likely to occur in. The probabilities between each document and topic are updated. This step determines more precisely which topics each document contains. These steps are repeated until the parameters of the topic and word distributions are learned [46].

There are also more complex and flexible models. These include methods such as non-negative matrix factorisation (NMF) and latent semantic analysis (LSA). NMF factors the text data and determines the contributions of each document and word. LSA reduces the data size by using eigenvalue decomposition to understand similarities between words [47].

4. Results

Before starting the data analysis process, a comprehensive text preprocessing phase was carried out to ensure that the tweets obtained from platform X could be processed appropriately. In this process, repetitive, irrelevant, noisy, and incomplete content was identified and removed from the dataset. The elements considered as noise included advertisements, automated messages from bot accounts, tweets consisting only of links or emojis, and off-topic content. In addition, punctuation marks, special characters, and unnecessary spaces were removed to make the texts grammatically correct. Thus, the dataset was made more coherent and meaningful, and a suitable basis was created to improve the accuracy of the analysis methods.

The raw dataset, which initially consisted of a total of 14,916 English and 1358 Turkish tweets, was optimised for analysis after the text preprocessing stage was completed. In this process, repetitive, off-topic, incomplete, or noisy tweets were removed from the dataset. As a result of text preprocessing, the number of data available for analysis decreased to 7950 for English tweets and 679 for Turkish tweets. This cleaned and structured dataset was made ready and suitable for the next analysis stages.

4.1. Sentiment Analysis

In this study, the support vector machines (SVM) method is used to classify sentences in the dataset according to their emotional states. This method is a machine learning-based classification algorithm that gives successful results in linear or non-linear classification problems. The dataset is divided into two parts: training and test. While training data were used to learn the parameters of the model, test data were used to evaluate the performance of the developed model. After trying different ratios during the parsing of the

dataset, the best result was obtained with a ratio of 70% training and 30% testing. Accordingly, 5566 comments were used for training the model for English tweets, and 2385 comments were used for the testing phase, while 475 comments were used for training and 204 comments were used for the testing phase for Turkish tweets. This process was carried out carefully to increase the overall success and accuracy of the model.

Metrics used to evaluate the performance of classification models play a critical role in measuring the predictive accuracy and efficiency of the model. In this context, metrics such as precision, recall, accuracy, and F1 score (F-measure) are widely preferred. These metrics evaluate different aspects of the model and may have different degrees of importance in various application scenarios.

Precision refers to how accurate the model's positive class predictions are and indicates how many of the positively predicted instances are actually positive. This metric focuses on the accuracy of predictions and is calculated as follows [48]:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{7}$$

Here, TP (True Positive) refers to true positive predictions, and FP (False Positive) refers to false positive predictions.

Recall indicates the rate at which the model correctly predicts the positive class and measures the proportion of samples that should have been predicted as positive. This metric provides information on the cost of missed positive samples and is calculated by the formula [48]:

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{8}$$

Here, FN (False Negative) represents the number of false negative predicted samples.

Accuracy refers to the overall correct classification rate of the model and is calculated as the ratio of total correct predictions to the total number of samples in the dataset [48]:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \tag{9}$$

Here, TN (True Negative) refers to the number of true negative predictions.

The F1 score is a measure of the balance between precision and sensitivity and is calculated as the harmonic mean of these two metrics [48]:

$$\text{F1} = \frac{2 * \text{Recall} * \text{Precision}}{\text{Precision} + \text{Recall}} \tag{10}$$

These metrics were used to comprehensively evaluate the performance of the classification model. The precision, sensitivity, and F1 score produced more meaningful results on imbalanced datasets, while the accuracy metric was used to evaluate the overall success of the model.

It was observed that the model developed with support vector machines gave successful results in user emotion classification. Performance metrics showed that the model was effective in correctly identifying positive examples, and the false positive or false negative prediction rates remained at low levels. These results show that the model is a reliable method for classifying user comments according to their emotional state. Table 2 shows the performance evaluation of the support vector machines.

Table 2. Performance measurement.

Metric	English Tweet	Turkish Tweet
Precision	0.871	0.931
Recall	0.913	0.953
Accuracy	0.905	0.943
F Score (F1)	0.891	0.942

Table 3 shows the comparative results of the metrics used in the performance and success evaluation of the model developed using the support vector machines (SVM) method. The model, which was evaluated separately for English and Turkish tweets, achieved successful results in user emotion classification. The precision metrics are calculated as 0.971 and 0.959, respectively, indicating that the model correctly classifies positive examples in both English and Turkish tweets, and false positive classifications are quite low. Recall metrics show a high value of 0.983 for English tweets, while this value is calculated as 0.908 for Turkish tweets. The high sensitivity in English tweets indicates that the majority of positive examples were successfully detected, and the false negative rate was low. On the other hand, the relatively lower value in Turkish tweets indicates that some positive examples were not detected. The accuracy metrics were 0.958 for English tweets and 0.918 for Turkish tweets. These results show that the overall correct classification rate of the model is higher for English tweets, but a very satisfactory success rate is also achieved for Turkish tweets. The F1 score was calculated as 0.977 for English tweets and 0.933 for Turkish tweets. These values show that while the model correctly detects positive examples in both languages, the false positive and false negative prediction rates remain low.

Table 3. Most frequently used words in English tweets categorised by sentiment.

Positive	Negative	Neutral
(‘clean’, 230)	(‘power’, 246)	(‘Water’, 27)
(‘power’, 201)	(‘fossil’, 197)	(‘Fuel’, 20)
(‘renewable’, 190)	(‘renewable’, 196)	(‘power’, 20)
(‘solar’, 187)	(‘climate’, 189)	(‘government’, 19)
(‘climate’, 170)	(‘coal’, 118)	(‘climate’, 15)

In general, the model developed using the SVM method performed effectively in user emotion classification for both English and Turkish tweets. However, it was observed that the success rate of the model was higher for English tweets compared to Turkish tweets. This may be due to the smaller size of the Turkish dataset or difficulties specific to the language structure. However, it was concluded that the model can be used reliably in both languages.

Figure 1 presents the statistics of English and Turkish tweets according to their sentiment. A total of 3314 positive, 4091 negative, and 545 neutral emotions were detected in English tweets. In Turkish tweets, 130 positive, 133 negative, and 43 neutral sentiments were observed. In both languages, it is noteworthy that tweets containing negative emotions are more numerous than positive ones. This shows that users’ criticisms or concerns about green energy are widespread. Although the rate of neutral sentiment in English tweets is higher than in Turkish tweets, the general trend is similar in both languages.

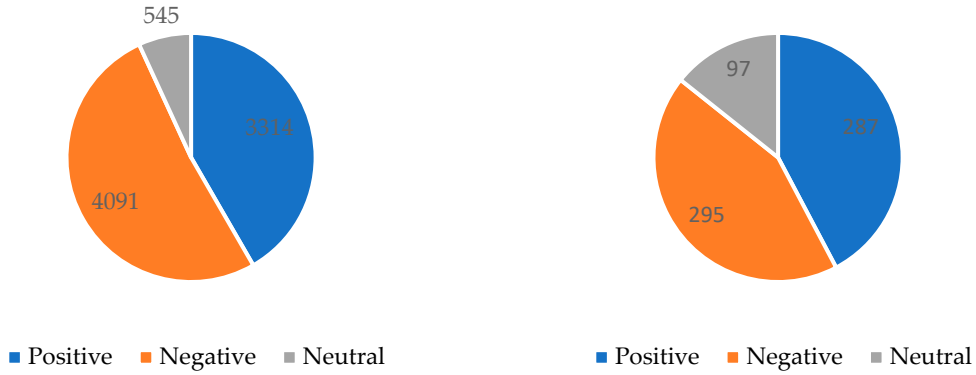


Figure 1. Distribution of sentiments in English and Turkish tweets. (**a**) Number of English tweets, (**b**) number of Turkish tweets.

4.2. Text Mining

In this study, b-gram and TF-IDF methods are used for text mining. N-gram is a grouping method consisting of n consecutive words or characters in a text document. For example, 2-g (bigrams) in a text document represent two consecutive words, and 3-g (trigrams) represent three consecutive words. In this study, the bigram grouping method is preferred.

Term Frequency (TF) refers to how often a particular term occurs in a document. This frequency is usually calculated as the ratio of the number of terms to the total number of words. Inverse Document Frequency (IDF) determines how rare or common a term is. Rare terms receive a higher weight, while common terms receive a lower weight. IDF is usually calculated as the ratio of the number of all documents to the number of documents containing a given term [49].

TF-IDF (Term Frequency-Inverse Document Frequency) is a statistical measure obtained by normalising the frequency of a word in a document by its frequency in all documents. This method is used to determine how important a word is to a document. High TF-IDF values indicate that a word is unique and less associated with a document compared to other documents [50].

The analysis of English tweets about green energy reveals different patterns of word usage in positive, negative, and neutral moods. In positive sentiments, words such as ‘clean’ (230), ‘power’ (201), ‘renewable’ (190), ‘solar’ (187), and ‘climate’ (170) stand out, reflecting a positive perception of the environmental benefits and sustainability of renewable energy sources. On the other hand, negative sentiments were characterised by words such as ‘power’ (246), ‘fossil’ (197), ‘renewable’ (196), ‘climate’ (189), and ‘coal’ (118), reflecting concerns about the dependence on fossil fuels, the viability of renewable energy, and the impacts of climate change. Neutral sentiments are represented by words such as ‘water’ (27), ‘fuel’ (20), ‘power’ (20), ‘government’ (19), and ‘climate’ (15), often reflecting topics that do not carry an emotional tone, such as information transfer or political debates. These findings suggest that an optimistic approach to green energy coexists with criticisms of practical challenges and neutral discussions.

The analysis of Turkish tweets about green energy shows that different themes are focused on in positive, negative, and neutral moods. In positive sentiments, words such as ‘solar’ (13), ‘companies’ (11), ‘technology’ (11), ‘renewable’ (10), and ‘clean’ (10) stand out, showing optimism about solar energy, renewable energy technologies, and the role of companies in this field. Negative sentiments are characterised by the words ‘company’ (28), ‘hydrogen’ (27), ‘renewable’ (9), ‘investments’ (7), and ‘coal’ (7), reflecting concerns

such as criticism of companies, the challenges of hydrogen energy, and the limitations of renewable energy investments. The neutral sentiment category is dominated by words such as 'production' (27), 'storage' (14), 'sustainable' (8), 'battery' (7), and 'procurement' (6), reflecting more objective discussions on energy production, storage technologies, and sustainability. These data reveal that Turkish tweets are a combination of hopes for green energy, criticism, and neutral approaches focusing on technical issues. (See Table 4).

Table 4. Most frequently used words in Turkish tweets categorised by sentiment.

Positive	Negative	Neutral
('solar', 13)	('company', 28)	('production', 27)
('companies', 11)	('hydrogen', 27)	('storage', 14)
('technology', 11)	('renewable', 9)	('sustainable', 8)
('renewable', 10)	('investments', 7)	('battery', 7)
('clean', 10)	('coal', 7)	('procurement', 6)

A general comparison of the most frequently mentioned words related to green energy in English and Turkish tweets reveals that similar themes are prominent in both languages, but the points of emphasis differ.

In English tweets, environmentally friendly, renewable energy sources such as 'clean', 'renewable', 'solar', and 'climate' and positive perceptions about climate change are predominant, while words used in a negative context such as 'fossil', 'coal', and 'power' point to the difficulties of the energy transition and the current fossil fuel dependency. Among the neutral words, political and practical topics such as 'government' and 'fuel' are prominent, suggesting that English-speaking users make connections between politics and energy sources.

In Turkish tweets, words such as 'solar', 'technology', and 'renewable' reflect positive sentiments, while negative sentiment terms such as 'company' and 'hydrogen' focus on criticism of companies and the potential challenges of hydrogen energy. Among neutral words, technical and operational themes such as 'production', 'storage', and 'sustainable' dominate, suggesting that Turkish-speaking users are more focused on energy production and technological infrastructure.

This comparison reveals that English tweets emphasise general awareness and environmental benefits, while Turkish tweets discuss more specific topics such as technology, production, and corporate responsibility. It can be argued that there is a balance of hope and criticism of renewable energy in both languages, but that discourses are shaped by regional and cultural priorities.

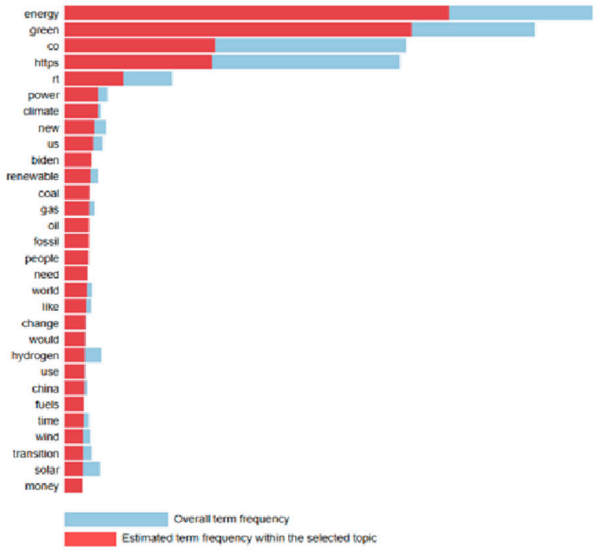
4.3. Topic Modeling

Latent Dirichlet Allocation (LDA) on Turkish and English tweets yielded a consistency score of 0.71 for Turkish tweets and 0.67 for English tweets when the number of topics was set to three. These scores indicate that the topics in both languages have a high semantic similarity and meaningfulness. The coherence score is a metric that measures the quality of a topic model and takes a value between 0 and 1, where 1 indicates perfect coherence, and 0 indicates no coherence. These results suggest that three topics provide an effective balance to capture a sufficient level of detail without overly fragmenting the data in tweets.

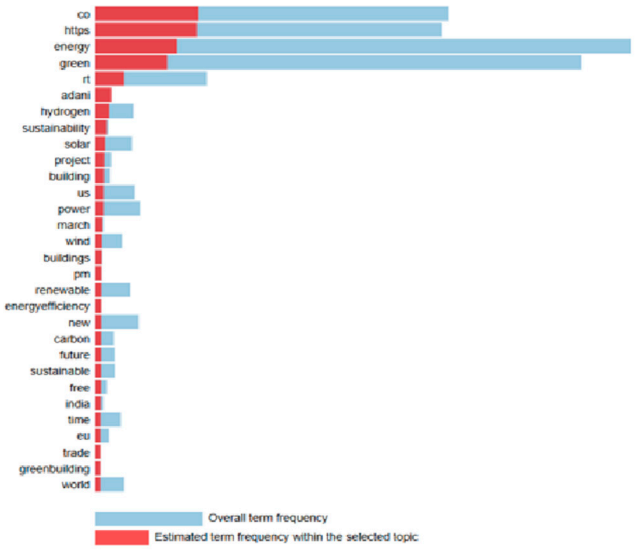
When more topics (e.g., 10 or 15) were attempted, the coherence score decreased, and the resulting topics were ambiguous and difficult to interpret. This suggests that the collection of text in both Turkish and English tweets consists of three distinct and separated topics, confirming the robustness of the modelling approach.

In Figures 2 and 3, the predicted frequency of terms in each topic is visualised in comparison with the overall model frequency. The PyLDAvis library was used for data visualisation. PyLDAvis provides a better interpretation of the topic modelling results for a text data collection. In the visualisations, the blue colour indicates the importance of a word in the overall dataset, and the red colour indicates the importance of the word within the selected topic. The larger the red part of a word, the more specific and meaningful that word is for that topic.

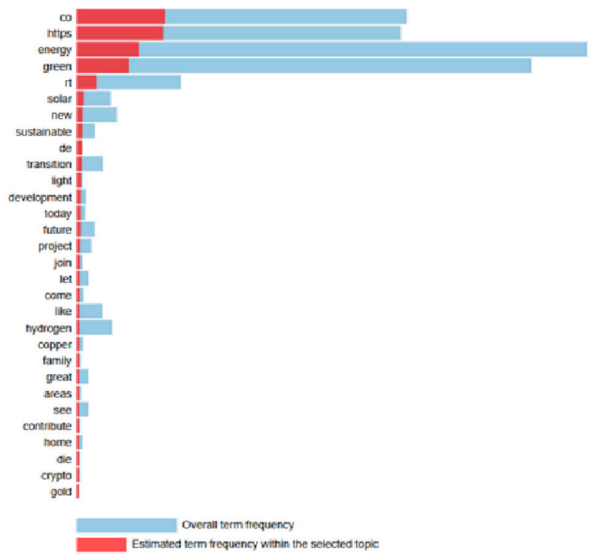
In this context, the words analysed through PyLDAvis in both Turkish and English tweets provide valuable insight to determine user perception and concerns about green energy by revealing the key terms of each topic and the context of these terms within the dataset.



(**a**)



(**b**)



(**c**)

Figure 2. LDA terms (English tweets). (**a**) Topic 1, (**b**) topic 2, (**c**) topic 3.

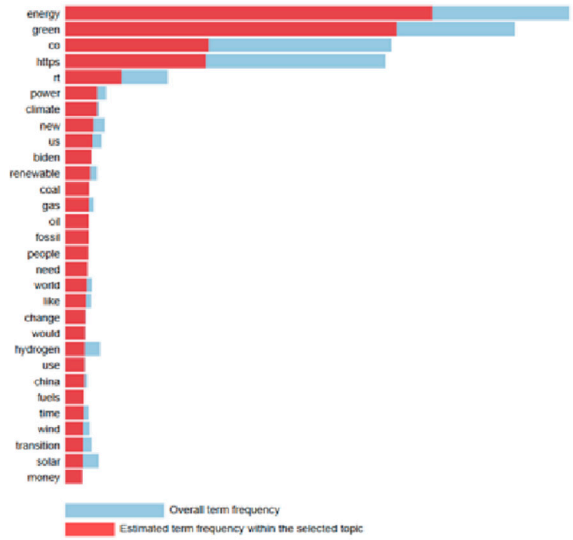
As a result of the analysis of English tweets, three main themes were identified. These themes reveal the multidimensional structure of the discussions on green energy.

The first theme is 'Global Energy Transition and Climate Awareness'. This theme focuses on energy transition and climate change awareness. Terms such as 'energy', 'green', 'power', 'climate', and 'renewable' emphasise the positive approach in the transition to renewable energy. On the other hand, the frequent use of terms such as 'coal', 'fossil', and 'gas' indicates that fossil fuels still play an important role in the energy system and the difficulties in the transition process. Furthermore, the word 'climate' indicates that climate change is a fundamental element shaping energy policies.

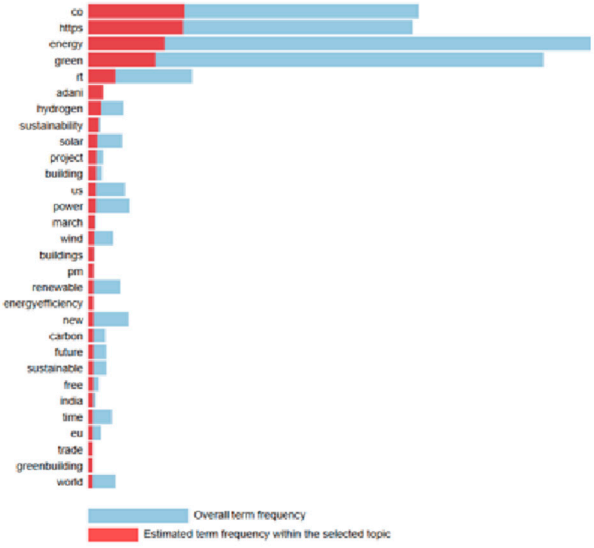
The second theme is 'Sustainability and Infrastructure Development'. This theme focuses on the sustainability objectives and infrastructure development requirements of energy projects. Words such as 'sustainability', 'project', 'building', 'renewable', 'wind', and 'solar' emphasise the growing interest in renewable energy projects and the physical infrastructure requirements of these projects. In particular, the frequent mention of renewable energy sources such as 'wind' and 'solar' reflects the importance of these energy types in sustainability goals. This theme expresses how green energy projects are shaped by future investments.

The third theme is 'New Technologies and Alternative Energy Solutions'. This theme focuses on innovative technologies and alternative energy solutions in the energy sector. Terms such as 'hydrogen', 'transition', 'development', and 'future' reveal the potential of new energy technologies such as hydrogen energy in the energy transition. In addition, words such as 'copper' point to technical and material requirements in the energy infrastructure. This theme reflects the growing interest in innovative technologies in future energy systems and their role in the energy sector.

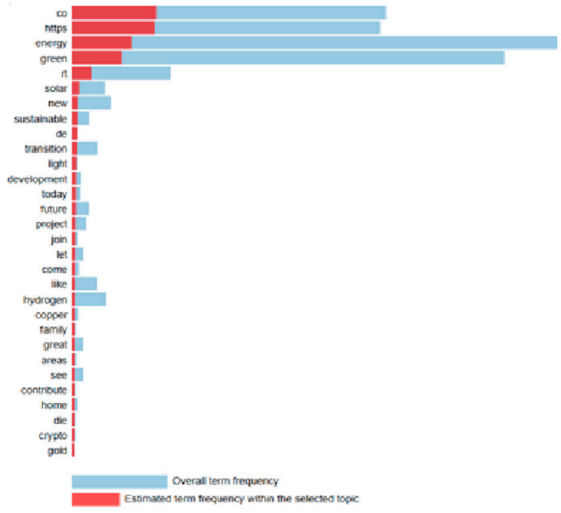
Overall, these three themes show that discussions on green energy in English tweets offer a multifaceted perspective on the global energy transition, sustainable projects, and innovative technologies. This clearly demonstrates the societal and sectoral interest in green energy.



(**a**)



(**b**)



(**c**)

Figure 3. LDA terms (Turkish tweets). (**a**) Topic 1, (**b**) topic 2, (**c**) topic 3.

Three main themes stand out in the analysis of Turkish tweets. The first theme can be defined as 'Energy and Climate Awareness'. This theme is heavily associated with words such as 'energy', 'green', 'power', and 'climate' and points to a strong emphasis on energy transformation and climate change awareness. The fossil fuel-related terms 'coal' and 'fossil' reveal concerns about the difficulties in this transformation process and the still significant role of fossil fuels. This theme reflects a discussion focusing on environmental awareness and climate change in general.

The second theme is 'Renewable Energy Projects and Sustainability'. Here, the prominence of terms such as 'sustainability', 'renewable', 'solar', and 'wind' indicates that renewable energy-based projects and sustainability targets are emphasised. Words such as 'building' and 'project' also indicate the importance of the infrastructure and investments required for the construction and implementation of these projects. This theme expresses a strong social demand for increasing renewable energy projects and a strong sense of environmental responsibility.

The third theme is 'Alternative Technologies and Future Perspective'. This theme is characterised by terms such as 'hydrogen', 'development', 'transition', and 'future' and emphasises the interest in innovative technologies for the energy sector. Alternative energy solutions such as hydrogen energy play an important role in the energy transition in the future. Words such as 'copper' and 'sustainable' emphasise the material and sustainability requirements for the infrastructure of these new technologies. This theme underlines the technological developments in the field of green energy and their future potential.

Overall, these three themes show that green energy discussions in Turkish tweets are multidimensional, focusing on key issues such as energy transition, sustainable projects, and innovative technologies. This clearly reflects the societal and sectoral interest in green energy.

5. Discussion

According to the sentiment statistics of English and Turkish tweets, it is seen that the number of tweets containing negative sentiment is higher than the number of tweets containing positive sentiment in both languages. While 4091 negative, 3314 positive, and 545 neutral emotions were recorded in English tweets, 133 negative, 130 positive, and 43 neutral emotions were detected in Turkish tweets. This shows that users express their concerns and criticisms about green energy, but positive opinions are generally less widespread. The higher rate of neutral sentiment in English tweets compared to Turkish tweets may indicate that English-speaking users may have developed a more analytical and objective perspective on this issue. The higher proportion of negative sentiment in both languages suggests that negative perceptions of green energy are widespread and are more likely to be met with criticism.

When the most frequently mentioned words related to green energy in English and Turkish tweets are compared, it is observed that while similar themes related to environmentally friendly and renewable energy sources come to the forefront in both languages, the points of emphasis differ. In English tweets, terms such as 'clean', 'renewable', 'solar', and 'climate' reflect a positive view of environmentally friendly energy sources and climate change, while words such as 'fossil', 'coal', and 'power' emphasise the difficulties in energy conversion and the continued dependence on fossil fuels. Similarly, in Turkish tweets, the words 'solar' and 'renewable' indicate a positive energy transition, while terms such as 'company' and 'hydrogen' reflect a critical perspective, especially corporate responsibility and the challenges faced by hydrogen energy. Furthermore, operational issues such as energy generation and storage are more prominent in the Turkish data, suggesting that users are having a more technical discussion about energy infrastructure. In

conclusion, it can be said that there is a balance of positive and critical views on green energy in both languages, but each language focuses on different areas of discussion due to its own cultural and regional context.

When the results of modelling two different topics are compared with the data obtained from Turkish and English green energy tweets, significant differences and similarities emerge. In English tweets, Topic 1 emphasises a strong focus on energy and climate change, Topic 2 focuses on renewable energy projects and sustainability discussions, and Topic 3 addresses future alternative energy solutions and technological developments. In this model, the words mostly refer to the global energy transition, innovative technologies to replace fossil fuels, and the need for sustainable energy infrastructures.

In Turkish tweets, similarly, Topic 1 focuses on energy transition and climate awareness, while Topic 2 focuses on renewable energy projects and sustainability. However, in the Turkish model, Topic 3 includes more words related to alternative energy technologies and strategies for the future, shaping a discussion focused on technology and innovation. The differences between the English and Turkish models point to different ways of interpretation arising from linguistic and cultural contexts. While the English data include more industrial and commercial terms as well as energy changes, especially on a global scale, the Turkish data emphasise more socially and environmentally orientated discussions. These differences reveal that the perceptions and expectations of users in both languages regarding green energy are shaped by local economic and environmental contexts.

6. Conclusions

The aim of this study is to comparatively analyse the views on green energy in English and Turkish social media posts and to identify the perceptions, sentiments, and discussion foci of users in both languages. This article analyses the tweets obtained from social media data using sentiment analysis, word frequency analysis, and topic modelling methods using artificial intelligence-based text mining methods. This study aims to gain an in-depth understanding of user views on green energy in Turkish and English contexts and to explore how social media dynamics in both languages are reflected in green energy debates.

The main purpose of this study is to comparatively analyse tweets about green energy shared on English and Turkish social media platforms. For this purpose, users' sentiments about green energy, which themes are prominent, and which words are used more frequently were analysed. By identifying existing green energy discourses in both languages, this study investigates how cultural and linguistic differences shape these discourses. The results present similar and different perspectives of users of two different languages on green energy, considering social, environmental, and economic contexts.

This study aims to fill an important gap in the field by analysing social media data on green energy in a linguistic and cultural context. While previous studies on green energy have generally focused on a specific language or geography, the originality of this study lies in analysing social perceptions and sentiments in two different languages by comparing Turkish and English tweet data. Furthermore, combining sentiment analysis and topic modelling methods to conduct an in-depth thematic analysis of social media data is an innovative approach. This innovation allows the findings of this study to have important implications for social media influence, environmental policies, and green energy policies.

The findings of this study show that in both languages, tweets containing negative emotions are more numerous than those containing positive emotions. While 4091 negative, 3314 positive, and 545 neutral sentiments were recorded in English tweets, 133 negative, 130 positive, and 43 neutral sentiments were detected in Turkish tweets. This reveals that users' concerns and criticisms about green energy are widespread, but positive views

are shared less. Linguistic differences show that similar themes are prominent in both languages, but the points of emphasis differ. While positive themes related to environmentally friendly and renewable energy sources are prominent in English tweets, more technical and operational themes such as technology, production, and corporate responsibility are discussed in Turkish tweets.

The topic modelling results also reflect linguistic differences. In English tweets, themes such as energy transformation, dependence on fossil fuels, renewable energy projects, and technological developments are prominent, whereas in Turkish tweets, criticisms on energy production, storage, and hydrogen energy are particularly prominent. These differences reveal the different perceptions and emphases of users in both languages on green energy. While Turkish-speaking users discuss green energy from a social and environmental responsibility perspective, English-speaking users approach it from a more global and industrial perspective.

This study provides important findings on how social media perceptions of green energy can be related to the climate research, media, and digital marketing sectors. Especially from a digital marketing perspective, analysing user perceptions and emotions in different languages and cultures makes it possible to develop strategies sensitive to regional differences in the marketing of environmentally friendly products and services. In this context, marketing communication strategies can be supported by such insights derived from social media data to create more effective messages for target audiences. Increasing brand loyalty with environmental awareness themed messages can be at the centre of these strategies. The findings show that the themes of environmental awareness and climate change are prominent in English tweets, while technical and operational issues are more prominent in Turkish tweets. This indicates that communication strategies for two different audiences should be diversified in terms of message content and emphasis points.

The results of this article can be summarised as follows:

- The sentiment analysis results indicate that negative perceptions about green energy are more dominant in both Turkish and English tweets.
- The topic modelling results reveal that English tweets focus on global energy policies, while Turkish tweets highlight local technical challenges.
- The comparison suggests that cultural and linguistic factors play a significant role in shaping consumer views on green energy.
- These findings provide insights for policymakers and marketers aiming to develop region-specific green energy communication strategies.

This study provides important findings on how social media perceptions of green energy can be related to climate research, media, and digital marketing sectors. Analysing user perceptions and emotions in different languages and cultures, especially from a digital marketing perspective, makes it possible to develop strategies that are sensitive to regional differences in the marketing of environmentally friendly products and services. The findings show that environmental awareness and climate change themes are prominent in English tweets, while technical and operational issues are more prominent in Turkish tweets. This indicates that communication strategies for two different audiences need to be diversified in terms of message content and emphasis points.

Author Contributions: Conceptualization, Y.B.; Data curation, Y.B., M.K., and F.Y.A.; Formal analysis, F.Y.A. and M.G.; Funding acquisition, F.K., A.B., and M.G.; Investigation, F.K., M.K., and A.B.; Methodology, Y.B., M.K., F.Y.A., and M.G.; Project administration, F.K. and Y.B.; Resources, A.B. and M.G.; Software, F.K., M.K., and F.Y.A.; Supervision, F.K.; Validation, A.B. and M.G.; Visualization, F.K.; Writing—original draft, Y.B., M.K., and F.Y.A.; Writing—review and editing, F.K., A.B., and M.G. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The dataset is available from the authors upon reasonable request.

Conflicts of Interest: The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References

1. Marks-Bielska, R.; Bielski, S.; Pik, K.; Kurowska, K. The importance of renewable energy sources in Poland's energy mix. *Energies* **2020**, *13*, 4624.
2. Omer, A.M. Green energies and the environment. *Renew. Sustain. Energy Rev.* **2008**, *12*, 1789–1821.
3. Panwar, N.L.; Kaushik, S.C.; Kothari, S. Role of renewable energy sources in environmental protection: A review. *Renew. Sustain. Energy Rev.* **2011**, *15*, 1513–1524.
4. Owusu, P.A.; Asumadu-Sarkodie, S. A review of renewable energy sources, sustainability issues and climate change mitigation. *Cogent Eng.* **2016**, *3*, 1167990.
5. Sovacool, B.K. Rejecting renewables: The socio-technical impediments to renewable electricity in the United States. *Energy Policy* **2009**, *37*, 4500–4513.
6. Wolsink, M. Planning of renewables schemes: Deliberative and fair decision-making on landscape issues instead of reproachful accusations of non-cooperation. *Energy Policy* **2007**, *35*, 2692–2704.
7. Ellis, G.; Barry, J.; Robinson, C. Many ways to say 'no', different ways to say 'yes': Applying Q-methodology to understand public acceptance of wind farm proposals. *J. Environ. Plan. Manag.* **2007**, *50*, 517–551.
8. Gross, C. Community perspectives of wind energy in Australia: The application of a justice and community fairness framework to increase social acceptance. *Energy Policy* **2007**, *35*, 2727–2736.
9. Wüstenhagen, R.; Menichetti, E. Strategic choices for renewable energy investment: Conceptual framework and opportunities for further research. *Energy Policy* **2012**, *40*, 1–10.
10. Foust, C.R.; O'Shannon Murphy, W. Revealing and reframing apocalyptic tragedy in global warming discourse. *Environ. Commun.* **2009**, *3*, 151–167.
11. Whitmarsh, L.; O'Neill, S.; Lorenzoni, I. Public engagement with climate change: What do we know and where do we go from here? *Int. J. Media Cult. Politics* **2013**, *9*, 7–25.
12. Johnson, N.; Krey, V.; McCollum, D.L.; Rao, S.; Riahi, K.; Rogelj, J. Stranded on a low-carbon planet: Implications of climate policy for the phase-out of coal-based power plants. *Technol. Forecast. Soc. Change* **2015**, *90*, 89–102.
13. Douglas, M. *Natural Symbols: Explorations in Cosmology*; Barrie & Rockliff: London, UK, 1970.
14. Hulme, M. Why we disagree about climate change. *Zygon* **2015**, *50*(4), 893-905.
15. Pasqualetti, M.J. Opposing wind energy landscapes: A search for common cause. *Ann. Assoc. Am. Geogr.* **2011**, *101*, 907–917.
16. Wu, S.; Tahri, O.; Shen, S.; Zhang, W.; Muhunthan, B. Environmental impact evaluation and long-term rutting resistance performance of warm mix asphalt technologies. *J. Clean. Prod.* **2021**, *278*, 123938.
17. Castillo-Manzano, J.I.; Castro-Nuño, M.; López-Valpuesta, L.; Sanz-Díaz, M.T.; Yñiguez, R. To take or not to take the laptop or tablet to classes, that is the question. *Comput. Hum. Behav.* **2017**, *68*, 326–333.
18. Devine-Wright, P. Energy citizenship: Psychological aspects of evolution in sustainable energy technologies. In *Governing Technology for Sustainability*, Routledge: New York, NY, USA, 2012; pp. 63–86.
19. Sloan, L.; Morgan, J.; Burnap, P.; Williams, M. Who tweets? Deriving the demographic characteristics of age, occupation and social class from Twitter user meta-data. *PLoS ONE* **2015**, *10*, e0115545.
20. Talib, R.; Hanif, M.K.; Ayesha, S.; Fatima, F. Text mining: Techniques, applications and issues. *Int. J. Adv. Comput. Sci. Appl.* **2016**, *7*, 414–418.
21. Karami, A.; Lundy, M.; Webb, F.; Dwivedi, Y.K. Twitter and research: A systematic literature review through text mining. *IEEE Access* **2020**, *8*, 67698–67717.
22. Munková, D.; Munk, M.; Vozár, M. Data pre-processing evaluation for text mining: Transaction/sequence model. *Procedia Comput. Sci.* **2013**, *18*, 1198–1207.

23. Chai, C.P. Comparison of text preprocessing methods. *Nat. Lang. Eng.* **2023**, *29*, 509–553.
24. Schonlau, M.; Guenther, N. Text mining using n-grams. *Schonlau M., Guenther N. Sucholutsky I. Text Min. Using N-Gram Var. Stata J.* **2017**, *17*, 866–881.
25. Gupta, S.; Singh, A.; Kumar, V. Emoji, text, and sentiment polarity detection using natural language processing. *Information* **2023**, *14*, 222.
26. Liu, B. *Sentiment Analysis and Opinion Mining*; Springer Nature: Berlin/Heidelberg, Germany, 2022.
27. Yenkikar, A.; Babu, N.; Sangve, S. R-SA: A Rule-based Expert System for Sentiment Analysis. In Proceedings of 2019 IEEE Pune Section International Conference (PuneCon), Pune, India, 18–20 December 2019; pp. 1–7.
28. Çaylak, P.Ç.; Kayakuş, M.; Eksili, N.; Yiğit Açikgöz, F.; Coşkun, A.E.; Ichimov, M.A.M.; Moiceanu, G. Analysing Online Reviews Consumers' Experiences of Mobile Travel Applications with Sentiment Analysis and Topic Modelling: The Example of Booking and Expedia. *Appl. Sci.* **2024**, *14*, 11800.
29. Hassan, S.U.; Ahamed, J.; Ahmad, K. Analytics of machine learning-based algorithms for text classification. *Sustain. Oper. Comput.* **2022**, *3*, 238–248.
30. Hammi, S.; Hammami, S.M.; Belguith, L.H. A hybrid method for enhancing aspect sentiment classification in the french language: Combining rule-based and learning methods. *Int. J. Data Sci. Anal.* **2024**, 1–21. Doi: 10.1007/s41060-024-00667-x
31. Mehrotra, D. *Basics of Artificial Intelligence & Machine Learning*; Notion Press: Chennai, India, 2019.
32. Urso, A.; Fiannaca, A.; La Rosa, M.; Ravì, V.; Rizzo, R. Data mining: Classification and prediction. *Encycl. Bioinform. Comput. Biol. ABC Bioinform.* **2018**, *384*. Doi: 10.1016/B978-0-12-809633-8.20461-5
33. Cervantes, J.; Garcia-Lamont, F.; Rodríguez-Mazahua, L.; Lopez, A. A comprehensive survey on support vector machine classification: Applications, challenges and trends. *Neurocomputing* **2020**, *408*, 189–215.
34. Bhavsar, H.; Panchal, M.H. A review on support vector machine for data classification. *Int. J. Adv. Res. Comput. Eng. Technol. (IJARCET)* **2012**, *1*, 185–189.
35. Kayakuş, M.; Yiğit Açikgöz, F.; Dinca, M.N.; Kabas, O. Sustainable Brand Reputation: Evaluation of iPhone Customer Reviews with Machine Learning and Sentiment Analysis. *Sustainability* **2024**, *16*, 6121.
36. Salcedo-Sanz, S.; Rojo-Álvarez, J.L.; Martínez-Ramón, M.; Camps-Valls, G. Support vector machines in engineering: An overview. *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.* **2014**, *4*, 234–267.
37. Jäkel, F.; Schölkopf, B.; Wichmann, F.A. A tutorial on kernel methods for categorization. *J. Math. Psychol.* **2007**, *51*, 343–358.
38. Chauhan, V.K.; Dahiya, K.; Sharma, A. Problem formulations and solvers in linear SVM: A review. *Artif. Intell. Rev.* **2019**, *52*, 803–855.
39. Chu, K.E.; Keikhosrokiani, P.; Asl, M.P. A topic modeling and sentiment analysis model for detection and visualization of themes in literary texts. *Pertanika J. Sci. Technol* **2022**, *30*, 2535–2561.
40. Rubin, T.N.; Chambers, A.; Smyth, P.; Steyvers, M. Statistical topic models for multi-label document classification. *Mach. Learn.* **2012**, *88*, 157–208.
41. Negara, E.S.; Triadi, D. Topic modeling using latent dirichlet allocation (LDA) on twitter data with Indonesia keyword. *Bull. Soc. Inform. Theory Appl.* **2021**, *5*, 124–132.
42. Hagen, L. Content analysis of e-petitions with topic modeling: How to train and evaluate LDA models? *Inf. Process. Manag.* **2018**, *54*, 1292–1307.
43. Maier, D.; Waldherr, A.; Miltner, P.; Wiedemann, G.; Niekler, A.; Keinert, A.; Pfetsch, B.; Heyer, G.; Reber, U.; Häussler, T. Applying LDA topic modeling in communication research: Toward a valid and reliable methodology. In *Computational Methods for Communication Science*, Routledge: New York, NY, USA, 2021; pp. 13–38.
44. Yiğit Açikgöz, F.; Kayakuş, M.; Zăbavă, B.-Ș.; Kabas, O. Brand Reputation and Trust: The Impact on Customer Satisfaction and Loyalty for the Hewlett-Packard Brand. *Sustainability* **2024**, *16*, 9681.
45. Onan, A.; Korukoglu, S.; Bulut, H. LDA-based topic modelling in text sentiment classification: An empirical analysis. *Int. J. Comput. Linguist. Appl.* **2016**, *7*, 101–119.
46. Suominen, A.; Toivanen, H. Map of science with topic modeling: Comparison of unsupervised learning and human-assigned subject classification. *J. Assoc. Inf. Sci. Technol.* **2016**, *67*, 2464–2476.
47. Hassani, A.; Iranmanesh, A.; Mansouri, N. Text mining using nonnegative matrix factorization and latent semantic analysis. *Neural Comput. Appl.* **2021**, *33*, 13745–13766.
48. Miao, J.; Zhu, W. Precision–recall curve (PRC) classification trees. *Evol. Intell.* **2022**, *15*, 1545–1569.

49. Qaiser, S.; Ali, R. Text mining: Use of TF-IDF to examine the relevance of words to documents. *Int. J. Comput. Appl.* **2018**, *181*, 25–29.
50. Nafis, N.S.M.; Awang, S. An enhanced hybrid feature selection technique using term frequency-inverse document frequency and support vector machine-recursive feature elimination for sentiment classification. *IEEE Access* **2021**, *9*, 52177–52192.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.