Renewable and Sustainable Energy Reviews 124 (2020) 109781

Contents lists available at ScienceDirect

Renewable and Sustainable Energy Reviews

journal homepage: http://www.elsevier.com/locate/rser

# Energy choices in Alaska: Mining people's perception and attitudes from geotagged tweets

Moloud Abdar [a,1], Mohammad Ehsan Basiri [b], Junjun Yin [c], Mahmoud Habibnezhad [d], Guangqing Chi [e], Shahla Nemati [b], Somayeh Asadi [d,*]

[a] Institute for Intelligent Systems Research and Innovation (IISRI), Deakin University, Australia
[b] Department of Computer Engineering, Shahrekord University, Shahrekord, Iran
[c] Department of Social Science Research Institute and an associate in the Institute for Cyberscience, Pennsylvania State University, University Park, PA, 16802, USA
[d] Department of Architectural Engineering, Pennsylvania State University, 104 Engineering Unit A, University Park, PA, 16802, USA
[e] Department of Agricultural Economics, Sociology, and Education, Population Research Institute, and Social Science Research Institute, The Pennsylvania State University, 112E Armsby, University Park, PA, 16802, USA

A R T I C L E I N F O

*Keywords:*
Social media
Twitter
Sentiment analysis
People's perception
Energy and renewable energy

A B S T R A C T

Alaska is at the forefront of climate change and subject to salient challenges including energy consumption. It is important to understand Alaskans' perceptions and opinions about energy consumption to solve Alaska's domestic energy problems and creating a sustainable future. However, it is challenging to collect public opinions about energy consumption using conventional survey methods, which are often expensive, labor-intensive, and slow. This study utilizes information-rich Twitter data to investigate Alaskans' perceptions and opinions on various energy sources and in particular clean energy sources. Using the geotagged Twitter data collected in Alaska from 2014 to 2016, a lexicon-based sentiment analysis approach was first applied to analyze the polarity in the expressed opinions. Further, a novel fuzzy-based theory is employed to derive the sentiment of the opinion in each tweet. The results indicate that there is a valuable growth rate for a set of energy-related keywords, such as "sun", "power", and "nuclear". The rank of top 20 renewable energy-related keywords shows the word "Tidal" has the highest ranking followed by "solar panel". Moreover, the attention to various types of energy is increasing dramatically among Alaskans. Importantly, Alaskans' attitudes toward energy and renewable energy changed positively from 2014 to 2016, indicating that Alaskans' energy choices are more acceptive towards or even favor renewable energy in the future.

## 1. Introduction

Worldwide energy consumption is anticipated to increase by up to 50% by 2035 [1]. In Alaska, the total energy consumption per capita was 809 million BTU in 2016 which is ranked third after Louisiana and Wyoming [1]. The statewide weighted average residential rate for electricity was 17.6 cents per kWh, higher than the U.S. average of 11.8 cents per kWh [2]. Not only many Alaskans live in energy poverty and pay an average of $800 per month just to electrify and heat their homes, but also most communities in rural Alaska depend on volatile and expensive fossil fuels for electricity generation. This high price of energy is due to the cost of hauling fossil fuels (primarily diesel) by plane or barge to these remote areas [2]. As Alaska is at the forefront of climate change, it is subjected to salient challenges including energy consumption and sources. To overcome these challenges, various strategies including energy conservation and efficiency [3–6] and adoption of clean energy [7–9] should be implemented considering the fact that Alaska is uniquely endowed with a full range of renewable energy opportunities. On one hand, Alaska has a nonbinding goal to generate 50% of its electricity from renewable sources by 2025 [1]. On the other hand, using these sources to develop practical renewable energy solutions will inevitably have effects on the security of food and water systems. Therefore, the critical first step in solving Alaska's domestic energy problems and creating a sustainable future is to understand Alaskans' perceptions and opinions about energy consumption and adoption of various energy sources, in particular renewable energy solutions.

* Corresponding author.
*E-mail address:* sxa51@psu.edu (S. Asadi).
[1] The first and second authors (M. Abdar and M. E. Basiri) contributed equally to this study.

https://doi.org/10.1016/j.rser.2020.109781
Received 16 March 2019; Received in revised form 12 November 2019; Accepted 13 February 2020
Available online 29 February 2020

However, an important factor that hinders the widespread adoption of renewable energy systems, besides its cost, is the lack of proper public information about renewable energy resources and the lack of "social acceptance" assessment. Social acceptance evaluates the "degree of readiness of citizens to invest in renewable energy in their area" [10]. Another study defined social acceptance as a "degree of the active or passive attitude of citizens towards different clean technologies or products" and willingness to pay for clean electricity [10].

A typical approach to understanding people's perceptions is to conduct surveys [11]. However, it is costly to conduct surveys in rural communities of Alaska; also, trust needs to be gained before conducting surveys in Indigenous communities, which would take time. As social media platforms are becoming major arenas for communication and information exchange, social media data can help detect communities' changing needs by tracking topical changes and identifying the influential factors. The microblogging platform Twitter provides an exponentially advantageous venue to study people's interactions with various topics [12,13].

Twitter allows its users to post short, 280-character messages [14,15] and to follow messages from other users while 140-character messages were previously allowed [14,16]. Interactions among Twitter users lead to a network topology characterized by opinion leaders followed by average users, placing Twitter in between a purely social network and a purely informational network. The information network nature of Twitter simplifies and expedites information dissemination; its social network nature eases the access to geographically and personally relevant information [12,17]. Due to the increasing importance of energy efficiency measures, availability of energy sources, and the application of clean energy sources, people are more likely to share their thoughts and sentiments on Twitter.

In view of the aforementioned gaps and discussions, the goal of this study is to exploit information-rich geotagged Twitter data for mining Alaskans' perceptions and opinions about energy efficiency measures, availability of energy sources, and the application of clean energy sources. To achieve this goal, the opinions of Alaskans (geotagged Twitter data) from 2014 to 2016 were collected and their opinions were analyzed using a lexicon-based sentiment analysis approach. Moreover, a novel fuzzy-based Dempster-Shafer (DS) theory is applied to investigate Alaskans' tweets. This study aims to answer the following questions related to energy and renewable energy separately:

1 Research question 1: What is the perception of Alaskans about various energy sources?
2 Research question 2: Do Alaskans use Twitter to make their voice on various sources of energy and renewable energy or lack of proper energy heard?
3 Research question 3: How do the online public sentiments about various energy sources and the application of renewable energy in Alaska change over time?

Responding to these gaps and questions could lead to the following advances. First, based on the experimental results, more popular energy resources (the crowd preference for specific energy resource(s)) can be identified. Second, our newly proposed approach can be benchmarked against other machine learning algorithms used in geotagged Tweets (or even Tweets in other domains such as disease outbreaks and disaster analysis) and serve as a universal perception and opinion mining model that can be applied based on daily Tweets worldwide.

As a summary, the main contributions of this study can be listed as follows:

1. This study proposes a new social media-based analysis approach (here Twitter) to mine public opinions towards energy-related topics. The proposed approach, which is validated and tested using real-world Twitter data for Alaskans' perceptions and opinions, reveals what the crowd preferences are for a specific energy resource.
2. The results of this research can be used by energy companies, utilities, and policy makers to provide proper energy resources to people based on their preferences and/or assist them to make an inform decision about the best sources of energy in terms of cost or other criteria.
3. In this research, the crowd preference can help find out the geographic preference of users. In other words, we can categorize the preference of users based on their locations. This benefits to provide useful services based on the user's needs in their location.
4. A lexicon-based method for sentiment analysis of tweets is implemented in this study. In this method, the effect of negations and intensifiers on the overall sentiment of the tweets is considered. The proposed method, in comparison to machine learning methods, is fast, scalable, and domain-independent.
5. An evidential fuzzy method is proposed to fuse phrase-level sentiment scores into tweet-level overall sentiment scores. The proposed method exploits a Gaussian fuzzy membership function as the mass function of the Dempster-Shafer aggregation method to assign four probabilities to each sentiment-bearing term. The two-point aggregation method is employed to consider both the first and the second most probable emotion class predicted by the proposed method.
6. A comprehensive analysis of Alaskans' tweets in three years is presented to show the frequency and change rate of positive, negative, and neutral sentiments toward energy in Alaska in 2014, 2015, and 2016. Moreover, the trend of changing four different sentiments in three years are analyzed.

This study open doors for future research by identifying the concepts that can be used to initiate a research agenda on analyzing the semantics of conversations in social media regarding energy and climate change with the aim of propelling further research in this domain. To the best of our knowledge, existing studies on public attitude and response to energy and climate change-related issues are based on offline studies with a limited number of survey participants and further are limited with geographical locations. This study overcomes this vital challenge and covers a larger group of individuals in Alaska by analyzing twitter data collected over three years. Very little has been done to use this massive resource to investigate perceptions and attitudes about energy-related topics at a public level. The proposed method in this study for clustering the tweets into four emotional categories has not been previously applied to similar problems. Our proposed approach will allow for global analysis of public response allowing for further assessment of energy and climate change issues.

Moreover, the proposed method has the following novelty in compare with the existing methods for sentiment analysis of users' tweets:

1. A new fuzzy mapping of sentiment scores into two–dimensional valence–arousal (VA) space.
2. A new evidential sentiment aggregation method based on the Dempster-Shafer theory for aggregation of phrase-level sentiment scores into tweet-level scores.
3. An improved aggregation mechanism based on the two-point structure for considering the second-best sentiment predicted class in addition to the most probable one in the process of sentiment score aggregation.

The remaining paper is structured as follows. The second section reviews the main findings of the relevant literature on opinion mining. Section 3 discusses the proposed methodology used in this study. The obtained results and discussion are presented in Section 4. And finally, Section 5 presents the main conclusions.

## 2. Background

Climate change causes substantial environmental, social, and economic risks for current and future generations. Recent studies emphasize the primary role of human activities on the fast changes experienced in the global climate during the last several decades which resulted in emissions of greenhouse gases [18]. According to data published by International Energy Agency, between 1971 and 2014, worldwide primary energy use has increased by 2.5 and carbon dioxide emissions have been doubled. Bach [19] reported a rising trend in global population, energy consumption, and economic activity which would increase the average temperature by 1.5 °C to 3 °C by 2050, due to the increase of anthropogenic $CO_2$ emissions.

The need for adaptation to and mitigation of climate change is identified as a significant challenge for scientists, decision-makers, and the general public. Studies have been done to conduct an evaluation and thorough examination of renewable energy technologies and their current and potential role in the extenuation of greenhouse gas emissions ranging from technology-specific studies, policy, characteristics and technical potentials of different resources, the challenges of their integration and social and environmental impacts of their use and cost [20]. The Kyoto Protocol and the Paris Agreement have directed some of the countries in an extraordinary shift towards renewable energy sources [21].

Existing studies on public attitudes toward energy consumption have primarily relied on opinion polls, interview, and/or surveys at the group level. A substantial amount of studies focused on public support for different energy sources. In the 1970s, solar energy and energy conservation policies have been used in various public opinion polls as alternatives for oil, coal, and nuclear energy [22–25]. Later in 2008, a poll was performed by International Policy Attitudes to investigate the application of renewable energy in 21 countries with 20,790 respondents [26]. The results indicated strong public support (77%) of governmental efforts to shift to increase renewable energy sources especially solar energy. A similar survey poll was conducted in the US in 2008 and 90% of the participants were in favor of renewable energy [22]. Another study was conducted in Malaysia to understand the views and perceptions of the local population towards solar energy and the installation of photovoltaics. The results indicated the lack of information about all the potential incentives and the socio-economic benefits of investing in solar panels [27]. In a similar study, Eshchanov et al. [28] conducted a closed-form of a questionnaire to examine the perception and opinion of rural and urban people of the Khorezm Province, Uzbekistan on renewable energy generation. Respectively, 95 and 55 people from rural and urban areas participated in this study. Results indicated that the cost of renewable energy facilities and incapability to entirely substitute fossil fuel energy sources could be a hindrance while the availability of crediting and public awareness may serve as an enticement. In India, an interview of 1675 homeowners and businesses was performed to understand their awareness and opinion about India's renewable energy and power situation. The interview process took several months, and it was found that people still need to be educated about renewable energy sources to make more informed decisions. Another survey study was conducted in Canada, the United States, and Mexico to understand public perception and policy preferences for renewable energy and its relation to global climate change. Three countries had different opinions and perceptions about climate change with Mexico indicating the highest levels of concern and the US the lowest. In addition, Mexico strongly supports the adoption of renewable energy sources [29]. Yet, the sample size was considerably small compared to the population with a total number of 2312. The small size sample was one of the key barriers in more informed decision-making. Hence, a larger sample size per country is needed to better understand the perceptual differences between countries and identify the affecting variables. This considers as a severe challenge that all the survey studies face. In addition, such surveys have failed to sufficiently elucidate the nature and complexity of people's perceptions about energy sources, energy efficiency measures, renewable and clean energy sources and cannot attain a thorough examination of issues, beliefs, perceptions, and attitude as functions of sustainable behavior relating to the public.

Besides the above research areas, recently social media technology has become pervasive which encourages researchers to leverage the potential value of the massive amount of information posted online about myriad topics. Social media is used as a vital communication channel for countless users to exchange information at any time and place. With billions of Global Positioning System (GPS)-enabled smartphones in use around the world, every single person can perform as an intelligent agent collecting data about the environment and shares opinions and feelings on social media in real-time and at different locations. Researchers used social media data such as Twitter focusing on locational property and content. Twitter data has been leveraged successfully to expedite the knowledge discovery process in a wide variety of fields due to its accessibility to the enormous real-time, large-scale, and fast propagation data, along with researchers' growing capability to process "Big Data."

One strand of research in the engineering domain used Twitter as a valuable data source for disaster management, emergency information dissemination, situational awareness, and crisis-related responses. Goodchild and Glennon [30] used geospatial data to investigate the main issues associated with volunteered geographic information (VGI) and its potential role in disaster management during the four wildfires that impacted the Santa Barbara area in 2007–2009. In another study, Earle et al. [31] compared the potential capacity of Twitter in reporting an earthquake and predicting its impacts with traditional monitoring methods. It was found that Twitter is a more effective and faster tool in identifying affected areas. Some studies conducted textual content analysis to identify whether a tweet is related to the disaster or not. These tweets are further utilized to provide information and awareness to people about precautionary measures. Another example is the photographs uploaded to Flickr to correlate with physical parameters that typify natural disasters.

Building on a similar concept, other studies confirm the association between the spatiotemporal distribution of tweets and the physical scope of floods [32]. In addition, the relation between the occurrence of disaster-related tweets and the distribution of Hurricane Sandy damage projected from simulation models was determined [33]. Another strand focuses on the use of social media as a communication tool in the different phases of the construction project. Leung et al. [34] suggested that social media sites, such as Facebook, can be used to successfully engage more public in construction development projects. In a similar topic, Russell et al. [35] reinvigorated construction companies to take advantage of social media, including Facebook, Twitter, Pinterest, and LinkedIn to present and share what they do in their projects to engage more public and how they do it and to shape relationships with current and potential clients. However, very few studies focus on the application of social media and in particular Twitter to study energy-related topics in any form and shape. As an example, Miles et al. [36,37] analyzed the post-event power outage in San Diego County, California in 2011 in which its residents were without electricity for up to 12 h. They used different data collection tools including interview transcripts, content from news and social media, as well as government documents and databases. They determined that social media is an effective communication tool during power outages to broadcast and find information. Tanielian [38] found that social media can be used to encourage fundamental change in energy-consumption behaviors, exchange ideas, and broadcast theories.

Studying prior studies showed the importance of renewable energy production and the usage of such energy resources could substantially contribute to positive environmental impacts and economic growth worldwide [39]. The main reason behind the importance of renewable energy resources is the environmentally benevolent nature of their production and consumption. Moreover, renewable energy resources, as a natural endowment, can be considered as a more sustainable source to build a stable industrial and environmental development. There are several studies that have found the decisive importance of energy consumption in the economic growth of the G7 (Group of Seven, including the United States, Canada, Germany, France, Italy, and the United Kingdom) countries [40–42]. Behera and Mishra [43] studied the relationship between non-renewable and renewable energy consumption and economic growth in these countries from 1990 to 2015. The results showed that renewable energy had a statically remarkable impact on economic growth among G7 countries.

The above-mentioned studies have considered the economic aspect while ignoring the collective opinion and crowd preference over energy. However, people's opinions (views) on energy sources can also be considered to find out what is the collective opinion of people in a particular region about various types of energy resources. Qazi et al. [44] have reviewed over 300 papers to find out the importance of public opinions towards renewable energy sources and technologies. The obtained results in Ref. [44] showed that global energy issues could be managed by adding renewable energy sources to the current power generation. In addition, Qazi et al. [44] have stated that public opinion and analysis of public tweets play an essential role in facilitating the development of different renewable energy technologies.

Reviewing the body of literature on opinion mining, it was found that sentiment analysis (opinion mining) using various machine learning algorithms could be a very efficient approach to find the crowd's opinion/preference regarding renewable energy resources. For example, Jain and Jain [45] found that solar, wind, and hydropower, bioenergy as well as geothermal energy words were used more often than other words in tweets. Due to the importance of energy, Gupta et al. [46] studied Twitter messages among competing advocacy groups about nuclear energy policy in the United States from January 2014 to May 2014. The obtained results validated the use of Twitter data for more work on the NPF (Narrative Policy Framework). Ikoro et al. [47] investigated the public opinion posted on Twitter by the United Kingdom (UK) energy consumers. They combined functions from two sentiment lexica to improve the accuracy of the classification results compared to using only one lexicon.

On the basis of our review, it can be inferred that public perceptions and attitudes about energy-related topics at an individual level have not been sufficiently studied or analyzed and consequently vital nuances missed. To bridge this gap, we analyze the application of social media to understand the general public's perception and opinion about various energy sources and the choices they make at the individual level. The other challenge is people's perception changes over time due to factors such as natural disasters, economic condition, media coverage, their knowledge, etc. in which we need to perform a new survey to capture their attitudes and beliefs that are very expensive and time-consuming. In this study, we use Twitter data to show how Alaskans' perception of energy-related topics changes over time from 2014 to 2017.

## 3. Materials and methods

The research methodology we applied in the current study is based on sentiment analysis and big data. Specifically, in order to analyze millions of tweets published by Alaskan users, we employed a lexicon-based sentiment analysis approach [48]. This approach enables the fast and scalable analysis of vast volumes of tweets. Moreover, it may be applied to both online and offline sentiment analysis of tweets, making the approach usable in applications that need to monitor public opinion towards different phenomena. This approach will be described in more detail in the following subsections but first, a precise specification of the data analyzed in the current study has been discussed.

### 3.1. Proposed methodology

Suppose we have a collection of tweets $C = \{c_1, c_2, \ldots, c_{|i|}\}$ where each tweet $c_i = (W_{c_i}, u_{c_i}, g_{c_i}, t_{c_i})$ consists of a sequence of words $W_{c_i} = \{w_1, w_2, \ldots, w_{|c_i|}\}$, the user $u_{c_i}$who sent $c_i$, and the time $t_{c_i}$ on which $c_i$ is sent. The task is to assign a sentiment label $l_i \in \{l_1, l_2, \ldots, l_{|L|}\}$ to every $c_i$where, in the current study, $L$ is the set of all emotion classes (see 3.2). To perform the task, the following steps are applied in the proposed model.

- Filter out $C^{'} = \{c_i | g_{c_i} \neq \text{"}Alaska\text{"}\}$ from $C$.
- For each $c_i \in C$:
- Remove irrelevant and useless characters.
- Remove links from $c_i$.
- For each $w_i \in W_{c_i}$:
  - $ps = polarity\ scores(w_i)$
  - $ps_{w_i} = avg(ps)$
    - $ps_{c_i}$ = polarity score of $c_i$ using Eq. (25).
    - $l_i = \underset{x}{\text{argmax}}(ps_{c_i}(x) | x \in \{HP, LP, LN, HN\})$

The overall structure of the proposed methodology is depicted in Fig. 1. After crawling relevant tweets, a geo-filtering is applied to filter out tweets concerning other regions rather than Alaska. This step is necessary since the focus of all three research questions is on Alaska.

The next phase in the proposed system is pre-processing which is performed with two goals: 1) decreasing the size of tweets by removing irrelevant and useless characters, and 2) increasing the accuracy of the matching method by removing probable sentiment-bearing words in links. To achieve these goals, a Natural Language Processing (NLP) approach based on regular expressions is employed [49]. In this approach, as illustrated in Fig. 2, regular expressions are used to find irrelevant parts of tweets. A regular expression is a common technique for specifying and matching patterns in NLP-related tasks [49] and the designed regular expressions in Python programming language were used in this study. The output of the pre-processing module is then a Twitter dataset containing user, time, and textual body of tweets.

The next module in the proposed system, parser, is employed to extract two tweet-related information, namely tweet's text and tweet's creation time, and two user-related information, namely @user messages and user IDs. @user messages are those messages that are directed at, or replied to, other users [50]. Text-related information is analyzed in the sentiment analysis module and user-related information is used in the user analysis module for further processing.

In order to analyze the opinions of Alaskan in their tweets, a lexicon-based approach [51] is used in the current study. The core of each lexicon-based method is the lexicon it employs to assign sentiment intensity scores to words [48]. Following the approach proposed in Ref. [52], in the current study we used a subjectivity lexicon of English adjectives (hereafter called ADJLex) that has been enriched with some sentiment-bearing nouns and verbs. ADJLex is used in the TextBlob [53] that is a python library for natural language processing tasks. In this lexicon, adjectives have polarity scores in the range of $[-1, +1]$ and a subjectivity score in the range of $[0, +1]$. The reason for using ADJLex in the current study is that this lexicon, containing about 2900 terms, has high coverage of sentiment-bearing words.

Words in ADJLex are tagged according to their WordNet [54] sense and this caused to have more than one polarity and subjectivity score for some words. In order to take this into consideration, an aggregation step is needed to combine scores for each word [55]. To this aim, arithmetic averaging is used in the current study. As an example, consider the word "great" that has four polarity scores: 1.0, 1.0, 0.4, and 0.8. If this word appears in a tweet, its probability score would be 0.8 that is the average of the four above-mentioned values.

When a lexicon-based approach is employed for sentiment analysis, the role of negation and intensifiers must be considered explicitly [51]. In order to apply the effect of such modifiers, we considered three possible situations as follows:

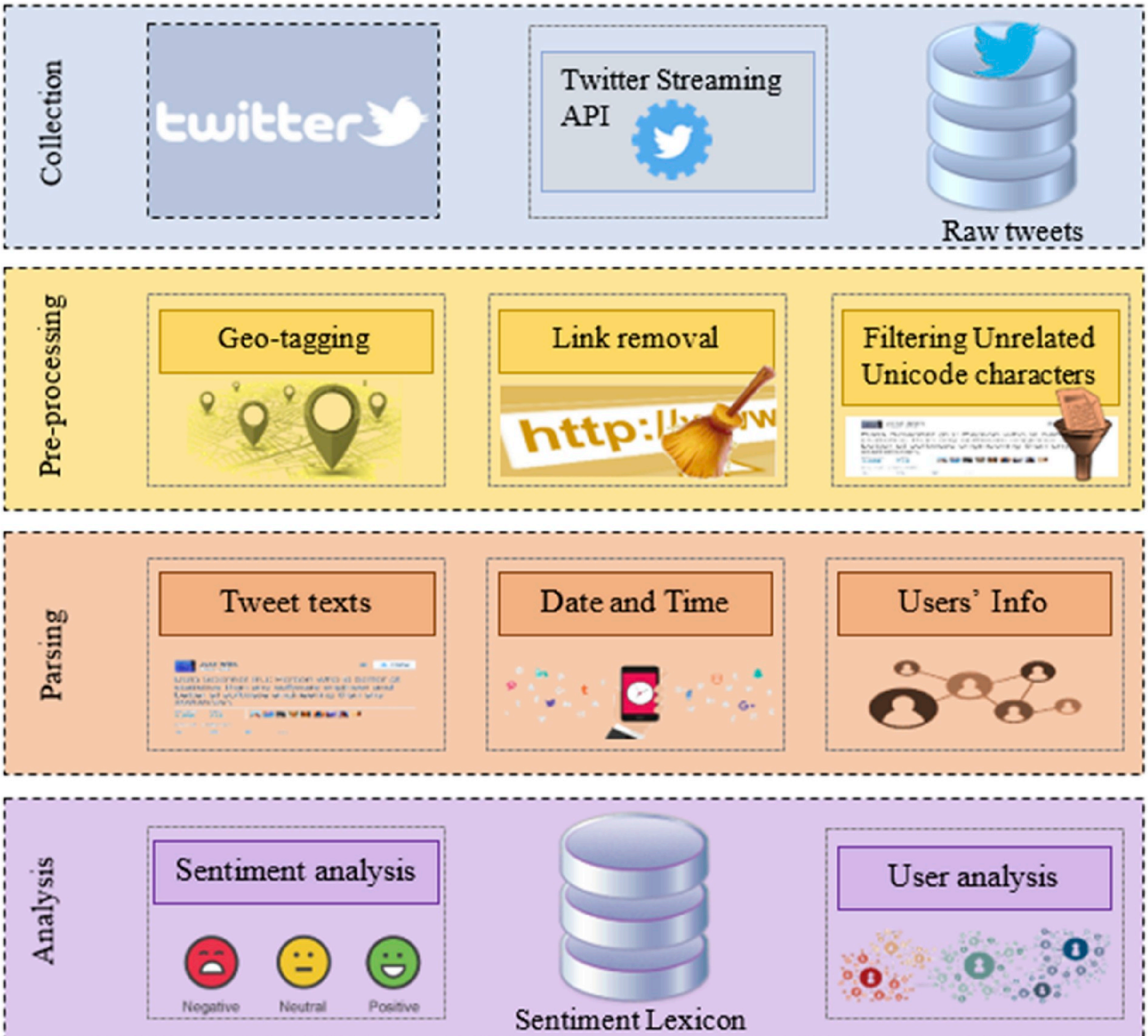



**Fig. 1.** The overall structure of the proposed system.

1) Having only negations: in some cases, there is only a negation word such as "not" before a word and no intensifier follows it. In such cases, we multiplied the polarity of the word by $-0.5$.
2) Having only intensifiers: in some cases, there is only an intensifier adverb such as "very" before a word and no negation word precedes it. In this case, we multiplied the polarity and subjectivity of the word by the intensity of the intensifier (for example, 1.3 for the word "very").
3) Having both negations and intensifiers: in this case, for example, when a phrase such as "not very good" exists, the multiplication by $-0.5$ is applied to preserve the effect of negation. Moreover, the inverse intensity of the modifier is multiplied to maintain the effect of the intensifier. For example, consider the phrase "not very good" in which the polarity of "good" is $+0.7$. The final polarity will be: $-0.5 \times \frac{1}{1.3} \times 0.7 = -0.27$.

### 3.2. Mapping polarity scores to class labels

In order to show the opinion and attitude of people towards energy and renewable energy, we employed two–dimensional valence–arousal (VA) space which is a standard dimensional model of emotion [56]. In this model, valence shows positive versus negative attitude and arousal represents low versus a high level of activation. According to this model, different feelings are located in different areas on VA space as shown in Fig. 3 (adopted from Ref. [57]).

Based on the polarity scores calculated in the previous step, four class labels are defined in the current study: HP, LP, LN, and HN corresponding to high positive, low positive, low negative, and high negative, respectively. These class labels are shown in four regions of VA space in Fig. 3.

As mentioned in the previous subsection, polarity scores are represented in the range of $[-1, +1]$. To map this interval into four class labels, we divided it into four equal parts as follows: $[-1, -0.5]$ for HN, $(-0.5, 0]$ for LN, $(0, 0.5]$ for LP, and $(0.5, +1]$ for HP.

After calculating the polarity of phrases and assigning class labels to them, the next step is the fusion of polarity scores of all sentiment-bearing phrases of the tweet. Different strategies are proposed for score aggregation in the literature [58]. Although simple fusion methods such as arithmetic mean, majority voting, sum, product, and maximum [59] are widely used in different applications, it has been shown that more formal fusion methods, such as Dempster-Shafer (DS) theory of evidence, outperforms simple heuristic methods [60]. DS theory is an evidential theory for addressing the problem of uncertainty and can be seen as a generalization of Bayesian fusion rule [57]. This theory was first proposed by Dempster and later developed by Shafer [61]. DS-based fusion method has been used for aggregating sentence-level

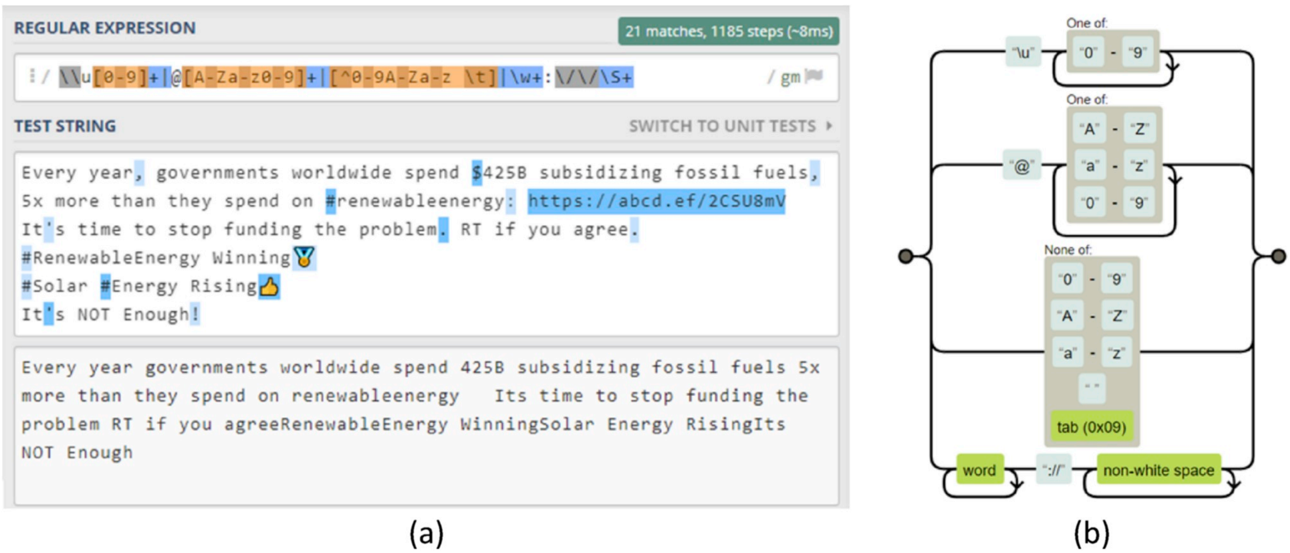



**Fig. 2.** The regular expression for pre-processing of tweets and a sample tweet on which the regular expression is applied (a), and the railroad diagram of the regular expression (b). The subfigure shown in (a) was generated using an online tool in https://regex101.com and the subfigure shown in (b) was generated using a similar online tool in https://regexper.com/" \h https://regexper.com.

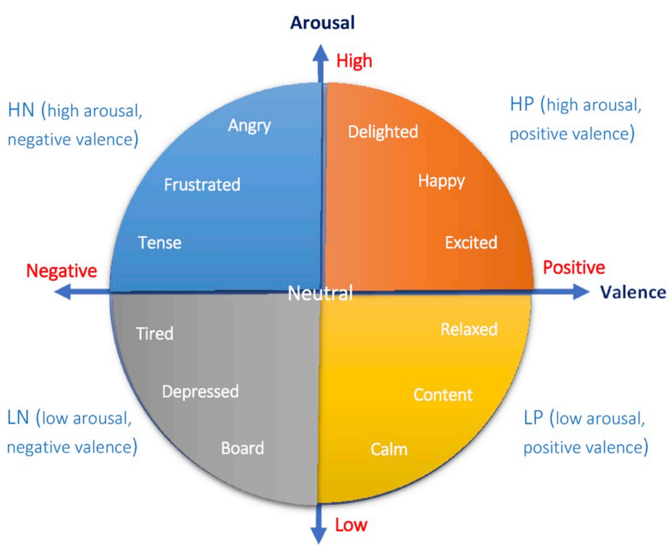



**Fig. 3.** Distribution of different fillings in the two–dimensional VA space (adopted from Ref. [55]).

sentiment scores into a document-level score [59], but it has not been previously used for aggregation of opinion expressed in tweets. This theory is exploited in the current study to fuse phrase-level sentiment polarity scores into tweet-level scores. The details of the proposed method are presented in the next subsection.

### 3.3. Proposed evidential fusion method

In order to apply the DS theory for the fusion of sentiment polarity scores, the first step is to define the frame of discernment, $\theta$, which is a mutually exclusive set of hypotheses for determining the scope of the problem as follows:

$$\theta = \{\theta_1, \theta_2, \cdots, \theta_n\} \tag{1}$$

It is evident that the power set of $\theta$ which is denoted as $2^\theta$ has $2^n$ elements, each of which showing a possible subset of $\theta$. In the current study, we considered the inclusion of each sentiment-bearing phrase in one of the possible categories as a hypothesis.

The next step is to define the mass function, $m(A)$, for assigning a probability to each evidence supporting a subset of $\theta$, such as $A \subseteq \theta$. In the current study, we are interested in particular subsets of $\theta$, namely HP, LP, LN, and HN as defined in the previous subsection. It should be noted that the mass function must be a basic probability assignment (BPA) having the following properties:

$$m : P(X) \rightarrow [0, 1] \tag{2}$$

$$m(\varnothing) = 0 \tag{3}$$

$$\sum_{A \in 2^\theta} m(A) = 1 \tag{4}$$

In order to define the mass function, we followed a fuzzy-based approach [62,63]. Explicitly, we first defined fuzzy membership functions as depicted in Fig. 4. As shown in this figure, the Gaussian membership function is selected. This function has several interesting features including being smooth and natural and having non-zero values at all points.

The Gaussian membership function is defined as follows:

$$Gaussian(x,\ c, \sigma) = e^{-\frac{1}{2}\left(\frac{x-c}{\sigma}\right)^2} \tag{5}$$

Where $c$ and $\sigma$ are two adjustable parameters for specifying the center

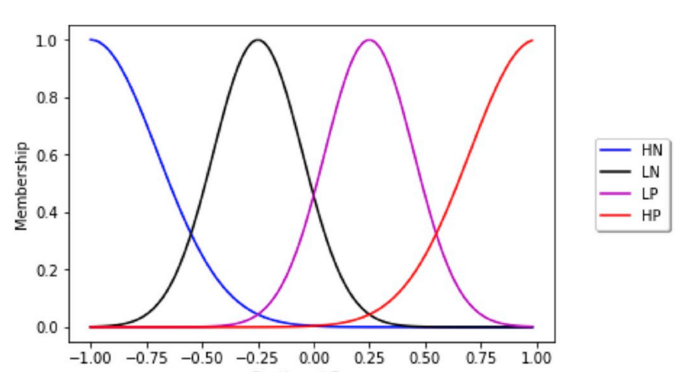



**Fig. 4.** Membership function of sentiment score with four linguistic terms.

and the width of the function.

Using the membership function defined in the Equation (5), four probabilities are generated for each sentiment-bearing term in a tweet. In order to satisfy the condition of Equation (4), the sum of these four probabilities for each term must be one. To this aim, the normalized form of the membership function is used.

The final step in the proposed fusion method is to fuse the mass functions of sentiment-bearing terms for each tweet. The DS combination rule, also called orthogonal sum, may be used for aggregating to mass function as follows:

$$(m_1 \oplus m_2)(A) = \begin{cases} \dfrac{\sum_{X\cap Y=A} m_1(X)m_2(Y)}{1-K_{12}} & A \neq \varnothing \\ 0 & A = \varnothing \end{cases} \tag{6}$$

$$K_{12} = \sum_{X\cap Y=\varnothing} m_1(X)m_2(Y) \tag{7}$$

Where relationship $K_{12}$ is a normalizing factor ensuring that $m_1 \oplus m_2$ remains BPA. Due to its commutativity and associativity, the above fusion rule may be applied iteratively when more than two m's should be fused.

The DS-based fusion function has been proposed for aggregating sentence-level sentiment scores into document-level scores [48]. Although this function may also be applied for the fusion of phrase-level sentiment scores into a tweet-level score, it has the problem of considering only one element in the resulted mass function [57]. This problem is more common in phrase-level aggregation than a sentence- and document-level. Therefore, in the current study, we adopted the two-point method proposed by Bi et al. [64] which also considers the second most probable decision in each step of fusion. Specifically, we specify the first and second most probable classes predicted by two mass functions as follows:

$$u = \underset{x}{\mathrm{argmax}}(m_1(x)|x \in \{HP, LP, LN, HN\}) \tag{8}$$

$$v = \underset{x}{\mathrm{argmax}}(m_1(x)|x \in \{HP, LP, LN, HN\} - u) \tag{9}$$

$$w = \underset{x}{\mathrm{argmax}}(m_2(x)|x \in \{HP, LP, LN, HN\}) \tag{10}$$

$$z = \underset{x}{\mathrm{argmax}}(m_2(x)|x \in \{HP, LP, LN, HN\} - w) \tag{11}$$

In the two-point method, in each step of the fusion, only two most probable classes are considered. Therefore, the mass function is changed as follows:

$$m^{\sigma}(u) + m^{\sigma}(v) + m^{\sigma}(c) = 1 \tag{12}$$

In other words, $m^{\sigma}$ is a triplet mass function. Based on the values of $u$, $v$, $w$, and $z$, there are three possible cases as follows:

1) Two equal pairs: this case happens when $u = w$ and $v = z$, or $u = z$ and $v = w$.
2) Only one equal pair: this case happens in one of four situations; $u = w$ and $v \neq z$ or $u = z$ and $v \neq w$ or $v = w$ and $u \neq z$ or $v = z$ and $u \neq w$.
3) No equal pairs: this happens only when $u \neq w \neq v \neq z$.

The combination rule shown in Equations (6) and (7) should also be changed according to the three cases above. In the first case, the following four equations are used for the fusion of two mass functions.

$$(m_1 \oplus m_2)(u) = K[m_1(u)m_2(u) + m_1(u)m_2(c) + m_1(c)m_2(u)], \tag{13}$$

$$(m_1 \oplus m_2)(v) = K[m_1(v)m_2(v) + m_1(v)m_2(c) + m_1(c)m_2(v)], \tag{14}$$

$$K^{-1} = 1 - [m_1(u)m_2(v) + m_1(v)m_2(u)]. \tag{15}$$

In the second case, suppose that the equal pair is denoted by $u$, then the following four equations are used for the fusion.

$$(m_1 \oplus m_2)(u) = K[m_1(u)m_2(u) + m_1(u)m_2(c) + m_1(c)m_2(u)], \tag{16}$$

$$(m_1 \oplus m_2)(v) = Km_1(v)m_2(c), \tag{17}$$

$$(m_1 \oplus m_2)(w) = Km_1(c)m_2(w), \tag{18}$$

$$K^{-1} = 1 - [m_1(u)m_2(w) + m_1(v)m_2(w) + m_1(v)m_2(u)]. \tag{19}$$

Finally, in the last case, the following five equations are used for the fusion.

$$(m_1 \oplus m_2)(u) = Km_1(u)m_2(c), \tag{20}$$

$$(m_1 \oplus m_2)(v) = Km_1(v)m_2(c), \tag{21}$$

$$(m_1 \oplus m_2)(w) = Km_1(c)m_2(w), \tag{22}$$

$$(m_1 \oplus m_2)(z) = Km_1(c)m_2(z), \tag{23}$$

$$K^{-1} = 1 - [m_1(u)m_2(w) + m_1(u)m_2(z) + m_1(v)m_2(w) + m_1(v)m_2(z)]. \tag{24}$$

In all cases, the fusion of m's for c is obtained using the following equation.

$$(m_1 \oplus m_2)(c) = Km_1(c)m_2(c). \tag{25}$$

After the fusion step, the class with the highest probability is selected as the label of the tweet.

## 4. Results and discussions

In this section, the results of implementing the proposed system to answer the aforementioned research questions are presented.

At first, we defined 260 words that were deemed relevant to the subject of energy. Then, the word cloud of keywords, which illustrates the most frequently appearing keywords for three years, is illustrated in Fig. 5. Word clouds have been used as useful tool in illustrating textual content, where the font size of a keyword could indicate its frequency in the text. As can be seen in this figure, the words with bigger sizes suggest that they have been used more often than the words with smaller sizes. For example, "power", "oil", "dam", "sun", "gas", and "heat" are more repeated words in 2014. According to the word cloud of keywords, "power", "oil", and "dam" are three more repeated words in 2014, whereas, "fuel" and "dam" in 2015 and "sun", "volcano", "power" and "wind" in 2016 are repeated more than other words. These results show the importance of energy-related discussions among Alaskans on Twitter.

As an example, "volcano" was one of the most repeated words in 2016. Due to Alaska's location, a volcanic arc spinning the Pacific Ocean, several opportunities exist for geothermal energy development and investment in the state. Alaska contains more than 130 active volcanoes and volcanic fields in the last two million years, and approximately 50 of them have been active within historical time. Moreover, more than 100 sites with thermal springs and wells have been recognized across the state. All these resources make Alaska as a pioneer state in generating electricity from geothermal resources (one of nine states). However, a vital issue is most of these resources are located in isolated areas, which are far from a population center that would use the electricity generated.

High energy prices distressing many rural areas and the lack of proper energy infrastructure entice Alaska to become a front-runner in the expansion of renewable energy resources and help local communities to produce stably-priced, environmentally responsible energy. Alaska has significant potential for renewable energy production. According to the Electric Power Research Institute [Ref 41], more than 50% of the nation's wave energy resources and more than 90% of the

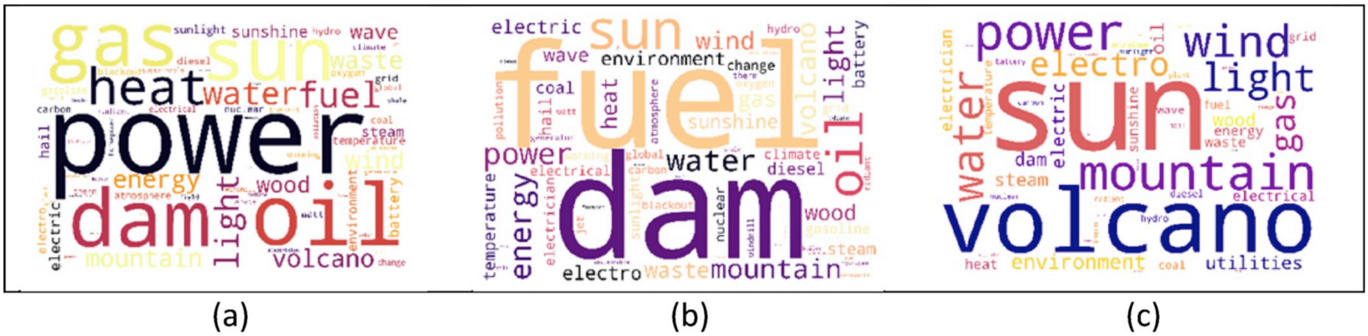



**Fig. 5.** The word cloud of the year (a) 2014, (b) 2015, and (c) 2016 showing the frequency of energy-related keywords in Alaskan tweets.

nation's river current and tidal energy resources are in Alaska. Moreover, Alaska has some of the best wind resources in the United States, a large number of volcanoes and hot springs, significant potential for bioenergy and solar energy.

To investigate the level of discussions about renewable energy sources in Alaska on Twitter, a more detailed analysis was performed to identify the most frequently keywords related to renewable energy sources. For this purpose, 83 renewable energy-related words were chosen, and their word cloud is illustrated in Fig. 6. It can be observed that "dam", "sun", and "wind" are more repeated words in these three years, while "sun" is repeated in all three years and "dam" is more repeated in 2014 and 2015. These keywords are aligned with existing renewable sources in Alaska.

The potential for using solar energy in Alaska has long agonized from the belief that the sun basically does not offer any hope for Alaskans while there are 230 h more of possible sunlight at the Arctic Circle than at the equator. In addition, the price to harvest solar energy has not only decreased dramatically over the last decade but also solar photovoltaic (PV) has minimal ongoing maintenance costs compared to wind turbines. This explains why Alaskans have considered solar energy as a potential source of energy in recent years. We also found that there are more discussions about solar energy in 2016 compared to previous years. This can be explained by the report published by the Department of Energy in February 2016 emphasizing the unique advantages in solar energy despite the northern latitude[1]. These include long daylight hours in the summer and "low ambient temperatures that improve the efficiency of solar modules and the reflectivity of sunlight off of snow cover on the ground." The same report also showed that for many communities, solar power would be less expensive than diesel fuel due to the declining cost of PV cells used to produce electricity. The other interesting aspect of this report highlighted that the solar resources in some regions of Alaska are at least comparable to that of Germany, which is leading the world in PV installations with more than 38,500 MW (MW) of solar installed as of October 2015. We believe this report brought more discussions in the potential of solar power generation in Alaska in 2016.

More recently, rural Alaskans found that wind turbines could produce power at a cheaper rate than diesel generators. Wind energy has been the primary goal of public investment in Alaska renewables, containing the largest share of grants (35%) under the Renewable Energy Fund because there are plentiful wind resources in Alaska, mostly along the coastal regions of the state. In Alaska, wind power is a very auspicious resource to generate power in both small and large scales. As of mid-2012, there were about 30 wind installations in Alaska, and all but three are in rural communities outside the Railbelt, the region extending from the Kenai Peninsula to Fairbanks and a similar number in the permitting process or under construction.

The other two most repeated words were dam and water. Alaska has a robust record of developing successful hydroelectric projects that deliver clean, reliable energy across the state. Alaska produces about a third of its power from hydroelectric dams in the Southeast, Southcentral and Southwest parts of the state and hydroelectric power is Alaska's largest source of renewable energy, delivering 21% of the state's electrical energy in an average water year. This explains why Alaskan's used these words in their tweets and shows the importance of it in their daily life.

Research question 1: What is the perception of Alaskans about various energy sources?

To answer this question, the four-class systems described in 3.3 was used to categorize the perception of Alaskans into four main classes according to the polarity and intensity of the sentiment expressed in their tweets in 2014, 2015, and 2016, respectively. The details can be observed in Fig. 7.

As can be seen in this figure, in 2014, about one-half (47%) of the tweets are positive, less than 20% of them are harmful, and the remaining are neutral. However, in 2015 and 2016, a significant increase is seen in positive sentiment (Fig. 8). In order to track the Alaskans' sentiment towards energy, the frequency and change rate of positive, negative, and neutral tweets in three consecutive years are compared in Table 1.

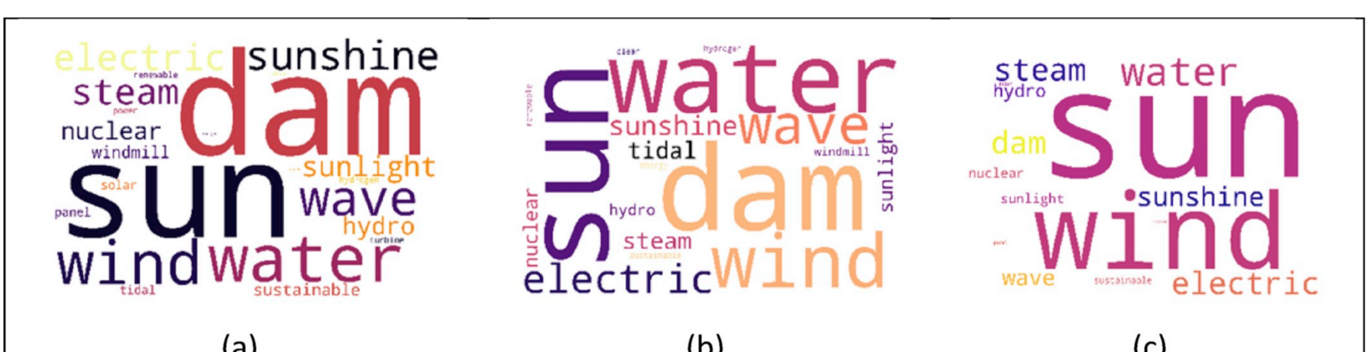
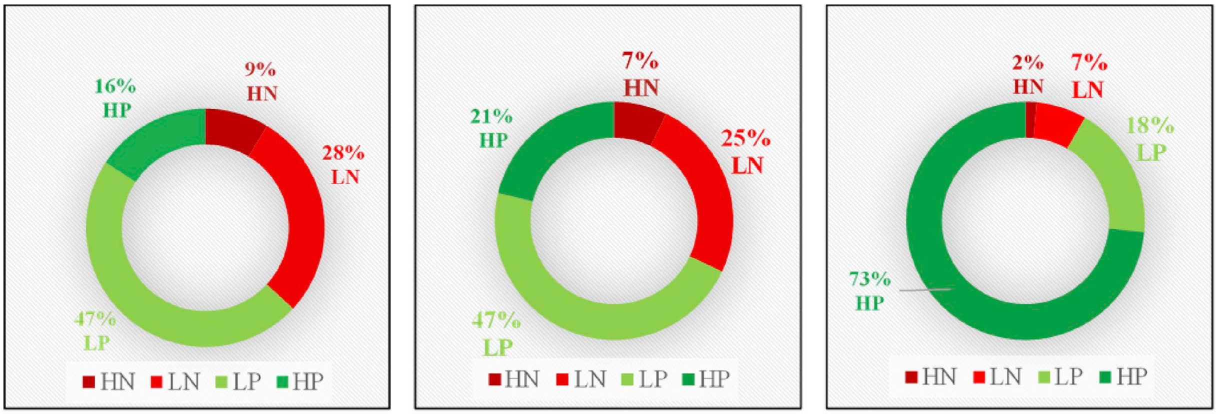
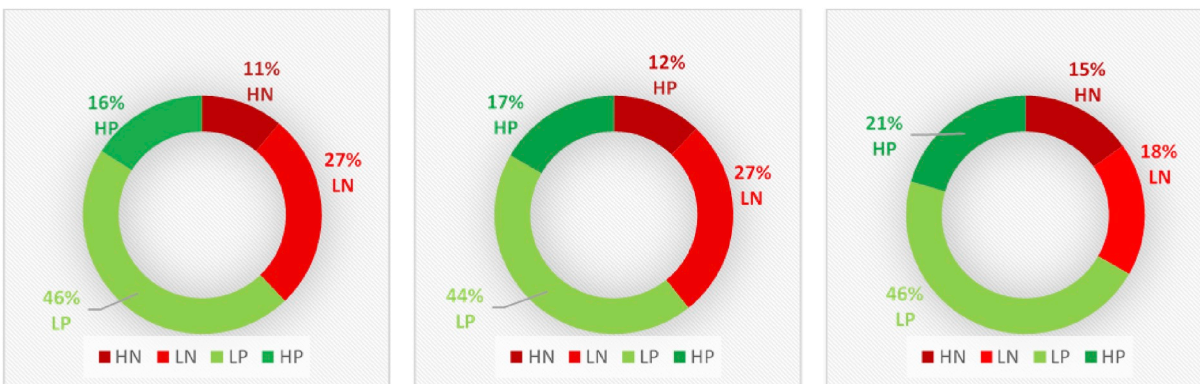



**Fig. 6.** The word cloud of the year (a) 2014, (b) 2015, and (c) 2016 showing the frequency of renewable energy-related keywords in Alaskan tweets.




(a) (b) (c)

**Fig. 7.** Distribution of four different energy-related sentiments in 2014 (a), 2015 (b), and 2016 (c).




(a) (b) (c)

**Fig. 8.** Distribution of five different renewable energy-related sentiments in 2014 (a), 2015 (b), and 2016 (c).

**Table 1**
Comparison of the frequency and change rate of positive, negative, and neutral sentiments towards energy in Alaska in 2014, 2015, and 2016.

| | Positive | | Negative | |
|---|---|---|---|---|
| | Frequency | Change | Frequency | Change |
| 2014 | 63% | 0 | 37% | 0 |
| 2015 | 68% | 7% | 32% | −16% |
| 2016 | 91% | 33% | 9% | −255% |

**Table 2**
Comparison of the frequency and change rate of positive, negative, and neutral sentiments towards renewable energy in Alaska in 2014, 2015, and 2016.

| | Positive | | Negative | |
|---|---|---|---|---|
| | Frequency | Change | Frequency | Change |
| 2014 | 52% | 0 | 38% | 0 |
| 2015 | 61% | 17% | 39% | 2% |
| 2016 | 67% | 10% | 33% | −18% |

According to Table 1, the frequency of positive sentiments towards energy increased in two successive years while the negative sentiments decreased at the same time. This reflects an increase in the positive opinion of Alaskans about energy over time. It is worth to notice that in Table 1, the change rate of negative sentiment is significantly higher than that of positive sentiment, showing a significant change in Alaskans' perceptions of energy.

According to Figs. 4–6, it can be seen that LP is the most sentiment followed by LN in 2014, 2015, and 2016 whereas HN sentiment had the lowest value in these three years. An interesting point is that the distribution of HN has fallen to 2% from 2014 to 2016. The same behavior happened to LN sentiment. Based on findings from our extensive data analysis, it can be observed that the distribution of 2-star sentiment in 2014 was 28% while in 2015 and 2016 were 25% and 7%, respectively. This, besides the results shown in Table 2, justifies a significant change in the overall view of Alaskans towards energy.

Research question 2: Do Alaskans use Twitter to make their voice on various sources of energy and renewable energy or lack of proper energy heard?

To respond to this question, the distribution of both sent messages in different months and the number of users who sent messages in different months of three years (2014, 2015, and 2016) are presented in Table 3. Note that each user was counted just once even if he/she had sent more than one tweet in a month. As can be seen in this table, there are different user behaviors across three years. However, there is a similarity between 2014 and 2016 regarding the distribution of both energy-related sent messages in different months. In 2014 and 2016, more energy-related sent messages happened in February, March, June, July, and August. Based on the outcomes obtained, the distribution rates for February, March, June, July, and August in 2014 are 10%, 10%, 13%,

**Table 3**
Monthly distribution of message and users in three years.

| | Distribution of Messages | | | Distribution of Users | | |
|---|---|---|---|---|---|---|
| | 2014 | 2015 | 2016 | 2014 | 2015 | 2016 |
| Jan | 0.22% | 15.82% | 8.16% | 0.58% | 15.58% | 6.58% |
| Feb | 10.77% | 11.36% | 6.19% | 9.89% | 12.03% | 5.11% |
| Mar | 11.53% | 19.52% | 7.85% | 10.11% | 17.91% | 6.9% |
| Apr | 7.9% | 17.03% | 8.43% | 8.1% | 16.52% | 8.02% |
| May | 9.85% | 4.96% | 9.74% | 9.55% | 6% | 9.26% |
| Jun | 13.04% | 4.93% | 5.1% | 12.78% | 6.67% | 7.74% |
| Jul | 12.6% | 0% | 14.57% | 12.73% | 0% | 15% |
| Aug | 11.78% | 5.69% | 10.47% | 11.44% | 6.42% | 12.37% |
| Sep | 5.68% | 6.05% | 10.2% | 6.62% | 6.15% | 9.86% |
| Oct | 5.15% | 4.74% | 9.03% | 5.8% | 4.37% | 8.26% |
| Nov | 4.38% | 4% | 6.92% | 5.51% | 4.15% | 6.82% |
| Dec | 7.12% | 5.91% | 3.34% | 6.89% | 4.2% | 4.07% |

13%, and 11% while in 2016 are 12%, 15%, 8%, and 15%, respectively. In all three years, the number of energy-related text messages was high in July. The distribution of energy-related sent messages in 2015 is significantly different from 2014 to 2016. As indicated in distribution of messages in 2015, most of the messages were sent in the first four months of the year, January 16%, February 11%, March 19%, and April %17 which is significantly higher than similar months in other two years. The same trend can be seen for the distribution of the number of users who sent energy-related tweets in 2015, January 20%, February 15%, March 23%, and April 21% which is noticeably higher than similar months in the other two years. This can be explained by an exceptionally warm winter of 2015 in Alaska as Alaskans experienced record-setting warmth during the cold season.

The top 3 distribution rates of renewable energy-related tweets are as follows:

1) 2014: June (13%), July (13%), and August (11%),
2) 2015: January (16%), March (18%), and April (17%),
3) 2016: July (15%), August (12%), and September (10%).

These results show that Alaskans use Twitter as a communication tool to talk about various energy sources ensuring their voice is heard.

Research question 3: How do the online public sentiments about various energy sources and the application of renewable energy in Alaska change over time?

To answer this question, the trend of changing four different sentiments (HP, LP, LN, and HN) of users towards energy in 2014, 2015 and 2016 is illustrated in Figs. 9–11, respectively.

As can be seen, there is a remarkable difference between the number of tweets in 2014, 2015 and 2016. In 2016, most numbers of tweets were related to HP, however, fewer tweets were LP, LN, and HN. As indicated in Figs. 9–11, the total number of tweets in 2014 is higher than the number of tweets in 2015 and 2016. This is due to changes in Twitter.com for enabling geo-tagged tweets from the second half of the year 2015 which resulted in a larger number of geo-tagged tweets from the year 2014 or earlier.

The trend of changing four different renewable energy-related sentiments (HP, LP, LN, and HN) in 2014, 2015 and 2016 is illustrated in Figs. 12–14, respectively.

According to Figs. 12–14, there are different behaviors in these three years. A remarkable point is that users showed more positive (either low or high) sentiments towards renewable energy. Fig. 12 illustrates the low negative (LN) sentiments among people compared to other sentiments (HP, HN, and LN). As Fig. 13 shows, users had high negative (HN) and low negative (LN) in the first half of 2015 of renewable energy-related sentiments while in the second half of 2015 they had higher positive (HP) and low positive (LP) sentiments. However, it is evident that individuals indicated HP and LP sentiments while they had less HN and LN sentiments in 2016.

In the following, more details on sentiment analysis of tweets about keywords chosen are discussed. In this regard, the top-20 energy-related keywords and the top-20 renewable energy-related keywords were identified and their frequency rank and amount of growth were compared in 2014, 2015, and 2016 as shown in Figs. 15–18. As can be seen in Fig. 15, “energy waste” had the highest rank in both 2014 and 2015 while the words “heat” and “steam” had the highest rank in 2016. The word “sun” had the lowest rank in all three years.

Fig. 16 compares the amount of growth in the rank of top-20 energy-related keywords in successive years. As can be seen, there was a significant growth rate for word “sun” in both 2016/2015 and 2016/2014 which is about 200%. The growth rate of word “power” was the highest in compare with other words (more than 700%) in 2016/2014.

Fig. 17 compares the rank of the top 20 renewable energy-related keywords for three years. As can be seen “Tidal” has the highest ranking in all three years and “solar panels” take the second rank in all three years. The amount of growth in the rank of top-12 renewable energy-related keywords in successive years is shown in Fig. 18. “Tidal” and “Nuclear” had the highest growth rate in 2015/2014 in compare with other keywords and other years. The amount of growth for “solar panel” in 2016/2015 was significantly high compared to previous years (approximately 300%). In addition, an increasing trend was seen for the word “sun” from 2015/2014 to 2016/2014.

## 5. Discussions

This study aimed to conduct sentiment analysis to find the crowd opinion/preference regarding renewable energy resources and understand Alaskans’ perceptions and opinions about energy efficiency measures, availability of energy sources, and the application of clean energy sources. One of the interesting points in our study was the rising trend in the frequency of positive sentiments among Alaskans which stipulate their preference for renewable energy resources compared to other energy types. Interestingly, the same energy preferences worldwide were

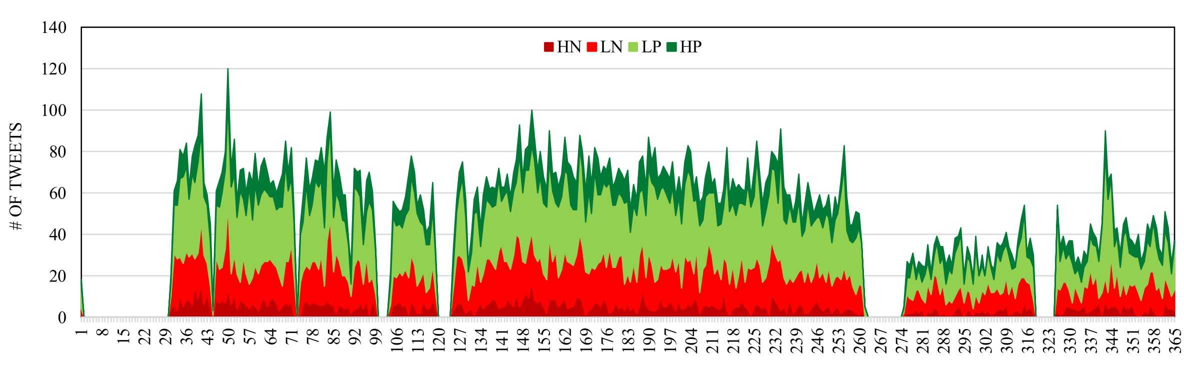



**Fig. 9.** Trend of changing four different sentiments in 2014.

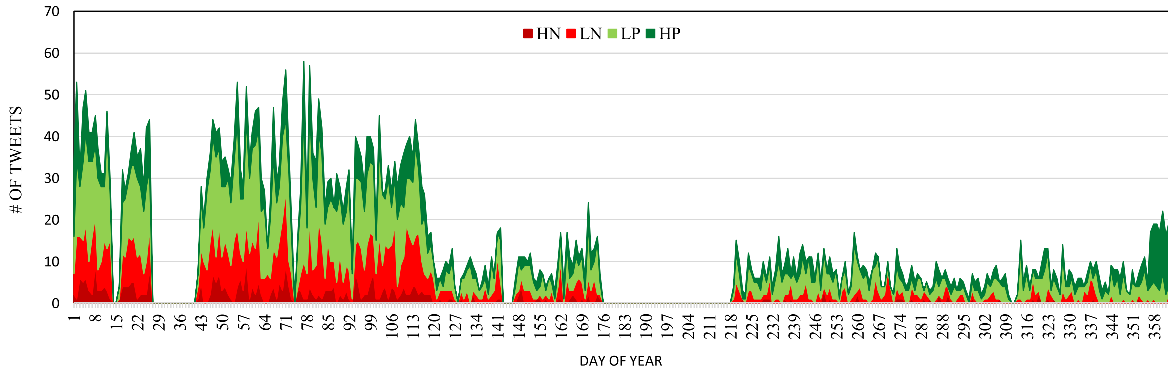



**Fig. 10.** Trend of changing four different sentiments in 2015.

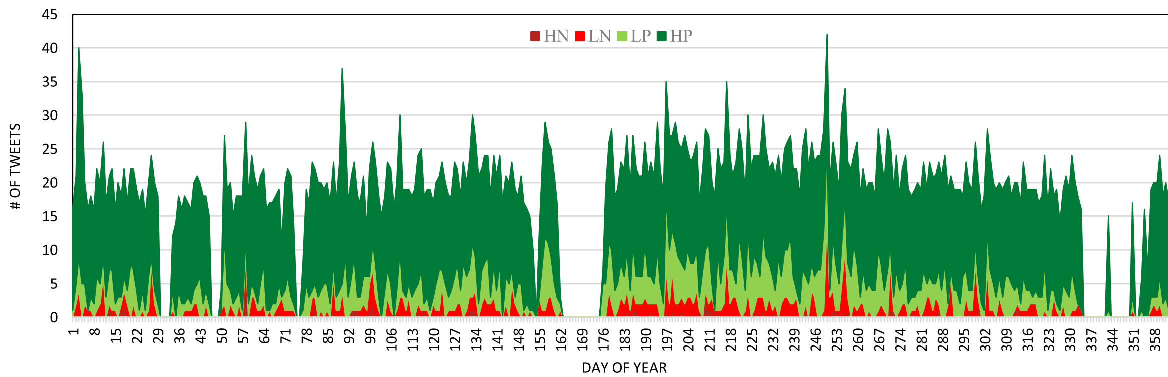



**Fig. 11.** Trend of changing four different sentiments in 2016.

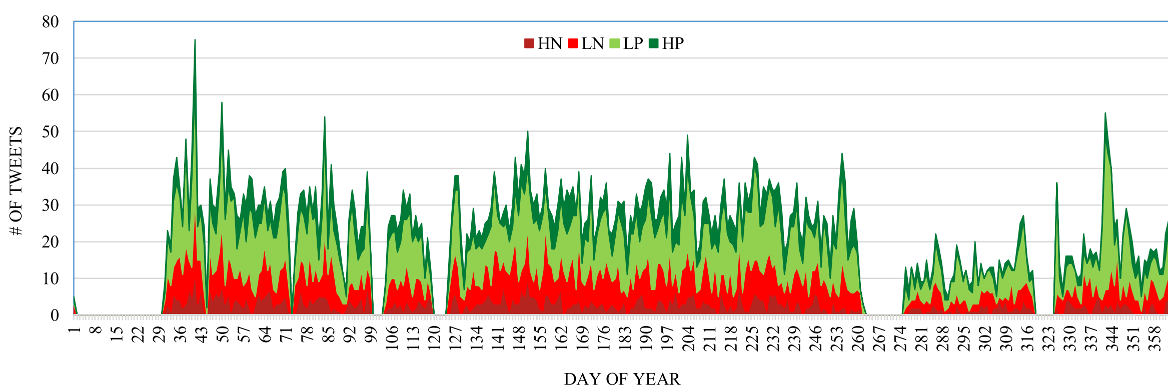



**Fig. 12.** Trend of changing four different renewable energy-related sentiments in 2014.

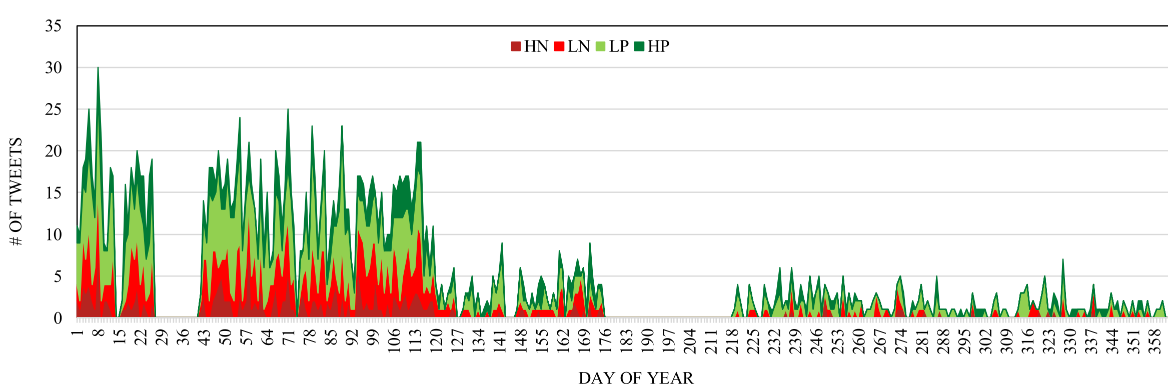



**Fig. 13.** Trend of changing four different renewable energy-related sentiments in 2015.

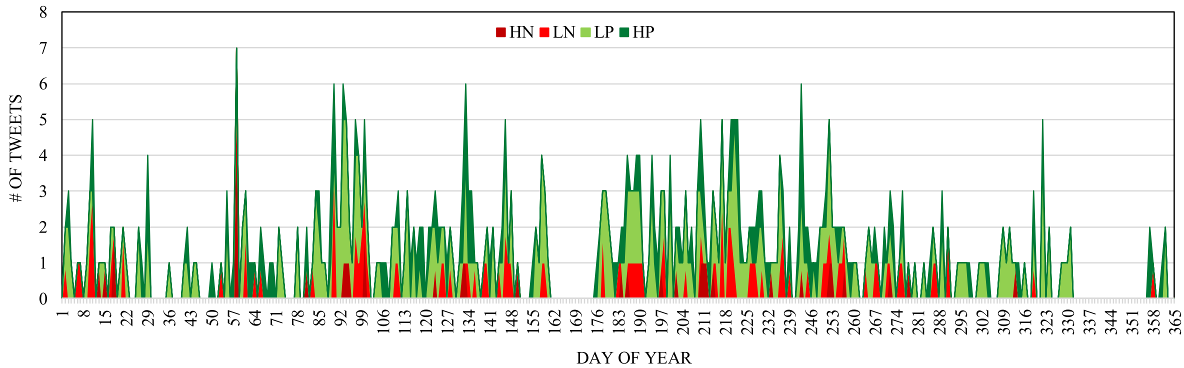



**Fig. 14.** Trend of changing four different renewable energy-related sentiments in 2016.

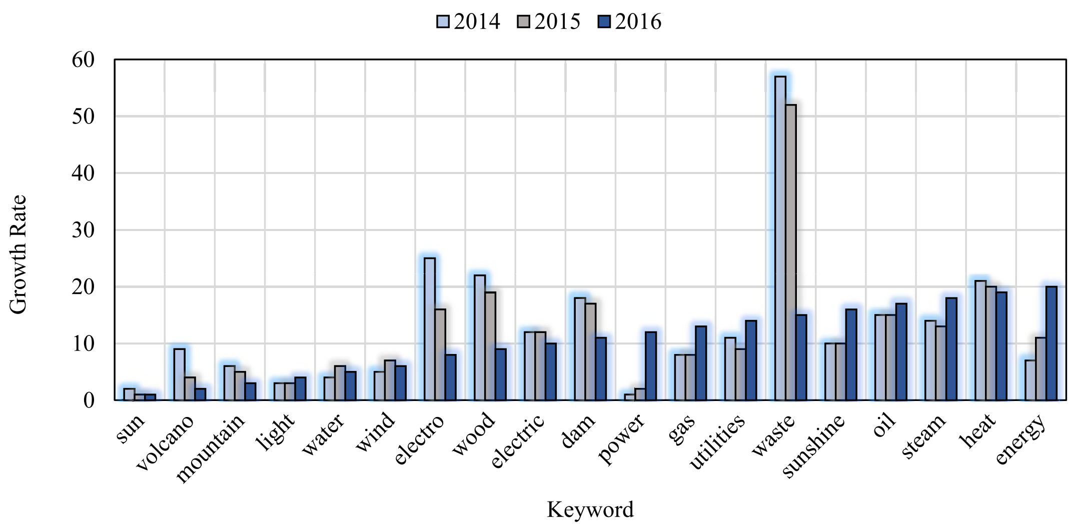



**Fig. 15.** Comparison of the rank of the most used energy-related keywords for three years.

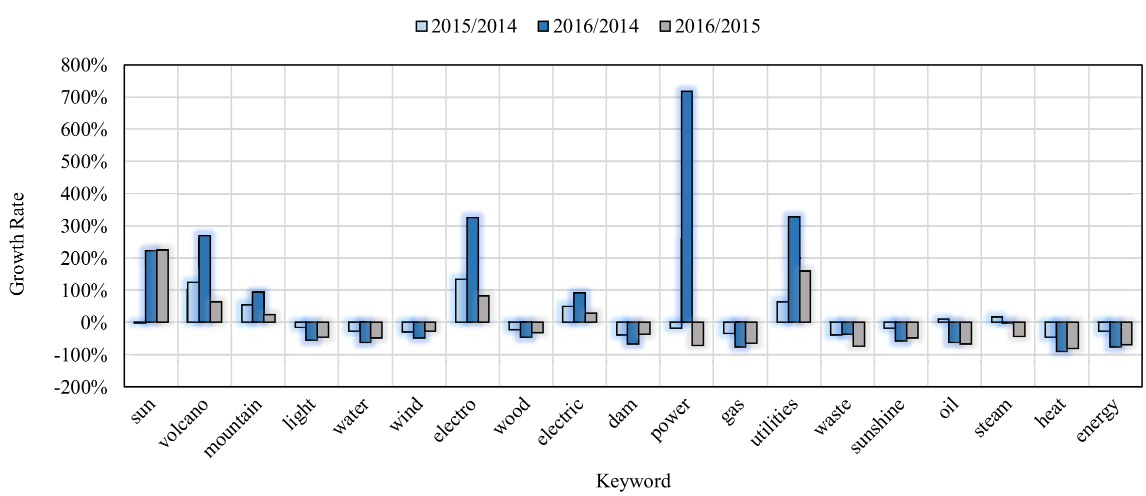



**Fig. 16.** Comparison of the amount of growth in rank of the most used energy-related keywords in successive years.

indicated by Ref. [44] emphasizing on managing the global energy issues by adding renewable energy sources to the current power generation. This study has highlighted the importance of public opinion and analysis of public tweets in facilitating the development of different renewable energy technologies. The general public' opinion can be an essential factor for social acceptance and derived from that aspect, also for political decision making.

In view of the study findings, it can be clearly seen that the Alaskans use Twitter as a communication tool to talk about various energy sources ensuring their voice is heard. Hence, it is reasonable to suggest that twitter can be used as a tool to explore public views and social attitudes towards the use of various energy sources. This resonates with other studies indicating the interest of the public to share their opinion about various topics but in particular energy resources on twitter [13,17,31].

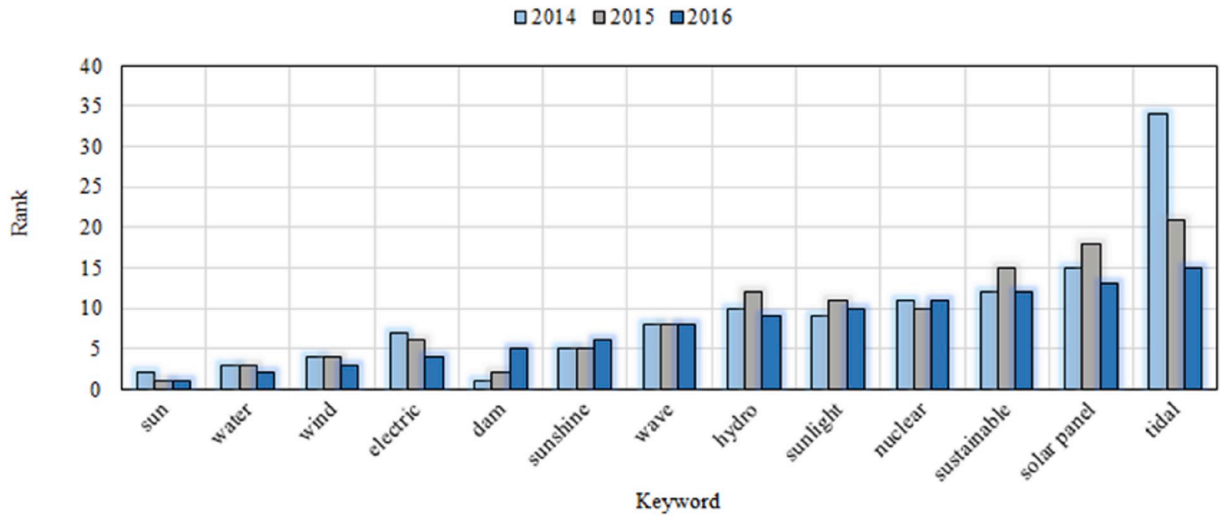



**Fig. 17.** Comparison of the rank of the most used renewable energy-related keywords for three years.

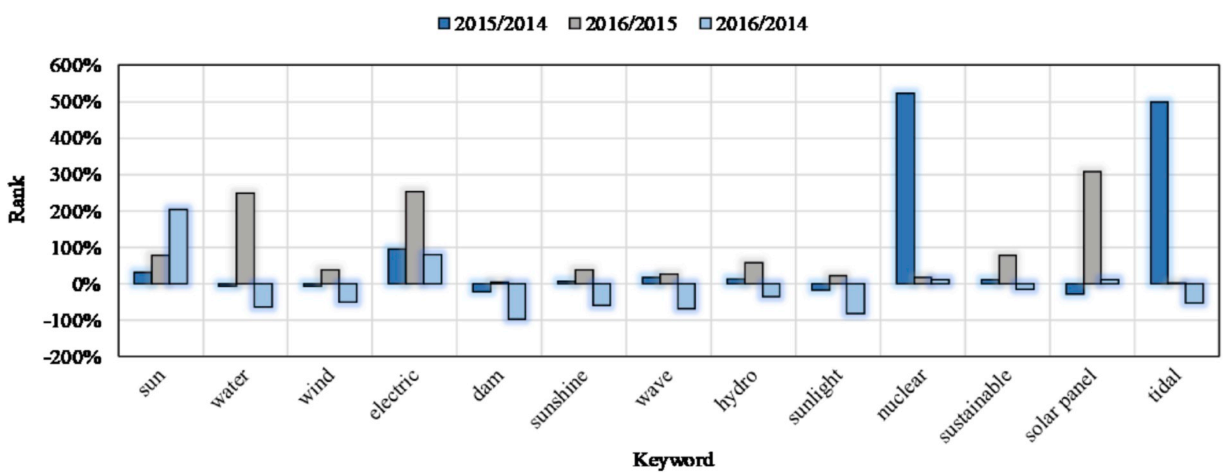



**Fig. 18.** Comparison of the amount of growth in rank of the most used renewable energy-related keywords in successive years.

Another result requiring discussion is how public opinion changes over time and how it can be affected by various external factors such as climate conditions, oil price, politics and legislation, cyber-attacks, etc. For example, the exceptionally warm winter in 2015 in Alaska in which people experienced record-setting warmth during the cold season resulted in more discussions about application of renewable energy resources since they are infinite, safe, and fuel free and cost-effective. This was further supported by another finding of our study. Most users considered it necessary to reduce fossil fuel dependence for electricity production and transition to renewables. The results showed more positive (either low or high) sentiments towards renewable energy.

The analysis of energy-related keywords indicated that there was a valuable growth rate for the word "sun", "solar panel" and the effect towards solar power was mostly positive, thus further strengthening the image of solar power. The reported preference could be attributable to a higher familiarity with solar and solar panels since this is one of the popular forms of renewable energy in Alaska in recent years. This can also be attributed to the recent investment in solar power in Alaska and a tremendous increase in requests for solar panel installation. More research is needed to identify the reasons accounting for Alaskans' familiarity and preference for solar power and investigate whether the same discussion exists in other States in the US.

In light of these findings, specific recommendations and policy implications can be considered. Primarily, the findings can be used as a groundwork to extend the research on awareness about renewable energy resources in different regions. In this respect, it is worthwhile that different scientific fields become more involved in developing policies and improving the existing financial, legislative, and social context, so that public support for and information about renewables are strengthened.

In future research, we will conduct similar studies in other states in the US and other countries to understand the crowd's opinion/preference regarding renewable energy resources. Studies have shown that worldwide energy issues could be managed by adding renewable energy sources to the current power generation and public opinion and analysis of public tweets play an essential role in facilitating the development of different renewable energy technologies. This would help researchers to understand the critical decision factors in implementing renewable energy resources and identify the hindering factors in using these resources in various geographic areas and most likely relate them to their socio-economic status. These results show the importance of energy-related discussions among people on Twitter.

## 6. Conclusions

The goal of this study is to exploit information-rich geotagged Twitter data for mining Alaskans' perceptions and opinions about energy efficiency measures, availability of energy sources, and the application of clean energy sources. This study analyzed the application of social media to understand the general public's perception and opinion about various energy sources and the choices they make at the individual level. In addition, people's perception of various energy resources and renewable energy over time was investigated in this study. The Twitter data was used to show how Alaskans' perception of energy-related topics changes over time from 2014 to 2016. The main contributions of this work are as follows:

1. This research proposes a new social media-based analysis (here Twitter) to understand the opinion of different individuals about both energy-related sources and renewable energy.
2. The results of this study reveal what is the crowd preference for specific energy resource.
3. The results of this research can be used by energy companies or government to provide proper energy resources to people based on their preference and/or inform them about the best sources of energy for them in terms of cost or other criteria.
4. In this research, the crowd preference can help to find out the geographic preference of users. In other words, we can categorize the preference of users based on their locations. This benefits to provide useful services based on the user's needs in their location.

The results of the energy-related keywords indicated that there is a valuable growth rate for the word "sun" in both 2016/2015 and 2016/2014 which is about 200%. The growth rate of the word "power" was the highest in comparison with other words (more than 700%) in 2016/2014. The rank of top 20 renewable energy-related keywords for three years shows the word "Tidal" has the highest ranking in all three years and "solar panel" takes the second rank in all three years. Both Tidal and Nuclear had the highest growth rate in 2015/2014 in comparing with other keywords and other years. The amount of growth for the word "solar panel" in 2016/2015 was significantly higher than in previous years (approximately 300%). In addition, an increasing trend was seen for the word "sun" from 2015/2014 to 2016/2014. As a result, it can be said that the attitude of Alaskans toward energy in general and renewable energy, in particular, was changed positively from 2014 to 2016. This means that attention to various types of energy is increasing dramatically among Alaskans.

## Author contributions

Moloud Abdar: Collected the data, Other contribution. Mohammad Ehsan Basiri: Collected the data, Other contribution. Junjun Yin: Conceived and designed the analysis, Other contribution. Mahmoud Habibnezhad: Conceived and designed the analysis, Collected the data, Contributed data or analysis tools, Other contribution. Guangqing Chi: Conceived and designed the analysis, Performed the analysis, Other contribution. Shahla Nemati: Conceived and designed the analysis, Collected the data, Contributed data or analysis tools, Other contribution. Somayeh Asadi: Collected the data, Other contribution.

## Acknowledgment

The National Science Foundation (Awards # CMMI-1541136, # OPP-1745369, # SES-1823633, and # ICER-1927827), the USDA National Institute of Food and Agriculture and Multistate Research Project #PEN04623 (Accession #1013257), and the Eunice Kennedy Shriver National Institute of Child Health and Human Development (Award # P2C HD041025).

## References

[1] Department of Energy. International energy outlook. 2018. Available from, https://www.eia.gov/outlooks/ieo/.
[2] Fay G, Villalobos Meléndez A, Pathan S. Alaska fuel price Projections 2012-2035. Alaska Energy Authority; 2012.
[3] Fesanghary M, Asadi S, Geem ZW. Design of low-emission and energy-efficient residential buildings using a multi-objective optimization algorithm. Build Environ 2012;49:245–50.
[4] Mostavi E, Asadi S, Boussaa D. Framework for energy-efficient building envelope design optimization tool. J Architect Eng 2018;24(2):04018008.
[5] Charles A, Maref W, Ouellet-Plamondon CM. Case study of the upgrade of an existing office building for low energy consumption and low carbon emissions. Energy Build 2019;183:151–60.
[6] Feng W, Zhang Q, Ji H, Wang R, Zhou N, Ye Q, Hao B, Li Y, Luo D, Lau S. A review of net zero energy buildings in hot and humid climates: experience learned from 34 case study buildings. Renew Sustain Energy Rev 2019;114:109303.
[7] Terlouw T, AlSkaif T, Bauer C, Van Sark W. Multi-objective optimization of energy arbitrage in community energy storage systems using different battery technologies. Appl Energy 2019;239:356–72.
[8] Rafiei SA, Mohammadi-ivatloo B, Asadi S, Goldani S, Falaghi H. Bi-level model for generation expansion planning with contract pricing of renewable energy in the presence of energy storage. IET Renew Power Gener 2019;13(9):1544–53.
[9] Nazari-Heris M, Mohammadi-Ivatloo B, Asadi S, Kim J, Geem ZW. Harmony search algorithm for energy system applications: an updated review and analysis. J Exp Theor Artif Intell 2019;31(5):723–49.
[10] Ntanos S, Kyriakopoulos G, Chalikias M, Arabatzis G, Skordoulis M. Public perceptions and willingness to pay for renewable energy: a case study from Greece. Sustainability 2018;10(3).
[11] Joy J, Joy D, Panwar TS. People's perception study: renewable energy in India. WWF-India; 2014.
[12] Haewoon K., Park,H., Moon, S., What is Twitter, a social network or a news media?, in *Proceedings of the 19th international conference on World wide web*. 2010, Acm: Raleigh, North Carolina, USA. p. 591-600.
[13] Alp ZZ, Öğüdücü ŞG. Influence factorization for identifying authorities in twitter. Knowl Base Syst 2019;163:944–54.
[14] Rosen A. Tweeting made Easier. 2017. Available from, https://blog.twitter.com/en_us/topics/product/2017/tweetingmadeeasier.html.
[15] Dahal B, Kumar S, Li Z. Topic modeling and sentiment analysis of global climate change tweets. Social Network Analysis and Mining 2019;9(1):24.
[16] Lee H, Abdar M, Yen NY. Event-based trend factor analysis based on hashtag correlation and temporal information mining. Appl Soft Comput 2018;71:1204–15.
[17] Ryong L, Kazutoshi S. Measuring geographical regularities of crowd behaviors for Twitter-based geo-social event detection. Proceedings of the 2nd ACM SIGSPATIAL International Workshop on location based social networks. California: ACM: San Jose; 2010. p. 1–10.
[18] Intergovernmental Panel on Climate Change (IPCC). Climate change 2013: the physical science basis. In: Contribution of working group I to the fifth assessment report of the intergovernmental panel on climate change; 2013. Cambridge, UK, report.
[19] Bach W. Impact of increasing atmospheric CO2 concentrations on global climate: potential consequences and corrective measures. Environ Int 1979;2(4):215–28.
[20] Edenhofer O, Pichs-Madruga R, Sokona Y, Seyboth K, Matschoss P, Kadner S, Zwickel T, Eickemeier P, Hansen G, Schloemer S, Von Stechow C. Renewable energy sources and climate change mitigation. IPCC; 2011. p. 1075.
[21] United Nation Climate Change., The Paris Agreement: work Programme under the Paris Agreement.
[22] Greenberg M. Energy sources, public policy, and public preferences: analysis of US national and site-specific data. Energy Pol 2009;37(8):3242–9.
[23] Farhar BC. Energy and the environment: the public view. 1996.
[24] Nuortimo K, Härkönen J, Karvonen E. Exploring the global media image of solar power. Renew Sustain Energy Rev 2018;81:2806–11.
[25] Nuortimo K, Härkönen J. Opinion mining approach to study media-image of energy production. Implications to public acceptance and market deployment. Renew Sustain Energy Rev 2018;96:210–7.
[26] Opinion.org, W.P.. World public strongly favor requiring more wind and solar energy, more efficiency, even if it increases costs. 2015.
[27] Muhammad-Sukki F, Ramirez-Iniguez R, Abu-Bakar S, McMeekin SG, Stewart BG. An evaluation of the installation of solar photovoltaic in residential houses in Malaysia: past, present, and future. Energy Pol 2011;39(12):7975–87.
[28] Eshchanov BR, Stultjes M, Eshchanov R, Salaev S. People's perceptions on renewable energy sources penetration prospects in the Khorezm Province, Uzbekistan. J. Knowledge Management, Economics and Information Technol. 2011;1(7).
[29] Hagen B, Pijawka D. Public perceptions and support of renewable energy in north America in the context of global climate change. Int. J. Disaster Risk Sci 2015;6(4):385–98.
[30] Goodchild MF, Glennon JA. Crowdsourcing geographic information for disaster response: a research frontier. Int. J. Digital Earth 2010;3(3):231–41.
[31] Earle P, Bowden D, Guy M. Twitter earthquake detection: earthquake monitoring in a social world. Ann Geophys 2011;54(6).
[32] Sit MA, Koylu C, Demir I. Identifying disaster-related tweets and their semantic, spatial and temporal context using deep learning, natural language processing and spatial analysis: a case study of Hurricane Irma. Int. J. Digital Earth 2019:1–25.
[33] Wang, H., Hovy, E., Dredze, M., The Hurricane Sandy twitter corpus, in Advancement of Artificial intelligence 2015, AAAI Workshop.
[34] Leung M, Yu J, Liang Q. Improving public engagement in construction development projects from a stakeholder's perspective. J Construct Eng Manag 2013;139(11):04013019.
[35] Russell SM, Menassa CC, McCloskey E. Lifelong learning to leverage project and career success: 21st-century imperative. Pract Period Struct Des Construct 2014;19(1):137–41.

[36] Miles BG, Huxford J. Restoration and impacts from the september 8, 2011, san Diego power outage. J Infrastruct Syst 2014;20(2):05014002.
[37] Miles B, Jagielo N, Gallagher H. Hurricane isaac power outage impacts and restoration. J Infrastruct Syst 2016;22(1):05015005.
[38] Tanielian Adam R. Balancing harmful impacts of the petroleum industry with internal dispute prevention and external sustainable development initiatives. J Leg Aff Dispute Resolut Eng Constr 2014;6(2):03013003.
[39] Sharif A, Raza A, Ozturk I, Afshan S. The dynamic relationship of renewable and nonrenewable energy consumption with carbon emission: a global study with the application of heterogeneous panel estimations. Renew Energy 2019;133:685–91.
[40] Soytas U, Sari R. Energy consumption and GDP: causality relationship in G-7 countries and emerging markets. Energy Econ 2003;25(1):33–7.
[41] Soytas U, Sari R. Energy consumption and income in G-7 countries. J Pol Model 2006;28(7):739–50.
[42] Narayan P, Smyth R. Energy consumption and real GDP in G7 countries: new evidence from panel cointegration with structural breaks. Energy Econ 2008;30(5): 2331–41.
[43] Behera J, Mishra A. Renewable and non-renewable energy consumption and economic growth in G7 countries: evidence from panel autoregressive distributed lag (P-ARDL) model. International Economics and Economic Policy; 2019.
[44] Qazi A, Hussain F, Rahim N, Hardaker G, Alghazzawi D, Shaban K, Haruna K. Towards sustainable energy: a systematic review of renewable energy sources, technologies, and public opinions. IEEE Access 2019;7:63837–51.
[45] Jain A, Jain V. Renewable energy sources for clean environment: opinion mining. Asian J Water Environ Pollut 2019;16(2):9–14.
[46] Gupta K, Ripberger J, Wehde W. Advocacy group messaging on social media: using the narrative policy framework to study twitter messages about nuclear energy policy in the United States. Pol Stud J 2018;46(1):119–36.
[47] Ikoro V, Sharmina M, Malik K, Batista-Navarro R. Analyzing sentiments expressed on twitter by UK energy company consumers. Fifth International conference on social networks analysis, management and security. SNAMS); 2018. 2018.
[48] Basiri ME, Kabiri A. Sentence-level sentiment analysis in Persian. 3rd International conference on pattern recognition and image analysis. IPRIA); 2017. 2017.
[49] Bird S, Klein E, Loper E. natural language processing with Python: analyzing text with the natural language Toolkit. first ed. O'Reilly; 2009.
[50] da Silva NFF, Hruschka ER. Tweet sentiment analysis with classifier ensembles. Decis Support Syst 2014;66:170–9.
[51] Taboada M, Brooke J, Tofiloski M, Voll K, Stede M. Lexicon-based methods for sentiment analysis. Comput Ling 2011;37(2):267–307.
[52] Basiri ME, Kabiri A. Translation is not enough: comparing Lexicon-based methods for sentiment analysis in Persian. International Symposium on computer science and Software engineering conference. CSSE); 2017. 2017.
[53] TextBlob. Available from: https://textblob.readthedocs.io/en/dev/.
[54] Miller GA. WordNet: a lexical database for English. Commun ACM 1995;38(11): 39–41.
[55] Nemati S, Naghsh-Nilchi A. An evidential data fusion method for affective music video retrieval. Intell Data Anal 2017;21(2):427–41.
[56] Nicolaou MA, Gunes H, Pantic M. Continuous prediction of spontaneous affect from multiple cues and modalities in valence-arousal space. IEEE Trans. on Affective Computing 2011;2(2):92–105.
[57] Nemati S, Naghsh-Nilchi A. Incorporating social media comments in affective video retrieval. J Inf Sci 2015;42(4):524–38.
[58] Xu Z. Intuitionistic fuzzy aggregation operators. IEEE Trans Fuzzy Syst 2007;15(6): 1179–87.
[59] Basiri ME, Naghsh-Nilchi A, Ghasem-Aghaee N. Sentiment prediction based on dempster-shafer theory of evidence. Mathematical Problems in Engineering; 2014. p. 13. 2014.
[60] Schouten K, Frasincar F. Survey on aspect-level sentiment analysis. IEEE Trans Knowl Data Eng 2016;28(3):813–30.
[61] Shafer G. Dempster-shafer theory. Encyclopedia of artificial intelligence 1992;1: 330–1.
[62] Velusamy D, Pugalendhi G. Fuzzy integrated Bayesian Dempster–Shafer theory to defend cross-layer heterogeneity attacks in communication network of Smart Grid. Inf Sci 2019;479:542–66.
[63] Hatefi M, Basiri ME, Tamošaitienė J. An evidential model for environmental risk assessment in projects using dempster–shafer theory of evidence. Sustainability 2019;11(22).
[64] Bi Y. The impact of diversity on the accuracy of evidential classifier ensembles. Int J Approx Reason 2012;53(4):584–607.