# Energy Choices in Alaska: Mining People's Perception and Attitudes from Geotagged Tweets

**Moloud Abdar, Mohammad Ehsan Basiri, Junjun Yin, Mahmoud Habibnezhad, Guangqing Chi, Shahla Nemati, Somayeh Asadi**

[1] Département d'Informatique, Université du Québec à Montréal, Montréal, QC, Canada

[2] Department of Computer Engineering, Shahrekord University, Shahrekord

[3] Social Science Research Institute and Population Research Institute, Pennsylvania State University, University Park, PA 16802, USA

[4] Department of Architectural Engineering, Pennsylvania State University, 104 Engineering Unit A, University Park, PA 16802, USA

[5]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, U.S.A.

The first and second authors (M. Abdar and M. E. Basiri) contributed equally to the manuscript.

## Abstract

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 in order to solving 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 investigates 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 applied 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". 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 Alaskans' energy choices are more acceptive towards or even favor renewable energy in the future.

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

## 1. Introduction

The world is quickly urbanizing, and society depends on food, water, and energy. More than 50 percent of the world's population now lives in cities, and they are projected to house approximately 60 percent of the global population by 2030 [1]. Worldwide energy use is anticipated to increase up to 50 percent by 2035 [2] and total global irrigation water use is anticipated to increase by 10 percent by 2050 [3]. This condition is projected to be aggravated soon as 60 percent more food will need to be produced in order to satisfy the needs of the world's population in 2050.

In Alaska in particular, nearly 103,000 Alaskans—roughly 1 in 7—struggle with hunger, and many of Alaska's rural communities lack running water and sewer systems [4]. Total energy consumption per capita in Alaska was 809 million BTU in 2016 which ranked third after Louisiana and Wyoming [2]. In Alaska, 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 [5]. 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. To overcome these challenges, various strategies including energy conservation, energy efficiency, and adoption of clean energy should be implemented considering this fact that Alaska is uniquely endowed with a full range of renewable energy opportunities. Indeed, Alaska has a nonbinding goal to generate 50% of its electricity from renewable sources by 2025 [2]. Using these sources to develop practical renewable energy solutions will inevitably have effects on the security of food and water systems. The key first step in solving Alaska's domestic energy problems and creating a sustainable future is to understand Alaskans' perceptions and opinions about energy consumption, various energy sources in particular renewable energy solutions. However, a key factor that hinders the widespread adoption of renewable energy systems, besides its cost, is the lack of proper public information about renewable energy systems and lack of "social acceptance" assessment. Social acceptance evaluates the "degree of readiness of citizens to invest in renewable energy in their area". 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 [6-9].

A typical approach to understanding people's perception is to conduct surveys [7]. However, it is very expensive 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 especially useful venue to study people's interactions with various topics.

Twitter allows its users to post short, 280-character messages and to follow messages from other users. 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 properties of Twitter simplify and expedite information dissemination; its social network properties ease the access to geographically and personally relevant information. With the increasing importance of energy efficiency measures, availability of energy sources, and the application of clean energy sources, people are likely to share their thoughts and sentiments on Twitter.

To this end, 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, we 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?*

## 2. Background

Climate change cause substantial environmental, social, and economic risks for current and future generations. Recent studies emphasize the key role of human activities on the fast changes experienced in global climate during the last several decades which resulted in emissions of greenhouse gases [10]. According to data published by International Energy Agency, between 1971–2014, worldwide primary energy use has increased by 2.5 and carbon dioxide emissions have been doubled. Bach [11] 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 major challenge to 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 [12]. The Kyoto Protocol and the Paris Agreement have directed some of the courtiers in an extraordinary shift towards renewable energy sources [13].

However, existing studies on public attitudes about energy have primarily focused on an opinion poll, interview, and/or survey research of public attitudes at the general 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 [14-17]. 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 [18]. The results indicated strong public support (77%) of governmental efforts to shift to increase renewable energy sources especially solar. Similar survey poll was conducted in the US in 2008 and 90% of the participants were in favor of renewable energy [14]. 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 their lack of information about all the potential incentives and the socio-economic benefits of investing in solar panels [19]. In a similar study, Eshchanov et al. [20] 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 the fossil fuel energy sources can be as 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 renewable energy and India's 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 application of renewable energy sources [21]. However, the sample size was considerably small compare to the population with a total number of 2312. The small size sample was one the key barriers to making a more informed decision; hence, we need to have larger sample sizes per country to better understand the perceptual differences between countries and identify the affecting variables. This considers as a serious 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.

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. Having access to the enormous Twitter data and due to its real-time feature, large-scale and fast-propagation as well as researchers' growing capability to process "Big Data," Twitter data has been leveraged successfully to expedite the knowledge discovery process in a wide variety of fields.

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 [22] 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. [23] 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 disaster. Building on a similar concept, other studies confirm the association between the spatiotemporal distribution of tweets and the physical scope of floods [24]. In addition, the relation between the occurrence of disaster-related tweets and the distribution of Hurricane Sandy damage projected from simulation models was determined [25]. Another strand focuses on the use of social media as a communication tool in the different phases of the construction project. Leung et al. [26] 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. [27] 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. [28] [29] analyzed the post-event power outage in San Diego County, California in 2011 in which its residents were without electricity for up to 12h. 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 [30] found that social media can be used to encourage fundamental change in energy-consumption behaviors, exchange ideas, and broadcast theories.

We believe that perceptions and attitudes about energy-related topics at a public 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 [31]. This approach enables the fast and scalable analysis of huge 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 details in the following subsections but first, a precise specification of the data analyzed in the current study has been discussed.

### 3.1.Proposed Methodology

The overall structure of the proposed methodology is depicted in Figure 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.

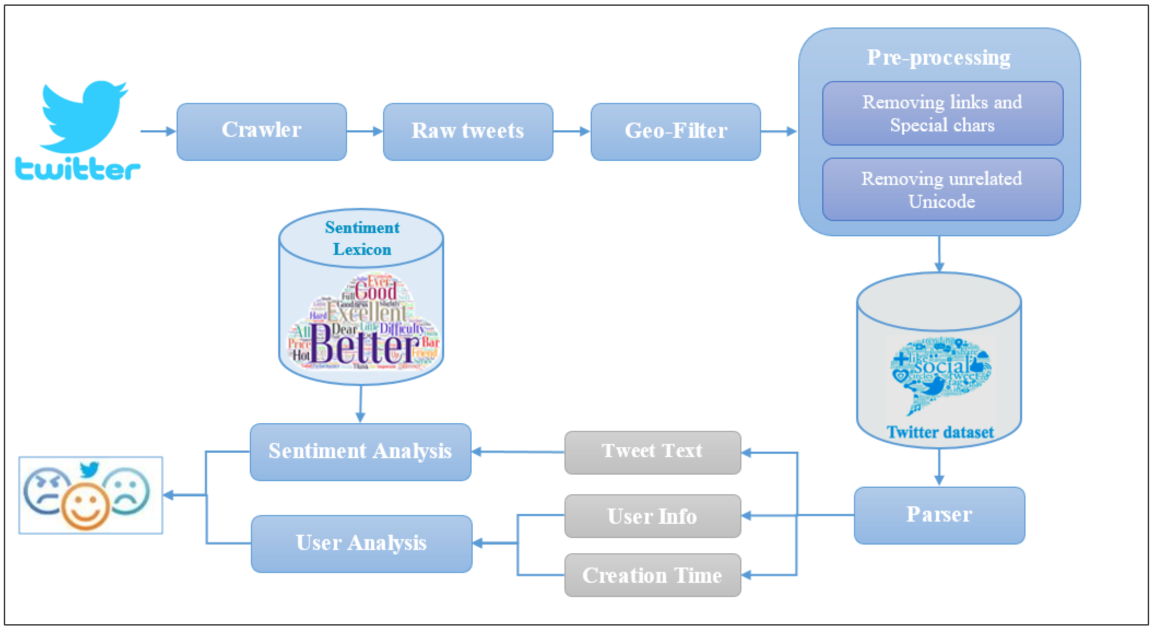
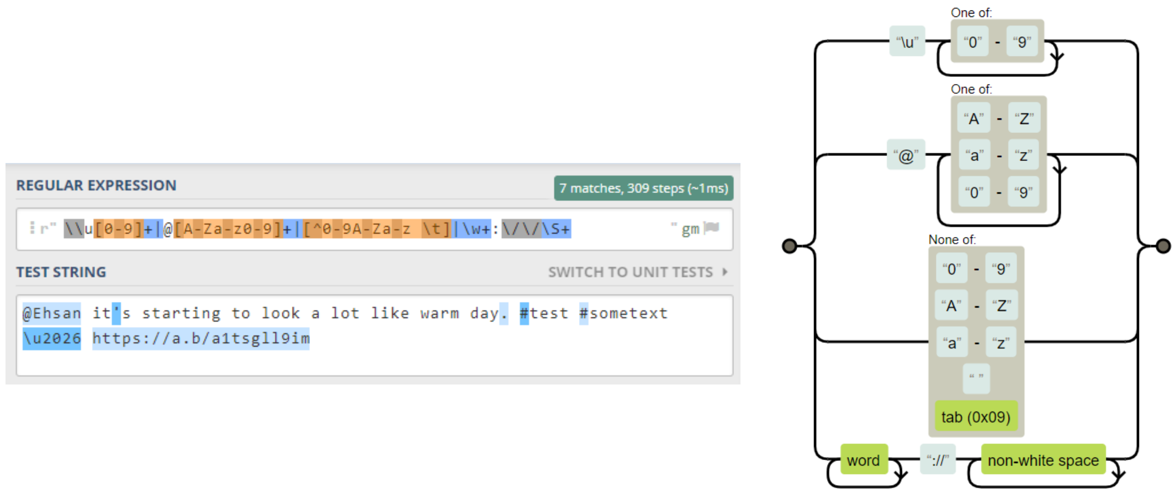



Figure 1. The overall structure of the proposed system.

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. In order to achieve these goals, a Natural Language Processing (NLP) approach based on regular expressions is employed [32]. In this approach, as illustrated in Figure 2, regular expressions are used to find irrelevant parts of tweets. A regular expression is a common language for specifying and matching patterns in NLP-related tasks [32] and the designed regular expressions in Python programming language was 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 [33]. Text-related information is analyzed in the sentiment analysis module and user-related information is used in the user analysis module for further processing.




(a) (b)

Figure 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 figure shown in (a) was generated using an online tool in https://regex101.com and the figure shown in (b) was generated using a similar online tool in https://regexper.com.

In order to analyze the opinions of Alaskan in their tweets, a lexicon-based approach [34] 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 [31]. Following the approach proposed in [35], 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 [36] that is a python library for natural language processing tasks. In this lexicon, adjectives have polarity score 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 a high coverage of sentiment-bearing words.

Words in ADJLex are tagged according to their WordNet [37] 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 [38]. 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 [34]. In order to apply the effect of such modifiers, we considered three possible situations as follows:

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 common dimensional model of emotion [39]. 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 Figure 3 (adopted from [40]).

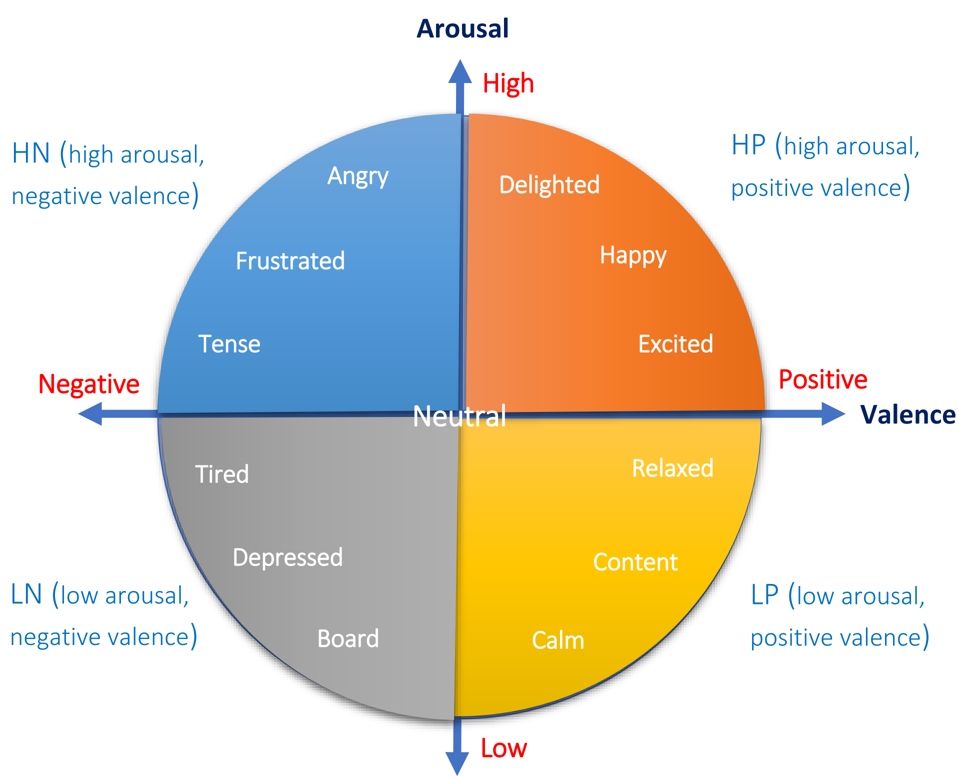



Figure 3. Distribution of different fillings in the two–dimensional VA space (adopted from [Nemati and Naghsh-Nilchi 2016]).

In the current study, based on the polarity scores calculated in the previous step, four class labels are defined: 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 Figure 3.

As mentioned in the previous subsection, polarity scores are represented in the range of $[-1, +1]$. Therefore, in order 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 [41]. Although simple fusion methods such as arithmetic mean, majority voting, sum, product, and maximum [42] are widely used in different applications, it has been shown that more formal fusion methods including Dempster-Shafer (DS) theory of evidence outperforms simple heuristic methods [43]. DS theory is an evidential theory for addressing the problem of uncertainty and can be seen as a generalization of Bayesian fusion rule [40]. This theory was first proposed by Dempster and later developed by Shafer [44]. DS-based fusion method has been used for aggregating sentence-level sentiment scores into a document-level score [42], 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 obvious 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 special 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:

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

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

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

In order to define the mass function, we followed a Fuzzy-based approach. Specifically, we first defined fuzzy membership functions as depicted in Figure 4. As shown in the 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.

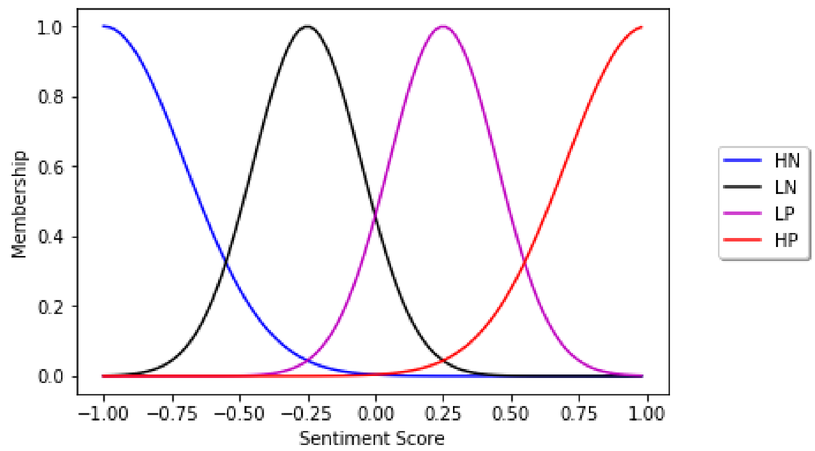



Figure 4. Membership function of sentiment score with four linguistic terms.

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 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:

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

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

Where relationship $K_{12}$ is a normalizing factor to ensure 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 [31]. 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 [40]. 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.[45] 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:

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

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

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

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

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 word, $m^{\sigma}$ 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 is shown in Equations (6) and (7) should also be changed according to the above three cases. In the first case, the following four equations are used for the fusion of two mass functions.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

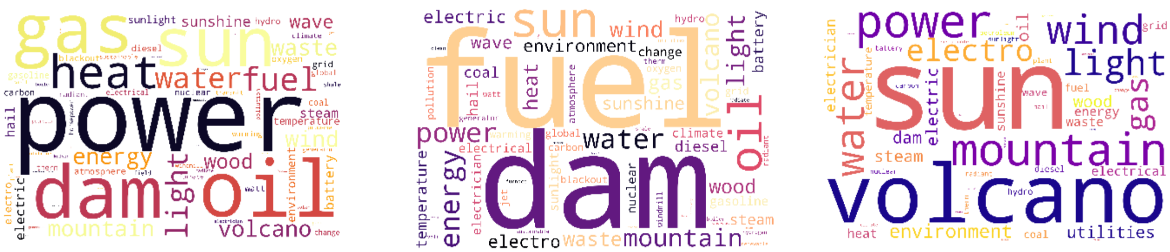
After the fusion step, the class with the highest probability is selected as the label of the tweet.

## 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 relevant to the subject of energy. Then, the word cloud of keywords, the most frequently appearing keywords, for three years is illustrated in Figure 5. Word clouds have been used as an effective 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 represent that they have been used more than 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" is 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 key issue is most of these resources are located in isolated areas which is far from a population center that would use the electricity generated.

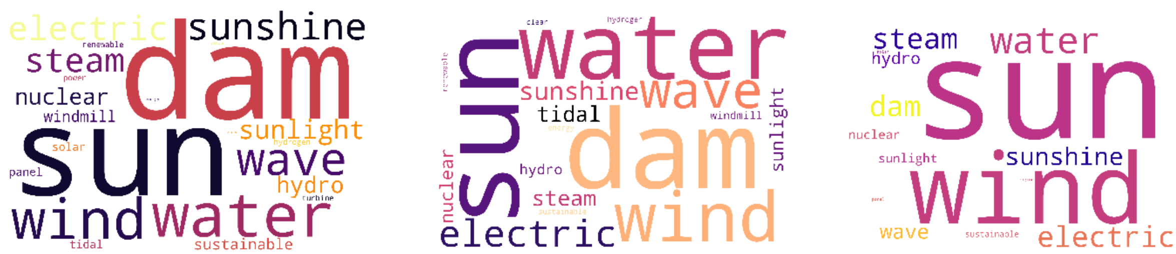



(a) (b) (c)

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

With high energy prices distressing many rural areas and 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 {Institute, #49}, more than 50% of the nation's wave energy resources and more than 90% of the 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 appearing keywords related to renewable energy sources. For this purpose, 83 renewable energy-related words were chosen, and their word cloud is illustrated in Figure 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.




(a) (b) (c)

Figure 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.

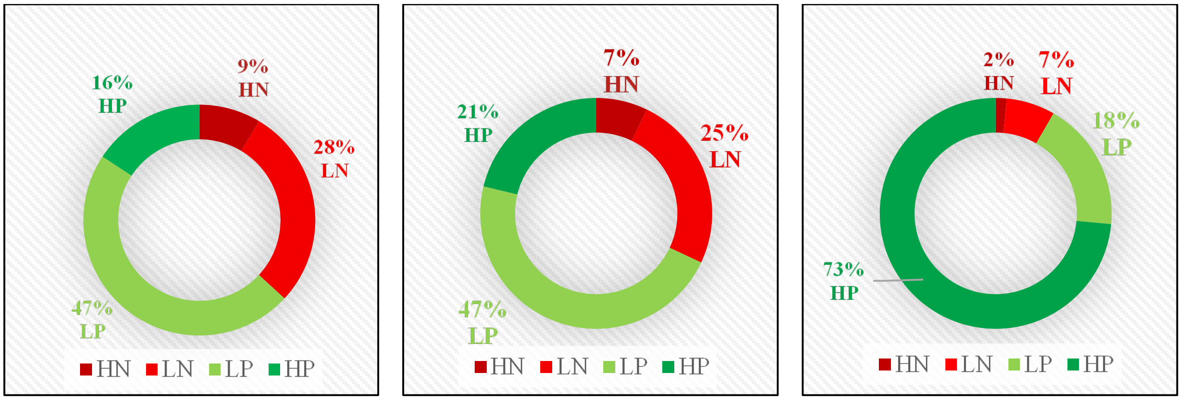
The potential for using solar energy in Alaska has long agonized from the belief that the sun basically does not offer any hope for Alaskans while there are 230 hours more of possible sunlight at the Arctic Circle than at the equator. In addition, the price to harvest solar energy has not only decreased dramatically over the last decade but also solar photovoltaic (PV) has minimal ongoing maintenance costs compare to wind turbine. This explains why Alaskans have considered solar energy as a potential source of energy in recent years. We also found there are more discussions about solar energy in 2016 compared to previous years. This can be explained by the report published by Department of Energy in February 2016 emphasizing the unique advantages in solar energy despite the northern latitude. 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 region of Alaska are at least comparable to that of Germany, which is leading the world in PV installations with more than 38,500 megawatts (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 main 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 Figure 7.




(a) (b) (c)

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

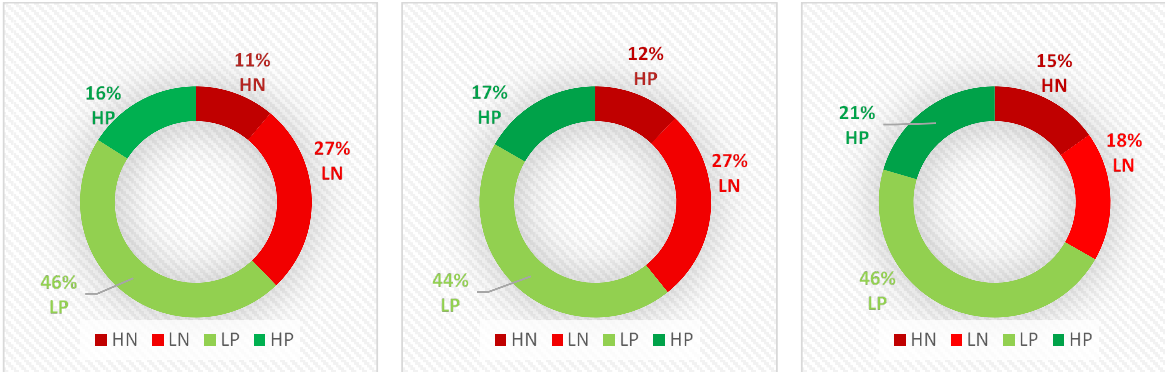
As can be seen in this figure, in 2014, about one-half (47%) of the tweets are positive, less than 20% of them are negative, and the remaining are neutral. However, in 2015 and 2016, a significant increase is seen in positive sentiment. In order to track the Alaskans' sentiment toward energy, the frequency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

Table 1. Comparison of the frequency and change rate of positive, negative, and neutral sentiments toward 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% |

According to Table 1, the frequency of positive sentiments toward 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' percept about energy.

According to Figures 4, 5, and 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 toward energy.




(a) (b) (c)

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

Table 2. Comparison of the frequency and change rate of positive, negative, and neutral sentiments toward 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% |

*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?*

In order to respond to this question, the distribution of both sent messages in different months and 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 has sent more than one tweet in a month. As can be seen in this table, there are different behaviors in 2014, 2015, and 2016; 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%, 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 and 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 distribution of a 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 other two years. This can be explained by exceptionally

warm winter of that year in Alaska which Alaskans experienced record-setting warmth during its cold season in 2015.
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.

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% |

*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 Figures 9, 10, and 11, respectively.

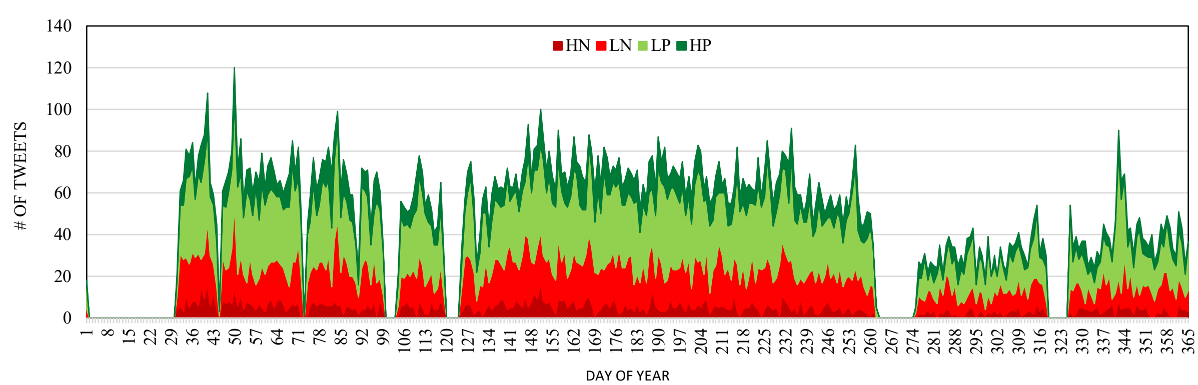


Figure 9. Trend of changing four different sentiments in 2014.

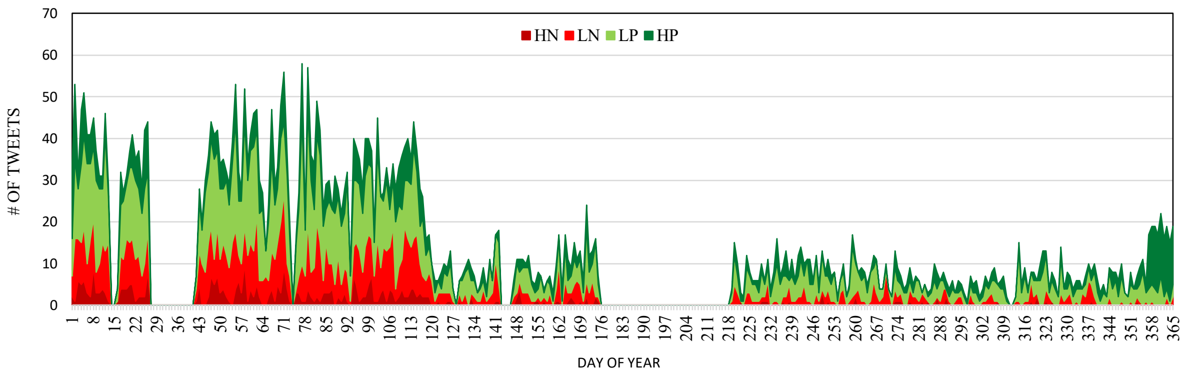



Figure 10. Trend of changing four different sentiments in 2015

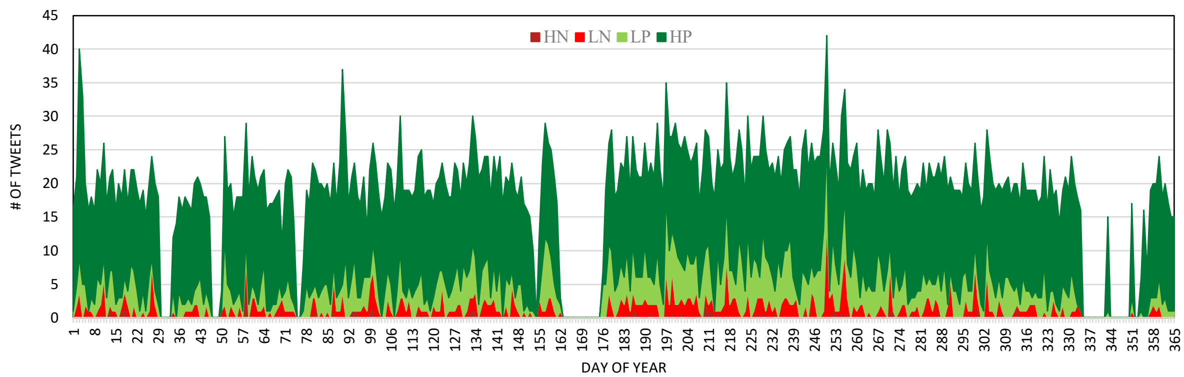



Figure 11. Trend of changing four different sentiments in 2016

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 Figures 9, 10, and 11, the total number of tweets in 2014 is higher than a 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 smaller 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 Figures 12, 13, and 14, respectively.

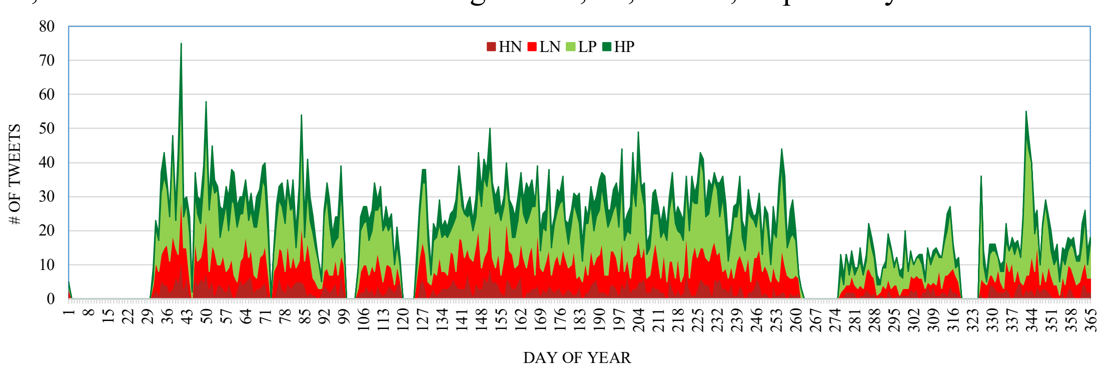


Figure 12. Trend of changing four different renewable energy related sentiments in 2014

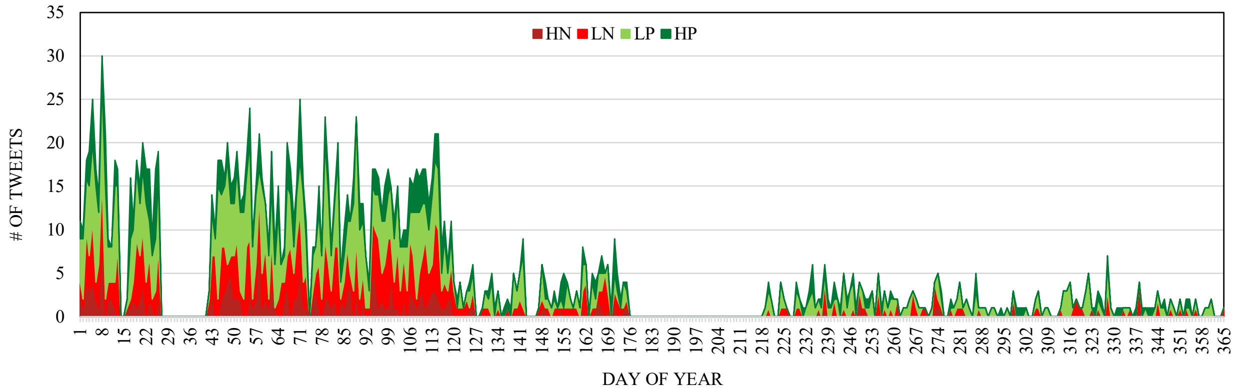



Figure 13. Trend of changing four different renewable energy related sentiments in 2015

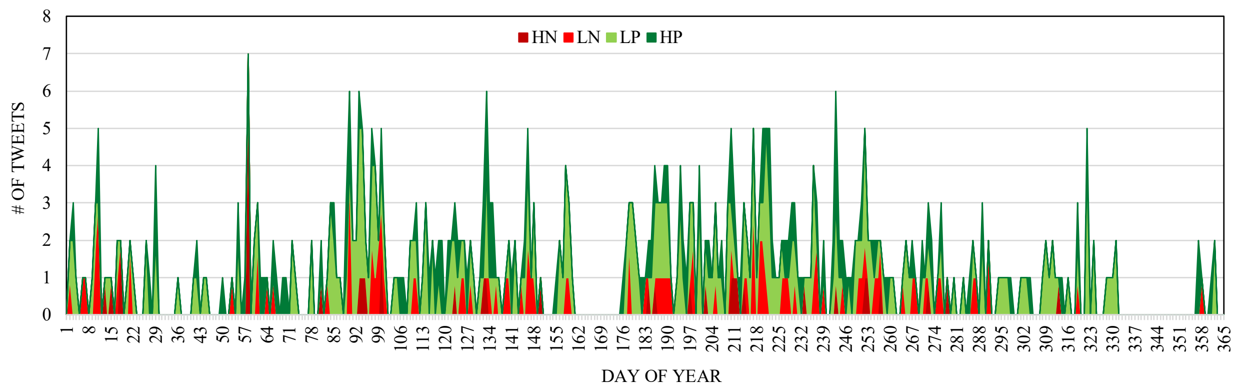



Figure 14. Trend of changing four different renewable energy related sentiments in 2016

According to Figures 12, 13, and 14, it can be seen that 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. Figure 12 illustrates people had showed low negative (LN) sentiments compared to other sentiments (HP, HN, and LN). As Figure 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 obvious 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 Figures 15, 16, 17, and 18. As can be seen in Figure 15, energy waste had the highest rank in both 2014 and 2015 while words heat and steam had the highest rank in 2016. The word sun had the lowest rank in all three years.

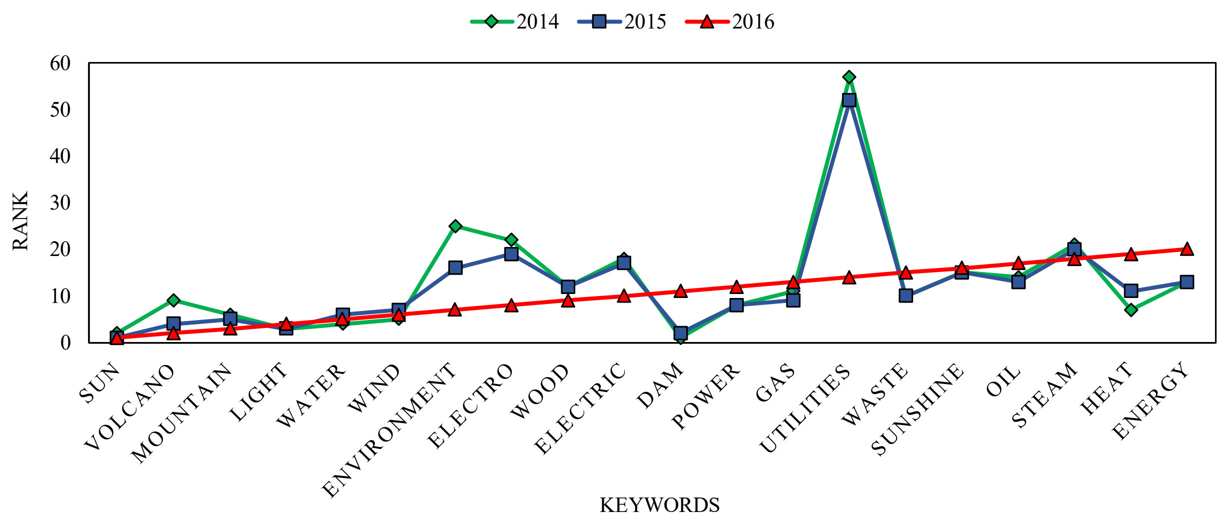



Figure 15. Comparison of the rank of top-20 energy-related keywords for three years.

Figure 16 compares the amount of growth in the rank of top-20 energy-related keywords in successive years. As can be seen, there is 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.

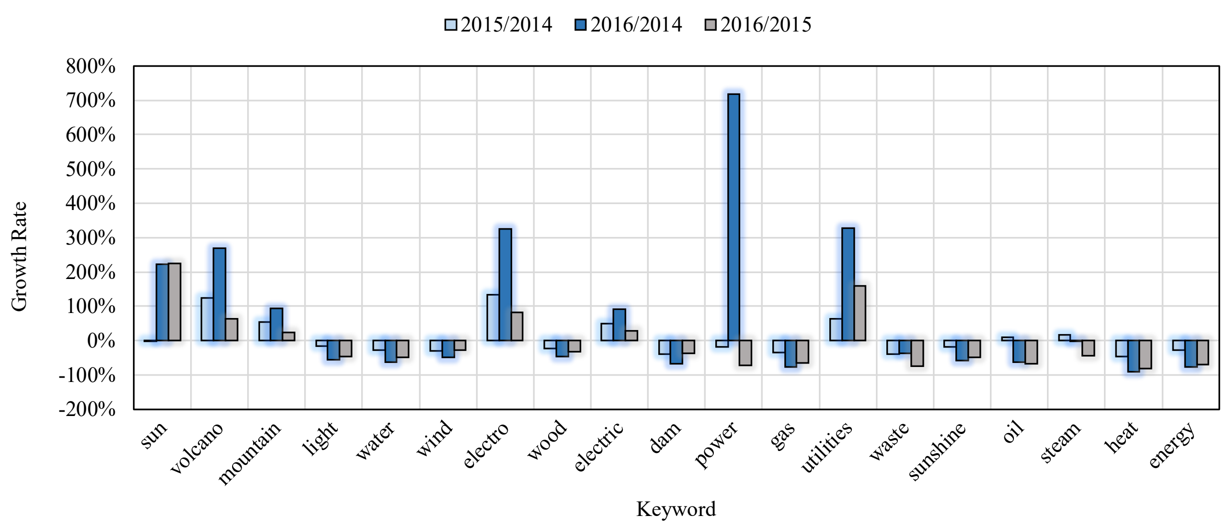



Figure 16. Comparison of the amount of growth in rank of top-20 energy-related keywords in successive years.

Figure 17 compares the rank of top 20 renewable energy-related keywords for three years. As can be seen Tidal has the highest ranking in all three years and solar panel 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 Figure 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 in compare with previous years (approximately 300%). In addition, an increasing trend was seen for word "sun" from 2015/2014 to 2016/2014.

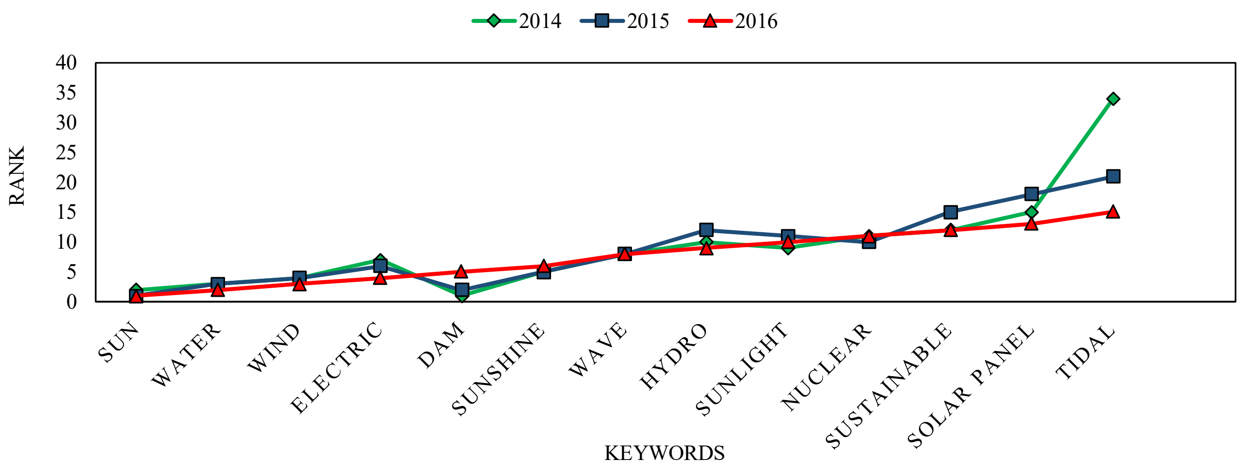



Figure 17. Comparison of the rank of top-20 renewable energy-related keywords for three years.

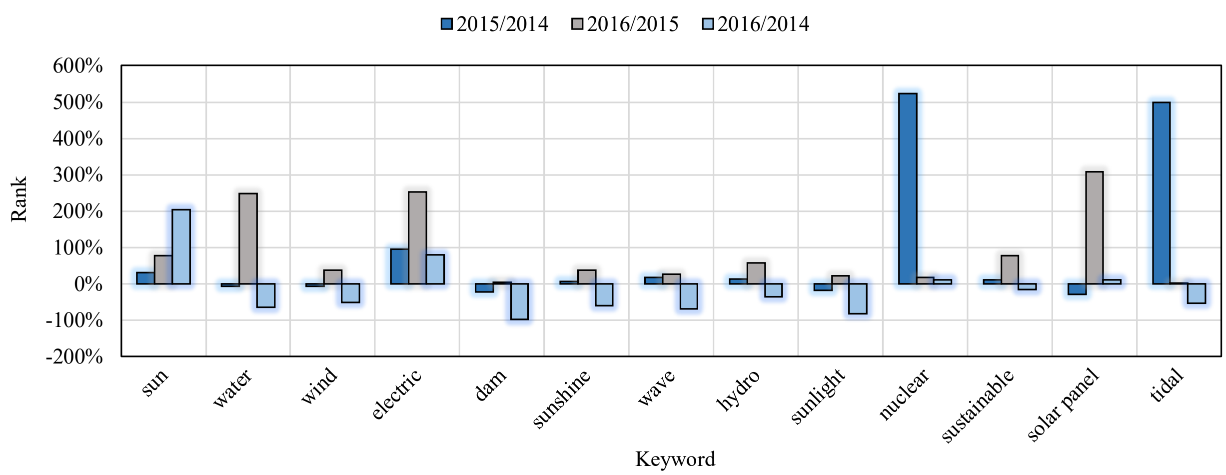



Figure 18. Comparison of the amount of growth in rank of top-12 renewable energy-related keywords in successive years.

## 5. 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 analyzes 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 2017. The main contributions of this work are as follows:

1. This research proposes a new social media-based analysis (here Twitter) to find out 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 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 show word "Tidal" has the highest ranking in all three years and "solar panel" take 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.

**References**


1. PBS. *More than half the world's population lives in urban areas, UN report finds*. 2014; Available from: https://www.pbs.org/newshour/world/half-worlds-population-live-urban-areas-un-report-finds.
2. Energy, D.o. *International Energy Outlook*. 2018; Available from: https://www.eia.gov/outlooks/ieo/.
3. Nations, F.a.A.O.o.t.U., *How to Feed the World in 2050.* 2009.
4. Alaska, F.B.o., *Facts About Hunger*
5. Fay, G., Villalobos Meléndez, A., Pathan, S., *Alaska Fuel Price Projections 2012-2035*. 2012, Alaska Energy Authority.
6. Ntanos, S., et al., *Public Perceptions and Willingness to Pay for Renewable Energy: A Case Study from Greece.* Sustainability, 2018. **10**(3).
7. Joy, J., Joy, D., Panwar, T.S., *People's Perception Study: Renewable Energy in India* 2014: WWF-India
8. Liu, W., C. Wang, and A.P.J. Mol, *Rural public acceptance of renewable energy deployment: The case of Shandong in China.* Applied Energy, 2013. **102**: p. 1187-1196.
9. Caporale, D. and C. De Lucia, *Social acceptance of on-shore wind energy in Apulia Region (Southern Italy).* Renewable and Sustainable Energy Reviews, 2015. **52**: p. 1378-1390.
10. Change), I.I.P.o.C., *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.

11. Bach, W., *Impact of increasing atmospheric CO2 concentrations on global climate: Potential consequences and corrective measures.* Environment International, 1979. **2**(4): p. 215-228.
12. Ottmar Edenhofer, R.P.-M., Youba Sokona, Kristin Seyboth, Patrick Matschoss, Susanne Kadner, Timm Zwickel, Patrick Eickemeier, Gerrit Hansen, Steffen Schloemer, Christoph von Stechow, *Renewable Energy Sources and Climate Change Mitigation*. 2011, IPCC. p. 1075.
13. Change, U.N.C., *The Paris Agreement: Work Programme under the Paris Agreement*.
14. Greenberg, M., *Energy sources, public policy, and public preferences: Analysis of US national and site-specific data.* Energy Policy, 2009. **37**(8): p. 3242-3249.
15. Farhar, B.C., *Energy and the environment: The public view*. 1996.
16. Nuortimo, K., J. Härkönen, and E. Karvonen, *Exploring the global media image of solar power.* Renewable and Sustainable Energy Reviews, 2018. **81**: p. 2806-2811.
17. Nuortimo, K. and J. Härkönen, *Opinion mining approach to study media-image of energy production. Implications to public acceptance and market deployment.* Renewable and Sustainable Energy Reviews, 2018. **96**: p. 210-217.
18. WorldPublicOpinion.org, *World public strongly favor requiring more wind and solar energy, more efficiency, even if it increases costs*. 2015.
19. Muhammad-Sukki, F., et al., *An evaluation of the installation of solar photovoltaic in residential houses in Malaysia: Past, present, and future.* Energy Policy, 2011. **39**(12): p. 7975-7987.
20. Eshchanov, B.R., et al., *People's Perceptions on Renewable Energy Sources Penetration Prospects in the Khorezm Province, Uzbekistan.* Journal of Knowledge Management, Economics and Information Technology, 2011. **1**(7).
21. Hagen, B. and D. Pijawka, *Public Perceptions and Support of Renewable Energy in North America in the Context of Global Climate Change.* International Journal of Disaster Risk Science, 2015. **6**(4): p. 385-398.
22. Goodchild, M.F. and J.A. Glennon, *Crowdsourcing geographic information for disaster response: a research frontier.* International Journal of Digital Earth, 2010. **3**(3): p. 231-241.
23. Paul S. Earle, D.C.B., Michelle Guy, *Twitter earthquake detection: earthquake monitoring in a social world.* Annals Of Geophysics, 2011. **54**(6).
24. Sit, M.A., C. Koylu, and I. Demir, *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.* International Journal of Digital Earth, 2019: p. 1-25.
25. Wang, H., Hovy, E., Dredze, M., *The Hurricane Sandy Twitter Corpus*, in *Advancement of Artificial Intelligence* 2015, AAAI Workshop.
26. Leung, M.-y., J. Yu, and Q. Liang, *Improving Public Engagement in Construction Development Projects from a Stakeholder's Perspective.* Journal of Construction Engineering and Management, 2013. **139**(11): p. 04013019.
27. Russell Jeffrey, S., C. Menassa Carol, and E. McCloskey, *Lifelong Learning to Leverage Project and Career Success: 21st-Century Imperative.* Practice Periodical on Structural Design and Construction, 2014. **19**(1): p. 137-141.

28. Miles Scott, B., H. Gallagher, and J. Huxford Charles, *Restoration and Impacts from the September 8, 2011, San Diego Power Outage.* Journal of Infrastructure Systems, 2014. **20**(2): p. 05014002.
29. Miles Scott, B., N. Jagielo, and H. Gallagher, *Hurricane Isaac Power Outage Impacts and Restoration.* Journal of Infrastructure Systems, 2016. **22**(1): p. 05015005.
30. Tanielian Adam, R., *Balancing Harmful Impacts of the Petroleum Industry with Internal Dispute Prevention and External Sustainable Development Initiatives.* Journal of Legal Affairs and Dispute Resolution in Engineering and Construction, 2014. **6**(2): p. 03013003.
31. Basiri, M.E. and A. Kabiri. *Sentence-level sentiment analysis in Persian*. in *2017 3rd International Conference on Pattern Recognition and Image Analysis (IPRIA)*. 2017.
32. Bird, S., Klein, E., Loper, E., *Natural Language Processing with Python: Analyzing Text with the Natural Language Toolkit*. 1st ed. 2009: O'Reilly.
33. da Silva, N.F.F., E.R. Hruschka, and E.R. Hruschka, *Tweet sentiment analysis with classifier ensembles.* Decision Support Systems, 2014. **66**: p. 170-179.
34. Taboada, M., et al., *Lexicon-Based Methods for Sentiment Analysis.* Computational Linguistics, 2011. **37**(2): p. 267-307.
35. Basiri, M.E. and A. Kabiri. *Translation is not enough: Comparing Lexicon-based methods for sentiment analysis in Persian*. in *2017 International Symposium on Computer Science and Software Engineering Conference (CSSE)*. 2017.
36. TextBlob. Available from: https://textblob.readthedocs.io/en/dev/.
37. Miller, G.A., *WordNet: a lexical database for English.* Commun. ACM, 1995. **38**(11): p. 39-41.
38. Nemati, S. and A.R. Naghsh-Nilchi, *An evidential data fusion method for affective music video retrieval.* Intelligent Data Analysis, 2017. **21**(2): p. 427-441.
39. Nicolaou, M.A., H. Gunes, and M. Pantic, *Continuous Prediction of Spontaneous Affect from Multiple Cues and Modalities in Valence-Arousal Space.* IEEE Transactions on Affective Computing, 2011. **2**(2): p. 92-105.
40. Nemati, S. and A.R. Naghsh-Nilchi, *Incorporating social media comments in affective video retrieval.* Journal of Information Science, 2015. **42**(4): p. 524-538.
41. Xu, Z., *Intuitionistic fuzzy aggregation operators.* IEEE Transactions on fuzzy systems, 2007. **15**(6): p. 1179-1187.
42. Basiri, M.E., A.R. Naghsh-Nilchi, and N. Ghasem-Aghaee, *Sentiment Prediction Based on Dempster-Shafer Theory of Evidence.* Mathematical Problems in Engineering, 2014. **2014**: p. 13.
43. Schouten, K. and F. Frasincar, *Survey on Aspect-Level Sentiment Analysis.* IEEE Transactions on Knowledge and Data Engineering, 2016. **28**(3): p. 813-830.
44. Shafer, G., *Dempster-shafer theory.* Encyclopedia of artificial intelligence, 1992. **1**: p. 330-331.
45. Bi, Y., *The impact of diversity on the accuracy of evidential classifier ensembles.* International Journal of Approximate Reasoning, 2012. **53**(4): p. 584-607.