

# Renewable Energy Sources for Clean Environment: Opinion Mining

Achin Jain and Vanita Jain<sup>1\*</sup>

University School of Information, Communication & Technology, GGSIPU, Delhi

<sup>1</sup>Bharati Vidyapeeth's College of Engineering, New Delhi

✉ vanita.jain@bharativedyapeeth.edu

*Received January 25, 2019; revised and accepted March 1, 2019*

**Abstract:** Use of conventional energy sources causes enormous amount of pollution leading to the problem of global warming. Most of the countries in Asia are facing the problem of pollution leading to lots of health problems. To curb these problems, renewable energy sources play a very important role. Sentiment analysis is used to analyze opinions of the user for decision making with the help of natural language processing techniques. Opinions are the sentiment or emotions about a subject that is expressed by a user. To study about the emotions of people about alternate energy sources we have carried out comparative sentiment analysis on various renewable energy sources using Twitter data. In our paper we have considered five sources namely: Solar energy, Bioenergy, Wind power, Hydro power and Geothermal energy. Data has been collected from Twitter which is approximately 20,000 tweets for each energy source amounting approximately to 100,000 data which have been considered for analysis. Eight sentiments are calculated for each renewable energy source. It has been found that the people's opinion about renewable energy sources—mainly solar and wind energy—fetches most tweets and people are more positive towards renewable energy sources for better environment.

**Key words:** Sentiment analysis, Twitter, renewable energy sources, social mining.

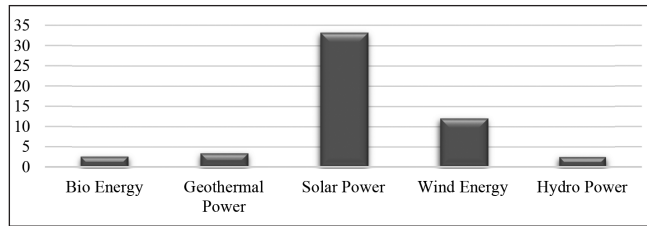
## Introduction

Conventional energy sources (coal, oil, fossil fuels etc.) drove economic progress of the world for a significant amount of time. But with the passing of time it also caused adverse effect on the environment and eventually leading to degrade in human health. This degradation of world environment leads the researchers to constantly search for alternate energy sources. In this direction, renewable energy sources proves to be the best alternative of conventional energy sources. Renewable energy sources boasted enormous potential as they'll meet over the world's energy demand. Apart from clean and green power, renewable energy sources also offer additionally low value construction of power plant. The share of

renewable energy sources has nearly doubled within the last few years, from around 8.5% in 2004 up to 16.7% in 2015 (Renewable Energy Statistics (RES), 2017). Figure 1 shows the percentage increase in the production of various renewable energy sources in year 2016. From the figure, maximum increase in production of energy is seen in solar energy source and wind power.

In this paper, the focus is regarding the sentiment of people on different renewable energy sources. By analysis of the Tweets we can find the opinion of individuals about the environment. Analysis of person's characteristics, emotions, reactions, etc. about concerning entities like products, social issues, political proceedings and their characteristics supported reaction from web sites, blogs, social network is known as Opinion Analysis or Sentiment Mining. In modern

\*Corresponding Author



**Figure 1: Percentage increase of energy production (From year 2015 to 2016).**

times people tend to express their feeling without any fear on social media platform, opinion analysis becomes essential for impact analysis and it also aids in creating decision on productive process guidelines. It is an area of research handling common strategies of sentiment finding and mining of opinions bestowed in a social media unstructured text data.

There is ample amount of text data present on the social media and web portals in the form of following data sets:

- Social network data from Twitter, Facebook etc.
- Forum discussion
- Blogs, microblogs

Every side of humans like management sciences, government, economic science and social sciences are affected by the sentiments and feelings of individual. An opinion is either negative (–ve) or positive (+ve) sentiment, feeling of term, attribute and feature (Pang et al., 2008).

In this paper, the work has been carried out as comparative sentiment analysis of various renewable energy sources. For the analysis purpose the five major renewable energy sources considered are:

1. Bioenergy
2. Solar energy
3. Wind energy
4. Hydro power
5. Geothermal energy

The paper is divided into five sections. The second section discusses about the related works carried out by researchers in renewable energy and sentiment analysis. The third section gives a brief overview of the proposed methodology which discusses about the process carried out in this paper. Experimental work and result analysis of the proposed work is carried out in the penultimate section. Conclusion of the paper is provided in the last section.

## Literature Survey

Web data analysis is used in various domains such as social network mining, sentiment analysis, prediction of movie success, behavioural analysis and lot more.

In Tayal et al. (2017) the authors have performed opinion mining on social campaign “Swachh Bharat Abhiyan”. The authors have developed a tool named “SENTI-METER” that is used to estimate the success rate of social campaigns using sentiment analysis. The paper focuses on the keyword “Swachh Bharat Abhiyan” and data was collected from Twitter. Jiliang Tang et al. (2016) worked on signed network mining in social media where signed network means positive and negative links.

Apart from just mining the opinion of the user, web data analysis can also be used in social cause. Kunz et al. (2017) have worked on active water administration in the mining sector by analyzing social network. After the analysis of social network of a mining company, the researchers found that there is communication gap which is preventing effective water management.

W. Christian Crannell et al. (2016) in their research worked on sentiments of US cancer-patients tweets. The study reveals that Twitter can be used as a tool to identify patient needs as means to measure cancer patient experience. Twitter analysis is widely used as web data analysis in various fields of life. Shane Pill et al. (2017) worked on teaching which is based on gaming. The authors in their paper collected tweets from 18 countries and perform both quantitative and qualitative analysis.

In Óskarsdóttir et al. (2016), the authors carried out the comparative study of social network classifiers for the prediction of churn in the telecommunication industry. The researchers showed that using these methods give more accurate predictions. In Manek et al. (2017) authors proposed Gini Index based feature selection method with SVM classifier for sentiment classification on very large movie review data set. The authors displayed that the proposed methodology has better classification results in terms of low error rate and accuracy. The authors have carried out a detailed survey on various sentiment analysis algorithms and applications in Medhat et al. (2014). The paper presents various classification and algorithm techniques that are applied in sentiment analysis. Social media is one of the major source of web data analysis and Injadat et al. (2016) presented a survey on various data mining techniques in social media.

## Methodology

In our work, Lexicon analysis using Unigram Approach (Taboada et al., 2011) is used. Lexicon analysis is an unsupervised approaching in which we infer the opinion carried by a piece of text on the ground of the polarity of the word (or the phrases) which compose it.

The flow of the process is shown in Figure 2. The complete process is divided into five phases. First, the extraction of the tweets for each renewable energy source type has been done. In the second phase preprocessing of the data has been done to remove the noisy and irrelevant data. After cleaning the data, tokenization process is applied to extract the meaningful words and finally sentiment, polarity of tweets are calculated.

### Tweets Extraction

The first step is extracting the renewable energy based tweets from the Twitter using the Application Programming Interface (API). The Twitter API uses three basic parameters to extract the tweets: Keyword, Language and Number of Tweets to be extracted. In the first parameter, we have used the keywords related to renewable energy sources. Second parameter was set to language English so that only the tweets written in English language are extracted from the Twitter. Third parameter was set to 2000 tweets for a given day per keyword. Pseudocode to make connection and extracting tweets from Twitter is given below:

```
Make Twitter Application and write the values of
"consumer key", "consumer secret", "access token"
and "access secret".
```

```
Setup Twitter authentication
```

```
{
```

```
    Tweet Extraction using keywords
```

```
    Write tweets to CSV file
```

```
}
```

```
Connection is closed
```

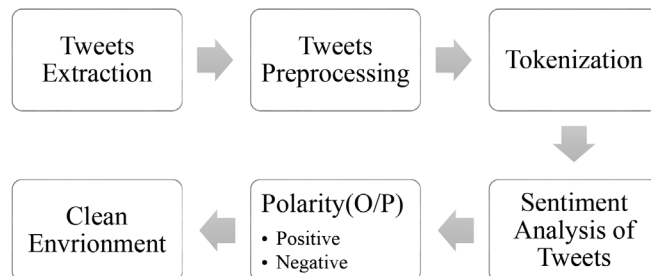


Figure 2: Proposed methodology.

### Tweets Preprocessing

It is a very important step in sentiment analysis to achieve better results. Tweets are preprocessed to remove irrelevant and noisy data. There are lot of features in a tweet that are not useful in the analysis and preprocessing is done to remove those features. First step is to convert the entire Tweet corpus in to Lower case. There are certain words in Tweet like RT (Retweet), @, punctuations (full stop(.), comma(,) etc.), numbers, links, tabs, blank spaces. All these words are removed from tweet corpus to produce clean and ready to analyze tweets. Pseudocode to preprocess the Tweets is shown below:

```
Open Tweets CSV file created in Algorithm 1
```

```
{
```

```
    Remove RT from Tweets[Retweet]
```

```
    Remove @ from Tweets[AT]
```

```
    Remove Punctuations from Tweets [Full stop(.),
```

```
    Comma (,), Question Mark(?) etc.]
```

```
    Remove Numbers [1, 2, 3,...]
```

```
    Remove Links[https]
```

```
    Remove Tabs and blank spaces
```

```
}
```

```
Save the CSV File
```

### Tokenization

In this step, complete tweet corpus is broken into words called tokens. We have created tokens from entire Tweet Corpus for each renewable energy source. These tokens are passed as input to calculate the sentiment from the tweets.

### Sentiment Analysis of Tweets

We have carried out sentiment analysis on the Twitter data for various renewable energy sources. Library used for the reference of the words is NRC sentiment dictionary. A list of words was taken from the library and was saved in R object for use. A total of eight sentiments (anger, anticipation, disgust, fear, joy, sadness, surprise and trust) are calculated for each tweet. Pseudocode to calculate sentiment of the tweets is given below:

```
Create Text Corpus of Tweet
```

```
Sentiment = get_nrc_sentiment(Tweet_Corpus_Text)
```

### Polarity of Tweets

We have calculated the polarity of tweets which is used for tweets classification which can be either positive or negative. Polarity are calculated using the get\_nrc\_sentiment() function in R code. The function returns a

total of 10 emotions out of which last two are polarity score i.e., positive and negative. Pseudocode to calculate polarity of the tweets is given below:

```
Create Text Corpus of Tweet
polarity = get_nrc_sentiment(Tweet_Corpus_Text)
polarity = polarity[,9:10]
```

### Experimental Work

For the analysis purpose, we extracted tweets using the keywords and hashtags listed below for various renewable energy sources for a time span of 15 days from 5<sup>th</sup> Sept 2017 till 20<sup>th</sup> Sept 2017.

- *Solar Power* – “Solar Power”, “Solar Energy”, “#SolarPower”
- *Wind Power* – “Wind Power”, “Wind Energy”, “#WindPower”
- *Hydro Power* – “Hydro Power”, “Hydro Energy”, “#HydroPower”
- *Bio Energy* – “Bio Energy”, “Bio Energy”, “#BioEnergy”
- *Geothermal Energy* – “GeoThermal Energy”, “Geo Thermal Energy”, “#GeothermalEnergy”

Since most of the users retweet a tweet which means lots of duplicate entries in the corpus. So, for the analysis purpose the duplicate tweets have been removed and analysis has been done only on unique tweets. Table 1 shows the details of total tweets fetched for each renewable energy source and percentage of

duplicate tweets. Figure 3 shows the total number of tweets and duplicate tweets. In sentiment analysis frequency of word is very important and to show that which word is mostly used by the user, we have plotted comparative Word Cloud of the five renewable energy sources tweet corpus. From all the words that are visible above in Figure 4, we have found top eight most frequently used words from corpus of all renewable energy sources tweets which are listed in Table 2. In sentiment analysis frequency of word is very important and to show that which word is mostly used by the user, we have plotted comparative Word Cloud of the five renewable energy sources tweet corpus.

There are around eight sentiments that are commonly found in a sentence which are – “anger”, “anticipation”, “disgust”, “fear”, “joy”, “sadness”, “surprise” and “trust”. Table 3 shows the number of tweets expressing an emotion.

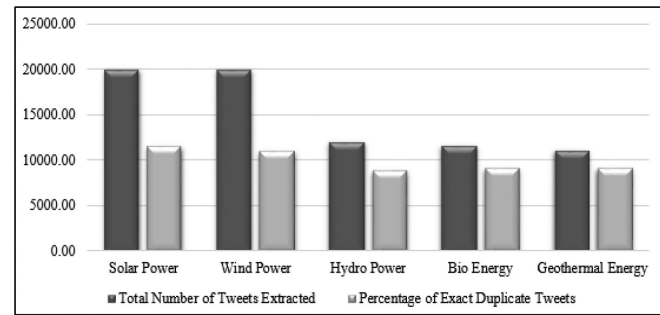


Figure 3: Details of various renewable energy sources tweets.

Table 1: Details of extracted tweets for each renewable energy source

Parameter	Solar power	Wind power	Hydro power	Bio energy	Geothermal energy
Total number of tweets extracted	20000	20000	12000	11526	11079
Number of tweets considered for analysis (After removing duplicate tweets)	8589	9165	3216	2537	2107
Percentage of exact duplicate tweets	57.05%	54.17%	73.2%	77.98%	80.98%

Table 2: Frequently used words from various renewable energy tweet corpus

Solar power		Wind power		Hydro power		Bio energy		Geothermal energy	
Word	Frequency	Word	Frequency	Word	Frequency	Word	Frequency	Word	Frequency
Solar	8721	Wind	8772	Power	2965	Bioenergy	1985	Geothermal	1996
Power	7583	Power	8644	Hydro	2772	Biofuel	369	Energy	1800
Energy	1071	Solar	1707	Project	372	Biomass	347	Power	262
Wind	903	Energy	1421	Energy	245	Energy	258	First	256
Amp	637	Nuclear	1378	Wind	244	Amp	252	Scotland	236
Panels	555	Offshore	1233	Plant	232	Solar	204	Deep	212
Grid	419	Cheaper	912	Electric	224	Plant	189	Solar	150
World	386	Amp	800	Solar	204	Digest	149	Renewable	130

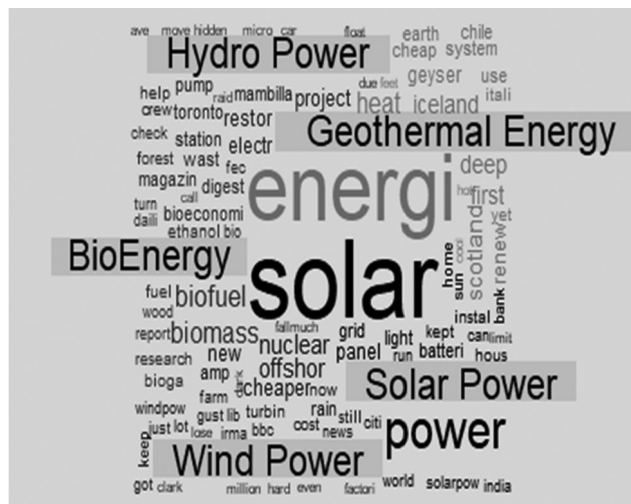


**Table 3: Number of tweets for each emotion**

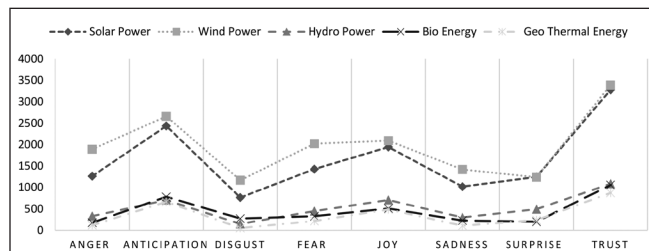
Energy Source	Anger	Anticipation	Disgust	Fear	Joy	Sadness	Surprise	Trust
Solar power	1265	2437	766	1430	1945	1020	1249	3279
Wind power	1893	2661	1170	2023	2093	1421	1241	3391
Hydro power	332	708	150	450	709	301	495	1087
Bio energy	179	783	275	330	512	229	202	1059
Geo thermal energy	119	666	55	218	488	109	239	876

The polarity of tweets has been shown in Table 4.

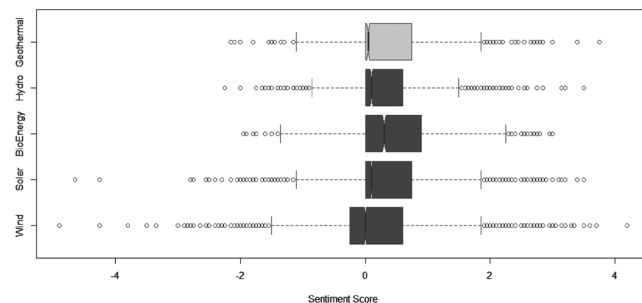
The plot of all the eight sentiments against the total number of tweets have been shown in Figure 5.



**Figure 4: Comparative WordCloud (Count) of various energy resource tweets.**



**Figure 5: Comparative sentiment score for renewable energy sources.**



**Figure 6: BoxPlot of sentiment score of renewable energy tweet.**

**Table 4. Polarity of tweets**

Energy Source	Negative (%)	Positive (%)
Solar power	29.72	70.28
Wind power	38.69	61.31
Hydro power	27.72	72.28
Bio energy	26.01	73.99
Geo thermal energy	22.59	77.41

## Conclusion

In this paper, we have conducted sentiment analysis on more than 20,000 tweets for each renewable energy source. After the analyses of renewable energy tweets, the following conclusions are made:

- Table 1 shows the details of extracted tweets and percentage of duplicate tweets. So it is clearly visible that people are more enthusiast when tweeting about solar power and wind power as the number of tweets fetched for these two sources are the highest.
- From Figure 3 we can see total number of tweets fetched for geothermal and bio energy are very less as compared to other sources and most of the tweets are duplicated also. It shows lack of interest and information among the users about this alternative source of energy.
- Figure 4 shows the comparative WordCloud that shows most frequent words used in each tweet data source such as tweets related to geothermal energy source contain words such as “heat”, “iceland” and similarly tweets related to solar power contain words such as “nuclear” and “panel”. We have presented top 10 most used words for each renewable energy source in Table 2.
- Table 3 and Figure 5 show the sentiment extracted from the tweets which clearly shows that people are tweeting with positive emotion such as anticipation, joy, surprise and trust. This leads to the conclusion that in the people’s opinion renewable energy

sources can become clean and green alternative to the conventional energy sources.

- Polarity (positive, negative) of the tweets are listed in Table 4 which shows the attitude of users for alternative source of energy. Maximum number of positive tweets are from geothermal and bio energy even though these two sources contain minimum number of tweets as shown in Figure 3. From this we conclude that people are more positive towards geothermal and bio energy sources but there is lack of communication and information among the users related to these sources.
- Figure 6 shows the boxplot of sentiment score calculated for each renewable energy source. Most of the tweets are having positive score and very few outliers can be seen having negative score.
- Overall, we conclude that people are more positive towards renewable energy sources to prevent the environment from further harm including global warming and pollution problem.

## References

- Crannell, W.C., Clark, E., Jones, C., James, T.A. and J. Moore (2016). A pattern-matched Twitter analysis of US cancer-patient sentiments. *Journal of Surgical Research*, **206(2)**: 536-542.
- Injadat, M., Salo, F. and A.B. Nassif (2016). Data mining techniques in social media: A survey. *Neurocomputing*, **214**: 654-670.
- Kunz, N.C., Kastle, T. and C.J. Moran (2017). Social network analysis reveals that communication gaps may prevent effective water management in the mining sector. *Journal of Cleaner Production*, **148**: 915-922.
- Manek, A.S., Shenoy, P.D., Mohan, M.C. and K.R. Venugopal (2017). Aspect term extraction for sentiment analysis in large movie reviews using Gini Index feature selection method and SVM classifier. *World wide web*, **20(2)**: 135-154.
- Medhat, W., Hassan, A. and H. Korashy (2014). Sentiment analysis algorithms and applications: A survey. *Ain Shams Engineering Journal*, **5(4)**: 1093-1113.
- Óskarsdóttir, M., Bravo, C., Verbeke, W., Sarraute, C., Baesens, B. and J. Vanthienen (2016). A comparative study of social network classifiers for predicting churn in the telecommunication industry. *IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, 2016: 1151-1158.
- Pang, B. and L. Lee (2008). Opinion mining and sentiment analysis. *Foundations and Trends® in Information Retrieval*, **2(1-2)**: 1-135.
- Pill, S., Harvey, S. and B. Hyndman (2017). Novel research approaches to gauge global teacher familiarity with game-based teaching in physical education: An exploratory# Twitter analysis. *Asia-Pacific Journal of Health, Sport and Physical Education*, **8(2)**: 161-178.
- Renewable Energy Statistics – Accessed from [http://ec.europa.eu/eurostat/statistics-explained/index.php/Renewable\\_energy\\_statistics](http://ec.europa.eu/eurostat/statistics-explained/index.php/Renewable_energy_statistics)
- Taboada, M., Brooke, J., Tofiloski, M., Voll, K. and M. Stede (2011). Lexicon-based methods for sentiment analysis. *Computational Linguistics*, **37(2)**: 267-307.
- Tang, J., Chang, Y., Aggarwal, C. and H. Liu (2016). A survey of signed network mining in social media. *ACM Computing Surveys (CSUR)*, **49(3)**: 42.
- Tayal, D.K. and S.K. Yadav (2017). Sentiment analysis on social campaign “Swachh Bharat Abhiyan” using unigram method. *AI & Society*, **32(4)**: 633-645.