ICCEPM 2022
The 9th International Conference on Construction Engineering and Project Management
Jun. 20-23, 2022, Las Vegas, NV, USA

Social Media Analytics to Understand the Construction Industry Sentiments

K. Joseph Shrestha [1]*, Nirajan Mani[2], Krishna P. Kisi[3], Ahmed Abdelaty[4]

[1] *Department of Engineering, Engineering Technology, and Surveying, East Tennessee State University, Johnson City, TN 37614, USA*, E-mail address: shresthak@etsu.edu
[2] *Engineering Technology Department, Fitchburg State University, 160 Pearl Street, Fitchburg, MA 01420, USA*, E-mail address: nmani@fitchburgstate.edu
[3] *Department of Engineering Technology, Texas State University, 601 University Drive San Marcos, TX 78666, USA*, E-mail address: kpkisi@txstate.edu
[4] *Department of Civil and Architectural Engineering and Construction Management, University of Wyoming, 1000 E University Ave, Laramie, WY 82071, USA*, E-mail address: aahmed3@uwyo.edu

Abstract: The use of social media to disseminate news and interact with project stakeholders is increasing over time in the construction industry. Such social media data can be analyzed to get useful insights of the industry such as demands of new housing construction and satisfaction of construction workers. However, there has been a limited attempts to analyze social media data related to the construction industry. The objective of this study is to collect and analyze construction related tweets to understand the overall sentiments of individuals and organizations about the construction industry. The study collected 87,244 tweets from April 6, 2020, to April 13, 2020, which had hashtags relevant to the construction industry. The tweets were then analyzed to evaluate its sentiments polarity (positive or negative) and sentiment intensity or scores (-1 to +1). Descriptive statistics were produced for the tweets and the sentiment scores were visualized in a scatterplot to show the trend of the sentiment scores over time. The results shows that the overall sentiment score of all the tweets was slightly positive (0.0365). Negative tweets were retweeted and marked as favorite by more users on average than the positive ones. More specifically, the tweets with negative sentiments were retweeted by 2,802 users on average compared to the tweets with positive sentiments (247 average retweet count). This study can potentially be expanded in the future to produce a real time indicator of the construction market industry such as the increased availability of construction jobs, improved wage rates, and recession.

Key words: social media, construction, tweets, information, sentiments

1. INTRODUCTION

The use of the social media platforms to disseminate information and connect with past, current, and potential customers has been increasing rapidly over time [1], [2], [3], [4]. Users are sharing their experiences and views on current news on politics, economics, businesses, and other global issues [5]. In the construction industry, companies are using various social media for recruitment, disseminating companies and projects' news, client networking, brand awareness, and showcasing innovations [1]. The information available in various social media platforms provides a tremendous opportunity to understand the opinions of users from across the globe on a specific topic or industry sector such as the construction industry. However, a large amount of social media data is produced every second by millions of users and hence analyzing such data manually to get insights on a specific topic is not practical. In such cases, Sentiment Analysis can be used to automatically extract, analyze, and summarize such user-generated data [6]. A sentiment analysis or opinion mining is “the field of study that analyzes people’s opinions, sentiments, evaluations, appraisals, and emotions towards entities, such as products, services, organizations, individuals, issues, events, topics, and their attributes" [7]. Sentiment analysis identifies emotion expressed in a natural language text and mostly focuses on analyzing polarity value of subjective sentences [8]. The polarity value is determined based on positive or negative sentiments expressed by users.

The sentiments expressed on social media can provide useful insights about the construction industry such as availability of jobs, demands of new housing construction, and status of transportation infrastructure. Despite the potential benefits, limited efforts have been made to collect and analyze construction industry related social media data. This study takes the first step to achieve this goal. The major objectives of this study are to a) collect and analyze social media data related to the construction industry and b) obtain insights on the overall trend of the social media users’ sentiments on the construction industry. More specifically, this study utilizes construction related tweets to get such insights

2. BACKGROUND

The word sentiment analysis was introduced by Das and Chen in 2001 [9] for stock market sentiment analysis. Since then, sentiment analysis has been used for various purposes such as: i) analyzing product reviews [10], ii) forecasting sales and stock markets [11], iii) analyzing product marketing or political issues [12]. Other uses of the sentiment analysis includes: i) web-page classification for content advertising [13], ii) developing intelligence system interface [14], iii) predicting movie sales [15], iv) predicting hostile or negative sources [16], v) E-rule making [17] opinions on a law before its approval, vi) classification of email on the basis of emotion like anger email, depresses email, normal email [18], and vii) visual sentiment analysis for abstraction of subjectivity in the human recognition process [19].

2.1. Types of Sentiment Analysis Techniques

Three types of sentiment analysis techniques are available: i) machine learning based approach, ii) lexicon-based approach, and iii) hybrid approach [19, 20]. In the machine learning based approach, sentiment analysis models are developed by using supervised, semi-supervised, or unsupervised techniques [6, 21]. Such technique has been used for sentiment analysis of diverse types of texts such as a microblog post and market analysis post [6, 19]. In the lexicon-based approach, manually crafted rules and lexica are used to evaluate the sentiment of a text [9]. Lexica are a set of idioms, pre-compiled sentence terms, or phrases used in a specific profession, subject, or style. A hybrid technique combines both machine learning based and lexicon-based approach [22, 23]. This study utilizes lexicon-based approach to analyze Twitter posts related to the construction industry.

3. METHODOLOGY

The overall methodology of this research is divided into four steps: a) Data Collection, b) Data Filtering, c) Sentiment Analysis, and d) Data Analysis and Visualization (Figure 1).

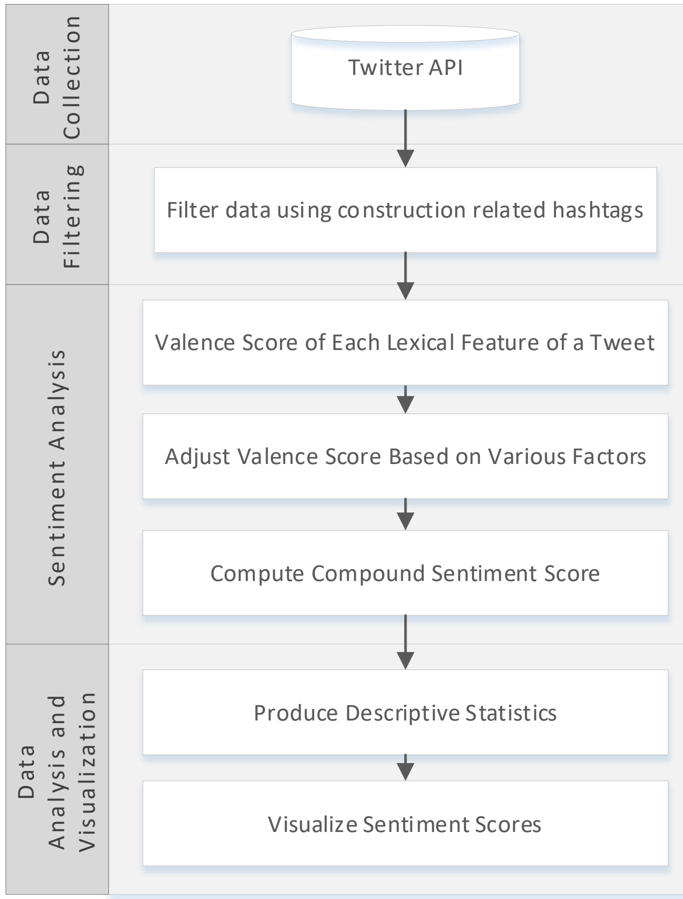

Figure 1. Methodology for the Sentiment Analysis

3.1. Data Collection

Twitter provides a public Application Programming Interface (API) to obtain tweets using various parameters, such as a specific keyword. Compared to other social media, such as Facebook, Twitter users tend to post more information publicly. As such, Twitter was used as the source of social media data for this study. Tweepy library was used to collect tweets related to the construction industry in real time using the Twitter API for 7 days continuously [24]. Some of the available data attributes for each tweet include tweet ID, tweet text, retweet count, tweet time, hashtags, and favorite count.

3.2. Data Filtering

Only tweets that contains construction related words (such as #construction, #build, #housing, and #building) is filtered and collected for further analysis. The filtered data is continuously added in a Comma Separated Values (CSV) file for the duration of the data collection period.

3.3. Sentiment Analysis

Each tweet is analyzed to evaluate whether the tweet indicates a positive or negative sentiment. Further, the analysis also quantifies the degree of positivity or negativity. For this study, the sentiment score ranges from -1 (most extremely negative) to +1 (most extremely positive). A score of zero or closed to zero indicates a neutral sentiment. This study utilizes VADER (Valance Aware Dictionary and sEntiment Reasoner) sentiment analysis tool to compute sentiment scores [25].

The VADER sentiment analysis tool is a “lexicon and rule-based sentiment analysis tool that is specifically attuned to sentiments expressed in social media and works well on texts from other domains.” The tool has been shown to outperform individual human raters at correctly classifying the sentiments of texts. The VADER algorithm is fast, performant, domain agnostic, and accurate algorithm and does not require training data. The algorithm consists of three main components: a) lexical features with Valence scores, b) adjustment of Valence scores based on a several factors, and c) generating a compound score.

The VADER algorithm consists of 7,520 lexical features with valid Valence score that indicates the sentiment polarity (i.e., positive, or negative) and sentiment intensity on a scale from -4 to +4. These intensity values for the lexical features are assigned by multiple human raters while developing the algorithm. For example, the word “okay” has a positive valence of 0.9 while the word “great” has a positive valence of 3.1 based on multiple human raters. The lexical features are based on well-established sentiment word-banks such as Linguistic Inquiry and Word Count (LIWC) and Affective Norms for English Words (ANEW). In addition to those, emoticons (e.g., “:-)”), sentiment-related acronyms and initialisms (e.g., “LOL”), and commonly used slangs with sentiment values (e.g., “meh”) are also added to the list. Other lexical features that had zero average rating or had standard variation of 2.5 or more when rated by human raters were removed from the list.

Several factors such as punctuation mark (e.g., exclamation mark), capitalization, and degree modifier (e.g., “extremely”), contrastive conjunction (e.g., “but”), and trigrams (that flips the polarity) are used to adjust the Valence score of the words.

Using the lexical features, an overall score of a text is computed by adding the valence scores of each word in the lexicon and are adjusted for several factors described above. The overall score is then normalized to be between -1 to +1 to get the final compound score. Although the score of 0 is the best value for neutral sentiment, typically the sentiment scores higher than -0.05 and lower than 0.05 are considered as neutral.

3.4. Data Analysis and Visualization

Descriptive statistics are calculated for the tweets collected for the study. The sentiment values of tweets also at over time are plotted in a scatter plot using matplotlib [26].

4. DATA ANALYSIS

Data was collected from April 6, 2020, to April 13, 2020. The data included 87,244 tweets that has hashtags relevant to the construction industry. The overall average sentiment score for all the tweets was 0.0365. Table 1 below provides the descriptive analysis of the collected tweets and their sentiment. For this study, tweets with sentiment values less than 0.05 and more than -0.05 was considered as neutral as a perfectly zero value of the sentiment score does not occur frequently. Overall, there are more tweets with positive sentiments than the negative.

Table 1. Descriptive Analysis of the Tweets

Sentiment Category	Count	Percentage	Average Score
Positive Sentiment (>= 0.05)	34,986	40%	0.5226
Neutral Sentiment (> -0.05 to < 0.05)	18,095	21%	0.0001
Negative Sentiment (<= -0.05)	34,163	39%	-0.4419

All	87,244	100%	0.0365

5. RESULTS AND DISCUSSION

This section presents the results of the data collected and analyzed for this study in three subsections a) sentiment trend over time, b) popularity of tweets with various levels of sentiments, and c) sample tweets that are extremely positive, neutral, or negative.

5.1. Sentiment Trend Over Time

Figure 2 shows a sentiment of construction related tweets collected for this study. It shows that the sentiments of the tweets range from close to -1 to close to +1. Overall, it has more positive tweets and/or higher degree of positive sentiments than the negative sentiments. As such the rolling mean (red line) and expanding mean (yellow line) are generally positive with the exception of the earlier part of the week.

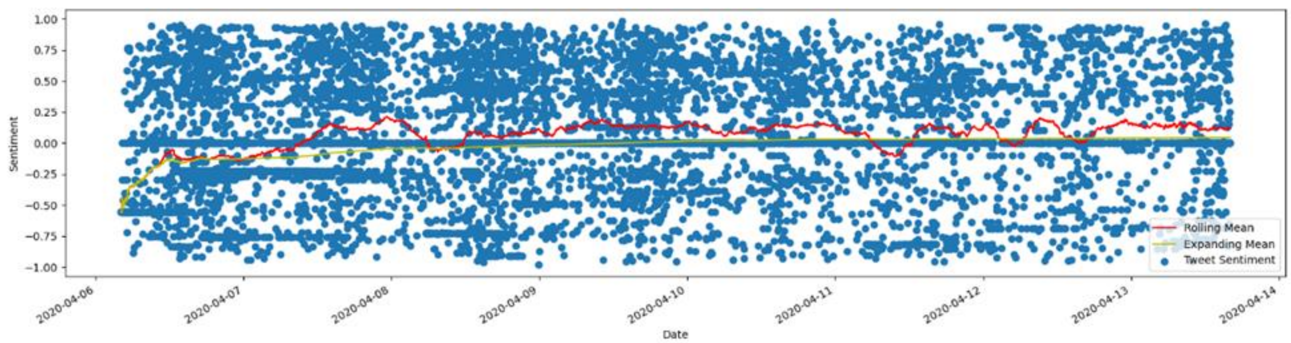

Figure 2. Sentiment of Construction Tweets Over Time

5.2. Popularity of Tweets with Various Levels of Sentiments

Table 2 shows the popularity of the sentiments quantified in terms of the retweet and favorite counts and their averages. It shows that the negative tweets are more popular with average retweet count of 2,802 compared to the positive tweets (average retweet count of 247). Similarly, the negative tweets are marked as favorite by 3.18 users on average compared to 2.33 for positive tweets.

Table 2. Retweets and Favorite Count Statistics

Sentiment Category	Total Retweet Count	Average Retweet Count	Total Favorite Count	Average Favorite Count
Positive Sentiment (>= 0.05)	8,658,334	247	81,629	2.33
Neutral Sentiment (> -0.05 to < 0.05)	6,795,634	376	49,702	2.75
Negative Sentiment (<= -0.05)	95,717,800	2,802	108,639	3.18
Total	111,171,768	1,274	239,970	2.75

5.2. Sample Tweets That are Extremely Positive, Neutral, or Negative

Table 3 below lists a few examples of tweets that are most extremely positive, neutral, and most extremely negative sentiments.

Table 3. Portion of tweets and corresponding sentiment score

Tweets	Sentiment Score
… I bless the construction workers the ones who put a roof over our heads the hardest workers out there AR the construction workers they're the heroes' great great grandfather…	0.9723
…WE'RE HIRING! This is a great opportunity to work for an award-winning company!! We're looking for a Quantity Surveyor to join our team, to find out more or apply for the role…	0.9594
…Here's a look at the construction of a field hospital that @USACEHQ is building at the TCF Center in Detroit, Mich., in response to #COVID19…	0.0000
…Here we explore the 5 tallest buildings currently under construction. We also uncover the 20 largest buildings currently completed…	0.0000
…CONSTRUCTION SITES ARE DEATH TRAPS! NO MASKS, and NO SOCIAL DISTANCING! BEING FORCED TO WORK OR BE FIRED! …	-0.9764
… Increased job losses under him. More violent attacks on trucks, construction companies. Crime has increased and all economic indicators show we are nowhere near recovery…	-0.9600
… A construction worker was unfortunately killed due to faulty structural integrity in the building he was working on.....	-0.9432

6. LIMITATIONS

The keyword-based tweet filtering algorithm does not ensure that the collected data are related to the construction industry. For example, the term "construction' can be used in a different context. Further, not all tweets are necessarily related to the current time (see the tweet below).

"...at first worked at Monowitz where IG Farben factory was in construction. In July 1942 he became a musician…"

The results of the social media analysis may not necessarily be representative of the actual construction market. For example, on one hand, the construction companies might be interested in tweeting more about their success rather than failures. On the others, the road users might be more inclined to post about the negative impacts of the road closure and failure of public agencies. While these biases are opposite and can possibly negate each other, they might also create an overall bias in one way or another.

Finally, some tweets were extremely popular with retweet counts as much as 26,240. As such, small number of such tweets had a significant impact on the average values. Such data may be considered as outliers by some readers. However, since they are real data, they are not removed in this study.

7. CONCLUSIONS

This study developed an algorithm to collect, filter, and process tweets that had hashtags related to the construction industry. The tweets were analyzed to compute their sentiments, and a chart was produced to visualize the results. The results show that overall sentiments related to the construction is slightly positive. Tweets with negative sentiments were shared more often with 2,802 retweet counts on average compared to the tweets with positive sentiments (247 average retweet count). Similarly, negative tweets were marked as favorite by 3.18 users on average compared to the positive tweets that are marked as favorite by 2.33 users on average.

The methodology and framework developed for this study can be extended in the future for more detailed analysis. Such detailed analysis of social media data can assist construction company owners, project managers, and policy makers in extracting valuable insights about the current status of the construction industry that can aid them to develop effective strategies for the overall growth of the industry.

The research can be extended further by collecting data for a specific location over a longer period of time. The long-term data and the sentiment of the tweets can be analyzed against other indicators of the market conditions such as a construction inflation index, construction job growth, or stock price of a specific construction company or composite index of construction companies. The indexes such as job growth are often available after the fact, but if the sentiment for the tweet has a significant correlation with such statistics, then the tweets can be analyzed in the real time and be used as an indicator of the current market conditions with little to no lags.

Sentiment analysis of data from various locations could provide other benefits as well. For example, comparison of the computed sentiment scores could provide insights such as availability of construction related jobs and working conditions for construction professionals in those locations. Negative sentiments could indicate the abundance of delayed construction projects, cost growth, disputes, litigations, inferior quality jobs, and excess road closures. Similarly, the positive scores could indicate successfully completed projects that reduces road users travel time, reduced congestion, and more jobs.

ACKNOWLEDGEMENT

The authors would like to thank Mr. Matthew Sweets for writing the Python script and collecting the data.

REFERENCES

[1] S. Azhar, J. M. Abeln, “Investigating social media applications for the construction industry”, Procedia Engineering, Elsevier, vol. 85, pp. 42-51, 2014.

[2] P. Ndlovu, P. Simbanegavi, “The use of social media platforms for business branding in the South African construction industry”, Construction Education: Live the Future, Australasian Universities Building Education Association Conference, 44th AUBEA, pp. 450 - 459, 2021.

[3] Y. Song, S. Jeong, H. Kim, “Semi-automatic construction of a named entity dictionary for entity-based sentiment analysis in social media”, Multimedia Tools and Application, vol. 76, no. 9, pp. 11319-11329, 2017.

[4] L. Razmerita, K. Kirchner, T. Nabeth, "Social media in organizations: Leveraging personal and collective knowledge processes", Journal of Organizational Computing & Electronic Commerce, vol. 24, no. 1, pp. 74-93, 2014.

[5] R. Y. K. Lau, C. Li, S. S. Y. Liao, "Social analytics: Learning fuzzy product ontologies for aspect-oriented sentiment analysis", Decision Support Systems, ScienceDirect, vol. 65, pp. 80-94, 2014. https://doi.org/10.1016/j.dss.2014.05.005

[6] N. K. Singh, D. S. Tomar, A. K. Sangaiah, "Sentiment analysis: A review and comparative analysis over social media", Journal of Ambient Intelligence and Humanized Computing, vol. 11, no. 1, pp. 97-117, 2020.

[7] A. P. Rajan, S. P. Victor, "Web sentiment analysis for scoring positive or negative words using twitter data", International Journal of Computer Applications (0975-8887), vol. 96, no. 6, 2014.

[8] L.Y.N. Tang, Y. M. Zhang, F. Dai, Y.J. Yoon, "Sentiment analysis for the construction industry: A case study of weibo in China", ASCE International Workshop on Computing in Civil Engineering, June 25-27, Seattle, Washington, 2017. https://doi.org/10.1061/9780784480823.033

[9] S. Das, M. Chen, "Yahoo! for amazon: Extracting market sentiment from stock message boards", Asia Pacific Finance Association Annual Conference (APFA), 2001.

[10] E. A. Stepanov, G. Riccardi, "Detecting general opinions from customer surveys", IEEE 11th International Conference on Data Mining Workshops, pp 115 – 122, 2011.

[11] L. C. Yu, J. L. Wu, P. C. Chang, H. S. Chu, "Using a contextual entropy model to expand emotion words and their intensity for the sentiment classification of stock market news", Knowledge Based System, ScienceDirect, vol. 41 (Supplement C), pp. 89 -972, 2013.

[12] R. Feldman, "Techniques and applications for sentiment analysis", Communications of the ACM, vol. 56, no. 4, pp. 82 – 89, 2013.

[13] X. Jin, Y. Li, T. Mah, J. Tong, "Sensitive webpage classification for content advertising", Proceedings of the 1st International Workshop on Data Mining and Audience Intelligence for Advertising, New York, NY, USA, ACM, pp. 28 – 33, 2007.

[14] H. Liu, H. Lieberman, T. Selker, "A model of textual affect sensing using real-world knowledge", Proceedings of the 8th International Conference on Intelligent User Interfaces, New York, NY, USA, pp. 125 – 132, 2003.

[15] G. Mishne, N. S. Glance, "Predicting movies sales from blogger sentiment", AAAI Spring Symposium: Computational Approaches to Analyzing Weblogs, pp. 155 – 158, 2006.

[16] A. Abbasi, H. Chen, "Affect intensity analysis of dark web forums", IEEE Intelligence and Security Informatics, New Brunswick, NJ, USA, pp. 282 – 288, 2007.

[17] C. Cardie, C. Farina, T. Bruce, "Using natural language processing to improve erulemaking: Project highlight", Proceedings of the 2006 International Conference on Digital Government Research, Digital Government Society of North America, pp. 177 – 178, 2006. https://doi.org/10.1145/1146598.1146651

[18] R. M. Carro, F. J. Ballesteros, A. Ortigosa, G. Guardiola, E. Soriano, "Angryemail? Emotion-Based Email Tool Adaptation", Proceedings of the 4th International Conference on Ambient Assisted Living and Home Care, Springer, Berlin, pp. 399 – 406, 2012.

[19] A. Balahur, J. M. Perea-Ortega, "Sentiment analysis system adaptation for multilingual processing: The case of tweets", Information Processing and Management: An International Journal, vol. 51, no. 4, pp. 547 - 556, 2015. https://dl.acm.org/doi/10.1016/j.ipm.2014.10.004

[20] A. Montejo-Rez, M. C. Daz-Galiano, F. Martinez-Santiago, L. A. Urea-Lpez, "Crowd explicit sentiment analysis", Knowledge Based System, vol. 69, Supplement C, pp. 134 - 139, 2014. https: //doi.org/10.1016/j.knosys.2014.05.007

[21] G. Yan, W. He, J. Shen, C. Tang, “A bilingual approach for conducting Chinese and English social media sentiment analysis. Computer Networks, Elsevier, vol. 75, no. B 24, pp. 491 - 503, 2014. https://doi.org/10.1016/j.comnet.2014.08.021

[22] D. Zhang, X. Hua, S. Zengcai, X. Yunfeng, “Chinese comments sentiment classification based on word2vec and svmperf”, Expert Systems with Applications, Elsevier, vol. 42, no. 4, pp. 1857 - 1863, 2015.

[23] M. Ghiassi, J. Skinner, D. Zimbra, “Twitter brand sentiment analysis: A hybrid system using n-gram analysis and dynamic artificial neural network. Expert Systems with Applications, Elsevier, vol. 40, no. 16, pages 6266 - 6282, 2013. https://doi.org/10.1016/j.eswa.2013.05.057

[24] J. Roesslein, “Tweepy: Twitter for python!” https://github.com/tweepy/tweepy , 2020.

[25] C. J. Hutto and E.E. Gilbert. “VADER: A parsimonious rule-based model for sentiment analysis of social media text.” Eighth International Conference on Weblogs and Social Media (ICWSM-14). Ann Arbor, MI, June 2014.

[26] J. D. Hunter, "Matplotlib: A 2D graphics environment", Computing in Science & Engineering, vol. 9, no. 3, pp. 90-95, 2007.