Sentiment Analysis

Gaging Opinions of Large Groups

Andrei P. Kirilenko, Luyu Wang, and Svetlana O. Stepchenkova

Learning Objectives

- Define sentiment analysis goals
- Describe variety of data for sentiment analysis
- Explain main approaches used in text sentiment analysis
- Apply sentiment analysis to tourism domain data
- Indicate popular software used for sentiment analysis

1 Introduction

"Sentiment analysis or opinion mining is the computational study of people's opinions, appraisals, attitudes, and emotions toward entities, individuals, issues, events, topics and their attributes" (Liu & Zhang, 2012, p. 215). The word "sentiment" represents peoples' feelings such as joy, sadness, anger, and similar. With the explosive popularity of social media leading to the necessity of fast processing of huge volumes of data, e.g., from customer reviews, the traditional methodologies of manual estimation of people's opinion about topics of products of interest are being increasingly replaced with the automated sentiment analysis (Liu, 2012). Consequently, the scholarship on the methodologies and practices of the computer-based sentiment analysis is in demand and exhibits fast growth. For example, the paper by Bakshi et al. (2016) on using sentiment analysis of tweets to predict changes in stock prices was cited over 10,000 times.

A. P. Kirilenko (✉) · L. Wang · S. O. Stepchenkova
Department of Tourism, Hospitality and Event Management, University of Florida, Gainesville, United States
e-mail: Andrei.kirilenko@ufl.edu; luyuw.93@ufl.edu; Svetlana.step@ufl.edu

R. Egger (ed.), *Applied Data Science in Tourism*, Tourism on the Verge,
https://doi.org/10.1007/978-3-030-88389-8_17

Sentiment analysis attempts to measure the emotional valence of the text using a one-dimensional numerical scale from positive to negative sentiment. Depending on the goal of the analysis, it can be applied to the entire document, separate sentences, or the aspects of interest (e.g., different aspects of a consumer product). In addition, the comparative sentiment analysis attempts to compare different sentiment estimates, e.g., "Bonaire is better than Aruba." Liu (2012) and Hu and Bing (2004) postulate the following essential elements of an expressed opinion: the sentiment (s), the opinion target (g), the aspect of the target on which the opinion is expressed (a), the opinion holders (those who holds the opinion) (h), and the time when the opinion is expressed (t). The opinion then can be formally written as a 5D vector (g, a, s, h, t). For example, a hotel review sentence "I hated beddings in the hotel, but liked the view" written on May 5, 2020, by user cat1967 could be expressed as two vectors (hotel, bed, negative, cat1967, 05052020) and (hotel, view, positive, cat1967, 05052020). This makes possible a variety of probes such as aspect sentiment analysis, comparative sentiment analysis, evolution of sentiment over time, and so on.

The document-level analysis is the simplest one. Its goal is finding the sentiment of an entire document; thus, it assumes that opinions expressed in an analyzed document are coming from a single person and related to a single event or product (Liu, 2020). This assumption mainly holds for the review-type documents since they are typically authored by one person and express an opinion on one product and for microblogs such as Twitter, but generally, it is too restrictive.

The sentence-level analysis is free of the abovementioned restriction and hence can be applied to many more types of documents. The drawback is that the amount of information used to determine the sentiment is much smaller compared to the document-level classification, making the problem more complex. In addition, while it is generally possible to classify documents into two classes, positive and negative, many sentences contain no sentiment. Hence, instead of the two-class classification of a document, a three-class sentence-level classification is a must. The latter drastically reduces accuracy of the classification algorithms (for comparison, see Ribeiro et al., 2016).

The sentence-level analysis, however, does not assign the sentiment to a specific target. For example, in the sentence "I liked Disneyland but driving there was terrible" the sentiment "terrible" relates only to driving experience, but not to Disneyland. Complicating the analysis, in a sentence "Bonaire diving was excellent" the sentiment "excellent" relates only to the target "diving," but not to Bonaire as a whole. The approaches aiding in finding the target or an entity of the sentiment are described in detail by Liu (2020).

The sentiment itself may be characterized by its orientation (also called polarity or valence) and intensity. In terms of orientation, the sentiment can be positive or negative, with some researchers also including neutral sentiment. The intensity can be measured using a variety of scales; however, for practical purposes Liu (2020) advises no more than five levels, with two levels frequently being adequate. For example, the sentiment of the statement "I hated beddings" could be -4 intensity on a one-dimensional scale $[-5, 5]$ due to the presence of word "hate" with negative valence. This method was accepted by the authors of the popular software

SentiStrength (Thelwall et al., 2010) that is based on a large dictionary containing word stems rated according to their sentiment scores (Thelwall, 2016).

The lexicon-based approach is based on a list of words and phrases together with their sentiment orientation and strength; this list is referred to as a sentiment (opinion) lexicon. In the most simplistic implementation, the software performs sentiment analysis by matching each word with the lexicon, thus, extracting the sentiment score. This sentiment score would then be reversed if negation words (such as "not") are present. In addition, modifiers may weaken or strengthen the sentiment: compare "suspicious person" with "deeply suspicious person" (Polanyi & Zaenen, 2006). The document's sentiment is then defined as a sum of sentiment scores for all words in the document or as two separate sums of positive and negative sentiments.

A specific problem in the lexicon-based approach is how to generate the sentiment lexicon (for detail, see Liu, 2020). The most straightforward approach is dictionary based. In this approach, a small manually collected seed set of sentiment carrying words with known orientation is used to search a dictionary in order to extract the synonyms, which in turn are used as new seeds. When no new candidate sentiment words are found, the generated list is manually cleaned. The shortcoming of this approach is, however, that the obtained sentiment list is generic and lacks the context. For example, the word "cold" in phrases "cold beer" and "cold person" carry opposite sentiment. This problem is tackled with the corpus-based approach, which applies a variety of approaches to extract sentiment from a collection of representative texts from the field of interest (a corpus). For example, provided that the corpus contains the phrase "he is a cold and greedy person" and that the sentiment of "greedy" is negative, we could conclude that the word "cold" is also negative. A better approach further enhances specificity by including the context in which the adjective "cold" is used.

Essentially, sentiment analysis is a classification problem. The dictionary-based approach is frequently described as an unsupervised classification, that is, classification performed without providing additional external information regarding classification patterns. A competing supervised classification approach is based on machine learning (Liu, 2020). Here, a sample of documents from the same domain is manually classified according to the sentiment expressed (those documents are called "labeled"). This sample is then used to train and validate a machine learning algorithm which is finally applied to the rest of the documents (which are called "unlabeled"). Notice that this approach does not require a list of sentiment carrying words or phrases. Instead, the sentiment is learned by the algorithm during the training process on a pre-processed huge dataset of representative documents. In terms of the machine learning models, many papers apply Naïve Bayes or SVM (Alpaydin, 2020). Recently, a new crop of machine learning models optimized for natural language processing are being successfully used to improve sentiment analysis process; among these models, the most visible is BERT (Bidirectional Representation for Transformers) developed by a Google team. The idea of BERT is to simplify learning process by introducing a new pre-training step which uses a model that is already pre-trained on generic texts. The results can then be fine-tuned using the field-specific data, resulting in lower training data requirements and faster training process.

The comparative analysis of two approaches typically demonstrates that the machine learning algorithms outperform the lexicon-based ones when the formers are properly trained (Hailong et al., 2014). The lexicon-based methods, however, have distinct advantage by being transparent compared to the "black-box" machine learning algorithms; they require no human- and computer-intensive model training and, therefore, are not sensitive to training quality (Ibid.). The latter point illustrates critical dependency of the machine learning approach on high-quality labeling of a sample of documents by human raters; when a model is pre-trained on documents from a somewhat different domain, the advantage of the machine learning approach disappears (Kirilenko et al., 2018). Even though the lexicon-based methods rely on generic language dictionaries and, hence, are less effective in recognizing emotions in specialized texts such as tweets, they are easier to use and more robust, which frequently make them preferable.

Recently, a new crop of semi-supervised methods has appeared that radically reduces demands of the machine learning methods by injecting the unlabeled documents into the algorithm training process (Van Engelen & Hoos, 2020). These algorithms can be applied to sentiment analysis as well (Lee et al., 2019) and make the machine learning methods more user-friendly. Finally, machine-based methods can be used to improve outcomes of the lexicon-based approach (Zhang et al., 2011).

As a final note, one area closely related to sentiment analysis is emotion detection. While sentiment can be expressed in a single "negative to neutral to positive" dimension, emotion recognition involves classification into multiple emotion classes, for example, Happiness, Sadness, Fear, Disgust, Anger and Surprise (Eckman, 1992). Some researchers experiment with lexicon-based approaches, similar to those used in sentiment analysis; for example, Mohammad and Turney (2013) developed a large multi-language emotion dictionary based on Plutchik (1980) "wheel of emotions." Nevertheless, it seems that currently emotion detection is better progressing in image and audio analysis (Gajarla & Gupta, 2015), as opposed to text analysis. Indeed, one could imagine the difficulties in recognizing the emotion in a sentence "I work mostly over Zoom nowadays," which could express happiness, sadness, or be just neutral. For this reason the emoticons and emojis are frequently used in social media to aid convening writer's emotions. In a distinct line of research, the emoticons as indicators of emotions are used to successfully train emotion recognition models (Felbo et al., 2017).

In tourism and hospitality, sentiment analysis is an emerging field. The existing reviews found only 26 (Ma et al., 2018), 24 (Alaei et al., 2019), and 68 (Jain & Pamula, 2021) articles; the latter review mostly included papers published in non-tourism journals. The most comprehensive upcoming publication by Mehraliyev et al. (2021) used a systematic search and uncovered 70 articles published in hospitality and tourism journals that used sentiment analysis up to June 2020. The main venues include *Tourism Management*, followed by the *International Journal of Hospitality Management*, *International Journal of*

Contemporary Hospitality Management and *Journal of Travel Research*. Notably, ¼ of all articles was published in the first half of 2020, indicating that the interest toward sentiment analysis in tourism scholarship is very recent. Further, the absolute majority of scholarship was focused on market intelligence, with very few papers dealing with other fields such as destination management, strategic management, or social media management. Methodologically, the majority (72%) of the papers used the lexicon approach; half of those papers employed one of the four most popular packages: SentiStrength (Thelwall et al., 2010), AFINN (Nielsen, 2011), LIWC (Pennebaker et al., 2001), or SentiWordNet (Baccianella et al., 2010). Overall, it seems that tourism and hospitality academics only recently discovered sentiment analysis and methodology is mainly based on the most accessible and widely available approaches and packages.

2 Theoretical Foundations

The problem of unearthing sentiment in texts was recognized as a distinct aspect of content analysis in the first half of the twentieth century. To differentiate on people's evaluative judgments and affective responses to stimuli (issues, topics, etc.) conveyed in texts, Osgood et al., 1957) identified three aspects of meaning: Evaluation, Potency, and Activity (EPA system) which, taken together, make three-dimensional space where the meaning of each word can be located. Evaluation dimension represents cognitive appraisals on the good-bad continuum. Potency reflects the intensity of the evaluative judgments on the strong-weak continuum. The last dimension, Activity, is represented by the active-passive pair of anchors. The EPA three-factor system was determined through a factor analysis of a large collection of semantic-differential scales and provided the foundation to the attitude research, and numerous studies supported validity of the approach (Heise 1970). Research has also found the stability of EPA structure across various cultures (Osgood 1964; Jakobovits 1966). Not only adjectives but also concepts can be tagged with the meaning along the EPA dimensions. For example, the concept of "war" would score very high on bad, strong, and active dimensions, while the word "baby" would likely score as highly positive, highly weak, and somewhat passive.

Currently, a large amount of works on sentiment analysis involves determining valence, which can be roughly equated with the evaluative EPA dimension; that is, where the sentiment is identified as good/bad; positive/negative or favorable/unfavorable (e.g., Pang and Lee 2008; Liu 2015). Valence, arousal, and dominance are the three dimensions of the Russell's (1980) core affect framework for study of emotions, where valence is associated with pleasure and is also placed on the positive/negative scale. For example, joy is considered as carrying positive valence and, thus, indicates positive sentiment, while anger is indicative of a negative sentiment. The intensity of the emotion can be measured by how far from a neutral point on the positive–negative scale it is located: e.g., wrath is judged as a stronger emotion than anger. This idea that various concepts, descriptors, and affective states have valence and, thus, can be assigned a score on a positive–negative dimension, lies at the foundation of the automated sentiment analysis (e.g., Pang & Lee, 2008).

3 Practical Demonstration

This section provides a brief explanation of the methodology steps, while detailed implementation will be covered in the case study discussed in the next section. Generally, the analysis starts with data cleaning and normalization. The goal of this step is broadly described as increasing data quality and cohesiveness. That may include the following:

- Removal of noise and artifacts such as HTML tags, pictograms, and unwanted characters.
- Tokenization and decapitalization, which breaks textual data into the atomic analysis units, for example, lower-case words.
- Stopword removal: examples include the words like "in," "of," "are," "the," and "it"; One popular list of stopwords comes from the Natural Language Toolkit (nltk.org).
- Resolving the attached words such as encountered in hashtags, e.g., "#AwesomeDay".
- Spelling and grammar correction.
- Resolving negations (e.g., "no good").
- Part-of-speech (POS) tagging with retaining the words of interest (e.g., adjectives and nouns only).
- Lemmatization or stemming. This step reduces the inflectional and derivational forms of words to a common base form, which in turn increases data cohesiveness. This step is especially important when the machine learning approach is used but may be skipped otherwise.

In no way those steps should be applied without validation. For example, multiple recommendations of a popular tourist guide Mr. Luck or Ms. Grim may dramatically skew distribution of park visitors' sentiment. Spell checking reviews of Manuel Antonio National Park may replace "Manuel" for "manual." As a solution, customization of data normalization algorithms is a must.

When the rule-based lexicon approach is used, the next step includes matching the tokens with one of the sentiment or emotion dictionaries, as discussed in Software section. The machine learning approach will include manual processing of a sample of documents classifying them according to expressed emotions. To improve reliability, it is recommended to attract multiple raters. The classified ("labeled") data then are used to train a classifier such as Naïve Bayes, SVM, or many others, followed by algorithm validation. Finally, the trained algorithm is used to process the unlabeled data.

During the final step, the outcomes are validated, analyzed, and interpreted. The following two sections present a case study demonstrating how those steps are realized in practice.

4 Research Case 1: Lexicon-Based Sentiment Analysis[1]

In this section, we demonstrate how the lexicon-based sentiment analysis is used to understand the sentiment expressed by visitors to Manuel Antonio National Park, Costa Rica. Manuel Antonio is the smallest Costa Rica national park (land area 6.8 km^2) famous for its beaches, wildlife viewing opportunities, beauty of landscapes, and hiking opportunities. Owing to the park's proximity to the national capital (130 km), the park is visited by 150,000 tourists annually, making it the busiest park in the country (govisitcostarica.com).

The following case study shows how sentiment analysis was applied to TripAdvisor data to measure the polarity of tourists' reviews covering personal opinions and real travel events. To demonstrate both approaches covered in this article, this section covers both the supervised feature-based machine learning and the rule-based lexicon approaches. The data includes 2700 TripAdvisor park reviews from February 2016 to September 2020 in all languages. All non-English reviews were translated to English by Goggle Cloud Translate. Then, reviews were normalized following the steps discussed in the previous section.

The scope of the project did not allow us to do the manual classification of sentiment reflected in customer reviews as required by the machine learning approach; hence, the decision was made to use the lexicon-based approach. Specifically, two widely used lexicon methods, SentiWordnet and VADER (Valence Aware Dictionary for Sentiment Reasoning) were applied to extract tourists' sentiment about the park. Both methods are based on opinion (sentiment) lexicons which contain the words with positive sentiment such as happy or enjoyable and negative sentiment such as terrible or bad. The sentiment is then defined by mapping the text into the respective lexicon (Al-Shabi, 2020). SentiWordNet (Baccianella et al., 2010) is based on the WordNet (wordnet.princeton.edu) lexical database of English language (Bonta & Janardhan, 2019). The algorithm assigns each text three scores: objectivity, positivity, and negativity, which range from 0 to 1.

As opposed to SentiWordNet, optimized for texts written in general English, VADER (Valence Aware Dictionary for Sentiment Reasoning) is specifically optimized for microblogs (Gilbert & Hutto, 2014). For each review, Vader generates four sentiment scores: text neutrality, positivity, negativity score, and a compound summary score. The compound score ranges between -1 for the most negative sentiment and 1 for the most positive. A typical sentence with positive sentiment

[1] The sentiment analysis code used in this article is publicly available at https://github.com/luyuwang1993/Sentiment-Analysis/tree/dev-sentiment

Table 1 Sentiment analysis validation for SentiWordNet and Vader algorithms

	SentiWordNet	Vader
Accuracy	0.681	0.681
Precision	0.711	0.710
Recall	0.872	0.990
F1 measure	0.783	0.827

would have a compound score greater than 0.05, and a negative sentiment sentence would have a compound score lesser than −0.05.

In order to validate sentiment predictions, we manually labeled 300 reviews (Table 1). Notice a slightly better performance of VADER; this is to be expected since this algorithm is optimized for social media as opposed to SentiWordNet, which would be preferable for texts written in standard English. Also, notice multiple metrics used for performance evaluation; the data distribution and intended application of the sentiment analysis indicate which metrics is the most useful. In our case, the sentiments were highly imbalanced with many more positive reviews than the negative ones, which makes F1 measure a preferable indicator of model quality. Another good choice of classification quality is Cohen's kappa.

Finally, the reviews carrying negative sentiment were manually processed to find the main topics of dissatisfaction shared by park visitors. The analysis revealed five shared areas of complaint: overcrowding, unprofessional staff, trail condition, opportunistic locals selling parking tickets, and monkeys thieving personal belongings.

5 Research Case 2: Machine Learning Sentiment Analysis

In this section, we demonstrate how the machine learning sentiment analysis is used to understand the sentiment expressed by the airline travelers. The dataset[2] represents scraped Twitter data representing six US airlines, subsequently processed by volunteers who classified the tweets into three categories: positive, negative, and neutral, together with the volunteer's confidence score. For this case study, we selected only the tweets with 0.6 or better confidence scores, which removed 1.6% of tweets. Together, that constituted 14,402 airline reviews.

The reviews were pre-processed as described in the How-to section and then vectorized using the Term Frequency-Inverse Document Frequency (TF-IDF) metrics (Liu, 2020). Further, the data was split into training and testing sets with 90% of tweets used for training and 10% reserved for testing. Three models, Bernoulli and Multinomial Naïve Bayes and SVM, were trained on the training dataset; then, models were validated on the testing dataset.

[2]https://www.kaggle.com/crowdflower/twitter-airline-sentiment

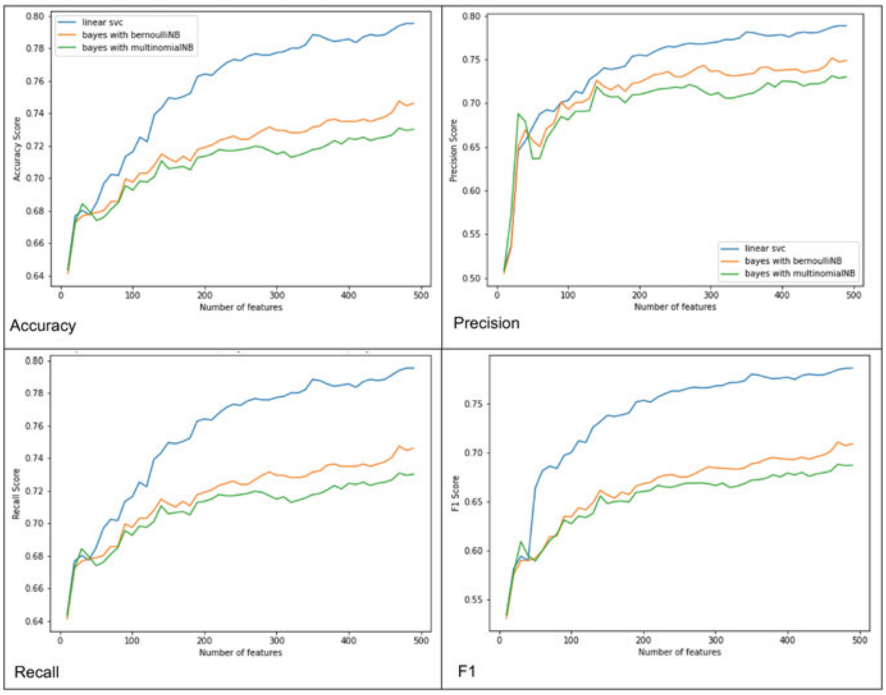

Fig. 1 Performance comparison of algorithms

An important step in the machine learning approach is feature selection. The higher the number of features (e.g., words) selected for model training, the better model predictions on the training data are. However, model performance on the testing dataset follows the bell shape and is reduced when the number of features is too high ("model overfitting"). In addition, a large number of features negatively affect model complexity, requiring expensive computer resources. A performance comparison of models utilizing a progressively increasing number of features (Fig. 1) was used to make decision on the optimal number of features. Notice that after the initial fast growth the performance curve eventually flattens down as more and more features (words) are taken into account by the machine learning algorithm. Hence, the decision was made to limit the number of features at $N = 300$.

Similar to the lexicon-based approach, the final decision on satisfactory model implementation was made based on the analysis of multiple indicators of model performance on an independent dataset selected for model testing (Table 2). Similarly to the Manuel Antonio park, the dataset is highly unbalanced: while the natural park reviews are predominantly positive, the airline reviews are predominantly negative. In our case, 61% of the tweets were negative, 22% neutral, and only 17% positive, which makes F1 measure preferable for judging model performance. Overall, SVM model was selected over two other.

Table 2 Overall performance of all tested algorithms (at 300 features for machine learning approaches)

	Bernoulli NB	Multinomial NB	SVM
Accuracy	72.94%	71.48%	77.72%
Precision	73.63%	70.92%	76.90%
Recall	72.94%	71.48%	77.72%
F1 measure	68.47%	66.65%	76.85%

Service Section

Main Application Fields: Computational study of people's emotions, attitudes, and opinions, usually expressed in a written text. In tourism, the primary area of application is the analysis of visitors' reviews of the hotels, destinations, points of interest, and similar.

Limitations and Pitfalls: Uncritical use of the computational sentiment analysis without deep understanding of the methods results in unwarranted predictions. For the lexicon-based approach, the dictionary used by the algorithm and the analyzed data much originate from similar domains (e.g., social media). For the machine learning approach, manual classification of a sample of data from same domain is a must. Both approaches require accurate validation on an independent manually classified dataset using multiple performance indices; the latter should account for data distribution and the purpose of analysis.

Similar Methods and Methods to Combine with: The sentiment analysis is frequently used together with content analysis and share many approaches and methods.

Code: The Python code is available at: https://github.com/DataScience-in-Tourism/Chapter-17-Sentiment-Analysis

Further Readings and Other Sources

Books: "Sentiment Analysis: Mining Opinions, Sentiments, and Emotions" by Bing Liu (2020) is a good introductory text covering all important aspects of computational analysis of sentiment and emotions as well as the most popular algorithmic approaches and major developments in the field.

Videos: "Sentiment Analysis: extracting emotion through machine learning" by Andy Kim. A 10-minutes TED talk introducing sentiment analysis. https://www.youtube.com/watch?v=n4L5hHFcGVk

Web sites: Medium.com, towardsdatascience.com, and KDnuggets.com sites have an excellent set of AI articles including those covering sentiment analysis.

References

Alpaydin, E. (2020). *Introduction to machine learning*. MIT Press.
Alaei, A. R., Becken, S., & Stantic, B. (2019). Sentiment analysis in tourism: Capitalizing on big data. *Journal of Travel Research, 58*(2), 175–191.
Al-Shabi, M. A. (2020). Evaluating the performance of the most important lexicons used to sentiment analysis and opinions mining. *IJCSNS, 20*(1), 1.
Baccianella, S., Esuli, A., & Sebastiani, F. (2010, May). Sentiwordnet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining. In *Lrec* (Vol. 10, no. 2010, pp. 2200–2204).
Bakshi, R. K., Kaur, N., Kaur, R., & Kaur, G. (2016, March). Opinion mining and sentiment analysis. In *2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom)* (pp. 452–455). IEEE.
Bonta, V., & Janardhan, N. K. N. (2019). A comprehensive study on lexicon based approaches for sentiment analysis. *Asian Journal of Computer Science and Technology, 8*(S2), 1–6.
Eckman, P. (1992). An argument for basic emotions. *Cognitive Emotions, 6*, 169–200.
Felbo, B., Mislove, A., Søgaard, A., Rahwan, I., & Lehmann, S. (2017). Using millions of emoji occurrences to learn any-domain representations for detecting sentiment, emotion and sarcasm. *arXiv preprint arXiv:1708.00524*.
Gajarla, V., & Gupta, A. (2015). *Emotion detection and sentiment analysis of images*. Georgia Institute of Technology.
Gilbert, C. H. E., & Hutto, E. (2014, June). Vader: A parsimonious rule-based model for sentiment analysis of social media text. In *Eighth International Conference on Weblogs and Social Media (ICWSM-14)* (Vol. 81, p. 82). *Available at (20/04/16) http://comp.social.gatech.edu/papers/icwsm14.vader.hutto.pdf*.
Hailong, Z., Wenyan, G., & Bo, J. (2014, September). Machine learning and lexicon based methods for sentiment classification: A survey. In *2014 11th web information system and application conference* (pp. 262–265). IEEE.
Heise, D. R. (1970). The semantic differential and attitude research. *Attitude Measurement*, 235–253.
Hu, M., & Bing, L. (2004). Mining and summarizing customer reviews. In *Proceedings of ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD-2004)*.
Jain, P. K., & Pamula, R. (2021). A systematic literature review on machine learning applications for consumer sentiment analysis using online reviews. *Computer Science Review, 41*. Available at: https://arxiv.org/pdf/2008.10282.pdf
Jakobovits, L. A. (1966). Comparative psycholinguistics in the study of cultures. *International Journal of Psychology, 1*(1), 15–37.
Kirilenko, A. P., Stepchenkova, S. O., Kim, H., & Li, X. (2018). Automated sentiment analysis in tourism: Comparison of approaches. *Journal of Travel Research, 57*(8), 1012–1025.
Lee, V. L. S., Gan, K. H., Tan, T. P., & Abdullah, R. (2019). Semi-supervised learning for sentiment classification using small number of labeled data. *Procedia Computer Science, 161*, 577–584.
Liu, B. (2012). Sentiment analysis and opinion mining. *Synthesis Lectures on Human Language Technologies, 5*(1), 1–167.
Liu, B. (2015). *Sentiment analysis: Mining opinions, sentiments, and emotions*. Cambridge University Press.
Liu, B. (2020). *Sentiment analysis: Mining opinions, sentiments, and emotions*. Cambridge University Press.
Liu, B., & Zhang, L. (2012). A survey of opinion mining and sentiment analysis. In *Mining text data* (pp. 415–463). Springer.
International Journal of Contemporary Hospitality Management, In second review.
Ma, E., Cheng, M., & Hsiao, A. (2018). Sentiment analysis – A review and agenda for future research in hospitality contexts. *International Journal of Contemporary Hospitality Management, 30*(11), 3287–3308.

Mohammad, S. M., & Turney, P. D. (2013). Crowdsourcing a word–emotion association lexicon. *Computational Intelligence, 29*(3), 436–465.

Mehraliyev, F., Chan, I. C. C., & Kirilenko, A. P. (2021). Sentiment analysis in hospitality and tourism: A thematic and methodological review. *International Journal of Contemporary Hospitality Management*.

Nielsen, F. Å. (2011). A new ANEW: Evaluation of a word list for sentiment analysis in microblogs. *arXiv preprint arXiv:1103.2903*.

Osgood, C. E. (1964). Semantic differential technique in the comparative study of cultures. *American Anthropologist, 66*(3), 171–200.

Osgood, C. E., Suci, G. J., & Tannenbaum, P. H. (1957). *The measurement of meaning* (No. 47). University of Illinois Press.

Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. *Found Trends Inf Retr, 2*(1–2), 1–135.

Pennebaker, J. W., Francis, M. E., & Booth, R. J. (2001). Linguistic inquiry and word count: LIWC 2001. *Mahway: Lawrence Erlbaum Associates, 71*(2001), 2001.

Plutchik, R. (1980). A general psychoevolutionary theory of emotion. In R. Plutchik & H. Kellerman (Eds.), *Emotion: Theory, research and experience, theories of emotion* (Vol. v. 1, pp. 3–33). Academic Press.

Polanyi, L., & Zaenen, A. (2006). Contextual valence shifters. In *Computing attitude and affect in text: Theory and applications* (pp. 1–10). Springer.

Ribeiro, F. N., Araújo, M., Gonçalves, P., Gonçalves, M. A., & Benevenuto, F. (2016). Sentibench – a benchmark comparison of state-of-the-practice sentiment analysis methods. *EPJ Data Science, 5*(1), 1–29.

Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology, 39*(6), 1161–1178.

Thelwall, M., Buckley, K., Paltoglou, G., Cai, D., & Kappas, A. (2010). Sentiment strength detection in short informal text. *Journal of the American Society for Information Science and Technology, 61*(12), 2544–2558.

Thelwall, M. (2016). *Sentiment analysis for small and big data. The SAGE handbook of online research methods* (pp. 344–355).

Van Engelen, J. E., & Hoos, H. H. (2020). A survey on semi-supervised learning. *Machine Learning, 109*(2), 373–440.

Zhang, L., Ghosh, R., Dekhil, M., Hsu, M., & Liu, B. (2011). *Combining lexicon-based and learning-based methods for twitter sentiment analysis*. HP Laboratories, Technical Report HPL-2011, 89.