

REVIEW OF SENTIMENT ANALYSIS STUDIES USING SOCIAL MEDIA DATA

Kalpana S Kumaran¹

Amit Dass²

¹. Associate Professor, ITM Business School, Navi Mumbai, India

²Student, ITM Business School, Navi Mumbai, India

¹ kalpanas@itm.edu

² 2022.amitdass@itm.edu

Abstract—This paper provides a summary and an examination of the analyses and findings from various studies on sentiment analysis using data from social media. People express their views, critics, feelings, opinions as well as provide feedback in the form of writings, photos, videos, and audio. These data are transformed into meaningful information using sentiment analysis and make better decisions based on the output. Studies published between 2018 and 2023 from eminent and reliable databases like Science Direct, Scopus, Emerald Insight, Web of Science, Springer, and IEEE Xplore were taken into consideration for the review of this research. The study's primary focus areas were global news, business, politics, and healthcare. After the extensive screening of 86 papers 33 papers were chosen for the review process. From each article the objectives, trends, values, concepts, techniques used, and conclusions were considered as important components to review the papers and to map the relevance between them from the papers. Most of the articles used lexicon-based approach then machine learning approach irrespective of the topics of the articles and the platforms used for the data collection were Twitter, Facebook, Reddit, Instagram, News forum blogs and community portal.

Keywords—Sentiment analysis, social media, Lexicon-based, Machine learning.

I. INTRODUCTION:

Consumer behavior when looking for information about products and services has evolved dramatically. Customer's worth to a business is determined not just by their financial impact but also by how satisfied they are. In either an online or offline setting, happy consumers spread good word of mouth, while unhappy ones do the opposite. Their opinions influence a company's reputation, which is one of the important factors to consider when selecting a good, a service, or an organization. Questionnaires are the common method of obtaining consumer feedback. However, the manner in which users share their ideas on social media creates new possibilities and makes it simpler for businesses to reach a broader audience, get interesting data, and continuously monitor their brand. Social media platforms have had a huge impact on how people engage, communicate, and share information. As a result, consumer behavior has

altered. Consumers surf numerous websites and examine the material and user-generated feedback before making a purchasing decision. Users can freely express their opinions, describe their consumer experiences, describe problems they have encountered and how those problems were resolved, and highlight the features of products or services they are satisfied or dissatisfied with in these contents—texts that users voluntarily post on the Internet and make it publicly available. Future consumer's attitudes and purchasing behavior are greatly influenced by these materials. The influence these sources have on customer behavior has been supported by numerous studies[1]–[4].

Companies can gather crucial information for advancing their businesses by doing an adequate analysis of the online content that people produce, and particularly by examining the emotions concealed therein. Sentimental analysis is one method for uncovering information regarding attitudes, feelings, and views that are concealed in the text [4], [5].

II. LITERATURE REVIEW

A. Types of sentimental analysis

- Fine-grained sentiment analysis: This is one of the simplest and most popular methods for determining the attitudes of the customers. Positive, neutral, and negative categories that are easily accessible are employed to analyze the sentiments. Another way to scale consumer feedback is to offer a range of 1 to 5[6], [7].

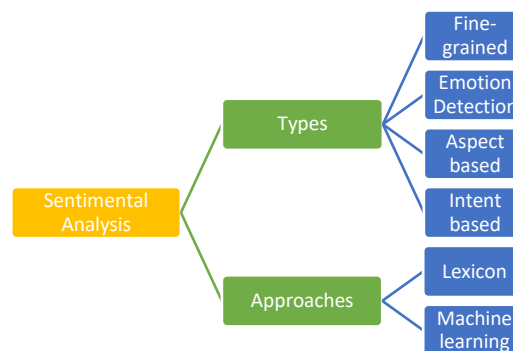


Figure 1: Sentimental Analysis types and approaches[4], [8]

- Emotion detection sentiment analysis: This is a more accurate technique for identifying the emotion in a text. This type of analysis helps to identify and comprehend people's emotions. Anger, sadness, joy, frustration, fear, panic, and concern are a few examples of possible emotions[6].
- Aspect-based analysis: This kind of sentiment analysis is primarily concentrated on the features of a specific good or service. Aspect-based sentiment analysis is crucial because it enables businesses to automatically classify and analyze client data. By automating jobs like customer care, we can acquire important insights instantly[6].
- Intent-based sentiment analysis: The automatic classification of textual material based on the goal of the client is known as intent classification. A natural intent classifier can break down the writings and reports into intents like "purchase," "downgrade," "unsubscribe," and other categories. This helps to comprehend the purposes behind a lot of the client's inquiries, automates processes, and gains valuable experience[6].

B. Lexicon approach

Lexicons are collections of tokens, each of which has a predetermined score indicating the text's neutral, positive, or

Tokens are given a score depending on their polarity, such as +1, 0, or 1 for positive, neutral, or negative, or the score may be determined by the degree of polarity, with values ranging from $[-1, 1]$, where +1 denotes a highly positive state and -1 a strongly negative state. For a particular review or text, the Lexicon Based Approach performs the aggregation of scores for each token, i.e., positive, negative, and neutral values are summed independently. The text is given an overall polarity in the final stage based on the highest value of the individual scores. As a result, the document is first broken into single-word tokens, after which the polarity of each token is determined and then totaled [3], [8]–[16].

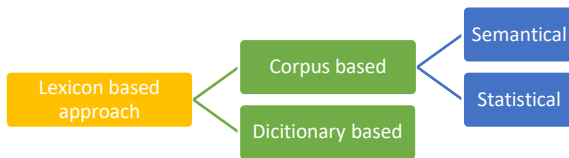


Figure 2: Lexicon based approach types classification[8]

- Corpus-Based Approach: The corpus-based approach begins with a seed list of opinion words and then mines a big corpus of words for additional concepts to elicit opinions from particular angles. In order to identify additional words, including their orientations, the initial step was to construct a seed list and use it with a variety of language limitations. The following is an illustration of how we apply the

corpus-based method using the statistical approach and the semantic approach [4], [6], [16], [17].

- a) Statistical approach: It is utilized in numerous applications that relate to sentimental analysis. The most well-known of them is the runs test, a statistical evaluation of randomization that can reveal review tampering [8], [10]
 - b) Semantic approach: It assigns values to feelings while using more than just a single premise to determine how closely related distinct words are to one another. Supporting the sentiment value in the words and words close forms the foundation of this principle [8], [10].
- Dictionary-Based Approach: The dictionary-based method gave a thorough plan for the dictionary-based approach. An exclusive selection of terms with well-established trends is used in this well-known tactic. Once we have found all possible synonyms and antonyms for this group of words, we can plant them. Following the discovery of new words, the seed list is updated, and the subsequent repetition starts. Once no new words are found, the repeating process comes to an end [8], [10].

B. Machine Learning approach

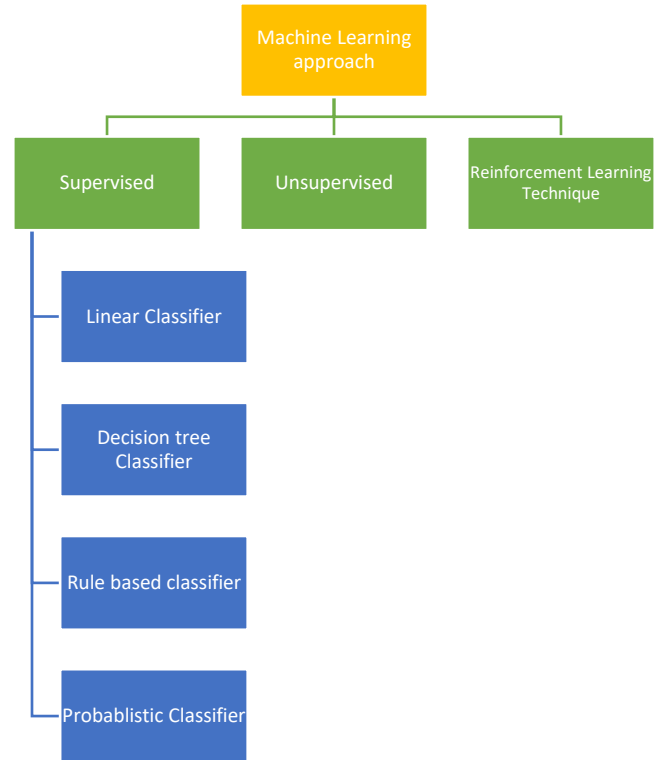


Figure 3: Classification of Machine Learning approaches[17], [18] negative tone

The classification of people's emotions has been dealt with using a variety of machine learning techniques. In contrast to lexicon-based approaches, sentiment analysis based on machine learning approaches trains the models with emotive cues in the text to make them capable of automatically recognizing emotions.[3]

Machine Learning approach is divided into Reinforcement Learning, Unsupervised Learning, and Supervised Learning[18].

- Reinforcement Learning Technique: Its totality demonstrates how to use an important method that differs from its counterpart, unsupervised learning, to make the best possible choice. To demonstrate the significance and prominence of the reinforcement learning strategy, this method places a strong priority on increasing the effectiveness of text classification[18], [19].
 - Unsupervised Approach: It is a distinct kind of machine learning technique and is frequently used to infer varied groupings of data using input data that do not include any labeled replies. When labeled training documents are not available, this method is employed[18], [19].
 - Supervised Learning: It is a type of machine learning approach that uses a data set called training data set to make predictions. These data set contain input data as well as response values. In supervised learning methods, it makes use of a large number of assorted training documents[18], [19].
- A. Probabilistic Classifiers: It classifies data using a variety of models. There are different kinds of mixture models, and each one needs to include integrated mixing components. This method is known as a generative classifier since each type of this mixture functions as generative and can support the specific taking term for one component or another [9].[1], [2], [8], [18]–[24]
- a) Maximum Entropy Classifier: A common classification method used in NLP, speech, data, and problem solving is the maximum entropy classifier. For a range of natural language problems, including language modeling, part-of-speech tagging, and text segmentation, maximum entropy, also known as probability distribution estimation, is a significant and well-known technique. Maximum entropy's fundamental tenet is the absence of external knowledge.
- b) Bayesian Network classifier: The Bayesian network classifier bases its classification on a number of variables, each of which contains a finite set of mutual examples. It is unaffected by characteristics. The actual premise is an inference that aims to have all the properties that are totally dependent. This

undoubtedly results in a particular Bayesian network model that is a guided graph and symbolizes a random contract.

- c) Naïve Bayes: The most widely used method for classifying texts recently is naive Bayes. The Nave Bayes classifier model divides the words in the adopted document into categories to get the back probability of the class.
- B. Rule-based Classification: Any classification technique that builds the classification based on IF and THEN rules is referred known as rule-based classification. A vector of real-valued numerical input features makes up a linear classifier [9].[1], [2], [8], [18]–[24].
- C. Linear classifier: It is based on the importance of a linear combination of attributes. The qualities of an object are also known as feature values and are often supplied to the computer in a vector called a feature vector. There are two linear classifier methods [9].[1], [2], [8], [18]–[24]:
- a) Neural Network: A neural network is a series of algorithms that recognize the connections between various sets of data in a manner akin to how the human brain thinks.
- b) Support vector machine: The SVM machine learning method is used to examine datasets for classification and regression analysis. It leads to the automatic processing of data.
- D. Decision Tree Classifiers (DTC): The DTC uses multi-valued attributes, data features, and class label discrete prediction to form a class label discrete prediction. Its goals are to split huge data into tiny groups for easy handling. It is a widely used method in the sentiment analysis industry and is reasonably straightforward [9].[1], [2], [8], [18]–[24].

III.METHODOLOGY AND ANALYSIS

Text is gathered from Kaggle dataset on “Twitter Sentiment Analysis - EDA and ML/DL” where we compared the positive, negative and neutral sentiment reason of 6 airlines ('Delta', 'United', 'Southwest', 'US Airways', 'Virgin America', 'American'). We have used MS-Excel to derive our interpretation.

Airline	Sentiment	Count	Total Tweet	Negative reason %
Delta	Positive	544	2222	42%
	Neutral	723		
	Negative	955		
United	Positive	492	3822	68%
	Neutral	697		
	Negative	2633		
Southwest	Positive	570	2420	48%
	Neutral	664		
	Negative	1186		
US Airways	Positive	269	2913	76%
	Neutral	381		
	Negative	2263		
Virgin America	Positive	152	504	35%
	Neutral	171		
	Negative	181		
American	Positive	336	2759	70%
	Neutral	463		
	Negative	1960		

Table 1: Data table signifying the sentiment and total tweet counts for the 6 airlines

Delta Airlines: Out of 2222 customers, 42% customer feel negative

United Airlines: Out of 3822 customers, 68% customer feel negative

Southwest Airlines: Out of 2420 customers, 48% customer feel negative

US Airways: Out of 2913 customers, 76% customer feel negative

Virgin Airways: Out of 504 customers, 35% customer feel negative

American: Out of 2759 customers, 70% customer feel negative

The provided data represents customer feedback and complaints for several major airlines, including Delta, United, Southwest, US Airways, Virgin America, and American. The first notable observation is the variation in the

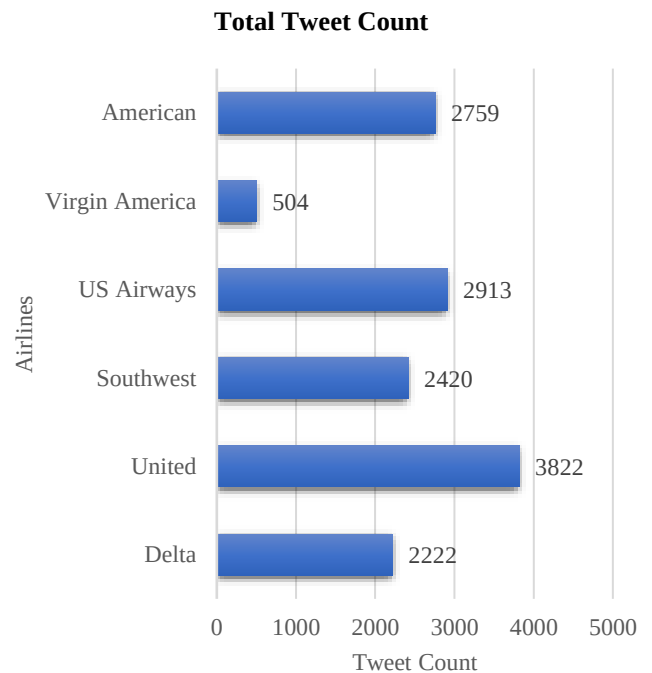


Figure 4: Visual Representation of tweet comparison

percentage of negative reasons among these airlines. US Airways and American Airlines stand out with the highest negative reasons percentage at 76% and 70%, respectively. This suggests that a significant proportion of their customers had issues and complaints about their flights. United Airlines also has a high negative reasons percentage at 68%. In contrast, Delta and Virgin America have comparatively lower negative reasons percentages at 42% and 35%, respectively, indicating a more positive customer experience.

The specific reasons for customer complaints also vary between airlines. Customer service issues, flight booking problems, and late flights are common grievances across all airlines. United Airlines appears to have the most substantial number of complaints in most categories, particularly in categories like late flights, customer service issues, and lost luggage. Southwest Airlines, on the other hand, has relatively lower negative reasons percentages compared to United and US Airways, indicating a better overall customer experience. Overall, this analysis highlights the importance of customer service, operational efficiency, and luggage handling in

shaping the perception of airlines among passengers. Airlines with a higher negative reasons percentage should consider

focusing on these areas to improve their customer satisfaction and loyalty.



Figure 5: Comparing the negative reason for different airlines

III.CONCLUSION

This work has presented a thorough overview and analysis of sentiment analysis research that have been carried out utilizing data from social media. One cannot exaggerate how crucial sentiment analysis is to comprehending consumer behavior and how it affects businesses. The unrestricted expression of customer thoughts and experiences on various social media platforms in the current digital era makes it essential for businesses to use this information to inform their decisions and uphold the reputation of their brands.

Various types of sentiment analysis, such as fine-grained sentiment analysis, emotion detection sentiment analysis, aspect-based analysis, and intent-based sentiment analysis, were highlighted in the literature study. The lexicon-based technique and the machine learning approach, the two main approaches for sentiment analysis, were also covered.

Furthermore, a practical illustration of sentiment analysis using consumer comments and grievances from significant airlines was provided in the methodology and analysis part. This investigation demonstrated the various degrees of unfavorable attitude among different airlines and pinpointed frequent consumer complaints, including poor customer service, trouble with flight reservations, and delayed

flights. The results highlight how critical it is to solve these problems in order to increase customer happiness and loyalty.

In essence, sentiment analysis is a potent tool that helps organizations better comprehend the thoughts and feelings of their clients. Companies may proactively solve issues, develop their goods and services, and ultimately strengthen their brand image in the cutthroat industry by utilizing sentiment analysis methodologies.

IV.REFERENCES

[1]–[33]

[1] C. A. Iglesias and A. Moreno, “Sentiment Analysis for Social Media.” [Online]. Available: www.mdpi.com/journal/applsci

[2] F. Neri, C. Aliprandi, F. Capecci, M. Cuadros, and T. By, “Sentiment analysis on social media,” in *Proceedings of the 2012 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, ASONAM 2012*, 2012, pp. 919–926. doi: 10.1109/ASONAM.2012.164.

[3] Q. A. Xu, V. Chang, and C. Jayne, “A systematic review of social media-based sentiment analysis: Emerging trends and challenges,” *Decision Analytics Journal*, vol. 3, p. 100073, Jun. 2022, doi: 10.1016/j.dajour.2022.100073.

[4] O. Grljević and Z. Bošnjak, “Olivera Grljević et al. Sentiment Analysis of Customer Data 39 STRATEGIC MANAGEMENT,” 2018.

[5] N. Rizun and A. Revina, “Business Sentiment Analysis. Concept and Method for Perceived Anticipated Effort Identification.”

- [6] P. Sudhir and V. D. Suresh, "Comparative study of various approaches, applications and classifiers for sentiment analysis," *Global Transitions Proceedings*, vol. 2, no. 2, pp. 205–211, Nov. 2021, doi: 10.1016/j.gltp.2021.08.004.
- [7] M. Birjali, M. Kasri, and A. Beni-Hssane, "A comprehensive survey on sentiment analysis: Approaches, challenges and trends," *Knowl Based Syst*, vol. 226, Aug. 2021, doi: 10.1016/j.knosys.2021.107134.
- [8] M. Wankhade, A. C. S. Rao, and C. Kulkarni, "A survey on sentiment analysis methods, applications, and challenges," *Artif Intell Rev*, vol. 55, no. 7, pp. 5731–5780, Oct. 2022, doi: 10.1007/s10462-022-10144-1.
- [9] A. A. Q. Aqlan, B. Manjula, and R. Lakshman Naik, "A study of sentiment analysis: Concepts, techniques, and challenges," in *Lecture Notes on Data Engineering and Communications Technologies*, vol. 28, Springer Science and Business Media Deutschland GmbH, 2019, pp. 147–162. doi: 10.1007/978-981-13-6459-4_16.
- [10] M. Wankhade, A. C. S. Rao, and C. Kulkarni, "A survey on sentiment analysis methods, applications, and challenges," *Artif Intell Rev*, vol. 55, no. 7, pp. 5731–5780, Oct. 2022, doi: 10.1007/s10462-022-10144-1.
- [11] S. Shayaa *et al.*, "Sentiment analysis of big data: Methods, applications, and open challenges," *IEEE Access*, vol. 6, pp. 37807–37827, Jun. 2018, doi: 10.1109/ACCESS.2018.2851311.
- [12] S. Shayaa *et al.*, "Sentiment analysis of big data: Methods, applications, and open challenges," *IEEE Access*, vol. 6, pp. 37807–37827, Jun. 2018, doi: 10.1109/ACCESS.2018.2851311.
- [13] E. O. Omuya, G. Okeyo, and M. Kimwele, "Sentiment analysis on social media tweets using dimensionality reduction and natural language processing," *Engineering Reports*, vol. 5, no. 3, Mar. 2023, doi: 10.1002/eng2.12579.
- [14] J. R. Grande-Ramírez, E. Roldán-Reyes, A. A. Aguilar-Lasserre, and U. Juárez-Martínez, "Integration of Sentiment Analysis of Social Media in the Strategic Planning Process to Generate the Balanced Scorecard," *Applied Sciences (Switzerland)*, vol. 12, no. 23, Dec. 2022, doi: 10.3390/app122312307.
- [15] Z. Drus and H. Khalid, "Sentiment analysis in social media and its application: Systematic literature review," in *Procedia Computer Science*, Elsevier B.V., 2019, pp. 707–714. doi: 10.1016/j.procs.2019.11.174.
- [16] I. El Alaoui, Y. Gahi, R. Messoussi, Y. Chaabi, A. Todoskoff, and A. Kobi, "A novel adaptable approach for sentiment analysis on big social data," *J Big Data*, vol. 5, no. 1, Dec. 2018, doi: 10.1186/s40537-018-0120-0.
- [17] H. Singh and D. Srivastava, "Sentiment Analysis Using Various Approaches: A Review." [Online]. Available: <https://ssrn.com/abstract=4157625>
- [18] P. Mehta and S. Pandya, "A Review On Sentiment Analysis Methodologies, Practices And Applications," *INTERNATIONAL JOURNAL OF SCIENTIFIC & TECHNOLOGY RESEARCH*, vol. 9, p. 2, 2020, [Online]. Available: www.ijstr.org
- [19] M. Bordoloi and S. K. Biswas, "Sentiment analysis: A survey on design framework, applications and future scopes," *Artif Intell Rev*, Nov. 2023, doi: 10.1007/s10462-023-10442-2.
- [20] K. L. Tan, C. P. Lee, and K. M. Lim, "A Survey of Sentiment Analysis: Approaches, Datasets, and Future Research," *Applied Sciences (Switzerland)*, vol. 13, no. 7. MDPI, Apr. 01, 2023. doi: 10.3390/app13074550.
- [21] J. Cui, Z. Wang, S. B. Ho, and E. Cambria, "Survey on sentiment analysis: evolution of research methods and topics," *Artif Intell Rev*, vol. 56, no. 8, pp. 8469–8510, Aug. 2023, doi: 10.1007/s10462-022-10386-z.
- [22] C. Ahmed, A. ElKorany, and E. ElSayed, "Prediction of customer's perception in social networks by integrating sentiment analysis and machine learning," *J Intell Inf Syst*, vol. 60, no. 3, pp. 829–851, Jun. 2023, doi: 10.1007/s10844-022-00756-y.
- [23] K. Mathur, P. Jain, S. Gupta, and P. Mathur, "SGVU International Journal of Convergence of Technology and Management Sentiment Analysis of Social Media Text Data using Machine Learning-A Review".
- [24] X. Huang, Y. Zhou, and Y. Du, "A Novel Bi-Dual Inference Approach for Detecting Six-Element Emotions," *Applied Sciences (Switzerland)*, vol. 13, no. 17, Sep. 2023, doi: 10.3390/app13179957.
- [25] Y. Kong, Z. Xu, and M. Mei, "Cross-Domain Sentiment Analysis Based on Feature Projection and Multi-Source Attention in IoT," *Sensors*, vol. 23, no. 16, Aug. 2023, doi: 10.3390/s23167282.
- [26] S. Sohangir, D. Wang, A. Pomeranets, and T. M. Khoshgoftaar, "Big Data: Deep Learning for financial sentiment analysis," *J Big Data*, vol. 5, no. 1, Dec. 2018, doi: 10.1186/s40537-017-0111-6.
- [27] M. Rambocas and B. G. Pacheco, "Online sentiment analysis in marketing research: a review," *Journal of Research in Interactive Marketing*, vol. 12, no. 2. Emerald Group Holdings Ltd., pp. 146–163, May 25, 2018. doi: 10.1108/JRIM-05-2017-0030.
- [28] P. C. Shilpa, R. Shereen, S. Jacob, and P. Vinod, "Sentiment analysis using deep learning," in *Proceedings of the 3rd International Conference on Intelligent Communication Technologies and Virtual Mobile Networks, ICICV 2021*, Institute of Electrical and Electronics Engineers Inc., Feb. 2021, pp. 930–937. doi: 10.1109/ICICV50876.2021.9388382.
- [29] S. Kumar, M. Gahalawat, P. P. Roy, D. P. Dogra, and B. G. Kim, "Exploring impact of age and gender on sentiment analysis using machine learning," *Electronics (Switzerland)*, vol. 9, no. 2, Feb. 2020, doi: 10.3390/electronics9020374.
- [30] M. Hamza and S. Gupta, "A COMPARISON OF SENTIMENTAL ANALYSIS ALGORITHMS ON TWITTER DATA USING MACHINE LEARNING."
- [31] K. Francisco and M. Panguila, "Sentiment Analysis on Social Media Data Using Intelligent Techniques," 2019. [Online]. Available: <http://www.irphouse.com>
- [32] V. Dragos, J. Besombes, and A. Mascaro, "Trend analysis in online data with unsupervised classification and appraisal categories," p. 10, 2020, doi: 10.23919/FUSION45008.2020.9190451i.
- [33] A. D. 'Andrea, F. Ferri, and P. Grifoni, "Approaches, Tools and Applications for Sentiment Analysis Implementation," 2015. [Online]. Available: <http://messenger.yahoo.com/features/emoticons>