

Online panel data quality: a sentiment analysis based on a deep learning approach

Youb Ibtissam^{1,3,4}, Azmani Abdallah², Hamlich Mohamed¹

¹Department of Electronic, CCPS Laboratory, ENSAM, University of Hassan II, Casablanca, Morocco

²LIST, Faculty of Science and Technology of Tangier, University Abdelmalek Essaadi, Tetouan, Morocco

³Department of Computer Science and Numerical Analysis, University of Cordoba, Cordoba, Spain

⁴CENTRE DE RECHERCHE CERIM, HESTIM Engineering and Business School, Casablanca, Morocco

Article Info

Article history:

Received Jun 29, 2022

Revised Aug 30, 2022

Accepted Sep 19, 2022

Keywords:

Deep learning

Online panel

Representativeness

Sentiment analysis

ABSTRACT

The rise of online access panels has profoundly changed the market research landscape. Often presented by their owners as very powerful tools, they nevertheless raise important scientific questions, particularly regarding the representativeness of the samples they produce and, consequently, the validity of the information they provide. In this paper, we present an innovative approach, based on deep learning and sentiment analysis techniques, to assess in real time the representativeness of an online panel sample. The idea is to measure the extent to which the opinions of an online panel converge with opinions on social networks. To validate the proposed method, we conducted a case study on the emerging discussion on coronavirus disease (COVID-19) vaccination. The results not only proved the representativeness of online panel sample, but also demonstrated the feasibility and effectiveness of our approach.

This is an open access article under the [CC BY-SA](#) license.



Corresponding Author:

Youb Ibtissam

Department of Electronics, CCPS Laboratory, ENSAM, University of Hassan II

Km. 150 Bd du Nil, Casablanca 20670, Morocco

Email: ibtissam.youb@etu.uae.ac.ma

1. INTRODUCTION

Academic consumer research is highly dependent on having participants in its surveys; typical theories of consumer behavior usually require testing on human participants [1] over time this has created a pervasive need for survey respondents. This need is even greater in the field of marketing research, as marketers constantly need answers to their questions. Traditional “one-off” sampling techniques fail to meet this need. They present many practical and theoretical challenges for researchers who depends on human participants. As an example, researchers find difficulties replicating studies [2] which makes answering questions about the evolution of the same population challenging. Traditional sampling techniques can also present challenges of a more practical nature. Researchers are often required to exert considerable effort before organizational gatekeepers agree to participate. Online panels appear to have solve many of these problems by offering a continuous type of research. Online panels are the natural evolution of consumer panels. It can be defined as a group of selected research participants who have agreed to provide information at specified intervals over an extended period of time [3].

The use of Internet panels to collect survey data is increasing because it is cost-effective enables access to large and diverse samples quickly, takes less time than traditional methods to get data back for analysis, and the standardization of data collection process makes studies easy to replicate [2]. Over time online consumer panels, have become a valuable tool for decision-making. They are widely used in many fields, including market research [1], [4]–[6], management research [2], social research [7], psychological

research [8], and election research [6]. Online panel data already represents more than half of global research revenues; they are the most adopted methods for survey data collection by the commercial sector [6]. Online panels are expected to continue to be the primary methodology for market and social research professionals [3]. Unfortunately, the growth of online panel data was not necessarily accompanied by a development in terms of research methodology. Several scholars have raised important scientific concerns with online panels [3], [8], [9], particularly regarding the representativeness of the samples they produce and, consequently, the validity of the information obtained. Since the beginning of online panel research, a typical study of panel data quality has been to compare online panel results with known benchmarks. In [4] presented a critical review of published studies using comparative approaches to address online panel quality. Typical studies involved running the same questionnaire with online panels and known benchmarks. Two key aspects were considered: comparison of point estimates and comparison of relationships between variables. The results obtained from these studies helped to improve the representativeness and quality of the panels studied. A major limitation of these studies is that they do not consider the dynamic aspect of the panel. The panel is constantly evolving and requires continuous refreshment to achieve representativeness and data quality. This requires frequent benchmarking surveys, which is a difficult and demanding task. Access to high quality statistical surveys is expensive and time consuming. To overcome the above limitations, we propose an alternative to benchmark polls: social media data. Spontaneously expressed opinions on social media such as Twitter may be more representative of real opinions than survey data. To be able to interpret the raw textual information, we will use sentiment analysis based on deep learning models to extract the polarity of sentiments from the text. The idea is to measure the extent to which opinions published on an online panel about a particular product or brand converge with opinions published on a social network such as Twitter. To validate our proposed method, we conducted a case study on the emergent discussion of coronavirus disease (COVID-19). This paper is structured: first, we will explain in detail the proposed methodology. Next section three is devoted to the case study based on the conversation on COVID-19 vaccination. Finally, we will discuss the results obtained, and indicate the limitations of the research as well as future perspectives.

2. PROPOSED APPROACH

The focus of this paper is to propose a new method to verify the representativeness of an online consumer panel. The approach is based on comparing the trend of the opinions of the population in question with the opinions of the panel population, towards a specific brand or product. This methodology will use online sentiment analysis techniques to extract valuable and important information from raw text data. In this section we will explain in detail the proposed approach. Sentiment analysis is not a single problem, but in fact a set of research problems that require tackling many natural language processing (NLP) tasks. In order to extract sentiment from raw text data, several steps are required. The generic process of our proposed method, whose center is the generic sentiment analysis process, is illustrated in Figure 1.

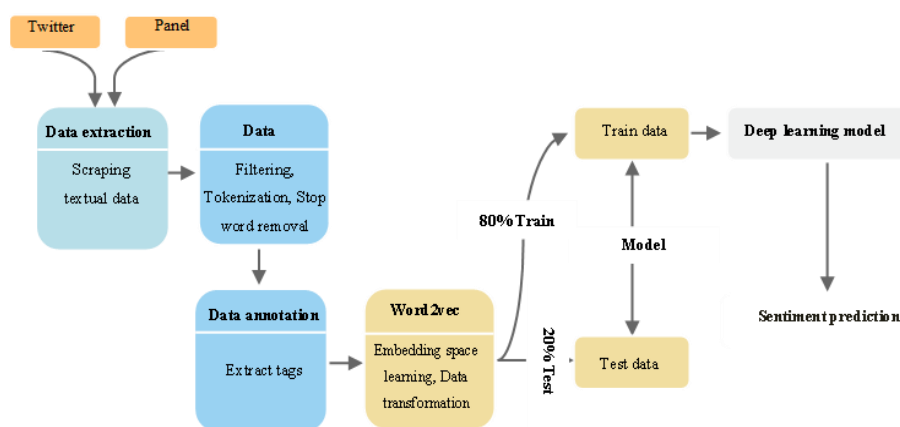


Figure 1. The general process of the proposed method

The process starts with parallelly collecting two corpora: i) Corpus1: this corpus is based on the free discussions of the target populations, and these spontaneous opinions may be more representative of reality than survey data, ii) Corpus2: this corpus is based on the opinions and reactions shared by the panelists. The collected data will then be converted to text and pre-processed using different NLP techniques; the goal of

these steps is to classify the text using deep learning algorithms into three categories: positive, negative and neutral. The details of this process are explained in the next section. The proposed methodology is then applied to the analysis of the emerging discussion on COVID-19.

3. CASE STUDY

It has been a long year of illness, devastation, grief, and despair, but the global rollout of COVID-19 vaccines has brought a sense of relief and new optimism to many. The discussion about vaccine progress, accessibility, efficacy, and side effects is ongoing, and it permeates Tweets and news every day. This discussion generates rich data that researchers can use in text mining and sentiment analysis. To implement the proposed method, we decided to conduct a case study based on this emerging discussion about vaccination against COVID-19. We will employ web-scraping techniques in order to collect in parallel the history of publications that correspond to the vaccination against COVID-19 on twitter.

3.1. Data collection

3.1.1. Corpus1

We will use Twitter as a reference of public opinion. The justification for choosing to work with Twitter is because it has a very large social network [10] and has proven its ability to predict very difficult phenomena such as stock market prediction [11], movies revenue [12], and election results [13]. Using Twitter application programming interface (API), we have collected 30024 valid tweets about vaccination from 21/03/2022 to 29/03/2022 that were published in India in the English language.

3.1.2. Corpus2

As in this project we do not work on a specific panel sample, we decided to simulate the panel simple by choosing randomly twitter accounts from India, this chosen twitter accounts will constitute our panel simple and thus. The opinions published by this twitter accounts will be the base of the Corpus2. The Corpus2 is then composed of 4029 valid tweets.

3.2. Data preprocessing

Because the two corpuses share the same characteristics, we concatenate them into one final Corpus that contains 34,053 on which we will conduct the analysis. Being published online tweeter texts are generally noisy and are characterized by their limited length and the use of informal language, hence the need to go through a step of preprocessing whose goal is to prepare text for further analysis. The process involves several tasks: i) Data cleaning: to remove noise from textual data, such as hashtags, user, http tags, emojis, and punctuation. ii) Tokenization: the focus of this step is to split text to single tokens; in the case of Unigram these tokens are simply single words. iii) Stop word removal: stop words are words frequently used in a language and do not contribute a lot in the meaning of the sentence for example in English: 'in', 'the', 'this'... iv) Lemmatization: is reducing the words derived from their original tags into their original format.

3.3. Labeling the sentiment dataset

In our study we adopted tree-level scale of tweet sentiment: positive negative and neutral. The selected 34,053 tweets were labeled following this algorithm: first, we get the annotation of Vader. Then, we get the annotation of TextBlob. If a text has received the same sentiment annotation from Vader and TextBlob, we keep the sentiment value. If the difference between the sentiment polarities given by the two algorithms is only 0.5, we keep the sentiment given by Vader. In other cases, the tweet is annotated manually. This procedure resulted in 27,000-labeled words with Vader and TextBlob, 2018 agreement with Vader only, and 890 were manually annotated. Figure 2 shows the distribution of the data according to their classes.

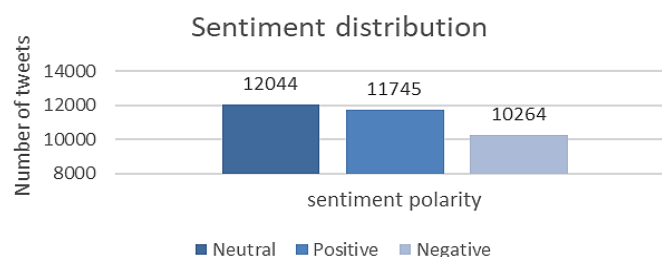


Figure 2. Distribution of data in the dataset by class

3.3.1. Word cloud

Word clouds are a simple and visually appealing method of visualizing text. They are used in a variety of contexts as a means of providing an overview by narrowing down text to the words that occur most frequently. Figures 3(a) to 3(c) shows the Word Cloud produced by words in each class. We can easily identify how the lexicon of data in a contains words such as "death" and "risk", while in positive Word Cloud we find the words "hope" and "life", in the Neutral Word Cloud we observe that the lexicon is general.

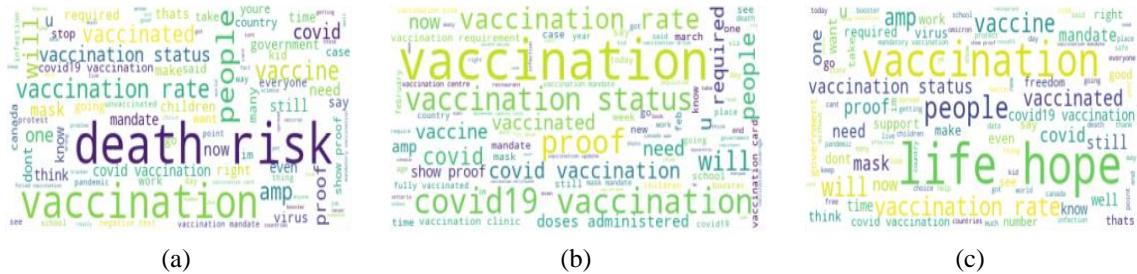


Figure 3. Word cloud of the most frequent words in the dataset, (a) the word cloud of negative tweets, (b) the word cloud of neutral tweets, and (c) the word cloud of positive tweets

3.4. Feature extraction

As mentioned before textual tweet data are very noisy, hence, to perform better sentiment analysis a fundamental step is needed: feature extraction. Feature extraction is the most important task in the text classification process [14] especially in the sentiment classification task the feature extraction technique can have a great impact of the performance of the classifier. The objective of this task is to convert the dataset from its instructed text sequences into a structured numerical space, which can then be fed into the machine learning or deep learning algorithm. The challenge of this task is to convert the text while retaining important information about the semantics of the words. In this project, we will compare two popular approaches to preparing inputs for classification algorithms: term frequency inverse document frequency, which is a weighted word technique, and glove, a word embedding based technique.

3.4.1. Term frequency-inverse document frequency

The term frequency inverse document frequency [15] is a very popular feature extraction technique; it is essentially a measure of the importance of a word in a document and in all the documents of the corpus. The fundamental equation of the weight of a term in a document is given by,

$$W(d, t) = TF(d, t) * \log \left(\frac{|D|}{df(t)} \right) \tag{1}$$

where |D| is the total number of documents, $TF(d, t)$ is the frequency of a word t in a document d, and $df(t)$ is the number of documents where the term t appears.

3.4.2. Glove word embedding

Global vector [16] is a very powerful word embedding technique. What glove does is representing a word by a high-dimensional vector trained on the context of the word. The pre-trained model of glove, which was trained on the Wikipedia Corpus, is particularly used in academic research for its high performance. The objective function of glove is,

$$f(w_i - w_j, \tilde{w}_k) = \frac{P_{ik}}{P_{jk}} \tag{2}$$

where w_i is the vector corresponding to word i , and P_{ik} is the probability that word k appears in the context of word i .

3.5. Sentiment analysis technique

Sentiment analysis as sub field of NLP, it is an active and flourishing research area, and can be applied in many domains [17], [18]. For this reason, researchers are constantly proposing evaluating and comparing new approaches with the aim of increasing the performance of sentiment analysis. Recently, deep learning has

emerged as a powerful tool for machine learning. Since it is appearing, it has produced state of the art results in many research fields [19], such as computer vision, speech recognition, and NLP. In this project, we will test two deep learning architectures for the sentiment classification task: deep neural network (DNN) and recurrent neural network (RNN).

3.5.1. Deep neural network

Deep learning [19] is the implementation of an artificial neural network for learning tasks using a multi-layer approach. A deep neural network is primarily the concatenation of more than two layers, some of which are hidden layers, it uses sophisticated mathematical modeling to process the input data. Neural networks are essentially adjustable functions of inputs that consist of several layers: an input layer that takes the input data, this layer learns the local characteristics of the input data. Hidden layers that process the data and learn features that are more complex, and a final output layer that usually uses the SoftMax function as output neurons. The Figure 4 describes a standard deep learning model.

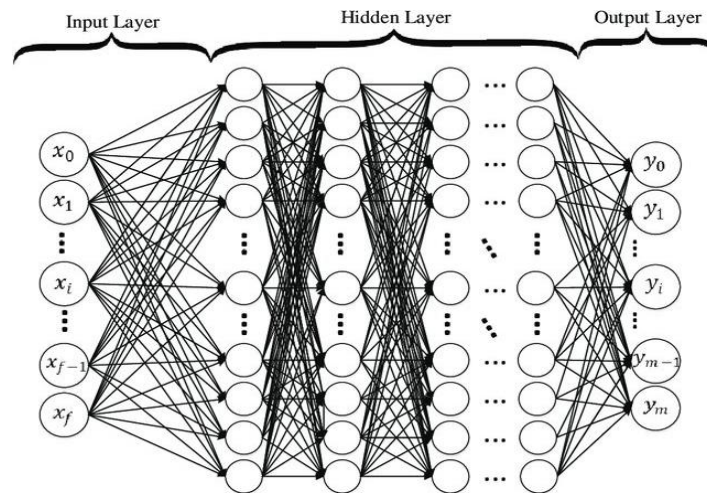


Figure 4. A fully connected DNN [14]

3.5.2. Recurrent neural network

The recurrent neural network [20] is a special deep learning architecture. The particularity of this architecture lies in the utilization of a memory cell to process a sequence of inputs. RNN could remember information about long sequences, which makes it very effective for NLP tasks. RNN is widely used to solve the sentiment classification task, and it produces highly good predictions.

3.6. Evaluation metrics

The performance of classifiers is evaluated using the accuracy metric [21]. Accuracy is the most popular metric for classification problems and is a good choice for sentiment classification. Accuracy presents the ratio of correctly predicted values to the total number of examples.

$$accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

where TP are the true positives, TN are the true negative, FP are the false positive and FN are the false negative.

3.7. Experiments results

To perform the tests, the Keras [22] and TensorFlow [23] libraries were used, the DNN and RNN architectures were implemented and tested using both term frequency inverse document frequency (TF-IDF) and pre-trained glove embedding [16]. We used 80% of the data as training data and the remaining 20% as test data. In all experiments, the parameters of our code were the same, we chose epochs=20 and batch size=128. To avoid overfitting, we adopted a 10-fold cross validation, we also kept our models very simple with only one hidden layer, and we also employed heavy dropout to render the model more generalizable [24]. Tables 1 and 2 resume the experimental results on the datasets.

Table 1. Experiments results on the dataset without removing stop words

Feature extractor	Deep learning model	Accuracy
TF-IDF	DNN	0.7973
	RNN	0.5301
Glove embedding	DNN	0.8121
	RNN	0.8340

Table 2. Experiments results on the dataset with removing stop words

Feature extractor	Deep learning model	Accuracy
TF-IDF	DNN	0.7887
	RNN	0.4574
Glove embedding	DNN	0.8016
	RNN	0.7880

3.8. Discussion

We can see that RNN is more sensitive to changing the type of feature extractor than DNN, in fact the combination of RNN and the TF-IDF weighting system gives the worst classification results (0.5301, 0.4574). This confirms the recommendation in the literature [25] to use deep learning algorithms with a word beautification feature extractor and not with a weighted word feature extractor. The combination of pretrained Glove and RNN gives satisfactory results with and without stop word removing (0.8340, 0.7880). The influence of the feature extractor technique is not very significant in the case of DNN, the usage of pretrained Glove and DNN gives average results (0.8126, 0.8016).

Experimentation on the dataset with and without stop word removal, allows us to understand the impact of this preprocessing step on the performance of deep learning algorithms. Stop word removal leads to a reduction in the performance of the sentiment model, which can be explained by the fact that stop word removal consequently eliminates important features that help the classifier understand the text better. However, the removing stop words stage significantly reduced the training time of the model.

3.9. Conclusions about the representativity of the panel

After classifying the text into three categories, using these results to address the problem of representativeness of the panel sample. Figure 5 illustrates the sentimental trend in both the panel population and the target population. In this figure we can clearly see that the opinions of the panel population are mostly consistent with the opinions of the reference population. This is a strong indicator of the representativeness of the panel sample. These results are not surprising as the panel population was randomly selected from the twitter accounts of the Indian population, so theoretically the sample chosen was a well representative sample of the population.

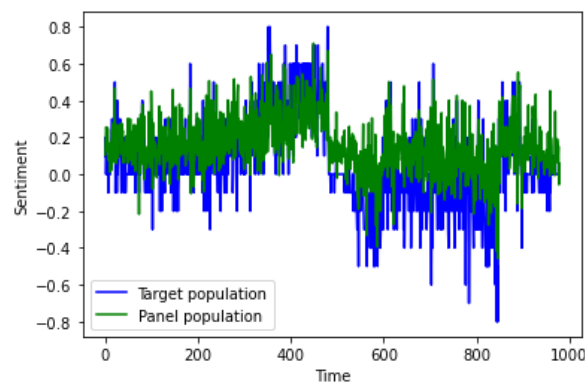


Figure 5. Evolution of the sentimental tendency towards vaccination in the two populations

4. CONCLUSION

In this paper, we addressed online panel data quality from an “opinion perspective”. Our goal was to measure how opinions published in a panel can correspond to opinions published on a large and representative social media platform. Using deep learning-based sentiment analysis, we were able to extract interpretable information from the corpora published in the panel and in twitter regarding the vaccination process. The results

of our case study show that the proposed approach is in fact feasible and very effective. Of course, this approach needs to be further developed. In fact, one of the perspectives of this project is to consider various data sources other than twitter on which our study is based, and we therefore plan to integrate other languages than English on which our study is based. The implementation of machine learning-based sentiment analysis to solve survey methodology problems, raises important questions about the possibility of using social media as an alternative to opinion surveys, and more broadly about whether big data can supplant or even replace survey data. This question is the subject of many discussions, which have been addressed, in the literature about big data and survey science. What we can add is that if big data offers a new window to understand consumer behavior, it nevertheless generates fundamental ethical concerns, mainly regarding the preservation of users' privacy. Big data does not consider the fact that users can refuse to participate in a survey, while surveyors, on the other hand, are much more ethical, since their work relies on the will of the participants. Finally, it seems that big data can provide a valuable enrichment to survey science and vice versa.

ACKNOWLEDGEMENTS

This work is fully supported by the School of Engineering and Business Sciences and Technics (HESTIM).




REFERENCES

- [1] J. K. Goodman and G. Paolacci, "Crowdsourcing consumer research," *Journal of Consumer Research*, vol. 44, no. 1, pp. 196–210, 2017, doi: 10.1093/jcr/ucx047.
- [2] C. O. L. H. Porter, R. Outlaw, J. P. Gale, and T. S. Cho, "The use of online panel data in management research: A review and recommendations," *Journal of Management*, vol. 45, no. 1, pp. 319–344, 2019, doi: 10.1177/0149206318811569.
- [3] M. Callegaro, K. L. Manfreda, and V. Vehovar, "Web survey methodology," *Revista Espanola de Investigaciones Sociologicas*, vol. 157, pp. 165–169, 2017, doi: 10.1515/jos-2016-0053.
- [4] M. Callegaro, R. Baker, J. Bethlehem, A. S. Göritz, J. A. Krosnick, and P. J. Lavrakas, "Online panel research: A data quality perspective," *Online Panel Research: A Data Quality Perspective*, pp. 1–477, 2014, doi: 10.1002/9781118763520.
- [5] S. Van Auken, "From consumer panels to big data: An overview on marketing data development," *Journal of Marketing Analytics*, vol. 3, no. 1, pp. 38–45, 2015, doi: 10.1057/jma.2015.2.
- [6] K. Brosnan, A. Kemperman, and S. Dolnicar, "Maximizing participation from online survey panel members," *International Journal of Market Research*, vol. 63, no. 4, pp. 416–435, 2021, doi: 10.1177/1470785319880704.
- [7] A. G. Blom, J. M. E. Herzing, C. Cornesse, J. W. Sakshaug, U. Krieger, and D. Bossert, "Does the recruitment of offline households increase the sample representativeness of probability-based online panels? Evidence from the German internet panel," *Social Science Computer Review*, vol. 35, no. 4, pp. 498–520, 2017, doi: 10.1177/0894439316651584.
- [8] S. L. Walter, S. E. Seibert, D. Goering, and E. H. O'Boyle, "A tale of two sample sources: Do results from online panel data and conventional data converge?," *Journal of Business and Psychology*, vol. 34, no. 4, pp. 425–452, 2019, doi: 10.1007/s10869-018-9552-y.
- [9] M. Callegaro, A. Villar, D. Yeager, and J. A. Krosnick, "A critical review of studies investigating the quality of data obtained with online panels based on probability and nonprobability samples1," *Online Panel Research: A Data Quality Perspective*, vol. 9781119941, pp. 23–53, 2014, doi: 10.1002/9781118763520.ch2.
- [10] D. Zimbra, A. Abbasi, D. Zeng, and H. Chen, "The state-of-the-art in twitter sentiment analysis: A review and benchmark evaluation," *ACM Transactions on Management Information Systems*, vol. 9, no. 2, 2018, doi: 10.1145/3185045.
- [11] V. S. Pagolu, K. N. Reddy, G. Panda, and B. Majhi, "Sentiment analysis of Twitter data for predicting stock market movements," *International Conference on Signal Processing, Communication, Power and Embedded System, SCOPES 2016 - Proceedings*, pp. 1345–1350, 2017, doi: 10.1109/SCOPES.2016.7955659.
- [12] Y. Kim, M. Kang, and S. R. Jeong, "Text mining and sentiment analysis for predicting box office success," *KSI Transactions on Internet and Information Systems*, vol. 12, no. 8, pp. 4090–4102, 2018, doi: 10.3837/tiis.2018.08.030.
- [13] B. Bansal and S. Srivastava, "On predicting elections with hybrid topic based sentiment analysis of tweets," *Procedia Computer Science*, vol. 135, pp. 346–353, 2018, doi: 10.1016/j.procs.2018.08.183.
- [14] K. Kowsari, K. J. Meimandi, M. Heidarysafa, S. Mendu, L. Barnes, and D. Brown, "Text classification algorithms: A survey," *Information (Switzerland)*, vol. 10, no. 4, 2019, doi: 10.3390/info10040150.
- [15] K. S. Jones, "A statistical interpretation of term specificity and its application in retrieval," *Journal of Documentation*, vol. 28, no. 1, pp. 11–21, 1972, doi: 10.1108/eb026526.
- [16] J. Pennington, R. Socher, and C. Manning, "Glove: Global Vectors for Word Representation," 2014, doi: 10.3115/v1/d14-1162.
- [17] M. Birjali, M. Kasri, and A. Beni-Hssane, "A comprehensive survey on sentiment analysis: Approaches, challenges and trends," *Knowledge-Based Systems*, vol. 226, 2021, doi: 10.1016/j.knsys.2021.107134.
- [18] M. Hamlich, L. Bellatreche, A. Mondal, and C. Ordonez, *Smart applications and data analysis*, vol. 1207, no. 1. Cham: Springer International Publishing, 2020.
- [19] L. Zhang, S. Wang, and B. Liu, "Deep learning for sentiment analysis: A survey," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 8, no. 4, 2018, doi: 10.1002/widm.1253.
- [20] D. P. Mandic and J. A. Chambers, "Recurrent neural networks for prediction: Learning algorithms, architectures, and stability," *John Wiley & Sons, Ltd*, 2001, doi: 10.1002/047084535x.
- [21] S. Makridakis, "Accuracy measures: Theoretical and practical concerns," *International Journal of Forecasting*, vol. 9, no. 4, pp. 527–529, 1993, doi: 10.1016/0169-2070(93)90079-3.
- [22] F. Chollet, "Keras: The Python deep learning API. 2015," URL: <https://keras.io> [accessed 2020-01-31], 2020, [Online]. Available: <https://keras.io/>.
- [23] TensorFlow, "Create production-grade machine learning models with TensorFlow," 2022, [Online]. Available: <https://www.tensorflow.org/>.




- [24] I. El Alaoui and Y. Gahi, "The impact of big data quality on sentiment analysis approaches," *Procedia Computer Science*, vol. 160, pp. 803–810, 2019, doi: 10.1016/j.procs.2019.11.007.
- [25] N. C. Dang, M. N. Moreno-García, and F. De la Prieta, "Sentiment analysis based on deep learning: A comparative study," *Electronics (Switzerland)*, vol. 9, no. 3, 2020, doi: 10.3390/electronics9030483.

BIOGRAPHIES OF AUTHORS






Youb Ibtissam    is a statistical engineer graduated from Abd el Malek Saadi University in 2021; she is pursuing a Ph.D in big data and artificial intelligence for text mining and sentiment analysis, in cotutelle between Hassan II University and the University of Cordoba. She is currently working on big data and text mining. She can be contacted at email: Yibtissam7756@gmail.com.



Dr. Abdellah Azmani    received his Ph.D in Industrial Computing at the University of Science and Technology of Lille in 1991. He worked as a professor at the Ecole Centrale de Lille and at the Institute of Computer and Industrial Engineering from Lens. He is a member of the Laboratory of Automatics and Informatics of Lille (LAIL). He is a professor at Faculty of Sciences and Technology of Tangier, Morocco. He has contributed to many scientific research. He can be contacted at email: azmani.students@gmail.com.



Dr. Mohamed Hamlich    is Professor-Researcher at ENSAM (UH2C) Casablanca, Morocco since 2014. Currently, he is director of the research laboratory: "Complex Cyber Physical Systems". He is the author of several research papers in the field of machine learning, robotics and electrical energy. He can be contacted at email: moha.hamlich@gmail.com.