Initial submission 18 July 2023, final submission 28 September 2023, accepted 23 October 2023, date of publication 1 November 2023, date of current version 8 November 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3329049

RESEARCH ARTICLE

Recognition of Opinion Leaders in Social Networks Using Text Posts' Trajectory Scoring and Users' Comments Sentiment Analysis

WIDED OUESLATI[1], SIWAR MEJRI[1], SHAHA AL-OTAIBI[2], (Member, IEEE), AND SARRA AYOUNI[2]

[1]Ecole Supérieure de Commerce de Tunis, BESTMOD Laboratory, Université de la Manouba, Manouba 2010, Tunisia
[2]Department of Information Systems, College of Computer and Information Sciences, Princess Nourah bint Abdulrahman University, Riyadh 11671, Saudi Arabia

Corresponding author: Shaha Al-Otaibi (stalotaibi@pnu.edu.sa)

This work was supported by Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia, through the Princess Nourah bint Abdulrahman University Researchers Supporting Project under Grant PNURSP2023R136.

ABSTRACT Identifying opinion leaders in social networks, particularly in social media, is a crucial marketing strategy. These individuals have a considerable influence on the purchasing decisions of their communities. Companies can benefit from collaborating with relevant opinion leaders in their market as this can increase their visibility, establish their credibility, and gain consumer trust, leading to increased sales, improved brand perception, and an expanded market share. Additionally, by gaining a comprehensive understanding of opinion leaders, companies can better comprehend the trends and preferences of their target audience. This allows them to tailor their marketing and product strategies more effectively. Identifying suitable influencers to endorse their products or services is a significant challenge for companies. The identification of opinion leaders is complicated by their informal and unstructured nature, as well as the varying selection criteria depending on the marketing campaign's goals. While numerous research studies have focused on detecting opinion leaders in social networks based on content, interactions, or a combination of both, few have explored sentiment analysis of post content, received interactions, and user comments in relation to published posts. The purpose of this paper is to present an hybrid approach to detect opinion leaders in Facebook. This approach involves analyzing the trajectory of post content by examining interactions on the post, as well as mining the text content of the post itself and analyzing the users' comments sentiments.

INDEX TERMS Opinion leader, post trajectory, social signals score, sentiment analysis, comments polarity.

I. INTRODUCTION

Detecting opinion leaders within social networks represents a recent and pressing area of research, given the significant impact of these leaders on shaping public opinion. Social networks have become an integral component of daily life, and are now crucial in a range of fields and industries. With over 2.7 billion users accessing social networks worldwide, companies have committed considerable resources and investments towards leveraging these platforms to increase their revenue and enhance their brand reputation. In addition, social networks have surpassed the influence of traditional media outlets such as television, radio, and magazines, particularly among younger consumers, thus demonstrating their unparalleled reach in disseminating information and promoting products. As mentioned above, The emergence of social networks especially social media [1] such as Facebook, Instagram, and YouTube as primary channels for disseminating information has spurred the creation of novel marketing strategies such as viral or digital marketing. As a result, social networks have become more influential than traditional communication tools, leading to their widespread utilization across diverse fields including politics and healthcare. For instance, political candidates have taken advantage of social networks during presidential campaigns as a means of establishing closer connections with potential voters. The surge in social network usage has given rise to a new phenomenon known as the "Influencer" or "Opinion

The associate editor coordinating the review of this manuscript and approving it for publication was Barbara Guidi.

© 2023 The Authors. This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License.
For more information, see https://creativecommons.org/licenses/by-nc-nd/4.0/

Leader''. The Influencer is an individual who possesses the power to sway a community to embrace an idea or purchase a product. As such, numerous corporations concentrate their efforts on identifying the most effective Opinion Leader, with a considerable following or sizeable community, to promote their products quickly and inexpensively. Traditional tools are often costly and may not always yield desired outcomes. Companies also rely on Opinion Leaders or Influencers to gain insight into user behavior, identify potential customers, and assess a product's feasibility, competitiveness, and potential for success within the market. Numerous companies actively seek out Opinion Leaders to leverage their unique advantages. In fact, the recognition of opinion leaders within social networks holds significance for a variety of reasons:

- Targeted Marketing [2]: Opinion leaders frequently specialize in specific niches or areas of expertise. Recognizing them enables businesses to enhance the precision of their marketing strategies. For instance, a fitness expert could serve as an excellent influencer for a sportswear brand.
- Influence and Persuasion [3]: Opinion leaders wield considerable sway over their followers, capable of convincing them to embrace specific viewpoints, behaviors, or products. Identifying these individuals enables marketers, advertisers, and organizations to harness their influence for promotional objectives.
- Viral Content [4]: Opinion leaders possess the ability to propel content to viral status. When they distribute or support content, it gains exposure to a wider audience, thereby enhancing its chances of spreading organically throughout the network.
- Building Trust [3]: Opinion leaders typically enjoy the trust of their followers. When they give their endorsement to a product or idea, it imparts credibility and reliability to that product or idea. This sense of trust can be instrumental in managing reputation and fostering brand loyalty.
- Feedback and Insights [5]: Opinion leaders have the capacity to offer invaluable feedback and perspectives. They frequently remain well-informed about the latest developments within their niche, rendering them a valuable wellspring of information and trend analysis.
- Crisis Handling [6], [7]: During crisis situations, opinion leaders have the potential to influence public perception positively. When they align with a company's crisis response, it can alleviate harm to the company's image.
- Content Generation [2]: Opinion leaders often produce their own content, representing a valuable resource for content marketers. Collaborating with them enables companies to co-create authentic and captivating content.
- Market Analysis [2]: The identification of opinion leaders can enhance the process of market research. Through an understanding of the opinions expressed by these leaders and the interests of their followers, companies can glean valuable insights into consumer preferences and emerging market trends.
- Competitive Edge [2]: When a business identifies opinion leaders ahead of its competitors, it has the opportunity to secure a competitive advantage by establishing partnerships or collaborations with these influential individuals.

To sum up, the identification of opinion leaders within social networks is essential for leveraging their impact, enhancing precision in marketing endeavors, fostering trust, acquiring valuable insights, and ultimately accomplishing a range of business goals, spanning from brand promotion to product adoption.

The aim of this research work is to present a novel approach for detecting Opinion Leaders within social networks by assessing the trajectories (that is composed of social signals interactions) of Facebook posts and performing sentiment analysis on user comments. In fact, existing methods do not comprehensively account for all the necessary parameters to detect opinion leaders. Typically, these methods tend to focus either on the analysis of social interactions associated with a post or on sentiment analysis of user comments on a given post. However, there is a notable gap in the literature where none of the existing approaches systematically consider the entirety of relevant parameters in this context. In practice, identifying opinion leaders in dynamic social networks requires a multifaceted approach that takes into consideration not only either social interactions or sentiment analysis but both of them at the same time. As a result, a more comprehensive and holistic methodology is needed to capture the full spectrum of opinion leadership dynamics within dynamic social networks. By integrating analysis of social interactions, post text and users comments sentiments, we can develop more accurate and robust models for opinion leader detection, providing a more nuanced understanding of the influence landscape in online communities.

The remainder of this paper is structured as follows: Section II puts the accent on the importance of detecting opinion leaders in social networks in various domains. Section III provides an overview of various approaches to identifying Opinion Leaders. In section IV, we propose a new approach to improve upon the method developed by Oueslati et al [5]. Section V presents the approach validation and section VI summarize the research work and propose new perspectives to be done in future works.

II. THE IMPORTANCE OF OPINION LEADER DETECTION IN DIFFERENT FIELDS

In the upcoming section, we will delve into the paramount importance of opinion leaders within the realm of social networks. These individuals assume a pivotal role in shaping the dissemination of information, the formulation of decisions, and the influence on behaviors across a spectrum of domains. Whether it pertains to marketing, healthcare, education, industry, agriculture, tourism, or politics, social media opinion leaders exert a substantial impact on their respective audiences. The identification and comprehension of these influential figures are quintessential for the attainment of specific objectives, the mobilization of communities, and the shaping of ever-evolving trends in our dynamic digital landscape. here are concrete examples that highlight the importance of detecting opinion leaders in various fields:

Medicine and Healthcare: Prominent doctors and leading researchers are often regarded as opinion leaders in the medical field. Their recommendations and publications can influence clinical practices, treatments, and medication choices. They played a crucial role during the COVID-19 pandemic [8] by providing accurate and credible information about vaccines, thus helping to counter misinformation and false information that circulated, urging people to get vaccinated for their protection.

Agriculture: Successful farmers who share their experiences on social media can become opinion leaders in agriculture. Their advice on farming techniques, crop varieties, and fertilizer usage can influence other farmers. In fact, the research article [9], delves into the distribution of agricultural knowledge in the southern regions of Ethiopia. This study utilizes network analysis techniques to pinpoint the authentic opinion leaders within the community who hold a central role in spreading agricultural insights. The primary objective of this research is to unveil the individuals with the most significant influence in the network of knowledge-sharing. This, in turn, can facilitate the precise and targeted distribution of agricultural information, contributing to advancements in farming practices and community development in southern Ethiopia.

Tourism: Travel bloggers and social media influencers are frequently considered opinion leaders in the tourism sector. Their recommendations on destinations, hotels, restaurants, and tourist activities can significantly impact travelers' choices [10].

Industry: In the realm of industry, experts in manufacturing, technology, or supply chain management can be opinion leaders. Their insights and guidance can shape strategic decisions for businesses [11].

Education: Distinguished educators, education researchers, and innovative educators can serve as opinion leaders in the field of education [12], [13]. Their teaching methods, recommended resources, and pedagogical ideas can influence educational practices.

Marketing: Social media influencers have become major opinion leaders in the marketing domain [14], [15]. Their product and service recommendations can influence consumer purchasing decisions [16], making them valuable partners for businesses.

In all these domains, identifying and collaborating with suitable opinion leaders can enhance credibility, visibility, and impact for an individual or organization. These opinion leaders are often well-informed, have significant influence over their target audience, and can play a pivotal role in disseminating information, promoting products or services, and driving change within their respective sectors.

III. AN OVERVIEW ON OPINION LEADER DETECTION METHODS

In this section, we aim to expound upon a range of studies that have sought to identify opinion leaders [17] within the context of social networks. Our analysis discerns three distinct categories in this research literature: namely, detection methods based on users' interactions, methods based on users' content, methods that combine both interaction and content-based features and methods based on sentiment analysis for opinion leader detection. Detecting opinion leaders in social networks has emerged as a focal point of scholarly inquiry, given their considerable influence over the opinions and behaviors of other users. As a result, discerning the identities of such individuals holds significant implications for an array of fields, including marketing, public health, and political campaigns [18]. Furthermore, a comprehensive grasp of the efficacy and limitations of various opinion leader detection methods can furnish valuable insights into the mechanics of social networks and their potential utility for diverse applications.

A. DETECTION OF THE OPINION LEADER BASED ON INTERACTIONS

In their work, Chen et al. [19] introduced an algorithm known as Social Network User Rank (SNURank), which focuses on the analysis of user attention behaviors, including commenting, point praise, and other concerns. This algorithm defines the concept of attention degree and adopts the notion of voting from the PageRank algorithm to determine a user's influence. The authors consider the attention degree as a transfer factor for distributing the influence value.

The algorithm measures the attention degree of users based on their engagement behavior, such as commenting, liking, and other activities. The algorithm uses a voting scheme, similar to the PageRank algorithm, to calculate the influence of users, taking into account their attention degree as a transfer factor of influence value distribution. The authors considered both the network topology and user attributes in their analysis, making the SNURank algorithm an effective tool for identifying opinion leaders. To evaluate the performance of SNURank, the authors compared it with the PageRank-based opinion leader mining algorithm (PR), based on the influence coverage rate of the top 20 users identified by both methods.

In their experimental setup, the authors utilized a dataset sourced from "Sina" that contained information on 63,641 users, their friendships with 1,391,718 other users, and forwarding relationships in 27,759 microblogs obtained from the Datatang platform.

In [20], authors developed a model for detecting positive opinion leader groups in a given forum's user network, which includes both explicit and implicit links. They employed the Longest Sequence Phrase Tree (LSP-Tree) method to represent the comments on the forum, which allowed them to measure the similarity between comments and detect implicit links. Based on this approach, they were able to identify positive opinion leader groups in the forum network. In their research they defined, a first definition of explicit link and implicit link by supposing that ''ci'' is published earlier than ''cj''. If ''cj'' is a reply of ''ci'', ''cj'' links to ''ci''. If ''cj'' isn't a reply of ''ci'' but has semantic similarity with ''ci'', ''cj'' is considered as having an implicit link to ''ci''. The second definition is about positive link and negative link. In fact, if ''cj'' has the same sentiment attitude with ''ci'', the link is considered as ''positive link'', otherwise as ''negative link''. The third definition is about the multi-themes user network which can be represented with a graph UNG(V, E), Where E is edge set including all links of definitions 1 and 2, and V is the vertex set that contains all users corresponding every comment. The fourth definition concerns the positive opinion leader group by ranking these users by the user score. For their experiments, they used data from Chinese website.

The researchers cited in [21] have presented a method for identifying Opinion Leaders through the analysis of node importance in multi-relationship online social networks. The authors utilized a signalling approach based on the Multi-subnet Composited Complex Networks Model, taking into account the interconnectedness of various relationship characteristics. The approach is founded on two fundamental principles.

The first principle involves a signalling process, whereby every node in a network, consisting of n nodes, is regarded as a system capable of sending, receiving, and recording signals. Thus, a node can only influence its neighbouring nodes, which, in turn, affect their respective neighbours. Through this process, the entire network is affected. Initially, a source node is selected, which possesses one unit of signal, whereas the other nodes have none. The source node then sends the signal to all of its neighbour nodes, including itself. Subsequently, the node and its neighbours receive a signal, and each node records the quantity of signals it receives. Then, every node sends an equivalent quantity of signals during the next process, causing the signalling process to repeat continually throughout the network. The second principle concerns the Signalling process in multi-subnet composited complex networks. Specifically, when the source node sends a signal concerning the relationship ''rf'' to all of its neighbour nodes, only those neighbour nodes connected to the source node, where there exists a relationship ''rf,'' can receive signals. The process of forming links among other neighbor nodes connected to a source node, where no relationship ''rf'' exists, results in the inability of said nodes to receive signals. Subsequent to the signaling process, a vector containing all relationships in the network is generated for every source node. Each vector component represents an m-tuple of all types of relationships in the network. In order to evaluate the effectiveness of their algorithm, the researchers utilized a dataset derived from a social network comprising 30,453 users. The dataset captured user relationships and comments over a period of two days. The resulting experimental network consisted of 30,453 nodes, 55,865 links, and 1,789,567 book comments that encompassed both friendship relationships and book-related comments. This experimental network is a representative example of a multi-relationship online social network. The researchers conducted experiments to compare the performance of their algorithm against other established algorithms such as the PageRank algorithm and HITS algorithm.

B. DETECTION OF OPINION LEADER BASED ON THE CONTENT

Authors in [22] proposed a Dynamic Opinion Rank algorithm that aims to identify opinion leaders based on comments from Chinese news articles. Specifically, the algorithm aims to identify the most influential comments and users across the entire network of comments and users. The study focuses on a single topic and aims to develop a unified view and a mathematical model of users based on the network topology. To achieve this, the authors defined three distinct layers: the first layer represents all the themes of news, the second layer represents the network of comments, and the third layer represents the network of single-topic users. In addition, the authors proposed three definitions to support their analysis.

The initial definition entails the assumption that score1 is greater than score2, which in turn is greater than score3. Consequently, the comments that receive the highest score are identified as the most influential ones. The second definition builds upon the first, wherein the user who garners the highest score is considered the most influential and is deemed the opinion leader among users. The comments are segmented into statements, and the ''ICTCLAS'' method is utilized to split each statement into individual words. Subsequently, words that convey emotions are extracted from a dictionary, and the number of statements containing negative words is statistically determined.

The authors subsequently delved into the modeling of the Single-Topic Network, utilizing both explicit and implicit link algorithms to establish all link relationships within the set. Their proposed approach involved employing sentiment analysis, whereby they designated any reply that demonstrated a tendency as a positive link with a value of 1, while negative links were assigned a value of 0. Importantly, the authors also incorporated the temporal factor into their analysis, assuming that the length of the time interval between reading a review or comment and replying to it would impact the sentiment expressed.

In order to assess the accuracy of their algorithm, the authors gathered data from Weibo news. By monitoring this news over a two-day period and dividing it into four distinct time periods, each of these intervals was subjected to analysis. Furthermore, the authors incorporated dynamic changes into their methodology to identify opinion leaders.

In their publication, ''Willingness to Follow Opinion Leaders on Social Media: An Empirical Study of Weibo,'' Luqiu and colleagues 2019 examined the impact of both follower count and comment tone on the identification of potential opinion leaders on Weibo, a Chinese social media platform. Specifically, the authors tested two distinct types of negative comments: "constructive but negative feedback," which they referred to as critical, and "true flaming comments." They posited two hypotheses: H1, that participants would perceive an opinion leader's account with a smaller number of followers to provide lower quality content, and H2, that participants would be less inclined to follow an opinion leader's account with a relatively low number of followers.

C. DETECTION OF THE OPINION LEADER BASED ON THE INTERACTIONS AND CONTENTS

In 2017, authors put forth a methodology to identify opinion leaders on Facebook in order to promote a sport product on a marathon page. To establish the network structure between users, the authors utilized the Degree centrality technique, which involves calculating the number of direct connections each individual has with others within the same network. To model the network, the authors constructed a matrix that classified social network actors based on their level of influence using two complementary metrics: Betweeness centrality and Eigenvector centrality. Individuals who exhibited the highest values for both metrics were classified as potential influencers. The resulting matrix was then represented as a two-dimensional scatter plot, with the actors plotted based on their Betweeness and Eigenvector centrality thresholds, which were selected using a relevant criterion. This plot categorized the actors into four groups: potential influencers with high scores in both Betweeness and Eigenvector centrality, brokers or individuals with high Betweeness centrality but low Eigenvector centrality, actors with high connections but low Betweeness centrality, indicating a limited path to reach groups outside their local community, and secondary actors. For their experiments, the authors used a Facebook page as a data source, consisting of 977 vertices or actors, including the page administrator, linked by 25,613 edges corresponding to comments. The authors evaluated the performance of their proposed model for detecting influencers using various techniques. First, they compared the average number of comments received and provided by individuals selected by the model with those of unselected individuals and the total dataset. The authors then proceeded to analyze the selected users in more detail. They found that the high number of comments both received and issued by the selected users indicates the amount of traffic they generate in their network and thus their influence. They also performed a separate analysis of each quadrant of the scatter plot and noted that the average number of comments produced and received by actors in each quadrant was higher than when using centrality measures to classify individuals independently. However, the combination of both dimensions outperformed them independently, demonstrating that the model efficiently identifies influential actors.

In their work [25], the authors proposed a method for identifying opinion leaders using sentiment analysis and Leader PageRank. They computed the score of each opinion and classified them into positive, negative, or neutral opinions. For their experiments, they collected data from an automotive Chinese forum. To classify the opinions, they employed the LDA-SVM algorithm and compared its performance with other methods. Their results showed that the SVM algorithm outperformed the other algorithms in identifying opinion leaders.

D. DETECTION OF THE OPINION LEADER BASED ON SENTIMENT ANALYSIS

The paper [26] introduces a novel approach for identifying prominent Instagram users using sentiment analysis and specific criteria, including likes, followers, keyword frequency, and post positivity. The primary goal is to analyze the sentiment of a user's content to help select the right individuals for campaigns or collaborations. This approach achieved an 85.05% accuracy rate. Another study presented in [2] highlights the importance of influencer marketing, stressing the need to choose the right influencer to enhance marketing efforts effectively. The study suggests that follower and engagement metrics may not provide a complete picture of an influencer's impact, as these numbers can be manipulated. Therefore, sentiment analysis is proposed as a tool for better influencer selection. In [27], a hybrid approach is presented that combines lexicon-based sentiment analysis with social network analysis to identify opinion leaders in social media networks. A BP neural network is used to train the emotional social media opinion leader model, resulting in improved accuracy. In the paper by [28], The authors propose a technique that combines sentiment analysis, similarity metrics, and graph representation to detect influencers in social networks. The approach considers similarity metrics to identify opinion communities and employs graphs to represent interactions within these communities. Influence value models are established for each user in the opinion community, with the user having the highest influence value designated as the opinion leader.

The table 1 presents a comparison between the aforementioned studies. Both articles [2] and [26] focused on analyzing the influence of users in social networks, specifically by considering social signals and comments. However, one crucial aspect that was overlooked in both studies is the analysis of content. In [27], the focus was primarily on analyzing user comments and their impact on others, while other metrics such as social signals and post

TABLE 1. Comparative study.

Article	Social Network	Social signals	Post content	Comments sentiment analysis
[26]	Instagram	X		X
[2]	Online platforms(Reviews)	X		X
[27]	Blogs			X
[28]	Forum		X	

content were overlooked. On the other hand, [28] primarily considered content similarity between users while neglecting the importance of social signals and comments.

To achieve a more comprehensive and accurate detection of opinion leaders, it is crucial to know the significance of the three key metrics (social signals, post content, and comments). By integrating these metrics, we can gain a deeper understanding of how users can shape opinions.

IV. IDENTIFYING OPINION LEADERS IN SOCIAL NETWORKS USING TEXT POSTS' TRAJECTORY SCORING AND USERS' COMMENTS SENTIMENT ANALYSIS

The proliferation of smart phones has made social networks a crucial part of our daily lives. This trend has led to the emergence of a new form of information sharing that facilitates the exchange of knowledge and experience not only between individuals but also between people and companies. Consequently, marketing strategies have started to leverage the power of social networks, particularly due to the immense influence they wield.

Marketing strategies focus on engaging influential users to promote products or ideas. This approach has become increasingly popular due to the potential of social networks to reach a wider audience and generate higher engagement rates. As a result, companies are actively investing in social media marketing to capitalize on the immense opportunities presented by these platforms. Recommendations on social media are powerful in product promotion, as they serve as persuasive testimonials that build trust. Opinion leaders possess the unique ability to inspire confidence, and are more effective at convincing people than commercial advertisements. This explains why most users tend to check reviews and feedback posted by community members before making any purchases. Therefore, it is vital for companies to identify opinion leaders on social media networks as part of their efforts to target potential clients. Companies used to overlook consumer feedback, as communication was mostly one-way [29]. However, this approach is no longer effective. Nowadays, companies acknowledge the importance of recruiting opinion leaders to connect with a wider audience [30], seeing their significant impact on brand promotion.

Numerous studies have aimed to identify opinion leaders on social networks. Yet, many focus solely on the network structure, neglecting temporal aspects, vast content, and diverse user perspectives. In a previous study, the Dynamic Detection of Opinion Leaders (DDOL) approach [5] was introduced. It's based on users' interactions and the counting of opinion terms in text posts. The approach focuses on users' text posts on Facebook communities, considering social signals (like, comment, and sharing) and opinion terms.

To evaluate the approach, it was compared with other methods such as Betweenness centrality and PageRank, using the precision metric and other criteria. DDOL performed well compared to other methods. However, it didn't consider user comments on a given post.

Facebook user comment sentiment analysis can be an effective tool to identify opinion leaders. It provides valuable insights for businesses, organizations, and researchers. Analyzing user comments' sentiment can help identify individuals consistently providing opinions on specific topics. Consequently, sentiment analysis can be useful in identifying opinion leaders. s an extension to DDOL [5], we propose to:

- Ameliorating the computing of the social signals score (ScoreImp).
- Computing the engagement score of the text post.
- Computing the user's comments positive score based on users comments sentiment analysis.

In the following subsections, we will present in detail the process and the different steps of the New DDOL approach (NDDOL) that allows to detect the Opinion Leader in the Social Network based on users interactions, text posts' contents and users comments sentiment analysis. The opinion leader detection process is composed of 2 phases that are the text post score computing based on the Facebook post text's trajectory and the users comments positive score computing based on sentiment analysis. Each phase is composed of a set of steps.

A. POST SCORE COMPUTING FROM TEXT POST TRAJECTORY COMPONENTS

A facebook post text has a trajectory. This latter has 4 components [31], [32] [33] that are the begin, the stops, the moves and the end [34], [35], [36]. Indeed, the begin corresponds to the creation of the post text in a facebook group or profile. While it receives social interactions such as likes, comments and shares during a continious time interval, the post text evolves and moves internally or exeternally. In fact, when the facebook post text receives likes and comments, it moves internally but if it is shared by another facebook profile, here the post text will move from a facebook profile to another. Such move is qualified of an external move. During a time interval that is no empty, the facebook post text may not receive social interactions, here we talk about the stop as element of the trajectory. After a determined period of time, the post text can be disappeared from the newsfeed or completely removed by its owner and therefore, no longer receives social interactions. This corresponds to the end of the Facebook post text whole trajectory. The following figure 1 describes the trajectory components of a Facebook post text.

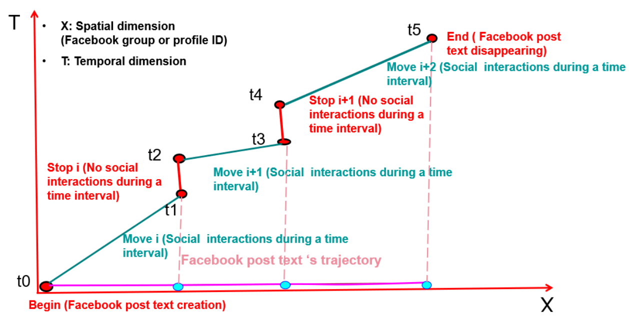


FIGURE 1. Facebook user text post trajectory.

Based on the Facebook text post trajectory, we can generate the post score. The computing of this latter requires going through different stages that are posts extraction, data filtering, interaction and engagement scores computing and post's opinion terms computing. The figure 2 describe the main steps to follow in the New DDOL approach to compute the post score. Each step will be discussed in detail in a subsection.

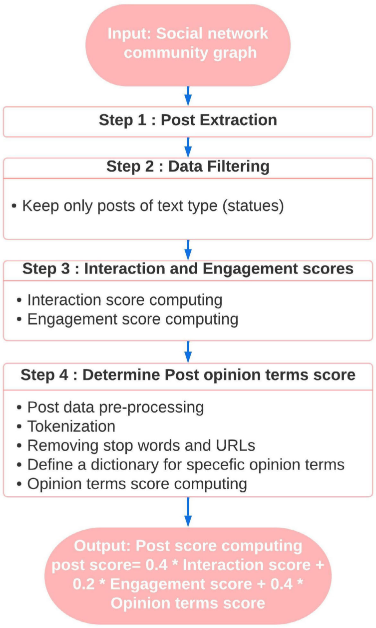


FIGURE 2. Post score computing.

1) TEXT POST EXTRACTION AND FILTERING

The first step of our approach of detecting opinion leaders from social networks is extracting posts from a public Facebook group in the french language called "LIRE, REVER et PARTAGER"[1] and use it as a dataset to detect the opinion leader. To collect data containing a set of posts, we can utilize the web crawler APIFY. This tool is a programming interface offered for developers and applications that allows for the retrieval of real-time data for various web projects and applications. It facilitates communication between multiple software programs. APIFY provides developers and users with a range of tasks, such as data processing (e.g., users, posts, comments, etc.) and downloading data. The web crawler APIFY is an essential tool for retrieving and collecting data from Facebook pages or groups in an automated manner. It provides a streamlined and efficient method for developers and users to access the information they need and analyze it for various purposes. Following the extraction step, it is common to have a large amount of extraneous data that requires filtering. This is because a post can take various forms such as video, image, URL, or text. Even in the case of text, it may not express an opinion and could be a question or statement that does not provide any relevant information. As a result, during the filtering process, we focus solely on posts that are of type text and express opinions, while disregarding other forms of posts. This helps to eliminate irrelevant data and ensure that we are only analyzing data that is useful for our research purposes.

[1] https://www.facebook.com/groups/281587007113184

2) TEXT POST INTERACTION SCORE COMPUTING

In this section, we are interested in the text posts score based on post text trajectory components (stops and moves) that represents Facebook social signals such as reactions (like, dislike, adore), comments and shares. First we will define a score for the reactions received by a post.

According to Pew Research Center [37], the"likes" are not considered as indicator of a strong link and its polarity does not indicate if it is a positive or a negative sentiment. Authors in [24], saw that the using of likes increases the volume of important data but don't add a value to the analysis. Therefore, they did not use it as indicator but they pretend that"likes" could be potentially be added to the analysis by giving them a low weighting to reflect their lower information quality without dumping them. The "share" and the "comment" are more complex and specify the reaction of the user to the content. In fact, comments generate texts that can be analyzed in order to detect opinions. The share is most important given that it entails the reproduction of the content via the user's own page. Indeed, by sharing a post, the user increases its visibility and thereafter the number of other social signals. That means that the content of the shared post is very interesting and the user wants to benefit others and keep track of the publication for himself.

Based on what we found in the literature [37], [38], [39], [40] (priority share > priority comment > priority like) and according to the reasons that we already described, the following weights 0.2, 0.3, 0.5 were given respectively to the social signals "like", "comment" and" share" as priorities.

The highest priority order is given for sharing because it is more important for a post than comment and like. According to [41] the reactions (like, adore) are the lowest behaviour of the users comparing to commenting on post or sharing a post. The ScoreInteraction is calculated according to the following formula:

$$\mathrm{ScoreInteraction} = \sum React \text{ x } 0.2 + \sum com \text{ x } 0.3 + \sum Share \text{ x } 0.5 \quad (1)$$

The obtained interaction score will be multiplied two times if the time of user reaction or comment or share on the post was made is in the week day and not the weekend. In fact, generally people consult their facebook when they have time so it is the weekend. On the other hand, during the week days, they are too busy and even if they consult their facebook it is just to see roughly the news so if they react to a post during the week day that is means that they are really interested in the post and their interactions will be doubly calculated compared to interactions on weekends because the level of interest to the post is double. Thus the reactions in week days have a higher score than the reactions in the weekend.

3) TEXT POST ENGAGEMENT SCORE COMPUTING

The engagement score is a measure of the level of involvement and interest of the subscribers of a Facebook group in a particular post. It takes into account two key factors: the number of members in the group and the interaction score for the post. The interaction score is a measure of the number of likes, comments, and shares a post receives. To calculate the engagement score, we divide the interaction score for a post by the number of members in the group, as mentioned in formula 2 This allows us to assess the level of engagement and interest of the group members in the post, while also taking into account the size of the group. By using the engagement score, we can gain insights into the effectiveness of a post in engaging the group's members, and identify posts that are particularly successful in generating interest and interaction. This information can be valuable for marketers and social media managers, as it can help them to optimize their content and engagement strategies, and improve their overall reach and impact on the platform.

$$ScoreEngagement = \frac{ScoreInteraction}{\sum membership} \text{x}100 \tag{2}$$

Algorithm 1 presents the pseudo code for calculating a post's interaction score and engagement score

Algorithm 1 Post's Interactions Score and Engagement Score Computing

Input: P : Posts, M1 : Matrix of post, R: reactions, C : Comments, SH:share, M[weekday]: contains the time of reactions in weekday, M[weekend]: contains the time of reactions in weekend, ScoreInteraction : Score of reaction, comments and shares
Output: ScoreEngagement
LOOP Process
for *i to n* **do**
  for *j to n* **do**
    $scoreReaction \leftarrow countM1[Reactions]$
    $ScoreSH \leftarrow countM1[Shares]$
    $ScoreCom \leftarrow countM1[Comment]$
    if $(c \in M[weekday])$ **then**
      $ScoreInteraction \leftarrow ((scoreReact * 02) + (ScoreCom * 0,3) + (scoreSH * 0,5) * 2)$ $ReturnScoreInteraction$
    end if
  else
    if $(c \in M[weekend])$ **then**
      $ScoreInteraction \leftarrow (scoreReact * 02) + (ScoreCom * 0,3) + (scoreSH * 0,5)$ $ReturnScoreInteraction$
    end if
  end for
end for
$ScoreEngagement \leftarrow \frac{ScoreInteraction}{\sum(members)} * 100$
return $ScoreEngagement$

4) TEXT POST OPINION TERMS SCORE COMPUTING

Once we have the interaction score and the engagement scores, we will determine a third score for the opinion terms. In this step we will create a dictionary for opinion terms relative to the Facebook group subject. Then we will compare the stems of the filtered post with the list of opinion terms presented in the dictionary. A filtering and pre-processing process of the text post is applied before the opinion terms score determination as follows: tokenization, removing stop words and stemming. The following algorithm 2 describes the data pre-treatment process of the text post.

The dictionary of french opinion terms is created manually, and their stems are generated using the Porter Stemmer. This latter is a commonly used algorithm in natural language processing for generating stems. An example of this is presented in Table 2, which lists several opinion terms and their corresponding stems or roots. It should be noted that this is only a small sample of the dictionary.

TABLE 2. Opinion terms stems.

Words	Roots or Stems
"Bel","belle","bels","belles"	bel
"joli","jolie","jolis","jolies"	joli
"adore","adorer","adorent","adorais"	ador
"nul","nulle","nuls","nulles"	nul

The methodology of algorithm 3 entails a systematic approach to determine the opinion terms included in a given post by leveraging a lexicon of pre-defined opinion terms contained within a dictionary. The algorithm begins by segmenting the post into individual terms, and subsequently compares each term against the lexicon. If a match is identified between a term in the post and an opinion term within the lexicon, the algorithmic score for the corresponding opinion term is incremented. Subsequently, the algorithm outputs the cumulative score for all identified opinion terms in the post. The effectiveness of the proposed algorithm depends heavily on the accuracy and comprehensiveness of the opinion term lexicon within the dictionary.

5) TEXT POST SCORE COMPUTING

Once we have the interaction score, the engagement score and the opinion terms score, we can calculate the final score called scorePost. In fact, it is the sum of the weighted three scores. The ScorePost is calculated as follow:

$$ScorePost = ((SI\text{x}0.4) + (SE\text{x}0.2) + (SOT\text{x}0.4)) \tag{3}$$

where SI is the ScoreInteraction, SE is the ScoreEngagement, and SOT is the score of opinion terms.

Authors in [5] have demonstrate in their research work that ScoreInteraction and Score Opinion Terms(OT) have the same importance in the detection of opinion leaders process. For this reason we gave the same weight that

Algorithm 2 Data Pre-Treatement

Input: M1:matrix contains the posts of type text, Td[]: table contains the determinants Tp[]: table contains the pronounces Tsw[]: table contains the stopwords P:posts of type text

Output: M2 matrix of pre treatment posts

LOOP Process

for *i to n* **do**
  $Tmsg[] \leftarrow split(M1[i][j])$
  $Tmsg[] \leftarrow RemoveURL(M1[i][j])$
  for *j to* $lengthT[j]$ **do**
    if $(Tmsg[j] \in Td[])$ **then**
      $Delete(Tmsg[j])$
    end if
  end for
  for *k to* $length(Tmsg[k])$ **do**
    if $Tmsg[k] \in Tp[]$ **then**
      $DeleteTmsg[k]$
    end if
  end for
  for *l to* $length(Tmsg[l])$ **do**
    if $(Tmsg[l] \in Tsw[])$ **then**
      $DeleteTmsg[l]$
    end if
  end for
  $M2[i][j] \leftarrow stem.Tmsg[i]$
  $Tmsg[] \leftarrow (Tmsg[] - (Tp[] + Td[] + Tsw[])$
  $M2[i][j] \leftarrow Concat(Tmsg[i]; \text{JJ})$
end for
return $M2$

Algorithm 3 Post Opinion Terms Score Computing

Input: M2=contains the posts after pre-treatment MO: Matrix Contains the list of opinion terms,Valid:boolean

Output: ScoreOT

$Initialisation : ScoreOT \leftarrow 0$
$Initialisation : TM2[] \leftarrow StemM2[i][j]$
$Initialisation : TMO[] \leftarrow Split(MO, \text{JJ})$

LOOP Process

for *i to* $lengthTM2[i], TMO[i]$ **do**
  if $(TM2[i] \in TMO[i])$ **then**
    $ScoreOT++$
  end if
end for
return $ScoreOT$

is 0.4 for both ScoreInteraction and ScoreOT. In contrast, we gave the weight 0.2 to the Score Engagement because the SoreInteraction is more important than Score Engagement since it is more significant. Here are arguments in favor of this perspective:

- Measurement of Tangible Action: Social signals, such as shares, comments, and recommendations, represent concrete actions taken by subscribers in response to your content. These actions go beyond mere observation and indicate an active level of engagement and genuine interest in your content.
- Content Amplification: Social signals, especially shares, have the potential to significantly amplify the reach of your content. When someone shares your content, they bring it to the forefront of their own network, thereby increasing its visibility and dissemination. This can have a significant impact on brand awareness or message propagation.
- Personal Recommendation: When a subscriber shares your content, they essentially recommend it to their social circle, which can enhance trust and credibility. Personal peer recommendations are often perceived as more influential than direct promotional messages.
- Quality Measurement: Social signals can be regarded as a measure of the quality of your content. If the content elicits active reactions, it suggests that it has a positive impact on your audience and resonates with them.
- Alignment with Objectives: If the goal is to create content that prompts specific actions (such as shares to increase reach or comments to stimulate discussion), then social signals are a more direct measure of the success in achieving those objectives.

In this context, you can argue that social signals are more important than the simple engagement score because they reflect real action and influence of the content on the audience.

The following algorithm 4 presents the computing of the post score:

Algorithm 4 Post Score Computing

Input: S: Set of snapshots, P : post, ScoreInteraction, ScoreEngagement, ScoreOT : score of opinion terms

Output: ScorePost

LOOP Process

for *si in s* **do**
  $ScorePost = ScoreInteraction * 0.4 + ScoreEngagement * 0.2 + ScoreOT * 0.4$
end for
return $ScorePost$

B. COMPUTING OF TEXT POSTS' POSITIVE SCORE THROUGH SENTIMENT ANALYSIS OF USERS' COMMENTS

Sentiment analysis [42] is a highly effective area of study for analyzing and interpreting text or human language in its natural form. In fact, Sentiment analysis [43] involves the systematic examination of a collection of text, such as sentences or paragraphs, with the goal of identifying and categorizing human emotions as positive, negative, or neutral [44]. Using sentiment analysis of user comments to detect opinion leaders on Facebook can be useful for several reasons:

- Trend Identification: Sentiment analysis helps determine how users react to posts and comments, aiding in identifying topics that generate the most interest or controversy, which can be related to the influence of an opinion leader.
- Engagement Detection: Opinion leaders often have strong engagement power with their audience. Sentiment analysis can reveal emotional reactions and engagement generated by a user's comments, indicating their influence.
- Monitoring Post Reactions: By tracking positive and negative reactions to a user's posts, it's possible to assess the impact of their comments and publications on public opinion. This can help identify opinion leaders with significant impact.
- Distinguishing Supporters from Opponents: Sentiment analysis can help distinguish supporters of an opinion leader from their opponents, especially in politically or controversially charged discussions where public opinion is polarized.
- Measuring Public Perception: By evaluating sentiments associated with a particular opinion leader, you can measure the public perception of that individual. Overall positive perception may indicate strong positive influence, while negative perception may suggest the opposite.
- Trend Prediction: Sentiment analysis can be used to predict emerging trends and topics gaining popularity. Opinion leaders often play a role in creating and disseminating these trends.
- Reputation Assessment: Sentiment analysis of user comments can contribute to assessing the online reputation of an opinion leader, which is crucial for understanding their influence and potential impact on user opinions and behaviors.

In summary, sentiment analysis of user comments on Facebook can provide valuable insights for opinion leader detection by revealing emotional reactions, trends, engagement, and public perception associated with these leaders. It complements other detection methods and contributes to a better understanding of their influence.

The process of sentiment analysis employs text analysis techniques to determine the sentiment expressed in the text. Sentiment analysis employs natural language processing techniques to perform qualitative analysis on human language [45], with the aim of extracting insights about their emotions and sentiments towards a particular subject of interest. This approach holds significant importance as it enables the understanding of followers' perspectives, feedback, and emotions in response to a text posted on Facebook. Indeed, Sentiment analysis has the capability to assess significant aspects of human emotions, including feelings, emotions and intentions. Feelings are the result of actions that impact a person consciously such as feeling Happy, Sad or Angry. Emotions express thoughts, behavioral responses, or levels of pleasure. they can be classify into Positive, Negative or Neutral. Intentions allow to have an idea about the itent of the person, that means that if he/she is Interested or Not Interested.

After calculating the text post scores, we will proceed to the sentiment analysis of users'comments to determine their polarities that can be positive, negative or neutral. In fact Sentiments encompass a wide range of subjective experiences, including feelings, emotions, opinions, and intentions. Different types of sentiment analysis techniques can reveal latent or hidden sentiments within a text, providing insights into the emotions, opinions, feelings and intentions that may not be explicitly expressed. These hidden sentiments can be identified using techniques such as emotion detection, opinion mining, and intent analysis, among others. Our proposed approach for user comments emotion detection is performed using sentiment analysis with the help of lexicons that are a list of words that are classified into positive, negative or neutral sentiments based on their inherent connotations or emotional valence. The proposed approach follows several steps to determine the polarity or the emotions of user's comments on a given text post. The first step is filtering and pre-processing of user's comment on a given post. The second step is the definition of positive emotion terms dictionary, negative emotion terms dictionary and neutral emotion terms dictionary. Emotion terms are related to a specific context. In fact, emotions are inherently subjective, being influenced by an individual's unique viewpoints, beliefs, and encounters. They are a culmination of one's personal perceptions and experiences in relation to a particular product, human being, or anything else in the world. The third step is the determination of user's comment positive score. Each step will be detailed in the next sub sections.

1) USER'S COMMENTS FILTERING AND PRE-PROCESSING

Upon completion of the extraction process of user comments, it is necessary to carry out a data filtering and pre-processing procedure. This stage involves several steps, including tokenization, the removal of URLs, the elimination of stop words, and the application of stemming.

- **Tokenization:** The tokenization process involves breaking down a text into smaller units, such as words, with each unit being considered a token. The principle of tokenization is to comprehend the meaning of a text by analyzing the tokens that constitute the text. By breaking a text down into its constituent tokens, it becomes easier to analyze and process the text, which is particularly important in natural language processing applications such as sentiment analysis and machine translation.
- **Removing stop words:** The step of removing stop words involves eliminating words that do not contribute significant meaning to the text, such as pronouns and determiners. This is because these words are not considered to be semantically important. In order to accomplish this task, the Natural Language Toolkit

(NLTK) can be used. The NLTK provides a list of commonly occurring stop words that can be used to filter out such words from the text, enabling a more focused analysis of the meaningful content.

- **Stemming:** The third step in the process involves stemming the text, which entails reducing words to their root form. Stemming is considered one of the most common data pre-processing operations, as it helps to standardize the text and reduce the dimensionality of the data. By reducing words to their root form, variations of a word can be treated as a single word, which can aid in tasks such as information retrieval and text classification.

Below is the algorithm 5 for the pre-processing of users' comments:

Algorithm 5 Users'comments Pre-Treatement

Input: M2:matrix contains the post's comments, Tcom[]; Td[]: table contains the determinants; Tp[]: table contains the pronounces; Tsw[]: table contains the stopwords; Tsc[]: contains self comments; Tdc[]:contains duplicate comments

Output: M3: matrix of pre treatment comments of posts

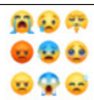
LOOP Process

for *i to n* **do**
 if $Tcom[i] \in Tsc[]$ **then**
  $DeleteTcom[i]$
 end if
else
 if $Tcom[i] inTdc[]$ **then**
  $DeleteTcom[i]$
  $Tcom[] \leftarrow split(M2[i][j])$
  $Tcom[] \leftarrow RemoveURL(M2[i][j])$
 end if
end for
for *j to lengthTcom[j]* **do**
 if $Tcom[j] \in Td[]t$ **then**
  $Delete(Tcom[j])$
 end if
end for
for *k to length(Tcom[k])* **do**
 if $Tcom[k] \in Tp[]$ **then**
  $DeleteTcom[k]$
 end if
end for
for *l to length(Tcom[l])* **do**
 if $Tcom[l] \in Tsw[]$ **then**
  $DeleteTcom[l]$
 end if
end for
$M3[i][j] \leftarrow stem.Tcom[i]$
$Tcom[] \leftarrow (Tcom[] - (Tp[] + Td[] + Tsw[])$
$M3[i][j] \leftarrow Concat(Tcom[i];$ JJ)
return $M3$

2) DICTIONARY ELABORATION

In this step we will define three dictionaries manually. The first for the positive terms, the second for the negative terms and the third for the neutral words. In addition, we will use the emoticons as an indicator of the sentiment of users. In fact, user usually uses emoticons to express its sentiment in the text. Thus, we will add these emoticons in our sentiment analysis dictionaries, and we will present them as punctuation since the emoticons can be written by combining a set of punctuation. Next, we will analyze the comments and extract the positive, the negative and the neutral words that has been used in a given user's comment. After that, we will compare the words of the user comment to the words defined in the three dictionaries to see if there is a match between them. The table 3 represents the dictionaries of the positive, negative and neutral terms and emoticons.

TABLE 3. Examples of user comment positive, negative and neutral terms.

Emoticons	Polarity	Exemples of terms	stemms
	Positive	magnifique, excellent, fabuleux, extraordinaire, belle, jolie, super, beau, superbe, fantastique, adorable, efficace, formlidable, merveilleux, génial, sensationnel, admirable, bon, bien, plais, apprecis, aime, adore, admiré, top	magnifique, excellen, fabuleu, extraordinaire, bel, joli, super, beau, superbe, fantastique, ador, efficace, formlidable, merveil, geni, sensationnel, admir, bon, bien, plais, appreci, aim, ador, admir, top
	Negative	nul, décevant, terrible, catastrophique, déteste, moche, mal, pas bon, flop	nul, deçu, terrible, catastrophique, détest, moch, mal, flop
	Neutral	normal, acceptable, moyen	normal, acceptabl, moyen

3) USERS' COMMENTS POSITIVE SCORE COMPUTING BASED ON SENTIMENT ANALYSIS

After completing the data pre-processing stage, the next step is to classify the users' comments into three categories: positive, negative, and neutral. While many research studies focus only on the positive and negative classes, the neutral class should not be ignored as it can provide valuable insights into the opinions of users. In fact, the neutral class can be particularly important in the detection of opinion leaders, as it may indicate a lack of strong sentiment or a more balanced viewpoint. Therefore, we insist that the classification process includes all three classes in order to provide a more comprehensive analysis of the users' sentiments.

For this aim we will use the Lexicon Based Approach in order to classify the user comment. Firstly, We create our dictionaries which contains positive terms, negative terms and neutral terms and also the positive emoticons, negative emoticons and neutral emoticons. Secondly, we will compare word of the user comment with the words defined in the three dictionaries and assign to each term its polarity. For each user comment that has positive polarity, a score positive comment will be calculated. This latter represents the number of positive words in a user comment that has positive polarity. For each text post, the process of user comments sentiment analysis to extract scores of positive comments based on a dictionaries of positive, negative, and neutral terms is as follows:

- Define the three dictionaries of positive, negative, and neutral terms.
- Tokenize each comment into words.
- Initialize the ScorePositive, ScoreNegative, and ScoreNeutral to 0.
- For each word in the comment, check if it exists in the positive dictionary, negative dictionary, or neutral dictionary.
- If the word is in the positive dictionary, increment the ScorePositive of the comment.
- If the word is in the negative dictionary, increment the ScoreNegative of the comment.
- If the word is in the neutral dictionary, increment the ScoreNeutral of the comment.
- Compare the ScorePositive, of the comment with the ScoreNegative and ScoreNeutral scores.
- If the ScorePositive is higher than both, consider the comment as positive and return its ScorePositive.
- If the ScorePositive is not higher than both, consider the comment as neutral or negative and return 0.
- Repeat steps 2-10 for all comments received by the text post.
- Sum up the positive scores of all positive comments to obtain the total positive score of all comments received by the post.

The following flow chart (figure 3) describes how we calculate the score positive comments resulting from the classification of the user's comments.

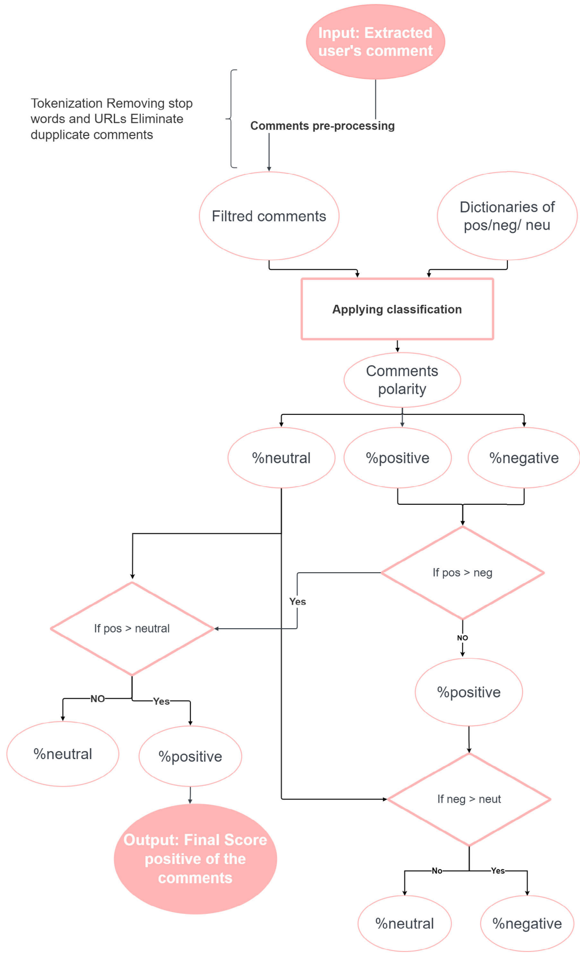


FIGURE 3. Users' positive comments score.

C. OPINION LEADER SCORE COMPUTING

Sentiment analysis and social signals each play important roles in detecting opinion leaders in social networks, but their significance can vary depending on the context and goals of the analysis. In fact, sentiment analysis might be more important. Here are some considerations:

- Influence on Public Opinion: Opinion leaders often have a significant impact on shaping public opinion. Sentiment analysis can help identify how their comments and content are affecting the emotions and attitudes of their followers and the broader audience.
- Content Impact: Opinion leaders are known for their ability to drive engagement and interactions. Sentiment analysis can reveal whether their content is creating positive or negative sentiment, which is a crucial indicator of their influence.
- Trust and Credibility: Opinion leaders are typically trusted sources of information. Analyzing sentiment can help gauge the level of trust and credibility associated with their posts and comments.
- Content Creation: Opinion leaders often create content that resonates with their audience. Sentiment analysis can provide insights into the emotional appeal of their content and help identify patterns that contribute to their influence.

According to [46], The positive score of user comments on a Facebook post is considered more significant than the score of social interactions received by that post because user comments express their opinions and reactions to the post, providing richer information on how users perceive it. Social interactions such as likes and shares also indicate user engagement, but do not provide as much information on user opinion as comments do. Among other reasons and arguments why the positive score of user comments on a Facebook post is more important than the score of social interactions received by that post, we can say that:

- User comments provide more contextual information: While likes and shares can indicate user engagement, they do not provide any contextual information about why a user may have engaged with the post. On the other hand, user comments can provide specific feedback and context about what users liked or didn't like about the post, and how it may have affected them.
- User comments are more reliable indicators of user sentiment: Likes and shares can be influenced by a number of factors such as peer pressure, trends, and algorithms. However, user comments are usually more direct and honest expressions of user sentiment as they allow users to express their opinions and emotions in their own words.
- User comments can provide valuable insights for improvement: User comments can provide valuable feedback to content creators and marketers about what users like or don't like about a post, and how it can be improved in the future. Likes and shares, on the other hand, do not provide as much actionable feedback.
- User comments are more engaging: User comments provide a platform for users to interact with each other and the content creator, which can result in longer and more engaging conversations. This engagement can help build a community around the content and increase its reach and impact. Overall, user comments provide more valuable insights into user sentiment, engagement, and feedback than likes and shares. Therefore, the positive score of comments is deemed more important in assessing the quality of a post and in determining the opinion leader.

According to the reasons described above and after the calculation of the comment's score we will calculate a final score called opinion leader score, which sum up the score of positive comments with the score of the post (ScorePost). Since the sentiment analysis process is more important than social interactions in detecting opinion leaders, We will give higher weight for ScorePositive (0.6) than the ScorePost (0.4). Here are the steps to determine the opinion leader

1) Calculate the opinion leader score (ScoreOP) for each post owner based on the ponderation of the ScorePost and the ScorePositive.
2) Find the post owner with the highest opinion leader score (ScoreOP).
3) Declare the owner found in step 2 as the opinion leader of the Facebook group.

The opinion leader score (ScoreOP) is obtained according to the following algorithm 6:

Algorithm 6 Opinion Leader Score

Input: :ScorePositive, ScorePost, S : set of snapshots, P:Post

Output: ScoreOP

LOOP Process

```
if si ∈ s then
    for Pi ∈ P do
        ScoreOP ← ScorePost ∗ 0.4 + ScorePositive ∗ 0.6
    end for
end if
return ScoreOP
```

V. APPROACH VALIDATION

A. REAL DATASET COLLECTION

To assess the efficiency of our proposed approach, we gathered a dataset comprising of posts from a Facebook group called "LIRE, REVER et PARTAGER".[2] For each post, we retrieved users' reviews, engagement metrics (such as likes, shares, and comments count), and the timestamps of the comments and post creation.

To collect a set of posts, we used a web crawler called Apify.[3] Apify is a web crawler that offers flexible and efficient solution for scraping data from various websites through a web interface or an API. It allows users to utilize distributed queues of URLs to crawl and store structured data or binary files. The platform also enables users to export datasets in formats such as CSV, JSON, Excel, and more. It provides easy access to compute instances (Actors), convenient request and result storages. Furthermore, Apify enable developers to build, deploy, and monitor web scraping and browser automation tools.

All the acquired data was saved into a CSV file to facilitate further analysis. To ensure the accuracy of our results, we conducted a cleaning process, eliminating any irrelevant data such as images and videos. Our focus was solely on text-based posts, which were subject to pre-treatment processing. Ultimately, we identified a total of 895 text-based posts from August 02, 2022, to Mai 09, 2023, which served as the foundation for our algorithm testing.

Table 4 presents some examples of real posts extracted from the Facebook group.

B. MACHINE LEARNING MODELS

Machine learning [47] consists of developping models to make predictions and recommendations [48] without being explicitly programmed. These machines are based on learning from a large datasets and through repetition to produce a particular result. This process called training phase, in which the machine iteratively adjust its parameters, compare the predicted outputs to the inputed data, and minimizing the discrepancy between them.

[2]https://www.facebook.com/groups/281587007113184
[3]https://apify.com/

TABLE 4. Examples of real posts.

Example of textual posts	likesCount	commentsCount	sharesCount
Coucou 😃 Je vous invite à découvrir le nouveau roman de Elodie Lheureux. Sa sortie officielle est le 6 mai, mais les précommandes sont déjà ouvertes. Un livre captivant et émouvant que je vous conseille vivement. Attention, vous n'en sortirez pas indemne. ❤❤❤	1	0	0
Hello. Il était une fois, dans un monde imparfait, une jeune fille plutôt rebelle, qui préférait la compagnie des gens de la rue aux amis de son âge. Mais à se confronter à un monde plutôt glauque ça ne promet rien de bon. C'est conté par Clément Milian avec un style percutant et c'est présenté sur le blog ➡➡➡ https://madosedencre.over-blog.com/.../un-conte-parisien...	2	0	0
Coucou tout le monde ! J'ai lu Astra de Cedar Bowers paru chez Gallmeister. Cedar Bowers nous brosse un portrait très singulier d'Astra à travers les personnes qui l'ont rencontrée. Encore une héroïne fascinante qui comme toutes celles qui l'ont précédée chez Gallmeister, nous restera longtemps en mémoire. Un très beau roman, très original, avec une protagoniste énigmatique et solaire.	3	1	0

The types of machine learning models include Supervised learning and unsupervised learning. Supervised learning is a training set that relies on labelled input, so the classifier predicts the output which is already known. Unlike supervised learning approaches, unsupervised learning approaches do not require any labelled input and must learn by itself.

Machine learning involves two different types of predictions: Classification and regression. Classification consists of classifying data into predefined classes. This type of prediction is used in spam detection, fake news detection, sentiment analysis,etc. On the other hand, regression consists of predicting a continuous numerical value or quantity. It is commonly used for predicting continuous variables such as detecting house prices, time, temperature, size, etc.

The type of prediction depends on the nature of the problem and the type of output variable being predicted. In our case, since the output of our data is known in advance, the training will be supervised and the choice of prediction is regression because the desired output is in the form of numerical values.

The machine learning algorithms employed in this study are as follows:

1) K-NEAREST NEIGHBOR CLASSIFIER (KNN)

KNN is a supervised machine algorithme used for both classification and regression tasks. K-nearest neighbor makes predictions by finding the k nearest neighbors to a given data point and using their labels (for classification) or values (for regression) to determine the prediction. Finaly, The algorithm calculates the distance between the target data and all its neighbors, then it selects the nearest k neighbors. The predicted value, for regression, is calculated as the average of the target values of the k nearest neighbors. KNN is used for its efficiency in supporting multi-labels [49].

2) SUPPORT VECTOR MACHINES(SVM) CLASSIFIER

SVM is another supervised machine used in for both classification and regression tasks. This algorithm works by taking data points and finding an optimal hyperplan that best separates the data points of each class. The objective of the SVM algorithm is to maximize the margin between the hyperplane and the nearest data points, known as support vectors. SVM has proven to be effective in sentiment classification tasks and often outperforms other methods such as Naïve Bayes, KNN, Logistic Regression and others [49], [50], [51]. Furthermove, the study [52] showed that svm has shown great performance when it is combined with dictionaries as it reached 90.4% accuracy.

3) DECISION TREE REGRESSION

Decision Tree Regression is a variant of Decision tree, it is used to predict continuous numerical values. It structures the data in the form of a tree where nodes represent a feature or attribute, branches represent a decision and leaf nodes represents a predicted value.

4) ENSEMBLE MODELS

Ensemble models combines several individual models to improve the overall performance and accuracy of predictions. This technique can involves various approaches such bagging, boosting, and stacking. Bagging methods,or Bootstrap Aggregating, underlies the random forests machine learning algorithm. The technique entails training an ensemble of decision trees on different set of training data and combining their predictions. Unlike Bagging, Boosting iteratively train models to improve the weaknesses of the previous models. As to Stacking, is an ensemble learning method that combine the predictions of the base models in an optimal way.

Nowadays, There has been a growing interest in ensemble learning techniques. Thus, many researches have demonstrated in the past few years that ensemble model works better than individual models such as sentiment analysis [53], [54], [55]. The effective performance of the Ensemble models is due to the following reasons [56]:

- **Overfitting avoidance:** When the available data is limited, learning algorithms are prone to finding numerous hypotheses that perfectly predict the training data but perform poorly on unseen instances.
- **Computational advantage:** Single learners that rely on local searches can become stucked in local optima. Ensemble methods reduce the risk of reaching a local minimum by combining multiple learners.

C. LEXICON-BASED APPROACH

Lexicon-based approach [57] relies on a previously constructed dictionary of positive, negative or neutral words where each word is labelled with its polarity and a corresponding sentiment score. The sentiment polarity of the document (e.g positive, negative or neutral) is determined by summing the sentiment scores of all the words existing in the document. Since we are interested with a particular theme which is reading books, dictionary-based approach will be suitable for our case to estimate and figure out the number of terms used per post and the sentiment behind the comments.

For this task, we will utilize VADER, a rule-based sentiment analysis tool that employs a dictionary to determine the polarity (positive, negative, or neutral) of text. It is widely used for sentiment analysis tasks due to its high performance and availability as a free open-source tool. The decision to use VADER in our research was influenced by previous studies, specifically the research presented in [26], which demonstrated the effectiveness of VADER in sentiment analysis tasks.

D. EVALUATION METRICS

In machine learning, evaluation metrics are used to measure the performance and effectiveness of the used models. There are several measures commonly used such as Precision, Recall, F-1, and Confusion matrix, but, since our models are regression models, we will use other measures such as Mean Absolute Error (MAE), Mean Squared Error (MSE) and R-squared (R^2). Note here, that the choice of the evaluation metrics depends on factors such as the size of the dataset, the complexity of the model and others.

1) MEAN SQUARED ERROR (MSE)

In Statistics, MSE calculates the average squared difference between the actual (observed) and predicted values. The larger the value the larger the error. Thus, lower values indicating better performance. The MSE formula 4 is as following:

$$MSE = \frac{\sum(yi - \hat{y}i)^2}{n} \tag{4}$$

where:

yi– the observed value

ŷi– the predicted value

n– the number of observations

2) R-SQUARED (R^2)

R-squared is a statistical measure that determines the proportion of variance in the dependent variable that can be explained by the independent variable. In other words, r-squared shows how well the data fit the regression model (the goodness of fit). It ranges from 0 to 1, with higher values indicating a better fit of the model to the data.

The formula for calculating R-squared is 5:

$$R - squared = 1 - \frac{SS^{regression}}{SS^{total}} \tag{5}$$

where:

SSregression— the sum of squares due to regression (explained sum of squares).

SStotal— the total sum of squares.

3) MEAN ABSOLUTE ERROR (MAE)

MAE is utilized as a metric to evaluate the performance of a regression model. It is calculated as the average absolute difference between the predicted values and the actual values. The closer the value of the MAE is to 0, the better. The formula for calculating R-squared is 6:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|yi - \hat{y}i| \tag{6}$$

where:

n— the number of observations in the dataset.

yi— the true value.

ŷi— the predicted value.

E. RESULTS AND INTERPRETATIONS

1) SENTIMENT ANALYSIS: CASE STUDY

According to our approach, the identified opinion leader of the group is the writer of the text post in Figure 4.

The text post contains numerous opinion terms, mostly of a positive nature. Furthermore, the post received significant engagement, with 21 reactions (including likes and love), 2 positive comments, and 1 share as mentionned in figure 5.

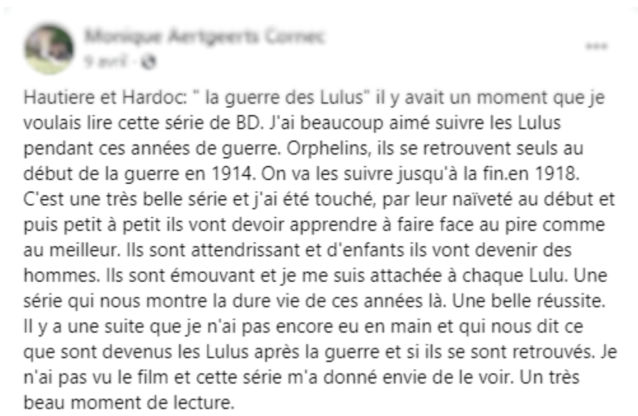


FIGURE 4. The text post of the opinion leader of the group.

FIGURE 5. The text post reactions.

The author of this post writes in an engaging and descriptive style, sharing their thoughts and feelings about the comic series "Hautiere et Hardoc: La Guerre des Lulus." They use a variety of opinion words and phrases to express their sentiments. An analysis of the opinion words and phrases used is presented in table 5

TABLE 5. Opinion terms.

The opinion Terms	Polarity of the opinion term
"beaucoup aimé" (really liked)	Positive opinion
"très belle série" (very beautiful series)	Positive emotion
"très belle série" (very beautiful series)	Highly positive opinion
"touché" (moved)	Positive emotion
"attendrissant" (endearing)	Positive opinion
"émouvant" (moving)	Positive sentiment
"attachée" (attached)	Positive connection to the characters
"belle réussite" (beautiful success)	Highly positive evaluation
"donné envie" (made me want)	Positive influence on the reader
"très beau moment de lecture" (very beautiful reading moment)	Highly positive description

The author employed sentiment analysis by predominantly using positive and appreciative language to express their profound enjoyment of the comic series. Their adept use of positive words and phrases not only makes for a compelling review but also acts as a powerful tool to attract and engage readers on Facebook. Through their expressive writing, readers are influenced and encouraged to explore the series themselves, fostering a sense of community among those who share these sentiments. This connection, conveyed through glowing evaluations and endearing sentiments, is likely to resonate deeply with readers, leading them to be influenced by the author's opinions and emotions and potentially inspiring them to become part of this literary community.

2) EVALUATION OF THE RESULTS

The experiments were conducted using a dataset containing text posts, social interactions, and comments. The dataset was split into 80% for training and 20% for test. Different combinations of feature extraction techniques and classifiers were implemented. BERT and TF-IDF were used to extract features, and then was combined with Stacking Regression, SVM Regression, Decision Tree Regression and KNN Regression for the evaluation. All the obtained results are mentioned in the table 6 accompanied by figures 6 and 7 which present the MAE, MSE, and R^2 evaluation metrics for each model with Bert and TF-IDF features. Note that as the MAE and MSE metrics approach 0, the quality of the results improves. These metrics measure error, and lower values signify better performance. In contrast, R^2 assesses how well the data fits the regression model and falls within the range of 0 to 1.

TABLE 6. Comparison between models using Bert and TF-IDF features.

Models	Features	MAE	MSE	R^2
SVM Regression	Bert	0.21	0.19	0.32
	TF-IDF	0.04	0.10	0.60
KNN	Bert	0.24	0.19	0.20
	TF-IDF	0.25	0.30	0.17
Decision Tree Regression	Bert	0.25	0.20	0.16
	TF-IDF	0.20	0.22	0.10
Stacking Regression	Bert	0.20	0.17	0.96
	TF-IDF	0.20	0.23	0.19

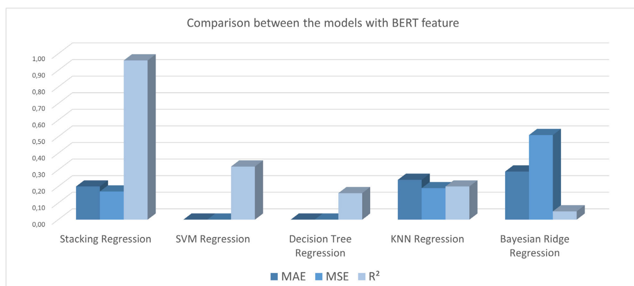


FIGURE 6. Comparison between the models using BERT feature.

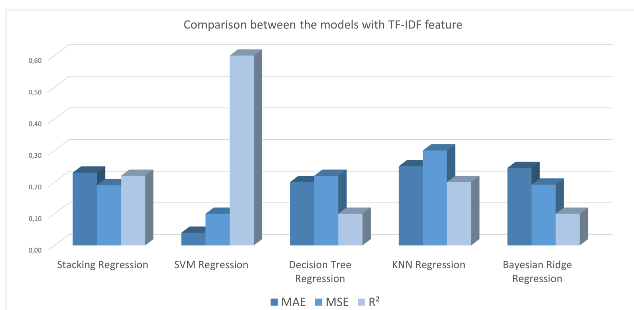


FIGURE 7. Comparison between the models using TF-IDF feature.

By comparing the performance of the models using the TF-IDF feature, SVM stood out with the highest R^2 value, reaching a commendable 0.60. Additionally, it achieved MAE and MSE scores of 0.04 and 0.10, respectively, which are favorable due to their proximity to 0. In contrast, KNN displayed less accurate performance, evident in its relatively high MAE and MSE scores of 0.25 and 0.30, respectively. Its R^2 score of 0.17 further indicated suboptimal fit to the data. While Decision Tree Regression and Stacking Regression outperformed some other algorithms, the distinctions were not particularly significant. When using the Bert feature, table 6 shows that Stacking Regression achieves the best performance followed by decision tree regression and SVM while KNN performs poorly. In the same time, the difference between Stacking and decision tree is minimal. finally, we cannot argue that one of these two algorithms performs better than the other. As the SVM and stacking create better models, all subsequent results will be discussed in terms of them. Figures 8 illustrate the results obtained from running the four models using both the Bert and TF-IDF feature extraction methods. The results indicate that the Bert feature performs slightly better than TF-IDF. Therefore, it is the best feature extraction method to be used for this research purpose.

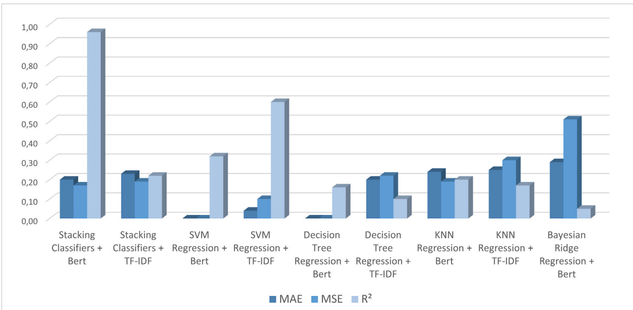


FIGURE 8. **Comparison between the models.**

In order to demonstrate the significance of the selected features, we utilized three distinct datasets (as presented in Table 7) and three performance measures. The first dataset comprises of score interaction (including the number of likes, shares, and comments), score engagement, and the number of opinion terms in a text post. The second dataset includes all the features of the first dataset, as well as the positive score of users' comments. Lastly, the third dataset includes only the positive score of users' comments and the score interaction. By analyzing these datasets, we aimed to evaluate the relevance of the selected features in our approach. We employed three performance measures to determine the accuracy of our model and to validate the effectiveness of our chosen features.

TABLE 7. **Data sets'features.**

Datasets	Features
DS1	ScoreInteraction, ScoreEngagement, ScoreOT
DS2	ScoreInteraction, ScoreEngagement, ScoreOT,ScorePositive
DS3	ScoreInteraction,ScorePositive

The experiments have demonstrated that the dataset Ds2 delivers the best performance among the three datasets. Ds2 comprises of four features, which include all the features of Ds1 along with the positive score of users' comments. The findings presented in Table 8 indicate that the use of Ds2 (including all the features) leads to better accuracy and more precise results for detecting the Opinion Leader.

TABLE 8. **Comparison between models using Bert and TF-IDF features.**

Datasets	Evaluation Metrics	SVM with BERT	Stacking with BERT
Dataset1	MAE	0.08	**0.04**
	MSE	0.01	**0.01**
	R^2	0.93	**0.96**
Dataset2	MAE	0.08	**0.00**
	MSE	0.01	**0.00**
	R^2	0.94	**1.00**
Dataset3	MAE	0.10	**0.02**
	MSE	0.01	**0.01**
	R^2	0.95	**0.99**

Our study findings provide compelling evidence that incorporating sentiment classification has led to a substantial improvement in the performance of the DDOL approach, as compared to our previous work [5]. Specifically, our results indicate that the amalgamation of post score and sentiment polarity of comments results in better accuracy and precision in identifying opinion leaders within the context of the Facebook group "LIRE, REVER et PARTAGER" Table 9 presents the results of our analysis, which support this conclusion unequivocally.

TABLE 9. **Precision comparison between NDDOL and DDOL.**

Approach	Precision
DDOL	80.13%
NDDOL	94.67%

F. DISCUSSION AND EVALUATION

The motivation behind our study was rooted in the ever-growing importance of identifying opinion leaders in social networks, particularly on platforms like Facebook. Opinion leaders play a pivotal role in shaping public discourse, influencing opinions, and driving engagement with content. To address this need, we developed a hybrid approach known as the New Dynamic Detection of Opinion Leaders in Social Networks (NDDOL). This approach combines the analysis of social signals, post text content, and sentiment analysis of user comments to detect opinion leaders effectively.

In our study, we first explained the three main steps of the NDDOL approach. The initial step involves computing a score for the text post of a candidate "opinion leader" by considering various factors such as interaction, engagement, and the presence of opinion terms. The second step focuses on classifying user comments and generating a ScorePositive for the post by employing a sentiment lexicon we specifically designed for the domain. This step uses dictionaries to detect sentiment polarity within comments, ultimately determining if a comment is positive. Finally, the results from these two steps are combined to identify the opinion leader within the targeted Facebook community, "LIRE, REVER et PARTAGER."

In light of these steps and the thorough evaluation of our approach, we can draw meaningful conclusions. Our hybrid approach effectively harnesses social signals, post text content analysis, and sentiment analysis of user comments to identify opinion leaders on social networks. The results align with our initial motivation, which was to provide a robust method for detecting opinion leaders. Moreover, we have surpassed existing state-of-the-art methods, which often rely solely on social signals or basic sentiment analysis. Our approach offers a more comprehensive and accurate means of recognizing opinion leaders, ultimately contributing to a deeper understanding of online influence dynamics and facilitating better decision-making in the realm of social networks.

To assess the effectiveness and quality of our NDDOL method, it is crucial to conduct a comparative analysis with baseline approaches. We validate our work by subjecting our proposed NDDOL method to rigorous testing alongside established methods from prior research. Our comparative evaluation includes benchmarking NDDOL against a range of techniques, including PageRank, lexical-based approaches, maximization techniques, DDOL, and data mining methods based on the following criteria:

- **Network dynamicity:** dynamic networks are characterized by the presence of nodes that can appear and disappear, as well as edges that can experience disruptions and then recover. In contrast, static networks maintain consistent node presence, and edges retain their operational status indefinitely. The primary distinction between dynamic and static networks lies in the consideration of temporal aspects when analyzing structures and interactions. Dynamic networks exhibit changes in their structural composition over time, including the introduction of new members, the cessation of interactions among some members, the formation of new connections, and the obsolescence of existing ones.
- **Context of the data:** considering the context of data is a fundamental aspect of data analysis and interpretation, as it involves examining the broader environment and factors surrounding the data points. This practice is essential for gaining a comprehensive and meaningful understanding of the data and its implications.
- **Sentiment polarity of the comments:** taking into account the sentiment polarity of comments refers to considering the emotional tone or attitude expressed in the comments associated with a piece of content, such as a social media post, product review, or article. Sentiment polarity analysis aims to determine whether the sentiment expressed in these comments is positive, negative, or neutral. This analysis provides valuable insights into how people feel about the content or subject matter.
- **Social signals of the post:** taking into account the social signals of a post refers to considering various indicators and interactions associated with a piece of content, typically on social media platforms. These signals provide insights into how the content is being received and engaged with by the audience. Social signals can include likes, shares, comments, and other forms of user engagement.
- **Opinion terms in Facebook text post:** taking into account the opinion terms means recognizing and considering words or phrases within a text that express a particular sentiment, attitude, or opinion. Opinion terms are linguistic elements that indicate whether the author or speaker has a positive, negative, or neutral viewpoint regarding a subject or topic.
- **Time consuming:** it means assessing how much time and computational resources are required for the approach to complete a task or process.
- **Applicable to any size of data:** it means that the method or technique can be used effectively regardless of the volume or scale of the dataset it is applied to. In other words, the approach is designed to handle datasets of varying sizes, from small to very large, without significant loss of performance or scalability issues.
- **Precision:** Precision is a commonly used statistical measure in fields such as information retrieval, machine learning, and data analysis. Its primary function is to assess the accuracy of classifications or predictions by calculating the ratio of true positive outcomes (instances that are correctly identified) to the total number of positive results (including both correctly and incorrectly identified instances.

Table 10 provides an overview of the various methods and their corresponding comparison criteria.

When comparing the NDDOL approach to several other methods, including PageRank, Lexical-based approaches, Machine Learning methods, Maximization techniques, and DDOL, several key observations come to light.

Firstly, both NDDOL and DDOL demonstrate their versatility by being suitable for various network types, particularly dynamic networks. This adaptability sets them apart from other methods that do not consider network dynamicity, rendering them less suitable for complex graphs. Secondly, NDDOL and DDOL, along with the lexical-based approach, outshine PageRank, Machine Learning methods, and Maximization techniques by taking context into account when analyzing networks. This contextual understanding enhances their ability to uncover meaningful insights. Furthermore, only NDDOL incorporates the evaluation of social signals and sentiment polarity within comments, offering a more comprehensive analysis of user-generated content. Moreover, opinion term recognition, a valuable feature, is uniquely found in the lexical-based approach, DDOL, and NDDOL,

TABLE 10. NDDOL versus Topological methods, lexical-basd approach, machines learning methods, DDOL, and maximization techniques.

Criteria	Page Rank	Lexical-based approach	Machine Learning methods	Maximization technique	DDOL	NDDOL
Network dynamicity	No	No	No	No	Yes	Yes
Take into account the context of data	No	Yes	No	No	Yes	Yes
Take into account the sentiment polarity of the comments	No	No	No	No	No	Yes
Take into account the social signals of the post	No	No	No	No	No	Yes
Take into account the opinion terms	No	Yes	No	No	Yes	Yes
Time Consuming	Time consuming when dealing with a large community	vary depending on several factors	Vary widely depending on several factors	Require considerable computational resources and time to carry out calculations	No	No, but it vary depending on the size of the dataset, and the efficiency of the implemented algorithms.
Applicable to any size of data	Often used to analyse small sized networks	Yes	Yes	Vary depending on the analysis or the goal	Yes	Yes
Precision	87%	85.5%	86.5%	N/A	81%	94.67%

as they rely on predefined dictionaries for this purpose. In terms of precision, NDDOL exhibits significantly higher values compared to PageRank and performs better than Betweenness Centrality. However, no precision value is available for the Maximization technique. Regarding efficiency, NDDOL and DDOL prove to be less time-consuming than the Maximization technique. Still, their efficiency can vary, similar to PageRank, Lexical-based approaches, and Machine Learning methods. Various factors influence their performance. Lastly, Lexical-based approaches, Machine Learning methods, DDOL, and NDDOL demonstrate scalability by being applicable to datasets of various sizes. However, PageRank is often favored for smaller datasets, while the suitability of the Maximization technique depends on specific goals.

These observations are summarized in Figure 9, which depicts the precision metric results for PageRank, Lexical-based approaches, Machine Learning methods, Maximization techniques, NDDOL, and DDOL.

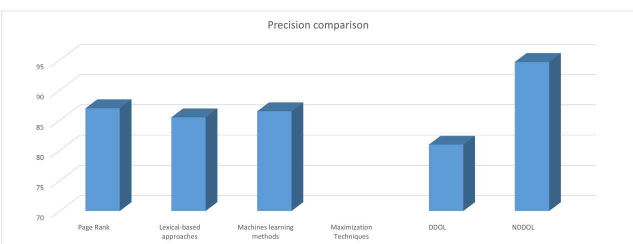


FIGURE 9. Precision comparison.

VI. CONCLUSION

In this research study, we proposed an advanced hybrid approach for detecting opinion leaders on Facebook by integrating Facebook post interactions, text post content analysis, and user comment sentiment analysis. Our approach builds upon the existing approach proposed in [5] and enhances it by considering the post trajectory, which includes elements such as social signals, comments, and shares that constitute the move component of the post trajectory. To improve the score interaction, we incorporated the time of user interaction in order to increase the value of the score interaction in case of interactions were made in the week days and added a new engagement score to further improve the score post. Moreover, we introduced a user comment positive score, which was derived using a sentiment analysis based lexicon approach. Our findings demonstrated that our approach outperformed the previous work presented in [5] by improving the precision metric. This improvement was primarily achieved by taking into account the post trajectory and adding the user comment positive score. These enhancements have enabled us to accurately identify opinion leaders in the Facebook group ''LIRE, REVER et PARTAGER''.

In conducting this study, several limitations were encountered that deserve careful consideration. First, the sample size utilized was relatively small, potentially excluding genuine opinion leaders within the group due to limited data. Additionally, our data collection methods posed challenges, including potential issues concerning data accuracy and completeness. For instance, to improve data quality, additional information about users' account history and their previous posts may have been required. Moreover, confounding variables, such as uncontrolled factors in network structure and temporal changes within dynamic social networks, could influence the outcomes. Furthermore, the context of comments and social signals revealed complexities, as some emotional expressions like 'sad' and 'angry' reactions might reflect support and empathy, while seemingly negative words could carry a positive connotation within specific contexts, such as horror books.

In summary, the field of opinion leader detection based on social signals and user comment analysis is constantly evolving to meet the changing needs of our digital world. It offers a vast realm of research with numerous intriguing prospects. Social networks encompass diverse data types, including text, images, and videos. Our future work can explore how multimodal analysis, which involves examining various data types such as images and videos, can enhance opinion leader detection by considering the entirety of user generated content. Furthermore, we can delve into methods for real-time detection of opinion leaders, a crucial aspect in areas such as real-time marketing and crisis management. Additionally, understanding how the influence of opinion leaders evolves over time is essential. In our future research, we will investigate models of influence change based on social signals and comment analysis, including seasonal fluctuations or shifts related to specific events.

ACKNOWLEDGMENT

Princess Nourah bint Abdulrahman University Researchers Supporting Project number (PNURSP2023R136), Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia.

REFERENCES

[1] B. Jlifi, C. Sakrani, and C. Duvallet, "Towards a soft three-level voting model (soft T-LVM) for fake news detection," *J. Intell. Inf. Syst.*, vol. 2022, pp. 1–21, Dec. 2022.
[2] S. Agarwal and M. Damle, "Sentiment analysis to evaluate influencer marketing: Exploring to identify the parameters of influence," *PalArch's J. Archaeol. Egypt/Egyptol.*, vol. 17, no. 6, pp. 4784–4800, 2020.
[3] S. A. Mohammadi and A. Andalib, "Using the opinion leaders in social networks to improve the cold start challenge in recommender systems," in *Proc. 13rd Int. Conf. Web Res. (ICWR)*, Apr. 2017, pp. 62–66.
[4] N. S. Khan, M. Ata, and Q. Rajput, "Identification of opinion leaders in social network," in *Proc. Int. Conf. Inf. Commun. Technol. (ICICT)*, Dec. 2015, pp. 1–6.
[5] W. Oueslati, S. Arrami, Z. Dhouioui, and M. Massaabi, "Opinion leaders' detection in dynamic social networks," *Concurrency Comput., Pract. Exper.*, vol. 33, no. 1, p. e5692, Jan. 2021.
[6] L. Rocha, F. Mourão, R. Vieira, A. Neves, D. Carvalho, B. Bandyopadhyay, S. Parthasarathy, and R. Ferreira, "Connecting opinions to opinion-leaders: A case study on Brazilian political protests," in *Proc. IEEE Int. Conf. Data Sci. Adv. Anal. (DSAA)*, Oct. 2016, pp. 716–725.
[7] J. C. Lynn-Sze and A. Kamaruddin, "Online opinion leaders in the health promotion digital era," *Jurnal Komunikasi, Malaysian J. Commun.*, vol. 37, no. 4, pp. 295–309, Dec. 2021.
[8] D. W. Baker, "Trust in health care in the time of COVID-19," *JAMA*, vol. 324, no. 23, pp. 2373–2375, Dec. 2020.
[9] N. de Roo, T. Amede, E. Elias, C. Almekinders, and C. Leeuwis, "Diffusion of agricultural knowledge in southern Ethiopia: Finding the real opinion leaders through network analysis," *J. Agricult. Educ. Extension*, vol. 29, no. 1, pp. 99–115, Jan. 2023, doi: 10.1080/1389224X.2021.1987282.
[10] M. T. Troncalli and J. R. Thompson, "Relative importance of information sources used by travel opinion leaders," *J. Travel Res.*, vol. 14, no. 4, pp. 11–14, Apr. 1976, doi: 10.1177/004728757601400404.
[11] A. J. Khan, W. Ul Hameed, J. Iqbal, A. A. Shah, M. A. U. R. Tariq, and S. Ahmed, "Adoption of sustainability innovations and environmental opinion leadership: A way to foster environmental sustainability through diffusion of innovation theory," *Sustainability*, vol. 14, no. 21, p. 14547, Nov. 2022.
[12] J. Zhou and J.-M. Ye, "Sentiment analysis in education research: A review of journal publications," *Interact. Learn. Environments*, vol. 31, no. 3, pp. 1252–1264, Apr. 2023.
[13] C. Erl, R. M. McGregor, J. Lucas, and C. Anderson, "Resentment and admiration: Public opinion toward teachers and public sector employees in Ontario," *Can. J. Educ./Revue Canadienne L'éducation*, vol. 46, no. 3, pp. 597–627, 2023.
[14] F. Fakhreddin and P. Foroudi, "Instagram influencers: The role of opinion leadership in consumers' purchase behavior," *J. Promotion Manage.*, vol. 28, no. 6, pp. 795–825, Aug. 2022.
[15] Z. Li, C. Chan, Y.-F. Chen, W. W. H. Chan, and U. L. Im, "Millennials' hotel restaurant visit intention: An analysis of key online opinion leaders' digital marketing content," *J. Quality Assurance Hospitality Tourism*, vol. 10, no. 2, pp. 1–30, 2023, doi: 10.1080/1528008X.2023.2219467.
[16] R. Fitriati, I. S. M. Siwi, and M. Rudiyanto, "Mega-influencers as online opinion leaders: Establishing cosmetic brand engagement on social media," *J. Promotion Manage.*, vol. 29, no. 3, pp. 359–382, 2023, doi: 10.1080/10496491.2022.2143992.
[17] M. Kang, T. Liang, B. Sun, and H.-Y. Mao, "Detection of opinion leaders: Static vs. dynamic evaluation in online learning communities," *Heliyon*, vol. 9, no. 4, Apr. 2023, Art. no. e14844. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2405844023020510
[18] S. E. Bestvater and B. L. Monroe, "Sentiment is not stance: Target-aware opinion classification for political text analysis," *Political Anal.*, vol. 31, no. 2, pp. 235–256, Apr. 2023.
[19] Y. Chen, X. Wang, B. Tang, R. Xu, B. Yuan, X. Xiang, and J. Bu, "Identifying opinion leaders from online comments," in *Proc. Chin. Nat. Conf. social media Process.* Cham, Switzerland: Springer, 2014, pp. 231–239.
[20] K. Song, D. Wang, S. Feng, D. Wang, and G. Yu, "Detecting positive opinion leader group from forum," in *Proc. Int. Conf. Web-Age Inf. Manage.* Cham, Switzerland: Springer, 2012, pp. 95–101.
[21] G. Sun and S. Bin, "A new opinion leaders detecting algorithm in multi-relationship online social networks," *Multimedia Tools Appl.*, vol. 77, no. 4, pp. 4295–4307, Feb. 2018.
[22] B. Huang, G. Yu, and H. R. Karimi, "The finding and dynamic detection of opinion leaders in social network," *Math. Problems Eng.*, vol. 2014, pp. 1–7, Jan. 2014.
[23] L. R. Luqiu, M. Schmierbach, and Y.-L. Ng, "Willingness to follow opinion leaders: A case study of Chinese Weibo," *Comput. Hum. Behav.*, vol. 101, pp. 42–50, Dec. 2019.
[24] A. M. Litterio, E. A. Nantes, J. M. Larrosa, and L. J. Gómez, "Marketing and social networks: A criterion for detecting opinion leaders," *Eur. J. Manage. Bus. Econ.*, vol. 26, no. 3, pp. 347–366, Oct. 2017.
[25] M. Zhu, X. Lin, T. Lu, and H. Wang, "Identification of opinion leaders in social networks based on sentiment analysis: Evidence from an automotive forum," in *Proc. Int. Conf. Modeling, Simulation Optim. Technol. Appl. (MSOTA)*, 2016, pp. 412–416.
[26] A. A. Al-Rasheed, "Finding influential users in social networking using sentiment analysis," *Informatica*, vol. 46, no. 5, Mar. 2022.
[27] Y. Zhai, Z. Wang, H. Zeng, and Z. Hu, "Social media opinion leader identification based on sentiment analysis," in *Proc. Int. Conf. Bioinf. Intell. Comput.*, Jan. 2021, pp. 436–440.
[28] C. Li, J. Bai, L. Zhang, H. Tang, and Y. Luo, "Opinion community detection and opinion leader detection based on text information and network topology in cloud environment," *Inf. Sci.*, vol. 504, pp. 61–83, Dec. 2019.
[29] A. Benedetti, *Marketing en Redes Sociales: Detrás de Escena*. Buenos Aires, Argentina: AMDIA, 2015.
[30] Z. Wang and H. G. Kim, "Can social media marketing improve customer relationship capabilities and firm performance? Dynamic capability perspective," *J. Interact. Marketing*, vol. 39, pp. 15–26, Aug. 2017.
[31] W. Oueslati and J. Akaichi, "Trajectory data warehouse modeling based on a trajectory UML profile: Medical example," in *Proc. IWBBIO*, 2014, pp. 1527–1538.
[32] W. Oueslati and J. Akaichi, "Mobile information collectors trajectory data warehouse design," *Int. J. Manag. Inf. Technol.*, vol. 2, no. 3, Aug. 2010. [Online]. Available: https://ssrn.com/abstract=3396983

[33] J. Akaichi, "Extending UML for trajectory data warehouses conceptual modelling," *Int. J. Adv. Comput. Sci. Appl.*, vol. 3, no. 11, pp. 1–10, 2012.

[34] W. Oueslati, S. Tahri, H. Limam, and J. Akaichi, "A systematic review on moving objects' trajectory data and trajectory data warehouse modeling," *Comput. Sci. Rev.*, vol. 47, Feb. 2023, Art. no. 100516.

[35] W. Oueslati, S. Tahri, H. Limam, and J. Akaichi, "A new approach for predicting the future position of a moving object: Hurricanes' case study," *Appl. Artif. Intell.*, vol. 35, pp. 1–30, Dec. 2021.

[36] W. Oueslati and J. Akaichi, "A trajectory UML profile for modeling trajectory data: A mobile hospital use case," 2011, *arXiv:1102.4429*.

[37] D. Mochon, K. Johnson, J. Schwartz, and D. Ariely, "What are likes worth? A Facebook page field experiment," *J. Marketing Res.*, vol. 54, no. 2, pp. 306–317, Apr. 2017.

[38] L. K. John, O. Emrich, S. Gupta, and M. I. Norton, "Does 'liking' lead to loving? The impact of joining a brand's social network on marketing outcomes," *J. Marketing Res.*, vol. 54, no. 1, pp. 144–155, Feb. 2017.

[39] K. Subbian, B. A. Prakash, and L. Adamic, "Detecting large reshare cascades in social networks," in *Proc. 26th Int. Conf. World Wide Web*, Apr. 2017, pp. 597–605.

[40] H. Tang, Y. Yin, and H. Shen, "A model for vessel trajectory prediction based on long short-term memory neural network," *J. Mar. Eng. Technol.*, vol. 21, no. 3, pp. 136–145, May 2022, doi: 10.1080/20464177.2019.1665258.

[41] C. Kim and S.-U. Yang, "Like, comment, and share on Facebook: How each behavior differs from the other," *Public Relations Rev.*, vol. 43, no. 2, pp. 441–449, Jun. 2017.

[42] S. M. H. Bamakan, I. Nurgaliev, and Q. Qu, "Opinion leader detection: A methodological review," *Expert Syst. Appl.*, vol. 115, pp. 200–222, Jan. 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0957417418304950

[43] P. Sharma, S. Joshi, S. Gautam, S. Maharjan, S. R. Khanal, M. C. Reis, J. Barroso, and V. M. de Jesus Filipe, "Student engagement detection using emotion analysis, eye tracking and head movement with machine learning," in *Proc. Int. Conf. Technol. Innov. Learn., Teach. Educ.* Cham, Switzerland: Springer, 2022, pp. 52–68.

[44] J. Duan, J. Zeng, and B. Luo, "Identification of opinion leaders based on user clustering and sentiment analysis," in *Proc. IEEE/WIC/ACM Int. Joint Conf. Web Intell. (WI) Intell. Agent Technol. (IAT)*, vol. 1, Aug. 2014, pp. 377–383.

[45] A. Qazi, R. G. Raj, G. Hardaker, and C. Standing, "A systematic literature review on opinion types and sentiment analysis techniques: Tasks and challenges," *Internet Res.*, vol. 27, no. 3, pp. 608–630, Jun. 2017.

[46] R. Singh, R. Bagla, and H. Kaur, "Text analytics of web posts' comments using sentiment analysis," in *Proc. Int. Conf. Workshop Comput. Commun. (IEMCON)*, Oct. 2015, pp. 1–5.

[47] B. Ramzan, I. S. Bajwa, N. Jamil, R. U. Amin, S. Ramzan, F. Mirza, and N. Sarwar, "An intelligent data analysis for recommendation systems using machine learning," *Sci. Program.*, vol. 2019, pp. 1–20, Oct. 2019.

[48] M. Ibrahim, I. S. Bajwa, N. Sarwar, H. A. Waheed, M. Z. Hasan, and M. Z. Hussain, "Improved hybrid deep collaborative filtering approach for true recommendations," *Comput., Mater. Continua*, vol. 74, no. 3, pp. 5301–5317, 2023.

[49] S. S. Aljameel, D. A. Alabbad, N. A. Alzahrani, S. M. Alqarni, F. A. Alamoudi, L. M. Babili, S. K. Aljaafary, and F. M. Alshamrani, "A sentiment analysis approach to predict an individual's awareness of the precautionary procedures to prevent COVID-19 outbreaks in Saudi Arabia," *Int. J. Environ. Res. Public Health*, vol. 18, no. 1, p. 218, Dec. 2020.

[50] C. Troussas, M. Virvou, K. J. Espinosa, K. Llaguno, and J. Caro, "Sentiment analysis of Facebook statuses using Naive Bayes classifier for language learning," in *Proc. IISA*, Jul. 2013, pp. 1–6.

[51] J. Akaichi, Z. Dhouioui, and M. J. López-Huertas Pérez, "Text mining Facebook status updates for sentiment classification," in *Proc. 17th Int. Conf. Syst. Theory, Control Comput. (ICSTCC)*, Oct. 2013, pp. 640–645.

[52] S. Trinh, L. Nguyen, M. Vo, and P. Do, "Lexicon-based sentiment analysis of Facebook comments in Vietnamese language," in *Recent Developments in Intelligent Information and Database Systems* (Studies in Computational Intelligence), vol. 642, D. Król, L. Madeyski, and N. Nguyen, Eds. Cham, Switzerland: Springer, 2016, pp. 263–276, doi: 10.1007/978-3-319-31277-4_23.

[53] D. Sunitha, R. K. Patra, N. V. Babu, A. Suresh, and S. C. Gupta, "Twitter sentiment analysis using ensemble based deep learning model towards COVID-19 in India and European countries," *Pattern Recognit. Lett.*, vol. 158, pp. 164–170, Jun. 2022.

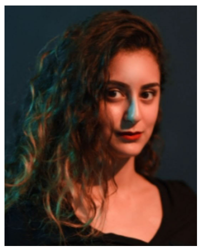
[54] A. Chopra and P. Bhilare, "Application of ensemble models in credit scoring models," *Bus. Perspect. Res.*, vol. 6, no. 2, pp. 129–141, Jul. 2018.

[55] F. Idrees, M. Rajarajan, M. Conti, T. M. Chen, and Y. Rahulamathavan, "PIndroid: A novel Android malware detection system using ensemble learning methods," *Comput. Secur.*, vol. 68, pp. 36–46, Jul. 2017.

[56] O. Sagi and L. Rokach, "Ensemble learning: A survey," *WIREs Data Mining Knowl. Discovery*, vol. 8, no. 4, p. e1249, Jul. 2018.

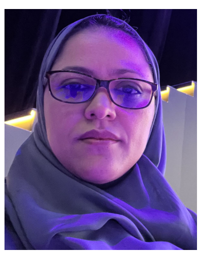
[57] B. Verma and R. S. Thakur, "Sentiment analysis using lexicon and machine learning-based approaches: A survey," in *Proc. Int. Conf. Recent Advancement Comput. Commun.* Cham, Switzerland: Springer, 2018, pp. 441–447.

WIDED OUESLATI received the Ph.D. and Habilitation degrees in business computing from the University of Tunis, Tunisia, in 2013 and 2022, respectively. She is currently an Associate Professor with l'Ecole Supérieure de Commerce de Tunis, l'Université de la Manouba, and the Information Technology Unit Coordinator with the Quantitative Methods Department. Beyond her academic pursuits, she has also been actively involved in various cooperation projects, collaborating with international institutions to foster knowledge exchange and promote research collaboration on a global scale. Her participation in these projects has not only enriched her expertise but has also facilitated valuable cross-cultural and interdisciplinary collaborations. She has published in international journals, book-edited chapters, and conferences, presented various tutorials, and served on the program committees for several international conferences and journals.



SIWAR MEJRI received the master's degree in business intelligence, in July 2023. She is currently pursuing the Ph.D. degree in artificial intelligence and social networks. Her master's degree provided her with a strong foundation in programming, data structures, and algorithms. Her enthusiasm for data analysis and decision-making is a significant driving force in her academic journey. Her unwavering dedication and well-rounded educational background have led her into the world of research.

SHAHA AL-OTAIBI (Member, IEEE) received the M.S. degree in computer science and the Ph.D. degree in artificial intelligence from KSU. She is currently an Associate Professor with the Department of Information Systems, College of Computer and Information Sciences, Princess Nourah bint Abdulrahman University, Saudi Arabia. Her main research interests include data science, artificial intelligence, machine learning, bio-inspired computing, cybersecurity, and information security. She is a Senior Fellow of the U.K. Higher Education Academy (SFHEA). She is a reviewer of some journals and an editorial board member of other journals.



SARRA AYOUNI received the M.S. degree from the Faculty of Sciences of Tunis (FST) and the Ph.D. degree in computer science from the University of Montpellier 2, France. She is currently an Assistant Professor with the Department of Information Systems, College of Computer and Information Sciences (CCIS), Princess Nourah bint Abdulrahman University (PNU), Saudi Arabia. She is also the Coordinator of distance education with CCIS. Her main research interests include data science, artificial intelligence, fuzzy data mining, and e-learning.