IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received October 1, 2020, accepted October 12, 2020, date of publication October 19, 2020, date of current version October 29, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3032216

Sentiment Analysis for e-Payment Service Providers Using Evolutionary eXtreme Gradient Boosting

DANA A. AL-QUDAH[1], ALA' M. AL-ZOUBI[1,2], PEDRO A. CASTILLO-VALDIVIESO[3], AND HOSSAM FARIS[1]

[1]King Abdullah II School for Information Technology, The University of Jordan, Amman 11942, Jordan
[2]School of Informatics and Telecommunications Engineering, University of Granada, 18071 Granada, Spain
[3]Department of Computer Architecture and Technology, University of Granada, 18071 Granada, Spain

Corresponding author: Hossam Faris (hossam.faris@ju.edu.jo)

This work was supported in part by the Project grant from the Deanship of Scientific Research, The University of Jordan, and in part by the Ministerio español de Economía y Competitividad under Project TIN2017-85727-C4-2-P (UGR-DeepBio).

ABSTRACT Online services depend primarily on customer feedback and communications. When this kind of input is lacking, the overall approach of the service provider can shift in unintended ways. These services rely on feedback to maintain consumer satisfaction. Online social networks are a rich source of consumer data related to services and products. Well developed methods like sentiment analysis can offer insightful analyses and aid service providers in predicting outcomes based on their reviews—which, in turn, enables decision-makers to develop effective strategic plans. However, gathering this data is more challenging on Arabic online social networks, due to the complexity of the Arabic language and its dialects. In this study, we propose an approach to sentiment analysis that combines a neutrality detector model with eXtreme Gradient Boosting and a genetic algorithm to effectively predict and analyze customers' opinions of an e-Payment service through an Arabic social network. The proposed approach yields excellent results compared to other approaches. Feature analysis is also conducted on consumer reviews to identify influencing keywords.

INDEX TERMS Evolutionary, genetic algorithm, neutrality detector model, sentiment analysis, social network, XGBoost.

I. INTRODUCTION

With the evolution of the internet and the increasing number of users that access it, the competition to create online services has grown. Web 2.0 innovation has enabled the development of these services through websites that emphasize user-generated content, participatory culture, ease of use, and interoperability. What was once just a vision is now a reality [1].

The governmental sector and private companies have also come to recognize the importance of entering this competition, which will eventually be one of the success factors of increased income. During the COVID-19 pandemic in Jordan, e-government services have been enhanced in various ways to include everything from purchasing services to online appointment booking for medical examinations. However, the changes implemented have not been to the users' satisfaction, and the provision of services is not on the same level as that of many advanced countries. A lack of communication between providers and clients is at least partly to blame [2].

Accessing, analyzing, and understanding client opinions are primary goals for many service providers. Studying client reviews in detail is commonplace not just in Jordan, but all over the world [3], [4], and it benefits both clients and providers. However, analyzing these reviews manually is time-consuming and demanding. For this reason, sentiment analysis — a new method that extracts opinions on a specific topic or service and then analyzes them — has emerged [5]–[8]. Sentiment analysis uses text analysis and Natural Language Processing (NLP) to extract sentiments from reviews [9], [10]. In other words, *"it is a technique used to detect favorable and unfavorable opinions toward specific*

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

products and services using large numbers of textual data sources" [11].

Nowadays, the best and largest platforms for writing reviews or opinions about services are online social networks (OSNs), where users can write whatever and whenever they please. OSNs are user-friendly web-based platforms where content can be written and published with ease — and more quickly than on any other web - based platform. Thus, analyzing and extracting reviews from OSNs is considered not only helpful but essential for modern service providers [1], [12]. In recent years, the literature has investigated different approaches to analyzing online reviews using various techniques [2], [13], [14].

Sentiment analysis in the Arabic context — especially when it includes different dialects, as it does in online social networks — is challenging for many reasons. For example, the Arabic language uses inflection to convey gender, mood, syntactic case, person, tense, number, voice, definiteness, and grammatical aspect [15]. Many common words share meanings but are differentiated by these features. This complexity is only magnified by incorporating different dialects; to date, there is no standardized orthography for Arabic dialects. There are also considerable differences in the linguistic aspects of the dialects within the Arabic language, including their lexicography, syntax, pragmatics, morphology, and semantics [15]. Therefore, lower frequency words inevitably appear on global platforms like online social networks, and the different possible interpretations of these words make it incredibly difficult to detect neutrality. For example, the word الصحة means wellness العافية in most Arabic dialects; however, in the Moroccan dialect, it is النار which means fire.

A machine learning technique that has received a great deal of attention in recent years is the XGBoost (eXtreme Gradient Boosting) algorithm. XGBoost saves time, optimizes memory resources, and can be implemented in a parallel environment. However, it has its drawbacks as well, including a large number of hyperparameters that require a tremendous amount of effort in their tuning.

In this study, we propose a sentiment analysis approach for online e-services in the Arabic context — one that addresses the complexities of neutrality detection and boosts predictive power using a hybrid of XGBoost and a genetic algorithm (GA). The GA enables the automatic tuning of the hyperparameters of the XGboost model. This mitigates the drawback of XGBoost — the tremendous effort required in tuning its hyperparameters. In the proposed approach, the Neutrality Detector Model (NDM) detects neutral opinions in the data; then, the XGBoost — in conjunction with the GA — classifies the remaining opinions as positive or negative. The approach, which will be referred to as NDMGA-XGB, is applied to datasets from the OSNs Twitter and Facebook during two different periods: 2017 and 2019. These data are used to analyze the development of the services, as well as to examine variation in user satisfaction with the main e-Payment service provider in Jordan.

This article is structured as follows: Section II introduces previous work done in the areas of sentiment analysis and, more specifically, Arabic sentiment analysis in e-services. It also explores some of the challenges of using the Arabic language in this context. Section III presents a brief background of the algorithms utilized in the proposed approach. The description of the proposed approach is presented in section IV. Experiments and results are conducted and analyzed in section V, while future work is discussed in section VI.

II. PREVIOUS WORKS

Sentiment analysis (SA) has been the subject of a great deal of interesting research over the last decade, exploring such domains as prediction [16]–[18], politics [19]–[22], products [23], [24], polarity disambiguation [25], [26] and services [27], [28]. Such extensive analysis is made possible by the enormous amount of text and data that already exists and continues to be generated on different platforms, including social networks, blogs, and websites. Given its capacity to process large amounts of data, sentiment analysis — along with affective computing — has become central to analyzing and extracting useful information from text, especially for understanding emotions and expanding human-computer interactions [29].

Though sentiment analysis (SA) is popular in a variety of areas, social networking platforms are the focus of this study due to their large numbers of registered users in recent years [30], [31]. According to the Global Digital Report, these numbers reached 3.48 billion users in 2019 [32]. Several studies have taken advantage of this trove of user metadata, extracting text and analyzing opinions. For example, one study [33] proposed a semi-supervised learning model in order to produce a knowledge base system, then developed a lexical representation by combing through the graphical representation with linguistic resources. This would not have been possible without the benefits of the unstructured bank of social data. Another recent work [34] attempted to predict the intensities of emotions and sentiments through a stacked ensemble method—namely, convolutional neural networks, gated recurrent units, long- and short-term memory, and support vector regression. The proposed model achieved an impressive result, even compared to other state-of-the-art systems.

English-language SA has been applied in many domains in the literature and has made significant advancements [15]. However, SA using other languages — such as Arabic — has received considerably less attention and therefore made considerably less progress, despite a stable and fast-growing network of Arabic users online. In the recent years, interest in Arabic SA has begun to increase.

One study [35] suggests that Arabic SA can help to improve different aspects of e-government services - for example, by combining the technology with the services themselves to better meet citizens' quality needs. This can be done using the deep neural network to automatically apply SA and classify Arabic text as positive or negative. The performance of the deep neural network model achieved a 95.98% accuracy rate. Another study [36] proposed an Arabic dataset consisting of user reviews of government smart applications in Dubai. Their approach, Arabic Aspect-Based Sentiment Analysis (ABSA), combined rule-based models, and lexicon-based methods to measure the SA of these reviews and provide insights into the expectations and needs of clients. The results of the ABSA showed improvements in f-measure and accuracy of 17% and 6%, respectively, against the baseline of the model.

The authors of another work [37] presented a class-specific sentiment analysis (CLASENTI) method for governmental services. Their approach used a new annotation to generate a multi-faceted Arabic lexicon and corpus for different domains, such as polarity strengths, linguistic issues, and dialects — each of which included its own class. For the classification task, they proposed a hybrid of lexicon-based and corpus-based models. Their dataset was collected and annotated from Twitter, Facebook, and surveys. Additionally, they developed a web-based application to execute their proposed framework. The performance of the CLASENTI framework achieved 95% accuracy and a 93% F1-Score. Another study [38] utilized big data sentiment analysis alongside soft computing techniques to assess the e-government domain. They argued that sentiment analysis benefits and improves e-government services and ultimately contributes to its goals of operating in a corruption-free, transparent, and trustworthy manner.

Arabic text requires pre-processing before SA can be conducted. The most complicated consideration is dialect — an understanding of which can help to identify the context. Although the Arabic language has a standard form, known as Modern Standard Arabic (MSA), this form is not popular among OSN users; most Arabic online users tend to utilize dialects (Dialectal Arabic) [39]. The consequent lack of standardization makes meaning difficult to detect.

In the literature, researchers address some of the challenges they have faced when applying Arabic SA, especially in terms of Dialectal Arabic. One research team [40] has stated that stemming is useful when MSA is used, especially for words that have correct roots — but that the same cannot be said for dialectal Arabic, in which most words do not have roots. Another team [41] has discussed the challenges of Arabic SA in social media, where colloquial language and compound terms are common and resources are lacking. The language used on these platforms is continuously evolving and dynamic [42]. Additional problems arise when transliterated English is used. For instance, words like كيوت which is the transliteration of the English word "cute", and اوفر which means "over" (referring to exaggeration) are used frequently by younger generations [43]. The authors of other studies [44], [45] have also noted that the dialectal variations and morphological complexities of the Arabic language necessitate more extensive pre-processing than is required with the English language. A recent work [46] asserts that dialectal Arabic does not undergo sufficient processing in Arabic SA studies. This is a major drawback of current Arabic SA, as this language is used with high frequency on social media [47].

A unique study [14] explored how the extraction of emotion from comments can determine the sentiment ratio of the sharing economy. This work investigated positive/negative sentiments and then established a ratio to determine the risk levels of decisions based on consumer reviews. The outcomes of the proposed approach suggest that the ranking of the sentiment ratio shifts when some customers see, for instance, both information and a picture online. For other customers, seeing details about cost will result in a different decision, and the sentiment ratio will, again, shift accordingly. Different consumers are affected by different information. Measuring sentiment here helped to pinpoint which information was of consequence to which consumers.

One study noted that analyzing and interpreting consumer opinions is important for specific people [48]. The researchers focused on increasing marketing intelligence by examining applicable online comments; in the process, they collected 9,625 tweets about the UK market—more specifically, about fast fashion retailers. The analysis showed a contrast between consumers' responses to various retailers and proposed enhancing their marketing intelligence. Another study [49] suggested that purchase decisions improve when reviews are taken into consideration. Reviewing consumers' feedback can also improve sales, leading to increased income.

For these reasons, it can be useful to implement new solutions and regularly measure consumers' satisfaction levels for different products and services. With the right tools, sentiment analysis has unimaginable capabilities. However, online services do not get nearly as much attention from researchers as other applications do. Although some studies have examined these kinds of services, they have not been investigated in detail, or another issue was faced that prevented services from being purely automatic and available online.

In India, a study was conducted [27] using 153,651 tweets from five different telecom companies: Idea Cellular, Bharti Airtel, Vodafone, Reliance Jio, and Aircel. Using the sentiment score method, the author developed a model for predicting new subscribers using five months' worth of tweets. The results showed that customers' positive opinions about a specific telecom company affected that company's growth in comparison to other companies, with growth being reflected in numbers of new subscribers. Sentiment analysis helped the companies to make timely actions in order to avoid consumer churn and improve consumer experience. Another work used sentiment analysis to review consumer experiences and opinions with cloud services [28]. Their opinions were classified as neutral, negative, or positive. Using two data mining tools — RapidMiner and KNIME — the results were compared against four machine learning algorithms: Random Tree, Naive Bayes, Random Forest, and K-Nearest Neighbour. The Random Forest classifier achieved the best results, with a 97.06% accuracy rate.

Several studies have been proposed that would examine what might be defined as online services. One of them [50], for instance, involved reviewing online forums for the detection and forecasting of hotspots using sentiment analysis. The authors' proposed algorithm would automatically analyze for sentiments and determine a value for each line of text, then combine those values with the algorithm of support vector machine (SVM) and K-means clustering to develop an unsupervised model. The extracted text would be grouped into different clusters and set each center as a hotspot forum. The experimental results showed that the SVM obtained better results than the K-means with a rate of 80%.

In the last decade, many governments have made massive investments in their electronic services to keep up with modern progress and put a solid infrastructure in place for the future in this domain [51]. Electronic government (e-government) services are more necessary than ever before due to the COVID-19 pandemic; services such as the provision of medical records, e-licensing (obtaining licenses electronically), e-tax, and e-payments are suddenly in high demand. These services offer citizens better insights, more choice, and increased efficiency. However, citizens' trust and acceptance of such services remain low [52]. Thus, to build trust between the service provider and consumer, governments must identify the determinants of consumer acceptance. Sentiment analysis techniques offer access to clients' opinions and might be the government's key to effective strategic planning.

This work differs from existing approaches that involve studying feedback during two different time periods and using XGBoost to classify the extracted opinions. To the best of our knowledge, no works have previously combined the XGBoost with Arabic SA. Further, a Neutrality Detector Model (NDM) is a new feature that improves the accuracy of the analysis by filtering neutral opinions. Finally, the genetic algorithm (GA) is utilized to select and tune the XGBoost parameters.

III. PRELIMINARIES

A. XGBoost

XGBoost (eXtreme Gradient Boosting) was first introduced by Tianqi Chen [53], and has since been developed and expanded upon by other developers. Its implementation intended to improve and expedite the performance of machines using gradient-boosted decision trees that maximize the available hardware and memory resources. In other words, it is a cutting-edge and extensible application that uses gradient boosting machines to push computing power to its limits and create boosted tree algorithms. The gradient boosting can also be described as an algorithm that creates new models to predict from previous models; the models are then combined to establish a final prediction, and loss is minimized using the gradient descent algorithm when adding novel models. The XGBoost can assist in tuning the model and in algorithm enhancement, and it can perform and execute the main gradient boosting techniques, including stochastic boosting, gradient boosting, and regularized boosting. Further, the algorithm is known for its ability to optimize time consumption, memory resources, handling of missing values, and the performance of parallel executions in tree construction [53]. One of the main reasons we have chosen this algorithm — aside from its previously mentioned benefits — is its excellent performance against different real-world problems, in both classification and regression applications. Notably, this algorithm won in 17 of the 29 machine learning tasks in the Kaggle Competition in 2015.

The XGBoost functions as a series of tree algorithms that take into account the features of the dataset. These features perform either as an internal node or a conditional node. The root node of the tree splits into several branches and continues dividing until it reaches the leaf node, at which point it cannot produce any more branches or edges. This means the splitting is complete and the decision is made, as shown in Figure 1.

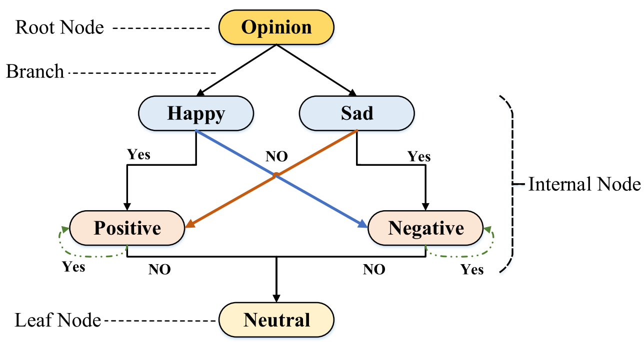

FIGURE 1. Decision tree structure.

The idea of XGBoost revolves around gradient-boosted trees (see Figures 2, 3 and 4) utilizing supervised learning as their primary method. Supervised learning is a strategy in which the data input is used to predict the target value. In more detail, suppose we have a dataset D where $D = \{x, y\}$ with n observations, y refers to the features, and x refers to the class target. K denotes the number of boosting in the XGBoost, and B (the additive function) is applied to predict the output. Additionally, $\hat{y}_i$ refers to the prediction of the i_{th} instance at the b_{th} boost, f_b denotes the tree structure q, and j denotes the leaf with the weighted score w_j. The final prediction is calculated for a given instance x_i by collecting the total scores of all the leaves, as shown in Equation 1.

$$\hat{y}_i = \sum_{b=1}^{B} f_b(x_i) \tag{1}$$

Since the XGBoost base learner come from the decision tree algorithm, various hyper-parameters are linked to the tree structures — such as max leaves, sub-sample, and max depth. This can help to enhance the model while also reducing the over-fitting problem. The learning rate parameter is responsible for controlling the weights of the new trees that will be added to the model, as well as decreasing the

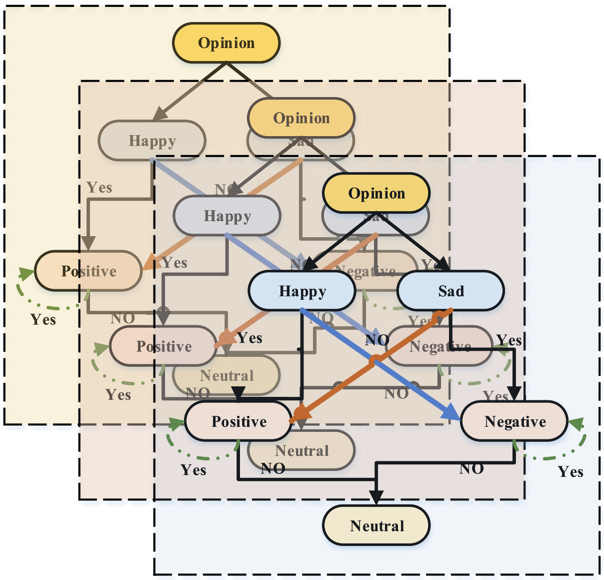

FIGURE 2. Random forest tree structure.

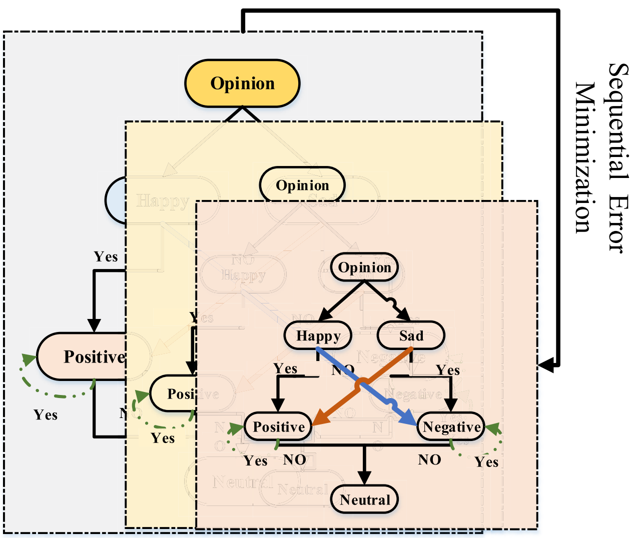

FIGURE 3. Gradient boosting tree structure.

rate of adaption of the model with the training data. Aside from the aforementioned hyper-parameters, there are several additional parameters that can be used to improve the model, including gamma (γ), which is the minimum loss reduction needed for the partition; colsample-bytree, the subsample ratio of the attributes/features to fit the tree; min-child-weight (Wmc), which denotes the instance minimum weights of a specific leaf; and others that will be introduced shortly.

To achieve the best performance from XGboost, the model requires careful tuning of its parameters. Tuning the XGBoost can be a complex task, given the huge number of hyper-parameters; there are different techniques that can be used, such as random search, grid search, and so on. In this article, we applied a new method, using the meta-heuristic algorithm because of its efficiency in finding the optimum parameters for any model.

B. GENETIC ALGORITHM

Genetic algorithms (GA) are popular inspired algorithms and search heuristics based on the natural evolution theory [54]. Inspired algorithms involve higher-level processes designed to select, generate, and find a heuristic that offers an appropriate solution. They are considered part of the Evolutionary Algorithms (EA) family, which generates solutions that mimic natural selection (in which the fittest individuals are encouraged to reproduce and develop the next generation, while others are eliminated). In short, GAs search for optimal solutions among a set of candidate solutions, then the set is repeatedly refined until the best options are selected for the next generation. GA optimization is carried out using three processes—namely, mutation, crossover, and selection [54], [55]. The quality of the solutions (populations) is evaluated using a fitness function; each individual in the population is characterized by a group of variables (genes). These genes are grouped to form a string of solutions (a chromosome).

The three functions — mutation, crossover, and selection — are the primary operators impacting the fitness function. In the selection phase, the fittest individuals or genes (two pairs) are selected to progress to the next level of generation. The second phase — the crossover, which identifies the most efficient operator in the GA — allows for a new region of the solution in the search space. It works by pointing a crossover point or more within the genes that are randomly chosen for exchanging genes between chromosomes. While in the mutation phase, some genes may be subjected to an alteration to limit their premature convergence and maintain the diversity of the population. For example, in binary gene form, the genes may change from 1 to 0, and vice versa — thereby replacing the old population through the use of a diversity replacement or elitism strategy and generating a new population. Each of these phases is summarized in Figure 5.

IV. METHODOLOGY

In this section, the proposed approach for opinion detection in e-services is discussed. The approach consists of several phases, each of which will be described in detail. These phases are in chronological order: data collection, data preparation, model development, and evaluation.

A. DATA DESCRIPTION AND COLLECTION

Efawateercom is the first well-established online payment service in Jordan. It is operated and monitored by the Central Bank of Jordan. Citizens use the service to pay electronic bills for telecommunications, banking, education, insurance, healthcare, trading, water, and electricity, among other things. This online system was created to facilitate payments while reducing the amount of time consumed by regular processes. The traditional methods of payment required a significant amount of time and effort from citizens, including traveling and waiting in lines.

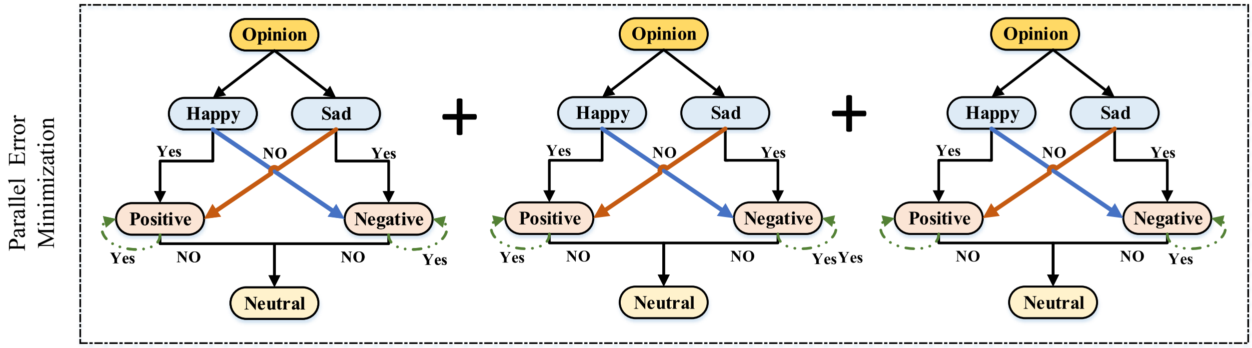

FIGURE 4. XGBoost tree structure.

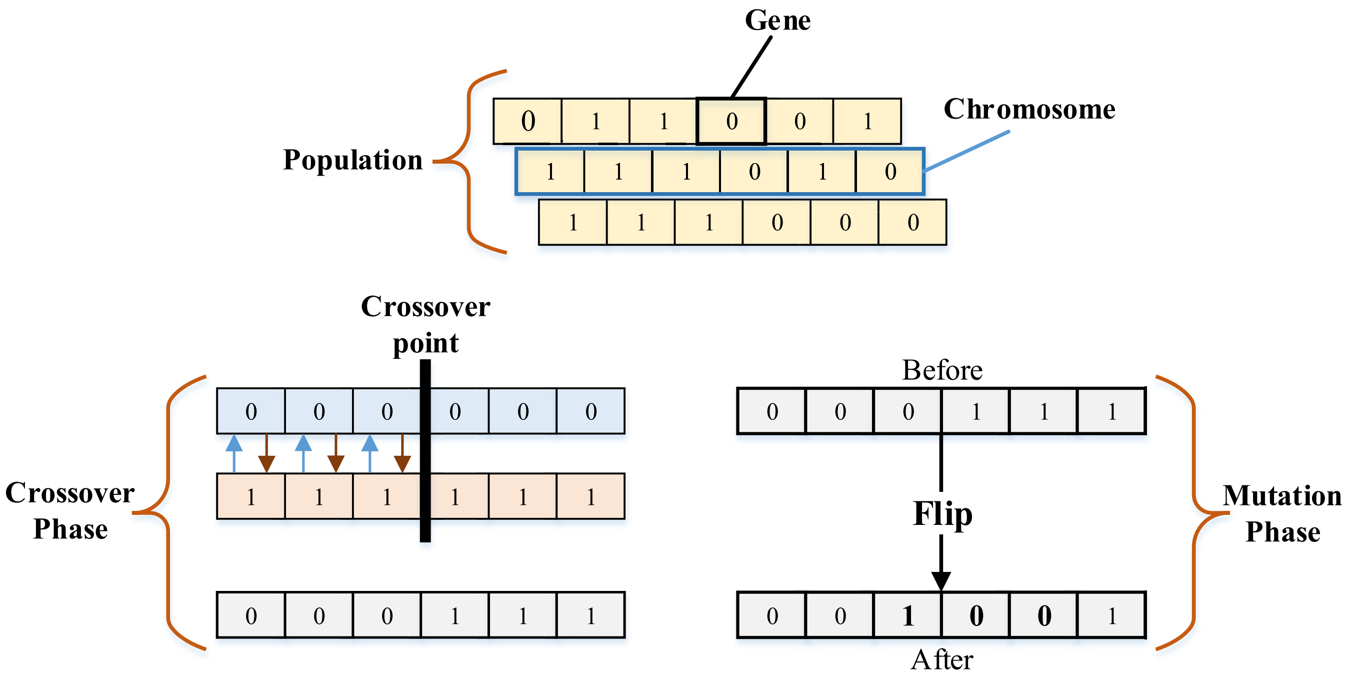

FIGURE 5. Genetic algorithms optimization process.

The slow development of online services in places like the Middle East, and especially in Jordan, has caused numerous users to abandon such services altogether and return to traditional methods. This failure can be attributed to a lacking awareness of users' needs and preferences. Though this failure can be reversed, the process will demand a considerable amount of time and money from the organizations responsible. These costs can be reduced, however, using techniques like text analysis—particularly sentiment analysis.

Our dataset was collected from the online social networks Facebook and Twitter. These platforms were selected due to their popularity in Jordan, their simple interfaces, and the minimal restrictions they place on users' writing. Comments were extracted using a different approach for each social network. On Facebook, we took reviews directly from the Efawateercom Facebook page; meanwhile, on Twitter, we used the term search criteria to locate applicable reviews. We also applied the "Rfacebook" and "twitteR" packages, both of which require a handshake process to secure access and enable comment collection, as shown in Figure 6.

We collected the data from two different periods: 2017 and 2019. This was done to allow for analyses of the development of the services and variations in user satisfaction. The first dataset contained 546 instances and 1743 features, while the second contained 1128 instances and 3385 features. Additional details about the datasets can be found in Table 1.

TABLE 1. Datasets description.

Datasets	Instances	Features	Neutral	Negative	Positive
2017	546	1743	190	165	191
2019	1128	3385	488	303	337

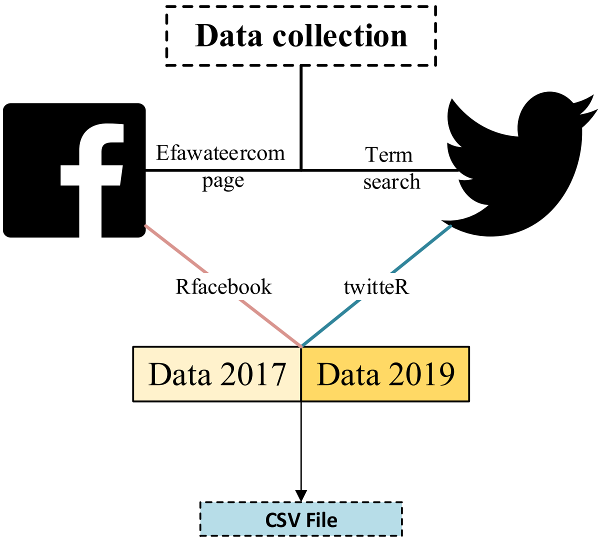

FIGURE 6. Data collection process from each social network.

B. DATA PREPARATION

In this phase, the collected reviews went through pre-processing steps before they could be read. These steps included cleaning, removing missing values, and formatting the data.

After collecting the data in text format in a CSV file, we asked experts to label the data. To facilitate this process, we gave the experts a brief description of the data and what we needed to be done. These individuals were selected based on their experience with the Jordanian dialect.

The comments and reviews were classified into three groups: positive, neutral, and negative. Then, when the labeling was completed for the original text, the preparation process began. The first step was to remove the "stop words" from the text — an example of which might be " كان ، أصبحقال انت". These words do not offer any insight into the true meaning of the phrase and could therefore be removed. A normalization procedure was also performed to remove unwanted non-letters, special characters, and non-Arabic letters to further reduce the number of extracted terms. Then, a stemming process was applied to the text to eliminate any duplication generated during the prior extractions. For instance, letters like " ال,كال,ب" were removed from words; these letters would be referred to as "prefixes" in English. Other letters ("suffixes") were considered and removed by the stemming algorithm as well, including "ون,ان,ها". We employed several stemming techniques but found the Arabic Light Stemming Algorithm to be most effective after standardizing some of the Jordanian dialect words.

Finally, a tokenization process was performed. This process was responsible for breaking the words, punctuation marks, and strings of characters into tokens through linguistic analysis and separating them into individual tokens. We also employed feature extraction techniques, including Term Frequency-Inverse Term Frequency (TF-IDF), that used a numerical statistic process to calculate the relevance of the keywords in each line of our CSV file. TF-IDF was calculated as per Equations 2 & 3.

$$IDF(t) = log(\frac{N}{DF(t)}) \tag{2}$$

where $TF(t, d)$ denotes the number of word t shown in document d, while $IDF(t)$ denotes the inverse document frequency.

$$TF - IDF(d, t) = TF(d, t) \times IDF(t) \tag{3}$$

In Equation 3, N is the whole number of documents, and the document frequency ($DF(t)$) represents the number of documents in which the word t occurs.

In the end, TF-IDF modified the data into a matrix format that was ready for the classifier (the classification model). Although noisy data and missing values could still be found, we noted that this could be solved by either normalizing the data or manually filling in missing values with the "majority vote" among existing values.

In the end, the TF-IDF modified the data into a matrix format that was ready for the classifier (classification model). However, sometimes noisy data and missing values can be found, this can be solved by, normalizing the data.On the other hand, missing value cases can be handled by manually filling the blanks with the majority vote of the existed values.

The data preparation steps are summarized in Figure 7.

C. PROPOSED APPROACH (NDMGA-XGB) MODEL DEVELOPMENT

In this subsection, the classification model was applied to the prepared data. Then, the developed classification model (sentiment identification) was assessed using the confusion matrix. The confusion matrix is a table used for explaining classification performance.

After completing all the preparation steps, the data were ready to be run through the classification model. However, the use of three classes in the datasets reduced the classification accuracy; also, several experiments have been done to ensure this, and the goal of our study was to differentiate negative opinions from positive opinions. Thus, although removing the neutral class would have been logical, it was necessary to retain all three classes to simulate a model for a real-world problem.

Therefore, we implemented an auto-filter known as the Neutrality Detector Model (NDM) on Python. This filter removed the neutral text (tweets) from the dataset based on a pre-defined dictionary of applicable terms. This dictionary was designed by analyzing the dataset as a whole (text), in which positive and negative words were identified, and neutral instances were then removed. Detecting positive and negative words is easier than detecting neural words, which has been a problem in the Arabic language for a long time. Once removed, the neutral instances were transferred to another dataset (*Dataset A*) that would later be used for comparison with the results of the new dataset (*Dataset B*)

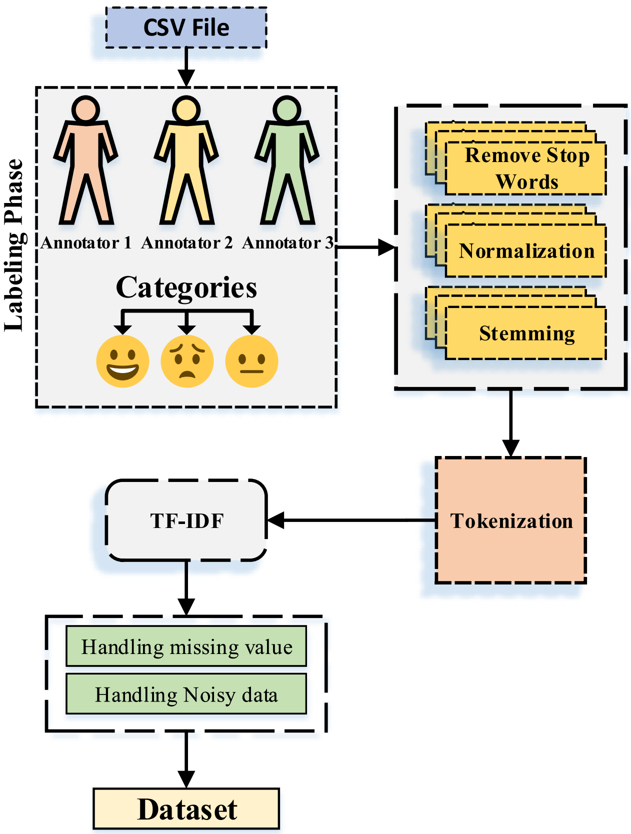

FIGURE 7. Data preparation phases.

in order to investigate the performance of the classification model. Like the original data, *Dataset B* went through the preparation steps. However, the filter — as expected — did not catch all the neutral instances; hence, the classification model was required.

Once the new dataset was prepared, it was time to run it through the classification model. In this work, we used XGBoost as our main classifier and compared it with traditional models, such as J84, NB, and k-NN.

In recent years, XGBoost has been used with increasing frequency; it is now considered one of the best-developed classification models available. However, due to its numerous hyper-parameters, it can be challenging to tune the system and find the optimal combination for every problem. Therefore, we used the GA to solve this problem, as it can determine optimal parameters in a more timely manner than other applied methods such as grid search. Grid search is an algorithm that searches for the best parameters for a given model; its downfall, however, is its time-consuming operation. Therefore, the GA is considered a preferable solution for problems with multiple hyper-parameters to tune. The GA consists of three fundamental components: the search algorithm, the learning algorithm, and the parameter

Regarding which hyper-parameters required tuning, not all parameters enhance the classification model; we chose only those used most frequently in the literature. The selected parameters can be found in Table 2.

1) DESIGN ISSUES

When an optimizer algorithm is applied to a given problem, two design aspects must be considered: the design representation of the solution and the fitness function.

a: SOLUTION REPRESENTATION

The chromosome of the GA is designed to represent the solution to the problem, in which searching for the best parameter of the XGBoost as shown in Figure 8. In our case, the chromosome consists of one-dimensional randomly generated numbers corresponding to the parameter's value. These real numbers have lower and upper boundaries that must be scaled into numbers between 0 to 1 to simplify parameter selection. The scaled criteria can be applied using the following equation:

$$B = \frac{A - min_A}{max_A - min_A}(max_B - min_B) + min_B \tag{4}$$

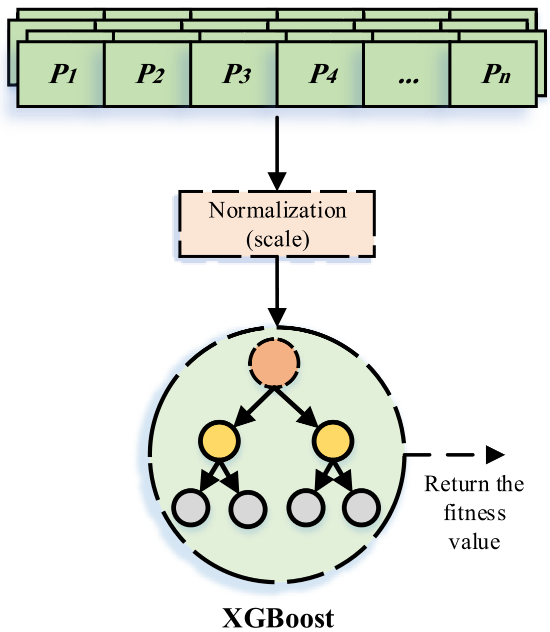

FIGURE 8. Structure of the solution and fitness evaluation.

where B is the new scaled value and A represents the value that needs to be scaled. max_A and min_A denote the upper and lower bounds of the old range, respectively. The upper and lower bounds of the new interval are represented as max_B and min_B, respectively.

b: FITNESS FUNCTION

In this phase, the evaluation criteria were applied to assess and enhance the solution generated by the GA. Fitness value feedback was provided by XGBoost for every iteration. In this work, classification accuracy was selected as a fitness function for the GA and therefore operated as a modification employed in the GA to maximize the fitness value.

TABLE 2. List of all parameters used and their descriptions, ranges, and best values.

#	Parameters	Description	Range	Best value
1	min_child_weight	Denotes the instance minimum weights of a specific leaf	[1-10]	1
2	gamma γ	It is the minimum loss reduction needed for the partition	[0.1-5.0]	1.0
3	subsample	Responsible of the ratio of the training instances	[0.1-1.0]	0.8
4	colsample_bytree	Subsample ratio of the attributes/features to fit the tree	[0.1-1.0]	1.0
5	max_depth	Maximum number of tree depth	[1-20]	4
6	learning_rate	Step size	0.02	-
7	n_estimators	The number of trees utilized in the model	100	-

2) SYSTEM ARCHITECTURE

The procedure of the proposed approach began with the splitting of the dataset (*Dataset B*) into a training set and a testing set. The splitting criteria, in this case, using the 10 folds, where the dataset partition into k parts. The training set was divided into $k-(1/k)$, while the remaining $(1/k)$ parts were utilized for testing to ensure the maximum diversity of both sets and develop the optimal model.

The GA generated a random vector of real numbers in the first iteration. Then, the XGBoost began its training using the selected parameters from the GA. After the XGBoost completed its training, it returned the fitness value to the GA. All steps were repeated for the exact training set until the maximum number of iterations was reached. As long as the number of iterations reaches the maximum value, the best chromosome produced by the GA can be used for testing. All prior procedures were thus repeated k times, then the average of all values was calculated.

D. EVALUATION

The results from *Dataset A* and *Dataset B* were compared to evaluate the performance of the model as a whole, from the NDM to the classification model. This was completed using the confusion matrix table depicted in Figure 9; here, True Positive (TP) is the count of all data points in Class A that were correctly predicted, while False Negative (FN) represents the number in Class A that were wrongly predicted. False Positive (FP), meanwhile, denotes the number of Class B data points that were wrongly predicted as Class A, and True Negative (TN) is the number in Class B that were correctly predicted to be Class B.

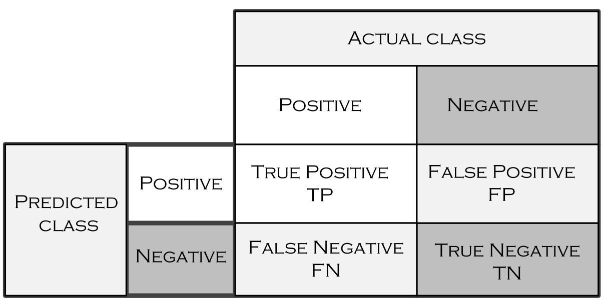

FIGURE 9. Confusion matrix.

Three evaluation measures were used to calculate and assess the model's performance: accuracy, recall, and precision. All of these measures can be calculated using the following equations:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{5}$$

$$Recall(Sensitivity) = \frac{TP}{TP + FN} \tag{6}$$

$$Precision = \frac{TP}{TP + FP} \tag{7}$$

A summary of the framework can be found in Figure 10.

V. EXPERIMENT AND RESULTS

After extracting the data from both OSNs for 2017 and 2019, we split each set into two types: the first (original), with three classes — including positive, negative, and neutral — and the second, which was the filtered data.

For each dataset, there were also two phases of experiments — one for the original data and the other for the filtered data. For the original data, the default XGBoost was compared with other classic machine learning classifiers popular in the literature — namely, k-NN, J48, and NB. Two new models — NDM-Grid and NDMGA-XGB — were added in the filtered data phase, alongside the classic classifiers. These models were responsible for tuning the XGBoost parameters using two different techniques. The first model combined XGBoost with grid search, while our proposed model combined it with the genetic algorithm.

All experiments were conducted on a personal computer with Intel Core i5-6400, 8GB RAM, and Windows 7 specifications. Our model was implemented on Python 3.7.

A. DATA ONE: 2017

First, XGBoost, NDM-Grid, and NDMGA-XGB were applied to the 2017 dataset and measured against the classic machine learning models (k-NN, J48, and NB). In this experiment, there were two phases of three-class data (original), and the data that were filtered by the NDM technique. A description of the classification model's performance is provided in the following subsections.

1) PHASE 1: THREE-CLASSES DATA (ORIGINAL)

In this phase of the 2017 analysis, the original version of the data was evaluated using four classifiers: XGBoost, k-NN,

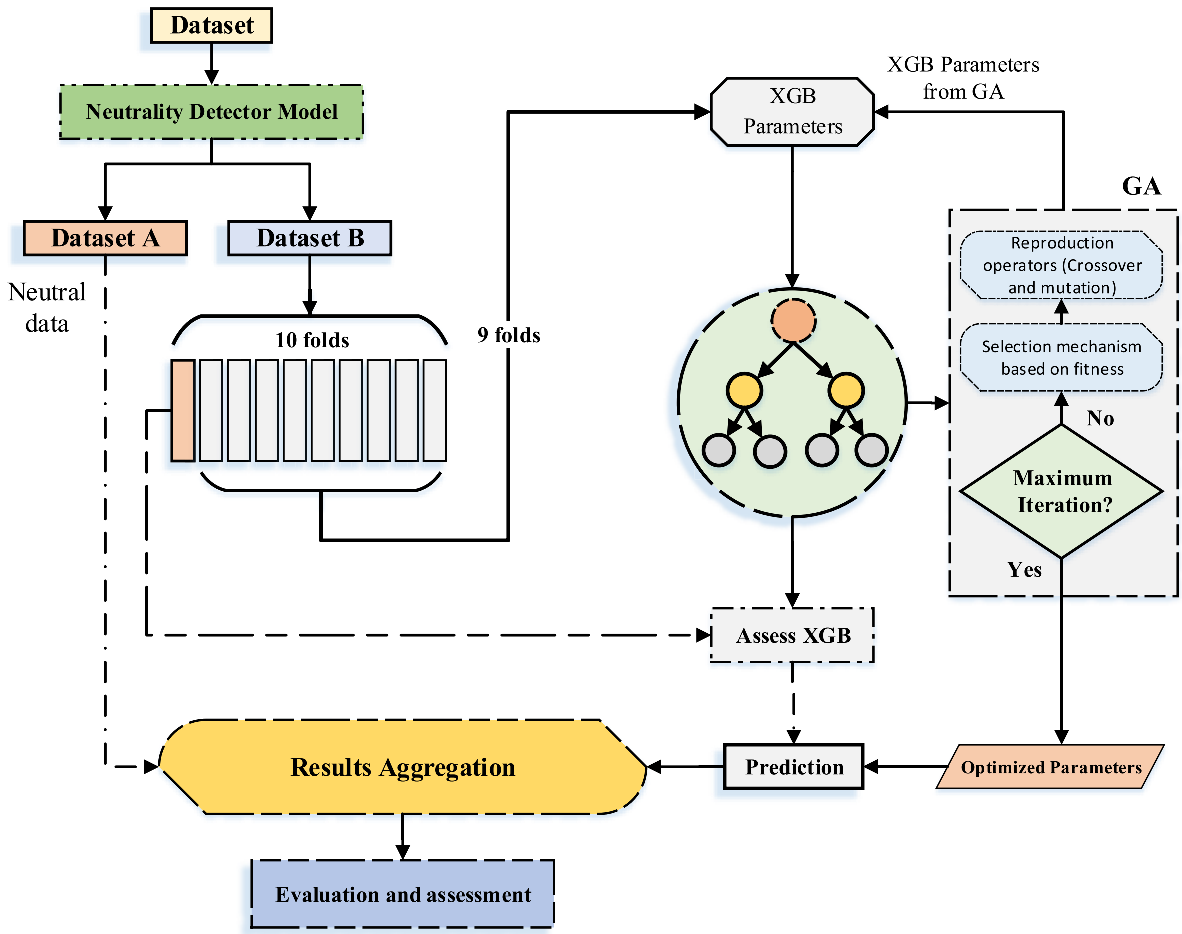

FIGURE 10. General overview of the applied framework.

J48, and NB. As shown in Table 3, three measurements were conducted on the classification models to determine accuracy, recall and precision.

TABLE 3. Results for XGBoost and classic classifiers on three-class data (2017 data).

Algorithm	Accuracy	Recall	Precision
k-NN	0.531	0.332	0.649
J48	0.584	0.416	0.950
NB	0.698	0.558	0.721
XGBoost	**0.709**	0.514	0.604

The classifier with the highest result in terms of accuracy was XGBoost with a rate of 0.706; the second highest was NB with 0.698. For the recall measure, NB obtained the best results, with a rate of 0.558, and XGBoost placed second with 0.514. This shows just how close the competition is between these two models. For precision, it was J48 that had the best results, followed by NB, then k-NN.

This proved that XGBoost offers superior accuracy when compared to the other classifiers and provides competitive results in other measurement areas.

2) PHASE 2: DATA FILTRATION BY THE NDM

In this second phase, as discussed, the data went through a filtering process using the NDM. Two new classification models were also added: NDM-Grid and NDMGA-XGB.

In table 4, the results show that our proposed model, NDMGA-XGB, outperformed all the other classifiers in terms of accuracy and recall. XGBoost ranked second in terms of accuracy, with a rate of 0.769. For the recall measure, J48 placed second and NB third. Nevertheless, k-NN achieved the best results in precision with a rate of 0.888.

These observations indicate that the filtered results have been improved in all measures for most classifiers, as shown in Figures 11, 12 and 13. In other words, all results increased in terms of accuracy except with the NB classifier. Similarly, precision was enhanced for all but one classifier, which was J48. For the recall measure, all classifiers showed improvement in performance.

The NDM improved results by filtering and removing unnecessary opinions — i.e., most of the neutral instances. Furthermore, the NDMGA-XGB achieved better results than the other classifiers, especially concerning accuracy when compared with the unmodified XGBoost. This indicates that

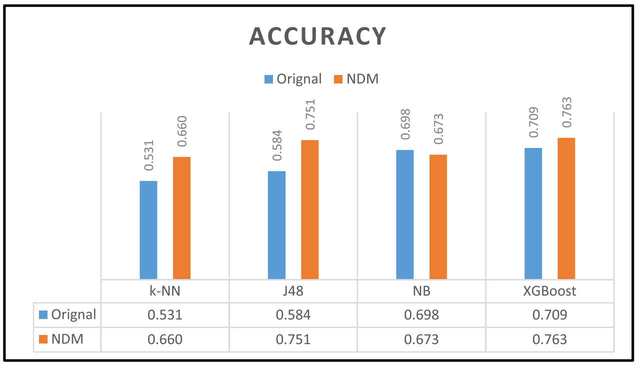

FIGURE 11. Comparison between original and NDM based on average accuracy results for 2017 dataset.

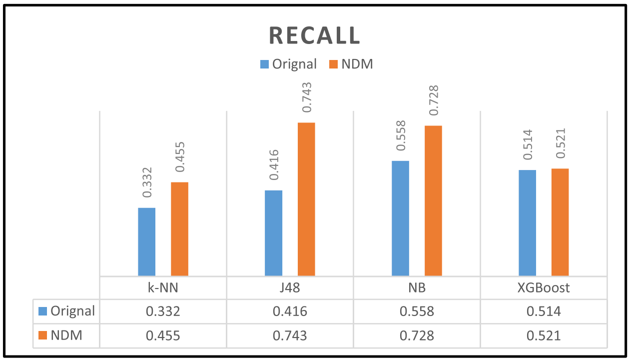

FIGURE 12. Comparison between original and NDM based on average recall results for 2017 dataset.

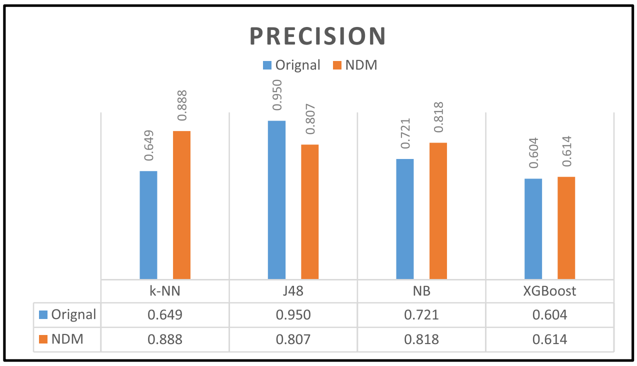

FIGURE 13. Comparison between original and NDM based on average precision results for 2017 dataset.

using the GA to tune the XGBoost parameters improved the classification performance.

B. DATA TWO: 2019

In order to investigate improvements in service, opinions were also extracted for the 2019 period. The 2019 dataset went through the same process as the 2017 dataset.

1) PHASE 1: THREE-CLASSES DATA (ORIGINAL)

In this phase, as shown in Table 5, unmodified XGBoost achieved the best results in terms of accuracy, with a rate of 0.668. The second-best accuracy was achieved by k-NN, with 0.639. In recall, k-NN and XGBoost achieved the best results with 0.852 and 0.828, respectively. In terms of precision, NB has the best results with 0.720, followed by J48 with 0.585.

TABLE 4. Results for GA, Grid, XGBoost and classic classifiers on filtered data (2017 data).

Algorithm	Accuracy	Recall	Precision
k-NN	0.660	0.455	0.888
J48	0.751	0.743	0.807
NB	0.673	0.728	0.818
XGBoost	0.763	0.521	0.614
NDM-Grid	0.714	0.710	0.804
NDMGA-XGB	**0.781**	0.789	0.802

TABLE 5. Results for XGBoost and classic classifiers on three-class data (2019 data).

Algorithm	Accuracy	Recall	Precision
k-NN	0.639	0.852	0.571
J48	0.623	0.764	0.585
NB	0.598	0.453	0.720
XGBoost	**0.668**	0.828	0.578

TABLE 6. Results for GA, Grid, XGBoost and classic classifiers on filtered data (2019 data).

Algorithm	Accuracy	Recall	Precision
k-NN	0.529	0.926	0.564
J48	0.724	0.861	0.711
NB	0.737	0.926	0.690
XGBoost	0.752	0.622	0.715
NDM-Grid	0.758	0.519	0.661
NDMGA-XGB	**0.761**	0.523	0.512

2) PHASE 2:DATA FILTRATION BY THE NDM

In the filtered phase, the results of the 2019 dataset showed more increased values than the previous phase in most models. NDMGA-XGB had the highest results, with rates of 0.761, while NDM-Grid ranked second in terms of accuracy. Both k-NN and NB achieved the best results in recall measure, with 0.926. As for the precision, XGBoost came first, followed by J48.

Similar to Phase 2 of the 2017 data, the results of this phase improved for most classifiers, as shown in Figures 14, 15 and 16.

The NDM criteria, when combined with GA, primarily showed significant improvement in classification performance. This confirms the superiority of our proposed approach relative to the other approaches.

C. FEATURES ANALYSIS OF 2017 AND 2019 DATASETS

To analyze the differences between the two periods of service, we applied feature importance to the two datasets. Additionally, and for this experiment only, the data were transformed to TF-IDF without removing the stop words, to allow for study of the features in all aspects and scenarios.

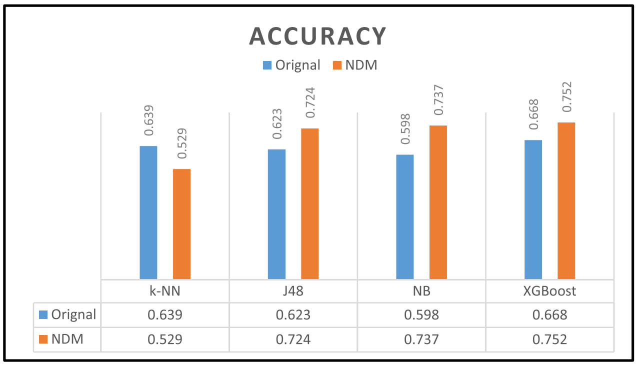

FIGURE 14. Comparison between original and NDM based on average accuracy results for 2019 dataset.

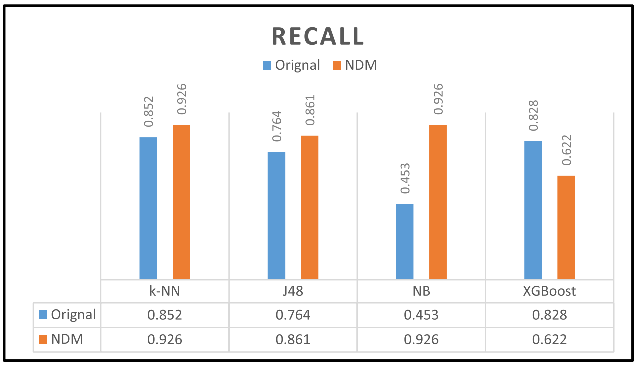

FIGURE 15. Comparison between original and NDM based on average recall results for 2019 dataset.

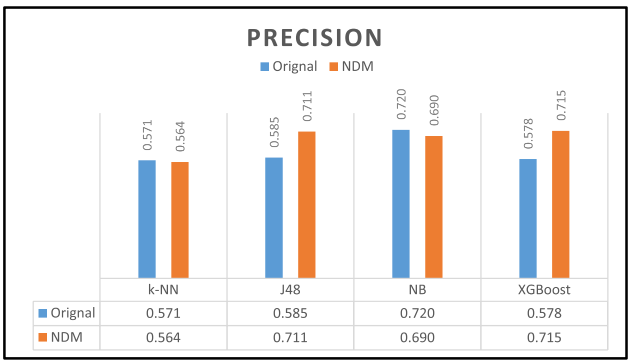

FIGURE 16. Comparison between original and NDM based on average precision results for 2019 dataset.

2019			2017		
#	Features	Translation	#	Features	Translation
F0	رائع	Wonderful	F0	كيف	How?
F1	طريق	Way or Rout	F1	فاتورة	Bill
F2	مش	No or non	F2	اكيد	Sure
F3	عمولة	Commission	F3	رائع	Wonderful
F4	ممتاز	Excellent	F4	دفع	Pay
F5	اختر	Select	F5	ممتاز	Excellent
F6	تطور	Evolve	F6	ممكن	Possible
F7	مبارك	Congrats	F7	نعم	Yes
F8	رمز	symbol	F8	حساب	Account
F9	فاتورة	Bill	F9	وقت	Time
F10	تمنيات	Regards	F10	اي_فواتيركم	eFAWATEERCOM
F11	عنوانك	Address	F11	رقم	Number
F12	زبائنكم	Customers	F12	مش	No or non
F13	مطبق	Applied	F13	الي	Automatic
F14	مميز	Distinctive	F14	افضل	Better or Best
F15	تعاملكم	Interaction	F15	شو	What?
F16	عبيتهم	Fill them	F16	جهد	Effort
F17	يعطيكم	Give you	F17	دور	Line
F18	مبروك	Congratulation	F18	اشتراك	Subscription
F19	مشكلة	Problem	F19	مريح	Comfortable

FIGURE 17. List of features and their translation.

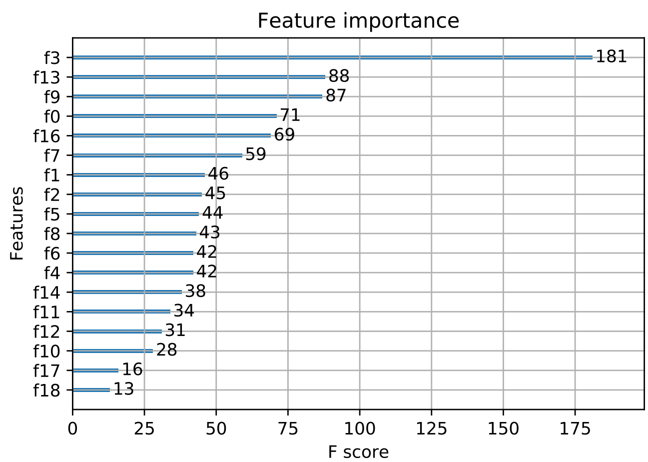

FIGURE 18. Features importance of 2017 dataset.

The selection of the features was performed by the XGBoost algorithm. A list of the features and their translations can be found in Figure 17. The order of the features for both datasets is shown in Figures 18 and 19.

In the first dataset (2017), the most important feature was رائع (f3), meaning "wonderful". This indicates that the majority of users liked the service and found it to be efficient. The second feature was intended to describe one of the main characteristics of the services, which was automation الي. This characteristic helps to facilitate and execute some required activities automatically, on the user's behalf. The third feature, time, is considered an advantage for the user, as it reduces the amount of time wasted while waiting in line to be served. The fourth feature indicated that users had queries about certain activities that needed to be known. This implies that additional direction was needed along with the service provided. Similar to the third feature — time — the fifth feature — effort — indicates that the use of the services reduced the effort required in comparison to traditional methods.

For the 2019 dataset, the first feature was also "wonderful", indicating that the service remained popular among citizens. The symbol feature ranked second; this symbol is employed when the user performs an action and the service needed a confirmation that the same user done the procedure, this feature will increase the security and reduce the risk

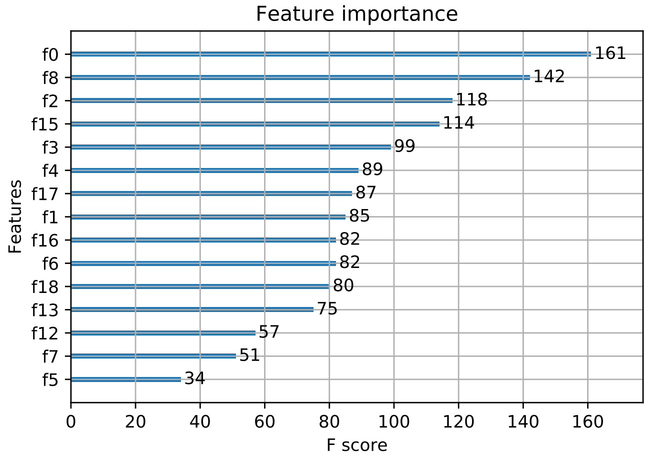

FIGURE 19. Features importance of 2019 dataset.

of hacking their accounts. Some users struggled with this activity and required further information. The third feature was a stop word meaning "no" or "non"; therefore, it could not be interpreted. The fourth feature was interaction, which primarily occurs between users and customer services staff; some users found it useful, and others did not. The fifth feature was commission. In the 2017 data, this feature was not a major consideration; however, the increase in commission percentages over time caused users to request more information about each transaction and payment.

In both datasets, we can observe improvements and increases in certain characteristics. This analysis will help the organizations involved to see their businesses from their users' point of view.

VI. CONCLUSION AND FUTURE DIRECTIONS

The main purpose of this research was to study the development of e-services in Jordan by examining users' satisfaction with the available services. We focused on e-government services—specifically, an online platform called eFawateercom. This platform serves multiple governmental sectors in Jordan in its capacity as an online payment system. User satisfaction with the services was examined using a sentiment analysis approach. Two datasets of client feedback were harvested—one from 2017 and another from 2019—to compare the different periods and determine whether service improved. The data were extracted from both Facebook and Twitter, the two most commonly used social network platforms in Jordan. The data were harvested in the form of text from user reviews of the e-service provided by eFawateercom. In order to extract and analyze the data, this research used three algorithms. The first algorithm, NDM (Neutrality Detector Model), was used to filter the data. Using this model helped to refine source data and prepare it for processing. The next stage involved analyzing the processed dataset using two algorithms: XGBoost (extreme Gradient Boosting) and GA (genetic algorithms). The need for GAs was a consequence of XGBoost's numerous parameters, which GAs can tune based on selected features. Given that the main thrust of this research was to examine different data filtering and processing algorithms for investigating user satisfaction, the proposed approach has yielded excellent results; the research will thus be extended. The future direction of this research will involve examining other e-services in Jordan using a different platform. The datasets to be used in upcoming research will be labeled in advance and ready for processing.

REFERENCES

[1] T. O'Reilly, "What is Web 2.0: Design patterns and business models for the next generation of software," *Commun. Strategies*, vol. 1, p. 17, Mar. 2007.

[2] J. A. Morente-Molinera, G. Kou, Y. Peng, C. Torres-Albero, and E. Herrera-Viedma, "Analysing discussions in social networks using group decision making methods and sentiment analysis," *Inf. Sci.*, vol. 447, pp. 157–168, Jun. 2018.

[3] W. Duan, B. Gu, and A. B. Whinston, "Do online reviews matter?—An empirical investigation of panel data," *Decis. Support Syst.*, vol. 45, no. 4, pp. 1007–1016, Nov. 2008.

[4] F. Tang, L. Fu, B. Yao, and W. Xu, "Aspect based fine-grained sentiment analysis for online reviews," *Inf. Sci.*, vol. 488, pp. 190–204, Jul. 2019.

[5] A. R. Alaei, S. Becken, and B. Stantic, "Sentiment analysis in tourism: Capitalizing on big data," *J. Travel Res.*, vol. 58, no. 2, pp. 175–191, Feb. 2019.

[6] M. A. Hassonah, R. Al-Sayyed, A. Rodan, A. M. Al-Zoubi, I. Aljarah, and H. Faris, "An efficient hybrid filter and evolutionary wrapper approach for sentiment analysis of various topics on Twitter," *Knowl.-Based Syst.*, vol. 192, Mar. 2020, Art. no. 105353.

[7] M. Al-Ayyoub, A. A. Khamaiseh, Y. Jararweh, and M. N. Al-Kabi, "A comprehensive survey of Arabic sentiment analysis," *Inf. Process. Manage.*, vol. 56, no. 2, pp. 320–342, Mar. 2019. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0306457316306689

[8] H. H. Do, P. Prasad, A. Maag, and A. Alsadoon, "Deep learning for aspect-based sentiment analysis: A comparative review," *Expert Syst. Appl.*, vol. 118, pp. 272–299, Mar. 2019. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0957417418306456

[9] J. Yi, T. Nasukawa, R. Bunescu, and W. Niblack, "Sentiment analyzer: Extracting sentiments about a given topic using natural language processing techniques," in *Proc. 3rd IEEE Int. Conf. Data Mining*, Nov. 2003, pp. 427–434.

[10] M. Al-Smadi, B. Talafha, M. Al-Ayyoub, and Y. Jararweh, "Using long short-term memory deep neural networks for aspect-based sentiment analysis of Arabic reviews," *Int. J. Mach. Learn. Cybern.*, vol. 10, no. 8, pp. 2163–2175, Aug. 2019.

[11] D. Delen, *Real-World Data Mining: Applied Business Analytics and Decision Making*. Upper Saddle River, NJ, USA: FT Press, 2014.

[12] A. M. Al-Zoubi, H. Faris, J. Alqatawna, and M. A. Hassonah, "Evolving support vector machines using whale optimization algorithm for spam profiles detection on online social networks in different lingual contexts," *Knowl.-Based Syst.*, vol. 153, pp. 91–104, Aug. 2018.

[13] J. Serrano-Guerrero, J. A. Olivas, F. P. Romero, and E. Herrera-Viedma, "Sentiment analysis: A review and comparative analysis of Web services," *Inf. Sci.*, vol. 311, pp. 18–38, Aug. 2015.

[14] W.-L. Chang and J.-Y. Wang, "Mine is yours? Using sentiment analysis to explore the degree of risk in the sharing economy," *Electron. Commerce Res. Appl.*, vol. 28, pp. 141–158, Mar. 2018.

[15] M. Abdul-Mageed, M. Diab, and S. Kübler, "SAMAR: Subjectivity and sentiment analysis for Arabic social media," *Comput. Speech Lang.*, vol. 28, no. 1, pp. 20–37, Jan. 2014.

[16] T. H. Nguyen, K. Shirai, and J. Velcin, "Sentiment analysis on social media for stock movement prediction," *Expert Syst. Appl.*, vol. 42, no. 24, pp. 9603–9611, Dec. 2015.

[17] A. Mittal and A. Goel, "Stock prediction using Twitter sentiment analysis," Stanford Univ., Palo Alto, CA, USA, Tech. Rep. CS229, 2012, vol. 15. [Online]. Available: http://cs229.stanford.edu/proj2011/GoelMittal-StockMarketPredictionUsingTwitterSentimentAnalysis.pdf

[18] T. H. Nguyen and K. Shirai, "Topic modeling based sentiment analysis on social media for stock market prediction," in *Proc. 53rd Annu. Meeting Assoc. Comput. Linguistics 7th Int. Joint Conf. Natural Lang. Process.*, vol. 1, 2015, pp. 1354–1364.

[19] H. Wang, D. Can, A. Kazemzadeh, F. Bar, and S. Narayanan, "A system for real-time Twitter sentiment analysis of 2012 us presidential election cycle," in *Proc. ACL Syst. Demonstrations*. Stroudsburg, PA, USA: Association Computational Linguistics, 2012, pp. 115–120.

[20] A. Bermingham and A. Smeaton, "On using Twitter to monitor political sentiment and predict election results," in *Proc. Workshop Sentiment Anal. AI Psychol. (SAAIP)*, 2011, pp. 2–10.

[21] M. Choy, M. L. F. Cheong, M. N. Laik, and K. P. Shung, "A sentiment analysis of Singapore presidential election 2011 using Twitter data with census correction," 2011, *arXiv:1108.5520*. [Online]. Available: http://arxiv.org/abs/1108.5520

[22] E. T. K. Sang and J. Bos, "Predicting the 2011 Dutch senate election results with Twitter," in *Proc. Workshop Semantic Anal. Social Media*, 2012, pp. 53–60.

[23] W. Zhang, H. Xu, and W. Wan, "Weakness finder: Find product weakness from Chinese reviews by using aspects based sentiment analysis," *Expert Syst. Appl.*, vol. 39, no. 11, pp. 10283–10291, 2012.

[24] S. Mukherjee and P. Bhattacharyya, "Feature specific sentiment analysis for product reviews," in *Proc. Int. Conf. Intell. Text Process. Comput. Linguistics*. Berlin, Germany: Springer, 2012, pp. 475–487.

[25] F. Yin, Y. Wang, J. Liu, and L. Lin, "The construction of sentiment lexicon based on context-dependent part-of-speech chunks for semantic disambiguation," *IEEE Access*, vol. 8, pp. 63359–63367, 2020.

[26] Y. Xia, E. Cambria, A. Hussain, and H. Zhao, "Word polarity disambiguation using Bayesian model and opinion-level features," *Cognit. Comput.*, vol. 7, no. 3, pp. 369–380, Jun. 2015.

[27] S. Ranjan, S. Sood, and V. Verma, "Twitter sentiment analysis of real-time customer experience feedback for predicting growth of Indian telecom companies," in *Proc. 4th Int. Conf. Comput. Sci. (ICCS)*, Aug. 2018, pp. 166–174.

[28] A. M. Alkalbani, L. Gadhvi, B. Patel, F. K. Hussain, A. M. Ghamry, and O. K. Hussain, "Analysing cloud services reviews using opining mining," in *Proc. IEEE 31st Int. Conf. Adv. Inf. Netw. Appl. (AINA)*, Mar. 2017, pp. 1124–1129.

[29] E. Cambria, "Affective computing and sentiment analysis," *IEEE Intell. Syst.*, vol. 31, no. 2, pp. 102–107, Mar. 2016.

[30] A. Al-Zoubi, J. Alqatawna, H. Faris, and M. A. Hassonah, "Spam profiles detection on social networks using computational intelligence methods: The effect of the lingual context," *J. Inf. Sci.*, pp. 1–25, Aug. 2019, doi: 10.1177/0165551519861599.

[31] J. Alqatawna, A. Madain, A.-Z. Ala'M, and R. Al-Sayyed, "Online social networks security: Threats, attacks, and future directions," in *Social Media Shaping E-Publishing and Academia*. London, U.K.: Springer, 2017, pp. 121–132.

[32] S. Kemp. (2019. *Digital Trends 2019: Every Single Stat You Need to Know About The Internet*. [Online]. Available: https://thenextweb.com/contributors/2019/01/30/digital-trends-2019-every-single-stat-you-need-to-know-about-the-internet/

[33] A. Hussain and E. Cambria, "Semi-supervised learning for big social data analysis," *Neurocomputing*, vol. 275, pp. 1662–1673, Jan. 2018.

[34] M. S. Akhtar, A. Ekbal, and E. Cambria, "How intense are you? Predicting intensities of emotions and sentiments using stacked ensemble," *IEEE Comput. Intell. Mag.*, vol. 15, no. 1, pp. 64–75, Feb. 2020.

[35] O. S. Al-Mushayt, "Automating e-government services with artificial intelligence," *IEEE Access*, vol. 7, pp. 146821–146829, 2019.

[36] S. Areed, O. Alqaryouti, B. Siyam, and K. Shaalan, "Aspect-based sentiment analysis for Arabic government reviews," in *Recent Advances in NLP: The Case of Arabic Language*. Cham, Switzerland: Springer, 2020, pp. 143–162.

[37] A. Hamdi, K. Shaban, and A. Zainal, "CLASENTI: A class-specific sentiment analysis framework," *ACM Trans. Asian Low-Resource Lang. Inf. Process.*, vol. 17, no. 4, pp. 1–28, Aug. 2018.

[38] A. Dwivedi, R. Pant, S. Pandey, M. Pande, and A. K. Mittal, "Benefits, of using big data sentiment analysis and soft computing techniques in e-governance," *Int. J. Recent Technol. Eng.*, vol. 8, no. 3, pp. 3038–3044, 2019.

[39] S. Harrat, K. Meftouh, and K. Smaili, "Machine translation for Arabic dialects (survey)," *Inf. Process. Manage.*, vol. 56, no. 2, pp. 262–273, Mar. 2019.

[40] L. Albraheem and H. S. Al-Khalifa, "Exploring the problems of sentiment analysis in informal Arabic," in *Proc. 14th Int. Conf. Inf. Integr. Web-Based Appl. Services (IIWAS)*, 2012, pp. 415–418.

[41] S. Volkova, T. Wilson, and D. Yarowsky, "Exploring sentiment in social media: Bootstrapping subjectivity clues from multilingual Twitter streams," in *Proc. 51st Annu. Meeting Assoc. Comput. Linguistics*, vol. 2, 2013, pp. 505–510.

[42] S. R. El-Beltagy and A. Ali, "Open issues in the sentiment analysis of Arabic social media: A case study," in *Proc. 9th Int. Conf. Innov. Inf. Technol. (IIT)*, Mar. 2013, pp. 215–220.

[43] H. ElSahar and S. R. El-Beltagy, "A fully automated approach for Arabic slang lexicon extraction from microblogs," in *Proc. Int. Conf. Intell. Text Process. Comput. Linguistics*. New York, NY, USA: Springer, 2014, pp. 79–91.

[44] S. O. Alhumoud, M. I. Altuwaijri, T. M. Albuhairi, and W. M. Alohaideb, "Survey on Arabic sentiment analysis in Twitter," *Int. Sci. Index*, vol. 9, no. 1, pp. 364–368, Feb. 2015.

[45] M. Korayem, D. Crandall, and M. Abdul-Mageed, "Subjectivity and sentiment analysis of Arabic: A survey," in *Proc. Int. Conf. Adv. Mach. Learn. Technol. Appl.* Berlin, Germany: Springer, 2012, pp. 128–139.

[46] A. Assiri, A. Emam, and H. Aldossari, "Arabic sentiment analysis: A survey," *Int. J. Adv. Comput. Sci. Appl.*, vol. 6, no. 12, pp. 75–85, 2015.

[47] C. Qwaider, S. Chatzikyriakidis, and S. Dobnik, "Can modern standard Arabic approaches be used for Arabic dialects? Sentiment analysis as a case study," in *Proc. 3rd Workshop Arabic Corpus Linguistics*, 2019, pp. 40–50.

[48] E. Pantano, S. Giglio, and C. Dennis, "Making sense of consumers' tweets: Sentiment outcomes for fast fashion retailers through big data analytics," *Int. J. Retail Distrib. Manage.*, vol. 47, no. 9, pp. 915–927, Sep. 2019.

[49] V. Vyas and V. Uma, "Approaches to sentiment analysis on product reviews," in *Sentiment Analysis and Knowledge Discovery in Contemporary Business*. Hershey, PA, USA: IGI Global, 2019, pp. 15–30.

[50] N. Li and D. D. Wu, "Using text mining and sentiment analysis for online forums hotspot detection and forecast," *Decis. Support Syst.*, vol. 48, no. 2, pp. 354–368, Jan. 2010.

[51] J. D. Twizeyimana and A. Andersson, "The public value of E-government–a literature review," *Government Inf. Quart.*, vol. 36, no. 2, pp. 167–178, Apr. 2019.

[52] H. Al-Khatib, H. Lee, C. Suh, and V. Weerakkody, "E-government systems success and user acceptance in developing countries: The role of perceived support quality," Brunel Univ. Res. Arch., London, U.K., Tech. Rep. 2438/18514, 2019.

[53] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Aug. 2016, pp. 785–794.

[54] R. L. Kadri and F. F. Boctor, "An efficient genetic algorithm to solve the resource-constrained project scheduling problem with transfer times: The single mode case," *Eur. J. Oper. Res.*, vol. 265, no. 2, pp. 454–462, Mar. 2018.

[55] H. Faris, A. M. Al-Zoubi, A. A. Heidari, I. Aljarah, M. Mafarja, M. A. Hassonah, and H. Fujita, "An intelligent system for spam detection and identification of the most relevant features based on evolutionary random weight networks," *Inf. Fusion*, vol. 48, pp. 67–83, Aug. 2019.

DANA A. AL-QUDAH received the B.Sc. degree from The University of Jordan, Amman, the M.Sc. degree from the University of Sunderland, U.K., and the Ph.D. degree from the University of Warwick, U.K. She has been an Assistant Professor with the Department of Information Technology, The University of Jordan, since 2016. Her research interests include adaptation, personalization, and data mining in web services and related applications. She has many publications in these fields and is working on funded projects, such as Blogforever, from 2011 to 2014, at Warwick University; FORC, from 2017 to 2020, an Eurasmus+ Project; and studying and analyzing consumers' behaviors using sentiment analysis, from 2019 to 2021, a locally funded project from The University of Jordan. She has also worked as the Assistant Dean for quality and development with the King Abdullah II for Information Technology.

ALA' M. AL-ZOUBI received the B.Sc. degree in software engineering from Al-Zaytoonah University, in 2014, and the M.Sc. degree in web intelligence from The University of Jordan, Jordan, in 2017. He is currently pursuing the Ph.D. degree with the School of Informatics and Telecommunications Engineering, University of Granada. During his graduate studies, he has published several publications in well-recognized journals and conferences. He has worked as a Teacher and a Research Assistant on several projects funded by The University of Jordan. He is a member of the JISDF Research Group, which focuses on bridging the gap between academic and industry mechanisms in security. His research interests include evolutionary computation, machine learning, and security in social network analysis and other research fields.

PEDRO A. CASTILLO-VALDIVIESO was born in Granada, in 1974. He received the B.Sc. degree in computer science and the Ph.D. degree from the University of Granada, Spain, in 1997 and 2000, respectively. He has worked as a Teaching Assistant with the Computer Science Department, University of Jaén, Spain, and a Visiting Researcher with Napier University, Edinburg, U.K., in July 1998, and the Santa Fe Institute, Santa Fe, NM, USA, in September 2000. He has led several research projects and directed five Ph.D. students. He is currently an Associate Professor with the Department of Computer Architecture and Technology, University of Granada. His main research interests include bio-inspired systems, hybrid systems, and the combination of evolutionary algorithms and neural networks.

HOSSAM FARIS received the B.A. degree in computer science from Yarmouk University, in 2004, the M.Sc. degree (Hons.) in computer science from Al-Balqa' Applied University, in 2008, and the Ph.D. degree in e-business from the University of Salento, Italy, in 2011. He is currently a Full Professor with the Information Technology Department/King Abdullah II School for Information Technology, The University of Jordan, Jordan. In 2016, he has worked as a Postdoctoral Researcher with the Information and Communication Technologies Research Center (CITIC), GeNeura Team, University of Granada, Spain. His research interests include applied computational intelligence, evolutionary computation, knowledge systems, data mining, semantic web, and ontologies. He was awarded a full-time, competition-based Ph.D. Scholarship from the Italian Ministry of Education and Research to pursue his Ph.D. degree.