IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received 11 November 2023, accepted 5 January 2024, date of publication 12 January 2024, date of current version 19 January 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3353312

RESEARCH ARTICLE

The Impact of the Weighted Features on the Accuracy of X-Platform's User Credibility Detection Using Supervised Machine Learning

NAHID R. ABID-ALTHAQAFI[1], HESSAH A. ALSALAMAH[1], AND WALAA N. ISMAIL[2]
[1]Information Systems Department, College of Computer and Information Sciences, King Saud University, Riyadh 14511, Saudi Arabia
[2]Department of Management Information Systems, College of Business Administration, Al Yamamah University, Riyadh 13541, Saudi Arabia

Corresponding authors: Nahid R. Abid-Althaqafi (442204695@student.ksu.edu.sa) and Hessah A. Alsalamah (halsalamah@ksu.edu.sa)

This work was supported by the Deputyship for Research and Innovation, Ministry of Education, Saudi Arabia, under Project IFKSUOR3–409–2.

ABSTRACT Social media represent a vital actor in our lives, often serving as a primary source of information, surpassing traditional sources. Among these platforms, the X-Platform, which used to be called Twitter, has emerged as a leading space for the exchange of opinions and emotions. In this study, we introduced a supervised machine learning system designed to detect user credibility in this influential platform. User credibility detection depends largely on the features of the users on the platform. Feature weighting plays a pivotal role in identifying the significance of each feature in a dataset. It can indicate irrelevant features, which can lead to better performance in classification problems. This study aims to highlight the impact of weighted features on the accuracy of X-Platform User Credibility Detection (XUCD) using supervised machine learning methods, such as Principal Component Analysis (PCA) and correlation-coefficient algorithms, and tree-based methods, such as (ExtraTressClarifier) to extract new weighted features in the dataset and then use them to train our model to discover their impact on the accuracy of user credibility detection issues. As a result, we measured the effectiveness of different feature-weighting methods on different dataset categories to determine which obtained the best detection accuracy. Experiments were conducted on real user profiles, and statistical and emotional information was extracted from a publicly available dataset called (ArPFN). The improvement in XUCD accuracy using different weighting methods was dependent on the method and dataset category used.

INDEX TERMS Feature engineering, feature weighting, social network, supervised machine learning, user credibility, X-platform.

I. INTRODUCTION

Detecting credibility among online social network (OSNs) accounts is a crucial task, as it identifies trustworthy sources of information that the audience can rely on. This distinction is vital for mitigating the dissemination of misinformation and fake news, which can have detrimental effects on people in this era, where online platforms replaced conventional sources of information. X-Platform is considered a significant source of information appealing to a wide range of audiences. Consequently, the detection of untrustworthy X-platform users holds particular significance for combating the spread of misinformation within this audience.

Automatic detection of X-Platform User Credibility (XUC) is essential because of the large amount of data required to be processed and handled on such a platform. Machine Learning (ML) algorithms are frequently used to effectively identify patterns and extract valuable insights from data [1]. However, it is crucial to acknowledge that their effectiveness depends heavily on the quality of the dataset. When a dataset includes irrelevant or noisy information, deriving trustworthy knowledge becomes challenging [2]. As a result, the process of data preprocessing, which involves converting raw data into a useful and comprehensible format,

The associate editor coordinating the review of this manuscript and approving it for publication was N. Ramesh Babu.

has emerged as a pivotal phase in the application of ML algorithms [2]. Traditionally, it is assumed that all features have equal significance when estimating the output. Nonetheless, when certain features exhibit greater importance than others, they can influence the results and potentially impact the overall algorithm's performance and accuracy [3].

Not all features contribute equally to predicting the correct class. It is necessary to weigh the features in the dataset and use them to prove the accuracy of the classifier. Many weighting techniques are available, such as ExtraTreesClassifier, Principal Component Analysis (PCA), and Correlation-Coefficient-based algorithms. However, the effectiveness of these techniques in detecting user credibility remains unclear. In this study, we focused on X-Platform User Credibility Detection (XUCD); therefore, experiments were conducted to evaluate the impact of feature weighting methods on the performance of user credibility detection using Supervised Machine Learning (SML). By investigating these techniques, we aimed to comprehensively explore different facets of feature weighting and assess their effectiveness in enhancing X-Platform User Credibility (XUC) detection within the context of our study. Additionally, different categories of features and their combinations were considered. them yourself, just to ensure that the right heading style is applied.

II. RESEARCH BACKGROUND

A. USER CREDIBILITY DETECTION

Credibility can be defined in many ways, but it essentially means being seen as believable and trustworthy. In other words, it is about offering reasonable grounds for belief. User credibility in OSNs concerns the positive qualities of a user or news provider that makes their followers trust them. User credibility can be described as the willingness of people to trust the user of an OSN as a reliable source of information [3], [4], [5]. It is important for online communication, especially in social media, where people can share anything they want.

User Credibility Detection is the process of understanding and measuring user credibility in an OSNs. It involves identifying the different features that can be used to determine whether a user is credible or not, and can be categorized as

1) CONTENT-BASED FEATURES

These features include content quality and relevance as well as language and tone [4], [5], [6].

2) INTERACTION-BASED FEATURES

Interaction-based features are composed of information on the following details and the reaction of the audience to the posted content [5].

3) PROFILE-BASED FEATURES

Including the details about the users that include their demographic information and their status [4], [6].

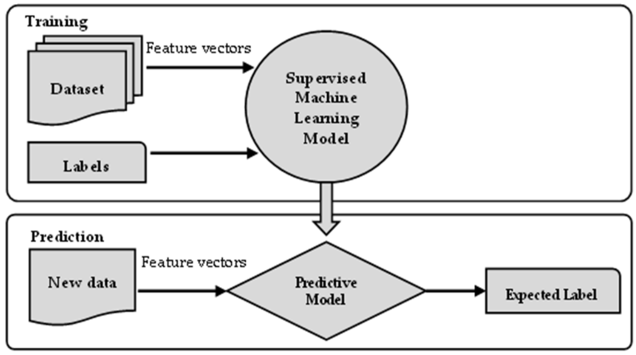

FIGURE 1. Supervised machine learning.

4) SENTIMENT-BASED FEATURES

Describe the user's views, emotions, feelings, opinions, or assessments regarding products, events, news, or services [7], [8], [9].

5) STATISTICAL INFORMATION

Some of the features listed above can be quantified as statistical information [9].

Machine learning can play a significant role in automating the extraction and analysis of UCD features. This can help improve the quality and reliability of online communication.

B. SUPERVISED MACHINE LEARNING (SML)

Supervised Machine Learning (SML) is a type of artificial intelligence that allows computers to learn from data and make predictions without explicit programming. SML algorithms are trained on a set of labeled data, where the algorithm learns to identify patterns in the data and uses them to predict the label of new data, [9] as shown in Figure 1.

SML is idle for classification problems. It has been widely used for user-credibility detection, where each user is labeled as either credible or non-credible. The algorithm learns to identify features of users that are highly associated with credibility, such as the quality of their content, engagement with other users, and reputation within the platform. Once trained, an SML algorithm can be used to predict the credibility of a new user by analyzing its features.

The most popular SML approaches used in user credibility detection include logistic regression (LR) [6], [9], [10], Support Vector Machine (SVM) [6], [9], [10], Naïve Based (NB) [6], [9], [10], Decision Tree (DT) [9], and Random Forest (RF) [6], [9], [10].

1) SUPPORT VECTOR MACHINE (SVM)

SVM, which is primarily used for classification tasks, accomplishes this by segregating the data into distinct groups. It identifies a hyperplane, often referred to as a decision boundary, that effectively divides the dataset into different groups [11], [12].

2) LOGISTIC REGRESSION (LR)

The LR method is used to estimate discrete values based on a given set of independent variables. It helps predict the probability of the occurrence of an event by fitting the data to a logit function. Its output value was between 0 and 1. LR overcomes the limitations of linear regression for better classification [12], [13].

3) NAÏVE BAYES (NB)

The NB model calculates the probabilities of the categories assigned to a given dataset based on Bayes' theorem. Subsequently, it classifies the test data [12], [14]. In the Naive Bayesian (NB) model, the process involves computing the conditional probability for each class label, and subsequently identifying the label with the highest probability as the predicted label [15].

4) DECISION TREE (DT)

DTs classify instances by sorting them according to their feature values. In this approach, each node in the tree corresponds to a feature of an instance that requires classification and each branch emanating from a node represents a possible value that the node can take on. Decision trees are a commonly employed technique for classification, in which the classification itself is represented as a tree structure known as a decision tree [12], [16].

5) RANDOM FOREST (RF)

RF is a machine learning algorithm that employs an ensemble of decision trees. This is based on the collective work of a large number of decision trees. This ensemble was generated from a casually selected subset of the training data. The model gathers votes from several decision tree approaches to determine the final class of the test dataset [12], [17], [18].

6) BOOSTING ALGORITHMS

Boosting algorithms operate as greedy techniques. Unlike RF, boosting models do not grow decision trees simultaneously. Instead, they sequentially train individual decision trees, each of which is an improved version of the previous one, to reduce the error rate. XGB is considered a powerful machine-learning algorithm. where a regularization term is used to control the complexity of the model, leading to better prevention of overfitting [19].

C. FEATURE WEIGHTING

The extraction of feature weights from a dataset can be performed through traditional methods that require expert input or by employing filter selection algorithms, which can assess the significance of each feature in the dataset. Some widely used feature-weighting methods are as follows.

1) ANALYTIC HIERARCHY PROCESS (AHP)

The AHP serves as a decision-making tool that facilitates the comparison of several criteria and determines their relative importance [20], [21]. It has been used to produce feature weights in machine learning, and recent research has indicated that AHP can effectively assign weights to features, ultimately improving the performance of machine learning models across diverse applications [9], [22], [23].

2) INFORMATION GAIN

This type of filter method is used to measure feature weights in high-dimensional feature spaces. The information gain ratio is valuable for assessing features with a large number of distinct values. However, it should be noted that this method may result in a bias favoring features with low information values [22], [24].

3) CHI-SQUARED TEST

The chi-square test is a statistical tool used to determine the significant association between a categorical feature and the target variable. It can be used to assess the importance of features in the dataset [22], [24].

4) EXTRA TREES CLASSIFIER

The extra-tree classifier is a type of DT algorithm that is primarily used for classification tasks. A key advantage of feature weighting is its ability to rank features based on their importance [25].

D. USER CREDIBILITY DATASETS

Several online datasets can be used to train the SML model to detect X-Platform User Credibility (XUC). The most suitable datasets were as follows.

1) CREDBANK

This dataset contained tweets from October 2014 to February 2015. Tweets about events, and they were classified based on the credibility rating of these events [26].

2) FAKENEWSNET

This dataset contains fake news stories and X-Platform users who share them, along with their profiles, timelines, followers, and following information [27].

3) ARPFN

This dataset contains X-Platform users and the number of fake and true tweets they share, along with 39 features for each user, grouped into profile, text, emotional, and statistical features [28].

4) PHEME

This dataset contains tweets related to breaking news events along with their credibility and veracity ratings. It also has tweet-level features, such as sentiment, source, and retweet count [29].

Table 1 presents a comparison of the main characteristics of the aforementioned datasets.

TABLE 1. Dataset comparison.

Dataset	Description	Size	# Of Features	Feature's Types				Labels
				Content-based	*User-based*	*Interaction-based*	*Sentiment-based*	
CredBank [26]	Real-world events	60 M tweets	6	√				• Certainly inaccurate • Most likely inaccurate • Uncertain • Most likely accurate • Certainly accurate
FakeNewsNet [27]	Tweets from two domains: politics and entertainment and the corresponding images	510 users	7	√	√	√		• Real • Fake
ArPFN [28]	Arabic tweets include articles related to politics, health, and other topics.	1546 users	39	√	√	√	√	• Prone to spread fake news • Not prone to spread fake news
PHEME [29]	Tweets related to five breaking news stories	5,802 tweets	5	√		√		• Rumors • Non-rumors

III. LITERATURE REVIEW

Users' credibility is a significant factor in deciding whether the information is trusted. Given that in OSNs, a large amount of information comes from unknown individuals who lack a proven indicator of their credibility, it is challenging to determine the trustworthiness of this information. Hence, the credibility of information relies on the reliability of its source. Therefore, automated UCD techniques have been addressed by a significant number of scientific papers in literature. For example, by a simple search in the Google Scholar database with the keywords (user's credibility + detection + X-Platform, between 2015 and 2023), 17300 related articles were obtained. In this section, we discuss only those studies that are most relevant to our work.

A. X-PLATFORM USER CREDIBILITY DETECTION METHODS

The literature addresses user credibility detection in OSNs using different methods. Many studies use machine learning techniques, particularly SML algorithms, such as SVMs [4], [30], [31], [32], [33], [34], [35], NB [33], RF [6], [9], [36], [37], [38], [39], XGBoost [40], [41], [42], [43], LR [42], [44], [45], and DT [4], [46], [47], [48]. Some studies have also used ensemble models [41], [49] that combine the predictions of multiple SML algorithms.

Hybrid approaches that combine SML with other techniques are popular. For example, [33] used a graph-based approach to analyze the relationships among users, products, and shops to calculate the credibility of customers. In [47] A node2vec graph embedding algorithm was used to extract features from the X-Platform followers/followed graph, and then these features were combined with user features provided by X-Platform to create a hybrid model that reflects both the user's features and their social graph. Other hybrid approaches are based on feature hybridization. For example, [41] sentiment analysis combined with social network features was used to identify features that can be used in XUCD. They applied a reputation-based technique to each user profile and assigned a sentiment score based on the user history. The CredRank algorithm proposed in [50] analyzes user behavior to measure user credibility in OSNs. In the same context, [51] the User Credibility (UCred) model uses both machine learning and deep learning methods to improve the accuracy of X-platform user credibility detection. In their study, they generated their output using RoBERT (Robustly optimized BERT), Bi-LSTM (Bidirectional LSTM), and RF (Random Forest), and then fed this output into a voting classifier to achieve their goal. In addition, feature hybrids have been used in [42] by combining sentiment analysis and social networks to find features that can be used for XUCD. This study applied reputation-based techniques and assigned a sentiment score to each user based on the history of the user's profile. Reputation features have also been shown to be useful for UCD [41] when a probabilistic reputation feature model is proposed. This model outperformed raw reputation features in terms of accuracy. In addition, [43] integrated semantic and sentiment analyses are used to estimate and predict domain-based analysis of user content in social big data. Reference [35] used a hybrid approach that combined sentiment analysis and machine learning to identify the credibility of both user profiles and content.

B. FEATURE WEIGHTING

Feature weighting has been a subject of study in various studies. In [9], a credibility formula was introduced for Facebook users involving several parameters, each of which was assigned a specific weight. These weights were determined through the application of the Analytical Hierarchical Process (AHP) approach, which is grounded in credibility theory. Using this formula, users' accounts are ranked based on their credibility, allowing for the prediction of trust and credibility levels among Facebook users. Furthermore, in a related context, [52] introduced an enhanced version of the AHP called the Interval Type-2 Fuzzy Analytical Hierarchy Process. This method was used to rank online reviewers concerning their credibility and address the issue of reviewer credibility.

Moreover, [53] proposed a model for assessing the credibility of publications on various social networks. The credibility analysis was based on three metrics: text credibility, user credibility, and social credibility. They introduced a model for measuring text credibility from social network information sources (TCred), represented as follows:

$TCred(t) = (weight_{text} * TextCred) + (weight_{user} \times UserCred) + (weight_{social} \times SocialCred)$

Here, $weight_{text,}$ $weigh_{user}$, and $weight_{social}$ represent the weights assigned to the text credibility, user credibility, and social credibility, respectively. TextCred, UserCred, and SocialCred represent credibility measures associated with text, user, and social impact, respectively. The user of the model is responsible for defining the values of the parameters and weights within the system.

Another study [54] aimed to detect fake news using opinion mining in which user credibility scores were calculated and used. The user credibility equation comprises of three components: user reputation, user influence, and user comments. Each component was assigned a specific weight. User comments carried a lower weight (0.2) because they did not directly reflect a user's credibility, whereas user reputation and influence both had the same weight (0.4) because they more directly indicated a user's credibility.

CredRank, as proposed in [50], evaluates user credibility by identifying similarities in online behavior. It is designed to detect coordinated behavior on social media and assign lower credibility weights to users engaged in such behavior. Coordinated users can suppress other users and hinder the spread of their content, potentially leading to the dissemination of misleading information.

In addition, [55] the information entropy method is employed to assign weights to various feature items. They considered four factors in their model for evaluating user credibility: social relationship strength, social influence scope, information value, and information transmission control. However, determining the optimal weights for these factors remains a challenge [56].

C. FEATURES RELATED TO XUCD

The concept of user credibility on social media platforms, such as X-Platform, has garnered substantial research attention. In [57], a set of language-independent features, extracted from four different languages, was utilized to investigate the nature and characteristics of spam profiles in X-Platform and to enhance spam detection.

Reference [41] introduced a novel probabilistic reputation feature model that focused on user reputation. The analysis of user reputation within a social network was also addressed by [8], who delved into a user's reputation regarding a specific topic while also assessing the user's profile and sentiment to identify trustworthy sources of topic-related information.

In another approach, [30] introduced a method for rating user credibility in his/her X-platform profile. This method considers the content, images, links, sentiment, and profile features. Moreover, [46] the identified key tweet features affect credibility, including the **user's** duration time on X-Platform**,** posting frequency, friend/follower counts, and the number of retweets received.

Examining tweets related to various events revealed [58] that credibility is strongly associated with the inclusion of URLs, mentions, retweets, and tweet length. Additionally, [59] users often base their credibility perceptions on easily identifiable information such as their username and profile picture.

Another study [39] aimed to calculate users' credibility scores based on factors such as users' social profiles, tweet credibility, number of likes and retweets, and sentiment scores. They suggested that a higher user-credibility score signifies a greater influence and credibility.

The detection of XUCD was also addressed in [60], where sentiment features, the presence of emojis, hashtags, and political bias in user tweets were considered in the detection process. Conversely, [61] features such as the number of followers, tweet production volume, and ratio of tweet count to the account's creation duration in days significantly influenced credibility judgments, with the number of followers being the most impactful feature.

D. LITRATURE OVERVIEW

Finally, Table 2 provides a comparative overview of the most relevant studies in the domain of XUCD, focusing on the primary objectives of the proposed solutions.

IV. MATERIALS AND METHODS

This research aims to investigate the impact of feature weighting on the accuracy of XUCD by using embedded methods such as the ExtraTreeClassifier, Correlation coefficient, and PCA to determine the importance of each feature in our dataset, which has been used to calculate feature weights that enabled us to transform features into weighted features. The feature-weighting process was performed between feature extraction and classification. This is the process of generating new weighted datasets that have been used to evaluate their impact on prediction results. Our hypothesis was based on the fact that treating all features equally may reduce the accuracy of the model. It is hoped that feature weighting will increase model accuracy in XUCD. Figure 3 illustrates the main stages of the research methodology.

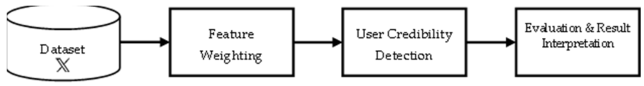

FIGURE 2. Research stages.

A. DATASET

We plan to use the ArPFN dataset [28] for our experiments because it is the most recent dataset and has the most features. The ArPFN is a real-world dataset created [42] in three stages. First, they collect a set of verified Arabic claims from

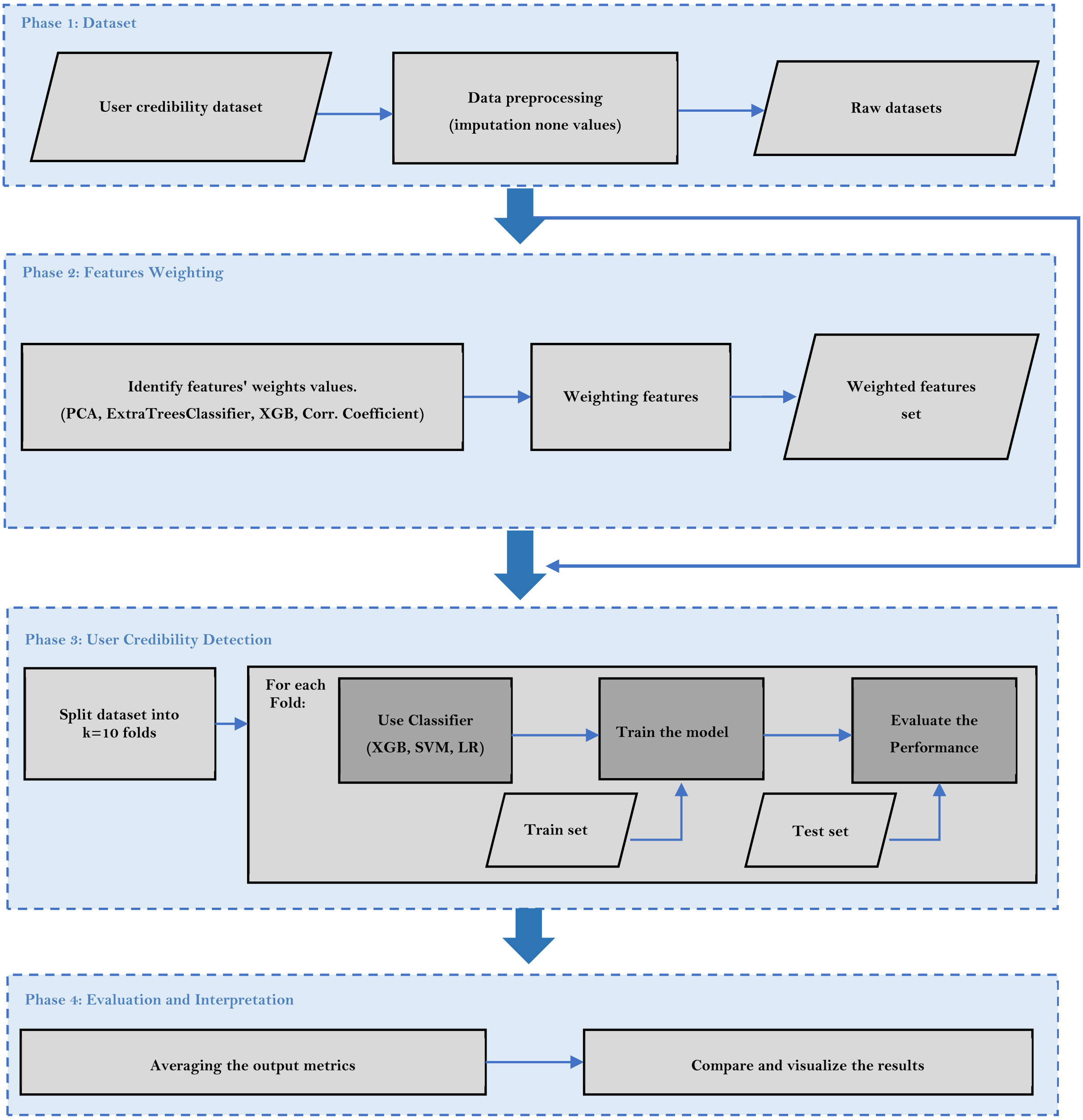

FIGURE 3. Flow diagram of the proposed model.

various sources. They then used these claims to find tweets that spread them. Finally, they identified the users associated with these tweets and labeled them credible or non-credible based on how often they tweeted fake news. The ArPFN contains 1,546 X-platform user accounts, of which 541 are non-credible (prone to spreading fake news) and 1,005 are credible (not prone to spreading fake news).

As shown in Table 3, the dataset had three feature categories for each user: profile, emotional, and statistical.

B. FEATURE WEIGHTING

This phase focused on estimating the importance of the features in the UCD. Each category of feature is processed individually and combined with the other categories, which results in seven different sub-datasets of features as follows:

Datasets: {(profile features), (emotional features), (statistical features), (Profile and Emotional features), (Emotional and Statistical features), (Profile and Statistical features), (Profile, Emotional, and Statistical features)}.

Two alternatives have been considered for each sub-dataset through this phase.

First: considering raw data.

Second: Using ML feature importance estimator methods, such as ExtraTreeClassifier, correlation coefficient, and principal component analysis (PCA), are used to generate weighted feature datasets. This approach consists of the following steps:

TABLE 2. Credibility detection in literature.

Ref	Year	Purpose	Classifier	Weighting	Data Categories	Performance
[62]	2022	Content credibility	ensemble regression algorithm	√	-	96.29%. by Ensemble model
[63]	2022	Spam detection	DT, SVM, RF	-	√	93% by All features and RF
[51]	2022	User credibility	RoBERT, Bi-LSTM, RF	-	-	98.96%
[48]	2022	Content credibility	DT, KNN, RF, SVM	-	-	Best improvement after Feature Selection with DT approximately 21%
[42]	2022	User credibility	XGB, RF, LR, NN	-	√	79% by All features and XGB
[60]	2022	Content credibility	KMEANS, LSI, NMF, LDA	√	-	74% by NMF
[55]	2021	User credibility	SVM	√	-	93.5%
[57]	2021	Spam detection	KNN, RF, NB, DT, MLP	-	-	An average of 97.1% by KNN
[44]	2021	Spam detection	DT, SVM, KNN, LR, AdaBoost, NB	-	-	88% by LR
[39]	2021	User credibility	RF, MLP, SVM, LR	-	-	96% by RF
[40]	2021	Content credibility	RF, SVM, LR, DT, XGB, KNN, NB, LDA	-	√	77% by All features and LR
[6]	2021	Content credibility	NB, SVM, KNN, LR, RF, ME, CRF	-	√	83.4% by All features and RF
[64]	2020	Content credibility	LR, SVM, MLP, KNN, DT, LSVM, XGB, AdaBoost	-	-	On average, individual learners achieved an accuracy of 85%, but ensemble learners achieved an accuracy of 88.16%
[65]	2020	User credibility	SVM, LR, RF	-	√	72% by All features and SVM
[47]	2020	User credibility	DT, RF, KNN, Bayes, QDA, SVM, LR, LDA	-	√	98% by All features and LDA

TABLE 2. *(Continued.)* Credibility detection in literature.

Ref	Year	Purpose	Classifier	Weighting	Data Categories	Performance
[4]	2020	Content credibility	NB, SVM, LR, DT	-	√	88.4% by All features and DT
[66]	2020	User credibility	LR, KNN, RF, DT, SVM	-	√	68% by All features and DT
[54]	2020	Content credibility	RNN	√	-	97%
[49]	2020	User credibility	LR, SVM, DT, MLP, KNN LSTM, Bi-LSTM	-	-	70%
[45]	2020	User credibility	LR, SVM, RF, XGB	-	-	77.75% by ensemble model
[43]	2020	User credibility	NB, LR, DT, RF, XGB, GLM	-	-	GLM achieves the best results
[38]	2020	Content credibility	RF, NB, SVM		√	97.7% by RF
[35]	2020	User credibility	SVM, N, DT	-	-	68% by SVM
[32]	2020	User credibility	SVM, LR , RF, NB, KNN	-	√	84.9%
[24]	2020	Content credibility	DT, NB, SVM	-	-	85.29% by SVM with Chi-square feature selection
[9]	2020	User credibility	DT, RF, LR, KNN, NB, SVM, NN	√	-	94% by NN
[30]	2020	User credibility	NB, DT, SVM, KNN	-	-	68% by SVM with
[53]	2019	Content credibility	-	√	-	-
[67]	2019	Spam detection	KNN, SVM	-	-	87.6% by KNN
[36]	2019	Spam detection	RF, DT, BN, KNN, SVM	-	-	91.2% RF
[10]	2018	Content credibility	SVM, LR, RF, KNN, NB	-	√	78.4% by All features and RF
[68]	2017	Spam detection	SVM, DT, NB, RF	-	-	64.84% by NB
[8]	2017	User credibility	NB, LR	-	-	82.91%
[7]	2017	User credibility	LR, SVM, NB, RF, NN	-	√	78.42% by All features and NN
[33]	2015	User credibility	SVM, NB	-	√	90.8% by Base + Emotion features
[50]	2013	User credibility	CredRank algorithm	√	-	-
[46]	2011	Content credibility	SVM, DT, DR, NB	-	√	86% by All features and DT

TABLE 3. ArPFN feature types [28].

Feature Type	Number of Features	Description
Profile Features	**17**	Such as the user's identifier, verification status, follower and following counters, and the frequency of the user's tweets.
Emotional Features	**11**	Include anger, anticipation, disgust, fear, joy, love, optimism, pessimism, sadness, surprise, and trust.
Statistical Features	**11**	Describe the users' influence, and activities such as the ratio of user tweets that contain hashtags, the average number of hashtags per tweet, the ratio of user tweets that are replies, the ratio of user tweets that contain URLs or media such as images or videos, number of tweets that are retweets.

Finding the weight for each feature by using the following equation.

$$Wi = Ii/\sum\nolimits_{i=1}^{T} Ii \tag{1}$$

where:

- Wi: Feature(i) weight,
- T: Number of features in sub-dataset, and
- Ii: Feature importance score

While $\sum_{i=1}^{T} \mathrm{Wi} = 1$

Calculate the weighted features based on their weights by multiplying each feature value by its weight.

$$WFi = Fi^{*}Wi \tag{2}$$

where:

- WFi: Weighted Feature(i),
- Fi: Feature (i) value, and
- Wi: Feature (i) weight.

User Credibility Detection

This phase of the research focused on developing an SML model that can distinguish between credible and non-credible users on the X platform. We chose to use SML because it has been proven to be highly accurate for classification problems, as shown in the literature review. To obtain a better and more generalizable model, we trained it using a 10-fold cross-validation method, which has been shown to reduce overfitting. We then compared the most common classification algorithms, such as XGB, SVM, and LR, to determine the most accurate for our datasets.

C. IMPLEMENTATION, EVALUATION, AND INTERPRETATION OF THE RESULTS

Python has been used for model implementation using open-source libraries such as Scikit-learn and Matplotlib. In this phase, we use various evaluation metrics to determine the effectiveness of the proposed method. These evaluation metrics, including the accuracy, precision, recall, and F1 score, were used to validate each alternative from the previous phase. Python visualization tools, such as bar plots, heatmaps, and confusion matrix visualization, were used to analyze and visualize the results. Figure 3 shows the flow diagram of the proposed model.

V. RESULTS

Using the proposed methodology, we determined the impact of various feature weighting methods, different feature categories, and three SML classifiers on the accuracy of XUCD, which can be summarized as follows:

A. FEATURE WEIGHTING

1) FEATURE WEIGHTING METHODS

In general, feature-weighting algorithms do not reduce the dataset dimensionality. Instead, they assigned weights to each feature based on their importance in predicting the correct labeled class. Unless features with extremely low weights are explicitly removed from the dataset at the outset, we assume that each feature has some level of importance in the induction process, with the magnitude of its weight reflecting its degree of significance. In this study, we used several machine learning algorithms to weigh features according to their importance in detecting X-platform user credibility (XUC). As described in Appendix, we considered three of the most popular methods for calculating the feature importance score.

a: CORRELATION COEFFICIENTS

A Logistic regression algorithm was used to assess the correlation coefficients of the model. The significance of a feature in detecting XTUC is determined by the magnitude of its coefficient, whether it is positive or negative. A coefficient of zero indicates that the feature has no influence on the detection. In this approach, we considered the absolute value of each significant score.

b: TREE-BASED MODEL

ExtraTreeClassifier and XGBClassifier were employed to train the model and obtain the important scores for each feature.

c: PRINCIPAL COMPONENT ANALYSIS (PCA)

The importance of features within the datasets was determined using the first principal component (PC1), which signifies variance.

2) XUCD ACCURACIES USING WEIGHTING METHODS

We utilized the various methods mentioned above for the sub-datasets within our model. As depicted in Table 4 and Figure 4, the results revealed that in the most favorable scenario, the ExtraTreeClassifier was able to positively

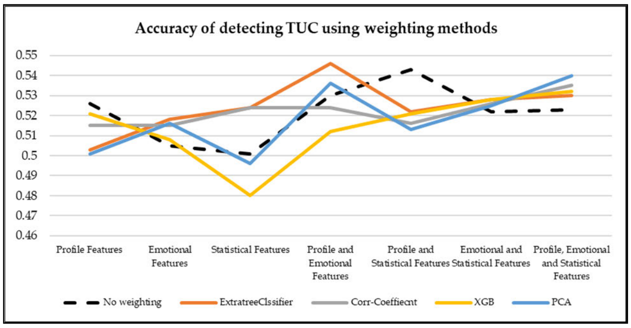

FIGURE 4. Accuracy comparison after features weighting.

TABLE 4. Accuracy after applying weighting methods.

Dataset Category	No weighting	Weighting Methods			
		ExtraTree Clssifier	*Corr-Coefficient*	*XGB*	*PCA*
Profile Features	**0.526**	0.503	0.515	0.521	0.501
Emotional Features	0.505	**0.518**	0.515	0.508	0.516
Statistical Features	0.501	**0.524**	**0.524**	0.480	0.496
Profile and Emotional	0.530	**0.546**	0.524	0.512	0.536
Profile and Statistical	**0.543**	0.522	0.516	0.521	0.513
Emotional and Statistical	0.522	**0.528**	0.526	**0.528**	0.525
Profile, Emotional, and Statistical	0.523	0.530	0.535	0.532	**0.540**

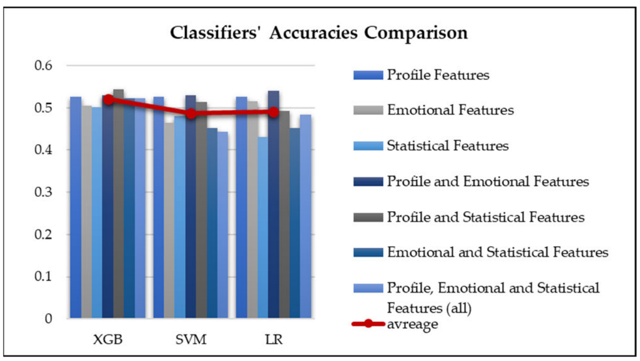

FIGURE 5. Classifiers' accuracies comparison.

influence five out of seven groups, enhancing the accuracy of XUCD.

3) DATASET CATEGORY IMPACT ON XUCD

As seen in Table 4 and Figure 4, the most significant dataset category was the combination of weighted profile and emotional features using ExtraTreeClassifier, which achieved the highest accuracy score (0.546) among all datasets. It is important to note that using a combination of feature categories

TABLE 5. SML classifier metrics for XUCD.

	Dataset Category	Accuracy	F1_score	Recall	Precision
XGB	Profile	0.526	0.346	0.356	0.336
	Emotional	0.505	0.345	0.355	0.325
	Statistical	0.501	0.331	0.361	0.311
	Profile and Emotional	0.530	0.300	0.290	0.320
	Profile and Statistical	**0.543**	0.333	0.333	0.343
	Emotional and Statistical	0.522	0.312	0.302	0.332
	Profile, Emotional and Statistical	0.523	0.333	0.333	0.343
SVM	Profile	0.526	0.506	0.526	0.526
	Emotional	0.465	0.455	0.465	0.475
	Statistical	0.481	0.471	0.481	0.481
	Profile and Emotional	**0.530**	0.530	0.530	0.530
	Profile and Statistical	0.513	0.503	0.513	0.513
	Emotional and Statistical	0.452	0.452	0.452	0.462
	Profile, Emotional and Statistical	0.443	0.443	0.443	0.443
LR	Profile	0.526	0.526	0.526	0.526
	Emotional	0.515	0.515	0.515	0.515
	Statistical	0.431	0.431	0.431	0.431
	Profile and Emotional	**0.540**	0.540	0.540	0.540
	Profile and Statistical	0.493	0.493	0.493	0.493
	Emotional and Statistical	0.452	0.452	0.452	0.452
	Profile, Emotional and Statistical	0.483	0.483	0.483	0.483

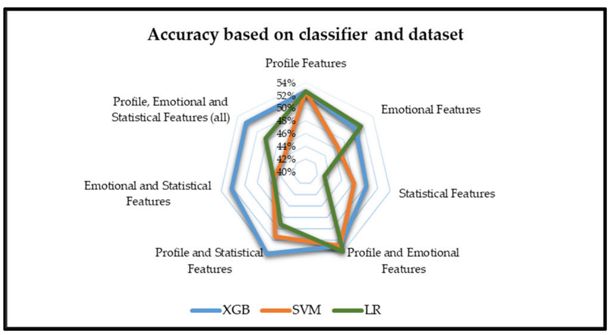

FIGURE 6. Influence of dataset on classifiers.

TABLE 6. Classifiers' accuracies comparisons.

Dataset Category	XGB	SVM	LR
Profile Features	**0.526**	**0.526**	**0.526**
Emotional Features	0.505	0.465	**0.515**
Statistical Features	**0.501**	0.481	0.431
Profile and Emotional Features	0.530	0.530	**0.540**
Profile and Statistical Features	**0.543**	0.513	0.493
Emotional and Statistical Features	**0.522**	0.452	0.452
Profile, Emotional, and Statistical Features (all)	**0.523**	0.443	0.483
Average	**0.521**	**0.487**	**0.491**
Range	**0.042**	**0.087**	**0.109**

TABLE 7. Profile features importance.

Feature	Importance			
	Extra-Tree-Classifier	PC1	XGB-Classifier	Coefficient
statuses_count	0.069467	0.22249	0.064523	0.019530
followers_count	0.068724	-0.47140	0.052990	0.148797
following_count	0.067434	0.552899	0.055653	0.228387
favourites_count	0.067370	0.195748	0.057379	0.138202
listed_count	0.059010	-0.39781	0.054106	0.241255
default_profile	0.055609	0.826927	0.048254	0.081021
verified	0.023388	-0.20844	0.073554	0.027216
tweet_freq	0.076701	0.299807	0.082639	0.345976
follower_growth_rate	0.070163	-0.31962	0.060078	0.226247
following_growth_rate	0.064454	0.576509	0.068358	0.250871
listed_growth_rate	0.059918	-0.40486	0.052574	0.157493
followers_following_ratio	0.071033	-0.22221	0.064449	0.089700
screen_name_length	0.057932	0.835513	0.061314	0.197955
digits_in_screen_name	0.051480	0.543096	0.050996	0.124075
name_length	0.056769	0.756989	0.063265	0.318861
digits_in_name	0.015965	0.062482	0.034975	0.058524
description_length	0.064581	0.592333	0.054893	0.153656

provides higher accuracy scores than using the individual category dataset in the detection of XUC, as shown in Table 4.

B. SML CLASSIFIERS COMPARISON

Three supervised machine-learning classifiers, SVM, LR, and XGBoost, were trained using the datasets. The metrics for detecting the credibility of X-platform users are listed in Table 5.

Table 6 and Figure 5 present a comparison of the accuracies of the three classifiers. The XGB surpasses the others with the highest accuracy of 0.543 when utilizing a combination of profile and statistical features. LR comes in close second, with an accuracy of 0.540, when working with a dataset that incorporates a combination of profile and emotional features. Bringing up the rear is an SVM, with its peak accuracy reaching 0.530, which is achieved when using the profile and emotional features.

TABLE 8. Emotional features importance.

Feature	Importance			
	Extra_Tree_Classifier	PC1	XGB_Classifier	Coefficient
anger	0.08714036	-0.748032	0.091850	0.079197
anticipation	0.09008217	-0.021049	0.091316	0.007921
disgust	0.09003102	-0.691841	0.092702	0.000230
fear	0.09169413	-0.305098	0.087790	0.009871
joy	0.08867907	0.678513	0.079992	0.010210
love	0.08863126	0.688775	0.089939	0.017048
optimism	0.08706432	0.792609	0.090151	0.090849
pessimism	0.0927801	-0.607379	0.095095	0.073498
sadness	0.0886634	-0.408814	0.089391	0.109658
surprise	0.10087149	-0.149469	0.091200	0.085795
trust	0.09436268	0.620704	0.100576	0.097664

TABLE 9. Statistical features importance.

Feature	Importance			
	Extra_Tree_Classifier	PC1	XGB_Classifier	Coefficient
non_duplicate	0.081137	0.938383	0.081370	0.053175
hashtags_prop	0.083582	0.721684	0.080343	0.030145
hashtags_per_twt_prop	0.084789	0.887875	0.075781	0.119263
mentions_prop	0.081542	0.296642	0.098349	0.189141
n_unique_mentions	0.080722	-0.078769	0.082353	0.035708
replies_prop	0.083115	0.245487	0.095809	0.201123
urls_prop	0.084691	0.800061	0.079636	0.134987
qt_prop	0.084959	0.165531	0.071311	0.212587
media_prop	0.086006	0.714525	0.089208	0.031207
count_RT	0.075064	-0.405484	0.084631	0.112783
avg_engagement	0.089295	0.071536	0.082040	0.035847
avg_days_between_tweets	0.085098	0.824354	0.079169	0.244553

Our experiments confirmed that XGB achieved the best performance among the examined supervised machine learning classifiers, with an average accuracy of 0.521 in XUCD, as shown in Figure 5.

The accuracy of the classifiers varied based on the processed dataset. Figure 6 illustrates the impact of different dataset categories on the detection accuracy of the three classifiers. This visual representation highlights how the choice

TABLE 10. Profile and emotional features importance.

Feature	Importance			
	Extra-Tree-Classifier	PC1	XGB-Classifier	Coefficient
statuses_count	0.039327	0.267008	0.031282	0.044307
followers_count	0.037847	-0.48678	0.030799	0.130062
following_count	0.039896	0.494249	0.030602	0.246210
favourites_count	0.038407	0.216109	0.043984	0.162859
listed_count	0.033375	-0.37319	0.031046	0.240590
default_profile	0.032085	0.722128	0.026503	0.079294
verified	0.017155	-0.18015	0.033121	0.027203
tweet_freq	0.041360	0.291773	0.049919	0.358294
follower_growth_rate	0.036749	-0.34956	0.035556	0.226614
following_growth_rate	0.030480	0.500790	0.029958	0.247251
listed_growth_rate	0.033085	-0.39340	0.026564	0.154487
followers_following_ratio	0.035892	-0.2502	0.045346	0.105857
screen_name_length	0.029171	0.725944	0.041202	0.221478
digits_in_screen_name	0.031001	0.467226	0.026393	0.133762
name_length	0.032804	0.660009	0.036053	0.334829
digits_in_name	0.009674	0.030969	0.018002	0.054162
description_length	0.034760	0.520620	0.030237	0.145394
anger	0.036031	-0.31885	0.033006	0.092516
anticipation	0.037643	-0.08138	0.036552	0.010996
disgust	0.032316	-0.3109	0.035198	0.010568
fear	0.034952	-0.19821	0.034864	0.034413
joy	0.031191	0.261158	0.028675	0.006191
love	0.040167	0.367017	0.035657	0.052126
optimism	0.033693	0.464623	0.032898	0.146854
pessimism	0.042665	-0.31707	0.033348	0.034322
sadness	0.035468	-0.20401	0.034867	0.105359
surprise	0.039459	-0.18181	0.038967	0.109934
trust	0.041257	0.358613	0.049685	0.106253

of dataset category can affect the performance of the classifiers.

VI. DISCUSSION

This section discusses the findings of our research on feature-weighting XUCD. We investigated the impact of various feature-weighting methods including logistic regression coefficients, tree-based models, and PCA. We used the ArPFN dataset, which contains profile, emotional, and statistical features, to conduct the experiments. The key findings and their implications are as follows:

A. FEATURE WEIGHTING

1) EFFECTIVENESS OF THE METHOD

In this study, the effectiveness of feature weighting techniques, such as logistic regression coefficients, tree-based models, and PCA, proved pivotal in assigning weights or importance scores to features. These weights were subsequently utilized by the model to generate new weighted sub-datasets, which were then employed in the training to assess their influence on the accuracy of XUCD. Notably, the use of these methods, particularly tree-based algorithms, had a positive impact on the detection accuracy of our model. As depicted in Table and Figure 4, our findings indicate that five out of seven sub-datasets demonstrated improved performance when utilizing feature weighting methods. Conversely, two out of the seven sub-datasets showed a decrease in performance when any of the four weighted methods were applied. Hence, this finding confirms that the tree-based weighting methods outperformed the other methods in most cases, except in one case (using all 39 features), where PCA achieved the highest performance in XUCD.

TABLE 11. Profile and statistical features importance.

Feature	Importance			
	Extra-Tree-Classifier	PC1	XGB-Classifier	Coefficient
statuses_count	0.041236	0.084003	0.033235	0.040307
followers_count	0.036925	-0.13366	0.028960	0.139386
following_count	0.034167	0.354876	0.029456	0.218529
favourites_count	0.040598	-0.10842	0.040080	0.144842
listed_count	0.031300	-0.15813	0.033971	0.233072
default_profile	0.033121	0.759889	0.027994	0.016460
verified	0.015099	-0.09722	0.044911	0.024245
tweet_freq	0.042969	0.144848	0.044463	0.347872
follower_growth_rate	0.033766	-0.05000	0.029796	0.232094
following_growth_rate	0.036537	0.391909	0.035088	0.197840
listed_growth_rate	0.032559	-0.14539	0.032735	0.139385
followers_following_ratio	0.038404	-0.01585	0.043053	0.070163
screen_name_length	0.031611	0.783362	0.037070	0.498140
digits_in_screen_name	0.030902	0.470015	0.028569	0.234822
name_length	0.035326	0.737413	0.025384	0.529860
digits_in_name	0.010578	0.024306	0.013113	0.067666
description_length	0.032611	0.500808	0.038921	0.235759
non_duplicate	0.027948	0.680480	0.028817	0.144370
hashtags_prop	0.030399	0.543442	0.030746	0.007219
hashtags_per_twt_prop	0.035238	0.708320	0.035947	0.111654
mentions_prop	0.028636	0.278510	0.032539	0.227481
n_unique_mentions	0.035464	-0.084702	0.022373	0.027578
replies_prop	0.032083	0.496452	0.042613	0.186706
urls_prop	0.035930	0.465246	0.035679	0.147474
qt_prop	0.034555	0.328534	0.034133	0.305182
media_prop	0.034041	0.514044	0.028495	0.020190
count_RT	0.035785	-0.36320	0.041448	0.088047
avg_engagement	0.035764	0.138188	0.029598	0.028589
avg_days_between_tweets	0.028920	0.747824	0.034576	0.177819

2) FEATURE CATEGORY INFLUENCE

The impact of feature weighting varies across feature categories. The profile and emotional features showed the greatest improvements in accuracy, especially when using tree-based models for feature weighting. In contrast, the profile features category and the combination of profile and

TABLE 12. Emotional and statistical features importance.

Feature	Importance			
	Extra-Tree-Classifier	PC1	XGB-Classifier	Coefficient
anger	0.042134	0.234522	0.045045	0.080836
anticipation	0.042845	0.101445	0.043737	0.009514
disgust	0.042468	0.167148	0.043223	0.022644
fear	0.042615	0.182302	0.031313	0.005544
joy	0.044740	-0.24489	0.041067	0.024728
love	0.040203	-0.29100	0.036553	0.020198
optimism	0.044281	-0.34111	0.042694	0.117359
pessimism	0.044858	0.183691	0.049197	0.069348
sadness	0.043377	0.253049	0.041561	0.127879
surprise	0.048230	0.063020	0.049150	0.079603
trust	0.050192	-0.28827	0.033807	0.093783
non_duplicate	0.045310	0.895915	0.042745	0.059719
hashtags_prop	0.039221	0.695770	0.045116	0.033143
hashtags_per_twt_prop	0.038856	0.842032	0.048219	0.134912
mentions_prop	0.044805	0.278531	0.040087	0.198384
n_unique_mentions	0.040481	-0.07684	0.045512	0.056923
replies_prop	0.050344	0.208170	0.053190	0.186186
urls_prop	0.042917	0.769143	0.044459	0.127970
qt_prop	0.043552	0.148173	0.035913	0.218984
media_prop	0.042609	0.677389	0.039890	0.030810
count_RT	0.041624	-0.38273	0.061598	0.125646
avg_engagement	0.043087	0.064416	0.039700	0.034877
avg_days_between_tweets	0.041254	0.778946	0.046225	0.222520

statistical feature categories were negatively affected by the application of weighting techniques on XUCD accuracy.

B. FEATURE CATEGORY

The experimental results shown in Table 6 and Figures 5 and 6 prove that the effect of the data categories changes depending on the classifier used in the model. Using a combination of feature categories can provide the best performance in XUCD in all three classifiers, and all three classifiers confirmed that profile features must be included for the best detection of TUC.

C. XUCD CLASSIFICATION USING SML

Referring to Table 6. The best performance for XUCD was achieved using the XGB classifier because it provides the highest accuracy result with an average of 0.521. it is also important to note that XGB was least affected by dataset categories, as the difference between the accuracies obtained from different categories of datasets was the less range among other classifiers, it does not exceed 0.042 Next, in the second place, LR classifier comes with accuracy average 0.491. Finally, the worst accuracy resulted from the SVM, with an average accuracy of 0.487.

VII. CONCLUSION

A supervised machine learning framework for detecting XUC based on feature weighting techniques was proposed.

TABLE 13. Profile, emotional, and statistical features importance.

Feature	Importance			
	Extra-Tree-Classifier	PC1	XGB-Classifier	Coefficient
statuses_count	0.030106	0.080959	0.026124	0.064020
followers_count	0.024609	-0.12748	0.031755	0.123537
following_count	0.024656	0.353195	0.015623	0.237029
favourites_count	0.027679	-0.11170	0.022944	0.168775
listed_count	0.023162	-0.15373	0.020474	0.232264
default_profile	0.023110	0.758855	0.017316	0.029226
verified	0.012173	-0.09536	0.029413	0.026588
tweet_freq	0.033875	0.140759	0.035551	0.359834
follower_growth_rate	0.024563	-0.04555	0.023921	0.234860
Following-growth_rate	0.027542	0.389317	0.023809	0.193007
listed_growth_rate	0.022845	-0.14132	0.022676	0.134935
followers_following_ratio	0.022890	-0.01383	0.031082	0.086252
screen_name_length	0.019046	0.782132	0.028012	0.538029
digits_in_screen_name	0.022456	0.468250	0.021853	0.244956
name_length	0.023946	0.736520	0.024751	0.552056
digits_in_name	0.007043	0.022847	0.003564	0.062822
description_length	0.022611	0.501664	0.024843	0.231364
anger	0.021109	-0.01595	0.025957	0.096108
anticipation	0.026858	0.054585	0.027782	0.015316
disgust	0.027857	-0.05823	0.029255	0.030006
fear	0.024007	0.039109	0.023639	0.036505
joy	0.027348	-0.03449	0.027363	0.011767
love	0.024413	-0.02580	0.024403	0.046304
optimism	0.027400	0.029306	0.025594	0.172086
pessimism	0.027648	-0.04647	0.026896	0.036000
sadness	0.027583	0.056129	0.020823	0.117929
surprise	0.032147	-0.04166	0.025797	0.099754
trust	0.026645	-0.00575	0.029925	0.103833
non_duplicate	0.022248	0.682295	0.017311	0.145944
hashtags_prop	0.025037	0.544661	0.023930	0.003664
hashtags_per_twt_prop	0.022958	0.709800	0.024208	0.125918
mentions_prop	0.022059	0.278374	0.025982	0.238118
n_unique_mentions	0.026199	-0.08607	0.019488	0.042160
replies_prop	0.018867	0.495193	0.028939	0.173246
urls_prop	0.024644	0.467680	0.022216	0.127133
qt_prop	0.022778	0.327814	0.023473	0.315798
media_prop	0.020711	0.515558	0.017170	0.024007
count_RT	0.025039	-0.36270	0.029109	0.095035
avg_engagement	0.026549	0.138899	0.022947	0.024109
avg_days_between_tweets	0.018892	0.749822	0.028672	0.156742

These weighting techniques were examined to determine their impact on the detection accuracy of the proposed model. Three weighting methods were investigated: tree-based, correlation coefficient, and principal component analysis (PCA). Our findings confirmed that tree-based methods, such as ExtraTreeClassifier and XGBClassifier, were the most effective methods for achieving the highest accuracy in XUCD.

Furthermore, the ArPFN dataset provides seven different datasets according to the feature categories. Our experiments also focused on determining the impact of different feature categories on XUCD, and the results revealed that the inclusion of profile features is essential for enhancing the accuracy of XUCD.

In our model, we employ three ML classifiers: XGB, LR, and SVM. The highest performance was attained using XGB, followed by LR, with SVM ranking last in terms of the TUC detection accuracy.

In conclusion, our research provides valuable insights into the role of feature weighting methods, feature categories, and the three ML classifiers in the accuracy of XUCD. This was achieved by comparing the accuracy of the results obtained in terms of the accuracy, precision, recall, and F1 score. This paragraph has been added to the Conclusion section.

Moreover, the findings of this study not only contribute to the theoretical understanding of feature weighting in machine learning, but also provide actionable insights for practitioners seeking to implement robust credibility detection systems. As online platforms continue to evolve, the adaptability and efficacy of such systems have become paramount for safeguarding the integrity of information dissemination.

APPENDIX

Importance scores for all 7 datasets in the proposed model: See Tables 7–13.

REFERENCES

[1] J. A. Marshall, "Neural networks for pattern recognition," *Neural Netw.*, vol. 8, no. 3, pp. 493–494, 1995.

[2] S. García, J. Luengo, and F. Herrera, *Data Preprocessing in Data Mining*, vol. 72. Cham, Switzerland: Springer, 2015, pp. 59–139.

[3] M. Daszykowski, K. Kaczmarek, Y. V. Heyden, and B. Walczak, "Robust statistics in data analysis—A review: Basic concepts," *Chemometric Intell. Lab. Syst.*, vol. 85, no. 2, pp. 203–219, 2007.

[4] E. B. Setiawan, D. H. Widyantoro, and K. Surendro, "Measuring information credibility in social media using combination of user profile and message content dimensions," *Int. J. Electr. Comput. Eng. (IJECE)*, vol. 10, no. 4, p. 3537, Aug. 2020.

[5] G. Sarna and M. P. S. Bhatia, "Content based approach to find the credibility of user in social networks: An application of cyberbullying," *Int. J. Mach. Learn. Cybern.*, vol. 8, no. 2, pp. 677–689, Apr. 2017.

[6] M. Azer, M. Taha, H. H. Zayed, and M. Gadallah, "Credibility detection on Twitter news using machine learning approach," *I. J. Intell. Syst. Appl.*, vol. 3, pp. 1–10, Jun. 2021.

[7] R. Kurniati and D. H. Widyantoro, "Identification of Twitter user credibility using machine learning," in *Proc. 5th Int. Conf. Instrum., Commun., Inf. Technol., Biomed. Eng. (ICICI-BME)*, Nov. 2017, pp. 282–286.

[8] M. Alrubaian;, M. Al-Qurishi, M. Al-Rakhami, and M. M. Hassan, "Reputation-based credibility analysis of Twitter social network users," *Concurrency Comput., Pract. Exp.*, vol. 29, no. 7, p. e3873, 2017.

[9] E. A. Afify, A. Sharaf, and A. E. Khedr, "Facebook profile credibility detection using machine and deep learning techniques based on user's sentiment response on status message," *Int. J. Adv. Comput. Sci. Appl.*, vol. 11, no. 12, pp. 622–637, 2020.

[10] N. Y. Hassan, W. H. Gomaa, G. A. Khoriba, and M. H. Haggag, "Supervised learning approach for Twitter credibility detection," in *Proc. 13th Int. Conf. Comput. Eng. Syst. (ICCES)*, Dec. 2018, pp. 196–201.

[11] S. Tong and D. Koller, "Support vector machine active learning with applications to text classification," *J. Mach. Learn. Res.*, vol. 2, no. 11, pp. 45–66, 2001.

[12] S. Amin, I. Uddin, H. H. Al-Baity, A. Zeb, and A. Khan, "Machine learning approach for COVID-19 detection on Twitter," *Comput., Mater. Continua*, vol. 68, no. 2, pp. 2231–2247, 2021.

[13] C.-y. J. Peng, K. L. Lee, and G. Ingersoll, "An introduction to logistic regression analysis and reporting," *J. Educ. Res.*, vol. 96, no. 1, pp. 3–14, 2002.

[14] B. Liu, E. Blasch, Y. Chen, D. Shen, and G. Chen, "Scalable sentiment classification for big data analysis using Naïve Bayes classifier," in *Proc. IEEE Int. Conf. Big Data*, Oct. 2013, pp. 99–104.

[15] G. I. Webb, E. Keogh, and R. Miikkulainen, "Naïve Bayes," in *Encyclopedia of Machine Learning and Data Mining*, vol. 15. Cham, Switzerland: Springer, 2010, pp. 7013–7014.

[16] I. D. Mienye, Y. Sun, and Z. Wang, "Prediction performance of improved decision tree-based algorithms: A review," *Procedia Manuf.*, vol. 35, pp. 698–703, Jan. 2019.

[17] Y. Al Amrani, M. Lazaar, and K. E. El Kadiri, "Random forest and support vector machine based hybrid approach to sentiment analysis," *Procedia Comput. Sci.*, vol. 127, pp. 511–520, Jan. 2018.

[18] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, pp. 5–32, Oct. 2001.

[19] S. E. Jozdani, B. A. Johnson, and D. Chen, "Comparing deep neural networks, ensemble classifiers, and support vector machine algorithms for object-based urban land use/land cover classification," *Remote Sens.*, vol. 11, no. 14, p. 1713, Jul. 2019.

[20] X. Wang, N. Liu, and K. Me, "A novel AHP-based image retrieval interface," in *Proc. Chin. Control Decis. Conf.*, Jul. 2008, pp. 2334–2337.

[21] X. Wang and X. Kanglin, "Content-based image retrieval incorporating the AHP method," *Int. J. Inf. Tech.*, vol. 11, no. 1, pp. 25–37, 2011.

[22] N. K. Rout, M. K. Ahirwal, and M. Atulkar, "Analytic hierarchy process-based automatic feature weight assignment method for content-based satellite image retrieval system," *Soft Comput.*, vol. 27, no. 2, pp. 1105–1115, Jan. 2023.

[23] G. Bhattacharya, K. Ghosh, and A. S. Chowdhury, "Granger causality driven AHP for feature weighted kNN," *Pattern Recognit.*, vol. 66, pp. 425–436, Jun. 2017.

[24] S. Bahassine, A. Madani, M. Al-Sarem, and M. Kissi, "Feature selection using an improved Chi-square for Arabic text classification," *J. King Saud Univ.-Comput. Inf. Sci.*, vol. 32, no. 2, pp. 225–231, 2020.

[25] P. Geurts, D. Ernst, and L. Wehenkel, "Extremely randomized trees," *Mach. Learn.*, vol. 63, pp. 3–42, Mar. 2006.

[26] GitHub. (Oct. 10, 2016). *CREDBANK-Data*. Accessed: Dec. 11, 2022. [Online]. Available: https://github.com/compsocial/CREDBANK-data

[27] GitHub. (Sep. 23, 2021). *FakeNewsNet*. Accessed: Dec. 5, 2022. [Online]. Available: https://github.com/KaiDMML/FakeNewsNet

[28] GitLab. (Sep. 9, 2022). *ArPFN*. Accessed: Dec. 27, 2022. [Online]. Available: https://gitlab.com/bigirqu/ArPFN

[29] FigShare. (Oct. 24, 2016). *PHEME Dataset of Rumours and Non-Rumours*. Accessed: Dec. 25, 2022. [Online]. Available: https://figshare.com/articles/dataset/PHEME_dataset_of_rumours_and_non-rumours/4010619

[30] M. Wijesekara and G. U. Ganegoda, "Source credibility analysis on Twitter users," in *Proc. Int. Res. Conf. Smart Comput. Syst. Eng.*, Sep. 2020, pp. 96–102.

[31] H. S. Al-Khalifa and R. M. Al-Eidan, "An experimental system for measuring the credibility of news content in Twitter," *Int. J. Web Inf. Syst.*, vol. 7, no. 2, pp. 130–151, Jun. 2011.

[32] N. Y. Hassan, W. H. Gomaa, G. A. Khoriba, and M. H. Haggag, "Credibility detection in Twitter using word N-gram analysis and supervised machine learning techniques," *Int. J. Intell. Eng. Syst.*, vol. 13, no. 1, pp. 291–300, Feb. 2020.

[33] R. Zhang, M. Gao, X. He, and A. Zhou, "Learning user credibility for product ranking," *Knowl. Inf. Syst.*, vol. 46, no. 3, pp. 679–705, Mar. 2016.

[34] G. Alfian, M. Syafrudin, I. Fahrurrozi, N. L. Fitriyani, F. T. D. Atmaji, T. Widodo, N. Bahiyah, F. Benes, and J. Rhee, "Predicting breast cancer from risk factors using SVM and extra-trees-based feature selection method," *Computers*, vol. 11, no. 9, p. 136, Sep. 2022.

[35] N. C. Wickramarathna, T. D. Jayasiriwardena, M. Wijesekara, P. B. Munasinghe, and G. U. Ganegoda, "A framework to detect Twitter platform manipulation and computational propaganda," in *Proc. 20th Int. Conf. Adv. ICT for Emerg. Regions (ICTer)*, Nov. 2020, pp. 214–219.

[36] N. S. Murugan and G. U. Devi, "Feature extraction using LR-PCA hybridization on Twitter data and classification accuracy using machine learning algorithms," *Cluster Comput.*, vol. 22, no. S6, pp. 13965–13974, Nov. 2019.

[37] G. Dong and H. Liu, Eds. *Feature Engineering for Machine Learning and Data Analytics*, 1st ed. USA: CRC Press, 2018. [Online]. Available: https://doi.org/10.1201/9781315181080
[38] F. Ahmad and S. A. M. Rizvi, "Features identification for filtering credible content on Twitter using machine learning techniques," in *Proc. Social Netw. Comput. Intell. (SCI)*, Singapore, 2020, pp. 137–149.
[39] T. Khan and A. Michalas, "Seeing and believing: Evaluating the trustworthiness of Twitter users," *IEEE Access*, vol. 9, pp. 110505–110516, 2021.
[40] T. Thaher, M. Saheb, H. Turabieh, and H. Chantar, "Intelligent detection of false information in Arabic tweets utilizing hybrid Harris hawks based feature selection and machine learning models," *Symmetry*, vol. 13, no. 4, p. 556, Mar. 2021.
[41] J. Buda and F. Bolonyai, "An ensemble model using N-grams and statistical features to identify fake news spreaders on Twitter," in *Proc. CLEF*, 2020. [Online]. Available: https://ceur-ws.org/Vol-2696/paper_189.pdf
[42] Z. S. Ali, A. Al-Ali, and T. Elsayed, "Detecting users prone to spread fake news on Arabic Twitter," in *Proc. 5th Workshop Open-Source Arabic Corpora, Process. Tools Shared Tasks Qur'an QA Fine-Grained Hate Speech Detection*, 2022, pp. 12–22.
[43] B. Abu-Salih, K. Y. Chan, O. Al-Kadi, M. Al-Tawil, P. Wongthongtham, T. Issa, H. Saadeh, M. Al-Hassan, B. Bremie, and A. Albahlal, "Time-aware domain-based social influence prediction," *J. Big Data*, vol. 7, no. 1, pp. 1–37, Dec. 2020.
[44] P. K. Jain, R. Pamula, and S. Ansari, "A supervised machine learning approach for the credibility assessment of user-generated content," *Wireless Pers. Commun.*, vol. 118, no. 4, pp. 2469–2485, Jun. 2021.
[45] E. D. Raj and L. D. D. Babu, "An enhanced trust prediction strategy for online social networks using probabilistic reputation features," *Neurocomputing*, vol. 219, pp. 412–421, Jan. 2017.
[46] C. Castillo, M. Mendoza, and B. Poblet, "Information credibility on Twitter," in *Proc. 20th Int. Conf. World Wide Web*, 2011, pp. 675–684.
[47] T. Hamdi, H. Slimi, I. Bounhas, and Y. Slimani, "A hybrid approach for fake news detection in Twitter based on user features and graph embedding," in *Proc. 16th Int. Conf. Distrib. Comput. Internet Technol. (ICDCIT)*, 2020, pp. 266–280.
[48] U. Sharma and S. Kumar, "Feature-based comparative study of machine learning algorithms for credibility analysis of online social media content," in *Proc. Data Eng. Smart Syst. (SSIC)*, Singapore, 2022, pp. 13–25.
[49] U. Saeed, H. Fahim, and F. Shirazi, "Profiling fake news spreaders on Twitter," in *Proc. CLEF*, 2020.
[50] M.-A. Abbasi and H. Liu, "Measuring user credibility in social media," in *Proc. 6th Int. Conf. Social Comput., Behav.-Cultural Modeling Predict.*, Washington, DC, USA, 2013, pp. 441–448.
[51] P. K. Verma, P. Agrawal, V. Madaan, and C. Gupta, "UCred: Fusion of machine learning and deep learning methods for user credibility on social media," *Social Netw. Anal. Mining*, vol. 12, no. 1, p. 54, Dec. 2022.
[52] H. Abbasimehr and M. J. Tarokh, "A novel interval type-2 fuzzy AHP-TOPSIS approach for ranking reviewers in online communities," *Scientia Iranica*, vol. 23, no. 5, pp. 2355–2373, 2016.
[53] I. Dongo, Y. Cardinale, and A. Aguilera, "Credibility analysis for available information sources on the Web: A review and a contribution," in *Proc. 4th Int. Conf. Syst. Rel. Saf. (ICSRS)*, Nov. 2019, pp. 116–125.
[54] V. Sabeeh, M. Zohdy, A. Mollah, and R. Al Bashaireh, "Fake news detection on social media using deep learning and semantic knowledge sources," *Int. J. Comput. Sci. Inf. Secur.*, vol. 15, no. 2, pp. 45–68, 2020.
[55] K. Zhao, L. Xing, and H. Wu, "S3UCA: Soft-margin support vector machine-based social network user credibility assessment method," *Mobile Inf. Syst.*, vol. 2021, pp. 1–10, Nov. 2021.
[56] S. Liu, L. Zhang, and Z. Yan, "Predict pairwise trust based on machine learning in online social networks: A survey," *IEEE Access*, vol. 6, pp. 51297–51318, 2018.
[57] A. M. 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.*, vol. 47, no. 1, pp. 58–81, 2021.
[58] J. ODonovan, B. Kang, G. Meyer, T. Höllerer, and S. Adalii, "Credibility in context: An analysis of feature distributions in Twitter," in *Proc. Int. Conf. Privacy, Secur., Risk Trust Int. Conf. Social Comput.*, Sep. 2012, pp. 293–301.
[59] M. R. Morris, S. Counts, A. Roseway, A. Hoff, and J. Schwarz, "Tweeting is believing: Understanding microblog credibility perceptions," in *Proc. ACM Conf. Comput. Supported Cooperat. Work*, Feb. 2012, pp. 441–450.
[60] M. Hernandez-Mendoza, A. Aguilera, I. Dongo, J. Cornejo-Lupa, and Y. Cardinale, "Credibility analysis on Twitter considering topic detection," *Appl. Sci.*, vol. 12, no. 18, p. 9081, Sep. 2022.
[61] J. Meinert, A. Aker, and N. Krämer, "The impact of Twitter features on credibility ratings—An explorative examination combining psychological measurements and feature based selection methods," in *Proc. Annu. Hawaii Int. Conf. Syst. Sci.*, 2019, pp. 2600–2609.
[62] M. Gayakwad, S. Patil, A. Kadam, S. Joshi, K. Kotecha, R. Joshi, S. Pandya, S. Gonge, S. Rathod, K. Kadam, and M. Shelke, "Credibility analysis of user-designed content using machine learning techniques," *Appl. Syst. Innov.*, vol. 5, no. 2, p. 43, Apr. 2022.
[63] P. Jayashree, K. Laila, K. S. Kumar, and A. Udayavannan, "Social network mining for predicting users' credibility with optimal feature selection," in *Proc. Intell. Sustain. Syst. (ICISS)*, 2022, pp. 361–373.
[64] I. Ahmad, M. Yousaf, S. Yousaf, and M. O. Ahmad, "Fake news detection using machine learning ensemble methods," *Complexity*, vol. 2020, pp. 1–11, Oct. 2020.
[65] X. Duan, E. Naghizade, D. Spina, and X. Zhang, "RMIT at PAN-CLEF 2020: Profiling fake news spreaders on Twitter," in *Proc. CLEF*, 2020. [Online]. Available: https://ceur-ws.org/Vol-2696/paper_233.pdf
[66] M. S. Espinosa, R. Centeno, and Á. Rodrigo, "Analyzing user profiles for detection of fake news spreaders on Twitter," in *Proc. CLEF*, 2020. [Online]. Available: https://ceur-ws.org/Vol-2696/paper_132.pdf
[67] M. S. Karakaşlı, M. A. Aydin, S. Yarkan, and A. Boyaci, "Dynamic feature selection for spam detection in Twitter," in *Proc. Int. Telecommun. Conf. (ITelCon)*, Istanbul, Turkey, 2019, pp. 239–250.
[68] W. Etaiwi and A. Awajan, "The effects of features selection methods on spam review detection performance," in *Proc. Int. Conf. New Trends Comput. Sci. (ICTCS)*, Oct. 2017, pp. 116–120.

NAHID R. ABID-ALTHAQAFI is currently pursuing the Ph.D. degree with King Saud University, Riyadh, Saudi Arabia. Her research interests include feature engineering, machine learning, data mining, and decision support systems.

HESSAH A. ALSALAMAH is currently an Associate Professor with the Information Systems Department, College of Computer and Information Sciences, King Saud University (KSU). She is also the Dean of the College of Engineering and Architecture, Al Yamamah University. She specializes in business process management and workflow technologies, including process discovery, modeling, analysis, re-engineering, and automation. She also focuses on the emerging requirements of collaborative environments involving human aspects, such as communication, collaboration, and coordination, and technical aspects, such as heterogeneity and distribution. Additionally, she has been exploring the domains of information security and privacy requirements in collaborative environments, such as eHealth and eGovernment.

WALAA N. ISMAIL received the M.S. degree in information systems and computer science from the Department of Information Systems, Cairo University, Egypt, and the Ph.D. degree from the Information Systems Department, College of Computer and Information Sciences, King Saud University, Riyadh, Saudi Arabia. Currently, she serves as an Assistant Professor with the Department of Information Systems, Minia University, Egypt, and also holds a position at the College of Business, al Yamamah University. Her research interests include pervasive healthcare, anomaly detection, cloud computing, mobile healthcare, and data mining.