Journal of Data Acquisition and Processing
ISSN: 1004-9037
https://sjcjycl.cn/
DOI: 10.5281/zenodo.7750865

A SUPERVISED APPROACH FOR SCAMMER DETECTION FOR ONLINE SOCIAL NETWORKS FRAUD BASED ON USER INFORMATION INTERESTS

Smita Bharne[1]

[1]School of Computer Sciences and Engineering, Sandip University Nashik, /Ramrao Adik Institute of Technology, D. Y. Patil Deemed to be University, Navi Mumbai, India, ORCID ID:0000-0003-1869-5552, smita146@gmail.com

Pawan Bhaladhare[2]

[2]School of Computer Sciences and Engineering, Sandip University Nashik, India
ORCID ID:0000-0003-1640-0984

ABSTRACT:

The online social networking phenomenon has grown tremendously over the last twenty years. As social networking platforms have evolved, numerous online activities have emerged that have captured the attention of a large number of users.People increasingly rely on the credibility of the information presented on Online Social Networks (OSN). Conversely, online social networks have experienced a rise in the number of compromised, false accounts, scam profiles that do not correspond to real individuals.OSN operators are now using a variety of resources to detect such kind of scam profiles and accounts.Scammers in OSN are taking advantage of this for performing various OSN frauds. It is difficult to detect scammers due to the wide range of OSN platforms and the variety of OSN frauds. In this paper, an effort has been made to detect a scammer by designing a scammer detection model which will blacklist scammer profiles through user profile-based features. The proposed approach also differentiates between the scammer and real profiles. The experimental result and analysis show that the proposed model demonstrates better performance compared to other competing models, achieving an accuracy and f1 score of 98.75% and 97.95%, respectively for the dataset created for the study. This work aims to increase early-stage detection of scammers in dating frauds, compromised accounts, and fake profiles to provide safety to women and society.

Keywords: Scammer profiles, Online social network,OSN frauds, Scammer detection model, social threats, Compromised accounts, Fake profiles, Dating Fraud, Machine Learning.

1.INTRODUCTION

Millions of people use social networking sites around the world. Users' interactions with social media sites like Twitter, Facebook, Instagram, Tinder etc.have a huge impact on their everyday lives, with sometimes negative consequences [1-2]. The prevalence of modern electronic devices such as smartphones and laptops has led to a significant portion of the

global population accessing the OSN, with estimates indicating that more than half of the world's population uses it. [3]. Among the 4.66 billion internet users worldwide, more than 4.14 billion individuals utilize social media apps like Facebook, Twitter, Instagram, and others [4]. Millions of people use such social media apps every day [5-6]. Popular social networking sites have become a target platform for scammers to disseminate a large volume of irrelevant and harmful content. The spread of false information or fake news through online social media networks is increasing due to internet fraud (OSN). However, social network information and identity theft are the most common types of illegal activities that occur on the internet. [7].More particular, fake or cloned accounts, where scammers mimic victims through identity theft, are a key source of incorrect information on OSN [8]. Social scammers are individuals or groups who use social media platforms to deceive and defraud others.OSN frauds also includes human (specifically women) targeted frauds, which refers to fraudulent activities that are specifically targeted towards individuals to deceive and defrauding them. The frauds like online dating, compromised accounts, and false identityoften create fraudulent accounts to impersonate legitimate usersto gain trust and then steal money. Many times, the emotional harm to the users is more than the financial loss in such types of fraud.Scammers use a variety of tactics such as phishing, and fake investment schemes to trick their victims into providing sensitive information or sending money. Social scammers can cause serious harm to their victims and can undermine trust in social media platforms.OSNaccount profiles are often cloned and used for deceptive activities such as false advertising, blackmailing, money laundering, terrorist propaganda, spamming, scamming, and other types of malicious behavior. These activities aim to steal information, tarnish the victim's reputation and credibility, or gain the trust of the victim's friends and followers, leading to further fraudulent activities. [9].

More than 95,000 customers reported losses exceeding 770 million dollars as a result of OSN theft in 2021.By 2021, the losses will account for roughly a quarter of all reported fraud losses, an increase of eighteenfold from 2017. Those between the ages of 18 and 39 are going to be nearly twice as likely as older persons to report losing money to these scams in 2021 [10]. Scammers use social networking sites to present bogus possibilities to connect with them and even establish direct contact with former friends to persuade them to invest and utilize their personal information to trap them.

To prevent online social networking fraud, we proposed a methodology for detecting scammers and blacklisting scammer profiles using user profile-based attributes. The study has made the following noteworthy contributions: a) Our study employs advanced techniques in text classification and image caption analysis to excerpt valuable insights from a large collection of user profiles to detect scammer profiles.b) We have used a multi-classifier approachthat analyzes separate aspects of public profile characteristics. c) Wedesign a scammer detection model that can effectively identify and blacklist scammer profiles in online social networks (OSNs) to help the users before they become victims at early stages.

The remainingpaper is structured asfollows: Section 2 presents a depth literature survey on state of art systems. The proposed methodology is explained in section 3. Results and discussion are presented in section 4. Section 5 concludes the paper with future scope.

2.RELATED WORK

In literature, various authors have contributed their work for the detection of scammer profiles, compromised accounts, and false and duplicated profiles on social networking platforms such as Facebook, Twitter, LinkedIn, Tinder, etc. The author in [11] proposeda methodology using simple statistical analysis to discover cloned profiles by comparing the similarities of the profiles and validating the behaviors and IP addressesTypically, when comparing user profiles on social networks, two main similarity metrics are used: attribute similarity and network similarity. Attribute similarity measures the similarity of demographic information between profiles, while network similarity measures the similarity of friends' lists. [12-13].The author in [15] focused on similarity index parameter computation that uses weights for the features based on their usefulness in classifying the profiles. On LinkedIn, a duplicated profile detection algorithm based on profile feature similarities was proposed. For determining the similarity value, this technique employs a straightforward string-matching strategy. Nevertheless, these models lack accuracy and cannot be utilized alone to detect false profiles.

Authors in [16] proposed a method for detecting scammers on online dating sites using a machine learning (ML) classifier. ML-based classifiers such as support vector machines (SVM) and decision trees are used for recognizing images used in romance scams. Later on, a comparison of the performance of ML-based classifiers is given for the most accurate results. Researchers in [17] proposed a method for detecting a scammer from an online dating site using image-based detection with profile descriptors.To recognize images, the author constructed a generative model using a deep neural network (NN). Based on the SVM prediction model feature, the images are captured. Authors in [18] use a method to distinguish the profile photos on dating sites between celebrity and non-celerity categories. Many scammers use celebratory photos as profile pictures to hide their identity. The focus of this study was on scraping data from websites and detecting faces using ML technology.Numerous authors offered their detection technique by utilizing various qualities and assert that including those attributes significantly increases the performance of distinguishing fake profiles from authentic ones. The characteristics of online profiles are typically classified into five categories, namely network-based, content-based, temporal-based, profile-based, and action-based [19]. Author [20] proposed a detection model that analyzes multimedia data and found that content-based and profile-based characteristics resulted in higher accuracy in identifying scam profiles [20]. Furthermore, in the classification of scam profiles, various machine learning classification and clustering algorithms are tested [21-22].

In [23], a new technique for evaluating trustworthy and distrusted relations in OSNs is devised. Several algorithms attempted to overcome this problem by segmenting social graphs depending on user identities [24-25]. Another approach is Sybil Infer and Sybil Rank, which returns the probability based on ranking each node in the social graph according to their estimated odds of being fake nodes [26-27]. The authors of [28] proposed a detection strategy for Sybil accounts in Renren-OSN by observing Sybil’sbehavior in the wild. Other approaches to detecting fake accounts based on profile features and behaviors are introduced. An author in [29] presented an automated Feature-based Fake Profile

Detection Algorithm based on Machine Learning Considerations is initiated. Author in [30]proposed a new method for identifying profile cloning based on profile attributes similarity &Facebook network similarity is presented. The author of the paper [31] introduces a five-step automated technique for detecting malicious users & social spam campaigns. The authors of [32] described how to apply the Exclusive Shared Knowledge technique among friends to identify their close friends in an OSN. In [33], a novel approach for Sybil detection is described that is based on the core behavioral patterns of Click-Stream models. Crowdsourcing [34] is yet another stand-alone solution that relies on online human experts to discover Sybil accounts in OSNs.In astudy by the author [35], a machine-learning pipeline and a random forest classifier were used to propose a simple approach for detecting fake profiles. A method proposed by the author in [36]usedan agglomerative hierarchical clustering methodwith Jaccard similarity metrics and weighted characteristics to identify cloned profiles. Despite the authors' claims that the procedures are effective, the paper lacks in-depth experimental investigation and comparison.

For the detection of scammer and false profiles, there is no specific dataset available. As there are multiple OSN platforms available, investigating the scammer profile detection methods needs extra effort to extract a large number of attributes. Nevertheless, not all of these techniques are useful for classification.The associationis the most significant subject of study in statistics for assessing dependence between two sets of data that are commonly employed in feature selection [37]. Hence, before running the classification method, a few feature selection approaches such as correlation and principal component analysis were used [38]. To identify important features, the study employed based on metrics-based feature weighting and evaluated the effectiveness of this approach using various classifiers including Random Forest, Decision tree, Naive Bayes, neural network, and Support vector machines. Only 7 of the 22 gathered attributes were shown to be efficient in detecting phony profiles [39].

To detect the cloned profile, a method that measures the relationship strength among two profiles having active friend lists as well as the number of likes was presented [40]. A smart system named FBChecker has been proposed that detects phony Facebook profiles by combining behavioral and informational aspects with supervised learning algorithms. The procedure was carried out by using the KNN schema to fill in the missing data and filtering the records with missing values [41]. The above studies, however, lack rigorous experimental investigation to back up the findings. The authors further extended their research with unsupervised clustering algorithms, and the findings show that ID3 improves detection accuracy [42]. Similarly, with 30 profile attributes, the analysis was performed using powerful algorithms for machine learning such as boosting and bagging, with AdaBoost demonstrating enhanced detection accuracy of false detection [43].

To detect false identities, scam profiles artificial intelligence and natural language processing (NLP) was recommended. The model was proposed to use principal component analysis for feature selection. The machine learning model was tested using classifiers such as “Random Forest”, (RF), “Support vector machines” (SVM), and the optimized Nave Bayes algorithm. The author concludes that the SVM algorithm provides betteraccuracy than others [44]. Instead of classifiers, a novel notion of employing the PageRank method was developed, in which the model gathered features and used a clustering procedure to group comparable

qualities [45]. Finally, the PageRank method was utilized to detect copied profiles. Even though the model was built with the MapReduce framework, the evaluation was done using the celebrity's profile. With the data mining algorithm, a privacy-protected system for detecting susceptible and false users and cloned profiles were presented by the author [46]. According to the author, the model reduces the erroneous rate to 1%. Only a handful of works support large data among all of these strategies. Many of these methods demand more time to categorize data by employing algorithms of machine learning to compare profiles using similarity metrics.

Afterstudying the existing work from the literature review, it is analyzed that each method has certain limitations based on the paraments used for the study like less accuracy or high computational intricacy. This research is motivated to propose a scammer detection model that detects scammers to simplify computation.

3. PROPOSED METHODOLOGY

Figure 1 depicts the generalized workflow of the proposed machine learning-based scammer detection model. The scammer creates a phony identity and personally communicates with the victim in human-targeted frauds like online dating fraud, false equivalence, and compromised accounts. Due to the nature of human-targeted fraud and the way scammers exploit the trust of their victims, traditional detection methods are often ineffective.Users on OSN sites and dating sites offer as much information as possible to find their ideal match. Scammer detection performance can be increased by employing modern machine learning algorithms. The following steps are included in the machine learning-based scammer identification.

The dataset is constructed based on user profile attributes for the real and scammer user profile categories. After data collection, data must be pre-processed. Before transmitting the acquired raw data to an ML model, the data must be cleaned and organized using pre-processing procedures. Data normalization, data cleaning, noise removal, and other processes are some of the stages involved in the pre-processing of text data. After data preprocessing, the next step is feature extraction, which is used to form a matrix from the best set of features extracted and selected using feature extraction and feature selection. Our proposed model selects a subset of features from entire features. The machine learning models are trained on a constructed dataset of the user profile attribute’s extracted features and tested on a test data set. It has performance evaluations parameters like accuracy (A), precision (P), recall (R), and F1-measure/score (F) based on the confusion matrix

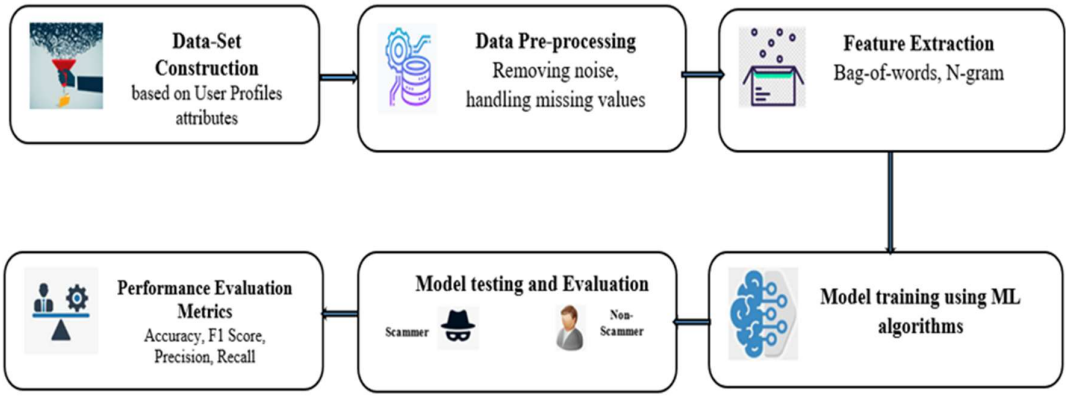

Figure 1. Workflow of the scammer detection model of the proposed system

3.1 Construction of the Dataset

Detection of scammers using machine learning algorithms requires a high volume of datasets. Researchers are in search of a good dataset for research purposes. But due to data privacy, legal and ethical issues, scammer evasion tacticsandlimited publicly available dataset, it is difficult to find the finding scammer's profiles dataset for online social network platforms. Since there is no standard dataset available for scammers in OSNfraudster'sprofiles, we construct the dataset from the websites scamdigger.com [47] forthe creation of scammers profiles and datingnmore.com [48] for construction of the real user profiles attributes. The real user profiles taken from websites are able to differentiate themselves that only authentic profiles are registered with themselves. Online social networks typically use the range of user profile features that allow users to share information about themselves and connect with others. The user profile attributes are profile image, short descriptions, interests, occupation, locations etc. We have collected the dataset of the scammers and real user profiles till march 2022. The dataset consists of fraudster profiles as well as a big sample of authentic profiles.For ethical reasons, only the URLs that link to the images were retained for both datasets. The URLs of the images were extracted in anautomated way using python library. We also preserved the user's data privacy; no personal identification information is revealed including those reported in the category of scammers profiles.

3.2User Profile Characteristics

Although different types of OSN are available in the market the typical form of user profile consist of the user profile image, some basic demographic information along userself-description. The major difference observed between the probable scammer profile and real user is the "showcase of information" in the short-description attribute is more compared to the real user. Our approach for scammer detection is by evaluating these common profile attributes present on the OSN user profiles. Besides that, we are comparing the characteristics of the real profiles with the scammer's profiles in detail with profile demographics, profile image recognition, and profile descriptions. The attributes of

scammer profiles are usually different than real userprofiles, as most of the information like email id, multiple phone no, and locations are provided additively. These attributes are typically publicly not seen on real user profiles. Focusing on the features that are available for both types of profiles, we categorize profile attributes into three classes and apply the three different classifiers with the information it need.Figure 2 shows the in-detail architecture of the proposed scammer detection model. The detailing of all the steps is described in the below sections.

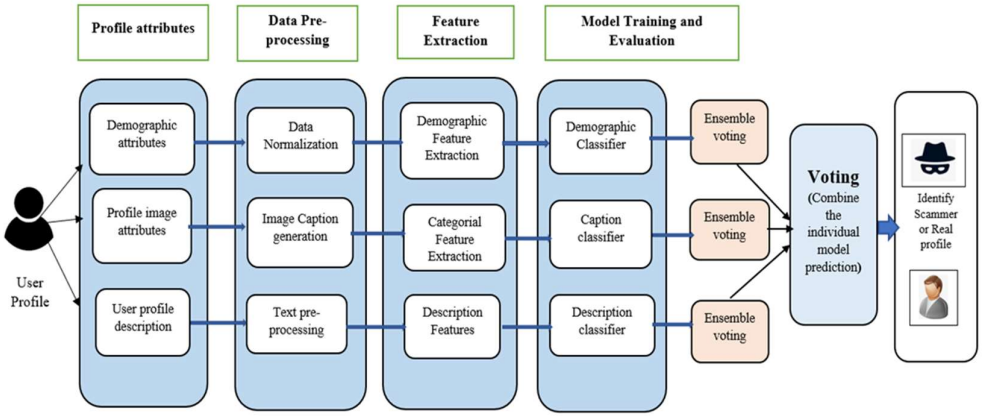

Figure 2. Detailed architecture

3.2.1 Describing profile attribute features

We have divided the user profile attributes into three different setsof attributes i.e. demographics, profile image and short profile description.

1. **Demographic attributes:**includes name, age, gender, ethnicity, location, occupation, country, status etc.The profile image is one of the important attributes, scammers use profile image to gain attention towards them. From the dataset, it is observed that there are multiple profile images per user in the constructed dataset.
2. **Profiles images:**are illustrating their hobbies, interest,and workplaces to make their profile more attractive compared to the real user profile. In both types of profiles, a prevalent trend is the inclusion of photos that not only showcase the user's physical features but also express their hobbies or interests.
3. **A short textual self-description:** from the user that highlights their most salient characteristics and interests.

In both genuine and fraudulent profiles, there was an equal distribution of genders, with approximately 65% of profiles belonging to males. This finding emphasizes that OSN frauds are not restricted to a particular gender, which aligns with the conclusions drawn by previous research[49]. In terms of average age, both genuine and fraudulent profiles had a similar average age of approximately 40 to 44 years. Nevertheless, the distribution of ages varied

notably between the two groups. While the average age of male and female genuine profiles was identical, the average age of female scammers was around 30 to 35, and male scammers averaged around 50 to 55 years old. Although both genuine and fraudulent users were predominantly single, scammers tended to portray themselves as widowed, rather than any other marital status. This outcome is unsurprising, given that female victims of OSN frauds like dating fraud or fake profiles frequently report that scammers leverage this attribute to garner their sympathy and trust.

Distinct approaches are necessary to derive significant insights from the attributes present in these profiles. In the following sections, we discuss the pre-processing steps for each category and the significant characteristics that distinguish scammer profiles from genuine user profiles.

3.3 Data Preprocessing

The primary objective of this model is to deliver precise and reliable predictions, and its algorithms should possess the ability to swiftly comprehend the features of the data. Most datasets obtained from the real world contain missing, inconsistent, or noisy data due to their diverse sources. Our dataset for the scammer and real user categories is somewhat different as some of the scammer data profile images havemultiple profile pictures and some of the real and scammer users are not using any profile images.Such kinds of profiles are discarded. Hence after preprocessing the no. of scammer and real user profiles is reduced.The data preprocessing consists of several stages, which include the following steps:

1. **Data cleaning***:* after downloading the real and scammer user profiles, separation of the datawith scam and real profiles in JSON (JavaScript Object Notation) files with all attribute data. JSON file data is represented using key-value pairs, where the key is a string and the valueincludes name, age, email, address, location, etc.all the demographic details are stored in this file.
2. **Encoding categorical variables:** Some attributes, such as gender, ethnicity, and status, are categorical variables. They need to be encoded numerically for machine learning algorithms to work properly.
3. **Data transformation:**refers to the process of converting data from one form to another. Data normalization is done to fill in the missing values in the user profile attributes.
4. **Data reduction**: a process of reducing the size of the dataset byeliminating the irrelevant scam and real user profiles with respect to the profile images. Finally, the profile.csv dataset consisting of the scam and user profiles is ready for the machine learning model.

3.3.1 Image Caption Generation

Image captions have been generated from the user profile images to understand the semantics.Both types of profiles, real and fake, tend to use images that not only showcase the user's physical appearance but also provide a glimpse into their interests and hobbies. The pictures are carefully selected to convey a certain message about the user's personality and lifestyle.Many times, the scammers are using public figures'images, in the other context. Hence image caption generation from the profile's dataset is help to understand the choices of

the users while selecting the profile image. The model generates a description that captures the underlying meaning of each image in a profile, based on its analysis of the image content.Once we have obtained the most appropriate caption for each image, we eliminate the least informative components, keeping only the entitiesthatcarry the most significant meaning. For a generation of image captions, we have used the Blip-Image-captioning model [50]. It is a deep learning model that generates captions for images. The model uses a convolutional neural network (CNN) to extract visual features from images and a recurrent neural network (RNN) to generate captions based on those features. In the context of scammer detection, the Blip-Image-captioning model is used to analyze profile pictures and generate captions that describe the content of the images. These captions are the additional features for scammer detection, along with demographic information, text descriptions, and other relevant data.

3.3.2 Text preprocessing

To gain attraction onsocial networking platforms, many userseasily share intimate details about themselves, such as their life experiences, interests, and preferences, with complete strangers. The personal description section of a user's profile is strongly encouraged by these websites because it helps attract the attention of potential matches and increases the likelihood of finding a compatible partner or friends. We obtained description features from the text content of the descriptions that were tokenized. Our feature extraction process included techniquessuchas character n-grams, and word n-grams. The frequency of occurrence of each n-gram in the text is used as a feature for the machine learning model.

3.4 Feature Extraction

We extracted three sets of features based on the three profile attributes categories which will the input to the classification system. Table 1 shows the different classes of attributes from the profile.csv dataset.

1. *Demographic attributes* (*Xde)*-refer to the numerical and categorical features. A numerical attribute refers to a type of attribute or feature in a dataset that contains numerical values.

2. Categorical attribute (*Xca)*-is a type of attribute or feature in a dataset that takes on values from a limited set of categories or classes that are extracted.

Table 1: Different classes of attributes with a feature set

Class of attributes	Name of the attributes	Type of the attributes
Demographic features	Age, gender, ethnicity, occupation, latitude, longitude, country, status	Numerical and categorical attribute
Categorical features	Image captions objects	Categorical attribute
Description features	Word n-grams, character n-grams, and text tokens	Nominal attributes

3. Description-basedfeatures (Xds)-refer to the features that are extracted from the textual descriptions or captions associated with an image or profile. From descriptions, word features are extracted based on the textual contents. The purpose of extracting these features is to gain insight into the context and meaning of the textual content, with the goal of enhancing the performance of machine learning models.

3.5 Machine Learning Model Prediction

To identify fraudsters, the system undergoes initial training using a dataset consisting of both real and fraudulent profiles. The primary objective of this phase is to extract crucial components that will be utilized in the subsequent stages for scammer detection. Following steps of the proposed algorithm are used for the identifying of scammer profile.

Algorithm:

Step 1. Dataset loading

Step 2. preprocess the dataset profile.csv

Step 3: ***Xi*** *(Xde, Xca, Xds)* is a set of user profiles and $Y_i$is corresponding set of labels indicating the profile is scammer or real profile.

Step 4: for each profile X as a feature vector calculate the probabilities {P} of scam using classifier

$$P(X_i) = (P\{Xde, Xca, Xds\})$$

Step 5. for each $\mathrm{P(X_i)}$ in $\mathrm{P_{de}, P_{ca}, P_{ds}}\{\mathrm{X} = 0|1\}$ indicate corresponding classifier probabilities
of demographics, caption and description classifiers.

Step 6: The dataset was divided into three separate subsets for training, testing, and validation purposes.

Step 7. A voting model $fP(X) = \{\ 0|1\ \}$ used to combine the individual classifier probabilities such that $f(W_i) = w_{de}, w_{ca,} w_{ds}$ be the weights assigned to demographics, caption and description classifiers respectively.

$$f\{P(X)\ =\ \sum w_i * P_i(Pde, Pca, Pds)\ \ \{X =\ 0\ |\ 1\ \}$$

where P_i is the prediction of the classifiers i and

w_i its corresponding vote based on the accuracy of validation set. The summation runs over all classifiers in the set $P(X_i)$ for all individual ensemble classifier's probabilities $\mathrm{P_{de}, P_{ca}, P_{ds}}$.

Step 8: A function f: X → {0, 1} that maps each feature vector X_i to a binary label Y_i indicating whether the profile is scam (1) or not (0), where f is a performance metric that measures the accuracy of the ensemble model.

3.5.1 Multi-classifiers used:

OSN user profiles typically have incomplete information as many users choose not to disclose personal details or prioritize contacting others over sharing information about themselves. Therefore, an effective detection system for fraudulent profiles should be able to handle incomplete profiles with flexibility.This section describes three separate classifiers that are used to determine the likelihood of fraudulent profiles.The design of each classifier is focused on effectively modeling a specific section of the profile attributes describe in section 3. 4. We combine the probability of the outputs from each classifier to provide a stable decision. By designing multiple classifiersmodels on distinct sections of the profiles, we can provide a more reliable and accurate solution to classification problems compared to a single classifier.Ensemble classifiers are generally preferred over single classifiers due to their ability to combine the outputs of multiple models and achieve better overall performance.

1. Demographic classifier: In this study, we employed three popular machine learning algorithms Random Forest, Naive Bayes, and XGBoost (Extreme Gradient Boosting) along with ensemble classificationto develop a demographic classifier.In the Random Forest classifier, we can build multiple decision trees, each using a random subset of the demographic attributes. Each decision tree makes a prediction, and the final prediction is based on the majority vote of all decision trees. In XGBoost, we can build an ensemble of weak learners, each focusing on a different subset of the demographic attributes.It constructs an ensemble of weak decision trees, where each new tree is trained to correct the errors made by the previous trees. The weak learners are combined to form a strong learner that makes the final prediction. The demographic classifier utilizes a diverse range of original profile attributes, including location, ethnicity, age, and gender. It is unique in its ability to handle non-binary missing data situations where some information may be missing for a profile, while other information is still available. In most scenarios, more information is available within the demographics data, which the classifier uses to make informed decisions.However, the results given by the Naive Bayes classifier is not the most effective for profiles with complete data. In this study, we found a Random Forests model was found to be more effective. To provide the final prediction with function f: X → {0, 1), a joint model using both Random Forests, Naive Bayes, and Random Forest, XGBoost was trained. The Random Forests model was used to make predictions when all demographic data was available. The final approach is to provide the probability using a joint model that combines the output probabilities of both the Random Forests, Naive Bayes models, and Random Forest, XGBoost model. In our approach the ensemble of Random Forest, XGBoost is outperform well compared to other models.

2. Caption classifier: This model is used to predict the likelihood of a profile being a scam based on the features extracted from the image captions. To build the caption classifier module we employed the SVM (support vector machine), Random Forest, linear SVM, andXTree

(Extra randomized trees). Previous studies show that the [51]SVMis an effective method for fraud detection. SVM can be used to build a model that predicts whether a generated image caption is associated with a real or fraudulent profile. The SVM algorithm works by finding a hyperplane that separates the training data into different classes. The hyperplane is chosen such that it maximizes the margin between the two classes, meaning it tries to find the largest possible distance between the closest data points of each class.SVM has been shown to be effective for a variety of classification tasks, including image caption classification.Extra randomized trees (XTree) is an ensemble learning algorithm based on decision trees. It is an extension of the Random Forest algorithm, where each tree is built on a random subset of features and a random subset of the training data, but with an additional step of randomly selecting the split points for each node [52]. Compared to other classifiers, linear SVM [53] has the advantage of being computationally efficient and easy to implement. It can also handle high-dimensional data well, which is important when dealing with large text datasets. The final approachto providing the probability using a joint model which combines the output probabilities of the SVM, Random Forest, linear SVM, and XTree (Extra randomized trees).

3. Description classifier

Description features are extracted as short text in the user profiles in the form of short textual content. The description classifier is built using Lib Short Text [54] which is a text classification library that is specifically designed for short and sparse text. It uses a bag-of-words approach and a linear SVM model to classify the text into scam or real profiles. The textual descriptions from the user profiles are preprocessed by tokenization and stop-word removal and then transformed into numerical features using the bag-of-words model. The trained linear SVM model is then used to classify the transformed text features into scams or real profiles.

4. Ensemble voting classifier

Ensemble learning is a method in which multiple models are combined to produce a more accurate and robust prediction. In the context of scammer detection in OSN frauds like online dating fraud, fake profiles, and compromised accountsan ensemble of classifiers can be used to combine the predictions from multiple classifiers to achieve a better overall prediction.In this case, the ensemble approach can be applied by combining the predictions of the demographic classifier, the caption classifier, and the description classifier. Each of these classifiers can independently predict the probability of a profile being fraudulent, based on the features that it considers.

The outputs of these classifiers can be combined using a weighted average or a voting scheme, where the prediction with the highest confidence is selected as the final prediction. By combining the strengths of multiple classifiers, the ensemble approach can lead to a more accurate and reliable prediction of fraudulent profiles.

3.6 Evaluation methodology

The evaluation methodology forthe scammer detection system involved the following steps:

1. **Data splitting:** The dataset was split into three parts: training set (60% of the data), validation set (20% of the data), and testing set (20% of the data).
2. **Individual classifier training:** Each component classifier (i.e., demographic classifier, caption classifier, and description classifier) was trained within the 60% training set. The individual performance levels were established through k-fold cross-validation within this set.
3. **Ensemble classifier training**: The probability outputs from each classifier were combined to provide one balanced judgment using the training set. The ensemble classifier was trained using these probability outputs.
4. **Validation:** The validation set was used to tune the hyperparameters of the ensemble classifier to improve its performance.
5. **Testing:** The testing set was used to evaluate the performance of the ensemble classifier. The following evaluation metrics were used: precision, recall, F1-score, and ROC.

6. **Cross-validation**: The entire process was repeated several times using different random seeds to perform cross-validation and obtain the average performance metrics.

Overall, this evaluation methodology ensured that the ensemble classifier was robust and generalizable, and could accurately detect fraudulent profiles even in incomplete profile data.

4. RESULTS AND DISCUSSION

Our findings are evaluated along with performance parameters addressing four classification performance parameters (i) accurately classified scammer profiles (TP), (ii) accurately classified real profiles (TN), (iii) misclassified real profiles (FP), and (iv) misclassified scammer profiles (FN).

4.1 Performance Parameters

The performance metrics used to evaluate the scammer profile classification system include precision, recall, accuracy, F1 score, and ROC (Receiver Operating Characteristic) curve.

1. Precision is the fraction of true scam profiles among the total number of profiles classified as scams.

$$Precision = TP / (FP + TP) \quad (1)$$

2. Recall is the fraction of true scam profiles classified as scams among the total number of actual scam profiles.

$$Recall\ Score = TP / (FN + TP) \quad (2)$$

3. Accuracy is the fraction of correctly classified profiles among the total number of profiles.

$$Accuracy\ Score\ =\ (TP\ +\ TN)/\ (TP\ +\ FN\ +\ TN\ +\ FP) \quad (3)$$

4. F1 score is a harmonic mean of precision and recall.

$$F1\ Score\ =\ 2 * Precision\ Score\ *\ Recall\ Score/\ (Precision\ Score\ +\ Recall\ Score) \quad (4)$$

The ROC curve depicts the relationship between the true positive rate (TPR) and false positive rate (FPR) across various classification thresholds.The TPR is the proportion of real profiles that are correctly classified as real and the FPR is the proportion of scammer profiles that are incorrectly classified as real. A common metric used to assess the overall performance of a classifier is the area under the ROC curve (AUC). A higher AUC value indicates better performance in distinguishing between scammer and real profiles.By analyzing these performance metrics, the effectiveness of the scammer profile classification system can be evaluated and improved.

4.2 Classifiers Results and analysis

We have calculated the results for classifier (demographics, caption, description and ensemble voting classifier) as shown in figure 3. Figure 3 shows the all the three classifier results with overall ensemble voting for individual classifiers. The results achieved after overall ensemble voting with f1 score is 97.95% with accuracy 98.75%.

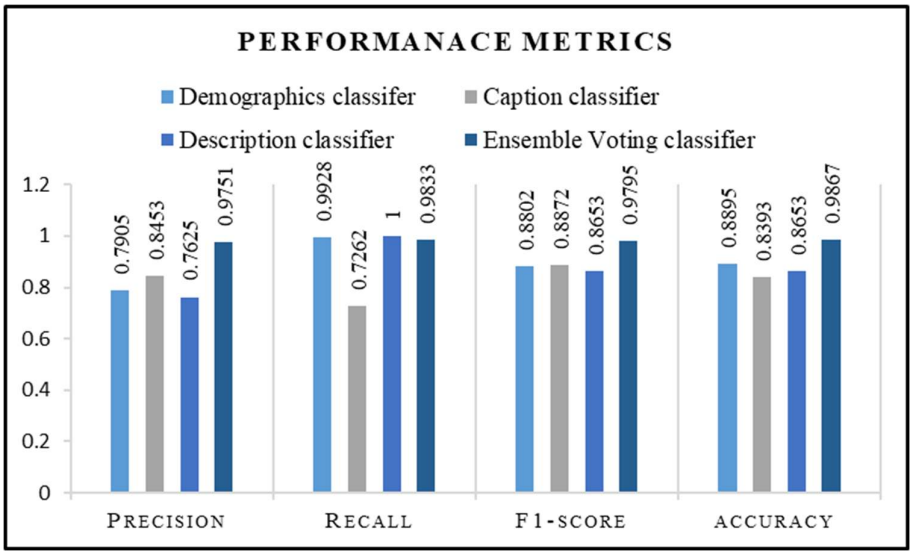

Figure 3. Final results for each component classifier with Ensemble classifier

We have implemented the state of art algorithm to find out their performance in caption classifiers. The accuracy achieved with caption classifier by SVM is 83.25%, RF is 83.73%, Linear SVM is 82.94% and XTREE is 83.41% with ensemble voting. Analysis of this is depicted in the figure 4.

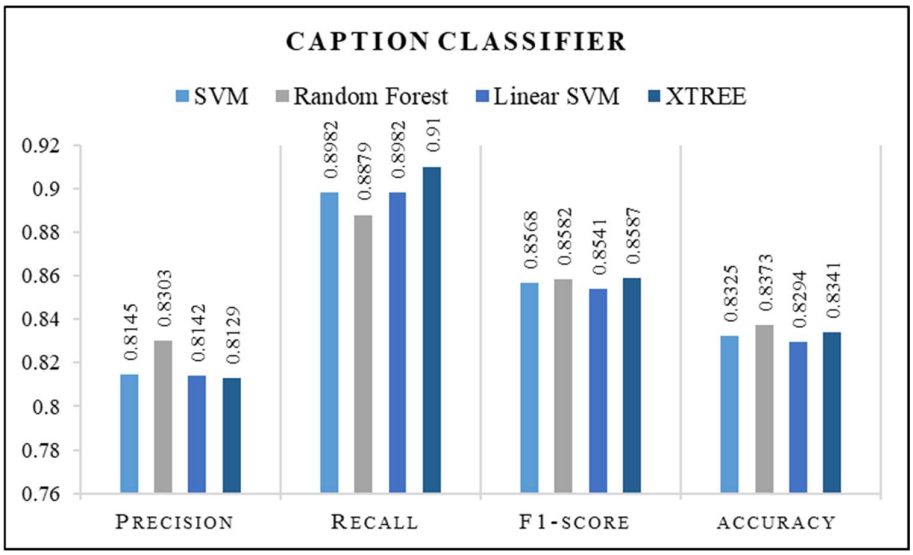

Figure 4.Comparative Analysis of SVM, RF Linear SVM and XTREE

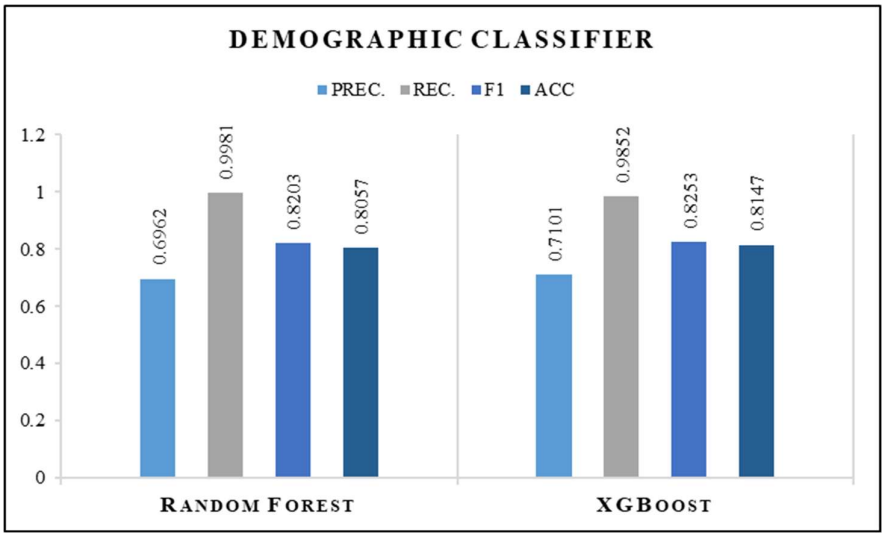

Figure 5. Demographic classifier with Random Forest with XGBoost

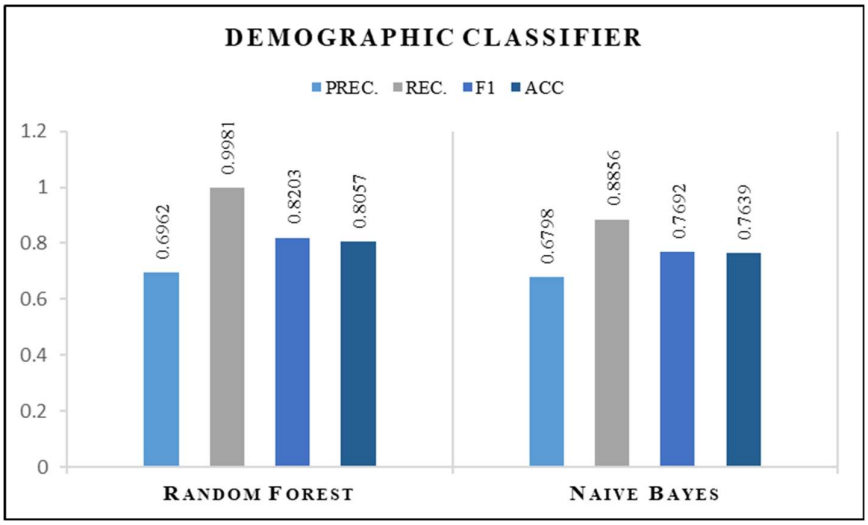

Figure 6.Demographic classifier with Random Forest with Naive Bayes

The accuracy achieved with the demographics classifier by RF is 80.57% and is compared with NaïveBayes and XGBoost which has achieved the accuracy of 76.39% and 81.47 respectively. Analysis of this is depicted in the below figure 5 and figure 6. The ROC curve demonstrates the balance between the true positive rate (sensitivity) and the false positive rate (1 – specificity) in a classification model. ROC curve of the Random Forest and Naive bias is shown in Figure 7(a) and for Random Forest and XGBoost is shown in figure 7(b). From both the figure it is clearly differentiated the curves of XGBoost algorithm is closer to top left corner which depicts high performance that the performance of XGBoost is good as compare to another algorithm.

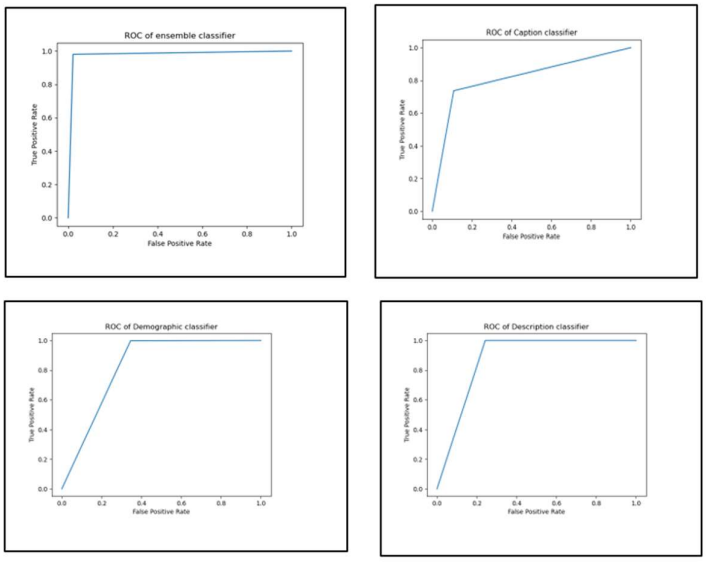

Figure 7(a). ROC curve with Random Forest and Naive Bayes

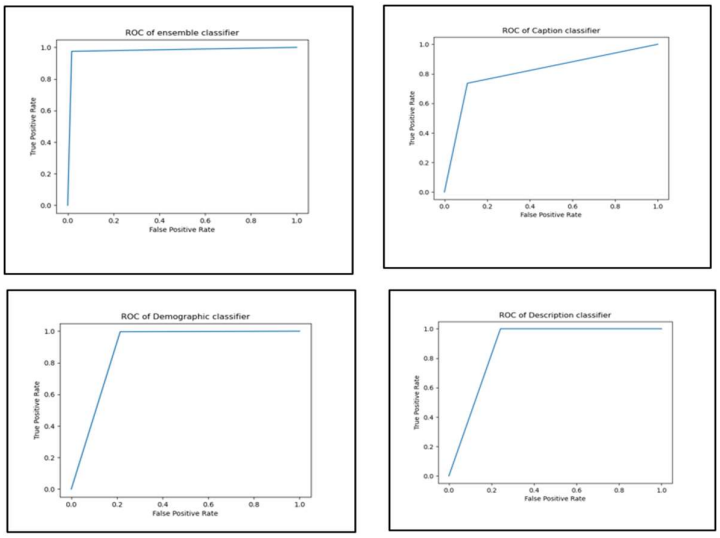

Figure 7(b). ROC curve with Random Forest and XGBoost

Table 2. Performance analysis of the proposed system

Classifier	**Proposed System**				**Existing System [17]**			
	Precision	Recall	F1-score	Accuracy	Precision	Recall	F1-score	Accuracy
Demographic classifier	0.7905	0.9928	0.8802	0.8895	0.858	0.822	0.848	0.913
Caption classifier	0.8453	0.7262	0.8872	0.8393	0.997	0.546	0.705	0.874
Description classifier	0.7625	1.000	0.8653	0.8653	0.884	0.804	0.842	0.917
EnsembleVoting classifier	0.9751	0.9833	0.9795	0.9875	0.962	0.929	0.945	0.971

A SUPERVISED APPROACH FOR SCAMMER DETECTION FOR ONLINE SOCIAL NETWORKS FRAUD BASED ON USER INFORMATION INTERSETS

Table 2 shows the performance analysis of the proposed system with existing system [17]. Proposedsystem shows the better results in ensemble voting classifier. Although the accuracy of the individual classifiers is less but the f1 score is high in proposed system as compared with the existing system.F1 score can be a better metric than accuracy because it takes into account both precision and recall, which are equally important in evaluating the performance of a classifier.

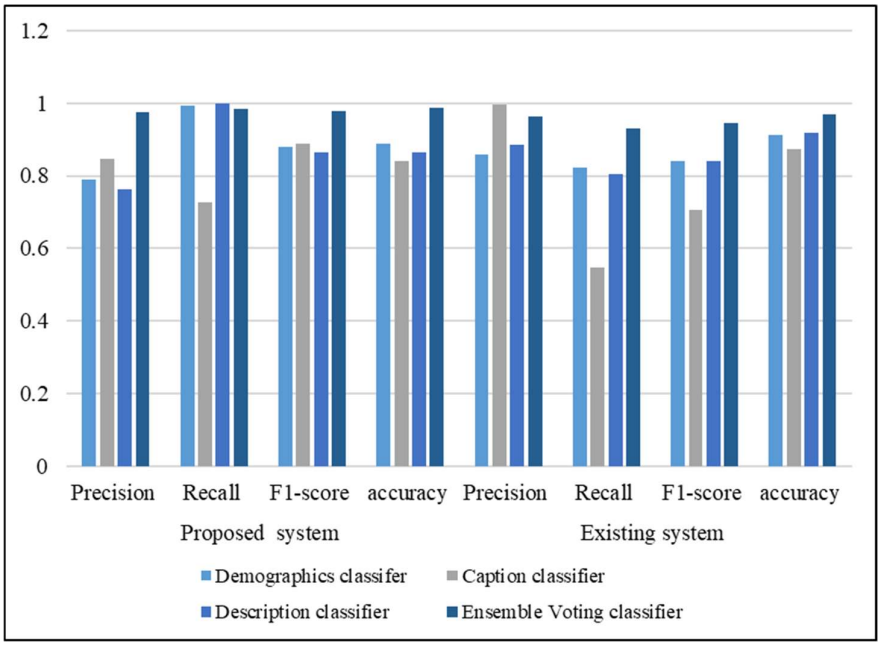

Figure 8. Evaluation of the classifier with existing system

We have compared the results with existing system [17] forall classifiers demographic, captions and description classifiers as shown in figure 8. The existing system achieved the overall accuracy of 97% and f1 score is 94.50%, with the dataset used as up to March, 2017. The proposed system uses the large dataset of user profiles up to the Feb 2022 consisting of themore no. of user profiles. The overall accuracy of the proposed system is 98.75% and f1 score is 97.95% with the more no. of samples in dataset.

5. CONCLUSION AND FUTURE SCOPE

In this paper, we present a scammer detection model for online social networks that utilize a demographic classifier, caption classifier, description classifier, and ensemble classifierhas shown promising results. The output of our proposed approach is tested and evaluated on the dataset which is created by us using the evaluation parameters like precision, recall, f1 score, and accuracy. This multi-classifier approach allows for a more comprehensive assessment of potential scammer accounts, utilizing various classifiers to analyze different aspects of a user's profile and content.We showed that our proposed method has achieved 98.67% accuracy in detecting thescammer's profile in OSN frauds like online dating, compromised accounts,and fake profiles at early stages. We investigated other methods also and compared their results with our proposed model. The experimental analysis and results indicate that the proposed model outperforms as compared to state of art systems.

However, it's important to note that no model is infallible, and there may still be instances where scammers can evade detection. Therefore, continuous evaluation and improvement of the model is necessary to stay ahead of evolving scamming tactics.Overall, the use of a multi-classifier approach for scammer detection in online social networks is a step towards creating

a safer online environment for users, and future research and development in this area may further improve the effectiveness of such models.

Further research and development in this area to enhance the effectiveness of the model, such as incorporating more advanced machine learning techniques, social network graph analysis, real-time monitoring, natural language generation techniques, and user feedback.

REFERENCES

1. C. C. Yang, S. M. Holden, M. D. Carter, and J. J. Webb, "Social media social comparison and identity distress at the college transition: A dual-path model," *Journal of Adolescence*, vol. 69, pp. 92-102, 2018.
2. A. K. Jain, S. R. Sahoo, and J. Kaubiyal, "Online social networks security and privacy: comprehensive review and analysis," *Complex & Intelligent Systems*, vol. 7, no. 5, pp. 2157-2177, 2021.
3. T.R. Soomro, and M.Hussain, "Social Media-Related Cybercrimes and Techniques for Their Prevention," *Appl. Comput. Syst*, vol. 24, pp.9-17, 2019.
4. J.Johnson, "Global digital population", https://www.statista.com/statistics/617136/digitalpopulation-worldwide. Jan 27 2021.
5. A. Homsi, J. Al Nemri, N. Naimat, , H.A. Kareem, M. Al-Fayoumi, and M.A. Snober, "Detecting Twitter Fake Accounts using Machine Learning and Data Reduction Techniques", *In DATA*, pp. 88-95, 2021.
6. J. Heidemann, M. Klier, and F. Probst, "Online social networks: A survey of a global phenomenon", *Computer networks*, vol.56, no.18, pp. 3866-3878, 2012.
7. P. Patel, K. Kannoorpatti, B. Shanmugam, S. Azam, and K.C. Yeo, "A TheoreticalRreview of Social Media Usage by Cyber-Criminals", in 2017 International Conference on Computer Communication and Informatics (ICCCI), IEEE, pp. 1-6, 2017.
8. S. Maniraj, P. Harie Krishnan, G. T. Surya, and R. Pranav, "Fake Account Detection using Machine Learning and Data Science", International Journal of Innovative Technology and Exploring Engineering (IJITEE), vol. 9, no. 1, 2019.
9. M. Zabielski, Z. Tarapata, R. Kasprzyk, and K. Szkółka, "Profile cloning detection in online social networks," Computer Science and Mathematical Modelling, 2016
10. New Data Shows FTC Received 2.8 Million Fraud Reports from Consumers in 2021 | Fed-eral Trade Commission. Federal Trade Commission, www.ftc.gov, 22 Feb. 2022, https://www.ftc.gov/news-events/news/press-releases/2022/02/new-data-shows-ftc-re-ceived-28-million-fraud-reports-consumers-2021-0.
11. A. Rauf, S. Khusro, S. Mahfooz, and R. Ahmad, "A Robust System Detector for Clone Attacks on Facebook Platform", Journal of Research, vol.13, no.4, pp. 71-80, 2016.
12. P. Sowmya and Chatterjee, "Detection of Fake and Cloned Profiles in Online Social Networks", Proceedings 2019: Conference on Technologies for Future Cities (CTFC), March 2019.
13. P. Bródka, M. Sobas, and H. Johnson, "Profile cloning detection in social networks", in 2014 European Network Intelligence Conference, pp. 63-68, September 2014.

14. N. Kumar, and P. Dabas, “Detection and Prevention of Profile Cloning in Online Social Networks”, in 2019 5th International Conference on Signal Processing, Computing and Control (ISPCC), pp. 287-291, October 2019.
15. G. Kontaxis, I. Polakis, S. Ioannidis, and E.P. Markatos, “Detecting social network profile cloning”, in 2011 IEEE international conference on pervasive computing and communications workshops (PERCOM Workshops), pp. 295-300, March 2011
16. K. De Jong, ‘Detecting the online romance scam : Recognizing images used in fraudulent dating profiles’, 2019.
17. Suarez-Tangil, Guillermo, et al. "Automatically dismantling online dating fraud." IEEE Transactions on Information Forensics and Security 15 (2019): 1128-1137.’.
18. S. Al-Rousan, A. Abuhussein, F. Alsubaei, O. Kahveci, H. Farra, and S. Shiva, ‘Social-Guard: Detecting Scammers in Online Dating’, IEEE Int. Conf. Electro Inf. Technol., vol. 2020-July, no. August, pp. 416–422, 2020, doi:10.1109/EIT48999.2020.9208268
19. A.N. Hakimi, S. Ramli, , M. Wook, N. Mohd Zainudin, N.A. Hasbullah, N. Abdul Wahab, and N.A. Mat Razali, “ Identifying Fake Account in Facebook Using Machine Learning”, in International Visual Informatics Conference”, pp. 441-450, Springer, Cham, November,2019.
20. S.R. Sahoo, B.B Gupta,” Fake profile detection in multimedia big data on online social networks”, International Journal of Information and Computer Security, vol. 12, no. 2-3, pp.303-331.
21. Y. Elyusufi, Z. Elyusufi, and M.H.A Kbir, “Social networks fake profiles detection based on account setting and activity”, in Proceedings of the 4th International Conference on Smart City Applications, pp. 1-5, October,2019.
22. S. Kiruthiga, and A. Kannan, “Detecting cloning attack in Social Networks using classification and clustering techniques”, in 2014 International Conference on Recent Trends in Information Technology, pp. 1-6, April 2014.
23. Thomas D, Jennifer G,and Aravind S, Predicting Trust and Distrust in Social Net-works,Privacy, Security, Risk and Trus (PASSAT),Third Inernational Conference on Social Computing (SocialCom),IEEE, Boston, Massachusetts, USA, (2011).
24. Yu, H, M , Kaminsky, P, B. Gibbons, and Flaxman, A, SybilGuard: Defending Against Sybil Attacks Via Social Networks,Networking, IEEE/ACM Transactions on Vol.16(3), (2008), PP.576-589.
25. Yu, H, Gibbons , P, B, Kaminiski, M , Feng,X. SybilLimitANearOptimal Social Network Defense against Sybil Attacks , IEEE Symposium on Security and Privacy Conference, Okland, CA, 18-22, May, (2008). PP:3-17.
26. Cao, Q ,Sirivianos, M, Yang, X, and Pregueiro, T, Aiding the Detection of Fake Accounts in Large Scale Social Online Services, 9th USENIX Conference on Networked Systems Design and Implementation, San Jose, CA, USA ,(2012), pp:15-15
27. Danezis, G. , and Mittall, P SybillInfer: Detecting Sybil Nodes Using Social Net-works Network and Distributed Syste Security Symposium -(2009),[Online].Available: http://libra.msra.cn/Publication/4727139/sybilinfer-detecting-sybilnodes-using-social-networks,(Consult May 2015).

28. Yang,Z,Wilson C ,Wang, X , Gao, T , Zhao,B, Y, and Dai, Y., Uncovering Social Network Sybils in theWildACM Transactions on Knowledge Discovery from Data (TKDD) 8(1) (2014).
29. Fire, M, Katz, G, and Elovici,Y, Strangers Intrusion Detection-Detecting Spammers and Fake Profiles in Social Networks Based on Topology Anomalies, HUMAN 1 (1), (2012) ,pp.26-39.
30. Khayyambashi,M,R, and Rizi, F,S, An Approach for Detecting Profile Cloning in Online Social Networks, 7th International Conference on e-Commerce in Developing Countries with Focus on e-Security, IEEE, , Kish Island, Iran. (2013), PP. 1-12.
31. Gao,H, Hu,J, and Wilson,C. Detecting and Characterizing Social Spam Campaigns Proceedings of The 10th ACM SGCOMM International Measurement Conference ,ACM, ,Melbourne, Australia (2010). PP.35-75.
32. Baden, R ,Spring, N , and Bhattacharjee, B Identifying Close Friends on the Internet, 8th ACM Workshop on Hot Topics in Networks , HotNets , New York, NY, USA, (2009).
33. Wang, G ,Konolige, T ,Wilson,C, Wang, X ,Zheng, H , and Zhao, Y , You are How You Click: ClickStream Analysis for Sybil Detection, Proc. Of the 22nd USENIX Security Symposium (USENIX Security 2013),Washungton,USA ,(2013) PP.1-15.
34. Wang G. ,Mohanlal, M, Wilson, C, Wang , X, Metzger, M, Zheng,H, and Zhao,B.Y. Social Turning Tests: Croudsourcing Sybil Detection , The 20th Network and Distributed System Security Symposium NDSS,San Diego, CA United States(2013).
35. S. Ranjana, R. Sathian, and M.D. Kamalesh, " Fake Profile Detection in Facebook", In International Conference on Emerging Trends and Advances in Electrical Engineering and Renewable Energy, pp. 725-732, Springer, Singapore, 2020, March.
36. C.R Liyanage, and S.C. Premarathne, "Clustered Approach for Clone Detection in social media", International Journal of Advanced Science Engineering Information technology, vol. 11, no. 1, pp. 99-104, 2021.
37. S.S. Bama, M.I. Ahmed, and A. Saravanan, "A mathematical approach for mining web content outliers using term frequency ranking", Indian Journal of Science and Technology, vol. 8, no. 14, pp.1-5, 2015.
38. A. Homsi, J. Al Nemri, N. Naimat, , H.A. Kareem, M. Al-Fayoumi, and M.A. Snober, "Detecting Twitter Fake Accounts using Machine Learning and Data Reduction Techniques", In DATA, pp. 88-95, 2021
39. A. ElAzab,"Fake accounts detection in twitter based on minimum weighted feature. World Academy of Science, Engineering and Technology", International Journal of Computer, Electrical, Automation, Control and Information Engineering, vol. 10, no. 1, pp. 13-18, 2016.
40. M.Y. Kharaji, and F.S. Rizi, "An iac approach for detecting profile cloning in online social networks", International Journal of Network Security & Its Applications, vol. 6, no.1, pp. 75-90, 2014.
41. M.B. Albayati, and A.M. Altamimi, "An empirical study for detecting fake Facebook profiles using supervised mining techniques", Informatica, vol. 43, no. 1, pp. 77–86.03, 2019.

42. M.B. Albayati, and A.M. Altamimi, "Identifying Fake Facebook Profiles Using Data Mining Techniques", Journal of ICT Research & Applications, vol. 13, no. 2, pp. 107-117, 2019.
43. Singh, and Banerjee, "Fake (Sybil) Account Detection Using Machine Learning", Proceedings of International Conference on Advancements in Computing & Management (ICACM), Available at SSRN:https://ssrn.com/abstract=3462933 or http://dx.doi.org/10.2139/ssrn.3462933, October 2, 2019.
44. F. Ajesh, S.U. Aswathy, F.M. Philip, and V. Jeyakrishnan , "A hybrid method for fake profile detection in social network using artificial intelligence", Security Issues and Privacy Concerns in Industry 4.0 Applications, pp.89-112, 2021.
45. M. Zare, S.H. Khasteh, and S. Ghafouri, "Automatic ICA detection in online social networks with PageRank", Peer-to-Peer Networking and Applications, vol. 13 no.5, 1297-1311, 2020.
46. M. Suriakala, and S. Revathi, "Privacy protected system for vulnerable users and cloning profile detection using data mining approaches", in 2018 Tenth International Conference on Advanced Computing (ICoAC), pp. 124-132, December, 2018.
47. http://scamdigger.com/ accessed 2nd Jan 2022
48. https://www.datingnmore.com/ accessed 6nd Jan 2022
49. M. T. Whitty "The scammers' persuasive techniques model:Development of a stage model to explain the online datingromance scam", British Journal of Criminology, 53(4):665–684, 2013.
50. https://huggingface.co/Salesforce/blip-image-captioning-base-"BLIP:Bootstrapping Language-Image Pre-training for Unified Vision-Language Understanding and Generation"
51. Rahman, M.S., Halder, S., Uddin, M.A. et al. "An efficient hybrid system for anomaly detection in social networks. Cybersecurity", (2021). https://doi.org/10.1186/s42400-021-00074-w
52. Ampomah, Ernest Kwame, Zhiguang Qin, and Gabriel Nyame. 2020. "Evaluation of Tree-Based Ensemble Machine Learning Models in Predicting Stock Price Direction of Movement" Information 11, no. 6: 332. https://doi.org/10.3390/info11060332
53. Maryamsadat Hejazi & Yashwant Prasad Singh (2013) ONE-CLASS SUPPORT VECTOR MACHINES APPROACH TO ANOMALY DETECTION, Applied Artificial Intelligence, 27:5, 351-366, DOI: 10.1080/08839514.2013.785791
54. H. Yu, C. Ho, Y. Juan, and C. Lin. Libshorttext: A libraryfor short-text classification and analysis. Technical report,2013.