JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

A Smart Framework for Detecting Instagram Fake Accounts using Machine Learning and Correlation & Singular Value Decomposition Techniques

Apoorva Dwivedi[*1], Prof. (Dr.) Devendra Agarwal[2], Km Divya[3], Upasana Dugal[4], Vipin Rawat[5], Dr. Yusuf Perwej[6]

[1*]Assistant Professor, Department of Computer Science & Engineering, IIMT College of Engineering, Greater Noida, U.P
[2]Dean (Academics), Goel Institute of Technology & Management, Lucknow, U.P
[3]Assistant Professor, Department of Computer Science & Engineering, Ambalika Institute of Management & Technology, Lucknow
[4]Assistant Professor, Department of Computer Science & Engineering, Babu Banarasi Das University, Lucknow (BBDU), Lucknow
[5]Assistant Professor, Department of Computer Science & Engineering, Ambalika Institute of Management & Technology, Lucknow
[6]Professor, Department of Computer Science & Engineering, Goel Institute of Technology & Management, Lucknow, U.P

Abstract: We now have access to a wealth of knowledge thanks to the Internet and social media. On the other hand, these developments have made it easier for scammers to operate. Fake accounts are used to disseminate spam, post fake news, offer bogus product evaluations, and even meddle in political campaigns. False accounts may cause a great deal of harm in the corporate world, including financial waste, reputational harm, legal issues, and a host of other issues. Due to the massive expansion of the online social network, the number of false accounts is significantly rising; as a result, these accounts must be identified. This study focuses on the identification of automated and phony Instagram profiles that generate phony engagement. It's simple to replicate the appearance of having a large following on social media by making a phony Instagram account. It is impractical and infeasible to do data analysis and mining on a large volume of data since it takes a very long time to process. Data reduction techniques are used to generate a reduced version of the dataset that is significantly smaller in volume while keeping the integrity of the original data. This can speed up the processing time for data analysis. In order to rapidly and effectively identify bogus accounts, academics have been working to create and improve machine learning (ML) and correlation & SVD techniques. Correlation and Singular Value Decomposition (SVD), together with seven machine learning classifier algorithms J48, Random Forest, K-Nearest Neighbors (KNN), Neural Networks, logistic regression, and Naive Bayes Algorithm were employed to discover these accounts. Results indicate that using a neural network algorithm and correlation together produces the greatest accuracy 99.43%.

Keywords: Fake Identity, Instagram, Detection, Classification, Kaggle Dataset, Fake Profile, Singular Value Decomposition (SVD), Correlation, Machine Learning, Phishing.

1. Introduction

Online social networks (OSNs) [1] have been incredibly popular over the past decade among users of all demographics because they allow users to cooperate, meet new people, organize events, exchange news, and express personal viewpoints. Digital networking has spread widely in recent years. Globally, social media profiles [2] are common, especially those that identify as celebrities and high-profile individuals. False social media profiles often and intentionally assume the personas of famous people and real brand champions. Such strategies are used to rob and undermine customer interest and loyalty bases and to foster mistrust among the ardent fans of original brands. Future cyberattacks, particularly those that target business executives, political figures, and leaders, sometimes start with fake profiles. To undertake fraudulent actions using social media, fake accounts are made [3]. On Instagram, fake and automated accounts pose a number of problems and cast doubt on the legitimacy of the service. These accounts are made with the intention of doing harm, participating in things like spamming, phishing, identity theft, and spreading false information. Additionally, automated accounts, sometimes known as bots, have the ability to alter engagement numbers, skew social interactions [4], and trick people into thinking they are authentic accounts. As a result, the platform's reputation is threatened and user confidence is diminished [5]. Additionally, several government organizations utilize OSNs [6] to deliver their own services, communicate official news and announcements to their constituents, and disseminate details about various activities.

The process of authenticating user accounts becomes increasingly difficult as a result of fake profiles, often known as Cyber-Bots [7], which are created by online criminals. Some false accounts are created to mimic the profile of an actual

person, while others are created as generic profiles to act as fake followers. In order to make choices more quickly, machine learning (ML) [8] algorithms are employed to automate and enhance the process of identifying bogus accounts. In order to identify false and spam Instagram accounts, this project intends to create and deploy a machine learning [9], correlation, and singular value decomposition model. The goal of this research is to create a machine learning model, train it, and then propose a detection technique for automatic spamming accounts and false Instagram account [10] profiles. The suggested method produced insightful and effective performance assessments while achieving a greater true positive rate and a lower true negative rate to conduct a thorough analysis of the pertinent techniques.

2. Background

On how to recognize fraudulent or copied accounts on social networking sites like Facebook, Twitter, LinkedIn, etc., several writers have contributed their study. A few writers recommended using straightforward statistical analysis to find duplicate profiles by comparing the similarity of the profiles, checking the behaviours, and checking the IP address [11]. Social media is now changing very fast; many people in society depend on these services, particularly for marketing [12] campaigns and for celebrities and politicians who use social media to try to attract followers and supporters [13]. False accounts created on behalf of people and organizations therefore have the ability to damage and ruin their reputations, ultimately leading to a fall in the number of their genuine likes and followers. Several identification algorithms are used to categories social media profiles by looking at specific existing qualities. ML algorithms [14] are one of the other detecting techniques for better account categorization. Numerous writers have suggested using various variables in their detection models and maintain that doing so significantly enhances the efficiency of separating bogus profiles from authentic ones. Network-based, content-based, temporal-based, profile-based, and action-based characteristics are the five categories that these traits fall under [15].

To identify duplicated profiles, the PageRank method was eventually employed. The celebrity's profile was used for the assessment even though the model was developed using the MapReduce [16] framework. A data mining method was suggested as a way to detect duplicated profiles and susceptible and fraudulent users while protecting user privacy [17]. Support vector machine (SVM) [18], random forest, and deep neural network [19] are just a few of the classification methods that have been proposed. The popularity of social media platforms is correlated with the proliferation of fraudulent accounts. Such a phony account or identity is created for a variety of nefarious reasons [20]. Due to the potential for encouraging participation in several online and offline crimes, the use of these false identities poses a particular threat to society [21]. A study of Twitter hackers was suggested in order to better understand their behavioral traits. Over the course of a month, 100,000 communications were collected for the analysis [22]. Two distinct spammer kinds utilizing various trolling methods were evaluated. Additionally, three important organizations [23] discovered a number of measures for detecting spammers, including account assets, social connections, and profile attributes. For categorizing profiles based on topological data and meta data, a two-layer method is also recommended [24]. By using a supervised technique and a finite number of seed cases, Ritter et al.'s [25] weekly supervised extraction of computer security events from Twitter makes it straightforward to create and train extractors for additional categories of events.

On the other hand, Wang et al. [26] developed an approach for differentiating tweets produced by legitimate users from those made by spammers by employing n-grams in combination with additional sets of characteristics. Last but not least, the feature vector for each tweet included n-grams, emotion, content, and user metadata (details about the individual who sent the particular tweet). They experimented with binary term-frequency (tf) and tf-idf (i.e., Term Frequency times Inverse Document Frequency) approaches to create uni+bi-gram or bi+tri-gram n-grams. They employed two freely accessible datasets for their investigations [27]. They used Recall, Precision, and F-Measure to assess their studies after training a Random Forest Classifier with various sets of features.

However, a realistic, effective ML-based defense must be meticulously constructed with traits that are impervious to manipulation by the opposition, substantial amounts of labelled ground truth data for model training, and a system that can scale to all active users on an OSN (perhaps billions of them). To get around these issues, they provide Deep Entity Classification (DEC), an ML [28] framework that locates abusive accounts in OSNs that have evaded earlier, more traditional abuse detection techniques. Karami, et al. [29] proposed the idea of profiling fake news spreaders on social media using psychological and motivational factors. In Herzallah et al. [30], the effectiveness of methods that gauge each feature's influence on the dataset was compared to the effectiveness of machine learning classifiers trained with all conceivable characteristics. In specifically, they employed Information Gain, CoM, and Relief-F to rank features depending on their usefulness towards the classification after categorizing users into spammers or non-spammers using conventional machine learning methods [31]. They said that the repute of the account, account age, average time between tweets, average tweet length, and average mentions per tweet make up the finest combination of attributes.

Spam on social networks was detected using a number of current features, and new features were added to increase performance. As classification techniques, multi-layer perceptron’s (MLPs), support vector machines (SVM), and random forests were trained [32]. The best outcomes were obtained using the random forest approach, which had an accuracy rate of 96.30%. Wanda, et al. [33] proposed dynamic deep learning as a technique for identifying abnormal nodes in online social networks. By using connection data from nodes and training huge features using dynamic deep learning [34] architecture, the authors offer a model to categories harmful vertices. To identify cloned profiles, a technique that compares the strength of the connections between two profiles with active buddy lists and like counts was suggested [35]. A clever method named FBChecker was suggested that uses supervised learning algorithms and behavioral and informational aspects to identify bogus Facebook profiles [36]. The approach was applied by filtering the records with missing values and filling in the missing values using the KNN schema [37]. The aforementioned papers, however, lack comprehensive experimental investigation to back up the conclusions. The results show that ID3 [38] gives increased detection accuracy [39], and the authors have expanded their work with

unsupervised clustering techniques. Spam detection methods include fuzzy logic and multilayer perceptron evaluation through neural networks. It was discovered that the fuzzy [40] applied logic managed the enormous data set effectively and took very little time to identify spammers in seconds [41]. Three separate feature sets and three different machine learning techniques [42] decision trees, random forests, and SVM are used in the process of real-time cybersecurity [7] account identification on Twitter [43]. Each user's activity data is divided into several continuous time periods using the DeepScan model. DeepScan uses timeseries features that are comparatively more inclusive and descriptive than regular features [44] by utilizing deep learning technology.

3. Fake Accounts

A sort of automated account fraud called fake account creation involves hackers using bots to establish fake accounts expressly for carrying out fraudulent activities including skewing product ratings, disseminating misleading information, [45] or disseminating malware. False accounts are now frequently linked to social media (such as Twitter bots attempting to sway elections), but there are numerous other applications as well. For instance, the use of bogus accounts to sway customer ratings is now so pervasive that it has given rise to a thriving business of its own. False accounts are frequently employed in the gaming and gambling industries to get access to welcome bonuses, discount codes, or other rewards that may then be turned into cash [46]. Other instances include the falsification of survey findings, the use of free services improperly, or the mining of cryptocurrencies via a free cloud computing account. Lastly, attackers can log into a large number of "legitimate" accounts using well-known usernames and passwords to decrease the attack's login failure rate by using phony accounts to mask credential stuffing attempts.

4. Fake Instagram Accounts

With more than a billion users globally, Instagram has emerged as one of the most widely used social media platforms in recent years. However, the increase of phony accounts coincides with its popularity. These accounts could be formed for phishing, spamming, or even identity theft. Because of this, it's critical to understand how to recognize phony Instagram accounts and remain secure. To fool people into thinking they are the person being impersonated, the individual who establishes the bogus account frequently utilizes the name, photographs, and other personal information of another person. However, creating false accounts is not always done for impersonation. There are several reasons why someone could establish fake Instagram accounts [47]. Others may use them for more nefarious goals, such as distributing false information or harassing others, while other individuals build them for amusement or as a joke. People occasionally use fictitious accounts in order to track or snoop on others without disclosing their genuine identities [48]. Instagram's rules of service forbid the creation of false accounts, and doing so can result in significant repercussions including account suspension or legal action. Fake Instagram accounts are created for a variety of purposes [49].

4.1 Anonymity

Some people who establish phony accounts do so to conceal their genuine identities because they wish to remain anonymous. This could be for a variety of reasons, including as protecting their privacy, not wanting to share personal information, or preventing cyberbullying.

4.2 Impersonation

A username and other personal information may be used to establish a fake account and steal your identity. Using this, someone may pose as you and publish defamatory content that would harm your reputation.

4.3 Cyberbullying

False accounts might be set up by cyberbullies so they can stalk or threaten you in secret. They may leave offensive remarks or send you harsh direct messages [50]. Your well-being and mental health may suffer as a result.

4.4 Catfishing

In order to trick individuals into dating or friendships, someone establishes a phony account known as catfishing. There are several reasons why someone could do this, including attention, affirmation, or money scams.

4.5 Spamming

In an effort to steal their personal information or gain financial gain, some phony accounts are made to bombard other users with promotional or harmful content, such spam messages or phishing links.

5. Challenges with Fake Instagram Accounts

Undoubtedly, Instagram is among the most widely used social networking sites on the planet. The Datar portal estimates that by January 2023, there will be 1.318 billion users worldwide. Fake postings and user accounts are among Instagram's major issues. If you've used Instagram for any length of time, you've undoubtedly come across a bogus account or post. However, it's possible that they [51] were unaware that any of the Instagram postings or profiles were phony. The attention given to an account's follower level is the major cause of fraudulent Instagram accounts. And the challenging task of naturally gaining such following. The people who purchase phony accounts with large follower numbers are just trying to get ahead of the competition faster than usual without putting in any effort. Instagram challenges are a specific kind of competition where participants enter by posting a photo. Although each challenge has its own features, they are frequently based on a theme, such as submitting a photo of your pet or your attire. Participants are frequently urged to include the brand's product or service in their submissions when a challenge is sponsored by a company.

5.1 Fake Bio

In order to imitate an authentic Instagram account and gain followers, bogus accounts frequently use a photo of an attractive guy or woman, frequently in undress, as their profile photo. They are hoping that by posting this image, anyone who come across the account would find it to be more appealing and fascinating. Nevertheless, if you look at their posts, you won't find the person from their profile photo. As hardly everyone who creates phony accounts bothers to make them appear natural, it's fairly unusual to also utilize images [52][53] of

models or celebrities, or to not have a profile photo at all. The bio is another item you ought to look at. It frequently makes irrational claims or has an absurd backstory.

5.2 Fake Followers

One of the most reliable signs that an Instagram account is phony is this. You'll discover a highly improbable scenario when you contrast the account's follower count with the quantity of articles and stories it has shared. It is quite implausible that a profile with only 10 posts would have amassed 10,000 organic followers. An extraordinarily high figure in the subsequent category is another item to watch out for [54]. The following feature on fake Instagram accounts frequently reaches its limit of 7,500.

5.3 Fake Comments

A bogus account won't publish anything that features the person in their profile photo, as was already established. In most cases, the posts won't be about the occupation they identify in their bio. Even with a cursory scan at their profile, these discrepancies ought to be pretty obvious. Take note of how frequently they post and the time period in which the majority of the posts are published if you wish to investigate if the account is false further [55]. To achieve their aim as fast as possible, those who create and maintain false accounts with the intention of selling them frequently create the account and share a number of posts to give the impression that the account is active. You'll see that their posts are rarely read, frequently unconnected, and occasionally outright bizarre.

6. Kaggle Dataset

A Kaggle dataset, often referred to as a Kaggle Kernel, is a collection of data offered by organizations, pupils, and graduates. Competitors can solve issues using these data sets, or they can utilize them as a practice simulation. There are a lot of datasets [56] on Kaggle's site. Users of Kaggle may submit datasets, collaborate with other data scientists and machine learning experts, identify datasets they wish to employ in AI [57] model construction, and compete to solve data science problems. Spammers and fake accounts are a big issue on Instagram and other social networking sites. My senior year research focused on this, and I sought to identify methods of identifying them using machine learning. Although there may be a few mistaken accounts in the spammers list, as shown in figure 1, we have personally recognized the spammer/fake accounts included in this dataset after carefully reviewing each case. As a result, the dataset has a high degree of accuracy. We continued by the dataset while taking several categorization techniques into account. We also give numerical feature types some thought. The numerical aspects of other categorical features had also been altered. We attempted to evaluate the most important features because the dataset has several attributes. We excluded non-significant attributes from our model. Applying several ML algorithms [58] to the dataset is crucial. After that, as part of the data pretreatment stage, we had normalized the data. In order to preserve the information, it is important to convert the dispersed huge numerical values to a common scale of [0,1] without distorting the value range-differences. Additionally, certain algorithms require this phase in order to properly represent the data.

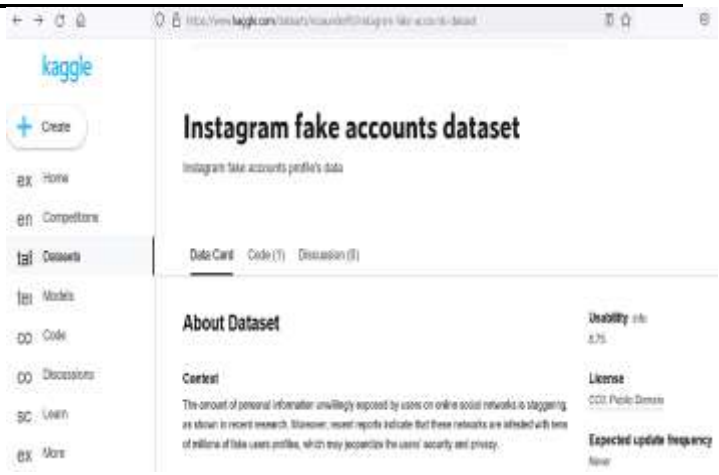

Figure 1. The Kaggle Dataset

7. Data Reduction Techniques

Data reduction is the process of correcting, organizing, and simplifying numerical or alphabetical digital information that has been obtained empirically or experimentally. Data reduction can have two separate goals: reducing the quantity of data records by removing useless material, or producing summary data and statistics at various levels of aggregation for different purposes [59]. Data reduction is a method used in data mining to shrink a dataset while keeping the most crucial details intact. This can be useful when the dataset is too big to analyses effectively or when the dataset contains a lot of information that is redundant or useless. Singular Value Decomposition (SVD) and correlation are two data reduction techniques that are covered in this section.

7.1 Singular Value Decomposition (SVD)

A matrix's Singular Value Decomposition (SVD) is a decomposition into three different matrices. It communicates significant geometrical and theoretical insights regarding linear transformations and has several intriguing algebraic characteristics. It also has a few significant uses in data science [60]. I'll try to explain the mathematical reasoning behind SVD and its geometrical significance in this essay. The factorization of a matrix A into the product of three matrices A = U DV ^{T}, where the columns of U and V are orthonormal and the matrix D is diagonal with positive real entries, is known as singular value decomposition. Many tasks benefit from using the SVD. Here, we provide a few illustrations. Firstly, the data matrix A is frequently close to a matrix of low rank, therefore it might be helpful to choose a low rank matrix that closely approximates the data matrix. We'll demonstrate how to obtain the matrix B of rank k that most closely resembles A from the singular value decomposition of A. In fact, we can accomplish this for all k. A should be a m x n matrix. The A, B, and C Singular Value Decompositions (SVD).

$$A = U\Sigma V^T,$$

Where is a m n diagonal matrix with nonnegative diagonal elements, U is m m and orthogonal, V is n n and orthogonal, and

$$\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_p, \quad p = \min\{m, n\}$$

The decomposition of A's singular values, sometimes referred to as the singular values of A, is very helpful and reveals many details about the variable, including its range, null space, rank, and 2-norm condition number.

7.2 Correlation

A statistical technique called correlation analysis is used to determine the link between two variables and gauge how strongly two variables are linearly related. The degree of change in one variable as a result of the other's change is determined [61] via correlation analysis. A high correlation indicates a strong association between the two variables, whilst a low correlation indicates a poor correlation between the two variables. The correlation coefficient is always between -1 and +1, with -1 denoting a negative connection between X and Y and +1 denoting a positive correlation.

$$r = \frac{n(\sum xy) - (\sum x)(\sum y)}{\sqrt{\left[n\sum x^2 - (\sum x)^2\right]\left[n\Sigma y^2 - (\sum y)^2\right]}}$$

Where r is the correlation coefficient.

The correlation coefficient is the name given to this output number. When the correlation coefficient is positive, it indicates that the two variables are positively correlated, meaning that when the value of variable X rises, the value of variable Y rises as well. However, if the correlation coefficient is negative, it indicates that there is a negative correlation between the two variables (i.e., as variable X grows, variable Y declines). However, when the correlation coefficient has a value of 0, it indicates that there is no correlation between the two variables. Even if they come from diverse areas of the organization, correlation analysis may help you comprehend the connections between various data. You may get fresh perspectives and become aware of interdependencies as a result of this information.

8. Machine Learning

Machine learning is a branch of artificial intelligence, which is broadly characterized as a machine's ability to emulate intelligent human behavior [62]. Systems using artificial intelligence (AI) are used to complete difficult jobs in a way that is similar to how individuals solve problems. With AI, computer models are designed to exhibit "intelligent behaviours" that are comparable to those of people. These are machines that can read a text written in natural language, recognize a scene in a picture, or perform an action. One of the most intriguing technologies ever developed is machine learning. A wide range of applications for machine learning exist to solve issues and automate several industries. The development of machine learning [63] methodologies, the growth of informational resources, and advancements in computer power are the key reasons of this. Without a doubt, machine learning has been used to a wide range of sophisticated and contemporary network management and operation challenges. Several machine learning studies have been undertaken for specific networking technologies or specialized networking businesses. Machine learning, a method of data analysis [64], automates the development of analytical models. Machine learning enables data filtering and inference. Going beyond just learning or absorbing information, it requires putting knowledge to use and developing it through time [65]. Finding and using hidden patterns in "training" data is the primary objective of machine learning. The patterns discovered can be used to categories new data or match it to already existing categories [66]. There are several categorization algorithms available today, but it is impossible to judge which one is superior than the others. Its depends on the kind of application being utilized and the nature of the dataset being used.

8.1 K-Nearest Neighbors (KNN)

The supervised learning technique K-nearest neighbor's (KNN) is used for both regression and classification. By computing the distance between the test data and all of the training points, KNN tries to predict the proper class for the test data. Then choose the K spots that are closest to the test data. The KNN method determines which of the classes of the 'K' training data the test data will belong to, and the class with the highest probability is chosen. The value in a regression situation is the average of the 'K' chosen training points.

8.2 J48

A straightforward C4.5 decision tree for classification is the J48 classifier. A binary tree is produced. The classification problem is where the decision tree method shines. Using this method, a tree is built to represent the categorization process. Each tuple in the database is classified once the tree has been constructed and applied to it. J48 ignores the missing values while creating a tree, meaning that the value of the item may be anticipated based on the knowledge of the attribute values for the other entries. The fundamental concept is to categories the data according to the attribute values for each item that may be found in the training sample. J48 supports categorization using either rules derived from decision trees or by them [67].

8.3 Random Forest

Random Forest is an ensemble methodology that can handle both regression and classification tasks, using a number of decision trees and a technique called Bootstrap and Aggregation, sometimes known as bagging. The key tenet of this approach is to combine many decision trees to obtain the outcome rather than relying just on one decision tree [68]. The core learning models of Random Forest are multiple decision trees. By picking rows and attributes from the dataset at random, we produce sample datasets for each model. This part is referred to as Bootstrap.

8.4 Artificial Neural Networks (ANNs)

Artificial neural networks contain artificial neurons, sometimes referred to as units. These units, which are layered in various levels, make up the system's overall Artificial Neural Network. Whether a layer includes a dozen units or millions of units depends on how the complex neural networks will be employed to find the hidden patterns in the dataset [69]. In addition to input, output, and output layers, artificial neural networks often incorporate hidden layers. The input layer receives data from the outside world that the neural network needs to analyses or learn. This data is then turned into information that is suitable for the output layer after going through one or more hidden layers. Last but not least, the output layer generates a response to incoming input that is represented by an artificial neural network.

8.5 Naive Bayes (NB)

The naive Bayes classifier, a probabilistic classifier based on the Bayes theorem [70], assumes that each feature contributes equally and independently to the target class. Assuming that each feature is independent of the others and does not interact, the NB classifier states that each feature separately and equally influences the likelihood that a sample belongs to a certain class. The NB classifier works well on big datasets with high dimensionality and is simple to build. The NB classifier is noise-resistant and appropriate for real-time applications.

8.6 Logistic Regression

Early in the 20th century, the biological sciences began to employ logistic regression. Then, it was put to many different social sciences uses. When the dependent variable (target) is categorical, logistic regression is utilized. Based on a given dataset of independent variables, logistic regression calculates the likelihood that an event will occur, such as voting or not voting. Given that the result is a probability, the dependent variable's range is 0 to 1. In logistic regression, the odds that is, the likelihood of success divided by the probability of failure are transformed using the logit formula. The natural logarithm of odds or the log odds are other names for this.

8.7 Decision Tree Algorithm

The non-parametric supervised learning approach used for classification and regression applications is the decision tree. It creates a tree structure resembling a flowchart where each internal node represents a test on an attribute, each branch a test result, and each leaf node (terminal node) a class label. A stopping requirement, such as the maximum depth of the tree or the least number of samples needed to split a node, is reached by repeatedly separating the training data into subsets depending on the values of the attributes. The amount of impurity or unpredictability in the subsets is measured using metrics like entropy or Gini impurity, and the Decision Tree algorithm chooses [71] the optimum attribute to split the data depending on these metrics during training. Finding the property that maximizes information gain or impurity reduction following the split is the objective.

9. Proposed Approach

Our suggested model is thoroughly presented in this section. The model, which comprises three primary phases data preprocessing, data reduction, and data classification is depicted in figure 2. Processing the dataset came first in our effort, and in the second stage, we included additional reduction strategies. The data was filtered and reduced using several reduction mechanisms in the reduction phase to prepare it for the classification phase, where the filtered data was run through various classification algorithms and the results were shown. There are several ways to filter data, ranging from simple procedures like removing properties to more complex ones like singular value decomposition (SVD). the numerous hierarchical clustering algorithms and clustering schemes, such as k-means, that facilitate unsupervised learning. To identify the collection of critical classes for the classification performance, attribute selection uses a variety of search techniques and selection requirements. We continued by the dataset while taking several categorization techniques into account. We attempted to evaluate the most important features because the dataset has several attributes. We did not incorporate non-significant attributes in our model. Applying several ML algorithms to the dataset is crucial. After that, as part of the data pretreatment stage, we had normalized the data. Additionally, certain algorithms require this phase in order to properly represent the data. Using normalization is a helpful strategy when the data distribution is uncertain, unpredictable, or not Gaussian, as normalization involves rescaling the chosen characteristics.

9.1 Dataset

33,935 influencers are represented in the dataset, totaling 10,180,500 Instagram posts (300 posts per influencer). Post information and picture files are both included in the collection. The following data is contained in post metadata files in JSON format: the caption, user tags, hashtags, timestamp, sponsorship, likes, and comments, among other things. Since a post may contain more than one picture file, as shown in table 1, the collection comprises 12,933,406 image files in JPEG format. A JSON file and the related image files have the same name if a post only contains one picture file. However, the names of the JSON file and the associated image files alter if a post has several images. As a result, we also provide you access to a JSON-Imagemapping file, which lists all the image files that correspond to each post's information.

Table 1. The List of Collected Features

Index	Feature	Description
1	**user_id**: (Integer)	Primary ID
2	**username**: (String)	Username of the user
3	**email**: (String)	Email address of the user
4	**password**: (String)	Hashed password for the user
5	**bio**: (String)	Short bio of the user
6	**profile_picture**: (String)	URL to the profile picture
7	**post_id**: (Integer)	Primary ID
8	**user_id**: (Integer)	ID of the user who made the post
9	**datetime_added**: (Datetime or Timestamp Integer)	When was this post added?
10	**image_url**: (String)	URL to the image
11	**caption**: (String)	Caption for the post
12	**comment_id**: (Integer)	Primary ID
13	**user_id**: (Integer)	ID of the user who made this comment
14	**post_id**: (Integer)	ID of the post the comment was made on
15	**datetime_added**: (Datetime or Timestamp Integer)	When was this comment added?
16	**comment**: (String)	Text of the comment
17	**like_id**: (Integer)	Primary ID
18	**user_id**: (Integer)	ID of the user who liked the post
19	**post_id**: (Integer)	ID of the post that was liked
20	datetime_added: (Datetime or Timestamp Integer)	When was this like added?
21	**follower_id**: (Integer)	Primary ID
22	**user_id**: (Integer)	ID of the user who is following
23	**following_user_id**: (Integer)	ID of the user being followed
24	**sender_id**: (Integer)	ID of the user who sent the message
25	**receiver_id**: (Integer)	ID of the user who received the message
26	**hashtag_id**: (Integer)	Primary ID
27	**hashtag**: (String)	The actual hashtag text, unique

Finally, we have looked at the experimental results that were achieved following the advised methods. These results demonstrated that the recently created approach performs

remarkably well in terms of accuracy and has a significant capacity to detect fake accounts. In order to reliably identify the Kaggle Dataset and discover data patterns, the experimental system depicted in figure 2 was developed. The techniques used to train the system include K-Nearest Neighbor's (KNN), J48 [72], Random Forest, Artificial Neural Network (ANN) [73], Naive Bayes (NB), Logistic Regression, and Decision Tree Algorithm. By taking into account the Kaggle training dataset, these models properly create trained models. Once these models have been developed, they may be used to accurately categories data using the four test Kaggle Datasets that were produced [73]. Table 2 and table3 below displays the performance results for this model. We also compared the proposed Machine Learning classifier to the benchmark models in terms of performance metrics like classification accuracy for the job of recognizing Instagram phony accounts. Instagram accounts were included in the Kaggle Dataset, of which around 80% were utilized for training and 20% for testing.

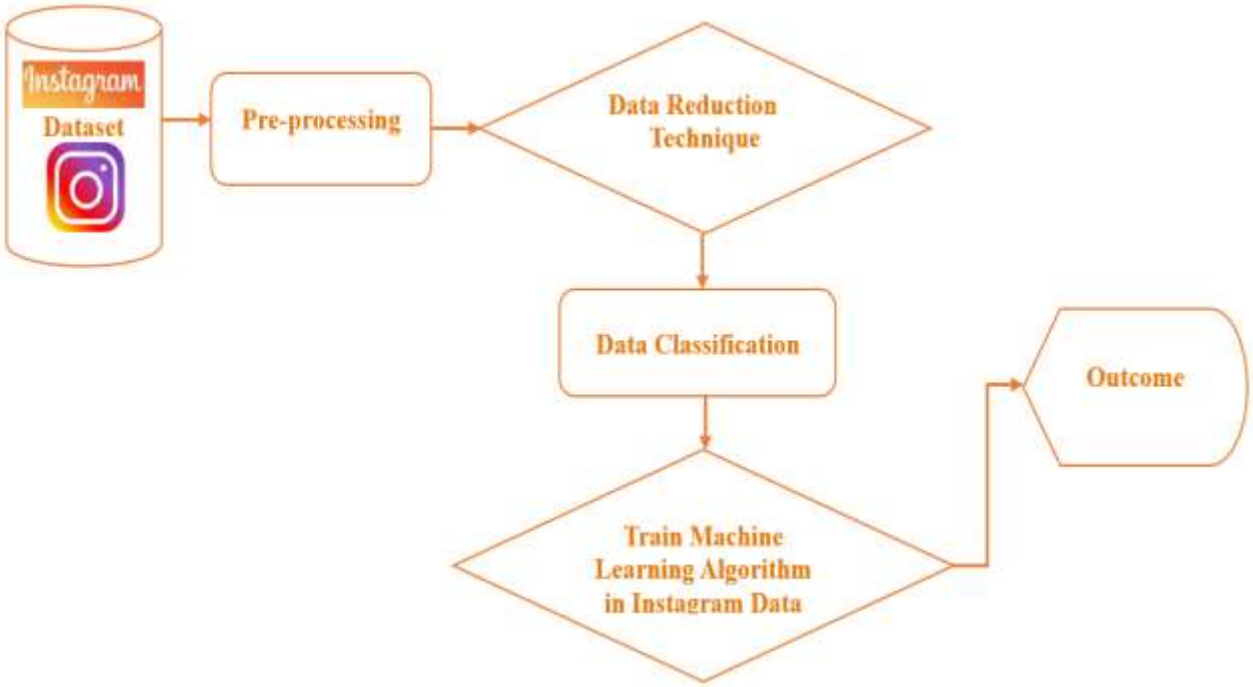

Figure 2. The Architecture of Instagram Fake Accounts Detection Proposed Model

10. Outcome Evaluation

This section compares a few things and discusses how well the machine learning algorithm works to spot Instagram fake accounts. Certain performance measures are measured for their performance studies [75]. We also used the collected dataset, which includes information on 33,935 Kaggle Dataset users, in all of our experiments. Singular Value Decomposition (SVD) and correlation were used as two reduction approaches in the experiment, along with seven classification algorithms.

Table 2. The Model's Performance of Classification Outcomes for the Kaggle Test Datasets with Correlation Reduction Technique

Machine Learning Algorithms with Correlation Reduction Technique	Performance Summary for 80% - 20%	
	Imperfection Rate	Precision
K-Nearest Neighbour's (KNN)	6.44%	93.56 %
Naive Bayes (NB)	17.87%	82.13 %
Random Forest	1.34%	98.66 %
Logistic Regression	11.31%	88.69%
Artificial Neural Network	0.57%	99.43%
J48	2.00%	98.00 %
Decision Tree	1.52%	98.48%

Table 3. The Model's Performance of Classification Outcomes for the Kaggle Test Datasets with Singular Value Decomposition (SVD) Reduction Technique

Machine Learning Algorithms with Singular Value Decomposition (SVD) Reduction Technique	Performance Summary for 80% - 20%	
	Imperfection Rate	Precision
K-Nearest Neighbour's (KNN)	7.34%	92.66 %
Naive Bayes (NB)	21.87%	78.13 %
Random Forest	4.24%	95.76 %
Logistic Regression	13.31%	86.69%
Artificial Neural Network	1.87%	98.13%
J48	6.58%	93.42%
Decision Tree	5.52%	94.48%

All seven classification methods are then used after the data has been reduced using one of the procedures. The accuracy of the methods for the total precision is shown in figure 3 and figure4.

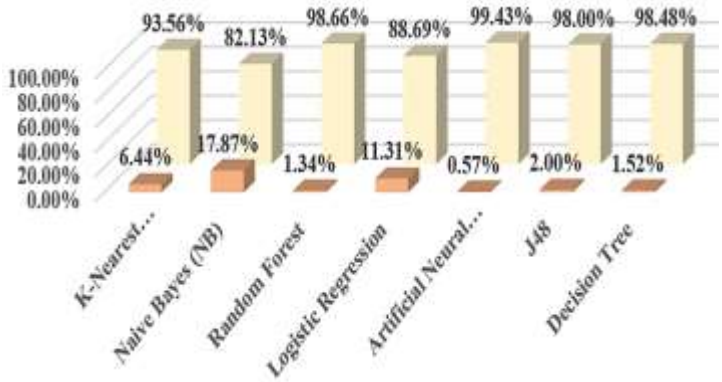

Figure 3. The Model's Performance Summary with Correlation Reduction Technique

The accuracy of the algorithm is shown in this example as a percentage (%). With an 80 to 20 ratio, the findings demonstrate the algorithm's effectiveness [76]. The algorithm's imperfection rate, which displays how frequently [77] the method is misclassified, is used to gauge performance. That might be calculated using this equation.

$$\textbf{Imperfection Rate} = 100 - \textbf{Precision}$$

Furthermore, we found that ANNs outperform the other used approaches (see figure 5). As a result, plans for putting the recommended data model into practice may be explored soon.

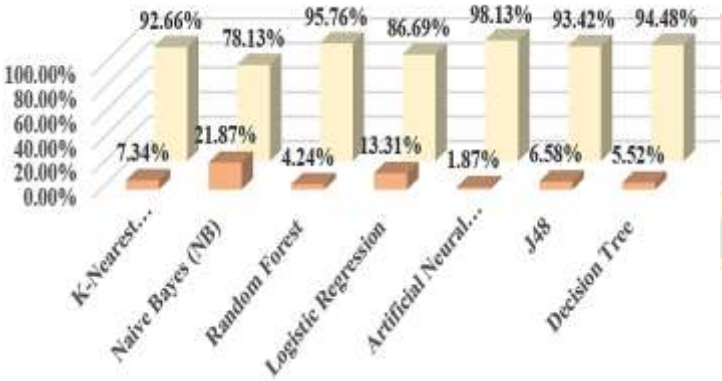

Figure 4. The Model's Performance Summary with Singular Value Decomposition (SVD) Reduction Technique

10.1 Achievement Evaluations

A substantial number of Instagram accounts were inputted into the proposed technique in order to identify Instagram phony accounts. The following equations were used to evaluate the performance measures accuracy (A), precision (P), recall (R), and F-measure (F) after extracting the classification results in terms of the confusion matrix.

$$\textbf{Accuracy (A)} = \frac{T_p + T_n}{T_p + F_p + T_n + F_n}$$

$$\textbf{Precision (P)} = \frac{T_p}{T_p + F_p}$$

$$\textbf{Recall (R)} = \frac{T_p}{T_p + F_n}$$

$$\textbf{F-Measure (F)} = \frac{2 * P * R}{P + R}$$

The actual negative figure is the number of Instagram fake accounts that were correctly detected as being unfavorable for a certain Instagram fake accounts category, Tn. Tp is the quantity of Instagram fake accounts items correctly categorized [78] as positive for a certain news category. The false positive value (Fp) is the proportion of news items [79] that should not be classified as belonging to the given category but are yet given that label. False negative value (Fn) is the proportion of objects that should not be classed as belonging to a certain category but are nonetheless. Based on this criteria, a number of query articles were sent to the system, and the A,P,R, and F values were noted. This conclusion is summarized across several classification algorithms kinds in Table 2, which also assesses the average accuracy values for the ANNs [80] models.

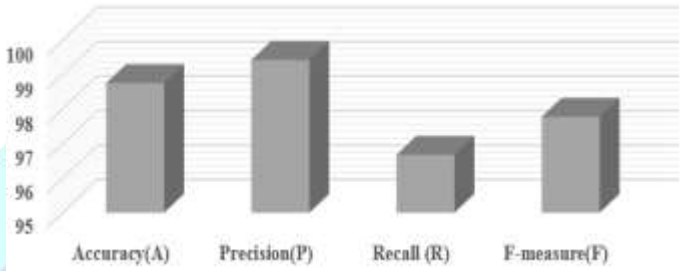

Figure 4. The ANNs Model Performance Metrics Accuracy(A), Precision(P), Recall(R), and F-measure(F)

11. Conclusion

People use social media sites like Instagram, Facebook, Twitter, and others to do a number of tasks nowadays, including learning new things, buying, selling, and marketing consulting, teaching, developing communities, and hanging out with friends. As a result, social media platforms gradually cover a. They have a big impact on how we conduct our lives every day and their influence grow as we age. There is growing worry over certain accounts manipulating and spreading misleading information on social media networks. Our major objective is to have a better knowledge of how classifier algorithms work in conjunction with correlation to identify phony Instagram profiles on social media. Our model's data pre-processing and reduction stages were created to make the dataset usable for categorization. The data was then classified using a variety of machine learning methods in order to identify the optimum accuracy. An additional reduction strategy has also been shown to considerably improve the effectiveness, accuracy, and training time of ANNs. These methods are used to increase the Artificial Neural Networks' (ANNs) classification accuracy for bogus Instagram accounts. Singular Value Decomposition (SVD) and correlation were applied using seven classifier methods throughout the data reduction phase. The greatest accuracy of 99.43%, according to the results, is provided by the Artificial Neural Networks (ANNs) algorithm and correlation data reduction. The least accurate algorithm, the Naive Bayes

algorithm, with correlation data reduction, reach 82.1% accuracy. The testing of several alternative experiments, approaches, and algorithms is still on the future to-do list. By doing more thorough research and analysis, we intend to evaluate various reduction approaches and classification algorithms.

Data Availability

The study used open-source dataset and is accessed from the Weblink- https://www.kaggle.com/datasets/rezaunderfit/instagram-fake-and-real-accounts-dataset

References

[1] Ms Farah Shan, Versha Verma, Apoorva Dwivedi, Dr. Y. Perwej, Ashish Kumar Srivastava, "Novel Approaches to Detect Phony Profile on Online Social Networks (OSNs) Using Machine Learning", International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT), Volume 9, Issue 3, Pages 555-568, May-June 2023, DOI: 10.32628/CSEIT23903126

[2] Edosomwan, Simeon & Prakasan, S.K. & Kouame, D. & Watson, J. & Seymour, T.,"The history of social media and its impact on business", Journal of Applied Management and Entrepreneurship. 16. 79-91, 2011

[3] Vishal Verma, Apoorva Dwivedi, Kajal, Prof. (Dr.) Devendra Agarwal, Dr. Fokrul Alom Mazarbhuiya, Dr. Y. Perwej, "An Evolutionary Fake News Detection Based on Tropical Convolutional Neural Networks (TCNNs) Approach", International Journal of Scientific Research in Science and Technology (IJSRST), Print ISSN: 2395-6011, Online ISSN: 2395-602X, Volume 10, Issue 4, Pages 266-286, July-August-2023, DOI: 10.32628/IJSRST52310421

[4] Sachin Bhardwaj, Apoorva Dwivedi, Ashutosh Pandey, Dr. Y. Perwej, Pervez Rauf Khan, "Machine Learning-Based Crowd Behavior Analysis and Forecasting", International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT), ISSN: 2456-3307, Volume 9, Issue 3, Pages 418-429, May-June 2023-2023, DOI: 10.32628/CSEIT23903104

[5] Jaana Isohatala, Hanna Jarvenoja and Sanna Jarvela, "Participation in cognitive-oriented and socio-emotionally oriented interaction during collaborative learning", Earli SIG 27-Online Measures of Learning Processes conference, 2016

[6] Z. Chu, S. Gianvecchio, H. Wang and S. Jajodia, "Detecting automation of twitter accounts: Are you a human bot or cyborg*f*", IEEE Transactions on Dependable and Secure Computing, vol. 9, no. 6, pp. 811-824, 2012

[7] Y. Perwej, Prof. (Dr.) Syed Qamar Abbas, Jai Pratap Dixit, Nikhat Akhtar, Anurag Kumar Jaiswal, "A Systematic Literature Review on the Cyber Security", International Journal of Scientific Research and Management (IJSRM), ISSN (e): 2321-3418, Volume 9, Issue 12, Pages 669 - 710, 2021, DOI: 10.18535/ijsrm/v9i12.ec04

[8] Mathews Chibuluma and Josephat Kalezhi, "Application of a modified perceptron learning algorithm to monitoring and control", 2017 IEEE PES Power Africa

[9] Venkata K. S. Maddala, Dr. Shantanu Shahi, Dr. Yusuf Perwej, H G Govardhana Reddy, "Machine Learning based IoT application to Improve the Quality and precision in Agricultural System", European Chemical Bulletin (ECB), ISSN: 2063-5346, SCOPUS, Hungary, Volume 12, Special Issue 6, Pages 1711 – 1722, May 2023, DOI: 10.31838/ecb/2023.12.si6.157

[10] Aleksei Romanov, Alexander Semenov, Oleksiy Mazhelis and Jari Veijalainen, "Detection of Fake Profiles in Social Media - Literature Review", WEBIST 2017 - 13th International Conference on Web Information Systems and Technologies, vol. 1, pp. 363-369, 2017

[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] Prof. Kameswara Rao Poranki, Dr. Yusuf Perwej, Dr. Asif Perwej," The Level of Customer Satisfaction related to GSM in India ", TIJ's Research Journal of Science & IT Management – RJSITM, International Journal's-Research Journal of Science & IT Management of Singapore, ISSN: 2251-1563, Singapore, Volume 04, Number: 03, Pages 29-36, 2015

[13] Blair, S.J., Bi, Y., Mulvenna, M.D., "Aggregated topic models for increasing social media topic coherence", Applied Intelligence, 50(1): 138-156, 2020

[14] Shweta Pandey, Rohit Agarwal, Sachin Bhardwaj, Sanjay Kumar Singh, Dr. Y. Perwej, Niraj Kumar Singh, "A Review of Current Perspective and Propensity in Reinforcement Learning (RL) in an Orderly Manner" , International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT), Volume 9, Issue 1, Pages 206-227, January-February-2023, DOI: 10.32628/CSEIT2390147

[15] 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

[16] Y. Perwej, Md. Husamuddin, Fokrul Alom Mazarbhuiya,"An Extensive Investigate the MapReduce Technology", International Journal of Computer Sciences and Engineering (IJCSE), E-ISSN : 2347-2693, Volume-5, Issue-10, Page No. 218-225, 2017, DOI: 10.26438/ijcse/v5i10.218225

[17] 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, 2018

[18] Asif Perwej, Prof. (Dr.) K. P. Yadav, Prof. (Dr.) Vishal Sood, Dr. Yusuf Perwej, " An Evolutionary Approach to Bombay Stock Exchange Prediction with Deep Learning Technique", IOSR Journal of Business and Management (IOSR-JBM), e-ISSN: 2278-487X, p-ISSN: 2319-7668, USA,

Volume 20, Issue 12, Ver. V, Pages 63-79, 2018, DOI: 10.9790/487X-2012056379

[19] Y. Perwej, "The Bidirectional Long-Short-Term Memory Neural Network based Word Retrieval for Arabic Documents", Transactions on Machine Learning and Artificial Intelligence (TMLAI), Society for Science and Education, United Kingdom (UK), ISSN 2054-7390, Volume 3, Issue 1, Pages 16 - 27, 2015, DOI: 10.14738/tmlai.31.863

[20] Y. Boshmaf, D. Logothetis, G. Siganos, J. Lería, J. Lorenzo, M. Ripeanu, K. Beznosov, H. Halawa, "Íntegro: Leveraging victim prediction for robust fake account detection in large scale osns", Computers & Security, vol. 61, pp. 142-168, 2016

[21] Asif Perwej, Kashiful Haq, Y. Perwej, "Blockchain and its Influence on Market", International Journal of Computer Science Trends and Technology (IJCST), ISSN 2347 – 8578, Volume 7, Issue 5, Pages 82- 91, 2019, DOI: 10.33144/23478578/IJCST-V7I5P10

[22] H. Faris and Aljarah, "Improving email spam detection using content-based feature engineering approach,"in Jordan conference on applied electrical engineering and computing technologies (AEECT), pp. 1–6, Aqaba, Jordan, 2017

[23] Y. Perwej, Kashiful Haq, Uruj Jaleel, Firoj Perwej, "Block Ciphering in KSA, A Major Breakthrough in Cryptography Analysis in Wireless Networks", International Transactions in Mathematical Sciences and Computer, India, ISSN-0974-5068, Volume 2, No. 2, Pages 369-385, 2009

[24] Firoj Parwej, Nikhat Akhtar, Dr. Yusuf Perwej, "A Close-Up View About Spark in Big Data Jurisdiction", International Journal of Engineering Research and Application (IJERA), ISSN: 2248-9622, Volume 8, Issue 1, (Part -I1), Pages 26-41, 2018, DOI: 10.9790/9622-0801022641

[25] Ritter, A.; Wright, E.; Casey, W.; Mitchell, T. Weakly Supervised Extraction of Computer Security Events from Twitter. In Proceedings of the 24th International Conference on World Wide Web, Geneva, Switzerland, 18–22, pp. 896–905, 2015

[26] B. Wang, A. Zubiaga, M. Liakata, R. Procter, Making the most of tweetinherent features for social spam detection on twitter, 2015, arXiv preprint arXiv:1503.07405

[27] Kajal, Uttam Kumar Singh, Dr. Nikhat Akhtar, Satendra Kumar Vishwakarma, Niranjan Kumar, Dr. Y. Perwej, "A Potent Technique for Identifying Fake Accounts on Social Platforms", International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT), Volume 9, Issue 4, Pages 308 - 324, 2023, DOI: 10.32628/CSEIT2390425

[28] Y. Perwej, Firoj Parwej, "A Neuroplasticity (Brain Plasticity) Approach to Use in Artificial Neural Network", International Journal of Scientific & Engineering Research (IJSER), France , ISSN 2229 – 5518, Volume 3, Issue 6, Pages 1- 9, 2012, DOI: 10.13140/2.1.1693.2808

[29] Karami, Mansooreh, Tahora H. Nazer, and Huan Liu. "Profiling Fake News Spreaders on Social Media through Psychological and Motivational Factors." Proceedings of the 32nd ACM Conference on Hypertext and Social Media, 2021

[30] W. Herzallah, H. Faris, O. Adwan, Feature engineering for detecting spammers on twitter: Modelling and analysis, J. Inf. Sci. 44 (2) (2018) 230–247

[31] Saurabh Sahu, Km Divya, Dr. Neeta Rastogi, Puneet Kumar Yadav, Dr. Y. Perwej, "Sentimental Analysis on Web Scraping Using Machine Learning Method" , Journal of Information and Computational Science (JOICS), ISSN: 1548-7741, Volume 12, Issue 8, Pages 24-29, August 2022, DOI: 10.12733/JICS.2022/V12I08.535569.67004

[32] Neha Kulshrestha, Nikhat Akhtar, Dr. Y. Perwej, "Deep Learning Models for Object Recognition and Quality Surveillance", International Conference on Emerging Trends in IoT and Computing Technologies (ICEICT-2022), ISBN 978-10324-852-49, SCOPUS, Routledge, Taylor & Francis, CRC Press, Chapter 75, pages 508-518, Goel Institute of Technology & Management, Lucknow, May 2022, DOI: 10.1201/9781003350057-75

[33] Wanda, Putra, and Huang J. Jie. "Deep Friend: finding abnormal nodes in online social networks using dynamic deep learning." Social Network Analysis and Mining 11.1 , , : 1- 12, 2021

[34] Y. Perwej, "An Evaluation of Deep Learning Miniature Concerning in Soft Computing", International Journal of Advanced Research in Computer and Communication Engineering (IJARCCE), ISSN (Online): 2278-1021, ISSN (Print): 2319-5940, Volume 4, Issue 2, Pages 10 - 16, 2015 , DOI: 10.17148/IJARCCE.2015.4203

[35] 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

[36] Y. Perwej , Asif Perwej , "Forecasting of Indian Rupee (INR) / US Dollar (USD) Currency Exchange Rate Using Artificial Neural Network", International Journal of Computer Science, Engineering and Applications (IJCSEA), Academy & Industry Research Collaboration Center (AIRCC), USA , Volume 2, No. 2, Pages 41- 52, 2012, DOI: 10.5121/ijcsea.2012.2204

[37] 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

[38] Y. Perwej, Firoj Parwej, Nikhat Akhtar, "An Intelligent Cardiac Ailment Prediction Using Efficient ROCK Algorithm and K- Means & C4.5 Algorithm", European Journal of Engineering Research and Science (EJERS), Bruxelles, Belgium, ISSN: 2506-8016 (Online), Vol. 3, No. 12, Pages 126 – 134, 2018, DOI: 10.24018/ejers.2018.3.12.989

[39] 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

[40] Dr. F. A. Mazarbhuiya, Dr. Y. Perwej, "The Mining Hourly Fuzzy Patterns from Temporal Datasets", International Journal of Engineering Research & Technology (IJERT), ISSN: 2278-0181, Volume 4, Issue 10, Pages 555-559, 2015, DOI: 10.17577/IJERTV4IS100576

[41] P. Tehlan, R. Madaan, K.K. Bhatia, A spam detection mechamism in social media using Soft computing, in: 6th International Conference on Computing for Sustainable Global Development, INDIACom), pp. 950-955, 2019

[42] Nikhat Akhtar, Y. Perwej, Firoj Parwej, Jai Pratap Dixit, "A Review of Solving Real Domain Problems in Engineering for Computational Intelligence Using Soft Computing" Proceedings of the 11th INDIACom; INDIACom-2017; SCOPUS, IEEE Conference ID: 40353, 2017 4th International Conference on "Computing for Sustainable Global Development", ISSN 0973-7529; ISBN 978-93-80544-24-3, Pages 706–711, Bharati Vidyapeeth's Institute of Computer Applications and Management (BVICAM), New Delhi (INDIA), 01st - 03rd March, 2017

[43] Ç.B. Aslan, R.B. Saglam, S. Li, Automatic detection of cyber security related accounts on online social networks: twitter as an example, in: Proceedings of the 9th International Conference on social media and Society, 2018, pp. 236e240, https:// doi.org/10.1145/3217804.3217919

[44] G. Qingyuan, Y. Chen, X. He, Z. Zhuang, T. Wang, H. Huang, X. Wang, X. Fu, DeepScan: Exploiting deep learning for malicious account detection in location-based social networks, IEEE Commun. Mag. 56 (2018) 21e27, https://doi.org/ 10.1109/MCOM.2018.1700575

[45] F.A. Ozbay and B. Alatas, "Fake news detection within online social media using supervised artificial intelligence algorithms", Physica A: Statistical Mechanics and its Applications, vol. 540, pp. 123174, 2020

[46] Aleksei Romanov, Alexander Semenov, Oleksiy Mazhelis and Jari Veijalainenin, "Detection of fake profiles in social media-Literature review", International Conference on Web Information Systems and Technologies, vol. 2, pp. 363-369, 2018

[47] B. Erşahin, Ö. Aktaş, D. Kılınç and C. Akyol, "Twitter fake account detection", 2017 International Conference on Computer Science and Engineering (UBMK), pp. 388-392, 2017

[48] Y. Perwej, Dr. Nikhat Akhtar, Neha kulshrestha, Pavan Mishra, "A Methodical Analysis of Medical Internet of Things (MIoT) Security and Privacy in Current and Future Trends", Journal of Emerging Technologies and Innovative Research (JETIR), ISSN-2349-5162, Volume 09, Issue 1, Pages 346 - 371, January 2022, DOI: 10.6084/m9.figshare.JETIR2201346

[49] A. El Azab, A. M. Idrees, M. A. Mahmoud and H. Hefny, "Fake account detection in twitter based on minimum weighted feature set", Int. Sch. Sci. Res. Innov, vol. 10, no. 1, pp. 13-18, 2016

[50] Van Der Walt, Estée and Jan Eloff, "Using machine learning to detect fake identities: bots vs humans", IEEE Access, vol. 6, pp. 6540-6549, 2018

[51] P. G. Efthimion, S. Payne and N. Proferes, "Supervised machine learning bot detection techniques to identify social twitter bots", SMU Data Science Review, vol. 1, no. 2, pp. 5, 2018

[52] Y. Perwej , Firoj Parwej, Asif Perwej, "Copyright Protection of Digital Images Using Robust Watermarking Based on Joint DLT and DWT ", International Journal of Scientific & Engineering Research (IJSER), France, ISSN 2229-5518, Volume 3, Issue 6, Pages 1- 9, June 2012

[53] Y. Perwej, Asif Perwej, Firoj Parwej, "An Adaptive Watermarking Technique for the copyright of digital images and Digital Image Protection", International journal of Multimedia & Its Applications (IJMA), which is published by Academy & Industry Research Collaboration Center (AIRCC) , USA , Volume 4, No.2, Pages 21- 38, 2012, DOI: 10.5121/ijma.2012.4202

[54] W. Zhang and H. Sun, "Instagram spam detection", 2017 IEEE 22nd Pacific Rim International Symposium on Dependable Computing (PRDC), pp. 227-228, Jan 2017

[55] N Reputation, M Forwarding and H A. Rate, "A Fuzzy Collusive Attack Detection Mechanism for Reputation Aggregation in Mobile Social Networks: A Trust Relationship Based Perspective", Mobile Information Systems, vol. 2016, no. 4, pp. 1-16, 2016

[56] G. Gousios and D. Spinellis, "GHTorrent: Github's data from a firehose", MSR '12: Proc. of the 9th Working Conference on Mining Software Repositories, pp. 12-21, 2012

[57] Y. Perwej, Asif Perwej, "Forecasting of Indian Rupee (INR) / US Dollar (USD) Currency Exchange Rate Using Artificial Neural Network", International Journal of Computer Science, Engineering and Applications (IJCSEA), Academy & Industry Research Collaboration Center (AIRCC), USA, Volume 2, No. 2, Pages 41- 52, 2012, DOI: 10.5121/ijcsea.2012.2204

[58] Y. Perwej, "The Bidirectional Long-Short-Term Memory Neural Network based Word Retrieval for Arabic Documents", Transactions on Machine Learning and Artificial Intelligence (TMLAI), Society for Science and Education, United Kingdom (UK), ISSN 2054-7390, Volume 3, Issue 1, Pages 16 - 27, 2015, DOI: 10.14738/tmlai.31.863

[59] Z. A. Bhuiyan, G. Wang, T. Wang, A. Rahman and J. Wu, "Content-centric event-insensitive big data reduction in internet of things", 2017 IEEE Glob. Commun. Conf. GLOBECOM 2017 - Proc., 2017

[60] F. Anowar, S. Sadaoui and B. Selim, "Conceptual and empirical comparison of dimensionality reduction algorithms (pca kpca lda mds svd lle isomap le ica t-sne)", Computer Science Review, vol. 40, pp. 100378, 2021

[61] avid R. Hardoon, Sandor Szedmak and John Shawe-Taylor, "Canonical correlation analysis; An overview with application to learning methods", Neural Computation, vol. 16, no. 12, pp. 2639-2664, 2004

[62] Y. Perwej, "An Evaluation of Deep Learning Miniature Concerning in Soft Computing", International Journal of Advanced Research in Computer and Communication Engineering (IJARCCE), ISSN (Online): 2278-1021, ISSN (Print): 2319-5940, Volume 4, Issue 2, Pages 10 - 16, 2015, DOI: 10.17148/IJARCCE.2015.4203

[63] Li, D. Li and C. Zhang, "An Application of Machine Learning in the Criterion Updating of Diagnosis Cancer", IEEE, pp. 187-190, 2005

[64] Y. Perwej, Bedine Kerim, Mohmed Sirelkhtem Adrees, Osama E. Sheta, "An Empirical Exploration of the Yarn in Big Data", International Journal of Applied Information Systems (IJAIS) – ISSN: 2249-0868, Foundation of Computer Science FCS, New York, USA, Volume 12, No.9, Pages 19-29, December 2017, DOI: 10.5120/ijais2017451730

[65] A. El Azab, A. M. Idrees, M. A. Mahmoud and H. Hefny, "Fake account detection in twitter based on minimum weighted feature set", Int. Scholarly Sci. Res. Innov., vol. 10, no. 1, pp. 13-18, 2016

[66] Y. Perwej, Dr. Shaikh Abdul Hannan, Firoj Parwej, Nikhat Akhtar, "A Posteriori Perusal of Mobile Computing", International Journal of Computer Applications Technology and Research (IJCATR), ATS (Association of Technology and Science), India, ISSN 2319–8656 (Online), Volume 3, Issue 9, Pages 569 - 578, 2014, DOI: 10.7753/IJCATR0309.1008

[67] Margaret H. Danham,S. Sridhar, " Data mining, Introductory and Advanced Topics", Person education , 1st ed., 2006

[68] Gareth James, Daniela Witten, Trevor Hastie and Robert Tibshirani, An introduction to statistical learning, Springer, pp. 204, 2013

[69] B. K. Bose, "Neural network applications in power electronics and motor drives – an introduction and perspective", IEEE Trans. Ind. Electron., vol. 54, no. 1, pp. 14-33, 2007

[70] N. Akhtar, Devendera Agarwal, "An Efficient Mining for Recommendation System for Academics", International Journal of Recent Technology and Engineering (IJRTE), ISSN 2277-3878 (online), SCOPUS, Volume-8, Issue-5, Pages 1619-1626, 2020, DOI: 10.35940/ijrte.E5924.018520

[71] F. Takahashi and S. Abe, "Decision-tree-based multiclass support vector machines", Proc. 9th Int. Conf. Neural Inf. Process. (ICONIP), pp. 1418-1422, Nov. 2002

[72] Y. Perwej, Firoj Parwej, Nikhat Akhtar, "An Intelligent Cardiac Ailment Prediction Using Efficient ROCK Algorithm and K- Means & C4.5 Algorithm", European Journal of Engineering Research and Science (EJERS), Bruxelles, Belgium, ISSN: 2506-8016 (Online), Vol. 3, No. 12, Pages 126 – 134, 2018, DOI: 10.24018/ejers.2018.3.12.989

[73] Y. Perwej, Nikhat Akhtar, Firoj Parwej, "The Kingdom of Saudi Arabia Vehicle License Plate Recognition using Learning Vector Quantization Artificial Neural Network", International Journal of Computer Applications (IJCA), USA, ISSN 0975 – 8887, Volume 98, No.11, Pages 32 – 38, http://www.ijcaonline.org., 2014, DOI: 10.5120/17230-7556

[74] https://www.kaggle.com/datasets/rezaunderfit/instagram-fake-and-real-accounts-dataset

[75] Davis, J., and Goodrich, M. : The relationship between Precision- Recall and ROC curves. 23rd international conference on machine learning, 233-240, 2006

[76] Jesse Davis and Mark Goadrich. The relationship between precision-recall and roc curves. In Proceedings of the 23rd international conference on Machine learning, pages 233–240. ACM, 2006

[77] Takaya Saito and Marc Rehmsmeier. The precision-recall plot is more informative than the roc plot when evaluating binary classifiers on imbalanced datasets. PloS one, 10(3):e0118432, 2015

[78] Y. Perwej, Dr. Faiyaz Ahamad, Dr. Mohammad Zunnun Khan, Nikhat Akhtar, "An Empirical Study on the Current State of Internet of Multimedia Things (IoMT)", International Journal of Engineering Research in Computer Science and Engineering (IJERCSE), ISSN (Online) 2394-2320, Volume 8, Issue 3, Pages 25 - 42, March 2021, DOI: 10.1617/vol8/iss3/pid85026

[79] Akyon, F. C., & Esat Kalfaoglu, M. (2019). Instagram Fake and Automated Account Detection. Proceedings- 2019 Innovations in Intelligent Systems and Applications Conference, ASYU 2019

[80] J.-W. Lin and J.-S. Chiou, "Active probability backpropagation neural network model for monthly prediction of probabilistic seismic hazard analysis in Taiwan", IEEE Access, vol. 7, pp. 108990-109014, 2019