Identifying Fake Facebook Profiles Using Data Mining Techniques

Mohammed Basil Albayati* **& Ahmad Mousa Altamimi**

Department of Computer Science,
Applied Science Private University, Al Arab St. 21 Amman 11931, Jordan
*E-mail: mohammed.sabri@asu.edu.jo

Abstract. Facebook, the popular online social network, has changed our lives. Users can create a customized profile to share information about themselves with others that have agreed to be their 'friend'. However, this gigantic social network can be misused for carrying out malicious activities. Facebook faces the problem of fake accounts that enable scammers to violate users' privacy by creating fake profiles to infiltrate personal social networks. Many techniques have been proposed to address this issue. Most of them are based on detecting fake profiles/accounts, considering the characteristics of the user profile. However, the limited profile data made publicly available by Facebook makes it ineligible for applying the existing approaches in fake profile identification. Therefore, this research utilized data mining techniques to detect fake profiles. A set of supervised (ID3 decision tree, k-NN, and SVM) and unsupervised (k-Means and k-medoids) algorithms were applied to 12 behavioral and non-behavioral discriminative profile attributes from a dataset of 982 profiles. The results showed that ID3 had the highest accuracy in the detection process while k-medoids had the lowest accuracy.

Keywords: *Facebook; fake profiles; machine learning; supervised algorithms; unsupervised algorithms.*

1 Introduction

In 2003, Mark Zuckerberg started work on a new concept, which eventually turned into the global social network known as Facebook. Since then, Facebook has expanded over the whole world, reaching more than 2.3 billion monthly active users as of December 2018 [1]. A tool such as this changes the way people interact with each other.

Facebook removes communication boundaries so that people can easily connect with others to share life events, stories, or social activities with high availability, reliability, and accessibility [2,3]. This has resulted in a huge number of registered users subscribed to this network [4]. According to the American Academy of Pediatrics, 84% of adolescents in America have a Facebook account, with a total of 2.2 billion users worldwide according to the latest official announcements [5-7]. Facebook being the preferred communication

Received October 13th, 2018, Revised December 12th, 2018, Accepted for publication August 20th, 2019.
Copyright © 2019 Published by ITB Journal Publisher, ISSN: 2337-5787, DOI: 10.5614/itbj.ict.res.appl.2019.13.2.2

platform for so many people, the privacy of users can be the target of scammers [3], for example by creating fake profiles using false information to impersonate the victim in order to steal valuable information or using the user's contacts for abusive actions such as financial fraud [8,9].

This work aimed to address this problem by utilizing data-mining techniques to detect fake profiles on Facebook. Three supervised algorithms (k-NN, SVM, and ID3 Decision Tree) and two unsupervised algorithms (k-Means, and k-medoids) were implemented using RapidMiner Studio [10] on a set of 12 profile attributes and a dataset of 982 profiles (781 real and 201 fake) to validate the conceptual idea. The results showed that the supervised algorithms outperformed the unsupervised algorithms with respect to accuracy. More details about the obtained experimental results are given in Section 4.

The rest of this paper is structured as follows. Section 2 presents related works, the material and methodology are discussed in Section 3, while Section 4 illustrates the experiment and the obtained results. In Section 5, a discussion about the experimental results is given. Finally, the conclusion of this paper is given in Section 6.

2 Related Works

Many approaches have been proposed for detecting the phenomenon of fake profiles on online social networks. Most of them employed supervised algorithms to analyze fake profiles from different perspectives. The authors of [11] proposed a model that employs supervised algorithms (SVM, Naïve Bayes, and Decision Tree) to exploit profile attributes (e.g. 'number of friends', 'education and work', 'gender', and others). The proposed model was implemented using Python scripts on a dataset of 975 profiles extracted from one Facebook account. However, collecting profiles from one account may lead to inaccurate results and may give mistaken observations.

In contrast, the authors of [12] collected their dataset using the Facebook API. The dataset consisted mainly of behavioral attributes ('user online activities' and 'user interactions'). These attributes were characterized through a set of 17 attributes, after which a total of 12 supervised machine-learning techniques were applied to the dataset. The results showed an accuracy of 79%, which is not sufficient. The works [13,14] used a similar approach. Ref. [13] for example utilized three supervised algorithms (Naive Bayes, Jrip, and Decision Tree J48) to identify spam profiles on Facebook and Twitter based on a set of 14 generic features (attributes). Moreover, the algorithms were also used to discover the impact of each attribute on the classification process. On the other hand, Ref. [14] proposed add-on software implemented in the Firefox browser. SW utilized

eight supervised algorithms on a set of fifteen connection features to detect fake profiles on Facebook. None of these works considered unsupervised techniques.

Unsupervised techniques have been considered in [15,16]. Ref. [15] investigated three unsupervised methods to predict whether multiple accounts belong to the same Facebook user. Ref. [16] presents an anomaly detection model that aggregates three major types of attributes (temporal, spatial, spatio-temporal features) to calculate a fourth one (multiple features) represented as one vector passed to the proposed approach. To test their conceptual idea, three popular networks (Facebook, Yelp, and Twitter) were considered. The results showed an accuracy of 66%.

Our work is different from the presented related works in several aspects. Firstly, our work utilized both supervised and unsupervised learning techniques in order to detect fake Facebook profiles. Secondly, most of the presented works utilized behavioral-based attributes, for example [13,14], whereas other works used non-behavioral attributes, for example [17] for detecting fake LinkedIn profiles using supervised mining techniques. In this work, the two types of attributes (behavioral and non-behavioral) were considered using both supervised and unsupervised techniques.

3 Methodology

In this research, supervised and unsupervised machine-learning techniques were utilized to build a model with different attributes and predefined labels of known classes (*fake* and *real*). This facilitates the classification or prediction of unlabeled new data. To do so, 12 of the behavioral and non-behavioral attributes listed in Tables 1-3 were considered in our model, using a dataset consisting of 982 profiles. The data were pre-processed, and missing values were handled using the k-NN algorithm. These were imputed by finding the k-nearest neighbor of the current missing value in the dataset based on the other available information.

Table 1 Attributes previously used.

Attribute	Type	References
Education, Workplace, Introduction "Bio."	Non-Behavioural \ Profile Content Attributes	[12]
No. of tags	Behavioural \ Numerical Attributes	[13] [12]
No. of Mutual Friends	Behavioural \ Numerical Attributes	[18]
No. of posts (Wall Activities)	Behavioural \ Numerical Attributes	[12] [16]
No. of Pages	Behavioural \ Numerical Attributes	[14] [19]

Table 2 The new employed attributes.

Attribute	Type
Profile Picture	Non-Behavioural \ Profile Content Attributes
Living Place	Non-Behavioural \ Profile Content Attributes
Check-In	Non-Behavioural \ Profile Content Attributes
Family Member/ Relationsh	Non-Behavioural \ Profile Content Attributes
No. of Groups	Behavioural \ Numerical Attributes

Table 3 Attributes used in the FFPD model.

Attribute	Description	Justification
Profile Picture*	Visual identification of the user.	Real users use their real pictures more often than fake users.
Work place	Workplace or job title's information,	Real users more often use their real workplace information than fake users.
Education	Attended (school, college, university…etc.) information.	Real users mentioned their education information in their Facebook profiles more often than fake users.
Living Place*	Living place address (city, town, state…etc.) information.	Real users more often use their real living place information than fake users
Check In*	Information for announcing user location.	Real users check into places in their Facebook's profiles more often than fake users.
No. of Posts	Social online activities shared on Facebook	Real users have more online activities than fake users.
No. of Tags	Identify the user by someone else on his/ her wall.	Real Users tagged more often than fake users.
Introduction "Bio."	Introduction information about Facebook's users.	Real users are more often write something about themselves than fake users.
No. of Mutual Friends	Number of the people who are Facebook friends with both users and the target profiles.	Real users have more mutual friends with target profile than fake users, hence gives profile more incredibility.
No. of Pages	Number of pages liked.	Real users usually liked more pages than fake users.
No. of Groups*	Number of groups joined.	Real users usually join groups more than fake users.
Family\ Relationship*	Social relation Information\Status	Real users share their real social relation status than fake users.

Note: the * indicates the new attributes introduced in this work.

After preparing the dataset, a set of supervised and unsupervised mining algorithms (k-NN, SVM, ID3 Decision Tree, k-Means, and k-medoids) were implemented using RapidMiner Studio v.8.0.1. In supervised mining, the classifier is trained with known class data (*fake* and *real*). However, for unsupervised mining, statistically significant measures were defined for each cluster. Accordingly, profiles with similar attributes were grouped together in the same cluster, while the other profiles were grouped together in a different cluster.

4 Experiment and Results

Before discussing the obtained results, a brief description of the used dataset is given, followed by the performance metrics.

4.1 Dataset Description

A total of 906 profiles were collected for use in the experiments. Of those, 125 profiles were excluded because they were irrelevant or duplicates. This resulted in 781 real profiles, among which 19 profiles were found to be fake, so these were labeled in the dataset as fake. To collect more fake profiles, 250 more profiles were purchased. However, only 182 profiles of those were considered as some profiles were found deactivated or blocked. This resulted in 201 fake profiles. Thus, the collection process resulted in a total of 982 profiles (781 real and 201 fake). Out of this total, 86 (61 real and 25 fake) had values missing from some of their attributes. A k-NN model for data imputation was employed for handling the missing values. Finally, manual labeling was applied to the collected dataset to label profiles as fake or real for training and testing purposes.

4.2 Performance Metrics

A group of common metrics can be applied in the validation process. In this work, the following metrics were used, taken from [19]:

1. Accuracy: Measure the performance of the detection model
   Accuracy = (correct predictions) / (total examples).
2. Recall: true positive rate
   Recall = (true positive predictions) / (positive examples),
3. Precision: Measure the probability that the positive predication is correct
   Precision = (true positive predictions) / (positive predictions).
4. Specificity: true negative rates
   Specificity = (true negative predictions) / (negative examples).

4.3 Experimental Results

4.3.1 Supervised Algorithms Experiment

After handling the missing values using the k-NN estimator, the supervised algorithms were applied to the 982 profiles. Table 4 shows the results for the supervised algorithms.

Table 4 Results of supervised algorithms.

Metrics	Accuracy	Precision	Recall	Specificity
ID3	0.9776	0.9872	0.9846	0.9502
SVM	0.9572	0.9780	0.9680	0.9154 0.9403
K-NN with k = 3	0.9145	0.9520	0.9398	0.8159

4.3.2 Unsupervised Algorithms Experiment

Following the same approach, the unsupervised algorithms were applied using the same dataset. It is important to note that the training data were an unlabeled dataset, which made evaluation problematic because there was nothing to which the model's results could be meaningfully compared. Thus, there was no straightforward way to evaluate the accuracy of the applied algorithm [20]. To evaluate the clustering techniques, we formed an evaluating model using RapidMiner's special operators, which can be exploited in flexible ways. For example, the operator Map Clustering on Labels maps clustering and prediction processes by adjusting the given clusters with class labels. This let us adjust the dataset and evaluate our model. Both algorithms were applied to (k = 2) or 2 clusters (C0, C1), where C0 represents the real profiles, while C1 represents the fake profiles. The k-Means algorithm partitioned the dataset and showed an accuracy of 0.6731, while k-medoids showed an accuracy of 0.6701, as shown in Table 5.

Table 5 Results of unsupervised algorithms.

Algorithm	Actual States		Real	Fake	Clusters
	Predicted States	Real	661	201	C_0
		Fake	120	0	C_1
	Accuracy	Precision	Recall	Specificity	
k-Means	0.6731	0.7668	84.64%	0.0000	
k-medoids	0.6701	0.8840	0.6735	0.6567	

5 Discussion

As shown in the previous section, the supervised algorithms outperformed the unsupervised algorithms. However, before justifying these results, some important points should be mentioned. Firstly, the model depends on the informative attributes to make a decision. These attributes are illustrated in Figure 1. As can be seen, the 'mutual friends' attribute is the most informative, while the 'introduction' attribute is the least informative. Secondly, we note that some attributes had the same values in both real and fake profiles. For example, fake profiles typically have zero tags, zero posts, and high liking activity. Unfortunately, many real profiles have the same values, which misleads the classification techniques.

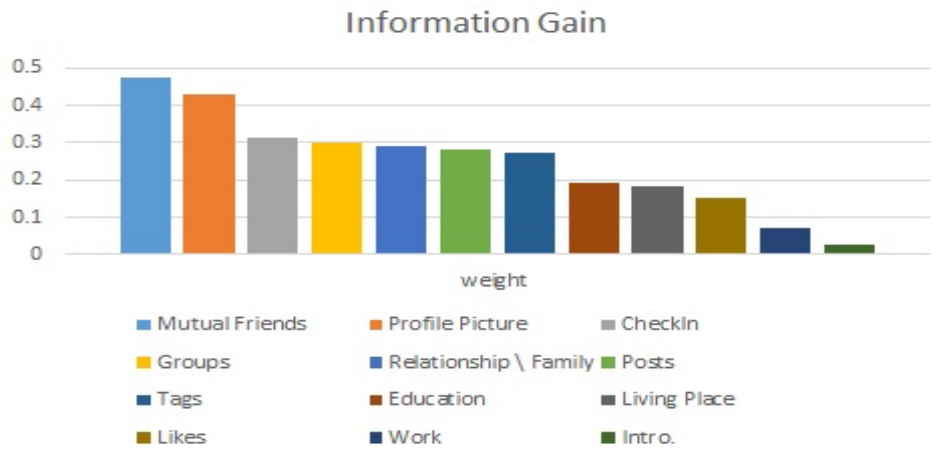

Figure 1 Information gain of the attributes.

Figure 2 (1-5) illustrates the histogram charts for the interfered attributes with respect to the two-class labels (*fake* and *real*). Thus, the algorithm that is capable of handling the interfered attributes correctly will make the most accurate decisions. Accordingly, in the next subsections, we will justify the performance by explaining how each technique resolved the interfered attributes.

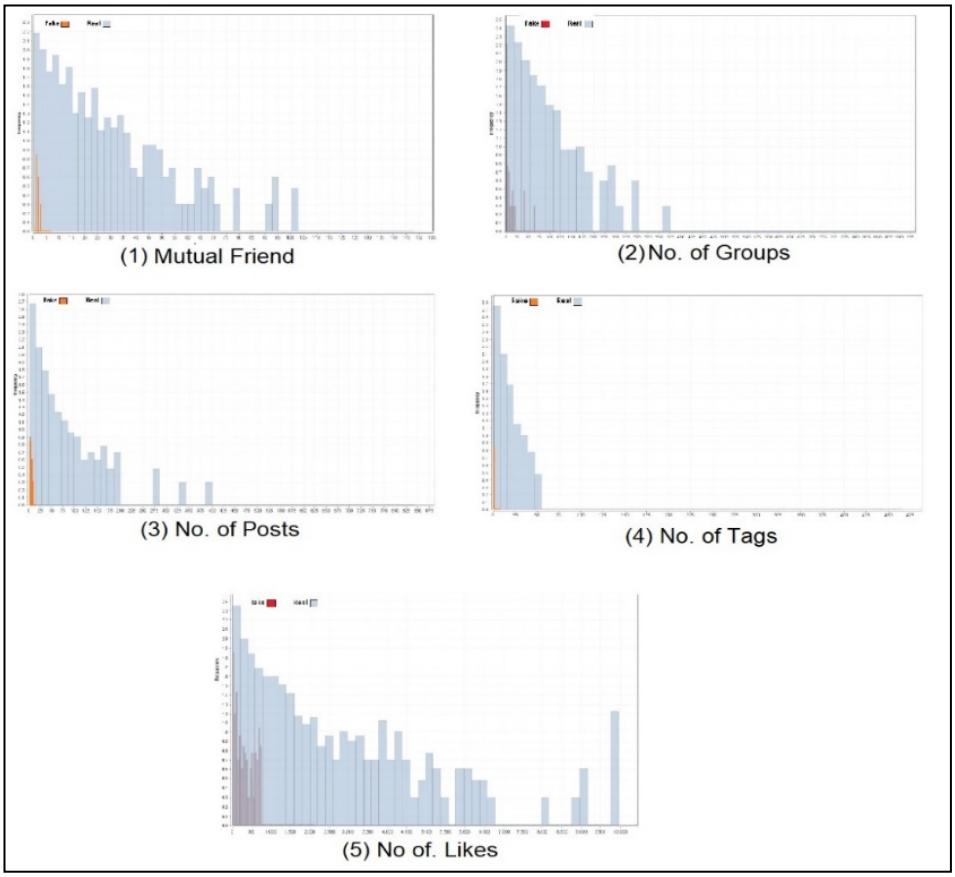

Figure 2 Attribute distribution.

5.1 Why the Supervised Algorithms Outperformed the Unsupervised Algorithms

The supervised algorithms outperformed the unsupervised algorithms because the training set was already labeled in the case of supervised techniques, which makes an essential difference in the detection process. This process minimizes the interference factor in the attributes and gives the model the necessary experience in the detection process. However, the unsupervised techniques deal with all profiles as a single unit without class labels to separate the dataset. This makes interference detection more difficult, as the attributes not only have to be labeled in fake and real profiles but also in profiles of the same class. This distracts the detection model and groups the profiles with similar attributes in clusters, ignoring any profile with interfered attributes.

Moreover, the unsupervised algorithms had low accuracy rates because these clustering techniques handle the dataset as a single unit and group profiles with similar attributes in one cluster. Because of this, a problem with the interfered attributes emerged, where some of the informative attributes were not clustered into different clusters. Thus, these techniques could not correctly cluster profiles into *fake* and *real*.

5.2 Why Did the Supervised Algorithms Have a High Accuracy Rate?

The supervised algorithms have a high accuracy rate because they use the k-NN estimator, which has proved its efficiency in handling missing values. The k-NN estimator inputs data by assigning values from the k-most similar profiles to the missing values. This can be seen from the experiment, where the model exhibited stable performance with nearly identical accuracy for all algorithms. We note that most of the missing values were in the 'groups' and 'likes' attributes, which had the highest interference factor. k-NN excludes these attributes from the calculation process, which positively affects the accuracy rate.

6 Conclusion

This work considered the detection of fake Facebook profiles using data-mining techniques. A model was proposed that utilizes 5 supervised and unsupervised techniques with 12 discriminative (behavioral and non-behavioral) attributes. RapidMiner Studio 8.0.1 was employed to conduct an experiment to evaluate the accuracy of the model based on a dataset with 982 profiles (781 real, and 201 fake). The supervised algorithms outperformed the unsupervised algorithms and showed high and promising accuracy rates in all experiments. More specifically, the ID3 decision tree exhibited the highest accuracy among all algorithms and all unsupervised algorithms showed a relatively similar low accuracy. A deep explanation of these results was given at the end of this paper.

Acknowledgment

The authors are grateful to the Applied Science Private University, Amman-Jordan, for the full financial support granted to cover the publication fee of this research article.

References

[1] Smith, A.N., Fischer, E. & Yongjian, C., *How Does Brand-related User-generated Content Differ Across YouTube, Facebook, and Twitter?*, Journal of Interactive Marketing, **26**(2), pp. 102-113, 2012.

[2] Romero, D.M., Galuba, W., Asur, S. & Bernardo, A., *Influence, and Passivity in Social Media*, in Proceedings of the 20th International Conference Companion on World Wide Web, ACM, pp. 113-114,2011.

[3] Obar, J.A. & Wildman, S.S., Social Media Definition, and The Governance Challenge: An Introduction to the Special Issue, 2015. DOI: 10.1016/j.telpol.2015.07.014.

[4] Kaplan, A.M. & Haenlein, M., *Users of the World, Unite! The Challenges and Opportunities of Social Media*, Business Horizons, **53**(1), pp. 59-68, 2010.

[5] Eugene, A., Castillo, C., Donato, D., Gionis, A. & Mishne, G., *Finding High-Quality Content in Social Media*, In Proceedings of the 2008 International Conference on Web Search and Data Mining, ACM, pp. 183-194, 2008

[6] O'Keeffe, Schurgin, G. & Pearson, K.C., *The Impact of Social Media on Children, Adolescents, and Families*, Pediatrics. **127**(4), pp. 800-804, 2011.

[7] Qian, T., Gu, B. & Whinston, A.B., *Content Contribution for Revenue Sharing and Reputation in Social Media: A Dynamic Structural Model*, Journal of Management Information Systems, **29**(2), pp. 41-76, 2012.

[8] Kontaxis, Georgios, Polakis, I., Ioannidis, S. & Markatos, E.P., *Detecting Social Network Profile Cloning*, In Pervasive Computing and Communications Workshops (PERCOM Workshops), 2011 IEEE International Conference on, pp. 295-300. IEEE, 2011.

[9] Wani, M.A, Jabin, S. & Ahmad, N., *A Sneak into the Devil's Colony-Fake Profiles in Online Social Networks*, arXiv preprint arXiv:1705.09929 ,2017.

[10] RapidMiner. https://rapidminer.com/ (3rd August 2019).

[11] Kumar, N. & Reddy, R.N., *Automatic Detection of Fake Profiles in Online Social Networks*." Ph.D. diss., National Institute of Technology Rourkela, 2012.

[12] Gupta, A. & Kaushal, R., *Towards Detecting Fake User Accounts in Facebook*, In Asia Security and Privacy (ISEASP), 2017 ISEA, pp. 1-6. IEEE, 2017.

[13] Ahmed, F. & Abulaish, M., *A Generic Statistical Approach for Spam Detection in Online Social Networks*, Computer Communications, **36**(10), pp. 1120-1129, 2013.

[14] Fire, M., Kagan, D., Elyashar & Elovici, Y, *Friend or Foe? Fake Profile Identification in Online Social Networks*, Social Network Analysis and Mining, **4**(1), pp. 194-210, 2014.

[15] Xiaoyun, W., Lai, C.M., Hong, Y., Hsieh, C.J. & Wu, S.F., *Multiple Accounts Detection on Facebook Using Semi-Supervised Learning on Graphs*, arXiv preprint arXiv:1801.09838, 2018.

[16] Bimal, V., Bashir, M.A., Crovella, M., Guha, S., Gummadi, K.P., Krishnamurthy, B. & Mislove, A., *Towards Detecting Anomalous User Behavior in Online Social Networks*, In USENIX Security Symposium, pp. 223-238, 2014.

[17] Shalinda, A. & Dutta, K., *Identifying Fake Profiles in LinkedIn*, In PACIS, pp. 278. 2014.

[18] Nazir. A., Raza, S., Chuah, C.N., Schipper, B. & Davis, C.A., *Ghostbusting Facebook: Detecting and Characterizing Phantom Profiles in Online Social Gaming Applications*, In WOSN, 2010.

[19] Yousuf, B.S. & Abulaish, M., *Community-Based Features for Identifying Spammers in Online Social Networks*, In Advances in Social Networks Analysis and Mining (ASONAM), 2013 IEEE/ACM International Conference on, pp. 100-107. IEEE, 2013.

[20] Jiawei, H., Pei, J. & Kamber, M., *Data Mining: Concepts and Techniques*, Elsevier, 2011.

[21] Karel, H., Templ, M. & Filzmoser, P., Imputation of Missing *Values for Compositional Data Using Classical and Robust Methods*, Computational Statistics & Data Analysis, **54**(12), pp. 3095-3107, 2010.