TSDL: A framework for Tip Spam Detection in Location Based Social Network

Adewole K. S
Department of Computer Science, University of Ilorin, Nigeria
adewole.ks@unilorin.edu.ng

Jimoh R. G
Department of Computer Science, University of Ilorin, Nigeria
jimoh_rasheed@yahoo.com

Isiaka R. M.
Department of Computer, Library and Information Science, Kwara State University, Malete, Nigeria
abdulrafiu.isiaka@kwasu.edu.ng

Ajiboye A. R.
Department of Computer Science, University of Ilorin, Nigeria
ajibabdulraheem@gmail.com

Abstract—In Web 2.0 systems, Location Based Social Network (LBSN) has become increasingly popular because of its ability for locating places, such as restaurants, hotels and nearby facilities that can render services to the users. Among such LBSN is Apontador, a popular social network introduced in Brazil a couple of years ago. However, despite the numerous opportunities offered by Apontador LBSN, it has become attractive social network for tips spam distribution. In this work, an intelligent machine learning framework is proposed to detect tips spam in Apontador LBSN. The proposed framework explored three classification algorithms: Random Forest, Multilayer Perceptron, and Decorate. Bio-inspired feature identification was studied using Evolutionary algorithm (EA) and Particle Swamp Optimization (PSO) to identify the discriminating features for tip spam detection in LBSN. Three experiments were conducted based on sixty (60), twenty-two (22), and ten (10) features to ascertain the performance of the proposed framework for tip spam detection. Based on the various experiments conducted, Random Forest classification algorithm produced accuracy and F-measure of 90.5% and ROC of 96.7% using 60 features extracted from Apontador LBSN. The algorithm also outperformed the two other classifiers during EA evaluation. However, Decorate classifier produced the best results during the PSO evaluation, achieving F-measure and ROC of 86.3% and 92.9% respectively. The experimental results show that the proposed framework improved in performance for tips spam detection in Apontador LBSN.

Keywords— Location based social network; tip spam; evolutionary computation; machine learning.

I. INTRODUCTION

Location Based Social Network (LBSN), such as Apontador and Yelp, are now becoming attractive social media for the new set of users to obtain information about new locations. LBSN enables users to share information, which may include hotel nearby, restaurants and a host of other places about a specific location to the new users using smartphone with embedded GPS capability. LBSN is a type of online social network (OSN) based on the Web 2.0 technology that is increasingly attracting attention from visitors who are unaware of the layout of the new locations. For instance, the popular Brazil based LBSN, Apontador, includes features that allow users to locate different places for relaxation, add new locations, check-in in locations, and read users tips about a location, such as interesting things to buy or products to avoid when visiting a specific location as recommended by the users on the platform (Costa et al., 2014). This tip allows new users to explore the area and gain more insight on the geography of the new locations. Despite the benefits offer by LBSN to the users of the system they have become attractive platforms for new form of spam message distribution, which propagates wrong tips about a location. For instance, tip spam in LBSN may appear inform of opinion spam, which targets to mislead readers by posting positive reviews about a specific area in order to promote the location. At the same time, spammer can post negative review to damage the reputation of the area or link users to other external web pages that are completely unrelated to the users need (Guo & Chen, 2014; Molavi Kakhki et al., 2013). This misleading tip can expose LBSN to different attacks from spammers and endanger the trust relationship existing on the platform. Thus, compromising the efforts of promoting LBSN and indeed waste readers' time in filtering spam tips from legitimate information about places (Costa et al., 2013). While existing studies on spam detection in OSN focus more on social networks like Twitter (Cresci et al., 2015) and Facebook (Ahmed & Abulaish, 2012), little efforts have been committed to detecting tip spam in LBSN. For example, Costa et al. (2013) presented machine learning technique to identify tip spam in LBSN using Apontador as test bed. They manually annotate 1,260 spam tips and obtain 1,260 non-spam tips from Apontador. These labeled collections were used to train and test machine learning

classifiers. An extended version of this research was presented in (Costa et al., 2014). The authors classified tip spammers in Apontador LBSN as Pollution, bad-mouthing, and local marketing. However, much effort is needed to investigate the performance of the models presented in these studies. In this paper, a more robust framework is presented using bio-inspired evolutionary computation approach. The proposed framework explored 60 features for tip spam detection and applied bio-inspired evolutionary search method to identify the most discriminating attributes that can produce compact model for tip spam detection in LBSN. Based on this objective, three classification algorithms are investigated to produce suitable model for the proposed framework. The remaining part of this paper is organized as follows: Section 2 presents related works in social spam detection. Section 3 highlights the proposed framework presented in this paper. Section 4 discusses the evaluation metrics considered in this study to determine the performance of the proposed framework. Section 5 focuses on the experimental results and Section 6 concludes the paper and presents future directions.

II. RELATED WORK

Spam represents all forms of unsolicited posts that appear in personal feeds, comments, trending topics, and users' walls. Spammers distribute spam messages with the goal of overpowering legitimate users on the network. The aim of spammers may be to hijack existing legitimate accounts or to steal the identities of users on the network in order to create fake accounts that mimic the real one. For instance, Bilge et al. (2009) investigated the use of cloned accounts to launch different attacks on social networks. They examined whether the use of CAPTCHA can reduce identity clone on social networking sites. However, their findings revealed that CAPTCHA approach is vulnerable to manipulation using complex algorithms that can analyze and decode CAPTCHA contents. Grier et al. (2010) applied blacklist based method to identify spammers on Twitter. The researchers analyzed the clickstream data from URLs click patterns on Twitter. Aside the use of CAPTCHA and blacklist based methods; researchers have addressed social spam detection using graph and machine learning methods. In Graph-based method, social network is modeled as a network consisting of nodes (users) and edges (relationships). The relationships between nodes are analyzed in order to detect users with unusual characteristics. This method has proved suitable for separating spam accounts from non-spam. For example, Ahmed and Abulaish (2012) applied Markov clustering (MCL) algorithm to group set of profiles as spam and non-spam on Facebook. The MCL algorithm takes a weighted graph as input and uses random walk approach to assign probabilities to each node on the network. Based on these probabilities, the algorithm is able to cluster set of accounts as spam or non-spam. Zhang and Lu (2016) proposed machine learning and social network analysis methods for discovering millions of fake followers in Weibo social network. The authors identified near-duplicate accounts used for spam contents distribution in Weibo. Ruan et al. (2016) identified compromised accounts that are utilized for spamming activities using social behavioral analysis. Mulamba et al. (2016) proposed SybilRader algorithm based on social network analysis to identify fake accounts on social networks. Echeverría and Zhou (2017) introduced machine learning method to identify Star wars botnets on Twitter social network. Atluri and Tran (2017) provided a comprehensive analysis of botnets threat, infections, and their detection approaches. The use of machine learning method has played a major role in detecting spammers on social network. For instance, in the domain of supervised machine learning, Chu et al. (2012b) proposed Random Forest algorithm to identify spam campaign on Twitter. They combined tweet, account property, link, and campaign features to train machine learning algorithm. Costa et al. (2013) presented machine learning technique to identify tip spam in LBSN. Martinez-Romo and Araujo (2013) explored the use of language and content based features to detect spam tweets using SVM algorithm. Benevenuto et al. (2010) also applied SVM algorithm to detect spammers on Twitter.

In unsupervised machine learning domains, Egele et al. (2015) applied statistical model and anomaly detection technique to identify compromised accounts on social networks. Their approach extracts different features from messages, such as time of the day, message source, message text, message topic, message link, and direct user interaction. They developed user behavioral model, which distinguished known behavior from its slight deviation. Lee and Kim (2014) proposed hierarchical clustering approach to initially group spammers with malicious account names. They trained Markov chain algorithm with valid account names from Twitter. Therefore, any account name that violates this pattern is categorized as malicious. Chen et al. (2014) analyzed domain and social behavioral features for malicious URL detection on Facebook. Their results revealed that Bayesian classifier achieved good performance accuracy. Bhat et al. (2014) trained ensemble machine learning algorithms to detect spam accounts on Facebook.

However, as opposed the aforementioned studies, this paper investigates the different features for tips spam detection in LBSN. A framework for tip spam detection is proposed and bio-inspired feature investigation is explored to produce compact model for tip spam detection in Apontador LBSN.

III. METHODOLOGY

Machine learning (ML) is a field that evolved from the study of pattern recognition and computational learning theory in artificial intelligence. ML enables computers to learn without being explicitly programmed. In this paper, an intelligent ML framework shown in

Fig. 1 is proposed. The framework consists of four (4) main stages: data collection, feature analysis, bio-inspired computation, and classification. The data collection stage involves gathering of location based social network data from Apontador online social network. The goal is to distinguish spam tips from legitimate tips posted on the network. The features extracted from the data are analyzed in order to identify the users' behavioral patterns along their posting patterns. The bio-inspired evolutionary search stage considers two evolutionary algorithms to determine the discriminative ability of the various features considered in this study. In addition, the classification stage applied three classification algorithms to detect spam tips in LBSN.

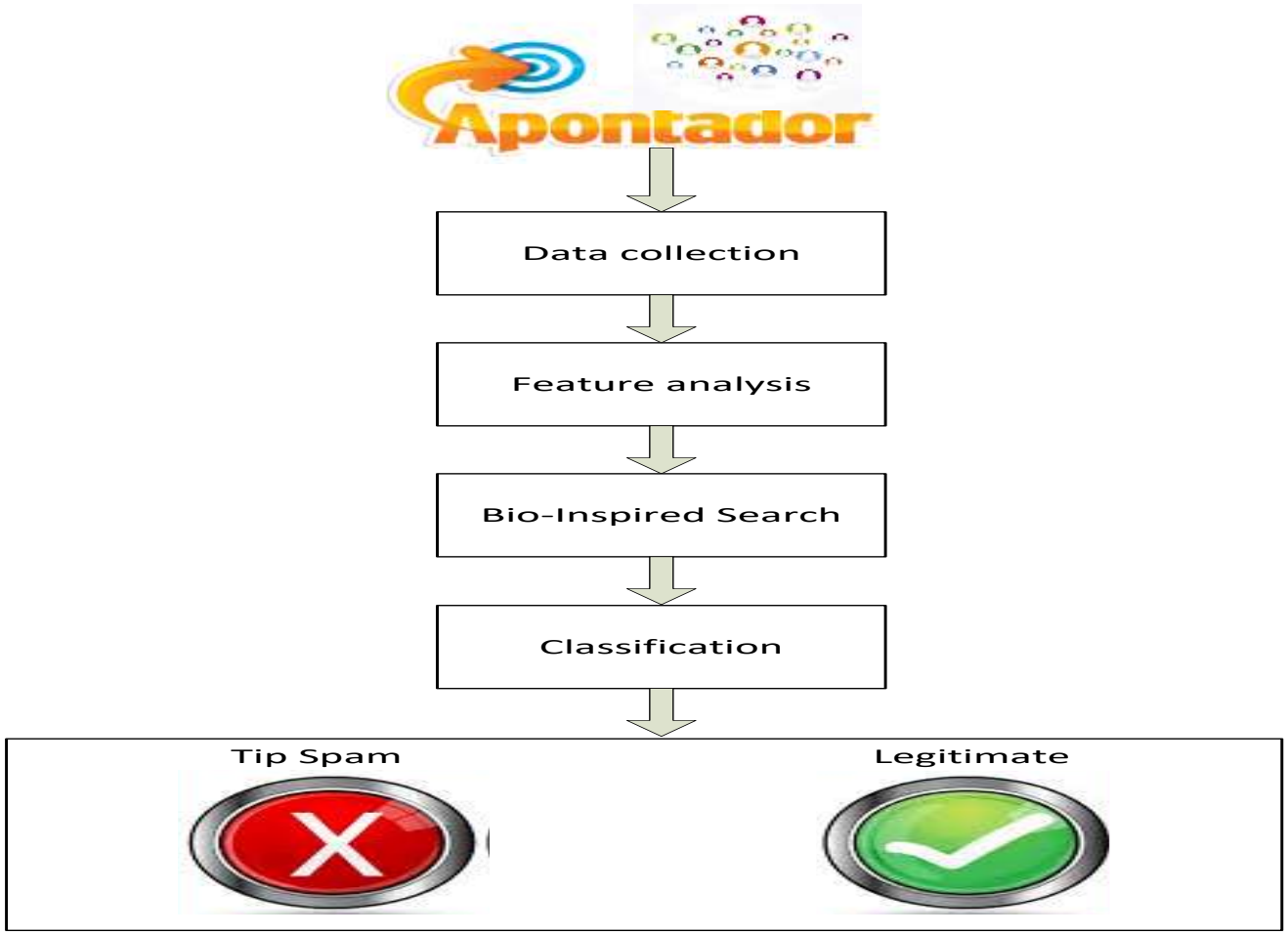


Fig. 1. Proposed framework for tip spam detection

A. Data Collection

Spam in LBSN usually appears in the form of tips with the aim of spreading fake information or advertising fake products. The dataset used in this study was collected by Costa et al. (2014) from Apontador location based social network. This dataset contains tips about places manually labeled as spam or non-spam by three Apontador's moderators during the month of September, 2011. The assessors identified 3668 tips as spam while Apontador provided the non-spam tips during the same period. The detail discussion on the dataset can be found in (Costa et al., 2014).

Table 1 shows the statistics of the dataset used in this study.

Table 1: Apontador dataset collected by Costa et al. (2014).

Description of items	Number of items
Number of spam tips	3538
Number of non-spam tips	3538
Number of features	60
Total labeled samples	7076

The 60 features are divided into four features categories containing 32 content attributes, 11 user attributes, 5 place attributes, and 12 social attributes as discussed in the subsequent section.

B. Analysis of features

As discussed in the previous section, the dataset collected contains 60 features. The features, also known as attributes, enable separation of spam behaviors from legitimate ones. This section highlights the various features employed for tip spam detection, which are divided into four main categories.

Contents features

Contents features are those attributes extracted from the text tips posted by the users as shown in Table 2. The detail discussion on each of the feature categories can be found in (Costa et al., 2014).

Table 2: Contents features

Feature description
Number of email addresses on the text
Number of contact information on the text
Number of URLs on the text
Number of phone numbers on the text
Number of numeric characters
SentiStrength
Combined-method
Number of words
Number of distinct 1-gram
Number of capital letters
SentiWordNet
SenticNet
Similarity score (avg, median, max, min, sd)
Happiness index
SASA
Fraction of 1-gram
Number of spam words and spam rules
Value of "Has offensive word"
Number of words in capital
Number of offensive words on the text
Clicks on the link "This tip helped me"
PANAS-t
Clicks on the link "Report abuse"
Emoticons
Number of distinct 3-gram
Fraction of 3-gram
Number of distinct 2-gram
Fraction of 2-gram

User based features

These attributes are specific to particular user behavior. They capture variation in tips posting patterns in order to separate anomalous from legitimate posting behavior. Eleven (11) user based features utilized in this study are shown in Table 3.

Table 3: User based features

Feature description
Number of tips posted by the user
Number of photos posted by the user
Number of places registered by the user
Distance among all places reviewed by the user (avg, median, max, min, sd)
Tip entropy of a user
Number of different areas where the user posted a tip
Tip focus of a user

Place features

These attributes relate to places where the tips are posted. It is expected that the behavior of spammer will deviate from legitimate users as it relates to their posting location. Five (5) features considered in this study are highlighted in Table 4.

Table 4: Place features

Feature description
Number of tips on the place
Place rating
Clicks on the link "Thumbs up"
Clicks on the link "Thumbs down"
Number of clicks on the place page

Social features

The social features are those attributes, which relate to the interaction or relationship among users on the network. The connection among users is modeled as a graph where the nodes represent users on the network and the links between nodes are the relationship established by the users. It is expected that the network connection of legitimate users will differ from spammers, thus, twelve (12) features are investigated to separate spam tips from legitimate tips. Table 5 shows the social features considered in this study.

Table 5: Social features

Feature description
Fraction of followers per followees
Number of followers (in-degree)
Assortativity (in–in, in–out, out–in, out–out)
Degree
Clustering coefficient
Number of followees (out-degree)
Pagerank
Reciprocity
Betweenness

C. Bio-Inspired Search

This paper considers two bio-inspired algorithms to investigate the discriminative power of the 60 features. Evolutionary algorithm (EA) and Particle Swarm Optimization (PSO) algorithm were utilized, which are all based on population estimation. The parameters used for the EA and Particle Swarm Optimization (PSO) algorithms are highlighted in Figure 2 and Figure 3.

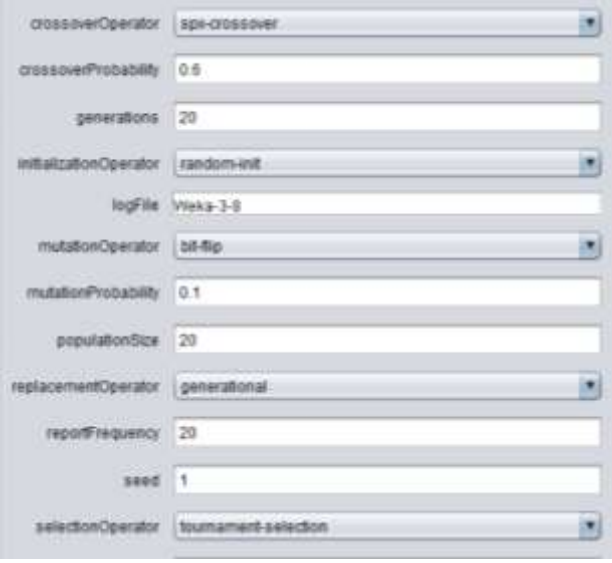


Figure 2: Evolutionary algorithm (EA) parameters

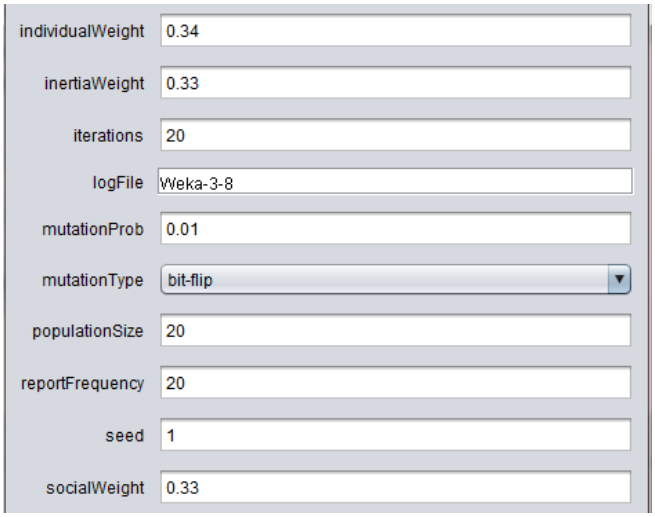


Figure 3: Particle Swarm Optimization (PSO) parameters

EA utilizes biologically inspired evolutionary search methods, such as recombination, mutation, fitness, and selection. The algorithm stochastically explores the search space to prevent local maxima (Manurung, 2004). PSO is a population based stochastic optimization algorithm inspired by social behavior of bird flocking or fish schooling. PSO shares many similarities with evolutionary computation techniques such as Genetic Algorithms (GA) and EA. The algorithm is initialized with a population of random solutions and searches for optima by updating generations. However, unlike EA, PSO has no evolution operators such as crossover and mutation. In PSO, the potential solutions, called particles, fly through the problem space by following the current optimum particles. In this paper, EA and PSO algorithms are both applied to search for the discriminative features for detecting tip spam.

D. Classification

The classification stage applied classification algorithms to detect tip spam and non-tip spam in Apontador LBSN. At this stage, three classifiers were considered, which include Random Forest, Multilayer perceptron (MLP), and Decorate.

Random Forest

Random Forest is a class of decision tree algorithms based on ensemble approach. Decision tree algorithms classify instances using a tree structure. In this tree, a node represents the test of an attribute value and a branch denotes the result of the test. Random Forest decision creates an ensemble of classifiers by constructing different decision trees using random feature selection and bagging approach at training stage. The decision tree produces two types of nodes: the leaf node labeled as a class and the interior node associated with a feature. A different subset of training data is selected with a replacement to train each tree (Chu et al., 2012a; Narudin et al., 2014). Entropy is applied to compute the information gain contributed by each feature. Let D represents the dataset with the labeled instances and C as the class such that $C = \{C_1, C_2, C_3, ..., C_j\}$, where j is the number of classes considered. In this paper, j is set to 2, which represent spam or non-spam tips. Thus, the information needed to identify the class of an instance in the dataset D is denoted as *Info(D) = Entropy(P)*, where P is the class probability distribution such that:

$$P = \left\{ \frac{|C1|}{|D|}, \frac{|C2|}{|D|}, \frac{|C3|}{|D|}, ..., \frac{|Cj|}{|D|} \right\} \quad (1)$$

By partitioning D based on the value of a feature F according to subsets $\{D_1, D_2, D_3, ..., D_n\}$, *Info(F,D)* according to F is computed as:

$$Info(F, D) = \sum_{i=1}^{n} \frac{|D_i|}{|D|} Info(D_i) \quad (2)$$

The corresponding information gain after obtaining the value of F is computed as:

$$Gain(F,D) = Info(D) - Info(F,D) \quad (3)$$

Then the *GainRatio* is defined as:

$$GainRatio(F, D) = \frac{Gain(F, D)}{SplitInfo(F, D)} \quad (4)$$

where *SplitInfo(F,D)* represents the information due to the splitting of D according to the feature F. Random Forest uses

the majority voting of all the individual decision to obtain the final decision (Chu et al., 2012a).

Multilayer perceptron (MLP)

Multilayer perceptron (MLP) is a class of feedforward artificial neural networks that consists of activation units, usually referred to as artificial neurons and weights (Noriega, 2005). MLP modifies the standard linear perceptron by including multiple layers, such as input, hidden, and output layers to solve both linear and non-linear classification problems. The algorithm maps input data to appropriate outputs. During the training stage, MLP applies a learning algorithm, mostly backpropagation, to adjust the weights so that the network can acquire the required knowledge to classify new unseen data. The process of finding the correct values of weights is known as the learning rule, and this involves initializing the weight matrix to a set of random numbers between −1 and +1. As the network learns, these values are changed until it has been shown that the network has solved the problem under consideration. This is done to minimize the error generated by the output unit when compared with the expected result. MLP algorithm has been used to solve a number of problems including detection of a fake account on the LinkedIn social network (Adikari & Dutta, 2014).

Decorate

Decorate is a meta-learner for building diverse ensembles of classifiers by using specially constructed artificial training instances. The artificial training data is constructed by randomly generating examples using an approximation of the training data distribution, such as a Gaussian distribution determined by estimating the mean and standard deviation of the training set for numeric features. Thus, an ensemble is generated iteratively, learning one class at each iteration and adding it to the current ensemble. The ensemble is initialized with the classifier trained on the given data. At each iteration, a specified number of artificial training examples are generated based on a simple model of the data distribution (Melville & Mooney, 2003; Roli et al., 2004).

IV. EVALUATION METRICS

To measure the performance of the proposed framework discussed earlier, a number of metrics were used based on the numbers of correct and incorrect predictions recorded during the experiments. These are utilized to determine the accuracy of the system as well as the error committed during predictions. A table known as confusion matrix in

Table 6 is employed to measure the True Positive (TP) and True Negative (TN) referred to correctly classified spam and non-spam tips respectively. False Positive (FP) represents non-spam tips that are classified as spam, while False Negative (FN) represents spam tips classified as non-spam.

Table 6. Confusion matrix

		Predicted	
		Class = Spam	Class = Non-spam
Actual	Class = Spam	TP	FN
	Class = Non-spam	FP	TN

The parameters TP, TN, FP, and FN in the table was used to derive some standard metrics, such as True Positive Rate (TPR), True Negative Rate (TNR), False Positive Rate (FPR), and False Negative Rate (FNR) as shown in Eqns. 5, 6, 7, and 8 respectively. TPR is also called sensitivity or recall, and can be used to indicate the accuracy of a classification model on the labeled samples. A combined metric known as F-measure or F1-score has been widely used to measure the performance of a classification system. This metric is calculated as the harmonic mean of precision and recall as shown in Eqn. (10). The highlighted metrics were considered during the results evaluation.

$$TPR / Sensitivity / \mathrm{Re}call(R) = \frac{TP}{TP + FN} \quad (5)$$

$$TNR / Specificity = \frac{TN}{TN + FP} \quad (6)$$

$$FPR = \frac{FP}{TN + FP} = 1 - \frac{TN}{TN + FP} \quad (7)$$

$$FNR = \frac{FN}{TP + FN} \quad (8)$$

$$\mathrm{Pr}ecision(P) = \frac{TP}{TP + FP} \quad (9)$$

$$F - measure = \frac{2PR}{(P + R)} \quad (10)$$

V. RESULTS AND DISCUSSION

In this section, different experiments were conducted to ascertain the performance of the proposed framework for tip spam detection in LBSN. An experiment based on all the 60 features was first conducted followed by the bio-inspired search experiments. As stated earlier, three classification algorithms were tested to identify the best classifier for detecting tip spam. In each of the experiments, the TRP, FPR, Precision, Recall, F-measure, and Receiver-operating characteristics (ROC) were recorded for performance evaluation of the proposed framework. 10-fold cross-validation technique was employed, which evaluates predictive model by partitioning the original data samples into a training and testing set. In 10-fold cross-validation, the original sample is randomly partitioned into 10 equal size. Of the 10 subsamples, a single subsample is retained as the validation data for testing the model, and the remaining 9

subsamples are used as training data. The cross-validation process is then repeated 10 times, with each of the subsamples used exactly once as the validation data. The 10-fold results are averaged to produce a single result.

A. Results based on all features

Considering the 60 features earlier discussed, Random Forest classifier gives the best performance out of the three classifiers considered. Random Forest produced accuracy and F-measure of 90.5% and ROC of 96.7% as shown in Table 7 and Figure 4. This result is followed by Decorate classifier with accuracy and F-measure of 89.2% and ROC of 95.1%.

Table 7: Result based on 60 features

Classifiers	10-fold cross-validation					
	TPR	FPR	Precision	Recall	F1	ROC
Random Forest	0.905	0.095	0.907	0.905	0.905	0.967
MLP	0.853	0.147	0.854	0.853	0.853	0.929
Decorate	0.892	0.108	0.893	0.892	0.892	0.951

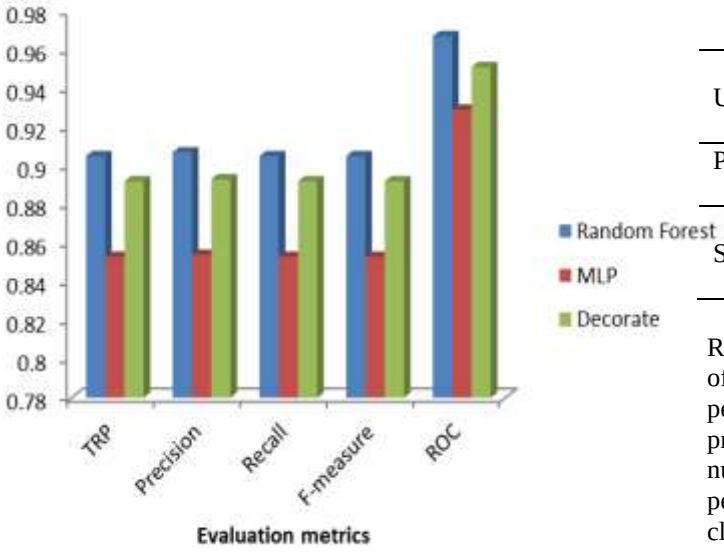


Figure 4: Results based on 60 features

The implication of these results is that with a confidence of over 90%, the proposed framework can effectively identify tip spam in LBSN. The performance of the proposed framework in terms of ROC also improved significantly producing ROC area above 95%.

B. Bio-inspired feature search

In this section, the results based on bio-inspired feature search are presented using both EA and PSO algorithms.

Results using EA algorithm

The EA algorithm produces 22 discriminating features as shown in Table 8. Based on these features, the three classifiers were examined to ascertain the performance of the proposed framework for tip spam detection in LBSN. The results according to Figure 5 shows that Random Forest classifier still outperforms the other two classification algorithms for tip spam detection. In this result, the bio-inspired evolutionary computation method is investigated to produce compact model for tip spam detection in LBSN.

Table 8: Discriminating features based on EA algorithm

Feature category	Feature description
Content based	Number of email addresses on the text Number of contact information on the text Number of URLs on the text Number of phone numbers on the text Number of numeric characters Combined-method SentiWordNet SenticNet Similarity score (median, sd) SASA Number of spam words and spam rules Number of words in capital Clicks on the link "This tip helped me"
User based	Number of photos posted by the user Number of places registered by the user Tip focus of a user
Places	Number of tips on the place Place rating
Social	Fraction of followers per followees Pagerank Reciprocity

Random produces accuracy and F-measure of 88.9% and ROC of 95.6. By exploring 22 features, the reduced model achieved performance slightly close to 60 features. This shows that the proposed framework in this paper can identify tip spam with a number of minimal features without much compromise to the performance accuracy. This in turn will generate tip classification model with minimal execution time.

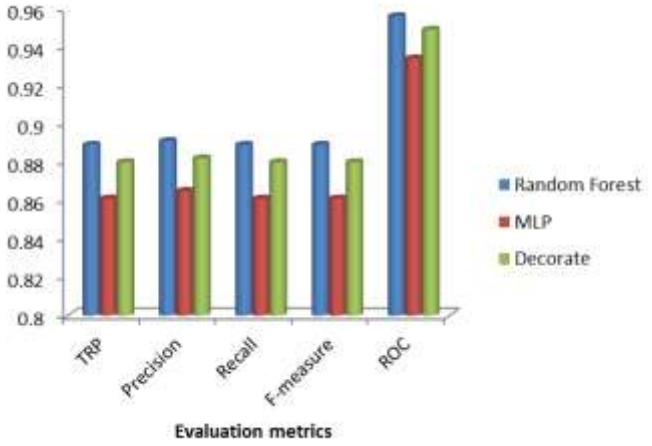


Figure 5: Results based on EA reduced features

Results using PSO algorithm

Conversely, PSO algorithm produces 10 discriminating features as shown in Table 9. It was noticed that EA algorithm produces improved results over the PSO. During the PSO evaluation, Decorate ensemble classifier produced the best result with accuracy and F-measure of 86.3% and ROC of 92.9% as shown in Figure 6.

Table 9: Discriminating features based on PSO algorithm

Feature category	Feature description
Content based	Number of email addresses on the text Number of contact information on the text Number of URLs on the text Number of phone numbers on the text Emoticons SenticNet Clicks on the link "This tip helped me"
Places	Number of tips on the place Place rating
Social	Pagerank

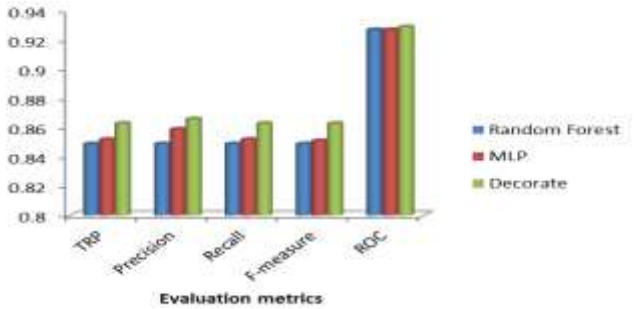


Figure 6: Results based on PSO reduced features

VI. CONCLUSION

LBSN provided opportunity for users to share different locations using GPS technology on smartphones. The availability of LBSN has attracted diverse attentions from social network users who are interested in searching for different places. Existing LBSN users can leverage the functionalities of the system to get in touch with different places and share new information about places using GPS. However, in the recent time, attackers have targeted LBSN as a platform to perpetrate malicious activities, such as tip spam distribution. In this paper, a framework for tip spam detection in LBSN is proposed, which is based on intelligent machine learning. The performances of three classification algorithms, Random Forest, MLP, and Decorate, were investigated for tip spam detection using 60 features classified into four major categories: contents, user, places, and social relationship features. In addition, two bio-inspired evolutionary algorithms, EA and PSO, were applied to search for the most discriminating features that can be employed for tip spam detection in LBSN. The results obtained from the various experiments show that the proposed framework is promising for tip spam detection in LBSN. Random Forest algorithm has demonstrated better performance based on the experiment with all the features in the dataset, producing accuracy and F-measure of 90.5% and ROC of 96.7%. Similarly, Random Forest achieved the best result with EA features evaluation. However, Decorate classifier outperformed other classification algorithms during the PSO experiment. The various results obtained from the different analysis, confirmed the suitability of the proposed framework for tip spam detection in LBSN.

In the future, more analysis is needed to further improve the detection performance of the proposed framework. In addition, the aim is to evaluate the framework using other location based social networks to ascertain the generalization ability of the proposed method.

REFERENCES

[1] Adikari, S., & Dutta, K. (2014). *Identifying fake profiles in LinkedIn.* In: PACIS.

[2] Ahmed, F., & Abulaish, M. (2012). *An MCL-based approach for spam profile detection in online social networks.* In: 2012 IEEE 11th International Conference on Trust, Security and Privacy in Computing and Communications (TrustCom).

[3] Atluri, A. C., & Tran, V. (2017). Botnets Threat Analysis and Detection *Information Security Practices* (pp. 7-28): Springer.

[4] Benevenuto, F., Magno, G., Rodrigues, T., & Almeida, V. (2010). *Detecting spammers on Twitter.* In: 7th Annual Collaboration, Electronic Messaging, Anti-Abuse and Spam Conference, CEAS 2010.

[5] Bhat, S. Y., Abulaish, M., & Mirza, A. A. (2014). *Spammer classification using ensemble methods over structural social network features.* In: Proceedings of the 2014 IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT)-Volume 02.

[6] Bilge, L., Strufe, T., Balzarotti, D., & Kirda, E. (2009). *All your contacts are belong to us: automated identity theft attacks on*

social networks. In: Proceedings of the 18th international conference on World wide web. ACM.

[7] Chen, C.-M., Guan, D., & Su, Q.-K. (2014). Feature set identification for detecting suspicious URLs using Bayesian classification in social networks. *Information Sciences, 289*, 133-147.

[8] Chu, Z., Gianvecchio, S., Wang, H., & Jajodia, S. (2012a). Detecting automation of twitter accounts: Are you a human, bot, or cyborg? *IEEE Transactions on Dependable and Secure Computing, 9*(6), 811-824. doi:10.1109/TDSC.2012.75

[9] Chu, Z., Wang, H., & Widjaja, I. (2012b). *Detecting social spam campaigns on Twitter* (Vol. 7341 LNCS).

[10] Costa, H., Benevenuto, F., & Merschmann, L. H. (2013). *Detecting tip spam in location-based social networks.* In: Proceedings of the 28th Annual ACM Symposium on Applied Computing.

[11] Costa, H., Merschmann, L. H., Barth, F., & Benevenuto, F. (2014). Pollution, bad-mouthing, and local marketing: The underground of location-based social networks. *Information Sciences, 279*, 123-137.

[12] Cresci, S., Di Pietro, R., Petrocchi, M., Spognardi, A., & Tesconi, M. (2015). Fame for sale: Efficient detection of fake Twitter followers. *Decision Support Systems, 80*, 56-71. doi: 10.1016/j.dss.2015.09.003

[13] Echeverría, J., & Zhou, S. (2017). TheStar Wars' botnet with> 350k Twitter bots. *arXiv preprint arXiv:1701.02405*.

[14] Egele, M., Stringhini, G., Kruegel, C., & Vigna, G. (2015). Towards Detecting Compromised Accounts on Social Networks. *IEEE Transaction on Dependable and Secure Computing*. doi: 10.1109/TDSC.2015.2479616

[15] Grier, C., Thomas, K., Paxson, V., & Zhang, M. (2010). @spam: the underground on 140 characters or less. *Proceedings of the 17th ACM conference on Computer and communications security*, 27-37.

[16] Guo, D., & Chen, C. (2014). Detecting Non-personal and Spam Users on Geo-tagged Twitter Network. *Transactions in GIS, 18*(3), 370-384. doi: 10.1111/tgis.12101

[17] Lee, S., & Kim, J. (2014). Early filtering of ephemeral malicious accounts on Twitter. *Computer Communications, 54*, 48-57.

[18] Manurung, H. M. (2004). *An evolutionary algorithm approach to poetry generation.* (Doctor of Philosophy PhD), University of Edinburgh. Retrieved from https://www.era.lib.ed.ac.uk/handle/1842/314

[19] Martinez-Romo, J., & Araujo, L. (2013). Detecting malicious tweets in trending topics using a statistical analysis of language. *Expert Systems with Applications, 40*, 2992-3000. doi: 10.1016/j.eswa.2012.12.015

[20] Melville, P., & Mooney, R. J. (2003). *Constructing diverse classifier ensembles using artificial training examples.* In: IJCAI.

[21] Molavi Kakhki, A., Kliman-Silver, C., & Mislove, A. (2013). *Iolaus: Securing online content rating systems.* In: Proceedings of the 22nd international conference on World Wide Web.

[22] Mulamba, D., Ray, I., & Ray, I. (2016). *SybilRadar: A Graph-Structure Based Framework for Sybil Detection in On-line Social Networks.* In: IFIP International Information Security and Privacy Conference.

[23] Narudin, F. A., Feizollah, A., Anuar, N. B., & Gani, A. (2014). Evaluation of machine learning classifiers for mobile malware detection. *Soft Computing*, 1-15.

[24] Noriega, L. (2005). Multilayer perceptron tutorial. *School of Computing. Staffordshire University*.

[25] Roli, F., Kittler, J., & Windeatt, T. (2004). *Multiple Classifier Systems: 5th International Workshop, MCS 2004, Cagliari, Italy, June 9-11, 2004, Proceedings* (Vol. 5): Springer Science & Business Media.

[26] Ruan, X., Wu, Z., Wang, H., & Jajodia, S. (2016). Profiling Online Social Behaviors for Compromised Account Detection. *IEEE TRANSACTIONS ON INFORMATION FORENSICS AND SECURITY, 11*(1), 176-187. doi: 10.1109/tifs.2015.2482465

[27] Zhang, Y., & Lu, J. (2016). Discover millions of fake followers in Weibo. *Social Network Analysis and Mining, 6*(1), 1-15.