Received: October 13, 2020. Revised: November 19, 2020.

International Journal of
Intelligent Engineering & Systems

http://www.inass.org/

Ontology Meter for Twitter Fake Accounts Detection

Mohammed Hussein Jabardi[1,2]* **Asaad Sabah Hadi[1]**

[1]*College of Information Technology, the University of Babylon, Babylon, Iraq*
[2]*Information Technology Research and Development Center (ITRDC), the University of Kufa, Najaf, Iraq*
* Corresponding author's Email: mohammed.jabardi@student.uobabylon.edu.iq

Abstract: One of the most popular social media platforms, Twitter is used by millions of people to share information, broadcast tweets, and follow other users. Twitter is an open application programming interface and thus vulnerable to attack from fake accounts, which are primarily created for advertisement and marketing, defamation of an individual, consumer data acquisition, increase fake blog or website traffic, share disinformation, online fraud, and control. Fake accounts are harmful to both users and service providers, and thus recognizing and filtering out such content on social media is essential. This study presents a new approach to detect fake Twitter accounts using ontology and Semantic Web Rule Language (SWRL) rules. SWRL rules-based reasoner is utilized under predefined rules to infer whether the profile is trust or fake. This approach achieves a high detection accuracy of 97%. Furthermore, ontology classifier is an interpretable model that offers straightforward and human-interpretable decision rules.

Keywords: Social media, Semantic web, Twitter fake account, Web ontology language, Semantic web rule language, Ontology, Reasoner.

1. Introduction

In today's era, online social networking is one of the most popular applications. People of all ages spend much time on social networking, creating and exchanging large quantities of data across the globe. Millions of users are tweeting on Twitter, posting on Facebook, or sharing videos and photos on other online social networks (OSN). However, people using OSN unwillingly expose a staggering amount of personal information. Recent reports indicate that networks such as Facebook and Twitter are infested with tens of millions of fake user profiles that can collect personal details about real users and their friends. In the first quarter of 2018, statistics show that the average monthly active Twitter users amounted to 336 million worldwide, with approximately 5% fake or spam accounts [1]. Also, 9% to 15% of tweets originate from fake accounts [2]. Twitter began its battle against disinformation campaigns by removing over one million fake accounts in 2018 [3].

Fake accounts play a massive role in advanced persistent threats and other malicious activities. However, OSNs do not rigorously test whether account owners dominate their profiles. If not, another person can use the identity. Profiles can also be easily generated using openly created names and other information that cannot be attached to any specific person.

Fake accounts on OSNs can endanger user safety and cause severe damage in the online and the real worlds. OSNs entice the attention of malicious entities that attempt to exploit unintended network vulnerabilities. Threats in social networking over the Internet that users are unaware of include loss of privacy, identity theft, malware, Sybil's bots, and sexual harassment [4].

Fake account identification in the social network thus gained recent considerable attention, and different strategies to solve the problem have been proposed. Previous works are divided into three general solutions, namely, Crowdsourcing, Graph-based, and Feature-based detection. For Twitter, several works have recently been developed to recognize fake accounts using machine learning algorithms. Despite advanced data mining algorithms, ML methods open up a vast array of data uses stored in databases. They allow, among others, the systematic acquiring of the knowledge of designers, thus being a solutionthat helps to make use of what has been learned through experience [5].

The ontology classifier, on the other hand, is an interpretable construct, and can therefore provide insights into how the process makes a decision. The main contributions of the technique proposed are described below:

1. A new semantic modelling meter approach has been presented based on features of Twitter profiles account.
2. Ontology engineering and semantic web rule language rules are utilized as a classifier to differentiate bot account from a real account.
3. Area Under the ROC Curve (AUC) technique used for ranking the top essential features according to a specific threshold (the cutoff point).
4. The semantic web rule language rules have been building on the thresholds of features profile account.
5. The points scoring from rules are used as a meter for Twitter fake account detection.

The remainder of the paper is structured as follows: Section 2 is reserved for the analysis of the literature on techniques of detection and classification. Section 3 deals with the methods used in this article. Section 4 defines and assesses the degree to which the success of the model has proved successful. Section 5 provides a description of the results obtained, accompanied by a brief overview of the results predicted. Lastly, Section 6 contains offers future work conclusions and directions.

2. Related work

Fake account identification in the social network has gained considerable recent attention, and different strategies to solve the problem have been proposed. Previous works are divides into three general solutions, namely, Crowdsourcing, Graph-based, and Feature-based detection. In this study, we focus on feature-based detection. Recently, machine-learning algorithms are used to distinguish spam from non-spam accounts on Twitter. Features are extracted and selected from accounts or tweets and then applied with supervised, semi-supervised, and unsupervised machine-learning techniques to determine or classify the types of fake accounts.

I. Rojek and E. Dostatni [5] use different machine-learning techniques on content-based features extracted from tweets to classify fake and trust accounts. The five machine classifiers are Decorate, Random Forest, AdaBoost, Decision Tree and Naïve Bayesian, with the best results recorded using the first classifier.

The researchers suggest quantity approaches calculate the shift in the interest of the account and decide whether the account has a concentrated interest or a broad interest. Then, based on the degree of concentrated interest, we reflect users by characteristics. By integrating unsupervised and supervised learning, we establish a mechanism to distinguish between bots and trust account [6].

To enhance the identification of fake profiles on OSNs, a new classification model was presented, where SVM trained model decision values were used to train a NN model, and SVM test decision values were used to evaluate the NN model [7]. S. Cresi et al.[8] claim that an effective identification of spambots can be accomplished by an in-depth study of their collective activities using the digital DNA method to model the activities of users of social networks. Inspired by its biological equivalent, the behavioral lifespan of a digital account is encoded into a series of characters in the digital DNA representation. Then, for such digital DNA sequences, they establish a similarity metric. To define both real accounts and spambots, they build upon digital DNA and the similarity between groups of users. the Social Fingerprinting method using such classification, which can discriminate between spambots and legitimate accounts in both a supervised and an unsupervised manner.

A deep neural network based on the architecture of contextual Long Short-Term Memory (LSTM) that utilizes both content and metadata to identify bots at the tweet stage: semantic characteristics are extracted from user metadata and fed as an extra input to the processing of tweet by LSTM deep networks [9].

By hybridizing an established meta-heuristic technique called Whale Optimization Algorithm (WOA) with Support Vector Machines (SVM), a hybrid machine learning model was proposed with a view to identifying spammers in Twitter. The concept behind using WOA in this hybrid technique was to improve SVM's parameters along with the task of selecting the proper spam recognition features[10].

A feature-based model was suggested to classify fraudulent accounts on social media sites utilizing twenty-four features in order to effectively classify fake accounts. Three learning techniques are used (SVM, Random Forest and Logistic Regression algorithms) to validate the classification results. Experimental findings show that, using the Random Forest algorithm, our model was able to achieve 97.9 percent accuracy[11].

A. Balestrucci et al.[12] introduced a supervised classification algorithm to discriminate against credulous users, i.e., human-operated accounts with a significant proportion of bots as mates. The classifier brings about very good results and bypass the extraction process of a weak attributes. In order to improve Twitter bot detection.

J. Rodríguez et al [13] suggests to use a one-class classification, as this enables new bot accounts to be detected, and requires only samples of real accounts. The experiment results show that different types of bots with an output above 0.89 calculated using AUC can be reliably identified by the proposed model without needing previous knowledge about them. The researchers have confirmed that a supervised classification method is a viable choice for the identification of Twitter bot, because when perceptive between bots and genuine user accounts, our findings achieve an output higher than 0.95 AUC. Another model for detecting fake profiles in OSNs based on graph query and classification algorithm using the similarity between user "friends" networks proposed by Mohammadrezaei et al.[14]. These similarity measures are calculated using Jaccard, standard profile, and cosine. PCA is used for feature extraction, and SVM is used as a classification technique.

N. Sun et al.[15] Propose a semi real-time Twitter spam detection framework that offers data collection, light-weight features extraction from a single Twitter account, training detection model, and online visualization of detection results. In this method, parallel computing technology is applied to train and upgrade the models. On the basis of their datasets, empirical findings verify that the model can achieve acceptable efficiency. In addition, this framework also serves as a collection instrument for tweets. Crawler model has been developed utilize URL based detection plus machine learning techniques for the detection of malicious users based on user characteristics. Crawler is designed to include user profiles, notifications and follow up on Twitter 's website with approximately 22 k. Some user-based attributes have been evaluated, such as follower count, friend count, tweet count, etc. which are useful for fake account classification [16] .

Selvam et al. [17] present a new algorithm for spam detection based on Ontology. Numerous steps are implemented, starting with the construction of the ontology feature extraction, comparison of words with current class context, and finishing with the classification of whether spam or not spam.

The researchers[18] introduce a new approach for identification of fake accounts using a multi-objective hybrid feature selection model that enables the selection of features with optimum classification efficiency. First, the nominee feature set was defined by the Minimum Redundancy-Maximum Relevance Algorithm (mRMR) due to the highest connection to the target class and the least redundancy among the features.

An alternative ontology-based approach for detecting spam tweets was proposed exclusively through content analysis. This was accomplished because classification accuracy was much lower than anticipated. The suggested ontology has disabled the dependence on both private and user relationship information that most of the current spam detection procedures use. Study findings showed that the proposed approach outperformed nearly 200 per cent of the current spam message identification techniques [19].

Our study presents a new approach to detect fake Twitter accounts on the basis of ontology and SWRL rules by exploiting SW technologies. SWRL rules-based reasoning is used under predefined rules to infer whether the profile is fake or trust. This study is based on the Fake Project dataset released by the Institute of Informatics and Telematics of the Italian National Research Council (IIT-CNR) Lab[20].

3. Methodology

The block diagram of the proposed approach is divided into three tasks: data preprocessing and features selection, ontology construction, and SWRL rules creation and semantic reasoner classifier (Fig. 1).

3.1 Data preprocessing

The data set (Fake Project dataset) used in this study is based on the Fake Project dataset released by the Institute of Informatics and Telematics of the

Table 1. The dataaset description. 70% fake accounts and 30 trust accounts

Grouping of accounts	Number of Accounts	Number of Tweets
Fake Account	8,263	3,653,371
Trust Account	3,474	8,377,522
Total	**11,737**	**12,030,893**

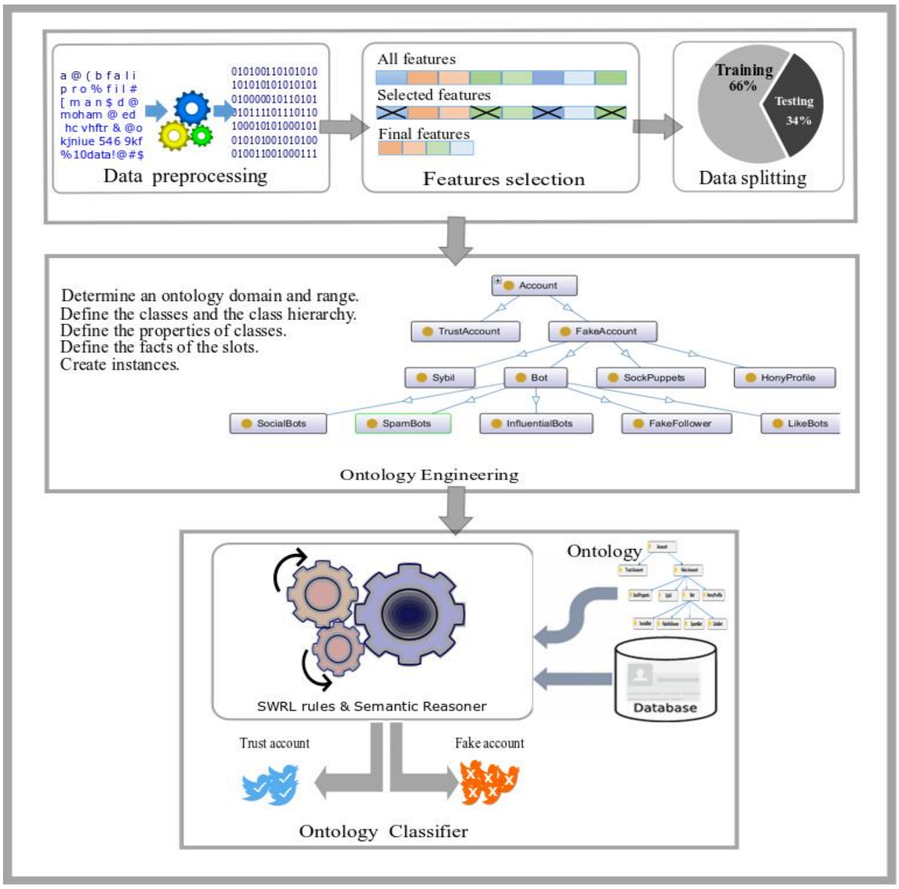

Figure. 1 Block diagram of the proposed approach

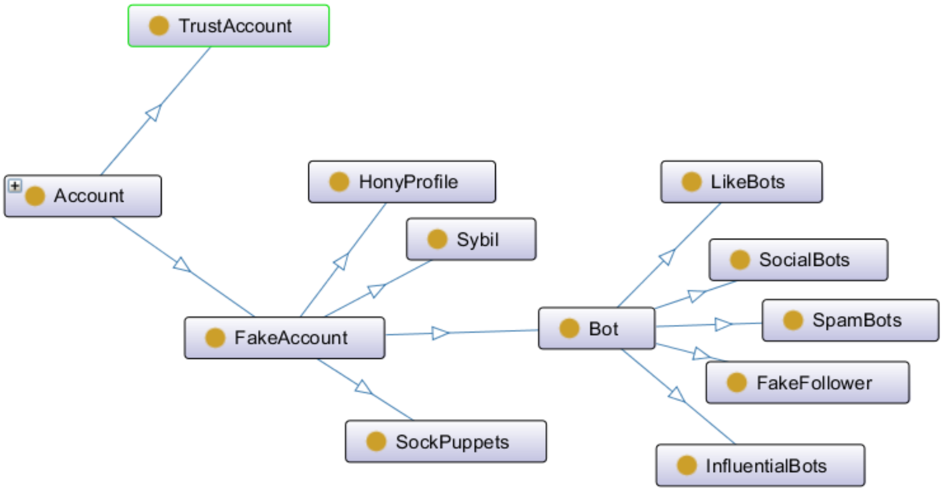

Figure. 2 The tree hierarchy of account class visualized in protégé

Italian National Research Council (IIT-CNR) Lab [20]. The data set consists of 11737 accounts, with trust (n=3474) and fake ones (n=8263). The data set is divided in two, with 7746 (66%) for training and 3991 (34%) for testing. The original data set has 11737 accounts with 12030893 tweets (Table 1).

The data can be inaccurate, inconsistent, and incomplete. Thus, we need to find the missing, corrupted, unreliable, or irrelevant parts of the data. Then, the data set is pre-processed, and the unimportant features are removed. Finally, the data set is summarized, and only the essential features are selected from the original features.

3.2 Ontology construction

Ontologies are a system for the representation of information that can be shared and reused around a domain. Their ability to explain relationships and high interconnectivity create the foundations for high quality, connected, and coherent data modelling. The architecture of ontology consists of the following steps.

- Determine an ontology domain and range.
- Define the classes and class hierarchy.
- Define the class properties.
- Define the facts of the slots.
- Create instances.

The Stanford Protégé ontology editor is utilized to develop a web ontology language to represent the classes and properties of the model. The main classes are Account and Agent, and the latter is an object that can own the former. (Fig. 2) shows the class hierarchy of the ontology as visualized in the WebVOWL ontology visualization tool [21]. WebVOWL is a plugin tab in Protégé's editor for the interactive visualization of ontologies and carries out the Visual Notation for OWL Ontologies (VOWL) by providing graphical representation for OWL elements.

3.3 SWRL rules and semantic reasoner

The SWRL is the standard guideline language of the Semantic Web. SWRL can be used to express rules and logic, incorporating OWL DL or OWL Lite with a Rule Markup Language subset (RuleML). SWRL rules form as couples of antecedent–consequent[22]. The antecedent points to the body (rules), and the subsequent part is referred to as the head. The head and body consist of one or more atoms joining together.

The SWRL rule syntax follows Eq. (1) and Eq. (2) [23]:

$$B_1 \wedge B_2 \wedge .... \wedge B_n \rightarrow H \quad (1)$$

Where, H: head (an atom) and B_n: body (all atoms), while the atoms in SWRL are defined as follows:

$$C(j)|O(j,k)|D(j,v)|\ T(v)|Built\text{-}Ins\ (f,v_1,...,v_n) \rightarrow Atom \quad (2)$$

Where C = Class, T = Data type, O = Object Property, D = Data type Property, *J, k* = Object individual names or object variable names, $v_1... v_n$ = Data type variable names or Data type value names and f = Built-in name.

SWRL is considered a major cornerstone for the realization of the semantic web and the support of creative applications based on the rules. SWRL can thus be used to infer new information from given facts [24]. All rules are expressed in terms of ontology concepts (classes, properties, and individuals) and stored as OWL syntax in the domain ontology.

OWL reasoner such as Pellet, HermiT, ELK, and FaCT++ is the most common in ontology for executing SWRL rules and inferring new ontology axioms [25, 26]. The reasoning derives information that is not presented directly in ontology or knowledge base, and classification is one of its widest usages. Pellet reasoner supports several vital logical services such as consistencies, description, and instance testing. Pellet reasoner is used in the proposed ontology because of its more direct functionality for working with OWL and SWRL rules and allowance for defining custom SWRL built-ins.

4. Evaluation

Assessment methods play a critical role in the design of a classification model. The performance of our model is assessed to obtain better results, and here is where the Confusion matrix comes to the spotlight. A confusion matrix is a method of summing up a classification algorithm results [27, 28], and its most basic terms for a binary classifier are:

True positive (TP) - the number of accounts correctly identified as Fake.

False positive (FP) - the number of accounts incorrectly identified as Fake.

True negative (TN) - the number of accounts correctly identified as Trusted.

False negative (FN) - the number of accounts incorrectly identified as Trusted.

The assessment metrics often computed from a confusion matrix are:

Precision: the proportion of fake accounts in those assumed fake as seen in Eq. (3).

$$\text{Precision} = \frac{\text{TP}}{\text{TP+FP}} \quad (3)$$

Recall: expresses relevant accounts that are correctly detected as seen in Eq. (4).

$$\text{Recall} = \frac{\text{TP}}{\text{TP+FN}} \quad (4)$$

F-**Score**: measures the quality of a prediction as seen in Eq. (5).

$$\text{F-Score} = 2 \times \frac{\text{Precision} * \text{Recall}}{\text{Precision+Recall}} \quad (5)$$

Accuracy: state the accounts are identified correctly in the total as seen in Eq. (6).

$$\text{Accuracy} = \frac{\text{TP+TN}}{\text{TP+FP+TN+TP}} \quad (6)$$

5. Results and discussion

5.1 Results

This paper presents a new approach to detect fake Twitter accounts using ontology engineering and SWRL rules. We use a published data set from the MIB website, with a testing set of 3991 accounts, distributed as 2830 fake and 1161 trust ones. New relationships (driven features) are inferred from given features (Table 2), and thereby can be used for the classification.

Area Under the ROC Curve (AUC) technique is a performance measurement for the binary classification problem at different thresholds [27, 28]. AUC–ROC Curve can be used for ranking the top and essential features from different ones according to a specific threshold (the cutoff point). Thresholded classifications are monotonic; and therefore, any feature that ranks positive for a given threshold also ranks positive for all lower thresholds. A threshold (or cutoff) value determines how expected posterior probabilities translate to class labels for binary scoring classification [29], and represents the range of criterion values that determine a positive condition giving the highest accuracy. The top ten features are listed according to their effect on the account classification process starting by most significant feature (highest impact) to least significant feature that was above of 79% as shown in the Table 3.

Table 2. Driven features. New relationships (features) ware extracted from existing features

Relationship	Description
FriendShip	Ratio of friendsCount to followersCount
FollowerShip	Ratio of followersCount to friendsCount
Interestingness	Ratio of favouritesCount to statusesCount
Activeness	Ratio of statusesCount to AccountAge
FriendRate	Ratio of Friends count to AccountAge
FollowerRate	Ratio of followersCount to AccountAge
Reputation	Ratio of followersCount to sum of friendsCount and followersCount

Table 3. The ranking of the most important features according thresholds (the cut-off point)

Feature name	Ranking	weight
favouritesCount	1	0.931
interest	2	0.893
statusesCount	3	0.888
geoEnabled	4	0.877
followersCount	5	0.87
accountAge	6	0.862
friendRate	7	0.844
reputation	8	0.835
friendShip	9	0.814
listedCount	10	0.792

Table 4. A confusion matrix of faked and trusted accounts classification

		Actual values	
		Faked	Trusted
Predictive value	Faked	TP= 2768	FP= 37
	Trusted	FN= 62	TN= 1124

The thresholds are utilized to build the tested rules, which are used to calculate scoring of each account. The score of each account that greater than or equal five can be estimated as fake account. All test rules are converted into SWRL rules using a conversion technique called rolification. A Pallet reasoner is applied to infer whether the account is fake or trusts on the basis of its final score, being interested in false account detection, we consider the ability of each rule to detect and classify a fake account. The results show that of the 2805 accounts, 2768 are correctly classified as fake, and 37 are classified as real. Meanwhile, 62 of 1186 real accounts are classified as fake. Total accuracy is 97.5 (Table 4).

5.2 Discussion

Table 5 shows the classification performance of different methods, including the proposed ontology meter, on the bot detection method based based on profile features. Classic off-the-shelf baseline approaches reflect machine learning techniques: their accuracy ranges from 80% to 98%. To prove that profile features-based bot identification can be performed with extremely great precision, small number of features, and limited-size training datasets, the ontology meter approaches precision averaging above 97% accuracy. The Random Forest classifier appears to consistently provide the best output across all feature-based bot detection benchmarks among conventional machine learning models in [6-9, 11, 18].

If we discuss the results from the view of interpretable and clarity (How & Why), the ontology outcomes considered the best because the make of decision is understandable and easily described. Machine learning is an artificial intelligence method wherein machines use mathematical algorithms to

Table 5. Evaluation metrics clarify that the accuracy of ontology meter approach are very close or achieve better performance compared to machine learning techniques of other related works

Technique	**accuracy**	**precision**	**recall**	**F-M**
Alghamdi et al. 2018 [6]				
Random Forest	0.962	0.962	0.962	0.962
J48	0.953	0.954	0.953	0.949
Decision Table	0.949	0.949	0.95	0.949
Sarah et al. 2018 [7]				
Correlation	0.983	0.977	0.998	0.987
Regression	0.960	0.973	0.976	0.975
Wrapper-SVM	0.960	0.973	0.976	0.975
S. Cresci et al. 2017 [8]				
Unsupervised	0.976	0.982	0.972	0.977
Supervised	0.977	0.982	0.977	0.977
S. Kudugunta and E. Ferrara – 2018 [9]				
Random Forest	0.984	0.98	0.98	0.98
Logistic Regression	0.957	0.94	0.93	0.93
AdaBoost Classifier	0.982	0.98	0.98	0.98
Jyoti Kaubiyal et al. 2019 [11]				
Random Forest	0.977	0.98	0.98	0.98
Logistic Regression	0.957	0.94	0.96	0.95
SVM	0.808	0.82	100	0.90
R. Rostami and S. Karbasi 2020 [18]				
Technique	***accuracy***	***precision***	***recall***	***F-M***
Random Forest	0.97	--	--	0.969
Naïve Bayes	0.971	--	--	0.97
SVM	0.965	--	--	0.966
Our study 2020				
Ontology meter	**0.975**	**0.975**	**0.986**	**0.983**

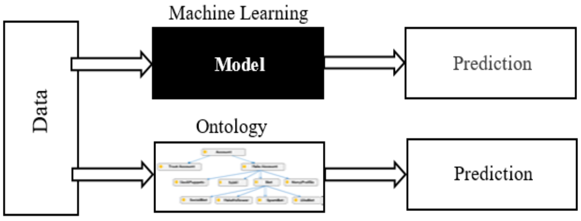

Figure. 3 Machine learning classifier vs onnology classifier. Machine learning algorithms are "black boxes" in the manner that they can make excellent predictions but that the logic behind those predictions is not understandable. In contrast, our ontology classifier is an interpretable model

learn from the data. Despite using accurate mathematical models, understanding the criteria or rules used for decision making or to determine the output source (black box) can be challenging. Black-box classifiers do not offer straightforward and human-interpretable decision rules [30], and thus understanding every piece of the model or generating a set of rules for making decisions is not necessarily easy. By contrast, the ontology classifier is an interpretable model and thus can provide insights into how the model arrives at a decision. The results of this approach are very close and comparable with those of machine learning techniques. Furthermore, the findings are human-interpretable, and decision rules can be easily described (Fig. 3).

The significant rules with the best outcomes are selected according to important related features (see Table 4). The rules are translated into SWRL and Pallet reasoner is then applied to infer whether the account is fake or trust. The final classification decision depends on the final aggregated score. The SWRL rules classification criteria are listed in (Table 5).

Fake account estimation depends on point-scoring, which is a cumulative summation outcome of all rules. Each rule gains zero point or one point according to specific predefined criteria and the final score falls within the range [0, 10]. The final score is represented as a threshold that determines whether the account is fake (score is greater than or equal to 5) or trust (score is less than 5), as shown in Eq. (7).

$$\text{Scoring (account)} = \begin{matrix} < \mathbf{5} & \text{Trust account} \\ \geq \mathbf{5} & \text{Fake account} \end{matrix} \qquad (7)$$

6. Conclusion

In this study, we present a new approach for fake account detection based on domain ontology and SWRL rules. The ontology contains the relevant concepts related to the discovery and data of Twitter profile accounts. SWRL rules are created from strong relationships between ontology concepts to estimate the fake account. The functions of the approach are divided into four tasks, namely, data preparation, classification and relationship, ontology building, and finally, application of SWRL rules to infer whether the account is fake or not. All detecting rules and relevant knowledge are extracted from the domain ontology (FakAccOnt). The reasoner uses profile account data and rules to draw the inference and provides the final decision to show that the nature of the account is fake or trust. Experiments are carried out to test the approach performance in recognizing 3991 Twitter accounts. System evaluation depends on SWRL rules and standard evaluation metrics, and

Table 5. The SWRL testing rules are crated from existing facts and new driven relationships for getting final score. The final score used as threshold to distinguish a count is fake or not

Rule	SWRL rules
Rule-1	Account(?a) ^ hasFavouritesCount(?a, ?v) ^ swrlb:lessThanOrEqual(?v, 3) ^ swrlb:add(?s1,1) → Score1(?a, ?s1)
Rule-2	Account(?a) ^hasFavouritesCount(?a, ?fv) ^ hasStatusesCount(?a, ?st) ^ swrlb:add(?m2, ?st, 0.01) ^ swrlb:add(?m1, ?fv, 0.01) ^swrlb:divide(?m, ?m1, ?m2) ^ swrlb:greaterThan(?m, 0.0296) ^ swrlb:add(?s2,1) → Score2(?a, ?s2)
Rule-3	Account(?a) ^ hasStatusesCount(?a, ?st) ^ swrlb:lessThanOrEqual(?st, 144) ^ swrlb:add(?s3,1) → Score3(?a, ?s3)
Rule-4	Account(?a) ^ hasGeoEnabled(?a, ?g) ^ swrlb:equal(?g, 0) ^ swrlb:add(?s4,1) →Score4(?a, ?s4)
Rule-5	Account(?a) ^ hasFollowersCount(?a, ?fl) ^ swrlb:lessThanOrEqual(?fl, 26) ^ swrlb:add(?s5, 1) → Score5(?a, ?s5)
Rule-6	Account(?a) ^ hasAge(?a, ?ag) ^ swrlb:lessThanOrEqual(?ag, 77) ^ swrlb:add(?s6, 0, 1) → Score6(?a, ?s6)
Rule-7	Account(?a) ^ hasAge(?a, ?ag) ^ hasFriendsCount(?a, ?fr) ^ swrlb:add(?m1, ?fr, 0.01)^ swrlb:add(?m2, ?ag, 0.01) ^ swrlb:divide(?m, ?m1, ?m2) ^ swrlb:greaterThan(?m, 0.446) ^ swrlb:add(?s7, 0, 1) →Score7(?a, ?s7)
Rule-8	Account(?a) ^ hasFollowersCount(?a, ?fl) ^ swrlb:add(?m1, ?fl, 0.01) ^ hasFriendsCount(?a, ?fr) ^swrlb:add(?m2, ?fr, 0.01) ^ swrlb:add(?m, ?m2, ?m1)^ swrlb:divide(?r,?fl,?m)^swrlb:lessThan(?r,0.192) ^ swrlb:add(?s8, 1) →Score8(?a, ?s8)
Rule-9	Account(?a) ^ hasFollowersCount(?a, ?fl) ^ hasFriendsCount(?a, ?fr) ^ swrlb:add(?m1, ?fl, 0.01) ^ swrlb:add(?m2, ?fr, 0.01) ^ swrlb:divide(?m, ?m1, ?m2) ^ swrlb:lessThanOrEqual(?m, 0.22) ^ swrlb:add(?s9, 0, 1) →Score9(?a, ?s9)
Rule-10	Account(?a) ^ haslistedCount(?a, ?ls) ^ swrlb:lessThanOrEqual(?ls, 1) ^ swrlb:add(?s10,1) -> Score10(?a, ?s10)
Scoring rule	Account(?a) ^ Score1(?a, ?x1) ^ Score2(?a, ?x2) ^ Score3(?a, ?x3) ^ Score4(?a, ?x4) ^ Score5(?a, ?x5) ^ Score6(?a, ?x6) ^ Score7(?a, ?x7) ^ Score8(?a, ?x8) ^ Score9(?a, ?x9) ^ Score10(?a, ?x10) ^ swrlb:add(?sum, ?x1, ?x2, ?x3, ?x4, ?x5, ?x6, ?x7, ?x8, ?x9, ?x10) ^ swrlb:greaterThanOrEqual(?sum, 5) →FakeAccount(?a)

results show that the proposed approach can correctly identify 2768 out of the 2805 fake accounts, amounting to an accuracy of 97.5%. The main contribution of this study is using ontology and SWRL rules to detect Twitter profile accounts according to the meter function.

Machine learning models are described as black boxes that are ambigiuose. In several applications, while they have outstanding results, different researchers request attention to the issue of interpretability. Recently, many studies have aimed to shed light on the inner rules those used for decisions making. With all that in view, by restructured the knowledge represantation, we try to present a clear procedure to understand and interpret on how decision making and how extraction hidden knowledge. Also, ontology classifier is an interpretable model that offers straightforward and human-interpretable decision rules

Conflicts of Interest

The authors declare no conflict of interest.

Author Contributions

The paper conceptualization, methodology, software, validation, formal analysis, investigation, resources, writing—original draft preparation, writing—review and editing, visualization, have been done by 1st author. The supervision, and project administration, have been done by 2nd author.

Acknowledgments

Acknowledgments are to show that the article is supported by what organization. For example, “This work was supported by the National Nature Science Foundation under Grant No. 405”.

References

[1] C. Timberg and E. Dwoskin, *Twitter is sweeping out fake accounts like never before, putting user growth at risk,* 2018 Available: https://www.washingtonpost.com/technology/2018/07/06/twitter-is-sweeping-out-fake-

accounts-like-never-before-putting-user-growth-risk/

[2] Z. Chong, *Up to 48 million Twitter accounts are bots, study says*, 2017. Available: https: //www.cnet.com/ news/new-study-says-almost-15-percent- of-twitter-accounts-are-bots/

[3] O. Gonzalez, *Twitter removes thousands of fake, state-backed accounts*, 2019. Available: https://www.cnet.com/news/twitter-removes-thousands-of-fake-state-backed-operation-accounts/

[4] S. Deliri and M. Albanese, "Security and privacy issues in social networks", In *Data Management in Pervasive Systems, Springer*, pp. 195-209, 2015.

[5] I. Rojek and E. Dostatni, "Machine learning methods for optimal compatibility of materials in ecodesign", *Bulletin of the Polish Academy of Sciences: Technical Sciences*, Vol. 68, No. 2, pp. 199-206, 2020.

[6] B. Alghamdi, Y. Xu, and J. Watson, "A Hybrid Approach for Detecting Spammers in Online Social Networks", In: *Proc. of International Conf. on Web Information Systems Engineering*, pp. 189-198, 2018.

[7] S. Khaled, N. El-Tazi, and H. M. Mokhtar, "Detecting Fake Accounts on Social Media!", In: *Proc. of IEEE International Conf. on Big Data (Big Data)*, pp. 3672-3681, 2018.

[8] C. Stefano, D. P. Roberto, P. Marinella, S. Angelo, and M. Tesconi, "Social fingerprinting: detection of spambot groups through DNA-inspired behavioral modeling", *IEEE Transactions on Dependable Secure Computing ACM Transactions on the Web*, Vol. 15, No.4, pp. 561-576, 2017.

[9] S. Kudugunta and E. Ferrara, "Deep neural networks for bot detection", *Information Sciences*, Vol. 467, pp. 312-322, 2018.

[10] A. Al-Zoubi, F. Hossam, A. Ja'far, and H. Mohammad, "Evolving support vector machines using whale optimization algorithm for spam profiles detection on online social networks in different lingual contexts", *Knowledge-Based Systems*, Vol. 153, pp. 91-104, 2018.

[11] K. Jyoti and J. A. Kumar, "A Feature Based Approach to Detect Fake Profiles in Twitter", In: *Proc. of the 3rd International Conf. on Big Data and Internet of Things*, pp. 135-139, 2019.

[12] A. Balestrucci, R. De Nicola, M. Petrocchi, and C. Trubiani, "Do you really follow them? automatic detection of credulous Twitter users", In: *Proc. of International Conf. on Intelligent Data Engineering and Automated Learning*, pp. 402-410, 2019.

[13] R. Jorge, M. Javier, M. Raúl, L. Octavio, and L. Armando, "A one-class classification approach for bot detection on Twitter", *Computers Security*, Vol. 91, pp, 101715, 2020.

[14] M. Mohammadrezaei, M. E. Shiri, A. M. J. S. Rahmani, and C. Networks, "Identifying fake accounts on social networks based on graph analysis and classification algorithms", *Security and Communication Networks*, Vol. 91, pp. 1-8, 2018.

[15] S. Nan, L. Guanjun, Q. Junyang, R. Paul, and Applications, "Near real-time twitter spam detection with machine learning techniques", *International Journal of Computers and Applications*, pp. 1-11, 2020.

[16] Yosef Jbara and H. Mohamed, "Twitter Spammer Identification using URL based Detection", In: *Proc. of IOP Conf. Series: Materials Science and Engineering*, Vol. 925, No.1, pp 012014, 2020.

[17] S. Sheba, B. Ramadoss, and R. Balasundaram, "Social event detection-A systematic approach using ontology and linked open data with significance to semantic links", *The International Arab Journal of Information Technology*, Vol. 15, No. 4, pp. 729-738, 2018.

[18] R. Ramzanzadeh and K. Soheila, "Detecting Fake Accounts on Twitter Social Network Using Multi-Objective Hybrid Feature Selection Approach", *Webology*, Vol. 17, No.1, 2020.

[19] B. Halawi, A. Mourad, H. Otrok, and E. Damiani, "Few are as good as many: an Ontology-based tweet spam detection approach", *IEEE Access*, Vol. 6, pp. 63890-63904, 2018.

[20] S. Cresci, R. Di Pietro, M. Petrocchi, A. Spognardi, and M. Tesconi, "The paradigm-shift of social spambots: Evidence, theories, and tools for the arms race", In: *Proc. of the 26th International Conf. on World Wide Web companion*, pp. 963-972, 2017.

[21] S. Lohmann, V. Link, E. Marbach, and S. Negru, "WebVOWL: Web-based visualization of ontologies", In: *Proc. of International Conf. on Knowledge Engineering and Knowledge Management*, pp. 154-158, 2014.

[22] W. Li, H. Kang, D. Ma, and W. Wei, "SWRL Parallel Reasoning Method with Spark SQL", In: *Proc. of IEEE/ACIS 18th International Conference on Computer and Information Science (ICIS)*, pp. 270-273, 2019.

[23] V. Karimi, "Semantic Web Rule Language (SWRL)", pp. 1-37, 2008. Available: https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.222.6385&rep=rep1&type=pdf

[24] P. Benedikt and F. Hans-Georg, “A visual modeling approach for the Semantic Web Rule Language”, *Semantic Web*, Vol. 11, No. 2, pp. 361-389, 2020.

[25] E. Sirin, B. Parsia, B. C. Grau, A. Kalyanpur, and Y. Katz, “Pellet: A practical owl-dl reasoner”, *Journal of Web Semantics*, Vol. 5, No. 2, pp. 51-53, 2007.

[26] A. Sunitha, “A survey on ontology reasoners and comparison”, *International Journal of Computer Applications*, Vol. 57, No. 17, 2012.

[27] D. M. Powers, “Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation”, *International Journal of Machine Learning Technology*, Vol. 1, No. 2 pp. 37-63, 2020

[28] M. Sokolova, N. Japkowicz, and S. Szpakowicz, “Beyond accuracy, F-score and ROC: a family of discriminant measures for performance evaluation”, In: *Proc. of Australasian Joint Conf. on Artificial Intelligence*, pp. 1015-1021, 2006.

[29] A. del-Río-Ortega, F. García, M. Resinas, E. Weber, F. Ruiz, and A. Ruiz-Cortés, “Enriching decision making with data-based thresholds of process-related KPIs”, In: *Proc. of International Conf. on Advanced Information Systems Engineering*, pp. 193-209, 2017.

[30] R. Guidotti, A. Monreale, S. Ruggieri, F. Turini, F. Giannotti, and D. Pedreschi, “A survey of methods for explaining black box models”, *ACM Computing Surveys*, Vol. 51, No. 5, pp. 1-42, 2018.