future internet

Article

Automatic Detection of Online Recruitment Frauds: Characteristics, Methods, and a Public Dataset

Sokratis Vidros [1,*,†], Constantinos Kolias [2], Georgios Kambourakis [1,2] and Leman Akoglu [3]

[1] Department of Information & Communication Systems Engineering, University of the Aegean, Karlovassi, Samos 83200, Greece; gkamb@aegean.gr
[2] Computer Science Department, George Mason University, Fairfax, VA 22030, USA; kkolias@gmu.edu
[3] H. John Heinz III College, Carnegie Mellon University, Pittsburgh, PA 15213, USA; lakoglu@cs.cmu.edu
* Correspondence: sokratis.vidros@aegean.gr; Tel.: +30-22730-82256
† Current address: Laboratory of Information and Communication Systems Security,
Department of Information and Communication Systems Engineering, University of the Aegean,
Karlovassi, Samos 83200, Greece.

Academic Editor: Dino Giuli
Received: 16 December 2016; Accepted: 17 February 2017; Published: 3 March 2017

Abstract: The critical process of hiring has relatively recently been ported to the cloud. Specifically, the automated systems responsible for completing the recruitment of new employees in an online fashion, aim to make the hiring process more immediate, accurate and cost-efficient. However, the online exposure of such traditional business procedures has introduced new points of failure that may lead to privacy loss for applicants and harm the reputation of organizations. So far, the most common case of Online Recruitment Frauds (ORF), is employment scam. Unlike relevant online fraud problems, the tackling of ORF has not yet received the proper attention, remaining largely unexplored until now. Responding to this need, the work at hand defines and describes the characteristics of this severe and timely novel cyber security research topic. At the same time, it contributes and evaluates the first to our knowledge publicly available dataset of 17,880 annotated job ads, retrieved from the use of a real-life system.

Keywords: fraud detection; online recruitment; employment scam; job scam; data mining; machine learning; natural language processing; dataset

1. Introduction

Nowadays, the most critical procedures of corporations are already being conducted through Software as a Service (SaaS) products. To this direction, the critical procedure of hiring has been successfully ported to the cloud. Hiring can be modelled as a multistep process that starts with composing and advertising job ads and ends with successful hires. To accomplish their mission more effectively, hiring managers rely on a wide range of cloud-based solutions, namely the Applicant Tracking Systems (ATS).

On the downside, the increasing adoption of ATS has also attracted the interest of scammers. In most cases, this phenomenon (a) jeopardizes job seekers' privacy; (b) leads to financial loses; and (c) vitiates the credibility of organizations. Today, as detailed in Section 3, job frauds have become rather resourceful. Phony content is hard to be distinguished from benign, so countermeasures are usually ad-hoc and their practical value is often questionable. Furthermore, as discussed in Section 4, the peculiarities of this particular field render the application of existing solutions developed for relevant problems difficult to adapt. Specifically, although considerable work has been carried out to tackle relevant problems such as email spam [1–3], phishing [4], Wikipedia vandalism [5,6], cyber bullying [7,8], trolling [9], and opinion fraud [10]. Nevertheless, when it comes to employment scam, the proposed solution have proven inadequate in practice. Our preliminary study leads to the conclusion that text and metadata based classification can only provide a stepping stone towards an effective scam prediction system. A fully-fledged solution will require a composite approach that involves deep inspection and analysis of user, organization and network data.

The contribution of this work is twofold: on one hand, we examine the diverse aspects of Online Recruitment Fraud (ORF) and more specifically employment scam, underlining its severity and distinct characteristics. Major similarities and differences between the investigated problem and related ones are also identified. On the other hand, this work offers the first of its kind publicly available dataset [11] containing real-life legitimate and fraudulent job ads, published in English spoken countries. Moreover, we offer an exploratory evaluation of the dataset by different means, including empirical evaluation, bag of words modeling, and machine learning analysis.

The remainder of this paper is comprised of the following sections: Section 2 elucidates the value of ATS. Section 3 describes the problem at hand and elaborates on its consequences. Section 4 outlines the relevant problems and solutions proposed in the literature so far. The contributed dataset is described in full detail in Section 5. Section 6 contains the experiments conducted using this dataset. The final section contains the conclusions and closing remarks.

2. Applicant Tracking Systems

Before we delve into the problem of employment scam, we consider it necessary to succinctly discuss the context around ATS. An ATS, also known as a candidate management system, is a software designed to help organisations identify and recruit employees more efficiently by enabling the electronic management of recruitment phases. Such systems bear similarities to Customer Relationship Management (CRM) systems, but are tailored to meet the special requirements of the recruitment process.

In further detail, ATS automate the most time consuming procedures of hiring such as (a) scaffolding job ads; (b) publishing them; (c) collecting incoming resumes; (d) managing candidate communication; (e) collaborating on hiring decisions; and (f) reducing the hassle of managing candidates by sorting through thousands of resumes to determine which ones are the best fit for a given position. In fact, recent surveys show that 75% of the industry is using ATS to review and screen candidate profiles [12]. Although initially such systems were solely utilized by large corporations, nowadays small businesses take advantage of their benefits as well.

One of the core features of an ATS is the streamlined dissemination of new openings to multiple job boards (e.g., Indeed, Monster, CareerBuilder, etc.), social networks (e.g., Linkedin, Facebook, Twitter, etc.) and email addresses. According to Bloomberg, there were around 5 million job openings in USA in 2015, most of which were published online [13]. Furthermore, 51% of the employed workers are either actively seeking, or open to a new career opportunity using web resources [14]. Actually, this automated job distribution is achieved through (a) direct feeds and API integrations; (b) shared URLs and dedicated mailbox addresses [15].

3. The Problem of Online Recruitment Fraud

ORF is a form of malicious behaviour aiming to inflict loss of privacy, economic damage or harm the reputation of the stakeholders of ATS by manipulating the functions of these systems, most frequently the job-advertisement publishing process. The latter case is also referred to as employment scam or job scam and it comprises the main axis of this work.

In 2012 a job seeker acquired more than 600 resumes in one day after they decided to post a fake job ad on Craigslist in order to identify their competitors [16]. In the same year, the Australian Bureau of Statistics published a report about personal fraud stating that 6 million people per year were exposed to several forms of scam including employment scam [17]. This exposure resulted in financial losses for the Australian economy. According to Workable, a popular ATS that supports all activities pertaining to the recruitment process [18], well-crafted fraudulent job ads for blue-collar or secretarial positions in highly populated countries will reach to large audience rapidly, allowing scammers to effortlessly collect around 1000 resumes per day.

All the above facts demonstrate the serious repercussions of ORF. Although as discussed in Section 4, related problems are well-studied, tackling ORF remains largely unexplored until now. Most of the relevant documented mentions can be found in recruiting blog posts [19], discussions on hiring fora [20], published articles from business consultants [21] and security researchers [15]. Such sources educate job seekers on identifying fraudulent job opening while job boards warn job seekers on the consequences of employment scams and provide reporting tools for any malicious activity.

In practice, ATS and job boards generally build ad-hoc defences against employment scams and design countermeasures according to their sales policy. For example, ATS with free registration employ in-house, inflexible solutions of questionable value. Other ATS enforce customers to use valid corporate email addresses upon signup and perform additional verification steps including DNS Mail Exchanger (DNS MX) lookups to prevent spoofing. Conversely, ATS with restricted registration have no particular need for fraud detection systems as such systems rely on outbound sales where a salesperson contacts the potential client and verifies their profile in advance.

The ORF Facets

The methods followed by scam practitioners can be classified into two main groups of increasing severity. The first category comprises of rogue job advertisements that aim at harvesting contact information. By luring the user into filling application forms corresponding to non-existing positions, an ill-motivated individual is able to build a database usually containing the full name, phone number, address and ZIP code of that user. More sophisticated scammers may also elicit the educational and working experience profile of their victims and end up aggregating this information with other contextual socioeconomic data. Comprehensive databases and statistical results can then be re-sold to third parties such as cold-callers, aggressive marketeers, communicators for political campaigns or even to website administrators who plan to send out targeted bulk email messages containing links to generate page views and inbound traffic. In other cases, the collected email addresses can be used to forward spam emails.

The second class of job scamming is sleazier as it aims towards complete identity theft that can later be used as part of economic chicaneries such as money-laundering and reshipping fraud [22]. In this case, scammers assume the role of a legitimate or fictional employer and use the ATS as a medium to disseminate content about fake job positions. These posts re-direct the users to external methods of communication (i.e., site, email address or telephone number). From that point on, they may engage into a series of actions such as dissemination of fake skills test, scheduling of phony interviews, transmission of congratulatory emails for successful onboarding, etc. The ulterior purpose is to convince the victim to hand out extremely sensitive documents such as Social Security Numbers, identity cards, and passports or unwittingly become a "money mule" and use their bank accounts to help criminals launder money [23]. Alternatively, scammers may even trick candidates into filling out an inquiry that looks like a direct deposit form, aiming to steal their bank information and routing numbers or drive them to a direct wire transfer usually under the guise of working visa and travel expenses [24].

4. Employment Scam Detection & Relevant Problems

The problem of employment scam detection can be defined as the process of distinguishing the subset among the sum of content of an ATS that aims at being used for fraudulent activities instead of valid recruiting. Such a process is typically achieved by correlating information about the textual, structural and contextual attributes of that content. One can easily notice that employment scam detection shares common characteristics with relevant problems such as email spam, phishing, Wikipedia vandalism, cyber bullying, trolling, and opinion fraud. This section analyses the relevant problems and presents significant works proposed as countermeasures for each of them. It is stressed however that the aim of the current section is not to provide a complete review of these problems. Instead, it is intended to provide the reader with a solid grasp of the similarities and differences that these problems present when compared to employment scam.

4.1. Email Spam

Email spam is an unsolicited bulk email traffic posted blindly to numerous recipients [25]. Spammers exploit the fact that the Simple Mail Transfer Protocol (SMTP) lacks a reliable mechanism for verifying the identity of the message source, and as a result, craft spam emails that usually contain hyperlinks that redirect to phishing web sites, conceal malicious executable code or transfer attached malware. This practice is also related to several types of online fraud and frequently constitutes the stepping stone for identity theft attacks, wire scam, and employment fraud.

Spam filtering is a well-studied problem and can be applied in every phase of the email communication starting from transmission routers [26] to recipient's mailbox. It is also driven by the social characteristics of recipients such as their circle of contacts. That is, contacts of a user are less likely to send the user an unwanted message even if that message has advertising content. In addition, technical characteristics that imply abuse of the protocols (e.g., spoof of email addresses) or senders who are responsible for dissemination of large volumes of email can also be utilized.

The proposed solutions range from various sender authentication protocols [27–30] to trained classifiers that discriminate between normal and junk emails. The employed features are extracted from the message body and the message headers [31,32].

Spamcop, a Naive Bayes classifier proposed by Pantel and Lin [33] was an early spam filter. Few years later, Naive Bayes became prominent in spam filtering due to Graham's article "A Plan for Spam" [34]. Androutsopoulos et al. [35], Kanaris et al. [36], Ciltik and Gungor [37] continued by using sequences of characters obtained through the application of a sliding window (n-gram models). Drucker et al. [38] introduced Support Vector Machines (SVM) to tackle spamming. Sculley and Wachman [39] reduced the computational cost of updating the hypothesis of SVM classifiers by training only on actual errors. Yeh et al. [40], and Hershkop [41] trained classifiers with significant fraudulent behaviors such as incorrect dates in message body or noticeable discrepancies in user's past email activity (behavior-based filtering). Bratko [42] focused on adaptive statistical compression models used as probabilistic text classifiers that work on character streams. For a holistic review of the spam filtering countermeasures the reader should refer to the work of Blanzieri and Bryl [1], Guzella and Caminhas [2] and Saadat [3].

4.2. Phishing

Phishing combines social engineering with sophisticated attack vectors to direct users to bogus websites in an attempt to (a) increase website traffic; (b) spread malware; (c) unleash Cross-site scripting (XSS) attacks; and (d) acquire sensitive information. Usability tests proved that participants were unable to differentiate between legitimate and fake web sites while anti-phishing security indicators in major web browsers were almost invisible to the layman [43].

Phishing content detection can be benefited from technical information attesting unauthorized redirects to other domains, the level of visual or structural similarity among online services as well as from previously reported bad user experience. Similarly to spam filtering, supervised classification algorithms such as regression trees, SVM, Random Forest (RF), and Neural Networks (NN) have been used extensively for phishing detection. The classifiers rely on features extracted from the page URL (punctuation and random tokens in URLs), the page content (spam words detection), the page layout and design (sloppy HTML markup and clumsy stylesheets) and network characteristics (blacklisted domain or IP address, spoofed DNS records). Abu-Nimeh et al. [4] published an in-depth comparison of various anti-phishing machine learning techniques. Moreover, Cantina [44], a TF-IDF approach to detect fake web sites, evaluated lexical signatures extracted from the content of the suspicious web page. Lastly, a distinct anti-phishing approach presented by Wenyin et al. [45], introduced Dynamic Security Skins that depend on the visual similarity between fake and legitimate websites.

4.3. Wikipedia Vandalism

Crowdsourced online encyclopedias like Wikipedia are susceptible to vandalism; in other words, blatantly unproductive false edits that undermine entries credibility and integrity, thereby forcing administrators to manually amend the content. Reported incidents vary from easily spotted vulgar language to inconspicuous alterations in articles such as placing arbitrary names in historical narratives and tampering with dates.

The Wikipedia platform provides access to full revision history where spiteful alterations of the context may be located easily, and if necessary reverted. Metadata such as whether an article has been edited anonymously may be indicative of cases of vandalism, while the contributions of eponymous users (or users with a static IP address) are stored and can be analyzed to discover systematic attempts to undermine platform's objectiveness. Thus, the reputation of a user inside the platform as well as the extent of alteration of an article across time may serve as additional strong signs of ill-motivated content.

The proposed solutions combine the aforementioned characteristics with Natural Language Processing (NLP) and machine learning classification. Potthast et al's. preliminary study [5] and PANWikipedia vandalism corpus using Amazon's Mechanical Turk [6] set a common ground for researchers working in the field. Wang and McKeown [46] proposed a shallow syntactic semantic modelling based on topic specific n-tags and syntactic n-grams models trained on web search results about the topic in question. Other researchers such as Chin et al. [47] trained an active learning statistical language model to address existing incomplete datasets and suggested a three type taxonomy of vandalism instances: (a) "Blanking" or "Large-scale Editing", defined as a 90% difference in context length between two consecutive revisions, (b) "Graffitti", namely the insertion of unproductive, irrelevant or unintelligible text and (c) "Misinformation" that involves changes in existing entities such as names, brands or locations. Harpalani et al. [48] boosted vandalism detection by using stylometric and sentiment analysis. In particular, their study was based on the fact that Wikipedia authors strive to maintain a neutral and objective voice in contrast to vandals who aim at polarization and provocation. Meanwhile, latest researches [49] based on spatial (e.g., edit timestamp, revision length, user registration timestamp) and temporal features (e.g., geographical location, country reputation, content categories) resulted in lightweight and robust solutions yet to be thoroughly field-tested and evaluated.

4.4. Cyber Bullying

Cyberbullying is defined as an aggressive, intentional act carried out by a group or individual systematically, using electronic forms of contact. The victims of cyberbullying are usually users who are unable to carry out the proper legal actions as a response, due to, say, their young age.

Early approaches to tackle the problem attempted to detect threatening and intimidating content by focusing on individual comments. Dinakar et al. [7], applied assorted binary and multiclass classifiers to a manually labelled corpus of YouTube comments modelled as "bag of words". Chen et al. [8] proposed the Lexical Syntactic Feature (LSF) architecture to identify offensive users in social media by incorporating hand-authoring syntactic rules into the presented feature set.

State-of-the-art studies are concentrated around unified approaches where bullying detection relies on broader, heterogeneous features and text mining paradigms. The proposed feature sets combine profane content [50], gender information [51], and user activity history across multiple social networks [52]. For instance, if someone gets bullied on Facebook, later on, Twitter postings can be an indication of victim's feelings and state of the mind. Unlike previous approaches, Potha and Maragoudakis [9] addressed this issue using time series modelling. That is, instead of monitoring an online conversation in a fixed window, they took advantage of the whole thread and modelled it as a signal whose magnitude is the degree of bullying content.

4.5. Trolling

Users who disrupt the on-topic discussions at social media, chat rooms, fora and blogs, namely trolls, attempt to provoke readers into an emotional response. This can degrade the quality of the content of web services or inflict psychological trauma to the users.

Trolling detection systems follow common text mining paradigms and utilize conventional supervised classifiers trained with statistical and syntactic features extracted from inapt messages posted by users with known identifiers. Cheng et al. [53] presented a data-driven study of antisocial behavior in online communities and designed a system that predicts trolling by monitoring user behavior at an early stage, that is, by observing a user's first ten posts after signing up. Santos et al. [54] worked on trolling detection through collective classification, a semi-supervised approach that blends the relational structure of labelled and unlabelled datasets to increase the algorithm's accuracy.

4.6. Opinion Fraud

Opinion fraud, also known as review spamming, is the deliberate posting of deceptive and misleading fake reviews to promote or discredit target products and services such as hotels, restaurants, publications and SaaS [55] products. The main obstacle while designing countermeasures is the unpredictability of human reviewing methods. Popular review hosting sites such as Yelp.com have built proprietary domain-specific review spamming filters [56]. According to Heydari et al. the great mass of the proposed methodologies focus on spam review detection compared to spammer and spammer groups detection [10].

Supervised learning is the dominant technique in opinion fraud detection. Most of the employed features fall into three groups namely (a) linguistic features; (b) behavioral features and (c) network features. Such features derive from the content of the review, the metadata information and the information about the targeted product or service. Previous work in the field pieced together content based features, genre identification, POS analysis, psycholinguistic deception detection, n-gram-based text categorization techniques [57–59] as well as deep syntactic stylometry patterns based on context free grammar parse trees [60]. Li et al. [61] analysed a real-world dataset provided by Dianping.com and generated features based on spatial and temporal patterns by leveraging the dependencies among reviews, users and network characteristics such as IP addresses. Fei et al. studied the burst patterns of review scammers and employed Markov random fields to model such behaviors [62].

Other researchers focused on identifying novel detection techniques that can be generalized across domains. They also tried to overcome the main obstacle in opinion fraud detection, that is the lack of ground trust information by employing unsupervised classification models and co-training algorithms with unlabelled data. To this direction, Akoglu et al. [63] proposed FRAUDEAGLE, a framework where fake review detection is modeled as a network classification task on a signed bipartite network containing three type of nodes, namely users, reviews and products. Li et al. used collective Positive-Unlabeled Learning (PU learning) trained on language independent features [55]. For a systematic review of the opinion fraud detection techniques the reader should refer to the work of Heydari et al. [10].

4.7. Discussion

Having evaluated all the above, let us analyze our initial observations of employment scam. To begin with, one can immediately grasp that employment scam detection is a non-trivial, primarily text-based problem that is closely affiliated with the aforementioned problems, but still presents several peculiarities. Most of them derive from the limited context surrounding a job ad, the brief user interaction with the ATS, and most importantly the fact that the malicious content aims by definition to be as indistinguishable as possible from the legitimate one.

As a matter of fact, employment scam lacks strong contextual information. Furthermore, the activity of the composer of a post within an ATS through time is limited, that is, the user may generate a single advertisement, broadcast it and then not further interact with the ATS. In some cases, assailants impersonate existing businesses or recruiting agencies, which makes it harder to deduce the real origin of the job posting. On the contrary, in problems such as trolling or cyber bullying detection, the analyst is able to compose additional contextual and temporal information, regarding the reputation of the misbehaving user, their sequence of actions, and their online footprint mined from multiple open social platforms.

At the same time, ATS are offered as web applications over HTTP, which typically do not entail any dedicated network communication protocol as for example in email spam. As in phishing or wikipedia vandalism, this fact alone makes it impossible to rely on multiple protocol layers for additional indications. As for the application layer, structural anomalies (e.g., invalid HTML markup or CSS rules), visual contrivances, or incomplete company profiles are in most cases products of low-skilled practitioners and serious attackers can easily circumvent them. Moreover, information such as the location of a job or uploading the corporate logo are often neglected even by expert users.

As with opinion fraud detection discussed in Section 4.6, relying just on the raw content often proves to be insufficient. Added to that, our experimentation presented in Section 6 also confirmed that unilateral classifiers will mislabel at least one out of ten job ads. More precisely, opinion fraud heavily relies on detecting the outliers among a large number of available reviews about the same product on the same or similar websites. In other words, the cardinality and the coherence of legitimate reviews are of the essence, whereas in employment scam maintaining a consistent hiring history for legitimate companies is neither straightforward nor practical. Lastly, it is questionable whether alternate approaches such as sentiment analysis could be effectively applied to employment scam, as compared to biased reviews trying to hype or defame products or businesses, the content of a job ad is usually written in neutral language.

In summary, Table 1 presents the feature categories used for detecting malicious content in all six relevant problems discussed in Section 4. Employment scam detection features used in Section 6 are also added. In future work, we would like to experiment with more feature categories.

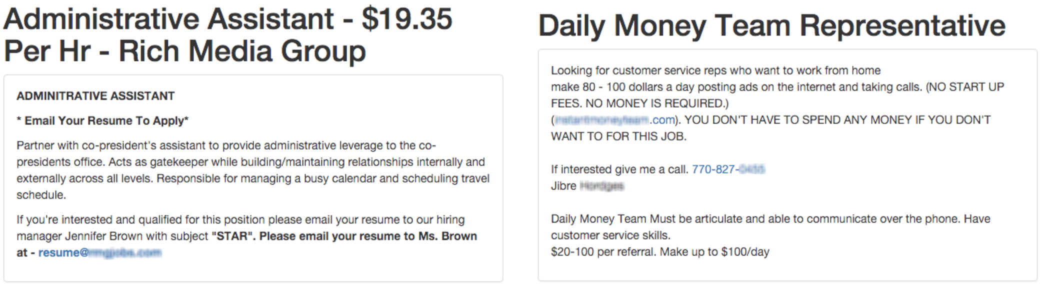
Table 1. Summary of feature categories used by classification algorithms in related problems and employment scam.

Problem	Contextual	Linguistic	Metadata	Protocol	User History	Social Footprint	Dependency Graphs	Revisions
Spamming	♦	♦		♦				
Phishing	♦	♦	♦					
Vandalism	♦	♦	♦					♦
Cyberbullying	♦				♦	♦		
Trolling	♦				♦	♦		
Opinion fraud	♦	♦	♦		♦		♦	
Employment scam	♦	♦	♦					

5. Dataset Description

In our effort to provide a clear picture of the problem to the research community we decided to release a set of relevant data. The Employment Scam Aegean Dataset (EMSCAD) [11] contains real-life job ads posted by Workable [18]. We anticipate that the EMSCAD dataset will act as a valuable testbed for future researchers while developing and testing robust job fraud detection systems.

EMSCAD contains 17,014 legitimate and 866 fraudulent job ads (17,880 in total) published between 2012 to 2014. All the entries were manually annotated by specialized Workable employees. The annotation process pertained to out-of-band quality assurance procedures. The criteria for the classification were based on client's suspicious activity on the system, false contact or company information, candidate complaints and periodic meticulous analysis of the clientele. Two characteristic examples of fraudulent jobs are given in Figure 1.



Administrative Assistant - $19.35 Per Hr - Rich Media Group

ADMINITRATIVE ASSISTANT

* Email Your Resume To Apply*

Partner with co-president's assistant to provide administrative leverage to the co-presidents office. Acts as gatekeeper while building/maintaining relationships internally and externally across all levels. Responsible for managing a busy calendar and scheduling travel schedule.

If you're interested and qualified for this position please email your resume to our hiring manager Jennifer Brown with subject **"STAR". Please email your resume to Ms. Brown at - resume@**[redacted]

Daily Money Team Representative

Looking for customer service reps who want to work from home
make 80 - 100 dollars a day posting ads on the internet and taking calls. (NO START UP FEES. NO MONEY IS REQUIRED.)
([redacted].com). YOU DON'T HAVE TO SPEND ANY MONEY IF YOU DON'T WANT TO FOR THIS JOB.

If interested give me a call. 770-827-[redacted]
Jibre [redacted]

Daily Money Team Must be articulate and able to communicate over the phone. Have customer service skills.
$20-100 per referral. Make up to $100/day

Figure 1. Examples of fraudulent job ads.

Each record in the dataset is represented as a set of structured and unstructured data. It is formally described by a set of fields $F = \{F1, \ldots, Fn\}_{n=16}$, and a binary class field $C\{+,-\}$ indicating whether the current job ad entry is fraudulent or not. Fields can be of four types, namely string as in the job title, HTML fragment like the job description, binary such as the telecommuting flag, and nominal as in the employment type (e.g., full-time, part-time). The detailed list of field types is displayed in Table 2.

Table 2. Detailed list of fields types in the dataset.

Type	Name	Description
String	Title Location Department Salary range	The title of the job ad entry. Geographical location of the job ad. Internal corporate department (e.g., sales). Indicative salary range (e.g., $50,000-$60,000)
HTML fragment	Company profile Description Requirements Benefits	A brief company description. The details description of the job ad. Enlisted requirements for the job opening. Enlisted offered benefits by the employer.
Binary	Telecommuting Company logo Questions Fraudulent	True for telecommuting positions. True if company logo is present. True if screening questions are present. Classification attribute.
Nominal	Employment type Required experience Required education Industry Function	Full-type, Part-time, Contract, etc. Executive, Entry level, Intern, etc. Doctorate, Master's Degree, Bachelor, etc. Automotive, IT, Health care, Real estate, etc. Consulting, Engineering, Research, Sales, etc.

The original dataset is highly unbalanced. Furthermore, it contains duplicates and entries with blank fields due to the fact that fraudsters can quickly and repeatedly try to post the same job ad in identical or different locations. As a result, for our experimentation, we created a balanced corpus of 450 legitimate and 450 fraudulent job ads by randomly selecting among the entries that contained significant information in most fields for both classes and by skipping duplicates. Then, we trained two different classification models as presented in Sections 6.1 and 6.3.

At this point we must underline that some entries in the full dataset may have been misclassified. For example, fraudulent entries may have managed to slip away from the manual annotation process and were thus misclassified as legitimate or on the other hand legitimate entries may have been marked as fraudulent due to an error in judgement. Overall, we expect their number to be insignificant.

6. Analysis and Experimentation

In order to gain better insight into the dataset and provide a baseline to the research community, we subjected our balanced dataset to a multistep experiment. First off, we sanitized all entries and filtered out any unexpected non-English words by identifying non-ascii character sequences in texts using regular expression pattern matching. Then, we've removed standard English stop-words such as "the, that, with, etc..." using the Apache Lucene's StopAnalyzer stop-words list [64].

Afterwards, we used the long-established bag of words modelling, we trained six popular WEKA classifiers [65] and we evaluated their performance (Section 6.1). At a next step (Section 6.2), we performed an empirical analysis on the balanced dataset and we generated a preliminary ruleset. The ruleset was then converted into a binary feature vector that was tested against the same WEKA's classifiers (Section 6.3). Finally we compared the results.

6.1. Bag of Words Modeling

The first experiment consists of the bag of words (bow) modeling of the job description, benefits, requirements and company profile HTML fields shown in Table 2. Before feeding our data to six classifiers, namely ZeroR, OneR , Naives Bayes, J48 decision trees, random forest and logistic regression (LR), we applied stopword filtering excluding most common English parts of speech such as articles and propositions. For each run, the corpus was split into training and cross-validation subsets using the k-fold cross-validation strategy (k = 10). The results are displayed in Tables 3 and 4.

Table 3. Confusion matrices of the six classifiers for the bow model.

Legitimate	Fraudulent	Classified as
(**a**) ZeroR		
450	0	Legitimate
450	0	Fraudulent
(**b**) Logistic regression		
304	146	Legitimate
59	391	Fraudulent
(**c**) OneR single rule		
323	127	Legitimate
77	373	Fraudulent
(**d**) J48 decision trees		
379	71	Legitimate
66	384	Fraudulent
(**e**) Naive Bayes		
416	34	Legitimate
89	361	Fraudulent
(**f**) Random forest for 100 trees		
424	26	Legitimate
53	397	Fraudulent

As shown, the random forest classifier had the highest precision (0.914) and recall (0.912). Naive Bayes and J48 decision trees followed, both achieving similar F-measures of 0.863 and 0.848 accordingly. Logistic regression performed poorly and its training time proved to be about six times slower than J48 even on a small dataset.

Although the ordinary random forest classifier showed promising results, it is important to emphasize that as described in Section 5 the preliminary balanced corpus is curated and its size is small in order to rush to firm conclusions.

Table 4. Classification evaluation of the bag of words model.

Algorithm	Correctly Classified %	Incorrectly Classified %	TP Rate	FP Rate	Precision	Recall	F-measure	ROC Area	Time (s)
ZeroR	50.000	50.000	0.500	0.500	0.250	0.500	0.333	0.500	0.000
Logistic regression	77.222	22.778	0.772	0.228	0.783	0.772	0.770	0.810	30.860
OneR	77.333	22.667	0.773	0.227	0.777	0.773	0.773	0.773	0.270
J48	84.778	15.222	0.848	0.152	0.848	0.848	0.848	0.848	5.620
Naive Bayes	86.333	13.667	0.863	0.137	0.869	0.863	0.863	0.901	0.620
Random forest	91.222	8.778	0.912	0.088	0.914	0.912	0.912	0.970	4.040

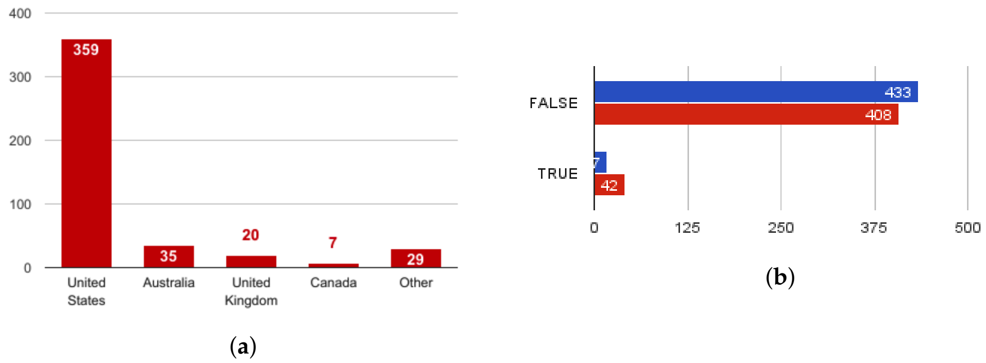
6.2. Empirical Analysis

The goal of the second step was to build a preliminary ruleset consisting of contextual, linguistic and metadata features that derive from statistical observations and empirical evaluation of the balanced dataset. Those features are summarized in Table 1 and are presented in detail in the following sections. In the subsequent diagrams, legitimate job ads are displayed in blue, whereas fraudulent ones in red. Although it can be argued that the following rules are specific to EMSCAD dataset, it is an interesting topic of future work to prove whether or not these rules apply in general.

6.2.1. Geography

As illustrated in Figure 2a, the dataset indicates the vast majority of scammers (86.7%) were published in USA and Australia in contrast to European countries where employment scam is less frequent. It is important to point out though, that EMSCAD is location biased as it only contains entries from English spoken countries.

Moreover, a good indicator of a fraudulent posting is whether it advertises a telecommuting (work from home) position. As observed from Figure 2b, about 10% of the fraudulent postings in the chosen sample contain this characteristic. The predictive power of this feature is stronger if one takes into account that the amount malicious postings that contain the characteristic is over two times greater than the corresponding benign ones.



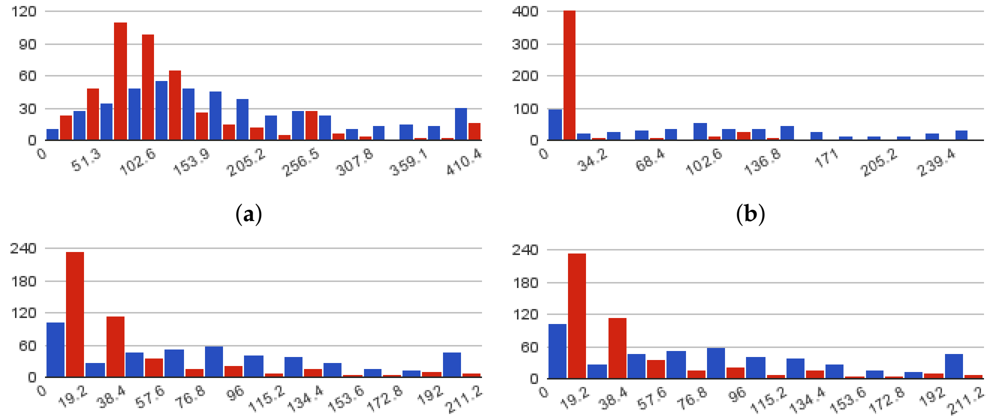
(**b**)

(**a**)

Figure 2. Geographical distribution. (**a**) Top four counties with fraudulent job postings (US, Australia, United Kingdom, Canada); (**b**) Number of telecommuting job ads in fraudulent (red) and non fraudulent (blue) dataset entries. Notice that the amount of malicious job ads that contain the characteristic is over two times greater than the benign one.

6.2.2. Text Analysis

Scammers are less likely to compose a meticulous job ad for the company they claim to be hiring for. As a matter of fact, fraudulent job and company descriptions tend to be short and sketchy. Furthermore, while job requirements and benefits are present in most legitimate job ads, scammers tend to leave them blank. Figure 3 depicts the term counts histograms for both classes for each of the HTML attributes contained in the dataset (see Table 2). To reduce noise in the results, all the HTML tags were stripped before computing the term counts. As shown in Figure 3a, more than half of the fraudulent job postings on the examined sample have job descriptions that do not extend beyond 100 terms. This contradicts the majority of normal postings, only 25% of which are short. According to Figure 3b, 88% of fraudulent postings (around 400), have very short descriptions. The same applies to job requirements and benefits where as observed from Figure 3c,d, around 240 fraudulent job ads have these fields blank.



(**a**)

(**b**)

(**c**)

(**d**)

Figure 3. Term count (TC) histogram computed on the HTML attributes contained in Table 2. (**a**) Job description TC; (**b**) Company description TC; (**c**) Job requirements TC; (**d**) Job benefits TC.

According to human experts, another common method fraudsters adopt, to capture job seekers' attention is to capitalize words in the title or use consecutive punctuation to make it stand out. In EMSCAD, 17% of fraudulent job titles and 15% of legitimate job titles contain capitalized words. Added to that, 3% of fraudulent titles contain consecutive punctuation such as multiple exclamations marks.

Moreover, token sequences such as 'work from home', 'be your own boss', 'income', 'easy money', 'no experience', inside a job ad indicate alarming scam content. Figure 4 illustrates occurences of common spamwords in both classes. For instance, tokens like 'online', 'home' are more likely to be present in fraudulent job ads, whereas words such as 'business', 'opportunity' are more probable to coexist in legitimate contexts.

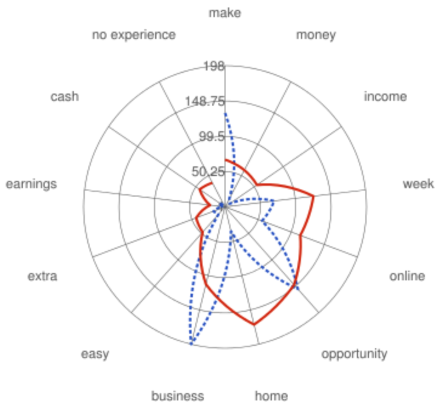


Figure 4. Spamword occurences in legitimate and fraudulent classes.

6.2.3. HTML Analysis

Job ads composed by ATS systems are written in WYSIWYG editors that output HTML markup. HTML markup inspection conducted on EMSCAD resulted in the following classification rules:

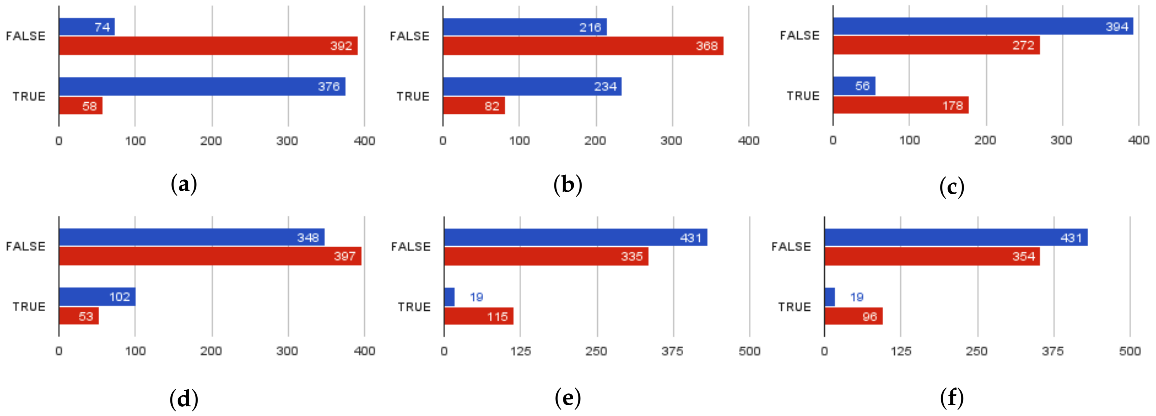
1. **High emphasized word ratio.** Fraudulent job ads have a high ratio of emphasized text, text wrapped in `<b>`, `<strong>`, `<em>`, and heading HTML tags. Table 5 shows emphasized text statistics in core job fields obtained from our dataset. One can easily notice that fraudulent job descriptions have four times the ratio (40.51%) of legitimate ones (9.8%), whereas in fraudulent job requirements the ratio is tripled (26.38%) compared to legitimate job ads (8.71%).
2. **Absence of valid HTML list formatting in requirements or benefits.** Legit job ads enlist job requirements and benefits wrapped in HTML list elements defined by the `<ul>`, `<ol>` and `<li>` tags, whereas fraudulent ones tend to contain raw text lists separated with dashes or asterisks. With reference to our dataset, 55.78% of non-fraudulent job ads have HTML lists in job requirements and 6.89% in benefits in contrast to 28.00% and 3.11% of non fraudulent job ads in the same fields.

Table 5. Emphasized text averages. Aggregated percentages are computed on every record of the dataset, whereas normalized percentages exclude job ads missing the corresponding field.

Field	Aggregated %	Normalized %
(**a**) Legitimate job ads.		
Description	4.9	9.8
Requirements	2.46	8.71
Benefits	2.95	20.42
(**b**) Fraudulent job ads.		
Description	8.64	40.51
Requirements	3.75	26.38
Benefits	3.35	15.70

6.2.4. Binary Analysis

The analysis of the corpus and the information retrieved by the remaining job attributes resulted in seven more rules: (a) opportunistic career pages usually do not have a corporate logo; (b) scammers omit adding screening questions; (c) usually mention salary information even in their title to lure candidates; (d) skip designated job attributes (i.e. industry, function, candidate's education level, and experience level) used for job board categorization; (e) prompt defrauded candidates to apply at external websites, bypassing the ATS pipeline; (f) or force them send their resumes to their personal email addresses directly and (g) address lower educational level. The results acquired for each of the aforementioned rules are displayed in Figure 5. As shown in Figure 5a, the vast majority of fraudulent job postings (392 out of 450 included in the sample) do not have a corporate logo. On the other hand, only a mere of 16% of the legitimate job postings neglect to include a logo. A similar trend applies to screening questions presence displayed in Figure 5b. Moreover, according to Figure 5c,e,f, 176 fraudulent job postings mention salary, 115 redirect applicants to apply to other websites bypassing the ATS, and 96 prompt them to forward their CV to untrustworthy email addresses. Lastly, 30% of fraudulent ads state that higher education is not mandatory.



(**a**) (**b**) (**c**)

(**d**) (**e**) (**f**)

Figure 5. Binary rules from non text attributes and their distribution among the classes. (**a**) Corparate logo; (**b**) Screening questions; (**c**) Salary mentions; (**d**) Fulfilled designated job attributes; (**e**) Prompts external application; (**f**) Contains email links.

6.3. Machine Learning Analysis of the Empirical Ruleset

For the second experiment, we transformed each record of the balanced dataset to a vector of binary features. As shown in Table 6, each feature derives from the rules presented in Section 6.2. For example, corporate logo presence shown in Figure 5a was transformed into the has-no-company-logo binary feature for the new model.

Table 6. Ruleset based binary features.

Category	Name	Description
Linguistic	contains-spamwords has-consecutive-punctuation contains-money-in-title contains-money-in-description	Set if a spamword such as "easy money" is present in job text. Set when consecutive punctuation such as "!!!" is spotted. Set when money symbols are detected in job title. Set when money symbols are detected in job description.
Contextual	has-no-company-profile has-short-company-profile has-no-long-company-profile has-short-description has-short-requirements contains-email-link prompts-for-external-application addresses-lower-education has-incomplete-extra-attributes	Set when company profile field is empty. Set if company profile is less than 10 words. Set if company profile is not short and is less than 100 words. Set if job description is less than 10 words. Set if job requirements are less than 10 words. Set when email links are present in job text. Set when phrases such as "apply at" or "send resume" are detected. Set when phrases such as "High School or equivalent" are detected. Set when designated job attributes are empty.
Metadata	located-in-us is-telecommuting has-no-company-logo has-no-questions has-emphasized-description has-emphasized-requirements has-emphasized-benefits has-no-html-lists-in-requirements has-no-html-lists-in-benefits	Set if job is located in the US. Set when the job is marked as a telecommuting job. Set if company logo is missing. Set if screening questions are missing. Set when high emphasized word ratio in job description is higher than 0.5. Set when high emphasized word ratio in job requirements is higher than 0.5. Set when high emphasized word ratio in job benefits is higher than 0.5. Set when job requirements do not contain HTML lists. Set when job benefits do not contain HTML lists.

The new model is more compact and can scale better on large datasets as it requires significantly lower space and computational resources in comparison to a bow model. The model was tested against the same six classifiers. The dataset was partitioned into training and cross-validation subsets using the k-fold cross-validation strategy (k = 10). The results are summarized in Tables 7 and 8.

Table 7. Confusion matrices of the six classifiers for the ruleset model.

Legitimate	Fraudulent	Classified as
(**a**) ZeroR		
450	0	Legitimate
450	0	Fraudulent
(**b**) OneR single rule		
366	84	Legitimate
57	393	Fraudulent
(**c**) Naive Bayes		
386	64	Legitimate
43	407	Fraudulent
(**d**) Logistic regression		
394	56	Legitimate
34	416	Fraudulent
(**e**) J48 decision trees		
397	53	Legitimate
32	418	Fraudulent
(**f**) Random forest for 100 trees		
401	49	Legitimate
36	414	Fraudulent

Table 8. Classification evaluation for the empirical rules model

Algorithm	Correctly Classified %	Incorrectly Classified %	TP Rate	FP Rate	Precision	Recall	F-measure	ROC Area	Time (s)
ZeroR	50.000	50.000	0.500	0.500	0.250	0.500	0.333	0.500	0.010
OneR	84.333	15.667	0.843	0.157	0.845	0.843	0.843	0.843	0.010
Naive Bayes	88.111	11.889	0.881	0.119	0.882	0.881	0.881	0.947	0.040
LR	90.000	10.000	0.900	0.100	0.901	0.900	0.900	0.956	0.130
J48	90.556	9.444	0.906	0.094	0.906	0.906	0.906	0.812	0.912
Random forest	90.556	9.444	0.906	0.094	0.906	0.906	0.906	0.946	0.450

In comparison to bow modeling presented in Section 6.1, the second experiment showed increased performance with higher accuracy for all but one of the tested classifiers. In further detail, all classifiers significantly increased their achieved accuracy by a margin of 2%–13%. Only RF presents a small decrease of 0.5% which could be product of the sample chosen. Also, notice that ZeroR is not a real classifier as it naively assigns labels to the majority class (in this case the first class) and was included to provide a baseline.

In order to evaluate in detail the effectiveness of each feature deriving from the empirical ruleset, we've also performed Pearson's correlation feature analysis using Weka for the random forest classifier. The analysis concluded that company related features such as short or blank company profiles as well as the lack of company logo are the most effective. The absence of questions and the lack of HTML formatting in job fields follow. On the contrary, it is evident that consecutive punctuation and short job descriptions can be found in fraudulent and legitimate job ads. The results are summarized in Figure 6.

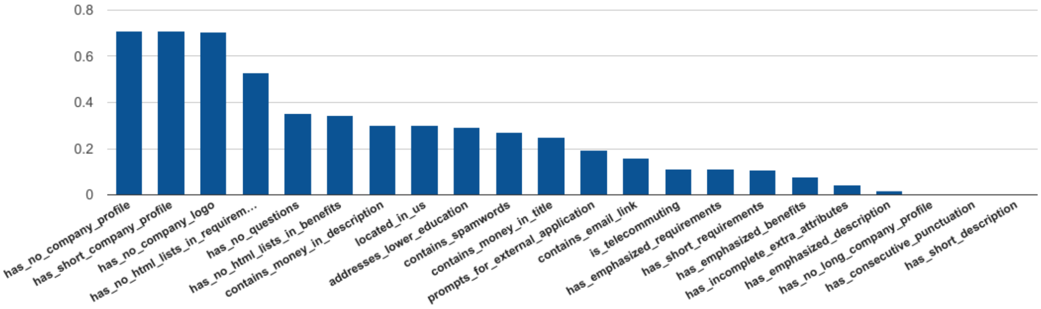


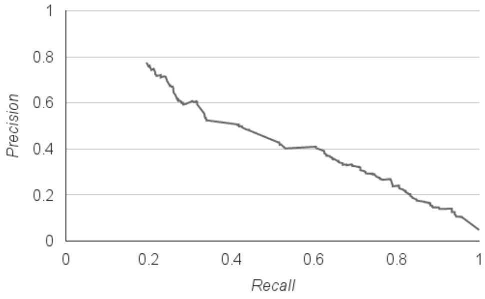
Figure 6. Pearson's correlation feature selection.

6.4. Complete Dataset Evaluation

During the last step, we tested the random forest classifier trained on the empirical ruleset against the complete imbalanced dataset of 17,880 records. The results showed that 89.5% of the entries were classified correctly. In further detail though, the model's precision and recall for the dominant, non-fraudulent class were 0.986 and 0.903 accordingly. On the contrary, for the fraudulent class, precision was 0.282 and recall was 0.751. The confusion matrix for the final experiment is displayed in Figure 7a and the Precision-Recall curve for the fraudulent class is given in Figure 7b. Note that while designing an effective job ad classifier to prevent the dissemination of fraudulent postings, high recall is of the essence.

Legitimate	Fraudulent	Classified as
15.361	1.653	Legitimate
216	650	Fraudulent

(**a**)



(**b**)

Figure 7. Evaluation of the empirical Random Forest classifier on the complete imbalanced dataset. (**a**) Confusion matrix; (**b**) Precision-Recall curve for the fraudulent class.

Taking into account all the above observations, it can be stated that using one-sided classifiers can give us an accuracy of 90% on a fully curated balanced dataset. When it comes to recall, our experiments showed that in a balanced dataset one out of ten fraudulent job ads will evade the detection process. More importantly, as recall drops in the complete imbalanced dataset the amount of undetectable malicious postings increases. To achieve better results and ease out the high volatility of job ad entries, we strongly believe that composite data from multiple domains about users and companies should be incorporated. Our intention is to include these pieces of data in a future feature engineering step.

7. Conclusions and Future Work

In this paper, we analysed the possible aspects of employment scam, an unexplored up to now research field that calls for further investigation, and we introduced EMSCAD, a publicly available dataset containing both real-life legitimate and fraudulent job ads. As shown, ORF is a relative new field of variable severity that can escalate quickly to extensive scam. What is clear from our work is that employment scam bears resemblance to well-studied problems such as email spam, phishing, Wikipedia vandalism, cyber bullying, trolling and opinion fraud, but is characterised by several peculiarities that hinder reliable scam detection through known methodologies, thus requiring composite approaches while designing countermeasures.

We also experimented with the EMSCAD dataset. Preliminary, yet, detailed results show that text mining in conjunction with metadata can provide a preliminary foundation for job scam detection algorithms. We strongly believe that the provided dataset can be used as a part of an automated anti-scam solution by ATS to train classifiers or gain deeper knowledge to the characteristics of the problem. It is also anticipated to trigger and fuel further research efforts to this very interesting, yet still in its infancy area.

In future works, we intend to expand EMSCAD and enrich the ruleset by focusing on user behavior, company and network data as well as user-content-IP collision patterns. Moreover, we would like to employ graph modeling and explore connections between fraudulent job ads, companies, and users. Ultimately, our goal is to propose an applicable employment fraud detection tool for commercial purposes.

Acknowledgments: We would like to thank Workable for allowing us to use public job ads that were published through their system as well as the fraud prevention specialists for the valuable insights.

Author Contributions: The authors contributed equally to this research.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

SaaS	Software as a service
ATS	Applicant Tracking Systems
CRM	Customer Relationship Management
ORF	Online Recruitment Fraud

References

1. Blanzieri, E.; Bryl, A. A survey of learning-based techniques of email spam filtering. *Artif. Intell. Rev.* **2008**, *29*, 63–92.
2. Guzella, T.S.; Caminhas, W.M. A review of machine learning approaches to spam filtering. *Expert Syst. Appl.* **2009**, *36*, 10206–10222.
3. Saadat, N. Survey on spam filtering techniques. *Commun. Netw.* **2011**, *3*, 153–160.
4. Abu-Nimeh, S.; Nappa, D.; Wang, X.; Nair, S. A comparison of machine learning techniques for phishing detection. In Proceedings of the Anti-Phishing Working Groups 2nd Annual eCrime Researchers Summit, Pittsburgh, PA, USA, 4–5 October 2007; ACM: New York, NY, USA, 2007; pp. 60–69.
5. Potthast, M.; Stein, B.; Gerling, R. Automatic vandalism detection in Wikipedia. In *Advances in Information Retrieval*; Springer: Berlin/Heidelberg, Germany, 2008; pp. 663–668.
6. Potthast, M. Crowdsourcing a wikipedia vandalism corpus. In Proceedings of the 33rd International ACM SIGIR Conference on Research and Development in Information Retrieval, Geneva, Switzerland, 19–23 July 2010; ACM: New York, NY, USA, 2010; pp. 789–790.
7. Dinakar, K.; Reichart, R.; Lieberman, H. Modeling the detection of Textual Cyberbullying. In Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media, Barcelona, Spain, 17–21 July 2011.
8. Chen, Y.; Zhou, Y.; Zhu, S.; Xu, H. Detecting offensive language in social media to protect adolescent online safety. In Proceedings of the 2012 International Conference on Privacy, Security, Risk and Trust and 2012 International Confernece on Social Computing, Amsterdam, The Netherlands, 3–5 September 2012.
9. Potha, N.; Maragoudakis, M. Cyberbullying Detection using Time Series Modeling. In Proceedings of the 2014 IEEE International Conference on Data Mining Workshop (ICDMW), Dallas, TX, USA, 7–10 December 2013; pp. 373–382.
10. Heydari, A.; ali Tavakoli, M.; Salim, N.; Heydari, Z. Detection of review spam: A survey. *Expert Syst. Appl.* **2015**, *42*, 3634–3642.

11. Laboratory of Information and Communication Systems, University of the Aegean, Samos, Greece. EMSCAD Employment Scam Aegean Dataset, 2016. Available online: http://icsdweb.aegean.gr/emscad (accessed on 22 February 2017)
12. Peggs, M. Applicant Tracking Systems Solved. 2015. Available online: http://www.michaelpeggs.com/applicant-tracking-systems-solved (accessed on 14 March 2015).
13. Bloomberg. There Are Now More Than Five Million Job Openings in America. 2015. Available online: http://www.bloomberg.com/news/articles/2015-02-10/job-openings-in-u-s-rose-by-181-000-in-december-to-5-03-million (accessed on 11 February 2015).
14. Jobvite. Social recruiting survey for 2014. 2014. Available online: http://web.jobvite.com/rs/jobvite/images/2014%20Job%20Seeker%20Survey.pdf (accessed on 15 March 2015).
15. Vidros, S.; Kolias, C.; Kambourakis, G. Online recruitment services: Another playground for fraudsters. *Comput. Fraud Secur.* **2016**, *2016*, 8–13.
16. Auld, E. Man Posts Fake Job on Craigslist, Gets 600+ Resumes. 2012. Available online: http://chemjobber.blogspot.gr/2012/08/man-posts-fake-job-on-craigslist-gets.html (accessed on 19 March 2015).
17. Australian Bureau of Statistics. Personal Fraud. 2012. Available online: http://www.abs.gov.au/AUSSTATS/abs@.nsf/mf/4528.0 (accessed on 19 March 2015).
18. Workable. Available online: https://www.workable.com (accessed on 19 July 2015).
19. CareerBuilder. Think You Can Spot a Fake Resume? 2015. Available online: http://thehiringsite.careerbuilder.com/2012/05/04/think-you-can-spot-a-fake-resume (accessed on 7 May 2015).
20. Indeed. Job Forum. 2015. Available online: http://www.indeed.com/forum (accessed on 8 May 2015).
21. Mashable. 10 Signes a Job Is a Scam. 2015. Available online: http://mashable.com/2013/10/05/10-signs-a-job-is-a-scam (accessed on 8 May 2015).
22. Monster.com. Money-Laundering and Reshipping Scams. 2016. Available online: http://inside.monster.com/money-laundering/inside2.aspx (accessed on 22 November 2016)
23. Malwarebytes. Money Mules, If It Looks Too Good to Be True... 2013. Available online: https://blog.malwarebytes.com/threat-analysis/2013/10/money-mules-if-it-looks-too-good-to-be-true/ (accessed on 28 October 2013).
24. Straus, R.R. Fake Job Offer Scam Dupes Thousands into Laundering Money for Criminal Gangs. 2013. Available online: http://www.thisismoney.co.uk/money/news/article-2284737/Fake-job-offer-scam-dupes-thousands-laundering-money-criminal-gangs.html (accessed on 13 April 2015).
25. Androutsopoulos, I.; Koutsias, J.; Chandrinos, K.V.; Spyropoulos, C.D. An Experimental Comparison of Naive Bayesian and Keyword-based Anti-spam Filtering with Personal e-Mail Messages. In Proceedings of the 23rd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, Athens, Greece, 24–28 July 2000; ACM: New York, NY, USA, 2000; pp. 160–167.
26. Agrawal, B.; Kumar, N.; Molle, M. Controlling spam emails at the routers. In Proceedings of the ICC 2005—2005 IEEE International Conference on Communications, Seoul, Korea, 15–19 May 2005; Volume 3, pp. 1588–1592.
27. SPF Council. Sender Policy Framework. 2015. Available online: http://www.openspf.org (accessed on 3 May 2015).
28. Schiavone, V.; Brussin, D.; Koenig, J.; Cobb, S.; Everett-Church, R. *Trusted Email Open Standard*; White Paper, May 2003; Copyright ePrivacy Group: Malvern, PA, USA, 2003.
29. Microsoft corporation. Sender ID. 2015. Available online: http://www.microsoft.com/mscorp/safety/technologies/senderid/default.mspx (accessed on 3 May 2015).
30. Kuipers, B.J.; Liu, A.X.; Gautam, A.; Gouda, M.G. Zmail: Zero-sum free market control of spam. In Proceedings of the 25th IEEE International Conference on Distributed Computing Systems Workshops, Columbus, OH, USA, 6–10 June 2005; pp. 20–26.
31. Leiba, B.; Ossher, J.; Rajan, V.; Segal, R.; Wegman, M.N. SMTP Path Analysis. In Proceedings of the CEAS 2005 Second Conference on Email and Anti-Spam, Stanford, CA, USA, 21–22 July 2005.
32. Oscar, P.; Roychowdbury, V. Leveraging social networks to fight spam. *IEEE Comput.* **2005**, *38*, 61–68.
33. Pantel, P.; Lin, D.; others. Spamcop: A spam classification & organization program. In Proceedings of the AAAI-98 Workshop on Learning for Text Categorization, Madison, WI, USA, 26–27 July 1998; pp. 95–98.
34. Graham, P. A Plan for Spam. Available online: http://www.paulgraham.com/spam.html (accessed on 17 March 2016).

35. Androutsopoulos, I.; Paliouras, G.; Michelakis, E. *Learning to Filter Unsolicited Commercial E-mail*; National Center for Scientific Research "DEMOKRITOS": Athens, Greece, 2004.
36. Kanaris, I.; Kanaris, K.; Houvardas, I.; Stamatatos, E. Words versus character n-grams for anti-spam filtering. *Int. J. Artif. Intell. Tools* **2007**, *16*, 1047–1067.
37. Çıltık, A.; Güngör, T. Time-efficient spam e-mail filtering using n-gram models. *Pattern Recogn. Lett.* **2008**, *29*, 19–33.
38. Drucker, H.; Wu, S.; Vapnik, V.N. Support vector machines for spam categorization. *IEEE Trans. Neural Netw.* **1999**, *10*, 1048–1054.
39. Sculley, D.; Wachman, G.M. Relaxed online SVMs for spam filtering. In Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, Amsterdam, The Netherlands, 23–27 July 2007; ACM: New York, NY, USA, 2007; pp. 415–422.
40. Yeh, C.Y.; Wu, C.H.; Doong, S.H. Effective spam classification based on meta-heuristics. In Proceedings of the 2005 IEEE International Conference on Systems, Man and Cybernetics, Waikoloa, HI, USA, 10–12 October 2005; Volume 4, pp. 3872–3877.
41. Hershkop, S. Behavior-Based Email Analysis with Application to Spam Detection. Ph.D. Thesis, Columbia University, New York, NY, USA, 2006.
42. Bratko, A.; Filipič, B.; Cormack, G.V.; Lynam, T.R.; Zupan, B. Spam filtering using statistical data compression models. *J. Mach. Learn. Res.* **2006**, *7*, 2673–2698.
43. Akhawe, D.; Felt, A.P. Alice in Warningland: A Large-Scale Field Study of Browser Security Warning Effectiveness. In Proceedings of the 22nd USENIX conference on Security Washington, DC, USA, 14–16 August 2013; pp. 257–272.
44. Zhang, Y.; Hong, J.I.; Cranor, L.F. Cantina: A content-based approach to detecting phishing web sites. In Proceedings of the 16th international conference on World Wide Web, Banff, AB, Canada, 8–12 May 2007; ACM: New York, NY, USA, 2007; pp. 639–648.
45. Wenyin, L.; Huang, G.; Xiaoyue, L.; Min, Z.; Deng, X. Detection of phishing webpages based on visual similarity. In Proceedings of the Special Interest Tracks and Posters of the 14th International Conference on World Wide Web, Chiba, Japan, 10–14 May 2005; ACM: New York, NY, USA, 2005; pp. 1060–1061.
46. Wang, W.Y.; McKeown, K.R. Got you!: Automatic vandalism detection in Wikipedia with web-based shallow syntactic-semantic modeling. In Proceedings of the 23rd International Conference on Computational Linguistics, Beijing, China, 23–27 August 2010; Association for Computational Linguistics: Stroudsburg, PA, USA, 2010; pp. 1146–1154.
47. Chin, S.C.; Street, W.N.; Srinivasan, P.; Eichmann, D. Detecting Wikipedia vandalism with active learning and statistical language models. In Proceedings of the 4th Workshop on Information Credibility, Raleigh, NC, USA, 26–30 April 2010; ACM: New York, NY, USA, 2010; pp. 3–10.
48. Harpalani, M.; Hart, M.; Singh, S.; Johnson, R.; Choi, Y. Language of vandalism: Improving Wikipedia vandalism detection via stylometric analysis. In Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies: Short Papers-Volume 2, Portland, Oregon, 19–24 June 2011; Association for Computational Linguistics: Stroudsburg, PA, USA, 2011; pp. 83–88.
49. West, A.G.; Kannan, S.; Lee, I. Detecting Wikipedia vandalism via spatio-temporal analysis of revision metadata? In Proceedings of the Third European Workshop on System Security, Paris, France, 13–16 April 2010; ACM: New York, NY, USA, 2010; pp. 22–28.
50. Dadvar, M.; De Jong, F. Cyberbullying detection: A step toward a safer Internet yard. In Proceedings of the 21st International Conference Companion on World Wide Web, Lyon, France, 16–20 April 2012; ACM: New York, NY, USA, 2012; pp. 121–126.
51. Dadvar, M.; de Jong, F.; Ordelman, R.; Trieschnigg, R. Improved cyberbullying detection using gender information. In Proceedings of the Twelfth Dutch-Belgian Information Retrieval Workshop, DIR 2012, Ghent, Belgium, 23–24 February 2012.
52. Dadvar, M.; Trieschnigg, D.; Ordelman, R.; de Jong, F. Improving cyberbullying detection with user context. In *Advances in Information Retrieval*; Springer: Berlin/Heidelberg, Germany, 2013; pp. 693–696.
53. Cheng, J.; Danescu-Niculescu-Mizil, C.; Leskovec, J. Antisocial Behavior in Online Discussion Communities. *arXiv* **2015**, arXiv:1504.00680.
54. Santos, I.; De-La-PeñA-Sordo, J.; Pastor-López, I.; Galán-García, P.; Bringas, P.G. Automatic categorisation of comments in social news websites. *Expert Syst. Appl.* **2012**, *39*, 13417–13425.

55. Li, H.; Chen, Z.; Liu, B.; Wei, X.; Shao, J. Spotting fake reviews via collective positive-unlabeled learning. In Proceedings of the 2014 IEEE International Conference on Data Mining (ICDM), Shenzhen, China, 14–17 December 2014; pp. 899–904.
56. Mukherjee, A.; Venkataraman, V.; Liu, B.; Glance, N.S. What yelp fake review filter might be doing? In Proceedings of the Seventh International AAAI Conference on Weblogs and Social Media, Cambridge, MA, USA, 8–11 July 2013.
57. Ott, M.; Choi, Y.; Cardie, C.; Hancock, J.T. Finding deceptive opinion spam by any stretch of the imagination. In Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies-Volume 1, Portland, Oregon, 19–24 June 2011; Association for Computational Linguistics: Stroudsburg, PA, USA, 2011; pp. 309–319.
58. Ott, M.; Cardie, C.; Hancock, J. Estimating the prevalence of deception in online review communities. InProceedings of the 21st international conference on World Wide Web, Lyon, France, 16–20 April 2012; ACM: New York, NY, USA, 2012; pp. 201–210.
59. Banerjee, S.; Chua, A.Y. Applauses in hotel reviews: Genuine or deceptive? In Proceedings of the Science and Information Conference (SAI), London, UK, 27–29 August 2014; pp. 938–942.
60. Feng, S.; Banerjee, R.; Choi, Y. Syntactic stylometry for deception detection. In Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Short Papers-Volume 2, Jeju Island, Korea, 8–14 July 2012; Association for Computational Linguistics: Stroudsburg, PA, USA, 2012; pp. 171–175.
61. Li, H.; Chen, Z.; Mukherjee, A.; Liu, B.; Shao, J. Analyzing and Detecting Opinion Spam on a Large-scale Dataset via Temporal and Spatial Patterns. In Proceedings of the Ninth International AAAI Conference on Web and Social Media, Oxford, UK, 26–29 May 2015.
62. Fei, G.; Mukherjee, A.; Liu, B.; Hsu, M.; Castellanos, M.; Ghosh, R. Exploiting Burstiness in Reviews for Review Spammer Detection. *ICWSM* **2013**, *13*, 175–184.
63. Akoglu, L.; Chandy, R.; Faloutsos, C. Opinion Fraud Detection in Online Reviews by Network Effects. *ICWSM* **2013**, *13*, 2–11.
64. Apache Lucene. 2016. Available online: http://lucene.apache.org/core/ (accessed on 13 March 2016).
65. The university of Waikato. Weka. 2015. Available online: http://www.cs.waikato.ac.nz/ml/weka (accessed on 17 June 2015).