University of Kent

Kent Academic Repository

Hernandez-Castro, Julio C. and Roberts, David L. (2015) *Automatic detection of potentially illegal online sales of elephant ivory via data mining.* PeerJ Computer Science, 1:e10 . pp. 1-11. ISSN 2376-5992.

Downloaded from

https://kar.kent.ac.uk/49885/ The University of Kent's Academic Repository KAR

The version of record is available from

https://doi.org/10.7717/peerj-cs.10

This document version

Publisher pdf

DOI for this version

Licence for this version

CC BY (Attribution)

Additional information

Versions of research works

Versions of Record

If this version is the version of record, it is the same as the published version available on the publisher's web site. Cite as the published version.

Author Accepted Manuscripts

If this document is identified as the Author Accepted Manuscript it is the version after peer review but before type setting, copy editing or publisher branding. Cite as Surname, Initial. (Year) 'Title of article'. To be published in *Title of Journal* , Volume and issue numbers [peer-reviewed accepted version]. Available at: DOI or URL (Accessed: date).

Enquiries

If you have questions about this document contact ResearchSupport@kent.ac.uk. Please include the URL of the record in KAR. If you believe that your, or a third party's rights have been compromised through this document please see our Take Down policy (available from https://www.kent.ac.uk/guides/kar-the-kent-academic-repository#policies).

PeerJ Computer Science

Automatic detection of potentially illegal online sales of elephant ivory via data mining

Julio Hernandez-Castro[1,2] and David L. Roberts[1,3]

[1] Interdisciplinary Centre for Cyber Security Research, School of Computing, University of Kent, Canterbury, Kent, United Kingdom
[2] School of Computing, University of Kent, Canterbury, Kent, United Kingdom
[3] Durrell Institute of Conservation & Ecology, School of Anthropology & Conservation, University of Kent, Canterbury, Kent, United Kingdom

ABSTRACT

In this work, we developed an automated system to detect potentially illegal elephant ivory items for sale on eBay. Two law enforcement experts, with specific knowledge of elephant ivory identification, manually classified items on sale in the Antiques section of eBay UK over an 8 week period. This set the "Gold Standard" that we aim to emulate using data-mining. We achieved close to 93% accuracy with less data than the experts, as we relied entirely on metadata, but did not employ item descriptions or associated images, thus proving the potential and generality of our approach. The reported accuracy may be improved with the addition of text mining techniques for the analysis of the item description, and by applying image classification for the detection of Schreger lines, indicative of elephant ivory. However, any solution relying on images or text description could not be employed on other wildlife illegal markets where pictures can be missing or misleading and text absent (e.g., Instagram). In our setting, we gave human experts all available information while only using minimal information for our analysis. Despite this, we succeeded at achieving a very high accuracy. This work is an important first step in speeding up the laborious, tedious and expensive task of expert discovery of illegal trade over the internet. It will also allow for faster reporting to law enforcement and better accountability. We hope this will also contribute to reducing poaching, by making this illegal trade harder and riskier for those involved.

Subjects Data Mining and Machine Learning
Keywords eBay, Elephas, Loxodonta, Wildlife trade, Internet, Machine learning

INTRODUCTION

The illegal trade in wildlife is fourth only to narcotics, human trafficking and counterfeiting in terms of transnational organised crimes, with an estimated value of $19–26.5 billion per year (*Haken, 2011*). The illegal trade of wildlife is of concern as it can lead to the extinction of targeted species, incidental impacts through "by-catch" on non-target species (*Broad, Mulliken & Roe, 2001*) and fuels the growth of organised crime syndicates

Submitted 17 April 2015
Accepted 5 July 2015
Published 29 July 2015

Corresponding author
Julio Hernandez-Castro, jch27@kent.ac.uk

Academic editor
Elena Marchiori

Additional Information and Declarations can be found on page 9

DOI **10.7717/peerj-cs.10**

Copyright
2015 Hernandez-Castro and Roberts

Distributed under
Creative Commons CC-BY 4.0

OPEN ACCESS

(*Haken, 2011*). In 2013 the UN[1] recognised environmental crime, including illegal wildlife trade, as part of an emerging form of transnational organised crime requiring a greater response by governments. Subsequently, the UK government hosted the London Conference in 2013 where world governments signed the declaration on Illegal Wildlife Trade (https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/281289/london-wildlife-conference-declaration-140213.pdf). The follow-up conference to the London declaration was held in Botswana, in March 2015,[2] where governments were asked to commit to the following framework:

1. Eradicating the market for illegal wildlife products,
2. Ensuring effective legal frameworks and deterrents,
3. Strengthening law enforcement, and
4. Sustainable livelihoods and economic development.

Research in this area is, however, relatively rare. This is in part because the trade in wildlife, both legal and illegal, is very dynamic and influenced by a variety of factors that have important consequences for biodiversity (*Van Balen et al., 2000*; *O'Brien et al., 2003*; *Shepherd & Magnus, 2004*) and people (*Roe et al., 2002*; *Roe, 2008*; *McNeill & Lichtenstein, 2003*).

Of particular concern is the trade in elephant ivory. Overharvesting is the second leading driver of global biodiversity decline and local extinction, including in elephants. It has recently been shown (*Wittemyer et al., 2014*) that current elephant ivory consumption is unsustainable, with elephant populations having reached a tipping point where the rate of mortality exceeds the birth rate.

Illegal online trade in wildlife is particularly of concern as it is difficult to detect and identify due to the significant amount of legal trade. Contributing to this difficulty is that some words, such as ivory, are not specific to the target in question; ivory is a generic name for material from teeth, and also a colour. As a result, much of the current trade detection by law enforcement officers and NGOs is done by hand, and is therefore an extremely time-consuming process. This makes it generally a tedious, slow and very costly task. Limited resources for this important task means that the results of law enforcement officers and NGOs are, both in terms of localizing and prosecuting offenders, inadequate. Detection is largely expert-based, using certain keywords and detecting the presence of Schreger lines; the latter being indicative of elephant ivory.

Here, we present a method that employs a series of data-mining algorithms and compare it to human experts in automatically finding examples of potentially illegal elephant ivory items on sale. The illegal trade in elephant ivory is of particular concern as not only its harvesting requires the individual to be dead, unlike rhino horn, but also because current consumption of elephant ivory is unsustainable.

The rest of the paper is organised as follows: In the "Methods" section, we describe the methodology used to get the data and its preparation and processing, with a discussion on the importance of metadata. In the "Results" section, we briefly present the results and

[1] In a resolution adopted by the Economic and Social Council on 25 July 2013 titled *On the recommendation of the Commission on Crime Prevention and Criminal Justice responses to illicit trafficking in protected species of wild fauna and flora* that can be found at https://www.unodc.org/documents/commissions/CCPCJ/Crime_Resolutions/2010-2019/2013/ECOSOC/Resolution_2013-40.pdf

[2] The Statement of the Kasane Conference on Illegal Wildlife Trade can be found at https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/417231/kasane-statement-150325.pdf

in the "Discussion" section, we discuss them in detail, and displaying some of the insights gathered, after which we end with "Conclusions and Future Works and Bibliography."

METHODS

In this section we present all relevant information of how we acquired the relevant data for our analysis, and why metadata is crucial to our problem.

Data source

Every Friday, and for 8 consecutive weeks (beginning on the 28th March 2014), items posted on eBay UK and listed in the UK were downloaded from the Antiques category based on the search for the keyword "*ivory*." This represented in total 1,159 unique items. Two former law enforcement officers with specialist knowledge in illegal wildlife trade, including ivory, classified the items manually as being elephant ivory, or some other material (e.g., hippopotamus ivory, bone or resin). These were then classified as potential illegal or potential legal, as only when the item is available to an expert to examine and the vendor has been questioned can the legal status of the item be more precisely determined. In essence, we attempted to replicate an investigation into illegal ivory trade, where law enforcement officers would first identify items of interest for further study.

Metadata

It should be noted that in this paper metadata refers to all the information associated with an item offered for sale, except the two that are arguably the most informative for the identification of elephant ivory items, namely the item description and item images. Often used in expert investigations, they are in our view unreliable as there are new platforms, such as Instagram, where there is no item description and image recognition is still not capable of dealing accurately with the problem of detecting ivory in images with a free context and background (i.e., b/n, colour, studio like pics or blurry ones, after colour filters have been applied, different size, etc.), and of varying quality, though this may change quickly due to the recent advances in deep-learning. Further, we believe metadata is the minimum amount of data we can reliably encounter for classification, therefore we limited our main effort to dealing with it exclusively; however, some interesting results with a more encompassing approach are described in the Results section. By doing so, we also tackle the problem with strictly less information than the human experts will use for comparison, thus proving the potential of the technique. In this case, metadata included data such as the postage costs associated with the item, its price, whether it was offered through an auction or a buy-it-now, the number of bids received, the number of reviews and feedback of the vendor, etc. We collected a total of 37 metadata attributes, which were used in the data-mining algorithms.

RESULTS

After cleaning the data received from the experts, we used the Orange (*Demšar, Curk & Erjavec, 2013*) data mining toolbox to apply multiple data mining techniques to it. We divided the problem of whether an elephant ivory item is potentially illegal in two phases,

Table 1 Confusion Matrix for our classifier. Class 1 represents illegal items, of which there are 214. A total of 257 items were labeled as "elephant ivory."

Class	0	1	Total real
0	**901**	44	945
1	38	**176**	214
Total predicted	939	220	1,159

as this approach produced by far the best results (going from an accuracy of 87% to a 93%). The first stage simply tries to decide whether items on sale contained elephant ivory. For this we used the CN2 classifier, which extracts information about the class based on the CN2 induction algorithm (*Clark & Niblett, 1989*), and produces a set of if-then rules. This classifier, apart from achieving the best performance presents another advantage: its results can be easily analysed and interpreted due to its readability. For verifying the generality of the rules inferred, we used 10-fold cross-validation, and achieved an accuracy of around 88%. At the next stage, we employed those rules for predicting a new attribute of each item, and used this new predicted value for aiding the classification of whether it was potentially illegal. With this combined approach, we obtained an overall accuracy of 92.92%. The overall scheme can be seen in Fig. 1. The resulting confusion matrix can be seen in Table 1, and relevant performance scores are shown in Table 2.

It is worthy to point out that, if we were to use a classical 90–10% training–testing random sampling approach instead of the 10-fold cross-validation employed above, the accuracy obtained would have been very similar (92.5%).

DISCUSSION

In this section we present a detailed analysis of the results, covering the confusion matrix and showing some of the most effective insights found.

Confusion matrix

The confusion matrix in Table 1 shows that the true positives and true negatives add up to 1,077 out of 1,159 items. We chose this classifier over similar ones because it minimizes the number of false negatives to 38 (or 3.27%), which is one of our main objectives, as we do not want to miss potentially illegal items. Rather we prefer to err on the side of caution (false positives), misclassifying as potentially illegal something that is not.

Rules for potentially illegal elephant ivory items

A brief description and analysis follows, covering some of the top rules found:

Listing 1: Top rule for discarding potentially illegal items

```
1 if cn2-rules(i) = 0 then
2     if price(i) in (10,386] then
3         i = not-illegal
4      end if
5 end if
```

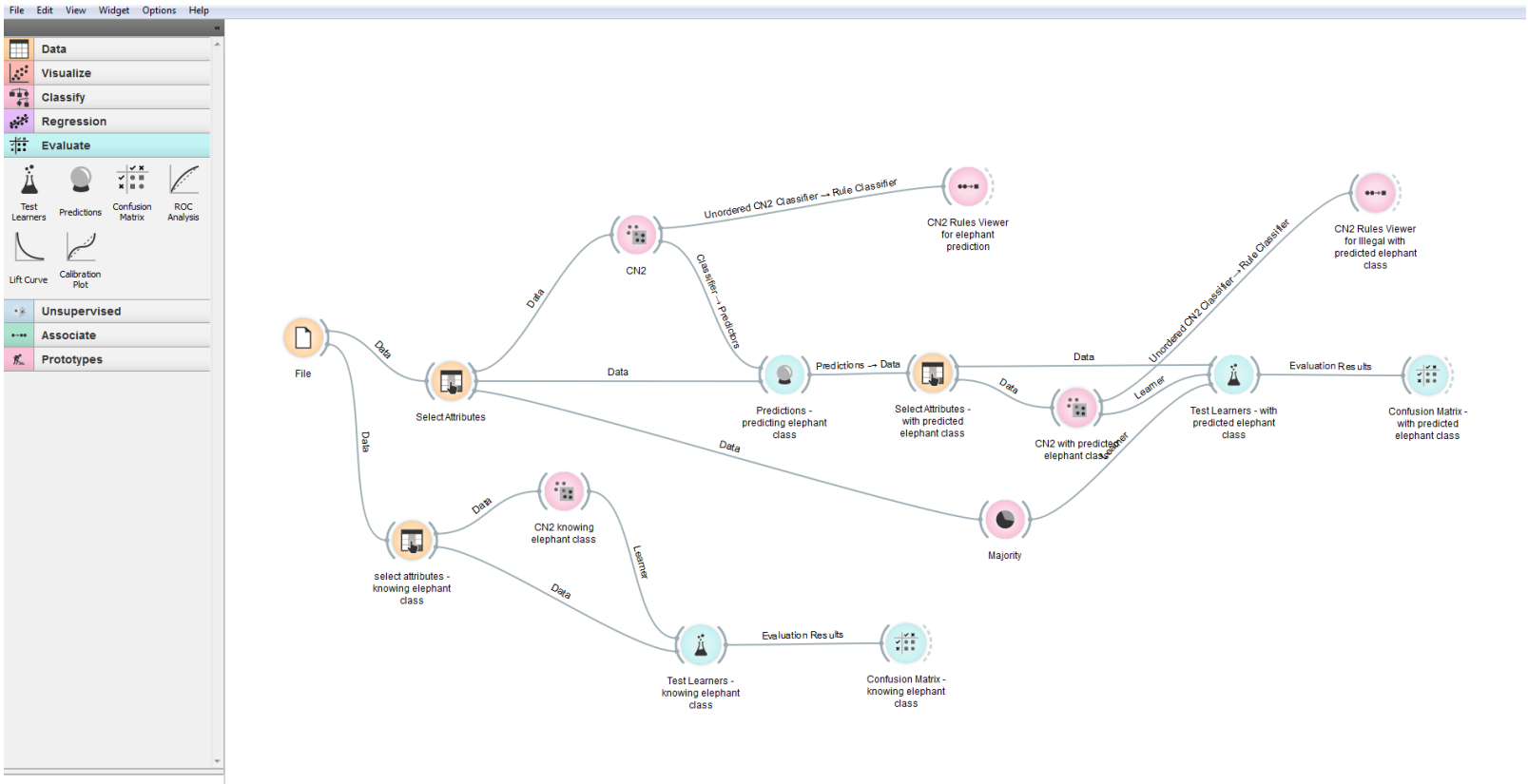


Figure 1 Overall architecture of the classifier. The overall scheme, with its two phases: first we learn about predicting elephant ivory, then we use these results for predicting whether items are potentially illegal.

Table 2 Performance scores for the proposed classifier.

Accuracy	Sensitivity	Specificity	F1	Precision	Recall	Brier
0.9292	0.934	0.8224	0.9565	0.9595	0.9534	0.1182

This rule has an accuracy of 100% and applies to 461 instances that it perfectly classifies as not illegal. On top of it, it is easy to read and interpret, and offers some insights about a criminal's sales practice. It is clear that this rule relies on the previous set of rules used to determine whether the item in question *i* was elephant ivory. It uses this attribute as the main criterion (*if cn2-rules(i)=0* means the previous CN2 classifier best guess is that the class for item *i* is 0, i.e., not elephant ivory). After this, the rule checks for the price in a curious and revealing way. Prices in that range seems to be strongly associated with non-illegal products. This particular rule is extremely helpful in quickly and very accurately discarding many items that will require no further inspection, thus greatly speeding up item processing and contributing to an optimal allocation of expert's time.

Let us now consider a rule aimed at identifying potentially illegal items.

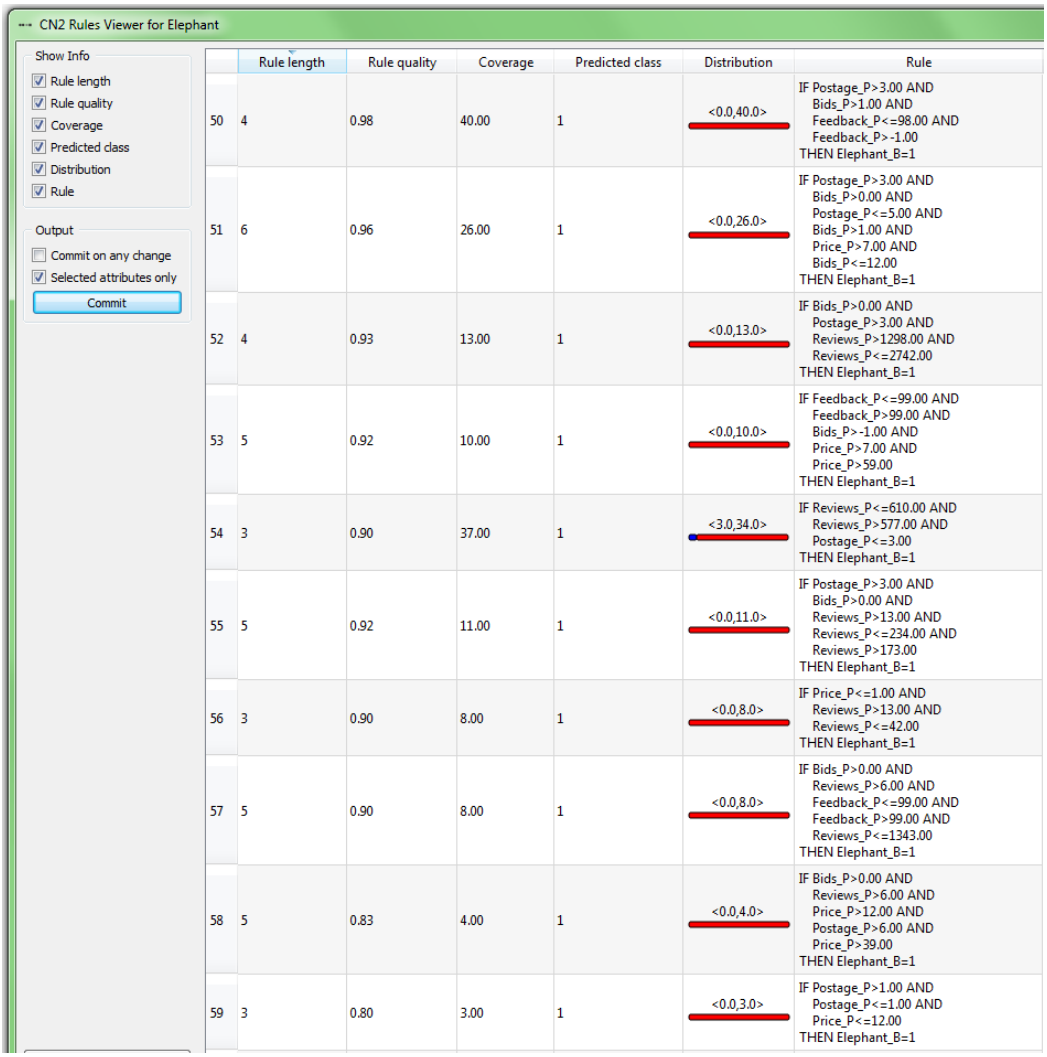
Listing 2: Top rule for finding potentially illegal items

```
1 if cn2-rules(i) = 1 then
2     if reviews(i) in (197,509] then
3         if postage(i) > 4 then
4             i = illegal
5         end if
6     end if
7 end if
```

This rule applies to exactly 57 items that it correctly classifies as potentially illegal. Similarly to the previous case, the main criteria here to identify a potentially illegal item is for it to have been previously classified as elephant ivory (*CN2-rules(i)=1*) and then some other extra conditions. In this case, that the number of reviews is not too low nor too high and that the costs associated with postage are higher than £4.

Rules for elephant ivory items

We were positively surprised by the rules found for this problem. They tended to be quite good and of good generality, although the overall accuracy at around 88% was just fair. What surprised us more was the algorithm's heavy reliance for many of the rules on a relatively unexpected and obscure attribute such as *Bids*. This attribute represents the number of bids a particular item has received, and initially we were not particularly hopeful of its value for the classification problem. Contrary to our expectations, it was used by many of the best rules found. After further analysis, we found this was actually a very clever classification-by-proxy of sorts, where the algorithm basically lets humans (in this case, potential buyers) to show an interest in an item and uses that as an element for the detection of elephant ivory, as this is potentially a scarce and valuable material, thus



	Rule length	Rule quality	Coverage	Predicted class	Distribution	Rule
50	4	0.98	40.00	1	<0.0,40.0>	IF Postage_P>3.00 AND Bids_P>1.00 AND Feedback_P<=98.00 AND Feedback_P>-1.00 THEN Elephant_B=1
51	6	0.96	26.00	1	<0.0,26.0>	IF Postage_P>3.00 AND Bids_P>0.00 AND Postage_P<=5.00 AND Bids_P>1.00 AND Price_P>7.00 AND Bids_P<=12.00 THEN Elephant_B=1
52	4	0.93	13.00	1	<0.0,13.0>	IF Bids_P>0.00 AND Postage_P>3.00 AND Reviews_P>1298.00 AND Reviews_P<=2742.00 THEN Elephant_B=1
53	5	0.92	10.00	1	<0.0,10.0>	IF Feedback_P<=99.00 AND Feedback_P>99.00 AND Bids_P>-1.00 AND Price_P>7.00 AND Price_P>59.00 THEN Elephant_B=1
54	3	0.90	37.00	1	<3.0,34.0>	IF Reviews_P<=610.00 AND Reviews_P>577.00 AND Postage_P<=3.00 THEN Elephant_B=1
55	5	0.92	11.00	1	<0.0,11.0>	IF Postage_P>3.00 AND Bids_P>0.00 AND Reviews_P>13.00 AND Reviews_P<=234.00 AND Reviews_P>173.00 THEN Elephant_B=1
56	3	0.90	8.00	1	<0.0,8.0>	IF Price_P<=1.00 AND Reviews_P>13.00 AND Reviews_P<=42.00 THEN Elephant_B=1
57	5	0.90	8.00	1	<0.0,8.0>	IF Bids_P>0.00 AND Reviews_P>6.00 AND Feedback_P<=99.00 AND Feedback_P>99.00 AND Reviews_P<=1343.00 THEN Elephant_B=1
58	5	0.83	4.00	1	<0.0,4.0>	IF Bids_P>0.00 AND Reviews_P>6.00 AND Price_P>12.00 AND Postage_P>6.00 AND Price_P>39.00 THEN Elephant_B=1
59	3	0.80	3.00	1	<0.0,3.0>	IF Postage_P>1.00 AND Postage_P<=1.00 AND Price_P<=12.00 THEN Elephant_B=1

Figure 2 Elephant ivory identification. Some rules that help spot elephant ivory items on eBay.

generating interest and multiple bids. Letting many humans check the pictures for Schreger lines, decide it is really elephant ivory and bid for the item is a particularly 'lazy' but very smart and effective way of discovering these items. Other attributes such as postage price, merchant feedback, item price and number of reviews were also heavily used by the classification algorithm. An example of other rules found for depicting elephant ivory items are shown in Fig. 2.

CONCLUSIONS AND FUTURE WORKS

In this project, we were able to mimic with great accuracy (close to 93%) human expert classification (the closest we are to a gold standard) of potentially illegal elephant ivory items sold online. We achieved this at a fraction of the cost and at around 10^3 times an expert's speed. Further, we used a worst case scenario, a minimal information approach to prove the generality of our method, which operated with strictly less information than the one used by the experts. By employing only metadata, and discarding text and

Figure 3 Webchat sales. Images displaying elephant ivory for sale on the Chinese Wechat social mobile platform.

images in the item description and pictures, we prove that even by using advanced tactics potentially illegal items will still be located with our method. We also show how the current accuracy of 93%, enough for most practical purposes, can be improved by adding more (non-metadata) info.

Practically speaking, it is arguable that the most useful statistical metric for a high throughput machine learning algorithm is sensitivity. A highly sensitive computational analysis of eBay reports would then enable law enforcement officials to focus on websites that are likely to be advertising illegal products. In this scenario, specificity or accuracy can be considered less important. We showed in Table 2 that this approach also is able to obtain very high sensitivity values.

In addition, our method has the added benefit of being very lightweight in terms of computational requirements, and hence very fast. This is, in our opinion, a major step in simplifying, speeding up and automating what has been until now a very arduous, time consuming and expensive task. In addition, this at the same time will help deter criminals from making easy profits from the sales of elephant ivory products.

In the future, we aim to apply other text data mining techniques to the item description, together with advanced deep-learning algorithms for image classification.

Furthermore, metadata is likely to be very robust to changes in human behaviour and linguistics attempts to code or disguise certain keywords. In many cases, metadata is automatically generated by the selling platform and the trader has very little control over it, making it a reliable source of information not subject to easy manipulation.

There is a large market for ivory products in China, typically traded on Chinese social media (see a sample image in Fig. 3). The texts describing these products are frequently absent or simply describe "raw material" or "dragon pendant," etc., with no key words such as "ivory" ever mentioned. Most of the images do not have close ups, making the detection of Shreger lines very complex or impossible. It is our opinion that these difficult circumstances will become increasingly common in most markets, and that only metadata will be able to reliable help locating illegal items.

Finally, we have reason to believe the same or similar techniques could be used successfully at a number of other online merchants (i.e., Gumtree, Craigslist, CQOUT, eBid, Artfire, etc.) to detect online criminal sales, not only for wildlife trade but for other types (firearms, stolen merchandise, etc.) of illegal items.

ACKNOWLEDGEMENTS

We want to thank Chenyue Ma, a Program Analyst with IFAW in China for discussing some of the characteristics and trends in the Chinese markets, and providing us with other supporting materials.

ADDITIONAL INFORMATION AND DECLARATIONS

Funding

This work was partially funded by the IDEAS Factory program of the University of Kent, NERC Pathfinder grant NE/L00075X/1, and Chester Zoo's Conservation and Research

grant. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Grant Disclosures

The following grant information was disclosed by the authors:
IDEAS Factory program of the University of Kent.
NERC Pathfinder: NE/L00075X/1.
Chester Zoo's Conservation and Research grant.

Competing Interests

Julio Hernandez-Castro has no competing interests. David L. Roberts is an Academic Editor for PeerJ.

Author Contributions

- Julio Hernandez-Castro conceived and designed the experiments, performed the experiments, analyzed the data, wrote the paper, prepared figures and/or tables, performed the computation work, reviewed drafts of the paper.
- David L. Roberts conceived and designed the experiments, performed the experiments, wrote the paper, reviewed drafts of the paper.

REFERENCES

Broad S, Mulliken T, Roe D. 2001. The nature and extent of legal and illegal trade in wildlife. In: Oldfield S, ed. *The trade in wildlife: regulation for conservation*. London: Routledge, 3–22.

Clark P, Niblett T. 1989. The CN2 induction algorithm. *Machine Learning* **3(4)**:261–283.

Demšar J, Curk T, Erjavec A. 2013. Orange: data mining toolbox in python. *Journal of Machine Learning Research* **14(Aug)**:2349–2353.

Haken J. 2011. *Transnational crime in the developing world*. Global financial integrity. *Available at http://www.gfintegrity.org/storage/gfip/documents/reports/transcrime/gfi_transnational_crime_web.pdf*.

McNeill D, Lichtenstein G. 2003. Local conflicts and international compromises: the sustainable use of vicuña in Argentina. *Journal of International Wildlife Law & Policy* **6**:233–253 DOI 10.1080/13880290390437328.

O'Brien S, Emahalala ER, Beard V, Rakotondrainy RM, Reid A, Raharisoa V, Coulson T. 2003. Decline of the Madagascar radiated tortoise Geochelone radiata due to overexploitation. *Oryx* **37**:338–343 DOI 10.1017/S0030605303000590.

Roe D. 2008. *Trading nature: a report, with case studies, on the contribution of wildlife trade management to sustainable livelihoods and the Millennium Development Goals*. Cambridge: TRAFFIC International.

Roe D, Mulliken T, Milledge S, Mremi J, Mosha S, Grieg-Gran M. 2002. Making a killing or making a living. In: *Wildlife trade, trade controls and rural livelihoods*. Cambridge: IIED, London and TRAFFIC.

Shepherd CR, Magnus N. 2004. *Nowhere to hide: the trade in Sumatran tiger*. Petaling Jaya: TRAFFIC Southeast Asia.

Van Balen S, Dirgayusa IWA, Adi Putra IMW, Prins HHT. 2000. Status and distribution of the endemic Bali starling *Leucopsar rothschildi*. *Oryx* **34**:188–197 DOI 10.1017/S0030605300031185.

Wittemyer G, Northrup JM, Blanc J, Douglas-Hamilton I, Omondi P, Burnham KP. 2014. Illegal killing for ivory drives global decline in African elephants. *Proceedings of the National Academy of Sciences of the United States of America* **111(36)**:13117–13121 DOI 10.1073/pnas.1403984111.