A Text-based Deception Detection Model for Cybercrime

A.Mbaziira[1] and J.Jones[1]

[1]George Mason University, Fairfax, VA, 22030

Abstract. Incidents of cybercrime exploiting text-based deception discourse are increasing due to popularity of text messages. We use machine learning and linguistic approaches to detect deception within text messages in cybercriminal networks. We develop cybercrime detection models by web genre. Our contributions are: models trained in scams in social media web genre detect fraud in messages in the email web genre with 60% predictive accuracy; models trained on fraud in the email genre can predict scams in social media web genre with 50% predictive accuracy. The prediction for the email model is promising due to the linguistic variations of cybercriminals in this study. We also demonstrate that cybercrime detection models can be constructed using features from natural language processing and linguistic psychological processes linked to cybercrime.

1 Introduction

Innovations in computer and mobile technology as well as its wide adoption continues to make the Internet ubiquitous and text messaging the most popular channel for communication (Shropshire, 2016). There are different types of channels for text messaging on the Internet as well as variation in linguistic styles used in generating text messages for communication. These variations in writing styles are called web genres and include: web (Mehler, Sharoff, & Santini, 2011), Twitter (Westman & Freund, 2010) and email (Wollman-Bonilla, 2003)

The surge in text messaging, on the one hand, and inherent vulnerabilities in the Internet architecture driving this mode of communication is attracting cybercriminals to exploit victims using this communication medium (McGrath, 2015; Ponemon Institute, 2015; Vergelis, Shcherbakova, Demidova, & Gudkova, 2016). Two of the techniques cybercriminals use to bypass content-filtering systems to launch successful cybercrime campaigns are: changing sequences of lexicons to make messages unique, and exploiting unprotected email clients and messengers on mobile devices since these lack content filters etc (Vergelis et al., 2016).

Cybercrime that involves planning, generating and disseminating malicious content to commit crime is called content-based cybercrime (ITU, 2009). This includes spam, fraud, scams, untruthful online product reviews, child pornography, xenophobia, cyber-terrorism, espionage, etc (Engel, 2015; Hall, 2015). In this paper, we focus on fraud and scams because unlike spam, cybercriminals use deception as a technique for exploiting their victims and also because this problem is not well studied. In this paper, we define cybercrime as content-based crime in the form of scams and fraud where cybercriminals use text messages to plan amongst themselves and exploit their victims. These two uses reflect two types of cybercriminal networks when it comes to text messaging: the first type is cybercriminal-to-cybercriminal while the second type is cybercriminal-to-victim. In a cybercriminal-to-cybercriminal network, cybercriminals send messages to each other to plan and execute crime, while in a cyber-criminal-to-victim network, criminals send malicious and/or deceptive messages to exploit their victims. In this paper, we use the term cybercriminal network to refer to a cybercriminal-to-victim network where cybercriminals send deceptive messages to exploit their victims. This is because deceptive cybercriminal-to-victims messages have more linguistic and research value compared to those in cybercriminal-to-cybercriminal networks which may be coded with phrases and specific to a cybercriminal network.

We argue that with text-based deception discourse, we can develop a model with linguistic features for detecting and analyzing cybercrime. To achieve this, we use computational linguistics in natural language processing and psycholinguistics to identify features for this model. We use computational linguistics to identify lexical and syntactical features for the model. Furthermore, psycholinguistics enables us to identify features whose linguistic processes can be mapped to psychological cybercrime processes.

Table 1 Summary of CL Features

Features	CL Process Linked to Deception and Cybercrime
Quantity of words	Deceptive messages have more amount of lexicons i.e. verbs, modifiers
Lexical diversity	Deceptive messages have fewer ratio of unique words
Expressivity	Deceptive messages have higher frequencies of adjectives and adverbs
Non-immediacy	Deceptive messages have fewer self-references
Sentence complexity	Deceptive messages have less complex sentence complexity i.e., lower average sentence length, average word length and punctuation marks

Our goal is to create a model for detecting deception in messages constituting fraud and scams within cybercriminal networks. We address the following research questions:

1. Can we detect cybercrime in different web genres by detecting deception?
2. What linguistic features indicate deception in text messages?
3. Can we generalize deception detection in text messages?

This paper extends two earlier papers on detecting cybercrime in criminal networks (Abozinadah, Mbaziira, & Jones, 2015; Mbaziira, Abozinadah, & Jones, 2015). This work shows that it is possible to detect cybercrime with 60% accuracy using deception detection based on linguistic features. Our contributions are a model trained on scam in one web genre that can detect scam and fraud messages from another web genre. We also have a model trained on messages with fraud in one web genre that can detect scams in messages of another web genre. Fraud is a serious crime that involves use of deception while scamming, though a deceptive scheme too, is legally a minor crime (Muireann, n.d.; Theoharis, 2016). The remainder of the paper is organized as follows. In Section 2 we discuss related work in text-based cybercrime detection and deception detection. Section 3 describes our methodology and the experimental set up used to collect and analyze data for this paper. Section 4 describes experimental results and analysis, and Section 5 presents the conclusions of the paper.

2 Related Work

Research on deception detection in text messages in general but cybercrime in particular is still limited (Hancock, Curry, Goorha, & Woodworth, 2007; Rowe, n.d.; Zhou, Burgoon, Twitchell, Qin, & Nunamaker, 2004). In this section, we review related work on cybercrime detection, especially linguistic processes and psycholinguistic processes. We also review studies on machine learning, especially text classification in cybercrime.

2.1 Computational Linguistic (CL) Processes for Detecting Cybercrime

Earlier research applied the bag-of-words approach to detect patterns of cybercrime in text-based Computer Mediated Communication (CMC) (Abozinadah et al., 2015; Li, Huang, Yang, & Zhu, 2011; Mbaziira et al., 2015; Mukherjee, Liu, & Glance, 2012). In this approach, individual and combined lexicons from text messages are used as features to detect cybercrime. Although this approach is popular in text classification, some studies reveal that it is not robust enough to detect patterns of cybercrime (Chen, Zhou, Zhu, & Xu, 2012; Reynolds, Kontostathis, & Edwards, 2011). Some research using this approach attempted to boost classifiers with graph-based features from communicating nodes in cybercriminal networks (Abozinadah et al., 2015). However, we avoid this approach because models developed from bag-of-words in one web genre cannot generalize well with data in a different web genre.

We consider another CL process which uses lexical and syntactical features. Lexical features are character-based and word-based features. Character-based features are in the form of frequencies and ratios of characters used within the messages, while word-based syntactic features include frequency of punctuation marks, occurrence of function words, and parts-of-speech (POS) tagging (Afroz, Brennan, & Greenstadt, 2012; Shojaee, Murad, Azman, Sharef, & Nadali, 2013). Table 1 shows lexical and syntactical features which are relevant to our work but which also distinguish our work from other research.

2.1 Psycholinguistic (PL) Processes for Detecting Cybercrime

Cybercrime in text-based communication can be also detected using PL features. This is because linguistic behavior in text-based communication can be mapped to criminal psychological processes. Tools like Linguistic Inquiry and Word Count (LIWC) have been widely used in studying various relationships between psychology and linguistics. (Crawford, Khoshgoftaar, Prusa, Richter, & Najada, 2015; Hancock et al., 2007; Tausczik & Pennebaker, 2010). LWIC supports a number of linguistic processes for various psychological processes; we identify features which are relevant to scams and fraud in text messages. For instance, linguistic features like word quantity, average sentence length, first-person singular and exclusive words have a relationship with psychological processes like talkativeness, cognitive complexity and truthfulness, respectively (Tausczik & Pennebaker, 2010). Table 2 is summary of psycholinguistic features relevant to our work.

2.2 Machine Learning and Cybercrime

There is a growing body of research addressing various text classification problems on cyber-crime (Firte, Lemnaru, & Potolea, 2010; Pearl & Steyvers, 2012; Shojaee et al., 2013, 2013). Some of the popular algorithms for text classification are: Naïve Bayes (NB), Support Vector Machines (SVM) and k-Nearest Neighbor (kNN). We also use these classifiers to analyze and detect cybercrime using text-based deception detection.

NB is applied to a number of text classification problems including detection and analyzing cybercrime (Fette, Sadeh, & Tomasic, 2007; Sommer & Paxson, 2010). NB classifies records by computing posterior probabilities for every class *C* given a document *d* (Tan, Steinbach, & Kumar, 2006). SVM uses a maximal margin hyper-plane to linearly separate instances into two classes (Chang & Lin, 2001). Eager learners, like *kNN*, delay mapping the input data attributes to class labels until that time when the training data is available. *kNN* uses a distance function to determine which instances are closest to the new instance.

The main distinction between our work and that of other research is that we use a text-based deception detection model comprising features from CL and PL processes. The features are used in constructing NB, SVM and kNN text classifiers to detect deception, and hence cybercrime, in text messages.

3 Methodology and Experimental Setup

3.1 Data Description

We collect text messages from two web genres: Facebook and email, because these communication mediums are asynchronous and text messages do not have vast linguistic variations. For the Facebook web genre, we use a publicly leaked dataset of 1036 email addresses of Nigerian cybercriminals that were subscribing to an online data theft service called *PrivateRecovery* (Sarvari, Abozinadah, Mbaziira, & McCoy, 2014). We conducted Facebook look-ups to identify corresponding public profiles linked to each email address. We also collected public data from each of 43,125 friends linked to each Facebook profile.

For the email web genre, we use the Enron email dataset which contains 500,000 emails. This dataset was made public by the Federal Energy Commission during the Enron case (Cohen, 2015). We also collected 89 emails that were part of the evidence used to prosecute Mr. Kenneth Lay, the former Enron Chairman, and Mr. Jeffrey Skilling, the Enron Chief Executive Officer, for securities and wire fraud in the scandal. This evidence was made public by the Department of Justice due to public interest in this case.

The data we use to detect and analyze content-based cyber-crime is real word data and it has issues like: noise, missing values, inconsistency, and redundancy, hence it has to be preprocessed (Larose, 2014). Data from the Facebook (FB) web genre contained: non-ASCII symbols which that were used in expressing emotion, emoticons, phrases in non-English languages like local Nigerian dialects and

pidgin, and accented words in languages like Spanish and French. Data from the email web genre was also preprocessed by removing noise like email headers, email threads, and text-based emoticons. We also use Principal Component Analysis (PCA) and normalization to determine features that provide better variability.

3.2 Feature Selection

Since real world data is imbalanced, we use random sampling to select instances of datasets in each web genre. We then manually labeled all the instances as either truthful or deceptive because we are using supervised learning to construct our cybercrime detection models. Deceptive instances are positive instances that have been manually identified as either scams or fraud while truthful instances are those which are benign. We extract 31 features for training and testing sets in each web genre comprising CL and PL features. We use natural language processing to extract CL features while PL features are extracted from the datasets using Linguistic Inquiry and Word Count (LIWC). LIWC is a text analysis tool for analyzing words with respect to PL processes (Tausczik & Pennebaker, 2010).

3.3 Metrics for Evaluating Classifier Performance

Measures that we adopt to evaluate performance of the classifiers are: precision (P), recall (R), f-measure (F), and the Receiver Operator Characteristic (ROC) Curve. We are addressing a binary classification problem where the classes are either *deceptive* or *truthful* for cybercrime and benign messages respectively. Precision measures the proportion of instances that are actually deceptive in the group which the classifier has declared as deceptive. Recall measures the proportion of actual deceptive messages that are predicted correctly. F measure is a combined measure reflecting the trade-off between precision and recall. A ROC curve is a graph that illustrates the trade-off between the benefits (true positive rate) and costs (false positive rate) of a binary classifier. We also use accuracy to determine the closeness in the predictions to the instances being measured.

4 Experiment Results and Discussion

In this section we present performance results of the three classifiers using the algorithm implementations in WEKA (Frank et al., 2005). We use three classifiers to solve our classification problem: NB, SVM, kNN. For each web genre we construct classifiers as shown in Figure 1. The Enron model represents the email genre with fraud messages while the FB model corresponds to the FB genre with scam messages. Each model is tested on a labelled test set of another web genre. The NB classifier for FB model predicts fraud with 60% accuracy while FB's kNN and SVM models predict fraud with 40% and 20% accuracy. Also, the NB and kNN models predict fraud with 50% accuracy while SVM predicts scams with 40% accuracy. SVM has lower predictions in both models because PCA components eliminate linear relationships in correlated attributes making it hard to define a hyper plane that separates the class labels. We believe the model for email web genre is promising because it is constructed on a data of native English intellectual cybercriminals. We observed that messages of the non-native English cybercriminals are much less complex in structure, ambiguous and redundant compared to those of native English cybercriminals.

Table 2 Summary of PL Features

Linguistic Feature	PL Process Linked to Deception and Cybercrime
Quantity of words	Criminals are talkative to make their scams and fraud forgettable
Average sentence length	Criminals are verbally fluent to make to scams fraud forgettable hence high cognitive complexity
First person pronoun singular *(i.e. I, me, mine)*	Criminals use less first pronouns to avoid accountability in their messages
Exclusive words *(i.e. but, without, exclude)*	Criminals use more exclusive words to be more imprecise hence high cognitive complexity & deception

Table 2 is a summary of the metrics we use to evaluate performance of the classifiers in the training model. The classifiers in both web genres perform well given the CL and PL features for cybercrime detection. For the Enron web genre, NB has precision of 0.939, recall of 0.938, f measure of 0.938 and ROC of 0.976. SVM has a precision of 0.961, recall of 0.961, f measure of 0.961 and ROC of 0.961. IBK has a precision of 0.878, recall of 0.876, f-measure of 0.876 and ROC of 0.883. For the FB web genre, NB has a precision of 0.898, recall of 0.890, f-measure of 0.91 and ROC of 0.917. SVM has a precision of 0.914, recall of 0.91, f-measure of 0.91 and ROC of 0.91. IBK has a precision of 0.9, recall of 0.9 and ROC of 0.9.

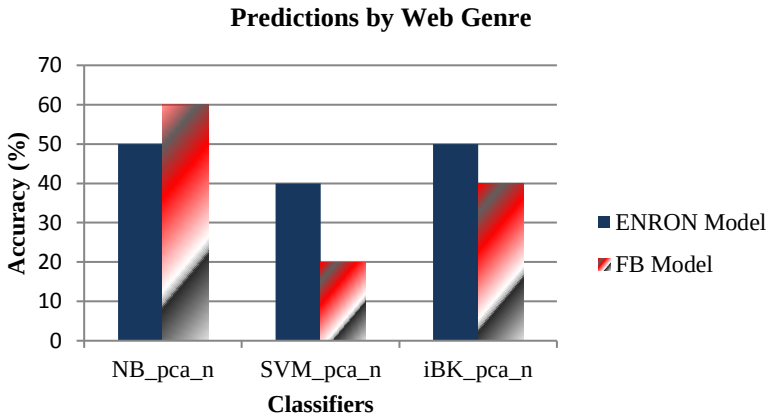

Fig. 1. Predictions of Classifiers by Web Genre.

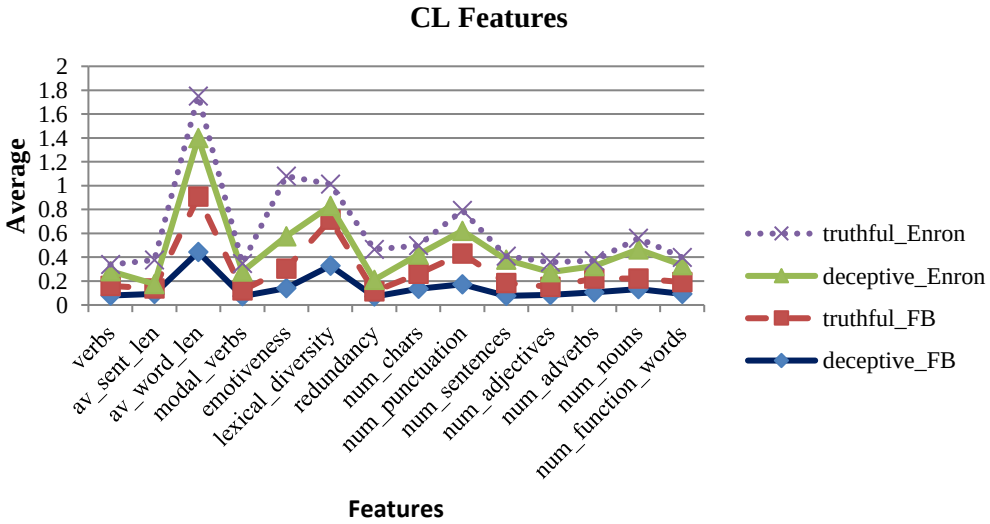

Fig. 2. Computational Linguistic Features by Web Genre.

Table 3 Evaluation of Classifier Performance for Training Model

Web Genre	Classifier	Precision (P)	Recall (R)	F	ROC
Enron	NB	0.939	0.938	0.938	0.976
Enron	SVM	0.961	0.961	0.961	0.961
Enron	kNN	0.878	0.876	0.876	0.883
FB	NB	0.898	0.89	0.889	0.917
FB	SVM	0.914	0.91	0.91	0.91
FB	kNN	0.903	0.9	0.9	0.9

Figure 2 dissects the features set for Enron and FB models into CL features. Both models reveal that deceptive messages are more verbose and expressive but with less complex structure compared to truthful ones. However, scams are even more verbose than fraud because cybercriminals are not native English speakers and neither are they as intellectual as the native English speaking fraudsters. Hence the scams will be more verbose and redundant, with lower lexical diversity and less complex sentence structures compared to fraud messages. On the contrary, fraud messages have higher expressivity than scams because intellectual fraudsters are more eloquent than scammers as indicated by the higher lexical diversity measures. This category of fraudsters has better linguistic styles for expressing emotion to be more deceptively convincing in messages to their victims.

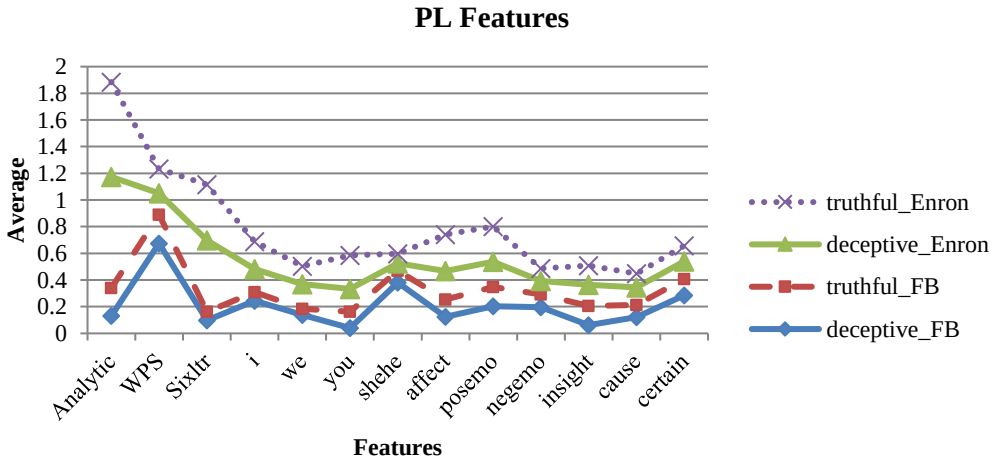

Fig. 2. Psycholinguistic Features by Web Genre

Figure 3 depicts the feature set for both models with respect to PL features. Both models reveal that deceptive messages have lower cognitive complexity, fewer self-references, with shorter phrases to make the messages forgettable compared to truthful ones. We also observe that fraud messages have relatively higher cognitive complexity than scam messages with respect to analytical phrases with evidence of phrases for expressing logical arguments, hence more words per sentence (WPS). Fraud messages have fewer self references than scams which implies that criminal actors do not want to be held accountable for their communication. We also observe that fraud messages express affection, positive, and negative emotion better than scam messages.

5 Conclusion

The popularity and convenience of text messaging is leading cybercriminals to exploit victims using this method of communication. Vulnerabilities in content filtering mechanisms continue to make it difficult to protect victims from cybercrime, leading us to explore a linguistic approach that use computational linguistics, psycholinguistics, and machine learning. Our experiments demonstrate that it is possible to use text-based deception detection to recognize messages representing cybercrime. Our research also demonstrates that it is possible construct a cybercrime detection model

using features from natural language processing and psychological processes linked to cybercrime. In our research, we also demonstrate we can construct models in different web genres that generalize detection of scams and fraud in text messages in other web genres.

References

1. Abozinadah, E., Mbaziira, A., & Jones, J. (2015). Detection of Abusive Accounts with Arabic Tweets. *International Journal of Knowledge Engineering*, *1*(2), 6. http://doi.org/10.7763/IJKE.2015.V1.19
2. Afroz, S., Brennan, M., & Greenstadt, R. (2012). Detecting Hoaxes, Frauds, and Deception in Writing Style Online. In *2012 IEEE Symposium on Security and Privacy (SP)* (pp. 461–475). http://doi.org/10.1109/SP.2012.34
3. Chang, C., & Lin, C.-J. (2001). *LIBSVM: a Library for Support Vector Machines*.
4. Chen, Y., Zhou, Y., Zhu, S., & Xu, H. (2012). Detecting Offensive Language in Social Media to Protect Adolescent Online Safety. In *Privacy, Security, Risk and Trust (PASSAT), 2012 International Conference on and 2012 International Confernece on Social Computing (SocialCom)* (pp. 71–80). http://doi.org/10.1109/SocialCom-PASSAT.2012.55
5. Cohen, W. (2015, May 8). Enron Email Dataset. Retrieved March 29, 2016, from http://www.cs.cmu.edu/~enron/
6. Crawford, M., Khoshgoftaar, T. M., Prusa, J. D., Richter, A. N., & Najada, H. A. (2015). Survey of review spam detection using machine learning techniques. *Journal of Big Data*, *2*(1), 1–24. http://doi.org/10.1186/s40537-015-0029-9
7. Engel, P. (2015, May 9). ISIS has mastered a crucial recruiting tactic no terrorist group has ever conquered. Retrieved March 16, 2016, from http://www.businessinsider.com/isis-is-revolutionizing-international-terrorism-2015-5
8. Fette, I., Sadeh, N., & Tomasic, A. (2007). Learning to Detect Phishing Emails. In *Proceedings of the 16th International Conference on World Wide Web* (pp. 649–656). New York, NY, USA: ACM. http://doi.org/10.1145/1242572.1242660
9. Firte, L., Lemnaru, C., & Potolea, R. (2010). Spam detection filter using KNN algorithm and resampling. In *2010 IEEE International Conference on Intelligent Computer Communication and Processing (ICCP)* (pp. 27–33). http://doi.org/10.1109/ICCP.2010.5606466
10. Frank, E., Hall, M., Holmes, G., Kirkby, R., Pfahringer, B., Witten, I. H., & Trigg, L. (2005). Weka. In O. Maimon & L. Rokach (Eds.), *Data Mining and Knowledge Discovery Handbook* (pp. 1305–1314). Springer US. Retrieved from http://link.springer.com/chapter/10.1007/0-387-25465-X_62
11. Hall, E. (2015, March 11). How ISIS Uses Twitter To Recruit Women [BuzzFeed News]. Retrieved March 16, 2016, from http://www.buzzfeed.com/ellievhall/how-isis-uses-twitter-to-recruit-women
12. Hancock, J. T., Curry, L. E., Goorha, S., & Woodworth, M. (2007). On Lying and Being Lied To: A Linguistic Analysis of Deception in Computer-Mediated Communication. *Discourse Processes*, *45*(1), 1–23. http://doi.org/10.1080/01638530701739181
13. ITU. (2009). *Understanding Cybercrime:A Guide For Developing Countries* (p. 225). Switzerland: International Telecommunications Union. Retrieved from http://www.itu.int/ITU-D/cyb/cybersecurity/docs/itu-understanding-cybercrime-guide.pdf
14. Larose, D. T. (2014). Why do We Need to Preprocess the Data? In *Discovering Knowledge in Data: An Introduction to Data Mining, 2nd Edition* (2nd ed.). John Wiley & Sons. Retrieved from http://proquest.safaribooksonline.com/book/databases/business-intelligence/9781118873571/chapter-2-data-preprocessing/c02_xhtml
15. Li, F., Huang, M., Yang, Y., & Zhu, X. (2011). Learning to identify review spam. *Proceedings of International Joint Conference on Artificial Intelligence*, *22*(3).
16. Mbaziira, A., Abozinadah, E., & Jones, J. (2015). Evaluating Classifiers in Detecting 419 Scams in Bilingual Cybercriminal Communities. *International Journal of Computer Science and Information Security*, *13*(7), 7.

17. McGrath, S. (2015, June 11). Winning The War: The Evolution Of Cybercrime [Technology]. Retrieved from http://www.informationweek.com/interop/winning-the-war-the-evolution-of-cybercrime/a/d-id/1320817

18. Mehler, A., Sharoff, S., & Santini, M. (Eds.). (2011). *Genres on the Web* (Vol. 42). Dordrecht: Springer Netherlands. Retrieved from http://link.springer.com/10.1007/978-90-481-9178-9

19. Muireann. (n.d.). English Vocabulary: The Subtle Difference Between Fraud and Scam. Retrieved May 10, 2016, from http://www.skypeenglishclasses.com/english-vocabulary-the-subtle-difference-between-fraud-and-scam/

20. Mukherjee, A., Liu, B., & Glance, N. (2012). Spotting Fake Reviewer Groups in Consumer Reviews. In *Proceedings of the 21st International Conference on World Wide Web* (pp. 191–200). New York, NY, USA: ACM. http://doi.org/10.1145/2187836.2187863

21. Pearl, L., & Steyvers, M. (2012). Detecting authorship deception: a supervised machine learning approach using author writeprints. *LLC*, *27*, 183–196.

22. Ponemon Institute. (2015). *2015 Cost of Cyber Crime Study: Global* (p. 29). Retrieved from https://ssl.www8.hp.com/ww/en/secure/pdf/4aa5-5207enw.pdf

23. Reynolds, K., Kontostathis, A., & Edwards, L. (2011). Using Machine Learning to Detect Cyberbullying. In *2011 10th International Conference on Machine Learning and Applications and Workshops (ICMLA)* (Vol. 2, pp. 241–244). http://doi.org/10.1109/ICMLA.2011.152

24. Rowe, N. (n.d.). Detecting Online Deception and Responding to It. Retrieved March 16, 2016, from http://www.au.af.mil/au/awc/awcgate/nps/decep_detec.htm

25. Sarvari, H., Abozinadah, E., Mbaziira, A., & McCoy, D. (2014). Constructing and Analyzing Criminal Networks. *IEEE Security and Privacy Workshops*, 8. http://doi.org/DOI 10.1109/SPW.2014.22

26. Shojaee, S., Murad, M. A. A., Azman, A. B., Sharef, N. M., & Nadali, S. (2013). Detecting deceptive reviews using lexical and syntactic features. In *2013 13th International Conference on Intelligent Systems Design and Applications (ISDA)* (pp. 53–58). http://doi.org/10.1109/ISDA.2013.6920707

27. Shropshire, C. (2016, March 26). Americans prefer texting to talking, report says [News]. Retrieved April 1, 2016, from http://www.chicagotribune.com/business/ct-americans-texting-00327-biz-20150326-story.html

28. Sommer, R., & Paxson, V. (2010). Outside the Closed World: On Using Machine Learning for Network Intrusion Detection. In *2010 IEEE Symposium on Security and Privacy (SP)* (pp. 305–316). http://doi.org/10.1109/SP.2010.25

29. Tan, P.-N., Steinbach, M., & Kumar, V. (2006). *Introduction to data mining* (Vol. 1). Pearson Addison Wesley Boston.

30. Tausczik, Y. R., & Pennebaker, J. W. (2010). The Psychological Meaning of Words: LIWC and Computerized Text Analysis Methods. *Journal of Language and Social Psychology*, *29*(1), 24–54. http://doi.org/10.1177/0261927X09351676

31. Theoharis, M. (2016). Fraud - Laws, Sentencing, Penalties | Criminal Law. Retrieved May 10, 2016, from http://www.criminaldefenselawyer.com/crime-penalties/federal/Fraud.htm

32. Vergelis, M., Shcherbakova, T., Demidova, N., & Gudkova, D. (2016, February 5). Kaspersky Security Bulletin. Spam And Phishing In 2015 - Securelist. Retrieved March 31, 2016, from https://securelist.com/analysis/kaspersky-security-bulletin/73591/kaspersky-security-bulletin-spam-and-phishing-in-2015/

33. Westman, S., & Freund, L. (2010). Information Interaction in 140 Characters or Less: Genres on Twitter. In *Proceedings of the Third Symposium on Information Interaction in Context* (pp. 323–328). New York, NY, USA: ACM. http://doi.org/10.1145/1840784.1840833

34. Wollman-Bonilla, J. E. (2003). E-mail as genre: A beginning writer learns the conventions. *Language Arts*, *81*(2), 126–134.

35. Zhou, L., Burgoon, J. K., Twitchell, D. P., Qin, T., & Nunamaker, J. F. (2004). A Comparison of Classification Methods for Predicting Deception in Computer-Mediated Communication. *Journal of Management Information Systems*, *20*(4), 139–165.