Machine Training for Intelligent Analysis of Text for the Identification of the Author

Nadiia Pasieka[1], Vasyl Sheketa[2], Myroslava Kulynych[3], Yulia Romanyshyn[2], Svitlana Chupakhina[1], and Olha Khytrova[4]

[1] *Vasyl Stefanyk Precarpathian National University, Ivano-Frankivsk, 76000, Ukraine*
[2] *National Tech. University of Oil & Gas, Ivano-Frankivsk, 76068, Ukraine*
[3] *Ukraine Academy of Printing, Lviv, 79020, Ukraine*
[4] *Chernivtsi Trade and Economic Institute Kiev National University of Trade and Economics, Ukraine*

Abstract

The continuous development of information technology has led to an increasing danger and critical cyberattacks, which have recently developed and penetrated unimpeded in various institutions that have a sophisticated infrastructure of information technology use. Based on the analysis of the last three years, there have been critical cases of cybercrime around the world, primarily involving significant leaks of critical information, the spread of fake messages, cyberbullying, and cloud-based cryptojacking. As a result, scientific research has sprung up around the world to unambiguously identify the cybercriminal. For this purpose, various agencies have improvised innovative methods to combat this vice, as well as on the possibility of bringing the perpetrators to justice, in connection with such critical cybersecurity issues. As one option to effectively address this problem, the Forensic Writer Identification system, which works on the principles of stylometry is being considered. Indeed, the intellectual analysis of text for belonging to one or another author is a complex technological task, at its base uses artificial intelligence technology to identify, protect, recognize, create, extract and document digital evidence, which can then be used as evidence of wrongdoing regarding social media users or simply to analyze critical data. Thus, the main goal of this study is to examine in detail the capabilities of Forensic Writer Identification technology to analyze the tweets of different users around the world and unequivocally and apply it to reduce the search time for criminals by providing the police with the most accurate methodology. As well as to compare the accuracy of different methodologies. Conducted analytical research behind a logical literature review that examines the most important methods of text analysis. The study used texts from various Twitter users for intelligent analysis. Various online and offline databases were used to expedite the study, and information systems were used to efficiently search for relevant scholarly results. As systems analysts have recently emphasized computer methods for rapid analysis of digital text in order to establish authorship, the results presented are very encouraging. Thus, this research provides a general framework and rationale for the use of text and author identification methods. This article reviews current research methods and software applications, and touches on the issues of evaluating the performance of such research. Various research strategies for digital text research are summarized, and a more detailed description of two combined methods is presented. Thus, through the use of textures, algorithms, and polygraphs, new technologies are beginning to show valuable levels of performance. Nevertheless, the use of combined methods to analyze text for its identity will play a vital role in future technologies. In this regard, the goal of formulating a project proposal is to create an analytical analysis system that automatically recognizes authorship of all aspects of technology on a global scale, which may partially solve the problems of modern cybercrime.

Keywords

Machine learning, tweet analysis, cyberbullying, cybercrime, text analysis, stylometry.

CPITS-II-2021: Cybersecurity Providing in Information and Telecommunication Systems, October 26, 2021, Kyiv, Ukraine
MAIL: pasyekanm@gmail.com (N. Pasieka); vasylsheketa@gmail.com (V. Sheketa); kumyr@ukr.net (M. Kulynych); yulromanyshyn@gmail.com (Y. Romanyshyn); cvitlana2706@gmail.com (S. Chupakhina); olga_hitrova@ukr.net (O. Khytrova)
ORCID: 0000-0002-4824-2370 (N. Pasieka); 0000-0002-1318-4895 (V. Sheketa); 0000-0002-9271-7855 (M. Kulynych); 0000-0001-7231-8040 (Y. Romanyshyn); 0000-0003-1274-0826 (S. Chupakhina); 0000-0003-2253-4356 (O. Khytrova)

1. Introduction

The permanent process, the development of innovations in digital communications, such as network technologies (social networking platforms), SMS, online forums, blog posts and emails, have defined the initial conditions for a faster as well as effective transmission of digital information through an innovative intelligent IT-infrastructure. Under certain conditions, the author of texts can remain incognito, but unfortunately, this status can also lead to various cybersecurity problems on all available social networks. Digital forensics of anonymous texts for authorship is the process of using information technology, namely the storage, extraction and transmission of data from various technical platforms to identify evidence and combine it into useful information that can be effectively used to solve network problems like cyberbullying around the world. Thus, effective digital, authentication and attribution of a specific author of an online source text is a critical tool in combating cybercrime. Unfortunately, in our study, the length of the original text is limited to a few steps, which complicates the process of effective testing. [1, 2, 4, 6] Thus, the goal of our study is to process the differentiation of triggering messages from Twitter that are limited to one hundred and forty characters in length. In our study we consider and evaluate the basic stylometric signs of the digital text belonging to a particular author, for this purpose we use abstract methods for examination, as well as explicit signs for the social network Twitter, such as URLs, hashtags, responses or statements. [3, 5, 11, 16, 28] Subsequently, such analytical approaches are able to achieve an optimum accuracy of about eighty percent in determining text authorship. Thus, with the increasing number of cybercrimes against social network users around the world, there is a critical need to identify the author of digital text with an acceptable and adequate quality of author identification, there is an important issue for scientific expertise in order to effectively combat cybercrime, such as cyberbullying, emasculative notes and deception on the Internet by various criminals.

Of course, the anonymity provided by phones with ready-made SIM cards, public networks, public Wi-Fi zones and decentralized network resources, for example such as Tor, an application that hides the personal attributes of the user, can make the task of identifying online customers much more difficult. Thus, sometimes the content of a separately published message text becomes the only effective information to identify its creator. At the same time, it is also worth noting that the real existence of online text messaging clients sometimes turns out to be quite different from what they appear to be and how they look on the Internet. [7, 9, 12, 24, 36] However, it is necessary to emphasize the permanent processes that change the nature as well as the results of this and are constantly evolving. In addition, in various publications circulating in the Internet space there is an example of a media organization which in its activities has allegedly created and used pseudonymous virtual personages to conduct a coordinated campaign of disinformation through the network media systems. Along with this, it is also expected that some of these tasks to determine the authorship of text messages in order to stop illegal actions will be supported by state structures.

To correctly identify an anonymous author from a published text message (which looks like a cyberbullying) through the analytical process of authorship identification in social and public networks is presented in the model Figure 1.

The permanent development of the Internet offers more and more opportunities for cybercriminals to secretly spread their malicious deeds, such as phishing, cyberbullying, fraud and spam. To this end, an original study has been proposed to determine the authorship of such digital works on the Internet (e.g., discussion comments, analysis of text messages, tweets, SMS) The use of scientific phonetics provides a special mathematical apparatus for the examination of such digital text messages based on a certain linguistic evidence of belonging to a particular author Thus, online identity is ways of orthographic and phonetic composition of original digital text, and an This study also involves computer analysis of text messages for combining styles or style metrics from an archive of such malformed messages. [8, 13, 18, 40] The study solves several different problems, namely the automated text message extraction analysis, in order to determine and identify the real author of the analyzed content. Thus, the largest subjects of storage and generation of digital text messages make considerable efforts involving data analysts. [10, 17, 31] However, the use of applications without advanced artificial intelligence (AI) functions, does not provide sufficient information to unambiguously identify the authorship of digital text messages, both real and

suspected cybercriminals. The Internet provides a useful environment (arena) for cybercriminals to covertly carry out their intended activities, such as phishing, fraud and spam.

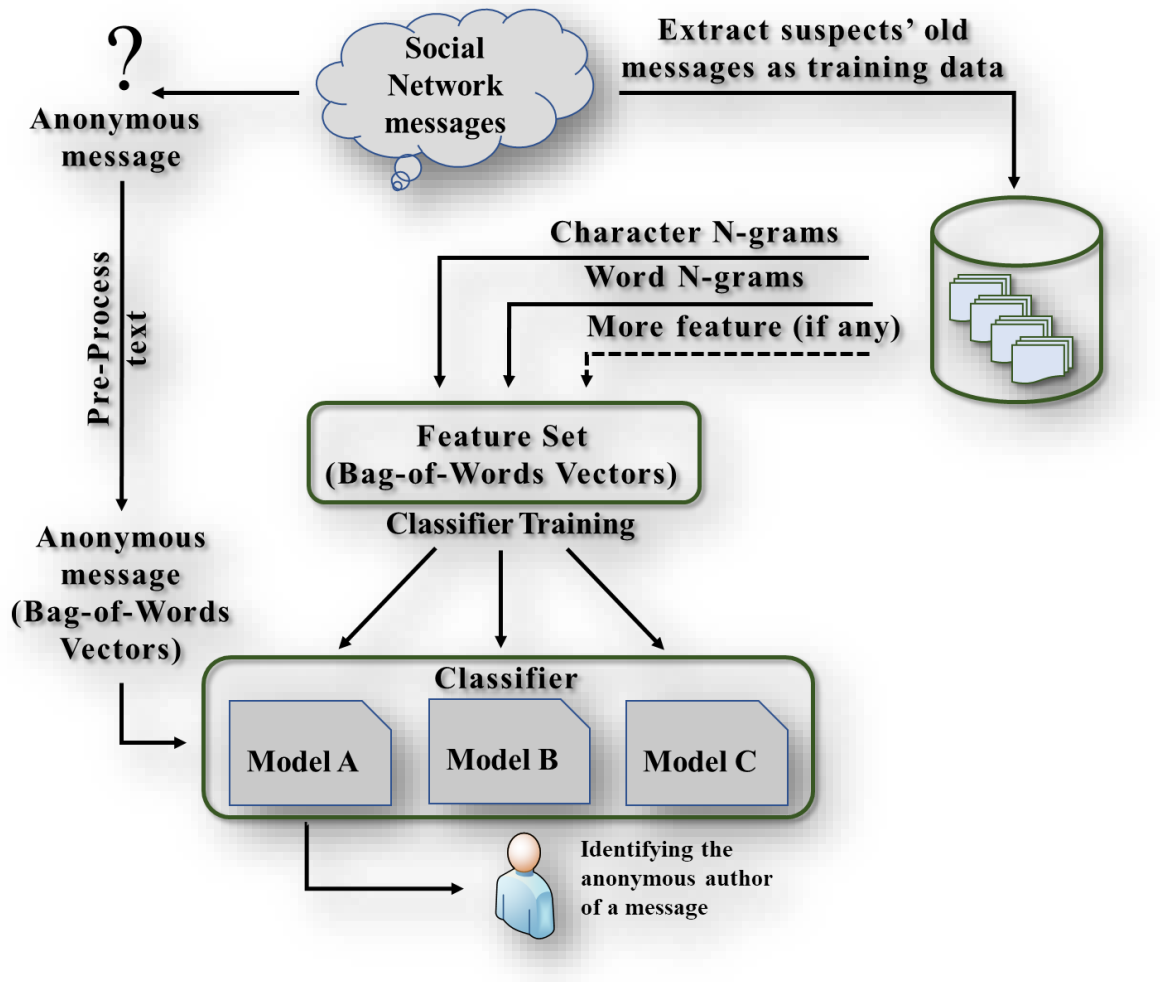

Figure 1: A model for identifying the author of a text message

Thus, the relevance of original research on anonymous digital text messages such as, discussions, comments on them tweets. Therefore, the qualitative determination of the authorship of anonymous text messages directed against users in the Internet environment and received serious direction in the analytical analysis of the available information. Scientific phonetics interrogates experts who investigate linguistic turns (unknown or authorized works used in cybercrimes) during the preliminary analysis of digital text messages. In addition, the analytical examination of digital text messages, which malicious actors tend to obscure problems with online content, are called creation checks. [14, 15, 21] Essentially, online digital text message originality checks are an examination of voice or computational attributes that may be recorded by a combination of known or unknown software systems. In addition, the process of digital text message author identification can also include checking the composition style or style accents in the investigated signature. [19, 22] It should also be emphasized that there are various challenges to implementing this process, for example, manual extraction and automatic text analysis to identify the true author of a digital text message has become a significant challenge for various specialized units. Based on the paradigm that most of the specialized units that are engaged in deep analysis of big data, use computer systems that do not have innovative capabilities with elements of artificial intelligence, and therefore cannot cope with the task of identifying the author of a digital text message, both real and perceived.

In addition, various projected applications around the world lack basic artificial intelligence algorithms for analyzing digital text messages for anonymous author identification, which can solve some delicate

problems, such as language attribute recognition in electronic communication services. Internet short messages, such as the signature set in Twitter characters, have some attributes that make the origin of the location test comparison, and moreover, the official content of the artistic function is written as follows. [19, 26, 32, 37] It should be emphasized that the web digital text message is short or long in most cases, indicating that a specific language complex measure that depends on the number of words in the content may not be appropriate.

In addition, some computer applications may not recognize parts of speech. In fact, part-of-speech tagging classifies the parts of speech for each word in online content, taking into account the quality of the word and the specific circumstances in which it appears. In addition, in most cyberattacks, cybercriminals use false and fraudulent methods to gain access to various Internet and Web technologies, manipulating user information without user authorization. Undoubtedly, the constant growth in the use of information and communication systems and technological equipment has led to an increase in global cybercrime and conflicts between governments and criminals. [20, 23, 35, 41] Ultimately, in order to correctly identify an anonymous user, a significant volume of banned digital text messages must be processed for analysis. However, cybercrime is generating more and more new methods and techniques for its unlawful acts and as a result there are more and more cyberattacks on various Internet sites. [25, 27] As a rule, cybercriminals, for their unlawful deeds, use mass distribution in social networks specially programmed false accounts, which they use as:

- a means to legally carry out cybercrime business;
- organizing the embedding of special information flows with the purpose of discrediting both organizations and private users of social networks;
- unlawful possibility of mass theft of personal data of social network users;
- artificial creation of certain conditions for deterioration of trust in social networks;
- the artificial creation of fake news feeds and fake votes (ratings);
- artificial creation of problems for social marketing.

To achieve the task of identifying an anonymous author of a digital text message we used methods of computer visualization (for graphical representation of social networks), methods of collecting personal data from various Internet services with analytical analysis, mathematical methods of visual pattern recognition, methods of piece intelligence, methods of hierarchical clustering, neural network methods, fuzzy clustering methods (to conduct clustering Web users into groups with the same attributes), as well as simulation modeling (to analyze the results obtained).

2. Analysis Literature

In fact, some of the nastiest digital text messages posted in the Internet environment and on various social media platforms can be used in criminal investigations of content authorship, especially through the method of forensic phonetic identification of anonymous authors. Unfortunately, many authors of such illegal online content messages still remain out of reach. Despite the fact that on some digital platforms where web technologies are hosted, the problem of identifying the anonymous author of a particular digital textual content can be quite a challenge, since the length of digital text messages on the Internet has limitations on a certain number of characters. [29, 30] Consequently, the goal of a responsible data analyst is to identify the anonymous sender of a tweet or the owner of a virtual Twitter account, where a tweet digital text message can be up to two hundred and eighty characters long. Therefore, it became a matter of developing an algorithm to evaluate the key metrics used for initial attribution, which can identify the key characteristics of each cybercriminal's compositional style. Specifically, research data analysts applied stylistic measurement techniques to various digital text messages for natural language processing (NLP), including specifying an anonymous author, author profile, author verification, detecting changes in a particular text style, and grouping different tweets. [22, 34, 39] Obviously, stylistic algorithms were used to identify the anonymous author of the digital textual content or the blog of a given tweet from the so-called list of investigated authors in order to achieve a positive result at the center of this study. The aforementioned methods were used to test

different data sets, and finally, ideas for future work were considered. In fact, this research topic has important implications for the application of data analysis tools to help reduce cybercrime worldwide.

Obviously, with offline and online digital text messages, it is not easy to track any direct intra-species handwritten creations to identify or verify an anonymous author. Nevertheless, many research papers have been published on author-specific technology for identifying an anonymous author from digital text messages, which has become a basic tool for big data analytics. Therefore, they have been used to use stylistic measurement techniques to analyze social media texts for this article. In addition, online social networks (OSN), such as Facebook, LinkedIn and Twitter, provide increasingly new factors for triggering anonymous attribution. [33, 38–42] These online tools are believed to provide effective and quick virtual linking methods that help anonymous social media users commit cybercrimes. Essentially, users can use screen names, aliases, or VPNs in these places, and other users may not provide the correct credentials for the record. In fact, to obtain meaningful test results, the data scientist or analyst should use an information validation method based on reliable information that contains known, validated models and key points applicable to the test being performed. Will be most accurate as shown in Figure 2 below. In addition, to compare multiple literature articles, the report should consider the following aspects.

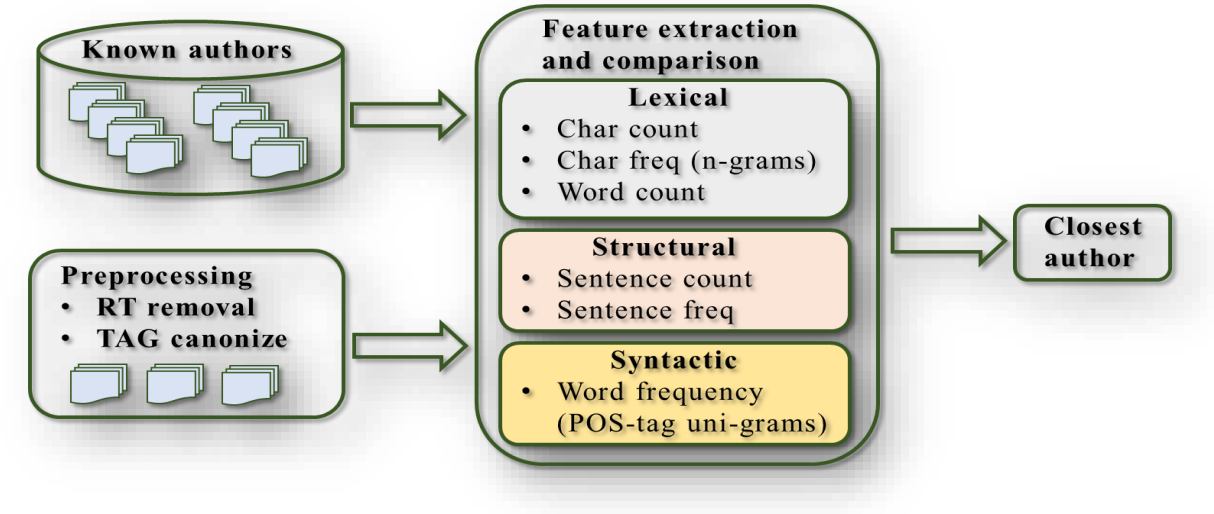

Figure 2: Model of the information-analytical system for identifying an anonymous author

3. Analytical identification versus verification

In essence, semantic language properties are focused on obtaining information about the creators of digital textual content in the field of stylistics. In addition, most big data analysts use a variety of stylistic identity verification methods to identify anonymous customers based on their digital text messages. This study details the digital content originality verification methods used for sequential validation using unstructured online partitions based on message content. The online file is then split into continuous grids of short messages, and (sequential) confirmations are selected in these grids to separate true and self-proclaimed practices. To achieve the exact goal of the study, analysts are encouraged to analyze content containing one hundred forty, two hundred eighty, and five hundred characters based on the proposed algorithm, which is designed to extract the digital text message in the study. In addition, the list of features includes traditional digital text selection such as morphology, syntax, explicit selection, as well as innovative selection created in the study of n-grams. In addition, the proposed method includes a number of methods to circumvent the emerging problems associated with unbalanced sets of big data, and uses the obtained information and the analyzed information as component selection methods and support vector mechanisms (SVMs) for grouping digital text messages. Experimental evaluation of the proposed method and text message parsing algorithm based on Enron and Twitter emails gave promising results: according to the syntax of the algorithm used, the EER is approximately 23.86%. In addition, the clustering of the analyzed digital text message can be obtained by the formula (1) shown in Figure 3 below, where A and λ

(sigma) represent the normalization of the big data set, S_K represents the Gaussian distribution of the data set, and Du represents the u-th digital text message.

$$P_{t,u}(S) = \frac{1}{A}\sum_{k=1}^{n_0} G(S_k\lambda) = \frac{1}{A}\sum_{k=1}^{n_0} \frac{1}{\sqrt{2\pi}\lambda} exp\left[1\frac{(S-S_k)}{\lambda^2}\right] \quad (1)$$

4. Methodologies for analyzing digital text messages

As one of the criteria for identifying digital text messages or tweets, the research uses stylistic methods to accurately perform the post-creation task of identifying anonymous authors of content. In this section, we will comprehensively describe the main methods that can be used to automatically identify an anonymous author across multiple Twitter accounts. Essentially, the desired methodology involves the selection of software that uses stylometric, semantic algorithms and the fleeting parts of authoring mechanisms.

Stylometric mechanisms for identifying an anonymous author by content.

Of course, different analysts use stylistic variables to identify the anonymous authorship of amateur bloggers on social media as one way to tackle the growing problem of cybercrime. In addition, the stylistic analysis of the corpus is usually one of the main methods of identifying the author of a Twitter post. Usually their consideration depends on the field of application. For example, the source attribution method and other short online posts include basic style tags such as good news, document links, specific HTML tags, etc., or bizarre style tags such as misspellings. Since our research area is abstract writing, we decided to include a wider range of style markers. The stylistic variables commonly used in this study are specific aspects such as punctuation in tweets and n-gram text arrays.

In fact, the corpora used in this study usually consists of multiple digital text messages from several anonymous authors. Like any corpus-based method, corpus configuration is strictly considered to produce reasonable research results. Regarding source attribution, the characteristics of the corpus (text type, language, time, case) affect the accuracy of attribution. In addition, the ideal consensus recommendation is based on the summary of the author's company, taking into account the narrowest possible language range. In addition, we assume that we have received a set of predetermined posts or tweets as information in online media, and each method will generate a similarity score for each pair of accounts analyzed. In fact, the corpus used in this study usually consists of multiple posts by multiple authors. As with every model-based method, the configuration of the model is considered from a critical point of view to produce the most reasonable research results. Regarding the attribution of provenance, model characteristics (text type, language, time, case) affect the accuracy of attribution. In addition, it has been clearly suggested that it is ideal to accumulate a corpus of digital text messages based on anonymous authors, who have considered the narrowest possible language range. In addition, we assume that we have obtained a set of time-stamped posts or digital text messages (tweets) as information in online media. Therefore, each suggested determination method will generate a similarity score for each pair of accounts analyzed.

In addition, this technology is used in conjunction with the spry information mining program that is essential for text extraction, and the check also corresponds to the four important stages of tweet recognition, such as preprocessing different tweets, extracting tweet features, and comparing the same tweets And to determine the identity of the author.

Logically speaking, the stylistic research of tweets is similar to the stylistic research of different types of short messages, such as online discussion posts or online text conversations. It is worth noting that they are random and comparative in design and syntax. For our tests, a complete feature list has been developed, taking style data into account in any case, and assuming that creators unknowingly follow a certain design and are predictable in their decisions. In the analysis process, the algorithm will use iterative techniques to search the array tree and identify various characteristic vocabulary features-this will help produce the most accurate results.

In addition, preprocessing tweets involves deleting suspicious tweets from some authors. In fact, the proposed tweet preprocessing technique relies on the association of slang words with other matching words to check the importance of slang words and their presumed interpretation. We use n-gram to detect connections and condition arbitrary fields to check the meaning of slang terms. However, an

important issue in this field is data matching about hype, relevance, emoji, folklore classification, and slang.

Recognizing the characteristics of tweets involves analyzing pre-suggested tweets to determine the total number of words in a particular tweet, the special characters used, the consonants and vowels, the total number of characters, and the regularity of the use of certain words or phrases. In fact, recognizing appropriate selections for learning opinions in tweets remains an open area of research, as text-ordering methods face the problem of parsimony and part-of-speech (POS) labeling strategies bombard the lack of linguistic construction of tweets. The character-based selection, namely n-gram characters, is now very mature because they do not contain any language.

Logically speaking, the analyst must find content that matches the content stored in the database in order to use feature matching techniques to detect similarities when possible. In addition, the feature evaluation algorithm of selected tweets can be used to directly identify friendships and find evaluations. For example, decision-making evaluation is an important area, and presumption checking of tweets has been widely used in the past few years.

Lexical

This is the technique of Twitter sentiment investigation. In fact, it attempts to quantify the diversity of popular assessments of retail signs. The first is a vocabulary-based strategy, which uses word references and semantic scores of words to calculate the final endpoint of a tweet and incorporate grammatical feature tags. In addition, with the rapid spread of interpersonal organizations, Weibo applications, and gatherings, its main role is developing significantly.

Structural

Basically, Twitter is a Weibo platform that spreads about 450 million tweets every day. It requires a structured algorithm to accurately analyze all data. In addition, it also solves the important information hotspots of disease fighting and control in local areas. This research investigates the structural algorithm methods used to separate and analyze Twitter information, including the attributes and representativeness of the information; information sources, access, and cost; inspection methods; information boards and cleanliness; standardized measurements; and examinations.

Syntactic

In essence, the language used on Twitter has some peculiarities, such as the use of hashtags or client references. Therefore, data preprocessing technology will use syntactic algorithms or data structures to speed up the analysis process. In order to improve the effectiveness of language preparation techniques (such as morphological restoration and syntactic analysis), we have performed some standardization steps. We eliminate #images, all @ notices, and connect and perform lowercase conversion. Similarly, if a vowel is repeated multiple times in a word, we reduce it to a single event, and reduce the different back-to-back accent marks to a single event. Finally, we lemmatize the standardized content.

N-grams

With the advancement of global Web technology, the Internet has been used as a hot spot to get news about the latest developments. Recently, Twitter is probably the most famous online media platform that allows public customers to share news. This stage develops rapidly, especially among young people who may be affected by data from mysterious sources. In this way, foreseeing the credibility of information on Twitter becomes a need, especially in the event of a crisis. This paper proposes an arrangement model based on managed AI strategy and word-based N-gram analysis to naturally rank Twitter messages as sounds instead of entities. Applied and specifically analyzed five different management characterization programs: Linear Support Vector Machine (LSVM), Logistic Regression (LR), Random Forest (RF), Naive Bayes (NB) and K-Nearest Neighbor (KNN). This exam explores two descriptions (TF and TF-IDF) and different word N-gram ranges. For model preparation and testing, a 10-crease cross-approval was performed on two data sets in different dialects (such as English). In addition, the application of syntactic concepts will also be used to relieve the sentence structure about stress and the grammatical form-part of speech. According to the theory, punctuation is an important guide to characterize restrictions and make tweets or posts meaningful by dividing paragraphs into sentences and dividing each sentence into different tags, such as exclamation marks, quotations, and periods. The arrangement of basic accents includes single sentences, commas, periods, colons, semicolons, question marks, and shout signs. In addition, according to the Unicode design with code and specific grammar, a series of notable emphasis is made using various symbols, such as (!, "). The function is subject-specific and captures the creator’s style in various themes These highlights

include the number of sentences per square of text, the number of normal characters, words and sentences in a square text, and the number of normal sentences beginning with uppercase and lowercase.

Text Representation

Logically, for each style variable, the data analyst creates a loop vector, where each measurement is related to an alternative component. In addition, in most experiments, data analysts will select different combinations of factors, and then treat each report as a link to a repeating vector related to these factors. Basically, some of the important features include the total number of characters in each tweet, the total number of words in each sentence in the tweet, the frequency of words in the tweet, the total number of common words in the dictionary, emoji in the tweet and its appearance in the tweet. In the meaning of Google and various online applications, trend words are for specific reasons-maybe a day or a week or so. Fundamentally, the main period of Twitter content confirmation will be pre-measurement of different tweets from certain clients, where untrusted parts will be removed from content written by others, for example, other web-based media-Twitter-customers Tweets forwarded by the end.

Many etymological characters have been recommended for origin checking, for example, the determination of specific words and syntactic structure. In sharp contrast to topic-based content arrangements (the essential issue is a set of texts), key arrangements are expanded by adding vocabulary, syntax, and application explicit focus. Such a mix can better convey the creator's style. In this way, the expert will use the created calculations to coordinate the highlights of the indistinguishable content characters. This is done after the component is discovered, and it is compared with the data sets of different authors to find similarities. It is worth emphasizing that the grouping model includes various configuration files created independently for a single customer.

5. Results

Thus, analyzing the results obtained by using different models to determine the anonymous authorship of digital text messages and machine learning to automate this process, we found that the best model is the model of increasing the gradient with an accuracy of 37.43%. We can state that the probability of detection is low, but we can also note that this figure can be directly related to the nature of the digital text messages processed. Since these anonymous authors can change on bot farms, and it is very likely that some authors are constantly changing agents, that is why the prediction is so low. Which will ultimately affect the accuracy of the proposed model.

Due to the cumbersome nature of analytical calculations, we present only the general values of accuracy of the models considered by the anonymous author by numerical analysis of text messages, which was for: Gradient Boosting 37.43%; Decision Tree 23.17%; Bayesian Network 19.89%.

As the results of the different models show, they are in line with our expectations, along with demonstrating the significant potential of different machine learning algorithms in identifying anonymous social media authors. And some differences in the percentage of detection for each algorithm successfully shows us that machine learning algorithms can be used to extract useful information from significant amounts of digital text message data and predict the author of a particular textual content. This could play a huge role in the future in the fight against cybercrime, cyberbullying, and identity theft from social media users.

6. Conclusions

The main purpose of the study is to automate the identification of the author of a digital message using various analytics technologies. To identify the anonymous author using the recognition of his text messages on the social network Twitter to minimize and prevent cybercrime. In fact, the science of data analytics has activated the development of a unique stylistic evaluation model for digital text messages that has the ability to use computer analysis of unique linguistic and stylistic "fingerprints" for affiliation and author identification. The study attempted to test the proposed concept of digital text message authorship by testing and evaluating the accuracy of the proposed model, the principle approaches that were tried to test, proves that they can significantly help to automate these mechanisms for the relevant structures. The results of the study clearly delineate the research topics. In addition, the evidence presented in the study subsequently confirmed that the proposed model is an acceptable mathematical

model for the process of author identification using stylometry technology. In fact, empirical research on determining the author of a digital text message has answered almost all of the questions posed and achieved the basic goals of being reviewed and investigated through computer dialogue and analytic analysis. However, there is still room for improvement in method models and algorithms, which requires additional research to achieve better results that will have a positive impact on reducing cybercrime. Due to the limited amount of data sets used in the proposed model, the ones used in this study are not highly accurate. These datasets consist of popular Twitter users who are likely to find a way to continue generating digital text messages or pay penalties to the organizers and continue participating in social media. Future research will explore other models, methods, and algorithms to obtain more robust metrics, and will greatly expand the data set that is now being accumulated on chimera resources. This research is also a successful starting point for further research, and applying models and methods to digital text messaging data significantly reduces the potential for cybercrime, which can be useful for various specialized units.

7. References

[1] Adak, C., Chaudhuri, B. B., & Blumenstein, M. (2019). An empirical study on writer identification and verification from intra-variable individual handwriting. IEEE Access, 7, 24738-24758. doi: 10.1109/access.2019.2899908

[2] Agile Business Consortium Limited. (2021). Agilebusiness.org. agilebusiness.org. https://www.agilebusiness.org/page/ProjectFramework_06_Process

[3] Alonso-Fernandez, F., Belvisi, N. M., Hernandez-Diaz, K., Muhammad, N., & Bigun, J. (2020). Writer Identification Using Microblogging Texts for Social Media Forensics. International Journal of Recent Trends in Engineering and Research, pp. 1-22

[4] Analysis of Stylometric variables in long and short texts. Procedia - Social and Behavioral Sciences, 95, 604-611. Retrieved from https://doi.org/10.1016/j.sbspro.2013.10.688

[5] B. Durnyak, B. H. O. Tymchenko, O. Tymchenko and D. Anastasiya, "Research of image processing methods in publishing output systems," 2018 XIV-th International Conference on Perspective Technologies and Methods in MEMS Design (MEMSTECH), 2018, pp. 178-181, doi: 10.1109/MEMSTECH.2018.8365728.

[6] B. Durnyak, B. Havrysh, O. Tymchenko, M. Zelyanovsky, O. O. Tymchenko and O. Khamula, "Intelligent System for Sensor Wireless Network Access: Modeling Methods of Network Construction," 2018 IEEE 4th International Symposium on Wireless Systems within the International Conferences on Intelligent Data Acquisition and Advanced Computing Systems (IDAACS-SWS), 2018, pp. 93-97, doi: 10.1109/IDAACS-SWS.2018.8525792.

[7] B. Durnyak, O. Tymchenko, O. Tymchenko and B. Havrysh, "Applying the Neuronetchic Methodology to Text Images for Their Recognition," 2018 IEEE Second International Conference on Data Stream Mining & Processing (DSMP), 2018, pp. 584-589, doi: 10.1109/DSMP.2018.8478482.

[8] Belvisi, N. M., Muhammad, N., & Alonso-Fernandez, F. (2020). Forensic authorship analysis of Microblogging texts using N-grams and Stylometric features. 2020 8th International Workshop on Biometrics and Forensics (IWBF), 1-6. Retrieved from https://arxiv.org/pdf/2003.11545.pdf

[9] Dronyuk I., Nazarkevych M., Fedevych O. (2016) Synthesis of Noise-Like Signal Based on Ateb-Functions. In: Vishnevsky V., Kozyrev D. (eds) Distributed Computer and Communication Networks. DCCN 2015. Communications in Computer and Information Science, vol 601. Springer, Cham https://doi.org/10.1007/978-3-319-30843-2_14

[10] Gomez Adorno, H. M., Rios, G., Posadas Durán, J. P., Sidorov, G., & Sierra, G. (2018). Stylometry-based approach for detecting writing style changes in literary texts. Computación y Sistemas, 22(1). Retrieved from http://www.scielo.org.mx/scielo.php?script=sci_arttext&pid

[11] Joshi, S., & Deshpande, D. (2018). Twitter Sentiment Analysis System. International Journal of Computer Applications (IJCA), 180(47), 35-39. Retrieved from https://arxiv.org/ftp/arxiv

[12] Killian, A., Brounstein, T., Skryzalin, J., & Garcia, D. (2019). Stylometric and Temporal Techniques for Social Media Account Resolution. Sandia National Laboratories Journal, 1-8.

https://www.osti.gov/servlets/purl/1456316 Kramer, S. (2020, June 9). Tracking coronavirus disinformation on Twitter.
[13] Kula, S., Choraś, M., Kozik, R., Ksieniewicz, P., & Woźniak, M. (2020). Sentiment Analysis for Fake News Detection by Means of Neural Networks. Springer, Cham. Retrieved from https://link.springer.com/chapter/10.1007/978-3-030- 50423-6_49
[14] M. Pasyeka, V. Sheketa, N. Pasieka, S. Chupakhina and I. Dronyuk, "System Analysis of Caching Requests on Network Computing Nodes," 2019 3rd International Conference on Advanced Information and Communications Technologies (AICT), Lviv, Ukraine, 2019, pp. 216-222, doi: 10.1109/AIACT.2019.8847909.
[15] Mariya Nazarkevych, Andrii Marchuk, Lesia Vysochan, Yaroslav Voznyi, Hanna Nazarkevych and Anzhela Kuza Ateb-Gabor Filtering Simulation for Biometric Protection Systems. CPITS 2020 pp. 14-22
[16] Medykovskyy M., Pasyeka M., Pasyeka N. & Turchyn O. (2017). Scientific research of life cycle perfomance of information technology. 12th International Scientific and Technical Conference on Computer Sciences and Information Technologies, CSIT 2017, , 1 425-428. doi:10.1109/STC-CSIT.2017.8098821
[17] Mishchuk, O., R. Tkachenko, & I. Izonin Missing Data Imputation through SGTM Neural-Like Structure for Environmental Monitoring Tasks. Advances in Intelligent Systems and Computing. Vol. 938. 2020, pp. 142-151, doi:10.1007/978-3-030-16621-2_13
[18] Mykhailyshyn H., Pasyeka N., Sheketa V., Pasyeka M., Kondur O. & Varvaruk M. (2021). Designing network computing systems for intensive processing of information flows of data doi:10.1007/978-3-030-43070-2_18
[19] N. Pasieka, V. Sheketa, Y. Romanyshyn, M. Pasieka, U. Domska & A. Struk «Models, Methods and Algorithms of Web System Architecture Optimization» IEEE International Scientific-Practical Conference Problems of Infocommunications, Science and Technology (PIC S&T), Kyiv, Ukraine, 2019, pp. 147-153, doi: 10.1109/PICST47496.2019.9061539.
[20] Nazarkevych M., Logoyda M., Dmytruk S., Voznyi, Y. & Smotr O. (2019). Identification of biometric images using latent elements. Paper presented at the CEUR Workshop Proceedings, 2488 pp. 99-108.
[21] Nazarkevych M., Logoyda M., Troyan O., Vozniy Y. & Shpak Z. (2019, September). The Ateb-Gabor Filter for Fingerprinting. In International Conference on Computer Science and Information Technology pp. 247-255. Springer, Cham.
[22] Nazarkevych M., Logoyda, M., Troyan, O., Vozniy, Y., & Shpak, Z. (2019, September). The Ateb-Gabor Filter for Fingerprinting. In Conference on Computer Science and Information Technologies (pp. 247-255). Springer, Cham.
[23] Nazarkevych M., Lotoshynska N., Klyujnyk I., Voznyi Y., Forostyna S. & Maslanych I. (2019, July). Complexity Evaluation of the Ateb-Gabor Filtration Algorithm in Biometric Security Systems. In 2019 IEEE 2nd Ukraine Conference on Electrical and Computer Engineering (UKRCON) pp. 961-964
[24] Nazarkevych M., Lotoshynska, N., Brytkovskyi, V., Dmytruk, S., Dordiak, V., & Pikh, I. (2019). Biometric identification system with ateb-gabor filtering. Paper presented at the 2019 11th International Scientific and Practical Conference on Electronics and Information Technologies, ELIT 2019 - Proceedings, 15-18. doi:10.1109/ELIT.2019.8892282
[25] Nazarkevych M., Oliiarnyk R., Nazarkevych H., Kramarenko O., & Onyshschenko I. (2016, August). The method of encryption based on Ateb-functions. In 2016 IEEE First International Conference on Data Stream Mining & Processing (DSMP) pp. 129-133.
[26] Nazarkevych, M., Oliarnyk, R., & Dmytruk, S. (2017, September). An images filtration using the Ateb-Gabor method. In 2017 12th International Scientific and Technical Conference on Computer Sciences and Information Technologies (CSIT) (Vol. 1, pp. 208-211
[27] Neal, T., Sundararajan, K., Fatima, A., Yan, Y., Xiang, Y., & Woodard, D. (2018). Surveying stylometry techniques and applications. ACM Computing Surveys, 50(6), 1-36. doi: 10.1145/3132039

[28] Nirkhi, S., Dharaskar, R., & Thakare, V. (2016). Authorship verification of online messages for forensic investigation. Procedia Computer Science, 78, 640-645. Retrieved from https://www.sciencedirect.com/science/article/pii/S1877050916001137
[29] Pascual, F. (2020, August 19). Twitter sentiment analysis with machine learning. MonkeyLearn Blog. Retrieved from https://monkeylearn.com/blog/sentiment- analysis-of-twitter/
[30] Pasieka, N., Sheketa, V., Romanyshyn, Y., Pasieka, M., Domska, U., & Struk, A. (2019). Models, methods and algorithms of web system architecture optimization. Paper presented at the 2019 IEEE International Scientific-Practical Conference: Problems of Infocommunications Science and Technology, PIC S and T 2019 – pp. 147-152. doi:10.1109/PICST47496.2019.9061539
[31] Pasyeka M., Sheketa V., Pasieka N., Chupakhina S. & Dronyuk, I. (2019). System analysis of caching requests on network computing nodes. 3rd International Conference on Advanced Information and Communications Technologies, AICT2019 - Proceedings, pp. 216-222, doi:10.1109/AIACT.2019.8847909
[32] Pasyeka M., Sheketa V., Pasieka N., Chupakhina S. & Dronyuk, I. (2019). System analysis of caching requests on network computing nodes. Paper presented at the 2019 3rd International Conference on Advanced Information and Communications Technologies, AICT 2019 - Proceedings, 216-222. doi:10.1109/AIACT.2019.8847909
[33] Riznyk O., Povshuk O., Kynash Y., Nazarkevich M., & Yurchak I. (2018). Synthesis of non-equidistant location of sensors in sensor network. 14th International Conference on Perspective Technologies and Methods in MEMS Design, MEMSTECH 2018 - Proceedings, 204-208. doi:10.1109/MEMSTECH.2018.8365734
[34] S. Babichev, A. Sharko, B. Durnyak, V. Zhydetskyy and I. Izonin, "Application of Huang Transform and Wavelet Analysis for Acoustic Emission Signal Filtering," 2019 IEEE 2nd Ukraine Conference on Electrical and Computer Engineering (UKRCON), 2019, pp. 859-863, doi: 10.1109/UKRCON.2019.8879839.
[35] Sharon Belvisi, N. M., Muhammad, N., & Alonso-Fernandez, F. (2020). Forensic authorship analysis of Microblogging texts using N-grams and Stylometric features. 2020 8th International Workshop on Biometrics and Forensics (IWBF), 1-6. doi: 10.1109/iwbf49977.2020.9107953
[36] Sikora, L., Lysa, N., Fedyna, B., Durnyak, B., Martsyshyn, R., & Miyushkovych, Y. (2018). Technologies of development laser based system for measuring the concentration of contaminants for ecological monitoring. Paper presented at the 2018 IEEE 13th International Scientific and Technical Conference on Computer Sciences and Information Technologies, CSIT 2018 - Proceedings, 1 93-96. doi:10.1109/STC-CSIT.2018.8526602
[37] Singh, T., & Kumari, M. (2016). Role of text pre-processing in Twitter sentiment analysis. Procedia Computer Science, 89, 549-554. https://doi.org/10.1016/j.procs.2016.06.095
[38] Tkachenko, R., Izonin, I., Kryvinska, N., Dronyuk, I., & Zub, K. (2020). An approach towards increasing prediction accuracy for the recovery of missing iot data based on the grnn-sgtm ensemble. Sensors (Switzerland), 20(9) doi:10.3390/s20092625
[39] Tkachenko, R., Izonin, I., Vitynskyi, P., Lotoshynska, N., & Pavlyuk, O. (2018). Development of the non-iterative supervised learning predictor based on the ito decomposition and sgtm neural-like structure for managing medical insurance costs. Data, 3(4) doi:10.3390/data3040046
[40] V. Buriachok, et al., Invasion Detection Model using Two-Stage Criterion of Detection of Network Anomalies, Cybersecurity Providing in Information and Telecommunication Systems (CPITS), pp. 23–32, Jul. 2020.
[41] Y. Romanyshyn, V. Sheketa, L. Poteriailo, V. Pikh, N. Pasieka and Y. Kalambet Social-communication web technologies in the higher education as means of knowledge transfer. IEEE 2019 14th International Scientific and Technical Conference on Computer Sciences and Information Technologies (CSIT). – Vol.3. – 2019. – Lviv, Ukraine. – pp. 35–39.
[42] Zharikova M. & Sherstjuk, V. (2017). “Academic integrity support system for educational institution,” 2017 IEEE 1st Ukraine Conference on Electrical and Computer Engineering, UKRCON 2017 - Proceedings, 1212-1215. doi:10.1109/UKRCON.2017.8100445