International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 08 Issue: 05 | May 2021 www.irjet.net p-ISSN: 2395-0072

# De-Anonymization of Electronic Mail

**Akshay Goel[1], Mandeep Singh Narula[2]**

[1]*Student, Dept. of ECE, Jaypee Institute of Information Technology, Noida, India*
[2]*Assistant Professor, Dept. of ECE, Jaypee Institute of Information Technology, Noida, India*

---

**Abstract -** *Everyone has their own style of writing, the way of focusing on a particular part of the sentence, vocabulary level and most of the time this can be used to infer the author of a particular text. This is affected by one's learning source, region, country and from style one often read etc. Often one's writing style, it really is a term that is used to describe the author of a specific document. Electronic mail assumes a significant function in day by day close to home correspondence. Likewise, email is regularly used to send official warning. Email makes it simpler for individuals to impart, yet tragically, it is more defenseless to assault[11] simultaneously. It thus becomes necessary to develop a mechanism that can de-anonymize such emails and identify any sort of spam, if present. One of the ongoing and rising patterns in initiation examination is stylometric[8] data is extracted using a computer programmed, highlights from the content of any text and look for a robust classification technique that can identify author of a text based on its content.*

***Key Words***: Stylometric, Lexical, Syntactic, CNN, NLP, Recall, Precision.

## 1.INTRODUCTION

The proposed solution approach is a complete pipeline for making a secure email platform. The first phase of the complete solution is the de-anonymization of electronic mail which will then be followed by author verification, author profiling and abuse detection.

Research has been done in the field using basic NLP techniques which perform well but still does not give desired outputs. The proposed approach utilises a combination of NLP and deep learning techniques for better results. The proposed approach takes into account various discriminating stylometric features like ratio of unique words to total words, number of adjectives, average word-length, average sentence-length, total number of characters, total number of function words, total number of personal pronouns, total number of adjectives etc. The work is based not only on lexical features but syntactic features as well.

No work has been done in the past where a complete approach has been presented for a secure email platform. The first task of authorship attribution[6] in electronic mail is proposed to be done using a convolutional neural network and the second task of spam detection is done using a robust classification model.

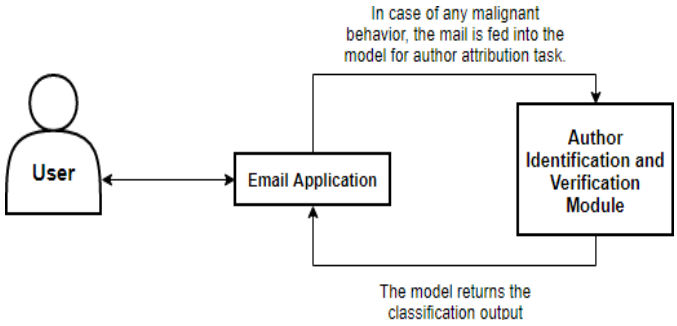


**Fig-1**: Author Identification/Verification

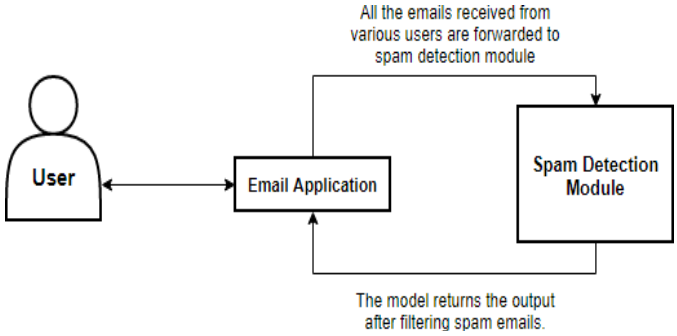


**Fig -2**: Spam Detection

### 1.1 Proposed Solution

The project can be summarised as a text-based analysis of email data by extracting various lexical and syntactic features and then using these features proceed to use cases like finding author of an email, author verification and spam detection. Although the basic approach can be extended to achieve further tasks such as priority assignment and abusive content detection, the current proposed solution takes into account the first two important tasks to build a secure pipeline. Figure-1 & 2 depicts the flow chart of propose approach.

Author Identification/Attribution: It is the task of identifying the author of a given email from a set of suspects. The main concern of the task is to define an appropriate characterization of text that capture the writing styles of authors.

Author Verification: It is an extension of the author attribution problem where the problem statement is to verify whether a particular email was written by the sender who claims to have sent it.

Spam Detection: It is the task of filtering emails based on the content of the mail. It detects unwanted and unsolicited mails.

### 1.2 Author Identification/Attribution

The proposed approach aims to implement CNN based classification model for the purpose of authorship identification. The CNN[2] can identify commonly used groups of words and phrases by an author. There are three layers to the CNN. Figure-3 Author Identification

1. First, an embedding layer is present.
2. Second, the convolution layer performs the convolution operation using 128 5x5 filter
3. Third, the dense layer is used for classification

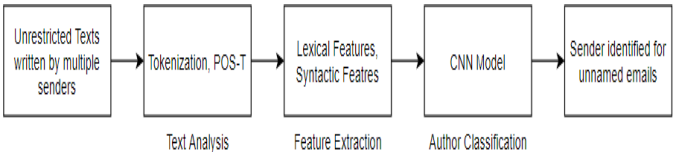


**Fig -3**: Author Identification

It is the task of identifying[3] the author of a given email from a set of suspects for which we already have a stored database that is used to analyse the writing style of a particular person. The work is based not only on lexical features but syntactic features as well. The various extracted features include

I. Ratio of unique words to total words: A total word count is maintained for the body of the email and the number of unique words is identified and used as a feature of author writing style.

II. Number Of adjectives: The use of adjectives in the sentence is found to be unique by a number of studies and hence, this feature is used.

III. Average word length: It has been observed by authors that some people tend to use longer words than others and hence, average word length vary extensively among users.

IV. Average sentence-length: Just like average word length, average sentence length is a unique feature for each person.

V. Total number of characters: Some people like to mention only relevant information while others ramble on and about a lot.

VI. Total number of personal pronouns: Just like the use of function words, use of personal pronouns is also an indicator of nature of the style of writing. Some people tend to talk in third person, while others prefer the use of personal pronouns.

### 1.3 Author Verification

The task of the author is an extension to the task of author identification[1] only. While identifying an author for a given email, we make use of the feature extracted and try to classify the unknown email into a known author category. However, while verification of the author, we need to check whether the particular mail was actually sent by the person who claims to have sent it. The task of author verification can be leveraged by making use of the already available sender name and details and that information could be used to optimize the search space. Figure-4 Author verification

Here, for a given email with a sender address that is stored in the database, the developed model of author identification is used and the output received is cross-referenced with the sender who claims to have sent the mail.

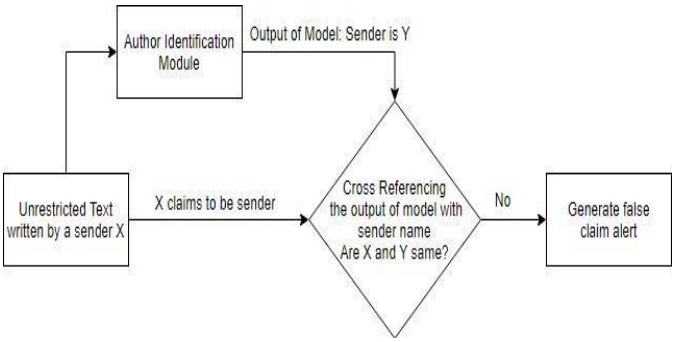


**Fig -4**: Author Verification

### 2.3 Spam Detection

The second module is the spam filtering module which detects any unwanted mail possibly from a spammer. Broadly speaking, mail filtering is the method of organizing mail based on a set of parameters. The word can refer to human intelligence intervention, but it most often refers to the automated processing of incoming message with resentful-spam techniques, which includes both outgoing and incoming emails. Figure-5

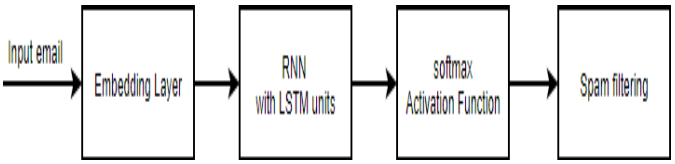


**Fig -5**: Spam Detection

The 1- point is actively-trained embedding zone that converts each word into a 100-dimensional N-dimensional vector of real numbers. The vectors of two terms with identical meanings are usually very near. A recurrent neural network with LSTM units makes up the second layer. Finally, the output layer consists of two neurons, each of which represents "spam" - "non-spam". The model is trained for 20 epochs.

### 1.4 Requirement Analysis

**Efficient Feature Extraction -**The requirement is to process an electronic mail content which includes both the body and metadata from the sender.
- It should be able to remove the metadata from the content of the mail.

- It must be feasible to select the relevant features from the body of the mail.

- It shall have the potential to be able to extract those features from the said mail body and store them in a proper organised manner for further use.

**Identifying the Author** - The requirement is to identify the sender of the mail based on the body of the mail. The following requirements are to be satisfied in order to fulfil this requirement:
- It must be feasible to load the extracted features from a given path for classification purpose.

- It should be able to attribute the author associated with the mail.

- It must be attainable to record the sender's name in the database and use this new information for further training of the model.

- It shall have the potential to cross-reference this identified author name with the sender that claims to have sent the mail.

- It shall produce an alert when a mismatch is found during the verification purpose.

**Detecting Spams** - The requirement is to filter the electronic mails that can be potential spam messages. The following requirements are to be satisfied in order to fulfil this requirement:

- It must be feasible to load the mail content and metadata from the given path.

- It should be able to detect the presence of unwanted or potentially threatening content.

- It must be attainable to redirect the spam emails to a separate location.

**Table-1:** List of test cases

| S No | Component tested | Input | Expected Output | Status |
|---|---|---|---|---|
| 1 | Feature Extraction | Feed the electronic mail text input along with metadata. | Exclusion of metadata and list of extracted features for all the emails. | Pass |
| 2 | Author Identification and Verification | Extracted features from the first phase of pre-processing for the test case. | Correctly identify the author of the email based on the feature and verify it with the actual sender. | Pass |
| 3 | Spam Classification | E-mail text input for the test email. | Correctly identify the email identifying as spam and irrelevant. | Pass |

## 2. LIMITATIONS

The proposed model, although sound in its implementation, still has some limitations

1. Not able to identify if an email belongs to anonymous author outside the testing data, to identify the email as unknown. i.e. the model is not able to identify the entry by a new user and tries to adjust the text association with an existing author from the database. Rather, it should learn from the entry of a new text and modify the algorithm accordingly to accommodate the new author writing style.

2. Classification for texts with small sizes such as single line emails could not be achieved.

3. It does not consider gradual change over time as an author's writing style evolves.

4. As the size of the dataset increases, and a greater number of authors are incorporated in the company, the complexity of the problem increases and time efficiency decreases.

## 3. FINDING

The model performs well in the given scenario. The evaluation metrics used for the various models differ according to the functionality that they serve. Accuracy score is calculated for the author identification and verification which is an indicator of the goodness of fit of the model. Accuracy of 78.23% is achieved by the model, indicating that out of 100 test emails, the model was able to correctly identify the authors of about 78 emails.

For the spam classification/filtering model, the evaluation metrics like accuracy, precision and recall were calculated. Figure-6 result of spam email.

The results obtained from the model were promising. An accuracy score of 98.21% was observed whereas the precision value was 99.1 % and the recall value was 98.75 %, Here, the value of F score is found to be 0.98 which is a promising outcome.

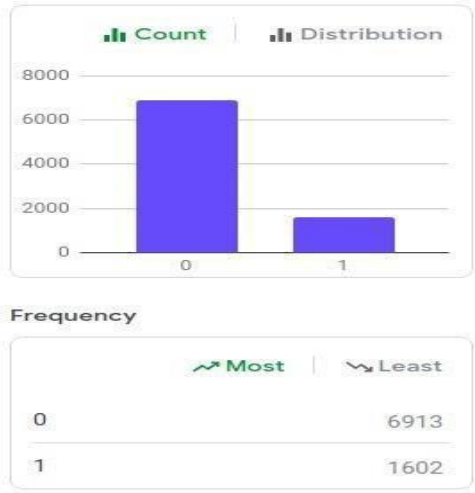


**Fig -6**: Count and Frequency of Spam v/s Non-Spam emails

## 4. CONCLUSIONS

Given an email as an input, we were able to successfully remove metadata from the body of the email and extract the relevant features for classification purpose. An extensive study was carried out to identify the optimal features which could bring about the maximum accuracy and yet does not increase the dimensionality of the CNN[2] model. Using the eight identified features, we were able to successfully understand the practice of stylometry and correctly identify the sender of the mail based on the content.

The task of author verification was carried out by cross referencing the name of the sender and the identified author which can be used in the cases where a derogatory/abusive/inappropriate email was sent within an organization. In such cases, now the author can't make a false claim of the email account being hacked or falsify evidence regarding the same.

The final task of spam detection was achieved by use of appropriate algorithm and a model is developed to filter any unsolicited and unwanted email. The emails which possibly contain virus threats or security threats are also identified. It is observed through evaluation metrics, that the performance of the model is satisfactory and in close competition with the existing approaches as proposed by various other researchers.

## REFERENCES

[1] Fang, Yong, Yue Yang, and Cheng Huang. "EmailDetective: An Email Authorship Identification And Verification Model." The Computer Journal 63.11 (2020): 1775-1787.

[2] Ma, Weicheng, et al. "Towards Improved Model Design for Authorship Identification: A Survey on Writing Style Understanding." arXiv preprint arXiv:2009.14445 (2020).

[3] Radhakrishnan Iver, Rahul, and Carolyn Penstein Rose. "A Machine Learning Framework for Authorship Identification From Texts." arXiv e-prints (2019): arXiv-1912.

[4] Tamboli, Mubin Shoukat, and Rajesh S. Prasad. "Authorship identification with multi sequence word selection method." International Conference on Intelligent Systems Design and Applications. Springer, Cham, 2018.

[5] Ahmad, Sumnoon Ibn, Lamia Alam, and Mohammed Moshiul Hoque. "An Empirical Framework to Identify Authorship from Bengali Literary Works." International Conference on Cyber Security and Computer Science. Springer, Cham, 2020.

[6] Martins, Ricardo, et al. "A sentiment analysis approach to increase authorship identification." Expert Systems (2019): e12469.

[7] Hinh, Robert, Sangmi Shin, and Julia Taylor. "Using frame semantics in authorship attribution." 2016 IEEE International Conference on Systems, Man, and Cybernetics (SMC). IEEE, 2016.

[8] Abbasi, Ahmed, and Hsinchun Chen. "Writeprints: A stylometric approach to identity-level identification and similarity detection in cyberspace." ACM Transactions on Information Systems (TOIS) 26.2 (2008): 1-29.

[9] Alhijawi, Bushra, Safaa Hriez, and Arafat Awajan. "Text-based authorship identification-A survey." 2018 Fifth International Symposium on Innovation in Information and Communication Technology (ISIICT). IEEE, 2018.

[10] Anwar, Waheed, Imran Sarwar Bajwa, and Shabana Ramzan. "Design and implementation of a machine learning-based authorship identification model." Scientific Programming 2019 (2019).

[11] Kaur, Ravneet, Sarbjeet Singh, and Harish Kumar. "Authorship analysis of online social media content." Proceedings of 2nd International Conference on Communication, Computing and Networking. Springer, Singapore, 2019.

[12] Nizamani, Sarwat, and Nasrullah Memon. "CEAI: CCM-based email authorship identification model." Egyptian Informatics Journal 14.3 (2013): 239-249.