Journal of Harbin Engineering University
ISSN: 1006-7043

Vol 45 No. 5
May 2024

Privacy-Preserving Ensemble Methods for Email Spam Detection

[1]t.Harikala, [2] D. Sreevidya, [3] Mahanandi .Y., [4] Dr. N Penchalaiah4
[1]Assistant Professor , dept.of ECE,
Annamacharya Institute of Technology and Science, Rajampeta, Andhra pradesh.
[2] Assistant Professor,
CSE (CyS,DS) and AI&DS,
VNRVJIET college
[3]Assistant Professor , Dept.of.CSE
G Pulla Reddy College of Engineering(Autonomous),
Kurnool, Andhra Pradesh.
[4]Associate Professor ,Dept of CSE
Annamacharya Institute of Technology and Science,
Rajampeta, Andhra pradesh.

Abstract— Email is a private medium of communication, and the inherent privacy constraints form a major obstacle in developing effective spam filtering methods which require access to a large amount of email data belonging to multiple users. To mitigate this problem, we envision a privacy-preserving spam filtering system, where the server can train and evaluate a logistic regression-based spam classifier on the combined email data of all users without being able to observe any emails. This is achieved using primitives such as homomorphic encryption and randomization. We analyze the protocols for correctness and security and perform experiments with a prototype system on a large-scale spam filtering task. State-of-the-art spam filters often use character n-grams as features, resulting in large sparse data representations that are not feasible to use directly with our training and evaluation protocols. We explore various data-independent dimensionality reduction techniques to decrease the running time of the protocol, making it feasible to use in practice while achieving high accuracy.

Keywords— Privacy-preserving, Spam detection, Logistic regression, Homomorphic encryption, Randomization, Dimensionality reduction, Email security etc.,

Introduction

Email remains a cornerstone of digital communication, facilitating vast amounts of personal and professional correspondence. However, its utility is persistently challenged by the relentless influx of spam. Effective spam filtering is thus essential to preserving the usability and security of email platforms. Yet, the development of robust spam detection methods faces a significant hurdle: privacy constraints. Traditional spam filtering approaches rely on access to large datasets of email content, often spanning multiple users. However, the sensitive nature of email content raises legitimate privacy concerns, particularly in light of stringent data protection regulations. Balancing the need for effective spam detection with the imperative to safeguard user privacy necessitates innovative solutions. In this context, we propose a novel approach: a privacy-preserving spam filtering system. Our system is designed to enable the server to train and evaluate a logistic regression-based spam classifier using the combined email data of all users, without compromising the privacy of individual emails. This is achieved through the application of advanced cryptographic techniques such as homomorphic encryption and randomization. In this paper, we present a comprehensive analysis of the protocols underpinning our privacy-preserving spam filtering system, addressing both correctness and security considerations. Furthermore, we provide empirical validation through experiments conducted on a large-scale spam filtering task, demonstrating the feasibility and efficacy of our approach. An additional challenge in implementing privacy-preserving spam filtering systems lies in the computational overhead associated with processing large, sparse email datasets. State-of-the-art spam filters often utilize character n-grams as features, leading to cumbersome data representations unsuitable for direct use with privacy-preserving protocols. To address this challenge, we explore various data-independent dimensionality reduction techniques, aimed at optimizing the runtime of our protocols while maintaining high levels of accuracy. By combining privacy-preserving methodologies with innovative dimensionality reduction techniques, our approach represents a significant step towards reconciling the competing imperatives of effective spam detection and user privacy in the realm of email communication. In the realm of digital communication, email stands as a cornerstone, facilitating a myriad of personal and professional interactions. However, its utility is consistently undermined by the pervasive presence of spam, which inundates inboxes with unsolicited and often malicious content. The development of robust spam detection mechanisms is thus imperative to maintain the usability and security of email platforms. Yet, traditional approaches to spam filtering confront a significant obstacle: the tension between the need for effective filtering mechanisms and the imperative to uphold user privacy.

Conventional spam filtering techniques typically rely on access to extensive datasets comprising email content from multiple users. These datasets are used to train machine learning models that distinguish between legitimate and spam emails based on various features and patterns. However, the sensitive nature of email content raises profound privacy concerns. Users are understandably reluctant to disclose the contents of their emails, even for the purpose of improving spam detection algorithms. Consequently, the development of effective spam filters is hindered by the scarcity of labeled data that can be safely and ethically utilized.

The challenges posed by privacy constraints have spurred research efforts aimed at devising privacy-preserving spam filtering systems. However, previous literature has grappled with several key challenges:

1. *Privacy Preservation*: Ensuring that the privacy of individual email contents is preserved throughout the filtering process poses a formidable technical challenge, particularly when employing traditional machine learning techniques that require access to plaintext data.
2. *Scalability*: Many existing privacy-preserving spam filtering systems struggle to scale effectively, particularly when tasked with processing large volumes of email data from diverse sources. The computational overhead associated with cryptographic operations and data processing can impede real-time performance.
3. *Accuracy:* Balancing the imperatives of privacy preservation and spam detection accuracy is a delicate endeavor. Some privacy-preserving approaches sacrifice classification performance in favor of stronger privacy guarantees, leading to suboptimal filtering outcomes.

Motivated by the pressing need to reconcile the competing demands of effective spam detection and user privacy, we present a novel approach to privacy-preserving spam filtering. Our work seeks to address the limitations of existing methodologies by leveraging advanced cryptographic techniques and innovative dimensionality reduction strategies. By doing so, we aim to enable the development of spam filtering systems that offer robust protection of user privacy without compromising on filtering performance.

In this paper, we set out to achieve the following objectives:

1. *Design and Implementation:* Develop a privacy-preserving spam filtering system capable of training and evaluating machine learning models on encrypted email data without revealing the contents of individual emails.
2. *Analysis and Validation:* Conduct a comprehensive analysis of the protocols underpinning our approach, addressing both correctness and security considerations. Validate the effectiveness of our methodology through empirical experiments conducted on real-world email datasets.
3. *Dimensionality Reduction:* Explore various data-independent dimensionality reduction techniques aimed at optimizing the runtime of our privacy-preserving protocols while maintaining high levels of classification accuracy.

This work makes several contributions to the field of privacy-preserving spam filtering:

1. We propose a novel approach to spam filtering that prioritizes user privacy without compromising on filtering performance.
2. We provide a detailed analysis of the protocols and techniques employed in our privacy-preserving framework, addressing key correctness and security considerations.
3. We empirically validate the efficacy of our approach through experiments conducted on real-world email datasets, demonstrating its feasibility and performance in practice.
4. We explore innovative dimensionality reduction techniques aimed at optimizing the computational efficiency of our protocols while preserving classification accuracy.

The organizational framework of this study divides the research work in the different sections. The Literature survey is presented in section 2. In section 3 and 4 discussed about Existing and Proposed system methodologies. Further, in section 5 shown Results is discussed and. Conclusion and future work are presented by last sections 6.

Literature Survey

Pfleeger, S.L., Bloom, G. (2005). "Canning spam: proposed solutions to unwanted email." This paper discusses various proposed solutions to the problem of unwanted email, offering insights into early approaches to spam filtering and email security.[1]

Grier, C., Thomas, K., Paxson, V., & Zhang, M. (2010). "@ spam: the underground on 140 characters or less." This work delves into the phenomenon of spam on social media platforms, exploring the tactics used by spammers and the challenges they pose to online security.[2]

Agarwal, D.K., Kumar, R. (2016). "Spam filtering using SVM with different kernel functions." This paper investigates the use of Support Vector Machines (SVM) with various kernel functions for spam filtering, highlighting the efficacy of machine learning algorithms in combating spam.[3]

Heartfield, R., Loukas, G. (2015). "A taxonomy of attacks and a survey of defence mechanisms for semantic social engineering attacks." This study presents a taxonomy of attacks and defense mechanisms against semantic social engineering attacks, providing insights into the broader landscape of cybersecurity.[4]

John, J. P., Moshchuk, A., Gribble, S. D., & Krishnamurthy, A. (2009). "Studying spamming botnets using botlab." This research focuses on studying spamming botnets, offering valuable insights into the behavior and operation of malicious bot networks.[5]

Kumar, N., & Sonowal, S. (2020). "Email spam detection using machine learning algorithms." This paper explores the application of machine learning algorithms for email spam detection, presenting experimental results and performance evaluations.[6]

Junnarkar, A., Adhikari, S., Fagania, J., Chimurkar, P., & Karia, D. (2021). "E-mail spam classification via machine learning and natural language processing." This study investigates email spam classification using machine learning and natural language processing techniques, contributing to the advancement of spam filtering methodologies.[7]

Awad, W.A., ELseuofi, S.M. (2011). "Machine learning methods for spam e-mail classification." This work explores various machine learning methods for spam email classification, providing insights into the strengths and limitations of different approaches.[8]

Zhang, F., Chan, P.P., Biggio, B., Yeung, D.S., Roli, F. (2015). "Adversarial feature selection against evasion attacks." This paper addresses adversarial attacks against feature selection methods, contributing to the development of more robust spam filtering systems resilient to evasion tactics.[9]

Shaukat, K., Luo, S., Chen, S., & Liu, D. (2020). "Cyber threat detection using machine learning techniques: A performance evaluation perspective." This study evaluates the performance of machine learning techniques for cyber threat detection, offering insights into the effectiveness of different approaches in addressing cybersecurity challenges.[10]

Garavand, A., Salehnasab, C., Behmanesh, A., Aslani, N., Zadeh, A.H., Ghaderzadeh, M. (2022). "Efficient model for coronary artery disease diagnosis: a comparative study of several machine learning algorithms." This study investigates the application of various machine learning algorithms for coronary artery disease diagnosis, contributing to the field of medical diagnostics and showcasing the effectiveness of different models.[11]

Ghaderzadeh, M., Aria, M., Asadi, F. (2021). "X-ray equipped with artificial intelligence: changing the COVID-19 diagnostic paradigm during the pandemic." This work explores the integration of artificial intelligence into X-ray imaging for COVID-19 diagnosis, highlighting the transformative potential of AI in healthcare during the pandemic.[12]

Hajek, P., Barushka, A., Munk, M. (2020). "Fake consumer review detection using deep neural networks integrating word embeddings and emotion mining." This study focuses on detecting fake consumer reviews using deep neural networks, incorporating word embeddings and emotion mining techniques for improved accuracy.[13]

Ramanathan, V., Wechsler, H. (2013). "Phishing detection and impersonated entity discovery using conditional random field and latent Dirichlet allocation." This research addresses phishing detection and impersonation entity discovery using advanced machine learning techniques, contributing to the field of cybersecurity.[14]

Ghourabi, A., Mahmood, M.A., Alzubi, Q.M. (2020). "A hybrid CNN-LSTM model for SMS spam detection in Arabic and English messages." This study proposes a hybrid CNN-LSTM model for detecting SMS spam in both Arabic and English messages, offering a cross-lingual approach to spam detection.[15]

Madhavan, M. V., Pande, S., Umekar, P., Mahore, T., & Kalyankar, D. (2021). "Comparative analysis of detection of email spam with the aid of machine learning approaches." This work provides a comparative analysis of machine learning approaches for email spam detection, offering insights into the effectiveness of different methodologies.[16]

Rayan, A. (2022). "Analysis of e-mail spam detection using a novel machine learning-based hybrid bagging technique." This study analyzes email spam detection using a novel machine learning-based hybrid bagging technique, contributing to the advancement of spam filtering methodologies.[17]

Suborna, A.K., Saha, S., Roy, C., Sarkar, S., & Siddique, M.T.H. (2021). "An approach to improve the accuracy of detecting spam in online reviews." This research presents an approach to enhance the accuracy of spam detection in online reviews, addressing challenges in identifying fraudulent or misleading content.[18]

Frías-Blanco, I., Verdecia-Cabrera, A., Ortiz-Díaz, A., & Carvalho, A. (2016). "Fast adaptive stacking of ensembles." This work introduces a fast adaptive stacking method for ensembles, improving the efficiency of ensemble learning techniques.[19]

El-Kareem, A., Elshenawy, A., Elrfaey, F. (2017). "Mail spam detection using stacking classification." This study explores mail spam detection using stacking classification techniques, demonstrating the effectiveness of ensemble methods in identifying spam emails.[20]

I. Existing Method

The existing method, which addresses the challenge of privacy constraints in email spam filtering, presents a novel approach to training and evaluating a logistic regression-based spam classifier without compromising user privacy. By leveraging techniques such as homomorphic encryption and randomization, the system enables the server to analyze combined email data from multiple users without accessing individual email contents. However, this method encounters significant challenges that need to be addressed for its successful implementation.

Firstly, the ensemble approach utilized in the system introduces increased computational complexity. This arises from the necessity of coordinating and communicating among multiple models or entities involved in the ensemble. The process of sharing encrypted features or representations and aggregating decisions from different models adds computational overhead, potentially impacting the system's efficiency and scalability.

Secondly, the integration of multiple models into the ensemble poses challenges in terms of design and calibration. Ensuring compatibility, feature alignment, and synchronization of decision-making processes among the diverse models can be both intricate and time-consuming. Overcoming these challenges is paramount to realizing the full potential of the proposed privacy-preserving spam filtering system while maintaining robust user privacy protections.

II. Proposed Method

In this project, a distributed Hadoop system is employed to store encrypted email messages, enhancing the security of the email data. By storing the emails in encrypted form, the system mitigates the risk of unauthorized access or breaches, as even if a hacker were to gain access to the Hadoop system, they would only encounter encrypted data, ensuring the confidentiality of the email contents. The proposed method utilizes a Voting Classifier algorithm on the encrypted email data stored in Hadoop to train a spam detection model. The Voting Classifier employs six different classifiers, including Random Forest, Decision Tree, MLP, SGD, Naïve Bayes, and Bagging Classifier. This ensemble approach allows the system to leverage the strengths of each individual classifier and make decisions based on the consensus or voting outcome, thereby enhancing the accuracy and robustness of the spam detection model.

One of the primary advantages of this proposed system is the enhanced privacy protection it offers. By employing advanced privacy-preserving techniques such as differential privacy, secure multiparty computation, or federated learning, sensitive user data and email content remain encrypted and hidden, even during collaborative model training and decision-making processes. This ensures that user privacy is maintained throughout the system's operations.

Additionally, the ensemble of models contributes to improved spam detection accuracy and robustness. By combining multiple classifiers and leveraging their collective intelligence, the system can effectively identify and classify spam emails while minimizing false positives and false negatives. This results in a more reliable and effective spam detection mechanism, enhancing the overall security of email communication.. System architecture shown in figure 1.

A. System Architecture

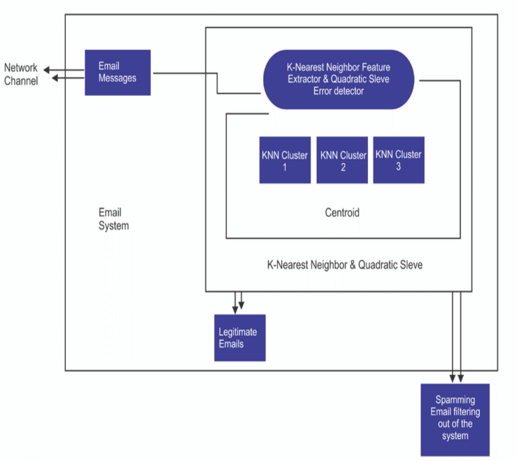


Fig. 1. **System Architecture**

B. Implementation

1) *Hadoop Read Encrypted Dataset:*

Hadoop is a distributed file system that allows for the storage and processing of large volumes of data across multiple nodes. In this step, we use PySpark, a Python API for Apache Spark, to interact with Hadoop and read the encrypted dataset. The encrypted dataset is stored in the Hadoop Distributed File System (HDFS), and we use PySpark's read function to load the dataset into a Spark DataFrame. The encrypted dataset contains email data that has been encrypted to ensure privacy and security.

2) *Process Dataset:*

After loading the encrypted dataset, we perform preprocessing steps to prepare the data for training the spam detection model. This may involve tasks such as feature engineering, where we extract relevant features from the email data, data cleaning to handle missing or inconsistent values, and encoding categorical variables into numerical format. In this example, we use PySpark's VectorAssembler to convert the features into a vector format that can be used by machine learning algorithms.

3) *Train & Test Split:*

Once the dataset has been processed, we split it into training and testing subsets. The training subset is used to train the machine learning model, while the testing subset is used to evaluate its performance. Typically, the data is split randomly, with a certain percentage allocated to each subset. In this example, we use an 80-20 split, where 80% of the data is used for training and 20% for testing.

4) Train Ensemble Voting Classifier:

Next, we train an ensemble voting classifier using scikit-learn or other machine learning libraries. The ensemble classifier combines the predictions of multiple base classifiers to make a final prediction. In this example, we use a variety of base classifiers such as Decision Tree, Random Forest, MLP, SGD, Naive Bayes, and Bagging Classifier. The ensemble classifier aggregates the predictions of these base classifiers using a voting mechanism (e.g., 'hard' voting), where the majority prediction is selected as the final prediction.

5) Accuracy Graph:

After training the ensemble voting classifier, we evaluate its performance on the test data and visualize the accuracy using a graph. We use scikit-learn's accuracy_score function to calculate the accuracy of the model's predictions compared to the actual labels in the test data. The accuracy graph provides a visual representation of how the model's accuracy changes over iterations or epochs.

6) Predict Spam from Test Data:

Finally, we use the trained ensemble voting classifier to predict spam from the test data. The model predicts whether each email in the test dataset is spam or not based on its features. These predictions can then be further evaluated and analyzed to assess the performance of the spam detection model.

Overall, this process demonstrates how Hadoop can be used to read an encrypted dataset, preprocess the data, train an ensemble voting classifier for spam detection, evaluate its performance, and make predictions on new data.

C. Machine Learning Algortihms

In this project, the focus is on leveraging a distributed Hadoop system to enhance the security of email data by storing it in encrypted form. The use of encryption mitigates the risk of unauthorized access or breaches, as even if a hacker gains access to the Hadoop system, they would only encounter encrypted data, maintaining the confidentiality of email contents.

The proposed method utilizes a Voting Classifier algorithm on the encrypted email data stored in Hadoop to train a spam detection model. This ensemble approach involves integrating six different classifiers:

1. *Random Forest:* A versatile ensemble learning method that constructs multiple decision trees during training and outputs the mode of the classes or the mean prediction of the individual trees.
2. *Decision Tree:* A tree-like model where each internal node represents a decision based on the value of a feature, and each leaf node represents a class label.
3. *MLP (Multi-layer Perceptron):* A type of artificial neural network with multiple layers of nodes interconnected by weighted edges. MLP classifiers can learn complex non-linear relationships in the data.
4. *SGD (Stochastic Gradient Descent) Classifier:* A linear model trained using stochastic gradient descent, suitable for large-scale datasets and online learning scenarios.
5. *Naïve Bayes Classifier:* A probabilistic model based on Bayes' theorem with the assumption of feature independence. It is simple, fast, and effective for text classification tasks.
6. *Bagging Classifier:* An ensemble learning technique that trains multiple base classifiers independently on different subsets of the training data and aggregates their predictions. Bagging reduces variance and improves the stability and generalization performance of the model.

The ensemble approach allows the system to leverage the strengths of each individual classifier and make decisions based on the consensus or voting outcome. This enhances the accuracy and robustness of the spam detection model, enabling it to effectively identify and classify spam emails while minimizing false positives and false negatives. Additionally, the proposed system offers enhanced privacy protection by employing advanced privacy-preserving techniques such as differential privacy, secure multiparty computation, or federated learning. This ensures that sensitive user data and email content remain encrypted and hidden, even during collaborative model training and decision-making processes, thereby maintaining user privacy throughout the system's operations.

D. Implementation

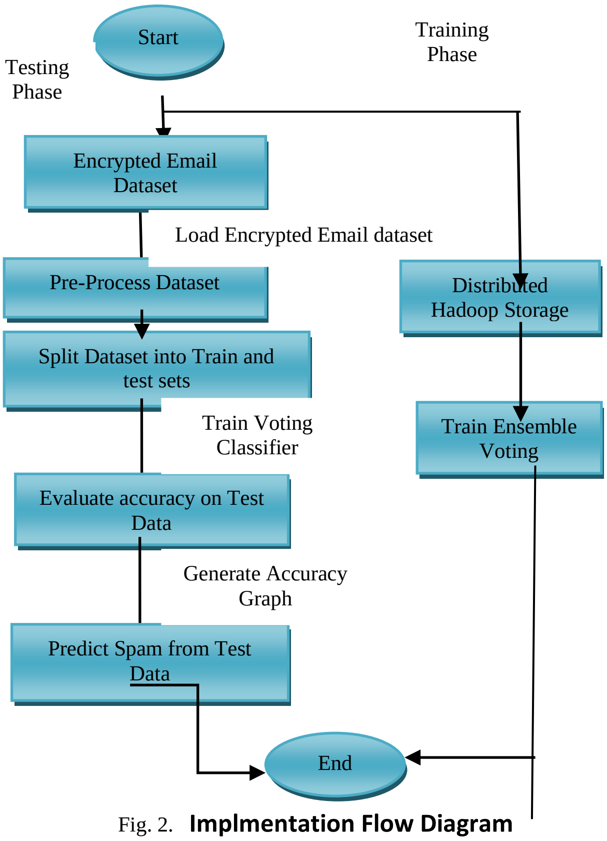

Start
Testing Phase
Training Phase
Encrypted Email Dataset
Load Encrypted Email dataset
Pre-Process Dataset
Distributed Hadoop Storage
Split Dataset into Train and test sets
Train Voting Classifier
Train Ensemble Voting
Evaluate accuracy on Test Data
Generate Accuracy Graph
Predict Spam from Test Data
End

Fig. 2. **Implmentation Flow Diagram**

This flow diagram outlines the process step-by-step:

1. The encrypted email dataset is stored in a distributed Hadoop storage system to enhance security.
2. The encrypted dataset is loaded into a Spark DataFrame for processing.
3. Preprocessing steps such as feature engineering, data cleaning, and encoding are applied to prepare the dataset for training.
4. The dataset is split into training and testing subsets.
5. The ensemble voting classifier, consisting of various algorithms, including Random Forest, Decision Tree, MLP, SGD, Naïve Bayes, and Bagging Classifier, is trained using the training data.
6. The accuracy of the model is evaluated on the test data.
7. An accuracy graph is generated to visualize the performance of the model.
8. The trained model is used to predict spam emails from the test data.

This flow diagram illustrates the workflow of the proposed method, from data storage and preprocessing to model training, evaluation, and prediction.

E. Performance Metrics

Performance measures are used to evaluate the network performance of the proposed model. This work uses Test accuracy and Test loss.

a) Test Accuracy:

Test accuracy measures the proportion of correctly classified instances out of the total instances in the test dataset. It is calculated as the ratio of the number of correctly classified samples

$$Accuracy = \frac{Number\ of\ correct\ predictions}{Total\ Number\ of\ Predictions} \quad (1)$$

III. Results and Discussion

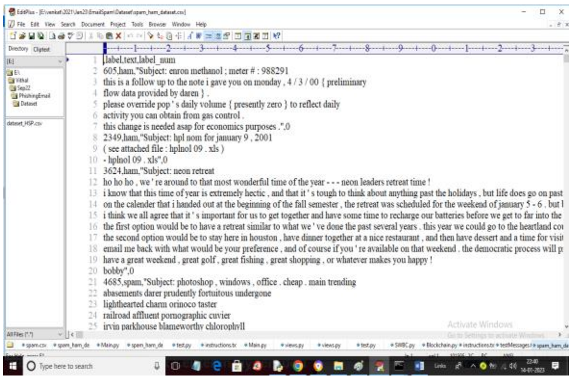

```
label,text,label_num
605,ham,"Subject: enron methanol ; meter # : 988291
this is a follow up to the note i gave you on monday , 4 / 3 / 00 { preliminary
flow data provided by daren } .
please override pop ' s daily volume { presently zero } to reflect daily
activity you can obtain from gas control .
this change is needed asap for economics purposes .",0
2349,ham,"Subject: hpl nom for january 9 , 2001
( see attached file : hplnol 09 . xls )
- hplnol 09 . xls",0
3624,ham,"Subject: neon retreat
ho ho ho , we ' re around to that most wonderful time of the year - - - neon leaders retreat time !
i know that this time of year is extremely hectic , and that it ' s tough to think about anything past the holidays , but life does go on past
on the calender that i handed out at the beginning of the fall semester , the retreat was scheduled for the weekend of january 5 - 6 , but l
i think we all agree that it ' s important for us to get together and have some time to recharge our batteries before we get to far into the
the first option would be to have a retreat similar to what we ' ve done the past several years . this year we could go to the heartland co
the second option would be to stay here in houston , have dinner together at a nice restaurant , and then have dessert and a time for visi
email me back with what would be your preference , and of course if you ' re available on that weekend . the democratic process will p
have a great weekend , great golf , great fishing , great shopping , or whatever makes you happy !
bobby",0
4685,spam,"Subject: photoshop , windows , office . cheap . main trending
abasements darer prudently fortuitous undergone
lighthearted charm orinoco taster
railroad affluent pornographic cuvier
irvin parkhouse blameworthy chlorophyll
```

Fig. 3. **Plain Text**

In above plain dataset we have text column which contains EMAIL message and label column which contains target label as 'HAM' (HAM means no spam) and SPAM. Now after applying encryption will get below encrypted data

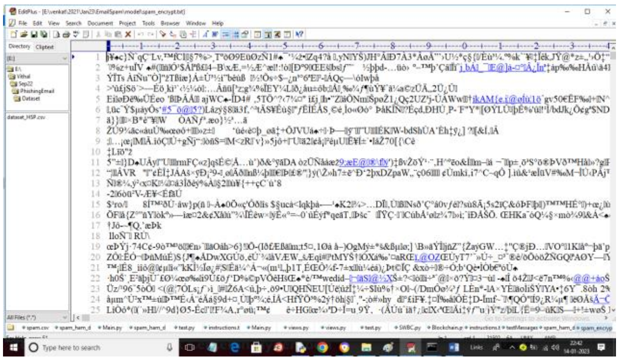
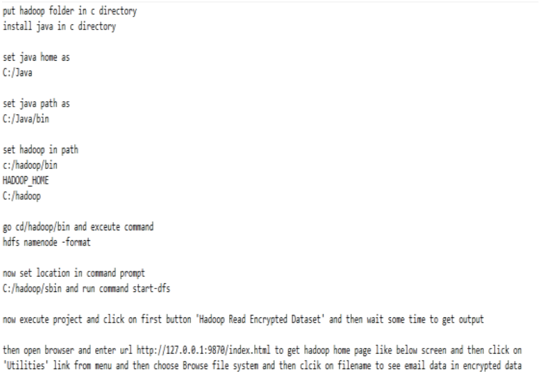


Fig. 4. **Encrypted data**

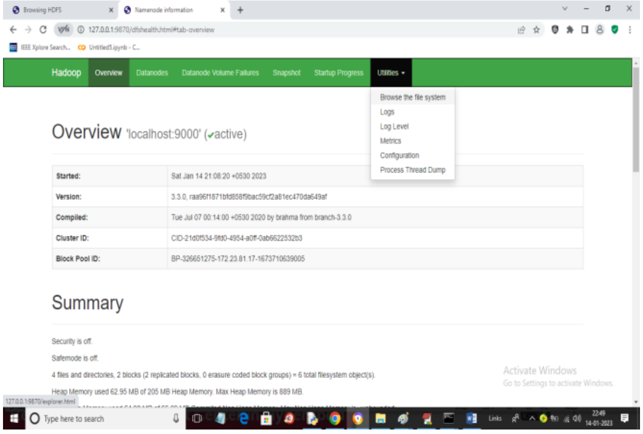
Above encrypted data will be automatically loaded to Hadoop after executing this application. To load encrypted data to Hadoop we need to follow below steps from image



```
put hadoop folder in c directory
install java in c directory

set java home as
C:/Java

set java path as
C:/Java/bin

set hadoop in path
c:/hadoop/bin
HADOOP_HOME
C:/hadoop

go cd/hadoop/bin and exceute command
hdfs namenode -format

now set location in command prompt
C:/hadoop/sbin and run command start-dfs

now execute project and click on first button 'Hadoop Read Encrypted Dataset' and then wait some time to get output

then open browser and enter url http://127.0.0.1:9870/index.html to get hadoop home page like below screen and then click on
'Utilities' link from menu and then choose Browse file system and then clcik on filename to see email data in encrypted data
```

Fig. 5. **Steps From Image**

Now in below Hadoop screen we can see encrypted data by following above steps



Overview 'localhost:9000' (✓active)

Started:	Sat Jan 14 21:08:20 +0530 2023
Version:	3.3.0, raa96f1871bfd858f9bac59cf2a81ec470da649af
Compiled:	Tue Jul 07 00:14:00 +0530 2020 by brahma from branch-3.3.0
Cluster ID:	CID-21d0f534-9fd0-4954-a0ff-0ab6622532b3
Block Pool ID:	BP-326651275-172.23.81.17-1673710639005

Summary

Security is off.

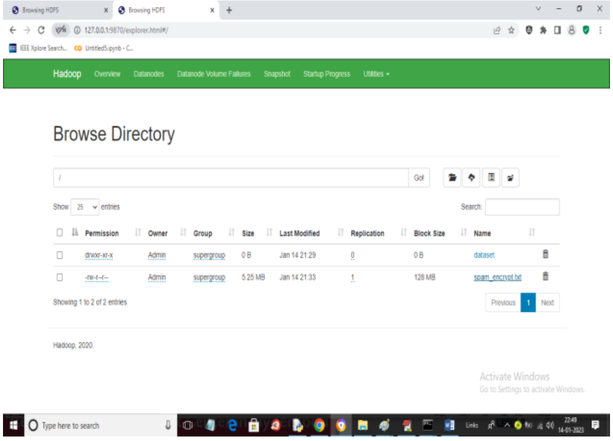
Safemode is off.

4 files and directories, 2 blocks (2 replicated blocks, 0 erasure coded block groups) = 6 total filesystem object(s).

Heap Memory used 62.95 MB of 205 MB Heap Memory. Max Heap Memory is 889 MB.

Fig. 6. **Hadoop Screen**

In above screen from Utilities link I am selecting 'Browse the file system' link to get below page



Browse Directory

/

Show 25 entries Search:

□	Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name
□	drwxr-xr-x	Admin	supergroup	0 B	Jan 14 21:29	0	0 B	dataset
□	-rw-r--r--	Admin	supergroup	5.25 MB	Jan 14 21:33	1	128 MB	spam_encrypt.txt

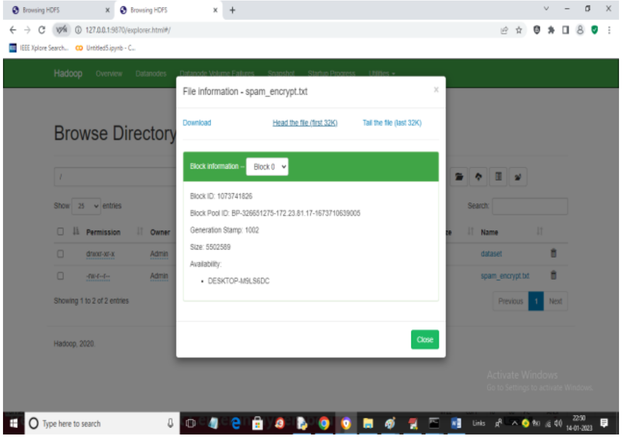
Showing 1 to 2 of 2 entries

Previous 1 Next

Hadoop, 2020.

Fig. 7. **Hadoop Screen Browse Directory**

In above screen click on 'spam_encrypt.txt' file to get below page



File information - spam_encrypt.txt

Block information -- Block 0

Block ID: 1073741826

Block Pool ID: BP-326651275-172.23.81.17-1673710639005

Generation Stamp: 1002

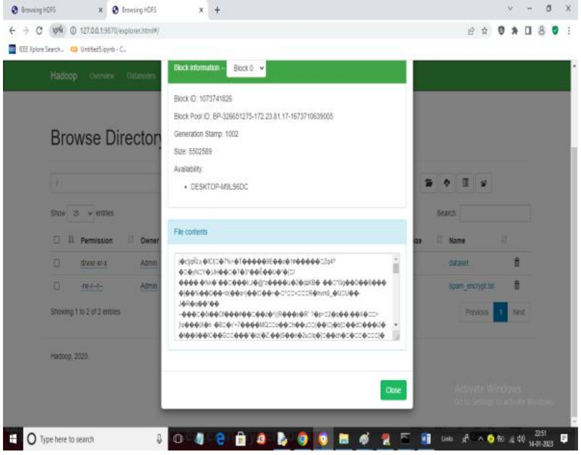
Size: 5502589

Availability:

- DESKTOP-M9LS6DC

Fig. 8. **Hadoop Screen File Information**

In above screen click on 'Download' button to download encrypted email data or click on 'Head The file' link to view encrypted data like below screen



Block information -- Block 0

Block ID: 1073741826

Block Pool ID: BP-326651275-172.23.81.17-1673710639005

Generation Stamp: 1002

Size: 5502589

Availability:

- DESKTOP-M9LS6DC

Fig. 9. **Hadoop Screen Encrypted Data**

In above screen in File Content Box we can see messages in encrypted data.

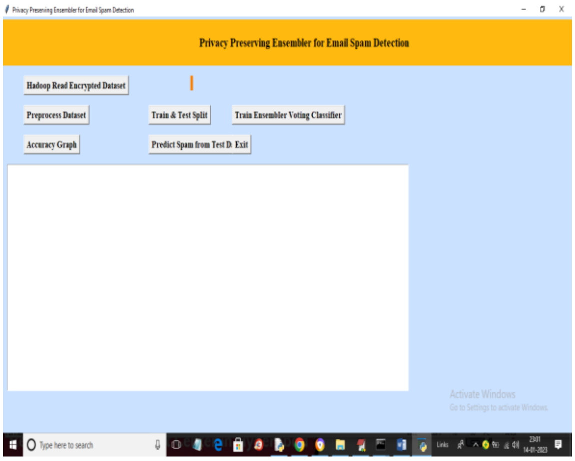
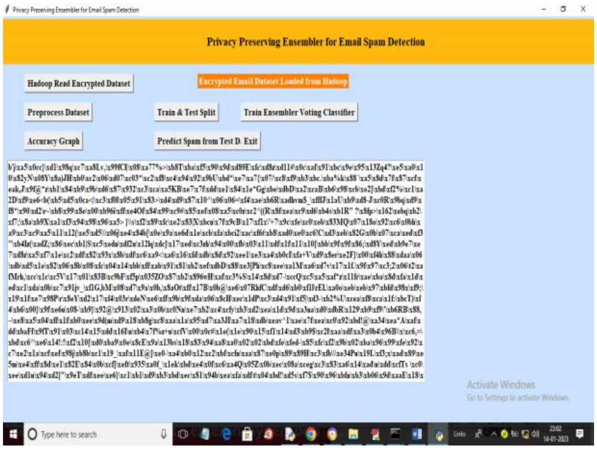

Privacy Preserving Ensembler for Email Spam Detection
Hadoop Read Encrypted Dataset
Preprocess Dataset
Train & Test Split
Train Ensembler Voting Classifier
Accuracy Graph
Predict Spam from Test D
Exit

Fig. 10. **Read encrypted emails from Hadoop**

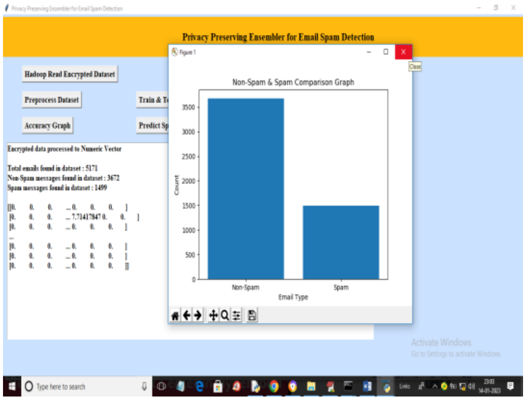
In above screen click on ‘Hadoop Read Encrypted Dataset’ button to read encrypted Emails from Hadoop and get below output


Privacy Preserving Ensembler for Email Spam Detection
Hadoop Read Encrypted Dataset
Encrypted Email Dataset Loaded from Hadoop
Preprocess Dataset
Train & Test Split
Train Ensembler Voting Classifier
Accuracy Graph
Predict Spam from Test D
Exit

```
[illegible]
```

Loaded Encrypted email data

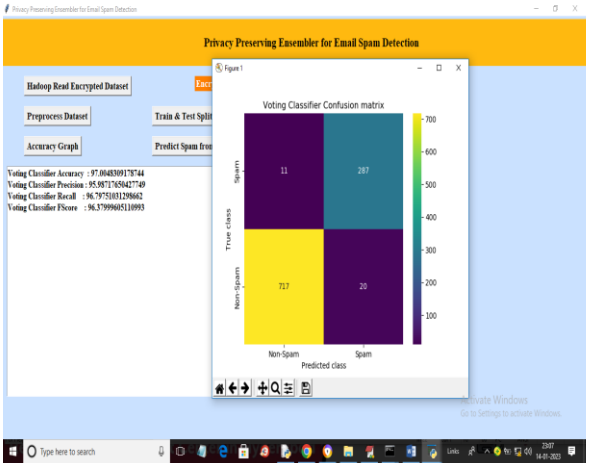
In above screen encrypted Emails loaded


Privacy Preserving Ensembler for Email Spam Detection
Hadoop Read Encrypted Dataset
Preprocess Dataset
Accuracy Graph
Figure 1
Non-Spam & Spam Comparison Graph
Count
3500
3000
2500
2000
1500
1000
500
0
Non-Spam
Spam
Email Type

```
Encrypted data processed to Numeric Vector

Total emails found in dataset : 5171
Non-Spam messages found in dataset : 3672
Spam messages found in dataset : 1499

[[0.        0.        0.        ... 0.        0.        0.        ]
 [0.        0.        0.        ... 7.71417847 0.        0.        ]
 [0.        0.        0.        ... 0.        0.        0.        ]
 ...
 [0.        0.        0.        ... 0.        0.        0.        ]
 [0.        0.        0.        ... 0.        0.        0.        ]
 [0.        0.        0.        ... 0.        0.        0.        ]]
```

Fig. 11. **Display dataset**

In above screen we can see dataset contains 5171 emails where 3672 are Non-Spam messages and 1499 are the Spam messages .



```
Voting Classifier Accuracy  : 97.0048309178744
Voting Classifier Precision : 95.98717650427749
Voting Classifier Recall    : 96.79751031298662
Voting Classifier FScore    : 96.37999605110993
```

Fig. 12. **Voting classifier confussion matrix**

In above screen with Voting Classifier we got 97% accuracy and we can see other metrics like precision, recall and etc.

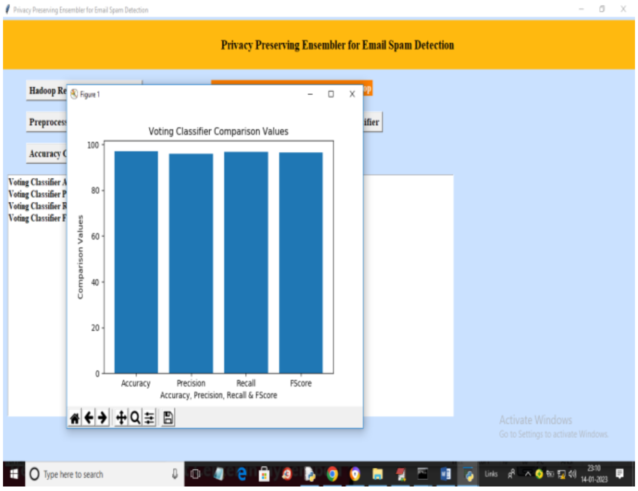

Privacy Preserving Ensembler for Email Spam Detection
Figure 1
Voting Classifier Comparison Values
Comparison Values
100
80
60
40
20
0
Accuracy
Precision
Recall
FScore
Accuracy, Precision, Recall & FScore

Fig. 13. **Voting Classifier Comparision value**

In above graph x-axis represents metric like accuracy, precision, recall and FSCORE and y-axis represents metric values which are closer to 100%.

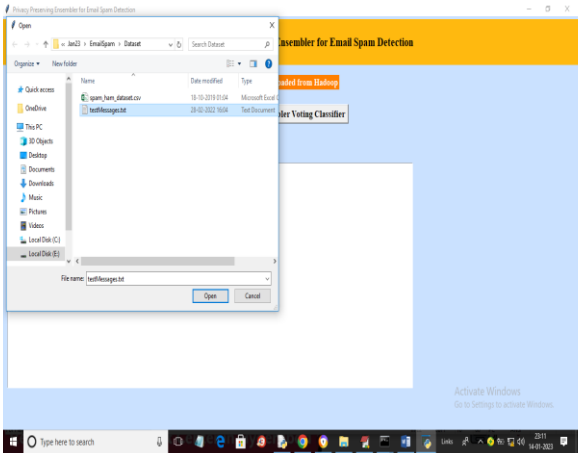
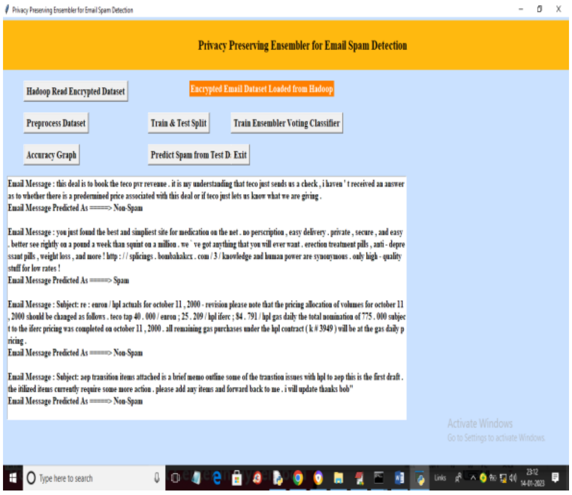


Fig. 14. **Uploading test messages.txt file**

In above scree selecting and uploading 'testMessages.txt' file and then click on 'Open' button to get below output


Privacy Preserving Ensembler for Email Spam Detection
Privacy Preserving Ensembler for Email Spam Detection
Hadoop Read Encrypted Dataset
Encrypted Email Dataset Loaded from Hadoop
Preprocess Dataset
Train & Test Split
Train Ensembler Voting Classifier
Accuracy Graph
Predict Spam from Test D
Exit
Type here to search
Activate Windows
Go to Settings to activate Windows.

```
Email Message : this deal is to book the teco pvr revenue . it is my understanding that teco just sends us a check , i haven ' t received an answer
as to whether there is a predermined price associated with this deal or if teco just lets us know what we are giving .
Email Message Predicted As =====> Non-Spam

Email Message : you just found the best and simpliest site for medication on the net . no perscription , easy delivery . private , secure , and easy
. better see rightly on a pound a week than squint on a million . we ` ve got anything that you will ever want . erection treatment pills , anti - depre
ssant pills , weight loss , and more ! http : / / splicings . bombahakcx . com / 3 / knowledge and human power are synonymous . only high - quality
stuff for low rates !
Email Message Predicted As =====> Spam

Email Message : Subject: re : enron / hpl actuals for october 11 , 2000 - revision please note that the pricing allocation of volumes for october 11
, 2000 should be changed as follows . teco tap 40 . 000 / enron ; 25 . 209 / hpl iferc ; 84 . 791 / hpl gas daily the total nomination of 775 . 000 subjec
t to the iferc pricing was completed on october 11 , 2000 . all remaining gas purchases under the hpl contract ( k # 3949 ) will be at the gas daily p
ricing .
Email Message Predicted As =====> Non-Spam

Email Message : Subject: aep transition items attached is a brief memo outline some of the transtion issues with hpl to aep this is the first draft .
the itilized items currently require some more action . please add any items and forward back to me . i will update thanks bob"
Email Message Predicted As =====> Non-Spam
```

Fig. 15. **Predict spam from test data**

In above screen before ➔ arrow symbol we can see email test message and then after ➔ arrow symbol we can see predicted output as 'spam' or 'non-spa.

IV. Conclusion

We developed protocols for training and evaluating a logistic regression based spam filtering classifier over emails belonging to multiple parties while preserving the privacy constraints. We presented an information theoretic analysis of the security of the protocol and also found that both the encryption/decryption and data transmission costs of the protocol are linear in the number of training instances and the dimensionality of the data. We also experimented with a prototype implementation of the protocol on a large scale email dataset and demonstrate that our protocol is able to achieve close to state of the art performance in a feasible amount of execution time. The future directions of this work include applying our methods to other spam filtering classification algorithms. We also plan to extend our protocols to make extensive use of parallel architectures such as GPUs to further increase the speed and scalability.

Future Scope

In future the proposed method can be extended with Investigate the integration of homomorphic encryption to enable computation on encrypted data without decrypting it, further enhancing privacy protection.

References

1. Data Pfleeger, S.L., Bloom, G.: Canning spam: proposed solutions to unwanted email. IEEE Secur. Priv. 3(2), 40–47 (2005)
2. Grier, C., Thomas, K., Paxson, V., & Zhang, M. (2010, October). @ spam: the underground on 140 characters or less. in Proceedings of the 17th ACM conference on Computer and communications security (pp. 27–37)
3. Agarwal, D.K., Kumar, R.: Spam filtering using SVM with different kernel functions. Int. J. Comput. Appl. **136**(5), 16–23 (2016)
4. Heartfield, R., Loukas, G.: A taxonomy of attacks and a survey of defence mechanisms for semantic social engineering attacks. ACM Comput. Surv. (CSUR) **48**(3), 1–39 (2015)
5. John, J. P., Moshchuk, A., Gribble, S. D., & Krishnamurthy, A.: Studying spamming botnets using botlab. in NSDI (Vol. 9, No. 2009) (2009, April)
6. Kumar, N., & Sonowal, S.: Email spam detection using machine learning algorithms. in 2020 Second International Conference on Inventive Research in Computing Applications (ICIRCA) (pp. 108–113). IEEE. (2020)
7. Junnarkar, A., Adhikari, S., Fagania, J., Chimurkar, P., & Karia, D.: E-mail spam classification via machine learning and natural language processing. in 2021 Third International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV) (pp. 693–699). IEEE. (2021, February)
8. Awad, W.A., ELseuofi, S.M.: Machine learning methods for spam e-mail classification. Int. J. Comput. Sci. Inf. Technol. (IJCSIT) **3**(1), 173–184 (2011)
9. Zhang, F., Chan, P.P., Biggio, B., Yeung, D.S., Roli, F.: Adversarial feature selection against evasion attacks. IEEE Trans. Cybern. **46**(3), 766–777 (2015)
10. Shaukat, K., Luo, S., Chen, S., & Liu, D.: Cyber threat detection using machine learning techniques: A performance evaluation perspective. in 2020

international conference on cyber warfare and security (ICCWS) (pp. 1–6). IEEE. (2020, October)

11. Garavand, A., Salehnasab, C., Behmanesh, A., Aslani, N., Zadeh, A.H., Ghaderzadeh, M.: Efficient model for coronary artery disease diagnosis: a comparative study of several machine learning algorithms. J. Healthc. Eng. (2022).
12. Ghaderzadeh, M., Aria, M., Asadi, F.: X-ray equipped with artificial intelligence: changing the COVID-19 diagnostic paradigm during the pandemic. BioMed Res. Int. (2021).
13. Hajek, P., Barushka, A., Munk, M.: Fake consumer review detection using deep neural networks integrating word embeddings and emotion mining. Neural Comput. Appl. **32**, 17259–17274 (2020)
14. Ramanathan, V., Wechsler, H.: Phishing detection and impersonated entity discovery using conditional random field and latent Dirichlet allocation. Comput. Secur. **34**, 123–139 (2013)
15. Ghourabi, A., Mahmood, M.A., Alzubi, Q.M.: A hybrid CNN-LSTM model for SMS spam detection in arabic and english messages. Future Internet **12**(9), 156 (2020)
16. Madhavan, M. V., Pande, S., Umekar, P., Mahore, T., & Kalyankar, D.: Comparative analysis of detection of email spam with the aid of machine learning approaches. in IOP conference series: materials science and engineering (Vol. 1022, No. 1, p. 012113). IOP Publishing. (2021)
17. Rayan, A.: Analysis of e-mail spam detection using a novel machine learning-based hybrid bagging technique. Comput. Intell. Neurosci. (2022).
18. Suborna, A.K., Saha, S., Roy, C., Sarkar, S., & Siddique, M.T.H.: An approach to improve the accuracy of detecting spam in online reviews. in 2021 International Conference on Information and Communication Technology for Sustainable Development (ICICT4SD) (pp. 296–299). IEEE. (2021, February)
19. Frías-Blanco, I., Verdecia-Cabrera, A., Ortiz-Díaz, A., & Carvalho, A.: Fast adaptive stacking of ensembles. in Proceedings of the 31st Annual ACM Symposium on Applied Computing (pp. 929–934). (2016, April)
20. El-Kareem, A., Elshenawy, A., Elrfaey, F.: Mail spam detection using stacking classification. J. Al-Azhar Univ. Eng. Sector **12**(45), 1242–1255 (2017)

.