Using Machine Learning to Prioritize Forensic Artifacts in Email Evidence

Mohammed AbuJarour
XU University of Applied Sciences
Mediadesign University of Applied Sciences
Berlin, Germany
m.abujarour@mediadesign.de

Raad Bin Tareaf
XU University of Applied Sciences
Potsdam, Germany
r.bintareaf@xu-university.de

Viktoria Hafner
XU University of Applied Sciences
Potsdam, Germany
v.hafner@student.xu-university.de

***Abstract*—This paper addresses one of the challenges of digital forensics by enhancing the analysis phase, focusing on prioritizing forensic artifacts in email evidence using machine learning. The urgency stems from the growing volume and variety of digital evidence, leading to backlogs in forensic laboratories. The research explores how categorizing forensic artifacts by priority can optimize the criminal investigation. The study develops and evaluates three machine learning models: Decision Tree, Support Vector Machine (SVM), and Fully Connected Neural Network (FCNN). These models are assessed based on their accuracy, strengths, limitations, and ability to prioritize email evidence. Results indicate that the SVM model is the most accurate and consistently performing well, while the FCNN model uniquely classifies all validation set emails correctly.**

***Index Terms*—Artificial Intelligence, Classification, Decision Trees, SVM, FCNN**

I. Introduction

Crimes that affect thousands of victims, families, and businesses each year like homicide, kidnapping, and burglary are increasingly characterized by digital evidence [1]. This shift is underscored by the ubiquity of electronic devices, which reinforces the importance of digital devices for detecting evidence related to cyber and non-cyber criminal activities [2]. According to the [3], over 90% of all crimes have a proven digital connection. However, the digital forensic investigation process faces many challenges. The increasing amount of digital evidence, its diverse sources, and the evidence-centric nature of industry tools contributes to delays and backlogs in forensics labs [4]. Technological advancements in storage have resulted in a significant rise in the amount of data, thereby exacerbating the case backlog. This has had an adverse effect on criminal investigations and court proceedings, impeding their progress [5].

Despite the variety of tools, digital forensics remains largely a manual process that requires careful analysis by experts in the field [6]. The urgency of extracting information quickly from digital devices has reached a critical point due to the increasing number of investigations involving different computer systems and vast amounts of data [7]. Investigators struggle with cognitive challenges and time-consuming processes and are often overwhelmed by the size and volume of cases [8], [9]. The increase in crime and the associated complexities increase the pressure on investigators to deliver timely results [10].

The overall objective of this work is to develop an innovative approach that enables the prioritization of forensic artifacts to enhance the efficiency and efficacy of the crime investigation process in the domain of digital forensics. This approach allows for a swifter and more targeted examination of pertinent evidence with the potential to optimize the analysis phase of an investigative procedure.

II. Theoretical Foundation

A. Digital Evidence and its Significance

Digital evidence is any digital data that reliably supports or disproves a hypothesis related to an incident [11]. Homem expands this definition to include data generated, transmitted, or stored on digital devices, which can reconstruct the course of a suspected criminal event. Digital evidence encompasses “digital data” used during transmission, storage, or processing to verify or refute theories associated with malicious events [4].

This type of evidence applies to both digital and traditional crimes, such as robberies, burglaries, and homicides, where digital footprints like chat messages, emails, and phone records can provide critical insights into alibis, timelines, and incident details [4], [12].

Currently, around 85% of criminal investigations involve electronic evidence, as estimated by the European Union. Retrieving digital evidence from sources like email, cloud services, online payments, and portable devices has become fundamental in criminal proceedings [13]. Mark Stokes, Head of Digital Forensics at the Metropolitan Police Service, estimates that around 90% of criminal activities involve a digital component, significantly impacting legal outcomes [14].

B. Automation in Digital Forensics Investigations

The integration of automation techniques in digital forensics has become essential for data collection, analysis, and interpretation. Automation efficiently handles repetitive tasks, reducing manual effort in digital forensic investigations [15]. As digital crime increases, automation is crucial for managing the workload [13]. Al Fhadi et al. (2013, p. 5) reported that 58% of law enforcement and organizations recognize automation’s potential to reduce manual efforts, viewing it as critical for the field’s future.

979-8-3315-0694-0/25/$31.00 ©2025 IEEE

Automation in digital forensics aims to accelerate evidence processing and address lab backlogs, easing forensic practitioners' workload [4]. Proper application can reduce data analysis burdens, allowing analysts to focus on other tasks and improving investigation accuracy [16]. The significant rise in machine learning (ML) research publications from 2010 to 2021 highlights the growing interest in applying ML to digital forensics [17].

With increasing data volumes, optimizing analysis processes to swiftly identify relevant evidence is crucial. Studies suggest using ML to prioritize file artifacts based on metadata and timeline events. Du et al. proposed training models with previously analyzed files to generate relevance values for new artifacts, enabling immediate analysis of seized devices. However, this method's effectiveness depends on the size of the pre-analyzed dataset and does not involve content analysis [18]. Du and Scanlon's supervised ML approach automates metadata-based classification, addressing the challenge of manually finding suspicious files [19]. Al Fahdi et al. demonstrated that unsupervised pattern recognition with Self-Organizing Maps could cluster significant artifacts, although adaptation based on investigator decisions was limited [8].

These approaches aim to expedite investigations but primarily focus on file metadata, potentially overlooking crucial file content information. Clemens cautions that metadata alone can be misleading or incomplete [6]. This paper explores using ML to prioritize email analysis based on content and emotional tone, addressing the identified research gap.

III. Research Method

This chapter presents a comprehensive overview of the research methodology employed in this study to automate the process of prioritizing email evidence. It encompasses the procedure of creating an appropriate dataset and provides information about the ML model, as well as the evaluation metrics employed to compare the models.

A. Overview of the Approach

The core methodology includes categorizing email evidence into content-based priority levels using ML models. This approach requires not only the development and training of the ML models as well as the extraction of comparative information but also the development of a label for the dataset created by expert knowledge. For this purpose, the Enron email dataset was chosen as the basis. This dataset is particularly suited because of the real events surrounding the Enron scandal and thus reflects the real application of ML models in the field of digital forensics. The ML model types of DT, SVM, and FCNN were selected for this approach due to the classification requirement of the problem. In order to conduct a comprehensive evaluation of models, it is advised by [20] to utilize a standardized collection of performance metrics throughout the evaluation procedure. Hence, the evaluation phase incorporates the employment of performance measures such as Accuracy, Precision, Recall, as well as F1-Score. Furthermore, the Confusion Matrix is generated for each model to assess the quality of predictability and subsequently compared to determine a potentially superior model. After the individual models have been developed, a comparative analysis of the model types is carried out in order to identify their strengths and limitations as well as to provide recommendations regarding the suitability of the different model types during an investigation.

B. Dataset Selection

Hilmand et al. emphasize the importance of text recordings and emails as critical evidence sources in digital forensic investigations [21]. To reflect real-world investigations, the Enron email corpus was selected for this study. The Enron corpus comprises approximately 500,000 emails collected over 3.5 years from the Enron energy company, which filed for bankruptcy in December 2001 following a major corporate scandal [22], [23].

This study trained supervised models using a modified Enron email corpus, incorporating a subset of around 1,700 emails annotated by UC Berkeley students, focusing on business emails and those related to the California energy crisis. The experiment utilized 1,641 labeled emails, categorized into four primary categories and 53 subcategories. Emails were assigned categories based on their content and given weight values indicating the frequency of category assignment.

1) Data Cleaning and Preprocessing: To ensure the quality and reliability of the dataset, meticulous data cleaning was performed. After finishing the data cleaning, the One Hot Encoding method was employed to create distinct columns for each primary category, as well as for the subcategories and the respective weightings. This step was necessary as the underlying dataset suffers an absence of an inherent order of the category designations.

After performing One Hot Encoding, category values were adjusted such that a value of one indicates a positive allocation to a content category, and zero indicates unassigned categories, aiding in more accurate model training despite increased data dimensionality.

2) Priority Score Value and Priority Recommendation: The existing dataset lacks priority labels, so a "Priority_Recommendation" label was created based on category assignments. Emails were assigned low, medium, and high priorities within the "Coarse Genre" and "Included/Forwarded Information" categories, excluding non-representative categories. Expert input from a digital forensics specialist identified and prioritized subcategories, assigning eleven to low, five to medium, and five to high priority. A Priority_Score_Value, ranging from 0.0 to 1.0, was calculated using weights (Low=0, Medium=0.5, High=1) based on subcategory assignments and frequency. This score was then used to derive the Priority_Recommendation (Low, Medium, High) for each email, resulting in a labeled dataset for model training and providing clear insights into email significance in investigations.

3) Feature Selection: Features have a decisive impact on the execution time and predictive performance of an ML model. For example, training and testing with a large number of irrelevant features that show no relevance to the prediction can negatively affect the performance of the model [24]. With small datasets, as is the case in the approach of this paper, a few features may be sufficient to achieve the desired results [18]. Thus, the features for the development of the ML models are limited to the four main category and the 53 sub-categories as well as the weightings. This results in a total of 159 features per email.

The performance of the classification machine learning models is evaluated using four metrics: Accuracy, Precision, Recall, and F1-Score.

IV. Developing Machine Learning Models for Evidence Priority Categorization

This section covers the development process of DT, SVM, and FCNN models. Moreover, it delves into the process of hyperparameter tuning and the selection of the most suitable model within each category by evaluating diverse performance metrics.

A. Decition Tree

Decision Tree (DT) models are popular for solving classification problems due to their interpretability and insight into feature influence on priority labels [25]. The used DT model, implemented in Python using Scikit-learn, requires categorical labels to be encoded numerically using LabelEncoder. This conversion improves model interpretability and prediction accuracy by integrating qualitative information into the quantitative framework.

Three DT models with varying depths were trained and evaluated:

- Model 1 (max_depth=12): Achieved 96.6% accuracy, indicating strong pattern learning but a slight risk of overfitting.
- Model 2 (max_depth=8): Achieved 94.8% accuracy, balancing performance and generalization.
- Model 3 (max_depth=5): Achieved 93.9% accuracy, with the least risk of overfitting but lower performance.

Learning curves indicated that all models converged with increasing training examples. Confusion matrices revealed minimal misclassifications, with Model 1 having the fewest errors but a slight risk of overfitting. Model 2 balanced performance and stability, while Model 3 had the most errors but the smallest risk of overfitting.

All three DT models demonstrated strong predictive capabilities (as shown in Table I). Model 1 had the highest accuracy but required attention to overfitting. Model 3 showed the least overfitting but had the lowest accuracy. Model 2 offered a balanced performance, making it a reliable option for generalization with good precision and recall. Therefore, Model 2 was selected for further comparison with other ML models in this study.

B. Support Vector Machine

Support Vector Machines (SVMs) create decision boundaries to classify data. This is crucial for handling noisy, nonlinearly separable data [26], making SVMs advantageous for prioritizing complex email evidence.

The SVM model was implemented in Python using Scikit-learn. Unlike DT models, SVMs do not require converting categorical labels to numeric. Both linear and radial basis function (RBF) kernels were tested to evaluate SVM performance across different data relationships. The regularization parameter (C) was adjusted to balance margin maximization and training error minimization, aiding in generalization, and reducing overfitting [27].

Three SVM models were trained and evaluated:

- Model 1 (linear kernel, C=1): Achieved 99.7% accuracy, effectively capturing dataset patterns with a clear decision boundary.
- Model 2 (RBF kernel, C=1): Achieved 97.2% accuracy, handling more complex relationships but with lower accuracy.
- Model 3 (RBF kernel, C=5): Achieved 98.2% accuracy, improving over Model 2 but slightly overfitting.

Learning curves indicated that Model 1 had excellent generalization with a small gap between training and cross-validation accuracy. Models 2 and 3 showed larger gaps, indicating potential overfitting. Confusion matrices revealed that Model 1 had the fewest misclassifications, followed by Models 3 and 2, respectively.

All three SVM models demonstrated strong classification capabilities (as shown in Table II). Model 1 performed exceptionally well, with high accuracy, precision, recall, and F1-scores, but risked overfitting. Models 2 and 3 had slightly lower performance but better generalization. Validation tests confirmed that all models could correctly classify high-priority emails, essential for digital forensic investigations. Model 3 was selected for further comparison with other ML models due to its balanced performance and improved generalization.

C. Fully Connected Neural Network

Neural networks (NNs) training involves the backpropagation algorithm, which optimizes the model by adjusting weights based on prediction errors [28]. The Fully Connected Neural Network (FCNN) was implemented in Python using TensorFlow and Keras. The categorical label “Priority Recommendation” was converted to numerical format using Scikit-learn’s LabelEncoder. The FCNN architecture includes an input layer, two hidden layers (64 and 32 units), and an output layer (three units for “High”, “Middle”, and “Low” priorities). Hidden layers use Rectified Linear Unit (ReLU) activation functions to capture nonlinear relationships, and the output layer uses a softmax activation function. Two optimization approaches, Adaptive Moment Estimation (Adam) and Stochastic Gradient Descent (SGD), were tested. Models were compiled with these optimizers and the sparse categorical cross-entropy loss function. Early stopping callback was employed to prevent overfitting by monitoring validation loss and halting training when improvements ceased, ensuring generalization and efficiency.

Two FCNN models were developed:

TABLE I
OVERVIEW OF DT MODELS' PERFORMANCE METRICS

	Precision			**Recall**			**F1-Score**		
	High	Middle	Low	High	Middle	Low	High	Middle	Low
First DT Model (max_depth=12)	0.95	0.96	0.98	0.98	0.93	0.98	0.97	0.95	0.98
Second DT Model (max_depth=8)	0.94	0.92	0.98	0.96	0.92	0.96	0.95	0.92	0.97
Third DT Model (max_depth=5)	0.93	0.93	0.95	0.96	0.88	0.97	0.95	0.90	0.96

TABLE II
OVERVIEW OF SVM MODELS' PERFORMANCE METRICS

	Precision			**Recall**			**F1-Score**		
	High	Middle	Low	High	Middle	Low	High	Middle	Low
First SVM Model (kernel=linear; C=1)	0.99	1.00	1.00	1.00	0.99	1.00	1.00	1.00	1.00
Second SVM Model (kernel=rbf; C=1)	0.99	0.95	0.97	0.97	0.96	0.98	0.98	0.96	0.98
Third SVM Model (kernel=rbf; C=5)	0.99	0.97	0.98	0.97	0.97	1.00	0.98	0.97	0.99

- Model 1: Utilized the Adam optimizer with 100 epochs and early stopping, terminating at 23 epochs, achieving 97.9% accuracy.
- Model 2: Utilized the SGD optimizer with 200 epochs and early stopping, terminating at 52 epochs, achieving 99.1% accuracy.

Learning curves for both models demonstrated efficient training and rapid learning. Model 1 showed a quick decline in loss and convergence of accuracy curves at around 96% after five epochs. Model 2 achieved better convergence and higher accuracy.

Confusion matrices revealed seven misclassifications in Model 1 and three in Model 2. Both models correctly predicted all high-priority emails, essential for time-critical investigations. Performance metrics indicated that Model 2 consistently outperformed Model 1, though both demonstrated robust classification capabilities.

Both FCNN models effectively prioritized email evidence (as shown in Table III). Despite Model 2's higher metrics, Model 1 performed better on the validation set, suggesting better generalization. Therefore, Model 1 with the Adam optimizer is selected for further comparison with other machine learning models.

V. COMPARATIVE ANALYSIS

In the final comparative analysis, we selected the most promising model from each model type. The DT model type is represented by the second model with a maximum tree depth of eight, which demonstrated a solid accuracy of 94.8% and was capable of accurately classifying priorities and generalizing to unknown instances. As for the SVM model type, the third model with the RBF kernel and the regularization parameter of five was selected to be compared with the other model types. With an accuracy of 98.2%, it outperformed the DT model by 3.4 basis points. Among the FCNN models, the first model with the Adam optimizer was chosen. With an accuracy of 97.9%, this model performs in the higher midrange compared to the other two types. As seen in Table IV, the DT model type performs less effectively than the other ML types in terms of performance metrics. Nevertheless, all values exceeding 90% indicate that the model performs satisfactorily. The SVM model type performs best in terms of the performance metrics with five of the highest values and three averages. However, the values of the FCNN model type are also in the high range, thus underlining the ability to accurately prioritize emails. Among the selected models, the FCNN type was able to correctly classify all emails in the validation set while the other two types misclassified one medium-prioritized email.

Ensuring the accurate identification of high-prioritized emails is critical to prevent relevant information from being omitted in later stages of the investigation and, most importantly, during the legal process. Hence, the accurate classification of the validation set by the FCNN model type is a significant accomplishment.

The SVM type exhibits a slight advantage in terms of overall accuracy, outperforming both the DT and FCNN types. Accuracy is a critical metric in digital forensics, especially when dealing with large amount of email evidence. A highly accurate model ensures that emails are accurately categorized, allowing investigators to promptly prioritize important items and thus streamline the investigative process. This allows to optimize the utilization of the resources like time and manpower towards the most pertinent emails. Therefore, inaccuracies in digital forensic analysis can reduce its effectiveness, as they can result in false positives, where irrelevant items are incorrectly identified as relevant, or false negatives, where relevant items are missed. Although the FCNN model has slightly lower accuracy compared to the SVM, it displays strong learning convergence and similar classification accuracy.

It is worth emphasizing that the DT model type has a

TABLE III
Overview of FCNN Models' Performance Metrics

	Precision			Recall			F1-Score		
	High	Middle	Low	High	Middle	Low	High	Middle	Low
First FCNN Model (Adam Optimizer)	0.99	0.98	0.97	1.00	0.95	0.98	1.00	0.97	0.97
Second FCNN Model (SGD Optimizer)	1.00	0.990	0.980	1.000	0.980	0.990	1.000	0.990	0.99

TABLE IV
Performance Metrics Overview of ML Model Types

	Precision			Recall			F1-Score		
	High	Middle	Low	High	Middle	Low	High	Middle	Low
DT Model Type	0.94	0.92	0.98	0.96	0.92	0.96	0.95	0.92	0.97
SVM Model Type	0.99	0.97	0.98	0.97	0.97	1.00	0.98	0.97	0.99
FCNN Model Type	0.99	0.98	0.97	1.00	0.95	0.98	1.00	0.97	0.97

significant advantage compared to other types. The utilization of a graphical representation of the DT model type provides a means to comprehend the decision-making process. Consequently, the predicted priority recommendations can be derived based on the tree visualization, thereby enabling traceability.

Compared to DT models, FCNN and SVM models suffer from interpretability issues due to their black-box nature, complicating their use in court proceedings [29]. DT models, while more interpretable, face accuracy challenges as complexity increases, influenced by stop criteria and pruning methods [25]. SVM models are scalable and suitable for high-dimensional data but require careful kernel function selection [30]. FCNN models, adept at capturing patterns in complex data, are prone to overfitting and depend on appropriate hyperparameter selection [31].

VI. Summary and Outlook

Digital forensics faces numerous challenges due to the increasing volume, diversity, and complexity of electronic evidence. Forensic professionals must rapidly identify and prioritize crucial evidence, which is time-consuming. This research introduces a methodology using machine learning (ML) to optimize the analysis of email evidence. The study comprehensively compares various ML models, evaluating their performance, strengths, and limitations, highlighting the practical challenges in model selection for forensic experts [29].

The findings demonstrate that ML models can accurately prioritize email evidence, identifying emails with pertinent information based on content. Specifically, SVM and FCNN models enhance efficiency by effectively handling large data volumes.

One of the key strengths of this study is its research approach, which surpasses the use of metadata-based models in comparison to other studies. As highlighted by [6], relying solely on metadata analysis is not recommended due to the potential for misleading or incomplete information. Therefore, this study introduces a new approach to predicting recommendations by taking into account the categories of email content.

Despite the existing strengths, the study has several limitations that are typical characteristics of the field of digital forensics. One of the main limitations in this area of research is the lack of availability of datasets that can be used for the development of forensic tools. According to [8] it is made clear that this limitation arises due to legal and data protection restrictions. [18] and [32] extend this statement by stating that this limitation additionally arises because of ethical reasons and non-sharing policies. The limited amount of data available overstretches individual datasets such as the Enron dataset, causing doubts about their ability to be applied to a wider context [32]. Additionally, the Enron dataset mainly consists of messages from employees of a single company, which amplifies concerns about its generalizability [33].

To further advance the research approach of this paper, the following future research recommendations are given: By utilizing data augmentation techniques, the dataset could be expanded and used to improve the ML capabilities. Another strategy for optimization involves broadening the scope of the email content categories to expand the performance of the ML models. Furthermore, by incorporating emails from different companies instead of solely relying on the Enron dataset, it becomes feasible to introduce greater diversity in terms of content and context. Ultimately, this can enhance the adaptability and flexibility of the ML models.

References

[1] IACP, “Digital evidence task force,” https://www.theiacp.org/resources/document/digital-evidence-task-force, 2024.

[2] A. Jarrett and K.-K. R. Choo, “The impact of automation and artificial intelligence on digital forensics,” *WIREs Forensic Science*, vol. 3, no. 6, p. e1418, 2021. [Online]. Available: https://wires.onlinelibrary.wiley.com/doi/abs/10.1002/wfs2.1418

[3] National Police Chiefs’ Council (NPCC), “Digital forensic science strategy,” https://www.npcc.police.uk/SysSiteAssets/media/downloads/publications/publications-log/2020/national-digital-forensic-science-strategy.pdf, 2020.

[4] I. Homem, “Advancing automation in digital forensic investigations,” Ph.D. dissertation, Stockholm University, Department of Computer and Systems Sciences, 2018.

[5] R. Montasari and R. Hill, “Next-generation digital forensics: Challenges and future paradigms,” in *2019 IEEE 12th International Conference on Global Security, Safety and Sustainability (ICGS3)*, 2019, pp. 205–212.

[6] J. Clemens, “Automatic classification of object code using machine learning,” *Digital Investigation*, vol. 14, pp. S156–S162, 2015, the Proceedings of the Fifteenth Annual DFRWS Conference. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1742287615000523

[7] E. Casey, “Triage in digital forensics,” *Digital Investigation*, vol. 10, no. 2, pp. 85–86, 2013, triage in Digital Forensics. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1742287613000765

[8] M. Al Fahdi, N. Clarke, F. Li, and S. Furnell, “A suspect-oriented intelligent and automated computer forensic analysis,” *Digital Investigation*, vol. 18, pp. 65–76, 2016. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1742287616300792

[9] J. I. James and P. Gladyshev, “Challenges with automation in digital forensic investigations,” 2013.

[10] A. Irons and H. S. Lallie, “Digital forensics to intelligent forensics,” *Future Internet*, vol. 6, no. 3, pp. 584–596, 2014. [Online]. Available: https://www.mdpi.com/1999-5903/6/3/584

[11] B. Carrier and E. Spafford, “Defining Event Reconstruction of Digital Crime Scenes,” *Journal of Forensic Sciences*, vol. 49, no. 6, pp. JFS2004127–8, 11 2004. [Online]. Available: https://doi.org/10.1520/JFS2004127

[12] J. Sammons, “The basics of digital forensics: The primer for getting started in digital forensics,” *The Basics of Digital Forensics: The Primer for Getting Started in Digital Forensics*, pp. 1–177, 2 2012. [Online]. Available: http://www.sciencedirect.com:5070/book/9781597496612/the-basics-of-digital-forensics

[13] F. Casino, T. K. Dasaklis, G. P. Spathoulas, M. Anagnostopoulos, A. Ghosal, I. Boröcž, A. Solanas, M. Conti, and C. Patsakis, “Research Trends, Challenges, and Emerging Topics in Digital Forensics: A Review of Reviews,” *IEEE Access*, vol. 10, pp. 25 464–25 493, 2022.

[14] Science and Technology Select Committee (UK Parliament), “Forensic science and the criminal justice system: a blueprint for change,” https://publications.parliament.uk/pa/ld201719/ldselect/ldsctech/333/33302.htm, May 2019.

[15] G. Horsman, C. Laing, and P. Vickers, “A case based reasoning system for automated forensic examinations,” 06 2011.

[16] C. Horan and H. Saiedian, “Cyber crime investigation: Landscape, challenges, and future research directions,” *Journal of Cybersecurity and Privacy*, vol. 1, no. 4, pp. 580–596, 2021. [Online]. Available: https://www.mdpi.com/2624-800X/1/4/29

[17] T. Nayerifard, H. Amintoosi, A. G. Bafghi, and A. Dehghantanha, “Machine learning in digital forensics: A systematic literature review,” 2023.

[18] X. Du, Q. Le, and M. Scanlon, “Automated artefact relevancy determination from artefact metadata and associated timeline events,” in *2020 International Conference on Cyber Security and Protection of Digital Services (Cyber Security)*. IEEE, Jun. 2020. [Online]. Available: http://dx.doi.org/10.1109/CyberSecurity49315.2020.9138874

[19] X. Du and M. Scanlon, “Methodology for the automated metadata-based classification of incriminating digital forensic artefacts,” in *Proceedings of the 14th International Conference on Availability, Reliability and Security*, ser. ARES ’19. New York, NY, USA: Association for Computing Machinery, 2019. [Online]. Available: https://doi.org/10.1145/3339252.3340517

[20] D. Shah, “Top performance metrics in machine learning: A comprehensive guide,” 2023. [Online]. Available: https://www.v7labs.com/blog/performance-metrics-in-machine-learning

[21] K. Hilmand, H. Sarmad, and M. Bakht, “A survey of machine learning applications in digital forensics,” *Trends in Computer Science and Information Technology*, pp. 020–024, Apr. 2021. [Online]. Available: http://dx.doi.org/10.17352/tcsit.000034

[22] W. W. Cohen, “Enron Email Dataset,” https://www.cs.cmu.edu/˜./enron/, 2015. [Online]. Available: https://www.cs.cmu.edu/˜./enron/

[23] J. Diesner and K. Carley, “Exploration of communication networks from the enron email corpus,” 01 2005.

[24] C. Serhal and N.-A. Le-Khac, “Machine learning based approach to analyze file meta data for smart phone file triage,” *Forensic Science International: Digital Investigation*, vol. 37, p. 301194, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2666281721001025

[25] L. Rokach and O. Maimon, *Decision Trees*. Boston, MA: Springer US, 2005, pp. 165–192. [Online]. Available: https://doi.org/10.1007/0-387-25465-X_9

[26] IBM Corporation, “About SVM - IBM Documentation,” https://www.ibm.com/docs/en/spss-modeler/18.2.2?topic=models-about-svm, 2021. [Online]. Available: https://www.ibm.com/docs/en/spss-modeler/18.2.2?topic=models-about-svm

[27] ——, “SVM Node Expert Options - IBM Documentation,” https://www.ibm.com/docs/en/spss-modeler/18.2.2?topic=node-svm-expert-options, 2021. [Online]. Available: https://www.ibm.com/docs/en/spss-modeler/18.2.2?topic=node-svm-expert-options

[28] I. H. Sarker, “Deep learning: A comprehensive overview on techniques, taxonomy, applications and research directions,” *SN Computer Science*, vol. 2, pp. 1–20, 11 2021. [Online]. Available: https://link.springer.com/article/10.1007/s42979-021-00815-1

[29] D. Dunsin, M. C. Ghanem, K. Ouazzane, and V. Vassilev, “A comprehensive analysis of the role of artificial intelligence and machine learning in modern digital forensics and incident response,” *Forensic Science International: Digital Investigation*, vol. 48, p. 301675, Mar. 2024. [Online]. Available: http://dx.doi.org/10.1016/j.fsidi.2023.301675

[30] V. R. Jakkula, “Tutorial on support vector machine (svm),” 2011. [Online]. Available: https://api.semanticscholar.org/CorpusID:15115403

[31] S. Kohli, S. Miglani, and R. Rapariya, “Basics of artificial neural network,” *International Journal of Computer Science and Mobile Computing*, vol. 3, no. 9, Sep 2014.

[32] A. A. Solanke and M. A. Biasiotti, “Digital forensics ai: Evaluating, standardizing and optimizing digital evidence mining techniques,” *KI - Künstliche Intelligenz*, vol. 36, pp. 143–161, 2022.

[33] S. Ozcan, M. Astekin, N. K. Shashidhar, and B. Zhou, “Centrality and scalability analysis on distributed graph of large-scale e-mail dataset for digital forensics,” in *2020 IEEE International Conference on Big Data (Big Data)*, 2020, pp. 2318–2327.