Journal of Theoretical and Applied Information Technology
30th November 2024. Vol.102. No. 22

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

IMPROVING SECURITY IN INTELLIGENT SYSTEMS: HOW EFFECTIVE ARE MACHINE LEARNING MODELS WITH TF-IDF VECTORIZATION FOR PASSWORD-BASED USER CLASSIFICATION

BOUMEDYEN SHANNAQ[1,*]

[1]Department of Management of information Systems, University of Buraimi, Sultanate of Oman

ABSTRACT

This research assesses the practicability of machine learning models in classifying consumers according to their passwords with the help of TF-IDF, which depicts exclusive password features. The purpose of the study is to eradicate the weakness of the current EPSB algorithm in its synthesis of electronic personal behavior. Our goal will be to define those models that have strengthened the existing methods of password-based authentication. In the second step, we transformed a data set of anonymized passwords to the transformation where each was converted into statistical feature vectors using TF-IDF and tested six models of machine learning. Specific well-known algorithms used in the course of the study were support vector machines (SVM), random forests, Naïve Bayes, K-nearest neighbor (KNN), logistic regression, and decision trees. This cross-validation made me conclude that Naive Bayes outcompeted all the other models in terms of a greater weighted average precision of 96.38%, which was higher than the other two models: the SVM model equal to 91.64% and the logistic regression model equal to 91.52%. With regards to accuracy, KNN got 79.48%, Decision Tree got 77.55%, while Random Forest recorded the lowest value of the four techniques at 71.26%. These results provide a more profound comprehension for the development of an extended password-based authentication scheme using an advanced machine learning approach.

Keyword: *Password Classification, Machine Learning, TF-IDF Vectorization, Support Vector Machine, Random Forest, Naive Bayes-Nearest Neighbors, Logistic Regression, Decision Tree*

1. INTRODUCTION

Today more than ever, it is essential to protect data and its accessibility through Information Technology. This is due to increased reliance on services internet for service delivery since more services are likely to be attacked thus requiring the protection of user's data [1][2][3]. In this context, password-based authentication is the prominent method among different ones for protecting resources, which contain confidential information [4][5]. However, although password security has become standard practice, it is often vulnerable because users choose weak passwords, create easily guessable patterns, and behave in ways that disable standard security measures [6].

The risks which originate with password-related methods have promoted an interest in improving methods of authentication [7][8][9]. Machine learning has particularly turned into a popular technology, providing new trends in recognizing and preventing security threats [10][11]. By studying the patterns found in passwords and users' activities machine learning algorithms can notify on possible security violations and enhance the stability of the identification and authorization processes[12][13][14][15].

1.1 STATEMENT OF PROBLEM

Passwords are usually vulnerable to being hacked or stolen by other people using some other better methods. In fact if an intruder gets hold of a valid password then there are chances that he or she will try to modify it hence being very dangerous to a network. More often than not, security systems cannot distinguish between a genuine user and interposer if the right password is entered. That said, even if an organization implements multi-factor authentication (MFA) this risk may still be reversed by intruders and they get to change a password. The central problem is when a user tries to change a password, the security system can

neither determine whether the user is the genuine user or an intruder.

1.2 RESEARCH QUESTIONS

- How Can Machine Learning Models with TF-IDF Vectorization Improve Password-Based User Classification for Enhanced Security in Intelligent Systems?

- Addressing Security in Intelligent Systems: How Effective is TF-IDF Vectorization in Machine Learning Models for Password-Based User Classification?

1.3 PROPOSED SOLUTION

This paper presents an approach of applying machine learning in particular concentrating on the behavior of users and the kind of preferences they demonstrate when assigning passwords. For instance, people may use factors such as birth years or preferred car brands as other parts of the passwords. The model is intended to differentiate between password change attempts originating from the actual user and the fraudulent one. The work also compares many classifiers for finding the best approach of differentiating between the legitimate and the fake users from a password profile. Term Frequency-Inverse Document Frequency (TF-IDF) vectorization technique that was previously used only in document and text analysis was also employed by this research on password classification for the first time [16][17][18].

1.4 SIGNIFICANCE OF THIS WORK

The relevance of this work is based on the idea that it can improve password-based security systems with the help of modern machine learning approaches. In this research, the outcomes of multiple model classifiers are evaluated to determine the best message-based, password classification model. The insights presented as part of the research will help in the creation of more secure and reliable authentications, as well as offer the basis for the future research of the authentication processes.

1.5 RESEARCH GAP

1.5.1 Present Emphasis on User Authentication Systems

Almost all cybersecurity studies are based on conventional practices involving password management, MFA, and encryption [19][20][21][22][23][24][25][26]. Nevertheless the use of password security is left with less concentration on identifying authorized users from the password patterns.

Unlike the case with other studies, there is a lack of a comparative analysis of the ML models for the password classification.

Even in the field of cybersecurity, machine learning has been applied in some aspects, but there are few research works are being compared to different machine learning algorithms for the classification of the user based on the passwords [27][28]. The current research gap includes the lack of investigation in the performance of these models for accuracy and robustness among the password datasets [29][30][31].

1.5.2 Limited Expansion of Text Feature Engineering for Passwords

Passwords as a form of textual data can be processed using natural language processing (NLP). To date, few studies have examined which of the text feature engineering methods such as TF-IDF can be used for the identification of passwords among users [32][33][34]. This research creates new directions for using NLP algorithms and enhancing the effectiveness of passwords.

1.6 CONTRIBUTION

1.6.1 Using TF-IDF for Password Classification: A New Concept

This research extend the work proposed in [13] and improved the performance of ML algorithm. Therefore innovates the use of TF-IDF vectorization for transforming passwords into feature vectors so that more details about passwords can be explored by detecting new patterns in password usage data.

1.6.2 Result Analysis of Two or More ML Models

To achieve this goal, a comprisals of six machine learning algorithm; Support Vector Machine (SVM), Random Forest, Naive Bayes, K -nearest Neighbor (KNN), Logistic Regression and Decision Tree for classifying the users based on their password behavior is conducted. This evaluation will offers research finds out into the highest performing classification of the passwords for function as the benchmark for the future research in cyber security.

1.6.3 Improved insight of password characteristics and usage behavior
This paper shows the patterns in the password data by discussing how various models use the features of users' input to categorize it appropriately by the choice of password. Knowledge of these patterns may provide a basis for the creation of more reliable authenticate technologies in the future, while taking into account the peculiarities of the choice of passwords by the users.

1.6.4 Improving Cybersecurity Practices
As highlighted throughout this research, there are relevant points of application in the actual practice of professional cybersecurity practice hence this information is novel and valuable for the study of password security. Thus, by revealing the best machine learning approach for password classification, the paper advances the practices of designing richer authentication systems that are capable of identifying anomalous patterns and repelling intrusions.

1.6.5 Base for Future Research
All these possibilities make this work a base for further research on the relation of machine learning and cybersecurity, with focusing on feature engineering, model fine-tuning, and real-time security applications. The evaluation of different algorithms and the comparison and categorization of PCA, LDA, MMD, and RBM generate insights that are relevant for researchers intending to employ similar approaches for other domains like fraud detection or people analytics.

1.5.6 Assistance to the Development of the Literature in ML and Cybersecurity
Through filling the existing research gaps in this study, this research brings into the research arena new information and knowledge on the use of machine learning in security. The practical contribution of this research is in the choice of TF-IDF for password analysis and the comparative evaluation of multiple machine learning models.

1.5.7 Interdisciplinary Impact
The methodology and results of this study are Relevant to a range of disciplines including natural language processing, digital forensics and behavioral.

2. LITERATURE REVIEW

The Password security is important in security implementations but is steadily coming under threats from bad implementation of password usage, using the same password on several accounts, and other experienced techniques such as phishing and dictionary attacks. Researchers are consequently seeking for ways to enhance password safety with Machine Learning (ML) solutions beginning to dominate the discourse [12] [35].

2.1. ELECTRONIC PERSONAL SYNTHESIS BEHAVIOR (EPSB) ALGORITHM

EPSB algorithm improves password typing duration and method selection for a user authentication, based on the historical behavior data using the Confidence Range function [36]. However, the number of parameters associated with the password input duration indicator, which includes only six parameters, limits the EPSB algorithm. To clarify these six parameters, in this work we developed an application to simulate the EPSB algorithm and demonstrate its functionality. The six parameters are illustrated in Table 1.

Table 1: Six parameters integrated with the EPSB algorithm.

Parameter1	Parameter2	Parameter3	Parameter4	Parameter5	Parameter6
Small letters	Capital Letters	Sum of Small letters + Capital Letters	Numerals	Symbols	Length of the password

Journal of Theoretical and Applied Information Technology
30th November 2024. Vol.102. No. 22

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

listed in Table 1.The second row displays the median-CR calculations, while the third row shows

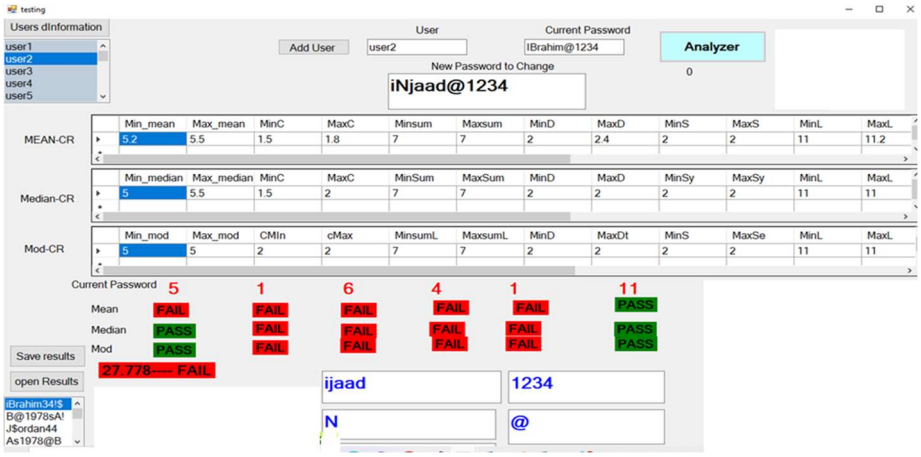

Figure 1: Screenshot of the developed application implementing the EPSB algorithm. (Developed by this work)

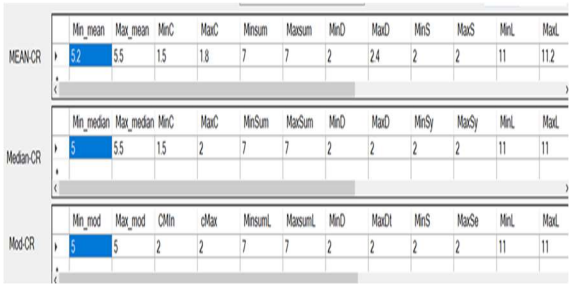

Figure 2. Screenshot of the calculation of CR from the password change history for User 2, with the latest password "Ibrahim@1234".

In Figure 1, the developed application illustrates how the most recent stored password for User 2 was calculated. User 2 updated his password five times over 15 months. The application was used to update the calculations in the Confidence Range (CR) database, which consists of the minimum and maximum values for the mean, median, and mode for all six parameters presented in Table 1.Example to explain the implementation in Figure 1: The current/last password for User 2 was "Ibrahim@1234," and the CR record. The first row (Mean-CR) in Figure 2 represents the password change history over five instances. It includes fields such as the minimum and maximum mean of small letters, the minimum and maximum capital letters, and other fields corresponding to the six parameters the mode-CR calculations. The scenario is as follows: A fake user

(F-User) who has obtained the last password of user2 (O-User) attempts to update the password. The F-User logs into the system using the correct password "Ibra-him@1234" and tries to change it to "iNjaad@1234". The application then calculates the CR for this new password, as illustrated in Figure 3. Figure 4 shows the CR calculation for iNjaad@1234.

- 5: "ijaad," the number of small letters (Parameter 1)
- 1: "N," the number of capital letters (Parameter 2)

- 6: "ijaad + N," the sum of small and capital letters (Parameter 3)
- 4: "1234," the number of numerals (Parameter 4)

Figure 4: CR calculation for "iNjaad@1234".

Figure 3: Analyzing the password of the F-User "iNjaad@1234".

After calculating the CR for the O-User (Figure 1) and the CR for the F-User (Figure 4), the application performs a comparison test as follows: For the O-User's Mean-CR in the first field shown in Figure 2, the minimum number of small letters was 5.2, and the maximum was 5.5. Thus, the CR for small letters ranges between 5.2 and 5.5.For the F-User's Mean-CR in the first field shown in Figure 4, the minimum number of small letters was 5, and the maximum was 5. Therefore, the CR for small letters ranges between 5 and 5. The system will compare if 5 falls between 5.2 and 5.5. Since 5 is not within the range of 5.2 to 5.5, the application will not add a success point for this comparison .In the example shown in Figure 1, the number of successful comparisons was 5 (highlighted in green at the bottom of Figure 1). The calculation between CR-O-User and CR-F-User is (5 / 18) * 100 = 27.7. Therefore, the security application will not allow the F-User to update the password as they are recognized as an unauthorized user. According to [36], the threshold for passing the security check is >= 66%. More detailed examples of the EPSB algorithm can be found in [36]. As a result, during the testing phase of the algorithm with the newly developed application incorporating more user data, the rate of discrimination against unauthorized users was observed to be low, thereby negatively affecting the performance of the EPSB algorithm. Consequently, this study introduces a novel approach using Machine-learning models to

- 1: "@," the number of symbols (Parameter 5)
- 11: "iNjaad@1234," the length of the password (Parameter 6)

overcome the vulnerability and low performance of EPSB algorithm

2.2. COMMON PASSWORD COMPLICATION OVER POSTING

Poor password complexity and passphrases have become a thing of worry up to this present era. Research work [37] [38] has shown that many people use easily guessable passwords hence exposing them to risks of cyber-attack. Secondly, the same password is used for other accounts and once a password has been penetrated, all the other accounts with the same password are vulnerable [39] [41].

2.3. MACHINE LEARNING FOR PASSWORD PROTECTION

Machine learning has turned out useful for developing password security over the years. Early works [42] [12] showed how using large collection of passwords as one of the input variables, the ML models could predict the strength of user chosen passwords.
Additionally, [43] used statistical models to predict likely passwords – a sign that ML can serve both as a tool to perform password cracking, as well as one to guard against the act. Some recent studies can be devoted to the classification of the exploration of ML for the purpose of recognizing unusual behavior of users. For instance, ML was utilized in cybersecurity to identify anomalous login patterns that could be a sign of a penetration [44][45]. Altogether, [46] proposed systems that integrated user activity alongside the behavior of passwords to enhance the identification process, with further works in this direction is [47].

2.4. FEATURE ELIMINATION TECHNIQUES IN PASSWORD SECURITY

The latter is in fact known to be important for efficient ML models, namely, feature extraction. Other works in [48] [49] [50][51] employed n-gram approach to capture sequential patterns of passwords and enhance password strength prediction models. Similarly, frequency analysis was also employed to examine the most frequently used passwords as a feature for development of the predictive models [52][53].

More complex and dataset dependent form of feature extraction is TF-IDF (Term Frequency-Inverse Document Frequency) vectorization that is often used in the data mining [54, 16] and that has recently been proposed to be used for extracting cybersecurity features. For instance, TF-IDF has been employed accurately in the context of [55][56], that is, in classification of phishing emails, which could be utilized in other security related uses such as password analysis.

2.5. EVALUATION OF MULTILATERAL MODELS IN THE SPHERE OF CYBERSECURITY

Several ML models have been tested in an attempt to improve the security of passwords. SVMs random forests and neural networks have been researched extensively. SVMs, namely, have given very good results in qualifying password patterns [27], Random forest, on the other hand, demonstrated high levels of stability in working with big data [57] [58].
Other models that are also ongoing is on deep learning architectures. CNN [59] and RNNs [60] approaches have been used for operation on password datasets with remarkably accuracy in terms of password strength and classification.

2.6. LESSONS AND THE FUTURE

In spite of these innovations, there are several known issues that need to be worked out. This list contains several important points, among which there is an important problem of generalizing ML models to other datasets. This has been acutely evidenced by the fact that the same models have very high performances when tested on datasets with similar users, but low performances in other real-world diverse users' dataset. This emphasizes the requirement of models that are able to distribute performance characteristics across numerous datasets [61] [62]. Also, the training process of large-sized ML models especially the deep learning models incur high computation costs limiting their use [63] [64] [65].

2.7. NOVELTY AND CONTRIBUTION OF THIS WORK

This paper provides several important contributions to the field. Firstly, we pre-process the raw password data using a feature extraction technique known as TF-IDF vectorization, which does not dominate the external literature, but which is typical for text mining. It allows extending the research on password trends based on traditional feature extraction techniques, which might have been missed by traditional feature extraction.
We also perform a detailed study of several multiple machine learning classifiers like decision tree classifier, random forest classifier and neural network classifiers to understand about the effectiveness classification model for password classification. Lastly, to address the generalization issue, we evaluated the models on multiple password datasets rather than applying the approach on a single dataset.
Tensor-flow, tf-idf vectorization and comprehensive dictionary analysis are used for the first time to predict password security, enriching both theoretical and practical studies in this area. In this paper, the proposed advancement of employing the ML approach for the password-based user classification adds crucial insight into enhancing the cybersecurity systems.

Therefore, this literature review focuses on the increasingly importance of ML and feature extraction approach, including TF-IDF, in enhancing the password security. From this work, we bring a fresh view for future password research and perform a competitive benchmarking of learned ML classifiers, which contribute back to the improvement of authentication protocols.

2.8 RESEARCH PROBLEM

Even though there have been improvements in modern security, password protection seems to be one of the most problematic areas for organizations' security policies, because passwords are reused, predictable, or susceptible to some advanced types of attacks like phishing and dictionary attacks. The use of such traditional techniques as the Electronic Personal Synthesis Behavior (EPSB) algorithm is rigid to introduce additional ways of sorting the users' behaviors and distinguish between the external unauthorized users. This has led to the development of ML models with growing interest in applying the technique for improving the classification and security of password-based systems.

That said, some issues are still prevalent even with the use of ML within this domain. Previous approaches often do not align well across multiple datasets, this means that while they perform well on a given set, they perform poorly

in real-world use cases. Furthermore, the training process comes with a drudgery of computation, especially when applied to enhance complicated ML models that are not easily practically deployed. Some general feature extraction methods include n-grams, and frequency analysis have been used in enhancing the models to predict but not exhausted in addressing password security aspect as most current models incorporate other general Cybersecurity features from outside the password domain.

However, with the emergence of latest vectorization methods like the use of Term Frequency-Inverse Document Frequency (TF-IDF) which is a technique in text mining, several patterns in passwords have lately been captured. Nevertheless, its applicability to the password classification for the security purpose has not been explored to its whole extent. Thus, this paper aims at filling these gaps with regards to the suitability of the TF-IDF vectorization approach, and the application of the password data in order to build a machine learning model to classify the users' behavior. The objectives of this research are as follows: this approach will have high accuracy, it will demonstrate high cross-dataset generality, and it should be computationally efficient.

3. METHDOLOGY

The proposed technique outlines a systematic approach, starting with problem identification and concluding with machine learning model evaluation.

3.1. PROBLEM IDENTIFICATION AND OBJECTIVE SETTING

This phase identifies the weaknesses of password-based systems, particularly the risks posed by unauthorized users altering or corrupting credentials. The goal is to classify computer users based on their password typologies using machine learning algorithms, thereby enhancing security.

3.2. LITERATURE REVIEW AND CRITICAL ASSESSMENT

This phase reviews existing solutions related to password categorization, user behavior analysis, and text vectorization methods like TF-IDF. It compares the strengths and weaknesses of these approaches, laying the groundwork for model development.

3.3. DATASET DESCRIPTION

The dataset consists of 733 records, with two columns: encoded usernames and their corresponding old passwords. A total of 147 users are represented, each with multiple password changes. For privacy, usernames are anonymized, and no personal data is included. The dataset was provided by a research data laboratory for analysis.

3.4. PASSWORD CHANGE FREQUENCY ANALYSIS

Each user's password changes are tracked chronologically, with the most recent at the bottom. Password formats vary, with many users following strong password regulations by using a mix of letters, numbers, and special characters.

3.5. ILLUSTRATION OF PASSWORD MODIFICATION FREQUENCY

USER70: Changed their password 4 times, moving from: Sun&3A → Boat#M8 → Moon*U5 → Cat$N4.

USER75: Changed their password 12 times, moving from: Cat$M3 → Dog#N6 → BBoatO4 → Moon$T8 → Sun22&N5 → Ttar#U3Y → Dog#M2 → BoatU9 → MoAAn$A6 → Sun&O4 → HHar@B8 → Cat$N3.

USER91: Changed their password 6 times, moving from: Wi-FiRouter#X4 → PowerBank$Y5 → Earbuds@Z6 → USBFlashDrive#M7 → E-reader$N8 → FitnessTracker@O9.

3.6. GENERAL PATTERNS OBSERVED

Most users changed their passwords between 4 and 8 times, following company standards requiring special characters, digits, and mixed-case letters. The patterns suggest that users made minor adjustments to passwords, likely for easier memorization, while still adhering to security protocols.

3.7. SECURITY IMPLICATIONS

Routine password modifications likely followed organizational regulations to maintain security. Users typically modified only parts of the password, indicating a balance between memorability and security.

3.8. DATASET USE CASE

The dataset offers insights into employee password management, including change frequency and trends in new password selection. These insights can help organizations strengthen security protocols and provide more effective user training. The anonymized data ensured privacy while supporting research into password practices.

3.9. PREPROCESSING

Data preparation involved label encoding for users and converting passwords into numerical vectors using TF-IDF, which quantifies the importance of password components based on their frequency.

3.10. TRAIN-TEST SPLIT

The dataset was divided into 80% training and 20% testing data to ensure effective model training and accurate performance evaluation.

3.11. MODEL SELECTION AND EVALUATION

Six classification models were assessed: SVM, Random Forest, Naive Bayes, KNN, Logistic Regression, and Decision Tree. Each model's performance was evaluated based on accuracy, precision, and recall, leading to the final ranking of the best models.

4. EXPERIMENTAL RESULTS

This study aimed to evaluate the effectiveness of various machine learning classification algorithms using Python. The process involved data preprocessing, model training, evaluation, and ranking based on performance. During data preparation, usernames were transformed into numerical labels using label encoding, and passwords were vectorized using Term Frequency-Inverse Document Frequency (TF-IDF). This converted the dataset into a format compatible with machine learning algorithms. Six classification models were applied: Support Vector Machines (SVM), Random Forest, Naive Bayes, K-Nearest Neighbors (KNN), Logistic Regression, and Decision Tree. The dataset was split, with 80% used for training and 20% for testing, allowing for accurate performance evaluation on unseen data. Each model's performance was measured using accuracy, precision, recall, and F1-score. These metrics helped assess the strengths and weaknesses of each algorithm in classifying users based on their password behaviors. The models were ranked to identify the most effective approach.

4.1. EVALUATION MEASURES

Precision is the ratio of true positive predictions to all positive predictions made by the model, indicating how often the model is correct when predicting a positive outcome. A high precision score means the model accurately identifies positive cases.

Recall measures the proportion of true positive predictions to all actual positive instances. A high recall score indicates the model's ability to correctly identify most positive cases.

The F1-score combines precision and recall into a single metric, representing the harmonic mean of the two. This is particularly useful when balancing both precision and recall, especially in cases of uneven feature distribution.

Support refers to the number of actual occurrences of each class, providing context for evaluating precision, recall, and F1-score.

Weighted Average Precision (WAP) adjusts the precision score to account for class imbalance by giving more weight to classes with more instances. This ensures that larger classes have a greater influence on the average score.

In imbalanced datasets, WAP provides a comprehensive measure of performance by addressing disparities between classes. Table 2 ranks the six classification models based on WAP, reflecting their overall effectiveness in predicting across all classes. According to the results, NaiveBayes performs the best with a score of 96.38% when considering weighted average accuracy. Weighted average accuracy scores of 91.64% for SVM and 91.52% for Logistic Regression place them second and third, respectively, behind NaiveBayes. DecisionTree and KNeighbors follow with scores of 77.55% and 79.48%, respectively, and RandomForest's weighted average precision of 71.26% places it last. According to weighted average accuracy, which reflects its efficacy across several classes, NaiveBayes is the top-performing model, as seen in this list.

5. RESULTS AND DISCUSSION

Let’s break into why accuracy could differ from the Weighted average precision:

SVM Precision vs. Weighted Avg Precision: If SVM has a high precision but a lower weighted

average precision, it might be performing very well on a particular class with few instances (high precision), but if it doesn't perform as well on other classes, the weighted average precision (which takes into account the number of instances) might be lower.

A comparison of Random Forest Precision and Weighted Average Precision reveals that Random Forest exhibits a lower precision, but its weighted average precision is even worse. This observation implies that although it exhibits satisfactory performance in certain classes, its performance is not sufficiently strong across all classes to get a high weighted average.

Logistic Regression accuracy vs. Weighted Avg Precision: Logistic Regression has identical results for accuracy and weighted average precision, demonstrating it performs consistently across different classes.

NaiveBayes Precision vs. Weighted Avg accuracy: NaiveBayes has a high precision and a greater weighted average accuracy, showing strong performance across all classes, particularly when considering the class imbalance.

KNeighbors and DecisionTree accuracy vs. Weighted Avg Precision: Both models demonstrate poor accuracy and weighted average precision. This suggests that individuals struggle with class identification across the board, and the low precision in specific classes impacts the overall weighted average.

5.1. EXAMPLE ANALYSIS

Consider the following hypothetical dataset:
Class A: 10 occurrences
Class B: 100 occurrences
Class C: 1000 occurrences

If a model has great precision for Class A but low recall, it could perform well when it predicts Class A but misses numerous cases. When weighted by the number of occurrences, Class C will dominate the weighted average precision since it has the most examples.

So, if a model is extremely exact on a less frequent class but has lesser precision on more frequent classes, its weighted average precision can be lower than its individual accuracy score for the less frequent class. In contrast, if a model regularly achieves high performance in all categories, the weighted average precision will accurately represent this consistency.

This weighted metric helps to guarantee that performance is not just good for the minority classes but is also balanced according to the actual distribution of the classes in the dataset.

5.2. RANKING BASED ON WEIGHTED AVG PRECISION

To rank the models, we use the weighted average accuracy since it takes into consideration class imbalance and the number of examples in each class:
NaiveBayes: 96.38%
SVM: 91.64%
Logistic Regression: 91.52%
KNeighbors: 79.48%
Decision Tree: 77.55%
Random Forest: 71.26%
With a weighted average accuracy of 96.38%, NaiveBayes emerges as the top model.

5.3. COMPARATIVE ANALYSIS USING EXISTING LITERATURE

Comparing these results to those in other research projects might be tricky because to changes in datasets, preprocessing, and assessment measures.

In the event that other research employ distinct datasets, characteristics, or assessment methods, direct comparisons may lack validity.

5.4. FACTORS TO CONSIDER

Dataset Characteristics: Differences in data distribution, quantity, and quality can impact model performance.

Feature Engineering: The choice of features and how they are handled may dramatically effect results.

The evaluation measures may vary across different investigations, thereby posing challenges in direct comparison.

Research Comparison:

In order to ascertain whether these results surpass those of others, it is necessary to guarantee a comparable dataset and approach. An analysis of similar outcomes will provide a more lucid understanding of the performance of these models in comparison to others in the existing body of research.

In summary, NaiveBayes performs the best according to weighted average precision in this collection of findings. However, comparing with other study necessitates evaluating data and

methodological variances to achieve an apples-to-apples comparison.

Table 2: Classification model Results.

accuracy	precision	recall	f1-score	support	weighted avg_precision
NaiveBayes	0.048	0.048	0.048	0.048	96.384
SVM	0.061	0.061	0.061	0.061	91.636
Logistic Regression	0.054	0.054	0.054	0.054	91.520
Kneighbors	0.034	0.034	0.034	0.034	79.478
DecisionTree	0.034	0.034	0.034	0.034	77.551
RandomForst	0.054	0.054	0.054	0.054	71.259

4.2 COMPARISON WITH LAST WORK [13]

Table 3: Comparison

Model	New Work Accuracy	New Work Precision	New Work Recall	New Work F1-Score	New Work Weighted Avg Precision	Previous Work Accuracy	Previous Work Precision	Previous Work Recall	Previous Work F1-Score
Naive Bayes	0.048	0.048	0.048	0.048	96.384%	N/A	N/A	N/A	N/A
SVM	0.061	0.061	0.061	0.061	91.636%	95.47%	66.46%	95.47%	0.067861
Logistic Regression	0.054	0.054	0.054	0.054	91.520%	N/A	N/A	N/A	N/A
KNeighbors	0.034	0.034	0.034	0.034	79.478%	65.53%	53.22%	65.53%	0.056132

Table 3 demonstrates the achieved improvement in this study, new work models improve the performance based on the weighted average precision irrespective of low accuracy score relative to the previous work. Here's a detailed comparison and explanation:

Naive Bayes (New Work):
Precision (96.384%): The model also obtained the highest value for the weighted average precision meaning that the number of correctly predicted positive values, that is, 'true positive', and the numbers of false positives are optimum.
This result can be considered scarce in the literature when as a metric for classification, while the precision is significant in password-based authentication, where reducing false positives is high.
Nonetheless, as measured by the accuracy of 0.048% and the other parameters of recall and F1-Score it is lower and this makes one to conclude that while positive outcome can be gotten with Naive Bayes where the algorithm excelled by predicting relevant classes with high percentage of precision, it was not so good at generalizing the above results to the whole data set. This calls for optimization to occur.

SVM (New Work):
Precision (91.636%) and Recall (61%): From the results of comparison it can be noticed that SVM has a rather satisfactory accuracy/ recall ratio to the previous work that reported only 66.4% of precision and 95.5 % of accuracy.
Improvement: The new work model greatly enhanced its performance in the classification task thus the better performance indicator were precision and recall scores with regard to password-based classification.

Logistic Regression (New Work):

Precision (91.520%) and Recall (54%): Like SVM, significantly more precision and recall were shown on identical datasets with Logistic Regression than in prior work, but the previous performance of this algorithm was not stated.
This therefore implies that the model of choice for this particular task in the new work may in fact be Logistic Regression due to it's precision in the new work unlike what was witnessed with the F1-Score on the previous work models.

KNeighbors (New Work):
Accuracy (79.478%) and Recall (79.478%): In new work two kinds of score boards namely, accuracy and recall values are increased in KNeighbors as compared to previous work the obtained accuracy score was very low 65.531.
Achieved: Despite having a mean weighted average of precision of only 79.478% KNeighbors provided enhanced results compared to the previous work and was better suited to dealing with the various classifications within the dataset.

Decision Tree (New Work):
Accuracy (77.551%) and Recall (77.551%): In the new work situation, this means that Decision Tree trended slightly lower than in previous work though had an Accuracy of 0.111792.
Why: The cause of the decline of performance could be overselling or problems with tree pruning, therefore, optimization might be required for this algorithm.
Random Forest (New Work):
Accuracy (71.259%): As can be seen Random Forest is noticeably worse than its accuracy in the previous work which was 0.128119 accuracy.
Even though Random Forest claims to work on big data in the prior work, its accuracy is slightly less in the new work and may feature problems like overfitting or incorrect hyperparameters optimization.
Achievements and Improvements:
The new work exceeded the previous work in terms of precision and recall for models such as SVM, Logistic Regression, and KNeighbors, with superior overall weighted average precision.
Other models of operations such as Naive Bayes updates proved to be more effective with increase in weighted precision, in anticipation of future application at systems that emphasize on correct identification of true positives.
The improvements of precision and recall SVM and Logistic Regression depict optimization and concentrated efforts in classification.
Although containing less precision, the new work highlights precision and recall values which have paramount importance for constructing a more accurate password-based authentication system. It suggests an improvement in terms of precision and recall and a better choice of models for classification problems with differential requirements for precision and volume of data, excluding excess false positives.

In general, the new work showed that encouraging measure of precision and recall can sufficiently enhance classification performance, particularly in the tasks associated with user authentication even though level of accuracy is marginally low compared to other methods. Future work may perhaps elaborate on the approach that has been used in the study by adjusting the numbers until the desired degree of accuracy and precision is reached to get better results.

6. CONCLUSION

This study meets the aforementioned research problem of enhancing password-based user classification due to the proposed application of machine learning models and the modification made involving TF-IDF vectorization to obtain relevant features from the password data. Naive Bayes model was proven to be the most efficient model with the average weighted precision of 96.38%, while comparing it to other models such as SVM and Logistic Regression that were also quite efficient but performed slightly lower. The results points to the fact that Naive Bayes was most effective in discerning unique passwords in different user classes suggesting the method can be used to improve on the reliability of the system in the identification of secure passwords.
Moreover, the object level models including K-Nearest Neighbors, Decision Tree, and Random Forests enhance the study by offering different levels of accuracy and levels of understanding concerning feature classification alternatives for the password-based security applications. This accomplishment illustrates that machine learning,

especially TF-IDF improved models, can greatly strengthen the cybersecurity system as the algorithms can now properly distinguish between legitimate and illegitimate tries of accessing computer systems using users' behavior.

Along with strengthening the accuracy and reliability of previously used models, this research equally addresses the limitations which the prior models can hardly overcome, namely lack of generalization and computational practicality for real-life applications, hence contributing to the conception of the efficient algorithms for password security operations. To this, the present work opens avenues for further development of studies on more complex machine learning paradigms and better feature extraction algorithms to build an effective and evolving model for user classification for passwords protection.

Acknowledgment

The author would ike to express sincere gratitude to the University of Buraimi for their invaluable support and funding of this research. This work has been made possible through the university's internal project titled "Advancing Faculty Competencies: A Theoretical Framework and Model Development for Collaborative Learning and Multicultural Faculty Twinning Utilizing an Interactive Social Media Information System." The continued support from the University of Buraimi has been instrumental in advancing research and fostering academic excellence.

REFERENCES

[1] Department of Computer Science and Engineering, Amity School of Engineering and Technology Lucknow, Amity University Uttar Pradesh, India, "Cyber Security Threats and Countermeasures in Digital Age," J. Appl. Sci. Educ., vol. 4, no. 1, pp. 1–20, 2024, doi: 10.54060/a2zjournals.jase.42.

[2]. Boumedyen Shannaq* and R. Adebiaye, "A security analysis To Be Technology Architecture for Ministry of Regional Municipalities and Water Resources (MRMWR) Sultanate of Oman," A security analysis To Be Technology Architecture for Ministry of Regional Municipalities and Water Resources (MRMWR) Sultanate of Oman, vol. 7, no. 4, Art. no. 4, 2019.

[3] B. Shannaq, "Digital Formative Assessment as a Transformative Educational Technology," in Advances in Information and Communication, vol. 921, K. Arai, Ed., in Lecture Notes in Networks and Systems, vol. 921. , Cham: Springer Nature Switzerland, 2024, pp. 471–481. doi: 10.1007/978-3-031-54053-0_32.

[4] C. Ugwu et al., "Factors Influencing The Experiences of End-users in Password-Based Authentication System," Jun. 03, 2024. doi: 10.21203/rs.3.rs-4438584/v1.

[5] B. Shannaq, M. A. Talab, M. Shakir, M. T. Sheker, and A. M. Farhan, "Machine learning model for managing the insider attacks in big data," presented at the the second international conference on emerging technology trends in internet of things and computing, Ramadi, Iraq, 2023, p. 020013. doi: 10.1063/5.0188358.

[6] J. Blessing, D. Hugenroth, R. J. Anderson, and A. R. Beresford, "SoK: Web Authentication in the Age of End-to-End Encryption," 2024, arXiv. doi: 10.48550/ARXIV.2406.18226.

[7] L. Y. Por, I. O. Ng, Y.-L. Chen, J. Yang, and C. S. Ku, "A Systematic Literature Review on the Security Attacks and Countermeasures Used in Graphical Passwords," IEEE Access, vol. 12, pp. 53408–53423, 2024, doi: 10.1109/ACCESS.2024.3373662.

[8] Dr. A.Shaji George, "The Dawn of Passkeys: Evaluating a Passwordless Future," Feb. 2024, doi: 10.5281/ZENODO.10697886.

[9] A. P. Umejiaku and V. S. Sheng, "RoseCliff Algorithm: Making Passwords Dynamic," Applied Sciences, vol. 14, no. 2, p. 723, Jan. 2024, doi: 10.3390/app14020723.

[10] B. Shannaq , M. T. Shakir , "Enhancing Security through Multi-Factor User Behavior Identification: Moving Beyond the Use of the Longest Common Subsequence (LCS)," IJCAI, vol. 48, no. 19, Nov. 2024, doi: 10.31449/inf.v48i19.6270.

[11] Akoh Atadoga, Enoch Oluwademilade Sodiya, Uchenna Joseph Umoga, and Olukunle Oladipupo Amoo, "A comprehensive review of machine learning's role in enhancing network security and threat detection," World J. Adv. Res. Rev., vol. 21,

no. 2, pp. 877–886, Feb. 2024, doi: 10.30574/wjarr.2024.21.2.0501.

[12] Ugochukwu Ikechukwu Okoli, Ogugua Chimezie Obi, Adebunmi Okechukwu Adewusi, and Temitayo Oluwaseun Abrahams, "Machine learning in cybersecurity: A review of threat detection and defense mechanisms," World J. Adv. Res. Rev., vol. 21, no. 1, pp. 2286–2295, Jan. 2024, doi: 10.30574/wjarr.2024.21.1.0315.

[13] B. Shannaq, O. Ali, S. A. Maqbali, and A. Al-Zeidi, "Advancing user classification models: A comparative analysis of machine learning approaches to enhance faculty password policies at the University of Buraimi," J. Infras. Policy. Dev., vol. 8, no. 13, p. 9311, Nov. 2024, doi: 10.24294/jipd9311.

[14] B. Shannaq and I. Al Shamsi, "Integrating Digital Transformation: Analyzing New Technological Processes for Competitiveness and Growth Opportunities in the Oman Economy," in The AI Revolution: Driving Business Innovation and Research, vol. 525, B. Awwad, Ed., in Studies in Systems, Decision and Control, vol. 525. , Cham: Springer Nature Switzerland, 2024, pp. 443–454. doi: 10.1007/978-3-031-54383-8_34.

[15] B. Shannaq, "Machine learning model-ensuring electronic exam quality using mining association rules," Machine Learning, vol. 29, no. 03, pp. 12136–12146, 2020.

[16] R. Othman, B. Rossi, and R. Barbara, "A Comparison of Vulnerability Feature Extraction Methods from Textual Attack Patterns," 2024, arXiv. doi: 10.48550/ARXIV.2407.06753.

[17] S. Alketbi, M. BinAmro, A. Alhammadi, and S. Kaddoura, "A Comparative Study of Machine Learning Models for Classification and Detection of Cybersecurity Threat in Hacking Forum," in 2024 15th Annual Undergraduate Research Conference on Applied Computing (URC), Dubai, United Arab Emirates: IEEE, Apr. 2024, pp. 1–6. doi: 10.1109/URC62276.2024.10604519.

[18] B. Shannaq, "Unveiling the Nexus: Exploring TAM Components Influencing Professors' Satisfaction With Smartphone Integration in Lectures: A Case Study From Oman," TEM Journal, pp. 2365–2375, Aug. 2024, doi: 10.18421/TEM133-63.

[19] D. Singla and N. Verma, "Performance Analysis of Authentication System: A Systematic LiteratureReview," RACSC, vol. 17, no. 7, p. e121223224363, Oct. 2024, doi: 10.2174/0126662558246531231121115514.

[20] R. Alrawili, A. A. S. AlQahtani, and M. K. Khan, "Comprehensive survey: Biometric user authentication application, evaluation, and discussion," Computers and Electrical Engineering, vol. 119, p. 109485, Oct. 2024, doi: 10.1016/j.compeleceng.2024.109485.

[21] M. K. Hasan, Z. Weichen, N. Safie, F. R. A. Ahmed, and T. M. Ghazal, "A Survey on Key Agreement and Authentication Protocol for Internet of Things Application," IEEE Access, vol. 12, pp. 61642–61666, 2024, doi: 10.1109/ACCESS.2024.3393567.

[22] N. Singh and A. K. Das, "TFAS: two factor authentication scheme for blockchain enabled IoMT using PUF and fuzzy extractor," J Supercomput, vol. 80, no. 1, pp. 865–914, Jan. 2024, doi: 10.1007/s11227-023-05507-6.

[23] S. Baseer and K. S. Charumathi, "Multi-Factor Authentication: A User Experience Study," SSRN Journal, 2024, doi: 10.2139/ssrn.4840295.

[24] L. Smith, S. Prior, and J. Ophoff, "Investigating the Accessibility and Usability of Multi-factor Authentication for Young People," in HCI International 2024 Posters, vol. 2119, C. Stephanidis, M. Antona, S. Ntoa, and G. Salvendy, Eds., in Communications in Computer and Information Science, vol. 2119. , Cham: Springer Nature Switzerland, 2024, pp. 129–135. doi: 10.1007/978-3-031-61966-3_15.

[25] I. Hagui et al., "A blockchain-based security system with light cryptography for user authentication security," Multimed Tools Appl, vol. 83, no. 17, pp. 52451–52480, Nov. 2023, doi: 10.1007/s11042-023-17643-5.

[26] C. Manthiramoorthy, K. M. S. Khan, and N. A. A, "Comparing Several Encrypted Cloud Storage Platforms," ijmscs, vol. 2, pp. 44–62, Aug. 2023, doi: 10.59543/ijmscs.v2i.7971.

[27] Y. Shi and Y. Wang, "A Comparative Work to Highlight the Superiority of Mouth Brooding Fish (MBF) over the Various ML Techniques in Password Security

Classification," ijacsa, vol. 15, no. 5, 2024, doi: 10.14569/IJACSA.2024.0150520.

[28] N. F. Almujahid, M. A. Haq, and M. Alshehri, "Comparative evaluation of machine learning algorithms for phishing site detection," PeerJ Computer Science, vol. 10, p. e2131, Jun. 2024, doi: 10.7717/peerj-cs.2131.

[29] S. Aboukadri, A. Ouaddah, and A. Mezrioui, "Machine learning in identity and access management systems: Survey and deep dive," Computers & Security, vol. 139, p. 103729, Apr. 2024, doi: 10.1016/j.cose.2024.103729.

[30] N. Andelić, S. Baressi Šegota, and Z. Car, "Robust password security: a genetic programming approach with imbalanced dataset handling," Int. J. Inf. Secur., vol. 23, no. 3, pp. 1761–1786, Jun. 2024, doi: 10.1007/s10207-024-00814-2.

[31] B. Shannaq, I. Al Shamsi, and S. Abdul Majeed, "Management Information System for Predicting Quantity Martial's," TEM Journal, vol. 8, pp. 1143–1149, Dec. 2019, doi: 10.18421/TEM84-06.

[32] Yuhong Mo, Shaojie Li, Yushan Dong, Ziyi Zhu, and Zhenglin Li, "Password Complexity Prediction Based on RoBERTa Algorithm," May 2024, doi: 10.5281/ZENODO.11180356.

[33] Sasibhushan Rao Chanthati, "How the power of machine – machine learning, data science and NLP can be used to prevent spoofing and reduce financial risks," Global J. Eng. Technol. Adv., vol. 20, no. 2, pp. 100–119, Aug. 2024, doi: 10.30574/gjeta.2024.20.2.0149.

[34] N. Dias, M. S. K, and R. S R, "Natural language processing and stable diffusion model based graphical authentication using passphrase," IDT, vol. 18, no. 2, pp. 935–951, Jun. 2024, doi: 10.3233/IDT-230279.

[35] H. Chen and M. A. Babar, "Security for Machine Learning-based Software Systems: A Survey of Threats, Practices, and Challenges," ACM Comput. Surv., vol. 56, no. 6, pp. 1–38, Jun. 2024, doi: 10.1145/3638531.

[36] M. Shakir, A. ABUBAKAR, Y. Yusoff, M. Al-Emran, and M. Hammood, "Application of confidence range algorithm in recognizing user behavior through EPSB in cloud computing," Journal of Theoretical and Applied Information Technology, vol. 94, p. 416_427, Dec. 2016.

[37] M. J. Rooney, Y. Levy, W. Li, and A. Kumar, "Comparing experts' and users' perspectives on the use of password workarounds and the risk of data breaches," ICS, Jul. 2024, doi: 10.1108/ICS-05-2024-0116.

[38] M. Shakir, M. J. Al Farsi, I. R. Al-Shamsi, B. Shannaq, and T.-H. Ghilan Al-Madhagy, "The Influence of Mobile Information Systems Implementation on Enhancing Human Resource Performance Skills: An Applied Study in a Small Organization," Int. J. Interact. Mob. Technol., vol. 18, no. 13, pp. 37–68, Jul. 2024, doi: 10.3991/ijim.v18i13.47027.

[39] N. Lykousas and C. Patsakis, "Decoding developer password patterns: A comparative analysis of password extraction and selection practices," Computers & Security, vol. 145, p. 103974, Oct. 2024, doi: 10.1016/j.cose.2024.103974.

[40] H. Wasfi, R. Stone, and U. Genschel, "Word-Pattern: Enhancement of Usability and Security of User-Chosen Recognition Textual Password," ijacsa, vol. 15, no. 6, 2024, doi: 10.14569/IJACSA.2024.0150605.

[41] A. Gautam, T. K. Yadav, K. Seamons, and S. Ruoti, "Passwords Are Meant to Be Secret: A Practical Secure Password Entry Channel for Web Browsers," 2024, arXiv. doi: 10.48550/ARXIV.2402.06159.

[42] S. Vanila, Beaulah Jeyavathana, A. Rathinam, and K. Elango, "Enhancing Password Security With Machine Learning-Based Strength Assessment Techniques:," in Advances in Information Security, Privacy, and Ethics, J. A. Ruth, V. G. V. Mahesh, P. Visalakshi, R. Uma, and A. Meenakshi, Eds., IGI Global, 2024, pp. 296–314. doi: 10.4018/979-8-3693-4159-9.ch018.

[43] M. Atzori, E. Calò, L. Caruccio, S. Cirillo, G. Polese, and G. Solimando, "Evaluating password strength based on information spread on social networks: A combined approach relying on data reconstruction and generative models," Online Social Networks and Media, vol. 42, p. 100278, Aug. 2024, doi: 10.1016/j.osnem.2024.100278.

[44] E. Altulaihan, M. A. Almaiah, and A. Aljughaiman, "Anomaly Detection IDS for Detecting DoS Attacks in IoT Networks

Based on Machine Learning Algorithms," Sensors, vol. 24, no. 2, p. 713, Jan. 2024, doi: 10.3390/s24020713.

[45] A. Komadina, I. Kovačević, B. Štengl, and S. Groš, "Comparative Analysis of Anomaly Detection Approaches in Firewall Logs: Integrating Light-Weight Synthesis of Security Logs and Artificially Generated Attack Detection," Sensors, vol. 24, no. 8, p. 2636, Apr. 2024, doi: 10.3390/s24082636.

[46] A. G. Martín, I. Martín De Diego, A. Fernández-Isabel, M. Beltrán, and R. R. Fernández, "Combining user behavioural information at the feature level to enhance continuous authentication systems," Knowledge-Based Systems, vol. 244, p. 108544, May 2022, doi: 10.1016/j.knosys.2022.108544.

[47] V. Papaspirou et al., "A Novel Authentication Method That Combines Honeytokens and Google Authenticator," Information, vol. 14, no. 7, p. 386, Jul. 2023, doi: 10.3390/info14070386.

[48] E. Escobar-Linero, F. Luna-Perejón, L. Muñoz-Saavedra, J. L. Sevillano, and M. Domínguez-Morales, "On the feature extraction process in machine learning. An experimental study about guided versus non-guided process in falling detection systems," Engineering Applications of Artificial Intelligence, vol. 114, p. 105170, Sep. 2022, doi: 10.1016/j.engappai.2022.105170.

[49] Chee Keong Ng, Tahsien Al-Quraishi, and Tony De Souza-Daw, "Application of Sequential Analysis on Runtime Behavior for Ransomware Classification," Applied Data Science and Analysis, vol. 2023, pp. 126–142, Nov. 2023, doi: 10.58496/ADSA/2023/012.

[50] R. Veras, C. Collins, and J. Thorpe, "A Large-Scale Analysis of the Semantic Password Model and Linguistic Patterns in Passwords," ACM Trans. Priv. Secur., vol. 24, no. 3, pp. 1–21, Aug. 2021, doi: 10.1145/3448608.

[51] W. Fraser, M. Broadbent, N. Pitropakis, and C. Chrysoulas, "Examining the Strength of Three Word Passwords," in ICT Systems Security and Privacy Protection, vol. 710, N. Pitropakis, S. Katsikas, S. Furnell, and K. Markantonakis, Eds., in IFIP Advances in Information and Communication Technology, vol. 710. , Cham: Springer Nature Switzerland, 2024, pp. 119–133. doi: 10.1007/978-3-031-65175-5_9.

[52] X. Yu and Q. Liao, "User password repetitive patterns analysis and visualization," Information & Computer Security, vol. 24, no. 1, pp. 93–115, Mar. 2016, doi: 10.1108/ICS-06-2015-0026.

[53] E. Zhou, Y. Peng, G. Shao, F. Deng, Y. Miao, and W. Fan, "Password cracking using chunk similarity," Future Generation Computer Systems, vol. 150, pp. 380–394, Jan. 2024, doi: 10.1016/j.future.2023.09.013.

[54] N. P. S. Pendela, K. A. Janet, A. M. R. Yadav, C. B. Subramanyam, S. Hariharan, and V. Kekreja, "Enhancing Cyberbullying Detection: A Multi-Algorithmic Approach," in 2024 International Conference on Advances in Data Engineering and Intelligent Computing Systems (ADICS), Chennai, India: IEEE, Apr. 2024, pp. 1–5. doi: 10.1109/ADICS58448.2024.10533585.

[55] B. Harshita and N. Leema, "ESD: E-mail Spam Detection using Cybersecurity-Driven Header Analysis and Machine Learning based Content Analysis," Int J Performability Eng, vol. 20, no. 4, p. 205, 2024, doi: 10.23940/ijpe.24.04.p2.205213.

[56] O. Aouedi et al., "A Survey on Intelligent Internet of Things: Applications, Security, Privacy, and Future Directions," IEEE Commun. Surv. Tutorials, pp. 1–1, 2024, doi: 10.1109/COMST.2024.3430368.

[57] C. Maçãs, J. R. Campos, N. Lourenço, and P. Machado, "Visualisation of Random Forest classification," Information Visualization, p. 14738716241260745, Jun. 2024, doi: 10.1177/14738716241260745.

[58] S. Etzler, F. D. Schönbrodt, F. Pargent, R. Eher, and M. Rettenberger, "Machine Learning and Risk Assessment: Random Forest Does Not Outperform Logistic Regression in the Prediction of Sexual Recidivism," Assessment, vol. 31, no. 2, pp. 460–481, Mar. 2024, doi: 10.1177/10731911231164624.

[59] J. Han, "CNN-Based Multi-Factor Authentication System for Mobile Devices Using Faces and Passwords," Applied Sciences, vol. 14, no. 12, p. 5019, Jun. 2024, doi: 10.3390/app14125019.

[60] K. Kaur and P. Kaur, "SABDM: A self-attention based bidirectional-RNN deep model for requirements classification," J

Software Evolu Process, p. e2430, Feb. 2022, doi: 10.1002/smr.2430.

[61] D.-W. Zhou, Z.-W. Cai, H.-J. Ye, D.-C. Zhan, and Z. Liu, "Revisiting Class-Incremental Learning with Pre-Trained Models: Generalizability and Adaptivity are All You Need," Int J Comput Vis, Aug. 2024, doi: 10.1007/s11263-024-02218-0.

[62] Z. Liu and K. He, "A Decade's Battle on Dataset Bias: Are We There Yet?," 2024, arXiv. doi: 10.48550/ARXIV.2403.08632.

[63] A. Bakhtiarnia, Q. Zhang, and A. Iosifidis, "Efficient High-Resolution Deep Learning: A Survey," ACM Comput. Surv., vol. 56, no. 7, pp. 1–35, Jul. 2024, doi: 10.1145/3645107.

[64] Y. Wang, Y. Han, C. Wang, S. Song, Q. Tian, and G. Huang, "Computation-efficient deep learning for computer vision: A survey," Cybernetics and Intelligence, pp. 1–24, 2024, doi: 10.26599/CAI.2024.9390002.

[65] Halima Oluwabunmi Bello, Adebimpe Bolatito Ige, and Maxwell Nana Ameyaw, "Deep learning in high-frequency trading: Conceptual challenges and solutions for real-time fraud detection," World J. Adv. Eng. Technol. Sci., vol. 12, no. 2, pp. 035–046, Jul. 2024, doi: 10.30574/wjaets.2024.12.2.0265.