Predicting Individuals' Vulnerability to Social Engineering Attacks

Tesfaye Abaya
HiLCoE School of Computer Science & Technology
jadon.tesfaye@gmail.com

Tibebe Beshah
School of Information Science, Addis Ababa University
tibebe.beshah@aau.edu.et

Abstract

The increasing prevalence of social engineering attacks (SEAs) poses a significant threat to individuals and organizations [1]. This study aims to develop and evaluate a predictive model for identifying individuals susceptible to SEAs. A dataset comprising 505 responses from university and colleges students in Addis Ababa, Ethiopia, was collected through an online survey. To address data insufficiency, synthetic data generation was employed, expanding the dataset to 3000 instances [2]. Employing a Design Science Research (DSR) methodology, multiple machine learning algorithms were compared, with Random Forest demonstrating superior performance in predicting SEA vulnerability. Key predictors identified include age, income, occupation, curiosity, and trust factors. The findings contribute to enhancing cybersecurity measures by enabling targeted interventions to protect vulnerable individuals.

Keywords- Social engineering attacks, Predictive model, Machine Learning, Random Forest, Cybersecurity

1. Introduction

The escalating reliance on digital technologies has ushered in an era characterized by unprecedented connectivity and data exchange. Concurrently, the landscape of cyber threats has evolved, with social engineering attacks (SEAs) emerging as a formidable challenge [3] [4]. These attacks, which exploit human psychology rather than technical vulnerabilities, have proven highly effective in compromising individuals and organizations alike. Despite advancements in cybersecurity, the human element remains a critical weak link [4].

The increasing sophistication of SEAs, coupled with the lack of comprehensive research on individual vulnerability, underscores the urgency of this study. Existing research often overlooks the complex interplay between demographic, psychological, and behavioral factors that contribute to susceptibility. This research aims to develop and evaluate a predictive model capable of identifying individuals at higher risk of falling victim to SEAs. By understanding these vulnerabilities, targeted interventions can be implemented to enhance cybersecurity defences.

To achieve this objective, a dataset comprising 505 responses from university students in Addis Ababa, Ethiopia, was collected through an online survey. Due to data insufficiency, synthetic data generation techniques were employed to expand the dataset to 3000 instances [2]. A Design Science Research (DSR) methodology was adopted for model development and evaluation. Multiple machine learning algorithms were compared, with Random Forest demonstrating superior performance in predicting SEA vulnerability.

2. Literature Review

The escalating prevalence of social engineering attacks (SEAs) has prompted extensive research into identifying factors influencing individual vulnerability. Previous studies have explored various dimensions of user susceptibility, including demographic, psychological, and behavioral factors.

Albladi and Weir [5] investigated the impact of user behavior, perceptions, and socio- emotional factors on social engineering vulnerability within social networks. Their study revealed that factors such as social media involvement, risk perception, and trust influence susceptibility. However, the focus on a specific social media platform and academic population limits the generalizability of their findings.

Huseynov and Ozdenizci [6] employed machine learning algorithms to predict social engineering susceptibility based on demographics, technology usage, and personality traits. While their study demonstrated the potential of machine learning in this domain, it focused primarily on email and SMS phishing, limiting its scope.

Wang et al. [7] examined the role of demographics, personality, cognitive processes, knowledge, and experience in predicting phishing vulnerability. Their research highlighted the importance of these factors in susceptibility but was constrained by a focus on email phishing. These studies collectively contribute to our understanding of social engineering vulnerability. However, they also reveal significant gaps in knowledge. Existing research often adopts a narrow perspective, focusing on specific attack types or limited demographic groups. Furthermore, the role of app usage behavior and social media habits in influencing susceptibility remains underexplored.

To address these limitations, this study expands the scope of inquiry by investigating a broader range of user characteristics. By incorporating app usage behavior, social media habits, and diverse demographic factors, this research aims to provide a more comprehensive understanding of individual vulnerability to SEAs. Additionally, the study employs a comparative analysis of multiple machine learning algorithms to identify the most effective model for predicting susceptibility.

This research builds upon the foundation laid by previous studies while venturing into new territory to provide a more comprehensive picture of social engineering vulnerability.

3. Methodology

3.1 Overview

This section outlines the methodology employed to develop a predictive model for assessing individual vulnerability to social engineering attacks (SEAs). It encompasses data collection strategies, sampling methods, experimentation techniques, and evaluation metrics utilized to assess the model's performance.

3.2. General Approach and Research Design

This research adopts a Design Science Research (DSR) framework, specifically following the process outlined by Offermann et al [8]. DSR is characterized by its iterative nature, focusing on the creation and evaluation of innovative artifacts to resolve real-world challenges. The DSR process consists of several key phases: problem identification, design and development, experimentation, and evaluation [8].

3.2.1 Problem Identification

The research commenced with an extensive literature review to identify existing challenges in cybersecurity, particularly the rising incidence of social engineering attacks, which are reported to constitute up to 98% of cyber incidents globally [9]. Insights from interviews with cybersecurity professionals further underscored the urgency of addressing the vulnerabilities associated with these attacks, particularly in the Ethiopian context, where tactics are becoming increasingly sophisticated.

3.2.2 Solution Design

The solution design phase involved several critical steps:

Data Preparation and Exploration

The core artifact of this research is a predictive model aimed at identifying individuals at risk of SEAs. The development process includes:

Technology Stack: The implementation utilized Python, leveraging libraries such as scikit- learn for machine learning, alongside tools like Visual Studio and Jupyter Notebook for data exploration and coding.

Data Acquisition and Synthetic Data Generation: Due to the lack of publicly available datasets relevant to the Ethiopian context, an online survey was conducted, yielding 505 responses. To enhance this dataset, the SMOTE (Synthetic Minority Over-sampling Technique) algorithm was applied, increasing the dataset to 3,000 instances. This step was crucial for balancing the dataset and improving model accuracy [2].

Data Preprocessing: The collected data underwent preprocessing to prepare it for machine learning applications. The dataset was divided into training (80%) and testing (20%) subsets.

Model Development

This stage focused on training and testing the predictive model using three machine

learning algorithms: Logistic Regression (LR), Decision Tree (DT), and Random Forest (RF). Each algorithm was rigorously tuned through hyperparameter [10] optimization to enhance performance. This comprehensive approach aimed to provide insights into the model's capabilities across various scenarios.

3.3 Evaluation

The evaluation of the model's performance involved several established metrics, including accuracy, precision, recall, and F1 score, particularly suited for binary classification tasks. To ensure robust performance estimates and mitigate overfitting, the following techniques were employed:

Data Splitting: The dataset was divided into training and testing sets using an 80/20 split, ensuring that the model was trained on a representative sample while retaining a separate set for unbiased evaluation.

Cross-Validation: A 5-fold cross-validation method was implemented, allowing for iterative training and evaluation on different data subsets. This technique enhances the model's generalizability to unseen data.

After conducting the evaluations, the findings were meticulously documented and analyzed, providing a comprehensive understanding of the model's strengths, weaknesses, and potential limitations. The F1 macro score was selected as the primary evaluation metric, reflecting the model's performance in predicting individual vulnerability to social engineering attacks.

4. Solution Design/ Data Experimentation/- Model Development

4.1 Experimental Setup

Data Splitting:

To ensure a fair evaluation of our machine learning models, we divided our dataset of **3000 instances** into training and testing sets. The **training set** comprised **80%** of the data, which was used to train the models. The remaining **20%** was allocated to the **testing set**, serving as a holdout set for evaluating the models' performance on unseen data.

Class Distribution Analysis:

Before proceeding with the experimentation, we carefully analyzed the class distribution within the training and testing sets. The specific class labels "high," "moderate," and "low" were defined based on domain expertise and predefined criteria, such as the frequency of phishing attempts, password breaches, or social media scams.

Table 1 Training and Test Data Class Distribution

Class	Training Data (Instances)	Training Data (Percentage)	Testing Data (Instances)	Testing Data (Percentage)
Low	789	32.90%	199	33.20%
Medium	1004	41.80%	248	41.30%
High	607	25.30%	153	25.50%

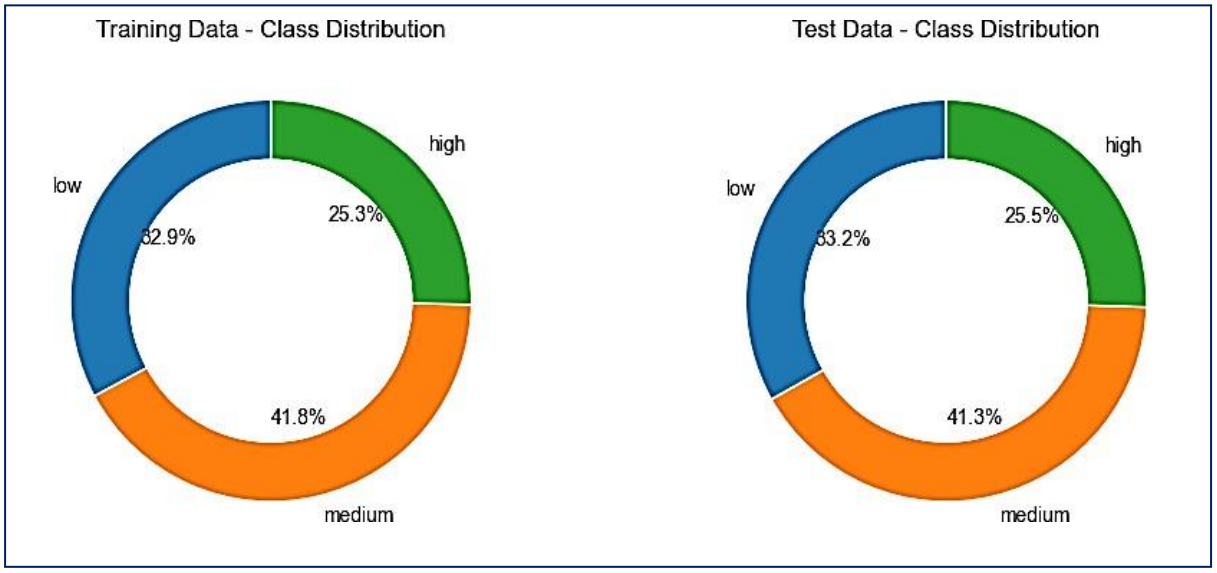

Figure 1 Training and Test Data Class Distribution

Evaluation Metrics:

We selected the following evaluation metrics to assess the performance of our machine learning models:

- **Accuracy:** This measures the overall proportion of correct predictions.
- **Precision:** This indicates the proportion of positive predictions that were actually correct.
- **Recall:** This measures the proportion of actual positive cases that were correctly predicted.
- **F1-score:** This is the harmonic mean of precision and recall, providing a balanced measure of both.
- **Cross-Validation:** To obtain reliable performance estimates and mitigate biases, we incorporated cross- validation techniques. In this study, we utilized a 5-fold cross-validation scheme, where the data was split into five subsets and the model was trained and evaluated on each subset in turn. This approach ensured a more

robust evaluation of the model's generalizability.

- **Hyperparameter Tuning for Model Optimization**

Hyperparameter [10] tuning was a critical component of our research. By carefully adjusting the hyperparameters of logistic regression, decision trees, and random forests, we were able to optimize their performance for social engineering vulnerability prediction. This process ensured a balance between model complexity and generalizability, allowing us to extract the maximum potential from each model.

Through hyperparameter tuning [10], we were able to tailor each model's configuration to the specific characteristics of our dataset. This customization helped us identify the most effective model for our task, ensuring that our final choice was not solely based on default settings. By mitigating the influence of hyperparameter choices on the final performance comparison, we were able to make a more objective and fair selection of the best model for social engineering vulnerability prediction.

4.2 Laboratory Experiment

4.2.1 Model 1 Logistic Regression

This research leverages logistic regression (LR) to develop a predictive model for assessing individual vulnerability to social engineering attacks (SEA). To address the challenge of missing data, we employed mean imputation to replace missing values with the average of observed values within each feature. This enriched dataset allowed the LR model to capture a broader range of relationships and interactions between variables, potentially improving its predictive accuracy and generalizability.

Data Splitting and Model Training

We utilized a dataset of 3000 instances, split into an 80/20 ratio for training and testing. This division provided sufficient data for model training while reserving a significant portion for unbiased evaluation.

Model Performance:

Our LR model, optimized through hyperparameter tuning, achieved a moderate level of predictive capability. The model exhibited an accuracy of 75.17%, correctly classifying approximately three-quarters of the data points. However, a more nuanced analysis revealed class-specific performance variations.

- **Class-Specific Performance:**
    - **Class 0 (low vulnerable):** Achieved the highest precision (91.08%) and recall (94.76%), indicating a low rate of false positives and false negatives.
    - **Class 1 (highly vulnerable):** Demonstrated the lowest precision (62.69%) and recall (55.63%), suggesting a higher risk of misclassification.

- **Class 2 (moderately vulnerable):** Achieved a moderate F1-score of 64.20%.

Impact of Class Imbalance:

The class-wise support values (number of instances) revealed an imbalanced distribution, with Class 1 having the highest support. This imbalance may have influenced the overall model performance, particularly for Class 2.

Our logistic regression model, optimized through hyperparameter tuning after data imputation, exhibits moderate predictive capability for identifying individuals vulnerable to social engineering attacks (SEA). While the achieved accuracy is encouraging, further investigation and potential refinements are needed to improve the model's generalizability and performance across all vulnerability categories.

Key Findings of LR:

- The LR model achieved an overall accuracy of 75.17%.
- Class-specific performance varied, with stronger results for low vulnerable individuals and weaker results for highly vulnerable individuals.
- Class imbalance may have influenced the model's performance, particularly for the highly vulnerable class.

4.2.2 Model 2 Decision Tree

In this study, after addressing missing values through data imputation, the second machine learning model we used to experiment on our data is decision tree. We applied hyperparameter tuning to a decision tree (DT) model to predict individual vulnerability to social engineering attacks (IVSEA). The tuning process focused on optimizing the model’s performance by identifying the best values for parameters such as maximum depth, minimum samples per leaf, and minimum samples per split. The optimal hyperparameters were determined to be: max_depth = 7, min_samples_leaf = 1, and min_samples_split = 2.

The dataset, consisting of 3,000 instances, was divided into an 80/20 split for training and testing. This split ensured a robust training process while reserving enough data for unbiased model evaluation. The DT model achieved an accuracy of 86.17%, demonstrating strong predictive capability in identifying individuals vulnerable to social engineering attacks.

The model’s performance across different classes varied. Class 1, which represented individuals most vulnerable to attacks, exhibited the highest precision (96.76%) and recall (96.37%), indicating accurate and consistent identification of this group. Class 2, representing a lower vulnerability, had the lowest precision (77.55%) and recall (75.50%), suggesting a higher rate of misclassification in this category. The F1-scores, which balance precision and recall, were highest for Class 1 (96.57%), followed by Class 0 (80.99%) and Class 2 (76.51%). Support values, representing the number of instances in each class, showed that Class 1 had the highest representation (248

instances), followed by Class 0 (199 instances) and Class 2 (151 instances). These values reflect the distribution of the dataset across different vulnerability categories.

In conclusion, the decision tree model, with optimized hyperparameters, showed strong overall performance in predicting individual vulnerability to social engineering attacks. However, the variation in precision, recall, and F1-scores across the classes indicates that further refinement of the model could improve its consistency, particularly for Class 2.

4.2.3 Model 3 Random Forest

In this study, a Random Forest (RF) model was implemented to predict individual vulnerability to social engineering attacks (IVSEA), with a particular focus on enhancing model performance through hyperparameter tuning. The dataset, comprising 3,000 instances, was divided into an 80/20 split for training and testing, ensuring a sufficient portion of data for model training while preserving a significant amount for unbiased evaluation.

Hyperparameter tuning played a crucial role in optimizing the Random Forest model. By adjusting key parameters such as the number of trees, maximum depth, and minimum samples per leaf, the model's performance was significantly enhanced, leading to superior predictive accuracy. The final model achieved an impressive accuracy of 99.67%, correctly classifying vulnerability status for nearly all instances in the dataset.

The precision and recall scores across all vulnerability classes were equally remarkable. For Class 0 (non-vulnerable individuals), both precision and recall were near perfect at 0.97, signifying that the model successfully identified almost all non-vulnerable cases. In Class 1 (vulnerable individuals), the model achieved a precision score of 0.992 and a perfect recall score of 1.00, indicating that it effectively captured all vulnerable individuals. For Class 2, both precision and recall scores were flawless at 1.00, highlighting the model's exceptional performance in this category.

The F1-scores, which measure the balance between precision and recall, were similarly outstanding. Class 0 achieved a perfect F1-score of 1.00, reflecting its strong classification performance. Class 1 had an F1-score of 0.996, and Class 2 achieved an F1-score of 0.993, further underscoring the model's balanced and robust predictive capability across all classes. Support values, representing the number of instances per class, revealed that Class 1 had the largest representation with 248 instances, followed by Class 0 with 199 instances, and Class 2 with 151 instances. This distribution helped illustrate the consistency of the model across varying class sizes.

Overall, the implementation of hyperparameter tuning significantly enhanced the Random Forest model's predictive power. The near-perfect accuracy, precision, recall, and F1-scores demonstrate the model's effectiveness in identifying vulnerable individuals, making it a reliable tool for predicting individual vulnerability to social engineering attacks. These findings suggest that the Random Forest model, when fine-

tuned with optimized hyperparameters, is an exceptionally accurate and robust method for addressing cybersecurity vulnerability prediction.

5. Results and Discussions

5.1 Results

This study aimed to assess the effectiveness of a Random Forest model with optimized hyperparameters in predicting individual susceptibility to social engineering attacks (SEAs). By fine-tuning the model's parameters, we sought to enhance its predictive power and identify key factors influencing vulnerability. The analysis revealed several significant predictors, ranked by their importance in the model:

Age: Age emerged as the most influential factor, contributing 38% to the model's predictive power. The findings indicate that older individuals are more susceptible to SEAs, highlighting the need for targeted interventions for this demographic.

Average Monthly Income: With a 10% importance value, income was another major predictor of vulnerability. The data suggests that individuals with higher incomes may be more prone to social engineering tactics, perhaps due to increased financial targets or overconfidence in their digital interactions.

Occupation: Occupation accounted for 6% of the predictive power, with individuals working in science and business fields being more vulnerable to SEAs. Further investigation is required to explore the specific factors within these professions that may contribute to this heightened susceptibility.

Curiosity: This trait was associated with a 5% predictive value. Individuals with higher levels of curiosity exhibited an increased susceptibility to SEAs, suggesting that attackers may exploit curiosity-driven behavior through deceptive tactics.

Trust in Apps: Trusting apps had a 5% importance value, indicating that individuals who have high trust in apps are moderately more vulnerable to SEAs. This underscores the need for awareness regarding app-related threats, as attackers may leverage users' trust in seemingly legitimate applications.

Trust in Friends from Contact Lists: Trust placed in contacts also had a 5% predictive value. This finding highlights the potential for social engineering attackers to exploit pre-existing trust relationships, often relying on contacts or impersonation tactics.

Self-Efficacy and Extroversion: Both traits contributed 4% to the model's predictive power. The data suggests that individuals with higher self-efficacy and extroversion may be more vulnerable to SEAs, possibly due to overconfidence or social interactions that create opportunities for manipulation.

Field of Study: Educational background, reflected by the field of study, also had a 4% predictive value. This suggests that certain educational disciplines may correlate

with varying levels of susceptibility to SEAs.

Duration of Social Media Usage: Daily social media usage contributed 3% to the predictive power. Individuals with higher usage showed a moderate increase in vulnerability, possibly due to increased exposure to attack vectors commonly found on social media platforms.

Frequency of App Updates: Surprisingly, the frequency of updating apps had a lower importance value (2%) in predicting SEA susceptibility. The data indicated that those who occasionally update their apps might be more vulnerable than those who update regularly, although the overall effect was minimal.

In summary, the optimized Random Forest model identified age (38%) and average monthly income (10%) as the most critical factors in predicting susceptibility to SEAs. These findings suggest that cybersecurity awareness initiatives could be particularly effective if tailored to these demographics, emphasizing the importance of educating older and higher-income individuals about the risks of social engineering. Additionally, while other factors such as occupation, curiosity, and social media usage play smaller roles (2% to 6%), they remain relevant in understanding overall vulnerability. The model's results provide a comprehensive framework for developing targeted interventions aimed at mitigating the risks posed by social engineering attacks.

5.2 Discussions

This study provides a comprehensive analysis of individual susceptibility to social engineering attacks (SEAs) using hyperparameter-tuned machine learning models (MLMs), particularly the Decision Tree (DT), Random Forest (RF), and Logistic Regression (LR) models. The focus was on identifying demographic, behavioral, and psychological factors that contribute to vulnerability, and evaluating the models' performance in predicting individuals' susceptibility to social engineering attacks in Ethiopia.

Key features, such as age, gender, educational level, field of study, occupation, average monthly income, social media usage patterns, and frequency of updating apps and operating systems (OS), were analyzed. These features exhibited similar distributions across all target classes, with some variations across different levels of vulnerability. The analysis found that factors like age, number of social media contacts, and duration of social media usage per day displayed distinct distributions across all target classes, indicating their significance in differentiating vulnerability levels.

In terms of model performance, the Random Forest (RF) model emerged as the best performer, achieving an accuracy of 99.67%, surpassing both the Logistic Regression (LR) model with 75.1% accuracy and the Decision Tree (DT) model with 86.7% accuracy. This high accuracy suggests that the RF model is highly effective in identifying individuals who are vulnerable to SEAs. The RF model’s success can be attributed to hyperparameter tuning and data imputation, which improved its performance over other models.

Several factors emerged as key predictors of vulnerability. Age (38% importance value) and average monthly income (10%) were the most significant contributors. Older individuals and those with higher incomes were more vulnerable to SEAs, likely due to factors such as increased financial exposure and less familiarity with modern online threats. Occupation (6%), particularly individuals in science and business fields, also played a crucial role in determining vulnerability, likely due to the nature of their work making them more susceptible to certain social engineering tactics. Other factors such as curiosity (5%), trust in apps (5%), and trust in social connections (5%) demonstrated moderate predictive power, emphasizing the role of behavioral traits in influencing susceptibility.

The findings underscore the need for cybersecurity awareness programs tailored to these key factors, particularly focusing on older individuals and those with higher incomes, as well as individuals in specific professional fields. By addressing the behavioral tendencies that make individuals more vulnerable, such as curiosity and trust in apps and contacts, targeted interventions can be designed to improve resilience against SEAs.

While accuracy is a critical metric, other evaluation measures such as precision, recall, and F1- score should be considered to gain a comprehensive understanding of the model's performance. Future research should validate the findings across broader datasets and populations to ensure the model's generalizability and reliability in predicting individual vulnerability to SEAs in different contexts.

In conclusion, this research advances the understanding of social engineering attack vulnerability by identifying key demographic and behavioral factors that influence susceptibility. The Random Forest model, with its optimized hyperparameters, demonstrated superior performance in predicting vulnerability, highlighting its potential as a reliable tool for cybersecurity risk assessment. The findings also emphasize the importance of tailored security awareness programs that consider individuals' unique demographic and behavioral profiles to mitigate the risks posed by SEAs effectively.

6. Conclusions

The research successfully addressed the core questions and objectives outlined at the beginning of the study, providing valuable insights into the prediction of individual vulnerability to social engineering attacks (SEAs) using machine learning models (MLMs).

What type of individuals are most vulnerable to SEA?

This study identified several key demographic and behavioral characteristics that contribute to an individual's susceptibility to SEAs. Age emerged as the most influential factor, with older individuals being more vulnerable. Income, occupation, and curiosity also played significant roles, highlighting that individuals with higher incomes and those working in fields such as science and business were more likely to

fall victim to these attacks. Additionally, trust in apps and social connections further heightened vulnerability, as attackers exploit these trusted relationships.

Which factors primarily influence a person's vulnerability to SEA?

The research demonstrated that demographic factors like age and income are primary predictors of vulnerability, but behavioral traits such as curiosity, trust in technology, and social media usage patterns also significantly influence susceptibility. The importance of trust in apps and frequency of social media usage underlines the role of user behaviors in creating opportunities for social engineers to exploit.

Which MLMs are suitable for predicting social engineering attack vulnerability, and under what circumstances are they best suited for doing so?

In answering this question, the research compared multiple machine learning models, including Random Forest (RF), Logistic Regression (LR), and Decision Trees (DT). The Random Forest model, with hyperparameter tuning, emerged as the most effective model, achieving an accuracy of 99.67%, far surpassing the other models. This demonstrates that RF is highly suitable for predicting vulnerability to SEAs, especially in contexts where diverse demographic and behavioral data are available. The model's performance was further enhanced through the use of data imputation techniques, ensuring reliable predictions even in the presence of missing data.

Research Objectives Achieved:

The general objective of the study was to develop a model for predicting individual vulnerability to social engineering attacks and propose the most suitable machine learning model. This objective was fully achieved through the successful experimentation with various MLMs, identifying Random Forest as the best performer. The research further advanced the understanding of which demographic, behavioral, and personality traits are critical in determining SEA vulnerability, fulfilling the first two specific objectives.

The predictive model developed as part of the study is capable of identifying individuals who are most susceptible to SEAs based on factors such as age, income, curiosity, and trust in apps. This directly addressed the third specific objective. The final objective, examining which MLMs are most suitable, was thoroughly explored, with RF being identified as the optimal choice under the tested conditions.

In conclusion, this research provides a robust model for predicting individual vulnerability to social engineering attacks, with Random Forest proving to be the most suitable machine learning model for this task. The study successfully answered the research questions by identifying key factors influencing vulnerability and establishing a predictive framework that can be used for targeted security awareness and interventions. As a result, the research offers a meaningful contribution to both the academic understanding of social engineering vulnerabilities and practical efforts to mitigate them.

7. Future Work

This research has successfully identified key factors that contribute to individual vulnerability to social engineering attacks (SEAs), using machine learning models to analyze demographics, personality traits, and security awareness. Notably, age and income were found to be the most influential predictors, providing a foundation for developing targeted prevention strategies to mitigate these risks.

For future work, several opportunities arise to enhance the model and its practical applications. First, creating a user-friendly interface would enable organizations to easily assess the vulnerability of employees or customers, facilitating more tailored security awareness training programs. Second, refining the model by incorporating additional data sources, exploring new machine learning algorithms, and testing its generalizability across different populations will help improve its accuracy and versatility. Third, developing a dynamic risk assessment system that monitors real-time online behavior could allow for continuous vulnerability scoring, triggering timely interventions when an individual's risk increases.

Lastly, an exciting direction for future research involves expanding the model's capabilities to detect and respond to ongoing social engineering attacks in real-time. This would enhance the system's defensive capabilities, enabling organizations to not only predict vulnerabilities but also actively defend against evolving SEAs. Pursuing these avenues will further advance this research into a powerful, actionable tool that can be used by both individuals and organizations to proactively combat the growing threat of social engineering attacks.

References

[1] S. Morgan, "Cybercrime To Cost The World 8 Trillion Annually In 2023," Cybercrime Magazine, Oct. 13, 2022. https://cybersecurityventures.com/cybercrime-to-cost-the-world-8- trillion-annually-in-2023/

[2] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic Minority Over-sampling Technique," Journal of Artificial Intelligence Research, vol. 16, no. 16, pp. 321–357, Jun. 2002, doi: https://doi.org/10.1613/jair.953.

[3] Wikipedia Contributors, "Cyberattack," Wikipedia, Sep. 24, 2019. https://en.wikipedia.org/wiki/Cyberattack

[4] JPMorgan, "How Criminals Use Social Engineering to Target Your Company."

[5] S. Albladi and G. Weir, "Predicting individuals' vulnerability to social engineering attacks," International Journal of Advanced Computer Science and Applications, vol. 10, no. 2, pp. 144-152, 2019. [Online]. Available: https://thesai.org/Downloads/Volume10No2/Paper_19-

[6] Predicting_Individuals_Vulnerability_to_Social_Engineering_Attacks.pdf

[7] F. Huseynov and B. Ozdenizci Kose, "Using machine learning algorithms to

predict individuals' tendency to be victim of social engineering attacks," Information Development, p. 026666692211163, Aug. 2022, doi: https://doi.org/10.1177/02666669221116336.

[8] R. Yang, K. Zheng, B. Wu, D. Li, Z. Wang, and X. Wang, "Predicting User Susceptibility to Phishing Based on Multidimensional Features," Computational Intelligence and Neuroscience, vol. 2022, pp. 1–11, Jan. 2022, doi: https://doi.org/10.1155/2022/7058972.

[9] P. Offermann, O. Levina, M. Schönherr, and U. Bub, "Outline of a design science research process," Proceedings of the 4th International Conference on Design Science Research in Information Systems and Technology - DESRIST '09, 2009, doi: https://doi.org/10.1145/1555619.1555629.

[10] M. Security, "Are Social Engineering Attacks on the

[11] Rise?," www.mitnicksecurity.com. https://www.mitnicksecurity.com/blog/are-social- engineering-attacks-on-the-rise

[12] G. James, D. Witten, T. Hastie, and R. Tibshirani, An Introduction to Statistical Learning. New York, NY: Springer New York, 2013. doi: https://doi.org/10.1007/978-1- 4614-7138-7.