IS 603 Final Report

CYBERATTACKS ON TEENAGERS

Submitted To

Dr. Faisal Quader

Submitted By:

Pranahith Babu Yarra (IW23456)

Sai Rajesh Rapelli (HD49179)

Ameenur Rahman Khan (DJ46492)

Sai Vikas Amaraneni (FS93533)

Venkat Kowshik Maram (MO76733)

M.S. Information Systems

Department of Information Systems

University of Maryland, Baltimore County

Table of Contents

1. Abstract

The increasing use of technology and social media use by teenagers has made them more vulnerable to cyberattacks. Teenagers are exposed to a variety of cyberattacks, including identity theft, phishing, hacking, and cyberbullying. These assaults may result in severe consequences such as mental distress, loss of money, and damage to academic and personal reputations. Teenagers should be aware of the warning signs of cyberbullying and report any incidents to an adult they can trust or an appropriate authority person. Additionally, they must beware when disclosing private information online and should only do so on reliable, safe websites.

As part of this project, we will use Kaggle datasets to identify the various sorts of cyberattacks that teenagers are vulnerable to and predict their results using machine learning techniques. The study's significance originates from its potential to bring insight into the patterns and trends of cyberattacks against minors, as well as to develop effective techniques for preventing and mitigating such attacks. The findings of the study will also help lawmakers design legislation and policies to protect teenagers from cyberattacks and we will discuss the various types of cyberattacks that teenagers may encounter and their potential impact. Using the various machine learning algorithms, we will convey the best algorithm to detect the type of cyber-attack a person is facing based on given information.

2. Introduction

Teenagers are particularly at risk for cyberattacks as a result of their extensive use of technology and social media. In fact, over 70% of teenagers report experiencing cyberbullying at some or other time in their life, according to recent research. In addition, it is found that one in three youth had fallen victim to identity theft or fraud online, and one in four have been scammed. Hackers can take advantage of software problems to take personal information or mislead. To avoid cyberattacks on minors, it is critical to educate them on safe internet behaviors such as choosing strong passwords, not disclosing personal information, and being cautious of suspicious emails or texts. It is also necessary to maintain software and devices up to date with the latest security updates since obsolete software might be exposed to assaults. For this problem, various approaches have been tried, including educational campaigns, resources aimed at teaching teenagers about safe online practices, parental controls, monitoring software to limit access to specific websites and activities, and cybersecurity training for teachers and school administrators.

Numerous strategies have been attempted to address the issue of cyber safety among teenagers, including educational campaigns, resources aimed at teaching teenagers about safe online habits, and the implementation of parental controls. Additionally, monitoring software can be used to restrict access to specific websites and activities, and cybersecurity training for

teachers and school administrators can help create a safer online environment in educational settings. Teenagers must thus be aware of the possible dangers associated with using technology and social media and take precautions to safeguard themselves against cyberattacks. This includes being cautious about sharing personal information online, using strong passwords, and being aware of the signs of cyberbullying and other forms of cybercrime.

3. Objectives

The primary objectives of discussing cyberattacks on teenagers are to enhance internet safety, reduce the incidence of cyberattacks, and mitigate the impact of cybercrime on adolescents. By addressing these objectives, we aim to create a safer online environment and empower teenagers to navigate the digital world with confidence and security.

3.1 Increase awareness: The first objective is to raise awareness about the potential dangers associated with technology and social media usage among teenagers. By discussing various types of cyberattacks that teenagers may encounter, we can help them understand the risks involved and encourage them to adopt preventive measures against cyber threats. Increased awareness will enable teenagers to make informed decisions about their online activities and recognize potential threats before they cause harm.

3.2 Educate teenagers: Equipping teenagers with the necessary knowledge and skills to defend themselves against cyberattacks is essential for their overall safety. This objective focuses on providing comprehensive information about the potential consequences of cyberattacks and offering strategies to maintain online security. By educating teenagers on the best practices for online safety, such as using strong passwords, being cautious when sharing personal information, and recognizing signs of cyberbullying, we empower them to take control of their digital lives and minimize their vulnerability to cyber threats.

3.3 Develop and implement effective prevention strategies: Another crucial objective is to develop and implement effective prevention strategies to reduce the incidence of cyberattacks on teenagers. This involves researching and understanding the underlying causes of cybercrime, identifying potential vulnerabilities in online platforms and systems, and creating targeted interventions to address these weaknesses. Collaborating with stakeholders, such as parents, educators, and policymakers, will be vital in developing comprehensive prevention strategies that address the specific needs of teenagers.

3.4 Support and empower victims of cyberattacks: Finally, it is essential to provide support and resources to teenagers who have been affected by cyberattacks. This objective includes developing intervention programs and counseling services to help victims of cybercrime cope with the emotional and psychological consequences of their experiences. Additionally,

empowering teenagers to report incidents of cybercrime and seek help when needed will contribute to creating a supportive and secure online environment for all users.

By addressing these objectives, we aim to create a more secure and supportive online environment for teenagers, ultimately reducing the incidence of cyberattacks and mitigating their lives.

4. Related Work

This section reviews existing literature on cybercrime awareness among teenagers, exploring the factors influencing their vulnerability to cybercrimes, existing awareness programs, and the effectiveness of prevention strategies.

4.1 Factors Influencing Teenager's Vulnerability to Cybercrimes:

Teenagers' vulnerability to cybercrimes is influenced by several factors. Firstly, their high digital engagement (Hinduja & Patchin, 2018) exposes them to various cyber threats. Their inclination to share personal information online (Livingstone & Haddon, 2014) and lack of knowledge about the potential consequences (O'Keeffe & Clarke-Pearson, 2011) further exacerbate their vulnerability.

4.2 Existing Cybercrime Awareness Programs:

Several awareness programs target teenagers to enhance their understanding of cybercrimes and online safety. Programs such as the NetSmartz Workshop by the National Center for Missing & Exploited Children (2017) and CyberSmart! Education Program (iKeepSafe, 2017) focuses on increasing awareness of cyber threats, promoting safe online behavior, and fostering digital citizenship.

4.3 Effectiveness of Prevention Strategies:

Research on the effectiveness of prevention strategies has produced mixed results. Some studies suggest that awareness programs positively impact teenagers' knowledge of cyber threats and help develop safe online behavior (Navarro & Jasinski, 2012; Patchin & Hinduja, 2010). Others argue that the effectiveness of these programs is limited by their short-term nature, lack of reinforcement, and inadequate focus on specific cybercrime types (Wright, 2015).

4.4 Gaps in the Literature:

While existing literature has contributed to our understanding of cybercrime awareness among teenagers, several gaps remain. First, there is a need for more empirical research to assess the effectiveness of prevention strategies and identify the best practices for increasing cybercrime awareness among teenagers. Second, research should explore the role of parents, educators, and

policymakers in promoting cybercrime awareness and reducing teenagers' vulnerability to online threats.

5. Implementation

Data Preparation:

The dataset retrieved from Kaggle is first cleaned by removing all the unwanted columns and duplicate rows. Later, the null values are identified and then the respective rows are removed. The columns are also renamed to be short and precise.

```
In [1]: import numpy as np
        import pandas as pd
        from sklearn import preprocessing
        import matplotlib.pyplot as plt
```

```
In [2]: df = pd.read_csv('dataset.csv')
        df.head()
```

Out[2]:

	Gender	Age	Received suspicious email?	Encountered phishing attempt?	Aware of cybercrime?	Experienced cybercrime?	Type of cybercrime
0	Male	21	Yes	Yes	Yes	Yes	PhishingEmails
1	Male	20	Yes	Yes	Yes	Yes	PhishingEmails
2	Male	22	Yes	Yes	Yes	Yes	IdentityTheft
3	Male	22	Yes	Yes	Yes	Yes	PhishingEmails
4	Male	25	No	No	Yes	No	PhishingEmails

Data Exploration:

We explored the data by using various numpy and pandas functions to understand the data. Using the shape attribute we found out there are 500 rows and 7 columns.

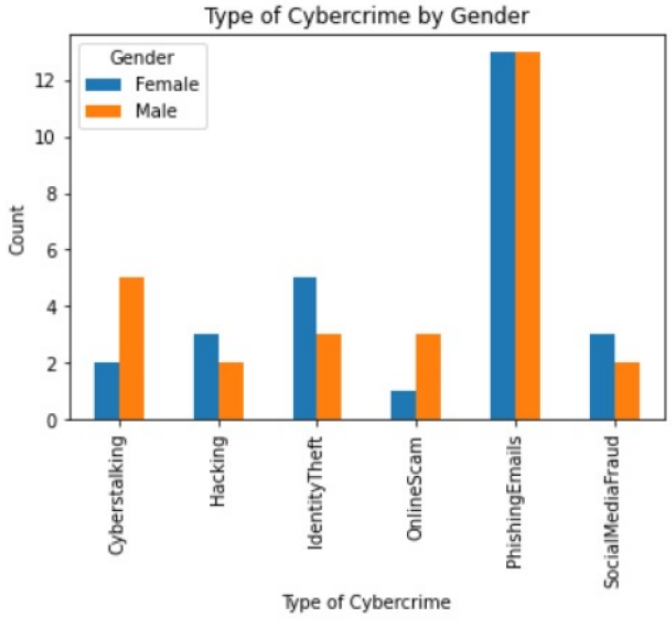
```
In [3]: df.shape
Out[3]: (500, 7)
```

Data Visualization:

For a quick capture of how the dataset is, the below graph depicts the information about different types of cybercrimes found vulnerable by male and female. From the below graph it can be understood that both males and females are mostly vulnerable towards phishing emails compared to other cyberattacks. Besides, the males are then more likely to be vulnerable by cyber stalking, whereas females are more likely to be vulnerable by identity theft.

```
df_counts = df.groupby(['Gender', 'Type of cybercrime']).size().reset_index(name='Counts')
df_pivot = df_counts.pivot(index='Type of cybercrime', columns='Gender', values='Counts')

ax = df_pivot.plot(kind='bar')
ax.set_xlabel('Type of Cybercrime')
ax.set_ylabel('Count')
ax.set_title('Type of Cybercrime by Gender')
plt.show()
```



Encoding the categorical data:

Among the 7 columns available, only the age column is numerical and the rest are non-numerical columns. For training the dataset further, we require all the columns to be numerical. Hence, we conducted label encoding where the columns with Yes/No values are changed to 1/0 respectively and the type of cybercrime columns are of 5 unique values and they are randomly assigned from 0 to 4.

In [5]:
```
label_encoder = preprocessing.LabelEncoder()

df['Gender'] = label_encoder.fit_transform(df['Gender'])
df['Received suspicious email?'] = label_encoder.fit_transform(df['Received suspicious email?'])
df['Encountered phishing attempt?'] = label_encoder.fit_transform(df['Encountered phishing attempt?'])
df['Aware of cybercrime?'] = label_encoder.fit_transform(df['Aware of cybercrime?'])
df['Experienced cybercrime?'] = label_encoder.fit_transform(df['Experienced cybercrime?'])
df['Type of cybercrime'] = label_encoder.fit_transform(df['Experienced cybercrime?'])
df.head()
```

Out[5]:

	Gender	Age	Received suspicious email?	Encountered phishing attempt?	Aware of cybercrime?	Experienced cybercrime?	Type of cybercrime
0	1	21	1	1	1	1	1
1	1	20	1	1	1	1	1
2	1	22	1	1	1	1	1
3	1	22	1	1	1	1	1
4	1	25	0	0	1	0	0

Split the dataset:

As we have our dataset completely structured, so now we started splitting the dataset. It is divided into 80% training and 20% testing, where initially the algorithms will be trained on 80% of the dataset and they find a pattern. Later this pattern is tested on the rest 20% of the dataset.

Testing:

As we have split the dataset, we now use them to calculate the accuracy with various algorithms. Based on our problem statement our ultimate goal is to predict the type of cybercrime based on the input values, hence it is supervised learning as there are labels associated here. Out of which it is a classification type as we are trying to classify the type of cybercrime. Now we will look into different classification algorithms.

1. Logistic regression:

It is a classification algorithm used to predict the type of category with input values. We have tweaked the model by passing various inputs, so max_iter=10000 will iterate the dataset many times to give the final score.

```
from sklearn.linear_model import LogisticRegression
log = LogisticRegression(max_iter=10000)
log_fit = log.fit(x_train,y_train)
log_pred = log.predict(x_test)
log_score = log.score(x_train,y_train)
print("Logistic Regression Accuracy: %f"%log_score)
```

```
Logistic Regression Accuracy: 0.772727
```

2. Random Forest Classifier:

It is another type of classification algorithm. Each decision tree in a random forest is constructed using a random subset of the input features and the training data. This procedure aids in lowering overfitting and enhancing model correctness. Following the construction of each individual tree, the random forest aggregates the predictions made by the trees by averaging the outputs for regression issues or selecting the majority vote for classification problems.

```
from sklearn.ensemble import RandomForestClassifier
random_forest = RandomForestClassifier()
random_fit = random_forest.fit(x_train,y_train)
random_pred = random_forest.predict(x_test)
random_score = random_forest.score(x_train,y_train)
print("Random Forest Classifier Accuracy: %f"%random_score)
```

```
Random Forest Classifier Accuracy: 0.977273
```

3. Decision Tree Classifier:

A decision tree classifier is an approach where we can even manually find the pattern to conclude the type of cybercrime it is by splitting all the categorical columns into an organized way such as from the parent node (root) to the child node (leaf).

```
from sklearn.tree import DecisionTreeClassifier
tree = DecisionTreeClassifier(max_depth=100)
tree_fit = tree.fit(x_train,y_train)
tree_pred = tree.predict(x_test)
tree_score = tree.score(x_train,y_train)
print("Decision Tree Classifier Accuracy: %f"%tree_score)
```

```
Decision Tree Classifier Accuracy: 0.977273
```

4. Support Vector Machine:

It is used for classification, regression, and outlier detection tasks. The SVM classifier is used while segregating the two classes(hyper-plane/line). They mainly separate data until a hyperplane with a high minimum distance is found and it is used to classify two or more data types.

```
from sklearn import svm
svm_mod = svm.SVC()
svm_mod_fit = svm_mod.fit(x_train,y_train)
svm_mod_pred = svm_mod.predict(x_test)
svm_score = svm_mod.score(x_train,y_train)
print("Support Vector Machine Accuracy: %f"%svm_score)
```

```
Support Vector Machine Accuracy: 0.704545
```

5. K-Nearest Neighbors:

This algorithm works by locating the mentioned K nearest data points in the training dataset to its test data, and then based on the kind of the problem it will predict the output. For a classification problem, it will look for the majority of instances, whereas for the regression problem, it will take the average. So for our scenario, it considered different similar input data and then considered the majority occurrences of cybercrime type.

```
from sklearn.neighbors import KNeighborsClassifier
neigh = KNeighborsClassifier(n_neighbors=3)
neigh_fit = neigh.fit(x_train,y_train)
neigh_pred = neigh.predict(x_test)
neigh_score = neigh.score(x_train, y_train)
print("K Nearest Neighbor Classifier Accuracy: %f"%neigh_score)
```

```
K Nearest Neighbor Classifier Accuracy: 0.795455
```

6. Experimental Results

From the above classification models, we can observe that Random Forest Classifier and Decision Tree Classifier have the best accuracy compared to other algorithms. Based on the runtime of both the algorithms Random Forest Classifier is more efficient compared to the Decision Tree Classifier, hence we choose the Random Forest Classifier for predicting the type of cybercrime for a given input.

Below is the snapshot indicating how the trained model is performing based on the given input. As the best accuracy is found in the random forest classifier algorithm, hence the prediction is done alongside this model. Where for a male of 22 years, who experienced cybercrime is predicted as he is vulnerable to phishing emails.

```
rfc = RandomForestClassifier(n_estimators=100)
rfc.fit(x_train, y_train)

testinput = [[1,22,1,1,1,1]]
testoutput = rfc.predict(testinput)

testoutput
```

```
array([0], dtype=int64)
```

7. Future work

To further address the issue of cyberattacks targeting teenagers, several avenues of future work should be considered, such as;

1. Developing innovative technologies and tools: To identify and stop cyberattacks targeting teenagers, this includes developing AI-based threat detection systems that make use of machine learning and natural language processing methods. Teenagers' internet accounts may be made more secure by developing sophisticated password encryption techniques and supporting multi-factor authentication.

2. Implementing comprehensive educational programs and awareness campaigns: In order to create an age-appropriate cybersecurity curriculum that includes experiential learning and real-world examples, policymakers, educators, and parents should work together. Teenagers' awareness of possible online hazards should be increased through these programs, which should also encourage safe online conduct and digital resilience.

3. Longitudinal research: Studies that evaluate the long-term efficacy of preventative techniques and educational initiatives, as well as their effects on teens' online behavior and vulnerability to cybercrimes, should be the main focus of future studies.

8. Conclusion

The comparative analysis of various classification algorithms revealed that the Random Forest Classifier and Decision Tree Classifier are the most suitable choices for predicting the type of cybercrime based on input data. The Random Forest Classifier's superior runtime efficiency and accuracy make it the optimal choice for our dataset. If efficiency is not a concern, then to construct a decision tree manually, input labels can be classified sequentially within the tree model. To generate predictions, the input data will be fitted into the selected model, allowing the algorithm to output predictions based on the given input data. This research contributes to our understanding of cybercrime awareness among teenagers and highlights the need for innovative solutions and comprehensive educational programs to enhance their online safety.

9. References

1. Hinduja, S., & Patchin, J. W. (2018). Connecting adolescent suicide to the severity of bullying and cyberbullying. Journal of School Violence, 17(3), 333-346.

2. Livingstone, S., & Haddon, L. (2014). EU kids online. LSE, London: EU Kids Online.

3. National Center for Missing & Exploited Children (2017). NetSmartz Workshop. Retrieved from http://www.netsmartz.org

4. Navarro, J. N., & Jasinski, J. L. (2012). Going cyber: Using routine activities theory to predict cyberbullying experiences. Sociological Spectrum, 32(1), 81-94.

5. O'Keeffe, G. S., & Clarke-Pearson, K. (2011). The impact of social media on children, adolescents, and families. Pediatrics, 127(4), 800-804.

6. Patchin, J. W., & Hinduja, S. (2010). Cyberbullying and self-esteem. Journal of School Health, 80(12), 614-621.

7. iKeepSafe (2017). CyberSmart! Education Program. Retrieved from https://ikeepsafe.org

8. Wright, M. F. (2015). Adolescents' cyber aggression perpetration and cyber victimization: The longitudinal associations with school functioning. Social Psychology of Education, 18(4), 653-666.