Pranahith Babu Yarra
Sai Vikas Amaraneni
Sai Rajesh Rapelli
Ameenur Rahman Khan
Venkat Kowshik Maram

CYBERATTACKS ON TEENAGERS

Advisor: Dr. Faisal Quader

ABSTRACT

The research being investigated will look into cyberattacks on teenagers, such as online abuse, cyberbullying, and identity theft. The study will use Kaggle data sets 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.

PROBLEM STATEMENT

In today's digital world, cyberattacks on minors are an increasing problem. Teenagers are more at risk of phishing emails, cyberbullying, online harassment, identity theft, and other types of cybercrime as social media and technology usage increase. As a result of these attacks teenagers may suffer from mental suffering, bad reputations, and financial loss. Therefore, it is critical to understand the type of cyberattacks that target kids, their effects, and how to stop and prevent them.

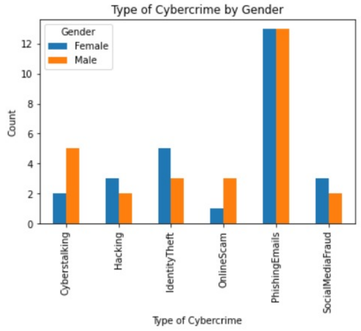
IMPLEMENTATION

Step-1 Data Cleaning :
The data obtained is cleaned by removing duplicate rows and unwanted columns. Null value rows are removed.

Step-2 Data Exploration :
Explored data with various numpy and pandas functions to learn more about dataset.

```
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()
```



Step-3 Label Encoding :
All the non-numerical columns are then label encoded to 0 or 1.

Step-4 Split the Dataset :
All the data is here divided into training and testing data, where trained data is 80% and test data is 20%.

Step-5 Testing :
Based on the problem statement and dataset we are working on it is a supervised learning as there are inputs with different labels and a target label as well. As we will be predicting a class label, hence we shall proceed with classification algorithms.

Logistic Regression :

```
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
```

Random Forest Classifier :

```
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
```

Decision Tree Classifier :

```
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
```

Support Vector Machine :

```
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
```

CONCLUSION

Based on the various accuracy achieved from different classification algorithms, we can conclude that random forest classifier and decision tree classifier are the one's to be opted for predicting the type of cybercrime for an input data.

For our dataset we can achieve the decision tree manually as well by classifying the input labels one after the other in the tree model. For predicting we will be taking the input data and fit in to the model and predict the output for the given input data.

FUTURE WORK

There are several areas of future work that can help to address the issue of cyberattacks on teenagers.

1. Creating new technologies and tools, such AI-based threat detection systems and more password encryption techniques, that can identify and stop cyberattacks on teens.

2. Teenagers need to be educated and made more aware of the dangerous cyberattacks and how to be safe.