Phisher

Konam Akhil Vamshi, Donthireddy Manoj Reddy
2020513, 2020375

Introduction

Phisher, a browser extension that leverages machine learning and human feedback to identify and block phishing attempts. Phisher analyzes website content, landing page URLs, linked resources, and other features to detect signs of phishing like suspicious links, spoofed domains, and known phishing tactics. It then aggregates reports from users to determine the legitimacy of new websites and links. By combining AI and crowdsourced insights, Phisher achieves a high detection accuracy while maintaining a low false positive rate.

Phisher functions as a defensive shield, warning users when they encounter potential phishing attempts in their web browser or email client. It works to prevent phishing at the point of user interaction, eliminating the need for users to manually report phishing messages after being targeted. The ultimate goal of Phisher is to shut down phishing campaigns at scale and curb the spread of malware, fraud, and account compromise.

Features:

The CheckLink feature allows users to check whether a link is malicious. When a user clicks on CheckLink, a pop-up screen appears on the current page, where they can paste the link they want to check. The algorithm, based on a RoBERTa-based transformer model, then analyzes the link to determine whether it is malicious or not. The algorithm classifies the link as either Label 0-Non-Malicious or Label 1-Malicious.

The SummarizeEmail feature allows users to identify malicious activity in recent emails. The machine learning algorithm reads the email body and determines the percentage of recent emails that are malicious.

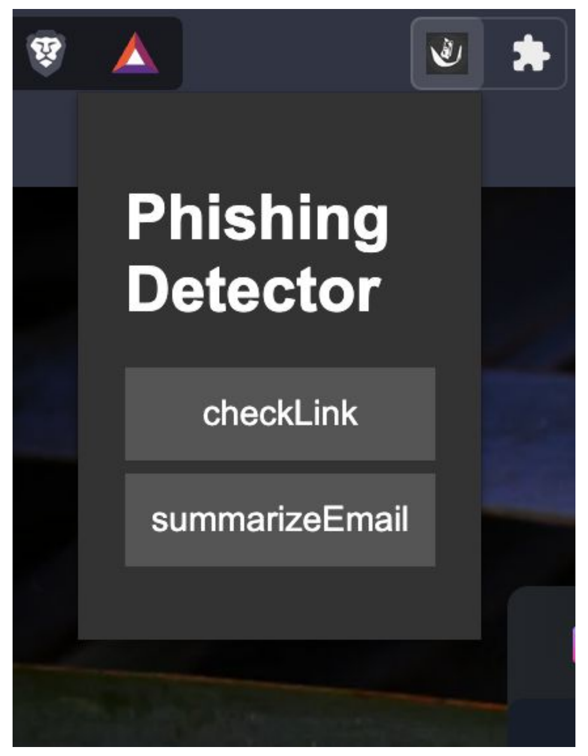


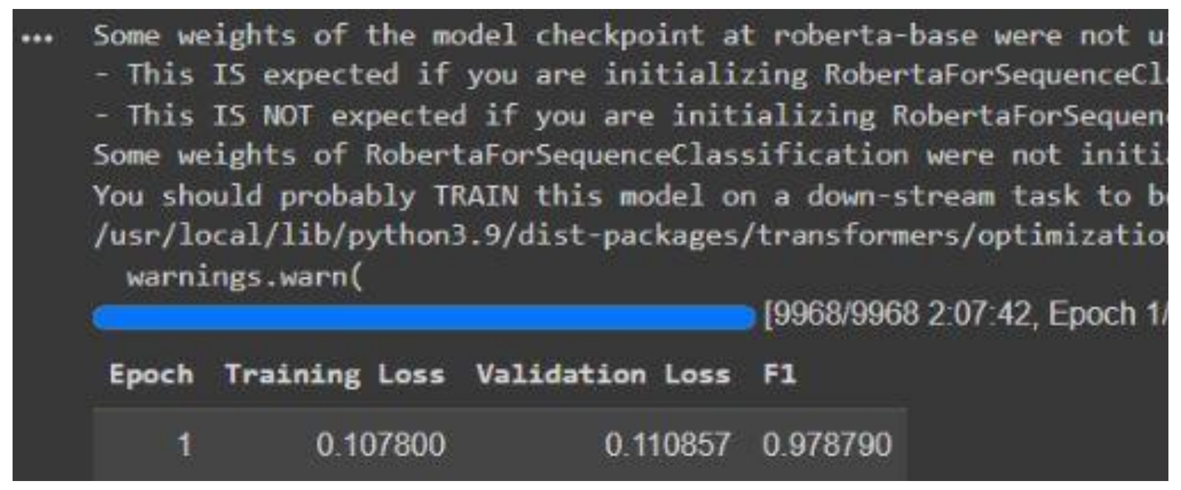
Fig 1 : Here is the Glimpse of the Extension

Comparison:

Compared to other phishing detection tools, Phisher stands out for its integration of user feedback and its use of machine learning to analyze messages and websites. Other phishing detection tools may rely on rules-based algorithms or blacklists to detect phishing attempts, but these methods are less effective as phishing tactics evolve over time. Phisher's machine learning approach allows it to adapt to new phishing techniques and identify previously unknown threats.

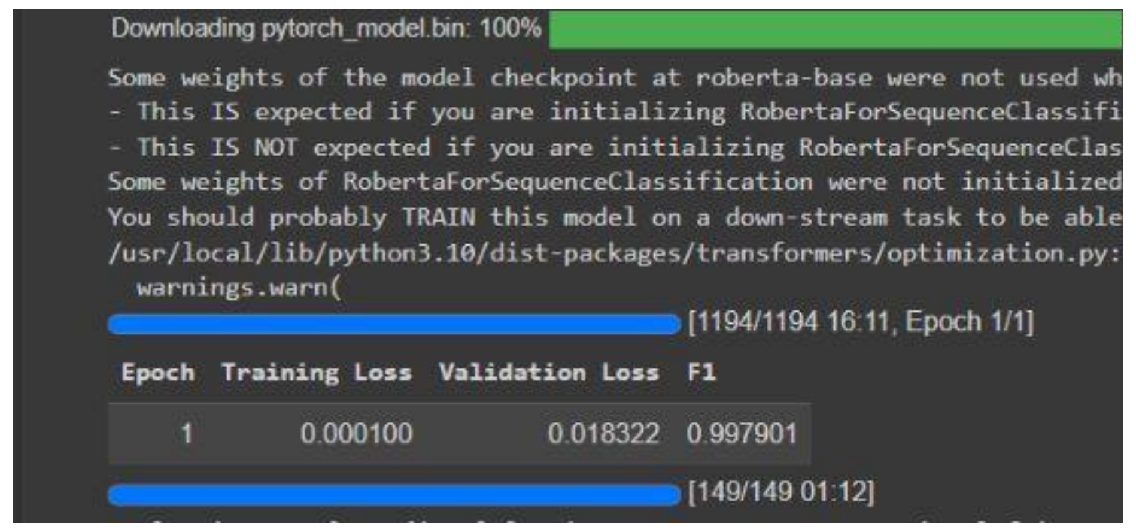
Algorithm:

Phisher's algorithm is based on a RoBERTa-based transformer model trained on 100,000 links for training. The model extracts features from the link and classifies it as either malicious or non-malicious. The algorithm also reads the email body to determine the percentage of recent emails that are malicious.



```
... Some weights of the model checkpoint at roberta-base were not u
    - This IS expected if you are initializing RobertaForSequenceCl
    - This IS NOT expected if you are initializing RobertaForSequen
    Some weights of RobertaForSequenceClassification were not initi
    You should probably TRAIN this model on a down-stream task to b
    /usr/local/lib/python3.9/dist-packages/transformers/optimizatio
      warnings.warn(
                                               [9968/9968 2:07:42, Epoch 1/
     Epoch  Training Loss  Validation Loss  F1
         1       0.107800         0.110857  0.978790
```

Fig 2 : ML Model F1 Score which checks Malicious Links



```
Downloading pytorch_model.bin: 100%
Some weights of the model checkpoint at roberta-base were not used wh
- This IS expected if you are initializing RobertaForSequenceClassifi
- This IS NOT expected if you are initializing RobertaForSequenceClas
Some weights of RobertaForSequenceClassification were not initialized
You should probably TRAIN this model on a down-stream task to be able
/usr/local/lib/python3.10/dist-packages/transformers/optimization.py:
  warnings.warn(
                                         [1194/1194 16:11, Epoch 1/1]
 Epoch  Training Loss  Validation Loss  F1
     1       0.000100         0.018322  0.997901
                                         [149/149 01:12]
```

Fig 3 : ML Model F1 Score which checks Malicious Mail Body

Other Phishing Detections:

Other phishing detection tools include anti-phishing software, web filters, and spam filters. These tools typically use a combination of rule-based algorithms and blacklists to identify phishing attempts. However, these methods are less effective than Phisher's machine learning approach, as they may not detect previously unknown threats.

Future Work:

In the future, Phisher could expand its features to include real-time monitoring of user activity and machine learning-based analysis of user behavior to identify potential phishing attempts. Additionally, Phisher could integrate with other cybersecurity tools to provide a more comprehensive protection suite for users.

Implementation

Phisher is a browser extension that can be downloaded from the extension toolbar. It is compatible with most popular web browsers, including Google Chrome, Mozilla Firefox, and Microsoft Edge.

Conclusion:

Phisher is a powerful tool for identifying and preventing phishing attacks. Its use of machine learning and user feedback allows it to adapt to new threats and provide accurate results. The CheckLink and SummarizeEmail features make it easy for users to identify malicious links and emails. With continued development, Phisher could become an essential tool for protecting users from phishing scams.

Working Model