Digisurksha Parhari Foundation Final Internship Submission 2025-26 Presentation

Topic Name : **AI in Social Engineering and Phishing Campaigns**

College Name : Mulund College of Commerce

Team Work By :

- Prathamesh Shinde SYCS 248360
- Ajinkya Mohite SYCS 248335
- Aditya Gaikwad SYCS 248311
- Rutuja Kurane SYCS 248429
- Sneha Kamble SYCS 248425

❑ Introduction

- **today's digital world, cyberattacks are no longer just about breaking firewalls or exploiting software bugs—they are increasingly targeting human behavior.**
- **Social engineering has become one of the most effective techniques used by attackers. It involves manipulating people into revealing sensitive information or unknowingly performing actions that compromise security.**
- **One of the most common and dangerous forms of social engineering is phishing, where attackers send fake emails, messages, or calls to trick users into clicking malicious links or sharing confidential data.**
- **With the rise of Artificial Intelligence (AI) and especially powerful language models like GPT, phishing attacks have become more dangerous and harder to detect.**
- **Attackers can now use AI to create phishing emails that sound more realistic, are better tailored to the victim, and can even mimic a specific writing style or tone of voice.**
- **Not just emails—AI can now generate deepfake audio or video, making voice phishing (vishing) and video-based scams even more convincing.**
- **This research explores how AI is being used to enhance social engineering attacks and what can be done to defend against them.**
- **As part of this study, a tool is developed to simulate AI-powered phishing content and test how well detection systems perform against these new threats.**
- **The goal is to understand the growing role of AI in social engineering, evaluate its impact on cybersecurity, and propose possible solutions to counter such attacks responsibly.**

❑ What is the need of these topic And which kind of problem it sovles!

As Artificial Intelligence (AI) technologies continue to evolve, they are increasingly being exploited by cyber attackers to carry out advanced social engineering and phishing schemes. These AI-assisted threats are more sophisticated, personalized, and difficult to identify compared to traditional cyber attacks. This trend presents serious challenges for cybersecurity, emphasizing the need for a deeper understanding of how AI is misused and the development of effective strategies to combat such threats.

❑ How Can We Tackle the Problem?

- **Tackling phishing especially AI-powered phishing 321requires a multi-layered defense strategy combining technology, policy, and user awareness. Here's a breakdown of effective approaches you can include in your research:**

How to Tackle the Problem of Phishing

1. **AI-Powered Detection Systems Use machine learning models trained on phishing datasets to detect suspicious content. Natural Language Processing (NLP) can analyze email tone, structure, and intent to flag AI-generated or malicious messages. Tools like BERT, GPT detectors, or custom classifiers can identify unnatural patterns in phishing emails.**
2. **2. Email Authentication Protocols Implement protocols like SPF (Sender Policy Framework), DKIM (DomainKeys Identified Mail), and DMARC to verify email senders and reduce spoofing.**

3. User Awareness & Training Conduct phishing simulation campaigns to train employees to recognize suspicious emails. Teach users to check for signs like: Unusual sender addresses Poor grammar (less common in AI-generated emails now)Urgent tone or unusual requests Unexpected links or attachments

4. Multi-Factor Authentication (MFA)Even if credentials are stolen, MFA adds a second layer of protection that blocks unauthorized access.

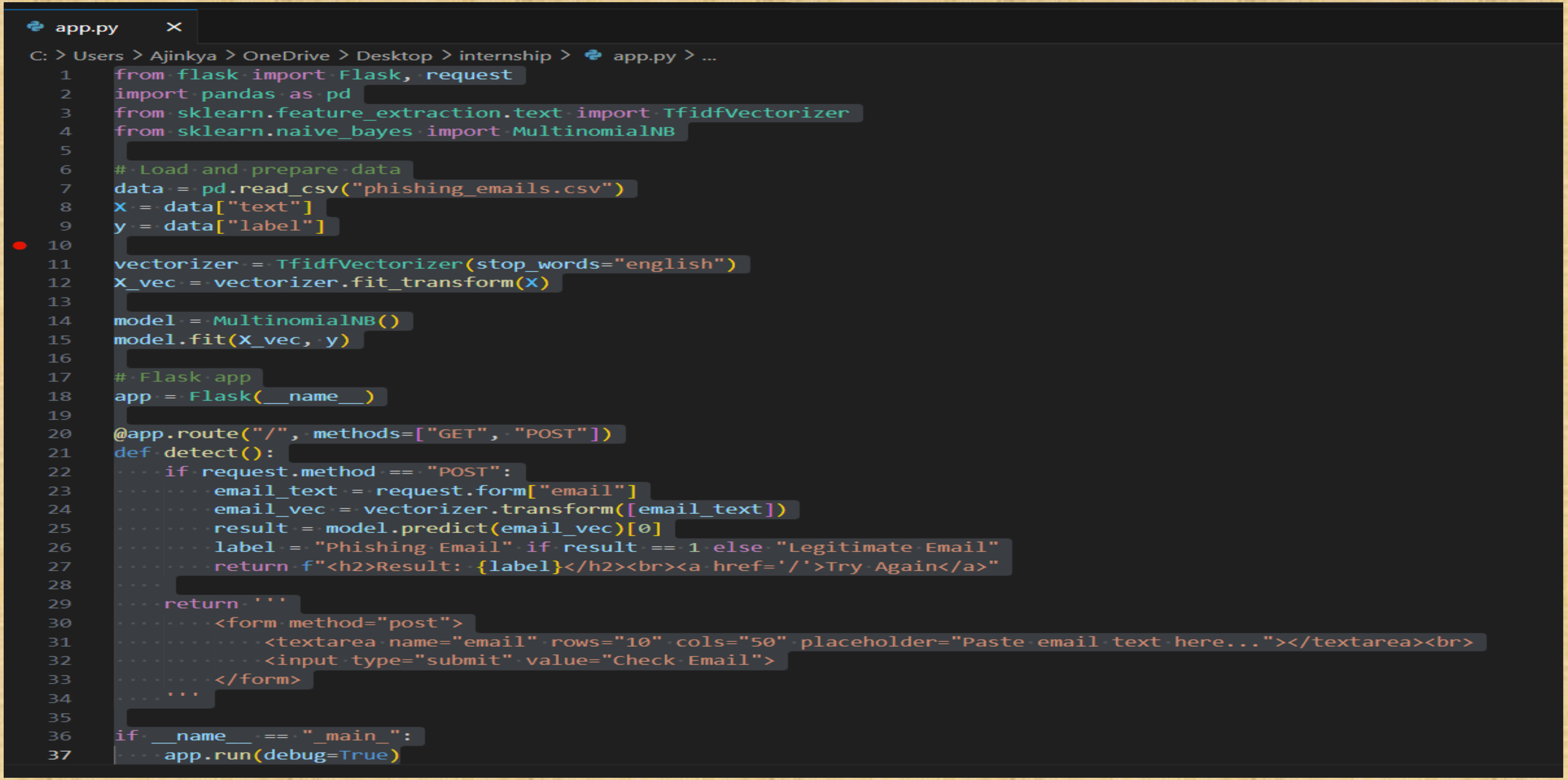
5. Browser & Endpoint Protection Use security tools that warn users when visiting fake or malicious websites (URL filtering, anti-phishing plugins).Keep browsers, email clients, and antivirus software updated to defend against known phishing tactics.

6. Real-Time Threat Intelligence Integrate threat feeds to stay updated on newly reported phishing sites and campaigns. Use AI to correlate threats and flag phishing domains in real time.

❑ What we do to solve these problem ! How?

- **First we have to download python version 3.13.3 (3.+) than install. Open command Prompt and type "pip install pandas scikit-learn nltk flask" and run, than our requried module will download. Than in VS Code I created two files such as app.py and phishing_emails.csv I provided that source code. And for phishing source code content via "https://www.kaggle.com/datasets/naserabdullahalam/phishing-email-dataset" this link. Than run the app.py file, So in the terminal we got a url or IP address as "http://127.0.0.1:5000". Open it and check the messages by pasting here weither it is Legitimate Email or Phishing Email.**

➢ Python Source Code



app.py ✕
C: > Users > Ajinkya > OneDrive > Desktop > internship > app.py > ...

```python
from flask import Flask, request
import pandas as pd
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.naive_bayes import MultinomialNB

# Load and prepare data
data = pd.read_csv("phishing_emails.csv")
X = data["text"]
y = data["label"]

vectorizer = TfidfVectorizer(stop_words="english")
X_vec = vectorizer.fit_transform(X)

model = MultinomialNB()
model.fit(X_vec, y)

# Flask app
app = Flask(__name__)

@app.route("/", methods=["GET", "POST"])
def detect():
    if request.method == "POST":
        email_text = request.form["email"]
        email_vec = vectorizer.transform([email_text])
        result = model.predict(email_vec)[0]
        label = "Phishing Email" if result == 1 else "Legitimate Email"
        return f"<h2>Result: {label}</h2><br><a href='/'>Try Again</a>"
    
    return '''
        <form method="post">
            <textarea name="email" rows="10" cols="50" placeholder="Paste email text here..."></textarea><br>
            <input type="submit" value="Check Email">
        </form>
    '''

if __name__ == "_main_":
    app.run(debug=True)
```

❑ Real-World Use Cases

- **Deepfake Voice & Video Scams**
- **Use Case: Voice cloning for CEO fraud or business email compromise (BEC)**
- **Description: AI tools can mimic a person’s voice using a small audio sample, enabling fraudsters to impersonate executives or colleagues over the phone or video calls.**
- **Real-World Example: In 2020, criminals used deepfake voice to trick a UK energy firm into transferring €220,000, believing they were speaking with the CEO.**

- **Chatbots for Real-Time Scamming**

Use Case: AI-driven customer support impersonation

Description: Malicious actors deploy AI chatbots on fake websites or via messaging apps to impersonate customer service representatives and harvest user credentials.

Real-World Example: Phishing websites mimicking banks or crypto exchanges have used AI chatbots to answer questions in real time and lure victims into entering sensitive data.

❑ Future Enhancements

- **1. Advanced AI-Powered Detection EnginesDevelop more robust phishing detection models using context-aware AI that can detect not just keywords but the intent and emotional tone behind messages.**
- **2. Integration with Real-Time Communication PlatformsExtend protection beyond email to platforms like Slack, WhatsApp, Microsoft Teams, and LinkedIn, where social engineering is growing rapidly.**
- **3. Deepfake and Voice Phishing Defense SystemsIntegrate deepfake detection algorithms into phone systems to catch AI-generated voice scams (vishing) and impersonation attempts.**