FUNDAÇÃO DE ASSISTÊNCIA E EDUCAÇÃO
CENTRO UNIVERSITÁRIO ESPÍRITO-SANTENSE
CURSO DE GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

NATHAN ALEXANDRE VIDIGAL DE SOUZA

DETECTOR DE PHISHING INTELIGENTE

VITÓRIA
2025

NATHAN ALEXANDRE
VIDIGAL DE SOUZA

DETECTOR DE PHISHING INTELIGENTE

Trabalho acadêmico do Curso de Graduação em Ciência da Computação, apresentado ao Centro Universitário Espírito-santense como parte das exigências da disciplina Projeto Integrador IV, sob orientação do Professor Howard Cruz Roatti.

VITÓRIA
2025

SUMÁRIO

1 INTRODUÇÃO

A segurança digital se estabeleceu como um pilar essencial na sociedade contemporânea, onde a maior parte das transações e comunicações ocorre em plataformas virtuais. Dentro deste ecossistema, o phishing persiste como uma das ameaças mais difundidas e perigosas, sendo responsável pela grande maioria dos ataques de roubo de dados pessoais e corporativos.

Frequentemente, os golpes de phishing evoluem em sofisticação e são desenhados para explorar vulnerabilidades humanas, fazendo com que usuários comuns não consigam identificar mensagens fraudulentas, seja por falta de conhecimento técnico ou pela alta qualidade das imitações. Este cenário ressalta a urgência de ferramentas que não apenas detectem, mas também expliquem o risco de forma acessível.

Nesse contexto, o presente trabalho, intitulado Detector de Phishing Inteligente e desenvolvido na área de Cibersegurança, propõe uma solução inovadora. O projeto consiste no desenvolvimento de uma ferramenta que utiliza modelos avançados de linguagem natural (LLMs) para analisar o conteúdo textual de e-mails ou mensagens. O objetivo central do projeto é duplo: classificar o risco de uma mensagem em tempo real e, crucialmente, fornecer ao usuário uma explicação clara e em linguagem simples sobre os indícios que tornam aquela comunicação suspeita.

Ao unir detecção tecnológica e educação ativa, este projeto visa mitigar perdas e, simultaneamente, aumentar a literacia digital dos usuários, impactando positivamente o cidadão comum e as pequenas empresas ao aumentar sua segurança passiva e capacidade de autoproteção.

O relatório a seguir detalha a justificativa do tema, o escopo da primeira versão funcional (MVP), a metodologia a ser aplicada, o plano de trabalho para o desenvolvimento e as métricas de sucesso do projeto.

2 JUSTIFICATIVA

O phishing continua sendo uma das principais ameaças digitais, responsável por grande parte dos ataques de roubo de dados pessoais e corporativos. Muitas vezes, usuários não conseguem identificar e-mails ou mensagens fraudulentas, seja por falta de conhecimento técnico ou pela sofisticação crescente dos golpes.

Com o uso de modelos de linguagem natural (LLMs), é possível analisar o conteúdo das mensagens e fornecer ao usuário não apenas a detecção da fraude, mas também uma explicação clara e em linguagem simples sobre os motivos que tornam aquela mensagem suspeita. Isso contribui para educar o usuário e aumentar a segurança digital.

3 ESCOPO DO PROJETO (MVP)

Funcionalidades previstas (MVP):

Upload ou inserção de texto de mensagem/e-mail.
• Análise em tempo real com API/LLM.
• Retorno ao usuário com classificação de risco.
• Explicação em linguagem simples sobre os indícios de fraude.
• Interface web responsiva e fácil de usar.

Funcionalidades futuras (fora do MVP):

• Integração com sistemas de e-mail (ex.: Gmail API).
• Histórico de mensagens analisadas.
• Estatísticas sobre ataques detectad

4 IMPACTO SOCIAL

Sociedade Impactada: O projeto impacta diretamente os usuários individuais de internet e e-mail e as pequenas e médias empresas (PMEs), que são frequentemente alvos vulneráveis de ataques de phishing. O foco principal é o cidadão comum, que muitas vezes não possui o conhecimento técnico necessário para identificar golpes sofisticados.

A Contribuição Social: O detector contribui para a sociedade de duas maneiras principais:

Segurança Passiva: Reduzindo o número de incidentes bem-sucedidos de roubo de dados pessoais e financeiros, minimizando perdas e o estresse associado.

Educação Ativa: Fornecendo explicações claras e em linguagem simples sobre os motivos da suspeita , o sistema atua como uma ferramenta educacional, aumentando o nível de literacia digital e a capacidade do usuário de se proteger no futuro.

5 MÉTRICAS DE MEDIÇÃO DO IMPACTO

• O sucesso do projeto e seu impacto social serão medidos através de métricas tanto na fase de protótipo/MVP quanto em um cenário de uso real (funcionalidades futuras):

• Métricas de Desempenho do Modelo (Técnicas):

• Taxa de Detecção (Recall): Percentual de e-mails de phishing identificados corretamente pelo modelo.

• Precisão (Precision): Percentual de alertas emitidos que são realmente phishing (minimizando falsos positivos).

• Acurácia: Medida geral da corretude da classificação.

• Métricas de Usabilidade e Impacto (Foco no Usuário):

• Taxa de Aceitação da Explicação: Porcentagem de usuários que avaliam a explicação fornecida (item 4) como "clara" ou "útil" para o entendimento da fraude. (Medido via pesquisa de satisfação após o uso do MVP).

Redução de Clicks em Fraudes: Em um cenário futuro de integração (funcionalidade fora do MVP), o impacto social poderá ser medido pela diminuição da taxa de usuários que clicam em links maliciosos após a análise e alerta do sistema.

6 PLANO DE TRABALHO

Fase	Período	Entrega esperada	Responsáveis
C1 – Relatório (Entrega 1)	02/09 a 06/09	Documento com ideia, justificativa, escopo e plano de trabalho	Nathan
C2 – Protótipo (Entrega 2)	21/10 a 25/10	Vídeo no YouTube com protótipo funcional	Nathan
C3 – Produto Final (Entrega 3)	25/11 a 29/11	Vídeo no YouTube com MVP completo e documentação	Nathan