EXTRAÇÃO E CLASSIFICAÇÃO DE DADOS DO REDDIT COMO FORMA DE IDENTIFICAÇÃO AUTOMÁTICA DE GOLPES CIBERNÉTICOS

FERNANDA POLGA SOUZA[1]; CHARLLYNSON CARVALHO CAXIAS[2]; JUNIOR VINICIOS PREDIGER[3]; BRENDA SALENAVE SANTANA[4]

[1]Universidade Federal de Pelotas – fpsouza@inf.ufpel.edu.br
[2]Universidade Federal de Pelotas – cccaxias@inf.ufpel.edu.br
[3]Universidade Federal de Pelotas – jvprediger@inf.ufpel.edu.br
[4]Universidade Federal de Pelotas – bssalenave@inf.ufpel.edu.br

1. INTRODUÇÃO

Conforme o Anuário de Segurança Pública de 2024, o aumento de fraudes eletrônicas tem mudado o perfil dos crimes patrimoniais no Brasil, com golpes virtuais crescendo enquanto crimes de rua diminuem. Nesse contexto, este trabalho se propõe a coletar dados da rede social Reddit por meio de sua API, com o objetivo de identificar e categorizar golpes através da análise das descrições dos relatos de usuários. A área de estudo insere-se no campo da ciência de dados aplicada à segurança digital, com auxílio do Processamento de Linguagem Natural (PLN), área de pesquisa voltada para o desenvolvimento de sistemas que permitam a interação entre humanos e computadores por meio da linguagem natural. De acordo com CASELI; NUNES; PAGANO (2023), o PLN busca investigar e propor métodos de processamento computacional da linguagem humana, permitindo que grandes volumes de dados textuais possam ser analisados de forma automática e eficiente.

O objetivo geral deste trabalho é realizar uma categorização automatizada dos golpes cibernéticos publicados no Reddit, utilizando técnicas de PLN. Essa categorização tem como finalidade identificar padrões recorrentes entre os tipos de fraude, fornecendo subsídios para o desenvolvimento de estratégias de conscientização e prevenção. O projeto se encaixa com os objetivos propostos pelo grupo de estudos GEPESC - Grupo de Ensino, Pesquisa e Extensão em Segurança Cibernética - tendo como fundamentações teóricas fontes exploratórias de técnicas PLN na detecção de fraudes em ambientes digitais. A análise é realizada com base em técnicas de PLN, para que se possa entender como os usuários relatam fraudes online e de que forma esses dados podem ser utilizados para prevenir novos golpes.

2. METODOLOGIA

O trabalho é estruturado em três algoritmos principais, implementados na linguagem de programação Python, cada um desempenhando uma funcionalidade específica (etapas) na coleta, tratamento e análise dos dados extraídos da rede social Reddit, por meio de sua API oficial.

A primeira etapa é a extração de dados da plataforma Reddit por meio da API PRAW - Python Reddit API Wrapper - configurada para acessar postagens realizadas no subreddit ‘r/golpe’, fórum destinado a divulgação de relatos e tipos de golpes. Dessa forma, foi possível coletar 907 amostras de postagens, este

procedimento de extração também é conhecido como *web scraping* (Raspagem de dados).

De acordo com Glez-Peña et al., (2013), o web scraping pode ser definido como o processo de extrair e combinar conteúdos de interesse da Web de forma sistemática, em tal procedimento, um agente de software, imita a interação entre um servidor e o usuário humano. De maneira similar, na primeira etapa do trabalho, foram selecionadas diversas informações que pudessem ser pertinentes a associações durante as análises, tais como: index (para manter a contagem de posts coletados), data de postagem, identificador (ID), titulo da postagem, autor, flair (marcadores que categorizam o tipo de postagem), número de *upvotes (*votos de suporte à publicação) e *downvotes* (votos contrários à publicação) para contagem de interações classificando o nível de relevância, URL de origem, e conteúdo textual do post.

A segunda etapa inclui uma pré-classificação dos posts para garantir a integridade e relevância dos dados. Dessa forma, inicialmente foram eliminadas instâncias obtidas com posts vazios, isto é, sem textos associados, ou informações incompletas. Em seguida, considerando que o subreddit ‘r/golpe’ possui Flairs com significados bastante similares, uma filtragem foi realizada, onde as dez tags iniciais foram organizadas em cinco: ‘informativo’ engloba “informativo”, “explicação de golpe”, “reverti o golpe”, “aconteceu” e “notícia”; a tag ‘socorro’ engloba “ajuda” e “socorro”; e as que permanecem como únicas são ”meme”, “discussão” e *“scambait”*. Para a pós-classificação, foi criado um dicionário com categorias de golpes, utilizando palavras-chave comuns em relatos de fraudes, para serem percorridas pelo algoritmo em cada postagem e se encontradas adicionar uma nova informação ao arquivo final, classificando o post com sua categoria ideal. As palavras-chave utilizadas para identificação dos tipos de golpe foram definidas com base em observações empíricas e termos extraídos com o uso do YAKE! (Campos et al., 2020). Dessa forma, conseguimos reduzir o número de posts relevantes para 255 e padronizá-los baseados em suas semelhanças, o que serve de grande valia para *insights* entre as *flairs* existentes e tipos de golpe registrados.

A terceira etapa consiste na análise dos dados, já coletados e tratados, baseando-se na visualização das relações encontradas. Dessa maneira, o uso de bibliotecas de plotagem gráfica como o *Matplotlib* e o *Seaborn*, permitem uma visualização clara e acessível de padrões e tendências relacionadas aos golpes reportados. Essa etapa é importante por permitir lidar com a correlação de dados para obter resultados, como volume de relatos por período de tempo em esfera de macro e micro-classificação ou nível de engajamento das postagens por classificação delas, possibilitando assim uma análise mais aprofundada das fraudes relatadas no fórum.

3. RELATOS E IMPACTOS GERADOS

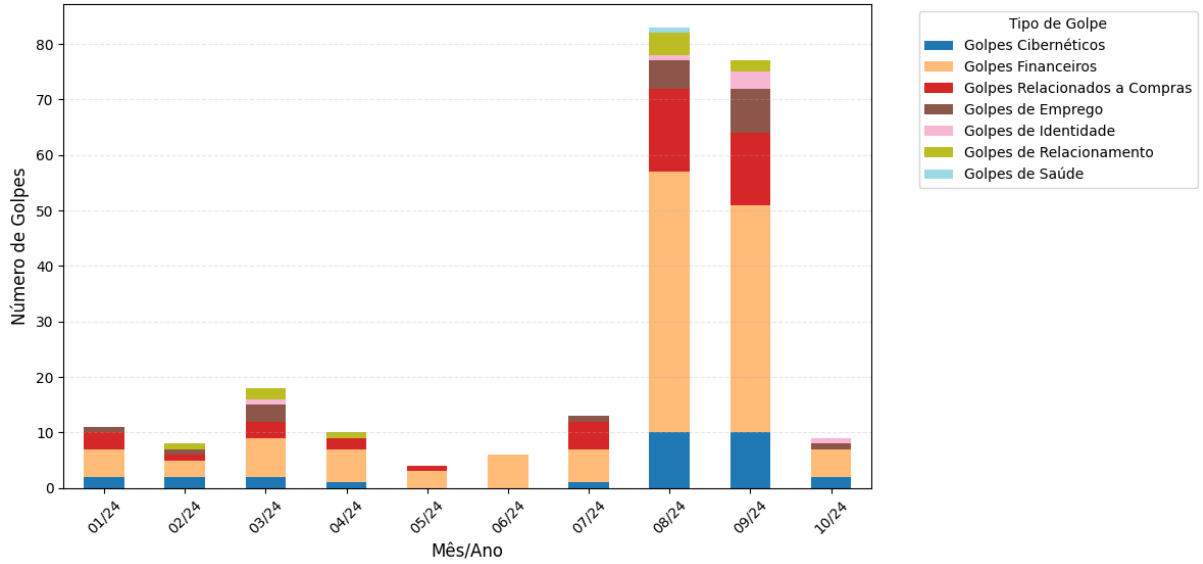
Até o presente momento, o projeto seguiu todas as etapas anteriormente definidas, sendo elas: a coleta de dados da API do Reddit e pré-processamento dos textos, agrupamento dos relatos de golpe por similaridade e a categorização automatizada das publicações realizadas no fórum.

A análise dos posts, utilizando um modelo baseado em palavras-chave junto com o uso de técnicas de agrupamento permitiu a identificação de padrões entre os diferentes tipos de golpes. A base de dados inicial contava com 907 posts que após filtragem para remoção de posts vazios, foi reduzida à 727. Desses, apenas 255

puderam ser classificados em tipos de golpes específicos, correspondendo a 35,07% dos posts válidos e 28,11% do total extraído. Embora a proporção de dados classificados seja baixa, demonstra o potencial da metodologia aplicada e como pode ser melhorada. Como melhorias futuras, planeja-se fazer uso de abordagens mais avançadas para reconhecimento de padrões e processamento dos dados, assim aumentando a precisão e cobertura da classificação de golpes e ampliando a eficiência da análise.

Com isso, considerando os dados filtrados e categorizados foi feito um agrupamento dos posts em sete categorias distintas, baseadas em suas recorrências e similaridades. A Figura 1 apresenta a ocorrência dos tipos de golpes identificados ao longo do ano corrente (janeiro a outubro de 2024).

Figura 1: Relatos de Golpe em 2024 por categoria.



Fonte: Autoria própria

Ao analisar a Figura 1, observa-se que, independentemente do mês, os golpes financeiros e relacionados a compras são os mais recorrentes, destacando-se em relação aos demais. A prevalência dos golpes financeiros não é surpreendente, dado o valor central que o dinheiro ocupa em nossa sociedade. No entanto, com o avanço da tecnologia e a conveniência proporcionada pelas plataformas digitais, esses crimes têm ganhado ainda mais força, conforme ilustrado pela Figura.

A crescente dependência de sistemas digitais, combinada com o baixo nível de alfabetização econômica e tecnológica em nosso país, agrava ainda mais esse cenário. Muitas pessoas, sem o conhecimento necessário, tornam-se alvos fáceis para fraudes e esquemas como pirâmides financeiras, promessas de retorno financeiro rápido, promoções falsas e golpes de “valores a receber”. Isso reforça a importância da conscientização e educação digital, que são essenciais para capacitar a população a identificar e evitar esses crimes, protegendo tanto suas finanças quanto a segurança digital.

A partir da identificação dos golpes mais recorrentes no ambiente digital é possível criar campanhas de divulgação ou palestras, de forma a conscientizar a comunidade local. Essas iniciativas podem ser complementadas com as ações já existentes no âmbito institucional do GEPESC, como o catálogo de fraudes - um repositório dedicado a documentar os mais variados tipos de golpes. Além disso, um

atendimento humanizado direcionado a pessoas vítimas desses delitos, com o fito de oferecer um suporte especializado e, consequentemente, uma possível solução para o problema.

4. CONSIDERAÇÕES

O projeto de identificação de golpes filtrados do Reddit se mostrou promissor no que diz respeito à identificação e organização automática dos mais variados tipos de fraudes relatadas. O catálogo de fraudes desenvolvido é uma ferramenta importante para conscientização da comunidade e para o aperfeiçoamento dos projetos desenvolvidos pelos núcleos do GEPESC. Ao fornecer orientações acessíveis e atualizadas, essa iniciativa capacita os indivíduos a proteger seus dados e dispositivos, reduzindo o risco de golpes e promovendo um ambiente digital mais seguro. A conscientização é essencial para a proteção de informações pessoais e coletivas, contribuindo para a estabilidade e segurança da sociedade.

Entre os dados mais relevantes, destaca-se que os golpes financeiros representaram a maior parte das ocorrências. Além disso, observa-se uma crescente onda de crimes digitais, o que ressalta a urgência de ações preventivas e de conscientização. Dada a gravidade e o impacto desse tipo de golpe na comunidade, é essencial priorizar iniciativas voltadas à educação e à proteção digital. Nesse contexto, ações de extensão ligadas à tecnologia desempenham um papel fundamental, pois abordam essas questões de forma prática e acessível, capacitando a população a se proteger e a navegar de maneira mais segura no ambiente digital.

Como trabalhos futuros, pretende-se realizar aumentar o conjunto de dados utilizado como base para este estudo, de forma a desenvolver uma abordagem mais fortemente embasada em dados e com resultados mais confiáveis. Ainda, pretende-se realizar a implementação de abordagens de aprendizado de máquina para realização de categorização automática dos dados, autônoma e mais precisa dos relatos coletados.

5. REFERÊNCIAS BIBLIOGRÁFICAS

Campos, R., Mangaravite, V., Pasquali, A., Jatowt, A., Jorge, A., Nunes, C. e Jatowt, A. (2020). **YAKE! Keyword Extraction from Single Documents using Multiple Local Features**. In Information Sciences Journal. Elsevier, Vol 509, pp 257-289.

Caseli, H.M.; Nunes, M.G.V. (org.) **Processamento de Linguagem Natural: Conceitos, Técnicas e Aplicações em Português.** 2 ed. BPLN, 2024. Disponível em: https://brasileiraspln.com/livro-pln/2a-edicao.

FÓRUM BRASILEIRO DE SEGURANÇA PÚBLICA. **Anuário Brasileiro de Segurança Pública 2024**. Acesso em: 08 out. 2024. Disponível em: https://publicacoes.forumseguranca.org.br/handle/123456789/253.

Daniel Glez-Peña, Anália Lourenço, Hugo López-Fernández, Miguel Reboiro-Jato, Florentino Fdez-Riverola, **Web scraping technologies in an API world, *Briefings in Bioinformatics*,** Volume 15, Issue 5, September 2014, Pages 788–797, https://doi.org/10.1093/bib/bbt026