PUC Minas

PONTIFÍCIA UNIVERSIDADE CATÓLICA DE MINAS GERAIS

Disciplina: [IEC_IAAM_O5_T1_Online] Processamento de Linguagem Natural

Alunos:

Jonas Aguiar Junior
Keli Tauana Prass Ruppenthal
Leonardo de Jesus Diz Conde

1. Título da Proposta

Análise Automatizada de Indicadores de Ameaça Cibernética Utilizando Técnicas de Processamento de Linguagem Natural

2. Resumo

Este projeto desenvolve um sistema automatizado que integra TheHive, VirusTotal e Processamento de Linguagem Natural (PLN) para analisar indicadores de ameaça (IPs, domínios, hashes) e enriquecer alertas com comentários gerados automaticamente. O código apresentado corresponde à etapa de análise textual via NLP, classificando relatórios do VirusTotal em maliciosos, suspeitos ou benignos com base em técnicas supervisionadas (Machine Learning) e regras heurísticas (palavras-chave), reduzindo a carga manual de analistas de segurança.

3. Introdução e Caracterização do Problema

Em ambientes de Security Operations Center (SOC), analistas enfrentam um volume crescente de alertas, muitos deles envolvendo a verificação de indicadores de comprometimento (IOCs) em ferramentas como VirusTotal. Esse processo é manual, repetitivo e demorado, levando a atrasos na resposta a incidentes. Além disso, a interpretação de relatórios técnicos exige conhecimento especializado, sobrecarregando equipes. A solução proposta automatiza a análise textual desses relatórios, combinando NLP e integração com APIs para acelerar a triagem de ameaças.

4. Proposta de Solução

O sistema segue um fluxo modular:

I. **Recepção no TheHive**: Alertas com IOCs são disparados.
II. **Consulta ao VirusTotal**: A API do VirusTotal obtém relatórios completos sobre os IOCs.
III. **Análise com NLP**:
    A. **Pré-processamento**: Limpeza do texto (tokenização, remoção de stopwords, normalização).

B. **Classificação**:

1. **Modelos supervisionados**: Pipelines com BoW/TF-IDF + Naive Bayes/SVM.
2. **Regras heurísticas**: Pontuação baseada em palavras-chave (ex: "malware" = +2 pontos).

C. **Ensemble**: Combina previsões dos modelos e regras para decisão final.

IV. **Geração de Comentários**: Explicações automáticas baseadas em estatísticas do relatório (ex: número de AVs que detectaram malícia).

V. **Inserção no TheHive**: O comentário é adicionado ao alerta via API.

5. Fontes de Dados

- **Relatórios do VirusTotal**: JSONs estruturados com campos como `last_analysis_results` (detalhes de antivírus) e `last_analysis_stats` (contagem de detecções).
- **Dados de Treino**: Relatórios históricos rotulados manualmente (ex: "malicioso" se ≥5 AVs detectaram ameaça).
- **Palavras-Chave**: Listas curradas (ex: `KEYWORDS_MALICIOUS = {"malware", "phishing"}`).

6. Experimentos realizados

Pré-processamento:

- Testou-se a eficácia da limpeza com spaCy e nltk (ex: impacto da remoção de stopwords).

Modelos de ML:

- Compararam-se **Bag-of-Words (BoW) + Naive Bayes** vs. **TF-IDF + SVM Linear**.
- Avaliação com métricas (precision, recall, F1-score) via classification_report.

Regras Heurísticas:

- Validação manual para ajuste de pesos (ex: "malicioso" se score ≥4).

Ensemble:

- Verificou-se a acurácia da votação majoritária entre ML e regras.

7. Resultados alcançados

```
*** bow_nb ***
              precision    recall  f1-score   support

     benigno       0.93      1.00      0.97        14
   malicioso       0.67      1.00      0.80         2
    suspeito       1.00      0.50      0.67         4

    accuracy                           0.90        20
   macro avg       0.87      0.83      0.81        20
weighted avg       0.92      0.90      0.89        20


*** tfidf_svc ***
              precision    recall  f1-score   support

     benigno       0.82      1.00      0.90        14
   malicioso       0.50      0.50      0.50         2
    suspeito       1.00      0.25      0.40         4

    accuracy                           0.80        20
   macro avg       0.77      0.58      0.60        20
weighted avg       0.83      0.80      0.76        20
```

```
Indicador suspeito com 0 alertas. IP 202.131.82.244, ASN 'Cambo TechnologyISP Co.,Ltd' (KH). Monitoramento contínuo é recomendado.
 benigno_66-249-64-0.txt -> BENIGNO
Indicador benigno com 63 detecções limpas. IP 66.249.64.0, ASN 'GOOGLE' (US). Sem sinais de risco atuais.
 benigno_13-70-0-0.txt -> BENIGNO
Indicador benigno com 62 detecções limpas. IP 13.70.0.0, ASN 'MICROSOFT-CORP-MSN-AS-BLOCK' (HK). Sem sinais de risco atuais.
 malicioso_64-62-197-238.txt -> MALICIOSO
Indicador classificado como malicioso com 11 detecções confirmadas. O IP 64.62.197.238 pertence ao ASN 'HURRICANE' (US), com reputação -2. Tags: . Recomenda
do bloqueio e investigação.
 suspeito_103-241-67-157.txt -> SUSPEITO
Indicador suspeito com 2 alertas. IP 103.241.67.157, ASN 'KAMATERA' (ES). Monitoramento contínuo é recomendado.
 benigno_13-66-0-0.txt -> BENIGNO
Indicador benigno com 62 detecções limpas. IP 13.66.0.0, ASN 'MICROSOFT-CORP-MSN-AS-BLOCK' (US). Sem sinais de risco atuais.
 benigno_42-179-217-67.txt -> BENIGNO
Indicador benigno com 0 detecções limpas. IP 42.179.217.67, ASN 'CHINA UNICOM China169 Backbone' (CN). Sem sinais de risco atuais.
 benigno_13-72-0-0.txt -> BENIGNO
Indicador benigno com 0 detecções limpas. IP 13.72.0.0, ASN 'MICROSOFT-CORP-MSN-AS-BLOCK' (US). Sem sinais de risco atuais.
 benigno_13-76-0-0.txt -> BENIGNO
Indicador benigno com 63 detecções limpas. IP 13.76.0.0, ASN 'MICROSOFT-CORP-MSN-AS-BLOCK' (SG). Sem sinais de risco atuais.
 benigno_101-32-214-229.txt -> BENIGNO
Indicador benigno com 0 detecções limpas. IP 101.32.214.229, ASN 'Tencent Building, Kejizhongyi Avenue' (HK). Sem sinais de risco atuais.
 benigno_13-75-0-0.txt -> BENIGNO
Indicador benigno com 62 detecções limpas. IP 13.75.0.0, ASN 'MICROSOFT-CORP-MSN-AS-BLOCK' (HK). Sem sinais de risco atuais.
 suspeito_41-141-11-43.txt -> BENIGNO
Indicador benigno com 59 detecções limpas. IP 41.141.11.43, ASN 'MT-MPLS' (MA). Sem sinais de risco atuais.
 suspeito_27-112-79-160.txt -> SUSPEITO
Indicador suspeito com 2 alertas. IP 27.112.79.160, ASN 'PT Cloud Hosting Indonesia' (ID). Monitoramento contínuo é recomendado.
 suspeito_103-243-242-61.txt -> SUSPEITO
Indicador suspeito com 0 alertas. IP 103.243.242.61, ASN 'PacketFabric Japan Co., Ltd.' (JP). Monitoramento contínuo é recomendado.
 suspeito_201-231-83-229.txt -> BENIGNO
Indicador benigno com 60 detecções limpas. IP 201.231.83.229, ASN 'Telecom Argentina S.A.' (AR). Sem sinais de risco atuais.
 benigno_13-80-0-0.txt -> BENIGNO
Indicador benigno com 62 detecções limpas. IP 13.80.0.0, ASN 'MICROSOFT-CORP-MSN-AS-BLOCK' (NL). Sem sinais de risco atuais.
 benigno_13-74-0-0.txt -> BENIGNO
Indicador benigno com 62 detecções limpas. IP 13.74.0.0, ASN 'MICROSOFT-CORP-MSN-AS-BLOCK' (IE). Sem sinais de risco atuais.
 benigno_13-82-0-0.txt -> BENIGNO
Indicador benigno com 63 detecções limpas. IP 13.82.0.0, ASN 'MICROSOFT-CORP-MSN-AS-BLOCK' (US). Sem sinais de risco atuais.
```

8. Conclusões e trabalhos futuros

A solução reduz significativamente o tempo de triagem de IOCs ao automatizar a análise textual com NLP, garantindo **consistência** e **rastreabilidade** (via comentários no TheHive). A abordagem híbrida (ML + regras) melhora a robustez, especialmente para casos limítrofes. No entanto, ainda há muito a ser melhorado, visto que quando comparado ao desempenho de um LLM neste cenário, os resultados são bastante completos. Aqui, o link do repositório onde se encontra o projeto: https://github.com/jonasaguiarj/thehive-pln-alert-classifier.git

Trabalhos Futuros:

Expansão do Dataset: Incorporar mais relatórios rotulados para refinar os modelos.
Integração em Tempo Real: Acionar o pipeline automaticamente via webhooks do TheHive.
Análise Multimodal: Combinar NLP com metadados (ex: reputação de IP) para maior precisão.
Feedback de Analistas: Usar classificações manuais para ajuste contínuo (active learning).

9. Referências

Jurafsky, D., & Martin, J. H. (2024).
Speech and Language Processing (3rd ed.). Pearson.

Bird, S., Klein, E., & Loper, E. (2009).
Natural Language Processing with Python (1st ed.). O'Reilly.

ROCHA, R. SOAR Automation for CSIRT Teams. LinkedIn, 2024. Disponível em: https://www.linkedin.com/posts/romrocha_soar-automation-csirt-activity-7224103678842986497-XfBA/. Acesso em: 28 abril. 2024.