

# Detecção Automatizada de Negociações de Substâncias Ilícitas em Mensagens de WhatsApp

Silvio R. Tassini Borges

08 de julho de 2025



PUC Minas

# Sumário

- 1 Motivação
- 2 Objetivos
- 3 Hipóteses
- 4 Perguntas de Pesquisa
- 5 Justificativa
- 6 Trabalhos Relacionados
- 7 Metodologia
- 8 Referências



PUC Minas

- WhatsApp é amplamente utilizado para comunicações criminosas devido ao seu alcance e percepção de privacidade
- Crescimento do tráfico de drogas através de plataformas digitais
- Necessidade de ferramentas automatizadas para análise pericial de grandes volumes de conversas



- Aplicações em investigações criminais:
  - Identificação de padrões de negociação de substâncias ilícitas
  - Mapeamento de redes de tráfico
  - Coleta de evidências digitais
- Limitações das análises manuais: tempo, volume de dados e subjetividade
- Lacuna tecnológica na perícia criminal para detecção automática de negociações ilícitas



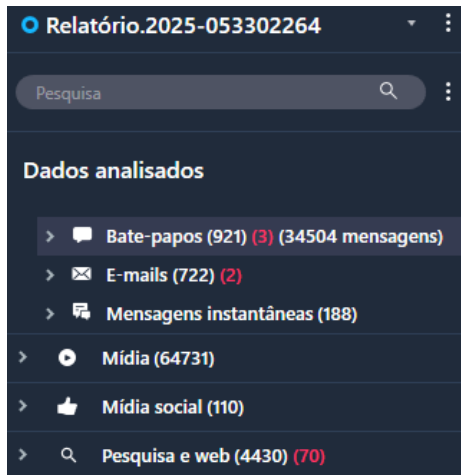


Figura: Volume de mensagens em uma extração rotineira



## Objetivo Principal

Desenvolver um modelo de Processamento de Linguagem Natural para a identificação automática de padrões de negociação em mensagens de WhatsApp vinculadas ao comércio de substâncias ilícitas.



# Objetivos Específicos

- 1 Coletar e preprocessar dados de conversas do WhatsApp
- 2 Identificar características linguísticas indicativas de comércio de substâncias ilícitas
- 3 Implementar modelos de classificação para detecção
- 4 Avaliar a performance dos modelos propostos
- 5 Comparar diferentes abordagens de vetorização
- 6 Analisar padrões linguísticos e contextuais das negociações



PUC Minas

## H1 - Termos Indicativos

Mensagens que apresentam combinações de termos relacionados a transações financeiras e logísticas são mais frequentemente classificadas como negociações?

## H2 - Padrão linguístico

O uso de linguagem cifrada ou gírias específicas aumenta a probabilidade de uma mensagem estar relacionada à negociação de substâncias ilícitas





- ① **RQ1:** Quais características linguísticas são mais indicativas de mensagens relacionadas à negociação de substâncias ilícitas no WhatsApp?
- ② **RQ2:** Como diferentes técnicas de vetorização (TF-IDF, Word2Vec, BERT) se comparam na detecção de negociações?
- ③ **RQ3:** É possível identificar diferentes tipos de negociação (preço, prazo, condições)?



- **Relevância Técnico-Científica:**

- Contribuição para área de PLN aplicado a conversas informais
- Automatização da análise de evidências digitais
- Redução do tempo de análise pericial

- **Relevância para Segurança Pública:**

- Combate ao tráfico de drogas em plataformas digitais
- Modernização das técnicas de perícia criminal
- Agilização de investigações policiais



- **(SURYAKUMAR et al., 2025)** - Visualization of Hidden Patterns in WhatsApp Chat Using Sentimental Analysis
  - Revelar padrões de comunicação
  - Identificar os tópicos de discussão
- **(SUBBULAKSHMI et al., 2024)** - Forensic Investigation and WhatsApp Chat Analysis using Named Entity Recognition with Web based Visualization
  - Reconhecimento de Entidades Nomeadas



- **(YAMAGUCHI et al., 2023)** - Dialogue Act-based Breakdown Detection in Negotiation Dialogues
  - Identificação de atos de fala em diálogos
  - Identificar falhas em negociações
- **(SHAH et al., 2022)** - An unsupervised machine learning approach for the detection and characterization of illicit drug-dealing comments and interactions on Instagram
  - Biterm Topic Model para extrair e sumarizar
  - Identificação de sinal e ruído



- 1 Coleta de Dados
- 2 Pré-processamento
- 3 Anotação e Rotulação
- 4 Extração de Features
- 5 Modelagem
- 6 Avaliação
- 7 Análise dos Resultados



## Coleta de Dados

- Conversas exportadas do WhatsApp (formato .txt)
- Critérios de inclusão/exclusão

## Pré-processamento

- Limpeza de metadados
- Normalização de texto
- Tratamento de emojis e abreviações



## Trecho de conversa

[10:30] A: Oi, tem aí?

[10:31] B: Tenho sim, qto quer?

[10:32] A: 2g

[10:33] B: 50 reais, fechou?

[10:34] A: uhummm

[18:45] A: E aquele outro produto?

[18:46] B: Qual?

[18:47] A: O mais forte

[18:48] B: Ah sim, 100 pau

## Extração de Features

- **Linguísticas:** N-grams, POS tags, entidades nomeadas
- **Semânticas:** Embeddings





## Modelos

- **Baseline:** TF-IDF + SVM/Random Forest
- **Word2Vec:** + Redes Neurais
- **BERT:** Fine-tuning para classificação
- **Ensemble:** Combinação de modelos



## Métricas de Avaliação

- Coherence Score
- Curva ROC e AUC
- Acurácia, Precisão, Recall, F1-Score

## Análise

- Análise de features mais importantes
- Estudo de casos específicos
- Análise de erros



# Referências



LSS, Suryakumar et al. Visualization of Hidden Patterns in WhatsApp Chat Using Sentimental Analysis. In: 2025 3rd International Conference on Intelligent Data Communication Technologies and Internet of Things (IDCIoT). IEEE, 2025. p. 8-13.



SUBBULAKSHMI, N. et al. Forensic Investigation and WhatsApp Chat Analysis Using Named Entity Recognition with Web Based Visualization. In: 2024 International Conference on Sustainable Communication Networks and Application (ICSCNA). IEEE, 2024. p. 1683-1688.



KAUSHAL, Royal; CHADHA, Raman. A Survey of Various Sentiment Analysis Techniques of Whatsapp. In: 2023 2nd International Conference for Innovation in Technology (INOCON). IEEE, 2023. p. 1-6.



YAMAGUCHI, Atsuki; IWASA, Kosui; FUJITA, Katsuhide. Dialogue act-based breakdown detection in negotiation dialogues. In: Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume. 2021. p. 745-757.



SHAH, Neal; LI, Jiawei; MACKEY, Tim K. An unsupervised machine learning approach for the detection and characterization of illicit drug-dealing comments and interactions on Instagram. Substance abuse, v. 43, n. 1, p. 273-277, 2022.



HUTTO, Clayton; GILBERT, Eric. Vader: A parsimonious rule-based model for sentiment analysis of social media text. In: Proceedings of the international AAAI conference on web and social media. 2014. p. 216-225.



PUC Minas

# Obrigado!

Perguntas?



PUC Minas