Classificação Automática de Relatos de Ocorrências Policiais para Crimes Cibernéticos em Português Brasileiro: Uma Abordagem Comparativa com Modelos Transformer

Victor Carvalho Soares de Araujo[1], **José Victor Vieira de Oliveira**[1], **Pedro Feitosa Soares**[1], **Yasmine Martins da Costa e Silva**[1], **Rogério Figueredo Sousa**[2], **Raimundo Santos Moura**[1]

[1] Universidade Federal do Piauí (UFPI) – Teresina, PI – Brasil
[2] Instituto Federal do Piauí (IFPI) – Picos, PI – Brasil

{victor.araujo, jose.victor, pedro.soares, yasmine.silva, rsm}@ufpi.edu.br, rogerio.sousa@ifpi.edu.br

Abstract. *The study evaluates the use of Transformers for automatic classification of cybercrime police reports in Brazilian Portuguese. BERTimbau-large and XLM-RoBERTa-large were fine-tuned with balanced samples from more than ten thousand police records from Piauí. The evaluation showed that BERTimbau-large achieved superior and more stable performance, especially under strong class imbalance. Most errors stem from semantic overlap among similar crime types. The results indicate that Portuguese-specialized models are a robust solution for automating police triage.*

Resumo. *Este trabalho investiga o uso de Transformers para classificar automaticamente boletins de ocorrência de crimes cibernéticos em português brasileiro. Foram ajustados os modelos BERTimbau-large e XLM-RoBERTa-large com amostras balanceadas de mais de 10 mil registros da polícia do Piauí. A avaliação mostrou que o BERTimbau-large teve desempenho superior e mais estável, sobretudo em cenários com forte desbalanceamento. Os principais erros decorrem da sobreposição semântica entre tipos de crime semelhantes. Os resultados indicam que modelos especializados em português são uma solução robusta para automação de triagem policial.*

1. Introdução

A digitalização dos serviços públicos tem gerado um crescimento significativo no volume de registros de ocorrências em plataformas virtuais. A Secretaria de Segurança Pública do Estado do Piauí (SSP-PI) contabilizou 48.282 relatos via Delegacia Virtual[1] (DEVIR) entre agosto de 2024 e julho de 2025 — uma média de 132 notificações diárias. Tal volume impõe desafios operacionais à classificação e homologação manual desses registros. O "relato histórico", campo textual não estruturado onde o cidadão descreve livremente o fato, constitui a fonte informacional mais rica do boletim, porém, seu processamento manual representa um gargalo crítico.

Um subconjunto complexo dessas notificações refere-se aos crimes cibernéticos, investigados pela Delegacia de Repressão e Combate aos Crimes de Informática (DRCI),

[1] https://delegaciavirtual.sinesp.gov.br/portal/

com predominância de delitos como “Estelionato”, “Invasão de Dispositivo Informático”e “Falsa Identidade”. Tais ocorrências frequentemente apresentam complexidade semântica e sobreposição conceitual: um estelionato pode envolver crimes-meio subsequentes, como invasão de dispositivo informático que, por sua vez, se vale de falsa identidade. A correta classificação jurídica do fato típico constitui, portanto, um desafio não trivial.

Após a etapa de classificação, é possível realizar análises posteriores, incluindo a detecção de inconsistências entre o relato textual e a natureza jurídica atribuída, extração de informações estruturadas (*modus operandi*, dados de suspeitos, objetos subtraídos) e identificação de padrões relacionais entre ocorrências de mesma natureza.

Este artigo avalia a aplicabilidade de Grandes Modelos de Linguagem (LLMs) baseados na arquitetura *Transformer* [Vaswani et al. 2017] para classificação automática da natureza jurídica de crimes cibernéticos, utilizando exclusivamente o texto do campo “relato histórico”. O objetivo consiste em estabelecer um *baseline* de desempenho para essa tarefa em corpus do mundo real, caracterizado por ruído textual, linguagem coloquial e terminologia específica do domínio jurídico-policial brasileiro.

Utilizou-se um corpus de 14.188 registros de boletins de ocorrências da DRCI. Das cinco classes mais prevalentes no conjunto de dados, totalizando 10.340 registros, extraíram-se 1.500 amostras balanceadas, 300 por categoria, para *fine-tuning*. Comparou-se o desempenho do modelo monolíngue *BERTimbau-large* [Souza et al. 2020], especializado em português brasileiro, com o modelo multilíngue *XLM-RoBERTa-large* [Conneau et al. 2020], visando identificar a abordagem mais adequada para classificação de ocorrências policiais neste domínio específico. A avaliação foi conduzida em dois cenários complementares: um conjunto balanceado com 375 instâncias para análise equitativa do desempenho por classe, e um conjunto desbalanceado com 2.068 instâncias refletindo a distribuição natural dos dados, permitindo avaliar a robustez operacional dos modelos em condições realistas de implantação.

Este trabalho está estruturado: a Seção 2 examina trabalhos correlatos; a Seção 3 descreve o corpus; a Seção 4 detalha a configuração experimental; a Seção 5 apresenta os resultados e as análises; e a Seção 6 apresenta conclusões e direções futuras.

2. Trabalhos Relacionados

A análise de dados criminais tem sido impulsionada pelo uso de técnicas de aprendizado de máquina aplicadas a boletins de ocorrência, que reúnem campos estruturados e narrativas textuais. Esses relatos apresentam desafios de processamento, mas também permitem extrair padrões relevantes da dinâmica criminal. A literatura tem explorado múltiplas estratégias para classificar, prever e recuperar informações a partir desses textos.

Algoritmos tradicionais continuam relevantes na área. [Amorim and Pereira 2019] aplicaram modelos como C4.5 e Ripper para tipificar o “histórico do fato”, obtendo até 99% de acurácia em um cenário binário e balanceado. Eles reportaram uma limitação ao lidar com 381 classes, inviabilizando a classificação completa. [de Castro 2020], por sua vez, utilizou *Random Forest* e *XGBoost* para prever tendências criminais a partir de fontes oficiais e não oficiais. Enquanto o primeiro foca na classificação do texto individual, o segundo busca padrões agregados, mostrando usos distintos das técnicas clássicas.

Modelos baseados em redes neurais ampliaram o desempenho, sobretudo em ba-

ses maiores. [Matos et al. 2022] utilizou uma CNN para classificar 463 classes de BOs no Pará, com acurácia geral de 78%, prejudicada pelo forte desbalanceamento. Em contraste, [Alves et al. 2024] restringiram a tarefa aos dez crimes mais frequentes e aplicaram modelos *Transformer* (*BERT* e *RoBERTa*), alcançando cerca de 90% de acurácia. Nessa direção, [de Sousa and Moura 2025] analisaram três métodos — *Random Forest* com TF-IDF, uma *CNN-LSTM* e um LLM ajustado (*Qwen 2.5 7B*) — e o LLM apresentou o melhor F1-Score (0,69), reforçando a vantagem dos *transformers*. A comparação evidencia que a redução do número de classes, aliada a modelos modernos, favorece o desempenho.

Outras pesquisas abordaram tarefas mais específicas. [Araújo 2023] investigou a similaridade entre BOs, mostrando que TF-IDF foi mais eficaz para a similaridade sintática, enquanto SBERT [Reimers and Gurevych 2019] e USE (*Universal Sentence Encoder*) [Cer et al. 2018] superaram na semântica. Lins e Saraiva (2024) propuseram uma arquitetura multimodelo para extrair atributos como localização, modus operandi e itens subtraídos, combinando CNN, LSTM e técnicas para lidar com desbalanceamento, como *Focal Loss* [Lin et al. 2017].

No geral, a análise da literatura revela uma evolução das técnicas, que passaram de algoritmos clássicos aplicados a subconjuntos de dados para redes neurais em larga escala e, mais recentemente, para modelos *Transformer* e abordagens multimodais voltadas à extração detalhada de informações. Apesar das diferenças quanto às tarefas e aos tipos de dados, os desafios recorrentes permanecem, como o desbalanceamento e a alta dimensionalidade. Isso reforça a necessidade de alinhar o método às características da base e aos objetivos analíticos.

3. Corpus de Trabalho

O corpus utilizado neste estudo compreende registros de boletins de ocorrências policiais coletados pela DRCI da Secretaria de Segurança Pública do Piauí (SSP-PI). O conjunto inicial continha 20.216 registros estruturados em duas colunas: *naturezas*, com a tipificação jurídica atribuída pelos agentes de polícia no momento do registro, e *"relato_historico"*, contendo a narrativa textual da ocorrência, possuindo em média 91,64 palavras cada.

3.1. Análise de Qualidade e Processo de Limpeza

A inspeção preliminar revelou inconsistências críticas que comprometiam o treinamento confiável dos modelos. Implementou-se processo de depuração em duas etapas: (i) exclusão de 4.531 registros com valores ausentes na coluna *naturezas*; e (ii) remoção de 1.497 entradas com campo *relato_historico* vazio ou preenchido apenas com "*Não informado*", evitando o aprendizado de correlações espúrias entre ausência textual e categorias específicas.

Os rótulos em *naturezas* refletem a interpretação técnica do servidor responsável pela homologação do boletim com base na análise preliminar do relato fornecido pelo cidadão. Portanto, representam decisões reais de triagem policial, com suas variações e eventuais inconsistências inerentes ao processo operacional.

3.2. Corpus Final

Após a depuração, o conjunto de dados continha 14.188 registros válidos, correspondendo a aproximadamente 70% da base original. A Tabela 1 exibe a distribuição das cinco

categorias mais prevalentes, evidenciando acentuado desbalanceamento: "Estelionato" representa 41,7% das ocorrências, enquanto "Extorsão" apenas 2,6%.

Tabela 1. Distribuição das cinco classes mais frequentes no corpus limpo

Classe	Frequência	Percentual (%)
Estelionato	5.920	41,7
Perda ou Extravio de Documento e/ou Objeto	1.834	12,9
Invasão de Dispositivo Informático	1.461	10,3
Falsa Identidade	750	5,3
Extorsão	375	2,6
Total	**10.340**	**72,8**

Cabe destacar que "Perda ou Extravio de Documento e/ou Objeto", embora não constitua crime cibernético em sentido estrito, integra o corpus por fazer parte do fluxo operacional da DRCI, frequentemente associada a documentos digitais ou dispositivos eletrônicos. Sua inclusão mostra-se relevante para avaliar a capacidade dos modelos em distinguir ocorrências não-criminosas de crimes cibernéticos propriamente ditos, aspecto fundamental para sistemas de triagem automatizada.

Estas cinco categorias mais frequentes, totalizando 10.340 registros, foram selecionadas para o desenvolvimento do classificador, assegurando volume amostral adequado para treinamento e avaliação robusta dos modelos, conforme detalhado na Seção 4.

4. Metodologia

4.1. Modelos de Linguagem

Para este estudo, selecionaram-se dois modelos baseados em arquitetura *Transformer*, ambos instanciados via biblioteca `transformers`, utilizando a classe `AutoModelForSequenceClassi` com cabeça de classificação configurada para 5 categorias.

- **BERTimbau-large:** `neuralmind/bert-large-portuguese-cased`, especializado em português brasileiro e pré-treinado em extenso corpus nativo.
- **XLM-RoBERTa-large:** `xlm-roberta-large`, modelo multilíngue de alta capacidade, empregado como baseline comparativo.

4.2. Estratégia de Balanceamento e Particionamento

Devido ao acentuado desbalanceamento observado na Tabela 1, adotou-se uma estratégia de amostragem balanceada para o treinamento. Do conjunto total de 10.340 registros das cinco classes selecionadas, extraíram-se aleatoriamente 1.500 instâncias balanceadas (300 por categoria) para compor o conjunto de treinamento. Este conjunto de 1.500 amostras foi particionado internamente em:

- **Treino:** 1.200 amostras (80%)
- **Validação interna:** 300 amostras (20%)

Para avaliação final dos modelos, foram criados dois conjuntos de teste independentes, mantidos isolados durante todo o processo de treinamento:

- **Conjunto balanceado:** 375 instâncias (75 por classe), permitindo análise equitativa do desempenho em todas as categorias.
- **Conjunto desbalanceado:** 2.068 instâncias mantendo a distribuição natural do corpus original, refletindo o cenário real de operação da SSP-PI. Este conjunto contém 1.184 casos de Estelionato, 367 de Perda ou Extravio, 292 de Invasão de Dispositivo, 150 de Falsa Identidade e 75 de Extorsão.

A Figura 1 ilustra o esquema completo de particionamento dos dados.

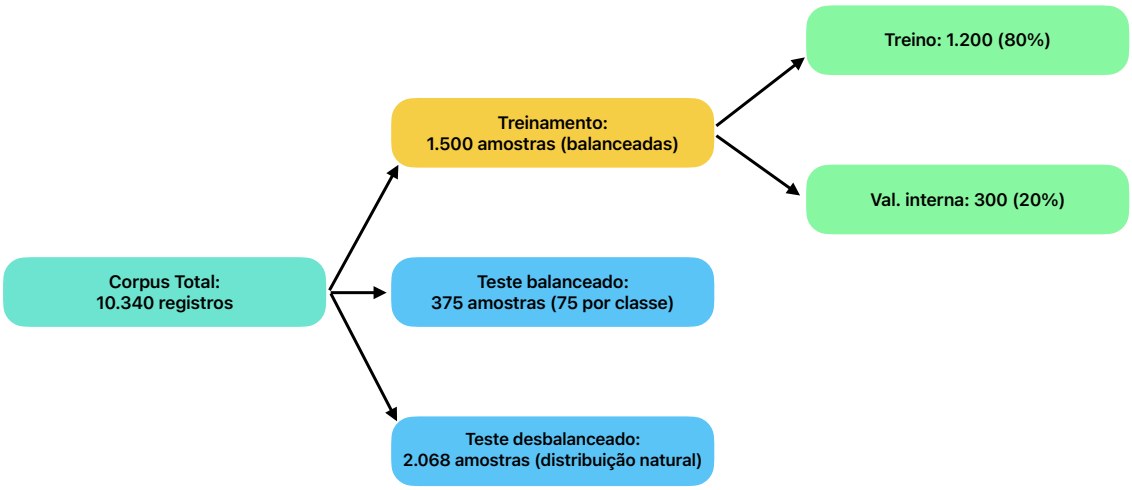

Figura 1. Particionamento dos dados

4.3. Configuração do Fine-Tuning

O fine-tuning foi executado em ambiente local devido à sensibilidade dos dados, utilizando GPU NVIDIA GeForce RTX 3060 (12 GB VRAM), processador AMD Ryzen 5 5600GT (3.60 GHz) e 32 GB de memória RAM, com precisão mista (FP16) habilitada para otimização computacional. Os hiperparâmetros de treinamento, implementados via classe `TrainingArguments` e posteriormente passados como parâmetro para o `Trainer` da biblioteca `transformers`, foram idênticos para ambos os modelos: taxa de aprendizado de 5e-5, 10 épocas (máximo), *batch size* de 8, 500 *warmup steps* e *weight decay* de 0.01.

Para seleção do modelo final, implementou-se *early stopping* com paciência de 2 épocas (`early_stopping_patience=2`), monitorando o *F1-Score* ponderado no conjunto de validação interna a cada época. O `Trainer` foi configurado para salvar automaticamente apenas o *checkpoint* de melhor desempenho (`load_best_model_at_end=True`).

O treinamento do BERTimbau convergiu na época 6 em aproximadamente 10 minutos e 35 segundos, enquanto o XLM-RoBERTa convergiu na época 4 em aproximadamente 1 hora e 44 minutos. A inferência no conjunto de validação (375 amostras) foi executada a uma taxa média de aproximadamente 35 amostras por segundo para ambos os modelos.

4.4. Protocolo de Avaliação

A avaliação dos modelos foi conduzida em duas etapas complementares:

Etapa 1 - Avaliação em conjunto balanceado: Os modelos foram avaliados no conjunto de 375 instâncias balanceadas para análise equitativa do desempenho por classe, eliminando o efeito da prevalência de categorias majoritárias.

Etapa 2 - Avaliação em conjunto desbalanceado: Para validação da robustez em condições realistas, os mesmos modelos foram avaliados no conjunto de 2.068 instâncias com distribuição natural, simulando o cenário operacional da SSP-PI onde o Estelionato representa a maioria das ocorrências policiais registrada.

5. Resultados

Esta seção apresenta a avaliação comparativa dos modelos BERTimbau-large e XLM-RoBERTa-large em dois cenários distintos: um conjunto balanceado com 375 instâncias e um conjunto desbalanceado com 2.068 instâncias, refletindo a distribuição natural do corpus.

5.1. Avaliação em Conjunto Balanceado

A primeira avaliação foi conduzida em um conjunto de instâncias balanceadas, permitindo uma análise equitativa do desempenho dos modelos em todas as categorias.

5.1.1. Desempenho Geral

A Tabela 2 apresenta as métricas gerais de desempenho dos dois modelos avaliados. O BERTimbau-large obteve desempenho superior em todas as métricas globais, superando o XLM-RoBERTa-large em 3,74 pontos percentuais de acurácia. Ambos os modelos demonstraram alta confiança média nas predições, com valores superiores a 97%.

Tabela 2. Comparação do desempenho geral dos modelos (conjunto balanceado)

Modelo	Acurácia	Precisão	Recall	F1-Score	Confiança Média
BERTimbau-large	0,8987	0,9011	0,8987	0,8987	0,9798
XLM-RoBERTa-large	0,8613	0,8855	0,8613	0,8630	0,9739

5.1.2. Desempenho por Classe

A Tabela 3 apresenta a comparação detalhada do F1-Score por classe para ambos os modelos. Observa-se comportamento heterogêneo entre as classes, com desempenho notavelmente superior na categoria "Perda ou Extravio de Documento e/ou Objeto" para ambos os modelos.

O BERTimbau-large superou o XLM-RoBERTa-large em todas as cinco classes, com diferenças variando entre 2 e 5 pontos percentuais. Destaca-se que a classe "Perda ou Extravio de Documento e/ou Objeto" atingiu desempenho quase perfeito, indicando vocabulário semanticamente distinto e facilmente discriminável pelos modelos.

5.2. Avaliação em Conjunto Desbalanceado

Para avaliar a robustez dos modelos em condições mais próximas da distribuição real dos dados, realizou-se uma segunda avaliação utilizando o restante do corpus completo com 2.068 instâncias, mantendo o desbalanceamento natural.

Tabela 3. Comparação do F1-Score por classe (conjunto balanceado)

Classe	BERTimbau	XLM-RoBERTa	Δ
Estelionato	0,84	0,80	+0,04
Extorsão	0,95	0,92	+0,03
Falsa Identidade	0,87	0,84	+0,03
Invasão de Dispositivo	0,84	0,79	+0,05
Perda ou Extravio	0,99	0,97	+0,02

5.2.1. Desempenho Geral

A Tabela 4 apresenta os resultados neste cenário mais desafiador. Os modelos demonstraram notável consistência em relação ao cenário balanceado, com o BERTimbau-large mantendo acurácia de 89,51% (apenas 0,36 pontos percentuais inferior ao cenário balanceado) e o XLM-RoBERTa-large alcançando 83,70% de acurácia.

Tabela 4. Desempenho geral dos modelos (conjunto desbalanceado)

Modelo	Acurácia	Precisão (weighted)	Recall (weighted)	F1-Score (weighted)
BERTimbau-large	0,8951	0,92	0,90	0,90
XLM-RoBERTa-large	0,8370	0,89	0,84	0,85

A diferença de desempenho entre os modelos se acentuou neste cenário, com o BERTimbau-large superando o XLM-RoBERTa-large em 5,81 pontos percentuais de acurácia, comparado aos 3,74 pontos no cenário balanceado. Este resultado sugere uma maior capacidade de generalização do modelo monolíngue (BERTimbau-large) para distribuições de teste desbalanceadas, mesmo quando treinado em dados balanceados.

5.2.2. Desempenho por Classe

A Tabela 5 apresenta o detalhamento do desempenho por classe no conjunto desbalanceado, revelando comportamentos distintos em relação ao cenário balanceado.

Tabela 5. Desempenho detalhado por classe - BERTimbau-large (conjunto desbalanceado)

Classe	Precisão	Recall	F1-Score	Support
Estelionato	0,98	0,86	0,92	1184
Extorsão	0,65	0,95	0,77	75
Falsa Identidade	0,68	0,90	0,77	150
Invasão de Dispositivo Informático	0,73	0,88	0,80	292
Perda ou Extravio de Doc. e/ou Obj.	1,00	1,00	1,00	367

A Figura 2 apresenta as matrizes de confusão para o cenário desbalanceado, evidenciando os padrões de erro de cada modelo.

Tabela 6. Desempenho detalhado por classe - XLM-RoBERTa-large - conjunto desbalanceado

Classe	Precisão	Recall	F1-Score	Support
Estelionato	0,98	0,77	0,86	1184
Extorsão	0,68	0,89	0,77	75
Falsa Identidade	0,79	0,75	0,77	150
Invasão de Dispositivo Informático	0,54	0,96	0,69	292
Perda ou Extravio de Doc. e/ou Obj.	0,95	0,99	0,97	367

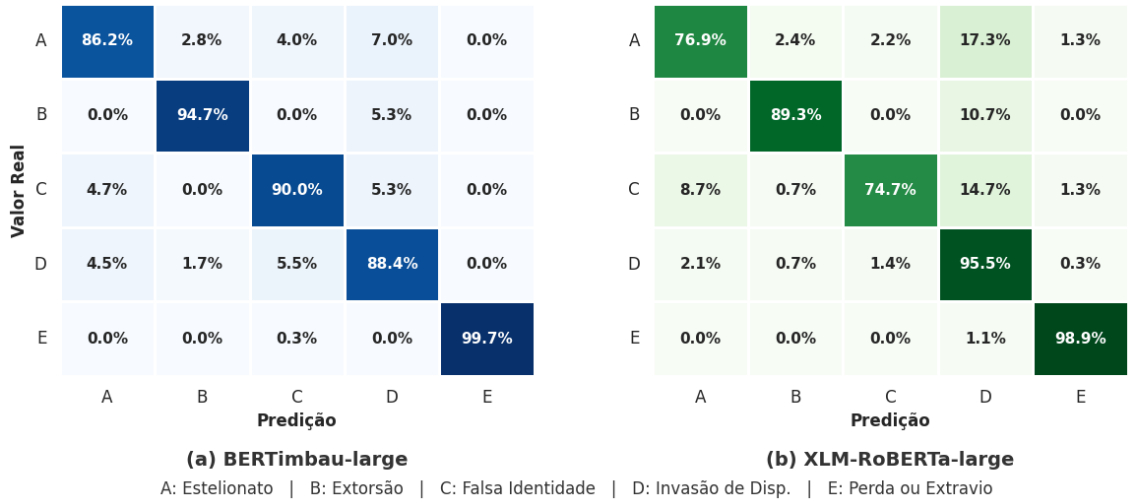

Figura 2. Matrizes de confusão no conjunto desbalanceado.

5.3. Análise Comparativa dos Cenários

A comparação entre os dois cenários de avaliação revela aspectos importantes sobre a robustez e o comportamento dos modelos:

Consistência do BERTimbau-large: O modelo monolíngue manteve desempenho notavelmente estável entre os cenários balanceado (89,87%) e desbalanceado (89,51%), com queda de apenas 0,36 pontos percentuais. Esta consistência demonstra uma alta robustez de generalização: mesmo tendo sido treinado em dados balanceados, o modelo é capaz de manter seu desempenho ao ser avaliado em um cenário desbalanceado, sugerindo que suas regras de classificação aprendidas são mais gerais.

Sensibilidade do XLM-RoBERTa-large: O modelo multilíngue apresentou maior variação, com acurácia de 86,13% no cenário balanceado e 83,70% no desbalanceado, indicando menor robustez à mudança de distribuição. A performance do XLM-R degradou mais visivelmente quando a composição do conjunto de teste diferiu daquela vista durante o treino.

Comportamento por classe: Observam-se padrões distintos:

- **Estelionato** (classe majoritária): Ambos os modelos mantiveram alta precisão (0,98), porém o BERTimbau apresentou recall superior (0,86 vs 0,77), resultando em melhor F1-Score (0,92 vs 0,86).
- **Extorsão e Falsa Identidade** (classes minoritárias): Ambos os modelos convergiram para F1-Score de 0,77, evidenciando desafios similares nestas categorias

com menor representação.

- **Invasão de Dispositivo Informático:** Observou-se comportamento complementar entre os modelos — o XLM-RoBERTa priorizou recall elevado (0,96) às custas de precisão reduzida (0,54), enquanto o BERTimbau manteve equilíbrio superior (precisão: 0,73, recall: 0,88).
- **Perda ou Extravio:** O BERTimbau alcançou desempenho perfeito (1,00), enquanto o XLM-RoBERTa manteve desempenho próximo ao cenário balanceado (0,97).

A análise das matrizes de confusão revela que o XLM-RoBERTa-large tende a classificar mais instâncias nas categorias "Estelionato" e "Invasão de Dispositivo Informático", aumentando o recall destas classes mas comprometendo a precisão. O BERTimbau-large demonstra distribuição de erros mais equilibrada.

5.4. Análise de Erros

A análise dos erros de classificação em ambos os cenários revela que as confusões mais frequentes ocorrem entre classes semanticamente relacionadas e que frequentemente concorrem no mundo real: "Estelionato", "Falsa Identidade" e "Invasão de Dispositivo Informático". Essa sobreposição reflete a própria estrutura típica dos delitos cibernéticos, nos quais um crime principal, como o estelionato, é frequentemente viabilizado por crimes-meio, como invasão de dispositivo informático ou falsa identidade digital. A Figura 2 ilustra esses padrões de forma clara.

Tal interpenetração entre condutas evidencia uma das maiores dificuldades na classificação automatizada desses ilícitos, uma vez que, mesmo para o intérprete jurídico, a delimitação entre o tipo penal principal e os crimes instrumentais pode apresentar margens de subjetividade, exigindo a análise do contexto fático e probatório integral da investigação. Em contraste, a classe "Perda ou Extravio de Documento e/ou Objeto" apresentou desempenho excepcional em ambos os cenários devido ao seu vocabulário característico e ausência de sobreposição semântica com crimes intencionais, facilitando sua identificação pelos modelos.

6. Conclusão

Este trabalho mostrou que é viável realizar *fine-tuning* de modelos Transformer para classificar automaticamente ocorrências policiais em português brasileiro. A avaliação nos cenários balanceado e desbalanceado evidenciou tanto a capacidade discriminativa quanto a robustez operacional dos modelos.

O BERTimbau-large superou o XLM-RoBERTa-large em ambos os cenários, mantendo consistência (89,87% vs 89,51% de acurácia) e evidenciando ausência de viés significativo para classes majoritárias. O modelo monolíngue alcançou desempenho perfeito em "Perda ou Extravio de Documento" (F1=1,00), enquanto as confusões entre "Estelionato", "Falsa Identidade" e "Invasão de Dispositivo" refletem a interpenetração conceitual inerente aos crimes cibernéticos. Em contraste, o XLM-RoBERTa demonstrou *trade-off* desfavorável entre precisão e recall, classificando 71% dos casos de "Invasão de Dispositivo" como "Estelionato".

Os resultados indicam que modelos especializados em português brasileiro constituem abordagem promissora para automação da triagem na SSP-PI, com potencial para

redução de carga operacional. Como trabalhos futuros, sugere-se: (i) arquiteturas hierárquicas para relações crime-meio/crime-fim; (ii) sistemas de classificação multi-rótulo; (iii) avaliação de transferibilidade entre unidades federativas; e (iv) análise de termos discriminativos e elementos constitutivos do delito para maior interpretabilidade jurídica.

Agradecimentos

O presente trabalho foi financiamento parcialmente pela SSP/PI, com apoio da fundação Cultural e de Fomento à Pesquisa, Ensino, Extensão e Inovação (FADEX). Agradecemos também a equipe técnica da SSP/PI - Rogerio Paulo, Venceslau Felipe, Joaquim Carvalho e aos demais colegas - pelo apoio e disponibilidade constante em conversar com os pesquisadores do Departamento de Computação da UFPI.

Referências

Alves, D., Marques, M., Santos, R., and Santos, A. (2024). Classificaçao de boletins de ocorrências através de modelos de linguagem baseados em bert. In *Workshop de Computação Aplicada em Governo Eletrônico (WCGE)*, pages 169–179. SBC.

Amorim, M. d. S. and Pereira, J. R. S. (2019). Tipificação de ocorrências policiais utilizando machine learning.

Araújo, J. A. F. (2023). Busca por similaridade de boletins de ocorrência via embeddings: um estudo de caso.

Cer, D., Yang, Y., yi Kong, S., Hua, N., Limtiaco, N., John, R. S., Constant, N., Guajardo-Cespedes, M., Yuan, S., Tar, C., Sung, Y.-H., Strope, B., and Kurzweil, R. (2018). Universal sentence encoder.

Conneau, A., Khandelwal, K., Goyal, N., Chaudhary, V., Wenzek, G., Guzmán, F., Grave, E., Ott, M., Zettlemoyer, L., and Stoyanov, V. (2020). Unsupervised cross-lingual representation learning at scale.

de Castro, U. R. M. (2020). Explorando aprendizagem supervisionada em dados heterogêneos para predição de crimes.

de Sousa, R. F. and Moura, R. S. (2025). Avaliação preliminar de técnicas de pln para classificação de relatos em boletins de ocorrência policial. In *Encontro Unificado de Computação do Piauí (ENUCOMPI)*, pages 79–88. SBC.

Lin, T.-Y., Goyal, P., Girshick, R., He, K., and Dollár, P. (2017). Focal loss for dense object detection. In *Proceedings of the IEEE international conference on computer vision*, pages 2980–2988.

Matos, H., Souza, S., Santos, R., Costa, J. W., and Costa, C. (2022). A supervised classifier for police reports at the state of pará, brazil. In *ERAD-NO2 e ERAMIA-NO2*.

Reimers, N. and Gurevych, I. (2019). Sentence-bert: Sentence embeddings using siamese bert-networks.

Souza, F., Nogueira, R., and Lotufo, R. (2020). Bertimbau: pretrained bert models for brazilian portuguese. In *Brazilian conference on intelligent systems*. Springer.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., and Polosukhin, I. (2017). Attention is all you need. *Advances in neural information processing systems*, 30.