

Classificação de Boletins de Ocorrências através de Modelos de Linguagem baseados em BERT

Dhênnifer Alves¹, Marcílio Marques², Reginaldo Santos³, Adam Santos^{1,2}

¹Faculdade de Sistemas de Informação
Universidade Federal do Sul e Sudeste do Pará (UNIFESSPA)
Marabá - PA - Brasil

²Programa de Pós-Graduação em Ciências Forenses
Universidade Federal do Sul e Sudeste do Pará (UNIFESSPA)
Marabá - PA - Brasil

³Faculdade de Computação - Universidade Federal do Pará (UFPA)
Belém - PA - Brasil

{dhennifer, marciliiodsm, adamdreyton}@unifesspa.edu.br, regicsf@ufpa.br

Abstract. *The Public Security sector has adopted trends in the application of data science and mining, driven by the volume of data generated daily and the automation and improvement of internal processes. This study proposes the application of language models based on BERT to classify crimes in police reports in Marabá, Pará. The results highlighted that BERT and RoBERTa achieved the best performances, with accuracy between 89% and 90% for ten classes related to more frequent crimes. This classification automation can speed up the process of consolidating police reports, reducing the need for manual work.*

Resumo. *O setor da Segurança Pública tem adotado as tendências de aplicação de ciência e mineração de dados, impulsionado pelo volume de dados gerados diariamente e pela automatização e aprimoramento dos processos internos. Este estudo propõe a aplicação de modelos de linguagem baseados em BERT para classificar crimes em relatos de boletins de ocorrências em Marabá, Pará. Os resultados mostraram que BERT e RoBERTa alcançaram as melhores performances, com acurácia entre 89% e 90% para dez classes relacionadas com crimes de maior ocorrência. Tal automação da classificação pode agilizar o processo de consolidação de boletins de ocorrências, reduzindo a necessidade de trabalho manual.*

1. Introdução

A Segurança Pública é uma das áreas de maior interesse para a administração pública do Brasil. Logo, é de extrema importância ter dados e informações criminais bem organizados, padronizados e disponíveis, tarefa esta de responsabilidade do Sistema Nacional de Informações de Segurança Pública, Prisionais, de Rastreabilidade de Armas e Munições, de Material Genético, de Digitais e de Drogas (SINESP) [BRASIL 2019]. O SINESP auxilia a comunicação entre os entes do Sistema Único de Segurança Pública (SUSP). Cada estado do Brasil está encarregado pela coleta, normalização e disponibilização dos dados extraídos das delegacias policiais. Essa integração de informações demanda especial esforço, por exemplo, nos interiores, onde tais sistemas tendem a falhar devido a falta

de manutenção, delegacias computadorizadas ou profissionais capacitados a cadastrar e rotular um grande volume de registros policiais.

A crescente quantidade de dados gerados por registros policiais trouxe a necessidade da digitalização de processos no setor da Segurança Pública. Paralelamente, a aplicação de soluções de ciência de dados e aprendizado de máquina emergiu como uma tendência a ser explorada nesse setor. A Segurança Pública tem potencial de ser beneficiada por ferramentas de extração de conhecimento em bases de dados, incluindo a classificação de relatos policiais [MATOS 2022], a descoberta de regras associativas [SOUZA 2022], visualização da informação, etc.

Neste contexto, este trabalho propõe a classificação supervisionada de relatos policiais via modelos de linguagem baseados em *Bidirectional Encoder Representations from Transformers* (BERT), usando dados coletados junto a Secretaria Adjunta de Inteligência e Análise Criminal (SIAC), uma instituição vinculada à Secretaria de Segurança Pública e Defesa Social (SEGUP) do Estado do Pará, Brasil. Especificamente, os dados dos dez crimes com mais ocorrências na Cidade de Marabá, Pará, serão empregados para os treinamentos, validações e testes dos referidos modelos.

Este estudo está organizado tal que: Na Seção 2, os trabalhos mais relacionados a este estudo são evidenciados; A metodologia empregada neste trabalho é detalhada na Seção 3; Na Seção 4, os resultados são apresentados e analisados considerando três modelos de linguagem baseados em BERT; Finalmente, na Seção 5, as conclusões derivadas deste trabalho são sintetizadas e os possíveis trabalhos futuros que poderão advir deste estudo são elencados.

2. Trabalhos Relacionados

Na literatura, diversos estudos podem ser encontrados acerca do processamento e análise de dados de crimes, considerando técnicas que auxiliam na extração de padrões e na classificação de diferentes crimes, de forma que essas metodologias possam promover ações de prevenção e combate ao crime, além de auxiliar na formulação de políticas públicas para área de Segurança Pública.

Uma metodologia de predição da tendência e número de ocorrências de tipos de crimes por regiões geográficas é evidenciada em [CASTRO 2020], buscando elucidar o padrão comportamental criminal. No referido estudo, algoritmos de aprendizado de máquina, como *Long Short Term Memory* (LSTM) e *eXtreme Gradient Boosting*, são empregados em bases de dados de furto e roubo no Estado de Minas Gerais. Os resultados experimentais revelaram que o algoritmo LSTM obteve o melhor desempenho, no que tange a acurácia de 91% na predição.

Através de um estudo quantitativo em bases de dados de boletins de ocorrências policiais dos anos de 2017, 2018 e 2019, englobando os crimes de furto, roubo, homicídio, lesão corporal e outros, Regateiro [REGATEIRO 2021] apresenta uma metodologia baseada na estimação de quatro índices de criminalidade, para avaliar a criminalidade no Estado do Pará e na Cidade de Belém/PA. Os índices propostos estão relacionados com médias bayesianas aplicadas por município e bairro. Os experimentos conduzidos nesses dados mostraram que sete bairros estavam classificados como de alta criminalidade ao longo dos três anos de estudo, destacando que nesses bairros haveriam condições proeminentes para o crescimento da criminalidade.

Os trabalhos de Souza *et al.* [SOUZA et al. 2022a, SOUZA et al. 2022b] consideraram crimes violentos (mortes, roubos e furtos), para a aplicação de ferramentas de ciência e mineração de dados em bases públicas de segurança provenientes da SIAC (Estado do Pará), ao longo dos anos de 2019, 2020 e 2021, incluindo boletins de ocorrências policiais. Com o emprego da metodologia CRISP-DM, uma análise exploratória dos dados e a geração de regras associativas foram realizadas, subsidiando as autoridades em processos de tomada de decisão através de índices criminais no período especificado. As evidências da análise exploratória mostraram, por exemplo, que houveram decréscimos de cerca de 41% dos furtos e de 25% dos roubos para o período entre 2019 e 2020 (pandemia de COVID-19). Os autores evidenciaram que os oito primeiros meses da pandemia no Brasil (abril a dezembro/2020) possuem correlação com baixos índices de registros de crimes violentos. Além disso, os furtos tem alta incidência durante a manhã, enquanto roubos e homicídios são mais predominantes durante a noite. Em relação às regras associativas, destacam-se as correlações do crime de furto com: ausência de instrumento empregado para o desenvolvimento da ação criminosa e tendência das vítimas serem homens entre 35 e 64 anos.

Os trabalhos relacionados apresentados anteriormente aplicaram técnicas de mineração de dados e aprendizado de máquina em dados de ocorrências policiais. Especificamente, estudos como os conduzidos por [MATOS 2022, SOUZA 2022] que classificaram crimes por meio da leitura de um relato (*i.e.*, boletim de ocorrência), concentraram-se em algoritmos tradicionais de aprendizado de máquina, enquanto neste trabalho está sendo proposta a utilização de modelos de linguagem baseados em BERT, os quais demonstraram-se mais apropriados para processar, aprender e generalizar padrões em dados de relatos contidos em boletins de ocorrências. Ademais, o mesmo recorte temporal dos dados utilizados nos estudo de Souza *et al.* [SOUZA et al. 2022a, SOUZA et al. 2022b] foi assumido, porém considerando os dez crimes com maior ocorrência na Cidade de Marabá/PA.

3. Classificação de Boletins de Ocorrências

Nesta seção, serão abordados o pré-processamento e a classificação dos boletins de ocorrências utilizando técnicas de mineração de dados e aprendizado de máquina. Especificamente, serão descritos os processos de seleção e limpeza dos dados, visando garantir sua qualidade e relevância. Posteriormente, será detalhada a metodologia de classificação utilizada, incluindo divisão da base de dados, algoritmos empregados e métricas de avaliação.

3.1. Dataset

Os dados utilizados neste estudo foram disponibilizados pela SIAC (Estado do Pará), a partir de um conjunto de registros policiais dos anos 2019, 2020 e 2021. O conjunto de dados original contém 1.450.999 amostras e 92 atributos em arquivos tabulares com tamanho total de 1,06 GB. Os atributos abrangem uma variedade de informações, incluindo identificação de registro, informações temporais e de localização, descrição do *modus operandi*, detalhes sobre as vítimas e autores, além do relato descrevendo o evento. Especificamente, em relação à tipologia criminal, dois atributos são destacados:

- “registros”: referem-se às classes de eventos atribuídas nas delegacias no momento do registro do boletim de ocorrência. No entanto, devido ao caráter emer-

gencial do preenchimento, nem sempre reflete precisamente o evento descrito no relato. Essas classes podem incluir referências ao Código Penal Brasileiro (por exemplo, “ART. 147 - AMEAÇA”), e também subconjuntos com erros ou duplicidades de grafia que podem ser agrupadas em uma única classe, por exemplo, “ART. 157 - FURTO”, “ARTIGO 157 - FURTO”, “FURTO”, entre outros.

- “consolidado”: representa a classe de eventos atribuídos na base de dados da SIAC, considerada uma alternativa mais precisa e confiável em comparação com as classes de registros. Essas classes são utilizadas principalmente na geração de relatórios estatísticos para gestores da Segurança Pública no Estado do Pará, mídia interessada em dados quantitativos e nos portais de transparência da Secretaria.

Internamente, o atributo “consolidado” surge como uma solução para a baixa consistência dos registros. O departamento de estatística da SIAC utiliza um grupo com um pouco mais de uma dezena de analistas criminais para ler e rotular os relatos, gerando as classes de “consolidados” com base em seus conhecimentos empíricos da legislação brasileira. Essa análise se concentra principalmente em crimes violentos, devido ao grande volume de dados gerados diariamente nas delegacias. Cerca de 10% das classes de registro não são compatíveis com as classes finais de “consolidados”, o que indica uma margem de erro, a qual é o foco principal para melhoria da qualidade dos dados.

Dada a necessidade de uma ferramenta automatizada para confirmar as classes de eventos criminais e processar o volume massivo de registros policiais diários, juntamente com a necessidade de definir um conjunto descritivo das classes de relatos para fins de pesquisa, o atributo “consolidado” foi escolhido como a coluna-alvo.

3.2. Seleção de Dados e Limpeza

Devido à grande quantidade de dados e à presença de informações irrelevantes ou duplicadas, foi necessário realizar uma cuidadosa seleção e limpeza dos dados para garantir a qualidade e relevância para os objetivos deste trabalho, facilitando a aplicação da mineração de dados.

Inicialmente, foi aplicado um filtro para a Cidade de Marabá/PA, selecionando os dez maiores grupos de crime registrados, ordenados de forma decrescente conforme mostra a Tabela 1. Em seguida, foram avaliados dois atributos como importantes para o desenvolvimento do classificador, sendo eles:

- “relato”: descrição completa dos acontecimentos, sendo a fonte primária para extração de conhecimento do *dataset*.
- “consolidado”: classe atribuída pelo analista criminal após ler o relato. É a propriedade alvo para os problemas de classificação, pois contém uma descrição precisa e concisa dos relatos da base de dados. A base de dados original contém 977 classes, com distribuição desbalanceada, mas neste trabalho foram utilizadas apenas as dez classes com mais ocorrências para Cidade de Marabá/PA.

Ao concluir a seleção, temos 33.422 registros e 2 atributos.

Na etapa de limpeza, os dados foram processados para garantir a qualidade e a viabilidade do processamento e análise de dados. Essa etapa englobou as ações de:

- Remoção de amostras duplicadas: ao analisar a base de dados, identificamos e removemos valores duplicados usando o campo “nro_bop” como identificador

Tabela 1. Classes e quantitativos dos dez crimes mais comuns em Marabá/PA, de 2019 a 2021.

Classes de Crime	Quantidade de Registros
FURTO	10717
ROUBO	6710
DANO NO TRÂNSITO	3837
AMEAÇA	3169
LESÃO CORPORAL	2397
ESTELIONATO	1983
OUTROS FATOS ATÍPICOS	1905
LESÃO NO TRÂNSITO	1253
ACIDENTE DE TRÂNSITO SEM VÍTIMA	751
ROUBO DE VEÍCULO	700

único dos boletins de ocorrências. Isso nos permite preservar apenas a primeira ocorrência de cada evento, considerando também o campo “relato” para garantir a integridade dos dados. Esse procedimento é crucial para garantir a precisão das análises subsequentes e fornecer uma base confiável para tomadas de decisão.

- Remoção de caracteres especiais e pontuações: pontuação e caracteres especiais nos relatos são removidos, simplificando o conteúdo textual e tornando sua análise mais fácil.
- Remoção de tags HTML: os relatos na base de dados original podem incluir elementos de marcação HTML provenientes da página da web, onde o sistema está hospedado. Essas marcações são indesejadas e podem introduzir ruído nos dados. Utilizando expressões regulares, essas tags HTML são identificadas e removidas, preservando apenas o conteúdo textual relevante dos relatos.

Ao final da limpeza, os dados processados contém 21.786 registros e 2 atributos.

Durante o processo de limpeza, foi também realizado a conversão dos caracteres dos relatos para minúsculo, assegurando representação uniforme no texto. Após todo o processo de seleção e limpeza, os dados foram armazenados e transformados em um *Huggingface dataset* (<https://huggingface.co/>), que é compatível com o modelo BERT. Essa ação facilitou todo o processo de pré-processamento, treinamento e avaliação dos modelos de linguagem, assegurando resultados mais consistentes e confiáveis em análises e experimentos.

3.3. Classificação e Métricas

A base de dados pós-processada foi dividida em três subconjuntos não-estratificados: 80% para treino, 10% para validação e 10% para o teste. Para o processo de classificação, foram usados o algoritmo BERT e mais duas variantes, sendo elas: RoBERTa e Albert. Uma vez que não foram realizadas modificações significativas nos algoritmos de classificação neste estudo, esses modelos de linguagem são brevemente discutidos a seguir.

- *Bidirectional Encoder Representations from Transformers* (BERT): desenvolvido por Jacob Devlin e sua equipe na *Google AI Language* em 2018 [DEVLIN et al. 2018], é um modelo de linguagem que introduziu a abordagem

- bidirecional para o pré-treinamento de representações de linguagem natural. Treinado em grande quantidade de texto não rotulado, o BERT aprende contextos à esquerda e à direita de cada palavra, capturando nuances semânticas e sintáticas.
- *Robustly optimized BERT approach* (RoBERTa): desenvolvido pela equipe do Facebook AI em 2019 [LIU et al. 2019], o RoBERTa é uma variação do BERT que otimiza o pré-treinamento na tarefa de preenchimento de lacunas. Com técnicas como treinamento em larga escala e em sequências mais longas, o RoBERTa supera o BERT em várias tarefas de processamento de linguagem, enfatizando a robustez e a eficácia em uma variedade de contextos. Por exemplo, em comparação com o BERT, o RoBERTa demonstrou melhorias significativas em tarefas como modelagem de linguagem, classificação de texto e tradução automática, devido à sua capacidade de capturar relações mais complexas entre as palavras e entender contextos mais amplos.
 - *A Lite BERT* (Albert): é uma criação da *Google Research* em 2019 [LAN et al. 2019], sendo uma versão eficiente do BERT que visa reduzir o número de parâmetros do modelo. Isso é alcançado através de técnicas como o compartilhamento de parâmetros entre camadas e fatores de escala de *embedding*, resultando em uma arquitetura mais compacta e escalável. Embora mantenha a qualidade das representações de linguagem, o Albert é projetado para ser mais leve e adequado para ambientes com recursos limitados.

Os três modelos foram executados com parâmetros de treinamento iguais, sendo eles: “tamanho do lote para avaliação” igual a 1, “taxa de decaimento” igual a 0,01 e “número de épocas de treinamento” igual a 3.

Para avaliação e verificação do desempenho dos resultados encontrados com as técnicas de classificação, foram usadas as seguintes métricas: *precision*, *recall*, acurácia, e *f1-score* [PADUA 2020]. Tais métricas são amplamente conhecidas e confiáveis na literatura de computação no que tange principalmente às tarefas de classificação.

4. Resultados

Os modelos de linguagem usados para classificação (ver Subseção 3.3), foram implementados na linguagem Python, versão 3.11.5, com as bibliotecas PyTorch [PASZKE et al. 2019] e Scikit-Learn [PEDREGOSA et al. 2011]. Todos os experimentos computacionais deste trabalho foram executados em um computador com o processador AMD Ryzen 7 3700X, 8 núcleos de 4.4GHz, e 64GB de memória RAM. Além disso, para aproveitar o poder computacional da placa de vídeo Nvidia RTX 3060 de 12GB disponível nesse sistema, os algoritmos foram executados na GPU, utilizando as capacidades de processamento paralelo para acelerar o treinamento e a inferência dos modelos.

Os algoritmos de classificação baseados em BERT foram configurados para executar durante três épocas, considerando pré-treinamentos em bases de dados disponíveis no *Hugging Face* (<https://huggingface.co/>), especificamente “bert-base-uncased”, “roberta-base” e “albert-base-v2”. Os resultados apresentados a seguir assumem médias de vinte execuções de cada algoritmo.

Os resultados da etapa de teste com os modelos BERT, RoBERTa e Albert, são apresentados, respectivamente, nas Tabelas 2, 3 e 4. Analisando os resultados obtidos pelos modelos BERT, RoBERTa e ALBERT, é possível identificar padrões interessantes

em relação ao desempenho de cada modelo em diferentes classes de crimes. Inicialmente, é evidente que para as classes “dano no trânsito”, “furto” e “roubo”, todos os modelos obtiveram um desempenho consistente, apresentando tanto *precision* quanto *recall* em níveis satisfatórios. Isso sugere que as classes estão bem representadas nos dados e são distintas o suficiente para serem corretamente identificadas pelos modelos, independentemente de suas arquiteturas específicas.

Tabela 2. Resultados da etapa de teste com o modelo BERT.

Classe	<i>precision</i>	<i>recall</i>	<i>f1-score</i>
acidente de trânsito sem vítima	0,62	0,86	0,72
ameaça	0,87	0,74	0,80
dano no trânsito	1,00	0,95	0,97
estelionato	0,89	0,93	0,91
furto	0,95	0,95	0,95
lesão corporal	0,72	0,89	0,80
lesão no trânsito	0,93	0,88	0,91
outros fatos atípicos	0,61	0,50	0,55
roubo	0,92	0,95	0,94
roubo de veículo	0,71	0,62	0,67
Média	0,82	0,83	0,82

Além disso, foi observado que o modelo RoBERTa teve um desempenho aceitável principalmente para classes com um número maior de registros. Isso pode indicar que o RoBERTa tende a se destacar em classes mais comuns ou bem representadas nos dados de treinamento, enquanto pode ter dificuldades com classes menos frequentes.

Por outro lado, o modelo Albert, apesar de ter uma quantidade considerável de registros disponíveis em algumas classes, mostrou uma precisão abaixo do esperado em algumas classes. Isso sugere que o desempenho do Albert pode ser influenciado por fatores como o número de épocas ou a complexidade da arquitetura do modelo.

Tabela 3. Resultados da etapa de teste com o modelo RoBERTa.

Classe	<i>precision</i>	<i>recall</i>	<i>f1-score</i>
acidente de trânsito sem vítima	0,46	0,70	0,56
ameaça	0,91	0,86	0,88
dano no trânsito	0,99	0,92	0,95
estelionato	0,83	0,88	0,85
furto	0,97	0,96	0,96
lesão corporal	0,83	0,88	0,85
lesão no trânsito	0,81	0,80	0,80
outros fatos atípicos	0,58	0,55	0,57
roubo	0,95	0,95	0,95
roubo de veículo	0,70	0,86	0,77
Média	0,8	0,84	0,81

Uma possível explicação para o desempenho abaixo do esperado do Albert,

em comparação com o RoBERTa, pode ser atribuída à sua estratégia de redução de parâmetros, que visa tornar o modelo mais eficiente e escalável. No entanto, a redução de parâmetros pode levar a uma representação menos expressiva dos dados, especialmente em classes menos representadas ou com características complexas [LAN et al. 2019]. Ademais, o compartilhamento de parâmetros entre camadas e fatores de escala de *embedding* implementados no Albert, podem influenciar sua capacidade de capturar nuances sutis nos dados, o que pode afetar seu desempenho em tarefas de classificação de texto mais desafiadoras.

Um resultado que merece destaque foi observado na classe “acidente de trânsito sem vítima”, onde o modelo BERT foi o único a apresentar um desempenho satisfatório, enquanto RoBERTa e Albert tiveram precisões significativamente mais baixas, sendo 0,46 e 0,43, respectivamente. Isso sugere que o BERT pode ser mais adequado para identificar essa classe específica com base nas características dos dados ou na arquitetura do modelo, isto é, considerando as três épocas de aprendizado e validação dos modelos, o BERT não performou de forma inferior quando comparado com suas variantes, fato corroborado pela Tabela 5.

Tabela 4. Resultados da etapa de teste com o modelo Albert.

Classe	<i>precision</i>	<i>recall</i>	<i>f1-score</i>
acidente de trânsito sem vítima	0,43	0,53	0,47
ameaça	0,80	0,86	0,82
dano no trânsito	1,00	0,94	0,97
estelionato	0,88	0,83	0,86
furto	0,93	0,93	0,93
lesão corporal	0,78	0,76	0,77
lesão no trânsito	0,85	0,82	0,83
outros fatos atípicos	0,43	0,46	0,44
roubo	0,91	0,93	0,92
roubo de veículo	0,31	0,24	0,27
Média	0,73	0,73	0,73

Tabela 5. Resultados de acurácia no processo de classificação obtidos pelos algoritmos Albert, BERT e RoBERTa.

Modelo	<i>Treino</i>	<i>Validação</i>	<i>Teste</i>
Albert	0,88	0,86	0,86
BERT	0,92	0,89	0,89
RoBERTa	0,92	0,90	0,90

A análise dos resultados de acurácia, conforme demonstrado na Tabela 5, juntamente com as características distintas de cada arquitetura, permite uma compreensão mais profunda do desempenho dos modelos. A arquitetura do modelo Albert, com ênfase no compartilhamento e redução de parâmetros, pode ter contribuído para uma leve redução na acurácia em comparação com BERT e RoBERTa. Em contraste, as arquiteturas mais tradicionais do BERT e do RoBERTa podem ter permitido uma representação mais rica dos dados, refletindo-se em uma acurácia superior. Essas particularidades destacam a

importância de considerar não apenas os resultados numéricos, mas também as características específicas de cada modelo ao interpretar os resultados.

5. Conclusões

O avanço da criminalidade no Brasil é um problema que pode ensejar elevados prejuízos para a economia e qualidade de vida da população. Logo, na última década, observaram-se grandes investimentos no setor de Segurança Pública, de forma que ações e políticas públicas possam ser criadas para superar esse obstáculo no caminho do desenvolvimento do país. Entretanto, muito do trabalho que é realizado, por exemplo, em delegacias, ainda é pouco apoiado por tecnologias da informação. Há grande diversidade de dados disponíveis a partir de registros policiais, que poderiam ser utilizados para tomadas de decisões em prol da população no que tange o combate ao crime.

Este trabalho buscou então automatizar a classificação de crimes a partir da aplicação de modelos de linguagem baseados em BERT nos dados extraídos de boletins de ocorrências policiais dos anos de 2019, 2020 e 2021, fornecidos pela SIAC (Estado do Pará). Especificamente, os dez crimes mais frequentes na Cidade de Marabá/PA, como furto, roubo, dano no trânsito, ameaça, lesão corporal, etc.

Os resultados experimentais demonstraram que os modelos de linguagem baseados em BERT são promissores para classificar e generalizar variados crimes, com nível de acurácia factível, variando entre 86% e 90%. Destaca-se que os melhores resultados obtidos, de forma geral e em análise por classe, foram provenientes dos modelos BERT e RoBERTa. O desempenho abaixo do esperado, alcançado pelo modelo Albert, pode ser justificado pela necessidade de um maior número de épocas para treinamento e validação.

É imperativo frisar que a utilização dos modelos propostos neste trabalho para classificação de crimes, a partir de relatos de boletins de ocorrências, pode automatizar o processo de consolidação de boletins de ocorrências, isto é, reduzir a força de trabalho humana empregada entre o horário de início do processamento do boletim e a atribuição da classe correta a qual pertence aquele relato.

Como trabalhos futuros que poderão derivar deste estudo, destacam-se: (i) Mais experimentos, considerando médias de execuções, do modelo de linguagem BERT e as variantes destacadas neste estudo, aumentando o número de épocas para treinamento e validação; (ii) Experimentos com outras variantes do modelo de linguagem BERT não consideradas neste trabalho, como BERTimbau, ELECTRA, DistilBERT, SpanBERT, e TinyBERT; (iii) Aumentar a base de dados, para novos experimentos, considerando mais classes de crimes com maiores ocorrências na Cidade de Marabá e/ou solicitar a SIAC (Estado do Pará) dados dos anos de 2022 e 2023.

Referências

- BRASIL (2019). O Sistema Nacional de Informações de Segurança Pública, Prisionais, de Rastreabilidade de Armas e Munições, de Material Genético, de Digitais e de Drogas (Sinesp). Disponível em: <https://www.gov.br/mj/pt-br/assuntos/sua-seguranca/seguranca-publica/sinesp-1/>.
- CASTRO, U. R. M. (2020). Explorando aprendizagem supervisionada em dados heterogêneos para predição de crimes. Dissertação de Mestrado, Programa de Pós-Graduação em Informática, Pontifícia Universidade Católica de Minas Gerais.

- DEVLIN, J., Chang, M., Lee, K., and Toutanova, K. (2018). BERT: pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, Minneapolis, USA.
- LAN, Z., Chen, M., Goodman, S., Gimpel, K., Sharma, P., and Soricut, R. (2019). ALBERT: a lite BERT for self-supervised learning of language representations. In *Proceedings of the 37th International Conference on Machine Learning (ICML)*, Vienna, Austria.
- LIU, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L., and Stoyanov, V. (2019). RoBERTa: a robustly optimized bert pretraining approach. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, Hong Kong, China.
- MATOS, H. M. R. (2022). Um Classificador Supervisionado para Relatos Policiais no Estado do Pará. Trabalho de Conclusão de Curso, Curso de Ciência da Computação, Universidade Federal do Pará.
- PADUA, M. (2020). Machine Learning - Métricas de avaliação: Acurácia, Precisão e Recall, F1-score. Disponível em: <https://medium.com/@mateuspdua/machine-learning-m%C3%A9tricas-de-avalia%C3%A7%C3%A3o-acur%C3%A1cia-precis%C3%A3o-e-recall-d44c72307959> (Acesso em: 02/04/2024).
- PASZKE, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., Desmaison, A., Kopf, A., Yang, E., DeVito, Z., Raison, M., Tejani, A., Chilamkurthy, S., Steiner, B., Fang, L., Bai, J., and Chintala, S. (2019). PyTorch: An Imperative Style, High-Performance Deep Learning Library. In Wallach, H., Larochelle, H., Beygelzimer, A., d'Alché-Buc, F., Fox, E., and Garnett, R., editors, *Advances in Neural Information Processing Systems 32*, pages 8024–8035. Curran Associates, Inc.
- PEDREGOSA, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830.
- REGATEIRO, H. A. S. (2021). Avaliação da criminalidade em Belém e no Estado do Pará. Dissertação de Mestrado, Programa de Pós-Graduação em Segurança Pública, Universidade Federal do Pará, Universidade Federal do Pará.
- SOUZA, S. (2022). Mineração de Dados em Bancos de Dados de Segurança Pública no Estado do Pará. Dissertação de Mestrado, Programa de Pós-Graduação em Ciência da Computação, Universidade Federal do Pará.
- SOUZA, S., Matos, H., Costa, C., and Filho, R. S. (2022a). Data Mining in Public Security Databases in Belém, Pará, Brazil. *Anais da Escola Regional de Alto Desempenho Norte 2 (ERAD-NO2) e Escola Regional de Aprendizado de Máquina e Inteligência Artificial Norte 2 (ERAMIA-NO2)*, pages 33–36.

SOUZA, S., Matos, H., Filho, R., and Costa, J. (2022b). Data Mining and Analysis Applied to Public Security Data in Belém of Pará, Brazil. *CONTECSI USP - International Conference on Information Systems and Technology Management - ISSN 2448-1041*.