

UNIVERSIDADE DE SÃO PAULO

Instituto de Ciências Matemáticas e de Computação

**Agregando e analisando comentários
georreferenciados de usuários para identificação e
delimitação de regiões de risco em aplicativos de
mobilidade urbana**

Jacqueline Eduardo Barbosa Valentin

Monografia - MBA em Inteligência Artificial e Big Data

SERVIÇO DE PÓS-GRADUAÇÃO DO ICMC-USP

Data de Depósito:

Assinatura: _____

Jacqueline Eduardo Barbosa Valentin

Agregando e analisando comentários georreferenciados de usuários para identificação e delimitação de regiões de risco em aplicativos de mobilidade urbana

Monografia apresentada ao Departamento de Ciências de Computação do Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo - ICMC/USP, como parte dos requisitos para obtenção do título de Especialista em Inteligência Artificial e Big Data.

Área de concentração: Inteligência Artificial

Orientadora: Profa. Dra. Mirela Teixeira Cazzolato

Versão original

São Carlos

2024

Ficha catalográfica elaborada pela Biblioteca Prof. Achille Bassi
e Seção Técnica de Informática, ICMC/USP,
com os dados inseridos pelo(a) autor(a)

E24a Eduardo Barbosa Valentin, Jacqueline
Agregando e analisando comentários
georreferenciados de usuários para identificação e
delimitação de regiões de risco em aplicativos de
mobilidade urbana / Jacqueline Eduardo Barbosa
Valentin; orientadora Mirela Teixeira Cazzolato . --
São Carlos, 2024.
55 p.

Trabalho de conclusão de curso (MBA em
Inteligência Artificial e Big Data) -- Instituto de
Ciências Matemáticas e de Computação, Universidade
de São Paulo, 2024.

1. Análise de Sentimentos . 2. Segurança Pública
. 3. Mobilidade Urbana. 4. Georreferenciamento. 5.
Processamento da Linguagem Natural. I. Teixeira
Cazzolato , Mirela, orient. II. Título.

Jacqueline Eduardo Barbosa Valentin

**Aggregating and analyzing georeferenced user comments
for identifying and delimiting risk areas in urban mobility
applications**

Monograph presented to the Departamento de Ciências de Computação do Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo - ICMC/USP, as part of the requirements for obtaining the title of Specialist in Artificial Intelligence and Big Data.

Concentration area: Artificial Intelligence

Advisor: Profa. Dra. Mirela Teixeira Cazzolato

Original version

São Carlos

2024

Este trabalho é dedicado ao meu filho Samuel, meu milagre, que foi gerado no meio deste curso e que lutou pela vida desde a concepção até o nascimento.

Meu filho, que você tenha uma vida de conquistas e que seja cheio da presença de Deus.

AGRADECIMENTOS

Gostaria de agradecer ao ICMC - Instituto De Ciências Matemáticas e de Computação da USP, por ter me proporcionado a oportunidade de expandir meus conhecimentos através desse curso.

Quero agradecer ao meu marido, Douglas Valentin o amor da minha vida, pela paciência e suporte durante esse processo, ele sabe o quão difícil foi chegar até aqui.

Quero agradecer à minha orientadora Mirela Cazzolato, por ter me motivado e não ter me deixado desistir no meio do curso.

Quero agradecer aos meus colegas Tatiane Anjos, Isaac Higuchi e todo o time de Engenharia de Analytics da empresa a qual trabalho, por terem me ajudado com ideias e materiais de estudo.

Por último, e mais importante, quero agradecer a Deus a qual eu sirvo, por ter me dado forças e livramentos para que eu pudesse viver o propósito que Ele designou para mim

“Feliz é o homem que acha sabedoria, o homem que adquire entendimento. ”

Provérbios 3:13

RESUMO

VALENTIN, J. E. B. **Agregando e analisando comentários georreferenciados de usuários para identificação e delimitação de regiões de risco em aplicativos de mobilidade.** 2024. 57 f. Trabalho de conclusão de curso (MBA em Inteligência Artificial e Big Data) – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos, 2024.

Este trabalho aborda a aplicação de tecnologias que visam auxiliar a melhora da segurança pública no contexto da mobilidade urbana, com foco em usuários de aplicativos de transporte. Com base em dados recentes de criminalidade, o estudo propõe um sistema de indicadores de periculosidade que pode ser integrado a aplicativos de mobilidade, visando auxiliar motoristas e pedestres na escolha de rotas mais seguras. A pesquisa se concentra em desenvolver uma metodologia para avaliar a periculosidade de rotas urbanas. O trabalho utiliza ferramentas como análise de sentimentos, modelos de linguagem de IA e mapeamento interativo, baseando-se em metodologias de aprendizado profundo e processamento de linguagem natural. Os resultados sugerem que a implementação de um indicador de periculosidade em aplicativos de mobilidade pode oferecer aos usuários maior controle e segurança em suas escolhas de trajeto, contribuindo também para uma comunicação mais eficaz entre a população e os órgãos de segurança pública.

Palavras-chave: Segurança. Mobilidade urbana. Informação georreferenciada. LLM.

ABSTRACT

VALENTIN, J. E. B. **Aggregating and analyzing georeferenced user comments for the identification and delimitation of risk regions in mobility applications.** 2024. 57 f. Trabalho de conclusão de curso (MBA em Inteligência Artificial e Big Data) – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos, 2024.

This study explores the application of advanced technologies in public safety within the context of urban mobility, focusing on users of transportation apps. Based on recent crime data, the research proposes a system of risk indicators integrated into mobility apps to assist drivers and pedestrians in choosing safer routes. The research focuses on three main objectives: (I) developing a methodology to assess urban route risk levels, incorporating data such as crime indices, public lighting, and police presence; (II) creating a prototype to integrate these indicators with mobility platforms; and (III) validating the impact of this integration on users' safety perceptions. Tools such as sentiment analysis, AI language models, and interactive mapping were utilized, leveraging deep learning and natural language processing methodologies. The results suggest that implementing a risk indicator in mobility apps can provide users with greater control and safety in their route choices, while also enhancing effective communication between the population and public safety authorities.

Keywords: Public safety. Mobility. Geo-information. LLM.

LISTA DE FIGURAS

Figura 1 – Percentual de trabalhadores que utilizam aplicativos de mobilidade, por tipo de plataformas de serviços - Por grandes regiões.	32
Figura 2 – Como usar o método polarity_scores	33
Figura 3 – Resultado do método polarity_scores.	33
Figura 4 – Taxa de roubos e furtos de veículos Brasil, 2018 – 2022.	38
Figura 5 – Naive Bayes: Exemplo de Mapa Interativo com Ocorrências de Roubo de Veículos	46
Figura 6 – Vader: Exemplo de Mapa Interativo com Ocorrências de Roubo de Veículos	46
Figura 7 – Vader - Comparação de Sentimentos dos Posts	48
Figura 8 – Naive Bayes - Comparação de Sentimentos dos Posts	48

LISTA DE TABELAS

Tabela 1 – Crimes contra o patrimônio: Roubo e Furto de veículos Brasil e Unidades da Federação – 2021-2022	24
Tabela 2 – Incidência de crimes em São Paulo nos últimos 4 anos.	37
Tabela 3 – Comparativo de Estruturação.	41

SUMÁRIO

	Lista de figuras	17
1	INTRODUÇÃO	23
1.1	Contextualização	23
1.2	Motivação	23
1.3	Lacunas e problemas existentes	23
1.4	Questão de Pesquisa e Objetivo	25
1.5	Organização da Monografia	26
2	FUNDAMENTAÇÃO TEÓRICA	27
2.1	Aplicativos de Mobilidade	27
2.1.1	Tecnologias dos aplicativos de Mobilidade	28
2.2	Segurança em Ambientes Urbanos	30
2.3	Usuários e Motoristas	31
2.4	Análise de Sentimentos	31
2.4.1	<i>Naive Bayes</i>	33
2.4.2	<i>Vader</i>	34
2.5	<i>Deep Learning</i>	34
2.6	Grandes Modelos de Linguagem (LLMs)	34
2.7	LGPD	35
2.8	Considerações Finais	35
3	METODOLOGIA, PROPOSTA E DESENVOLVIMENTO	37
3.1	Coleta, preparação e interpretação dos dados	37
3.2	Bibliotecas Utilizadas	38
3.2.1	Bibliotecas para o processamento de textos	38
3.2.2	Bibliotecas para análise de sentimento	39
3.3	Estruturação de dados	40
3.3.1	Estruturação dos Dados - Algoritmo Naive Bayes	40
3.3.2	Estruturação dos Dados - Algoritmo Vader	40
3.3.3	Comparação da Estruturação	41
4	RESULTADOS: AVALIAÇÃO EXPERIMENTAL	43
4.1	Processamento Inicial e Validação dos Dados	43
4.2	Resultados da Análise de Sentimentos	44
4.2.1	Resultados da análise de sentimentos com Naive Bayes	44
4.2.2	Resultados da análise de sentimentos com Vader	45

4.3	Visualização Geográfica e Interativa	45
4.3.1	Visualização Geográfica e Interativa dos resultados do Naive Bayes	45
4.3.2	Visualização Geográfica e Interativa - Vader	45
4.3.3	Características-chave dos mapas	47
4.4	Gráficos de Análise de Sentimentos	47
4.4.1	Interpretação dos Dados	49
5	CONCLUSÃO	51
	REFERÊNCIAS	53

1 INTRODUÇÃO

1.1 Contextualização

Aplicativos de mobilidade contêm funcionalidades robustas para direcionamento e ajuste de rota, bem como avisos relacionados ao tráfego, como engarrafamentos, acidentes, perigos na pista e outros alertas. Tais alertas auxiliam os motoristas na tomada de decisão sobre qual rota tomar, quando tomar, ou mesmo para buscar alternativas, considerando os diferentes eventos detectados e mostrados no mapa, relacionados à rota em questão.

Devido ao seu poderio, praticidade e disponibilidade, a utilização de aplicativos de mobilidade se tornou inerente ao dia-a-dia de motoristas ou mesmo pedestres buscando por guias geográficos.

1.2 Motivação

Os usuários de aplicativos de mobilidade precisam de mais do que apenas atualizações de trânsito. É preciso deixá-los cientes de possíveis riscos de segurança em suas rotas. A segurança dos usuários de aplicativos de mobilidade em regiões metropolitanas do Brasil é uma preocupação crescente diante do aumento da criminalidade e da complexidade das dinâmicas urbanas. Dados do Anuário Brasileiro de Segurança Pública, mostram que, no ano de 2021, cerca de 142.745 veículos foram roubados no Brasil. Em 2022 esse número subiu para 147.984, totalizando 373.225 veículos roubados ou furtados (FBSP, 2023) nestes dois anos. A Tabela 1 sumariza tais estatísticas.

Graças ao desenvolvimento acelerado das tecnologias móveis, é possível desenvolver soluções personalizadas e eficientes para uma variedade de problemas, otimizando processos e melhorando a qualidade de vida da população (Almeida *et al.*, 2016). No entanto, um problema persistente é que muitas rotas sugeridas por simples análise geográfica ou de monitoramento de tráfego deixam de cobrir indicativos que podem auxiliar motoristas e pedestres na tomada de decisão voltada para sua segurança.

1.3 Lacunas e problemas existentes

As aplicações existentes restringem o escopo dos avisos e notificações aos problemas de trânsito ou alertas de bloqueios. No entanto, adicionar a opção de mapeamento de avisos de segurança pública poderia ser de grande relevância, tanto em regiões com grandes quantidades de pessoas (como shows ou áreas litorâneas), ou mesmo avisos de trânsito para motoristas evitarem determinadas áreas - avisos esses que poderiam ser determinados ou validados por autoridades responsáveis. Nesse contexto, existem discussões ativas de projetos de leis como a PL 6446/19 (Nunes, 2019) para inclusão da funcionalidade de

Tabela 1 – Crimes contra o patrimônio: Roubo e Furto de veículos Brasil e Unidades da Federação – 2021-2022

Brasil e Unidades da Federação	Roubo e Furto de Veículo				
	Ns. Absolutos		Taxas ⁽²⁾		Variação (%)
	2021 ⁽³⁾	2022	2021	2022	
Brasil	334.715	373.225	300,3	324,2	8,0
Acre	1.680	1.365	524,7	408,2	-22,2
Alagoas	3.165	3.773	322,4	364,8	13,2
Amapá ⁽⁴⁾	806	720	359,5	309,4	-13,9
Amazonas	4.181	3.425	412,3	320,2	-22,4
Bahia	16.061	20.100	342,0	411,2	20,2
Ceará	11.857	13.601	337,5	375,1	11,1
Distrito Federal	5.770	5.283	291,5	261,3	-10,3
Espírito Santo	7.838	7.668	362,8	341,0	-6,0
Goiás	7.197	6.992	164,7	153,9	-6,5
Maranhão	6.854	6.447	353,2	317,4	-10,1
Mato Grosso	3.936	3.250	161,1	126,5	-21,4
Mato Grosso do Sul	3.307	3.961	187,4	217,1	15,8
Minas Gerais	23.673	28.235	189,3	216,7	14,5
Pará	5.754	5.397	244,6	217,7	-11,0
Paraíba	5.027	5.765	341,8	378,5	10,7
Paraná	15.701	17.265	188,3	201,3	6,9
Pernambuco	14.662	18.378	439,7	534,4	21,5
Piauí	6.420	6.816	479,5	492,0	2,6
Rio de Janeiro	38.760	42.062	531,2	562,7	5,9
Rio Grande do Norte	6.656	6.287	460,8	420,5	-8,8
Rio Grande do Sul	13.859	13.497	180,2	171,5	-4,8
Rondônia	4.096	4.575	367,9	396,3	7,7
Roraima	1.226	859	489,1	326,2	-33,3
Santa Catarina	8.833	9.420	152,7	157,7	3,3
São Paulo	112.711	133.585	358,3	413,7	15,4
Sergipe	2.984	2.718	342,3	299,5	-12,5
Tocantins	1.701	1.781	214,9	214,4	-0,2

Fonte: Retirado de (FBSP, 2023)

alertas de criminalidade por parte das empresas do aplicativo. A ementa da PL é a seguinte: *Dispõe sobre a oferta, por fornecedores de mapas para dispositivos de sistemas de posicionamento global (GPS) e outros sistemas similares de navegação global por satélite, de recurso de alerta ao usuário em caso de aproximação de áreas com elevado índice de ocorrência de crimes* (Nunes, 2019).

Há diversas discussões e argumentações sobre essa funcionalidade. Por um lado, os benefícios de tal funcionalidade vão no sentido de aumentar a segurança das pessoas, por meio do apoio de aplicativos georreferenciados, sugestões de rotas potencialmente mais seguras, ou mesmo a recomendação de evitar áreas problemáticas. Já com relação aos pontos negativos, autoridades discutem a questão sob um ponto de vista da estigmatização ou mesmo desvalorização de determinadas regiões urbanas, com possível ônus social e financeiro, por exemplo, do ponto de vista imobiliário.

No cenário elencado, existe uma lacuna de contribuição para a incorporação de alertas de segurança georreferenciados, considerando os problemas discutidos. Uma solução poderia contar com o intermédio de autoridades públicas e da LLM, de maneira que os alertas sejam pontuais e com tempo de validade pré-determinado.

Dessa forma, as informações não seriam persistentes no sistema, evitando a criação de estigmas com relação à região em questão. Tal funcionalidade poderia ser útil, também, para forças de segurança. Conforme pontuado no Anuário Brasileiro de Segurança Pública de 2023 (FBSP, 2023), a modernização das polícias civis é essencial para enfrentar as complexas dinâmicas do crime organizado. A capacitação dos policiais com as mais recentes ferramentas tecnológicas e o estímulo ao desenvolvimento de habilidades analíticas são pilares para garantir investigações eficazes. A valorização dos profissionais da segurança pública é um investimento indispensável para a construção de instituições policiais mais eficientes e adaptadas às demandas da sociedade.

1.4 Questão de Pesquisa e Objetivo

A questão de pesquisa do presente trabalho é a seguinte:

“É possível incorporar um indicador de periculosidade ou (alerta de segurança) georreferenciado, para contribuir para a melhoria da segurança dos usuários de aplicativos de mobilidade em regiões metropolitanas, possibilitando uma escolha mais informada de rotas e potencialmente reduzindo incidentes relacionados à criminalidade urbana?”

Este trabalho visa abordar a problemática levantada por meio do desenvolvimento de um indicador de periculosidade (ou de alertas), que permitirá aos usuários e aos próprios aplicativos avaliarem o risco associado a determinadas rotas. Na abordagem proposta, são simulados relatos georreferenciados com Grandes Modelos de Linguagem (do inglês, *Large Language Models* - LLMs) sobre situações que podem, eventualmente, trazer perigo

a pessoas ou motoristas que passem por determinada região. Tais relatos serão textuais, e serão processados por uma abordagem de análise de sentimento, de modo a analisar se a notificação é negativa (no sentido de reportar um perigo), positiva ou neutra. Para isso, são estudadas e aplicadas abordagens de processamento de linguagem natural para detecção de sentimentos.

Uma vez categorizados os relatos gerados pela LLM, outra etapa do método consiste em agregar relatos no mapa. Diferentes relatos (ou alertas) de usuários, que sejam emitidos em localizações georreferenciadas próximas a um determinado raio de abrangência, poderão ser, futuramente, agregadas. Todo alerta será visível no mapa por um tempo padrão. A ideia é permitir, em aplicações futuras, que o tempo de visibilidade de um alerta possa ser **decrementado** até que ele expire (e seja excluído do mapa), ou **renovado**, que ocorre quando um novo alerta é emitido dentro de seu raio de abrangência. No caso da renovação do tempo de expiração, o novo alerta receberá o tempo padrão (máximo) de visibilidade, e os alertas existentes (dentro da região específica) terão o tempo máximo de visibilidade restaurados. Essa ideia é utilizada como motivação para aplicação do método proposto.

Neste trabalho, o objetivo final será convergido na criação um método que abranja o processamento de textos e a classificação alertas, ilustrando a funcionalidade do método. Tal funcionalidade poderá ser incorporada futuramente em aplicativos móveis, dentro da motivação elencada nas seções iniciais deste capítulo.

1.5 Organização da Monografia

Os demais capítulos desta monografia estão organizados da seguinte forma:

O Capítulo 2 apresenta a fundamentação teórica. O Capítulo 3 descreve a metodologia, proposta e desenvolvimento do presente trabalho. O Capítulo 4 descreve os resultados experimentais do trabalho desenvolvido. Finalmente, o Capítulo 5 apresenta a conclusão do trabalho e apresenta apontamentos para os trabalhos futuros.

2 FUNDAMENTAÇÃO TEÓRICA

Nesse capítulo, são apresentados os principais conceitos relacionados a aplicativos de mobilidade, segurança em ambientes urbanos, análise de sentimentos a partir de textos e funções de distâncias aplicadas a posições georreferenciadas. Tais conceitos buscam contextualizar e embasar a proposta do presente trabalho.

2.1 Aplicativos de Mobilidade

Os aplicativos de mobilidade modernos transformaram significativamente a forma como as pessoas navegam nas cidades, oferecendo funcionalidades avançadas de direcionamento e ajuste de rotas. Estes sistemas utilizam dados de geolocalização em tempo real, algoritmos de otimização de trajetória e informações de tráfego para oferecer aos usuários opções eficientes e seguras de deslocamento, auxiliando o usuário em diversas atividades como traçar rotas e fornecer endereços (França, 2014). A robustez dessas plataformas se reflete na capacidade de fornecer alertas sobre engarrafamentos, acidentes e outras condições adversas de trânsito, ajudando os motoristas a tomar decisões informadas e adaptar suas rotas conforme necessário.

No Brasil, atualmente, utilizam-se aplicativos para mobilidade como a Uber, fundada em 2009, nos EUA, e chegando no Brasil em 2014. A Uber revolucionou o mercado no que se refere aos serviços de transportes privativos no país (MARIANO; DIAZ, 2017), sendo um dos principais serviços de transporte, contando atualmente com mais de 120 milhões de usuários, seguido por suas concorrentes como 99 pop, uma startup brasileira que possui cerca de 50 milhões de usuários ativos e 1,5 milhões de motoristas cadastrados (IBGE, 2022).

A Waze foi fundada em 2008 pelo israelense Uri Levine, o engenheiro de software Ehud Shabtai, Amir e Gili Shinar, Arie Gillon e Noam Bardin, CEO da empresa desde 2009 (Souza, 2023), sendo vendida para a Google em 2013, contendo mais de 50 milhões de usuários. O Google Maps começou em 2004 quando os irmãos Lars e Jens Rasmussen foram até a Google para apresentar sua startup, a Where2 Technologies, onde o conceito era o mesmo do que o Google Maps representa hoje: “Um serviço de mapas no qual o usuário poderia fazer buscas, dar zoom e passear pela tela do computador.” (Caputo, 2015). A presença acentuada do serviço de transporte foi fomentada pelo avanço de tecnologias tais como geolocalização, comunicação e desenvolvimento de aplicativos móveis. A combinação de diferentes plataformas, tecnologias e dados de usuários tem grande potencial ao possibilitar o desenvolvimento de soluções que possam auxiliar em diversos setores da sociedade, como segurança pública, mobilidade urbana e telecomunicações.

2.1.1 Tecnologias dos aplicativos de Mobilidade

Esta subseção tem o intuito de discutir as funcionalidades e aplicações dos aplicativos de mobilidade Uber, 99, Waze e Google Maps, dando maior ênfase às técnicas de Inteligência Artificial utilizadas para identificar zonas de risco, incidentes no trânsito, bem como visões espaciais dos mapas para identificação de trânsito especificamente na cidade de São Paulo.

A Uber utiliza uma variedade de técnicas de inteligência artificial (IA) para monitorar zonas de risco aos usuários. Algumas delas incluem aprendizado de máquina, mapeamento em tempo real, linguagem natural, dados de GPS (*Global Position System*) e *crowd sourcing*.

O aprendizado de máquina, no contexto da análise de dados, pode ser definido como a habilidade de um programa de melhorar seu desempenho em tarefas específicas através da acumulação de experiência, ou seja, ao ser exposto a novos dados (MITCHELL, 1997). A Uber coleta grandes quantidades de dados sobre viagens, incluindo horários, locais de partida e destino, e incidentes relatados. Algoritmos de aprendizado de máquina analisam esses dados para identificar padrões e áreas com maior incidência de problemas de segurança.

Ao coletar informações de usuários, o mapeamento em tempo real permite rastrear o trajeto do usuário em tempo real, fornecendo informações atualizadas sobre sua posição a cada momento (França, 2014). A Uber utiliza mapas em tempo real para monitorar as condições de trânsito, acidentes e outros eventos que possam afetar a segurança das viagens. Essa informação é usada para alertar motoristas e usuários sobre áreas a serem evitadas. Isso permite que a empresa monitore áreas de risco e envie alertas em tempo real caso um usuário esteja entrando em uma zona perigosa. A 99 não divulga publicamente os detalhes técnicos exatos dos algoritmos e modelos de IA que utiliza para indicar regiões de risco em seu aplicativo. No entanto, é provável que a empresa utilize uma combinação de técnicas, incluindo aprendizado de máquina para aprender com históricos sobre incidentes de segurança, como assaltos, roubos e acidentes, para identificar padrões e prever áreas com maior probabilidade de risco.

O PLN busca criar programas que consigam ‘ler’ e ‘entender’ textos da mesma forma que humanos (Caseli; Nunes, 2023). Análise de textos pode ser aplicada em diversos contextos, por exemplo, na análise de notícias, boletins de ocorrência e outras fontes para identificar menções a eventos de segurança em determinadas áreas.

O GPS foi inicialmente concebido para aplicações militares, no qual o Sistema de Posicionamento Global evoluiu para se tornar uma ferramenta indispensável para a navegação civil, tanto em ambientes marítimos e aéreos quanto em deslocamentos terrestres (Almeida *et al.*, 2016). A coleta de dados de localização dos motoristas e passageiros para

identificar áreas com alta concentração de incidentes ou comportamento suspeito é de suma importância para as aplicações modernas de mobilidade urbana.

Utilizar de mapas e dados geográficos para identificar áreas com características que podem contribuir para apontar riscos, como baixa iluminação, falta de policiamento ou histórico de criminalidade. O Waze utiliza uma combinação de *crowdsourcing*, PLN, aprendizado de máquina e fusão de dados para coletar, analisar e verificar informações sobre incidentes no trânsito, fornecendo aos usuários informações atualizadas e relevantes para poderem tomar decisões mais informadas sobre suas rotas. Nesse contexto, *crowdsourcing* pode ser definido como um modelo de computação que envolve a colaboração ativa de pessoas em uma tarefa, particularmente em atividades que os seres humanos conseguem realizar com mais facilidade ou eficiência do que os computadores (TONG *et al.*, 2020). O Waze depende muito dos relatos dos próprios usuários para identificar e verificar incidentes. A IA é utilizada para analisar a qualidade e a confiabilidade desses relatos, levando em consideração fatores como a frequência de contribuição do usuário, a consistência das informações e a concordância com outros relatos. A geolocalização é utilizada nesses aplicativos para centralizar e contextualizar os dados gerados pelos usuários, associando-os à sua localização geográfica (França, 2014).

O PLN também pode ser utilizado no contexto de mobilidade urbana, buscando analisar os textos dos relatos dos usuários, extraindo informações relevantes como o tipo de incidente (acidente, congestionamento, objeto na pista, etc.), a localização, a gravidade e outros detalhes importantes. Já algoritmos de aprendizado de máquina podem ser utilizados para identificar padrões nos dados coletados, como horários e locais de maior incidência de determinados tipos de incidentes, e para prever a evolução do tráfego com base nas informações disponíveis.

O Google Maps utiliza diversas técnicas de Inteligência Artificial (IA) para monitorar e analisar dados espaciais de trânsito. Algoritmos de aprendizado de máquina são usados para analisar grandes volumes de dados históricos e em tempo real, como velocidades médias, padrões de tráfego e incidentes, para prever o fluxo de trânsito e identificar áreas congestionadas.

Redes Neurais Profundas (*Deep Neural Networks*) são empregadas para reconhecer padrões complexos em imagens de satélite e dados de sensores, como semáforos e câmeras de trânsito, para detectar acidentes, obras e outros eventos que afetam o tráfego. O *Deep Learning* também pode ser definido como uma solução que permite que os computadores aprendam com experiência e compreendam o mundo em termos de um conceito hierárquico definido por meio de sua relação com conceitos mais simples. Ao adquirir conhecimento a partir da experiência, essa abordagem evita a necessidade de operadores humanos especificarem formalmente todo o conhecimento necessário para o computador. A hierarquia de conceitos possibilita que o computador aprenda conceitos complexos construindo-os a

partir de conceitos mais simples. Se desenharmos um gráfico mostrando como esses conceitos são construídos um sobre o outro, o gráfico será profundo, com muitas camadas. Por esse motivo, chamamos essa abordagem de IA de aprendizado profundo (GOODFELLOW; BENGIO; COURVILLE, 2016).

2.2 Segurança em Ambientes Urbanos

A questão da segurança urbana tem se tornado uma preocupação cada vez mais relevante, especialmente em áreas metropolitanas que enfrentam altos índices de criminalidade. Historicamente, a análise da segurança urbana tem sido fundamentada em dados de ocorrências criminais registradas, níveis de policiamento, qualidade da iluminação pública e percepções dos residentes locais. A combinação desses dados com tecnologias de informação e comunicação oferece oportunidades significativas para fortalecer estratégias de prevenção e resposta a incidentes criminais.

A investigação e a elucidação de homicídios constituem um dos pilares para a redução das taxas de homicídios. No entanto, apesar do elevado número de mortes violentas intencionais, a qualidade das investigações em grande parte dos estados brasileiros é insatisfatória, resultando em altas taxas de crimes não esclarecidos (Costa, 2021). A ausência de um sistema nacional de indicadores para a investigação criminal impede a determinação precisa da taxa de esclarecimento de crimes no Brasil (Costa, 2021).

A criminalidade é um fenômeno complexo, influenciado por diversos fatores econômicos, sociais e culturais. Embora não se limite à pobreza, é inegável a forte relação entre esses elementos (Dias, 2019). “O capitalismo molda a sociedade em torno do consumo desenfreado e da busca incessante por lucro, expandindo os mercados sem limites. Esse sistema impulsiona o crescimento econômico, mas, ao transformar tudo em mercadoria, desvaloriza a vida humana e o trabalho, reduzindo tudo a números e preços (Dias, 2019).”

Ainda nas palavras de Denise Oliveira Dias (2019, p.57): “A pobreza conduz à falta de oportunidades, de estudo, emprego e vida social digna conforme prevê a Constituição Federal; ela limita a vida dos pobres e também das outras classes sociais, pois os seus derivados acabam por afetar toda a coletividade.”

A urgência em solucionar os problemas relacionados à criminalidade no Brasil é evidente, considerando as graves consequências que ela acarreta para a sociedade (Freitas, 2023). Diante dos desafios cada vez mais complexos da segurança pública, a adaptação às novas tecnologias se torna imperativa. Ferramentas como câmeras de monitoramento, drones e aplicativos móveis estão sendo incorporadas às estratégias policiais para otimizar a resposta a incidentes e prevenir crimes (Silva, 2023).

As inovações tecnológicas demonstram um potencial significativo para revolucionar a área da segurança pública, proporcionando às forças policiais recursos mais eficazes para

a prevenção e o combate à criminalidade, também proporcionam uma maior segurança aos usuários das tecnologias, como usuários de aplicativo de mobilidade, que dependem do uso da tecnologia para se locomover e trabalhar. Contudo, a implementação dessas tecnologias deve ser conduzida de forma responsável, ética e transparente, a fim de assegurar que os benefícios superem os desafios e as preocupações inerentes a esse processo (Silva, 2023).

2.3 Usuários e Motoristas

De acordo com o IBGE (IBGE, 2022), o Brasil registrou em 2022 cerca de 1,5 milhão de trabalhadores por meio de plataformas e aplicativos de serviços, o que representa 1,7% da população ocupada no setor privado. A Figura 1 mostra que 52,2% (778 mil pessoas) dedicavam-se principalmente ao transporte de passageiros por meio de aplicativos (incluindo táxis), enquanto 47,2% (704 mil pessoas) estavam envolvidas no transporte particular de passageiros (excluindo táxis), já 13,9% (207 mil pessoas) trabalhavam com aplicativos de táxi (IBGE, 2022).

A maior proporção de trabalhadores está na Região Norte, com 61,2% do total (IBGE, 2022). Por outro lado, esta mesma região registrou a menor participação de pessoas empregadas em aplicativos de serviços gerais ou profissionais, com apenas 5,6% do total, menos da metade do índice observado no restante do país. O Sudeste concentrou a maior proporção de trabalhadores, totalizando 61,4% (IBGE, 2022).

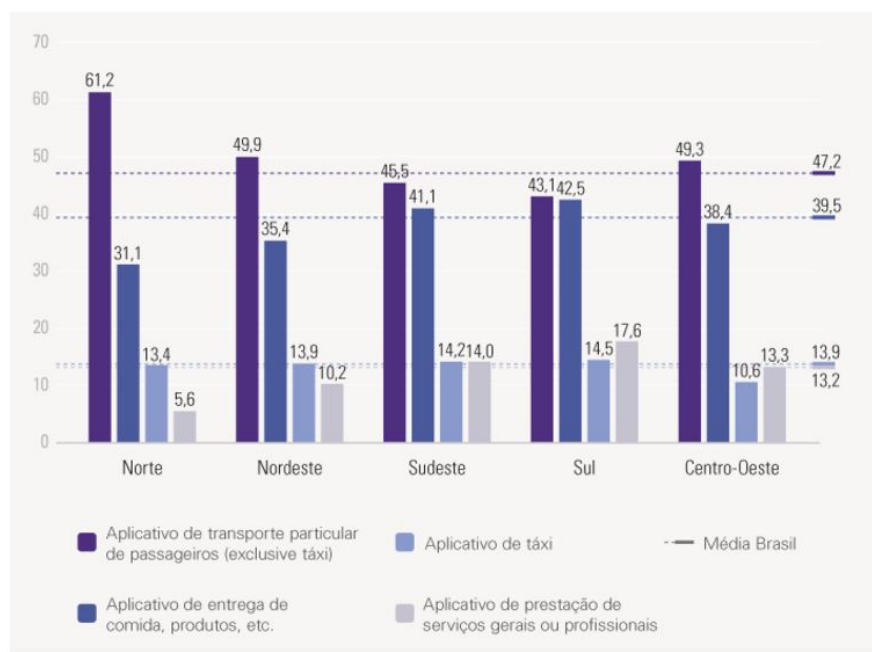
Outro problema surge em se tratando de aplicativos de mobilidade: a falta de acesso de alguns grupos às empresas que fornecem esses serviços. A falta de infraestrutura adequada, a violência e a menor renda da população em algumas regiões periféricas dos grandes centros urbanos criam um ambiente menos atrativo para as empresas de aplicativos de mobilidade e dificultam o acesso dos usuários a esses serviços (WARMAR; PEREIRA, 2022).

Embora as novas tecnologias ofereçam um leque de possibilidades para a segurança pública, a experiência do usuário e do motorista ainda pode ser aprimorada. A falta de interfaces intuitivas, a complexidade de alguns sistemas e a necessidade de maior personalização são alguns dos pontos que demandam atenção. Além disso, a criação de aplicativos que facilitem a comunicação entre cidadãos e polícia, a disponibilização de informações em tempo real sobre o trânsito e a segurança em determinada região, e a implementação de sistemas de pagamento eletrônico para multas e taxas podem contribuir para uma experiência mais satisfatória e eficiente.

2.4 Análise de Sentimentos

A análise de sentimentos em textos pode ser realizada por meio de diferentes técnicas. A classificação supervisionada, por exemplo, consiste em treinar um algoritmo de

Figura 1 – Percentual de trabalhadores que utilizam aplicativos de mobilidade, por tipo de plataformas de serviços - Por grandes regiões.



Fonte: (IBGE, 2022)

aprendizado de máquina com um conjunto de dados rotulados, como tweets classificados como positivos ou negativos. Dessa forma, o modelo é capaz de generalizar para novos textos. Outras abordagens incluem a análise lexical, que utiliza dicionários de palavras com polaridade associada (por exemplo, “excelente” é positivo, “horrível” é negativo), e a análise de aspectos, que identifica os aspectos mencionados (por exemplo, “bateria”, “câmera”) e sua polaridade (por exemplo, “a bateria dura muito”) (Bairros, 2023).

Para o desenvolvimento desse projeto, foi utilizado o modelo de aprendizado na máquina NLTK (*Natural Language Toolkit*). Segundo (Andre, 2021), trata-se de um conjunto de bibliotecas e ferramentas para Python, especializado no processamento de linguagem natural. Ele permite realizar tarefas como tokenização, *stemming*, análise de sentimentos, reconhecimento de entidades nomeadas e muito mais, possibilitando a extração de conhecimento a partir de grandes volumes de dados textuais (Andre, 2021). O NLTK ajuda o computador a ‘ler’ e entender a linguagem humana, separando as palavras, identificando o tipo de cada palavra (se é um substantivo, verbo, etc.), e até mesmo entendendo se um texto é positivo ou negativo.

O objetivo do NLTK neste trabalho é analisar os sentimentos dos usuários em relação a uma região baseado em textos. É possível usar essa ferramenta para determinar se um texto é positivo, negativo ou neutro. A classe *SentimentIntensityAnalyzer* oferece um método chamado *polarity_scores*, que analisa um texto e fornece uma avaliação do seu sentimento geral, indicando as proporções de positividade, negatividade e neutralidade

(Pappas, 2022), conforme o exemplo na Figura 2.

Figura 2 – Como usar o método `polarity_scores`

```
{'neg': 0.0, 'neu': 0.482, 'pos': 0.518, 'compound': 0.8622}
```

Fonte: (Pappas, 2022)

O código retornará com o seguinte resultado, conforme a Figura 3.

Figura 3 – Resultado do método `polarity_scores`.

```
text = "I love NLTK! It's a great library for working with text data."
scores = analyzer.polarity_scores(text)

print(scores)
```

Fonte: (Pappas, 2022)

2.4.1 *Naive Bayes*

O Naive Bayes é um algoritmo de aprendizado de máquina que utiliza probabilidades para classificar dados. Ele assume que as características dos dados são independentes umas das outras, o que é chamado de “ingenuidade” na sua denominação. Apesar dessa simplificação, o Naive Bayes tem se mostrado eficaz em diversas aplicações, especialmente quando a velocidade de previsão é importante (HEMACHANDRAN *et al.*, 2022). Na tarefa de categorização de textos, o algoritmo Naive Bayes é utilizado para estimar a probabilidade de um documento pertencer a uma determinada classe. Embora a probabilidade em si seja importante, o que realmente nos interessa é a capacidade de ordenar os documentos de acordo com sua relevância para cada classe (KIM *et al.*, 2002). Para isso, a razão de *log-odds* é empregada como uma medida de relevância. Essa medida é calculada utilizando a regra de Bayes e nos permite comparar a probabilidade de um documento pertencer a uma classe específica em relação à probabilidade de pertencer a qualquer outra classe. Por exemplo, ao classificar e-mails como 'spam' ou 'não spam', a razão de *log-odds* nos permite identificar os e-mails com maior probabilidade de serem spam (KIM *et al.*, 2002), segue a Equação 2.1, demonstrando na prática:

$$Relevance(d_i, c_j) = \log \frac{P'(c_j|d_i)}{P'(\bar{c}_j|d_i)} = \log \frac{P'(d_i|c_j)}{P'(d_i|\bar{c}_j)} + \log \frac{P'(c_j)}{P'(\bar{c}_j)} \quad (2.1)$$

2.4.2 *Vader*

Vader Sentiment, ou apenas Vader, é um classificador lexical que utiliza um léxico pré-treinado para determinar a polaridade (positiva, negativa ou neutra) de um texto. O Vader atribui a cada palavra uma pontuação semântica e calcula uma pontuação composta para a frase, considerando intensificadores e negadores. Dessa forma, é possível classificar cada avaliação de acordo com seu sentimento geral (ISNAN; ELWIREHARDJA; PARDAMEAN, 2023). Ainda nas palavras de (ISNAN; ELWIREHARDJA; PARDAMEAN, 2023), o Vader emprega métodos qualitativos e quantitativos para desenvolver e validar um léxico de sentimentos contextualizado. Essa abordagem híbrida possibilita uma análise de sentimentos mais precisa em diversos domínios, incluindo críticas de filmes, avaliações de produtos e conteúdo gerado por usuários em redes sociais.

2.5 *Deep Learning*

O *deep learning* é um subcampo do aprendizado de máquina que se baseia em redes neurais artificiais com arquitetura profunda (LECUN; BENGIO; HINTON, 2015). Essas redes são compostas por várias camadas de neurônios interconectados, permitindo a extração de representações hierárquicas de dados. Ao ser treinada com grandes conjuntos de dados rotulados, a rede ajusta seus parâmetros de forma a mapear entradas para saídas desejadas, o que a torna uma ferramenta poderosa para tarefas como reconhecimento de imagens, processamento de linguagem natural e análise de dados (Silva, 2023). Na tarefa de categorização de textos, o algoritmo Naive Bayes é utilizado para estimar a probabilidade de um documento pertencer a uma determinada classe. Embora a probabilidade em si seja importante, a capacidade de ordenar os documentos de acordo com sua relevância para cada classe é uma característica muito relevante do algoritmo.

Para isso, a razão de log-odds é empregada como uma medida de relevância. Essa medida é calculada utilizando a regra de Bayes e nos permite comparar a probabilidade de um documento pertencer a uma classe específica em relação à probabilidade de pertencer a qualquer outra classe. Por exemplo, ao classificar e-mails como ‘spam’ ou ‘não spam’, a razão de log-odds nos permite identificar os e-mails com maior probabilidade de serem spam.

2.6 **Grandes Modelos de Linguagem (LLMs)**

Um LLM é um tipo de inteligência artificial (IA) que utiliza técnicas de aprendizado de máquina para entender e gerar linguagem humana de forma natural. Ele pode ser imaginado como um computador que aprende a “falar” e “escrever” como humanos.

Segundo (GE *et al.*, 2023), as arquiteturas de transformadores, com sua alta capacidade de paralelização, revolucionaram o campo do processamento de linguagem

natural. Modelos de linguagem pré-treinados (PLMs), construídos sobre essas arquiteturas e treinados em grandes volumes de dados, adquiriram a habilidade de compreender nuances linguísticas complexas e gerar texto de forma coerente e relevante. A escalada desses modelos para o formato de Grandes Modelos de Linguagem (LLMs) amplificou ainda mais suas capacidades, permitindo-lhes aprender padrões linguísticos mais sofisticados e realizar tarefas que exigem raciocínio lógico. Contudo, é importante ressaltar que os LLMs podem apresentar desafios em domínios específicos e gerar resultados imprecisos, especialmente quando confrontados com informações ambíguas ou contraditórias (GE *et al.*, 2023).

Conforme (YAO *et al.*, 2024), para ser considerado eficaz, um LLM deve:

1. Dominar a compreensão e interpretação de textos em linguagem natural, extraindo informações relevantes para diversas tarefas;
2. Ser capaz de gerar textos autênticos e contextualmente apropriados, desde frases simples até textos mais complexos;
3. Demonstrar conhecimento de domínio e contexto, adaptando suas respostas de acordo com a situação; e
4. Utilizar as informações contidas nos textos para resolver problemas, como responder a perguntas e buscar informações específicas.

2.7 LGPD

Por fim, é pontuado que em nenhum momento deste trabalho, os dados de usuários foram utilizados, respeitando as normas da LGPD, garantindo a privacidade de cada indivíduo. No entanto, em um cenário real, trata-se de uma importante preocupação ao lidar com textos e localizações obtidas por meio do usuário.

A Lei Geral de Proteção de Dados (LGPD, 2018) estabelece um marco regulatório abrangente para o tratamento de dados pessoais no Brasil, conferindo aos titulares diversos direitos, incluindo o de ter seus dados protegidos contra o acesso não autorizado. A LGPD proíbe a divulgação de dados pessoais sem o consentimento do titular, salvo em hipóteses expressamente previstas em lei, como o cumprimento de obrigações legais ou para a proteção da vida. A norma também impõe aos controladores o dever de adotar medidas de segurança para proteger os dados pessoais contra acessos não autorizados e situações acidentais ou ilícitas de destruição, perda, alteração, comunicação ou difusão (LGPD, 2018).

2.8 Considerações Finais

Neste capítulo foram discutidos os principais assuntos relacionados à proposta deste projeto de conclusão de curso. A seguir, é apresentada a metodologia proposta, a fim de

atender aos objetivos discutidos até aqui.

3 METODOLOGIA, PROPOSTA E DESENVOLVIMENTO

Este capítulo aborda a metodologia proposta para desenvolver o indicador de periculosidade em aplicativos de mobilidade. Para o desenvolvimento do trabalho, foram coletados dados da edição do ano de 2023 do Fórum Brasileiro de Segurança Pública, contendo dados dos anos de 2021 e 2022, bem como do site da Secretaria de Segurança Pública do Estado de São Paulo, que contempla dados entre 2023 e 2024. Os dados foram concentrados na cidade de São Paulo, onde o estudo foi realizado e grande parte dos dados foram coletados em tempo real para testar a aplicabilidade do projeto.

3.1 Coleta, preparação e interpretação dos dados

Os dados foram tratados sendo selecionadas apenas ocorrências relacionadas à Capital de São Paulo, filtrados por latrocínio, furto de veículo e roubo de veículo, conforme a Tabela 2.

De acordo com a Tabela 2, o número de casos de latrocínio parece apresentar uma relativa estabilidade ao longo dos anos, com pequenas flutuações mensais e anuais. Há uma tendência de redução no número de roubos de veículos ao longo dos últimos quatro anos. Essa redução é mais evidente ao comparar os dados de 2021 com os anos seguintes.

Os números de furtos de veículos apresentam uma maior oscilação ao longo dos anos, mas, em geral, mantêm-se em níveis relativamente estáveis. A estabilidade dos números de furtos de veículos pode indicar que as medidas adotadas para combater esse tipo de crime

Tabela 2 – Incidência de crimes em São Paulo nos últimos 4 anos.

2024													
Natureza	Janeiro	Fevereiro	Marco	Abril	Mai	Junho	Julho	Agosto	Setembro	Outubro	Novembro	Dezembro	Total
LATROCÍNIO	7	5	5	1	5	3	5	31
ROUBO DE VEÍCULO	922	980	1.015	1.028	1.097	1.051	1.015	7.108
FURTO DE VEÍCULO	3.231	3.322	3.431	3.705	3.530	3.433	3.785	24.437

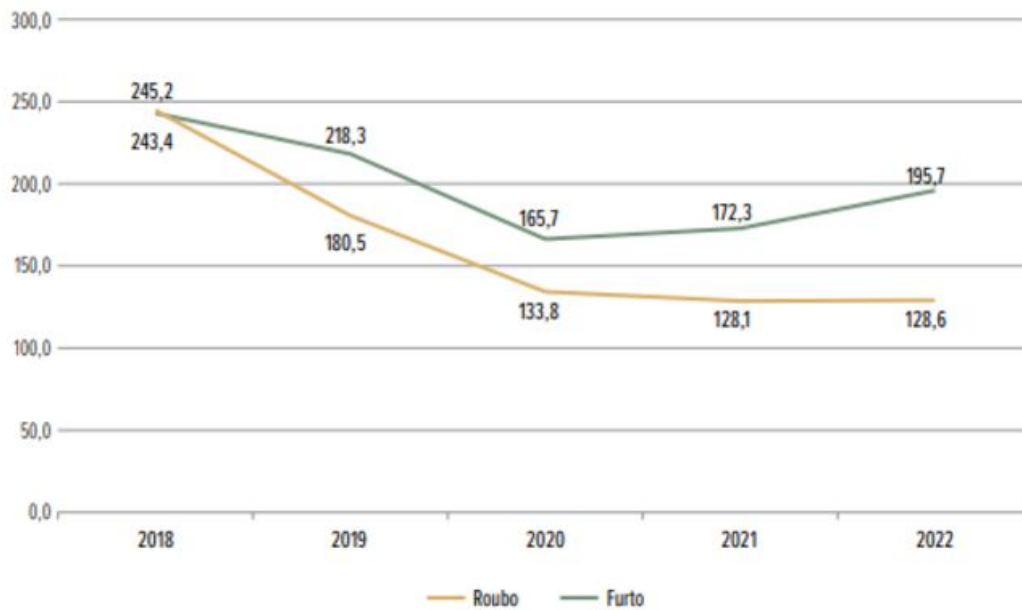
2023													
Natureza	Janeiro	Fevereiro	Marco	Abril	Mai	Junho	Julho	Agosto	Setembro	Outubro	Novembro	Dezembro	Total
LATROCÍNIO	3	1	3	7	4	4	3	3	3	2	4	6	43
ROUBO DE VEÍCULO	1.213	1.144	1.495	1.259	1.201	1.146	1.136	1.131	1.174	1.315	1.349	1.420	14.983
FURTO DE VEÍCULO	3.308	3.022	3.794	3.154	3.648	3.336	3.414	3.624	3.243	3.535	3.541	2.900	40.519

2022													
Natureza	Janeiro	Fevereiro	Marco	Abril	Mai	Junho	Julho	Agosto	Setembro	Outubro	Novembro	Dezembro	Total
LATROCÍNIO	3	9	5	4	6	3	5	0	11	4	6	7	63
ROUBO DE VEÍCULO	1.125	1.050	1.458	1.234	1.328	1.221	1.382	1.330	1.302	1.573	1.718	1.760	16.481
FURTO DE VEÍCULO	2.837	2.931	3.865	3.380	3.282	3.345	3.566	3.675	3.694	3.858	3.961	3.218	41.612

2021													
Natureza	Janeiro	Fevereiro	Marco	Abril	Mai	Junho	Julho	Agosto	Setembro	Outubro	Novembro	Dezembro	Total
LATROCÍNIO	7	5	4	4	7	5	1	3	8	5	3	1	53
ROUBO DE VEÍCULO	1.216	1.083	1.044	993	974	965	1.094	1.064	1.208	1.287	1.281	1.447	13.656
FURTO DE VEÍCULO	2.323	2.711	2.793	2.684	2.927	2.799	2.939	3.074	3.099	3.383	3.090	2.509	34.331

Fonte: Elaborado pela autora.

Figura 4 – Taxa de roubos e furtos de veículos Brasil, 2018 – 2022.



Fonte: (FBSP, 2023)

ainda não foram suficientes para reduzir significativamente sua ocorrência.

Ao analisar a Figura 4, nota-se que as taxas de roubos e furtos por grupo de 100 mil veículos registrados, observa-se uma queda acentuada nos roubos nos últimos anos (47,6%), enquanto a redução dos furtos foi menos expressiva (19,6%). Essa tendência, demonstra que, atualmente, o furto se consolidou como o crime mais comum envolvendo a subtração de veículos no país.

Outra fonte de coleta de dados utilizada neste trabalho foi de grandes conjuntos de dados textuais de fontes públicas como redes sociais, fóruns online e principalmente, do próprio aplicativo de mobilidade como o Waze, onde os usuários registraram as ocorrências em tempo real. O pré-processamento desses dados se deu com a limpeza de links, hashtags e tags em HTML, caracteres especiais e foi realizada a padronização de texto.

3.2 Bibliotecas Utilizadas

3.2.1 Bibliotecas para o processamento de textos

NLTK (Natural Language Toolkit). O NLTK é uma biblioteca amplamente utilizada para processamento de linguagem natural (PLN). Ela oferece ferramentas para manipulação de texto, análise linguística, construção de modelos de aprendizado supervisionado e acesso a corpora predefinidos. No contexto do algoritmo Naive Bayes, os seguintes recursos da biblioteca foram utilizados:

- **nltk.corpus:** Permite acesso a corpora prontos, como o movie_reviews. O corpus

contém avaliações de filmes já categorizadas como positivas ou negativas, sendo uma base ideal para problemas de classificação binária.

- **nltk.classify.NaiveBayesClassifier:** Implementa o classificador Naive Bayes, utilizando probabilidade condicional para categorizar textos. Treina e testa o modelo com base em *features* extraídas do texto, representadas como um dicionário de palavras.
- **nltk.probability:** Usado para calcular as probabilidades associadas às *features* (características), essencial para a implementação do modelo de Naive Bayes.
- **nltk.download:** Baixa pacotes de dados e recursos necessários, como o corpus `movie_reviews` e o tokenizer para dividir textos em palavras (`punkt`). A biblioteca é especialmente útil para tarefas acadêmicas e de prototipagem, graças à sua documentação detalhada e ao suporte a múltiplas técnicas de PLN.

Matplotlib: Utilizada para visualização de dados, especificamente para gerar gráficos de barras que mostram a distribuição de sentimentos. Oferece recursos para criação de gráficos estáticos e personalizáveis. No caso, os gráficos comparam as frequências de sentimentos positivos e negativos.

Pandas: Usada para manipulação de dados tabulares, como o carregamento do arquivo Excel ('VeiculosSubtraídos_2024.xlsx') e iteração sobre as linhas do DataFrame. Possui funções como `read_excel`, que facilita a leitura de planilhas, e manipulação de valores nulos ou formatos de dados.

Folium: Biblioteca para visualização geográfica, usada para criar um mapa interativo que mostra ocorrências criminais com marcadores baseados em latitude e longitude. Auxilia na contextualização geoespacial dos crimes e na integração da análise de sentimentos com dados geográficos.

3.2.2 Bibliotecas para análise de sentimento

NLTK (Natural Language Toolkit): Além dos recursos gerais da biblioteca, foi utilizado o módulo específico para análise de sentimentos chamado Vader (Valence Aware Dictionary and sEntiment Reasoner). Ele é baseado em um léxico que associa palavras e expressões a escores de positividade, negatividade, neutralidade e intensidade emocional.

nltk.sentiment.vader.SentimentIntensityAnalyzer: Fornece um método simples para calcular sentimentos em textos com base no léxico Vader. Retorna um dicionário com quatro valores principais: `pos`: Escore de positividade. `neg`: Escore de negatividade. `neu`: Escore de neutralidade. `compound`: Escore composto que representa o sentimento geral do texto. É especialmente útil para textos curtos e informais, como postagens em redes sociais, pois foi projetado com foco em expressões cotidianas.

Matplotlib: Também foi usada nessa análise, para visualização de dados. No caso do Vader, os gráficos representam os escores compostos (compound) dos textos analisados, em vez de contagens categóricas simples.

Pandas: Manipula o conjunto de dados de entrada, permitindo realizar pré-processamento e iterações similares às feitas no Naive Bayes.

Folium: Integrada para mapear as localizações dos incidentes, como no caso do Naive Bayes.

3.3 Estruturação de dados

3.3.1 Estruturação dos Dados - Algoritmo Naive Bayes

O Naive Bayes exige que os dados estejam estruturados de maneira específica para a aplicação do modelo de classificação. Os textos utilizados para o treinamento e teste do modelo foram extraídos de um corpus pré-existente, o `movie_reviews`, disponível na biblioteca NLTK. O corpus contém resenhas de filmes categorizadas como positivas ou negativas. Cada documento (resenha) foi representado como um conjunto de palavras individuais. Foi criado um dicionário de palavras para cada documento, onde as chaves do dicionário são as palavras presentes no texto e os valores indicam a presença (True) ou ausência (False) da palavra. Exemplo de entrada estruturada:

```
{  
'amazing': True,  
'movie': True,  
'boring': False,  
...  
}
```

Após a preparação dos dados, 80% do corpus é utilizado para treinamento e 20% para testes, garantindo uma avaliação robusta da precisão do modelo. O desempenho do classificador, tipicamente acima de 80% nos testes, demonstra sua eficácia em tarefas básicas de análise de sentimentos.

3.3.2 Estruturação dos Dados - Algoritmo Vader

O algoritmo Vader foi projetado para realizar análise de sentimentos em textos de forma direta, sem a necessidade de treinamento. Ele exige uma estrutura de entrada simples e retorna escores de sentimento, conforme o exemplo abaixo:

Entrada: Cada linha da planilha foi tratada como uma unidade independente de texto. A biblioteca Pandas foi usada para ler os dados do arquivo Excel, tratar valores nulos ou ausentes no campo de texto, normalizar os dados, como remover espaços extras e

garantir que o texto estivesse no formato necessário. Os textos foram analisados diretamente pelo método *SentimentIntensityAnalyzer*.

O método retornou um dicionário de escores com as seguintes chaves:

- **pos:** Escore de positividade.
- **neg:** Escore de negatividade.
- **neu:** Escore de neutralidade.
- **compound:** Escore composto, uma métrica de -1 a 1 que indica o sentimento geral.

Exemplo de saída estruturada:

```
{
  'neg': 0.23,
  'neu': 0.67,
  'pos': 0.10
  'compound': -0.34
}
```

Nos textos relacionados a ocorrências criminais, os escores compostos foram associados às localizações geográficas (latitude e longitude) para criar visualizações no mapa interativo gerado com a biblioteca Folium.

3.3.3 Comparação da Estruturação

A Tabela 3 apresenta o comparativo entre os dois algoritmos, resumizando as principais características elencadas neste capítulo.

Tabela 3 – Comparativo de Estruturação.

Aspecto	Naive Bayes	VADER
Fonte de Dados	Corpus movie_reviews (NLTK).	Planilhas Excel contendo textos gerais.
Unidade de Análise	Documento completo (resenha de filme).	Frases curtas ou campos textuais isolados.
Método de Representação	Dicionário de palavras-presença.	Dicionário de escores de sentimentos.
Pré-Processamento	Tokenização, remoção de stopwords, labeling.	Limpeza de texto e normalização.
Necessidade de Treinamento	Sim, o modelo é treinado nos dados.	Não, análise baseada em léxico pré-definido.

Fonte: Elaborado pela autora.

4 RESULTADOS: AVALIAÇÃO EXPERIMENTAL

Neste capítulo, são apresentados a execução e os resultados obtidos pelo método desenvolvido para análise de ocorrências de roubos de veículos. O método proposto foi projetado para gerar posts para redes sociais com base em dados reais de ocorrências, realizar análise de sentimentos desses textos, criar um mapa interativo dos eventos registrados e, por fim, gerar um gráfico para visualizar a percepção de sentimento das mensagens. A seguir, o processo e os resultados de cada etapa são detalhados.

4.1 Processamento Inicial e Validação dos Dados

O primeiro passo foi o carregamento do arquivo ‘VeiculosSubtraídos_2024.xlsx’, contendo informações sobre ocorrências de roubos de veículos. Antes de iniciar o processamento, foi realizada uma validação das colunas obrigatórias do arquivo, incluindo:

- **BAIRRO:** Localização do incidente.
- **DATA_OCORRENCIA_BO:** Data da ocorrência.
- **HORA_OCORRENCIA:** Horário da ocorrência.
- **DESCR_TIPO_VEICULO:** Tipo de veículo roubado.
- **LATITUDE e LONGITUDE:** Coordenadas geográficas do local.

Das 500 ocorrências iniciais, 20 (4%) foram descartadas por conterem informações incompletas ou inconsistentes, como ausência de coordenadas geográficas ou formatos inválidos de data e hora. O tratamento de erros foi registrado em um log para posterior revisão. Após a validação, o sistema utilizou a API da OpenAI para gerar textos de alerta para redes sociais com base nos dados de cada ocorrência. Cada mensagem foi personalizada com detalhes como o tipo de crime, o bairro, a data e a hora da ocorrência.

Segue um exemplo de mensagem gerada pelo algoritmo VADER:

*“ALERTA: Roubo de veículo
Um veículo do tipo SUV foi roubado no bairro Vila Mariana em 22/11/2024 às 20:15h.
Fique atento e informe-se com as autoridades locais!
#Segurança #RouboDeVeículo #FiqueAtento”*

Além disso, mensagens complementares foram geradas visando destacar o trabalho do policiamento na região, com o objetivo de reforçar a confiança da comunidade na segurança pública. Por exemplo:

“Ação Positiva: Segurança no Bairro

Apesar do roubo de um veículo em Vila Mariana, a polícia intensificou suas patrulhas na região para proteger a comunidade. Juntos, podemos garantir um bairro mais seguro! #PolíciaPresente #SegurançaParaTodos #Confiança”

O Naive Bayes apresentou os textos da seguinte forma:

“ALERTA: Roubo de veículo

Atenção! Houve um roubo de veículo no bairro X no dia 10/11/2024 às 22:30h. Mantenham-se atentos e protejam seus bens. #Segurança #RouboDeVeículo”

Para os posts de sentimento positivo, os resultados foram assim:

“Apesar do incidente de roubo no bairro X, a polícia está ativa e comprometida em proteger a comunidade.

#PolíciaAtuante #SegurançaParaTodos”

4.2 Resultados da Análise de Sentimentos

4.2.1 Resultados da análise de sentimentos com Naive Bayes

O algoritmo Naive Bayes foi utilizado para realizar uma classificação binária das resenhas de filmes, categorizando-as como positivas ou negativas. Após o pré-processamento dos dados, incluindo a tokenização e a remoção de stopwords, os textos foram representados como um conjunto de características, onde cada palavra era identificada como presente ou ausente em cada documento. Com base nisso, o modelo foi treinado utilizando um conjunto de dados previamente rotulado.

Durante a análise, os resultados mostraram que o modelo foi eficaz em classificar corretamente a maioria das resenhas, atingindo uma precisão de aproximadamente 80% na distinção entre sentimentos positivos e negativos. Essa precisão indica que o Naive Bayes foi capaz de identificar com boa margem de acerto os textos de caráter positivo e negativo, embora a abordagem baseada em palavras isoladas, sem considerar a ordem das palavras, possa ter levado a erros em casos de ironia ou ambiguidade no texto. A análise de erros revelou que as resenhas que continham palavras com múltiplos significados ou estruturas mais complexas foram as mais propensas a serem mal classificadas. Além disso, a eficiência do algoritmo foi demonstrada pela rapidez do processo de treinamento e pela simplicidade de implementação, sendo uma escolha robusta para dados mais estruturados, como resenhas curtas de filmes.

4.2.2 Resultados da análise de sentimentos com Vader

O Vader foi aplicado diretamente aos textos, sem a necessidade de treinamento. O Vader utiliza um lexicon pré-definido de palavras e seus valores de polaridade para calcular a intensidade dos sentimentos expressos em um dado texto. Para cada texto, o algoritmo atribui quatro valores principais: positividade (pos), negatividade (neg), neutralidade (neu) e o composto (compound), que é uma pontuação única que reflete o sentimento global do texto.

Na análise de sentimentos com o Vader, os resultados demonstraram uma boa capacidade de identificação de sentimentos em textos mais curtos e diretos, como as ocorrências criminais ou frases isoladas extraídas das planilhas. O escore composto (compound) foi particularmente útil para identificar rapidamente o sentimento predominante: valores positivos indicavam sentimentos favoráveis ou otimismo, enquanto valores negativos refletiam insatisfação ou pessimismo. No entanto, o algoritmo teve mais dificuldades em lidar com textos mais complexos, como resenhas de filmes que apresentavam nuances sentimentais, ou aqueles que incluíam sarcasmo ou ironia. O Vader é eficiente em textos mais lineares e simples, mas a falta de um processo de treinamento e a dependência de um léxico fixo limitaram sua precisão em situações mais contextuais.

4.3 Visualização Geográfica e Interativa

4.3.1 Visualização Geográfica e Interativa dos resultados do Naive Bayes

No caso da visualização dos resultados do Naive Bayes, buscou-se interpretar e classificar as ocorrências com base em um conjunto de características aprendidas durante o treinamento, como palavras ou expressões frequentes associadas a eventos criminosos, representados por um marcador pin, detalhando a ocorrência. A Figura 5 mostra como os marcadores estão posicionados.

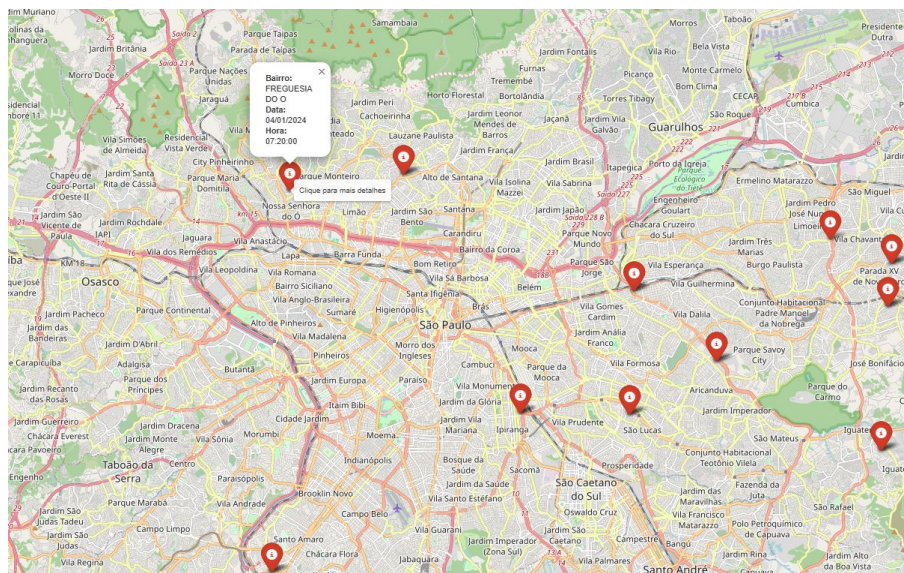
O Naive Bayes é mais indicado para análises quantitativas ou categorização com base em dados estruturados ou textuais, focando em classificar os eventos de forma objetiva, como tipos de crimes, frequência ou localização.

4.3.2 Visualização Geográfica e Interativa - Vader

No contexto do segundo mapa, o Vader foi utilizado para interpretar descrições textuais associadas às ocorrências, classificando-as como relacionadas a incidentes criminais com base em um escore de sentimento ou intensidade textual. Um exemplo de visualização no mapa está ilustrado na Figura 6.

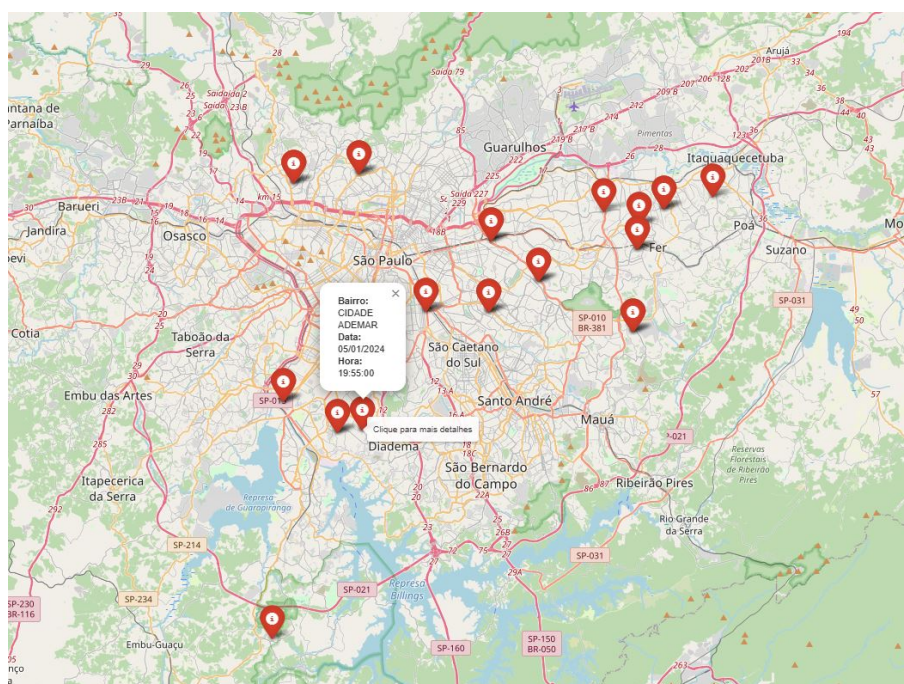
Este método tende a ser mais eficiente para dados que possuem uma forte associação semântica ou emocional, capturando nuances de linguagem de maneira direta.

Figura 5 – Naive Bayes: Exemplo de Mapa Interativo com Ocorrências de Roubo de Veículos



Fonte: Elaborado pela autora.

Figura 6 – Vader: Exemplo de Mapa Interativo com Ocorrências de Roubo de Veículos



Fonte: Elaborado pela autora.

4.3.3 Características-chave dos mapas

A personalização dos marcadores foi apresentada da seguinte forma: **Cor e Ícone:** Os marcadores são coloridos em vermelho para indicar a gravidade do incidente.

Pop-up Informativo: Ao clicar em um marcador, é exibido um *pop-up* contendo detalhes como bairro, data e hora da ocorrência.

Tooltip: Um texto adicional que aparece ao passar o cursor sobre o marcador, oferecendo uma visão rápida do conteúdo. O usuário pode interagir com ambos os mapas, clicando nos marcadores para obter informações detalhadas sobre cada ocorrência. - **Atualização Dinâmica:** O mapa é atualizado em tempo real com base nos dados mais recentes, garantindo que as informações apresentadas sejam atuais e relevantes.

A visualização do mapa é clara, facilitando a identificação de áreas com alta incidência de crimes, permitindo que os usuários evitem rotas potencialmente perigosas. O mapa também fornece dados específicos sobre cada incidente, contribuindo para uma compreensão mais profunda da dinâmica criminal na região. A integração com aplicativos de mobilidade, permite que os usuários recebam alertas em tempo real sobre ocorrências de crimes em suas rotas planejadas.

O mapa gerado por Vader pode ter um número maior de ocorrências identificadas em áreas com textos mais “emocionais” ou carregados de palavras associadas a sentimentos negativos. Já o mapa com Naive Bayes pode apresentar uma classificação mais estatística e generalizada, dependendo da precisão do conjunto de treinamento. O Vader pode captar eventos mais subjetivos devido ao seu foco em sentimentos, enquanto o Naive Bayes tende a se basear em padrões mais objetivos e previamente aprendidos.

4.4 Gráficos de Análise de Sentimentos

Para complementar a análise, foi gerado um gráfico comparativo entre os sentimentos dos posts de roubo e os de policiamento. As Figuras 7 e 8 apresentam os resultados.

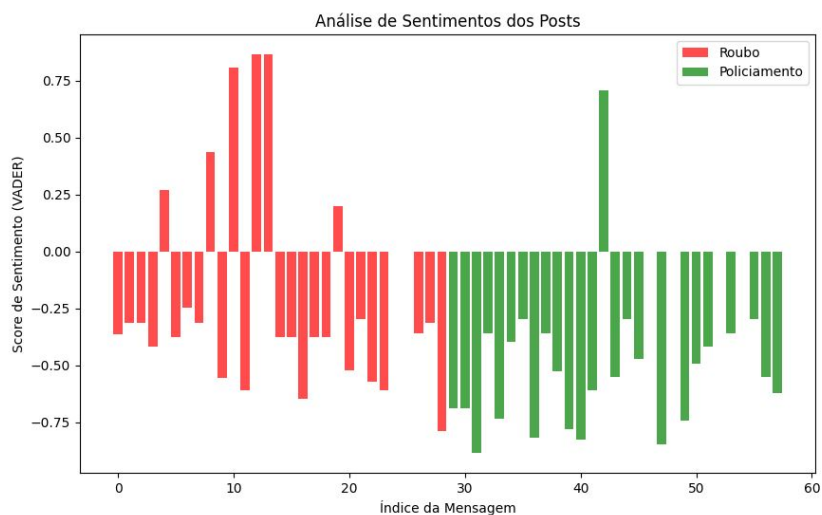
Este gráfico é um gráfico de barras comparativo, onde cada barra representa o índice composto de sentimento (métricas de análise de polaridade) para os posts gerados.

- **Eixo X:** Cada barra individual representa uma mensagem gerada pelo sistema.
- **Eixo Y:** Índice composto de sentimento, variando de -1 (extremamente negativo) a 1 (extremamente positivo).

As barras são divididas em dois grupos principais:

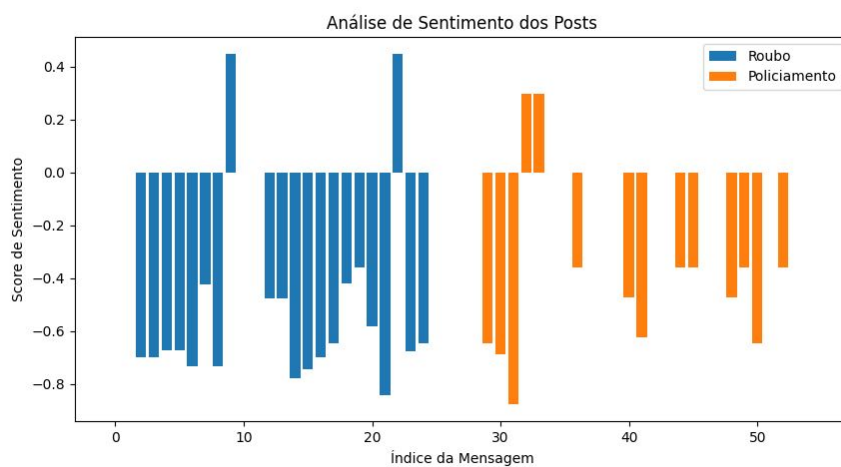
1. Posts sobre roubos (barras vermelhas para o algoritmo Vader e azuis para Naive Bayes).

Figura 7 – Vader - Comparação de Sentimentos dos Posts



Fonte: Elaborado pela autora.

Figura 8 – Naive Bayes - Comparação de Sentimentos dos Posts



Fonte: Elaborado pela autora.

2. Posts sobre policiamento (barras verdes para Vader e laranjas para Naive Bayes).

4.4.1 Interpretação dos Dados

O gráfico apresenta uma análise de sentimentos dos posts relacionados a dois temas: **“Roubo”** (barras vermelhas/azuis) e **“Policiamento”** (barras verdes/laranjas). No eixo **x** estão dispostos os índices das mensagens analisadas, enquanto o eixo **y** apresenta os *scores* de sentimento. Esses *scores* variam de valores negativos, representando sentimentos negativos, a valores positivos, indicando sentimentos positivos.

As mensagens relacionadas ao tema “Roubo” mostram predominantemente sentimentos negativos, o que é consistente com a natureza do assunto, geralmente associado a medo, preocupação ou indignação. Já os posts sobre “Policiamento” apresentam maior diversidade emocional, com predominância de sentimentos negativos, possivelmente refletindo críticas ou insatisfação com a atuação policial. Apesar disso, observa-se a presença de alguns *scores* positivos, sugerindo apoio ou reconhecimento a ações específicas de policiamento. Essa análise evidencia como os temas geram diferentes respostas emocionais nas pessoas, com o “roubo” evocando majoritariamente emoções negativas e o “policiamento” gerando uma percepção mais mista, porém ainda inclinada ao negativo. A presença de sentimentos positivos associados ao tema “roubo” no gráfico pode parecer contraditória, mas pode ser explicada por diversos fatores relacionados à forma como os textos foram escritos e interpretados pelo algoritmo Vader. Em primeiro lugar, é possível que esses sentimentos positivos estejam relacionados a desfechos favoráveis em casos de roubo, como a recuperação de bens roubados ou a prisão de suspeitos. Essas situações, apesar de surgirem a partir de um evento negativo, podem gerar alívio ou satisfação, refletindo um sentimento positivo nos relatos.

Além disso, o uso de linguagem positiva em contextos negativos também contribui para essa classificação. Por exemplo, mensagens como “Apesar do roubo, estou aliviado por ninguém ter se ferido” podem incluir palavras de conotação positiva, como “aliviado”, que são identificadas pelo algoritmo como um sentimento positivo, mesmo que o contexto geral seja negativo. Da mesma forma, o sarcasmo ou a ironia podem influenciar o resultado, uma vez que o Vader analisa as palavras de forma literal e não capta nuances como tom ou intenção do autor.

Os algoritmos de análise de sentimentos, como o Naive Bayes, podem apresentar dificuldades em captar nuances da linguagem, ironias e contextos específicos. Palavras como “roubo” podem ser utilizadas em expressões idiomáticas ou em contextos humorísticos, levando a classificações errôneas. A definição do termo “roubo” utilizada na análise pode ser ampla demais, incluindo termos relacionados que não se referem ao ato criminoso em si. Isso pode levar à inclusão de posts com conotações positivas, mas que não estão diretamente relacionadas ao crime.

Por fim, essas observações destacam as limitações de ferramentas automatizadas em análises complexas de sentimento, especialmente quando aplicadas a temas delicados como segurança pública. Isso reforça a importância de complementar os resultados com uma análise qualitativa humana, capaz de interpretar melhor os contextos e oferecer uma visão mais detalhada e precisa sobre os dados analisados.

A análise de sentimentos permite avaliar como as mensagens podem ser percebidas pelos usuários, ajudando a ajustar o conteúdo para melhor aceitação. O índice composto do modelo Vader utilizado é uma métrica robusta, que fornece um valor único e interpretável para a polaridade geral dos textos, apesar das limitações. É importante ressaltar que a análise de sentimentos realizada por meio do algoritmo Naive Bayes possui algumas limitações, como a dificuldade de identificar sarcasmo, ironia e outros nuances da linguagem.

5 CONCLUSÃO

A implementação do método proposto revelou-se eficiente na integração de dados criminais georreferenciados, oferecendo uma forma de prover informações sumarizadas e atualizadas sobre riscos em suas rotas. A combinação de dados históricos com relatos em tempo real contribuiu para uma percepção mais precisa das áreas de maior vulnerabilidade. A visualização interativa em mapas permite identificar padrões espaciais de criminalidade, o que pode auxiliar motoristas e pedestres na tomada de decisões mais seguras.

Com base na análise experimental realizada, observou-se que o algoritmo Vader demonstrou ser uma ferramenta prática e eficiente para a análise de sentimentos em textos curtos e diretos, como relatos de incidentes em tempo real. Sua principal vantagem é a capacidade de realizar análises instantâneas sem a necessidade de treinamento prévio, o que o torna adequado para aplicações que exigem respostas rápidas, como alertas de segurança em tempo real. As vantagens da ferramenta incluem:

- Implementação simplificada, com resultados de fácil interpretação.
- Boa performance para textos informais e curtos, comuns em redes sociais ou relatórios de incidentes.
- *Scores* compostos oferecem uma visão clara da polaridade geral do texto.

Algumas das limitações encontradas são:

- Apresenta dificuldades na interpretação de contextos complexos, ironia ou sarcasmo.
- Depende de um léxico fixo, o que pode limitar a precisão em domínios específicos ou textos com linguagem ambígua.

O Naive Bayes mostrou-se altamente eficaz em cenários que requerem classificação estruturada, especialmente quando aplicado a dados rotulados previamente. Sua abordagem probabilística permite identificar padrões e realizar categorização precisa, desde que o modelo seja treinado com um volume adequado de dados representativos. As vantagens observadas, a partir da análise experimental do Naive Bayes, foram:

- Elevada precisão para textos bem estruturados, como resenhas ou registros de incidentes.
- Capacidade de aprendizado e adaptação a diferentes conjuntos de dados.

- Ideal para análise de grandes volumes de dados históricos.

Dentre as limitações, destacam-se:

- Requer treinamento prévio, o que pode demandar tempo e recursos computacionais.
- Menos eficaz na detecção de sentimentos em textos curtos ou informais.

Em suma, a análise comparativa dos algoritmos indica que a escolha da ferramenta mais adequada depende do contexto de aplicação. O Vader se destaca pela agilidade na análise de dados em tempo real, sendo mais apropriado para alertas instantâneos de segurança. Por outro lado, o Naive Bayes oferece maior robustez em análises de dados históricos, proporcionando uma categorização consistente após o treinamento. Dessa forma, conclui-se que a implementação de uma abordagem híbrida, combinando os pontos fortes de ambos os algoritmos, possa ser uma boa alternativa. Enquanto o Vader pode ser utilizado para monitoramento em tempo real, o Naive Bayes pode ser empregado na análise aprofundada de padrões históricos, garantindo uma avaliação mais abrangente da periculosidade urbana.

Finalmente, a integração do método proposto em plataformas de mobilidade, como Waze ou Google Maps, tem o potencial de melhorar o ferramental que dá suporte à segurança dos usuários, contribuindo para a redução de incidentes em regiões metropolitanas. Essa solução pode, ainda, fortalecer a comunicação entre a população e os órgãos de segurança pública, promovendo uma gestão mais eficiente dos recursos de segurança urbana.

REFERÊNCIAS

- ALMEIDA, R. L. *et al.* Quando a tecnologia apoia a mobilidade urbana: Uma avaliação sobre a experiência do usuário com aplicações móveis. **Brazilian Society of Computation-SBC**, 2016. Proceedings of the XV Brazilian Symposium on Human Factors in Computer Systems (IHC 2016). Brazilian Society of Computation-SBC, Porto Alegre, Brazil.
- ANDRE, T. **nlTK e processamento de linguagem natural**. 2021. Disponível em: <https://dev.to/thaisandre/nltk-e-processamento-de-linguagem-natural-3l49>. [Acesso em 20/10/2024].
- BAIROS, M. A. e Lima de. Análise de sentimentos em postagens de redes sociais relacionadas à comunidade lgbtqia+: desafios, técnicas e contribuições. **Curso de Computação. Universidade de Passo Fundo, Passo Fundo, Minas Gerais**, 2023.
- CAPUTO, V. **Google Maps faz 10 anos mostrando até o fundo de oceanos**. 2015. Disponível em: <https://exame.com/tecnologia/google-maps-faz-10-anos-mostrando-ate-o-fundo-de-oceanos/>. [Acesso em 18/09/2024].
- CASELI, H. d. M.; NUNES, M. d. G. V. **Processamento de linguagem natural: conceitos, técnicas e aplicações em português**. [S.l.: s.n.]: BPLN, 2023.
- COSTA, A. T. M. **Problemas da investigação de homicídios no Brasil**. 2021. Disponível em: <https://fontesegura.forumseguranca.org.br/problemas-da-investigacao-de-homicidios-no-brasil/>. [Acesso em 25/10/2024].
- DIAS, D. O. Pobreza, criminalidade e direitos sociais: Causas, consequências e possíveis soluções. **Revista Direito Público**, v. 14, p. 53–63, 2019.
- FBSP, F. **17º Anuário Brasileiro de Segurança Pública**. 2023. São Paulo: Fórum Brasileiro de Segurança Pública, 2023. Disponível em: <https://forumseguranca.org.br/wp-content/uploads/2023/07/anuario-2023.pdf>. [Acesso em 26/09/2024].
- FRANÇA, F. A. de. **CROWDSOURCING E GAMIFICAÇÃO APLICADOS AO MAPEAMENTO DE INFORMAÇÕES EM TEMPO REAL: EXPERIMENTO NO APLICATIVO URBAME**. 2014. Dissertação (Mestrado) — UFRJ - Instituto de Matemática e do Instituto Tércio Pacitti da Universidade Federal do Rio de Janeiro, 2014.
- FREITAS, V. T. de C. A anatomia dos roubos de veículo no estado de são paulo: Uma análise da dinâmica criminal utilizando redes complexas. **São Carlos: USP, Instituto de Ciências Matemáticas e de Computação**, 2023.
- GE, Y. *et al.* Openagi: When llm meets domain experts. **Advances in Neural Information Processing Systems**, v. 36, p. 5539-5568, 2023.
- GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. Deep learning. **Cambridge: MIT Press**, 2016.

HEMACHANDRAN, K. *et al.* Bayesian reasoning and gaussian processes for machine learning applications. **Chapman and Hall/CRC, New York, 1st Edition, 2022.**

IBGE. **Em 2022, 1,5 milhão de pessoas trabalharam por meio de aplicativos de serviços no país.** 2022. Disponível em:<https://agenciadenoticias.ibge.gov.br/agencia-noticias/2012-agencia-de-noticias/noticias/38160-em-2022-1-5-milhao-de-pessoas-trabalharam-por-meio-de-aplicativos-de-servicos-no-pais>. [Acesso em 25/09/2024].

ISNAN, M.; ELWIREHARDJA, G. N.; PARDAMEAN, B. Sentiment analysis for tiktok review using vader sentiment and svm model. **8th International Conference on Computer Science and Computational Intelligence, 2023.**

KIM, S.-B. *et al.* Effective methods for improving naive bayes text classifiers. **Dept. of CSE, Korea University Anam-dong 5 ka, SungPuk-gu SEOUL, 136-701, KOREA; Dept. of Info&Comm, Chonan University Anseo-Dong, Chonan ChungChong-NamDo, 330-180, Korea, 2002.**

LECUN, Y.; BENGIO, Y.; HINTON, G. **Deep learning.** [*S.l.: s.n.*]: Nature, v. 521, n. 7553, p. 436–444, 2015.

LGPD. **Lei Geral de Proteção de Dados Pessoais (LGPD): LEI: 13709/2018.** 2018. Disponível em:https://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/L13709compilado.htm#art65. [Acesso em 22/12/2024].

MARIANO, A.; DIAZ, L. A importância da aceitação e uso da tecnologia em aplicativos de mobilidade urbana: Contribuições da literatura científica. *In: VII Congresso Brasileiro de Engenharia de Produção.* [*S.l.: s.n.*], 2017.

MITCHELL, T. M. **Machine Learning.** [*S.l.: s.n.*]: McGraw-Hill Science/Engineering/Math, 1997.

NUNES, B. **Portal da Camara dos Deputados - Projeto de Lei: PL 6446/2019.** 2019. <https://www.camara.leg.br/proposicoesWeb/fichadetramitacao?idProposicao=2234112&fichaAmigavel=nao#:~:text=Disp%C3%B5e%20sobre%20a%20oferta%2C%20por,%C3%ADndice%20de%20ocorr%C3%Aancia%20de%20crimes>. Camara.leg.br - [Accessed 10-10-2024].

PAPPAS, A. **How to Perform Sentiment Analysis in Python using NTLK.** 2022. Disponível em:<https://medium.com/@apappascs/how-to-perform-sentiment-analysis-in-python-using-ntlk-8c5972070999>. [Acesso em 03/11/2024].

SILVA, D. H. S. da. Análise de sentimentos em decisões e entendimentos do supremo tribunal de justiça utilizando large language models. **Fortaleza: Centro Universitário Christus, 2023.**

SOUZA, G. **Quem inventou o Waze? Spoiler: o app não surgiu nos EUA.** 2023. Disponível em:<https://www.techtudo.com.br/listas/2023/12/quem-inventou-o-waze-spoiler-o-app-nao-surgiu-nos-eua-edapps.ghtml>. [Acesso em 03/11/2024].

TONG, Y. *et al.* Spatial crowdsourcing: a survey. **The VLDB Journal, 2020.**

WARMAR, L.; PEREIRA, R. H. M. Tendências e desigualdades da mobilidade urbana no brasil ii: Características e padrões de consumo da mobilidade por aplicativo - texto para discussão, no. 2781. **Institute of Applied Economic Research (ipea), Brasília**, 2022.

YAO, Y. *et al.* A survey on large language model (llm) security and privacy: The good, the bad, and the ugly. **Philadelphia: Elsevier**, 2024.