TERMÔMETRO DE SENTIMENTOS: DO MONITORAMENTO REATIVO À GESTÃO PROATIVA DE REPUTAÇÃO COM PROCESSAMENTO DE LINGUAGEM NATURAL

121 SENTIMENT THERMOMETER: FROM REACTIVE MONITORING TO PROACTIVE REPUTATION MANAGEMENT USING NATURAL LANGUAGE PROCESSING

DANIEL ZULAR
INSTITUTO DE TECNOLOGIA E LIDERANÇA - INTELI

MARCELO LUIZ DO AMARAL GONÇALVES
INSTITUTO DE TECNOLOGIA E LIDERANÇA - INTELI

LUCAS DA SILVA BARBOSA
INSTITUTO DE TECNOLOGIA E LIDERANÇA - INTELI

PEDRO FARIA SANTOS
INSTITUTO DE TECNOLOGIA E LIDERANÇA - INTELI

Comunicação:
O XIV SINGEP foi realizado em conjunto com a 14th Conferência Internacional do CIK (CYRUS Institute of Knowledge) e com a SPARK (Society for Promoting Academic Research and Knowledge), em formato híbrido, com sede presencial na UNINOVE - Universidade Nove de Julho, no Brasil.

Agradecimento à orgão de fomento:
Agradecemos ao Instituto de Tecnologia e Liderança pelo apoio à pesquisa e inovação.

TERMÔMETRO DE SENTIMENTOS: DO MONITORAMENTO REATIVO À GESTÃO PROATIVA DE REPUTAÇÃO COM PROCESSAMENTO DE LINGUAGEM NATURAL

Objetivo do estudo
Desenvolver uma solução de Processamento de Linguagem Natural para classificar automaticamente comentários no X em sentimentos positivos, negativos ou neutros. O sistema integra-se ao Slack para automatizar alertas e viabilizar uma gestão reputacional proativa..

Relevância/originalidade
Soluciona os gargalos do monitoramento reputacional manual e terceirizado com inteligência artificial aplicada a dados reais altamente desbalanceados. Inova ao integrar-se diretamente ao Slack (Nexus Bot), eliminando barreiras organizacionais para adoção prática diária da tecnologia.

Metodologia/abordagem
Pesquisa qualitativa e aplicada estruturada sob o protocolo CIMO. O desenvolvimento tecnológico utilizou Scrum em cinco sprints, realizando pré-processamento textual, representação densa por Word2Vec e testes comparativos de múltiplos algoritmos de classificação supervisionada.

Principais resultados
O modelo combinando Word2Vec, Regressão Logística e oversampling atingiu 86% de acurácia, 0,80 de F1-Score e 0,78 de recall, sem overfitting. A entrega gerou duas ferramentas integradas: um painel Front-end interativo e o Nexus Bot no Slack.

Contribuições teóricas/metodológicas
Valida empiricamente que a representação vetorial densa (Word2Vec) supera limitações de matrizes esparsas (Bag of Words) em redes sociais. Consolida a utilidade do protocolo CIMO para relatos de produção técnica organizacional aplicada.

Contribuições sociais/para a gestão
Permite uma gestão reputacional proativa e independente, eliminando a lentidão e custos de relatórios mensais terceirizados. Demonstra que projetos de IA corporativos ganham efetividade quando desenvolvidos de forma ágil e integrados às ferramentas de trabalho cotidiano.

Palavras-chave: Análise de Sentimentos, Processamento de Linguagem Natural, Monitoramento de Redes Sociais, Gestão de Projetos, Relato Técnico

121 SENTIMENT THERMOMETER: FROM REACTIVE MONITORING TO PROACTIVE REPUTATION MANAGEMENT USING NATURAL LANGUAGE PROCESSING

Study purpose
To develop a Natural Language Processing solution to automatically classify X (formerly Twitter) comments as positive, negative, or neutral in sentiment. The system integrates with Slack to automate alerts and enable proactive reputation management.

Relevance / originality
It addresses the bottlenecks associated with manual and outsourced reputation monitoring by applying artificial intelligence to highly imbalanced real-world data. It innovates through direct integration with Slack (Nexus Bot), removing organizational barriers to the technology's practical, day-to-day adoption.

Methodology / approach
Applied qualitative research structured according to the CIMO protocol. Technical development employed the Scrum framework across five sprints, involving text preprocessing, dense representation via Word2Vec, and comparative testing of multiple supervised classification algorithms.

Main results
The model—combining Word2Vec, Logistic Regression, and oversampling—achieved 86% accuracy, an F1-Score of 0.80, and a recall of 0.78, without overfitting. The project delivered two integrated tools: an interactive front-end dashboard and the Nexus Bot for Slack.

Theoretical / methodological contributions
It empirically validates that dense vector representation (Word2Vec) overcomes the limitations of sparse matrices (Bag of Words) in the context of social media. It reinforces the utility of the CIMO protocol for reporting on applied organizational technical development.

Social / management contributions
It enables proactive, independent reputation management, eliminating the delays and costs associated with outsourced monthly reports. It demonstrates that corporate AI projects are more effective when developed using agile methods and integrated into daily workflow tools.

Keywords: Sentiment Analysis, Natural Language Processing, Social Media Monitoring, Project Management, Technical Report

TERMÔMETRO DE SENTIMENTOS: DO MONITORAMENTO REATIVO À GESTÃO PROATIVA DE REPUTAÇÃO COM PROCESSAMENTO DE LINGUAGEM NATURAL

1. Introdução

A organização identificada neste relato como ETM, empresa de tecnologia de mobilidade de grande porte, conduzia o monitoramento dos comentários de usuários em redes sociais de forma terceirizada, manual e com periodicidade mensal. Esse modelo apresentava escalabilidade reduzida e ciclos longos de atualização, o que limitava a identificação de tendências negativas antes que se tornassem críticas. Em decorrência, a empresa atuava de forma reativa diante de reclamações e episódios de crise de imagem, sem mecanismo para categorizar os comentários de modo padronizado.

Diante desse cenário, o objetivo do projeto foi desenvolver uma solução de Processamento de Linguagem Natural capaz de classificar automaticamente comentários de usuários na rede social X em sentimentos positivos, negativos ou neutros e de gerar alertas automáticos integrados à plataforma de comunicação corporativa Slack. A intervenção caracterizou-se como força-tarefa de desenvolvimento, conduzida por equipe de sete estudantes de graduação em parceria com a ETM, sob metodologia ágil, em cinco sprints de duas semanas, com duração total de dez semanas.

O relato foi construído com base na participação direta dos autores no projeto, que integraram a equipe responsável pela concepção, implementação e entrega da solução. Essa participação possibilitou a observação contínua das etapas, do entendimento do negócio à integração da solução com a plataforma de comunicação do parceiro. As informações utilizadas compreendem fontes primárias, como o documento de projeto, a documentação técnica das sprints e os resultados de desempenho do modelo, e fontes secundárias, correspondentes à literatura empregada no referencial teórico.

O relato está organizado em seis seções além desta introdução: referencial teórico, método de produção técnica, contexto do projeto, tipo de intervenção e mecanismos adotados, apresentação dos resultados e análise, e conclusão.

2. Referencial Teórico

Esta seção apresenta os principais conceitos teóricos que fundamentam a intervenção descrita neste relato: a análise de sentimentos como subárea do Processamento de Linguagem Natural (PLN), as técnicas de vetorização de texto aplicadas, os modelos de aprendizado de máquina utilizados para classificação, o papel estratégico do monitoramento de redes sociais na gestão da reputação organizacional e a abordagem de gerenciamento de projetos adotada.

2.1 Processamento de Linguagem Natural e Análise de Sentimentos

A análise de sentimentos é uma subárea do Processamento de Linguagem Natural voltada à identificação e à classificação de opiniões, emoções e posicionamentos expressos em texto (Liu, 2012; Wankhade et al., 2022).

O crescimento das plataformas de redes sociais ampliou o volume de conteúdo gerado por usuários e motivou o desenvolvimento de métodos automatizados para o tratamento desse conteúdo (Giachanou & Crestani, 2016). A tarefa de classificação organiza os textos em categorias de polaridade, definidas com frequência como positiva, negativa e neutra (Wankhade et al., 2022). No contexto do Twitter, atual rede social X, a informalidade da linguagem, o uso de gírias e a extensão reduzida das mensagens constituem fatores que afetam o desempenho dos classificadores (Giachanou & Crestani, 2016; Naseem et al., 2020).

2.2 Técnicas de Pré-Processamento Textual

O pré-processamento antecede a representação vetorial e a classificação, com a função de reduzir ruído e padronizar o texto (Manning, Raghavan, & Schütze, 2008). Etapas como segmentação de frases, remoção de links, tokenização, remoção de stop-words e lematização compõem o conjunto de técnicas aplicadas a dados de redes sociais (Symeonidis, Effrosynidis, & Arampatzis, 2018). Symeonidis et al. (2018) avaliaram dezesseis técnicas de pré-processamento em dois conjuntos de dados do Twitter e registraram que lematização, remoção de números e substituição de palavras alongadas se associam a aumento de acurácia, enquanto outras técnicas apresentam efeito reduzido. A seleção das etapas depende das características do conjunto de dados analisados (Symeonidis et al., 2018).

2.3 Vetorização: Bag of Words e Word2Vec

A vetorização converte texto em representações numéricas processáveis por algoritmos de aprendizado de máquina. O modelo Bag of Words representa documentos pela frequência de termos, o que gera matrizes esparsas de alta dimensionalidade e não captura relações semânticas entre palavras (Manning et al., 2008). O Word2Vec representa palavras em vetores densos cujas distâncias refletem similaridade semântica e sintática, a partir do treinamento sobre os contextos de ocorrência (Mikolov et al., 2013). Essa propriedade aplica-se a tarefas de classificação de sentimentos em linguagem informal, nas quais o significado depende do contexto de uso (Mikolov et al., 2013).

2.4 Modelos de Classificação

A classificação de sentimentos por aprendizado supervisionado emprega algoritmos treinados sobre dados rotulados (Wankhade et al., 2022). Entre os modelos aplicados a essa tarefa estão Naive Bayes, Regressão Logística e Random Forest (Giachanou & Crestani, 2016; Symeonidis et al., 2018). A Regressão Logística estima a probabilidade de pertencimento a uma classe por meio de uma função logística (Hosmer & Lemeshow, 2000). O Random Forest combina múltiplas árvores de decisão e reduz a variância por agregação (Breiman, 2001), com possibilidade de overfitting quando o ajuste ao conjunto de treino supera a capacidade de generalização para dados não vistos. A avaliação dos modelos utiliza métricas como acurácia, precisão, recall e F1-Score (Wankhade et al., 2022).

2.5 Monitoramento de Redes Sociais e Gestão de Reputação

As redes sociais constituem fonte de informação sobre a percepção pública de marcas e serviços (Giachanou & Crestani, 2016). O monitoramento dessa percepção fornece subsídios para a gestão da reputação organizacional e para a resposta às manifestações dos usuários (Islam et al., 2021). A reputação corporativa associa-se à satisfação, à confiança e à lealdade do cliente (Islam et al., 2021), o que estabelece relação entre o acompanhamento de sentimentos negativos e a mitigação de riscos de imagem. A automação desse monitoramento por meio de Processamento de Linguagem Natural amplia a escala e a frequência da análise em comparação com processos manuais (Naseem et al., 2020).

2.6 Gerenciamento Ágil de Projetos

O desenvolvimento ágil organiza o trabalho em ciclos iterativos e incrementais, com entregas frequentes e revisão contínua junto às partes interessadas (Dybå & Dingsøyr, 2008). O Scrum estrutura esses ciclos em sprints e define cerimônias de planejamento, acompanhamento diário, revisão e retrospectiva (Schwaber & Beedle, 2002). A revisão sistemática de Dybå e Dingsøyr (2008) reuniu evidências empíricas sobre benefícios e limitações do desenvolvimento ágil, com registro de efeitos sobre a coordenação da equipe e a

adequação às mudanças de requisitos. A validação ao final de cada ciclo permite o ajuste do produto às necessidades identificadas pela organização (Dybå & Dingsøyr, 2008).

3. Método de Produção Técnica

Para a elaboração deste relato técnico foi adotado o protocolo para elaboração de relatos de produção técnica proposto por Biancolino et al. (2012), seguindo a lógica CIMO, que organiza a produção técnica em quatro dimensões: Contexto, Intervenção, Mecanismos adotados e Resultados obtidos. Trata-se de uma pesquisa de natureza predominantemente qualitativa e aplicada, voltada à descrição e análise de uma intervenção técnica realizada em uma organização real, com foco na prescrição de soluções que possam ser replicadas em contextos semelhantes.

O problema relatado neste trabalho consiste na ausência de um mecanismo automatizado de monitoramento de sentimentos nas redes sociais por parte da organização parceira, que atuava de forma reativa diante de reclamações e crises de imagem, sem capacidade de identificar tendências negativas antes que se tornassem críticas.

O relato foi construído com base na participação direta dos autores no projeto, que integraram a equipe de desenvolvimento responsável pela concepção, implementação e entrega da solução de PLN. Essa modalidade de participação direta possibilitou a observação contínua do processo em todas as suas etapas, desde o entendimento do negócio e levantamento de requisitos até o desenvolvimento técnico, avaliação dos modelos e integração da solução de PLN com a plataforma de comunicação do parceiro. As habilidades mobilizadas ao longo do projeto abrangeram análise de negócios, pré-processamento de dados textuais, desenvolvimento de modelos de aprendizado de máquina, engenharia de software e gestão de projetos em ambiente ágil.

As informações utilizadas para a construção do relato são de duas naturezas. As fontes primárias compreendem o documento de projeto elaborado em conjunto com a organização parceira, a documentação técnica produzida pela equipe ao longo das cinco sprints, os resultados de desempenho do modelo obtidos em cada ciclo de desenvolvimento e os requisitos definidos pelo parceiro para a solução de PLN. As fontes secundárias compreendem a literatura acadêmica utilizada no referencial teórico, que serve para situar a intervenção dentro do estado da arte das áreas de PLN, análise de sentimentos e gestão de projetos.

A orientação acadêmica do projeto foi conduzida por professor da instituição de ensino responsável pelo módulo, com acompanhamento ao longo de todo o período de desenvolvimento. A interlocução com a organização parceira ocorreu por meio de representantes das áreas de Comunicação, Engenharia de Software e Gestão de Projetos da Empresa de Tecnologia de Mobilidade (ETM), que participaram das cerimônias de revisão ao final de cada sprint e validaram os entregáveis produzidos pela equipe.

4. Contexto do Projeto

Esta seção apresenta o contexto no qual foi desenvolvida a solução de PLN para análise de sentimentos em redes sociais, descrevendo a organização parceira e a situação-problema que motivou a intervenção técnica.

4.1 Caracterização da Organização

A empresa observada neste estudo é uma empresa de tecnologia de mobilidade de grande porte, de capital aberto e de propriedade estrangeira, sediada em São Francisco, nos Estados Unidos. Fundada oficialmente em junho de 2010, a ETM atua no setor de serviços digitais, operando uma plataforma integrada de mobilidade presente em mais de 10 mil cidades

do planeta, com o objetivo de aproximar pessoas e revolucionar o modo de se movimentar nas cidades, seja para transporte de passageiros ou entrega de produtos.

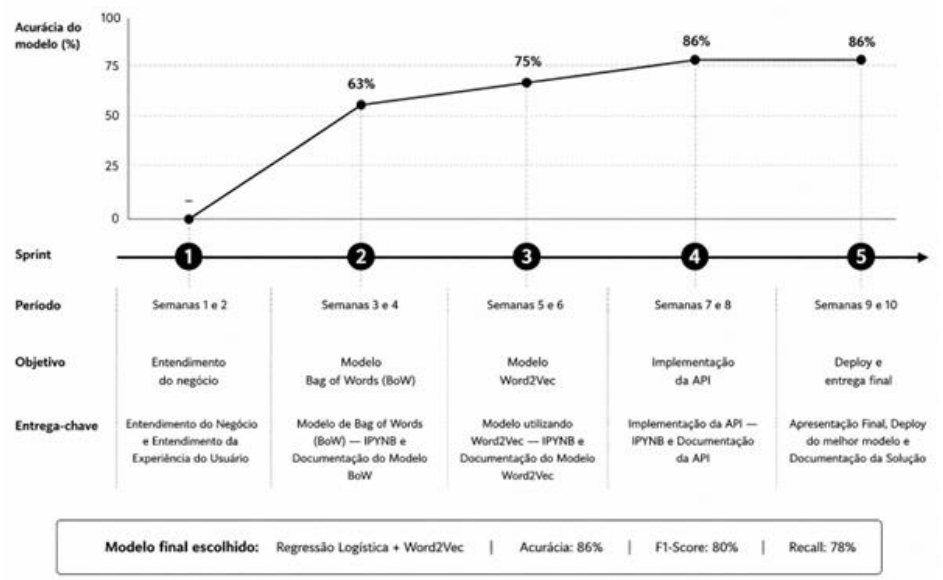
As áreas da ETM diretamente envolvidas neste projeto foram Comunicação, Engenharia de Software e Gestão de Projetos, responsáveis, respectivamente, pelo monitoramento da percepção pública da marca, pelo suporte técnico à integração da solução e pela supervisão do processo de desenvolvimento. Essas áreas atuaram em conjunto para definir os requisitos do projeto, acompanhar o andamento das entregas e validar os resultados ao longo das sprints.

Anteriormente à execução do projeto aqui relatado, o processo de análise dos comentários nas redes sociais era conduzido de forma terceirizada, manual e com periodicidade mensal. Esse modelo apresentava baixa escalabilidade e limitada capacidade de resposta, uma vez que os ciclos longos de atualização impediam a identificação rápida de tendências negativas ou crises emergentes.

4.2 Caracterização do Projeto e do Problema Analisado

O projeto desenvolvido foi denominado Termômetro de Sentimentos, uma aplicação de Processamento de Linguagem Natural para Monitoramento de Redes Sociais, executado por uma equipe de sete estudantes de graduação de uma instituição de ensino superior privada brasileira, em parceria com a ETM, no âmbito de um módulo de projeto com metodologia ágil, dividido em 5 sprints, com duração de 2 semanas cada, totalizando 10 semanas. A linha do tempo de execução do projeto é demonstrada na Figura 1.

Figura 1. Linha do tempo do projeto Termômetro de Sentimentos.



Fonte: Elaborado pelos autores (2026).

A figura evidencia a progressão das entregas ao longo das dez semanas, com cada sprint concentrada em um conjunto específico de objetivos que evoluíram do entendimento do negócio até a entrega da solução completa, refletindo a abordagem iterativa adotada no projeto.

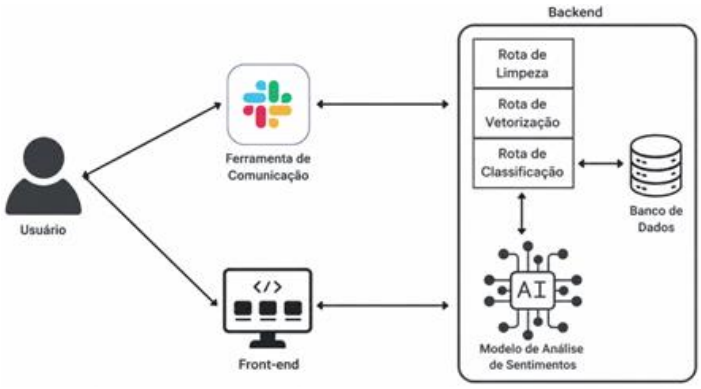
O tema central do projeto é de caráter empresarial, voltado ao fortalecimento da gestão de relacionamento com clientes e à mitigação de riscos de imagem da ETM por meio de automação e inteligência artificial. Em relação às fontes de financiamento e valores envolvidos, a análise financeira elaborada pela equipe estimou um investimento total para implementação e manutenção da solução ao longo de um ano variando entre R$ 645.600,00 e R$ 1.245.600,00, a depender da estrutura de equipe e da inclusão de API de PLN. A economia estimada

decorrente de melhorias em eficiência e gestão de crises foi projetada entre R$ 96.840,00 e R$ 186.840,00 anuais, com ponto de equilíbrio estimado em aproximadamente 3,45 anos.

O problema central identificado era a postura predominantemente reativa da ETM diante de reclamações e episódios de crise nas redes sociais. Conforme relatado pela própria organização parceira, a empresa atuava de forma reativa ao receber quaisquer reclamações dos clientes, sem mecanismo capaz de identificar oportunidades de melhoria antes que uma questão viesse a se tornar uma crise de grandes proporções, afetando negativamente a percepção de usuários e parceiros sobre a marca. O processo vigente, que era terceirizado, manual e de baixa frequência, não era compatível com o volume e a velocidade de geração de conteúdo nas plataformas digitais. Agravava ainda a situação a ausência de parâmetros claros para a categorização dos comentários e a falta de escalabilidade das análises então realizadas.

Diante desse cenário, o objetivo do projeto foi desenvolver uma solução de PLN capaz de classificar automaticamente comentários de usuários na rede social X (anteriormente Twitter) em sentimentos positivos, negativos ou neutros, e de gerar alertas automáticos integrados à plataforma de comunicação corporativa Slack. A arquitetura da solução desenvolvida é apresentada na Figura 2.

Figura 2. Arquitetura da solução.



Fonte: Elaborado pelos autores (2026).

O usuário interage com o sistema por dois caminhos: pela Ferramenta de Comunicação ou pelo Front-end. Em ambos os casos, a requisição é enviada ao Backend, que concentra três rotas sequenciais: a Rota de Limpeza, responsável pelo pré-processamento textual; a Rota de Vetorização, que transforma os comentários em vetores numéricos; e a Rota de Classificação, que aciona o Modelo de Análise de Sentimentos para identificar o sentimento expresso. O resultado é armazenado no Banco de Dados e retornado ao usuário simultaneamente pela Ferramenta de Comunicação e pelo Front-end, mantendo ambas as interfaces sincronizadas.

Esses alertas permitiriam à equipe de Comunicação da ETM agir de forma proativa diante de situações críticas, como sentimentos negativos severos, risco de abandono da plataforma por usuários ou parceiros, ataques à imagem da marca ou problemas identificados na qualidade do serviço. Os benefícios esperados pelo parceiro eram a capacidade de tomar decisões estratégicas com base nas informações apresentadas e mitigar crises severas antes que ocorressem.

5. Tipo de Intervenção e Mecanismos Adotados

A intervenção realizada caracteriza-se como uma força-tarefa de desenvolvimento tecnológico, na qual uma equipe foi formada especificamente para conceber e entregar uma solução capaz de superar as limitações do processo existente na organização parceira. Partindo do diagnóstico de que o processo de análise de sentimentos era conduzido de forma terceirizada, manual e com periodicidade mensal, o objetivo da solução foi criar um agente de Inteligência Artificial baseado em Processamento de Linguagem Natural capaz de classificar automaticamente comentários de usuários em sentimentos positivos, negativos ou neutros e de notificar proativamente a equipe responsável por meio de alertas integrados à plataforma de comunicação corporativa, substituindo o mecanismo de baixa frequência e escalabilidade limitada por um processo contínuo e automatizado.

A escolha pelo Processamento de Linguagem Natural como abordagem central justificou-se pela natureza do problema: era necessário processar grandes volumes de texto em linguagem natural de forma automatizada e com capacidade de capturar nuances semânticas que uma análise baseada em regras ou palavras-chave não seria capaz de identificar com precisão. Além disso, a ausência de parâmetros claros para categorização dos comentários, identificada como um dos agravantes do problema, reforçou a necessidade de um modelo treinado sobre dados reais, capaz de aprender os padrões de linguagem específicos dos usuários da plataforma. O projeto foi conduzido sob a metodologia ágil Scrum, organizada em cinco sprints de duas semanas cada. Ao início de cada sprint era realizada a cerimônia de Planning, na qual as atividades do ciclo eram priorizadas e distribuídas entre a equipe. Diariamente eram realizadas reuniões de acompanhamento (Daily Meeting) para alinhamento do progresso individual. Ao final de cada sprint, as cerimônias de Review e Retrospective reuniram a equipe e os representantes das áreas de Comunicação, Engenharia de Software e Gestão de Projetos da ETM para validação dos entregáveis e definição de ajustes para o ciclo seguinte.

Na Sprint 1, a equipe concentrou-se no entendimento do negócio e na análise da experiência do usuário. Foram desenvolvidas personas representativas dos perfis que utilizariam a solução dentro da ETM, além de user stories que traduziram as necessidades do parceiro em requisitos funcionais. Essa etapa foi determinante para garantir que as decisões técnicas subsequentes estivessem ancoradas nas dores reais identificadas junto à organização.

Na Sprint 2, a equipe realizou a análise descritiva da base de dados e definiu o pipeline de pré-processamento textual. A análise dos comentários evidenciou a presença de URLs, variações morfológicas de palavras e referências a entidades nomeadas de formas distintas, o que orientou a escolha das etapas de tratamento. O pipeline adotado compreendeu, nessa ordem, as etapas ilustradas na Figura 3 abaixo: segmentação das frases, remoção de links, tokenização, remoção de stop-words, lematização e reconhecimento de entidades nomeadas.

Figura 3. Pipeline de pré-processamento textual.



Fonte: Elaborado pelos autores (2026).

A sequência das etapas foi definida a partir da análise descritiva dos dados, que identificou as principais características do texto que precisavam ser tratadas antes da vetorização: a presença de links, a variação morfológica das palavras e a ocorrência de entidades nomeadas referenciadas de formas distintas.

A análise descritiva da base também revelou que 27,7% das palavras presentes nos comentários eram stop-words, conforme apresentado na Figura 4 abaixo, reforçando a necessidade de incluir essa etapa de remoção no pipeline.

Figura 4. Distribuição de palavras e stop-words na base de dados.



Fonte: Elaborado pelos autores (2026).

O dado evidencia que mais de um quarto das palavras da base não agregavam valor semântico à análise, o que justificou a inclusão obrigatória da etapa de remoção de stop-words no pipeline para reduzir o ruído e melhorar a qualidade das representações vetoriais.

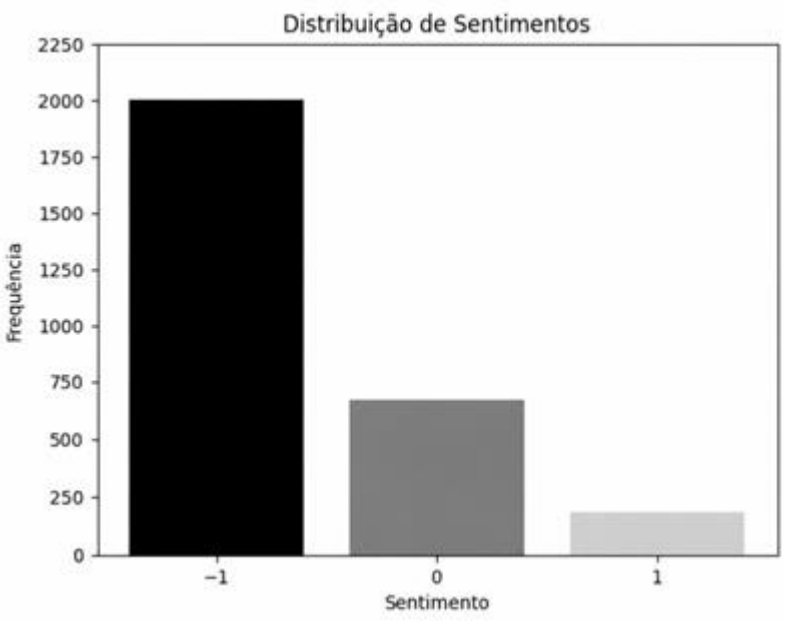
Ao final da sprint, foi desenvolvido o primeiro modelo de classificação com a técnica de vetorização Bag of Words, cujos resultados iniciais, com acurácia de 63%, evidenciaram as limitações dessa abordagem para o problema em questão, especialmente pela incapacidade de

capturar relações semânticas entre palavras e pela geração de matrizes esparsas de alto custo computacional.

Na Sprint 3, a equipe adotou o Word2Vec como técnica de vetorização em substituição ao Bag of Words. Ao representar palavras em vetores densos de 300 dimensões que capturam similaridade semântica e sintática, o Word2Vec mostrou-se significativamente mais adequado para a tarefa de análise de sentimentos em linguagem informal de redes sociais. Foram testados quatro modelos de classificação em conjunto com essa vetorização: Naive Bayes Gaussiano, Random Forest, Camada de Embedding e Regressão Logística. A acurácia dos modelos nessa sprint atingiu 75%, confirmando o avanço proporcionado pela nova abordagem de vetorização.

Na Sprint 4, foram endereçados dois problemas centrais: o desequilíbrio da base de dados e a necessidade de integração da solução com a infraestrutura da ETM. Conforme ilustrado na Figura 5 abaixo, a base apresentava distribuição assimétrica entre as classes, com aproximadamente 2.010 comentários negativos, 688 neutros e 198 positivos, o que tendia a enviesar os modelos.

Figura 5. Distribuição de sentimentos na base de dados



Fonte: Elaborado pelos autores (2026).

O desequilíbrio expressivo entre as classes, com a categoria negativa representando cerca de 70% da base, foi o fator determinante para a aplicação do oversampling na sprint seguinte, sem o qual os modelos tenderiam a classificar a maioria dos comentários como negativos independentemente do conteúdo real.

Para corrigir isso, foi aplicada a técnica de oversampling, que gerou dados sintéticos nas classes sub-representadas e equilibrou o ambiente de treinamento. Com esse ajuste, a Regressão Logística combinada com Word2Vec atingiu acurácia de 86%, F1-Score de 0,80 e recall de 0,78, destacando-se pelo melhor equilíbrio entre capacidade preditiva, ausência de overfitting e eficiência computacional entre todos os modelos avaliados. Em paralelo, foi desenvolvida a API com três rotas sequenciais: Rota de Limpeza, responsável pelo pré-processamento dos textos recebidos; Rota de Vetorização, que transforma os textos em vetores Word2Vec; e Rota de Classificação, que aciona o modelo e retorna o sentimento identificado. A conexão com a

plataforma de comunicação corporativa foi estabelecida nesta sprint, possibilitando o envio automático de alertas à equipe da ETM.

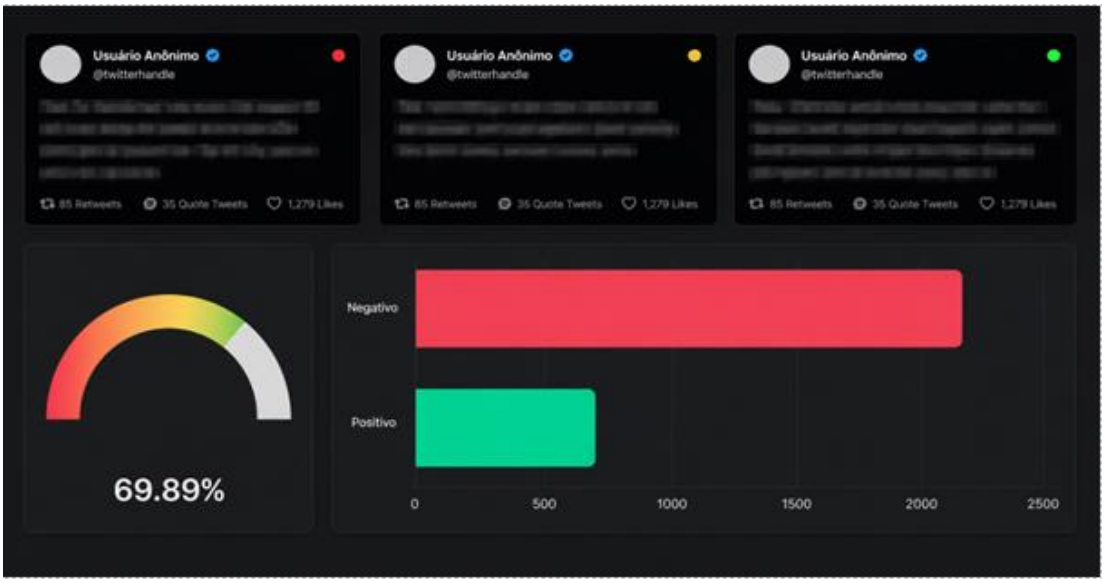
Na Sprint 5, foram implementados os alertas automáticos e criado o painel de visualização em Front-end. A solução completa foi apresentada e validada pelos representantes das três áreas da ETM na cerimônia de revisão final, encerrando o ciclo de desenvolvimento com a entrega de um sistema integrado, funcional e capaz de classificar comentários em tempo reduzido, endereçando diretamente a limitação de escalabilidade que caracterizava o processo anterior.

6. Apresentação dos Resultados e Análise

A solução entregue ao término do projeto materializou-se em dois componentes integrados e funcionais: um painel de visualização em Front-end e um bot integrado ao Slack corporativo da ETM. Esses dois componentes respondem diretamente ao problema central identificado no diagnóstico, qual seja, a incapacidade de identificar tendências negativas ou crises emergentes em tempo hábil, dado que o processo anterior era terceirizado, manual e de periodicidade mensal.

O painel web, acessível via navegador, apresenta três cards superiores que exibem exemplos de comentários classificados por categoria de sentimento, identificados por indicadores coloridos: vermelho para negativos, amarelo para neutros e verde para positivos. Abaixo dos cards, um velocímetro semicircular exibe a porcentagem de comentários negativos presentes na base analisada, e um gráfico de barras horizontais apresenta a contagem comparativa entre as categorias negativa e positiva. O painel em funcionamento com dados carregados é apresentado na Figura 6 abaixo.

Figura 6. Painel de visualização do Termômetro de Sentimentos.



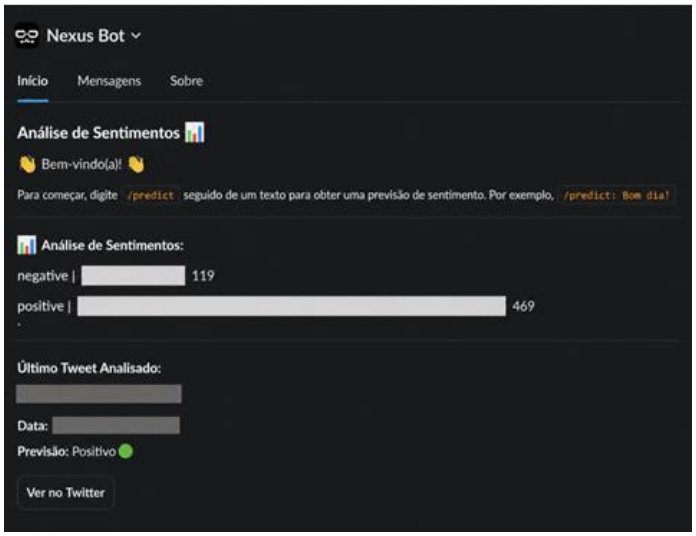
Fonte: Elaborado pelos autores (2026).

A figura evidencia o painel após o processamento de uma base de comentários: o velocímetro registra 69,89% de sentimentos negativos e o gráfico de barras permite visualizar a proporção entre as classes, tornando acessível para a equipe de Comunicação da ETM uma

leitura rápida do estado atual da percepção pública. O upload de novos comentários é feito diretamente pelo botão na interface, sem necessidade de acesso a sistemas externos ou conhecimento técnico.

A segunda interface de acesso à solução é o Nexus Bot, aplicativo integrado ao workspace do Slack da ETM. Na aba inicial do bot, a equipe encontra um painel resumido com a contagem acumulada de comentários classificados como negativos e positivos, o último comentário analisado, sua data de processamento, a previsão atribuída pelo modelo e um botão de acesso direto ao tweet original na rede social X. O bot também permite o envio de arquivos CSV diretamente pela conversa, processando o arquivo e retornando, em aproximadamente 10 a 15 segundos, uma tabela com cada comentário e seu respectivo sentimento classificado. A interface do Nexus Bot na aba inicial é apresentada na Figura 7 abaixo.

Figura 7. Interface inicial do Nexus Bot no Slack.



Fonte: Elaborado pelos autores (2026).

A figura mostra o painel do bot após o processamento de uma base, com 102 comentários negativos e 402 positivos registrados no banco de dados, e o último tweet analisado com previsão positiva. Esse painel é atualizado automaticamente a cada nova análise realizada, seja pelo Slack ou pelo Front-end, mantendo ambas as interfaces sincronizadas. A integração direta ao ambiente de comunicação interna da ETM elimina a necessidade de acesso a uma plataforma externa, reduzindo a barreira de uso para as equipes de Comunicação e Gestão de Projetos.

O desenvolvimento da solução ao longo das cinco sprints resultou em um agente de Inteligência Artificial funcional, integrado à plataforma de comunicação corporativa e ao painel de visualização, capaz de classificar automaticamente comentários de usuários em sentimentos negativos ou positivo-neutros. A evolução do desempenho do modelo ilustra o processo iterativo adotado: a primeira versão, baseada em Bag of Words com Naive Bayes, atingiu acurácia de 63%; com a adoção do Word2Vec na Sprint 3, a acurácia avançou para 75%; e após o ajuste do desequilíbrio da base por *oversampling* na Sprint 4, o modelo final atingiu acurácia de 86%, F1-Score de 0,80 e recall de 0,78.

A comparação entre os quatro modelos avaliados com Word2Vec após a aplicação do *oversampling* é apresentada na Tabela 1 abaixo.

Tabela 1. Comparação de desempenho dos modelos de classificação.

Modelo	Acurácia	Precisão	Recall	F1-Score	Uso de Memória	Tempo de Execução
Naive Bayes Gaussiano	0.83	0.83	0.82	0.82	Baixo	Alto
Random Forest	0.78	0.73	0.74	0.73	Alto	Moderado
Regressão Logística	0.86	0.80	0.78	0.80	Baixo a moderado	Rápido
Camada de Embedding	0.83	0.83	0.82	0.82	Muito Alto	Lento

Fonte: Elaborado pelos autores (2026).

A Regressão Logística destacou-se pelos melhores resultados gerais, com acurácia de 0,86, precisão de 0,80, recall de 0,78 e F1-Score de 0,80. O Naive Bayes Gaussiano e a Camada de Embedding apresentaram acurácia de 0,83, enquanto o Random Forest obteve 0,78. Um fator decisivo na escolha da Regressão Logística foi a menor diferença entre as métricas de treino e teste em comparação com os demais modelos, o que indicou ausência de overfitting e maior capacidade de generalização para dados não vistos. O Random Forest, em contraste, apresentou acurácia de treino de 0,99 contra 0,78 no teste, sinalizando forte overfitting e limitando sua aplicabilidade ao contexto real.

A matriz de confusão do modelo final, apresentada na Figura 8 abaixo, evidencia que, dos comentários negativos presentes no conjunto de teste, 549 foram classificados corretamente e 69 de forma equivocada. Dentre os comentários positivos e neutros, 148 foram classificados corretamente e 97 de forma incorreta. Esses números refletem um desempenho robusto especialmente na identificação de sentimentos negativos, que é a classe de maior relevância estratégica para a organização parceira, dado que são os comentários negativos que sinalizam riscos de crise e demandam ação imediata.

Figura 8. Matriz de confusão do modelo final.



Fonte: Elaborado pelos autores (2026).

Do ponto de vista organizacional, o resultado mais relevante para a ETM não é apenas a acurácia do modelo em si, mas a mudança estrutural no processo de monitoramento que ela viabilizou. Essa transformação endereça diretamente o problema de escalabilidade identificado no diagnóstico inicial e elimina a dependência de ciclos mensais e de terceiros para obter informações sobre a percepção dos usuários.

Os resultados obtidos confirmam aspectos descritos no referencial teórico. A evolução da acurácia de 63% com Bag of Words para 75% com Word2Vec é coerente com a limitação das representações esparsas apontada por Manning et al. (2008) e com a propriedade de captura

de similaridade semântica atribuída aos vetores densos por Mikolov et al. (2013). A inclusão da remoção de stop-words e da lematização no pipeline alinha-se aos achados de Symeonidis et al. (2018), que associaram a lematização a aumento de acurácia em conjuntos de dados do Twitter. O comportamento dos modelos também dialoga com a literatura: a Regressão Logística apresentou a menor diferença entre as métricas de treino e teste, em consonância com a estimação probabilística descrita por Hosmer e Lemeshow (2000), enquanto o Random Forest registrou acurácia de treino de 0,99 contra 0,78 no teste, situação de overfitting que se distingue da redução de variância por agregação descrita por Breiman (2001).

No plano organizacional, a relação entre o acompanhamento de sentimentos negativos e a gestão da reputação corresponde ao vínculo entre reputação corporativa, satisfação e confiança apresentado por Islam et al. (2021). A informalidade da linguagem dos comentários, tratada na etapa de pré-processamento, reflete as dificuldades de classificação em redes sociais descritas por Giachanou e Crestani (2016) e por Naseem et al. (2020). A condução do projeto em cinco sprints, com validação ao final de cada ciclo, corresponde à estrutura de cerimônias do Scrum definida por Schwaber e Beedle (2002) e às evidências sobre adequação às mudanças de requisitos reunidas por Dybå e Dingsøyr (2008).

Do ponto de vista das contribuições para casos similares, a experiência relatada demonstra que organizações com alto volume de menções em redes sociais e processos manuais de monitoramento podem se beneficiar de soluções de PLN estruturadas de forma incremental, com validações frequentes junto às áreas de negócio. A combinação de Word2Vec com Regressão Logística mostrou-se uma abordagem equilibrada e replicável para tarefas de classificação de sentimentos em linguagem informal, especialmente quando a base de dados apresenta desbalanceamento entre classes, situação comum em dados reais de redes sociais. A integração da solução diretamente à plataforma de comunicação interna da organização, em vez de exigir acesso a uma interface externa, foi um fator identificado como determinante para a adoção e a utilidade prática da solução.

O ensinamento profissional central que esta experiência oferece à comunidade de profissionais de tecnologia e gestão é que projetos de Inteligência Artificial aplicados à comunicação organizacional ganham em efetividade quando conduzidos com metodologia ágil e validação contínua dos stakeholders. A cada ciclo de revisão, os representantes das áreas de negócio da ETM puderam direcionar ajustes que mantiveram a solução alinhada às necessidades reais da organização, evitando o risco comum nesses projetos de entregar um modelo tecnicamente satisfatório mas operacionalmente inadequado ao contexto de uso.

7. Conclusão

O objetivo proposto, desenvolver uma solução de Processamento de Linguagem Natural capaz de classificar automaticamente comentários de usuários em sentimentos positivos, negativos ou neutros e de gerar alertas integrados à plataforma de comunicação corporativa, foi atingido. A solução foi entregue em dois componentes integrados, um painel de visualização em Front-end e um bot no Slack corporativo, e o modelo final, que combinou Word2Vec e Regressão Logística com oversampling, atingiu acurácia de 86%, F1-Score de 0,80 e recall de 0,78. A substituição do processo terceirizado, manual e de periodicidade mensal por um fluxo automatizado e contínuo endereçou a limitação de escalabilidade identificada no diagnóstico inicial.

Para empresas e projetos em situação semelhante, a experiência indica que organizações com alto volume de menções em redes sociais e processos manuais de monitoramento podem adotar soluções de Processamento de Linguagem Natural estruturadas de forma incremental,

com validação frequente junto às áreas de negócio. A combinação de Word2Vec com Regressão Logística e o tratamento do desbalanceamento por oversampling constitui abordagem aplicável a bases com distribuição assimétrica entre classes, e a integração à plataforma de comunicação interna reduz a barreira de uso para as equipes.

Entre os aspectos que limitaram o trabalho, registra-se a classificação final em duas categorias agregadas, negativos e positivo-neutros, em vez das três classes inicialmente previstas, e o tamanho e a distribuição da base de dados, que demandaram a geração de dados sintéticos por oversampling. O escopo do relato restringe-se à intervenção realizada em uma organização, o que delimita a generalização dos resultados.

Como desdobramentos futuros, indicam-se a ampliação da base de dados rotulada, a avaliação de modelos baseados em arquiteturas de aprendizado profundo para a classificação de sentimentos e o desenvolvimento de trabalhos acadêmicos a partir do aprofundamento na literatura e da ampliação do escopo do estudo.

Referências

Biancolino, C. A., Kniess, C. T., Maccari, E. A., & Rabechini Jr., R. (2012). Protocolo para elaboração de relatos de produção técnica. *Revista Gestão e Projetos, 3*(2), 294-307. https://doi.org/10.5585/gep.v3i2.121

Breiman, L. (2001). Random forests. *Machine Learning, 45*(1), 5-32. https://doi.org/10.1023/A:1010933404324

Dybå, T., & Dingsøyr, T. (2008). Empirical studies of agile software development: A systematic review. *Information and Software Technology, 50*(9-10), 833-859. https://doi.org/10.1016/j.infsof.2008.01.006

Giachanou, A., & Crestani, F. (2016). Like it or not: A survey of Twitter sentiment analysis methods. *ACM Computing Surveys, 49*(2), 1-41. https://doi.org/10.1145/2938640

Hosmer, D. W., & Lemeshow, S. (2000). *Applied logistic regression* (2nd ed.). John Wiley & Sons.

Islam, T., Islam, R., Pitafi, A. H., Xiaobei, L., Rehmani, M., Irfan, M., & Mubarak, M. S. (2021). The impact of corporate social responsibility on customer loyalty: The mediating role of corporate reputation, customer satisfaction, and trust. *Sustainable Production and Consumption, 25*, 123-135. https://doi.org/10.1016/j.spc.2020.07.019

Liu, B. (2012). *Sentiment analysis and opinion mining*. Morgan & Claypool. https://doi.org/10.2200/S00416ED1V01Y201204HLT016

Manning, C. D., Raghavan, P., & Schütze, H. (2008). *Introduction to information retrieval*. Cambridge University Press.

Mikolov, T., Sutskever, I., Chen, K., Corrado, G., & Dean, J. (2013). Distributed representations of words and phrases and their compositionality. *Advances in Neural Information Processing Systems, 26*, 3111-3119.

Naseem, U., Razzak, I., Musial, K., & Imran, M. (2020). Transformer based deep intelligent contextual embedding for Twitter sentiment analysis. *Future Generation Computer Systems, 113*, 58-69. https://doi.org/10.1016/j.future.2020.06.050

Schwaber, K., & Beedle, M. (2002). *Agile software development with Scrum*. Prentice Hall.

Symeonidis, S., Effrosynidis, D., & Arampatzis, A. (2018). A comparative evaluation of pre-processing techniques and their interactions for Twitter sentiment analysis. *Expert Systems with Applications, 110*, 298-310. https://doi.org/10.1016/j.eswa.2018.06.022

Wankhade, M., Rao, A. C. S., & Kulkarni, C. (2022). A survey on sentiment analysis methods, applications, and challenges. *Artificial Intelligence Review, 55*(7), 5731-5780. https://doi.org/10.1007/s10462-022-10144-1