

FATEC – FACULDADE DE TECNOLOGIA DE RIO CLARO
CURSO: Inteligência Artificial

**RELATÓRIO DE PESQUISA – WEB SCRAPING E ANÁLISE DE SENTIMENTOS
APLICADOS À REPUTAÇÃO EMPRESARIAL**

GRUPO: SOUL CARE
ALUNO: Vladimir Queiroz Sejas

Rio Claro – 2025

1. INTRODUÇÃO

A reputação de uma empresa no ambiente digital tornou-se um dos principais ativos intangíveis da atualidade. Com o aumento exponencial de plataformas de avaliação, redes sociais e fóruns de consumidores, compreender o sentimento do público sobre marcas e produtos passou a ser essencial para a tomada de decisão estratégica. Nesse cenário, o uso de técnicas de Web Scraping e Análise de Sentimentos, baseadas em Inteligência Artificial (IA) e Processamento de Linguagem Natural (PLN), representa uma abordagem robusta e escalável para coletar, organizar e interpretar grandes volumes de textos oriundos de múltiplas fontes.

O presente relatório, desenvolvido pelo grupo SOUL CARE, tem como objetivo analisar a integração entre essas tecnologias e discutir suas aplicações práticas no contexto empresarial, com foco em soluções que possam ser incorporadas no projeto Feeling AI.

2. FUNDAMENTOS TEÓRICOS

O Web Scraping consiste na coleta automatizada de dados disponíveis publicamente na web. Essa técnica é amplamente utilizada em pesquisas acadêmicas, monitoramento de mercado, jornalismo de dados e sistemas de reputação corporativa. As bibliotecas Python mais empregadas são:

- BeautifulSoup – ideal para páginas estáticas e de fácil estruturação HTML.
- Scrapy – poderosa para extrações em larga escala, com controle de requisições e integração a bancos de dados.
- Playwright e Selenium – voltadas para páginas dinâmicas que utilizam JavaScript, simulando a navegação humana.

Em paralelo, a Análise de Sentimentos emprega modelos de PLN para identificar a emoção ou opinião expressa em textos. Ela pode ser dividida em dois tipos principais: (a) análise de polaridade global (positiva, negativa ou neutra) e (b) análise baseada em aspectos, que busca identificar o sentimento associado a temas específicos (como “preço”, “entrega” ou “atendimento”).

3. ANÁLISE DE SENTIMENTOS EM PORTUGUÊS

Modelos de linguagem desenvolvidos especificamente para o português têm se destacado por sua precisão e relevância cultural. Dentre os principais estão:

- BERTimbau – modelo treinado em corpus brasileiro, disponível na plataforma Hugging Face.
- pysentimiento – oferece classificação de polaridade e emoções com suporte nativo a português.
- RoBERTuito – modelo leve e rápido, ideal para aplicações em tempo real.

Serviços de nuvem também oferecem APIs robustas para essa finalidade: Google Cloud Natural Language, AWS Comprehend e Azure Text Analytics. Esses serviços apresentam suporte multilíngue, escalabilidade e integração direta com sistemas corporativos, permitindo a análise de milhares de textos por minuto.

4. ARQUITETURA E ETAPAS DO PROCESSO

Um pipeline típico de coleta e análise de sentimentos é composto por cinco etapas:

1. Planejamento da coleta – definição das fontes (Google Reviews, Reclame Aqui, redes sociais, lojas virtuais) e das palavras-chave.
2. Raspagem e armazenamento – uso de Scrapy ou BeautifulSoup para extração e pandas para organização dos dados.
3. Pré-processamento linguístico – remoção de stopwords, correção ortográfica, lematização e normalização de textos.
4. Classificação de sentimentos – aplicação de modelos pré-treinados ou APIs de nuvem.
5. Visualização e relatórios – criação de dashboards interativos (Looker Studio, Power BI, Superset) para acompanhamento e tomada de decisão.

5. APLICAÇÃO NO CONTEXTO EMPRESARIAL

O uso dessas técnicas no ambiente corporativo traz benefícios significativos, como:

- Monitoramento em tempo real da percepção da marca.
- Identificação de padrões de satisfação e insatisfação.
- Suporte à estratégia de marketing e atendimento.
- Detecção de possíveis crises de imagem antes que se ampliem.

Empresas de varejo, telecomunicações, tecnologia e serviços financeiros já aplicam soluções baseadas em IA para captar insights de consumidores e melhorar a experiência do cliente.

6. CONSIDERAÇÕES ÉTICAS E LEGAIS

Toda coleta de dados deve observar a Lei Geral de Proteção de Dados (LGPD) e os termos de uso das plataformas de origem. É fundamental garantir a anonimização das informações e respeitar o princípio da minimização de dados, coletando apenas o necessário. Além disso, sistemas de Web Scraping devem seguir as boas práticas indicadas em arquivos robots.txt, respeitando limites de requisição e horários de acesso para evitar sobrecarga nos servidores.

7. APLICAÇÃO AO PROJETO SOUL CARE

O projeto Feeling AI, desenvolvido pelo grupo SOUL CARE, propõe a criação de um sistema inteligente de análise emocional voltado à compreensão de opiniões públicas. Nesse contexto, a integração entre Web Scraping e Análise de Sentimentos se torna essencial. A coleta será realizada em fontes como Reclame Aqui, Google Maps e redes sociais, e os dados serão tratados por modelos de PLN em português. O resultado alimentará dashboards que exibirão, em tempo real, a proporção de sentimentos positivos, negativos e neutros, permitindo que empresas parceiras acompanhem sua reputação de forma visual e dinâmica.

8. CONCLUSÃO

A combinação entre Web Scraping e Análise de Sentimentos representa um avanço relevante na área de Inteligência Artificial aplicada à gestão da reputação. Ao permitir o monitoramento contínuo e automatizado de opiniões, essas técnicas fortalecem a capacidade analítica das empresas e oferecem bases sólidas para decisões estratégicas. Para o grupo SOUL CARE, essa pesquisa reforça o compromisso com soluções inovadoras e eticamente responsáveis, consolidando o projeto Feeling AI como um instrumento capaz de unir tecnologia, empatia e inteligência de dados.

9. REFERÊNCIAS

BRASIL. Lei nº 13.709, de 14 de agosto de 2018 (Lei Geral de Proteção de Dados Pessoais – LGPD). Diário Oficial da União, Brasília, 2018. HUGGING FACE. Modelos de linguagem BERTimbau e RoBERTuito para português. Disponível em: <https://huggingface.co>. Acesso em: nov. 2025. MICROSOFT AZURE. Text Analytics for Sentiment Analysis (pt-BR). Disponível em: <https://azure.microsoft.com>. Acesso em: nov. 2025. AMAZON WEB SERVICES. Comprehend Sentiment Analysis API. Disponível em: <https://aws.amazon.com/comprehend/>. Acesso em: nov. 2025. GOOGLE CLOUD. Natural Language API – Sentiment Analysis. Disponível em: <https://cloud.google.com/natural-language>. Acesso em: nov. 2025. RECLAME AQUI. RA Reviews e API institucional. Disponível em: <https://www.reclameaqui.com.br>. Acesso em: nov. 2025.