

FATEC – Faculdade de Tecnologia de Rio Claro

Curso: Inteligência Artificial

Projeto: FEELING IA

Grupo: SOULCARE

Relatório de Pesquisa – Integração do Flask com Web Scraping e Análise de Sentimentos no Projeto Feeling IA

Rio Claro – SP
2025

Sumário

1. Introdução
 2. Visão Geral do Projeto
 3. Fundamentação Teórica
 4. Arquitetura do Sistema
 5. Modelo de Dados e JSON
 6. Web Scraping
 7. API Flask
 8. Pipeline de Análise de Sentimentos
 9. Validação, Testes e Métricas
 10. Segurança, Ética e Privacidade
 11. Desempenho e Escalabilidade
 12. Logging, Observabilidade e Manutenção
 13. Estudo de Caso Ilustrativo
 14. Resultados e Discussão
 15. Plano de Evolução e Roadmap
 16. Conclusão
- Referências

1. Introdução

O projeto *Feeling IA* é uma iniciativa educacional do grupo SoulCare na FATEC para construir uma plataforma de coleta e análise de opiniões on-line. A proposta combina técnicas de web scraping, um serviço web com Flask e representação de dados em JSON para entregar uma visão clara e auditável da polaridade dos textos.

Este relatório apresenta a motivação, a base teórica, as decisões de engenharia e a implementação do sistema, com ênfase em clareza, reprodutibilidade e ética no uso dos dados.

2. Visão Geral do Projeto

O sistema é organizado em quatro camadas: coleta de dados, pré-processamento, classificação de sentimentos e exposição por API. A interoperabilidade é garantida por contratos JSON simples e estáveis, consumidos por clientes web ou ferramentas analíticas.

3. Fundamentação Teórica

A análise de sentimentos deriva de estudos clássicos e modernos que variam de léxicos a modelos neurais. Em paralelo, a arquitetura REST orienta o desenho de APIs escaláveis, e princípios de engenharia sustentam modularidade e confiabilidade.

4. Arquitetura do Sistema

A arquitetura define fronteiras claras: o scraper coleta, o processador limpa e estrutura texto, o classificador atribui rótulos e a API entrega resultados. Cada módulo é testável de forma independente e comunica-se por JSON.

5. Modelo de Dados e JSON

Os objetos principais são Documento, Amostra e Resultado. O JSON inclui campos para rastreabilidade (URL, timestamp), conteúdo original e saídas do classificador (polaridade e score).

6. Web Scraping

A coleta é feita com requests e BeautifulSoup, respeitando limites e termos de uso. Os seletores são parametrizados para facilitar ajustes entre sites.

7. API Flask

A API usa rotas REST e retorno em JSON. O Flask foi adotado pela simplicidade e robustez da comunidade. Cabeçalhos, códigos de status e erros são padronizados.

8. Pipeline de Análise de Sentimentos

O pipeline inclui normalização, tokenização, remoção de stopwords e classificação por léxico ponderado. Thresholds definem positivo, neutro e negativo.

9. Validação, Testes e Métricas

Acurácia, precisão, revocação e F1 avaliam o classificador. Testes unitários validam rotas da API e contratos JSON. Amostras duplicadas são filtradas na coleta.

10. Segurança, Ética e Privacidade

Minimização de dados, transparência de finalidade e logs sem PII são diretrizes centrais. O projeto utiliza somente conteúdo público em demonstrações acadêmicas.

11. Desempenho e Escalabilidade

Cache, paginação e compressão são adotados para reduzir latência. Para alto volume, filas assíncronas e múltiplas instâncias WSGI são caminhos de evolução.

12. Logging, Observabilidade e Manutenção

Logs estruturados, métricas e rastreamento de requisições facilitam diagnóstico. A manutenção prevê atualização de seletores e dicionários léxicos.

13. Estudo de Caso Ilustrativo

Em um conjunto de 2.000 avaliações simuladas, 54% foram positivas, 31% neutras e 15% negativas, com menções a 'entrega' e 'atendimento' correlacionando positividade.

14. Resultados e Discussão

O *Feeling IA* demonstra viabilidade técnica e didática: integra fluxos modernos da web com tratamento linguístico para apoiar decisões baseadas em dados.

15. Plano de Evolução e Roadmap

Próximos passos incluem integração com APIs oficiais, *fine-tuning* supervisionado, dashboards interativos e internacionalização do pipeline.

16. Conclusão

A combinação de scraping, JSON e Flask, aliada a um classificador transparente, cria uma base sólida para ensino, pesquisa e aplicações práticas na FATEC.

Referências

PANG, B.; LEE, L. Opinion Mining and Sentiment Analysis. Foundations and Trends in Information Retrieval, 2008.

CAMBRIA, E. Affective Computing and Sentiment Analysis. IEEE Intelligent Systems, 2020.

FIELDING, R. Architectural Styles and the Design of Network-based Software Architectures. 2000.

FOWLER, M. Patterns of Enterprise Application Architecture. 2002.

BASS, L.; CLEMENTS, P.; KAZMAN, R. Software Architecture in Practice. 2012.

CAVOUKIAN, A. Privacy by Design. 2011.