

 INSTITUTO FEDERAL Piauí	INSTITUTO FEDERAL DO PIAUÍ CURSO: TECNOLOGIA EM ANÁLISE E DESENVOLVIMENTO DE SISTEMAS DISCIPLINA: TÓPICOS ESPECIAIS EM PROGRAMAÇÃO PROFESSOR: HÉLCIO DE ABREU SOARES Teresina (PI) 10 de dezembro de 2025
--	--

ALUNOS:

- *Marcos Araújo Silva*
- *José Macêdo dos Santos Junior*
- *Wadson Tardelle Dias de Lima*
- *Leonardo Vitorio da Silva*
- *Emanoel Sousa do Carmo*

Relatório

O projeto visou criar um *pipeline* de *Machine Learning* para classificação binária de textos (SMS). O principal diferencial foi o uso de **LIME** e **SHAP** para mitigar a natureza "caixa-preta" dos modelos e aumentar a interpretabilidade das decisões. O fluxo de trabalho foi dividido em 5 etapas: Carregamento/pré-processamento, Vetorização, Modelagem (Linear), Avaliação de Desempenho e Interpretabilidade/Análise Crítica.2. Caracterização dos Dados e Pré-processamento

- **Dataset:** SMS Spam Collection, composto por textos curtos, informais e com alta redundância.
- **Desbalanceamento:** O conjunto é altamente desbalanceado, com cerca de **4800 amostras ham** (legítimas) e apenas **750 amostras spam**.
- **Pipeline de Tratamento:** Envolveu limpeza básica (minúsculas, remoção de caracteres especiais) e vetorização.

3. Metodologia de Vetorização e Desempenho

O projeto comparou a eficácia das técnicas de vetorização **TF-IDF** e **Bag-of-Words (BOW)** para alimentar modelos lineares (Regressão Logística/LinearSVC).

Métrica	TF-IDF (Melhor Desempenho)	Bag-of-Words (BOW)
Accuracy	98.02%	97.58%
Precision	1.00	1.00
Recall (Spam)	0.85	0.82
F1-Score	0.92	0.90

- **Conclusão da Performance:** A escolha da vetorização (**TF-IDF**) influenciou mais o resultado do que a troca do algoritmo, demonstrando superioridade, especialmente na detecção de *spam* (**Recall de 0.85**).
- **Matriz de Confusão (TF-IDF):** Em 1115 testes, houve 22 Falsos Negativos (Spam classificado como Ham), e o modelo teve **100%** de acerto na classe *ham*.

4. Análise de Interpretabilidade

A análise buscou entender *quais palavras* levam o modelo a classificar uma mensagem como *spam*.LIME (Local Interpretable Model-agnostic Explanations)

- **Foco:** Explicações locais (para uma única mensagem).
- **Principais Features (Spam):** Palavras como "call", "to", "txt", "free", "mobile" e "stop" apareceram com os maiores pesos positivos para *spam*.
- **Limitações:** Instabilidade em textos muito curtos e dependência de aleatoriedade.

SHAP (SHapley Additive exPlanations)

- **Foco:** Abordagem baseada em teoria dos jogos, oferecendo consistência matemática para análises globais e locais.
- **Visão Global:** Identificou que "to", "call", "me", "my" e "ok" são as palavras com maior impacto médio na saída do modelo.
- **Custo:** Exige mais memória e poder computacional, necessitando limitar o vocabulário.

Comparativo LIME vs. SHAP

Critério	LIME	SHAP
Foco	Local	Global + Local
Estabilidade	Média	Alta
Interpretação	Muito intuitiva	Pode ser mais complexa
Custo Computacional	Médio	Alto
Melhor Uso	Explicar previsão individual	Analizar importância geral de <i>features</i>

5. Conclusões e Recomendações

Análise Crítica e Conclusão Consolidada do Projeto1. O Triunfo da Simplicidade Algorítmica

A pesquisa demonstrou que, para a tarefa específica de classificação de mensagens curtas (SMS) com vocabulário limitado, a complexidade algorítmica não se traduz em performance superior. Modelos lineares, quando combinados com a representação TF-IDF, atingiram um desempenho de excelência, caracterizado por uma Acurácia de aproximadamente 98% e Precisão de 100%.

- **Destaque:** A fase de representação dos dados foi mais determinante para o sucesso do que a escolha do classificador. A utilização de técnicas avançadas como *Deep Learning* ou *Transformers* se mostrou desnecessária, elevando o custo computacional sem trazer ganhos de performance proporcionais.

2. Interpretabilidade (XAI): O *Trade-off* entre Estabilidade e Agilidade

A aplicação de ferramentas de Inteligência Artificial Explicável (XAI) evidenciou um dilema de estabilidade e custo entre as abordagens LIME e SHAP:

- **LIME (Agilidade com Instabilidade):** É uma ferramenta rápida, ideal para validações pontuais. Contudo, a geração de perturbações aleatórias em mensagens muito concisas levou a uma instabilidade notável em suas explicações.
- **SHAP (Padrão Ouro de Confiança):** Embora seja mais exigente em termos de memória e tempo de processamento, o SHAP estabeleceu-se como a referência, fornecendo uma visão global e matematicamente consistente. Ele confirmou de forma robusta quais palavras (e.g., "free", "call", "txt") são os preditores mais fortes e confiáveis de spam.