**Marcos Vinicius Peixoto Moura**

**Álgebra Linear**

**Trabalho sobre Similaridade de Cosseno**

**Processamento de Linguagem Natural (PNL)**

**INTRODUÇÃO**

**Atualmente é comum que pessoas compartilhem suas experiências com empresas em sites de avaliação. No setor aéreo, essas opiniões ajudam outros usuários a decidirem qual companhia escolher. Mas com tantos comentários, como identificar qual empresa tem as avaliações mais positivas e consistentes?**

Uma das formas de fazer isso é usando técnicas de Processamento de Linguagem Natural (PNL), que permitem analisar grandes volumes de texto. Entre essas técnicas, a similaridade de cosseno é usada para medir o quão parecidas são duas avaliações. Esta técnica ajuda a entender se os passageiros estão dizendo coisas semelhantes sobre uma empresa — por exemplo, se todos elogiam o conforto ou criticam atrasos, isso fica evidente.

Neste trabalho, foi usada essa técnica para descobrir quais companhias aéreas são mais bem avaliadas, tanto pelas notas que recebem quanto pela consistência dos comentários.

**Objetivo**

1. Os principais objetivos desse projeto foram:
2. Analisar os textos das avaliações feitas por passageiros sobre diversas companhias aéreas.
3. Usar a similaridade de cosseno para medir o quanto as opiniões se parecem — o que pode indicar se os elogios (ou críticas) são constantes.
4. Criar uma métrica combinada que considere:
5. A coerência entre os comentários (baseada na similaridade entre textos);
6. A nota média que cada companhia recebe.
7. Classificar as companhias com base nessa análise para apontar as que são mais recomendadas.

**Como a análise foi feita**

Coleta dos dados: utilização de um conjunto de dados disponível no site Kaggle, com avaliações de passageiros sobre várias companhias aéreas.

Limpeza dos textos: remoção de pontuações, palavras irrelevantes (como “the”, “and”) e transformamos tudo em minúsculas para facilitar a análise.

Vetorização com TF-IDF: transformação dos textos em números usando uma técnica chamada TF-IDF, que valoriza palavras mais importantes em cada avaliação.

Cálculo da similaridade de cosseno: Com os textos convertidos em vetores, comparamos as avaliações entre si para ver o quanto se pareciam. Quanto maior a média dessa similaridade para uma empresa, mais consistentes eram os comentários.

Criação de um “score” final: unir a similaridade média com a nota média das avaliações para gerar uma pontuação final. Assim, conseguimos classificar as companhias de acordo com a satisfação dos clientes

**Conclusão**

Usar similaridade de cosseno em avaliações de texto foi uma forma eficiente de entender a reputação das companhias aéreas além das notas numéricas. Essa abordagem permite identificar padrões nos comentários, trazendo insights valiosos tanto para clientes quanto para as empresas.

Esse tipo de análise pode ser aplicado em qualquer área onde existem muitas opiniões escritas — seja em lojas, restaurantes, hotéis ou aplicativos.

***Dataset***: <https://www.kaggle.com/datasets/juhibhojani/airline-reviews>