

Nome: Arthur Bispo Reis

## **Relatório sobre Análise de Similaridade com o Termo “food”**

### **1. Introdução**

Com a crescente utilização de plataformas digitais para registrar opiniões, entender o conteúdo textual dessas avaliações se tornou fundamental. Neste projeto, empregamos técnicas de Processamento de Linguagem Natural (PLN) com o intuito de verificar quais comentários de restaurantes possuem maior relação semântica com a palavra “food”.

### **2. Objetivo**

O propósito desta análise é identificar os comentários de usuários que fazem menção direta ou possuem alto grau de relação com o termo “food”, por meio de um algoritmo de similaridade textual. Essa estratégia facilita a seleção de avaliações mais pertinentes para análises focadas na experiência gastronômica.

### **3. Características do Conjunto de Dados**

A base de dados utilizada é composta por avaliações de estabelecimentos alimentícios e conta com as seguintes colunas:

`restaurant_id`: Código único de identificação do restaurante.

`review_period`: Intervalo de tempo em que a avaliação foi registrada.

`review_period_description`: Descrição textual do período da avaliação.

`reviewer_type`: Perfil do avaliador (por exemplo, cliente ocasional, visitante).

`review_text`: Texto da avaliação feita pelo cliente.

`rating`: Nota dada ao estabelecimento.

A análise se concentrou na coluna `review_text`, por conter o conteúdo textual a ser examinado.

### **4. Metodologia**

A abordagem adotada envolveu os seguintes passos:

Importação dos dados: O arquivo CSV foi carregado e os comentários foram extraídos.

Tratamento prévio: Textos em branco foram substituídos por strings vazias, assegurando o bom funcionamento do processo de vetorização.

Transformação com TF-IDF: As avaliações foram convertidas em vetores numéricos com o método TF-IDF, que pondera os termos conforme sua importância no documento e no conjunto total.

Criação do vetor de consulta ("food"): A palavra-chave foi vetorizada no mesmo espaço para comparação.

Cálculo da Similaridade do Cosseno: Foi aplicada essa métrica para medir o quão próximos semanticamente estão os textos em relação ao termo "food", sendo valores mais próximos de 1 indicativos de maior similaridade.

Ordenação dos resultados: As avaliações foram classificadas do maior para o menor grau de similaridade com a palavra analisada.

## 5. Resultados

Os cinco comentários mais relacionados ao termo "food" foram:

Texto da Avaliação	Similaridade com "food"
As a food blogger always eager to explore new ...	0.1676
As a food blogger exploring affordable dining ...	0.1426
I recently visited this cozy spot during my fo...	0.1347
As a food blogger constantly scouting for hidd...	0.1336
As a food blogger always hunting for hidden ge...	0.1331

Observa-se que todas as avaliações destacam diretamente a temática "food", evidenciando que o algoritmo conseguiu identificar corretamente os textos mais relevantes ao contexto alimentar.

## 6. Conclusão

A técnica de Similaridade do Cosseno aliada à vetorização TF-IDF se mostrou eficaz para detectar quais avaliações têm maior afinidade com a palavra "food". A mesma metodologia pode ser aplicada a outras palavras-chave, como "serviço", "preço" ou "ambiente", contribuindo para análises temáticas automatizadas em grandes volumes de dados textuais.