

## ESTUDO DE CASO: ANÁLISE DE SENTIMENTO DE AVALIAÇÕES DE PRODUTOS

**OBJETIVO:** Permitir com que o aluno possa aplicar os conceitos teóricos iniciais relacionados ao Processamento de Linguagem Natural.

**INSTRUÇÕES:** A atividade pode ser realizada em duplas, porém a entrega deve ser realizada de forma individual.

Uma empresa de e-commerce está enfrentando dificuldades para entender o feedback de seus clientes. A empresa vende uma ampla gama de produtos, e os clientes frequentemente deixam avaliações e comentários sobre suas compras. No entanto, com milhares de avaliações sendo postadas diariamente, torna-se impossível para a equipe de suporte ao cliente ler e analisar manualmente todas elas.

A empresa deseja automatizar a análise de sentimentos das avaliações dos clientes para entender melhor como os clientes se sentem em relação aos produtos. Isso permitirá que a empresa identifique rapidamente problemas com produtos específicos, melhore a satisfação do cliente e tome decisões mais informadas sobre quais produtos promover ou descontinuar.

Suponha que você foi contratado para desenvolver uma solução usando os conceitos relacionados a Processamento de Linguagem Natural (PLN) tendo como objetivo principal classificar automaticamente as avaliações dos clientes como positivas, negativas ou neutras. Além disso, a empresa gostaria de identificar as principais palavras-chave e temas mencionados nas avaliações. Com base nas informações fornecidas, realize as atividades descritas abaixo:

a) Cada dupla deverá escolher uma empresa real que utilize um e-commerce ao qual seja possível coletar informações das avaliações reais realizadas por seus clientes.

b) Cada dupla deverá reunir um conjunto de avaliações de texto (reviews) que já estão rotuladas como positivas, negativas ou neutras (no mínimo 20 avaliações). Cada avaliação deve incluir o comentário (feedback) do cliente em forma textual e a classificação atribuída pelos clientes (por exemplo, 1 – 5 estrelas). Para isso, considere a seguinte classificação:

- 1 a 2 estrelas: avaliação negativa
- 3 estrelas: avaliação neutra

- 4 a 5 estrelas: avaliação positiva.

c) Realizar a limpeza dos textos das avaliações para remover ruídos, como pontuação, números, caracteres especiais e palavras irrelevantes. Em PLN, **palavras irrelevantes** são termos que não contribuem significativamente para o entendimento do contexto ou significado do texto. Exemplos de palavras irrelevantes:

- Artigos: "o", "a", "os", "as", "um", "uma" (em português), "the", "a", "an" (em inglês).
- Preposições: "de", "para", "em", "com", "por" (em português), "in", "on", "at", "by" (em inglês).
- Conjunções: "e", "mas", "ou", "se", "quando" (em português), "and", "but", "or", "if", "when" (em inglês).
- Pronomes: "eu", "você", "ele", "ela", "nós" (em português), "I", "you", "he", "she", "we" (em inglês).
- Outras palavras de uso frequente: Palavras como "isso", "aquilo", "então", "porque" (em português), "this", "that", "so", "because" (em inglês).

d) Divida cada avaliação em palavras (tokens). Por exemplo, a frase "O produto é excelente" se tornaria ["produto", "excelente"].

e) Criar uma base de dados que contenha os termos encontrados após a etapa anterior bem como a classificação das avaliações em positivas, negativas e neutras. Essa base deve ser dividida em 2 partes, treinamento e teste, utilizando a técnica de holdout estratificado.

f) A entrega da atividade deve ser realizada de forma individual, da seguinte forma: O aluno deverá entregar a base de dados criada (base de treinamento e teste) e um relatório técnico no formato da norma ABNT descrevendo todo o processo de solução deste estudo de caso.