



Sistemas de Informação
Centro de Informática
Universidade Federal de Pernambuco

Proposta de Projeto

Augusto Lima (arl4)
Carlos Melo (chnm)

Introdução

O site do “Reclame AQUI” é um canal em que pessoas podem expor suas insatisfações e alertar outros consumidores sobre empresas que já tenham tido algum tipo de experiência. Diariamente, mais de 600 mil pessoas pesquisam as reputações das empresas antes de realizar uma compra, contratar um serviço ou resolver um problema.

Pessoas que utilizam o Reclame AQUI (RA), são mais de 15 milhões de consumidores, procuram o site antes mesmo de ligar no SAC. Já as empresas, são mais de 120 mil cadastradas, utilizam o RA como referência para aprimorar seu atendimento. E o melhor de tudo é que o acesso a todo esse conteúdo não tem nenhum custo para o consumidor.

Além de fornecer um feed onde consumidores podem relacionar-se com empresas para solucionar seus conflitos o RA também fornece dados estatísticos entre outras informações que são de extrema importância para se avaliar uma empresa. Quantidade e porcentagem de reclamações feitas, de reclamações respondidas, de pessoas que voltariam a fazer negócio, índice de solução, palavras chaves de reclamações, tempo médio de resposta, nota da empresa, etc, são alguns dados encontrados no site.

Observando todas essas informações que são fornecidas pelo RA, surgiu nosso interesse em realizar uma análise mais detalhada sobre esses dados, utilizando conceitos visto na cadeira de “Data Science” que vão desde a extração dos dados no site, passando por análises exploratórias até chegar na parte de visualização e poder contar uma história sobre esses dados.

Objetivo

Nosso objetivo é coletar dados do Reclame AQUI, mais especificamente das empresas que são listadas nos principais rankings do site e a partir destes obter informações gerais das reclamações registradas para essas empresas para que posteriormente possamos realizar um estudo baseando-se nesses dados e conseguir confirmar ou detectar novas hipóteses, extrair padrões que definem a reputação das empresas e no final apresentar um resultado consolidado do nosso estudo que entregue algum valor para todos que tenham acesso esse trabalho.

Metodologia

Com BeautifulSoup iremos acessar e coletar os dados que iremos usar para o projeto, com auxílio da ferramenta Jupyter, e a linguagem de programação Python, que auxiliará com a extração, processamento, aplicação de técnicas e apresentação dos dados.

Utilizaremos da biblioteca Bokeh para gerar gráficos de dispersão para avaliar a influência das palavras chaves utilizadas nas reclamações de acordo com ranking das empresas. Também será utilizado para gerar gráficos de linhas para compararmos o uso

destas palavras nos rankings, e qual o índice de resolução de reclamação em contraste com a avaliações de estrelas das empresas.

E por fim tentaremos utilizar árvore de decisão para criar uma apresentação dos dados de como eles se comportam em meio aos rankings e visualizar de forma fácil a explicação para atribuição de um determinado valor ou classe para uma amostra.

Resultados Esperados

Gráficos, tabelas e árvores de decisão para a que a apresentação dos dados seja fácil de ser lida, compreendida e completa para o que objetivamos, um documento completo em Jupyter que traduzirá códigos em python e suas demais bibliotecas assim como Markdown para textos de resumo explicando cada gráficos, etc.