

Proposta de Tema Data Science

Arthur Bezerra de Freitas, Leonardo do Rêgo Espíndola, Matheus Barbosa
Domingues Feliciano

*Universidade Federal de Pernambuco – Centro de Informática
Caixa Postal 7.851 – 50.732-970 – Recife, PE – Brasil
abf2@cin.ufpe.br, lre@cin.ufpe.br, mbd@cin.ufpe.br*

Abstract

O Sistema Nacional de Informações de Defesa do Consumidor – Sindec é um sistema informatizado que integra processos e procedimentos, relativos ao atendimento aos consumidores nos Procons, visando proporcionar um instrumento de gestão. Este trabalho tem como objetivo estudar os dados obtidos pelo sistema das reclamações dos brasileiros à companhias e extrair informações utilizando as abordagens de data science a fim de entender como esse relacionamento entre cliente e fornecedor se dá.

1. Coleta

A base de dados escolhida para estudo foi coletada no repositório online dados.gov.br e descreve os dados de reclamações fundamentadas feitas por consumidores ao Procon entre os anos de 2009 e 2016. A base contém mais de 100
5 mil reclamações fundamentadas por ano, com informações detalhadas sobre a reclamação, a empresa e informações gerais sobre o consumidor, além de descrever se a reclamação foi atendida ou não, segundo os critérios do Procon. Os dados estão disponíveis em formato CSV separados por arquivos anuais.

2. Pré-Processamento

10 Os dados estão divididos em arquivos separados, sendo cada um com um ano especificado. Todos os arquivos serão compilados em um só, tendo assim os dados de todos os anos disponíveis. O conjunto de dados contém algumas

colunas com dados não relevantes para o nosso estudo, portanto serão removidas em conjunto com as colunas que contém dados repetidos. Serão criadas novas
15 colunas a partir das colunas presentes, como por exemplo período do atendimento a partir das colunas de “data de abertura” e “data de arquivamento”. Serão também analisadas colunas que tenham dados faltantes, para que sejam substituídos por NaN ou mesmo ter a remoção dessas linhas específicas.

3. Análise exploratória

20 A base reflete o histórico das empresas brasileiras em relação ao tratamento e resolução dos problemas trazidos pelos consumidores. Nosso intuito com esse estudo é entender as correlações dos dados, suas causas e a influência no resultado da resolução dos problemas e além disso verificar se há casos de discriminação com determinado tipo de consumidores.

25 Com esse fim, elaboramos algumas hipóteses para tentar entender melhor qual a relação das empresas com os consumidores:

- Faixa etária e Gênero tem influência no número total de reclamações
- Há mais problemas não resolvidos do que resolvidos
- Na indústria de cartões de crédito houve uma aumento de resoluções ao
30 longo do período de 2012 até 2016
- O sexo do consumidor, descrição do problema, período da reclamação são determinantes para deduzir a faixa etária do consumidor
- É possível inferir a partir do período da reclamação, companhia e descrição do problema se a reclamação foi resolvida ou não
- 35 • Dada a resolução, a indústria e a descrição do problema, determinar qual o sexo do consumidor
- O sexo do consumidor e faixa etária são determinantes para inferir se a reclamação teve como descrição “Venda Enganosa”

Iremos elaborar visualizações e modelos de regressão e classificação para
40 analisar e comprovar ou refutar as hipóteses propostas acima