**CLASSIFICAÇÃO AUTOMÁTICA DE RECLAMAÇÃO DA ANATEL**

Jozias Rolim de Araújo Júnior

Analista Programador Sênior - iColabora

**Resumo - Introdução**

Diariamente vários clientes realizam reclamações junto a Anatel, essas reclamações podem se originar de vários canais, como exemplos podemos citar o “Fale Conosco”, “0800”, entre outros. Todas essas reclamações são classificadas em uma arvore que define, entre outras coisas, o tipo do serviço, o motivo da reclamação e o submotivo da reclamação. Após a reclamação classificada ela é disponibilizada para os reclamados solucionarem o problema. A fase de classificação da reclamação se apresenta com uma das fases mais importante do processo, pois define entre outras coisas, o setor responsável dentro das empresas para solução da reclamação. A classificação demanda tempo e esforço do operador e ainda gera classificações equivocadas, o que acaba prejudicando o processo. Neste sentido, esse trabalho propõe uma forma automatizada para classificar reclamações da Anatel.

# Classificando com o Elasticsearch

### Como se comporta o Elasticsearch

O Elasticsearch se apresentou como uma possibilidade de classificação de texto, devido à forma com que ele indexa seus documentos, assim uma abordagem para o problema foi traçada. Os textos das reclamações foram submetidos a um “*analyzer”* que tratou as palavras de forma individual, como podemos observar na Figura1, e indexou o documento completo por meio delas. Neste sentido, as novas reclamações que possuírem os radicais “reclam” e “verdad”, possuirão uma grande semelhança com o anterior, sugerindo ao algoritmo suas características para classificar o novo documento, seguindo o calculo de similaridade do Elasticsearch que é a representação TF-IDF. Podemos observar ainda que algumas palavras não foram indexadas como é o exemplo dos temos “eu” e “de”, isso ocorre por serem termos comuns e não serem suficientemente representativos para a classificação do documento, ainda foram excluídos as pontuações e acentos, todas as palavras foram transformadas em minúsculas, palavras que possuem sentido com termos separadas foram mantidos, como exemplo, podemos apresentar “wi-fi”, “São Paulo”.

Tipo de serviço: Serviço

Motivo: Motivo

Submotivo: Submotivo

Reclamação: eu **reclam**o, de **verdad**e

reclam

verdad

Muitas vezes, os termos presentes nos documentos são distintos, assim o Elasticsearch retorna os documentos com maior relevância, uma consulta com apenas a palavra “reclamação”, retornaria ainda assim o documento citado como exemplo anteriormente, mesmo não sendo 100% similar ao anterior.

### Definição da base de reclamações

Para o teste de classificação de reclamações, foi escolhida a base de reclamações da OI, do canal Fale Conosco, totalizando 18887 reclamações, sendo 12000 para treino e treino e 6887 para validação da estratégia. As reclamações foram classificadas em xxx classes, quanto ao tipo de serviço.

### Resultados obtidos

Para avaliar os resultados foram utilizadas quantidade de documentos similares distintos, variando de 1 a 16 documentos e verificando suas respectivas classes. A escolha da classe se deu pelo método de classe majoritária, ou seja a classe que se repetiu mais vezes nos documentos retornados, quando havia empate entra a quantidade de classes uma era escolhida aleatoriamente e retornada como resposta. Os resultados podem ser vistos no gráfico da Figura2.