

Grupo 3

Sprint 1

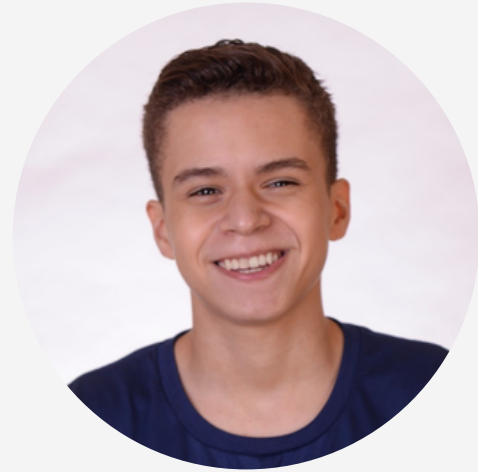
MoshiMoshiBot

Agenda

1 Análise Exploratória

2 Pré-processamento

Time



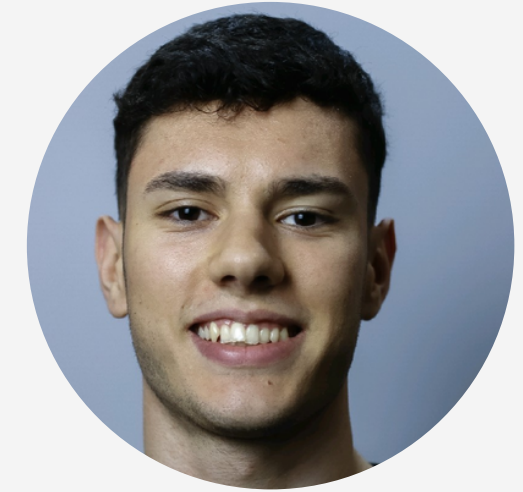
Gabriel Carneiro



Pedro Romão



João Alcaraz



Gabriel Nhoncanse



Yasmin Rocha



Sarah Ribeiro



Pedro Munhoz

Análise exploratória

A análise



EXEMPLOS + FAQ

Utilizamos os dados disponíveis nos exemplos fornecidos (~500 P/R) e no FAQ encontrado no site oficial.



OBJETIVO

Foram definidas perguntas sobre os dados, a fim de entendermos melhor a base que temos à disposição e identificarmos possíveis padrões nas mensagens.

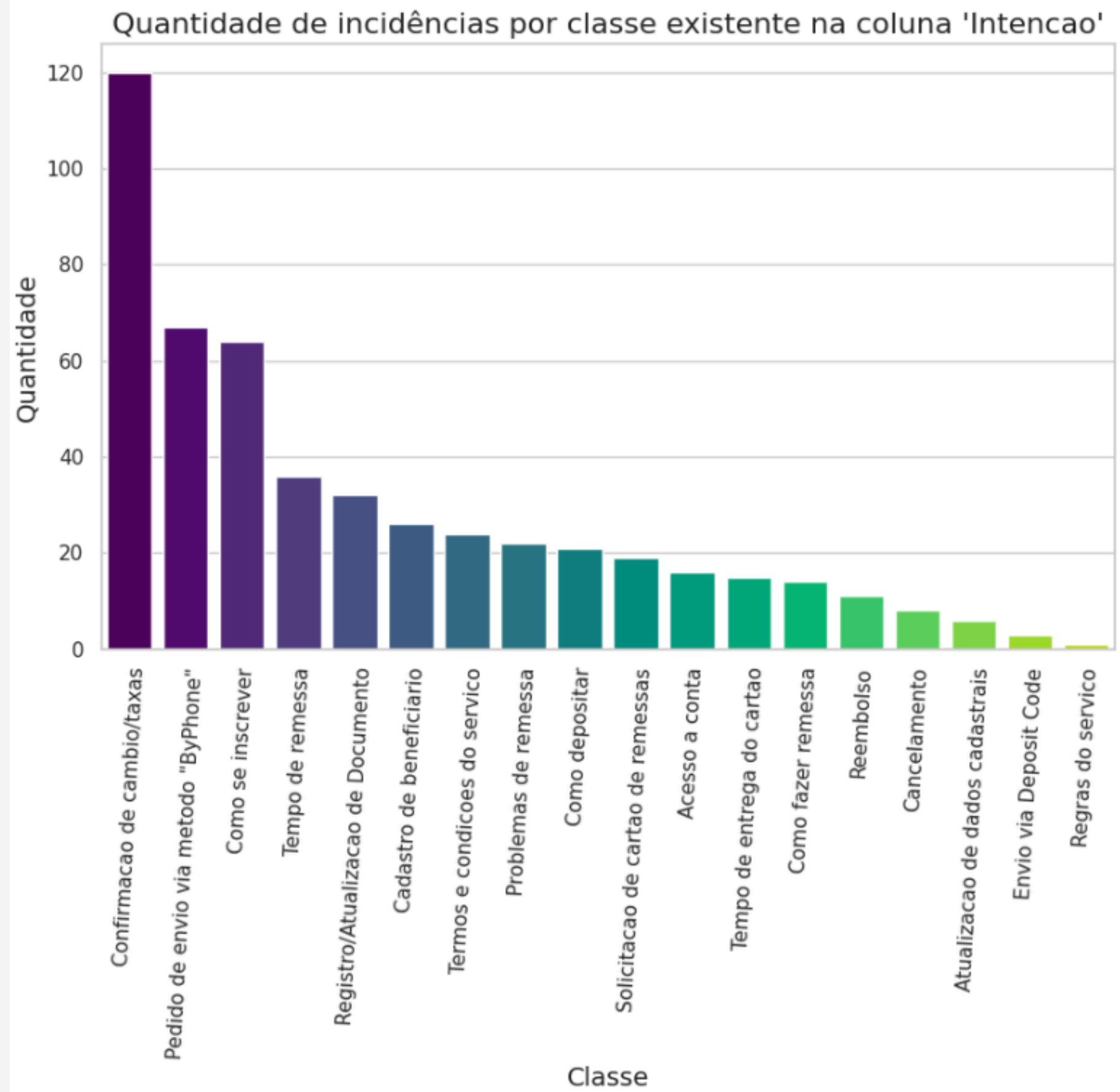


TESTES E RESPOSTAS

Por meio da análise exploratória, obtivemos respostas às nossas perguntas.

Exemplo

- Quantidade de perguntas feitas por classe de intenção.



Pré-Processamento

Correção da Gramática

- Remoção de abreviações
- Correção de ortografia

Exemplo

Original: “Bom dia. Sim, o sr pode utilizar o cartão de depósitos nesta máquinas 24 horas exclusivas do JP Bank.”

Corrigida: “Bom dia. Sim, o senhor pode utilizar o cartão de depósitos nestas máquinas 24 horas exclusivas do JP Bank.”

Reconhecimento de Nomes de Entidades

- Detecta nomes próprios na frase de entrada
- Retorna todas as entidades

Exemplo

Original: “Bom dia. Sim, o senhor pode utilizar o cartão de depósitos nestas máquinas 24 horas exclusivas do JP Bank.”

Resultado: JP Bank

Tokenização

- Divisão de um texto em partes menores, palavras, no nosso caso
- Remoção de pontuação

Exemplo

Original: “Bom dia. Sim, o senhor pode utilizar o cartão de depósitos nestas máquinas 24 horas exclusivas do JP Bank.”

Resultado: Bom | dia | Sim | o | sr | pode | utilizar | o | cartão | de | depósitos | nesta | máquinas | 24 | horas | exclusivas | do | JP | Bank

Remoção de Stop Words

- Eliminação de palavras irrelevantes para o processamento do modelo
- A dimensionalidade do problema é diminuída

Exemplo

Original: Bom | dia | Sim | o | senhor | pode | utilizar | o | cartão | de | depósitos | nesta | máquinas | 24 | horas | exclusivas | do | JP | Bank

Resultado: Bom | dia | Sim | senhor | pode | utilizar | cartão | depósitos | nesta | máquinas | 24 | horas | exclusivas | JP | Bank

Lematização

- Processo de simplificar palavras
- Importante para a rede reconhecer de forma mais simples os padrões entre os tokens

Exemplo

Original: Bom | dia | Sim | senhor | pode | utilizar | cartão | depósitos | nesta | máquinas | 24 | horas | exclusivas | JP | Bank

Resultado: Bom | dia | Sim | senhor | poder | utilizar | cartão | depósito | em | esta | máquina | 24 | hora | exclusivo | JP | Bank

**MUITO
OBRIGADO!**