



Pyght Sentiment Time

Apresentação do projeto e
análise exploratória

INTRODUÇÃO



- **Proposta do Projeto**

Busca-se a criação de aplicação baseada em análise de sentimentos que apresente graficamente a relação entre a polaridade e o período.

- O que é “análise de sentimento” no sentido do projeto?

A *análise de sentimento* (do inglês, *sentiment analysis*) ou *mineração de Opiniões*, identifica e analisa opiniões, sentimentos, avaliações, atitudes e emoções das pessoas em relação a algum tema específico ou de acordo com parâmetros previamente estabelecidos.

- Como será utilizada?

No projeto que ora se apresenta, a análise se sentimento avalia mensagens da rede social Twitter, por meio de *classificação por polaridade (positivo, neutro e negativo)* dos tweets ou temas centrais extraídos deles.

- **Conjunto de dados**

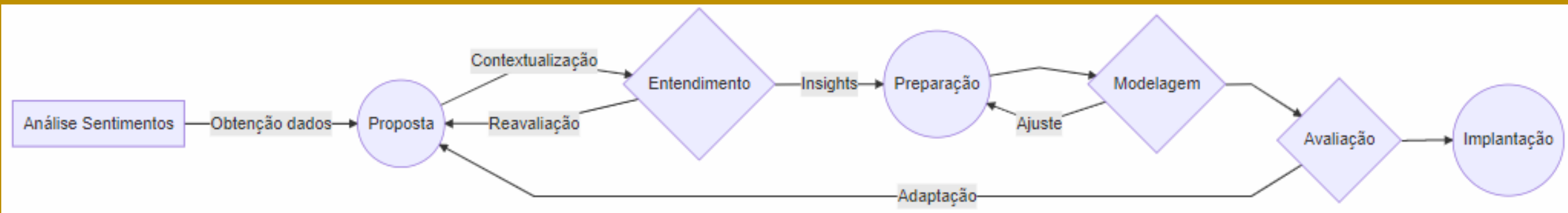
Conjunto de dados obtidos do Twitter entre 01/08/2018 e 20/10/2018. Contém 5 variáveis, inclusive classificação de sentimentos, no caso por polaridade (*positivo, negativo, neutro*).



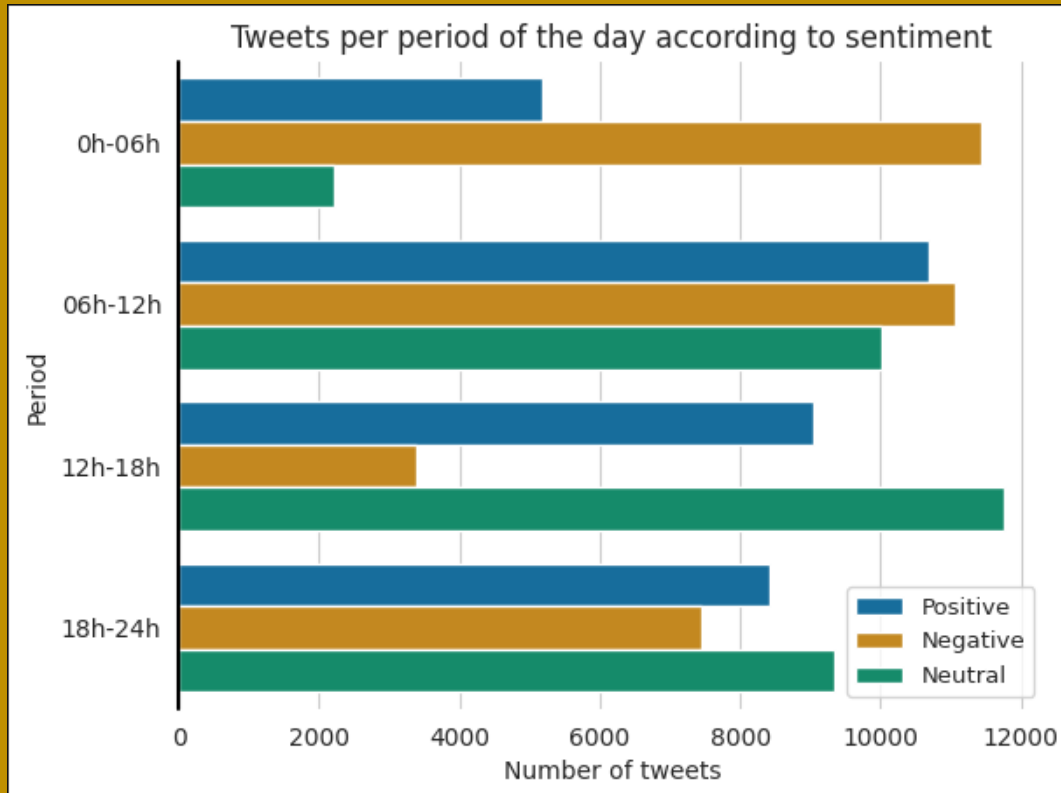
METODOLOGIA



- O projeto adota a **metodologia CRISP-DM** para desenvolvimento das etapas de transformação de dados e produção de conhecimento de forma gerenciável.
- Com esse fundamento, terá como propósito as seguintes etapas:
 - 1ª Fase: entrega da proposta e entendimento dos dados;
 - 2ª Fase: entrega da preparação e modelagem dos dados;
 - 3ª Fase: avaliação do projeto e implantação da aplicação.
- Fluxo de execução do projeto:



ANÁLISE EXPLORATÓRIA



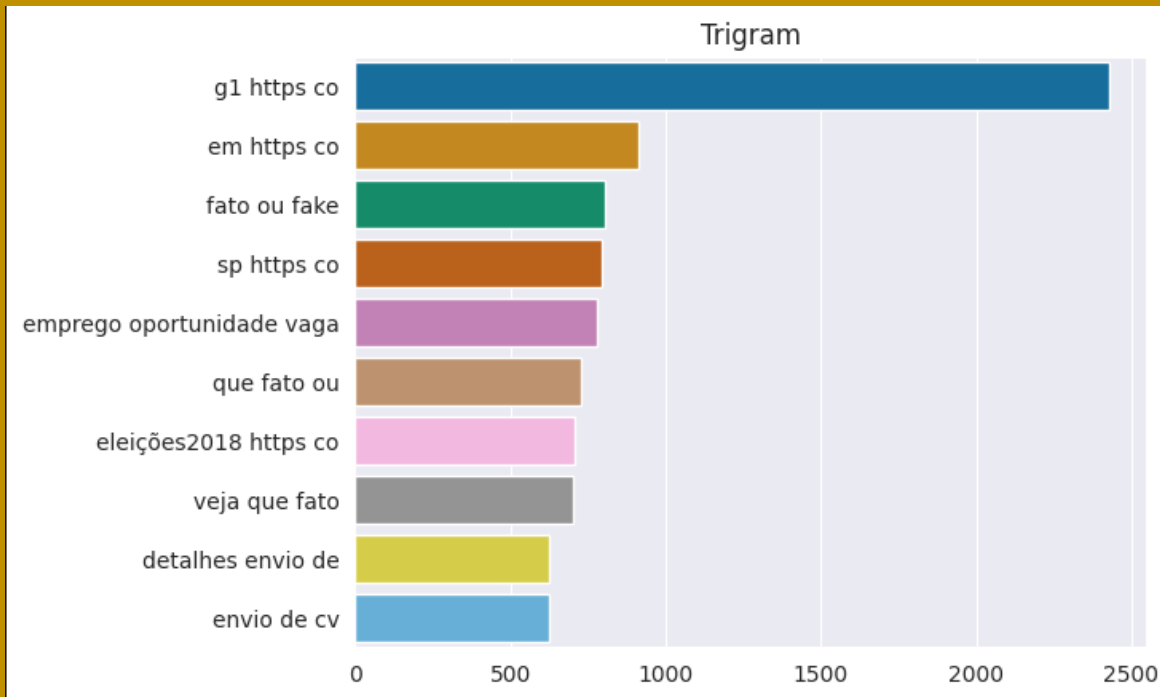
- **Explorando os dados**

- “ Na fase inicial, extrai-se que 5 variáveis foram utilizadas, representando: *identificação única de usuário, texto da mensagem publicada, horário de publicação da mensagem, classificação por polaridade do sentimento e palavra relevante.*
- “ Necessidade de padronização das variáveis (*lowercase*) e a conversão dos *horários de publicação da mensagens* para UTC-3, adequando ao horário nacional.
- “ Agrupamento dos *horários das publicações das mensagens* em intervalos, para simplificar a análise inicial.
- “ Contagem dos *tweets* por intervalo de horas e classificação de sentimento por polaridade.
- “ Identificação de *stopwords* atravessando as análises e necessidade de removê-las.
- “ O período entre 0-06h aconteceu o maior número de publicações negativas, entre 12h-18h de publicações positivas e 06h-12h publicações totais.

- **Questionamentos**

- ◇ *Em alguns intervalos do dia existem mais tweets do que em outros?*
- ◇ *Qual correlação existente entre polaridade do sentimento e o período do dia?*
- ◇ *O conteúdo do tweet é relevante para a avaliação do sentimento e/ ou período do dia?*

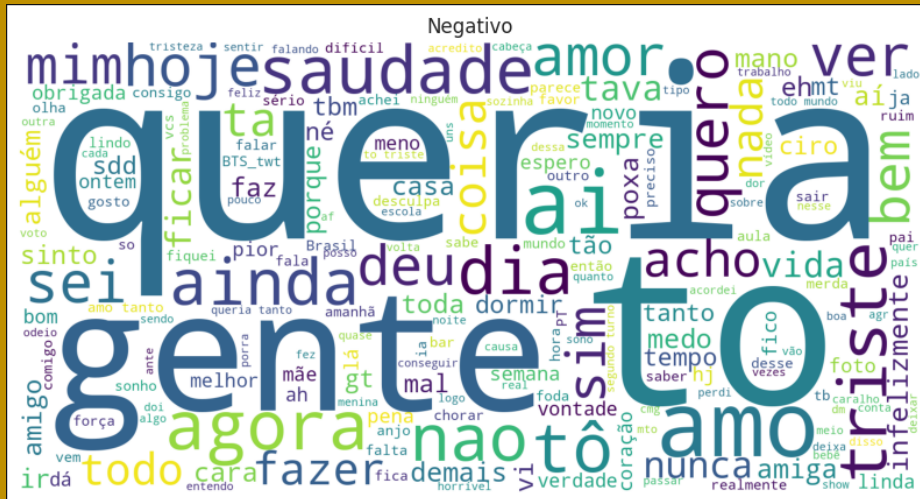
ANÁLISE EXPLORATÓRIA



• Hipóteses

- γ Disponibilidade das pessoas pode variar conforme o tempo e influenciar no envio de mensagens.
- γ Fatores externos podem influenciar nos assuntos mais utilizados em tweets.
- γ Conteúdos variam com o passar dos dias, dando dinamicidade às análises.
- γ Conteúdo é avaliado por palavras-chave e mostram uma tendência de polaridade do sentimento em relação ao assunto dos tweets.
- γ Condições de estabilidade emocional podem influenciar, como insônia, estresse ou estado de ânimo, podem influenciar na disposição de publicações e na respectiva polaridade.

ANÁLISE EXPLORATÓRIA



- Insights

- ✦ Como observado nos gráficos, os períodos de positividade e negatividade oscilam durante o dia.
- ✦ A variabilidade de assuntos durante períodos do dia e dias da semana difere para sentimento positivos e negativos.
- ✦ Fatores pessoais, sociais e regionais podem influenciar na evolução de cada assunto, alterando conforme determinados períodos e dias.
- ✦ Palavras-chave usualmente são utilizadas e repetidas, permitindo identificar sentimentos, em uma classificação por polaridade.
- ✦ Da mesma forma, podem auxiliar a estabelecer padrões de sentimentos e assuntos, conforme períodos do dia e dias da semana.
- ✦ Tendência de maior disponibilidade de publicações (totais) no período entre 06h-12h e publicações negativas entre 0h-06h, por exemplo.



OBRIGADO!!!

Integrantes da Equipe:

- Amanda
- Ítalo
- José Aurelio
- Kaio
- Leonardo
- Rayanne
- Renato
- Tarciano

Github

github.com/atlantico-academy/twitter-sentiment-analysis