




Eleições, Redes Sociais e Data Science



Agenda

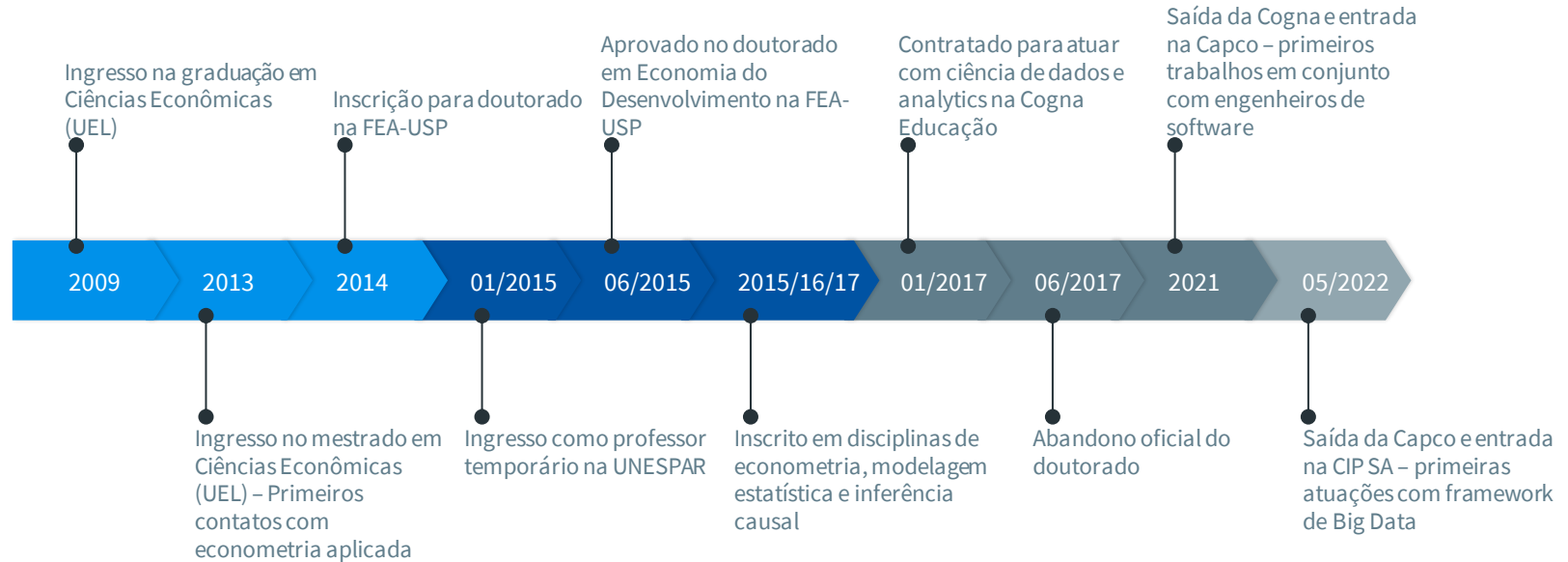
1. Sobre o palestrante
 2. O que é esta apresentação?
 3. Ideia original (e porque ela não foi executada)
 4. Como ter acesso à API do Twitter
 5. Campo de atuação do projeto
 6. Extração, organização e armazenamento dos dados
 7. Análises e resultados iniciais
 8. Ideias para o futuro
- 

A decorative network diagram in the top-left corner, featuring a complex web of interconnected nodes and lines, with some nodes highlighted in blue.

1.

Sobre o palestrante

Minha linha do tempo



A decorative network diagram in the top-left corner, featuring a complex web of interconnected nodes and lines, with some nodes highlighted in blue.

2.

**O que é esta
apresentação?**

É e Não É

Não é

- ⊙ Uma aula de modelos de Machine Learning
- ⊙ Propaganda a favor de um candidato específico
- ⊙ Uma análise prescritiva
- ⊙ Um julgamento de qual a melhor campanha

É

- ⊙ A história sobre como uma ideia surgiu e foi aplicada
- ⊙ A demonstração dos resultados obtidos até o momento
- ⊙ A apresentação dos próximos passos

A decorative network diagram in the top-left corner, featuring a complex web of interconnected nodes and lines. The nodes are represented by small circles, some of which are larger and have concentric circles, suggesting a hierarchical or central structure. The lines are thin and gray, connecting the nodes in a non-linear fashion.

3. **A ideia original**

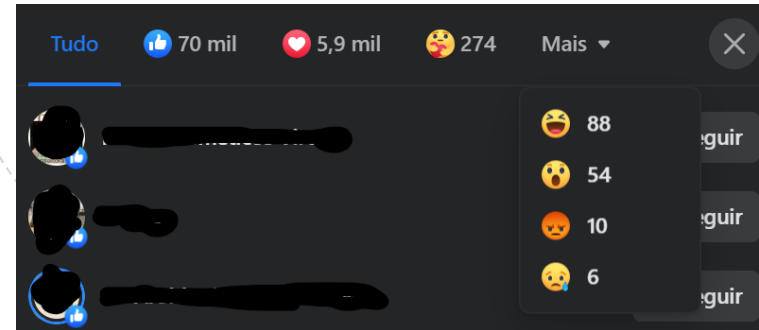
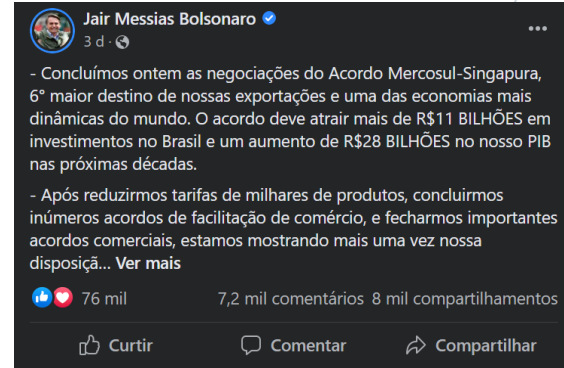


A política é quase tão excitante como a guerra e não menos perigosa. Na guerra a pessoa só pode ser morta uma vez, mas na política diversas vezes.

Winston Churchill

Facebook: reações e comentários

Ser capaz de extrair a contagem de reações e comentários das postagens de políticos no facebook seria uma análise interessante. Uma espécie de “termometro” de rede social



Porém...

Após o escândalo da Cambridge Analytica as APIs do facebook se tornaram mais restritas. Enquanto o twitter se mostrou uma rede mais interessante.





4.

Acesso à API do Twitter

Twitter para desenvolvedores

A pessoa interessada pode criar uma conta gratuita no site do Twitter para Desenvolvedores e criar diferentes projetos que vão permitir acesso às APIs do Twitter

Twitter para devs:

<https://developer.twitter.com/en>

Overview sobre como os projetos funcionam:

<https://developer.twitter.com/en/docs/projects/overview>

Diferentes níveis de acesso às APIs do twitter

Essential access

Nível de acesso fornecido a qualquer pessoa que tenha se cadastrado no Twitter para desenvolvedores. Permite a extração de 500 mil tweets por mês

Elevated access

Nível de acesso fornecido para pessoas que tenham sido aprovadas ao acesso adicional. A requisição de acesso deve ser feita no portal para desenvolvedores.

Permite a extração de 2 milhões de tweets por mês

Academic research access

Nível de acesso fornecido para pesquisadores que tenham sido aprovados ao acesso adicional e que atendam critérios específicos (motivação da pesquisa, contribuição social, etc).

Permite a extração de 10 milhões de tweets por mês

A decorative network diagram in the top-left corner, featuring a complex web of interconnected nodes and lines, with some nodes highlighted in blue.

5.

Campo de atuação do projeto

Quem analisar e o que analisar?

Candidatos

- ◎ Jair Bolsonaro
- ◎ Lula da Silva

O que analisar

- ◎ Textos das postagens dos próprios candidatos
- ◎ Textos das respostas dadas a estes candidatos

Duas análises iniciais

Timelines: identificação de assunto/tópico

“Nós vamos melhorar os indicadores de educação do Brasil” = **EDUCAÇÃO**

“Precisamos lutar contra as fakenews” = **FAKENEWS**

“Nunca antes na história deste país os índices de criminalidade foram tão baixos” = **CRIMINALIDADE**

Replies: análise de polaridade/sentimento

“Quando você acredita que teremos a aprovação de um novo marco institucional” = **NEUTRO**

“Você é um imbecil, jamais vou votar em você” = **NEGATIVO**

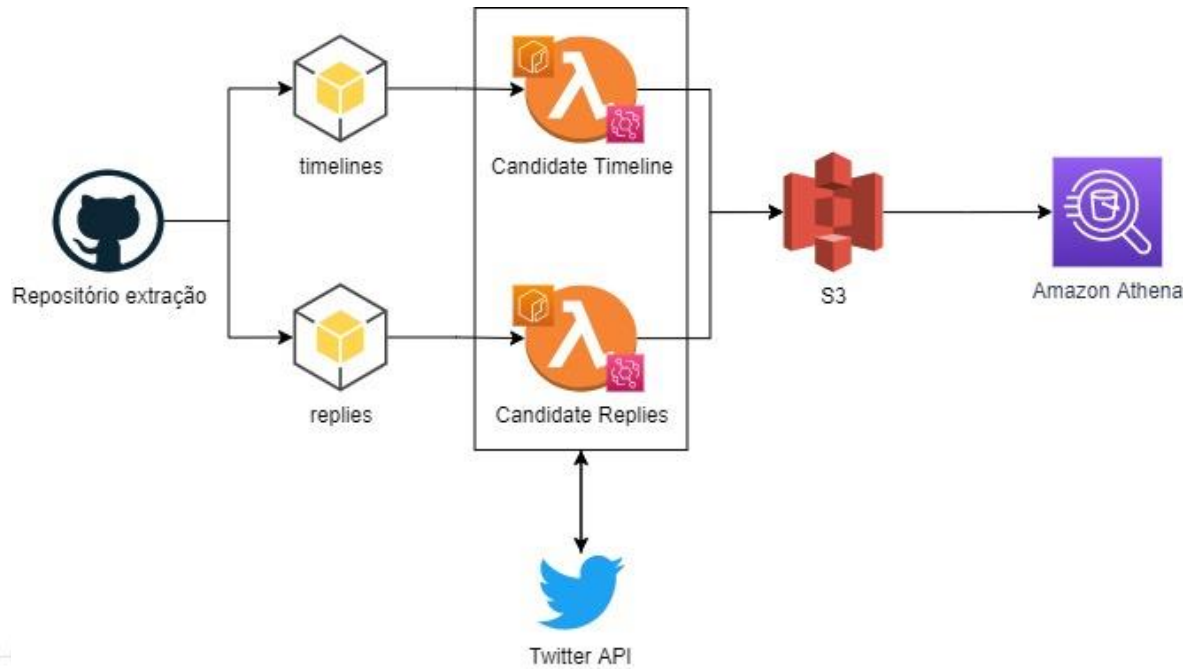
“Avante meu candidato, seremos vencedores em 2022” = **POSITIVO**

A decorative network diagram in the top-left corner, featuring a cluster of interconnected nodes and lines, rendered in a light gray color.

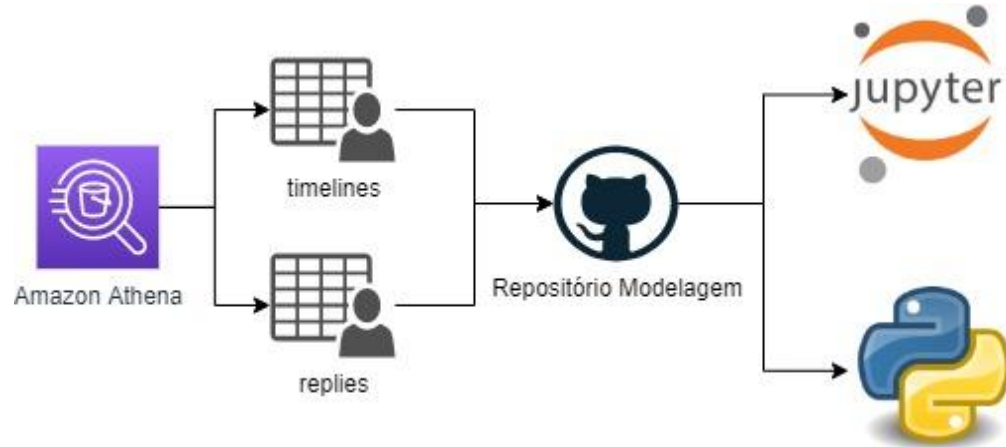
6.

Extração, organização e armazenamento dos dados

Captura dos dados



Gestão dos dados para análise



A decorative network diagram in the top-left corner, featuring a complex web of interconnected nodes and lines, rendered in a light gray color.

7.

Análises e resultados iniciais

Tratamento do texto

Post original:

“Sergio Rolando Lero Moro, sob domínio do MBL, continua fugindo do debate, mas agora com pose de falso machão. A máscara de enganador, no entanto, não desprega do rosto. Deixa de frouxidão, cara, e escolha hora, local e tema !”

Texto após limpeza:

“sergio rolando lero moro dominio mbl continua fugindo debate agora pose falso machao mascara enganador entanto desprega rosto deixa frouxidao cara escolha hora local tema”

Tokens:

sergio / sergio rolando / rolando / rolando lero / lero / lero moro / moro / moro dominio / mbl / mbl continua / continua / continua fugindo / fugindo / fugindo debate / debate / debate agora / agora / agora pose / pose / pose falso / falso / falso machao / machao / machao mascara [...]

Análise de tópicos: modelagem

Após a criação dos tokens:

- ⦿ É calculada a frequência de aparição de cada token
- ⦿ Os resultados da contagem são transformados em uma matriz
- ⦿ Um algoritmo (LSA) de redução de dimensionalidade é treinado para encontrar conjuntos de tokens que pertencem a um determinado tópico

Análise de tópicos: batismo dos tópicos

1	2	3	4
bom dia	povo brasileiro	quero representar	gasolina gas diesel
bom dia boa	trabalhar ter	reconstruir brasil	gas diesel
dia boa	nesse pais	dar oportunidade	gasolina gas
boa semana	voltar ser	dar esperanca	aumento gasolina
bom dia boa semana	ser grande	apenas pt quero representar movimento quero reconstruir brasil	pessoas faziam domingo simbolo dias melhores quanto custando

1. Bom dia
2. País voltar a ser grande
3. Reconstruir o Brasil
4. Preço dos combustíveis

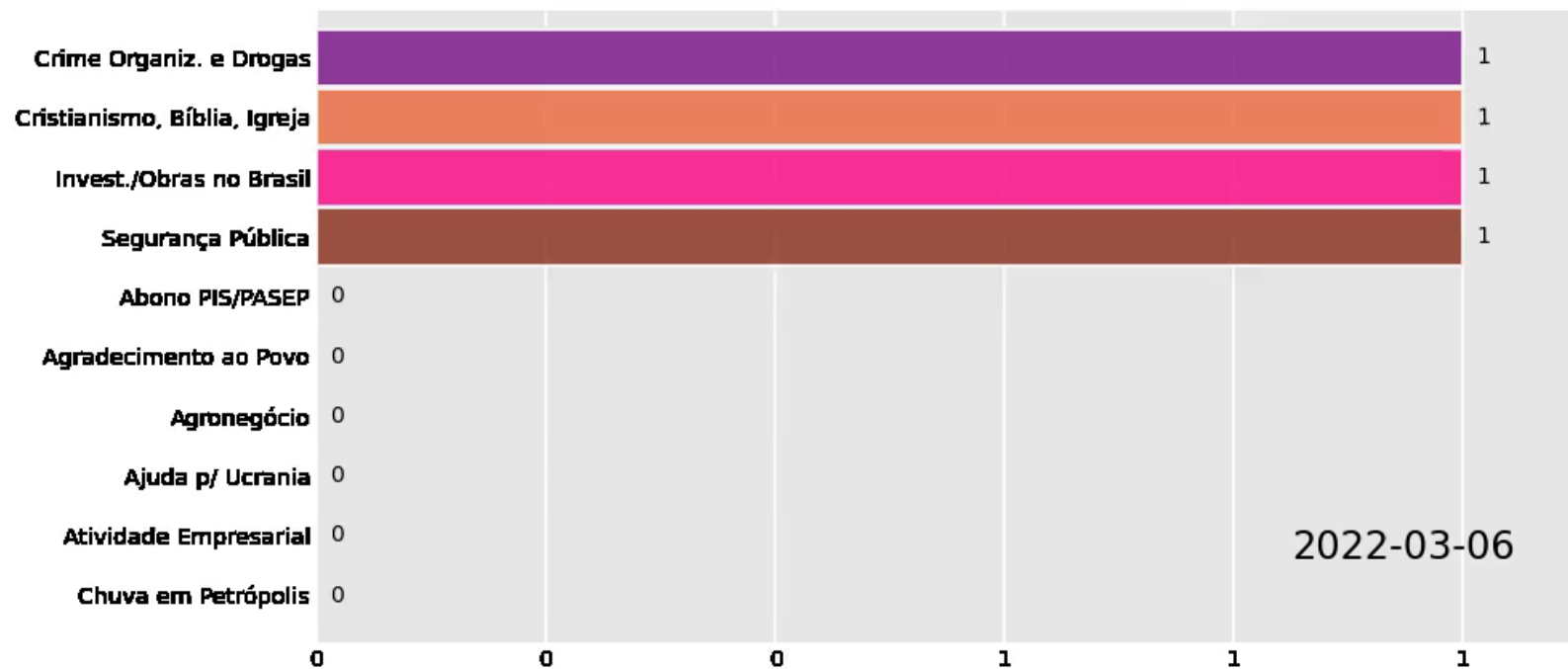
Análise de tópicos: reaplicação do modelo aos tweets

Após a criação da lista de tópicos:

- ⦿ Cada tweet do candidato é lido pelo modelo
- ⦿ O modelo informa qual o tópico mais presente em cada tweet
- ⦿ É gerada uma contagem de tweets-tópicos abordados pelo candidato para cada dia

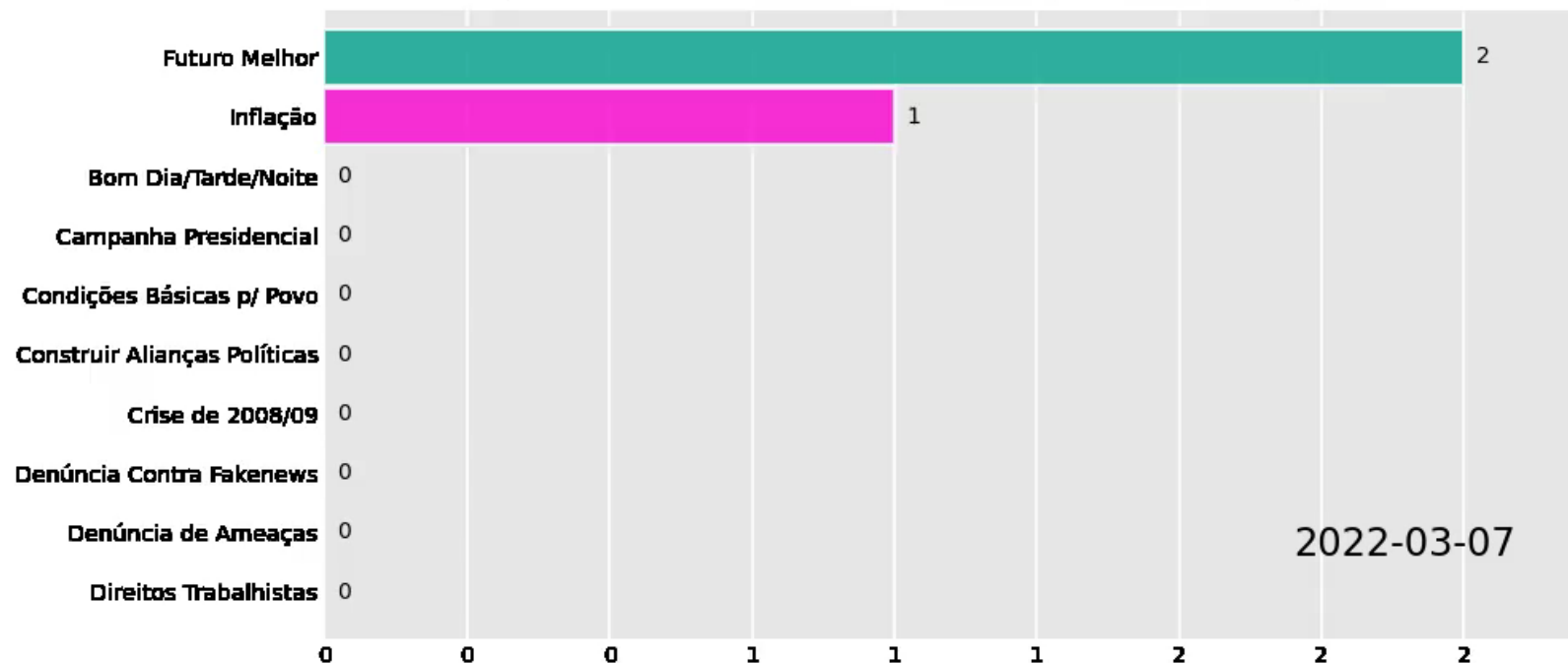
Análise de tópicos: resultados

Top 10 Assuntos do Bolsonaro no Twitter em Março 2022



Análise de tópicos: resultados

Top 10 Assuntos do Lula no Twitter em Março 2022



Análise de sentimento: tweets para treinamento do modelo

Amostra:

- ⦿ Dados de respostas extraídos durante 7 dias em fevereiro (09/02 a 15/02)
- ⦿ Tweets lidos um a um e classificados em sentimento negativo, neutro ou positivo

Análise de sentimento: classificação manual de tweets

Bolsonaro

“Vale ressaltar que o comunismo matou bem mais do que o nazismo” = **NEUTRO**

“Hipócrita: cê plagiou o lema de Hitler na Campanha eleitoral de 2018.” = **NEGATIVO**

“A gente que agradece tudo o que o senhor vem fazendo pelo povo nordestino” = **POSITIVO**

“Obrigado presidentes Lula e Dilma.” = **NEGATIVO**

“Cadê o ladrão safado, que lidera todas as pesquisas, vem Lula dá uma volta no nordeste” = **POSITIVO**

Lula

“Porque o Capitalismo é mais forte que Jesus.” = **NEUTRO**

“Vai dizer milhares de vezes. Em todas as obras que vocês iniciaram, saquearam e não concluíram, seu corrupto.” = **NEGATIVO**

“Minha filha vai tirar o título de eleitor esse ano para votar em você!! Vamos em frente e de preferência com a vitória no 1º turno!!!” = **POSITIVO**

“Viva nosso MITO, Jair Messias Bolsonaro, que, FINALMENTE, levou ÁGUA para o Nordeste. E os nordestinos são agradecidos!” = **NEGATIVO**

“DESMASCARANDO O GOVERNO CORRUPTO DO BOLSONARO” = **POSITIVO**

Análise de sentimento: classificação manual de tweets

Bolsonaro

- Total de tweets: 39.944
- Tweets classificados: 1.968
- Positivos: 32%
- Negativos: 31%
- Neutros: 37%

Lula

- Total de tweets: 85.180
- Tweets classificados: 4.631
- Positivos: 08%
- Negativos: 67%
- Neutros: 25%

Análise de sentimento: modelagem

Após a criação dos tokens:

- ⦿ É calculada a frequência de aparição de cada token
- ⦿ Os resultados da contagem são transformados em uma matriz
- ⦿ Um algoritmo (Reg. Log) de classificação é treinado para aprender as características dos tweets negativos e não negativos

Análise de sentimento: modelagem

Bolsonaro

- ◎ Precisão (Neg): 79%
- ◎ Revocação (Neg): 68%
- ◎ Acuracidade: 84%

Lula

- ◎ Precisão (Neg): 90%
- ◎ Revocação (Neg): 90%
- ◎ Acuracidade: 87%

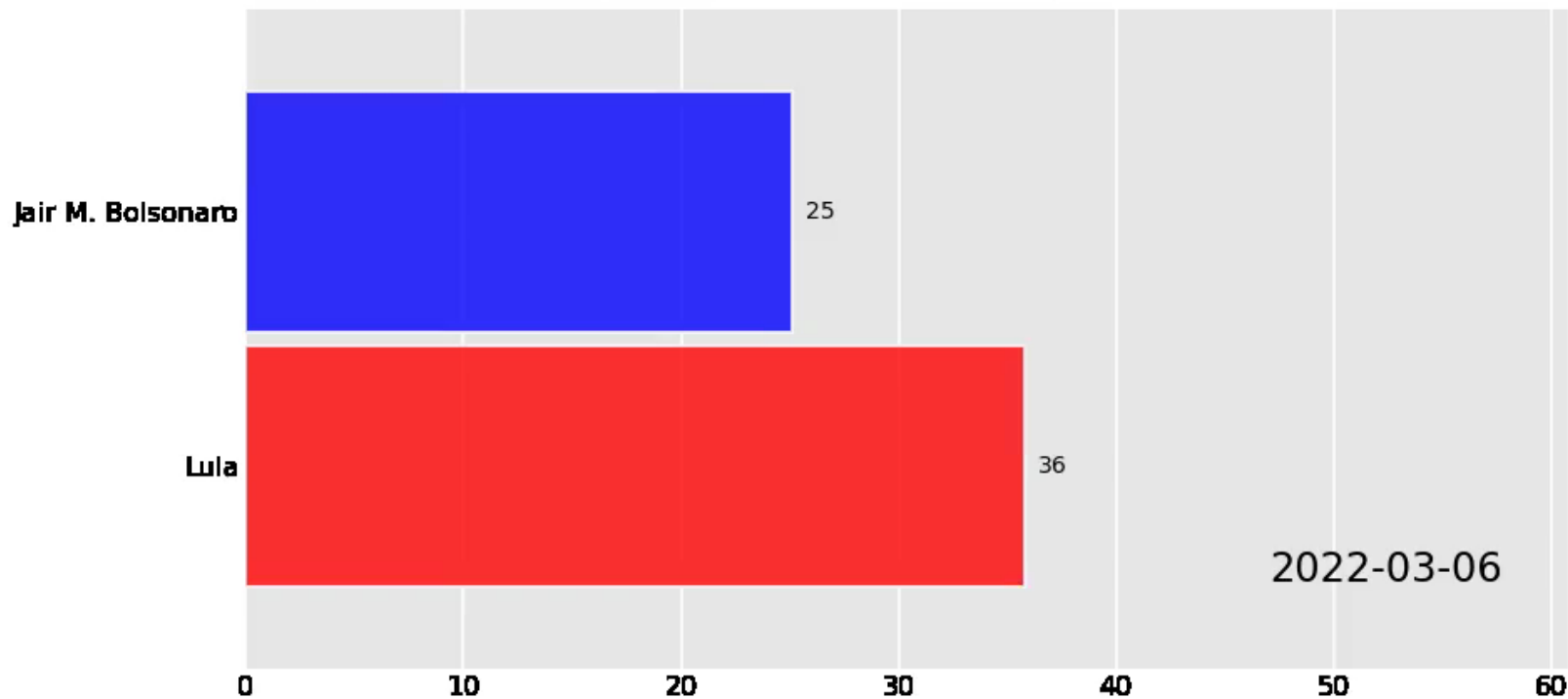
Análise de sentimento: modelagem

Após o treinamento do modelo:

- ◎ O modelo de machine learning é salvo no repositório
- ◎ Tweets mais recentes, sem classificação manual, são fornecidos para o modelo armazenado
- ◎ A partir destas previsões é possível analisar a evolução do percentual de negatividade (hate) recebido pelos candidatos

Análise de sentimento: resultados

Hate recebido pelos candidatos Lula e Bolsonaro no Twitter



A decorative network diagram in the top-left corner, featuring a complex web of interconnected nodes and lines, with some nodes highlighted in blue.

8.

Ideias para o futuro

Possibilidades

- ◎ Mapa de calor dos horários em que os candidatos postam
- ◎ Índice que sintetiza a similaridade dos tópicos dos tweets com os tópicos dos planos de governo
- ◎ Adicionar análises do Ciro Gomes
- ◎ Aumentar tweets classificados manualmente



Obrigado!

Para contato:

flaviokfiuza@gmail.com

LinkedIn: Flavio Fiuza de Moura

