

# Atividade 03 - Projeto e Correlatos

## Equipe

**Nome:** Detetives Virtuais

**Repositório GitLab da equipe:** [https://gitlab.com/jr\\_waine/csb-51](https://gitlab.com/jr_waine/csb-51)

**Membros:**

- Waine Barbosa de Oliveira Junior, 1905120, jr\_waine, [waine@alunos.utfpr.edu.br](mailto:waine@alunos.utfpr.edu.br), Eng. Comp, UTFPR
- Thiago Schinda Bubniak, 1540778, thiagobubniak, [thiagobubniak@alunos.utfpr.edu.br](mailto:thiagobubniak@alunos.utfpr.edu.br), Eng. Comp, UTFPR

## Tema

Compreender quais os temas de fake news mais relevantes em determinado período e comparar com eventos relacionados no período

## Perguntas de pesquisa

- Quais foram os tópicos mais comentados nas notícias falsas?
- Como esses tópicos evoluíram com o passar do tempo?

## Hipóteses

- Os tópicos das notícias falsas evoluíram conforme eventos políticos (e.g. perto do julgamento do ex-presidente, as notícias falsas sobre eles aumentaram)
- Os tópicos das notícias falsas e verdadeiras, num mesmo intervalo de tempo, tendem a ser os mesmos

## Dados e modelos

Com relação aos dados, pretendemos usar o [corpus Fake.br](https://corpus.fake.br) como base. Caso seja necessário, para uma análise com dados mais representativos de um intervalo de tempo, cogitamos limitar o intervalo de tempo das notícias utilizadas. Porém isso requer uma investigação melhor com relação a adequação dos dados ao modelo.

Para a detecção de tópicos, devemos utilizar o [LDA \(latent Dirichlet allocation\) \(Blei et. al\)](#). Para as alterações temporais, modelos que utilizam como base o LDA (ou semelhante), como [Topics Over Time \(Wang, McCallum\) \(implementação\)](#), o qual apresenta a evolução de tópicos no tempo. E também modelos como [Adaptive Window \(Bifet, Gavalda\) \(implementação\)](#), o qual apresenta as mudanças dos tópicos em um intervalo de tempo.

Os modelos pesquisados estão abertos a alterações conforme recomendações do professor.

# Cronograma

1. Investigar mais precisamente os resultados obtidos da aplicação do LDA;
  - a. Correlacionar os tópicos encontrados com as datas dos fatos ocorridos;
  - b. Otimizar os parâmetros do modelo a fim de obter resultados bons.
2. Utilizar modelo de *Topics Over Time*;
3. Realizar um comparativo dos resultados obtidos dos diferentes métodos aplicados;
4. Montar relatório com visualizações e insights encontrados.