Apresentação TCC 2º Semestre

O que foi feito:

- Rodei de novo o crawler de fake news para aumentar a amostragem

- Análise dos dados 🡪 mostrar gráficos

- Mostrar pipeline de NLP

- Limpeza

- Vetorização

- TF-IDF

- Word2Vec

- Modelo

- Arquitetura da aplicação

- Aplicativo

- API Rest

# Stripping accents

# Removing characters that aren't alphanumeric

# Removing tokens that intercalate between letters and digits

# Removing digits

# Removing tokens with letters that appear more than twice in a row

# Removing extra spaces

# Removing words with length equal or lower than 2 or are in STOPWORDS