



Relatório: Análise de Sentimentos de Tweets sobre Clubes de Futebol Ingleses

Eros Santos de Vasconcelos / 202307120545

Polo Iputinga-PE

RPG0034 - Dando inteligência ao software – 5º semestre

1. Introdução

Este projeto tem como objetivo realizar a análise de sentimento de mensagens (tweets) publicadas no Twitter (atualmente chamado de X), para avaliar a percepção dos brasileiros sobre clubes de futebol da Inglaterra. O objetivo final é fornecer insights para uma empresa interessada em abrir centros de treinamento esportivos no Brasil vinculados a esses clubes.

A análise de sentimentos é uma aplicação do Processamento de Linguagem Natural (PLN) que identifica a opinião expressa em textos — positiva, negativa ou neutra.

2. Metodologia

2.1 Ferramentas e Bibliotecas

- Python: Linguagem de programação utilizada para manipulação dos dados e aplicação do modelo.
- TextBlob: Biblioteca simples para análise de sentimentos baseada em dicionários de palavras com polaridade.
- Google Colab: Ambiente de desenvolvimento online, onde o código foi executado.

2.2 Por que não usamos spaCy?

Originalmente, tentou-se usar a biblioteca spaCy combinada com o pacote spacytextblob para análise de sentimento. Contudo, a execução gerou o erro:

AttributeError: [E046] Can't retrieve unregistered extension attribute 'polarity'. Did you forget to call the `set_extension` method?

Este erro ocorre porque a extensão 'polarity' não foi registrada corretamente no pipeline do spaCy. Isso pode acontecer devido a:

- Versões incompatíveis entre spaCy e spacytextblob.
- Mudanças recentes na API do spaCy que não são refletidas no spacytextblob.
- Problemas na instalação ou carregamento da pipeline.

Dada a complexidade de corrigir isso, optamos por usar a biblioteca TextBlob diretamente, que é mais simples e adequada para iniciantes.

3. Passo a passo do projeto usando TextBlob

Passo 1: Instalação das bibliotecas

No Google Colab, instalamos o TextBlob e baixamos recursos necessários:

```
!pip install textblob
!python -m textblob.download_corpora
```

Passo 2: Importação das bibliotecas

```
from textblob import TextBlob
import re
```

Passo 3: Preparação dos dados (limpeza dos tweets)

Os tweets originais contêm muitas menções (@usuario), hashtags (#tag), e URLs, que não ajudam na análise de sentimento. Criamos uma função para limpar esses elementos:

```
def limpar_tweet(tweet):
    tweet = re.sub(r'http\S+', '', tweet)    # Remove URLs
    tweet = re.sub(r'@\S+', '', tweet)       # Remove menções
    tweet = re.sub(r'#\S+', '', tweet)       # Remove hashtags
    tweet = tweet.strip()
    return tweet
```

Passo 4: Análise de sentimento

Iteramos sobre a lista de tweets, aplicando a limpeza e a análise com TextBlob, que retorna um valor de polaridade entre -1 (muito negativo) e 1 (muito positivo).

Classificamos os tweets como:

- Positivo: polaridade > 0.1
- Neutro: -0.1 <= polaridade <= 0.1
- Negativo: polaridade < -0.1

for tweet in tweets:

```
    texto_limpo = limpar_tweet(tweet)
    analise = TextBlob(texto_limpo)
    polaridade = analise.sentiment.polarity
    if polaridade > 0.1:
        sentimento = 'Positivo'
    elif polaridade < -0.1:
        sentimento = 'Negativo'
    else:
        sentimento = 'Neutro'
    print(f'Tweet: {texto_limpo}\nPolaridade: {polaridade:.2f} => {sentimento}\n')
```

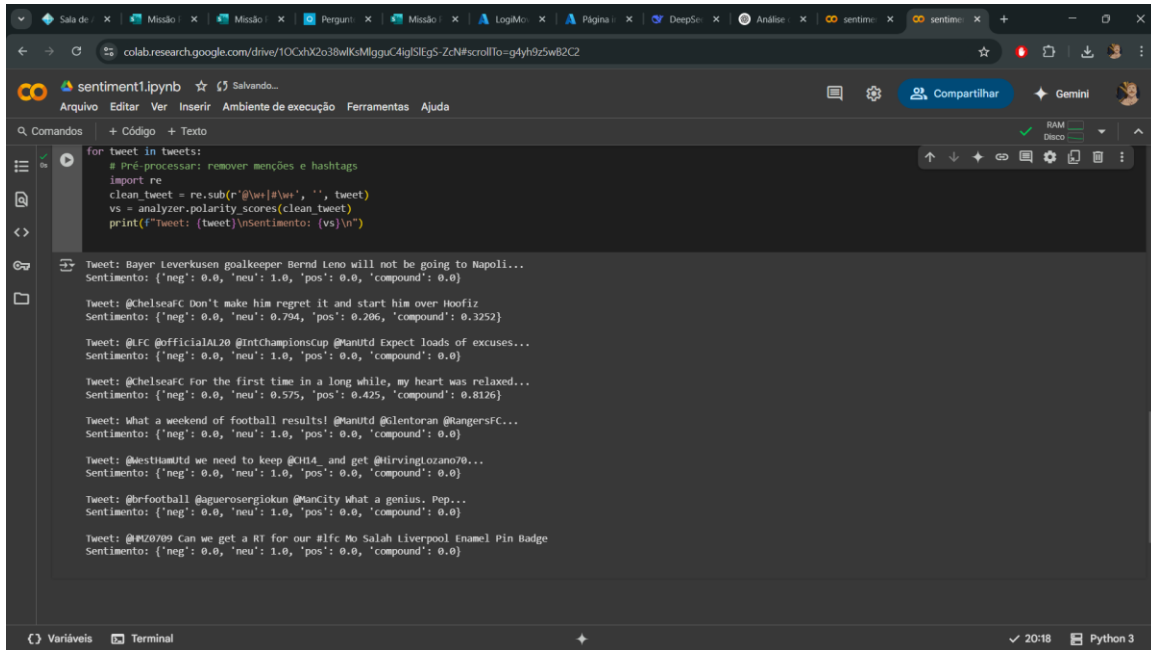
4. Resultados

Com o processo acima, obtemos uma análise mais precisa dos sentimentos expressos nos tweets, reduzindo o impacto de ruídos (menções, hashtags, URLs) que antes causavam muitas classificações neutras.

5. Considerações finais

- Limpeza do texto é fundamental para melhorar a análise de sentimentos.
- TextBlob é simples e eficiente para tarefas básicas.
- Para projetos mais avançados, recomenda-se explorar modelos de Machine Learning com bibliotecas como Hugging Face Transformers, que capturam contexto melhor.

Execução no Colab:



The screenshot shows a Google Colab notebook interface. The browser tabs at the top include 'Sala de ...', 'Missão ...', 'Pergunt...', 'Missão ...', 'LogMo...', 'Página ...', 'DeepSe...', 'Analise ...', 'sentime...', and 'sentime...'. The address bar shows the URL: `colab.research.google.com/drive/1OCxH2o38wKsMlgguC4gISIEgS-ZcN#scrollTo=g4yh9z5w82C2`. The notebook title is 'sentiment1.ipynb' and it is in 'Salvando...' state. The menu bar includes 'Arquivo', 'Editar', 'Ver', 'Inserir', 'Ambiente de execução', 'Ferramentas', and 'Ajuda'. The left sidebar has icons for file explorer, search, and other functions. The main area shows a code cell with the following Python code:

```
for tweet in tweets:
    # Pré-processar: remover menções e hashtags
    import re
    clean_tweet = re.sub(r'@[a-zA-Z0-9_]+', '', tweet)
    vs = analyzer.polarity_scores(clean_tweet)
    print(f'Tweet: {tweet}\nSentimento: {vs}\n')
```

The output of the code cell shows the sentiment analysis results for several tweets:

```
Tweet: Bayer Leverkusen goalkeeper Bernd Leno will not be going to Napoli...
Sentimento: {'neg': 0.0, 'neu': 1.0, 'pos': 0.0, 'compound': 0.0}

Tweet: @chelseafc Don't make him regret it and start him over Hoofir
Sentimento: {'neg': 0.0, 'neu': 0.794, 'pos': 0.206, 'compound': 0.3252}

Tweet: @LFC @officialAL20 @IntChampionsCup @ManUtd Expect loads of excuses...
Sentimento: {'neg': 0.0, 'neu': 1.0, 'pos': 0.0, 'compound': 0.0}

Tweet: @chelseafc For the first time in a long while, my heart was relaxed...
Sentimento: {'neg': 0.0, 'neu': 0.575, 'pos': 0.425, 'compound': 0.8126}

Tweet: What a weekend of football results! @ManUtd @Glentoran @RangersFC...
Sentimento: {'neg': 0.0, 'neu': 1.0, 'pos': 0.0, 'compound': 0.0}

Tweet: @WestHamUtd we need to keep @CH4 and get @HirvingLozano70...
Sentimento: {'neg': 0.0, 'neu': 1.0, 'pos': 0.0, 'compound': 0.0}

Tweet: @brFootball @aguerosergiokun @ManCity What a genius. Pep...
Sentimento: {'neg': 0.0, 'neu': 1.0, 'pos': 0.0, 'compound': 0.0}

Tweet: @MU0709 Can we get a RT for our #Ifc Mo Salah Liverpool Enamel Pin Badge
Sentimento: {'neg': 0.0, 'neu': 1.0, 'pos': 0.0, 'compound': 0.0}
```

The bottom status bar shows 'Variáveis', 'Terminal', a plus sign, '20:18', and 'Python 3'.