O BNDES (Banco Nacional de Desenvolvimento Econômico e Social) possui indicadores estratégicos para medir o sentimento dos artigos e postagens publicadas nas redes sociais (Facebook, Twitter, Instagram etc.) ou na imprensa. Atualmente existe uma empresa contratada para realizar o clipping e medição do sentimento (positivo, neutro e negativo).

A proposta do trabalho é a utilizar os textos obtidos pelas ferramentas de clipping e suas respectivas análises de sentimento para treinar uma rede neural capaz de avaliar o sentimento de textos postados sobre o BNDES na internet.

O BNDES como instituição pública responsável por investimentos em setores relacionados ao desenvolvimento econômico e social do país tem alta exposição na imprensa e nas mídias sociais. A polarização dos sentimentos nas publicações é considerável:

* As reportagens sobre a atuação do banco são em sua maioria neutras porque relatam os objetivos de programas e linhas de crédito criadas.
* Os responsáveis pelo investimento e beneficiados diretos ou indiretos pelas linhas de financiamento normalmente tem opiniões positivas.
* Grupos que não foram beneficiados e a oposição geralmente tem opiniões negativas ou entendem que o investimento deveria priorizar outro tipo de atividade.
* Existem os grupos que criticam a existência do banco, a remuneração dos funcionários e condenam qualquer tipo de ação realizada pelo banco, gerando mensagens negativas que buscam denegrir a imagem do banco.

Priorizada pela alta hierarquia do BNDES, a melhora da sua imagem externa é um indicador estratégico, sendo a análise de sentimentos sobre artigos e posts publicados uma das variáveis utilizadas nessa medição.

Atualmente o BNDES possui contrato com uma empresa que realiza o clipping e análise de sentimentos. Nosso projeto utilizou os textos extraídos e os rótulos atribuídos (positivo, neutro, negativo) para treinar os modelos.

**Carregamento dos Textos**

Textos da imprensa não estavam disponíveis diretamente nas planilhas que foram encaminhadas, para consultar o conteúdo foi necessário navegar por links informados nas planilhas que continham dados de 2020 e 2021. Um programa foi construído com essa finalidade, gerando um arquivo CSV (separador |) com o texto do artigo e o respectivo sentimento.

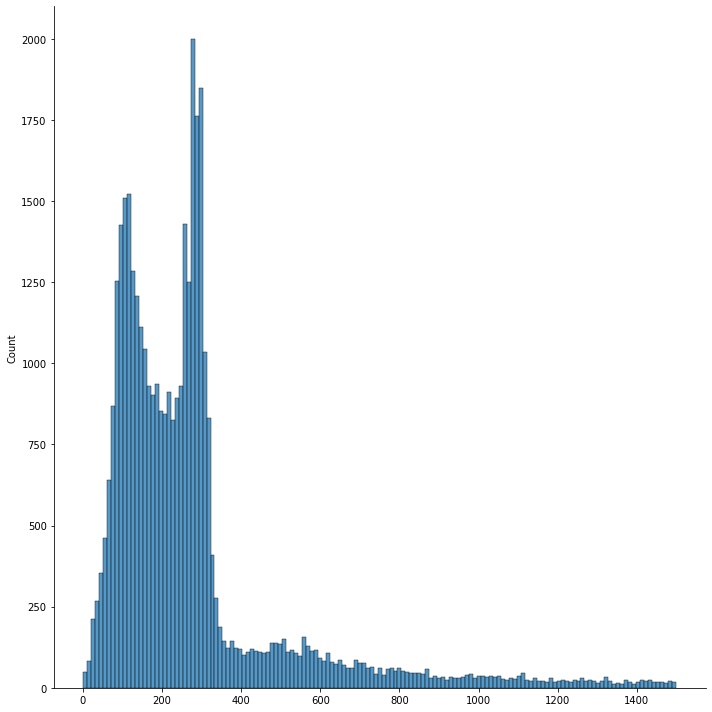
Os textos das mídias sociais estavam disponíveis em planilhas, sem necessidade de consultar links, no entanto continham colunas desnecessárias para nossa análise, linhas vazias ou repetidas (geralmente repostagens de tweets). Criamos outro programa para extrair o conteúdo gerando um arquivo CSV pronto para ser processado. Importante ressaltar que a manipulação inicial para criação dos *datasets* foi executada somente uma vez.

O carregamento extraiu 45.440 textos utilizando dados de 2020 (imprensa e mídias) e 2021 (imprensa e mídias), e maio de 2022 (somente mídias sociais).

**Tamanho**

Os artigos de imprensa são bem grandes, a maior parte continha aproximadamente 1.000 tokens, contudo existiam artigos com mais de 14.000 tokens. Já as publicações nas redes sociais possuem, normalmente, tamanho bem menor, variando entre 100 e 350 tokens.

No gráfico abaixo limitamos o número máximo de tokens para facilitar a visualização.

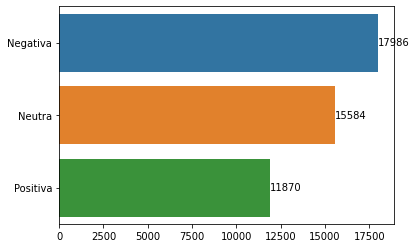


Definimos 1.000 tokens como o tamanho máximo que o modelo poderia receber para análise, caso o texto fosse maior então ele seria truncado*.*

**Balanceamento**

A base não estava completamente balanceada, mas a quantidade de textos adquiridos permitiu um treinamento adequado, no gráfico abaixo pode-se visualizar as proporções entre os sentimentos.

Importante ressaltar que artigos da imprensa são na sua maioria neutra, contudo elas compõem aproximadamente 15% do conjunto de textos.



**Treinamento**

As bases de treino, validação e teste foram separadas conforme o gráfico abaixo:

Gráfico, Gráfico de barras

Descrição gerada automaticamente

**Modelos**

Foram utilizados dois tipos de modelos:

* LSTM – *Long Short Term Memory*
* BERT - *Bidirectional Encoder Representations from Transformer*

**Modelo 1 - LSTM**

Para auxiliar a contextualização dos texto a rede neural utilizou uma camada de *Embedding* pré-treinado, disponibilizados em português pelo NILC - Núcleo Interinstitucional de Linguística Computacional. Diferentes dimensões e tipos de *embedding* foram utilizados e apresentaremos um comparativo entre os resultados.

As tarefas de pré-processamento aplicadas no texto foram as seguintes:

* Remoção de *stop words*
* Remoção de palavras que não existem no *Embedding*
* Transformação em minúscula.
* Correção de erros comuns de escrita com alta incidência, exemplo: negociacao, inflacao, etc.
* Lemetização e Stemming não foram aplicados.

A tabela abaixo contém o comparativo dos modelos testados, importante ressaltar os seguintes pontos:

* Nossos testes iniciais foram realizados com a base de imprensa porque os textos de mídias ainda não haviam sido disponibilizados.
* A métrica utilizada para comparação foi a acurácia, sendo que a nota final do modelo deveria ser baseada na base de teste (nunca antes analisada).
* As LSTM’s de dois níveis, bem como as LSTM’s bidirecionais, em nossos testes, não conseguiram superar o modelo de uma camada de LSTM.

Nosso melhor resultado com os textos de imprensa foi 83,24%, já com imprensa e mídias sociais conseguimos 83,67%, resultado bastante satisfatório. Segue a descrição do modelo com o melhor resultado:

1. Input de textos com tamanho de 1.000 posições
2. Embedding Word2Vec (300)
3. LSTM (300, RecurrentDroput=0,3)
4. Dropout (0.3)
5. GlobalMaxPoolling1D
6. Dense(300, ReLu)
7. Dense (3)

O treinamento para esse modelo demora em média 16 horas.

Observações sobre os modelos testados:

devido ao grande volume de textos e o tamanho máximo de cada um (1.000). Em um teste diminuindo o tamanho máximo tivemos perdas significativas de performance, alcançando somente 71,46 %.

Interessante ressaltar também que com *embeddings* com dimensão superior a 300 pioraram o resultado do modelo, com 1.000 posições tivemos o resultado de 78,8% e com 600 81,52%

**Resultados**

**Conclusões**

O resultado obtido pela rede construído é satisfatório, mas o tempo de treinamento é muito longo.

Não realizamos testes com *embeddings* próprios, mas dada a quantidade de textos poderíamos construir um *embedding* para comparar os resultados, contudo o tempo não foi suficiente para realização desses testes.

Também poderíamos realizar testes com modelos BERT, contudo para classificação de sentimentos não encontramos na literatura exemplos de resultados