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 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, gerando mensagens negativas que buscam denegrir a imagem do BNDES.

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

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

**Carregamento dos Textos**

A equipe de comunicação do BNDES encaminhou planilhas com artigos da imprensa, mas estes não continham o conteúdo necessário para o processamento, somente os links. Para consultar o conteúdo foi necessário navegar pelas referências, sendo criado um programa com essa finalidade.

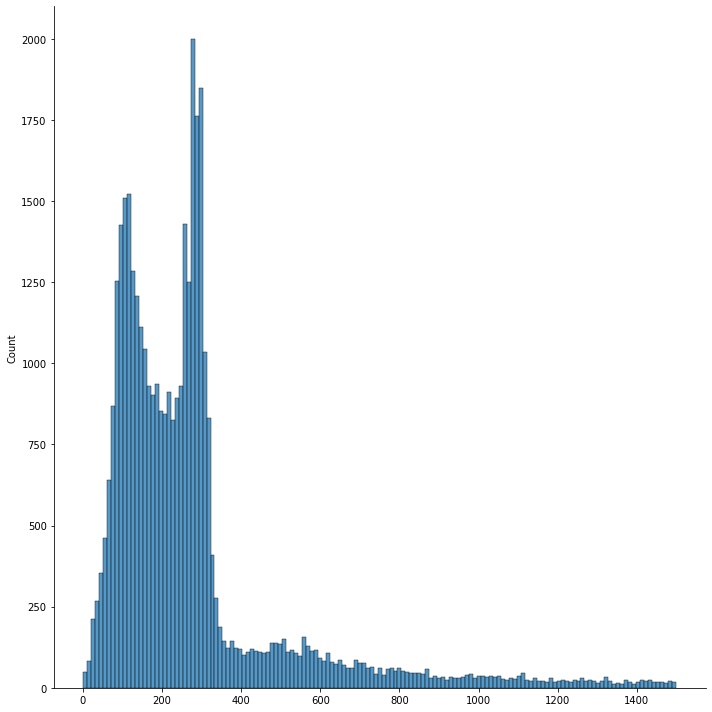
Os textos das mídias sociais estavam disponíveis em planilhas, sem necessidade de consultar links, no entanto continham colunas desnecessárias para nossa análise, linhas vazias ou repetidas (geralmente tweets repostados).

O carregamento extraiu 45.440 textos utilizando dados de 2020 (imprensa e mídias), 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.

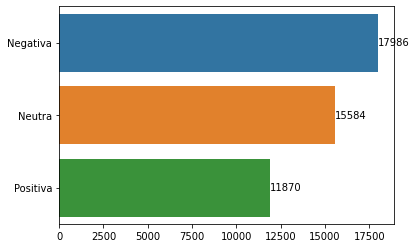
No gráfico abaixo limitamos o número máximo de tokens para facilitar a visualização.



**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 na contextualização dos textos 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 testados e apresentaremos posteriormente um comparativo entre os resultados.

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

* Transformação em minúscula
* Remoção de *stop words*
* Remoção de palavras que não existem no *Embedding*
* Correção de erros comuns de escrita com alta incidência, exemplo: negociacao, inflacao, etc.
* Limitar o tamanho do texto a 1.000 tokens, truncando aqueles com tamanho superior ao limite estipulado.
* 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, que não foi vista ainda pelo modelo.
* As LSTM’s de dois níveis, bem como as LSTM’s bidirecionais, em nossos testes, não conseguiram superar o modelo com uma camada de LSTM.

Uma imagem contendo Texto

Descrição gerada automaticamente

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)

**Observações sobre os modelos testados:**

* No teste que diminuiu o tamanho máximo tivemos perdas significativas de performance, alcançando somente 71,46 %.
* *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%
* O resultado obtido pela rede construído é satisfatório, mas o tempo de treinamento é muito longo, em média 16 horas.

**Modelo 2 - BERT**

Utilizamos o modelo BERT da Neuralmind (BERTimbau) e treinamos a última camada para que o modelo aprendesse a classificação de textos com sentimento positivo, negativo e neutro.

Nosso primeiro teste não tratou os textos e conseguimos o resultado de 86% de acurácia. Posteriormente achamos que o resultado melhoraria com outros tratamentos de texto, mas foi inferior.

Abaixo mostramos a tabela comparando alguns testes que realizamos com o BERT.

Interface gráfica do usuário, Texto, Aplicativo, Email

Descrição gerada automaticamente

Para os modelos BERT também medimos o F1-score.

Os modelos BERT eram treinados em 2 horas, 8 vezes menor que o tempo de treinamento das redes LSTM.

**Conclusões**

Não realizamos testes com *embeddings* próprios, mas dada a quantidade de textos poderíamos construir um *embedding* para comparar os resultados, contudo não acreditamos que mesmo com um *Embedding* próprio as redes LSTM’s consigam superar o BERT, contudo isso não foi testado no trabalho.

A construção de um BERT próprio, dada a enorme quantidade de textos (o trabalho utilizou somente 2020 e 2021), seria um teste interessante porque acredito que isso poderia melhorar consideravelmente o resultado. O BNDES por exemplo não é um token reconhecido pelo *Tokenizador*, sendo gerado dois tokens [B, #NDES].