

# Classificação de Sentimentos no Dataset de Sentenças Financeiras de Notícias

## I. DATASET

O dataset utilizado [1] neste estudo foi desenvolvido no artigo desta referência: [2] Nesse artigo, os autores propõem o modelo Linearized Phrase Structure (LPS) para análise de sentimentos em textos financeiros.

O dataset contém aproximadamente 5.000 frases extraídas de notícias financeiras e comunicados de empresas, todas classificadas como positivas, negativas ou neutras por 16 anotadores com formação adequada em negócios. Esse conjunto de dados se mostra relevante ao fornecer uma alternativa para medir o sentimento do mercado sem depender de pesquisas mensais sobre desenvolvimentos econômicos.

## II. CLASSIFICATION PIPELINE

Para melhorar a performance da classificação, foi criada uma lista de stopwords manualmente, removendo palavras que apareciam frequentemente no dataset, mas que não agregavam valor ao problema de classificação. Essas palavras incluíam termos como “the”, “to”, “in”, “of”, e “eur”, entre outras. A remoção dessas palavras foi necessária para evitar que influenciassem negativamente o modelo, pois não carregam informações relevantes para determinar o sentimento das sentenças financeiras.

Optou-se pelo uso do modelo Multinomial Naive Bayes (MultinomialNB) [3] e para a vetorização do texto foi utilizado o CountVectorizer. Esse modelo é ideal para problemas de classificação com várias classes, como o caso presente, onde as frases são classificadas como positivas, negativas ou neutras.

A metodologia de testagem consistiu na divisão do dataset em conjuntos de treinamento e teste, alocando 20 por cento dos dados para o conjunto de teste. Destaca-se que essa divisão variou em cada iteração, permitindo que o modelo fosse avaliado em diferentes subconjuntos de dados. Para garantir uma avaliação do modelo, esse procedimento de divisão e treinamento foi repetido 100 vezes, registrando as accuracias em cada execução.

## III. EVALUATION

A avaliação do modelo revelou uma acurácia média de 69 por cento, com um erro padrão de 0,0013 e um valor de  $p$  de 0,017. Com um nível de significância ( $\alpha$ ) de 0,05, esses resultados indicam que há mais de 95 por cento de confiança de que a média das acurácias obtidas não é diferente de 0,69 por cento.

Para analisar os resultados de forma mais detalhada, foi gerada uma matriz de confusão. [4] Os dados mostraram que o modelo apresenta um desempenho significativamente

melhor na identificação de sentimentos neutros e positivos em comparação aos sentimentos negativos.

Além disso, foi realizada uma análise das palavras com maior probabilidade de ocorrência em cada classificação. Observou-se que palavras como “profit”, “down”, e “sales” foram frequentemente associadas à classificação negativa, enquanto termos como “with”, “sales”, e “year” estavam ligados à classificação positiva. Essa sobreposição sugere que o modelo pode estar classificando coincidências, uma vez que várias palavras são comuns em ambas as categorias.

Para aprofundar essa análise, foi adotada uma metodologia em que se calculou a diferença entre a probabilidade de uma palavra estar na classe positiva e a probabilidade de estar na classe negativa, utilizando o módulo dessa diferença. Esse procedimento permitiu identificar as palavras que apresentaram os maiores valores, indicando aquelas que têm maior impacto nas previsões do modelo. Analisando as palavras com esse módulo elevado, foram identificadas palavras com significados claros, como “down”, “up”, “rose”, “decreased”, e “new”. Essa relevância semântica reflete que o modelo está capturando nuances significativas no texto. O primeiro caso de sobreposição, com a palavra “sales”, pode ser explicado pela alta frequência dessa palavra na base de dados utilizada.

## DATASET SIZE

Para avaliar a adequação do tamanho do dataset, foram plotadas as curvas de aprendizado utilizando a classificação de erro em vez da acurácia. Os resultados indicaram que os erros do conjunto de treinamento estão muito próximos de 0, enquanto os erros do conjunto de teste são significativamente maiores. Essa discrepância sugere que, se tivéssemos acesso a mais dados, seria possível melhorar a acurácia do modelo e que estamos overfitting o modelo.

## TOPIC ANALYSIS

A análise de tópicos não se mostrou efetiva para a melhoria do modelo, uma vez que os resultados das acurácias revelaram uma diferença mínima, mantendo uma taxa de erro quase constante ao longo do dataset para cada tópico analisado. Essa homogeneidade nos erros sugere que a segmentação dos dados em tópicos não trouxe informações adicionais significativas para a classificação de sentimentos. Além disso, a análise realizada não conseguiu determinar de forma clara o que cada tópico representava, dificultando a interpretação dos resultados. Essa limitação implica que a abordagem de tópicos pode não ser a mais adequada para este tipo de tarefa,

## REFERENCES

- [1] **Sbhatti, S.** Dataset: *Financial Sentiment Analysis*. Disponível em: <https://www.kaggle.com/datasets/sbhatti/financial-sentiment-analysis>. Acesso em: 3 out. 2024.
- [2] **Malo, P., Sinha, A., Takala, P., Korhonen, P.** (2013). Good Debt or Bad Debt: Detecting Semantic Orientations in Economic Texts. Disponível em: <https://arxiv.org/pdf/1307.5336>. Acesso em: 3 out. 2024.
- [3] **Scikit-learn.** Multinomial Naive Bayes. Disponível em: [https://scikit-learn.org/stable/modules/naive\\_bayes.html#multinomial-naive-bayes](https://scikit-learn.org/stable/modules/naive_bayes.html#multinomial-naive-bayes). Acesso em: 3 out. 2024.
- [4] **Wikipedia.** Matriz de Confusão. Disponível em: [https://pt.wikipedia.org/wiki/Matriz\\_de\\_confus%C3%A3o](https://pt.wikipedia.org/wiki/Matriz_de_confus%C3%A3o). Acesso em: 3 out. 2024.