Exploring financial sentiment analysis with the Financial Phrasebank dataset

Hugo Veríssimo
Complements of Machine Learning 24/25
University of Aveiro
Aveiro, Portugal
hugoverissimo@ua.pt

João Cardoso
Complements of Machine Learning 24/25
University of Aveiro
Aveiro, Portugal
joaopcardoso@ua.pt

***Abstract*—abstratotootottot**

Keywords: key, word, number, 1

I. Introduction

With the ever increasing volume of information created and distributed by the minute, it is more important than ever to have access to fast and reliable analysis of any information we may come across. Especially with the democratized access to financial instruments and capital markets, where individuals have the possibility to invest in virtually any company on the stock exchange, it is important to have ways to leverage against giant institutions with hundreds of financial analysts at their disposal.

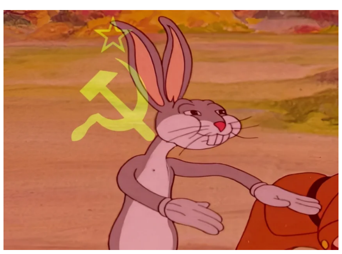

Fig. 1: Power to the people, colorized (circa 1917).

Historically, financial analysis relied heavily on fundamental analysis (examining earnings, balance sheets, annual financial reports), which required extensive knowledge in the field (also the strategy that made Warren Buffett one of the richest men in the world), along with technical analysis (studying price and volume trends). Around 2010, after the 2008 global financial crisis, there was a surge in news analysis to evaluate the tone and derive investment strategies from it. Due to the lack of domain specific lexicon these analysis were falible, until the work by Loughran and McDonald was published, a financial lexicon based on 10-K forms (i.e., annual financial reports) and dictionaries. This allowed to use more sofisticated analysis rather than using the presence of negative words as a signal to sell.

Upon the launch of Twitter, information streams increased dramatically, making more and more data available for analysis. But, machine learning was not heavily used, as most data was not annotated, or there was very little data with high-quality annotations. In 2014, P. Malo *et al.* published a fundamental dataset for financial sentimental analysis, that is still used, the Financial Phrasebank. It is unique, for the inclusion of important aspects as directional expressions (e.g., profits decreased), entity polarity shifts (e.g. profits may be negative if decreased), and phrase level context.

With this, machine learning models started finding their place in research, as the field of natural language processing grew and niche fields such as financial investments found useful data. In this work we explore the Financial Phrasebank dataset, by implementing different machine learning and deep learning models.

II. State of The Art

III. Methodology

A. Dataset & EDA

The Financial PhraseBank is a widely used benchmark dataset for financial sentiment analysis. It consists of roughly 4,840 English sentences (mostly news headlines or short statements) about companies, drawn from financial news articles and press releases. Each sentence is labeled with one of three sentiment classes – positive, negative, or neutral – representing the sentence's sentiment from the perspective of an investor.

+ 50 66 75 percento

ns q concordancia bla bla

https://huggingface.co/datasets/takala/financial_phrasebank [1], [2]

B. Preprocessing

split teste e treino

IV. Models

abcbdhdhhd

A. model 1

model 1

B. model 2

model 2

C. model 3

model 3

V. Exploring BERT

A. base model

TABLE I: Hyperparameter space for ......

Hyperparameter	Possible Values
Epochs	$\{1, 2, 3, 4, 5\}$
Learning rate	$[10^{-5}, 10^{-2}]$
Weight decay	$[0, 0.5]$

best hyperameters:
- Num train epochs: 2.0
- Learning rate: 0.0001
- Weight decay: 0.1

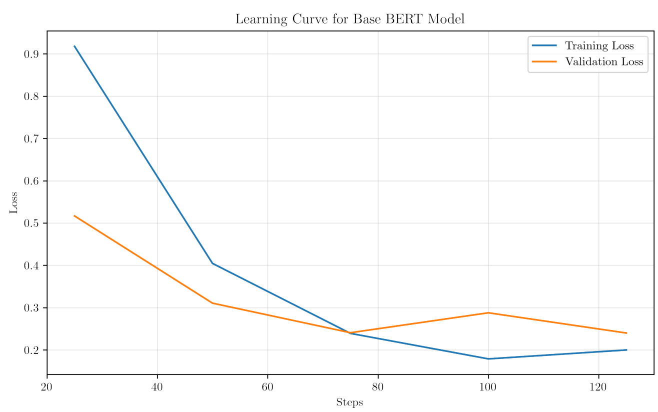

Fig. 2: learning curve

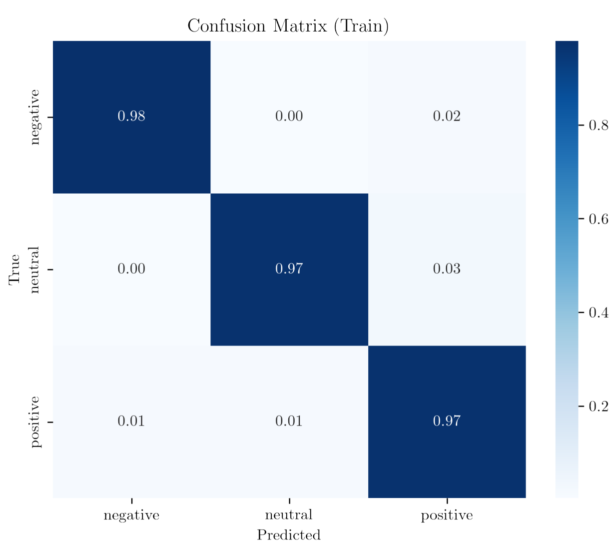

Fig. 3: base bert conf matri train

TABLE II: Classification report for BASE BERT on training data.

Class	Precision	Recall	F1-Score	Support
Negative	0.98	0.98	0.98	336
Neutral	0.98	0.97	0.98	336
Positive	0.96	0.97	0.96	336
Accuracy			0.97	1008
Macro avg	0.97	0.97	0.97	1008
Weighted avg	0.97	0.97	0.97	1008

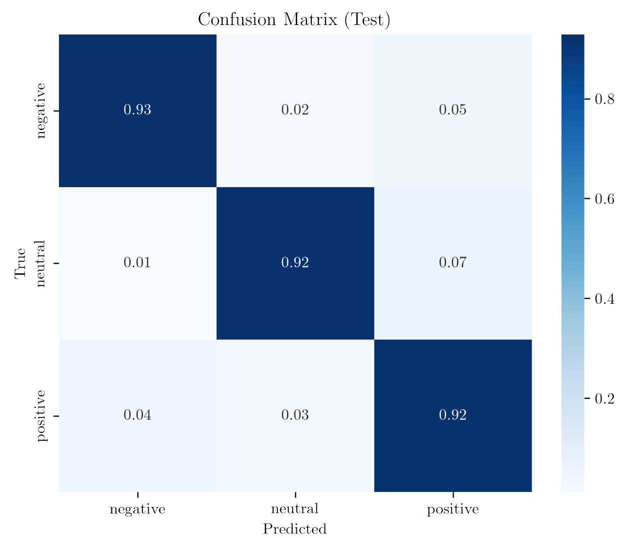

Fig. 4: base bert cong matrix test

TABLE III: Classification report for BASE BERT on test data.

Class	Precision	Recall	F1-Score	Support
Negative	0.86	0.93	0.89	84
Neutral	0.98	0.92	0.95	429
Positive	0.84	0.92	0.88	178
Accuracy			0.92	691
Macro avg	0.89	0.92	0.91	691
Weighted avg	0.93	0.92	0.92	691

VI. Discussion

results discuttion
eemplo de comparacao

TABLE IV: Error metric (MAE) for the fine-tuned models, along with the best performers in the competition.

Model	MAE (Test Set)
Hybrid	0.8434
Obj. Det.	1.2645
Inst. Seg.	1.3415
Team Lacuna (1st)	0.3299
K_Junior (2nd)	0.5698

VII. Conclusion

conclisao

Fig. 5: Enter Caption

Work Load

Both authors contributed equally to the project.

References

[1] P. Malo, A. Sinha, P. Takala, P. Korhonen, and J. Wallenius, "Financialphrasebank-v1.0," 07 2013.

[2] ——, "Good debt or bad debt: Detecting semantic orientations in economic texts," 2013. [Online]. Available: https://arxiv.org/abs/1307.5336