

Problem Statement

In the fast-paced forex market, traders often rely on headlines for quick decisions without reading full articles. However, some news agencies may exaggerate headlines to capture attention, leading to sentiment misalignment with the article's content. Headlines are short in nature and may oversimplify or omit nuances present in the full article, distorting sentiment. This discrepancy can contribute to market inefficiencies, increased volatility, and misguided trading decisions.

This project investigates the discrepancy between news **headline sentiment** and **full-text sentiment**. The study uses a manually labeled dataset of headline sentiment and evaluates its alignment against full-text sentiment, which is predicted using a fine-tuned FinBERT model. By treating full-text sentiment as a reference point, this analysis aims to assess the accuracy and reliability of forex news headlines in reflecting broader article sentiment.

Data Sourcing & Preparation

As seen in Fig 1, the dataset contains news headlines ("title") and full text content("text") relevant to key forex pairs: AUD-USD, EUR-CHF, EUR-USD, GBP-USD, and USD-JPY. The data was extracted from renowned platforms such as Forex Live and FXstreet over a period of 86 days, from January to May 2023.

"true_sentiment" is manually labeled based on how traders would likely react to a headline, it took consideration of how the headlines impacted currency movements at the time of publication rather than just its wording. It reflects real-world trader sentiment upon reading the news headline. As it is in text, I encoded it to numerical values for modelling.

	published_at	ticker	true_sentiment	title	author	url	source	text
0	2023-01-12 07:47:00	EURCHF	Positive	Euro to benefit from the ECBs pronounced hawkish...	FXStreet Insights Team	https://www.fxstreet.com/news/euro-to-benefit-...	FX Street	The Euro was able to appreciate particularly s...
1	2023-01-12 10:34:00	EURCHF	Positive	EURCHF Trend higher may remain in place – ING	FXStreet Insights Team	https://www.fxstreet.com/news/eur-chf-trend-hi...	FX Street	EUR/CHF yesterday broke above 1.00. Economists...
2	2023-01-12 11:40:00	EURCHF	Neutral	Does a jump in EURCHF point to a break above 1...	FXStreet Insights Team	https://www.fxstreet.com/news/does-a-jump-in-e...	FX Street	EUR/CHF vaults parity for the first time since...

Fig 1. Snapshot of dataset

(Source:<https://paperswithcode.com/dataset/forex-news-annotated-dataset-for-sentiment>)

Model Choice & Fine-tuning

In this analysis, I fine-tuned a **FinBERT** model on **news headlines** to predict their **true sentiment**. I then used this fine-tuned model to classify sentiment on **full-text** articles with

adjustment to text length. The results of full-text sentiment are then compared against headline sentiment in a confusion matrix to assess the reliability in headlines.

Why BERT (FinBERT)

BERT, a transformer-based model, outperforms LSTMs and CNNs in NLP by leveraging bidirectional context to analyze both past and future words. Fine-tuning a pre-trained BERT model also requires less data and computing power, making it ideal for the dataset with only 2,291 rows.

FinBERT, specifically pre-trained on financial texts, excels at finance-specific sentiment analysis, avoiding misinterpretations common in general NLP models. For example, it can correctly identify "bearish" as negative in finance, whereas a general model might classify it as neutral.

Fine-tuning FinBERT

While FinBERT is strong in financial sentiment analysis, it may not fully capture Forex-specific nuances. To improve alignment with market dynamics, it is fine-tuned on headlines against true sentiment labels.

Fine-tuning leverages the **AdamW optimizer** and **Cross-Entropy Loss**. AdamW efficiently optimizes large transformers by dynamically adjusting learning rates and preventing overfitting through weight decay. Cross-Entropy Loss, ideal for multi-class classification (Negative, Neutral, Positive), minimizes errors by comparing predicted logits with actual sentiment labels.

The training loop starts with `model.train()`, updating all layers. Each epoch processes mini-batches of tokenized headlines, generates logits for sentiment classes, and computes Cross-Entropy Loss. Gradients are back propagated (`loss.backward()`), AdamW updates weights, and `optimizer.zero_grad()` clears previous gradients to ensure stable learning.

Applying Headline-trained Model on Full-text

After tuning the FinBERT model from headlines, I used it to classify sentiment for full-text. Since headline and the full text of a news article often share similar vocabulary, the model is transferable.

However, longer texts dilute sentiment, as neutral or mixed content reduces intensity. To mitigate this, I used **length-normalized sentiment adjustment** (Amplayo et al., 2019), which segments long text into headline-sized chunks and weights sections based on their contribution to overall sentiment, following these steps:

1. Convert full text to random chunks, each contains length L_i and Sentiment Probabilities from FinBERT softmax output: $P_i = [p_i^{(neg)}, p_i^{(neu)}, p_i^{(pos)}]$

2. Compute sentiment score for each chunk: $S_i = -p_i^{(neg)} + p_i^{(pos)}$

$$W_i = \frac{1}{1 + \log(1 + L_i)}$$

3. Compute overall sentiment score based on weighted score of each chunk

Evaluation, Limitation & Conclusion

Performance of Tuning

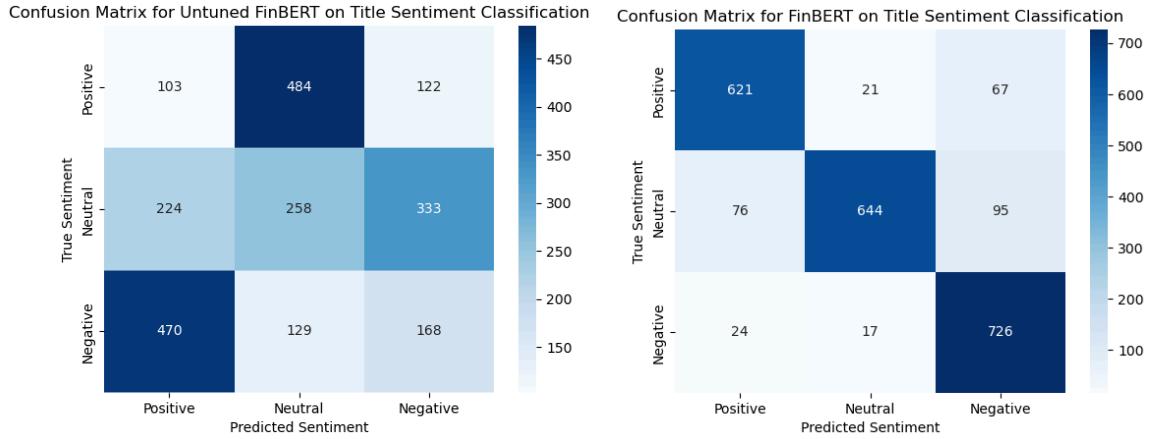


Fig 2. Confusion Matrix of pre-trained (Left) and tuned (Right) models

Model	Accuracy	Precision	Recall	F1-Score	AUC Score
Pre-trained	0.2309	0.2356	0.2309	0.2321	0.4237
Tuned	0.8691	0.8762	0.8691	0.8686	0.9024

Fig 3. Evaluation Matrix comparing pre-tuned and tuned model

Fine-tuning FinBERT was highly effective, drastically improving accuracy, precision, recall, and AUC (Fig 3). It correctly classifies sentiment in 86.91% of cases, compared to just 23.09% for the untuned model. The fine-tuned model does not just predict more correctly overall but also distributes predictions across sentiment classes more accurately (Fig 2).

Given the strong performance of the fine-tuned FinBERT model in classifying Forex news sentiment, it is appropriate to extend its use for full-text sentiment classification for the next step.

Performance of Headline sentiment

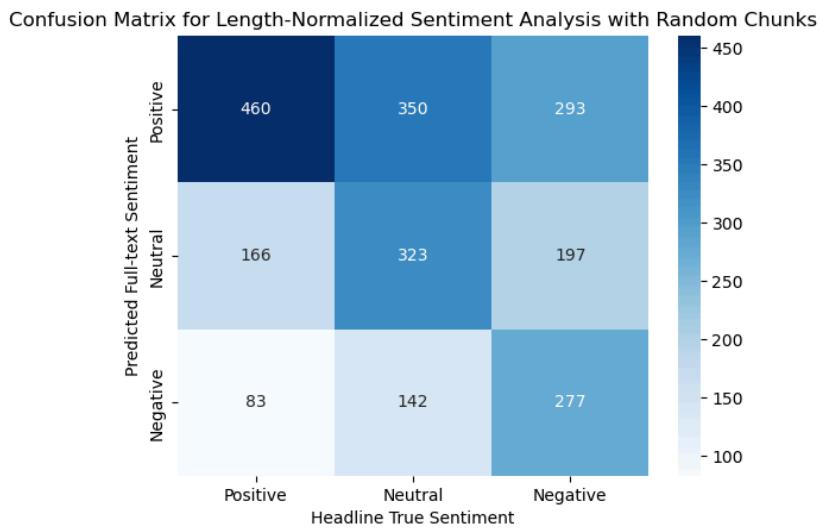


Fig 4. Confusion Matrix of Headline True sentiment against Predicted Full-text sentiment

After comparing headline sentiment against model-classified full-text sentiment, the confusion matrix (Fig 4) showed poor performance of Headline sentiment. An overall low accuracy (0.4627) suggests that headline sentiment often diverges from full-text sentiment, reinforcing the hypothesis that headlines may not always reflect the true market condition conveyed in the article.

Note that the reference point, full-text sentiment, is model-derived, potential limitations and improvements are discussed in the Limitation section.

Sentiment Class	Precision	Recall	F1-score
Positive	0.649	0.417	0.507
Neutral	0.396	0.471	0.430
Negative	0.361	0.552	0.437

Fig.5 Evaluation Matrix by Sentiment Class

High precision (64.9%) but low recall (41.7%) for Positive sentiments, suggests headlines are good at predicting Positive when it exists but fail to capture all instances of full-text positivity. Neutral sentiment is the hardest to classify, with the lowest F1-score (43.0%). Headlines struggle to maintain neutrality, often exaggerating sentiment. Negative sentiment is underreported, with low precision(36.1%) but relatively high recall (55.2%), meaning that Negative sentiment is often softened in headlines.

Different news agencies also have similar accuracy (FX Street: 0.4627, Forex Live: 0.4623), suggesting that headline sentiment misalignment with full-text sentiment is a systemic issue rather than being specific to a single source.

Limitations & Improvements

Unlike traditional evaluations using manually labeled ground truth, this approach assumes full-text sentiment as the true value for assessing headline accuracy. This relies on the assumption that full-text sentiment provides a more accurate reflection of market conditions than headlines. However, this may not always hold true, as full-text articles can also introduce subjective framing, editorial bias, or conflicting viewpoints that dilute or distort sentiment rather than clarify it.

Another key limitation is that full-text sentiment itself is a predicted value rather than a manually labeled ground truth. Since it is derived from a fine-tuned model, it may carry inherent biases or errors, especially if the model misclassified sentiment and the effect propagates into headline sentiment evaluation.

Therefore, instead of relying solely on model-predicted full-text sentiment, incorporating human-labeled sentiment annotations for a subset of the full-text data can validate whether full-text sentiment accurately represents market sentiment. This could serve as a benchmark to assess the model's reliability. Additionally, using multiple sentiment models (e.g., FinBERT + LLM-based classifier + sentiment lexicons) could help cross-validate sentiment labels, reducing the risk of relying on one model's potential biases.

Conclusion & Trading Opportunities

This study highlights the limitations of headlines in accurately reflecting full-text sentiment of a news. Using the tuned FinBERT model, I found that while headlines effectively detect positive sentiment, they often exaggerate its intensity, while negative sentiment is frequently softened, making headlines appear more neutral than the full text suggests.

Headline sentiments are more indicative of immediate market reactions, while full articles provide a more balanced reflection of actual market conditions. Forex traders can exploit this gap, identifying cases where sensational headlines drive short-term movements that later correct as full-text information is absorbed. A sample trading strategy leveraging this sentiment divergence could be short-term momentum trades on exaggerated headlines followed by mean-reversion trades based on full-text sentiment.

Citation

Amplayo, R. K., Lim, S., & Hwang, S. (2019, September 18). *Text length adaptation in sentiment classification*. arXiv.org. <https://arxiv.org/abs/1909.08306>