

PHÂN TÍCH CẢM XÚC TRÊN TIN TỨC TÀI CHÍNH SỬ DỤNG LARGE LANGUAGE MODEL

Phạm Phúc Hậu - 230101042

Tóm tắt

- Lớp: CS2205. APR2024
- Link Github:
<https://github.com/PhamPhucHau/CS2205.APR2024>
- Link YouTube video: <https://youtu.be/IFPoT4Z2Ej4>
- Phạm Phúc Hậu



Giới thiệu

- Phân tích cảm xúc tài chính từ các văn bản như báo cáo tài chính, tin tức kinh tế.
- Mô hình large language model ngày càng phổ biến như BERT, GPT...

Input: A Financial Text.

Example :

1. 'Growth is strong and we have plenty of liquidity.'

2. 'There is a shortage of capital, and we need extra financing.'

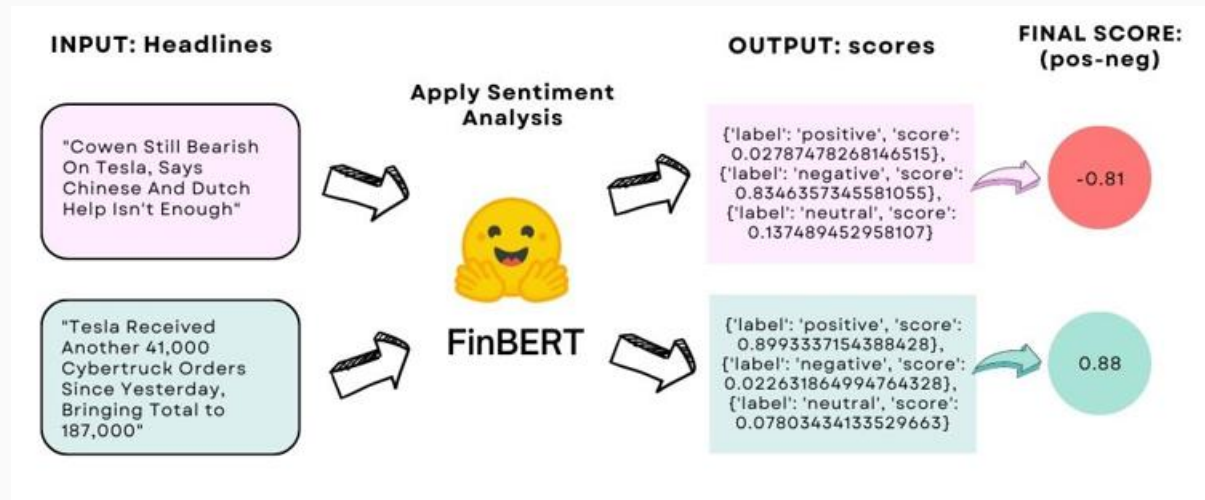
3. 'Formulation patents might protect Vasotec to a limited extent.'



Output: Positive, Neutral or Negative.

Mục tiêu

- Phân tích cảm xúc tài chính từ các văn bản như báo cáo tài chính, tin tức kinh tế.
- Sử dụng mô hình FinBERT, một biến thể của BERT, tối ưu cho lĩnh vực tài chính.



Nội dung và Phương pháp

- Nội dung :
 - Thu thập và chuẩn bị dữ liệu.
 - Tìm hiểu mô hình large language model
 - Chọn và xây dựng mô hình.
 - Tinh chỉnh và đánh giá mô hình.

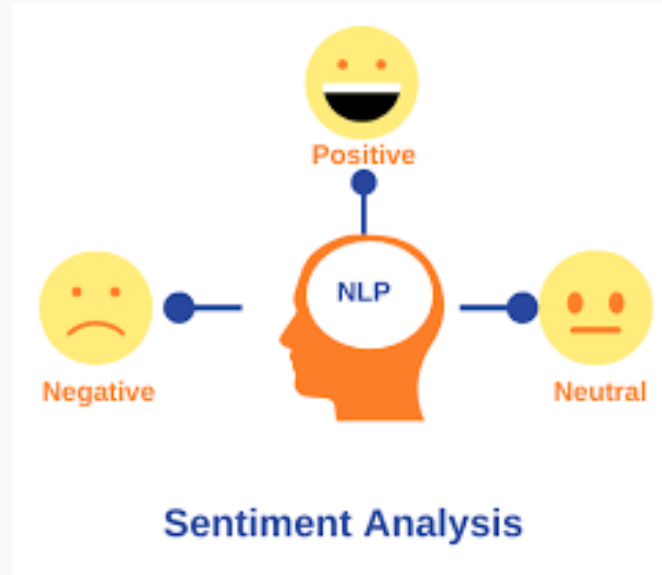


Nội dung và Phương pháp

- Phương pháp :
- Tìm hiểu mô hình BERT và FinBERT: Tham khảo research paper và video.
- Thu thập và chuẩn bị dữ liệu : Sử dụng dataset của Kaggle và StockEmotions Dataset. Sử dụng thư viện Python: để xử lý văn bản.
- Huấn luyện FinBERT : Sử dụng thư viện Hugging Face Transformers và sử dụng Grid Search hoặc Random Search để tìm ra các tham số tối ưu cho mô hình.
- Tinh chỉnh và Đánh giá:
 - Fine-tuning trên tập dữ liệu có gắn nhãn: Sử dụng tập dữ liệu có gắn nhãn cảm xúc để tinh chỉnh FinBERT.
 - Đánh giá mô hình : Cross-validation

Kết quả dự kiến

- Phân tích cảm xúc, quan điểm văn bản chính xác.
- Xây dựng được mô hình hiệu quả hơn các mô hình tự tạo.



Tài liệu tham khảo

- [1] Araci, D. (2019). FinBERT: A Pretrained Language Model for Financial Communications. arXiv preprint arXiv:1908.10063.
- [2] Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers) (pp. 4171-4186).
- [3] Howard, J., & Ruder, S. (2018). Universal Language Model Fine-tuning for Text Classification. In Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers) (pp. 328-339).