Financial Sentiment Across News Sources

PROJECT PROPOSAL

Repository:
https://github.com/Roee104/FinancialSentimentAnalysis.git

Liel Ziv - 212541353
Roee Levi - 207496852
Adir Havakuk - 208636597

Problem Statement

- **Motivation:**

  Financial news shapes market dynamics, yet current models fail to extract the company and sector level sentiment insights required for detailed financial analysis.

- **Problem Definition:**
  - Input: Full financial article (headline + body)
  - Output:
    - Overall sentiment: Positive / Neutral / Negative
    - Per-ticker sentiment + mention weight
    - Sector-level sentiment (aggregated)

- **Challenges**
  - Multiple companies per article
  - Match sentiment to correct ticker
  - Lack of labeled datasets
  - Complex aggregation logic

Training and Test Data

❖ **Data Type:**

- Financial news articles (headline + body).
- Target labels:
  - Overall sentiment: Positive / Neutral / Negative.
  - Ticker-level sentiment: per company mentioned.
  - Sector-level sentiment: aggregated by industry.

❖ **Data Source**:

- EODHD News API (2020–2025) :
  15,000 headlines + body financial articles.
- Synthetic Data:
  GPT-generated headlines to improve class balance.

❖ **NLP Tasks**

- Ticker extraction (NER + regex)
- Sentence splitting
- Chunk-level sentiment (FinBERT)
- Aggregation: ticker-level , sector-level , document-level

Example I/O

Input: full financial article – headline + body text

:Output


```
{
  "overall_sentiment": "Negative",
  "sectors": [
    {
      "name": "Information Technology",
      "sentiment": "Positive",
      "relative_weight": 0.65,
      "confidence": 0.89
    },
    {
      "name": "Financials",
      "sentiment": "Neutral",
      "relative_weight": 0.25,
      "confidence": 0.75
    },
    {
      "name": "Health Care",
      "sentiment": "Negative",
      "relative_weight": 0.10,
      "confidence": 0.80
    }
  ],
  "tickers": [
    {
      "symbol": "AAPL",
      "sector": "Information Technology",
      "sentiment": "Positive",
      "relative_weight": 0.25,
      "confidence": 0.87
    },
    {
      "symbol": "MSFT",
      "sector": "Information Technology",
      "sentiment": "Positive",
      "relative_weight": 0.40,
      "confidence": 0.90
    },
    {
      "symbol": "JPM",
      "sector": "Financials",
      "sentiment": "Neutral",
      "relative_weight": 0.15,
      "confidence": 0.72
    },
    {
      "symbol": "PFE",
      "sector": "Health Care",
      "sentiment": "Negative",
      "relative_weight": 0.10,
      "confidence": 0.80
    }
  ]
}
```

Evaluation

Metrics:

- Ticker-level: Precision, Recall, F1
- Sector-level: F1
- Document-level: Accuracy

Evaluation Strategy:

- 500 manually annotated articles (ticker + sentence level)
- Train/Test split for evaluation
- Compare chunk-based aggregation performance

Baseline:

- Rule-based ticker extraction + VADER
- Baseline F1 ≈ 0.55
- FinBERT pipeline compared against baseline