Week 1 Final Report – Predicting Price Moves with News Sentiment

Author: Haile Weleslassie Gebru

1. Introduction

This report summarizes the accomplishments from Week 1 of the project, which focuses on evaluating whether financial news sentiment can predict short-term stock price movements. The approach integrates NLP-based sentiment analysis with technical stock indicators and historical return analysis. The work completed covers all three initial tasks: Git setup, technical analysis of stock data, and sentiment correlation, as outlined in the Week 1 guidelines.

2. Git & Project Structure

The project is version-controlled and structured using GitHub for collaboration and traceability.

Project structure:

```
├── .github/
├── data/
├── .gitignore
├── requirements.txt
├── README.md
├── src/
│   └── __init__.py
├── notebooks/
│   ├── __init__.py
│   └── README.md
├── reports/
├── tests/
│   └── __init__.py
└── scripts/
    ├── __init__.py
    └── README.md
```

A consistent branching strategy was used to isolate features and track progress across tasks.

3. Technical Analysis of Stock Data

Using the `yFinance` API and supplementary CSVs, historical stock data for the following tickers was collected: `META`, `TSLA`, `AMZN`, `GOOG`, `MSFT`, and `NVDA`.

Three key technical indicators were computed and visualized using `matplotlib`:

- **Moving Average (MA)**
- **Relative Strength Index (RSI)**
- **MACD (Moving Average Convergence Divergence)**

These indicators provided insights into price trends and momentum changes.

4. Sentiment Analysis and Correlation

Sentiment scores were extracted from the `raw_analyst_ratings.csv` dataset using `TextBlob`. Headlines were tokenized and assigned polarity scores ranging from -1 (negative) to 1 (positive). The sentiment scores were aggregated per stock and date.

Daily stock returns were then merged with these sentiment scores. A Pearson correlation analysis was conducted to assess the relationship between sentiment and price movement.

5. Observations & Interpretation

In most cases, the correlation between news sentiment and price changes was weak. However, `NVDA` and `AMZN` showed stronger positive correlations, suggesting they may be more reactive to sentiment or investor expectations.

It is important to note that correlation does not imply causation—many external factors (macroeconomic indicators, earnings reports, etc.) also impact stock prices.

6. Challenges & Solutions

a. Module Import Errors

- **Issue:** `ModuleNotFoundError` for modules like `textblob` and `yfinance`.
- **Solution:** Still unresolved; installation verification needed.

b. Path Handling Issues

- **Issue:** File-saving functions failed due to missing directories.
- **Solution:** Used `os.makedirs()` and `pathlib.Path` to create directories dynamically.

c. Function Import Errors

- **Issue:** Functions weren't importable due to incorrect definitions or file structure.
- **Solution:** Verified function placement, saved files correctly, and restarted Jupyter/Python environment.

d. Sentiment–Return Merge Issues

- **Issue:** Merge failures caused by mismatched date formats or missing stock tickers.
- **Solution:** Standardized dates using `pd.to_datetime(...).dt.date` and ensured ticker info was included during aggregation.

7. Recommendations

- **Adopt Financial-Specific Sentiment Models:** Replace `TextBlob` with domain-specific tools like **FinBERT** or **VADER**, which are better suited for financial text.
- **Incorporate Additional Features:** Add trading volume, volatility, and index-level features to assess if they interact with sentiment in predicting returns.

8. Conclusion

This week's work successfully combined natural language processing with technical stock analysis. The project included:

- Data ingestion and cleaning
- Feature engineering through technical indicators
- Basic correlation between sentiment and stock returns

The foundations have been laid for a robust sentiment-based financial forecasting tool. The integration of data science, finance, and NLP in this project showcases the potential for interdisciplinary solutions in financial analytics.