

Project: Market Mood & Moves – Sentiment-Driven Stock Prediction

Duration Covered: Week 1 & Week 2

1. Introduction

During Weeks 1 and 2 of the Market Mood & Moves project, I gained a comprehensive understanding of how market sentiment can be quantified and integrated into stock prediction systems. The learning combined concepts from behavioral finance, natural language processing, and quantitative finance, while also introducing advanced transformer-based models such as BERT and FinBERT.

2. Behavioral Finance and Market Sentiment

I learned that short-term stock price movements are significantly influenced by investor psychology rather than pure fundamentals. Behavioral biases such as herd behavior, loss aversion, overconfidence, and confirmation bias create inefficiencies that sentiment analysis can exploit.

I understood sentiment as a short-term market signal that captures collective investor emotions reflected in news headlines and reports.

3. Overall System Architecture

I learned the complete workflow of a sentiment-driven trading system:

- News and market data collection
- Text preprocessing and sentiment extraction
- Time alignment with trading sessions
- Financial feature engineering
- Performance evaluation and backtesting

This clarified how individual technical components integrate into a real-world quantitative pipeline.

4. Financial Data Processing with Python, Pandas & NumPy

I learned how Python libraries are used to process large-scale financial time-series data. Pandas enabled structured handling of stock prices, while NumPy allowed efficient numerical computations for returns and risk estimation.

I also learned how raw price data is transformed into meaningful features such as returns and volatility, which are essential for modeling and evaluation.

5. Natural Language Processing Fundamentals

I learned the importance of preprocessing unstructured textual data before applying sentiment models. This included:

- Tokenization
- Stop-word removal
- Lemmatization

These steps reduce noise and improve sentiment model accuracy.

6. Sentiment Analysis Techniques

I learned and compared two sentiment analysis approaches:

- VADER, a rule-based baseline model
- FinBERT, a transformer-based model trained specifically on financial text

This comparison highlighted the importance of domain-specific NLP models in finance.

7. Word Embeddings and Contextual Understanding

In Week 2, I learned the limitations of traditional word embeddings such as Word2Vec and GloVe, particularly their inability to handle polysemy. BERT introduced contextual embeddings, where word meaning dynamically changes based on surrounding words.

This concept was critical in understanding why BERT-based models outperform traditional NLP methods.

8. Transformer Architecture and Self-Attention

I studied the transformer encoder architecture used in BERT. The self-attention mechanism allows each word to attend to all other words in a sentence, enabling deep contextual understanding.

This architectural design was a key learning milestone.

9. Pretraining and Domain Adaptation in FinBERT

I learned how FinBERT is developed through:

- General language pretraining
- Further pretraining on financial corpora

- Supervised fine-tuning on labeled financial sentiment data

This process explained why FinBERT accurately interprets finance-specific terms that generic models misclassify.

10. Quantitative Finance Metrics

I learned that evaluating trading strategies requires both return and risk analysis. Metrics such as CAGR, Sharpe Ratio, and Maximum Drawdown provide a comprehensive view of performance.

11. Stationarity and Data Integrity

I learned the importance of stationarity in time-series modeling and applied the Augmented Dickey-Fuller test to validate assumptions. I also learned to align news timestamps with trading hours to prevent look-ahead bias.

12. Conclusion

Overall, Weeks 1 and 2 provided a strong foundation in sentiment-driven quantitative finance. I gained both conceptual clarity and practical understanding of how NLP and financial theory combine to build predictive trading systems. This learning prepares me for advanced modeling and strategy development in subsequent phases of the project.