**Sprint-1: Multi-Source Sentiment-Driven Stock Prediction and Trading Strategy**

**Introduction**

In this sprint, the goal was to integrate sentiment analysis from multiple sources, including financial news, social media, and financial reports, into a stock market prediction model. The objective was to enhance trading decisions and improve investment returns by leveraging sentiment-driven insights. The system was designed to preprocess and analyze sentiment data, integrate it into predictive models, and validate trading strategies through backtesting and forward testing.

[US291: As a trader, I want to integrate sentiment analysis from diverse sources, including financial news, social media, and financial reports, into my stock market prediction model so that I can enhance my investment returns, make better-informed trading](https://github.com/Rivier-Computer-Science/AI-Agent-Stock-Prediction/issues/291)

**Conditions of Satisfaction**

**Data Coverage & Quality:** Data pipelines must collect timely, relevant, and clean data from multiple sources.

**Sentiment Analysis Efficacy:** Models should accurately capture sentiment signals, including sarcasm and irony.

**Predictive Model Integration**: Sentiment features must show measurable improvements in stock predictions.

**Trading Strategy & Validation:** Sentiment-driven trading signals should be backtested and forward-tested.

**Performance Metrics:** Achieve target risk-adjusted returns and maintain controlled drawdowns.

**Documentation & Explainability:** Clear documentation of the entire process and model decisions.

**Definition of Done**

1. Data pipelines successfully collect, clean, and store sentiment data from multiple sources.

2. Sentiment analysis models achieve predefined accuracy benchmarks.

3. Predictive models incorporating sentiment data outperform baseline models.

4. Trading strategies demonstrate positive returns in backtesting, out-of-sample testing, and forward testing.

5. Risk management metrics confirm stable performance.

6. Complete documentation of models, trading strategies, and evaluation methods.

**Tasks**

[MSS 1.1: Data Collection and Processing (20 ph)](https://github.com/Rivier-Computer-Science/AI-Agent-Stock-Prediction/issues/292)

Subtask 1: Develop pipelines for news articles (4 ph)

Subtask 2: Collect social media data (4 ph)

Subtask 3: Gather financial reports (4 ph)

Subtask 4: Clean and preprocess data (4 ph)

Subtask 5: Conduct exploratory data analysis (4 ph)

**Tasks I Worked On**

Data Collection Agent Implementation

MSS 1.1: Data Collection and Processing (20 ph) 292

Developed automated pipelines to collect and preprocess sentiment data from news, social media, and financial reports.

Implemented preprocessing steps to handle missing data, noise, and multilingual content.

Conducted exploratory data analysis to ensure data quality and relevance.

**Sprint-2: Sentiment Analysis Model Development and Feature Engineering for Stock Prediction**

**Introduction**

In this sprint, the goal was to develop a comprehensive sentiment analysis model to process and interpret sentiment data, followed by engineering features that would be integrated into stock market prediction models. The sentiment models aimed to analyze and quantify the sentiment extracted from various sources, while the feature engineering process focused on creating meaningful features to enhance predictive accuracy. Additionally, model stacking techniques were explored to further optimize the performance of the stock prediction model.

[US291: As a trader, I want to integrate sentiment analysis from diverse sources, including financial news, social media, and financial reports, into my stock market prediction model so that I can enhance my investment returns, make better-informed trading](https://github.com/Rivier-Computer-Science/AI-Agent-Stock-Prediction/issues/291)

**Conditions of Satisfaction:**

**1. Sentiment Analysis Model Development:**

- Lexicon-based, machine learning, and deep learning models should effectively process and analyze sentiment data.

- Each model type (lexicon-based, machine learning, deep learning) must meet specific accuracy benchmarks.

**2. Feature Extraction and Engineering:**

- Extracted sentiment features must be relevant and representative of market movements.

- Feature engineering techniques should demonstrate a clear improvement in model performance when integrated.

**3. Model Stacking:**

- Implement model stacking to combine different models and improve prediction accuracy.

- Stacked models should outperform individual models in predictive accuracy.

**4. Addressing Data Challenges:**

- Address challenges such as handling missing data, noise, and multilingual content in sentiment analysis models.

**5. Integration with Predictive Models:**

- Sentiment features must be successfully integrated into stock prediction models.

- Models incorporating sentiment features must outperform baseline predictive models.

**6. Validation:**

- Validate the models using appropriate evaluation techniques such as backtesting and forward testing to ensure reliable predictions and actionable trading signals.

**Definition of Done:**

1. Lexicon-based models, machine learning models, and deep learning models are trained and validated against predefined accuracy benchmarks.

2. Sentiment features are extracted, engineered, and integrated into the stock prediction models.

3. Model stacking is successfully implemented, and stacked models outperform individual models.

4. Addressed challenges related to missing data, noise, and multilingual sentiment data.

5. Predictive models incorporating sentiment data demonstrate improved stock market forecasting and actionable trading signals.

6. Complete documentation of models, feature engineering process, and integration into the prediction pipeline.

**Tasks:**

**[MSS 1.2: Sentiment Analysis Model Development (25 ph) 371](https://github.com/Rivier-Computer-Science/AI-Agent-Stock-Prediction/issues/371)**

Agent: Sentiment Analysis Agent

**- Subtask 1: Train Lexicon-based Models (5 ph)**

- Develop and implement lexicon-based sentiment models to analyze sentiment in financial news, social media, and reports.

- Train the lexicon models to score sentiment (positive/negative/neutral) based on predefined lexicons.

- Evaluate model accuracy based on its performance on a labeled sentiment dataset.

**- Subtask 2: Train Machine Learning Models (5 ph)**

- Train machine learning models (e.g., logistic regression, SVM) using features derived from sentiment-labeled data.

- Fine-tune the hyperparameters of the machine learning models to improve accuracy.

- Validate the machine learning models using cross-validation techniques.

**- Subtask 3: Train Deep Learning Models (10 ph)**

- Implement and train deep learning models (e.g., LSTM, BERT) for sentiment analysis to capture more complex patterns and relationships.

- Preprocess the data for deep learning (e.g., tokenization, embedding layers).

- Train the models on large datasets and evaluate their performance against baseline models.

**- Subtask 4: Address Data Challenges (5 ph)**

- Handle missing data, noisy data, and multilingual content by applying appropriate preprocessing techniques.

- Use data augmentation techniques to improve the robustness of the models.

- Develop a strategy to handle imbalanced sentiment data if needed (e.g., oversampling/undersampling).

**[MSS 1.3: Feature Engineering and Model Integration (20 ph) 372](https://github.com/Rivier-Computer-Science/AI-Agent-Stock-Prediction/issues/372)**

Agent: Feature Engineering Agent

**- Subtask 1: Extract Sentiment Features (5 ph)**

- Extract sentiment-related features (e.g., sentiment polarity, intensity) from the output of sentiment analysis models.

- Identify key sentiment indicators that can be directly tied to stock market trends.

- Preprocess and structure the features in a format suitable for predictive models.

**- Subtask 2: Integrate Features into Prediction Models (5 ph)**

- Integrate the sentiment features with other technical and fundamental features used in stock prediction models.

- Develop a seamless pipeline to feed sentiment data into existing prediction models (e.g., regression, random forests).

- Evaluate the improvement in model performance with the integration of sentiment features.

**- Subtask 3: Explore Feature Engineering Techniques (5 ph)**

- Research and explore different feature engineering techniques (e.g., sentiment aggregation, temporal analysis) that can better capture the nuances in sentiment data.

- Test various combinations of sentiment features to optimize model performance.

- Identify feature importance and discard irrelevant or redundant features.

**- Subtask 4: Implement Model Stacking (5 ph)**

- Implement a model stacking strategy by combining different sentiment analysis models (lexicon-based, machine learning, deep learning) to improve prediction accuracy.

- Use ensemble methods such as stacking, boosting, or bagging to combine predictions from different models.

- Evaluate the performance of the stacked model in comparison to individual models.

**Tasks I Worked On**

**Sentiment Analysis Agent Implementation:**

**[- MSS 1.2: Sentiment Analysis Model Development (25 ph)](https://github.com/Rivier-Computer-Science/AI-Agent-Stock-Prediction/issues/371)**

- Developed and trained lexicon-based, machine learning, and deep learning models for sentiment analysis, addressing challenges like multilingual content and noisy data.

- Implemented and evaluated deep learning models such as LSTM and BERT for sentiment extraction.

- Focused on refining sentiment extraction processes to improve prediction performance.

**Feature Engineering Agent Implementation:**

**[- MSS 1.3: Feature Engineering and Model Integration (20 ph)](https://github.com/Rivier-Computer-Science/AI-Agent-Stock-Prediction/issues/372)**

- Worked on extracting sentiment features (e.g., polarity, intensity) from sentiment analysis outputs and integrated them into predictive models.

- Explored feature engineering techniques to refine the sentiment features and enhance model accuracy.

- Collaborated on implementing model stacking to combine the strengths of individual models, improving overall prediction results.

**Summary Table of Work**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **User Story ID** | **User Story** | **Story Points** | **Task ID** | **Task** | **Task Hours** | **Status** | **Actual Hours** |
| [291](https://github.com/Rivier-Computer-Science/AI-Agent-Stock-Prediction/issues/291) | As a trader, I want to integrate sentiment analysis from diverse sources into my stock market prediction model. | 3 | [292](https://github.com/Rivier-Computer-Science/AI-Agent-Stock-Prediction/issues/292) | MSS 1.1: Data Collection and Processing | 20 | Completed | 20 |
| [291](https://github.com/Rivier-Computer-Science/AI-Agent-Stock-Prediction/issues/291) | As a trader, I want to integrate sentiment analysis from diverse sources into my stock market prediction model. | 3 | [371](**MSS 1.2: Sentiment Analysis Model Development (25 ph)** " \l "371) | MSS 1.2: Sentiment Analysis Model Development (25 ph) | 25 | Completed | 25 |
| [291](https://github.com/Rivier-Computer-Science/AI-Agent-Stock-Prediction/issues/291) | As a trader, I want to integrate sentiment analysis from diverse sources into my stock market prediction model. | 3 | [372](https://github.com/Rivier-Computer-Science/AI-Agent-Stock-Prediction/issues/372) | MSS 1.3: Feature Engineering and Model Integration (20 ph) | 20 | Completed | 19 |

## **Summary Table of Commits**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Date | Commit Number | Commit Description (exactly as in github) | User Story | Task |
| 2/6/2025 | [d098a8b9fd840470223bf270e685c212543e9e48](https://github.com/Rivier-Computer-Science/AI-Agent-Stock-Prediction/pull/462/commits/d098a8b9fd840470223bf270e685c212543e9e48) | Update app.py | [291](https://github.com/Rivier-Computer-Science/AI-Agent-Stock-Prediction/issues/291) | [292](https://github.com/Rivier-Computer-Science/AI-Agent-Stock-Prediction/issues/292) |
| 2/26/2025 | [05468e5b12672956f910299b98c1a1db6b5fe47a](https://github.com/Rivier-Computer-Science/AI-Agent-Stock-Prediction/commit/05468e5b12672956f910299b98c1a1db6b5fe47a) | MSS 1.2: Sentiment Analysis Model Development (25 ph) | [291](https://github.com/Rivier-Computer-Science/AI-Agent-Stock-Prediction/issues/291) | [371](**MSS 1.2: Sentiment Analysis Model Development (25 ph)** " \l "371) |
| 3/8/2025 | [337947e5cf1863a21b33fbc6a0ae84073cb74e36](https://github.com/Rivier-Computer-Science/AI-Agent-Stock-Prediction/commit/337947e5cf1863a21b33fbc6a0ae84073cb74e36) | MSS 1.3: Feature Engineering and Model Integration | [291](https://github.com/Rivier-Computer-Science/AI-Agent-Stock-Prediction/issues/291) | [372](https://github.com/Rivier-Computer-Science/AI-Agent-Stock-Prediction/issues/372) |