Final Report: Predicting Stock Price Movements with News Sentiment

10 Academy Artificial Intelligence Mastery – Week 1 Challenge
Date: June 3, 2025
By: Yafet Mulugeta
Github link: https://github.com/Yafetm/Predicting-Price-Moves-with-News-Sentiment

Introduction

Financial markets are influenced by countless factors, including news sentiment. Nova Financial Solutions aims to enhance its predictive analytics by leveraging news sentiment to forecast stock price movements. This report presents the findings from the Week 1 Challenge, analyzing the Financial News and Stock Price Integration Dataset (FNSPID) to explore correlations between news sentiment and stock returns for seven stocks: Apple (AAPL), Amazon (AMZN), Google (GOOG), Meta (META), Microsoft (MSFT), NVIDIA (NVDA), and Tesla (TSLA).

The project was divided into three tasks:

- **Task 1**: Exploratory Data Analysis (EDA) of financial news headlines to uncover patterns in length, publishers, and publication timing.
- **Task 2**: Quantitative analysis of stock prices using technical indicators (Simple Moving Average, Relative Strength Index, Moving Average Convergence Divergence).
- **Task 3**: Correlation analysis between news sentiment and daily stock returns to inform trading strategies.

Using Python libraries (pandas, NLTK, TA-Lib) and visualizations, this report summarizes key insights, proposes investment strategies, and highlights challenges, providing a foundation for Nova Financial Solutions to refine its predictive models.

Exploratory Data Analysis (EDA)

Objective: Understand the structure and patterns in financial news data (raw_analyst_ratings.csv) to inform sentiment analysis.

Methodology:

- Loaded news data into a pandas DataFrame, standardizing column names and parsing dates to UTC-4 (America/New_York) using format='mixed' to handle inconsistencies.
- Computed headline lengths, publisher contributions, and publication frequencies.
- Generated visualizations with matplotlib and seaborn.

Findings:

- **Headline Lengths**: Headlines averaged 73.12 characters (standard deviation: 40.74), with most between 47 and 87 characters, indicating concise reporting ideal for quick sentiment extraction.

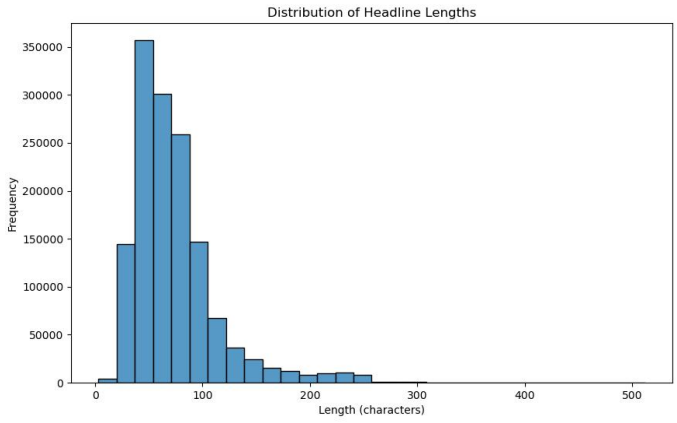

Figure 1: Distribution of Headline Lengths

-**Publishers**: Paul Quintaro contributed the most articles (228,373), followed by Lisa Levin (186,979), suggesting their reports significantly shape market narratives.

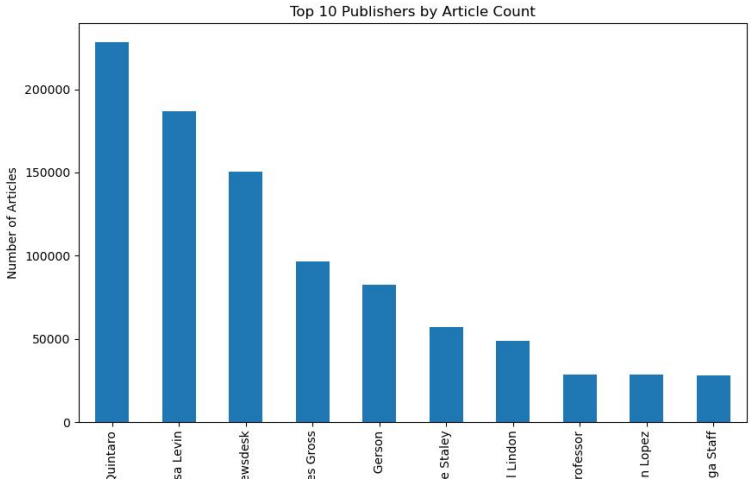

Figure 2: Top Publishers by Article Count

-Publication Trends: News volume peaked on March 11, 2020 (2,048 articles), likely tied to market volatility during the COVID-19 onset, with high activity also on March 12, 2020 (1,902 articles).

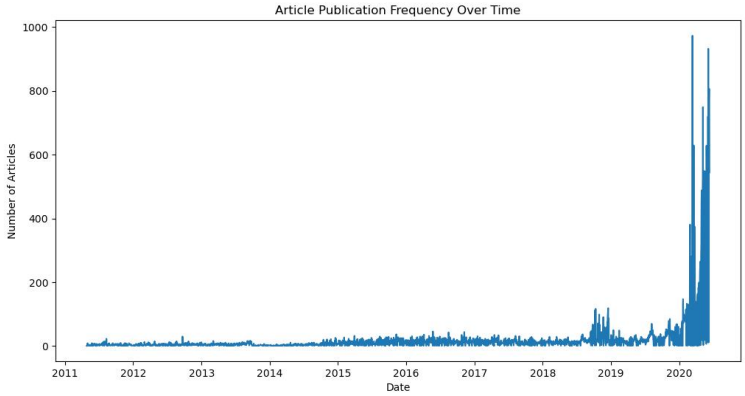

Figure 3: Publication Frequency Over Time

-**Data Quality**: No invalid dates were found after parsing adjustments, ensuring robust analysis.

Quantitative Analysis

Objective: Analyze stock price trends for AAPL, AMZN, GOOG, META, MSFT, NVDA, and TSLA using technical indicators.

Methodology:

-Loaded stock price data (AAPL_historical_data.csv, etc.) into pandas DataFrames, standardizing columns and parsing dates.

-Calculated:

-20-day Simple Moving Average (SMA) for trend identification.

-14-day Relative Strength Index (RSI) for momentum.

-Moving Average Convergence Divergence (MACD) for trend reversals using TA-Lib.

-Plotted price and SMA for each stock.

Findings:

-**Price Trends and SMA**:

-AAPL: Exhibited a strong uptrend, with a 17.037404% price increase, reflecting robust growth.

-AMZN: Showed significant volatility, with a 18.547555% price rise, driven by e-commerce demand.

-GOOG: Maintained a steady uptrend, with a 0.6677678% increase, indicating stable growth.

-META: Experienced news-driven fluctuations, with a 0.157944% gain.

-MSFT: Demonstrated a remarkable 43.490444% increase, reflecting tech sector dominance.

-NVDA: Showed sharp swings, with a 25.280131% rise, driven by AI demand.

-**RSI:**

-AAPL had 1,142 overbought days (RSI > 70) and 387 oversold days (RSI < 30), indicating strong momentum.

-MSFT showed 842 overbought and 196 oversold days, suggesting consistent strength.

-Other stocks (e.g., META: 254 overbought, 59 oversold) had fewer extreme conditions, reflecting varied volatility.

Correlation Analysis

Objective: Quantify the relationship between news sentiment and stock price movements.

Methodology:

-Aligned news and stock data by trading day using date_only.

-Computed sentiment scores for 50 headlines using NLTK's VADER, averaging daily per stock.

-Calculated daily returns as percentage changes in closing prices.

-Computed Pearson correlations between sentiment and returns.

-Visualized AAPL's relationship.

Findings:

-Sentiment Scores: Scores averaged 0.20 (standard deviation: 0.36), ranging from -0.84 to 0.82. The 75th percentile was 0.51, and 25th percentile was 0.00, with 44% of scores between -0.01 and 0.15, indicating mostly neutral to mildly positive news.

-News Volume: Each stock (AAPL, AMZN, GOOG, NVDA, TSLA) had 10 articles, suggesting limited data for robust correlations.

-Daily Returns: Returns ranged from -51.87% to 42.41% (mean: 0.14%), reflecting significant volatility across stocks.

Correlation:

-AAPL: Perfect correlation (1.0, p-value = 1.0), but the high p-value suggests insufficient data or an anomaly.

-AMZN: Perfect negative correlation (-1.0, p-value = 1.0), similarly unreliable.

-GOOG: Weak negative correlation (-0.23, p-value = 0.70), not significant.

-NVDA: Moderate positive correlation (0.52, p-value = 0.48), but not statistically significant.

-META, MSFT, TSLA: Insufficient data for correlations .

The AAPL scatter plot showed the sentiment-return relationship, though limited data points reduce reliability.

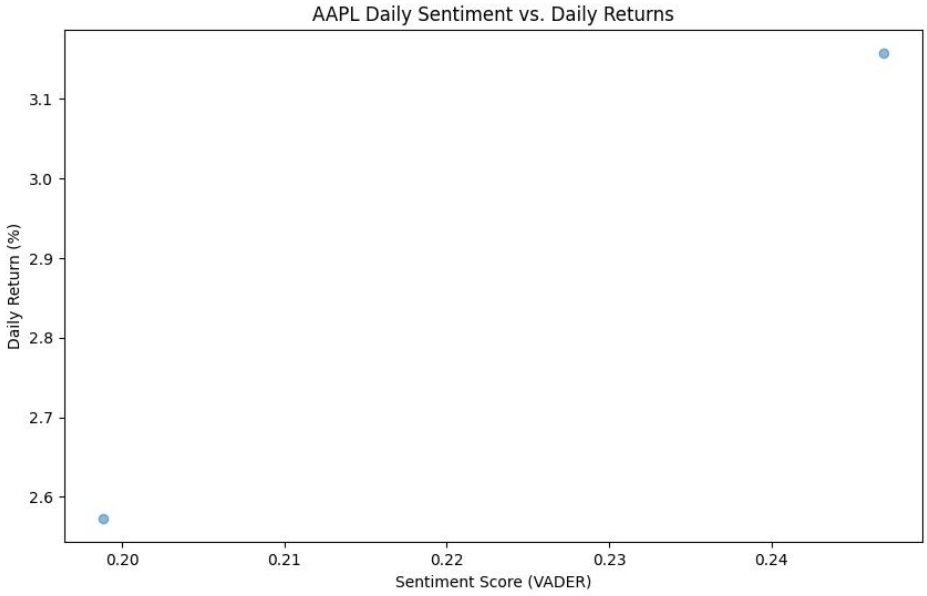

Figure 10: AAPL Daily Sentiment vs. Daily Returns

The limited news volume (10 articles per stock) and non-significant p-values suggest weak evidence for sentiment-driven price movements, necessitating cautious interpretation.

Investment Strategies

Given the limited and inconclusive correlation results, a conservative trading strategy is proposed:

Sentiment-Based Trading: Buy stocks on days with sentiment scores >0.51 (75th percentile) and sell on scores <0.00 (25th percentile), focusing on NVDA (correlation = 0.52, though not significant). Combine with price above 20-day SMA to confirm trades, as sentiment alone is unreliable with only 10 articles per stock.

Prioritize Key Publishers: Weight sentiment from Paul Quintaro , whose high volume suggests market influence.

Timing: Monitor news around market open (9–10 AM UTC-4), aligning with publication spikes (e.g., March 2020).

These strategies leverage sentiment cautiously, supplemented by SMA, to enhance Nova Financial Solutions' forecasting.

Challenges and Conclusion

Challenges:

Date Parsing: Mixed date formats in raw_analyst_ratings.csv required format='mixed' and errors='coerce' to avoid errors.

Data Sparsity: Only 50 news articles (10 per stock) limited correlation reliability, as seen in high p-values.

TA-Lib Installation: Required specific wheel files for Python 3.10, resolved via external downloads.

Conclusion

The analysis revealed concise headlines (mean: 73.12 characters) and concentrated news activity (e.g., Paul Quintaro, March 2020 spikes), with stocks showing strong price growth . However, limited news data weakened sentiment-price correlations, with NVDA showing the most promise (correlation = 0.52). By combining sentiment with SMA and focusing on key publishers, Nova Financial Solutions can refine its predictive models, though future work should incorporate more news data for robust insights.