

---

**Amr Elhag Aly Y00866881**

amelhagaly@student.ysu.edu

## **Stock Prediction with Sentiment Analysis and Technical Indicators**

### **Introduction**

The objective of this project is to develop a stock prediction system based on real-time financial news sentiment analysis and stock technical indicators. The system fetches data, processes news articles for sentiment, merges with stock data, and uses machine learning models to predict stock price movements using AWS services such as Lambda, SageMaker Autopilot, and S3. This work contributes to financial decision-making by offering timely insights from multiple data sources and technical analysis.

### **Background**

The domain of stock price prediction is a well-researched one that combines time series analysis, machine learning, and sentiment analysis. Stock prediction models based on past stock price data are traditional, but adding alternative data sources such as news sentiment can increase prediction accuracy. Natural language processing (NLP) has advanced in recent years to the point where sentiment analysis from text is easier to analyze and quantify, and can be a valuable source of information about market movements. The AWS suite of services makes it easy to integrate machine learning and data processing workflows so that prediction systems can be created that are robust.

---

## Problem Statement

The stock price movements are volatile and hence it is a complex task to predict. Traditional models based on historical price data are limited in their ability to capture the effect of real-world events such as news releases. As a result, this project seeks to improve prediction accuracy by combining news sentiment and technical analysis indicators and using AWS services to process data at scale and efficiency.

## Data Sources

- Alpha Vantage News API: It fetches real-time news data about stock tickers and captures sentiment information directly from news sources.
- Yahoo Finance API: To represent stock movements, stock prices and associated technical indicators (SMA, EMA, RSI, MACD, Bollinger Bands) are fetched.

## Preprocessing and Integration

- Sentiment Analysis: News article text is cleaned, tokenized, and lemmatized, and sentiment scores are derived from the text using Alpha Vantage's sentiment API.
- Technical Analysis: Common technical indicators such as Simple Moving Average (SMA), Exponential Moving Average (EMA), Relative Strength Index (RSI), and Bollinger Bands are calculated using stock prices fetched from Yahoo Finance.

## Modeling

- Machine Learning: The sentiment data and technical indicators are preprocessed and passed into AWS SageMaker Autopilot for automatic machine learning model selection, training, and evaluation. The dataset characteristics are used to choose the best model by SageMaker.
- Data Flow: By integrating both news sentiment and technical analysis data into the model, the system can predict stock price movements more accurately by combining news impact and historical price data.

---

## AWS Services Used

- Lambda: Fetching of stock and news data is automated by AWS Lambda functions, as well as the model training and inference process is triggered without the need to provision or manage servers.
- S3: Amazon S3 buckets store data, making it easy to access, scalable, and secure data management.
- SageMaker Autopilot: It automates the process of model selection, training, and tuning so that you can easily develop accurate predictive models without having deep expertise in machine learning.

## Experiments

### Data Preprocessing

Text Preprocessing Pipeline: News articles were text cleaned, tokenized and lemmatized. To have high-quality input for sentiment analysis, we removed non-informative words and special characters.

### Sentiment Analysis

Alpha Vantage's News API was used to derive sentiment scores, which are labeled as "Bullish," "Bearish," and "Neutral."

The stock price information is stored along with the sentiment data so that we have a complete picture of the market's mood at the time of each news release.

### Model Training

The combined dataset of stock prices and sentiment scores was used to automatically train a model using AWS SageMaker Autopilot. Model selection (e.g. XGBoost) and hyperparameter tuning were handled by SageMaker Autopilot.

---

## Discussion

### Technology Choices

- Alpha Vantage API for News Sentiment: This API allows you to get real-time access to news articles, adding sentiment data from financial news sources directly into stock prediction models. Sentiment scores represent the market's general reaction to news events and the likely impact of the news on stock prices.
- Yahoo Finance for Stock Prices and Technical Indicators: Stock price data, including essential technical indicators, is widely used and reliable from Yahoo Finance. These indicators (SMA, EMA, RSI, MACD, and Bollinger Bands) are often used in trading strategies and provide the numerical basis for forecasting future stock price movements.
- AWS Lambda: Lambda was chosen because of its serverless, event-driven nature, which is perfect for automating data fetching, processing, and triggering of model inference without the need to manage server infrastructure. It also reduces operational overhead.
- AWS SageMaker Autopilot: The reason it was chosen was that SageMaker Autopilot can automatically select a model, train it, and tune it. With the complexity of picking the best machine learning algorithm for such a dataset, SageMaker makes it easy for people who aren't experts in machine learning.

### Model Performance

Initial results are promising, but the model needs to be fine-tuned and more data to improve prediction accuracy. One of the current challenges is to ensure that news sentiment has a true reflection on stock price changes and to address the time lag between news release and stock price reactions

---

## Training Data Sample

Date: 2024-11-29

- **Title:** *Amazon Rises 10.4% Since Q3 Earnings: Time to Buy the Stock?*
- **Summary:**  
"AMZN's soaring profits, AWS dominance, and AI leadership make it a compelling buy despite premium valuation, as growth momentum signals strong upside potential."
- **Sentiment:** Bullish (Score: 0.498544)
- **Source:** Zacks Commentary
- **Topics:** Retail & Wholesale, Financial Markets, Earnings, Technology
- **Related Ticker Sentiment:**
  - AMZN: *Bullish* (Score: 0.507017)
- **Stock Price (AMZN):**
  - Open: 205.83
  - Close: 207.89
  - Volume: 24,892,400
  - SMA-20: 204.50
  - RSI: 58.68
  - Bollinger High: 214.51
  - Bollinger Low: 194.50

## Conclusion

This project shows how sentiment analysis and technical indicators can be combined for stock price prediction using AWS services. The system combines news sentiment and technical indicators in a machine-learning model to produce stock price predictions that may be more accurate than traditional models. The system has been made scalable and flexible using AWS services such as Lambda, SageMaker Autopilot, and S3.

---

## References

1. Alpha Vantage. (2024). **News Sentiment API**. Retrieved from <https://www.alphavantage.co/documentation/#news-sentiment>
2. Yahoo Finance. (2024). **Yahoo Finance API**. Retrieved from <https://www.yahoofinanceapi.com/>
3. Amazon Web Services (AWS). (2024). **AWS Lambda**. Retrieved from <https://aws.amazon.com/lambda/>
4. Amazon Web Services (AWS). (2024). **Amazon SageMaker Autopilot**. Retrieved from <https://aws.amazon.com/sagemaker/autopilot/>