Stock Analysis using AI and Machine Learning Models

Elizabeth Flynn, *UCLA*. Kevin Gerges, *USC*
Jasper Garcia, *La-Tech,* Stephanie Manalo, *SMC*
Christopher Osgood, *Pierce College,* Kleber Ordonez, *Pierce College*

1. INTRODUCTION

Stock price prediction is an important tool that helps investors and analysts make informed financial decisions. Traditional forecasting methods often rely solely on historical data, limiting their responsiveness to real world events and changing public sentiment. Our project aims to improve prediction accuracy by integrating advanced AI techniques, including LLM for smarter reasoning, natural language processing to understand news and public opinion. The primary objective was to develop and compare the performance of two models: a machine learning model trained on time series data, and an AI model enhanced with LangChain agents, sentiment analysis, and Groq. This report outlines the methodologies, system architecture performance metrics, and key findings.

2. SYSTEM DESIGN AND ARCHITECTURE

We collected daily closing prices for four companies: Tesla, Nvidia, Dow Inc., and ExxonMobil. Historical stock data was gathered using the yfinance python library, which retrieves information directly from Yahoo Finance. During the preprocessing phase, financial data was cleaned to remove null values, normalized for consistency, and organized chronologically. Sentiment data was extracted from news headlines and earnings call transcripts using NLP techniques assigning each a daily sentiment score – positive, neutral or negative, based on the tones of the articles. For model development, we used the open-source Prophet library to build a time series forecasting baseline. Prophet is designed to handle seasonality, trends and other patterns in time series data. We integrated sentiment analysis by analyzing financial news and headlines related to each stock for more context to our predictions. To enhance the model's contextual understanding, we used LangChain agents that processed and interpreted information from multiple sources, including news and earnings calls. These agents leveraged LLMs to perform reasoning tasks, generate insights and refine predictions. We also utilized vector databases for storing text embeddings and Groq for accelerating inference. This multi-layered architecture allowed the AI model to combine quantitative data with unstructured information, leading to not only relevant but more context-aware predictions.

3. RESULTS AND DATA COLLECTION

Both the ML-based time series model and the AI model with LLM integration were tested on four companies: Tesla, Nvidia, Dow Inc., and ExxonMobil. Daily closing prices were collected over a 14 day period. Predictions were logged alongside actual closing prices in a shared spreadsheet for evaluation. To assess model performance, we used three key metrics: Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and percentage error. MAE measured the difference between predicted and actual prices, RMSE penalized larger errors more heavily, offering insight into prediction reliability. Percentage error standardized the differences across various stock prices, making cross comparison easier. These metrics were calculated for both models across all stocks and results were visualized using Tableau, showing how closely the predictions matched actual values. The AI model outperformed the ML model, particularly during days with market-moving news or earnings calls. For stocks more sensitive to market sentiment, like Tesla and Nvidia, the AI model had a lower MAE and RMSE, indicating more precise and stable forecasting. The ML model performed reliably during calm market periods but struggled to respond to real time sentiment or sudden market shifts. Data collection suggests that while both models were effective, the AI model's ability to factor real time sentiment and contextual data led to greater forecasting precision.

4. CONCLUSION

This project demonstrates the value of integrating language models and sentiment data into financial forecasting systems. While both models offer strengths, the AI's model's responsiveness to market events shows promise for future development. Further testing with longer timeframes and additional variables could enhance reliability and provide deeper insights into model behavior. Our approach could improve real time investment strategies and inform future financial AI applications by demonstrating the advantage of combining structured and unstructured data in predictive models.

GitHub Repository:
https://github.com/elizabethf1706/FinancialStockPricePredictor

Excel AI Data Sheet:
Stock Accuracy Testing(regular ai)

Excel ML Data Sheet:
Stock Data Accuracy Testing2