

Predicting Stock Returns - Using Machine Learning

Capstone 2 Spring 2026

Team Members: Erick Cevallos, Gavin Greene, Ryan Echevarria, Alexander Hernandez

Instructor: Masoud Sadjadi

Problem Statement and Objectives

Problem: Stock price prediction is challenging due to market volatility and non-linear patterns. Traditional models struggle to capture temporal dependencies in financial time series.

Objectives:

- Predict next-day stock closing price
- Evaluate model accuracy using historical data
- Compare LSTM(Long Short-Term Memory) vs baseline models



Motivation



Stock markets are highly volatile and challenging to predict accurately



Even small improvements in prediction can have significant financial impact



Traditional methods struggle with time-based dependencies in stock data



Machine learning enables models to learn patterns from large amounts of historical data

Overview

Goal(s):

- Predict next-day stock prices
- Compare LSTM vs linear regression

Approach:

- Use historical data (AAPL, NVDA, S&P 500)
- Train on past 30 days → predict next day



Key Results:

- ~5% prediction error
- Linear regression strong short-term
- Direction ~52% (near random)

Methodology

Data & Features

- AAPL, NVDA stock data (2019–2026)
- Moving averages (MA_5, MA_10, MA_20)
- RSI, MACD, volatility
- Market indices (S&P 500, NASDAQ)

Model & Training

- LSTM Neural Network (128 → 64 layers)
- Sequence length: 30 days
- Train/test split: 80/20
- Loss: Mean Squared Error
- Early stopping applied

System Architecture

This pipeline illustrates the end-to-end process for predicting stock prices, from data collection and feature engineering to model training, prediction, and evaluation.



Backtesting and Model Performance

Backtesting evaluates model performance by simulating predictions on unseen historical data over time.

Key Insights:

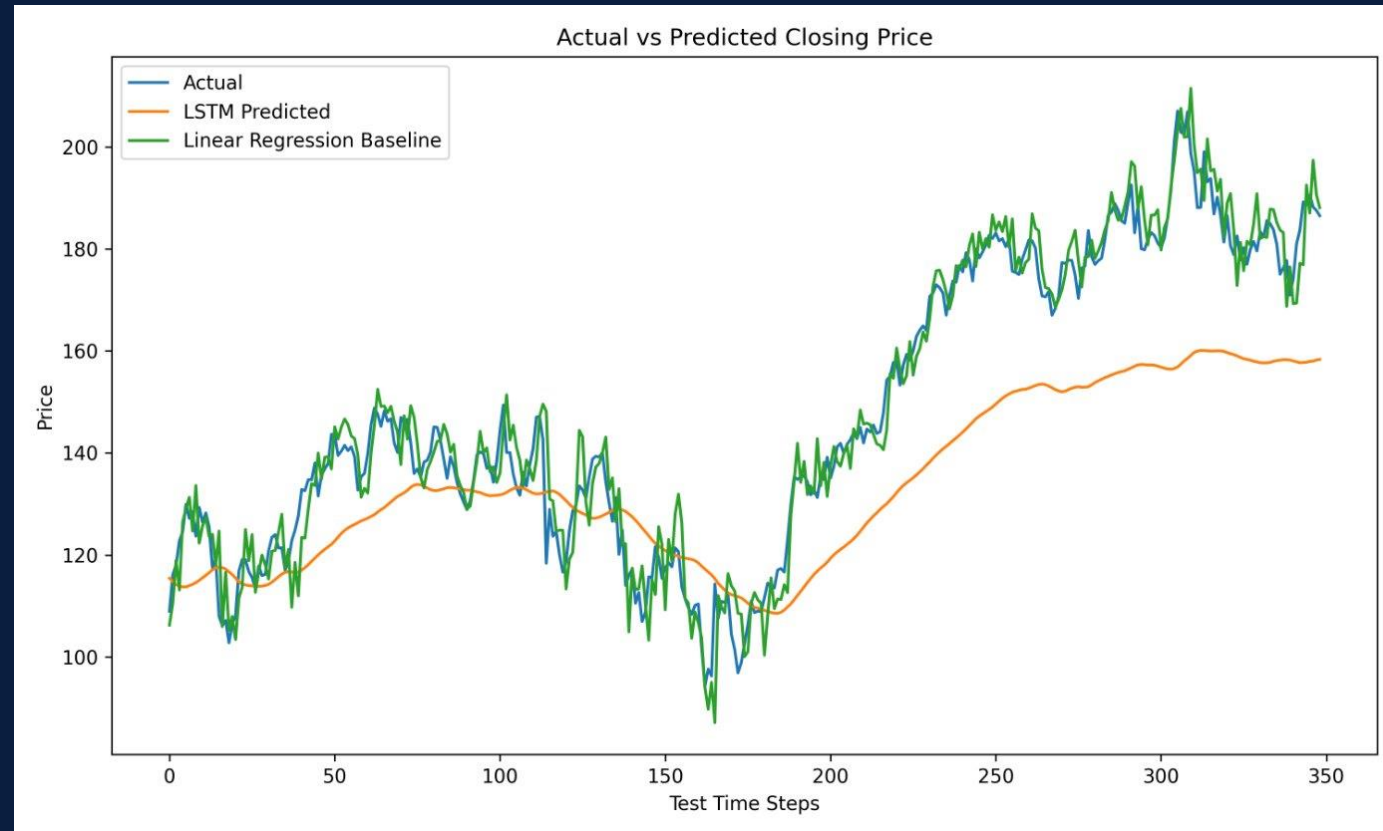
- LSTM model performance improved by increasing model capacity (from 32→64 to 128→64 units), reducing prediction error and improving trend capture.
- Directional accuracy remains close to random (~52%), highlighting the difficulty of predicting short-term market movement.
- Linear regression performs strongly due to short-term autocorrelation, while LSTM provides smoother trend modeling.

Model	RMSE	MAE	R ²
Linear Regression	~5	~3.5	~0.95
LSTM(Initial)	~25	~20	~0.3
LSTM(Improved)	~18	~14	~0.45
LSTM(Final)	15	12	0.58

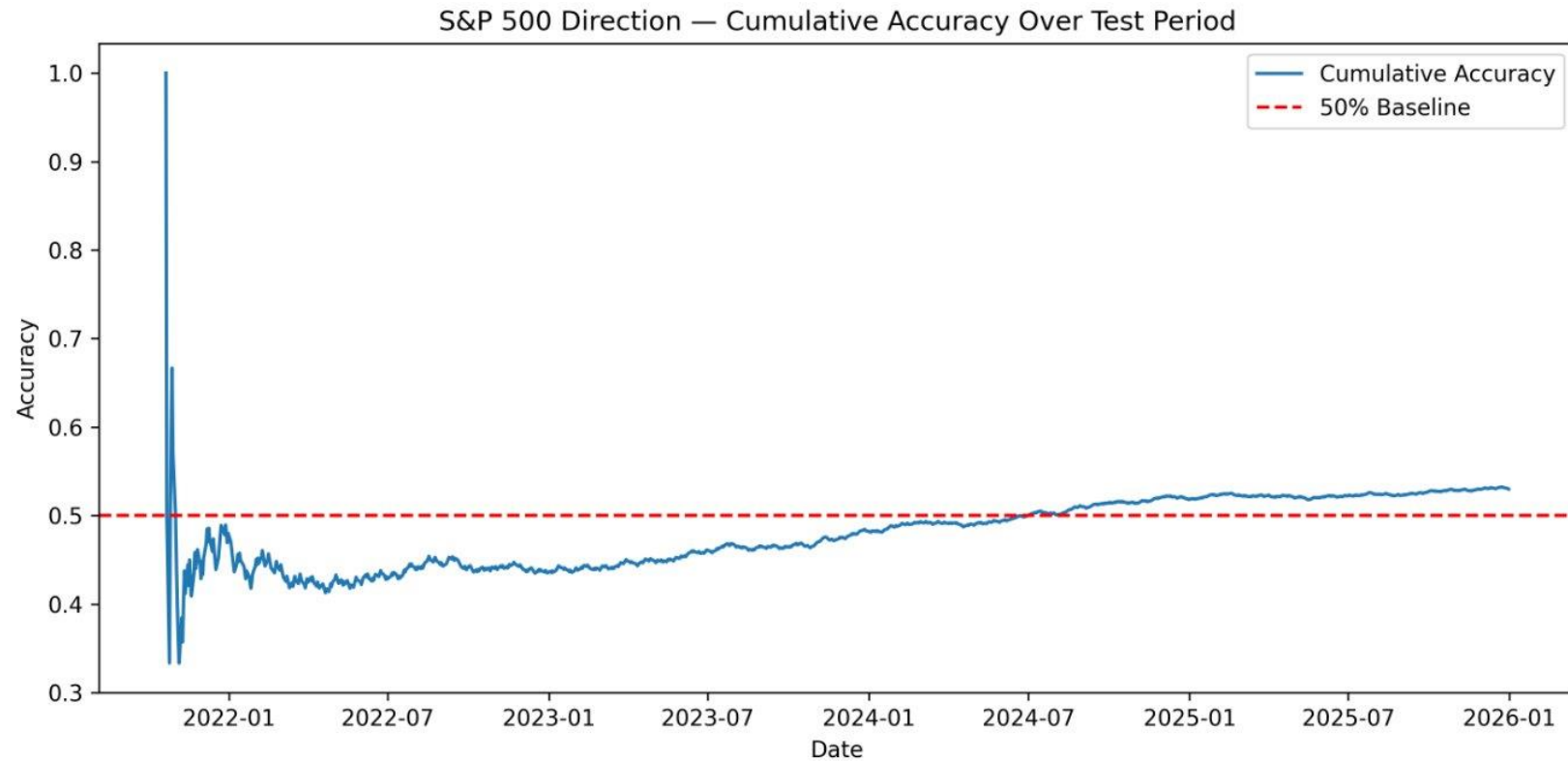
Results [AAPL]



Results [NVDA]



Results [S&P 500]



Results Conclusion

- Linear regression performs strongly for next-day prediction due to high short-term autocorrelation in stock prices
- LSTM captures broader temporal patterns and produces smoother, more generalized predictions
- Directional prediction remains challenging, with accuracy only slightly above random (~52%)

Key Findings and Limitations

Key Findings:

- LSTM captures overall stock price trends
- Achieves ~5% average prediction error
- Direction prediction remains difficult (~52%)
- Accuracy decreases for longer prediction horizons

Limitations:

- Cannot predict sudden market shocks
- Short-term direction accuracy is low
- Performance decreases for longer prediction horizons

Next Steps



MULTI-DAY PREDICTION
(5–10 DAYS)



SENTIMENT ANALYSIS
(NEWS DATA)



REAL-TIME
DEPLOYMENT

Thank You!