Stock Price Prediction

Introduction

In this project, the aim is to utilize machine learning techniques to develop a profitable algorithmic trading strategy by predicting future stock prices based on historical daily Open, High, Low, and Close (OHLC) price data. To enhance the predictive power of the model, historical Twitter sentiment analysis scores will also be incorporated.

Task

The primary task involves creating a predictive model capable of accurately forecasting stock prices by combining historical stock price data and sentiment analysis scores from Twitter. The goal is to generate actionable trading signals and assess the model's performance in a real-world trading scenario.

Prepare

Before implementing the project, it's essential to ensure the necessary tools and data are available. Here's what is required:

- **Python:** Python installation is required for data manipulation, machine learning, and sentiment analysis.

- **Required Python Packages:** Essential Python packages such as pandas, numpy, scikit-learn, tweepy (for Twitter data), and yfinance (for stock data) need to be installed and imported. The yfinance package can be found at https://pypi.org/project/yfinance/.

- **Historical Stock Data:** Download historical daily OHLC stock price data from Yahoo Finance using the yfinance package. This data will serve as the primary input for the model.

- **Twitter Data:** Utilize the Twitter API (through tweepy or similar libraries) to collect relevant tweets related to the stocks of interest. Perform sentiment analysis on these tweets to generate sentiment scores.

Steps

1. Data Import and Exploratory Data Analysis.

2. Twitter Sentiment Analysis Score.

3. Data Pre-Processing.

4. Model Implementation.

5. Model Evaluation.

6. Bibliography.

Result

The results are as follows:

Model	Stock	Lag Period	EVR	CoD	RMSE	MAPE
LSTM	AMCR	5 days	0.7516	0.7229	0.1656	0.0110
LSTM	AMCR	10 days	0.7779	0.7772	0.1488	0.0099
LSTM	AMCR	20 days	0.7926	0.7900	0.1512	0.0100
LSTM	NVR	5 days	0.9447	0.9447	115.40	0.0184
LSTM	NVR	10 days	0.9445	0.9434	117.77	0.0191
LSTM	NVR	20 days	0.8991	0.8965	148.70	0.0245
LSTM + Sentiment	AMCR	5 days	0.7388	0.7358	0.1607	0.0109
LSTM + Sentiment	AMCR	10 days	0.7573	0.7463	0.1604	0.0106
LSTM + Sentiment	AMCR	20 days	0.7348	0.7257	0.1748	0.0120
LSTM + Sentiment	NVR	5 days	0.9469	0.9465	112.80	0.0178
LSTM + Sentiment	NVR	10 days	0.9336	0.9101	144.52	0.0240
LSTM + Sentiment	NVR	20 days	0.9428	0.9407	108.86	0.0177
Random Forest	AMCR	5 days	0.7612	0.7612	0.1528	0.0101
Random Forest	AMCR	10 days	0.7352	0.7352	0.1622	0.0110
Random Forest	AMCR	20 days	0.7303	0.7303	0.1713	0.0122
Random Forest	NVR	5 days	0.7146	0.5962	311.91	0.0414

Model	Stock	Lag Period	EVR	CoD	RMSE	MAPE
Random Forest	NVR	10 days	0.6987	0.5852	318.74	0.0417
Random Forest	NVR	20 days	0.7262	0.6683	266.13	0.0351