Stock Price Prediction Using ARIMA, LSTM, and Moving Averages

Luv Valecha
Dheeraj Kumar
Dhruv Sharma
Arafat Ahmed Sheikh

OVERVIEW

This project aims to predict the future of stock price using previous data using models like ARIMA, LSTM and Moving Averages. Each model brings unique strengths to address the challenges of time-series forecasting in stock price prediction.

ARIMA (AutoRegressive Integrated Moving Average) is a statistical model suited for capturing patterns in historical stock prices by analyzing
autocorrelations in the time series. It's particularly useful for stationary data and provides short-term predictions by considering past values and error terms.

LSTM (Long Short-Term Memory) networks, a type of recurrent neural network, excel in capturing long-term dependencies in sequential data. By leveraging the neural network's architecture, LSTM models can handle complex non-linear patterns, making them valuable for predicting future prices based on historical data trends and seasonality

MOVING AVERAGES are used to smooth stock prices, removing noise and highlighting trends. Simple Moving Average (SMA) and Exponential Moving Average (EMA) provide insights into stock price direction, helping to indicate long- and short-term trends that traders use for strategy development.

Each model brings unique perspectives and is applied to provide a robust, well-rounded approach to predicting stock price movements. Combining statistical and machine learning models allows for both interpretability and flexibility in stock prediction.

ARIMA Model for Stock Price Prediction

Model

The ARIMA model, short for AutoRegressive Integrated Moving Average, is a statistical approach commonly used for time-series forecasting. It combines three components:

- **AR (AutoRegression):**
  The model uses the relationship between an observation and several lagged observations (past values).
- **I (Integrated) :**
  Differencing the data helps make it stationary, ensuring consistency over time.
- **MA (Moving Average):**
  The model utilizes dependency between observations and residual error terms from a moving average model applied to lagged errors.

ARIMA is parameterized by three values, written as ARIMA(p, d, q):

- **P:** Order of the autoregressive part (number of lag observations).
- **d:** Degree of differencing (how many times the data is differenced to make it stationary).
- **q:** Order of the moving average part (number of lagged forecast errors).

The ARIMA model works well with stationary data, where statistical properties remain consistent over time. For non-stationary stock prices, differencing (d) helps stabilize the mean.

Model Implementation:

To use ARIMA:

1. Make the data stationary by differencing as needed.
2. Determine (p) and (q) through plots like the ACF (Autocorrelation Function) and PACF (Partial Autocorrelation Function).
3. Fit the model and evaluate its performance on historical stock data.

By accounting for past values and error terms, ARIMA is suitable for short-term stock price prediction, especially in stable markets.

LSTM Model for Stock Price Prediction

LSTM Model

Long Short-Term Memory (LSTM) networks are a specialized form of recurrent neural networks (RNNs) designed to model temporal sequences and learn long-term dependencies effectively. Traditional RNNs face challenges with the vanishing gradient problem, which limits their ability to learn from long sequences. LSTMs address this by incorporating a memory cell that can maintain information over time.

The core architecture of an LSTM consists of three main gates:

- **Forget Gate:**
  This gate decides what information from the cell state should be discarded. It takes the previous output and current input, processes them through a sigmoid activation function, and generates a value between 0 and 1 for each number in the cell state. A value closer to 0 means to forget the information, while a value closer to 1 indicates to retain it.

- **Input Gate:**
  The input gate determines which new information should be stored in the cell state. It also employs a sigmoid activation function to decide which values to update. Additionally, a tanh activation function is used to create a vector of new candidate values that could be added to the cell state.

- **Output Gate:**
  This gate controls the output from the cell state. It applies a sigmoid function to decide which parts of the cell state will be output. The cell state is then passed through a tanh activation function to ensure the output is within a certain range.

The combination of these gates enables LSTMs to selectively retain or discard information, making them highly effective for time-series forecasting tasks, such as stock price prediction. By processing sequences of historical stock prices, LSTMs can identify complex patterns and trends that traditional models might miss.

Moving Averages for Stock Price Prediction

Moving Averages

Moving Averages (MA) are fundamental tools in time series analysis, particularly in financial markets, for smoothing price data and identifying trends. The primary goal of moving averages is to filter out noise from price fluctuations, allowing traders and analysts to focus on the underlying trend.

There are two common types of moving averages:

- **Simple Moving Average(SMA):** The SMA is calculated by averaging a set number of past price data points. For instance, a 10-day SMA sums the closing prices of the last 10 days and divides by 10. This moving average shifts over time, providing a dynamic view of the stock's trend.
- **Exponential Moving Average (EMA):** Unlike the SMA, which gives equal weight to all prices in the average, the EMA assigns more weight to recent prices, making it more responsive to new information. The EMA is calculated using a smoothing factor, which can be adjusted based on the desired sensitivity.

Moving averages are primarily used for trend identification, allowing traders to distinguish between upward, downward, and sideways movements. A rising moving average indicates an upward trend, while a falling moving average suggests a downward trend. Crossovers between short-term and long-term moving averages are also significant trading signals; for example, a bullish signal occurs when a short-term MA crosses above a long-term MA, and a bearish signal occurs when the opposite happens.

In stock price prediction, moving averages help establish support and resistance levels. Traders often use them in conjunction with other technical indicators to refine their strategies. By analyzing the convergence and divergence of moving averages with stock prices, analysts can gain insights into potential future price movements.

EXPLORATORY DATA ANALYSIS

Based on the plots mentioned in the EDA notebook, here are some specific observations and inferences:

1. **Time Series of Stock Prices (Open, High, Low, Adjusted Close)**:

- **Observation**: Each price metric (Open, High, Low, Adjusted Close) shows an upward trend over time, with some visible volatility at specific periods.
- **Inference**: The general upward trend suggests that Apple's stock has appreciated in value over the examined period, although certain points of volatility may indicate market events, earnings reports, or economic factors influencing the stock price.

2. **Moving Average for High Prices** (10, 20, and 30-day moving averages):

- **Observation**: The 10-day moving average is more responsive to short-term price changes, while the 20- and 30-day moving averages show smoother, less volatile trends.
- **Inference**: Shorter moving averages capture recent price changes more closely, while longer moving averages highlight the broader trend, filtering out short-term fluctuations. This is useful for identifying longer-term trends and potential support or resistance levels for trading.

3. **Volume Analysis**:

- **Observation**: The volume plot likely shows spikes in trading activity at specific times.
- **Inference**: These peaks might correspond with major market events, such as earnings announcements, product launches, or economic news. Higher volume during these events can indicate increased interest and possibly a price trend following the event.