Name: Sai Ram Gunturu
ID: C00313478

Note: This is a completely new notebook created by me

Linear Regression: Stock Price Prediction Notebook Project Log

---

Phases	Techniques Used	Reason for the Technique Used	Duration	Difficulty level (1-10)
Dataset selection	Used the Apple stock price real time data from the yahoo finance website	To predict the real-world data which will get updated every day.	30 min	7
Algorithm Version	Simple Linear model and Ridge Regression	To test the performance, I have used the baseline model and to improvise the model I have used Ridge regression	1 hour	6
Further Enhancements	Trained with Baseline Simple Linear Regression Model	To test how the dataset work with the existing model	4 hours	9
	7 - Day Simple Moving Average	Although results were pretty impressive with baseline model, but it has huge MSE with me lead to wrong predictions, So I implemented 7 SMA		
	3 - Day Simple Moving Average	Great improvement from 7 SMA with MSE as only 20%, but this is not ideal to build a robust model. To improvise the model, I implemented 3 SMA		

	Added Volatility Indicator	Amazing performance from 3 SMA as it generated only 4.9 % MSE with 99% accuracy. But its sensitivity might be a concern if we look for long term predictions, to mitigate this I added volatility indicator		
	Ridge Regression	Although Volatility doesn't made any drastic improvements but it is very important during market fluctuations, to overcome this I have choose ridge regression		
Conclusions		Because of its simple L2 regularization (which is relevant to our use case) it helped to build a robust model. The results might look the same for ridge and normal regression, Ridge plays a crucial role when it comes to long term predictions. Decision making is the thing I have learnt from this learning. It helped me to select the best version of linear regression also constant fine tuning definitely helps the model to predict better.	30 min	6

1. Business Understanding

In this work, I want to make predictions about the closing price of Apple Inc. (AAPL) based on historical daily data of stocks. I started with a baseline model fitting a simple linear regression using the date (converted to its ordinal value) as the sole predictor. While the model gave a first glance to the trend, its prediction was not perfectly matched with the recent market trends. This difference indicated a further model improvement to include more market dynamics.

2. Data Understanding & Preparation

I sourced daily stock price data for AAPL from Yahoo Finance. This dataset contains daily closing price and other related metrics. After examining the data, I enhanced it by calculating technical indicators:

- I calculated a 7-day Simple Moving Average (SMA) to identify short-term trends but it suffered from a bit of lag.
- Then I had to calculate a more responsive version of the one above, a 3-day SMA, to track the recent price changes.
- I also derived the 3-day volatility (rolling standard deviation) to measure short-term fluctuations in the market. These additional features expanded the dataset and offered the model with a more comprehensive set of inputs.

3. Modeling & Training

3.1 Baseline Model

At first, I implemented a baseline linear regression model with only the date (converted into ordinal) as predictor. This model achieved an MSE of around 193.42 and an R^2 of approximately 0.78, predicting a closing price that deviated significantly from recent market trends.

3.2 Improved Model using Technical Indicators

To improve the model, I added some more features:

- First, I added the 7-day SMA, which reduced the error but still lagged behind the market movements.
- Then, I substituted it with a 3-day SMA, which proved more responsive and also lowered the MSE further to about 4.98, with R^2 around 0.9940. The prediction for March 31, 2025, was approximately $221.06.
- At the end, I added the 3-day volatility indicator to account for market fluctuations. The final model, using the date (ordinal), 3-day SMA, and 3-day volatility as predictors, yielded an MSE of 5.11 and an R^2 of 0.99347. The prediction for March 31, 2025, was approximately $221.08.

4. Evaluation

The final model performed exceptionally well:

- Mean Squared Error (MSE): 5.11
- R^2 Score: 0.99347

- Predicted closing price for March 31, 2025: Approximately $221.08

The high R^2 score indicates that the model explains over 99% of closing prices variance and the prediction aligns well with the recent market data. Although the introduction of the 3-day SMA and volatility has greatly improved the model, I acknowledge that the volatility measure may impose sensitivity to short-term noise. This will be an area for continuous monitoring and further refinement if necessary.

5. Conclusion

Moving from a basic linear regression model to an enhanced version with both a 3-day SMA and a 3-day volatility indicator has led to a significant improvement in the predictive power of the model. The final model is robust, capturing both the trend and the short-term fluctuations effectively. The prediction of around $221.08 for March 31, 2025, closely aligns with the recent market trends. The next steps will involve preparing the model for deployment in a live environment, where it can be integrated into a minimal web application for real-time predictions and performance evaluation.