Choice of machine learning model for implementation 1: Long Short Term Memory (LSTM) network
LSTM was chosen to predict stock prices for NVDA for a few reasons:

First, LSTM are better at handling long-term dependencies in sequential data, as they can capture information from earlier time steps and remember it from extended periods of time. Stock prices often exhibit complex patterns and dependencies over time, and LSTMs can effectively model these relationships, due to their ability to remember information for long periods of time.

Second, LSTMs are less susceptible to the vanishing gradient problem that is commonly faced with RNNS, because of the use of a different kind of activation function, that allow them to control the flow of gradients through the network during backpropagation, preventing the gradients from becoming too small. allowing them to preserve information over long sequences. Given that the stock data gathered from NVDA is from a long time period, it is important that the machine learning model is able to effectively capture information from earlier time sets.

Finally, LSTMs can handle noisy or missing data better than other types of RNN, since they can selectively filter our noisy information using the forget gate, allowing LSTMs to remember important information while discarding irrelevant information. Given that financial data can be noisy and influenced by various external factors, LSTM is an effective model to handle noisy input and filter our irrelevant information while focusing on important trends.

Choice of ARIMA model for implementation 2: Why the ARIMA model of order (2, 1, 2) was chosen over others

I decided to use Auto ARIMA to automate the decision of fitting different models and deciding which is best. It suggested that the model of order (2, 1, 2) minimises the AIC score. Thus, the model suggests a lagged value of order 2 for autoregressive model, 1 for degree of differencing and 2 for moving-average model will give the lowest value for the AIC score, which means the model has captured all patterns that exist in the data and leave out the white noise (unpredictable part of the actual occurrence).