XGBOOST
What is XGBoost? XGBoost is a complex machine-learning algorithm that can help us understand our data and make better decisions. It is an implementation of gradient-boosting decision trees that push the limits of computing power for the boosted tree algorithms. It was built primarily for energizing machine learning model performance and speed.

How is Xgboost different?
the full name of xgboost is eXtreme Gradient Boosting Algorithim and so the difference between XGB and regular GradientBoosting is that XGB uses a regularization technique in it.

One of the main advantages of XGB over other machine learning algorithims is scalability. what does that mean? XGB was designed for efficient and scalable training of machine learning models, which in turn makes it an ideal model to utilize when using large datasets.

Another advantage, Customizable. XGB has a wide range of hyperparameters that can be adjusted to optimize performance. An example of those are 50 DMA & RSI

NEXT SLIDE

50 day moving average, what is it?
A 50 day moving average, aka "50 DMA", simply put, it is the average closing price of a stock over the previous 50 days. The primary reason why the 50-day moving average feature is popular is because it's a realistic and effective trend indicator in the stock market.

The 50-day moving average is a dividing line that shows the stocks' technical health on the upper line and not technically healthy on the lower line. Which means, the percentage of stocks above their 50-day moving average helps gauge the market's overall health/growth.
Several stock market traders use moving day averages to measure the profitable entry and exit points into specific "securities" and is one of the most widely respected technical indicators among investors.

How would you calculate this?
One can calculate the 50 DMA by adding up the closing prices from the last ten weeks and divide the sum by the total number of days that is 50 [(Day 1 + Day 2 + Day 3 + ... + Day 49 + Day 50)/50].

RSI(Relative Strength Index)
what is it?
The relative Strength Index (RSI) is a technical indicator that measures the momentum on a scale of 0 to 100 over a specific time period Which means, it helps identify whether a stock was overbought or oversold. Investors use the results with price indicators because it can potentially provide a higher quality stock buy or sell signal. What it does is measure the difference between two moving averages.

NEXT SLIDE

How do we calculate it?
RSI uses two parts for the calculation and it starts with the following formula:
Step 1
the Relative strength is determined by calculating the average gain divided by the average loss. The calculation uses a 14-day time frame. Keep in mind that losses are reported as positive values not negative values.

Step 2
RSI = 100 subtrated by 100 divided by 1+ RS

A stock is considered overbought when the RSI indicator reaches 70 or above and oversold when it drops below 30.

NEXT SLIDE

Once those were calculated, testing the features, here as you can see adding the 50 dma to the data spliting, an important parameter to set was "shuffle=False"

NEXT SLIDE

importing XGboost, making sure to add 50 DMA and the RSI, training the model, setting parameters and then initiating the model.

NEXT SLIDE

Testing the model here utilizing the predictions and of course, the mean squared error into consideration as well....

and then plotting it

NEXT SLIDE

Another feature that could be added would be a safety factor column that takes into consideration the difference between the actual and the predicted to create a safety net so that the predictions will come in lower than the actual (hopefully)

Moving on,

NEXT SLIDE

Testing XGB with the Netflix data

So again, starting off with the same strategy as before, adding the two features, 50 dma and RSI,

NEXT SLIDE

importing XGB, setting our features and target once again, the parameters are set to a max depth of 30 and learning rate of 0.3, this can be customizable based on the amount of days to compare.

NEXT SLIDE

once again, calculating the mean squarred error as a factor in testing our data

NEXT SLIDE

and then plotting it, as you can see there were some differences in the actual and the predicted but not as accurate as it was with amazon.

NEXT SLIDE handoff to justin