

## Predicting Costco Prices Using an LSTM

### Overview Of Project:

We aim to make an accurate predictor of NASDAQ: COST  $\text{Day}_t$  stock prices using predictors that can all be found on  $\text{Day}_{t-1}$ . We will train on five years of stock data, with daily periods, to get 1255 days of trading. We continue with some calculations for moving averages, RSI, and other indicators to better enhance our model. Finally, we will train an LSTM to calculate the  $\text{Day}_t$  price using historical data and  $\text{Day}_{t-1}$  data. Our goal is to maximize the return of our model on new data, and only allow it to buy and sell, without shorts or options.



### Single Day ( $\text{Day}_t$ ) Predictors:

- Open, Low, High, Close of the previous day
- 5, 10, and 20 day moving averages
  - May need to chop off the first twenty days of data because they won't have moving averages.
- Volume of the previous day
- Relative Strength Index - used for overbought and oversold indicators

- $RSICostco = 100 - (100 / (1 + (Nup / Ndown)))$  where Nup and Ndown are the number of up days and down days in the last fourteen days.
  - Oversold when  $<30$ , should be a buy indicator
  - Overbought when  $>70$ , should be a sell indicator
- Average True Range - Used for volatility
  - $TR = \text{Max}[(\text{Today High} - \text{Today Low}), \text{abs}(\text{Today high} - \text{yesterday close}), \text{abs}(\text{today low} - \text{yesterday close})]$
  - Calculate for everyday, then make a 10 day moving average
- Bollinger Band with RSI  
[https://www.youtube.com/watch?v=pCmJ8wsAS\\_w&ab\\_channel=TradingLab](https://www.youtube.com/watch?v=pCmJ8wsAS_w&ab_channel=TradingLab)
- 200 EMA MACD: [https://www.youtube.com/watch?v=rf\\_EQvubKlk&ab\\_channel=TradingLab](https://www.youtube.com/watch?v=rf_EQvubKlk&ab_channel=TradingLab)
  - Blue: 12 day moving average
  - Red: 26 day moving average
  - Green: 200-day moving average (EMA)
  - Price Action with key support
  - For visualization: histogram displaying the difference between Blue and Red
  - Conditions for predicting the upward trend:
    - Momentum in the market
    - Current shares are above the green line
    - When blue and red are below 0, blue crosses the red line
    - Share bounces off of the key support value, and all conditions above are satisfied
  - Conditions for predicting a downward trend are the versa of conditions for an upward trend:
    - When current shares are below the green line
    - When blue and red are above 0, and red crosses the blue line
- Stochastic Oscillator - captures momentum
  - $\%K = 100 * (\text{Recent Close} - \text{Lowest price of past 14 periods}) / (\text{Highest price last 14 periods})$ 
    - Known as the fast stochastic indicator
  - %D is same formula but using last three periods instead of 14
- Average Directional Index
  - Composed of a few elements
  - Plus Directional Movement is current high - previous high
  - Minus Directional Movement is previous low - current low

- $SMA = \text{Smoothed } TR_{t-1} - (\text{Smoothed } TR_{t-1}/n) + TR$
- $DI14+ = (SMA + DM14) / (ATR) * 100$
- $DI14- = (SMA - DM14) / (ATR) * 100$
- $DX = \text{abs}((DI14+) - (DI14-)) / \text{abs}((DI14+) + (DI14-)) * 100$

### **Long Term Predictors - historical data used for long term investing - will need to manually enter**

- Price / Earnings - Stock price relative to its earnings
  - gives us a good indicator if a security is undervalued or overvalues
- Price / Book - Stock price relative to its book value per share
  - $BVPS = (\text{Shareholder's Equity} - \text{Preferred Equity}) / \text{Total Outstanding Shares}$
  - $\text{Shareholder's Equity} = \text{Total Assets} - \text{Total Liabilities}$
  - Gives us a good indicator if a security is underpriced or overpriced
  - A value of 0.5 means that the book value is twice the current stock price
- Debt / Equity - Gives us an idea of how much debt a company is in and if they can pay it
  - A higher value is typically bad, with exceptions with high fixed cost industries
  - $D/E = (\text{Short Term Debt} + \text{Long Term Debt} + \text{Other fixed payments}) / \text{Shareholders Equity}$
- Free Cash Flow
  - Absolute number normally, could do FCF / ppS
- Price / Earnings / Growth
  - PEG ratio uses expected growth, which may pose a challenge in data gathering
  - $PEG = P/E / \text{Expected EPS Growth Rate}$
  - $EEPSGR = (EPSt / EPSt-1) - 1$
  - $EPSt = (\text{Net Income} - \text{Preferred Dividends}) / (\text{Average Number of Shares Outstanding})$
- Could use EBITDA if able to find a good source online
  - Earnings + Interest + Income Tax + Depreciation/Amortization
  - <https://rowzero.io/blog/yfinance>
    - Looks like yfinance has quarterly data on company fundamentals
    - Does not have historical financials beyond 7 quarters ago.
  - <https://www.macrotrends.net/stocks/charts/COST/costco>
  - Just manually made a fuckton of data
  - Sec-edgar returns XML docs, could parse and get financials
- Could get VIX if yfinance has the correct calls/puts

## List of Columns:

Open, High, Low, Close, Volume, RSI, ATR, Bol1, Bol2, MACD, StoOsc%K, StoOsc%D, ADI, P/E, BVPS, P/B, D/E, FCF/ppS, PEG, EBITDA

## Workflow

1. Gather Historical Data from yfinance python package
2. Clean the data
3. Add our predictors to our dataframe
  - a. TODO: Combine Short and Long Term data into a CSV
  - b. Future Column that is the close of the next day
4. Normalize the data and preprocess it
  - a. When to use a minmax vs a standardscaler?
  - b. Train, Validation, Test
  - c. Segmented into sets of 103 trading days: 80 training days, 13 validation days, and 10 testing days
  - d. Feature Selection**
5. Train the LSTM and make predictions
  - a. Need to make predictions and a better graphical representation
    - i. IDEA: Predict 5 trading days in the future. If it is up, then we buy one share, keep for 5 days, and sell at close after 5 days, no matter the price. If the value is down, we do nothing.
6. Analysis:
  - a. Output will need to be inverse scaled to get a human readable numerical output
  - b. Real price versus predicted price
  - c. Models' potential gains versus holding from buy time to sell time
  - d. Make a GOOD latex writeup on the topic and put it on my GitHub. Could also linkedin farm
7. TODO
  - a. Code Review
  - b. Feature Selection
  - c. LATEX writeup
8. Future Work
  - a. Predict trendedness of the data? Or something that does not fall with day t-1 errors?

- b. Find value investments using company financials? Given an API is found
  - i. Might need to make models to predict growth, then use those as inputs to predict value of the investment
- c. Application to intraday trading?
- d. Apply to larger variety of stocks

### **Why LSTMs don't work for stock market prediction**

- Randomness in the stock market
  - There is too much uncertainty inherent in the stock market for a model to accurately fit to a certain stock or any set of stocks, and it will never be able to predict a future price
  - My model predicts 5 days in advance, and using our predictors, it guesses if the price is going to go up or down, and it holds for those 5 days no matter what
  - This means that it sometimes gets super lucky in its gains, but if it notices even a slight downtrend in future activity, it says not to buy, so it mitigates some of the risk
  - External factors play a huge role in statistics
  - Costco's repetitive data Q3 of every costco year in a huge increase in earnings, my model could be capitalizing on those earnings.

### **Sources**

#### [Stanford CS230 Quantitative Analysis using LSTM](#)

1. Use an attention mechanism for better results
2. They only used 20 days of past history to predict the next day
3. They used the top 10 stocks for stability
4. They used company financial statements