DS105 Project Proposal
Predicting Stock Prices

Edwin Wan

PROBLEM STATEMENT

Forecast stock prices up to 1 week ahead to see how forecasted prices will interact with resistance/support to decide on options strategy.
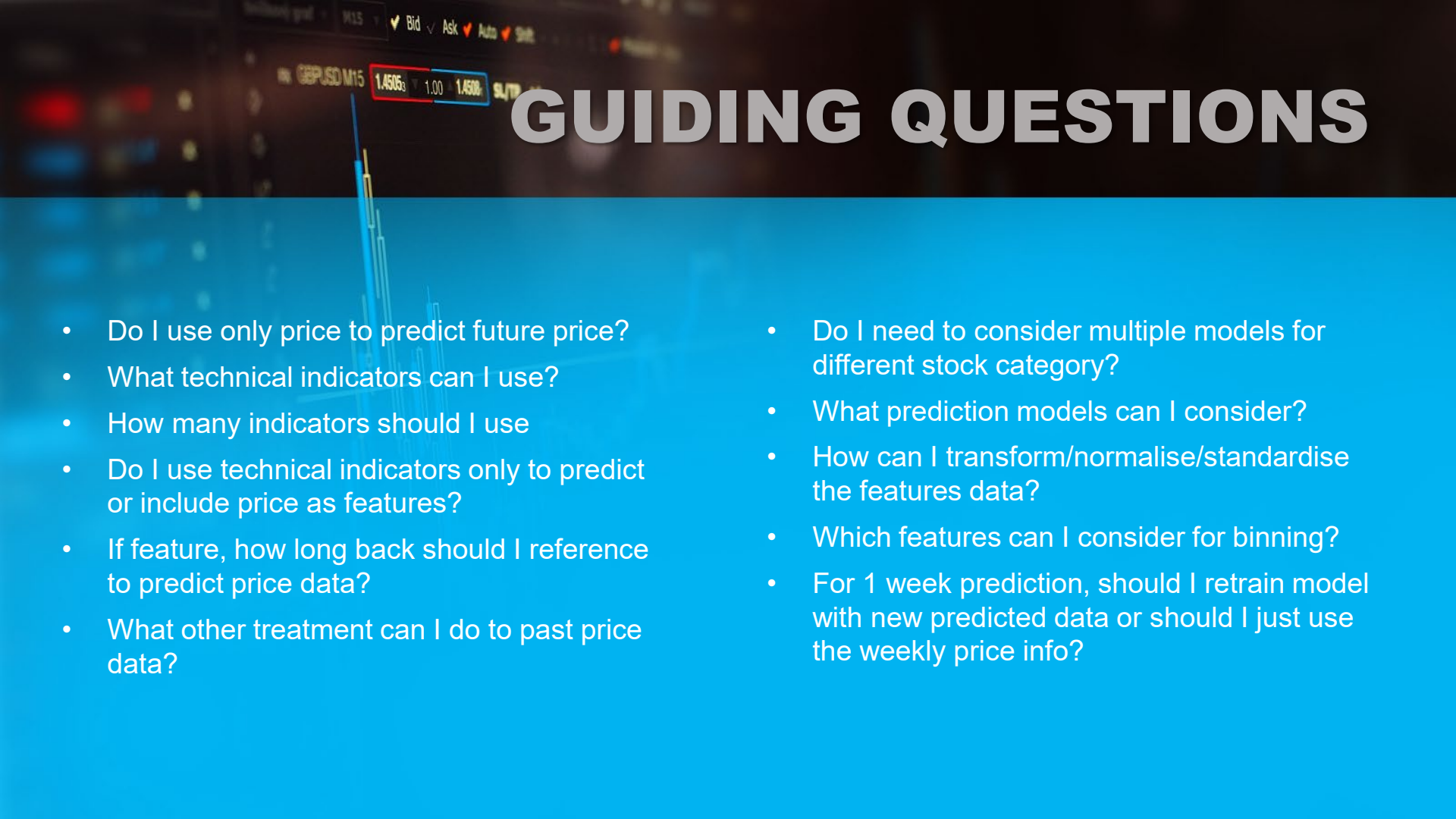
BACKGROUND

To decide which options strategy to employ, there few criteria to meet such as the company fundamentals, overall trend of stock, implied volatility and IV percentile etc.

Price action and its interaction with support and resistance line is also one of the criteria to decide the timing of strategy.

Focus for this ML project is to attempt to predict future prices between 3 days to a week if possible

GUIDING QUESTIONS

- Do I use only price to predict future price?

- What technical indicators can I use?

- How many indicators should I use

- Do I use technical indicators only to predict or include price as features?

- If feature, how long back should I reference to predict price data?

- What other treatment can I do to past price data?

- Do I need to consider multiple models for different stock category?

- What prediction models can I consider?

- How can I transform/normalise/standardise the features data?

- Which features can I consider for binning?

- For 1 week prediction, should I retrain model with new predicted data or should I just use the weekly price info?

Prediction Plan

- Predict future price using past stock prices, MACD values, RSI values

Moving Avg Convergence Divergence (MACD)

- The difference between the MACD line and signal line.
- It shows the momentum of the price in a certain direction

Relative Strength Index (RSI)

- Measures magnitude of recent price changes
- Range from 0 to 100
- Indicates oversold or overbought

Prediction Plan

Money Flow Index (MFI)

- Similar to RSI
- Takes into account trading volume

Bollinger Bands (BB)

- Takes into account the standard deviation from the mean price
- Offer unique insights into price and volatility.

Steps for Prediction

- Gather and consolidate required data from 1 Jan 2004 to 31 Mar 2022
- Process the data – Treat null values, transform data to be used as features
- Bin RSI/MFI data to 4 categories
    - > 70, < 30 , 30 to 50, 50 to 70
- Train-Validation-Test split by slicing
- Standardise MACD, RSI and % change of price or Log transform price data
- Test with different models to see which has better performance
    - LSTM (Neural Network)
    - Time Series Forecasting (Auto TS)
    - Linear Regression
- Metrics
    - RSME and R-squared
- Plot for visual comparison
- Test model with test data
- Test model and observe on actual market trading

Datasets

- 5 to 6 features for Prediction
- Price data from Yahoo Finance API
- Focus on **<Adj Close>** price data
- May convert to **<Pct Change in Price>**

Date	Open	High	Low	Close	Adj Close	Volume
2003-12-31	40.125000	40.264999	40.029999	40.215000	26.606417	8070200
2004-01-02	39.875000	40.215000	39.455002	39.544998	26.163160	16897000
2004-01-05	39.660000	39.799999	39.360001	39.660000	26.239231	14535400
2004-01-06	39.560001	39.695000	39.400002	39.595001	26.196226	15083600
2004-01-07	39.525002	39.584999	39.404999	39.505001	26.136683	13346200
...	...	...	...	...	...	...
2022-03-25	43.480000	44.259998	43.330002	43.730000	43.730000	38968100
2022-03-28	43.709999	43.750000	42.830002	43.549999	43.549999	37428600
2022-03-29	44.250000	44.389999	43.110001	43.439999	43.439999	46355800
2022-03-30	43.439999	43.650002	42.750000	43.000000	43.000000	36601800
2022-03-31	42.840000	42.889999	41.200001	41.220001	41.220001	67902500

4595 rows × 6 columns

Datasets

Date	rsi	MACD_12_26	Adj Close
2004-02-06	66.525397	0.191077	27.377193
2004-02-09	55.837839	0.181767	27.019922
2004-02-10	55.306162	0.170817	27.000071
2004-02-11	57.964094	0.170056	27.122480
2004-02-12	54.742436	0.158549	27.009996
...	...	...	...
2022-03-25	52.738906	-0.323436	43.730000
2022-03-28	51.772619	-0.242699	43.549999
2022-03-29	51.155774	-0.185452	43.439999
2022-03-30	48.658447	-0.173587	43.000000
2022-03-31	40.124607	-0.304307	41.220001

- Calculate MACD and RSI values as one of the 3 features for prediction

- Use python TA-lib

- May also consider including other indicators such as:
  - Bollinger's Band (Volatility)
  - Money Flow Index (Volume)

Challenges

- May need to create different model for different sector stocks (eg: Tech stocks, FMCG, Auto, Pharma)
- Which stock/ETF to use to aggregate for sector.
- Need to bring in new indicators and consider how to treat them (Do we bin and categorise them or maintain as continuous)
  - Bollinger band to take into account volatility
- Model only predicts one day in advance
  - Need to formulate how to get prediction for 7 days ahead
- Predictions do not take into account sentiments, news, black swan events

Comments from Sijie:

- To think about: if I will be adding other features to the time series model
  - Need to explore ARIMA as time series can only compare prices to its own lag prices. May not be able to include other features
- What stocks or ETFs to use for the modelling
  - visually compared to see how similar they are.
  - May also want to narrow down to few stocks that I will trade on