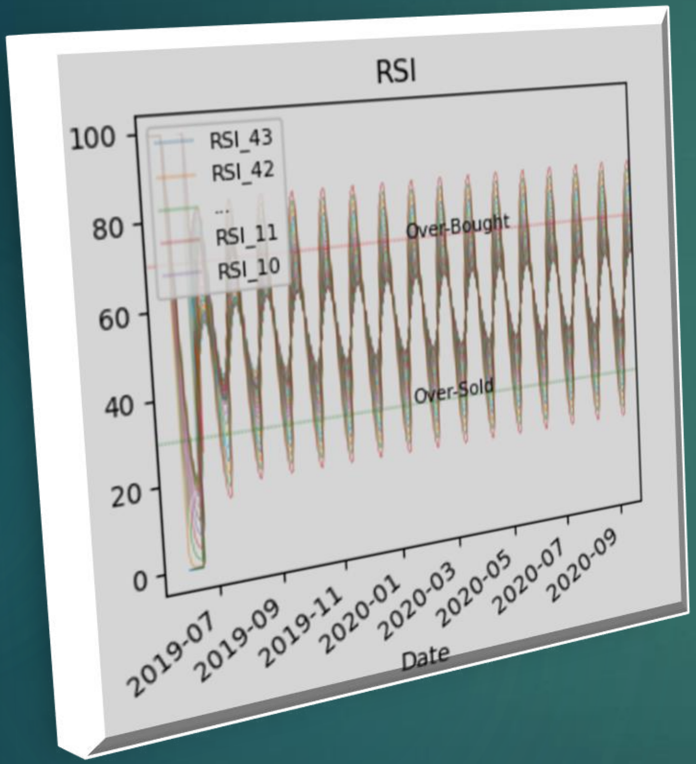
Project Report

Investment R_{deep}

Stock Recommender

Group 3
Chng Yan Hao
A0024023A

Contents


RSI
100
80
60
40
20
0
RSI_43
RSI_42
...
RSI_11
RSI_10
Over-Bought
Over-Sold
2019-07
2019-09
2019-11
2020-01
2020-03
2020-05
2020-07
2020-09
Date

Executive Summary

At this point, can't it go higher?

[1]Image <http://observationsinanundemocraticworld.blogspot.com/2017/06/buy-lowsell-high.html>

At this point, can't it go lower?

Can we truly "Buy Low and Sell High" with this application? There are trillions of dollars sitting in stock market and a series of good investments can change one's financial well-being for life. Just stock analysis alone, there are hundreds of ways to perform on at one single stock. So which is right which is wrong when one result is against another?

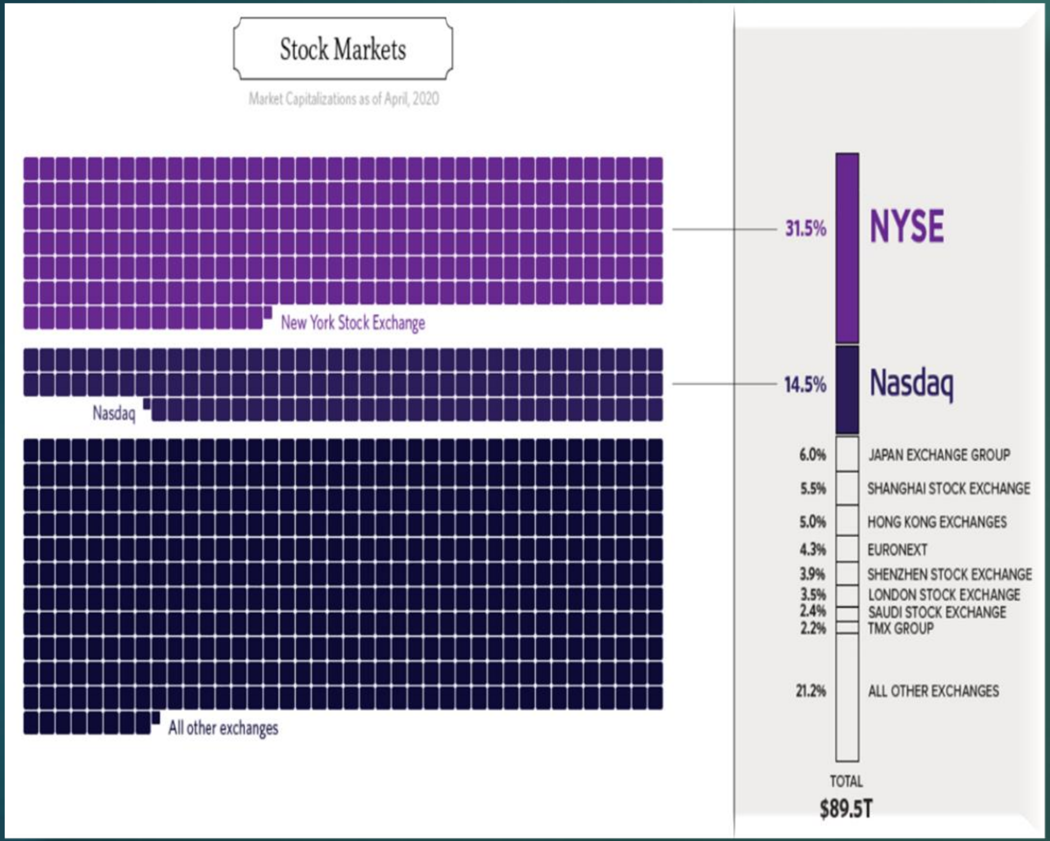
Many times, a lot speculation, pressure from buying and selling, momentum and news are affecting the stock prices. Many of those, we have no control and calculation over.

In this application, we will just zoom into facts, facts derived on daily closing value in hope to find patterns.

If you see a series 1,2,3,4,5,4,3,2,1,2, will you want to believe the next is 3 and next next is 4? Will others believe this too? Will you take action to buy if you are 'quite certain' the pattern is correct? The application can only answer at best the first line.

So can we really "Buy Low and Sell High" with this application? I'll have to leave this for the user to find out.

Business Problem Background



[2]All of the World's Money and Markets in One Visualization <https://www.visualcapitalist.com/all-of-the-worlds-money-and-markets-in-one-visualization-2020/>

Stock market is a money bucket worth (USD 89.5T) with different companies and individual price tags.

You are effectively owning a part of the company when you buy stock shares.

1) Looking for bargains? (Value Investing)

2) Buy low sell high? (Capital Growth)

3) If you are already invested, who's recommendation to trust?

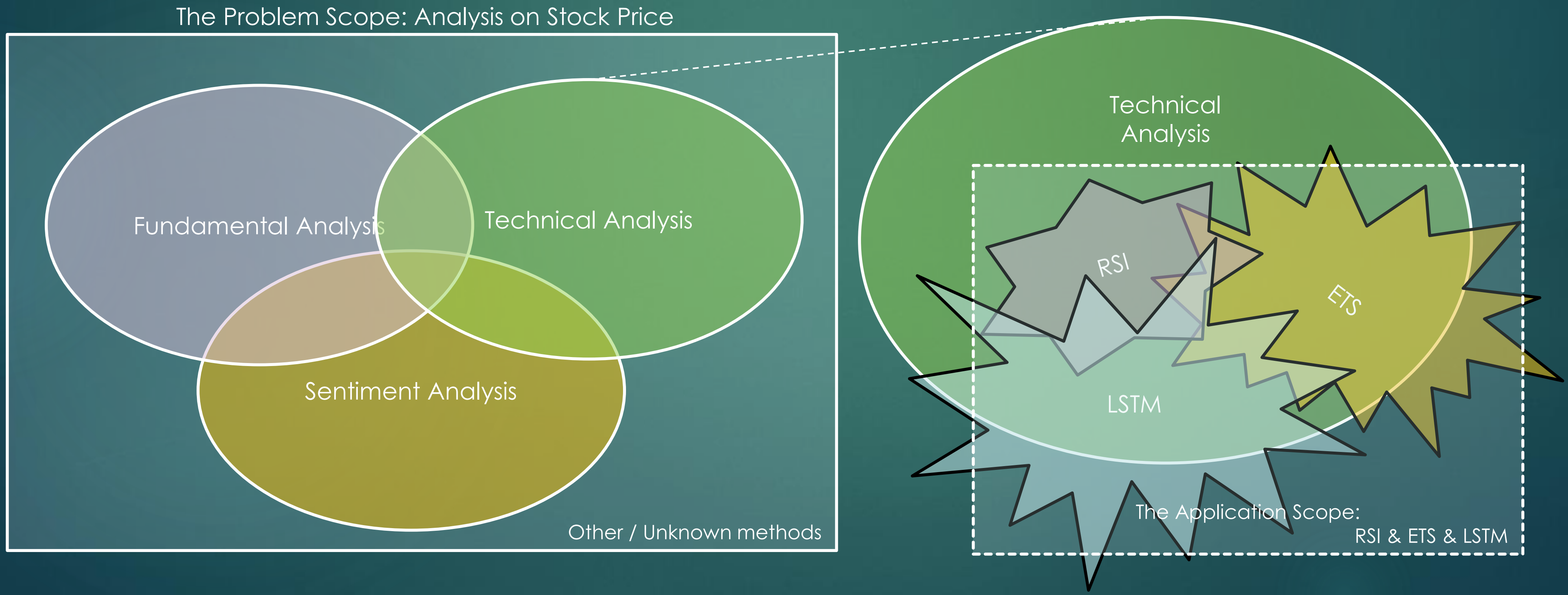
4) When is a good time to buy?

5) Where to start?

Business Problem Background

Buy low sell high!

1) Potential investors want a quick start - answer to 'when can I buy'.
2) There are too many scattered ways, methods are not transparent to trust.
3) A lot of good applications required a premium to access or upfront investments.
4) Even the most experienced stock reminiscer cannot be totally trusted.



Project Objectives

Key objectives:

1) Find earning opportunities in stock investing.
2) To allow users to explore more intensively with lower entry barriers.
3) Less hassle on data, interpretation and its analysis.
4) To provide an additional option, opinion and platform for comparison. i.e. against investment platforms, friends, brokers and even robo-advisors.

Technical Objectives:

1) Supervised learning / Unsupervised learning
2) Machine learning / Deep Learning
3) Hybrid machine learning / Ensemble approach
4) Intelligent Sensing / Sense Making

Project Solution (Part 0)

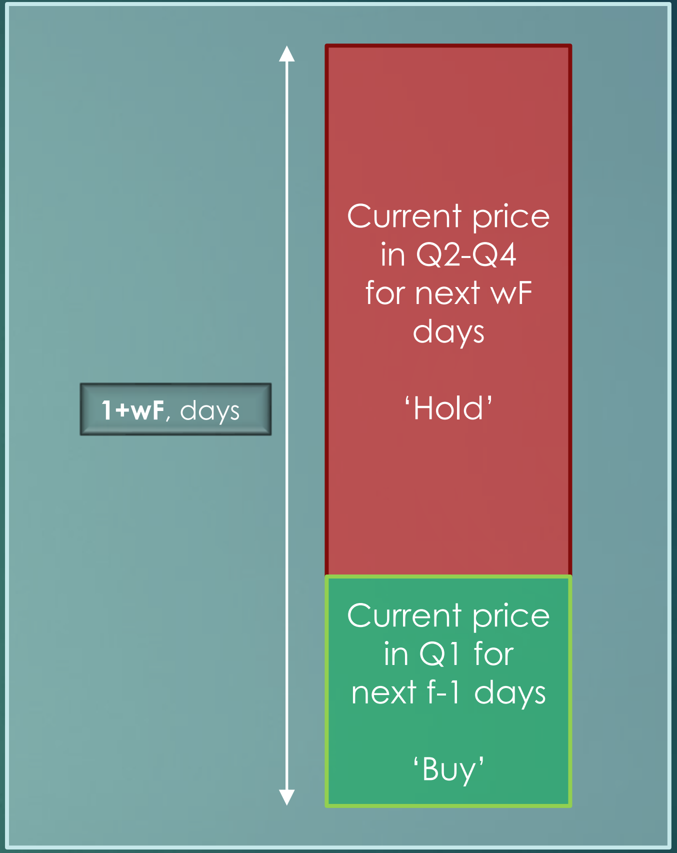
Get Data, define Data Rows
1) dR, 504 (2 stock years) <= dR <= 1260 (5 stock years)
2) Only required 'Date' and daily 'Close' values.
3) Application already setup with yfinance (Python), standalone option available using .csv file

Define 'Buy' & 'Hold' used (the labels for historical data)
1) A day is labelled 'Buy' if its current day price is <= 25th percentile in the next wF days.
2) i.e. Price($) today is 2, next 9 days is {1,3,4,5,6,7,8,9}.
3) Today is at 20th percentile hence 'Buy'
4) i.e. Price($) today is 2, next 4 days is {1,3,4,5,}.
5) Today is at 40th percentile hence 'Hold'

Define Window Length (wL) aka 'look back' period.
1) No. of days, 15 (3 stock weeks) <= wL <= 63 (3 stock months).
2) wL will be used in all three models.

Define Window Future (wF) no of days to forecast.
1) No. of days, 5 (1 stock week) <= wF <= 21 (1 stock month)
2) wF has to be <= wL, making thing more meaningful.

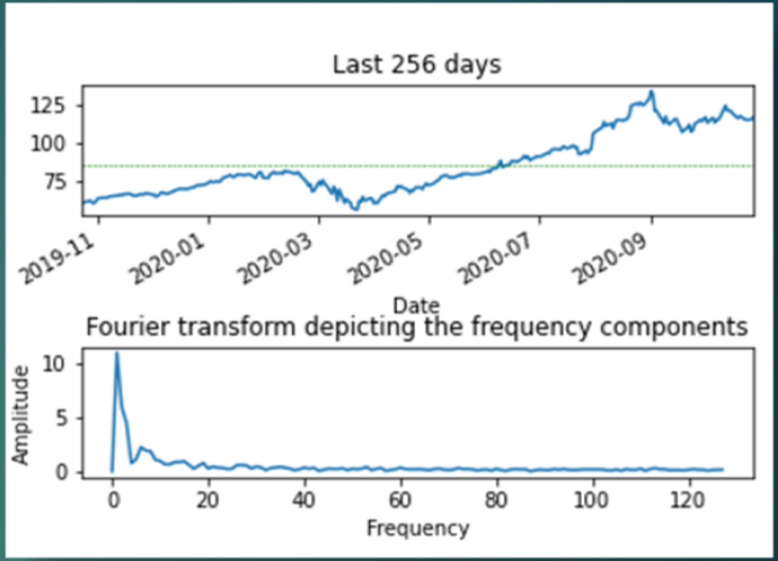
The ranges have been carefully chosen, taking into consideration from both compatibility to domain and speed of data processing. Using 15 minutes as guide.



Project Solution (Part 1)

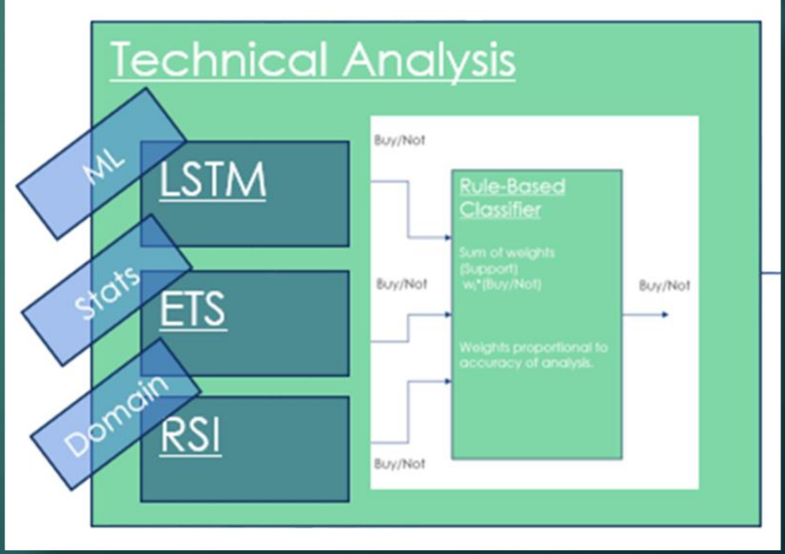
Fast Fourier Transformation (Before data is fed to models)

1) One of the challenges is how long to 'look back'.
2) In this project, let's call it Window Length, wL.
3) Fast Fourier Transformation is chosen to do a quick estimation on Window Length, wL.
4) The frequency within the last 256 days is computed. The first suitable frequency is read to return a suggested wL.
5) 1/freq = period, largest period in 15 <= period <= 63.
6) This feature can be switched off.
7) It was preferred over Fourier Transformation as even a full Fourier Transform remains as a guess.



Overall: Hybrid model (after all models are run)

1) A simple rule-based model with weight x support from 3 default models.
2) As different models are from different approaches, all models will be considered so to keep recommendation diversified. (regardless of whether support is enough)
3) Each individual model represent some insights different from another.



Project Solution (Part 2)

RSI (Relative Strength Index – more domain related)

1) Stock Technical Analysis can be done from various angles.
2) Approaches include: Simple/Exponential Moving Averages, Buying/Selling Pressure, Relative Strength Index (RSI) …
3) Relative Strength Index of Stock A requires analyst to imagine a 'look back' period. i.e. RSI_(look back)
4) RSI = 100 - (100 / (1 + RS)), RS simply said is average up / average down in the number of look back days.
5) RSI <=30 or RSI >= 70 in practice is seldom for a stable stock.

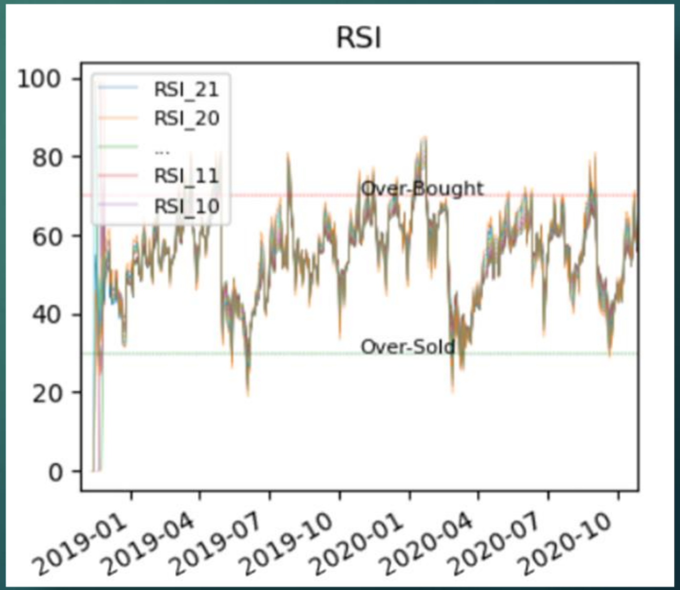
KEY TAKEAWAYS

- The relative strength index (RSI) is a popular momentum oscillator developed in 1978.
- The RSI provides technical traders signals about bullish and bearish price momentum, and it is often plotted beneath the graph of an asset's price.
- An asset is usually considered overbought when the RSI is above 70% and oversold when it is below 30%.

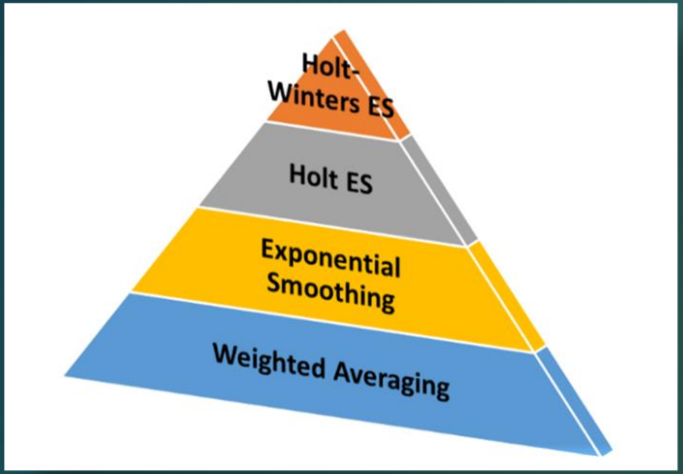
[3]Relative Strength Index (RSI) <https://www.investopedia.com/terms/r/rsi.asp>

Approach

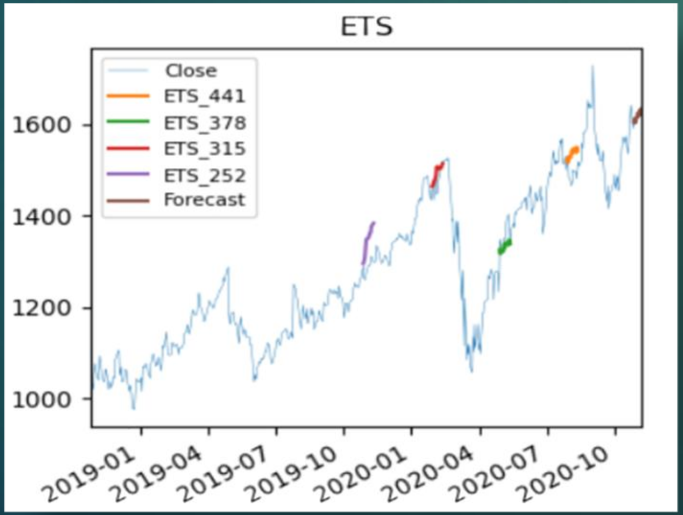
1) Investment Rdeep will take RSI from range wF – wL.
2) Window future, wF will be the number of days we want to forecast ahead.
3) For all RSIs, if RSI <= 30, return recommendation 'Buy' else 'Hold'
4) These chunks of recommendations is fed into MLP and trained against our 'buy definition'. (Supervised Learning)



Project Solution (Part 3)



[4]Holt-Winters Exponential Smoothing
<https://towardsdatascience.com/holt-winters-exponential-smoothing-d703072c0572>



ETS (Exponential Smoothing – more statatistics-like)

1) We already know approaches include: Simple/Exponential Moving Averages, Buying/Selling Pressure, Relative Strength Index (RSI) …
2) For this part of the solution, we will let the Holt Winter's Exponential Smoothing algorithm handle the Exponential Moving Averages.

Approach

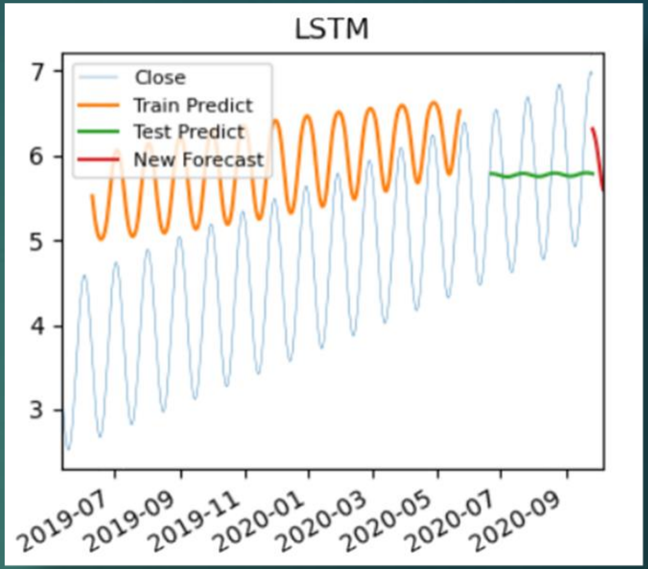
1) The settings for trend and seasonal will be additive, with season period = Window Length, wL.
2) More than 100 ETS models will be created, each used to forecast Window Future, wF days ahead.
3) The last day of each model is taken with wF days of forecast.
4) If last day that fall within 25 percentile, 'Buy'. Same as our 'buy definition'
5) All the ETS models are tested whether it made the correct forecast. (Unsupervised Learning)

Project Solution (Part 4)

LSTM (Long Short-Term Memory – Deep Learning)

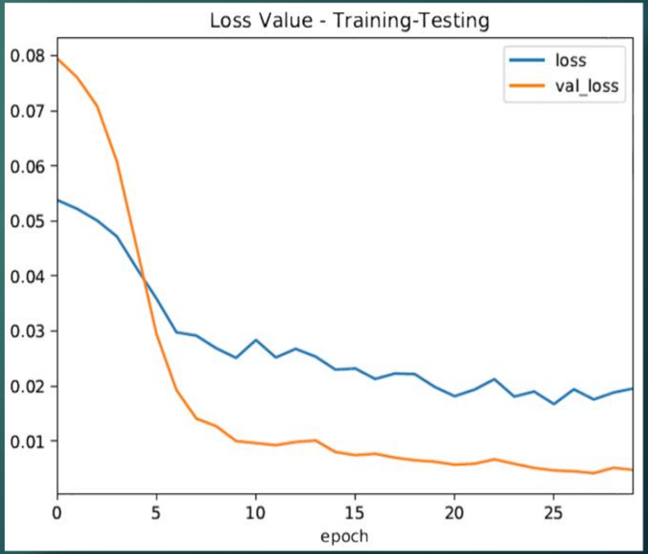
1) LSTM is an artificial recurrent neural network architect used in deep learning.
2) LSTM networks are well-suited to classifying, processing and making predictions based on time series data.[5]
3) This method is chosen to complement the lacking in the previous two models, primarily to remember pattern over an arbitrary period.

[5]Long short-term memory <https://en.wikipedia.org/wiki/Long_short-term_memory>



Approach

1) 2-layer stacked LSTM.
2) 30 epochs (consider time vs vast possibilities)
3) Look back = Window Length.
4) A train-test for illustration purpose as it will be computational intensive to run many forwarding validating models.
5) Support is derived from test predict.
6) Purpose is to catch the short term pattern.
7) Final model consist of max data.



Solution Design

1. **Data Input**

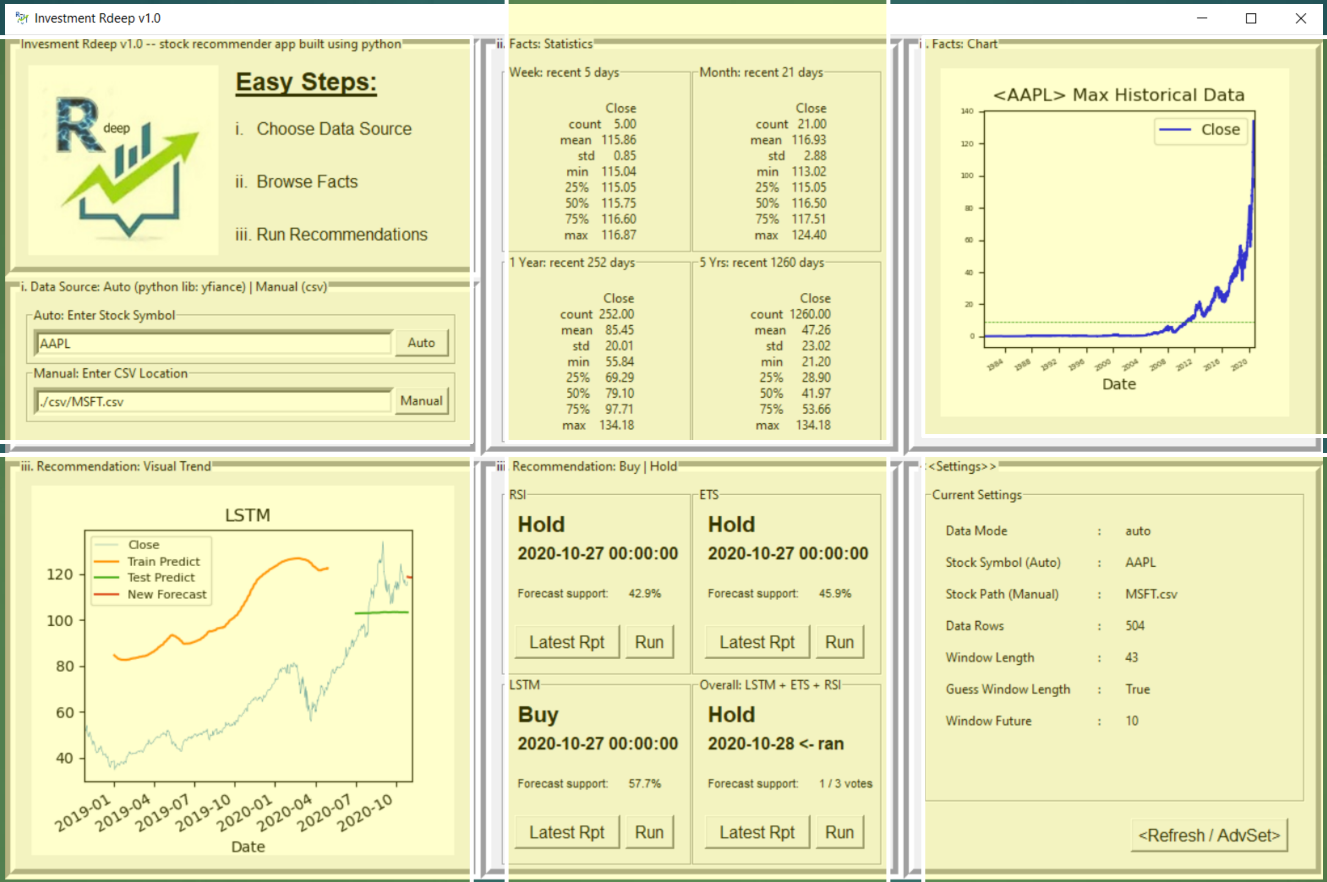
Auto:
Data taken from yfinance (Python)
As good as API.

Manual:
Open <data>.csv

2. Statistics

3. Max Data Chart

Close data in blue
Mean line in green



4. Last Model Visual (RSI, ETS & LSTM)

Refreshes once a new model is ran.

5. Run Panel

Run model

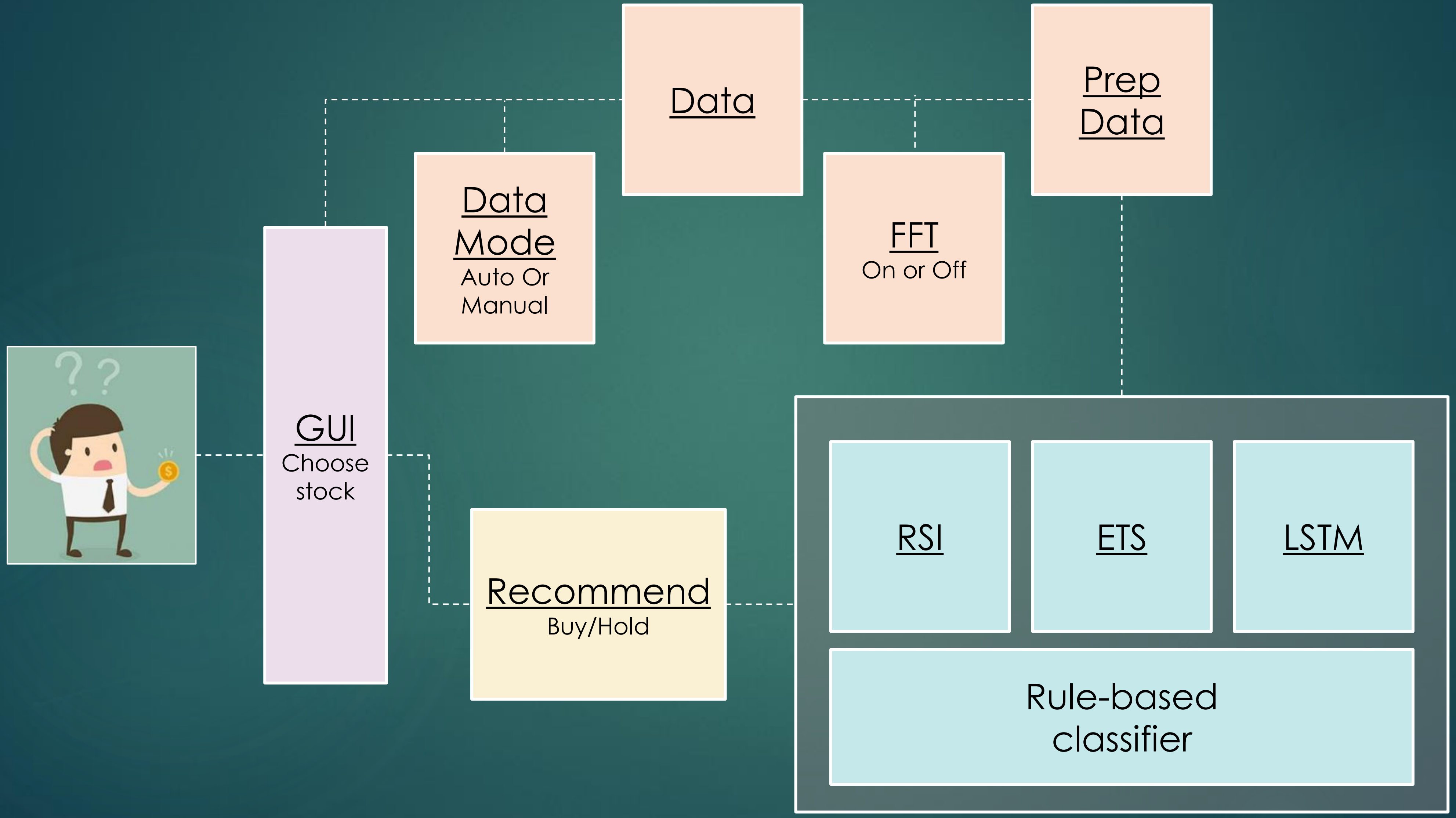
Lastest Rpt:
Open report .pdf

6. Settings Panel

Displays current setting

Allow changes: dR, wL, wF,
Toggle FFT: on or off on wL

Project Implementation (Part 1)



Project Implementation (Part 2)

Filename	Main function	Return
GUI	All user input and outputs. Final rule-base classifier.	All visual outputs: Historical Statistics Historical Chart Visual Trend of Model Recommendation Settings
DataSource	Check for 'Close' & 'Date' Columns. Load data into memory.	Initial data or throws error
DataPrep	Extract dataframe subsets and require parameters.	Parameters and first prepared data
cDB	Create database to store parameters. Query database to return entries.	Parameters: auto/manual data mode stock symbol stock filepath dR, wL,wF, FFT(True/False)
FFT	Apply Fast Fourier Transform to data.	Period calculated from frequency
RSI	Calls RSI calculations.	Charts Reports
ETS	Calls ETS	Charts Reports
LSTM	Calls LSTM and its architecture	Charts Reports (include loss and model summary)
Speech	Some greetings and instructions	Sound

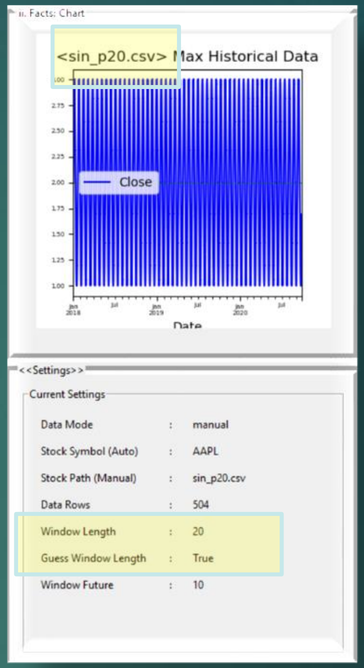
Project Performance & Validation (1)

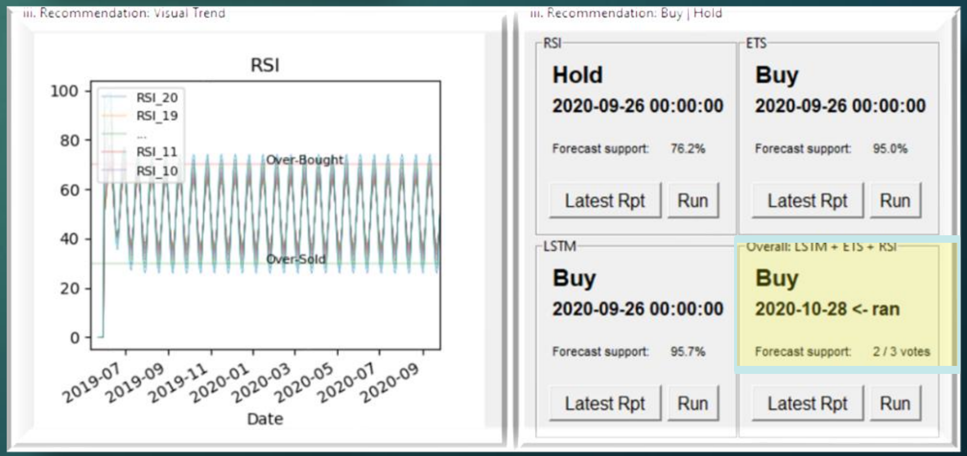
FFT, Fast Fourier Transformation (before data loaded to models)

1) The period guess is consistently close match to original wave.
2) Estimated performance of FFT has error <=3 periods (converted from frequency) caused due to no actual sinusoid pattern in real stock data and slight rounding during calculation.
3) Overall, it is a good estimate to guess Window Length.

Final rule-based classifier

1) Model recommendation ‘Buy’ as 1 and ‘Hold’ as 0
2) Final recommendation = Σ Recommendation x Support(%) / 3
3) Mainly produce ‘Hold’ (<0.5) results due to:
4) Demanding RSI, stable stocks hardly touches domain standard of under 30. Thus easily 1 vote out.
5) Dealing with radical stocks, models greatly affected and hover with supports at slightly above 50%.
6) Easily 1 vote to decline ‘Buy’.
7) Overall is a ‘very cautious’ recommender.

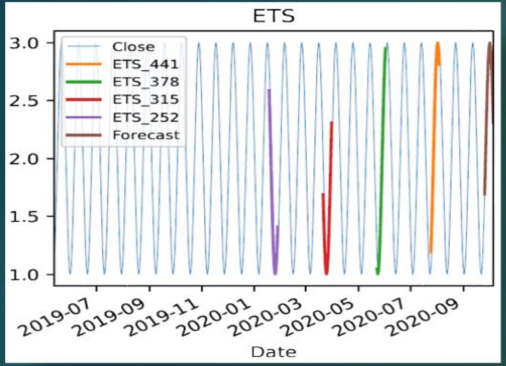


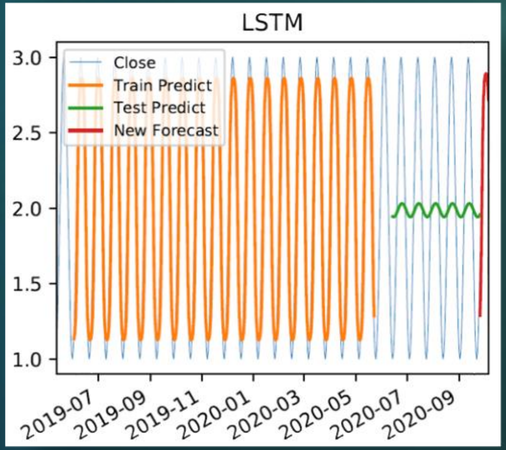


Investment Rdeep

Project Performance & Validation (2)

RSI <= 30	Pred: Hold	Pred: Buy
True: Hold	78	0
True: Buy	30	18
	Accuracy:	76.2%
	Mean-Squared-Error:	0.238





<u>RSI</u>

1) It is a very tough indicator for stable stock.
2) Most stable stock price do not fall under 30.
3) High false negatives in relation to our ‘Buy’ definition.

<u>ETS</u>

1) Very sensitive, to ups and downs near to last rows. There is little chance of it forecasting ‘U’ shape if season period is out.
2) Usually above 60% accuracy of correct guesses over more than hundreds of individual models’ forecast on non-turning points.

<u>LSTM</u>

1) Slowest individual model among the three hence offset by rerunning less.
2) Most accurate in catching the ‘shape’.
3) Requires refreshing the model periodically to get best forecast.
4) Training – Test sets are only for illustration purpose.
5) Even so, the long test sets still performs fairly well over long period of unseen data.

** Reports are available after each runs, feel free to trust / reject them.
** It is also good when individual model is against each other, hence warning to user there is no clear signal to ‘Buy’.

Investment Rdeep

Project Conclusions: Findings & Recommendation

First and foremost, the recommender in no ways is liable to any loss incurred during actual stock investment using the application.

Models and techniques:
FFT gave a great Window Length guess, despite its true strength to split waves and find frequency has greater use in other areas.
RSI will trigger usually when stock hits bottom, excellent to use in conjunction with other models.
ETS relies heavily on pattern in the designated seasonal period, else its performance is just close to any linear regressor.
LSTM is the strongest all rounder among the models.
The major drawback is that it is slower and parallel processing is not an option.

Overall:
The models has minimal problem working on synthetic sinusoid related data.
Stock prices may have some patterns, but not all signals from the model will flag at the same time.
Stock prices requires monitoring, the application has limited resource to create its own database and perform deep learning onto its own daily results.
Nonetheless, for any one stock recommendation, the application has sufficed its project objectives.

-- End of Report --