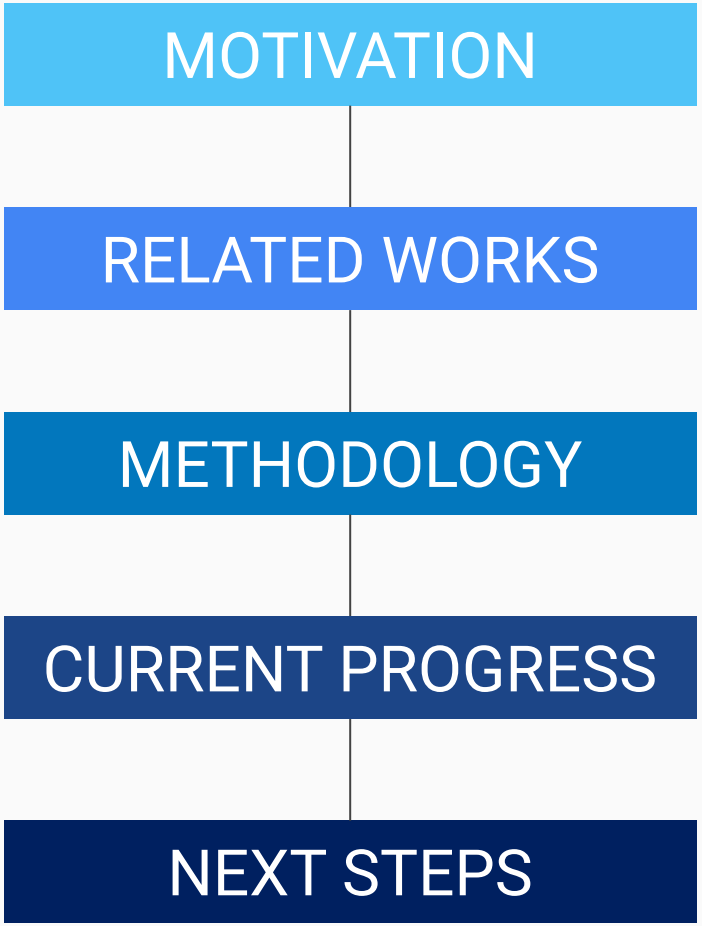
FINANCIAL DATA FORECASTER

Shubhankar Agrawal (Shub)
Karan Mahajan

Supervisor: Dr Beta Yip

23 January 2020

OUTLINE


MOTIVATION
RELATED WORKS
METHODOLOGY
CURRENT PROGRESS
NEXT STEPS

WHY FORECAST?

WHY FORECAST

INFER

- To obtain analysis of trends, seasonality and other factors for a better understanding of past data.
- To deduce the features, and by what extent, they influence the data.

PREDICT

- To obtain future information on the data including the direction of movement or the actual values.
- To act on the data to create impactful changes in strategies in markets.

OBJECTIVES

OBJECTIVES

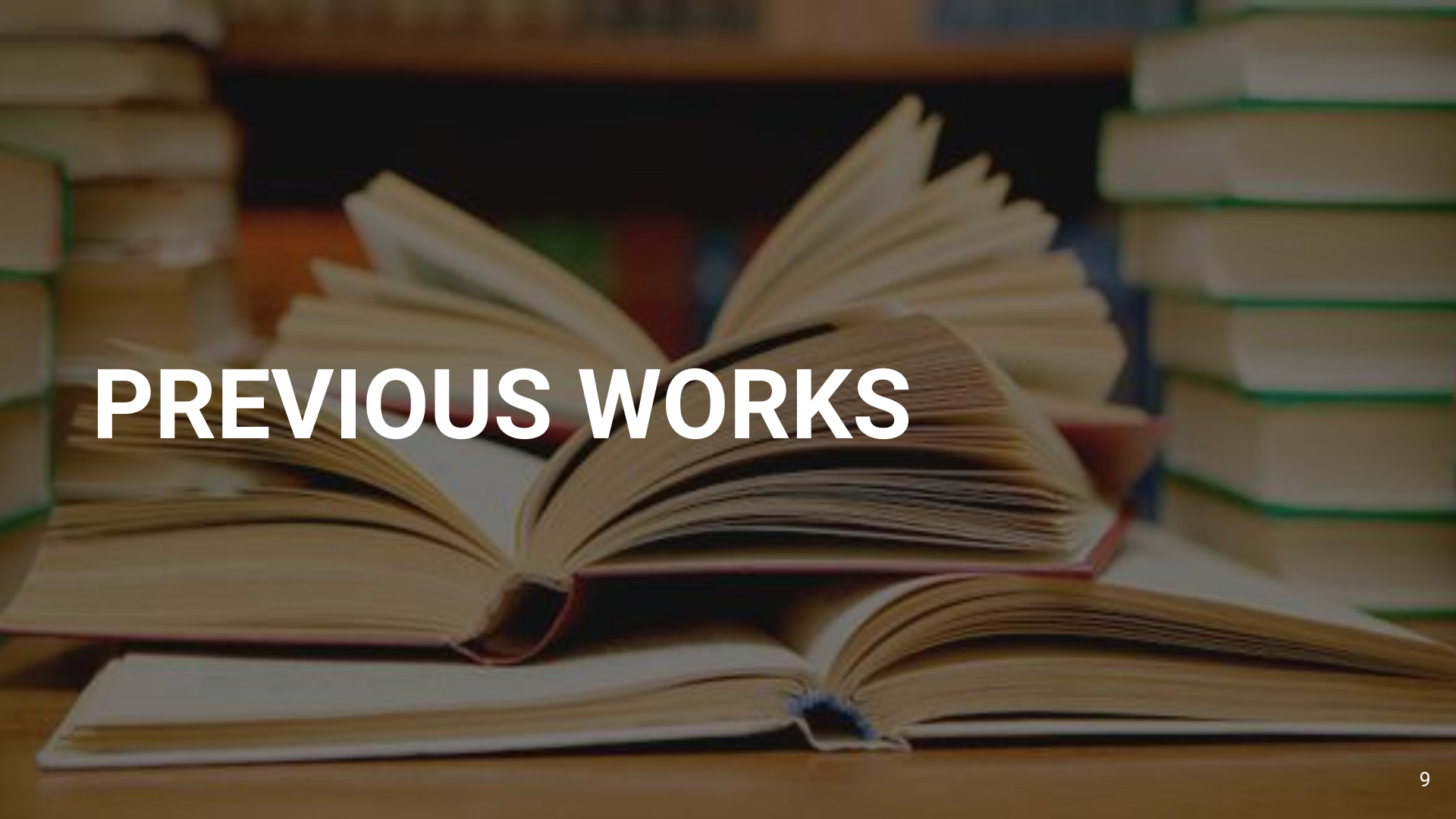
- Analyse time series of historical financial data and determine the combinations of most effective indicators as input.
- Identify trends and relationships between financial data of different markets.
- Identify relationships between the time series' of financial data, market indicators and non-numerical data and achieve a favourable accuracy rate for financial predictions.

POTENTIAL USES

Predictions can be used to analyse futures of economies and can act as an early warning systems for economic downturns.

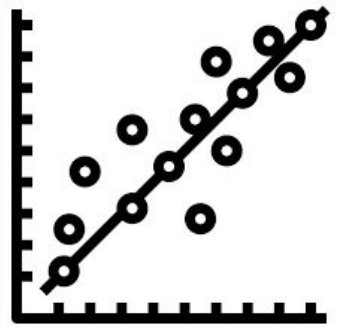
POTENTIAL USES

Predictions can also help businesses make strategic and operational decisions based on the insights from our study



PREVIOUS WORKS

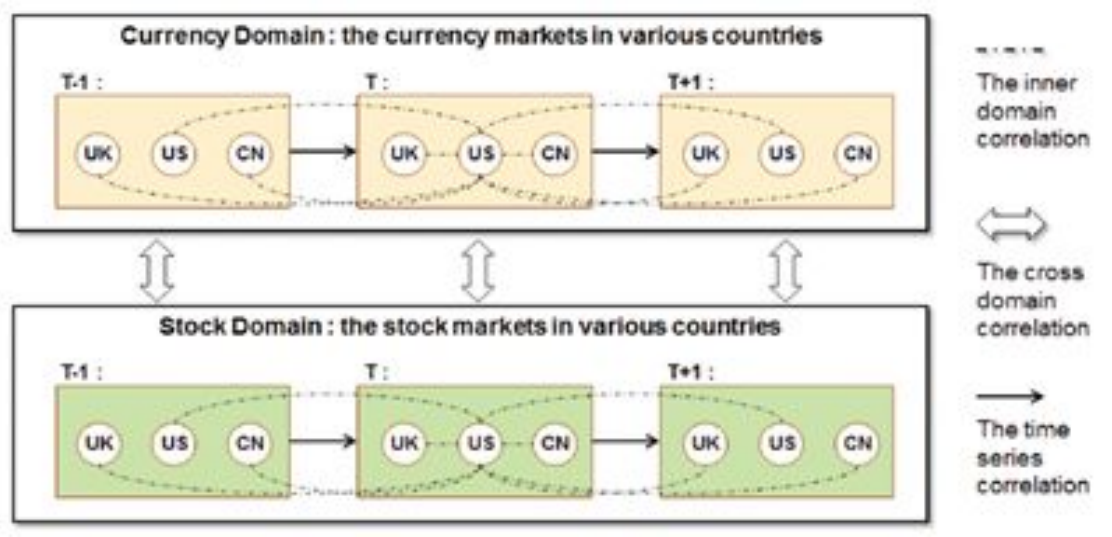
PAST WORKS



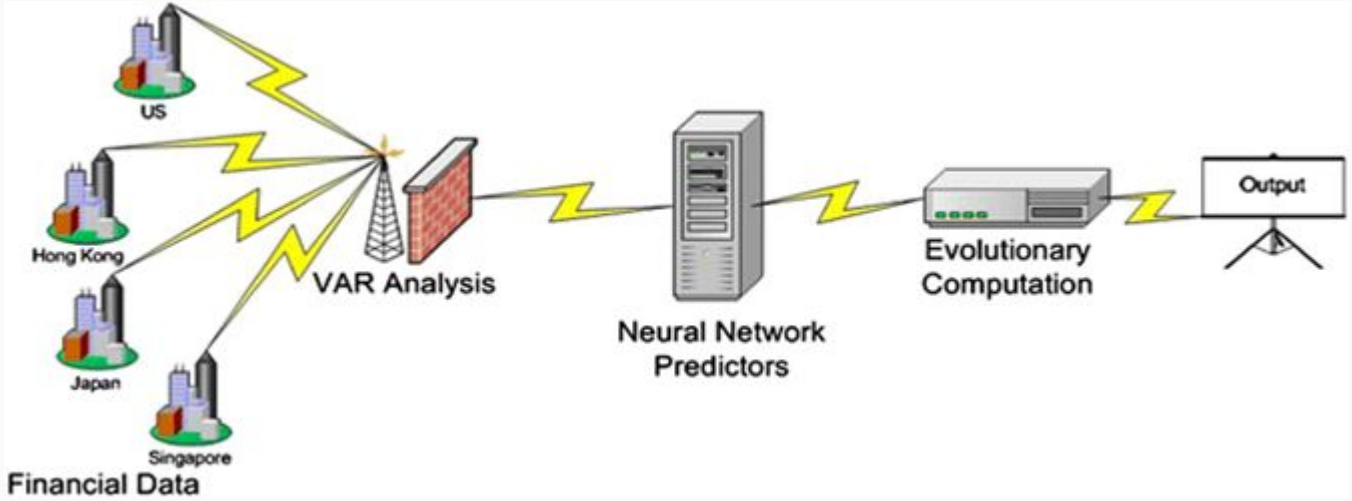
PAST WORKS : Machine Learning

- Neural Networks
- Random Forest
- Support Vector Machines
- Autoregressive Moving Average
- Genetic Algorithms
- Multiple Kernel Learning
- K-means Clustering

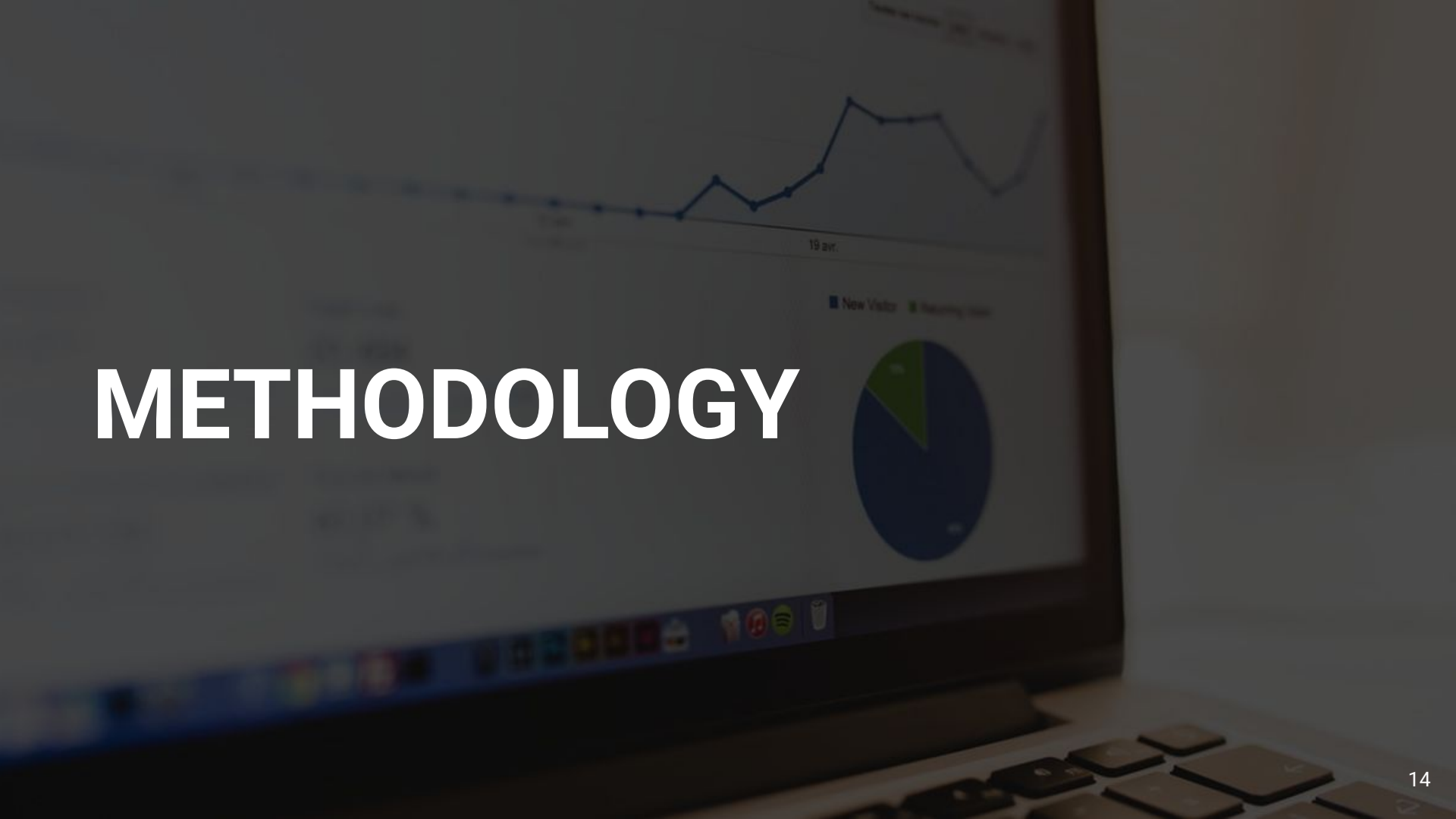
MAIN FOCUS: Cross Domain Approach



MAIN FOCUS: Cross Domain Approach



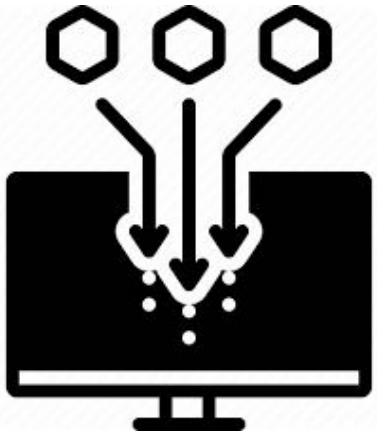
METHODOLOGY



OUR SOLUTION

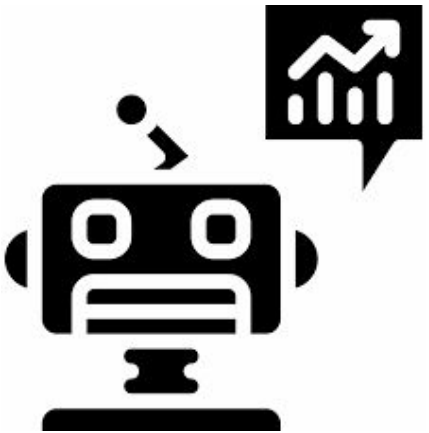
Build a model to predict the movement of market using various parameters and trends analyses.

1



2

3



Switch to Shubhankar

1 DATA COLLECTION

Social Media Data

- Scrape historical data from Twitter and Reddit

Financial Data

- Get market index and forex data for historic time periods

2 DATA PREPROCESSING

Social Media Data

- Sentiment Analysis to get numerical scores

Financial Data

- Handling missing data and creating additional features

2 DATA ANALYSIS

Understand data better

- Get correlations between financial instruments
- Perform seasonal decomposition
- Mathematical transformations
- Obtain stationary data

3 MACHINE LEARNING

Modelling approach

- Classification and Regression approach
- Statistical time series methods
- Neural networks

EVALUATION METRICS

$$RMSE = \sqrt{\sum_{i=1}^{n} \frac{(\hat{y}_i - y_i)^2}{n}}$$

- RMSE
- R2 score
- F1
- ROC
- Benchmark ML models

$$\begin{aligned} precision &= \frac{TP}{TP + FP} \\ recall &= \frac{TP}{TP + FN} \\ F1 &= \frac{2 \times precision \times recall}{precision + recall} \end{aligned}$$

CURRENT PROGRESS

DATA COLLECTION & PROCESSING

Social Media Queries

- Political leaders of Asia- Pacific countries and their terms.
- Financial capitals of countries in scope.

Financial Data Queries

- Currency exchange rates for each USA and Asia-Pacific country pair - 18 pairs
- Market Index values for all stock exchanges of nations in scope.

DATA COLLECTION: SOCIAL MEDIA


```
{"usernameTweet": "iiiitsandrea",
"ID": "11866219979812782l0",
"text": "I want this for when we need to shield our biometrics á la  Hong Kong
https:// twitter.com/singareddynm/s tatus/11866175586921840 66   …",
"url": "/iiiitsandrea/status/11866219979812782l0",
"nbr_retweet": 1,
"nbr_favorite": 1,
"nbr_reply": 1,
"datetime": "2019-10-22 20:35:05",
"is_reply": false,
"is_retweet": false,
"user_id": "539234194"}
```

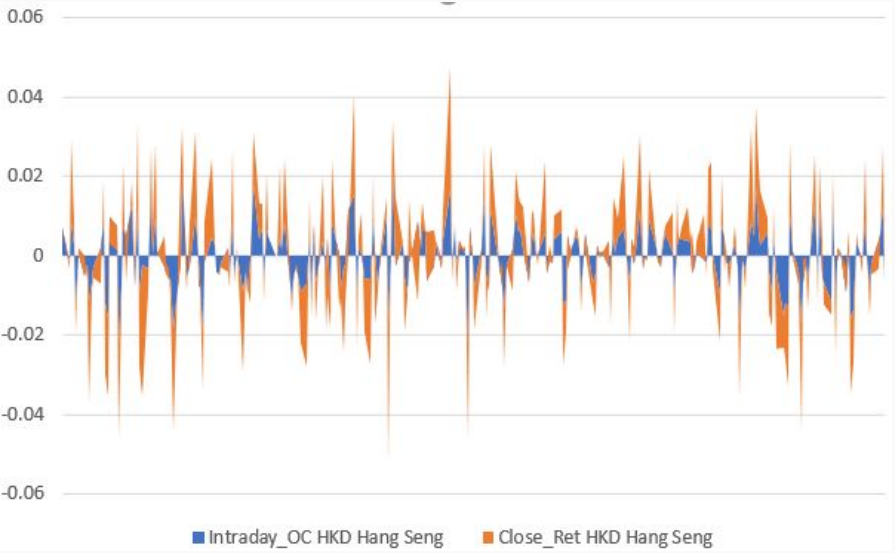
DATA PROCESSING: SENTIMENT ANALYSIS

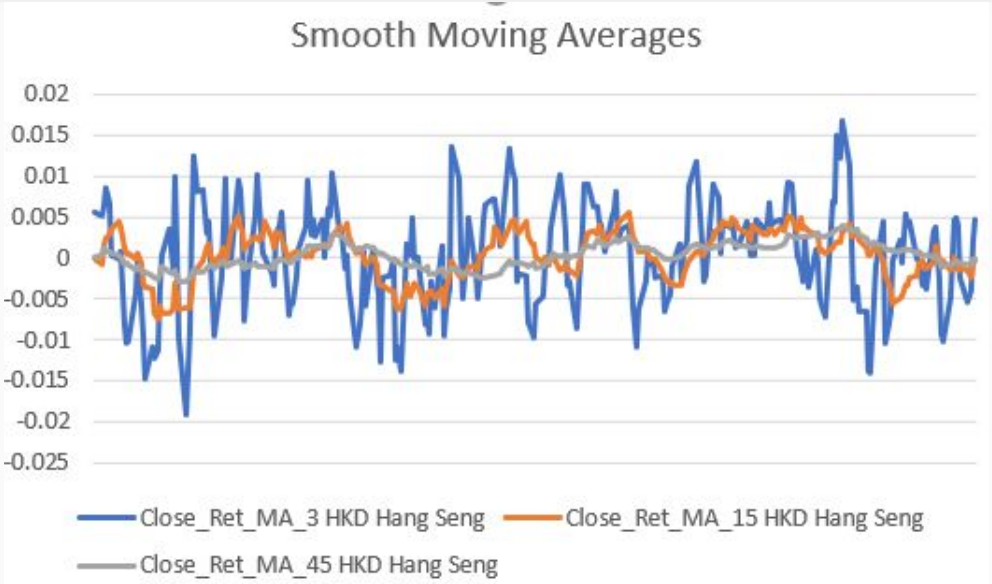

```
{"author": "xtrabola_official",
"created_utc": 1564741263,
"subreddit": "Jakarta",
"title": "Ciri Ciri Dan Kelebihan Ayam Bangkok Suro Tanjung Karang",
"url": "https://xtrabola.com/ciri-ciri-dan-kelebihan-ayam-bangkok-suro-tanjung-karang/",
"created": 1564712463.0,
"timestamp": "02-08-2019",
"test_score": ["0.067687325", "0.93231267"]}
```

DATA COLLECTION: MARKET DATA

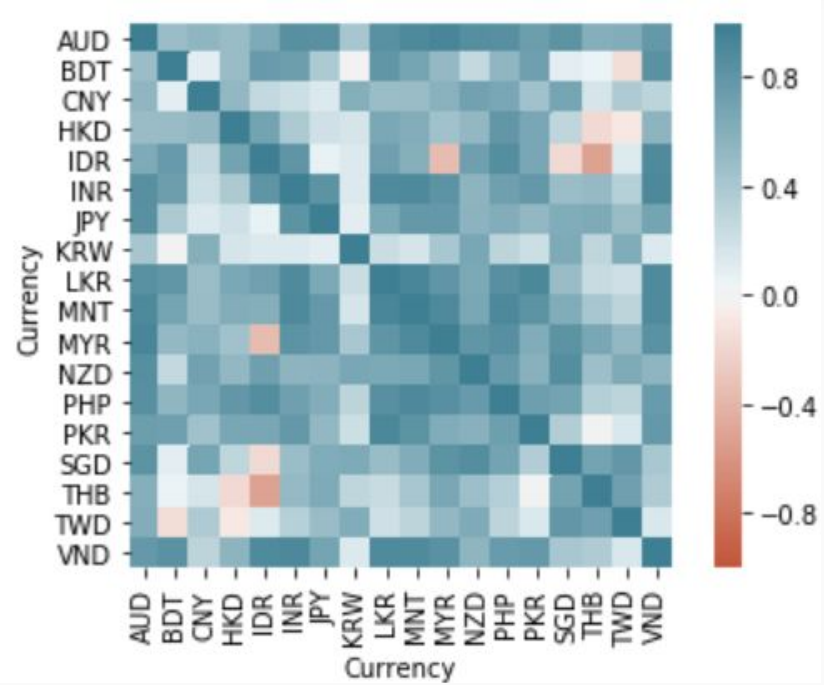
Date	Open	High	Low	Close	Volume	Currency
2019-09-23	7815.23	7842.98	7790.29	7818.61	138570000	USD
2019-09-24	7855.97	7873.25	7684.80	7710.04	205300000	USD
2019-09-25	7706.50	7822.43	7647.63	7803.54	154550000	USD
2019-09-26	7794.28	7798.12	7718.69	7771.99	146930000	USD
2019-09-27	7786.03	7790.80	7626.82	7681.58	170020000	USD

DATA PROCESSING: NEW FEATURES





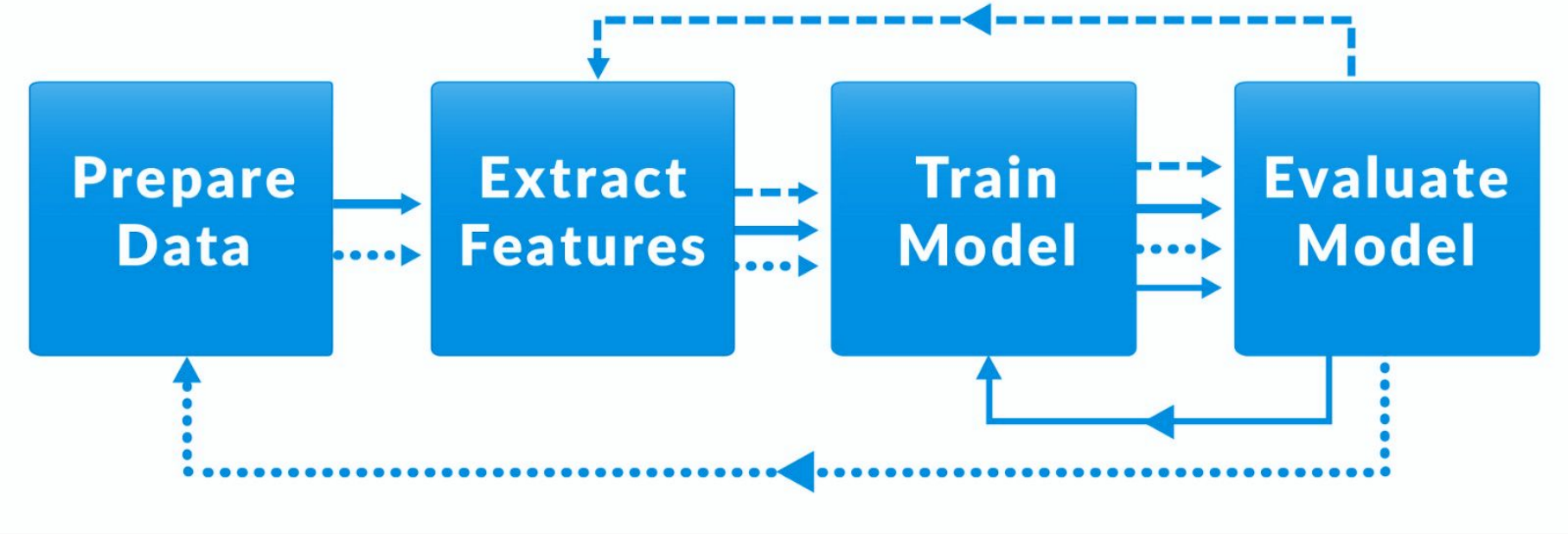
DATA PROCESSING: EDA



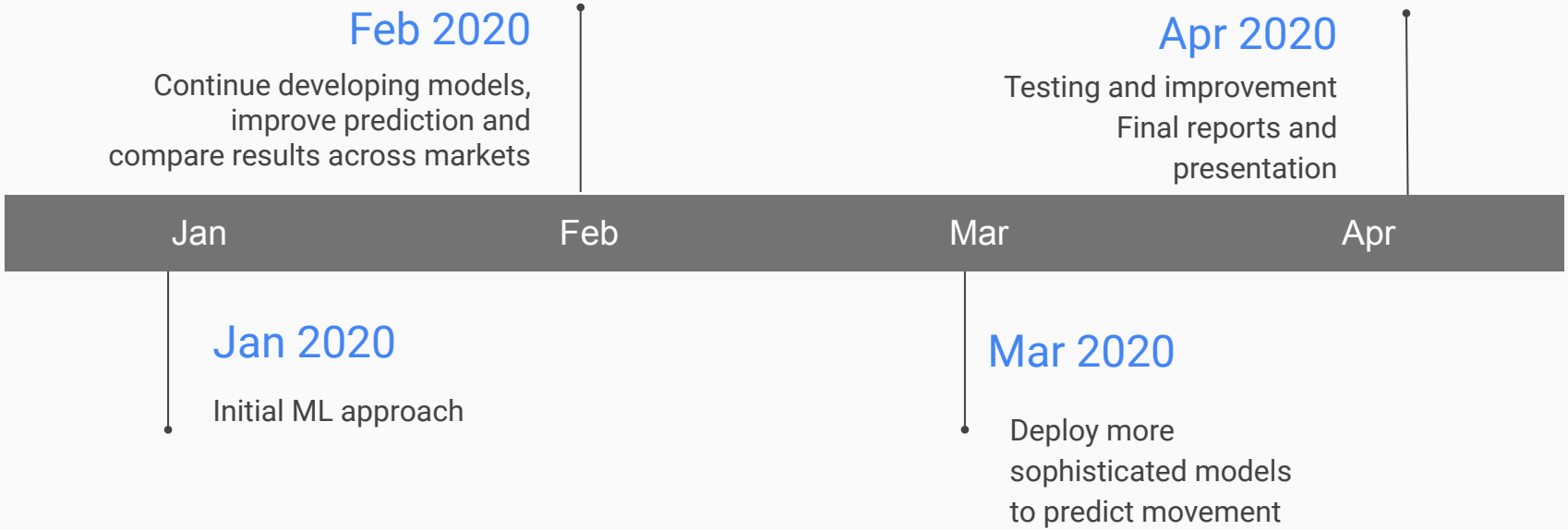
		ADF Statistic	P - Value	Lag
Close	HKD	-1.915315876	0.324891	7
Intraday_OC	HKD	-47.98682064	0	0
Close_Ret	HKD	-22.1239645	0	6
Close_Ret_MA_3	HKD	-9.904748278	3.29E-17	28
Close_Ret_MA_1	HKD	-10.00621455	1.83E-17	25
Close_Ret_MA_4	HKD	-7.983655549	2.58E-12	22
Close_MTD	HKD	-9.651401785	1.43E-16	14
Close_YTD	HKD	-3.654486706	0.004801	6

UPCOMING PHASE

MACHINE LEARNING



NEXT STEPS


Feb 2020
Continue developing models, improve prediction and compare results across markets
Apr 2020
Testing and improvement
Final reports and presentation
Jan
Feb
Mar
Apr
Jan 2020
Initial ML approach
Mar 2020
Deploy more sophisticated models to predict movement

Conclusion

BIBLIOGRAPHY

1. S. I. Ao, "A hybrid neural network cybernetic system for quantifying cross-market dynamics and business forecasting," *Soft Computing,* journal article vol. 15, no. 6, pp. 1041-1053, June 01 2011, doi: 10.1007/s00500-010-0580-4.
2. S. P. X. Jiang, J. Jiang and G. Long, "Cross-Domain Deep Learning Approach For Multiple Financial Market Prediction," presented at the 2018 International Joint Conference on Neural Networks (IJCNN), Rio de Janeiro, 2018.
3. E. K. Laitinen and T. Laitinen, "Bankruptcy prediction: Application of the Taylor's expansion in logistic regression," *International review of financial analysis,* vol. 9, no. 4, pp. 327-349, 2000.
4. M. R. Hassan and B. Nath, "Stock market forecasting using hidden Markov model: a new approach," in *5th International Conference on Intelligent Systems Design and Applications (ISDA'05)*, 2005: IEEE, pp. 192-196.
5. G. E. Box, G. M. Jenkins, G. C. Reinsel, and G. M. Ljung, *Time series analysis: forecasting and control*. John Wiley & Sons, 2015.
6. P. Vats and K. Samdani, "Study on Machine Learning Techniques In Financial Markets," in 2019 IEEE International Conference on System, Computation, Automation and Networking (ICSCAN), 2019: IEEE, pp. 1-5.

Deliverables

Inception

Detailed project plan

Project web page

September 30th, 2019

Elaboration

Preliminary implementation

Detailed interim report

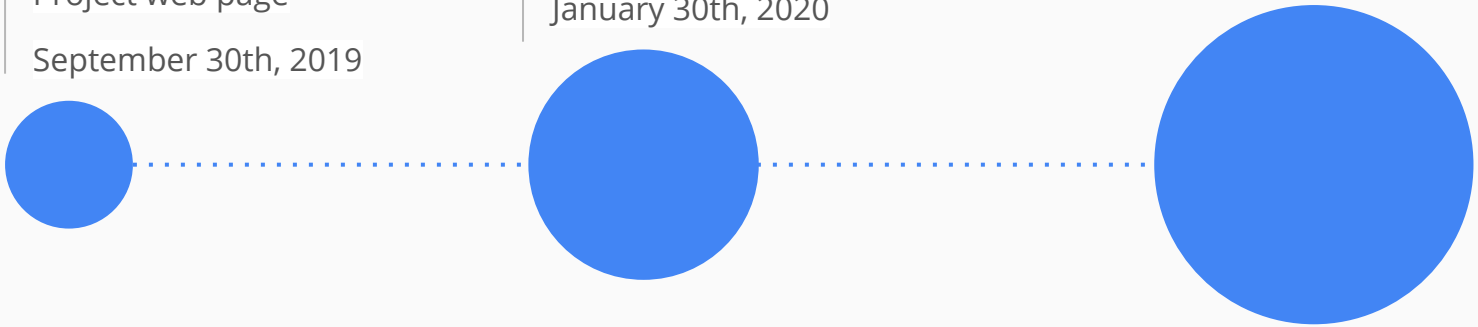
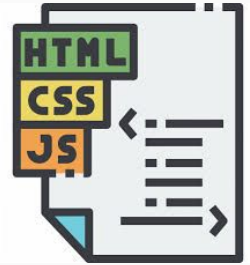
January 30th, 2020

Construction

Finalized tested implementation

Final report

April 19th, 2020



The technology: