



**ANLY482-Anlytics Practicum**  
**Project Proposal**  
**Group 06- KYY Market**



**Chong Zi Hao Kevin**

**Lim Yan Ling**

**Tee Yu Xuan**

**Professor:**

**Kam Tin Seong**

# Table of Contents

Overview .....	2
Project Sponsor Information .....	2
Motivation .....	2
Objectives .....	3
Data .....	3
Methodology.....	4
Scope of Work .....	5
Client Requirements .....	5
Privacy.....	5
Work Plan .....	6
References .....	7

## Overview

Proprietary trading has long relied on computers to help automate and execute trades. Data scientists, or more commonly known as Quants by Wall Street, have developed huge statistical models for the purpose of this automation. These models though complex, are somewhat static and as the market changes, a commonality in finance markets, they do not work as well as they do in the past.

As technology advances, we are entering an era of Artificial Intelligence and Machine Learning. Systems have capabilities to analyse large amounts of data at enormous speed and improve themselves through the process. This evolutionary computation and deep learning are seen to be able to automatically recognize changes in the market and adapt in ways the previous statistical models fail to do so.

## Project Sponsor Information

pH7 is a private investment and consultancy firm that serves clients who are keen to appreciate their wealth and grow their capital. It has its humble beginnings in 2013 in Singapore and has been working hard to build relationship with clients to understand their business and personal concerns. With its strong information analysis capabilities and experience, pH7 provides business opportunities and solutions that are customised to clients' needs. In addition, pH7 leverages on cutting-edge technology in their work, excelling in professionalism and productivity.

By partnering with market platforms that boasts state-of-the-art technology and competitive market access, pH7 aims to capitalize on every investment and business opportunity present in the markets. It aspires to be a firm of excellence and distinction which boasts of its professionalism in dealings and partnerships with clients.

## Motivation

The team's motivation for doing this project is primarily an interest in undertaking a challenging project in an interesting area of research which has been a hot topic among the finance industry, Algorithmic-centric Funding. The opportunity to learn and put into practice a new area of machine learning not covered in our academics was appealing. Machine Learning/ Artificial Intelligence is expected to take a huge role in trading, causing a notable shift the the trading markets. Utilising past data,the opportunity pH7 Global has given us allows us to tap on their expertise in trading of financial instruments and the existing market

data they have collected. This gives us a whole new experience of applying analytics in the financial markets.

## Objectives

Utilising the minute tick data from our sponsor, we would like to discover useful and practical insights which will allow traders to make more informed decisions in their trading. We would be coming up with a predictive modelling for currency pair.

The team and our sponsor pH7 Global have identified 2 areas of focus for this project.

1. Preliminary Data Analysis and Information Research
2. Predictive Algorithm Modeling and Strategy Testing

At the end of the project, the teams aims to design a unique predictive model from the data insights discovered during the analysis.

## Data

The dataset given to us includes a data of 1 minute ticks for 1 month from YYYY/MM/DD to YYYY/MM/DD. It includes XXX rows of data with the fields including:

- Timestamp (timestamp of the data)
- High (High point of the currency pair for the minute)
- Low (Low point of the currency pair for the minute)
- Open (open price of the currency pair for the minute)
- Close (closing price of the currency pair for the minute)

Timestamp	Open	High	Low	Close
2017-10-25 00:01:00	1.31303	1.31308	1.31303	1.31308
2017-10-25 00:02:00	1.31308	1.31313	1.31306	1.31307

# Methodology

Our methodology will be a XX step approach to data prediction, explanation modelling for USD/JPY 1 minute chart.

## Exploratory Segment

### 1. Data Collection

At the initial phases of data collection, we must ensure that we have the sufficient fields that are needed for modelling in the later stage.

### 2. Data Cleaning + Transformation

In the data cleaning and transformation phase, the data would be tweaked into necessary statistical and analytics parameters necessary for prediction later.

### 3. Initial Data Exploration

In this area, the data would be initially explored and we would determine the approach of modelling based on the nature of the dataset. Necessary preparations such as checking for multicollinearity of the variables would be taken into consideration before modelling of the variables would be done. Due to the nature of our dataset, careful data exploration must be done.

## Iterative Segment

### 4. Model Building

Creating model, determining predictor and target variables. In this area, we would be experimenting with multiple different approaches based on our initial understanding of the dataset after the exploration. It could range from visualizations to machine learning algorithms to achieve the objectives of our client.

### 5. Model Validation

We would be proposing a multi-variate methodology of sampling data in order to validate our model. In this aspect, we would be using the 3 way of approach of model validation called "train, test and validate". Due to the nature of the project, we would like to avoid overfitting and bias in our models. Hence, we will be aiming for a more rigorous testing process with a larger sample data size to avoid such issues.

We would also be using benchmark metrics to test our predictive modelling to ensure that it is satisfactory. Should it not be satisfactory, we would go back to phase 4 of model building or phase 2 to rebuild the model till the results is satisfactory

## Scope of Work

We intend to adopt the following steps in our analysis:

- Discover insights within the provided data
- To collect and ensure the data of currency pair is relevant in modelling
- Ensure accuracy of data by checking for multicollinearity during data exploration stage
- Identification of approaches that range from visualization to machine learning algorithms to determine predictor and target variables
- Validate model through “train, test and validate”
- Use a large sample data to prevent overfitting and bias in our model
- Design a unique predictive model
- Utilisation of benchmark metrics to test the success rate of the predictive model
- It is important to note that the scope of the project is versatile and can be furthered to address additional questions pH7 might have on the dataset

## Client Requirements

- Understanding of the variables identified during the data analysis process and include them into considerations during the predictive model building and designing phases.
- Regular meetings with client to discuss and analyse the models and strategy development.
- Developed model and strategy should be capable of live usage through testing and deployment on client provided brokerage account(s).

## Privacy

As our team will be using our client, pH7's data. We have signed the non-disclosure agreement.

# Work Plan

S/N	Task	Allocation	Pre School	Week 0	Week 1	Week 2	Week 3	Week 4	Week 5	Week 6	Week 7	Week 8	Week 9	Week 10	Week 11	Week 12	Week 13	Week 14	
Preliminary																			
1	Project Sourcing	All	✓																
2	Initial Client Meeting	All	✓																
3	Data Receiving	All	✓																
Milestone 1: Project Proposal Submission-14th Jan																			
4	Proposal Preparation	All		✓															
5	Wiki Setup	yanling		✓															
Data Collection																			
6	Checking completeness of data	Kevin			✓														
Data Cleaning + Transformation																			
7	Planning of data usage	kevin																	
8	Data Cleaning	yuxuan																	
9	ETL	yuxuan																	
Initial Data Exploration																			
10	understanding of forex	All																	
11	Data Exploration	All																	
Model Building																			
12	Model Creation	kevin, yuxuan																	
13	Type of Algorithm/Visualization	kevin, yuxuan																	
14	Determining variables	All																	
Milestone 2.1 : Interim Report Submission-25th Feb																			
Milestone 2.2: Interim Report Presentation-26th Feb to 4th March																			
15	Slides	yanling																	
16	Report	yanling																	
17	Wiki	yanling																	
18	Presentation	All																	
Model Validation																			
19	Train, test validate	All																	
20	Benchmark Metrics	Kevin																	
21	Model Tweaking	Kevin, Yuxuan																	
Milestone 3: Abstract Submission-1st April																			
22	Slides	yanling																	
23	Presentation	All																	
24	Report	yanling																	
25	Wiki page	yanling																	
26	Poster	yanling																	
Milestone 4: Full Paper Submission- 8th April																			
Milestone 5: Conference Day- 14th to 15 April																			
Milestone 6: Final Submission-22nd April																			

## References

Rise of the billionaire robots: how algorithms have redefined hedge funds. (2016, May 15). Retrieved from <https://www.theguardian.com/business/us-money-blog/2016/may/15/hedge-fund-managers-algorithms-robots-investment-tips>

Satariano, A., & Kumar, N. (2017, September 27). The Massive Hedge Fund Betting on AI. Retrieved from <https://www.bloomberg.com/news/features/2017-09-27/the-massive-hedge-fund-betting-on-ai>