SMART TRADER

* *Help Trading Smarter and More Fruitful -*

****

Project Report

By

|  |  |
| --- | --- |
| Weiyang Lim | A0107648Y |
| Gao Yong | A0107658X |
| Sathasivaiyer Nadarajasarma | A0107664A |
| Shriranga Kulkarni | A0107666Y |
| Tang Hui-Fen, Elizabeth | A0107689N |

**THE INSTITUTE OF SYSTEMS SCIENCE**

**IN PARTIAL FULFILMENT OF THE REQUIREMENTS**

**FOR THE**

**MASTER OF TECHNOLOGY IN KNOWLEDGE ENGINEERING**

**THE INSTITUTE OF SYSTEMS SCIENCE**

**NATIONAL UNIVERSITY OF SINGAPORE**

**Distribution List**

|  |  |
| --- | --- |
| Name | Role |
| Weiyang Lim | Team member |
| Gao Yong | Team member |
| Sathasivaiyer Nadarajasarma | Team member |
| Shriranga Kulkarni | Team member |
| Tang Hui-Fen, Elizabeth | Team member |
| Pang T-Howe Charles | Supervisor |
| Ding Li Ya | Supervisor |
| Zhu Fangming | Supervisor |
| Fan Zhenzhen | Supervisor |
| George Varghese | CIO PiQuant Capital |

**Acknowledgement**

It is our radiant sentiment to place on record our best regards, deepest sense of gratitude to our guide Lecturers (Mr. Charles Pang, Dr. Liya Ding, Dr. Zhu Fangming and Ms. Fan Zhenzhen) for their exemplary guidance, monitoring and constant encouragement, their careful and precious guidance which were extremely valuable for our project both theoretically and practically, throughout the course of this FYP.

We also take this opportunity to express a deep sense of gratitude to Project Sponsor Mr George Varghese PiQuant Capital Ltd, Singapore, who believed and approved sufficient resources to ensure the completion of the project.

We express our thanks to George Varghese, CIO of Pi-Quant Capital Ltd., Singapore, for taking part in useful decision & giving necessary advices and guidance and arranged all facilities to accomplish this project. We choose this moment to acknowledge his contribution gratefully.

Our team members are thankful to, Weiyang Lim, Project manager and domain expert at PiQuant Capital Ltd., Singapore, for constant ineteractions between team members and our sponsors.

Many people, especially our classmates, team members and friends , have made valuable comment suggestions on this proposal which gave us an inspiration to improve our project/research. We thank all the people for their help directly and indirectly to complete our FYP assignment.

**Project Endorsement Certificate**

Following is the letter of endorsement from PiQuant Capital Pte. Ltd (http://www.piquantcapital.com).

We extend our thanks to PiQuant members and Weyang Lim for the project endorsement.



Table of Contents

[1 PROBLEM DEFINITION 7](#_Toc407792024)

[1.1 Problem Description 7](#_Toc407792025)

[1.2 Project Scope and Objectives 7](#_Toc407792026)

[1.3 Benefits and Costs 7](#_Toc407792027)

[1.4 Solution Outline 8](#_Toc407792028)

[2 KNOWLEDGE ACQUISITION & MODELLING, PROBLEM ANALYSIS 9](#_Toc407792029)

[2.1 Domain Familiarization 9](#_Toc407792030)

[2.1.1 Domain familiarization steps – Iterative method: 9](#_Toc407792031)

[2.2 Knowledge Acquisition/Elicitation: 10](#_Toc407792032)

[2.2.1 About the Interviewee: 10](#_Toc407792033)

[3 SYSTEM DESIGN 12](#_Toc407792034)

[3.1 OPERATIONAL CONTEXT 12](#_Toc407792035)

[3.2 FUNCTIONAL DESCRIPTION 12](#_Toc407792036)

[3.3 USER INTERFACE DESIGN 12](#_Toc407792037)

[3.3.1 Graphic Interface 12](#_Toc407792038)

[3.3.2 Functionality Modularization 13](#_Toc407792039)

[3.4 KNOWLEDGE STRUCTURE & REPRESENTATION 13](#_Toc407792040)

[3.4.1 Model Factors/Variables 13](#_Toc407792041)

[3.4.2 Model Process 15](#_Toc407792042)

[3.4.3 Model Outcome 16](#_Toc407792043)

[3.4.4 Data Exploration and Pre-processing 16](#_Toc407792044)

[3.4.5 FACTOR SELECTION 21](#_Toc407792045)

[3.5 PROBLEM SOLVING PARADIGM 21](#_Toc407792046)

[3.6 TECHNICAL ARCHITECTURE AND DESIGN 22](#_Toc407792047)

[3.6.1 Data Architecture 22](#_Toc407792048)

[3.6.2 Application Architecture 22](#_Toc407792049)

[3.6.3 Core Components 23](#_Toc407792050)

[3.7 HARDWARE & SOFTWARE REQUISITE 24](#_Toc407792051)

[3.7.1 Technology Candidacy 24](#_Toc407792052)

[3.7.2 Technology Selection 24](#_Toc407792053)

[3.7.3 Development Principle 24](#_Toc407792054)

[3.7.4 Development Platform, Framework and Tool 25](#_Toc407792055)

[4 IMPLEMENTATION AND RESULTS 26](#_Toc407792056)

[4.1 Weighted Window Based Modelling 26](#_Toc407792057)

[4.2 Random Forest based Modelling 26](#_Toc407792058)

[4.2.1 Modelling approach 26](#_Toc407792059)

[4.2.2 Data preparation 27](#_Toc407792060)

[4.2.3 Algorithm/Prediction 27](#_Toc407792061)

[4.3 SVM based modelling 28](#_Toc407792062)

[4.3.1 Modelling approach 28](#_Toc407792063)

[4.3.2 Data preparation 28](#_Toc407792064)

[4.3.3 Algorithm and parameters 29](#_Toc407792065)

[4.4 Ada-Boost 29](#_Toc407792066)

[4.4.1 Modelling approach - Classify the stocks using Ada-Boost Algorithm 29](#_Toc407792067)

[4.4.2 Data preparation 30](#_Toc407792068)

[4.4.3 Ada-Boost Algorithm 30](#_Toc407792069)

[4.4.4 Flow Chart 32](#_Toc407792070)

[4.4.5 Weighted window based modelling for Ada-Boost 33](#_Toc407792071)

[4.5 Neural Network 35](#_Toc407792072)

[4.5.1 Modelling Approach 35](#_Toc407792073)

[4.5.2 Regression Modelling 35](#_Toc407792074)

[4.5.3 Genetic-Neural Hybrid Modelling 36](#_Toc407792075)

[4.5.4 Classification Modelling 37](#_Toc407792076)

[4.5.5 Weighted Window based Model 37](#_Toc407792077)

[4.6 Results & Comparison 37](#_Toc407792078)

[5 FINDINGS AND RECOMMENDATIONS 41](#_Toc407792079)

[5.1 PROJECT MANAGEMENT 41](#_Toc407792080)

[5.2 RESULTS 41](#_Toc407792081)

[5.2.1 Modified Scope 41](#_Toc407792082)

[5.2.2 Organizational Impact 42](#_Toc407792083)

[5.2.3 Implementation of the Delivery system 42](#_Toc407792084)

[5.2.4 Cost Estimation 42](#_Toc407792085)

[5.2.5 Risks 42](#_Toc407792086)

[5.3 CONCLUSION 43](#_Toc407792087)

[5.3.1 Objective Recap 43](#_Toc407792088)

[5.3.2 Challenge Summarization 43](#_Toc407792089)

[5.3.3 Modelling Highlight 43](#_Toc407792090)

[5.3.4 Future Plan 44](#_Toc407792091)

[5.4 REFERENCES 45](#_Toc407792092)

[6 Appendix A: User Requirements Specification 46](#_Toc407792093)

[6.1 Executive Summary 46](#_Toc407792094)

[6.1.1 Project Description 46](#_Toc407792095)

[6.1.2 Purpose and Scope of this Specification 46](#_Toc407792096)

[6.1.3 Benefits and Costs 47](#_Toc407792097)

[6.1.4 Solution Outline 47](#_Toc407792098)

[6.2 Product/Service Description 48](#_Toc407792099)

[6.2.1 Product Context 48](#_Toc407792100)

[6.2.2 User Characteristics 48](#_Toc407792101)

[6.2.3 Constraints 49](#_Toc407792102)

[6.3 Dependencies 49](#_Toc407792103)

[6.4 Functional requirements 50](#_Toc407792104)

[6.5 User Interface requirements 50](#_Toc407792105)

[6.6 Performance Measurement 51](#_Toc407792106)

[6.7 Maintenance 51](#_Toc407792107)

[6.8 Data requirements 51](#_Toc407792108)

[7 Appendix B: US Stock Knowledge Model 53](#_Toc407792109)

[8 Appendix C : Knowledge Acquisition Transcript 98](#_Toc407792110)

# PROBLEM DEFINITION

## Problem Description

Portfolio equity managers/traders are in the business of managing clients’ assets and need to generate reasonable returns to satisfy their clients and generate fees for themselves. They perpetually face difficult decisions such as which stocks to buy/sell, when to buy/sell and how much capital to commit at any point in time. They need to understand the possible underlying factors that drive both the stock market as a whole and individual stock returns. The ability to determine which factors work and when they work can give the portfolio manager/trader an edge over his competitors.

## Project Scope and Objectives

Overall build a dynamic and robust US stock trading system based on a multi-factor model that is systematic, rigorously and carefully tested across time and thus giving consistent and profitable returns in real life that can closely match the back test. Both long and short positions should be considered and combined wherever possible.

1. Customizable based on user’s risk profile and style preference through a user interface. This can be done through the use of appropriate screens combining various factors/metrics, leverage, number of stocks chosen (though this has to be of a sufficient size to achieve diversification and model efficacy) etc.
2. Use of appropriate KE techniques to enhance performance if possible though not at the expense of sacrificing the intuition (through appropriate use of factors) behind what drives stock performance

## Benefits and Costs

Artificial intelligence and knowledge systems technology first surfaced in the latter part of the last century when they were used to predict market trends. Since then, there has been limited literature or widespread use of such technologies in the financial markets due to various reasons such as a lack of trust in its predictive accuracy or a greater belief in individual judgment rather than using the computer to make a trading decision.

However, the benefits of using these technologies are multifold. First, the technologies and models used have become increasingly comprehensive and therefore complicated, with the ability to synthesize a large amount of data particularly with comprehensive stock level information involving a large universe such as S&P1500 and multiple attributes that the human being will be unable to perform easily by himself.

Second, non-linear decision making computer programs that attempt to make the “best” decision are attempting to more closely model the type of decision making used by humans in an effort to better reflect our comprehensive problem solving capabilities. This may eventually result in the closest we have come to technologically reproducing human logic and intelligence. Third, these systems can be dynamic and evolve constantly with the markets while being able to learn new models and relationships between data with the arrival of new financial data.

There are also possible disadvantages of a system based on artificial intelligence and predictive modeling. There are limitations as to how far these technologies can go because of the inseparable human element in the fluctuations in the stock market and how well these fluctuations can be modeled accordingly. In addition, it can be difficult to use such technologies appropriately and effectively without a solid understanding.

## Solution Outline

The ability to accurately determine the crucial factors affecting stock returns at any point in time, their importance and how to model them effectively can give the portfolio manager an edge. The portfolio manager/domain expert’s role is to determine, based on his understanding of the stock market what such factors can be. Due to a multitude of explanatory factors and the difficulty in understanding precisely the mathematical relationship between factors and stock returns, artificial intelligence or knowledge systems technology can then be applied to better attempt to model or make sense of such relationships, possible in cases where there is non-linearity and some randomness.

Computational Intelligent techniques that will be explored include Linear Regression, Decision Trees, Ada-Boost, Support Vector Machines, and Neural Networks. These methods will be used on the resulting set of factors (through a technique like random forest and other methods like correlation analysis etc.) and used to predict individual stock returns to understand which method may be more effective.

Clustering is another method that can be used on the entire set of factors to analyze relationships between stock returns and factor attributes.

Allowing the change of weights to vary across time for different composite factor groups such as valuation, profitability, price related attributes can be explored as well.

# KNOWLEDGE ACQUISITION & MODELLING, PROBLEM ANALYSIS

## Domain Familiarization

The domain familiarization is an iterative and on-going process. The team members performed it both individually and collectively in order to get familiar with the domain and understand the various concepts involved in the domain.

### Domain familiarization steps – Iterative method:

1. General background information, knowledge acquisition and elicitation from sponsors or other channels in regular and ad-hot manners with formal and informal interview, conversation, emailing, and WhatsApp.
2. Formulation of an initial set of stock level data needed.
3. Downloading the stock data at regular interval of time, to get familiar with the data.
4. Exploratory data analysis to understand the data better.
5. Data cleaning/factor scaling and modeling.
6. Appropriate scoring/normalization and relative ranking of stocks by each factor score and grouping into respective portfolio quadrants. This factor scoring has to be sector specific/even capitalization specific (each sector has its own characteristics) and based on what we understand about each factor.
7. Generated daily factor returns for all portfolio quadrants for each factor. The best quadrant should preferably perform best over time.
8. Combined and applied appropriate weightings to individual factors, then formed broad composite factors. A final factor was determined and “factor returns” then was generated for this. To form a final factor, different broad composite factors were combined at different periods based on some clustering technique. However, this was approached carefully as factor rotation is not an easy task to achieve. Possibly weightings of composite factors could change across time. Alternatively, equal weighting usually would be quite well. Multiple combinations of broad composite factors were constructed and explored to see if performing well at different times.
9. Determined market favorability which could be primarily based on stock/sector level aggregate fundamental/technical information. Economic/Monetary policies/Sentiment data was used wherever appropriate. This helped determine the level of risk to take at any point in time.
10. Portfolio construction/implementation. This was where appropriate costs and actual rebalancing of the portfolio/execution of trades has to be considered. Can be a fixed duration such as monthly or dynamic rebalancing. Appropriate leverage, net long/net short positions (if we also consider short positions) or quantity of stocks to recommend is also important based on appropriate risk to take.

*Please refer to appendix B for the Domain Familiarization model details.*

## Knowledge Acquisition/Elicitation:

Knowledge acquisition and knowledge representation are the fundamental building blocks of knowledge-based systems (KBSs). How to efficiently elicit knowledge from experts and transform this elicited knowledge into a machine usable format is a significant and time consuming problem for KBS developers. Object-orientation provides several solutions to persistent knowledge acquisition and knowledge representation problems including transportability, knowledge reuse, and knowledge growth.

One of the initial and important processes in building the expert system is the Knowledge Acquisition for stock analysis and recommender in a certain area of interest. Since we had to get domain knowledge in the field of stocks, finance and related technology, we went through following procedures for accomplishing the knowledge elicitation process.

1. Primarily, interviewing Mr Yingwei Heng (Stanley Wang), a domain expert in the field of stocks trading and finance in general.
2. Discussions with Mr. Johnny Kwon, Investment manager for Woodsford Capital Management Ltd, Singapore.
3. Also do our own research in the stocks investment and financial domain using online/library resources.
4. Take guidance/assistance of our professor, Mr. Charles (for technical feasibility and clarification regarding implementation ways).

### About the Interviewee:

Stanley Wang has a wealth of experience in the finance industry having worked in banks for a few years.

* He has considerable experience in trading equities, forex and warrants having traded for years.
* He has taken part and came in second in the Singapore round of an equity research competition organized by the CFA Institute.
* He believes that one can predict future market trend by analyzing historical data.

The interview was conducted online using Google hangouts air application and the further clarifications were discussed via mobile/email. The interview video was uploaded onto YouTube (<https://www.youtube.com/watch?v=UjAOcsTF1Aw> ) with detailed transcript.

We followed mixture of unstructured interview/Semi-structured interview techniques for the domain knowledge acquisition.

During the early stage of the interview, elicitation followed **unstructured interview mode**, where session was free flowing which produced basics of knowledge domain [basically a broad discussion about the stocks and finance domain].

For later/final stages of the interview, **Semi-structured interview** mode was followed which is the main technique for elicitation, In this mode, Pre-defined questions were sent to expert prior to interview, supplementary questions asked at the interview. We followed this format since outcome of this kind of interview, can be used as part of validation.

Apart from the Interview with domain experts, we carried out our own research in finding the information about different aspects of stocks investment. Please refer to Reference section of the report for online references/resources used for the knowledge acquisition.

*Please refer to appendix C for the summary of Knowledge acquisition transcripts.*

*Please refer to appendix A for the full User requirements specification.*

# SYSTEM DESIGN

## OPERATIONAL CONTEXT

This system can be installed in a windows based device with Windows 7 or high and .NET framework 4.5 and above. The system should have the connectivity to market data server to download the data every day.

The target user will be mainly the portfolio manager. Institutional traders also can be benefitted by using this system.

## FUNCTIONAL DESCRIPTION

The system has the following functionalities:

**Everyday basis:**

1. Download Data
   1. Downloading data from a set of servers everyday using batch scripts.
2. Data Verification
   1. Verifying the data for possible errors and missing values.
3. Data Cleanup and preparation
   1. Dirty data removal and data cubing
4. Data Standardization
   1. Rescale and Normalize
5. Data warehousing
   1. Store the data in data base.

**End of month:**

1. Generate models based on windows.
2. Apply the past window data to predict the portfolio for next month.

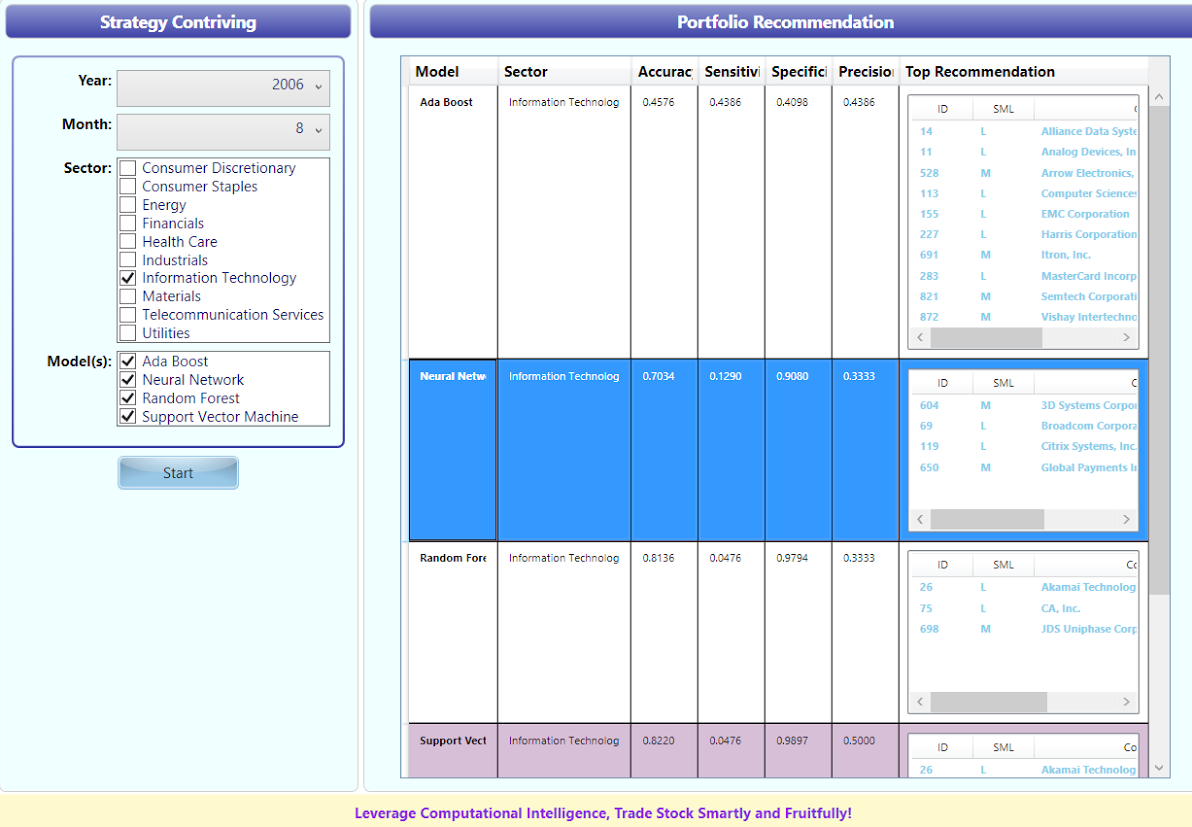
## USER INTERFACE DESIGN

The Windows Presentation Framework (WPF) based user interface is designed to wrap complex modelling and setting behind the scene and provide intuitive and simple user interaction experience.

### Graphic Interface

Primarily there are 2 regions in the GUI:

* Input panel: capture input like targeting year, month, sector and models
* Output panel: display the model performances in terms of confusion matrix metrics and top performing securities.



### Functionality Modularization

The system is modularized with distinctive front and back end components for separation of concerns and maintainability:

* Front end module:
  + Capture use customizable parameters to feed into the backend module together with other configurable system settings, i.e., NN/GA model settings
  + Once the processing is done, present the results to user
* Back end:
  + Collaborate the 3 main modeling platforms: R, C++, and .NET in terms of leveraging command line and SQL database for inter-process integration
  + Build model in the real time with user specified data window
  + Evaluate the models with stock recommendations

## KNOWLEDGE STRUCTURE & REPRESENTATION

### Model Factors/Variables

#### Objective Variable

Future stock returns over 20-day period from the point where underlying predictive factor values are used. The holding period was chosen as it allows for realistic trading when accounting for costs and system profitability. Depending on which algorithm is used to train the model, this variable will be processed and labelled accordingly.

#### Predictive Variables

We started off with a wide range of 100+ explanatory factors for stocks including forward looking estimates and historical data. They spanned six categories of stock related factors, namely valuations, profitability, quality, analyst actions, growth and technical. This initial list was used for knowledge exploration, data discovery pattern and analysis in 3.4.4. Finally, this list was narrowed down to 31 key factors based on the methodology in 3.4.5 as shown in the following table and used for predictive modelling:

|  |  |  |  |
| --- | --- | --- | --- |
| **No** | **Factor** | **Category** | **Risk Adjusted Return (Long-Short Portfolio)** |
| **1** | EarningsFY2UpDnGrade\_1M | Analyst Grading | 4.13 |
| **2** | EarningsFY1UpDnGrade\_1M | Analyst Grading | 3.64 |
| **3** | EarningsRevFY1\_1M | Analyst Revision | 3.21 |
| **4** | NMRevFY1\_1M | Analyst Revision | 2.91 |
| **5** | PriceMA10 | Technical (Reversal) | 2.69 |
| **6** | RSI14D | Technical (Reversal) | 2.25 |
| **7** | PriceMA20 | Technical (Reversal) | 2.17 |
| **8** | EarningsFY2UpDnGrade\_3M | Analyst Grading | 2.13 |
| **9** | FERating | Analyst Grading | 2.01 |
| **10** | PriceSlope10D | Technical (Reversal) | 1.99 |
| **11** | PMOM10 | Technical (Reversal) | 1.98 |
| **12** | SalesRevFY1\_1M | Analyst Revision | 1.87 |
| **13** | PMOM20 | Technical (Reversal) | 1.74 |
| **14** | NMRevFY1\_3M | Analyst Revision | 1.72 |
| **15** | PriceMA50 | Technical (Reversal) | 1.67 |
| **16** | Price52WHigh | Technical (Reversal) | 1.67 |
| **17** | EarningsFY2UpDnGrade\_6M | Analyst Grading | 1.60 |
| **18** | PriceSlope20D | Technical (Reversal) | 1.55 |
| **19** | RSI50D | Technical (Reversal) | 1.51 |
| **20** | EarningsRevFY1\_3M | Analyst Revision | 1.50 |
| **21** | PriceTStat200D | Technical (Trend) | 1.48 |
| **22** | PriceTStat100D | Technical (Trend) | 1.42 |
| **23** | Volatility12M | Technical (Others) | 1.33 |
| **24** | MoneyFlow14D | Technical (Reversal) | 1.31 |
| **25** | PEGFY1 | Valuations | 1.30 |
| **26** | Volatility6M | Technical (Reversal) | 1.30 |
| **27** | SharesChg12M | Quality | 1.30 |
| **28** | EarningsFY1Cov | Analyst Coverage | 1.28 |
| **29** | SharesChg6M | Quality | 1.27 |
| **30** | SalesYieldFY1 | Valuations | 1.12 |
| **31** | EarningsYieldFY2 | Valuations | 0.84 |

### Model Process

#### Initial steps

Following steps are formulated to cover the Data Model and Analysis Process:

* First, data files containing stock level information and underlying factors will be downloaded on a daily basis between 2004 and 2014.
* Factor values will evolve on a daily basis across the universe of stocks.
* Predictive models will be applied consistently over time but will dynamically evolve due to time-varying factor values. This ensures that models keeps up to date with constantly evolving market information, always selecting the best set of factor inputs for modeling.
* Models will be created on a monthly basis at the end of each month.
* Models will be evaluated for their accuracy and return profile over time.

#### Possible data models

We believe that the data that goes into the algorithms matter as much if not more so than the algorithms being used. Hence, we have to decide the different aspects of the data that are used as input into the algorithms. Below are various components that determine which data to use as input for the algorithms. Each of the components can be used on its own or combined.

**Moving window**

A moving window of data since the start of the entire period should be used for training and testing to take into consideration the dynamic nature of the stock market over time. What length of window should we use? On one hand, using a longer rolling window will give us more training data and thus can increase the model accuracy, but on the other hand using longer a rolling window introduces more out-of-date data which might not be consistent with the current financial environment. We choose a window of 12 months as a good cut-off point to train our model. For example, a first model to train will be to use stock input variables at the beginning of each month between January and December 2004 and use forward 20 day return just one day after the corresponding beginning of each month as the output variable.

**Sector differentiation**

A second variation of the model is to create various sector-based models, i.e. each sector will have its own model with the corresponding window used for training.

**Seasonality differentiation**

A third possible enhancement is to use seasonal based models. This is based on the idea that stocks can behave quite differently during different months of the year. Therefore, for each month across the training period (say between 2004-2009), we can aggregate cross sectional stock input factor data with their corresponding forward return and train a monthly model. The result is 12 monthly models for the available stock universe.

### Model Outcome

“Future returns” of each stock for a specified duration will be predicted.

The duration used will be 20 days such that there will be sufficient time for stocks to perform and to allow for more realistic trading including costs. The stocks with the best predicted returns will be evaluated for how they perform in real life.

The accuracy rate will be calculated based on how well the selected stocks actually performed in returns terms compared to the average stock in the selected universe. If it beats the return of the average stock, it is considered as being successfully predicted. The accuracy rate can be determined by the percentage of successfully predictions.

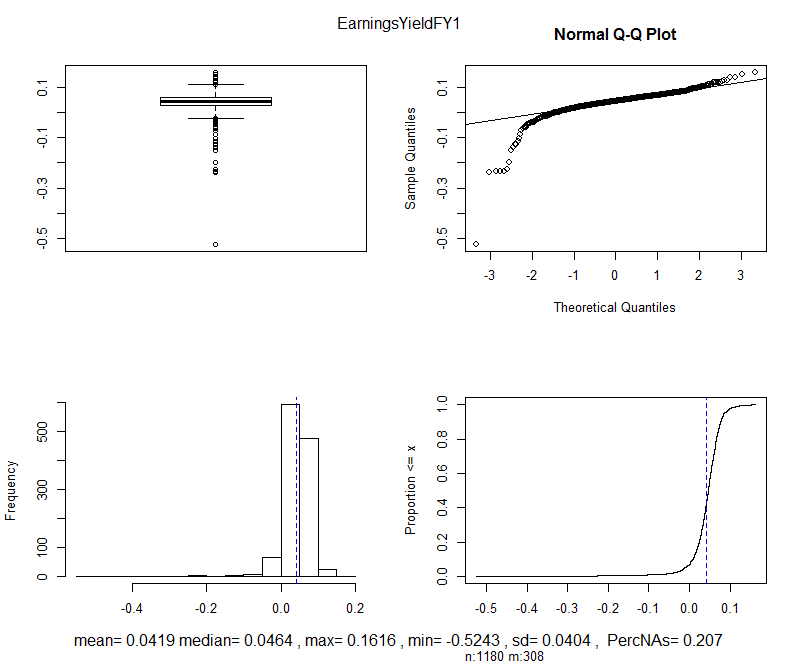
### Data Exploration and Pre-processing

Two tables were created to store various summary statistics. First is for each factor for **various market capitalizations** (Large, Mid and Small Cap) on **each day** across time. Second is for each factor for **each sector** (for all 10 sectors) on **each day** across time.

The summary statistics include the mean, the median, the maximum, the minimum, standard deviation, median absolute deviation and percentage with NAs/missing values, lower and upper limits with two mean absolute standard deviations.

This will allow us to understand the characteristics of stocks in various sectors and market capitalizations.

Various summary plots were also created on a single day as shown below. They include box plots, QQ-plots, histograms and cumulative density plots. Below shows the plots of the factor “EarningsYieldFY1” for all stocks on Jan 15, 2004. This was done before any data pre-processing was carried out.



To determine whether we can use all the stocks in the S&P1500, we examine the number of missing values for various important factors. Below shows the average percentage of missing values for three of these factors in year 2004. The year 2004 was used as this is one of the earliest years of data where data might be less readily available. It can be observed that large cap stocks have the least number of missing values followed by mid cap and small cap stocks. As small cap stocks have a large number of missing values and they are less similar in characteristics compared to large and mid-cap stocks, we decide to omit these stocks for consideration. We include the mid cap names as well as we will like to have a sufficient number of observations for our modelling.

#### Handling of outliers/missing values

A uniform approach is applied to handle outliers and missing values for all input variables and all stocks across time. The reason is that the tremendous amount of data over time does not allow us to specifically examine different periods and have different rules for achieving this purpose at any point in time. Also, such an approach allows for better comparison of results over time.

**Handling of outliers**

An outlier is determined as being three median absolute deviations above or below the median. Such outliers will be restricted within a threshold of +/- three median absolute deviations. The threshold was set to be loose primarily only to curtail the extreme outliers and to give due consideration to values which deviate significantly from the median but within the outlier limits but are of predictive value.

We use the median absolute deviation instead of standard deviation as we do not want the mean to be unnecessarily skewed by outlier values.

**Handling of missing values**

This was done via a three step approach on any trading day as described below.

1. *Remove stocks*

First, we determine if any stock needs to be removed for prediction if it has too many missing values for its input variables on any particular trading day. To achieve this, a threshold for the percentage of missing values across all input variables for each stock on any trading day was set. This was arbitrarily set as 30% as a stock observation is only considered meaningful for prediction if it has at least a percentage of input variable values that are available. Otherwise it will be removed for prediction for that trading day. Alternately, if a stock satisfies the threshold on other trading days, we will still use it for prediction for those days.

1. *Avoid using and normalizing input variables*

Next, we determine if any input variable should be removed for consideration on any trading day. The input variable will be considered meaningful for prediction if it has a minimum percentage of available observations. We arbitrarily determine the threshold for the percentage of missing values to be 30%. If the input variable has less than 70% of available values on any trading day, we will not use the input variable for prediction on that trading day. Alternately, if the input variable satisfies the threshold on other trading days, it will still be used for prediction for those days.

1. *Set missing value*

Finally, after the two steps above are done, missing values are set to be median values. A median value is considered to be neutral.

#### Normalizing input variables

Stocks from different sectors usually have large differences in terms of factor distribution and performance. For example, for a factor variable such as debt capital ratio, the telecommunications/utilities sector will have a higher average value compared to other sectors as they are more capital intensive. The technology sector on the other hand is less capital-intensive and hence has less debt. Such factor variable value will naturally tilt towards certain sectors. Therefore, it is more ideal to neutralize the factor variable for sector bias. This means when normalizing factor variables, each sector of stocks should be considered on its own. Below shows a comparison of average debt capital ratio in year 2010 across sectors.

This step is done after the data has been handled for outliers and missing values on each trading day. To normalize each input variable for each sector, we simply use the z-score. The resulting values for each input variable will have a value between -3 and 3.

#### Identifying distribution of stocks across sectors

After the data has been pre-processed and normalized by sector, we examine the number of stocks in each sector that can be used for modelling. The figure below shows the number of stocks in each sector in year 2004. The average total number of stocks that can be used was 633. The telecommunications sector had the least average number of stocks at 7. Even if subsequent years where there were more available data, the number of stocks for this sector remained small. Given the small size of the telecommunications sector, we will omit it for modelling if we were to consider the use of sector models.

### FACTOR SELECTION

#### Determining Factor Returns

To determine how effective each factor is, a return series for each factor can be constructed. The following steps are carried out.

* On any trading day, use normalized factor data and rank values for each factor for each stock in ascending/descending order.
* We will buy stocks with normalized values >= 0 and sell-short stocks that have normalized values <0, for factors where experience tells us that the higher value it is the better. For example, for a factor like earnings yield, generally the higher the value the better.
* We use the one day forward return and determine an average return based on the difference between the average return of stocks being bought and the average return of stocks being shorted.
* This process is carried out every day and the resulting daily spread returns are linked together to form a cumulative return series for the factor.
* Finally, various statistics are calculated for each return series and the effectiveness of each factor is primarily determined by the risk adjusted return. (Annualized return divided by annualized volatility averaged over all years)
* Factors with the best risk adjusted return are selected.

## PROBLEM SOLVING PARADIGM

There are two fundamental paradigms of problem solving across the whole project:

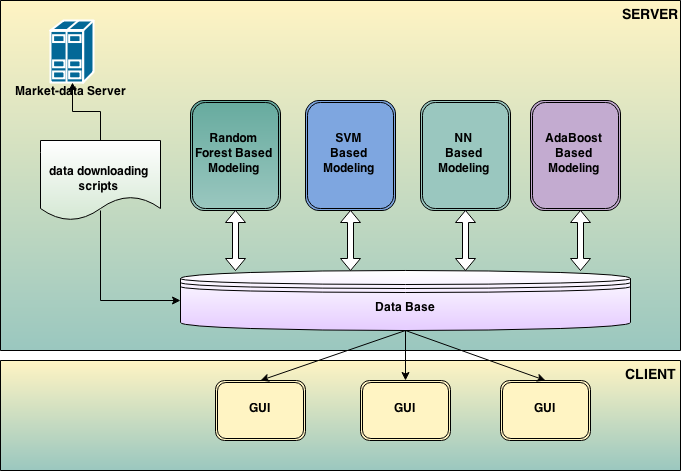
* Divide-and-conquer
* Separation-of-concern

By following and adapting the guidance of Cross Industry Standard Process for Data Mining (CRISP-DM), the project is decomposed of integral parts to carry out:

1. Data retrieval
   1. Data capture: downloading from external marketing data server.
   2. Data cleanup and normalization in database.
2. Data understanding
3. Data preprocessing
4. Modeling
   1. 4 main modeling algorithms used: Random Forest, SVM, NN and Ada-Boost.
   2. Each modeling technique running as a separate process for separation of concern.
5. Validation and Evaluation
6. Graphical User Interface Development
   1. Client program which will be connecting to data base to derive the list of stocks predicted and for getting other analysis.

## TECHNICAL ARCHITECTURE AND DESIGN

### Data Architecture

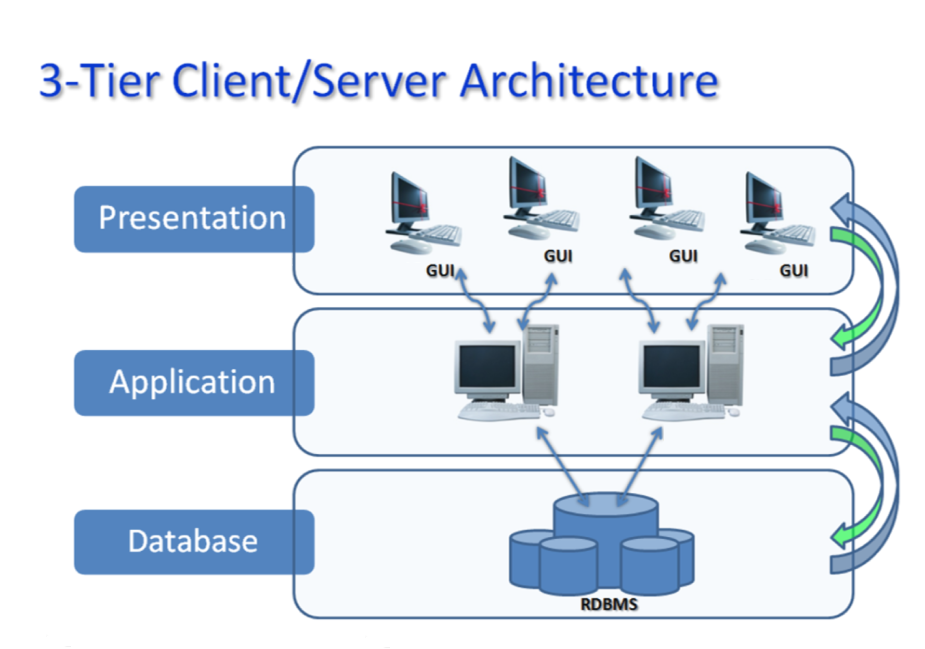


### Application Architecture

The system will be based on typical 3-tier Client/Server architecture for scalability, modularity, robustness and maintainability. The 3 layers are as shown below.

* Presentation layer
* Application layer
* Repository layer

The overall architecture can be represented as follows:



### Core Components

There will be a couple of core functional components within each layer and on top of them façade will be defined to expose functionality and service for the adjacent upper layer to consume.

* + - 1. ***Repository Layer***

This layer will mainly manage incoming/outgoing data flow on the repository level to ensure data integrity, availability, consistency and maintainability:

* Data importing component: import stock data update on a periodic or ad-hoc basis
* Data accessing component: provide data access for upper and other components
  + - 1. ***Application Layer***

This layer will encapsulate business inferential rule/logic, perform the various processes on request, and generate sensible and concrete outcomes as expected

* Inferential engine component
* Data intelligence enablement component
* User investment profiling component
* Factoring analytical and computational component
  + - 1. ***Presentation Layer***

This layer is what administrator and end user will liaise with to initiate informative queries and perform predictive tasks:

* Interactive graphical user interface
* User input collection, validation, representation and transformation component
* Reporting/exporting component

## HARDWARE & SOFTWARE REQUISITE

### Technology Candidacy

To align with project objectives and fulfil the architectural requirements with the context of limited manpower and time, the technology candidates should possess the following merits to outstand:

* Maturely adoptable
* RAD (rapid application development) capable
* Rich open source resource available
* Flatter learning curve

### Technology Selection

It consists of selection of technologies for the fundamental aspects/parts of the system design.

|  |  |  |
| --- | --- | --- |
| Aspects/Parts | Selection | Assessment |
| Database | MS SQL Server 2008 | Prerequisite of sponsor for compatibility and integration |
| Presentation | Windows Presentation Foundation 4.5 | * Mature * RAD well supported |
| Programming Language & Framework | R  Visual C++  C#  R packages  NMath.NET  AForge.NET  Encog.NET | * .NET background and foundation of team * Open source resource richness |

### Development Principle

* Agile adherent
* CRISP
* Agile modeling, processing, and delivering
* Scrum planning and scheduling

### Development Platform, Framework and Tool

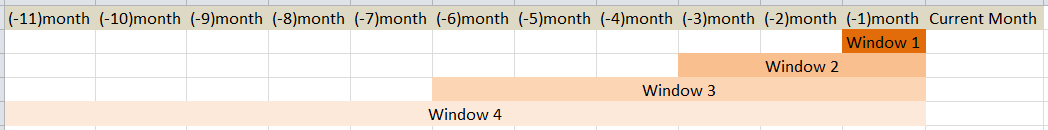
There are 2 heterogeneous technologies proposed to take advantage of: R and .NET framework. In-between there is bridging framework/library like R.NET and NMath.NET.

The major tools used include but not limited to:

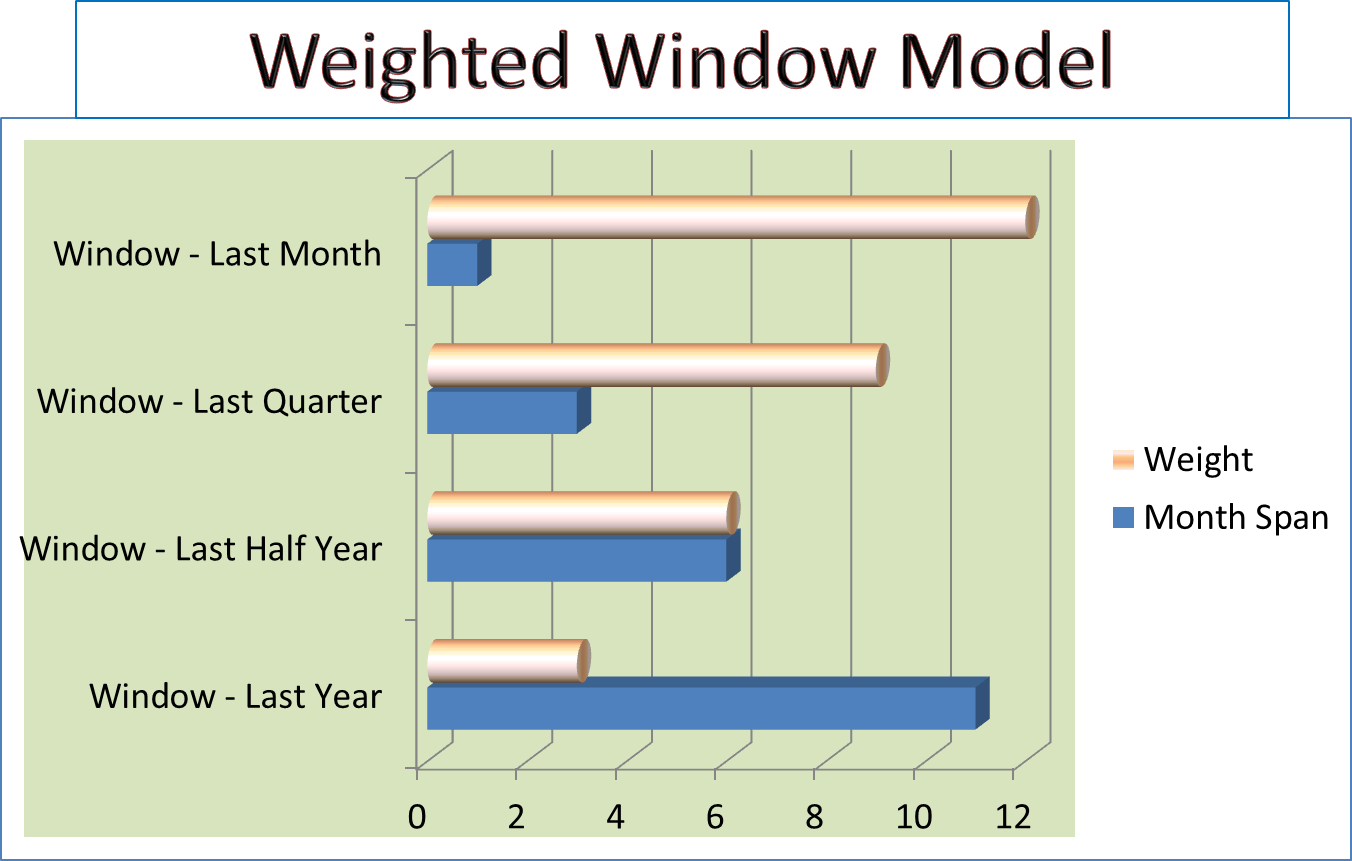
* Microsoft Visual Studio.
* R/RStudio IDE
* Microsoft SQL Server Management Studio

# IMPLEMENTATION AND RESULTS

## Weighted Window Based Modelling

Weighted windowing aims to expand the data to “look back” as attempt to gain the insight and pattern in the time series and therefore hopefully achieve better performance. Each window will build a model and these models will be assembled to product resultant prediction via voting mechanism. The system creates models from four windows:

Each window is associated with a certain weightage. Generally the more recent, the more weight assigned. This reflects the essential trading fact: short term fluctuation momentums have higher impacts for trading decision:



## Random Forest based Modelling

### Modelling approach

The RF technique is based on ensemble learning of the decision tree. Ensemble learning is useful for improving model accuracy with the integration of multiple decision trees. Random forests differ from decision tree in that they use a modified tree learning algorithm that selects at each candidate split in the learning process, a random subset of the features. If some features are very good predictors for the target variable, these features will be selected in many of the resulting trees, hence potentially increasing predictive ability than using a single decision tree.

Random forest has the following key features.

1. It can deal with high-dimensional data.
2. It allows the user to estimate the importance of each variable.
3. Variables are randomly selected at each node of a tree which provides a better classification compared to CART.
4. It stops growing the tree when the number of nodes reaches a certain threshold.

### Data preparation

We use the normalized factor score itself as the input and the normalized one month forward return as the output. The normalization is done across each factor for each sector on each day for the available stocks.

We set the stocks in the top 20%, measured by one-month forward return, as the outperformers where we labelled 1. Similarly we set the remaining stocks as 0. We perform this labelling exercise on a monthly basis, and pile the previous three months’ data together as training data.

We have only chosen the three month window due to the speed and size limitation of running the algorithm in R.

### Algorithm/Prediction

The algorithm was implemented in R using RandomForest library. Some of the parameters are as follow:

* Number of trees to grow: 500
* Number of predictors sampled for splitting at each node: 14

At the end of each calendar month between the between from April 2004 to October 2014, the model is fitted to the available stocks in each sector to predict the likely performing stocks. A three month moving window is used.

## SVM based modelling

### Modelling approach

SVMs can be a useful tool for trading applications, when non-regularity exists in the data, i.e. when the data are not regularly distributed or have an unknown distribution. This can be the case with stock level metrics. Some advantages of the SVM technique include:

1. By introducing the kernel, SVMs is flexible in the choice of the type of the threshold separating factors separating good and bad companies which may not be linear.
2. As the kernel implicitly contains a non-linear transformation, no assumptions about the functional form of the transformation, which makes data linearly separable, is necessary. The transformation is based on a robust theoretical basis without the need for human expertise judgment.
3. SVMs provide a good out-of-sample generalization, if the parameters C and r (in the case of a Gaussian kernel) are chosen appropriately. By having an appropriate generalization grade, SVMs can be robust, even when the training sample has some bias.
4. SVMs deliver a unique solution, since the optimality problem is convex. This is an advantage compared to Neural Networks, which have multiple solutions associated with local minima and for this reason may not be robust over different samples.

### Data preparation

We use the normalized factor score itself as the input and the normalized one month forward return as the output. The normalization is done across each factor for each sector on each day for the available stocks. We then assemble this data across the past three months before the current month end of data used for prediction.

We set the stocks in the top 20%, measured by one-month forward return, as the outperformers where we labelled 1. Similarly we set the remaining stocks as 0. We perform this labelling exercise on a monthly basis, and pile the previous three months’ data together as training data.

We have only chosen the three month window due to the speed and size limitation of running the algorithm in R.

### Algorithm and parameters

The algorithm was implemented in R using kernlab library. Some of the parameters are as follow:

* Kernel: Radial Basis Function
* C-constraint:1

At the end of each calendar month between the between from April 2004 to October 2014, the model is fitted to the available stocks in each sector to predict the likely performing stocks. A three month moving window is used.

## Ada-Boost

### Modelling approach - Classify the stocks using Ada-Boost Algorithm

Machine learning methods have become more and more popular in the quant world to predict stock returns. The main idea in using machine learning methods to predict future stock returns is to automatically learn to recognize complex data patterns and capture the hidden relationships among financial data, which are normally difficult to see with the human eye.

The Ada-Boost model treats stock selection as a binary classification problem using supervised learning. We classify the stocks in our universe into outperformers and underperformers based on one-month forward stock returns. The construction of the confidence score has two steps: the training step and the prediction step. In the training step, we use the end-of-month factor scores for each stock and one-month forward returns as training data to build the classifiers. In the prediction step, we use the current month factor score as the input for the classifiers we built in the training step, and the output is a confidence score.

### Data preparation

We use the cross-sectional ranking of the factors, rather than the factor score itself as the input, and we calculate the factor ranking each month for all the available stocks. Then, we divide the factor ranking by the number of stocks to normalize the factor ranking to between (0, 1]. After normalizing the factors by rank and coverage, we then assemble the training data. We set the stocks in the top n%, measured by one-month forward return, as the outperformers. Similarly we set the bottom m% as the underperformers. Note that stocks not classified in the top or bottoms are disregarded. We perform this labelling exercise on a monthly basis, and pile the different months’ data together as training data.

### Ada-Boost Algorithm

We use a machine learning algorithm called Ada-Boost (Schapire [1998]) to build classifiers that can combine the best performing factors in previous months and get a confidence score. The higher the confidence score, the more likely the stock is going to outperform in the next month. The main idea of Ada-Boost is that it adaptively builds a sequence of classifiers that are constantly being tweaked to emphasize misclassified stocks, thereby slowly improving the classification of stocks that would normally be incorrectly classified. Although certain classifiers can be weak, as long as their performance is not random, the performance of the final model will improve.

In our case, a weak classifier is simply defined by a factor. We divide the factor into quantiles, and calculate the weights of outperformers and underperformers in each quantile. Intuitively, the most effective factors are those which have the largest difference between the weights of outperformers and underperformers in each quantile. The higher the weight of outperformers relative to the weights of underperformers in that quantile, the higher value the output of that weak classifier will be. This way of defining the weak classifier can transform the non-linear factors into linear factors.

Initially, we equally-weighted each observation in the training data, and then the weights are updated in each round after a new weak classifier is found. The weight of each incorrectly classified stock is increased and the weight of each correctly classified stock is decreased, so that the next classifier focuses on the stocks which have so far not been correctly classified. In each round we choose the most effective weak classifier, which can distinguish the outperformers and underperformers the most with the current set of weights. This means the current best performing factor is less correlated with the previously selected factors. Essentially, what our machine learning system does is select those factors that are complementary to each other. The output of the strong classifier is the sum of all the weak classifiers; it is a real value confidence score of how likely the stock is to be an outperformer.

After we train our machine learning model, we can predict month-ahead returns using the constructed classifiers with new data. First we transform factor scores into quantiles, and get the corresponding value of the weak classifier. Adding up all the weak classifiers, we get the final value of the strong classifier. We use the value of strong classifier as a new composite factor, which has better performance than any of the factors comprising it.

### Flow Chart

Normalize the factor value to a real value [0,1] according to the rank of the factor in each month

Add the selected weak classifier to the strong classifier and updates the weights. Lower the weight for correctly classified samples and increase the weight for wrongly classified samples.

The final output of the strong classifier is the value of all the weak classifiers add up together.

Classify outperformers and underperformers.

Initially dually weight each stock.

Build weak classifiers for each factor.



### Weighted window based modelling for Ada-Boost

The weights used for the windows:

|  |  |
| --- | --- |
| Window | Weight |
| Window 1 | 0.4 |
| Window 2 | 0.3 |
| Window 3 | 0.2 |
| Window 4 | 0.1 |

## Neural Network

### Modelling Approach

As a computational model, NN features as follows:

* Strength
  + High tolerance to noisy data
  + Ability to classify untrained patterns
  + Well-suited for continuous-valued inputs and outputs
  + Algorithms are inherently parallel
* Weakness
  + Long training time and large training data
  + Poor interpretability

In our stock trading scenario, 2 paradigms of NN models are investigated:

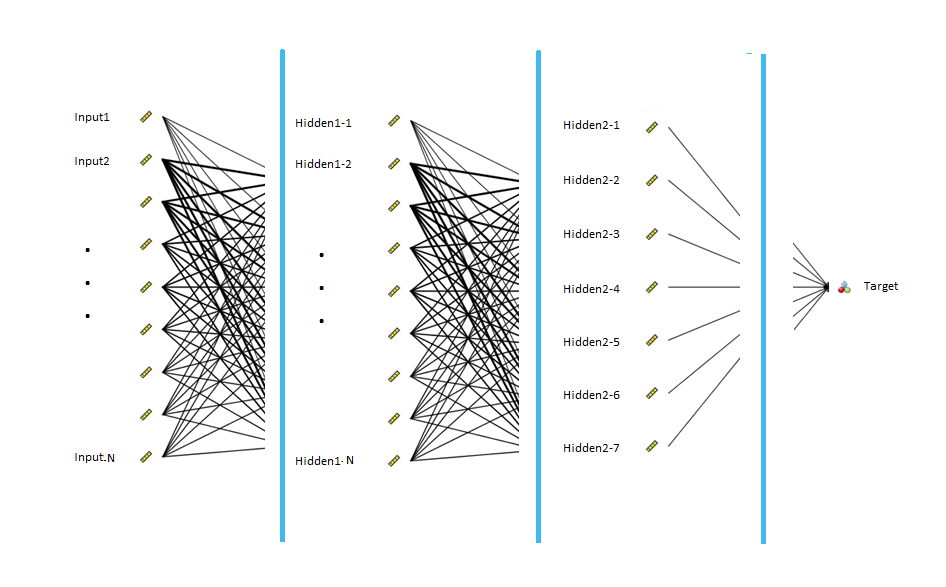
* Classification model: to predict stock price fluctuation trend, either up or down,
* Regression model: to predict the stock price fluctuation amplitude, the quantity

### Regression Modelling

The network is designed as Two-hidden-layer Feed-forward Network with back-propagation learning with the architecture below:

* Input layer: 32 nodes, representing the collective input data and bias
* Hidden layer 1: 32 / 2 = 16 nodes, representing the compositional factors
* Hidden layer 2: 6 + 1 = 7 nodes, representing the 6 aggregative factors and bias respectively
* Output layer: 1 node, representing the numerical target
* Activation function: Bipolar Sigmoid Function
* Learning: Back-propagation

The architecture is shown below:



### Genetic-Neural Hybrid Modelling

#### Rationale

As computational model, NN takes long training time with large training data. Still with the NN architecture above, instead of using Back-propagation learning to fine tune the weights of neuros, Genetic algorithm, which is capable to generate optimization solution generally and globally, is incorporated to optimize the weights as an attempt to accomplish better predictability in a shorter converging time.

#### Genetic Algorithm Design

**Chromosome Design**

Chromosome will be an array (of double-floating point number) of the Weight Adjustment (WA) of NN nodes in a fixed order, i.e., from first hidden layer to last, and finally the output layer.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| WA11 | WA12 | … | WAi1 | WAi2 | … | WAij | … |
| WA: weight Adjustment in term of percentage  New-weight = Old-weight \* (1 + WA) | | | | | | | |

**GA Operators**

* Selection: Elite selection
* Crossover: Double-number swopping
* Mutation: Mutation by “flip” randomly with adjustment

**Fitness Function**

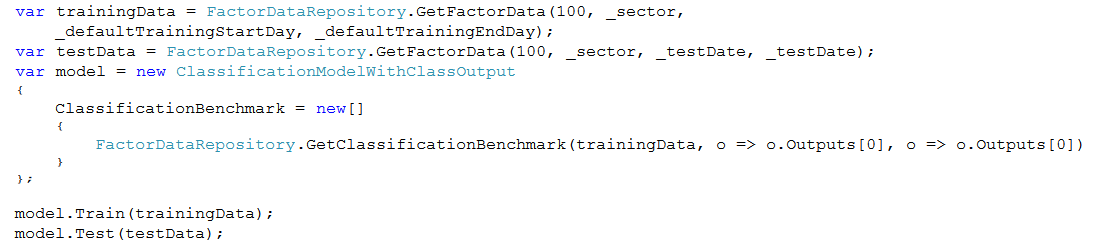
The fitness function is defined as the Least MSE of the NN against all the test data set after weight adjustments and the least is regarded the best chromosome.

### Classification Modelling

Two classification networks are designed:

* With NN regression as the backbone, a binary classifier is attached with the output node to transform the results into classes
* Probabilistic Neural Network (PNN)

For the first model, a benchmark of the output is obtained from the training data and then classifying transformer is designed as step function against the benchmark:

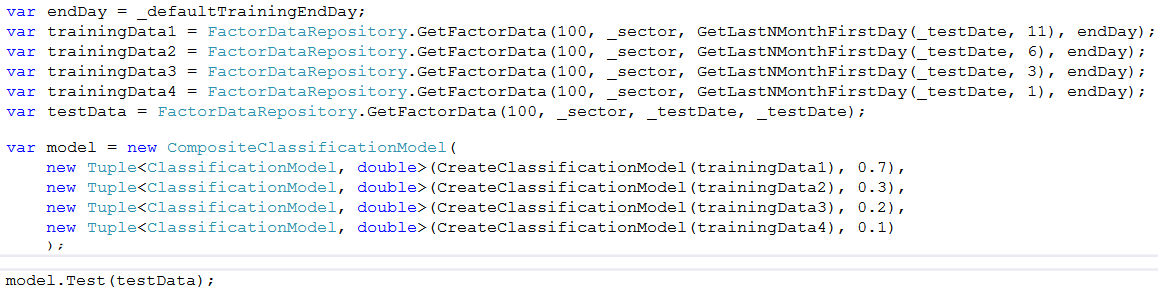


PNN features with the architecture below:

* Input layer: 32 nodes, representing the collective input data
* Output layer: 2 node, representing the categorical trend, up or down
* Activation function: Gaussian function

### Weighted Window Based Model

With the 4 predefined windows, 4 classification models are built and combined into a composite model with the majority-win voting mechanism. The C# code snippet is as follows:



## Results & Comparison

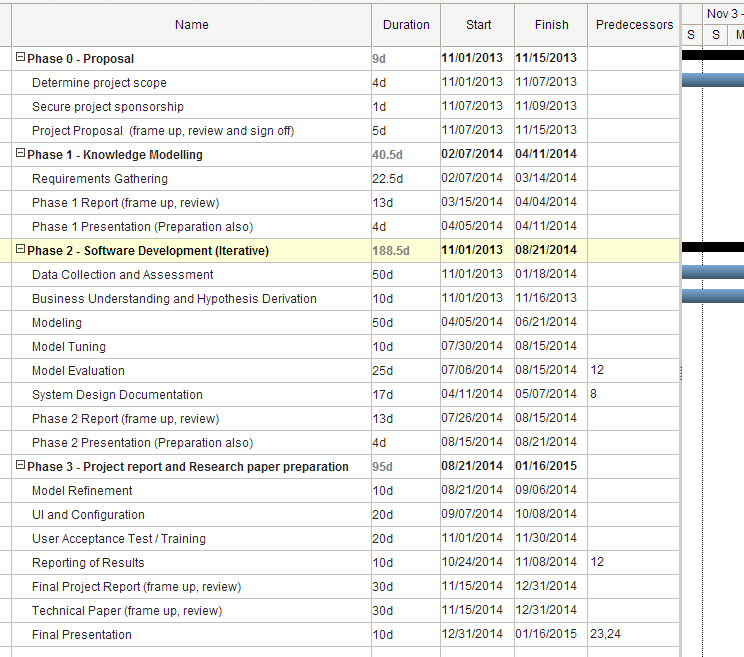
The results are measured based the confusion matrix metrics:

* Accuracy
* Precision
* Sensitivity
* Specificity

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Model | Year | Accuracy | Precision | Sensitivity | Specificity |
| Ada-Boost | 2005 | 0.507398 | 0.482923 | 0.524954 | 0.490587 |
| Ada-Boost | 2006 | 0.500866 | 0.480099 | 0.526706 | 0.486215 |
| Ada-Boost | 2007 | 0.486679 | 0.468949 | 0.508198 | 0.472625 |
| Ada-Boost | 2008 | 0.484871 | 0.484853 | 0.519157 | 0.545895 |
| Ada-Boost | 2009 | 0.524261 | 0.501385 | 0.547148 | 0.528444 |
| Ada-Boost | 2010 | 0.49652 | 0.492347 | 0.501753 | 0.537233 |
| Ada-Boost | 2011 | 0.494343 | 0.483907 | 0.519338 | 0.50331 |
| Ada-Boost | 2012 | 0.506518 | 0.503839 | 0.516051 | 0.524038 |
| Ada-Boost | 2013 | 0.517794 | 0.505235 | 0.526854 | 0.500811 |
| NN | 2005 | 0.656524 | 0.355095 | 0.188795 | 0.840184 |
| NN | 2006 | 0.655441 | 0.332077 | 0.220353 | 0.82254 |
| NN | 2007 | 0.653865 | 0.294841 | 0.177815 | 0.837617 |
| NN | 2008 | 0.653291 | 0.32126 | 0.191818 | 0.837571 |
| NN | 2009 | 0.67447 | 0.368442 | 0.203122 | 0.854632 |
| NN | 2010 | 0.629363 | 0.314757 | 0.221826 | 0.788945 |
| NN | 2011 | 0.639749 | 0.292449 | 0.176961 | 0.82811 |
| NN | 2012 | 0.648196 | 0.324148 | 0.182603 | 0.834397 |
| NN | 2013 | 0.661623 | 0.321402 | 0.188522 | 0.848532 |
| RF | 2005 | 0.787966 | 0.354523 | 0.071992 | 0.969423 |
| RF | 2006 | 0.790471 | 0.328158 | 0.057747 | 0.975323 |
| RF | 2007 | 0.781792 | 0.292999 | 0.041838 | 0.963611 |
| RF | 2008 | 0.784738 | 0.224417 | 0.033726 | 0.970124 |
| RF | 2009 | 0.776605 | 0.37778 | 0.070578 | 0.96017 |
| RF | 2010 | 0.787453 | 0.290567 | 0.043091 | 0.96545 |
| RF | 2011 | 0.782004 | 0.255398 | 0.047426 | 0.970342 |
| RF | 2012 | 0.786126 | 0.312967 | 0.040742 | 0.972721 |
| RF | 2013 | 0.787685 | 0.298247 | 0.053089 | 0.971842 |
| SVM | 2005 | 0.785733 | 0.337634 | 0.065491 | 0.968201 |
| SVM | 2006 | 0.789249 | 0.360234 | 0.058 | 0.973175 |
| SVM | 2007 | 0.78603 | 0.240781 | 0.046433 | 0.968882 |
| SVM | 2008 | 0.789937 | 0.251022 | 0.030869 | 0.976771 |
| SVM | 2009 | 0.782962 | 0.362618 | 0.058281 | 0.972481 |
| SVM | 2010 | 0.790443 | 0.257138 | 0.038005 | 0.970852 |
| SVM | 2011 | 0.782342 | 0.287183 | 0.045027 | 0.971896 |
| SVM | 2012 | 0.785543 | 0.316074 | 0.034986 | 0.973287 |
| SVM | 2013 | 0.786218 | 0.234753 | 0.045079 | 0.972099 |

# FINDINGS AND RECOMMENDATIONS

## PROJECT MANAGEMENT



## RESULTS

### Modified Scope

From the beginning our intended scope was assist the Trader/portfolio manager to generate a list of stocks which are expected to give more return on next month. From the prototype to the final version we kept the main scope unchanged. Only some small changes are done in the implementation of the system and the GUI of the system.

#### Changes in the Implementation

First our plan was implementing the whole system using R. But, because of the different libraries are available in different platforms and the consideration of handling huge database, we are pushed to use different technologies and languages. So, that we used C++, C# apart from R to complete this system. The integration of technologies and the GUI migrated to .Net platform from R.

#### Changes in GUI

We had planes to display some graphs and analysis in the GUI along with the prediction of stocks. But according to the client’s requirement, we are displaying very precise metrics (Accuracy, Sensitivity, Specificity and Precision) other than graphs.

### Organizational Impact

In Piquant capital Traders and Portfolio managers have their own set of R and S+ scripts to analyse the past data and create the stock list for next month. Our system will be changing their working style for a certain amount. The system will generate the list of stocks every month. Traders take the list and analyse do a further analysis in their own. So the system simplifies the stock section process.

### Implementation of the Delivery system

Implementation and demonstration of the system is finished. Delivering the system and knowledge transfer will be on done before 15th Jan 2015.

Support will be provided extensively throughout 1 year as we agreed with PiQuant.

### Cost Estimation

|  |  |  |
| --- | --- | --- |
|  | Component | Cost(S$) |
| Hardware | Dell PowerEdge T110 II compact tower server | 5000 |
|  | Network Connections | 400 |
|  |  |  |
| Software | Windows Server Edition | In-built with server |
|  | MS SQL server | In-built with server |
|  |  |  |
| Human Effort | Development – 30 man days | Not sponsored |
|  | Research – 150 man days | Not sponsored |

### Risks

#### System Level Risks:

1. Process down

2. Data Base down

3. System crash

4. Network down

Handling - It should be monitored by watch dog processes. It will be deployed during the actual deployment.

#### Management Level Risks

It is mainly because of the inappropriate prediction of stocks. Since it is not a automated system, this risk is minimal. It is only a support system for Traders and Portfolio Managers. This is inevitable since models will be generating output according to the data. Some sudden changes in trend will affect the system.

## CONCLUSION

### Objective Recap

The system is designed as a supportive tool to traders and portfolio managers to sort out list of the stocks which hopefully yields more return over the next one month.

### Challenge Summarization

* The main challenge faced is how to select the appropriate factors from the pool of factors. We used several techniques such as correlation and PCA analysis to select the correct set of factors. Expert Knowledge also incorporated and prioritized as a pre-heuristic part of the selection process.
* Another challenge faced is the implementation of various algorithms. Team members’ strength and expertise diversify in different techniques, framework, and programming languages. Therefore integration of the models was a big challenge ahead. Finally we proposed a multi process architecture interconnected by database. This process flow is very cost effective and easy to develop and maintain. We proposed this architecture to client who consents to move forward.
* Additionally, there are 4 different modelling techniques with multiple moving windows used to build the whole system. Multiple moving window technique is quite resource intensive process over tens of million data records. Memory overflow is an issue not avoidable. Thus seeking for data/feature reduction and resource-efficient alternative is the future direction for the project to succeed in real scenario.

### Modelling Highlight

Advantages of this architecture are easy to add another model without changing the existing implementation because the modelling techniques are decoupled fully and only have the interaction via the data base.

To improve the performance, we managed to create a moving window based modelling. 11 months used to train the model and 12th month will be the testing month. Also, we used another 4 windows inside this main window, 1 month, 3 months, 6 months and 11 months together with according weights for each window. Last month window gets higher weight and the 11 month window gets lowest weight, which reflects the trading expertise taken advantage of currently.

The performance of the system will be measured gradually and time to time we agreed to fine tune the models. To accomplish this client will have a standard format to record the system output and actual return. That document will be analysed time to time by the system team to fine tune the models.

### Future Plan

As for the proof-of-concept nature of the project currently most efforts were focused on the modelling aspect. In future the project intends to extend in the following aspects:

* Evolve the modelling for refined performance
* Attempt new models
* New functionality exposed and accessible from UI and transform the research based solution into practical product.

## REFERENCES

* <http://piquantcapital.com/home.html>
* <http://www.kdnuggets.com/datasets/index.html>
* <http://en.wikipedia.org/wiki/Data_mining>
* <http://scholarbank.nus.edu.sg/handle/10635/35562>
* <http://www.nasdaq.com/investing/glossary/>
* <http://teamgantt.com/>
* <http://www.r-project.org/>
* <http://rdotnet.codeplex.com/>
* <http://www.microsoft.com/en-us/server-cloud/products/sql-server/>
* <http://finance.yahoo.com/stock-center/>
* <http://www.stocknod.com/stock-market-for-beginners/>
* Various NUS Library Resources

# Appendix A: User Requirements Specification

## Executive Summary

### Project Description

Portfolio equity managers/traders are in the business of managing clients’ assets and need to generate reasonable returns to satisfy their clients and generate fees for themselves. They perpetually face difficult decisions such as which stocks to buy/sell, when to buy/sell and how much capital to commit at any point in time. They need to understand the possible underlying factors that drive both the stock market as a whole and individual stock returns. The ability to determine which factors work and when they work can give the portfolio manager/trader an edge over his competitors.

### Purpose and Scope of this Specification

The aim of this project is to develop an intelligent and dynamic software system which will generate a portfolio of US stock information from the S&P1500 based on a set of most relevant and important explanatory factors and accompanying models on any particular day, and help the traders/portfolio managers to make informed decisions about selecting the most ideal stocks to generate their desired absolute return. Below show the various scope of the project.

* Output will be a list of stocks which are expected to give high return relative to other stocks within universe on a particular day.
* User has the control over the number of maximum stocks in the list and the minimum possibility for the prediction.
* Factors selected by the models used in stock selection at any point in time and possible importance/explanatory ability.

#### Possible Extensions

* Integration with trading systems for real time trading of stocks, based on the historical models and results generated. Incorporate costs, use appropriate/realistic holding periods and risk control mechanisms and back-test strategy over historical period
* Incorporate market favorability/direction which can possibly be based on a predictive model and advice on stock allocation accordingly.
* User preference over the selection of factors.
* Other stock selection methods based on various styles such as value or dividend investing, growth investing and trend/momentum trading etc.

### Benefits and Costs

Artificial intelligence and knowledge systems technology first surfaced in the latter part of the last century when they were used to predict market trends. Since then, there has been limited literature or widespread use of such technologies in the financial markets due to various reasons such as a lack of trust in its predictive accuracy or a greater belief in individual judgment rather than using the computer to make a trading decision.

However, the benefits of using these technologies are multifold. First, the technologies and models used have become increasingly comprehensive and therefore complicated, with the ability to synthesize a large amount of data particularly with comprehensive stock level information involving a large universe such as S&P1500 and multiple attributes that the human being will be unable to perform easily by himself.

Second, non-linear decision making computer programs that attempt to make the “best” decision are attempting to more closely model the type of decision making used by humans in an effort to better reflect our comprehensive problem solving capabilities. This may eventually result in the closest we have come to technologically reproducing human logic and intelligence. Third, these systems can be dynamic and evolve constantly with the markets while being able to learn new models and relationships between data with the arrival of new financial data.

There are also possible disadvantages of a system based on artificial intelligence and predictive modeling. There are limitations as to how far these technologies can go because of the inseparable human element in the fluctuations in the stock market and how well these fluctuations can be modeled accordingly. In addition, it can be difficult to use such technologies appropriately and effectively without a solid understanding.

### Solution Outline

The ability to accurately determine the crucial factors affecting stock returns at any point in time, their importance and how to model them effectively can give the portfolio manager an edge. The portfolio manager/domain expert’s role is to determine, based on his understanding of the stock market what such factors can be. Due to a multitude of explanatory factors and the difficulty in understanding precisely the mathematical relationship between factors and stock returns, artificial intelligence or knowledge systems technology can then be applied to better attempt to model or make sense of such relationships, possible in cases where there is non-linearity and some randomness.

KE techniques that will be explored include linear regression, decision trees, Ada-Boost, support vector machines, neural networks. These methods will be used on the resulting set of factors (through a technique like random forest and other methods like correlation analysis etc) and used to predict individual stock returns to understand which method may be more effective.

Clustering is another method that can be used on the entire set of factors to analyze relationships between stock returns and factor attributes.

Allowing the change of weights to vary across time for different composite factor groups such as valuation, profitability, price related attributes can be explored as well.

## Product/Service Description

As this is a prediction system, and the result of the prediction is based on a set of underlying explanatory factors, the output produced may differ from actual results depending on the unforeseen events in the market. System will be tested with the past data and validated for maximum reliability. The accuracy of the system is dependent on the real time data collected as trained data.

### Product Context

The system is an independent system having no direct dependencies with other systems such as live stock data feed or brokerage systems. The analysis results produced by systems. This system may be used as input for trading strategies which are currently out of scope of the system.

.

### User Characteristics

This system has two types of users.

* Admin
  + This mode is used to train the system and modify the core functionalities of the system. Only the system developer and system administrator have the privileges of Admin user.
* End User
  + This mode is used by Traders/Portfolio Managers to generate the portfolio every day. This is the target mode (end user mode) of this system.

### Constraints

Because of the huge data to be analyzed to generate the relevant models and resulting portfolios, a significant processing time and processing power will be required.

## Dependencies

* This system will require a daily download of data from the market data server.
* Downloading the data and transferring to appropriate place to be completed before starting this process.

## Functional requirements

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Requirement No.** | **Requirement** | **Comments** | **Priority** | **Reviewed/**  **Approved** |
| **1** | Data Availability | Downloading data from servers using appropriate tools/scripts | 5 |  |
| **2** | Data Verification | Verifying the data for possible errors and missing values. | 3 |  |
| **3** | Data Cleanup and preparation | Clean the data. | 2 |  |
| **4** | Data Standardization | Rescale and Normalize. | 2 |  |
| **5** | Data Analysis and Model Design and Implementation | Data mining and modeling | 1 |  |
| **6** | Prediction | Predicted stocks with higher return. | 1 |  |

**1** - Highest priority, **6** - Lowest priority

## User Interface requirements

User has control over the following items:

1. No. of stocks in the output list.

2. Minimum expected return over a desired period

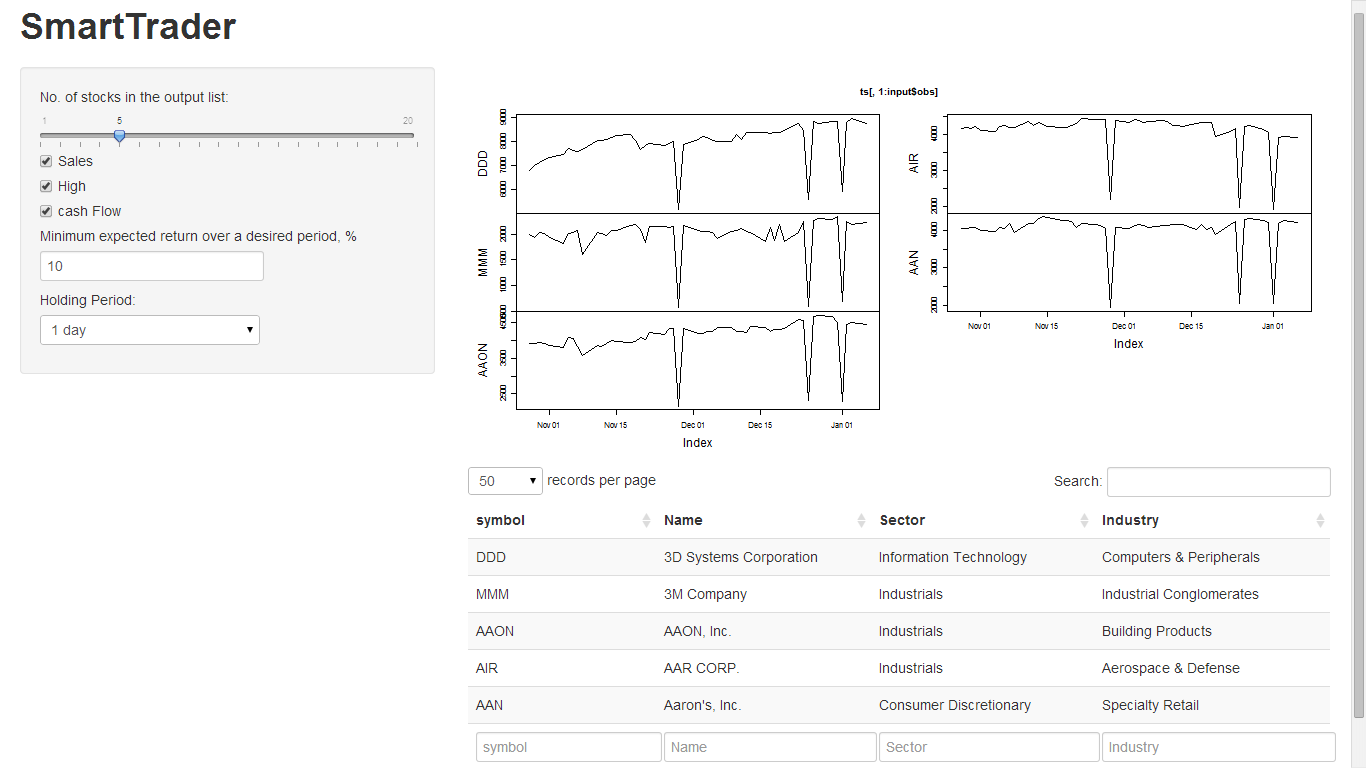
3. Number and/or Type of variables to include/exclude for the prediction. (Possibly add on and for advanced users)

Possible Outputs:

1. List of stocks

2. Graphs

3. Variables analyzed, analyzed periods and details.

Schematic Representation of The system Output

## Performance Measurement

Performance will be measured as below:

1. Number of records processed and analyzed per unit time

2. System throughput

## Maintenance

1. All the technical design and source code should be properly documented.

2. System should have interfaces to scale/extend in future.

3. Different levels of logging mechanisms are essential.

## Data requirements

|  |  |
| --- | --- |
| **Requirement** | **Description** |
| **Types of information used by various functions** | Market data ticks. |
| **Frequency of use** | Everyday |
| **Data Security** | Data should be protected from unauthorized access. |
| **Data entities and relationships** | Relationship and association between different data objects and variables should be designed |
| **Integrity constraints** | Data integrity refers to maintaining and assuring the accuracy and consistency of data over its entire life-cycle |
| **Data retention** | N/A |
| **Valid range, accuracy, and/or tolerance** | Range is different for different variables. |
| **Units of measure** | N/A |
| **Data formats** | CSV/XML, MSDB, or other applicable types |
| **Default or initial values** | N/A |

# Appendix B: US Stock Knowledge Model

Please find below the Domain Familiarization of the US Stock Knowledge Model

**/\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/**

**/\* US Stock Knowledge Model \*/**

**/\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/**

/\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\* S T A R T \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/

/\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/

**KNOWLEDGE-MODEL** us-stock;

/\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/

/\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/

**DOMAIN-KNOWLEDGE** us-stock-domain;

/\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/

/\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/

**DOMAIN-SCHEMA** stock-selection-schema;

/\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/

/\* 1st of 3 main concepts stock-list \*/

**CONCEPT** stock-list**;**

**DESCRIPTION**:

"A description of the basic characteristics of each stock";

**ATTRIBUTES**:

company: STRING;

symbol: STRING;

market-cap (millions): FLOAT;

volume (millions): FLOAT;

cap-category: {large, mid, small};

style-selection: {dividend, growth, momentum, value};

price-chg%10d: FLOAT

price-chg%1m: FLOAT

price-chg%3m: FLOAT

price-chg%6m: FLOAT

pos-earning-growth2y: BOOLEAN

pos-earning-growth4y: BOOLEAN

pos-earning-growth6y: BOOLEAN

earning-growth3y: FLOAT

earning-growth6y: FLOAT

earning-growth9y: FLOAT

pe-fy1: FLOAT

p-sales-fy1: FLOAT

div-yield%: FLOAT

pos-div-growth5y: BOOLEAN

pos-div-growth10y: BOOLEAN

pos-div-growth15y: BOOLEAN

risk-match-cf: {-1 - 1}

stock-attractiveness-cf: {-1 - 1}

recommendation-rank: {-1 - 1}

/\* RULE-TYPE 1 determines ‘sector’ \*/

sector: {consumer-staples, consumer-discretionary, energy,

financials, health-care, industrials, materials, technology,

telecommunication-services, utilities};

/\* RULE-TYPE 2 determines ‘risk-match-certainty-factor’ \*/

risk-match-certainty-factor: {between -1 and 1}

/\* RULE-TYPE 3 determines ‘stock-attractiveness-certainty-factor’ \*/

stock-attractiveness-certainty-factor: {between -1 and 1}

/\* RULE-TYPE 4 determines ‘recommendation-rank’ \*/

recommendation-rank: {between -1 and 1};

**END CONCEPT** stock-list**;**

/\* SUB SET OF 1st of 3 main concepts stock-list \*/

**CONCEPT** stock-risk-feature**;**

**DESCRIPTION**:

"features of a stock that can affect the riskiness of a stock";

**SUB-TYPE-OF:** stock-list

**ATTRIBUTES**:

cap-category: {large, mid, small}

liquidity: volume-category;

price-volatility: FLOAT;

earning-volatility: FLOAT;

sector: {consumer-staples, consumer-discretionary, energy,

financials, health-care, industrials, materials, technology,

telecommunication-services, utilities};

stock-risk-cf: {-1 - 1}

**END CONCEPT** stock-risk-feature;

**VALUE-TYPE** volume-category**;**

**TYPE:** ORDINAL;

**VALUE-LIST:** {<500000,>=500000};

**END VALUE-TYPE** volume-category;

/\* SUB SET OF 1st of 3 main concepts stock-list \*/

**CONCEPT** stock-attractiveness-feature-inherent-market-feature;

**DESCRIPTION**:

"features of a stock that can affect the attractiveness of a stock";

**SUB-TYPE-OF:** stock-list

**ATTRIBUTES**:

peg: FLOAT;

ebitda-margin: FLOAT;

roic: FLOAT;

net-debt-equity-ratio: FLOAT;

stock-attractiveness-inherent-market-feature-cf: {0 - 1};

**END** CONCEPT stock-attractiveness-feature-inherent-market-feature;

/\* SUB SET OF 1st of 3 main concepts stock-list \*/

**CONCEPT** stock-attractiveness-feature-inherent-derived**;**

**DESCRIPTION:**

"derived features of a stock that can affect the attractiveness of a stock";

**SUB-TYPE-OF:** stock-list

**ATTRIBUTES:**

market-favourability: {high, medium, low}

//(derived from Market Favourability CF)

cap-category-attractiveness-cf: {0 - 1}

sector-attractiveness-cf: {0 - 1}

stock-attractiveness-from-market-favourability-cf: {0 - 1}

**END CONCEPT** stock-attractiveness-feature-inherent-derived;

/\* SUB SET OF 1st of 3 main concepts stock-list \*/

**CONCEPT** stock-style-feature;

**DESCRIPTION:**

"Stock level style feature to be matched with Customer Style Profile ";

**SUB-TYPE-OF:** stock-list

**ATTRIBUTES:**

style-selection: {dividend, growth, momentum, value};

price-momentum-preference: {<4weeks,4weeks-12weeks,>12weeks};

no-consecutive-years-increasing-earnings: {2,4,6};

earning-growth-rate: {10%-20%,20%-30%,>30%};

yrs-used-for-earning-growth: {3,6,9}

price-earning-ratio: {<6,6-10,10-15};

price-sales-ratio: {0-0.5,0.5-1,1-1.5};

dividend-yield-required: {0-2%,2%-4%,>4%};

dividend-growth-desired: {5%-8%,8%-12%,>12%};

no-consecutive-positive-years-dividend-growth: {5-10,10-15,>15,};

**END CONCEPT** stock-style-feature;

/\* 2nd of 3 main concepts macroeconomic-and-stock-market-conditions \*/

**CONCEPT** macroeconomic-and-stock-market-conditions**;**

**DESCRIPTION:**

"overall macroeconomic and stock market conditions";

**ATTRIBUTES:**

market-favourability-cf: {between 0 and 1};

macroeconomic-conditions-cf: {between -1 and 1};

stock-market-conditions-cf: {between -1 and 1};

**END CONCEPT** macroeconomic-and-stock-market-conditions;

/\* SUB SET OF 2nd of 3 main concepts macroeconomic-and-stock-market-conditions \*/

**CONCEPT macroeconomic-conditions;**

**DESCRIPTION:**

"Economic factors that influence the state of the whole economy, such as changes in employment levels, gross national product (GNP), and prices (deflation or inflation)";

**SUB-TYPE-OF:** macroeconomic-and-stock-market-conditions

**ATTRIBUTES:**

inflation-rate**:** {<0% or >4%: bearish, 0% - 2%: bullish, 2% - 4%: neutral};

unemployment-rate: {<5%: bullish, 5%-6%: neutral, >6%: bearish};

unemployment-trend: {uptrend,flat,downtrend}

average-weekly-unemployment-claim: {<350,000: bullish, 350,000 - 400,000: neutral, >400,000: bearish};

unemployment-claim-trend: {uptrend,flat,downtrend};

yield-curve-slope: {upward-sloping,flat,downward-sloping};

retail-sale-trend: {uptrend,flat,downtrend};

pmi: {<50,>=50};

**END CONCEPT** macroeconomic-conditions;

/\* SUB SET OF 2nd of 3 main concepts macroeconomic-and-stock-market-conditions \*/

**CONCEPT** stock-market-conditions;

**DESCRIPTION:**

"Market conditions is a term that refers to the state of an stock";

**SUB-TYPE-OF:** macroeconomic-and-stock-market-conditions

**ATTRIBUTES:**

indicator-1: {very-bearish,bearish,neutral,bullish,very-bullish};

indicator-2: {very-bearish,bearish,neutral,bullish,very-bullish};

indicator-3: {very-bearish,bearish,neutral,bullish,very-bullish};

indicator-4: {very-bearish,bearish,neutral,bullish,very-bullish};

indicator-5: {very-bearish,bearish,neutral,bullish,very-bullish};

**END CONCEPT** stock-market-conditions;

/\* 3rd of 3 main concepts customer-profile \*/

**CONCEPT** customer-profile;

**DESCRIPTION:**

"Customer level profile; each customer has one profile”;

**ATTRIBUTES:**

customer-name: STRING;

customer-address: STRING;

customer-phone: NATURAL;

**END CONCEPT customer-profile;**

/\* SUB SET OF 3rd of 3 main concepts customer-profile \*/

**CONCEPT** customer-risk-profile;

**DESCRIPTION:**

"features of a customer's ability to buy stock";

**SUB-TYPE-OF:** customer-profile

**ATTRIBUTES:**

overall-risk-tolerance-cf: {between -1 and 1};

risk-tolerance-ability-cf: {between -1 and 1};

risk-tolerance-willingness-cf: {between -1 and 1};

**END CONCEPT** customer-risk-profile;

/\* SUB SET OF customer-risk-profile, which is SUB SET OF 3rd of 3 main concepts customer-profile \*/

**CONCEPT** customer-risk-profile-ability;

**DESCRIPTION:**

"features of a customer's ability to buy stock";

**SUB-TYPE-OF:** customer-risk-profile

**ATTRIBUTES:**

age: {20-35,36-45,46-55,56-65}

net-assets: {<25000,25000-100000,100000-250000,250000-500000,>500000};

job-stability: {very-unstable,rather-unstable,a-bit-stable,stable,very-stable};

monthly-income: {<2500,2500-5000,5000-7500,7500-10000,>10000};

liquidity-needs: {very-low,low,medium,high,very-high};

**END CONCEPT** customer-risk-profile-ability;

/\* SUB SET OF customer-risk-profile, which is SUB SET OF 3rd of 3 main concepts customer-profile \*/

**CONCEPT** customer-risk-profile-willingness;

**DESCRIPTION:**

"features of a customer's willingness to buy a stock";

**SUB-TYPE-OF:** customer-risk-profile

**ATTRIBUTES:**

thrill-factor: {very-low,low,medium,high,very-high};

job-security-preference: {very-high,high,medium,low,very-low};

investment-risk-tolerance: {very-low,low,medium,high,very-high};

job-compensation-preference: {all-salary,mainly-salary,equal-mix-salary-commission,mainly-commission,all-commission};

aversion-to-previous-stock-loss: {very-high,high,medium,low,very-low};

**END CONCEPT** customer-risk-profile-willingness;

/\* SUB SET OF 3rd of 3 main concepts customer-profile \*/

**CONCEPT** customer-style-profile;

**DESCRIPTION:**

"features of customer's personality";

**SUB-TYPE-OF:** customer-profile

**ATTRIBUTES:**

style-selection: {dividend, growth, momentum, value};

price-momentum-preference: {<4weeks,4weeks-12weeks,>12weeks};

no-consecutive-years-increasing-earnings: {2,4,6};

earning-growth-rate: {10%-20%,20%-30%,>30%};

yrs-used-for-earning-growth: {3,6,9}

price-earning-ratio: {<6,6-10,10-15};

price-sales-ratio: {0-0.5,0.5-1,1-1.5};

dividend-yield-required: {0-2%,2%-4%,>4%};

dividend-growth-desired: {5%-8%,8%-12%,>12%};

no-consecutive-positive-years-dividend-growth: {5-10,10-15,>15,};

**END CONCEPT** customer-style-profile;

/\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/

/\* 4 RULE TYPEs \*/

/\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/

/\* 1st of 4 RULE TYPEs \*/

**RULE-TYPE** update-stock-risk;

**ANTECEDENT**: customer-risk-profile; CARDINALITY: l;

stock-risk-feature; CARDINALITY: l+;

**CONSEQUENT**: stock-risk-feature; CARDINALITY: 1+;

**CONNECTION-SYMBOL**: update-stock-risk;

**END RULE-TYPE** update-stock-risk**;**

/\* 2nd of 4 RULE TYPEs \*/

**RULE-TYPE** filter-stock-style;

**ANTECEDENT**: customer-style-profile;

**CARDINALITY**: l;

style-feature; CARDINALITY: l+;

**CONSEQUENT**: stock-style-feature: 1+;

**CONNECTION-SYMBOL**: filter-stock-style;

**END RULE-TYPE** filter-stock-style;

/\* 3rd of 4 RULE TYPEs \*/

**RULE-TYPE** update-stock-favourability;

**ANTECEDENT**: macroeconomic-and-stock-market-conditions;

CARDINALITY: l;

stock-attractiveness-feature-derived;

CARDINALITY: l+;

**CONSEQUENT**: stock-attractiveness-feature-derived;

CARDINALITY: 1+;

**CONNECTION-SYMBOL**: update-stock-favourability;

**END RULE-TYPE** update-stock-favourability;

/\* 4th of 4 RULE TYPEs \*/

**RULE-TYPE** recommend-stock;

**ANTECEDENT**: stock-list;

CARDINALITY: l;

**CONSEQUENT**: stock-list;

CARDINALITY: l;

**CONNECTION-SYMBOL**: recommend-stock;

**END RULE-TYPE** recommend-stock;

/\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/

**END DOMAIN-SCHEMA** stock-selection-schema;

/\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/

/\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/

/\* Detailed RULEs \*/

/\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/

/\* 1st of 4 KNOWLEDGE-BASEs \*/

**KNOWLEDGE-BASE** filter-stock-style-model;

USES:

filter-stock-style FROM filter-stock-style-kb;

/\* Filter By dividend Style \*/

customer-style-profile.style-selection = ‘dividend’

AND

/\* First By dividend Yield \*/

(customer-style-profile.dividend-yield-required = ‘0% - 2%’

filter-stock-style stock-list.div-yield% >=0 AND stock-list.div-yield% < 2

customer-style-profile.dividend-yield-required = ‘2% - 4%’

filter-stock-style stock-list.div-yield% >= 2

AND stock-list.div-yield% < 4

customer-style-profile.dividend-yield-required = ‘>=4%’

filter-stock-style CHARACTERISTICS stock-list.div-yield% >= 4

)

AND

/\* Second By minimum no of years of consecutive dividend growth \*/

(

customer-style-profile.no-consecutive-positive-

years-dividend-growth = ‘>=5 years’

filter-stock-style CHARACTERISTICS stock-list.pos-div-growth5y = TRUE

OR

customer-style-profile.no-consecutive-positive-

years-dividend-growth = ‘>=10 years’

filter-stock-style CHARACTERISTICS stock-list.pos-div-growth10y = TRUE

OR

customer-style-profile.no-consecutive-positive-

years-dividend-growth = ‘>=15 years’

filter-stock-style stock-list.pos-div-growth15y = TRUE

AND

/\* Third By dividend growth desired \*/

(customer-style-profile.dividend-growth-desired = ‘>=5%’

filter-stock-style CHARACTERISTICS stock-list.Divgrowth% >=5

OR

customer-style-profile.dividend-growth-desired = ‘>=8%’

filter-stock-style CHARACTERISTICS stock-list.Divgrowth% >=8

OR

customer-style-profile.dividend-growth-desired = ‘>=12%’

filter-stock-style stock-list.Divgrowth% >=12

)

/\* Filter By growth Style \*/

customer-style-profile.style-selection = ‘growth’

AND

/\* First By no of consecutive years of increasing earnings \*/

(customer-style-profile.no-consecutive-years-increasing-earnings = 2

filter-stock-style stock-list.pos-earning-growth2y = TRUE

OR

customer-style-profile.no-consecutive-years-

increasing-earnings = 4

filter-stock-style stock-list.pos-earning-growth4y = TRUE

OR

customer-style-profile.no-consecutive-years-

increasing-earnings = 6

filter-stock-style stock-list.pos-earning-growth6y = TRUE

)

AND

/\* Second By earnings growth rate \*/

(customer-style-profile.earnings-growth-rate = ‘10% - 20%’

AND

customer-style-profile.yrs-used-for-earnings-growth = ‘3 years’

filter-stock-style stock-list.earning-growth3y > 10 and stock-list.earning-growth3y <= 20

OR

(customer-style-profile.earnings-growth-rate = ‘20% - 30%’

AND

customer-style-profile.yrs-used-for-earnings-growth = ‘3 years’)

filter-stock-style stock-list.earning-growth3y > 20 and

stock-list.earning-growth3y <= 30

OR

(customer-style-profile.earnings-growth-rate = ‘>=30%’

AND

customer-style-profile.yrs-used-for-earnings-growth = ‘3 years’)

filter-stock-style stock-list.earning-growth3y >= 30%

OR

(customer-style-profile.earnings-growth-rate = ‘10% - 20%’

AND

customer-style-profile.yrs-used-for-earnings-growth = ‘6 years’)

filter-stock-style stock-list.earning-growth6y > 10 and

stock-list.earning-growth6y <= 20

OR

(customer-style-profile.earnings-growth-rate = ‘20% - 30%’

AND

customer-style-profile.yrs-used-for-earnings-growth = ‘6 years’)

filter-stock-style stock-list.earning-growth6y > 20 and

stock-list.earning-growth6y <= 30

OR

(customer-style-profile.earnings-growth-rate = ‘ >=30%’

AND

customer-style-profile.yrs-used-for-earnings-growth = ‘6 years’)

filter-stock-style stock-list.earning-growth6y >= 30

OR

(customer-style-profile.earnings-growth-rate = ‘10% - 20%’

AND

customer-style-profile.yrs-used-for-earnings-growth = ‘9 years’)

filter-stock-style stock-list.earning-growth9y > 10 and stock-list.earning-growth9y <= 20

OR

(customer-style-profile.earnings-growth-rate = ‘20% - 30%’

AND

customer-style-profile.yrs-used-for-earnings-growth = ‘9 years’)

filter-stock-style stock-list.earning-growth9y > 20 and stock-list.earning-growth9y <= 30

OR

(customer-style-profile.earnings-growth-rate = ‘ >=30%’

AND

customer-style-profile.yrs-used-for-earnings-growth = ‘9 years’)

filter-stock-style stock-list.earning-growth9y >= 30

/\* Filter By momentum Style \*/

customer-style-profile.style-selection = ‘momentum’

AND

(customer-style-profile.price-momentum-preference = ‘< 4 weeks’

filter-stock-style stock-list.price-chg%10d > 5 and stock-list.price-chg%1m > 10

OR

customer-style-profile.price-momentum-preference = ‘4 weeks - 12 weeks’

filter-stock-style stock-list.price-chg%1m > 10 and stock-list.price-chg%3m > 20

OR

customer-style-profile.price-momentum-preference = ‘> 12 weeks’

filter-stock-style stock-list.price-chg%3m > 20 and stock-list.price-chg%6m > 40

/\* Filter By value Style \*/

customer-style-profile.style-selection = ‘value’

AND

/\* First by price-earnings-ratio \*/

(customer-style-profile.price-earnings-ratio = ‘5-8’

filter-stock-style stock-list.pe-fy1 >5 and stock-list.pe-fy1 <=8

OR

customer-style-profile.price-earnings-ratio = ‘8-11’

filter-stock-style stock-list.pe-fy1 >8 and stock-list.pe-fy1 <=11

OR

customer-style-profile.price-earnings-ratio = ‘11-15’

filter-stock-style stock-list.pe-fy1 >11 and stock-list.pe-fy1 <=15

)

AND

/\* Second by price-sales-ratio \*/

(customer-style-profile.price-sales-ratio = ‘0 - 0.5’

filter-stock-style stock-list.p-sales-fy1 >0 and stock-list.p-sales-fy1 <=0.5

OR

customer-style-profile.price-sales-ratio = ‘0.5 - 1’

filter-stock-style stock-list.p-sales-fy1 >0.5 and stock-list.p-sales-fy1 <=11

OR

customer-style-profile.price-sales-ratio = ‘1 - 1.5’

filter-stock-style stock-list.p-sales-fy1 >1

and stock-list.p-sales-fy1 <=1.5

)

**END KNOWLEDGE-BASE** filter-stock-style-model;

/\* 2nd of 4 KNOWLEDGE-BASEs \*/

**KNOWLEDGE-BASE** update-stock-favourability-model;

USES:

update-stock-favourability FROM stock-selection-schema;

EXPRESSIONS:

/\* Abstraction rules \*/

inflation-rate**:** {bearish,bullish,neutral};

unemployment-rate: {bullish,neutral,bearish};

unemployment-trend: {uptrend,flat,downtrend}

average-weekly-unemployment-claim: {bullish,neutral,bearish};

unemployment-claim: {uptrend,flat,downtrend};

yield-curve-slope: {upward-sloping,flat,downward-sloping};

retail-sale: {uptrend,flat,downtrend};

pmi: {<50,>=50};

/\* Macroeconomic conditions \*/

/\* 1. Inflation Rate for the past 6 months \*/

macroeconomic-conditions.inflation-rate < 0%

OR

macroeconomic-conditions.inflation-rate > 4%

update-stock-favourability inflation-rate-cf = -0.8

macroeconomic-conditions.inflation-rate = ‘0% - 2%’

update-stock-favourability inflation-rate-cf = +0.9

macroeconomic-conditions.inflation-rate = ‘2% - 4%’

update-stock-favourability inflation-rate-cf **=** +0.5

/\* 2. Unemployment Rate over the past 6 months (Absolute) \*/

macroeconomic-conditions.unemployment-rate < 5%

update-stock-favourability unemployment-rate-cf = +0.7

macroeconomic-conditions.unemployment-rate = ‘5% - 6%’

update-stock-favourability unemployment-rate-cf = +0.5

macroeconomic-conditions.unemployment-rate > 6%

update-stock-favourability unemployment-rate-cf = +0.3

/\* 3. Unemployment trend over the past 6 months \*/

macroeconomic-conditions.unemployment-trend = uptrend

update-stock-favourability unemployment-trend-cf = -0.7

macroeconomic-conditions.unemployment-trend = flat

update-stock-favourability unemployment-trend-cf = +0.5

macroeconomic-conditions.unemployment-trend = downtrend

update-stock-favourability unemployment-trend-cf = +0.8

/\* 4. Average-weekly-unemployment-claim for the past 4 weeks \*/

macroeconomic-conditions.average-weekly-unemployment-claim < 350,000

update-stock-favourability unemployment-claim-cf = +0.7

macroeconomic-conditions.average-weekly-unemployment-claim =

‘350,000 - 400,000’

update-stock-favourability unemployment-claim-cf = +0.5

macroeconomic-conditions.average-weekly-unemployment-claim > 400,000

update-stock-favourability unemployment-claim-cf = -0.6

/\* 5. Unemployment-claims over the past 6 months - trend \*/

macroeconomic-conditions.unemployment-claim-trend = uptrend

update-stock-favourability unemployment-claim-trend-cf = -0.7

macroeconomic-conditions.unemployment-claim-trend = flat

update-stock-favourability unemployment-claim-trend-cf = +0.5

macroeconomic-conditions.unemployment-claim-trend = downtrend

update-stock-favourability unemployment-claim-trend-cf = +0.8

/\* 6. Yield-curve-slope \*/

macroeconomic-conditions.yield-curve-slope = upward-sloping

update-stock-favourability yield-curve-slope-cf = +0.9

macroeconomic-conditions.yield-curve-slope = flat

update-stock-favourability yield-curve-slope-cf = +0.5

macroeconomic-conditions.yield-curve-slope = downward-sloping

update-stock-favourability yield-curve-slope-cf = -0.8

/\* 7. Retail Sales Trend \*/

macroeconomic-conditions.retail-sale-trend = Uptrend

update-stock-favourability retail-sale-trend-cf = +0.8

macroeconomic-conditions.retail-sale-trend = Flat

update-stock-favourability retail-sale-trend-cf = +0.5

macroeconomic-conditions.retail-sale-trend = Downtrend

update-stock-favourability retail-sale-trend-cf = -0.7

/\* 8. ISM Manufacturing PMI \*/

macroeconomic-conditions.pmi > 50

update-stock-favourability pmi-cf = 0.7

macroeconomic-conditions.pmi <= 50

update-stock-favourability pmi-cf = 0.4

/\* Working Goal - macroeconomic-conditions \*/

inflation-rate-cf

AND

unemployment-rate-cf

AND

unemployment-trend-cf

AND

unemployment-claim-cf

AND

unemployment-claim-trend-cf

AND

yield-curve-slope-cf

AND

retail-sale-trend-cf

AND

pmi-cf

CREATE WORKING GOAL =

macroeconomic-and-stock-market-conditions.macroeconomic-conditions-cf

/\* stock market conditions \*/

/\* 1. Indicator 1 - *20D MA Ratio of no of positive returns/no of negative*

*returns* \*/

stock-market-conditions.indicator-1 = ‘0 - 0.3 (very-bearish)’

update-stock-favourability indicator-1-cf = -0.8

stock-market-conditions.indicator-1 = ‘0.3 - 0.8 (bearish)’

update-stock-favourability indicator-1-cf = -0.6

stock-market-conditions.indicator-1 = ‘0.8 - 1.2 (neutral)’

update-stock-favourability indicator-1-cf = +0.5

stock-market-conditions.indicator-1 = ‘1.2 - 2 (bullish)’

update-stock-favourability indicator-1-cf = +0.6

stock-market-conditions.indicator-1 = ‘>2 (very-bearish)’

update-stock-favourability indicator-1-cf = +0.8

/\* 2. Indicator 2 - *20D MA Ratio of no of daily +4% moves/no of -4% moves*

\*/

stock-market-conditions.indicator-2 = 0 - 0.3 (very-bearish)

update-stock-favourability indicator-2-cf = -0.8

stock-market-conditions.indicator-2 = 0.3 - 0.8 (bearish)

update-stock-favourability indicator-2-cf = -0.6

stock-market-conditions.indicator-2 = 0.8 - 1.2 (neutral)

update-stock-favourability indicator-2-cf = +0.5

stock-market-conditions.indicator-2 = 1.2 - 2 (bullish)

update-stock-favourability indicator-2-cf = +0.6

stock-market-conditions.indicator-2 = >2 (very-bearish)

update-stock-favourability indicator-2-cf = +0.8

/\* 3. Indicator 3 - *20D MA Ratio of no of quarterly +25% moves/no of -25%*

*moves* \*/

stock-market-conditions.indicator-3 = 0 - 0.3 (very-bearish)

update-stock-favourability indicator-3-cf = -0.8

stock-market-conditions.indicator-3 = 0.3 - 0.8 (bearish)

update-stock-favourability indicator-3-cf = -0.6

stock-market-conditions.indicator-3 = 0.8 - 1.2 (neutral)

update-stock-favourability indicator-3-cf = +0.5

stock-market-conditions.indicator-3 = 1.2 - 2 (bullish)

update-stock-favourability indicator-3-cf = +0.6

stock-market-conditions.indicator-3 = >2 (very-bearish)

update-stock-favourability indicator-3-cf = +0.8

/\* 4. Indicator 4 - *No of names above 200D MA/Number of names below 200D*

*MA \*/*

stock-market-conditions.indicator-4 = 0 - 0.3 (very-bearish)

update-stock-favourability indicator-4-cf = -0.8

stock-market-conditions.indicator-4 = 0.3 - 0.8 (bearish)

update-stock-favourability indicator-4-cf = -0.6

stock-market-conditions.indicator-4 = 0.8 - 1.2 (neutral)

update-stock-favourability indicator-4-cf = +0.5

stock-market-conditions.indicator-4 = 1.2 - 2 (bullish)

update-stock-favourability indicator-4-cf = +0.6

stock-market-conditions.indicator-4 = >2 (very-bearish)

update-stock-favourability indicator-4-cf = +0.8

/\* 5. Indicator 5 - *VIX (measure of implied volatility of S&P500 out of*

*the money options) \*/*

stock-market-conditions.indicator-5 = <10

OR (stock-market-conditions.indicator-5 = >=25

AND stock-market-conditions.indicator-5 = <35)

update-stock-favourability indicator-5-cf = -0.6

stock-market-conditions.indicator-5 = >=10

AND stock-market-conditions.indicator-5 = <18

update-stock-favourability indicator-5-cf = +0.8

stock-market-conditions.indicator-5 = >=18

AND stock-market-conditions.indicator-5 = <25

update-stock-favourability indicator-5-cf = +0**.5**

stock-market-conditions.indicator-5 = >=35

update-stock-favourability indicator-5-cf = -0.9

/\* Working Goal - stock-market-conditions indicators \*/

indicator-1-cf

AND

indicator-2-cf

AND

indicator-3-cf

AND

indicator-4-cf

AND

indicator-5-cf

CREATE WORKING GOAL =

macroeconomic-and-stock-market-conditions.stock-market-conditions-cf

/\* Working Goal - market favourability-cf \*/

macroeconomic-and-stock-market-conditions.macroeconomic-conditions-cf

AND

macroeconomic-and-stock-market-conditions.stock-market-conditions-cf

CREATE WORKING GOAL =

macroeconomic-and-stock-market-conditions.market-favourability-cf

/\* Determine market favourability \*/

macroeconomic-and-stock-market-conditions.market-favourability-cf >= 0.8

AND

macroeconomic-and-stock-market-conditions.market-favourability-cf < 1

update-stock-favourability market-favourability = high

macroeconomic-and-stock-market-conditions.market-favourability-cf >= 0.5

AND

macroeconomic-and-stock-market-conditions.market-favourability-cf < 0.8

update-stock-favourability market-favourability = medium

macroeconomic-and-stock-market-conditions.market-favourability-cf >= 0

AND

macroeconomic-and-stock-market-conditions.market-favourability-cf < 0.5

update-stock-favourability market-favourability = low

/\* Determine stock features attractiveness as a result of market

favourability \*/

/\* Determine stock cap category (Feature 1) attractiveness \*/

/\* High market favourability \*/

market-favourability = high

AND

(stock-list.cap-category = large

update-stock-favourability

stock-attractiveness-feature-inherent-derived.cap-category-attractiveness-

cf = 0.6

OR

stock-list.cap-category = mid

update-stock-favourability

stock-attractiveness-feature-inherent-derived.cap-category-attractiveness

-cf = 0.6

OR

stock-list.cap-category = small

update-stock-favourability

stock-attractiveness-feature-inherent-derived.cap-category-attractiveness

-cf = 0.8

)

/\* Medium market favourability \*/

market-favourability = medium

AND

(stock-list.cap-category = large

update-stock-favourability

stock-attractiveness-feature-inherent-derived.cap-category-attractiveness

-cf = 0.7

OR

stock-list.cap-category = mid

update-stock-favourability

stock-attractiveness-feature-inherent-derived.cap-category-attractiveness

-cf = 0.6

OR

stock-list.cap-category = small

update-stock-favourability

stock-attractiveness-feature-inherent-derived.cap-category-attractiveness

-cf = 0.6

)

/\* Low market favourability \*/

market-favourability = low

AND

(stock-list.cap-category= large

update-stock-favourability

stock-attractiveness-feature-inherent-derived.cap-category-attractiveness

-cf = 0.8

OR

stock-list.cap-category = mid

update-stock-favourability

stock-attractiveness-feature-inherent-derived.cap-category-attractiveness

-cf = 0.4

OR

stock-list.cap-category = small

update-stock-favourability

stock-attractiveness-feature-inherent-derived.cap-category-attractiveness

-cf = 0.2

)

/\* Determine sector category (Feature 2) attractiveness \*/

/\* High market favourability \*/

market-favourability = high

AND

(stock-list.sector = ’consumer-staples’

OR stock-list.sector = ’health-care’

OR stock-list.sector = ‘utility’)

update-stock-favourability

stock-attractiveness-feature-inherent-derived.sector-attractiveness-cf

= 0.4

market-favourability = high

AND

(stock-list.sector = ’materials’

OR stock-list.sector = ’industrials’

OR stock-list.sector = ‘Telecommunications’)

update-stock-favourability

stock-attractiveness-feature-inherent-derived.sector-attractiveness-cf

= 0.6

market-favourability = high

AND

(stock-list.sector = ’technology’

OR stock-list.sector = ’consumer-discretionary’

OR stock-list.sector = ‘financials’

OR stock-list.sector = ‘energy’)

update-stock-favourability

stock-attractiveness-feature-inherent-derived.sector-attractiveness-cf

= 0.8

/\* Medium market favourability \*/

market-favourability = medium

AND

(stock-list.sector = ’consumer-staples’

OR stock-list.sector = ’health-care’

OR stock-list.sector = ‘utility’)

update-stock-favourability

stock-attractiveness-feature-inherent-derived.sector-attractiveness-cf

= 0.5

market-favourability = medium

AND

(stock-list.sector = ’materials’

OR stock-list.sector = ’industrials’

OR stock-list.sector = ‘Telecommunications’)

update-stock-favourability

stock-attractiveness-feature-inherent-derived.sector-attractiveness-cf

= 0.5

market-favourability = medium

AND

(stock-list.sector = ’technology’

OR stock-list.sector = ’consumer-discretionary’

OR stock-list.sector = ‘financials’

OR stock-list.sector = ‘energy’)

update-stock-favourability

stock-attractiveness-feature-inherent-derived.sector-attractiveness-cf

= 0.7

/\* Low market favourability \*/

market-favourability = low

AND

(stock-list.sector = ’consumer-staples’

OR stock-list.sector = ’health-care’

OR stock-list.sector = ‘utility’)

update-stock-favourability

stock-attractiveness-feature-inherent-derived.sector-attractiveness-cf

= 0.9

market-favourability = low

AND

(stock-list.sector = ’materials’

OR stock-list.sector = ’industrials’

OR stock-list.sector = ‘Telecommunications’)

update-stock-favourability

stock-attractiveness-feature-inherent-derived.sector-attractiveness-cf

= 0.4

market-favourability = low

AND

(stock-list.sector = ’technology’

OR stock-list.sector = ’consumer-discretionary’

OR stock-list.sector = ‘financials’

OR stock-list.sector = ‘energy’)

update-stock-favourability

stock-attractiveness-feature-inherent-derived.sector-attractiveness-cf

= 0.2

/\* Stock attractiveness inherent market features \*/

/\* Valuation (PEG) \*/

stock-attractiveness-feature-inherent-market-feature.peg >= 0

And stock-attractiveness-feature-inherent-market-feature.peg < 1

update-stock-favourability peg-cf = 0.9

stock-attractiveness-feature-inherent-market-feature.peg >= 1

AND stock-attractiveness-feature-inherent-market-feature.peg < 1.7

update-stock-favourability peg-cf = 0.6

stock-attractiveness-feature-inherent-market-feature.peg >= 1.7

AND stock-attractiveness-feature-inherent-market-feature.peg < 2.4

update-stock-favourability peg-cf = 0.3

stock-attractiveness-feature-inherent-market-feature.peg >= 2.4

AND stock-attractiveness-feature-inherent-market-feature.peg < 3

update-stock-favourability peg-cf = -0.2

stock-attractiveness-feature-inherent-market-feature.peg < 0

OR stock-attractiveness-feature-inherent-market-feature.peg >= 3

update-stock-favourability peg-cf = -0.9

/\* EBITDA-Margin (%) \*/

stock-attractiveness-feature-inherent-market-feature.ebitda-margin >= 25

update-stock-favourability ebitda-margin-cf = 0.9

stock-attractiveness-feature-inherent-market-feature.ebitda-margin >= 18

AND stock-attractiveness-feature-inherent-market-feature.ebitda-margin < 25

update-stock-favourability ebitda-margin-cf = 0.7

stock-attractiveness-feature-inherent-market-feature.ebitda-margin >= 10

AND stock-attractiveness-feature-inherent-market-feature.ebitda-margin < 18

update-stock-favourability ebitda-margin-cf = 0.4

stock-attractiveness-feature-inherent-market-feature.ebitda-margin >= 0

AND stock-attractiveness-feature-inherent-market-feature.ebitda-margin < 10

update-stock-favourability ebitda-margin-cf = -0.2

stock-attractiveness-feature-inherent-market-feature.ebitda-margin < 0

update-stock-favourability ebitda-margin-cf = -0.7

/\* ROIC (Return on invested capital %) \*/

stock-attractiveness-feature-inherent-market-feature.roic > 15

update-stock-favourability roic-cf = 0.9

stock-attractiveness-feature-inherent-market-feature.roic >= 10

AND

stock-attractiveness-feature-inherent-market-feature.roic <= 15

update-stock-favourability roic-cf = 0.7

stock-attractiveness-feature-inherent-market-feature.roic >= 5

AND

stock-attractiveness-feature-inherent-market-feature.roic < 10

update-stock-favourability roic-cf = 0.3

stock-attractiveness-feature-inherent-market-feature.roic >= 0

AND

stock-attractiveness-feature-inherent-market-feature.roic < 5

update-stock-favourability roic-cf = -0.3

stock-attractiveness-feature-inherent-market-feature.roic < 0

update-stock-favourability roic-cf = 0.9

/\* Net debt equity ratio \*/

stock-attractiveness-feature-inherent-market-feature.net-debt-equity-ratio < -20

update-stock-favourability net-debt-equity-ratio-cf = 0.9

stock-attractiveness-feature-inherent-market-feature.net-debt-equity-ratio >= -20

AND

stock-attractiveness-feature-inherent-market-feature.net-debt-equity-ratio < 0

update-stock-favourability net-debt-equity-ratio-cf = 0.7

stock-attractiveness-feature-inherent-market-feature.net-debt-equity-ratio >= 0

AND

stock-attractiveness-feature-inherent-market-feature.net-debt-equity-ratio < 20

update-stock-favourability net-debt-equity-ratio-cf = 0.4

stock-attractiveness-feature-inherent-market-feature.net-debt-equity-ratio >= 20

AND

stock-attractiveness-feature-inherent-market-feature.net-debt-equity-ratio < 40

update-stock-favourability net-debt-equity-ratio-cf = -0.4

stock-attractiveness-feature-inherent-market-feature.net-debt-equity-ratio >= 40

update-stock-favourability net-debt-equity-ratio-cf = -0.8

/\* Compute stock-attractiveness-inherent-market-feature-cf \*/

peg-cf

ebitda-margin-cf

roic-cf

net-debt-equity-ratio-cf

AVERAGES TO

stock-attractiveness-inherent-market-feature.stock-attractiveness-inherent-market-feature-cf

/\* Determine stock attractiveness inherent derived due to market favourability \*/

stock-list.risk-match-cf

AND

stock-attractiveness-inherent-derived.cap-category-attractiveness-cf

AND

stock-attractiveness-inherent-derived.sector-attractiveness-cf

AVERAGES TO

stock-attractiveness-inherent-derived.stock-attractiveness-from-

market-favourability-cf

/\* Determine overall stock attractiveness \*/

stock-attractiveness-inherent-market-feature.stock-attractiveness-inherent-market-feature-cf

AND

stock-attractiveness-inherent-derived.stock-attractiveness-from-

market-favourability-cf

AVERAGES TO stock-list.stock-attractiveness-cf

**END KNOWLEDGE-BASE** update-stock-favourability-model;

/\* 3rd of 4 KNOWLEDGE-BASEs \*/

**KNOWLEDGE-BASE** update-stock-risk-model

USES:

update-stock-risk FROM update-stock-risk-kb;

/\* market capitalization size\* /

stock-risk-feature.cap-size = large

update-stock-risk cap-size-cf = 0.9

stock-risk-feature.cap-size = mid

update-stock-risk cap-size-cf = 0.5

stock-risk-feature.cpa-size = small

update-stock-risk cap-size-cf = 0.3

/\* liquidity\* /

stock-risk-feature.liquidity >= 500000

update-stock-risk liquidity-cf = 0.7

stock-risk-feature.liquidity <500000

update-stock-risk liquidity-cf = 0.3

/\* sector \*/

stock-risk-feature.sector = consumer-staples

OR stock-risk-feature.sector = health-care

OR stock-risk-feature.sector = utility

update-stock-risk sector-cf = 0.8

stock-risk-feature.sector = materials

OR stock-risk-feature.sector = industrials

update-stock-risk sector-cf = 0.5

stock-risk-feature.sector = technology

OR stock-risk-feature.sector = consumer-discretionary

OR stock-risk-feature.sector = financials

update-stock-risk sector-cf = 0.3

/\* price volatility \*/

stock-risk-feature.price-volatility <=18%

update-stock-risk price-volatility-cf = 0.9

stock-risk-feature.price-volatility >18%

AND stock-risk-feature.price-volatility <=24%

update-stock-risk price-volatility-cf = 0.7

stock-risk-feature.price-volatility >24%

AND stock-risk-feature.price-volatility <=40%

update-stock-risk price-volatility-cf = 0.5

stock-risk-feature.price-volatility >40%

AND stock-risk-feature.price-volatility <= 60%

update-stock-risk price-volatility-cf = -0.1

stock-risk-feature.price-volatility >60%

update-stock-risk price-volatility-cf = -0.6

/\* earning volatility \*/

stock-risk-feature.earning-volatility <=14%

update-stock-risk earning-volatility-cf = 0.9

stock-risk-feature.earning-volatility >14%

AND

stock-risk-feature.earning-volatility <=24%

update-stock-risk earning-volatility-cf = 0.7

stock-risk-feature.earning-volatility >24%

AND

stock-risk-feature.earning-volatility <=40%

update-stock-risk earning-volatility-cf = 0.4

stock-risk-feature.earning-volatility >40%

AND

stock-risk-feature.earning-volatility <=60%

update-stock-risk earning-volatility-cf = -0.2

stock-risk-feature.earning-volatility >60%

update-stock-risk earning-volatility-cf = -0.7

/\* Compute stock-risk-cf \*/

market-cap-cf

liquidity-cf

price-volatility-cf

earning-volatility-cf

AVERAGES TO stock-risk-feature.stock-risk-cf

/\*Determine customer’s overall risk tolerance\* /

/\* Risk ability \*/

/\* Age\*/

customer-risk-profile-ability.age = ‘20 - 35’

update-stock-risk age-cf = 0.8

customer-risk-profile-ability.age = ‘36 - 45’

update-stock-risk age-cf = 0.6

customer-risk-profile-ability.age = ‘46 - 55’

update-stock-risk age-cf = 0.4

customer-risk-profile-ability.age = ‘56 - 65’

update-stock-risk age-cf = -0.2

customer-risk-profile-ability.age = ‘>65’

update-stock-risk age-cf = -0.4

/\* Net assets \*/

customer-risk-profile-ability.net-assets = ‘<$25,000’

update-stock-risk net-assets-cf = -0.2

customer-risk-profile-ability.net-assets = ‘$25,000 - $100,000’

update-stock-risk net-assets-cf = 0.2

customer-risk-profile-ability.net-assets = ‘$100,000 - $250,000’

update-stock-risk net-assets-cf = 0.4

customer-risk-profile-ability.net-assets = ‘$250,000 - $500,000’

update-stock-risk net-assets-cf = 0.6

customer-risk-profile-ability.net-assets = ‘>$500,000’

update-stock-risk net-assets-cf = 0.8

/\* Job stability \*/

customer-risk-profile-ability.job-stability = ‘very-unstable’

update-stock-risk job-stability-cf = -0.2

customer-risk-profile-ability.job-stability = ‘rather-unstable’

update-stock-risk job-stability-cf = 0.2

customer-risk-profile-ability.job-stability = ‘a-bit-unstable’

update-stock-risk job-stability-cf = 0.4

customer-risk-profile-ability.job-stability = ‘stable’

update-stock-risk job-stability-cf = 0.6

customer-risk-profile-ability.job-stability = ‘very-stable’

update-stock-risk job-stability-cf = 0.8

/\* Monthly income \*/

customer-risk-profile-ability.monthly-income = ‘<$2,500’

update-stock-risk monthly-income-cf = -0.2

customer-risk-profile-ability.monthly-income = ‘$2,500 - $5,000’

update-stock-risk monthly-income-cf = 0.2

customer-risk-profile-ability.monthly-income = ‘$5,000 - $7,500’

update-stock-risk monthly-income-cf = 0.4

customer-risk-profile-ability.monthly-income = ‘$7,500 - $10,000’

update-stock-risk monthly-income-cf = 0.6

customer-risk-profile-ability.monthly-income = ‘>$10,000’

update-stock-risk monthly-income-cf = 0.8

/\* Liquidity needs \*/

customer-risk-profile-ability.liquidity-needs = ‘very-low’

update-stock-riskliquidity-needs-cf = -0.2

customer-risk-profile-ability.liquidity-needs = ‘low’

update-stock-riskliquidity-needs-cf = 0.2

customer-risk-profile-ability.liquidity-needs = ‘medium’

update-stock-riskliquidity-needs-cf= 0.4

customer-risk-profile-ability.liquidity-needs = ‘high’

update-stock-riskliquidity-needs-cf = 0.6

customer-risk-profile-ability.liquidity-needs = ‘very-high’

update-stock-riskliquidity-needs-cf = 0.8

/\* Compute customer risk ability cf \*/

age-cf

net-assets-cf

job-stability-cf

monthly-income-cf

liquidity-needs-cf

AVERAGES TOcustomer-risk-profile.risk-tolerance-ability-cf

/\* Risk willingness \*/

/\* Thrill factor \*/

customer-risk-profile-willingness.thrill-factor = very-low

update-stock-riskthrill-factor-cf = -0.5

customer-risk-profile-willingness.thrill-factor = low

update-stock-riskthrill-factor-cf = 0.2

customer-risk-profile-willingness.thrill-factor = medium

update-stock-riskthrill-factor-cf = 0.4

customer-risk-profile-willingness.thrill-factor = high

update-stock-riskthrill-factor-cf = 0.6

customer-risk-profile-willingness.thrill-factor = very-high

update-stock-riskthrill-factor-cf = 0.8

/\* Job security preference \*/

customer-risk-profile-willingness.job-security-preference = very-high

update-stock-risk job-security-preference.cf = -0.5

customer-risk-profile-willingness.job-security-preference = high

update-stock-risk job-security-preference.cf = 0.2

customer-risk-profile-willingness.job-security-preference = medium

update-stock-risk job-security-preference.cf = 0.4

customer-risk-profile-willingness.job-security-preference = low

update-stock-risk job-security-preference.cf = 0.6

customer-risk-profile-willingness.job-security-preference = very-low

update-stock-risk job-security-preference.cf = 0.8

/\* Investment risk tolerance \*/

customer-risk-profile-willingness.investment-risk-tolerance = very-low

update-stock-risk investment-risk-tolerance-cf = -0.2

customer-risk-profile-willingness.investment-risk-tolerance = low

update-stock-risk investment-risk-tolerance-cf = 0.1

customer-risk-profile-willingness.investment-risk-tolerance = medium

update-stock-risk investment-risk-tolerance-cf = 0.2

customer-risk-profile-willingness.investment-risk-tolerance = high

update-stock-risk investment-risk-tolerance-cf = 0.4

customer-risk-profile-willingness.investment-risk-tolerance = very-high

update-stock-risk investment-risk-tolerance-cf = 0.6

customer-risk-profile-willingness.investment-risk-tolerance = Very, very-high

update-stock-risk investment-risk-tolerance-cf = 0.8

/\* Job compensation preference \*/

customer-risk-profile-willingness.job-compensation-preference = ‘all-salary’

update-stock-risk job-compensation-preference-cf = -0.2

customer-risk-profile-willingness.job-compensation-preference = ‘mainly-salary’

update-stock-risk job-compensation-preference-cf = 0.1

customer-risk-profile-willingness.job-compensation-preference = ‘equal-mix-of-salary-and-commission’

update-stock-risk job-compensation-preference-cf = 0.2

customer-risk-profile-willingness.job-compensation-preference = ‘mainly-commission’

update-stock-risk job-compensation-preference-cf = 0.5

customer-risk-profile-willingness.job-compensation-preference = ‘all-commission’

update-stock-risk job-compensation-preference-cf = 0.7

/\* Aversion to previous stock loss \*/

customer-risk-profile-willingness.aversion-to-previous-stock-loss = very-low

update-stock-risk aversion-to-previous-stock-loss-cf = -0.2

customer-risk-profile-willingness.aversion-to-previous-stock-loss = low

update-stock-risk aversion-to-previous-stock-loss-cf = 0.1

customer-risk-profile-willingness.aversion-to-previous-stock-loss = medium

update-stock-risk aversion-to-previous-stock-loss-cf = 0.2

customer-risk-profile-willingness.aversion-to-previous-stock-loss = high

update-stock-risk aversion-to-previous-stock-loss-cf = 0.5

customer-risk-profile-willingness.aversion-to-previous-stock-loss = very-high

update-stock-risk aversion-to-previous-stock-loss-cf = 0.7

/\* Compute customer risk willingness cf \*/

thrill-factor-cf

AND

job-security-preference-cf

AND

investment-risk-tolerance-cf

AND

job compensation-preference-cf

AND

aversion-to-previous-stock-loss-cf

AVERAGES TOcustomer-risk-profile.risk-tolerance-willingness-cf

/\* Compute customer overall risk tolerance cf \*/

customer-risk-profile.risk-tolerance-ability-cf

AND

customer-risk-profile.risk-tolerance-willingness-cf

CREATES WORKING GOAL = customer-risk-profile.overall-risk-tolerance-cf

/\* Determine risk match \*/

customer-risk-profile.overall-risk-tolerance-cf

AND

stock-risk-feature.stock-risk-cf

update-stock-risk stock-list.risk-match-cf

**END KNOWLEDGE-BASE** update-stock-risk-model

/\* 4th of 4 KNOWLEDGE-BASEs \*/

**KNOWLEDGE-BASE** update-recommendation-rank-model

USES:

update-recommendation-rank FROM update-recommendation-rank-kb;

/\* Determine recommmendation rank based on risk match and stock attractiveness \*/

stock-list.risk-match-cf

AND

stock-list.stock-attractiveness-cf

recommend-stock stock-list.recommendation-rank

**END KNOWLEDGE-BASE** update-recommendation-rank-model

/\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/

**END DOMAIN-KNOWLEDGE** us-stock-domain

/\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/

/\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/

**INFERENCE-KNOWLEDGE** us-stock-recommendation;

/\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/

/\* 1st of 4 INFERENCEs: filter-stock-style \*/

KNOWLEDGE-ROLE customer-style-profile;

TYPE: DYNAMIC;

DOMAIN-MAPPING: customer-style-profile;

END KNOWLEDGE-ROLE customer-style-profile;

KNOWLEDGE-ROLE stock-style-feature;

TYPE: DYNAMIC;

DOMAIN-MAPPING: stock-style-feature SET OF stock-list;

END KNOWLEDGE-ROLE stock-style-feature;

**KNOWLEDGE-ROLE** filter-stock-style-model;

**TYPE**: STATIC;

**DOMAIN-MAPPING**: filter-stock-style FROM filter-stock-style-kb;

**END KNOWLEDGE-ROLE** filter-stock-style-model;

**INFERENCE** filter-stock-style**;**

ROLES:

INPUT: customer-style-profile;

stock-style-feature;

OUTPUT: stock-style-feature;

STATIC: filter-stock-style;

SPECIFICATION: “Input1 is a set of customer style profile (customer level), Input2 is a set of stock style features (stock level). Output is the same set of stock list rated with style features (stock level).";

**END INFERENCE** filter-stock-style**;**

/\* 2nd of 4 INFERENCEs: update-stock-risk \*/

KNOWLEDGE-ROLE customer-risk-profile;

TYPE: DYNAMIC;

DOMAIN-MAPPING: customer-risk-profile;

END KNOWLEDGE-ROLE customer-risk-profile;

KNOWLEDGE-ROLE stock-risk-feature;

TYPE: DYNAMIC;

DOMAIN-MAPPING: stock-risk-feature SET OF stock-list;

END KNOWLEDGE-ROLE stock-risk-feature;

**KNOWLEDGE-ROLE** update-stock-risk-model;

**TYPE**: STATIC;

**DOMAIN-MAPPING**: update-stock-risk FROM update-stock-risk-kb;

**END KNOWLEDGE-ROLE** update-stock-risk-model;

**INFERENCE** update-stock-risk;

ROLES:

INPUT: customer-risk-profile;

stock-risk-feature;

OUTPUT: stock-risk-feature;

STATIC: update-stock-risk;

SPECIFICATION: “Input1 is a set of customer risk profile (customer level), Input2 is a set of stock risk features (stock level). Output is the same set of stock list rated with risk features (stock level).";

**END INFERENCE** update-stock-risk;

/\* 3rd of 4 INFERENCEs: update-stock-favourability \*/

KNOWLEDGE-ROLE macroeconomic-and-stock-market-conditions;

TYPE: DYNAMIC;

DOMAIN-MAPPING: macroeconomic-and-stock-market-conditions;

END KNOWLEDGE-ROLE macroeconomic-and-stock-market-conditions;

KNOWLEDGE-ROLE stock-attractiveness-feature-inherent-Market-Feature;

TYPE: DYNAMIC;

DOMAIN-MAPPING: stock-attractiveness-feature-inherent-Market-Feature SET OF stock-list;

END KNOWLEDGE-ROLE stock-attractiveness-feature-inherent-Market-Feature;

KNOWLEDGE-ROLE stock-attractiveness-feature-inherent-derived;

TYPE: DYNAMIC;

DOMAIN-MAPPING: stock-attractiveness-feature-inherent-serived SET OF stock-list;

END KNOWLEDGE-ROLE stock-attractiveness-feature-inherent-derived;

**KNOWLEDGE-ROLE** update-stock-favourability-model;

**TYPE**: STATIC;

**DOMAIN-MAPPING**: update-stock-favourability FROM

update-stock-favourability-kb;

**END KNOWLEDGE-ROLE** update-stock-favourability-model;

**INFERENCE** update-stock-favourability;

ROLES:

INPUT: macroeconomic-and-stock-market-conditions;

stock-attractiveness-feature-inherent-Market-Feature;

OUTPUT: stock-attractiveness-feature-inherent-derived;

STATIC: update-stock-favourability;

SPECIFICATION: “Input1 is a set of macroeconomic-and-stock-market-conditions (world level & stock market/sector level), Input2 is a set of stock Favourability (attractiveness) features (stock level). Output is the same set of stock list rated with Favourability features (stock level).";

**END INFERENCE** update-stock-favourability;

/\* 4th of 4 INFERENCEs: update-recommendation-rank \*/

KNOWLEDGE-ROLE stock-risk-feature;

TYPE: DYNAMIC;

DOMAIN-MAPPING: stock-risk-feature SET  
OF stock-list;

END KNOWLEDGE-ROLE stock-risk-feature;

KNOWLEDGE-ROLE stock-attractiveness-feature-inherent-derived;

TYPE: DYNAMIC;

DOMAIN-MAPPING: stock-attractiveness-feature-inherent-derived  
SET OF stock-list;

END KNOWLEDGE-ROLE stock-attractiveness-feature-inherent-derived;

**KNOWLEDGE-ROLE** update-recommendation-rank-model;

**TYPE**: STATIC;

**DOMAIN-MAPPING**: update-recommendation-rank

FROM update-recommendation-rank-kb;

**END KNOWLEDGE-ROLE** update-recommendation-rank-model;

**INFERENCE** update-recommendation-rank;

ROLES:

INPUT: stock-risk-feature;

stock-attractiveness-feature-inherent-derived;

OUTPUT: stock-risk;

STATIC: update-recommendation-rank;

SPECIFICATION:  
“Input1 is a set of stock risk feature (stock level), Input2 is a set of stock favourability  
features (stock level). Output is the same set of stock list rated with final/overall  
Recommendation Rank (stock level).";

**END INFERENCE** update-recommendation-rank;

/\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/

**END INFERENCE-KNOWLEDGE** us-stock-recommendation;

/\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/

/\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/

**TASK-KNOWLEDGE** us-stock-recommendation-task;

/\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/

/\* Main Task & Task Method: us-stock-assessment-process-task \*/

/\* Main Task & Task Method have 4 respective SUB-Task & Task Method\*/

**TASK** us-stock-assessment-process-task;

DOMAIN-NAME: us-stock-domain;

GOAL: “Update/Rank all the stock in stock list base on Customer Risk profile, Customer Style profile, and parameters of Macroeconomic-and-Stock-market-conditions, ”;

ROLES:

INPUT: stock-list;

customer-risk-profile SET OF customer-profile;

customer-style-profile SET OF customer-profile;

macroeconomic-and-stock-market-conditions;

OUTPUT: stock-list.recommendation-rank: “with Recommendation Rank";

END TASK us-stock-assessment-process;**;**

**TASK-METHOD** us-stock-assessment-process-task-method**;**

REALIZES: us-stock-assessment-process-rask;

DECOMPOSITION:

TASKS: filter-stock-style, update-stock-risk, update-stock-favourability, update-recommendation-rank;

ROLES:

INTERMEDIATE:

stock-list.recommendation-rank: “with Recommendation Rank";

CONTROL-STRUCTURE:

filter-stock-style(stock-list.stock-style-feature + customer-profile.customer-style-profile -> stock-list.style-selection);

update-stock-risk(stock-list.stock-risk-feature + customer-profile.customer-risk-profile -> stock-list.risk-match-cf);

update-stock-favourability(stock-list.stock-attractiveness-feature-inherent-market-feature + macroeconomic and stock-market-conditions -> stock-list.stock-attractiveness-cf);

update-recommendation-rank(stock-list.risk-match-cf + stock-list.stock-attractiveness-cf -> stock-list.recommendation-rank);

**END TASK-METHOD** us-stock-assessment-process-task-method**;**

/\* 1st of 4 SUB Task/Method of Main Task & Task Method: filter-stock-style \*/

**TASK** filter-stock-style**;**

DOMAIN-NAME: us-stock-domain;

GOAL: "Compare between the stock-style-feature and customer-style-features, identify and retain stock in stock-list if the stock-style-feature matches customer-style-features.";

ROLES:

INPUT: stock-style-feature SET OF stock-list;

customer-style-profile SET OF customer-profile;

OUTPUT: stock-list.style-selection: "dividend, growth, momentum, value";

**END TASK** filter-stock-style**;**

**TASK-METHOD** filter-stock-style-task-method;

REALIZES: filter-stock-style;

DECOMPOSITION:

INFERENCES: filter-stock-style;

ROLES:

INTERMEDIATE:

stock-list.style-selection: "dividend , growth, momentum, value";

CONTROL-STRUCTURE:

WHILE NOT END OF stock-list DO

IF stock-list.style-selection =

customer-style-profile.style-selection;

RETAIN the stock in stock-list

ELSE

REMOVE the stock from stock-list

END WHILE

**END TASK-METHOD** filter-stock-style-task-method;

/\* 2nd of 4 SUB Task/Method of Main Task & Task Method: update-stock-risk \*/

**TASK** update-stock-risk;

DOMAIN-NAME: us-stock-domain;

GOAL: "Compare between the stock-risk-feature and customer-risk-features, identify and update stock-risk-feature when it’s matched with customer-risk-features.";

ROLES:

INPUT: stock-risk-feature SET OF stock-list;

customer-risk-profile SET OF customer-profile;

OUTPUT: stock-list.risk-match-cf: {-1 - 1};

**END TASK update-stock-risk;**

**TASK-METHOD update-stock-risk-task-method;**

REALIZES: update-stock-risk;

DECOMPOSITION:

INFERENCES: update-stock-risk;

ROLES:

INTERMEDIATE:

stock-list.risk-match-cf: {-1 - 1};

CONTROL-STRUCTURE:

WHILE NOT END OF stock-list DO

stock-list.risk-match-cf := normalized-gap between customer-risk-profile and stock-list.stock-risk-feature;

END WHILE

**END TASK-METHOD** update-stock-risk-task-method;

/\* 3rd of 4 SUB Task/Method of Main Task & Task Method: update-stock-favourability \*/

**TASK** update-stock-favourability;

DOMAIN-NAME: us-stock-domain;

GOAL: "Compare between stock-list.stock-attractiveness-feature-inherent-Market-Feature and stock attractiveness due to macroeconomic-and-stock-market-conditions, to update stock-attractiveness certainty factor.";

ROLES:

INPUT:

stock-attractiveness-feature-inherent-market-feature SET OF stock-list

macroeconomic-and-stock-market-conditions

OUTPUT: stock-list.stock-attractiveness-cf: {-1 - 1};

**END TASK** Update-Stock-Favourability;

**TASK-METHOD** update-stock-favourability-task-method;

REALIZES: update-stock-favourability;

DECOMPOSITION:

INFERENCES: update-stock-favourability;

ROLES:

INTERMEDIATE:

stock-list.stock-attractiveness-cf: {-1 - 1};

CONTROL-STRUCTURE:

WHILE NOT END OF stock-list DO

stock-list.stock-attractiveness-certainty-factor:= normalized-gap between macroeconomic-and-stock-market-conditions and stock-list.stock-attractiveness-feature-inherent-market-feature;

END WHILE

**END TASK-METHOD** update-stock-favourability-task-method;

/\* 4th of 4 SUB Task/Method of Main Task & Task Method: update-stock-recommendation-rank \*/

**TASK** update-stock-recommendation-rank;

DOMAIN-NAME: us-stock-domain;

GOAL: " compute the final stock-list.recommendation-rank based on CF merging between: stock-list.risk-match-certainty-factor and stock-list.stock-attractiveness-certainty-factor.";

ROLES:

INPUT: stock-list.risk-match-cf;

stock-list.stock-attractiveness-cf;

OUTPUT: stock-list.recommendation-rank: {-1 - 1};

**END TASK update-stock-recommendation-rank;**

**TASK-METHOD update-stock-recommendation-rank-task-method;**

REALIZES: update-stock-recommendation-rank;

DECOMPOSITION:

INFERENCES: update-stock-recommendation-rank;

ROLES:

INTERMEDIATE:

stock-list.recommendation-rank: {-1 - 1};

CONTROL-STRUCTURE:

WHILE NOT END OF stock-list DO

stock-list.recommendation-rank-match := CF-merger between stock-list.risk-match-cf and stock-list.stock-attractiveness-cf;

END WHILE

**END TASK-METHOD** update-stock-recommendation-rank-task-method;

/\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/

**END TASK-KNOWLEDGE** us-stock-recommendation-task;

/\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/

/\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/

**END KNOWLEDGE-MODEL** us-stock;

/\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/

/\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\* E N D \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/

# Appendix C : Knowledge Acquisition Transcript

*Following is the summary of the transcript of the knowledge elicitation and acquisition process.*

**1. We will like to find out more about investing in stocks. What are the advantages and disadvantages of investing the stock market?**

The stock market is a great way of putting your capital to work for you. In the long run, the right stocks will create returns that should more than sufficiently beat inflation and help to build a portfolio of wealth that should put one in a strong financial situation and possibly help with retirement needs. It should preferably at least be considered in a part of anyone’s portfolio given its strong return potential compared to other instruments such as savings or even bonds. That said, the stock market is a treacherous place and is not for the fainted hearted with its constant swing and gyrations and even occasional market crash. If one chooses the wrong stocks or invest at the wrong time, he can even cause his wealth to erode rapidly. As such, it is important that one takes the trouble to educate himself about how best to go about investing in stocks. The effort in doing so can potentially pay dividends in the long term.

**2. We have heard that the U.S market is a great place to start an investing journey. Is that true? What advantages does the U.S market offer?**

Indeed, the U.S market is one of the most attractive, if not the most attractive market in the world to invest in. There are several advantages that I can mention here. They include the fact that the U.S market is the most highly capitalized market in the world, highly liquid and efficient and is home to many of the top companies in the world such as McDonalds, Google, Starbucks, Microsoft and Procter and Gamble to name a few. Any investor should be able to find his niche or favoured method of investing in this market given such a wide range of choices.

**3. Is the stock market for everyone? Are there any personal circumstances that one should evaluate when considering an investment in the stock market?**

Well, I guess it is not for everyone. As the stock market can involve a high level of risk, it is definitely not for the uninformed and risk adverse person. What I mean is if you do not understand what to look up for when selecting stocks and rely frequently on hearsay or advice from friends who are equally as clueless as you are, then you might be better off not touching it. When considering any investment in the stock market, it is important to first identify traits that will have a bearing on the type of risk that you can undertake. I will say there are two main components to evaluate one’s overall risk tolerance. There are his ability to take risk as well as his willingness to take risk. Ability to take risk will include factors such as his age, his personal wealth, future liquidity needs and their size, stability of his job and size of his paycheck and so on. This aspect of risk tolerance has more to do with a person’s existing circumstances. On the other hand, willingness to take risk is a person’s inherent characteristic and can be determined based on how risk adverse the person is when it comes to losing money, the amount of drawdown in his stock trading account he can accept before feeling very uncomfortable, how often he invest simply for the thrill of seeing his stock trading account goes up in value or how adverse he is towards making a new investment in a stock which he previously lost money in but which appears to have huge potential again.

**4. Can stocks be classified as having different level of risks? What can affect the risks affecting a stock?**

Definitely. Various attributes of a stock can affect its riskiness. Not all stocks are the same in terms of return and risk profile. Factors that can affect stock risk include its market capitalization (Generally, large cap stocks are safer than small cap stocks not least because they are more liquid, such to less manipulation because of their large size as well as being more well established and less prone to failures.), the sector it is in (Consumer staples, health care, utility stocks are generally safer compared to technology and financials or even consumer discretionary stocks whose performance can vary closely to the state of the economy). Earnings volatility over the years is an important factor as well as a high value on this measure means that the stock can be highly cyclical and very dependent on the economy, causes it to possibly have more wild swings in price. Likewise, price volatility can also affect stock risk as it may potentially mean a large fall in price.

**5. What are the criteria that can be used to select a stock investment or how attractive it is?**

This is where fundamental analysis comes in. It is the favoured method used by many investors, including investment greats like warren buffet and peter lynch. It considers various components such as the financial health of the company, how profitability the company is as well as how reasonably valued it is. In considering financial health, the amount of debt relative to equity can determine if a company is likely to experience financial hardship when times are hard. In considering profitability, profit margins and return on invested capital can indicate how good the company is in turning each dollar of sales into actual earnings and how much earnings the company is generating on the capital used. Finally, even with strong profitability and financial health, it is important to understand if the stock of the company is reasonably valued and whether there is continued good growth going forward. After all, a good company does not necessarily mean a good stock and a good company might be unnecessarily bid up in price and becomes too expensive. A good valuation metric that combines both valuation and growth is the price-earnings growth ratio which measures the price-earnings per percentage of growth. Typically, a PEG under 1 is considered to be a favourable value.

**6. Are there various styles of investing in the stock market? If so, what are they and what should an investor look out for?**

Yes. There are different ways that an investor can play the stock market. There are quite a few dominant themes in the stock market. There are themes such as value investing, growth investing, dividend investing as well as momentum trading. Each has its merits and suits different kinds of investors. In value investing, the key is to identify stocks which are cheap so that there is a chance of buying them low and selling them high. That said, prices can continue going lower without necessarily turning higher and other fundamental criteria need to be evaluated as well. Metrics to determine valuation can be price- earnings ratio, price-sales ratio, price-cashflow ratio etc. For price earnings ratio, a value of 15 is considered reasonable and a value below that can be considered cheap. For price-sales ratio, 1.5 is about fair and below 1 is considered cheap. For growth investing, the key is to identify stocks that experience earnings growth year over year. These stocks can experience a good surge in price especially in earnings continue to stay strong in the coming years. However, growth stocks tend to have higher expectations embedded in them and as such can turn lower in price quickly if they do not perform as well as expected in earnings growth in future. In dividend investing, an investor identifies stocks that distributes a good amount of dividends every year and preferably increase its dividend year over year over the past years. This means the investor can enjoy much profit in terms of dividends over the years with a low cost basis even as the price does not appreciate much. Finally, in momentum trading, the investor’s tries to ride on strong upwards price momentum for a relatively short period of time without necessarily considering the company’s underlying fundamental characteristics and instead rely on strong price movement. The investor has to be nimble enough to get out of the market when the tide turns against him as momentum can fade as quickly as it goes up.

**7. Are there external factors that can affect the price of a stock besides the inherent characteristics of each stock?**

Yes. While the stock market averages tend to move up over a long period of time, there exist secular bull and bear markets along the way. One should vary his allocation to stocks according to how favourable the market is. To have a sense of whether the market is favourable, two broad factors can be taken into consideration. First, macroeconomic factors should be considered. This will affect the corporate earnings of companies, how much companies are willing to invest and hence the price of their stocks. There are various important economic factors to consider, mainly the leading indicators. They include the inflation rate, the interest rate which can be determined by the yield curve and fed reserve actions, the unemployment rate and its recent month-over-month trend, weekly unemployment claims, trend of retail sales and manufacturing PMI. All these factors can foretell potential turning points in the stock market and alert the investor to the possibility of reducing exposure to equities. The second broad factor more directly measures stock market conditions and the underlying bullishness of the market. This has to do with aggregates of metrics relating to stock prices. They include measures such as 20 Day moving average of no of positive returns versus negative returns, 20 Day moving average of no of +4% moves versus -4% moves, 20 Day moving average of quarterly +25% versus -25% moves and ratio of no of names above 200 Day moving average versus no of names below 200 Day moving average. VIX which is the implied volatility of various S&P500 out of the money options otherwise known as the fear factor is also a good gauge of a potential turning point in the stock market.

**8. Does market favourability impact different types of stock differently? Are they stocks which generally perform better during bull markets and versus during bear markets?**

Yes. When the market is favourable, it is better to invest in cyclical stocks. Such stocks should be in sectors such as consumer discretionary (including housing stocks), technology and financials. On the other hand, when the market is not favourable, it is better to invest in defensive stocks in sectors such as health care, utilities and consumer staples. The stocks in these sectors tend to do better because they deal with products that are still consumed whether or not times are good. Also, small cap stocks will do better in a strong market environment compared to large cap stocks because investors are more optimistic and willing to take on more risk when they invest in stocks.