CPF INVESTMENT ADVISOR

INTELLIGENT REASONING SYSTEMS PROJECT REPORT

IRS – GROUP 8
KAH GHI LIM (A0100172)
YAN BO NG (A0186871N)
NIRAV PARIKH (A0213573J)

Contents

1.0 Executive Summary

Life expectancies over the world have generally increased over the years, allowing us to be able to spend more time with loved ones and enjoy various activities. Most importantly, there are more years to enjoy after retirement. To ensure the desired standard of living during our golden years, a significant retirement nest egg has to be maintained.

Though the Singapore government has implemented a compulsory social security savings scheme for every working Singapore Citizen and Permanent Resident, it would be good to further boost our retirement nest by investing.

However, all investment comes with risk and various important decisions that have to be made. We have to be careful to ensure that our retirement savings are not overwhelmingly at risk while maximizing our expected returns.

In this project, our group has proposed a solution, named CPF Investment Advisor. CPF Investment Advisor is able to address the various concerns that one may have before investing – should I invest, or should I just leave the money in my CPF account? How much should I invest in each stock? We will also discuss limitations of the solution and suggest possible improvements for future work to handle investment products of increased complexity.

2.0 Problem Description

Globally, life expectancy has increased over the years from an average of 64.9 years in 1995 to 71.4 years in 2014 and is projected to increase to 75.9 years by 2040 (Roser, Ortiz-Ospina and Ritchie, 2019). Similarly, Singapore residents are expected to live longer compared to a decade ago. Life expectancy increased from 80.9 years in 2008 to 83.2 years in 2018 ("Life Expectancy at Birth", 2018).

To an individual, longer life expectancies entail both good and bad news. Each of us now have more time to spend with our loved ones and enjoy our golden retirement years. However, this would also mean that our retirement duration has increased, and sufficient savings have to be set aside to maintain our basic standard of living while keeping a small sum for rainy days.

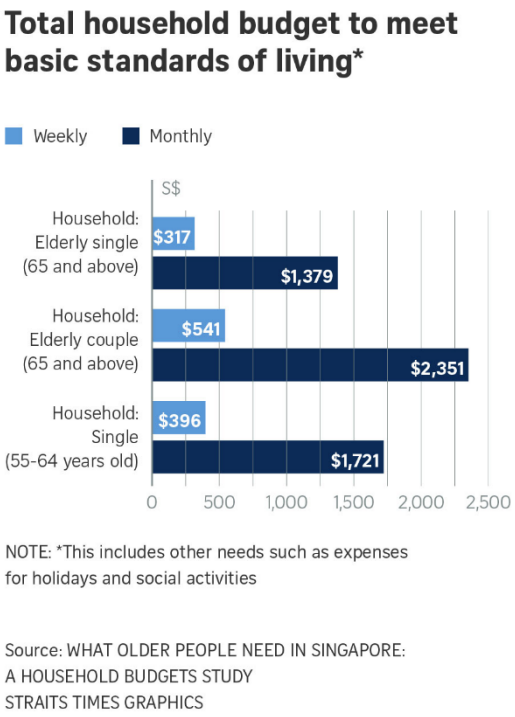

Figure 1: Total household budget to meet basic standards of living (Sin, 2019)

From a household budgets study reported by Straits Times (Figure 1), for an elderly single aged 65 years and above, approximately $1,400 is needed per month to meet their basic standards of living. With the minimum retirement age defined by the Ministry of Manpower ("Retirement", 2019) as 62 years, this would equate to having at least $30,000 in your retirement nest by the time of retirement. This amount does not include future medical expenditure nor unpaid mortgage loans, if any.

To help achieve a sufficient retirement nest, the Singapore government implemented a social security savings scheme, also known as Central Provident Fund (CPF). CPF aims to enable working Singapore Citizens and Permanent Residents to have a secure retirement, by setting aside funds for retirement, and also serves to meet our retirement, housing and healthcare needs.

Both employees and employers make monthly CPF contributions and the contributions are split into 3 accounts – Ordinary Account (OA), Special Account (SA) and Medical Account (MA). OA is typically used for housing, insurance, investment and education needs, with a lower interest rate of 3.5% per annum. SA, with a higher interest rate of 5% per annum, is for old age and investment in retirement-related financial products. MA is used for hospitalisation expenses and approved medical insurance, with an interest rate of 5% per annum.

Although the interest rates in each of the CPF accounts are higher than that of savings accounts (typically in the range of 0.5%) in Singapore, we can still do more to optimize our CPF accounts and enhance our retirement nest egg through the CPF Investment Scheme.

The CPF Investment Scheme provides an option of investing the OA and SA savings in a wide range of investments to enhance our retirement nest egg. Up to 35% of the amount in OA can be invested into stocks. SA, being the account targeted for old age, will be locked in till retirement and hence have very limited investment options. MA is typically used to cover medical expenses and should not be used for investment.

As with all investments, there will definitely be risk involved, where there is a possibility of losing part or all of the investment due to financial market changes. As such, several crucial decisions would have to be made – should I transfer part of OA to SA or should I invest in stocks? How much risk should I take with my retirement savings and can I afford this risk? Which stocks should I buy, how much should I invest in total and how much should I invest in each stock? What returns can I expect to get and how much would I potentially lose on this investment?

These questions are what we would like to answer through our proposed solution, the CPF Investor Advisor (CIA).

2.1 Project Scope and Objective

CIA is an expert advisory service that appraises how much risk you can afford to take based on your personal and financial profile, which then recommends stocks listed in Singapore

Exchange that match your preference and are within your risk appetite. CIA will then advise how much to transfer to CPF SA versus how much to invest in the stock market, while evaluating which stocks to buy and how much to invest in each specific stock. Finally, CIA will then predict the expected performance of the portfolio and quantify the potential for loss.

CIA is a hybrid reasoning model that consists of a rule based system to evaluate the user's risk profile and recommend stocks to select based on their preferred industries, a decision tree that then predicts the performance of the selected stocks before passing the high performance stocks into a genetic algorithm that then optimizes the amount to invest in each stock and the amount to be transferred to SA.

The entire hybrid reasoning model is packaged into React WebApp, making it simple and easy for everyday users to use.

3.0 Knowledge-Based System

Knowledge-Based Systems relies primarily on the interaction between an inference engine and knowledge base to generate a solution for new but similar scenarios. The knowledge base provides relevant information in a specific application domain, such as rules, facts and decision trees. The inference engine then implements a reasoning model and makes use of the knowledge base to explicit solutions to problems.

3.1 Knowledge Acquisition

As part of the knowledge base, we have identified various sources of information as shown in Table 1.

Charles Schwab Corporation is an American multinational financial services company and is ranked third in Best Overall Online Brokers in Investopedia's 2019 Best Online Brokers Awards. They are also customer-focused, receiving the highest numerical score in a tie in the J.D. Power 2019 US Wealth Management Mobile App Satisfaction Study of customers' satisfaction with their wealth management mobile app ("Industry Awards", 2020).

Table 1: Sources of knowledge

No.	Sources of information	Insights from information source
1	CPF website	Basic information such as minimum user requirements (age, minimum amounts in OA/SA) and interest rates of each account
2	Charles Schwab Questionnaire	Score system of time horizon and risk tolerance levels Decision table of overall risk profile Maximum risk appetite of a user
3	Singapore Exchange data	Database for decision tree to learn from
4	Expert with background in financial portfolio	Subject matter expert Provides background on function to optimize for in Genetic Algorithm

3.2 Knowledge Model

The basic information obtained from CPF website such as age requirement and minimum amounts in OA and SA was used to set up rules in CIA. Interest rates and maximum amount that can be transferred from OA were also obtained as inputs for the genetic algorithm.

The Charles Schwab Questionnaire ("Investor profile questionnaire", 2018) was used to assess the user's risk profile by asking questions that evaluates the user's time horizon and risk tolerance score. The time horizon and risk tolerance score were then combined to allocate a risk profile to the user, from a range of conservative to aggressive, based on a decision table shown in Figure 2. With the assigned risk profile, the user's risk tolerance can then be derived based on the maximum loss the user could expect in a given year, represented by the worst year's annual return rate shown in Figure 3. This risk tolerance is then passed to the GA Optimizer as a constraint towards identifying the best portfolio with the highest returns.

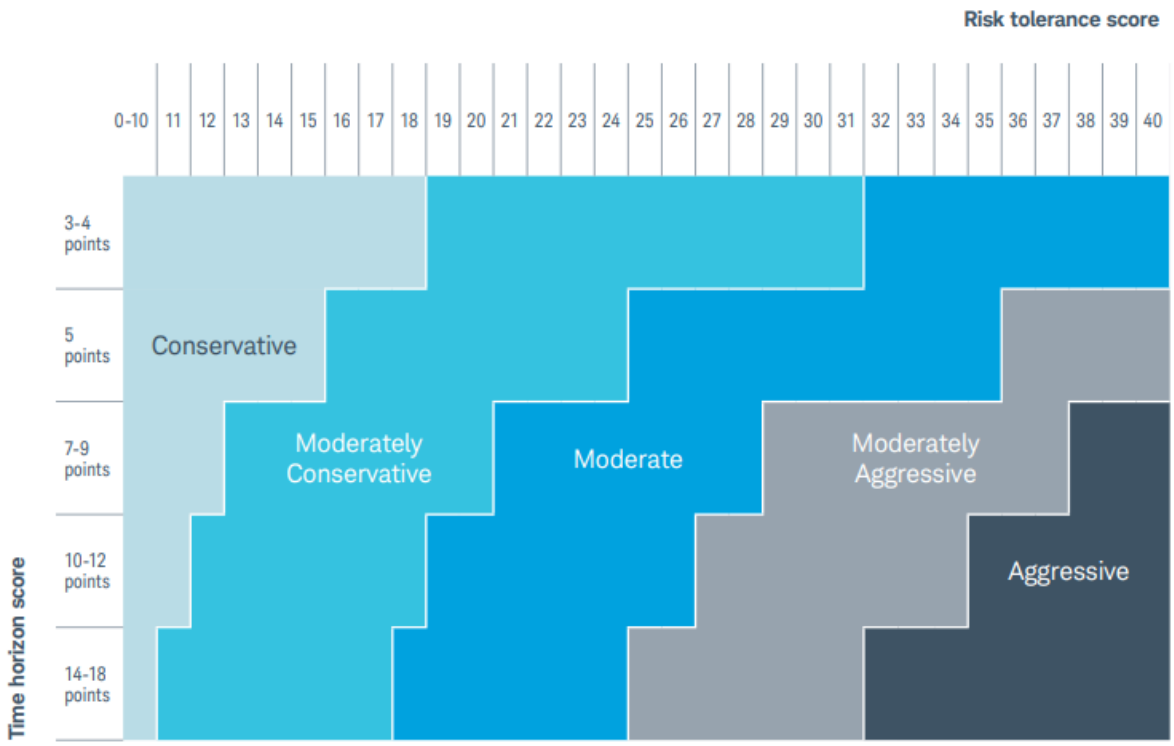

Figure 2: Allocation of risk profile based on time horizon and risk tolerance score

Conservative allocation	Moderately conservative	Moderate allocation	Moderately aggressive	Aggressive allocation
Average annual return: 7.6% Best year: 22.8% Worst year: -4.6%	Average annual return: 8.8% Best year: 27.0% Worst year: -12.5%	Average annual return: 9.5% Best year: 30.9% Worst year: -20.9%	Average annual return: 10.0% Best year: 34.4% Worst year: -29.5%	Average annual return: 10.3% Best year: 39.9% Worst year: -36.0%
For investors who seek current income and stability and are less concerned about growth.	For investors who seek current income and stability, with modest potential for increase in the value of their investments.	For long-term investors who don't need current income and want some growth potential. Likely to entail some fluctuations in value, but presents less volatility than the overall equity market.	For long-term investors who want good growth potential and don't need current income. Entails a fair amount of volatility, but not as much as a portfolio invested exclusively in equities.	For long-term investors who want high growth potential and don't need current income. May entail substantial year-to-year volatility in value in exchange for potentially high long-term returns.

Figure 3: Risk appetite for user based on worst year's annual return rate

For the next module that helps the user select suitable stocks to invest, we used the 2019 historical financial data of all stocks listed on Singapore Exchange (SGX). The data was pre-processed and labelled as 'high performance' (or 'low performance') if the historic yield was greater than 4%. The classification threshold of 4% was decided based on the fact that CPF-SA provides a guaranteed return of 4% effective serving as the risk-free rate, and a floor for any investment returns. There is limited value in investing in a stock that yields less than the CPF-SA account.

We then used this labelled data to train several classification models available in sklearn. The accuracy results of predicted labels based on test data against actual labels as shown in Figure 4 below.

```
Accuracy - DT: 0.5238095238095238
Accuracy - KNN: 0.5555555555555556
Accuracy - RF: 0.47619047619047616
Accuracy - AB: 0.5079365079365079
Accuracy - NB: 0.5555555555555556
Accuracy - XGB: 0.5714285714285714
```

Figure 4: Accuracy results of various training models

Extreme gradient boosting (XGB) was eventually selected as it gave the highest accuracy score and the trained decision tree is shown in Figure 5. The XGB package also allows us to evaluate the importance of a feature.

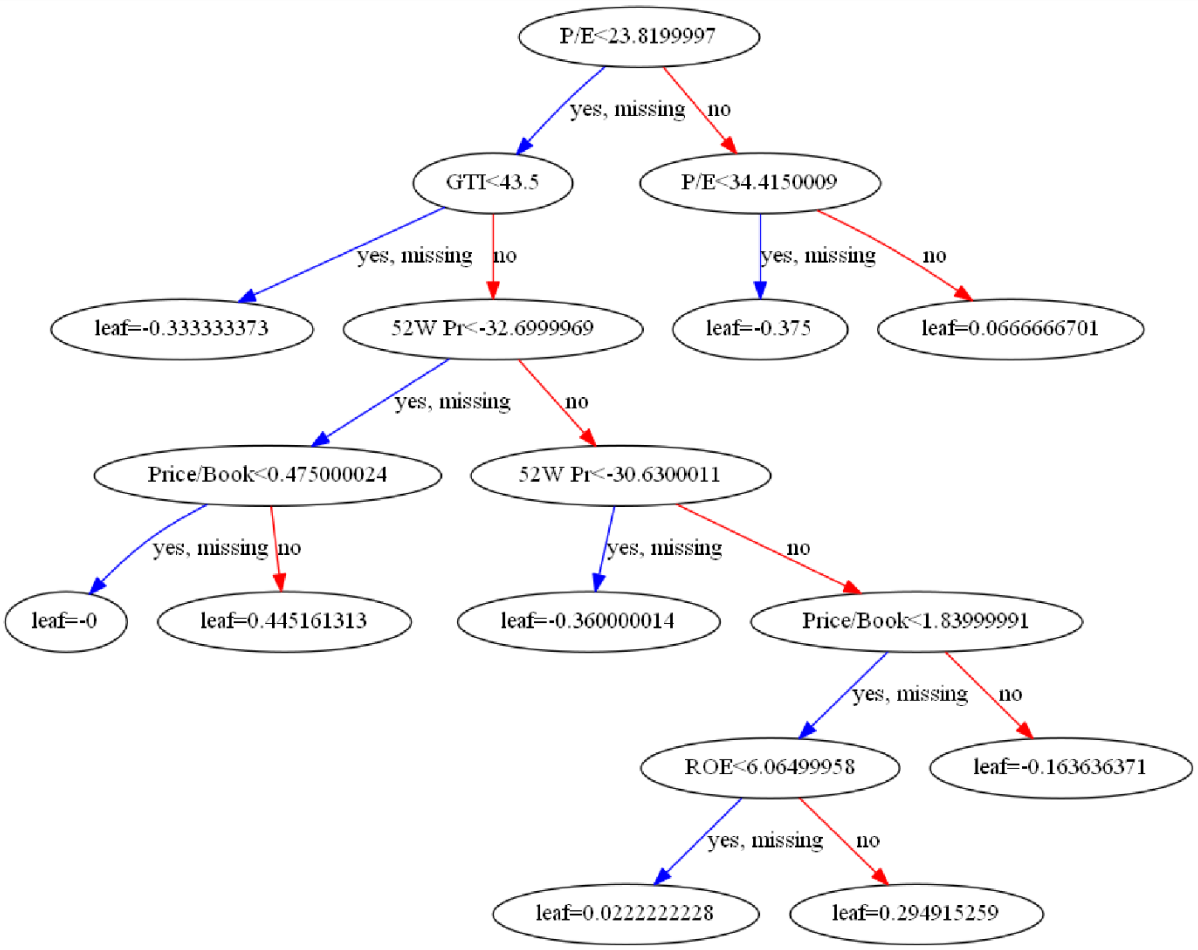

Figure 5: Decision Tree as generated by XGB

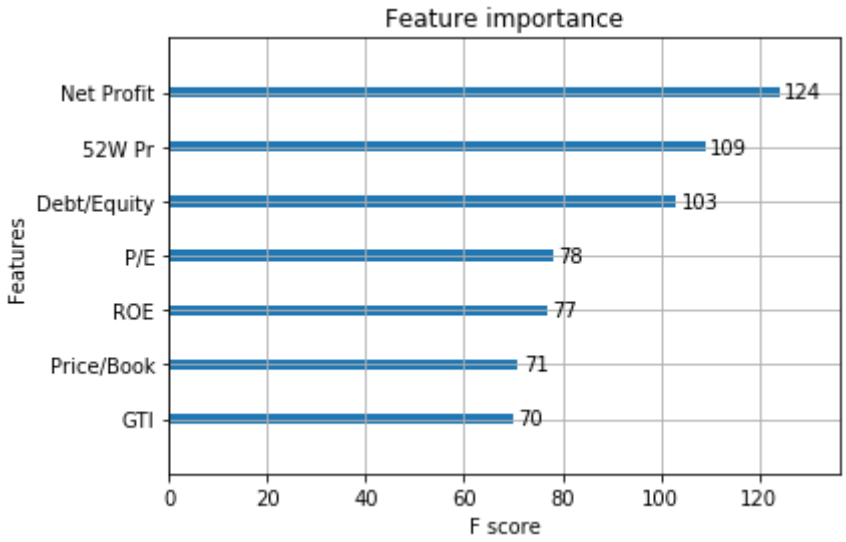

Figure 6: Importance of individual training features

Finally, we use a Genetic Algorithm (GA) to search for the optimal portfolio i.e. one which provides the highest returns while keeping the variance or risk within the user's risk profile. The GA models the Markowitz Mean-Variance model to determine the most efficient portfolio by analysing various possible portfolios of the given stocks.

An efficient portfolio is one that provides maximum return for a given risk or volatility. Put differently, for portfolios that have the same variance or risk, the investor will prefer the portfolio which provides the highest returns.

By plotting the return-risk of all possible combinations of portfolios, we get the plot of the efficient frontier. Efficient portfolios are ones that lie on the upper edge of the efficient frontier. From Figure 7, the red star indicates the portfolio with the lowest risk/variance, while the blue star identifies the portfolio with the Sharpe Ratio, which is the highest ratio of returns over risk. The green star is the portfolio recommended to the user, with the highest return while keeping the risk variance below the user's risk tolerance - in this example under 10%.

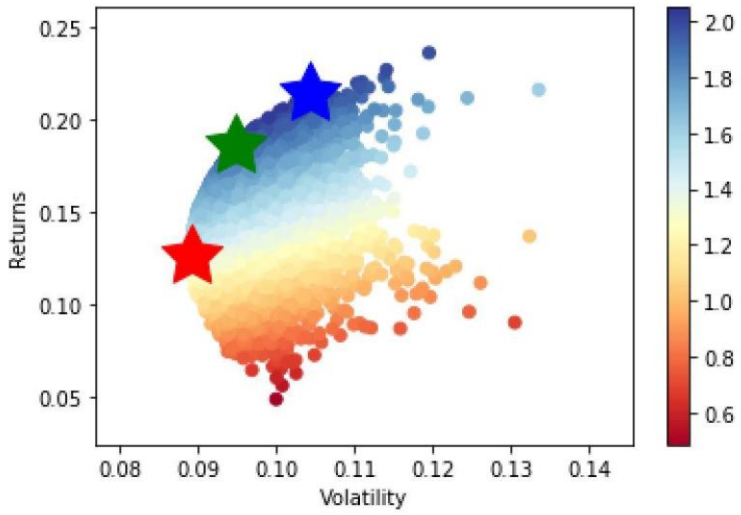

Figure 7: Markowitz Bullet

In our system, we aim to recommend portfolios that lie on the efficient frontier by setting a risk value and optimising for maximum returns at that particular risk value. The maximum risk value that the user can accept is as derived in the Risk Profile Assessment from the first module.

The maximum risk value is thus set as the hard constraint within the genetic algorithm (the other constraint being the sum of weights should total to 1) while optimizing for the weight assignment for the individual stocks selected by the user to achieve a maximum rate of return for the user's respective risk profile.

4.0 Solution

4.1 System Architecture

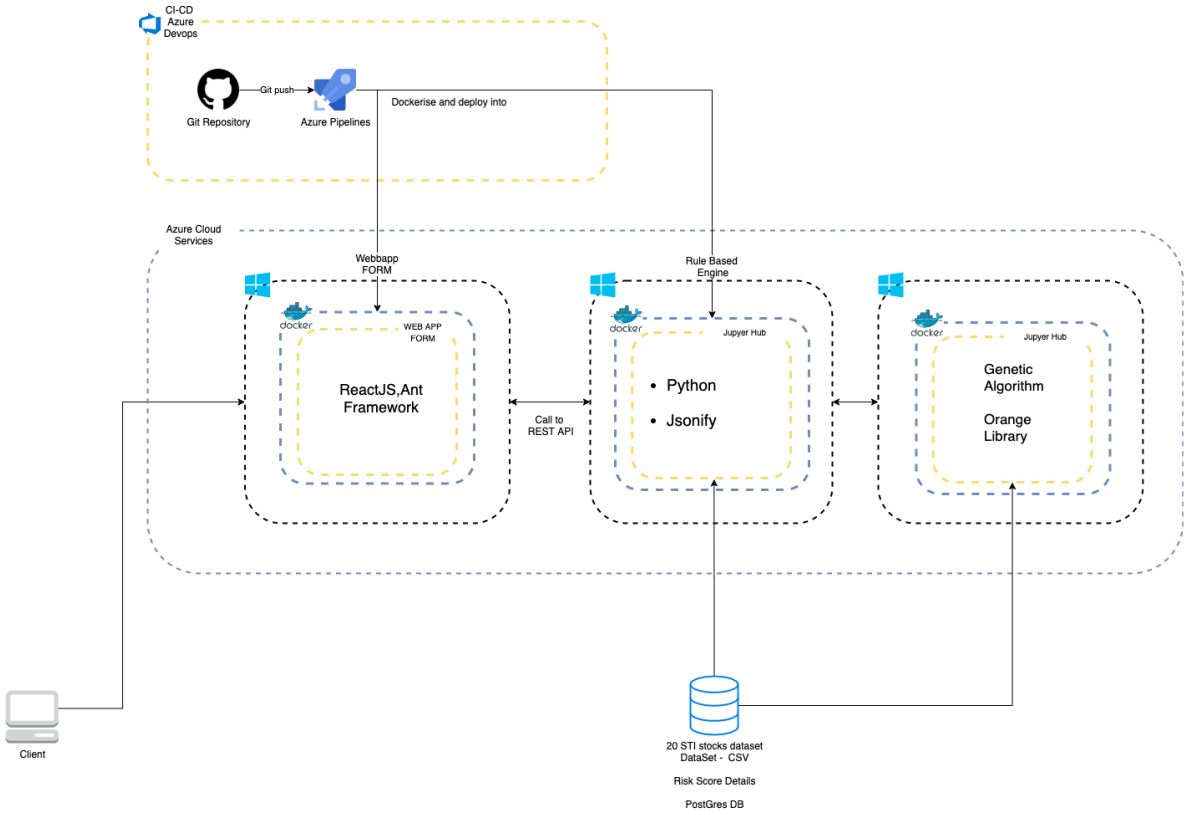

Figure 8: CPF Investment Advisor – System Architecture

The initial plan was to utilise cloud computing as servers instead of deploying onto a docker container, but it has since changed as the CICD is non-existent other than the connection to the Azure container registry. An Azure-pipeline YAML exists in the codebase. However, the framework of the core packages stays the same.

The frontend uses ReactJS and AntD (https://ant.design/) for its component library, while in the backend, the codes are served in a monolithic service, fronted by the flask middleware. The codes are in python and flask will take in the specified parameters from the frontend and consume it accordingly.

4.2 System Flow

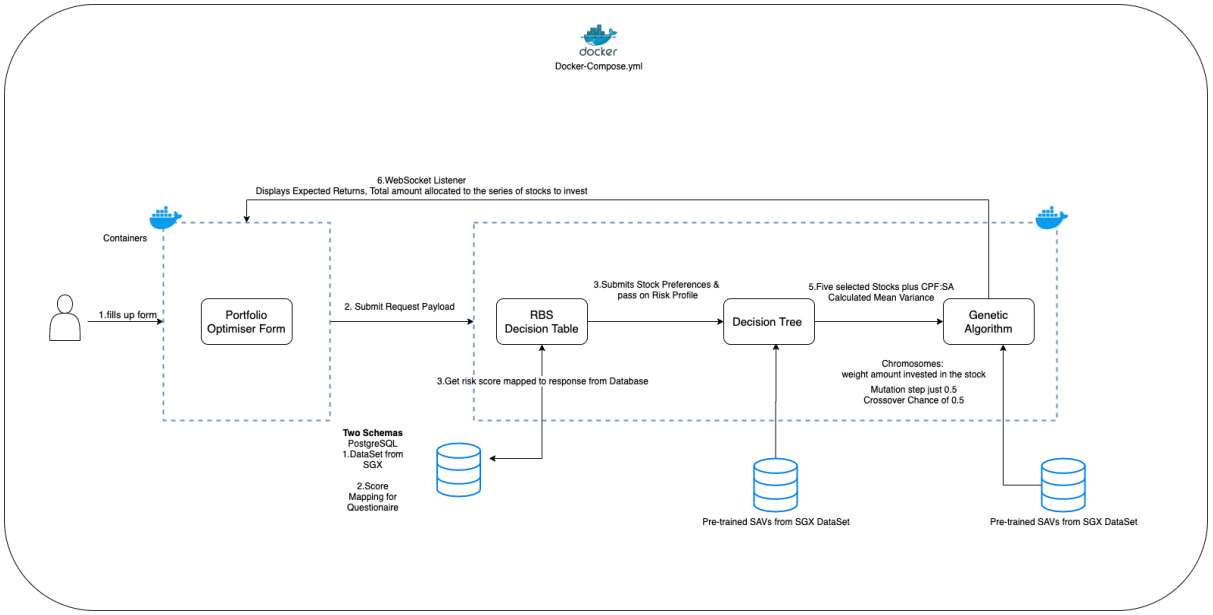

Figure 9: Detailed Inference Structure Diagram for CPF Investment Advisor

From Figure 9, 3 containers are used – React WebApp that hosts the form for user inputs, the Python/Flask backend where the various rules, decision tree and genetic algorithm engines are stored and lastly, the PostgreSQL that returns the output to the form for users to view.

The user inputs for the questionnaire will be sent from the frontend application to the rule-based engine that resides in the monolith, where the inputs are mapped to a specific score in the database. To retrieve the scores, an array/list of answers are submitted to the flask middleware. Subsequently, it will call on the handler which will then formulate the SQL query accordingly and send the results back to a rule-based engine utility function.

The payload to be consumed by the rule-based engine has two parameters, the time horizon score and the risk score. These two are then mapped to a specific profile type (moderate, aggressive etc.) in a decision table. Finally, the risk threshold associated with the persona type is then stored as a parameter to be consumed later on.

For the Database we are using PostgreSQL, it will be packaged together with the docker compose together and you will have to have docker installed in your computer to be able to deploy them automatically.

4.3 System Scope

CIA only covers companies listed in SGX although there are other investment products included under the CPF Investment Scheme such as Unit Trusts, Singapore Government Bonds

etc. The team believes that the current dataset of around 200 SGX stocks is sufficient for the decision tree to be trained.

4.4 Assumptions

Some of the key underlying assumptions are:

1. User has a basic familiarity with investing e.g. understands concepts like sectors, stocks, returns, risk etc
2. SGX stocks financial & performance data is updated periodically to keep it current & relevant for prediction (Disclaimer: Past performance does not 'guarantee' Future performance)
3. Assumes the user does not have existing CPF Investments i.e. entire CPF balance is in Cash.
4. Assumes the user does not have any requirement to set aside cash for Housing or Education i.e. only sets aside the min mandatory $20K
5. Assumes that information retrieved from CPF website is accurate and up to date

4.5 System Features

Knowledge & Intelligence

- Knowledge Representation: The Risk Profile Questionnaire has been represented as a Rules Based Representation to assess Investor Risk Profile.
- Supervised Learning: Uses labelled past financial data to learn the relevant financial performance attributes (features) and their correlations to predict future Yield Performance.
- Search Space: Models the space using a simple chromosome representation of the individual investment weights, and then leverages a Genetic Algorithm to search for an optimal solution

Technical Features

- Ease of access – available in both online and offline versions
- Scalability – possible to input bigger sets of data such as the complete list of investment products within CPF Investment Scheme

4.6 Limitations

Trained data is based on historical data and may not be able to address unexpected situations such as economic downturn brought about by pandemic (i.e. CoVID-19) and cyclical recessions.

5.0 Conclusions & References

5.1 Future Roadmap

1. Commercialize the Service: Implement the functionality behind the 'Invest' button - where the application will place the order for the relevant stocks with the broker. The application will get a percentage of the brokerage for this, and this will be the revenue model to commercialize this application i.e. offer the Investment Advisory as a complementary service and monetize the brokerage commission earned.
2. Wealth of Data to offer Wealth Services (pun intended): In return for offering the complementary Investment Advisory, we collect a rich set of data on the Investors personal, financial & risk profile which can be used to target future investment products suitable for the Investor (e.g. Retirement Annuities, Insurance Products etc). With a growing client base, this can be a significant distribution channel for new financial products and the commissions earned on selling these new products would be a significant revenue stream for the product.
3. Expand to other Products: while the current implementation is focused on SGX stocks, there are several other products that can be invested via the CPF-OA Investment account. Including these new products will expand the product offering and make the service more useful (& additional revenues)

6.0 Bibliography

Roser, M., Ortiz-Ospina, E., & Ritchie, H. (2019). Life Expectancy. Retrieved from https://ourworldindata.org/life-expectancy

Life Expectancy at Birth. (2018). Retrieved from https://www.singstat.gov.sg/-/media/files/visualising_data/infographics/population/life-expectancy.pdf

Sin, Y. (2019) Single elderly Singaporeans need $1,379 a month to meet basic living standard: Study. *The Straits Times.* Retrieved from https://www.straitstimes.com/singapore/study-finds-1379-a-month-needed-to-meet-basic-living-standard-for-single-elderly

Retirement (2019). Retrieved from https://www.mom.gov.sg/employment-practices/retirement

Industry Awards (2020). Retrieved from https://www.schwab.com/awards

Investor profile questionnaire (2018). Retrieved from https://www.schwab.com/public/file/P-778947/InvestorProfileQuestionnaire.pdf

7.0 Appendices

1. Mapped System Functionalities

No.	System Function	Modular Course	Knowledge/Technique/Skills used
1	Rules	Machine Reasoning	Data Mining, Knowledge Acquisition, Knowledge Modelling
2	Decision Tree	Machine Reasoning	Data Mining, Knowledge Discovery, Knowledge Acquisition, Knowledge Modelling
3	Portfolio Optimizer	Reasoning System	Knowledge Modelling, Genetic Algorithm
4	Overall System	Reasoning System	Machine Inference, Hybrid Reasoning System

2. Project Proposal (uploaded into GitHub)

3. Installation and User Guide (uploaded into GitHub)

4. Individual project report (uploaded into GitHub)