

CDS 590:

**Consultancy Project and Practicum**

Consultancy Project Proposal:

**Real Private Consumption**

**Time Series Prediction and Forecasting**

Prepared by:

**Muhammad Azzubair bin Azeman P-COM0019/19**

Submitted for:

**Dr Azam Osman**

Date of Submission:

**1 November 2020**

School of Computer Science

**Academic Session 2019 / 2020**

**CHAPTER 1**

**INTRODUCTION & RELATED WORKS**

**1.1 Background of Practicum Company**

DataMicron Systems Sdn Bhd is a technology company which offers consultant services for business intelligence and big data-related solutions. As many companies nowadays have their own databases, they face challenges and difficulties in gaining insights from their big and complex data using traditional techniques. Therefore, this is where DataMicron comes in whereby their managing director said in his interview with *The Star* newspaper publication:

“*We extract data from various databases provided to us by our clients and merge them in the data warehouse, where from there, we do analysis of the data to provide our clients with business intelligence and predictive analysis, which in turn, would help them in their decision-making process*” (Hooi, 2014).

On 2004, DataMicron company was granted by Government of Malaysia through Malaysia Digital Economy Corporation (MDEC) with Malaysia Status Services (MSC) status which enables their company to enhance their product and service developments on multimedia technologies. As a result, the company has extended their scope of services to more than five countries as in 2014. The success of this company was reflected by their Microsoft Asia Pacific Keystone Award on 2005, and SME Corp Innovation Award (ICT) on 2013 (Hooi, 2014).

DataMicron provides innovative solutions for Big Data, and Business Intelligence for many local and international organisations. As data value is significantly increasing, DataMicron offers three types of data-related services which are Training, Consultancy, and Support. In terms of training, DataMicron together with other industry partners agreed to develop future talents by conducting one-year placement under their company for Bachelor students of Universiti Teknologi Malaysia under 2u2i mode programme (MDEC, 2019).

**1.2 Background of Domain**

Level of economic advancement of one country to another differs by the main macroeconomics indicator which is the Gross Domestic Product (GDP). As such, world countries are categorised into three main categories which are Developed Economies, Economies in Transition, and Developing Economies, with additional Least Developed Countries listed (United Nations, 2020). In determining country classification, World Gross Product (WGP) (derived form of GDP) is included as one of the indicators.

Gross Domestic Product (GDP)is definedas “the market value of all final goods and services produced in an economy annually” (Hashim et. al, 2018). It is measured based on three main approaches which are the Production, Expenditure, and Income approaches (Department of Statistics Malaysia (DoSM), 2020). In terms of Production approach, it reflects on economic activities of individuals towards GDP as an overall; while for Expenditure approach, it determines the values of services and products consumed by consumers. As for Income approach, it includes all income sources and amounts gained in economy. Therefore, in order to determine the economic values of each approach, macroeconomic indicators (econometrics) are used as input to calculate the values of each approach.

Expenditure approach plays a crucial role in overall GDP as it contribute the most to the overall GDP since 2013 until 2018 (Asada et.al, 2019). This approach is dependent on five main macroeconomic indicators (econometrics) namely as Real Private Consumption (RPC), Real Government Consumption (RGC), Fixed Capital Formation (FCF), Changes in Inventories and Valuables (CIV), and Net Export (NE) (DoSM, 2020). Therefore, forecasting of econometrics are crucial to be applied as reference for Ministry of Finance (MoF) Malaysia in making decisions for future financial planning. Currently, statistical techniques such as vector autoregression (VAR) and Autoregressive Integrated Moving Average (ARIMA) model are commonly used for econometrics forecasting in Malaysia (Razak, Khamis & Abdullah, 2017).

Econometrics forecasting using machine learning has been a major topic discussed in many literatures in the last two decades (Taieb, 2014). Several machine learning models such as Neural Network, Support Vector Machine, and K-Nearest Neighbour were proposed and discussed. However, machine learning are foreign among Malaysians until it was recommended by Minister of International Trade and Industry (MITI) (Bernama, 2018). As a result, machine mearning models and techniques are gradually being learned by Malaysians in many online courses recently (Fadzil, Latif, & Munira, 2015).

**1.2 Problem Statement**

Shifting from statistical models into machine learning models for econometrics forecasting definitely requires some time to be adopted. Plus, machine learning models have to compete with current statistical models used in government sectors. In particular, its ability to forecast future values is still questionable whether it could outperform current statistical models or not. Model evaluations such as model prediction accuracy, time taken for modelling and prediction, and model’s complexity and transparency must be taken into account in weighing their reliability for econometrics forecasting. If machine learning is proved to be more reliable, then government sectors should reconsider in adapting to the new technology emerged nowadays. Otherwise, statistical models are remained.

**1.3 Research Question**

This research proposal makes an attempt to model and forecast time series of one of the econometrics which is the Real Private Consumption. This will consequently results in determining whether machine learning models or statistical models is better for econometrics forecasting. This project proposes machine learning models used for econometrics forecasting between Support Vector Machine (SVM) and Random Forest (RF) models with ARIMA and VAR models. This brings to the following research questions:

* Between RF and SVM, which model is better for RPC forecasting?
* What would be the RPC forecast of machine learning models for the third quarter of 2020?

Throughout this proposal, the research question answers will be reviewed from literatures to offer valid justifications, and to propose the best machine learning model for RPC forecasting.

**1.4 Objectives**

Therefore, the aim of this project is to propose machine learning techniques as a new approach to improve econometric forecasting in Malaysia. This project will be focusing on forecasting one of the econometric indicators which is the RPC published quarterly by DoSM. Typically, to achieve the aim of this project, there are two objectives listed below which are:

1. To determine the best machine learning model for RPC forecasting
2. To forecast Real Private Consumption for third quarter of 2020

**1.5 Benefits of the project**

This problem is actually a consultation project between DataMicron with Malaysia’s Ministry of Finance (MoF). Therefore, this project will benefit DataMicron in providing proposed solution for their client’s problem. In particular, this project will deliver the insights of previous RPC trends and will develop a reliable model for suggesting RPC forecasts as a reference for MoF’s top management in making effective decisions. In return, MoF could use these insights to optimise their financial planning and reduce financial loss.

**1.6 Related Works**

In this section, econometric literatures will be referred to explore the details of established and proposed methods in econometrics modelling typically for RPC. Published researches demonstrating on the theories and analysis behind econometrics modelling will be reviewed and proposed models potentially applicable in the future will be discussed.

**1.6.1 Review on Domain**

Since 2015, RPC contributed the most to Malaysia’s GDP followed by FCF and NE. However, RPC has shown a fluctuating trend between 6.0% – 8.0% in annual variation (Central Bank of Malaysia (BNM), 2019; BNM, 2018; BNM, 2017). This trend is contributed mostly by the growth of employments and wages, thus this shows that affecting this sector resulted much to the overall annual RPC. Other than that, RPC is also being contributed by other variables such as imports of consumption goods, narrow money, and loans disbursed by banks (BNM, 2016).

**1.6.2 Review on Data Science & Analytics techniques**

Statistical techniques such as ARIMA, and VAR were reported and compared for Malaysia’s econometrics forecasting. (Razak, Khamis & Abdullah, 2017). Both of these models can be used for univariate time series (UTS) forecasting. This means both models can be used to forecast a variable for several periods ahead in future. The findings of this study found out that VAR is more accurate than ARIMA due VAR’s less mean absolute percentage error (MAPE) compared to ARIMA (Razak, Khamis & Abdullah, 2017). In addition, they also highlighted that VAR outweighed ARIMA by having multivariate time series (MTS) forecasting which enables for more dynamic forecasting using multiple variables to forecast a certain value.

In other parts of the world, machine learning techniques were began to be proposed for econometrics forecasting since 20th century. The earliest attempt was done by Yu (1999) whereby she compared model performance between ARIMA and Artificial Neural Network (ANN) in forecasting stock index. The outcome of this study found out that ANN produced lower MAPE and Root Mean Squared Error (RMSE) compared to ARIMA. This reflects that nonlinear trend of stock index is better to be forecasted using machine learning models than linear models such as ARIMA. Similar observation was obtained by Dematos et. Al (1996) in which they found that Recurrent Neural Network (RNN) outperformed ARIMA in forecasting Japanese yen / U.S. dollar exchange rate. This summarised that nonlinear trend of economic indicators are better to be forecasted using machine learning instead of statistical methods.

Recently, a comparative study was done by Kumar et. al (2018) in comparing between machine learning model performances in predicting stock market trend. Machine learning used were support vector machine (SVM), random forest (RF), k-nearest neighbour, naive bayes, and softmax. Interestingly, they found out that when large dataset (4500 entries) were input, random forest outperformed other models by having the highest accuracy and f-measure followed by support vector machine. This indicates that RF and SVM are suitable models to be used for predicting nonlinear trends of economic growths. In addition, RF is also good to used for prediction due to its robustness against outliers because of its bagging principle in learning the training set and predicting the test set (Roy & Larocque, 2015). Meanwhile for SVM, it is known to be highly effective and efficient in forecasting values of stock prices (Vo et al., 2016).

**1.6.3 Review on Data Science & Analytics tools**

Analytical tools for time series prediction and forecasting are abundant nowadays. They are available online either as a free software or as an advance premium software. An example of a free software is the Python, a well-integrated, and popular data science tool among data scientists. Multiple studies have been using Python for time series prediction and forecasting because there are many available libraries that are created for the purpose of dealing with time series data and forecasts. One of the popular libraries for time series forecasting is the statsmodels library. According to McKinney et al. (2011), this library provide many statistical models made available to python users such as ordinary least squares, VAR, ARIMA and many more. Another library for time series forecasting is the Facebook’s Prophet as proposed by Usher and Dondio (2020) for short term forecast on the pound sterling with respect to euro and dollar currency. They forecasted that the pound sterling would rise against dollar and euro by ± 0.02.

Another approach of time series forecasting is by using machine learning approach. This approach recognises time series data as a supervised learning through the use of sliding window for model training and testing. Brownlee (2020) explained about the sliding window in detail whereby he described that time series dataset can be restructured into a supervised learning dataset by using the value of previous data to predict for future data. In short, historical data are taken as input and future data is treated as the output. In python, this windowing feature is available is several libraries such tslearn, cesium-ml, ts-fresh, and seglearn. Burns and Whyne (2018) compared these libraries features for time series forecasting. Their comparison shows that all of the libraries has forecasting feature except tslearn. On top of that, they found out that seglearn library is the only library that has the most features such as multivariate time series, sliding window, and compatible with machine learning models.

Another important library for tasks using machine learning models is the scikit learn library. This library contains most of the common machine learning models for data scientist. According to Hackeling (2017), scikit learn library provided linear models such as linear and multiple linear regressions, non linear models such as decision trees and random forests, and perceptron derived models such as support vector machines and artificial neural networks. Hence, combination of scikit learn and seglearn libraries are sufficient for time series forecasting.

**CHAPTER 2**

**RESEARCH METHODOLOGY**

**2.1 Activities plan and Gantt Chart**

Throughout semester 1 2020 / 2021, the activities for project consultancy and practicum will be conducted based on the plans as illustrated in figure 1 below. This project will be done individually with the supervision of project supervisor and guided by a mentor from the practicum company. With the limitations of the current COVID-19 situation, all of the process for the project consultancy and practicum will be conducted online via emails, phone calls, and conference calls. 96 contact hours with the practicum company will be recorded in a logbook and weekly meetings with supervisor will also be recorded to update about project progress.

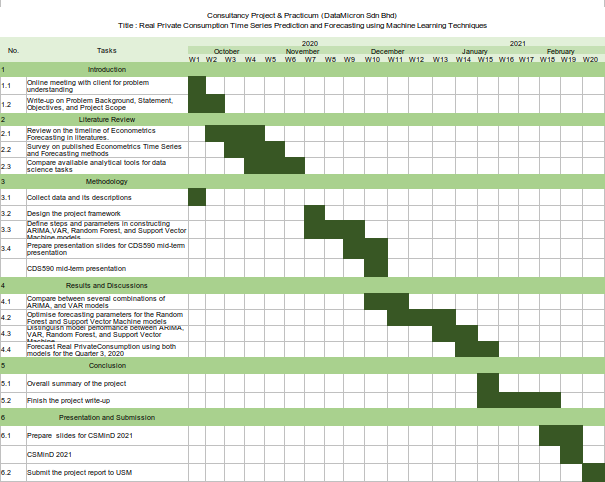
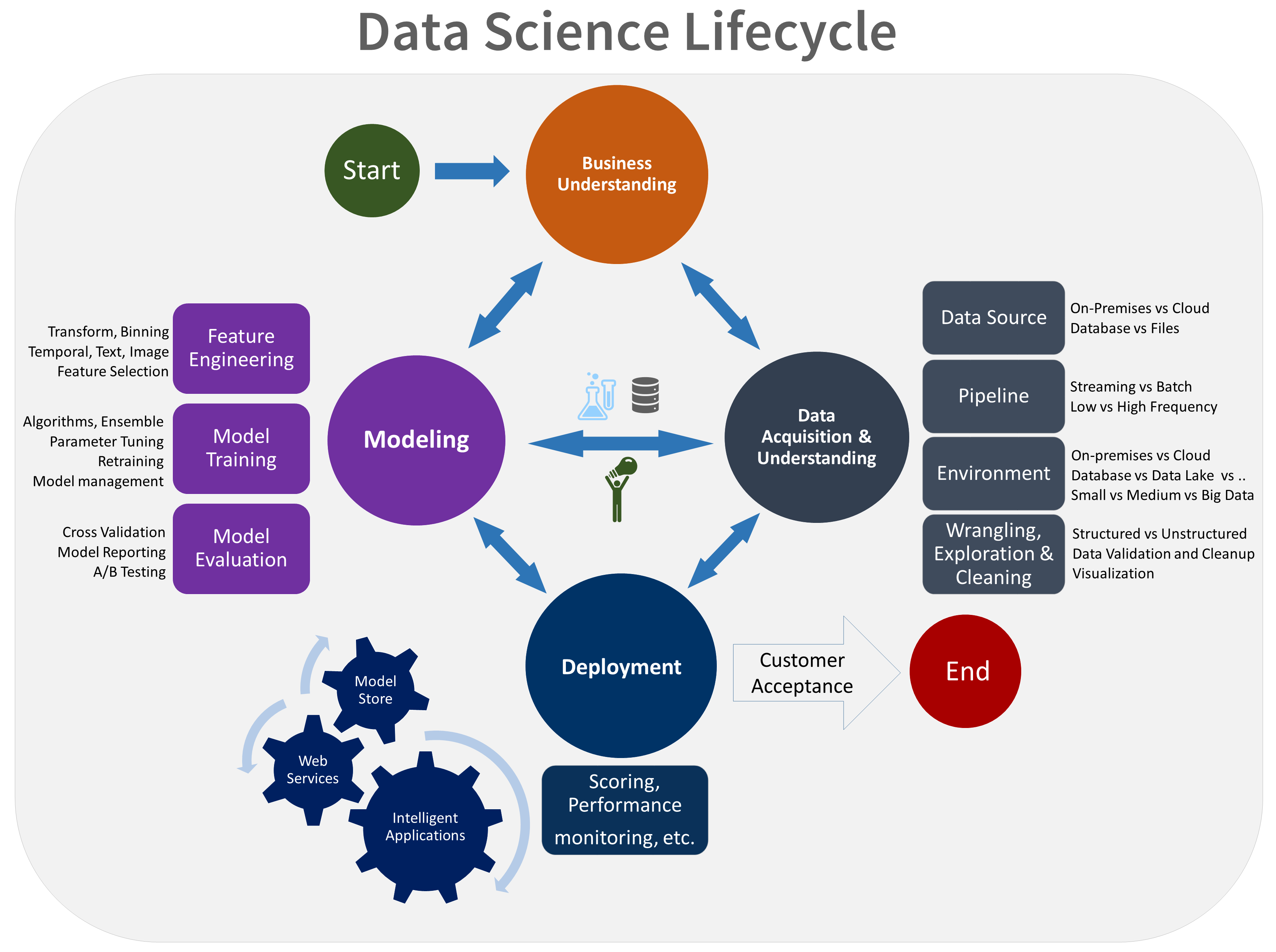


Figure 1 Gantt Chart of Project Consultancy and Practicum

**2.2 Data Science Project Lifecycle**

As a data scientist consultant, it is important to impliment the fundamentals of data science project lifecycle in daily life. The reason is to structure the process of data science projects so that these projects provide beneficial insights for clients effectively and the outcomes of these project are deliverable on time.According toMicrosoft (2020),there are five main stages of data science project lifecycle which are Business Understanding, Data Acquisition and Understanding, Modelling, Deployment, and Customer Acceptance. Figure 2 illustrates the Data Science Lifecycle stages. Throughout practicum, all of these stages will be went through.

Figure 2 Lifecycle of Data Science Projects

In short, Business Understanding stage will be conducted in a consultation meeting whereby MoF will explain the background of their RPC problem, analyse the problem together in the form of research questions, and describe their expected solutions. In order to determine the success of the proposed solution, it will be measured, and ensured to be within clients’ expectations using success metrics that are specific, measurable, achievable, relevant, and time-bound (Microsoft, 2020). Next stages will be conducted individually and lastly the proposed solution will be presented to MoF during the Customer Acceptance stage.

**2.3 Problem Analysis**

Analysis of client’s problem will be conducted during Business Understanding stage. At this stage, a list of formulated questions will be prepared and asked to the client. The purpose of these questions are mainly to determine the problem framing, identify target and predictor variables, and pinpoint for data source. Before coming into the main questions, general questions related to domain background will be prepared. The listed questions are as follow. After the questions are answered, exploratory data analysis and final analysis will be done.

**2.3.1** **Initial Questions**

Questions listed below will be prepared to the client for getting to know of the domain background, methods currently being used for RPC forercasting, and solution expectations. The answers of these questions will be used as a reference throughout this consultancy project.

* What is Real Private Consumption?
* What are the variables considered in forecasting RPC?
* How are the variables collected before going into data analytics?
* Why is it important to forecast Real Private Consumption?
* Among all of the mentioned variables, which variable would be the target variable?
* What are the current methods of forecasting RPC?
* How good are those methods in modelling and forecasting RPC?
* What are the limitations of current methods in forecasting RPC?
* What type machine learning models would be expected for RPC forecasting?
* What is the expected model performance metrics to be implimented?
* What is the threshold of acceptance for the RPC forecast values?

**2.3.2 Specific use case to be addressed**

According to the World Bank Group (2020), Malaysia’s annual private consumption is projected to be declining from 1.2% in 2019 into - 4.9% in 2020 due to the recent COVID-19 pandemic. Although Malaysia government had already provide financial support to its citizens through *Prihatin Rakyat* and *Penjana* packages, real private consumption will still be affected due to social restrictions which reduced the household demands of purchasing wants carefreely.

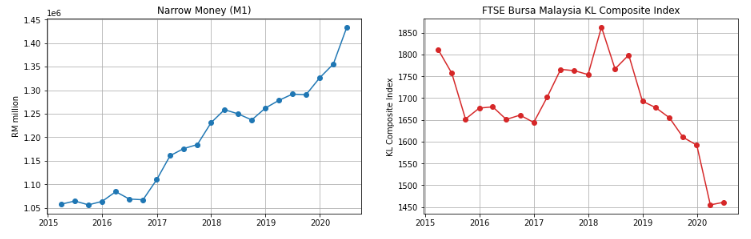
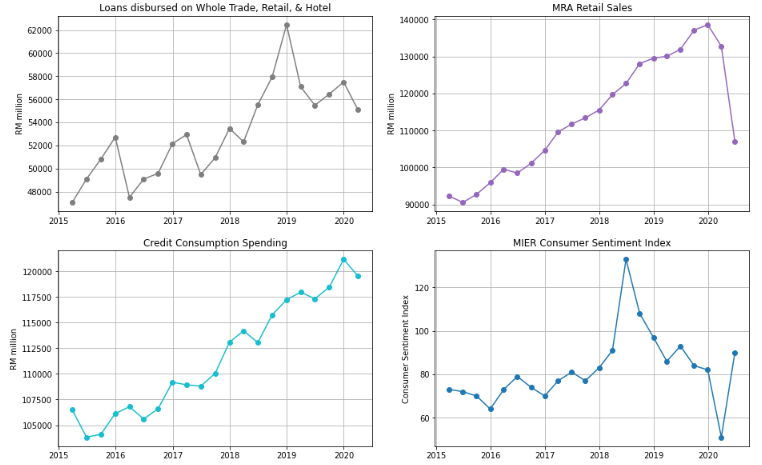
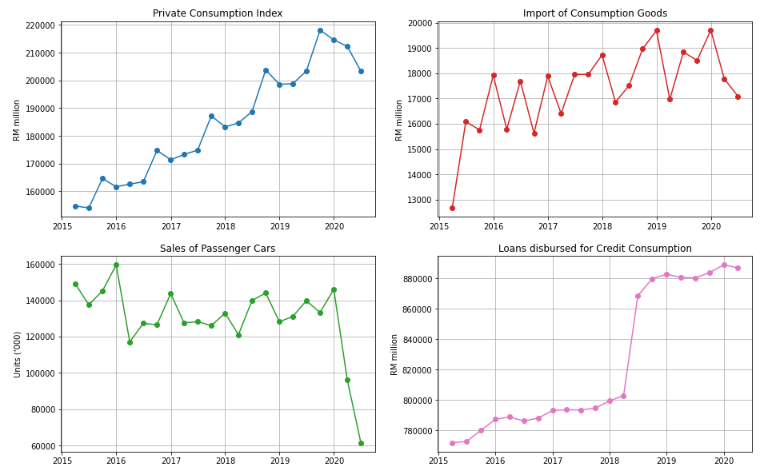
**2.3.3 Exploratory Data Analysis**

Upon receiving dataset from MoF, an exploratory data analysis will be done to overview the variable distributions, insights, and trends. In terms of Data Science Lifecycle, this step is categorised under Data Acquisition and Understanding. Firstly, the dataset will be given in an excel file containing variables of real private consumption indicators as published in BNM (2016). These variables is collected from data published by DoSM, and BNM from 2015 until 2020. All of the numerical variables will be cleaned, and transformed into quarterly data for them to be aligned with quarterly RPC. Table 1 shows the variables types and descriptions in detail.

Table 1 Description of Real Private Consumption and its indicators

|  |  |  |
| --- | --- | --- |
| No. | Variable | Descriptions |
| 1. | Private Consumption Index (PCI) | Measures consumer spending on goods and services in RM billions |
| 2. | Imports of Consumption Goods (ICG) | Import of any tangible commodity produced and purchased by consumers in RM billion amount |
| 3. | Sales of Passenger Cars (SOPC) | Amount of sold cars manufactured by local Malaysian brands in ‘000 units |
| 4. | Loans disbursed for Consumption Credit (LCC) | Amount of RM billions lended by banks for loans in Consumption Credits |
| 5. | Loans disbursed to Wholesale & Retail Trade, Restaurant, & Hotels (LOWT) | Amount of RM billions lended by banks for loans in consumers’ Consumption Credits |
| 6. | MRA retail sales (MRS) | Amount of sales collected by registered retailers under MRA in RM billions |
| 7. | Credit Card Turnover Spending (CCS) | Total amount of credits spent in RM billions amount |
| 8. | MIER Consumer Sentiment Index (MIER) | Measure consumer confidence on Malaysia’s economy status |
| 9. | Narrow Money (NM) | Aggregate amount of monetary assets available in Malaysia in RM billions |
| 10. | FBM KLCI | Capitalised-weighted stock market index comprised of 30 largest companies on Bursa Malaysia |

After describing each attributes, python will be used to be visualise the trends of each attributes. Figure 3 below shows the overall visualisation of RPC and its indicators.

Figure 3 Overview of Real Private Consumption trends and its indicators

Referring to Figure 3, prior to the COVID-19 pandemic, PCI shows an increasing trend with seasonal patterns from 2015 until 2019, while ICG, LOWT, LCC, MRS, CCS, and NM also show increasing trends. This shows that most of RPC indicators shows an improvement of RPC growth throghout the years. In addition, MIER also shows an increasing trend but only until second quarter of 2018, beyond than that, MIER declined. Because of Malaysia’s General Election held during the second quarter of 2018, consumer sentiments went skyrocketed until the end of the second quarter of 2018. Beyond that quarter, pessimists began outnumbered optimists due to the global challenges affecting Malaysia’s economic growth (Rasid, 2019). In contrast, FBM displays an overall decreasing trend with some exceptions in 2018. This attribute is heavily affected by Malaysia’s political issues in which investors did not want to take risk in investment while Malaysia is having political turbulence (Afandi & Khoo, 2020).

Upon the COVID-19 pandemic emergence in Malaysia, all RPC indicators show a significance decrease in the first quarter of 2020 in comparison with fourth quarter of 2019 excluding the narrow money attribute. This reflects that external virus has inflicted severely on Malaysia’s RPC especially when Malaysia enforced the Movement Control Order starting on March 2020. In second quarter of 2020, some of the RPC indicators such as the MRS, NM, and FBM show a rebound trend whereby the values are slightly improving during Malaysia’s Recovery Movement Control Order (RMCO). However, still most of the remaining variables such as ICG, SOPC, and MRS are having declining trends which resulted in the overall downward trend of RPC (PCI) in the second quarter of 2020. In contrast, NM shows an increasing trend throughout the years which means money supply for Malaysia is not affected by the pandemic. This indicates that more money are being supplied in the economy over time.

**2.3.4 Final Analysis**

Forecasting RPC using machine learning techniques will be conducted according the proposed method by Kumar et al. (2018) with some modifications. This method consists mainly of five steps: data cleansing, model training, model optimisation, model evaluation, and data forecasting. In particular, the windowing implimentation will be conducted using the steps stated by Rasel et al. (2015). Flow chart for the proposed methodology is illustrated in Figure 4. The flow of the final analysis and discussions will arranged based on this flow chart.

Data Preparation

Model Training

Data Cleansing

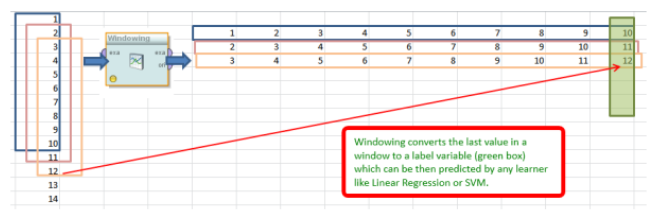
Model Optimisation

Data Forecasting

Model Evaluation

Figure 4 Flow Chart of RPC Forecasting

The first step to be done after obtaining the dataset will be **data cleansing** by replacing the empty valued columns with moving average values of the previous 2 quarters which refers to number of quarters with the same seasonal patterns of the time series. After data cleansing will be **data preparation** stage. At this stage, data will be splitted into training and testing sets with the ratio of 70 : 30. Both of these sets will be prepared for model training and testing by transforming them into window inputs. According to Rasel et al. (2015), these windows will be generated by transposing the column of RPC indicators into horizontal windows in which the last row will become the target variable to be predicted. This sliding window will move horizontally from the beginning of the time series until the end through a time-based cross-validation process. Figure 5 illustrates the concept of transposing a column into horizontal windows and Figure 6 displays the concept of sliding window during time-based cross validation process.

Figure 5 Generation of horizontal windows via column transpose

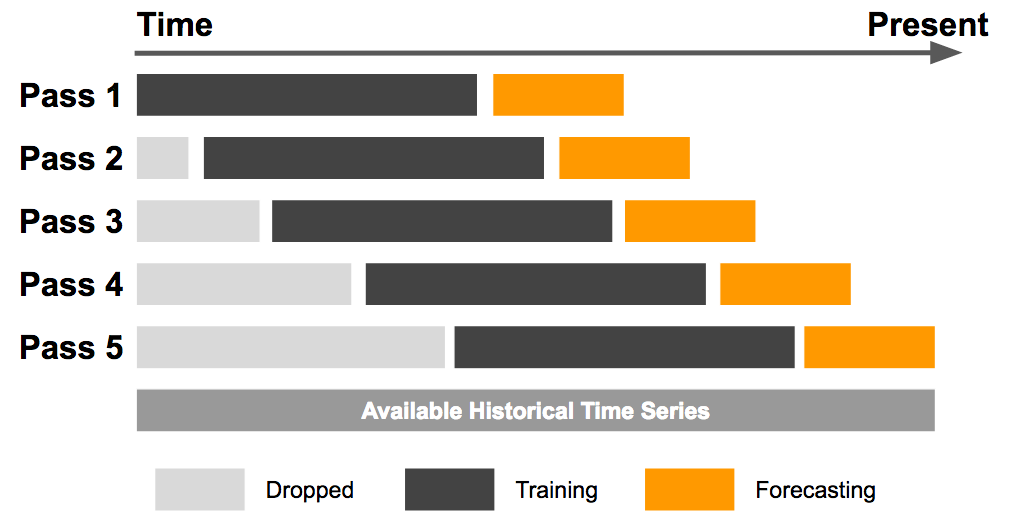


Figure 6 Illustration of sliding window concept for model training, and prediction

After the data has been prepared, the windowed data will be input into RF and SVM models for **model training** stage. Throughout the training, the models will be evaluated using RMSE evaluation metrics similar to the current evaluation metrics currently being used by MoF. Model training and testing will be repetitively conducted together with model optimization to achieve the best model performance. During **model optimization** stage, two components will be tested to come out with the best model performance which are feature selection, and window size. The best attributes to be used in model training will be determined using feature selection available in python’s scikit learn library while the best window size will be identified using loop function. Before forecasting, model performances will be evaluated in terms of their evaluation metrics during the **model evaluation** stage. Lastly, future data of RPC for third quarter of 2020 will be forecasted using the best model evaluated. This forecasted value will be presented to client during Customer Acceptance stage of Data Science Project Lifecycle.

**2.4 Proposed Solution**

The forecasted value of Malaysia’s RPC of third quarter of 2020 will be presented to the client together with the methods used in order to achieve to the forecasted value. This forecasted value will be evaluated by the client whether it is acceptable or not. Plus, the forecasted value also will be compared with the actual RPC value after it is officially published by DoSM.

**2.5 Justification on Selected Data Science Techniques and Tools**

Based on a comparative study done by Kumar et al. (2018), they found out that Random Forest and Support Vector Machine performed the best for stock market prediction which is related to economics. Therefore, since the target variable of this project is the Real Private Consumption which is also an econometrics of GDP, thus RF and SVM will be selected as they has the highest accuracy in predicting economics variables and robust against overfitting issues suffered by other algorithms (Kumar et al., 2018). Hence, it is justified that RF and SVM will be selected in this project due to their robustness against overfitting.

Despite of having abundant statistical tools online for RPC forecasting, python language is still the best among data scientist due to its dynamic and flexible usage. According to Burns and Whyne (2018), python has multiple open source libraries for time series forecasting using windowing techniques such as tslearn, cesium-ml, ts-fresh, and seglearn. However, they also found out that only the seglearn library is compatible to be used with machine learning models from the scikit learn library, plus with windowing feature. Other than seglearn and scikit learn libraries, other basic libraries such as “Pandas”, “Matplotlib”, and “Numpy”, will be included in this project for data cleansing, data preparation, and visualisations. Hence, it is justified that python will be selected in this project due its dynamic usage throughout this project flow chart.

**References**

Afandi, A. and Khoo, R. (2020). Ringgit will not face extreme volatility thanks to managed float. *Bernama*. Retrieved on Oct 18, 2020 from https://www.bernama.com/en/general/

news\_covid-19.php?id=1816684

Asada, H., Kiang, T. K., Espinoza, R., and Vandeweyer, M. (2019). *OECD Economic Surveys 2019: Malaysia*. Retrieved on Oct. 8. 2020, from http://www.oecd.org/economy/surveys/ Malaysia-2019-OECD-economic-survey-overview.pdf

Bank Negara Malaysia (2020). *Annual Report 2019*. Retrieved on Oct. 15, 2020 from https://www.bnm.gov.my/ar2019/files/ar2019\_en\_full.pdf

Bank Negara Malaysia (2019). *Annual Report 2018*. Retrieved on Oct. 15, 2020 from https://www.bnm.gov.my/files/publication/ar/en/2018/ar2018\_book.pdf

Bank Negara Malaysia (2018). *Annual Report 2017*. Retrieved on Oct. 15, 2020 from https://www.bnm.gov.my/files/publication/ar/en/2017/ar2017\_book.pdf  
Bank Negara Malaysia (2017). *Annual Report 2016*. Retrieved on Oct. 15, 2020 from https://www.bnm.gov.my/files/publication/ar/en/2016/ar2016\_book.pdf

Bernama (2018). *Azmin: Statistical Community needs to Embrace Digital Revolution*. The Edge Markets. Retrieved on Oct. 9, 2020, from https://www.theedgemarkets.com/article/azmin- statistical-community-needs-embrace-digital-revolution

Brownlee, J. (2020). *Introduction to Time Series Forecasting with Python: How to Prepare Data and Develop Models to Predict the Future*.Machine Learning Mastery. Retrieved on Oct. 16, 2020, from https://books.google.com.my/books?hl=en&lr=&id=-AiqDwAAQBAJ&oi=fnd&pg=PP1&dq=time+series+forecast+tools+machine+learning+ +python&ots=XfrnA3WACq&sig=JPma- RkDnY0Vcff2rjJQwU\_OUZg&redir\_esc=y#v=onepage&q=time%20series%20forecast %20tools%20machine%20learning%20%20python&f=false

Burns, D. M., and Whyne, C. M. (2018). Seglearn: A python package for learning sequences and time series. *Journal of Machine Learning Research*, 19(1), pp. 3238-3244.

Dematos, G., Boyd, M.S., Kermanshahi, B. (1996). Feedforward versus recurrent neural networks for forecasting monthly japanese yen exchange rates. *Financial Engineering and the Japanese Markets,* **3,** pp. 59–75.

Department of Statistics Malaysia (2020). *National Accounts FAQ.* Retrieved on Oct. 8, 2020 from https://www.dosm.gov.my/v1/index.php?r=column/cone&menu\_id=dUtRR1JYWjk

2TEJha1BrZml0REY4UT09

Fadzil, M., Latif, L. A., and Munira, T. A. (2015). MOOCsin Malaysia : A preliminary case study. *MOOCs and Educational Challenges around Asia and Europe,* 1(6), pp. 65-86.

Hackeling, G. (2017). *Mastering Machine Learning with scikit-learn*. United Kingdom: Packt Publishing Ltd.

Hashim, E., Ramli, N. R., Romli, N., Jalil, N. A., Bakri, S. M., and Ron, N. W. (2018). Determinants of Real GDP in Malaysia. *The Journal of Social Sciences Research*, No. 3, pp. 97-103.

Kumar, I., Dogra, K., Utreja, C., and Yadav, P. (2018). A Comparative Study of Supervised Machine Learning Algorithms for Stock Market Trend Prediction. *2018 Second International Conference on Inventive Communication and Computational Technologies (ICICCT)*, pp. 1003-1007

McKinney, W., Perktold, J., and Seabold, S. (2011). Time Series Analysis in Python with statsmodels. *Proceedings of the 10th Python in Science Conference,* pp 107-113

Microsoft (2020). *The Business Understanding Stage of the Team Data Science Process Lifecycle.* Retrieved on Oct 17, 2020, from https://docs.microsoft.com/en-us/azure/machine-learning/team-data-science-process/lifecycle-business-understanding

Rasic, A. H. (2019). Consumer Sentiment, Business Condition Indexes Down in Q4. *New Straits Times.* Retrieved on Oct. 17, 2020, from https://www.nst.com.my/business/2019/01/456084/consumer-sentiment-business-condition-indexes-down-q4

Rasel, R. I., Sultana, N., & Meesad, P. (2015). An efficient modelling approach for forecasting financial time series data using support vector regression and windowing operators. *International Journal of Computational Intelligence Studies*, *4*(2), pp. 134-150.

Razak, N. A. A., Khamis, A., and Abdullah, M. A. A. (2017). ARIMA and VAR Modeling to Forecast Malaysian Economic Growth. *Journal of Science and Technology: Special Issue on the Application of Science and Mathematics,* 9(3), pp. 16-24.

Roy, M., and Larocque, D. (2012).Robustness of Random Forests for Regression. *Journal of Nonparametric Statistics*, 24(4), pp. 993-1006

Taieb, S. B. (2014). *Machine learning Srategies For Multi-Step-Ahead Time Series Forecasting*. Université Libre de Bruxelles Belgium. Retrieved on Oct. 9, 2020 from http://souhaib- bentaieb.com/pdf/2014\_phd.pdf

United Nations, (2020). *World Economic Situation and Prospects 2020*. United Nations Publication. Retrieved on Oct. 7, 2020 from https://www.un.org/development/desa/dpad/ wp-content/uploads/sites/45/WESP2020\_FullReport.pdf

Usher, J., and Dondio, P. (2020). BREXIT Election: Forecasting a Conservative Party Victory through the Pound using ARIMA and Facebook’s Prophet. In *Proceedings of the 10th International Conference on Web Intelligence, Mining and Semantics*, pp. 123-128.

Vo, V., Luo, J., and Vo, B. (2016). Time Series Trend Analysis Based on K-Means and Support Vector Machine. *Computing and Informatics*, 35, pp. 111-127

World Bank Group (2020). *Malaysia Economic Indicator.* Retrieved on Oct. 18 from https://openknowledge.worldbank.org/bitstream/handle/10986/33960/149872.pdf? sequence=4&isAllowed=y

Yu, S. (1999), Forecasting and Arbitrage of the Nikkei Stock Index Futures: An Application of Backpropagation Networks. *Asia-Pacific Financial Markets,* **6,** pp.341–354 .