PREDICTIVE ANALYTICS FOR FAST MOVING ITEM USING NONLINEAR REGRESSSION MODELS

NUR ARISHA BINTI MOHD AZHAR

A project report submitted in partial fulfilment of
the requirements for the award of the degree of
Master of Science (Data Science)

School of Computing
Faculty of Engineering
Universiti Teknologi Malaysia

FEBRUARY 2021

ACKNOWLEDGEMENT

In preparing this thesis, I was in contact with many people, researchers, academicians, and practitioners. They have contributed towards my understanding and thoughts. In particular, I wish to express my sincere appreciation to my main project supervisor, Dr Nor Erne Nazira Binti Bazin, for encouragement, guidance, critics and friendship. Without her continued support and interest, this project report would not have been the same as presented here. I am also very thankful to DHL's data scientist Mr. Shawn Kwek and Mr. Yee Jia Hao for their guidance, advices and motivation in completing my internship and ensuring supports for data usage is available for the construction of this report.

My fellow classmates should also be recognised for their support, especially Natasha and Durkah. Thank you for the companionship that has made this journey a little less lonely. My sincere appreciation also extends to all my colleagues and others who have provided assistance at various occasions. I am grateful to all my family member for their undying support for me to complete my study at this level especially to both of my parent. Without their love, support and motivation, I could not have completed this study.

ABSTRACT

A supply chain and logistics company are trying to realign their stock placement in the warehouse based on the type of movement of the stock keeping units (SKU), fast moving or slow moving. This project is executed to construct a nonlinear regression models for order frequency per month of fast moving items using Python programming language. The variables used for this prediction model is the median order frequency per month for each warehouse, total quantity of item, total volume of item and total value of item. The project framework has been set up with the inclusion of data visualization for the type of movement of each SKU for each warehouse using Tableau software. SKU are segmented by comparing the average frequency of order for each SKU in the span of 33 months with the median frequency of order for each respective warehouse the SKU resides in. Three nonlinear regression based models are used to construct the predictive model which are Decision Tree Regression, Random Forest Regression and Extreme Gradient Boosting Algorithms. Parameters tuning for the model carried out by using RandomizedSearchCV from scikit-learn library. Random forest produce the smallest error rate for prediction by using mean square error with an average value of 1.2608 and mean absolute error with an average value of 0.4496 as model evaluation and holdout method as model validation in this study.

ABSTRAK

Sebuah syarikat rantaian bekalan dan logistik berusaha menyusun semula penempatan stok di dalam gudang berdasarkan jenis pergerakan unit penyimpanan stok inventori (*SKU*), bergerak pantas atau bergerak lambat dalam pengaturan inventori pelbagai item. Projek ini dilaksanakan untuk membina model regresi tidak linear untuk frekuensi pesanan setiap bulan bagi item bergerak pantas menggunakan bahasa pengaturcaraan *Python*. Pembolehubah yang digunakan untuk model ramalan ini adalah frekuensi pesanan median setiap bulan untuk setiap gudang, jumlah kuantiti barang, jumlah isipadu barang dan jumlah nilai barang. *SKU* bergerak pantas disegmentasikan dengan membandingkan purata kekerapan pesanan bagi setiap *SKU* dalam jangka masa 33 bulan dengan frekuensi pesanan median setiap gudang di mana *SKU* berada. Set data yang digunakan telah melalui pra-pemprosesan termasuk ketanpanamaan data. Tiga model berasaskan regresi tidak linear digunakan untuk membina model ramalan iaitu *Decision Tree Regression, Random Forest Regression* dan *Extreme Gradient Boosting Algorithms*. Kerangka projek disusun dengan merangkumi visualisasi data bagi jenis pergerakan setiap *SKU* untuk setiap gudang menggunakan perisian Tableau. Penetapan parameter dilakukan dengan menggunakan *RandomizedSearchCV* dari perpustakaan *scikit-learn*. *Random Forest Regression* menghasilkan kadar kesalahan terkecil untuk ramalan dengan menggunakan ralat kuadrat min dengan nilai 1.2608 dan ralat mutlak min dengan nilai 0.4496 sebagai penilaian model dan *holdout method* sebagai pengesahan model dalam kajian ini.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ANN	-	Artificial Neural Network
BDA	-	Big Data Analytics
BI	-	Business Intelligence
BIA	-	Business Intelligence and Analytics
BLUE	-	Best Linear Unbiased Estimator
CART	-	Classification and Regression Trees
ETL	-	Extract, Transform and Load
GLC	-	Government-Linked Companies
MAE	-	Mean Absolute Error
MAPE	-	Mean Absolute Percentage Error
MSE	-	Mean Square Error
RF	-	Random Forest
RMSE	-	Root Mean Square Error
SCL	-	Supply Chain and Logistics
SKU	-	Stock Keeping Unit
SVR	-	Support Vector Regression
TM	-	Telekom Malaysia
VPN	-	Virtual Private Network
XGB	-	Extreme Gradient Boosting

LIST OF SYMBOLS

α	-	Intercept
β	-	Coefficient of x
ε	-	Error
n	-	Total number of sampled data
y_i	-	Actual value of the dependent variable
$\widehat{y_i}$	-	Predicted values of the dependent variable
∞	-	Infinity

LIST OF APPENDICES

CHAPTER 1

INTRODUCTION

1.1 Problem Background

Supply chains and logistics companies' are exposed to many types of risks while trying to maintain a high customer service level may involve a situation of keeping high stock of inventory. In achieving customer's expectation of expected service deliverance, sensitive decisions on number and point of inventory that should be secured and hold within the company's supply chain is highly important. This shows how warehousing functions importance in maintain the whole network in SCL business. Getting the right product at the exact time and exact location is an immeasurable importance highlighted in any supply chain network. It is deemed important as this perfect arrangement efficiently expands product time availability to customers hence boost up customer service while minimizing costs (Kasimov, 2016). Understanding this element in a supply chain and logistic company, segmentation of inventory helps figure out the demand and movement of items in the warehouse. A better understanding of the product's motion (fast or slow moving) and conduct are essential so that prediction of inventory can be executed effectively.

The trends in this type of industries are increasingly moving toward demand-driven supply-chain solutions to reduce issues such as time-to-market, stock-outs, overstocks, inventory levels and improve forecast accuracy and overall performance. Hence, analytics solutions available in market have to ensure a complex view of true demand signals across a network, predict demand across a product lifecycle, capture a demand signal closer to its source, analyze and steer demand in line with business objectives and apply proper strategy to demand segmentation. Data analytics techniques enhance business sectors with higher amount of prediction accuracy in

ever-changing environment and provide future insight by leveraging a data-driven approach.

Predictive analytics has helped supply chain and logistics companies to satisfy the customer's demands proficiently. Truth be told, predictive analytics has been singled out by logistics industry that massively influence the changes in SCL network. 93% of logistics based company and 98% of third-party logistics firms consider performance and quality of their network is improved using data-driven decision-making. This conducted inquiry performed by the Council of Supply Chain Management Professionals also disclosed that 71% of the inquired firms believes that that big data provide a more cost savvy solution through implementation of predictive models for their supply network (Langley, 2018). Through construction and implementation of a customized demand prediction models, supply chain and logistics companies can effectively accomplish a more accurate prediction in various ways. These models can help companies to understand precisely in managing their key performance index in achieving an efficient management in their supply chain.

1.2 Problem Statement

The company operates multi-site retail warehouses strategy with the objective of streamlining e-commerce operations and staying competitive in SCL domain. Fulfilment of orders from online purchases are directly shipped out to customers directly from warehouse without using distribution centers. Maintaining optimal level of stock in each warehouse inventory becomes a necessity. Excellent management in this division ensures a right amount of items are accessible in the right place and quantities at each point of supply channel. In maintaining high level of customer service, accurately planned inventory especially in fast moving items is crucial. Inventory management is highly liable towards an organizational performance index. In ensuring an effective answering to the customer needs, companies must ensure that the fast moving items are consistently accessible. Highly important aspect in customer's relationship management relies on outbound logistics as it represents the deliverance of item to the end front of the business which

are the customers. This specific logistics movement plays an important role on the success of the business as the performance played out on outbound logistics influence the decision of stocking up a product or not based on the numbers it sells out to the customer. However, high stock inventory from this strategy incurs a high cost in holding the stock in the warehouse. Hence, inventory segmentation in business sectors are necessary to identify the most prospective item and therefore able provide a strategic planning in managing their resources and production capacities with the main focus of meeting the demand ensued by customer without compromising the aspect of cost effectiveness. Company's fallout in maintaining an effective service to the customer will incur cost across organization due to possibility of loss sales but maintaining a high inventory level without a proper inventory identification incur a high investment. Hence, anticipation of number of order's that can come in a specified amount of time can be determine using order frequency metric. Amount of workload, job assignments to workers and effectiveness of sales strategies launched can be established by staying on top identified order frequency which in turn ensure stock allocation plan can be successfully streamlined.

Variation of product items in inventory demands specific inventory policies influenced by their specific characteristics which can be inferred in various condition such as storage requirements, sales volume, product value and also predictability of demand. Considering wide arrays of possibilities on managing various items of inventory, management of inventory system of the company is a complex situation. Therefore a small number of SKU classes based on the characteristics of these SKUs is usually regarded as useful to differentiate between each other in inventory setting. Same type of SKU does not has the same type of movement in each warehouse. Challenges in segmenting the multi items inventory based on type of movement in this case study also due to the factor of multi-site warehouse usage of the company. Techniques like ABC analysis and XYZ analysis are widely used for inventory segmentation (Biswas et al., 2017). These techniques disparate inventory items into numbers of arrangements dependent on the annual cost, quantity or volume of the SKU. Time series and regression based analysis are among the various statistical analysis techniques have been used for prediction in supply chain and management (Seyedan & Mafakheri, 2020). With the progressions in data advancements and improved computational efficiencies, big data analytics has arisen as a methods for

showing up at more exact prediction that better reflect client needs, encourage performance in execution, improve the productivity, lessen response time, and support risk evaluation.

1.3 Study Questions

In reaching the objectives of this proposed study, the following questions are established:

(a) What are the suitable parameter that can be used to segment each SKU?

(b) What is the appropriate representation that can be provided to the stakeholders to assess the different segmentation of SKU?

(c) What type of regression model that is suitable to predict order frequency per month for fast moving item?

1.4 Project Goals

The objectives of the project are:

(a) To categorize outbound SKU into two types of movement, fast moving or slow moving.

(b) To provide visualization dashboard that portray each SKU's type of movement for each warehouse.

(c) To predict order frequency per month for fast moving items using nonlinear regression model.

1.5 Scope

The focus of this study is on utilizing predictive analytics for fast moving items in a supply chain and logistic company. The study region of this project entails on investigating the inventory management aspect specifically on the outbound stock movement. The segmentation of stock keeping units are executed to identify fast moving items from the outbound data that need to be prioritized in a high multi-product inventory. Then, nonlinear regression model is used to predict the order frequency for these fast moving items that will assist the decision-making process for the stock management. Software that is utilized in this project for visualization is Tableau and JupyterHub is utilized for data transformation and run in order to code the problem and arrive to the analytics aspect.

1.6 Importance of Study

An efficient supply chain and logistics network significantly provides an upper hand and offers a competitive advantages for all related divisions of the organization. Two crucial aspects of SCL performance are efficiency and effectiveness in which this study will help to identify the balance between both aspects in inventory management specifically in SKU segmentation and predictive analytics measures for fast moving items. The provision of SKU segmentation analysis performed in this study enable SCL companies to carry out a data-driven decisions of inventory management related to stock allocation optimally, effectively and penultimate cost efficient.

1.7 Chapter Summary

This chapter conveys basic details concerning the study that is executed. To this end, first a background of the study is presented. Then, the problem of the study is raised based on observed necessities in the related sector domain. Once the problem is clarified, objectives of the study are defined. The objectives address the study problem. This chapter also discusses scope of the study within which the study is performed. Furthermore, significance of the study is discussed in detail. Finally, questions which are relied upon to be tended to by the end of this study are presented.

REFERENCES

Ajah, I. A., & Nweke, H. F. (2019). Big Data and Business Analytics: Trends, Platforms, Success Factors and Applications. *Big Data and Cognitive Computing*, *3*(2), 32. https://doi.org/10.3390/bdcc3020032

Asana, I. M. D. P., Radhitya, M. L., Widiartha, K. K., Santika, P. P., & Wiguna, I. K. A. G. (2020). *Inventory control using ABC and min-max analysis on retail management information system*. *1469*(1). Scopus. https://doi.org/10.1088/1742-6596/1469/1/012097

Biswas, S. K., Karmaker, C. L., Islam, A., Hossain, N., & Ahmed, S. (2017). *Analysis of Different Inventory Control Techniques: A Case Study in a Retail Shop*. *6*(3), 11.

Boulaksil, Y. (2016). Safety stock placement in supply chains with demand forecast updates. *Operations Research Perspectives*, *3*, 27–31. https://doi.org/10.1016/j.orp.2016.07.001

Ekawati, R., Kurnia, E., Wardah, S., & Djatna, T. (2019). Predictive Demand Analytics for Inventory Control in Refined Sugar Supply Chain Downstream. *2019 International Seminar on Application for Technology of Information and Communication (ISemantic)*, 100–104. https://doi.org/10.1109/ISEMANTIC.2019.8884293

Garre, A., Ruiz, M. C., & Hontoria, E. (2020). Application of Machine Learning to support production planning of a food industry in the context of waste generation under uncertainty. *Operations Research Perspectives*, *7*, 100147. https://doi.org/10.1016/j.orp.2020.100147

Hanafi, R., Mardin, F., Asmal, S., Setiawan, I., & Wijaya, S. (2019). *Toward a green inventory controlling using the ABC classification analysis: A case of motorcycle spares parts shop*. *343*(1). Scopus. https://doi.org/10.1088/1755-1315/343/1/012012

Ho, D. C. K., Mo, D. Y. W., Wong, E. Y. C., & Leung, S. M. K. (2019). Business intelligence for order fulfilment management in small and medium enterprises. *International Journal of Internet Manufacturing and Services*, *6*(2), 169–184. Scopus. https://doi.org/10.1504/IJIMS.2019.098231

Huang, Y., Yuan, Y., Chen, H., Wang, J., Guo, Y., & Ahmad, T. (2019). A novel energy demand prediction strategy for residential buildings based on ensemble learning. *Energy Procedia*, *158*, 3411–3416. https://doi.org/10.1016/j.egypro.2019.01.935

Igual, L., & Seguí, S. (2017). *Introduction to Data Science*. Springer International Publishing. https://doi.org/10.1007/978-3-319-50017-1

Jeong, J. H., Woo, J. H., & Park, J. (2020). Machine Learning Methodology for Management of Shipbuilding Master Data. *International Journal of Naval Architecture and Ocean Engineering*, *12*, 428–439. https://doi.org/10.1016/j.ijnaoe.2020.03.005

Kasimov, I. (2016). *Issues in Logistics and Supply Chain Management, Bullwhip Effect and Warehouse Management*.

Langley, J. (2018). *Capgemini Consulting (2017) 2017 Third-Party Logistics Study: The State of Logistics Outsourcing. Results and Findings of the 21st Annual Study*. Penn State University, available at: www. 3plstudy. com/media/downloads/2016 ….

Larson, D., & Chang, V. (2016). A review and future direction of agile, business intelligence, analytics and data science. *International Journal of Information Management*, *36*(5), 700–710. https://doi.org/10.1016/j.ijinfomgt.2016.04.013

Oo, M. C. M., & Thein, T. (2019). An efficient predictive analytics system for high dimensional big data. *Journal of King Saud University - Computer and Information Sciences*. https://doi.org/10.1016/j.jksuci.2019.09.001

Petropoulos, A., Siakoulis, V., Stavroulakis, E., & Vlachogiannakis, N. E. (2020). Predicting bank insolvencies using machine learning techniques. *International Journal of Forecasting*, *36*(3), 1092–1113. https://doi.org/10.1016/j.ijforecast.2019.11.005

Pinto, T., Praça, I., Vale, Z., & Silva, J. (2021). Ensemble learning for electricity consumption forecasting in office buildings. *Neurocomputing*, *423*, 747–755. https://doi.org/10.1016/j.neucom.2020.02.124

Ribeiro, M. H. D. M., & dos Santos Coelho, L. (2020). Ensemble approach based on bagging, boosting and stacking for short-term prediction in agribusiness time series. *Applied Soft Computing*, *86*, 105837. https://doi.org/10.1016/j.asoc.2019.105837

Saggi, M. K., & Jain, S. (2018). A survey towards an integration of big data analytics to big insights for value-creation. *Information Processing & Management*, *54*(5), 758–790. https://doi.org/10.1016/j.ipm.2018.01.010

Schmidt, A. F., & Finan, C. (2018). Linear regression and the normality assumption. *Journal of Clinical Epidemiology*, *98*, 146–151. https://doi.org/10.1016/j.jclinepi.2017.12.006

Seyedan, M., & Mafakheri, F. (2020). Predictive big data analytics for supply chain demand forecasting: Methods, applications, and research opportunities. *Journal of Big Data*, *7*(1), 53. https://doi.org/10.1186/s40537-020-00329-2

Wang, C.-H., & Chen, T.-Y. (2020). Combining biased regression with machine learning to conduct supply chain forecasting and analytics for printing circuit board. *International Journal of Systems Science: Operations & Logistics*, *0*(0), 1–12. https://doi.org/10.1080/23302674.2020.1859157

Wang, X. (Shane), Ryoo, J. H. (Joseph), Bendle, N., & Kopalle, P. K. (2020). The role of machine learning analytics and metrics in retailing research. *Journal of Retailing*. https://doi.org/10.1016/j.jretai.2020.12.001

Zhang, W., Wu, C., Zhong, H., Li, Y., & Wang, L. (2021). Prediction of undrained shear strength using extreme gradient boosting and random forest based on Bayesian optimization. *Geoscience Frontiers*, *12*(1), 469–477. https://doi.org/10.1016/j.gsf.2020.03.007

Zougagh, N., Charkaoui, A., & Echchatbi, A. (2020). Prediction models of demand in supply chain. *Procedia Computer Science*, *177*, 462–467. https://doi.org/10.1016/j.procs.2020.10.063