LEVERAGING MACHINE LEARNING TO UNDERSTAND THE EFFECT OF COVID-19 ON THE ECONOMY AND INDIVIDUAL

Khor Woon Shing[1] and Rathimala Kannan[2,*]
[1,2]*Department of Information Technology/ Faculty of Management, Multimedia University, Cyberjaya, Malaysia*
*rathimala.kannan@mmu.edu.my

Abstract

Globally COVID-19 pandemic has caused drastic changes in individual's life in terms of social and economic factors. Many countries have implemented various stimulus packages to combat the pandemic effects. In order to reduce the financial burdens of the citizens, the Malaysian government has created a series of stimulus packages. But there is a significant lag when considering the impact of COVID-19 on the economy and individuals' life. To address this knowledge gap, this paper aims to determine how the pandemic affects the economy and individuals by leveraging data analytics and machine learning techniques. The secondary data of Prihatin Rakyat Economic Stimulus Package first survey from DOSM is used. Research methodology is based on the cross-industry standard process for data mining(CRISP-DM). As the dataset is noisy with missing values and related issues, considerable time is spent on cleaning the dataset before applying descriptive analytics techniques. Supervised and unsupervised machine learning algorithms will be applied to uncover the insights from the dataset. The findings would help to determine the needs and preferences of various types of individuals to lessen their burdens in this pandemic period. Also, this knowledge could be useful to design a personalised recommendation system for future stimulus packages.

Keywords—COVID-19, Economic stimulus packages, data analytics and machine, B40 group

I. INTRODUCTION

The COVID-19 pandemic has caused a global crisis and has created a major impact on the health and economy of many countries (Lim, 2020). According to the statistics from the Ministry of Health, there were accumulated 298,315 confirmed infected COVID-19 cases and 1,121 deaths in 13 states and 3 federal territories of Malaysia until 27th February 2021(Pfordten & Ahmad, 2021). Malaysia had implemented Movement Control Order (MCO) to prevent and control the spread of the COVID-19 on 18th March 2020. Though MCO is the immediate need to combat the pandemic, it has generated multiple ripple effects, causing major economic disruptions as most economic activities were forced to cut back significantly, if not stop completely. Because of the lockdown, employees were forced to work from home and even some businesses were forced to close down due to losses and this made some people had lost their job in this pandemic (Shah et al., 2020). The third stimulus package was released to support SMEs and was called as Prihatin SME Economic Stimulus Package. This package was aimed to secure employees from getting terminated which otherwise will lead to job loss and a high unemployment rate.

Without a source of income, many families could quickly lose access to essentials needs like food or housing (Flanders, Nungsari, & Chuah, 2020). In order to reduce the financial burdens of the citizens, the Malaysian government has created a series of stimulus packages. The first stimulus package was aimed to support the business, tourism and hospitality sector. It also focused on promoting high value-added investments in both the public and private sectors. The second stimulus package is named the Prihatin Rakyat Economic Stimulus Package and implement on 27th March 2020. This stimulus package was created to support the M40 and B40 income groups in easing the financial burden of households and individuals in the form of a direct cash transfer program. Even though the government has taken many actions to reduce the pandemic effect on its people, there is limited knowledge on the real impact on citizens from various income groups and age levels. Department of Statistics Malaysia (DOSM) conducted two surveys to understand how the pandemic affects the economy and individuals. The first survey was conducted between 23rd to 31st of March 2020 prior implementing the MCO and had received 168,145 responses. The second survey was conducted between 10th to 24th 2020, after the first MCO and had received 41,386 responses. These two survey data were made open to the public and researchers (Prime Minister office, 2020) and hence this study is motivated to further extract the insights.

This paper aimed to apply data analytics and machine learning techniques on the first survey data to extract insights that would be helpful to gain a better understanding of COVID-19 impacts on the economy and individuals. The findings would help to determine the needs and preferences of various types of people to lessen their burdens. Also, this knowledge could be useful to design a personalised recommendation system for future stimulus packages.

II. METHODOLOGY

This research used Cross-Industry Standard Process-Data Mining (CRISP-DM) to plan and execute the project. This methodology consists of 6 phases which are business understanding, data understanding, data preparation, modelling, evaluation and deployment (Schröer, Kruse, & Gómez, 2021). In the business understanding phase, the study's objective identified was to extract the insight from the survey data and determine the effect of COVID-19 on the economy and individuals in Malaysia by applying machine learning techniques. The data understanding phase is to select the right dataset and understand the variables.

In this study, KNIME analytics platform used for executing machine learning and the online survey conducted by the Department of Statistics Malaysia to determine the effects of the COVID-19 pandemic on the economics and status of citizens had been applied. First, the whole dataset translated into the English language from the Malay language. After that, redefine and shorten the variables' names to be more understandable. There are 6,365 rows of missing records found in the dataset. These 6,365 missing records had been filtered and separated into the new table for visualization purposes. Then, the unused variables had been filtered and string manipulation is done in the dataset for more understandable. Lastly, the linear correlation also had been run to view the relationship between the variables. For the data preparation phase, the convert string records into numeric records to do the preparation before building the models. After that, the string variables filtered from the dataset and the normalizer node apply to normalize the dataset. Next, in the modeling phase, the different algorithms will be applied to achieve the objectives of the study. The prediction models will be applied such as decision tree, random forest, and other algorithms to predict the citizens' readiness to cope with the pandemic. Besides, the clustering models will be applied such as K-Means and K-Medoids to predict and identify the group of citizens that need the government's support. For the evaluation phase, the standard metrics results of each algorithm will be compared and evaluated. For the deployment phase, the best performance models will be deployed and review the performance with the study's objectives.

III. RESULTS/FINDINGS

This paper reports an ongoing study's results which is obtained from descriptive analytics. This study used the secondary data Prihatin Rakyat Economic Stimulus Package survey from DOSM. Survey data has totally 31 variables; 6 variables related to demographic information, 3 variables related to financial preparedness for MCO, 11 variables related to job status and income level, and 11 variables related to changes in lifestyle before and after COVID-19. The majority of the respondents were from Selangor (43,560), followed by Johor (33,341), Kuala Lumpur (13,787), and Sabah (11,359). From the remaining states on average, there were 5860 responses each. This survey has high participation from females (95,158) compared to males (66,622). In terms of financial preparedness to face the pandemic, almost 50% of the participants said not prepared. Their income level stated that the majority from the B40 group. Following the MCO, 30,903 said they faced a reduction in their income while 47,173 said it remained the same.

IV. CONCLUSION

The research explores the COVID-19 survey dataset by leveraging data analytics and machine learning techniques. The findings would help to determine the needs and preferences of various types of individuals to lessen their burdens in this pandemic period. Also, this knowledge could be useful to design a personalised recommendation system for future stimulus packages.

ACKNOWLEDGMENT

We thank the Department of Statistics Malaysia for allowing us to use "Effects of Covid-19 on the Economy and Individual'- Round 1 survey data to conduct this research.

REFERENCES

Flanders, S., Nungsari, M., & Chuah, H. Y. (2020). The COVID-19 Hardship Survey: An Evaluation of the Prihatin Rakyat Economic Stimulus Package, (April), 1–44. Retrieved from https://www.researchgate.net/publication/340769013_The_COVID-19_Hardship_Survey_An_Evaluation_of_the_Prihatin_Rakyat_Economic_Stimulus_Package

Lim, L. L. (2020). *The socioeconomic impacts of COVID-19 in Malaysia : Policy review and guidance for protecting the most vulnerable and supporting enterprises*. *International Labour Organization*.

Pfordten, D., & Ahmad, R. (2021). Covid-19: Cases up by 3,100 bringing total to 245,552 (updated daily). Retrieved February 28, 2021, from https://www.thestar.com.my/news/nation/2020/03/23/covid-19-current-situation-in-malaysia-updated-daily

Prime Minister office, P. (2020). *PRIHATIN RAKYAT ECONOMIC STIMULUS PACKAGE2020*.

Schröer, C., Kruse, F., & Gómez, J. M. (2021). A Systematic Literature Review on Applying CRISP-DM Process Model. *Procedia Computer Science*, *181*, 526–534. https://doi.org/10.1016/j.procs.2021.01.199

Shah, A. U. M., Safri, S. N. A., Thevadas, R., Noordin, N. K., Rahman, A. A., Sekawi, Z., … Sultan, M. T. H. (2020). COVID-19 outbreak in Malaysia: Actions taken by the Malaysian government. *International Journal of Infectious Diseases*, *97*, 108–116. https://doi.org/10.1016/j.ijid.2020.05.093