CHAPTER 5

CONCLUSION AND FUTURE WORKS

5.1 Conclusion

This project has analysed and predicted the health expenditure in Malaysia. The healthcare spending data is collected from Ministry of Health Malaysia, WHO Global Health Expenditure Database and World Development Indicators Database for the period of 2000 to 2022. The collected datasets are pre-processed and merged for analysis and forecasting. Explanatory data analysis shows a gradual increase in health expenditures from 2000 to 2019, and some fluctuations between 2019 to 2022 due to the COVID-19 pandemic. Key determinants affecting health expenditure have been identified by investigating their correlations. Feature engineering has been done to select features to train the machine learning model.

From the initial findings, it can be concluded that ARIMA outperform Random Forest in forecasting Malaysia's Total Health Expenditure (TEH) from 2019 to 2022, achieving a low MAE of RM 2,191 million, a low RMSE of RM 2,731 million and a high R^2 of 0.818, indicating strong predictive accuracy. However, it is notable that in 2021, the ARIMA struggled to predict accurately due to an unexpected surge in health spending caused by the COVID-19 pandemic. In contrast, Random Forest performance is poor due to the absence of lagged values, insufficient training data and lack of hyperparameter tuning. The results conclude that a time-series model like ARIMA is suitable for health expenditure forecasting when small datasets are provided and suggest that a complex model like Random Forest requires additional data and further optimisation to perform effectively in the forecasting task. In master's project, hyperparameter tuning and validation should be prioritised to improve the accuracy and reliability of the models before forecasting Malaysia's health expenditure up to 2035.

5.2 Future works

In this project, the health expenditures of Malaysia are predicted using machine learning models. Hyperparameter tuning can be conducted for the model to further improve the results. Currently, the random forest does not predict with good accuracy and requires further tuning. Also, the lagged features of health expenditures can be calculated and used as input for the random forest model, simulating how ARIMA model works to enhance its accuracy. Cross-validation should be done to ensure the reliability of the forecast by the models.

This study is conducted based on overall health expenditure data and macroeconomic data only. The health expenses prediction can be used for smaller components in the healthcare spending, for instance, outpatient and inpatient services, pharmaceutical expenditures, education and training, provided that the accessibility to detailed data is granted. In addition, individual factors like patients' age, gender, medical conditions, current medications, income level and family history of illness can be considered for individual healthcare cost prediction. This will provide a better understanding to the end-users of Malaysia's healthcare system and promote improved health outcomes for the public.

Moreover, the methodology of this study can be extended to ASEAN countries with similar health economic structures, for instance Thailand, Indonesia and Philippines. By applying the machine learning models across multiple countries, researchers can compare the differences in health expenditure trends, key determinants of health spending and most importantly, forecasting accuracy in different nations. This may improve the generalisability of the machine learning models and provide insight for improving healthcare budget planning based on varying health policies across counties.