CHAPTER 1

INTRODUCTION

1.1 Overview

Health expenditure is defined as all the money spent on health goods and services, including preventative measures, promotion and provision of health services, nutrition, pharmaceuticals, and emergency aid. (World Health Organization [WHO], 2025a). Health funding sources from the public and private sectors include the government, individuals, private health insurance, and other non-government organizations. Health expenditure can be further classified into Total Health Expenditure (TEH), Current Health Expenditure (CHE), which excludes health-related expenditure (e.g. personnel training, research and development), General Government Health Expenditure (GGHE), and household Out-Of-Pocket Health Expenditure. (OOP).

Machine learning approaches have been used to deepen understanding and provide insight into healthcare spending. For example, researchers from Jordan predicted total healthcare expenditure for their country using two neural network strategies: Adaptive Neuro-Fuzzy Inference System and Hybrid Neural Fuzzy Inference System (Saleh et.al., 2023). Support Vector Regression (SVR) and Random Forest (RF) are performed on American healthcare expenditure and enable the prediction of healthcare expenditure as a percentage of Gross Domestic Product (GDP) for 2050, and RF provides comparable results with Autoregressive Integrated Moving Average (ARIMA) model. (Wang et. al., 2024). These researches suggest that machine learning algorithms have great potential when used in modern health economies.

In this study, machine learning techniques will be applied to the health expenditure data from Malaysia to shed light on the future health expenditure. It is

expected to provide valuable insight for policy planning related to healthcare sector in Malaysia.

1.2 Problem Background

In Malaysia, total health expenditure has been increasing from 2011 to 2022, from RM 36.9 billion to RM 78.9 billion, and as % GDP from 3.94% to 4.41% (Ministry of Health Malaysia [MOH], 2024). Recently, the Malaysia government has allocated RM 45.3 billion in Budget 2025 for the Ministry of Health for spending on healthcare alone, which is the second highest after education (Ministry of Finance Malaysia, 2024). The growing spending in healthcare raises concerns as it may cause reduced allocation of the budget for other critical fields. It also means that people living in Malaysia are spending more on healthcare expenses, which might cause a burden to those living under poor economic conditions.

Rising healthcare expenditure is a global challenge. For example, the health spending by 38 countries participating in the Organisation for Economic Co-operation and Development (OECD) is predicted to peak at 11.8% of GDP in 2040. By that time, the increase in healthcare expenses from public sources is estimated to be twice the average growth in government revenues. (Organisation for Economic Co-operation and Development, 2024) With the inflation in medication prices, medical expenses, and the ageing population in Malaysia, it is anticipated that healthcare expenditure will continue to increase. However, limited academic research has been done to predict future health expenditures in Malaysia. Therefore, there is a need for an accurate predictive model to be developed to aid in the planning of future healthcare budgets.

1.3 Problem Statement

Determinants of healthcare expenditure are complex, and their application varies depending on the prediction models (micro, macro level, or component-based). These include demographic factors such as an ageing population, health-related factors

like prevalence of chronic disease, and economic factors, for instance GDP of the country. However, the appropriate determinants of healthcare expenditure have not been established. Identifying the key determinants to be used in this project is important for precise healthcare expenditure prediction.

Furthermore, while most analyses of healthcare expenditure are done in traditional models, they present a trade-off in terms of accuracy and the time frame of prediction. Accurate prediction for each component of health expenditure is required for informed decision-making. Machine learning approaches such as Random Forest and ARIMA can provide new insights into this issue by proposing a better model in estimating future health expenditure, which helps in decision making for policymakers.

In addition, the performance of different machine learning algorithms is inconsistent in existing studies when used to predict health expenditure. This is because machine learning algorithms perform differently depending on the context of the input data. A thorough comparison of performance by RF and ARIMA on the local data is required to determine the best model in predicting the health expenditure of Malaysia.

1.4 Research Question

a) What are the key determinants of health expenditure that contribute to accurate prediction in machine learning algorithms?

b) What is the predicted health expenditure of Malaysia from 2026 to 2035 using machine learning algorithms (Random Forest and ARIMA)?

c) How do different machine learning models perform on Malaysia's health expenditure data, and which model demonstrates the highest accuracy?

1.5 Research Aim

This research aims to predict health expenditure in Malaysia using machine learning techniques to provide insight for health financing and policy planning.

1.6 Research Objectives

The objectives of this research are:

a) To identify the key determinants of health expenditure to use as features for machine learning algorithms

b) To apply Random Forest and ARIMA for predicting health expenditure in Malaysia from 2026 to 2035

c) To evaluate and compare the performance metrics of the machine learning models and to identify the model with the highest accuracy in forecasting health expenditure in Malaysia

1.7 Research Scopes

a) The data will be sourced from the Ministry of Health Malaysia, Department of Statistics Malaysia, World Bank Group, and World Health Organization.

b) The data collected will only involve demographic data and data related to health economics. No individual data that reveals an individual's medical and medication history will be used.

c) The study will use data from 2000 to 2022, providing relevant and up-to-date data for health expenses forecasting

d) This research will apply two machine learning methods: Random Forest and ARIMA.

1.8 Expected Contribution

This project can provide insights for policymakers in the country in planning health expenditures and allow strategic allocation of budgets for health expenses. This helps to ensure the long-term sustainability of funding for Malaysia's healthcare system. Overall, this research is projected to contribute to better health outcomes for the patients and people in Malaysia. The findings of this project are also expected to provide insights for other countries with similar healthcare systems or income levels.

1.9 Thesis Organisation

The following chapters are presented as outlined below:

Chapter 2 covers extensive literature reviews regarding health expenditure. This chapter explores methodology and findings on the health expenditure forecasting from existing research and identifies research gaps.

Chapter 3 dives deep into research methodology. This chapter discusses data collection and data pre-processing steps. This chapter also includes details about proposed steps for exploratory data analysis, feature engineering and selection of machine learning algorithms.

Chapter 4 covers initial findings from this project. Exploratory data analysis is conducted to gain insight from the dataset collected. Feature engineering is conducted to select the appropriate determinants of health expenditure. Initial results from the application of machine learning models, RF and ARIMA, on the total health expenditure of Malaysia are evaluated, compared and discussed in detail.

Chapter 5 ends with a conclusion about this project. Directions for future work will be proposed to extend the research outcomes.