

CHAPTER 1

INTRODUCTION

1.1 Rationale

BPJS *Kesehatan* is a health insurance most commonly used by the people of Indonesia. Based on data uploaded by BPJS *Kesehatan*, the increase in the percentage of BPJS *Kesehatan* users from 2016 to 2020 is approximately 50.5 million BPJS *Kesehatan* participants [1]. One of the factors that increase BPJS *Kesehatan* users every year is the monthly fee that is affordable for all people. Heart disease has the most significant number of claims on BPJS *Kesehatan* medical costs. It is one of the diseases in the category of catastrophic or non-communicable disease. From 2014 to 2020, the expenditure of BPJS *Kesehatan* on these diseases increased by approximately 25% to 31% of the total JKN-KIS BPJS *Kesehatan* medical costs. In 2020, heart disease accounted for 49% of the financing for catastrophic diseases [2].

With a large number of participants and the increasing cost of health services, the projected total expenditure on medical costs is likely to exceed the total income received from contributions. This condition was experienced in 2017 when the total health expenditure amounted to 84.44 trillion rupiahs, while the total contribution income reached 74.25 trillion rupiahs [1]. To address issues, it is essential to analyze and predict the costs to inform future government policies. However, the most challenging aspect of cost prediction in the healthcare sector is the complexity of the process, which involves unique requirements that often vary between individuals based on their specific needs [3].

Currently, many researchers have conducted cost prediction research. One of the most commonly used techniques is data mining techniques. Wang, Tianyi, et al. [6] conducted research in predicting medical costs using several methods; they were Multilayer Perceptron (MLP), Generalized Regression Neural Network, Extreme Learning Machine (ELM), Random Forest Regression Tree, Regression Tree, Bagging Regression Tree, Regression SVM (Polynomial and Linear), Linear Regression (LR). Among those methods, the best performance in predicting costs was Random Forest. Data mining techniques primarily emphasize the analysis of data trends. Consequently, when processing event log data, these techniques may not be optimal because they are unable to fully leverage the hidden information within the event log data for making predictions.

Process mining is a highly effective technique for analyzing processes and addressing problems using event logs [4]. Many researchers in the medical field have explored the application of process mining and achieved promising results. For instance, a study [10] focused on predicting medical costs for gastric cancer patients and evaluating different process flow schemes to identify the most optimal one. The proposed scheme recommendations showed a 25% performance improvement compared to conventional methods. Tibeme, B., Shahriar, H., & Zhang, C. in study [5] carried out comparative analysis of workflows using alpha miners, heuristic miners, inductive miners and fuzzy miners. Heuristic miners are quietly good at dealing with noise data compared to other methods, so that the results obtained are more focused on the main process flow than the behavior details in the event log. Besides that, another process mining method, Inductive Miners, is considered to be able to produce a process model consistently that fits as it produces traces in the log. In study [9], Inductive Miners can produce process models that are free from deadlock, robust to noise, and can handle complex event log data such as medical data [19]. The fusion of process mining and data mining has been pursued by Julian Theis, et al. [7] to predict patient mortality rates and explore flowing models for inpatient processes. The study concluded that process mining played a significant role in accurately predicting patient mortality rates. Additionally, incorporating process mining techniques with prediction methods addresses the limitations of prediction techniques in extracting hidden information from event log data.

1.2 Theoretical Framework

Healthcare sector was facing some challenges such as the permanent and rapid adaptation of clinical processes based on the emerging scientific evidence and limited resources that impacted to high-quality care provision. Healthcare organizations have been aware of this problem and made some improvements to the process flow of the Healthcare sector [8]. Data collected in the healthcare sector information system can be used as the right resource to be utilized for decision making and system management especially for chronic diseases [9]. However, it is not that easy to utilize big data from the healthcare sector. It takes the right technique to process the big data resource into an effective analysis.

In addition to improvements in process flow, a prediction system is also needed in the healthcare sector, one of which is in cost optimization. This cost prediction system is an effort to reduce losses felt by government health insurance such as BPJS Kesehatan in

Indonesia. With the increase in many participants and the increasing cost of health services, it is likely that the total cost of BPJS *Kesehatan* health expenditure submitted will be higher than the total contribution income received.

Based on the rationale and theoretical framework of this research, the researchers have developed a BPJS *Kesehatan* medical cost prediction system using a combination of process mining and data mining techniques. The primary focus of this prediction system is to estimate the medical costs for patients with heart disease in the BPJS *Kesehatan* dataset, specifically for Advanced Level Referral Health Facilities (FKRTL) and First Level Health Facilities (FKTP) Non-Capital. The process mining method chosen for this system is heuristic miners and Inductive Miners, which is deemed suitable for handling unstructured processes data such as health sector data [16][19]. On the other hand, the data mining method employed is random forest. The researchers aim to leverage the strengths of both techniques and create a more accurate prediction model. Aside from making cost predictions, this study also involves analyzing each process flow within the data. This analysis will generate process flow recommendations with the best performance. The results obtained from the medical cost predictions can provide valuable insights for government decision-making, such as determining whether it is necessary to increase the monthly fee from BPJS *Kesehatan*. The objective is to ensure that the proposed fee burden remains lower than the contribution income received.

1.3 Statement of the Problem

With the number of BPJS *Kesehatan* participants and the cost of health services that continue to increase, the risk of higher expenditure costs compared to the contribution costs obtained is also increasing. So the formulation of the problem in this study is as follows:

1. How the performance of combination Inductive Miner method with Random Forest to predict the treatment costs for heart disease patients?
2. How the performance of combination Heuristic Miner method with Random Forest to predict the treatment costs for heart disease patients?
3. How does the comparison of result between the Heuristic miners-Random Forest and Inductive miners-Random Fores determine which method is superior in predicting BPJS *Kesehatan* Cost.

1.4 Objective and Hypotheses

In accordance with the formulation of the problem mentioned above, this research has the following objectives:

1. Knowing the performance of combination Inductive Miner method with Random Forest to predict the treatment costs for heart disease patients.
2. Knowing the performance of combination Heuristic Miner method with Random Forest to predict the treatment costs for heart disease patients.
3. Analyzing the comparative result of Heuristic miners and Inductive Miner as process mining methods with Random Forest as regression method for predicting BPJS Kesehatan Cost.

This research hypothesizes is that the use of process mining in the BPJS Kesehatan medical expenditure cost prediction path is expected to produce more accurate predictions because it participates in analyzing the information hidden in the event log data. The process mining technique does not only look at each treatment individually and the trend of the data but will focus on the sequence of data processing so that it will produce different medical cost predictions for each process sequence. In addition, other hypotheses from this study, namely the timestamp variable, referral city, type of BPJS Kesehatan class, diagnosis of disease, treatment procedure, type of drug, and type of poly are independent variables that affect the prediction of the total BPJS Kesehatan expenses.

1.5 Assumption

The assumptions of this study include:

1. The data recorded in the BPJS Kesehatan system is accurate and valid.
2. Timestamps and patient care activities are properly integrated into the database.
3. The process mining tools function effectively and produce reliable outcomes.

1.6 Scope and Delimitation

The scope and limitations of this research are:

1. The scope of this study is focused on the treatment of heart disease patients.
2. This study will rely on limited access to data sources from Indonesia's National Health Insurance data records from 2015 to 2018.
3. The study does not include severity.

4. The data excludes information on patients' health characteristics, other factors influencing the clinical course, as well as congenital diseases and additional complications.
5. This study solely utilized BPJS *Kesehatan* Health data. For time measurements, such as those based on travel time, remaining time, layover time, and waiting time, patients who received examinations at hospitals not affiliated with BPJS *Kesehatan* were excluded from the study.

1.7 Significance of the Study

This study was conducted to provide strategic recommendations to BPJS *Kesehatan*, especially in supporting a more precise and effective decision-making process related to financial management. By utilizing accurate cost predictions based on data analysis, this study aims to help BPJS *Kesehatan* minimize the risk of expenditure exceeding total income, which can disrupt the financial stability of the organization. If this challenge can be overcome through a systematic and data-based approach, it is hoped that BPJS *Kesehatan* can improve its operational efficiency, so that the quality of services provided to patients becomes better, faster, and more equitable, which will ultimately increase participant satisfaction and support the sustainability of the national health insurance program.

This study also contributes to the existing body of knowledge by combining the process mining method and data mining method which is expected to provide appropriate cost prediction results based on the process model of BPJS *Kesehatan* Indonesia event data. These approaches have the potential to enable subsequent research in healthcare analytics.