

1. Pendahuluan (Introduction)

Initially, process mining was used to facilitate business process management activities, but over time it has begun to be used in various fields, such as health [1], financial [2], manufacturing [3], etc. Process mining encounters numerous challenges in the healthcare domain, such as the inherent variability of processes and the suboptimal quality of the health data [4]. This study focuses on the health sector using sample data 2015-2018 from the Health Social Security Agencies (BPJS Kesehatan). This agency plays a crucial role in Indonesia, as the country boasts one of the largest single-player national insurance systems with population-based membership. Since its establishment in 2014, this insurance initiative has proven beneficial to an impressive 217 million patients. Consequently, by capturing valuable insights from the extensive electronic health records, the study holds the potential to contribute significantly to enhancing the quality healthcare process in Indonesia.[5] In this study, BPJS Kesehatan datasets were used to analyze and help to improve the efficiency of existing health procedures. In the sample data, there are four tables which are Membership, Primary Level Health Facilities (FKTP), Advanced Referral Health Facilities (FKRTL), and Non-Capitation Primary Level Health Facilities (Non-FKTP) [1]. The sample data is used to see what workflows are in the data, especially workflows related to patient diagnoses. However, this dataset must be traced first whether it has noises or not. In this study, a comparison is made of each treatment process based on the diagnosis group recorded in the dataset of BPJS Kesehatan. It is crucial to gain an understanding of the optimal processes and identify the ones that require further improvement is crucial. This knowledge empowers us to enhance these processes, allowing them to achieve values that are not only better than others but also potentially surpass the optimal values of those processes.

Process mining was used in our study to improve data analysis efficiency from information systems and reduce manual effort [6]. This study uses one of the algorithms in process mining, namely the Inductive Miner [2]. Inductive miner has been widely used in various studies. It basically detects cuts in a log and then loops through the sub-logs. This is done continuously until the base case is found. With the direct-follows variant, recursive paths can be avoided in sub logs. The Inductive Miner algorithm has the capability to handle invisible process tasks within a dataset. After carrying out process mining, process comparison is carried out where existing processes are compared with one another.

Our initial analysis identifies several challenges. Firstly, it is necessary to address how to process the BPJS Kesehatan sample data, particularly when dealing with noisy data. Additionally, the approach for conducting process discovery on the sample data needs to be determined. Furthermore, finding a suitable method to compare treatment processes for patients diagnosed with different diseases is essential. Lastly, analyzing the performance of processes using the sample data is another important aspect to consider.

In this research, process comparison is done by measuring process similarity with Graph Edit Distance (GED). We used GED considering its capacity to assess node weights in graphs, facilitating informed decision-making based on similarities and differences between graphs. Therefore, this research is expected to showcase the impact and benefits of recording process differences or similarities at the BPJS Kesehatan data. The conformance values of trace fitness, precision, generalization, and simplicity handling of patient treatments and improvement of the processes to be re-optimized.

The main objective of this research is to check the conformance of the process model to the records of event logs at Health Social Security Agencies. To achieve this objective, the research involves the following activities. Firstly, the Health Social Security Agencies sample data is preprocessed to prepare the event log data. Secondly, a process model is discovered using the inductive miner method, and a Petri net is used to visualize the sequence of the activities. Thirdly, GED is used to see similarities and differences between processes from different diagnostic groups. Lastly, the research measures four conformance metrics, namely trace fitness, precision, generalization, and simplicity. The expected outcome of this study is the analysis of the treatment processes for each diagnosis of the patient's disease and identify any variations that may exist between them. This research is carried out using the Python programming language assisted by a library, namely PM4PY to perform process discovery, conformance, and enhancement on the Health Social Security Agencies dataset.