

INTISARI

DETEKSI *SUBSEQUENCE OUTLIER* PADA DATA *TIME SERIES* BERBASIS *K-MEANS CLUSTERING* MENGGUNAKAN *EUCLIDEAN* *DISTANCE* DAN *MAHALANOBIS DISTANCE*

Oleh

LISA NURFAUZIYYA

14/364139/PA/15913

Perkembangan teknologi semakin cepat, berdampak pada bertambahnya jumlah data yang mengakibatkan ledakan data di berbagai bidang khususnya data *time series*. Hal tersebut tidak lepas dari alasan bahwa pada saat ini banyak proses seperti industri, stasiun pemantauan meteorologi atau pasar saham, menghasilkan data *time series* yang relevan secara terus menerus. Kondisi ini mendorong adanya usaha untuk mengolah serta menganalisis data menjadi sebuah informasi yang berguna. *Data mining* merupakan proses ekstraksi informasi tersembunyi dari kumpulan data yang besar untuk mendapatkan suatu informasi. Ada beberapa metode pada *data mining* seperti klasifikasi, *clustering*, prediksi, serta analisis *outlier*. Metode analisis *outlier* merupakan metode yang digunakan untuk mengetahui pola yang tidak sesuai dengan perilaku normal pada data. *Outlier* sendiri merupakan penyimpangan atau keanehan yang terjadi pada suatu data. Adanya *outlier* dapat menyebabkan perubahan hasil analisis data secara drastis. Oleh karena ini deteksi *outlier* bertujuan untuk mengidentifikasi suatu pola yang tidak sesuai. Sehingga perlu adanya keputusan yang harus dilakukan ketika *outlier* terdeteksi. *Clustering* merupakan salah satu metode yang dapat digunakan untuk melakukan deteksi *oulier* data *time series*.

Pada penelitian ini digunakan algoritma *K-Means* dengan metode *Euclidean Distance* dan *Mahalanobis Distance* pada proses pengukuran jarak. Dari hasil pengujian proses deteksi *outlier*, metode *Mahalanobis Distance* menghasilkan nilai akurasi dan sensitivitas yang lebih baik dibandingkan dengan *Euclidean Distance* ketika jumlah *cluster* semakin banyak dan ukuran *subsequence* semakin kecil. Dengan karakteristik *dataset time series* yaitu nilai varians yang dimiliki bervariasi, maka metode *Mahalanobis Distance* lebih tepat karena mempertimbangkan *variens* dari data.

Kata Kunci : *Time Series, Deteksi outlier, Clustering, K-Means, Mahalanobis Distance*

ABSTRACT

TIME SERIES SUBSEQUENCE OUTLIER DETECTION BASED ON K-MEANS CLUSTERING USING EUCLIDEAN DISTANCE AND MAHALANOBIS DISTANCE

By

LISA NURFAUZIYYA

14/364139/PA/15913

Technological developments increasingly affect the rapid increase in the amount of data that resulted in the explosion of data in various fields, especially time series data. This can not be separated from the reason that at present many processes such as industry, meteorological monitoring station or stock market, produces time series data that are relevant continuously. This condition encourages the effort to process and analyze the data into a useful information. Data mining is the process of extracting hidden information from large data sets to obtain information. There are several methods in data mining such as classification, clustering, prediction, and outlier analysis. Outlier analysis method is a method used to find patterns that are not in accordance with normal behavior on the data. Outlier itself is a deviation or peculiarities that occur in a data. The presence of an outlier can lead to drastic changes in the results of data analysis. Therefore this outlier detection aims to identify an unsuitable pattern. So the need for a decision to be made when the outlier is detected. Clustering is one method that can be used to detect outlier time series data.

In this research used K-Means algorithm with Euclidean Distance and Mahalanobis Distance method on distance measurement process. From the results of outlier detection test process, Mahalanobis Distance method yielded better accuracy and sensitivity value compared to Euclidean Distance when the number of cluster is more and the size of subsequence is smaller. With the characteristic of the time series dataset that is variant value owned, the Mahalanobis Distance method is more appropriate because considering the variance of the data.

Keywords: Time Series, Outlier Detection, Clustering, K-Means, Mahalanobis Distance