

ABSTRAK

Karyawan adalah orang yang bekerja di sebuah perusahaan atau lembaga dan menerima gaji. Karyawan merupakan aset paling penting yang harus dikelola dengan baik oleh perusahaan agar dapat memberikan kontribusi yang maksimal. Namun, banyak karyawan yang merasa tidak puas dengan hasil kontribusinya terhadap perusahaan, karena tidak mendapatkan balasan yang sesuai dengan harapan mereka. Oleh karena itu, banyak karyawan yang memutuskan untuk *Resign* dari perusahaan, dan perusahaan kesulitan dalam memprediksi karyawan yang akan *Resign*. Penelitian ini bertujuan untuk memprediksi faktor-faktor apa yang paling berpengaruh dalam menyebabkan karyawan *Resign* dari suatu perusahaan, dengan menerapkan metode K-means dengan dan tanpa *Principal Component Analysis*(PCA). Penelitian ini menggunakan dataset dari *Kaggle.com*, dengan total 14999 baris data dari 10 *attribute*. Dalam percobaan pertama, dataset direduksi menggunakan PCA sebelum dilakukan proses klustering dengan metode K-means. Sedangkan dalam percobaan kedua, dataset akan langsung dimasukkan ke dalam klustering dengan metode K-means tanpa PCA. Untuk melakukan evaluasi kluster (k) dalam metode K-means, penelitian ini menerapkan metode *sum of squared error* (SSE) dan metode *Silhouette Coefficient* untuk mengetahui kluster yang optimal. Penelitian ini bertujuan untuk mengevaluasi efektivitas K-means klustering dengan dan tanpa *Principal Component Analysis*(PCA) dalam mengidentifikasi faktor-faktor dominan yang mempengaruhi karyawan *Resign* dan menentukan jumlah kluster yang optimal dalam dataset. Penelitian ini menyimpulkan bahwa ada dua faktor yang paling dominan yang berkontribusi membuat karyawan *Resign* dari suatu perusahaan yaitu *last_evaluation* dan *average_monthly_hours*. Evaluasi SSE menunjukkan bahwa kedua metode memiliki titik siku (*koefisien*) pada 3 kluster, menunjukkan bahwa pembagian data menjadi lebih dari 3 kluster tidak memberikan informasi tambahan yang signifikan. Evaluasi *Silhouette Coefficient* menunjukkan bahwa K-mean tanpa PCA memperoleh nilai *Silhouette Coefficient* terbaik sebesar 0.5674 sedangkan K-means dengan PCA memperoleh nilai *Silhouette Coefficient* sebesar 0.5491. Meskipun K-means dengan PCA memiliki keuntungan mengurangi dimensi dataset, tetapi juga memiliki waktu eksekusi yang lebih lama dibandingkan dengan K-means tanpa PCA yang dimana untuk waktu eksekusi K-means dengan PCA yaitu 181.53 detik sedangkan waktu eksekusi K-means tanpa PCA yaitu 95.84 detik. Oleh karena itu, disarankan untuk menggunakan K-means tanpa PCA untuk mengelompokkan dataset ini.

Kata kunci : Karyawan, *Resign*, K-means dan *Principal Component Analysis* (PCA), *sum of squared error* (SSE), *Silhouette Coefficient*.

ABSTRACT

Employees are people who work in a company or institution and receive a salary. Employees are the most important assets that must be managed properly by the company in order to provide maximum contribution. However, many employees are dissatisfied with the results of their contribution to the company, because they do not get the response they expected. Therefore, many employees decide to resign from the company, and the company has difficulty predicting employees who will resign. This study aims to predict what factors are most influential in causing employees to leave a company, by applying the K-means method with and without Principal Component Analysis (PCA). This study uses a dataset from Kaggle.com, with a total of 14999 rows of data from 10 attributes. In the first experiment, the dataset was reduced using PCA before the clustering process was carried out using the K-means method. Whereas in the second experiment, the dataset will be directly entered into clustering using the K-means method without PCA. To evaluate clusters (k) in the K-means method, this study applies the sum of squared error (SSE) method and the Silhouette Coefficient method to determine the optimal cluster. This study aims to evaluate the effectiveness of K-means clustering with and without Principal Component Analysis (PCA) in identifying the dominant factors affecting Resigning employees and determining the optimal number of clusters in the dataset. This study concludes that there are two most dominant factors that contribute to making employees leave a company, namely last_evaluation and average_monthly_hours. The SSE evaluation shows that both methods have an angle (coefficient) in 3 clusters, indicating that dividing the data into more than 3 clusters does not provide significant additional information. The Silhouette Coefficient evaluation showed that K-means without PCA obtained the best Silhouette Coefficient value of 0.5674 while K-means with PCA obtained a Silhouette Coefficient value of 0.5491. Although K-means with PCA has the advantage of reducing dataset dimensions, it also has a longer execution time compared to K-means without PCA where the execution time for K-means with PCA is 181.53 seconds while the execution time for K-means without PCA is 95.84 second. Therefore, it is suggested to use K-means without PCA to classify this dataset.

Keywords : Employee, Resign, K-means and Principal Component Analysis (PCA), sum of squared error (SSE), Silhouette Coefficient.