ABSTRAK

Employee Turnover merupakan permasalahan penting bagi perusahaan karena berdampak pada peningkatan biaya rekrutmen dan pelatihan serta penurunan produktivitas. Penelitian ini bertujuan untuk membandingkan kinerja algoritma *Gaussian Naïve Bayes* dan *K-Nearest Neighbor* (KNN) dalam klasifikasi dataset *employee turnover*, serta menganalisis pengaruh teknik normalisasi dan standarisasi terhadap performa model. Dataset yang digunakan berasal dari *Kaggle* dengan total 9.540 data karyawan periode 2016-2020. Tahapan penelitian meliputi *preprocessing* data, penyeimbangan kelas menggunakan *Synthetic Minority Over-Sampling Technique* (SMOTE), ektraksi fitur dengan *Principal Component Analysis* (PCA), serta optimasi parameter menggunakan *GridSearchCV*. Evaluasi dilakukan menggunakan *Confusion Matrix* dengan metrik akurasi, *precision*, *recall*, dan *F1-Score*. Hasil penelitian menunjukkan bahwa *K-Nearest Neighbor* (KNN) memiliki performa lebih baik dibandingkan *Gaussian Naïve Bayes*. Pada skema terbaik dengan standarisasi menggunakan *StandardScaler*, KNN mencapai akurasi sebesar 85,46%, *precision* 81%, *recall* 92%, dan *F1-Score* 86%. Sementara itu, *Gaussian Naïve Bayes* hanya mencapai akurasi sebesar 67,18%, *precision* 63%, *recall* 81%, dan *F1-Score 71%*. Penggunaan *StandardScaler* terbukti lebih optimal dibandingkan *Min-Max Scaler* pada kedua algoritma, pada KNN akurasi meningkat dari 80,47% menjadi 85,46%, sedangkan pada *Gaussian Naïve Bayes* meningkat dari 55,86% menjadi 67,18%.

Kata kunci : *employee turnover*, klasifikasi, *gaussian naïve bayes*, *k-nearest neighbor*, *standardscaler*

ABSTRACT

Employee Turnover is an important issue for companies because it leads to increased recruitment and training costs and decreased productivity. This study aims to compare the performance of the Gaussian Naïve Bayes and K-Nearest Neighbor (KNN) algorithms in classifying employee turnover datasets, as well as to analyze the effect of normalization and standardization techniques on model performance. The dataset used is from Kaggle with a total of 9.540 employee data for the 2016-2020 period. The research stages include data preprocessing, class balancing using the Synthetic Minority Over-Sampling Technique (SMOTE), feature extraction with Principal Component Analysis (PCA), and parameter optimization using GridSearchCV. Model performance evaluation was carried out using a Confusion Matrix with accuracy, precision, recall, and F1-Score metrics. The results show that the K-Nearest Neighbor (KNN) algorithm performs better than Gaussian Naïve Bayes. In the best scheme with standardization using StandardScaler, KNN achieved an accuracy of 85,46%, precision of 81%, recall of 92%, and an F1-Score of 86%. Meanwhile, Gaussian Naïve Bayes only achieved an accuracy of 67,18%, precision of 63%, recall of 81%, and an F1-Score of 71%. The use of StandardScaler proved to be more optimal than Min-Max Scaler in both algorithms. In KNN, accuracy increased from 80,47% to 85,46%, while in Gaussian Naïve Bayes, it increased from 55,86% to 67,18%.

Keywords : *employee turnover, classification, gaussian naïve bayes, k-nearest neighbor, standardscaler.*