



Klasifikasi Diabetes Mellitus (DM) Menggunakan Metode Naïve Bayes Dengan Seleksi Variabel Chi-Square

Farhan Arizal Ginanjar¹, Ambar Winarni², Nur'aini Muhassanah³

¹²³Universitas Nahdlatul Ulama Purwokerto, Jalan Sultan Agung No. 42, Purwokerto, Indonesia

*corresponding author: farhanarizalginanjar@gmail.com

Article Info

Received

Revised

Accepted

Keywords: Diabetes Mellitus;
Classification; Naïve Bayes;
Chi-Square



Abstract

Diabetes mellitus (DM) is a chronic disease that can cause serious complications, making early detection essential. Technological advances enable the use of data mining techniques, particularly the Naïve Bayes classification method, to support early diabetes detection. Although Chi-Square variable selection is known to improve Naïve Bayes accuracy, studies examining the impact of different significance levels remain limited. Therefore, this study applies the Naïve Bayes method with and without Chi-Square variable selection at three significance levels ($\alpha = 0.05$, $\alpha = 0.01$, and $\alpha = 0.001$) to evaluate their effects on classification performance and identify the optimal significance level. The results show that Naïve Bayes without variable selection achieved an accuracy of 87.50%, precision of 93.01%, and recall of 86.21%. After applying Chi-Square selection, performance improved across all significance levels. At $\alpha = 0.05$, the accuracy reached 87.88%, with precision of 93.06% and recall of 86.85%. At $\alpha = 0.01$, accuracy increased to 88.46%, precision to 94.25%, and recall to 86.53%. The best performance was obtained at $\alpha = 0.001$, achieving an accuracy of 88.65%, precision of 94.19%, and recall of 86.86%. These findings indicate that Chi-Square variable selection effectively enhances the performance of the Naïve Bayes algorithm for diabetes classification.