GENERATION JOURNAL
Vol. 10 No. 2 Tahun 2026
e-ISSN: 2549-2233 / p-ISSN: 2580-4952
doi

SIMPULAN

Berdasarkan hasil penelitian yang dilakukan tentang klasifikasi penyakit diabetes mellitus menggunakan metode Naïve Bayes dengan penerapan seleksi variabel Chi-Square, diperoleh beberapa simpulan sebagai berikut:

1. Metode Naïve Bayes tanpa seleksi variabel menunjukkan kemampuan klasifikasi yang baik dengan nilai *accuracy* 87,50%, *precision* 93,01%, dan *recall* 86,21%. Hasil tersebut mengidentifikasi bahwa metode Naïve Bayes mampu mengenali pola data gejala diabetes dengan ketepatan prediksi 87,50%.
2. Penerapan seleksi variabel Chi-Square terbukti memberikan pengaruh positif terhadap peningkatan kinerja model. Pada $\alpha = 0{,}05$ *accuracy* model meningkat menjadi 87,88%, pada $\alpha = 0{,}01$, *accuracy* model meningkat menjadi 88,46%, dan pada $\alpha = 0{,}001$ mencapai 88,65%. Dengan demikian terjadi peningkatan *accuracy* sebesar 0,38% hingga 1,15% dibandingkan model tanpa seleksi variabel. Hal ini menunjukkan bahwa proses seleksi variabel membantu model fokus pada variabel yang paling relevan sehingga menghasilkan klasifikasi yang lebih tepat.
3. Temuan ini menunjukkan bahwa proses pemilihan variabel yang lebih ketat dan penggunaan nilai signifikansi yang lebih kecil meningkatkan kemampuan model untuk mengenali pola data yang signifikan. Secara keseluruhan, kombinasi klasifikasi Naïve Bayes dengan seleksi variabel Chi-Square terbukti efektif dalam mendeteksi dini penyakit diabetes karena mampu memberikan hasil yang lebih akurat.

SARAN

Beberapa saran dan rekomendasi guna pengembangan penelitian selanjutnya terkait penerapan metode klasifikasi Naïve Bayes dengan seleksi variabel Chi-Square sebagai berikut:

1. Penelitian selanjutnya direkomendasikan untuk mengkombinasikan metode Chi-Square dengan teknik seleksi variabel lainnya, seperti *Mutual Information, Forward Selection, Backward Elimination* dan *Recursive Feature Elimination (RFE)*.
2. Metode seleksi variabel yang telah dikembangkan disarankan diterapkan pada metode klasifikasi lainnya, seperti *Artificial Neural Network (ANN), Random Forest, K-Nearest Neighbor (KNN), Decision Tree (C4.5, C5.0),* dan *Support Vector Machine (SVM)*. Penggunaan ini bertujuan untuk memperluas cakupan penelitian sekaligus melakukan perbandingan terhadap hasil performa setiap algoritma, sehingga dapat diperoleh model klasifikasi yang paling optimal.
3. Penelitian selanjutnya direkomendasikan untuk menambahkan uji McNemar dan *paired t-test* dalam menguji signifikansi perbedaan performa model, serta melakukan pengujian asumsi independensi antar variabel untuk memastikan kesesuaian dengan metode Naïve Bayes, sehingga hasil yang diperoleh tidak hanya bersifat numerik, tetapi juga signifikan secara statistik.