

KOMPARASI EFEKTIVITAS ALGORITMA K-NEAREST NEIGHBORS (KNN) DAN NAIVE BAYES DALAM IDENTIFIKASI PASIEN DIABETES

Dibawah bimbingan Bapak Eko Budi Susanto, M.Kom dan Bapak Bambang Ismanto, M.Kom.

95 halaman + xcv halaman / 43 gambar / 22 tabel / 28 pustaka (2018-2024)

ABSTRAK

Diabetes merupakan penyakit metabolik kronis yang berpotensi menimbulkan komplikasi serius pada berbagai organ tubuh. Keterlambatan diagnosis penyakit diabetes menjadi salah satu penyebab meningkatnya angka kematian akibat komplikasi sebelum diagnosis ditegakkan. Dalam bidang medis, penerapan model machine learning telah membuka peluang signifikan dalam meningkatkan akurasi diagnosis dini diabetes. Penelitian ini bertujuan untuk membandingkan kinerja algoritma K-Nearest Neighbors (KNN) dan Naive Bayes Classifier (NBC) dengan menggunakan dataset sekunder berjumlah 128 record data dan memuat 8 variabel data yang relevan untuk prediksi diabetes. Proses penelitian menggunakan kerangka CRISP-DM, yang melibatkan tahapan pemahaman bisnis, pemahaman data, persiapan data, pemodelan, evaluasi, dan deployment. Hasil analisis menunjukkan bahwa algoritma KNN dengan iterasi $K=21$, berdasarkan evaluasi confusion matrix mencapai akurasi sebesar 76,92%, recall 100%, precision 77% dan F1-Score 87%. Sementara itu algoritma naïve bayes memiliki akurasi 65,38%, recall 55%, precision 100% dan F1-Score 71%. Pada evaluasi dengan metode k-fold cross validation menggunakan $K=10$ menghasilkan rata-rata akurasi sebesar 73% untuk algoritma KNN dan 70% untuk algoritma naïve bayes. Dengan demikian, algoritma KNN lebih unggul dan direkomendasikan untuk klasifikasi penyakit diabetes. Rekomendasi untuk penelitian selanjutnya adalah menguji algoritma klasifikasi lain guna mendapatkan hasil perbandingan yang lebih komprehensif, serta mengintegrasikan hasil penelitian ini ke dalam aplikasi yang dapat digunakan secara praktis.

Kata kunci: *Komparasi, Algoritma, K-Nearest Neighborn, Naïve Bayes, Confusion Matrix, K-Fold Cross Validation, Diabetes*

COMPARISON OF THE EFFECTIVENESS OF K-NEAREST NEIGHBORS (KNN) AND NAIVE BAYES ALGORITHMS IN IDENTIFYING DIABETES PATIENTS

Supervised by Mr. Eko Budi Susanto, M.Kom and Mr. Bambang Ismanto, M.Kom.

95 pages + xcv pages / 43 image / 22 tables / 28 references (2018-2024)

ABSTRACT

Diabetes is a chronic metabolic disease that poses significant risks of severe complications in various organs. Delayed diagnosis is one of the key factors contributing to increased mortality rates due to complications before proper medical intervention. In the medical field, the application of machine learning models offers substantial potential for improving the accuracy of early diabetes diagnosis. This study aims to compare the performance of the K-Nearest Neighbors (KNN) and Naive Bayes Classifier (NBC) algorithms using a secondary dataset comprising 128 records with 8 relevant predictive variables for diabetes diagnosis. The research follows the CRISP-DM framework, encompassing the stages of business understanding, data understanding, data preparation, modeling, evaluation, and deployment. The findings reveal that the KNN algorithm, with 21 neighbors ($K=21$), achieves an accuracy of 76,92%, recall of 100%, precision of 77% and an F1-score of 87% based on confusion matrix evaluation. In contrast, the Naive Bayes algorithm achieves an accuracy of 65.38%, recall of 55%, precision of 100%, and an F1-score of 71%. Further evaluation using 10-fold cross-validation indicates an average accuracy of 73% for the KNN algorithm and 70% for the Naive Bayes algorithm. Therefore, the KNN algorithm demonstrates superior performance and is recommended for diabetes classification tasks. Recommendations for further research are to test other classification algorithms to obtain more comprehensive comparison results, as well as to integrate the results of this research into applications that can be used practically.

Keywords: Comparison, Algorithm, K-Nearest Neighbors, Naive Bayes, Confusion Matrix, K-Fold Cross Validation, Diabetes