

ABSTRAK

Hasanah, Shafira Maudatul. 2024. Penggunaan Algoritma *K-nearest Neighbor* dengan Strategi *Cross Validation* dalam Menentukan Penyakit *Diabetes* pada Wanita. Tugas Akhir. Program Sarjana. Program Studi Teknik Informatika. Universitas Muhammadiyah Jember.

Pembimbing: (1) Hardian Oktavianto ; (2) Qurrota A'yun

Diabetes adalah suatu penyakit yang diakibatkan terlalu banyak jumlah kadar gula dalam darah dikarenakan produksi dan pelepasan insulin yang tidak normal. Penyakit tersebut tergolong penyakit yang menyebabkan komplikasi dan kematian. Seorang wanita dapat dikatakan memiliki penyakit tersebut jika hasil pengujian lab menunjukkan bahwa wanita tersebut positif diabetes. Beberapa Algoritma dapat digunakan untuk perhitungan proses *klasifikasi*. Dalam penelitian ini peneliti akan memakai metode *K-Nearest Neighbor (KNN)*. Tujuan dalam melakukan penelitian ini adalah ingin mengetahui nilai akurasi untuk mengetahui tingkat ketepatan suatu metode dalam mendiagnosis, presisi untuk menghasilkan pengujian berulang data positif, dan recall untuk menghasilkan pengujian berulang data negatif. Data yang akan digunakan pada penelitian ini yakni sebanyak 768 data dengan 8 atribut yang terdiri dari kehamilan, glukosa, tekanan darah, ketebalan kulit, insulin, bmi, keturunan diabetes, dan umur. Nilai K yang akan digunakan dalam penelitian ini yaitu 5, 7, 11, dan 13. Sedangkan untuk nilai *KFold Cross Validation* yang akan digunakan yaitu 4, 6, 8, 12. Dari hasil klasifikasi yang telah dilakukan oleh peneliti mendapatkan hasil yang akurat yaitu berada di nilai K=11 dengan Kfold 12 yakni akurasi sebesar 74,37%, Presisi sebesar 72,01% dan recall sebesar 69,91%.

Kata Kunci

Diabetes, K-Nearest Neighbor, Klasifikasi, Cross-Validation

ABSTRACT

Hasanah, Shafira Maudatul. 2024. Penggunaan Algoritma *K-nearest Neighbor* dengan Strategi *Cross Validation* dalam Menentukan Penyakit *Diabetes* pada Wanita. Tugas Akhir. Program Sarjana. Program Studi Teknik Informatika. Universitas Muhammadiyah Jember.

Mentor: (1) Hardian Oktavianto ; (2) Qurrota A'yun

Diabetes is a disease caused by too much sugar in the blood due to abnormal insulin production and release. The disease is classified as a disease that causes complications and death. A woman can be said to have the disease if the results of the lab test show that the woman is positive for diabetes. Several algorithms can be used to calculate the classification process. In this study, the researcher will use the K-Nearest Neighbor (KNN) method. The purpose of this research is to find out the accuracy value to determine the level of accuracy of a method in diagnosing, precision to produce repeated testing of positive data, and recall to produce repeated testing of negative data. The data that will be used in this study are 768 data with 8 attributes consisting of pregnancy, glucose, blood pressure, skin thickness, insulin, BMI, diabetes heredity, and age. The K values that will be used in this study are 5, 7, 11, and 13. Meanwhile, the KFold Cross Validation values that will be used are 4, 6, 8, 12. From the classification results that have been carried out by the researcher, accurate results were obtained, namely at a value of $K = 11$ with Kfold 12, namely an accuracy of 74.37%, Precision of 72.01% and recall of 69.91%.

Keywords

Diabetes, K-Nearest Neighbor, Classification, Cross-validation