

ABSTRAK

Diabetes Mellitus (DM) mengalami peningkatan setiap tahun, bahkan pada tahun 2045 penderita DM diprediksi mencapai 700 juta jiwa didunia (*International of Diabetic Federation*, 2019). Di Indonesia, prevalensi DM menurut Konsensus Perkeni 2015 bahwa pada tahun 2018 penderita DM terus bertambah hingga meningkat 10,9% (Riskesdas, 2018). Banyak masyarakat yang belum menyadari bahwa dirinya mengidap penyakit DM yang disebabkan banyak faktor dan salah satunya adalah kurangnya pengetahuan tentang penyakit tersebut. (Sholeh et al, 2022). DM perlu diidentifikasi sejak dini karena dapat memiliki konsekuensi yang fatal yaitu kematian. Penting bagi seseorang untuk mengetahui faktor-faktor penyakit DM agar dapat melakukan pencegahan atau mengurangi resiko komplikasi yang lebih serius. (Karo Karo dan Hendriyana, 2022).

Penelitian ini mengimplementasikan metode normalisasi data *Zscore*, *MinMax* dan *MaxAbs* terhadap metode klasifikasi *K-Nearest Neighbor* untuk memprediksi penyakit Diabetes Mellitus. *Dataset* penelitian ini adalah *dataset* diabetes yang didapat dari *kaggle* pada *website* <https://www.kaggle.com/datasets/mathchi/diabetes-data-set> yang berjumlah 768 data dengan 8 atribut serta satu atribut label yang menunjukkan seseorang mengidap penyakit DM atau tidak. *Dataset* tersebut kemudian dilakukan proses *cleaning data* untuk mendapatkan data yang lebih baik, selanjutnya dinormalisasi dan dibagi dengan perbandingan 90:10 untuk data *training* dan data *testing* terlebih dahulu sebelum dilakukan proses klasifikasi dengan KNN.

Hasil penelitian ini menggunakan pengujian *confusion matrix* pada setiap model normalisasi *Zscore*, *MinMax* dan *MaxAbs* serta tanpa normalisasi dengan klasifikasi KNN pada nilai K=1 sampai K=11. Hasil pengujian yang telah dilakukan menunjukkan bahwa normalisasi data dapat memberikan performa lebih baik dibandingkan tanpa normalisasi. Hasil pengujian pada data tanpa normalisasi mendapat performa terbaik pada K=7 dengan akurasi 76,47%, normalisasi *Zscore* pada K=3, 7 dan 11 mendapat akurasi 79,41%, dan normalisasi *MaxAbs* pada K=5, 9 dan 11 mendapat akurasi 82,35%, sedangkan pada normalisasi *MinMax* mendapat performa terbaik dengan akurasi 85,29% pada nilai K=5.

Kata kunci : Diabetes Mellitus, Prediksi, Normalisasi, *Zscore*, *MinMax*, *MaxAbs*, *K-Nearest Neighbor*

ABSTRACT

Diabetes Mellitus (DM) is experiencing an annual increase, and it is predicted that by the year 2045, the number of DM patients worldwide will reach 700 million (International of Diabetic Federation, 2019). In Indonesia, the prevalence of DM according to the Konsensus Perkeni 2015 shows that in 2018, the number of DM patients continued to increase to 10.9% (Riskseddas, 2018). Many people are unaware that they have DM, which can be caused by various factors, one of which is a lack of knowledge about the disease (Sholeh et al., 2022). Early identification of DM is crucial as it can have fatal consequences, including death. It is important for individuals to be aware of the risk factors for DM to prevent or reduce the risk of more serious complications (Karo Karo and Hendriyana, 2022).

This research implements the Zscore, MinMax, and MaxAbs data normalization methods on the K-Nearest Neighbor classification method to predict Diabetes Mellitus. The research dataset is a diabetes dataset obtained from Kaggle website <https://www.kaggle.com/datasets/mathchi/diabetes-data-set> consist of 768 data samples with 8 attributes and one label attribute indicating whether a person has DM or not. The dataset then undergoes data cleaning process to obtain better data, subsequently normalized, and divided with a 90:10 ratio for training and testing data before the classification process using KNN.

The results of this research use confusion matrix testing on each normalization model: Zscore, MinMax, and MaxAbs, as well as without normalization with KNN classification for values of K=1 to K=11. The conducted testing shows that data normalization can provide better performance compared to without normalization. The best performance for data without normalization is achieved at K=7 with an accuracy of 76.47%, Zscore normalization at K=3, 7, and 11 achieved an accuracy of 79.41%, and MaxAbs normalization at K=5, 9, and 11 achieved an accuracy of 82.35%, while MinMax normalization yielded the best performance with an accuracy of 85.29% at K=5.

Keywords: Diabetes Mellitus, Prediction, Normalization, Zscore, MinMax, MaxAbs, K-Nearest Neighbor.