

ABSTRACT

Unhealthy people's lifestyles such as consuming fast food and carbonated drinks in excess and not followed by exercise can lead to obesity and lead to diabetes. Diabetes or what is often referred to as diabetes, is a metabolic disorder caused by a high level of sugar in the blood that cannot be processed properly by the body. Classic symptoms that usually occur in general are the desire to urinate too often, thirst that is difficult to stop, and a high rhythm of hunger. Support Vector Machine (SVM) algorithm is one of the methods in Machine Learning which is included in Supervised Learning. SVM itself has the advantage of being able to generalize high data and SVM can produce better accuracy results even though it only uses a small amount of data. In this study, data normalization will be used to process raw datasets and then split the data to simplify the classification process. In the classification process itself uses the SVM algorithm. To see the final results of this study, an evaluation was carried out using the Confusion Matrix method. To find out how well the SVM accuracy rate is carried out various test scenarios during the classification process on the diabetes dataset. Tests were carried out with several scenarios, in the scenario of the effect of using preprocessing the best average accuracy value which was given was 82.50% by using the Z-Score preprocessing, in the scenario of the effect of using the data sharing method the best average level of accuracy was 87.94 % by using the Train test split data, and for the scenario of the effect of using the kernel the best average accuracy value is 80.49% on the use of the Linear kernel.

Keywords: Diabetes, Support Vector Machine, Machine Learning, Supervised Learning, Normalization, Confusion Matrix, Preprocessing, Kernel