

ABSTRAK

*International Diabetes Federation (IDF) Atlas edisi kesepuluh menyatakan jumlah penderita diabetes di dunia mencapai 537 juta jiwa, dengan total kematian hingga tahun 2021 sekitar 6,7 juta jiwa dalam rentang usia 20-79 tahun. Sedangkan di Indonesia, jumlah penderita diabetes mencapai 19,47 juta jiwa dalam rentang usia yang sama. Diperkirakan bahwa pada tahun 2030, jumlah penderita diabetes akan meningkat menjadi 643 juta dan kemungkinan akan mencapai 783 juta pada tahun 2045. Hal ini disebabkan karena faktor keturunan, perubahan gaya hidup, kurangnya berolahraga dan faktor usia, sehingga penderita diabetes akan terus meningkat setiap tahun. Maka penelitian ini bertujuan untuk menerapkan Algoritma *Logistic Regression* dengan SMOTE dalam membangun model klasifikasi penyakit diabetes tipe 2. Kemudian model tersebut diimplementasikan kedalam sebuah sistem yang dapat mendiagnosis apakah seorang pasien terkena penyakit diabetes atau tidak. Penelitian ini menggunakan metode CRISP-DM. Metode tersebut terbagi dalam beberapa proses dimulai dari *business understanding*, *data understanding*, *data preparation*, *modeling*, *evaluasi* dan *deployment*. Berdasarkan hasil penelitian rata-rata akurasi dari model tersebut sebesar 86%, presisi 85%, dan recall 88%. Sedangkan hasil dari data uji validasi sistem dan pakar sesuai yang diharapkan. Sehingga pada proses perhitungan manual dengan rumus *confusion matrix* menghasilkan nilai akurasi, presisi dan *recall* sebesar 100%.*

Kata Kunci: CRISP-DM, Diabetes, *Logistic Regression*, SMOTE

KARAWANG

ABSTRACT

The tenth edition of the International Diabetes Federation (IDF) Atlas states that the number of people with diabetes in the world reaches 537 million, with a total death toll until 2021 of around 6.7 million people in the age range of 20-79 years. Meanwhile, in Indonesia, the number of people with diabetes reaches 19.47 million in the same age range. It is estimated that by 2030, the number of people with diabetes will increase to 643 million and will likely reach 783 million by 2045. This is due to heredity, lifestyle changes, lack of exercise and age factors, so that diabetics will continue to increase every year. So this research aims to apply the Logistic Regression Algorithm with SMOTE in building a classification model for type 2 diabetes. Then the model is implemented into a system that can diagnose whether a patient has diabetes or not. This research uses the CRISP-DM method. The method is divided into several processes starting from business understanding, data understanding, data preparation, modeling, evaluation and deployment. Based on the research results, the average accuracy of the model is 86%, precision is 85%, and recall is 88%. While the results of the system and expert validation test data are as expected. So that the manual calculation process with the confusion matrix formula produces an accuracy, precision and recall value of 100%..

Keyword: CRISP-DM, Diabetes, Logistic Regression, SMOTE

