Klasifikasi Carrier Talasemia Alfa Berdasarkan Hasil Complete Blood Count (CBC) dengan Pendekatan Algoritma Pembelajaran Mesin = Classification of Alpha Thalassemia Carriers Based on Complete Blood Count (CBC) Results Using Machine Learning Algorithms

Viar Ghina Qatrunnada, author

Abstrak

Talasemia merupakan penyakit autosomal resesif yang menyebabkan tubuh tidak mampu memproduksi hemoglobin (Hb) secara normal, sehingga penderitanya membutuhkan transfusi darah seumur hidup. Skrining genetik bagi pasangan yang akan menikah merupakan langkah awal untuk menekan angka bayi lahir dengan gen talasemia. Namun, perhatian masyarakat masih rendah karena skrining ini tidak termasuk ke dalam prosedur pra-nikah yang dapat ditanggung oleh Jaminan Kesehatan Nasional (JKN), serta harganya cukup mahal. Penelitian ini memanfaatkan machine learning untuk memprediksi carrier dan mengklasifikasikan jenis talasemia berdasarkan hasil tes hematologi lengkap/Complete Blood Count (CBC) yang memiliki harga lebih terjangkau dari skrining genetik. Pada penelitian, digunakan beberapa algoritma pembelajaran mesin bersifat supervised classification seperti Logistic Regression, Random Forest, Support Vector Machine, Gradient Boosting, XGBoost, dan AdaBoost. Hasil menunjukkan penggunaan Support Vector Machine dengan oversampling menggunakan synthetic minority oversampling technique edited nearest neighbors (SMOTE-ENN), normalisasi dengan RobustScaler, hyperparameter tuning, dan 10-fold cross-validation berhasil mencapai nilai akurasi 98.84% dalam mengklasifikasikan carrier talasemia alfa berdasarkan hasil CBC.

.....Thalassemia is an autosomal recessive disease that unable the body to produce hemoglobin (Hb) normally, requiring lifelong blood transfusions. Genetic screening for future married couples is the first step to reduce the number of babies born with the thalassemia gene. However, public attention is still low because the screening is not included in the pre-marital procedures that can be covered by the Jaminan Kesehatan Nasional (JKN), despite the price is quite expensive. This study utilizes machine learning to predict the carrier and classify the type of alpha-thalassemia based on the results of the Complete Blood Count (CBC) test, which is more affordable than genetic screening. In the study, several supervised classification machine learning algorithms were utilized such as Logistic Regression, Random Forest, Support Vector Machine, Gradient Boosting, XGBoost, and AdaBoost. The results show the use of Support Vector Machine with oversampling with synthetic minority oversampling technique edited nearest neighbors (SMOTE-ENN), normalization with RobustScaler, hyperparameter tuning, and 10-fold cross-validation successfully achieved 98.84% accuracy in classifying alpha thalassemia carriers based on CBC results.