

## ABSTRAK

Penyakit genetik adalah penyakit yang disebabkan adanya kelainan dalam susunan gen pada seseorang. Adapun salah satu penyakit genetik adalah thalassemia. Thalassemia merupakan penyakit kelainan darah merah yang diturunkan dari kedua orang tua kepada anak dan keturunannya. Penyakit ini disebabkan karena berkurang atau tidak terbentuknya protein pembentuk hemoglobin utama manusia, hal ini dapat menyebabkan sel darah merah (eritrosit) yang mudah rusak dan pasien menjadi pucat karena kekurangan darah. Sampai saat ini, thalassemia belum dapat disembuhkan namun dapat diketahui sifat pembawanya dengan melakukan pemeriksaan skrining. Dengan adanya penderita penyakit thalassemia yang banyak, maka perlu tindakan awal sebagai solusi dari permasalahan penyakit thalassemia yaitu dengan melakukan prediksi untuk mendeteksi sedari dini penyakit thalassemia. Terdapat beberapa metode data mining yang sudah diterapkan dalam memprediksi penyakit untuk mendapatkan tingkat akurasi yang tinggi. Salah satu dari algoritma di dalam data mining yang dapat digunakan untuk melakukan prediksi adalah algoritma *Naïve Bayes*. Masalah pada metode *Naïve Bayes* perlu diperbaiki agar kinerja dan performa algoritma menjadi lebih baik dan akurasi dapat meningkat. Masalah ini dapat dihilangkan dengan cara melakukan optimasi fitur menggunakan algoritma *Particle Swarm Optimization (PSO)*. *PSO (Particle Swarm Optimization)* merupakan salah satu metode optimasi terbaik dalam melakukan klasifikasi. Penerapan *Particle Swarm Optimization (PSO)* untuk mengetahui seberapa akurat *feature selection Particle Swarm Optimization (PSO)* dalam meningkatkan akurasi algoritma *Naïve Bayes* untuk memprediksi diagnosis penyakit thalassemia.

Dalam penelitian ini, pengujian dilakukan dengan *confusion matrix* untuk menghitung nilai akurasi pada model algoritma *Naïve Bayes* dan model algoritma *Naïve Bayes* dan *PSO* untuk prediksi penyakit thalassemia. Pada model algoritma *Naïve Bayes* dan *PSO*, setelah dilakukan seleksi fitur didapatkan jumlah fitur yang relevan sebanyak 5 fitur dari total fitur yaitu 15 fitur. Pengujian dengan *confusion matrix* pada model algoritma *Naïve Bayes* mendapatkan akurasi sebesar 63% sedangkan akurasi dari algoritma *Naïve Bayes* dan *PSO* yaitu sebesar 81%. Dari hasil pengujian tersebut terjadi peningkatan nilai akurasi sebesar 18% setelah diterapkan seleksi fitur, hal ini menunjukkan bahwa seleksi fitur menggunakan *PSO* dapat meningkatkan performa model algoritma *Naïve Bayes* untuk prediksi diagnosis penyakit thalassemia.

**Kata Kunci : Thalassemia, Prediksi, Seleksi Fitur, *Naïve Bayes*, PSO.**

## ABSTRACT

Genetic diseases are diseases caused by abnormalities in a person's gene structure. One of the genetic diseases is thalassemia. Thalassemia is a red blood disorder that is passed from both parents to their children and offspring. This disease is caused by reduced or non-formation of the main human hemoglobin-forming protein, this can cause red blood cells (erythrocytes) to be easily damaged and the patient to become pale due to lack of blood. Until now, thalassemia cannot be cured, but the nature of the carrier can be identified by carrying out a screening examination. With so many people suffering from thalassemia, it is necessary to take initial action as a solution to the problem of thalassemia, namely by making predictions to detect thalassemia early. There are several data mining methods that have been applied to predict disease to obtain a high level of accuracy. One of the algorithms in data mining that can be used to make predictions is the Naïve Bayes algorithm. Problems with the Naïve Bayes method need to be fixed so that the performance and performance of the algorithm becomes better and accuracy can increase. This problem can be eliminated by optimizing features using the Particle Swam Optimization (PSO) algorithm. PSO (Particle Swarm Optimization) is one of the best optimization methods for classification. Application of Particle Swarm Optimization (PSO) to find out how accurate Particle Swarm Optimization (PSO) feature selection is in increasing the accuracy of the Naïve Bayes algorithm for predicting the diagnosis of thalassemia.

In this research, testing was carried out with a confusion matrix to calculate the accuracy value of the Naïve Bayes algorithm model and the Naïve Bayes and PSO algorithm models for predicting thalassemia disease. In the Naïve Bayes and PSO algorithm models, after feature selection, the number of relevant features is 5 out of a total of 15 features. Testing with the confusion matrix on the Naïve Bayes algorithm model obtained an accuracy of 63%, while the accuracy of the Naïve Bayes and PSO algorithms was 81%. From the test results, there was an increase in the accuracy value of 18% after feature selection was applied, this shows that feature selection using PSO can improve the performance of the Naïve Bayes algorithm model for predicting the diagnosis of thalassemia.

**Kata Kunci : Thalassemia, Prediction, Feature Selection, Naïve Bayes, PSO.**