

PERBANDINGAN METODE NAIVE BAYES DAN K-NEAREST NEIGHBOR PADA KLASIFIKASI MORFOLOGI GEN SEL DARAH PUTIH

Muhammad Nur'adli Hasbi Gumay

ABSTRAK

Dibidang kesehatan, mendiagnosis penyakit leukemia merupakan hal yang sulit karena masih didiagnosis secara manual dengan bantuan dokter. Diagnosis manual tersebut dapat mengalami kesalahan yang disebabkan oleh kelalaian manusia. Dari permasalahan tersebut, maka dibutuhkan diagnosis jenis penyakit leukemia menggunakan kecanggihan teknologi yaitu *machine learning* untuk mengatasi permasalahan tersebut. Dalam penelitian ini, *machine learning* tersebut mengolah data yang berasal dari jenis leukemia yaitu *Acute Myeloid Leukemia* (AML) dan *Acute Lymphoblastic Leukemia* (ALL) berdasarkan ciri morfologi gen sel darah putih tersebut. Metode pengklasifikasian data yang digunakan untuk penelitian ini yaitu K-Nearest Neighbor (K-NN) dan Naïve Bayes yang kemudian kedua metode klasifikasi tersebut dibandingkan untuk melihat metode klasifikasi yang terbaik. Penelitian ini menggunakan praproses *data cleaning*, seleksi fitur, dan *scaling* untuk meningkatkan nilai akurasi. Hasil dari penelitian ini adalah metode klasifikasi K-Nearest Neighbors (K-NN) merupakan klasifikasi yang terbaik dengan nilai akurasi yang menggunakan kurva ROC/AUC bernilai 0.952 jika dibandingkan dengan metode klasifikasi Naïve Bayes yaitu 0.912.

Kata Kunci: Perbandingan, Naïve Bayes, K-Nearest Neighbors, Leukemia, *Machine Learning*

COMPARISON OF NAIVE BAYES AND K-NEAREST NEIGHBOR METHODS IN WHITE BLOOD CELL GENE MORPHOLOGICAL CLASSIFICATION

Muhammad Nur'adli Hasbi Gumay

ABSTRACT

In the health sector, the diagnosis of leukemia is a difficult thing because it is still diagnosed manually with the help of a doctor. The diagnosis manual may suffer from errors caused by human negligence. From these problems, it is necessary to diagnose the type of leukemia using advanced technology, namely Machine learning to overcome these problems. In this study, the machine learning processes data from the types of leukemia, namely Acute Myeloid Leukemia (AML) and Acute Lymphoblastic Leukemia (ALL) based on the morphological characteristics of the white blood cell genes. The data classification methods used for this research are K-Nearest Neighbor (K-NN) and Naïve Bayes, then the two classification methods are compared to see the best classification method. This study uses preprocessing of data cleaning, feature selection, and scaling to increase the accuracy value. The results of this study are the K-Nearest Neighbors (K-NN) classification method is the best classification with an accuracy value using the ROC/AUC curve worth 0.952 when compared to the Naïve Bayes classification method, which is 0.912.

Keywords: *Comparasion, Naïve Bayes, K-Nearest Neighbors, Leukemia, Machine Learning*