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

SKRIPSI

MUHAMMAD NUR'ADLI HASBI GUMAY

1810511103

UNIVERSITAS PEMBANGUNAN NASIONAL "VETERAN" JAKARTA

FAKULTAS ILMU KOMPUTER

PROGRAM STUDI INFORMATIKA

2022

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

SKRIPSI

Diajukan Sebagai Salah Satu Syarat untuk Memperoleh

Gelar Sarjana Komputer

MUHAMMAD NUR'ADLI HASBI GUMAY

1810511103

UNIVERSITAS PEMBANGUNAN NASIONAL "VETERAN" JAKARTA

FAKULTAS ILMU KOMPUTER

PROGRAM STUDI INFORMATIKA

2022

LEMBAR PENGESAHAN

LEMBAR PENGESAHAN

Dengan ini dinyatakan bahwa Skripsi berikut:

Nama	: Muhammad Nur'adli Hasbi Gumay
Nim	: 1810511103
Program Studi	: S1 Informatika
Judul	: PERBANDINGAN METODE NAIVE BAYES DAN K-NEAREST NEIGHBOR PADA KLASIFIKASI MORFOLOGI GEN SEL DARAH PUTIH

Telah berhasil dipertahankan di hadapan Tim penguji dan diterima sebagai bagian persyaratan yang diperlukan untuk memperoleh gelar Sarjana Komputer pada Program Studi S1 Informatika, Fakultas Ilmu Komputer, Universitas Pembangunan Nasional Veteran Jakarta.

Iin Ernawati, S.Kom., M.Si.
Penguji 1

Noor Falih, S.Kom., M.T.
Penguji 2

Yuni Widiastiwi, S.Kom., M.Si
Pembimbing 1

Mayanda Mega Santoni, M.Kom
Pembimbing 2

Dr. Ermatita, M.Kom.
Dekan

Desta Sandya Prasvita, S.Kom., M.Kom.
Ketua Program Studi

Ditetapkan di : Jakarta
Tanggal Pengesahan : 27 Juli 2022

PERNYATAAN ORISINALITAS

Tugas Skripsi ini adalah hasil karya sendiri, dan semua sumber yang dikutip maupun yang dirujuk telah saya nyatakan dengan benar.

Nama : Muhammad Nur'adli Hasbi Gumay

NIM : 18105011103

Tanggal : Kamis, 23 Juni 2022

Bilamana di kemudian hari ditemukan ketidaksesuaian dengan pernyataan saya ini, maka saya bersedia dituntut dan diproses sesuai dengan ketentuan yang berlaku.

Jakarta, 23 Juni 2022
Yang Menyatakan,

(Muhammad Nur'adli Hasbi Gumay)

PERNYATAAN PERSETUJUAN PUBLIKASI SKRIPSI UNTUK KEPENTINGAN AKADEMIS

Sebagai civitas akademik Universitas Pembangunan Nasional "Veteran" Jakarta, saya yang bertanda tangan di bawah ini:

Nama	: Muhammad Nur'adli Hasbi Gumay
NIM	: 1810511103
Fakultas	: Ilmu Komputer
Program Studi	: Informatika

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Pembangunan Nasional "Veteran" Jakarta Hak Bebas Royalti Non eksklusif *(Non-exclusive Royalty Free Right)* atas karya ilmiah saya yang berjudul:

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

Beserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti ini Universitas Pembangunan Nasional "Veteran" Jakarta berhak menyimpan, mengalih media/formatkan, mengelola dalam bentuk pangkalan data (*database)*, merawat, dan mempublikasikan Tugas Skripsi saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilih Hak Cipta. Demikian pernyataan ini saya buat dengan sebenarnya.

Dibuat Di : Jakarta
Pada Tanggal : 23 Juni 2022

Yang Menyatakan,

(Muhammad Nur'adli Hasbi Gumay)

LEMBAR PERSETUJUAN

Dengan ini menyatakan bahwa laporan tugas akhir sebagai berikut:

Nama	: Muhammad Nur'adli Hasbi Gumay
NIM	: 1810511103
Program Studi	: Informatika
Judul	: Perbandingan Metode Naive Bayes dan K-Nearest Neighbor Pada Klasifikasi Morfologi Gen Sel Darah Putih.

Sebagai bagian persyaratan yang diperlukan untuk mengikuti ujian Sidang Tugas Akhir/Skripsi pada Program Studi Informatika, Fakultas Ilmu Komputer, Universitas Pembangunan Nasional Veteran Jakarta.

Menyetujui,

Dosen Pembimbing 1	Dosen Pembimbing 2
Yuni Widiastiwi, S.Kom., M.Si.	Mayanda Mega Santoni, M.Kom.

Mengetahui,

Ketua Program Studi

Desta Sandya Prasvita, S. Komp., M.Kom.

Ditetapkan : Jakarta

Tanggal Persetujuan : Kamis, 23 Juni 2022

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*

KATA PENGANTAR

Puji dan syukur peneliti panjatkan kehadirat Allah SWT karena berkat rahmat dan anugerahNya sehingga peneliti dapat menyelesaikan Laporan Tugas Akhir ini dengan baik dan lancar. Penulisan laporan ini adalah untuk memenuhi dan sebagai syarat untuk menempuh proses lebih lanjut pelaksanaan Tugas Akhir di Program Studi Informatika, Fakultas Ilmu Komputer, Universitas Pembangunan Nasional "Veteran" Jakarta. Judul yang dipilih dalam penelitian ini adalah "*Perbandingan Metode Naive Bayes dan K-Nearest Neighbor Pada Klasifikasi Morfologi Gen Sel Darah Putih"*.

Peneliti menyadari bahwa tanpa bantuan dan bimbingan dari semua pihak penyusunan laporan Tugas Akhir ini tidak dapat berjalan dengan baik. Oleh karena itu peneliti mengucapkan terima kasih kepada:

1. Kepada kedua orang tua peneliti yang telah memberikan doa dan semangat selama proses pembuatan Proposal Tugas Akhir.
2. **Ibu Dr. Ermatita, M. Kom**. selaku Dekan Fakultas Ilmu Komputer Universitas Pembangunan Nasional "Veteran" Jakarta.
3. **Pak Desta Sandya Prasvita, S. Komp., M. Kom.** selaku kepala program studi Informatika Universitas Pembangunan Nasional "Veteran" Jakarta
4. **Ibu Yuni Widiastiwi, S. Kom., M.Si**. selaku dosen Pembimbing 1
5. **Ibu Mayanda Mega Santoni, S. Kom., M.Kom.** selaku Dosen Pembimbing 2.
6. Kepada Lovya Neysa serta teman-teman mahasiswa Informatika angkatan 2018 yang selalu memberikan semangat, bantuan, serta dukungan untuk peneliti dalam menyelesaikan Proposal Tugas Akhir.

Jakarta, 27 Mei 2022

Penulis

Muhammad Nur'adli Hasbi Gumay

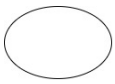
DAFTAR ISI

DAFTAR TABEL

DAFTAR GAMBAR

DAFTAR LAMPIRAN

DAFTAR SIMBOL

No.	Gambar Simbol	Nama	Keterangan
1.		*Start / End*	Simbol oval yang menggambarkan dimulainya atau berhenti suatu diagram alir.
2.		*Flow*	Simbol garis dengan tanda panah yang menggambarkan aliran data atau prosedur dari satu tahap ke tahap yang lain.
3.		*Process*	Simbol berbentuk persegi panjang dan menggambarkan proses atau kegiatan apa yang sedang terjadi.
4.		*Input / Output*	Menyatakan proses *input atau output*
5.		*Document*	Menyatakan bahwa *input* berasal dari dokumen dalam bentuk fisik atau *output* yang perlu dicetak