KLASIFIKASI DIAGNOSA PENYAKIT DEMAM BERDARAH DENGUE MENGGUNAKAN METODE HYBRID NAÜVE BAYES-K NEAREST NEIGHBOR

SKRIPSI



Disusun Oleh
ERA ALFI ZURROH
H72217022

PROGRAM STUDI MATEMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI SUNAN AMPEL
SURABAYA

2021

PERNYATAAN KEASLIAN

Saya yang bertanda tangan di bawah ini,

Nama : Era Alfi Zurroh

NIM : H72217022

Program Studi : Matematika

Angkatan : 2016

Menyatakan bahwa saya tidak melakukan plagiat dalam penulisan skripsi saya yang berjudul "Klasifikasi Diagnosa Penyakit Demam Berdarah Dengue Menggunakan Metode Hybrid Naïve Bayes-K Nearest Neighbor". Apabila suatu saat nanti terbukti saya melakukan tindakan plagiat, maka saya bersedia menerima sanksi yang telah ditetapkan.

Demikian pernyataan keaslian ini saya buat dengan sebenar-benarnya.

Surabaya, 25 Januari 2021

Yang menyatakan,

METERAL

MET

LEMBAR PERSETUJUAN PEMBIMBING

Skripsi oleh

Nama : Era Alfi Zurroh

NIM : H72217022

Judul Skripsi : Klasifikasi Diagnosa Penyakit Demam Berdarah Dengue

Menggunakan Metode Hybrid Naïve Bayes-K Nearest

Neighbor

telah diperiksa dan disetujui untuk diujikan.

Pembimbing I

Nurissaidah Ulinnuha, M.Kom

NIP. 199011022014032004

Pembimbing II

Dr. Abdulloh Hamid, M.Pd

NIP. 198508282014031003

Mengetahui,

Ketua Program Studi Matematika

UIN Sunan Ampel Surabaya

Aris Fanani, M.Kom

NIP. 198701272014031002

PENGESAHAN TIM PENGUJI SKRIPSI

Skripsi oleh

Nama

: Era Alfi Zurroh

NIM

: H72217022

Judul Skripsi

: Klasifikasi Diagnosa Penyakit Demam Berdarah Dengue

Menggunakan Metode Hybrid Naïve Bayes-K Nearest

Neighbor

Telah dipertahankan di depan Tim Penguji pada tanggal 28 Januari 2021

> Mengesahkan, Tim Penguji

Penguji I

Dian Candra Kini Novitasari, M.Kom

NIP. 198511242014032001

Penguji II

Putroue Keumala Intan, M.Si

NIP. 198805282018012001

Penguji III

Nurissaidah Climuha, M.Kom

NIP. 199011022014032004

Dr. Abdulloh Hamid, M.Pd

NIP. 197312252006041001

Mengetahui,

Dekan Fakultas Sains dan Teknologi

Sunan Ampel Surabaya

umatur Rusydiyah, M.Ag 37312272005012003



KEMENTERIAN AGAMA UNIVERSITAS ISLAM NEGERI SUNAN AMPEL SURABAYA PERPUSTAKAAN

Jl. Jend. A. Yani 117 Surabaya 60237 Telp. 031-8431972 Fax.031-8413300 E-Mail: perpus@uinsby.ac.id

LEMBAR PERNYATAAN PERSETUJUAN PUBLIKASI KARYA ILMIAH UNTUK KEPENTINGAN AKADEMIS

Sebagai sivitas akademika UIN Sunan Ampel Surabaya, yang bertanda tangan di bawah ini, saya:

Nama	ERA ALFI ZURROH
NIM	H72217022
Fakultas/Jurusan	SAINS DAN TEKNOLOGI / MATEMATIKA
E-mail address	eraalfi 29@gmail.com
UIN Sunan Ampe. ✓ Sekripsi yang berjudul:	gan ilmu pengetahuan, menyetujui untuk memberikan kepada Perpustakaan I Surabaya, Hak Bebas Royalti Non-Eksklusif atas karya ilmiah: Tesis Desertasi Lain-lain () I DIAGNOSA PENYAKIT DEMAM BERDARAH
PENGUE M	MENGGUNAKAN METODE HYBRID NAIVE

beserta perangkat yang diperlukan (bila ada). Dengan Hak Bebas Royalti Non-Ekslusif ini Perpustakaan UIN Sunan Ampel Surabaya berhak menyimpan, mengalih-media/format-kan, mengelolanya dalam bentuk pangkalan data (database), mendistribusikannya, dan menampilkan/mempublikasikannya di Internet atau media lain secara *fulltext* untuk kepentingan akademis tanpa perlu meminta ijin dari saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan atau penerbit yang bersangkutan.

NEIGHBOR

Saya bersedia untuk menanggung secara pribadi, tanpa melibatkan pihak Perpustakaan UIN Sunan Ampel Surabaya, segala bentuk tuntutan hukum yang timbul atas pelanggaran Hak Cipta dalam karya ilmiah saya ini.

Demikian pernyataan ini yang saya buat dengan sebenarnya.

BAYES - KNEAREST

Surabaya,

Penulis

ERA ALFI ZURROH)

ABSTRAK

Klasifikasi Diagnosa Penyakit Demam Berdarah Dengue Menggunakan

Metode Hybrid Naïve Bayes-K Nearest Neighbor

Demam Berdarah Dengue (DBD) adalah salah satu penyakit serius yang menular melalui gigitan nyamuk Aedes aegypti dan Aedes albopictus. Jumlah kasus demam berdarah yang semakin meningkat merupakan salah satu ancaman serius dalam sepuluh tahun terakhir yang akan terus berlanjut di wilayah tropis termasuk Indonesia. Indonesia termasuk negara tropis dengan jumlah kasus demam berdarah yang terbilang banyak dan terus mengalami peningkatan. World Health Organization (WHO) mencatat negara indonesia merupakan salah satu negara dengan kasus DBD tertinggi di Asia Tenggara. Pada penelitian ini akan digunakan metode *classifier* baru yang menggabungkan algoritma berbasis jarak K-Nearest Neighbor dan Naüve Bayes Classifier berbasis statistik. klasifikasi yang digunakan adalah menggunakan metode hybrid Naïve-K Nearest Neighbor. Sistem klasifikasi yang dibuat menghasilkan nilai akurasi, sensitivitas, dan spesifitas masing-masing, 95.4%, 96.2%, dan 94.4% dengan nilai K pada K-Nearest Neighbor=3 dan k-fold=10. Dalam penelitian ini dapat dibuktikan bahwa gabungan dari kedua metode tersebut menghasilkan peforma yang baik.

Kata kunci: DBD, Naïve Bayes, *K-Nearest Neighbor*, hybrid Naïve Bayes-*K-Nearest Neighbor*, Data Mining

ABSTRACT

Klasifikasi Diagnosa Penyakit Demam Berdarah Dengue Menggunakan

Metode Hybrid Naïve Bayes-K Nearest Neighbor

Dengue Hemorrhagic Fever (DHF) is a serious disease that is transmitted through the bites of the Aedes aegypti and Aedes albopictus mosquitoes. The increasing number of dengue cases is one of the serious threats in the last ten years which will continue in tropical regions including Indonesia. a tropical country with a fairly large number of dengue fever cases and the World Health Organization (WHO) noted that Indonesia is one of the countries with the highest DHF cases in Southeast Asia. In this study, a new classifier method will be used that combines the K-Nearest Neighbor distance-based algorithm and the statistical-based Naive Bayes Classifier. The classification process used is the Naive-K Nearest Neighbor hybrid method. The classification system created produces the respective accuracy, sensitivity, and specificity values, 95.4%, 96.2%, dan 94.4%,respectively, with the K value on the *K-Nearest Neighbor* =3 and k-fold=10. In study it can be proven that the combination of the two methods produces good performance.

Keywords: DBD, Naïve Bayes, *K-Nearest Neighbor*, hybrid Naïve Bayes-*K-Nearest Neighbor*, *Knowledge Discovery in Data Base*

DAFTAR ISI

HA	ALAN	MAN PERNYATAAN KEASLIAN	i
LF	EMBA	AR PERSETUJUAN PEMBIMBING	ii
PE	ENGE	SAHAN TIM PENGUJI SKRIPSI	iii
LF	EMBA	AR PERSETUJUAN PUBLIKASI	iv
ΑI	BSTR	AK	v
ΑI	BSTR	ACT	vi
DA	(FTA	R ISI	vii
DA	(FTA	R TABEL	ix
DA	(FTA	R GAMBAR	X
I	PEN	DAHULUAN	1
	1.1.	Latar Belakang Masalah	1
	1.2.	Rumusan Masalah	7
	1.3.	Tujuan Penelitian	8
	1.4.	Manfaat Penelitian	8
	1.5.	Batasan Masalah	9
	1.6.	Sistematika Penulisan	9
II	TINJ	JAUAN PUSTAKA	11
	2.1.	Demam Berdarah Dengue (DBD)	11
		2.1.1. Usia	12
		2.1.2. Hemoglobin	12
		2.1.3. Leukosit	13
		2.1.4. Trombosit	13
		2.1.5. Hematokrit	13
	2.2.	Data Mining	14
		2.2.1. Manfaat Data Mining	14
	2.3.	Normalisasi	16
	2.4.	K-fold cross validation	16

	2.5.	Principal Component Analysis(PCA)	17
	2.6.	Klasifikasi	19
	2.7.	Hybrid Naïve Bayes-KNN	20
	2.8.	K-Nearest Neighbor(KNN)	23
	2.9.	Confusion Matrix	24
	2.10.	Integrasi Keilmuan	25
Ш	MET	ODE PENELITIAN	32
	3.1.	Jenis Penelitian	32
	3.2.	Pengumpulan Data	32
	3.3.	Variabel Penelitian	32
	3.4.	Tahapan Penelitian	34
IV	HAS	IL DAN PEMBAHASAN	37
	4.1.	Pengelolahan Data	37
	4.2.	Normalisasi Data	37
	4.3.	Metode Naïve Bayes untuk Data Numerik	38
	4.4.	Menghitung Nilai Probabilitas Tiap Data	40
	4.5.	Metode Naïve Bayes K-Nearest Neighbor	41
	4.6.	Pengujian Algoritma	43
	4.7.	Evaluasi Model	45
		4.7.1. Metode K-Nearest Neighbor	45
		4.7.2. PCA- K Nearest Neighbor	45
		4.7.3. Hybrid Naïve-K Nearest Neighbor	46
	4.8.	Integrasi Keilmuan	47
V	PEN	UTUP	51
	5.1.	Kesimpulan	51
	5.2.	Saran	52
DA	FTA	R PUSTAKA	52

DAFTAR TABEL

4.1	Sampel data penelitian	37
4.2	Data sesudah dinormalisasi	38
4.3	Mean dan standart deviasi berdasarkan pengolahan data	40
4.4	Nilai probabilitas setiap data	41
4.5	Seleksi fitur menggunakan Naive Bayes	41
4.6	Subset fitur yang terpilih menggunakan Naive Bayes	42
4.7	Uji data testing dalam menentukan K mayoritas	43
4.8	Tabel Hasil Klasifikasi Hybrid Naüve-KNN	44
4.9	Tabel Hasil Evaluasi metode Hybrid Naüve-KNN	44
4.10	Tabel Hasil Evaluasi metode KNN	45
4.11	Tabel Hasil Evaluasi metode PCA-KNN	45
4.12	Tabel Hasil Klasifikasi Hybrid Naive-KNN, K-NN, dan PCA-KNN	46

DAFTAR GAMBAR

2.1	Siklus penularan penyakit demam berdarah dengue	1.
2.2	Tahapan-tahapan Knowledge Discovery Databases	1.
2.3	Model 3 k-fold cross validation	17
2.4	Confusion Matrix	24
3.1	Diagram alir metode hybrid Nave Bayes-K Nearest Neighbor	34

BABI

PENDAHULUAN

1.1. Latar Belakang Masalah

Demam Berdarah Dengue (DBD) adalah salah satu penyakit serius yang menular melalui gigitan nyamuk *Aedes aegypti* dan *Aedes albopictus*(Kinansi and Pujiyanti, 2020). Nyamuk Aedes merupakan salah satu nyamuk yang tersebar luas di seluruh wilayah, kecuali wilayah yang berada pada ketinggian lebih dari seratus meter di atas permukaan laut. Beberapa penelitian menyatakan bahwa pada suhu 28-32 derajat celcius dengan kelembaban yang tinggi, nyamuk Aedes akan dapat bertahan hidup dengan jangka waktu yang lama. Faktor-faktor yang dapat mempengaruhi perkembangbiakan tersebut diantaranya adalah perubahan suhu, curah hujan, kondisi lingkungan, mobilitas penduduk dan perilaku masyarakat yang tidak sehat(Handayani, 2015).

Demam Berdarah Dengue (DBD) merupakan penyakit yang disebabkan karena terdapat virus dengue yang berhasil masuk ke dalam sistem peredaran darah dan menyerang sistem imun tubuh yang menyebabkan presentase nilai sel neutrofil dan limfosit mengalami perubahan spesifik(Kartika, 2018). Virus dengue berasal dari genus *flavivirus* dari *flaviviridae* dengan 4 serotype virus yang berisiko menyebabkan kematian(Permatasari et al., 2020). Dalam Genus tersebut telah terkandung lebih dari 50% macam yang berhubungan dengan penyakit manusia. Infeksi oleh satu serotipe virus dengue akan menghasilkan imunitas jangka panjang pada virus tersebut .

Jumlah kasus demam berdarah yang semakin meningkat merupakan salah satu ancaman serius dalam sepuluh tahun terakhir yang akan terus berlanjut di wilayah tropis(Chatterjee et al., 2017). Biaya tindakan pencegahan yang terus meningkat seiring dengan situasi adanya wabah penyakit demam berdarah juga memberikan beban tambahan pada perekonomian daerah. Di berbagi negara, virus dengue sendiri telah merenggut jutaan nyawa dan terus terjadi di sebagian besar wilayah tropis di Dunia. World Health Organization(WHO) menyebutkan bahwa kasus demam berdarah telah terjadi lebih dari 50 juta kasus di Dunia pada setiap tahunnya(Akbar and Maulana Syaputra, 2019).

Indonesia termasuk negara tropis dengan jumlah kasus demam berdarah yang terbilang banyak dan terus mengalami peningkatan(Qureshi and Atangana, 2019). World Health Organization (WHO) mencatat negara indonesia merupakan salah satu negara dengan kasus DBD tertinggi di Asia Tenggara. Kementerian Kesehatan RI hingga februari 2020 menyatakan terdapat enam belas ribu kasus DBD di Indonesia, beberapa diantaranya tercatat telah memasuki tahapan klinis dengan tanda panas serta adanya tanda perdarahan(Priyono, P., & Nuris, 2020). Infeksi virus dengue dinyatakan dapat menyerang semua usia. Tercatatat sebanyak 95% penyakit DBD menyerang pada anak berusia dibawah 15 tahun. Banyaknya kasus serta dampak yang ditimbulkan karena adanya penyakit tersebut mengharuskan kita untuk mewaspadai dan melakukan pencegahan dini agar kasus penyakit ini dapat berkurang.

Penyakit demam berdarah ditandai dengan gejala demam selama dua hingga tujuh hari sejak gigitan nyamuk dan dapat berlangsung selama sepuluh hari. Terjadinya pendarahan dari hidung, gusi, atau di bawah kulit, penurunan trombosit, dan hemokonsentrasi yaitu mengalami pemekatan pembuluh darah(Adharani and Ambo, 2019). Gejala demam berdarah yang sering ditunjukkan pada penderita secara umum yaitu nyeri kepala, nyeri pada otot seluruh tubuh dan sakit pada bagian belakang mata, sesak nafas, batuk, mual, muntah, dan lain sebaginya(Hia, 2019). Tindakan yang tepat sangat dibutuhkan ketika seseorang mengalami gejala-gejala tersebut. Faktor utama terjadinya penyakit demam berdarah dipengaruhi oleh tingkat kebersihan lingkungan sekitar. Lingkungan yang kotor dapat menjadi sarang nyamuk *Aedes aegypti* untuk berkembang biak.

Munculnya penyakit dapat disebabkan oleh lingkungan yang kotor memberikan pelajaran bagi masyarakat bahwa menjaga kebersihan merupakan sebuah kebutuhan. Bahaya dari dampak yang ditimbulkan tersebut tentu tidak dapat disepelekan. Islam telah mengajarkan kepada umatnya agar selalu menjaga kebersihan serta untuk hidup sehat baik jasmani maupun rohani. Anjuran untuk selalu menjaga kebersihan telah banyak dibahas dalam Islam. Dalam Islam, menjaga kebersihan menjadi hal yang sangat penting dalam upaya menjaga kesehatan sehingga dapat terhindar dari bakteri maupun virus penyakit. Pentingnya dalam menjaga kebersihan dibuktikan dengan adanya firman Allah SWT yang menerangkan bahwa orang yang menjaga dan mencintai kebersihan akan dicintai Allah SWT yang terdapat dalam surat al-Hasyr ayat 23 sebagai berikut:

Artinya: "Dialah Allah Yang tiada Tuhan selain Dia, Raja, Yang Maha Suci, Yang Maha Sejahtera, Yang Mengaruniakan Keamanan, Yang Maha Memelihara, Yang Maha Perkasa, Yang Maha Kuasa, Yang Memiliki segala Keagungan, Maha

Suci Allah dari apa yang mereka persekutukan".

Berdasarkan ayat di atas diperjelas dengan kata Allah maha suci dapat diartikan bahwa keimanan seseorang dapat dilihat dari tingkatan seseorang dalam menjaga kebersihan. Rasulullah *Shallallahu 'alaihi wa sallam* juga menerangkan tentang betapa pentingnya kebersihan dalam kehidupan umat Islam karena sesungguhnya Allah SWT itu suci yang menyukai hal-hal yang suci, Dia Maha Bersih yang menyukai kebersihan, Dia Maha Mulia yang menyukai kemuliaan, Dia Maha Indah yang menyukai keindahan, karena itu sebagai umat muslim diwajibkan untuk membersihkan tempat-tempat yang ditempati.

Pentingnya menjaga kebersihan dimaksudkan untuk meminimalisir penularan penyakit demam berdarah. Proses penularan penyakit demam berdarah yaitu diawali dengan gigitan nyamuk Aedes dan kemudian ditularkan kepada manusia lain. Penularan penyakit demam berdarah disebarkan oleh nyamuk Aedes setelah menggigit manusia yang sedang mengalami viremia. Viremia adalah suatu masa dimana virus dengue telah berada di dalam aliran darah manusia. Virus dengue dapat tumbuh serta berkembangbiak pada tubuh nyamuk Aedes tanpa menyebabkan kematian pada nyamuk itu sendiri. Virus dengue mampu untuk mempertahankan keberadaannya di alam dengan melalui dua mekanisme yaitu dengan transmisi horizontal dan vertikal. Transmisi horizontal terjadi antara vertebrata viremia yang ditularkan oleh nyamuk Aedes sedangkan transmisi vertikal terjadi dari nyamuk infektif ke generasi berikutnya(Noshirma et al., 2020). Penyakit demam berdarah dapat dikatakan sebagai penyakit yang tidak boleh diabaikan karena tergolong merupakan penyakit yang berbahaya. Untuk mengobati penyakit demam berdarah ini salah satunya dibutuhkan diagnosa awal

yang tepat disertai pengobatan yang intensif. Demam berdarah merupakan penyakit yang perlu diwaspadai karena penyakit tersebut dapat menyebabkan kematian jika tidak dilakukan tindakan yang tepat(Nur Itsna, 2020). Perlunya dilakukan klasifikasi data mining secara otomatis adalah untuk mempermudah dalam melakukan diagnosa penyakit DBB sehingga dapat mempercepat tindakan terhadap penyakit demam berdarah.

Klasifikasi adalah salah satu teknik multivariat terpenting yang digunakan dalam statistik. Hal ini erat kaitannya dengan prediksi dan yang menarik masalah klasifikasi kadang disebut masalah prediksi khususnya pada data mining. Dalam statistik, klasifikasi adalah prosedur di mana objek individu di tempatkan ke dalam kelompok berdasarkan informasi kuantitatif pada satu atau lebih karakteristik yang melekat pada objek yang disebut sebagai sifat, variabel, karakter, dan lainnya(Arifin and Ariesta, 2019). Klasifikasi juga merupakan suatu teknik berdasarkan pada set pelatihan objek berlabel sebelumnya. Terdapat beberapa metode yang menangani masalah klasifikasi seperti algoritma berbasis statistik Naïve Bayes dan algoritma berbasis jarak *K-Nearest Neighbor*.

Pada penelitian sebelumnya dalam studi deteksi berita bohong menggunakan metode Naïve Bayes diperoleh akurasi sebesar 78,6%. Akurasi yang diperoleh pada penelitian tersebut mengatakan bahwa algoritma Naïve Bayes sangat disarankan untuk metode klasifikasi yang digunakan pada penelitian yang dilakukan oleh Inggrid Yanuar(Inggrid Yanuar, Rosa Andrie, 2017). Pada Penelitian selanjutnya dalam proses klasifikasi yang dilakukan oleh Shofia dalam mendiagnosa penyakit DBD menggunakan metode *K-Nearest Neighbor* diperoleh hasil akurasi sebesar 84.7%. Penelitan sebelumnya yang dilakukan oleh Safri

dengan menggunakan metode *K-Nearest Neighbor* dan Naïve Bayes untuk mengklasifikasikan penduduk yang berhak mendapatkan KIS (Kartu Indonesia Sehat) diperoleh bahwa proses pengklasifikasian menunjukkan akurasi dari algoritma *K-Nearest Neighbor* sebesar 64% sedangkan kombinasi antara *K-Nearest Neighbor-Naïve Bayes Classifier* dalam menentukan klasifikasi memiliki peningkatan akurasi sebesar 32% sehingga hasil akurasinya sebesar 96%(Safri et al., 2018). Penelitian terdahulu yang dilakukan oleh Annur juga menjelaskan tentang kombinasi metode KNN dan Naïve menggunakan seleksi fitur *information Gain* untuk klasifikasi penyakit jantung memperoleh hasil akurasi sebesar 92.31% (Syafitri Hidayatul AA, Yuita Arum S, 2018). Kombinasi metode Naïve Bayes-KNN juga didukung dalam melakukan klasifikasi penyakit kelamin pada wanita hasil akurasi yang diperoleh adalah sebesar 99.17%(Nazaruddin et al., 2019). Dalam penelitian tersebutmenunjukkan bahwa kombinasi metode Naïve Bayes-*K Nearest Neighbor* memberikan akurasi lebih baik dibandingkan dengan hasil akurasi metode dasarnya yaitu *K-Nearest Neighbor*.

Hasil perbandingan penelitian yang telah dikaji menunjukkan bahwa terdapat kelemahan dan kelebihan pada setiap metode. Berdasarkan penelitian yang dilakukan oleh Safri ditunjukkan bahwa kelemahan metode *K-Nearest Neighbor* berada pada proses perhitungan yang dilakukan hampir disetiap data pada tahap klasifikasi. Jumlah fitur yang banyak akan menjadi beban pada proses klasifikasi *K-Nearest Neighbor*. Gabungan metode menggunakan Naïve Bayes dilakukan agar dapat mengurangi beban dalam menyeleksi fitur untuk dilanjutkan pada proses klasifikasi *K-Nearest Neighbor* sehingga dihasilkan proses yang lebih cepat dengan performa yang lebih baik. Pada penelitian ini akan dikaji mengenai modifikasi algoritma yang merupakan kombinasi metode KNN dengan Naïve

Bayes untuk mengatasi kelemahan pada metode *K-Nearest Neighbor*. Naïve Bayes berperan sebagai proses seleksi fitur dengan mempertimbangkan probabilitas setiap atribut. Atribut yang memenuhi syarat minimum probabilitas akan dilanjutkan dengan proses klasifikasi menggunakan metode *K-Nearest Neighbor*.

Dengan mempertimbangkan penelitian terdahulu maka pada penelitian kali ini akan dilakukan proses klasifikasi dengan menggunakan gabungan metode Naïve Bayes berperan sebagai seleksi fitur dan *K-Nearest Neighbor* sebagai metode klasifikasi dengan judul Klasifikasi Diagnosa Penyakit Demam Berdarah Dengue Menggunakan Metode *Hybrid Naïve Bayes-K Nearest Neighbor*. Penelitian ini juga akan membandingkan metode *K-Nearest Neighbor* sendiri sebagai dasar metode tersebut dengan metode *Hybrid Naïve Bayes-K Nearest Neighbor*. Selain itu dalam penelitian ini akan dilakukan perbandingan peforma dari metode *naïve bayes* dengan seleksi fitur statistika yang paling populer yaitu *Principal Component Analysis* (PCA) dalam meningkatkan akurasi pada Metode *K-Nearest Neighbor* (Gupta and Mittal, 2018). Hasil dari proses klasifikasi diharapkan dapat membantu proses penanggulangan penyakit demam berdarah.

1.2. Rumusan Masalah

Berdasarkan permasalahan yang telah diuraikan pada latar belakang, perumusan masalah penelitian ini sebagai berikut:

- 1. Berapakah nilai K terbaik untuk melakukan klasifikasi diagnosa penyakit Demam Berdarah Dengue (DBD) menggunakan metode hybrid Naïve Bayes-K-Nearest Neighbor?
- 2. Fitur apa yang dihasilkan dengan menggunkan metode Naïve Bayes untuk

diagnosa penyakit Demam Berdarah Dengue (DBD) dari metode *hybrid Naïve Bayes-K-Nearest Neighbor*?

3. Bagaimana hasil dari klasifikasi untuk diagnosa penyakit Demam Berdarah Dengue (DBD) dari metode *hybrid* Naïve Bayes-*K-Nearest Neighbor* dibandingkan dengan metode *K-Nearest Neighbor* dan PCA - *K-Nearest Neighbor*?

1.3. Tujuan Penelitian

- 1. Dapat mengetahui nilai K terbaik dalam klasifikasi untuk diagnosa penyakit Demam Berdarah Dengue (DBD)menggunakan metode*hybrid Naïve Bayes-K Nearest Neighbor*.
- 2. Dapat mengetahui fitur yang dihasilkan oleh metode Naï Bayes-K untuk diagnosa penyakit Demam Berdarah Dengue (DBD)dari metode hybrid Naïve Bayes-K Nearest Neighbor dibandingkan metode K-Nearest Neighbor.
- 3. Dapat mengetahui hasil dari klasifikasi untuk diagnosa penyakit Demam Berdarah Dengue (DBD)dari metode *hybrid* Naïve Bayes-*K-Nearest Neighbor* dibandingkan metode *K-Nearest Neighbor* dan *PCA K-Nearest Neighbor*.

1.4. Manfaat Penelitian

Penelitian ini diharapkan bermanfaat bagi berbagai lapisan masyarakat, seperti yang tertulis sebagai berikut:

1. Manfaat Teoritis

Penelitian ini diharapkan menjadi referensi bagi penulis-penulis berikutnya

dalam klasifikasi menggunakan arsitektur dari algoritma Na \ddot{i} ve Bayes dan KNearest Neighbor untuk diagnosa.

2. Manfaat Praktis

a. Bagi Penulis

Menambah wawasan baru bagi penulis dalam menerapkan algoritma Naïve Bayes dan *K Nearest Neighbor* untuk klasifikasi penyakit DBD.

b. Bagi Masyarakat Umum

Membantu ahli medis dalam mendiagnosis penyakit DBD dengan lebih mudah dan hasil akurasi yang baik.

1.5. Batasan Masalah

Luasnya ruang lingkup permasalahan, maka pada penelitian ini diberikan batasan-batasan masalah sebagai berikut:

- Metode pembelajaran yang digunakan untuk data klasifikasi diagnosa penyakit DBD merupakan data pasien tes hermatologi untuk diagnosa penyakit DBD.
- 2. Kategori yang ditentukan pada proses klasifikasi ini yaitu indikasi pasien positif DBD dan pasien negatif DBD.
- 3. Nilai K pada metodde *K-Nearest Neighbor* yang digunakan pada penelitian ini adalah K=3, K=5, K=7, K=9.

1.6. Sistematika Penulisan

BAB I PENDAHULUAN

Berisi latar belakang masalah, perumusan masalah, tujuan penelitian, manfaat

penelitian, metodologi penelitian, dan sistematika penulisan.

BAB II TINJAUAN PUSTAKA

Berisi tentang kajian terhadap teori-teori yang digunakan untuk memahami permasalahan yang ada pada penelitian ini.Teori-teori tersebut meliputi Demam Berdarah Dengue (DBD), Klasifikasi, *Confusion matrix*, Data mining, algoritma Naïve Bayes, dan *K Nearest Neighbor* dan integrasi keilmuan mengenai penyakit DBD dalam Islam.

BAB III METODE PENELITIAN

Berisi jenis penelitian, pengumpulan data, dan kerangka penelitian.

BAB IV HASIL DAN PEM<mark>BA</mark>HASAN

Berisi tentang penjabaran hasil yang didapat dalam penelitian ini dan analisis data untuk menarik kesimpulan sesuai dengan tujuan penelitian.

BAB V PENUTUP

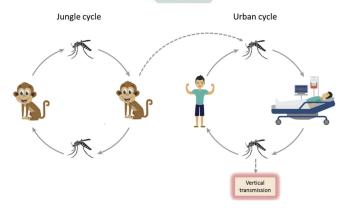
Berisi tentang kesimpulan dari hasil penelitian dan saran.

BAB II

TINJAUAN PUSTAKA

2.1. Demam Berdarah Dengue (DBD)

Demam berdarah atau yang memiliki istilah lain dengue haemoragic fever merupakan suatu penyakit yang dalam penularannya melibatkan perantara nyamuk Aedes. Cara penularan penyakit ini sendiri melalui gigitan nyamuk aedes yang telah terinfeksi virus dengue (Ningtyas et al., 2019). Virus dengue didukung oleh replikasi alternatif pada nyamuk dan manusia. Vektor terpenting dalam penyebaran virus dengue adalah nyamuk Aedes aegypti dan Aedes Albopictus karena efisiensi penularannya yang tinggi ke manusia. Aedes aegypti berasal dari Afrika dan sekarang telah tersebar di sebagian wilayah tropis dan sub- tropis di Dunia. Nyamuk Aedes Albopictus merupakan spesies endemik Asia Tenggara dan sekarang penyebarannya telah sampai ke berbagai belahan Dunia (Ahammad et al., 2019).



Gambar 2.1 Siklus penularan penyakit demam berdarah dengue

(sumber:Chen, 2018)

Berdasarkan ilustrasi pada Gambar 2.1 virus dengue secara alami ditularkan oleh nyamuk Aedes antara manusia atau monyet. Penularan transovarial atau yang disebut dengan penularan secara vertikal pada nyamuk itu sendiri juga kerap terjadi. Nyamuk *Aedes Aegypti* berjenis kelamin betina merupakan nyamuk yang sangat sering menggigit beberapa orang yang terkena virus dengue, sehingga hal tersebut dapat mempercepat penularan virus dengue. Setelah penularan ke inang manusia dan setelah masa inkubasi 3 sampai 14 hari, sebagian besar individu yang terkena dampak memasuki periode demam akut 2 sampai 10 hari(Butarbutar et al., 2019). Beberapa hal yang berhubungan dengan derajat klinis penyakit DBD adalah sebagai berikut:

2.1.1. Usia

Hubungan usia dengan terjadinya kasus DBD adalah mengenai resiko tertular oleh virus *dengue*. Penyakit DBD umumnya menyerang anak-anak. Anak usia di bawah 5 tahun memiliki resiko lebih tinggi dibanding anak usia diatas 5 tahun yang dikarenakan oleh tingkat imunitas tubuh yang lebih rendah(Sendangmulyo et al., 2017).

2.1.2. Hemoglobin

Hemaglobin merupakan protein eritrosit yang berfungsi untuk membantu membawa oksigen ke seluruh tubuh melalui paru-paru dan membawa karbon dioksida dari tubuh ke paru-paru agar dapat dikeluarkan pada saat proses mengeluarkan udara ke luar tubuh atau yang disebut dengan ekhalasi. Eritrosit atau yang disebut dengan sel darah merah merupakan komponen sel terbanyak dalam darah yaitu sebesar 40-50%. Eritrosit mudah berubah bentuk, hal tersebut menyebabkan eritrosit mampu berjalan di pembuluh darah dengan mudah. eritrosit

memiliki rata-rata umur 120 hari. Produksi eritrosit dikontrol oleh sebuah hormon yang dihasilkan oleh ginjal yang disebut dengan hormon eritropoetin (Dewi et al., 2020).

2.1.3. Leukosit

Leukosit adalah sel darah putih yang merupakan bagian terpenting terhadap sistem pertahanan tubuh dengan fungsi untuk melawan mikroorganisme penyebab infeksi, se tumor, serta zat-zat asing yang berbahaya. Pada penderita DBD dapat mengalami leukopenia atau disebut dengan kondisi rendahnya jumlah sel darah putih di dalam tubuh hingga mengalami leukositosis sedang atau disebut dengan kondisi dimana seseorang memiliki jumlah sel darah putih yang terlalu banyak (Bakhri, 2018).

2.1.4. Trombosit

Trombosit adalah keping darah yang berfungsi dalam mekanisme tubuh untuk menghentikan kehilangan darah yang berlebihan atau yang disebut dengan pendarahan. sel ini tidak memiliki nukleus dan dihasilkan oleh megakariosit dalam sumsum tulang. Kasus pada pasien DBD adalah terjadinya trombositopenia yang merupakan kondisi saat jumlah trombosit rendah atau di bawah nilai normal. Hal tersebut diakibatkan karena munculnya antibodi terhadap trombosit karena kompleks antigen serta antibodi yang terbentuk (Suparmono, 2021).

2.1.5. Hematokrit

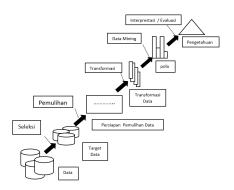
Hematokrit merupakan suatu nilai yang merupakan kosentrasi eritrosit. Nilai hematokrit dapat meningkat karena adanya penurunan volume plasma darah atau juga dapat disebabkan oleh adanya peningkatan kadar sel darah merah, proses peningkatan kadar hematokrit disebut dengan hemokonsentrasi.Peningkatan hematokrit menandakan bahwa akan terjadi perembesan plasma. Hal tersebut dibuktikan dengan adanya peningkatan kadar hematokrit hingga $\geq 20\%$ dari nilai awal. Nilai hematokrit juga dapat menurun, hal tersebut disebabkan oleh adanya penurunan seluler darah atau peningkatan kadar plasma darah (Cahyani et al., 2020).

2.2. Data Mining

Data Mining merupakan proses dalam mencari suatu pola yang menggunakan teknik statistik, matematika, dan kecerdasan buatan untuk mengidentifikasi informasi yang bermanfaat dan juga pengetahuan terkait dari *database* yang begitu besar (Bayes, 2019). Secara umum data mining bisa disebut dengan serangkaian proses yang bertujuan untuk menemukan nilai yang tidak dapat diketahui secara langsung dari sebuah data.

2.2.1. Manfaat Data Mining

Manfaat data mining adalah untuk mencari pengetahuan yang terdapat dalam basis data yang besar sehingga dalam mengarahkan pelaksanaan proses data mining adalah menggunakan *Knowledge Discovery Databases*(KDD). KDD merupakan proses menentukan informasi dan pola yang berguna didalam data(Nikmatun and Waspada, 2019).



Gambar 2.2 Tahapan-tahapan Knowledge Discovery Databases.

(sumber:Erwansyah, 2019)

Tahapan-tahapan *Knowledge Discovery Databases* pada Gambar 2.2 dapat dijelaskan sebagai berikut:

- 1. Seleksi data dengan melakukan pembersihan data yang tidak konsisten dan terdapat noise atau persebaran data yang tidak seimbang
- 2. Pemulihan data di mana beberapa sumber data dapat dikombinasikan
- 3. Transformasi data mengubah data sesuai bentuk yang diinginkann untuk dilakukan proses selanjutnya
- 4. Data mining merupakan pengaplikasian data mining dengan melakukan proses ekstraksi pola dari data yang ada
- 5. Interprestasi evaluasi merupakan proses pemahaman pola yang dapat digunakan sebagai pengetahuan dalam mendukung pengambilan keputusan.
- 6. Pengetahuan merupakan informasi yang didapat setelah dilakukan proses.

2.3. Normalisasi

Normalisasi data merupakan salah satu cara untuk mengatasi masalah pada set data yang memiliki rentang data tinggi. Rentang pada data yang cukup tinggi akan mempengaruhi hasil pada pengolahan data. Masalah ini dapat diatasi dengan cara menormalkan data. Normalisasi data dapat dilakukan dengan beberapa metode. Salah satunya dalam metode *Z-Score Normalization*. Metode ini menormalkan data dengan berdasarkan nilai rata-rata data dan nilai standar deviasi data (Nasution et al., 2019). *Z-score Normalization* diformulakan dengan rumus pada Persamaan 2.1 sebagai berikut:

$$Z_i = \frac{X_i - \mu}{\sigma} \tag{2.1}$$

Keterangan:

 Z_i = data hasil normalisasi

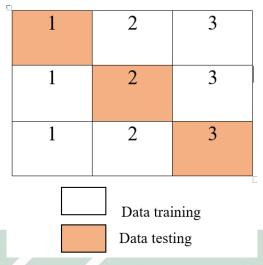
 X_i = data asli sebelum dinormalisasi

 μ = nilai rata-rata data

 σ = standar deviasi data

2.4. K-fold cross validation

K-fold cross validation merupakan suatu teknik validasi model yang bertujuan untuk melihat bagaimana hasil statistik analisis akan membentuk kumpulan data . Secara garis besar *k-fold cross validation* dimanfaatkan sebagai teknik perbaikan model sehingga dapat memperkirakan tingkat keakuratan sebuah model prediksi dalam praktiknya (Tjasyono, 2019). Penggunakan model *K-fold cross validation*dapat dilihat pada Gambar 2.3 berikut.



Gambar 2.3 Model 3 k-fold cross validation

Gambar di atas merupakan penggunaan *k-fold cross validation* untuk k sama dengan 3 yang memiliki arti bahwa kan dilakukan percobaan sebanyak 3 kali dan setiap subset data dapat memiliki kesempatan sebagai data testing atau sebagai data *training*. Percobaan pertama data pertama berperan sebagai data *testing* sedangkan data kedua dan ketiga sebagai data *training*. Percobaan kedua data kedua sebagai data *testing* sedangkan data pertama dan ketiga sebagai data *training*. Percobaan ketiga data ketiga sebagai data *testing* sedangkan data pertama dan kedua sebagai data *training*.

2.5. Principal Component Analysis(PCA)

Principal Component Analysis (PCA) atau analisis vaktor merupakan salah satu cara dalam teknik statistika yang digunakan dalam mencari hubungan atau korelasi antar variabel yang digunakan dalam penelitian. Berdasarkan ilmu statistika PCA merupakan suatu teknik yang digunakan untuk menyederhanakan suatu data. Proses penyederhanaan data dilakukan dengan transformasi linier sehingga terbentuk koordinat baru dengan variansi yang maksimum. Secara umum PCA

berfungsi untuk mereduksi suatu set variabel yang berdimensi tinggi menjadi lebih rendah tanpa menghilangkan informasi dari data awal (Annisa Halida, 2020).

PCA mengubah variabel asli yang saling berkorelasi satu dengan yang lainnya menjadi satu set variabel baru yang lebih kecil dan saling bebas. Analisis komponen utama merupakan analisis proses awalan dari analisis berikutnya. Sebagai contoh PCA digunakan sebagai masukan analisis faktor. Secara matematis PCA merupakan bentuk kombinasi linier tertentu dari p peubah acak $x1, x2, x3, \cdots, xp$. PCA terbentuk dari variansi dan kovariansi melalui suatu kombinasi linear pada vaiabel masukan. PCA dapat digunakan untuk melakukan pengecekan terhadap beberapa variabel yang saling berkorelasi. Langkah-langkah dari algoritma PCA adalah sebagai berikut (Putra et al., 2020).

1. Hitung matriks kovarian dengan menggunakan persamaan 2.2 sebagai berikut:

$$Kovarian(x,y) = \frac{\sum_{i=1}^{n} (x_i - \bar{x}) - ((y_i - \bar{y}))}{n-1}$$
(2.2)

Keterangan:

 y_i =data y ke i

 x_i =data x ke i

n =banyaknya data

 \bar{x} =rata-rata pada variabel x

 \bar{y} =rata-rata pada variabel y

2. Hitung nilai eigen dengan menggunakan persamaan 2.3 sebagai berikut:

$$det|C - \lambda I| = 0 (2.3)$$

Keterangan:

det= determinan

C= matriks kovarian

 λ = nilai eigen

I = matriks identitas

3. Hitung vektor eigen dengan menggunakan persamaan 2.4 sebagai berikut:

$$|C - \lambda I|X = 0 \tag{2.4}$$

Keterangan:

C= matriks kovarian

 λ = nilai eigen

I= matriks identitas

X= vektor eigen

- 4. Urutkan vektor eigen berdasarkan nilai eigen terbesar
- Dapatkan data yang ditransformasikan menggunakan vektor eigen yang diurutkan

2.6. Klasifikasi

Klasifikasi menurut bahasa disebut *classificatie* yang merupakan kata serapan dari bahasa Belanda. Istilah klasifikasi atau *classification* sendiri berasal dari Perancis yang diartikan sebagai suatu metode untuk menyusun data secara sistematis atau berdasarkan aturan yang telah ditetapkan. Menurut ilmu

pengetahuan klasifikasi adalah suatu proses terhadap penilaian objek data untuk dikelompokkan ke dalam kelas tertentu dari jumlah kelas yang telah tersedia. Untuk melakukan proses pada klasifikasi terdapat dua cara yaitu klasifikasi secara manual dan dengan bantuan teknologi. proses klasifikasi dengan bantuan teknologi diantaranya dapat menggunakan algoritma seperti Naive Bayes, Fuzzy, *Decission Tree,Support Vector Machine*, dan Jaringan Syaraf Tiruan (*Bacpropagation*). Komponen-komponen dasar pada proses klasifikasi yaitu, kelas, *predictor*, *training dataset*, *testing dataset*. Konsep dari proses *training* dan *testing* sendiri adalah untuk membangun sistem klasifikasi dengan training data set dan menguji metode dengan testing data set(Rohman, 2019).

2.7. Hybrid Naïve Bayes-KNN

Naïve Bayes merupakan suatu algoritma yang digunakan dalam proses klasifikasi dengan menggunakan metode statistika serta probabilitas yang telah dikembangkan oleh seorang ilmuwan inggris bernama Thomas Bayes. Teorema yang telah dikemukakan tersebut bertujuan untuk memprediksi peluang masa depan berdasarkan masa sebelumnya yang sekarang lebih dikenal dengan teorema Bayes. Naïve sendiri diasumsikan sebagai kondisi antar atribut yang saling bebas yang kemudian dikombinasikan dengan teorema Bayes sehingga dikenal dengan Naïve Bayes(Rahman, F., Ihsan, M., Pristianto, A., Khadijah, S., & Budi, 2019). Komponen-komponen dasar pada proses klasifikasi yaitu, kelas, *predictor, training dataset, testing dataset*. Konsep dari proses training dan testing sendiri adalah untuk membangun sistem klasifikasi dengan training data set dan menguji metode dengan testing data set. Metode *hybrid* merupakan gabungan dari algoritma Naïve bayes sebagai seleksi fitur dan *K Nearest Neighbor* sebagai proses klasifikasi. Gabungan dari kedua algoritma tersebut dilakukan dengan mencari nilai

probabilitas dari setiap data atribut yang akan diklasifikasikan dengan klasifikasi data kontinyu pada algoritma Na \ddot{i} ve Bayes. Data yang memiliki probabilitas lebih besar dari nilai alpha selanjutnya akan dijumlah berdasarkan masing-masing variabel. kemudian seleksi fitur menggunakan algoritma Na \ddot{i} ve Bayes dilakukan dengan cara merangking variabel yang terpilih. Variabel yang terpilih akan diuji menggunakan algoritma K- Nearest Neighbor dengan menghitung D (x,y) untuk setiap data yang disimpan. Langkah terakhir pada proses hybrid ini yaitu dengan menentukan urutan nilai minimum D(x,y) pada hasil perhitungan (Safri et al., 2018).

Pada proses klasifikasi di dalam algoritma Naïve Bayes dengan menggunakan data kontinyu maka digunakan rumus densitas gauss(Gultom et al., 2019)pada persamaan2.5 sebagai berikut:

$$P(X_i) = \frac{1}{\sqrt{2\pi\sigma}} e^{\frac{(x_i - \mu)^2}{2\sigma^2}}$$
(2.5)

Keterangan:

P= peluang

 X_i = atribut ke-i

 x_i = nilai Atribut ke-i

 μ = mean (rata-rata dari seluruh atribut)

 σ = standar deviasi (varian dari seluruh atribut)

Persamaan2.6 untuk mencari nilai rata-rata hitung (mean) sebagai berikut:

$$\mu = \frac{\sum_{i=1}^{n} x_i}{n} \tag{2.6}$$

Keterangan:

 μ = nilai rata-rata hitung (mean)

 $x_i = \text{nilai x ke-i}$

n = jumlah sampel

sedangkan Persamaan2.7 untuk menghitung nilai simpangan baku (standar deviasi) sebagai berikut:

$$\sigma = \frac{\sqrt{\sum_{i=1}^{n} x_i \mu}}{n-1} \tag{2.7}$$

Keterangan:

 σ = standar deviasi

 x_i = nilai x ke-i

 μ = nilai rata-rata hitung (mean)

n= jumlah sampel

Pada proses menghitung probabilitas pada data kontinyu tahap pertama adalah dengan menghitung nilai mean dan standar deviasi dari masing-masing parameter yang merupakan data numerik, kemudian setelah didapatkan nilai mean dan standar deviasi maka nilai probabilitas akan diperoleh dengan menggunakan persamaan 2.7. Selanjutnya akan dilakukan seleksi fitur dengan memilih nilai probabilitas lebih besar atau sama dengan nilai alpha yang diinginkan (Safri et al., 2018). Selanjutnya diambil fitur minimal yang dapat membantu dalam meringankan proses klasifikasi pada *K-Nearest Neighbor*(KNN). Fitur yang diambil berdasarkan rangking dari nilai probabilitas (Syafitri Hidayatul AA, Yuita Arum S, 2018).

2.8. K-Nearest Neighbor(KNN)

K-Nearest Neighbor dipandang sebagai salah satu dari sepuluh algoritma teratas dalam data mining karena memiliki keunggulan dari implementasi yang sederhana, efektifitas tinggi, dan intuisi yang mudah(Gou et al., 2019). Ide dasar dari *K-Nearest Neighbor* adalah bahwa sampel kueri termasuk dalam kelas dengan frekuensi paling banyak diantara K-tetangga terdekat. Tahapan-tahapan pada algoritma *K-Nearest Neighbor* menurut(Alghifari and Wibowo, 2019) ada tiga yaitu sebagai berikut:

- 1. Menghitung jarak dari data yang masuk dari semua catatan pelatihan.
- Mengatur catatan pelatihan berdasarkan jarak dan pemilihan K-tetangga terdekat.
- 3. Menggunakan kelas yang memiliki mayoritas diantara K-tetangga terdekat.

Klasifikasi berbasis *K-Nearest Neighbor* memiliki kinerja yang baik dalam banyak aplikasi praktis pengenalan pola. Masalah yang terkait dalam algoritma KNN adalah terkait dengan penentuan nilai K. Cara menentukan nilai K yang tepat merupakan peran yang penting dalam melakukan klasifikasi berbasis *K-Nearest Neighbor*, karena dapat mempengaruhi *sensitivitas* kinerja sistem itu sendiri. Secara umum prinsip kerja algoritma *K-Nearest Neighbor* adalah mencari jarak terdekat antar data dengan melakukan evaluasi terhadap tetangga terdekat dalam pelatihan(Nikmatun and Waspada, 2019). Untuk menghitung jarak dapat menggunakan rumus *euclidean distance* pada Persamaan 2.8 sebagai berikut.

$$d_i = \sqrt{\sum_{i=1}^p (X_{2i} - X_{1i})^2}$$
 (2.8)

Keterangan:

 X_1 = data latih

 X_2 = data uji

d = jarak

i = variabel data

p = dimensi data

2.9. Confusion Matrix

Confusion matrix merupakan suatu cara dalam melakukan analisis kinerja atau dalam melakukan perhitungan akurasi pada konsep data mining salah satunya pada proses klasifikasi(Nengsih, 2017). Konsep dasar confusion matrix adalah membandingkan hasil klasifikasi yang telah dilakukan oleh sistem dengan hasil klasifikasi yang seharusnya. Ilustrasi confusion matrix ditunjukkan pada Gambar 2.4

K L C L	Kelas Has	Kelas Hasil Klasifikasi		
Kelas Sebenarnya	Positive	Negative		
Positive	True Positive	False Negative		
	(TP)	(FN)		
Negative	False Positive	True Negative		
	(FP)	(TN)		

Gambar 2.4 Confusion Matrix

(sumber: Suniantara et al., 2020)

Keterangan:

TN = Jumlah kelas DBD yang terklasifikasi benar pada kelas DBD

TP = Jumlah kelas normal yang terklasifikasi benar pada kelas normal

FN = Jumlah kelas DBD yang terklasifikasi salah

Berdasarkan nilai-nilai True Negative (TN), True Positive (TP), False Negative (FN), dan False Positive (FP), evaluasi klasifikasi dianalisis dengan indikator akurasi. Akurasi merupakan rasio antara jumlah terprediksi benar dari semua data (Rivanthio et al., 2020).Maka rumus untuk menghitung akurasi klasifikasi pada Persamaan 2.9, Persamaan 2.10, dan Persamaan 2.11 sebagai berikut:

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN}$$
 (2.9)

$$Sensitifitas = \frac{TP}{TP + FN} \tag{2.10}$$

$$Spesifitas = \frac{TN}{TN + FP} \tag{2.11}$$

2.10. Integrasi Keilmuan

Faktor utama yang dapat mempengaruhi kesehatan adalah kebersihan. Banyak penyakit muncul akibat dari kebiasaan buruk manusia yang kurang memperhatikan kebersihannya. Dalam Islam telah dijelaskan mengenai kebersihan bahwa orang yang beriman adalah orang yang dapat menjaga kebersihan, mulai dari kebersihan diri sendiri sampai kebersihan lingkungan.Pentingnya menjaga kebersihan sebelumnya telah dijelaskan pada firman Allah didalam surat al-Maidah ayat 6 sebagai berikut:

Artinya :"Hai orang-orang yang beriman, apabila kamu hendak mengerjakan shalat, maka basuhlah mukamu dan tanganmu sampai dengan siku,

لِأَيُّهَا الَّذِيْنَ الْمَنُوْ آ اِذَا قُمْتُمْ اِلَى الصَّلُو قِ فَا غُسِلُوْ ا وُجُوْ هَكُمْ وَ اَيْدِكُمْ اِلَى الْمَرَا فِقِ وَامْسَحُوْ ا بِرُ ءُوْسِكُمْ وَ اَرْجُلَكُمْ اِلَى الْكَعْبَيْنِ ۗ وَانْ كُنْتُمْ جُنْبًا فَا طَّهَرُوْ ا ۗ وَ اِنْ كُنْتُمْ مَرْضَلَى اَوْ عَلَى سَفَرٍ اَوْ جَاءَ اَحَدٌ مِّنْكُمْ مِّنَ الْغَا بِطِ اَوْ لَمَسْتُمُ النِّسَاءَ فَلَمْ تَجِدُوْ ا مَاءً قَتَيَمَمُوْ ا صَعِيْدًا طَيِّبًا فَا مُسَحُوْ ا بِوُجُوْ هِكُمْ وَ اَيْدِ يُكُمْ مِّنْ هُ مِّنَ الْعَلَيْ لِيَجْعَلَ مَا يُرِيْدُ اللَّهُ لِيَجْعَلَ عَلَيْكُمْ مِّنْ حَرَج وَلْكِنْ يُرِيْدُ لِيُطَهِّرَ كُمْ وَ لِيُتِمَّ نِعْمَتَهُ عَلَيْكُمْ لَعَلَّكُمْ تَشْكُرُوْنَ ﴿ 6 ﴾ عَلَيْكُمْ مِّنْ حَرَج وَلْكِنْ يُرِيْدُ لِيُطَهِّرَ كُمْ وَ لِيُتِمَّ نِعْمَتَهُ عَلَيْكُمْ لَعَلَّكُمْ تَشْكُرُوْنَ ﴿ 6 ﴾

dan sapulah kepalamu dan (basuh) kakimu sampai dengan kedua mata kaki, dan jika kamu junub maka mandilah, dan jika kamu sakit atau dalam perjalanan atau kembali dari tempat buang air (kakus) atau menyentuh perempuan, lalu kamu tidak memperoleh air, maka bertayammumlah dengan tanah yang baik (bersih); sapulah mukamu dan tanganmu dengan tanah itu. Allah tidak hendak menyulitkan kamu, tetapi Dia hendak membersihkan kamu dan menyempurnakan nikmat-Nya bagimu, supaya kamu bersyukur".

Firman di atas telah menjelaskan bahwa Allah SWT tidak pernah mempersulit hambanya untuk bersuci dan menjaga kebersihan diri sendiri. Cara untuk bersuci juga telah dijelaskan pada firman di atas, mulai cara bersuci dari hadast kecil sampai hadast besar. Penyebab Islam memudahkan umatnya dalam bersuci adalah karena kebersihan merupakan salah satu hal sangat penting. Dalam Islam juga mengajarkan bahwa kebersihan lingkungan sekitar juga penting demi tetap menjaga kesucian tempat tersebut . Seperti yang telah dijelaskan pada firman Allah SWT didalam surat al-Baqarah ayat 125 sebagai berikut:

وَإِذْ جَعَلْنَا الْبَيْتَ مَثَا بَهُ لِّنَا سِ وَأَمْنًا وَتَّذِذُ وْا مِنْ مَّقَامِ إِبْرَاهِمَ مُصلِّى صُوعَهِدْ نَآلِكَى إِبْرَاهِمَ وَ إِسْمُعِيْلَ أَنْ طَهِرَا بَيْتِيَ لِطَّآ نِفِيْنَ وَالْعُكِفِيْنَ وَالرُّكَّعِ السُّجُوْدِ Artinya:"Dan (ingatlah), ketika Kami menjadikan rumah itu (Baitullah) tempat berkumpul bagi manusia dan tempat yang aman. Dan jadikanlah sebagian maqam Ibrahim tempat shalat. Dan telah Kami perintahkan kepada Ibrahim dan Ismail: "Bersihkanlah rumah-Ku untuk orang-orang yang thawaf, yang i'tikaf, yang ruku' dan yang sujud".

Manfaat dari menjaga kebersihan salah satunya adalah dapat menjauhkan diri dari virus dan bakteri karena penyakit bisa timbul dari lingkungan yang kotor. Penyakit merupakan suatu musibah yang dapat menimpa kapan saja. Macam-macam penyakit mulai dari penyakit menular, penyakit keturunan, penyakit mematikan, penyakit musiman dan masih banyak lagi. Allah menciptakan penyakit kepada umatnya tentu bukan tanpa alasan, segala bentuk ujian yang diturunkan akan menjadi tolak ukur tingkat keimanan manusia kepada Allah SWT. Seperti yang tertulis pada Al-Quran surat Al-Mulk ayat 2 sebagai berikut:

Artinya: "Yang menjadikan mati dan hidup, supaya Dia menguji kamu, siapa di antara kamu yang lebih baik amalnya. Dan Dia Maha Perkasa lagi Maha Pengampun".

Penyakit demam berdarah contohnya yang merupakan salah satu penyakit yang disebabkan oleh perantara nyamuk. Perantara dari penyakit tersebut adalah gigitan seekor nyamuk Aedes . Dalam Islam, nyamuk merupakan serangga yang membahayakan, memiliki dua sayap, mulut yang berbentuk seperti jarum yang digunakan untuk menggigit dan menghisap serta dapat menularkan penyakit

seperti demam berdarah. Allah berfirman dalam surat Al-Baqarah ayat 26 sebagai berikut:

Artinya:"Sesunggahnya Allah tiada segan membuat perumpamaan berupa nyamuk atau yang lebih rendah dari itu. Adapun orang-orang yang beriman, maka mereka yakin bahwa perumpamaan itu benar dari Tuhan mereka, tetapi mereka yang kafir mengatakan:"Apakah maksud Allah menjadikan ini untuk perumpamaan?". Dengan perumpamaan itu banyak orang yang disesatkan Allah, dan dengan perumpamaan itu (pula) banyak orang yang diberi-Nya petunjuk. Dan tidak ada yang disesatkan Allah kecuali orang-orang yang fasik.

Pada firman di atas dapat ditafsirkan bahwa adanya perumpaan sebuah nyamuk adalah untuk melihat tingkat keimanan seseorang. Orang-orang yang beriman akan yakin bahwa serangga kecil berupa nyamuk diciptakan oleh Allah dan tentu seluruh ciptaan Allah memiliki hikmah yang dapat dijadikan pelajaran. Keyakinan tersebut tentunya berhubungan dengan adanya penyakit DBD yang ditularkan oleh seekor nyamuk. Jika termasuk orang yang beriman maka seseorang akan berusaha untuk menanggulanginya dengan cara menjaga kebersihan, berusaha menemukan solusi agar terhindar dari penyakit DBD. Allah bahkan akan memberikan petunjuk kepada orang-orang yang beriman, dalam kasus ini orang yang diberi petunjuk seperti ide untuk melakukan klasifikasi penyakit DBD agar dapat memudahkan petugas medis dalam mendiagnosa penyakit demam berdarah. Berbeda dengan orang-orang yang tidak mempercayai adanya ciptaan Allah berupah nyamuk maka

akan disesatkan jalannya oleh Allah. Kata disesatkan memiliki arti bahwa seseorang tidak akan pernah sadar akan adanya penyakit DBD dan cara menanggulanginya yang akan berdampak pada kesehatannya karena terserang penyakit DBD itu sendiri. *Rasulullah shallallahualaihi wa sallam* bersabda didalam hadist-hadist berikut:

Artinya:"Banyak manusia merugi karena dua nikmat yaitu kesehatan dan waktu luang" (HR. Bukhari dari Ibnu Abbas).

Allah menurunkan ujian kepada hamba-Nya dengan dua bentuk yaitu ujian berupa kesenangan atau nikmat dan ujian berupa kesengsaraan atau niqmat. Penyakit merupakan salah satu bentuk ujian berupa kesengsaraan dari Allah SWT yang tidak dapat dihindari oleh siapapun(Kaltsum, 2018). Semua penyakit pasti ada obatnya sesuai dengan firman Allah dalam surat Al-Isra ayat 82 Allah SWT bahwa Alquran diturunkan untuk menjadi penawar dan rahmat bagi orang-orang yang beriman(Hasan, 2020). Bunyi dari firman Allah SWT dalam surat al-Isra ayat 82 sebagai berikut:

Artinya:"Dan Kami turunkan dari Alquran suatu yang menjadi penawar dan rahmat bagi orang-orang yang beriman dan Al-quran itu tidaklah menambah kepada orang-orang yang zalim selain kerugian".

Terdapat kata Penawar dalam firman Allah SWT yang dimaksudkan sebagai obat penyembuh untuk segala penyakit atau dalam bahasa arab disebut sebagai asy-Syifa. Dapat diartikan bahwa obat penawar untuk segala penyakit pasti ada(Rahman, F., Ihsan, M., Pristianto, A., Khadijah, S., & Budi, 2019). Allah SWT menurunkan musibah bersamaan dengan cara penyelesaiannya seperti yang diterangkan oleh Allah SWT didalam surah Asy-Syu'ara' ayat 80 sebagai berikut:

Artinya:"Dan apabila aku sakit, Dia-lah yang menyembuhkan aku".

Pada firman Allah SWT yang terdapat pada ayat di atas bertujuan untuk memberi seruan kepada umat manusia agar selalu berusaha dan berdoa karena segala penyakit dapat disembuhkan selama tetap berusaha dalam mencari kesembuhan. Firman Allah secara tidak langsung memberi petunjuk bahwa setiap penyakit pasti ada obatnya, ada upaya yang dapat dilakukan untuk menangkal dan menyembuhkannya. Setiap penyakit mempunyai karakteristik masing-masing dengan tingkat keparahan yang berbeda-beda. Sama dengan usaha manusia yang memiliki tingkatan ikhtiar masing-masing sesuai dengan cobaan yang dihadapi. Usaha yang dilakukan sebagai bentuk ikhtiar manusia merupakan bentuk kesabaran yang selalu dianjurkan dalam menghadapi musibah.

Berdasarkan firman Allah SWT dan hadist yang telah disertakan dapat diambil pelajaran bahwa rasa syukur harus selalu bertambah. Jika manusia dapat menanamkan rasa syukur dengan begitu besar maka akan muncul rasa menghargai terhadap suatu hal. Ketika rasa menghargai telah tertanam maka rasa untuk menjaga dan mencintai apa yang Allah SWT ciptakan akan muncul dengan

sendirinya. Hal tersebut berkaitan dengan penelitian kali ini yaitu tentang suatu penyakit yang disebabkan oleh perilaku manusia yang kurang mencintai diri sendiri dan juga lingkungannya sehingga dapat memicu terjadinya suatu kerugian dalam dirinya. Kerugian yang dimaksud adalah terserangnya virus demam berdarah dengue atau yang dikenal dengan DBD.



BAB III

METODE PENELITIAN

3.1. Jenis Penelitian

Jenis penelitian pada penelitian ini menggunakan jenis penelitian terapan dengan pendekatan kuantitatif. Penelitian terapan merupakan suatu penelitian yang dilakukan dengan tujuan memperoleh jawaban terhadap permasalahan dengan praktis. Penelitian ini tidak mengembangkan sebuah ide, gagasan atau teori tetapi membahas suatu permasalahan yang kerap terjadi di kehidupan sehari-hari. Hubungan terhadap penelitian ini yaitu memberikan solusi kepada petugas medis dalam memanfaatkan proses klasifikasi untuk diagnosa penyakit Demam Berdarah Dengue (DBD).

3.2. Pengumpulan Data

Sumber data pada penelitian ini merupakan data sekunder dari 240 pasien yang didapat dari data tes hematologi di Puskesmas Temindung Kota Samarinda, Kalimantan Timur tahun 2018 (Destiana et al., 2019). Sumber data yang diperoleh sebanyak seratus data yang terdiri dari lima atribut penunjang keputusan pada kelas atau yang disebut atribut umum dan satu label atau kelas yang terdiri dari 133 kelas positif demam berdarah dan 107 kelas negatif demam berdarah.

3.3. Variabel Penelitian

Variabel penelitian merupakan suatu parameter yang berfungsi sebagai objek penelitian sehingga dapat menghasilkan sebuah informasi dan dapat ditarik

kesimpulannya. Jenis variabel pada penelitian ini dibagi menjadi dua macam diantaranya:

1. Variabel bebas (*Independent*)

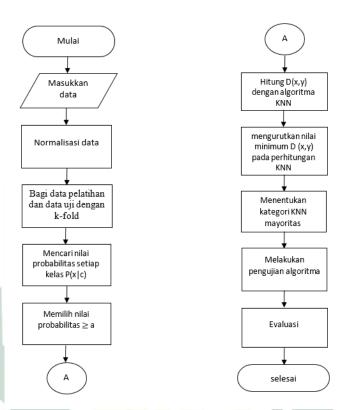
Variabel bebas adalah variabel yang mempengaruhi pada variabel respon (dependent). Variabel bebas pada penelitian ini terdiri dari:

- **a.** Usia yang merupakan variabel numerik dengan rentang usia pasien 1-90 tahun.
- b. Kadar Hb atau Hemoglobin merupakan variabel numerik dengan rentang 10-20 gram.
- c. Leukosit atau sel darah putih merupakan variabel numerik dengan rentang 0.1-20 mg/dl.
- d. Trombosit atau sel darah merah merupakan variabel numerik dengan rentang 1-500 mg/dl.
- e. Kadar Ht atau Hematokrit merupakan variabrel numerik untuk kadar sel darah merah dalam darah yang memiliki rentang 20-50 persen.

2. Variabel Respon (*Dependent*)

Variabel respon adalah variabel yang dipengaruhi oleh variabel bebas (*independent*). Variabel respon pada penelitian ini adalah diagnosa status pasien yang merupakan variabel numerik dengan perumpamaan nilai 1 sebagai pasien positif DBD dan nilai 0 sebagai pasien negatif penyakit DBD.

3.4. Tahapan Penelitian



Gambar 3.1 Diagram alir metode hybrid Nave Bayes-K Nearest Neighbor

Tahapan penelitian sesuai dengan diagram alir pada Gambar 3.1 diatas adalah sebagai berikut:

1. Data

Input data yang telah diperoleh sebanyak 240 data yang didapat dari data pasien tes hematologi di Puskesmas Temindung kota Samarinda, Kalimantan Timur tahun 2018 terdiri dari lima variabel penunjang keputusan pada kelas atau yang disebut variabel umum dan satu label atau kelas.

2. Normalisasi data

Normalisasi data dilakukan untuk merubah rentang agar menjadi lebih kecil.

Pada penelitian ini proses normalisasi data dilakukan menggunakan metode *Z-score Normalization* pada Persamaan ??.

3. Pembagian Data

Pembagian data dalam penelitian ini adalah dengan memanfaatkan model *k-fold cross validation*.

4. Hitung Probabilitas

Penentuan probabilitas akan digunakan metode Na*i*ve Bayes dengan menggunakan rumus pada persamaan 2.5.

5. Metode Naïve Bayes

Setelah nilai probabilitas setiap variabel diketahui tahap selanjutnya adalah dengan melakukan pemilihan nilai probabilitas yang memiliki nilai nilai probabilitas minimum (alpha) yang ditentukan yaitu sebesar 0.30 dengan jumlah fitur sebanyak 3 yang bertujuan untuk memperoleh variabel hasil seleksi fitur pada metode Naïve Bayes setelah itu akan dilakukan perhitungan selanjutnya pada algoritma K Nearest Neighbor .

6. Perhitungan algoritma *K-Nearest Neighbor*

Dalam tahap ini setiap data akan dilakukan perhitungan menggunakan algoritma K Nearest Neighbor dengan menggunakan rumus euclidean distance pada persamaan2.8. Pada Penelitian ini akan dilakukan uji coba nilai K terbaik yaitu dengan nilai k=3, k=5, k=7 dan k=9. Langkah selanjutnya yaitu dengan menghitung jarak antara data baru dan semua data yang ada di dalam data pelatihan.

7. Menentukan kategori KNN mayoritas

Setelah melakukan proses perhitungan kemudian data diproses untuk

mendapatkan nilai KNN mayoritas.

8. Melakukan pengujian algoritma

Pada tahap ini akan diperoleh hasil klasifikasi dengan metode yang digunakan yaitu metode hybrid Naïve Bayes-*K Nearest Neighbor* .

9. Evaluasi

Tahap ini merupakan tahap terakhir yang akan bertujuan untuk menghitung nilai akurasi, sensitifitas, dan spesifisitas menggunakan *confusion matrix* pada Persamaan 2.9, 2.10, 2.11 untuk mendapatkan sistem klasifikasi dengan performa terbaik diantara hybrid Naïve Bayes-*K Nearest Neighbor*, metode KNN dan hybrid PCA- *K Nearest Neighbor*.

BAB IV

HASIL DAN PEMBAHASAN

4.1. Pengelolahan Data

Data pada penelitian ini terdiri dari data Usia, Kadar Hemoglobin, Leukosit, Trombosit, kadar Hematokrit. Data-data tersebut merupakan data sekunder yang didapatkan dari data tes hematologi di Puskesmas Temindung kota Samarinda, Klimantan Timur tahun 2018. Data yang diperoleh dari 240 pasien dengan lima atribut penunjang keputusan pada kelas dan satu label atau kelas yang terdiri dari 133 kelas positif dan 107 kelas negatif. Sampel data ditunjukkan pada Tabel 4.1 sebagai berikut:

Tabel 4.1 Sampel data penelitian

Data ke-	Usia	Hemoglobin	Leukosit	Trombosit	Hematokrit	Label
1.	12	14.3	1.09	35.49	44.52	Negatif
2.	6	15.4	1.075	46.1	47.4	Negatif
3.	43	14.6	0.535	28.3	46.8	Negatif
4.	21	14.9	0.377	8.4	42.5	Positif
5.	32	17.6	1.102	33.8	54.99	Negatif
6.	5	14.5	0.591	24.7	45.58	Negatif
7.	10	15	0.31	8.5	42.3	Positif
:	:	:	:	:	:	:
239	2	11.5	6.2	487	36	Negatif
240	41	17.4	1.92	653	32	Negatif

4.2. Normalisasi Data

Normalisasi data dilakukan dengan merubah rentang data menjadi lebi kecil. Pada penelitian ini proses normalisasi data dilakukan menggunakan metode *Z-score* *Normalization*, sehingga hasil normalisasi data didapat dari Persamaan 2.1 sebagai berikut:

$$X_{Usia(i)}^{'} = \frac{12 - 25.83}{58.61} = -0.246$$

Dari hasil contoh perhitungan diatas, didapat data dengan skala 0 sampai 1 yang ditampilkan pada tabel 4.2.

			U . A			
No. Data	Usia	Hemoglobin	Leukosit	Trombosit	Hematokrit	Label
1.	-0.2467	0.50521	-0.7074	-0.5776	0.8287	Negatif
2.	-0.5666	1.0439	-0.7112	-0.4911	1.3218	Negatif
3.	1.4063	0.6521	-0.8461	-0.6363	1.2191	Negatif
4.	0.2332	0.7990	-0.8855	-0.7986	0.4828	Positif
5.	0.8197	2.1213	-0.70 <mark>44</mark>	-0.5914	2.6213	Negatif
6.	-0.6199	0.6031	-0.8321	-0. 6657	1.0102	Negatif
7.	-0.3533	0.8480	-0.9023	<mark>-0.7</mark> 978	0.4486	Positif
:	:	:	:	:	: /	:
239.	2.5261	0.5052	-0 .7 <mark>53</mark> 9	-0.5 645	1.0650	Negatif
240.	1.2997	2.0233	-0.5001	-0.5922	2.5032	Negatif

Tabel 4.2 Data sesudah dinormalisasi

4.3. Metode Naïve Bayes untuk Data Numerik

Mencari mean dan standart deviasi untuk masing- masing variabel menggunakan rumus pada Persamaan 2.6 dan 2.7 sebagai berikut:

Kelas negatif

$$Mean_{(Usia)} = \frac{-02467 + -0.5666 + \dots + 1.2997}{107}$$
$$= \frac{25.3103}{107}$$
$$= 0.2365$$

$$StandardDeviasi_{(Usia)} = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \mu)^2}{n - 1}}$$

$$= \sqrt{\frac{(-0.2467 - 0.2365)^2 + \dots + (1.2997 - 0.2365)^2}{107 - 1}}$$

$$= \sqrt{\frac{(-0.2467 - 0.2365)^2 + \dots + (1.2997 - 0.2365)^2}{106}}$$

$$= \sqrt{\frac{161.2749}{106}}$$

$$= 1.233447$$

Kelas positif

$$Mean_{(Usia)} = \frac{-0.2467 + -0.5666 + \dots + 1.2997}{133}$$
$$= \frac{-25.3103}{133}$$
$$= -0.1903$$

$$StandardDeviasi_{(Usia)} = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \mu)^2}{n - 1}}$$

$$= \sqrt{\frac{(0.2332 - (-0.1903))^2 + \dots + (1.29970 - (-0.1903))^2}{133 - 1}}$$

$$= \sqrt{\frac{(0.2332 - (-0.1903))^2 + \dots + (1.29970 - (-0.1903))^2}{132}}$$

$$= \sqrt{\frac{71.8630}{132}}$$

$$= 0.7120$$

Dari hasil contoh perhitungan diatas, didapat nilai mean dan standart deviasi dari masing-masing kategori yang ditampilkan pada Tabel 4.3.

Tabel 4.3 Mean da<mark>n standart deviasi</mark> berdas<mark>ark</mark>an pengolahan data

kategori	No. Data	Usia	Hemoglobin	Leukosit	Trombosit	Hematokrit
Negatif	Mean	0.23654	-0.11586	-0.13945	-0.01952	-0.01952
Negatii	Standart Devisai	1.23347	0.94224	1.15743	1.15743	1.00676
Positif	Mean	-0.19030	0.09321	0.11219	0.01570	0.01570
Positif	Standart Deviasi	0.71202	1.03827	0.84046	0.84046	0.99805

4.4. Menghitung Nilai Probabilitas Tiap Data

Menghitung nilai probabilitas setiap data menggunakan rumus pada Persamaan 2.5 sebagai berikut:

$$P_{(Usia)} = \frac{1}{\sqrt{2 \times 3.14 \times 1.2334}} e^{\frac{(-0.2467 - 0.2365)^2}{2 \times (1.2334)^2}}$$
$$= 0.3081$$

Dari hasil contoh perhitungan di atas, didapat nilai probabilitas setiap data

yang ditampilkan pada Tabel 4.4.

Tabel 4.4 Nilai probabilitas setiap data

Usia	Hemoglobin	Leukosit	Trombosit	Hematokrit	Label
0.1821	0.2577	0.0396	0.0021	0.3500	negatif
0.2858	0.3479	0.3492	0.3526	0.3466	negatif
0.1952	0.3591	0.0934	0.0072	0.3591	negatif
0.4311	0.4696	0.4714	0.3177	0.4724	positif
0.4334	0.4557	0.4221	0.2082	0.4370	positif
:	:		:	:	:
0.0189	0.3875	0.3913	0.2670	0.2205	Positif
0.2681	0.0345	0.4689	0.4355	0.1990	positif
0.2681	0.0345	0.4689	0.4355	0.1990	positif

4.5. Metode Naïve Bayes K-Nearest Neighbor

Dilakukan seleksi fitur menggunakan naive bayes dengan nilai probabilitas minimum adalah 0.30. Dalam melakukan seleksi fitur pada Naive Bayes dilakukan dengan cara memilih probabilitas lebih besar dari 0.30 contoh seperti Tabel 4.5 berikut ini.

Tabel 4.5 Seleksi fitur menggunakan Naïve Bayes

Usia	Hemoglobin	Leukosit	Trombosit	Hematokrit
0.1821	0.2577	0.0396	0.0021	0.3500
0.2858	0.3479	0.3492	0.3526	0.3466
0.1952	0.3591	0.0934	0.0072	0.3591
0.4311	0.4696	0.4714	0.3177	0.4724
0.4334	0.4557	0.4221	0.2082	0.4370
:	:	:	:	÷
0.0189	0.3875	0.3913	0.2670	0.2205
0.2681	0.0345	0.4689	0.4355	0.1990

Berdasarkan Tabel 4.5 seleksi fitur pada Naïve Bayes dengan memakai jumlah fitur sebanyak 3 maka dari itu dapat diketahui bahwa variabel hemoglobin, leukosit dan hematokrit memiliki probabilitas lebih besar atau sama dengan 0.3

lebih banyak dibandingkan dengan variabel lainnya. Dengan rincian variabel usia memiliki probabilitas yang memenuhi=83, hemoglobin=92, Leukosit=84, Trombosit=83, dan Hematokrit=91. Hasil dari subset fitur menggunakan metode Naïve Bayes dapat dilihat pada Tabel 4.6 sebagai berikut:

Tabel 4.6 Subset fitur yang terpilih menggunakan Naïve Bayes

Hemoglobin	Leukosit	Hematokrit
0.2577	0.0396	0.3500
0.3479	0.3492	0.3466
0.3591	0.0934	0.3591
0.4696	0.4714	0.4724
0.4557	0.4221	0.4370
/	:	
0.3875	0.3913	0.2205
0.0345	0.4689	0.1990

Setelah mengetahui variabel yang akan digunakan maka langkah selanjutnya adalah mencari nilai K mayoritas pada metode K-Nearest Neighbor. Setelah mendapatkan data training maka langkah selanjutnya adalah dengan melakukan pengujian menggunakan sampel menggunakan algoritma K-Nearest Neighbor untuk mendapatkan nilai K mayoritas pada tabel 4.6 Contoh perhitungan menggunakan data testing dengan nilai Hemoglobin=-0.8660, leukosit =0.5690, hematokrit= -0.6300 dengan menggunakan rumus *euclidien distance* sebagai berikut

$$d_{215} = \sqrt{((-0.131 - (-0.866))^2 + (-0.230 - 0.569)^2 + (-0.167 - (-0.630))^2}$$

$$= \sqrt{1.392}$$

$$= 1.18$$

Dari hasil contoh perhitungan diatas, didapat nilai euclidien setiap data dan K-mayoritas pada algoritma KNN yang ditampilkan pada Tabel 4.7.

Tabel 4.7 Uji data testing dalam menentukan K mayoritas

No. Data	Distance
1.	2.3744
2.	3.0161
3.	2.7797
4.	2.4753
5.	4.5954
6.	2.6100
7.	2.5033
4	
215.	1.1800
216.	1.1602

Berdasarkan tabel di atas dengan nilai K=3 maka yang diambil adalah 3 tetangga terdekat. Pada data ini diperoleh 3 jarak terdekat terdapat pada data ke 111, 112 dan 210 dan didapatkan label pada masing-masing indeks yaitu 0,1,0. Proses selanjutnya yaitu akan diambil jumlah label yang memiliki frekuensi yang mayoritas, sehingga dapat disimpulkan bahwa kelas dari data testing adalah kelas 0 atau negatif.

4.6. Pengujian Algoritma

Pengujian menggunakan metode hybrid Naïve-KNN adalam penelitian ini dilakukan dengan bantuan software komputer dan diperoleh hasil seperti pada Tabel 4.8 berikut:

Tabel 4.8 Tabel Hasil Klasifikasi Hybrid Naüve-KNN

K	k-fold	Data Latih	Data Uji	Akurasi	Sensitif	spesifitas
	2	120	120	73.3%	75.8%	70.4%
	4	181	59	78.0%	81.8%	73.1%
3	6	200	40	82.5%	90.9%	70.2%
3	8	209	31	87.1%	88.2%	85.7%
	10	216	24	95.4%	96.2%	94.4%
	Average	e		83.26%	86.58%	78.76%
	2	120	120	72.5%	75.8%	68.5%
	4	181	59	79.7%	81.8%	76.9%
5	6	200	40	80.0%	86.4%	72.2%
	8	211	29	83.9%	88.2%	78.6%
	10	216	24	87.5%	92.3%	81.8%
	Average			78.0%	84.90%	75.60%
	2	121	119	67.2%	74.2%	58.5%
	4	180	60	7 <mark>3.3</mark> %	78.8%	66.7%
7	6	200	40	75.0%	77.3%	72.2%
	8	210	30	7 6.7%	88.2%	61.2%
	10	216	24	79.2%	84.6%	72.7%
	Average	e		74.18%	80.62%	66.26%
	2	121	119	70.0%	73.1%	66.0%
	4	180	60	71.7%	72.7%	70.4%
9	6	200	40	72.5%	77.3%	66.7%
	8	210	30	73.3%	82.4%	61.5%
	10	217	23	78.3%	84.6%	70.0%
	Average	e		71.52%	78.02%	66.92%

Berdasarkan pengujian algoritma yang dapat dapat dilihat pada Tabel 4.8 bahwa nilai K terbaik pada algoritma hybrid Naïve- KNN adalah K=3 dengan nilai *k-fold cross validation*= 10. Dari hasil perhitungan diatas, didapat nilai akurasi, sensitifitas dan spesifitas pada *confusion matrix* pada Tabel 4.9 berikut ini :

Tabel 4.9 Tabel Hasil Evaluasi metode Hybrid Naëve-KNN

Kelas Sebenarnya	Kelas Hasil Klasifikasi		
Keias Sebellarilya	Positif DBD	Negatif DBD	
Positif DBD	101	5	
Negatif DBD	6	128	

4.7. Evaluasi Model

Evaluasi model ini dilakukan untuk mengetahui performa paling baik dari ketiga metode sebagai berikut:

4.7.1. Metode K-Nearest Neighbor

Berdasarkan pengujian algoritma dengan nilai K terbaik pada algoritma hybrid Naïve- KNN adalah K=3 dengan nilai *k-fold cross validation*= 10, selanjutnya akan dilakukan evaluasi untuk metode KNN pada Tabel 4.10.Pada proses KNN jumlah fitur yang dipakai=5 dengan variabel Usia, Hemaglobin, Leukosit, Trombosit dan Hematokrit.

Tabel 4.10 Tabel Hasil Evaluasi metode KNN

Kelas Sebenarnya	Kelas Hasil Klasifikasi			
Kelas Sebellatilya	Positif DBD	Negatif DBD		
Positif DBD	109	24		
Negatif DBD	5	102		

4.7.2. PCA- K Nearest Neighbor

Berdasarkan pengujian algoritma dengan nilai K terbaik pada algoritma hybrid Naive- KNN adalah K=3 dengan nilai *k-fold cross validation*= 10, selanjutnya akan dilakukan evaluasi untuk metode KNN pada Tabel 4.11.KNN jumlah fitur yang dipakai=5 dengan variabel Usia, Hemaglobin, Leukosit, Trombosit dan Hematokrit.

Tabel 4.11 Tabel Hasil Evaluasi metode PCA-KNN

Kelas Sebenarnya	Kelas Hasil Klasifikasi		
Kelas Scocharnya	Positif DBD	Negatif DBD	
Positif DBD	117	16	
Negatif DBD	12	95	

4.7.3. Hybrid Naïve-K Nearest Neighbor

Berdasarkan evaluasi algoritma yang digunakan dalam metode Hybrid Na $\ddot{\nu}$ e-KNN, untuk klasifikasi diagnosa penyakit DBD dapat ditunjukkan bahwa metode Naive Bayes- KNN memiliki performa lebih baik dibandingkan metode K-NN, dan PCA-KNN dengan hasil evaluasi metode seperti pada Tabel 4.12

Tabel 4.12 Tabel Hasil Klasifikasi Hybrid Naüve-KNN, K-NN, dan PCA-KNN

Metode	K	k-fold	Akurasi	Sensitif	spesifitas
Nave Bayes-KNN	3	10	95.4%	96.2%	94.4%
KNN	3	10	95.3%	82.0%	87.9%
PCA-KNN	3	10	88.8%	88.0%	88.3%

Penelitian sebelumnya terkait dengan data yang digunakan pada penelitian ini pernah dilakukan oleh Reza Destiana, dkk(Destiana et al., 2019).Penelitian yang dilakukan tersebut adalah mengklasifikasikan penyakit Demam Berdarah Dengue (DBD) dengan menggunakan 5 variabel yaitu usia, hemoglobin, trombosit leukosit dan hematokrit menggunakkan metode Probabilistic Neural Network (PNN). dengan pembagian data training 90 dan jumlah data testing 10 diperoleh hasil akurasi sebesar 93%. Dapat ditunjukkan bahwa dengan menggunakan data yang sama hasil akurasi dari metode hybrid Naïve-*K Nearest Neighbor* lebih baik dibandingkan dengan metode PNN untuk klasifikasi penyakit DBD. Dalam penelitian ini juga dapat ditunjukkan bahwa dalam menyeleksi fitur *Principal Component Analysis*(PCA)terdapat beberapa kekurangan salah satunya adalah kurang optimal dalam pemisahan antar kelas. Pada *covariance* matriks pemilihan vektor yang tidak sesuai untuk menangkap fenomena tersebut dengan menggunakan log ratio dari *covariance* determinan. Selain itu, metode tersebut sangat sensitif sehingga akan menghasilkan penduga parameter berbias sehingga

dapat mempengaruhi akurasi yang dihasilkan(Sanusi and Saputro, 2020).

4.8. Integrasi Keilmuan

Dari hasil klasifikasi diagnosa penyakit Demam Berdarah Dengue (DBD) menggunakan metode hybrid Nave Bayes-K Nearest Neighbor dilakukan dengan beberapa uji coba agar mendapatkan akurasi yang tinggi. Dalam proses ini banyak pelajaran yang dapat diambil bahwa dalam mencapai suatu tujuan diperlukan usaha dan kemauan yang tinggi oleh setiap individu itu sendiri agar dapat merubah suatu hasil yang sesuai dengan keinginanya. Seperti yang sudah dijelaskan pada surat Al-Anfal ayat 53 sebagai berikut:

Artinya:"(Siksaan) yang demikian itu adalah karena sesungguhnya Allah sekali-kali tidak akan merubah sesuatu nikmat yang telah dianugerahkan-Nya kepada suatu kaum, hingga kaum itu meubah apa-apa yang ada pada diri mereka sendiri, dan sesungguhnya Allah Maha Mendengar lagi Maha Mengetahui".

Berdasarkan ayat di atas dijelaskan bahwa Allah tidak akan merubah suatu keadaan apabila kaum itu sendiri tidak mau merubahnya. Ayat tersebut menjanjikan suatu kenikmatan kepada seorang kaum apabila kaum tersebut bersedia sabar dan berusaha keras dalam proses merubah suatu keadaan sesuai yang diinginkan. Arti dari kesabaran dapat diartikan pada setiap cobaan yang dihadapi pasti akan ada solusi jika kaum tersebut selalu berusaha. Dalam penelitian ini salah satu usaha yang harus dilakukan adalah dengan cara meenjaga kebersihan diri agar terhindar dari penyakit demam berdarah. Kaidah fiqih yang menjelaskan tentang cara bersuci

adalah sebagai berikut:

حَدَّ ثَنَا الْمَكِّيُّ بْنُ إِبْرَا حِيْمَ: أَخْبَرَنَا عَبْدُاللَّهِ بْنُ سَعِيْدٍ, هُوَ ابْنُ أَبِي هِنْدٍ, عَنْ أَبِيْهِ, عَنِ ابْنِ عَبَّاسٍ رَضِيَ اللَّهُ عَنْهُمَا قَالَ: قَالَ النَّبِيُّ صلى الله عليه وسلم: نِعْمَتَانِ مَعْبُونٌ فِيْهُمَا كَثِيْرٌ مِنَ النَّسِ: الصِّحَةُ وَالْفَرَاغُ

وَكَيْفِيَّةُ غَسْلِ النَّجَاسَةِ إِنْ كَانَتْ مُشَاهَدَةً بِالْعَيْنِ وَهِيَ الْمُسَمَّاةُ بِالْعَيْنِيَّةِ تَكُوْنُ بِزَوَالِ عَيْنِهَا وَمُحَا وَلَةِ زَوَالِ أَوْصَا فِهَا مِنْ طَعْمِ أَوْ رِيْحٍ فَإِنْ بَقِيَ طَعْمُ النَّجَاسَةِ ضَرَّ أَوْ لَوْنٌ أَوْرِيْحٌ عَسُرَزَوَاللَهُ لَمْ يَضُرَّ. وَإِنْ كَانَتِ النَّجَا سَةَ غَيْرَ مُشَا هَدَةٍ وَهِيَ الْمُسَمَّاةُ بِاللَّكُمْيَّةِ فَيَكُفِىْ جَرْئُ الْمَاءِ عَلَى الْمُتَنَجَّس بِهَا وَلَوْ مَرَّةً وَاحِدَةً.

Artinya: "Caranya mencuci najis , apabila najisnya bisa diketahui oleh mata (bisa diketahui dengan panca indera), yaitu najis yang biasa disebut "najis 'ainiyah", adalah dengan menghilangkan keberadaannya itu sendiri, dan berusaha menghilangkan sifat-sifatnya benda, yaitu rasanya, warnanya, atau baunya. Maka apabila rasa najis itu masih ada, hukumnya benda itu masih tetap najis. Atau warna, bau benda najis tersebut masih tetap ada, hal mana sukar dihilangkan, maka tidak dianggap najis (hal itu sudah suci). Dan apabila najis tersebut tidak bisa dilihat (diraba panca indera), yaitu yang biasa disebut najis "hukmiyah", maka cukup mengalirkan (meratakan) air pada benda yang terkena najis tersebut, walaupun hanya satu kali saja".

Maksud dari kaidah di atas adalah salah satu cara untuk menjaga kebersihan yaitu dengan mensucikan diri dari najis. Najis mempunyai makna *al qadzarah* yang artinya kotoran sedangkan ulama Syafi'iyah mendefinisikan najis merupakan sesuatu yang terbilang kotor. Definisi tersebut menjelaskan bahwa

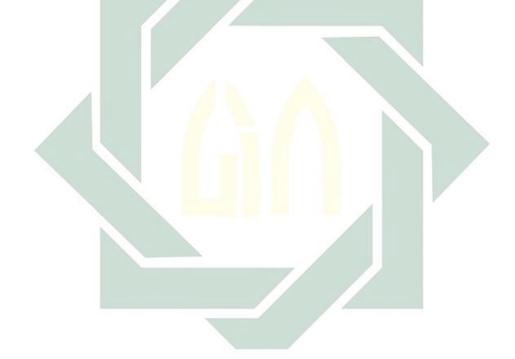
najis merupakan sesuatu yang kotor dan harus dibersihkan agar dapat menjaga kebersihan diri dan terhindar dari virus penyakit yang dapat timbul karena tingkat kebersihan seseorang salah satunya yaitu penyakit DBD. salah satu upaya dalam menjaga kebersihan diri sudah diterangkan dalam islam, tugas dari manusia sendiri adalah berusaha dan tidak berputus asa dalam menjalankan sesuatu seperti yang diterangkan pada surat Ghafir ayat 55 sebagai berikut:

Artinya: "Kami menyampaikan kabar gembira kepadamu dengan benar, maka janganlah kamu termasuk orang-orang yang berputus asa". Berdasarkan ayat di atas dijelaskan bahwa janji Allah pasti benar, maka dalam hal ini suatu kaum lah yang harus berusaha untuk merubah suatu keadaan sesuai yang diinginkan. Seperti halnya ujian suatu penyakit DBD yang saya bahas pada penelitian ini, yang telah dijelaskan bahwa Allah akan memberikan sesuai dengan apa yang kaum usahakan dalam kasus penyakit demam berdarah pada penelitian ini manusia harus berusaha dalam menjaga dan mencari solusi agar terhindar dari penyakit salah satunya yaitu dengan mempelajari ilmu yang telah didapatkan dalam ilmu sains dan teknologi untuk diimplementasikan pada kehidupan nyata seperti yang diterangkan pada surat At-Taubah ayat 105 sebagai berikut:

Artinya:"Dan Katakanlah: Bekerjalah kamu, maka Allah dan Rasul-Nya serta orang-orang mukmin akan melihat pekerjaanmu itu, dan kamu akan dikembalikan kepada (Allah) Yang Mengetahui akan yang ghaib dan yang nyata,

lalu diberitakan-Nya kepada kamu apa yang telah kamu kerjakan".

Berdasarkan ayat di atas dijelaskan bahwa dengan melakukan implementasi terhadap ilmu pendidikan, maka pekerjaan yang telah dilakukan akan dilihat dan memberikan manfaat oleh orang lain. Oleh sebab itu manfaatkan ilmu yang telah didapat sebaik mungkin, dalam kasus penelitian ini implementasi ilmu sains dan teknologi dimanfaatkan untuk mengklasifikasikan diagnosa penyakit Demm Berdarah Dengue (DBD).



BAB V

PENUTUP

Pada bab ini akan diberikan simpulan dan saran-saran yang dapat diambil berdasarkan materi-materi yang telah dibahas pada bab-bab sebelumnya.

5.1. Kesimpulan

Berdasarkan hasil pengelolahan data dengan memperhatikan tujuan penelitian maka didapat kesimpulan penelitian sebagai berikut:

- Hasil pengujian nilai K pada metode hybrid Naïve K Nearest Neighbor menunjukkan nilai K=3 dengan memperoleh akurasi yang baik sebesar 95.4% dengan sensitifitas sebesar 96.2% dan Spesifitas 94.4%.
- 2. Fitur yang dihasilkan oleh metode Naïve merupakan proses seleksi fitur yang menghasilkan hasil seleksi untuk variabel hemoglobin, leukosit, dan trombosit.
- 3. Hasil klasifikasi metode hybrid Naïve Bayes K Nearest Neighbor menunjukkan bahwa metode hybrid tersebut memiliki akurasi yang lebih tinggi dibandingkan dengan metode dasar yaitu K-NN yang memiliki akurasi sebesar 87.9% sensitifitas sebesar 82.0% dan spesifitas sebesar 95.3% Dalam Penelitian ini juga ditunjukkan bahwa Naïve Bayes memiliki performa lebih baik dalam melakukan seleksi fitur dibandingkan PCA. Hal tersebut dapat dibuktikan berdasarkan hasil klasifikasi metode hybrid Naïve Bayes K Nearest Neighbor memiliki akurasi yang lebih baik dibandingkan

dengan metode PCA-KNN dengan akurasi sebesar 88.3% sensitifitas sebesar 88.0% dan spesifitas sebesar 88.8%.

5.2. Saran

Penelitian selanjutnya diharapkan untuk memperluas data dengan menambah banyaknya data menggunakan variabel yang saling berpengaruh agar dapat memperoleh hasil akurasi yang lebih baik lagi serta dapat menggunakan metode lain untuk proses klasifikasi yang lebih optimal.

DAFTAR PUSTAKA

- Adharani, Y. and Ambo, S. N. (2019). Identifikasi Penyakit Balita Berdasarkan Gejala yang dialami dengan menggunakan Bayesian Network. pages 1–12.
- Ahammad, F., Abd Rashid, T. R. T., Mohamed, M., Tanbin, S., and Fuad,
 F. A. A. (2019). Contemporary Strategies and Current Trends in Designing
 Antiviral Drugs Against Dengue Fever Via Targeting Host-Based Approaches.
 Microorganisms, 7(9).
- Akbar, H. and Maulana Syaputra, E. (2019). Faktor Risiko Kejadian Demam Berdarah Dengue (DBD) di Kabupaten Indramayu. *MPPKI (Media Publ. Promosi Kesehat. Indones. Indones. J. Heal. Promot.*, 2(3):159–164.
- Alghifari, M. R. and Wibowo, A. P. (2019). Penerapan Metode K-Nearest Neighbor Untuk Klasifikasi Kinerja Satpam Berbasis Web. *J. Teknol. dan Manaj. Inform.*, 5(1).
- Annisa Halida (2020). Analisis kasus kemiskinan di provinsi kalimantan tengah dengan pendekatan. 5(2):134–141.
- Arifin, T. and Ariesta, D. (2019). Prediksi Penyakit Ginjal Kronis Menggunakan Algoritma Naive Bayes Classifier Berbasis Particle Swarm Optimization. 13(1):26–30.
- Bakhri, S. (2018). Analisis Jumlah Leukosit Dan Jenis Leukosit Pada Individu Yang Tidur Dengan Lampu Menyala Dan Yang Dipadamkan. *J. Media Anal. Kesehat.*, 1(1):83–91.

- Bayes, N. (2019). Implementasi Data Mining Dengan Algoritma Naive Bayes Untuk Memprediksi Angka Kelahiran. *J. Pelita Inform.*, 7(3):421–428.
- Butarbutar, R. N., Sumampouw, O. J., and Pinontoan, O. R. (2019). Trend Kejadian Demam Berdarah Dengue Di Kota Manado Demam Berdarah Dengue (DBD). KESMAS - J. Kesehat. Masy., 8(6):364–370.
- Cahyani, S., Rizkianti, T., and Susantiningsih, T. (2020). Hubungan Jumlah Trombosit, Nilai Hematokrit dan Rasio Neutrofil-Limfosit Terhadap Lama Rawat Inap Pasien DBD Anak di RSUD Budhi Asih Bulan Januari September Tahun2019. *Semin. Nas. Ris. Kedokt.* 2020, 1(1):49–59.
- Chatterjee, S., Dey, N., Shi, F., Ashour, A. S., and Fong, S. J. (2017). Clinical application of modified bag-of-features coupled with hybrid neural-based classifier in dengue fever classification using gene expression data.
- Chen, W. J. (2018). Dengue outbreaks and the geographic distribution of dengue vectors in Taiwan: A 20-year epidemiological analysis. *Biomed. J.*, 41(5):283–289.
- Destiana, R., Nasution, Y. N., and Wahyuningsih, S. (2019). Klasifikasi Probabilistic Neural Network (PNN) pada Data Diagnosa Penyakit Demam Berdarah. pages 15–21.
- Dewi, M. W. U., Herawati, S., and Subawa, N. (2020). Faktor-Faktor Yang Berhubungan Terhadap Derajat Berat Infeksi Virus Dengue Pada Pasien Dewasa Yang Dirawat Di Rumah Sakit Umum Pusat Sanglah Denpasar Bali. *J. Med. Udayana*, 9(4):2597–8012.
- Erwansyah, K. (2019). Implementasi Data Mining Untuk Menganalisa Hubungan

- Data Penjualan Produk Bahan Kimia Terhadap Persedian Stok Barang Menggunakan Algoritma FP (Frequent Pattern) Growth Pada PT . Grand Multi Chemicals. *J. Teknol. Sist. Inf. dan Sist. Komput. TGD*, 2(2):30–40.
- Gou, J., Qiu, W., Yi, Z., Xu, Y., Mao, Q., and Zhan, Y. (2019). A local mean representation-based K-nearest neighbor classifier. *ACM Trans. Intell. Syst. Technol.*, 10(3).
- Gultom, T. W., Tambunan, L. V., and Febryanto, J. (2019). Image Enhancement Kombinasi Metode Fuzzy Filtering dengan Metode Gaussian Filtering. *J. Mantik Penusa*, 3(1):110–116.
- Gupta, V. and Mittal, M. (2018). KNN and PCA classifier with autoregressive modelling during different ECG signal interpretation. *Procedia Comput. Sci.*, 125:18–24.
- Handayani, F. . (2015). Peramalan Jumlah Kasus Demam Berdarah Di Kabupaten
 Malang Menggunakan Metode Random Forest (Doctoral dissertation, Institut
 Teknologi Sepuluh Nopember). J. KomTekInfo Fak. Ilmu Komput., 2(1):47–61.
- Hasan, M. Z. (2020). Resepsi Al-Qur'an Sebagai Medium Penyembuhan Dalam Tradisi Bejampi Di Lombok. *J. Stud. Ilmu-ilmu Al-Qur'an dan Hadis*, 21(1), 133-152., 21(1):133.
- Hia, Y. (2019). Penerapan Proses Keperawatan Terhadap Penyakit Demam Berdarah Di Rumah Sakit. *Excell. Midwifery*, 2(2):1–6.
- Inggrid Yanuar, Rosa Andrie, F. R. (2017). Study of Hoax News Detection Using Naive Bayes Classifier in Indonesian Language. *Int. Conf. Inf. Commun. Technol. Syst.*, pages 73–78.

- Kaltsum, L. U. (2018). Al-Qur'an (Studi Ayat-Ayat Fitnah dengan Aplikasi Metode Tafsir Tematik). *Ilmu Ushuluddin*, 5(2):107–138.
- Kartika, E. A. (2018). Gambaran Presentase Nilai Neutrofil dan Limfosit Pada Pasien Penderita Demam Berdarah Dengue (Doctoral dissertation, STIKES Ngudia Husada Madura). STIKES, 53(9):1689–1699.
- Kinansi, R. R. and Pujiyanti, A. (2020). Pengaruh Karakteristik Tempat Penampungan Air Terhadap Densitas Larva Aedes sp. dan Risiko Penyebaran Demam Berdarah Dengue di Daerah Endemis di Indonesia The Effect of Characteristics of Containers On Larvae Aedes sp. Density and The Risk of Spreading. *BALABA*, 16(1):1–20.
- Nasution, D. A., Khotimah, H. H., and Chamidah, N. (2019). Perbandingan Normalisasi Data untuk Klasifikasi Wine Menggunakan Algoritma K-NN. Comput. Eng. Sci. Syst. J., 4(1):78.
- Nazaruddin, D. A., Bachtiar, F. A., and Dewi, R. K. (2019). Klasifikasi Penyakit Kelamin Pada Wanita Dengan Menggunakan Kombinasi Metode K-Nearest Neighbor Dan Naïve Bayes Classifier. 3(4):3266–3274.
- Nengsih, W. (2017). Analisa Akurasi Permodelan Supervised Dan Unsupervised. Sebatik, pages 285–291.
- Nikmatun, I. A. and Waspada, I. (2019). Implementasi Data Mining untuk Klasifikasi Masa Studi Mahasiswa Menggunakan Algoritma K-Nearest Neighbor. *J. SIMETRIS*, 10(2):421–432.
- Ningtyas, A. M., Lubis, I. K., and Herwanto, G. B. (2019). Monitoring Persebaran

- Penyakit Demam Berdarah Dengue dengan Memanfaatkan Data Berita Online. J. Kesehat. Vokasional, 4(2):105.
- Noshirma, M., Wadu, R., Kazwaini, M., Basuki, J., Km, R., Weri, P., and Timur, N. T. (2020). Deteksi Virus Dengue pada Nyamuk Aedes aegypti (Diptera: Culicidae) yang Tersebar di Kabupaten Sumba Timur dan Sumba Barat Daya Detection of Dengue Virus in Aedes aegypti (Diptera: Culicidae) Distributed in East and Southwest Sumba Districts. *J. Vektor Penyakit*, 14(1):57–64.
- Nur Itsna, I. (2020). Peningkatan Pengetahuan Masyarakat Dalam Menanggulangi Penyakit Demam Berdarah Dengue (DBD) di Desa Karangmalang Kedungbanteng. *JPKMI (Jurnal Pengabdi. Kpd. Masy. Indones.*, 1(1):35–41.
- Permatasari, R., Almurdi, A., and Tjong, D. H. (2020). Viral Load Terhadap Nilai Hematokrit Pada Infeksi Virus Dengue. *J. Kesehat. PERINTIS (Perintis's Heal. Journal)*, 7(1):57–63.
- Priyono, P., & Nuris, N. (2020). Sistem Pendukung Keputusan Metode TOPSIS Untuk Diagnosa Penyakit Demam Berdarah. *Inti Nusa Mandiri*, 15(1):51–58.
- Putra, R. E., Tjandrasa, H., and Suciati, N. (2020). Severity classification of non-proliferative diabetic retinopathy using convolutional support vector machine.
 Int. J. Intell. Eng. Syst., 13(4):156–170.
- Qureshi, S. and Atangana, A. (2019). Mathematical Analysis of Dengue Fever Outbreak by Novel Fractional Operators With Field Data. *Physica A*, 526:121–127.
- Rahman, F., Ihsan, M., Pristianto, A., Khadijah, S., & Budi, I. S. (2019). Terapui

- Latihan Komprehensif Untuk Penderita Penyakit Paru Obstruktif Kronik. *Urecol*, 9(1).
- Rivanthio, T. R., Ramdhani, M., and Sahi, A. (2020). Penerapan Teknik Clustering Data Mining untuk Memprediksi Kesesuaian Jurusan Siswa (Studi Kasus SMA PGRI 1 Subang). *Fakt. Exacta*, 13(2):125–131.
- Rohman, A. (2019). Komparasi Metode Klasifikasi Data Mining Untuk Prediksi Abstrak. *J. Neo Tek.*, 5(1):23–29.
- Safri, Y. F., Arifudin, R., and Muslim, M. A. (2018). K-Nearest Neighbor and Naive Bayes Classifier Algorithm in Determining The Classification of Healthy Card Indonesia Giving to The Poor. *Sci. J. Informatics*, 5(1):9–18.
- Sanusi, R. N. M. and Saputro, D. R. S. (2020). Metode Robust Principle Component Analysis (RPCA) dengan Algoritme Proyeksi dan Matriks Ragam Peragam. *Prism. Pros. Semin. Nas. Mat.*, 2(2019):52–57.
- Sendangmulyo, D. I. K., Pangestika, T. L., Cahyo, K., Tirto, B., and Nugraha, P. (2017). Faktor-Faktor Yang Mempengaruhi Perilaku Jumantik Dalam Sistem Kewaspadaan Dini Demam Berdarah Dengue Di Kelurahan Sendangmulyo. *J. Kesehat. Masy.*, 5(5):1113–1122.
- Suniantara, I. K. P., Suwardika, G., and Soraya, S. (2020). Peningkatan Akurasi Klasifikasi Ketidaktepatan Waktu Kelulusan Mahasiswa Menggunakan Metode Boosting Neural Network. *J. Varian*, 3(2):95–102.
- Suparmono, A. M. (2021). HUBUNGAN JUMLAH TROMBOSIT DAN NILAI HEMATOKRIT TERHADAP KEJADIAN SINDROM SYOK DENGUE (SSD). 02(02):533–536.

Syafitri Hidayatul AA, Yuita Arum S, A. A. (2018). Seleksi Fitur Information Gain untuk Klasifikasi Penyakit Jantung Menggunakan Kombinasi Metode K-Nearest Neighbor dan Naïve Bayes. *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, 2(9):2546–2554.

Tjasyono, B. (2019). B. Tjasyono, Ilmu Kebumian dan Antariksa, Bandung: Pascasarjana UPI, 2009. 5(05):577–584.

