

# PERBANDINGAN ALGORITMA NAÏVE BAYES DAN K-NEAREST NEIGHBOR PADA KLASIFIKASI JUDUL ARTIKEL PADA JURNAL ILMIAH

Anang Prayogo<sup>\*1)</sup>, Fauziah<sup>2)</sup>, Winarsih<sup>3)</sup>

1. Universitas Nasional, Indonesia
2. Universitas Nasional, Indonesia
3. Universitas Nasional, Indonesia

## Article Info

**Kata Kunci:** Klasifikasi; *K-Nearest Neighbor*; *Naïve Bayes*.

**Keywords:** *Classification*, *K-Nearest Neighbor*, *Naïve Bayes*

## Article history:

Received 22 May 2023

Revised 5 June 2023

Accepted 19 June 2023

Available online 1 December 2023

## DOI :

<https://doi.org/10.29100/jipi.v8i4.4141>

\* Corresponding author.

Anang Prayogo

E-mail address:

[anang1321@gmail.com](mailto:anang1321@gmail.com)

## ABSTRAK

Perkembangan teknologi Informatika membuat kebutuhan akan otomasi pada sistem kerja semakin meningkat. Termasuk kebutuhan akan klasifikasi artikel jurnal secara otomatis mulai dibutuhkan untuk menentukan kategori jurnal yang tepat. Metode text mining telah dapat digunakan untuk melakukan otomatisasi artikel jurnal berdasarkan kategori. Penelitian ini memakai beberapa metode text mining yakni Naïve Bayes (NB) dan K-Nearest Neighbor (K-NN). Dataset yang digunakan berupa judul artikel jurnal sejumlah 200 data yang kemudian dipisahkan menjadi data uji dan data latih. Hasil akurasi yang didapat dengan menggunakan K-NN sebesar 85,00% dengan data uji 10% dan k=3. Hasil akurasi menggunakan NB adalah sebesar 81,66% dengan menggunakan data uji 30%. Akurasi tidak terlalu berbeda jauh namun NB menciptakan akurasi yang sedikit lebih besar dibandingkan K-NN.

## ABSTRACT

The development of Informatics technology has made the fear of automation in the work system increasing. Including the need for automatic classification of journal articles began to be needed to determine the right journal categories. The text mining method can be used to automate journal articles by category. This research used several text mining methods, namely Naïve Bayes (NB) and K-Nearest Neighbor (K-NN). The dataset used is in the form of article titles and 200 data which are then separated into test data and training data. The accuracy obtained with KNN were 85,00% with test data of 10% and k = 3. The accuracy result using NB is 81,66% using 30% test data. Accuracy doesn't differ much but NB creates slightly greater accuracy than K-NN.

## I. PENDAHULUAN

**K**ebutuhan akan sistem kerja yang berbasis teknologi informasi (TI) telah berkembang seiring dengan perkembangan zaman. Pengembangan, pemeliharaan, dan pengoperasian jaringan komputer, perangkat lunak, serta sistem pemrosesan dan distribusi data adalah semua bidang teknologi informasi (TI) [1].

Situasinya mirip dengan klasifikasi jurnal yang sebelumnya dilakukan secara manual. Namun, seiring berjalannya waktu, mengklasifikasi jurnal berubah secara otomatis seiring dengan perkembangan teknologi informasi. Versi digital dari berbagai karya akademik tersedia, termasuk tesis, laporan penelitian, laporan kerja praktek [2].

Ditambah dengan kebutuhan untuk memperoleh informasi dalam bentuk dokumen yaitu berupa jurnal atau artikel ilmiah meningkat setiap tahun [3]. Kebutuhan akan jurnal atau artikel ilmiah pun semakin meningkat daya tamping mahasiswa yang terus meningkat setiap tahunnya disetiap universitas juga berpengaruh peneliti yang beranggotakan mahasiswa dan dosen [4]. Konsep text mining digunakan sebagai klasifikasi dokumen teks, di mana dokumen yang ada di klasifikasikan menurut dokumen yang akan diproses [5]. Dengan bantuan konsep ini, membuat artikel yang diteliti menjelaskan kategori jenis dengan memunculkan kata-kata dari artikel yang ada [6].

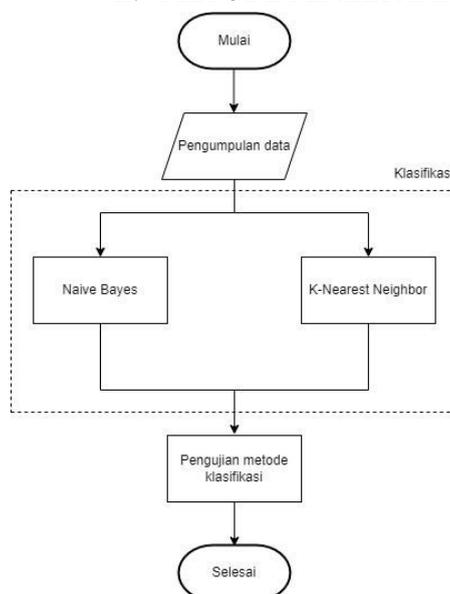
Pada penelitian terdahulu sudah ada beberapa yang mengembangkan metode ini salah satunya yaitu menghasilkan bahwa algoritma hasil *K-Nearest Neighbor* kinerja klasifikasi terbaik dicapai dengan persentase accuracy sebesar 72.91% dan precision mencapai 73,36% [7]. Kemudian ada yang menunjukkan bahwa algoritma

*Naïve Bayes* dan *K-Nearest Neighbor* dapat digunakan dalam mengklasifikasi teks, Pada penelitian ini akurasi *k-Nearest Neighbor* lebih baik dengan akurasi 80% dibandingkan *Naïve Bayes* yang mendapatkan akurasi 73% [8]. Adapun penelitian selanjutnya membandingkan ketiga metode yaitu dari hasil perbandingan algoritma klasifikasi yang terbaik adalah algoritma *Support Vector Machine* sebesar 81%. Sedangkan *Naïve Bayes* sebesar 67%, dan *K-Nearest Neighbor* sebesar 80% [9].

Penelitian ini akan dilakukan untuk membandingkan dua metode klasifikasi yaitu *Naïve Bayes* dan *K-Nearest Neighbor*. Kedua metode tersebut digunakan untuk klasifikasi jurnal menggunakan judul artikel. Penelitian ini juga akan menganalisis hasil akurasi dari kedua metode yang cocok dipakai untuk klasifikasi jurnal berdasarkan judul artikel. Mengapa menggunakan kedua metode ini dikarenakan diperlukannya sebuah metode klasifikasi dokumen yang dapat mengelompokkan artikel secara otomatis dan akurat. Terdapat banyak metode klasifikasi yang dapat digunakan. Metode yang digunakan dalam penelitian ini adalah *Naïve Bayes* dan sebagai baseline digunakan metode *K-Nearest Neighbor*. Metode *Naïve Bayes* dipilih karena dapat menghasilkan akurasi yang maksimal dengan data latih yang sedikit. Sedangkan metode *K-Nearest Neighbor* dipilih karena metode tersebut tangguh terhadap data noise.

Penelitian yang terkait dalam perbandingan tersebut yang sudah saya dapatkan ada tiga jurnal yang cocok dengan penelitian saya. Dengan itu saya kembangkan dalam penggunaan dataset artikel yang berjumlah lebih dari 100. Penelitian sebelumnya hanya menggunakan 20 dan 40 jurnal. Dan juga akurasi yang dihasilkan pada penelitian ini lebih besar dari 80%.

## II. METODE PENELITIAN



Gambar. 1. Tahapan Penelitian

Gambar 1 merupakan alur proses terjadinya pengklasifikasi kedua metode tersebut dari awal sampai selesai.

### 2.3 Pengumpulan Data

Pada titik ini, data dari internet yang bersumber dari google scholar digunakan sebagai metode pengumpulan jurnal untuk klasifikasi. Pada langkah ini proses pengumpulan data pada klasifikasi jurnal dilakukan dengan mengambil data dari google scholar. Data yang dikumpulkan berbentuk e-book dan e-jurnal. Setelah itu, data akan dipisahkan menjadi data uji dan data latih. Data tersebut akan digunakan untuk mengklasifikasikan *Naïve Bayes* dan *K-Nearest Neighbor* setelah dibagi.

## 2.4 Klasifikasi Menggunakan Algoritma *Naïve Bayes*

Metode ini adalah salah satu metode yang digunakan dalam melakukan klasifikasi. *Naive Bayes* adalah klasifikasi yang dibuat dengan metode probabilitas [10]. Algoritma ini dahulu dikenal dengan Teorema *Bayes*, karena mengacu pada kemampuan algoritma ini untuk mengantisipasi peluang masa depan berdasarkan pengalaman sebelumnya. Menurut klasifikasi *Naive Bayes*, nilai atribut kelas tidak dapat dipengaruhi oleh nilai atribut lainnya. Karena menganggap kemunculan satu kata dalam sebuah kalimat tidak dapat mempengaruhi kemunculan kata lain [11].

$$p(w_i|c_j) = \frac{1 + n_i}{n + |\text{kosakata}|}$$

Keterangan :

$p(w_i|c_j)$  menyatakan peluang untuk kata di setiap pengelompokan, lalu  $n_i$  menyatakan jumlah munculnya kata pada setiap pengelompokan, selain itu terdapat variable  $n$  menyatakan total kata dalam dokumen milik kelas tertentu, dan  $|\text{kosakata}|$  adalah total semua kata dalam dokumen latihan.

$$p(c_j) = \frac{(\text{doc}_j)}{n(\text{sampel})}$$

Keterangan :

$p(c_j)$  yaitu peluang dokumen kategori, selain itu  $n(\text{doc}_j)$  menyatakan total data dari suatu pengelompokan, dan  $n(\text{sampel})$  ialah total dokumen latihan.

## 2.5 Klasifikasi Menggunakan *K-Nearest Neighbor*

Metode ini melakukan klasifikasi berdasarkan (K) paling dekat dengan objek baru dapat digunakan untuk mengklasifikasikannya menggunakan Metode ini. Karena hasil dari *query instance* atau kategori diurutkan menurut jumlah kelas terbanyak yang muncul, *K-NN* dimasukkan ke dalam algoritma supervised learning [12].

Klasifikasi ini digunakan selaku nilai prediksi dari instance query baru oleh algoritma *K-NN*. Perhitungan melibatkan penjumlahan semua nilai kesamaan untuk satu kategori, yang akan dibandingkan dengan nilai mana pun yang lebih tinggi [13].

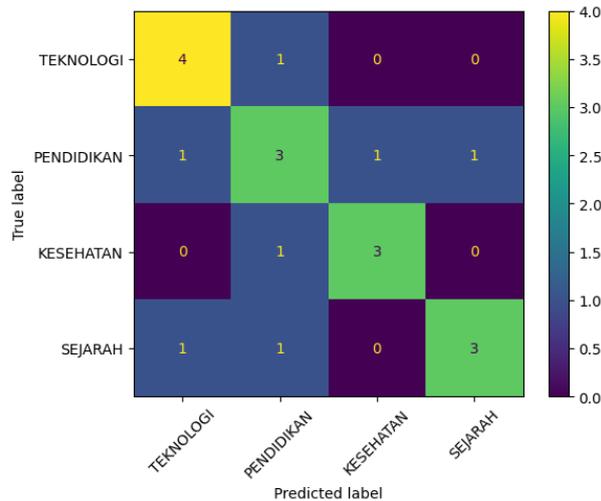
$$\text{Cos}(i, k) = \frac{\sum_k (d_i d_k)}{\sqrt{\sum_k d_i^2} \sqrt{\sum_k d_k^2}}$$

Keterangan :

$\sum_k (d_i d_k)$  adalah Vector pada  $i$  dan  $k$ , lalu  $\sqrt{\sum_k d_i^2}$  menyatakan panjangnya vector  $I$ , selain itu  $\sqrt{\sum_k d_k^2}$  menyatakan Panjangnya vector  $j$ , berikutnya variable  $i$  yaitu dokumen uji  $I$ , dan variable  $j$  itu adalah dokumen latihan  $j$ .

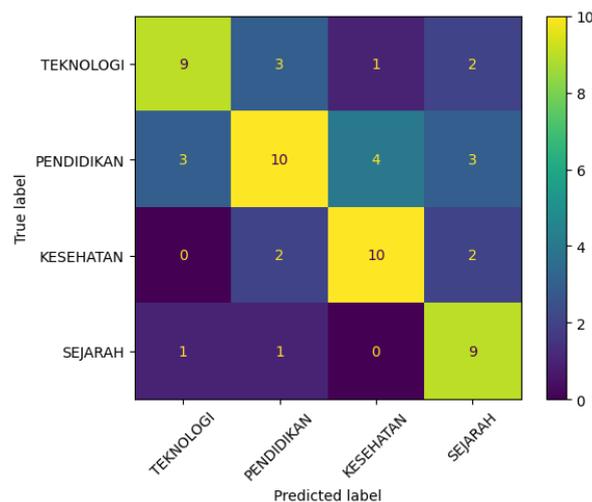
## 2.6 Pengujian Metode Klasifikasi

Pada pengujian metode klasifikasi merupakan pengujian dimana antara kedua algoritma yakni *NB* dan *K-NN* akan dipilih yang terbaik untuk mengklarifikasi data jurnal. Dalam pengujiannya memakai label *confusion matrix* pada jumlah data yang diklasifikasikan dengan benar atau salah dirinci dalam tabel ini [14]. *Confusion matrix* juga merupakan alat analisis prediktif yang dapat digunakan untuk membentuk metrik evaluasi seperti akurasi, presisi, daya ingat, dan *F1-Score* atau *F-Measure*. Selain itu, ini menampilkan dan mengontraskan nilai sebenarnya dengan nilai model yang diprediksi [15]. Berikut tampilan *confusion matrix* yang dapat dilihat pada gambar 2 dan 3 :



Gambar. 2. Confusion Matrix pada Naïve Bayes

Gambar 2 merupakan hasil data uji pada algoritma naïve bayes yang bisa dilihat dari 5 jurnal dengan kategori teknologi,terdapat 4 jurnal yang diklasifikasikan sesuai dengan kategorinya,seandainya 1 jurnal lainnya salah diklasifikasikan.



Gambar. 3. Confusion Matrix pada K-Nearest Neighbor

Gambar 3 merupakan hasil data uji pada algoritma k-nearest neighbor yang bisa dilihat dari 15 jurnal dengan kategori teknologi, terdapat 9 jurnal yang diklasifikasikan sesuai dengan kategorinya,seandainya 6 jurnal lainnya salah diklasifikasikan.

### III. HASIL DAN PEMBAHASAN

Metode pengumpulan ini yakni menarik data di internet yang bersumber dari google scholar. Data yang sudah ditarik dimasukkan kedalam excel kemudian dimasukkan ke dalam database. Dibawah ini merupakan contoh data yang diambil dari internet lalu diinput di excel yang dapat dilihat pada gambar 4 :

NO.	nama/jurnal	Penulis	Tahun	Kategori	Penerbit	file
1	A review of using natural rubber in the modification of bitumen and asphalt mixtures used for road construction	A Al Sabaei	2019	TEKNOLOGI	researchgate.net	<a href="https://www.researchgate.net/profile/Nur-Id">https://www.researchgate.net/profile/Nur-Id</a>
2	A review jenis dan pencetakan printing untuk pembuatan prototype	T Ruslianto	2019	TEKNOLOGI	journal.akpind.ac.id	<a href="https://journal.akpind.ac.id/index.php/jurte/">https://journal.akpind.ac.id/index.php/jurte/</a>
3	An empirical study on gamification for learning programming language website	FL Khaalae	2019	TEKNOLOGI	researchgate.net	<a href="https://www.researchgate.net/profile/Firas-I">https://www.researchgate.net/profile/Firas-I</a>
4	Analisis dampak perkembangan teknologi informasi dan komunikasi dalam bidang pendidikan	NYS Munti	2020	TEKNOLOGI	iptam.org	<a href="https://iptam.org/index.php/iptam/article/dc">https://iptam.org/index.php/iptam/article/dc</a>
5	Analisa Performa Mesin Motor Dengan Menggunakan Campuran BioetanolPertamax	SA Yudiastriani	2019	TEKNOLOGI	jurnal.umj.ac.id	<a href="https://jurnal.umj.ac.id/index.php/jurtek/arti">https://jurnal.umj.ac.id/index.php/jurtek/arti</a>
6	Analisis Risiko Kecelakaan Kerja Kayawan Menggunakan Metode Hazard and Operability HAZOP di PT PJB Services PLTU Tembilahan	I Rahmanto	2022	TEKNOLOGI	jurnal-tmtt.com	<a href="https://jurnal-tmtt.com/index.php/home/arti">https://jurnal-tmtt.com/index.php/home/arti</a>
7	Analisis Beban Kerja Mental Pekerja Train Distribusi PT Solusi Bangun Indonesia	FN Rahman	2022	TEKNOLOGI	jurnal-tmtt.com	<a href="https://jurnal-tmtt.com/index.php/home/arti">https://jurnal-tmtt.com/index.php/home/arti</a>
8	Analisis Clustering Untuk Recreatesaling Fasilitas Kesehatan Menggunakan Metode Fuzzy CMeans	V Herlinda	2021	TEKNOLOGI	jim.teknokrat.ac.id	<a href="http://jim.teknokrat.ac.id/index.php/sistemir">http://jim.teknokrat.ac.id/index.php/sistemir</a>
9	Analisis dan Perancangan Sistem Informasi Akuntansi Pengelolaan Tabungan Siswa pada SD ArRaudah Bandarlampung	D Damayanti	2021	TEKNOLOGI	ojs.unikom.ac.id	<a href="https://ojs.unikom.ac.id/index.php/jati/article">https://ojs.unikom.ac.id/index.php/jati/article</a>
10	Analisis Data Mining Untuk Clustering Kasus Covid Di Provinsi Lampung Dengan Algoritma KMeans	Z Nabila	2021	TEKNOLOGI	jim.teknokrat.ac.id	<a href="http://jim.teknokrat.ac.id/index.php/sistemir">http://jim.teknokrat.ac.id/index.php/sistemir</a>
11	Analisis Ergonomi Dalam Penggunaan Mesin Penggilingan Pupuk Menggunakan Metode Quick Exposure Checklist Pada PT Putra Manunggal Sakti	S Adji	2022	TEKNOLOGI	jurnal-tmtt.com	<a href="https://jurnal-tmtt.com/index.php/home/arti">https://jurnal-tmtt.com/index.php/home/arti</a>
12	Analisis Gamifikasi Learning Berbasis Teknologi Blockchain	Q Alini	2021	TEKNOLOGI	adi-journal.org	<a href="https://adi-journal.org/index.php/abd/article">https://adi-journal.org/index.php/abd/article</a>
13	Analisis jarak jangkauan lora dengan parameter rssi dan packet loss pada area urban	A Yanisah	2020	TEKNOLOGI	journal.akpind.ac.id	<a href="https://journal.akpind.ac.id/index.php/techr">https://journal.akpind.ac.id/index.php/techr</a>
14	Analisis Kapabilitas Proses untuk Pengendalian Kualitas Air Limbah di Industri Farmasi	D Rimantho	2019	TEKNOLOGI	jurnal.umj.ac.id	<a href="https://jurnal.umj.ac.id/index.php/jurtek/arti">https://jurnal.umj.ac.id/index.php/jurtek/arti</a>
15	Analisis Kebutuhan Bahan Ajar Digital Berbasis Flipbook untuk Membedakan Keterampilan Abad Peserta Didik pada Pembelajaran IPA Sekolah Dasar	FFK Sari	2021	TEKNOLOGI	jbasic.org	<a href="https://www.jbasic.org/index.php/basicedu/">https://www.jbasic.org/index.php/basicedu/</a>
16	Analisis Kelangkaan Ekonomi Menggunakan Metode Net Present Value NPV Metode Internal Rate Of Return IRR Payback Period PPB Pada Unit Stone Crusher	GM Abak	2020	TEKNOLOGI	ejournal.undana.ac.id	<a href="http://ejournal.undana.ac.id/jurnal_teknologi/">http://ejournal.undana.ac.id/jurnal_teknologi/</a>
17	Analisis komponen utama faktor-faktor pendahuluan antecedents berbagai pengetahuan pada usaha mikro kecil dan menengah UMKM di Indonesia	A Ilmaniaty	2019	TEKNOLOGI	jurnal.umj.ac.id	<a href="https://jurnal.umj.ac.id/index.php/jurtek/arti">https://jurnal.umj.ac.id/index.php/jurtek/arti</a>
18	Analisis pengaruh penerapan model pembelajaran sains teknologi masyarakat terhadap hasil belajar siswa	GL Suarni	2021	TEKNOLOGI	e-journal.undikma.ac.id	<a href="https://e-journal.undikma.ac.id/index.php/pe">https://e-journal.undikma.ac.id/index.php/pe</a>
19	Analisis Pengendalian Bahan Baku Ubi Jalar Jalar Menggunakan Metode Economic Order Quantity EOQ Dan HsIn Rau PT Galih Estetika Indonesia	YB Ismaya	2022	TEKNOLOGI	jurnal-tmtt.com	<a href="http://jurnal-tmtt.com/index.php/home/artic">http://jurnal-tmtt.com/index.php/home/artic</a>
20	Analisis Pengendalian Risiko Kecelakaan Kerja Bagian MekaniK Pada Proyek PLTU Ampana MW Menggunakan Metode Job Safety Analysis JSA	S Bailli	2022	TEKNOLOGI	jurnal-tmtt.com	<a href="https://jurnal-tmtt.com/index.php/home/arti">https://jurnal-tmtt.com/index.php/home/arti</a>
21	Analisis perkembangan ilmu pengetahuan dan teknologi IPTEK dalam pendidikan	F Mulyani	2021	TEKNOLOGI	jsk.org	<a href="http://www.jsk.org/index.php/jsk/article/6">http://www.jsk.org/index.php/jsk/article/6</a>
22	Analisis Pola Menggunakan Metode Untuk Peminatan Jurusan Siswa Berdasarkan Kurikulum Studi Kasus Sman Natar	AFO Pasaribu	2021	TEKNOLOGI	jim.teknokrat.ac.id	<a href="http://jim.teknokrat.ac.id/index.php/sistemir">http://jim.teknokrat.ac.id/index.php/sistemir</a>
23	Analisis Potensi Dan Strategi Penemuan Ekowisata Daerah Penyangga Taman Nasional Way Kambas	S Pramono	2020	TEKNOLOGI	jim.teknokrat.ac.id	<a href="http://jim.teknokrat.ac.id/index.php/sistemir">http://jim.teknokrat.ac.id/index.php/sistemir</a>
24	Analisis Produktivitas Operator Kasir Menggunakan Metode Work Sampling Studi Kasus Gerai Chatime Mangga Besar	H Tannady	2019	TEKNOLOGI	core.ac.uk	<a href="https://core.ac.uk/download/pdf/270220673">https://core.ac.uk/download/pdf/270220673</a>
25	Aplikasi EMarketplace Bagi Pengusaha Stainless Berbasis Mobile Di Wilayah Bandar Lampung	CA Febrina	2021	TEKNOLOGI	jim.teknokrat.ac.id	<a href="http://jim.teknokrat.ac.id/index.php/sistemir">http://jim.teknokrat.ac.id/index.php/sistemir</a>
26	Aplikasi EMarketplace Penjualan Hasil Panen Ikan Lela Studi Kasus Kabupaten Pingsu Kecamatan Pagelaran	A Afifah	2020	TEKNOLOGI	jim.teknokrat.ac.id	<a href="http://jim.teknokrat.ac.id/index.php/sistemir">http://jim.teknokrat.ac.id/index.php/sistemir</a>
27	Aplikasi pembelajaran audit sistem informasi dan tata kelola teknologi informasi berbasis Mobile	Y Rahmadhan	2020	TEKNOLOGI	ejurnal.teknokrat.ac.id	<a href="https://ejurnal.teknokrat.ac.id/index.php/tek">https://ejurnal.teknokrat.ac.id/index.php/tek</a>
28	Aplikasi Pengenalan Perangkat Keras Komputer Berbasis Android Menggunakan Augmented Reality Ar	AF Ramadhan	2021	TEKNOLOGI	jim.teknokrat.ac.id	<a href="http://jim.teknokrat.ac.id/index.php/sistemir">http://jim.teknokrat.ac.id/index.php/sistemir</a>
29	Aplikasi Smart School Untuk Kebutuhan Guru Di Era New Normal	Y Khadaffi	2021	TEKNOLOGI	jim.teknokrat.ac.id	<a href="http://jim.teknokrat.ac.id/index.php/sistemir">http://jim.teknokrat.ac.id/index.php/sistemir</a>
30	Application of the Picture and Picture Type of Cooperative Learning Model in Improving Student Learning Creativity	N Siregar	2021	TEKNOLOGI	jurnal.umj.ac.id	<a href="https://jurnal.umj.ac.id/unj/index.php/tp/arti">https://jurnal.umj.ac.id/unj/index.php/tp/arti</a>

Gambar. 4. Data jurnal yang diinput di excel

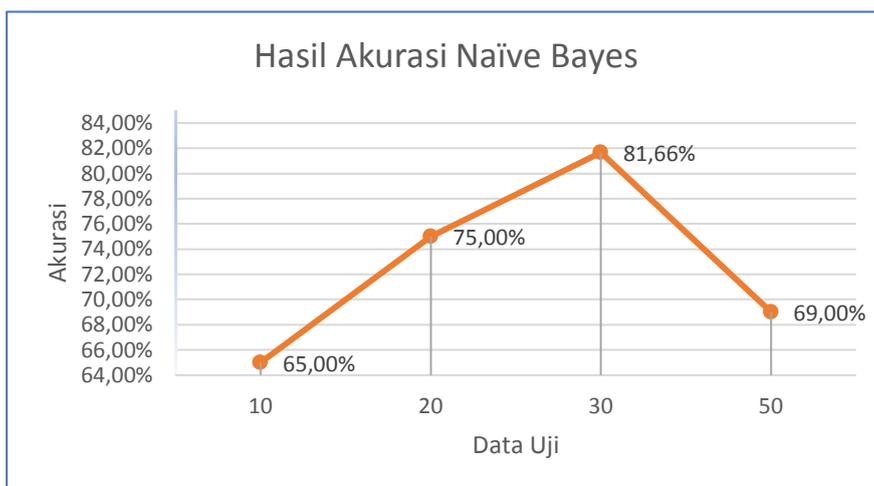
## 2. Klasifikasi Naïve Bayes

Pada klasifikasi jumlah data yang digunakan sebesar 200 data, kemudian akan dilakukan terlebih dahulu pembagian data training dan data uji dimana akan dibagi menjadi beberapa bagian, kemudian pada data uji akan di test untuk implementasi algoritma NB dalam mencari akurasi. Pembagian data tersebut antara lain data training 70% data uji 30%, data training 80% data uji 20%, dan data training 90% data uji 10%.

30% dengan akurasi sebesar 81,66%. Berikut data-data yang sudah di test dapat dilihat pada tabel 1 :

TABEL 1  
HASIL AKURASI NAÏVE BAYES

Data Training	Data Uji	Akurasi
50%(100 data)	50%(100 data)	69,00%
70%(140 data)	30%(60 data)	81,66%
80%(160 data)	20%(40 data)	75,00%
90%(180 data)	10%(20 data)	65,00%



Gambar. 5. Grafik Akurasi Naïve Bayes

Gambar 5 merupakan data grafik semua hasil akurasi algoritma Naïve Bayes yang dapat dilihat adanya kenaikan dan penurunan.

### 3. Klasifikasi *K-Nearest Neighbor*

Total data yang dipakai pada klasifikasi ini adalah 200. Kemudian dilakukan pemecahan data training dan data uji yang dipecah menjadi beberapa bagian, kemudian pada data uji akan di test untuk implementasi algoritma *K-Nearest Neighbor* dalam mencari akurasi. Pembagian data tersebut antara lain data training 70% data uji 30%, data training 80% data uji 20%, dan data training 50% data uji 50%. Dan pada jumlah (K)nya akan ditentukan antara lain 1, 3, 5, dan 7.

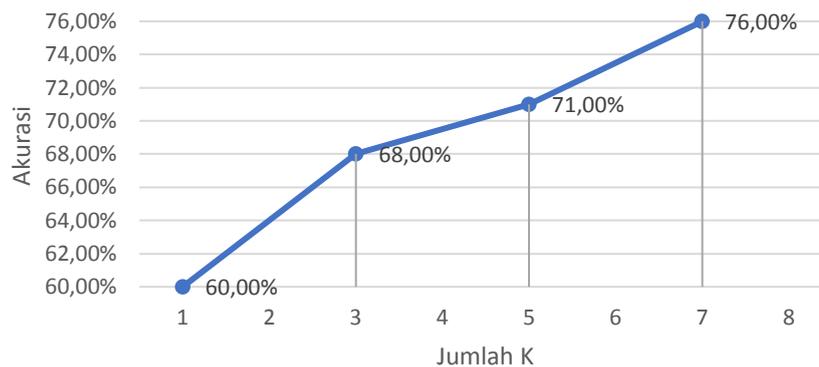
#### A. Klasifikasi *K-Nearest Neighbor* dengan 50% data training dan 50% data uji

Pengujian algoritma *K-Nearest Neighbor* menggunakan data uji 50% didapatkan akurasi terbaik sebesar 76,00% pada  $k = 7$ . Berikut hasil akurasi data uji 50% dapat dilihat pada tabel 2 :

TABEL 2  
 HASIL AKURASI K-NEAREST NEIGHBOR DATA UJI 50%

Jumlah K	Akurasi
1	60,00%
3	68,00%
5	71,00%
7	76,00%

Hasil Akurasi K-NN Data Uji 50%



Gambar 6 Grafik Akurasi K-NN Data Uji 50%

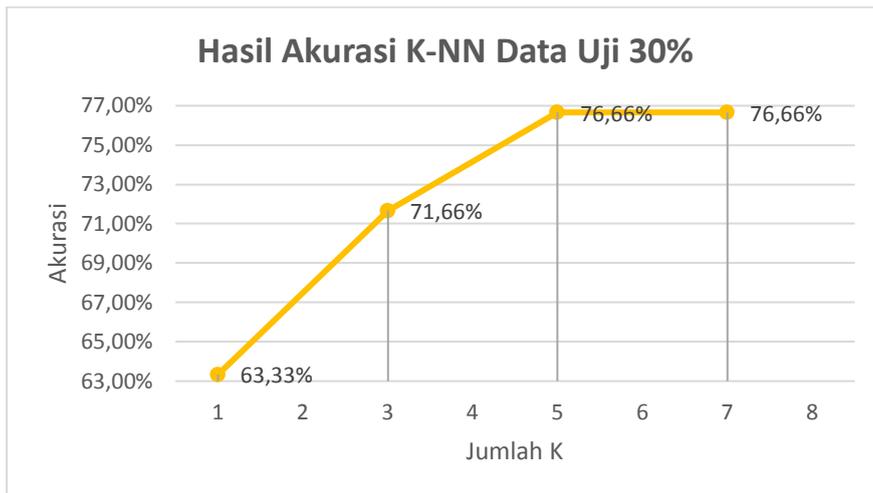
Gambar 6 merupakan data grafik hasil akurasi k-nearest neighbor yang menggunakan data uji 50% yang menghasilkan akurasi yang meningkat.

#### B. Klasifikasi *K-Nearest Neighbor* dengan 70% data training dan 30% data uji

Pengujian algoritma *K-NN* menggunakan data uji 30% didapatkan akurasi terbaik pada  $k= 5$  dan 7 dengan akurasi sebesar 76,66%. Berikut hasil akurasi data uji 30% dapat dilihat pada tabel 3 :

TABEL 3  
 HASIL AKURASI K-NEAREST NEIGHBOR DATA UJI 30%

Jumlah K	Akurasi
1	63,33%
3	71,66%
5	76,66%
7	76,66%



Gambar 7 Grafik Akurasi K-NN Data Uji 30%

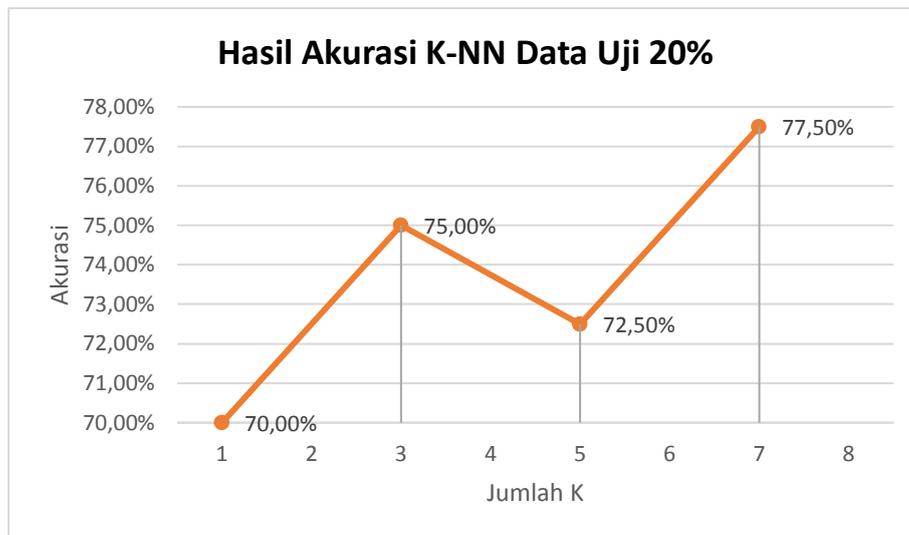
Gambar 7 merupakan data grafik hasil akurasi K-Nearest Neighbor yang menggunakan data uji 30% yang mengalami peningkatan.

**C. Klasifikasi *K-Nearest Neighbor* dengan 80% data training dan 20% data uji**

Pengujian algoritma *K-NN* menggunakan data uji 20% didapatkan akurasi terbaik pada  $k=7$  dengan akurasi sebesar 77,50%. Berikut hasil akurasi data uji 20% dapat dilihat pada tabel 4 :

TABEL 4  
 HASIL AKURASI K-NEAREST NEIGHBOR DATA UJI 20%

Jumlah K	Akurasi
1	70,00%
3	75,00%
5	72,50%
7	77,50%



Gambar. 8. Grafik Akurasi K-NN Data Uji 20%

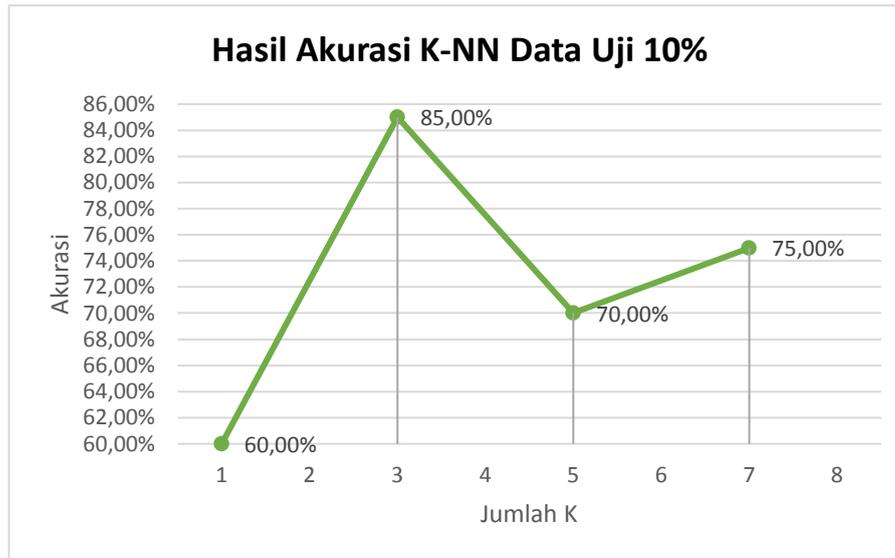
Gambar 8 merupakan data grafik hasil akurasi *K-Nearest Neighbor* yang menggunakan data uji 20% yang mengalami peningkatan.

**D. Klasifikasi *K-Nearest Neighbor* dengan 90% data training dan 10% data uji**

Pengujian *K-NN* pada data uji 10% didapatkan akurasi terbaik sebesar 85,00% pada  $k = 3$ . Berikut hasil akurasi data uji 10% dapat dilihat pada tabel 5 :

TABEL 5  
 HASIL AKURASI K-NEAREST NEIGHBOR DATA UJI 10%

Jumlah K	Akurasi
1	60,00%
3	85,00%
5	70,00%
7	75,00%



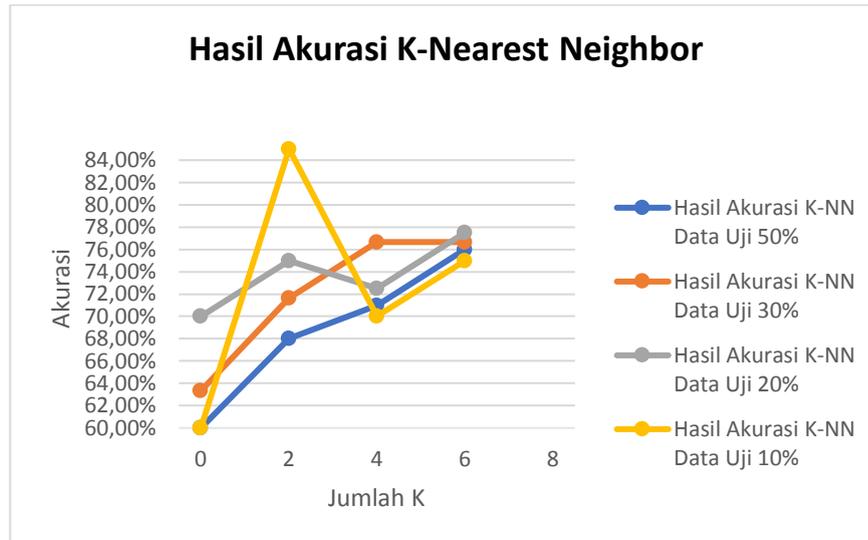
Gambar 9 Grafik Akurasi K-NN Data Uji 10%

Gambar 9 merupakan data grafik akurasi K-Nearest Neighbor yang menggunakan data uji 10% yang mengalami penurunan.

Pengujian algoritma *K-NN* menggunakan beberapa bagian data uji dengan beberapa jumlah *K* didapatkan akurasi terbaik pada jumlah *k* masing masing dari data uji tersebut. Namun akurasi terbaik terdapat pada data uji 10% dengan *K=3* sebesar 85,00%. Berikut yaitu hasil semua data testing K-Nearest Neighbor yang dapat dilihat pada tabel 6 :

TABEL 6  
HASIL AKURASI K-NEAREST NEIGHBOR

data train- ing	data uji	jumlah k	akurasi
50%	50%	1	60,00%
		3	68,00%
		5	71,00%
		7	<b>76,00%</b>
70%	30%	1	63,33%
		3	71,66%
		5	<b>76,66%</b>
80	20%	7	<b>76,66%</b>
		1	70,00%
		3	75,00%
		5	72,50%
90%	10%	7	<b>77,50%</b>
		1	60,00%
		3	<b>85,00%</b>
		5	70,00%
		7	75,00%



Gambar 10 Grafik Semua Akurasi K-NN

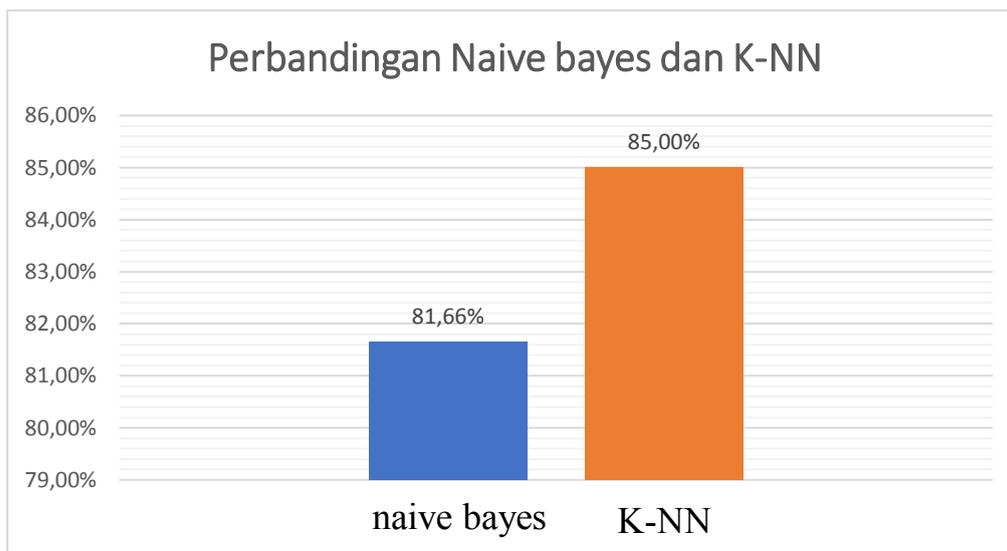
Gambar 10 merupakan data grafik hasil akurasi keseluruhan yang menggunakan algoritma K-Nearest Neighbor.

#### 4. Perbandingan Hasil Naïve Bayes dan K-Nearest Neighbor

Ketika akurasi dari *K-NN* dan *NB* dibandingkan, ditemukan akurasi *K-NN* kian unggul dari *NB*. Karena pada akurasi pengujian *K-NN* menghasilkan akurasi tertinggi yaitu 85,00% dengan data uji 10% pada  $k = 3$  jika dibandingkan dengan akurasi *NB* yang menghasilkan akurasi sebesar 81,66% pada data uji 30%. Berikut hasil akurasi kedua algoritma tersebut yang dapat dilihat pada tabel 7 :

TABEL 7  
 HASIL PERBANDINGAN AKURASI NAÏVE BAYES DAN K-NEAREST NEIGHBOR

Metode	Akurasi
<i>Naïve Bayes</i>	81,66%
<i>K-Nearest Neighbor</i>	85,00%



Gambar. 11. Grafik Perbandingan Akurasi K-NN dan Naïve Bayes

Gambar 11 merupakan data grafik perbandingan antara kedua algoritma K-NN dan Naïve Bayes yang menghasilkan akurasi terbesar terdapat di K-NN yaitu 85,00% dan Naïve Bayes mendapatkan akurasi 81,66%.

## 5. Visualisasi Web

Demonstrasi akurasi klasifikasi judul artikel dalam jurnal ilmiah dapat dilihat dalam visualisasi ini. Hasil akhirnya bisa terlihat seperti grafik, misalnya. Visualisasi data grafis berbasis web disediakan di bawah ini.



Gambar. 12. Visualisasi Klasifikasi pada Web

Gambar 12 merupakan sebuah tampilan output website yang terdapat di klasifikasi K-Nearest Neighbor dalam data grafik.

## 6. Hasil Penelitian

Berdasarkan penelitian sebelumnya dengan judul klasifikasi “Perbandingan Kinerja Metode Naive Bayes dan K-Nearest Neighbor Untuk Klasifikasi Arktikel Berbahasa Indonesia” dari 40 data uji yang dilakukan yang digunakan metode *naïve bayes* mampu mengklasifikasikan artikel jurnal berbahasa Indonesia sebanyak 28 dokumen. Sedangkan untuk metode *K-Nearest Neighbor* dari 40 data uji metode ini hanya dapat mengklasifikasikan artikel Bahasa Indonesia sebanyak 16 dokumen [16]. Dan pada penelitian ini jauh lebih berkembang dalam penggunaan data jurnal yang dipakai lebih banyak yaitu 200 data jurnal yang dibagi 4 kategori yaitu teknologi, sejarah, kesehatan, dan pendidikan. Dari 200 data jurnal masing-masing kategori mempunyai 50 dataset yang akan diklasifikasikan dalam pengujian data untuk metode *K-Nearest Neighbor* dilakukan pemecahan data training dan data uji yang dipecah menjadi beberapa bagian, kemudian pada data uji akan di test untuk implementasi algoritma *K- Nearest Neighbor* dalam mencari akurasi. Pembagian data tersebut antara lain data training 70% data uji 30%, data training 80% data uji 20%, dan data training 50% data uji 50%. Dan pada jumlah (K)nya akan ditentukan antara lain 1, 3, 5, dan 7. Untuk metode *Naïve Bayes* akan dilakukan terlebih dahulu pembagian data training dan data uji dimana akan dibagi menjadi beberapa bagian, kemudian pada data uji akan di test untuk implementasi algoritma NB dalam mencari akurasi. Pembagian data tersebut antara lain data training 70% data uji 30%, data training 80% data uji 20%, dan data training 50% data uji 50%.

## IV. KESIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan dalam menguji klasifikasi terhadap judul artikel menggunakan algoritma K-Nearest Neighbor dan Naïve Bayes. Maka hasil analisis pada penelitian ini dapat disimpulkan bahwa penelitian ini telah berhasil menguji perbandingan metode *K-Nearest Neighbor* dan *Naïve bayes* pada klasifikasi jurnal berbasis web. Dari hasil pengujian ini dapat di simpulkan bahwa pengujian analisis algoritma Naïve Bayes dan *K-Nearest Neighbor* ini didapatkan hasil akurasi terbaik dalam klasifikasi data yaitu pada algoritma K-Nearest Neighbor dimana mendapatkan akurasi sebesar 85,00% yang didapatkan pada data uji 10% dengan  $k = 3$  dibandingkan algoritma *Naïve Bayes* yang hanya mendapatkan akurasi sebesar 81,66% pada data uji 30%. Penelitian ini membuktikan bahwa pengklasifikasi judul artikel dengan menggunakan algoritma *K-NN* dan *NB* berjalan lancar dapat menghasilkan akurasi yang maksimal. Diharapkan penelitian ini kedepannya mungkin bisa mendapat peningkatan seperti pengujian klasifikasi menggunakan metode lain untuk mencapai akurasi yang lebih baik dan juga penambahan data jurnal.

## DAFTAR PUSTAKA

- [1] K. Akromunnisa and R. Hidayat, "Klasifikasi Dokumen Tugas Akhir (Skripsi) Menggunakan K-Nearest Neighbor," *JISKA (Jurnal Inform. Sunan Kalijaga)*, vol. 4, no. 1, p. 69, 2019, doi: 10.14421/jiska.2019.41-07.
- [2] M. D. Hendriyanto and B. N. Sari, "Penerapan Algoritma K-Nearest Neighbor Dalam Klasifikasi Judul Berita Hoax," *J. Ilm. Inform.*, vol. 10, no. 02, pp. 80–84, 2022, [Online]. Available: <https://ejournal.upbatam.ac.id/index.php/jif/article/view/5477>
- [3] A. C. Nisha, "Klasifikasi Abstrak Jurnal Repositor di Teknik Informatika UMM Menggunakan Metode Neighbor Weighted K-Nearest Neighbor," *J. Repos.*, vol. 3, no. 3, pp. 295–304, 2021, doi: 10.22219/repositor.v2i3.1225.
- [4] rahayu deny danar dan alvi furwanti Alwie, A. B. Prasetyo, R. Andespa, P. N. Lhokseumawe, and K. Pengantar, "Tugas Akhir Tugas Akhir," *J. Ekon. Vol. 18, Nomor 1 Maret 201*, vol. 2, no. 1, pp. 41–49, 2020.
- [5] F. Hasanah, T. Suprpti, N. Rahaningsih, and I. Ali, "Implementasi Algoritma K-Nearest Neighbor dalam Menentukan Buku Berdasarkan Peminatan," *J. Account. Inf. Syst.*, vol. 5, no. 1, pp. 102–111, 2022, doi: 10.32627/aims.v5i1.467.
- [6] A. Y. Muniar, P. Pasnur, and K. R. Lestari, "Penerapan Algoritma K-Nearest Neighbor pada Pengklasifikasian Dokumen Berita Online," *Inspir. J. Teknol. Inf. dan Komun.*, vol. 10, no. 2, p. 137, 2020, doi: 10.35585/inspir.v10i2.2570.
- [7] H. Ma'rifah, A. P. Wibawa, and M. I. Akbar, "Klasifikasi Artikel Ilmiah Dengan Berbagai Skenario Preprocessing," *Sains, Apl. Komputasi dan Teknol. Inf.*, vol. 2, no. 2, p. 70, 2020, doi: 10.30872/jsakti.v2i2.2681.
- [8] A. Indriani, "Analisa Perbandingan Metode Naïve Bayes Classifier Dan K-Nearest Neighbor Terhadap Klasifikasi Data," *Sebatik*, vol. 24, no. 1, pp. 1–7, 2020, doi: 10.46984/sebatik.v24i1.909.
- [9] R. Aziz Syahputro, Widodo, and H. Ajie, "Bayes Classifier dan Support Vector Machine dalam Klasifikasi Judul Karya Akhir Mahasiswa Program Studi PTIK UNJ," *PINTER J. Pendidik. Tek. Inform. dan Komput.*, vol. 3, no. 1, pp. 54–62, 2019, doi: 10.21009/pinter.3.1.9.
- [10] T. Adha, M. Putri, U. Enri, and B. N. Sari, "IJCIT (Indonesian Journal on Computer and Information Technology) Analisis Algoritma Naive Bayes Classifier untuk Klasifikasi Tweet Pelecehan Seksual dengan #MeToo," *IJCIT (Indonesian J. Comput. Inf. Technol.)*, vol. 5, no. 2, pp. 126–135, 2020.
- [11] D. T. Wiyanti, U. N. Semarang, P. Korespondensi, Z. Risiko, and P. Algoritma, "ANALISIS PERFORMA ALGORITMA DECISION TREE , NAÏVE BAYES , K- NEAREST NEIGHBOR UNTUK KLASIFIKASI ZONA DAERAH RISIKO COVID-19 DI INDONESIA PERFORMANCE ANALYSIS OF DECISION TREE , NAÏVE BAYES , K-NEAREST NEIGHBOR ALGORITHM FOR COVID-19 RISK ZONE CLASSIFICATI," vol. 10, no. 1, 2023, doi: 10.25126/jtiik.2023105935.
- [12] T. Rakasiwi, B. Rahayudi, and A. Ridok, "Klasifikasi Artikel Publikasi berdasarkan Judul pada Jurnal Teknologi Informasi dan Ilmu Komputer ( JTIK ) Universitas Brawijaya dengan menggunakan Metode Improved K-Nearest Neighbor," vol. 5, no. 10, pp. 4510–4516, 2021.
- [13] N. K-nn, L. Abd, R. Hakim, A. A. Rizal, and D. Ratnasari, "APLIKASI PREDIKSI KELULUSAN MAHASISWA BERBASIS K-NEAREST STUDENT GRADUATION PREDICTION USING K-NEAREST NEIGHBOR ( K-NN )," vol. 1, no. 1, pp. 30–36, 2019.
- [14] N. K. B. Kriteria, "TERHADAP LAYANAN IT PADA BANK MANDIRI MENGGUNAKAN ALGORITMA K-NEAREST ITSM TERHADAP LAYANAN IT PADA BANK MANDIRI MENGGUNAKAN ALGORITMA K-NEAREST NEIGHBOR ( K-NN ) BERDASARKAN KRITERIA ITSM Skripsi Sebagai Salah Satu Syarat untuk Memperoleh Gelar Sarjana S," 2019.
- [15] P. Seminar and N. Matematika, "Akurasi Algoritma Klasifikasi pada Software Rapidminer dan Weka," vol. 4, pp. 493–499, 2021.
- [16] R. N. Devita, H. W. Herwanto, and A. P. Wibawa, "Perbandingan Kinerja Metode Naive Bayes dan K-Nearest Neighbor untuk Klasifikasi Artikel Berbahasa indonesia," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 5, no. 4, p. 427, 2019, doi: 10.25126/jtiik.201854773.