

COMPARISON OF K-NEAREST NEIGHBOR AND NAIVE BAYES METHODS FOR CLASSIFICATION OF NEWS CONTENT

Andi Tejawati¹, Anindita Septiarini^{*2}, Rondongalo Rismawati³, Novianti Puspitasari⁴

^{1,2,3}Program Studi Informatika, Fakultas Teknik, Universitas Mulawarman

Email: ¹andi.tejawati@gmail.com, ²anindita@unmul.ac.id, ³rondongalorismawati@gmail.com,
⁴novipuspitasari@unmul.ac.id

(Naskah masuk: 8 November 2022, Revisi: 2 Desember 2022, Diterbitkan: 23 Maret 2023)

Abstract

With the development of technology, news reading via the internet or digital tends to increase. In addition, there are about 300 to 400 news articles in one month and many categories of news articles in a web portal. It makes the editor's performance more and more because an editor must be able to edit articles from various channels and at the same time have to categorize articles one by one manually into several specified categories. This study aims to compare the K-Nearest Neighbor (KNN) and Naive Bayes methods to classify news content in order to obtain the best method. The data used in this study are news articles from the web portal kaltimtoday.co from January 2022 to March 2022. Therefore 576 data are obtained. The results showed that the application of the KNN and Naive Bayes methods could be used to classify news content. The KNN method is able to produce a higher accuracy value than Naive Bayes, reaching 86% and 51% with test data of 100 news articles.

Keywords: News content classification, Text Processing, Naive Bayes, K-Nearest Neighbor, Confusion Matrix

PERBANDINGAN METODE K-NEAREST NEIGHBOR DAN NAIVE BAYES UNTUK KLASIFIKASI KONTEN BERITA

Abstrak

Semakin berkembangnya teknologi, pembacaan berita melalui internet atau digital cenderung meningkat. Selain itu, ada sekitar 300 sampai 400 artikel berita dalam satu bulan dan banyaknya kategori artikel berita dalam sebuah web portal, membuat kinerja editor semakin banyak karena seorang editor harus dapat mengedit artikel dari berbagai kanal dan sekaligus harus mengategorikan artikel satu persatu secara manual ke dalam beberapa kategori yang ditentukan. Penelitian ini bertujuan untuk melakukan perbandingan metode *K-Nearest Neighbor* (KNN) dan *Naive Bayes* untuk melakukan klasifikasi konten berita sehingga diperoleh metode yang terbaik. Data yang digunakan dalam penelitian ini berupa artikel berita dari portal web kaltimtoday.co bulan Januari 2022 hingga Maret 2022 sehingga diperoleh sebanyak 576 data. Hasil penelitian menunjukkan bahwa penerapan metode KNN dan *Naive Bayes* dapat digunakan untuk klasifikasi konten berita. Metode KNN mampu menghasilkan nilai akurasi yang lebih tinggi dibandingkan *Naive Bayes* yaitu mencapai 86% dan 51% dengan data uji sejumlah 100 artikel berita.

Kata kunci: Klasifikasi konten berita, *Text Processing*, *Naive Bayes*, *K-Nearest Neighbor*, *Confusion Matrix*

1. PENDAHULUAN

Berita merupakan informasi baru atau informasi mengenai sesuatu yang sedang terjadi, disajikan lewat bentuk cetak, siaran, internet, atau dari mulut ke mulut kepada orang ketiga atau orang banyak [1]. Teknologi komunikasi dan informasi saat ini telah berkembang mengikuti zaman dengan adanya beragam media termasuk media *online* [2]. Indonesia adalah salah satu negara dengan jumlah penduduk terpadat di dunia. Dukungan infrastruktur teknologi informasi yang memadai membuat tidak

mengherankan jika warga Indonesia kemudian menjadi pengakses media digital terbesar [3].

Tingkat pembaca berita online semakin hari semakin bertambah. Data survei pada tahun 2017 menunjukkan jumlah pembaca media online mencapai 6 juta orang dan terus mengalami peningkatan. Jika dibandingkan dengan jumlah pembaca media cetak sangat terpaut jauh yakni sebanyak 4,5 juta orang. Padahal pada tahun 2014 jumlah pembaca media cetak sebanyak 9,5 juta, sedangkan media daring hanya 3 juta [4].

Peningkatan menjadi hal menarik lainnya adalah, versi digital mampu menjangkau pembaca dari Generasi Z dengan rentang usia 10-19 tahun (17%). Mereka adalah konsumen media masa depan [5].

Data *analytics* kaltimtoday.co pada April 2022 menunjukkan bahwa pembaca berita *online* di kaltimtoday.co mencapai 228.680 pengguna dengan tayangan halaman sebanyak 339.122 dan rasio pantulan sebesar 86,42%. Hal ini menunjukkan tingginya peminat berita *online* dan tingkat kepercayaan masyarakat terhadap produk berita kaltimtoday.co. Selain itu, ada sekitar 300 sampai 400 artikel berita dalam satu bulan dan banyaknya kategori artikel atau kanal atau rubrik dalam sebuah web portal, membuat kinerja *editor* semakin banyak karena seorang *editor* harus dapat mengedit artikel dari berbagai kanal dan sekaligus harus melakukan pengkategorian artikel satu persatu secara manual ke dalam beberapa kategori yang ditentukan. Hal yang sama terjadi di kaltimtoday.co, sejak berdirinya tahun 2018 hingga saat ini, kaltimtoday.co menerbitkan berita sebanyak 150 sampai 300 berita dalam satu bulan. Semua artikel berita yang terbit akan dikategorikan secara manual oleh *editor*. Sehingga hal ini membuat kinerja *editor* di kaltimtoday.co semakin banyak.

Kategori pada sebuah artikel berita dapat diketahui secara otomatis tanpa harus dibaca satu persatu sehingga perlu dilakukan pengukuran kemiripan dokumen terkait dengan menggunakan metode *text mining* yaitu proses pengenalan teks dan dokumen. Seperti salah satu kasus terdahulu menggunakan metode tersebut untuk penelitian klasifikasi konten berita [6]. Penerapan metode klasifikasi dalam pengkategorian konten berita telah dilakukan oleh beberapa penelitian sebelumnya, dimana metode yang telah digunakan antara lain KNN [7], *Decision Tree* [8], dan *Naïve Bayes* [9].

Metode *Naïve Bayes* telah digunakan dalam penelitian tentang klasifikasi konten berita. *Naïve Bayes* merupakan teknik prediksi probabilitistik sederhana yang berdasar pada penerapan teorema *Bayes* dengan asumsi independensi yang kuat [10]. *Naïve Bayes* adalah salah satu algoritma yang paling efisien dan efektif untuk pembelajaran mesin dan penambahan data [11]. Metode ini memiliki sifat yang efektif dan cepat dalam mengolah data berjumlah besar [12]. Kemampuan tersebut membuat metode ini sering digunakan pada aplikasi seperti *spam filtering*, dan deteksi anomali di jaringan komputer [13].

Metode klasifikasi lain yang digunakan yaitu KNN. Metode KNN merupakan algoritma yang dapat digunakan untuk klasifikasi teks dan data [14], metode ini digunakan karena memiliki keunggulan dalam mengklasifikasi data berita yang tidak diketahui dengan adanya data latih dan data uji [15]. KNN dapat mem prosedur yang berbasis matematis untuk mengevaluasi nilai kriteria-kriteria tersebut menjadi sebuah keterangan klasifikasi [16]. Metode

KNN dapat mengklasifikasi data secara akurat dengan memilih terlebih dahulu nilai K tetangga terdekat dengan tepat [17], [18]. KNN juga dapat memilah kumpulan data calon yang dapat digolongkan menjadi baik, paling baik, dan kurang baik [15].

Penggunaan metode *Naïve Bayes* dilakukan dengan menghitung probabilitas data latih dan data uji yang diperoleh dari data berita *online* di portal kaltimtoday.co, sedangkan metode KNN dilakukan dengan pendekatan jarak dengan tetangga terdekat dimana dengan menggunakan metode tersebut dapat menghasilkan klasifikasi konten berita ke dalam 12 kategori yaitu *economics*, *sport*, *scholar*, *event*, *entertainment*, *healthy*, *incident*, *travel*, *education*, *politic*, *crime*, dan *figure*. Hal ini berguna dalam proses klasifikasi konten berita yang dapat membantu mempermudah kinerja *editor* dalam perusahaan pemberitaan.

Oleh sebab itu, penelitian ini mengangkat judul “Perbandingan Metode *Naïve Bayes* dan *K-Nearest Neighbour* untuk Klasifikasi Konten Berita”. Metode yang dapat digunakan pada pengklasifikasian konten berita ini adalah *naïve bayes* dan KNN. Penelitian ini bertujuan untuk melihat dan membandingkan metode yang lebih baik dalam proses klasifikasi konten berita.

2. METODE PENELITIAN

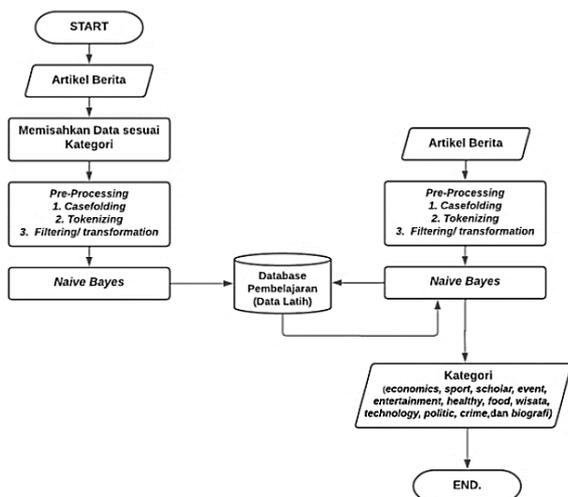
Data yang dikumpulkan pada penelitian ini yaitu data primer. Data primer merupakan data yang diperoleh langsung dari sumbernya, sedangkan data sekunder adalah data yang diperoleh bukan secara langsung dari sumbernya. Penelitian ini sumber data primer yang dipakai adalah sumber portal web kaltimtoday.co. Data latih yang dikumpulkan berupa berita yang terbit pada bulan Januari – Maret 2021 sebanyak 576 berita, sedangkan data uji yang digunakan berupa berita yang terbit pada bulan Oktober 2021 sebanyak 100 berita.

Adapun perancangan data pada penelitian ini adalah sebagai berikut:

- 1) Setelah melakukan pengambilan data dari website, masing-masing data dibagi sesuai dengan kategorinya. Kemudian data yang telah siap akan diolah pada proses *case folding*, *tokenizing*, *stemming*, dan *stopword*.
- 2) Data yang digunakan pada penelitian ini adalah data berita dari website kaltimtoday.co. Data tersebut digunakan sebagai data latih dan data uji pada model yang akan dibuat. Data latih yang digunakan berjumlah 12 data untuk setiap kategori.
- 3) Kelas yang akan digunakan pada penelitian ini berjumlah 12 (dua belas) kelas yaitu *economics*, *sport*, *scholar*, *event*, *entertainment*, *healthy*, *incident*, *travel*, *education*, *politic*, *crime*, dan *figure*.

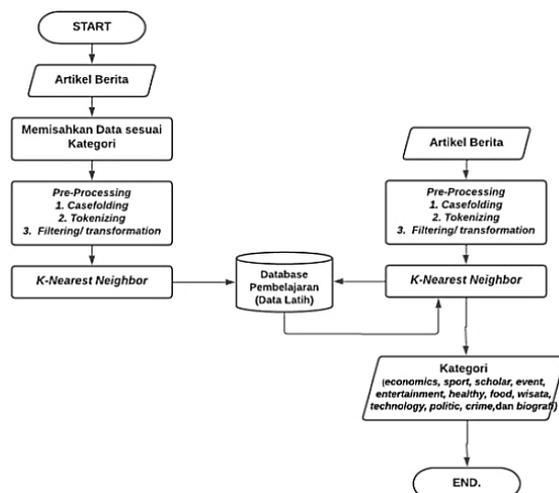
4) Penelitian ini menggunakan metode *Naive Bayes* dan *K-Nearest Neighbour* (KNN). Berikut penjelasan algoritma setiap metode.

Algoritma klasifikasi pada penelitian ini menggunakan metode *Naive Bayes Classifier*. Rancangan proses sistem dijelaskan dalam bentuk *flowchart* pada Gambar 1.



Gambar 1. Rancangan Proses Sistem Klasifikasi dengan Metode *Naive Bayes*

Gambar 1 merupakan rancangan proses sistem klasifikasi menggunakan metode *Naive Bayes* yang akan dibangun, terdapat dua proses yang berbeda yaitu proses pembelajaran dan proses pengujian. Algoritma klasifikasi pada penelitian ini menggunakan metode KNN rancangan proses sistem KNN dapat dilihat pada Gambar 2.



Gambar 2. Rancangan Proses Sistem Klasifikasi dengan Metode KNN

Klasifikasi konten berita dibuat menggunakan *software Orange* dengan menerapkan metode *Naive Bayes* dan KNN. Tampilan *software orange* 3.31 yang digunakan untuk melakukan klasifikasi konten berita dapat dilihat pada Gambar 3.



Gambar 3. Tampilan *Software Orange*

3. HASIL DAN PEMBAHASAN

Data yang digunakan dalam penelitian ini berupa data berita dari portal kaltimtoday.co pada bulan Januari – Maret 2022 sebanyak 576 data sebagai data latih kemudian berita pada bulan September 2022 sebanyak 100 data sebagai data uji. Data yang digunakan merupakan isi berita yang kemudian dimuat dalam format **txt* dan mengambil 12 kategori yaitu *economics, sport, scholar, event, entertainment, healthy, incident, travel, education, politic, crime, dan figure*.

3.1. Text Preprocessing

Setelah dilakukan pengambilan dan penyalinan data dari portal berita, masing-masing data dibagi sesuai dengan kategori. Tahap selanjutnya adalah mengubah data menjadi data yang siap dipakai. Kemudian data yang telah siap akan diolah pada proses *case folding, tokenizing, stemming, dan stopword*. Berikut adalah proses yang dilakukan:

a) *Tokenizing*

Berdasarkan data yang telah diperoleh terlebih dahulu dilakukan *tokenizing* untuk proses pemotongan dari kalimat sehingga terpisah tiap katanya serta menyaring tanda baca yang ada dalam sebuah kalimat. Proses *tokenizing* seperti sebagai berikut.

Sebelum:

Seorang pria yang sudah berusia lanjut BG (66) membunuh wanita simpanannya TSW (44) yang tinggal di Gang Assadiyah 4 RT 13, Kecamatan Sangatta Utara karena tersinggung disebut menjijikkan dan kere. BG kemudian emosi dan nekat menghabiskan TSW. Hal ini bermula saat TSW meminta uang dua juta kepada BG. Namun, BG belum mengabulkan. Apalagi sebelumnya BG sudah memberikan uang sebesar Rp 18 juta dan mendengar kabar kalau simpanannya itu selingkuh dengan pria lain. "Saat itu saya khilaf karena sebelumnya saya sudah kasih Rp 18 Juta kemudian minta lagi Rp 2 juta tapi saya lagi tidak punya uang, malah dikatakan kere dan menjijikkan, ditambah adanya kabar kalau sekarang dia ada main

dengan laki-laki lain," jelas BG yang dihadirkan dalam *press release*, Jumat (21/1/2022). Atas perkataan itu, BG naik pitam. Dia kemudian mencekik korban. Korban meninggal dunia. Selanjutnya, tersangka meninggalkan korban dengan menutup wajahnya dengan bantal dan selimut. "Usai membunuh, pintu rumah digembok dari luar dan tersangka meninggalkan lokasi," ujar Kapolres Kutim, AKBP Welly Djatmiko didampingi Kasat Reskrim Iptu I Made Jata Wiranegara, dan Humas Polres, Wahyu Winarko. Selanjutnya kata Kapolres Welly, pelaku berniat melakukan bunuh diri di Sungai Sangatta wilayah Kampung Kajang. Namun niatnya tak berhasil. Dia diselamatkan oleh warga. Kemudian dia dibawa ke Puskesmas terdekat. "Selesai diberikan pengobatan, tersangka kembali kabur. Awalnya kami lacak di Paser, namun berhasil diamankan oleh Tim Gabungan Jatanras Polda Kalsel dan Polda Kalteng di Palangkaraya," papar Kapolres. BG ditetapkan sebagai tersangka pembunuhan TSW berdasarkan hasil olah TKP dan pemeriksaan bukti-bukti yang ditemukan. Adapun, BG dan TSW berhubungan sebagai teman kencan atau simpanan dan selalu dibiayai kebutuhan sehari-harinya oleh BG sendiri. "Bisa dikatakan pasangan kencan, tak ada ikatan suami istri," tambah AKBP Welly. Atas perbuatannya, pelaku dijerat dengan pasal 338 KUHP tentang pembunuhan dengan ancaman maksimal hukuman penjara selama 15 tahun penjara. "Tetapi kami masih dalam lagi. Apakah masuk dalam pembunuhan berencana atau tidak. Kalau berencana ancamannya hukuman seumur hidup atau mati," tutupnya.

Sesudah:

Seorang pria yang sudah berusia lanjut BG (66) membunuh wanita simpanannya TSW (44) yang tinggal di Gang Assadiyah 4 RT 13, Kecamatan Sangatta Utara karena tersinggung disebut menjijikkan dan kere. BG kemudian emosi dan nekat menghabisi TSW. Hal ini bermula saat TSW meminta uang dua juta kepada BG. Namun, BG belum mengabdikan. Apalagi sebelumnya BG sudah memberikan uang sebesar Rp 18 juta dan mendengar kabar kalau simpanannya itu selingkuh dengan pria lain. "Saat itu saya khilaf karena sebelumnya saya sudah kasih Rp 18 Juta kemudian minta lagi Rp 2 juta tapi saya lagi tidak punya uang, malah dikatakan kere dan menjijikkan, ditambah adanya kabar kalau sekarang dia ada main dengan laki-laki lain," jelas BG yang dihadirkan dalam *press release*, Jumat (21/1/2022). Atas perkataan itu, BG naik pitam. Dia kemudian mencekik korban.

Korban meninggal dunia. Selanjutnya, tersangka meninggalkan korban dengan menutup wajahnya dengan bantal dan selimut. "Usai membunuh, pintu rumah digembok dari luar dan tersangka meninggalkan lokasi," ujar Kapolres Kutim, AKBP Welly Djatmiko didampingi Kasat Reskrim Iptu I Made Jata Wiranegara, dan Humas Polres, Wahyu Winarko. Selanjutnya kata Kapolres Welly, pelaku berniat melakukan bunuh diri di Sungai Sangatta wilayah Kampung Kajang. Namun niatnya tak berhasil. Dia diselamatkan oleh warga. Kemudian dia dibawa ke Puskesmas terdekat. "Selesai diberikan pengobatan, tersangka kembali kabur. Awalnya kami lacak di Paser, namun berhasil diamankan oleh Tim Gabungan Jatanras Polda Kalsel dan Polda Kalteng di Palangkaraya," papar Kapolres. BG ditetapkan sebagai tersangka pembunuhan TSW berdasarkan hasil olah TKP dan pemeriksaan bukti-bukti yang ditemukan. Adapun, BG dan TSW berhubungan sebagai teman kencan atau simpanan dan selalu dibiayai kebutuhan sehari-harinya oleh BG sendiri. "Bisa dikatakan pasangan kencan, tak ada ikatan suami istri," tambah AKBP Welly. Atas perbuatannya, pelaku dijerat dengan pasal 338 KUHP tentang pembunuhan dengan ancaman maksimal hukuman penjara selama 15 tahun penjara. "Tetapi kami masih dalam lagi. Apakah masuk dalam pembunuhan berencana atau tidak. Kalau berencana ancamannya hukuman seumur hidup atau mati," tutupnya.

b) *Stopword*

Pada tahap ini dilakukan proses *preprocessing* selanjutnya yakni *stopword*, penyaringan kata dari proses *tokenizing* dan membuang kata yang tidak penting. Adapun contoh dari kondisi sebelum dan sesudah dilakukan proses *stopword* sebagai berikut.

Sebelum:

| Seorang | pria | yang | sudah | berusia | lanjut | BG | 66 | membunuh | wanita | simpanannya | TSW | 44 | yang | tinggal | di | Gang | Assadiyah | 4 | RT | 13 | Kecamatan | Sangatta | Utara | karena | tersinggung | disebut | menjijikkan | dan | kere | BG | kemudian | emosi | dan | nekat | menghabisi | TSW | Hal | ini | bermula | saat | TSW | meminta | uang | dua | juta | kepada | BG | Namun | BG | belum | mengabdikan | Apalagi | sebelumnya | BG | sudah | memberikan | uang | sebesar | Rp | 18 | juta | dan | mendengar | kabar | jikalau | simpanannya | itu | selingkuh | dengan | pria | lain | Saat | itu | saya | khilaf | karena | sebelumnya | saya | sudah | kasih | Rp | 18 | Juta | kemudian | minta | lagi | Rp | 2 | juta | tapi |

saya | lagi | tidak | punya | uang | malah | dikatakan | kere | dan | menjijikkan | ditambah | adanya | kabar | kalau | sekarang | dia | ada | main | dengan | laki-laki | lain | jelas | BG | yang | dihadirkan | dalam | press | release | Jumat | 21 | 1 | 2022 | Atas | perkataan | itu | BG | naik | pitam | Dia | kemudian | mencekik | korban | Korban | meninggal | dunia | Selanjutnya | tersangka | meninggalkan | korban | dengan | menutup | wajahnya | dengan | bantal | dan | selimut | Usai | membunuh | pintu | rumah | digembok | dari | luar | dan | tersangka | meninggalkan | lokasi | ujar | Kapolres | Kutim | AKBP | Welly | Djatmiko | didampingi | Kasat | Reskrim | Iptu | I | Made | Jata | Wiranegara | dan | Humas | Polres | Wahyu | Winarko | Selanjutnya | kata | Kapolres | Welly | pelaku | berniat | melakukan | bunuh | diri | di | Sungai | Sangatta | wilayah | Kampung | Kajang | Namun | niatnya | tak | berhasil | Dia | diselamatkan | oleh | warga | Kemudian | dia | dibawa | ke | Puskesmas | terdekat | Selesai | diberikan | pengobatan | tersangka | kembali | kabur | Awalnya | kami | lacak | di | Paser | namun | berhasil | diamankan | oleh | Tim | Gabungan | Jatanras | Polda | Kalsel | dan | Polda | Kalteng | di | Palangkaraya | papar | Kapolres | BG | ditetapkan | sebagai | tersangka | pembunuhan | TSW | berdasarkan | hasil | olah | TKP | dan | pemeriksaan | bukti-bukti | yang | ditemukan | Adapun | BG | dan | TSW | berhubungan | sebagai | teman | kencan | atau | simpanan | dan | selalu | dibiayai | kebutuhan | sehari-harinya | oleh | BG | sendiri | Bisa | dikatakan | pasangan | kencan | tak | ada | ikatan | suami | istri | tambah | AKBP | Welly | Atas | perbuatannya | pelaku | dijerat | dengan | pasal | 338 | KUHP | tentang | pembunuhan | dengan | ancaman | maksimal | hukuman | penjara | selama | 15 | tahun | penjara | Tetapi | kami | masih | dalam | lagi | Apakah | masuk | dalam | pembunuhan | berencana | atau | tidak | Kalau | berencana | ancamannya | hukuman | seumur | hidup | atau | mati | tutupnya |.

Sesudah:

| berusia | membunuh | wanita | simpanannya | tersinggung | menjijikkan | kere | emosi | nekat | uang | juta | kabar | selingkuh | pria | khilaf | press | release | pitam | mencekik | korban | meninggal | dunia | tersangka | menutup | wajahnya | bantal | selimut | pintu | rumah | digembok | tersangka | meninggalkan | lokasi | kapolres | akbp | Kasat | Reskrim | Polres | pelaku | diselamatkan | pengobatan | kabur | pemeriksaan | bukti | kencan | dijerat | pasal | kuhp | ancaman | hukuman | penjara |.

c) *Stemming*

Proses *preprocessing* selanjutnya yaitu *stemming*, dimana pada tahap ini akan dilakukan proses pengubahan kata sebelumnya menjadi kata dasar dari proses *tokenizing*. Berikut contoh dari kondisi sebelum dan sesudah dilakukan proses *stemming*.

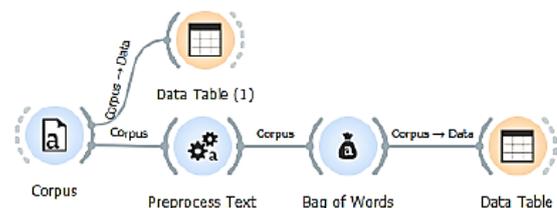
Sebelum:

| berusia | membunuh | wanita | tersinggung | menjijikkan | kere | emosi | nekat | uang | juta | kabar | selingkuh | pria | khilaf | press | release | pitam | mencekik | korban | meninggal | dunia | tersangka | menutup | wajahnya | bantal | selimut | pintu | rumah | digembok | tersangka | meninggalkan | lokasi | kapolres | akbp | Kasat | Reskrim | Polres | pelaku | diselamatkan | pengobatan | kabur | pemeriksaan | bukti | kencan | dijerat | pasal | kuhp | ancaman | hukuman | penjara |.

Sesudah:

| usia | bunuh | wanita | singgung | jijik | kere | emosi | nekat | uang | juta | kabar | selingkuh | pria | khilaf | press | release | pitam | cekik | korban | meninggal | dunia | tersangka | tutup | wajah | bantal | selimut | pintu | rumah | gembok | tersangka | lokasi | kapolres | akbp | kasat | reskrim | polres | pelaku | selamat | obat | kabur | periksa | bukti | kencan | jerat | pasal | kuhp | ancam | hukum | penjara |.

Proses pengerjaan *text preprocessing* dilakukan juga di dalam *software Orange*. Pengumpulan data dilakukan dengan menyalin 576 teks berita sebagai data latih dan 100 teks berita sebagai data uji dari portal berita kaltimtoday.co lalu di masukan ke dalam file excel. Selanjutnya setelah data mentah terkumpul dimasukan ke dalam *software Orange* untuk dilakukan *text preprocessing* dengan alur seperti pada Gambar 4.



Gambar 4. Proses *Text Preprocessing* Menggunakan *Software Orange*

Data yang telah melalui proses *text preprocessing* berupa potongan kata. Data yang telah melalui proses *tokenizing*, *stopword*, dan *stemming* selanjutnya data tersebut menjadi data yang terstruktur dan siap untuk dilakukan proses selanjutnya.

3.2. Perhitungan Metode Naive Bayes

a) Proses Pembelajaran

Pada proses pembelajaran, data berita yang telah direkap dalam *excel* akan diklasifikasikan berdasarkan kategori, setelah itu data diproses ke tahap *preprocessing*, kemudian tahap menghitung frekuensi kemunculan kata, dan terakhir menghitung nilai probabilitas. Hasil dari proses *text preprocessing*, maka didapat jumlah dokumen sebanyak 576 dan jumlah frekuensi dari masing-masing kategori *politic* = 89, jumlah frekuensi *economics* = 95, jumlah frekuensi *crime* = 48, jumlah frekuensi *healthy* = 92, jumlah frekuensi *education* = 33, jumlah frekuensi *travel* = 6, jumlah frekuensi *event* = 134, jumlah frekuensi *sport* = 7, jumlah frekuensi *entertainment* = 7, jumlah frekuensi *biografi* = 5, jumlah frekuensi *incident* = 55, dan jumlah frekuensi *scholar* = 5. Selanjutnya adalah menghitung nilai probabilitas dengan menggunakan persamaan 1 sebagai berikut.

$$P(W_k|V_j) = \frac{(n_k + 1)}{Jlm\ Frekuensi + Jlm\ Dokumen} \quad (1)$$

$P(W_k|V_j)$ merupakan probabilitas bobot kata sesuai kategori dan n_k adalah Nilai kemunculan frekuensi kata. Sebagai contoh digunakan data dari kategori *scholar*. Hasil *preprocessing* dari kategori *scholar* dapat dilihat sebagai berikut:

|formulir|negeri|universitas|pendidikan|syarat|program|lamar|kominfo|daftar|

Kemudian menghitung nilai probabilitas dengan hasil sebagai berikut:

$$\begin{aligned} P(\text{formulir} | \text{Scholar}) &= (1+1) / (5+576) = 0,0034 \\ P(\text{negeri} | \text{Scholar}) &= (1+1) / (5+576) = 0,0034 \\ P(\text{universitas} | \text{Scholar}) &= (1+1) / (5+576) = 0,0034 \\ P(\text{pendidikan} | \text{Scholar}) &= (1+1) / (5+576) = 0,0034 \\ P(\text{syarat} | \text{Scholar}) &= (1+1) / (5+576) = 0,0034 \\ P(\text{program} | \text{Scholar}) &= (1+1) / (5+576) = 0,0034 \\ P(\text{lamar} | \text{Scholar}) &= (1+1) / (5+576) = 0,0034 \\ P(\text{kominfo} | \text{Scholar}) &= (1+1) / (5+576) = 0,0034 \\ P(\text{daftar} | \text{Scholar}) &= (1+1) / (5+576) = 0,0034 \\ P(\text{beasiswa} | \text{Scholar}) &= (1+1) / (5+576) = 0,0034 \end{aligned}$$

Perhitungan probabilitas bobot kata yang sama dilakukan untuk semua kategori. Hasil perhitungan untuk semua kategori dapat dilihat pada Tabel 1.

Tabel 1. Nilai Probabilitas Kata untuk Seluruh Kategori

Probabilitas	Kata								
	Formula	Negeri	Universitas	Pendidikan	Syarat	Program	Lamar	Kominfo	Daftar
<i>Politic</i>	0,0015	0,0015	0,0015	0,0015	0,0015	0,0015	0,0015	0,0015	0,0015
<i>Economic</i>	0,0014	0,0014	0,0014	0,0014	0,0014	0,0014	0,0014	0,0014	0,0014
<i>Crime</i>	0,0016	0,0016	0,0016	0,0016	0,0016	0,0016	0,0016	0,0016	0,0016
<i>Healthy</i>	0,0014	0,0014	0,0014	0,0014	0,0014	0,0014	0,0014	0,0014	0,0014
<i>Education</i>	0,0016	0,0016	0,0016	0,0016	0,0016	0,0016	0,0016	0,0016	0,0016
<i>Travel</i>	0,0017	0,0017	0,0017	0,0017	0,0017	0,0017	0,0017	0,0017	0,0017
<i>Event</i>	0,0014	0,0014	0,0014	0,0014	0,0014	0,0014	0,0014	0,0014	0,0014

Setelah didapat nilai probabilitas kata pada setiap kategori, kemudian menghitung probabilitas kategori dengan menggunakan persamaan (2) sebagai berikut.

$$(PV_j) = \frac{Jml\ dokumen\ kategori}{Total\ Dokumen} \quad (2)$$

Diketahui:

- Jumlah dokumen *politic* = 89
 - Jumlah dokumen *economics* = 95
 - Jumlah dokumen *crime* = 48
 - Jumlah dokumen *healthy* = 92
 - Jumlah dokumen *education* = 33
 - Jumlah dokumen *travel* = 6
 - Jumlah dokumen *event* = 134
 - Jumlah dokumen *sport* = 7
 - Jumlah dokumen *entertainment* = 7
 - Jumlah dokumen *biografi* = 5
 - Jumlah dokumen *incident* = 5
 - Jumlah dokumen *scholar* = 5
- Jadi, probabilitas dari dokumen adalah:

$$\begin{aligned} P(\text{politic}) &= 89/576 &= 0.1545 \\ P(\text{economics}) &= 95/576 &= 0.1649 \\ P(\text{crime}) &= 48/576 &= 0.0833 \\ P(\text{healthy}) &= 92/576 &= 0.1597 \\ P(\text{education}) &= 33/576 &= 0.0572 \\ P(\text{travel}) &= 6/576 &= 0.0104 \\ P(\text{event}) &= 134/576 &= 0.2326 \\ P(\text{sport}) &= 7/576 &= 0.0121 \\ P(\text{entertainment}) &= 7/576 &= 0.0121 \\ P(\text{biografi}) &= 5/576 &= 0.0086 \\ P(\text{incident}) &= 55/576 &= 0.0954 \\ P(\text{scholar}) &= 5/576 &= 0.0086 \end{aligned}$$

b) Proses Pengujian

Proses pengujian dilakukan dengan menggunakan dokumen baru yang belum diketahui kategorinya. Adapun contoh data terstruktur yang digunakan dalam perhitungan *naive bayes* seperti pada Tabel 2.

Tabel 2. Contoh Data Terstruktur untuk Perhitungan *Naive Bayes*

KATA	DOK1	DOK2	DOK3	DOK4	DOK5	TES1	TES2	TES3
KATA1	anggaran	formulir	mahasiswa	alam	akademik	mahasiswa	program	politisi
KATA2	data	negeri	taahfidz	antisipasi	daring	kuliah	potensi	anggota
KATA3	ketua	universitas	prestasi	dinas	kuliah	miliar	izin	apbd
KATA4	langsung	belajar	rakyat	jalan	kampus	nilai	hubungan	didik
KATA5	bantu	syarat	juz	program	bijak	daftar	dinas	legeslatif
KATA6	resmi	program	angkatan	warga	tetap	reguler	perusahaan	dprd
KATA7	tim	lamar	jenjang	longsor	muka	anggaran	ekonomi	sidang
KATA8	program	kominfo	belajar	pinggir	praktik	belajar	bupati	Bijak
KATA9	bantu	daftar	program	kejadian	protokol	pemerintah	anggaran	dana
KATA10	tunggu	peserta	beasiswa	syarat	ukt	gratis	bandara	raperda
KATEGORI	<i>Economic</i>	<i>Scholar</i>	<i>Scholar</i>	<i>Insident</i>	<i>Education</i>	?	?	?

- c) Perhitungan Dokumen TES1
- Selanjutnya dihitung nilai probabilitas pada setiap kategori sebagai berikut:
- Kategori *economics*
 - $P(\text{mahasiswa} | \text{economics}) = 0,166666667$
 - $P(\text{kuliah} | \text{economics}) = 0,166666667$
 - $P(\text{miliar} | \text{economics}) = 0,166666667$
 - $P(\text{nilai} | \text{economics}) = 0,166666667$
 - $P(\text{daftar} | \text{economics}) = 0,166666667$
 - $P(\text{reguler} | \text{economics}) = 0,166666667$
 - $P(\text{anggaran} | \text{economics}) = 0,333333333$
 - $P(\text{belajar} | \text{economics}) = 0,166666667$
 - $P(\text{pemerintah} | \text{economics}) = 0,166666667$
 - $P(\text{gratis} | \text{economics}) = 0,166666667$
 - Jadi $P(/ \text{economics}) = 0,166666667 \times 0,166666667 \times 0,166666667 \times 0,166666667 \times 0,166666667 \times 0,166666667 \times 0,333333333 \times 0,166666667 \times 0,166666667 \times 0,166666667 = 2,61344 \times 10^{-10}$
 - Probabilitas = $P(\text{economics}) \times P(/ \text{economics}) = 0,2 \times 2,61344 \times 10^{-10} = 5,22688 \times 10^{-11}$
 - Kategori *education*
 - $P(\text{mahasiswa} | \text{education}) = 0,166666667$
 - $P(\text{kuliah} | \text{education}) = 0,333333333$
 - $P(\text{miliar} | \text{education}) = 0,166666667$
 - $P(\text{nilai} | \text{education}) = 0,166666667$
 - $P(\text{daftar} | \text{education}) = 0,166666667$
 - $P(\text{reguler} | \text{education}) = 0,166666667$
 - $P(\text{anggaran} | \text{education}) = 0,166666667$
 - $P(\text{belajar} | \text{education}) = 0,166666667$
 - $P(\text{pemerintah} | \text{education}) = 0,166666667$
 - $P(\text{gratis} | \text{education}) = 0,166666667$
 - Jadi $P(/ \text{education}) = 0,166666667 \times 0,333333333 \times 0,166666667 = 3,31 \times 10^{-8}$
 - Probabilitas = $P(\text{education}) \times P(/ \text{education}) = 0,2 \times 3,31 \times 10^{-8} = 6,61527 \times 10^{-9}$
 - Kategori *incident*
 - $P(\text{mahasiswa} | \text{incident}) = 0,166666667$
- $P(\text{kuliah} | \text{incident}) = 0,166666667$
- $P(\text{miliar} | \text{incident}) = 0,166666667$
- $P(\text{nilai} | \text{incident}) = 0,166666667$
- $P(\text{daftar} | \text{incident}) = 0,166666667$
- $P(\text{reguler} | \text{incident}) = 0,166666667$
- $P(\text{anggaran} | \text{incident}) = 0,166666667$
- $P(\text{belajar} | \text{incident}) = 0,166666667$
- $P(\text{pemerintah} | \text{incident}) = 0,166666667$
- $P(\text{gratis} | \text{incident}) = 0,166666667$
- Jadi $P(/ \text{incident}) = 0,166666667 \times 0,166666667 = 2,83211 \times 10^{-9}$
- Probabilitas = $P(\text{incident}) \times P(/ \text{incident}) = 0,2 \times 2,83211 \times 10^{-9} = 3,30763 \times 10^{-9}$
- Kategori *scholar*
 - $P(\text{mahasiswa} | \text{scholar}) = 0,285714286$
 - $P(\text{kuliah} | \text{scholar}) = 0,142857143$
 - $P(\text{miliar} | \text{scholar}) = 0,142857143$
 - $P(\text{nilai} | \text{scholar}) = 0,142857143$
 - $P(\text{daftar} | \text{scholar}) = 0,285714286$
 - $P(\text{reguler} | \text{scholar}) = 0,142857143$
 - $P(\text{anggaran} | \text{scholar}) = 0,142857143$
 - $P(\text{belajar} | \text{scholar}) = 0,285714286$
 - $P(\text{pemerintah} | \text{scholar}) = 0,142857143$
 - $P(\text{gratis} | \text{scholar}) = 0,142857143$
 - Jadi $P(/ \text{scholar}) = 0,285714286 \times 0,142857143 \times 0,142857143 \times 0,142857143 \times 0,285714286 \times 0,142857143 \times 0,142857143 \times 0,285714286 \times 0,142857143 \times 0,142857143 = 2,83 \times 10^{-9}$
 - Probabilitas = $P(\text{scholar}) \times P(/ \text{scholar}) = 0,4 \times 1,13284 \times 10^{-9} = 1,13284 \times 10^{-9}$
- Jadi, kategori dari dokumen berita baru di atas termasuk kategori *education* karena probabilitas paling tinggi yaitu $6,61527 \times 10^{-9}$.

3.3. Perhitungan Metode KNN

Perhitungan metode KNN dilakukan dengan dua tahap yaitu tahap pembobotan dengan *feature extraction* dan tahap perhitungan KNN. Adapun proses perhitungan manual klasifikasi konten berita menggunakan metode KNN dapat dilihat sebagai berikut:

a) *Feature Extraction*

Pada tahap awal sebelum dilakukan proses perhitungan KNN, terlebih dahulu dilakukan *feature extraction* yaitu mengubah data *text* hasil *preprocessing* menjadi nilai. Untuk menjadikan data berita menjadi nilai maka *feature extraction* yang digunakan adalah *Term Frequency (TF) Binary*. Berikut adalah langkah dalam proses *feature extraction*:

Pertama membuat atribut pada data *training* yang akan digunakan untuk proses TF Binary. Atribut didapatkan dari kata-kata yang terdapat pada semua data berita *training*. Tetapi tidak semua kata yang ada pada berita dijadikan atribut, kata yang diambil adalah kata yang berhubungan dengan beasiswa dan juga kata yang memiliki makna.

Atribut yang diambil dari kata berita *scholar* terdapat pada Tabel 3.

Tabel 3. Atribut dan Data *Training*

Nomor	Atribut
1	beasiswa
2	belajar
3	program
4	kuliah
5	formulir
6	universitas
7	syarat
8	anggaran
9	nilai
10	reguler

Kedua melakukan proses *TF Binary*, dimana pada tahap ini dilakukan pengecekan *record* yang terdapat pada data *training* dan data *testing*. Pengecekan *record* dilakukan satu persatu dengan mengacu pada atribut yang telah dibuat pada Tabel 6. Data berita dari hasil *text preprocessing* yang digunakan dapat dilihat pada Tabel 4. Maka hasil yang didapat setelah dilakukan proses *feature extraction* menggunakan TF Binary dapat dilihat pada Tabel 5. Apabila kata yang terdapat pada data *text* sesuai dengan atribut maka bernilai 1, sedangkan jika tidak sesuai bernilai 0.

Tabel 4. Data Berita Hasil *Text Preprocessing*

Dok	Kata1	Kata2	Kata3	Kata4	Kata5	Kata6	Kata7	Kata8	Kata9	Kata10
D1	anggaran	data	ketua	langsung	bantu	reuni	tim	program	bantu	tunggu
D2	formulir	negeri	universitas	belajar	syarat	program	lamar	kominfo	daftar	peserta
D3	mahasiswa	tahfidz	prestasi	rakyat	juz	angkatan	jenjang	belajar	program	beasiswa
D4	alam	antisipasi	dinas	jalan	program	warga	longsor	pinggir	kejadian	syarat
D5	akademik	daring	kuliah	kampus	bijak	tatap	muka	praktik	protokol	ukt
TES1	mahasiswa	kuliah	miliar	nilai	daftar	reguler	anggaran	belajar	pemerintah	gratis
TES2	program	petani	izin	hubung	dinas	perusahaan	ekonomi	bupati	anggaran	bandara
TES3	politisi	anggota	apbd	didik	legislatif	dprd	sidang	bijak	dana	raperda

Tabel 5. Hasil dari TF Binary

Dok	beasiswa	belajar	program	kuliah	formulir	universitas	syarat	anggaran	nilai	reguler
D1	0	0	1	0	0	0	0	1	0	0
D2	1	0	1	0	1	1	1	0	0	0
D3	1	1	1	0	0	0	0	0	0	0
D4	0	0	1	0	0	0	1	0	0	0
D5	0	0	0	1	0	0	0	0	0	0
TES1	0	1	0	1	0	0	0	1	1	1
TES2	1	0	0	0	0	0	0	0	1	0
TES3	0	0	0	0	0	0	0	0	0	0

b) Perhitungan Dokumen TEST1

Menetapkan parameter k (jumlah tetangga terdekat). Parameter k yang digunakan adalah 3. Perhitungan kuadrat jarak *euclidean distance* menggunakan persamaan (3) sebagai berikut.

$$dist = \sqrt{((train(i, k)) - (test(j, k)))^2} \quad (3)$$

Keterangan:

train(i, k) = seluruh *record* dan atribut yang ada pada data *training*

test(j, k) = seluruh *record* dan atribut yang ada pada data *testing*

Dokumen 1

$$dist = \sqrt{((0 - 0) + (0 - 1) + (1 - 0) + (0 + 1) + (0 - 0) + (0 - 0) + (0 - 0) + (1 - 1) + (0 - 1) + (0 - 1))^2} = 2,236067977$$

Dokumen 2

$$dist = \sqrt{((1 - 0) + (0 - 1) + (1 - 0) + (0 + 1) + (1 - 0) + (1 - 0) + (1 - 0) + (0 - 1) + (0 - 1) + (0 - 1))^2} = 3,16227766$$

Dokumen 3

$$dist = \sqrt{((1 - 0) + (1 - 1) + (1 - 0) + (0 - 1) + (0 - 0) + (0 - 0) + (0 - 0) + (0 - 1) + (0 - 1) + (0 - 1))^2} = 2,449489743$$

Dokumen 4

$$dist = \sqrt{((0 - 0) + (0 - 1) + (1 - 0) + (0 - 1) + (0 - 0) + (0 - 0) + (1 - 0) + (0 - 1) + (0 - 1) + (0 - 1))^2}$$

$$= 2,645751311$$

Dokumen 5

$$dist = \sqrt{((0 - 0) + (0 - 1) + (0 - 0) + (1 - 1) + (0 - 0) + (0 - 0) + (0 - 0) + (0 - 1) + (0 - 1) + (0 - 1))^2}$$

$$= 2$$

Kemudian mengurutkan hasil dari jarak data *testing* dengan data *training* secara *ascending* yaitu mengurutkan jarak paling kecil ke jarak yang terbesar. Selanjutnya menentukan tetangga terdekat berdasarkan jarak minimum K yang telah ditentukan di tahap pertama yaitu 3. Hasil pengurutan jarak terdekat dari jarak data *testing* dengan data *training* dapat dilihat pada Tabel 6.

Tabel 6. Pengurutan Jarak Terdekat Data *Testing* dan *Training*

Dokumen	dist	Ranking	Nilai K≤3
1	2,236067977	2	Ya (K<3)
2	3,16227766	5	Tidak (K>3)
3	2,449489743	3	Ya (K=3)
4	2,645751311	4	Tidak (K>3)
5	2	1	Ya (K<3)

Menentukan kategori dari tetangga terdekat. Kategori Ya diambil jika nilai $K \leq 3$. Jadi, dokumen 1, 3, dan 5 termasuk kategori Ya dan sisanya Tidak. Hasil dari penentuan kategori yang termasuk $K=3$ dapat dilihat pada Tabel 7.

Tabel 7. Penentuan Kategori yang Termasuk K=3

Dok	dist	Ranking	Nilai K≤3	Kategori Ya untuk KNN
1	2,23606798	2	Ya (K<3)	Beasiswa
2	3,16227766	5	Tidak (K>3)	-
3	2,44948974	3	Ya (K=3)	Beasiswa
4	2,64575131	4	Tidak (K>3)	-
5	2	1	Ya (K<3)	Pendidikan

Selanjutnya yaitu melakukan klasifikasi. Klasifikasi menggunakan kategori mayoritas dari tetangga terdekat sebagai nilai prediksi data yang terbaru atau data *testing*. Data yang masuk dalam 3 tetangga terdekat yaitu dokumen 1 adalah kategori *scholar*, dokumen 3 adalah kategori *scholar*, dan dokumen 5 adalah kategori *education*. Maka, dari jumlah mayoritas (beasiswa > pendidikan) tersebut, disimpulkan bahwa data TES1 tersebut termasuk dalam kategori *scholar*.

Proses uji coba dengan *software Orange* menghasilkan klasifikasi seperti terlihat pada Gambar 5 dan Gambar 6.

Gambar 5. Tampilan Hasil Klasifikasi *Naive Bayes* dengan *Software Orange*

Gambar 5 merupakan tampilan hasil pengujian perhitungan metode *Naive Bayes* menggunakan *software orange*. Terlihat hasil perhitungan klasifikasi metode *Naive Bayes* terhadap 100 data uji diklasifikasikan sebagai kategori *biografi*.

Gambar 6. Tampilan Hasil Klasifikasi KNN dengan *Software Orange*

Hasil dari prediksi kategori menggunakan *software Orange* dari 100 data *testing* yang diujikan dengan metode *Naive Bayes* terdapat hanya 1 data yang diprediksi sesuai dengan data aktual. Maka nilai akurasi dari metode *Naive Bayes* adalah sebagai berikut:

$$Akurasi = \frac{TP}{Jumlah\ Dataset}$$

$$= \frac{1}{100} \times 100\%$$

$$= 1\%$$

Hasil dari prediksi kategori menggunakan *software Orange*. Dari 100 data *testing* yang diujikan dengan metode KNN terdapat 63 data yang diprediksi sesuai dengan data aktual. Maka nilai akurasi dari metode KNN adalah sebagai berikut:

$$Akurasi = \frac{TP}{Jumlah\ Dataset}$$

$$= \frac{63}{100} \times 100\%$$

$$= 63\%$$

Prediksi untuk klasifikasi metode *naive bayes* dan KNN menghasilkan masing-masing nilai akurasi yang berbeda. Adapun perbandingan kinerja dari kedua metode tersebut dapat dilihat pada Tabel 8.

Tabel 8. Perbandingan Kinerja Metode menggunakan *Software Orange*

Metode	Akurasi
<i>Naive Bayes</i>	51%
KNN	86%

Berdasarkan Tabel 8 dapat diketahui bahwa kinerja dari metode *K-Nearest Neighbour* (KNN) dalam kasus ini lebih baik dari metode *Naive Bayes*. Akurasi klasifikasi tidak bisa mencapai hasil yang sempurna karena pasti ada nilai *error*. Hal tersebut dipengaruhi oleh banyaknya data uji dan data latih yang digunakan dalam proses simulasi yang dilakukan.

4. KESIMPULAN

Berdasarkan hasil penelitian perbandingan metode *Naive Bayes* dan KNN dalam mengklasifikasi konten berita, maka dapat disimpulkan bahwa perhitungan metode *Naive Bayes* menghasilkan nilai *error* yang lebih besar dibandingkan metode KNN dalam melakukan klasifikasi konten berita yaitu menghasilkan nilai akurasi sebesar 51%.

Berdasarkan pengujian akurasi klasifikasi konten berita dalam penelitian ini, metode KNN dapat dipilih sebagai metode yang terbaik dalam melakukan klasifikasi konten berita. Hasil pengujian *confusion matrix* metode KNN diperoleh nilai akurasi sebesar 86%.

DAFTAR PUSTAKA

- [1] A. Hadid, I. A. Alam, and M. Noor, "Peran Reporter dalam Meningkatkan Mutu Siaran pada Program Siaran Lampung Hari Ini di TVRI Lampung," *J. Visionist*, vol. 9, no. 1, pp. 1–5, 2020.
- [2] C. Juditha, "Interaksi Komunikasi Hoax di Media Sosial serta Antisipasinya," *J. Pekommas*, vol. 3, no. 1, pp. 31–44, 2018.
- [3] M. R. Kertanegara, "Penggunaan Clickbait Headline pada Situs Berita dan Gaya Hidup Muslim dream.co.id," *Mediat. J. Komun.*, vol. 11, no. 1, pp. 31–43, 2018, doi: 10.29313/mediator.v11i1.2751.
- [4] N. D. Sukmono, "Clickbait Judul Berita Online dalam Pemberitaan Covid-19," *Transform. J. Bahasa, Sastra, dan Pengajarannya*, vol. 5, no. 1, p. 1, 2021, doi: 10.31002/transformatika.v5i1.3643.
- [5] Nielsen, "Media Cetak Mampu Mempertahankan Posisinya," *Nielsen.com*, 2020. <https://www.nielsen.com/id/en/press-releases/2017/media-cetak-mampu-mempertahankan-posisinya/> (accessed Sep. 19, 2022).
- [6] A. N. Roifa, "Text Mining Dengan Metode Naive Bayes Classifier Untuk Mengklasifikasikan Berita Berdasarkan Konten." Institut Teknologi Sepuluh Nopember, 2018.
- [7] M. I. T. Rizal, "Klasifikasi Berita Olahraga pada Portal Berita Online dengan Metode K-Nearest Neighbour (KNN) Dan Levenshtein Distance," *J. Teknol. Terap. Sains 4.0*, vol. 2, no. 1, pp. 31–49, 2021, [Online]. Available: <https://ojs.unimal.ac.id/tts/article/view/3760> %0Ahttps://ojs.unimal.ac.id/tts/article/viewFile/3760/2161.
- [8] H. Tantyoko, Adiwijaya, and U. N. Wisesty, "Perbandingan Pembobot untuk klasifikasi Topik Berita Menggunakan Decision Tree," *J. Teknol.*, vol. 2, no. 1, pp. 97–113, 2019.
- [9] D. N. Chandra, G. Indrawan, and I. N. Sukajaya, "Klasifikasi Berita Lokal Radar Malang Menggunakan Metode Naive Bayes dengan Fitur N-Gram," *J. Ilmu Komput. Indones.*, vol. 4, no. 2, pp. 11–19, 2019.
- [10] C. Fadlan, S. Ningsih, and A. P. Windarto, "Penerapan Metode Naive Bayes dalam Klasifikasi Kelayakan Keluarga Penerima Beras Rastra," *J. Tek. Inform. Musirawas*, vol. 3, no. 1, p. 1, 2018, doi: 10.32767/jutim.v3i1.286.
- [11] M. Ardiansyah, A. Sunyoto, and E. T. Luthfi, "Analisis Perbandingan Akurasi Algoritma Naive Bayes an C4.5 untuk Klasifikasi Diabetes," *Edumatic J. Pendidik. Inform.*, vol. 5, no. 2, pp. 147–156, 2021, doi: 10.29408/edumatic.v5i2.3424.
- [12] E. Widodo and P. Sa'adah, "Klasifikasi Data Penjualan Alat Tulis Kantor (ATK) Terlaris untuk Optimasi Strategi Pemasaran di Toko Citramedia Menggunakan Metode Naive Bayes," *SIGMA J. Teknol. Pelita Bangsa*, vol. 10, no. 2, pp. 138–152, 2019.
- [13] R. Bahtiar, M. D. S. T, A. Setiawan, and P. Rosyani, "Analisis Perbandingan Detection Traffic Anomaly dengan Metode Naive Bayes dan DBSCAN," *J. Kreat. Mhs. Inform.*, vol. 1, no. 2, pp. 99–103, 2020.
- [14] A. C. Khotimah and E. Utami, "Comparison Naive Bayes Classifier, K-Nearest Neighbor and Support Vector Machine in the Classification of Individual on Twitter Account," *J. Tek. Inform.*, vol. 3, no. 3, pp. 673–680, 2022, [Online]. Available: <http://jutif.if.unsoed.ac.id/index.php/jurnal/article/view/254>.
- [15] N. N. Dzikrulloh, Indriati, and B. D. Setiawan, "Penerapan Metode K – Nearest Neighbor (KNN) dan Metode Weighted Product (WP) dalam Penerimaan Calon Guru dan Karyawan Tata Usaha Baru Berwawasan Teknologi (Studi Kasus : Sekolah Menengah Kejuruan Muhammadiyah 2 Kediri)," *Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 1, no. 5, pp. 378–385, 2017.

- [16] E. Yulianti and Y. A. Nurdin, "Sistem Pendukung Keputusan Penerimaan Bantuan Siswa Miskin (BSM) Berbasis Online dengan Metode Knn (K-Nearest Neighbor) (Studi kasus: SMPN 1 Koto XI Tarusan)," *J. Teknoif*, vol. 6, no. 1, pp. 12–17, 2018, doi: 10.21063/jtif.2018.v6.1.12-17.
- [17] A. N. Yuliarina and Hendry, "Comparison Of Prediction Analysis Of Gofood Service Users Using The Knn & Naive Bayes Algorithm With Rapidminer Software," *J. Teknik Informatika (JUTIF)*, vol. 3, no. 4, pp. 847–856, 2022.
- [18] P. R. Sihombing and A. M. Arsani, "Perbandingan Metode Machine Learning Dalam Klasifikasi Kemiskinan Di Indonesia Tahun 2018," *J. Teknik Informatika (JUTIF)*, vol. 2, no. 1, pp. 51–56, 2021.

