

KLASIFIKASI BERITA ONLINE PADA LIPUTAN 6 MENGUNAKAN METODE KNN

**Oleh
ANDIKA TRI SAPUTRA
T3117118**

SKRIPSI

**Untuk memenuhi salah satu syarat ujian
Guna memperoleh gelar Sarjana**



**PROGRAM SARJANA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS ICHSAN GORONTALO
GORONTALO
2023**

PERSETUJUAN PENELITIAN

KLASIFIKASI BERITA ONLINE PADA LIPUTAN 6 MENGUNAKAN METODE KNN

Oleh
ANDIKA TRI SAPUTRA
T3117118

Untuk memenuhi salah satu syarat ujian
Guna memperoleh gelar Sarjana
Program Studi Teknik Informatika,
Ini telah disetujui dan siap untuk diseminarkan

Gorontalo, Desember 2023

Pembimbing I

Pembimbing II


Yasin Aril Mustafa, M.Kom
NIDN: 0926088503


Yulianty Lasena, M.Kom
NIDN: 0907078603

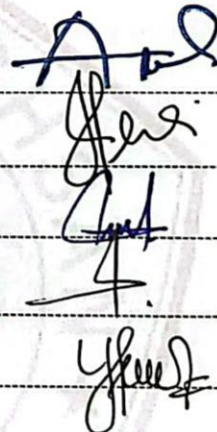
HALAMAN PENGESAHAN

KLASIFIKASI BERITA ONLINE PADA LIPUTAN 6 MENGUNAKAN METODE KNN

Oleh
ANDIKA TRI SAPUTRA
T3117118

Diperiksa Oleh Panitia Ujian Strata Satu (S1)
Universitas Ichsan Gorontalo

1. Ketua Penguji
Amiruddin, M.Kom
2. Anggota
Hastuti Dalai, M.Kom
3. Anggota
Maryam Hasan, M.Kom
4. Anggota
Yasin Aril Mustofa, M.kom
5. Anggota
Yulianti Lasena, M.Kom



Mengetahui :

Dekan Fakultas Ilmu Komputer

Irvan A. Salihi, M.Kom
NIND. 0928028101

Ketua Program Studi

Sudirman S. Panna, M.Kom
NIDN. 0924038205

HALAMAN PERNYATAAN

Dengan ini saya menyatakan bahwa:

1. Karya tulis saya (Skripsi) ini adalah asli dan belum pernah di ajukan untuk mendapatkan gelar akademik (Sarjana) baik di Universitas Ichsan Gorontalo maupun di perguruan Tinggi lainnya.
2. Karya tulis ini adalah murni gagasan, rumusan dan penelitian saya sendiri, tanpa bantuan pihak lain, kecuali arahan dari Tim Pembimbing.
3. Dalam karya tulis ini tidak dapat karya atau pendapat yang telah di publikasikan orang lain, kecuali secara tertulis dicantumkan sebagai acuan dalam naskah dengan disebutkan nama pengarang dan dicantumkan dalam daftar pustaka.
4. Pernyataan ini saya buat dengan sesungguhnya dan apabila dikemudian hari terdapat penyimpangan dan tidak kebenaran dalam pernyataan ini, maka saya tidak bersedia menerima sanksi akademik berupa pencabutan gelar yang telah di peroleh karena karya tulis ini, serta sanksi lainnya sesuai dengan norma yang berlaku di Perguruan Tinggi ini

Gorontalo, Desember 2023 Yang
Membuat Pernyataan,

Ar. 

Andilca Tri Saputra

KATA PENGANTAR

Alhamdulillah penulis dapat menyelesaikan penelitian dengan judul **“Klasifikasi Berita Online Pada Liputan 6 Menggunakan Metode KNN ”**, untuk memenuhi salah satu syarat penyusunan Skripsi Program Studi Teknik Informatika Fakultas Ilmu Komputer Universitas Ichsan Gorontalo.

Penulis menyadari bahwa penyusunan ini tidak dapat terwujud tanpa adanya bantuan dan dorongan dari ibu/bapak dosen. Untuk itu penulis mengucapkan banyak terima kasih dan penghargaan yang setinggi-tingginya kepada :

1. Ibu Dr. Dra. Juriko Abdussamad, M.Si, selaku Ketua Yayasan Pengembangan Ilmu Pengetahuan dan Teknologi (YPIPT) Ichsan Gorontalo;
2. Bapak Dr. Abdul Gaffar La Tjokke M.Si, selaku Rektor Universitas Ichsan Gorontalo;
3. Pak Irvan Abraham Salihi, S.Kom, M.Kom, selaku Dekan Fakultas Ilmu Komputer Universitas Ichsan Gorontalo;
4. Bapak Sudirman Melangi, M.Kom, selaku Wakil Dekan I Bidang Akademik Fakultas Ilmu Komputer Universitas Ichsan Gorontalo;
5. Ibu Irma Surya Kumala Idris, M.Kom, selaku Wakil Dekan II Bidang Administrasi Umum dan Keuangan Fakultas Ilmu Komputer Universitas Ichsan Gorontalo;
6. Bapak Sudirman S. Panna, M.Kom, selaku Ketua Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Ichsan Gorontalo;
7. Bapak Yasin Aril Mustafa, M.Kom, selaku Pembimbing I yang telah banyak membimbing penulis selama ini;
8. Yulianty Lasena, M.Kom, selaku Pembimbing II yang telah banyak membimbing penulis selama ini;
9. Bapak dan Ibu Dosen Universitas Ichsan Gorontalo yang sudah mendidik dan mengajarkan berbagai disiplin ilmu kepada penulis;

10. Kedua Orang Tua saya yang tercinta, atas segala kasih sayang, jerih payah dan doa restunya dalam membesarkan dan mendidik penulis;
11. Rekan-rekan seperjuangan yang telah banyak memberikan bantuan dan moril yang sangat besar kepada penulis;
12. Kepada semua pihak yang ikut membantu dalam penyelesaian skripsi ini yang tak sempat penulis sebutkan satu-persatu.

Semoga Allah SWT, membalas semua jasa-jasa mereka terhadap kami. Penulis sepenuhnya menyadari bahwa dalam penyusunan penelitian ini, terdapat banyak banyak kekurangan atau masih jauh dari. Untuk itu kritik yang membangun penulis sangat harapkan. Akhirnya penulis berharap agar hasil yang telah diperoleh ini membawa manfaat bagi kita semua. Aamiinn.

Gorontalo, November 2023

Penulis

ABSTRACT

ANDIKA TRI SAPUTRA. T3117118. CLASSIFICATION OF ONLINE NEWS ON LIPUTAN6 USING KNN METHOD

Online news is one of the vital sources of information for the community. However, the amount of online news available can make it difficult for users or news providers to classify online news, especially news on the liputan6.com page. The K-Nearest Neighbor (KNN) classification method is one of the classification methods used to classify online news. This method works by comparing news data with the other one as a known class. Previously data preprocessing and data transformation must be carried out using the Term Frequency-Inverse Document Frequency (TF-IDF) technique. This research aims to test the effectiveness of the KNN classification method in classifying online news on Liputan6. The data employed are online news data from Liputan6 published from November 19, 2023, to November 25, 2023, namely 788 news titles. The data includes news titles, dates, links, and news classes. The results indicate that the KNN classification method can classify online news on Liputan6 with an accuracy of 91% for six distinguished news classes, namely Business, Election, Islamic, Showbiz, Ball, and Techno classes. The research results indicate that the KNN classification method can be used to classify online news properly and correctly.

Keywords: data mining, classification, KNN, TF-IDF, online news, Liputan6



ABSTRAK

ANDIKA TRI SAPUTRA. T3117118. KLASIFIKASI BERITA ONLINE PADA LIPUTAN 6 MENGGUNAKAN METODE KNN

Berita online merupakan salah satu sumber informasi yang penting bagi masyarakat. Namun, banyaknya berita online yang tersedia dapat menyulitkan pengguna atau penyedia berita dalam melakukan klasifikasi berita online khususnya berita pada laman liputan6.com. Metode klasifikasi K-Nearest Neighbor (KNN) merupakan salah satu metode klasifikasi yang dapat digunakan untuk mengklasifikasi berita online. Metode ini bekerja dengan cara membandingkan data berita dengan data berita lain yang sudah diketahui kelasnya, dan sebelumnya harus dilakukan preprocessing data dan transformasi data dengan menggunakan teknik *Term Frequency-Inverse Document Frequency* (TF-IDF). Penelitian ini bertujuan untuk menguji efektivitas metode klasifikasi KNN dalam mengklasifikasi berita online pada Liputan6. Data yang digunakan dalam penelitian ini adalah databerita online dari Liputan6 yang diterbitkan pada tanggal 19 November 2023 s/d tanggal 25 November 2023 sebanyak 788 judul berita. Data tersebut meliputi judulberita, tanggal berita, link berita dan kelas berita. Hasil penelitian menunjukkan bahwa metode klasifikasi KNN dapat mengklasifikasi berita online pada Liputan6 dengan akurasi sebesar 91% untuk 6 kelas berita yang berbeda, yaitu kelas Bisnis, Pemilu, Islami, Showbiz, Bola dan Tekno Hasil penelitian ini menunjukkan bahwametode klasifikasi KNN dapat digunakan untuk melakukan klasifikasi berita onlinedengan baik dan benar.

Kata kunci: data mining, klasifikasi, KNN, TF-IDF, berita online, Liputan6



DAFTAR ISI

PERSETUJUAN PENELITIAN	ii
HALAMAN PENGESAHAN	iii
HALAMAN PERNYATAAN	iv
KATA PENGANTAR	v
<i>ABSTRACT</i>	vii
ABSTRAK	viii
DAFTAR ISI	ix
DAFTAR GAMBAR	xi
DAFTAR TABEL	xii
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Identifikasi Masalah	3
1.3 Rumusan Masalah	3
1.4 Tujuan Penelitian	3
1.5 Manfaat Penelitian	3
1.5.1 Manfaat Teoritis	4
1.5.2 Manfaat Praktis	4
BAB II LANDASAN TEORI	5
2.1 Tinjauan Studi	5
2.2 Tinjauan Pustaka	5
2.3 <i>Text Mining</i>	7
2.3.1 <i>Algoritma KNN</i>	7
2.3.2 Penerapan algoritma KNN	8
2.3.3 <i>Confusion Matrix</i>	9
2.4 Perangkat Lunak	12
2.5 Kerangka Pikir	13
BAB III METODE PENELITIAN	14
3.1 Jenis, Metode, Subjek, Objek, Waktu dan Lokasi Penelitian	14
3.1.1 Model Yang Diusulkan	14

BAB IV HASIL PEMBAHASAN	19
4.1 Hasil Pengumpulan Data.....	19
4.2 Pemodelan.....	33
BAB V PEMBAHASAN	48
5.1 Pembahasan Kinerja Model	48
5.2 Pembahasan Hasil Pemodelan.....	51
Bab VI Kesimpulan dan Saran.....	53
DAFTAR PUSTAKA	
LAMPIRAN	

DAFTAR GAMBAR

Gambar 2. 1 Proses <i>Knowledge Discovery</i> in Database (KDD) [6].....	6
Gambar 2. 3 Kerangka Pikir.....	13
Gambar 2. 2 Model yang Diusulkan	14
Gambar 5. 1 Confusion Matriks Data Testing	48
Gambar 5. 2 Nilai k Terbaik Perbandingan Data 80:20	52
Gambar 5. 3 Nilai k Terbaik Perbandingan Data 70:30	53

DAFTAR TABEL

Tabel 3.1 Data Berita Online	15
Tabel 4.1 Jumlah Dataset Berdasarkan Jenis Berita.....	19
Tabel 4. 2 Sampel Dataset Berita Online Kelas Pemilu.....	20
Tabel 4. 3 Sampel Dataset Berita Online Kelas Bisnis	22
Tabel 4. 4 Sampel Dataset Berita Online Kelas Bola.....	24
Tabel 4. 5 Sampel Dataset Berita Online Kelas Showbiz	26
Tabel 4. 6 Sampel Dataset Berita Online Kelas Tekno.....	28
Tabel 4. 7 Sampel Dataset Berita Online Kelas Islami	30
Tabel 4. 8 Dataset.....	34
Tabel 4. 9 <i>Case Folding Data Training</i>	36
Tabel 4. 10 <i>Stopword Removal</i>	36
Tabel 4. 11 <i>Steaming</i>	37
Tabel 4. 12 Hasil Ekstraksi Fitur TF-IDF Berita online	39
Tabel 4. 13 <i>Normalisasi TF-IDF</i>	41
Tabel 4. 14 Data Training Sebelum Preprosesing dan Pembobotan TF- IDF.....	42
Tabel 4. 15 Data Training Setelah melalui Preprosesing dan Pembobotan TF- I.	42
Tabel 4. 16 Data Training Sebelum Preprosesing data dan Pembobotan TF- IDF43	
Tabel 4. 17 Data Training Setelah melalui Preprosesing data dan Pembobotan TF- IDF	43
Tabel 4. 18 Rekap Confusion Matriks Hasil Pemodelan	47

BAB I

PENDAHULUAN

1.1 Latar Belakang

Berita merupakan sebuah informasi *public* yang turut ambil bagian bagi perkembangan peradaban manusia. Dikehidupan sehari-hari, manusia tidak akan pernah bisa terhindar dari yang namanya berita. Berbagai macam sumber berita yang bisa diperoleh oleh seseorang, seperti berita dari media cetak, media elektronik, sampai pada media online. Banyak orang yang setiap harinya mencari berita-berita terbaru baik sebelum maupun setelah beraktifitas bahkan dalam sementara ada kegiatan. Berita yang paling banyak dan mudah diakses merupakan berita melalui internet atau media online (Berita Online). Begitu banyaknya artikel berita yang dimuat di media internet dalam yang tempo singkat memberikan permasalahan tersendiri bagi seorang editor untuk mengelompokkan jenis berita secara manual. Begitu banyaknya berita yang akan diterbitkan setiap harinya membuat tim editor kewalahan dalam menentukan kategori berita yang akan diterbitkan, tentu saja kondisi tersebut membutuhkan waktu yang cukup lama untuk menentukan kategori tiap berita yang masuk.[1]

Masalah Klasifikasi Berita online menjelaskan tentang permasalahan yang terjadi dalam klasifikasi berita online. Masalah tersebut bisa berupa kekurangan sistem klasifikasi berita yang ada, keterbatasan dalam mengkategorikan berita ke dalam kategori yang sesuai, atau bahkan masalah-masalah lain yang berkaitan dengan klasifikasi berita. Penjelasan mengenai masalah tersebut akan membantu dalam menyusun tujuan dan manfaat dari penelitian ini. Misalnya, masalah klasifikasi berita yang sering terjadi adalah kesulitan dalam mengkategorikan berita ke dalam kategori yang tepat karena banyaknya topik yang muncul dalam sebuah berita. Hal ini dapat menyebabkan salah klasifikasi dan mengurangi keakuratan dari sistem klasifikasi berita. Oleh karena itu, diperlukan metode yang lebih efektif dalam mengelompokkan berita ke dalam kategori yang sesuai.

Tujuan dari penelitian ini adalah untuk mengklasifikasikan berita di online berdasarkan kategori menggunakan metode KNN (*K-Nearest Neighbors*). Metode

ini dipilih karena mampu mengelompokkan berita ke dalam kategori yang sesuai dengan akurasi yang tinggi. Dengan menggunakan metode KNN, diharapkan dapat meningkatkan keakuratan dalam klasifikasi berita di Indonesia. Selain itu, tujuan dari penelitian ini juga adalah untuk mengidentifikasi kategori-kategori yang paling sering muncul pada berita di Indonesia. Hasil dari penelitian ini diharapkan dapat memberikan informasi yang bermanfaat bagi para pengguna berita di Indonesia dalam mencari berita yang sesuai dengan kategori yang diinginkan. Keunggulan dari metode KNN dibuktikan dari peneliti sebelumnya yaitu penelitian yang dilakukan oleh Andi Ahmad Irfa 2018, dalam penelitiannya, peneliti menggunakan metode KNN untuk melakukan klasifikasi Topik Berita yang Berbahasa Indonesia hasil penelitian yang diperoleh yaitu mendapatkan *performa micro average f1-measure* sebesar 69,9%.[2], penelitian yang dilakukan oleh Nengah Widya Utami, 2022 menganalisis mengenai sentiment pembelajaran daring pada masa covid 19 dengan algoritma KNN, hasil akurasi yang diperoleh sebesar 82.48%.[3] dan penelitian yang dilakukan oleh Retno Sari, 2020 membahas tentang penerapan metode KNN dalam menganalisis sentimen pada Review Objek Wisata Dunia Fantasi, nilai akurasi yang diperoleh dengan nilai $k=7$ yaitu akurasi 77.01, *precision* 92.38, *recall* 61.56 dan nilai AUC 0.894.[4]

Variabel yang dibutuhkan dalam penelitian ini meliputi Berita (*Dependent Variable*), yaitu data berita yang akan diklasifikasikan ke dalam kategori yang sesuai dan Kategori (*Independent Variable*), yaitu kategori yang akan diterapkan pada setiap berita untuk mengelompokkannya.

Jenis data yang dibutuhkan dalam penelitian ini adalah data berita yang akan diambil dari berita liputan6. Data tersebut akan dibersihkan dan dipreprocessing sebelum diolah menggunakan metode KNN untuk proses klasifikasi berita. Selain itu, juga dibutuhkan data kategori asli dari setiap berita yang digunakan sebagai acuan dalam mengevaluasi akurasi dari hasil klasifikasi berita menggunakan metode KNN. Jenis data yang digunakan adalah data kuantitatif yang akan diolah secara numerik dan dianalisis menggunakan teknik statistik.

Berdasarkan penjelasan latar belakang maka peneliti akan melakukan penelitian dengan judul **“Klasifikasi Berita Online Pada Liputan6 Menggunakan Metode Knn”**.

1.2 Identifikasi Masalah

Pihak editor kesulitan dalam pengelompokan jenis berita dikarenakan banyaknya berita yang masuk dalam tempo yang singkat. Masalah tersebut bisa berupa kekurangan sistem klasifikasi berita yang ada, keterbatasan dalam mengkategorikan berita ke dalam kategori yang sesuai.

1.3 Rumusan Masalah

Berdasarkan identifikasi masalah diatas, maka permasalahannya dirumuskan yaitu:

- 1 Bagaimana hasil analisa pengelompokan berita dengan menggunakan metode KNN?
- 2 Seberapa besar tingkat akurasi dalam klasifikasi berita online dengan algoritma KNN

1.4 Tujuan Penelitian

Berdasarkan Rumusan permasalahannya diatas, maka tujuan dari penelitian ini yaitu:

1. Untuk mengetahui hasil analisa pengelompokan berita menggunakan metode KNN.
2. Untuk mengetahui tingkat akurasi dari penerapan algoritma KNN dalam melakukan klasifikasi berita online.

1.5 Manfaat Penelitian

Penelitian ini diharapkan mempunyai manfaat yaitu:

1.5.1 Manfaat Teoritis

Penelitian ini diharapkan dapat memberikan sumbangan pengembangan ilmu bidang kajian data mining tentang kemampuan metode KNN dalam penerapan analisteks mining.

1.5.2 Manfaat Praktis

Hasil dari penelitian ini diharapkan dapat digunakan sebagai *alternative* atau masukan kepada semua pihak yang berkepentingan khususnya dalam menganalisa kelompok berita.

BAB II

LANDASAN TEORI

2.1 Tinjauan Studi

Adapun penelitian terdahulu yang berhubungan dengan tema penelitian yang dijadikan sebagai bahan referensi dalam menentukan metode yang akan digunakan.

Tabel 2. 1 Referensi Penelitian

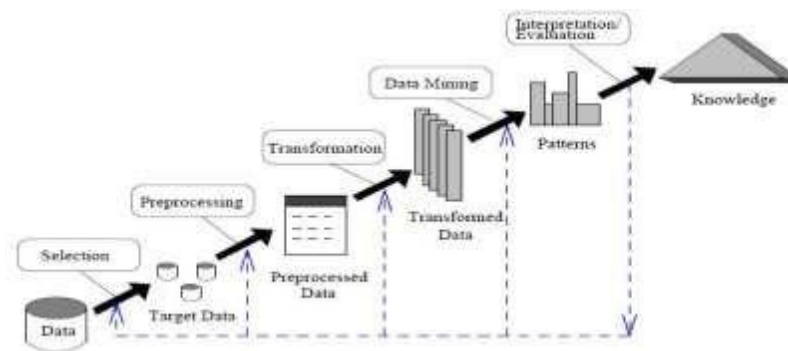
No	Nama Peneliti	Judul Penelitian	Deskripsi Singkat
1.	Andi Ahmad Irfa 2018[2]	Klasifikasi Topik Berita Berbahasa Indonesia Menggunakan <i>k-Nearest Neighbor</i>	menghasilkan performa <i>micro average f1-measure</i> sebesar 69,9%
2.	Nengah Widya Utami,2022[3]	<i>Text Mining</i> Dalam Analisis Sentimen Pembelajaran Daring Di Masa Pandemi Covid 19 Menggunakan Algoritma <i>K-Nearest Neighbor</i>	Hasil penelitiannya sangat bagus diaman nilai akurasi yang diperoleh dari metode KNN yaitu sebesar 82.48%.
3.	Retno Sari, 2020[4]	Analisis Sentimen Pada <i>Review</i> Objek Wisata Dunia Fantasi menggunakan Algoritma <i>K-Nearest Neighbor</i> (K-NN)	Hasil akurasi yang diperoleh dari Eksperimen adalah 77.01, <i>precision</i> 92.38, <i>recall</i> 61.56 dan nilai AUC 0.894.

2.2 Tinjauan Pustaka

2.2.1 Data Mining

Istilah data mining biasa juga di artikan sebagai *Knowledge Discovery* (Tan dalam Eko Prasetyo). Proses untuk mendapatkan informasi yang berguna dari gudang basis data yang besar di artikan sebagai menambang data (Data mining), selain itu diartikan sebagai pengekstrakan informasi baru yang diambil dari tumpukan data besar yang membantu dalam pengambilan keputusan.

Menurut *Daryl Pregibon* menyatakan bahwa “Data mining adalah campuran statistik, kecerdasan buatan, dan riset basis data” [5].



Gambar 2. 1 Proses *Knowledge Discovery in Database* (KDD) [6]

Proses *Knowledge Discovery in Database* (KDD) merupakan bagian dari data mining. KDD adalah proses terorganisir untuk mengidentifikasi pola yang valid, baru, berguna, dan dapat dimengerti dari sebuah dataset yang besar dan kompleks [5].

2.2.2 Fungsi Data Mining

Penambangan data sebenarnya mempunyai banyak sekali manfaat atau fungsi. Namun, secara umum, terdapat 2 fungsi data *mining* yaitu: [7]

1. Deskriptif: Kapasitas yang berbeda mengacu pada kapasitas dalam memahami informasi yang terkonsentrasi lebih lanjut. Melalui siklus ini, adalah normal bahwa perilaku informasi dapat diketahui. Informasi ini nantinya dapat dimanfaatkan untuk mengungkap kualitas informasi yang ideal. Dengan menggunakan kapasitas ilustratif dari penggalian informasi, contoh tertentu yang awalnya tersembunyi dalam suatu informasi dapat ditemukan. Dengan demikian, dengan asumsi disadari bahwa ada contoh yang suram dan memiliki harga diri, itu menyiratkan bahwa atribut informasi dapat diketahui.
2. Prediktif: Fungsi Prediktif dari data mining adalah elemen interaksi yang pada akhirnya akan mengungkap contoh informasi tertentu. Contoh ini harus terlihat dari beberapa faktor yang terkandung dalam informasi. Setiap kali sebuah contoh ditemukan, contoh tersebut dapat digunakan untuk mengukur berbagai faktor yang nilai atau jenisnya pada saat ini belum jelas. Oleh karena itu, fungsi prediktif dipandang sebagai fungsi prediktif yang sebanding dengan analisis prediktif. Kemampuan ini juga dapat digunakan

untuk menilai variabel luar biasa yang tidak ada dalam informasi. Selanjutnya, kapasitas ini dapat bekerja dengan dan menguntungkan setiap individu yang membutuhkan perkiraan yang tepat untuk meningkatkan hal-hal yang signifikan.

2.3 Text Mining

Text mining ialah cara yang dapat diterapkan untuk mengekstraksi sebuah data untuk diteliti yang mana data tersebut bersumber dari data dalam bentuk teks. salah satu disiplin ilmu yang diterapkan pada informasi *retrival*, *linguistic* komputasi, *machine learning*, *statistic* dan data mining adalah *Text Mining*. [8] untuk melakukan information extraction, klasifikasi, clustering dan information retrieval [9]. *Text mining* adalah upaya untuk mendapatkan informasi supaya pengguna mampu berinteraksi terhadap sekumpulan dokumen dari masa ke masa dengan menerapkan perangkat analisis [10]. Tujuan utama dari *text mining* adalah untuk memperoleh sebuah kata yang dapat menggantikan konten dari dokumen yang selanjutnya dilakukan proses pengkajian untuk mengetahui keterkaitan tiap dokumen.

2.3.1 Algoritma KNN

Algoritma KNN adalah sebuah metode klasifikasi terhadap sekumpulan data berdasarkan pembelajaran data yang sudah terklasifikasi sebelumnya. Termasuk dalam *supervised learning*, dimana hasil *query instance* yang baru diklasifikasi berdasarkan mayoritas kedekatan jarak dari kategori yang ada dalam KNN. [11]

$$d_i = \sqrt{\sum_{i=1}^p (x_{2i} - x_{1i})^2} \dots \dots \dots 2.1$$

Keterrangan

x_1 = Sampel Data
 x_2 = Data Uji/Testing
 i = Variabel Daya
 d = Jarak
 p = Dimensi Data

2.3.2 Penerapan algoritma KNN

Contoh penerapan algoritma KNN (*K-Nearest Neighbors*) dalam *text mining*:

1. Menentukan dataset yang akan digunakan. Misalnya, kita memiliki dataset berikut ini:

Tabel 2. 2 Contoh Data *Training*

No	Text
1	Aku suka makan nasi goreng
2	Anjing tersebut sangat lucu
3	Dia adalah teman terbaikku

2. Mengubah teks menjadi vektor *numerik*. Misalnya, menggunakan metode TF-IDF (*Term Frequency-Inverse Document Frequency*) untuk menghasilkan vektor *numerik* dari setiap teks dalam dataset.

Tabel 2. 3 Hasil Pengolahan TF-IDF Data *Training*

No	Text
1	0.7, 0.3, 0, 0, 0
2	0, 0, 0.5, 0.5, 0
3	0, 0, 0, 0, 0.7

Pada contoh ini, kita menggunakan 5 fitur *numerik*, dengan masing-masing fitur merepresentasikan kata-kata yang muncul dalam dataset.

3. Menentukan dokumen baru yang akan diklasifikasikan. Misalnya, kita memiliki dokumen baru dengan teks berikut:

Tabel 2. 4 Data *Testing*

No	Text
4	Aku suka anjing

4. Mengubah dokumen baru menjadi vektor *numerik* menggunakan metode yang sama seperti langkah sebelumnya.

Tabel 2. 5 Hasil TF-IDF Data *Testing*

No	Text
4	0.5, 0.2, 0, 0.3, 0

5. Menentukan parameter K , yaitu jumlah tetangga terdekat yang akan digunakan. Misalnya, kita menggunakan $K = 1$.
 6. Menghitung jarak antara dokumen baru dan setiap dokumen dalam dataset menggunakan metode yang sesuai, misalnya jarak *Euclidean*.
 - Jarak *NewText* dengan *Text1*: $\sqrt{((0.7-0.5)^2 + (0.3-0.2)^2 + (0-0)^2 + (0-0.3)^2 + (0-0)^2)} = 0.2$
 - Jarak *NewText* dengan *Text2*: $\sqrt{((0-0.5)^2 + (0-0.2)^2 + (0.5-0)^2 + (0.5-0.3)^2 + (0-0)^2)} = 0.854$
 - Jarak *NewText* dengan *Text3*: $\sqrt{((0-0)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2 + (0.7-0)^2)} = 0.7$
 7. Mengurutkan jarak secara menaik dan memilih K dokumen terdekat. Dalam contoh ini, kita memilih 1 dokumen terdekat.
 - Dokumen terdekat: *Text1*
 8. Melabeli dokumen terdekat berdasarkan mayoritas kelas tetangga terdekat. Misalnya, jika *Text1* memiliki label kelas "positif", maka dokumen baru juga akan diberi label kelas "positif".
- Dalam contoh ini, kita dapat mengklasifikasikan dokumen baru "Aku suka anjing" menjadi kelas "positif" berdasarkan KNN dengan $K = 1$.

2.3.3 Confusion Matrix

Confusion matrix adalah sebuah tabel yang digunakan untuk mengevaluasi performa dari suatu sistem klasifikasi. Tabel ini berisi empat jenis hasil klasifikasi yang mungkin terjadi dalam sebuah percobaan klasifikasi, yaitu *true positive (TP)*, *false positive (FP)*, *true negative (TN)*, dan *false negative (FN)*.

True positive (TP) adalah jumlah data positif yang benar diprediksi sebagai positif oleh sistem klasifikasi. *False positive (FP)* adalah jumlah data negatif yang salah diprediksi sebagai positif. *True negative (TN)* adalah jumlah data negatif yang benar diprediksi sebagai negatif. *False negative (FN)* adalah jumlah data positif yang salah diprediksi sebagai negatif.

Tabel 2. 6 Confusion Matrix

Predicted Class	Actual Class		
		Kelas=Positif	Kelas=Negatif
	Kelas = Positif	TP(True Positive)	FN(False Negatif)
	Kelas = Negatif	FP(False Positive)	TN(True Negatif)

$$akurasi = \frac{(TP+TN)}{TP+TN+FP+FN} \times 100 \dots\dots\dots 2.2$$

$$precision = \frac{TP}{TP+FP} \times 100 \dots\dots\dots 2.3$$

$$recall = \frac{TP}{TP+FN} \times 100 \dots\dots\dots 2.4$$

$$f-22222222 = \frac{(2 \times recall \times presisi)}{(recall + presisi)} \dots\dots\dots 2.5$$

Keterangan :

TP(True Positive) : jumlah data yang memiliki kategori positif diklasifikasikan positif.

TN(True Negatif) : jumlah data yang memiliki kategori negative diklasifikasikan negatif.

FN(False Negatif) : jumlah data yang memiliki kategori positif diklasifikasikan negatif.

TN(True Negatif) : jumlah data yang memiliki kategori negative diklasifikasikan positif.

Contoh perhitungan *confusion matrix* dapat diberikan sebagai berikut:

Misalkan kita memiliki dataset dengan 100 data, dengan 2 kelas yaitu "positif" dan "negatif".

Setelah melakukan prediksi, kita mendapatkan hasil sebagai berikut:

- 60 data diklasifikasikan dengan benar ke kelas positif (*true positive*)
- 5 data diklasifikasikan dengan salah ke kelas negatif (*false negative*)
- 25 data diklasifikasikan dengan benar ke kelas negatif (*true negative*)
- 10 data diklasifikasikan dengan salah ke kelas positif (*false positive*)

Dengan menggunakan informasi di atas, kita dapat membuat *confusion matrix* sebagai berikut:

Tabel 2. 7 *Confusion Matrix*

Aktual		
Prediksi	<i>Positive</i>	<i>Negative</i>
<i>Positive</i>	60	10
<i>Negative</i>	5	25

Dalam *Confusion matrix* tersebut, terdapat 4 nilai yang perlu diperhatikan:

True positive (TP) : Jumlah data yang diklasifikasikan dengan benar ke kelas positif (60). *False negative (FN)* : Jumlah data yang seharusnya diklasifikasikan ke kelas positif tapi salah diklasifikasikan ke kelas negatif (5). *True negative (TN)*: Jumlah data yang diklasifikasikan dengan benar ke kelas negatif (25). *False positive (FP)* : Jumlah data yang seharusnya diklasifikasikan ke kelas negatif tapi salah diklasifikasikan ke kelas positif (10)

$$akurasi = \frac{(60+25)}{60+25+10+5} \times 100 = 85$$

$$precision = \frac{60}{60+10} \times 100 = 85,71$$

$$recall = \frac{60}{60+5} \times 100 = 92,3$$

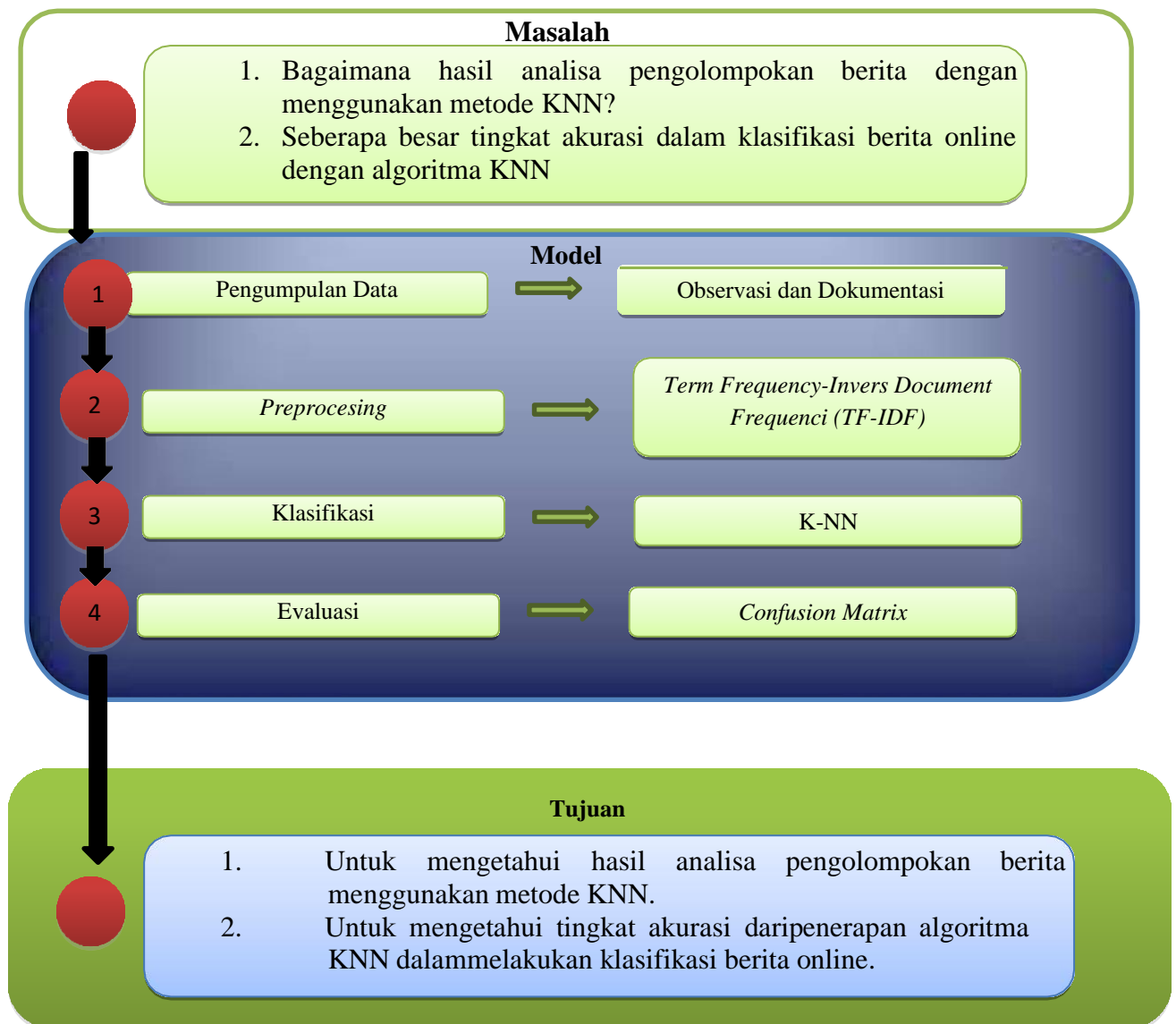
$$f - measure = \frac{(2 \times 92,3 \times 85,71)}{(92,3 + 85,71)} = 88,88$$

2.4 Perangkat Lunak

Tabel 2. 8 Perangkat Lunak Pendukung

No	Perangkat Lunak Pendukung	Berfungsi
1	<i>Phyton</i>	Suatu Bahasa Pemograman yang digunakan untuk melakukan olah dan analisa data
2	<i>Google Colab</i>	<i>Tools</i> yang digunakan untuk pengkodean
3	<i>MS.Excel</i>	Aplikasi <i>office</i> yang digunakan untuk menginput data

2.5 Kerangka Pikir



Gambar 2. 2 Kerangka Pikir

BAB III

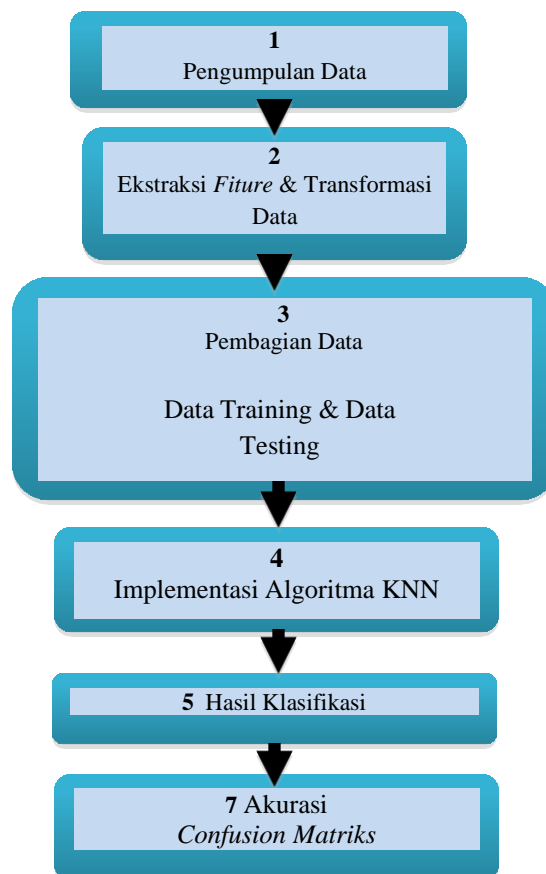
METODE PENELITIAN

3.1 Jenis, Metode, Subjek, Objek, Waktu dan Lokasi Penelitian

Penelitian ini menggunakan metode penelitian studi kasus media berita. Dengan demikian jenis penelitian ini adalah kuantitatif. Subjek penelitian ini adalah Analisa pengelompokan berita di Indonesia Menggunakan Metode KNN. Penelitian ini dimulai dari Agustus 2023 – November 2023 mencakup pengumpulan dan pengolahan data beserta bimbingan.

3.1.1 Model Yang Diusulkan

Berikut adalah tahapan penelitian yang akan dilakukan peneliti.



Gambar 2. 3 Model yang Diusulkan

1. Pengumpulan Data

Pengumpulan data merupakan tahapan awal yang dilakukan peneliti dalam memulai suatu penelitian baik data sekunder maupun data primer. Pengumpulan data ini dilakukan dengan beberapa cara yaitu baik dengan cara observasi maupun dengan cara dokumentasi.

Tabel 3. 1 Data Berita Online

NO	BERITA	KELAS
1	Ketum PAN Zulhas Rinci Strategi Pemenangan TKN Prabowo-Gibran Usai Rapat Dewan Pengarah	PEMILU
2	Survei Populi Center: Dukungan Pemilih di Bawah 35 Tahun kepada Prabowo-Gibran Meningkatkan Jadi 54,7 Persen	PEMILU
3	Ratusan Tokoh Masyarakat Jawa Tengah Deklarasi Dukung Prabowo-Gibran di Pilpres 2024	PEMILU
4	Alasan Forum Ulama Santri Indonesia Dukung Prabowo-Gibran di Pilpres 2024	PEMILU
5	Rapat TKN Prabowo-Gibran, Airlangga: Bahas Hal Strategis	PEMILU
6	Jelang Pemilu 2024, Sahroni NasDem Minta PPATK LebihAwat	PEMILU
7	Sebut Semua Butuh Proses, Jubir TPN Novi Basuki: Pasangan Ganjar-Mahfud Md Pernah Duduki BerbagaiJabatan	PEMILU
8	KPU Gelar Debat Capres Cawapres 5 Kali, Ini Temanya	PEMILU
9	Prabowo-Gibran Diprediksi Bisa Menang 1 Putaran di Pilpres2024	PEMILU
10	Survei Populi Center: Prabowo-Gibran 43,1%, Ganjar-Mahfud 23% dan Anies-Muhaimin 22,3%	PEMILU
...
51	Benahi Pinjol Dkk, OJK Luncurkan Peta Jalan Fintech Lending 2023-2028	BISNIS
52	Bank BTN Bidik Potensi Bisnis 5.000 Anggota HIPMI	BISNIS
53	Persiapkan Masa Pensiun, BRI Tumbuhkan Jiwa WirausahaAnggota dan ASN Polri	BISNIS
54	Rupiah Dibuka Melemah Tipis ke 15.695 per Dolar AS,Potensi Penurunan Lebih Dalam Masih Terbuka	BISNIS
55	Karya Unik “Cutemonster” Diminati Penikmat Fashion, BRILIANPRENEUR Jalan Perlebar Usaha	BISNIS
56	Gen Z Lebih Sulit Sukses Dibanding Orang Tuanya, Benarkah?	BISNIS

NO	BERITA	KELAS
57	Bentuk Ekosistem Kesehatan yang Setara, Eddy Sariaatmadja Sabet Penghargaan Filantropi	BISNIS
58	Raih Penghargaan Bidang Filantropi, Ini Pesan Eddy Sariaatmadja	BISNIS
...
101	Peristiwa dan Kisah Manusia di Bola Ganjil	OLAHRAGA
102	Jadwal, Hasil, dan Klasemen Lengkap Piala Dunia U-17 2023: Siapa Lolos ke 16 Besar?	OLAHRAGA
103	Piala Dunia U-17 2023 di Jakarta: Brasil Sebut Iran Lawan Paling Berat	OLAHRAGA
104	MotoGP Malaysia: Kalahkan Jorge Martin, Francesco Bagnaia Rebut Pole	OLAHRAGA
105	Prediksi Liga Italia Lecce vs AC Milan: Benahi Rapor Domestik!	OLAHRAGA
106	Top 3 Berita Bola: Ester Nurumi Tri Wardoyo ke Semifinal Korea Masters 2023	OLAHRAGA
107	Jadwal Korea Masters 2023, Sabtu 11 November: Siapa Lolos ke Final?	OLAHRAGA
108	Piala Dunia U-17 2023: Bima Sakti Bersyukur Timnas Indonesia U-17 Dapat Poin Lawan Ekuador	OLAHRAGA
...
151	iQOO 12 dengan Snapdragon 8 Gen 3 Rilis di China, Siap Masuk Juga ke Indonesia	TEKNOLOGI
152	Epic Games, Microsoft, EA, hingga Ubisoft Digugat karena Dianggap Bikin Anak Kecanduan Game	TEKNOLOGI
153	Top 3 Tekno: XL Axiata Ingin Setop 2G hingga Cuitan Selamat Hari Pahlawan Ramai di X Twitter	TEKNOLOGI
154	Konser BMTH di Jakarta Ricuh, Fans Khawatir Tidak Ada Kesempatan Lagi Menonton Oliver Sykes!	TEKNOLOGI
155	Apple Tunda Pengembangan iOS 18 hingga macOS 15 demi Perbaiki Bug	TEKNOLOGI
...
247	Obat dan Kunci Agar Tak Patah Hati ala Gus Iqdam, Murah dan Mujarab	ISLAM
248	5 Faktor Utama Dakwah Gus Iqdam Cepat Diterima Semua Kalangan	ISLAM
249	Lika-Liku Sejarah Kemerdekaan Palestina dan Perjuangan yang Belum Usai hingga Kini	ISLAM
250	4 Malaikat yang Datangi Orang Sakit dan Tugasnya, Simak Hikmah Tersembunyi di Balikinya	ISLAM

Sumber: www.liputan6.com, 202

2. Ekstraksi *Fiture* dan Transformasi Data

Ekstraksi *Fiture* dan Transformasi Data merupakan tahapan *pre_prosesing* data, dimana tahapan ini meliputi:

1. *Cleaning* (Pembersihan): Tahap ini mencakup menghilangkan spasi yang tidak perlu, tanda baca, karakter khusus, dan simbol yang tidak relevan. Selain itu, juga dapat melibatkan penghapusan kata-kata yang umum atau *stop words*, seperti "adalah", "dan", atau "dari".
2. Tokenisasi: Tahap ini melibatkan pemisahan teks menjadi unit-unit yang lebih kecil, seperti kata-kata atau *frase*. Misalnya, kalimat "Saya sedang belajar *text mining*" akan dipecah menjadi token-token seperti "Saya", "sedang", "belajar", dan "*text mining*".
3. Normalisasi: Tahap ini adalah mengubah semua token ke bentuk yang lebih standar atau normal. Contoh normalisasi termasuk mengubah semua huruf menjadi huruf kecil (*lowercase*), menggantikan sinonim dengan kata yang sama, atau menghapus sufiks atau prefiks yang tidak penting.
4. *Stop Word Removal*: Tahap ini melibatkan penghapusan kata-kata yang umum atau *stop words* yang tidak memberikan banyak informasi penting dalam analisis. Misalnya, kata-kata seperti "adalah", "dan", atau "dari" sering kali dihapus karena kemunculannya yang tinggi namun relevansinya yang rendah.
5. *Stemming* atau *Lemmatization*: Tahap ini adalah mengubah kata-kata ke bentuk kata dasarnya. *Stemming* melibatkan menghapus *sufiks* atau *prefiks* dari kata, sementara *lemmatization* melibatkan mengubah kata ke bentuk kata dasarnya menggunakan kamus atau aturan gramatikal.
6. *Data Transformation*: Tahap ini melibatkan mengubah data teks yang telah dipreproses menjadi representasi angka yang dapat dipahami oleh algoritma pemodelan. Salah satu teknik yang umum digunakan adalah *bag-of-words*, yang mengubah setiap dokumen menjadi vektor dengan jumlah kata-kata yang terjadi dalam dokumen tersebut.
7. *Feature Selection*: Tahap ini melibatkan memilih fitur-fitur atau kata-kata yang paling relevan untuk analisis. Misalnya, hanya memilih kata-kata dengan frekuensi tinggi atau kata-kata yang memiliki hubungan erat dengan tujuan analisis.

3. Pembagian data

Pembagian data *testing* dan *training* dalam *text mining* merupakan proses yang penting dalam mengembangkan model atau sistem *text mining* dengan kombinasi 80% dengan 20% atau 70% dengan 30% atau kombinasi *training* dan *testing* lainnya. Tujuan dari pembagian data ini adalah untuk menguji dan melatih model dengan menggunakan data yang berbeda.

2. Implementasi Algoritma KNN

Implementasi algoritma KNN merupakan tahapan menganalisa data dengan menggunakan algoritma KNN dengan mencari nilai K atau jarak terdekat untuk mendapatkan suatu kesimpulan dari proses klasifikasi.

3. Hasil Klasifikasi

Hasil klasifikasi merupakan hasil yang diperoleh dari hasil pengolahan data dari algoritma KNN dengan melihat nilai ketetanggaan yang diperoleh yang menunjukkan suatu hasil klasifikasi.

4. Akurasi

Akurasi merupakan nilai yang digunakan untuk mengukur kualitas dari suatu metode dalam melakukan pengolahan data. Ada beberapa cara yang dapat digunakan untuk mengukur akurasi antara lain, *confusion matrix*, MSE dan RMSE.

BAB IV

HASIL PEMBAHASAN

4.1 Hasil Pengumpulan Data

Dataset berita online yang digunakan pada penelitian ini adalah dataset yang dikumpulkan dari laman <https://www.liputan6.com/> dengan menggunakan teknik *Web scraping*. Teknik ini digunakan untuk mengambil data dari halaman web secara otomatis. Teknik *Web scraping* dilakukan sesuai jumlah jenis berita pada laman tersebut yang banyak diakses oleh pengunjung atau masyarakat. Berita online yang diunduh adalah berita yang diterbitkan dari tanggal 19 November 2023 s/d 25 November 2023. Adapun jumlah dataset berita online yang sudah dikumpulkan berdasarkan jenis berita dapat dilihat pada Tabel 4.1 berikut ini :

Tabel 4. 1 Jumlah Dataset Berdasarkan Jenis Berita

No	Jenis Berita	Jumlah Dataset
1	Pemilu	140
2	Bisnis	140
3	Bola	140
4	Showbiz	140
5	Tekno	121
6	Islami	106
Total Dataset		787

Berikut ini adalah sample dataset berita online yang sudah diunduh untuk masing-masing jenis berita :

Tabel 4. 2 Sampel Dataset Berita Online Kelas Pemilu

No	Judul Berita	Kelas	Tanggal Berita	Link Berita
1	Siap Menangkan Prabowo-Gibran, Ridwan Kamil: Harus Kerja Keras Semua	Pemilu	25 Nov 2023 23:00	https://www.liputan6.com/pemilu/read/5464661/siap-menangkan-prabowo-gibran-ridwan-kamil-harus-kerja-keras-semua
2	Siap Menangkan Prabowo-Gibran, Ridwan Kamil: Harus Kerja Keras Semua	Pemilu	25 Nov 2023 22:10	https://www.liputan6.com/pemilu/read/5464647/ridwan-kamil-yakin-pilpres-satu-putaran-jabar-tetap-prabowo
3	Bertemu Kaesang, Nelayan Mengadu Soal Amdal Rumah Mereka	Pemilu	25 Nov 2023 21:46	https://www.liputan6.com/pemilu/read/5464640/bertemu-kaesang-nelayan-mengadu-soal-amdal-rumah-mereka
4	Jalankan Konstitusi, Ganjar Siap Jamin Kebebasan Beragama dan Beribadah	Pemilu	25 Nov 2023 20:38	https://www.liputan6.com/pemilu/read/5464773/jalankan-konstitusi-ganjar-siap-jamin-kebebasan-beragama-dan-beribadah
5	Cak Imin Umumkan Timnas AMIN Jatim, Mendes Abdul Halim Didapuk Jadi Dewan Pengarah	Pemilu	25 Nov 2023 19:00	https://www.liputan6.com/pemilu/read/5464261/cak-imin-umumkan-timnas-amin-jatim-mendes-abdul-halim-didapuk-jadi-dewan-pengarah
6	KPU DKI Jakarta Pastikan Warga Disabilitas Nyaman saat Memilih di TPS	Pemilu	25 Nov 2023 18:01	https://www.liputan6.com/pemilu/read/5464253/kpu-dki-jakarta-pastikan-warga-disabilitas-nyaman-saat-memilih-di-tps
7	Mensos Risma Minta KPU Akomodasi Fasilitas bagi Difabel untuk Memilih pada Pemilu 2024	Pemilu	25 Nov 2023 17:30	https://www.liputan6.com/pemilu/read/5464246/mensos-risma-minta-kpu-akomodasi-fasilitas-bagi-difabel-untuk-memilih-pada-pemilu-2024

No	Judul Berita	Kelas	Tanggal Berita	Link Berita
8	Mardiono Resmikan Kantor DPW PPP Sulsel, Beri Semangat Kader Menangkan Pileg-Pilpres 2024	Pemilu	25 Nov 2023 17:02	https://www.liputan6.com/pemilu/read/5464537/mardiono-resmikan-kantor-dpw-ppp-sulsel-beri-semangat-kader-menangkan-pileg-pilpres-2024
9	Gerindra Klaim Dukungan Masyarakat ke Prabowo-Gibran Semakin Masif	Pemilu	25 Nov 2023 17:01	https://www.liputan6.com/pemilu/read/5464145/gerindra-klaim-dukungan-masyarakat-ke-prabowo-gibran-semakin-masif
10	Blusukan ke Pasar, Relawan Kawan Juang GP Gelar Baksos hingga Bagikan Sembako	Pemilu	25 Nov 2023 17:00	https://www.liputan6.com/pemilu/read/5464779/blusukan-ke-pasar-relawan-kawan-juang-gp-gelar-baksos-hingga-bagikan-sembaka
140	Anies Baswedan Bertemu Kiai Kampung, Bicara soal Keadilan	Pemilu	19 Nov 2023 12:00	https://www.liputan6.com/pemilu/read/5457216/anies-baswedan-bertemu-kiai-kampung-bicara-soal-keadilan

Tabel 4. 3 Sampel Dataset Berita Online Kelas Bisnis

No	Judul Berita	Kelas	Tanggal Berita	Link Berita
1	Sri Mulyani: Kesenjangan Gender di Indonesia Masih Jadi PR	Bisnis	25 Nov 2023 21:00	https://www.liputan6.com/bisnis/read/5464510/sri-mulyani-kesenjangan-gender-di-indonesia-masih-jadi-pr
2	Punya Banyak Talenta Unggul, Kinerja Unilever Masih Kinclong	Bisnis	25 Nov 2023 20:30	https://www.liputan6.com/bisnis/read/5464592/punya-banyak-talenta-unggul-kinerja-unilever-masih-kinclong
3	Berperan Penting di Sektor Industri, Pembangunan Pusat Data Harus Gencar	Bisnis	25 Nov 2023 20:15	https://www.liputan6.com/bisnis/read/5464584/berperan-penting-di-sektor-industri-pembangunan-pusat-data-harus-gencar
4	Kementan Dorong Program Upsus untuk Tingkatkan Produksi Jagung di Gorontalo	Bisnis	Informasi Tanggal Tidak Tersedia	https://www.liputan6.com/bisnis/read/5464591/kementan-dorong-program-upsus-untuk-tingkatkan-produksi-jagung-di-gorontalo
5	Kapan BJB Syariah Melantai di Bursa? Ini Bocoran Terbaru	Bisnis	25 Nov 2023 20:00	https://www.liputan6.com/bisnis/read/5464178/kapan-bjb-syariah-melantai-di-bursa-ini-bocoran-terbaru
6	Jumlah Penumpang KRL Diramal Turun 50% Saat Nataru 2023-2024	Bisnis	25 Nov 2023 19:00	https://www.liputan6.com/bisnis/read/5464185/jumlah-penumpang-krl-diramal-turun-50-saat-nataru-2023-2024
7	Pengamat: Idealnya UMP 2024 Naik 7%-10%	Bisnis	25 Nov 2023 18:00	https://www.liputan6.com/bisnis/read/5464169/pengamat-idealnya-ump-2024-naik-7-10

No	Judul Berita	Kelas	Tanggal Berita	Link Berita
8	Jamin Stok Pangan Aman, Pemerintah Siapkan 19 Gudang Bulog di IKN Nusantara	Bisnis	25 Nov 2023 17:00	https://www.liputan6.com/bisnis/read/5464498/jamin-stok-pangan-aman-pemerintah-siapkan-19-gudang-bulog-di-ikn-nusantara
9	Kenaikan UMP 2024 Tak Cukup Jaga Daya Beli Tahun Depan	Bisnis	25 Nov 2023 16:31	https://www.liputan6.com/bisnis/read/5463994/kenaikan-ump-2024-tak-cukup-jaga-daya-beli-tahun-depan
10	Mau Buka Lagi, TikTok Shop Bisa Gabung E-Commerce yang Sudah Ada	Bisnis	25 Nov 2023 15:38	https://www.liputan6.com/bisnis/read/5464139/mau-buka-lagi-tiktok-shop-bisa-gabung-e-commerce-yang-sudah-ada
....
140	Pengusaha: UMP 2024 Naik 15 Persen, Harga Barang Bakal Melambung	Bisnis	19 Nov 2023 11:00	https://www.liputan6.com/bisnis/read/5457198/pengusaha-ump-2024-naik-15-persen-harga-barang-bakal-melambung

Tabel 4. 4 Sampel Dataset Berita Online Kelas Bola

No	Judul Berita	Kelas	Tanggal Berita	Link Berita
1	Hasil La Liga Spanyol: Gol Bunuh Diri Selamatkan Barcelona Kalah dari Rayo Vallecano	Bola	25 Nov 2023 22:19	https://www.liputan6.com/bola/read/5464651/hasil-la-liga-spanyol-gol-bunuh-diri-selamatkan-barcelona-kalah-dari-rayo-vallecano
2	Newcastle United vs Chelsea Sudah Mulai, Ini Link Live Streaming Liga Inggris di Vidio	Bola	25 Nov 2023 22:00	https://www.liputan6.com/bola/read/5464590/newcastle-united-vs-chelsea-sudah-mulai-ini-link-live-streaming-liga-inggris-di-vidio
3	Hasil Babak 8 Besar Piala Dunia U-17 2023: Siapa Lolos ke Semifinal?	Bola	25 Nov 2023 21:36	https://www.liputan6.com/bola/read/5463124/hasil-babak-8-besar-piala-dunia-u-17-2023-siapa-lolos-ke-semifinal
4	Hasil MotoGP Valencia 2023: Juara Sprint Race, Jorge Martin Tunda Pesta Francesco Bagnaia	Bola	25 Nov 2023 21:35	https://www.liputan6.com/bola/read/5464635/hasil-motogp-valencia-2023-juara-sprint-race-jorge-martin-tunda-pesta-francesco-bagnaia
5	Hasil Liga Inggris: Gol Dianulir VAR, Manchester City Berbagi Angka dengan Liverpool	Bola	25 Nov 2023 21:28	https://www.liputan6.com/bola/read/5464629/hasil-liga-inggris-gol-dianulir-var-manchester-city-berbagi-angka-dengan-liverpool
6	Hasil Piala Dunia U-17 2023: Ibrahim Diarra Bawa Mali Bungkam Maroko dan Lolos ke Semifina...	Bola	25 Nov 2023 21:17	https://www.liputan6.com/bola/read/5464624/hasil-piala-dunia-u-17-2023-ibrahim-diarra-bawa-mali-bungkam-maroko-dan-lolos-ke-semifinal
7	Link Live Streaming Liga Inggris Newcastle United vs Chelsea di Vidio, Sabtu 25 November 2...	Bola	25 Nov 2023 20:30	https://www.liputan6.com/bola/read/5464586/link-live-streaming-liga-inggris-newcastle-united-vs-chelsea-di-vidio-sabtu-25-november-2023
8	Piala Dunia U-17 2023: Berbagai Cara Pelatih Uzbekistan Instruksikan Pemain dari Tribune S...	Bola	25 Nov 2023 20:26	https://www.liputan6.com/bola/read/5464601/piala-dunia-u-17-2023-berbagai-cara-pelatih-uzbekistan-instruksikan-pemain-dari-tribune-stadion-manahan

No	Judul Berita	Kelas	Tanggal Berita	Link Berita
9	LaLiga Spanyol Rayo Vallecano vs Barcelona Sedang Tayang di Vidio, Dapatkan Link Live Stre...	Bola	25 Nov 2023 20:00	https://www.liputan6.com/bola/read/5464321/laliga-spanyol-rayo-vallecano-vs-barcelona-sedang-tayang-di-vidio-dapatkan-link-live-streaming-nya-di-sini
10	Polisi Tahan Tersangka Pemalsuan Tiket Piala Dunia U-17 2023, Pelaku Terinspirasi Penipuan...	Bola	25 Nov 2023 19:53	https://www.liputan6.com/bola/read/5464583/polisi-tahan-tersangka-pemalsuan-tiket-piala-dunia-u-17-2023-pelaku-terinspirasi-penipuan-konser-coldplay
....
140	Hasil Japan Masters 2023: Gregoria Mariska Tunjung Raih Gelar Juara usai Hajar Jagoan Chin...	Bola	19 Nov 2023 11:22	https://www.liputan6.com/bola/read/5457229/hasil-japan-masters-2023-gregoria-mariska-tunjung-raih-gelar-juara-usai-hajar-jagoan-china-di-final

Tabel 4. 5 Sampel Dataset Berita Online Kelas Showbiz

No	Judul Berita	Kelas	Tanggal Berita	Link Berita
1	Heboh BCL dan Tiko Aryawardhana Akan Nikah Di Bali, Siap Lepas Status Janda Ashraf Sinclair...	Showbiz	25 Nov 2023 22:19	https://www.liputan6.com/showbiz/read/5464646/heboh-bcl-dan-tiko-aryawardhana-akan-nikah-di-bali-siap-lepas-status-janda-ashraf-sinclair
2	6 Reaksi Umi Pipik Soal Ammar Zoni Digugat Cerai: Yakin Irish Bella Istikharah dan Tak Bon...	Showbiz	25 Nov 2023 21:00	https://www.liputan6.com/showbiz/read/5464529/6-reaksi-umi-pipik-soal-ammar-zoni-digugat-cerai-yakin-irish-bella-istikharah-dan-tak-bongkar-aib-suami
3	Oklin Fia Tak Terima Dituding Pelakor, Yang Merusak Rumah Tangga Ammar Zoni dan Irish Bell...	Showbiz	25 Nov 2023 20:43	https://www.liputan6.com/showbiz/read/5464607/oklin-fia-tak-terima-dituding-pelakor-yang-merusak-rumah-tangga-ammar-zoni-dan-irish-bella
4	Saksikan Live Streaming SCTV Sinetron Di Antara Dua Cinta, Episode Sabtu 25 November 2023 ...	Showbiz	25 Nov 2023 20:40	https://www.liputan6.com/showbiz/read/5464244/saksikan-live-streaming-sctv-sinetron-di-antara-dua-cinta-episode-sabtu-25-november-2023-pukul-2110-wib
5	Saksikan Live Streaming Indosiar Mega Series Magic 5, Tayang Sabtu 25 November 2023 Pukul ...	Showbiz	25 Nov 2023 20:30	https://www.liputan6.com/showbiz/read/5464248/saksikan-live-streaming-indosiar-mega-series-magic-5-tayang-sabtu-25-november-2023-pukul-2100-wib
6	Jatuh Bangun Jefri Nichol Jadi Tulang Punggung Keluarga agar Adik Bisa Kuliah, Akui Pernah...	Showbiz	25 Nov 2023 20:00	https://www.liputan6.com/showbiz/read/5464474/jatuh-bangun-jefri-nichol-jadi-tulang-punggung-keluarga-agar-adik-bisa-kuliah-akui-pernah-kelaparan

No	Judul Berita	Kelas	Tanggal Berita	Link Berita
7	Saksikan Live Streaming SCTV Sinetron Takdir Cinta Yang Kupilih, Episode Sabtu 25 November...	Showbiz	25 Nov 2023 19:40	https://www.liputan6.com/showbiz/read/5464243/saksikan-live-streaming-sctv-sinetron-takdir-cinta-yang-kupilih-episode-sabtu-25-november-2023-pukul-1940-wib
8	Rio Dewanto Memimpin Aksi Spektakuler dalam Film 13 Bom di Jakarta, Sajikan Ketegangan Sej...	Showbiz	25 Nov 2023 19:30	https://www.liputan6.com/showbiz/read/5464008/rio-dewanto-memimpin-aksi-spektakuler-dalam-film-13-bom-di-jakarta-sajikan-ketegangan-sejak-dari-trailer-perdana
9	Keluarga Siap Mengawal Proses Hukum Dugaan Malpraktik Operasi Sedot Lemak Nanie Darham	Showbiz	25 Nov 2023 19:00	https://www.liputan6.com/showbiz/read/5464501/keluarga-siap-mengawal-proses-hukum-dugaan-malpraktik-operasi-sedot-lemak-nanie-darham
10	Abelle Sebarkan Semangat Perjuangan Lewat Single Pemenang Hati, Minta Orang Lain Jangan Me...	Showbiz	25 Nov 2023 18:30	https://www.liputan6.com/showbiz/read/5464496/abelle-sebarkan-semangat-perjuangan-lewat-single-pemenang-hati-minta-orang-lain-jangan-menyerah
140	KLa Project Merayakan Pergantian Tahun dengan Konser Intimate dan 20 Lagu Spektakuler	Showbiz	19 Nov 2023 14:30	https://www.liputan6.com/showbiz/read/5457232/kla-project-merayakan-pergantian-tahun-dengan-konser-intimate-dan-20-lagu-spektakuler

Tabel 4. 6 Sampel Dataset Berita Online Kelas Tekno

No	Judul Berita	Kelas	Tanggal Berita	Link Berita
1	Cara Aktifkan Sinkronisasi Otomatis di Aplikasi Samsung Notes	Tekno	25 Nov 2023 19:00	https://www.liputan6.com/teknoread/5463629/cara-aktifkan-sinkronisasi-otomatis-di-aplikasi-samsung-notes
2	Bos Google Wanti-Wanti Pengguna Bahaya HP Android Lakukan Praktik Sideload	Tekno	25 Nov 2023 18:00	https://www.liputan6.com/teknoread/5457255/bos-google-wanti-wanti-pengguna-bahaya-hp-android-lakukan-praktik-sideload
3	Tokopedia Hadirkan 2 Fitur Baru, Kini Belanja Banyak Barang Lebih Mudah	Tekno	25 Nov 2023 17:00	https://www.liputan6.com/teknoread/5462882/tokopedia-hadirkan-2-fitur-baru-kini-belanja-banyak-barang-lebih-mudah
4	Rumah BUMN Telkom Bantu UMKM Kutoarjo Kembangkan Bisnis lewat Digitalisasi	Tekno	25 Nov 2023 15:00	https://www.liputan6.com/teknoread/5464193/rumah-bumn-telkom-bantu-umkm-kutoarjo-kembangkan-bisnis-lewat-digitalisasi
5	Analisis LinkedIn Ungkap Karyawan Apple yang Resign Banyak Pindah ke Google	Tekno	25 Nov 2023 14:00	https://www.liputan6.com/teknoread/5463555/analisis-linkedin-ungkap-karyawan-apple-yang-resign-banyak-pindah-ke-google
6	Final League of Legends Worlds 2023 Cetak Rekor Jumlah Penonton Terbanyak Sepanjang Sejarah...	Tekno	25 Nov 2023 13:00	https://www.liputan6.com/teknoread/5462255/final-league-of-legends-worlds-2023-cetak-rekor-jumlah-penonton-terbanyak-sepanjang-sejarah

No	Judul Berita	Kelas	Tanggal Berita	Link Berita
7	Oppo Reno 11 Resmi Meluncur, Varian Pro Hadir dengan Snapdragon 8+ Gen 1	Tekno	25 Nov 2023 12:00	https://www.liputan6.com/teknoread/5463437/oppo-reno-11-resmi-meluncur-varian-pro-hadir-dengan-snapdragon-8-gen-1
8	Top 3 Tekno: Korea Utara Klaim Luncurkan Satelit Mata-Mata Militer hingga YouTube Lelet Ga...	Tekno	25 Nov 2023 11:00	https://www.liputan6.com/teknoread/5464302/top-3-tekno-korea-utara-klaim-luncurkan-satelit-mata-mata-militer-hingga-youtube-lelet-gara-gara-ad-blocker
9	Toge Productions Akuisisi Mojiken Studio Pembuat Game A Space for the Unbound	Tekno	25 Nov 2023 10:00	https://www.liputan6.com/teknoread/5463391/toge-productions-akuisisi-mojiken-studio-pembuat-game-a-space-for-the-unbound
10	Asus Rilis Zenbook Pro 14 OLED, Laptop Tipis dengan Performa Buas	Tekno	25 Nov 2023 09:00	https://www.liputan6.com/teknoread/5463896/asus-rilis-zenbook-pro-14-oled-laptop-tipis-dengan-performa-buas
...
121	Game Tarisland Buka Uji Coba Closed Beta ke-2 di Android dan PC, Indonesia Kebagian	Tekno	19 Nov 2023 08:00	https://www.liputan6.com/teknoread/5454922/game-tarisland-buka-uji-coba-closed-beta-ke-2-di-android-dan-pc-indonesia-kebagian

Tabel 4. 7 Sampel Dataset Berita Online Kelas Islami

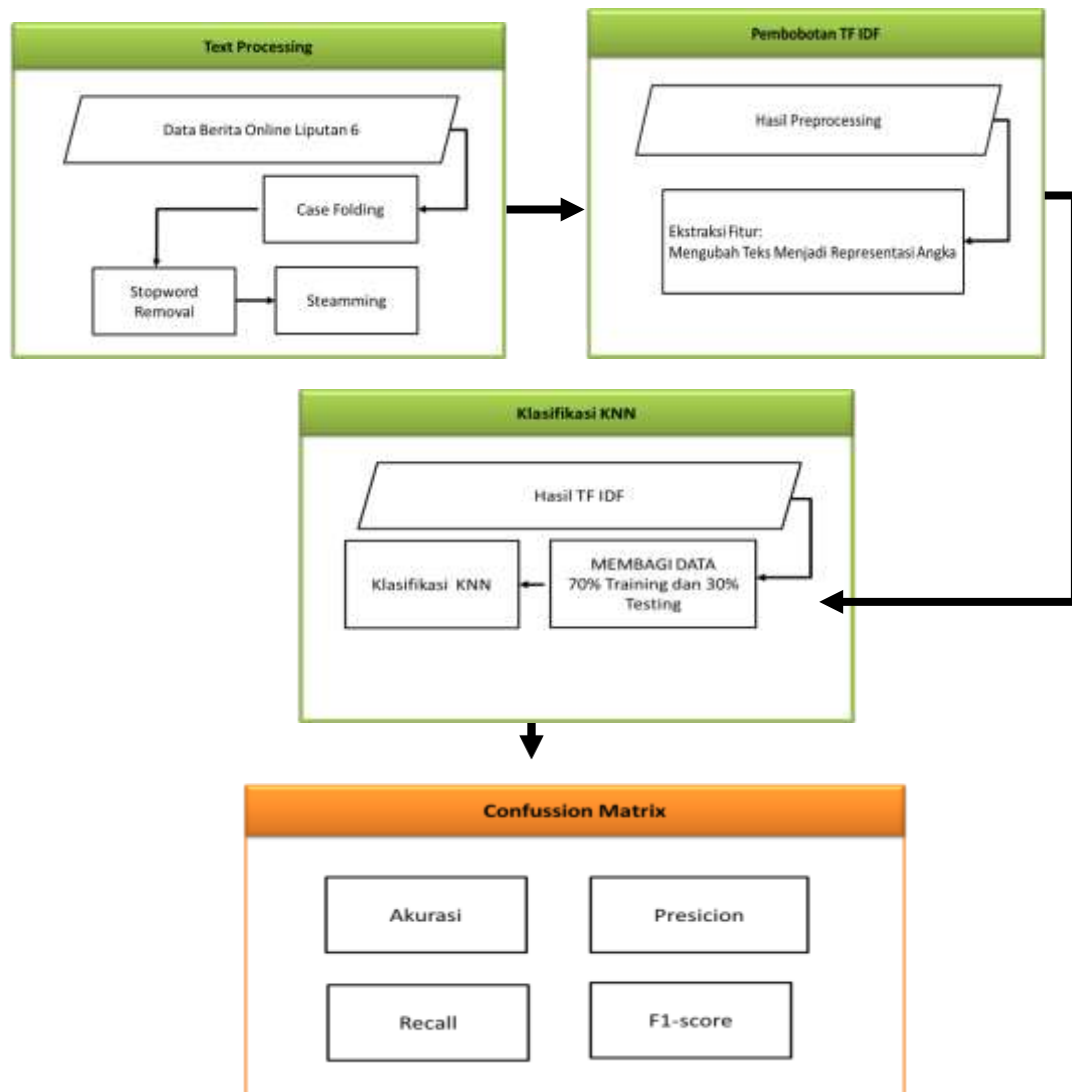
NO	Judul Berita	Kelas	Tanggal Berita	Link Berita
1	Biaya Haji 2024 Disepakati Rp93,4 Juta, Berapa yang Harus Dibayar Jemaah Haji?	Islami	25 Nov 2023 22:30	https://www.liputan6.com/islami/read/5464589/biaya-haji-2024-disepakati-rp934-juta-berapa-yang-harus-dibayar-jemaah-haji
2	Bacaan Niat, Tata Cara dan Doa Wudhu yang Benar: Agar Ibadah Sah	Islami	25 Nov 2023 20:30	https://www.liputan6.com/islami/read/5464580/bacaan-niat-tata-cara-dan-doa-wudhu-yang-benar-agar-ibadah-sah
3	Keutamaan Dahsyat Tidur Siang Sejenak dan Sholat Sunnah Sebelum Zuhur	Islami	25 Nov 2023 18:30	https://www.liputan6.com/islami/read/5464290/keutamaan-dahsyat-tidur-siang-sejenak-dan-sholat-sunnah-sebelum-zuhur
4	Tukang Ghibah Bakal Dijebloskan ke Neraka Hutamah, Begini Kengeriannya	Islami	25 Nov 2023 16:30	https://www.liputan6.com/islami/read/5463259/tukang-ghibah-bakal-dijebloskan-ke-neraka-hutamah-begini-kengeriannya
5	Kemuliaan Guru dalam Al-Qur'an dan Hadis, Salah Satunya Masuk Surga Tanpa Hisab	Islami	25 Nov 2023 14:30	https://www.liputan6.com/islami/read/5464264/kemuliaan-guru-dalam-al-quran-dan-hadis-salah-satunya-masuk-surga-tanpa-hisab
6	Termasuk Gus Iqdam, Siapapun Tak Berani Menolak Undangan Artis Ini	Islami	25 Nov 2023 12:30	https://www.liputan6.com/islami/read/5464236/termasuk-gus-iqdam-siapapun-tak-berani-menolak-undangan-artis-ini

No	Judul Berita	Kelas	Tanggal Berita	Link Berita
7	Hukum Mengganti Lafal Aamiin dengan Qobul dalam Berdoa, Bolehkah?	Islami	25 Nov 2023 10:30	https://www.liputan6.com/islami/read/5463262/hukum-mengganti-lafal-aamiin-dengan-qobul-dalam-berdoa-bolehkah
8	30 Ucapan Selamat Hari Guru Nasional, Berkesan dan Penuh Makna Mendalam	Islami	25 Nov 2023 09:30	https://www.liputan6.com/islami/read/5464285/30-ucapan-selamat-hari-guru-nasional-berkesan-dan-penuh-makna-mendalam
9	Tren Berdoa di Depan Ka'bah Lewat Video Call, Bagaimana Hukumnya?	Islami	25 Nov 2023 08:30	https://www.liputan6.com/islami/read/5463335/tren-berdoa-di-depan-kabah-lewat-video-call-bagaimana-hukumnya
10	Bacaan Doa untuk Guru Agar Diberi Ampunan, Disayangi dan Dimuliakan Allah SWT	Islami	25 Nov 2023 07:53	https://www.liputan6.com/islami/read/5464283/bacaan-doa-untuk-guru-agar-diberi-ampunan-disayangi-dan-dimuliakan-allah-swt

106	Habib Novel Ungkap Kekagumannya kepada Gus Iqdam: I Love You Full	Islami	19 Nov 2023 00:30	https://www.liputan6.com/islami/read/5456469/habib-novel-ungkap-kekagumannya-kepada-gus-iqdam-i-love-you-full

4.2 Pemodelan

Model yang dibangun pada penelitian ini akan dilakukan secara terstruktur mulai dari preprocessing, pembobotan TF.IDF, klasifikasi KNN, dan kemudian di valuasi menggunakan confusion matrix. Penulis Membuat flowchart untuk model yang dibangun:



Gambar 4.2.1 Pemodelan

Berdasarkan dataset yang sudah dikumpulkan pada Tabel 4.2 s/d Tabel 4.7 di atas dilakukan beberapa pengolahan data sebelum diterapkan pada model KNN untuk melakukan klasiifkasi jenis berita, diantaranya dilakukan tahapan pengolahan data yaitu dengan melakukan seleksi data. Pada tahapan seleksi atribut dari atribut Judul Berita,

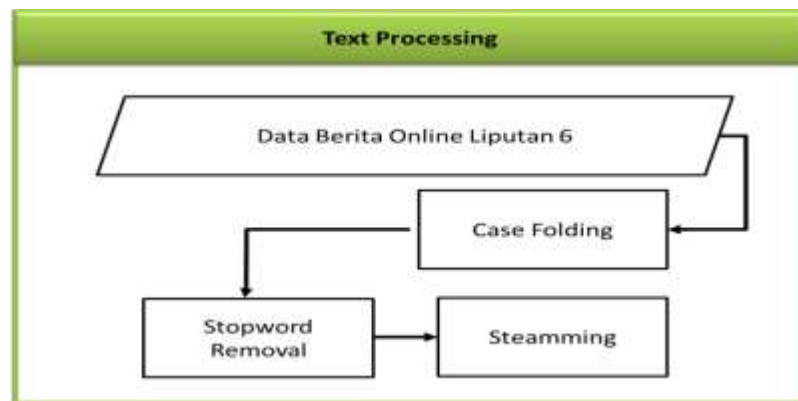
Kelas, Tanggal Berita dan Link Berita dipilih yaitu 2 (dua) atribut Judul Berita dan Kelas yang akan digunakan dalam pemodelan KNN selanjutnya. Sehingga diperoleh dataset seperti tabel 4.8

Tabel 4. 8 Dataset

No	Judul Berita	Kelas
1	Siap Menangkan Prabowo-Gibran, Ridwan Kamil: Harus Kerja Keras Semua	Pemilu
2	Ridwan Kamil Yakin Pilpres Satu Putaran: Jabar Tetap Prabowo	Pemilu
3	Bertemu Kaesang, Nelayan Mengadu Soal Amdal Rumah Mereka	Pemilu
4	Jalankan Konstitusi, Ganjar Siap Jamin Kebebasan Beragama dan Beribadah	Pemilu
5	Cak Imin Umumkan Timnas AMIN Jatim, Mendes Abdul Halim Didapuk Jadi Dewan Pengarah	Pemilu
6	KPU DKI Jakarta Pastikan Warga Disabilitas Nyaman saat Memilih di TPS	Pemilu
7	Mensos Risma Minta KPU Akomodasi Fasilitas bagi Difabel untuk Memilih pada Pemilu 2024	Pemilu
8	Mardiono Resmikan Kantor DPW PPP Sulsel, Beri Semangat Kader Menangkan Pileg-Pilpres 2024	Pemilu
9	Gerindra Klaim Dukungan Masyarakat ke Prabowo-Gibran Semakin Masif	Pemilu
10	Blusukan ke Pasar, Relawan Kawan Juang GP Gelar Baksos hingga Bagikan Sembako	Pemilu
788	Habib Novel Ungkap Kekagumannya kepada Gus Iqdam: I Love You Full	Islami

4.2.1 Preprocessing Data

Pada tahapan ini dilakukan sejumlah proses yaitu menghapus judul berita yang duplikat, mengubah huruf judul berita menjadi huruf kecil semuanya, dan menghapus kata-kata umum (di, ke, dan, atau, dari, dan lainnya) dalam judul berita. *Preprocessing* dilakukan setelah dataset siap untuk diklasifikasi. Pada tahapan *preprocessing* terdapat beberapa proses yang akan dilakukan sehingga data menjadi terstruktur dengan tahapan *Case folding*, *Stopword removal*, dan *Steaming*. Peneliti membuat *flowchart* mengenai *preprocessing*.



Gambar 4. 2 Preprocessing

Pada tahapan *preprocessing* peneliti menggunakan *framework google colab* berbasis *python*, untuk melakukan proses *Case folding*, *stopword removal* dan *steaming*.

1) *Case folding*

Proses Case folding merupakan tahapan proses yang dilakukan peneliti dalam mengolah data berita yaitu dengan mengubah semua huruf kapital menjadi huruf kecil pada semua data berita yang ada pada dokumen data *training* dan data *testing*. Berdasarkan dari hasil pengolahan case folding yang dilakukan peneliti dapat dilihat pada tabel 4.2 berikut

Tabel 4. 9 *Case Folding Data Training*

NO	BERITA	<i>Case Folding</i>
1	Ketum PAN Zulhas Rinci Strategi Pemenangan TKN Prabowo-Gibran Usai Rapat Dewan Pengarah	ketum pan zulhas rinci strategi pemenangan tkn prabowogibran rapat dewan pengarah
2	Survei Populi Center: Dukungan Pemilih di Bawah 35 Tahun kepada Prabowo-Gibran Meningkat Jadi 54,7 Persen	survei populi center dukungan pemilih 35 prabowogibran meningkat 547 persen
3	Ratusan Tokoh Masyarakat Jawa Tengah Deklarasi Dukung Prabowo-Gibran di Pilpres 2024	ratusan tokoh masyarakat jawa deklarasi dukung prabowogibran pilpres 2024
4	Alasan Forum Ulama Santri Indonesia Dukung Prabowo-Gibran di Pilpres 2024	alasan forum ulama santri indonesia dukung prabowogibran pilpres 2024
5	Rapat TKN Prabowo-Gibran, Airlangga: Bahas Hal Strategis	rapat tkn prabowogibran airlangga bahas strategis

2) *Stopword removal*

Setelah tahapan atau proses *case folding* berhasil dilakukan oleh peneliti, tahapan selanjutnya adalah mengolah data hasil dari pengolahan *case folding* ke dalam proses *Stopword removal*. Tahapan *Stopword removal* merupakan proses yang dilakukan peneliti untuk penghilang kata yang tidak berkaitan dengan topic dokumen. Dari hasilproses *Stopword removal* diperoleh hasil pengolahan data yang dapat dilihat pada tabel 4.3.

Tabel 4. 10 *Stopword Removal*

NO	<i>Case Folding</i>	<i>Stopword Removal</i>
1	ketum pan zulhas rinci strategi pemenangan tkn prabowogibran rapat dewan pengarah	ketum pan zulhas rinci strategi pemenangan tkn prabowogibran rapat dewan pengarah
2	survei populi center dukungan pemilih 35 prabowogibran meningkat 547 persen	survei populi center dukungan pemilih 35 prabowogibran meningkat 547 persen
3	ratusan tokoh masyarakat jawa deklarasi dukung prabowogibran pilpres 2024	ratusan tokoh masyarakat jawa eklarasi dukung prabowogibran pilpres 2024
4	alasan forum ulama santri indonesia dukung prabowogibran pilpres 2024	alasan forum ulama santri indonesia dukung prabowogibran pilpres 2024
5	rapat tkn prabowogibran airlangga bahas strategis	rapat tkn prabowogibran airlangga bahas strategis

3). ***Steaming***

Setelah proses *Stopword removal* berhasil dilakukan, dilanjutkan ke tahapan berikutnya yaitu proses steaming. Tahapan proses steaming iyalah proses penghapusan imbuhan dari setiap kata. Sehingga mendapatkan kata dasarnya. Hasil dari proses steaming dapat dilihat pada tabel 4.4 berikut:

Tabel 4. 11 *Steaming*

NO	<i>Stopword Removal</i>	<i>Steaming</i>
1	ketum pan zulhas rinci strategi pemenangan tkn prabowogibran rapat dewan pengarah	ketum pan zulhas rinci strategi pemenangan tkn prabowogibran rapat dewan pengarah
2	<i>survei populi center</i> dukungan pemilih 35 prabowogibran meningkat 547 persen	<i>survei populi center</i> dukungan pemilih 35 prabowogibran meningkat 547 persen
3	ratusan tokoh masyarakat jawa deklarasi dukung prabowogibran pilpres 2024	ratusan tokoh masyarakat jawa eklarasi dukung prabowogibran pilpres 2024
4	alasan forum ulama santri indonesia dukung prabowogibran pilpres 2024	alasan forum ulama santri indonesia dukung prabowogibran pilpres 2024
5	rapat tkn prabowogibran airlangga bahas strategis	rapat tkn prabowogibran airlangga bahas strategis

Berdasarkan dari hasil preprosesing menggunakan *google colaboraty* berbasis *python* didapatkan hasil preprosesing pada gambar berikut.

	NO	BERITA	KELAS
0	1	ketum pan zulhas rinci strategi pemenangan tkn...	PEMILU
1	2	survei populi center dukungan pemilih 35 prabo...	PEMILU
2	3	ratusan tokoh masyarakat jawa deklarasi dukung...	PEMILU
3	4	alasan forum ulama santri indonesia dukung pra...	PEMILU
4	5	rapat tkn prabowogibran airlangga bahas strategis	PEMILU

Gambar 4. 3 Tampilan Hasil Proses *Case Folding*

4.2.2 Pembobotan TF- IDF

Tahapan selanjutnya setelah *preprocessing* dilakukan, kemudian peneliti memberikan bobot pada setiap kalimat dengan menggunakan pembobotan TF-IDF, pembobotan TF-IDF dilakukan untuk dimplementasikan ke dalam metode KNN sehingga memudahkan metode tersebut untuk melakukan perhitungan jarak. Pada penggabungan dua konsep perhitungan yang digunakan dalam pembobotan TF-IDF yaitu inverse dan frekuensi kemunculan kata. Peneliti membuat *flowchart* untuk pembobotan TF-IDF berikut ini.



Gambar 4. 4 Pembobotan TF-IDF

Tabel 4. 12 Hasil Ekstraksi Fitur TF-IDF Berita online

HASIL EKSTRAKSI FITUR TF-IDF BERITA ONLINE	
(0, 814)	0.31080953051052124
(0, 261)	0.33403159063840226
(0, 922)	0.29433321664134726
(0, 887)	0.20509414949439395
(0, 1121)	0.27111115651346623
(0, 798)	0.29433321664134726
(0, 1042)	0.33403159063840226
(0, 942)	0.33403159063840226
(0, 1210)	0.31080953051052124
(0, 770)	0.29433321664134726
(0, 541)	0.31080953051052124
(1, 849)	0.30744423665613235
(1, 34)	0.3489109675582255
(1, 673)	0.3246544849530058
(1, 26)	0.3489109675582255
(1, 801)	0.3489109675582255
(1, 310)	0.3246544849530058
(1, 221)	0.30744423665613235
(1, 882)	0.30744423665613235
(1, 1057)	0.30744423665613235
(1, 887)	0.21423003136875587
(2, 18)	0.2772191583381771
(2, 868)	0.32054938763703117
(2, 309)	0.3029583988019973
(2, 255)	0.3597202189530804
:	:
(246, 389)	0.22863188047325678
(246, 710)	0.3413262926145909
(246, 578)	0.3232322556663605
(247, 497)	0.3922411702320599
(247, 292)	0.3922411702320599
(247, 223)	0.3922411702320599
(247, 247)	0.3922411702320599
(247, 332)	0.3922411702320599
(247, 453)	0.24831489285709557
(247, 389)	0.24447080716247319
(247, 1173)	0.33061769508848005
(248, 528)	0.48973689538951704
(248, 603)	0.48973689538951704
(248, 769)	0.38454277967496103
(248, 978)	0.4315335428378243
(248, 841)	0.4315335428378243

(249, 114)	0.34146875889099815
(249, 1106)	0.34146875889099815
(249, 413)	0.34146875889099815
(249, 1008)	0.34146875889099815
(249, 1139)	0.34146875889099815
(249, 956)	0.34146875889099815
(249, 251)	0.34146875889099815
(249, 623)	0.3177296627306483
(249, 761)	0.2878219385855743

1) *TF-IDF*

Untuk mencari *Term frequency* di atas dengan cara menghitung jumlah kata yang sama dan muncul pada se

tiap dokumen, kemudian menjumlahkan tiap kata yang muncul dari tiap-tiap dokumen yang ada. Dari hasil term frequency tersebut langkah selanjutnya melakukan proses normalisasi pada *Term Frequency* dengan cara membagi masing-masing dokumen dengan panjang dokumen. Untuk Dokumen pada tabel diberi keterangan T1- T11. Berikut ini tabel dari DF Normalisasi.

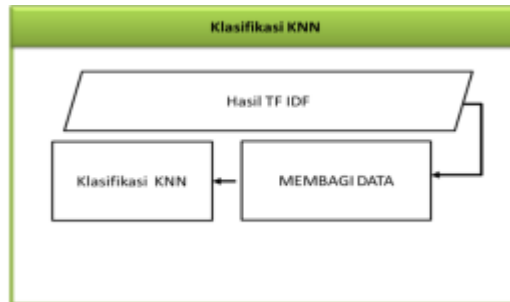
Sebelum melakukan perhitungan IDF, langkah selanjutnya adalah menghitung jumlah DF. Setelah dilakukan perhitungan DF kemudian melakukan perhitungan IDF. Dengan cara logaritma jumlah dokumen di bagi Dokumen Frekuensi. Berikut ini merupakan tabel dari perhitungan DF dan IDF.

Tabel 4. 13 Normalisasi TF-IDF

HASIL TF NORMALISASI							
No	T1	T2	T3	T4	T5	T11
0	0.31081	0.334032	0.294333	0.205094	0.271111	0.31081
1	0.307444	0.348911	0.324654	0.348911	0.348911	0
....	0
247	0.392241	0.392241	0.392241	0.392241	0.392241	0
248	0.489737	0.489737	0.384543	0.431534	0.431534	0
249	0.341469	0.341469	0.341469	0.341469	0.341469	0

4.2.3 Klasifikasi KNN

Proses Klasifikasi *KNN* dilakukan Setelah pembobotan menggunakan TF-IDF erhasildilakukan. Pada tahapan ini peneliti melakukan klasifikasi menggunakan algoritma *KNN classifier* dengan *google colab*. Dibawah ini adalah *flowchart* untuk proses klasifikasi *KNN*



Gambar 4. 5 Klasifikasi KNN

Membagi data adalah tahapan yang dilakukan peneliti untuk membagi dataset menjadi data *testing* dan data *training*, proses bagi data dilakukan sebelum proses klasifikasi. Dilanjutkan dengan proses klasifikasi, peneliti menggunakan *google colab* untuk proses klasifikasi *KNN*. Selanjtnya peneliti melakukan perhitungan untuk mendapatkan hasil klasifikasi berdasarkan model yang di bangun. Pada perhitungan ini peneliti. Untuk total data *training* menggunakan 10 dokumen dari hasil perhitungan TF-IDF dengan jumlah 2 dokumen berita pemilu dan 4 dokumen berita Islami. Untuk Dokumen pada tabel diberi keterangan T1-T11. Berikut adalah tabel dari data *training*.

Tabel 4. 14 Data *Training* Sebelum Preprosesing dan Pembobotan TF- IDF

BERITA	KELAS
Ketum PAN Zulhas Rinci Strategi Pemenangan TKN Prabowo-Gibran Usai Rapat Dewan Pengarah	PEMILU
Survei Populi Center: Dukungan Pemilih di Bawah 35 Tahun kepada Prabowo-Gibran Meningkat Jadi 54,7 Persen	PEMILU
Obat dan Kunci Agar Tak Patah Hati ala Gus Iqdam, Murah dan Mujarab	ISLAM
5 Faktor Utama Dakwah Gus Iqdam Cepat Diterima Semua Kalangan	ISLAM
Lika-Liku Sejarah Kemerdekaan Palestina dan Perjuangan yang Belum Usai hingga Kini	ISLAM
4 Malaikat yang Datangi Orang Sakit dan Tugasnya, Simak Hikmah Tersembunyi di Balikny	ISLAM

Tabel 4. 15 Data *Training* Setelah melalui *Preprosesing* dan Pembobotan TF- IDF

BERITA										
T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	T11
0.31081	0.334032	0.294333	0.205094	0.271111	0.294333	0.334032	0.334032	0.31081	0.294333	0.31081
0.307444	0.348911	0.324654	0.348911	0.348911	0.324654	0.307444	0.307444	0.307444	0.21423	0
0.392241	0.392241	0.392241	0.392241	0.392241	0.248315	0.244471	0.330618	0	0	0
0.489737	0.489737	0.384543	0.431534	0.431534	0	0	0	0	0	0
0.341469	0.341469	0.341469	0.341469	0.341469	0.341469	0.341469	0.31773	0.287822	0	0

Tabel 4. 16 Data *Training* Sebelum *Preprosesing* data dan Pembobotan TF- IDF

BERITA	KELAS
6 Amalan untuk Meraih Jannatul Firdaus, Surga dengan Level Kenikmatan Tertinggi	?

Tabel 4. 17 Data *Training* Setelah melalui *Preprosesing* data dan Pembobotan TF- IDF

BERITA										
T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	T11
0.353553	0.353553	0.353553	0.353553	0.353553	0.353553	0.353553	0	0	0	0

- Menentukan parameter K, yaitu jumlah tetangga terdekat yang akan digunakan. Misalnya, kita menggunakan $K = 1$.

6. Menghitung jarak antara dokumen baru dan setiap dokumen dalam dataset menggunakan metode yang sesuai, misalnya jarak *Euclidean*.

$$\text{Sqrt}((0.310809531-0.353553391)^2+(0.334031591-0.353553391)^2+(0.294333217-0.353553391)^2+(0.205094149-0.353553391)^2+(0.271111157-0.353553391)^2+(0.294333217-0.353553391)^2+(0.334031591-0.353553391)^2+(0.310809531-0.353553391)^2+(0.294333217-0.353553391)^2+(0.310809531-0.353553391)^2)=\mathbf{0.655632854}$$

$$\text{Sqrt}((0.307444237-0.353553391)^2+(0.348910968-0.353553391)^2+(0.324654485-0.353553391)^2+(0.348910968-0.353553391)^2+(0.348910968-0.353553391)^2+(0.324654485-0.353553391)^2+(0.307444237-0.353553391)^2+(0.307444237-0.353553391)^2+(0.307444237-0.353553391)^2+(0.214230031-0.353553391)^2+(0-0.353553391)^2)=\mathbf{0.490841605}$$

$$\text{Sqrt}((0.39224117-0.353553391)^2+(0.39224117-0.353553391)^2+(0.39224117-0.353553391)^2+(0.39224117-0.353553391)^2+(0.248314893-0.353553391)^2+(0.244470807-0.353553391)^2+(0.330617695-0.353553391)^2+(0-0.353553391)^2+(0-0.353553391)^2+(0-0.353553391)^2)=\mathbf{0.373852823}$$

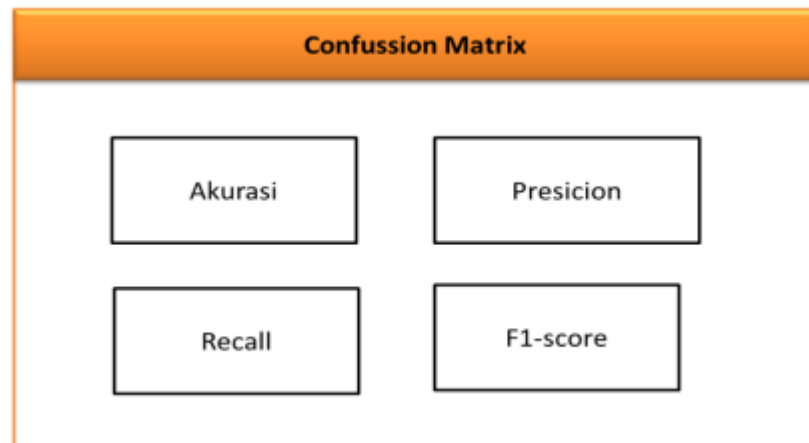
$$\text{Sqrt}((0.489736895-0.353553391)^2+(0.489736895-0.353553391)^2+(0.38454278-0.353553391)^2+(0.431533543-0.353553391)^2+(0.431533543-0.353553391)^2+(0-0.353553391)^2+(0-0.353553391)^2+(0-0.353553391)^2+(0-0.353553391)^2+(0-0.353553391)^2+(0-0.353553391)^2)=\mathbf{0.547917918}$$

$$\text{Sqrt}((0.341468759-0.353553391)^2+(0.341468759-0.353553391)^2+(0.341468759-0.353553391)^2+(0.341468759-0.353553391)^2+(0.341468759-0.353553391)^2+(0.341468759-0.353553391)^2+(0.317729663-0.353553391)^2+(0.287821939-0.353553391)^2+(0-0.353553391)^2+(0-0.353553391)^2)=\mathbf{0.429902169}$$

Berdasarkan dari hasil perhitungan jarak, maka diperoleh nilai K terkecil yaitu 0.373852823 dimana terdapat pada posisi data ke tiga pada data *training* yang berarti diperoleh hasil klasifikasi dengan kelas **ISLAM**.

4.2.4 Confusion Matrix

Untuk mengetahui kualitas dari hasil penerapan metode diperlukan evaluasi metode , dalam evaluasi metode atau algoritma peneliti menggunakan pengujian atau evaluasi *confusion matrix*. Berikut adalah *flowchart confusion matrix*.



Gambar 4. 6 *Confussion Matrix*

Confusion matrix digunakan untuk mengevaluasi sistem klasifikasi yang berdasarkan kemampuan parameter prediktif suatu model. parameter prediktif adalah parameter yang mengukur kualitas set uji statistik yang diberikan berdasarkan atribut yang ada dalam sampel. Di bawah ini adalah dataset yang digunakan untuk melakukan pengujian terhadap model yang telah dibangun. Pengujian telah dilakukan dengan dua kombinasi pembagian data antara data *testing* dengan data *training*, pengujian yang dilakukan dengangan kombinasi 20:80 dengan nilai K2, K3 dan K=4 memperoleh hasil akurasi terbaik pada nilai K=4 yaitu sebesar 91%, selanjutnya pengujian dilakukan dengan kombinasi 30:70 memperoleh hasil akurasi terbaik pada nilai K=3 yaitu sebesar 88%, sehingga model terbaik diperoleh pada kombinasi 20 : 80 dengan hasil akurasi yaitu 91%. Hasil pengujian *Confusion matrix* dapat dilihat pada tampilan berikut.

	precision	recall	f1-score	support
Bisnis	0.79	0.95	0.86	20
Bola	0.91	1.00	0.95	29
Islami	0.87	0.96	0.91	27
Pemilu	0.92	0.83	0.87	29
Showbiz	1.00	0.76	0.86	29
Tekno	0.96	0.96	0.96	24
accuracy			0.91	158
macro avg	0.91	0.91	0.90	158
weighted avg	0.91	0.91	0.90	158

Gambar 4. 7 Hasil *Confussion Matrix* pada *python*

Untuk memahami detail hasil kinerja model untuk setiap kelas berita yang berhasil diklasifikasikan, maka dengan melakukan pengukuran *Accuracy*, *Precision*, *recall* dan *f1-Score*.

- 1) *Accuracy* merupakan pengukuran yang menggambarkan seberapa akurat model dalam mengklasifikasikan dengan benar.

$$Akurasi = \frac{Jumlah\ Prediksi\ Benar(TP + TN)}{Jumlah\ Seluruh\ Sampel}$$

$$Akurasi = \frac{20 + 27 + 26 + 25 + 23 + 23}{158} = 0,91$$

- 2) *Precision* merupakan pengukuran yang menggambarkan akurasi antara data yang diminta dengan hasil prediksi yang diberikan oleh model. Pengukuran *Precision* dilakukan untuk masing-masing kelas berita karena kelasnya lebih dari 2 (dua) :

1. Perhitungan *Precision* untuk kelas berita Bisnis :

$$Precision = \frac{True\ Positif\ (TP)}{TP + FP}$$

$$Precision = \frac{20}{20 + 2 + 1 + 1 + 1 + 1} = 0,77$$

- 2 . Perhitungan *Precision* untuk kelas berita Bola :

$$Precision = \frac{27}{27 + 0 + 0 + 1 + 1 + 0} = 0,93$$

Perhitungan *Precision* untuk kelas lainnya dilakukan dengan cara sama, adapun hasil perhitungannya secara lengkap dapat dilihat pada Tabel 4.17.

- 3.) *Recall* merupakan pengukuran yang menggambarkan keberhasilan model dalam menemukan kembali sebuah informasi.

Pengukuran *Recall* dilakukan juga untuk masing-masing kelas berita :

1. Perhitungan *Recall* untuk kelas berita Bisnis

$$Recall = \frac{True\ Positif\ (TP)}{TP + FN}$$

$$Recall = \frac{20}{20 + 0 + 0 + 0 + 0 + 0} = 1,0$$

2. Perhitungan *Recall* untuk kelas berita Bola

$$Recall = \frac{27}{27 + 2 + 0 + 0 + 0 + 0} = 0,93$$

Perhitungan *Recall* untuk kelas lainnya dilakukan dengan cara sama, dan *F1-Score* merupakan pengukuran yang menggambarkan perbandingan rata-rata *precision* dan *recall*. Adapun hasil perhitungannya secara lengkap dapat dilihat pada Tabel 4.17.

Tabel 4. 18 Rekap *Confusion Matriks* Hasil Pemodelan

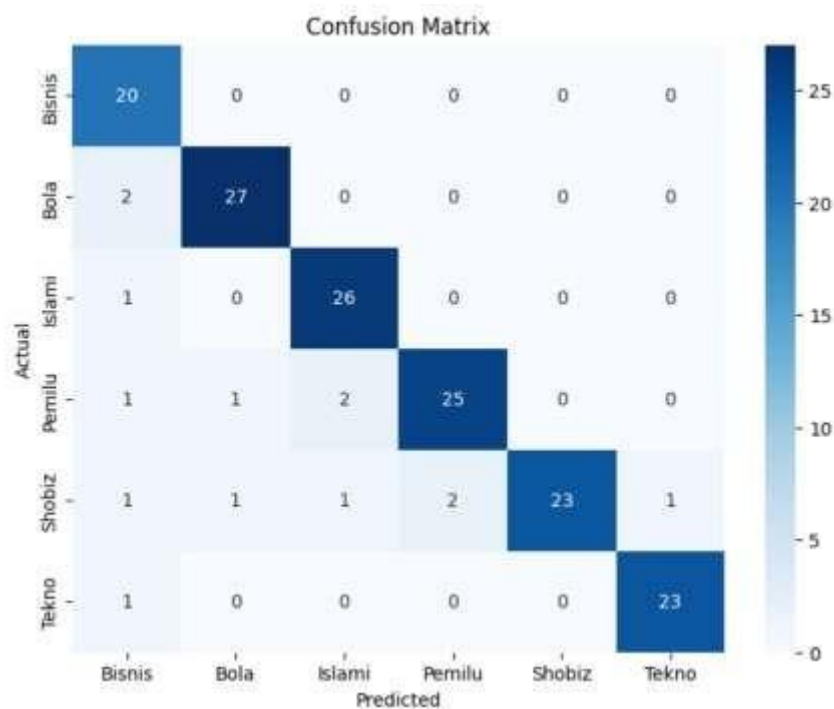
Kelas Berita	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>	Jumlah Data
Bisnis	0,77	1,00	0,87	20
Bola	0,93	0,93	0,93	29
Islami	0,90	0,96	0,93	27
Pemilu	0,93	0,86	0,89	29
Shohbiz	1,00	0,79	0,88	29
Tekno	0,96	0,96	0,96	24
<i>Accuracy</i>	0,91			

BAB V

PEMBAHASAN

1.1 Pembahasan Kinerja Model

Klasifikasi berita online pada Liputan 6 dengan menggunakan model yang sudah dilatih yaitu metode KNN dilakukan pengukuran kinerja model dengan menggunakan metode *Confusion Matriks* pada data testing sebanyak 25 data. Adapun hasil pengukuran kinerja model tersebut dalam melakukan klasifikasi berita online antaradata aktual dan hasil klasifikasi dapat dilihat pada Gambar 5.1 berikut :



Gambar 5. 1 *Confusion Matriks Data Testing*

Berdasarkan Gambar 5.1 di atas dapat dilihat bahwa model KNN sudah dapat melakukan klasifikasi berita online dengan baik, namun tetap masih ada kesalahan dalam melakukan klasifikasi misalnya pada berita kelas Bisnis, dimana model berhasil melakukan klasifikasi dengan benar sebanyak 20 judul berita tetapi terdapat 6 kesalahan klasifikasi yaitu 2 judul seharusnya kelas bola, masing-masing 1 judul seharusnya Islami, Pemilu, Shobiz dan Tekno, untuk mengetahui detail hasil kinerja model untuk

setiap kelas berita yang berhasil diklasifikasikan dengan diukur kinerja model yang terdiri dari Akurasi (menngambarkan seberapa akurat model dalam mengklasifikasikandengan benar), *Precision* (menggambarkan akurasi antara data yang diminta dengan hasil prediksi yang diberikan oleh model), *Recall* (menggambarkan keberhasilan model dalam menemukan kembali sebuah informasi) dan *F1-Score* (menggambarkan perbandingan rata-rata *precision* dan *recall* yang dibobotkan) dapat dihitung sebagai berikut :

$$Akurasi = \frac{Jumlah\ Prediksi\ Benar (TP + TN)}{Jumlah\ Seluruh\ Sampel}$$

$$Akurasi = \frac{20 + 27 + 26 + 25 + 23 + 23}{158} = 0,91$$

Pengukuran *Precision* dilakukan untuk masing-masing kelas berita karena kelasnya lebih dari 2 (dua) :

1) Perhitungan *Precision* untuk kelas berita Bisnis :

$$Precision = \frac{True\ Positif\ (TP)}{TP + FP}$$

$$Precision = \frac{20}{20 + 2 + 1 + 1 + 1 + 1} = 0,77$$

2) Perhitungan *Precision* untuk kelas berita Bola :

$$Precision = \frac{27}{27 + 0 + 0 + 1 + 1 + 0} = 0,93$$

Perhitungan *Precision* untuk kelas lainnya dilakukan dengan cara sama, adapun hasil perhitungannya secara lengkap dapat dilihat pada Tabel 5.1.

Pengukuran *Recall* dilakukan juga untuk masing-masing kelas berita :

3. Perhitungan *Recall* untuk kelas berita Bisnis

$$Recall = \frac{True\ Positif\ (TP)}{TP + FN}$$

$$Recall = \frac{20}{20 + 0 + 0 + 0 + 0 + 0} = 1,0$$

4. Perhitungan *Recall* untuk kelas berita Bola

$$Recall = \frac{27}{27 + 2 + 0 + 0 + 0 + 0} = 0,93$$

Perhitungan *Recall* untuk kelas lainnya dilakukan dengan cara sama, adapun hasil perhitungannya secara lengkap dapat dilihat pada Tabel 5.1.

Tabel 5. 1 Rekap *Confusion Matriks* Hasil Pemodelan

Kelas Berita	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>	Jumlah Data
Bisnis	0,77	1,00	0,87	20
Bola	0,93	0,93	0,93	29
Islami	0,90	0,96	0,93	27
Pemilu	0,93	0,86	0,89	29
Shohbiz	1,00	0,79	0,88	29
Tekno	0,96	0,96	0,96	24
<i>Accuracy</i>	0,91			

Catatan: isikan kolom-kolom di atas sesuai dari hasil coding sbb :

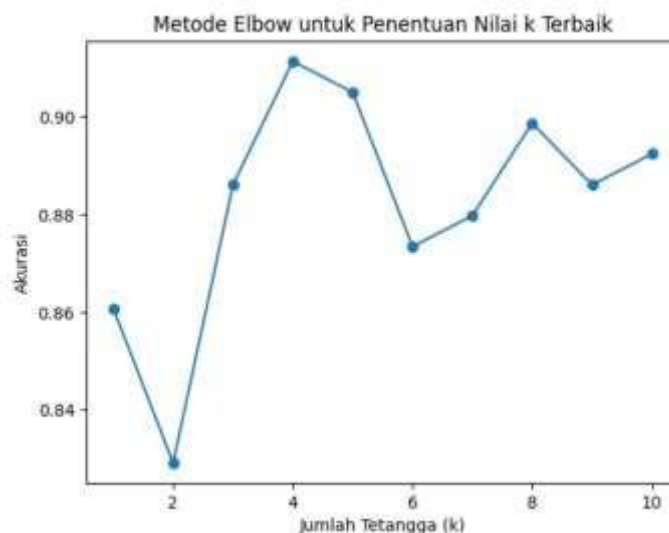
	precision	recall	f1-score	support
Bisnis	0.77	1.00	0.87	20
Bola	0.93	0.93	0.93	29
Islami	0.90	0.96	0.93	27
Pemilu	0.93	0.86	0.89	29
Showbiz	1.00	0.79	0.88	29
Tekno	0.96	0.96	0.96	24
accuracy			0.91	158
macro avg	0.91	0.92	0.91	158
weighted avg	0.92	0.91	0.91	158

Berdasarkan Tabel 5.1 di atas dapat dilihat bahwa hasil akurasi terhadap data testing sebanyak ... data yang sudah diklasifikasi didapatkan hasil akurasi sebesar 91 %.

5.2 Pembahasan Hasil Pemodelan

Model KNN dalam klasifikasi berita online yang paling optimal atau memiliki akurasi paling tinggi perlu dilakukan beberapa percobaan, untuk metode KNN harus dicari nilai k yang paling baik dan persentase pembagian data training dan data testing apakah digunakan perbandingan 80:20 atau 70:30. Hal ini bisa dilakukan dengan cepat dengan menggunakan teknik *Elbow* dalam penentuan nilai k terbaik.

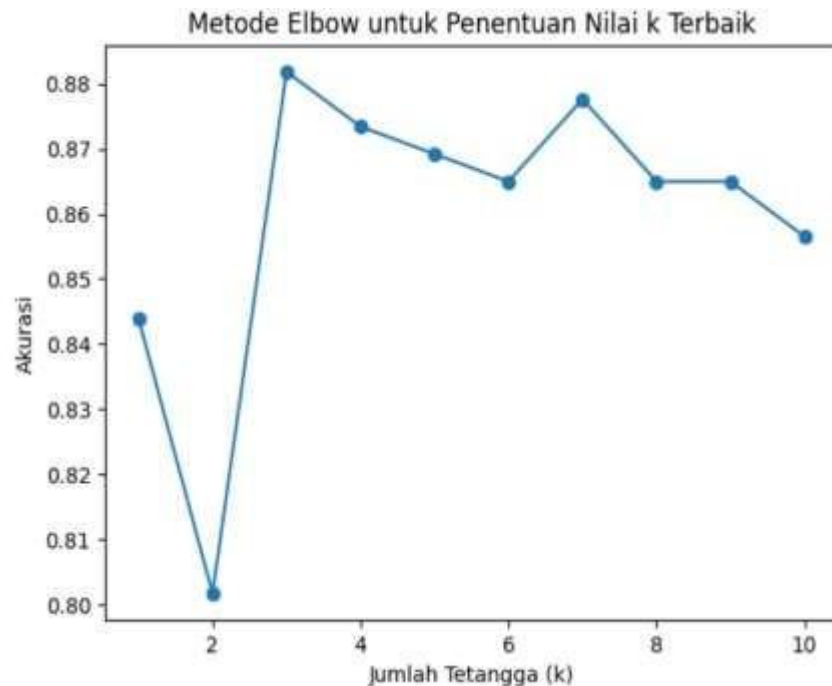
Percobaan pertama digunakan perbandingan data 80:20 dengan menggunakan teknik *Elbow*, maka dengan cepat bisa diketahui nilai k terbaik yang memiliki akurasi tertinggi tanpa perlu melakukan percobaan secara berulang untuk nilai $k=2$, $k=3$, $k=4$ dan seterusnya, dengan teknik *Elbow* cukup melihat titik nilai k yang memiliki akurasi yang tinggi, seperti pada Gambar 5.2 berikut ini :



Gambar 5. 2 Nilai k Terbaik Perbandingan Data 80:20

Berdasarkan Gambar 5.2 di atas langsung dapat diketahui nilai k terbaik yaitu $k=4$ dengan nilai akurasi sebesar 91%.

Percobaan kedua menggunakan perbandingan data 70:30, kemudian digunakan kembali teknik *Elbow* untuk menentukan nilai k terbaik, sehingga hasilnya dapat dilihat pada Gambar 5.3 berikut ini :



Gambar 5.3 Nilai k Terbaik Perbandingan Data 70:30

Berdasarkan Gambar 5.3 di atas langsung dapat diketahui nilai k terbaik yaitu $k=3$ dengan nilai akurasi sebesar 88%. Dari 2 (dua) percobaan yang dilakukan dapat ambil keputusan bahwa pada pemodelan KNN klasifikasi berita online ini digunakan pembagian data *training* dan data testing dengan persentase 80:20.

Setelah didapatkan model KNN untuk klasifikasi berita online dengan akurasi paing tinggi, maka model tersebut sudah dapat digunakan untuk melakukan klasifikasi berita online yang baru. Adapun data berita online yang dijadikan sebagai pengujian percobaan yaitu diambil dari laman liputan6.com yang diterbitkan pada tanggal 26 November 2023. Judul berita online tersebut diunduh dengan menggunakan teknik *Web Scraping* yang hasilnya disimpan dalam bentuk file *Excel*. Adapun judul berita yang berhasil diunduh sebanyak 25 judul berita, lengkapnya dapat dilihat pada Tabel 5.2 berikut ini :

Tabel 5. 2 Sample Judul Berita Baru

No	Judul Berita	Kelas
1	<i>nfografis</i> perempat final piala dunia u17 indonesia 2023 bagan jadwal <i>stadion live pe</i>	?
2	bola ganjil pencetak hattrick sempurna 3 tubuh berbeda beruntu	?
3	infografis daftar upah minimum provinsi 2024	?
4	viral tiktok ilusi optik ketakutan terbesar hidup	?
5	lee junho indonesia konser yaa	?
6	motaz azaiza jurnalis palestina dinobatkan majalah <i>gq man of the year</i>	?
7	5 tanda serangan jantung malam diwaspadai	?
8	timnas amin sebentar aniesmuhammad salip prabowogibran jabar	?
9	diet ketat tampil optimal antman paul rudd jadikan <i>sparkling water rewar</i>	?
10	mengejek kadal julurkan lidah biru intimidasi pemangsa	?
11	xl axiata polres garut bekuk kawanan petugas gadungan xl home	?
12	glass beams ajak penonton menari joyland festival 2023 diguyur hujan	?
13	hilang 2 pekan siswi smp pingsan ditemukan polisi salon	?
14	6 jenis olahraga cocok si introvert	?
15	7 tips raih skor ujian <i>toefl</i>	?
16	viral kisah wanita 37 alergi bertahan hidup	?
17	dabbah muncul <i>huraisy</i> hewan mengerikan neraka jahanam tanda kiamat	?
18	rusia serang ukraina 75 drone 6 jam rekor invasi	?
19	indodax binance berpeluang positif bitcoin halving	?
20	bawaslu luncurkan aplikasi pengawasan kampanye pemilu 2024	?
21	nissan investasi produksi 2 mobil listrik inggris	?
22	timnas amin hasil survei internal aniesmuhammad menang pilpres 2024	?
23	microsoft uji copilot windows 10 cek pakainya	?
24	ayat seribu dinar dibaca kali manfaat mengamalkannya	?
25	jadwal live streaming bri liga 1 arema vs persik vidio	?

Sebelum dilakukan klasifikasi dengan KNN judul berita baru tersebut di atas, terlebih dahulu dilakukan proses preprocessing data dan transformasi teks berita menjadi nilai numerik. Setelah itu baru dilakukan proses klasifikasi dengan metode KNN, maka hasilnya dapat dilihat pada Tabel 5.3 berikut ini :

Tabel 5. 3 Hasil Klasifikasi Judul Berita Baru

No	Judul Berita	Kelas
1	<i>nfografis</i> perempat final piala dunia u17 indonesia 2023 bagan jadwal <i>stadion live pe</i>	Bola
2	bola ganjil pencetak hattrick sempurna 3 tubuh berbeda beruntu	Bola
3	infografis daftar upah minimum provinsi 2024	Bisnis
4	viral tiktok ilusi optik ketakutan terbesar hidup	Bisnis
5	lee junho indonesia konser yaa	Showbiz
6	motaz azaiza jurnalis palestina dinobatkan majalah <i>gq</i> <i>man of the year</i>	Showbiz
7	5 tanda serangan jantung malam diwaspadai	Tekno
8	timnas amin sebentar aniesmuhammad salip prabowogibran jabar	Pemilu
9	diet ketat tampil optimal antman paul rudd jadikan <i>sparkling water rewar</i>	Bola
10	mengejek kadal julurkan lidah biru intimidasi pemangsa	Pemilu
11	xl axiata polres garut bekuk kawanan petugas gadungan xl home	Bola
12	glass beams ajak penonton menari joyland festival 2023 diguyur hujan	Bisnis
13	hilang 2 pekan siswi smp pingsan ditemukan polisi salon	Bola
14	6 jenis olahraga cocok si introvert	Showbiz
15	7 tips raih skor ujian <i>toefl</i>	Tekno
16	viral kisah wanita 37 alergi bertahan hidup	Islami
17	dabbah muncul huraisy hewan mengerikan neraka jahanam tanda kiamat	Islami
18	rusia serang ukraina 75 drone 6 jam rekor invasi	Bisnis
19	indodax binance berpeluang positif bitcoin halving	Bola
20	bawaslu luncurkan aplikasi pengawasan kampanye pemilu 2024	Pemilu
21	nissan investasi produksi 2 mobil listrik inggris	Bisnis
22	timnas amin hasil survei internal aniesmuhammad menang pilpres 2024	Pemilu
23	microsoft uji copilot windows 10 cek pakainya	Tekno
24	ayat seribu dinar dibaca kali manfaat mengamalkannya	Bola
25	jadwal live streaming bri liga 1 arema vs persik vidio	Bola

Berdasarkan Tabel 5.3 di atas dapat dilihat bahwa dari 25 judul berita baru dapat dilakukan klasifikasi jenis berita dengan baik dan benar.

BAB VI

KESIMPULAN DAN SARAN

A. Kesimpulan

Berdasarkan hasil penelitian yang diuraikan di atas mengenai klasifikasi berita online pada liputan6 dengan metode KNN dapat ditarik kesimpulan bahwa :

1. Penerapan metode KNN dalam klasifikasi berita online pada liputan6 dapat diterapkan dengan baik untuk melakukan klasifikasi berita kedalam 6 kelas berita yang berbeda yaitu : Bisnis, Pemilu, Islami, *Showbiz*, Bola dan Tekno
2. Metode klasifikasi KNN dapat mengklasifikasi berita online pada Liputan6 dengan akurasi sebesar 91% untuk 6 kelas berita yang berbeda

B. Saran

1. Penelitian selanjutnya dapat dilakukan dengan menggunakan data yang lebih besar dan lebih beragam. Hal ini akan dapat meningkatkan akurasi dari metode KNN.
2. Penelitian berikutnya juga dapat dilakukan untuk menguji efektivitas metode KNN dalam mengklasifikasi berita online di berbagai media berita. Hal ini akan dapat memberikan gambaran yang lebih luas tentang efektivitas metode ini.

DAFTAR PUSTAKA

- [1] D. Ariadi and K. Fithriasari, "Klasifikasi Berita Indonesia Menggunakan Metode Naive Bayesian Classification dan Support Vector Machine dengan Confix Stripping Stemmer," *J. SAINS DAN SENI ITS Vol. 4, No.2*, vol. 4, no. 2, pp. 248–253, 2015.
- [2] G. Tika, A. Adiwijaya, and N. D. W. Cahyani, "Klasifikasi Topik Berita Berbahasa Indonesia Menggunakan Multilayer Perceptron," *eProceedings Eng.*, vol. 6, no. 1, pp. 2137–2143, 2019.
- [3] 2021 Herfandi, Arman Diansyah, Eri Sasmita Susanto, "JINTEKS (Jurnal Informatika Teknologi dan Sains) ISSN 2686-3359 (Online) JINTEKS (Jurnal Informatika Teknologi dan Sains) ISSN 2686-3359 (Online)," (*Jurnal Inform. Teknol. dan Sains*, vol. 3, no. 1, pp. 301–307, 2021).
- [4] R. Sari, "Analisis Sentimen Pada Review Objek Wisata Dunia Fantasi Menggunakan Algoritma K-Nearest Neighbor (K-Nn)," *EVOLUSI J. Sains dan Manaj.*, vol. 8, no. 1, pp. 10–17, 2020.
- [5] A. Bengnga and R. Ishak, "Prediksi Jumlah Mahasiswa Registrasi Per Semester Menggunakan Linier Regresi Pada Universitas Ichsan Gorontalo," *Ilk. J. Ilm.*, vol. 10, no. 2, pp. 136–143, 2018.
- [6] I. A. Ahmad Sabri, M. Man, W. A. W. Abu Bakar, and A. N. Mohd Rose, "Web Data Extraction Approach for Deep Web using WEIDJ," *Procedia Comput. Sci.*, vol. 163, pp. 417–426, 2019.
- [7] "Data Mining: Pengertian, Fungsi, Penerapan, dan Metode Pengambilannya - Cermati.com."
- [8] I. S. K. Idris, Y. A. Mustofa, and I. A. Salihi, "Analisis Sentimen Terhadap Penggunaan Aplikasi Shopee Menggunakan Algoritma Support Vector Machine (SVM)," *Jambura J. Electr. Electron. Eng.*, vol. 5, no. 1, pp. 32–35, 2023.
- [9] A. Firdaus and W. I. Firdaus, "Text Mining Dan Pola Algoritma Dalam Penyelesaian Masalah Informasi : (Sebuah Ulasan)," *J. JUPITER*, vol. 13, no. 1, p. 66, 2021.
- [10] R. Feldman and J. Sanger, *The Text Mining Handbook*. 2006.
- [11] M. Learning, "Learning basic concept," pp. 1–39, 2019.
- [12] O. Sakti *et al.*, "Klasifikasi Teks Dengan Menggunakan Algoritma K-Nearest," vol. 5, no. 3, pp. 8237–8248, 2018.