

**KLASIFIKASI ARTIKEL BERITA BERDASARKAN JENIS
TOPIKNYA MENGGUNAKAN DECISION TREE**

SKRIPSI

Digunakan Sebagai Syarat Maju Ujian Diploma IV
Politeknik Negeri Malang

Oleh:

REZALDY RAFI NIM. 1941720213



**PROGRAM STUDI TEKNIK INFORMATIKA
JURUSAN TEKNOLOGI INFORMASI
POLITEKNIK NEGERI MALANG
2023**

HALAMAN PENGESAHAN

KLASIFIKASI ARTIKEL BERITA BERDASARKAN JENIS TOPIKNYA MENGGUNAKAN DECISION TREE

Disusun oleh:

REZALDY RAFLI. NIM. 1941720213

Laporan Skripsi ini telah diuji pada tanggal 07 Juli 2023

Disetujui oleh:

1. Pembimbing Utama : Mungki Astiningrum, S.T., MKom
NIP. 197710302005012001
2. Pembimbing Pendamping : Candra Bella Vista, S.Kom., MT.
NIP. 199412172019032020
3. Penguji Utama : Noprianto, S.Kom., M.Eng.
NIP. 198911082019031020
4. Penguji Pendamping : Dika Rizky Yuniarto, S.Kom,
M.Kom.
NIP. 199206062019031017

Mengetahui,

Ketua Jurusan
Teknologi Informasi

Ketua Program Studi
Teknik Informatika

Dr. Eng. Rosa Andrie Asmara, ST., MT.

NIP. 19801010 200501 1 001

Dr. Ely Setyo Astuti, ST., MT.

NIP. 19760515 200912 2 001

PERNYATAAN

Dengan ini saya menyatakan bahwa pada Skripsi ini tidak terdapat karya, baik seluruh maupun sebagian, yang sudah pernah diajukan untuk memperoleh gelar akademik di Perguruan Tinggi manapun, dan sepanjang pengetahuan saya juga tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang lain, kecuali yang secara tertulis disitasi dalam naskah ini serta disebutkan dalam daftar sitasi/pustaka.

Malang, 21 Juni 2023

Rezaldy Rafi.

ABSTRAK

Rafi., Rezaldy. “Klasifikasi Artikel Berita Berdasarkan Jenis Topiknya Menggunakan Decision Tree”.

Pembimbing: (1) Mungki Astiningrum, S.T., MKom. (2) Candra Bella Vista, S.Kom., MT.

Skripsi, Program Studi Teknik Informatika, Jurusan Teknologi Informasi, Politeknik Negeri Malang, 2023.

Klasifikasi teks berita merupakan salah satu aplikasi penting dalam bidang pemrosesan bahasa alami dan analisis teks. Dalam penelitian ini, penulis menggunakan algoritma decision tree sebagai metode klasifikasi untuk mengidentifikasi kategori yang tepat dari teks berita. Penelitian ini dimulai dengan mengumpulkan dataset teks berita yang telah diberi label kategori yang sesuai. Selanjutnya, penulis menerapkan metode ekstraksi fitur TF. Setelah itu, penulis membagi dataset menjadi dua bagian data pelatihan dan data pengujian. Hasil eksperimen dan evaluasi menunjukkan bahwa penggunaan decision tree dalam klasifikasi teks berita menghasilkan kinerja yang baik dengan akurasi sistem yang cukup tinggi yaitu sebesar 78%. Hasil penelitian ini dapat digunakan sebagai landasan untuk pengembangan lebih lanjut dalam bidang klasifikasi teks berita dan menjadi referensi bagi peneliti yang tertarik dalam menerapkan decision tree pada tugas-tugas klasifikasi teks serupa.

Kata Kunci : , Klasifikasi teks berita, Decision tree, Analisis teks

ABSTRACT

Rafi., Rezaldy. “Classification of News Articles Based on the Type of Topics Using a Decision Tree”. **Supervisor: Mungki Astiningrum, S.T., MKom. Candra Bella Vista, S.Kom., MT.**

Thesis, Informatics Engineering Study Program, Department of Information Technology, State Polytechnic of Malang, 2023.

News text classification is an important application in the field of natural language processing and text analysis. In this study, the authors used a decision tree algorithm as a classification method to identify the appropriate categories of news texts. This research begins by collecting news text datasets that have been labeled with the appropriate categories. Next, the authors apply the TF feature extraction method. After that, the authors divide the dataset into two parts, training data and testing data. The experimental and evaluation results show that the use of decision trees in news text classification produces good performance with a fairly high system accuracy of 78%. The results of this study can be used as a basis for further development in the field of news text classification and as a reference for researchers who are interested in applying decision trees to similar text classification tasks.

Keywords: *News text classification, Decision tree, Text analysis*

KATA PENGANTAR

Puji Syukur kami panjatkan kehadirat Allah SWT/Tuhan YME atas segala rahmat dan hidayah-Nya penulis dapat menyelesaikan skripsi dengan judul “Klasifikasi Artikel Berita Berdasarkan Jenis Topiknya Menggunakan Decision Tree”. Skripsi ini penulis susun sebagai persyaratan untuk menyelesaikan studi program Diploma IV Program Studi Teknik Informatika, Jurusan Teknologi Informasi, Politeknik Negeri Malang.

Kami menyadari bahwasannya dengan tanpa adanya dukungan dan kerja sama dari berbagai pihak, kegiatan laporan akhir ini tidak akan dapat berjalan baik. Untuk itu, kami ingin menyampaikan rasa terima kasih kepada:

1. Allah SWT yang selalu memberikan rahmat dan hidayah-Nya sehingga penulis dapat menyelesaikan skripsi ini dengan lancar.
2. Orang tua yang selalu memberikan dukungan moral, motivasi dan doa agar dimudahkan dalam mengerjakan skripsi.
3. Ibu Mungki Astiningrum, S.T., MKom., selaku dosen pembimbing utama yang telah memberi bimbingan, arahan, dan meluangkan waktu kepada penulis dalam menyusun skripsi ini dengan baik.
4. Ibu Candra Bella Vista, S.Kom., MT., selaku dosen pembimbing pendamping yang telah memberi bimbingan, arahan, dan meluangkan waktu kepada Penulis dalam menyusun skripsi ini dengan baik.
5. Bapak Dr. Eng. Rosa Andrie Asmara, ST., MT., selaku Ketua Jurusan Teknologi Informasi.
6. Ibu Dr. Ely Setyo Astuti, ST., MT., selaku Ketua Program Studi D-IV Teknik Informatika.
7. Semua teman-teman yang sudah memberikan support dan bantuannya selama pembuatan skripsi ini.
8. Dan seluruh pihak yang telah membantu dan mendukung lancarnya pembuatan Laporan Akhir dari awal hingga akhir yang tidak dapat kami sebutkan satu persatu.

Penulis menyadari bahwa dalam penyusunan laporan akhir ini, masih banyak terdapat kekurangan dan kelemahan yang dimiliki penulis baik itu sistematika penulisan maupun penggunaan bahasa. Untuk itu penulis mengharapkan saran dan

kritik dari berbagai pihak yang bersifat membangun demi penyempurnaan laporan ini. Semoga laporan ini berguna bagi pembaca secara umum dan penulis secara khusus. Akhir kata, penulis ucapkan banyak terima kasih.

Malang, 06 Juli 2023

Penulis

DAFTAR ISI

ABSTRAK	iv
ABSTRACT	vi
KATA PENGANTAR	vii
DAFTAR ISI	iii
DAFTAR GAMBAR	v
DAFTAR TABEL	vi
BAB I. PENDAHULUAN	1
1.1. Latar Belakang	1
1.2. Rumusan Masalah	2
1.3. Batasan Masalah	2
1.4. Tujuan	2
1.5. Manfaat	2
BAB II. LANDASAN TEORI	3
2.1. Studi Literatur	3
2.2. Dasar Teori	4
2.2.1. <i>Text Preprocessing</i>	4
2.2.2. Klasifikasi	5
2.2.3. <i>Decision Tree</i>	6
2.2.4. <i>Term Frequency</i>	6
2.2.5. <i>Confusion Matrix</i>	7
2.2.6. Website	8
2.2.7. PHP	8
2.2.8. MySQL	9
BAB III. METODOLOGI PENELITIAN	10
3.1. Waktu dan Tempat Penelitian	10
3.2. Teknik Pengumpulan Data	10
3.3. Teknik Pengolahan Data	10
3.4. Desain Sistem	12
3.4.1. Model <i>Decision Tree</i>	13
3.5. Uji Coba Sistem	14
BAB IV. ANALISIS DAN PERANCANGAN SISTEM	15
4.1 Deskripsi Sistem	15

4.2 Analisis Kebutuhan Fungsional.....	15
4.3 Analisis Kebutuhan Non-Fungsional	16
4.4 Perancangan Sistem.....	17
4.4.1. Perancangan Sistem Model.....	17
4.4.2. Perancangan Desain Database	18
4.4.3. Tahapan Pelatihan.....	20
4.4.4. Perancangan Antarmuka	21
BAB V. IMPLEMENTASI DAN PENGUJIAN	23
5.1 Implementasi Antarmuka	23
5.2 Implementasi Database.....	24
5.3 Implementasi Metode	24
5.4 Pengujian Sistem	26
5.4.1. Proses Pengujian Data Testing	26
5.4.2. Pengujian Model	32
5.4.3. Pengujian Fungsional Sistem.....	33
BAB VI. HASIL DAN PEMBAHASAN	35
6.1 Hasil dan Pembahasan	35
BAB VII. KESIMPULAN DAN SARAN	37
7.1 Kesimpulan.....	37
7.2 Saran	37
DAFTAR PUSTAKA	38

DAFTAR GAMBAR

Gambar 2.1 PHP	9
Gambar 3.1 Tahapan Pengolahan Data Dengan Text Processing.....	11
Gambar 3.2 Flowchart Pengolahan Data	12
Gambar 3.3 Desain Sistem Data Uji	13
Gambar 3.4 Decision Tree	14
Gambar 4.1 Desain Sistem Model	18
Gambar 4.2 Tahapan Pelatihan	20
Gambar 4.3 Halaman Perancangan Antarmuka Klasifikasi Teks.....	21
Gambar 4.4 Halaman Perancangan Antarmuka Riwayat Klasifikasi Teks.....	22
Gambar 5.1 Halaman Input Teks	23
Gambar 5.2 Halaman Riwayat Hasil Kategori.....	24
Gambar 5.3 Implementasi Database	24
Gambar 5.4 Implementasi Metode.....	25
Gambar 5.5 Akurasi Per-Kategori.....	29
Gambar 5.6 Recall Per-Kategori	30
Gambar 5.7 Precision Per-Kategori	31
Gambar 5.8 F1-Score Per-Kategori.....	32
Gambar 5.9 Hasil Pengujian Model	32
Gambar 5.10 Gambar Kode Output Akurasi.....	33
Gambar 6.1 Contoh Hasil <i>Testing</i>	35
Gambar 6.2 Hasil <i>balancing</i> data.....	36

DAFTAR TABEL

Tabel 2.1 Tabel <i>Confusion Matrix</i>	7
Tabel 3.1 Tabel Data Perkategori Topik	10
Tabel 5.1 Tabel <i>Confusion Matrix</i>	26
Tabel 5.2 Tabel <i>FP, TP, FN, dan TN</i> Kategori <i>Finance</i>	26
Tabel 5.3 Tabel <i>FP, TP, FN, dan TN</i> Kategori <i>Food</i>	26
Tabel 5.4 Tabel <i>FP, TP, FN, dan TN</i> Kategori <i>Health</i>	27
Tabel 5.5 Tabel <i>FP, TP, FN, dan TN</i> Kategori <i>News</i>	27
Tabel 5.6 Tabel <i>FP, TP, FN, dan TN</i> Kategori <i>Otomotif</i>	27
Tabel 5.7 Tabel <i>FP, TP, FN, dan TN</i> Kategori <i>Sport</i>	28
Tabel 5.8 Tabel <i>FP, TP, FN, dan TN</i> Kategori <i>Travel</i>	28
Tabel 5.9 Tabel Pengujian Halaman Uji	34

BAB I. PENDAHULUAN

1.1. Latar Belakang

Artikel berita merupakan salah satu jenis tulisan yang menyajikan informasi terkini mengenai peristiwa, kejadian, atau hal-hal lain yang terjadi di dunia. Artikel berita bisa ditemukan di media cetak seperti surat kabar atau majalah, media elektronik seperti televisi atau radio, maupun media online seperti situs berita atau aplikasi berita. Artikel berita ditulis oleh jurnalis yang melakukan riset dan mengumpulkan informasi melalui wawancara, observasi, dan pengamatan. Informasi yang dikumpulkan kemudian diolah dan ditulis dalam bentuk artikel yang menyajikan fakta-fakta yang relevan dan penting.

Pemrosesan dan pengolahan informasi merupakan salah satu kegiatan yang sangat penting dalam kehidupan sehari-hari. Di era digital seperti sekarang ini, informasi bisa dengan mudah diakses melalui berbagai media, seperti media cetak, media elektronik, maupun media online. Namun, dengan adanya banyaknya kategori yang tersedia, terkadang pembaca merasa kewalahan dalam menemukan informasi atau berita yang sesuai dengan kategori yang diinginkan oleh pembaca. Proses penentuan kategori manual memerlukan waktu yang sedikit lama karena harus membaca dahulu isi dari artikel berita tersebut, apalagi dengan banyaknya dokumen dari artikel berita yang ada di internet akan sangat merepotkan untuk mengkategorikan secara manual (Gallofré Ocaña & Opdahl, 2022).

Text processing merupakan proses pengolahan teks yang mencakup berbagai teknik dan algoritma untuk mengekstrak informasi dari teks. Ini termasuk tugas-tugas seperti pembersihan teks, ekstraksi fitur, klasifikasi teks, dan analisis sentimen. *Text processing* digunakan dalam berbagai aplikasi, seperti analisis opini, analisis teks, analisis konten media sosial, dan mesin penerjemah. Teknologi yang digunakan dalam *text processing* meliputi pemrosesan natural language, analisis komputasi, dan pembelajaran mesin.

Berdasarkan latar belakang tersebut, muncul ide untuk melakukan penelitian mengenai klasifikasi artikel berita berdasarkan jenis topiknya. Penelitian ini bertujuan untuk mengembangkan sistem yang dapat membantu dalam mengelompokkan artikel berita menurut topik yang dibahas, Dalam penelitian ini metode klasifikasi yang digunakan adalah metode *decision tree*.

Decision Tree merupakan suatu pendekatan yang sangat populer dan praktis untuk menyelesaikan permasalahan klasifikasi, dan dari penelitian sebelumnya yang membandingkan metode decision tree dengan metode lainnya yang dilakukan oleh (Maulidah et al., 2020) decision tree memiliki tingkat akurasi tertinggi, maka dari itu penelitian ini akan menggunakan decision tree untuk metode klasifikasi.

1.2 Rumusan Masalah

Rumusan masalah yang dapat diambil adalah sebagai berikut:

1. Bagaimana cara membangun sistem klasifikasi artikel berita berdasarkan jenis topiknya yang efektif menggunakan metode *decision tree*.
2. Seberapa baik akurasi dari metode *decision tree* dalam mengklasifikasi dokumen artikel berita.

1.3 Batasan Masalah

Permasalahan pada penelitian ini memiliki beberapa batasan sebagai berikut:

1. Aplikasi ini berbasis website.
2. Data yang digunakan adalah dataset dari detiknews
3. Penentuan kategori atau topik berita yang digunakan adalah *finance, food, health, news, otomotif, sport, dan travel*.

1.4 Tujuan

Tujuan dari dilakukannya skripsi dengan judul “**Klasifikasi Artikel Berita Berdasarkan Jenis Topiknya**”, adalah sebagai berikut:

1. Untuk membuat sistem yang dapat secara otomatis mengelompokkan artikel berita ke dalam kategori topik yang sesuai
2. Menguji seberapa baik akurasi metode decision tree untuk sistem klasifikasi artikel berita.

1.5. Manfaat

Manfaat dari penelitian ini adalah untuk:

1. Mempermudah dan mempercepat proses penentuan topik artikel berita.
2. Meningkatkan efisiensi dalam proses pengkategorian topik berita.

BAB II. LANDASAN TEORI

2.1 Studi Literatur

Pada penelitian ini penulis menggunakan beberapa penelitian sebelumnya sebagai penunjang penelitian ini.

Penelitian mengenai klasifikasi *Decision Tree* telah dilakukan oleh (Maulidah et al., 2020) dengan judul “Algoritma Klasifikasi *Decision Tree* Untuk Rekomendasi Buku Berdasarkan Kategori Buku”. Dalam penelitian ini sang penulis melakukan komparasi dengan beberapa model algoritma klarifikasi, dan *Decision Tree* mendapatkan hasil akurasi klasifikasi paling tinggi yaitu sebesar 99,95% dibandingkan dengan *Random Forest* dengan hasil akurasi 99,91%, *KNN* dengan akurasi 56,84%, *SVC* dengan akurasi 52,37%, dan *Naïve Bayes* dengan akurasi 10,96%. Hal itu menjadi bukti bahwa algoritma *Decision Tree* adalah algoritma yang baik untuk mengklasifikasikan kategori.

Penelitian mengenai *Decision Tree* juga dilakukan oleh (Sartika & Sensuse, 2017) dengan judul “Perbandingan Algoritma Klasifikasi *Naïve Bayes*, *Nearest Neighbour*, dan *Decision Tree* pada Studi Kasus Pengambilan Keputusan Pemilihan Pola Pakaian”. Penelitian ini bertujuan membandingkan algoritma klasifikasi *naïve bayes*, *nearest neighbour*, dan *decision tree*, ketiga algoritma metode tersebut akan dibandingkan berdasarkan tingkat akurasinya. Hasil pada penelitian tersebut menyatakan bahwa tingkat akurasi tertinggi didapatkan oleh *decision tree* yaitu mencapai 75,6% dibandingkan dengan *naïve bayes* dan *nearest neighbour*.

Penelitian mengenai klasifikasi berita juga dilakukan oleh (Pramudita et al., 2018) dengan judul “Klasifikasi Berita Olahraga Menggunakan Metode *Naïve Bayes* Dengan *Enhanced Confix Stripping Stemmer*”. Pada penelitian ini sang penulis bertujuan Menyusun berita olahraga sesuai dengan kategori yang ditentukan menggunakan metode *Naïve Bayes* dengan *Enhanced Confix Stripping Stemmer*. Hasil dari penelitian ini adalah kesalahan sistem mencapai 23% yang berarti tingkat keakuratan metode *Naïve Bayes* dengan *Enhanced Confix Stripping Stemmer* tidak cukup baik untuk pengklasifikasian suatu kategori.

Penelitian tentang *Preprocessing* pada klasifikasi telah dilakukan oleh (Ma’rifah et al., 2020) dengan judul “Klasifikasi Artikel Ilmiah Dengan Berbagai

Skenario *Preprocessing*”. Pada penelitian ini penulis menggunakan 9 skenario *preprocessing*. Hasilnya klasifikasi terbaik terdapat pada scenario ke-9 dengan hasil akurasi tertinggi dibandingkan dengan skenario lain yaitu 72,92% yang terdiri dari *Case Folding*, *Tokenizing*, *Stopwords Removal* berbasis kamus, *Stemming*, *Transformasi* ke bentuk *VSM*, *stopwords removal* berbasis frekuensi dokumen, dan *SMOTE*.

Penelitian tentang *Preprocessing* juga telah dilakukan oleh (Herwijayanti et al., 2018) dengan judul “Klasifikasi Berita *Online* dengan menggunakan Pembobotan *TF-IDF* dan *Cosine Similarity*”. Pada penelitian ini *preprocessing* digunakan untuk mendapatkan dataset yang dapat diolah dengan cepat dan menghasilkan kesimpulan yang tepat. Pengujian akurasi pada penelitian ini dilakukan sebanyak empat kali. Pada percobaan pertama dengan pembagian data latih 90% dan data uji 10% mendapatkan akurasi 80%. Pada percobaan kedua dengan pembagian data latih 80% dan data uji 20% mendapatkan akurasi 90%. Pada percobaan ketiga dengan pembagian data latih 70% dan data uji 30% mendapatkan akurasi 100%. Pada keempat dengan pembagian data latih 60% dan data uji 40% mendapatkan akurasi 95%.

Berdasarkan penelitian sebelumnya yang sudah dijelaskan pada studi literatur diatas, penelitian ini akan dilakukan menggunakan metode *decision tree* untuk melakukan klasifikasi artikel berita, dengan menggunakan metode *preprocessing Case Folding*, *Tokenizing*, *Filtering*, dan *Stemming*. Dengan metode yang akan digunakan tersebut diharapkan penelitian ini mendapatkan nilai akurasi yang tinggi.

2.2 Dasar Teori

2.2.1. Text Preprocessing

Text preprocessing menjadi tahap awal dalam *text mining*. Tujuan dari *text preprocessing* yakni menghasilkan sebuah set *term index* yang bisa mewakili dokumen. (Adhe et al., 2020). Berikut adalah tahapan proses *text processing*:

- 1) *Case Folding* merupakan proses mengubah semua huruf di dokumen menjadi huruf kecil, Sementara itu karakter selain huruf akan dihilangkan.
- 2) *Tokenizing* merupakan proses menghapus atau mengabaikan karakter yang dianggap sebagai tanda baca.

- 3) *Filtering* merupakan proses pengambilan kata yang penting dan mengabaikan kata umum dan tidak memiliki makna yang bisa disebut dengan *stopword*. Penghilangan kata *stopword* seperti yaitu, dan, serta, dan lainnya dapat mengurangi waktu pemrosesan.
- 4) *Stemming* merupakan proses untuk mengubah kata turunan menjadi kata dasar.

2.2.2. Klasifikasi

Klasifikasi adalah proses pembagian data menjadi kelompok-kelompok yang disebut kategori atau kelas. Proses klasifikasi digunakan untuk membuat kategori atau label data yang sesuai dengan karakteristik atau atribut yang dimilikinya. Proses klasifikasi dapat digunakan dalam berbagai bidang, seperti pengenalan pola, analisis data, dan pembelajaran mesin.

Klasifikasi merupakan teknik untuk mempelajari kumpulan data yang banyak sehingga menghasilkan aturan untuk mengenali data baru yang belum pernah dipelajari. Keluaran dari teknik klasifikasi adalah hasil akurasi yang menunjukkan seberapa benar teknik klasifikasi yang diterapkan dalam memprediksi. Semakin tinggi tingkat akurasi maka semakin bagus karena banyak data yang diklasifikasikan benar. (Tantyoko et al., 2019).

Dalam klasifikasi, data yang diterima akan dibandingkan dengan kategori yang sudah ditentukan sebelumnya. Data yang sesuai dengan kategori tersebut akan diklasifikasikan ke dalam kategori tersebut. Pada umumnya klasifikasi dokumen dilakukan dengan menentukan ciri-ciri atau fitur-fitur yang diwakili oleh kalimat-kalimat penting. Dalam dokumen yang berukuran besar, klasifikasi akan menjadi tantangan sistem. Klasifikasi dokumen secara otomatis dapat dilakukan dengan menggunakan ciri atau fitur kata yang muncul pada dokumen latih. (Nelly Indriani et al., 2017).

Klasifikasi merupakan suatu metode untuk mengelompokkan sebuah objek ke dalam kelompok atau kelas tertentu. Klasifikasi dokumen adalah bidang penelitian dalam perolehan informasi dengan mengembangkan metode untuk menentukan atau mengkategorikan dokumen ke dalam satu atau lebih kelompok yang sebelumnya telah diakui berdasarkan isi dokumen. (Yulia Muniar et al., 2020).

2.2.3. Decision Tree

Decision Tree merupakan pohon yang terstruktur dari kumpulan atribut untuk dapat diuji dengan mempunyai tujuan memprediksi output-nya. Dimana setiap simpul internal menunjukkan tes pada atribut, hasil dari tesnya diwakili oleh masing - masing cabang, dan label kelas dipegang oleh setiap node-nya. Di dalam pohon keputusan node paling atas merupakan simpul akar. Menentukan akar pohon menggunakan gain tertinggi dari masing - masing atribut atau dengan menurut nilai index entropy terendah.(Rizkia et al., 2019). Nilai Entropy dapat ditemukan dengan menggunakan rumus:

$$Entropy (S) = \sum_{j=1}^k - p_j \log_2 p_j \quad (1)$$

Keterangan:

S adalah ruang (data) sample yang digunakan untuk training

K adalah banyaknya partisi pada S

p_j adalah probabilitas yang didapat dari Sum (Ya) dibagi dengan total sample

Untuk nilai Gain dapat ditemukan dengan menggunakan rumus berikut ini:

$$Gain (A) = Entropy (S) - \sum_{j=1}^k \frac{|S_i|}{|S|} x Entropy (S_i) \quad (2)$$

Keterangan:

S adalah ruang (data) sample yang digunakan untuk training

A adalah atribut

$|S_i|$ adalah jumlah sample untuk nilai V

$|S|$ adalah jumlah seluruh sample data

$Entropy (S_i)$ adalah entropy untuk sample – sample yang memiliki nilai i

2.2.4. Term Frequency

Term Frequency (TF), yaitu faktor yang menentukan bobot term pada suatu dokumen berdasarkan jumlah kemunculannya dalam dokumen tersebut. Nilai jumlah kemunculan suatu kata (*term frequency*) diperhitungkan dalam pemberian bobot terhadap suatu kata. Semakin besar jumlah kemunculan suatu *term* (tf tinggi) dalam dokumen, maka akan semakin besar pula bobotnya dalam dokumen atau akan memberikan nilai kesesuaian yang semakin besar.(Yulia Muniar & Ria Lestari, 2020).

Perhitungan TF bisa menggunakan nilai dari pecahan *term* yang telah dilakukan normalisasi(Cahyono, 2017), rumus tersebut adalah:

$$TF(t, d) = 0,5 + 0,5 \frac{f(t, d)}{\max \{f(w, d): w \in d\}} \quad (3)$$

Keterangan:

(t,d) = Frekuensi kata t muncul pada dokumen d .

$\max \{f(w, d): w \in d\}$ = frekuensi maksimum dari term lain pada dokumen d .

2.2.5. Confusion Matrix

Confusion Matrix adalah suatu metode yang digunakan untuk melakukan perhitungan akurasi pada konsep *data mining*. Sehingga untuk mengetahui tingkat nilai akurasi, presisi dan recall dalam suatu klasifikasi. (Pratiwi et al., 2020).

Dalam *Confusion Matrix*, *Accuracy* menggambarkan seberapa akurat model dalam mengklasifikasi dengan benar. *Precision* menggambarkan akurasi antara data yang diminta dengan hasil prediksi yang diberikan oleh model. *Recall* atau *sensitivity* menggambarkan keberhasilan model dalam menemukan Kembali sebuah informasi.(Fauziningrum & Pd, 2021). Penjelasan *confusion matrix* dijelaskan pada Tabel 2.1 dibawah ini:

Tabel 2.1 Tabel *Confusion Matrix*

	Positive (Predicted)	Negative (Predicted)
Positive (Actual)	True Positive	False Negative
Negative (Actual)	False Positive	True Negative

Rumus *confusion matrix* untuk sebagai berikut:

Akurasi =

$$\frac{TP + TN}{TP + FP + TN + FN} \quad (4)$$

Recall =

$$\frac{TP}{TP + FN} \quad (5)$$

Precision =

$$\frac{TP}{TP + FP} \quad (6)$$

F-Measure =

$$2 * \frac{Precision * Recall}{Precision + Recall} \quad (7)$$

Keterangan:

TP = *True Positive*, adalah jumlah data positif yang terklasifikasi benar.

TN = *True Negative*, adalah jumlah data negatif yang terklasifikasi salah.

FP = *False Positive*, adalah jumlah data positif yang terklasifikasi salah.

FN = *False Negative*, adalah jumlah data negatif yang terklasifikasi salah.

2.2.6. Website

Website adalah salah satu jenis layanan yang disediakan di internet yang paling banyak digunakan dibandingkan layanan lainnya. Kelebihan website dibanding fasilitas internet lainnya adalah karena website bisa menyajikan informasi dengan fleksibel, website bisa menyajikan informasi berupa teks, gambar, suara, dan video.(Wibowo Soejono et al., 2018).

Website adalah kumpulan dari halaman-halaman situs yang terangkum dalam sebuah domain atau subdomain yang berada di dalam *World Wide Web* (WWW) di dalam internet, website juga dapat diartikan sebagai halaman yang berisi data yang dapat diakses secara online, ada banyak model pengembangan sistem yang bisa dimanfaatkan untuk membangun website.(Josi et al., 2017)

2.2.7. PHP

PHP merupakan script untuk pemrograman script web *server-side*. PHP awalnya merupakan program yang dikhususkan untuk menerima input melalui form yang ditampilkan dalam browser website. Hal yang dapat dikerjakan PHP secara mendasar antara lain mendapatkan data dari form, menghasilkan isi halaman web yang dinamik, dan menerima cookies, namun fitur PHP yang paling bisa diandalkan dan signifikan adalah dukungan kepada banyak *database*..(Sovia et al., 2017).

PHP merupakan Bahasa pemrograman berbasis web yang memiliki kemampuan untuk memproses dan mengolah data secara dinamis. PHP adalah *server-side embedded script language* yang berarti semua sintaks dan perintah program yang ditulis akan sepenuhnya dijalankan oleh server, tetapi dapat disertakan pada halaman HTML biasa.(Fadila et al., 2021).

PHP adalah bahasa pemrograman yang berbentuk script yang diletakkan didalam web server. PHP dapat diartikan sebagai Hypertext Preeprocessor. Ini merupakan bahasa yang dapat berjalan pada server yang hasilnya dapat ditampilkan pada klien.(Fadila et al., 2021)



Gambar 2.1 PHP

2.2.8. MySQL

MySQL adalah salah satu aplikasi DBMS (*Database Management System*) yang sudah banyak digunakan para pemrogram aplikasi website. MySQL merupakan sebuah sistem *database* yang relasional, MySQL bisa mengelompokkan informasi ke tabel-tabel yang berkaitan. MySQL menggunakan indeks untuk mempercepat proses pencarian terhadap baris informasi tertentu. MySQL memerlukan setidaknya satu indeks pada tiap tabel, biasanya menggunakan *primary key* sebagai pengenal unik untuk membantu penjejakan data. (Lutfi, 2017).



Gambar 2.2 MySQL

BAB III. METODOLOGI PENELITIAN

3.1. Waktu dan Tempat Penelitian

Penelitian ini dilakukan di Politeknik Negeri Malang dalam kurun waktu 6 bulan pada bulan Januari 2023 sampai Juni 2023.

3.2. Teknik Pengumpulan Data

Pengumpulan data pada penelitian ini menggunakan metode pengumpulan studi pustaka. Pada metode ini kegiatan yang dilakukan adalah mempelajari, mencari dan mengumpulkan data yang berhubungan dengan penelitian ini. Data yang digunakan untuk klasifikasi artikel berita berdasarkan jenis topiknya adalah data dari artikel berita *detiknews.com*, yang dimana data itu sudah berbentuk dataset. Dataset yang digunakan pada penelitian ini bersumber dari link berikut ini (https://github.com/ibamibrahim/dataset-judul-berita-indonesia/blob/master/detik_news_title.csv).

Dataset dibagi menjadi data latih sebanyak 80% dan data uji sebanyak 20%. Data untuk setiap topik memiliki jumlah yang berbeda beda. Jumlah dari data setiap topik ada pada Tabel 3.1 berikut:

Tabel 3.1 Tabel Data Perkategori Topik

Topik	Jumlah Data
News	25565
Finance	11341
Travel	6466
Health	3944
Otomotif	3481
Food	3444
Sport	1892
Total Data	54,809

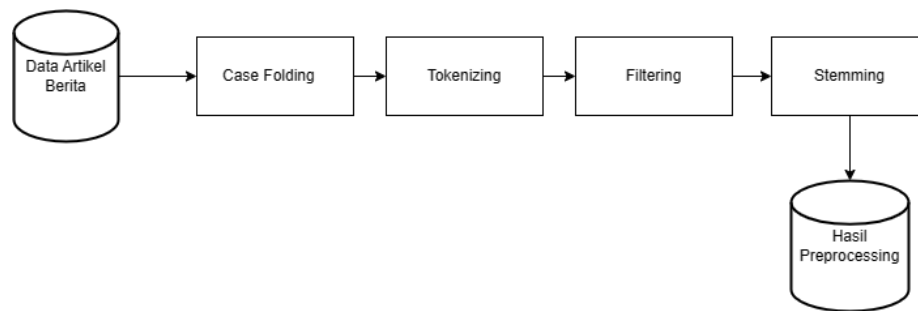
Selain menggunakan data yang ada pada dataset, data uji juga direncanakan menggunakan data dari data-data yang ada pada berita lain.

3.3 Teknik Pengolahan Data

Text Preprocessing menjadi tahapan penting dalam pengolahan data yang berupa teks, *text preprocessing* berfungsi untuk membuat data menjadi lebih

terstruktur sehingga lebih mudah untuk diproses lebih lanjut. Berikut adalah tahapan proses *text processing* yang akan digunakan pada penelitian ini:

1. Case Folding.
2. Tokenizing.
3. Filtering.
4. Stemming.

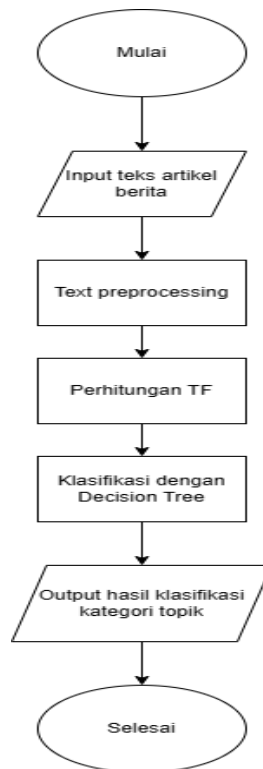


Gambar 3.1 Tahapan Pengolahan Data Dengan Text Processing

Setelah melalui tahap *text processing* yang sudah dijelaskan pada gambar 3.1 diatas, tahapan yang dilakukan selanjutnya adalah tahapan perhitungan TF. Tahapan ini digunakan untuk menghitung bobot dan frekuensi kemunculan kata dalam dokumen artikel berita. Nilai jumlah kemunculan suatu kata diperhitungkan dalam pemberian bobot terhadap suatu kata. Semakin besar jumlah kemunculan suatu *term* dalam dokumen, maka akan semakin besar pula bobotnya dalam dokumen. Perhitungan TF dapat dilakukan dengan menggunakan rumus berikut:

$$TF(t, d) = 0,5 + 0,5 \frac{f(t, d)}{\max \{f(w, d): w \in d\}} \quad (8)$$

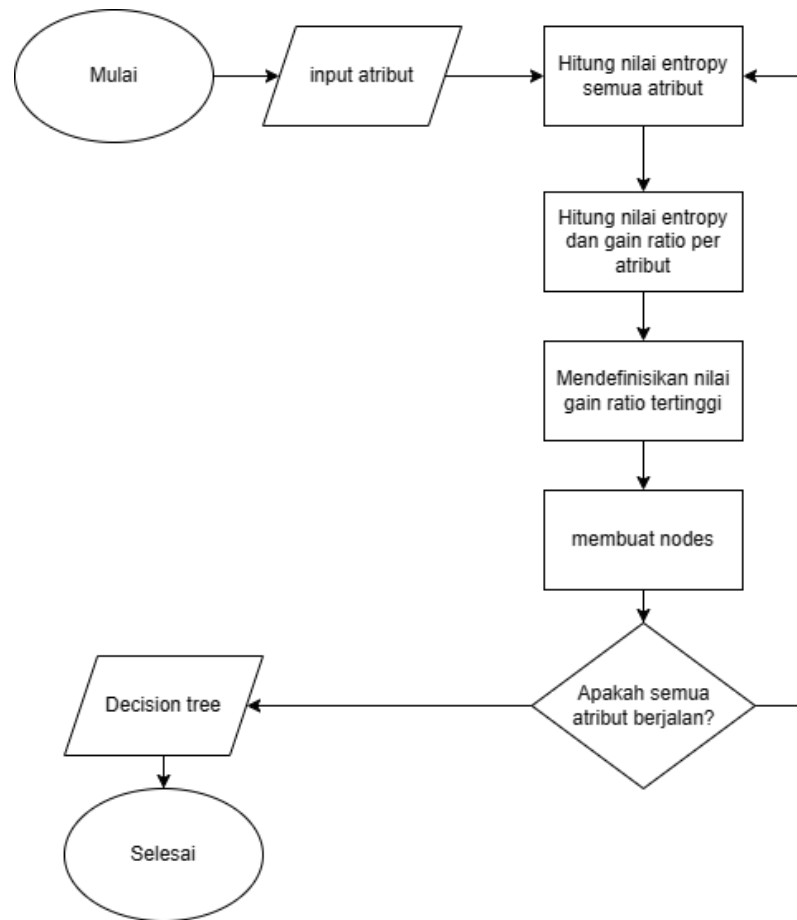
Hasil dari perhitungan TF akan digunakan untuk klasifikasi menggunakan metode *decision tree*. Berikut penjelasan pengolahan data secara lengkap:



Gambar 3.2 Flowchart Pengolahan Data

3.4 Desain Sistem

Sistem akan dimulai dengan *input* teks dari artikel berita, setelah itu teks dari artikel berita akan diolah dengan menggunakan metode *text processing* yang memiliki tahapan yaitu *case folding*, *tokenizing*, *filtering*, dan *stemming*. Setelah *text processing* selesai maka tahap selanjutnya adalah pembobotan kata yang ada di artikel berita yang telah di inputkan dengan menggunakan metode TF. Setelah pembobotan kata selesai maka tahap terakhir adalah pengklasifikasian artikel berita ke dalam topik yang sesuai dengan artikel berita tersebut, pengklasifikasian yang dilakukan menggunakan metode *decision tree*. Setelah pengklasifikasian selesai maka akan keluar *output* yang berupa hasil dari klasifikasi artikel berita yang sudah di inputkan oleh user. Hasil dari dari klasifikasi artikel berita adalah mengkategorikan artikel berita sesuai dengan topik dari artikel berita tersebut. Gambaran sistem untuk data uji akan dijelaskan pada gambar 3.3 berikut:

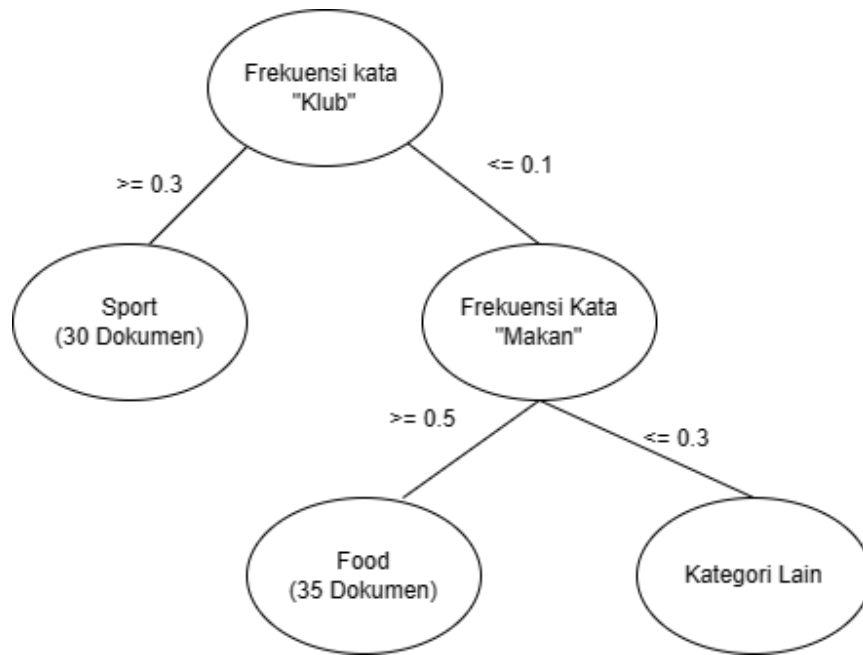


Gambar 3.3 Desain Sistem Data Uji

3.4.1. Model *Decision Tree*

Pada tahap *decision tree*, yang pertama dilakukan adalah menghitung Entropy untuk mengevaluasi kualitas potensi pemisahan dalam pohon keputusan untuk kategorisasi dokumen. Entropy dari setiap kumpulan dokumen dihitung pada setiap simpul di pohon, dan *information gain* dari setiap pemisahan kata. Perhitungan Entropy untuk mendapatkan *information gain* menggunakan rumus yang sudah dijelaskan pada bab 2. Contoh studi kasus:

- Frekuensi kata “Klub”.
- *Information Gain*: 0.5
- Total dokumen :100
- Pembagian data: 50 *Sport* dan 50 *Food*



Gambar 3.4 Decision Tree

3.5 Uji Coba Sistem

Pengujian sistem direncanakan menggunakan metode pengujian *black box*, pengujian ini bertujuan untuk menguji sistem yang dibangun berjalan sesuai fungsinya atau tidak. Pengujian *black box* menguji fungsi dari *input output* sebuah sistem. Sedangkan untuk pengujian akurasi, *recall*, *precision*, dan *f-measure* direncanakan menggunakan metode *confusion matrix* dengan menggunakan rumus yang sudah dijelaskan pada bab 2.

Skenario yang akan dilakukan untuk menguji coba sistem adalah dengan cara:

1. Menguji halaman uji dengan black box.
2. Menginputkan data uji yang sudah disiapkan ke sistem yang sudah dibuat dan sudah melatih model.
3. Memilih kategori asli dari berita tersebut untuk mencocokkan apakah hasil *predict* dari kategori model sesuai dengan kategori asli pada artikel berita tersebut.
4. Setelah semua data uji dimasukkan maka yang dilakukan selanjutnya adalah menentukan jumlah TP, TN, FN, dan FP.
5. Menghitung akurasi, *recall*, *precision*, dan *f-measure*.

BAB IV. ANALISIS DAN PERANCANGAN SISTEM

4.1 Deskripsi Sistem

Sistem klasifikasi artikel berita berdasarkan jenis topiknya menggunakan Decision Tree adalah sebuah sistem yang dirancang untuk mengelompokkan artikel berita ke dalam kategori topik yang relevan. Sistem ini menggunakan metode Decision Tree, yang merupakan metode pembelajaran mesin yang efektif dalam melakukan klasifikasi.

Sistem ini menerima input berupa teks artikel berita yang akan diklasifikasikan. Tahap pertama dalam sistem ini adalah *preprocessing* data, di mana teks artikel berita akan melalui beberapa langkah pengolahan teks. Langkah-langkah ini mencakup case folding, tokenizing, filtering, dan stemming.

Setelah *preprocessing*, proses yang dilakukan selanjutnya adalah perhitungan *TF* untuk menghitung nilai jumlah kemunculan suatu kata yang ada dalam dokumen dan memberi bobot pada suatu kata, Semakin besar jumlah kemunculan suatu term (*tf* tinggi) dalam dokumen, maka akan semakin besar pula bobotnya.

Selanjutnya, sistem akan membangun model *Decision Tree* menggunakan data latih. Data latih terdiri dari artikel berita yang telah diberi label berdasarkan jenis topiknya. Model *Decision Tree* akan belajar pola dari data latih dan membangun pohon keputusan yang dapat memprediksi jenis topik dari artikel berita yang belum diketahui.

Sistem ini akan diimplementasikan dalam bentuk aplikasi berbasis web, yang memungkinkan pengguna untuk secara interaktif memasukkan teks artikel berita dan mendapatkan hasil klasifikasi dengan cepat. Evaluasi kinerja sistem akan dilakukan dengan menggunakan metrik evaluasi seperti akurasi, presisi, *recall*, dan *F1-score*, untuk menghitung keakuratan klasifikasi yang dilakukan oleh sistem.

4.2 Analisis Kebutuhan Fungsional

Kebutuhan fungsional menggambarkan proses pada fitur-fitur tersedia yang disediakan aplikasi. Berikut adalah kebutuhan fungsional website Klasifikasi Berita:

- b. Sistem dapat mengklasifikasikan teks berita berdasarkan kategorinya.
- c. Sistem klasifikasi teks berita mengimplementasikan metode *Decision Tree*.

4.3 Analisis Kebutuhan Non-Fungsional

Kebutuhan non-fungsional adalah persyaratan atau karakteristik yang menjelaskan bagaimana sistem perangkat lunak harus beroperasi dan berinteraksi, daripada apa yang seharusnya dilakukan oleh sistem tersebut.

A. Kebutuhan Non Fungsional Sistem

- a) Sistem model dibuat dengan menggunakan bahasa pemrograman python.
- b) Sistem uji dibuat dengan menggunakan php dan *framework* Laravel.
- c) Sistem *testing* dapat dijalankan oleh beberapa *software* web browser diantaranya Microsoft Edge, dan Google Chrome.

B. Kebutuhan Non Fungsional *Usability*

Kebutuhan non-fungsional *usability* adalah hal-hal yang berhubungan dengan kemudahan penggunaan sistem klasifikasi artikel berita berdasarkan topiknya menggunakan *decision tree*. Beberapa contoh kebutuhan non-fungsional *usability* meliputi:

- a) Validasi Data: Sistem harus memiliki validasi data yang baik untuk menghindari kesalahan saat pengguna memasukkan informasi atau memilih opsi klasifikasi.
- b) Tampilan Hasil yang Mudah Dipahami: Hasil klasifikasi artikel berita ditampilkan dengan cara yang mudah dipahami
- c) Uji Pengguna (*User Testing*): Melakukan uji pengguna untuk mengidentifikasi masalah *usability*.

C. Kebutuhan Non Fungsional *Supportability*

Kebutuhan non fungsional *supportability* adalah hal-hal yang berhubungan dengan kemudahan pemeliharaan dan dukungan sistem klasifikasi artikel berita berdasarkan topiknya menggunakan *decision tree*. Beberapa contoh kebutuhan non-fungsional *supportability* meliputi:

- a) Logging dan Pencatatan Kejadian (*Logging and Auditing*): Sistem harus memiliki mekanisme logging dan pencatatan kejadian untuk memudahkan pemantauan dan analisis jika terjadi masalah.
- b) Dukungan untuk Pengujian (*Testing Support*): Sistem harus memiliki kemampuan untuk dilakukan pengujian secara otomatis untuk memastikan konsistensi dan keandalan fungsi klasifikasi.

4.4 Perancangan Sistem

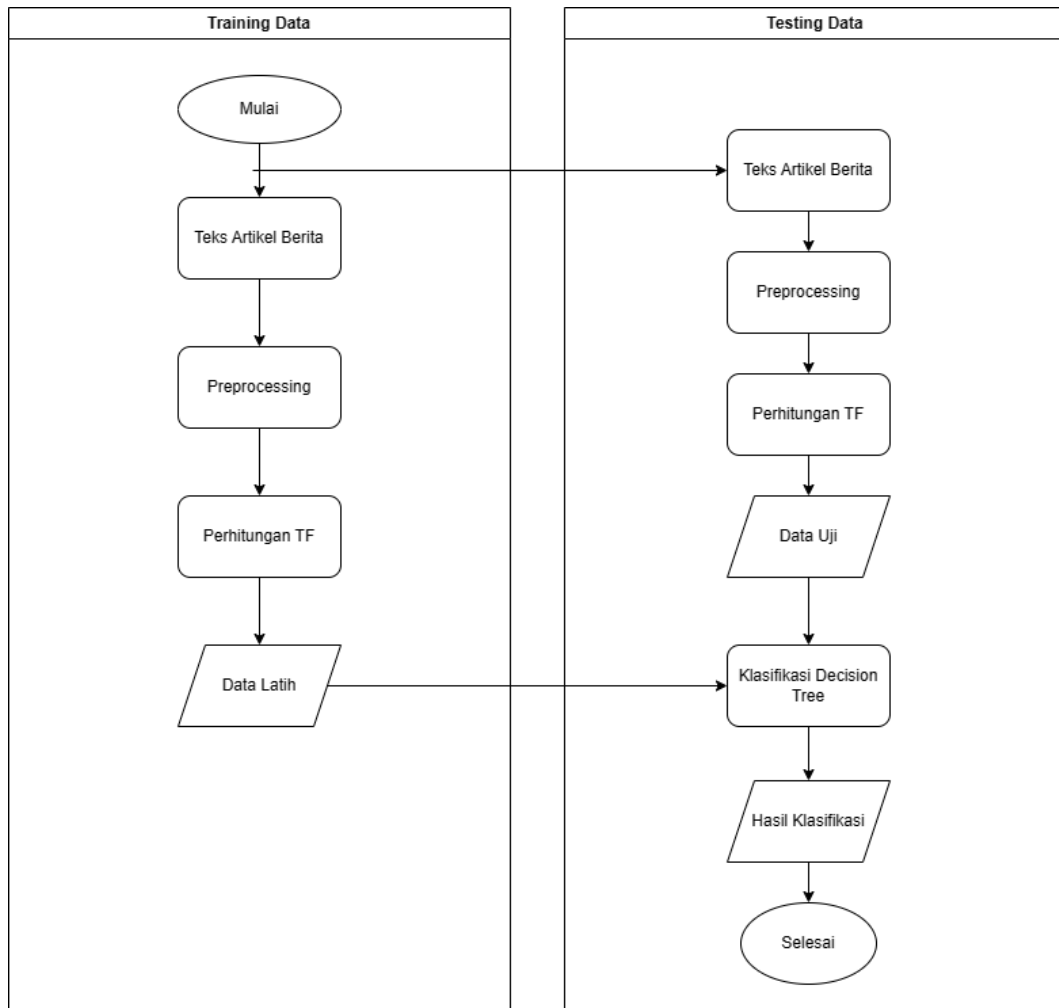
Pada sub bab ini menjelaskan rancangan sistem klasifikasi teks berita sesuai topiknya.

4.4.1. Perancangan Sistem Model

Perancangan sistem pada model klasifikasi dilakukan dengan langkah-langkah berikut:

1. Memasukkan dataset.
2. Melakukan pelabelan topik pada data.
3. Melakukan *preprocessing* pada data.
4. Menghitung *Term Frequency* pada teks.
5. Melakukan klasifikasi menggunakan *decision tree*.

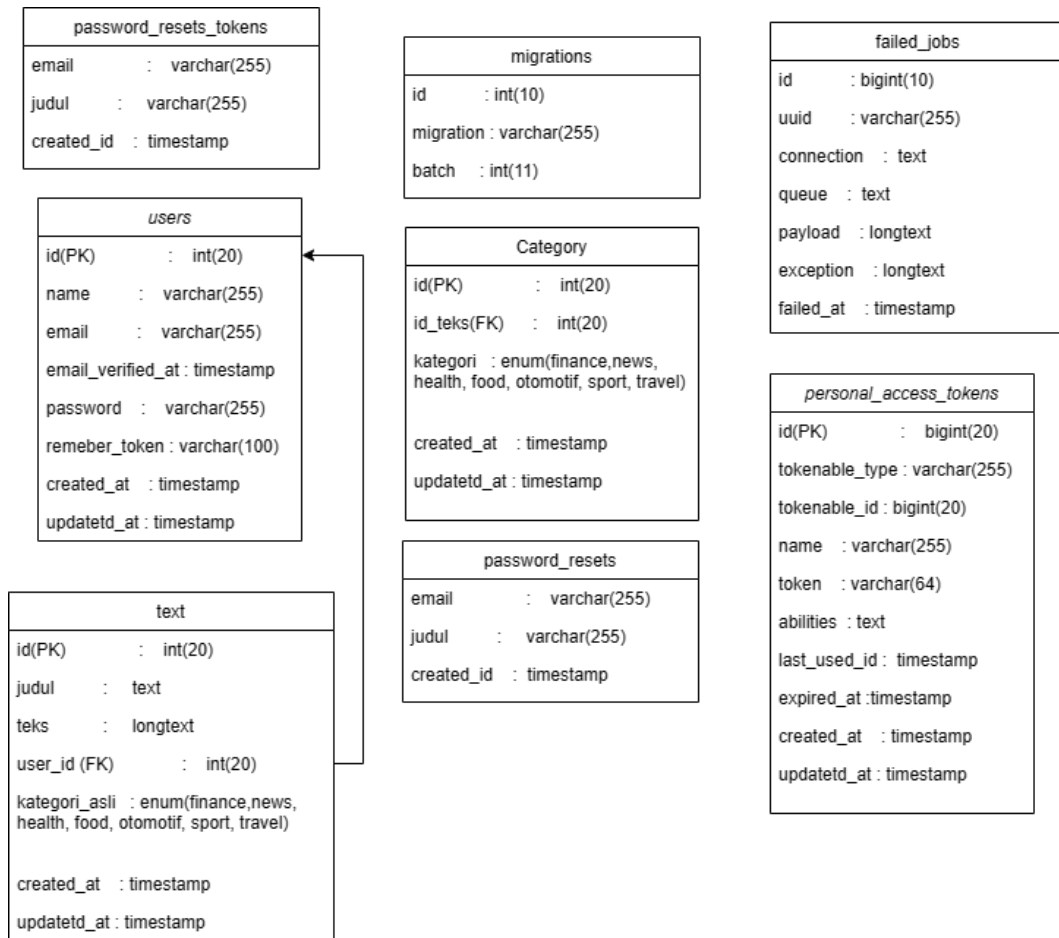
Langkah-langkah sistem pada model dijelaskan pada gambar 4.1 dibawah ini:



Gambar 4.1 Desain Sistem Model

4.4.2. Perancangan Desain Database

Desain database dibuat untuk memberi gambaran dari isi database yang akan dibuat seperti tabel, isi tabel, dan value. Kemungkinan besar isi dari database yang akan dibuat tidak jauh berbeda dengan desain database yang telah dirancang seperti gambar 4.2 dibawah ini:



Gambar 4.2 Desain Database

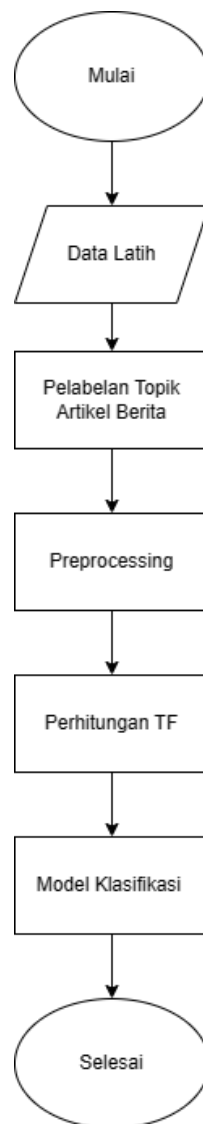
Pada desain database diatas terdapat 2 tabel utama yang akan digunakan untuk mengklasifikasikan artikel berita, yaitu tabel *text* dan *category*. Tabel *text* berfungsi untuk menyimpan data *input* dari judul, text, dan kategori asli. Sedangkan tabel *category* berfungsi untuk menyimpan kategori yang didapatkan dari hasil *predict* model. Relasi antara tabel *text* dan *category* adalah "one-to-many." Artinya, satu entri dalam tabel "*text*" dapat memiliki beberapa entri terkait dalam tabel "*category*," tetapi satu entri dalam tabel "*category*" hanya dapat terkait dengan satu entri dalam tabel "*text*." Hubungan ini diimplementasikan melalui *foreign key* "*id_text*" dalam tabel "*category*," yang mengacu pada *primary key* "*id*" dalam tabel "*text*". Ketika artikel berita dimasukkan ke dalam tabel "*text*" model klasifikasi akan memprediksi kategori untuk artikel tersebut. Hasil prediksi ini akan disimpan dalam tabel "*category*" bersama dengan referensi ke artikel yang sesuai dalam tabel

4.4.3. Tahapan Pelatihan

Tahapan pelatihan pada model klasifikasi dilakukan dengan langkah-langkah berikut:

1. Memasukkan dataset untuk data latih.
2. Melakukan pelabelan topik pada data.
3. Melakukan *preprocessing* pada data.
4. Menghitung *Term Frequency* pada teks.
5. Melakukan pelatihan data dengan model *decision tree*.

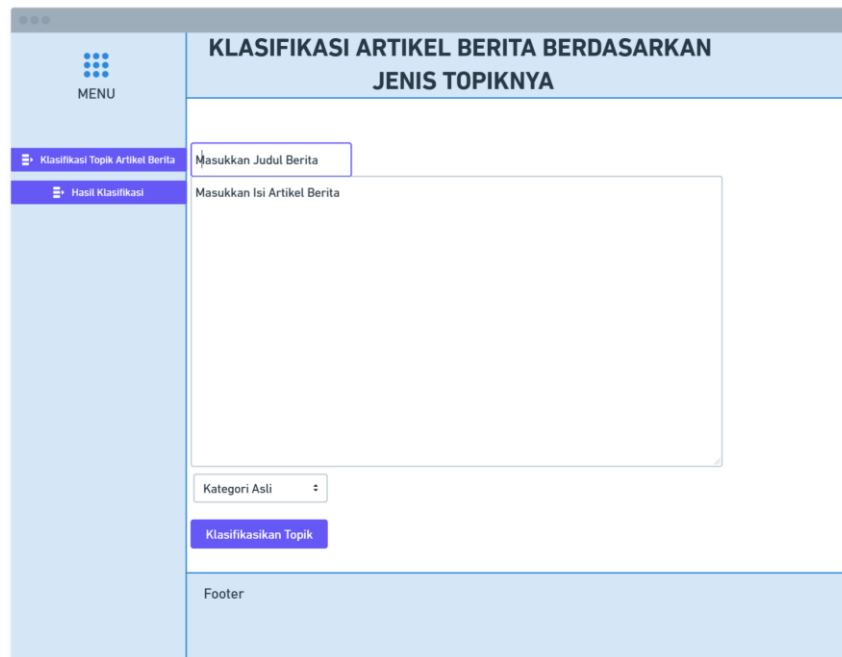
Tahapan-tahapan pelatihan dijelaskan pada gambar 4.2 dibawah ini:



Gambar 4.2 Tahapan Pelatihan

4.4.4. Perancangan Antarmuka

Gambaran umum halaman yang ditujukan kepada user untuk simulasi bagaimana cara mengklasifikasikan artikel berita berdasarkan isi dari artikel berita dijelaskan pada Gambar 4.3 dan Gambar 4.4 dibawah ini:



The screenshot shows a web application interface titled "KLASIFIKASI ARTIKEL BERITA BERDASARKAN JENIS TOPIKNYA". The interface is divided into several sections:

- Header:** A blue header bar with the title "KLASIFIKASI ARTIKEL BERITA BERDASARKAN JENIS TOPIKNYA".
- Left Sidebar:** A blue sidebar with a "MENU" icon and two menu items: "Klasifikasi Topik Artikel Berita" and "Hasil Klasifikasi".
- Main Content Area:**
 - A text input field labeled "Masukkan Judul Berita".
 - A larger text input area labeled "Masukkan Isi Artikel Berita".
 - A dropdown menu labeled "Kategori Asli".
 - A blue button labeled "Klasifikasikan Topik".
- Footer:** A blue footer bar with the text "Footer".

Gambar 4.3 Halaman Perancangan Antarmuka Klasifikasi Teks

Gambar diatas adalah gambar perancangan antarmuka untuk memasukkan input teks, judul, dan kategori asli untuk data testing



The screenshot shows a web application interface with a light blue theme. At the top, there is a header with the title "KLASIFIKASI ARTIKEL BERITA BERDASARKAN JENIS TOPIKNYA". Below the header, there is a search bar with the placeholder text "Search". The main content area contains a table with 5 rows of classified news items. The table has the following columns: "No", "Judul Artikel Berita", "Isi Artikel Berita", "Kategori Asli", and "Topik Hasil Klasifikasi".

No	Judul Artikel Berita	Isi Artikel Berita	Kategori Asli	Topik Hasil Klasifikasi
1	APU AMD Renoir Bakal Punya 8 Core?	Bocoran prosesor...	Internet	Internet
2	Hyundai Tucson Tampil Makin Segar	PT Hyundai Mobil Ind...	Otomotif	Otomotif
3	Mike Tyson Is Back! Kapan Naik Ring Lagi?	Mantan juara dunia...	Sport	Sport
4	Kastil 'Neraka' di Republik Ceko	Kastil Houska di Rep...	Travel	Travel
5	Resep Pembaca : Bala-Bala Ebi	Camilan rumahan ini....	Food	Food

At the bottom of the page, there is a footer area labeled "Footer".

Gambar 4.4 Halaman Perancangan Antarmuka Riwayat Klasifikasi Teks

Gambar diatas adalah gambar perancangan antarmuka untuk menampilkan riwayat hasil klasifikasi teks. Di halaman tersebut menampilkan *output* dari klasifikasi teks beserta kategori dari hasil klasifikasi teks.

BAB V. IMPLEMENTASI DAN PENGUJIAN

5.1 Implementasi Antarmuka

Halaman antarmuka yang sudah diimplementasikan tidak jauh berbeda dengan halaman yang sudah dirancang. Berikut adalah implementasi antarmuka dari aplikasi klasifikasi berita berdasarkan topik tertera pada Gambar 5.1 dan Gambar 5.2:

The screenshot shows a web application interface for classifying news by topic. The page title is 'KLASIFIKASI BERITA BERDASARKAN TOPIK'. On the left, there is a dark sidebar menu with the text 'testing' and options: 'Riwayat Hasil Klasifikasi Berita', 'Klasifikasi Kategori Berita', and 'Logout'. The main content area is titled 'Dashboard' and contains three input fields: 'Judul' (with a placeholder 'Enter ...'), 'Teks Berita' (with a placeholder 'Enter ...'), and 'Kategori Asli' (a dropdown menu currently showing 'finance'). Below these fields is a blue button labeled 'Klasifikasikan'. At the bottom of the page, there is a copyright notice 'Copyright © 2023 Klasifikasi Berita Berdasarkan Kategori.' and the name 'Rezaldy Rafli'.

Gambar 5.1 Halaman Input Teks

Pada gambar 5.1 ada halaman input teks yang berfungsi untuk memasukkan teks berita untuk diklasifikasi kan oleh model. Dan akan ditampilkan pada halaman Riwayat hasil kategori seperti pada gambar 5.2 dibawah ini:

The screenshot shows the 'Data Klasifikasi' page of the application. It features a table with columns: 'Nomor', 'Judul', 'Teks Berita', 'Kategori Asli', 'Kategori', and 'Action'. The table contains three rows of data. The first row has '1' in the 'Nomor' column, 'IHSG Berpotensi Menguat Lagi Ini' in the 'Judul' column, and a paragraph of Indonesian text in the 'Teks Berita' column. The 'Kategori Asli' and 'Kategori' columns both show 'finance', and the 'Action' column has a red 'Delete' button. The second row has '2' in the 'Nomor' column, 'Usai Brexit, Bahili Cari Investasi ke Inggris Bulan Depan' in the 'Judul' column, and another paragraph of Indonesian text in the 'Teks Berita' column. The 'Kategori Asli' and 'Kategori' columns both show 'finance', and the 'Action' column has a red 'Delete' button. The third row has '3' in the 'Nomor' column, 'Jokowi Janji Permudah Investasi Data Center Microsoft' in the 'Judul' column, and a paragraph of Indonesian text in the 'Teks Berita' column. The 'Kategori Asli' and 'Kategori' columns both show 'finance', and the 'Action' column has a red 'Delete' button. The page title is 'KLASIFIKASI BERITA BERDASARKAN TOPIK' and the sidebar menu is the same as in Gambar 5.1.

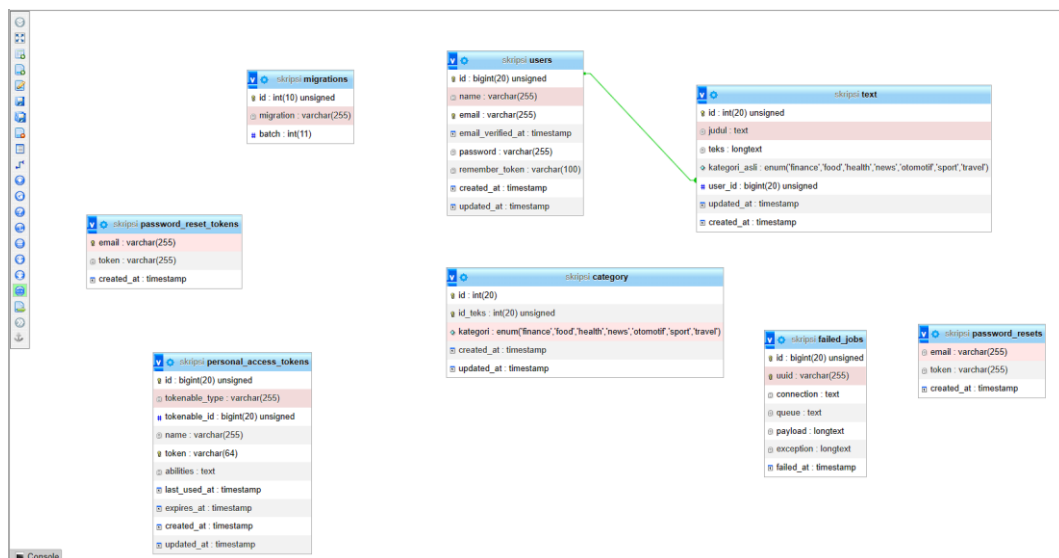
Nomor	Judul	Teks Berita	Kategori Asli	Kategori	Action
1	IHSG Berpotensi Menguat Lagi Ini	Pada perdagangan kemarin (18/2) IHSG ditutup menguat tipis 0.33% ke level 5.896.92. Sembilan dari sepuluh indeks sektoral ditutup dalam teritori positif, dimana sektor properti dan industri dasar memimpin penguatan masing-masing sebesar 1.17% dan 1.86%. Adapun saham yang menjadi penggerak indeks di antaranya BBRI, BBCA, BHRJ, PLIN, ACES. Pelaku pasar asing membukukan aksi jual bersih (netted) senilai Rp 260 miliar. Nilai tukar rupiah terdepresiasi sebesar 0,01% ke level Rp 13.694. Pada perdagangan semalam (18/2) bursa saham Wall Street ditutup mixed dengan kecenderungan melemah, dimana Dow Jones dan S&P 500 masing-masing turun sebesar 0,56% dan 0,29% sedangkan Nasdaq naik tipis 0,02%. ADVERTISEMENT SCROLL TO RESUME CONTENT Pelemahan yang terjadi pada dua indeks utama Wall Street terbebani oleh Apple Inc yang disebabkan adanya peringatan dari perusahaan tersebut yang menilai kinerja penjualan pada triwulan nanti akan meleset dari target sebelumnya, hal ini terjadi seiring dengan menurunnya permintaan dari China serta produksi dari iPhone yang sedikit melambat. Sementara itu pasar global masih mencermati dampak yang ditimbulkan dari virus Corona terhadap pertumbuhan ekonomi di China. IHSG ditutup menguat sebesar 0,33% ke level 5.896. Diperkirakan IHSG berpotensi menguat dengan pergerakan di kisaran 5.867-5.925. Menebak Arah IHSG Jelang Potensi Resesi	finance	finance	Delete
2	Usai Brexit, Bahili Cari Investasi ke Inggris Bulan Depan	Inggris resmi meninggalkan Uni Eropa (UE). Usai Brexit atau Britain Exit, Inggris masih harus melalui masa transisi hingga Desember mendatang. Artinya, negara Inggris masih akan menggunakan aturan UE dalam perdagangannya sampai masa transisi itu berakhir. Setelahnya, Britania Raya mungkin bakal menerapkan aturan baru dalam berhubungan dengan mitra-mitranya. Memanfaatkan momentum tersebut, Kepala Badan Koordinasi Penanaman Modal (BKPM) Bahili Lahadita mengungkapkan bahwa dirinya bulan depan akan langsung terbang ke Inggris demi menarik lebih banyak investasi dari sana. ADVERTISEMENT SCROLL TO RESUME CONTENT "Bulan depan bila ke Inggris," ujar Bahili ditemui di Gedung Balaiung Universitas Indonesia, Depok, Sabtu (1/2/2020). Selama ini, menurut Bahili, realisasi investasi terbesar UE ke Indonesia baru berasal dari Belanda. Sedangkan, porsi investasi Britania Raya hanya sebahagian kecilnya saja. "Gini pada tahun 2019, realisasi investasi Eropa yang paling besar itu Belanda, karena dari semua negara Eropa, pertumbuhan ekonomi yang agak sedikit baik itu ya Belanda, sedangkan Inggris dan Jerman itu menurun dari target pertumbuhan ekonomi di 2019," paparnya. Untuk itu, Bahili menyebut dirinya belum bisa menyampaikan seberapa besar potensi investasi yang bisa ditarik Indonesia dari Inggris pasca-Brexit ini. "Datanya seberapa besar dampak Brexit, akan kami sampaikan di Februari akhir atau bulan Maret. Sekarang lagi kita sikur," pungkasnya. Data BKPM menunjukkan, hanya dua negara Eropa yang masuk 10 besar investor di Indonesia yakni Belanda dan Inggris. Realisasi investasi Eropa pada kuartal ke-3 2019 sebesar US\$ 2,768 miliar. Meningkat dibandingkan periode sebelumnya 2018 sebesar US\$ 2,321 miliar. Dengan jumlah proyek Eropa dari 2015 sampai kuartal ke-3 2019 yang mencapai 15.770 proyek. Sebagian besar atau 61% investasi Eropa masuk ke Pulau Jawa, disusul Sumatera 19%, Bali dan Nusa Tenggara 11%, Sulawesi 4%, Papua 1%, dan Maluku 0,4%. Kelakar Bahili soal Pengusaha Hebat Sasati Aturan dan Taklukkan Pejabat	finance	finance	Delete
3	Jokowi Janji Permudah Investasi Data Center Microsoft	Microsoft memiliki niat untuk berinvestasi data center di Indonesia. Presiden Joko Widodo (Jokowi) berjanji akan memuluskan jalan Microsoft untuk investasi di RI. Jokowi mengaku, pagi tadi sudah berdiskusi dengan CEO Microsoft Satya Nadella. Perusahaan itu telah mengungkapkan ketertarikannya berinvestasi data center di Indonesia. "Tadi pagi setelah berdiskusi dengan CEO Microsoft Mr Satya. Intinya mereka ingin investasi di data center", tuturnya di Ritz Carlton, Pacific Place, Jakarta, Kamis (17/2/2020). ADVERTISEMENT SCROLL TO RESUME CONTENT Namun, niatan ini seoptimis masih terkendala terkait dasar peraturan. Pemerintah saat ini masih menaikan	finance	finance	Delete

Gambar 5.2 Halaman Riwayat Hasil Kategori

Pada gambar 5.2 adalah halaman yang menampilkan judul, teks berita, kategori asli artikel berita, dan kategori hasil *predict* dari model yang telah dilatih sebelumnya.

5.2 Implementasi Database

Implementasi dari database tidak jauh berbeda dengan desain database yang telah dibuat. Berikut adalah hasil dari implementasi database yang telah dibuat, tertera pada gambar 5.3 dibawah ini:



Gambar 5.3 Implementasi Database

Tabel-tabel tersebut diimplementasikan ke dalam database yang telah dibuat pada *MySQL*. Pada database yang telah dibuat, terdapat tabel tambahan dari *Laravel* seperti tabel *migrations*, *failed job*, dan *password_resets*. Database diatas memiliki relasi yaitu pada tabel *text* dan tabel *user* untuk mengambil id dari tabel *user* ke tabel *text*, yang berfungsi mencocokkan data inputan sesuai pengguna dengan teks yang telah dimasukkan ke sistem. "*text*" menggunakan *foreign key*.

5.3 Implementasi Metode

Berikut adalah penjelasan dari implementasi metode *decision tree* yang telah diterapkan sebagai berikut:

```
# Split data into features and target
X = data['text_stemmed']
y = data['Category']
```

```

# Convert text data to numerical features (Term Frequency)
vectorizer = CountVectorizer()
X = vectorizer.fit_transform(X)

# Split data into training and testing sets
X_train, X_test, y_train, y_test = train_test_split(X, y,
test_size=0.2, random_state=0)

# Train decision tree classifier
clf = DecisionTreeClassifier()
clf.fit(X_train, y_train)

```

Gambar 5.4 Implementasi Metode

Menurut gambar 5.4 diatas dapat diketahui bahwa, “X = data['text_stemmed]” Pada baris ini, atribut 'text_stemmed' diambil dari objek data yang berisi teks yang telah melalui proses stemming. Atribut (X) akan menjadi fitur input untuk proses klasifikasi. Sedangkan untuk “y = data['Category]” Pada baris kode ini, atribut 'Category' diambil dari objek data. Atribut ini berisi label atau kategori yang sesuai dengan teks pada atribut 'text_stemmed'. Atribut (y) akan menjadi output yang akan diprediksi oleh model. Untuk baris kode “vectorizer = CountVectorizer()”, CountVectorizer berfungsi mengubah teks menjadi representasi vektor berdasarkan frekuensi kemunculan kata-kata. Untuk baris kode “X = vectorizer.fit_transform(X)”, berfungsi untuk mengubah teks menjadi representasi vector dengan menggunakan metode “fit_transform”. Pada baris kode “clf = DecisionTreeClassifier()” berfungsi untuk memanggil model *decision tree* dari *library* sklearn yang akan digunakan untuk melakukan klasifikasi. Sedangkan baris kode “clf.fit(X_train, y_train)”, memiliki fungsi untuk melatih model *decision tree* dengan cara menggunakan *method* “fit” dengan menggunakan data latih yang ada pada “X_train, y_train). Untuk parameter *gain* yang digunakan adalah parameter *default* yang telah ditentukan oleh *library* dari sklearn.

5.4 Pengujian Sistem

5.4.1. Proses Pengujian Data Testing

Tabel 5.1 merupakan tabel *Confusion matrix* yang merupakan tabel yang berisikan data testing sebanyak 175 artikel. Dalam pengujian ini sistem diuji untuk mengetahui ketepatan klasifikasi artikel berita tersebut.

Tabel 5.1 Tabel *Confusion Matrix*

	Finance	Food	Health	News	Otomotif	Sport	Travel	TOTAL
Finance	17	0	0	6	1	0	1	25
Food	0	21	0	1	0	0	3	25
Health	5	1	13	4	0	1	1	25
News	1	0	0	22	2	0	0	25
Otomotif	1	0	1	5	16	1	1	25
Sport	1	1	0	5	1	16	1	25
Travel	7	2	2	2	1	0	11	25
								175

Setelah itu proses yang dilakukan adalah menghitung jumlah *FP*, *TP*, *FN*, dan *TN* berdasarkan nilai yang berasal dari data yang ada pada tabel 5.1 *confusion matrix* diatas. Berikut dibawah ini adalah tabel-tabel dari jumlah *FP*, *TP*, *FN*, dan *TN* perkategoriannya:

Tabel 5.2 Tabel *FP*, *TP*, *FN*, dan *TN* Kategori *Finance*

<i>Finance</i>	<i>FP</i>	15
	<i>TP</i>	17
	<i>FN</i>	8
	<i>TN</i>	135

Pada tabel 5.2 diatas, kategori *finance* memiliki jumlah data uji benar atau *TP* sebanyak 17 data dari 25 data *inputan*, *FP* sebanyak 15 data, *TN* sebanyak 135 dari total 175 data seluruh kategori, *FN* sebanyak 8 data. Dari empat metrik di atas, peneliti dapat menghitung beberapa metrik evaluasi tambahan yang penting untuk mengukur performa model, seperti akurasi, presisi, recall, dan F1-score.

Tabel 5.3 Tabel *FP*, *TP*, *FN*, dan *TN* Kategori *Food*

<i>Food</i>	<i>FP</i>	32
	<i>TP</i>	21

	<i>FN</i>	4
	<i>TN</i>	146

Pada tabel 5.3 diatas, kategori *food* memiliki jumlah data uji benar atau *TP* sebanyak 21 data dari 25 data *inputan*, *FP* sebanyak 32 data, *TN* sebanyak 146 dari 175 data, *FN* sebanyak 4 data.

Tabel 5.4 Tabel *FP*, *TP*, *FN*, dan *TN* Kategori *Health*

<i>Health</i>	<i>FP</i>	27
	<i>TP</i>	13
	<i>FN</i>	25
	<i>TN</i>	147

Pada tabel 5.4 diatas, kategori *health* memiliki jumlah data uji benar atau *TP* sebanyak 13 data dari 25 data *inputan*, *FP* sebanyak 27 data, *TN* sebanyak 147 dari 175 data, *FN* sebanyak 25 data.

Tabel 5.5 Tabel *FP*, *TP*, *FN*, dan *TN* Kategori *News*

<i>News</i>	<i>FP</i>	31
	<i>TP</i>	22
	<i>FN</i>	3
	<i>TN</i>	127

Pada tabel 5.5 diatas, kategori *news* memiliki jumlah data uji benar atau *TP* sebanyak 22 data dari 25 data *inputan*, *FP* sebanyak 31 data, *TN* sebanyak 127 dari 175 data, *FN* sebanyak 3 data.

Tabel 5.6 Tabel *FP*, *TP*, *FN*, dan *TN* Kategori *Otomotif*

<i>Otomotif</i>	<i>FP</i>	31
	<i>TP</i>	16
	<i>FN</i>	9
	<i>TN</i>	145

Pada tabel 5.6 diatas, kategori *otomotif* memiliki jumlah data uji benar atau *TP* sebanyak 16 data dari 25 data *inputan*, *FP* sebanyak 31 data, *TN* sebanyak 145 dari 175 data, *FN* sebanyak 9 data.

Tabel 5.7 Tabel *FP*, *TP*, *FN*, dan *TN* Kategori *Sport*

<i>Sport</i>	<i>FP</i>	31
	<i>TP</i>	16
	<i>FN</i>	9
	<i>TN</i>	132

Pada tabel 5.7 diatas, kategori *sport* memiliki jumlah data uji benar atau *TP* sebanyak 16 data dari 25 data *inputan*, *FP* sebanyak 31 data, *TN* sebanyak 132 dari 175 data, *FN* sebanyak 9 data.

Tabel 5.8 Tabel *FP*, *TP*, *FN*, dan *TN* Kategori *Travel*

<i>Travel</i>	<i>FP</i>	25
	<i>TP</i>	11
	<i>FN</i>	14
	<i>TN</i>	143

Pada tabel 5.8 diatas, kategori *travel* memiliki jumlah data uji benar atau *TP* sebanyak 11 data dari 25 data *inputan*, *FP* sebanyak 25 data, *TN* sebanyak 143 dari 175 data, *FN* sebanyak 14 data.

Setelah mendapatkan nilai metrik *FP*, *TP*, *FN*, dan *TN* per kategorinya proses yang dilakukan selanjutnya adalah menghitung akurasi, presisi, *recall*, dan *f1-score*.

Keterangan:

Jumlah data benar: $17 + 21 + 13 + 22 + 16 + 16 + 11 = 116$

Total Data: 175

Hasil Akurasi:

$$\text{Finance: } \frac{17+135}{17+15+135+8} = 0.86 = 86\%$$

$$\text{Food: } \frac{21+146}{21+32+146+4} = 0.82 = 82\%$$

$$\text{Health: } \frac{13+147}{13+27+147+25} = 0.75 = 75\%$$

$$\text{News: } \frac{22+127}{22+31+127+3} = 0.81 = 81\%$$

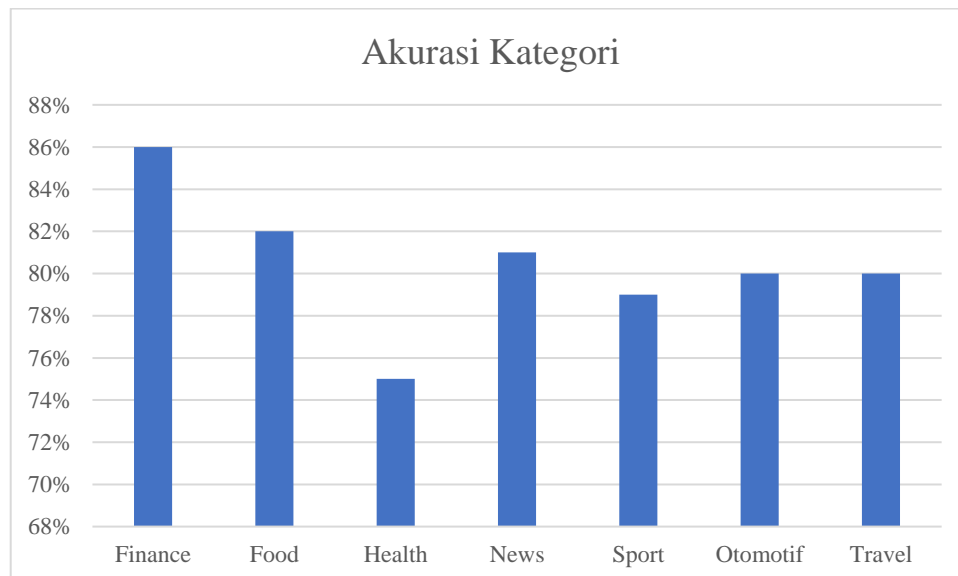
$$\text{Otomotif: } \frac{16+145}{16+31+145+9} = 0.80 = 80\%$$

$$\text{Sport: } \frac{16+132}{16+31+132+9} = 0.78 = 78\%$$

$$\text{Travel: } \frac{11+143}{11+25+143+14} = 0.79 = 79\%$$

Rata-rata akurasi = 80%

Untuk visualisasi grafik akurasi pada hasil uji ada pada gambar 5.5 dibawah ini:



Gambar 5.5 Akurasi Per-Kategori

Pada perhitungan *accuracy* diatas, kategori *finance* mendapatkan nilai paling tinggi yaitu sebesar 86%. Sedangkan untuk nilai terendah ada pada kategori *health* dengan nilai 75%.

Hasil perhitungan untuk *recall* untuk setiap kategori berita yang telah di uji pada penelitian ini adalah:

$$\text{Finance: } \frac{17}{17+8} = 0.11 = 11\%$$

$$\text{Food: } \frac{21}{21+4} = 0.13 = 13\%$$

$$\text{Health: } \frac{13}{13+25} = 0.08 = 8\%$$

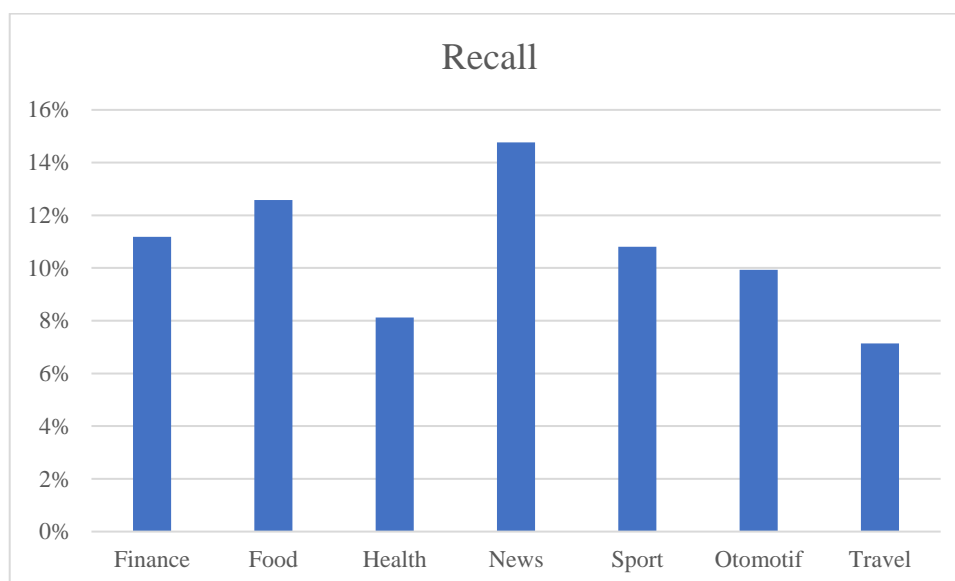
$$\text{News: } \frac{22}{22+3} = 0.15 = 15\%$$

$$\text{Otomotif: } \frac{16}{16+9} = 0.10 = 10\%$$

$$\text{Sport: } \frac{16}{16+9} = 0.11 = 11\%$$

$$\text{Travel: } \frac{11}{11+14} = 0.07 = 7\%$$

Untuk visualisasi grafik akurasi pada hasil uji ada pada gambar 5.6 dibawah ini:



Gambar 5.6 Recall Per-Kategori

Pada perhitungan *recall* diatas, kategori *news* mendapatkan nilai paling tinggi yaitu sebesar 15%. Sedangkan untuk nilai terendah ada pada kategori *travel* dengan nilai 7%.

Hasil perhitungan untuk *precision* untuk setiap kategori berita yang telah di uji pada penelitian ini adalah:

$$\text{Finance: } \frac{17}{17+15} = 0.53 = 53\%$$

$$\text{Food: } \frac{21}{21+32} = 0.39 = 39\%$$

$$\text{Health: } \frac{13}{13+27} = 0.32 = 32\%$$

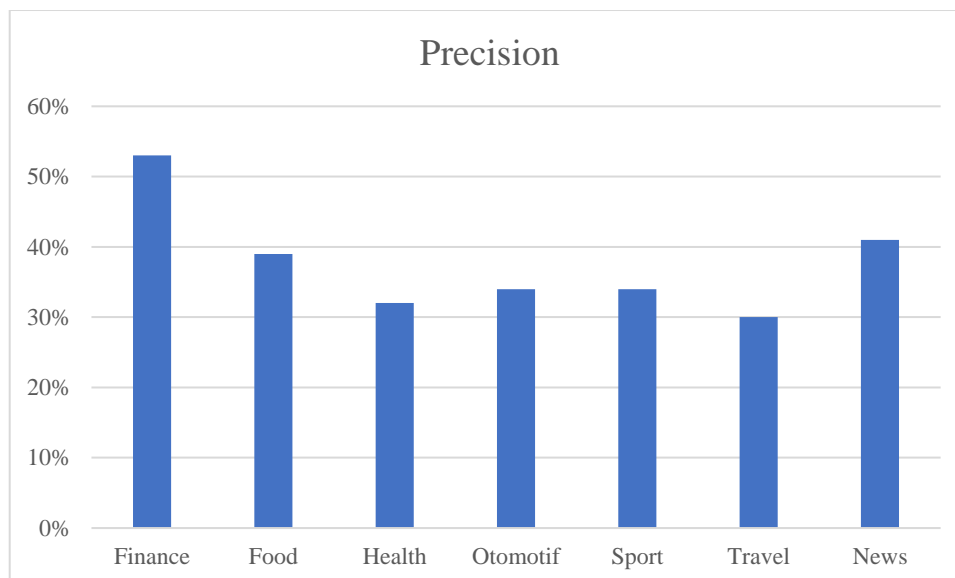
$$\text{News: } \frac{22}{22+31} = 0.41 = 41\%$$

$$\text{Otomotif: } \frac{16}{16+31} = 0.34 = 34\%$$

$$\text{Sport: } \frac{16}{16+31} = 0.34 = 34\%$$

$$\text{Travel: } \frac{11}{11+25} = 0.30 = 30\%$$

Untuk visualisasi grafik akurasi pada hasil uji ada pada gambar 5.7 dibawah ini:



Gambar 5.7 Precision Per-Kategori

Pada perhitungan *precision* diatas, kategori *finance* mendapatkan nilai paling tinggi yaitu sebesar 53%. Sedangkan untuk nilai terendah ada pada kategori *travel* dengan nilai 30%.

Hasil perhitungan untuk *f1-score* untuk setiap kategori berita yang telah di uji pada penelitian ini adalah:

$$\text{Finance: } 2 * \frac{0.53*0.11}{0.53+0.11} = 0.18 = 18\%$$

$$\text{Food: } 2 * \frac{0.39*0.12}{0.39+0.12} = 0.19 = 19\%$$

$$\text{Health: } 2 * \frac{0.32*0.08}{0.32+0.08} = 0.32 = 32\%$$

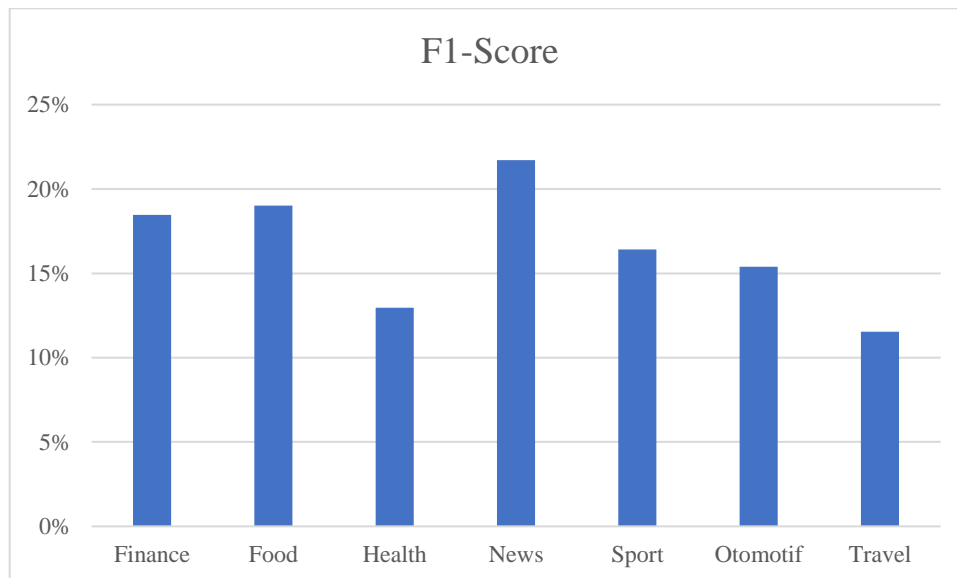
$$\text{News: } 2 * \frac{0.41*0.14}{0.41+0.14} = 0.21 = 21\%$$

$$\text{Otomotif: } 2 * \frac{0.34*0.09}{0.34+0.09} = 0.15 = 15\%$$

$$\text{Sport: } 2 * \frac{0.34*0.10}{0.34+0.10} = 0.16 = 16\%$$

$$\text{Travel: } 2 * \frac{0.3*0.07}{0.3+0.07} = 0.11 = 11\%$$

Untuk visualisasi grafik akurasi pada hasil uji ada pada gambar 5.8 dibawah ini:



Gambar 5.8 F1-Score Per-Kategori

Pada perhitungan *f1-score* diatas, kategori *news* mendapatkan nilai paling tinggi yaitu sebesar 21%. Sedangkan untuk nilai terendah ada pada kategori *travel* dengan nilai 11%.

5.4.2. Pengujian Model

Untuk Pengujian data *training* menggunakan *library* dari sklearn. Hasil dari Akurasi, *recall*, *precision*, dan *f-measure* dijelaskan pada gambar 5.9 dibawah ini:

	precision	recall	f1-score	support
finance	0.71	0.72	0.72	2249
food	0.84	0.84	0.84	705
health	0.72	0.69	0.70	805
news	0.85	0.87	0.86	5139
otomotif	0.69	0.62	0.65	714
sport	0.76	0.68	0.72	370
travel	0.66	0.63	0.65	980
accuracy			0.78	10962
macro avg	0.75	0.72	0.73	10962
weighted avg	0.78	0.78	0.78	10962

Gambar 5.9 Hasil Pengujian Model

Akurasi yang didapatkan oleh model yang sudah di latih lalu di uji adalah sebesar 78%. Sedangkan untuk hasil perhitungan dari recall, precision, dan f-measure dari setiap kategori yaitu pada perhitungan precision kategori news mendapatkan nilai paling tinggi sebesar 85% dan nilai precision paling rendah adalah travel dengan 66%. Untuk kategori yang mendapat nilai recall tertinggi adalah news. Sedangkan nilai tertinggi pada f-measure adalah kategori news.

Kode yang digunakan untuk menghasilkan nilai akurasi, presisi, *recall*, dan *f1-score* adalah sebagai berikut:

```
accuracy = accuracy_score(y_test, y_pred)
print('Accuracy:', accuracy)

report = classification_report(y_test, y_pred)
print(report)
```

Gambar 5.10 Gambar Kode Output Akurasi

Menurut gambar 5.10 diatas dapat diketahui bahwa kode memiliki tugasnya masing-masing. Seperti kode "accuracy_score(y_test, y_pred):" berfungsi untuk menghitung tingkat akurasi model. "y_test" adalah array yang berisi label sebenarnya dari data uji, sedangkan "y_pred" adalah array yang berisi prediksi model untuk data uji. Tingkat akurasi adalah jumlah prediksi benar dibagi dengan jumlah total prediksi. Hasilnya disimpan dalam variabel accuracy, "print('Accuracy:', accuracy):" berfungsi untuk mencetak nilai akurasi yang dihasilkan oleh model. "classification_report(y_test, y_pred)" berfungsi untuk membuat laporan klasifikasi yang berisi beberapa metrik evaluasi yang lebih rinci. y_test adalah array yang berisi label sebenarnya dari data uji, sedangkan y_pred adalah array yang berisi prediksi model untuk data uji. "print(report)" berfungsi mencetak laporan klasifikasi yang berisi informasi seperti akurasi, presisi, *recall*, dan *f1-score* untuk setiap kelas yang ada dalam data uji. *Report* ini sangat berguna untuk memahami sejauh mana model dapat mengklasifikasikan dengan benar setiap kelas.

5.4.3. Pengujian Fungsional Sistem

Pada tahap pengujian sistem blackbox ini akan menggunakan metode blackbox. Metode digunakan untuk mendemostrasikan jalanya aplikasi serta

menemukan kesalahan aplikasi yang sedang diuji. Metode ini akan menghasilkan kesimpulan berupa hasil pengujian apakah inputan yang dijalankan sesuai dengan outputan yang berjalan

Tabel 5.9 Tabel Pengujian Halaman Uji

No.	Halaman	Skenario Pengujian	Hasil yang diharapkan	Hasil Pengujian	Status
1	Halaman Pengujian	- User klik tombol “Klasifikasi Kategori Berita”	- Menuju ke halaman klasifikasi berita	-Sesuai pengujian	Berhasil
		-User memasukkan judul, teks berita dan memilih kategori asli dari berita tersebut, lalu klik tombol klasifikasikan.	- Menuju ke halaman Riwayat hasil klasifikasi yang menampilkan hasil kategori dari data yang sudah dimasukkan dan di klasifikasikan.	-Sesuai pengujian	Berhasil

BAB VI. HASIL DAN PEMBAHASAN

6.1 Hasil dan Pembahasan

Berdasarkan hasil dari data uji yang telah di klasifikasi kan, dari total 175 data artikel berita 25 data setiap kategorinya, akurasi yang didapatkan berbeda dari akurasi yang didapatkan dari hasil latih model. Akurasi yang didapat pada data uji sebesar 80% dibanding hasil akurasi dari model yaitu sebesar 78%. Kategori yang memiliki akurasi paling tinggi adalah kategori *finance* dengan nilai akurasi sebesar 86%. Data uji yang digunakan beragam yaitu ada data yang berasal dari portal berita lain yang berbeda dari portal berita yang digunakan di dataset yang digunakan sebagai data latih model.

Berikut adalah contoh hasil dari data uji yang sudah di klasifikasikan ada pada gambar 6.1 dibawah ini:

Nomor	Judul	Teks Berita	Kategori Asli	Kategori	Action
101	Cara Pemda di Aceh Tangani Corona: Potong SPPD Millaran Rupiah, Sumbang Gaji	Sejumlah daerah di Aceh mulai mengalokasikan dana hingga miliaran rupiah untuk penanganan pandemi Corona. Dana itu ada yang bersumber dari mengalihkan anggaran perjalanan dinas hingga sumbangan wali kota. Di Aceh Utara, misalnya, Pemerintah Kabupaten (Pemkab) dan DPR Kabupaten sepakat mengalihkan anggaran perjalanan dinas sebesar Rp 8,7 miliar untuk menangan Corona. Dana itu rencananya akan dipakai untuk tambahan pengadaan peralatan kesehatan, termasuk alat pelindung diri (APD) tenaga medis dan serta kebutuhan lain. "Dalam rapat disepakati untuk memotong anggaran SPPD (Surat Perintah Perjalanan Dinas), baik SPPD bupati, wakil bupati, SPPD pimpinan dan anggota Dewan, juga SPPD dinas, baik perjalanan luar daerah maupun dalam daerah," kata Juru Bicara Gugus Tugas Percepatan Penanganan Wabah COVID-19 Kabupaten Aceh Utara, Andree Prayuda, kepada wartawan. Andre mengatakan, bila melihat perkembangan situasi dan wabah saat ini, pengadaan peralatan kesehatan sangat penting dan khususnya kebutuhan untuk paramedis. Hal itu baik yang di rumah sakit maupun di Puskesmas di Aceh Utara. Sementara itu Bupati Aceh Utara, Muhammad Thalib mengingatkan kepada seluruh masyarakat untuk tetap waspada terhadap penularan wabah COVID-19. Dia mengimbau masyarakat untuk berada di rumah dan mengikuti imbauan pemerintah. "Kami sangat mengharapkan masyarakat mematuhi setiap imbauan yang dikeluarkan oleh pemerintah, khususnya imbauan untuk menetap di rumah, tidak bepergian atau tidak berkegiatan, kecuali untuk keperluan sangat mendesak," jelas Cek Mad, sapaan akrab Bupati Aceh Utara saat dihubungi terpisah. Anggota DPR Siap Dipotong Gajinya untuk Atasi Corona:	news	news	Delete
102	Bakal Lahir Mobil 3.000 Tenaga Kuda, Minum Bahan Bakar Tak Biasa	Era mesin kendaraan konvensional masih menarik itu dilirik. Untuk menarik konsumen berkecek tebal, pabrik otomotif terus melahirkan mobil yang mengusung mesin buas. Terbaru sebuah perusahaan otomotif asal Yunani, Spyros Panopoulos Automotive mengatakan tak cuma menjanjikan desain mobil yang ciamik, tenaganya pun sanggup bikin jantung berdebar. Horsepower besar memang menjadi obsesi banyak penggemar mobil dunia, karena itulah tidak salah bila banyak produsen mobil berlomba membuat mobil super dengan kekuatan besar, sebut saja Aston Martin, Bugatti, Konigsegg, Hennessey Venom, hingga Rimac. Dikutip dari Carbuzz, Spyros Panopoulos Automotive hendak melahirkan mobil dengan tenaga sebesar 3.000 hp. Mobil bernama 'Chaos' ini hendak menyanggar rekor sebagai mobil paling bertenaga di dunia. Bagaimana jantung mekanisnya? Chaos supercar dilengkapi dengan mesin 4.0L V10 twin-turbo untuk kapasitas sekitar 2.000-3.000 tenaga kuda. Versi 2.000 tenaga kuda akan menggunakan bahan bakar bensin biasa, putaran mesin maksimum 11.000 rpm. Sementara itu, versi 3.000 tenaga kuda menggunakan bahan bakar E85, mampu berada di putaran mesin hingga 12.000 rpm. Hypercar 3000 Tenaga Kuda Spyros Panopoulos Automotive Foto: Spyros Panopoulos Automotive Pabrikannya menggunakan bahan premium untuk mesin. Versi 2.000 tenaga kuda memiliki piston dan kopling piston dengan crankshaft yang terbuat dari titanium, sedangkan versi yang lebih tinggi menggunakan bahan keramik untuk piston dan serat karbon untuk injeksi bahan bakar. Tenaga dari mesin ditransmisikan ke roda melalui gearbox dual-clutch 8-speed dan diferensial Torson. Supercar diharapkan mampu berakselerasi 0-96 km/jam dalam waktu kurang dari 2 detik. Belum banyak yang bisa terungkap mengenai proyek dari Spyros Panopoulos. Sebab dalam akun instagramnya (@spyros_panopoulos, mobil masih tertutup dengan kain putih. Namun Spyros Panopoulos Automotive berencana untuk memperkenalkan supercar baru di Geneva Motor Show pada bulan Maret tahun depan. Review Dalhatsu New Terios: Tampang Baru, Layak Diburu?	otomotif	otomotif	Delete
103	Begini Jadinya Kalau Honda Jazz Kawin Silang dengan Suzuki Swift, dan MPV	Modifikasi selain memaksimalkan penampilan kendaraan ada juga yang dilakukan dengan mencangkok dari kendaraan lain. Bisa saja komponen dari mobil lain disambung ke mobil dengan merek berbeda seperti modifikasi satu ini. Mobil ini dimodifikasi dengan dua bentuk khas dari city car yang ada di Indonesia. Dalam sebuah video di akun instagram (@konten_otomotif memperlihatkan sebuah modifikasi yang membuat orang-orang kebingungan. Di awal video terlihat bagian belakang mobil yang cukup khas mobil itu memakai wajah Suzuki Swift. Jika dilihat dari platform mobil secara keseluruhan, bagian belakang dan depan mobil ini tidak asli karena mobil ini memiliki panjang seperti MPV tiga baris. Dilihat dari nomor kendaraan mobil ini jelas berada di Indonesia atau lebih tepatnya registrasi di kota Bandung, Jawa Barat. Kejanggalan juga ditemui dari suara mobil yang terdengar cukup tua ditambah dengan keputan asap yang keluar dari knalpotnya. Dilihat pada bagian kaki, paduan modifikasi ini semakin aneh karena pemilihan velgnya. Velg ini biasanya ditemukan pada mobil-mobil berotot ala Amerika tahun 80-an. Sayangnya tidak ada keterangan apa wujud asli dari mobil dengan modifikasi campuran ini. Modifikasi pun juga tidak terlalu rapi dan terlihat sebagian sisi mobil yang lecet hingga catnya terkelupas. (Gambar:Instagram) Kontes	otomotif	otomotif	Delete

Gambar 6.1 Contoh Hasil *Testing*

Pada halaman hasil *testing* seperti gambar diatas, halaman tersebut menampilkan judul telah diinputkan dan teks berita yang telah diinputkan dan diklasifikasikan, menampilkan kategori asli dari artikel berita tersebut untuk membandingkan hasil dari *predict* kategori yang telah dibuat dengan model *decision tree*.

Untuk skenario penyamaan jumlah data per kategori yaitu dengan data per kategori berjumlah 1892 data. Akurasi yang di dapatkan adalah sebesar 77% yang

berarti lebih kecil dari skenario awal dengan total data yang lebih banyak namun tidak seimbang per kategorinya. Hasil akurasi dari skenario ini ada pada gambar 6.2 dibawah ini.

```

... ['news' 'travel' 'otomotif' ... 'news' 'food' 'finance']
Accuracy: 0.7750094375235939

```

	precision	recall	f1-score	support
finance	0.63	0.60	0.62	356
food	0.92	0.90	0.91	369
health	0.77	0.81	0.79	393
news	0.64	0.65	0.65	351
otomotif	0.80	0.82	0.81	386
sport	0.89	0.91	0.90	410
travel	0.76	0.71	0.73	384

Gambar 6.2 Hasil *balancing* data

Pada hasil *balancing* data yang telah dilakukan, penurunan dan peningkatan nilai precision, recall, serta F1-score dalam model dapat dijelaskan oleh interaksi yang kompleks antara beberapa faktor. Pertama, ketika kita melakukan *balancing* data, distribusi kelas dalam dataset mengalami perubahan yang dapat mempengaruhi kemampuan model untuk mengenali dan memprediksi kelas yang sebelumnya kurang diwakili.

BAB VII. KESIMPULAN DAN SARAN

7.1 Kesimpulan

Berdasarkan penelitian yang telah dilakukan kesimpulan yang dapat diambil adalah hasil dari akurasi model yang didapatkan dari *preprocessing*, *TF*, dan *decision tree* adalah 78% dengan nilai *precision*, *recall*, dan *f-measure* yang paling tinggi adalah kategori *news*. Dikarenakan data latih dan data uji yang dimiliki oleh kategori *news* memiliki jumlah paling banyak yaitu sebanyak 25565 data diantara kategori lainnya. Faktor lain yang mempengaruhi akurasi dan nilai *precision*, *recall*, dan *f-measure* adalah kualitas data, seperti contohnya pada kategori *sport* walaupun memiliki data paling sedikit untuk di latih dan di uji namun tetap memiliki nilai lebih tinggi dari kategori *otomotif*, dan *sport*.

Untuk akurasi data uji yang didapatkan adalah sebesar 80% tidak jauh berbeda dengan akurasi dari hasil model, namun untuk perhitungan *confusion matrix* lainnya nilai yang didapatkan adalah nilai yang rendah.

Jumlah data juga bisa mempengaruhi nilai dan akurasi yang dihasilkan, semakin banyak data yang dilatih maka akan semakin baik akurasi dan nilai dari *confusion matrix*.

Skenario *Balancing* data mendapatkan akurasi yang cukup tinggi walaupun dengan data yang lebih sedikit dari skenario awal. Maka dari itu bisa disimpulkan bahwa dengan skenario *Balancing* data dengan data yang lebih banyak bisa membuat akurasi model lebih tinggi dari akurasi sebelumnya.

7.2 Saran

Penelitian ini masih memiliki kekurangan sehingga berikut adalah saran yang dibutuhkan untuk penelitian selanjutnya, yaitu untuk data pengujian bisa ditambah lagi jumlah datanya dengan rata per kategori untuk mendapatkan akurasi yang mungkin bisa menyamai akurasi dari hasil model. Hal ini penting karena dalam kasus ketidakseimbangan data, dimana beberapa kategori memiliki lebih banyak sampel daripada yang lain, model mungkin cenderung lebih baik dalam memprediksi kategori mayoritas sementara mengabaikan kategori minoritas

DAFTAR PUSTAKA

- Adhe, D., Rachman, C., Goejantoro, R., Deny, F., & Amijaya, T. (2020). Implementasi Text Mining Pengelompokkan Dokumen Skripsi Menggunakan Metode K-Means Clustering Implementation Of Text Mining For Grouping Thesis Documents Using K-Means Clustering. *Jurnal EKSPONENSIAL*, 11(2).
- Cahyono, Y. (2017). ANALISIS SENTIMENT PADA SOSIAL MEDIA TWITTER MENGGUNAKAN NAÏVE BAYES CLASSIFIER DENGAN FEATURE SELECTION PARTICLE SWARM OPTIMIZATION DAN TERM FREQUENCY. *JURNAL INFORMATIKA UNIVERSITAS PAMULANG*, 14(1).
- Fadila, R. R., Aprison, W., & Musril, H. A. (2021). Perancangan Perizinan Santri Menggunakan Bahasa Pemograman PHP/MySQL Di SMP Nurul Ikhlas. *CSRID (Computer Science Research and Its Development Journal)*, 11(2), 84. <https://doi.org/10.22303/csrid.11.2.2019.84-95>
- Fauziningrum, E., & Pd, M. (2021). *EVALUASI DAN PREDIKSI PENGUASAAN BAHASA INGGRIS MARITIM MENGGUNAKAN METODE DECISION TREE DAN CONFUSION MATRIX (STUDI KASUS DI UNIVERSITAS MARITIM AMNI)*.
- Gallofré Ocaña, M., & Opdahl, A. (2022). Supporting Newsrooms with Journalistic Knowledge Graph Platforms: Current State and Future Directions. *Technologies*, 10(3), 68. <https://doi.org/10.3390/technologies10030068>
- Herwijayanti, B., Ratnawati, D. E., & Muflikhah, L. (2018). *Klasifikasi Berita Online dengan menggunakan Pembobotan TF-IDF dan Cosine Similarity* (Vol. 2, Issue 1). <http://j-ptiik.ub.ac.id>
- Josi, A., Akuntansi, K., Prabumulih, S., Patra No, J. L., Sukaraja, K., & Selatan, K. P. (2017). STMIK-MUSIRAWAS LUBUKLINGGAU 50 PENERAPAN METODE PROTOTIPING DALAM PEMBANGUNAN WEBSITE DESA (STUDI KASUS DESA SUGIHAN KECAMATAN RAMBANG). In *JTI* (Vol. 9, Issue 1).

- Jurnal, W., Indriani, N., Rainarli, E., Evita Dewi, K., dan Ilmu Komputer, T., & Dipati Ukur, J. (2017). JURNAL INFOTEL Informatika-Telekomunikasi-Elektronika Peringkasan dan Support Vector Machine pada Klasifikasi Dokumen. *Jurnal Infotel*, 9(4). <https://doi.org/10.20895/infotel.v9i4>
- Lutfi, A. (2017). *SCHOOL USING PHP AND MYSQL* (Vol. 3, Issue 2).
- Ma'rifah, H., Prasetya Wibawa, A., & Akbar, M. I. (2020). *Sains, Aplikasi, Komputasi dan Teknologi Informasi Klasifikasi artikel ilmiah dengan berbagai skenario preprocessing*. 2(2), 70.
- Maulidah, M., Gata, W., Aulianita, R., Agustyaningrum, C. I., Studi, P., Komputer, I., & Mandiri, N. (2020). *Algoritma Klasifikasi Decision Tree Untuk Rekomendasi Buku Berdasarkan Kategori Buku*. 13(2), 89–96. <http://journal.stekom.ac.id/index.php/E-Bisnis/page89>
- Pramudita, Y. D., Putro, S. S., & Makhmud, N. (2018). Klasifikasi Berita Olahraga Menggunakan Metode Naïve Bayes dengan Enhanced Confix Stripping Stemmer. *Jurnal Teknologi Informasi Dan Ilmu Komputer*, 5(3), 269. <https://doi.org/10.25126/jtiik.201853810>
- Pratiwi, B. P., Handayani, A. S., & Sarjana. (2020). Pengukuran Kinerja Sistem Kualitas Udara Dengan Teknologi WSN Menggunakan Confusion Matrix. *JURNAL INFORMATIKA UPGRIS Vol. 6, No. 2*.
- Rizkia, S., Budi, E., Si, S. S., Puspendari, D., & Pd, M. (2019). *Analisis Sentimen Kepuasan Pelanggan Terhadap Internet Provider Indihome di Twitter Menggunakan Metode Decision Tree dan Pembobotan TF-IDF*.
- Sartika, D., & Senses, D. I. (2017). *Perbandingan Algoritma Klasifikasi Naive Bayes, Nearest Neighbour, dan Decision Tree pada Studi Kasus Pengambilan Keputusan Pemilihan Pola Pakaian* (Vol. 1, Issue 2).
- Sovia, R. (2017). MEMBANGUN APLIKASI E-LIBRARY MENGGUNAKAN HTML, PHP SCRIPT, DAN MYSQL DATABASE. *Jurnal Processor*, 6.
- Tantyoko, H., Adiwijaya, & Wisety, N. U. (2019). Perbandingan Pembobotan Untuk Klasifikasi Topik Berita Menggunakan Decision Tree. *Jurnal Teknologia*, Vol. 2, 97–113.
- Wibowo Soejono, A., Setyanto, A., & Fatah Sofyan, A. (2018). Evaluasi Usability Website UNRIYO Menggunakan System Usability Scale (Studi Kasus:

Website UNRIYO). *Jurnal Teknologi Informasi*, XIII, 29–37.
www.respati.ac.id

Yulia Muniar, A., & Ria Lestari, K. (2020). *Penerapan Algoritma K-Nearest Neighbor pada Pengklasifikasian Dokumen Berita Online* PENERAPAN ALGORITMA K-NEAREST NEIGHBOR PADA PENGKLASIFIKASIAN DOKUMEN BERITA ONLINE.