



## Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

## BAB IV

# ANALISA DAN PERANCANGAN

Pada tahap ini merupakan tahap yang menentukan hasil dari sistem yang akan dibuat. Dimana analisa yaitu tahap kegiatan dalam penentuan tindakan atau keputusan yang akan menentukan dari rancangan sistem.

Sementara tahap perancangan menjadi tahap kegiatan menentukan rincian sistem yang akan dibuat.

### 4.1 Analisa Klasifikasi Topik Berita

Tahapan ini akan menjelaskan gambaran sistem secara keseluruhan dalam menentukan topik berita . Adapun proses utama yang dapat dilakukan pada sistem ini terdiri dari 2 yaitu penentuan *query* yang akan ditentukan topik beritanya, dan menentukan topik berita menggunakan algoritma *Naive Bayes*.

Pada dasarnya setiap proses utama akan melakukan tahapan secara umum yaitu: *input*, proses dan *output*. Sistem dimulai ketika *user* ingin menentukan klasifikasi topik berita. *User* akan melakukan *input* dokumen yaitu berupa teks Berita. Sistem akan melakukan seleksi dokumen yang telah di *input user* sebagai bagian dari proses penentuan *query*. Proses penentuan *query* pada sistem akan bertindak menyeleksi dokumen yang masuk kemudian membuang dokumen yang dianggap tidak dapat ditentukan topiknya, yaitu dokumen *duplicate*. Setelah melewati proses ini maka *output* dari proses penentuan *query* akan menjadi inputan *query* pada 3 proses selanjutnya. *User* dapat menentukan proses selanjutnya.

Pada proses penentuan *query* menggunakan algoritma *Naive Bayes*, *query* hasil dari proses sebelumnya akan dilakukan proses *preprocessing* yaitu terdiri dari *case folding*, menghapus karakter selain UTF-8 hilangkan *link url*, pemisahan rangkaian kata (*tokenizing*), *stopword* dan *stemming* dan pembobotan. Proses



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

berlanjut dengan klasifikasi *query*. Hasil dari klasifikasi ini merupakan hasil akhir dari klasifikasi topik berita menggunakan algoritma *Naive Bayes*.

Sementara itu pada proses penentuan klasifikasi berita menggunakan pendekatan *Naive Bayes* yaitu identifikasi setiap kata pada *query* dan kalimat, identifikasi kalimat hiburan, identifikasi kalimat kesehatan, identifikasi kalimat olahraga dan menghitung *score* kelas. Hasil dari perhitungan *score* akan menentukan klasifikasi topik berita tersebut.

#### 4.1.1 Analisa Pengumpulan Data

Sebelum melakukan proses klasifikasi, hal pertama yang dilakukan pada penelitian ini adalah mengumpulkan berita yang nantinya akan digunakan sebagai dataset. Data dikumpulkan secara manual melalui *Detik.us*. Data berita yang dikumpulkan berjumlah 150 data, waktu pengumpulan data keluhan antara bulan agustus 2016 sampai dengan Januari 2017 dan di simpan dengan format *\*txt*. Selanjutnya menentukan kategori berita yang akan dijadikan sebagai kelas pada proses klasifikasi menggunakan metode yang dibahas dalam penelitian ini. Adapun kategori dan ciri-ciri adalah sebagai berikut:

##### 1. Kategori Olahraga

Kategori Olahraga ditandai dengan informasi yang ada dalam kategori seperti berikut:

- a. Sepak Bola
- b. Basket
- c. badminton
- d. Formula 1
- e. Moto gp
- f. Renang



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah,
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

2. Kategori Kesehatan

Kategori Olahraga ditandai dengan informasi yang ada dalam kategori seperti berikut:

- a. Diet
- b. Makanan Sehat
- c. Gaya hidup

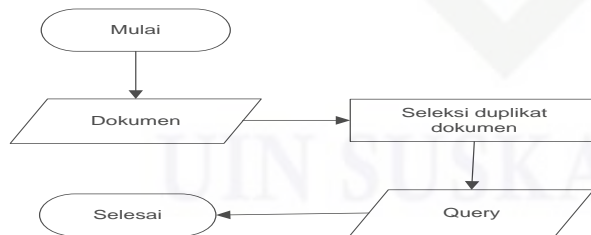
3. Kategori Hiburan

Kategori Hiburan ditandai dengan informasi yang ada dalam kategori seperti berikut:

- a. Travelling
- b. Film
- c. Musik
- d. infotainment

4.1.2 Analisa Penentuan Query

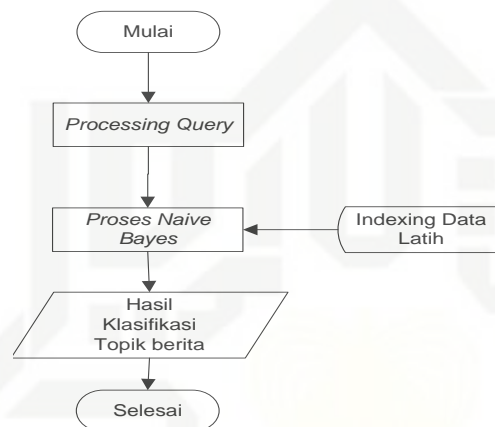
Pada tahap ini dokumen akan diseleksi, kemudian hasil seleksi akan dijadikan query pada proses selanjutnya. Tahapan ini diperlukan untuk melakukan *filtering* dikarenakan tidak setiap dokumen yang berhasil di *grab* pada sistem dapat dilakukan proses klasifikasi berita dan dokumen yang duplikat. Tahapan dari proses ini yaitu seleksi duplikat dokumen yakni berita duplikat dianggap sebagai *spamer*. *Flowchart* dari proses penentuan *query* ini dapat dijelaskan pada Gambar 4.1 berikut.



Gambar 4.1 Flowchart Penentuan Query

### 4.1.3 Analisa Menggunakan Algoritma Pembobotan Naive Bayes

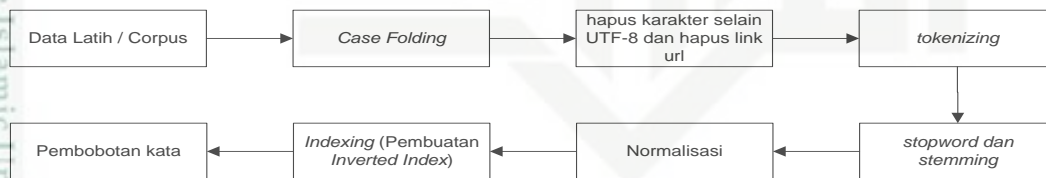
Analisa algoritma *Naive Bayes* merupakan tahapan menentukan proses pada klasifikasi topik berita. Proses ini merupakan tahapan yang akan menentukan hasil dari algoritma *Naive Bayes* dalam klasifikasi topik pada dokumen yang di *input*. Berikut Gambar 4.2 *flowchart diagram* analisa menggunakan algoritma *Naive Bayes*.



**Gambar 4.2 Flowchart Diagram Analisa Algoritma Naive Bayes**

Terdapat 3 proses utama dalam proses ini, yaitu :

1. Pembangunan index (pembuatan *inverted index*) dan pembobotan kata pada latih. Pada proses ini akan dilakukan pada Gambar 4.3 di bawah ini:



**Gambar 4.3 Proses Pembuatan *Inverted Index***

- a. Mengumpulkan dokumen yang didapat dari situs berita *online* detik.co berisi berita-beritasebagai latih yang telah diklasifikasikan sesuai dengan situs tersebut. Dalam hal ini kumpulan dokumen yang digunakan sebagai latih akan disusun secara berimbang. set yang digunakan dalam penelitaian ini



**Hak Cipta Dilindungi Undang-Undang**

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

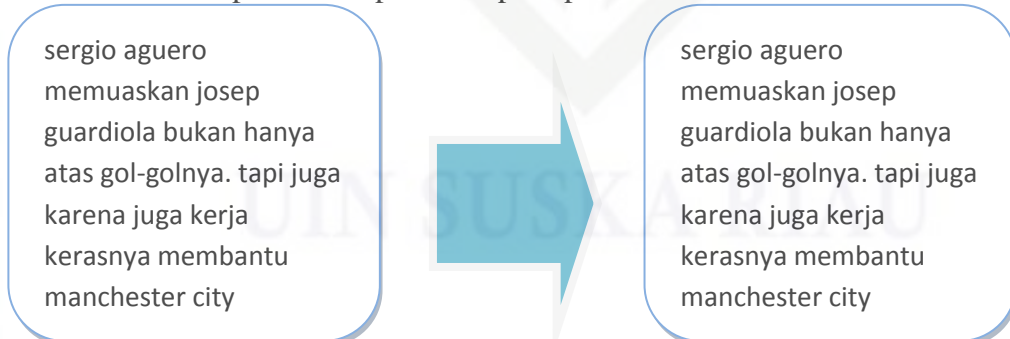
berjumlah 150 dokumen terdiri dari 135 dokumen sebagai latihan, sedangkan uji berjumlah 15 dokumen. latihan yang dibangun dalam penelitian ini berkenaan dengan berita hiburan, kesehatan dan olahraga pada bulan Agustus hingga Oktober. Jika topik diluar itu maka latihan harus di *training* kembali. Setelah tahapan ini maka tahapan berikutnya akan diproses menggunakan sistem.

- b. *Case folding* : tahapan ini merubah semua huruf kapital dengan huruf kecil. Tidak semua dokumen teks konsisten dalam penggunaan huruf. Oleh karena itu peran *case folding* dibutuhkan dalam mengkonversi keseluruhan teks dalam dokumen menjadi suatu bentuk standar (biasanya huruf kecil). Contoh proses *case folding* seperti pada Gambar 4.4 di bawah ini.



**Gambar 4.4 Contoh Proses Case Folding**

- c. Menghapus karakter selain UTF-8 dan *link url*. Proses ini akan menghapus karakter yang tidak relevan yang dianggap dapat mempengaruhi akurasi dari sistem. Contoh proses tahapan ini seperti pada Gambar 4.5 di bawah ini.



**Gambar 4.5 Contoh Proses Menghapus Karakter Selain UTF-8 dan link url**

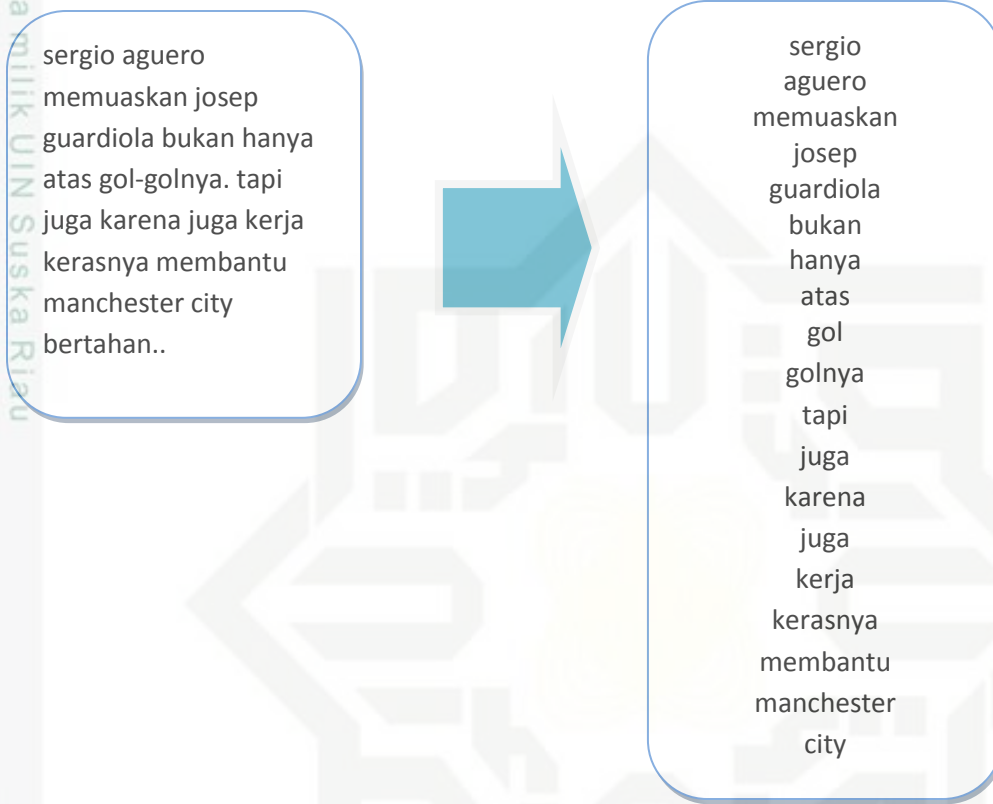
Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

d. Pemisahan rangkaian kata (*tokenizing*). Pada tahap ini seluruh kata pada dokumen akan dipisahkan menjadi token atau potongan kata tunggal. Contoh proses *tokenizing* seperti pada Gambar 4.6 di bawah ini.



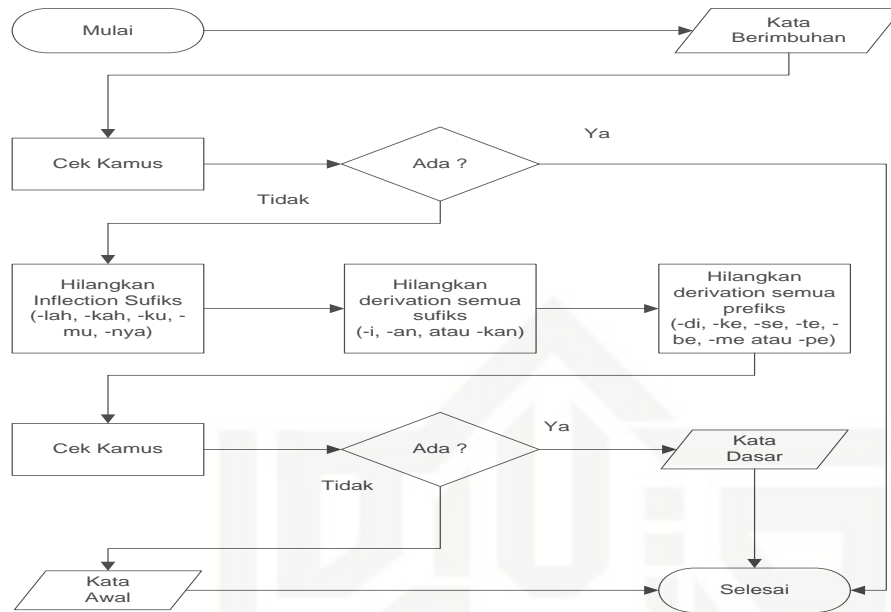
**Gambar 4.6 Contoh Proses *Tokenizing***

Proses *tokenizing* dari contoh Gambar 4.6 membuat dokumen yang sebelumnya terangkai menjadi terpisah seperti potongan kata tunggal.

e. Melakukan *linguistic preprocessing: stopword* dan *stemming*. Pada tahap ini kata-kata yang bersifat *stopword* akan dibuang hal ini di karnakan *stopword* dianggap tidak berguna dan tahapan ini juga dapat menekan penggunaan waktu dalam menentukan hasil. Selanjutnya dilakukan proses *stemming* dimana kata berimbuhan akan diubah dalam bentuk kata dasar dengan menggunakan algoritma nazief adriani. Proses *stemming* dengan algoritma Nazief dan Adriani dapat digambar dengan *flowchart* pada Gambar 4.7 di bawah ini:

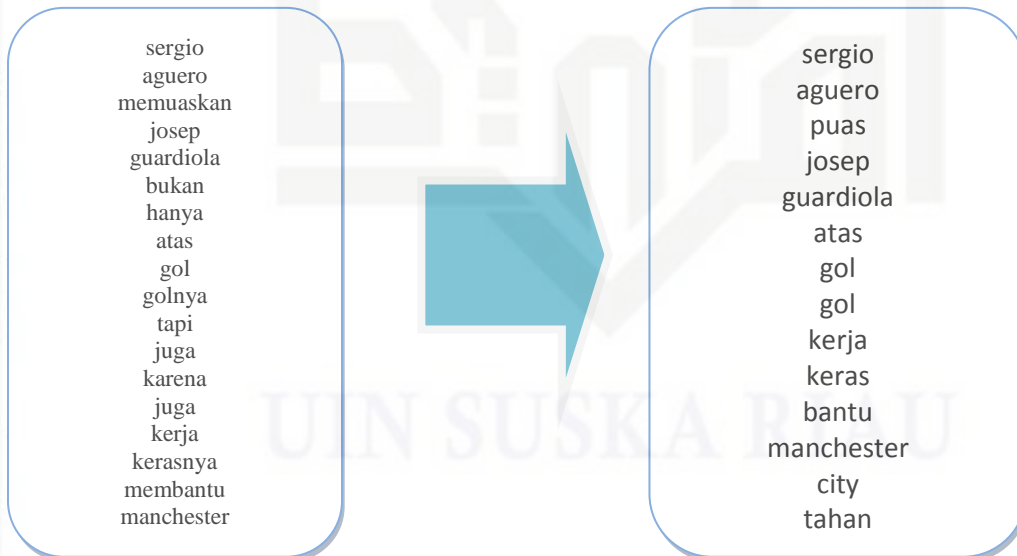
Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.



Gambar 4.7 Flowchart Algoritma Nazief dan Adriani

Contoh dari proses *linguistic preprocessing*: *stopword* dan *stemming* tahapan ini seperti pada Gambar 4.8 dibawah ini.



Gambar 4.8 Contoh Proses Linguistic Preprocessing: Stopword dan Stemming



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkannya dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Proses *linguistic preprocessing* dari contoh Gambar 4.8 terdapat pada kata “memuaskan” dirubah menjadi “puas” yang berarti telah dilakukan penghapusan imbuhan oleh sistem. Kemudian dilakukan penghapusan kata *stop words* yaitu “bukan” dimana sistem akan merubah kata tersebut menjadi “atas”

- f. Indexing (Pembuatan *Inverted Index*). Pada tahap ini, dilakukan proses pengindeksan pada setiap token. Pada *inverted index* akan tersimpan informasi berupa kata, id dokumen yang mengandung kata tersebut dan frekuensi kemunculan token tersebut dalam suatu dokumen.
- g. Pembobotan kata. Hasil dari *indexing* pada proses selanjutnya akan dihitung nilai bobotnya. Pada penelitian ini digunakan dua jenis perhitungan pembobotan. yakni pembobotan lokal dengan menggunakan *tf (term frequency)* dan pembobotan global dengan menggunakan *idf (inverse dokumen frequency)*.

Untuk lebih jelasnya, berikut adalah contoh pembuatan *inverted index* dan Pembobotan kata pada latihan secara keseluruhan.

- a. Mengumpulkan dokumen yang didapat dari kumpulan berita berisi topik beritasebagai latihan. Sebagai contoh digunakan 6 dokumen sebagai latihan. Berikut ini kumpulan dokumen latihan yang ditunjukkan pada Tabel 4.1.

**Tabel 4.1 Dokumen Latihan**

Dok(n)	Berita	Topik berita
Dok(1)	DETIK..COM , <b>Jakarta</b> , Pisau cukur menjadi salah satu alat yang digunakan untuk memangkas rambut tubuh, termasuk rambut kemaluan supaya lebih rapi dan tidak lebat. Sebab, seperti diketahui jika tak dirawat, rambut kemaluan yang lebat bisa jadi tempat berkembangnya kuman.	Kesehatan
Dok(2)	DETIK..COM , <b>Jakarta</b> - Setelah jeda nyaris sebulan gelaran F1 2016 akan kembali bergulir di Spa, Belgia. Ada sejumlah catatan terkait balapan itu, termasuk kans Lewis	Olahraga



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Dok(n)	Berita	Topik berita
	Hamilton menyamai torehan Juan Manuel Fangio dan Ayrton Senna.	
Dok(3)	DETIK..COM , <b>Jakarta</b> - Vin Diesel dan Dwayne Johnson kini dikabarkan tengah berseteru. Keduanya bahkan saling sindir di media sosial.Drama dimulai setelah Dwayne mengunggah sebuah video di Instagram yang mengkritik salah satu lawan mainnya di film 'Fast 8' .	Hiburan
Dok(4)	DETIK..COM, <b>Jakarta</b> - Bulan lalu, publik dihebohkan dengan video kekerasan yang dilakukan sang tunangan, Egor Tarabasov, kepada Lindsay Lohan. Kini, sang aktris kembali memamerkan cincin pertunangannya.	Hiburan
Dok(5)	DETIK..COM , <b>Brno</b> - Valentino Rossi tak menyangkal, Marc Marquez berada di pole position untuk menjuarai musim ini. Rossi kini mulai mengubah fokusnya, yakni merebut posisi dua	Olahraga
Dok(6)	DETIK..COM, <b>Jakarta</b> , Serangan jantung merupakan suatu kondisi jantung tidak dapat berfungsi setelah otot penggeraknya rusak, akibat kekurangan pasokan oksigen. Penyebab utama adalah aterosklerosis, penyempitan dan pengerasan pembuluh darah, yang menyebabkan darah pembawa oksigen terhambat sampai ke jaringan sel otot jantung. Salah satu penyebab terjadinya aterosklerosis yaitu kadar kolesterol darah yang tinggi, sehingga terjadi penumpukkan pada dinding pembuluh darah yang berlangsung lama menjadi gumpalan keras atau plak yang menyumbat pembuluh darah.	Kesehatan

- b. *Case folding* : tahapan ini merubah semua huruf kapital dengan huruf kecil. Berikut ini hasil *case folding* yang ditunjukkan pada Tabel 4.2

**Tabel 4.2 Hasil Case Folding Latih**

Dok(n)	Berita	Topik berita
Dok(1)	detik.com ,jakarta, pisau cukur menjadi salah satu alat yang digunakan untuk memangkas rambut tubuh, termasuk rambut kemaluan supaya lebih rapi dan tidak lebat. sebab, seperti diketahui jika tak dirawat, rambut kemaluan yang lebat bisa jadi tempat berkembangnya kuman	Kesehatan
Dok(2)	detik.com ,jakarta - setelah jeda nyaris sebulan gelaran fl 2016 akan kembali bergulir di spa, belgia. ada sejumlah catatan terkait balapan itu, termasuk kans lewis hamilton menyamai torehan juan manuel fangio dan ayrton senna.	Olahraga
Dok(3)	detik.com ,jakarta - vin diesel dan dwayne johnson kini dikabarkan tengah berseteru. keduanya bahkan saling sindir di media sosial.drama dimulai setelah dwayne mengunggah sebuah video di instagram yang mengkritik salah satu lawan mainnya di	Hiburan



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Dok(n)	Berita	Topik berita
	film 'fast 8'	
Dok(4)	detik.com, jakarta - bulan lalu, publik dihebohkan dengan video kekerasan yang dilakukan sang tunangan, egor tarabasov, kepada lindsay lohan. kini, sang aktris kembali memamerkan cincin pertunangannya.	Hiburan
Dok(5)	detik.com ,brno - valentino rossi tak menyangkal, marc marquez berada di pole position untuk menjuarai musim ini. rossi kini mulai mengubah fokusnya, yakni merebut posisi dua.	Olahraga
Dok(6)	detik.com, jakarta, serangan jantung merupakan suatu kondisi jantung tidak dapat berfungsi setelah otot penggeraknya rusak, akibat kekurangan pasokan oksigen. penyebab utama adalah aterosklerosis, penyempitan dan pengerasan pembuluh darah, yang menyebabkan darah pembawa oksigen terhambat sampai ke jaringan sel otot jantung. salah satu penyebab terjadinya aterosklerosis yaitu kadar kolesterol darah yang tinggi, sehingga terjadi penumpukkan pada dinding pembuluh darah yang berlangsung lama menjadi gumpalan keras atau plak yang menyumbat pembuluh darah.	Kesehatan

- c. Menghapus karakter selain UTF-8 dan *link url*. Berikut ini hasil penghapusan karakter selain UTF-8 dan *link url* yang ditunjukkan pada pada Tabel 4.3

**Tabel 4.3 Hasil Penghapusan Karakter selain UTF-8 dan link URL**

Dok(n)	Berita	Topik berita
Dok(1)	pisau cukur menjadi salah satu alat yang digunakan untuk memangkas rambut tubuh, termasuk rambut kemaluan supaya lebih rapi dan tidak lebat sebab seperti diketahui jika tak dirawat rambut kemaluan yang lebat bisa jadi tempat berkembangnya kuman	Kesehatan
Dok(2)	setelah jeda nyaris sebulan gelaran f1 2016 akan kembali bergulir di spa, belgia ada sejumlah catatan terkait balapan itu termasuk kans lewis hamilton menyamai torehan juan manuel fangio dan ayrton senna	Olahraga
Dok(3)	vin diesel dan dwayne johnson kini dikabarkan tengah berseteru keduanya bahkan saling sindir di media sosial drama dimulai setelah dwayne mengunggah sebuah video di instagram yang mengkritik salah satu lawan mainnya di film fast 8	Hiburan
Dok(4)	bulan lalu publik dihebohkan dengan video kekerasan yang dilakukan sang tunangan egor tarabasov kepada lindsay lohan kini sang aktris kembali memamerkan cincin pertunangannya	Hiburan



**Hak Cipta Dilindungi Undang-Undang**

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Dok(n)	Berita	Topik berita
Dok(5)	valentino rossi tak menyangkal marc marquez berada di pole position untuk menjuarai musim ini rossi kini mulai mengubah fokusnya yakni merebut posisi dua	Olahraga
Dok(6)	serangan jantung merupakan suatu kondisi jantung tidak dapat berfungsi setelah otot penggeraknya rusak akibat kekurangan pasokan oksigen penyebab utama adalah aterosklerosis penyempitan dan pengerasan pembuluh darah yang menyebabkan darah pembawa oksigen terhambat sampai ke jaringan sel otot jantung salah satu penyebab terjadinya aterosklerosis yaitu kadar kolesterol darah yang tinggi sehingga terjadi penumpukkan pada dinding pembuluh darah yang berlangsung lama menjadi gumpalan keras atau plak yang menyumbat pembuluh darah	Kesehatan

d. Pemisahan rangkaian kata (*tokenizing*). Berikut ini hasil tokenisasi yang ditunjukkan pada pada Tabel 4.4

**Tabel 4.4 Hasil Proses Tokenisasi Setiap Dokumen**

Dok(1)	Dok(2)	Dok(3)	Dok(4)	Dok(5)	Dok(6)
Pisau	Setelah	Vin	Bulan	Valentino	Serangan
Cukur	Jeda	Diesel	Lalu	Rossi	Jantung
menjadi	Nyaris	Dan	Publik	Tak	Merupakan
Salah	Sebulan	Dwayne	Dihebohkan	Menyangkal	Suatu
Satu	Gelaran	Johnson	Dengan	Marc	Kondisi
Alat	f1	Kini	Video	Marques	Jantung
Yang	2016	Dikabarkan	Kekerasan	Berada	Tidak
digunakan	Akan	Tengah	Yang	Dipole	Dapat
Untuk	Kembali	Berseteru	Dilakukan	Position	Berfungsi
memangkas	Bergulir	Keduanya	Sang	Untuk	Setelah
rambut	Dispa	Bahkan	Tunangan	Menjuarai	Otot
Tubuh	Belgia	Saling	Egor	Musin	Penggeraknya
termasuk	Ada	Sindir	Tarabasof	Ini	Rusak





**Hak Cipta Dilindungi Undang-Undang**

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Dok(1)	Dok(2)	Dok(3)	Dok(4)	Dok(5)	Dok(6)
					Terjadinya
					Aterosklerosis
					Yaitu
					Kadar
					Kolestrol
					Darah
					Yang
					Tinggi
					Sehingga
					Terjadi
					Penumpukan
					Pada
					Dinding
					Pembuluh
					Darah
					Yang
					Berlansung
					Lama
					Menjadi
					Gumpalan
					Keras
					Atau
					Plak
					Yang
					Menyumbat
					Pembuluh
					darah





Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Dok(1)	Dok(2)	Dok(3)	Dok(4)	Dok(5)	Dok(6)
					Jaring
					Sel

f. Indexing (Pembuatan *Inverted Index*) . Berikut ini hasil pembuatan *inverted index* yang ditunjukkan pada pada Tabel 4.6.

**Tabel 4.6 Hasil Proses Indexing (Pembuatan Inverted Index)**

Kosa Kata	(dokumen : frekuensi)
Pisau	1:1
Cukur	1:1
Alat	1:1
Guna	1:1
Pangkas	1:1
Rambut	1:2
Tubuh	1:1
Rambut	1:1
Kemaluan	1:2
Rapi	1:1
Lebat	1:2
Rawat	1:1
Tempat	1:1
Kembang	1:1
Kuman	1:1
Jeda	2:1
Nyaris	2:1
Bulan	2:1



**Hak Cipta Dilindungi Undang-Undang**

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Kosa Kata	(dokumen : frekuensi)
Gelar	2:1
F1	2:1
Gulir	2:1
Dispa	2:1
Belgia	2:1
Jumlah	2:1
Catat	2:1
Kait	2:1
Balap	2:1
Masuk	2:1
Torehan	2:1
Kabar	3:1
Tengah	3:1
Seteru	3:1
Sindir	3:1
Media	3:1
Sosial	3:1
Drama	3:1
Mulai	3:1
Video	3:1
Kritik	3:1
Lawan	3:1
Main	31
Film	3:1



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Kosa Kata	(dokumen : frekuensi)
Bulan	4:1
Publik	4:1
Heboh	4:1
Video	4:1
Keras	4:1
Laku	4:1
Sang	4:2
Tunangan	4:1
Artis	4:1
Mamer	4:1
Cincin	4:1
Nyangkal	5:1
Juara	5:1
Musim	5:1
Ubah	5:1
Fokus	5:1
Rebut	5:1
Posisi	5:1
Dua	5:1
Serang	6:1
Jantung	6:3
Kondisi	6:2
Fungsi	6:1
Otot	6:1



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Kosa Kata	(dokumen : frekuensi)
Gerak	6:1
Rusak	6:1
Kurang	6:1
Pasok	6:1
Oksigen	6:1
Utama	6:1
Aterosklerosis	6:1
Sempit	6:1
Keras	6:1
Darah	6:2

Keterangan :

Kosa kata : Kata tunggal yang terdapat dalam dokumen

(dokumen frekuensi) : Frekuensi kata dalam dokumen. Misal kata algoritma dalam dokumen 1 memiliki frekuensi 1 dan dalam dokumen 3 memiliki frekuensi 1.

g. Pembobotan kata. Berikut ini hasil pembobotan kata terhadap kata hasil pembobotan kata yang ditunjukkan pada Tabel 4.7

**Tabel 4.7 Hasil Pembobotan Kata**

Kosa Kata	$tf(D1)$	$tf(D2)$	$tf(D3)$	$tf(D4)$	$tf(D5)$	$tf(D6)$	$Df$
Pisau	1	0	0	0	0	0	1
Cukur	1	0	0	0	0	0	1
Alat	1	0	0	0	0	0	1
Guna	1	0	0	0	0	0	1



**Hak Cipta Dilindungi Undang-Undang**

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Kosa Kata	$tf(D1)$	$tf(D2)$	$tf(D3)$	$tf(D4)$	$tf(D5)$	$tf(D6)$	$Df$
Pangkas	1	0	0	0	0	0	1
Rambut	2	0	0	0	0	0	2
Tubuh	1	0	0	0	0	0	1
Rambut	1	0	0	0	0	0	1
Kemaluan	1	0	0	0	0	0	1
Rapi	1	0	0	0	0	0	1
Lebat	1	0	0	0	0	0	1
Rawat	1	0	0	0	0	0	1
Rambut	1	0	0	0	0	0	1
Kemaluan	2	0	0	0	0	0	2
Lebat	1	0	0	0	0	0	1
Tempat	1	0	0	0	0	0	1
Kembang	1	0	0	0	0	0	1
Kuman	1	0	0	0	0	0	1
Jeda	0	1	0	0	0	0	1
Nyaris	0	1	0	0	0	0	1
Bulan	0	1	0	0	0	0	1
Gelar	0	1	0	0	0	0	1
FI	0	1	0	0	0	0	1

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Kosa Kata	$tf(D1)$	$tf(D2)$	$tf(D3)$	$tf(D4)$	$tf(D5)$	$tf(D6)$	$Df$
Gulir	0	1	0	0	0	0	1
Dispa	0	1	0	0	0	0	1
Belgia	0	1	0	0	0	0	1
Jumlah	0	1	0	0	0	0	1
Catat	0	1	0	0	0	0	1
Kait	0	1	0	0	0	0	1
Balap	0	1	0	0	0	0	1
Masuk	0	1	0	0	0	0	1
Torehan	0	1	0	0	0	0	1
Kabar	0	0	1	0	0	0	1
Tengah	0	0	1	0	0	0	1
Seteru	0	0	1	0	0	0	1
Sindir	0	0	1	0	0	0	1
Media	0	0	1	0	0	0	1
Sosial	0	0	1	0	0	0	1
Drama	0	0	1	0	0	0	1
Mulai	0	0	1	0	0	0	1
Video	0	0	1	0	0	0	1
Kritik	0	0	1	0	0	0	1
Lawan	0	0	1	0	0	0	1



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Kosa Kata	$tf(D1)$	$tf(D2)$	$tf(D3)$	$tf(D4)$	$tf(D5)$	$tf(D6)$	$Df$
Main	0	0	1	0	0	0	1
Film	0	0	1	0	0	0	1
Bulan	0	0	0	1	0	0	1
Publik	0	0	0	1	0	0	1
Heboh	0	0	0	1	0	0	1
Video	0	0	0	1	0	0	1
Keras	0	0	0	1	0	0	1
Laku	0	0	0	2	0	0	2
Sang	0	0	0	1	0	0	1
Tunangan	0	0	0	1	0	0	1
Sang	0	0	0	1	0	0	1
Artis	0	0	0	1	0	0	1
Mamer	0	0	0	1	0	0	1
Cincin	0	0	0	1	0	0	1
Nyangkal	0	0	0	0	1	0	1
Juara	0	0	0	0	1	0	1
Musim	0	0	0	0	1	0	1
Ubah	0	0	0	0	1	0	1
Fokus	0	0	0	0	1	0	1



**Hak Cipta Dilindungi Undang-Undang**

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Kosa Kata	$tf(D1)$	$tf(D2)$	$tf(D3)$	$tf(D4)$	$tf(D5)$	$tf(D6)$	$Df$
Rebut	0	0	0	0	1	0	1
Posisi	0	0	0	0	1	0	1
Dua	0	0	0	0	1	0	1
Serang	0	0	0	0	0	1	1
Jantung	0	0	0	0	0	3	3
Kondisi	0	0	0	0	0	2	2
Jantung	0	0	0	0	0	1	1
Kondisi	0	0	0	0	0	1	1
Jantung	0	0	0	0	0	1	1
Fungsi	0	0	0	0	0	1	1
Otot	0	0	0	0	0	1	1

Keterangan :

Kosa kata : kata tunggal yang telah terpisah dari dokumen query

$Df$  : *Document frequency* atau jumlah dokumen dalam koleksi dokumen yang mengandung kosa kata

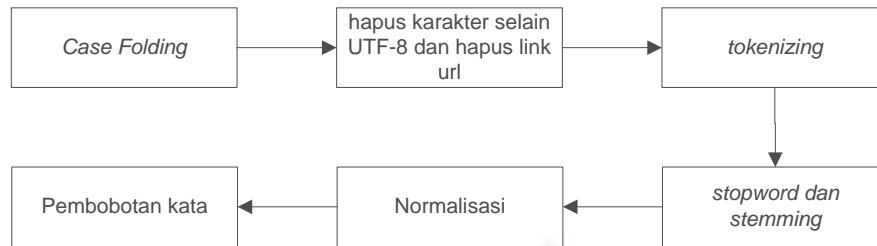
$idf$  : *inverse document frequency*(Persamaan 2.4)

$tf(n)$  : *Term frequency* (frekuensi kata) pada dokumen ke- $n$

## Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

2. *Preprocessing query*. Pada proses ini akan dilakukan:



**Gambar 4.10** Proses Pembuatan *Inverted Index Query*

- a. Proses *Case folding* yaitu merubah dokumen menjadi huruf kecil.
- b. Menghapus karakter selain UTF-8 dan *link url*.
- c. Pemisahan rangkaian kata (*tokenizing*).
- d. Melakukan *linguistic preprocessing*: *stopword* dan *stemming*.
- e. Normalisasi : tahapan identifikasi kata slang dan penulisan kata berlebihan kemudian di ganti dengan kata kamus KBBI.
- f. Pembobotan kata.

Untuk lebih jelasnya, akan dijelaskan tahapan *Preprocessing query*. Sebagai contoh *query* yang digunakan adalah

**Query** : DETIK..COM ,**Jakarta**, Pisau cukur menjadi salah satu alat yang digunakan untuk memangkas rambut tubuh, termasuk rambut kemaluan supaya lebih rapi dan tidak lebat. Sebab, seperti diketahui jika tak dirawat, rambut kemaluan yang lebat bisa jadi tempat berkembangnya kuman.

- a. Proses *Case folding*
  - detik..com ,**jakarta**, pisau cukur menjadi salah satu alat yang digunakan untuk memangkas rambut tubuh, termasuk rambut kemaluan supaya lebih rapi dan tidak lebat. sebab, seperti diketahui jika tak dirawat, rambut kemaluan yang lebat bisa jadi tempat berkembangnya kuman.
- b. Menghapus karakter selain UTF-8, *link url*
  - pisau cukur menjadi salah satu alat yang digunakan untuk memangkas rambut tubuh, termasuk rambut kemaluan supaya lebih rapi dan tidak



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

lebat. sebab, seperti diketahui jika tak dirawat, rambut kemaluan yang lebat bisa jadi tempat berkembangnya kuman.

c. Pemisahan rangkaian kata (*tokenizing*)

- pisau                    - cukur                    - menjadi                    -salah
- satu                    - alat                    - untuk                    -memangkas
- rambut                    -tubuh                    - termasuk                    -rambut
- kemaluan                    -supaya                    - lebih                    - rapi
- dan                    -tidak                    - lebat                    - sebab
- seperti                    -diketahui                    - jika                    - tidak
- dirawat                    -rambut                    -kemaluan                    -yang
- lebat                    -bisa                    -jadi                    -tempat
- berkembangnya                    -kuman
- 

d. Melakukan *linguistic preprocessing*: *stopword* dan *stemming*

- rambut                    - tubuh                    - masuk                    - rambut
- malu                    - rapi                    - lebat                    - rawat
- tempat                    -kembang                    -kuman

e. Normalisasi : tahapan identifikasi kata slang dan penulisan kata berlebihan kemudian di ganti dengan kata kamus KBBI

- rambut                    - tubuh                    - masuk                    - rambut
- malu                    - rapi                    - lebat                    - rawat
- tempat                    - kembang                    - kuman

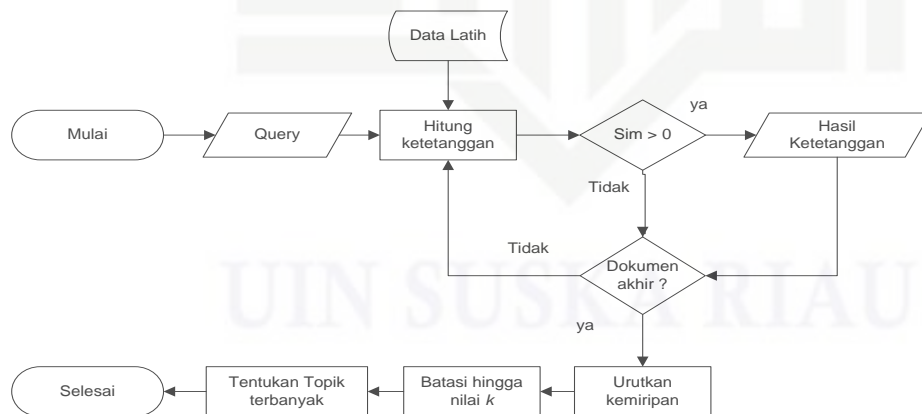
f. Pembobotan kata

- $tf_q$ rambut = 1      - $tf_q$ tubuh = 1
- $tf_q$ masuk = 1      - $tf_q$ rambut = 1
- $tf_q$  malu = 1      -  $tf_q$  rapi = 1
- $tf_q$  lebat = 1      -  $tf_q$  rawat = 1
- $tf_q$  tempat = 1      -  $tf_q$  kembang = 1
- $tf_q$  kuman = 1

3. Klasifikasi *query*. Pada proses ini akan dilakukan:

1. Menghitung ketetangaan terhadap *latih* yang telah dibangun. Setiap dokumen akan di hitung jaraknya dengan *query* menggunakan persamaan 2.5 *Cosine Similarity*.
2. Menentukan nilai pada algoritma *Naive Bayes*. Nilai *k* merupakan parameter yang akan membatasi ketetangaan terdekat yang akan diproses pada tahap selanjutnya.
3. Menyimpulkan klasifikasi topik berita dengan persamaan 2.6.

Tahapan klasifikasi ini dapat dilihat pada *flowchart* t gambar 4.11 berikut:



Gambar 4.11 Flowchart Proses Klasifikasi *Query*





1. Dilarang menjiplak atau menyalin sebagian atau seluruh karya tulis ini tanpa mengemukakan dan menyebutkan sumber.

2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.
- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

Kosa Kata	Tf								Idf	Wdt=Tf.idf						
	q	1	2	3	4	5	6	Df		Log n/df	W (Q)	W (D1)	W (D2)	W (D3)	W (D4)	W (D5)
Lebat	1	2		1				3	0,367	0,544	1,088	0	0,367	0	0	0
Rawat	1	1						2	0,544	0,544	0,544	0	0	0	0	0
Rambut	1	2				1		3	0,544	0,544	1,088	0	0	0	0,544	0
Kemaluan	1	1					1	2	0,544	0,544	0,778	0	0	0	0,544	0
Lebat	1	1						2	0,544	0,544	0,778	0	0	0	0	0
Tempat	1	1						2	0,544	0,544	0,778	0	0	0	0	0
Kembang	1	1		1				3	0,367	0,544	0,778	0	0	0	0	0
Kuman	1	1		2				3	0,367	0,544	0,778	0	0	0	0	0
Jeda	0		1					2	0,544	0	0	0,544	0	0	0	0
Nyaris			1					1	0,845	0	0	0,845	0	0	0	0
Bulan			1					1	0,845	0	0	0,845	0	0	0	0
Gelar			1					1	0,845	0	0	0,845	0	0	0	0
F1			1					1	0,845	0	0	0,845	0	0	0	0
Gulir			1					1	0,845	0	0	0,845	0	0	0	0
Dispa			1					1	0,845	0	0	0,845	0	0	0	0
Belgia			1					1	0,845	0	0	0,845	0	0	0	0
Jumlah			1					1	0,845	0	0	0,845	0	0	0	0
Catat			1					1	0,845	0	0	0,845	0	0	0	0
Kait			1					1	0,845	0	0	0,845	0	0	0	0



Kosa Kata	Tf								Idf	Wdt=Tf.idf						
	q	1	2	3	4	5	6	Df		Log n/df	W (Q)	W (D1)	W (D2)	W (D3)	W (D4)	W (D5)
Balap			1					1	0.845	0	0	0.845	0	0	0	0
Masuk			1			3		2	0.544	0	0	0.544	0	0	0	0
Forehan			1					1	0.845	0	0	0.845	0	0	0	0
Kabar				1				1	0.845	0	0	0	0.845	0	0	0
Tengah				1				1	0.845	0	0	0	0.845	0	0	0
Seteru				1				1	0.845	0	0	0	0.845	0	0	0
Sindir				1				1	0.845	0	0	0	0.845	0	0	0
Media				1				1	0.845	0	0	0	0.845	0	0	0
Sosial				1				1	0.845	0	0	0	0.845	0	0	0
Drama				1				1	0.845	0	0	0	0.845	0	0	0
Mulai				1				1	0.845	0	0	0	0.845	0	0	0
Video				1				1	0.845	0	0	0	0.845	0	0	0
Kritik				1				1	0.845	0	0	0	0.845	0	0	0
Lawan				1				1	0.845	0	0	0	0.845	0	0	0
Main				1				1	0.845	0	0	0	0.845	0	0	0
Film				1				1	0.845	0	0	0	0.845	0	0	0
Bulan					1			1	0.845	0	0	0	0	0.845	0	0
Publik					1			1	0.845	0	0	0	0	0.845	0	0



Kosa Kata	Tf								Idf	Wdt=Tf.idf						
	q	1	2	3	4	5	6	Df		Log n/df	W (Q)	W (D1)	W (D2)	W (D3)	W (D4)	W (D5)
Heboh					1			1	0.845	0	0	0	0	0.845	0	0
Video					1			1	0.845	0	0	0	0	0.845	0	0
Keras					1			1	0.845	0	0	0	0	0.845	0	0
Laku					1			1	0.845	0	0	0	0	0.845	0	0
Sang					2			1	0.845	0	0	0	0	0.845	0	0
Tunangan					1			1	0.845	0	0	0	0	0.845	0	0
Sang					1			1	0.845	0	0	0	0	0.845	0	0
Artis					1			1	0.845	0	0	0	0	0.845	0	0
Mamer					1			1	0.845	0	0	0	0	0.845	0	0
Cincin					1			1	0.845	0	0	0	0	0.845	0	0
Nyangkal						1		1	0.845	0	0	0	0	0	0.845	0
Juara						1		1	0.845	0	0	0	0	0	0.845	0
Musim						1		1	0.845	0	0	0	0	0	0.845	0
Ubah						1		1	0.845	0	0	0	0	0	0.845	0
Fokus						1		1	0.845	0	0	0	0	0	0.845	0
Rebut						1		1	0.845	0	0	0	0	0	0.845	0
Posisi						1		1	0.845	0	0	0	0	0	0.845	0
Dua						1		1	0.845	0	0	0	0	0	0.845	0



1. Dilarang menyalin atau menjiplak sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:  
 a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.  
 b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.  
 2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Kosa Kata	Tf							Idf	Wdt=Tf.idf							
	q	1	2	3	4	5	6		Df	Log n/df	W (Q)	W (D1)	W (D2)	W (D3)	W (D4)	W (D5)
Serang							1	1	0.845	0	0	0	0	0	0	0.845
Jantung							3	1	0.845	0	0	0	0	0	0	2,535
Kondisi							1	1	0.845	0	0	0	0	0	0	0.845
Fungsi							1	1	0.845	0	0	0	0	0	0	0.845
Otot							1	1	0.845	0	0	0	0	0	0	0.845
Gerak							1	1	0.845	0	0	0	0	0	0	0.845
Rusak							1	1	0.845	0	0	0	0	0	0	0.845
Kurang							1	1	0.845	0	0	0	0	0	0	0.845
Pasok							1	1	0.845	0	0	0	0	0	0	0.845
Oksigen							1	1	0.845	0	0	0	0	0	0	0.845
Darah							1	1	0.845	0	0	0	0	0	0	0.845
Sebab							1	1	0.845	0	0	0	0	0	0	0.845

## 4.2 Proses Prediksi dengan Naïve Bayes Classifier

Berdasarkan analisa topik berita hiburan, kesehatan, dan olahraga data yang telah dijabarkan sebelumnya, maka pada bagian ini dijelaskan teknik-teknik yang akan digunakan dalam klasifikasi data.

## Hak Cipta Dilindungi Undang-Undang

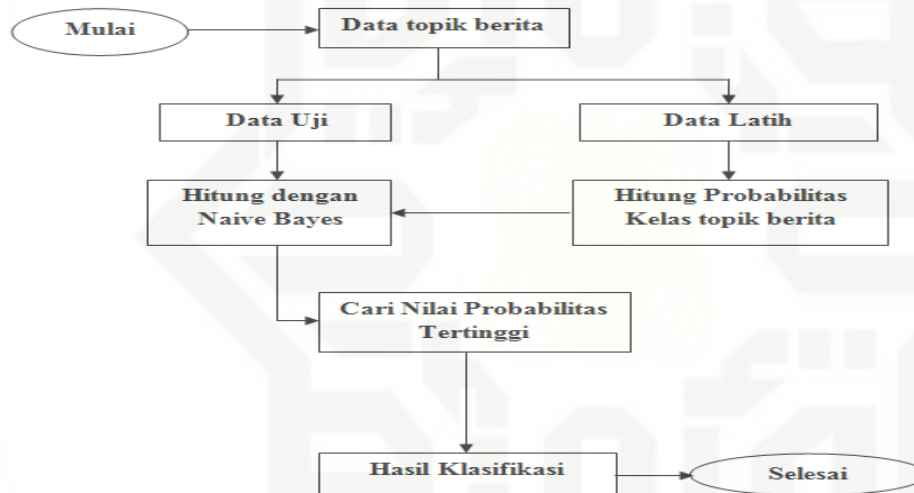
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkannya dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

#### 4.2.1 Algoritma Naive bayes classifier Classifier

Naïve Bayes merupakan salah satu metode machine learning yang menggunakan perhitungan probabilitas. Metode ini memanfaatkan teori probabilitas yang dikemukakan oleh ilmuwan Inggris Thomas Bayes , yaitu memprediksi probabilitas di masa depan berdasarkan pengalaman di masa sebelumnya. Naïve Bayes adalah algoritma yang termasuk ke dalam *supervised learning*, maka akan dibutuhkan pengetahuan awal untuk dapat mengambil keputusan. Langkah-langkah mengenai cara kerja algoritma *Naive bayes classifier* dalam klasifikasi topik berita ini akan dijelaskan pada *flowchart* yang diperlihatkan pada gambar 4. 12



**Gambar 4.12 Flowchart Klasifikasi Topik Berita Naïve Bayes**

Pada proses dengan *Naive bayes classifier* dibagi menjadi tahap *training* (latih) dan *testing* (uji). Berikut penjelasan dari tahapan tersebut

##### a. Tahap *training* (latih)

Pada tahap ini data yang telah didapatkan TF-IDF nya seperti sampel pada Tabel 4.7 dijadikan data latih untuk menjadi acuan dalam membentuk model klasifikasi. Pada fase ini akan dicari nilai probabilitas kelas dari data latih.

Menghitung nilai probabilitas (persamaan 2.8 dan persamaan 2.9) setiap *term* untuk setiap kelas dari data latih. Perhitungannya adalah sebagai berikut:



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkannya dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

1. Pertama hitung probabilitas setiap kategori menggunakan persamaan 2.8, dalam penelitian ini yang menjadi kategori ada dua yaitu, kategori iya dan tidak.

$$p(\text{hiburan}) = \frac{f_d(\text{hiburan})}{|D|} = \frac{50}{150} = \frac{1}{3}$$

$$p(\text{kesehatan}) = \frac{f_d(\text{kesehatan})}{|D|} = \frac{50}{150} = \frac{1}{3}$$

$$p(\text{olahraga}) = \frac{f_d(\text{olahraga})}{|D|} = \frac{50}{150} = \frac{1}{3}$$

2. Kemudian barulah hitung probabilitas setiap *term* dari semua dokumen dengan menggunakan persamaan 2.9. Jumlah seluruh term yang digunakan dalam penelitian ini adalah sebanyak 150, 50 term dari kategori topik berita hiburan, 50 term dari kategori topik berita kesehatan dan 50 term dari kategori topik berita olahraga. Banyaknya term berasal dari data yang telah melalui proses *preprocessing*. Perhitungan probabilitas dari setiap term adalah sebagai berikut, berdasarkan sampel yang ada pada Tabel 4.7.

- Probabilitas kata “pisau”

$$p(\text{"pisau"}|\text{"kesehatan"}) = \frac{(f(\text{"pisau"}|\text{"kesehatan"}) + 1)}{f(\text{"kesehatan"}) + |T|} = \frac{1 + 1}{50 + 150} = \frac{2}{200}$$

$$p(\text{"pisau"}|\text{"olahraga"}) = \frac{(f(\text{"pisau"}|\text{"olahraga"}) + 1)}{f(\text{"olahraga"}) + |T|} = \frac{0 + 1}{50 + 150} = \frac{1}{200}$$

$$p(\text{"pisau"}|\text{"hiburan"}) = \frac{(f(\text{"pisau"}|\text{"hiburan"}) + 1)}{f(\text{"hiburan"}) + |T|} = \frac{0 + 1}{50 + 150} = \frac{1}{200}$$

- Probabilitas kata “cukur”

$$p(\text{"cukur"}|\text{"kesehatan"}) = \frac{(f(\text{"cukur"}|\text{"kesehatan"}) + 1)}{f(\text{"kesehatan"}) + |T|} = \frac{1 + 1}{50 + 150} = \frac{2}{200}$$

$$p(\text{"cukur"}|\text{"olahraga"}) = \frac{(f(\text{"cukur"}|\text{"olahraga"}) + 1)}{f(\text{"olahraga"}) + |T|} = \frac{0 + 1}{50 + 150} = \frac{1}{200}$$



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber.

a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.

b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

$$p(\text{"cukur"}|\text{"hiburan"}) = \frac{f(\text{"cukur"}|\text{"hiburan"}) + 1}{f(\text{"hiburan"}) + |T|} = \frac{0 + 1}{50 + 150} = \frac{1}{200}$$

- Probabilitas kata “alat”

$$p(\text{"alat"}|\text{"kesehatan"}) = \frac{f(\text{"alat"}|\text{"kesehatan"}) + 1}{f(\text{"kesehatan"}) + |T|} = \frac{1 + 1}{50 + 150} = \frac{2}{200}$$

$$p(\text{"alat"}|\text{"olahraga"}) = \frac{f(\text{"alat"}|\text{"olahraga"}) + 1}{f(\text{"olahraga"}) + |T|} = \frac{0 + 1}{50 + 150} = \frac{1}{200}$$

$$p(\text{"alat"}|\text{"hiburan"}) = \frac{f(\text{"alat"}|\text{"hiburan"}) + 1}{f(\text{"hiburan"}) + |T|} = \frac{0 + 1}{50 + 150} = \frac{1}{200}$$

- Probabilitas kata “guna”

$$p(\text{"guna"}|\text{"kesehatan"}) = \frac{f(\text{"guna"}|\text{"kesehatan"}) + 1}{f(\text{"kesehatan"}) + |T|} = \frac{1 + 1}{50 + 150} = \frac{2}{200}$$

$$p(\text{"guna"}|\text{"olahraga"}) = \frac{f(\text{"guna"}|\text{"olahraga"}) + 1}{f(\text{"olahraga"}) + |T|} = \frac{0 + 1}{50 + 150} = \frac{1}{200}$$

$$p(\text{"guna"}|\text{"hiburan"}) = \frac{f(\text{"guna"}|\text{"hiburan"}) + 1}{f(\text{"hiburan"}) + |T|} = \frac{0 + 1}{50 + 150} = \frac{1}{200}$$

- Probabilitas kata “pangkas”

$$p(\text{"pangkas"}|\text{"kesehatan"}) = \frac{f(\text{"pangkas"}|\text{"kesehatan"}) + 1}{f(\text{"kesehatan"}) + |T|} = \frac{1 + 1}{50 + 150} = \frac{2}{200}$$

$$p(\text{"pangkas"}|\text{"olahraga"}) = \frac{f(\text{"pangkas"}|\text{"olahraga"}) + 1}{f(\text{"olahraga"}) + |T|} = \frac{0 + 1}{50 + 150} = \frac{1}{200}$$

$$p(\text{"pangkas"}|\text{"hiburan"}) = \frac{f(\text{"pangkas"}|\text{"hiburan"}) + 1}{f(\text{"hiburan"}) + |T|} = \frac{0 + 1}{50 + 150} = \frac{1}{200}$$

- Probabilitas kata “rambut”

$$p(\text{"rambut"}|\text{"kesehatan"}) = \frac{f(\text{"rambut"}|\text{"kesehatan"}) + 1}{f(\text{"kesehatan"}) + |T|} = \frac{2 + 1}{50 + 150} = \frac{3}{200}$$

$$p(\text{"rambut"}|\text{"olahraga"}) = \frac{f(\text{"rambut"}|\text{"olahraga"}) + 1}{f(\text{"olahraga"}) + |T|} = \frac{0 + 1}{50 + 150} = \frac{1}{200}$$


**Hak Cipta Dilindungi Undang-Undang**

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber.

a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.

b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

$$p(\text{"rambut"}|\text{"hiburan"}) = \frac{(f(\text{"rambut"}|\text{"hiburan"}) + 1)}{f(\text{"hiburan"}) + |T|} = \frac{0 + 1}{50 + 150} = \frac{1}{200}$$

- Probabilitas kata “tubuh”

$$p(\text{"tubuh"}|\text{"kesehatan"}) = \frac{(f(\text{"tubuh"}|\text{"kesehatan"}) + 1)}{f(\text{"kesehatan"}) + |T|} = \frac{1 + 1}{50 + 150} = \frac{2}{200}$$

$$p(\text{"tubuh"}|\text{"olahraga"}) = \frac{(f(\text{"tubuh"}|\text{"olahraga"}) + 1)}{f(\text{"olahraga"}) + |T|} = \frac{0 + 1}{50 + 150} = \frac{1}{200}$$

$$p(\text{"tubuh"}|\text{"hiburan"}) = \frac{(f(\text{"tubuh"}|\text{"hiburan"}) + 1)}{f(\text{"hiburan"}) + |T|} = \frac{0 + 1}{50 + 150} = \frac{1}{200}$$

- Probabilitas kata “kemaluan”

$$p(\text{"kemaluan"}|\text{"kesehatan"}) = \frac{(f(\text{"kemaluan"}|\text{"kesehatan"}) + 1)}{f(\text{"kesehatan"}) + |T|} = \frac{2 + 1}{50 + 150} = \frac{3}{200}$$

$$p(\text{"kemaluan"}|\text{"olahraga"}) = \frac{(f(\text{"kemaluan"}|\text{"olahraga"}) + 1)}{f(\text{"olahraga"}) + |T|} = \frac{0 + 1}{50 + 150} = \frac{1}{200}$$

$$p(\text{"kemaluan"}|\text{"hiburan"}) = \frac{(f(\text{"kemaluan"}|\text{"hiburan"}) + 1)}{f(\text{"hiburan"}) + |T|} = \frac{0 + 1}{50 + 150} = \frac{1}{200}$$

- Probabilitas kata “rapi”

$$p(\text{"rapi"}|\text{"kesehatan"}) = \frac{(f(\text{"rapi"}|\text{"kesehatan"}) + 1)}{f(\text{"kesehatan"}) + |T|} = \frac{1 + 1}{50 + 150} = \frac{2}{200}$$

$$p(\text{"rapi"}|\text{"olahraga"}) = \frac{(f(\text{"rapi"}|\text{"olahraga"}) + 1)}{f(\text{"olahraga"}) + |T|} = \frac{0 + 1}{50 + 150} = \frac{1}{200}$$

$$p(\text{"rapi"}|\text{"hiburan"}) = \frac{(f(\text{"rapi"}|\text{"hiburan"}) + 1)}{f(\text{"hiburan"}) + |T|} = \frac{0 + 1}{50 + 150} = \frac{1}{200}$$

- Probabilitas kata “lebat”

$$p(\text{"lebat"}|\text{"kesehatan"}) = \frac{(f(\text{"lebat"}|\text{"kesehatan"}) + 1)}{f(\text{"kesehatan"}) + |T|} = \frac{2 + 1}{50 + 150} = \frac{3}{200}$$



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.

b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

$$p(\text{"lebat"}|\text{"olahraga"}) = \frac{(f(\text{"lebat"}|\text{"olahraga"}) + 1)}{f(\text{"olahraga"}) + |T|} = \frac{0 + 1}{50 + 150} = \frac{1}{200}$$

$$p(\text{"lebat"}|\text{"hiburan"}) = \frac{(f(\text{"lebat"}|\text{"hiburan"}) + 1)}{f(\text{"hiburan"}) + |T|} = \frac{0 + 1}{50 + 150} = \frac{1}{200}$$

• Probabilitas kata “rawat”

$$p(\text{"rawat"}|\text{"kesehatan"}) = \frac{(f(\text{"rawat"}|\text{"kesehatan"}) + 1)}{f(\text{"kesehatan"}) + |T|} = \frac{1 + 1}{50 + 150} = \frac{2}{200}$$

$$p(\text{"rawat"}|\text{"olahraga"}) = \frac{(f(\text{"rawat"}|\text{"olahraga"}) + 1)}{f(\text{"olahraga"}) + |T|} = \frac{0 + 1}{50 + 150} = \frac{1}{200}$$

$$p(\text{"rawat"}|\text{"hiburan"}) = \frac{(f(\text{"rawat"}|\text{"hiburan"}) + 1)}{f(\text{"hiburan"}) + |T|} = \frac{0 + 1}{50 + 150} = \frac{1}{200}$$

• Probabilitas kata “tempat”

$$p(\text{"tempat"}|\text{"kesehatan"}) = \frac{(f(\text{"tempat"}|\text{"kesehatan"}) + 1)}{f(\text{"kesehatan"}) + |T|} = \frac{1 + 1}{50 + 150} = \frac{2}{200}$$

$$p(\text{"tempat"}|\text{"olahraga"}) = \frac{(f(\text{"tempat"}|\text{"olahraga"}) + 1)}{f(\text{"olahraga"}) + |T|} = \frac{0 + 1}{50 + 150} = \frac{1}{200}$$

$$p(\text{"tempat"}|\text{"hiburan"}) = \frac{(f(\text{"tempat"}|\text{"hiburan"}) + 1)}{f(\text{"hiburan"}) + |T|} = \frac{0 + 1}{50 + 150} = \frac{1}{200}$$

• Probabilitas kata “kembang”

$$p(\text{"kembang"}|\text{"kesehatan"}) = \frac{(f(\text{"kembang"}|\text{"kesehatan"}) + 1)}{f(\text{"kesehatan"}) + |T|} = \frac{1 + 1}{50 + 150} = \frac{2}{200}$$

$$p(\text{"kembang"}|\text{"olahraga"}) = \frac{(f(\text{"kembang"}|\text{"olahraga"}) + 1)}{f(\text{"olahraga"}) + |T|} = \frac{0 + 1}{50 + 150} = \frac{1}{200}$$

$$p(\text{"kembang"}|\text{"hiburan"}) = \frac{(f(\text{"kembang"}|\text{"hiburan"}) + 1)}{f(\text{"hiburan"}) + |T|} = \frac{0 + 1}{50 + 150} = \frac{1}{200}$$

• Probabilitas kata “kuman”



## Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber.

a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.

b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumunkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

$$p("kuman"|"kesehatan") = \frac{(f("kuman"|"kesehatan") + 1)}{f("kesehatan") + |T|} = \frac{1 + 1}{50 + 150} = \frac{2}{200}$$

$$p("kuman"|"olahraga") = \frac{(f("kuman"|"olahraga") + 1)}{f("olahraga") + |T|} = \frac{0 + 1}{50 + 150} = \frac{1}{200}$$

$$p("kuman"|"hiburan") = \frac{(f("kuman"|"hiburan") + 1)}{f("hiburan") + |T|} = \frac{0 + 1}{50 + 150} = \frac{1}{200}$$

- Probabilitas kata “jeda”

$$p("jeda"|"kesehatan") = \frac{(f("jeda"|"kesehatan") + 1)}{f("kesehatan") + |T|} = \frac{1 + 1}{50 + 150} = \frac{2}{200}$$

$$p("jeda"|"olahraga") = \frac{(f("jeda"|"olahraga") + 1)}{f("olahraga") + |T|} = \frac{0 + 1}{50 + 150} = \frac{1}{200}$$

$$p("jeda"|"hiburan") = \frac{(f("jeda"|"hiburan") + 1)}{f("hiburan") + |T|} = \frac{0 + 1}{50 + 150} = \frac{1}{200}$$

- Probabilitas kata “nyaris”

$$p("nyaris"|"kesehatan") = \frac{(f("nyaris"|"kesehatan") + 1)}{f("kesehatan") + |T|} = \frac{1 + 1}{50 + 150} = \frac{2}{200}$$

$$p("nyaris"|"olahraga") = \frac{(f("nyaris"|"olahraga") + 1)}{f("olahraga") + |T|} = \frac{0 + 1}{50 + 150} = \frac{1}{200}$$

$$p("nyaris"|"hiburan") = \frac{(f("nyaris"|"hiburan") + 1)}{f("hiburan") + |T|} = \frac{0 + 1}{50 + 150} = \frac{1}{200}$$

- Probabilitas kata “bulan”

$$p("bulan"|"kesehatan") = \frac{(f("bulan"|"kesehatan") + 1)}{f("kesehatan") + |T|} = \frac{1 + 1}{50 + 150} = \frac{2}{200}$$

$$p("bulan"|"olahraga") = \frac{(f("bulan"|"olahraga") + 1)}{f("olahraga") + |T|} = \frac{0 + 1}{50 + 150} = \frac{1}{200}$$

$$p("bulan"|"hiburan") = \frac{(f("bulan"|"hiburan") + 1)}{f("hiburan") + |T|} = \frac{0 + 1}{50 + 150} = \frac{1}{200}$$



**Hak Cipta Dilindungi Undang-Undang**

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

• Probabilitas kata “gelar”

$$p(\text{"gelar"}|\text{"kesehatan"}) = \frac{(f(\text{"gelar"}|\text{"kesehatan"}) + 1)}{f(\text{"kesehatan"}) + |T|} = \frac{1 + 1}{50 + 150} = \frac{2}{200}$$

$$p(\text{"gelar"}|\text{"olahraga"}) = \frac{(f(\text{"gelar"}|\text{"olahraga"}) + 1)}{f(\text{"olahraga"}) + |T|} = \frac{0 + 1}{50 + 150} = \frac{1}{200}$$

$$p(\text{"gelar"}|\text{"hiburan"}) = \frac{(f(\text{"gelar"}|\text{"hiburan"}) + 1)}{f(\text{"hiburan"}) + |T|} = \frac{0 + 1}{50 + 150} = \frac{1}{200}$$

• Probabilitas kata “F1”

$$p(\text{"F1"}|\text{"kesehatan"}) = \frac{(f(\text{"F1"}|\text{"kesehatan"}) + 1)}{f(\text{"kesehatan"}) + |T|} = \frac{1 + 1}{50 + 150} = \frac{2}{200}$$

$$p(\text{"F1"}|\text{"olahraga"}) = \frac{(f(\text{"F1"}|\text{"olahraga"}) + 1)}{f(\text{"olahraga"}) + |T|} = \frac{0 + 1}{50 + 150} = \frac{1}{200}$$

$$p(\text{"F1"}|\text{"hiburan"}) = \frac{(f(\text{"F1"}|\text{"hiburan"}) + 1)}{f(\text{"hiburan"}) + |T|} = \frac{0 + 1}{50 + 150} = \frac{1}{200}$$

• Probabilitas kata “gulir”

$$p(\text{"gulir"}|\text{"kesehatan"}) = \frac{(f(\text{"gulir"}|\text{"kesehatan"}) + 1)}{f(\text{"kesehatan"}) + |T|} = \frac{1 + 1}{50 + 150} = \frac{2}{200}$$

$$p(\text{"gulir"}|\text{"olahraga"}) = \frac{(f(\text{"gulir"}|\text{"olahraga"}) + 1)}{f(\text{"olahraga"}) + |T|} = \frac{0 + 1}{50 + 150} = \frac{1}{200}$$

$$p(\text{"gulir"}|\text{"hiburan"}) = \frac{(f(\text{"gulir"}|\text{"hiburan"}) + 1)}{f(\text{"hiburan"}) + |T|} = \frac{0 + 1}{50 + 150} = \frac{1}{200}$$

• Probabilitas kata “dispa”

$$p(\text{"dispa"}|\text{"kesehatan"}) = \frac{(f(\text{"dispa"}|\text{"kesehatan"}) + 1)}{f(\text{"kesehatan"}) + |T|} = \frac{1 + 1}{50 + 150} = \frac{2}{200}$$

$$p(\text{"dispa"}|\text{"olahraga"}) = \frac{(f(\text{"dispa"}|\text{"olahraga"}) + 1)}{f(\text{"olahraga"}) + |T|} = \frac{0 + 1}{50 + 150} = \frac{1}{200}$$



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.

b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

$$p("dispa"|"hiburan") = \frac{(f("dispa"|"hiburan") + 1)}{f("hiburan") + |T|} = \frac{0 + 1}{50 + 150} = \frac{1}{200}$$

- Probabilitas kata “belgia”

$$p("belgia"|"kesehatan") = \frac{(f("belgia"|"kesehatan") + 1)}{f("kesehatan") + |T|} = \frac{1 + 1}{50 + 150} = \frac{2}{200}$$

$$p("belgia"|"olahraga") = \frac{(f("belgia"|"olahraga") + 1)}{f("olahraga") + |T|} = \frac{0 + 1}{50 + 150} = \frac{1}{200}$$

$$p("belgia"|"hiburan") = \frac{(f("belgia"|"hiburan") + 1)}{f("hiburan") + |T|} = \frac{0 + 1}{50 + 150} = \frac{1}{200}$$

- Probabilitas kata “jumlah”

$$p("jumlah"|"kesehatan") = \frac{(f("jumlah"|"kesehatan") + 1)}{f("kesehatan") + |T|} = \frac{1 + 1}{50 + 150} = \frac{2}{200}$$

$$p("jumlah"|"olahraga") = \frac{(f("jumlah"|"olahraga") + 1)}{f("olahraga") + |T|} = \frac{0 + 1}{50 + 150} = \frac{1}{200}$$

$$p("jumlah"|"hiburan") = \frac{(f("jumlah"|"hiburan") + 1)}{f("hiburan") + |T|} = \frac{0 + 1}{50 + 150} = \frac{1}{200}$$

- Probabilitas kata “catat”

$$p("catat"|"kesehatan") = \frac{(f("catat"|"kesehatan") + 1)}{f("kesehatan") + |T|} = \frac{1 + 1}{50 + 150} = \frac{2}{200}$$

$$p("catat"|"olahraga") = \frac{(f("catat"|"olahraga") + 1)}{f("olahraga") + |T|} = \frac{0 + 1}{50 + 150} = \frac{1}{200}$$

$$p("catat"|"hiburan") = \frac{(f("catat"|"hiburan") + 1)}{f("hiburan") + |T|} = \frac{0 + 1}{50 + 150} = \frac{1}{200}$$

- Probabilitas kata “kait”

$$p("kait"|"kesehatan") = \frac{(f("kait"|"kesehatan") + 1)}{f("kesehatan") + |T|} = \frac{1 + 1}{50 + 150} = \frac{2}{200}$$



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.

b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

$$p("kait"|"olahraga") = \frac{(f("kait"|"olahraga") + 1)}{f("olahraga") + |T|} = \frac{0 + 1}{50 + 150} = \frac{1}{200}$$

$$p("kait"|"hiburan") = \frac{(f("kait"|"hiburan") + 1)}{f("hiburan") + |T|} = \frac{0 + 1}{50 + 150} = \frac{1}{200}$$

- Probabilitas kata “balap”

$$p("cukur"|"kesehatan") = \frac{(f("cukur"|"kesehatan") + 1)}{f("kesehatan") + |T|} = \frac{1 + 1}{50 + 150} = \frac{2}{200}$$

$$p("cukur"|"olahraga") = \frac{(f("cukur"|"olahraga") + 1)}{f("olahraga") + |T|} = \frac{0 + 1}{50 + 150} = \frac{1}{200}$$

$$p("cukur"|"hiburan") = \frac{(f("cukur"|"hiburan") + 1)}{f("hiburan") + |T|} = \frac{0 + 1}{50 + 150} = \frac{1}{200}$$

- Probabilitas kata “masuk”

$$p("cukur"|"kesehatan") = \frac{(f("cukur"|"kesehatan") + 1)}{f("kesehatan") + |T|} = \frac{1 + 1}{50 + 150} = \frac{2}{200}$$

$$p("cukur"|"olahraga") = \frac{(f("cukur"|"olahraga") + 1)}{f("olahraga") + |T|} = \frac{0 + 1}{50 + 150} = \frac{1}{200}$$

$$p("cukur"|"hiburan") = \frac{(f("cukur"|"hiburan") + 1)}{f("hiburan") + |T|} = \frac{0 + 1}{50 + 150} = \frac{1}{200}$$

- Probabilitas kata “torehan”

$$p("parah"|"iya") = \frac{(f("parah"|"iya") + 1)}{f("iya") + |T|} = \frac{9 + 1}{1665 + 3605} = \frac{10}{5270}$$

$$p("parah"|"tidak") = \frac{(f("parah"|"tidak") + 1)}{f("tidak") + |T|} = \frac{0 + 1}{2545 + 3605} = \frac{1}{6150}$$

$$p("pisau"|"hiburan") = \frac{(f("masuk"|"hiburan") + 1)}{f("hiburan") + |T|} = \frac{0 + 1}{50 + 150} = \frac{1}{200}$$

- Probabilitas kata “kabar”



## Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.

b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumunkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

$$p("kabar"|"kesehatan") = \frac{(f("kabar"|"kesehatan") + 1)}{f("kesehatan") + |T|} = \frac{1 + 1}{50 + 150} = \frac{2}{200}$$

$$p("kabar"|"olahraga") = \frac{(f("kabar"|"olahraga") + 1)}{f("olahraga") + |T|} = \frac{0 + 1}{50 + 150} = \frac{1}{200}$$

$$p("kabar"|"hiburan") = \frac{(f("kabar"|"hiburan") + 1)}{f("hiburan") + |T|} = \frac{0 + 1}{50 + 150} = \frac{1}{200}$$

- Probabilitas kata “tengah”

$$p("tengah"|"kesehatan") = \frac{(f("tengah"|"kesehatan") + 1)}{f("kesehatan") + |T|} = \frac{1 + 1}{50 + 150} = \frac{2}{200}$$

$$p("tengah"|"olahraga") = \frac{(f("tengah"|"olahraga") + 1)}{f("olahraga") + |T|} = \frac{0 + 1}{50 + 150} = \frac{1}{200}$$

$$p("tengah"|"hiburan") = \frac{(f("tengah"|"hiburan") + 1)}{f("hiburan") + |T|} = \frac{0 + 1}{50 + 150} = \frac{1}{200}$$

- Probabilitas kata “seteru”

$$p("seteru"|"kesehatan") = \frac{(f("seteru"|"kesehatan") + 1)}{f("kesehatan") + |T|} = \frac{1 + 1}{50 + 150} = \frac{2}{200}$$

$$p("seteru"|"olahraga") = \frac{(f("seteru"|"olahraga") + 1)}{f("olahraga") + |T|} = \frac{0 + 1}{50 + 150} = \frac{1}{200}$$

$$p("seteru"|"hiburan") = \frac{(f("seteru"|"hiburan") + 1)}{f("hiburan") + |T|} = \frac{0 + 1}{50 + 150} = \frac{1}{200}$$

- Probabilitas kata “sindir”

$$p("sindir"|"kesehatan") = \frac{(f("sindir"|"kesehatan") + 1)}{f("kesehatan") + |T|} = \frac{1 + 1}{50 + 150} = \frac{2}{200}$$

$$p("sindir"|"olahraga") = \frac{(f("sindir"|"olahraga") + 1)}{f("olahraga") + |T|} = \frac{0 + 1}{50 + 150} = \frac{1}{200}$$

$$p("sindir"|"hiburan") = \frac{(f("sindir"|"hiburan") + 1)}{f("hiburan") + |T|} = \frac{0 + 1}{50 + 150} = \frac{1}{200}$$



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber.

a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.

b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

- Probabilitas kata “media”

$$p("media"|"kesehatan") = \frac{(f("media"|"kesehatan") + 1)}{f("kesehatan") + |T|} = \frac{1 + 1}{50 + 150} = \frac{2}{200}$$

$$p("media"|"olahraga") = \frac{(f("media"|"olahraga") + 1)}{f("olahraga") + |T|} = \frac{0 + 1}{50 + 150} = \frac{1}{200}$$

$$p("media"|"hiburan") = \frac{(f("media"|"hiburan") + 1)}{f("hiburan") + |T|} = \frac{0 + 1}{50 + 150} = \frac{1}{200}$$

- Probabilitas kata “sosial”

$$p("sosial"|"kesehatan") = \frac{(f("sosial"|"kesehatan") + 1)}{f("kesehatan") + |T|} = \frac{1 + 1}{50 + 150} = \frac{2}{200}$$

$$p("sosial"|"olahraga") = \frac{(f("sosial"|"olahraga") + 1)}{f("olahraga") + |T|} = \frac{0 + 1}{50 + 150} = \frac{1}{200}$$

$$p("sosial"|"hiburan") = \frac{(f("sosial"|"hiburan") + 1)}{f("hiburan") + |T|} = \frac{0 + 1}{50 + 150} = \frac{1}{200}$$

- Probabilitas kata “drama”

$$p("drama"|"kesehatan") = \frac{(f("drama"|"kesehatan") + 1)}{f("kesehatan") + |T|} = \frac{1 + 1}{50 + 150} = \frac{2}{200}$$

$$p("drama"|"olahraga") = \frac{(f("drama"|"olahraga") + 1)}{f("olahraga") + |T|} = \frac{0 + 1}{50 + 150} = \frac{1}{200}$$

$$p("drama"|"hiburan") = \frac{(f("drama"|"hiburan") + 1)}{f("hiburan") + |T|} = \frac{0 + 1}{50 + 150} = \frac{1}{200}$$

- Probabilitas kata “mulai”

$$p("mulai"|"kesehatan") = \frac{(f("mulai"|"kesehatan") + 1)}{f("kesehatan") + |T|} = \frac{1 + 1}{50 + 150} = \frac{2}{200}$$

$$p("mulai"|"olahraga") = \frac{(f("mulai"|"olahraga") + 1)}{f("olahraga") + |T|} = \frac{0 + 1}{50 + 150} = \frac{1}{200}$$

$$p("mulai"|"hiburan") = \frac{(f("mulai"|"hiburan") + 1)}{f("hiburan") + |T|} = \frac{0 + 1}{50 + 150} = \frac{1}{200}$$



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

- Probabilitas kata “video”

$$p(\text{"video"}|\text{"kesehatan"}) = \frac{(f(\text{"video"}|\text{"kesehatan"}) + 1)}{f(\text{"kesehatan"}) + |T|} = \frac{1 + 1}{50 + 150} = \frac{2}{200}$$

$$p(\text{"video"}|\text{"olahraga"}) = \frac{(f(\text{"video"}|\text{"olahraga"}) + 1)}{f(\text{"olahraga"}) + |T|} = \frac{0 + 1}{50 + 150} = \frac{1}{200}$$

$$p(\text{"video"}|\text{"hiburan"}) = \frac{(f(\text{"video"}|\text{"hiburan"}) + 1)}{f(\text{"hiburan"}) + |T|} = \frac{0 + 1}{50 + 150} = \frac{1}{200}$$

- Probabilitas kata “kritik”

$$p(\text{"kritik"}|\text{"kesehatan"}) = \frac{(f(\text{"kritik"}|\text{"kesehatan"}) + 1)}{f(\text{"kesehatan"}) + |T|} = \frac{1 + 1}{50 + 150} = \frac{2}{200}$$

$$p(\text{"kritik"}|\text{"olahraga"}) = \frac{(f(\text{"kritik"}|\text{"olahraga"}) + 1)}{f(\text{"olahraga"}) + |T|} = \frac{0 + 1}{50 + 150} = \frac{1}{200}$$

$$p(\text{"kritik"}|\text{"hiburan"}) = \frac{(f(\text{"kritik"}|\text{"hiburan"}) + 1)}{f(\text{"hiburan"}) + |T|} = \frac{0 + 1}{50 + 150} = \frac{1}{200}$$

- Probabilitas kata “lawan”

$$p(\text{"lawan"}|\text{"kesehatan"}) = \frac{(f(\text{"lawan"}|\text{"kesehatan"}) + 1)}{f(\text{"kesehatan"}) + |T|} = \frac{1 + 1}{50 + 150} = \frac{2}{200}$$

$$p(\text{"lawan"}|\text{"olahraga"}) = \frac{(f(\text{"lawan"}|\text{"olahraga"}) + 1)}{f(\text{"olahraga"}) + |T|} = \frac{0 + 1}{50 + 150} = \frac{1}{200}$$

$$p(\text{"lawan"}|\text{"hiburan"}) = \frac{(f(\text{"lawan"}|\text{"hiburan"}) + 1)}{f(\text{"hiburan"}) + |T|} = \frac{0 + 1}{50 + 150} = \frac{1}{200}$$

- Probabilitas kata “main”

$$p(\text{"main"}|\text{"kesehatan"}) = \frac{(f(\text{"main"}|\text{"kesehatan"}) + 1)}{f(\text{"kesehatan"}) + |T|} = \frac{1 + 1}{50 + 150} = \frac{2}{200}$$

$$p(\text{"main"}|\text{"olahraga"}) = \frac{(f(\text{"main"}|\text{"olahraga"}) + 1)}{f(\text{"olahraga"}) + |T|} = \frac{0 + 1}{50 + 150} = \frac{1}{200}$$

$$p(\text{"main"}|\text{"hiburan"}) = \frac{(f(\text{"main"}|\text{"hiburan"}) + 1)}{f(\text{"hiburan"}) + |T|} = \frac{0 + 1}{50 + 150} = \frac{1}{200}$$



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

- Probabilitas kata “film”

$$p(\text{"film"}|\text{"kesehatan"}) = \frac{(f(\text{"film"}|\text{"kesehatan"}) + 1)}{f(\text{"kesehatan"}) + |T|} = \frac{1 + 1}{50 + 150} = \frac{2}{200}$$

$$p(\text{"film"}|\text{"olahraga"}) = \frac{(f(\text{"film"}|\text{"olahraga"}) + 1)}{f(\text{"olahraga"}) + |T|} = \frac{0 + 1}{50 + 150} = \frac{1}{200}$$

$$p(\text{"film"}|\text{"hiburan"}) = \frac{(f(\text{"film"}|\text{"hiburan"}) + 1)}{f(\text{"hiburan"}) + |T|} = \frac{0 + 1}{50 + 150} = \frac{1}{200}$$

- Probabilitas kata “bulan”

$$p(\text{"bulan"}|\text{"kesehatan"}) = \frac{(f(\text{"bulan"}|\text{"kesehatan"}) + 1)}{f(\text{"kesehatan"}) + |T|} = \frac{1 + 1}{50 + 150} = \frac{2}{200}$$

$$p(\text{"bulan"}|\text{"olahraga"}) = \frac{(f(\text{"bulan"}|\text{"olahraga"}) + 1)}{f(\text{"olahraga"}) + |T|} = \frac{0 + 1}{50 + 150} = \frac{1}{200}$$

$$p(\text{"bulan"}|\text{"hiburan"}) = \frac{(f(\text{"bulan"}|\text{"hiburan"}) + 1)}{f(\text{"hiburan"}) + |T|} = \frac{0 + 1}{50 + 150} = \frac{1}{200}$$

- Probabilitas kata “publik”

$$p(\text{"publik"}|\text{"kesehatan"}) = \frac{(f(\text{"publik"}|\text{"kesehatan"}) + 1)}{f(\text{"kesehatan"}) + |T|} = \frac{1 + 1}{50 + 150} = \frac{2}{200}$$

$$p(\text{"publik"}|\text{"olahraga"}) = \frac{(f(\text{"publik"}|\text{"olahraga"}) + 1)}{f(\text{"olahraga"}) + |T|} = \frac{0 + 1}{50 + 150} = \frac{1}{200}$$

$$p(\text{"publik"}|\text{"hiburan"}) = \frac{(f(\text{"publik"}|\text{"hiburan"}) + 1)}{f(\text{"hiburan"}) + |T|} = \frac{0 + 1}{50 + 150} = \frac{1}{200}$$

- Probabilitas kata “heboh”

$$p(\text{"heboh"}|\text{"kesehatan"}) = \frac{(f(\text{"heboh"}|\text{"kesehatan"}) + 1)}{f(\text{"kesehatan"}) + |T|} = \frac{1 + 1}{50 + 150} = \frac{2}{200}$$

$$p(\text{"heboh"}|\text{"olahraga"}) = \frac{(f(\text{"heboh"}|\text{"olahraga"}) + 1)}{f(\text{"olahraga"}) + |T|} = \frac{0 + 1}{50 + 150} = \frac{1}{200}$$


**Hak Cipta Dilindungi Undang-Undang**

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber.

a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.

b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

$$p("heboh"|"hiburan") = \frac{(f("heboh"|"hiburan") + 1)}{f("hiburan") + |T|} = \frac{0 + 1}{50 + 150} = \frac{1}{200}$$

- Probabilitas kata “keras”

$$p("keras"|"kesehatan") = \frac{(f("keras"|"kesehatan") + 1)}{f("kesehatan") + |T|} = \frac{1 + 1}{50 + 150} = \frac{2}{200}$$

$$p("keras"|"olahraga") = \frac{(f("keras"|"olahraga") + 1)}{f("olahraga") + |T|} = \frac{0 + 1}{50 + 150} = \frac{1}{200}$$

$$p("keras"|"hiburan") = \frac{(f("keras"|"hiburan") + 1)}{f("hiburan") + |T|} = \frac{0 + 1}{50 + 150} = \frac{1}{200}$$

- Probabilitas kata “laku”

$$p("laku"|"kesehatan") = \frac{(f("laku"|"kesehatan") + 1)}{f("kesehatan") + |T|} = \frac{1 + 1}{50 + 150} = \frac{2}{200}$$

$$p("laku"|"olahraga") = \frac{(f("laku"|"olahraga") + 1)}{f("olahraga") + |T|} = \frac{0 + 1}{50 + 150} = \frac{1}{200}$$

$$p("laku"|"hiburan") = \frac{(f("laku"|"hiburan") + 1)}{f("hiburan") + |T|} = \frac{0 + 1}{50 + 150} = \frac{1}{200}$$

- Probabilitas kata “sang”

$$p("sang"|"kesehatan") = \frac{(f("sang"|"kesehatan") + 1)}{f("kesehatan") + |T|} = \frac{1 + 1}{50 + 150} = \frac{2}{200}$$

$$p("sang"|"olahraga") = \frac{(f("sang"|"olahraga") + 1)}{f("olahraga") + |T|} = \frac{0 + 1}{50 + 150} = \frac{1}{200}$$

$$p("sang"|"hiburan") = \frac{(f("sang"|"hiburan") + 1)}{f("hiburan") + |T|} = \frac{0 + 1}{50 + 150} = \frac{1}{200}$$

- Probabilitas kata “tunangan”

$$p("tunangan"|"kesehatan") = \frac{(f("tunangan"|"kesehatan") + 1)}{f("kesehatan") + |T|} = \frac{1 + 1}{50 + 150} = \frac{2}{200}$$

$$p("tunangan"|"olahraga") = \frac{(f("tunangan"|"olahraga") + 1)}{f("olahraga") + |T|} = \frac{0 + 1}{50 + 150} = \frac{1}{200}$$


**Hak Cipta Dilindungi Undang-Undang**

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber.

a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.

b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

$$p(\text{"tunangan"}|\text{"hiburan"}) = \frac{(f(\text{"tunangan"}|\text{"hiburan"}) + 1)}{f(\text{"hiburan"}) + |T|} = \frac{0 + 1}{50 + 150} = \frac{1}{200}$$

- Probabilitas kata “artis”

$$p(\text{"artis"}|\text{"kesehatan"}) = \frac{(f(\text{"artis"}|\text{"kesehatan"}) + 1)}{f(\text{"kesehatan"}) + |T|} = \frac{1 + 1}{50 + 150} = \frac{2}{200}$$

$$p(\text{"artis"}|\text{"olahraga"}) = \frac{(f(\text{"artis"}|\text{"olahraga"}) + 1)}{f(\text{"olahraga"}) + |T|} = \frac{0 + 1}{50 + 150} = \frac{1}{200}$$

$$p(\text{"artis"}|\text{"hiburan"}) = \frac{(f(\text{"artis"}|\text{"hiburan"}) + 1)}{f(\text{"hiburan"}) + |T|} = \frac{0 + 1}{50 + 150} = \frac{1}{200}$$

- Probabilitas kata “pamer”

$$p(\text{"pamer"}|\text{"kesehatan"}) = \frac{(f(\text{"pamer"}|\text{"kesehatan"}) + 1)}{f(\text{"kesehatan"}) + |T|} = \frac{1 + 1}{50 + 150} = \frac{2}{200}$$

$$p(\text{"pamer"}|\text{"olahraga"}) = \frac{(f(\text{"pamer"}|\text{"olahraga"}) + 1)}{f(\text{"olahraga"}) + |T|} = \frac{0 + 1}{50 + 150} = \frac{1}{200}$$

$$p(\text{"pamer"}|\text{"hiburan"}) = \frac{(f(\text{"pamer"}|\text{"hiburan"}) + 1)}{f(\text{"hiburan"}) + |T|} = \frac{0 + 1}{50 + 150} = \frac{1}{200}$$

- Probabilitas kata “cincin”

$$p(\text{"cincin"}|\text{"kesehatan"}) = \frac{(f(\text{"cincin"}|\text{"kesehatan"}) + 1)}{f(\text{"kesehatan"}) + |T|} = \frac{1 + 1}{50 + 150} = \frac{2}{200}$$

$$p(\text{"cincin"}|\text{"olahraga"}) = \frac{(f(\text{"cincin"}|\text{"olahraga"}) + 1)}{f(\text{"olahraga"}) + |T|} = \frac{0 + 1}{50 + 150} = \frac{1}{200}$$

$$p(\text{"cincin"}|\text{"hiburan"}) = \frac{(f(\text{"cincin"}|\text{"hiburan"}) + 1)}{f(\text{"hiburan"}) + |T|} = \frac{0 + 1}{50 + 150} = \frac{1}{200}$$

- Probabilitas kata “nyangkal”

$$p(\text{"nyangkal"}|\text{"kesehatan"}) = \frac{(f(\text{"nyangkal"}|\text{"kesehatan"}) + 1)}{f(\text{"kesehatan"}) + |T|} = \frac{1 + 1}{50 + 150} = \frac{2}{200}$$

$$p(\text{"nyangkal"}|\text{"olahraga"}) = \frac{(f(\text{"nyangkal"}|\text{"olahraga"}) + 1)}{f(\text{"olahraga"}) + |T|} = \frac{0 + 1}{50 + 150} = \frac{1}{200}$$



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.

b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

$$p(\text{"nyangkal"}|\text{"hiburan"}) = \frac{f(\text{"nyangkal"}|\text{"hiburan"}) + 1}{f(\text{"hiburan"}) + |T|} = \frac{0 + 1}{50 + 150} = \frac{1}{200}$$

• Probabilitas kata “juara”

$$p(\text{"juara"}|\text{"kesehatan"}) = \frac{f(\text{"juara"}|\text{"kesehatan"}) + 1}{f(\text{"kesehatan"}) + |T|} = \frac{1 + 1}{50 + 150} = \frac{2}{200}$$

$$p(\text{"juara"}|\text{"olahraga"}) = \frac{f(\text{"juara"}|\text{"olahraga"}) + 1}{f(\text{"olahraga"}) + |T|} = \frac{0 + 1}{50 + 150} = \frac{1}{200}$$

$$p(\text{"juara"}|\text{"hiburan"}) = \frac{f(\text{"juara"}|\text{"hiburan"}) + 1}{f(\text{"hiburan"}) + |T|} = \frac{0 + 1}{50 + 150} = \frac{1}{200}$$

• Probabilitas kata “fokus”

$$p(\text{"fokus"}|\text{"kesehatan"}) = \frac{f(\text{"fokus"}|\text{"kesehatan"}) + 1}{f(\text{"kesehatan"}) + |T|} = \frac{1 + 1}{50 + 150} = \frac{2}{200}$$

$$p(\text{"fokus"}|\text{"olahraga"}) = \frac{f(\text{"fokus"}|\text{"olahraga"}) + 1}{f(\text{"olahraga"}) + |T|} = \frac{0 + 1}{50 + 150} = \frac{1}{200}$$

$$p(\text{"fokus"}|\text{"hiburan"}) = \frac{f(\text{"fokus"}|\text{"hiburan"}) + 1}{f(\text{"hiburan"}) + |T|} = \frac{0 + 1}{50 + 150} = \frac{1}{200}$$

• Probabilitas kata “rebut”

$$p(\text{"rebut"}|\text{"kesehatan"}) = \frac{f(\text{"rebut"}|\text{"kesehatan"}) + 1}{f(\text{"kesehatan"}) + |T|} = \frac{1 + 1}{50 + 150} = \frac{2}{200}$$

$$p(\text{"rebut"}|\text{"olahraga"}) = \frac{f(\text{"rebut"}|\text{"olahraga"}) + 1}{f(\text{"olahraga"}) + |T|} = \frac{0 + 1}{50 + 150} = \frac{1}{200}$$

$$p(\text{"rebut"}|\text{"hiburan"}) = \frac{f(\text{"rebut"}|\text{"hiburan"}) + 1}{f(\text{"hiburan"}) + |T|} = \frac{0 + 1}{50 + 150} = \frac{1}{200}$$

• Probabilitas kata “posisi”

$$p(\text{"posisi"}|\text{"kesehatan"}) = \frac{f(\text{"posisi"}|\text{"kesehatan"}) + 1}{f(\text{"kesehatan"}) + |T|} = \frac{1 + 1}{50 + 150} = \frac{2}{200}$$

$$p(\text{"posisi"}|\text{"olahraga"}) = \frac{f(\text{"posisi"}|\text{"olahraga"}) + 1}{f(\text{"olahraga"}) + |T|} = \frac{0 + 1}{50 + 150} = \frac{1}{200}$$



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.

b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

$$p("posisi"|"hiburan") = \frac{(f("posisi"|"hiburan") + 1)}{f("hiburan") + |T|} = \frac{0 + 1}{50 + 150} = \frac{1}{200}$$

- Probabilitas kata “dua”

$$p("dua"|"kesehatan") = \frac{(f("dua"|"kesehatan") + 1)}{f("kesehatan") + |T|} = \frac{1 + 1}{50 + 150} = \frac{2}{200}$$

$$p("dua"|"olahraga") = \frac{(f("dua"|"olahraga") + 1)}{f("olahraga") + |T|} = \frac{0 + 1}{50 + 150} = \frac{1}{200}$$

$$p("dua"|"hiburan") = \frac{(f("dua"|"hiburan") + 1)}{f("hiburan") + |T|} = \frac{0 + 1}{50 + 150} = \frac{1}{200}$$

- Probabilitas kata “serang”

$$p("serang"|"kesehatan") = \frac{(f("serang"|"kesehatan") + 1)}{f("kesehatan") + |T|} = \frac{1 + 1}{50 + 150} = \frac{2}{200}$$

$$p("serang"|"olahraga") = \frac{(f("serang"|"olahraga") + 1)}{f("olahraga") + |T|} = \frac{0 + 1}{50 + 150} = \frac{1}{200}$$

$$p("serang"|"hiburan") = \frac{(f("serang"|"hiburan") + 1)}{f("hiburan") + |T|} = \frac{0 + 1}{50 + 150} = \frac{1}{200}$$

- Probabilitas kata “jantung”

$$p("jantung"|"kesehatan") = \frac{(f("jantung"|"kesehatan") + 1)}{f("kesehatan") + |T|} = \frac{1 + 1}{50 + 150} = \frac{2}{200}$$

$$p("jantung"|"olahraga") = \frac{(f("jantung"|"olahraga") + 1)}{f("olahraga") + |T|} = \frac{0 + 1}{50 + 150} = \frac{1}{200}$$

$$p("jantung"|"hiburan") = \frac{(f("jantung"|"hiburan") + 1)}{f("hiburan") + |T|} = \frac{0 + 1}{50 + 150} = \frac{1}{200}$$

- Probabilitas kata “kondisi”

$$p("kondisi"|"kesehatan") = \frac{(f("kondisi"|"kesehatan") + 1)}{f("kesehatan") + |T|} = \frac{1 + 1}{50 + 150} = \frac{2}{200}$$

$$p("kondisi"|"olahraga") = \frac{(f("kondisi"|"olahraga") + 1)}{f("olahraga") + |T|} = \frac{0 + 1}{50 + 150} = \frac{1}{200}$$



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.

b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

$$p("kondisi"|"hiburan") = \frac{(f("kondisi"|"hiburan") + 1)}{f("hiburan") + |T|} = \frac{0 + 1}{50 + 150} = \frac{1}{200}$$

- Probabilitas kata “fungsi”

$$p("fungsi"|"kesehatan") = \frac{(f("fungsi"|"kesehatan") + 1)}{f("kesehatan") + |T|} = \frac{1 + 1}{50 + 150} = \frac{2}{200}$$

$$p("fungsi"|"olahraga") = \frac{(f("fungsi"|"olahraga") + 1)}{f("olahraga") + |T|} = \frac{0 + 1}{50 + 150} = \frac{1}{200}$$

$$p("fungsi"|"hiburan") = \frac{(f("fungsi"|"hiburan") + 1)}{f("hiburan") + |T|} = \frac{0 + 1}{50 + 150} = \frac{1}{200}$$

- Probabilitas kata “otot”

$$p("otot"|"kesehatan") = \frac{(f("otot"|"kesehatan") + 1)}{f("kesehatan") + |T|} = \frac{1 + 1}{50 + 150} = \frac{2}{200}$$

$$p("otot"|"olahraga") = \frac{(f("otot"|"olahraga") + 1)}{f("olahraga") + |T|} = \frac{0 + 1}{50 + 150} = \frac{1}{200}$$

$$p("otot"|"hiburan") = \frac{(f("otot"|"hiburan") + 1)}{f("hiburan") + |T|} = \frac{0 + 1}{50 + 150} = \frac{1}{200}$$

- Probabilitas kata “gerak”

$$p("gerak"|"kesehatan") = \frac{(f("gerak"|"kesehatan") + 1)}{f("kesehatan") + |T|} = \frac{1 + 1}{50 + 150} = \frac{2}{200}$$

$$p("gerak"|"olahraga") = \frac{(f("gerak"|"olahraga") + 1)}{f("olahraga") + |T|} = \frac{0 + 1}{50 + 150} = \frac{1}{200}$$

$$p("gerak"|"hiburan") = \frac{(f("gerak"|"hiburan") + 1)}{f("hiburan") + |T|} = \frac{0 + 1}{50 + 150} = \frac{1}{200}$$

- Probabilitas kata “rusak”

$$p("rusak"|"kesehatan") = \frac{(f("rusak"|"kesehatan") + 1)}{f("kesehatan") + |T|} = \frac{1 + 1}{50 + 150} = \frac{2}{200}$$

$$p("rusak"|"olahraga") = \frac{(f("rusak"|"olahraga") + 1)}{f("olahraga") + |T|} = \frac{0 + 1}{50 + 150} = \frac{1}{200}$$


**Hak Cipta Dilindungi Undang-Undang**

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber.

a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.

b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

$$p(\text{"rusak"}|\text{"hiburan"}) = \frac{(f(\text{"rusak"}|\text{"hiburan"}) + 1)}{f(\text{"hiburan"}) + |T|} = \frac{0 + 1}{50 + 150} = \frac{1}{200}$$

- Probabilitas kata “kurang”

$$p(\text{"kurang"}|\text{"kesehatan"}) = \frac{(f(\text{"kurang"}|\text{"kesehatan"}) + 1)}{f(\text{"kesehatan"}) + |T|} = \frac{1 + 1}{50 + 150} = \frac{2}{200}$$

$$p(\text{"kurang"}|\text{"olahraga"}) = \frac{(f(\text{"kurang"}|\text{"olahraga"}) + 1)}{f(\text{"olahraga"}) + |T|} = \frac{0 + 1}{50 + 150} = \frac{1}{200}$$

$$p(\text{"kurang"}|\text{"hiburan"}) = \frac{(f(\text{"kurang"}|\text{"hiburan"}) + 1)}{f(\text{"hiburan"}) + |T|} = \frac{0 + 1}{50 + 150} = \frac{1}{200}$$

- Probabilitas kata “pasok”

$$p(\text{"pasok"}|\text{"kesehatan"}) = \frac{(f(\text{"pasok"}|\text{"kesehatan"}) + 1)}{f(\text{"kesehatan"}) + |T|} = \frac{1 + 1}{50 + 150} = \frac{2}{200}$$

$$p(\text{"pasok"}|\text{"olahraga"}) = \frac{(f(\text{"pasok"}|\text{"olahraga"}) + 1)}{f(\text{"olahraga"}) + |T|} = \frac{0 + 1}{50 + 150} = \frac{1}{200}$$

$$p(\text{"pasok"}|\text{"hiburan"}) = \frac{(f(\text{"pasok"}|\text{"hiburan"}) + 1)}{f(\text{"hiburan"}) + |T|} = \frac{0 + 1}{50 + 150} = \frac{1}{200}$$

- Probabilitas kata “oksigen”

$$p(\text{"oksigen"}|\text{"kesehatan"}) = \frac{(f(\text{"oksigen"}|\text{"kesehatan"}) + 1)}{f(\text{"kesehatan"}) + |T|} = \frac{1 + 1}{50 + 150} = \frac{2}{200}$$

$$p(\text{"oksigen"}|\text{"olahraga"}) = \frac{(f(\text{"oksigen"}|\text{"olahraga"}) + 1)}{f(\text{"olahraga"}) + |T|} = \frac{0 + 1}{50 + 150} = \frac{1}{200}$$

$$p(\text{"oksigen"}|\text{"hiburan"}) = \frac{(f(\text{"oksigen"}|\text{"hiburan"}) + 1)}{f(\text{"hiburan"}) + |T|} = \frac{0 + 1}{50 + 150} = \frac{1}{200}$$

- Probabilitas kata “utama”

$$p(\text{"utama"}|\text{"kesehatan"}) = \frac{(f(\text{"utama"}|\text{"kesehatan"}) + 1)}{f(\text{"kesehatan"}) + |T|} = \frac{1 + 1}{50 + 150} = \frac{2}{200}$$

$$p(\text{"utama"}|\text{"olahraga"}) = \frac{(f(\text{"utama"}|\text{"olahraga"}) + 1)}{f(\text{"olahraga"}) + |T|} = \frac{0 + 1}{50 + 150} = \frac{1}{200}$$



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.

b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkannya dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

$$p(\text{"utama"}|\text{"hiburan"}) = \frac{(f(\text{"utama"}|\text{"hiburan"}) + 1)}{f(\text{"hiburan"}) + |T|} = \frac{0 + 1}{50 + 150} = \frac{1}{200}$$

- Probabilitas kata “aterosklerosis”

$$p(\text{"atersklerosis"}|\text{"kesehatan"}) = \frac{(f(\text{"atersklerosis"}|\text{"kesehatan"}) + 1)}{f(\text{"kesehatan"}) + |T|} = \frac{1 + 1}{50 + 150} = \frac{2}{200}$$

$$p(\text{"atersklerosis"}|\text{"olahraga"}) = \frac{(f(\text{"atersklerosis"}|\text{"olahraga"}) + 1)}{f(\text{"olahraga"}) + |T|} = \frac{0 + 1}{50 + 150} = \frac{1}{200}$$

$$p(\text{"atersklerosis"}|\text{"hiburan"}) = \frac{(f(\text{"atersklerosis"}|\text{"hiburan"}) + 1)}{f(\text{"hiburan"}) + |T|} = \frac{0 + 1}{50 + 150} = \frac{1}{200}$$

- Probabilitas kata “sempit”

$$p(\text{"sempit"}|\text{"kesehatan"}) = \frac{(f(\text{"sempit"}|\text{"kesehatan"}) + 1)}{f(\text{"kesehatan"}) + |T|} = \frac{1 + 1}{50 + 150} = \frac{2}{200}$$

$$p(\text{"sempit"}|\text{"olahraga"}) = \frac{(f(\text{"sempit"}|\text{"olahraga"}) + 1)}{f(\text{"olahraga"}) + |T|} = \frac{0 + 1}{50 + 150} = \frac{1}{200}$$

$$p(\text{"sempit"}|\text{"hiburan"}) = \frac{(f(\text{"sempit"}|\text{"hiburan"}) + 1)}{f(\text{"hiburan"}) + |T|} = \frac{0 + 1}{50 + 150} = \frac{1}{200}$$

- Probabilitas kata “keras”

$$p(\text{"keras"}|\text{"kesehatan"}) = \frac{(f(\text{"keras"}|\text{"kesehatan"}) + 1)}{f(\text{"kesehatan"}) + |T|} = \frac{1 + 1}{50 + 150} = \frac{2}{200}$$

$$p(\text{"keras"}|\text{"olahraga"}) = \frac{(f(\text{"keras"}|\text{"olahraga"}) + 1)}{f(\text{"olahraga"}) + |T|} = \frac{0 + 1}{50 + 150} = \frac{1}{200}$$

$$p(\text{"keras"}|\text{"hiburan"}) = \frac{(f(\text{"keras"}|\text{"hiburan"}) + 1)}{f(\text{"hiburan"}) + |T|} = \frac{0 + 1}{50 + 150} = \frac{1}{200}$$

- Probabilitas kata “darah”



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

$$p("darah"|"kesehatan") = \frac{(f("darah"|"kesehatan") + 1)}{f("kesehatan") + |T|} = \frac{1 + 1}{50 + 150} = \frac{2}{200}$$

$$p("darah"|"olahraga") = \frac{(f("darah"|"olahraga") + 1)}{f("olahraga") + |T|} = \frac{0 + 1}{50 + 150} = \frac{1}{200}$$

$$p("darah"|"hiburan") = \frac{(f("darah"|"hiburan") + 1)}{f("hiburan") + |T|} = \frac{0 + 1}{50 + 150} = \frac{1}{200}$$

b. Tahapan *testing*

Pada tahap ini, dilakukan pengujian dengan memasukkan data uji ke dalam model yang telah dibentuk pada tahap *training* tadi. Pada Tabel 4.9 adalah sebuah data yang akan dijadikan sebagai data uji.

**Tabel 4.9 contoh data *testing***

Berita	kelas
Sergio aguero memuaskan josep Guardiola bukan hanya atas gol golnya tapi juga karena juga kerja kerasnya membantu manchester city bertahan	olahraga

Berdasarkan data uji dari Tabel 4.8 di atas, maka akan dihitung nilai probabilitasnya berdasarkan nilai probabilitas setiap *term*. Perhitungan nilai probabilitas dengan metode *Naive bayes classifier* pada dokumen *testing* ini dilakukan dengan mengalikan nilai probabilitas semua kategori dengan probabilitas tiap *term* yang diambil dari semua data.

Sebelum masuk pada tahap perhitungan *testing*, data terlebih dahulu melewati proses *preprocessing*. Hasil data testing setelah di *preprocessing* adalah seperti pada Tabel 4.10

**Tabel 4.10 Data test setelah *preprocessing***

Berita	Kelas
puas atas gol gol kerja keras bantu tahan	olahraga

$$p(\text{test}|olahraga) = p(olahraga) \times p(\text{puas}|olahraga) \times p(\text{atas}|olahraga) \times$$

$$p(\text{gol}|olahraga) \times p(\text{kerja}|olahraga) \times p(\text{keras}|olahraga) \times p(\text{bantu}|olahraga) \times p(\text{tahan}|olah$$



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumunkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

$$p(\text{test}|\text{olahraga}) = \frac{1}{3} \times \frac{1}{150} \times \frac{1}{150} \times \frac{2}{150} \times \frac{1}{150} \times \frac{1}{150} \times \frac{1}{150} = 1.89016 \times 10^{-15}$$

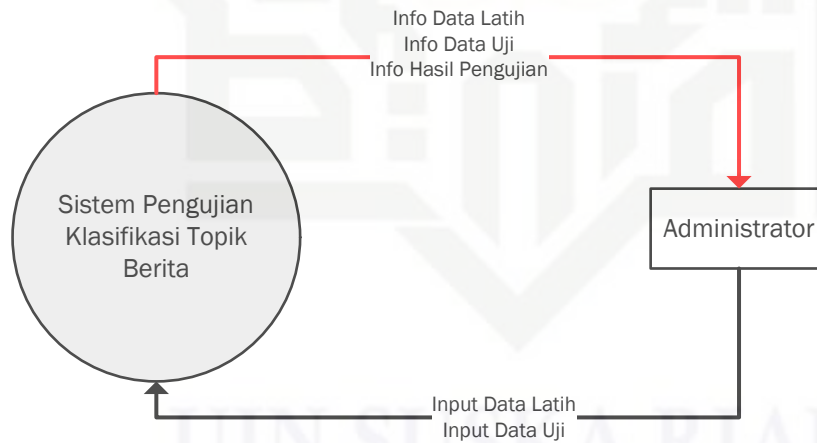
Nilai probabilitas tertinggi adalah pada kategori iya yaitu sebesar  $1.89016 \times 10^{-15}$ , sehingga berita tersebut diklasifikasikan ke dalam kategori berita olahraga.

### 4.1.3 Analisa Fungsional Sistem

Analisa fungsional sistem akan menjelaskan mengenai perancangan sistem yang akan dibuat menggunakan *Flow Diagram* (FD) meliputi *Context Diagram*, Data Flow Diagram (DFD).

a. *Context Diagram*

*Context Diagram* menggambarkan aliran fungsional dalam sebuah proses pada sistem/aplikasi. Berikut akan dijelaskan semua proses yang terjadi pada sistem. *Context Diagram* akan dijelaskan seperti pada gambar 4.13 di bawah ini.



**Gambar 4.13 Context Diagram**

Entitas yang berinteraksi dengan sistem ini adalah *Administrator*. Dalam sistem ini, *administrator* dapat melakukan penghapusan data.

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

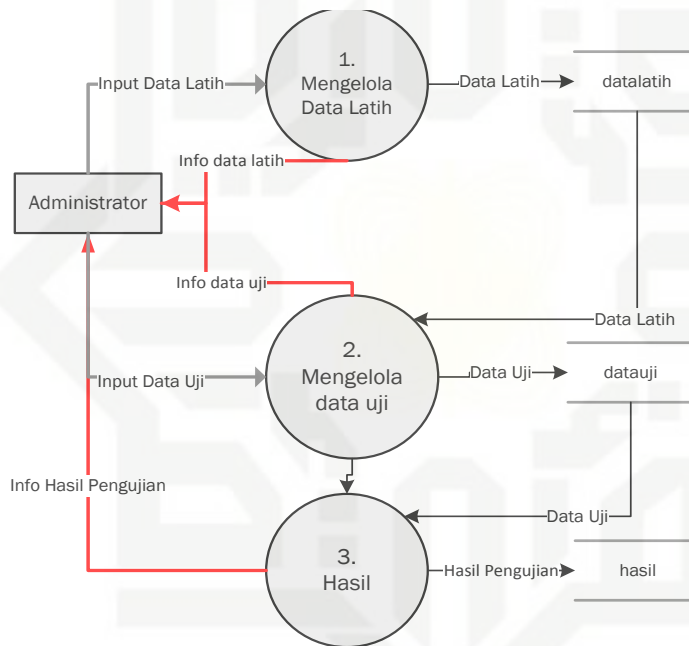
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

b. *Data Flow Diagram (DFD)*

*Data Flow Diagram (DFD)* sering digunakan untuk menggambarkan suatu sistem yang telah ada atau sistem baru yang akan dikembangkan secara logika tanpa mempertimbangkan lingkungan fisik dimana data tersebut mengalir, atau lingkungan fisik dimana data tersebut tersimpan.

**1. Data Flow Diagram (DFD) Level 1**

Gambar 4.14 dibawah ini adalah gambaran DFD *level 1* dari sistem pengujian klasifikasi topik berita.



**Gambar 4.14 DFD Level 1**

Pada Tabel 4.11 di bawah ini adalah proses DFD *Level 1*. Terdapat 3 (tiga) proses yaitu data latih, pengujian dan hasil pengujian.

**Tabel 4.11 proses DFD Level 1**

No	Nama Proses	Deskripsi
1.	Data Latih	Proses data latih dan normalisasi/tanpa

No	Nama Proses	Deskripsi
		normalisasi.
2.	Pengujian	Proses pengolahan data uji
3.	Hasil Pengujian	Tampilan hasil akurasi dalam <i>confusion matrix</i> .

Pada Tabel 4.12 dibawah ini adalah aliran data DFD *level 1*. Terdapat 5 (lima) aliran data yaitu data latih, info data latih, data uji, info data uji, dan hasil pengujian.

**Tabel 4.12 Aliran Data DFD level 1**

Nama Aliran Data	Deskripsi
Input Data Latih	Data latih yang diinputkan oleh admin
Input Data Uji	Data uji yang diinputkan oleh admin
Info Hasil Pengujian	Info data hasil pengujian

### 4.3 Perancangan Aplikasi

Pada tahap ini akan dibahas tentang perancangan aplikasi klasifikasi topik berita berdasarkan tahapan analisa yang telah dilakukan sebelumnya. Adapun perancangan yang akan dibuat adalah perancangan permodelan *flat file*, perancangan struktur menu dan perancangan *interface*.

#### 4.3.1 Perancangan File Teks (Flat File)

Dalam penelitian ini, juga akan menggunakan *flat file* yang bertindak sebagai penyimpanan yang akan mendukung pemrosesan klasifikasi topik berita. Adapun penggunaan *file teks (plain text)* akan digunakan sebagai penyimpanan dokumen uji, latih, hasil indexing, kumpulan kata *stopword* dan kamus bahasa Indonesia, kamus normalisasi, kumpulan kata kesehatan, kumpulan kata hiburan dan kumpulan kata olahraga. Untuk lebih detail dapat dijelaskan oleh Tabel 4.13 berikut.



**Tabel 4.13 Rancangan File Teks (Flat File)**

Nama File	Fungsi Penyimpanan
file-uji.txt	Sebagai penyimpanan uji
file-latih.txt	Sebagai penyimpanan latihan ( <i>Corpus</i> )
file-latih-index.txt	Sebagai penyimpanan hasil indexing
file-stopword.txt	Sebagai penyimpanan kata stopwords
kamus-ind.txt	Sebagai penyimpanan kata dasar
file-kesehatan_keyword.txt	Sebagai penyimpanan kata-kata topik beritakesehatan
file-hiburan_keyword.txt	Sebagai penyimpanan kata-kata topik beritahiburan
file-olahraga.txt	Sebagai penyimpanan kata-kata olahraga

Setiap file Teks memiliki fungsi penyimpanan tersendiri dan memuat informasi yang berkaitan terhadap fungsi penyimpanan. Pada file-uji.txt akan memuat informasi id dokumen, topik berita dokumen, topik, isi dokumen, hasil topik berita. Pada file-latih.txt akan memuat informasi id dokumen, topik berita dokumen, topik, isi dokumen. Sedangkan file-latih-index.txt merupakan penyimpanan dari hasil indexing latihan.txt dimana memuat informasi kata, id dokumen, *tremfrequenc (tf)*, *document frequency (df)*. Sementara itu file-stopword.txt, kamus-ind.txt yang akan digunakan pada *linguistic preprocessing* yaitu *stopword* dan *stemming* memuat informasi kata. Pada analisa juga membutuhkan beberapa file teks yaitu, file-kesehatan\_keyword.txt, file-hiburan\_keyword.txt dan file-olahraga.txt masing-masing memuat informasi kata.

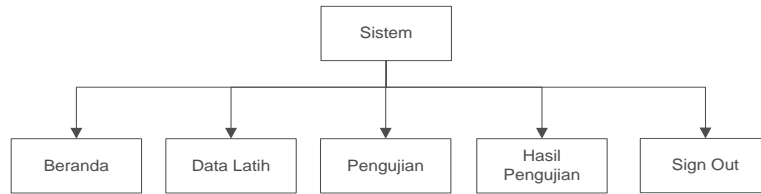
### 4.3.2 Perancangan Struktur Menu

Perancangan struktur menu menggambarkan hubungan halaman dengan halaman yang lainnya. Menu merupakan salah satu bagian penting dalam merancang sebuah sistem antar muka, karena melalui menu dapat dilihat bagaimana struktur sistem antar muka ini terbentuk.

Adapun struktur menu pada sistem ini terdiri dari Beranda, Latih, Pengujian, Klasifikasi Topik berita, Pengaturan dan *Sign Out*. Rancangan struktur menu dapat dilihat pada Gambar 4.15 berikut.

**Hak Cipta Dilindungi Undang-Undang**

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumunkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.



**Gambar 4.15 Rancangan Struktur Menu**

**4.3.3 Perancangan Interface**

*Interface* sistem adalah Antarmuka (*Interface*) merupakan mekanisme komunikasi antara pengguna (*user*) dengan sistem. Antarmuka (*Interface*) dapat menerima informasi dari pengguna (*user*) dan memberikan informasi kepada pengguna (*user*) untuk membantu mengarahkan alur penelusuran masalah sampai ditemukan suatu solusi. Berikut tampilan *interface* sistem klasifikasi topik berita ini :

**A. Tampilan Halaman Beranda**

Halaman beranda adalah halaman utama yang ditampilkan aplikasi kepada pengguna. Halaman ini berisi tentang informasi sistem klasifikasi topik berita, alur kerja sistem, dan tujuan sistem. Berikut ini adalah Gambar 4.16 rancangan *interface* halaman beranda :



**Gambar 4.16 Rancangan Interface Halaman Beranda**

**B. Tampilan Halaman Latih**

#### Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Halaman latihan merupakan halaman yang memuat koleksi dokumen yang menjadi latihan pada sistem. Dalam halaman ini terdapat tombol tambah dokumen dan lihat hapus dokumen. Tombol tambah berfungsi untuk menampilkan form inputan tambah dokumen. *Link* hapus berfungsi untuk menghapus dokumen latihan. Dalam menu ini juga terdapat fasilitas pencarian, yang berfungsi untuk melakukan pencarian dokumen. Berikut ini Gambar 4.17 adalah rancangan *interface* menu halaman latihan:



Topik Berita									
foto	<div style="text-align: right;"> <input type="text"/> </div> <div style="text-align: right;"> <input type="button" value="Tambah"/> <input type="button" value="Indexing"/> </div> <table border="1"> <thead> <tr> <th>No</th> <th>Data</th> <th>Topik</th> <th>Aksi</th> </tr> </thead> <tbody> <tr> <td colspan="4" style="text-align: center;">Data</td> </tr> </tbody> </table>	No	Data	Topik	Aksi	Data			
No	Data	Topik	Aksi						
Data									
Beranda									
Data latih									
Pengujian									
Hasil Pengujian									

**Gambar 4.17 Rancangan *Interface* Halaman Latihan**

Berikut ini Gambar 4.18 adalah rancangan *interface* halaman tambah latihan:



**Input data latih**

Dokumen

Topik

**Gambar 4.18 Rancangan *Interface* Halaman Tambah Latihan**

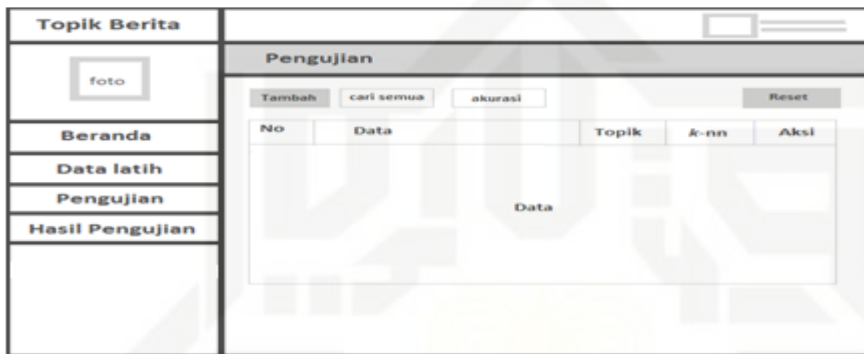
### C. Tampilan Halaman Pengujian

Halaman pengujian merupakan halaman yang memuat dokumen pengujian. Dalam halaman ini terdapat tombol tambah dokumen, cari topik berita, lihat akurasi, hapus dan *reset*. Tombol tambah berfungsi menambah dokumen pengujian sesuai

Hak Cipta Dilindungi Undang-Undang

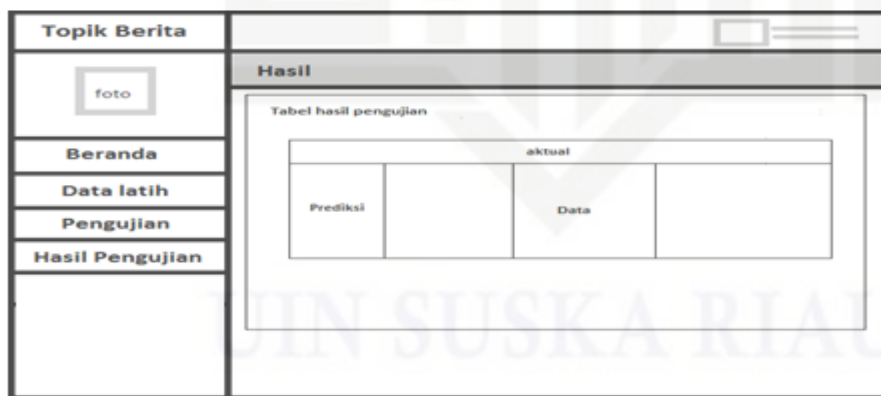
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

keinginan *user*. Tombol caritopik beritakan memberikan pilihan pencarian topik berita berdasarkan pilihan metode. Sedang pencarian topik berita secara detail dapat dilihat pada *field* aksi pada Tabel. Tombol lihat akurasi akan menampilkan halaman hasil pengujian sesuai pilihan metode. Sedangkan tombol *reset* akan memberikan aksi mengembalikan nilai topik berita pada keadaan awal. Berikut ini adalah Gambar 4.19 yaitu rancangan *interface* halaman pengujian :



**Gambar 4.19 Rancangan *Interface* Halaman Pengujian**

Jika *user* memilih lihat akurasi maka sistem akan menampilkan halaman hasil akurasi. Berikut ini adalah Gambar 4.20 yaitu rancangan *interface* halaman lihat akurasi :



**Gambar 4.20 Rancangan *Interface* Halaman Lihat Akurasi**

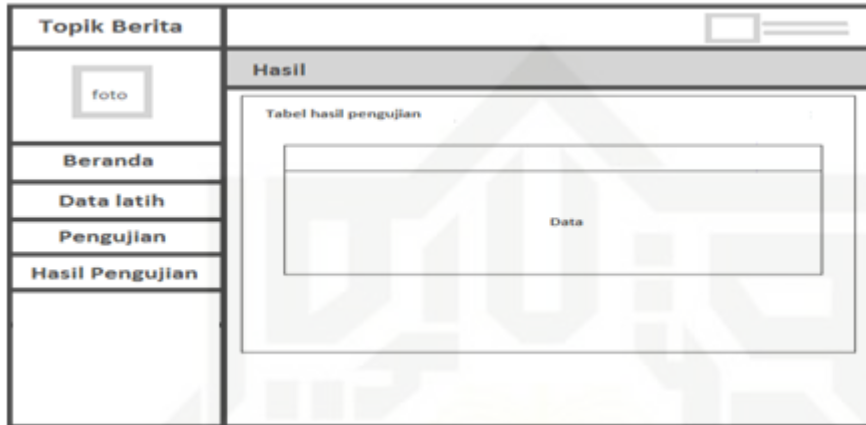


**Hak Cipta Dilindungi Undang-Undang**

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

**D. Tampilan Halaman Hasil Pengujian**

Halaman hasil pengujian memuat informasi hasil pengujian yang telah dilakukan. Terdapat informasi akurasi pengujian. Berikut halaman hasil pengujian dijelaskan Gambar 4.21



**Gambar 4.21 Rancangan Interface Halaman Hasil Pengujian**