

## BAB IV

### PEMBAHASAN

#### 4.1 Pengumpulan Data

Data dalam penelitian ini diperoleh dari Lembaga Pengolahan dan Penyedia Informasi (PPI), Dirjen Informasi dan Komunikasi Publik (IKP) yang berasal dari 5 media *online* yaitu detik.com, viva.co.id, inilah.com, antaranews.com, dan okezone.com yang diambil dari bulan Februari 2016 sampai dengan Mei 2016 untuk bidang Polhukam, Ekonomi dan Kesra. Jumlah data yang digunakan sebanyak 1007 data yang telah dikelompokkan ke dalam 15 kelas secara manual. Pembagian data berdasarkan jumlah kelas isu ditunjukkan pada Tabel 4.1.

Tabel 4.1 Data Penelitian

Kategori	Isu	Jumlah
Polhukam	Politik Dalam Negeri	67
	Politik Luar Negeri	87
	Kasus Korupsi	77
	Hukum	42
	Keamanan	42
	Kasus Narkoba	77
	Kekerasan Seksual	68
Perekonomian	Pajak	84
	Perbankan	78
Kesra	Infrastruktur	26
	Kecelakaan	78
	Bencana Alam	83
	Energi	67
	Lingkungan Hidup	79
	Transportasi Publik	52

Data-data ini kemudian dibagi untuk proses *training* dan proses *testing*. Karena data berita dipengaruhi oleh waktu, maka data-data berita yang terbit lebih awal dijadikan sebagai data *training* pada awal proses.

Data yang digunakan pada awal *training* sebanyak 395 data diperoleh dari masing-masing kategori berdasarkan waktu terbit yang lebih awal. Selanjutnya, proses *testing* menggunakan data yang terbit setelah data-data *training* sebelumnya yang dibagi menjadi 6 kali proses *testing*. Setiap data *testing* yang telah dilakukan, akan digunakan sebagai data *training* pada proses selanjutnya.

Pembagian jumlah data training dan testing pada penelitian ini dapat dilihat pada tabel berikut ini.

Tabel 4.2 Data *Training* dan *Testing*

Proses ke-n	Data <i>Training</i>	Data <i>Testing</i>
1	395	98
2	493	101
3	594	103
4	697	101
5	798	104
6	902	105

#### 4.2 Text Preprocessing

Tahap-tahap pada *text preprocessing* adalah sebagai berikut :

##### 1. Case Folding

Proses *case folding* dapat dilihat pada Gambar 4.1.

Petugas Komisi Pemberantasan Korupsi telah melakukan penggeladahan di sejumlah tempat terkait kasus dugaan suap pengajuan Peninjauan Kembali di Pengadilan Negeri Jakarta Pusat. Ketua KPK Agus Rahardjo, menyebut penggeladahan dilakukan di empat tempat. Di antaranya adalah Kantor Paramount Enterprise International di kawasan Gading Serpong Boulevard, Tangerang, Kantor Pengadilan Negeri Jakarta Pusat, Ruang Kerja Sekretaris MA serta rumah Sekretaris MA



petugas komisi pemberantasan korupsi telah melakukan penggeladahan di sejumlah tempat terkait kasus dugaan suap pengajuan peninjauan kembali di pengadilan negeri jakarta pusat ketua kpk agus rahardjo menyebut penggeladahan dilakukan di empat tempat di antaranya adalah kantor paramount enterprise international di kawasan gading serpong boulevard tangerang kantor pengadilan negeri jakarta pusat ruang kerja sekretaris ma serta rumah sekretaris ma

Gambar 4.1 Proses *Case Folding*

## 2. *Tokenization*

Proses *tokenization* adalah proses untuk mengubah kalimat menjadi *token-token* huruf. Proses *tokenization* dapat dilihat pada Gambar 4.2.

petugas komisi pemberantasan korupsi telah melakukan penggeladahan di sejumlah tempat terkait kasus dugaan suap pengajuan peninjauan kembali di pengadilan negeri jakarta pusat ketua kpk agus rahardjo menyebut penggeladahan dilakukan di empat tempat di antaranya adalah kantor paramount enterprise international di kawasan gading serpong boulevard tangerang kantor pengadilan negeri jakarta pusat ruang kerja sekretaris ma serta rumah sekretaris ma

Petugas – komisi – pemberantasan – korupsi – telah – melakukan – penggeladahan – di – sejumlah – tempat – terkait – kasus – dugaan – suap – pengajuan – peninjauan – kembali – di – pengadilan – negeri – jakarta – pusat – ketua – kpk – agus – rahardjo – menyebut – penggeladahan – dilakukan – di – empat – tempat – di – antaranya – adalah – kantor – paramount – enterprise – international – di – kawasan – gading – serpong – boulevard – tangerang – kantor – pengadilan – negeri – jakarta – pusat – ruang – kerja – sekretaris – ma – serta – rumah – sekretaris – ma

Gambar 4.2 Proses *Tokenization*

## 3. *Stemming*

Proses *stemming* bertujuan untuk mengubah kata menjadi bentuk kata dasar. Gambar 4.3 memperlihatkan hasil dari proses *stemming*.

petugas – komisi – pemberantasan - korupsi – telah – melakukan –  
 penggeledahan – di – sejumlah - tempat – terkait – kasus - dugaan –  
 suap – pengajuan – peninjauan - kembali – di – pengadilan – negeri –  
 jakarta - pusat - ketua - kpk - agus – rahardjo - menyebut –  
 penggeledahan - dilakukan - di – empat - tempat – di - antaranya –  
 adalah - kantor – paramount – enterprise – international – di –  
 kawasan – gading – serpong – boulevard - tangerang – kantor –  
 pengadilan - negeri – jakarta – pusat – ruang – kerja – sekretaris - ma  
 – serta - rumah – sekretaris – ma

tugas – komisi – berantas – korupsi – telah – laku – geledah – di –  
 jumlah – tempat – kait - kasus – duga – suap – aju – tinjau – kembali –  
 di – adil – negeri – jakarta - pusat – ketua - kpk – agus – rahardjo –  
 sebut – geledah – laku – di – empat – tempat – di – antara – adalah –  
 kantor - paramount – enterprise – international – di – kawasan –  
 gading – serpong – boulevard - tangerang - kantor - adil – negeri –  
 jakarta – pusat – ruang - kerja – sekretaris – ma – serta - rumah –  
 sekretaris – ma

Gambar 4.3 Proses *Stemming*

#### 4. *Filtering*

Proses *filtering* bertujuan untuk menghilangkan *stopwords*. Proses *filtering* ditunjukkan pada Gambar 4.4.

tugas – komisi – berantas – korupsi – telah – laku – geledah – di –  
 jumlah – tempat – kait - kasus – duga – suap – aju – tinjau – kembali –  
 di – adil – negeri – jakarta - pusat – ketua - kpk – agus – rahardjo –  
 sebut – geledah – laku – di – empat – tempat – di – antara – adalah –  
 kantor - paramount – enterprise – international – di – kawasan –  
 gading – serpong – boulevard - tangerang - kantor - adil – negeri –  
 jakarta – pusat – ruang - kerja – sekretaris – ma – serta - rumah –  
 sekretaris – ma

tugas - komisi – berantas – korupsi - laku - geledah - tempat - kait - kasus - duga - suap - aju - tinjau - adil - negeri - jakarta - pusat - ketua - kpk - agus - rahardjo - sebut - geledah - laku - empat – tempat - kantor - paramount - enterprise - international - kawasan – gading - serpong – boulevard - tangerang - kantor - adil - negeri – jakarta – pusat - ruang - kerja – sekretaris - ma - rumah - sekretaris – ma

Gambar 4.4 Proses *Filtering*

Setiap kata hasil *text preprocessing* ini kemudian disimpan di dalam database.

#### 4.3 Document Frequency Thresholding

Proses seleksi fitur dilakukan untuk seluruh data *training*. Tujuan dari proses ini adalah untuk mengurangi dimensi fitur, sehingga fitur-fitur yang digunakan pada proses klasifikasi merupakan fitur yang memiliki nilai kontribusi cukup tinggi terhadap suatu kelas. Salah satu proses seleksi fitur pada penelitian ini menggunakan metode *Document Frequency Thresholding*.

##### 1. Menghitung *Term Frequency*

*Term frequency* merupakan nilai frekuensi fitur kata pada dokumen tertentu. Contoh perhitungan *term frequency* ditunjukkan pada Tabel 4.3.

Tabel 4.3 Contoh Data Berita

1.	tugas komisi berantas korupsi laku geledah tempat kait kasus duga suap aju tinjau adil negeri jakarta pusat ketua kpk agus rahardjo sebut geledah laku empat tempat kantor paramount enterprise international kawasan gading serpong boulevard tangerang kantor adil negeri jakarta pusat ruang kerja sekretaris ma rumah sekretaris ma
2.	komisi berantas korupsi kpk resmi tahan panitera adil negeri pn jakarta pusat edy nasution en bahkan tetap bungkam kait periksa tunduk mobil tahan sekira pukul wib sangkut edy tahan hari pertama perlu sidi ujar laksana hari plh kabiyo humas kpk yuyuk andriati konfirmasi awak media Kamis

3.	nn laku intai sejak lama bupati ogan ilir ahmad wazir nofiadi kait penyalahgunaan narkoba nofiadi tahu lama alami gantung narkoba jenis sabu bagaimana cara transaksi sang bupati beli narkoba bnn duga nofiadi gantung narkoba jenis sabu lama lihat kondisi gigi rusak
4.	wakil ketua komisi berantas korupsi kpk saut situmorang beri klarifikasi kait nyata singgung himpun mahasiswa Islam hmi saut singgung kena banyak kader himpun mahasiswa islam hmi bukti korupsi jadi jabat

Tabel 4.4 Contoh Perhitungan *Term Frequency*

Kata	TF			
	Data-1	Data-2	Data-3	Data-4
korupsi	1	1	0	2

## 2. Menghitung *Document Frequency*

*Document frequency* merupakan banyaknya jumlah dokumen yang mengandung fitur tertentu. Contoh perhitungan *document frequency* ditunjukkan pada Tabel 4.5.

Tabel 4.5 Contoh Perhitungan *Document Frequency*

Kata	TF				DF
	Data-1	Data-2	Data-3	Data-4	
korupsi	1	1	0	2	3

## 3. Menentukan *Threshold*

Setelah didapatkan nilai DF untuk semua *term*, kemudian dilakukan pemilihan *threshold*. Penentuan *threshold* dilakukan dengan percobaan menentukan batas bawah dan batas atas pada data *training* awal. Berdasarkan pengamatan pada hasil perhitungan *document frequency*, maka dilakukan pengujian pada data *training* dengan beberapa *threshold* yang ditunjukkan pada Tabel 4.6.



Tabel 4.6 Pengujian *DF-Threshold*

<i>Threshold</i>	Jumlah Kata	Akurasi
1 – 55	8.737	99,74
1 – 60	8.865	99,74
2 – 50	7.769	98,97
2 – 55	7.940	99,22
2 – 60	8.068	99,22

Berdasarkan hasil uji pada data *training* tersebut, *threshold* yang dipilih untuk percobaan pada proses selanjutnya adalah 1-55.

#### 4.4 *Term Frequency – Inverse Document Frequency (TFIDF)*

Proses pembobotan TF-IDF ini dimulai dengan menghitung tiap *term* yang ada pada setiap dokumen (TF). Kemudian proses dilanjutkan dengan menghitung jumlah dokumen yang memiliki *term* tertentu (DF). Setelah itu proses menghitung *Inverse Document Frequeuncy* (IDF) dan yang terakhir nilai TF dikalikan dengan nilai IDF. Contoh perhitungan *TFIDF* dengan menggunakan data berita pada Tabel 4.2 dapat dilihat pada Tabel 4.7.

Tabel 4.7 Contoh Perhitungan *TFIDF*

Kata	TF				DF	IDF	TFIDF			
	D-1	D-2	D-3	D-4			D-1	D-2	D-3	D-4
korupsi	1	1	0	2	3	0,1249	0,1249	0,1249	0	0,2498

#### 4.5 **Klasifikasi**

Untuk proses klasifikasi terbagi menjadi dua yaitu *training* dan *testing*. Data *training* yang telah melewati seleksi fitur kemudian digunakan sebagai bahan pembelajaran pada proses *testing* untuk menentukan suatu data berita masuk pada kelas isu tertentu.

Data yang diinputkan pada proses *testing* sama halnya seperti data pada proses *training*, melewati tahap *text preprocessing* dan kemudian dihitung nilai tiap fitur

kata yang akan digunakan untuk proses klasifikasi menggunakan metode *Multinomial Naive Bayes*.

Tiap dokumen akan dihitung nilai probabilitasnya untuk setiap kelas isu. Nilai probabilitas didapatkan dari hasil perkalian antara nilai *prior* kelas isu dan nilai-nilai probabilitas kata ke-n dalam kelas isu tersebut berdasarkan frekuensi kemunculannya. Kelas dengan nilai probabilitas terbesar maka akan dipilih sebagai kelas dari dokumen tersebut.

Pada proses perhitungan *Multinomial Naive Bayes* dengan menggunakan TFIDF, nilai probabilitas didapatkan dari hasil perkalian antara nilai *prior* kelas isu dan nilai-nilai probabilitas kata ke-n dalam kelas isu berdasarkan bobot kata dalam dokumen.

#### 4.6.1 Contoh Perhitungan Multinomial Naive Bayes

Berikut ini merupakan contoh data yang akan digunakan dalam perhitungan *Multinomial Naive Bayes*.

Tabel 4.8 Contoh Data untuk Perhitungan Multinomial

	Dokumen	Data	Kelas
<i>Training</i>	1	korupsi pejabat	Kasus Korupsi
	2	pejabat jerat kasus korupsi	Kasus Korupsi
	3	polisi tangkap edar narkoba	Kasus Narkoba
<i>Testing</i>	4	tangkap kasus narkoba pejabat narkoba	?

1. Menghitung nilai prior :

$$P(\text{korupsi}) = \frac{2}{3}$$

$$P(\text{narkoba}) = \frac{1}{3}$$

2. Menghitung probabilitas kata ke-n data

$$P(\text{tangkap} \mid \text{korupsi}) = \frac{0+1}{6+8} = \frac{1}{14}$$

$$P(\text{kasus} \mid \text{korupsi}) = \frac{1+1}{6+8} = \frac{2}{14}$$



$$\begin{aligned}
 P(\text{narkoba} \mid \text{korupsi}) &= \frac{0+1}{6+8} = \frac{1}{14} \\
 P(\text{pejabat} \mid \text{korupsi}) &= \frac{2+1}{6+8} = \frac{3}{14} \\
 P(\text{tangkap} \mid \text{narkoba}) &= \frac{1+1}{4+8} = \frac{2}{12} \\
 P(\text{kasus} \mid \text{narkoba}) &= \frac{0+1}{4+8} = \frac{1}{12} \\
 P(\text{narkoba} \mid \text{narkoba}) &= \frac{1+1}{4+8} = \frac{2}{12} \\
 P(\text{pejabat} \mid \text{narkoba}) &= \frac{0+1}{4+8} = \frac{1}{12}
 \end{aligned}$$

3. Menghitung probabilitas dokumen pada kelas

$$P(\text{korupsi} \mid d4) = \frac{2}{3} \times \frac{1}{14} \times \frac{2}{14} \times \left(\frac{1}{14}\right)^2 \times \frac{3}{14} = 0,000007437$$

$$P(\text{narkoba} \mid d4) = \frac{1}{3} \times \frac{2}{12} \times \frac{1}{12} \times \left(\frac{2}{12}\right)^2 \times \frac{1}{12} = 0,00001071674$$

4. Menentukan kelas berdasarkan nilai probabilitas tertinggi

Nilai tertinggi adalah 0,00001071674, sehingga dokumen ke-4 tersebut diklasifikasikan ke dalam kategori Kasus Narkoba.

#### 4.6.2 Contoh Perhitungan Multinomial Naive Bayes dengan *TFIDF*

1. Menghitung bobot tiap kata

Tabel 4.9 Contoh Perhitungan *Multinomial Naive Bayes* dengan *TFIDF*

Term	TF			DF	IDF	W			$\Sigma W \text{ kata}$	
	D1	D2	D3			D1	D2	D3	Korupsi	Narkoba
korupsi	1	1	0	2	0,17609	0,17609	0,17609	0	0,35218	0
pejabat	1	1	0	2	0,17609	0,17609	0,17609	0	0,35218	0
jerat	0	1	0	1	0,47712	0	0,47712	0	0,47712	0
kasus	0	1	0	1	0,47712	0	0,47712	0	0,47712	0
polisi	0	0	1	1	0,47712	0	0	0,47712	0	0,47712
tangkap	0	0	1	1	0,47712	0	0	0,47712	0	0,47712
Edar	0	0	1	1	0,47712	0	0	0,47712	0	0,47712
narkoba	0	0	1	1	0,47712	0	0	0,47712	0	0,47712
Jumlah					3,2149				1,6586	1,90848

## 2. Menghitung probabilitas kata

$$P(\text{tangkap} \mid \text{korupsi}) = \frac{0+1}{1,6586+8} = 0,103534$$

$$P(\text{kasus} \mid \text{korupsi}) = \frac{0,47712+1}{1,6586+8} = 0,152933$$

$$P(\text{narkoba} \mid \text{korupsi}) = \frac{0+1}{1,6586+8} = 0,103535$$

$$P(\text{pejabat} \mid \text{korupsi}) = \frac{0,35218+1}{1,6586+8} = 0,243532$$

$$P(\text{tangkap} \mid \text{narkoba}) = \frac{0,47712+1}{1,90848+8} = 0,25$$

$$P(\text{kasus} \mid \text{narkoba}) = \frac{0+1}{1,90848+8} = 0,100924$$

$$P(\text{narkoba} \mid \text{narkoba}) = \frac{0,47712+1}{1,90848+3,2149} = 0,25$$

$$P(\text{pejabat} \mid \text{narkoba}) = \frac{0+1}{1,90848+3,2149} = 0,100924$$

## 3. Menghitung probabilitas dokumen pada kelas

$$\begin{aligned} P(\text{korupsi} \mid d4) &= \frac{2}{3} \times 0,103534 \times 0,152933 \times (0,103535)^2 \times 0,243532 \\ &= 0,00000275565 \end{aligned}$$

$$\begin{aligned} P(\text{narkoba} \mid d4) &= \frac{1}{3} \times 0,25 \times 0,100924 \times (0,25)^2 \times 0,100924 \\ &= 0,00000530499 \end{aligned}$$

## 4. Menentukan kelas berdasarkan nilai probabilitas tertinggi

Nilai tertinggi adalah 0.00000530499, sehingga dokumen ke-4 tersebut diklasifikasikan ke dalam kategori Kasus Narkoba.

#### 4.6 Evaluasi

Evaluasi dalam penelitian ini menggunakan perhitungan akurasi, *precision* dan *recall* dengan *confusion matrix* yang ditampilkan pada Lampiran 1. Hasil klasifikasi 6 kali *training* dan *testing* ditunjukkan pada tabel-tabel berikut ini.

Tabel 4.10 Hasil Evaluasi *Multinomial Naive Bayes*

<i>Data Training</i>	<i>Data Testing</i>	<i>Keyword</i>	Akurasi	Precision	Recall
395	98	10057	70,41	69,33	72,06
493	101	12645	86,14	82,41	85,76
594	103	15121	89,32	92,12	91,72
697	101	17648	88,12	91,37	88,87
798	104	20018	89,42	89,17	87,56
902	105	22512	94,29	95	93,05
Rata-rata			86,28	86,57	86,5

Tabel 4.11 Hasil Evaluasi *Multinomial Naive Bayes* dengan *DF-Thresholding*

<i>Data Training</i>	<i>Data Testing</i>	<i>Keyword</i>	Akurasi	Precision	Recall
395	98	8737	72,45	71,21	74,5
493	101	10742	85,15	79,96	85,24
594	103	12494	89,32	91,57	91,72
697	101	14228	90,09	91,67	90,58
798	104	15657	85,58	84,82	83,27
902	105	16766	93,33	91,55	92,09
Rata-rata			85,98	85,13	86,22

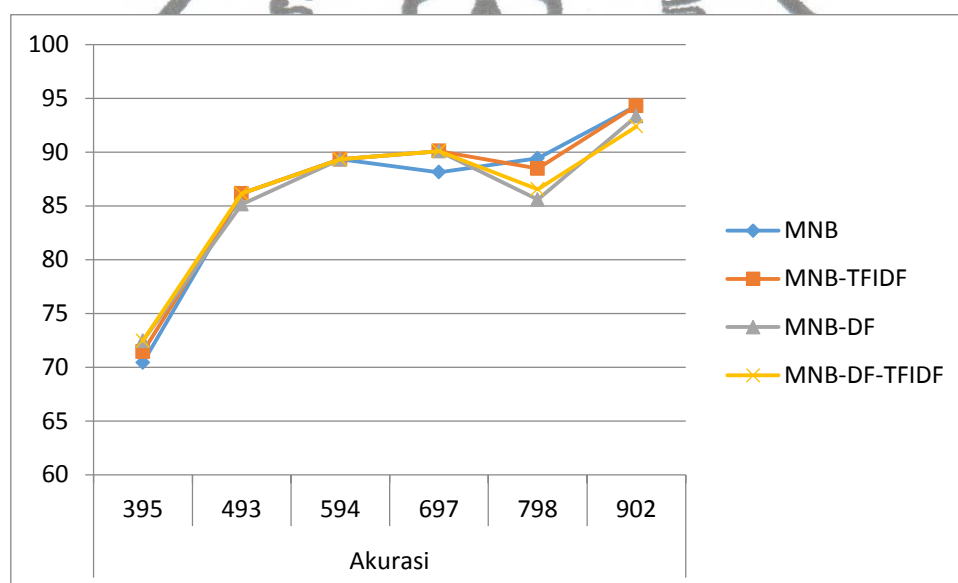
Tabel 4.12 Hasil Evaluasi *Multinomial Naive Bayes* dengan *TFIDF*

<i>Data Training</i>	<i>Data Testing</i>	<i>Keyword</i>	Akurasi	Precision	Recall
395	98	10057	71,43	68,04	72,88
493	101	12645	86,14	83,77	85,76
594	103	15121	89,32	92,12	91,72
697	101	17648	90,09	92,88	90,65
798	104	20018	88,46	88,06	86,44
902	105	22512	94,29	95	93,05
Rata-rata			86,62	86,64	86,75

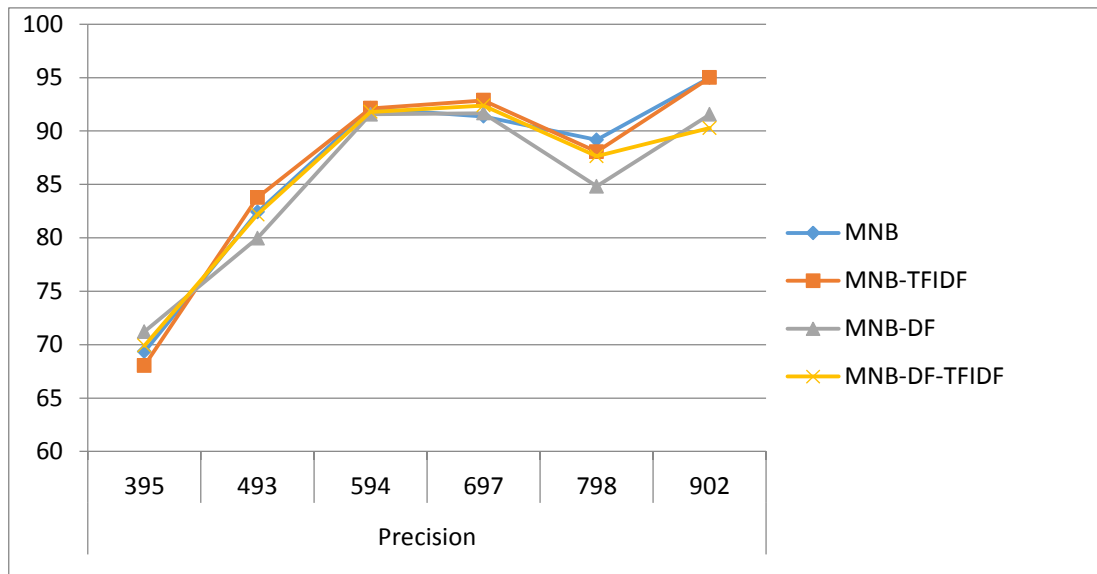
Tabel 4.13 Hasil Evaluasi *Multinomial Naive Bayes* dengan *DF Thresholding-TFIDF*

Data Training	Data Testing	Keyword	Akurasi	Precision	Recall
395	98	8737	72,45	69,9	74,5
493	101	10742	86,14	82,18	85,24
594	103	12494	89,32	91,57	91,65
697	101	14228	90,09	92,39	90,58
798	104	15657	86,54	87,64	84,38
902	105	16766	92,38	90,28	91,14
Rata-rata			86,15	85,73	85,07

Hasil evaluasi seluruhnya ditujukan pada gambar-gambar berikut ini.



Gambar 4.5 Hasil Akurasi

Gambar 4.6 Hasil *Precision*Gambar 4.7 Hasil *Recall*

Dari hasil proses pengujian, tingkat akurasi di awal proses cenderung kecil. Hal ini disebabkan karena data berita yang diinputkan sebagai data *testing* memiliki fitur kata yang belum muncul pada data berita sebelumnya yang dijadikan sebagai proses pembelajaran. Namun dari hasil tersebut, dapat dianalisis bahwa dengan semakin bertambahnya data *training*, akurasi cenderung naik dan stabil. Hal ini

dikarenakan fitur-fitur kata yang dijadikan sebagai bahan pembelajaran lebih banyak dan beragam.

Metode *Multinomial Naive Bayes* dengan menggunakan pembobotan *TFIDF* memiliki hasil yang lebih baik dari metode *Multinomial* maupun *Multinomial* dengan menggunakan fitur seleksi *DF-Threshold* dan juga *Multinomial* dengan kombinasi *DF-Threshold* dengan *TFIDF*. Penggunaan *DF-Thresholding* dengan *threshold* yang dipilih sebesar 1 - 55 justru mengurangi nilai akurasi disebabkan karena adanya penghilangan beberapa *term* pada pemotongan *threshold* yang mewakili suatu dokumen tertentu dan tidak digunakan dalam proses klasifikasi. Namun, meskipun demikian, penggunaan metode *DF-Thresholding* cukup efisien untuk mengurangi jumlah dimensi data ditunjukkan dengan pengurangan fitur sebesar 5746 pada *testing* akhir dengan hasil akurasi yang tidak jauh signifikan dari metode *Multinomial Naive Bayes*.

Penggunaan metode *TFIDF* dapat meningkatkan akurasi dilihat dari kenaikan nilai rata-rata akurasi metode *Multinomial* maupun dari kenaikan nilai rata-rata akurasi metode *Multinomial* dengan *DF-Threshold*.

Berikut ini merupakan contoh data *testing* yang tidak berhasil diklasifikasikan dengan baik.

Tabel 4.14 Contoh Data Berita yang Tidak Berhasil Diklasifikasikan

Teks Berita	<i>Term</i>	<i>TF</i>	<i>MNB</i>	<i>MNB-TFIDF</i>	<i>MNB-DF-Threshold</i>	<i>MNB-DF-TFIDF</i>
menteri luar negeri kemlu hari panggil kuasa usaha dubes china indonesia panggil kait kapal laku illegal fishing milik china km kway fey libat	menteri	1	Pajak	Politik Luar Negeri	Infrastruktur	Infrastruktur
	luar	1				
	negeri	1				
	kemlu	1				
	hari	1				
	panggil	2				
	kuasa	1				
	usaha	1				
	dubes	1				
	china	2				
	indonesia	1				



illegal fishing natuna pulau riau kepri	kait	1				
	kapal	1				
	laku	1				
	illegal	2				
	fishing	2				
	milik	1				
	km	1				
	kway	1				
	fey	1				
	libat	1				
	natuna	1				
	pulau	1				
	riau	1				
	kepri	1				

Pada data tersebut, kelas sebenarnya merupakan kelas “Politik Luar Negeri”. Sistem dengan *Multinomial Naive Bayes* dan *Multinomial Naive Bayes-DF Threshold* maupun *Multinomial Naive Bayes - DF-TFIDF* gagal mengklasifikasikan disebabkan karena adanya beberapa *term* pada kategori tersebut yang mewakili *term-term* data berita. *Term* tersebut dapat dilihat pada tabel berikut ini.

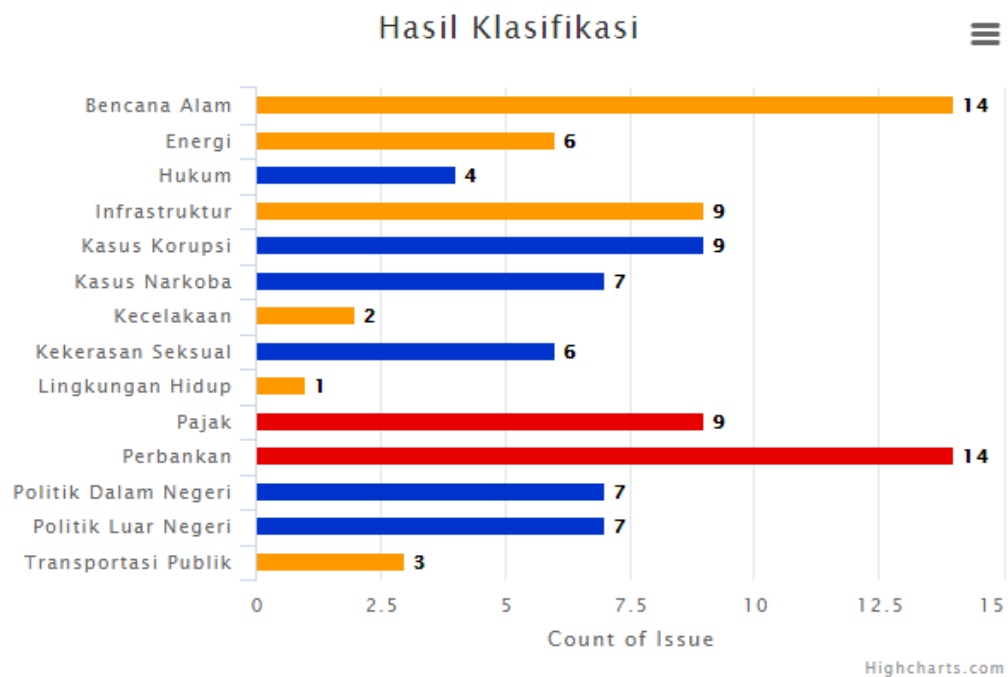
Tabel 4.15 Jumlah Term Kategori

<i>Term</i>	Jumlah <i>Term</i> pada Kategori -			<i>DF</i>	<i>IDF</i>
	Politik Luar Negeri	Pajak	Infrastruktur		
menteri	15	6	0	43	0.963129
Luar	32	8	0	34	1.06512
Negeri	14	7	0	33	1.07808
Kemlu	1	0	0	1	2.5966
Hari	13	1	0	46	0.933839
Usaha	0	22	0	32	1.09145
Dubes	2	0	0	2	2.29557
China	0	0	11	9	1.64235

indonesia	16	23	4	75	0.721536
Laku	6	2	3	59	0.825745
Milik	0	4	5	35	1.05253
Libat	0	1	0	9	1.64235

Pada perhitungan dengan metode *Multinomial Naive Bayes*, *term* pada data berita memiliki banyak kemiripan dengan *term-term* yang ada pada kategori Pajak, sehingga sistem mengklasifikan data tersebut sebagai kategori “Pajak”. Sementara itu, pada perhitungan dengan metode *Multinomial Naive Bayes-TFIDF*, *term-term* yang umum seperti “indonesia”, “laku” memiliki bobot yang lebih kecil dibandingkan *term-term* khusus seperti “kemlu” dan “dubes” yang hanya ada pada kategori “Politik Luar Negeri”, sehingga sistem mengklasifikasikan sesuai dengan kategori sebenarnya. Adapun untuk metode *Multinomial Naive Bayes-DF Threshold* maupun *Multinomial Naive Bayes - DF-TFIDF*, *term-term* yang memiliki nilai DF lebih dari 55 seperti *term* “indonesia” dan “laku” tidak diperhitungkan dalam proses klasifikasi. Selain itu, *term* “china” pada data berita berjumlah 2, sehingga probabilitas pada kelas “Infrastruktur” lebih tinggi dari kelas sebenarnya.

Contoh tampilan hasil dari klasifikasi dengan akurasi tertinggi yaitu menggunakan *Multinomial Naive Bayes* dengan *TFIDF* ditampilkan dalam bentuk diagram batang berikut ini.



Gambar 4.8 Tampilan Hasil Klasifikasi

Keterangan :

- Warna biru merupakan diagram untuk persebaran isu bidang Polhukam.
- Warna merah merupakan diagram untuk persebaran isu bidang Perekonomian.
- Warna kuning merupakan diagram untuk persebaran isu bidang Kesra.

## BAB V

### PENUTUP

#### 5.1 Kesimpulan

Dari hasil penelitian yang telah dilakukan dapat ditarik kesimpulan bahwa metode *Multinomial Naive Bayes* dapat digunakan untuk menganalisis berita dalam teks Bahasa Indonesia dengan cara pengelompokan ke dalam 15 kategori isu, yaitu Politik Luar Negeri, Politik Dalam Negeri, Keamanan, Kasus Narkoba, Kasus Korupsi, Kekerasan Seksual, Hukum, Pajak, Perbankan, Infrastruktur, Kecelakaan, Bencana Alam, Energi, Lingkungan Hidup dan Transportasi Publik, ditunjukkan dengan hasil akurasi akhir sebesar 94,29%. Penggunaan fitur seleksi dengan metode *DF-Thresholding* dapat mengurangi tingginya dimensi data, ditunjukkan dengan pengurangan fitur sebanyak 22.215 menjadi 16.766 dengan nilai akurasi akhir yang tidak jauh dari metode *Multinomial Naive Bayes* yaitu sebesar 93,33%. Sementara penggunaan *TFIDF* pada metode *Multinomial Naive Bayes* menunjukkan hasil akurasi akhir sebesar 94,29% dan dapat meningkatkan nilai rata-rata akurasi yang lebih besar dari metode *Multinomial Naive Bayes* dari 86,28% menjadi 86,62%. Penggunaan *DF-Threshold* pada *Multinomial Naive Bayes* dengan *TFIDF* pun menunjukkan nilai akurasi akhir yang lebih rendah dari *Multinomial Naive Bayes* dengan *TFIDF*, yaitu sebesar 92,38% namun dengan penggunaan jumlah fitur yang lebih sedikit dalam proses klasifikasi.

#### 5.2 Saran

Saran yang dapat diberikan untuk pengembangan pada penelitian selanjutnya:

1. Menambah fungsi untuk mengatur besar *threshold* pada seleksi fitur sehingga tahap pengujian dapat dilakukan pada beberapa *threshold* untuk mendapatkan nilai yang lebih maksimal.
2. Melakukan uji coba dengan beberapa metode seleksi fitur selain *DF-Thresholding* seperti *Mutual Information*, *Information Gain* dan *Chi Square*.