BAB 3
ANALISIS DAN PERANCANGAN

3.1 Analisis Sistem

Dalam proses pembuatan suatu sistem dilakukan analisis terhadap sistem yang akan dibangun. Analisis yang dilakukan untuk membangun aplikasi analisis sentimen pengguna twitter terhadap partai politik peserta PEMILU 2014 menggunakan pohon keputusan yang dibangun oleh algoritma C4.5.

3.1.1 Analisis Masalah

Masalah yang timbul dalam sistem analisis sentimen pengguna twitter terhadap partai politik peserta PEMILU 2014 adalah mengklasifikasikan *tweet* ke dalam opini positif dan negatif menggunakan algoritma C4.5.

3.1.2 Analisis Proses

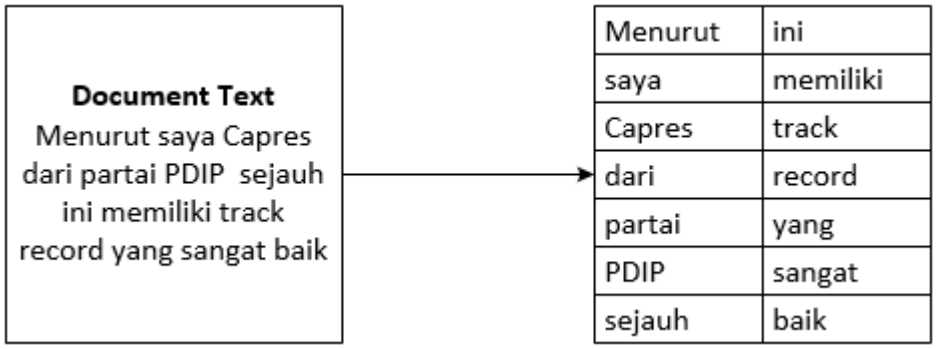
Analisis proses ini akan menjelaskan mengenai proses yang digunakan aplikasi analisis sentimen pengguna twitter terhadap partai politik peserta PEMILU 2014 menggunakan pohon keputusan yang dibangun oleh algoritma C4.5.

3.1.2.1 Proses Text Pre-processing

Proses *Text Pre-processing* dilakukan untuk mempersiapkan data sebelum ke proses selanjutnya, yaitu pemberian kelas kata untuk setiap kata dalam suatu *tweet* dan klasifikasi sentimen positif dan negatif. Terdapat beberapa sub proses dari *text pre-processing,* yaitu *tokenization, case folding,* dan *stemming* .

3.1.2.1.1 Tokenization

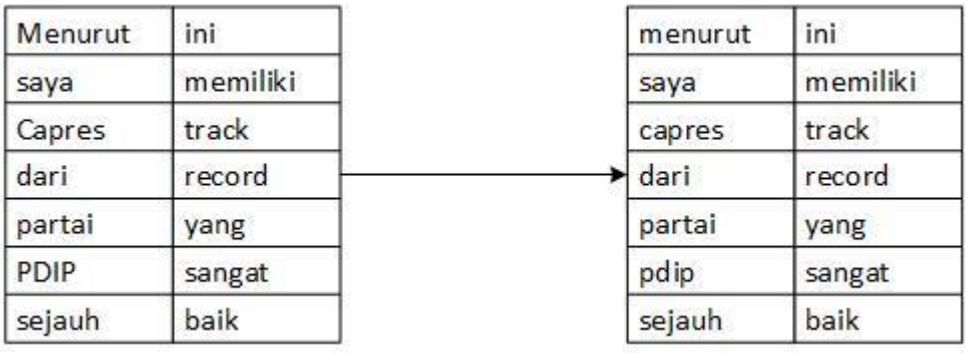
Tokenization dilakukan untuk memecah suatu kalimat dalam seluruh isi dokumen menjadi sekumpulan kata. Setiap kata yang telah dipecah dari kalimat/*tweet* akan disimpan ke basis data dengan atributnya yaitu id *tweet.*



Gambar 3.1 Tokenization

3.1.2.1.2 Casefolding

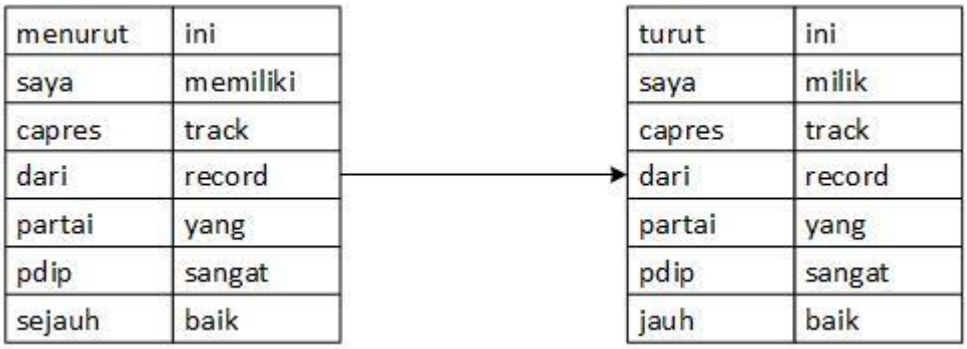
Case folding adalah proses mengubah semua huruf yang ada pada seluruh dokumen menjadi huruf kecil. Huruf "a" sampai dengan "z"yang akan diproses, karakter selain huruf akan dianggap *delimiter*. Proses ini dilakukan untuk menghindari duplikasi data sehingga kata yang memiliki perbedaan huruf kecil dan besar tetapi sama maka kata tersebut tidak akan disimpan.



Gambar 3.2 Casefolding

3.1.2.1.3 Stemming

Algoritma *stemming* yang digunakan pada penelitian ini adalah algoritma nazief dan adriani yang dibuat oleh Bobby Nazief dan Mirna Adriani



Gambar 3.3 Stemming

3.1.2.2 Proses Pembentukan Pohon Keputusan

Data hasi *text preprocessing* selanjutnya diklasifikasikan menggunakan pohon keputusan yang dibangun algoritma C4.5. Sebelum pohon keputusan digunakan untuk mengklasifikasikan ke data sebenarnya, model pohon keputusan belajar dari data *training* yang telah mempunyai kelas, yaitu negatif dan positif.

Pada seluruh dokumen *tweet* yang ada, setiap kata yang muncul dijadikan atribut untuk perhitungan entropi dan *gain.* nilai dari atribut tersebut adalah biner (0 dan 1)., dimana nilai 0 mengartikan tidak munculnya kata tersebut pada *tweet*, dan nilai 1 mengartikan bahwa kata tersebut muncul pada *tweet*.

Pada tabel III.1 adalah data latih yang digunakan dan tabel III.2 adalah implementasi perhitungan kasus algoritma C4.5 untuk membangun model pohon keputusan dengan atribut semua kata yang muncul dengan nilai 1 atau 0, dan atribut tujuan opini dengan nilai positif dan negatif.

Tabel 3.1 Training Data

No.	Training Data	Kelas
1	dukung caleg pejuang partai abc	Positif
2	masih muda sudah dibohongi partai abc	Negatif
3	tidak jadi Dukung partai abc	Negatif
4	Wiranto memiliki pergaulan internasional yang cukup baik.	Positif
5	gua bilang partai abc partai korupsi	Negatif
6	caleg dari partai abc bersih dari korupsi	Positif

7	yg siap dukung partai abc. Alhamdulillah dan smoga barokah. Amin.	Positif
8	dukung partai abc. tidak melakukan politk uang	Positif
9	partai abc Dukung Pemiskinan dan Hukum Mati Koruptor	Positif
10	tolak partai terkorupsi di Indonesia yaitu partai abc	Negatif
11	Mengajak saudaraku partai abc kembali baik. Amin ya Rob	Positif
12	oke cocok sama partai abc	Positif
13	Justru dg dukung Jokowi presiden diharapkan bs dorong kader2 muda partai abc yg bersih	Positif
14	partai abc itu bersih peduli tegas	Positif
15	Pemilih yg baik memilih politisi yg sdh teruji...	Positif
16	sikap partai abc menjadi hal yang tidak baik	Negatif
17	Ingin Indonesia lebih baik pilih partai abc	Positif
18	Koalisi ideal partai abc	Positif
19	saya anti partai partai abc	Negatif
20	disini banyak sekali yg tidak pilih partai abc	Negatif
21	jangan lupa pilih partai abc	Positif
22	kader partai abc yang anti korupsi	Positif
23	capres partai abc memliki elektabilitas tinggi	Positif
24	Pemimpin partai abc yang memeberikan harapan	Positif
25	capres pencitraan partai partai abc	Negatif
26	tanpa prabowo Indonesia damai partai abc	Negatif
27	yes mantap capres partai abc satu ini	Positif
28	partai abc mengutamakan intelektualitas tinggi calegnya	Positif
29	banyak caleg dengan intelektualitas jelek di partai partai abc	Negatif
30	prestasi ditonjolkan di partai partai abc	Positif
31	yakin dan percaya partai abc juaranya	Positif
32	Hasil survey menyebutkan partai abc tidak dipercaya	Positif
33	partai abc saya yakin prabowo bisa membuat indonesia semakin maju dan berkembang	Positif
34	partai abc yakin peroleh 10 persen kursi DPR	Positif
35	BAPAK jujur peduli tegas partai abc itu mah!	Positif
36	kader partai abc melakukan politik uang	Negatif
37	Cuma partai abc yg kadernya bisa membela tanpa masalah.	Positif
38	Saya tidak yakin partai abc akan selesaikan masalah RI	Negatif
39	Pendidikan Politik partai abc Berhasil	Positif
40	Elektabilitas partai abc Ungguli partai lain	Positif
41	kampanye partai abc kampanye partai yg ramah anak	Positif
42	partainya Gaul tapi sopan. partai abc !!!	Positif
43	partai abc tidak konsisten dalam pemilihan caleg	Negatif
44	Kalo saya mah pilih partai abc atuh, yg selalu konsisten	Positif
45	kader partai partai abc tidak mampu bersaing dengan partai lain	Negatif
46	partai partai abc masih tidak cocok untuk memimpin bangsa	Negatif
47	pasangan capres kubu partai abc mempunyai visi yang kuat	Positif
48	partai abc memiliki gagasan yang kurang baik mengenai	Negatif

	demokrasi	
49	visi misi partai partai abc kurang berpihak kepada rakyat	Negatif
50	ada yang tidak mendukung pencapresan dari partai partai abc	Negatif
51	partai abc punya gagasan sangat baik	Positif
52	kesalahan utama partai abc adalah sistem kampanye	Negatif
53	partai abc yang tidak terlibat kasus korupsi	Positif
54	kader partai abc berkomunikasi politik dengan baik	Positif
55	kita tidak mau dipimpin partai korupsi partai abc	Negatif
56	kader partai abc banyak yang tidak sopan berkampanye	Negatif
57	elektabilitas partai abc menurun	Negatif
58	strategi kampanye partai abc tidak berhasil	Negatif
59	partai abc anti pencitraan	Positif
60	partai abc sangat pro rakyat	Positif
61	visi partai abc tidak pro rakyat	Negatif
62	percaya partai abc mampu selesaikan masalah	Positif
63	siap menangkan partai abc	Positif
64	kita mantap pilih partai abc	Positif
65	elektabilitas partai abc tidak turun	Positif
66	banyak yang tak percaya kader capres partai abc bersih	Negatif
67	saya bukan pemilih partai abc	Negatif
68	partai abc bukan sarang koruptor	Positif
69	kader partai abc menolak politik uang	Positif
70	damai dengan pilih partai abc	Positif
71	partai abc kurang konsisten	Negatif
72	partai abc partai korupsi	Negatif
73	banyak yang menolak dukung partai abc	Negatif
74	partai abc gagal	Negatif
75	payah politik uang partai abc	Negatif
76	partai abc partai tak korupsi	Positif
77	partai abc bertahan tanpa korupsi	Positif
78	banyak kader partai abc terlibat korupsi	Negatif
79	partai abc kurang baik dalam berkampanye	Negatif
80	partai abc ga buruk lah	Positif
81	partai abc memiliki sistem yang baik	Positif
82	partai abc partai yang ga konsisten	Negatif
83	partai abc partai tidak korupsi	Positif
84	banyak kader yang ga baik di partai abc	Negatif
85	ga yakin deh sama partai abc	Negatif
86	kader partai abc tak pernah bohong	Positif
87	partai abc jujur !	Positif
88	saya mantap dukung partai abc	Positif
89	partai abc untuk perubahan	Positif
90	parlemen partai abc payah	Negatif
91	partai abc partai tak jujur	Negatif

92	partai abc tanpa pencitraan	Positif
93	yakin partai abc akan menang	Positif
94	partai abc tak kurang apapun	Positif
95	partai abc yang tidak berbohong	Positif
96	partai abc dipimpin pemimpin salah	Negatif
97	partai abc cermin politik tanpa pencitraan	Positif
98	politik jujur partai abc	Positif
99	sip semangat untuk partai abc	Positif
100	capres partai abc pemimpin yang dipercaya	Positif
101	partai abc ga baik untuk rakyat	Negatif
102	partai abc korupsi tertinggi	Negatif
103	partai abc adalah partai yang konsisten memberantas korupsi	Positif
104	Tapi Hanya partai abc yang konsisten	Positif
105	partai abc mengambil tindakan yang benar	Positif
106	partai abc tidak seburuk itu	Positif
107	tidak salah pilih partai abc	Positif
108	Komunikasi Politik partai abc Payah	Negatif
109	partai abc ga payah sebagai parpol	Positif
110	partai abc sarang koruptor	Negatif
111	partai abc kurangi korupsi	Positif
112	partai abc tolak korupsi	Positif
113	partai abc banyak yang ga jujur	Negatif
114	kader tidak jujur di partai abc	Negatif
115	partai abc , Ayo kita dukung !	Positif
116	partai abc partai yang ga salah	Positif
117	partai ga curang, partai abc lah!	Positif
118	partai abc tak pernah berbohong	Positif
119	partai abc tak pernah curang	Positif
120	partai abc penuh janji palsu	Negatif
121	kader partai abc anti korupsi	Positif
122	partai abc anti curang	Positif
123	kader partai abc anti politik uang	Positif
124	partai abc ? 1. Konsisten bela rakyat	Positif
125	partai abc ? 3. Anti Korupsi	Positif
126	partai abc ? 2. Berprestasi	Positif
127	partai abc tidak akan gagal	Positif
128	partai abc erat dengan politik uang	Negatif
129	partai abc tolak korupsi	Positif
130	tidak tolak dukung partai abc	Positif
131	tak tolak dukung partai abc	Positif
132	partai abc tak menolak kader korupsi	Negatif
133	partai abc tidak menolak kader korupsi	Negatif
134	partai abc menolak kader korupsi	Positif
135	partai abc tolak anti korupsi	Negatif

136	partai abc Jeblok karena kasus korupsi	Negatif
137	partai abc dikondisi yang buruk	Negatif
138	partai abc effect gagal.	Negatif
139	Yakin partai abc Konsisten	Positif
140	partai abc meningkat	Positif
141	partai abc curang	Negatif
142	partai abc kurang konsisten	Negatif
143	partai abc kurang konsisten memberantas korupsi	Negatif
144	partai abc tidak kurang baik	Positif
145	jangan pilih partai abc	Negatif
146	capres partai abc lembek	Negatif
147	partai abc tidak lembek tapi korupsi	Negatif
148	partai abc tidak lembek	Positif
149	partai abc tidak tegas	Negatif
150	partai abc fitnah partai lain	Negatif
151	pilih kader partai abc	Positif
152	partai abc ga !	Negatif
153	partai abc ga korupsi	Positif
154	partai abc partai anti suap	Positif
155	partai abc partai anti fitnah	Positif
156	partai abc lakukan kebaikan	Positif
157	partai bc anti damai	Negatif
158	partai abc ga bohong	Positif
159	partai abc jangan gentar	Positif
160	jangan mau dibohongi partai abc	Negatif

Proses pembentukan pohon keputusan dilakukan terhadap training data yang terdapat pada tabel *training data*. setiap atribut yang merupakan kata yang muncul pada seluruh *training data* akan memiliki nilai *information gain* dari hasil perhitungan tersebut. Atribut dengan nilai *information gain* tertinggi akan ditentukan sebuah simpul yang berisi kata pada atribut tersebut. Berikut tabel perhitungan simpul menggunakan algoritma C4.5.

Tabel 3.2 Perhitungan *Information Gain* Simpul Akar

Perhitungan Node Akar Total Entropy : 0.971							
Attribut	Neg(0)	Pos(0)	Entropy(0)	Neg(1)	Pos(1)	Entropy(1)	Gain
amanat	64	95	0.972	0	1	0	0.005
amin	64	94	0.974	0	2	0	0.009
anti	61	88	0.976	3	8	0.845	0.004
ayo	64	95	0.972	0	1	0	0.005
baik	59	87	0.973	5	9	0.94	0.001
bajak	64	95	0.972	0	1	0	0.005

bangsa	63	96	0.969	1	0	0	0.008
bapak	64	95	0.972	0	1	0	0.005
bela	64	94	0.974	0	2	0	0.009
benar	64	95	0.972	0	1	0	0.005
berantas	63	95	0.97	1	1	1	0.001
bersih	63	93	0.973	1	3	0.811	0.002
bisa	64	94	0.974	0	2	0	0.009
bohong	62	92	0.972	2	4	0.918	0.001
buat	64	95	0.972	0	1	0	0.005
bukan	63	95	0.97	1	1	1	0.001
buruk	63	94	0.972	1	2	0.918	0
citra	63	93	0.973	1	3	0.811	0.002
cocok	63	95	0.97	1	1	1	0.001
cukup	64	95	0.972	0	1	0	0.005
cuma	64	95	0.972	0	1	0	0.005
curang	63	93	0.973	1	3	0.811	0.002
damai	62	95	0.968	2	1	0.918	0.004
dena	63	96	0.969	1	0	0	0.008
dorong	64	95	0.972	0	1	0	0.005
dukung	61	87	0.978	3	9	0.811	0.006
fitnah	63	95	0.97	1	1	1	0.001
gagal	62	95	0.968	2	1	0.918	0.004
gagas	63	95	0.97	1	1	1	0.001
gaul	64	94	0.974	0	2	0	0.009
gentar	64	95	0.972	0	1	0	0.005
hara	64	96	0.971	0	0	0	0
harap	64	94	0.974	0	2	0	0.009
hasil	63	94	0.972	1	2	0.918	0
hukum	64	95	0.972	0	1	0	0.005
ideal	64	95	0.972	0	1	0	0.005
ingin	64	95	0.972	0	1	0	0.005
internasional	64	95	0.972	0	1	0	0.005
jangan	62	94	0.969	2	2	1	0.001
janji	63	96	0.969	1	0	0	0.008
jeblok	63	96	0.969	1	0	0	0.008
jelek	63	96	0.969	1	0	0	0.008
juang	64	95	0.972	0	1	0	0.005
juara	64	95	0.972	0	1	0	0.005
jujur	61	93	0.969	3	3	1	0.001
karena	63	96	0.969	1	0	0	0.008
kasus	63	95	0.97	1	1	1	0.001
koalisi	64	95	0.972	0	1	0	0.005
komunikasi	63	95	0.97	1	1	1	0.001
konsisten	59	91	0.967	5	5	1	0.002

korupsi	52	82	0.964	12	14	0.996	0.002
koruptor	63	94	0.972	1	2	0.918	0
kuat	64	95	0.972	0	1	0	0.005
kubu	64	95	0.972	0	1	0	0.005
kurang	58	93	0.961	6	3	0.918	0.012
kursi	64	95	0.972	0	1	0	0.005
laku	63	94	0.972	1	2	0.918	0
lebih	64	95	0.972	0	1	0	0.005
lembek	62	95	0.968	2	1	0.918	0.004
libat	63	95	0.97	1	1	1	0.001
lupa	64	95	0.972	0	1	0	0.005
maju	64	95	0.972	0	1	0	0.005
makin	64	95	0.972	0	1	0	0.005
mampu	63	95	0.97	1	1	1	0.001
mantap	64	93	0.975	0	3	0	0.014
masalah	63	94	0.972	1	2	0.918	0
mati	64	95	0.972	0	1	0	0.005
mau	62	96	0.966	2	0	0	0.017
menang	64	95	0.972	0	1	0	0.005
misi	63	96	0.969	1	0	0	0.008
nasional	64	95	0.972	0	1	0	0.005
palsu	63	96	0.969	1	0	0	0.008
parlemen	63	96	0.969	1	0	0	0.008
pasang	64	95	0.972	0	1	0	0.005
payah	61	95	0.965	3	1	0.811	0.01
peduli	64	94	0.974	0	2	0	0.009
percaya	63	92	0.975	1	4	0.722	0.004
pernah	64	93	0.975	0	3	0	0.014
persen	64	95	0.972	0	1	0	0.005
pihak	63	96	0.969	1	0	0	0.008
pilih	60	88	0.974	4	8	0.918	0.001
pimpin	61	94	0.967	3	2	0.971	0.004
politik	60	90	0.971	4	6	0.971	0
politis	64	95	0.972	0	1	0	0.005
presiden	64	95	0.972	0	1	0	0.005
prestasi	64	94	0.974	0	2	0	0.009
pro	63	95	0.97	1	1	1	0.001
punya	64	94	0.974	0	2	0	0.009
rakyat	61	94	0.967	3	2	0.971	0.004
ramah	64	95	0.972	0	1	0	0.005
saing	63	96	0.969	1	0	0	0.008
salah	62	94	0.969	2	2	1	0.001
sambil	64	95	0.972	0	1	0	0.005
sangat	64	94	0.974	0	2	0	0.009

sarang	63	95	0.97	1	1	1	0.001
saudara	64	95	0.972	0	1	0	0.005
sejahtera	64	95	0.972	0	1	0	0.005
sekal	63	96	0.969	1	0	0	0.008
selalu	64	95	0.972	0	1	0	0.005
selesai	63	95	0.97	1	1	1	0.001
semangat	64	95	0.972	0	1	0	0.005
siap	64	94	0.974	0	2	0	0.009
sikap	63	96	0.969	1	0	0	0.008
sip	64	95	0.972	0	1	0	0.005
sistem	63	95	0.97	1	1	1	0.001
strategi	63	96	0.969	1	0	0	0.008
suap	64	95	0.972	0	1	0	0.005
tahan	64	95	0.972	0	1	0	0.005
tak	61	90	0.973	3	6	0.918	0.001
tanpa	63	92	0.975	1	4	0.722	0.004
tapi	63	94	0.972	1	2	0.918	0
tegas	63	94	0.972	1	2	0.918	0
tidak	48	84	0.946	16	12	0.985	0.018
tinggi	63	94	0.972	1	2	0.918	0
tingkat	64	95	0.972	0	1	0	0.005
tolak	59	90	0.969	5	6	0.994	0
tonjol	64	95	0.972	0	1	0	0.005
turun	63	95	0.97	1	1	1	0.001
uang	61	93	0.969	3	3	1	0.001
ubah	64	95	0.972	0	1	0	0.005
uji	64	95	0.972	0	1	0	0.005
unggul	64	95	0.972	0	1	0	0.005
utama	63	95	0.97	1	1	1	0.001
visi	62	95	0.968	2	1	0.918	0.004
yakin	62	91	0.974	2	5	0.863	0.002
ga	58	90	0.966	6	6	1	0.002
Max Gain : 0.018 Node : tidak							

Setelah perhitungan untuk menentukan simpul berakhir maka selanjutnya ditentukan atribut dengan nilai gain tertinggi. Nilai gain tertinggi pada perhitungan simpuln adalah atribut tes dengan kata tidak dengan nilai 0.018.



Gambar 3.4 Pembentukan Pohon Keputusan Simpul Akar

Nilai entropy untuk atribut tes dengan kata tidak memiliki nilai tidak sama dengan nol. Maka simpul tersebut masih memiliki cabang. Selanjutnya akan dilakukan perhitungan nilai gain untuk menentukan simpul cabang.

Tabel 3.3 Perhitungan *Information Gain* Simpul Cabang Tidak

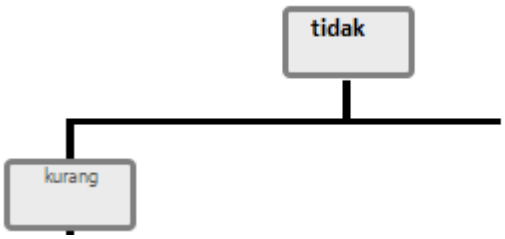
Perhitungan Cabang Node 'tidak' Value '0' Total Entropy : 0.943							
Attribut	Neg(0)	Pos(0)	Entropy(0)	Neg(1)	Pos(1)	Entropy(1)	Gain
amanah	49	85	0.947	0	2	0	0.01
amanat	49	86	0.945	0	1	0	0.005
amin	49	86	0.945	0	1	0	0.005
anak	49	86	0.945	0	1	0	0.005
anti	46	79	0.949	3	8	0.845	0.002
ayo	49	86	0.945	0	1	0	0.005
bagai	49	86	0.945	0	1	0	0.005
baik	45	79	0.945	4	8	0.918	0
bajak	49	86	0.945	0	1	0	0.005
bangsa	49	87	0.943	0	0	0	0
bapak	49	86	0.945	0	1	0	0.005
bela	49	85	0.947	0	2	0	0.01
benar	49	86	0.945	0	1	0	0.005
berantas	48	86	0.941	1	1	1	0.001
bersih	48	85	0.943	1	2	0.918	0.001
bisa	49	85	0.947	0	2	0	0.01
bohong	47	84	0.942	2	3	0.971	-0
buat	49	86	0.945	0	1	0	0.005
bukan	48	86	0.941	1	1	1	0.001
buruk	48	86	0.941	1	1	1	0.001
cena	48	87	0.939	1	0	0	0.011
cermin	49	86	0.945	0	1	0	0.005
citra	48	84	0.946	1	3	0.811	0.001
cocok	49	86	0.945	0	1	0	0.005
cukup	49	86	0.945	0	1	0	0.005
cuma	49	86	0.945	0	1	0	0.005
curang	48	84	0.946	1	3	0.811	0.001
damai	47	86	0.937	2	1	0.918	0.006
dukung	48	82	0.95	1	5	0.65	0.006
fitnah	48	86	0.941	1	1	1	0.001
gagal	47	87	0.935	2	0	0	0.022
gagas	48	86	0.941	1	1	1	0.001
gaul	49	85	0.947	0	2	0	0.01
gentar	49	86	0.945	0	1	0	0.005
hara	49	87	0.943	0	0	0	0
hasil	49	86	0.945	0	1	0	0.005

ideal	49	86	0.945	0	1	0	0.005
ingin	49	86	0.945	0	1	0	0.005
internasional	49	86	0.945	0	1	0	0.005
jangan	47	85	0.939	2	2	1	0.002
janji	48	87	0.939	1	0	0	0.011
jeblok	48	87	0.939	1	0	0	0.011
jelek	48	87	0.939	1	0	0	0.011
juang	49	86	0.945	0	1	0	0.005
juara	49	86	0.945	0	1	0	0.005
jujur	47	84	0.942	2	3	0.971	-0
kalo	49	86	0.945	0	1	0	0.005
karena	48	87	0.939	1	0	0	0.011
kasus	48	87	0.939	1	0	0	0.011
koalisi	49	86	0.945	0	1	0	0.005
komunikasi	48	86	0.941	1	1	1	0.001
konsisten	45	82	0.938	4	5	0.991	0.001
korupsi	40	75	0.932	9	12	0.985	0.003
koruptor	48	86	0.941	1	1	1	0.001
kuat	49	86	0.945	0	1	0	0.005
kubu	49	86	0.945	0	1	0	0.005
kurang	43	85	0.921	6	2	0.811	0.028
kursi	49	86	0.945	0	1	0	0.005
lain	48	86	0.941	1	1	1	0.001
laku	48	86	0.941	1	1	1	0.001
langgar	48	87	0.939	1	0	0	0.011
lebih	49	86	0.945	0	1	0	0.005
lembek	48	87	0.939	1	0	0	0.011
libat	48	87	0.939	1	0	0	0.011
lupa	49	86	0.945	0	1	0	0.005
maju	49	86	0.945	0	1	0	0.005
makin	49	86	0.945	0	1	0	0.005
mampu	49	86	0.945	0	1	0	0.005
manfaat	49	86	0.945	0	1	0	0.005
mantap	49	84	0.949	0	3	0	0.015
masalah	49	85	0.947	0	2	0	0.01
mau	48	87	0.939	1	0	0	0.011
menang	49	86	0.945	0	1	0	0.005
misi	48	87	0.939	1	0	0	0.011
nasional	49	86	0.945	0	1	0	0.005
palsu	48	87	0.939	1	0	0	0.011
parlemen	48	87	0.939	1	0	0	0.011
pasang	49	86	0.945	0	1	0	0.005
payah	46	86	0.933	3	1	0.811	0.014
peduli	49	86	0.945	0	1	0	0.005

percaya	48	84	0.946	1	3	0.811	0.001
pernah	49	84	0.949	0	3	0	0.015
persen	49	86	0.945	0	1	0	0.005
pihak	48	87	0.939	1	0	0	0.011
pilih	47	80	0.951	2	7	0.764	0.004
pimpin	48	85	0.943	1	2	0.918	0.001
politik	45	81	0.94	4	6	0.971	0.001
politis	49	86	0.945	0	1	0	0.005
prestasi	49	85	0.947	0	2	0	0.01
pro	49	86	0.945	0	1	0	0.005
punya	49	85	0.947	0	2	0	0.01
rakyat	47	85	0.939	2	2	1	0.002
ramah	49	86	0.945	0	1	0	0.005
saing	49	87	0.943	0	0	0	0
salah	47	86	0.937	2	1	0.918	0.006
sambil	49	86	0.945	0	1	0	0.005
sangat	49	85	0.947	0	2	0	0.01
santun	49	86	0.945	0	1	0	0.005
sarang	48	86	0.941	1	1	1	0.001
satu	49	86	0.945	0	1	0	0.005
saudara	49	86	0.945	0	1	0	0.005
sejahtera	49	86	0.945	0	1	0	0.005
sekal	49	87	0.943	0	0	0	0
selalu	49	86	0.945	0	1	0	0.005
selesai	49	86	0.945	0	1	0	0.005
semangat	49	86	0.945	0	1	0	0.005
siap	49	86	0.945	0	1	0	0.005
sikap	49	87	0.943	0	0	0	0
sip	49	86	0.945	0	1	0	0.005
sistem	48	86	0.941	1	1	1	0.001
strategi	49	87	0.943	0	0	0	0
suap	49	86	0.945	0	1	0	0.005
tahan	49	86	0.945	0	1	0	0.005
tak	46	81	0.944	3	6	0.918	0.001
tanpa	48	83	0.948	1	4	0.722	0.003
tapi	49	85	0.947	0	2	0	0.01
tegas	49	85	0.947	0	2	0	0.01
tinggi	48	85	0.943	1	2	0.918	0.001
tingkat	49	86	0.945	0	1	0	0.005
tolak	45	82	0.938	4	5	0.991	0.001
tonjol	49	86	0.945	0	1	0	0.005
turun	48	87	0.939	1	0	0	0.011
uang	46	85	0.935	3	2	0.971	0.007
ubah	49	86	0.945	0	1	0	0.005

uji	49	86	0.945	0	1	0	0.005
unggul	49	86	0.945	0	1	0	0.005
utama	48	86	0.941	1	1	1	0.001
visi	48	86	0.941	1	1	1	0.001
yakin	48	82	0.95	1	5	0.65	0.006
ga	43	81	0.931	6	6	1	0.006
Max Gain : 0.028 Node : kurang							

Setelah perhitungan untuk menentukan simpul berakhir maka selanjutnya ditentukan atribut dengan nilai gain tertinggi. Nilai gain tertinggi pada perhitungan simpul adalah atribut tes dengan kata kurang dengan nilai 0.028.



Gambar 3.5 Pembentukan Pohon Keputusan Cabang Tidak

Nilai entropy untuk atribut tes dengan kata kurang memiliki nilai tidak sama dengan nol. Maka simpul tersebut masih memiliki cabang. Selanjutnya akan dilakukan perhitungan nilai gain untuk menentukan simpul cabang.

Tabel 3.4 Perhitungan *Information Gain* Simpul Cabang Kurang

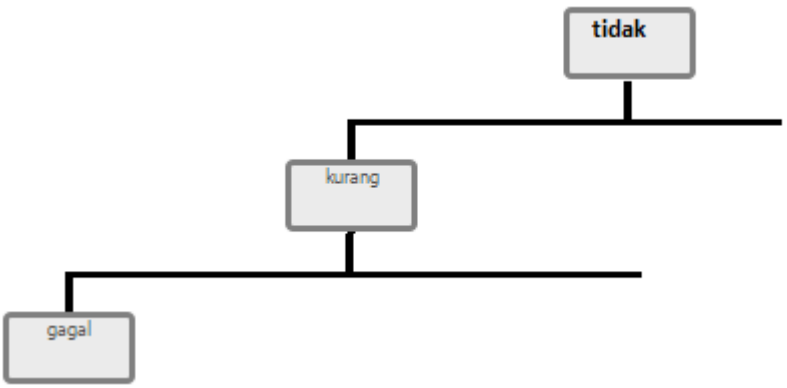
Perhitungan Cabang Node 'kurang' Value '0' Total Entropy : 0.921							
Attribut	Neg(0)	Pos(0)	Entropy(0)	Neg(1)	Pos(1)	Entropy(1)	Gain
amanah	43	83	0.926	0	2	0	0.009
amanat	43	84	0.923	0	1	0	0.005
amin	43	84	0.923	0	1	0	0.005
anak	43	84	0.923	0	1	0	0.005
anti	40	77	0.927	3	8	0.845	0.001
ayo	43	84	0.923	0	1	0	0.005
bagai	43	84	0.923	0	1	0	0.005
baik	41	77	0.932	2	8	0.722	0.005
bajak	43	84	0.923	0	1	0	0.005
bangsa	43	85	0.921	0	0	0	0
bapak	43	84	0.923	0	1	0	0.005
bela	43	83	0.926	0	2	0	0.009
benar	43	84	0.923	0	1	0	0.005
berantas	43	84	0.923	0	1	0	0.005

bersih	42	83	0.921	1	2	0.918	0
bisa	43	83	0.926	0	2	0	0.009
bohong	41	82	0.918	2	3	0.971	0.001
buat	43	84	0.923	0	1	0	0.005
bukan	42	84	0.918	1	1	1	0.002
buruk	42	84	0.918	1	1	1	0.002
cena	43	85	0.921	0	0	0	0
cermin	43	84	0.923	0	1	0	0.005
citra	42	82	0.924	1	3	0.811	0.001
cocok	43	84	0.923	0	1	0	0.005
cukup	43	84	0.923	0	1	0	0.005
cuma	43	84	0.923	0	1	0	0.005
curang	42	82	0.924	1	3	0.811	0.001
damai	41	84	0.913	2	1	0.918	0.008
dukung	42	80	0.929	1	5	0.65	0.005
fitnah	42	84	0.918	1	1	1	0.002
gagal	41	85	0.91	2	0	0	0.025
gagas	43	84	0.923	0	1	0	0.005
gaul	43	83	0.926	0	2	0	0.009
gentar	43	84	0.923	0	1	0	0.005
hara	43	85	0.921	0	0	0	0
hasil	43	84	0.923	0	1	0	0.005
ideal	43	84	0.923	0	1	0	0.005
ingin	43	84	0.923	0	1	0	0.005
internasional	43	84	0.923	0	1	0	0.005
jangan	41	83	0.916	2	2	1	0.002
janji	42	85	0.916	1	0	0	0.012
jeblok	42	85	0.916	1	0	0	0.012
jelek	42	85	0.916	1	0	0	0.012
juang	43	84	0.923	0	1	0	0.005
juara	43	84	0.923	0	1	0	0.005
jujur	41	82	0.918	2	3	0.971	0.001
kalo	43	84	0.923	0	1	0	0.005
karena	42	85	0.916	1	0	0	0.012
kasus	42	85	0.916	1	0	0	0.012
koalisi	43	84	0.923	0	1	0	0.005
komunikasi	42	84	0.918	1	1	1	0.002
konsisten	42	80	0.929	1	5	0.65	0.005
korupsi	35	74	0.906	8	11	0.982	0.004
koruptor	42	84	0.918	1	1	1	0.002
kuat	43	84	0.923	0	1	0	0.005
kubu	43	84	0.923	0	1	0	0.005
kursi	43	84	0.923	0	1	0	0.005
lain	42	84	0.918	1	1	1	0.002

laku	42	84	0.918	1	1	1	0.002
langgar	42	85	0.916	1	0	0	0.012
lebih	43	84	0.923	0	1	0	0.005
lembek	42	85	0.916	1	0	0	0.012
libat	42	85	0.916	1	0	0	0.012
lupa	43	84	0.923	0	1	0	0.005
maju	43	84	0.923	0	1	0	0.005
makin	43	84	0.923	0	1	0	0.005
mampu	43	84	0.923	0	1	0	0.005
manfaat	43	84	0.923	0	1	0	0.005
mantap	43	82	0.929	0	3	0	0.014
masalah	43	83	0.926	0	2	0	0.009
mau	42	85	0.916	1	0	0	0.012
menang	43	84	0.923	0	1	0	0.005
misi	43	85	0.921	0	0	0	0
nasional	43	84	0.923	0	1	0	0.005
palsu	42	85	0.916	1	0	0	0.012
parlemen	42	85	0.916	1	0	0	0.012
pasang	43	84	0.923	0	1	0	0.005
payah	40	84	0.907	3	1	0.811	0.017
peduli	43	84	0.923	0	1	0	0.005
percaya	42	82	0.924	1	3	0.811	0.001
pernah	43	82	0.929	0	3	0	0.014
persen	43	84	0.923	0	1	0	0.005
pihak	43	85	0.921	0	0	0	0
pilih	41	78	0.929	2	7	0.764	0.004
pimpin	42	83	0.921	1	2	0.918	0
politik	39	79	0.915	4	6	0.971	0.002
politis	43	84	0.923	0	1	0	0.005
prestasi	43	83	0.926	0	2	0	0.009
pro	43	84	0.923	0	1	0	0.005
punya	43	83	0.926	0	2	0	0.009
rakyat	42	83	0.921	1	2	0.918	0
ramah	43	84	0.923	0	1	0	0.005
saing	43	85	0.921	0	0	0	0
salah	41	84	0.913	2	1	0.918	0.008
sambil	43	84	0.923	0	1	0	0.005
sangat	43	83	0.926	0	2	0	0.009
santun	43	84	0.923	0	1	0	0.005
sarang	42	84	0.918	1	1	1	0.002
satu	43	84	0.923	0	1	0	0.005
saudara	43	84	0.923	0	1	0	0.005
sejahtera	43	84	0.923	0	1	0	0.005
sekal	43	85	0.921	0	0	0	0

selalu	43	84	0.923	0	1	0	0.005
selesai	43	84	0.923	0	1	0	0.005
semangat	43	84	0.923	0	1	0	0.005
siap	43	84	0.923	0	1	0	0.005
sikap	43	85	0.921	0	0	0	0
sip	43	84	0.923	0	1	0	0.005
sistem	42	84	0.918	1	1	1	0.002
strategi	43	85	0.921	0	0	0	0
suap	43	84	0.923	0	1	0	0.005
tahan	43	84	0.923	0	1	0	0.005
tak	40	80	0.918	3	5	0.954	0.001
tanpa	42	81	0.926	1	4	0.722	0.003
tapi	43	83	0.926	0	2	0	0.009
tegas	43	83	0.926	0	2	0	0.009
tinggi	42	83	0.921	1	2	0.918	0
tingkat	43	84	0.923	0	1	0	0.005
tolak	39	80	0.913	4	5	0.991	0.003
tonjol	43	84	0.923	0	1	0	0.005
turun	42	85	0.916	1	0	0	0.012
uang	40	83	0.91	3	2	0.971	0.009
ubah	43	84	0.923	0	1	0	0.005
uji	43	84	0.923	0	1	0	0.005
unggul	43	84	0.923	0	1	0	0.005
utama	42	84	0.918	1	1	1	0.002
visi	43	84	0.923	0	1	0	0.005
yakin	42	80	0.929	1	5	0.65	0.005
ga	37	79	0.903	6	6	1	0.009
Max Gain : 0.025 Node : gagal							

Setelah perhitungan untuk menentukan simpul berakhir maka selanjutnya ditentukan atribut dengan nilai gain tertinggi. Nilai gain tertinggi pada perhitungan simpul adalah atribut tes dengan kata gagal dengan nilai 0.025.



Gambar 3.6 Pembentukan Pohon Keputusan Cabang Kurang

Nilai entropy untuk atribut tes dengan kata gagal memiliki nilai tidak sama dengan nol. Maka simpul tersebut masih memiliki cabang. Selanjutnya akan dilakukan perhitungan nilai gain untuk menentukan simpul cabang.

Tabel 3.5 Perhitungan *Information Gain* Simpul Cabang Gagal

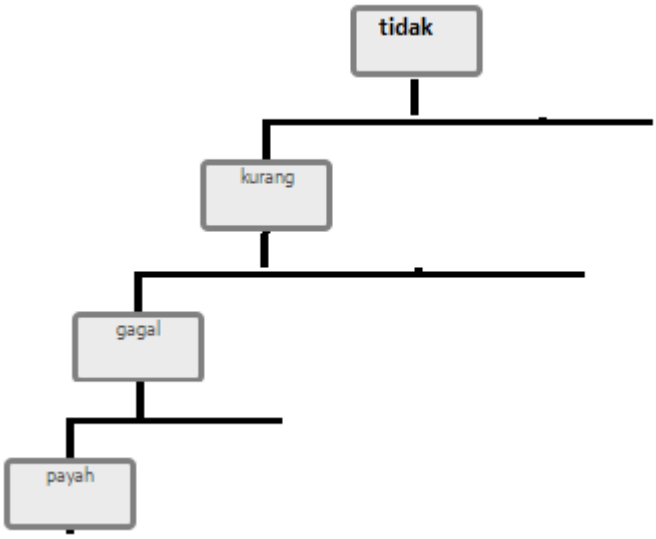
Perhitungan Cabang Node 'gagal' Value '0' Total Entropy : 0.91							
Attribut	Neg(0)	Pos(0)	Entropy(0)	Neg(1)	Pos(1)	Entropy(1)	Gain
amanah	41	83	0.916	0	2	0	0.009
amanat	41	84	0.913	0	1	0	0.004
amin	41	84	0.913	0	1	0	0.004
anak	41	84	0.913	0	1	0	0.004
anti	38	77	0.915	3	8	0.845	0.001
ayo	41	84	0.913	0	1	0	0.004
bagai	41	84	0.913	0	1	0	0.004
baik	39	77	0.921	2	8	0.722	0.005
bajak	41	84	0.913	0	1	0	0.004
bangsa	41	85	0.91	0	0	0	0
bapak	41	84	0.913	0	1	0	0.004
bela	41	83	0.916	0	2	0	0.009
benar	41	84	0.913	0	1	0	0.004
berantas	41	84	0.913	0	1	0	0.004
bersih	40	83	0.91	1	2	0.918	-0
bisa	41	83	0.916	0	2	0	0.009
bohong	39	82	0.907	2	3	0.971	0
buat	41	84	0.913	0	1	0	0.004
bukan	40	84	0.907	1	1	1	0.002
buruk	40	84	0.907	1	1	1	0.002
cena	41	85	0.91	0	0	0	0
cermin	41	84	0.913	0	1	0	0.004
citra	40	82	0.913	1	3	0.811	0
cocok	41	84	0.913	0	1	0	0.004
cukup	41	84	0.913	0	1	0	0.004
cuma	41	84	0.913	0	1	0	0.004
curang	40	82	0.913	1	3	0.811	0
damai	39	84	0.901	2	1	0.918	0.009
dukung	40	80	0.918	1	5	0.65	0.005
fitnah	40	84	0.907	1	1	1	0.002
gagas	41	84	0.913	0	1	0	0.004
gaul	41	83	0.916	0	2	0	0.009
gentar	41	84	0.913	0	1	0	0.004
hara	41	85	0.91	0	0	0	0
hasil	41	84	0.913	0	1	0	0.004
ideal	41	84	0.913	0	1	0	0.004

ingin	41	84	0.913	0	1	0	0.004
internasional	41	84	0.913	0	1	0	0.004
jangan	39	83	0.904	2	2	1	0.003
janji	40	85	0.904	1	0	0	0.013
jeblok	40	85	0.904	1	0	0	0.013
jelek	40	85	0.904	1	0	0	0.013
juang	41	84	0.913	0	1	0	0.004
juara	41	84	0.913	0	1	0	0.004
jujur	39	82	0.907	2	3	0.971	0
kalo	41	84	0.913	0	1	0	0.004
karena	40	85	0.904	1	0	0	0.013
kasus	40	85	0.904	1	0	0	0.013
koalisi	41	84	0.913	0	1	0	0.004
komunikasi	40	84	0.907	1	1	1	0.002
konsisten	40	80	0.918	1	5	0.65	0.005
korupsi	33	74	0.891	8	11	0.982	0.005
koruptor	40	84	0.907	1	1	1	0.002
kuat	41	84	0.913	0	1	0	0.004
kubu	41	84	0.913	0	1	0	0.004
kursi	41	84	0.913	0	1	0	0.004
lain	40	84	0.907	1	1	1	0.002
laku	40	84	0.907	1	1	1	0.002
langgar	40	85	0.904	1	0	0	0.013
lebih	41	84	0.913	0	1	0	0.004
lembek	40	85	0.904	1	0	0	0.013
libat	40	85	0.904	1	0	0	0.013
lupa	41	84	0.913	0	1	0	0.004
maju	41	84	0.913	0	1	0	0.004
makin	41	84	0.913	0	1	0	0.004
mampu	41	84	0.913	0	1	0	0.004
manfaat	41	84	0.913	0	1	0	0.004
mantap	41	82	0.918	0	3	0	0.014
masalah	41	83	0.916	0	2	0	0.009
mau	40	85	0.904	1	0	0	0.013
menang	41	84	0.913	0	1	0	0.004
misi	41	85	0.91	0	0	0	0
nasional	41	84	0.913	0	1	0	0.004
palsu	40	85	0.904	1	0	0	0.013
parlemen	40	85	0.904	1	0	0	0.013
pasang	41	84	0.913	0	1	0	0.004
payah	38	84	0.895	3	1	0.811	0.018
peduli	41	84	0.913	0	1	0	0.004
percaya	40	82	0.913	1	3	0.811	0
pernah	41	82	0.918	0	3	0	0.014

persen	41	84	0.913	0	1	0	0.004
pihak	41	85	0.91	0	0	0	0
pilih	39	78	0.918	2	7	0.764	0.003
pimpin	40	83	0.91	1	2	0.918	-0
politik	37	79	0.903	4	6	0.971	0.002
politis	41	84	0.913	0	1	0	0.004
prestasi	41	83	0.916	0	2	0	0.009
pro	41	84	0.913	0	1	0	0.004
punya	41	83	0.916	0	2	0	0.009
rakyat	40	83	0.91	1	2	0.918	-0
ramah	41	84	0.913	0	1	0	0.004
saing	41	85	0.91	0	0	0	0
salah	39	84	0.901	2	1	0.918	0.009
sambil	41	84	0.913	0	1	0	0.004
sangat	41	83	0.916	0	2	0	0.009
santun	41	84	0.913	0	1	0	0.004
sarang	40	84	0.907	1	1	1	0.002
satu	41	84	0.913	0	1	0	0.004
saudara	41	84	0.913	0	1	0	0.004
sejahtera	41	84	0.913	0	1	0	0.004
sekal	41	85	0.91	0	0	0	0
selalu	41	84	0.913	0	1	0	0.004
selesai	41	84	0.913	0	1	0	0.004
semangat	41	84	0.913	0	1	0	0.004
siap	41	84	0.913	0	1	0	0.004
sikap	41	85	0.91	0	0	0	0
sip	41	84	0.913	0	1	0	0.004
sistem	40	84	0.907	1	1	1	0.002
strategi	41	85	0.91	0	0	0	0
suap	41	84	0.913	0	1	0	0.004
tahan	41	84	0.913	0	1	0	0.004
tak	38	80	0.907	3	5	0.954	0
tanpa	40	81	0.916	1	4	0.722	0.002
tapi	41	83	0.916	0	2	0	0.009
tegas	41	83	0.916	0	2	0	0.009
tinggi	40	83	0.91	1	2	0.918	-0
tingkat	41	84	0.913	0	1	0	0.004
tolak	37	80	0.9	4	5	0.991	0.003
tonjol	41	84	0.913	0	1	0	0.004
turun	40	85	0.904	1	0	0	0.013
uang	38	83	0.898	3	2	0.971	0.009
ubah	41	84	0.913	0	1	0	0.004
uji	41	84	0.913	0	1	0	0.004
unggul	41	84	0.913	0	1	0	0.004

utama	40	84	0.907	1	1	1	0.002
visi	41	84	0.913	0	1	0	0.004
yakin	40	80	0.918	1	5	0.65	0.005
ga	35	79	0.89	6	6	1	0.01
Max Gain : 0.018 Node : payah							

Setelah perhitungan untuk menentukan simpul berakhir maka selanjutnya ditentukan atribut dengan nilai gain tertinggi. Nilai gain tertinggi pada perhitungan simpul adalah atribut tes dengan kata payah dengan nilai 0.018.



Gambar 3.7 Perhitungan Pohon Keputusan Cabang Gagal

Nilai entropy untuk atribut tes dengan kata payah memiliki nilai tidak sama dengan nol. Maka simpul tersebut masih memiliki cabang. Selanjutnya akan dilakukan perhitungan nilai gain untuk menentukan simpul cabang.

Tabel 3.6 Perhitungan *Information Gain* Simpul Cabang Payah

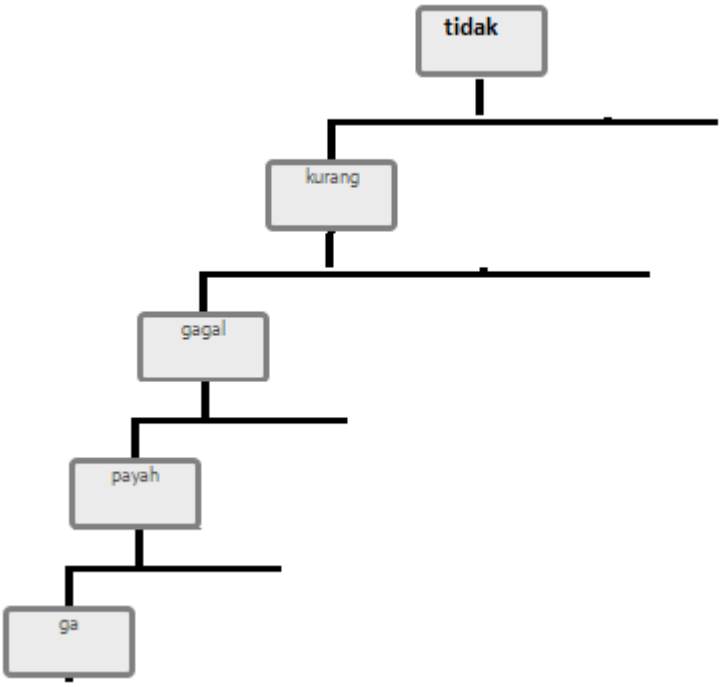
Perhitungan Cabang Node 'payah' Value '0' Total Entropy : 0.895							
Attribut	Neg(0)	Pos(0)	Entropy(0)	Neg(1)	Pos(1)	Entropy(1)	Gain
amanah	38	82	0.901	0	2	0	0.009
amanat	38	83	0.898	0	1	0	0.004
amin	38	83	0.898	0	1	0	0.004
anak	38	83	0.898	0	1	0	0.004
anti	35	76	0.899	3	8	0.845	0.001
ayo	38	83	0.898	0	1	0	0.004
bagai	38	84	0.895	0	0	0	0
baik	36	76	0.906	2	8	0.722	0.004

bajak	38	83	0.898	0	1	0	0.004
bangsa	38	84	0.895	0	0	0	0
bapak	38	83	0.898	0	1	0	0.004
bela	38	82	0.901	0	2	0	0.009
benar	38	83	0.898	0	1	0	0.004
berantas	38	83	0.898	0	1	0	0.004
bersih	37	82	0.894	1	2	0.918	0
bisa	38	82	0.901	0	2	0	0.009
bohong	36	81	0.89	2	3	0.971	0.002
buat	38	83	0.898	0	1	0	0.004
bukan	37	83	0.891	1	1	1	0.002
buruk	37	83	0.891	1	1	1	0.002
cena	38	84	0.895	0	0	0	0
cermin	38	83	0.898	0	1	0	0.004
citra	37	81	0.897	1	3	0.811	0.001
cocok	38	83	0.898	0	1	0	0.004
cukup	38	83	0.898	0	1	0	0.004
cuma	38	83	0.898	0	1	0	0.004
curang	37	81	0.897	1	3	0.811	0.001
damai	36	83	0.884	2	1	0.918	0.01
dukung	37	79	0.903	1	5	0.65	0.004
fitnah	37	83	0.891	1	1	1	0.002
gagas	38	83	0.898	0	1	0	0.004
gaul	38	82	0.901	0	2	0	0.009
gentar	38	83	0.898	0	1	0	0.004
hara	38	84	0.895	0	0	0	0
hasil	38	83	0.898	0	1	0	0.004
ideal	38	83	0.898	0	1	0	0.004
ingin	38	83	0.898	0	1	0	0.004
internasional	38	83	0.898	0	1	0	0.004
jangan	36	82	0.887	2	2	1	0.004
janji	37	84	0.888	1	0	0	0.014
jeblok	37	84	0.888	1	0	0	0.014
jelek	37	84	0.888	1	0	0	0.014
juang	38	83	0.898	0	1	0	0.004
juara	38	83	0.898	0	1	0	0.004
jujur	36	81	0.89	2	3	0.971	0.002
kalo	38	83	0.898	0	1	0	0.004
karena	37	84	0.888	1	0	0	0.014
kasus	37	84	0.888	1	0	0	0.014
koalisi	38	83	0.898	0	1	0	0.004
komunikasi	38	83	0.898	0	1	0	0.004
konsisten	37	79	0.903	1	5	0.65	0.004
korupsi	30	73	0.87	8	11	0.982	0.008

koruptor	37	83	0.891	1	1	1	0.002
kuat	38	83	0.898	0	1	0	0.004
kubu	38	83	0.898	0	1	0	0.004
kursi	38	83	0.898	0	1	0	0.004
lain	37	83	0.891	1	1	1	0.002
laku	37	83	0.891	1	1	1	0.002
langgar	37	84	0.888	1	0	0	0.014
lebih	38	83	0.898	0	1	0	0.004
lembek	37	84	0.888	1	0	0	0.014
libat	37	84	0.888	1	0	0	0.014
lupa	38	83	0.898	0	1	0	0.004
maju	38	83	0.898	0	1	0	0.004
makin	38	83	0.898	0	1	0	0.004
mampu	38	83	0.898	0	1	0	0.004
manfaat	38	83	0.898	0	1	0	0.004
mantap	38	81	0.904	0	3	0	0.013
masalah	38	82	0.901	0	2	0	0.009
mau	37	84	0.888	1	0	0	0.014
menang	38	83	0.898	0	1	0	0.004
misi	38	84	0.895	0	0	0	0
nasional	38	83	0.898	0	1	0	0.004
palsu	37	84	0.888	1	0	0	0.014
parlemen	38	84	0.895	0	0	0	0
pasang	38	83	0.898	0	1	0	0.004
peduli	38	83	0.898	0	1	0	0.004
percaya	37	81	0.897	1	3	0.811	0.001
pernah	38	81	0.904	0	3	0	0.013
persen	38	83	0.898	0	1	0	0.004
pihak	38	84	0.895	0	0	0	0
pilih	36	77	0.903	2	7	0.764	0.002
pimpin	37	82	0.894	1	2	0.918	0
politik	36	78	0.9	2	6	0.811	0.001
politis	38	83	0.898	0	1	0	0.004
prestasi	38	82	0.901	0	2	0	0.009
pro	38	83	0.898	0	1	0	0.004
punya	38	82	0.901	0	2	0	0.009
rakyat	37	82	0.894	1	2	0.918	0
ramah	38	83	0.898	0	1	0	0.004
saing	38	84	0.895	0	0	0	0
salah	36	83	0.884	2	1	0.918	0.01
sambil	38	83	0.898	0	1	0	0.004
sangat	38	82	0.901	0	2	0	0.009
santun	38	83	0.898	0	1	0	0.004
sarang	37	83	0.891	1	1	1	0.002

satu	38	83	0.898	0	1	0	0.004
saudara	38	83	0.898	0	1	0	0.004
sejahtera	38	83	0.898	0	1	0	0.004
sekal	38	84	0.895	0	0	0	0
selalu	38	83	0.898	0	1	0	0.004
selesai	38	83	0.898	0	1	0	0.004
semangat	38	83	0.898	0	1	0	0.004
siap	38	83	0.898	0	1	0	0.004
sikap	38	84	0.895	0	0	0	0
sip	38	83	0.898	0	1	0	0.004
sistem	37	83	0.891	1	1	1	0.002
strategi	38	84	0.895	0	0	0	0
suap	38	83	0.898	0	1	0	0.004
tahan	38	83	0.898	0	1	0	0.004
tak	35	79	0.89	3	5	0.954	0.001
tanpa	37	80	0.9	1	4	0.722	0.002
tapi	38	82	0.901	0	2	0	0.009
tegas	38	82	0.901	0	2	0	0.009
tinggi	37	82	0.894	1	2	0.918	0
tingkat	38	83	0.898	0	1	0	0.004
tolak	34	79	0.882	4	5	0.991	0.005
tonjol	38	83	0.898	0	1	0	0.004
turun	37	84	0.888	1	0	0	0.014
uang	36	82	0.887	2	2	1	0.004
ubah	38	83	0.898	0	1	0	0.004
uji	38	83	0.898	0	1	0	0.004
unggul	38	83	0.898	0	1	0	0.004
utama	37	83	0.891	1	1	1	0.002
visi	38	83	0.898	0	1	0	0.004
yakin	37	79	0.903	1	5	0.65	0.004
ga	32	79	0.866	6	5	0.994	0.017
Max Gain : 0.017 Node : ga							

Setelah perhitungan untuk menentukan simpul berakhir maka selanjutnya ditentukan atribut dengan nilai gain tertinggi. Nilai gain tertinggi pada perhitungan simpul adalah atribut tes dengan kata ga dengan nilai 0.017.



Gambar 3.8 Perhitungan Pohon Keputusan Cabang Payah

Nilai entropy untuk atribut tes dengan kata ga memiliki nilai tidak sama dengan nol. Maka simpul tersebut masih memiliki cabang. Selanjutnya akan dilakukan perhitungan nilai gain untuk menentukan simpul cabang.

Penentuan simpul akan selalu dilakukan untuk cabang kiri terlebih dahulu, jika simpul simpul sudah mencapai *leaf node* dengan syarat nilai entropy untuk kedua cabang sama dengan nol, maka penentuan simpul akan bergesar pada cabang kanan simpul *parent* dari simpul ditentukan sebelumnya.

Selanjutnya akan dilakukan perhitungan nilai gain untuk cabang kanan dari simpul ga. Dimisalkan cabang kiri dari simpul ga dan seterusnya sudah mencapai *leaf node*.

Tabel 3.7 Perhitungan *Information Gain* Simpul Cabang Ga

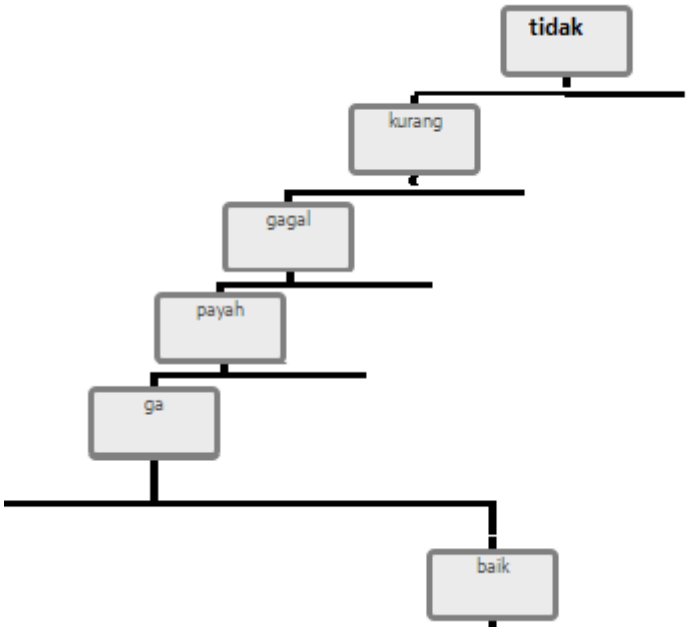
Perhitungan Cabang Node 'ga' Value '1' Total Entropy : 0.994							
Attribut	Neg(0)	Pos(0)	Entropy(0)	Neg(1)	Pos(1)	Entropy(1)	Gain
amanah	6	5	0.994	0	0	0	0
amanat	6	5	0.994	0	0	0	0
amin	6	5	0.994	0	0	0	0
anak	6	5	0.994	0	0	0	0
anti	6	5	0.994	0	0	0	0

ayo	6	5	0.994	0	0	0	0
bagai	6	5	0.994	0	0	0	0
baik	4	5	0.991	2	0	0	0.183
bajak	6	5	0.994	0	0	0	0
bangsa	6	5	0.994	0	0	0	0
bapak	6	5	0.994	0	0	0	0
bela	6	5	0.994	0	0	0	0
benar	6	5	0.994	0	0	0	0
berantas	6	5	0.994	0	0	0	0
bersih	6	5	0.994	0	0	0	0
bisa	6	5	0.994	0	0	0	0
bohong	6	4	0.971	0	1	0	0.111
buat	6	5	0.994	0	0	0	0
bukan	6	5	0.994	0	0	0	0
buruk	6	4	0.971	0	1	0	0.111
cena	6	5	0.994	0	0	0	0
cermin	6	5	0.994	0	0	0	0
citra	6	5	0.994	0	0	0	0
cocok	6	5	0.994	0	0	0	0
cukup	6	5	0.994	0	0	0	0
cuma	6	5	0.994	0	0	0	0
curang	6	4	0.971	0	1	0	0.111
damai	6	5	0.994	0	0	0	0
dukung	6	5	0.994	0	0	0	0
fitnah	6	5	0.994	0	0	0	0
gagas	6	5	0.994	0	0	0	0
gaul	6	5	0.994	0	0	0	0
gentar	6	5	0.994	0	0	0	0
hara	6	5	0.994	0	0	0	0
hasil	6	5	0.994	0	0	0	0
ideal	6	5	0.994	0	0	0	0
ingin	6	5	0.994	0	0	0	0
internasional	6	5	0.994	0	0	0	0
jangan	6	5	0.994	0	0	0	0
janji	6	5	0.994	0	0	0	0
jeblok	6	5	0.994	0	0	0	0
jelek	6	5	0.994	0	0	0	0
juang	6	5	0.994	0	0	0	0
juara	6	5	0.994	0	0	0	0
jujur	5	5	1	1	0	0	0.085
kalo	6	5	0.994	0	0	0	0
karena	6	5	0.994	0	0	0	0
kasus	6	5	0.994	0	0	0	0
koalisi	6	5	0.994	0	0	0	0

komunikasi	6	5	0.994	0	0	0	0
konsisten	5	5	1	1	0	0	0.085
korupsi	6	4	0.971	0	1	0	0.111
koruptor	6	5	0.994	0	0	0	0
kuat	6	5	0.994	0	0	0	0
kubu	6	5	0.994	0	0	0	0
kursi	6	5	0.994	0	0	0	0
lain	6	5	0.994	0	0	0	0
laku	6	5	0.994	0	0	0	0
langgar	6	5	0.994	0	0	0	0
lebih	6	5	0.994	0	0	0	0
lembek	6	5	0.994	0	0	0	0
libat	6	5	0.994	0	0	0	0
lupa	6	5	0.994	0	0	0	0
maju	6	5	0.994	0	0	0	0
makin	6	5	0.994	0	0	0	0
mampu	6	5	0.994	0	0	0	0
manfaat	6	5	0.994	0	0	0	0
mantap	6	5	0.994	0	0	0	0
masalah	6	5	0.994	0	0	0	0
mau	6	5	0.994	0	0	0	0
menang	6	5	0.994	0	0	0	0
misi	6	5	0.994	0	0	0	0
nasional	6	5	0.994	0	0	0	0
palsu	6	5	0.994	0	0	0	0
parlemen	6	5	0.994	0	0	0	0
pasang	6	5	0.994	0	0	0	0
peduli	6	5	0.994	0	0	0	0
percaya	6	5	0.994	0	0	0	0
pernah	6	5	0.994	0	0	0	0
persen	6	5	0.994	0	0	0	0
pihak	6	5	0.994	0	0	0	0
pilih	6	5	0.994	0	0	0	0
pimpin	6	5	0.994	0	0	0	0
politik	6	5	0.994	0	0	0	0
politis	6	5	0.994	0	0	0	0
prestasi	6	5	0.994	0	0	0	0
pro	6	5	0.994	0	0	0	0
punya	6	5	0.994	0	0	0	0
rakyat	5	5	1	1	0	0	0.085
ramah	6	5	0.994	0	0	0	0
saing	6	5	0.994	0	0	0	0
salah	6	4	0.971	0	1	0	0.111
sambil	6	5	0.994	0	0	0	0

sangat	6	5	0.994	0	0	0	0
santun	6	5	0.994	0	0	0	0
sarang	6	5	0.994	0	0	0	0
satu	6	5	0.994	0	0	0	0
saudara	6	5	0.994	0	0	0	0
sejahtera	6	5	0.994	0	0	0	0
sekal	6	5	0.994	0	0	0	0
selalu	6	5	0.994	0	0	0	0
selesai	6	5	0.994	0	0	0	0
semangat	6	5	0.994	0	0	0	0
siap	6	5	0.994	0	0	0	0
sikap	6	5	0.994	0	0	0	0
sip	6	5	0.994	0	0	0	0
sistem	6	5	0.994	0	0	0	0
strategi	6	5	0.994	0	0	0	0
suap	6	5	0.994	0	0	0	0
tahan	6	5	0.994	0	0	0	0
tak	6	5	0.994	0	0	0	0
tanpa	6	5	0.994	0	0	0	0
tapi	6	5	0.994	0	0	0	0
tegas	6	5	0.994	0	0	0	0
tinggi	6	5	0.994	0	0	0	0
tingkat	6	5	0.994	0	0	0	0
tolak	6	5	0.994	0	0	0	0
tonjol	6	5	0.994	0	0	0	0
turun	6	5	0.994	0	0	0	0
uang	6	5	0.994	0	0	0	0
ubah	6	5	0.994	0	0	0	0
uji	6	5	0.994	0	0	0	0
unggul	6	5	0.994	0	0	0	0
utama	6	5	0.994	0	0	0	0
visi	6	5	0.994	0	0	0	0
yakin	5	5	1	1	0	0	0.085
Max Gain : 0.183 Node : baik							

Setelah perhitungan untuk menentukan simpul berakhir maka selanjutnya ditentukan atribut dengan nilai gain tertinggi. Nilai gain tertinggi pada perhitungan simpul adalah atribut tes dengan kata baik dengan nilai 0.183.



Gambar 3.9 Perhitungan Pohon Keputusan Cabang Ga

Nilai entropy untuk atribut tes dengan kata ga memiliki nilai tidak sama dengan nol untuk cabang simpul kiri sedangkan nilai entropy cabang simpul kanan adalah sama dengan nol. Maka cabang simpul kanan sudah mencapai *leaf node* sedangkan cabang simpul kiri masih memiliki cabang. Selanjutnya akan dilakukan perhitungan nilai gain untuk menentukan simpul cabang.

Tabel 3.8 Perhitungan *Information Gain* Simpu Cabang Baik

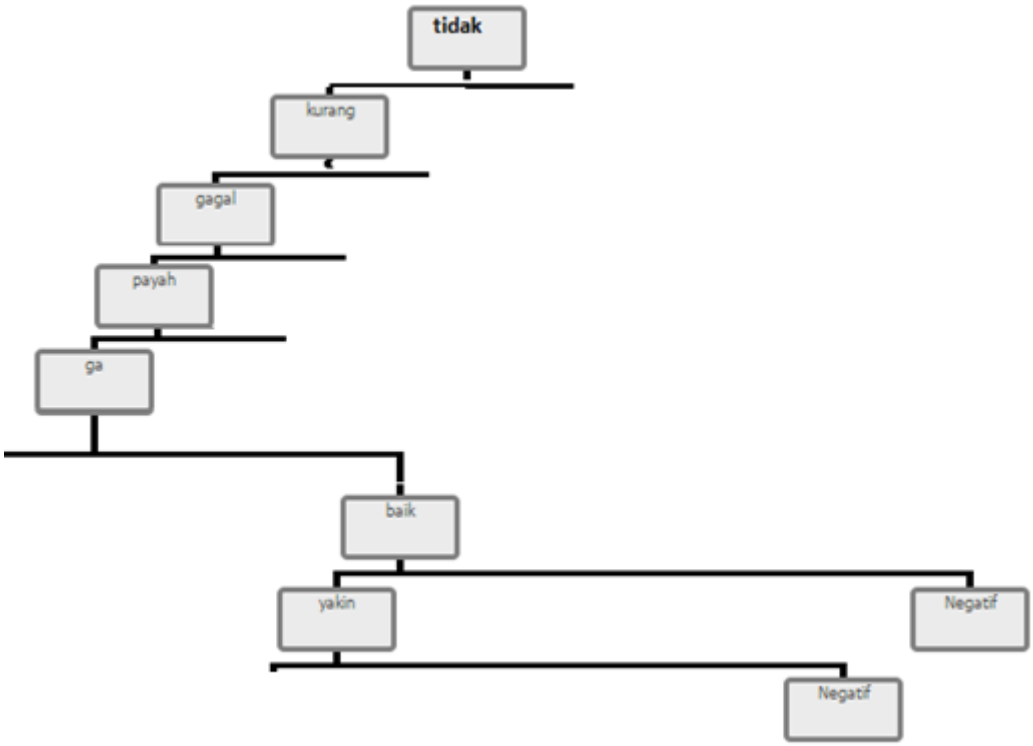
Perhitungan Cabang Node 'baik' Value '0' Total Entropy : 0.991							
Attribut	Neg(0)	Pos(0)	Entropy(0)	Neg(1)	Pos(1)	Entropy(1)	Gain
amanah	4	5	0.991	0	0	0	0
amanat	4	5	0.991	0	0	0	0
amin	4	5	0.991	0	0	0	0
anak	4	5	0.991	0	0	0	0
anti	4	5	0.991	0	0	0	0
ayo	4	5	0.991	0	0	0	0
bagai	4	5	0.991	0	0	0	0
bajak	4	5	0.991	0	0	0	0
bangsa	4	5	0.991	0	0	0	0
bapak	4	5	0.991	0	0	0	0
bela	4	5	0.991	0	0	0	0
benar	4	5	0.991	0	0	0	0
berantas	4	5	0.991	0	0	0	0
bersih	4	5	0.991	0	0	0	0

bisa	4	5	0.991	0	0	0	0
bohong	4	4	1	0	1	0	0.102
buat	4	5	0.991	0	0	0	0
bukan	4	5	0.991	0	0	0	0
buruk	4	4	1	0	1	0	0.102
cena	4	5	0.991	0	0	0	0
cermin	4	5	0.991	0	0	0	0
citra	4	5	0.991	0	0	0	0
cocok	4	5	0.991	0	0	0	0
cukup	4	5	0.991	0	0	0	0
cuma	4	5	0.991	0	0	0	0
curang	4	4	1	0	1	0	0.102
damai	4	5	0.991	0	0	0	0
dukung	4	5	0.991	0	0	0	0
fitnah	4	5	0.991	0	0	0	0
gagas	4	5	0.991	0	0	0	0
gaul	4	5	0.991	0	0	0	0
gentar	4	5	0.991	0	0	0	0
hara	4	5	0.991	0	0	0	0
hasil	4	5	0.991	0	0	0	0
ideal	4	5	0.991	0	0	0	0
ingin	4	5	0.991	0	0	0	0
internasional	4	5	0.991	0	0	0	0
jangan	4	5	0.991	0	0	0	0
janji	4	5	0.991	0	0	0	0
jeblok	4	5	0.991	0	0	0	0
jelek	4	5	0.991	0	0	0	0
juang	4	5	0.991	0	0	0	0
juara	4	5	0.991	0	0	0	0
jujur	3	5	0.954	1	0	0	0.143
kalo	4	5	0.991	0	0	0	0
karena	4	5	0.991	0	0	0	0
kasus	4	5	0.991	0	0	0	0
koalisi	4	5	0.991	0	0	0	0
komunikasi	4	5	0.991	0	0	0	0
konsisten	3	5	0.954	1	0	0	0.143
korupsi	4	4	1	0	1	0	0.102
koruptor	4	5	0.991	0	0	0	0
kuat	4	5	0.991	0	0	0	0
kubu	4	5	0.991	0	0	0	0
kursi	4	5	0.991	0	0	0	0
lain	4	5	0.991	0	0	0	0
laku	4	5	0.991	0	0	0	0
langgar	4	5	0.991	0	0	0	0

lebih	4	5	0.991	0	0	0	0
lembek	4	5	0.991	0	0	0	0
libat	4	5	0.991	0	0	0	0
lupa	4	5	0.991	0	0	0	0
maju	4	5	0.991	0	0	0	0
makin	4	5	0.991	0	0	0	0
mampu	4	5	0.991	0	0	0	0
manfaat	4	5	0.991	0	0	0	0
mantap	4	5	0.991	0	0	0	0
masalah	4	5	0.991	0	0	0	0
mau	4	5	0.991	0	0	0	0
menang	4	5	0.991	0	0	0	0
misi	4	5	0.991	0	0	0	0
nasional	4	5	0.991	0	0	0	0
palsu	4	5	0.991	0	0	0	0
parlemen	4	5	0.991	0	0	0	0
pasang	4	5	0.991	0	0	0	0
peduli	4	5	0.991	0	0	0	0
percaya	4	5	0.991	0	0	0	0
pernah	4	5	0.991	0	0	0	0
persen	4	5	0.991	0	0	0	0
pihak	4	5	0.991	0	0	0	0
pilih	4	5	0.991	0	0	0	0
pimpin	4	5	0.991	0	0	0	0
politik	4	5	0.991	0	0	0	0
politis	4	5	0.991	0	0	0	0
prestasi	4	5	0.991	0	0	0	0
pro	4	5	0.991	0	0	0	0
punya	4	5	0.991	0	0	0	0
rakyat	4	5	0.991	0	0	0	0
ramah	4	5	0.991	0	0	0	0
saing	4	5	0.991	0	0	0	0
salah	4	4	1	0	1	0	0.102
sambil	4	5	0.991	0	0	0	0
sangat	4	5	0.991	0	0	0	0
santun	4	5	0.991	0	0	0	0
sarang	4	5	0.991	0	0	0	0
satu	4	5	0.991	0	0	0	0
saudara	4	5	0.991	0	0	0	0
sejahtera	4	5	0.991	0	0	0	0
sekal	4	5	0.991	0	0	0	0
selalu	4	5	0.991	0	0	0	0
selesai	4	5	0.991	0	0	0	0
semangat	4	5	0.991	0	0	0	0

siap	4	5	0.991	0	0	0	0
sikap	4	5	0.991	0	0	0	0
sip	4	5	0.991	0	0	0	0
sistem	4	5	0.991	0	0	0	0
strategi	4	5	0.991	0	0	0	0
suap	4	5	0.991	0	0	0	0
tahan	4	5	0.991	0	0	0	0
tak	4	5	0.991	0	0	0	0
tanpa	4	5	0.991	0	0	0	0
tapi	4	5	0.991	0	0	0	0
tegas	4	5	0.991	0	0	0	0
tinggi	4	5	0.991	0	0	0	0
tingkat	4	5	0.991	0	0	0	0
tolak	4	5	0.991	0	0	0	0
tonjol	4	5	0.991	0	0	0	0
turun	4	5	0.991	0	0	0	0
uang	4	5	0.991	0	0	0	0
ubah	4	5	0.991	0	0	0	0
uji	4	5	0.991	0	0	0	0
unggul	4	5	0.991	0	0	0	0
utama	4	5	0.991	0	0	0	0
visi	4	5	0.991	0	0	0	0
yakin	3	5	0.954	1	0	0	0.143
Max Gain : 0.143 Node : yakin							

Setelah perhitungan untuk menentukan simpul berakhir maka selanjutnya ditentukan atribut dengan nilai gain tertinggi. Nilai gain tertinggi pada perhitungan simpul adalah atribut tes dengan kata yakin dengan nilai 0.143.



Gambar 3.10 Perhitungan Pohon Keputusan Cabang Baik

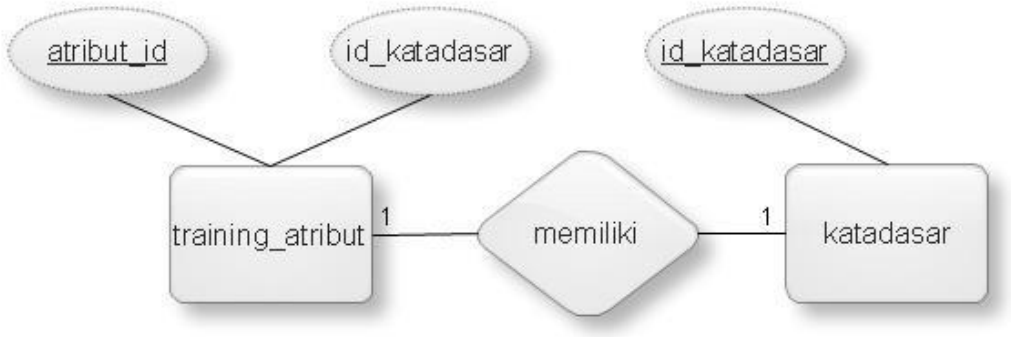
Nilai entropy untuk atribut tes dengan kata yakin memiliki nilai tidak sama dengan nol untuk cabang simpul kiri sedangkan nilai entropy cabang simpul kanan adalah sama dengan nol. Maka cabang simpul kanan sudah mencapai *leaf node* sedangkan cabang simpul kiri masih memiliki cabang.

Setelah perhitungan berakhir dengan memenuhi kondisi untuk mengakhiri proses rekursif, selanjutnya simpil-simpul yang telah ditentukan akan membentuk sebuah pohon keputusan.

3.1.3 Analisis Kebutuhan Perangkat Lunak

3.1.3.1 Analisis Basis Data

Basis data digunakan untuk menyimpan data twitter dan data untuk proses klasifikasi sentimen negatif fan positif. Gambar 3.5 adalah *Entity Relation Diagram* (ERD) yang digunakan pada pembangunan aplikasi.



Gambar 3.11 ERD yang digunakan pada Sistem Analisis Sentimen Pengguna Twitter Terhadap Partai Politik

3.1.3.2 Analisis Kebutuhan Non Fungsional

Analisis kebutuhan non fungsional adalah suatu analisis untuk mengetahui elemen-elemen yang berhubungan dengan sistem yang sedang berjalan.

3.1.3.2.1. Analisis Pengguna

Aplikasi ini digunakan oleh dua jenis user yaitu user dan user umum, user berperan menjalankan proses mengolah training data dan membentuk pohon keputusan yang dibentuk dari training data, sedangkan user umum atau masyarakat berperan mengakses data sentimen mengenai partai politik yang telah diklasifikasikan ke dalam kelas positif dan negatif dalam bentuk grafik.

Tabel 3.9 Analisis Pengguna

Pengguna	Tanggung Jawab
User	Mengelola training data, membentuk pohon keputusan yang akan digunakan untuk klasifikasi dan menjalankan *streaming twitter*.
User Umum	Mengakses data sentimen mengenai partai politik yang telah diklasifikasikan ke dalam sentimen positif dan negatif dalam bentuk grafik.

3.1.3.2.2. Analisis Kebutuhan Perangkat Keras

Adapun spesifikasi perangkat keras minimum yang dibutuhkan untuk membangun aplikasi ini adalah sebagai berikut :

1. Processor 1.6 Ghz.
2. Media Penyimpanan (Hardisk) 40 GB.
3. Memori RAM 256 Mb.
4. Monitor 16 inchi.

3.1.3.2.3. Analisis Perangkat Lunak

Berikut ini adalah spesifikasi perangkat lunak minimum yang dibutuhkan dalam membangun aplikasi:

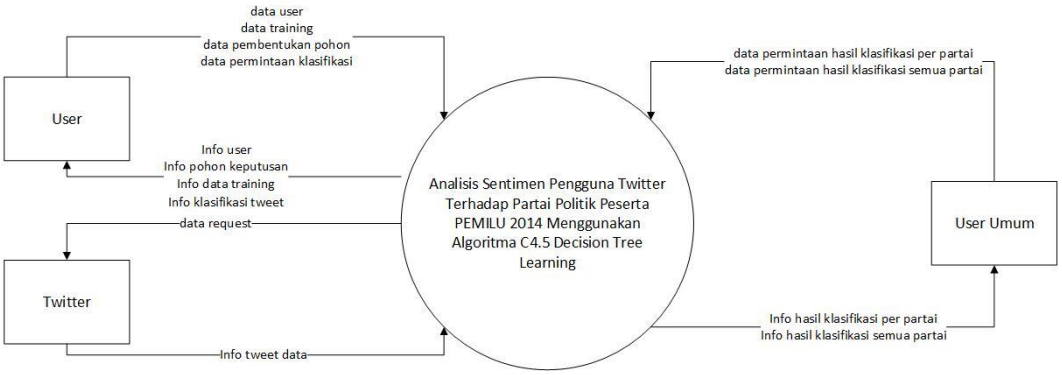
1. Sistem operasi Windows XP.
2. Wamp Server 2.4.
3. Notepad++.

3.1.3.3 Analisis Kebutuhan Fungsional

Analisis kebutuhan fungsional ini meliputi diagram konteks dan *Data Flow Diagram* (DFD).

3.1.3.3.1 Diagram Konteks

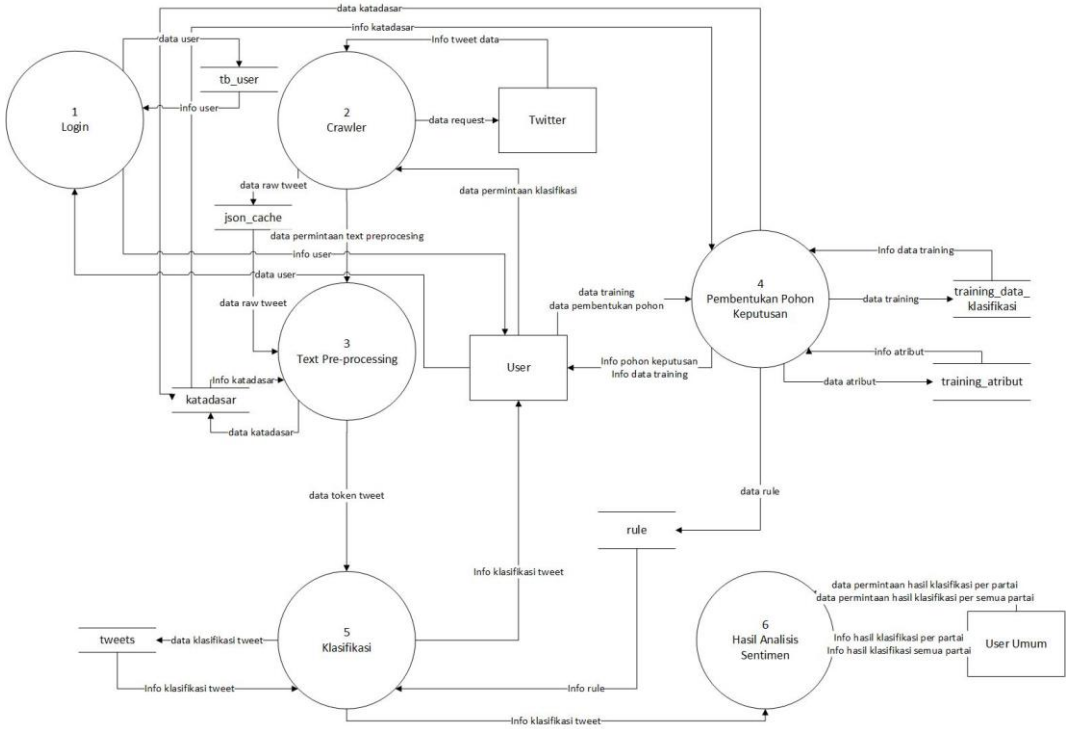
Diagram konteks adalah diagram yang menunjukkan keterhubungan antara perangkat lunak dengan konteks eksternal di luar program. Gambar berikut adalah diagram konteks yang dibangun dalam sistem.



Gambar 3.12 Diagram Konteks

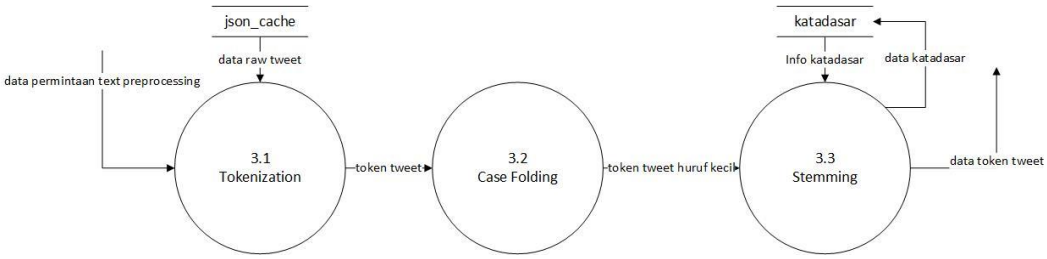
3.1.3.3.2 Data Flow Diagram

1. Data Flow Diagram Level 1



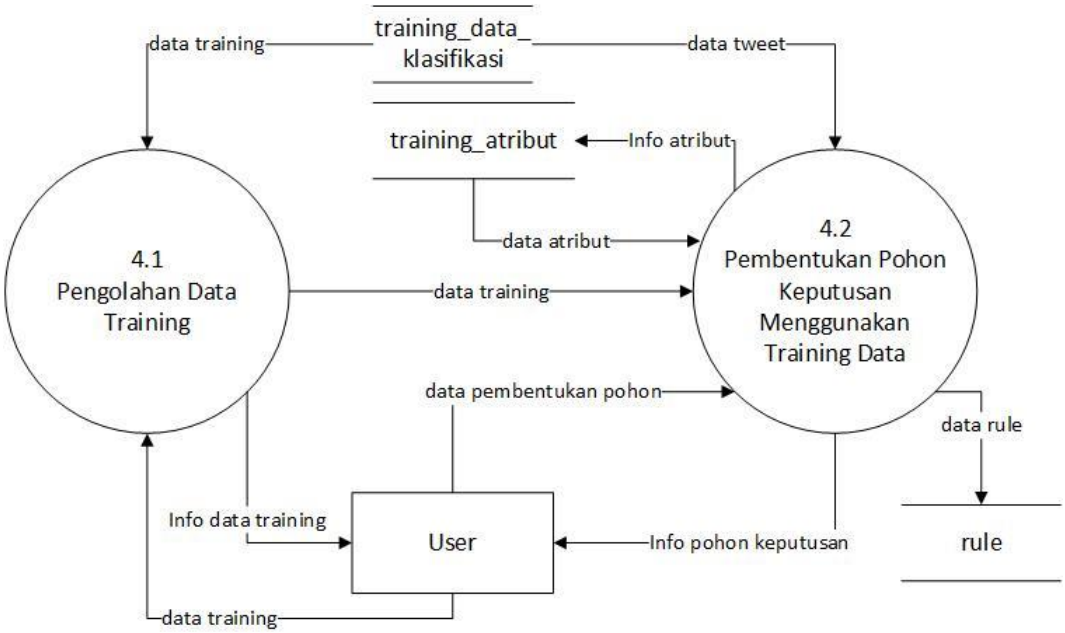
Gambar 3.13 DFD Level 1 Analisis Sentimen Pengguna Twitter Terhadap Partai Politik

2. Data Flow Diagram Level 2 Proses 3



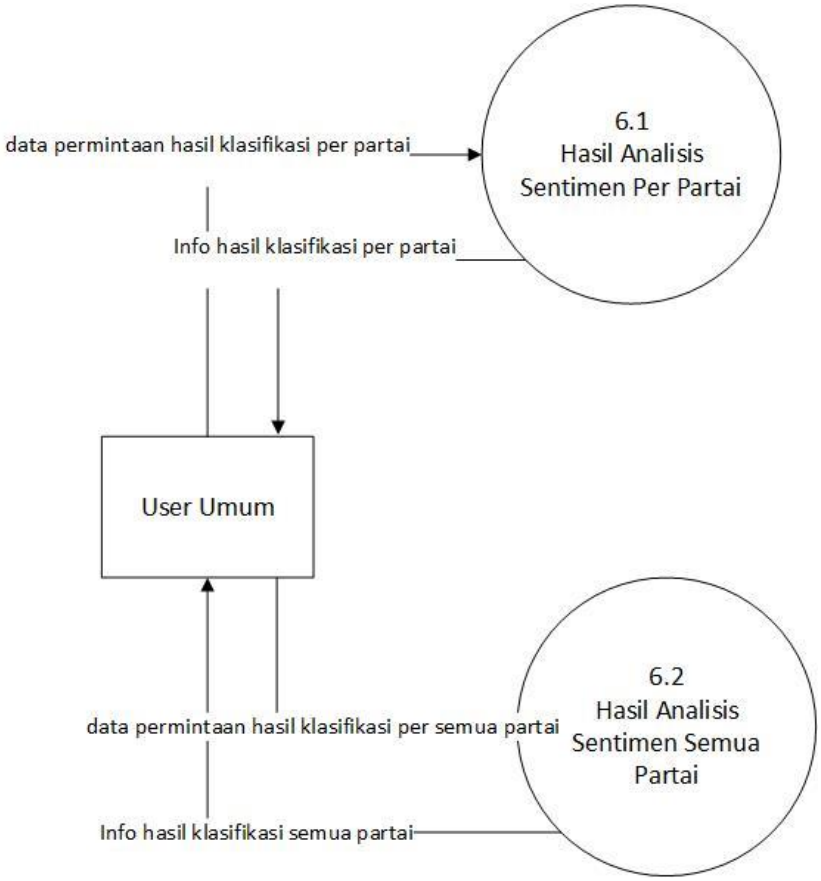
Gambar 3.14 DFD Level 2 Proses *Text Pre-processing*

3. Data Flow Diagram Level 2 Proses 4



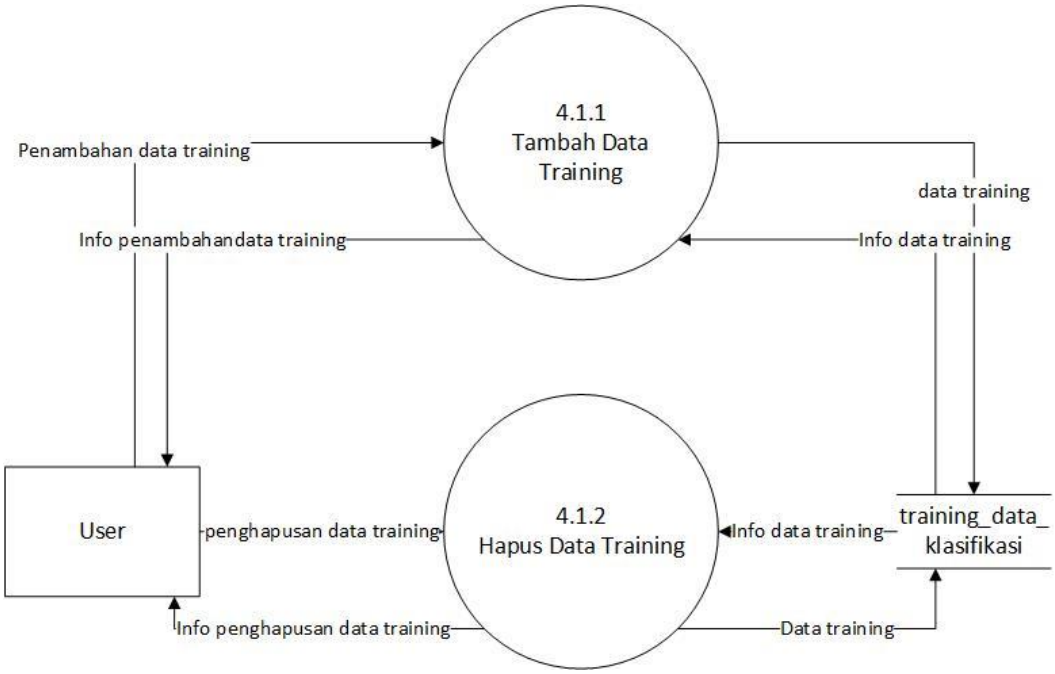
Gambar 3.15 DFD Level 2 Proses Pembentukan Pohon Keputusan

4. Data Flow Diagram Level 2 Proses 6



Gambar 3.16 DFD Level 2 Hasil Analisis Sentimen

5. Data Flow Diagram Level 3 Proses 4.1



Gambar 3.17 DFD Level 3 Proses Pengolahan Data Training

3.1.3.3.3 Spesifikasi Proses

Spesifikasi proses merupakan deskripsi dari setiap elemen proses yang terdapat dalam program yang meliputi nama proses, input, output, dan keterangan dari proses. Berikut ini spesifikasi proses dari sistem Analisis Sentimen Pengguna Twitter Terhadap Partai Politik.

Tabel 3.10 Spesifikasi Proses Analisis Sentimen Pengguna Twitter Terhadap Partai Politik

No.	Proses	Keterangan
1.	No. Proses	1
	Nama Proses	Login
	Input	info user
	Output	data user
	Logika Proses	1. Jika ada masukan data login. 2. Sistem akan mengecek apakah data login masih kosong. 3. Jika data kosong Sistem akan menampilkan pesan kesalahan belum

		mengisi username / password 4. Jika tidak kosong maka sistem akan memverifikasi data login 5. Jika data login valid maka pengguna akan masuk ke halaman kelola pohon keputusan.
	Keterangan Proses	Proses ini digunakan untuk verifikasi data login sebelum masuk ke halaman kelola pohon keputusan.
2.	No. Proses	2
	Nama Proses	Crawler
	Input	info tweet data
	Output	data request
	Logika Proses	1. Jika raw *tweet* diterima 2. Simpan *raw tweet* di tabel json_cache
	Keterangan Proses	Proses ini digunakan untuk streaming *tweet* dalam bentuk *raw tweet*.
3.	No. Proses	3
	Nama Proses	*Text Pre-Processing*
	Input	data permintaan text preproccesing
	Output	data_token_tweet
	Logika Proses	1. *Tokenization* (Proses 3.1) 2. *Casefolding* (Proses 3.2) 3. *Stemming* (Proses 3.3)
	Keterangan Proses	Proses tokenisasi, *casefolding*, dan *stemming* terhadap *tweet* yang akan diklasifikasi.
4.	No Proses	4
	Nama Proses	Pembentukan Pohon Keputusan
	Input	data training, data pembentukan pohon
	Output	Info pohon keputusan, Info data training

	Logika Proses	1. Pengolahan Data Training (Proses 4.1) 2. Pembentukan Pohon Keputusan Menggunakan Training Data (Proses 4.2)
	Keterangan	Proses digunakan untuk Pembentukan Pohon Keputusan Menggunakan Algoritma C4.5.
5.	No. Proses	5
	Nama Proses	Klasifikasi
	Input	data token tweet
	Output	Info klasifikasi tweet
	Logika Proses	1. Pengecekan data token tweet terhadap simpul akar, jika kata penguji pada simpul ada pada data token tweet maka cek simpul berikutnya pada ruas kanan, jika tidak ada cek simpul pada ruas kiri. 2. Jika simpul yang di cek merupakan simpul daun maka pengecekan simpul pada pohon keputusan berakhir dan menghasilkan kelas sesuai yang terdapat pada simpul daun. 3. Jika simpul yang di cek bukan simpul daun maka dilakukan pengecekan data token tweet terhadap simpul tersebut seperti pengecekan terhadap simpul akar. 4. Pengecekan simpul berhenti jika simpul yang dicek adalah simpul daun.
	Keterangan Proses	Proses ini digunakan untuk klasifikasi *tweet* yang belum diberi kelas.
6.	No. Proses	6
	Nama Proses	Hasil Analisis Sentimen

	Input	data permintaan hasil klasifikasi
	Output	info hasil klasifikasi
	Logika Proses	1. Jika user umum melakukan permintaan untuk melihat hasil klasifikasi, tampilkan hasil klasifikasi. 2. Hitung persentase sentimen positif dan negatif dari masing-masing partai pada *tweet* yang telah terklasifikasi. 3. Jika *tweet* kosong, maka tampilkan pesan data tidak tersedia.
	Keterangan	Proses ini digunakan user umum untuk melihat hasil analisis sentimen
7.	No. Proses	3.1
	Nama Proses	Tokenization
	Input	data text preprocessing
	Output	token tweet
	Logika Proses	1. Uraikan data *raw tweet* menjadi tweet beserta atributnya. 2. Cek *tweet* apakah tidak kosong. 3. Jika *tweet* tidak kosong, maka *tweet* diurai kata per kata.
	Keterangan Proses	Proses ini digunakan untuk memecah *tweet*
8.	No. Proses	3.2
	Nama Proses	Case Folding
	Input	token tweet
	Output	token tweet huruf kecil
	Logika Proses	1. Cek *tweet* apakah tidak kosong. 2. Jika terdapat huruf besar, maka huruf diubah menjadi kecil.
	Keterangan Proses	Proses ini digunakan untuk megubah bentuk

		karakter huruf A-Z menjadi a-z.
9.	No. Proses	3.3
	Nama Proses	Stemming
	Input	token tweet huruf kecil
	Output	data token tweet
	Logika Proses	1. Jika kata yang akan di-stem ditemukan dalam kamus katadasar, maka kata tersebut adalah *root word.* 2. Jika tidak ditemukan, maka dilakukan penghapusan *inflexion suffix* 3. Jika hasil sebelumnya bukan *root word,* maka dilakukan penghapusan *derivation suffix*, 4. jika pada langkah sebelumnya ada suffix yang dihapus maka dilakukan pemeriksaan kombinasi awalan akhiran yang tidak diijinkan. Jika ditemukan tentukan sebagai *root word.* 5. jika tidak, maka dilakukan penghapusan *derivation prefix* 6. jika semua langkah telah dilakukan dan tidak menemukan *root word.* Maka kata awal ditentukan sebagai *root word.*
	Keterangan Proses	Proses ini digunakan untuk mencari kata dasar dari kata berimbuhan.
10.	No Proses	4.1
	Nama Proses	Pengolahan Data Training
	Input	data training
	Output	Info data training
	Logika Proses	1. Tambah Data Training (Proses 4.1.1)

		2. Hapus Data Training (Proses 4.1.2)
	Keterangan Proses	Proses ini terdiri dari menambah atau menghapus data training.
11.	No Proses	4.2
	Nama Proses	Pembentukan Pohon Keputusan menggunakan Training Data
	Input	data training, data pembentukan pohon,
	Output	Info pohon keputusan
	Logika Proses	1. Terdapat masukan berupa training set yang setiap sampelnya telah diberi kelas positif atau negatif. 2. Jika seluruh sampel pada training set memiliki kelas yang sama maka pohon keputusan akan memiliki satu *node* berupa *leaf node* yang diberi label kelas yang terdapat pada semua sampel dalam training set. 3. Jika seluruh sampel tidak dalam satu kelas yang sama maka akan dicari *gain* tertinggi dari seluruh atribut untuk memilih atribut yang paling berpengaruh pada *training set*, dan akan dijadikan atribut penguji pada *node* tersebut. 4. Setiap *node* memiliki dua cabang yang dibangun berdasarkan nilai 0 dan 1 pada setiap atribut pengujian. Jika ada partisi nilai yang memiliki nilai entropy nol, maka cabang dari partisi nilai tersebut menjadi *leaf node* yang diisi kelas yang memiliki jumlah kemunculan tertinggi

		pada training data. 5. Jika ada cabang dari *node* yang dibentuk pada langkah sebelumnya belum mencapai *leaf node*, maka akan dicari nilai gain seperti pada langkah nomor 3 dimulai dari cabang paling kiri yang belum mencapai *leaf node*. 6. Jika seluruh cabang dari *node* yang dibentuk pada langkah sebelumnya sebelumnya telah mencapai *leaf node*, maka akan dicek cabang dari *node* diatas dari *node* yang dibentuk pada langkah sebelumnya, jika cabang tersebut belum mencapai *leaf node* maka akan dicari nilai *gain* seperti pada langkah nomor 3. 7. Proses yang sama akan dilakukan secara rekursif untuk membentuk pohon keputusan dari setiap sampel. 8. Proses rekursif akan berhenti jika semua sampel pada *node* memiliki kelas yang sama, semua simpul sudah mencapai *leaf node* atau semua atribut telah digunakan untuk mempartisi sampel,
	Keterangan Proses	Proses ini digunakan untuk membentuk pohon keputusan menggunakan algoritma C4.5.
12.	No Proses	4.1.1
	Nama Proses	Tambah Data Training
	Input	penambahan data training.
	Output	Info penambahan data training

	Logika Proses	1. Jika data training mengandung attribut klasifikasi, maka simpan data training 2. Jika data training mengandung attribut klasifikasi yang telah ada pada data training sebelumnya namun berbeda kelas, maka tampilkan pesan kesalahan.
	Keterangan Proses	Proses ini digunakan untuk menambah data training
13.	No Proses	4.1.2
	Nama Proses	Hapus Data Training
	Input	penambahan data training
	Output	Info penambahan data training
	Logika Proses	1. Jika data training dipilih, maka tampilkan pesan penghapusan data training. 2. Jika ya dipilih, maka hapus data training.
	Keterangan Proses	Proses ini digunakan untuk menghapus data training
14	No Proses	6.1
	Nama Proses	Hasil Analisis Sentimen Per Partai
	Input	data permintaan hasil klasifikasi per partai
	Output	info hasil klasifikasi per partai
	Logika Proses	1. Jika partai dipilih, maka tampilkan pilihan tahun. 2. Jika tahun dipilih, maka tampilkan grafik sentimen berdasarkan partai dan tahun yang dipilih.
	Keterangan Proses	Proses ini digunakan untuk melihat grafik sentimen berdasarkan partai dan tahun yang

		dipilih.
15	No Proses	6.2
	Nama Proses	Hasil Analisis Sentimen Semua Partai
	Input	data permintaan hasil klasifikasi semua partai
	Output	info hasil klasifikasi semua partai
	Logika Proses	1. Jika tanggal awal dan akhir tweet dipilih, maka ditampilkan grafik sentimen berdasarkan partai tanggal awal dan akhir
	Keterangan Proses	Proses ini digunakan untuk melihat grafik sentimen berdasarkan tanggal awal dan akhir yang dipilih.

3.1.3.3.4 Kamus Data

Kamus data digunakan untuk menjelaskan semua data yang digunakan dalam sistem. Berikut ini dijelaskan data yang digunakan dalam DFD (*Data Flow Diagram*).

Tabel 3.11 Kasus Data

No.	Data	Keterangan
1.	Nama	data training
	Struktur Data	kalimat_opini + kelas
	Deskripsi	kalimat_opini=[0..9|a..z|A..Z] kelas=[a..z|A..Z]
2.	Nama	info tweet data
	Struktur Data	raw_tweet
	Deskripsi	raw_tweet=[0..9|a..z|A..Z]
3.	Nama	data raw tweet
	Struktur Data	tweet_id + cache_date + raw_tweet
	Deskripsi	tweet_id[0..9] cache_date

		raw_tweet [0..9|a..z|A..Z]
4.	Nama	data katadasar
	Struktur Data	katadasar
	Deskripsi	[a..z]
5.	Nama	info katadasar
	Struktur Data	katadasar
	Deskripsi	[a..z]
6.	Nama	data token tweet
	Struktur Data	token + raw_tweet
	Deskripsi	token[0..9|a..z|A..Z] raw_tweet[0..9|a..z|A..Z]
7.	Nama	data klasifikasi tweet
	Struktur Data	tweet_id + tweet_text + created_at + screen_name + opinion
	Deskripsi	tweet_id[0..9] tweet_text[0..9|a..z|A..Z] created_at screen_name[0..9|a..z|A..Z] opinion[a..z|A..Z]
8.	Nama	info klasifikasi tweet
	Struktur Data	tweet_text + created_at + screen_name + opinion
	Deskripsi	tweet_text[0..9|a..z|A..Z] created_at screen_name[0..9|a..z|A..Z] opinion[a..z|A..Z]
9.	Nama	info data training
	Struktur Data	kalimat_opini + kelas
	Deskripsi	kalimat_opini=[0..9|a..z|A..Z] kelas=[a..z|A..Z]

10.	Nama	data atribut
	Struktur Data	term + count
	Deskripsi	term[a..z|A..Z] count[0..9]
10.	Nama	info atribut
	Struktur Data	term + count
	Deskripsi	term[a..z|A..Z] count[0..9]
11.	Nama	data rule
	Struktur Data	atribut + value + result + level + parent_id
	Deskripsi	atribut[a..z|A..Z] value[0..9] result[a..z|A..Z] level[0..9] parent_id[0..9]
12.	Nama	info rule
	Struktur Data	atribut + value + result + level + parent_id
	Deskripsi	atribut[a..z|A..Z] value[0..9] result[a..z|A..Z] level[0..9] parent_id[0..9]
12.	Nama	info pohon keputusan
	Struktur Data	atribut + value + result + level + parent_id
	Deskripsi	atribut[a..z|A..Z] value[0..9] result[a..z|A..Z]

		level[0..9] parent_id[0..9]
13.	Nama	data pembentukan pohon
	Struktur Data	kalimat_opini + kelas
	Deskripsi	kalimat_opini[a..z|A..Z|0..9] kelas[a..z|A..Z]
14.	Nama	data permintaan klasifikasi
	Struktur Data	kalimat_opini + kelas
	Deskripsi	kalimat_opini[a..z|A..Z|0..9] kelas[a..z|A..Z]
15.	Nama	data request
	Struktur Data	keyword
	Deskripsi	Keyword[a..z|A..Z]
16.	Nama	data permintaan hasil klasifikasi
	Struktur Data	nama_parpol
	Deskripsi	nama_parpol[a..z|A..Z|0..9]
17.	Nama	info hasil klasifikasi
	Struktur Data	tweet
	Deskripsi	tweet[a..z|A..Z|0..9]
18.	Nama	data permintaan text preprocessing
	Struktur Data	tweet_id + tweet
	Deskripsi	tweet_id[0..9] tweet[a..z|A..Z|0..9]
19.	Nama	token tweet
	Struktur Data	token
	Deskripsi	tweet_id[0..9] token[0..9|a..z|A..Z]
20.	Nama	token tweet huruf kecil
	Struktur Data	token_kecil
	Deskripsi	token[0..9|a..z|A..Z]

21.	Nama	penambahan data training
	Struktur Data	kalimat_opini + kelas
	Deskripsi	[0..9|a..z|A..Z]
22.	Nama	Info penambahan data training
	Struktur Data	kalimat_opini + kelas
	Deskripsi	Kalimat_opini[0..9|a..z|A..Z] Kelas[a..z|A..Z]
23.	Nama	Penghapusan data training
	Struktur Data	id_data_training
	Deskripsi	id_data_training[0..9]
24.	Nama	Info penghapusan data training
	Struktur Data	id_data_training
	Deskripsi	id_data_training[0..9]
25.	Nama	data permintaan hasil klasifikasi per partai
	Struktur Data	partai + tahun
	Deskripsi	partai[a..z|A..Z|0..9] tahun
26.	Nama	data permintaan hasil klasifikasi semua partai
	Struktur Data	tanggal awal + tanggal akhir
	Deskripsi	tanggal awal tanggal akhir
27.	Nama	info hasil klasifikasi per partai
	Struktur Data	partai + tahun
	Deskripsi	partai[a..z|A..Z|0..9] tahun
28.	Nama	Info hasil klasifikasi semua partai
	Struktur Data	tanggal awal + tanggal akhir
	Deskripsi	tanggal awal

		tanggal akhir

3.2 Perancangan Sistem

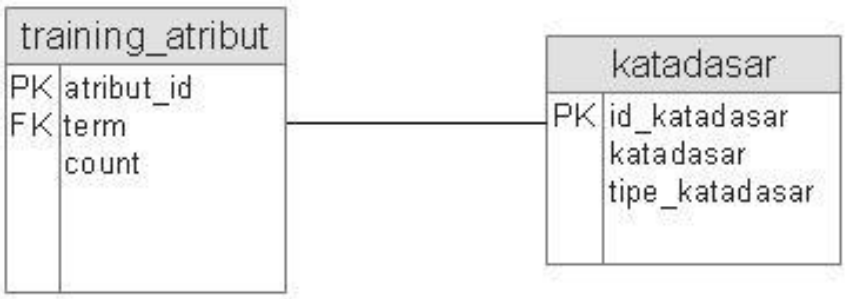
Perancangan sistem merupakan tahap lanjutan dari analisis sistem dimana digambarkan rancangan sistem yang akan dibangun sebelum dilakukan pengkodean menggunakan bahasa pemrograman.

3.2.1 Perancangan Data

Perancangan data terdiri dari perancangan struktur tabel dan skema relasi berdasarkan analisis basis data yang dibangun dalam sistem.

4.2.1.1 Skema Relasi

Gambar berikut adalah skema relasi yang dibangun pada aplikasi analisis sentimen pengguna twitter terhadap partai politik.



Gambar 3.18 Skema Relasi Analisis Sentimen Pengguna Twitter Terhadap Partai Politik

4.2.1.2 Struktur Tabel

Tabel berikut adalah struktur tabel berdasarkan skema relasi pada sistem

Tabel 3.12 Struktur Tabel katadasar

Nama Field	Tipe	Panjang	Keterangan
id_katadasar	int	10	primary_key
katadasar	varchar	70	

tipe_katadasar	int	25	

Tabel 3.13 Struktur Tabel training_atribut

Nama Field	Tipe	Panjang	Keterangan
atribut_id	int	10	primary_key
term	int	10	foreign_key, Ref katadasar(id_katadasar)
count	int	10	

Pada aplikasi ditambahkan lima tabel pendukung yaitu tabel training data untuk menyimpan data training, tabel json_cache digunakan untuk menyimpan *raw tweet* yang didapatkan dari proses *twitter streaming* dan tabel *tweets* untuk menyimpan *tweet* yang dihasilkan dari proses klasifikasi, tabel tree_rule untuk menyimpan *rule* pohon keputusan, tabel t_data untuk menyimpan data yang digunakan pada proses pembentukan pohon keputusan.

Tabel 3.14 Struktur Tabel tree_Rule

Nama Field	Tipe	Panjang	Keterangan
treerule_id	int	10	primary_key
attribut	int	10	
value	int	2	
result	int	10	
parent_id	int	10	

Tabel 3.15 Struktur Tabel data_training_klasifikasi

Nama Field	Tipe	Panjang	Keterangan
dtk_id	int	11	primary_key
kalimat_opini	varchar	160	
kelas	int	7	

Tabel 3.16 Struktur Tabel json_cache

Nama Field	Tipe	Panjang	Keterangan
cache_id	int	10	primary_key
tweet_id	bigint	20	
cache_date	timestamp		
raw_tweet	text		
parsed	tinyint	1	

Tabel 3.17 Struktur Tabel tweets

Nama Field	Tipe	Panjang	Keterangan
tweet_id	bigint	20	primary_key
tweet_text	varchar	140	
created_at	datetime		
screen_name	char	20	
profile_image_url	varchar	200	
opinion	int	3	

Tabel 3.18 Struktur Tabel t_data

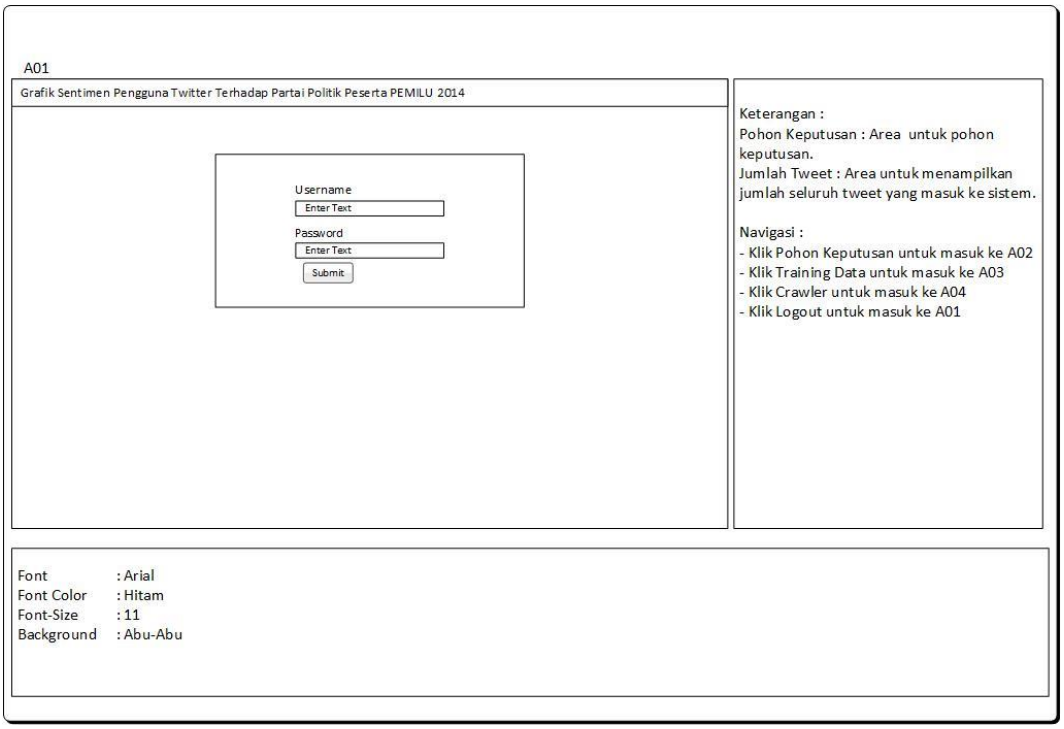
Nama Field	Tipe	Panjang	Keterangan
t_id	int	11	primary_key
atribut_value	int	3	
kelas	int	7	

Tabel 3.19 Struktur Tabel tb_user

Nama Field	Tipe	Panjang	Keterangan
usern_id	int	11	Primary_key
usern	varchar	30	
passw	varchar	32	

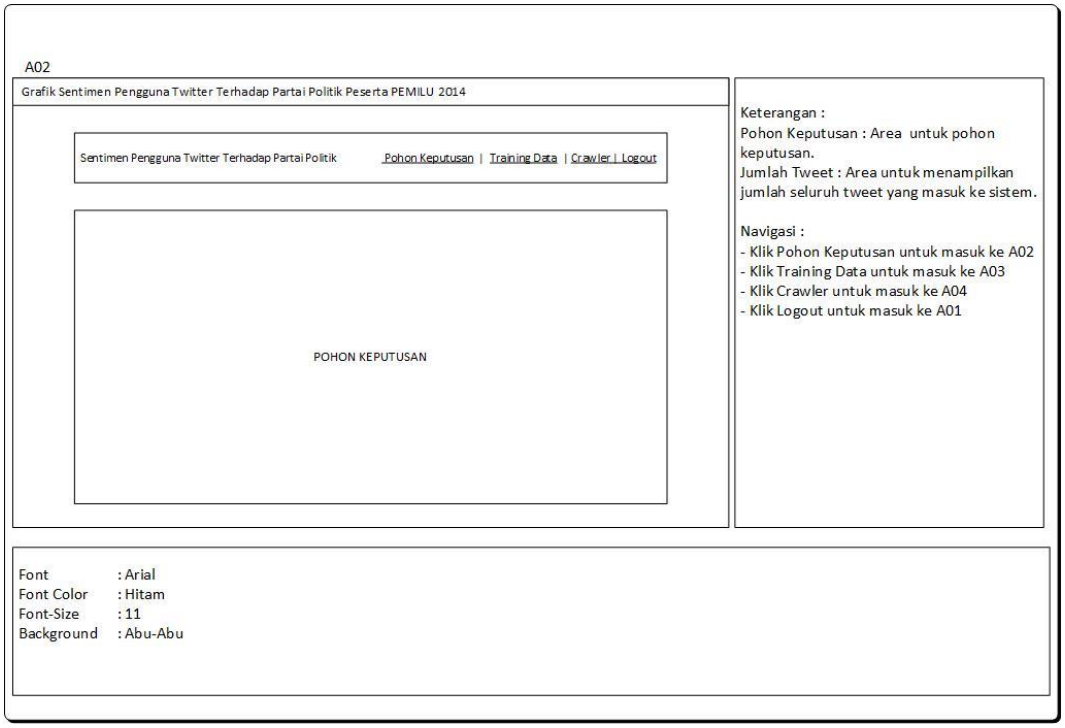
3.2.2. Perancangan Antarmuka

1. Perancangan Antarmuka Halaman Login Kelola Analisis Sentimen



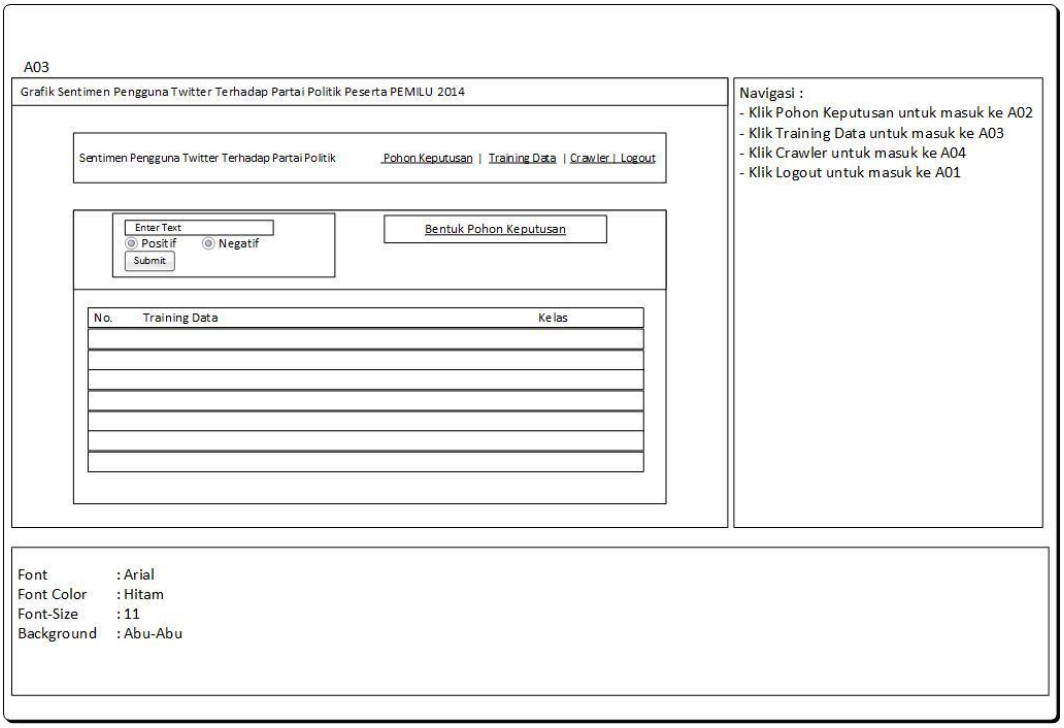
Gambar 3.19 Perancangan Antarmuka Halaman Login Kelola

2. Perancangan Antarmuka Halaman Pohon Keputusan



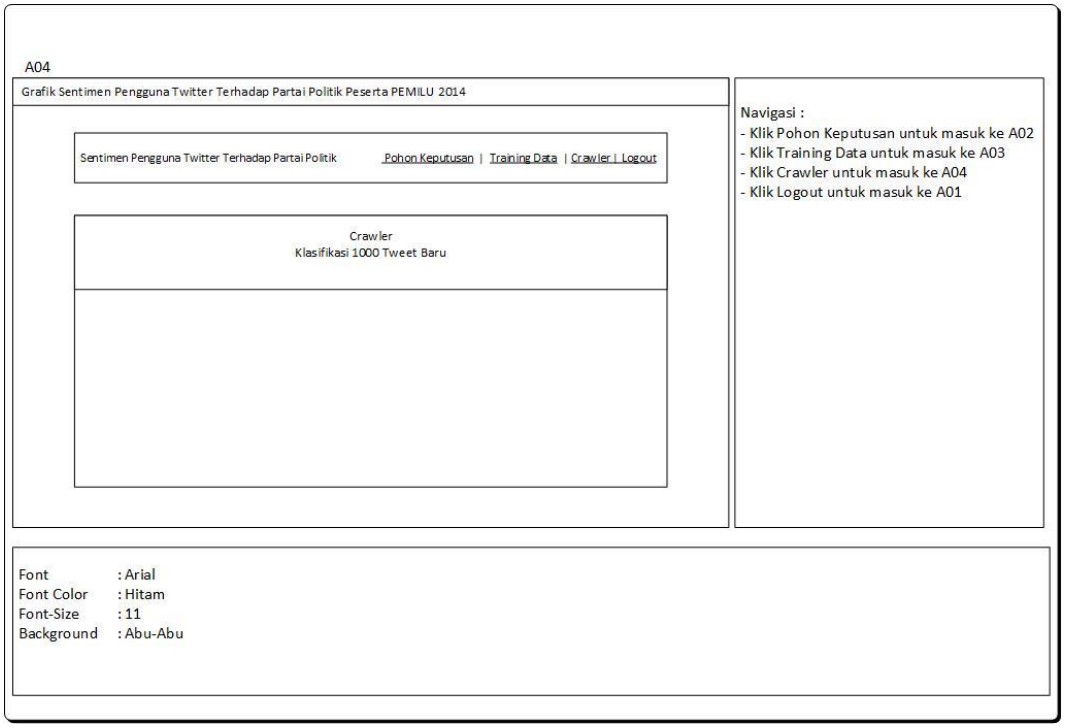
Gambar 3.20 Perancangan Antarmuka Halaman pohon Keputusan

3. Perancangan Antarmuka Halaman Training Data



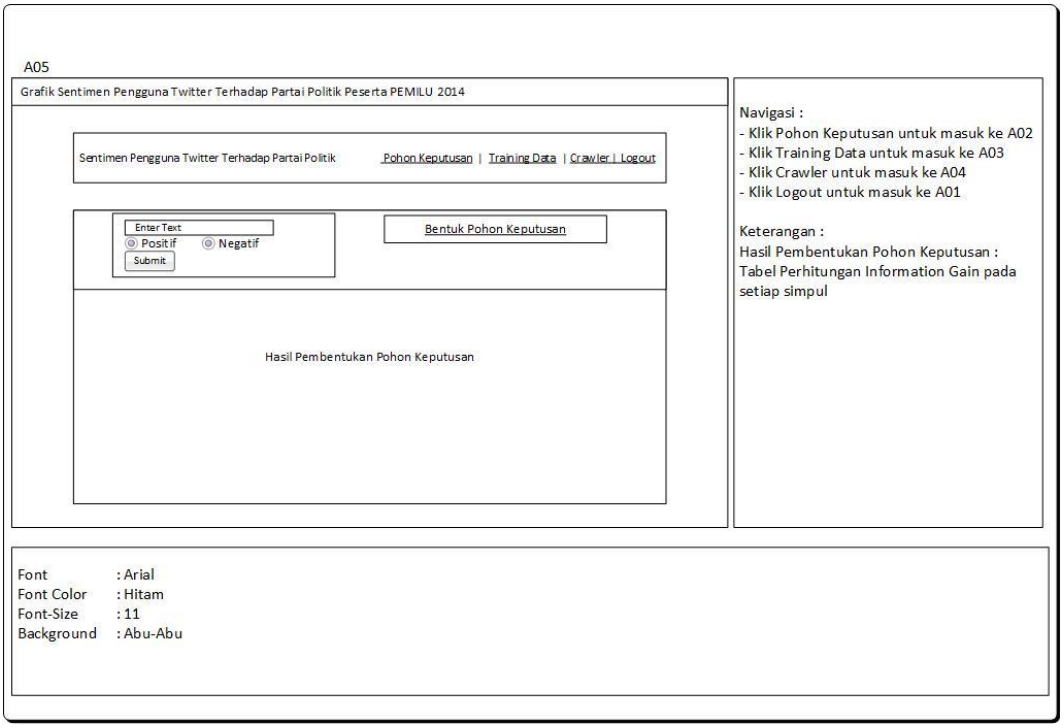
Gambar 3.21 Perancangan Antarmuka Halaman Training Data

4. Perancangan Antarmuka Halaman Crawler



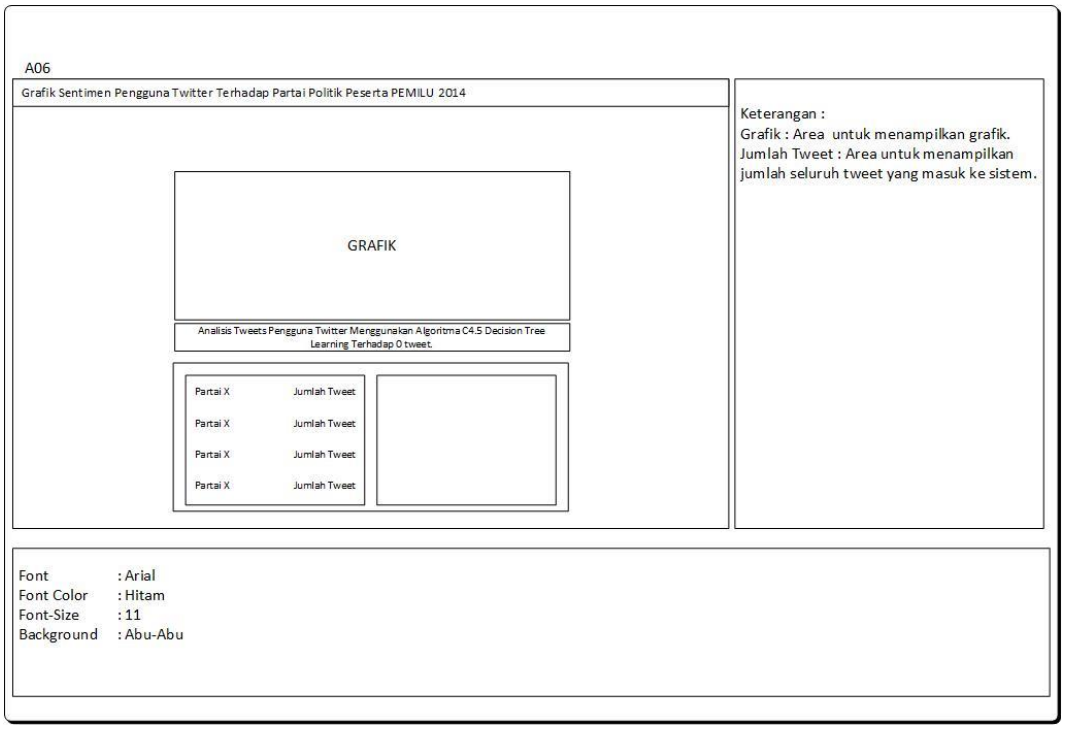
Gambar 3.22 Perancangan Antarmuka Halaman Crawler

5. Perancangan Antarmuka Halaman Hasil Pembentukan Pohon Keputusan



Gambar 3.23 Perancangan Antarmuka Halaman Hasil Pembentukan Pohon Keputusan

6. Perancangan Antarmuka Halaman Utama Analisis Sentimen Pengguna Twitter Terhadap Partai Politik



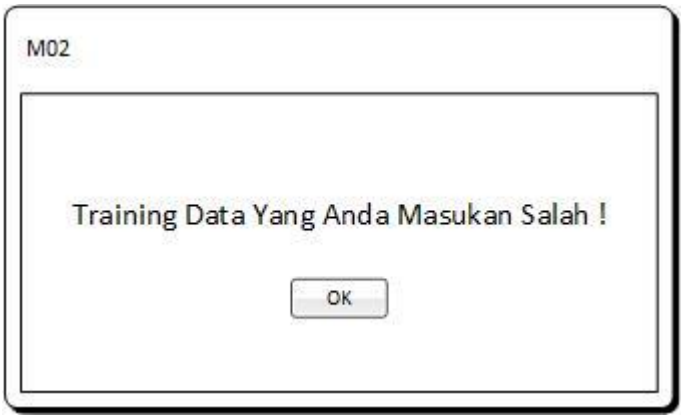
Gambar 3.24 Perancangan Antarmuka Halaman Utama Analisis Sentimen Pengguna Twitter Terhadap Partai Politik

7. Perancangan Antarmuka Pesan Tambah Training Data



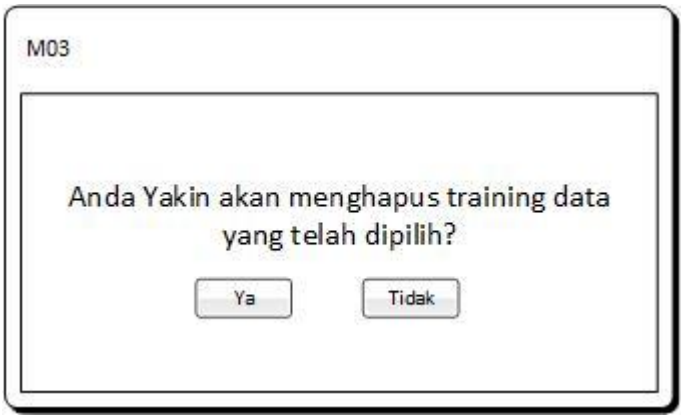
Gambar 3.25 Perancangan Antarmuka Pesan Tambah Training Data

8. Perancangan Antarmuka Pesan Kesalahan Training Data



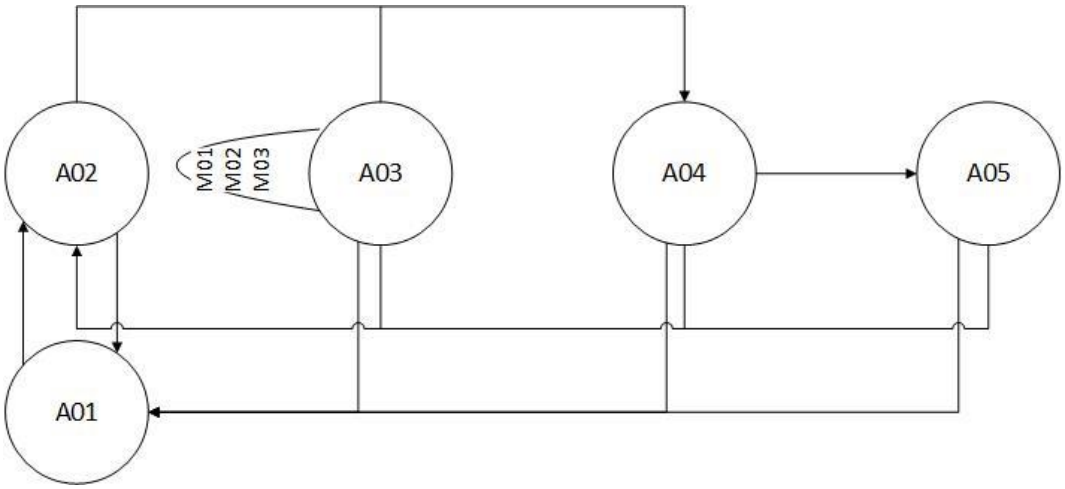
Gambar 3.26 Perancangan Antarmuka Pesan Kesalahan Training Data

9. Perancangan Antarmuka Pesan Hapus Training Data



Gambar 3.27 Perancangan Antarmuka Pesan Hapus Training Data

3.2.3. Jaringan Semantik

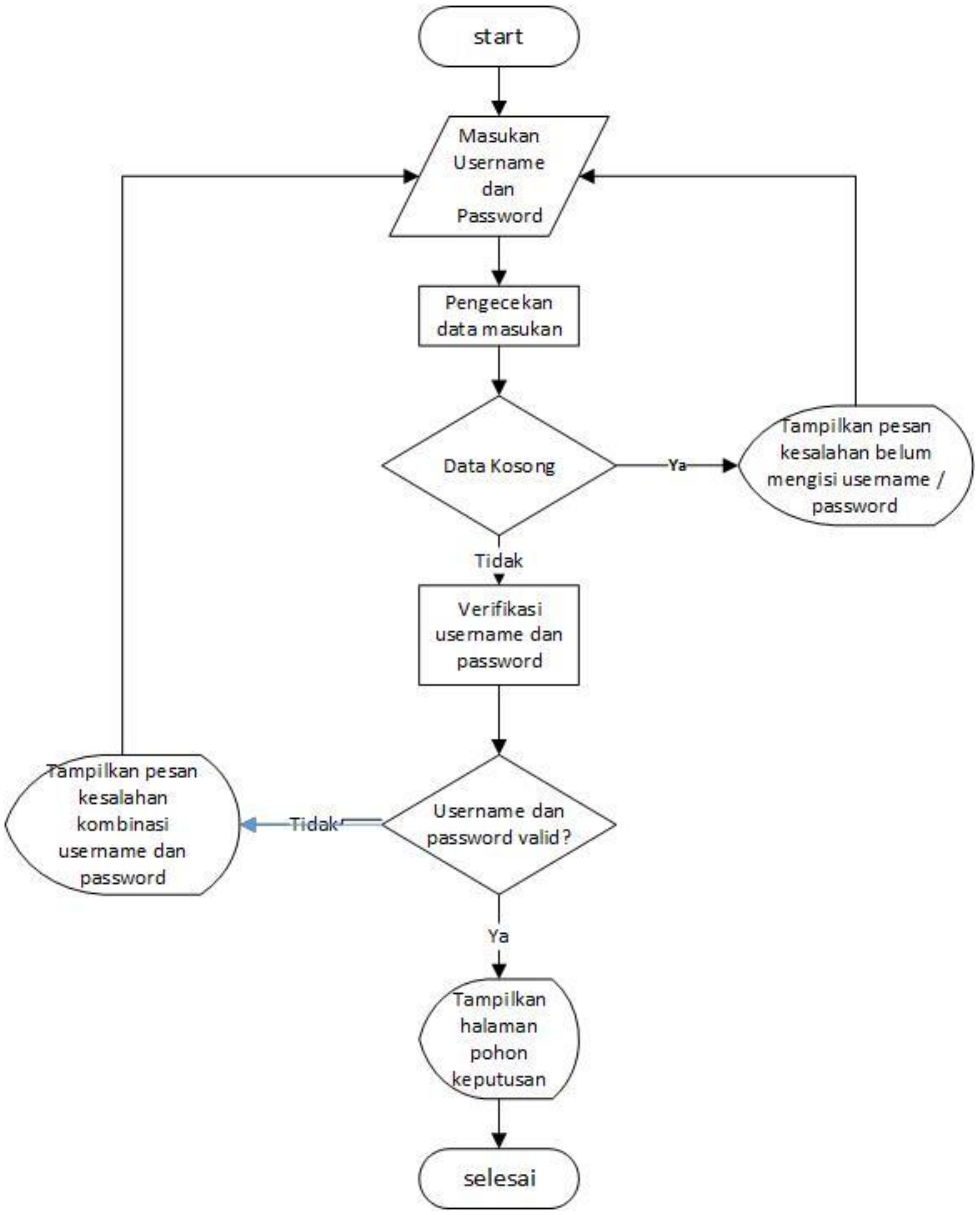


Gambar 3.28 Jaringan Semantik

3.2.4. Perancangan Prosedural

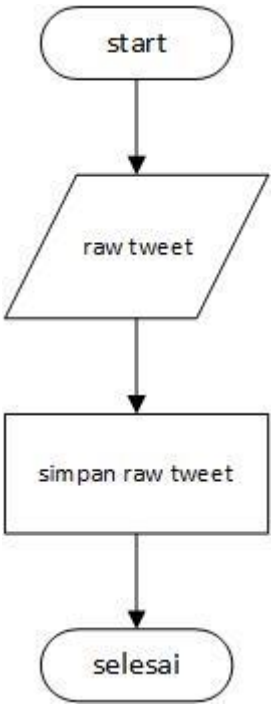
Perancangan prosedural merupakan perancangan yang dilakukan untuk menetapkan detail tahap-tahap yang akan dinyatakan ke dalam suatu perangkat lunak.

1. Prosedural Login Kelola



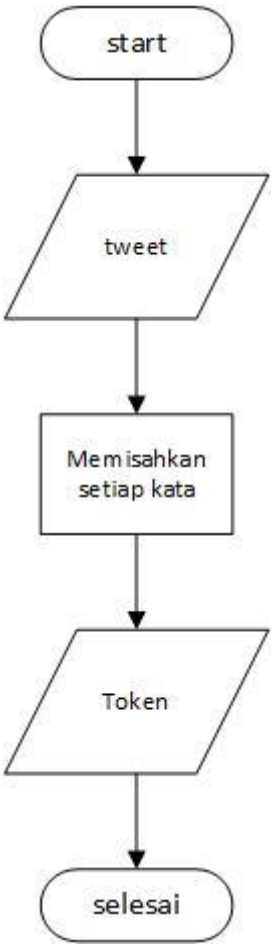
Gambar 3.29 Prosedural Login

2. Prosedural Twitter Streaming



Gambar 3.30 Prosedural Twitter Streaming

3. Prosedural Tokenization



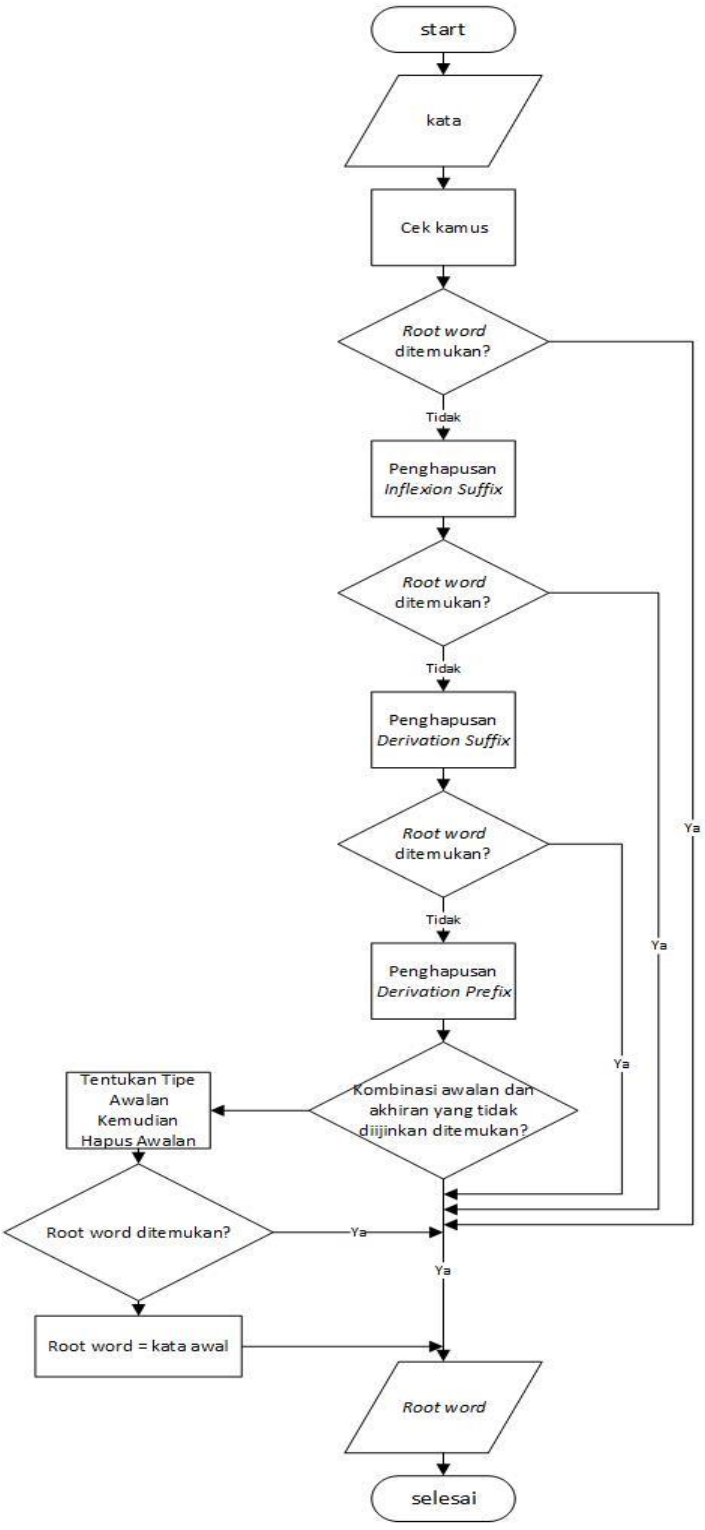
Gambar 3.31 Prosedural Casefolding

4. Prosedural Casefolding



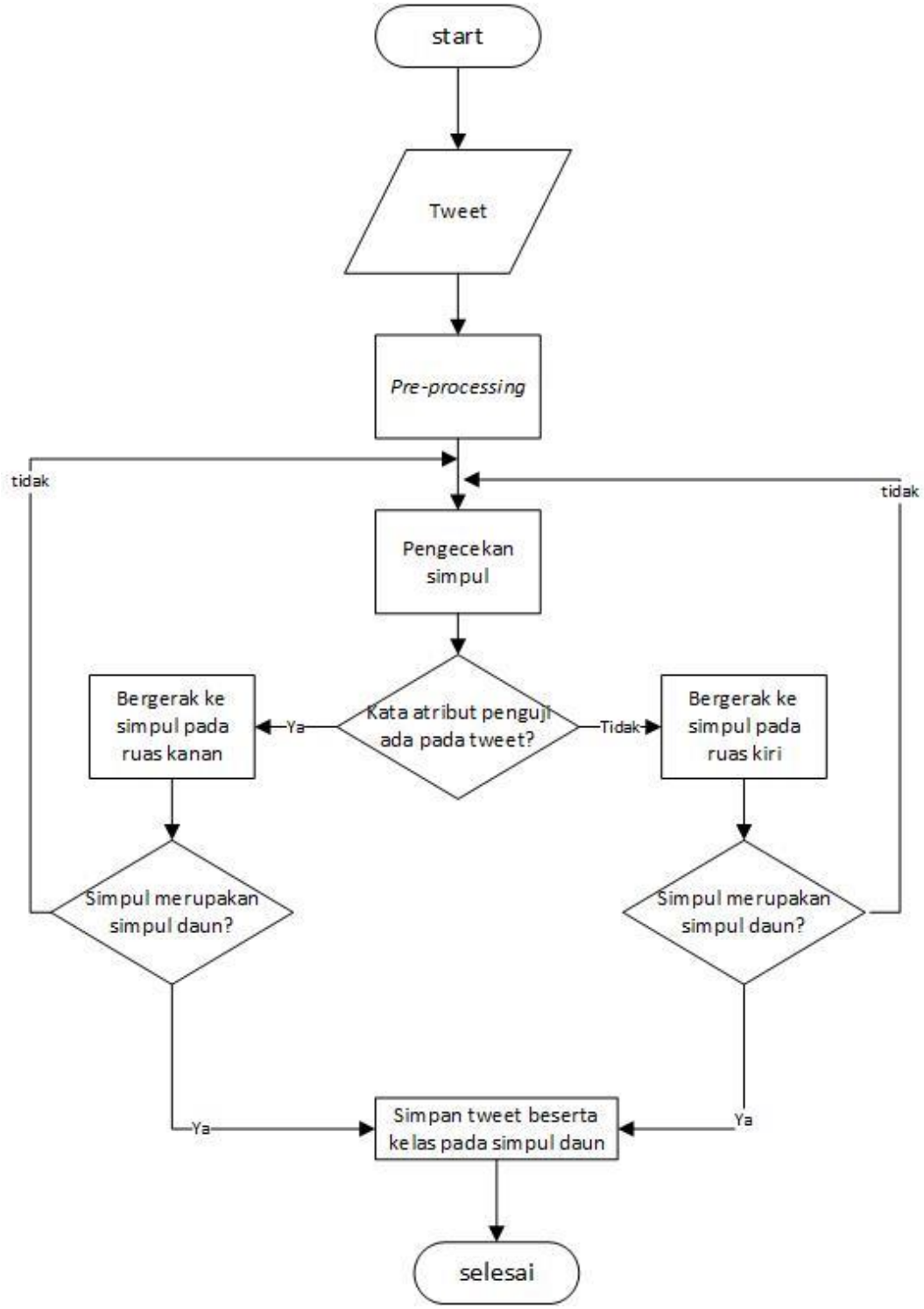
Gambar 3.32 Prosedural Casefolding

5. Prosedural Stemming



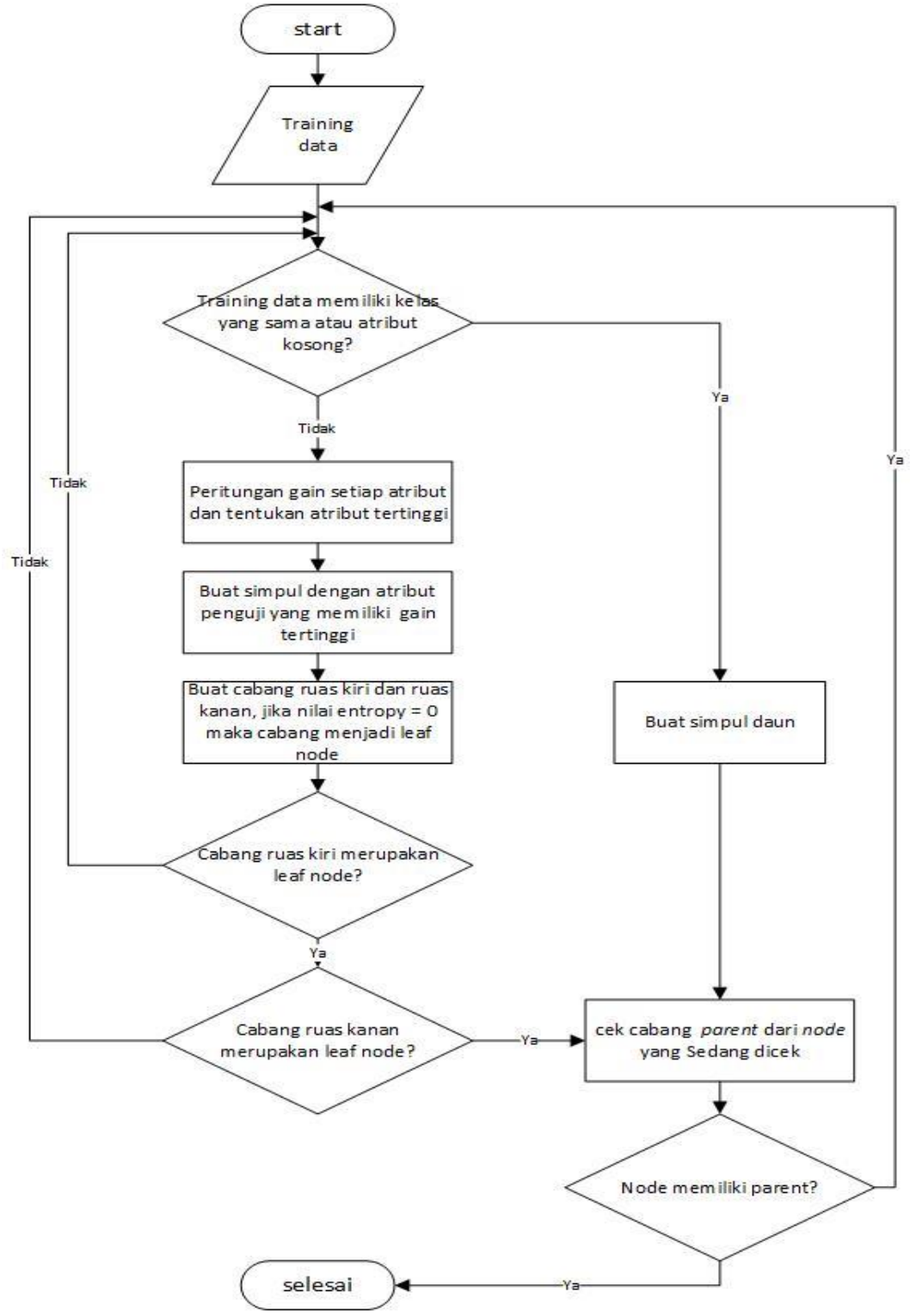
Gambar 3.33 Prosedural Stemming

6. Prosedural Klasifikasi



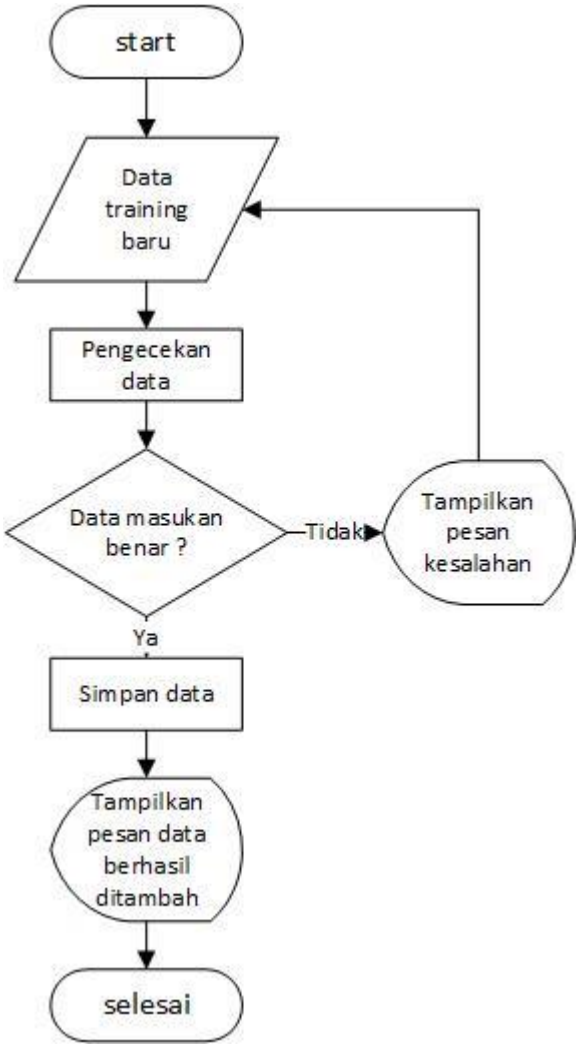
Gambar 3.34 Prosedural Klasifikasi

7. Prosedural Pembentukan Pohon Keputusan



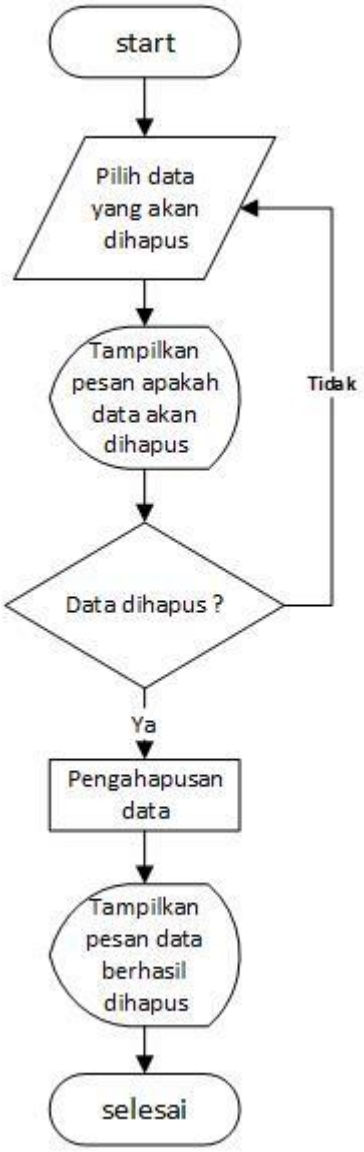
Gambar 3.35 Prosedural Pembentukan Pohon Keputusan

8. Prosedural Tambah Data Training



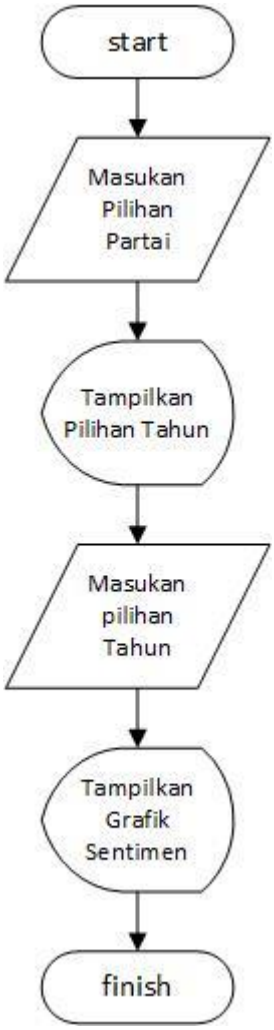
Gambar 3.36 Prosedural Tambah Data Training

9. Prosedural Hapus Data Training



Gambar 3.37 Prosedural Hapus Data Training

10. Prosedural Hasil Analisis Sentimen Per Partai



Gambar 3.38 Prisedural Hasil Analisis Sentimen Per Partai

11. Prosedural Hasil Analisis Sentimen Semua Partai



Gambar 3.39 Proseadural Hasil Analisis Sentimen Semua Partai