

BAB III ANALISIS SISTEM

3.1 Analisis Data mining

Didalam penulisan tugas akhir ini akan :

1. Dicari *node* dan *leaf*, serta *rule* yang dihasilkan dari *decision tree*.
2. Atribut pada kasus tidak semuanya akan digunakan hanya diambil 14 atribut dan 1 atribut target saja.
3. Uji data *inputan*.
4. Akurasi dan kesalahan data.
5. Di dalam data terdapat 4523 calon konsumen telemarketing bank yang akan ditawarkan untuk menandatangani uangnya.

Dengan jumlah konsumen yang cukup banyak dengan atribut target dari data ini adalah:

Tabel 3.1 Tabel Data Atribut Target

Atribut Target	Deskripsi
<i>Yes</i>	Nasabah akan menandatangani uangnya.
<i>No</i>	Nasabah tidak akan menandatangani uangnya.

3.2 Sumber Data

Data yang digunakan dalam tugas akhir ini adalah data konsumen bank yang diambil dari UCI WEKA. Data konsumen yang didapat mempunyai 4522 *record* yang terdiri dari 16 atribut dan 1 atribut target dan atribut target itu akan menjadi *Output* pada data konsumen telemarketing ini. Atribut yang ada dapat dilihat pada tabel 3.2.

Tabel 3.2 Tabel Data Konsumen Telemarketing

Atribut	Keterangan
<i>Age</i>	Merupakan atribut dari konsumen yang akan dikategorikan menjadi 3 dengan menggunakan referensi dari WHO tentang pengklasifikasian <i>age child</i> (≤ 14), <i>young - adult</i> (≥ 15 and ≤ 49), <i>old</i> (≥ 50).
<i>Job</i>	Merupakan atribut dari konsumen dengan jenis pekerjaan yang dikategorikan “ <i>admin</i> ”, “ <i>unknown</i> ”, “ <i>unemployed</i> ”, “ <i>management</i> ”, “ <i>housemaid</i> ”, “ <i>entrepreneur</i> ”, “ <i>student</i> ”, “ <i>blue-collar</i> ”, “ <i>self employed</i> ”, “ <i>retired</i> ”, “ <i>technician</i> ”, “ <i>services</i> ”.
<i>Marital</i>	Merupakan atribut dari konsumen dengan status perkawinan yang dikategorikan “ <i>married</i> ”, “ <i>divorced</i> ”, “ <i>single</i> ”.
<i>Education</i>	Merupakan atribut dari konsumen dengan status pendidikan yang dikategorikan “ <i>unknown</i> ”, “ <i>primary</i> ”, “ <i>secondary</i> ”, “ <i>tertiary</i> ”.
<i>Default</i>	Merupakan atribut dari konsumen dengan status memiliki kredit yang dikategorikan “ <i>yes</i> ”, “ <i>no</i> ”. nama atribut dirubah menjadi kredit.

Atribut	Keterangan
<i>Balance</i>	Merupakan atribut dari konsumen dengan status saldo rata-rata pertahun yang dikategorikan saldo dalam euro dan dikategorikan menggunakan GINI.
<i>Housing</i>	Merupakan atribut dari konsumen dengan status memiliki Kredit Perumahan Rakyat yang dikategorikan dengan “yes”, “no”.
<i>Loan</i>	Merupakan atribut dari konsumen dengan status memiliki pinjaman pribadi yang dikategorikan dengan “yes”, “no”.
<i>Contact</i>	Merupakan atribut dari konsumen dengan status dikontak melalui media apa yang dikategorikan “unknown”, “telephone”, “cellular”.
<i>Day</i>	Merupakan atribut dari konsumen tanggal dikontak dengan kategori numeric. Data tidak digunakan karena day terikat dengan atribut <i>month</i> sehingga variabel nya akan banyak.
<i>Month</i>	Merupakan atribut dari konsumen bulan dikontak dengan kategori januari, february, maret dan seterusnya. Data tidak digunakan karena nilai dari variabelnya banyak.
<i>Duration</i>	konsumen durasi lamanya konsumendikontak dengan kategori numeric dan data akan dikategorikan menggunakan GINI.
<i>Campaign</i>	atribut dari konsumen dengan jumlah kontak yang dilakukan selama kampanye dengan kategori <i>numeric</i>

Atribut	Keterangan
<i>Pdays</i>	Merupakan atribut dari konsumen dengan jumlah hari berlalu setelah klien dihubungi dari kampanye sebelumnya dengan kategori numeric -1 berarti klien tidak dihubungi. Atribut ini tidak digunakan karena data variabel yang unik terdapat nilai -1 yang berarti konsumen tidak dihubungi tetapi data variabel yang lain bernilai jumlah hari.
<i>Previous</i>	Merupakan atribut dari konsumen dengan jumlah kontak yang dilakukan sebelum kampanye ini dan untuk klien ini. Data akan dikategorikan menggunakan GINI.
<i>Poutcome</i>	Merupakan atribut dari konsumen dengan hasil kampanye yang dilakukan sebelum kampanye ini dilakukan dengan kategori “ <i>unknown</i> ”, “ <i>success</i> ”, “ <i>failure</i> ”, “ <i>success</i> ”.
<i>Output</i>	Atribut target konsumen yang akan mendepositokan dengan kategori <i>yes</i> atau <i>no</i> . data diubah dahulu menjadi numeric untuk mempermudah pengerjaan program yaitu <i>no</i> bernilai 1 dan <i>yes</i> bernilai 2.

3.3 Pra Proses

Untuk pra proses dan analisis data digunakan sample data yang diambil secara acak pada data dan hanya diambil 20 sample.

1. *Data selection*

Dari data yang ada atribut atau variabel yang diambil adalah:

1. *Age*
2. *Kredit*
3. *Housing*
4. *Loan*
5. *Marital*
6. *Education*
7. *Contact*
8. *Balance*
9. *Duration*
10. *Campaign*
11. *Pdays*
12. *Previous*
13. *Poutcome*
14. *Job*
15. *Output*

Pemilihan variabel-variabel tersebut dengan pertimbangan bahwa jumlah nilai variabelnya tidak banyak dan tidak unik sehingga diharapkan konsumen masuk dalam satu klasifikasi nilai variabel tersebut cukup banyak.

Tabel 3.3 Tabel Konsumen Telemarketing Atribut yang Digunakan

Age	Default	Housing	Loan	Marital	Education	Contact	Balance	Duration	Campaign	Pdays	Previous	Poutcome	Job	Output
30	no	no	no	married	primary	cellular	1787	79	1	-1	0	unknown	unemployed	1
31	no	no	no	married	secondary	cellular	132	148	1	152	1	other	services	1
28	no	yes	no	single	tertiary	unknown	11971	609	2	101	3	failure	management	1
77	no	no	no	divorced	tertiary	cellular	4659	161	1	-1	0	unknown	retired	2
46	no	yes	no	married	secondary	telephone	3098	192	9	-1	0	unknown	admin.	1
27	no	yes	no	single	primary	unknown	1128	88	4	-1	0	unknown	blue-collar	1
36	no	yes	yes	divorced	secondary	cellular	666	190	1	133	1	other	technician	1
58	no	yes	no	married	primary	cellular	1013	206	1	461	1	failure	self-employed	2
56	yes	no	no	divorced	primary	unknown	1238	1558	1	-1	0	unknown	housemaid	2
63	no	no	no	married	secondary	telephone	3904	250	2	187	2	success	entrepreneur	2
32	no	yes	no	married	secondary	cellular	1694	929	1	-1	0	unknown	student	1
60	no	no	no	married	unknown	cellular	7337	122	1	-1	0	unknown	unknown	1
78	no	yes	yes	divorced	tertiary	cellular	12100	1600	10	249	3	other	entrepreneur	1
38	no	yes	yes	married	secondary	cellular	593	1484	24	-1	0	unknown	entrepreneur	2
35	no	no	no	married	secondary	cellular	2971	429	4	164	2	failure	entrepreneur	2
37	no	yes	no	single	tertiary	cellular	145	799	2	-1	0	unknown	entrepreneur	2
52	no	yes	yes	married	tertiary	telephone	54	103	1	-1	0	unknown	entrepreneur	1
53	no	yes	no	married	primary	cellular	1034	854	2	170	3	other	entrepreneur	1
50	no	no	yes	divorced	tertiary	cellular	1790	748	3	-1	0	unknown	entrepreneur	1
54	no	no	no	married	secondary	telephone	3904	150	2	167	2	success	entrepreneur	2

2. Cleaning

Cleaning data pada kasus ini dilakukan ketika terlihat data yang tidak konsisten dan data duplikasi. Tahapan data *cleaning* yang digunakan pada tabel konsumen telemarketing yang akan dijadikan *input* pada *proses data mining* itu sendiri adalah sebagai berikut:

1. Menambahkan isi terhadap atribut yang kosong tersebut secara manual, namun pendekatan ini tidak efektif karena diterapkan pada data yang banyak.
2. Menghapus data yang sama secara manual, namun pendekatan ini tidak efektif karena diterapkan pada data yang banyak.

3. Transformation

Pada data ini terdapat beberapa variabel yang mengalami transformasi data karena data yang ada harus dirubah dulu ke dalam format untuk bisa diproses data-data yang harus ditransformasi diantaranya *age*, *balance*, *duration*, *campaign*, *previous* data-data tersebut harus ditransformasi karena bernilai *numeric* sehingga harus dikategorikan dan data *Output* dirubah kedalam bentuk *numeric* dahaulu yaitu *no* adalah 1 dan *yes* adalah 2 transformasi ini dilakukan secara manual didalam data yang akan diolah untuk mempermudah coding program. Cara mengkategorikan atribut *balance*, *duration*, *campaign*, *previous* itu adalah dengan dicari nilai GINI.

$$\text{Rumus mencari nilai gini} = 1 - \sum_j [p(j | t)]^2$$

Dalam analisis ini data diambil sample untuk mencari nilai GINI *balance*, sehingga dari hasil perhitungan *balance* yang digunakan untuk menggantikan data adalah ≤ 5998 dan > 5998 karena diantara data sample tersebut 5998 memiliki nilai GINI terkecil.

Tabel 3.4 Tabel Menentukan Nilai GINI Balance

output	no	no	no	no	no	no	no	yes	yes	no	yes	Yes	no	yes	yes	yes	no	no	no	yes																							
	54	132	145	593	666	1013	1034	1128	1238	1694	1787	1790	2971	3098	3904	3904	4659	7337	11971	12100																							
50	52	93	138	369	629	839	1023	1081	1183	1466	1740	1788	2380	3034	3501	3904	4281	5998	9654	12035																							
	≤	>	≤	>	≤	>	≤	>	≤	>	≤	>	≤	>	≤	>	≤	>	≤	>	≤	>																					
yes	0	8	0	8	0	8	1	7	2	6	2	6	3	5	3	5	3	5	3	5	4	4	4	4	4	4	4	4	4	5	3	5	3	7	1	7	1	8	0	8	0	8	0
no	0	12	1	11	2	10	2	10	2	10	3	9	3	9	4	8	5	7	5	7	6	6	7	5	8	4	8	4	9	3	9	3	9	3	9	3	9	3	10	2	11	1	
gini	0.48	0.463	0.444	0.478	0.475	0.48	0.471	0.449	0.479	0.476	0.48	0.476	0.466	0.479	0.471	0.468	0.468	<u>0.423</u>	0.444	0.463																							

Dalam analisis ini data diambil sample untuk mencari nilai GINI *duration*, sehingga dari hasil perhitungan *duration* yang digunakan untuk menggantikan data adalah ≤ 836 dan > 836 karena diantara data sample tersebut 154 memiliki nilai GINI terkecil.

Tabel 3.5 Tabel Menentukan Nilai GINI *Duration*

output	no	no	no	no	no	no	no	yes	yes	no	yes	Yes	no	yes	yes	yes	no	no	no	yes																			
	79	88	103	122	148	150	161	190	192	206	250	429	609	748	799	854	929	1484	1558	1600																			
75	77	83	95	112	135	149	155	175	191	199	228	339	519	678	773	826	891	1206	1521	1579																			
	\leq	$>$	\leq	$>$	\leq	$>$	\leq	$>$	\leq	$>$	\leq	$>$	\leq	$>$	\leq	$>$	\leq	$>$	\leq	$>$																			
yes	0	8	0	8	0	8	0	8	0	8	1	7	2	6	2	6	2	6	3	5	4	4	5	3	5	3	5	3	6	2	6	2	6	2	7	1	8	0	
no	0	12	1	11	2	10	3	9	4	8	5	7	5	7	5	7	5	7	6	6	6	6	6	6	6	7	5	8	4	9	3	10	2	11	1	11	1	11	1
gini	0.48	0.463	0.444	0.423	0.4	0.37	0.433	0.465	0.465	0.45	0.472	0.48	0.472	0.479	0.479	0.214	0.475	0.454	0.47	0.463																			

Dalam analisis ini data diambil sample untuk mencari nilai GINI *campaign*, sehingga dari hasil perhitungan *campaign* yang digunakan untuk menggantikan data adalah ≤ 17 dan > 17 karena diantara data sample tersebut 2 memiliki nilai GINI terkecil.

Tabel 3.6 Tabel Menentukan Nilai GINI *Campaign*

output	no	no	no	no	no	no	no	yes	yes	no	yes	Yes	no	yes	yes	yes	no	no	no	yes																		
	1	1	1	1	1	1	1	1	1	2	2	2	2	2	3	4	4	9	10	24																		
1	1	1	1	1	1	1	1	1	1	1	1	2	2	2	2	3	4	6	9	17																		
	\leq	$>$	\leq	$>$	\leq	$>$	\leq	$>$	\leq	$>$	\leq	$>$	\leq	$>$	\leq	$>$	\leq	$>$	\leq	$>$																		
yes	3	5	3	5	3	5	3	5	3	5	3	5	3	5	6	2	6	2	6	2	6	2	6	2	6	2	6	2	7	1	7	1	7	1	7	1		
no	6	6	6	6	6	6	6	6	6	6	6	6	6	6	8	4	8	4	8	4	8	4	8	4	8	4	8	4	9	3	10	2	10	2	11	1	12	0
gini	0.472	0.472	0.472	0.472	0.472	0.472	0.472	0.472	0.472	0.472	0.472	0.476	0.476	0.476	0.476	0.48	0.478	0.478	0.477	0.442																		

Dalam analisis ini data diambil sample untuk mentransformasikan *age*, menggunakan referensi dari WHO untuk pengklasifikasian *age child* (≤ 14), *young - adult* (≥ 15 and ≤ 49), *old* (≥ 50). Dengan melihat table sample data atribut *age* yang diklasifikasikan akan menjadi :

Tabel 3.9 Tabel Hasil Transformasi Klasifikasi Age

Age
young - adult (≥ 15 and ≤ 49)
young - adult (≥ 15 and ≤ 49)
young - adult (≥ 15 and ≤ 49)
old (≥ 50)
young - adult (≥ 15 and ≤ 49)
young - adult (≥ 15 and ≤ 49)
young - adult (≥ 15 and ≤ 49)
old (≥ 50)
old (≥ 50)
old (≥ 50)
young - adult (≥ 15 and ≤ 49)
old (≥ 50)
old (≥ 50)
young - adult (≥ 15 and ≤ 49)
young - adult (≥ 15 and ≤ 49)
young - adult (≥ 15 and ≤ 49)
old (≥ 50)
old (≥ 50)
old (≥ 50)
old (≥ 50)

4. Data mining

Apa yang dilakukan di *data mining* ini adalah memproses tabel master untuk menghasilkan klasifikasi data berupa *node* dan *leaf* untuk menghasilkan struktur tree lalu menghasilkan rule menggunakan algoritma c4.5

3.4 Penggunaan algoritma C4.5

Data yang ada terdapat 16 atribut namun dalam kasus ini dengan berbagai macam pertimbangan tidak digunakan semuanya, seperti contoh pada tabel 3.10. Untuk analisis data menggunakan algoritma C4.5 adalah dengan menggunakan 20 sample data yang diambil secara acak.

Tabel 3.10 Tabel Konsumen Telemarketing dengan seluruh atribut

Age	Job	Marital	Education	Default	Balance	Housing	Loan	Contact	Day	Month	Duration	Campaign	Pdays	Previous	Poutcome	Output
30	unemployed	married	primary	no	1787	no	no	cellular	19	oct	79	1	-1	0	unknown	1
31	services	married	secondary	no	132	no	no	cellular	7	jul	148	1	152	1	other	1
28	management	single	tertiary	no	11971	yes	no	unknown	17	nov	609	2	101	3	failure	1
77	retired	divorced	tertiary	no	4659	no	no	cellular	14	apr	161	1	-1	0	unknown	2
46	admin.	married	secondary	no	3098	yes	no	telephone	21	nov	192	9	-1	0	unknown	1
27	blue-collar	single	primary	no	1128	yes	no	unknown	30	jun	88	4	-1	0	unknown	1
36	technician	divorced	secondary	no	666	yes	yes	cellular	3	apr	190	1	133	1	other	1
58	self-employed	married	primary	no	1013	yes	no	cellular	11	aug	206	1	461	1	failure	2
56	housemaid	divorced	primary	yes	1238	no	no	unknown	5	jun	1558	1	-1	0	unknown	2
63	entrepreneur	married	secondary	no	3904	no	no	telephone	4	aug	250	2	187	2	success	2
32	student	married	secondary	no	1694	yes	no	cellular	28	jan	929	1	-1	0	unknown	1
60	unknown	married	unknown	no	7337	no	no	cellular	4	aug	122	1	-1	0	unknown	1
78	entrepreneur	divorced	tertiary	no	12100	yes	yes	cellular	3	apr	1600	10	249	3	other	1
38	entrepreneur	married	secondary	no	593	yes	yes	cellular	24	jul	1484	24	-1	0	unknown	2
35	entrepreneur	married	secondary	no	2971	no	no	cellular	4	may	429	4	164	2	failure	2
37	entrepreneur	single	tertiary	no	145	yes	no	cellular	18	jul	799	2	-1	0	unknown	2
52	entrepreneur	married	tertiary	no	54	yes	yes	telephone	18	nov	103	1	-1	0	unknown	1
53	entrepreneur	married	primary	no	1034	yes	no	cellular	19	nov	854	2	170	3	other	1
50	entrepreneur	divorced	tertiary	no	1790	no	yes	cellular	6	feb	748	3	-1	0	unknown	1
54	entrepreneur	married	secondary	no	3904	no	no	telephone	20	aug	150	2	167	2	success	2

Dari tabel 3.10 data diambil dari beberapa sample acak dan menghasilkan tabel baru untuk dianalisis dengan atribut yang telah dipilih. Tabel yang siap diolah terdapat pada table 3.11

Tabel 3.11 Tabel Konsumen Telemarketing Atribut yang digunakan

Age	Default	Housing	Loan	Marital	Education	Contact	Balance	Duration	Campaign	Pdays	Previous	Poutcome	Job	Output
30	no	no	no	married	primary	cellular	1787	79	1	-1	0	unknown	unemployed	1
31	no	no	no	married	secondary	cellular	132	148	1	152	1	other	services	1
28	no	yes	no	single	tertiary	unknown	11971	609	2	101	3	failure	management	1
77	no	no	no	divorced	tertiary	cellular	4659	161	1	-1	0	unknown	retired	2
46	no	yes	no	married	secondary	telephone	3098	192	9	-1	0	unknown	admin.	1
27	no	yes	no	single	primary	unknown	1128	88	4	-1	0	unknown	blue-collar	1
36	no	yes	yes	divorced	secondary	cellular	666	190	1	133	1	other	technician	1
58	no	yes	no	married	primary	cellular	1013	206	1	461	1	failure	self-employed	2
56	yes	no	no	divorced	primary	unknown	1238	1558	1	-1	0	unknown	housemaid	2
63	no	no	no	married	secondary	telephone	3904	250	2	187	2	success	entrepreneur	2
32	no	yes	no	married	secondary	cellular	1694	929	1	-1	0	unknown	student	1
60	no	no	no	married	unknown	cellular	7337	122	1	-1	0	unknown	unknown	1
78	no	yes	yes	divorced	tertiary	cellular	12100	1600	10	249	3	other	entrepreneur	1
38	no	yes	yes	married	secondary	cellular	593	1484	24	-1	0	unknown	entrepreneur	2
35	no	no	no	married	secondary	cellular	2971	429	4	164	2	failure	entrepreneur	2
37	no	yes	no	single	tertiary	cellular	145	799	2	-1	0	unknown	entrepreneur	2
52	no	yes	yes	married	tertiary	telephone	54	103	1	-1	0	unknown	entrepreneur	1
53	no	yes	no	married	primary	cellular	1034	854	2	170	3	other	entrepreneur	1
50	no	no	yes	divorced	tertiary	cellular	1790	748	3	-1	0	unknown	entrepreneur	1
54	no	no	no	married	secondary	telephone	3904	150	2	167	2	success	entrepreneur	2

Dalam kasus yang tertera pada tabel 3.11 akan dibuat pohon keputusan untuk menentukan konsumen mendepositokan uangnya atau tidak dengan melihat atribut *age*, *kredit*, *housing*, *loan*, *marital*, *education*, *contact*, *balance*, *duration*, *campaign*, *pdays*, *previous*, *poutcome*, *job*, *Output*. Untuk memilih atribut sebagai akar, didasarkan pada nilai *gain* tertinggi dari atribut-atribut yang ada. dan dicari entropinya.

Berikut ini adalah contoh dari analisis data mengenai langkah-langkah dalam pembentukan *decision tree* dengan menggunakan algoritma C4.5 untuk menyelesaikan kasus pada tabel 3.11.

1. Menghitung jumlah kasus, jumlah kasus untuk keputusan Ya, jumlah kasus untuk keputusan Tidak, dan *Entropy* dari semua kasus dan kasus yang dibagi berdasarkan atribut *age*, *marital*, *education*, *default*, *balance*, *housing*, *loan*,

<i>node</i>			Jumlah kasus	<i>no</i>	<i>yes</i>	entropy	gain
	<i>duration</i>	≤ 826	15	9	6	0.974	0.001
		> 826	5	3	2	0.974	
	<i>campaign</i>	≤ 17	19	12	7	0.953	0.07
		> 17	1	0	1	0	
	<i>Pdays</i>	≤ 158	14	10	4	0.866	0.093
		> 158	6	2	4	0.921	
	<i>previous</i>	≤ 2	17	9	8	1	0.125
		> 2	3	3	0	0	
	<i>Poutcome</i>	<i>Failure</i>	3	1	2	0.921	0.316
		<i>Unknown</i>	11	7	4	0.948	
		<i>Succes</i>	2	0	2	0	
		<i>Other</i>	4	4	0	0	
	<i>Job</i>	<i>unemployed</i>	1	1	0	0	
		<i>services</i>	1	1	0	0	
		<i>management</i>	1	1	0	0	
		<i>retired</i>	1	0	1	0	
		<i>admin.</i>	1	1	0	0	
		<i>blue-collar</i>	1	1	0	0	
		<i>technician</i>	1	1	0	0	
		<i>self-employed</i>	1	0	1	0	
		<i>housemaid</i>	1	0	1	0	
		<i>entrepreneur</i>	9	4	5	0.994	0.527
		<i>student</i>	1	1	0	0	
		<i>unknown</i>	1	1	0	0	

Baris TOTAL kolom entropy pada tabel 3.12 dihitung dengan persamaan 2 sebagai berikut:

$$Entropy (\text{Total}) = \left(-\frac{12}{20} * \log_2 \left(\frac{12}{20}\right)\right) + \left(-\frac{8}{20} * \log_2 \left(\frac{8}{20}\right)\right)$$

$$Entropy (\text{Total}) = 0.975$$

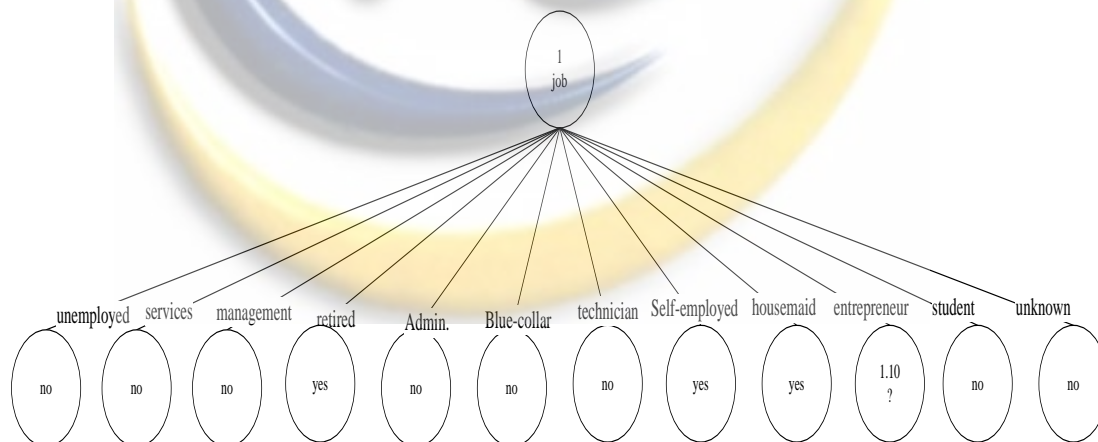
Sementara itu, nilai *Gain* pada baris *age* dihitung dengan menggunakan persamaan 1 sebagai berikut :

$$Gain(\text{Total}, \text{Age}) = Entropy(\text{Total}) - \sum_{i=1}^n \left(\left| \frac{Gain}{Total} \right| \right) * Entropy(\text{age})$$

$$Gain(\text{Total}, \text{Age}) = 0.975 - \left(\left(-\frac{10}{20} * 0.884 \right) + \left(-\frac{10}{20} * 1 \right) \right)$$

$$Gain(\text{Total}, \text{Age}) = 0.053$$

Dari hasil tabel 3.12 dapat diketahui bahwa atribut dengan *gain* tertinggi adalah *job*, yaitu sebesar 0,527 Dengan demikian *job* dapat dijadikan *node* akar. Ada 12 nilai atribut dari *job* yaitu *unemployed, services, management, retired, admin, blue-collar, technician, self-employed, housemaid, entrepreneur, student, unknown* dari ke 12 atribut itu *retired, self-employed, housemaid*, sudah mengklasifikasikan menjadi *yes* dan *unemployed, services, management, admin., blue-collar, technician student, unknown* menjadi *no* sehingga tidak memerlukan perhitungan lebih lanjut, tetapi untuk atribut *entrepreneur* masih diperlukan perhitungan lagi. Dari hasil tersebut dapat digambarkan pohon keputusan sementara Nampak seperti gambar 3.1

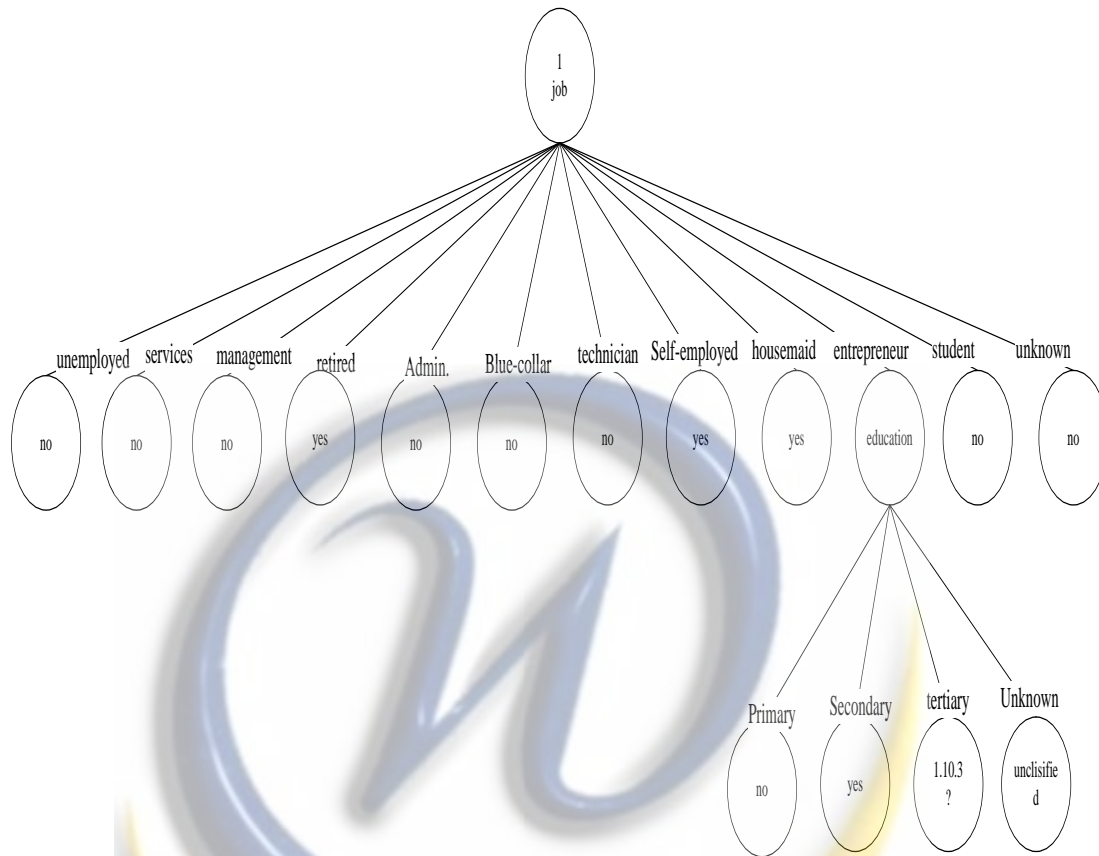


Gambar 3.1 Gambar *Tree* Hasil Perhitungan *Node* 1

2. Menghitung jumlah kasus, jumlah kasus untuk keputusan *yes* dan jumlah kasus untuk keputusan *no*, dan *entropy* dari semua kasus dan kasus yang

<i>node</i>		jumlah kasus	<i>no</i>	<i>yes</i>	entropy	gain
	<i>Phone</i>	3	1	2	0.921	
	<i>Unknown</i>	0	0	0	0	
<i>balance</i>	≤ 5988	8	3	5	0.958	0.143
	> 5988	1	1	0	0	
<i>duration</i>	≤ 826	6	2	4	0.921	0.021
	> 826	3	2	1	0.921	
<i>campaign</i>	≤ 17	8	4	4	1	0.106
	> 17	1	0	1	0	
<i>Pdays</i>	≤ 158	4	2	2	1	0.003
	> 158	5	2	3	0.974	
<i>previous</i>	≤ 2	7	2	5	0.866	0.321
	> 2	2	2	0	0	

Dari hasil tabel 3.13 dapat diketahui bahwa atribut dengan gain tertinggi yaitu *education* yaitu sebesar 0,633. Dengan demikian *education* dapat menjadi *node* cabang dari nilai atribut *failure*. Dari atribut *kredit* terdapat 4 atribut yaitu *primary*, *secondary*, *tertiary*, *unknown*. *primary* sudah mengkalsifikasikan dirinya menjadi *no* dan *secondary* sudah mengklasifikasikan dirinya menjadi *yes* sedangkan *unknown* unclisified karena pada *Output* bernilai 0. Pada atribut *tertiary* belum terklasifikasi sehingga masih memerlukan perhitungan selanjutnya. Dari hasil tersebut dapat digambarkan pohon keputusan sementara Nampak seperti gambar 3.2

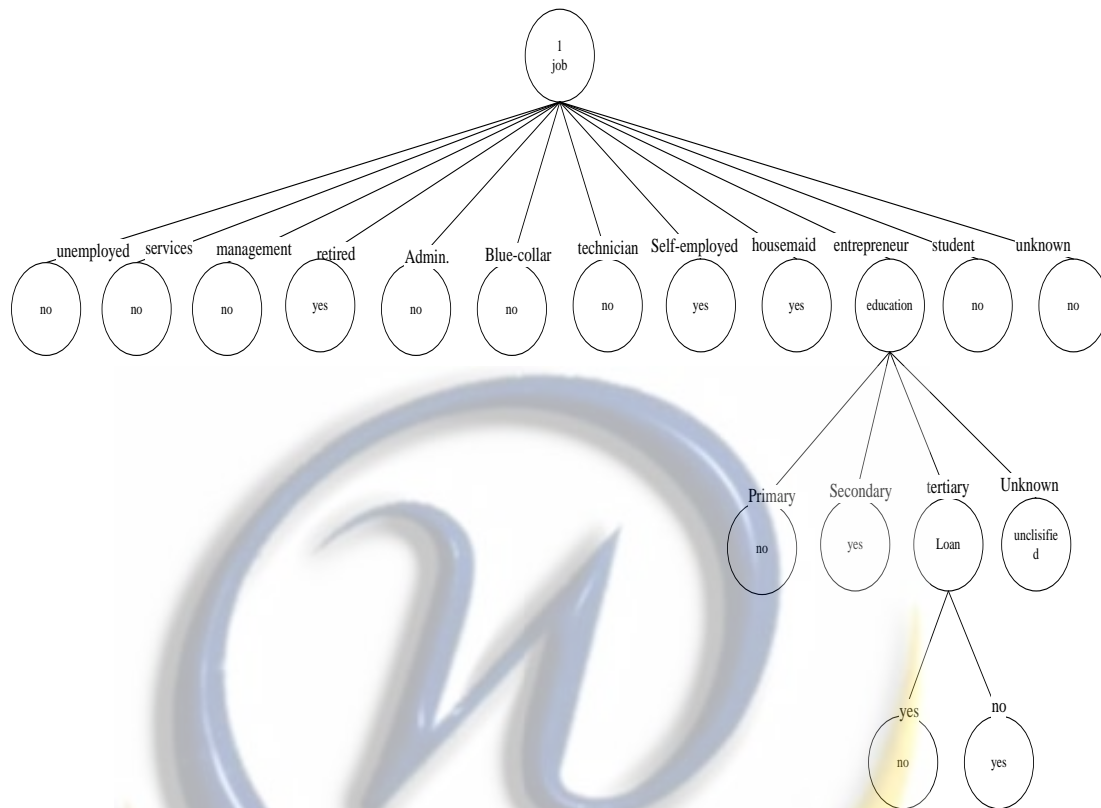


Gambar 3.2 Gambar *Tree* Hasil Perhitungan *Node 1.10*

3. Menghitung jumlah kasus, jumlah kasus untuk keputusan *yes* dan jumlah kasus untuk keputusan *no*, dan *entropy* dari semua kasus dan kasus yang dibagi ke dalam *age*, *kredit*, *housing*, *loan*, *marital*, *contact*, *balance*, *duration*, *campaign*, *pdays*, *previous*, *poutcome* yang dapat menjadi *node* akar dari nilai atribut *tertiary*. Setelah itu dilakukan perhitungan *gain* untuk tiap-tiap atribut hasil perhitungan ditunjukkan pada tabel 3.14.

<i>node</i>			jumlah kasus	<i>no</i>	<i>yes</i>	entropy	gain
	<i>balance</i>	≤ 5988	3	2	1	0.921	0.124
		> 5988	1	1	0	0	
	<i>duration</i>	≤ 826	3	2	1	0.921	0.124
		> 826	1	1	0	0	
	<i>campaign</i>	≤ 17	4	3	1	0.814	0
		> 17	0	0	0	0	
	<i>Pdays</i>	≤ 158	2	1	1	1	
		> 158	2	2	0	0	
	<i>Poutcome</i>	<i>Failure</i>	0	0	0	0	0.214
		<i>Unknown</i>	3	2	1	0.921	0.314
		<i>Succes</i>	0	0	0	0	
		<i>Other</i>	1	1	0	0	

Dari hasil tabel 3.14 dapat diketahui bahwa atribut dengan *gain* tertinggi yaitu *loan* yaitu sebesar 0.814. Dengan demikian *loan* dapat menjadi *node* cabang dari nilai atribut *tertiary*. Dari atribut *loan* terdapat 2 atribut yaitu *yes*, *no* dan dari ke 2 atribut *yes* sudah terklasifikasi *no* dan *no* sudah terklasifikasi *yes*. karena atribut sudah terklasifikasi maka tidak perlu dilakukan perhitungan lebih lanjut. Dari hasil tersebut dapat digambarkan pohon keputusan sementara nampak seperti gambar 3.3



Gambar 3.3 Gambar *Tree* Hasil Perhitungan *Node 1.2*

3.4.1 Rule Tree

Setelah tree terbentuk maka akan dihasilkan sejumlah rule dalam tree tersebut.

Contoh rule yang dapat terbentuk dari tree pada gambar 3.3 adalah sebagai berikut:

"IF Job = Unemployed THEN deposito = No"

"IF Job = Services THEN deposito = No"

"IF Job = Management THEN deposito = No"

"IF Job = Retired THEN deposito = Yes"

"IF Job = Admin. THEN deposito = No"

"IF Job = Blue-collar THEN deposito = No"

"IF Job = technician THEN deposito = No"

"IF Job = Self-employed THEN deposito = yes"

"IF Job = Housemaid THEN deposito = Yes"

"IF Job = Student THEN deposito = No"

"IF Job = Unknown THEN deposito = No"

"IF Job = Entrepreneur and Education = Primary THEN deposito = No"

"IF Job = Entrepreneur and Education = Secondary THEN deposito = Yes"

"IF Job = Entrepreneur and Education = Unknown THEN deposito = Unclisified"

"IF Job = Entrepreneur and Education = Tertiary and Loan = Yes THEN deposito = No"

"IF Job = Entrepreneur and Education = Tertiary and Loan = No THEN deposito = Yes"

3.5 Hasil Analisis Kebutuhan Sistem

1. Dibutuhkan database untuk menyimpan data untuk kasus ini.
2. Dibutuhkan suatu aplikasi untuk menghasilkan gain dan atribut untuk menghasilkan *decision tree* beserta rule-rule nya.

3.6 Pemodelan Sistem Yang Dibangun

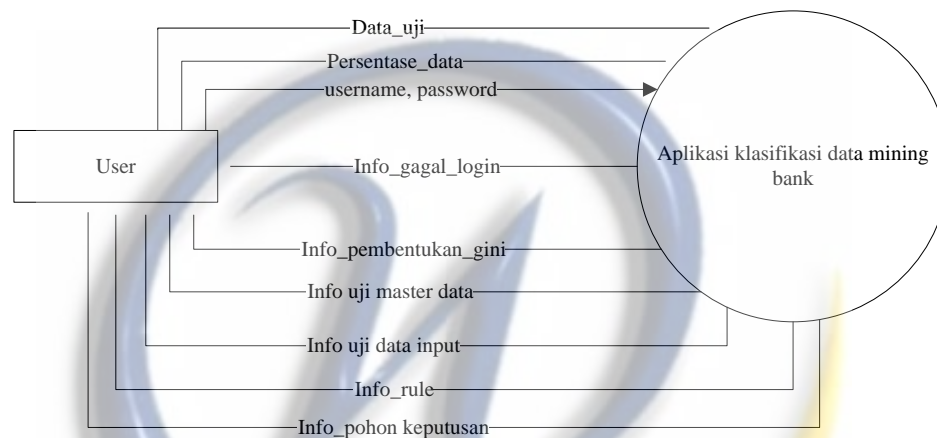
Kebutuhan fungsional sistem yang akan dibangun adalah :

1. Sistem mampu melakukan proses perhitungan GINI untuk dilakukan proses selanjutnya
2. Sistem mampu melakukan proses pembuatan struktur pohon keputusan.
3. Sistem mampu melakukan proses pembentukan *rule*.
4. Sistem mampu melakukan proses uji data *input*.
5. Sistem mampu melakukan proses pengujian master data untuk dilihat akurasi dari data yang diuji.
6. Sistem mempunyai proses pendukung login untuk keamanan program.

3.6.1 Deskripsi Proses DCD Level 0 yang disarankan

Deskripsi ini menggambarkan proses yang dilakukan oleh setiap modul sesuai hak akses pengguna aplikasi. Perancangan proses ini menggunakan tools *Data Flow Diagram* (DFD).

Data Context Diagram (DFD level 0) pada aplikasi klasifikasi *data mining* bank ini adalah sebagai berikut:



Gambar 3.4 Data Context Diagram (DCD) Level 0

Data Context Diagram (DCD) yang menggunakan aplikasi klasifikasi *data mining* bank ini adalah :

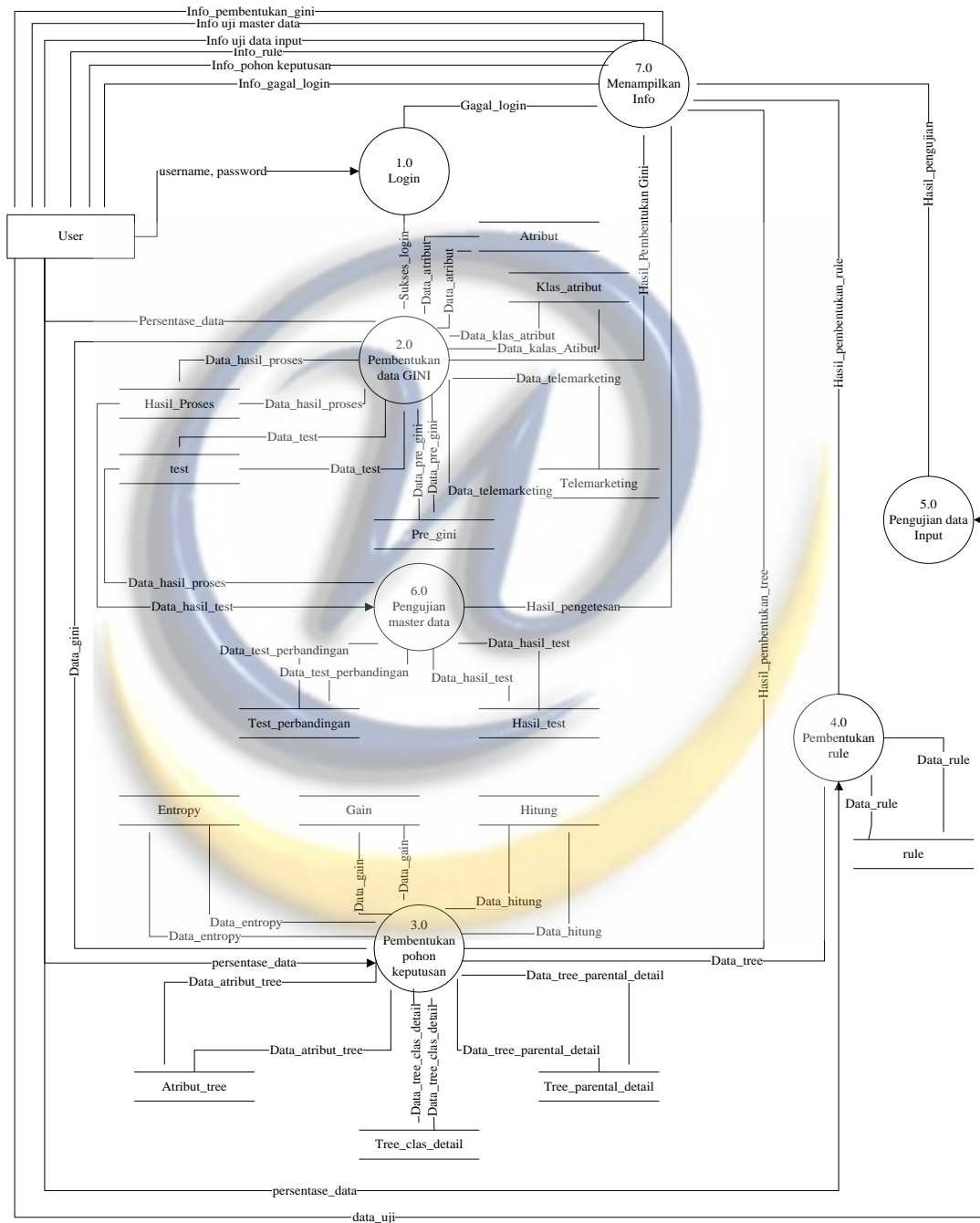
User adalah adalah pengguna dari aplikasi ini yang dapat mengakses data dan menginput data untuk mengetahui proses *data mining* klasifikasi bank ini.

3.6.2 Deskripsi Proses DFD

3.6.2.1 DFD Level 1 yang diusulkan

Pada data flow diagram dapat diketahui penjabaran dari diagram *context* dimana pada level ini terdapat proses-proses yang terdapat aplikasi. Aliran informasi dari entitas luar semakin jelas dapat diketahui, dimana volume arus informasi yang

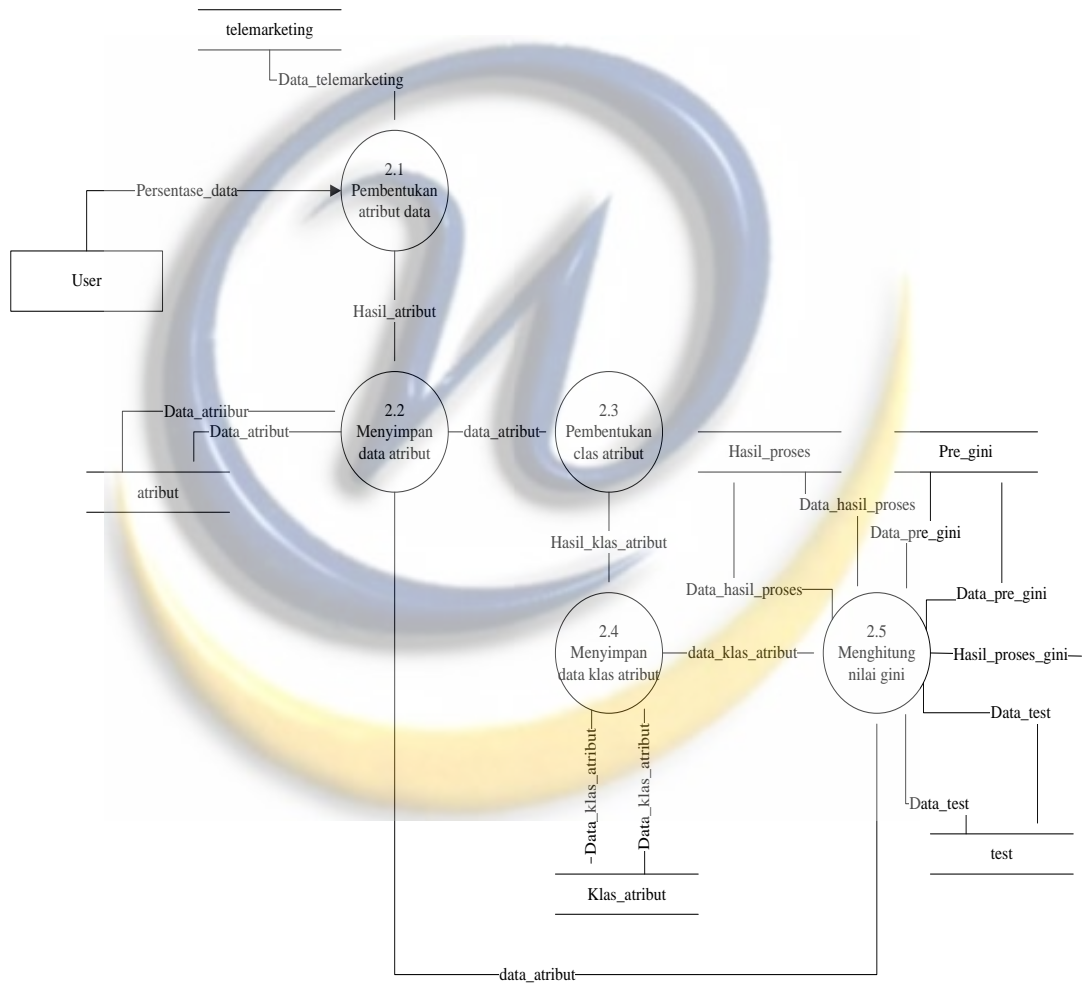
harus dikendalikan. Interaksinya berupa kerjasama antara sistem aplikasi klasifikasi *data mining* bank yang diusulkan dapat dilihat pada gambar dibawah ini:



Gambar 3.5 Data Flow Diagram (DFD) Level 1

3.6.2.2 DFD Level 2 Proses 2 yang diusulkan

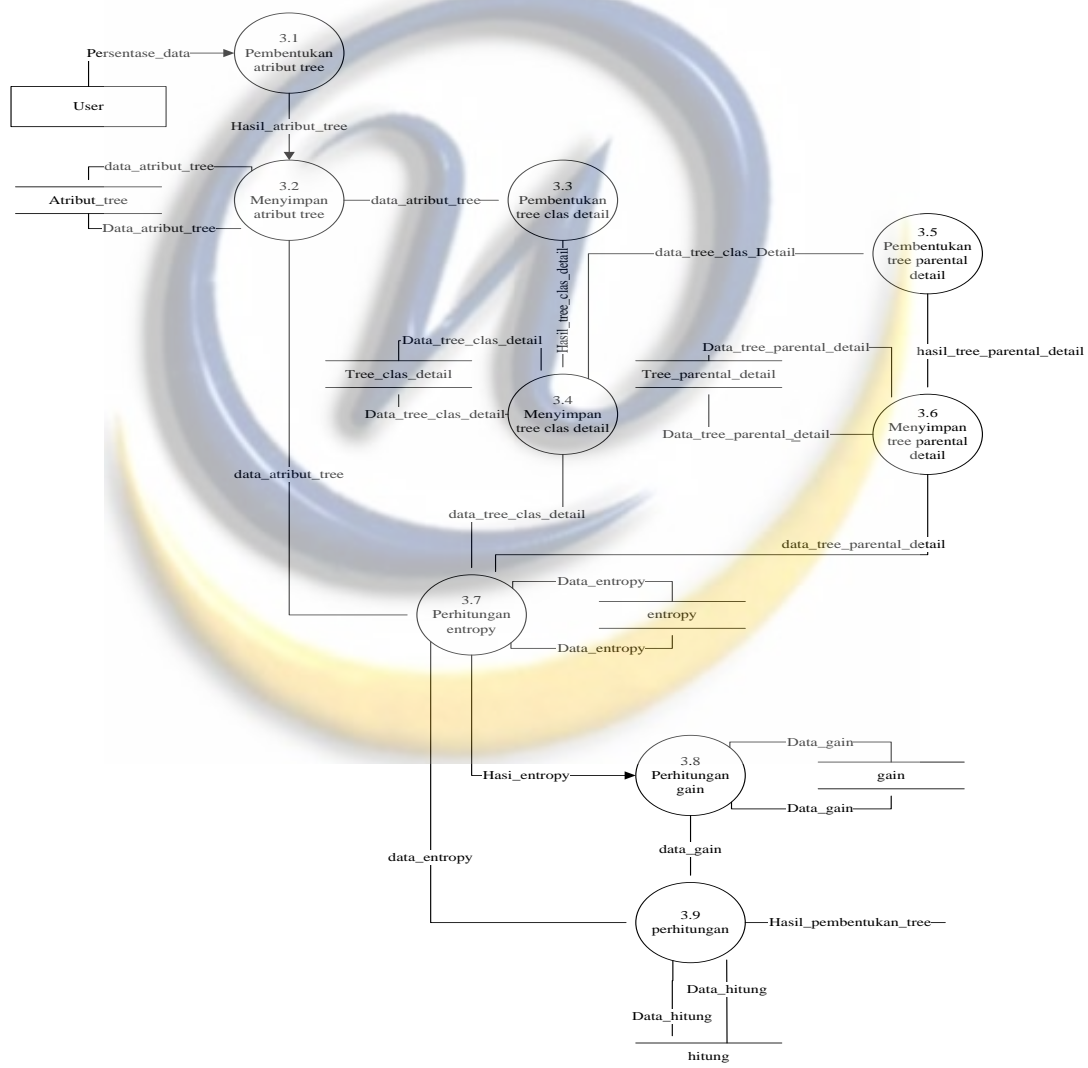
Pembentukan data GINI juga dapat dikembangkan menjadi Data Flow Diagram (DFD) Level 2. Terdapat 4 proses yang terlibat yaitu proses pembentukan atribut data, menyimpan data atribut, pembentukan clas atribut, menyimpan data clas atribut, menghitung nilai GINI. Gambaran proses tersebut dapat dilihat dalam diagram dibawah ini.



Gambar 3.6 Data Flow Diagram (DFD) Level 2 Proses 2

3.6.2.3 DFD Level 2 Proses 3 yang diusulkan

Proses pembentukan pohon keputusan dapat dikembangkan menjadi Data Flow Diagram (DFD) Level 2. Terdapat 9 proses yang terlibat yaitu proses pembentukan atribut tree, menyimpan atribut tree, pembentukan tree clas detail, menyimpan tree clas detail, pembentukan tree parental detail, menyimpan tree parental detail, perhitungan entropy, perhitungan gain, perhitungan. Gambaran proses tersebut dapat dilihat dalam diagram dibawah ini.



Gambar 3.7 Data Flow Diagram (DFD) Level 2 Proses 3

3.6.3 Kamus Data

Berikut ini akan dijelaskan data-data yang digunakan dalam perancangan proses (*Data Flow Diagram*)

Tabel 3.15 Kamus Data

No.	Istilah	Penjelasan
1	<i>User</i>	Pengguna yang secara langsung berhubungan dengan sistem.
2	<i>UserName, password</i>	<i>UserName</i> merupakan nama pengguna sistem yang telah terdaftar sebelumnya dan <i>password</i> merupakan kode, kata atau sandi unik yang digunakan sebelum masuk ke sebuah sistem misalnya pada proses <i>Login</i> .
3	Telemarketing	<i>Datastore</i> bank yang berisi data telemarketing bank.
4	Atribut	<i>Datastore</i> atribut yang berisi data atribut yang akan digunakan dalam proses mining.
6	<i>Atribut_tree</i>	<i>Datastore tree</i> yang berisi <i>kode_tree</i> yang berguna dalam pembentukan struktur <i>tree</i> .
7	<i>Pre_gini</i>	<i>Datastore</i> GINI yang berisi nama atribut dan nilai GINI yang digunakan untuk mensplit atribut yang belum terklasifikasi.
8	<i>gain</i>	<i>Datastore gain</i> yang berisi nama atribut

No.	Istilah	Penjelasan
		dan nilai gain dari atribut yang ada.
9	<i>Entropy</i>	<i>Datastore entropy</i> yang berisi nama atribut dan nilai entropy dari atribut yang ada.
10	<i>Klas_atribut</i>	<i>Datastore klas_atribut</i> yang berisi data untuk mengklasifikasikan atribut.
11	hitung	<i>Datastore</i> hitung digunakan untuk menyimpan data hasil perhitungan nilai gain dan entropy yang berguna untuk menentukan <i>node</i> tree.
12	<i>Tree_klas_detail</i>	<i>Datastore tree_klas_detail</i> digunakan untuk menentukan result hasil <i>Output</i> dari struktur tree.
13	<i>Tree_parental_detail</i>	<i>Datastore tree_parental_detail</i> digunakan untuk menentukan induk dari tree.
14	<i>Rules</i>	<i>Datastore rules</i> digunakan untuk menyimpan rules.
15	Hasil_proses	<i>Datastore</i> hasil_proses yang digunakan untuk menyimpan data hasil proses
16	<i>Test</i>	<i>Data store</i> untuk menyimpan data uji.
17	Hasil_test	<i>Data store</i> untuk menyimpan data hasil test
18	<i>Test_perbandingan</i>	<i>Data store</i> untuk membandingkan data latih dan data uji.
19	Data_telemarketing	Data yang berasal dari <i>datastore</i> telemarketing.

No.	Istilah	Penjelasan
20	Data_hitung	Data yang berasal dari <i>datastore</i> hitung.
21	Data_gain	Data yang berasal dari <i>datastore</i> gain.
22	Data_entropy	Data yang berasal dari <i>datastore</i> entropy.
23	Data_atribut	Data yang berasal dari <i>Datastore</i> Atribut.
24	Data_tree_klas_detail	Data yang berasal dari <i>datastore</i> tree_klas_detail.
25	Data_tree_parental_detail	Data yang berasal dari <i>datastore</i> tree_parental_detail.
26	Data_klas_atribut	Data yang berasal dari <i>datastore</i> klas_atribut.
27	Data_Pre_gini	Data yang berasal dari <i>datastore</i> pre_gini
28	Data_Atribut_tree	Data yang berasal dari <i>datastore</i> atribut_tree.
29	Data_Test	Data yang berasal dari <i>datastore</i> test.
30	Data_hasil_proses	Data yang berasal dari <i>datastore</i> hasil_proses.
40	Data_hasil_test	Data yang berasal dari <i>datastore</i> hasil_test.
41	Data_test_perbandingan	Data yang berasal dari <i>datastore</i> test_perbandingan.
42	Data_pohon_keputusan	Data yang diambil dari proses <i>data mining</i> untuk menampilkan struktur pohon keputusan.
43	Data_uji_data	Data yang diambil dari proses <i>data</i>

No.	Istilah	Penjelasan
		<i>mining</i> untuk menguji data yang dihasilkan oleh rule.
44	Data_rule	Data yang diambil dari proses <i>data mining</i> untuk menampilkan rule yang dihasilkan.
45	Gagal_login	Keterangan apabila <i>Id_User</i> dan password tidak tersedia dalam aplikasi sistem salah dalam memasukan <i>id_User</i> dan <i>password</i> .
46	Sukses_login	Keterangan apabila <i>id_User</i> dan password terdaftar dalam sistem atau pengguna (<i>admin</i>) telah berhasil memasukkan <i>id_User</i> dan <i>password</i> dengan benar.
47	Info_uji_data <i>input</i>	Semua data yang ditampilkan aplikasi sistem mengenai data yang diuji.
48	Info_uji_master_data	Semua data yang ditampilkan mengenai data latih dan data uji.
49	Info_rule	Semua data yang ditampilkan aplikasi sistem mengenai rule hasil pembacaan terhadap struktur tree.
50	Info_pohon keputusan	Semua data yang ditampilkan aplikasi sistem dalam bentuk struktur tree.
51	Info_gagal_login	Pesan yang ditampilkan apabila salah dalam memasukan <i>id_User</i> dan <i>password</i> .
52	Data_gini	Data yang telah diproses dan

No.	Istilah	Penjelasan
		menghasilkan data GINI yang diperlukan untuk proses selanjutnya.
53	Data_tree	Data yang telah diproses dan menghasilkan data tree yang diperlukan untuk proses selanjutnya.
54	Persentase_data	Besaran data yang diinput oleh User.
55	Data_uji	Data yang diinput oleh User untuk diuji kebenarannya.
56	Hasil_pembentukan_tree	Hasil data yang diambil dari proses pembentukan pohon keputusan.
57	Hasil_pembentukan_rule	Hasil data yang diambil dari proses pembentukan rule.
58	Hasil_pengujian	Hasil data yang diambil dari proses pengujian data input.

3.6.4 Proses *Spesification* (PSPEC)

Pada aplikasi yang akan dibuat dalam bentuk menggambarkan dan menjelaskan segala proses yang ada dalam aplikasi klasifikasi *data mining* bank. Beberapa spesifikasi proses yang akan dijelaskan.

3.6.4.1 Proses 7.0 menampilkan info

<i>Input</i>	Gagal_login, hasil_pembentukan_tree, hasil_pembentukan_rule, hasil_pengujian, hasil_pengetesan.
<i>Output</i>	Info_gagal_login, info_pohon_keputusan, info_rule, info_uji_data_input, info_uji_data_master,

	info_pembentukan_gini.
<i>Datastore</i>	-
Skenario	Sistem membaca <i>input</i> yang diberikan <i>User</i> maka sistem akan melakukan proses dan memberikan <i>Output</i> yang dipilih oleh <i>User</i> dengan menampilkan <i>Output</i> berupa informasi.

3.6.4.2 Proses 2.1 Pembentukan Atribut Data

<i>Input</i>	Persentase data, telemarketing
<i>Output</i>	Hasil_atribut
<i>Datastore</i>	telemarketing
Skenario	<i>User</i> melakukan <i>input</i> persentase data lalu program akan memproses <i>input</i> data <i>User</i> tadi. Program membaca <i>datastore</i> yang akan diproses untuk menghasilkan <i>Output</i> berupa hasil_atribut yang telah diberikan atribut.

3.6.4.3 Proses 2.2 Menyimpan Data Atribut

<i>Input</i>	hasil_atribut
<i>Output</i>	data_atribut
<i>Datastore</i>	atribut
Skenario	Hasil_atribut data yang telah diproses tadi akan disimpan ke dalam <i>datastore</i> atribut

3.6.4.4 Proses 2.3 Pembentukan klas Atribut

<i>Input</i>	data_atribut
<i>Output</i>	Hasil_klas_atribut
<i>Datastore</i>	

Skenario	Sistem akan membaca <i>data_atribut</i> lalu sistem akan melakukan proses pembeberian <i>clas_atribut</i> dan menghasilkan <i>Output</i> berupa data yang sudah terbentuk menjadi <i>klas_atribut</i> .
----------	---

3.6.4.5 Proses 2.4 Menyimpan Data klas Atribut

<i>Input</i>	Hasil_klas_atribut
<i>Output</i>	data_klas_atribut
<i>Datastore</i>	Klas_atribut
Skenario	Hasil_klas_atribut tadi kemudian akan disimpna ke dalam <i>datastore</i> <i>klas_atribut</i>

3.6.4.6 Proses 2.5 Menghitung Nilai GINI

<i>Input</i>	data_klas_atribut, data_atribut
<i>Output</i>	Hasil_proses_gini
<i>Datastore</i>	Hasil_proses, Pre_gini, test
Skenario	Sistem akan membaca data yang sudah diberikan <i>data_klas_atribut</i> dan <i>data_atribut</i> lalu dihitung nilai GINI nya dan data disimpan ke dalam 3 <i>datastore</i> yang berbeda.

3.6.4.7 Proses 3.1 Pembentukan Atribut Tree

<i>Input</i>	Persentase data
<i>Output</i>	Hasil_atribut_tree
<i>Datastore</i>	
Skenario	<i>User</i> melakukan <i>input</i> persentase data lalu program akan memproses <i>input</i> data <i>User</i> tadi. Program menghasilkan <i>Output</i> berupa <i>hasil_atribut_tree</i> yang

	telah diberikan atribut.
--	--------------------------

3.6.4.8 Proses 3.2 Menyimpan Atribut Tree

<i>Input</i>	Hasil_atribut_tree
<i>Output</i>	data_atribut_tree
<i>Datastore</i>	Atribut_tree
Skenario	Hasil_atribut_tree data yang telah diproses tadi akan disimpan ke dalam <i>datastore</i> atribut_tree

3.6.4.9 Proses 3.3 Pembentukan Tree Klas Detail

<i>Input</i>	data_atribut_tree
<i>Output</i>	Hasil_tree_klas_detail
<i>Datastore</i>	
Skenario	Sistem akan membaca data_atribut_tree lalu sistem akan melakukan proses pemberian tree_klas_detail dan menghasilkan <i>Output</i> berupa data yang sudah terbentuk menjadi tree_klas_detail

3.6.4.10 Proses 3.4 Menyimpan Tree Klas Detail

<i>Input</i>	Hasil_tree_klas_detail
<i>Output</i>	data_tree_clas_detail
<i>Datastore</i>	Tree_clas_detail
Skenario	data_tree_klas_detail data yang telah diproses tadi akan disimpan ke dalam <i>datastore</i> tree_clas_detail.

3.6.4.11 Proses 3.5 Pembentukan Tree Parental Detail

<i>Input</i>	data_tree_klas_detail
<i>Output</i>	Hasil_tree_parental_detail

<i>Datastore</i>	
Skenario	Sistem akan membaca <i>data_tree_klas_detail</i> lalu sistem akan melakukan proses pemberian <i>tree_parental_detail</i> dan menghasilkan <i>Output</i> berupa data yang sudah terbentuk menjadi <i>tree_parental_detail</i> .

3.6.4.12 Proses 3.6 Menyimpan Tree Parental Detail

<i>Input</i>	Hasil_tree_parental_detail
<i>Output</i>	data_tree_parental_detail
<i>Datastore</i>	Tree_parental_detail
Skenario	Hasil_tree_klas_detail data yang telah diproses tadi akan disimpan ke dalam <i>datastore</i> tree_klas_detail.

3.6.4.13 Proses 3.7 Perhitungan Entropy

<i>Input</i>	Data_atribut_tree, data_tree_klas_detail_, data_tree_parental_detail.
<i>Output</i>	Data_entropy
<i>Datastore</i>	entropy
Skenario	Sistem akan membaca Data_atribut_tree, data_tree_klas_detail_, data_tree_parental_detail lalu memproses data tersebut untuk menghasilkan <i>Output</i> berupa data_entropy dan disimpan dalam <i>datastore</i> entropy.

3.6.4.14 Proses 3.8 Perhitungan Gain

<i>Input</i>	Hasil_entropy
<i>Output</i>	Data_gain

<i>Datastore</i>	gain
Skenario	Sistem akan membaca hasil_entropy lalu memproses data tersebut untuk menghasilkan <i>Output</i> berupa data_gain dan disimpan dalam <i>datastore</i> gain.

3.6.4.15 Proses 3.9 Perhitungan Gabungan

<i>Input</i>	Data_entropy, data_gain
<i>Output</i>	Hasil_pembentukan_tree
<i>Datastore</i>	hitung
Skenario	Sistem akan membaca data_entropy dan data gain lalu data digabungkan untuk proses perhitungan dan disimpan dalam <i>datastore</i> hitung

3.6.4.16 Proses 4.0 Pembentukan rule

<i>Input</i>	Data_tree, persentase data
<i>Output</i>	Hasil_pembentukan_rule
<i>Datastore</i>	rule
Skenario	Sistem akan membaca <i>input</i> persentase data dan data_tree yang telah diproses. Dan data tersebut diproses untuk menghasilkan <i>Output</i> berupa hasil_pembentukan_tree dan disimpan dalam <i>datastore</i> rule

3.6.4.17 Proses 5.0 Pengujian Data *Input*

<i>Input</i>	Data_uji
<i>Output</i>	Hasil_pengujian
<i>Datastore</i>	
Skenario	Sistem akan menerima <i>input</i> dari <i>User</i> berupa

	data_uji lalu sistem akan memproses dan menghasilkan <i>Output</i> berupa hasil pengujian.
--	--

3.6.4.18 Proses 6.0 Pengujian Master Data

<i>Input</i>	Data_hasil_proses,data_hasil_test
<i>Output</i>	Hasil_pengetesan
<i>Datastore</i>	Test_perbandingan, hasil_test
Skenario	Sistem akan membaca data dari <i>input</i> sistem data_hasil_proses dan data_hasil_test lalu data diproses untuk dilihat akurasi data, data tersebut disimpan di dalam <i>datastore</i> test_perbandingan dan hasil_test dan <i>Output</i> data berupa hasil pengetesan.