



Hak cipta dan penggunaan kembali:

Lisensi ini mengizinkan setiap orang untuk menggubah, memperbaiki, dan membuat ciptaan turunan bukan untuk kepentingan komersial, selama anda mencantumkan nama penulis dan melisensikan ciptaan turunan dengan syarat yang serupa dengan ciptaan asli.

Copyright and reuse:

This license lets you remix, tweak, and build upon work non-commercially, as long as you credit the origin creator and license it on your new creations under the identical terms.

BAB III

METODOLOGI DAN PERANCANGAN APLIKASI

3.1 Metode Penelitian

Metodologi dalam penelitian ini dapat dijabarkan menjadi beberapa langkah yang terdiri dari,

a. Studi Literatur

Metode ini membantu peneliti dalam mencari teori-teori dasar yang diperlukan dalam penelitian, seperti teori tentang *data mining* dan algoritma C4.5. Selain itu, dilakukan juga pengumpulan *data training* dan *data testing* untuk proses prediksi kelulusan mahasiswa.

b. Perancangan dan Pembangunan Aplikasi

Perancangan dan pembangunan aplikasi dilakukan dengan menggunakan bahasa pemrograman C# untuk aplikasi berbasis *desktop* dengan menerima masukan berupa *file excel* untuk *data training* dan *data testing*.

c. Implementasi Aplikasi

Implementasi aplikasi yang dilakukan adalah memprediksi tingkat kelulusan berdasarkan *data* yang diperoleh.

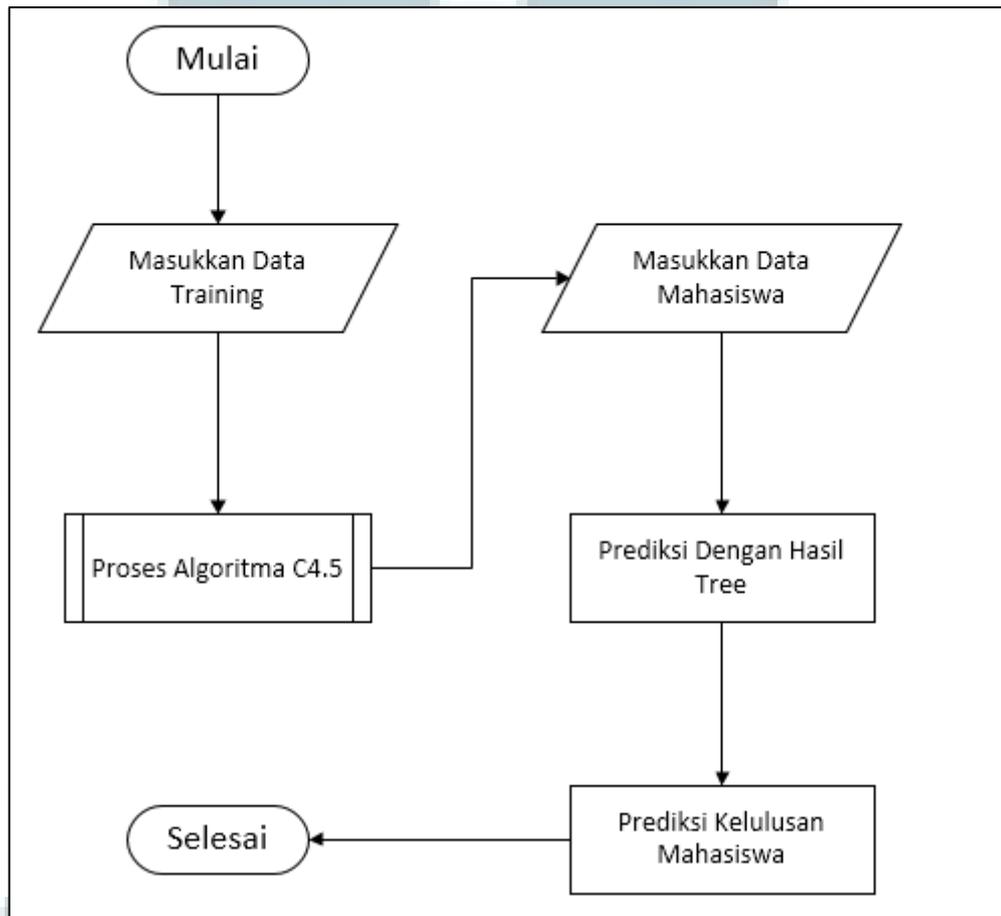
d. Uji Coba dan Evaluasi

Peneliti melakukan uji coba terhadap aplikasi yang telah dibuat disertai dengan hasil evaluasi. Uji coba dilakukan untuk memperlihatkan apakah algoritma C4.5 bisa memprediksi tingkat kelulusan mahasiswa.

3.2 Perancangan

3.2.1 Flowchart

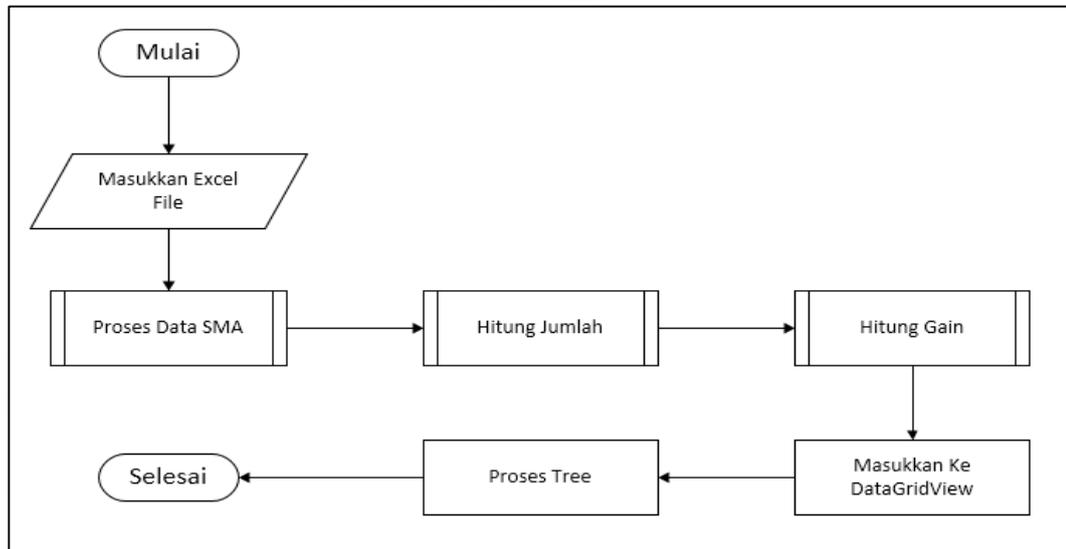
Perancangan aplikasi prediksi kelulusan mahasiswa tersebut dilakukan dengan perancangan *conceptual flowchart* seperti gambar berikut ini.



Gambar 3.1 *Flowchart* aplikasi

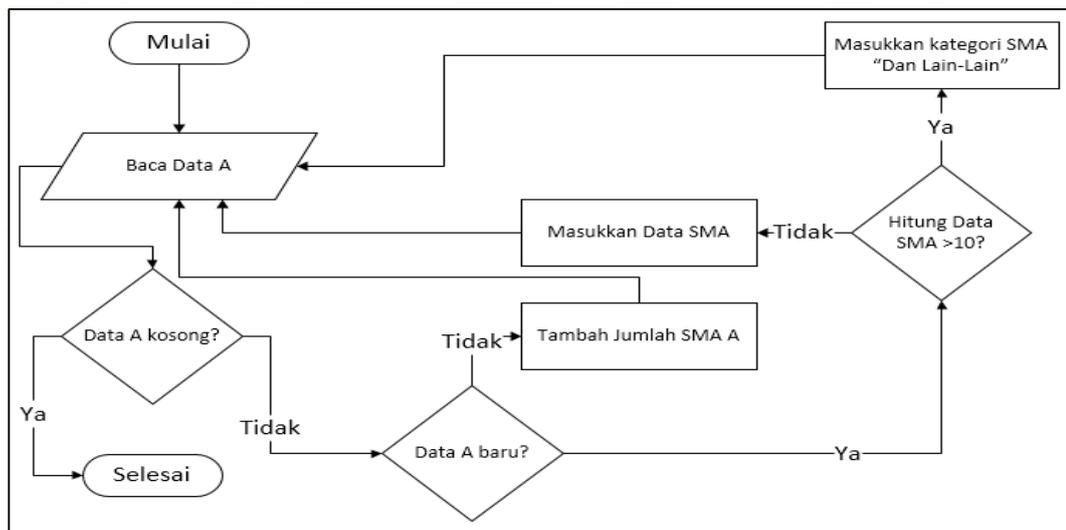
Pada gambar 3.1, proses aplikasi dimulai dengan memasukkan *data training* yang berupa *excel file* dengan format yang telah ditentukan. Setelah *data training* di-*input* maka aplikasi memproses data yang ada dengan algoritma C4.5 untuk menentukan cabang pertama. Subproses dari proses algoritma C4.5 dapat dilihat

pada gambar 3.7. Aplikasi dilanjutkan dengan memasukkan *data testing* dan cabang lainnya terbentuk untuk memprediksi kelulusan mahasiswa.



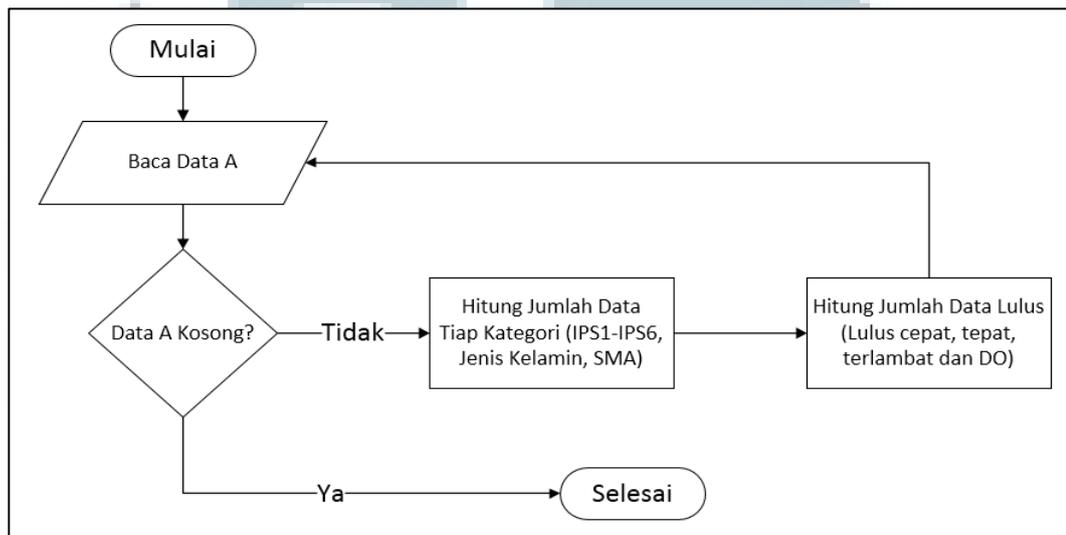
Gambar 3.2 *Flowchart data training*

Gambar 3.2 menggambarkan aliran proses dari *data training* yang diawali dengan memasukkan *excel file*. Setelah *file* tersebut berhasil dimasukkan, maka aplikasi melakukan proses *list data SMA*, algoritma C4.5 dan mendapatkan cabang pertama.



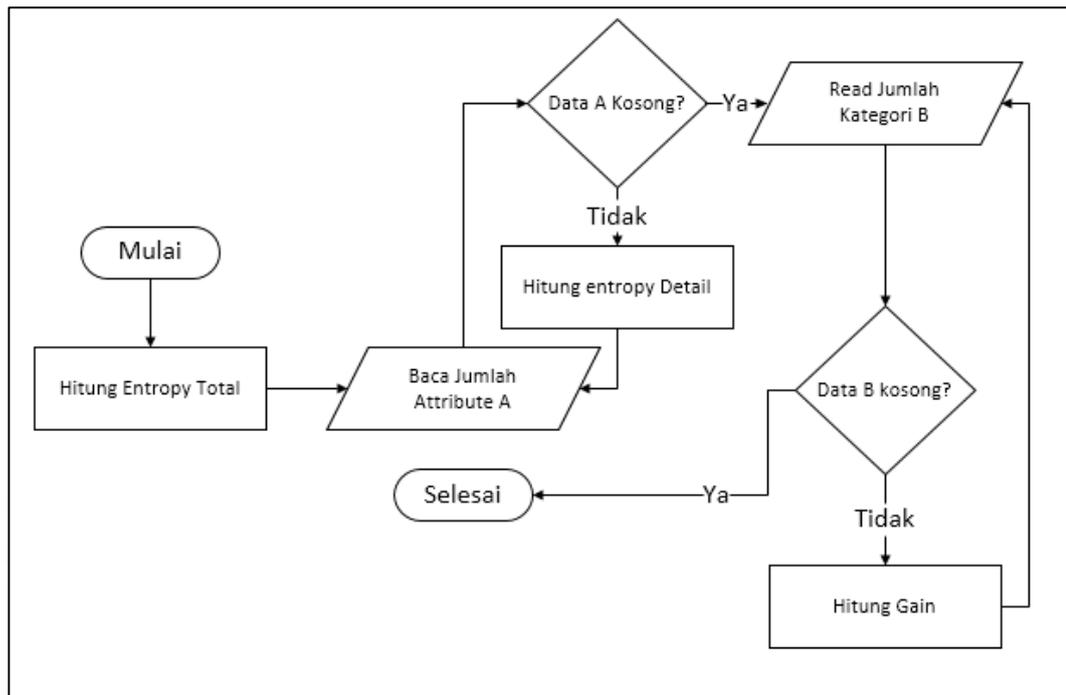
Gambar 3.3 *Flowchart subprocesses proses data SMA*

Proses data SMA dapat dilihat pada gambar 3.3. Alur proses dari *flowchart* tersebut membaca semua *data training* dan mendata SMA yang ada. Jika data SMA lebih dari 10 sekolah, maka dibuat *list data* SMA yang ke Sepuluh dengan nama “Dan Lain-Lain”, sedangkan data SMA dari *index* Satu sampai Sembilan merupakan *data* SMA yang terbanyak.



Gambar 3.4 *Flowchart* subproses perhitungan jumlah data tiap kategori

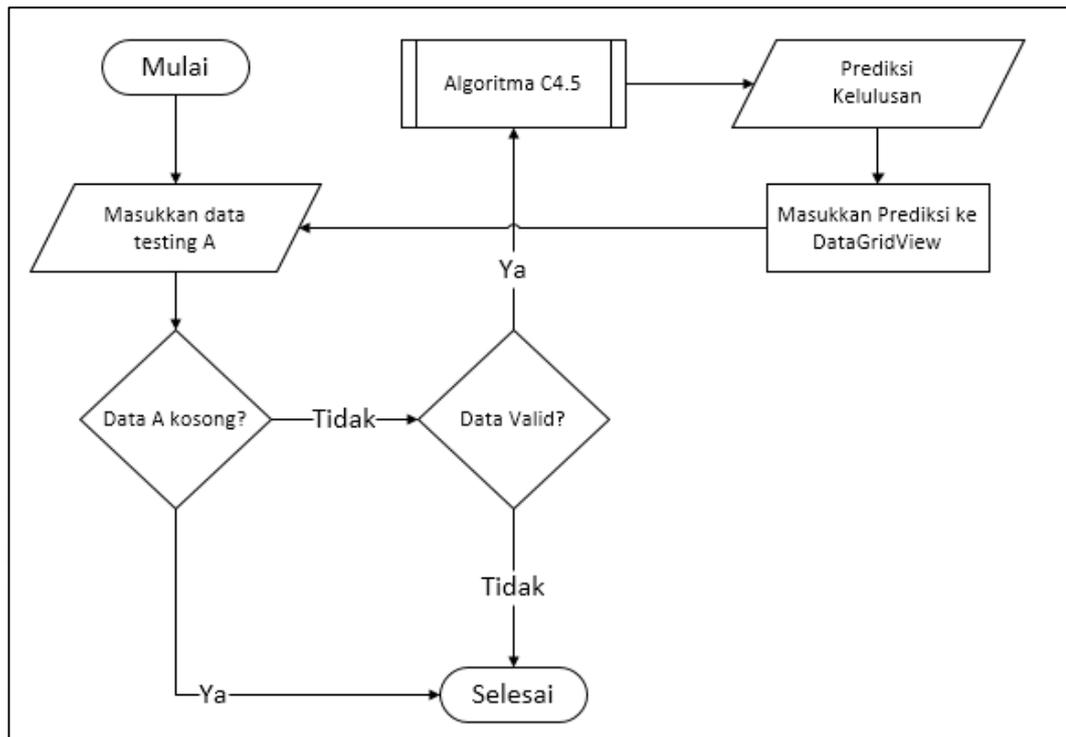
Gambar 3.4 menunjukkan alur dari proses hitung jumlah *data* untuk tiap kategori dengan *attribute*-nya yang berbeda-beda dimana IPS1 sampai IPS6 mempunyai empat *attribute*, jenis kelamin dua *attribute* dan SMA maksimal 10 *attribute*. Setelah itu dihitung jumlah kelulusan *data training* dengan empat kriteria yaitu lulus cepat, lulus tepat, lulus terlambat dan *Drop Out*.



Gambar 3.5 *Flowchart* subproses perhitungan *entropy* dan *gain*

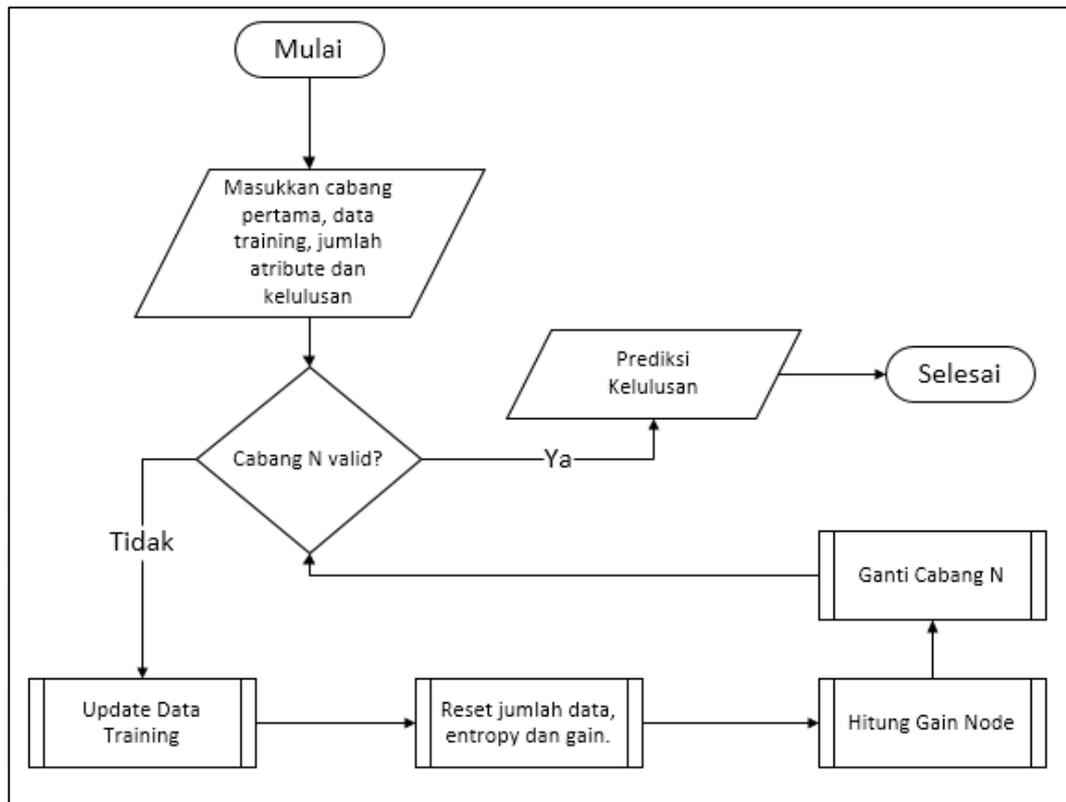
Gambar 3.5 menunjukkan alur proses perhitungan *entropy* total, *entropy* tiap *attribute* dan *gain* dari tiap kategori. Hasil *entropy* total didapat dari jumlah kelulusan empat *attribute* kelulusan yang dihitung dengan rumus pada gambar 2.2. Untuk perhitungan *entropy* tiap *attribute* dibutuhkan untuk menentukan nilai *gain*.

UMMN



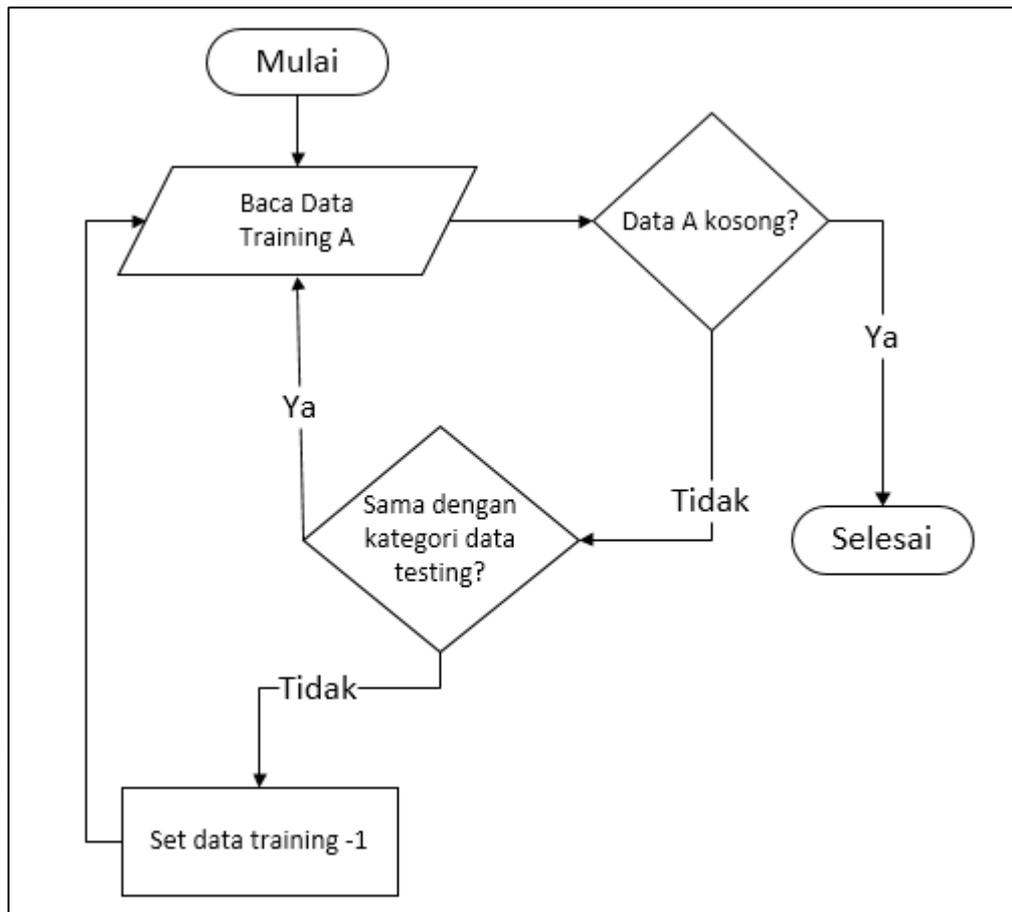
Gambar 3.6 *Flowchart data testing*

Gambar 3.6 menunjukkan alur dari proses *data testing*. Proses diawali dengan memasukkan *data testing* berupa *excel file* dengan format yang telah ditentukan. Jika format *file* tidak sesuai dan terjadi masalah maka proses berakhir, tetapi jika *file* benar maka diproses untuk mendapatkan hasil prediksi kelulusan dengan algoritma C4.5. Hasil prediksi data tersebut dimasukkan dalam *DataGridView* yang telah disediakan.



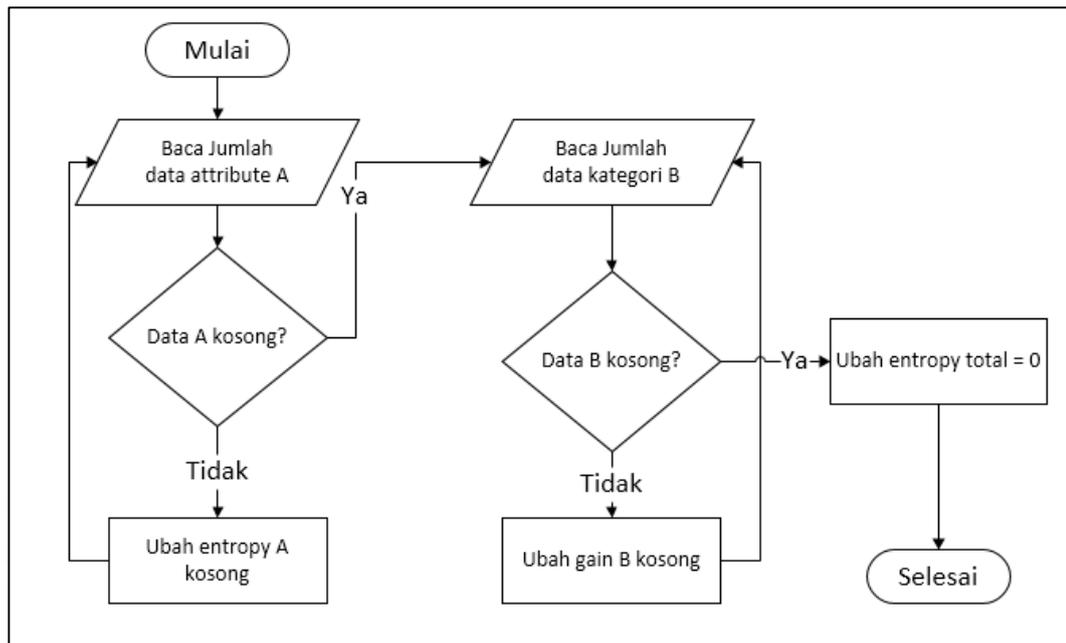
Gambar 3.7 Flowchart subproses perhitungan algoritma C4.5

Gambar 3.7 menjelaskan alur proses algoritma C4.5. Pada bagian awal aplikasi menerima masukkan cabang pertama, *data training*, jumlah tiap kategori, *attribute* dan kelulusan dari proses *data training* sebelumnya. Pada proses cek keputusan dilihat apakah *data testing* pada *attribute* yang ke N bisa diambil keputusan atau tidak. Jika tidak maka *data training* diubah dan aplikasi melakukan perhitungan kembali terhadap jumlah data tiap *attribute*, *entropy total*, *entropy* tiap *attribute* dan *gain*. Setelah itu cabang yang baru (*gain* yang tertinggi) dicek keputusannya berdasarkan *data testing* yang ada.



Gambar 3.8 Flowchart subproses update data training

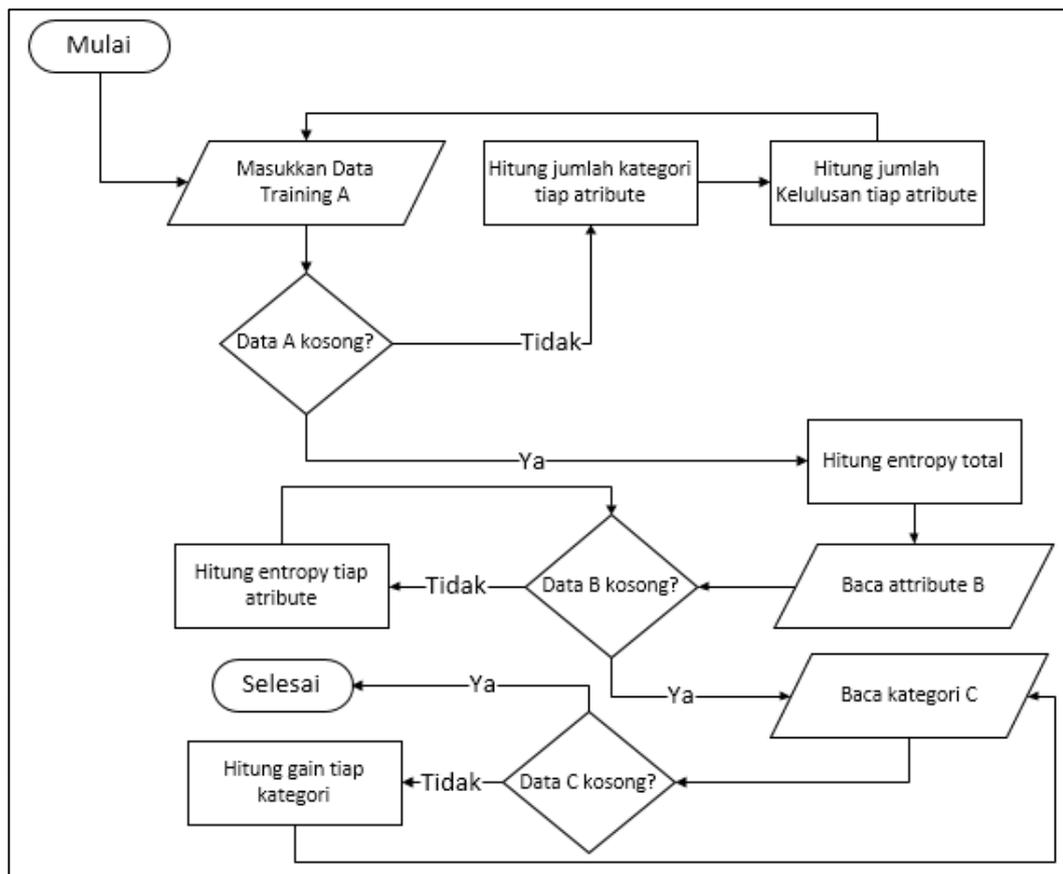
Alur poses perubahan data dapat dilihat pada gambar 3.8. Setiap *data training* yang tidak sesuai dengan kategori yang didapat di *data testing* diubah *data-*nya dengan nilai “-1”. Proses ini berguna untuk tidak menjumlahkan *data training* yang tidak sesuai dengan kategori dari *data testing*.



Gambar 3.9 Flowchart subproses *reset entropy, gain* dan jumlah data

Gambar 3.9 menunjukkan alur proses perubahan *data* pada tiap jumlah kategori dan *attribute*, *entropy total*, *entropy* tiap *attribute* dan *gain* tiap kategori menjadi *null*.

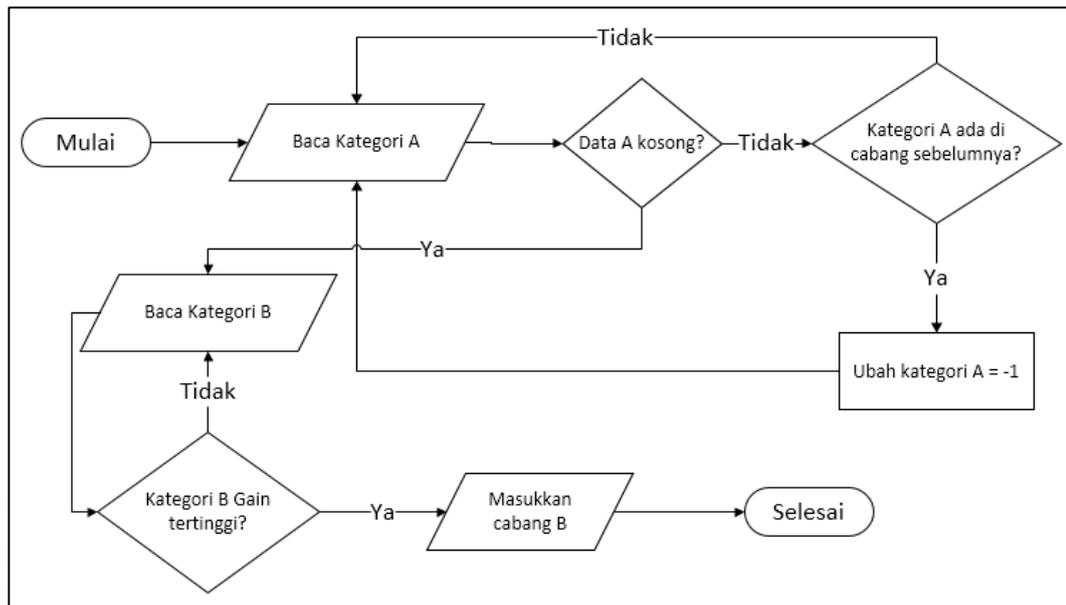
UMMN



Gambar 3.10 *Flowchart* subproses perhitungan nilai *gain*

Gambar 3.10 menunjukkan alur proses perhitungan nilai *gain* tiap kategori.

Perhitungan dimulai dengan menghitung jumlah data dan menghitung jumlah kelulusan untuk tiap *attribute* di masing-masing kategori. Proses dilanjutkan dengan menghitung *entropy* total, *entropy* tiap *attribute*, dan nilai *gain* untuk tiap kategori.

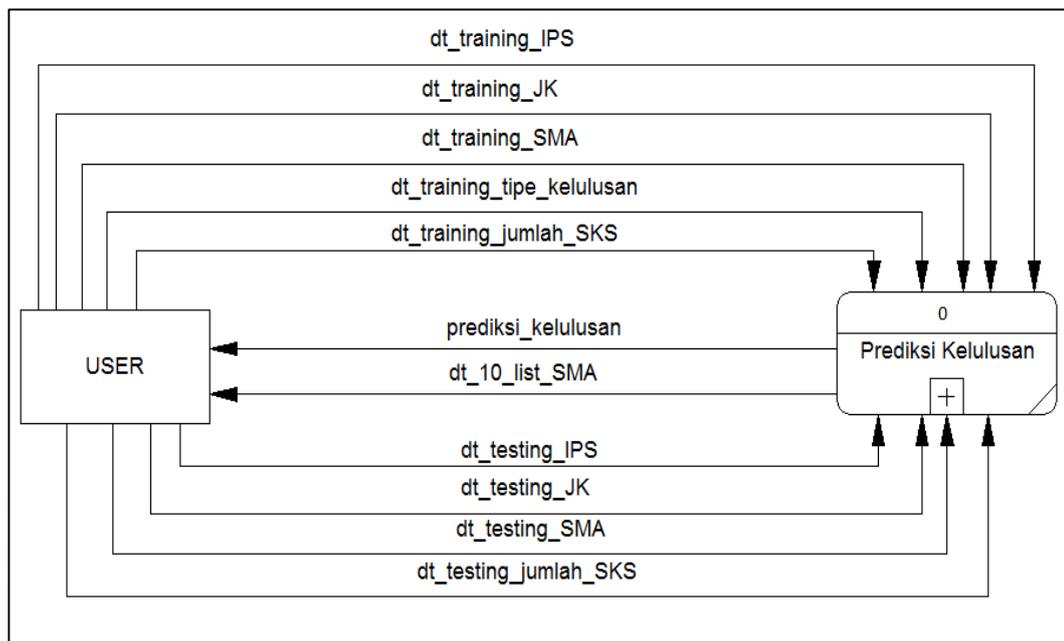


Gambar 3.11 *Flowchart* subproses pencarian cabang baru

Gambar 3.11 menunjukkan alur proses untuk mendapatkan cabang baru. Proses diawali dengan mengubah nilai *gain* menjadi “-1” untuk setiap cabang yang pernah ada di cabang sebelumnya. Dengan demikian proses ini dapat menemukan cabang baru yang didapat dari nilai *gain* yang tertinggi.

3.2.2 Dafa Flow Diagram

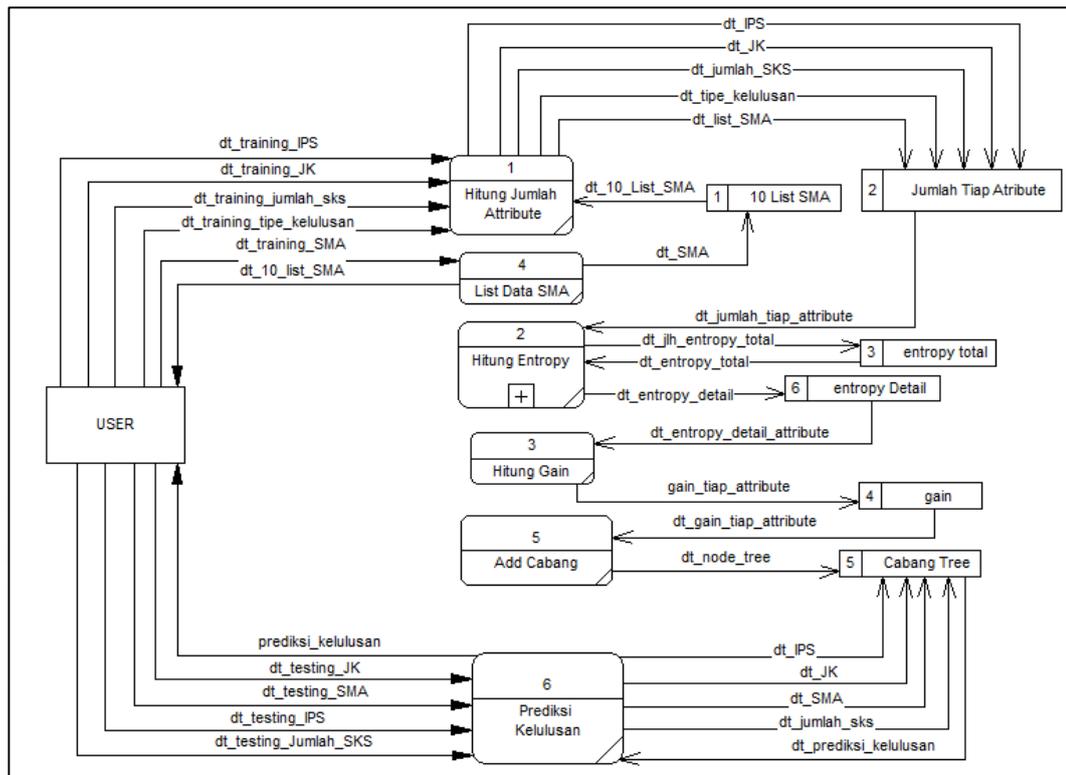
Proses aliran data pada aplikasi ini dapat diperjelas dengan *data flow diagram* pada *context diagram* di gambar 3.12 dan *data flow diagram level satu* di gambar 3.13.



Gambar 3.12 *Context Diagram*

Dari gambar 3.12 *user* memasukkan *data training* yang terdiri dari IPS (data IPS semester satu sampai dengan semester enam), JK (jenis kelamin), SMA, tipe kelulusan dan jumlah SKS pada semester enam. Selain itu, *user* juga memasukkan *data testing* yang terdiri dari IPS (data IPS semester satu sampai dengan semester enam), JK (jenis kelamin), SMA dan jumlah SKS pada semester enam.

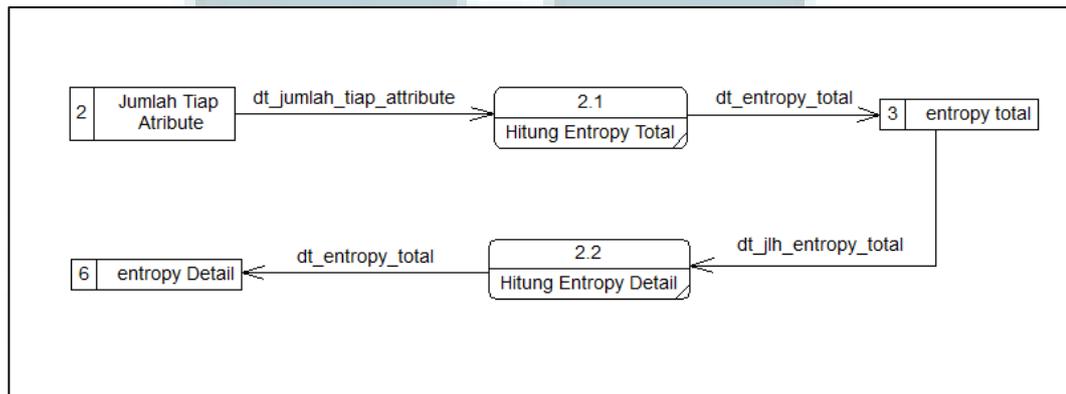
U
M
M
N



Gambar 3.13 DFD Level 1 aplikasi prediksi kelulusan

Pada *data flow diagram level satu* terjadi proses aliran data dengan enam proses yang terdiri dari hitung jumlah *attribute*, *list data SMA*, hitung *entropy*, hitung *gain*, *add cabang* dan prediksi kelulusan. Proses hitung jumlah *attribute* adalah proses yang akan menghitung jumlah tiap *attribute* berdasarkan data yang masuk dan hasil perjumlahan akan masuk dalam data jumlah tiap *attribute*. Proses *list data SMA* adalah proses untuk membuat sekumpulan data SMA menjadi sepuluh data SMA di mana SMA yang ke sepuluh merupakan SMA lain-lain. Proses hitung *entropy* berfungsi untuk menghitung nilai *entropy* total dan *entropy* masing-masing *attribute*. Proses hitung *gain* berfungsi untuk menghitung nilai *gain* untuk masing-masing *attribute*. Proses prediksi kelulusan adalah proses yang berfungsi memberikan keputusan kepada setiap *data testing* yang masuk.

Untuk DFD *level 2* dari proses hitung *entropy* dapat dilihat pada gambar 3.14. Pada proses ini, akan dihitung terlebih dahulu nilai *entropy total* berdasarkan jumlah tiap *attribute*. Dari hasil nilai *entropy total* akan dihitung nilai *entropy* dari masing-masing *attribute*.

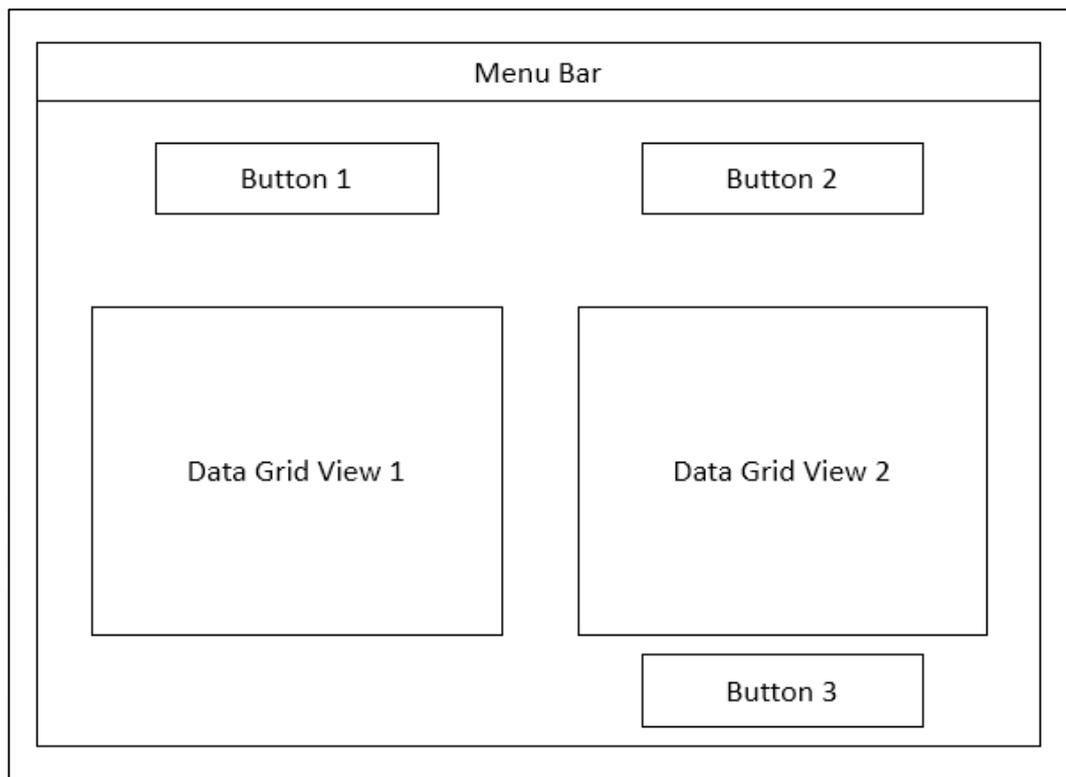


Gambar 3.14 DFD *level 2* proses hitung *entropy*

3.2.3 Sketsa Layar

Pada tampilan awal aplikasi dapat digambarkan seperti gambar berikut ini.

U
M
M
N



Gambar 3.15 Sketsa tampilan awal

Pada sketsa layar tersebut *menu bar* terdiri dari *File*, *Input*, *Help* dan *About Us*. Untuk *button 1* berfungsi untuk membuka *file excel* yang berisi *data training*, sedangkan untuk *button 2* berfungsi untuk memproses *data training*. *Data grid view 1* berfungsi untuk menampilkan *data training* yang di-*input*, sedangkan *data grid view 2* menampilkan hasil proses dari *data training*. *Button 3* digunakan untuk keluar dari aplikasi.

Untuk tampilan pada *menu bar input*, sketsa tampilan memiliki gambaran sebagai berikut.

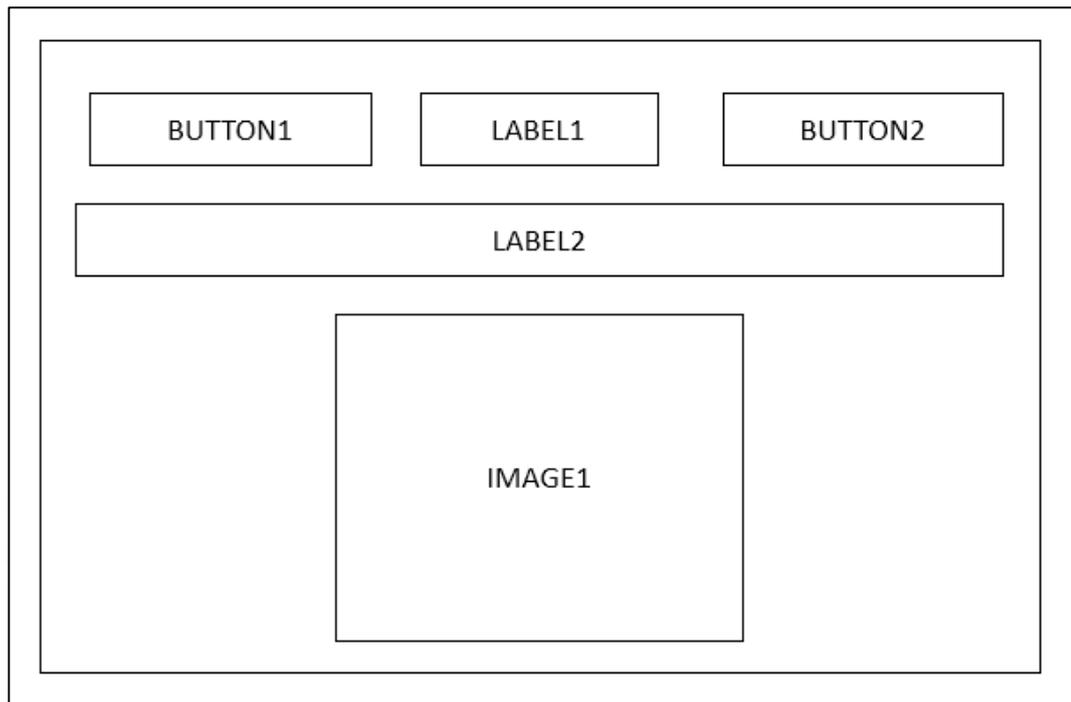
The sketch shows a rectangular form with the following layout:

- Row 1: TEXTBOX1, TEXTBOX2
- Row 2: TEXTBOX3, TEXTBOX4, TEXTBOX5
- Row 3: TEXTBOX6, TEXTBOX7, TEXTBOX8
- Row 4: CBX1 (with a downward arrow), TEXTBOX9, CBX2 (with a downward arrow)
- Row 5: LABEL1, BUTTON1
- Row 6: BUTTON2, BUTTON3, BUTTON4

Gambar 3.16 Sketsa tampilan *form input data*

Pada *textbox1* untuk *input* nama, *textbox2* untuk *input* nim, *textbox3* sampai dengan *textbox8* untuk *input* IPS1 sampai dengan IPS6, *textbox9* untuk jumlah SKS, *cbx1* untuk SMA dan *cbx2* untuk jenis kelamin. Untuk *button1* berfungsi memproses data yang sudah dimasukkan, *button2* berfungsi untuk membuka *file excel data* yang mau dimasukkan, *button3* untuk membuka *form help* dan *button4* untuk keluar dari *form input data*.

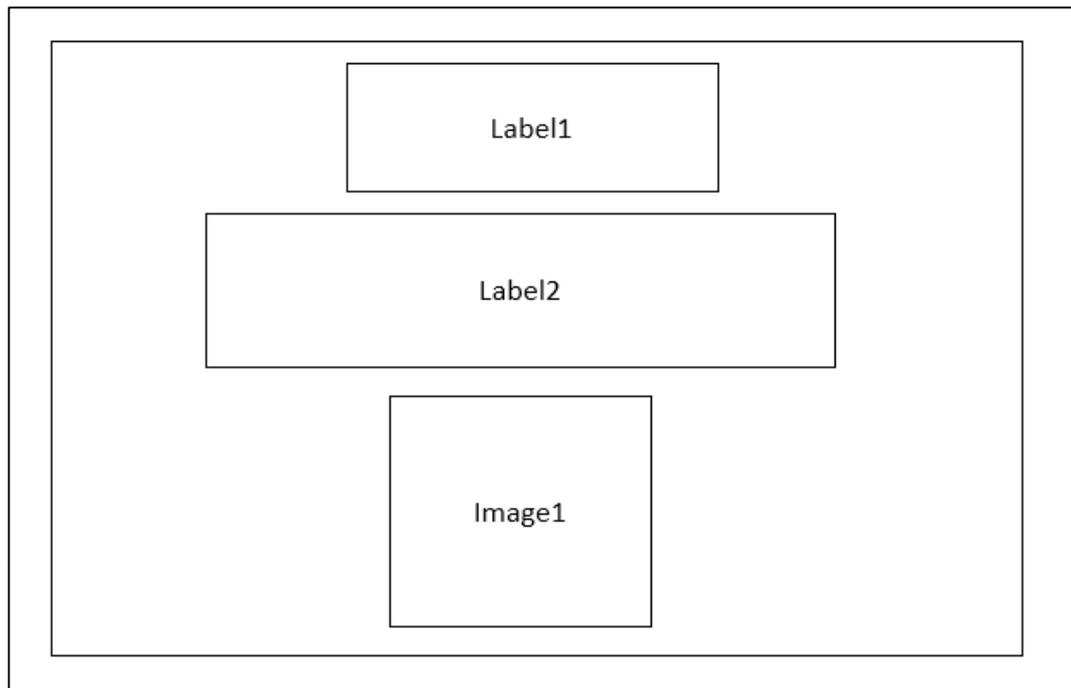
Untuk tampilan *help* memiliki sketsa tampilan sebagai berikut.



Gambar 3.17 Sketsa tampilan *form Help*

Pada gambar tersebut, *label1* merupakan *title* dari tampilan aplikasi tersebut. *Button1* dan *button2* merupakan *button next* dan *previous* yang berfungsi untuk melanjutkan atau kembali ke langkah-langkah yang lain. Untuk *label2* berisi tentang langkah-langkah yang diperlukan untuk menjalankan aplikasi, sedangkan *Image1* merupakan contoh tampilan dari langkah-langkah tersebut.

Sketsa tampilan pada *about us* dapat dilihat pada gambar berikut ini.

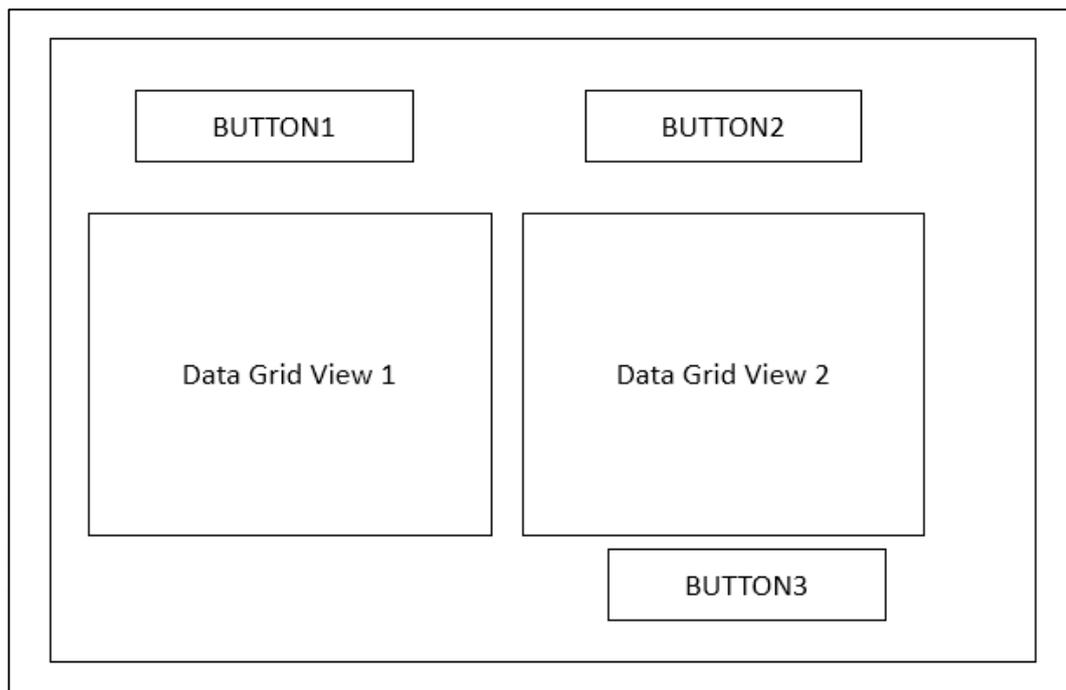


Gambar 3.18 Sketsa tampilan *form about us*

Pada *label1* menampilkan *title* dari tampilan aplikasi, sedangkan *label2* menampilkan informasi tentang data peneliti. Untuk *image1* menampilkan foto dari peneliti.

Tampilan untuk *input data testing* dengan menggunakan *excel file* dapat dilihat pada gambar 3.19.

UMMN



Gambar 3.19 Sketsa tampilan *form input excel*

Pada *button1* berfungsi untuk membuka *file excel data testing*, sedangkan *button2* untuk memproses hasil prediksi kelulusan *data testing* yang dimasukkan. Tampilan *data testing* yang dimasukkan dapat dilihat pada *Data Grid View 1*. Untuk tampilan hasil prediksi *data testing* yang telah diproses dapat dilihat pada *Data Grid View 2*. *Button 3* berfungsi untuk keluar dari *form* ini.

3.3 Analisis Model

Untuk memperjelas algoritma C4.5 yang digunakan penulis untuk memprediksi kelulusan mahasiswa, berikut adalah proses perhitungan secara manual algoritma C4.5 dengan menggunakan lima belas data alumni Universitas Multimedia Nusantara program studi Teknik Informatika angkatan 2008 dan alumni angkatan 2009. Lima belas data alumni tersebut merupakan data alumni

yang terdiri empat kategori data yaitu lulus cepat, lulus tepat, lulus terlambat dan *drop out*. Sebaran datanya dapat dilihat pada tabel.

Tabel 3.1 Tabel cabang pertama

No	Nama	NIM	IPS1	IPS2	IPS3	IPS4	IPS5	IPS6	JK	SMA	Lulus	SKS
1	###	###	3.57	3.14	3.70	3.40	3.14	3.71	L	SMA FRANSISKUS	2	124
2	###	###	3.71	2.95	3.70	3.75	3.50	3.38	L	SMA GONZAGA	2	127
3	###	###	2.67	2.24	3.12	3.83	3	3.38	L	SMA GONZAGA	2	119
4	###	###	3.38	3.52	3.85	3.65	3.43	3.57	L	SMA HARAPAN BANGSA	1	124
5	###	###	2.95	2.96	3.40	3.65	3.86	3.64	L	SMA KRISTEN YSKI	2	125
6	###	###	3.90	4	3.60	3.87	4	3.71	P	SMA MARDI YUANA	1	127
7	###	###	1.52	1	1.08	0	0	0	L	SMA MARKUS	4	50
8	###	###	2.62	2.33	2.72	3.40	3.14	3.38	P	SMA MATERDEI PAMULANG	3	118
9	###	###	3.10	2.76	3.10	3.10	3.43	3.55	L	SMA KRISTEN YSKI	2	125
10	###	###	3.86	3.71	3.85	3.80	4	3.57	L	SMA STRADA ST. THOMAS AQUINO	1	124
11	###	###	2.14	2.04	2	2.83	3.17	2.56	L	SMA SULUH	3	112
12	###	###	4	4	3.70	4	3.88	3.86	L	SMA XAVERIUS	1	127
13	###	###	3.43	2.76	3.30	3.50	3.57	3.68	L	SMAK 3 BPK PENABUR	2	125
14	###	###	3.57	3.86	3.55	4	4	3.90	L	SMAN 48	1	127
15	###	###	2.86	2.90	3.30	3.35	3.43	3.21	L	SMAN 6	2	125

Dari lima belas data tersebut, akan dilakukan proses perhitungan *entropy* dan *gain* untuk menentukan cabang dari pohon keputusan yang akan terbentuk sebagai aturan dalam memprediksi kelulusan mahasiswa. Sebelum menghitung nilai *gain* dengan Rumus 2.1, dihitung terlebih dahulu nilai *entropy* total dan *entropy detail attribute*. Nilai *entropy* total diperoleh sebagai berikut, seperti pada

Rumus II:

$$\begin{aligned}
 E &= - (5 / 15) \log_2 (5 / 15) - (7 / 15) \log_2 (7 / 15) - (2 / 15) \log_2 (2 / 15) - (1 / \\
 &15) \log_2 (1/15) \\
 &= 1.689482
 \end{aligned}$$

Untuk menghitung *entropy* setiap *attribute* menggunakan Rumus 2.2, sedangkan untuk menghitung *gain* menggunakan Rumus 2.1, sehingga mendapatkan hasil seperti tabel 3.2.

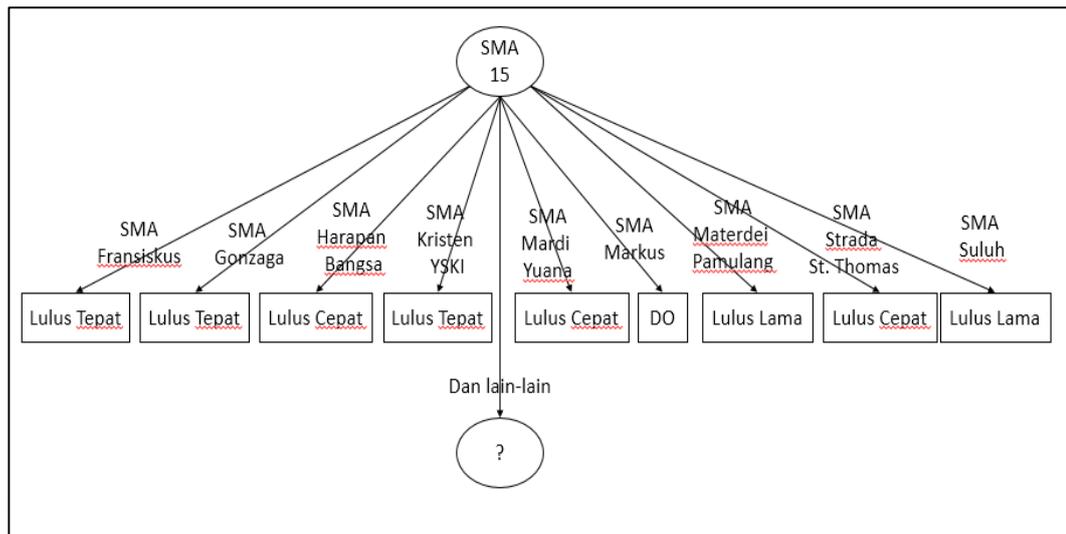
Tabel 3.2 Hasil perhitungan *entropy* dan *gain* (Bagian I)

Attribute	Kategori	Jumlah	Lulus			Drop Out	Entropy	Gain
			Cepat	Tepat	Lama			
		15	5	7	2	1	1.689482	
IPS 1	Nilai 3 - 4	9	5	4	0	0	0.991076	0.771186
	Nilai 2 - 3	5	0	3	2	0	0.970951	0.771186
	Nilai 1 - 2	1	0	0	0	1	0	
	Nilai 0 - 1	0	0	0	0	0	0	
IPS 2	Nilai 3 - 4	6	5	1	0	0	0.650022	0.996792
	Nilai 2 - 3	8	0	6	2	0	0.811278	0.996792
	Nilai 1 - 2	1	0	0	0	1	0	
	Nilai 0 - 1	0	0	0	0	0	0	
IPS 3	Nilai 3 - 4	12	5	7	0	0	0.979869	0.905587
	Nilai 2 - 3	2	0	0	2	0	0	
	Nilai 1 - 2	1	0	0	0	1	0	
	Nilai 0 - 1	0	0	0	0	0	0	
IPS 4	Nilai 3 - 4	13	5	7	1	0	1.295738	0.56651
	Nilai 2 - 3	1	0	0	1	0	0	
	Nilai 1 - 2	0	0	0	0	0	0	
	Nilai 0 - 1	1	0	0	0	1	0	
IPS 5	Nilai 3 - 4	14	5	7	2	0	1.43156	0.353359
	Nilai 2 - 3	0	0	0	0	0	0	
	Nilai 1 - 2	0	0	0	0	0	0	
	Nilai 0 - 1	1	0	0	0	1	0	

Tabel 3.3 Hasil perhitungan *entropy* dan *gain* (Bagian II)

Attribute	Kategori	Jumlah	Lulus			Drop Out	Entropy	Gain
			Cepat	Tepat	Lama			
IPS 6	Nilai 3 - 4	13	5	7	1	0	1.295738	0.56651
	Nilai 2 - 3	1	0	0	1	0	0	
	Nilai 1 - 2	0	0	0	0	0	0	
	Nilai 0 - 1	1	0	0	0	1	0	
JK	L	13	4	7	1	1	1.573402	0.192533
	P	2	1	0	1	0	1	
SMA	SMA FRANSISKUS	1	0	1	0	0	0	1.422816
	SMA GONZAGA	2	0	2	0	0	0	
	SMA HARAPAN BANGSA	1	1	0	0	0	0	
	SMA KRISTEN YSKI	2	0	2	0	0	0	
	SMA MARDI YUANA	1	1	0	0	0	0	
	SMA MARKUS	1	0	0	0	1	0	
	SMA MATERDEI PAMULANG	1	0	0	1	0	0	
	SMA STRADA ST. THOMAS	1	1	0	0	0	0	
	SMA SULUH	1	0	0	1	0	0	
	Dan Lain-lain	4	2	2	0	0	1	

Dari proses perhitungan tersebut, *gain* tertinggi terletak pada *attribute* SMA sehingga didapatkan cabang pertama seperti pohon keputusan pada gambar 3.20.



Gambar 3.20 Pohon keputusan cabang pertama

Karena masih ada satu node yang belum dapat keputusan maka akan dilakukan perhitungan *entropy* dan *gain* untuk mencari cabang berikutnya. Terdapat empat data pada kategori “dan lain-lain” dan dari data tersebut akan dicari nilai *entropy* dan *gain*-nya untuk mendapatkan cabang baru pada pohon keputusan. Pada tabel 3.4 ditampilkan empat data yang mempunyai kategori SMA “dan lain-lain”.

Tabel 3.4 Data pada cabang kedua

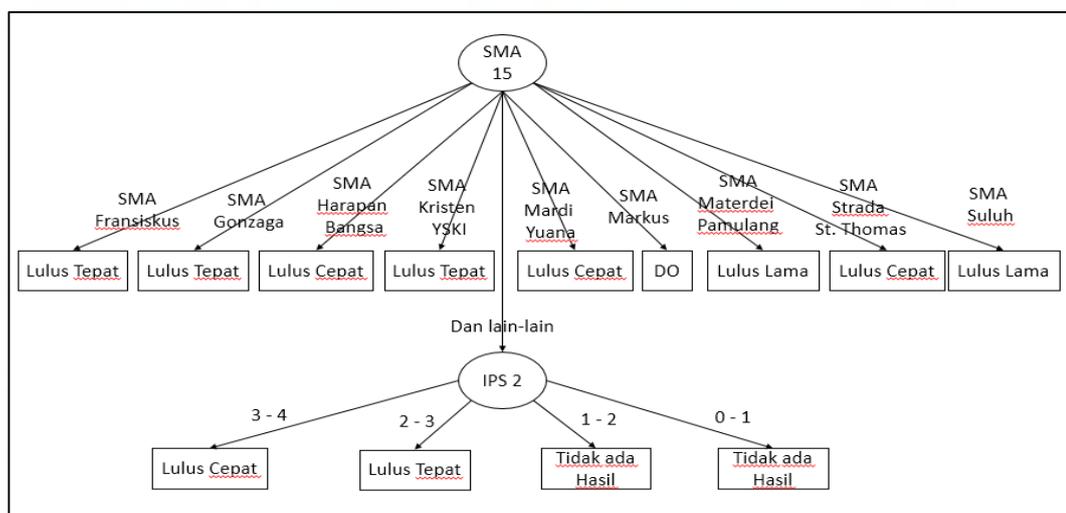
No	Nama	NIM	IPS1	IPS2	IPS3	IPS4	IPS5	IPS6	Jenis Kelamin	SMA	Lulus	Jumlah SKS
1	###	###	4	4	3.70	4	3.88	3.86	L	SMA XAVERIUS	1	127
2	###	###	3.43	2.76	3.30	3.50	3.57	3.68	L	SMK 3 BPK PENABUR	2	125
3	###	###	3.57	3.86	3.55	4	4	3.90	L	SMAN 48	1	127
4	###	###	2.86	2.90	3.30	3.35	3.43	3.21	L	SMAN 6	2	125

Berdasarkan data tersebut, didapatkan hasil perhitungan *entropy* dan *gain* seperti pada tabel 3.5 dimana IPS semester dua mempunyai nilai *gain* tertinggi dari *attribute* lainnya. Proses perhitungan pencarian cabang berakhir pada tahap ini, karena setiap cabang sudah mendapatkan hasil keputusannya masing-masing. Bentuk pohon keputusan akhirnya terdapat pada gambar 3.21.

Tabel 3.5 Hasil perhitungan cabang kedua

Attribute	Kategori	Jumlah	Lulus			Drop Out	Entropy	Gain
			Cepat	Tepat	Lama			
		4	2	2	0	0	1	
IPS 1	Nilai 3 - 4	3	2	1	0	0	0.9183	0.31128
	Nilai 2 - 3	1	0	1	0	0	0	
	Nilai 1 - 2	0	0	0	0	0	0	
	Nilai 0 - 1	0	0	0	0	0	0	
IPS 2	Nilai 3 - 4	2	2	0	0	0	0	1
	Nilai 2 - 3	2	0	2	0	0	0	
	Nilai 1 - 2	0	0	0	0	0	0	
	Nilai 0 - 1	0	0	0	0	0	0	
IPS 3	Nilai 3 - 4	4	2	2	0	0	1	0
	Nilai 2 - 3	0	0	0	0	0	0	
	Nilai 1 - 2	0	0	0	0	0	0	
	Nilai 0 - 1	0	0	0	0	0	0	
IPS 4	Nilai 3 - 4	4	2	2	0	0	1	0
	Nilai 2 - 3	0	0	0	0	0	0	
	Nilai 1 - 2	0	0	0	0	0	0	
	Nilai 0 - 1	0	0	0	0	0	0	
IPS 5	Nilai 3 - 4	4	2	2	0	0	1	0
	Nilai 2 - 3	0	0	0	0	0	0	
	Nilai 1 - 2	0	0	0	0	0	0	
	Nilai 0 - 1	0	0	0	0	0	0	
IPS 6	Nilai 3 - 4	4	2	2	0	0	1	0
	Nilai 2 - 3	0	0	0	0	0	0	
	Nilai 1 - 2	0	0	0	0	0	0	
	Nilai 0 - 1	0	0	0	0	0	0	
Jenis Kelamin	L	4	2	2	0	0	1	0
	P	0	0	0	0	0	0	

Pada hasil perhitungan di tabel 3.5, *attribute* yang menjadi cabang berikutnya adalah IPS 2 sehingga pohon keputusan yang terbentuk akan seperti gambar 3.



Gambar 3.21 Pohon keputusan cabang kedua