BAB IV
HASIL PENELITIAN DAN PEMBAHASAN

Penelitian ini dilakukan dalam beberapa tahap yaitu, pembentukan dataset *training* dan dataset *testing,* pembentukan dataset yang mengandung *missing value,* Imputasi *missing value* menggunakan Amelia II, pereduksian dan pengkategorian atribut menggunakan CART dan pemodelan menggunakan Regresi Logistik.

4.1 Pembentukan Data *Training* dan Data *Testing*

Tahap awal dari penelitian ini adalah pembentukan data *training* dan data *testing.* Data *training* digunakan untuk pembentukan model, sedangkan data *testing* digunakan untuk validasi model yang telah terbentuk. Data lengkap dibagi secara acak menjadi dua bagian yaitu sebanyak 90% sebagai data *training* dan 10% sebagai data *testing.* Hasil penarikan secara untuk 90% data *training* didapatkan 628 debitur dengan kategori baik (69.78%) dan 272 debitur dengan kategori buruk (30.22%). Hasil penarikan secara acak untuk 10% data *testing* didapatkan 72 debitur dengan kategori baik (72%) dan 28 debitur dengan kategori buruk (28.00%). Hasil dari pembagian data *training* dan data *testing* dapat dilihat pada tabel IV-1.

Tabel IV-1 Hasil Pembagian Data *Training* dan Data *Testing*

	Kategori Peubah Respon				Total
	Baik		Buruk		
	Jumlah Data	Persentase	Jumlah Data	Persentase	
Data *Training*	628	69.78%	272	30.22%	900
Data *Testing*	72	72.00%	28	28.00%	100
Total	700		300		1000

4.2 Pembentukan Dataset *Missing Value*

Tahap selanjutnya adalah pembentukan dataset yang mengandung *missing value.* Pembangkitkan *missing value* dilakukan dari data *training.* Nilai observasi dari setiap atribut pada data training dihilangkan sebanyak 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80% dan 90%. Proses pembangkitan *missing value* dilakukan secara acak. Pada tahap ini akan terbentuk 10 dataset yang akan digunakan untuk pembentukan model yaitu dataset lengkap dan dataset dengan *missing value* 10% - 90%. Peta sebaran *missing value* di setiap atribut dapat dilihat pada lampiran 1.

4.3 Imputasi *Missing Value* Menggunakan Amelia II

Imputasi *missing value* sangat penting dalam hal analisis data. Tidak tertanganinya *missing value* dengan baik dapat menyebabkan penarikan kesimpulan yang salah di akhir analisis. Beberapa metode analisis data tidak dapat menerima *missing value,* sehingga data yang memiliki *missing value* akan dibuang dan tidak diikutkan dalam analisis. Hal ini akan mengurangi tingkat akurasi dari hasil analisis kerena terdapat beberapa informasi yang dibuang. Selain itu dibuangnya *missing value* akan sangat merugikan terutama untuk penelitian yang untuk pengumpulan datanya dibutuhkan biaya yang mahal.

Perangkat lunak yang digunakan untuk imputasi *missing value* dalam penelitian ini adalah Amelia II. Imputasi dilakukan untuk dataset yang sebelumnya sudah dihilangkan datanya secara acak sebanyak 10% - 90%. Pada dataset 10% sampai dengan dataset 70% semua *missing value* berhasil diimputasi.

Pada dataset 80% terdapat 10 observasi yang semua data atributnya hilang. Dan pada dataset 90% terdapat 103 observasi yang semua data atributnya hilang. Amelia II tidak dapat melakukan imputasi jika semua data atribut hilang. Data yang semua atributnya hilang ini akan dibuang dan tidak dihitung dalam analisis berikutnya. Sehingga untuk dataset 80% ada 10 observasi yang dibuang dan untuk dataset 90% ada 103 observasi yang dibuang.

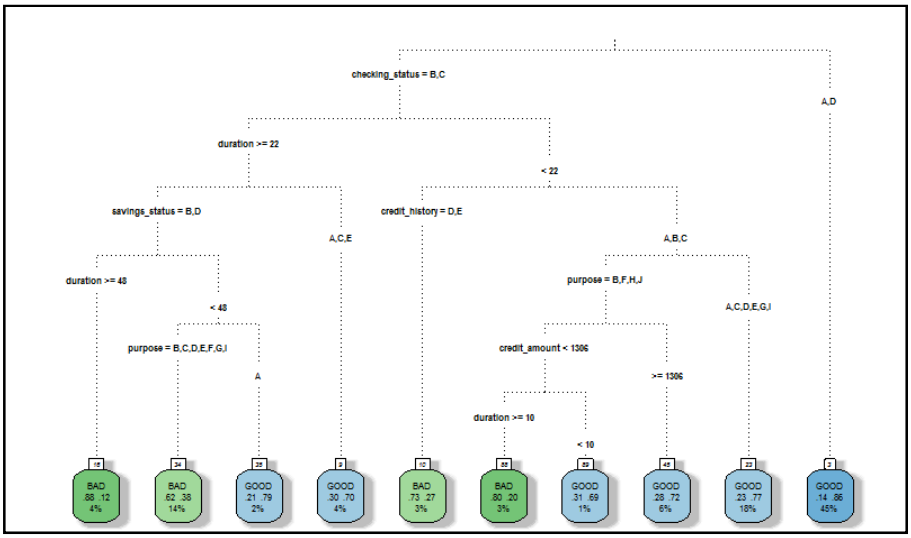
4.4 Pereduksian dan Pengkategorian Atribut Menggunakan CART

Langkah selanjutnya adalah pereduksian dan pengkategorian atribut menggunakan CART. Pada tahun 1984, Jerome Friedman, Richard Olsen dan Charles Stone mengembangkan metode yang dinamakan *Classification and Regression Tree* (CART). CART dapat digunakan untuk data numerik dan kategorik. Index Gini digunakan untuk menentukan atribut yang berpengaruh nyata terhadap peubah respon. Dengan Index Gini juga atribut – atribut tersebut akan dikategorikan ulang, sehingga model yang akan dihasilkan memiliki akurasi yang tinggi.

Penambahan atribut tidak selalu dapat memperkecil error. Jika penambahan atribut menyebabkan peningkatan error maka pembuatan pohon akan dihentikan. Sehingga akan terbentuk pohon optimal, yaitu pohon dengan jumlah simpul yang sedikit namun memiliki nilai error terkecil.

4.4.1 Dataset Lengkap

Pohon klasifikasi dari dataset lengkap dapat dilihat pada gambar IV-1.



Gambar IV-1 Pohon Klasifikasi Dataset Lengkap

Gambar IV-1 adalah pohon klasifikasi untuk dataset lengkap. Atribut yang berpengaruh terhadap peubah respon adalah checking_status, credit_amount, credit_history, duration, purpose, dan savings_status. Atribut – atribut tersebut kemudian dikategorikan ulang menjadi 2 kategori. Pengkategorian ulang dilakukan berdasarkan pohon klasifikasi yang terbentuk dengan menggunakan CART.

Berdasarkan pohon klasifikasi tersebut atribut checking_status yang sebelumnya terdiri dari 4 kategori (A, B, C, dan D dikategorikan ulang menjadi 2 kategori yaitu BC dan AD. Atribut credit_amount yang sebelumnya adalah atribut numerik kemudian dikategorikan ulang menjadi 2 kategori yaitu <1306 dan >= 1306. Atribut duration yang sebelumnya adalah atribut numerik kemudian dikategorikan ulang menjadi 2 kategori yaitu <22 dan >= 22. Atribut credit_history yang sebelumnya terdiri dari 5 kategori (A, B, C, D, dan E) dikategorikan ulang menjadi 2 kategori yaitu ABC dan DE. Atribut purpose yang

sebelumnya terdiri dari 10 kategori (A, B, C, D, E, F, G, H, I, dan J) dikategorikan ulang menjadi 2 kategori yaitu BFHJ dan ACDEGI. Atribut savings_status yang sebelumnya terdiri dari 5 kategori (A, B, C, D, dan E) dikategorikan ulang menjadi 2 kategori yaitu BD dan ACE.

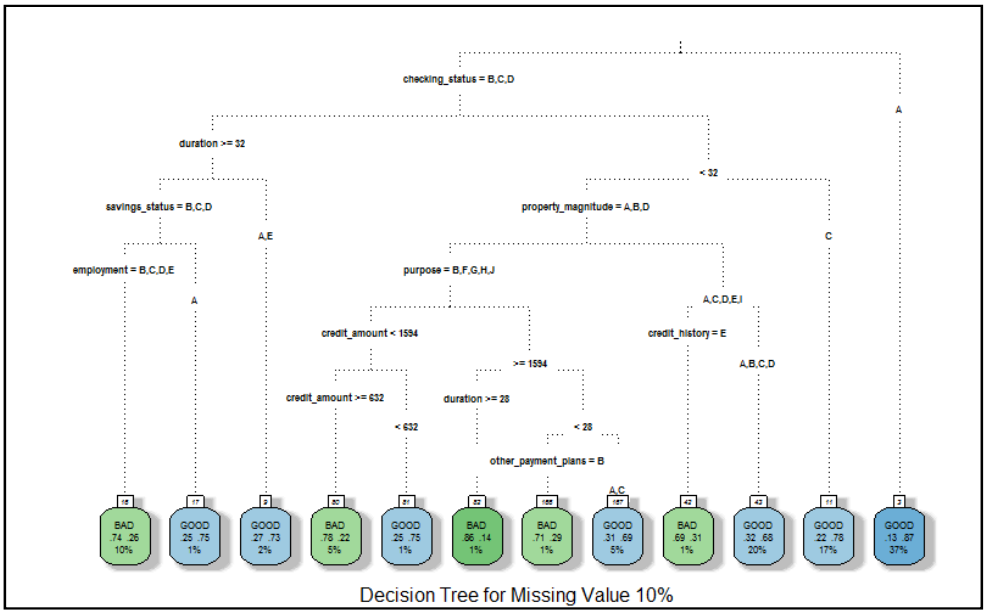
Atribut yang berpengaruh dan pengkategorian ulang hasil metode CART untuk data lengkap dapat dilihat pada tabel IV-2. Atribut berpengaruh yang telah dikategorikan ulang inilah yang selanjutnya akan dianalisis dengan Regresi Logistik untuk pembuatan model.

Tabel IV- 2 Pengkategorian Ulang Atribut Data Lengkap

Atribut yang Berpengaruh	Kategori Awal	Kategori Berdasarkan CART
checking_status	A; B; C; D	BC; AD
credit_amount	Numerik	<1306; >=1306
credit_history	A; B; C; D; E	ABC ; DE
duration	Numerik	<47; >=47
purpose	A; B; C; D; E; F; G; H; I; J	BFHJ; ACDGI
savings_status	A; B; C; D; E	BD; ACE

4.4.2 Missing Value 10%

Pohon klasifikasi dari dataset dengan *missing value* 10% dapat dilihat pada gambar IV-2.



Gambar IV-2 Pohon Klasifikasi *Missing Value* 10%

Gambar IV-2 adalah pohon klasifikasi untuk dataset dengan *missing value* 10%. Atribut yang berpengaruh terhadap peubah respon adalah checking_status, credit_amount, credit_history, duration, employment, property_magnitude, purpose, savings_status. Atribut – atribut tersebut kemudian dikategorikan ulang menjadi 2 kategori. Pengkategorian ulang dilakukan berdasarkan pohon klasifikasi yang terbentuk dengan menggunakan CART.

Berdasarkan pohon klasifikasi tersebut atribut checking_status yang sebelumnya terdiri dari 4 kategori (A, B, C, dan D) dikategorikan ulang menjadi 2 kategori yaitu BCD dan A. Atribut credit_amount yang sebelumnya adalah atribut numerik kemudian dikategorikan ulang menjadi 2 kategori yaitu <1594 dan >= 1594. Atribut credit_history yang sebelumnya terdiri dari 5 kategori (A, B, C, D, dan E) dikategorikan ulang menjadi 2 kategori yaitu ABCD dan E. Atribut duration yang sebelumnya adalah atribut numerik kemudian dikategorikan ulang

menjadi 2 kategori yaitu <32 dan >= 32. Atribut employment yang sebelumnya terdiri dari 5 kategori (A, B, C, D, dan E) dikategorikan ulang menjadi 2 kategori yaitu BCDE dan A. Atribut property_magnitude yang sebelumnya terdiri dari 4 kategori (A, B, C, dan D) dikategorikan ulang menjadi 2 kategori yaitu ABD dan C. Atribut purpose yang sebelumnya terdiri dari 10 kategori (A, B, C, D, E, F, G, H, I, dan J) dikategorikan ulang menjadi 2 kategori yaitu BFGHJ dan ACDEI. Atribut savings_status yang sebelumnya terdiri dari 5 kategori (A, B, C, D, dan E) dikategorikan ulang menjadi 2 kategori yaitu BCD dan AE.

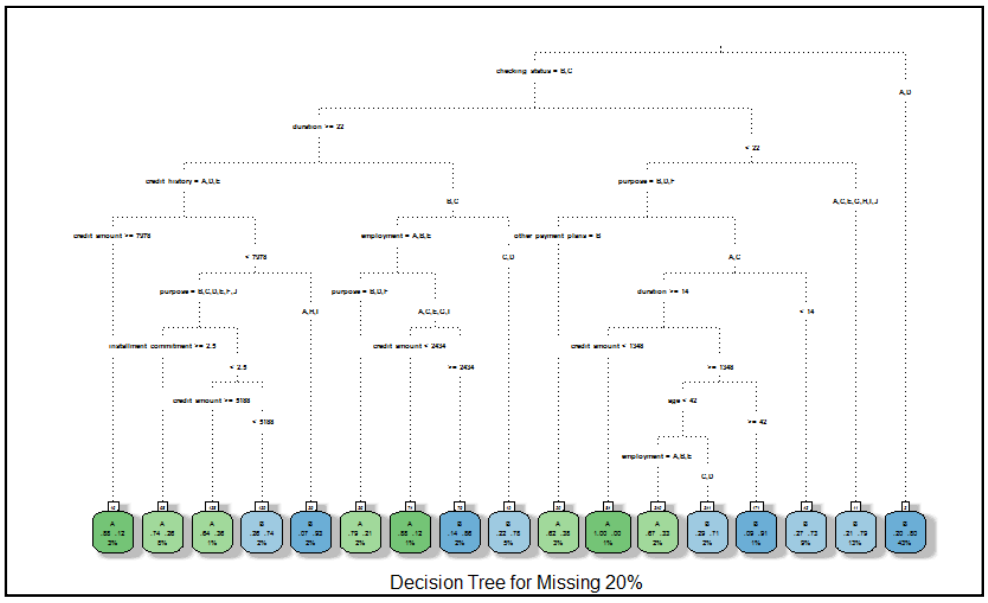
Atribut yang berpengaruh dan pengkategorian ulang hasil metode CART untuk data dengan *missing value* 10% dapat dilihat pada tabel IV-3. Atribut berpengaruh yang telah dikategorikan ulang inilah yang selanjutnya akan dianalisis dengan Regresi Logistik untuk pembuatan model.

Tabel IV-3 Pengkategorian Ulang Atribut Dataset *Missing Value* 10%

Atribut yang Berpengaruh	Kategori Awal	Kategori Berdasarkan CART
checking_status	A; B; C; D	A; BCD
credit_amount	Numerik	<1594; >=1594
credit_history	A; B; C; D; E	ABCD; E
duration	Numerik	<32; >=32
employment	A; B; C; D; E	A; BCDE
property_magnitude	A; B; C; D	ABD; C
purpose	A; B; C; D; E; F; G; H; I; J	BFGHJ; ACDEI
savings_status	A; B; C; D; E	BCD; AE

4.4.3 Missing Value 20%

Pohon klasifikasi dari dataset dengan *missing value* 20% dapat dilihat pada gambar IV-3.



Gambar IV-3 Pohon Klasifikasi *Missing Value* 20%

Gambar IV-3 adalah pohon klasifikasi untuk dataset dengan *missing value* 30%. Atribut yang berpengaruh terhadap peubah respon adalah age, checking_status, credit_amount, credit_history, duration, employment, installment_commitment, other_payment_plans, purpose. Atribut – atribut tersebut kemudian dikategorikan ulang menjadi 2 kategori. Pengkategorian ulang dilakukan berdasarkan pohon klasifikasi yang terbentuk dengan menggunakan CART.

Berdasarkan pohon klasifikasi tersebut atribut age yang sebelumnya adalah atribut numerik kemudian dikategorikan ulang menjadi 2 kategori yaitu <42.5 dan >= 42.5. Atribut checking_status yang sebelumnya terdiri dari 4 kategori (A, B, C, dan D) dikategorikan ulang menjadi 2 kategori yaitu BC dan AD. Atribut credit_amount yang sebelumnya adalah atribut numerik kemudian dikategorikan ulang menjadi 2 kategori yaitu <1348.5 dan >= 1348.5. Atribut

credit_history yang sebelumnya terdiri dari 5 kategori (A, B, C, D, dan E) dikategorikan ulang menjadi 2 kategori yaitu ADE dan BC. Atribut duration yang sebelumnya adalah atribut numerik kemudian dikategorikan ulang menjadi 2 kategori yaitu <21.5 dan >= 21.5. Atribut employment yang sebelumnya terdiri dari 5 kategori (A, B, C, D, dan E) dikategorikan ulang menjadi 2 kategori yaitu BCDE dan A. Atribut installment_commitment yang sebelumnya numerik dikategorikan ulang menjadi 2 kategori yaitu <2.5 dan >=2.5. Atribut other_installment_commitment yang sebelumnya terdiri dari 3 kategori (A, B, dan C) dikategorikan ulang menjadi 2 kategori yaitu AC dan B. Atribut purpose yang sebelumnya terdiri dari 4 kategori (A, B, C, dan D) dikategorikan ulang menjadi 2 kategori yaitu BD dan AC. Atribut purpose yang sebelumnya terdiri dari 10 kategori (A, B, C, D, E, F, G, H, I, dan J) dikategorikan ulang menjadi 2 kategori yaitu ACEGHIJ dan BDF.

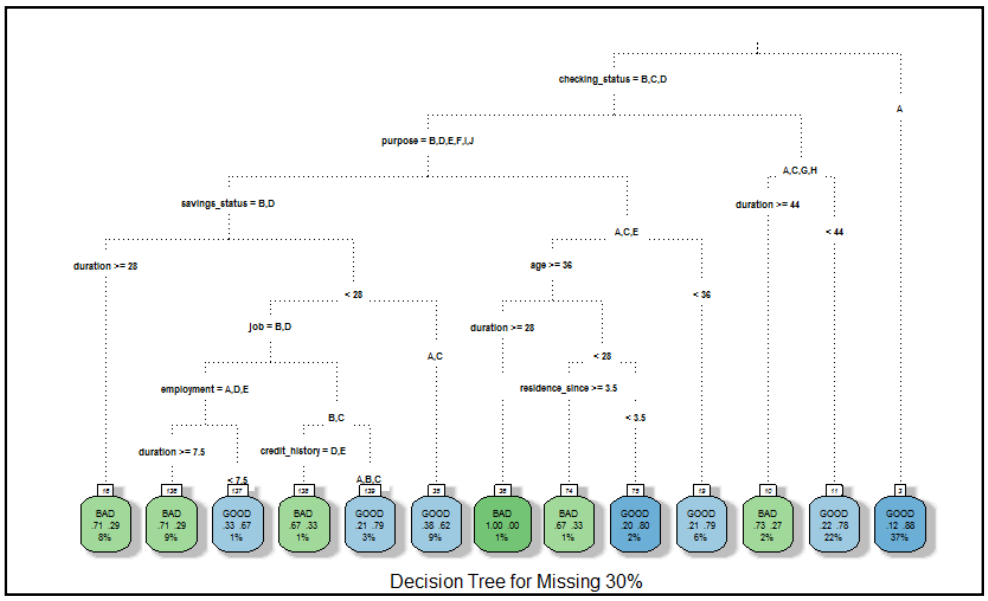
Atribut yang berpengaruh dan pengkategorian ulang hasil metode CART untuk data dengan *missing value* 20% dapat dilihat pada tabel IV-4. Atribut berpengaruh yang telah dikategorikan ulang inilah yang selanjutnya akan dianalisis dengan Regresi Logistik untuk pembuatan model.

Tabel IV- 4 Pengkategorian Ulang Atribut Dataset *Missing Value* 20%

Atribut yang Berpengaruh	**Kategori Awal**	**Kategori Berdasarkan CART**
age	Numerik	<42.5; >=42.5
checking_status	A; B; C; D	BC; AD
credit_amount	Numerik	<1348.5; >=1348.5
credit_history	A; B; C; D; E	ADE; BC
duration	Numerik	<21.5; >=21.5
employment	A; B; C; D; E	A; BCDE
installment_commitment	Numerik	<2.5; >=2.5
other_payment_plans	A; B; C	AC; BD
purpose	A; B; C; D; E; F; G; H; I; J	ACEGHIJ; BDF

4.4.4 Missing Value 30%

Pohon klasifikasi dari dataset dengan *missing value* 30% dapat dilihat pada gambar IV-4.



Gambar IV-4 Pohon Klasifikasi *Missing Value* 30%

Gambar IV-4 adalah pohon klasifikasi untuk dataset dengan *missing value* 30%. Atribut yang berpengaruh terhadap peubah respon adalah age, checking_status, credit_history, duration, employment, job, purpose, residence_since, savings_status. Atribut – atribut tersebut kemudian dikategorikan ulang menjadi 2 kategori. Pengkategorian ulang dilakukan berdasarkan pohon klasifikasi yang terbentuk dengan menggunakan CART.

Berdasarkan pohon klasifikasi tersebut atribut age yang sebelumnya adalah atribut numerik kemudian dikategorikan ulang menjadi 2 kategori yaitu <36.5 dan >= 36.5. Atribut checking_status yang sebelumnya terdiri dari 4 kategori (A, B, C, dan D) dikategorikan ulang menjadi 2 kategori yaitu A dan

BCD. Atribut credit_history yang sebelumnya terdiri dari 5 kategori (A, B, C, D, dan E) dikategorikan ulang menjadi 2 kategori yaitu ABC dan DE. Atribut duration yang sebelumnya adalah atribut numerik kemudian dikategorikan ulang menjadi 2 kategori yaitu <44 dan >= 44. Atribut employment yang sebelumnya terdiri dari 5 kategori (A, B, C, D, dan E) dikategorikan ulang menjadi 2 kategori yaitu ADE dan BC. Atribut job yang sebelumnya terdiri dari 4 kategori (A, B, C, dan D) dikategorikan ulang menjadi 2 kategori yaitu BD dan AC. Atribut purpose yang sebelumnya terdiri dari 10 kategori (A, B, C, D, E, F, G, H, I, dan J) dikategorikan ulang menjadi 2 kategori yaitu BDEFIJ dan ACGH. Atribut residence_since yang sebelumnya adalah atribut numerik kemudian dikategorikan ulang menjadi 2 kategori yaitu <3.5 dan >= 3.5. Atribut savings_status yang sebelumnya terdiri dari 5 kategori (A, B, C, D, dan E) dikategorikan ulang menjadi 2 kategori yaitu BD dan ACE.

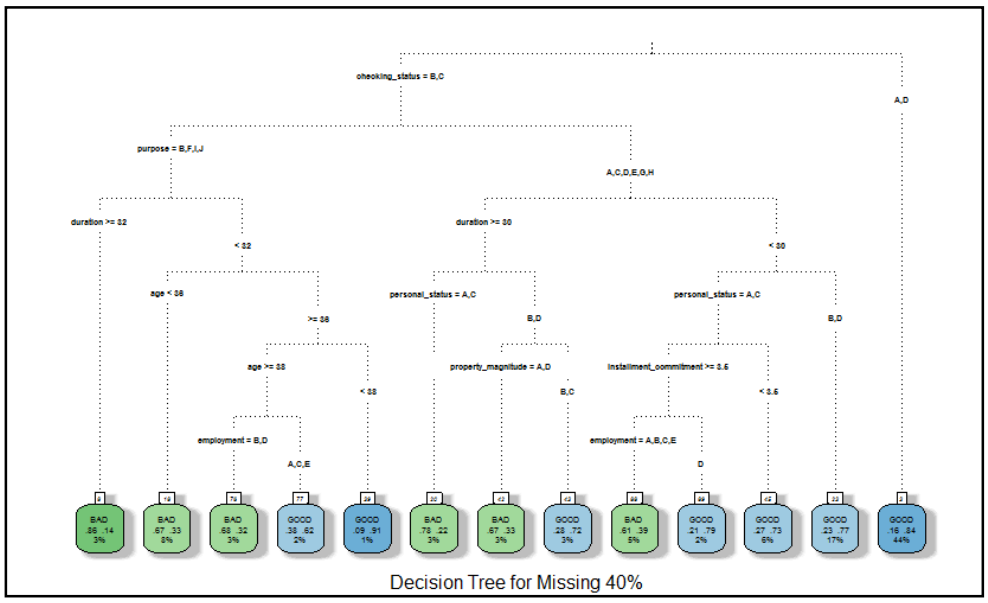
Atribut yang berpengaruh dan pengkategorian ulang hasil metode CART untuk data dengan *missing value* 30% dapat dilihat pada tabel IV-5. Atribut berpengaruh yang telah dikategorikan ulang inilah yang selanjutnya akan dianalisis dengan Regresi Logistik untuk pembuatan model.

Tabel IV-5 Pengkategorian Ulang Atribut Dataset *Missing Value* 30%

Atribut yang Berpengaruh	**Kategori Awal**	**Kategori Berdasarkan CART**
age	Numerik	<36.5; >=36.5
checking_status	A; B; C; D	A; BCD
credit_history	A; B; C; D; E	ABC; DE
duration	Numerik	<44; >=44
employment	A; B; C; D; E	ADE; BC
job	A; B; C; D	BD; AC
purpose	A; B; C; D; E; F; G; H; I; J	BDEFIJ; ACGH
residence_since	Numerik	<3.5; >=3.5
savings_status	A; B; C; D; E	BD; ACE

4.4.5 Missing Value 40%

Pohon klasifikasi dari dataset dengan *missing value* 40% dapat dilihat pada gambar IV-5.



Gambar IV-5 Pohon Klasifikasi *Missing Value* 40%

Gambar IV-5 adalah pohon klasifikasi untuk dataset dengan *missing value* 40%. Atribut yang berpengaruh terhadap peubah respon adalah age, checking_status, duration, employment, installment_commitment, personal_status, property_magnitude, purpose. Atribut – atribut tersebut kemudian dikategorikan ulang menjadi 2 kategori. Pengkategorian ulang dilakukan berdasarkan pohon klasifikasi yang terbentuk dengan menggunakan CART.

Berdasarkan pohon klasifikasi tersebut atribut age yang sebelumnya adalah atribut numerik kemudian dikategorikan ulang menjadi 2 kategori yaitu <37.5 dan >= 37.5. Atribut checking_status yang sebelumnya terdiri dari 4 kategori (A, B, C, dan D) dikategorikan ulang menjadi 2 kategori yaitu BC dan

AD. Atribut duration yang sebelumnya adalah atribut numerik kemudian dikategorikan ulang menjadi 2 kategori yaitu <31.5 dan >= 31.5. Atribut employment yang sebelumnya terdiri dari 5 kategori (A, B, C, D, dan E) dikategorikan ulang menjadi 2 kategori yaitu BD dan ACE. Atribut installment_commitment yang sebelumnya adalah atribut numerik kemudian dikategorikan ulang menjadi 2 kategori yaitu <3.5 dan >= 3.5. Atribut personal_status yang sebelumnya terdiri dari 4 kategori (A, B, C, dan D) dikategorikan ulang menjadi 2 kategori yaitu AC dan BD. Atribut property_magnitude yang sebelumnya terdiri dari 4 kategori (A, B, C, dan D) dikategorikan ulang menjadi 2 kategori yaitu AD dan BC. Atribut purpose yang sebelumnya terdiri dari 10 kategori (A, B, C, D, E, F, G, H, I, dan J) dikategorikan ulang menjadi 2 kategori yaitu BFIJ dan ACDEGH.

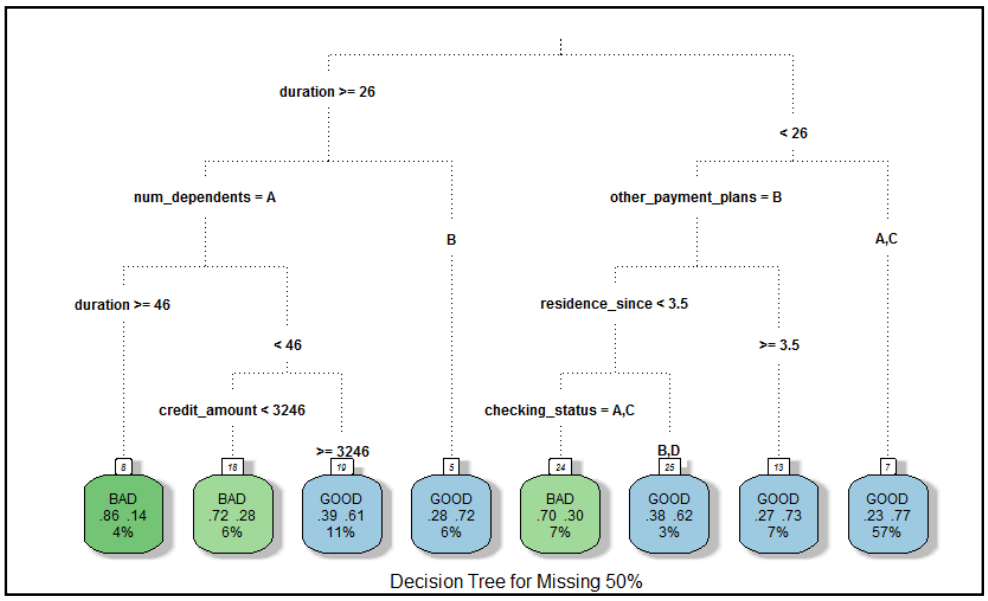
Atribut yang berpengaruh dan pengkategorian ulang hasil metode CART untuk data dengan *missing value* 40% dapat dilihat pada tabel IV-6. Atribut berpengaruh yang telah dikategorikan ulang inilah yang selanjutnya akan dianalisis dengan Regresi Logistik untuk pembuatan model.

Tabel IV-6 Pengkategorian Ulang Atribut Dataset *Missing Value* 40%

Atribut yang Berpengaruh	**Kategori Awal**	**Kategori Berdasarkan CART**
age	Numerik	<37.5; >=37.5
checking_status	A; B; C; D	BC; AD
duration	Numerik	<31.5; >=31.5
employment	A; B; C; D; E	BD; ACE
installment_commitment	Numerik	<3.5; >=3.5
personal_status	A; B; C; D	AC; BD
property_magnitude	A; B; C; D	AD; BC
purpose	A; B; C; D; E; F; G; H; I; J	BFIJ; ACDEGH

4.4.6 Missing Value 50%

Pohon klasifikasi dari dataset dengan *missing value* 50% dapat dilihat pada gambar IV-6.



Gambar IV-6 Pohon Klasifikasi *Missing Value* 50%

Gambar IV-6 adalah pohon klasifikasi untuk dataset dengan *missing value* 50%. Atribut yang berpengaruh terhadap peubah respon adalah checking_status, credit_amount, duration, num_dependents, other_payment_plans, residence_since. Atribut – atribut tersebut kemudian dikategorikan ulang menjadi 2 kategori. Pengkategorian ulang dilakukan berdasarkan pohon klasifikasi yang terbentuk dengan menggunakan CART.

Berdasarkan pohon klasifikasi tersebut atribut checking_status yang sebelumnya terdiri dari 4 kategori (A, B, C, dan D) dikategorikan ulang menjadi 2 kategori yaitu AC dan BD. Atribut credit_amount yang sebelumnya adalah atribut numerik kemudian dikategorikan ulang menjadi 2 kategori yaitu <3246 dan >=

3246. Atribut duration yang sebelumnya adalah atribut numerik kemudian dikategorikan ulang menjadi 2 kategori yaitu <26 dan >= 26. Atribut num_dependents tetap menjadi 2 kategori yaitu A dan B. Atribut other_payment_plans yang sebelumnya terdiri dari 3 kategori (A, B, dan C) dikategorikan ulang menjadi 2 kategori yaitu AC dan B. Atribut residence_since yang sebelumnya adalah atribut numerik kemudian dikategorikan ulang menjadi 2 kategori yaitu <3.5 dan >= 3.5.

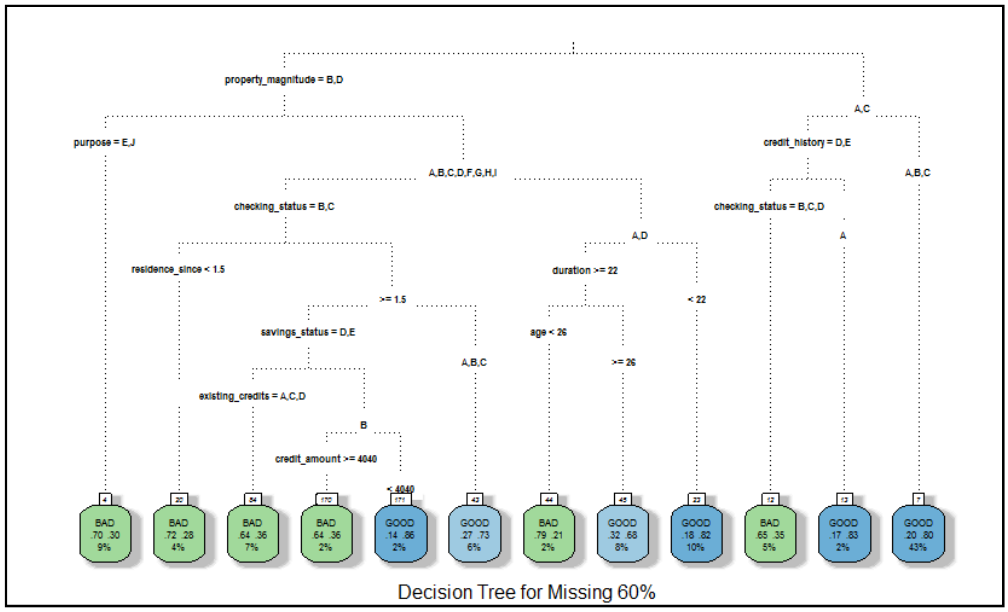
Atribut yang berpengaruh dan pengkategorian ulang hasil metode CART untuk data dengan *missing value* 50% dapat dilihat pada tabel IV-7. Atribut berpengaruh yang telah dikategorikan ulang inilah yang selanjutnya akan dianalisis dengan Regresi Logistik untuk pembuatan model.

Tabel IV-7 Pengkategorian Ulang Atribut Dataset *Missing Value* 50%

Atribut yang Berpengaruh	**Kategori Awal**	**Kategori Berdasarkan CART**
checking_status	A; B; C; D	AC; BD
credit_amount	Numerik	<3246; >=3246
duration	Numerik	<26; >=26
num_dependents	A; B	A; B
other_payment_plans	A; B; C	AC; B
residence_since	Numerik	<3.5; >=3.5

4.4.7 Missing Value 60%

Pohon klasifikasi dari dataset dengan *missing value* 60% dapat dilihat pada gambar IV-7.



Gambar IV-7 Pohon Klasifikasi *Missing Value* 60%

Gambar IV-7 adalah pohon klasifikasi untuk dataset dengan *missing value* 60%. Atribut yang berpengaruh terhadap peubah respon adalah age, checking_status, credit_amount, credit_history, duration, existing_credits, property_magnitude, purpose, residence_since, savings_status. Atribut – atribut tersebut kemudian dikategorikan ulang menjadi 2 kategori. Pengkategorian ulang dilakukan berdasarkan pohon klasifikasi yang terbentuk dengan menggunakan CART.

Berdasarkan pohon klasifikasi tersebut atribut age yang sebelumnya adalah atribut numerik kemudian dikategorikan ulang menjadi 2 kategori yaitu <25.5 dan >=25.5. Atribut checking_status yang sebelumnya terdiri dari 4 kategori (A, B, C, dan D) dikategorikan ulang menjadi 2 kategori yaitu BC dan AD. Atribut credit_amount yang sebelumnya adalah atribut numerik kemudian dikategorikan ulang menjadi 2 kategori yaitu <4040.5 dan >=4040.5. Atribut

credit_history yang sebelumnya terdiri dari 5 kategori (A, B, C, D, dan E) dikategorikan ulang menjadi 2 kategori yaitu ABC dan DE. Atribut duration yang sebelumnya adalah atribut numerik kemudian dikategorikan ulang menjadi 2 kategori yaitu <21.5 dan >= 21.5. Atribut existing_credit yang sebelumnya terdiri dari 4 kategori (A, B, C, dan D) dikategorikan ulang menjadi 2 kategori yaitu ACD dan B. Atribut property_magnitude yang sebelumnya terdiri dari 4 kategori (A, B, C, dan D) dikategorikan ulang menjadi 2 kategori yaitu AC dan BD. Atribut purpose yang sebelumnya terdiri dari 10 kategori (A, B, C, D, E, F, G, H, I, dan J) dikategorikan ulang menjadi 2 kategori yaitu ABCDFGHI dan EJ. Atribut residence_since yang sebelumnya adalah atribut numerik kemudian dikategorikan ulang menjadi 2 kategori yaitu <1.5 dan >= 1.5. Atribut savings_status yang sebelumnya terdiri dari 5 kategori (A, B, C, D, dan E) dikategorikan ulang menjadi 2 kategori yaitu DE dan ABC.

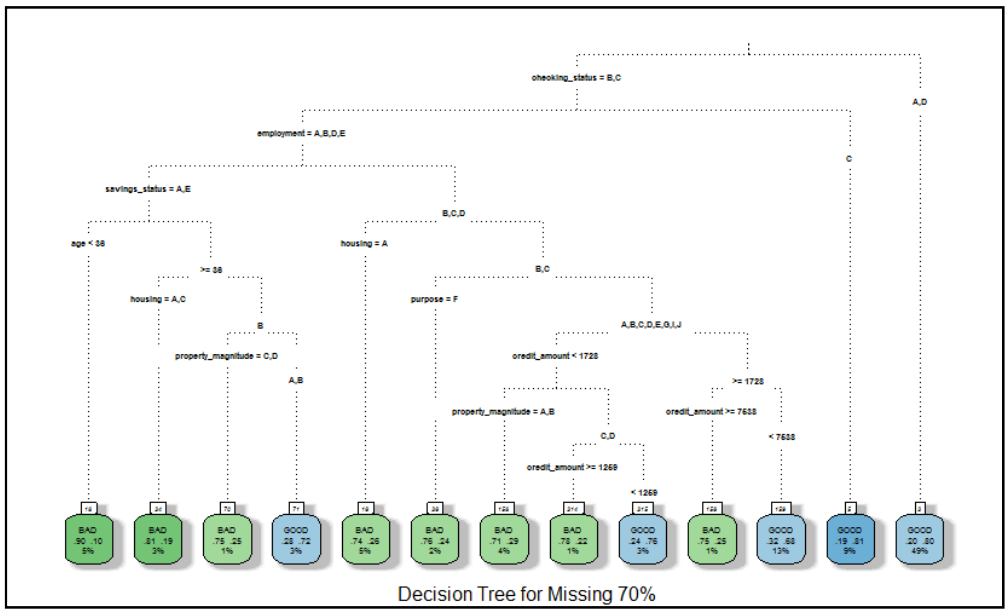
Atribut yang berpengaruh dan pengkategorian ulang hasil metode CART untuk data dengan *missing value* 60% dapat dilihat pada tabel 4.8. Atribut berpengaruh yang telah dikategorikan ulang inilah yang selanjutnya akan dianalisis dengan Regresi Logistik untuk pembuatan model.

Tabel IV-8 Pengkategorian Ulang Atribut Dataset *Missing Value* 60%

Atribut yang Berpengaruh	**Kategori Awal**	**Kategori Berdasarkan CART**
age	Numerik	<25.5; >=25.5
checking_status	A; B; C; D	BC; AD
credit_amount	Numerik	<4040.5; >=4040.5
credit_history	A; B; C; D; E	ABC ; DE
duration	Numerik	<21.5; >=21.5
existing_credits	A; B; C; D	ACD; B
property_magnitude	A; B; C; D	AC; BD
purpose	A; B; C; D; E; F; G; H; I; J	ABCGFGHI; EJ
residence_since	Numerik	<1.5; >=1.5
savings_status	A; B; C; D; E	DE; ABC

4.4.8 Missing Value 70%

Pohon klasifikasi dari dataset dengan *missing value* 70% dapat dilihat pada gambar IV-8.



Gambar IV-8 Pohon Klasifikasi *Missing Value* 70%

Gambar IV-8 adalah pohon klasifikasi untuk dataset dengan *missing value* 70%. Atribut yang berpengaruh terhadap peubah respon adalah age, checking_status, credit_amount, employment, housing, property_magnitude, purpose, savings_status. Atribut – atribut tersebut kemudian dikategorikan ulang menjadi 2 kategori. Pengkategorian ulang dilakukan berdasarkan pohon klasifikasi yang terbentuk dengan menggunakan CART.

Berdasarkan pohon klasifikasi tersebut atribut age yang sebelumnya adalah atribut numerik kemudian dikategorikan ulang menjadi 2 kategori yaitu <35.5 dan >=35.5. Atribut checking_status yang sebelumnya terdiri dari 4

kategori (A, B, C, dan D) dikategorikan ulang menjadi 2 kategori yaitu BC dan AD. Atribut credit_amount yang sebelumnya adalah atribut numerik kemudian dikategorikan ulang menjadi 2 kategori yaitu <1259 dan >=1259. Atribut employment yang sebelumnya terdiri dari 5 kategori (A, B, C, D, dan E) dikategorikan ulang menjadi 2 kategori yaitu ABDE dan C. Atribut housing yang sebelumnya terdiri dari 3 kategori (A, B, dan C) dikategorikan ulang menjadi 2 kategori yaitu A dan BC. Atribut property_magnitude yang sebelumnya terdiri dari 4 kategori (A, B, C, dan D) dikategorikan ulang menjadi 2 kategori yaitu AB dan CD. Atribut purpose yang sebelumnya terdiri dari 10 kategori (A, B, C, D, E, F, G, H, I, dan J) dikategorikan ulang menjadi 2 kategori yaitu ABCDEGHIJ dan J. Atribut savings_status yang sebelumnya terdiri dari 5 kategori (A, B, C, D, dan E) dikategorikan ulang menjadi 2 kategori yaitu BCD dan AE.

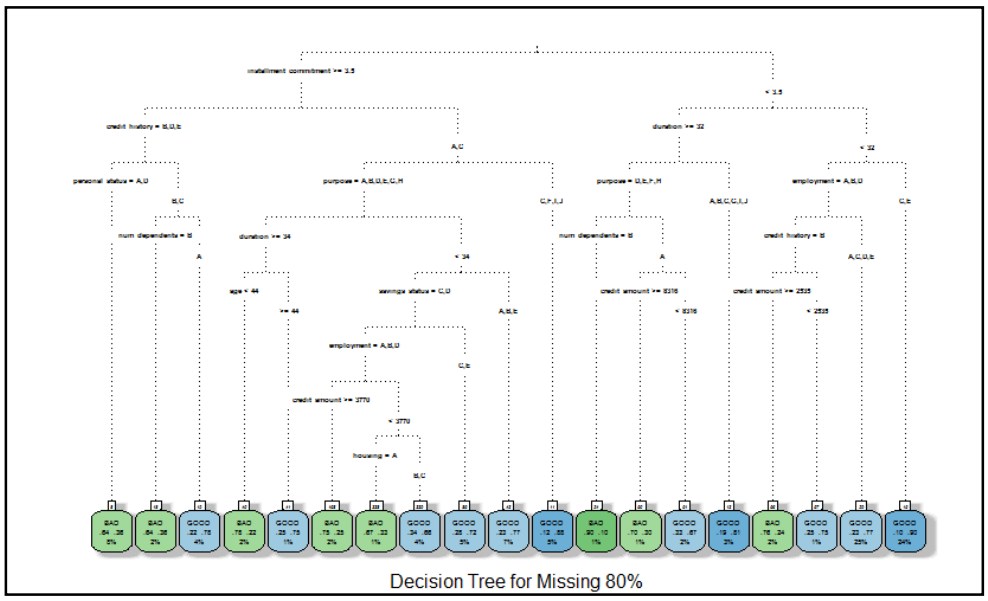
Atribut yang berpengaruh dan pengkategorian ulang hasil metode CART untuk data dengan *missing value* 70% dapat dilihat pada tabel 4.9. Atribut berpengaruh yang telah dikategorikan ulang inilah yang selanjutnya akan dianalisis dengan Regresi Logistik untuk pembuatan model.

Tabel IV-9 Pengkategorian Ulang Atribut Dataset *Missing Value* 70%

Atribut yang Berpengaruh	**Kategori Awal**	**Kategori Berdasarkan CART**
age	Numerik	<35.5; >=35.5
checking_status	A; B; C; D	BC; AD
credit_amount	Numerik	<1259; >=1259
employment	A; B; C; D; E	ABDE; C
housing	A; B; C	A; BC
property_magnitude	A; B; C; D	AB; CD
purpose	A; B; C; D; E; F; G; H; I; J	ABCDEGHIJ; F
savings_status	A; B; C; D; E	BCD; AE

4.4.9 Missing Value 80%

Pohon klasifikasi dari dataset dengan *missing value* 80% dapat dilihat pada gambar IV-9.



Gambar IV-9 Pohon Klasifikasi *Missing Value* 80%

Gambar IV-9 adalah pohon klasifikasi untuk dataset dengan *missing value* 80%. Atribut yang berpengaruh terhadap peubah respon adalah age, credit_amount, credit_history, duration, employment, housing, installment_commitment, num_dependents, personal_status, purpose, savings_status. Atribut – atribut tersebut kemudian dikategorikan ulang menjadi 2 kategori. Pengkategorian ulang dilakukan berdasarkan pohon klasifikasi yang terbentuk dengan menggunakan CART.

Berdasarkan pohon klasifikasi tersebut atribut age yang sebelumnya adalah atribut numerik kemudian dikategorikan ulang menjadi 2 kategori yaitu <44 dan >=44. Atribut credit_amount yang sebelumnya adalah atribut numerik

kemudian dikategorikan ulang menjadi 2 kategori yaitu <2535 dan >=2535. Atribut credit_history yang sebelumnya terdiri dari 5 kategori (A, B, C, D, dan E) dikategorikan ulang menjadi 2 kategori yaitu ABE dan CD. Atribut duration yang sebelumnya adalah atribut numerik kemudian dikategorikan ulang menjadi 2 kategori yaitu <31.5 dan >=31.5. Atribut employment yang sebelumnya terdiri dari 5 kategori (A, B, C, D, dan E) dikategorikan ulang menjadi 2 kategori yaitu ABD dan CE. Atribut housing yang sebelumnya terdiri dari 3 kategori (A, B, dan C) dikategorikan ulang menjadi 2 kategori yaitu A dan BC. Atribut installment_commitment yang sebelumnya adalah atribut numerik kemudian dikategorikan ulang menjadi 2 kategori yaitu <3.5 dan >=3.5. Atribut num_dependents tetap menjadi 2 kategori yaitu A dan B. Atribut personal_status yang sebelumnya terdiri dari 4 kategori (A, B, C, dan D) dikategorikan ulang menjadi 2 kategori yaitu AD dan BC. Atribut purpose yang sebelumnya terdiri dari 10 kategori (A, B, C, D, E, F, G, H, I, dan J) dikategorikan ulang menjadi 2 kategori yaitu ABCGIJ dan ACDE. Atribut savings_status yang sebelumnya terdiri dari 5 kategori (A, B, C, D, dan E) dikategorikan ulang menjadi 2 kategori yaitu ABE dan CD.

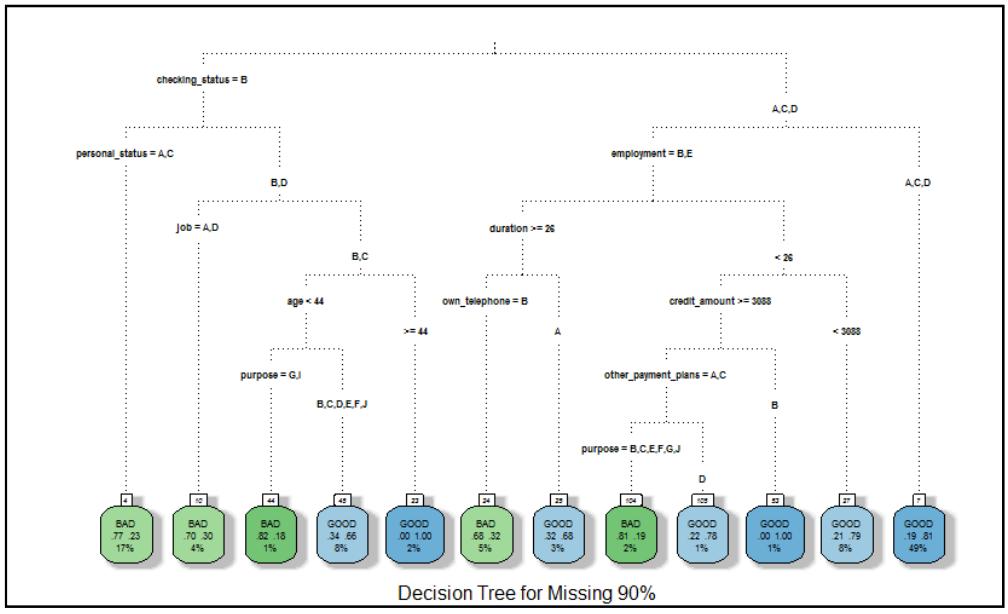
Atribut yang berpengaruh dan pengkategorian ulang hasil metode CART untuk data dengan *missing value* 80% dapat dilihat pada tabel IV-10. Atribut berpengaruh yang telah dikategorikan ulang inilah yang selanjutnya akan dianalisis dengan Regresi Logistik untuk pembuatan model.

Tabel IV- 10 Pengkategorian Ulang Atribut Dataset *Missing Value* 80%

Atribut yang Berpengaruh	Kategori Awal	Kategori Berdasarkan CART
age	Numerik	<44; >=44
credit_amount	Numerik	<2535; >=2535
credit_history	A; B; C; D; E	ABE ; CD
duration	Numerik	<31.5; >=31.5
employment	A; B; C; D; E	ABD; CE
housing	A; B; C	A; BC
installment_commitment	Numerik	<3.5; >=3.5
num_dependents	A; B	A; B
personal_status	A; B; C; D	AD; BC
purpose	A; B; C; D; E; F; G; H; I; J	ABCGIJ; ACDE
savings_status	A; B; C; D; E	ABE; CD

4.4.10 Missing Value 90%

Pohon klasifikasi dari dataset dengan *missing value* 90% dapat dilihat pada gambar IV-10.



Gambar IV-10 Pohon Klasifikasi *Missing Value* 90%

Gambar IV-10 adalah pohon klasifikasi untuk dataset dengan *missing value* 90%. Atribut yang berpengaruh terhadap peubah respon adalah age,

checking_status, credit_amount, duration, employment, job, other_payment_plans, own_telephone, personal_status, purpose. Atribut – atribut tersebut kemudian dikategorikan ulang menjadi 2 kategori. Pengkategorian ulang dilakukan berdasarkan pohon klasifikasi yang terbentuk dengan menggunakan CART.

Berdasarkan pohon klasifikasi tersebut atribut age yang sebelumnya adalah atribut numerik kemudian dikategorikan ulang menjadi 2 kategori yaitu <44.5 dan >=44.5. Atribut checking_status yang sebelumnya terdiri dari 4 kategori (A, B, C, dan D) dikategorikan ulang menjadi 2 kategori yaitu ACD dan B. Atribut credit_amount yang sebelumnya adalah atribut numerik kemudian dikategorikan ulang menjadi 2 kategori yaitu <3088.5 dan >=3088.5. Atribut duration yang sebelumnya adalah atribut numerik kemudian dikategorikan ulang menjadi 2 kategori yaitu <26.5 dan >=26.5. Atribut employment yang sebelumnya terdiri dari 5 kategori (A, B, C, D, dan E) dikategorikan ulang menjadi 2 kategori yaitu BE dan ACD. Atribut job yang sebelumnya terdiri dari 4 kategori (A, B, C, dan D) dikategorikan ulang menjadi 2 kategori yaitu AD dan BC. Atribut other_payment_plans yang sebelumnya terdiri dari 3 kategori (A, B, dan C) dikategorikan ulang menjadi 2 kategori yaitu AC dan B. Atribut own_telephone tetap menjadi 2 kategori yaitu A dan B. Atribut personal_status yang sebelumnya terdiri dari 4 kategori (A, B, C, dan D) dikategorikan ulang menjadi 2 kategori yaitu AC dan BD. Atribut purpose yang sebelumnya terdiri dari 10 kategori (A, B, C, D, E, F, G, H, I, dan J) dikategorikan ulang menjadi 2 kategori yaitu ABCDEFJ dan GI.

Atribut yang berpengaruh dan pengkategorian ulang hasil metode CART untuk data dengan *missing value* 80% dapat dilihat pada tabel 4.10. Atribut berpengaruh yang telah dikategorikan ulang inilah yang selanjutnya akan dianalisis dengan Regresi Logistik untuk pembuatan model.

Tabel IV-11 Pengkategorian Ulang Atribut Dataset *Missing Value* 90%

Atribut yang Berpengaruh	Kategori Awal	Kategori Berdasarkan CART
age	Numerik	<44.5; >=44.5
checking_status	A; B; C; D	ACD; B
credit_amount	Numerik	<3088.5; 3088.5
duration	Numerik	<26.5; >=26.5
employment	A; B; C; D; E	BE; ACD
job	A; B; C; D	AD; BC
other_payment_plans	A; B; C	AC; B
own_telephone	A; B	A; B
personal_status	A; B; C; D	AC; BD
purpose	A; B; C; D; E; F; G; H; I; J	ABCDEFJ; GI

4.4.11 Hasil Analisis CART

Atribut yang berpengaruh dari masing - masing dataset dapat dilihat pada tabel IV-12. Jumlah atribut yang berpengaruh dari masing – masing dataset berbeda – beda. Jumlah atribut yang berpengaruh paling sedikit adalah data lengkap dan data dengan *missing value* 50%. Sedangkan dataset dengan atribut yang berpengaruh paling banyak adalah dataset dengan *missing value* 80% yaitu sebanyak 11 atribut.

Tabel IV-12 Atribut yang Berpengaruh

Dataset	Jumlah Atribut yang Berpengaruh	Atribut yang Berpengaruh
Data Lengkap	6	checking_status, credit_amount, credit_history, duration, purpose, savings_status
Missing Value 10%	8	checking_status, credit_amount, credit_history, duration, employment, property_magnitude, purpose, savings_status
Missing Value 20%	9	age,checking_status,credit_amount, credit_history, duration, employment, installment_commitment, other_payment_plans, purpose
Missing Value 30%	9	age, checking_status, credit_history, duration, employment, job, purpose, residence_since, savings_status
Missing Value 40%	8	age,checking_status, duration, employment, installment_commitment, personal_status, property_magnitude, purpose
Missing Value 50%	6	checking_status, credit_amount, duration, num_dependents, other_payment_plans, residence_since
Missing Value 60%	10	age, checking_status, credit_amount, credit_history, duration, existing_credits, property_magnitude, purpose, residence_since, savings_status
Missing Value 70%	8	age,checking_status, credit_amount, employment, housing, property_magnitude, purpose, savings_status
Missing Value 80%	11	age, credit_amount, credit_history, duration, employment, housing, installment_commitment, num_dependents, personal_status, purpose, savings_status
Missing Value 90%	10	age,checking_status,credit_amount,duration,employment,job ,other_payment_plans,own_telephone,personal_status,purpo se

Atribut checking_status, duration, dan purpose adalah atribut yang paling banyak muncul di semua dataset. Ketiga atribut ini berpengaruh pada 9 dataset. Sedangkan atribut other_parties dan foreign_worker tidak berpengaruh di semua dataset. Jumlah kemunculan atribut dalam setiap dataset dapat dilihat pada tabel IV-13.

Checking_status adalah hasil pengecekan terhadap rekening debitur. Jumlah uang yang keluar masuk dalam rekening debitur. Jumlah uang yang keluar masuk di rekening dapat menggambarkan kekuatan finansial debitur dan kemampuan debitur untuk membayar kredit yang diberikan.

Duration adalah jangka waktu pinjaman yang diberikan kepada debitur. Semakin panjang jangka waktu yang diberikan maka cicilan yang harus dibayar setiap bulannya lebih kecil. Sedangkan semakin pendek jangka waktu yang diberikan maka semakin besar cicilan yang harus dibayar setiap bulannya semakin besar. Pemberian jangka waktu kredit ini biasanya disesuaikan dengan kemampuan finansial debitur.

Purpose adalah tujuan debitur mengajukan kredit. Tujuan debitur mengajukan kredit bisa untuk konsumsi seperti pembelian mobil, alat – alat rumah tangga, atau pembelian rumah. Bisa juga untuk keperluan bisnis seperti modal kerja, pembelian mesin, dan pembelian pabrik. Untuk keperluan konsumsi kemampuan membayar debitur bisa dilihat dari penghasilan bulanan debitur, sedangkan untuk keperluan bisnis kemampuan bayar debitur bisa dilihat dari laporan keuangan perusahaan.

Tabel IV- 13 Jumlah Kemunculan Atribut dalam Setiap Dataset

Atribut	Jumlah Kemunculan
checking_status	9
duration	9
purpose	9
credit_amount	8
age	7
employment	7
credit_history	6
savings_status	6
property_magnitude	4
personal_status	3
other_payment_plans	3
installment_commitment	3
residence_since	3
job	2
num_dependents	2
housing	2
own_telephone	1
existing_credits	1
other_parties	0
foreign_worker	0

4.5 Pemodelan Menggunakan Regresi Logistik

Pemodelan dilakukan dengan menggunakan regresi Logistik. Tujuan dari pemodelan ini adalah untuk mendapatkan persamaan regresi yang dapat memprediksi dua atau lebih kelompok objek yang dapat mengklasifikasikan debitur menjadi debitur baik dan buruk. Dalam penelitian ini peubah responnya adalah baik buruknya debitur dalam hal risikonya dalam kelancaran pembayaran kredit, sedangkan atributnya adalah profil debitur. Ketepatan model regresi dapat diukur menggunakan CCR.

Hasil analisis regresi logistik untuk seluruh dataset dapat dilihat pada tabel IV-14.

Tabel IV-14 Hasil Analisis Regresi Logistik

Dataset	Jumlah Atribut	CCR	Baik	Buruk	Atribut	β	Rasio Odds
Data lengkap	6	75%	91	9	(Intercept)	2.67	14.46
					checking_status_tBC	-1.40	0.25
					credit_amount_t>=1306	-0.01	0.99
					duration_t>=47	-1.13	0.32
					credit_history_tDE	-1.07	0.34
					purpose_tBFHJ	-0.61	0.55
					savings_status_tBD	-0.74	0.48
Missing Value 10%	8	75%	89	11	(Intercept)	2.57	13.06
					checking_status_tBCD	-1.50	0.22
					credit_amount_t>=1594	0.13	1.14
					duration_t>=32	-1.07	0.34
					credit_history_tE	-0.93	0.39
					purpose_tBFGHJ	-0.55	0.58
					savings_status_tBCD	-0.50	0.60
					employment_tBCDE	-0.14	0.87
					property_magnitude_tC	0.79	2.20

Tabel IV-14 Hasil Analisis Regresi Logistik (lanjutan)

Missing Value 20%	9	75%	89	11	(Intercept)	1.49	4.43
					checking_status_tBC	-0.96	0.38
					duration_t>=21.5	-0.72	0.49
					purpose_tBDF	-0.49	0.62
					age_t>=42.5	0.47	1.60
					credit_history_tBC	0.69	1.99
					employment_tCD	0.13	1.14
					installment_commitment_t>= 2.5	-0.19	0.83
					other_payment_plans_tB	-0.29	0.75
					credit_amount_t>=1348.5	0.20	1.23
Missing Value 30%	9	74%	87	13	(Intercept)	3.78	43.63
					checking_status_tBCD	-1.49	0.23
					duration_t>=44	-1.31	0.27
					purpose_tBDEFIJ	-1.03	0.36
					savings_status_tBD	-0.82	0.44
					age_t>=36.5	-0.30	0.74
					credit_history_tDE	-0.89	0.41
					employment_tBC	0.41	1.51
					job_tBD	-0.37	0.69
					residence_since_t>=3.5	-0.39	0.68
Missing Value 40%	8	75%	77	23	(Intercept)	1.83	6.20
					checking_status_tBC	-1.38	0.25
					duration_t>=31.5	-1.19	0.30
					purpose_tBFIJ	-0.94	0.39
					employment_tBD	0.12	1.13
					personal_status_tBD	0.32	1.38
					property_magnitude_tBC	0.34	1.41
					age_t>=37.5	0.11	1.12
					installment_commitment_t>= 3.5	-0.43	0.65
Missing Value 50%	6	63%	77	23	(Intercept)	0.99	2.70
					checking_status_tBD	0.07	1.07
					duration_t>=26	-0.99	0.37
					credit_amount_t>=3246	-0.06	0.95
					num_dependents_tB	0.65	1.91
					other_payment_plans_tB	-1.04	0.35
					residence_since_t>=3.5	0.26	1.29
Missing Value 60%	10	69%	72	28	(Intercept)	1.63	5.10
					checking_status_tBC	-0.68	0.51
					duration_t>=21.5	-0.58	0.56
					purpose_tEJ	-1.05	0.35
					savings_status_tDE	-0.56	0.57
					credit_history_tDE	-0.91	0.40
					credit_amount_t>=4040.5	0.07	1.07
					existing_credits_tB	0.28	1.32
					property_magnitude_tBD	-0.88	0.42
					age_t>=25.5	0.39	1.48
					residence_since_t>=1.5	0.44	1.55

Tabel IV-14 Hasil Analisis Regresi Logistik (lanjutan)

Missing Value 70%	8	66%	72	28	(Intercept)	-0.62	0.54
					checking_status_tBC	-1.46	0.23
					purpose_tF	-0.75	0.47
					savings_status_tBCD	0.85	2.35
					employment_tC	1.78	5.96
					credit_amount_t>=1259	0.03	1.03
					housing_tBC	1.13	3.10
					property_magnitude_tCD	0.06	1.06
					age_t>=35.5	0.61	1.85
Missing Value 80%	11	66%	95	5	(Intercept)	1.17	3.23
					duration_t>=31.5	-0.64	0.53
					purpose_tACDE	-0.44	0.64
					savings_status_tCD	-0.52	0.59
					credit_history_tCD	0.02	1.02
					employment_tCE	0.93	2.54
					credit_amount_t>=2535	-0.24	0.79
					housing_tBC	0.34	1.41
					personal_status_tBC	0.62	1.86
					age_t>=44	0.64	1.89
					installment_commitment_t>= 3.5	-0.77	0.46
					num_dependents_tB	-0.81	0.44
Missing Value 90%	10	66%	88	12	(Intercept)	0.78	2.17
					checking_status_tB	-1.25	0.29
					purpose_tBCDEFJ	0.12	1.13
					age_t>=44.5	1.04	2.83
					personal_status_tBD	1.15	3.15
					duration_t>=26.5	-0.14	0.87
					employment_tBE	-1.06	0.35
					job_tBC	0.46	1.58
					other_payment_plans_tB	0.45	1.56
					own_telephone_tB	-1.01	0.36
					credit_amount_t>=3088.5	0.02	1.02

Sehingga model yang terbentuk adalah sebagai berikut :

1) Model Data Lengkap

$$\pi(x) = \frac{\exp(2.67 - 1.4\,CS - 0.01\,CA - 1.13\,D - 1.07\,CH - 0.61\,P - 0.74\,SS)}{1 + \exp(2.67 - 1.4\,CS - 0.01\,CA - 1.13\,D - 1.07\,CH - 0.61\,P - 0.74\,SS)} \quad (4.1)$$

2) Model *Missing Value 10%*

$$\pi(x) = \frac{\exp(2.57 - 1.5\,CS + 0.13\,CA - 1.07\,D - 0.93\,CH - 0.55\,P - 0.5\,SS - 0.14\,E + 0.17\,PM)}{1 + \exp(2.57 - 1.5\,CS + 0.13\,CA - 1.07\,D - 0.93\,CH - 0.55\,P - 0.5\,SS - 0.14\,E + 0.17\,PM)} \quad (4.2)$$

3) Model *Missing Value 20%*

$$\pi(x) = \frac{\exp(1.49-0.96\,CS-0.72\,D-0.49\,P+0.47\,A+0.69\,CH+0.1\,E-0.19\,I-0.29\,OP+0.2CA)}{1+\exp(1.49-0.96\,CS-0.72\,D-0.49\,P+0.47\,A+0.69\,CH+0.1\,E-0.19\,I-0.29\,OP+0.2CA)} \quad (4.3)$$

4) Model *Missing Value 30%*

$$\pi(x) = \frac{\exp(3.78-1.49\,CS-1.31\,D-1.03\,P-0.82\,SS-0.3\,A-0.89\,CH+0.41\,E-0.37\,J-0.39\,RS)}{1+\exp(3.78-1.49\,CS-1.31\,D-1.03\,P-0.82\,SS-0.3\,A-0.89\,CH+0.41\,E-0.37\,J-0.39\,RS)} \quad (4.4)$$

5) Model *Missing Value 40%*

$$\pi(x) = \frac{\exp(1.83-1.38\,CS-1.19\,D-0.94\,P+0.12\,E+0.23\,PS+0.34\,PM+0.11\,A-0.43\,I)}{1+\exp(1.83-1.38\,CS-1.19\,D-0.94\,P+0.12\,E+0.23\,PS+0.34\,PM+0.11\,A-0.43\,I)} \quad (4.5)$$

6) Model *Missing Value 50%*

$$\pi(x) = \frac{\exp(0.99-0.07\,CS-0.99\,D-0.06\,CA+0.65\,ND-1.04\,OP+0.26\,RS)}{1+\exp(0.99-0.07\,CS-0.99\,D-0.06\,CA+0.65\,ND-1.04\,OP+0.26\,RS)} \quad (4.6)$$

7) Model *Missing Value 60%*

$$\pi(x) = \frac{\exp(1.63-0.68\,CS-0.58\,D-1.05\,P-0.56\,SS-0.91\,CH+0.07\,CA+0.28\,EC-0.88\,PM+0.39\,A+0.44\,RS)}{1+\exp(1.63-0.68\,CS-0.58\,D-1.05\,P-0.56\,SS-0.91\,CH+0.07\,CA+0.28\,EC-0.88\,PM+0.39\,A+0.44\,RS)} \quad (4.7)$$

8) Model *Missing Value 70%*

$$\pi(x) = \frac{\exp(-0.62-1.46\,CS-0.75\,P+0.85\,SS+1.78\,E+0.03\,CA+1.13\,H+0.06\,PM+0.61\,A)}{1+\exp(-0.62-1.46\,CS-0.75\,P+0.85\,SS+1.78\,E+0.03\,CA+1.13\,H+0.06\,PM+0.61\,A)} \quad (4.8)$$

9) Model *Missing Value 80%*

$$\pi(x) = \frac{\exp(1.17-0.64\,D-0.44\,P-0.52\,SS+0.02\,CH+0.93\,E-0.64\,CA+0.34\,H+0.62\,PS+0.64\,A-0.77\,I-0.81\,ND)}{1+\exp(1.17-0.64\,D-0.44\,P-0.52\,SS+0.02\,CH+0.93\,E-0.64\,CA+0.34\,H+0.62\,PS+0.64\,A-0.77\,I-0.81\,ND)} \quad (4.9)$$

10) Model *Missing Value 90%*

$$\pi(x) = \frac{\exp\,(0.78-1.25\,CS+0.12\,P+1.04\,A+1.15\,PS-0.14\,D-1.06\,E+0.46\,J+0.45\,OP-1.01\,T+0.02\,CA)}{1+\exp\,(0.78-1.25\,CS+0.12\,P+1.04\,A+1.15\,PS-0.14\,D-1.06\,E+0.46\,J+0.45\,OP-1.01\,T+0.02\,CA)} \quad (4.10)$$

Interpretasi koefisien model regresi logistik dapat dilihat dari nilai rasio odds. Sebagai contoh nilai rasio odds untuk atribut checking_status pada dataset lengkap adalah sebesar 0.25 artinya peluang debitur diklasifikasikan sebagai debitur baik adalah 0.25 lebih kecil apabila nilai checking_statusnya adalah BC dibandingkan dengan peluangnya apabila nilai checking_statusnya adalah AD. Nilai rasio odds untuk atribut credit_amount pada dataset *missing value 10%* adalah sebesar 1.14 artinya peluang debitur diklasifikasikan sebagai debitur baik adalah 1.14 lebih tinggi apabila nilai credit_amountnya >= 1306 dibandingkan dengan peluangnya apabila nilai credit_amountnya <1306. Interpretasi yang sama untuk seluruh atribut di setiap dataset.

Pengklasifikasian debitur menjadi debitur baik dan buruk dapat dilakukan dengan memasukkan seluruh nilai atribut pada model yang terbentuk. Nilai π(x) akan didapatkan dengan memasukkan seluruh nilai atribut kedalam model. Debitur diklasifikasikan sebagai debitur baik jika nilai π(x) >= 0.5 sebaliknya debitur diklasifikasikan sebagai debitur buruk jika nilai π(x) < 0.5 [19].

Sebagai contoh jika ada debitur dengan profil checking_status = *no checking* (D), duration = 20 bulan, credit_amount = 3000, num_dependents = 1 (A), other_payment_plans = *none* (C), dan residence_since = 2. Maka untuk model data missing *value 50%* nilai π(x) adalah sebagai berikut :

$$\pi(x) = \frac{\exp(0.99 + 0.07\,(1) - 0.99\,(0) - 0.06\,(0) + 0.65\,(0) - 1.04\,(1) + 0.26\,(0))}{1 + \exp(0.99 + 0.07\,(1) - 0.99\,(0) - 0.06\,(0) + 0.65\,(0) - 1.04\,(1) + 0.26\,(0))}$$

$$= \frac{0.02}{1 + 0.02}$$

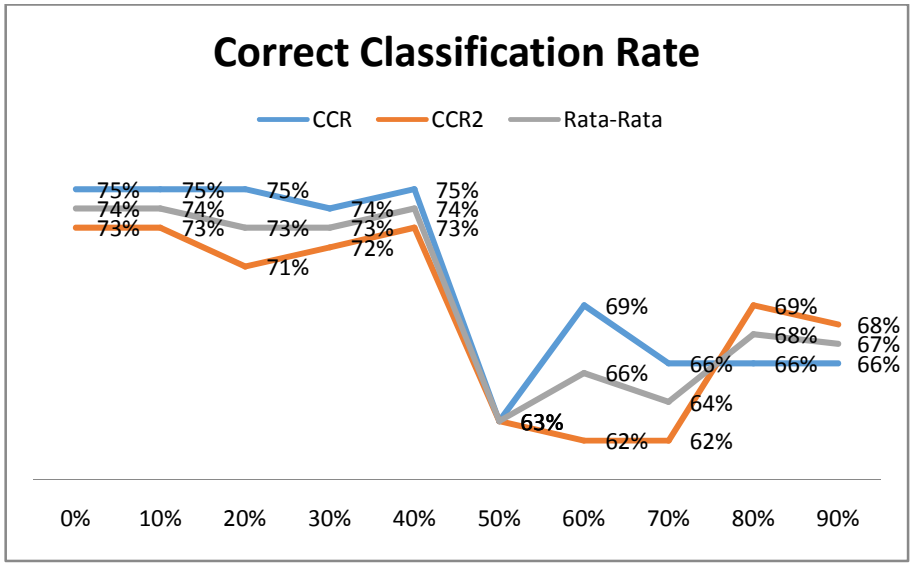
$$= 0.019$$

Karena nilai π(x) sebesar 0.019 lebih kecil dari 0.5 maka debitur tersebut dikategorikan sebagai debitur buruk.

Pembentukan model diulang sebanyak 2 kali untuk melihat pola nilai CCR untuk seluruh dataset. Nilai CCR percobaan pertama, percobaan kedua dan nilai rata-rata dari setiap dataset dapat dilihat pada tabel IV-15 dan pada gambar IV-11 dalam bentuk grafik.

Tabel IV- 15 Nilai CCR Regresi Logistik

Dataset	**CCR**	**CCR2**	**Rata-Rata**
Data Lengkap	75%	73%	74%
Missing Value 10%	75%	73%	74%
Missing Value 20%	75%	71%	73%
Missing Value 30%	74%	72%	73%
Missing Value 40%	75%	73%	74%
Missing Value 50%	63%	63%	63%
Missing Value 60%	69%	62%	66%
Missing Value 70%	66%	62%	64%
Missing Value 80%	66%	69%	68%
Missing Value 90%	66%	68%	67%



Gambar IV-11 Nilai CCR Regresi Logistik

Dari hasil tersebut terlihat bahwa nilai CCR percobaan pertama, percobaan kedua dan nilai rata-ratanya menunjukan pola yang sama. Nilai CCR untuk dataset dengan *missing value* 10% - 40% tidak berbeda signifikan dengan nilai CCR untuk data lengkap. Nilai CCR menurun pada dataset dengan *missing value* 50% - 90%. Rata – rata nilai CCR untuk dataset dengan *missing value* 10% - 40% adalah sebesar 73.5%, hanya selisih 0.5% dari nilai rata-rata CCR untuk data lengkap. Sedangkan rata - rata nilai CCR untuk dataset dengan *missing value* 50% - 90% adalah sebesar 65.6%, selisih sebesar 8.4% dari nilai rata-rata CCR untuk data lengkap. Output dari analisis regresi logistik dapat dilihat pada lampiran 2.

Model dari analisis regresi yang dihasilkan kemudian dapat digunakan pada saat ada debitur yang mengajukan aplikasi kredit. Atribut dari profil nasabah yang berpengaruh terhadap peubah respon dimasukkan kedalam model untuk

mendapatkan nilai $\pi(x)$. Jika nilai $\pi(x) >= 0.5$ maka debitur akan diklasifikasikan sebagai debitur baik. Sebaliknya, jika nilai $\pi(x) < 0.5$ maka debitur akan diklasifikasikan sebagai debitur buruk. Nilai rasio odds dapat menjadi informasi tambahan untuk melihat seberapa besar pengaruh dari masing-masing atribut. Pengklasifikasian debitur baik dan debitur buruk ini akan menjadi pertimbangan manajemen untuk menerima atau menolak pengajuan kredit dari debitur.