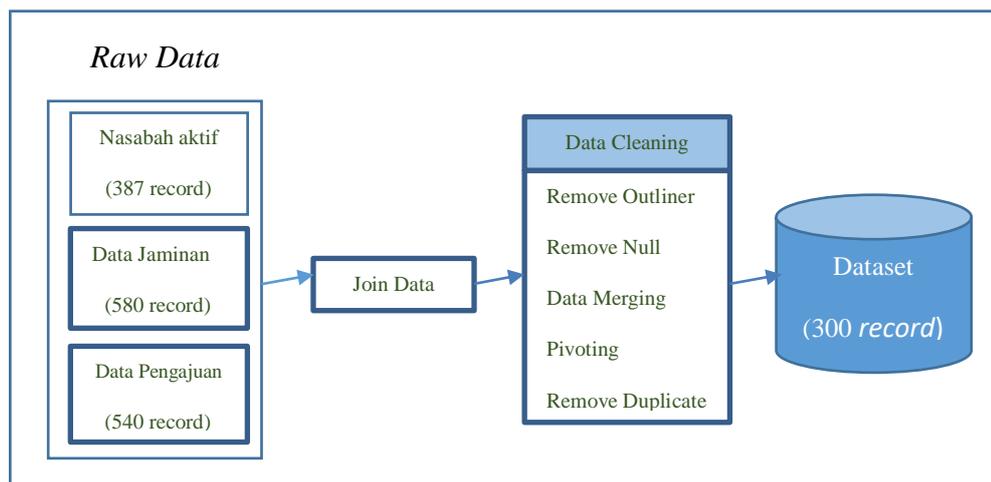


BAB IV

PEMBAHASAN HASIL PENELITIAN

4.1 Akuisisi Data

Akuisisi data merupakan sistem yang digunakan untuk mengambil, mengumpulkan dan menyiapkan data yang sedang berjalan, kemudian data tersebut diolah lebih lanjut dalam komputer untuk keperluan tertentu [2]. Sistem akuisisi data dapat didefinisikan sebagai suatu sistem yang berfungsi untuk mengambil, mengumpulkan dan menyiapkan data, hingga memprosesnya untuk menghasilkan data yang dikehendaki. Jenis serta metode yang di pilih pada umumnya bertujuan untuk menyederhanakan setiap langkah yang dilaksanakan pada keseluruhan proses. Suatu sistem akuisisi data pada umumnya dibentuk sedemikian rupa sehingga sistem tersebut berfungsi untuk mengambil, mengumpulkan dan menyimpan data dalam bentuk yang siap untuk diproses lebih lanjut.



Gambar 4.1 Proses Akuisisi Data

Gambar 4.1 menyajikan proses akuisisi data penelitian sehingga diperoleh dataset penelitian. Beberapa data sumber yang ditarik dari pusat penyimpanan data yaitu: (1) data nasabah aktif, (2) data jaminan nasabah, (3) data pengajuan. Seluruh data sumber selanjutnya digabungkan menjadi satu dataset dan dilakukan proses pembersihan data. Proses pembersihan data diperlukan untuk memastikan dataset yang diperlukan sudah sesuai. Berikut tahapan proses data *cleaning*:

1. *Remove outlier*, untuk menghilangkan data yang memiliki nilai yang salah dan diluar jangkauan nilai yang wajar.
2. *Remove null*, untuk menghilangkan record atau field yang kosong
3. *Data merging*, untuk menggabungkan beberapa record dengan data yang memiliki arti yang sama.
4. *Pivoting*, untuk mengubah dimensi data nasabah dari baris ke kolom.
5. *Remove duplicate*, untuk menghilangkan data ganda yang memiliki nilai dan arti yang sama.

Proses akuisisi data menghasilkan *dataset* penelitian sebanyak 300 *record* dimana terdapat 239 disetujui dan 61 ditolak. Pemilihan data dalam penelitian ini dituangkan dalam pembuatan model sistem. Model sistem yang akan dibuat memiliki batasan-batasan sebagai berikut:

1. Perancangan yang dibuat dengan menggunakan penalaran *data* dengan menggunakan metode *Naïve Bayes* dan *Decision Tree*.
2. Pembuatan aturan dalam basis pengetahuan dibantu oleh Kepala HRD, Kepala Marketing Manager dan Kepala GA di PT Batavia Prosperindo Finance Tbk.

4.2 Seleksi Atribut

Pada tahap awal penelitian, seleksi terhadap atribut evaluasi penentuan pengajuan pinjaman kredit pada Tabel 4.1 berikut dilakukan untuk mengurangi atribut awal. Pengurangan atribut dilakukan berdasarkan nilai *information gain* dari masing-masing atribut terhadap *entropi* kelas targetnya. Semakin tinggi nilai *information gain* dapat dikatakan bahwa atribut tersebut semakin penting atau semakin mempengaruhi keputusan kelas targetnya.

Tabel 4.1 Atribut Data Penelitian Berdasarkan *Information Gain*

No.	Atribut	Tipe Atribut	Nilai Atribut
1.	<i>Gender</i>	Binomial	Pria Wanita
2.	<i>Age</i>	Binomial	<=40 tahun >40 tahun
3.	Pendidikan_terakhir	Polinomial	Magister Sarjana Diploma SMA/SMK SMP SD
4.	Status	Polinomial	Belum Menikah Menikah Janda/Duda
5.	Jumlah_tanggungan	Binomial	<=2 >2
6.	Status_tempat_tinggal	Polinomial	Sewa Milik Sendiri Milik Keluarga Rumah Dinas
7.	Pekerjaan	Polinomial	PNS Wiraswasta TNI/Polri Guru/Dosen Petani Lain-lain
8.	Penghasilan_per_tahun	Binomial	<=45 juta >45 juta
9.	Nilai_jaminan	Binomial	<100 juta >100 juta
10.	Tujuan_Pinjaman	Binomial	Modal Usaha Kendaraan
11.	Pinjaman_max	Binomial	<=60 juta >60 juta
12.	Hasil <i>Survey</i>	Binomial	Disetujui(<i>Approval</i>) Ditolak (<i>Rejected</i>)

4.3 Pendefinisian Atribut Kelengkapan

Penentuan atribut-atribut yang digunakan dilakukan setelah melakukan wawancara dengan bagian terkait mengenai atribut apa saja yang paling penting digunakan untuk evaluasi nasabah yang melakukan pengajuan pinjaman. Hasilnya

ditetapkan 11 buah atribut data pengajuan yaitu: *gender*, *age*, *pendidikan_terakhir*, *status*, *jumlah_tanggungan*, *status_tempat_tinggal*, *pekerjaan*, *penghasilan_perthn*, *nilai_jaminan*, *tujuan_pinjaman*, dan *pinjaman_max*. Atribut-atribut ini menjadi *input* pada perancangan *prototype*, sementara untuk *output*-nya adalah Disetujui atau Ditolak.

Hal yang pertama dilakukan dalam penelitian ini adalah pemakaian dua algoritma yaitu *Naïve Bayes* dan *Decision Tree* untuk membentuk sebuah model. Model yang dihasilkan akan digunakan untuk mengetahui pola kelayakan pemberian kredit kepada calon nasabah yang mengajukan kredit dengan melihat data-data atribut yang telah ditentukan sebelumnya.

Pada saat penelitian ini dilakukan proses validasi untuk menemukan, dan mengkonversi data agar dapat digunakan dalam metode algoritma *data mining* dan memperoleh akurasi serta *performance* yang baik. Dalam *dataset* yang akan digunakan ini, validasi data yang digunakan adalah dengan menghapus data yang tidak lengkap atau kosong yang tidak memiliki nilai (*null*). Setelah itu dilakukan seleksi atribut untuk memilih atribut mana saja yang dibutuhkan dari dataset yang digunakan dalam proses menganalisis kelayakan pemberian kredit kepada calon nasabah. Berikut diberikan beberapa data training seperti tabel 4.2

Tabel 4.2 Sampel Data Training

No.	No Kontrak	Nama	Age	Gender	Pend.ter akhir	Status	Jumlah Tangg.Kel.	Kepemilikan Rumah	Pekerjaan	Penghasilan/tahun	Nilai Jaminan	Tujuan Pinjaman	Pinjaman max	HASIL
1	074372180004	TASRIP	KP	L	SMA	Menikah	K	Milik Sendiri	Wiraswasta	K	K	Kendaraan	B	Disetujui
2	074372180005	EKO SANTOSO	P	L	SMA	Menikah	K	Milik Sendiri	Wiraswasta	K	K	Modal Usaha	K	Disetujui
3	074372180010	YENI EFTIANINGSIH	KP	P	SMA	Menikah	K	Milik Sendiri	Wiraswasta	K	B	Kendaraan	K	Disetujui
4	074372180011	RESTU NURUL FATIMAH	KP	P	S1	Menikah	K	Milik Keluarga	Guru/Dosen	B	B	Kendaraan	B	Disetujui
5	074372180012	AGUS SURAHMAN	P	L	SMA	Menikah	B	Milik Sendiri	PNS	K	K	Kendaraan	B	Disetujui
6	074372180016	JAENURI	KP	L	SMA	Menikah	K	Milik Sendiri	Wiraswasta	B	B	Kendaraan	B	Disetujui
7	074372190002	ISRIYADI	P	L	SMA	Menikah	K	Milik Sendiri	Wiraswasta	K	K	Kendaraan	K	Disetujui
8	074372190004	CIPTO SUBAGIO	P	L	SMA	Menikah	K	Milik Sendiri	Wiraswasta	B	B	Kendaraan	B	Disetujui
9	074372190006	SUNGATNO	P	L	SMP	Menikah	K	Milik Sendiri	Petani	K	B	Kendaraan	B	Disetujui
10	074372190007	DEDE MAHENDRA (Nasabah GP)	KP	L	SMA	Menikah	K	Milik Keluarga	Petani	B	K	Kendaraan	B	Disetujui
11	074372190008	HENDRA RUDHIANATA	KP	L	SMA	Menikah	B	Milik Sendiri	Petani	K	K	Modal Usaha	K	Disetujui
12	074372190009	ISLAH *(Nasabah GP)	KP	L	SMA	Menikah	K	Milik Sendiri	Wiraswasta	B	B	Kendaraan	B	Disetujui
13	074372190012	YENIAR EFENDI	KP	L	S2	Menikah	B	Milik Keluarga	PNS	K	K	Modal Usaha	B	Ditolak
14	074372190013	DARMAWAN	KP	L	SMA	Menikah	B	Milik sendiri	Wiraswasta	K	K	Kendaraan	B	Ditolak
15	074372190014	NOVRIADI	KP	L	SMA	Menikah	B	Milik Sendiri	Wiraswasta	K	K	Modal Usaha	K	Ditolak
16	074372190016	RIZKI INDRA SETIAWAN (Nasabah GP)	KP	L	SMA	Menikah	B	Milik Keluarga	Wiraswasta	B	B	Kendaraan	B	Ditolak
17	074372190017	MUKHLISH *(Nasabah GP)	KP	L	SMP	Duda	B	Milik Sendiri	Wiraswasta	B	K	Modal Usaha	K	Ditolak
18	074372190018	RINI PUJI LESTARI	KP	P	SMP	Menikah	B	Milik Sendiri	Lain-lain	K	B	Kendaraan	B	Ditolak
19	074372190019	TUNUT	P	L	SMA	Belum Menikah	K	Milik Keluarga	Petani	B	B	Modal Usaha	B	Ditolak
20	074372190021	NANANG PURWANTO (Nasabah GP)	P	L	SMA	Menikah	K	Rumah Dinas	TNI/POLRI	B	K	Kendaraan	K	Ditolak
21	074372190025	DEDI ARPANI	P	L	SMA	Belum Menikah	K	Sewa	Wiraswasta	B	B	Kendaraan	B	Disetujui
22	074372190031	MIRHAN	P	L	SMA	Menikah	B	Sewa	Wiraswasta	K	K	Kendaraan	B	Disetujui
23	074372190032	NOPRIANDI	P	L	SMA	Menikah	K	Milik Keluarga	Wiraswasta	K	K	Kendaraan	B	Disetujui
24	074372190034	ZULYADI MUNZIRI	P	L	SMA	Menikah	B	Milik Sendiri	Wiraswasta	K	K	Kendaraan	B	Disetujui
25	074372190035	SYAHRUDDIN AL JAWI	P	L	SMA	Menikah	K	Milik Sendiri	Petani	K	K	Kendaraan	K	Disetujui
26	074372190036	MIMBAR WIBOWO (Nasabah GP)	P	L	SMA	Menikah	K	Milik Sendiri	Petani	B	B	Kendaraan	B	Disetujui
27	074372190037	FIRDAUS SE	P	L	SMA	Menikah	B	Milik Sendiri	Petani	B	B	Kendaraan	B	Disetujui
28	074372190038	CATUR DWIYATNO	P	L	SMA	Menikah	K	Milik Sendiri	PNS	B	B	Kendaraan	B	Disetujui
29	074372190041	HOTIB	KP	L	SMA	Menikah	K	Milik Sendiri	Petani	K	K	Kendaraan	K	Disetujui

Tabel 4.2 (lanjutan)

No.	No Kontrak	Nama	Age	Gender	Pend.ter akhir	Status	Jumlah Tangg.Kel.	Kepemilikan Rumah	Pekerjaan	Penghasilan/tahun	Nilai Jaminan	Tujuan Pinjaman	Pinjaman max	HASIL
30	074372190042	ELIA NOVITA WATI *(Nasabah GP)	KP	P	SMA	Menikah	K	Milik Sendiri	Petani	B	K	Kendaraan	B	Ditolak
31	074372190043	ADYANTO *(Nasabah GP)	P	L	SMA	Menikah	K	Milik Sendiri	Petani	K	K	Kendaraan	K	Disetujui
32	074372190046	TONI	P	L	S1	Menikah	K	Milik Sendiri	Wiraswasta	K	K	Kendaraan	K	Disetujui
33	074372190047	IMAM SOPINGI	P	L	S1	Menikah	K	Milik Sendiri	PNS	B	B	Kendaraan	B	Disetujui
34	074372190049	PONDI	P	L	SMA	Menikah	B	Milik Sendiri	Petani	B	B	Kendaraan	B	Disetujui
35	074372190050	ANTON PRAYOGO	P	L	SMA	Belum Menikah	K	Milik Sendiri	Petani	B	B	Kendaraan	B	Disetujui
36	074372190051	YULI ASMARA	P	P	S1	Belum Menikah	K	Sewa	Wiraswasta	K	K	Kendaraan	K	Disetujui
37	074372190053	MALA SARI	P	P	DIPLOMA	Belum Menikah	K	Milik Sendiri	PNS	K	K	Kendaraan	B	Disetujui
38	074372190054	JUNAIDI	P	L	SMA	Belum Menikah	K	Milik Keluarga	Wiraswasta	K	B	Kendaraan	B	Disetujui
39	074372190056	BAMBANG GUNAWAN	P	L	SMP	Menikah	B	Milik Keluarga	Wiraswasta	K	K	Kendaraan	K	Disetujui
40	074372190058	AMRAN SE	P	L	SMA	Menikah	K	Milik Keluarga	Wiraswasta	B	B	Modal Usaha	B	Disetujui
41	074372190059	SURA ATMAJA	P	L	SMA	Menikah	K	Milik Keluarga	Wiraswasta	B	B	Modal Usaha	B	Disetujui
42	074372190063	FITRI LESTARI	P	P	SMA	Menikah	K	Milik Keluarga	Wiraswasta	B	B	Kendaraan	B	Disetujui
43	074372190065	ROSIHUN *(Nasabah GP)	P	L	S1	Menikah	K	Milik Sendiri	Lain-lain	K	K	Modal Usaha	K	Disetujui
44	074372190066	KODAR AMINUDIN	KP	L	DIPLOMA	Menikah	K	Milik Sendiri	Lain-lain	K	K	Modal Usaha	K	Disetujui
45	074372190067	MASKUR	KP	L	S1	Menikah	B	Milik Sendiri	Wiraswasta	K	K	Modal Usaha	B	Disetujui
46	074372190071	SYAHRIL ANWAR	P	L	S1	Menikah	K	Milik Sendiri	Petani	K	K	Kendaraan	K	Disetujui
47	074372190072	IKA RACHIMAH	P	P	S1	Menikah	B	Milik Sendiri	Wiraswasta	B	B	Kendaraan	B	Ditolak
48	074372190074	ZIKRANI	P	L	SMA	Menikah	B	Milik Sendiri	Wiraswasta	B	B	Kendaraan	B	Disetujui
49	074372190075	NURHADI	P	L	SMA	Menikah	B	Milik Sendiri	Wiraswasta	B	K	Modal Usaha	B	Disetujui
50	074372190076	DEDI MARGONO	P	L	SMA	Belum Menikah	K	Milik Sendiri	Wiraswasta	B	B	Kendaraan	B	Ditolak
51	074372190079	KUSDAARMAJI	P	L	SMA	Belum Menikah	K	Milik Sendiri	Petani	K	K	Kendaraan	K	Ditolak
52	074372190080	BURHANUDIN RAIS	P	L	SMA	Belum Menikah	K	Milik Sendiri	Wiraswasta	K	K	Kendaraan	K	Ditolak
53	074372190082	HARDI YANTO *(Nasabah GP)	P	L	DIPLOMA	Belum Menikah	K	Milik Sendiri	Petani	B	K	Kendaraan	B	Ditolak
54	074372190084	ENDRIYANI	P	P	DIPLOMA	Belum Menikah	B	Sewa	Wiraswasta	B	B	Modal Usaha	B	Ditolak
55	074372190086	EDI PRASTYO *(Nasabah GP)	P	L	S1	Belum Menikah	K	Sewa	Petani	B	K	Modal Usaha	K	Ditolak
56	074372190087	JAMRONI	KP	L	SMP	Duda	K	Milik Sendiri	Wiraswasta	B	B	Kendaraan	B	Ditolak
57	074372190088	SUSANTO *(Nasabah GP)	KP	L	SMP	Menikah	K	Milik Sendiri	Petani	K	B	Kendaraan	B	Ditolak
58	074372190090	HARDIONO	KP	L	SD	Duda	K	Milik Sendiri	Petani	B	K	Kendaraan	B	Ditolak
59	074372190091	ARI JOHARI	P	L	SMA	Menikah	B	Milik Sendiri	Petani	B	K	Kendaraan	B	Ditolak
60	074372190092	IBNU SAYAF *(Nasabah GP)	P	L	S1	Menikah	K	Milik Sendiri	PNS	K	K	Kendaraan	K	Ditolak
...
300	074372200111	MUHAMMAD RAZIK	P	L	S1	Menikah	K	Milik Sendiri	Wiraswasta	B	B	Kendaraan	B	Disetujui

4.3.1 Pengujian Klasifikasi dengan Algoritma *Decision Tree*

Pada penelitian ini akan dilakukan perhitungan manual pada *data sample*. Perhitungan dilakukan pada 30 data sampel dari 300 *data sample* berdasarkan data yang terdapat pada tabel 4.2.

Berdasarkan tabel 4.2 dibuat tabel keputusan untuk menentukan kelayakan pemberian kredit kepada calon nasabah dengan melihat *atribut dataset* yang telah ditentukan. Langkah pertama menghitung total *entropi* dari jumlah kasus seluruhnya, jumlah keputusan “Disetujui” maupun “Ditolak”, dengan persamaan:

$$Entropy\ Total = \sum_{i=1}^n -p_i \log_2(p_i)$$

$$Entropy\ Total = \left(-\frac{21}{30} \log_2\left(\frac{21}{30}\right)\right) + \left(-\frac{9}{30} \log_2\left(\frac{9}{30}\right)\right) = 0.88129$$

Sehingga dari perhitungan diatas didapat total *entropi* = 0.88129

Kemudian menghitung total *entropi* dan *gain* dari semua kasus yang terbagi berdasarkan atribut “*age*”, “*gender*”, “*pend_terakhir*”, “*status*”, “*juml_tangg_kel*”, “*status_tempat_tinggal*”, “*pekerjaan*”, “*penghasilan_per_tahun*”, “*nilai_jaminan*”, “*tujuan_pinjaman*”, dan “*pinjaman_max*”.

1. Atribut *age*

Untuk atribut *age* pada tabel 4.2 terdapat 30 kasus dengan 15 nilai kategori ‘<=40’; dan 15 nilai kategori ‘>40’, dengan perincian untuk nilai kategori ‘<=40’ 13 ‘disetujui’ dan 2 ‘ditolak’, untuk nilai kategori ‘>40’ 8 ‘disetujui’ dan 7 ‘ditolak’, sehingga nilai *entropi age* dapat dihitung untuk setiap kasus sebagai berikut:

$$Entropy\ (" \leq 40") = \left(-\frac{13}{15} \log_2\left(\frac{13}{15}\right)\right) + \left(-\frac{2}{15} \log_2\left(\frac{2}{15}\right)\right) = 0.56536$$

$$Entropy\ (" > 40") = \left(-\frac{8}{15} \log_2\left(\frac{8}{15}\right)\right) + \left(-\frac{7}{15} \log_2\left(\frac{7}{15}\right)\right) = 0.9967$$

Kemudian menghitung nilai *gain age* sebagai berikut:

$$Gain_{(age)} = Entropy\ Total - \left(\frac{15}{30} \times Entropy\ (" \leq 40")\right) +$$

$$\frac{15}{30} \times Entropy("> 40")$$

$$Gain_{(age)} = 0.88129 - \left(\frac{15}{30} \times (0.56536) + \frac{15}{30} \times (0.9967)\right)$$

$$Gain_{(age)} = 0.10026$$

2. Atribut *gender*

Untuk atribut *gender* pada tabel 4.2 terdapat 30 kasus dengan 26 nilai 'laki-laki'; dan 4 nilai 'perempuan', dengan perincian untuk nilai 'laki-laki' 19 'disetujui' dan 7 'ditolak', untuk nilai 'perempuan' 2 'disetujui' dan 2 'ditolak', sehingga nilai *entropi gender* dapat dihitung untuk setiap kasus sebagai berikut:

$$Entropy ('laki-laki') = \left(-\frac{19}{26} \log_2 \left(\frac{19}{26}\right)\right) + \left(-\frac{7}{26} \log_2 \left(\frac{7}{26}\right)\right) = 0.84041$$

$$Entropy ('perempuan') = \left(-\frac{2}{4} \log_2 \left(\frac{2}{4}\right)\right) + \left(-\frac{2}{4} \log_2 \left(\frac{2}{4}\right)\right) = 1$$

Kemudian menghitung nilai *gain* jenis *gender* sebagai berikut:

$$Gain_{(gender)} = Entropy Total - \left(\frac{26}{30} \times Entropy('laki. laki') + \right.$$

$$\left. \frac{4}{30} \times Entropy('perempuan')\right)$$

$$Gain_{(gender)} = 0.88129 - \left(\frac{26}{30} \times (0.84041) + \frac{4}{30} \times (1)\right)$$

$$Gain_{(gender)} = 0.01965$$

3. Atribut *pend_terakhir*

Untuk atribut *pend_terakhir* pada tabel 4.2 terdapat 30 kasus dengan 3 nilai 'SMP'; 25 nilai 'SMA'; 1 nilai 'S1'; dan 1 nilai 'S2, dengan perincian untuk nilai 'SMP' 1 'disetujui' 2 'ditolak', untuk nilai 'SMA' 19 'disetujui' 6 'ditolak', untuk nilai 'S1' 1 'disetujui' 0 'ditolak', untuk

nilai 'S2' 0 'disetujui' 1 'ditolak' sehingga nilai entropi *pend_terakhir* dapat dihitung untuk setiap kasus sebagai berikut:

$$Entropy ('SMP') = \left(-\frac{1}{3} \log_2 \left(\frac{1}{3}\right)\right) + \left(-\frac{2}{3} \log_2 \left(\frac{2}{3}\right)\right) = 0.91825$$

$$Entropy ('SMA') = \left(-\frac{19}{25} \log_2 \left(\frac{19}{25}\right)\right) + \left(-\frac{6}{25} \log_2 \left(\frac{6}{25}\right)\right) = 0.79503$$

$$Entropy ('S1') = \left(-\frac{1}{1} \log_2 \left(\frac{1}{1}\right)\right) + \left(-\frac{0}{1} \log_2 \left(\frac{0}{1}\right)\right) = 0$$

$$Entropy ('S2') = \left(-\frac{0}{1} \log_2 \left(\frac{0}{1}\right)\right) + \left(-\frac{1}{1} \log_2 \left(\frac{1}{1}\right)\right) = 0$$

Kemudian menghitung nilai *gain* *pend_terakhir* sebagai berikut:

$$Gain_{(pend_terakhir)} = Entropy Total - \left(\frac{3}{30} \times Entropy("SMP") + \left(\frac{25}{30} \times Entropy("SMA") + \frac{1}{30} \times Entropy("S1") + \frac{1}{30} \times Entropy("S2")\right)\right)$$

$$Gain_{(pend_terakhir)} = 0.88129 - \left(\frac{3}{30} \times (0.91825) + \frac{25}{30} \times (0.79503) + \left(\frac{1}{30} \times (0) + \left(\frac{1}{30} \times (0)\right)\right)\right)$$

$$Gain_{(pend_terakhir)} = 0.12696$$

4. Atribut status

Untuk atribut status pada tabel 4.2 terdapat 30 kasus dengan 2 nilai 'belum_menikah', 27 nilai 'menikah', dan 1 nilai 'duda' dengan perincian untuk nilai 'belum_menikah' 1 'disetujui' dan 1 'ditolak', untuk nilai 'menikah' 20 'disetujui' dan 7 'ditolak', untuk nilai 'duda' 0 'disetujui' dan 1 'ditolak', sehingga nilai *entropi* status dapat dihitung untuk setiap kasus sebagai berikut:

$$Entropy ('belum_menikah') = \left(-\frac{1}{2} \log_2 \left(\frac{1}{2}\right)\right) + \left(-\frac{1}{2} \log_2 \left(\frac{1}{2}\right)\right) = 1$$

$$Entropy ('menikah') = \left(-\frac{20}{27} \log_2 \left(\frac{20}{27}\right)\right) + \left(-\frac{7}{27} \log_2 \left(\frac{7}{27}\right)\right) = 0.82557$$

$$Entropy ('duda') = \left(-\frac{0}{1} \log_2 \left(\frac{0}{1}\right)\right) + \left(-\frac{1}{1} \log_2 \left(\frac{1}{1}\right)\right) = 0$$

Kemudian menghitung nilai *gain* status sebagai berikut:

$$Gain_{(status)} = Entropy Total - \left(\frac{2}{30} \times Entropy('belum_menikah') + \frac{27}{30} \times Entropy('menikah') + \frac{1}{30} \times Entropy('duda')\right)$$

$$Gain_{(status)} = 0.88129 - \left(\frac{2}{30} \times (1) + \frac{27}{30} \times (0.82557) + \frac{1}{30} \times (0)\right)$$

$$Gain_{(status)} = 0.07157$$

5. Atribut *jumlah_tangg_kel*

Untuk atribut *jumlah_tangg_kel* pada tabel IV-2 terdapat 30 kasus dengan 19 nilai kategori ' ≤ 2 '; dan 11 nilai kategori ' > 2 ', dengan perincian untuk nilai kategori ' ≤ 2 ' 16 'disetujui' dan 3 'ditolak', untuk nilai kategori ' > 2 ' 5 'disetujui' dan 6 'ditolak', sehingga nilai entropi *jumlah_tangg_kel* dapat dihitung untuk setiap kasus sebagai berikut:

$$Entropy (" \leq 2 ") = \left(-\frac{16}{19} \log_2 \left(\frac{16}{19}\right)\right) + \left(-\frac{3}{19} \log_2 \left(\frac{3}{19}\right)\right) = 0.62922$$

$$Entropy (" > 2 ") = \left(-\frac{5}{11} \log_2 \left(\frac{5}{11}\right)\right) + \left(-\frac{6}{11} \log_2 \left(\frac{6}{11}\right)\right) = 0.99402$$

Kemudian menghitung nilai *gain* *jumlah_tangg_kel* sebagai berikut:

$$Gain_{(jumlah_tangg_kel)} = Entropy Total - \left(\frac{19}{30} \times Entropy(" \leq 2 ") + \frac{11}{30} \times Entropy(" > 2 ") \right)$$

$$Gain_{(jumlah_tangg_kel)} = 0.88129 - \left(\frac{19}{30} \times (0.62922) + \frac{11}{30} \times (0.99402)\right)$$

$$Gain_{(jumt_tang_kel)} = 0.11831$$

6. Atribut status_tempat_tinggal

Untuk atribut status_tempat_tinggal pada tabel IV-2 terdapat 30 kasus dengan 2 nilai 'sewa'; 21 nilai 'milik_sendiri', 6 nilai 'milik_keluarga', dan 1 nilai 'dinas', dengan perincian untuk nilai kategori 'sewa' semuanya "disetujui", untuk nilai kategori 'milik_sendiri' 16 'disetujui' dan 5 'ditolak', untuk nilai kategori 'milik_keluarga' 3 'disetujui' dan 3 'ditolak', dan untuk nilai kategori 'dinas', 0 'disetujui' dan 1 'ditolak', sehingga nilai entropi status_tempat_tinggal dapat dihitung untuk setiap kasus sebagai berikut:

$$Entropy('sewa') = \left(-\frac{2}{2} \log_2 \left(\frac{2}{2}\right)\right) + \left(-\frac{0}{2} \log_2 \left(\frac{0}{2}\right)\right) = 0$$

$$Entropy('milik_sendiri') = \left(-\frac{16}{21} \log_2 \left(\frac{16}{21}\right)\right) + \left(-\frac{5}{21} \log_2 \left(\frac{5}{21}\right)\right) = 0.79185$$

$$Entropy('milik_keluarga') = \left(-\frac{3}{6} \log_2 \left(\frac{3}{6}\right)\right) + \left(-\frac{3}{6} \log_2 \left(\frac{3}{6}\right)\right) = 1$$

$$Entropy('dinas') = \left(-\frac{0}{1} \log_2 \left(\frac{0}{1}\right)\right) + \left(-\frac{1}{1} \log_2 \left(\frac{1}{1}\right)\right) = 0$$

Kemudian menghitung nilai *gain* status_tempat_tinggal sebagai berikut:

$$Gain_{(status_temt_tinggal)} = Entropy\ Total - \left(\frac{2}{30} \times Entropy('sewa') +$$

$$\frac{21}{30} \times Entropy('milik_sendiri') +$$

$$\frac{6}{30} \times Entropy('milik_keluarga') + \frac{1}{30} \times Entropy('dinas'))$$

$$Gain_{(status_temt_tinggal)} = 0.88129 - \left(\frac{2}{30} \times (0) + \frac{21}{30} \times (0.79185) +$$

$$\left(\frac{6}{30} \times (1) + \left(\frac{1}{30} \times (0)\right)\right)$$

$$Gain_{(status_temt_tinggal)} = 0.12699$$

7. Atribut pekerjaan

Untuk atribut pekerjaan pada tabel 4.2 terdapat 30 kasus dengan 15 nilai kategori ‘wiraswasta’; dan 3 nilai kategori ‘PNS’; 9 nilai kategori ‘petani’; 1 nilai kategori ‘TNI/POLRI’, 1 nilai kategori ‘guru/dosen’, dan 1 nilai kategori ‘Lain-lain’, dengan perincian untuk nilai kategori ‘wiraswasta’, 11 “disetujui” dan 4 “ditolak”; untuk nilai kategori ‘PNS’ 2 “disetujui” dan 1 “ditolak”; untuk nilai kategori ‘petani’, 7 “disetujui” dan 2 “ditolak”; untuk nilai kategori ‘TNI/POLRI’ semuanya “ditolak”, untuk nilai kategori ‘guru/dosen semuanya “diterima”, untuk nilai kategori ‘lain-lain’ semuanya “ditolak”, sehingga nilai entropi pekerjaan dapat dihitung untuk setiap kasus sebagai berikut:

$$Entropy ('wiraswasta') = \left(-\frac{11}{15} \log_2 \left(\frac{11}{15}\right)\right) + \left(-\frac{4}{15} \log_2 \left(\frac{4}{15}\right)\right) = 0.83668$$

$$Entropy ('PNS') = \left(-\frac{2}{3} \log_2 \left(\frac{2}{3}\right)\right) + \left(-\frac{1}{3} \log_2 \left(\frac{1}{3}\right)\right) = 0.91825$$

$$Entropy ('petani') = \left(-\frac{7}{9} \log_2 \left(\frac{7}{9}\right)\right) + \left(-\frac{2}{9} \log_2 \left(\frac{2}{9}\right)\right) = 0.76412$$

$$Entropy ('tni_polri') = \left(-\frac{0}{1} \log_2 \left(\frac{0}{1}\right)\right) + \left(-\frac{1}{1} \log_2 \left(\frac{1}{1}\right)\right) = 0$$

$$Entropy ('guru') = \left(-\frac{1}{1} \log_2 \left(\frac{1}{1}\right)\right) + \left(-\frac{0}{1} \log_2 \left(\frac{0}{1}\right)\right) = 0$$

$$Entropy ('lain_lain') = \left(-\frac{0}{1} \log_2 \left(\frac{0}{1}\right)\right) + \left(-\frac{1}{1} \log_2 \left(\frac{1}{1}\right)\right) = 0$$

Kemudian menghitung nilai *gain* pekerjaan sebagai berikut:

$$\begin{aligned} Gain_{(pekerjaan)} = Entropy Total - & \left(\frac{15}{30} \times Entropy('wiraswasta') + \right. \\ & \frac{3}{30} \times Entropy('pns') + \frac{9}{30} \times Entropy('petani') + \\ & \frac{1}{30} \times Entropy('tni_polri') + \frac{1}{30} \times Entropy('guru') \\ & \left. + \frac{1}{20} \times Entropy('lain_lain')) \right) \end{aligned}$$

$$Gain_{(pekerjaan)} = 0.88129 - \left(\frac{15}{30} \times (0.83668) + \frac{3}{30} \times (0.91825) + \left(\frac{9}{30} \times (0.76412) + \left(\frac{1}{30} \times (0) + \left(\frac{1}{30} \times (0) + \left(\frac{1}{30} \times (0)\right)\right)\right)\right)\right)$$

$$Gain_{(pekerjaan)} = 0.11418$$

8. Atribut penghasilan_per_tahun

Untuk atribut penghasilan_per_tahun pada tabel IV-2 terdapat 30 kasus dengan 16 nilai kategori '<=45 juta'; dan 14 nilai kategori '>45 juta', dengan perincian untuk nilai kategori '<=45 juta' 12 'disetujui' dan 4 'ditolak', untuk nilai kategori '>45 juta' 9 'disetujui' dan 5 'ditolak', sehingga nilai *entropi* penghasilan_per_tahun dapat dihitung untuk setiap kasus sebagai berikut:

$$Entropy (" \leq 45 ") = \left(-\frac{12}{16} \log_2 \left(\frac{12}{16} \right) \right) + \left(-\frac{4}{16} \log_2 \left(\frac{4}{16} \right) \right) = 0.81128$$

$$Entropy (" > 45 ") = \left(-\frac{9}{14} \log_2 \left(\frac{9}{14} \right) \right) + \left(-\frac{5}{14} \log_2 \left(\frac{5}{14} \right) \right) = 0.94037$$

Kemudian menghitung nilai *gain* penghasilan_per_tahun sebagai berikut:

$$Gain_{(penghasilan_per_tahun)} = Entropy Total - \left(\frac{16}{30} \times Entropy (" \leq 45 ") + \right.$$

$$\left. \frac{14}{30} \times Entropy (" > 45 ") \right)$$

$$Gain_{(penghasilan_per_tahun)} = 0.88129 - \left(\frac{16}{30} \times (0.81128) + \frac{14}{30} \times (0.94037) \right)$$

$$Gain_{penghasilan_per_tahun} = 0.00977$$

9. Atribut nilai_jaminan

Untuk atribut nilai_jaminan pada tabel 4.2 terdapat 30 kasus dengan 17 nilai kategori '<=100 juta'; dan 13 nilai kategori '>100 juta', dengan perincian untuk nilai kategori '<=100' 11 'disetujui' dan 6 'ditolak', untuk nilai kategori '>100' 10 'disetujui' dan 3 'ditolak', sehingga nilai *entropi* nilai_jaminan dapat dihitung untuk setiap kasus sebagai berikut:

$$Entropy (" \leq 100 ") = \left(-\frac{11}{17} \log_2 \left(\frac{11}{17} \right) \right) + \left(-\frac{6}{17} \log_2 \left(\frac{6}{17} \right) \right) = 0.9367$$

$$Entropy (" > 100 ") = \left(-\frac{10}{13} \log_2 \left(\frac{10}{13} \right) \right) + \left(-\frac{3}{13} \log_2 \left(\frac{3}{13} \right) \right) = 0.7794$$

Kemudian menghitung nilai *gain* nilai_jaminan sebagai berikut:

$$Gain_{(nilai_jaminan)} = Entropy Total - \left(\frac{17}{30} \times Entropy (" \leq 100 ") + \frac{13}{30} \times Entropy (" > 100 ") \right)$$

$$Gain_{(nilai_jaminan)} = 0.88129 - \left(\frac{17}{30} \times (0.9367) + \frac{13}{30} \times (0.7794) \right)$$

$$Gain_{(nilai_jaminan)} = 0.01276$$

10. Atribut tujuan_pinjaman

Untuk atribut tujuan_pinjaman pada tabel 4.2 terdapat 30 kasus dengan 6 nilai kategori 'modal_usaha'; 24 nilai kategori 'kendaraan', dengan perincian untuk nilai kategori 'modal_usaha', 2 "disetujui" dan 4 "ditolak"; untuk nilai kategori 'kendaraan', 19 'disetujui' dan 5 "ditolak, sehingga nilai *entropi* tujuan_pinjaman dapat dihitung untuk setiap kasus sebagai berikut:

$$Entropy ('modal_usaha') = \left(-\frac{2}{6} \log_2 \left(\frac{2}{6} \right) \right) + \left(-\frac{4}{6} \log_2 \left(\frac{4}{6} \right) \right) = 0.91825$$

$$Entropy ('kendaraan') = \left(-\frac{19}{24} \log_2 \left(\frac{19}{24} \right) \right) + \left(-\frac{5}{24} \log_2 \left(\frac{5}{24} \right) \right) = 0.73831$$

Kemudian menghitung nilai *gain* tujuan_pinjaman sebagai berikut:

$$Gain_{(tujuan_pinjaman)} = Entropy Total - \left(\frac{6}{30} \times Entropy ('modal_usaha') + \frac{24}{30} \times Entropy ('kendaraan') \right)$$

$$Gain_{(tujuan_pinjaman)} = 0.88129 - \left(\frac{6}{30} \times (0.91825)\right) + \frac{24}{30} \times (0.73831)$$

$$Gain_{(tujuan_pinjaman)} = 0.10699$$

11. Atribut pinjaman_max

Untuk atribut pinjaman_max pada tabel 4.2 terdapat 30 kasus dengan 9 nilai kategori '<=60 juta'; dan 21 nilai kategori '>60 juta', dengan perincian untuk nilai kategori '<=60 juta' 6 'disetujui' dan 3 'ditolak', untuk nilai kategori '>60 juta' 15 'disetujui' dan 6 'ditolak', sehingga nilai *entropi* pinjaman_max dapat dihitung untuk setiap kasus sebagai berikut:

$$Entropy (" \leq 60 ") = \left(-\frac{6}{9} \log_2 \left(\frac{6}{9}\right)\right) + \left(-\frac{3}{9} \log_2 \left(\frac{3}{9}\right)\right) = 0.91825$$

$$Entropy (" > 60 ") = \left(-\frac{15}{21} \log_2 \left(\frac{15}{21}\right)\right) + \left(-\frac{6}{21} \log_2 \left(\frac{6}{21}\right)\right) = 0.86311$$

Kemudian menghitung nilai *gain* pinjaman_max sebagai berikut:

$$Gain_{(pinjaman_max)} = Entropy Total - \left(\frac{9}{30} \times Entropy (" \leq 60 ") + \frac{21}{30} \times Entropy (" > 60 ") \right)$$

$$Gain_{(pinjaman_max)} = 0.88129 - \left(\frac{9}{30} \times (0.91825)\right) + \frac{21}{30} \times (0.86311)$$

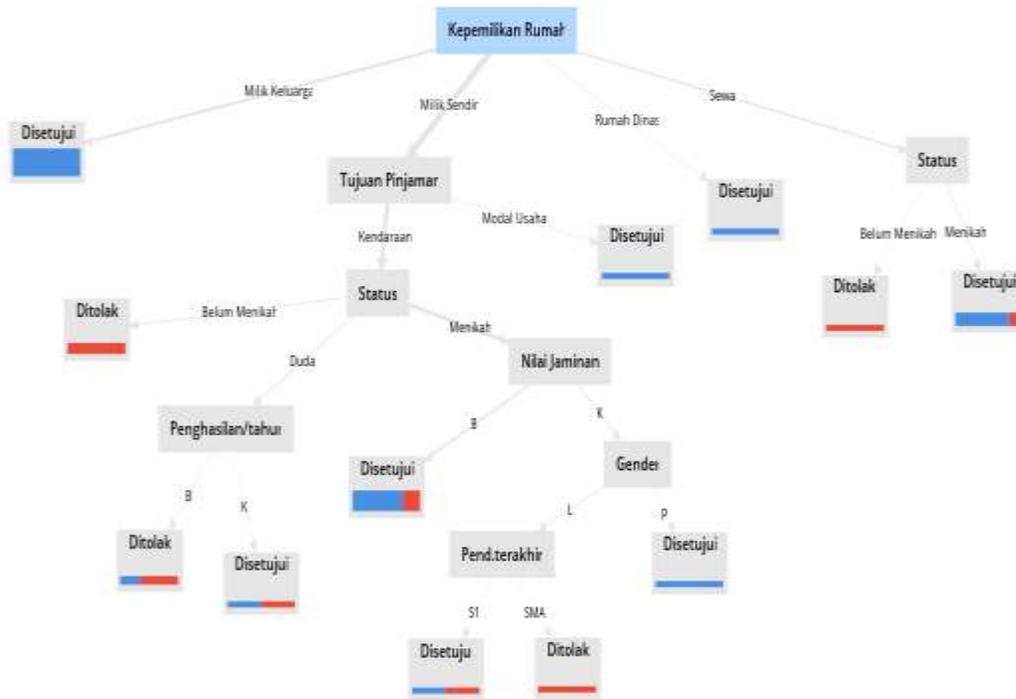
$$Gain_{pinjaman_max} = 0.00161$$

Dari hasil perhitungan total *entropi* dan *gain* setiap atribut diatas dapat dibuat dalam bentuk tabel 4.3 berikut ini:

Tabel 4.3 Hasil perhitungan *entropi* dan *gain* dari setiap atribut sampel

Atribut	Range	Jumlah kasus	Disetujui	Ditolak	Entropy	Gain
Total		30	21	9	0.88129	
Age						0.10026
	<=40	15	13	2	0.56536	
	>40	15	8	7	0.99672	
gender						0.01965
	L	26	19	7	0.84041	
	P	4	2	2	1	
Pend_terakhir						0.12696
	S2	1	0	1	0	
	S1	1	1	0	0	
	SMA	25	19	6	0.79503	
	SMP	3	1	2	0.91825	
status						0.07157
	belum menikah	2	1	1	1	
	menikah	27	20	7	0.82557	
	duda	1	0	1	0	
juml_tangg_kel						0.11831
	<=2	19	16	3	0.62922	
	>2	11	5	6	0.99402	
status_tempat_tinggal						0.12699
	sewa	2	2	0	0	
	milik sendiri	21	16	5	0.79185	
	milik keluarga	6	3	3	1	
	dinas	1	0	1	0	
pekerjaan						0.11418
	wiraswasta	15	11	4	0.83668	
	pns	3	2	1	0.91825	
	petani	9	7	2	0.76412	
	tni/polri	1	0	1	0	
	guru/dosen	1	1	0	0	
	lain-lain	1	0	1	0	
penghasilan_per_tahun						0.00977
	<=45	16	12	4	0.81128	
	>45	14	9	5	0.94037	
nilai_jaminan						0.01276
	<=100 juta	17	11	6	0.93671	
	>100 juta	13	10	3	0.7794	
tujuan_pinjaman						0.10699
	modal usaha	6	2	4	0.91825	
	kendaraan	24	19	5	0.73831	
pinjaman_max						0.00161
	<=60 juta	9.00	6	3	0.91825	
	>60 juta	21.00	15	6	0.86311	

Dari tabel 4.3 diperoleh bahwa atribut dengan nilai *gain* tertinggi adalah *status_tempat_tinggal* dengan nilai 0,12699, maka *status_tempat_tinggal* menjadi *node* akar. Atribut *status_tempat_tinggal* terdiri dari 4 kategori nilai yaitu “Milik Keluarga”, “Milik Sendiri”, “Rumah Dinas”, dan “Sewa”, oleh karena itu diagram pohon yang terbentuk memiliki empat cabang. Diagram pohon keputusan setelah dicabangkan dengan *node status_tempat_tinggal* digambarkan pada gambar 4.2



Gambar 4.2 Hasil Pohon Keputusan Menggunakan RapidMiner

Berdasarkan gambar 4.2 diatas, semua kasus sudah masuk pada kelasnya, sehingga pohon keputusan diatas merupakan pohon keputusan terakhir yang terbentuk dan diperoleh juga *rule-rule* untuk klasifikasi nasabah yang Diterima(A) dan Ditolak(R) dibawah ini yaitu:

```

IF status_tinggal== 'Milik Sendiri' THEN
  IF tujuan_pinjam== 'Kendaraan' THEN
    IF marital == 'Duda' THEN
      IF penghasilan_tahun == 'K' THEN
        pred_hasil = "Disetujui(A)"
      ELSEIF
        penghasilan_tahun == 'B' THEN
          pred_hasil = "Ditolak(R)"
        ENDIF
      ENDIF
    ENDIF
  ENDIF
ELSEIF status_tinggal== 'Milik Sendiri' THEN

```

```

IF tujuan_pinjam== 'Kendaraan' THEN
  IF marital == 'Menikah' THEN
    IF nilai_jaminan == 'K' THEN
      IF gender == 'L' THEN
        IF pendidikan == 'S1' THEN
          pred_hasil = "Disetujui(A)"
        ELSEIF pendidikan == 'SMA' THEN
          pred_hasil = "Ditolak(R)"
        ENDIF
      ENDIF
    ENDIF
  ENDIF
ENDIF
ENDIF
ELSEIF status_tinggal== 'Milik Sendiri' THEN
  IF tujuan_pinjam== 'Kendaraan' THEN
    IF marital == 'Menikah' THEN
      IF nilai_jaminan == 'K' THEN
        IF gender == 'P' THEN
          pred_hasil = "Disetujui(A)"
        ELSEIF nilai_jaminan == 'B' THEN
          pred_hasil = "Disetujui(A)"
        ENDIF
      ENDIF
    ENDIF
  ENDIF
ENDIF
ELSE IF status_tinggal== 'Milik Sendiri' THEN
  IF tujuan_pinjam== 'Kendaraan' THEN
    IF marital == 'Menikah' THEN
      IF nilai_jaminan == 'B' THEN
        pred_hasil = "Disetujui(A)"
      ENDIF
    ENDIF
  ENDIF
ENDIF

```

```

        ENDIF
    ENDIF
ENDIF
ENDIF
ELSE IF status_tinggal== 'Milik Sendiri' THEN
    IF tujuan_pinjam== 'Modal Usaha' THEN
        pred_hasil = "Disetujui(A)"
    ELSE IF status_tinggal== 'Milik Sendiri' THEN
        IF tujuan_pinjam== 'Kendaraan' THEN
            IF marital == 'Belum Menikah' THEN
                pred_hasil = "Ditolak(R)"
            ENDIF
        ENDIF
    ENDIF
ENDIF
ENDIF
ELSE IF status_tinggal== 'Rumah Dinas' THEN
    pred_hasil = "Disetujui(A)"
ELSE IF status_tinggal== 'Sewa' THEN
    IF marital == 'Menikah' THEN
        pred_hasil = "Disetujui(A)"
    ELSEIF marital == 'Belum Menikah' THEN
        pred_hasil = "Ditolak(R)"
    ELSE IF status_tinggal== 'Milik Keluarga' THEN
        pred_hasil = "Disetujui(A)" ENDIF

```

4.3.2 Pengujian Klasifikasi dengan Algoritma *Naive Bayes*

Penggunaan algoritma *Naive Bayes* menggunakan *data training* pada Tabel 4.2 dimulai dengan melakukan perhitungan *probabilitas prior* untuk mengetahui nilai yang disetujui dan ditolak untuk semua *record*. Pada *data training* jumlah *record* sebanyak 30, dimana nasabah yang disetujui sebanyak 21(dua puluh satu) dan yang ditolak sebanyak 9(sembilan). Berikut hasil perhitungan *prior probability* terlihat pada tabel dibawah.

Tabel 4.4 Perhitungan *Probabilitas Prior*

Atribut	Range	Jumlah kasus	Disetujui	Ditolak	p(x c1)	
					Disetujui	Ditolak
Total		30	21	9	0.7	0.3
Age						
	<=40	15	13	2	0.43333	0.06667
	>40	15	8	7	0.26667	0.23333
gender						
	L	26	19	7	0.63333	0.23333
	P	4	2	2	0.06667	0.06667
Pend_terakhir						
	S2	1	0	1	0.00000	0.03333
	S1	1	1	0	0.03333	0.00000
	SMA	25	19	6	0.63333	0.20000
	SMP	3	1	2	0.03333	0.06667
status						
	belum menikah	2	1	1	0.03333	0.03333
	menikah	27	20	7	0.66667	0.23333
	duda	1	0	1	0.00000	0.03333
juml_tangg_kel						
	<=2	19	16	3	0.53333	0.10000
	>2	11	5	6	0.16667	0.20000
status_tempat_tinggal						
	sewa	2	2	0	0.06667	0.00000
	milik sendiri	21	16	5	0.53333	0.16667
	milik keluarga	6	3	3	0.10000	0.10000
	dinas	1	0	1	0.00000	0.03333
pekerjaan						
	wiraswasta	15	11	4	0.36667	0.13333
	pns	3	2	1	0.06667	0.03333
	petani	9	7	2	0.23333	0.06667
	tni/polri	1	0	1	0.00000	0.03333
	guru/dosen	1	1	0	0.03333	0.00000
	lain-lain	1	0	1	0.00000	0.03333
penghasilan_per_tahun						
	<=45	16	12	4	0.40000	0.13333
	>45	14	9	5	0.30000	0.16667
nilai_jaminan						
	<=100 juta	17	11	6	0.36667	0.20000
	>100 juta	13	10	3	0.33333	0.10000
tujuan_pinjaman						
	modal usaha	6	2	4	0.06667	0.13333
	kendaraan	24	19	5	0.63333	0.16667
pinjaman_max						
	<=60 juta	9.00	6	3	0.20000	0.10000
	>60 juta	21.00	15	6	0.50000	0.20000

Untuk menentukan kasus baru termasuk disetujui atau ditolak, dilakukan perhitungan *probabilitas posterior* berdasarkan *probabilitas prior* yang telah dihitung sebelumnya pada Tabel 4.4. Perhitungan *probabilitas posterior* untuk menentukan *data testing* termasuk klasifikasi mana yang diterima atau ditolak terdapat disajikan pada Tabel 4.5. Misalkan diambil sebuah *data testing* X dengan nilai seperti pada Tabel 4.5 kolom dua, untuk menentukan kelas mana, dilakukan perhitungan *probabilitas posterior* yang hasilnya terdapat pada Tabel 4.5 kolom tiga dan empat.

Tabel 4.5 *Probabilitas Posterior Data*

Data X		p(X Ci)	
Atribut	Nilai	DISETUJUI	DITOLAK
Age	32	0.43333	0.06667
Gender	L	0.63333	0.23333
Pend.Terakhir	SMA	0.63333	0.20000
Status	Menikah	0.66667	0.23333
Jml Tangg Kel.	1	0.53333	0.10000
Status Tempat Tinggal	Milik Sendiri	0.53333	0.16667
Pekerjaan	Wiraswasta	0.36667	0.13333
Penghasilan_per_tahun	36,295,000	0.40000	0.13333
Nilai Jaminan	80,000,000	0.36667	0.20000
Tujuan Pinjaman	Kendaraan	0.63333	0.16667
Pinjaman_max	59,100,000	0.20000	0.10000

Dari tabel diatas terdapat beberapa langkah untuk menghitung, yaitu:

- a. $P(X|C_i)$ = P(X| remark = Disetujui)
 $= 0.43333 \times 0.63333 \times 0.63333 \times 0.66667 \times 0.53333$
 $\times 0.53333 \times 0.36667 \times 0.40000 \times 0.36667 \times 0.63333 \times 0.20000$
 $= 0,000224$
- b. $P(X|C_i)$ = P(X| remark = Ditolak)
 $= 0.06667 \times 0.23333 \times 0.20000 \times 0.23333 \times 0.10000$
 $\times 0.16667 \times 0.13333 \times 0.13333 \times 0.20000 \times$
 0.16667×0.10000
 $= 0,0000000007169$
- c. $P(X|C_i) P(C_i)$ = P(X| remark = Disetujui) P(remark = Disetujui)
 $= 0,7 \times 0,000224$
 $= \mathbf{0.000157}$
- d. $P(X|C_i) P(C_i)$ = P(X| remark = Ditolak) P(remark = Ditolak)
 $= 0,3 \times 0,0000000007169$
 $= 0,000000000215$

Dari hasil perhitungan di atas, didapat nilai $P(X|C_i)$ dan $P(X|C_i) P(C_i)$ lebih besar untuk *remark* = Disetujui sehingga dapat disimpulkan bahwa *data testing* tersebut termasuk klasifikasi “Disetujui”.

4.3.3 Confusion Matrix, F-measure dan Kurva ROC Decision Tree

Pengujian dilakukan dengan *confusion matrix* yang terdiri dari *accuracy*, *precision* dan *recall* dilakukan pada dataset sebanyak 300 data yang diolah dengan menggunakan *Decision Tree*. Pengujian *confusion matrix* untuk dataset yang diolah menggunakan *Decision Tree* dapat dilihat pada tabel dibawah ini.

Tabel 4.6 *Confusion Matrix* dengan *Decision Tree*

Dataset			
	A	R	Precision
Pred A	235	38	86.08%
Pred R	4	23	85.19%
Recall	98.33%	37.70%	

Nilai akurasi dari *confusion matrix* tersebut adalah sebagai berikut:

$$\begin{aligned} \text{akurasi} &= \frac{(\text{TN} + \text{TP})}{(\text{TN} + \text{FN} + \text{FP} + \text{TP})} \\ &= \frac{(23 + 235)}{(23 + 4 + 38 + 235)} \\ &= \frac{258}{300} = 0,8600 \end{aligned}$$

$$\text{akurasi} = \mathbf{86,00\%}$$

Nilai *precision* dari *confusion matrix* tersebut adalah sebagai berikut:

$$\begin{aligned} \text{precision} &= \frac{\text{TP}}{(\text{TP} + \text{FP})} \\ &= \frac{235}{(235 + 38)} \\ &= \frac{235}{273} = 0,8608 \end{aligned}$$

$$\text{precision} = \mathbf{86,08\%}$$

Nilai *recall* dari *confusion matrix* tersebut adalah sebagai berikut:

$$\begin{aligned} \text{recall} &= \frac{\text{TP}}{(\text{TP} + \text{FN})} \\ &= \frac{235}{(235 + 4)} \end{aligned}$$

$$= \frac{235}{239} = 0,9833$$

$$\mathbf{recall = 98,33\%}$$

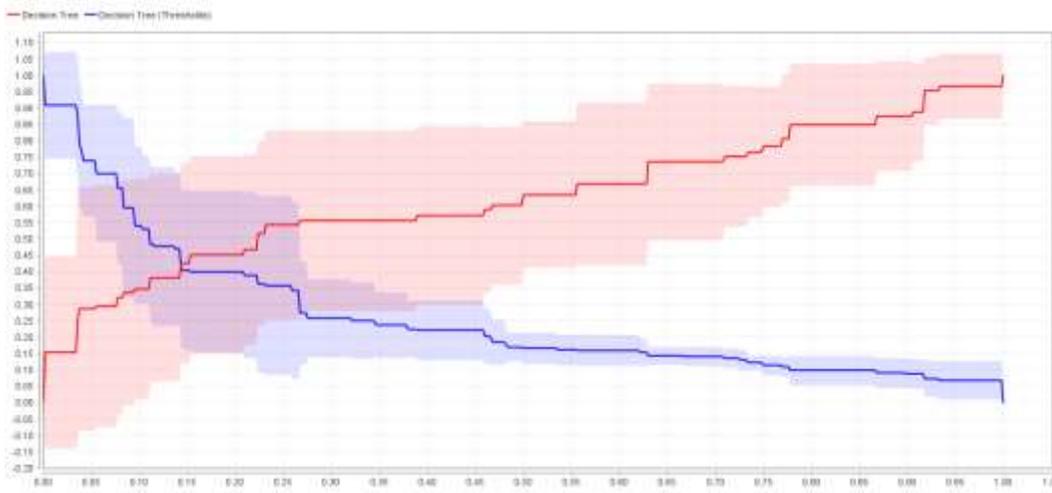
Nilai *F-measure* dapat dihitung dengan persamaan berikut:

$$f_measure = \frac{2 \times recall \times precision}{(recall + precision)}$$

$$= \frac{1,6928}{1,8441} = 0,9179$$

$$\mathbf{f_measure = 91,79\%}$$

Kurva ROC untuk *Decision Tree* sebagai berikut:



Gambar 4.3 Nilai AUC dengan *Decision Tree*

4.3.4 *Confusion Matrix, F-measure* dan Kurva ROC Algoritma *Naïve Bayes*

Pengujian *confusion matrix* untuk *dataset* yang diolah menggunakan algoritma *Naïve Bayes* dapat dilihat pada tabel dibawah ini.

Tabel 4.7 *Confusion Matrix* dengan algoritma *Naïve Bayes*

dataset			
	A	R	Precision
Pred A	234	53	81.53%
Pred R	5	8	61.54%

Recall	97.91%	13.11%	
---------------	--------	--------	--

Nilai akurasi dari *confusion matrix* tersebut adalah sebagai berikut:

$$\begin{aligned}
 \text{akurasi} &= \frac{(\text{TN} + \text{TP})}{(\text{TN} + \text{FN} + \text{FP} + \text{TP})} \\
 &= \frac{(8 + 234)}{(8 + 5 + 53 + 234)} \\
 &= \frac{242}{300} = 0,8067
 \end{aligned}$$

$$\text{akurasi} = \mathbf{80,67\%}$$

Nilai *precision* dari *confusion matrix* tersebut adalah sebagai berikut:

$$\begin{aligned}
 \text{precision} &= \frac{\text{TP}}{(\text{TP} + \text{FP})} \\
 &= \frac{234}{(234 + 53)} \\
 &= \frac{234}{287} = 0,81533
 \end{aligned}$$

$$\text{precision} = \mathbf{81,53\%}$$

Nilai *recall* dari *confusion matrix* tersebut adalah sebagai berikut:

$$\begin{aligned}
 \text{recall} &= \frac{\text{TP}}{(\text{TP} + \text{FN})} \\
 &= \frac{234}{(234 + 5)} \\
 &= \frac{234}{239} = 0,97907
 \end{aligned}$$

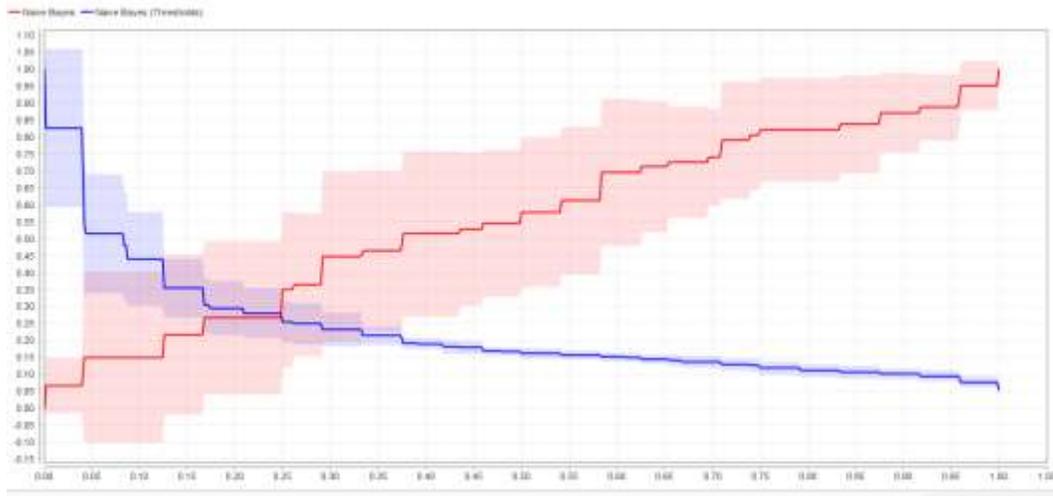
$$\text{recall} = \mathbf{97,91\%}$$

Nilai *F-measure* dapat dihitung dengan persamaan berikut:

$$\begin{aligned}
 f_measure &= \frac{2 \times \text{recall} \times \text{precision}}{(\text{recall} + \text{precision})} \\
 &= \frac{2 \times (0,97907) \times (0,81533)}{(0,97907 + 0,81533)} \\
 &= \frac{1,59653}{1,79440} = 0,88973
 \end{aligned}$$

$$f_{measure} = 88,97\%$$

Kurva ROC untuk algoritma *Naïve Bayes* sebagai berikut:



Gambar 4.4 Nilai AUC dengan Algoritma *Naïve Bayes*

4.4 Evaluasi dan Validasi

4.4.1 Evaluasi *Confusion Matrix*, *F-measure* dan Kurva ROC *Decision Tree*

Pengujian dilakukan dengan *confusion matrix* yang terdiri dari akurasi (*accuracy*), *precision* dan *recall* dilakukan pada 45 *data testing* yang diolah dengan menggunakan *Decision Tree*. Pengujian *confusion matrix* untuk *data testing* yang diolah menggunakan *Decision Tree* dapat dilihat pada tabel dibawah ini.

Tabel 4.8 *Confusion Matrix* dengan *Decision Tree*

Dataset			
	A	R	Precision
Pred A	35	6	85.37%
Pred R	1	3	75.00%
Recall	97.22%	33.33%	

Nilai akurasi dari *confusion matrix* tersebut adalah sebagai berikut:

$$akurasi = \frac{(TN + TP)}{(TN + FN + FP + TP)}$$

$$\begin{aligned}
 &= \frac{(3 + 35)}{(3 + 1 + 6 + 35)} \\
 &= \frac{38}{45} = 0,8444
 \end{aligned}$$

$$\mathbf{akurasi = 84,44\%}$$

Nilai *precision* dari *confusion matrix* tersebut adalah sebagai berikut:

$$\begin{aligned}
 precision &= \frac{TP}{(TP + FP)} \\
 &= \frac{35}{(35 + 6)} \\
 &= \frac{35}{41} = 0,8537
 \end{aligned}$$

$$\mathbf{precision = 85.37\%}$$

Nilai *recall* dari *confusion matrix* tersebut adalah sebagai berikut:

$$\begin{aligned}
 recall &= \frac{TP}{(TP + FN)} \\
 &= \frac{35}{(35 + 1)} \\
 &= \frac{35}{36} = 0,9722
 \end{aligned}$$

$$\mathbf{recall = 97.22\%}$$

Nilai *F-measure* dapat dihitung dengan persamaan berikut:

$$\begin{aligned}
 f_measure &= \frac{2 \times recall \times precision}{(recall + precision)} \\
 &= \frac{2 \times (0,9722) \times (0,8537)}{(0,9722 + 0,8537)} \\
 &= \frac{1,6599}{1,8259} = 0,9091
 \end{aligned}$$

$$\mathbf{f_measure = 90.91\%}$$

4.4.2 Evaluasi *Confusion Matrix*, *F-measure* dan Kurva ROC *Naïve Bayes*

Pengujian *confusion matrix* untuk *dataset* yang diolah menggunakan algoritma *Naive Bayes* dengan 45 *data testing* dapat dilihat pada tabel dibawah ini.

Tabel 4.9 *Confusion Matrix* dengan algoritma *Naïve Bayes*

dataset			
	A	R	Precision
Pred A	30	9	76.92%
Pred R	6	0	0.00%
Recall	83.33%	0.00%	

Nilai akurasi dari *confusion matrix* tersebut adalah sebagai berikut:

$$\begin{aligned}
 \text{akurasi} &= \frac{(\text{TN} + \text{TP})}{(\text{TN} + \text{FN} + \text{FP} + \text{TP})} \\
 &= \frac{(0 + 30)}{(0 + 6 + 9 + 30)} \\
 &= \frac{30}{45} = 0,6667
 \end{aligned}$$

$$\text{akurasi} = \mathbf{66.67\%}$$

Nilai *precision* dari *confusion matrix* tersebut adalah sebagai berikut:

$$\begin{aligned}
 \text{precision} &= \frac{\text{TP}}{(\text{TP} + \text{FP})} \\
 &= \frac{30}{(30 + 9)} \\
 &= \frac{30}{39} = 0.7692
 \end{aligned}$$

$$\text{precision} = \mathbf{76.92\%}$$

Nilai *recall* dari *confusion matrix* tersebut adalah sebagai berikut:

$$\begin{aligned}
 \text{recall} &= \frac{\text{TP}}{(\text{TP} + \text{FN})} \\
 &= \frac{30}{(30 + 6)} \\
 &= \frac{30}{36} = 0.8333
 \end{aligned}$$

$$\text{recall} = \mathbf{83.33\%}$$

Nilai *F-measure* dapat dihitung dengan persamaan berikut:

$$\begin{aligned}
 f_measure &= \frac{2 \times recall \times precision}{(recall + precision)} \\
 &= \frac{2 \times (0,8333) \times (0,7692)}{(0,8333 + 0,7692)} \\
 &= \frac{1,2819}{1,6025} = 0,7999 \\
 f_measure &= 79,99\%
 \end{aligned}$$

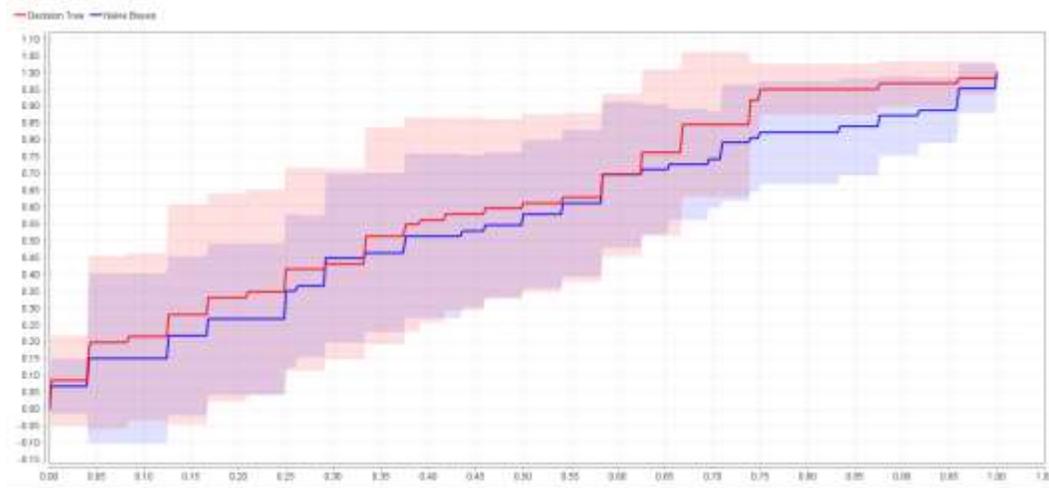
4.5 Hasil Komparasi

Model dengan algoritma *Decision Tree* dan *Naïve Bayes* untuk prediksi kelayakan pemberian kredit kepada nasabah yang diuji tingkat akurasi menghasilkan perbandingan nilai akurasi (*accuracy*) dan nilai AUC. Dengan dataset nasabah sebagai data uji. Algoritma *Decision Tree* mendapatkan akurasi yang paling tinggi yaitu sebesar 85,22%. Sedangkan algoritma *Naïve Bayes* mendapatkan akurasi sebesar 73,67%. Seperti yang ditunjukkan pada tabel 4.10.

Tabel 4.10 Perbandingan Akurasi dan AUC Metode *Decision Tree* dan *Naïve Bayes*

Metode Data Mining	Akurasi		AUC		Perbandingan	Perbandingan
	Training	Testing	Training	Testing	Nilai Akurasi	Nilai AUC
<i>Decision Tree</i>	86,00%	84,44%	93,80%	97,40%	85,22%	95,60%
<i>Naïve Bayes</i>	80,67%	66,67%	91,20%	87,90%	73,67%	89,55%

Kurva ROC untuk perbandingan dua metode algoritma *Decision Tree* dan *Naïve Bayes* sebagai berikut:



Gambar 4.5 Nilai Perbandingan AUC

Hasil perhitungan divisualisasikan dengan kurva ROC (*Receiver Operating Characteristic*) atau AUC (*Area Under Curve*). ROC memiliki tingkat nilai diagnosa yaitu [18]:

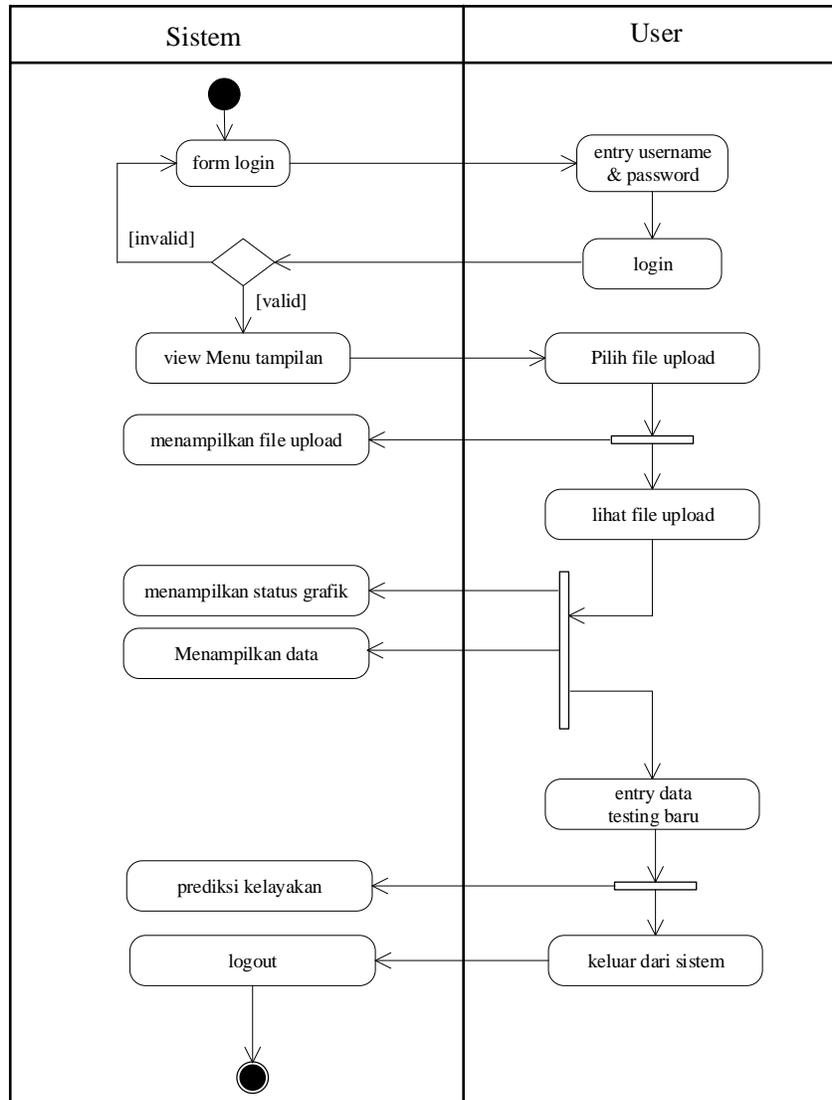
- a. Akurasi bernilai $0.90 - 1.00 = \textit{excellent classification}$
- b. Akurasi bernilai $0.80 - 0.90 = \textit{good classification}$
- c. Akurasi bernilai $0.70 - 0.80 = \textit{fair classification}$
- d. Akurasi bernilai $0.60 - 0.70 = \textit{poor classification}$
- e. Akurasi bernilai $0.50 - 0.60 = \textit{failure}$

Hasil yang didapat dari pengolahan ROC menggunakan data training untuk algoritma *decision tree* sebesar 0.9560 dengan tingkat diagnosa *excellent classification*, sedangkan algoritma *Naïve Bayes* sebesar 0.8955 dengan tingkat diagnosa *good classification*.

4.6 Penerapan Algoritma Terpilih

Dari hasil evaluasi dan validasi diatas dapat diketahui bahwa metode *Decision Tree* memiliki tingkat akurasi dan *perfoma* yang baik, sehingga *rule* yang dihasilkan oleh metode *Decision Tree* dapat dijadikan sebagai rule untuk pembuatan *prototype* yang dapat memudahkan dalam memprediksi pemberian pinjaman kembali kepada nasabah.

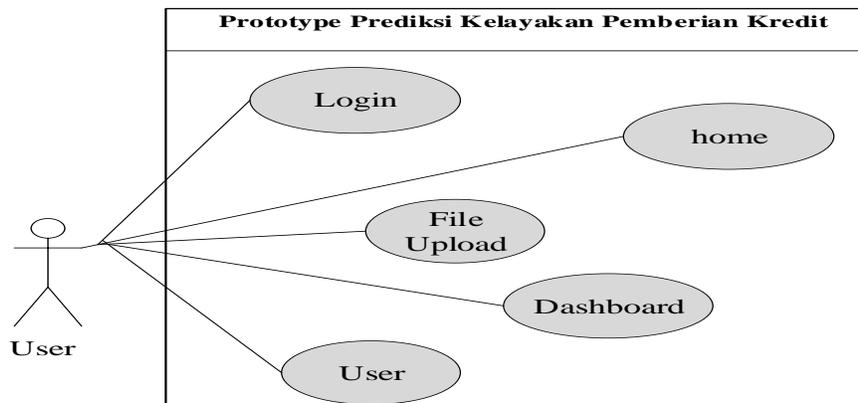
Berikut adalah alur dari *Activity Diagram*, *Use Case*, *Class Diagram* dan *Flow Chart* yang diimplementasikan kedalam *prototype* Prediksi Kelayakan Pemberian Kredit kepada nasabah.



Gambar 4.6 Activity Diagram Login

User membuka aplikasi dengan memasukkan *login* pada sistem dan akan menampilkan *menu* tampilan dengan menampilkan pilihan untuk *upload file* yang dipilih *user* pada saat yang sama sistem juga dapat menampilkan data dari *file* yang di *upload* yang dipilih *user*, selanjutnya sistem akan *import upload file* dan memberikan *summary* dari data *upload*. *User* akan meng-*upload* lagi pilihan *file* data *testing* dan menampilkan grafik prediksi dari data yang dipilih. Setelah selesai *user* akan keluar (*logout*) dari aplikasi. Dibawah ini adalah *use case* dari *menu login* dan *class diagram*

a. Use Case Diagram

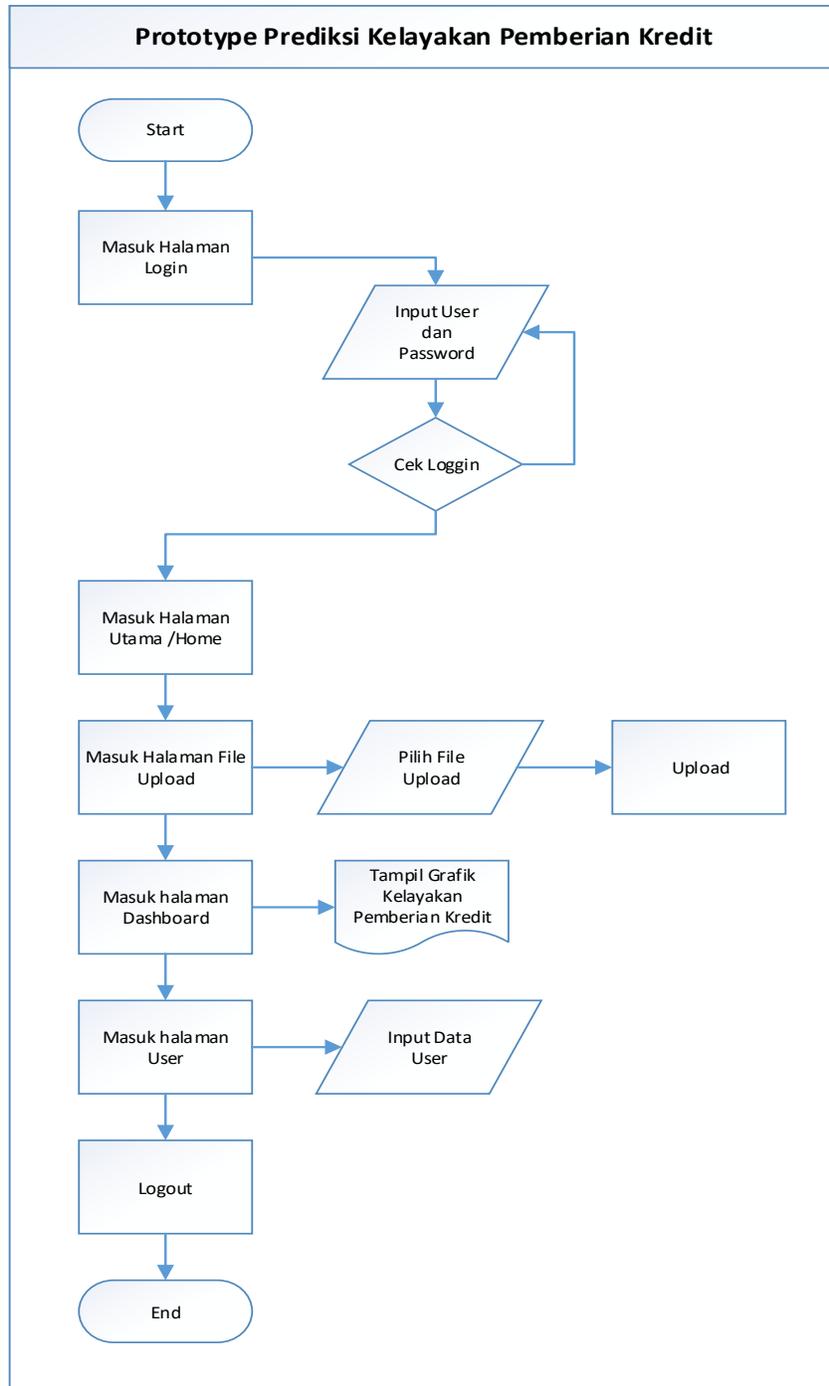


Gambar 4.7 Use Case Login

b. Class diagram



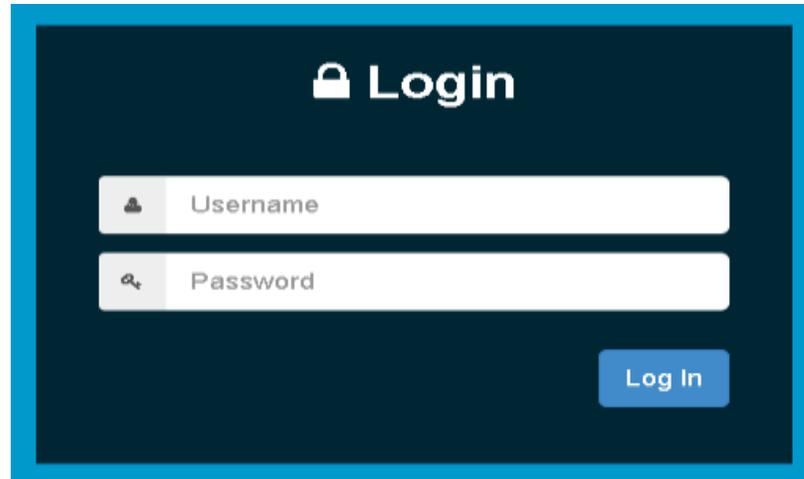
Gambar 4.8 Class Diagram



Gambar 4.9 *Flow Chart Prototype* Prediksi Kelayakan Pemberian Kredit

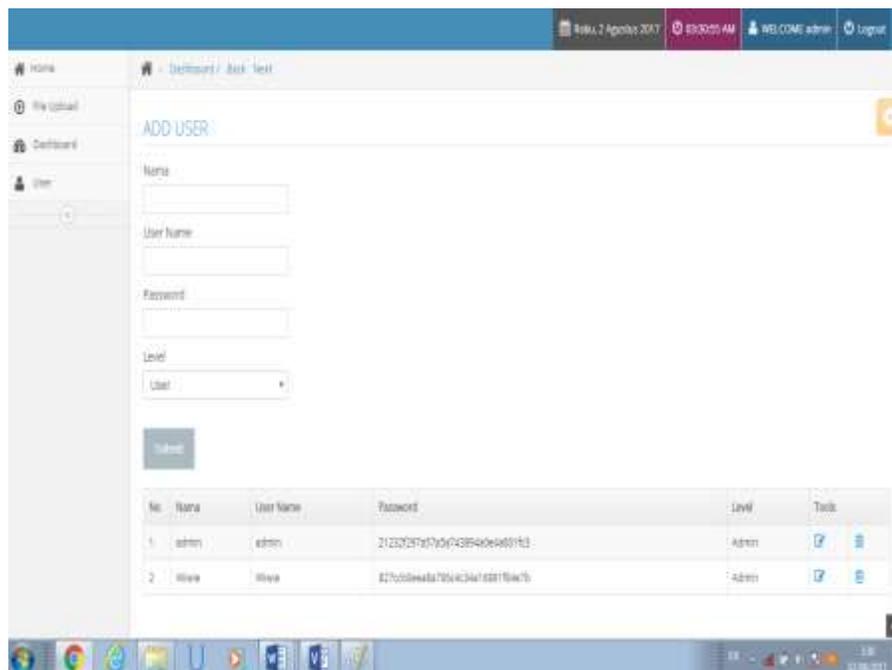
Prototype yang digunakan dalam penelitian ini dibuat berbasis *web* dengan menggunakan PHP dan *database* menggunakan *MySQL*. Tampilan untuk *Form* utama *Graphical User Interface (GUI)* *prototype* prediksi kelayakan pemberian kredit dapat dilihat pada gambar dibawah.

1. Tampilan Login



Gambar 4.10 *Form Login*

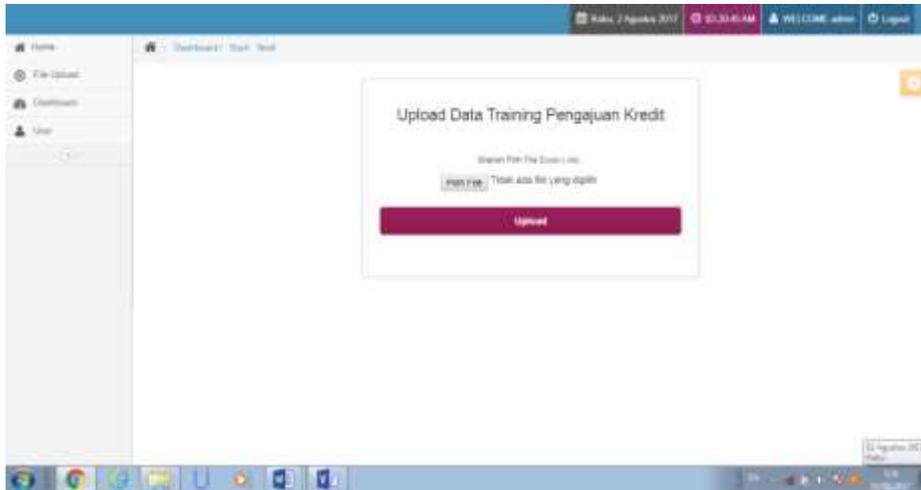
2. Tampilan Halaman Utama



Gambar 4.11 *Form User*

Pihak admin dapat menambahkan *user* pada menu *user*.

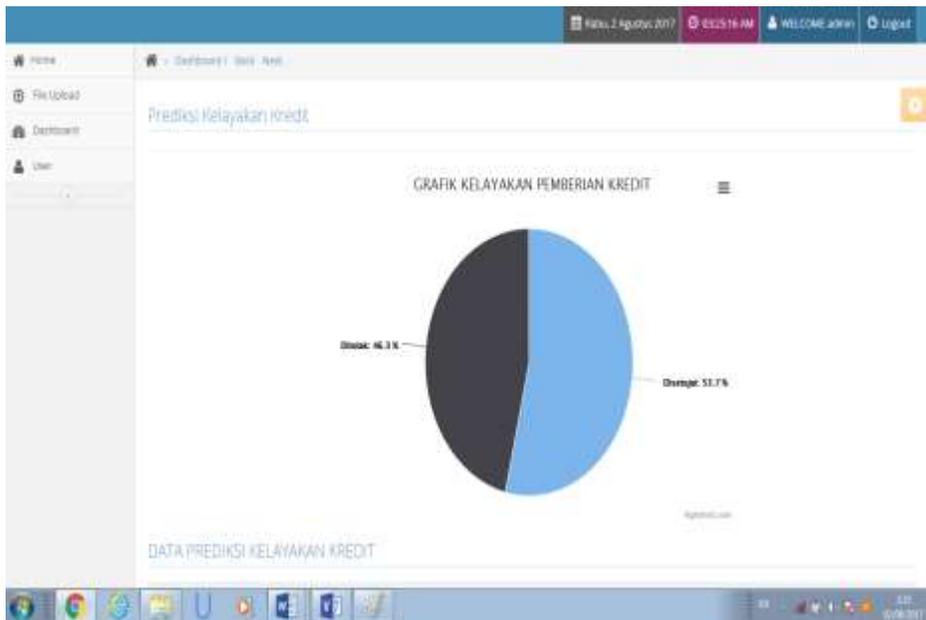
3. Tampilan Halaman *Upload*



Gambar 4.12 Form Upload Data Training Pengajuan Kredit

Pada menu file *upload*, user dapat meng-klik pilih *file* untuk memilih *file* data yang akan di uji. Selanjutnya akan ditampilkan hasil import data dari *file* yang dipilih.

4. Tampilan *Dashboard*



Gambar 4.13 Halaman *Dashboard*

Pada menu *Dashboard*, user dapat melihat grafik dari *file* yang telah dipilih, dengan mengklik *Dashboard*. Menu ini menampilkan juga detail dari hasil prediksi kelayakan kredit.

5. Tampilan *Detail Form*

DATA DETAIL KREDITUR Disetujui

No	Usia	Gender	Pendidikan Terakhir	Status	Jml Tanggungan	Status Timb. Tangg.	Pekerjaan	Jml Pekerjaan yg Dilakukan	Tujuan Pekerjaan	Income/Basar	Hasil	Pred. Hasil
1	52	M	S2	Merah	1	MRS Sendiri	Wiraswasta	3001	Personal Loan	11230	Disetujui	Disetujui
2	47	M	S1	Merah	2	MRS Keluarga	Guru	250	Car Loan	725	Disetujui	Disetujui
3	38	F	S1	Merah	4	MRS Keluarga	Pegawai Swasta	500	Houseing Loan	750	Disetujui	Disetujui
4	43	P	S2	Merah	3	MRS Sendiri	Wiraswasta	2700	Houseing Loan	4300	Disetujui	Disetujui
5	49	M	S2	Belum Merah	1	MRS Keluarga	Wiraswasta	800	Personal Loan	1300	Disetujui	Disetujui
6	48	P	S2	Merah	2	MRS Sendiri	Wiraswasta	800	Houseing Loan	875	Disetujui	Disetujui
7	52	M	S1	Merah	1	MRS Keluarga	Pegawai Swasta	300	Personal Loan	675	Disetujui	Disetujui
8	50	M	S1	Merah	2	MRS Sendiri	Pegawai Swasta	250	Car Loan	250	Disetujui	Disetujui
9	47	M	S3	Merah	5	MRS Sendiri	Dokter	500	Personal	250	Disetujui	Disetujui

Gambar 4.14 Detail Nasabah

Detail dari hasil prediksi akan ditampilkan secara rinci dan dapat dilihat dalam bentuk Microsoft Excel.

4.7 Pengujian *Prototype* Perangkat Lunak (*Software*)

4.7.1 Pengujian *Prototype*

Pengujian dilakukan dengan tujuan untuk mengetahui apakah aplikasi dibangun sesuai dengan fungsional yang di harapkan.

Tabel 4.11 Pengujian Kotak Hitam

Kelas Uji	Butir Uji	Jenis Pengujian
<i>File Upload</i>	<i>Pilih File</i>	<i>Black Box</i>
	<i>Upload File</i>	<i>Black Box</i>
<i>Dashboard</i>	Lihat Grafik	<i>Black Box</i>

A. Pengujian *File Upload*

Pada pengujian *file upload* terbagi atas dua bagian yaitu, *pilih file* dan *upload file*.

Tabel 4.12 Pengujian *File Upload*

Kasus dan Hasil Uji (Data sesuai)			
Data Masukan	Yang Diharapkan	Pengamatan	Kesimpulan
<i>File yang akan di upload</i>	File dalam bentuk <i>ext.xls</i>	Dapat melakukan <i>pilih file upload</i>	Diterima

Klik <i>Upload</i>	File berhasil diimport oleh sistem	Dapat melakukan <i>upload</i> data	Diterima
Kasus dan Hasil Uji (Data salah)			
Data Masukan	Yang Diharapkan	Pengamatan	Kesimpulan
<i>Extension</i> berbeda, tidak sesuai dengan sistem	Data tidak tersimpan	Data tidak tersimpan dan menampilkan kesalahan, sesuai yang diharapkan	Diterima

B. Pengujian *Dashboard*

Pada pengujian *Dashboard user* hanya meng-klik *Dashboard*.

Tabel 4.13 Pengujian *Dashboard*

Kasus dan Hasil Uji (Data sesuai)			
Data Masukan	Yang Diharapkan	Pengamatan	Kesimpulan
<i>File</i> yang telah di <i>upload</i>	Berhasil menampilkan Grafik persentase	Dapat menampilkan grafik presentase	Diterima
Klik Tombol Detail	Dapat menampilkan detail <i>customer</i> yang Disetujui ataupun Ditolak	Dapat menampilkan sesuai yang diharapkan	Diterima

Berdasarkan pengujian *Black Box* yang sudah dilakukan, menjelaskan bahwa aplikasi yang dibangun sudah berjalan dengan baik dan sesuai dengan yang diharapkan.

4.7.2 Lingkungan Pengujian

Lingkungan pengujian memberikan gambaran tentang spesifikasi *hardware* dan *software* yang digunakan oleh pengguna dalam proses pengujian sistem, baik pengujian validasi maupun kualitas. Spesifikasi tersebut diperoleh dalam proses observasi berdasarkan aspek sistem. Berikut ringkasan singkat spesifikasi dari perangkat keras dan perangkat lunak yang digunakan oleh pengguna untuk proses pengujian,

1. *Hardware*,

- a. Processor AMD Dual Core E1-6010 1.35GHz
 - b. Harddisk 500 GB
 - c. RAM / Memory 2 GB
2. *Software*,
- a. OS Win XP/7/8
 - b. XAMPP
 - c. PHP
 - d. MySQL
 - e. Web Browser (Mozilla Firefox/Google Chrome)

Perangkat lunak yang digunakan untuk membangun aplikasi *prototype* aplikasi prediksi kelayakan pemberian kredit adalah menggunakan *Software open source* yaitu PHP, XAMPP dan MySQL untuk *database*. Ditinjau dari segi kehandalan *software* tersebut sudah teruji dan sudah banyak digunakan untuk membangun sistem tanpa terkendala masalah lisensi Karena bersifat *open source*.

4.7.3 Hasil Pengujian Kualitas Perangkat Lunak (*Software*)

Untuk dapat mengetahui kesesuaian yang diharapkan pada *prototype* sistem informasi prediksi kelayakan pemberian kredit maka metode yang digunakan untuk pengukuran perangkat lunak tersebut adalah *Software Quality Assurance* (SQA). Adapun komponen yang menjadi tolak ukur yang digunakan pada pengujian ini dapat dilihat pada tabel 4.14.

Tabel 4.14 Komponen *Software Quality Assurance*

No	Metrik	Deskripsi
1	<i>Auditability</i>	Kesesuaian terhadap standar
2	<i>Accuracy</i>	Ketepatan perhitungan
3	<i>Completeness</i>	Kelengkapan kebutuhan
4	<i>Error Tolerance</i>	Toleransi terhadap kesalahan
5	<i>Execution Efficiency</i>	Performa <i>run-time</i>
6	<i>Operability</i>	Kemudahan untuk dioperasikan
7	<i>Simplicity</i>	Kemudahan untuk dipahami
8	<i>Training</i>	Kemudahan pembelajaran

Ada delapan komponen dalam *Software Quality Assurance* (SQA) yang digunakan. Dari delapan komponen tersebut akan dibuat delapan pertanyaan untuk angket yang akan disebarakan kepada lima orang *user* yang diambil secara acak.

Tabel 4.15 Skor Metrik

User	SKOR METRIK								SKOR
	1	2	3	4	5	6	7	8	
1	100	80	80	80	100	80	80	80	85
2	80	60	80	60	80	80	80	80	75
3	100	80	80	80	100	100	100	80	90
4	80	80	80	80	100	100	80	100	88
5	80	80	60	80	80	80	80	80	78
RATA - RATA									83

Untuk menghitung nilai rata-rata dari keseluruhan kuisioner adalah sebagai berikut:

$$\text{Rata - rata} = \frac{85 + 75 + 90 + 88 + 78}{5} = 83$$

Skor rata-rata yang dihasilkan adalah 83% merupakan hasil skor yang cukup tinggi, sehingga dapat disimpulkan bahwa kualitas perangkat lunak “*prototype* prediksi kelayakan pemberian kredit kepada nasabah” ini dapat dinyatakan baik untuk dilanjutkan implementasinya sebagai sebuah sistem Prediksi Kelayakan Pemberian Kredit dengan fitur yang lebih lengkap (*Full Version*).

4.8 Implikasi Penelitian

Implikasi penelitian ini berkisar pada aspek sistem dan aspek penelitian lanjut. Aspek sistem terkait dengan teknis operasional, desain *hardware* dan *software* yang diperlukan. Sedangkan aspek penelitian lanjut berkaitan dengan penelitian lanjutan yang diperlukan untuk meningkatkan kualitas penelitian sebelumnya.

4.8.1 Aspek Sistem

Untuk menerapkan hasil penelitian ini dibutuhkan dukungan sistem yang baik, sehingga pihak yang berkepentingan dapat menggunakan hasil penelitian ini untuk memprediksi kelayakan pemberian kredit untuk nasabah. Oleh karena itu dibutuhkan sarana dan prasarana yang memadai yang terdiri dari *hardware*, *software* (sistem operasi dan aplikasi yang dibuat pada tahap *deployment*), dan infrastruktur lainnya guna memberikan hasil yang terbaik, spesifikasi *hardware*

dan *software* yang dapat digunakan dalam penelitian ini memiliki spesifikasi minimal seperti pada butir 4.5.2 Lingkungan Pengujian Sistem.

4.8.2 Aspek Penelitian Lanjut

Penulis menyadari keterbatasan dalam penelitian saat ini, diharapkan pada penelitian selanjutnya yang sejenis dapat dipertimbangkan hal-hal berikut:

1. Mengkombinasikan lebih banyak metode dalam Analisa data dan penyelesaian masalah, sehingga didapat sebuah sistem yang lebih efektif dan efisien dalam pengolahan ataupun penyajian informasi.
2. Pengelolaan waktu penelitian agar dapat lebih dimaksimalkan, mengingat pendeknya waktu yang tersedia.
3. Peran dari responden sangatlah penting dalam mendukung penelitian ini terutama responden yang berkaitan langsung dengan objek penelitian.

Penelitian ini dapat dikembangkan dengan algoritma klasifikasi yang lain yang terdapat dalam *data mining*, seperti algoritma *Neural Network*, *K-Means* atau *SVM (Support Vector Machine)*.