

BAB IV HASIL DAN PEMBAHASAN

4.1 Pengolahan Data

1. *Business/Research Understanding Phase*

PT Batavia Prosperindo Tbk. Data yang dikumpulkan dengan beberapa atribut yaitu: no_kontrak, nama, accnumber, accname, tgl_lahir, Age, gender, alamat, kota, no_hp, pendidikan, status, jumlah_tanggungan, status_tempat_tinggal, pekerjaan, penghasilan_per_tahun, tgl_transaksi, tgl_bayar_akhir, tunggakan, merk_asset, jenis_asset, nama_asset, tahun_asset, no_rangka, no_mesin, no_polisi, bpkb, warna, nilai_jaminan, jumlah_pinjamanmax, tujuan_pinjaman, cmo.

Data yang dikumpulkan dengan beberapa *field* dengan jumlah atribut sebanyak 32 yaitu :

Tabel 4.1 Atribut *Data Master*

No	Atribut	Keterangan
1	no_kontrak	nomor kontrak data nasabah
2	Nama	Nama nasabah
3	Accnumber	No rekening nasabah
	Accname	
5	Tgl_lahir (<i>dob</i>)	tanggal, bulan dan tahun lahir
6	Age	Umur nasabah
7	Gender	Jenis kelamin
8	Alamat	Alamat nasabah
9	Kota	Kategori wilayah berdasarkan wilayah/area
10	No_hp	Mobile phone
11	Pendidikan	Pendidikan terakhir
12	Status	Status pernikahan

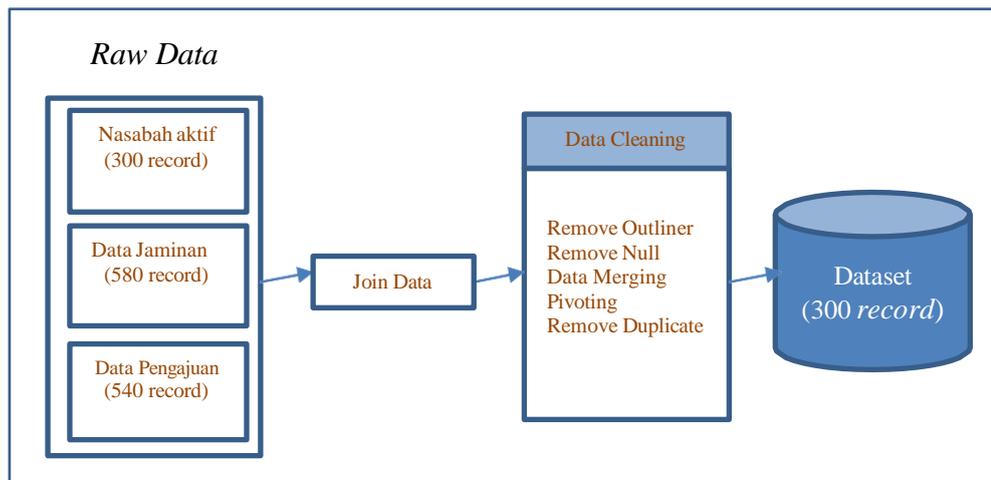
13	Jumlah_tanggungan	Jumlah tanggungan keluarga
14	Status_tempat_tinggal	Kepemilikan rumah
15	Pekerjaan	Jenis pekerjaan
16	Penghasilan_per_tahun	Penghasilan per tahun
17	Tgl_transaksi	Tanggal transaksi terakhir melakukan pinjaman
18	Tgl_bayar_akhir	Tanggal pelunasan terakhir
19	Tunggakan	Jumlah tunggakan
20	Merk_asset	Merek asset/jaminan
21	Jenis_asset	Jenis asset/jaminan
22	Nama_asset	nama asset/jaminan
23	Tahun_asset	Tahun pembuatan asset/jaminan
24	No_rangka	Nomer Rangka kendaraan
25	No_mesin	Nomer mesin kendaraan
26	No_polisi	Nomer polisi kendaraan
27	Bpkb	Nomer BPKB kendaraan
28	Warna	Warna kendaraan
29	Nilai_jaminan	Jumlah nilai taksiran jaminan
30	Jumlah_pinjamanmax	Jumlah pinjaman maximal yang diinginkan
32	Tujuan_pinjaman	Tujuan pinjaman
32	Cmo	Nama sales penjamin

2. *Data Understanding Phase* (Fase Pemahaman Data)

Pada fase pemahaman ini peneliti melakukan wawancara dengan staf *database* untuk mendalami data dan penentuan metode prediksi pengajuan kredit yang sedang berjalan. Dari 32 atribut yang ada maka diambil 12 atribut yang penting maka sisa atribut yang tidak diperlukan dihapuskan. Adapun 12 *atribut* dari data *training* yaitu: *gender*, *age*, pendidikan, status, jumlah_tanggungan, status_tempat_tinggal, pekerjaan, penghasilan_pertahun, nilai_jaminan, tujuan_pinjaman, jumlah_pinjaman, hasil.

3. 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. 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 4.3 Proses Akuisisi Data

Gambar 4.3 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 148 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*
2. Pembuatan aturan dalam basis pengetahuan dibantu oleh Kepala HRD, Kepala Marketing Manager dan Kepala GA di PT Batavia Prosperindo Finance Tbk.

4.2 Data Preparation Phase (Fase Pengolahan Data)

Dari fase pemahaman data akan menjadi modal peneliti untuk masuk ke fase pengolahan data, tahap ini meliputi semua kegiatan untuk membangun *dataset* akhir data yang akan diproses pada tahap pemodelan (*modeling*).

1. Data Cleaning

Untuk membersihkan nilai yang kosong, sebagai contoh *atribut* *tujuan_pinjaman*.

2. Data Integration.

Berfungsi menyatukan tempat yang berbeda kedalam satu data. Dalam kasus ini hanya ada satu tempat data yaitu alamat_nasabah.

3. Data Reduction

Jumlah atribut yang digunakan mungkin terlalu banyak, dari 32 atribut yang digunakan hanya 12 atribut yang diperlukan, dan atribut yang tidak diperlukan akan dihapus.

4. Seleksi Atribut

Pada tahap awal penelitian, seleksi terhadap atribut evaluasi penentuan pengajuan pinjaman kredit pada Tabel 4.1 berikut dilakukan untuk mengurangi atribut awal.

Tabel 4.1 Atribut Data Penelitian Berdasarkan Information

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>)

5. Preparatiom

Penentuan atribut-atribut yang digunakan dilakukan setelah melakukan wawancara dengan bagian terkait mengenai atribut apasaja yang paling penting digunakan. untuk evaluasi nasabah yang melakukan pengajuan pinjaman. Hasilnya ditetapkan 12 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*, hasil. Atribut-atribut ini menjadi *input* pada perancangan, sementara untuk *output*-nya adalah Disetujui atau Ditolak.

Hal yang pertama dilakukan dalam penelitian ini adalah pemakaian algoritma yaitu *Naïve Bayes* 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 adalah hasil penyesuaian atribut yang telah dilakukan data training seperti pada 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	DEDEMAHENDRA *(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	MUKHLIS HS *(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	M 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	ADIYANTO (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	ITRI 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	KA 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	KUSDARMAJI	P	L	SMA	Belum Menikah	K	Milik Sendiri	Petani	K	K	Kendaraan	K	Ditolak
52	074372190080	BURHANDJIN RAIS	P	L	SMA	Belum Menikah	K	Milik Sendiri	Wiraswasta	K	K	Kendaraan	K	Ditolak
53	074372190082	HARDIYANTO (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 Naive Bayes

Pada tahap ini memilih dan menerapkan teknik pemodelan klasifikasi *data mining* yaitu *Naive Bayes*. Data yang diperoleh dari fase pengolahan data akan digunakan pada proses ini.

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 12, 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*

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

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**”.

4.4 Evaluation Phase (Fase Evaluasi)

Pada tahap ini akan dilakukan evaluasi terhadap keefektifan dari perbandingan klasifikasi data mining, sebelum digunakan dan menentukan apakah model dapat mencapai tujuan yang diterapkan pada fase *business understanding*. Ketepatan atau ketelitian menjadi kriteria performance suatu metode peramalan. Ketepatan atau ketelitian tersebut dapat dinyatakan sebagai kesalahan dalam peramalan. Metode yang digunakan dalam fase ini adalah *k-fold cross validation*

Secara umum, kita akan membandingkan n model dalam *k-fold cross validation ini*, dalam arti lain fungsi dari penggunaan metode *k-fold cross validation* adalah

1. Untuk mengetahui performa dari suatu model algoritma dengan melakukan percobaan sebanyak k kali
2. Untuk meningkatkan tingkat performansi dari model tersebut
3. Untuk mengolah data set dengan kelas yang seimbang

4.5 Deployment Phase (Fase Penyebaran)

Tahap *deployment* dilakukan dengan pembuatan prototype klasifikasi *data mining* dengan menggunakan model yang terbentuk pada proses sebelumnya, pemodelan menggunakan UML dengan *use case diagram* dan *activity diagram*, Hasil dari fase ini diharapkan sebuah aplikasi data *mining* yang siap digunakan dan bermanfaat untuk PT Batavia Prosperindo Finance Tbk.

4.6 Confusion Matrix, F-measure dan Kurva ROC(Receiver Operating Characteristic) Algoritma Naïve Bayes

Pengujian *confusion matrix* untuk *dataset* akurasi data *training* 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:

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

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

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

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

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

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

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

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

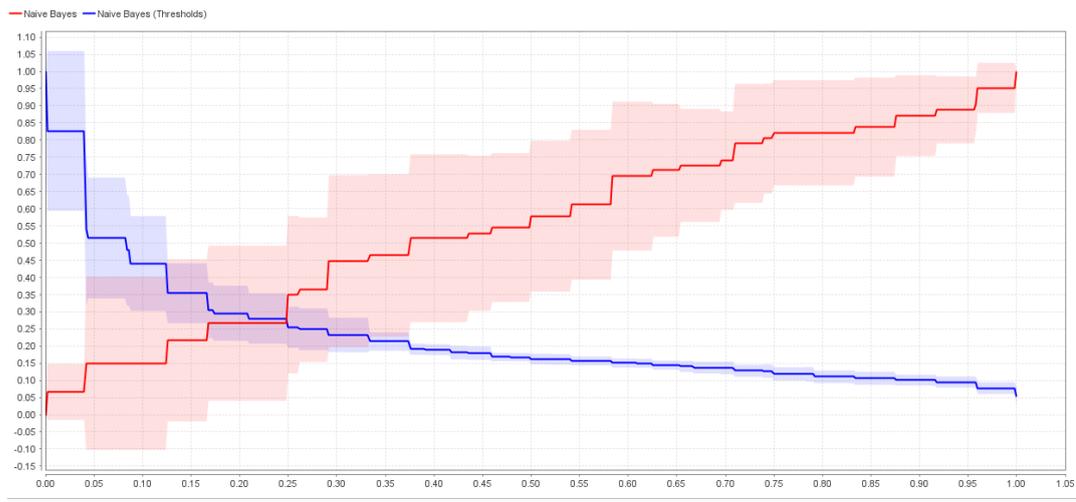
Nilai *F-measure* dapat dihitung dengan persamaan berikut:

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

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

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

Kurva ROC untuk algoritma *Naïve Bayes* sebagai berikut:



Gambar 4.4 Nilai AUC dengan Algoritma *Naïve Bayes*

4.7 Validasi Deployment

4.7.1 Validasi *Confusion Matrix*, *F-measure* dan Kurva ROC *Naïve Bayes*

Pengujian *confusion matrix* untuk *dataset* yang diolah menggunakan algoritma *Naive Bayes* dengan 60 *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:

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

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

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

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

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

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

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

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

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

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

$$f_measure = 79,99\%$$

4.8 Hasil

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

Tabel 4.10 Akurasi dan AUC Metode *Naïve Bayes*

Metode Data Mining	Akurasi		AUC		Perbandingan	Perbandingan
	<i>Training</i>	<i>Testing</i>	<i>Training</i>	<i>Testing</i>	Nilai Akurasi	Nilai AUC
<i>Naïve Bayes</i>	80,67%	66,67%	91,20%	87,90%	73,67%	89,55%

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 = *excellent classification*
- b. Akurasi bernilai 0.80 – 0.90 = *good classification*
- c. Akurasi bernilai 0.70 – 0.80 = *fair classification*
- d. Akurasi bernilai 0.60 – 0.70 = *poor classification*
- e. Akurasi bernilai 0.50 – 0.60 = *failure*

Hasil yang didapat dari pengolahan ROC menggunakan data training untuk algoritma *Naïve Bayes* sebesar 0.8955 dengan tingkat diagnosa *good classification*.