



**PERBANDINGAN ALGORITMA *NAÏVE BAYES* DAN *RANDOM FOREST* UNTUK KLASIFIKASI KELAYAKAN KREDIT PADA PT.
BPR MAGGA JAYA UTAMA**

SKRIPSI

Oleh:

CORNELLIUS ORLANDO CHRISTIADI

20201000039

**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS SAINS DAN TEKNOLOGI**

**UNIVERSITAS BUDDHI DHARMA
TANGERANG**

2025



PERBANDINGAN ALGORITMA *NAÏVE BAYES* DAN *RANDOM FOREST* UNTUK KLASIFIKASI KELAYAKAN KREDIT PADA PT. BPR MAGGA JAYA UTAMA

SKRIPSI

**Diajukan sebagai salah satu syarat untuk memperoleh gelar Sarjana pada
Program Studi Teknik Informatika Fakultas Sains dan Teknologi
Universitas Buddhi Dharma
Jenjang Pendidikan Strata 1**

Oleh:

CORNELLIUS ORLANDO CHRISTIADI

20201000039

FAKULTAS SAINS DAN TEKNOLOGI

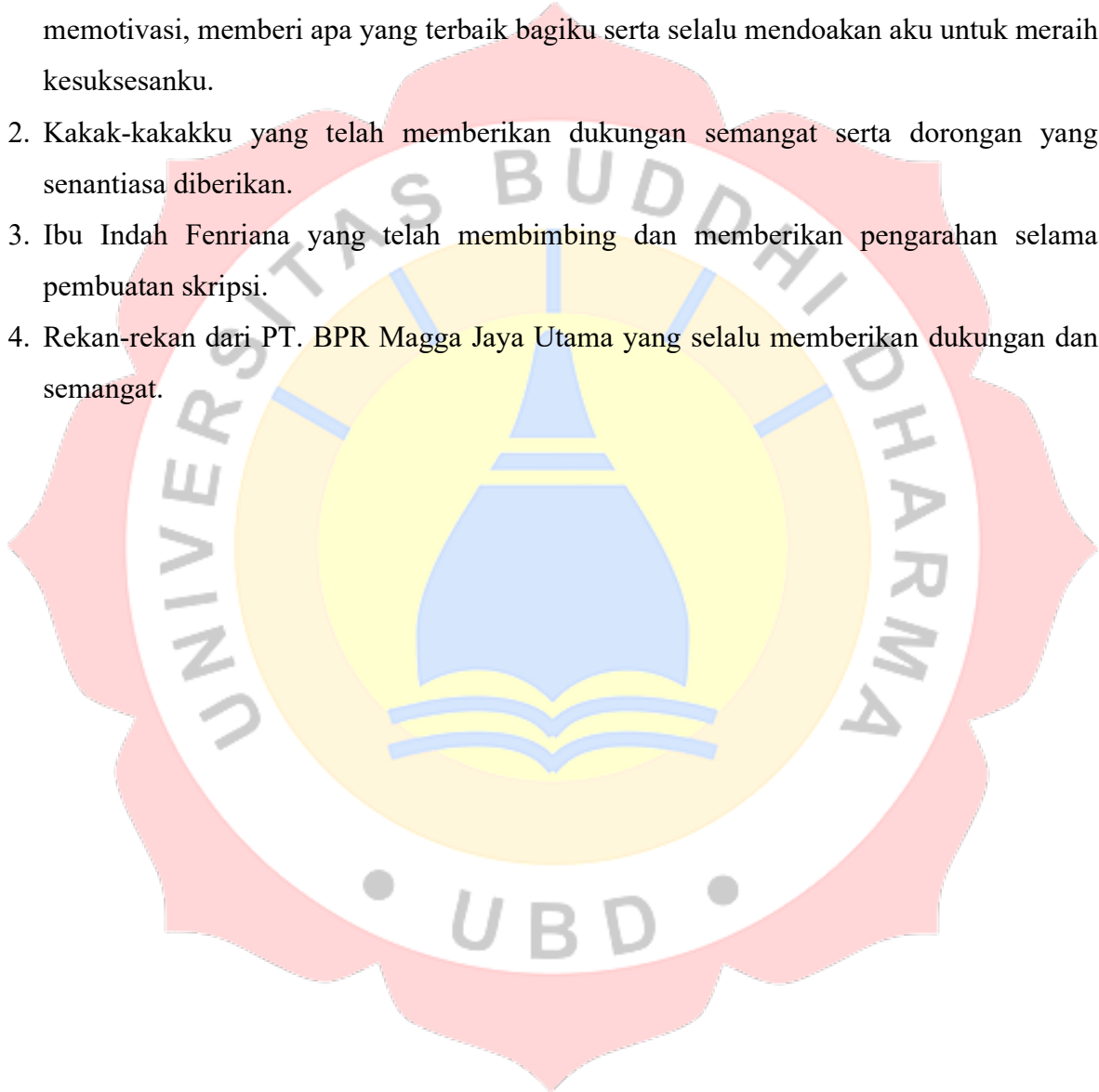
**UNIVERSITAS BUDDHI DHARMA
TANGERANG**

2025

LEMBAR PERSEMBAHAN

Dengan mengucapkan puji syukur kepada Tuhan Yang Maha Esa, Skripsi ini kupersembahkan untuk:

1. Kedua orang tua saya yang telah membesarkan aku dan selalu membimbing, mendukung, memotivasi, memberi apa yang terbaik bagiku serta selalu mendoakan aku untuk meraih kesuksesanku.
2. Kakak-kakakku yang telah memberikan dukungan semangat serta dorongan yang senantiasa diberikan.
3. Ibu Indah Fenriana yang telah membimbing dan memberikan pengarahan selama pembuatan skripsi.
4. Rekan-rekan dari PT. BPR Magga Jaya Utama yang selalu memberikan dukungan dan semangat.



LEMBAR PERNYATAAN KEASLIAN SKRIPSI

Yang bertanda tangan di bawah ini.

NIM : 20201000039
Nama : Cornellius Orlando Christiadi
Jenjang Studi : Strata 1
Program Studi : Teknik Informatika
Peminatan : *Database Development*

Dengan ini saya menyatakan bahwa :

1. Skripsi ini adalah asli dan belum pernah diajukan untuk mendapatkan gelar akademik Sarjana atau kelengkapan studi, baik di Universitas Buddhi Dharma maupun di Perguruan Tinggi lainnya.
2. Skripsi ini saya buat sendiri tanpa bantuan dari pihak lain, kecuali arahan dosen pembimbing.
3. Dalam Skripsi ini tidak terdapat karya atau pendapat yang telah ditulis atau dipublikasikan orang lain, kecuali secara tertulis dengan jelas dan dicantumkan sebagai acuan dalam naskah dengan disebutkan nama pengarang dan dicantumkan daftar Pustaka.
4. Dalam Skripsi ini tidak terdapat pemalsuan (kebohongan), seperti buku, artikel, jurnal, data sekunder, pengolahan data, dan pemalsuan tanda tangan dosen atau Ketua Program Studi Universitas Buddhi Dharma yang dibuktikan dengan keasliannya.
5. Lembar pernyataan ini saya buat dengan sesungguhnya, tanpa paksaan dan apabila dikemudian hari atau pada waktu lainnya terdapat penyimpangan dan ketidak beneran dalam pernyataan ini, saya bersedia menerima sanksi akademik berupa pencabutan gelar akademik yang telah saya peroleh karena Skripsi ini serta sanksi lainnya sesuai dengan peraturan dan norma yang berlaku.

Tangerang, 05-08-2025
Yang membuat pernyataan,



Cornellius Orlando Christiadi
20201000039

LEMBAR PERSETUJUAN PUBLIKASI KARYA ILMIAH

Yang bertanda tangan di bawah ini.

NIM	20201000039
Nama	: Cornellius Orlando Christiadi
Jenjang Studi	: Strata 1
Program Studi	: Teknik Informatika
Peminatan	: <i>Database Development</i>

Dengan ini menyetujui untuk memberikan izin kepada pihak Universitas Buddhi Dharma, Hak Bebas Royalti Non – Eksklusif (*Non-exclusive Royalty-Free Right*) atas karya ilmiah kami yang berjudul **“Perbandingan Algoritma Naïve Bayes dan Random Forest Untuk Klasifikasi Kelayakan Kredit Pada PT. BPR Magga Jaya Utama”**, beserta alat yang diperlukan (apabila ada).

Dengan Hak Bebas Royalti Non – Eksklusif ini pihak Universitas Buddhi Dharma berhak menyimpan, mengalih-media atau format-kan, mengelolanya dalam pangkalan data (*database*), mendistribusikannya, dan menampilkan atau mempublikasikannya di internet atau media lain untuk kepentingan akademis tanpa perlu meminta izin dari saya selama tetap mencantumkan nama saya sebagai penulis pertama atau pencipta karya ilmiah tersebut.

Saya bersedia untuk menanggung secara pribadi, tanpa melibatkan pihak Universitas Buddhi Dharma, segala bentuk tuntutan hukum yang timbul atas pelanggaran Hak Cipta dalam karya ilmiah saya ini.

Demikian pernyataan ini saya buat dengan sebenarnya.

Tangerang, 05-08-2025
Yang membuat pernyataan,



Cornellius Orlando Christiadi
20201000039

LEMBAR PENGESAHAN PEMBIMBING

PERBANDINGAN ALGORITMA *NAÏVE BAYES* DAN *RANDOM FOREST* UNTUK KLASIFIKASI KELAYAKAN KREDIT PADA PT. BPR MAGGA JAYA UTAMA

Dibuat Oleh:

NIM : 20201000039

Nama : Cornellius Orlando Christiadi

Telah disetujui untuk dipertahankan di hadapan Tim Penguji Ujian

Komprehensif

Program Studi Teknik Informatika

Peminatan Database Development

Tahun Akademik 2024/2025

Disahkan oleh,

Tangerang, 05-08-2025

Pembimbing,



(Indah Fenriana, S. Kom., M.Kom)

NUPTK. 6538766667230262

LEMBAR PENGESAHAN SKRIPSI

PERBANDINGAN ALGORITMA NAÏVE BAYES DAN RANDOM FOREST UNTUK KLASIFIKASI KELAYAKAN KREDIT PADA PT. BPR MAGGA JAYA UTAMA

Dibuat Oleh:

NIM : 20201000039

Nama : Cornellius Orlando Christiadi

Telah disetujui untuk dipertahankan di hadapan Tim Penguji Ujian
Komprehensif

Program Studi Teknik Informatika

Database Development

Tahun Akademik 2024/2025

Disahkan oleh,

Tangerang, 05-08-2025

Dekan,

Ketua Program Studi,


Dr. Yakub, M.Kom., M.M
NUPTK. 1836747648130172


Hartana Wijaya, S.Kom.,
NUPTK. 3844759660131192

LEMBAR PENGESAHAN TIM PENGUJI

Nama : Cornellius Orlando Christiadi
NIM : 20201000039
Fakultas : Sains dan Teknologi
Judul Skripsi : Perbandingan Algoritma *Naive Bayes* dan *Random Forest* Untuk Klasifikasi Kelayakan Kredit Pada PT. BPR Magga Jaya Utama

Dinyatakan LULUS setelah mempertahankan di depan Tim Penguji pada hari Selasa, 05/08/2025.

	Nama penguji :	Tanda Tangan :
Ketua Penguji :	Edy ST., M.Kom NUPTK. 7560760661130203	
Penguji I :	Hartana Wijaya M.Kom NUPTK. 3844759660131192	
Penguji II :	Indah Fenriana, S.Kom., M.Kom NUPTK. 6538766667230262	

Mengetahui,
Dekan Fakultas Sains dan Teknologi



Dr. Yakub, M.Kom., M.M.

NUPTK: 1836747648130172

KATA PENGANTAR

Dengan mengucapkan Puji Syukur kepada Tuhan Yang Maha Esa, yang telah memberikan Rahmat dan karunia-Nya kepada penulis sehingga dapat menyusun dan menyelesaikan skripsi ini dengan judul "**Perbandingan Algoritma *Naïve Bayes* dan *Random Forest* untuk Klasifikasi Kelayakan Kredit Pada PT. BPR Magga Jaya Utama**". Tujuan utama dari pembuatan skripsi ini adalah sebagai salah satu syarat kelengkapan dalam menyelesaikan program pendidikan Strata 1 Program Studi Teknik Informatika di Universitas Buddhi Dharma. Dalam penyusunan skripsi ini penulis banyak menerima bantuan dan dorongan baik moril maupun materiil dari berbagai pihak, maka pada kesempatan ini penulis menyampaikan rasa terima kasih yang sebesar-besarnya kepada:

1. Ibu Dr. Limajatini, S.E., M.M., B.K.P, sebagai Rektor Universitas Buddhi Dharma
2. Bapak Dr. Yakub, S.Kom., M.Kom., M.M, Dekan Fakultas Sains dan Teknologi
3. Bapak Hartana Wijaya, S.Kom., M.Kom, sebagai Ketua Program Studi Teknik Informatika
4. Ibu Indah Fenriana, S.Kom., M.Kom, sebagai pembimbing yang telah membantu dan memberikan dukungan serta harapan untuk menyelesaikan penulisan skripsi ini.
5. Orang tua dan keluarga yang selalu memberikan dukungan baik moril dan materiil.
6. Teman-teman yang selalu membantu dan memberikan semangat

Serta semua pihak yang terlalu banyak untuk disebutkan satu-persatu sehingga terwujudnya penulisan ini. Penulis menyadari bahwa penulisan skripsi ini masih belum sempurna, untuk itu penulis mohon kritik dan saran yang bersifat membangun demi kesempurnaan penulisan di masa yang akan datang.

Akhir kata semoga skripsi ini dapat berguna bagi penulis khususnya dan bagi para pembaca yang berminat pada umumnya.

Tangerang, 05 Agustus 2025

Penulis

PERBANDINGAN ALGORITMA *NAIVE BAYES* DAN *RANDOM FOREST* UNTUK KLASIFIKASI KELAYAKAN KREDIT PADA PT. BPR MAGGA JAYA UTAMA

115 Halaman + xviii / 23 Tabel / 62 Gambar / 4 Lampiran

ABSTRAK

Kredit menjadi salah satu fasilitas yang disediakan oleh perusahaan khususnya pada bidang keuangan dan cukup diminati oleh masyarakat. Tujuan menggunakan fasilitas tersebut adalah memenuhi keinginan atau kebutuhan baik individu maupun bisnis seperti modal usaha, investasi atau konsumsi pribadi. Penyaluran kredit kepada masyarakat terus meningkat setiap tahunnya dan penyaluran kredit oleh Bank Perkreditan Rakyat (BPR) sejalan dengan meningkatnya rasio *Non-Performing Loan (NPL)* yang berdampak pada kelayakan kredit debitur. Penelitian ini bertujuan untuk membandingkan algoritma *Naive Bayes* dan *Random Forest* dalam mengklasifikasikan kelayakan kredit pada PT. BPR Magga Jaya Utama. Sehingga dapat memberikan kemudahan dalam membuat keputusan kelayakan kredit calon debitur dan meminimalisir meningkatnya rasio *NPL*. Metode yang digunakan meliputi data *selection, cleaning, transformation*, dan pemodelan menggunakan *RapidMiner* dan implementasi *website* dengan *PHP Code Igniter*. *Dataset* yang belum diproses berjumlah 6983 data, dan setelah melakukan beberapa proses data yang digunakan berjumlah 5553 data nasabah, dibagi menjadi data *training* dan data *testing* dengan rasio 70:30. Evaluasi dilakukan menggunakan metrik akurasi, presisi, dan recall. Hasil menunjukkan bahwa algoritma *Random Forest* memberikan akurasi lebih tinggi (96,13% menggunakan *RapidMiner*, 95,50% menggunakan *website*) dibandingkan *Naive Bayes* (89,26% di *RapidMiner*, 89,68% menggunakan *website*). Penelitian ini menyimpulkan bahwa *Random Forest* lebih efektif dalam klasifikasi kelayakan kredit dan dapat dijadikan acuan dalam pengambilan keputusan oleh pihak bank.

Kata kunci: *Data Mining, Kelayakan Kredit, Klasifikasi, Naive Bayes, Random Forest*

*COMPARISON OF NAÏVE BAYES AND RANDOM FOREST ALGORITHMS FOR
CREDIT ELIGIBILITY CLASSIFICATION AT PT. BPR MAGGA JAYA UTAMA*

115 Pages + xviii / 23 Tables / 62 Images / 4 Attachment

ABSTRACT

Credit is one of the facilities provided by companies, especially in the financial sector, and is quite popular among the public. The purpose of using this facility is to fulfill the desires or needs of individuals or businesses, such as business capital, investment, or personal consumption. The distribution of credit to the public continues to increase every year, and credit distribution by Rural Banks (BPR) goes hand in hand with the rising ratio of Non-Performing Loans (NPL), which affects the creditworthiness of borrowers. This study aims to compare the Naive Bayes and Random Forest algorithms in classifying creditworthiness at PT. BPR Magga Jaya Utama. The goal is to facilitate decision-making regarding prospective borrowers' creditworthiness and to minimize the increasing NPL ratio. The methods used include data selection, cleaning, transformation, and modeling using RapidMiner, along with website implementation using PHP CodeIgniter. The unprocessed dataset consists of 6,983 records, and after several preprocessing steps, 5,553 customer records were used, split into training and testing datasets with a 70:30 ratio. Evaluation was conducted using accuracy, precision, and recall metrics. The results show that the Random Forest algorithm achieved higher accuracy (96.13% using RapidMiner, 95.50% using the website) compared to Naive Bayes (89.26% in RapidMiner, 89.68% using website). This study concludes that Random Forest is more effective in classifying creditworthiness and can serve as a reference in decision-making by the bank.

Keywords: Data Mining, Credit Eligibility, Classification, Naive Bayes, Random Forest

DAFTAR ISI

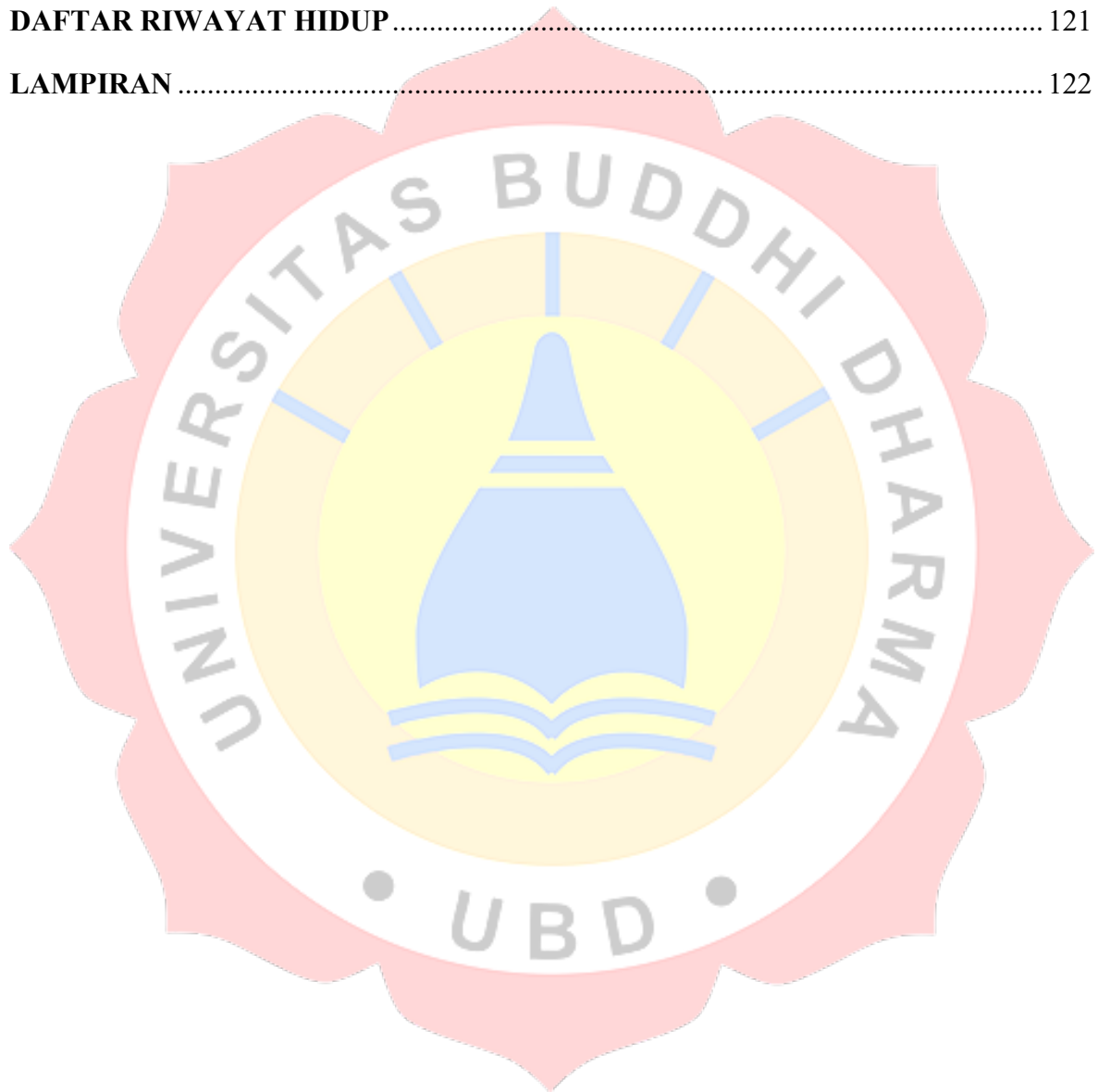
LEMBAR JUDUL DALAM SKRIPSI

LEMBAR PERSEMBAHAN	i
LEMBAR PERNYATAAN KEASLIAN SKRIPSI	ii
LEMBAR PERSETUJUAN PUBLIKASI KARYA ILMIAH	iii
LEMBAR PENGESAHAN PEMBIMBING	iv
LEMBAR PENGESAHAN SKRIPSI	v
LEMBAR PENGESAHAN TIM PENGUJI	vi
KATA PENGANTAR	vii
ABSTRAK	viii
<i>ABSTRACT</i>	ix
DAFTAR ISI	x
DAFTAR GAMBAR	xiv
DAFTAR TABEL	xvi
DAFTAR LAMPIRAN	xvii
BAB I PENDAHULUAN	1
1.1 Latar Belakang Masalah	1
1.2 Rumusan Masalah	4
1.3 Ruang Lingkup	4
1.4 Tujuan dan Manfaat	5
1.4.1 Tujuan	5
1.4.2 Manfaat	5
1.5 Metode Penelitian	6
1.6 Sistematika Penulisan	7
BAB II LANDASAN TEORI	9
2.1 Teori Umum	9
2.1.1 Data	9

2.1.2	Informasi.....	12
2.2	Teori Khusus.....	13
2.2.1	<i>Data Mining</i>	13
2.2.2	<i>Machine Learning</i>	18
2.2.3	Klasifikasi.....	19
2.2.4	Kredit.....	20
2.3	Teori Rancangan.....	21
2.3.1	Metode <i>Naïve Bayes</i>	21
2.3.2	Metode <i>Random Forest</i>	23
2.3.3	<i>RapidMiner</i>	24
2.3.4	<i>Use Case Diagram</i>	26
2.3.5	<i>Activity Diagram</i>	28
2.3.6	<i>Website</i>	30
2.3.7	<i>PHP (Hypertext Preprocessor)</i>	30
2.3.8	<i>Framework Codeigniter</i>	31
2.3.9	<i>Blackbox Testing</i>	33
2.3.10	<i>MySQL</i>	34
2.4	Tinjauan Studi.....	35
BAB III METODOLOGI PENELITIAN.....		48
3.1	Tinjauan Umum Perusahaan.....	48
3.1.1	Sejarah Singkat.....	48
3.1.2	Visi, Misi, dan Tujuan.....	48
3.1.3	Struktur Organisasi.....	49
3.1.4	Tugas dan Wewenang.....	50
3.2	Analisa Kebutuhan	54
3.2.1	RE (<i>Requirement Elicitation</i>)	54
3.2.2	<i>Data Selection</i>	54

3.2.3	<i>Data Cleaning</i>	57
3.2.4	<i>Data Transformation</i>	59
3.2.5	<i>Data Splitting</i>	60
3.3	<i>Model Training</i>	61
3.3.1	<i>Naïve Bayes</i>	61
3.3.2	<i>Random Forest</i>	61
3.4	<i>Model Evaluasi</i>	62
3.4.1	<i>Cross Validation</i>	63
3.5	<i>Kerangka Pemikiran</i>	63
3.6	<i>Perancangam UML</i>	64
3.6.1	<i>Use Case Diagram</i>	64
3.6.2	<i>Activity Diagram</i>	66
3.7	<i>Perancangan Database dan Layar</i>	68
3.7.1	<i>Perancangan Database</i>	68
3.7.2	<i>Perancangan Layar</i>	69
3.8	<i>Jadwal Penelitian</i>	77
BAB IV HASIL DAN PEMBAHASAN		78
4.1	<i>Spesifikasi Hardware dan Software</i>	78
4.1.1	<i>Hardware</i>	78
4.1.2	<i>Software</i>	78
4.2	<i>Penerapan Metode</i>	79
4.2.1	<i>Naïve Bayes</i>	79
4.2.2	<i>Random Forest</i>	80
4.3	<i>Implementasi Website</i>	80
4.4	<i>RapidMiner</i>	92
4.5	<i>Black Box</i>	97
4.6	<i>Kuesioner</i>	103

4.7	Pembahasan	112
BAB V SIMPULAN DAN SARAN		114
5.1	Simpulan.....	114
5.2	Saran	115
DAFTAR PUSTAKA		116
DAFTAR RIWAYAT HIDUP		121
LAMPIRAN		122



DAFTAR GAMBAR

Gambar 2. 1 Tahap-tahap dalam <i>data mining</i>	16
Gambar 2. 2 Proses <i>Mining</i>	17
Gambar 2. 3 Skema <i>machine learning</i>	19
Gambar 2. 4 Alur Kerja <i>Code Igniter</i>	31
Gambar 3. 1 Struktur Organisasi	49
Gambar 3. 2 Data Nasabah Kredit PT. BPR Magga Jaya Utama.....	55
Gambar 3. 3 Data setelah pemilihan atribut	57
Gambar 3. 4 Data yang terdapat nilai kosong atau <i>null</i>	58
Gambar 3. 5 Data setelah dilakukan <i>data cleaning</i>	58
Gambar 3. 6 Data status setelah diubah.....	60
Gambar 3. 7 Kerangka Pemikiran	63
Gambar 3. 8 <i>Use Case Diagram</i>	64
Gambar 3. 9 <i>Activity Diagram</i>	66
Gambar 3. 10 Perancangan <i>Database</i>	68
Gambar 3. 11 Halaman <i>Login</i>	69
Gambar 3. 12 Halaman <i>Dashboard Admin</i>	70
Gambar 3. 13 Halaman <i>Dataset Admin</i>	70
Gambar 3. 14 Halaman <i>Model Training Admin</i>	71
Gambar 3. 15 Halaman <i>Nasabah Admin</i>	71
Gambar 3. 16 Halaman <i>Prediksi Cepat Admin</i>	72
Gambar 3. 17 Halaman <i>Menu Admin</i>	72
Gambar 3. 18 Halaman <i>Role Admin</i>	73
Gambar 3. 19 Halaman <i>Permission Admin</i>	73
Gambar 3. 20 Halaman <i>User Admin</i>	74
Gambar 3. 21 Halaman <i>Divisi Admin</i>	74
Gambar 3. 22 Halaman <i>Dashboard User</i>	75
Gambar 3. 23 Halaman <i>Data Nasabah User</i>	75
Gambar 3. 24 Halaman <i>Prediksi</i>	76
Gambar 3. 25 <i>Jadwal Penelitian</i>	77
Gambar 4. 1 Halaman <i>Login User</i>	80
Gambar 4. 2 Halaman <i>Awal Setelah Login</i>	81

Gambar 4. 3 Halaman Menu <i>Dataset</i>	81
Gambar 4. 4 Halaman Perhitungan.....	82
Gambar 4. 5 Halaman Data Nasabah.....	83
Gambar 4. 6 Halaman Hasil Prediksi	83
Gambar 4. 7 Halaman Parameter Divisi.....	84
Gambar 4. 8 Halaman Input Nasabah Baru.....	85
Gambar 4. 9 Halaman Hasil Input	85
Gambar 4. 10 Halaman Ubah Data.....	86
Gambar 4. 11 Halaman Prediksi Cepat	87
Gambar 4. 12 Halaman Menu.....	87
Gambar 4. 13 Halaman Ubah Akses Parameter	88
Gambar 4. 14 Halaman Parameter <i>Roles</i>	89
Gambar 4. 15 Halaman Tambah <i>Roles</i>	89
Gambar 4. 16 Halaman Tambah <i>Permission</i> Baru	90
Gambar 4. 17 Halaman Parameter <i>User</i>	90
Gambar 4. 18 Halaman Menambahkan <i>User</i> Baru.....	91
Gambar 4. 19 Tampilan Seluruh Proses <i>RapidMiner</i>	92
Gambar 4. 20 Alur Proses pada Operator <i>CrossValidation</i> Naïve Bayes	93
Gambar 4. 21 Alur Proses <i>CrossValidation</i> Random Forest.....	94
Gambar 4. 22 Pengaturan Operator Split Data.....	95
Gambar 4. 23 Pengaturan Operator Filter <i>Examples</i> Macet.....	96
Gambar 4. 24 Pengaturan Operator Filter <i>Example</i> Lancar	97
Gambar 4. 25 Pertanyaan 1.....	103
Gambar 4. 26 Pertanyaan 2.....	104
Gambar 4. 27 Pertanyaan 3.....	105
Gambar 4. 28 Pertanyaan 4.....	106
Gambar 4. 29 Pertanyaan 5.....	107
Gambar 4. 30 Pertanyaan 6.....	108
Gambar 4. 31 Pertanyaan 7.....	109
Gambar 4. 32 Pertanyaan 8.....	110
Gambar 4. 33 Pertanyaan 9.....	111

DAFTAR TABEL

Tabel 2. 1 Keterangan Parameter <i>Naïve Bayes</i>	22
Tabel 2. 2 Simbol <i>Use Case Diagram</i>	26
Tabel 2. 3 Simbol <i>Activity Diagram</i>	29
Tabel 2. 4 Tinjauan Studi	35
Tabel 3. 1 RE (<i>Requirement Elicitation</i>).....	54
Tabel 3. 2 Keterangan Atribut untuk data yang digunakan.....	55
Tabel 3. 3 Keterangan atribut kolektibilitas	60
Tabel 3. 4 Jumlah Data <i>Training</i> dan Data <i>Testing</i>	61
Tabel 4. 1 Spesifikasi Perangkat.....	78
Tabel 4. 2 Spesifikasi Aplikasi dan Sistem	79
Tabel 4. 3 Hasil <i>Naïve Bayes</i> pada <i>RapidMiner</i> dan Website.....	79
Tabel 4. 4 Hasil <i>Random Forest</i> pada <i>RapidMiner</i> dan Website	80
Tabel 4. 5 Tabel Pengujian <i>Black Box</i>	98
Tabel 4. 6 Pertanyaan 1	103
Tabel 4. 7 Pertanyaan 2	104
Tabel 4. 8 Pertanyaan 3	105
Tabel 4. 9 Pertanyaan 5	106
Tabel 4. 10 Pertanyaan 5	107
Tabel 4. 11 Pertanyaan 6	108
Tabel 4. 12 Pertanyaan 7	109
Tabel 4. 13 Pertanyaan 8	110
Tabel 4. 14 Pertanyaan 9	111
Tabel 4. 15 Hasil Responden Pengguna Program	112

DAFTAR LAMPIRAN

Lampiran 1 <i>Requirement Elicitation</i>	122
Lampiran 2 Hasil Kuesioner.....	123
Lampiran 3 Surat Permohonan Penelitian.....	124
Lampiran 4 Kartu Bimbingan TA/Skripsi.....	125



BAB 1

PENDAHULUAN

1.1 Latar Belakang Masalah

Kredit perbankan merupakan fasilitas yang disediakan oleh Bank atau Lembaga Keuangan lainnya dalam bentuk pinjaman kepada individu atau badan usaha. Pinjaman ini diberikan dengan syarat bahwa penerima pinjaman akan mengembalikan jumlah yang dipinjam beserta bunga atau biaya lainnya dalam jangka waktu tertentu (Dasari et al., 2023). Kredit perbankan menjadi salah satu alat utama dalam kegiatan ekonomi, karena memungkinkan individu atau bisnis untuk mengakses dana atau modal tambahan yang mereka butuhkan untuk berbagai keperluan, seperti investasi, modal usaha atau konsumsi pribadi. Pada periode 2019 hingga 2023, Otoritas Jasa Keuangan mengeluarkan data Statistik Perbankan Indonesia, pemberian kredit oleh Bank Perkreditan Rakyat (BPR) mengalami kenaikan sebesar 29,63%, rasio ini menunjukkan bahwa jumlah kredit yang disalurkan bertumbuh secara signifikan selama periode tersebut (“Stat. Perbank. Indones.,” 2023).

Meskipun pemberian kredit terus meningkat setiap tahunnya, tetapi masih terdapat masalah seperti macetnya kredit yang terjadi pada nasabah sehingga menimbulkan dampak untuk perbankan yang ditandai dengan *ratio NPL (Ratio Non Performing Loan)*. *Ratio NPL* merupakan rasio yang menunjukkan kualitas sebuah Bank atau Lembaga Keuangan Lainnya dalam penyaluran kredit, yang dimana semakin rendah rasio *NPL* maka menggambarkan bahwa Bank atau Lembaga keuangan lainnya berstatus tidak bermasalah dalam pengembalian kredit oleh banyak debitur, tetapi sebaliknya jika rasio *NPL* tinggi maka berdampak pada jumlah kredit yang dapat disalurkan karena rasio pengembalian kredit yang telah disalurkan kepada debitur mengalami masalah (Gayo et al., 2022). Berdasarkan Data Statistik *Non Performing Loan* yang dikeluarkan oleh Otoritas Jasa Keuangan, Salah satu dari bank yang memberikan kredit ke masyarakat adalah Bank Perkreditan Rakyat (BPR)

yang cukup tinggi Rasio *NPL*-nya. Rasio *NPL* pada Bank Perkreditan Rakyat terus meningkat dari tahun 2019 hingga 2023 sebesar 71,98% (“Stat. Perbank. Indones.,” 2023). Hal ini menunjukkan bahwa sektor Lembaga keuangan mikro, khususnya Bank Perkreditan Rakyat (BPR) menghadapi tantangan dalam mengelola risiko kredit karena peningkatan rasio *NPL* yang perlu segera ditangani dalam manajemen kredit.

Untuk membantu dalam meminimalisir tingkat rasio *Non Performing Loan*, maka perlu dikembangkan sebuah model *Credit Scoring* untuk memprediksi dan menilai seorang calon debitur dalam menentukan kelayakannya ketika menerima fasilitas kredit ini. Untuk mengembangkan sebuah model *Credit Scoring*, maka diperlukan sebuah alat yang nantinya dapat digunakan untuk melakukan penilaian dan memprediksi. Skor kredit merupakan evaluasi yang harus dipertimbangkan oleh pemberi pinjaman sebelum menyalurkan dana pinjaman kepada peminjam.

Dari paragraf sebelumnya, dapat disimpulkan bahwa skor kredit adalah nilai yang dikumpulkan pemberi pinjaman dari orang-orang yang memasukkan detail permohonan pinjamannya. *Data Mining* menjadi salah satu metode yang dapat digunakan dalam penelitian yang dilakukan. *Data Mining* menurut (Sudirman et al., 2018) adalah proses yang menggunakan teknik-teknik dalam menganalisis data dan alat untuk mencari pola-pola yang sebelumnya belum diketahui.

Dalam *Data Mining* terdapat *Machine Learning* yang mampu untuk menganalisis data Calon Debitur sehingga membantu pemutus kredit mendapatkan informasi sebagai acuan untuk memutuskan apakah calon debitur tersebut layak untuk diberikan kredit atau tidak. Penggunaan *Machine Learning* yang tepat akan menghindarkan kesalahan pengambilan Keputusan akibat ketidakakuratan data akan penolakan kredit yang tidak akurat. Penolakan kredit akan berdampak pada kesempatan bisnis yang berpotensi hilang, menghambat pertumbuhan dan pengembangan usaha bagi calon debitur. Dalam hal ini

diperlukan sebuah metode yang tepat untuk menganalisis skor kredit untuk menghindari kesalahan.

Dengan menerapkan berbagai Teknik dan algoritma analisis data, *Machine Learning* memungkinkan kita untuk menggali potensi yang terkandung dalam dataset yang besar dan tentunya cukup kompleks, yang pada gilirannya dapat digunakan dalam berbagai aplikasi dalam kehidupan nyata. Dengan demikian, *Machine Learning* tidak hanya menghasilkan informasi yang berguna, tetapi juga memberikan wawasan yang mendalam dan informasi berharga sebagai pendukung dalam mengambil keputusan bagi bank untuk memutus kredit secara tepat.

Machine Learning memiliki banyak metode klasifikasi beberapa diantaranya adalah *Naïve Bayes* dan *Random Forest*, *Naïve Bayes* ialah klasifikasi yang didasarkan pada probabilitas dan statistik yang memprediksi kemungkinan di masa depan berdasarkan pengalaman atau masa yang sudah terjadi sehingga disebut sebagai Teorema Bayes dan secara alami dapat menangani data yang bersifat kategorial karena mengasumsikan independensi antara atribut-atribut (Ramadhan et al, 2022). Sedangkan *Random Forest* merupakan metode berbasis *ensemble decision tree* yang dimana mampu untuk menangani data dengan variabel yang kompleks dan menghasilkan akurasi tinggi. Hal ini membuatnya efektif digunakan dalam penelitian ini karena data kredit perbankan terdiri dari banyak atribut-atribut kategorial seperti jumlah plafond pengajuan, jangka waktu kredit, tujuan penggunaan, data pekerjaan, hingga data jaminan yang akan diserahkan.

Beberapa penelitian yang dilakukan sebelumnya menyatakan bahwa tingkat akurasi algoritma *Random Forest* lebih tinggi dibandingkan *Naïve Bayes*, seperti penelitian yang dilakukan oleh (Depari, 2022) menghasilkan tingkat akurasi *Naïve Bayes* sebesar 71% dan *Random Forest* sebesar 75%, kemudian (Indaryono et al., 2024) memberikan kesimpulan pada penelitiannya bahwa *Random Forest* menghasilkan persentase performa dan akurasi

sebesar 86.55%, sedangkan *Naïve Bayes* sebesar 36.61% dan jika dilihat selisih dari kedua algoritma tersebut sebesar 49.94%.

“Perbandingan Algoritma *Naïve Bayes* dan *Random Forest* untuk Klasifikasi Kelayakan Kredit Pada PT. BPR Magga Jaya Utama” menjadi judul dalam melakukan penelitian ini, sesuai dengan latar belakang yang sudah dijelaskan sebelumnya.

1.2 Rumusan Masalah

Berikut adalah permasalahan yang teridentifikasi berdasarkan dari penjabaran latar belakang masalah:

- a. Bagaimana dampak dari peningkatan rasio *NPL (Non Performing Loan)* terhadap kemampuan penyaluran kredit baru?
- b. Bagaimana cara untuk meminimalisir rasio *NPL (Non Performing Loan)* melalui sebuah sistem penilaian kelayakan kredit?
- c. Algoritma apa yang dapat membantu antara *Naive Bayes* atau *Random Forest* dalam memutuskan kelayakan calon debitur yang efektif dan efisien?
- d. Dari kedua algoritma yang digunakan, yaitu *Naive Bayes* dan *Random Forest*. Algoritma mana yang dapat memberikan performa dan tingkat akurasi yang baik dalam klasifikasi kelayakan kredit?

1.3 Ruang Lingkup

Ruang lingkup dalam penelitian ini adalah:

- a. Sumber data yang digunakan diambil dari histori kredit pada PT. BPR Magga Jaya Utama, dengan rentang waktu dari tahun 2019 hingga 2023.
- b. Data kredit yang digunakan memiliki batasan yang berkaitan dengan data pribadi nasabah.

- c. Memanfaatkan *Naïve Bayes* dan *Random Forest* sebagai metode dalam proses klasifikasi
- d. Dalam proses pembuatan program, PHP menjadi bahasa yang digunakan.

1.4 Tujuan dan Manfaat

1.4.1 Tujuan

Berikut adalah tujuan dari dilakukannya penelitian ini:

- a. Memberikan PT. BPR Magga Jaya Utama sebuah sistem yang tepat untuk menilai kelayakan calon debitur dengan lebih akurat, sehingga dapat mengurangi kenaikan rasio *NPL*.
- b. Memberikan kemudahan bagi pemutus kredit dalam memutuskan kelayakan kredit calon debitur.
- c. Mendorong pengembangan sistem evaluasi kredit yang lebih maju dan adaptif.
- d. Membandingkan performa dan tingkat akurasi dari kedua algoritma dalam mengklasifikasikan kelayakan kredit nasabah PT. BPR Magga Jaya Utama.
- e. Untuk mengetahui algoritma mana yang lebih efektif beserta kelebihan dan kekurangan dalam mengambil keputusan kelayakan kredit nasabah.

1.4.2 Manfaat

Tujuan yang sudah disebutkan sebelumnya tentunya memiliki manfaat yang dapat diberikan kepada PT. BPR Magga Jaya Utama. Berikut adalah manfaat yang akan diberikan:

- a. Meningkatkan efektivitas dan efisiensi pada PT. BPR Magga Jaya Utama dalam mengelola kredit dengan dengan cara meminimalisir risiko kredit *NPL*, sehingga dapat meningkatkan kredit yang dapat disalurkan dan profitabilitas perusahaan.

- b. Memastikan keakuratan data yang digunakan dalam pengambilan keputusan kredit, sehingga dapat mengurangi potensi kesalahan yang timbul akibat data yang tidak tepat.
- c. Membantu calon debitur dalam mendapatkan kredit dengan lebih mudah dan cepat, karena adanya sistem yang mampu menilai kelayakan mereka secara lebih akurat, sehingga meningkatkan akses terhadap penambahan modal bagi pengembangan usaha calon debitur.
- d. Dapat mengetahui performa dan tingkat akurasi dari algoritma *Naïve Bayes* serta *Random Forest*.
- e. Dapat memberikan gambaran kelebihan dan kekurangan dari algoritma *Naïve Bayes* dan *Random Forest* dalam mengklasifikasikan kelayakan kredit.

1.5 Metode Penelitian

a. Data Historis

Data historis merupakan data yang telah tercatat dan tersedia pada perusahaan. Data PT. BPR Magga Jaya Utama dari tahun 2019 hingga 2023 dijadikan sebagai bahan dalam penelitian ini.

b. Wawancara

Teknik ini dilakukan dengan kepala bagian kredit dan general manager kredit PT. BPR Magga Jaya Utama untuk mendapatkan pemahaman tentang faktor-faktor yang mempengaruhi penilaian kelayakan kredit calon debitur. Hasil dari wawancara akan diolah sebagai RE (*Requirement Elicitation*).

c. Metode Kuesioner

Untuk mengumpulkan data selanjutnya adalah dengan menggunakan kuesioner yang diberikan kepada pengguna sistem dengan tujuan memperoleh informasi terkait penggunaan sistem yang telah dibuat. Pengambilan data kuesioner bersifat *online* dengan menggunakan *Google Form*.

d. Studi Pustaka

Melalui berbagai media yang dimana dapat mendukung penelitian ini sebagai sumber informasi, seperti jurnal (baik nasional maupun internasional), buku elektronik, internet, dan sumber lainnya.

1.6 Sistematika Penulisan

Sistematika penulisan pada penelitian ini dibagi menjadi beberapa sub bab yang berisi materi-materi penelitian. Berikut pembagian sub bab:

BAB I PENDAHULUAN

Bab I menyajikan dasar terkait dilakukannya penelitian, rumusan masalah berdasarkan dasar permasalahan, batasan dalam penelitian, tujuan serta manfaat dari penelitian, teknik pengumpulan data yang digunakan, dan sistematika penulisan.

BAB II LANDASAN TEORI

Memuat lebih dari 1 teori yang berupa pengertian dan pandangan dari para ahli yang diambil dari literatur seperti buku dan jurnal yang saling berkaitan dengan topik penelitian dijadikan kutipan untuk mendukung penelitian ini, diletakkan pada Bab II.

BAB III METODOLOGI PENELITIAN

Bab III ini membahas terkait profil dari perusahaan yang diteliti, pengolahan data yang digunakan, dan pendekatan metode yang diterapkan.

BAB IV HASIL DAN PEMBAHASAN

Setelah melakukan perancangan baik data maupun aplikasi (program), hasil yang didapat diletakkan pada Bab IV. Hasil berupa pembahasan terkait perangkat yang digunakan, penjelasan setiap metode setelah digunakan, pengujian aplikasi menggunakan metode dan fungsi-fungsi tombol yang ada, serta pembahasan setelah dilakukannya penelitian.

BAB V SIMPULAN DAN SARAN

Bab V memuat rangkuman hasil dari penelitian yang dibuat menjadi lebih singkat dan disebut sebagai simpulan dan saran sebagai masukan yang dapat bermanfaat untuk penelitian yang akan datang.

BAB II

LANDASAN TEORI

2.1 Teori Umum

Setelah pemaparan dasar dari dilakukannya penelitian, dilakukan pencarian teori secara teoretis yang mendasari penelitian sehingga dapat menjadi pendukung penelitian ini.

2.1.1 Data

Data merupakan gabungan dari beberapa fakta yang digunakan sebagai *input* atau masukan yang diproses, kemudian memberikan sebuah informasi. Data dimanfaatkan sebagai bahan dalam melakukan diskusi dan menetapkan keputusan. Data dapat berupa gambar, suara, ataupun video, tidak hanya berbentuk huruf dan angka. Data juga merupakan sumber daya yang cukup penting di dalam sebuah organisasi yang dikelola sama halnya seperti mengelola aset penting dalam berbisnis. Tanpa data yang akurat terkait aktivitas baik internal maupun eksternal, perusahaan tidak akan bertahan lama atau bisa disebut juga perusahaan tersebut akan menghadapi tantangan yang besar kedepannya. Data memiliki dasar-dasar konsep, dimulai dari karakter huruf, karakter angka, atau karakter yang berupa simbol, sekelompok *field* yang saling berhubungan dapat membentuk sebuah catatan yang nantinya akan menjadi sebuah *file* jika catatan tersebut terkumpul (Susanto, 2015).

Data merupakan sebuah fakta yang dapat dijadikan bukti dari suatu kejadian pada waktu tertentu. Berikut ciri-ciri data yang digunakan (Lubis, 2016) :

1. Data dapat disimpan secara *integrated* atau terintegrasi dalam sebuah *database* yang memungkinkan penyimpanan secara menyeluruh, dimana semua bagian saling terhubung.
2. Data yang disimpan dalam sistem memungkinkan untuk diakses secara bersamaan oleh semua pengguna di waktu yang sama. Setiap pengguna dapat

melihat atau modifikasi data secara bersamaan tanpa mengganggu pengguna yang lain.

Data memiliki klasifikasi berdasarkan jenis, sifat dan sumber. Sehingga dapat dikatakan bahwa data adalah “Realitas yang menunjukkan suatu peristiwa dan kondisi yang sudah terjadi di kehidupan nyata”. Berikut klasifikasi yang dimiliki oleh data (Harismawan, 2019), sebagai berikut:

A. Jenis Data:

1. *Counting Data* disebut sebagai data hitung yang memiliki pengertian jenis data yang bersifat kuantitatif, yang dimana diperoleh melalui proses dari perhitungan atau jumlah sebuah objek, kejadian atau kategori tertentu, nilai dapat berupa bilangan bulat tidak memiliki nilai desimal.
2. *Measurement Data* disebut sebagai data ukur dengan pengertian jenis data yang bersifat kuantitatif, diperoleh dari hasil pengukuran, nilai dapat berupa bilangan pecahan atau desimal.

B. Sifat Data:

1. *Quantitative Data* adalah data yang bersifat kuantitatif dan berhubungan dengan angka atau nilai numerik yang dapat dihitung atau diukur secara sistematis. Data yang ada biasanya digunakan untuk menggambarkan kuantitas, jumlah, ukuran dari sebuah objek. Biasanya seperti penjumlahan, pengurangan, rata-rata dan perhitungan statistic lainnya.
2. *Qualitative Data* adalah data yang bersifat kualitatif dan tidak berhubungan dengan angka atau nilai numerik, melainkan menjelaskan atau mengelompokkan sebuah informasi berdasarkan kualitas atau sifat dari objek tersebut.

C. Sumber Data:

1. *Internal Data* merupakan data yang sifatnya asli, data tersebut diperoleh secara langsung yang berasal dari observasi, eksperimen, atau dokumentasi milik sendiri, bukan milik orang lain. Dikumpulkan secara langsung sesuai dengan kebutuhan.
2. *External Data* merupakan data yang bukan dari hasil observasi langsung oleh pihak sendiri atau internal. Data biasanya diambil dari karya orang lain yang sudah dipublikasikan secara terbuka. Jenis data eksternal terbagi menjadi 2 (dua), yaitu:
 - a. *Primary External Data* merupakan data eksternal yang jenisnya primer, dimana data berasal dari hasil karya orang lain, namun dilakukan secara langsung oleh peneliti baik dalam bentuk wawancara atau tertulis. Validitasnya cukup tinggi karena diperoleh secara langsung dari penelitinya, meskipun dari luar.
 - b. *Secondary External Data* merupakan data eksternal yang jenisnya sekunder, dimana data yang diperoleh dari karya orang lain dan bukan dari orang yang melakukan observasi, tetapi melalui seseorang atau orang lain. Dapat disebut juga sebagai hasil olahan atau ringkasan dari pihak ketiga.

Berdasarkan penjabaran dari beberapa sumber diatas, disimpulkan bahwa data adalah sebuah aset yang cukup penting bagi sebuah perusahaan untuk membuat perusahaan terus tumbuh dan berkembang, dimana data terdiri dari sekumpulan fakta dari sebuah peristiwa yang tidak hanya berupa karakter angka dan karakter huruf, namun bisa berupa visual, ataupun audio yang perlu diproses untuk memberikan sebuah informasi yang lebih bermanfaat.

2.1.2 Informasi

Informasi adalah data yang telah diolah menjadi sebuah pernyataan atau keterangan yang memiliki arti, dan menjadikan sebuah pesan, atau penjelasan yang dapat kita baca dan dengar. Sehingga memberikan pengetahuan kepada penerimanya. Disampaikan dalam beragam bentuk, baik melalui teknologi digital maupun cetak. (Ermawelis, 2018).

Informasi memiliki peran yang cukup penting dalam dunia bisnis dan informasi, yang dianggap setara dengan modal, bahan baku, dan tenaga kerja untuk mendukung kebutuhan internal dan keberlangsungan hidup organisasi atau perusahaan. Faktor-faktor yang dijadikan sebagai kualitas sebuah informasi, yaitu:

1. Relevansi, informasi yang disajikan harus sesuai dan berhubungan langsung dengan tujuan atau kebutuhan organisasi atau individu. Sehingga memberikan kontribusi yang nyata dalam mengambil keputusan dan informasi yang digunakan benar sesuai dengan topik yang dibahas.
2. Tepat Waktu, informasi yang berkualitas adalah informasi yang tersedia ketika dibutuhkan sehingga menjadi lebih bermanfaat. Ketepatan waktu cukup penting, karena informasi yang terlambat akan mengurangi nilai dan kemungkinan besar tidak lagi berguna.
3. Akurat/akurasi, mengacu pada ketepatan sebuah informasi yang disampaikan, akurat apabila bebas dari kesalahan, tidak menyesatkan, dan proses pengujian atau diverifikasi lebih dari satu pihak dengan hasil yang sama dan konsisten.
4. Lengkap, disajikan secara menyeluruh tanpa ada bagian penting pada data yang dilewatkan untuk memahami sebuah situasi atau mengambil keputusan. Informasi yang tidak lengkap dapat menimbulkan kesalahpahaman.

Informasi adalah data yang telah dikonversikan, fakta yang telah diolah dengan cara tertentu ke dalam sebuah konteks dimana memiliki makna dan berguna bagi penerimanya karena dari pengalaman atau instruksi menjadikan pengetahuan yang baru. (Anggraeni & Irviani, 2017).

Dapat penulis simpulkan bahwa informasi dapat menjadi lebih berguna bagi si penerima, karena memiliki nilai dan makna yang berasal dari pengolahan data yang sesuai dengan fakta-fakta yang telah dikumpulkan.

2.2 Teori Khusus

2.2.1 *Data Mining*

Data Mining ialah sebuah tahapan yang menggunakan sejumlah teknik dalam menganalisis data dan alat untuk menemukan hubungan dan pola yang tersembunyi (Sudirman et al., 2018). *Data mining* dapat diterapkan untuk mendapatkan nilai tambah dari kumpulan-kumpulan data yang selama ini belum diketahui dengan cara manual (Wanto, M Kom et al., 2020). *Data Mining* sebagai salah satu cabang ilmu komputer yang memiliki manfaat untuk menemukan pola-pola yang tersembunyi dari kumpulan data yang cukup besar, dengan menggabungkan metode dari *artificial intelligence* (kecerdasan buatan), *machine learning* (pembelajaran mesin), *statistics* (statistik), dan *database systems* (sistem basis data). Terbagi menjadi dua, yaitu deskriptif digunakan untuk menjelaskan karakteristik data dan prediktif digunakan untuk melakukan perkiraan suatu kejadian di masa depan (Suyanto, 2017).

Data mining menerapkan beberapa metode didalamnya, berikut adalah metode yang terdapat:

a. *Classification*

Proses yang bertujuan untuk menghasilkan sejumlah model atau pola yang mampu menjelaskan dan membedakan kelas-kelas data yang ada, sehingga dapat digunakan untuk memperkirakan atau mengklasifikasikan suatu objek atau kelas yang belum ada. Dikembangkan dengan menganalisa data pelatihan (*data training*), dan untuk mengevaluasi tingkat keakuratan model yang dihasilkan akan menggunakan data uji (*data testing*).

b. *Clustering*

Metode dalam *data mining* yang bertujuan untuk mengelompokkan data tidak memiliki label kelasnya (*unlabeled data*) ke dalam beberapa kelompok, dimana memiliki karakteristik yang serupa satu sama lain.

c. *Association*

Metode ini merupakan teknik *data mining* yang bertujuan untuk menghasilkan sejumlah *rule* (aturan) atau pola yang sering terjadi diantara item-item dalam sebuah kumpulan data yang dapat menghasilkan penjelasan terkait sejumlah data yang terhubung satu sama lain.

d. *Regression*

Metode prediktif yang mirip dengan klasifikasi, digunakan untuk memperkirakan nilai dari suatu variabel target yang bersifat kontinu berdasarkan satu atau lebih variabel input (independen). Perbedaannya terletak pada jenis yang dihasilkan.

Data Mining memiliki tujuan sebagai berikut (Kurniawan, 2024):

a. Explanatory

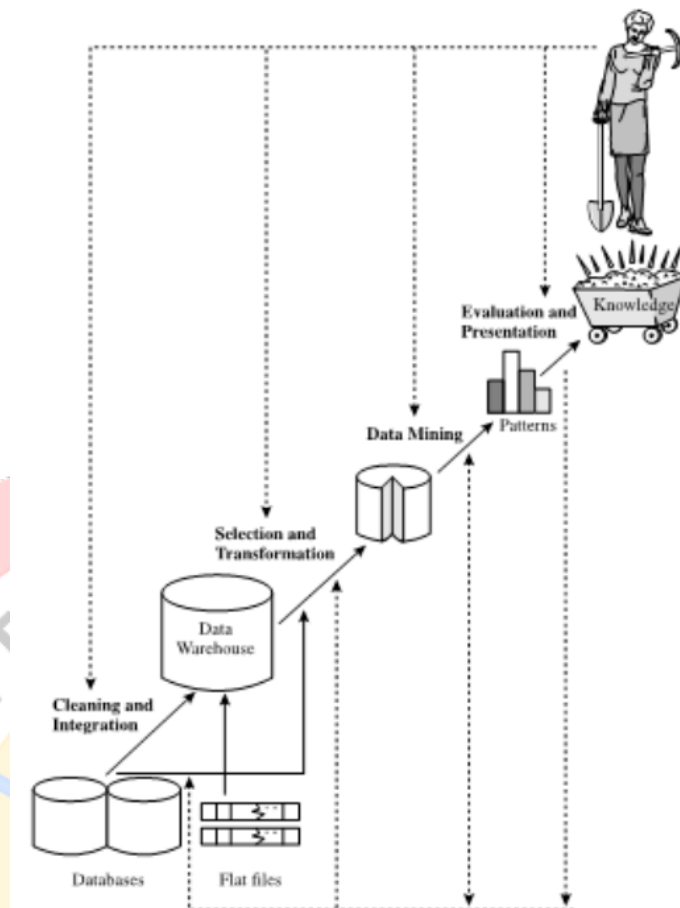
Untuk menjelaskan beberapa kondisi dalam penelitian secara mendalam terhadap suatu fenomena atau kondisi yang sedang dikaji. Berfokus pada menguji hubungan antar variabel dalam studi, mengapa dan bagaimana.

b. Confirmatory

Untuk mempertegas hipotesis atau teori yang telah ada sebelumnya. Peneliti sudah memiliki asumsi mengenai hubungan antar variable dan untuk mengkonfirmasi apa yang terjadi di dalam topik yang sedang diteliti.

c. Exploratory

Proses analisis data dengan tujuan sebuah pola, hubungan, atau struktur yang baru yang belum diketahui sebelumnya atau belum memiliki dasar teori yang kuat. Dengan tujuan untuk mendapatkan wawasan awal, mengenali masalah yang terjadi, atau mengembangkan teori berdasarkan temuan dari data yang dianalisis.



Gambar 2. 1 Tahap-tahap dalam *data mining*

(Sumber: Orisa & Ardita, 2021)

Berikut adalah tahapan dalam *data mining*:

a. Pembersihan Data (*Data Cleaning*)

Proses dimana mengeliminasi data yang tidak relevan dengan penelitian. Sebelum dihilangkan, dilakukan proses membandingkan hipotesa yang telah ditetapkan sebelumnya. Sehingga menjadi lebih mudah untuk menemukan hasil yang diinginkan.

b. Integrasi Data (*Data Integration*)

Sejumlah data yang terdapat dalam berbagai *database* yang dilakukan penggabungan ke dalam satu *database* baru. Volume data yang dibutuhkan cukup besar yang berasal dari berbagai sumber.

c. Seleksi Data (*Data Selection*)

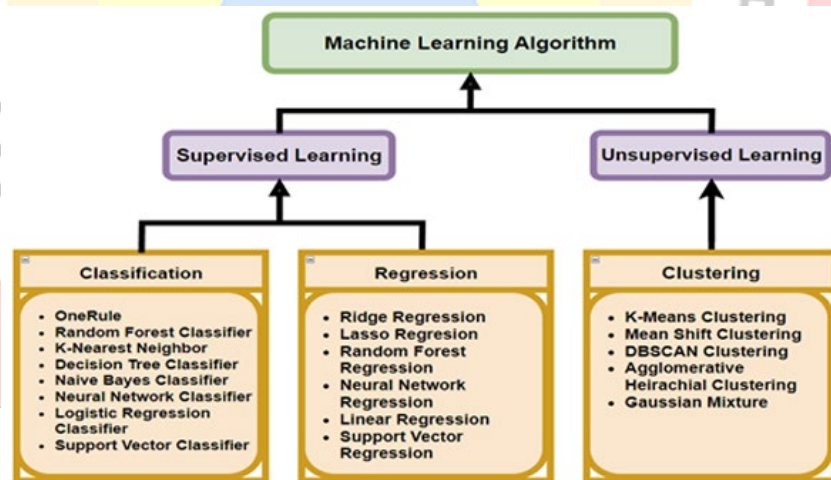
Tahap dimana peneliti memilih data yang relevan serta dibutuhkan untuk penelitian dan ada pada *database* yang sesuai untuk dianalisis, karena tidak seluruh data yang tersimpan di dalam *database* digunakan oleh peneliti.

d. Transformasi Data (*Data Transformation*)

Tahapan yang cukup penting, dimana data diubah, dikonversi, atau digabungkan ke dalam format khusus dan bisa untuk diolah dalam *data mining*. Karena memerlukan format yang khusus, terstruktur, atau terstandarisasi dalam pemrosesannya.

e. Proses *Mining*

Proses utama dalam *data mining*, dimana metode diterapkan untuk menemukan pengetahuan yang tersembunyi dari data. Pengelompokkan dapat dilihat pada gambar 2.2



Gambar 2. 2 Proses Mining

(Sumber: McMurray & Sodhro, 2023)

f. Evaluasi Pola (*Pattern Evaluation*)

Memperoleh kesimpulan dari pengimplementasian pada sistem. Dimana peneliti telah mengidentifikasi pola dalam *knowledge-based* yang telah ditemukan dan diuji pada hipotesa yang sudah dibuat sebelumnya dan dapat dilakukan proses selanjutnya

g. Presentasi Pengetahuan (*Knowledge Presentation*)

Memvisualisasikan dan menyajikan pengetahuan yang didapat mengenai metode dengan tujuan untuk memperoleh pengetahuan yang bersifat informatif.

2.2.2 *Machine Learning*

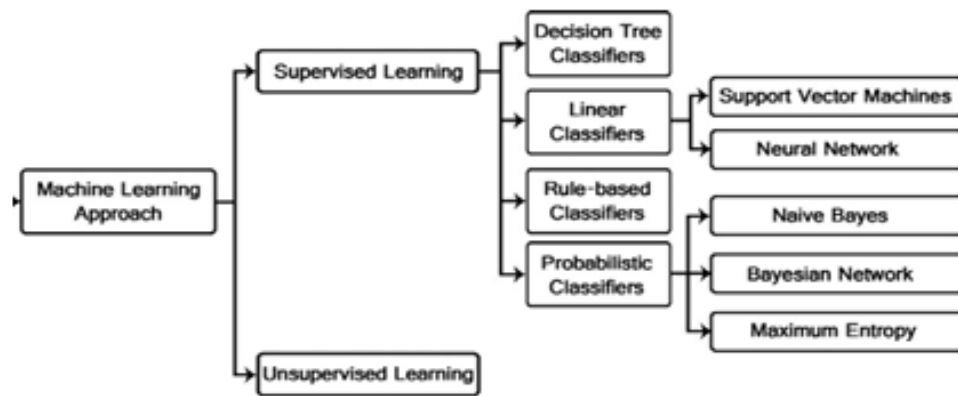
Machine Learning didefinisikan sebagai algoritma matematika yang cara pembelajarannya berasal dari data dan menghasilkan sebuah prediksi di masa yang akan datang dengan melalui 2 (dua) tahap yaitu Latihan (*training*) dan pengujian (*testing*). *Machine Learning* terbagi menjadi 2 (dua) kategori, yaitu:

1. *Supervised Learning*

Metode dalam *machine learning* yang mengandalkan data yang sepenuhnya sudah diberi label. Setiap sampel data yang digunakan dalam pelatihan memiliki pasangan antara *input* dan label atau target *output* yang diinginkan. Dengan tujuan mampu mempelajari hubungan atau pola-pola antara *input* dan *output* berdasarkan data pelatihan..

2. *Unsupervised Learning*

Metode dalam *machine learning* yang dimana datanya tidak memiliki label, dan tidak memiliki informasi terkait hasil dari data yang digunakan. Tujuan utamanya adalah menemukan struktur atau pola yang tersembunyi dalam data..



Gambar 2. 3 Skema *machine learning*

(Sumber: Chowdury et al., 2021)

Machine Learning adalah salah satu cabang dari *Artificial Intelligent* yang dimana memungkinkan computer untuk mempelajari dari sebuah pengalaman serta mengidentifikasi pola dan relasi antara input dan output dengan menggunakan berbagai metode statistic dan probabilitas (Chowdury et al., 2021)

2.2.3 Klasifikasi

Klasifikasi adalah langkah untuk membangun suatu model yang mampu membedakan antar kelas data, guna untuk mengidentifikasi atau memprediksi suatu kelas dari data baru yang labelnya belum diketahui. Beberapa algoritma yang secara umum digunakan oleh para peneliti adalah *Neural Network*, *Decision Tree*, *k-Nearest Neighbor*, dan *Naïve Bayes* (Mustafa et al., 2018).

Dalam klasifikasi variabel, tujuannya bersifat kategorikal dimana mengelompokkan data ke dalam kelas-kelas tertentu. Salah satu contohnya adalah, pengklasifikasian pendapatan dalam tiga kelas, yaitu pendapatan tinggi, sedang, dan rendah. Secara umum klasifikasi data memiliki 2 langkah utama, yang pertama adalah *learning* (fase pelatihan), proses untuk menganalisis data latihan lalu ditampilkan ke

dalam bentuk *rule* klasifikasi, yang kedua adalah *classification* (klasifikasi), yakni penerapan model terhadap data baru bertujuan untuk memperkirakan tingkat akurasi dan *rule* klasifikasi yang telah dibentuk.

Dapat disimpulkan bahwa klasifikasi ialah metode yang cukup penting dalam *data mining* dengan tujuan untuk memprediksi suatu objek menjadi lebih jelas dan untuk melakukannya menggunakan 2 langkah, langkah pertama menggunakan data latihan (*data training*) dan langkah kedua menggunakan data tes (*data testing*) untuk mendapatkan keakuratan dari prediksi data tersebut.

2.2.4 Kredit

Kredit adalah bentuk transaksi keuangan yang melibatkan kreditur sebagai pihak pertama yang memberikan sejumlah uang, barang, ataupun jasa kepada debitur sebagai pihak lain, dengan jangka waktu yang telah disepakati. Kredit juga merupakan fasilitas berupa dana atau tagihan yang dapat diberikan oleh bank atau lembaga keuangan lainnya berdasarkan perjanjian pinjam-meminjam antara bank dengan pihak lain yang mewajibkan untuk melunasi pinjaman tersebut dalam jangka waktu tertentu beserta bunga yang sudah disepakati sebelumnya ('UNDANG-UNDANG REPUBLIK INDONESIA NOMOR 7 TAHUN 1992 TENTANG PERBANKAN SEBAGAIMANA TELAH DIUBAH DENGAN UNDANG-UNDANG NOMOR 10 TAHUN 1998', Pasal 1 Ayat 11) (Abdullah & Wahjusaputri, 2018).

Beberapa unsur yang cukup penting dalam dunia perkreditan adalah kepercayaan, waktu, prestasi, dan risiko yang saling berkaitan satu sama lain. Kredit tidak dapat dilakukan jika tidak ada perjanjian periode pelunasannya (Suyatno, 2016).

Berdasarkan pemaparan sebelumnya, penulis dapat menyimpulkan bahwa kredit merupakan peminjaman uang, barang ataupun jasa yang disediakan oleh pihak bank dengan beberapa ketentuan dan peminjam diberikan waktu untuk melunasi hutang tersebut. Kredit dapat dilakukan jika terdapat perjanjian hitam diatas putih antar pihak bank dan peminjam.

2.3 Teori Rancangan

2.3.1 Metode *Naïve Bayes*

Metode *Naïve Bayes* yakni menjadi salah satu algoritma yang terdapat dalam teknik klasifikasi dan digunakan dengan metode probabilistik dan statistik berdasarkan teori yang dikemukakan oleh Thomas Bayes, ilmuwan asal Inggris. Metode ini digunakan untuk melakukan prediksi terhadap probabilitas suatu peristiwa di masa depan berdasarkan informasi atau pengalaman yang terjadi sebelumnya sehingga dikenal sebagai Teorema Bayes. Teorema ini dikombinasikan dengan *Naïve* yang dapat diasumsikan kondisi antar atributnya saling bebas, meskipun hal tersebut belum tentu terjadi pada kenyataannya (Fadlisyah & Eliyanda, 2021).

Metode *Naïve Bayes* merupakan klasifikasi sederhana dengan mengasumsikan bahwa dengan metode ini dapat mencari Nilai Probabilitas dari setiap atribut masing-masing kelas (Hayuningtyas, 2019).

Naïve Bayes merupakan sebuah metode pengklasifikasian yang digunakan untuk memprediksi peluang di masa depan berdasarkan pengalaman. Ciri utama dari metode ini adalah asumsi yang sangat kuat akan independensi dari masing-masing kondisi. Berikut adalah kegunaan *Naïve Bayes* dalam bidang klasifikasi dan analisis data (Widianto, 2019):

- Digunakan untuk mengklasifikasikan berbagai jenis dokumen seperti berita atau teks akademis
 - Digunakan sebagai salah satu metode dalam *machine learning* yang memanfaatkan probabilitas dalam prosesnya
 - Digunakan untuk mendiagnosa medis berdasarkan gejala pada sebuah sistem secara otomatis
 - Mendeteksi atau memfilter spam, biasanya digunakan pada pengelolaan *e-mail*
- Naïve Bayes* merupakan salah satu metode klasifikasi berbasis probabilitik yang bersifat sederhana dimana digunakan untuk menghitung peluang dari setiap kelas yang mengacu pada penjumlahan frekuensi dan kombinasi nilai dari dataset yang ada. Metode ini dianggap berpotensi baik dalam mengklasifikasikan dokumen, baik dari segi tingkat akurasi maupun efisiensi jika dibandingkan dengan metode pengklasifikasian yang lain. Persamaan dari teorema bayes sebagai berikut (Saputra et al., 2018):

$$P(H|X) = \frac{P(X|H) \cdot P(H)}{P(X)} \quad (1)$$

Keterangan:

Tabel 2. 1 Keterangan Parameter *Naïve Bayes*

Parameter	Keterangan
X	Data dengan class yang belum diketahui
H	Hipotesis data merupakan suatu class spesifik
P(H X)	Probabilitas Hipotesis H berdasarkan kondisi X (posterior probabilitas)
P(H)	Probabilitas hipotesis H (prior probabilitas)

$P(X H)$	Probabilitas X berdasarkan kondisi pada hipotesis X
$P(X)$	Probabilitas X

2.3.2 Metode *Random Forest*

Random Forest adalah algoritma yang terdapat dalam *machine learning* dengan tujuan untuk menyelesaikan sebuah permasalahan yang berhubungan dengan klasifikasi, regresi, dan estimasi dengan tingkat akurasi yang tinggi. Berikut penyebab disebutnya sebagai algoritma random, yaitu (Aziz, 2021):

1. Setiap pohon yang ada dibentuk menggunakan sampel yang berbeda, dimana sampel tersebut diambil dari data latih melalui Teknik *bootstrap sampling*, dilatih menggunakan sampel yang berbeda yang diambil secara acak dari dataset pelatihan.
2. Pemilihan Acak Subset Fitur dalam Setiap Node (Random Feature Selection), mengacak variable yang digunakan untuk melakukan split pada setiap node *decision tree*. Hanya Sebagian kecil dari seluruh variable yang tersedia.

Kombinasi dari beberapa *decision tree*, dimana setiap pohonnya akan bergantung pada nilai random vector yang dijadikan sampel secara bebas dan merata disebut sebagai algoritma *Random Forest*. Hasil prediksi menggunakan *Random Forest* didapatkan berdasarkan hasil terbanyak dari setiap individual *decision tree*. Adapun dirumuskan sebagai berikut:

$$l(y) = \arg \max_c \left(\sum_{n=1}^N l_{h_n}(y) = c \right)$$

(2)

Keterangan:

Tabel 2. 2 Keterangan Parameter *Random Forest*

Parameter	Keterangan
ly	Label hasil prediksi akhir dari <i>Random Forest</i> untuk input y
N	Jumlah total pohon (<i>tree</i>) dalam <i>Random Forest</i>
hny	Hasil Prediksi dari pohon ke-n terhadap input y
c	Kelas yang sedang diperiksa (dari semua kelas yang mungkin)
arg max	Fungsi untuk mengambil argument c yang memberikan nilai maksimum (jumlah voting terbanyak)

2.3.3 *RapidMiner*

RapidMiner merupakan sebuah *software* (perangkat lunak) yang dirancang untuk melakukan proses data dengan menerapkan berbagai prinsip dan algoritma dalam *data mining*. *RapidMiner* mampu mengekstraksi pola-pola dari data set dengan skala cukup besar, kemudian dikombinasikan dengan metode statistika, kecerdasan buatan, dan basis data. Keunggulan *RapidMiner* yaitu memudahkan penggunaanya untuk memproses data dalam skala besar dengan banyak bantuan operator-operatornya yang berfungsi untuk memodifikasi, mengolah, dan menginterpretasi data secara terstruktur (Rahmat et al., 2017).

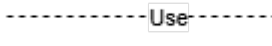
Dikarenakan *RapidMiner* yang bersifat terbuka atau biasa disebut dengan *open source*, maka perangkat lunak ini dijadikan sebagai sebuah solusi untuk kebutuhan analisis data, khususnya di bidang *data mining*, *text mining*, dan analisis prediksi. Menyediakan sekitar 500 operator yang mendukung dalam proses *data mining*, termasuk tahap *input*, tahap *output*, *data preprocessing*, and *visualitation*. Berbagai macam fitur-fitur tersebut membuat pengguna menjadi lebih mudah dalam merancang alur kerja analisis data dan menjadi lebih efisien. YALE (*Yet Another Learning Environment*) adalah nama *software* tersebut sebelum menjadi *RapidMiner*, dikembangkan di *Artificial Intelligence Unit University of Dortmund* pada tahun 2001 oleh RalfKlinkenberg, Ingo Mierswa, dan Simon Fischer. Beberapa fitur dari *RapidMiner* yang menjadikannya sebagai salah satu *software* yang komprehensif dalam *data mining* dan *machine learning*, antara lain:

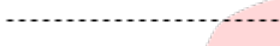
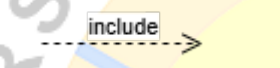
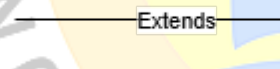
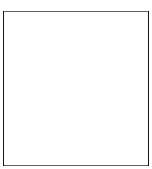
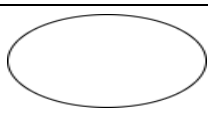
- a. Bentuk grafis yang canggih, sehingga memudahkan pengguna dalam merancang alur kerjanya
- b. Banyak algoritma *data mining*
- c. Banyaknya variasi plugin
- d. Menawarkan prosedur dalam penerapan *data mining* dan *machine learning* yang lebih lengkap untuk melaksanakannya
- e. Proses *data mining* tersusun atas operator-operator yang *nestable* atau dapat disusun secara hierarkis
- f. Dapat diintegrasikan dengan proyek *data mining* seperti weka dan bahasa pemrograman statistika R, yang dimana sangat memungkinkan analisis data secara lebih fleksibel

2.3.4 Use Case Diagram

Use Case Diagram adalah salah satu jenis diagram dalam *Unified Modeling Language* (UML) yang digunakan untuk memberikan gambaran fungsionalitas sebuah sistem dalam sudut pandang pengguna eksternal. *Use Case Diagram* berfungsi sebagai alat komunikasi antara pengembang sistem dan *skateholder* untuk memastikan bahwa kebutuhan sistem telah dipahami dengan benar (Nair et al., 2025). Dimana setiap *Use Case* merepresentasikan serangkaian Langkah atau aktivitas yang dilakukan untuk menrespon aksi yang dilakukan oleh aktor. *Use Case Diagram* sangat berguna dalam tahap analisis kebutuhan karena menyajikan gambaran sederhana namun komprehensif mengenai layanan yang disediakan. Selain itu, *Use Case Diagram* juga membantu proses pengujian sistem karena setiap *Use Case* dapat dikembangkan lebih lanjut menjadi skenario pengujian.

Tabel 2. 2 Simbol Use Case Diagram

NO	GAMBAR	NAMA	KETERANGAN
1		Actor	Menspesifikasikan himpunan peran yang pengguna mainkan ketika berinteraksi dengan <i>use case</i> .
2		Dependency	Hubungan dimana perubahan yang terjadi pada suatu elemen mandiri (<i>independent</i>) akan mempengaruhi elemen yang

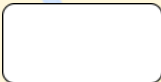
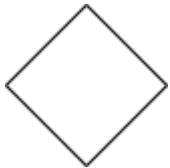
			bergantung padanya dan elemen yang tidak mandiri (<i>dependent</i>).
3		Generalization	Hubungan dimana objek anak (<i>descendent</i>) berbagi perilaku dan struktur data dari objek yang ada di atasnya yaitu objek induk (<i>ancestor</i>).
4		Include	Menspesifikasikan bahwa <i>use case</i> sumber secara <i>eksplisit</i> .
5		Extend	Menspesifikasikan bahwa <i>use case</i> target memperluas perilaku dari <i>use case</i> sumber pada suatu titik yang diberikan.
6		Association	Apa yang menghubungkan antara objek satu dengan objek lainnya.
7		System	Menspesifikasikan paket yang menampilkan sistem secara terbatas.
8		Use Case	Deskripsi dari urutan aksi-aksi yang ditampilkan

			sistem yang menghasilkan suatu hasil yang terukur bagi suatu aktor.
9		Collaboration	Interaksi antara aktor dan elemen lain yang bekerja sama untuk menyediakan perilaku yang lebih besar dari jumlah dan elemen-elemennya (<i>sinergi</i>).
10		Note	Elemen fisik yang eksis saat aplikasi dijalankan dan mencerminkan suatu sumber daya komputasi.

2.3.5 *Activity Diagram*

Activity Diagram Adalah salah satu jenis diagram dalam *Unified Modeling Language* (UML) yang digunakan untuk memvisualisasikan alur kerja (*workflow*) atau urutan aktivitas suatu sistem (Al-Fedaghi, 2021). Setiap aktivitas dalam sistem direpresentasikan sebagai sebuah aksi yang dapat dijalankan oleh sistem atau pengguna. Komponen utama yang sering digunakan ada *Activity*, *action*, *initial node*, *activity final node*, *fork node* dan *Decision Node*. Diagram ini berfungsi untuk mendeskripsikan proses bisnis atau alur logika dari suatu fungsi dalam perangkat lunak secara visual dan terstruktur. *Activity Diagram* sangat bermanfaat dalam perancangan sistem yang melibatkan banyak kondisi dan alur yang kompleks.

Tabel 2. 3 Simbol *Activity Diagram*

NO	GAMBAR	NAMA	KETERANGAN
1		Activity	Memperlihatkan bagaimana masing-masing kelas antarmuka saling berinteraksi satu sama lain.
2		Action	<i>State</i> dari sistem yang mencerminkan eksekusi dari suatu aksi.
3		Initial Node	Bagian awal <i>object</i> dibentuk atau diawali.
4		Activity Final Node	Bagian <i>object</i> dibentuk dan dihancurkan.
5		Fork Node	Satu aliran yang pada tahap <i>torrent</i> berubah menjadi beberapa aliran.
6		Decision Node	Menunjukkan aktivitas yang harus dipilih.

2.3.6 Website

Website merupakan sekumpulan halaman yang merupakan dokumen berisi teks yang terdapat di dalam *file* berformat HTML (*Hyper Text Markup Language*) dimana ditampilkan didalam sebuah domain dan disimpan dalam *server hosting*. Sehingga halaman tersebut dapat diakses oleh siapa saja menggunakan internet atau yang biasa disebut dengan URL (*Uniform Resource Locator*). *Website* saat ini sangat penting dimana semua Masyarakat dapat mengetahui sebuah informasi secara cepat, karena dapat diakses kapan dan dimanapun. Terbagi menjadi 2 (dua), yaitu statis dan dinamis. Statis memerlukan modifikasi secara langsung pada *code* secara berkala untuk sebuah konten. Sedangkan dinamis tidak perlu melakukan perubahan *code* untuk memperbaharui konten, karena konten tersebut telah disimpan ke dalam sebuah *database* (Widia & Asriningtias, 2021).

2.3.7 PHP (*Hypertext Preprocessor*)

PHP (*Hypertext Preprocessor*) digunakan untuk membuat sebuah halaman, *website*, aplikasi web, dsb. PHP digunakan karena bersifat gratis atau *open source*. Banyak web dan database yang mendukung Bahasa pemrograman ini, sehingga sangat mudah untuk diimplementasikan dan pemrograman PHP memiliki ekstensi “.php” dan dapat berisi teks, HTML, CSS, Javascript, dan tentunya kode PHP. Syntax PHP dimulai dengan <?php dan diakhiri dengan ?> (Rusli et al., 2019). Berikut adalah salah satu contoh penggunaan dasar PHP:

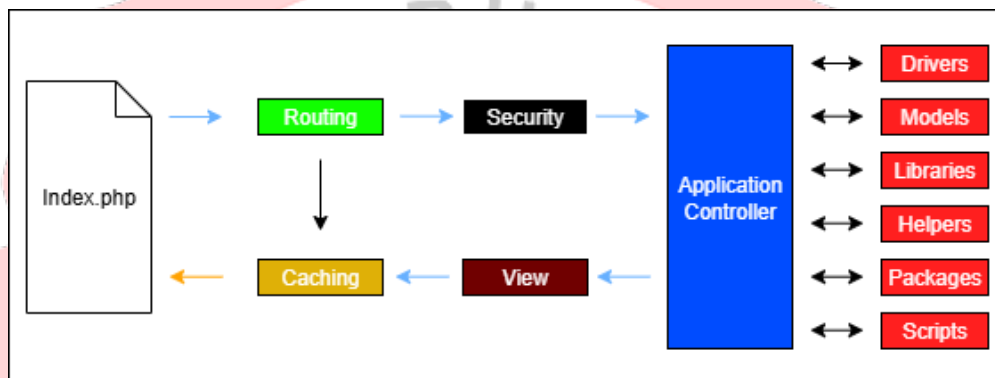
```
<?php
```

```
Echo “Hello New World!”; //Menampilkan tulisan Hello New World
```

```
?>
```

2.3.8 Framework Codeigniter

Codeigniter adalah *framework* PHP yang cukup populer dan banyak digunakan dalam pengembangan web, dimana memiliki fitur yang cukup lengkap dan dokumentasi yang dimiliki baik. Sehingga membuat para pengembang atau *developer* menjadi lebih efisien dalam membuat aplikasi (Herdiansah et al., 2024). *Codeigniter* menjadi *framework* PHP yang populer karena ringan dan kecepatan eksekusinya lebih tinggi disbanding dengan *framework* yang lain (Rawal, 2023).



Gambar 2. 4 Alur Kerja Code Igniter

(Sumber: Murtiningsih & Fatmasari, 2023)

Pada Gambar 2.4 merupakan alur dari cara kerja *codeigniter* (Murtiningsih & Fatmasari, 2023), yaitu:

1. *Index.php*

file pertama yang digunakan untuk mengidentifikasi kebutuhan untuk menjalankan *framework* ini. Berfungsi untuk memulai aplikasi dan mengarahkan request awal ke sistem inti *framework*. Terdapat penentuan *environment* (*development*, *testing*, atau *production*), pengaturan path sistem, serta pemanggilan *class loader* dan *core framework*. Sebagai titik utama untuk menyiapkan semua komponen yang digunakan agar aplikasi dapat berjalan.

2. *Router*

Berfungsi untuk menentukan tugas apa yang harus dikerjakan saat ada permintaan dari *client/browser* ke bagian yang sesuai dalam sistem aplikasi. *Router* akan menerjemahkan URL untuk menentukan metode apa yang harus dijalankan.

3. *Caching*

Digunakan untuk memvalidasi data apakah sudah pernah diminta atau belum. Mekanisme penyimpanan sementara data hasil pemrosesan, yang bertujuan untuk meningkatkan performa aplikasi. Jika data sudah pernah dihasilkan sebelumnya, maka data diambil dari cache bukan diproses ulang.

4. *Security*

Berperan untuk memfilter data yang masuk agar dapat mencegah adanya ancaman. Komponen ini sangat penting dalam *framework* karena berfungsi untuk melindungi aplikasi dari berbagai ancaman atau serangan, seperti SQL Injection.

5. *Controller*

Berguna untuk mengatur jalannya seluruh aplikasi dengan memproses sesuai permintaan dan menjadi penghubung, yaitu antara *models*, *helpers*, *plugins*, dan *script*. Ketika ada *request* dari *client*, *controller* akan menangkapnya dan akan memproses yang diperlukan.

6. *View*

Digunakan untuk menampilkan tampilan dan seluruh informasi berdasarkan permintaan *browser*. Berisi antarmuka pengguna (UI), seperti tampilan HTML, tabel data, form input atau grafik. Menggabungkan dengan elemen visual agar mudah dipahami.

2.3.9 *Blackbox Testing*

Black Box Testing merupakan jenis pengujian program yang berfokus pada kesesuaian antara masukan dan keluaran yang sudah dirancang sebelumnya, tanpa memeriksa struktur internal atau kode program. Dalam pengujian ini, penguji hanya memberikan input pada aplikasi dan menganalisa apakah output yang dihasilkan sesuai atau tidak. Jenis-jenis dalam *Black Box Testing* (Nayyar, 2019), yaitu:

1. *Functional Testing*

Berfokus pada *input* dan *output* saja, sesuai dengan spesifikasi kebutuhan penggunaanya. Dilakukan dengan memberikan berbagai macam input ke sistem, ketika data yang valid dapat disimpan kemudian ditampilkan sesuai dengan yang sudah dirancang dan ketika dimasukkan data yang tidak valid akan menampilkan notifikasi gagal atau pesan kesalahan.

2. *Non-Functional Testing*

Bagian ini berfokus pada aspek kualitas produk, seperti tampilan jika digunakan di berbagai perangkat dan resolusi yang berbeda, respon waktu, keamanan, keandalan, dan kinerja sistem secara keseluruhan. Berkaitan dengan aspek kualitas aplikasi.

3. Regression Testing

Pengujian yang lebih memperhatikan apakah terjadi penurunan fungsi program ketika ada perubahan versi akibat dari pembaruan sistem seperti perbaikan bug atau pembaruan versi.

2.3.10 *MySQL*

MySQL merupakan sistem manajemen basis data relasional (RDBMS – *Relational Database Management System*), berfungsi sebagai *server database* dan dapat digunakan tanpa pungutan biaya apapun, baik untuk pengembangan pribadi maupun aplikasi komersial karena berlisensi *GNU (General Public License)*, *Free Software* (Perangkat Lunak Bebas) dan *Shareware* (Perangkat Lunak berpemilik yang penggunaannya terbatas) adalah 2 (dua) jenis lisensi yang ditawarkan. Untuk mengelola dan memanipulasi data menggunakan bahasa yang paling sering digunakan yaitu *SQL (Structured Query Language)* (Fitri, 2020).

2.4 Tinjauan Studi

Tabel 2. 4 Tinjauan Studi

No.	Peneliti	Nama Jurnal dan Tahun	Judul	Metode	Kesimpulan
1.	H T Sihotang, D R Naidu	Jurnal Sains dan Teknologi, 2019	Klasifikasi Kelayakan Kredit Pada Calon Debitur CS Finance Dengan Metode <i>Naïve Bayes</i>	<i>Naïve Bayes</i>	Penambahan atribut menjadi 16 meningkatkan akurasi klasifikasi kelayakan kredit dengan <i>Naïve Bayes</i> menjadi 59%, sehingga prediksi keterlambatan pembayaran dapat dilakukan lebih baik sejak awal

2.	Diah Puspitasari, Syifa Sintia Al Khautsar, Wida Mustika	Jurnal Informatika Upgris, 2019	Algoritma <i>Naïve Bayes</i> Untuk Memprediksi Kredit Macet Pada Koperasi Simpan Pinjam	<i>Naïve Bayes</i>	Algoritma <i>Naïve Bayes</i> mampu mengklasifikasikan kredit macet dengan akurasi 59%, dan terbukti bahwa atribut pekerjaan, tempat tinggal, pendapatan, permohonan, dan cicilan berpengaruh dalam prediksi kredit macet
3.	Ratih Yulia Hayuningtyas	Jurnal Informatika, 2019	Penerapan Algoritma <i>Naïve Bayes</i> untuk Rekomendasi Pakaian Wanita	<i>Naïve Bayes</i>	<i>Naïve Bayes</i> efektif digunakan untuk klasifikasi dan

				mendukung pengambilan keputusan, seperti dalam rekomendasi pakaian wanita berdasarkan probabilitas posterior
4.	Rizal Rachman, Rissa Nurfitriana Handayani	Jurnal Informatika, 2021	Klasifikasi Algoritma <i>Naïve Bayes</i> Dalam Memprediksi Tingkat Kelancaran Pembayaran Sewa Teras UMKM	Algoritma Naice Bayes berhasil mengklasifikasikan status pembayaran sewa UMKM dengan akurasi 81,81% dan recall 100%, sehingga

					dapat membantu manajemen dalam mengambil keputusan untuk meminimalisir keterlambatan pembayaran
5.	L. Sathish kumar, V. Pandimurugan, D. Usha, M. Nageswara Guptha, M. S. Hema	Materials Proceedings, 2022	Today: Random Forest Classification Algorithm for Predicating Loan	Random Forest, SVM, Decision Tree, Logistic Regression	<i>Random Forest</i> menunjukkan kinerja terbaik dalam prediksi kelayakan pemberian pinjaman dibandingkan SVM, Decision Tree, dan Logistic Regression,

				serta memiliki kemampuan generalisasi yang kuat
6.	Nisha Arora, Pankaj Deep Kaur	Applied Soft Computing Journal, 2020	A Bolasso Based Consistent Feature Selection Enabled <i>Random Forest</i> Classification Algorithm: An Application to Credit Risk Assesment	Naive Bayes, SVM, KNN
				Bolasso terbukti sebagai metode seleksi fitur terbaik untukn meningkatkan akurasi klasifikasi resiko kredit, khususnya saat digunakan dengan algoritma <i>Random Forest</i>

7.	Yuelin Wang, Yihan Zhang, Yan Lu, Xinran Yu	Procedia Computer Science, 2020	A Comparative Assessment of Credit Risk Model Based on Machine Learning – A Case Study of Bank Loan Data	KNN, Decision Tree, <i>Random Forest</i> , <i>Naïve Bayes</i> , Logistic Regression	Model prediksi risiko kredit berbasis machine learning memperluas skala bisnis pinjaman, mengurangi biaya kredit macet, serta mendorong pengembangan sistem informasi kredit dan kesadaran pengguna untuk menjaga reputasi kredit pribadi
----	--	---------------------------------	--	---	---

8.	Tri Wahyuningsih, Danny Manongga, Irwan Sembiring, Sutarto Wijono	Procedia Computer Science, 2024	Comparison of Effectiveness of Logistic Regression, <i>Naïve Bayes</i> , and <i>Random Forest</i> Algorithms in Predicting Student Arguments	Logistic Regression, <i>Naïve Bayes</i> , <i>Random Forest</i>	Logistic Regression terbukti unggul dalam mengklasifikasikan argumen siswa dan daapt digunakan untuk sistem bimbingan otomatis yang lebih efektif, namun penelitian lebih lanjut dengan lagoritma dan variabel tambahan diperlukan untuk hasil yang lebih komprehensif
----	--	------------------------------------	---	---	---

9.	Devi Fitrianah, Saruni Dwiasnati, Hanny Hikmayanti H, Kiki Baihaqi	Ahmad	Faktor Exakta, 2021	Penerapan Metode Machine Learning untuk Prediksi Nasabah Potensial menggunakan Algoritma Klasifikasi <i>Naïve Bayes</i>	<i>Naïve Bayes</i>	Pemilihan algoritma data transformation tergantung pada kebutuhan proses data, dan <i>RapidMiner</i> menyediakan berbagai metode preprocessing yang fleksibel untuk mendukung pengembangan aplikasi data mining
10.	Kendall Lemons		International Journal of Undergraduate	A Comparison Between <i>Naïve Bayes</i> and <i>Random Forest</i> to Predict Breast Cancer	<i>Naïve Bayes</i> , <i>Random Forest</i>	<i>Random Forest</i> menghasilkan akurasi tertinggi

		Research and Creative Activities, 2020			97,82% dalam diagnosis kanker payudara dan mengungguli <i>Naïve Bayes</i> , meskipun lebih kompleks dan rentan overfitting jika tidak dituning dengan tepat
11.	Wei Chen, Yang Li, Weifeng Xue, Himan Shahabi, Shaojun Li, Haoyuan Hong, Xiaojing Wang,	Science of the Total Environment, 2020	Modeling Flood Susceptibility Using Data-driven Approaches of <i>Naïve Bayes Tree</i> , Alternating Decision Tree, and <i>Random Forest</i> Methods	<i>Naïve Bayes</i> , Decision Tree, <i>Random Forest</i>	Model <i>Random Forest</i> memberikan prediksi banjir terbaik, dengan faktor terpenting yang mempengaruhi banjir adalah

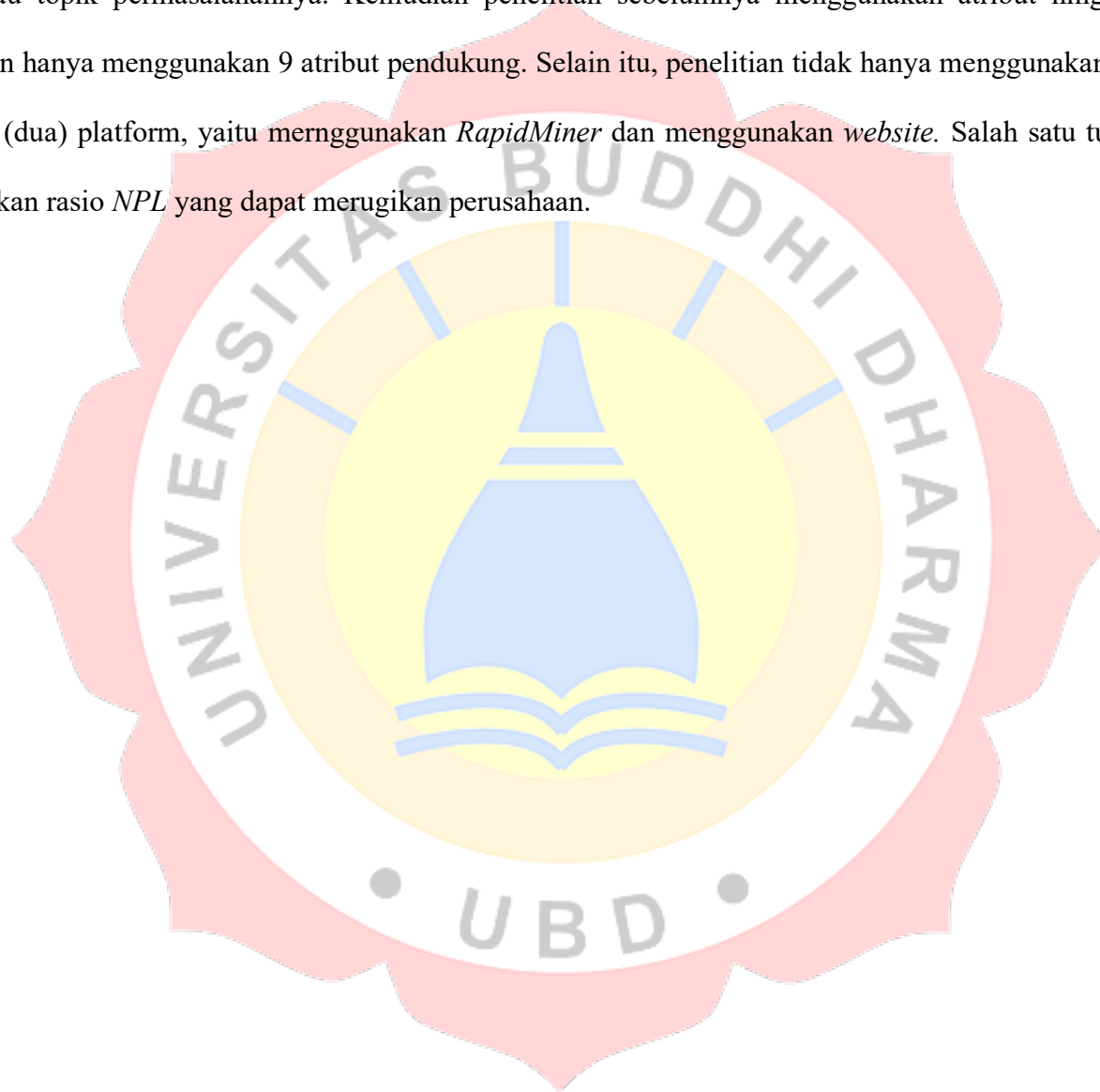
	Huiyuan Bian, Shuai Zhang, Biswajeet Pradhan, Baharin Bin Ahmad				penggunaan lahan NDVI, kemiringan, dan jarak ke sungai
12.	Evita Fitri, Yuri Yuliani, Susy Rosyida, Windu Gata	TRANSFORMTIKA, 2020	Analisis Sentimen Terhadap Aplikasi Ruangguru Menggunakan Algoritma <i>Naïve Bayes, Random Forest, Dan Support Vector Machine</i>	<i>Naïve Bayes,</i> <i>Random Forest,</i> Support Vector Machine	Model <i>Random Forest</i> memberikan performa terbaik dalam analisis sentimen aplikasi Ruangguru dengan akurasi 97,16% dan AUC 0,996, mengungguli SVM dan <i>Naïve Bayes</i>

13.	Atria, Wiwit Pura Nurmayanti, Umam Hidayaturohman, Ristu Haiban Hirzi4	ISST, 2021	Comparative Analysis of Classifier for BRI Customers Pattern Prediction Using <i>Naïve Bayes</i> , KNN, and <i>Random Forest</i>	<i>Naïve Bayes</i> , KNN, <i>Random Forest</i>	Metode <i>Random Forest</i> merupakan yang terbaik untuk mendeteksi penyaluran kredit di BRI Unit Sikur dengan akurasi tertinggi sebesar 96,15%
14.	Nurhidayah, Muhammad Gazali, Siti Arni Wulandya3	ISST, 2021	Analysis of Diagnosis Classification of Dengue Hemorrhagic Fever (DHF) Using <i>Random Forest</i> and Bagging Method With Imbalanced Class Resampling Technique	<i>Random Forest</i>	Metode <i>Over Random Forest</i> merupakan yang paling optimal untuk klasifikasi data diagnosis DBD di RS Kota Mataram

					dengan akurasi tertinggi sebesar 98,19%
15.	Omar Pahlevi, Amrin Amrin, Yopi Handrianto	Jurnal Infortech, 2023	Implementasi Algoritma Klasifikasi <i>Random Forest</i> untuk Penilaian Kelayakan Kredit	<i>Random Forest</i>	Model <i>Random Forest</i> berhasil mengklasifikasikan kelayakan kredit dengan akurasi 78,6% dan AUC 0,907, menunjukkan performa klasifikasi yang sangat baik

Berdasarkan penelitian-penelitian yang sudah dilakukan sebelumnya, sebagian menggunakan lebih dari satu algoritma dalam penelitiannya. Kemudian ada juga yang hanya 1 algoritma yaitu hanya menggunakan algoritma *naive bayes* atau *random forest*. Dari semua penelitian tersebut dicari algoritma yang memiliki potensi paling besar dalam memprediksi, meskipun terdapat beberapa penelitian yang berbeda

permasalahan utama atau topik permasalahannya. Kemudian penelitian sebelumnya menggunakan atribut hingga 16 atribut, sedangkan penelitian yang dilakukan hanya menggunakan 9 atribut pendukung. Selain itu, penelitian tidak hanya menggunakan satu platform, tetapi diuji dengan menggunakan 2 (dua) platform, yaitu menggunakan *RapidMiner* dan menggunakan *website*. Salah satu tujuan penelitian ini adalah untuk mengurangi kenaikan rasio *NPL* yang dapat merugikan perusahaan.



BAB III

METODOLOGI PENELITIAN

3.1 Tinjauan Umum Perusahaan

3.1.1 Sejarah Singkat

PT. BPR Magga Jaya Utama didirikan pada tanggal 18 Januari 2006 dengan berdasarkan Akta Notaris Pendy Tanzil, S.H No. 5. Kemudian pada tanggal 10 Maret 2006 mendapatkan pengesahan secara resmi melalui Surat KMK dan HAM RI No. C-07084 HT.01.01. Pada tanggal 17 April 2006 Bank Indonesia mengeluarkan izin usaha kepada PT. BPR Magga Jaya Utama dengan nomor 8/34/KEP.GBI/2006, Kemudian melakukan *soft Launching* pada tanggal 1 Mei 2006. Tujuan utama PT. BPR Magga Jaya utama didirikan adalah untuk mendukung program pemerintah melakukan pengembangan usaha kecil menengah dengan memberikan bantuan modal kepada yang membutuhkan untuk mengembangkan usahanya.

3.1.2 Visi, Misi, dan Tujuan

Berikut visi, misi dan tujuan dibalik didirikannya PT. BPR Magga Jaya Utama:

1. Visi

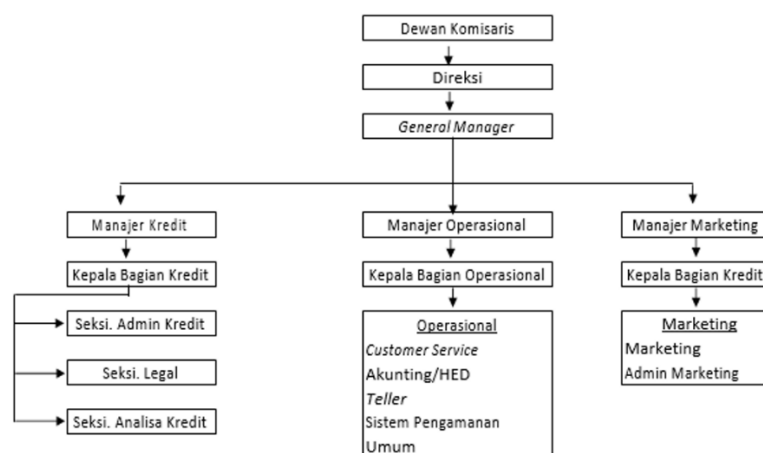
- a. Menjaga reputasi yang baik bagi PT. BPR Magga Jaya Utama, sehingga mendapatkan kepercayaan dari masyarakat.
- b. Memperbaiki kualitas layanan produk yang disediakan agar memberikan kepuasan yang lebih baik kepada pelanggan.
- c. Meningkatkan skala bisnis dan jaringan pemasaran perusahaan.
- d. Memperoleh keuntungan usaha yang adil dan sesuai dengan standar yang berlaku.

2. Misi

- a. Menyediakan layanan perbankan dan jasa keuangan berkualitas kepada nasabah dengan produk dan pelayanan unggulan.
- b. Menciptakan lingkungan kerja yang dinamis guna untuk mendukung pengembangan tenaga profesional berkualitas yang menunjukkan dedikasi dan integritas yang tinggi. Berkomitmen untuk menjunjung pembangunan nasional.
- c. Menjamin tingkat pengembalian keuntungan yang adil bagi para pemegang saham.

3.1.3 Struktur Organisasi

Setiap perusahaan butuh susunan yang jelas agar semua bagian bisa bekerja selaras, Struktur Organisasi menunjukkan alur tanggung jawab dan hubungan antar posisi. Berikut ini adalah bentuk Struktur yang digunakan dalam PT. BPR Magga Jaya Utama:



Gambar 3. 1 Struktur Organisasi

(Sumber: HRD PT. BPR Magga Jaya Utama, 2025)

3.1.4 Tugas dan Wewenang

Berdasarkan pada gambar 3.1 berikut adalah penjelasan secara umum mengenai tugas dan wewenang dari setiap jabatan:

a. Dewan Komisaris

Mengawasi kegiatan operasional yang dijalankan secara menyeluruh oleh Direksi, sebagai pendamping Direksi ketika menjalin hubungan dengan pihak lain, serta mengevaluasi kinerja direksi dan memastikan bahwa anggaran dan program kerja sesuai dengan peraturan perundang-undangan yang berlaku.

b. Direksi

Direktur utama bertanggung jawab sebagai koordinator atau ketua komite dan mengawasi pendanaan serta pemberian pinjaman. Direktur bertugas sebagai anggota kredit, mengkoordinasikan kegiatan operasional bank, memantau likuiditas bank. Direksi atas penetapan kebijakan perusahaan, perencanaan operasional, dan strategi dalam mencapai target perusahaan.

c. General Manager

Tugas General Manager mencakup beberapa aspek, termasuk kepemimpinan perusahaan dan memotivasi karyawan, mengelola operasional harian perusahaan, memastikan departemen-departemen melaksanakan strategi perusahaan secara efektif. Selain itu perencanaan, pelaksanaan, koordinator, pengawas, dan evaluasi seluruh aktivitas bisnis perusahaan, mengelola perusahaan sesuai dengan visi dan misi menjadi bagian sebagai GM.

d. Manager Kredit

Tugas Manager Kredit meliputi monitoring kredit, pengawasan kredit, serta pelaksanaan supervisi program pembiayaan dan pemasaran. Mereka bertanggung jawab dalam memantau dan mengevaluasi kinerja kredit, mengawasi proses pengajuan kredit.

e. Manager Operasional

Tugas Manager Operasional mencakup mengoordinasikan, mengamankan, dan mengarahkan kegiatan terkait sistem dan prosedur operasional, memberikan dukungan penuh dalam pelaksanaan audit internal dan eksternal.

f. Manager Marketing

Tugas Manager Marketing meliputi memastikan kebutuhan dan keinginan pelanggan terpenuhi, bertanggung jawab dalam meningkatkan keuntungan perusahaan, menganalisis peluang masa depan.

g. Marketing

Tugas Marketing meliputi: memimpin dan mengawasi kegiatan pemasaran, pelayanan nasabah, dan pencarian nasabah baru; memelihara disiplin kantor dan motivasi kerja pegawai.

h. Kepala Bagian Kredit

Kepala Bagian Kredit memiliki tanggung jawab untuk mengoordinasikan kegiatan dan proses kredit, mengawasi serta membina debitur bank, memberikan

usulan penanganan kredit bermasalah kepada Direksi, mengoordinasikan dan mengevaluasi tindakan untuk menjaga kolektibilitas kredit.

i. Administrasi Kredit

Meneliti kebenaran atau keakuratan dan kelengkapan sebuah dokumen yang diajukan oleh debitur melalui Account Officer, menyusun jadwal angsuran secara sistematis, dan mencatat semua kredit yang diberikan secara akurat.

j. Legal

Bagian Legal mempunyai tanggung jawab yang mencakup pemeriksaan kelengkapan dan validitas data debitur beserta pemilik jaminan, memverifikasi validitas dokumen jaminan seperti sertifikat kepemilikan, BPKB, dan dokumen lainnya untuk akad dan pencairan kredit.

k. Kepala Bagian Operasional.

Bertugas untuk mengkoordinasi serta mengawasi pelaksanaan pekerjaan seluruh karyawan di bagian operasional. Selain itu juga bertanggung jawab atas pemeliharaan kas, pemeliharaan rekening simpanan di bank lain, serta mengawasi surat-surat berharga yang dimiliki oleh bank.

l. *Customer Service*

Customer Service bertanggung jawab atas pemberian informasi tentang produk dan jasa perbankan kepada calon nasabah, bagaimana prosedur penyimpanan dan keuntungan yang didapat dari tabungan serta deposito, dan terkait produk kredit.

m. Teller

Tugas seorang Teller berkaitan dengan seluruh transaksi keuangan tunai dan non-tunai, memberikan pelayanan yang akurat, dan ramah kepada nasabah, serta memberikan informasi mengenai produk bank dan saldo rekening yang diminta kepada nasabah.

n. Akunting

Tugas departemen akunting meliputi mencatat seluruh transaksi sesuai dengan data atau bukti, mencatat penerimaan bunga dan biaya administrasi dari bank lain, melakukan *penginputan* data dan menyimpan mutasi pada sistem komputer secara akurat dan tepat waktu.

o. Personalia dan Umum

Bertanggung jawab untuk merekrut karyawan berdasarkan dengan kebutuhan perusahaan, menyusun surat-surat terkait karyawan, mengelola pengarsipan dokumen milik karyawan, serta menjaga kerahasiaannya.

p. Satuan Keamanan

Satuan Pengamanan (Satpam) mempunyai tugas utama dalam memberikan perlindungan terhadap karyawan dan properti milik perusahaan, serta menegakkan aturan yang berlaku di perusahaan.

q. *Office Boy*

Tugas tersebut melibatkan pelayanan minuman, kebersihan kantor, dan aspek logistik dalam mendukung kegiatan operasional perusahaan.

3.2 Analisa Kebutuhan

Penerapan metode dalam penelitian ini melalui beberapa tahapan, yaitu:

3.2.1 RE (*Requirement Elicitation*)

Untuk mengetahui apa saja yang diperlukan dalam perancangan sistem yang akan digunakan, dilakukan wawancara secara langsung kepada pengguna dan berikut adalah hasil dari wawancara:

Tabel 3. 1 RE (*Requirement Elicitation*)

No	Analisis Kebutuhan Sistem	Keterangan
	saya ingin sistem dapat :	
1	Tampilan mudah dipahami dan menarik.	
2	Menggunakan data seperti : Umur, Jenis Pekerjaan, Pendapatan Per Bulan , Riwayat pinjaman, Nominal Pengajuan Kredit, angsuran per bulan dan durasi pinjaman.	
3	Mampu mengolah data	
4	Aplikasi berbasis <i>Website</i>	
5	Menampilkan <i>Dashboard</i> Nasabah	
6	Memberikan tampilan keterangan layak atau tidak nasabah tersebut mendapatkan kredit.	
7	Dapat mengetahui jumlah data yang sudah dimasukkan	
8	Dapat Mengakses Riwayat data nasabah	

3.2.2 *Data Selection*

Dalam tahap ini, tentunya membutuhkan *dataset* untuk diolah. Data yang digunakan adalah data nasabah kredit pada PT. BPR Magga Jaya Utama dengan rentang waktu dari tahun 2019 hingga 2023 dengan jumlah data sebanyak 6983 nasabah.

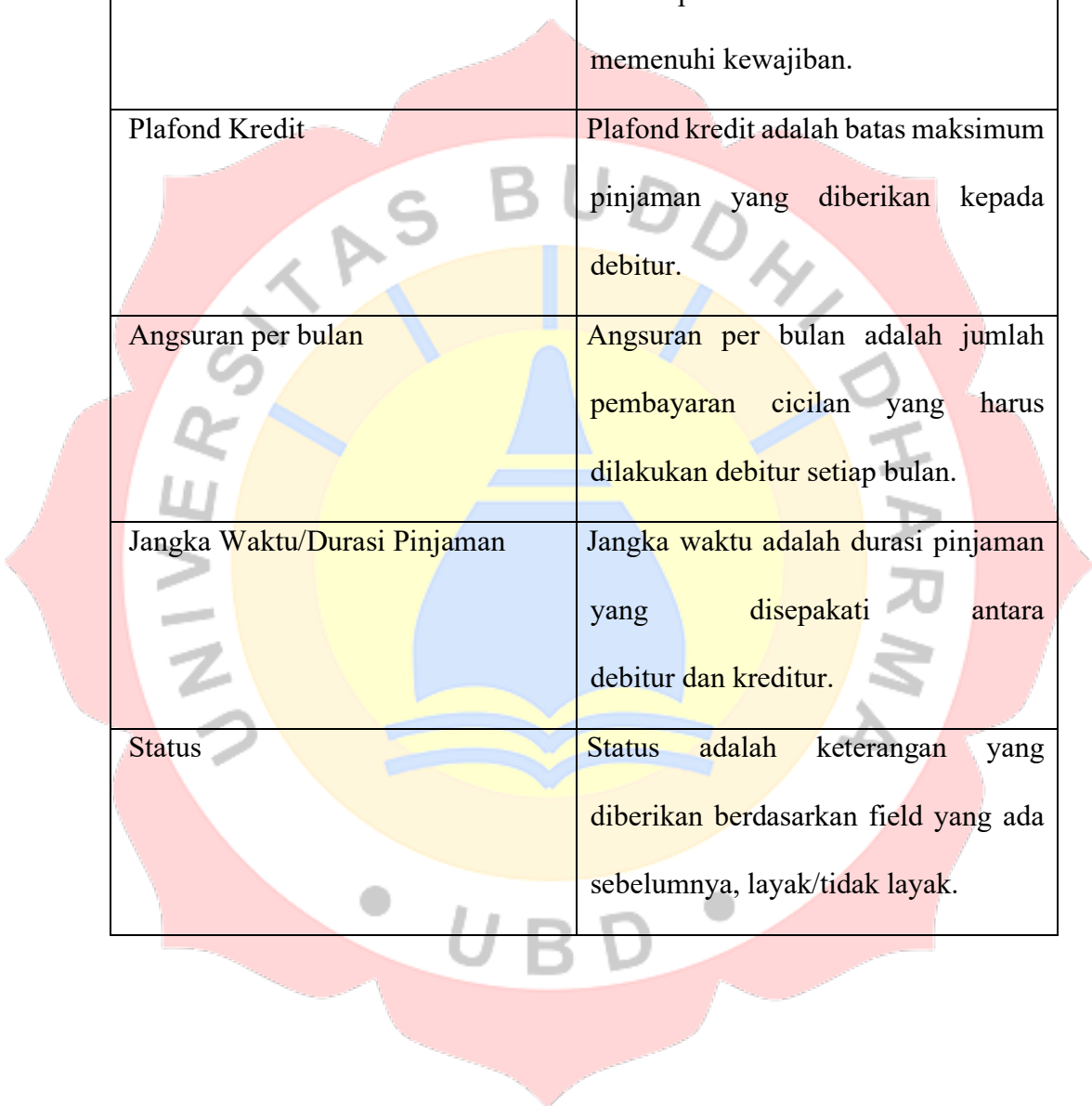
Kolektibilitas	umur	Profesi	pendapatan bulanan	plafond	angsuran bulanan	Jangka Waktu	Status
5	18		NULL	Rp 2.500.000.000,00	Rp 14.583.333,00	12	Tidak Layak
4	45	KARYAWAN SWASTA	Rp 108.067.747	Rp 500.000.000,00	Rp 7.500.000,00	12	Tidak Layak
1	46	WIRASWASTA	Rp 3.000.000	Rp 3.500.000,00	Rp 361.667,00	12	Layak
5	53	WIRASWASTA	Rp 1.148.653.509	Rp 700.000.000,00	Rp 9.100.000,00	12	Tidak Layak
5	18		NULL	Rp 5.000.000.000,00	Rp 29.166.667,00	12	Tidak Layak
5	44	WIRASWASTA	Rp 40.000.000	Rp 100.000.000,00	Rp 2.000.000,00	3	Tidak Layak
5	51	WIRASWASTA	Rp 100.000.000	Rp 250.000.000,00	Rp 5.000.000,00	2	Tidak Layak
5	55	KARYAWAN SWASTA	Rp 4.279.121	Rp 4.500.000,00	Rp 465.000,00	12	Tidak Layak
5	41	WIRASWASTA	Rp 12.158.875	Rp 20.000.000,00	Rp 400.000,00	4	Tidak Layak
5	41	WIRASWASTA	Rp 12.158.875	Rp 20.000.000,00	Rp 400.000,00	4	Tidak Layak
3	42	KARYAWAN SWASTA	Rp 4.471.000	Rp 10.000.000,00	Rp 983.333,00	12	Tidak Layak
3	44	KARYAWAN SWASTA	Rp 3.376.000	Rp 10.000.000,00	Rp 983.333,00	12	Tidak Layak
5	39	KARYAWAN SWASTA	Rp 3.680.000	Rp 5.000.000,00	Rp 908.333,00	6	Tidak Layak
5	37	KARYAWAN SWASTA	Rp 3.680.000	Rp 10.000.000,00	Rp 983.333,00	12	Tidak Layak
3	38	KARYAWAN SWASTA	Rp 3.376.000	Rp 8.000.000,00	Rp 786.667,00	12	Tidak Layak
3	33	KARYAWAN SWASTA	Rp 4.000.000	Rp 7.000.000,00	Rp 709.333,00	12	Tidak Layak
5	18		NULL	Rp 5.000.000.000,00	Rp 29.166.667,00	12	Tidak Layak
5	18		NULL	Rp 5.000.000.000,00	Rp 29.166.667,00	12	Tidak Layak
5	18		NULL	Rp 7.000.000.000,00	Rp 40.833.333,00	12	Tidak Layak
3	45	KARYAWAN SWASTA	Rp 5.955.000	Rp 10.000.000,00	Rp 983.333,00	12	Tidak Layak
5	33	KARYAWAN SWASTA	Rp 2.000.000	Rp 15.000.000,00	Rp 325.000,00	60	Tidak Layak
5	57	WIRASWASTA	Rp 7.500.000	Rp 3.000.000,00	Rp 590.000,00	6	Tidak Layak
5	46	WIRASWASTA	Rp 12.000.000	Rp 3.000.000,00	Rp 590.000,00	6	Tidak Layak

Gambar 3. 2 Data Nasabah Kredit PT. BPR Magga Jaya Utama

Kemudian penulis melakukan pemilihan atribut atau *field* yang akan digunakan dalam penelitian ini. Berikut adalah yang sudah dipilih untuk digunakan:

Tabel 3. 2 Keterangan Atribut untuk data yang digunakan

Atribut	Keterangan
Kolektibilitas/Riwayat Pinjaman	Kolektibilitas adalah status pembayaran kewajiban debitur kepada kreditur.
Jenis Kelamin	Jenis Kelamin calon debitur juga digunakan sebagai pertimbangan pengeluaran tiap bulannya
Umur	Umur debitur dapat menjadi faktor penting dalam penilaian kemampuan dan potensi pembayaran.



Profesi/Pekerjaan	Profesi debitur dapat menjadi indikator sumber pendapatan dan stabilitas finansial.
Pendapatan perbulan	Pendapatan per bulan mencerminkan kemampuan finansial debitur untuk memenuhi kewajiban.
Plafond Kredit	Plafond kredit adalah batas maksimum pinjaman yang diberikan kepada debitur.
Angsuran per bulan	Angsuran per bulan adalah jumlah pembayaran cicilan yang harus dilakukan debitur setiap bulan.
Jangka Waktu/Durasi Pinjaman	Jangka waktu adalah durasi pinjaman yang disepakati antara debitur dan kreditur.
Status	Status adalah keterangan yang diberikan berdasarkan field yang ada sebelumnya, layak/tidak layak.

Berikut adalah data yang akan digunakan setelah pemilihan atribut:

Kolektibilitas	umur	Profesi	pendapatan_bulanan	plafond	angsuran_bulanan	Jangka_Waktu	Status
5	18		NULL	Rp 2.500.000.000,00	Rp 14.583.333,00	12	Tidak Layak
4	45	KARYAWAN SWASTA	Rp 108.067.747	Rp 500.000.000,00	Rp 7.500.000,00	12	Tidak Layak
1	46	WIRASWASTA	Rp 3.000.000	Rp 3.500.000,00	Rp 361.667,00	12	Layak
5	53	WIRASWASTA	Rp 1.148.653.509	Rp 700.000.000,00	Rp 9.100.000,00	12	Tidak Layak
5	18		NULL	Rp 5.000.000.000,00	Rp 29.166.667,00	12	Tidak Layak
5	44	WIRASWASTA	Rp 40.000.000	Rp 100.000.000,00	Rp 2.000.000,00	3	Tidak Layak
5	51	WIRASWASTA	Rp 100.000.000	Rp 250.000.000,00	Rp 5.000.000,00	2	Tidak Layak
5	55	KARYAWAN SWASTA	Rp 4.279.121	Rp 4.500.000,00	Rp 465.000,00	12	Tidak Layak
5	41	WIRASWASTA	Rp 12.158.875	Rp 20.000.000,00	Rp 400.000,00	4	Tidak Layak
5	41	WIRASWASTA	Rp 12.158.875	Rp 20.000.000,00	Rp 400.000,00	4	Tidak Layak
3	42	KARYAWAN SWASTA	Rp 4.471.000	Rp 10.000.000,00	Rp 983.333,00	12	Tidak Layak
3	44	KARYAWAN SWASTA	Rp 3.376.000	Rp 10.000.000,00	Rp 983.333,00	12	Tidak Layak
5	39	KARYAWAN SWASTA	Rp 3.680.000	Rp 5.000.000,00	Rp 908.333,00	6	Tidak Layak
5	37	KARYAWAN SWASTA	Rp 3.680.000	Rp 10.000.000,00	Rp 983.333,00	12	Tidak Layak
3	38	KARYAWAN SWASTA	Rp 3.376.000	Rp 8.000.000,00	Rp 786.667,00	12	Tidak Layak
3	33	KARYAWAN SWASTA	Rp 4.000.000	Rp 7.000.000,00	Rp 709.333,00	12	Tidak Layak
5	18		NULL	Rp 5.000.000.000,00	Rp 29.166.667,00	12	Tidak Layak
5	18		NULL	Rp 5.000.000.000,00	Rp 29.166.667,00	12	Tidak Layak
5	18		NULL	Rp 7.000.000.000,00	Rp 40.833.333,00	12	Tidak Layak
3	45	KARYAWAN SWASTA	Rp 5.955.000	Rp 10.000.000,00	Rp 983.333,00	12	Tidak Layak
5	33	KARYAWAN SWASTA	Rp 2.000.000	Rp 15.000.000,00	Rp 325.000,00	60	Tidak Layak
5	57	WIRASWASTA	Rp 7.500.000	Rp 3.000.000,00	Rp 590.000,00	6	Tidak Layak
5	46	WIRASWASTA	Rp 12.000.000	Rp 3.000.000,00	Rp 590.000,00	6	Tidak Layak

Gambar 3. 3 Data setelah pemilihan atribut

3.2.3 Data Cleaning

Melakukan pembersihan data, yaitu dengan menghapus data yang bernilai *null* atau kosong, memperbaiki kesalahan pada data. Terdapat beberapa data yang memiliki nilai *null*, berikut kesalahan data yang dimaksud dapat dilihat pada gambar dibawah:

Kolektibilitas	umur	Profesi	pendapatan bulanan	plafond	angsuran bulanan	Jangka Waktu	Status
5	18		NULL	Rp 2.500.000.000,00	Rp 14.583.333,00	12	Tidak Layak
4	45	KARYAWAN SWASTA	Rp 108.067.747	Rp 500.000.000,00	Rp 7.500.000,00	12	Tidak Layak
1	46	WIRASWASTA	Rp 3.000.000	Rp 3.500.000,00	Rp 361.667,00	12	Layak
5	53	WIRASWASTA	Rp 1.148.653.509	Rp 700.000.000,00	Rp 9.100.000,00	12	Tidak Layak
5	18		NULL	Rp 5.000.000.000,00	Rp 29.166.667,00	12	Tidak Layak
5	44	WIRASWASTA	Rp 40.000.000	Rp 100.000.000,00	Rp 2.000.000,00	3	Tidak Layak
5	51	WIRASWASTA	Rp 100.000.000	Rp 250.000.000,00	Rp 5.000.000,00	2	Tidak Layak
5	55	KARYAWAN SWASTA	Rp 4.279.121	Rp 4.500.000,00	Rp 465.000,00	12	Tidak Layak
5	41	WIRASWASTA	Rp 12.158.875	Rp 20.000.000,00	Rp 400.000,00	4	Tidak Layak
5	41	WIRASWASTA	Rp 12.158.875	Rp 20.000.000,00	Rp 400.000,00	4	Tidak Layak
3	42	KARYAWAN SWASTA	Rp 4.471.000	Rp 10.000.000,00	Rp 983.333,00	12	Tidak Layak
3	44	KARYAWAN SWASTA	Rp 3.376.000	Rp 10.000.000,00	Rp 983.333,00	12	Tidak Layak
5	39	KARYAWAN SWASTA	Rp 3.680.000	Rp 5.000.000,00	Rp 908.333,00	6	Tidak Layak
5	37	KARYAWAN SWASTA	Rp 3.680.000	Rp 10.000.000,00	Rp 983.333,00	12	Tidak Layak
3	38	KARYAWAN SWASTA	Rp 3.376.000	Rp 8.000.000,00	Rp 786.667,00	12	Tidak Layak
3	33	KARYAWAN SWASTA	Rp 4.000.000	Rp 7.000.000,00	Rp 709.333,00	12	Tidak Layak
5	18		NULL	Rp 5.000.000.000,00	Rp 29.166.667,00	12	Tidak Layak
5	18		NULL	Rp 5.000.000.000,00	Rp 29.166.667,00	12	Tidak Layak
5	18		NULL	Rp 7.000.000.000,00	Rp 40.833.333,00	12	Tidak Layak
3	45	KARYAWAN SWASTA	Rp 5.955.000	Rp 10.000.000,00	Rp 983.333,00	12	Tidak Layak
5	33	KARYAWAN SWASTA	Rp 2.000.000	Rp 15.000.000,00	Rp 325.000,00	60	Tidak Layak
5	57	WIRASWASTA	Rp 7.500.000	Rp 3.000.000,00	Rp 590.000,00	6	Tidak Layak
5	46	WIRASWASTA	Rp 12.000.000	Rp 3.000.000,00	Rp 590.000,00	6	Tidak Layak

Gambar 3. 4 Data yang terdapat nilai kosong atau *null*

Pada data profesi dan pendapatan bulanan terdapat banyak data yang bernilai kosong atau *null*, sehingga data tersebut harus dihapus agar dapat dilakukan perhitungan untuk proses *data mining*.

Berikut adalah data yang sudah dilakukan pembersihan dan jumlah data menjadi sebanyak 5553 data:

Kolektibilitas	Umur	Profesi	pendapatan bulanan	plafond	angsuran bulanan	Jangka Waktu	Status
4	45	KARYAWAN SWASTA	108067747	500000000	7500000	12	Tidak Layak
1	46	WIRASWASTA	3000000	3500000	361667	12	Layak
5	53	WIRASWASTA	1148653509	700000000	9100000	12	Tidak Layak
5	44	WIRASWASTA	40000000	100000000	2000000	3	Tidak Layak
5	51	WIRASWASTA	100000000	250000000	5000000	2	Tidak Layak
5	55	KARYAWAN SWASTA	4279121	4500000	465000	12	Tidak Layak
5	41	WIRASWASTA	12158875	20000000	400000	4	Tidak Layak
5	41	WIRASWASTA	12158875	20000000	400000	4	Tidak Layak
3	42	KARYAWAN SWASTA	4471000	10000000	983333	12	Tidak Layak
3	44	KARYAWAN SWASTA	3376000	10000000	983333	12	Tidak Layak
5	39	KARYAWAN SWASTA	3680000	5000000	908333	6	Tidak Layak
5	37	KARYAWAN SWASTA	3680000	10000000	983333	12	Tidak Layak
3	38	KARYAWAN SWASTA	3376000	8000000	786667	12	Tidak Layak
3	33	KARYAWAN SWASTA	4000000	7000000	709333	12	Tidak Layak
3	45	KARYAWAN SWASTA	5955000	10000000	983333	12	Tidak Layak
5	33	KARYAWAN SWASTA	2000000	15000000	325000	60	Tidak Layak
5	57	WIRASWASTA	7500000	3000000	590000	6	Tidak Layak
5	46	WIRASWASTA	12000000	3000000	590000	6	Tidak Layak
2	50	KARYAWAN SWASTA	3376000	8000000	786667	12	Layak
2	39	KARYAWAN SWASTA	3376000	10000000	983333	12	Layak
2	32	KARYAWAN SWASTA	3376000	8000000	786667	12	Layak
3	64	WIRASWASTA	20589699	50000000	1936957	46	Tidak Layak
5	34	KARYAWAN SWASTA	1950000	3500000	635833	6	Tidak Layak

Gambar 3. 5 Data setelah dilakukan *data cleaning*

3.2.4 *Data Transformation*

Adapun dalam tahap ini data akan diubah menjadi data yang sesuai agar dapat diproses *data mining*. Atribut yang akan diberi label untuk mengikuti kondisi data-data adalah Klasifikasi data Kolektibilitas. Kolektibilitas pada data sebelumnya adalah kolek 1, 2, 3, 4, dan 5. Berikut adalah masing-masing penjelasan dari kolek tersebut:

1. Kolek 1: Lancar

Kredit dibayar tepat waktu dan tidak ada tunggakan. Sehingga debitur memiliki kemampuan untuk melunasi kredit yang dimiliki.

2. Kolek 2: Dalam Pengawasan Khusus

Terdapat keterlambatan pembayaran hingga kurang lebih 90 hari atau terdapat indikasi penurunan kemampuan bayar, namun masih dapat terkendali.

3. Kolek 3: Kurang Lancar

Pembayaran mengalami keterlambatan 91 hingga kurang lebih 120 hari, dan mulai terlihat masalah yang dialami oleh debitur.

4. Kolek 4: Diragukan

Adanya keterlambatan 121 hingga 180 hari, prospek pelunasan rendah, namun terdapat kemungkinan untuk diselesaikan.

5. Kolek 5: Macet

Keterlambatan pembayaran melebihi 180 hari, memiliki kemungkinan kecil untuk melakukan pelunasan dan biasanya kredit dianggap gagal bayar.

Berdasarkan keterangan kolek sebelumnya, maka status kolek dibagi menjadi 2 yaitu lancar dan macet. Kemudian selanjutnya adalah hasil data yang akan digunakan setelah dilakukan *data selection* dan *data cleaning*, maka terdapat 2 nilai yaitu sebagai berikut:

Tabel 3. 3 Keterangan atribut kolektibilitas

Kolektibilitas	Status
1 & 2	Lancar
3, 4 & 5	Macet

Gambar 3.6 adalah data yang sudah dilakukan beberapa tahapan dimulai dari *data selection*, *data cleaning*, dan *data transformation*.

Kolektibilitas	Umur	Profesi	pendapatan_bulanan	plafond	angsuran_bulanan	Jangka_Waktu	Status
4	45	KARYAWAN SWASTA	108067747	500000000	7500000	12	Macet
1	46	WIRASWASTA	3000000	3500000	361667	12	Lancar
5	53	WIRASWASTA	1148653509	700000000	9100000	12	Macet
5	44	WIRASWASTA	40000000	100000000	2000000	3	Macet
5	51	WIRASWASTA	100000000	250000000	5000000	2	Macet
5	55	KARYAWAN SWASTA	4279121	4500000	465000	12	Macet
5	41	WIRASWASTA	12158875	20000000	400000	4	Macet
5	41	WIRASWASTA	12158875	20000000	400000	4	Macet
3	42	KARYAWAN SWASTA	4471000	10000000	983333	12	Macet
3	44	KARYAWAN SWASTA	3376000	10000000	983333	12	Macet
5	39	KARYAWAN SWASTA	3680000	5000000	908333	6	Macet
5	37	KARYAWAN SWASTA	3680000	10000000	983333	12	Macet
3	38	KARYAWAN SWASTA	3376000	8000000	786667	12	Macet
3	33	KARYAWAN SWASTA	4000000	7000000	709333	12	Macet
3	45	KARYAWAN SWASTA	5955000	10000000	983333	12	Macet
5	33	KARYAWAN SWASTA	2000000	15000000	325000	60	Macet
5	57	WIRASWASTA	7500000	3000000	590000	6	Macet
5	46	WIRASWASTA	12000000	3000000	590000	6	Macet
2	50	KARYAWAN SWASTA	3376000	8000000	786667	12	Lancar
2	39	KARYAWAN SWASTA	3376000	10000000	983333	12	Lancar
2	32	KARYAWAN SWASTA	3376000	8000000	786667	12	Lancar
3	64	WIRASWASTA	20589699	50000000	1936957	46	Macet
5	34	KARYAWAN SWASTA	1950000	3500000	635833	6	Macet

Gambar 3. 6 Data status setelah diubah

3.2.5 Data Splitting

Dataset yang digunakan dalam penelitian dibagi menjadi 2 (dua) bagian, yaitu data untuk *training* dan *testing*. Pembagian dilakukan dengan menggunakan rasio yang sering digunakan dalam penelitian yaitu rasio 70:30 (Joseph, 2022). *Data*

Training digunakan untuk membangun model agar dapat mempelajari pola-pola dari data, sedangkan *data testing* digunakan untuk mengevaluasi atau menilai performa model. Distribusi data dapat dilihat pada Tabel 3.4.

Tabel 3. 4 Jumlah Data *Training* dan Data *Testing*

Metode	Data <i>Training</i> (70%)	Data <i>Testing</i> (30%)	Total
<i>Naïve Bayes</i>	3887	1666	5553
<i>Random Forest</i>	3887	1666	5553

3.3 Model *Training*

3.3.1 *Naïve Bayes*

Melakukan *training* dengan menggunakan data yang sebelumnya sudah dibagi, yaitu 70% dari total data yang digunakan. Data *training* sebanyak 3887 data. Setelah model selesai dilatih, model tersebut disimpan dan digunakan untuk melakukan prediksi selanjutnya dimana prediksi tersebut menggunakan data *testing* yang berjumlah 1666 data untuk melakukan evaluasi performa klasifikasinya,

3.3.2 *Random Forest*

Algoritma *Random Forest* digunakan untuk melatih model klasifikasi berdasarkan data *training* sebanyak 3887 data. Algoritma yang menggabungkan sejumlah *decision tree* untuk mendapatkan hasil prediksi yang lebih stabil dan akurat dengan jumlah data yang kompleks. Prediksi akhir didapatkan melalui pemilihan mayoritas dari seluruh pohon untuk klasifikasi. Setelah melakukan pelatihan, model ini digunakan untuk memprediksi data testing sebanyak 1666 data.

Kedua model tersebut diimplementasikan menggunakan *RapidMiner* untuk keperluan analisis awal, kemudian model *Naïve Bayes* dan *Random Forest* diimplementasikan juga dalam bentuk aplikasi *website* dengan PHP *framework CodeIgniter*. Dengan sejumlah atribut nasabah seperti:

1. Jenis Kelamin
2. Usia
3. Pendapatan Bulanan
4. Pekerjaan
5. Plafond
6. Angsuran
7. Riwayat Pinjaman/ Kolektibilitas
8. Durasi Peminjaman
9. Label/Status

3.4 Model Evaluasi

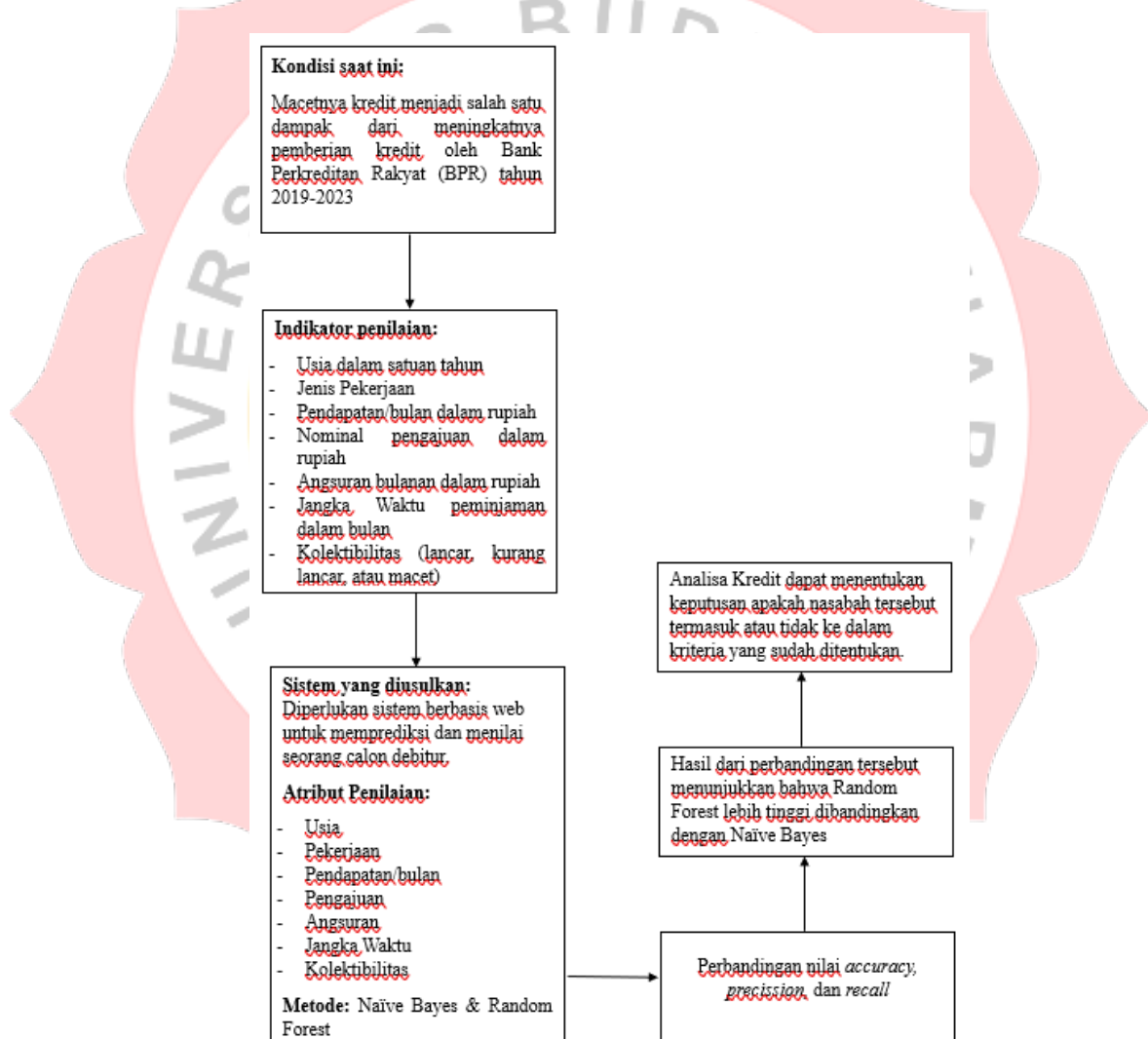
Tujuan dari model evaluasi ini untuk mengukur performa model klasifikasi terhadap data yang telah diuji. Setelah dilakukan proses *training* dan *testing*, evaluasi dilakukan untuk mengetahui sejauh mana model mampu melakukan prediksi secara akurat. Metrik yang digunakan sebagai berikut:

1. Akurasi, menggambarkan persentase prediksi yang benar secara menyeluruh.
2. Presisi, menunjukkan sejauh mana model mampu dalam mengklasifikasikan data positif yang benar.
3. Recall, mengukur kemampuan model dalam menemukan data yang benar-benar relevan.
4. F1-Score, nilai rata-rata dari presisi dan recall yang digunakan untuk menyeimbangkan.

3.4.1 Cross Validation

Selain menggunakan data *splitting* dengan persentase 70% *training* dan 30% *testing*. *Cross validation* juga digunakan agar hasil akhir menjadi lebih stabil. Dengan adanya model evaluasi ini dapat mengukur kemampuan model ketika menggunakan data yang baru, dan data yang digunakan dapat secara bergantian, yaitu data pernah menjadi data *training* dan pernah menjadi data *testing*.

3.5 Kerangka Pemikiran

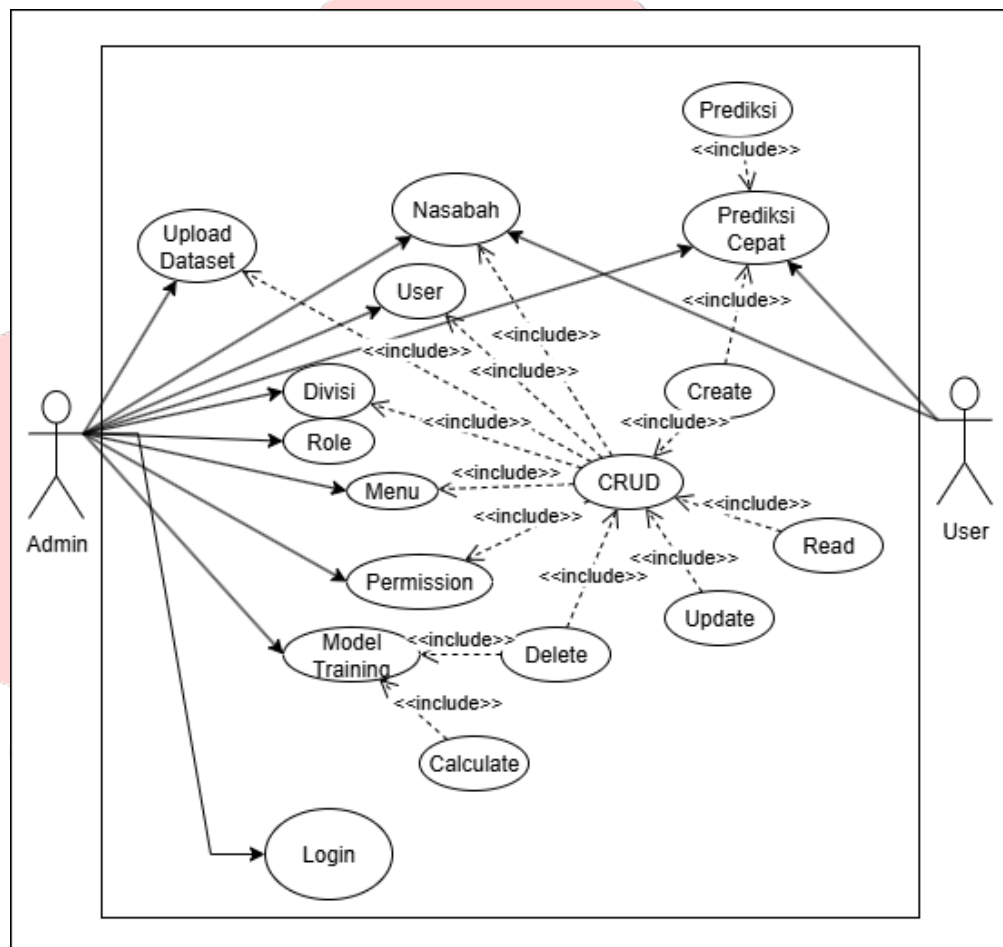


Gambar 3. 7 Kerangka Pemikiran

Gambar 3.7 adalah gambaran alur dari proses penelitian, dimulai dari mengidentifikasi masalah yang terjadi sampai tujuan yang ingin dicapai. Disusun berdasarkan teori, konsep, dan hasil penelitian yang diinginkan.

3.6 Perancangan UML

3.6.1 Use Case Diagram



Gambar 3. 8 Use Case Diagram

Gambar 3.8 menampilkan gambaran antara admin dan user dalam sistem yang dirancang. Terdapat dua (2) aktor, yaitu Admin (memiliki akses penuh) dan User (hanya fitur tertentu).

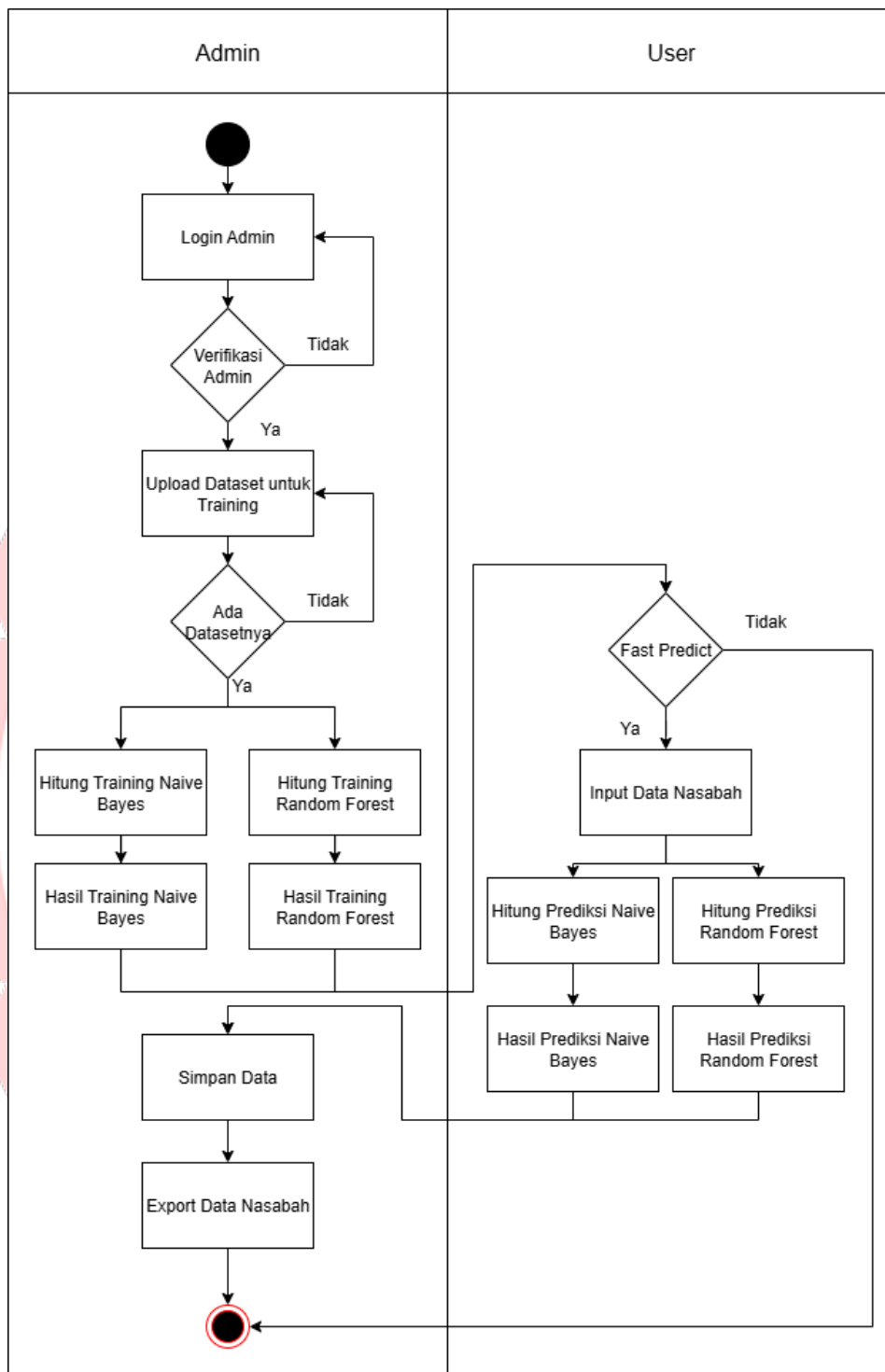
1. Admin, memiliki hak akses penuh terhadap seluruh fitur yang tersedia pada sistem, meliputi:

- Login, mengakses sistem menggunakan akun yang telah terdaftar
- Upload Dataset, mengunggah data yang diperlukan sebagai bahan penelitian atau untuk memproses prediksi
- Model Training, melakukan pelatihan model berdasarkan data yang diunggah
- Nasabah, mengelola terkait data nasabah dengan fungsi CRUD (*Create, Read, Update, Delete*)
- Prediksi cepat, melakukan prediksi terhadap data nasabah tanpa menyimpan ke dalam database
- *Permission/Role/Divisi/Menu/User*, manajemen sistem yang digunakan untuk mengatur secara keseluruhan dengan fungsi CRUD.

2. *User*, pengguna biasa yang hanya memiliki akses untuk menambahkan data nasabah dan melakukan prediksi cepat

- Nasabah, menambahkan data nasabah untuk melakukan prediksi
- Prediksi cepat, melakukan prediksi terhadap data nasabah tanpa menyimpan ke dalam database

3.6.2 Activity Diagram



Gambar 3. 9 Activity Diagram

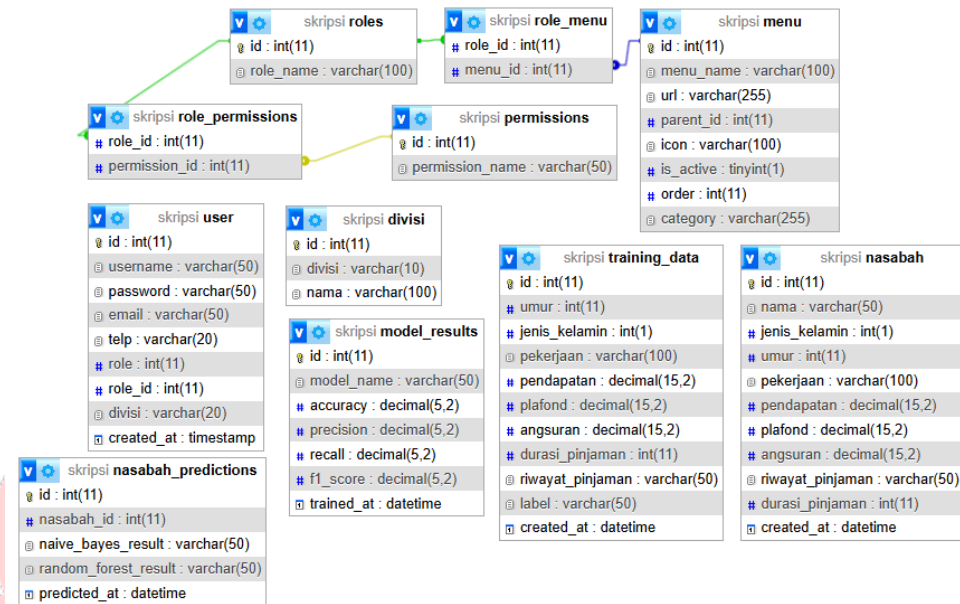
Gambar 3.9 menampilkan alur dari implementasi Website dengan admin dan user sebagai pelaku. Dimana terdapat dua aktor utama, yaitu Admin dan User. Proses dimulai dari Admin yang melakukan login ke dalam sistem. Setelah berhasil login,

dilakukan verifikasi terhadap identitas Admin. Jika verifikasi gagal, maka Admin harus mengulang proses login. Namun jika berhasil, Admin dapat melanjutkan untuk mengunggah dataset yang akan digunakan sebagai data latih (*training data*). Selanjutnya, sistem akan memverifikasi ketersediaan dataset. Jika dataset tidak tersedia, maka proses kembali ke tahap upload. Jika dataset tersedia, maka sistem akan menjalankan proses pelatihan model klasifikasi menggunakan dua algoritma yaitu *Naive Bayes* dan *Random Forest*. Setelah proses pelatihan selesai, sistem menampilkan hasil training dari masing-masing algoritma. Admin kemudian dapat menyimpan hasil training tersebut ke dalam database, dan selanjutnya memiliki opsi untuk mengekspor data nasabah yang telah diproses.

Sementara itu, di sisi *User*, proses dimulai dari keputusan apakah ingin menggunakan fitur *Fast Predict* atau tidak. Jika memilih untuk melanjutkan, *User* akan mengisi data nasabah yang ingin dinilai kelayakannya. Setelah data diinput, sistem akan melakukan proses klasifikasi menggunakan dua algoritma yang sama, yaitu *Naive Bayes* dan *Random Forest*. Hasil dari kedua metode tersebut akan ditampilkan kepada *User* sebagai hasil prediksi. Proses ini memungkinkan *User* untuk memperoleh informasi kelayakan kredit nasabah.

3.7 Perancangan Database dan Layar

3.7.1 Perancangan Database



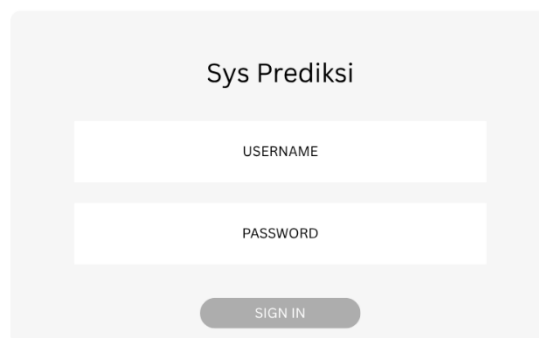
Gambar 3. 10 Perancangan Database

Berikut adalah penjelasan dari masing-masing tabel yang digunakan:

1. Tabel menu digunakan untuk menyimpan menu apa saja yang tersedia dalam sistem yang dibangun. Masing-masing menu memiliki status dan iconnya yang disimpan ke dalam *database*.
2. Tabel role_menu digunakan sebagai table relasi antara tabel roles dan tabel menu.
3. Tabel roles digunakan untuk menyimpan role pengguna sistem, contoh admin, user, dll.
4. Tabel permissions digunakan untuk menyimpan hak ases, contoh user a hanya dapat melihat data, sedangkan user b dapat melihat, membuat, mengubah, dan menghapus.
5. Tabel role_permissions digunakan sebagai tabel relasi antara tabel roles dan permissions.

6. Tabel user digunakan untuk menyimpan informasi siapa saja yang dapat masuk ke dalam sistem yang dibuat.
7. Tabel divisi digunakan untuk menyimpan nama-nama divisi apa saja yang ada dalam sistem, contoh divisi Analisa kredit, dll.
8. Tabel nasabah digunakan untuk menyimpan data-data nasabah pengajuan kredit, seperti nama, jenis kelamin, usia, pekerjaan, pendapatan, dsb.
9. Tabel training_data digunakan untuk menyimpan data yang akan digunakan untuk keperluan uji coba metode.
10. Tabel model_results digunakan untuk menyimpan hasil dari perhitungan menggunakan sistem, seperti hasil dari akurasi setiap metode, precision, recall, f1 score, dan kapan dilakukannya perhitungan.
11. Tabel nasabah_predictions digunakan untuk menyimpan hasil prediksi nasabah.

3.7.2 Perancangan Layar



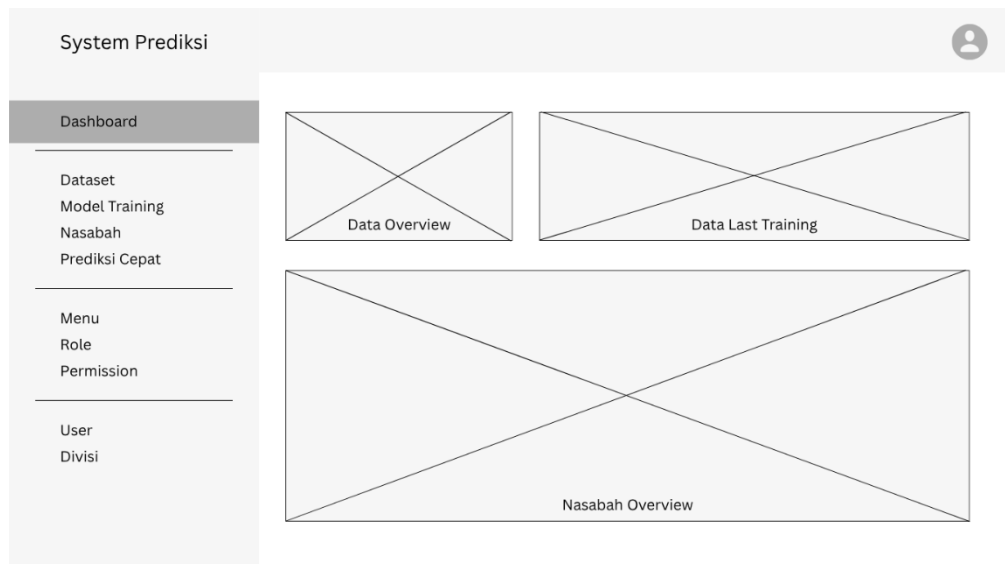
Sys Prediksi

USERNAME

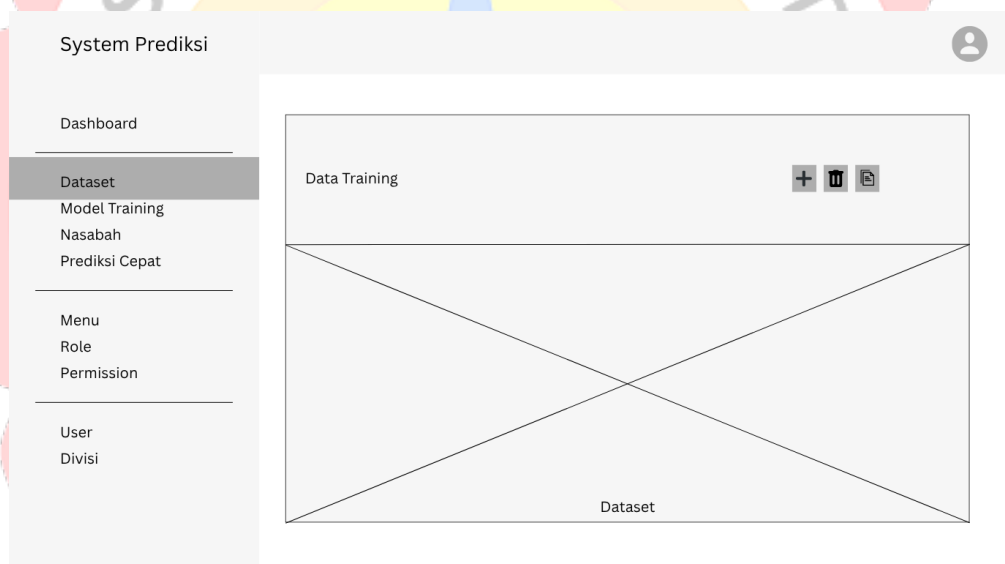
PASSWORD

SIGN IN

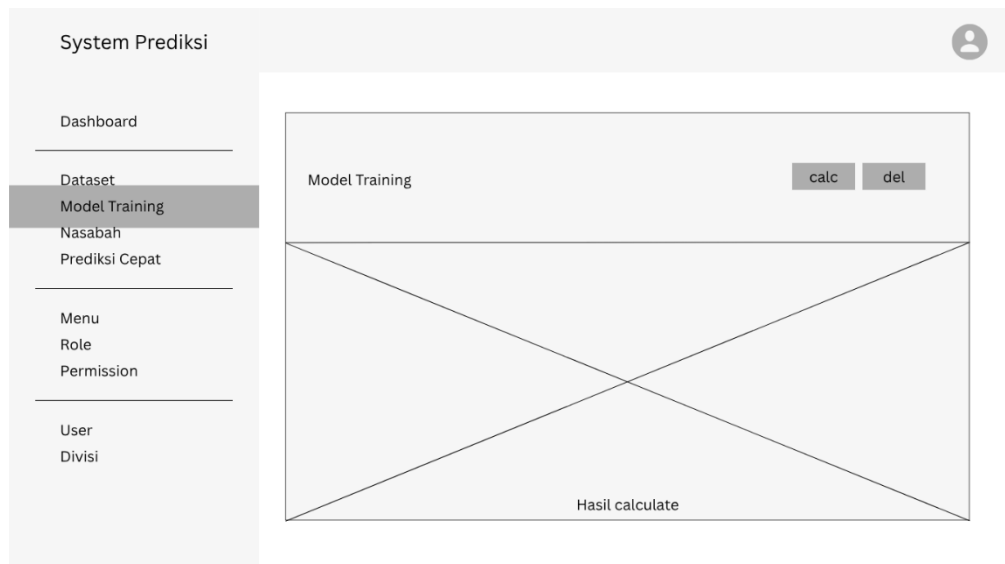
Gambar 3. 11 Halaman Login



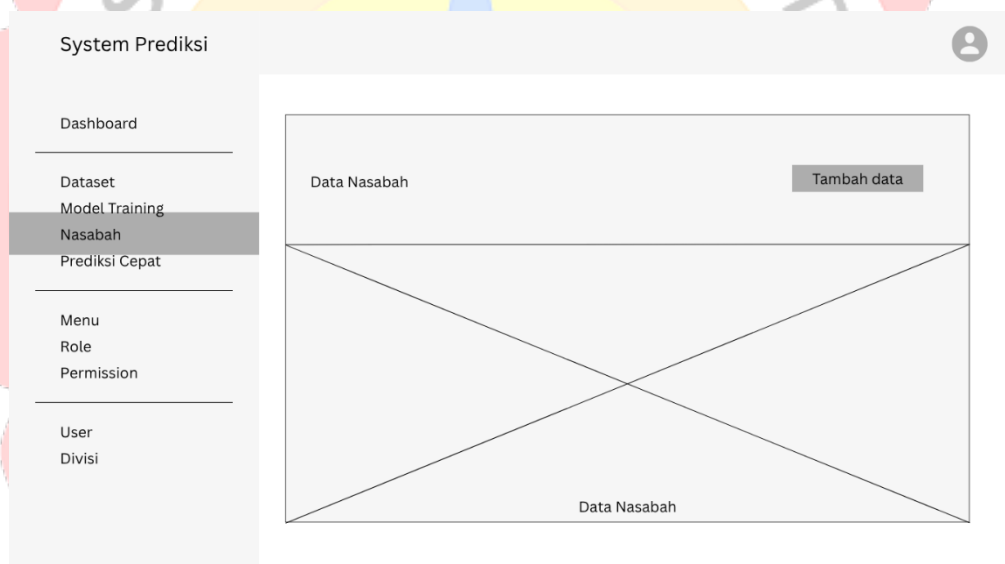
Gambar 3. 12 Halaman Dashboard Admin



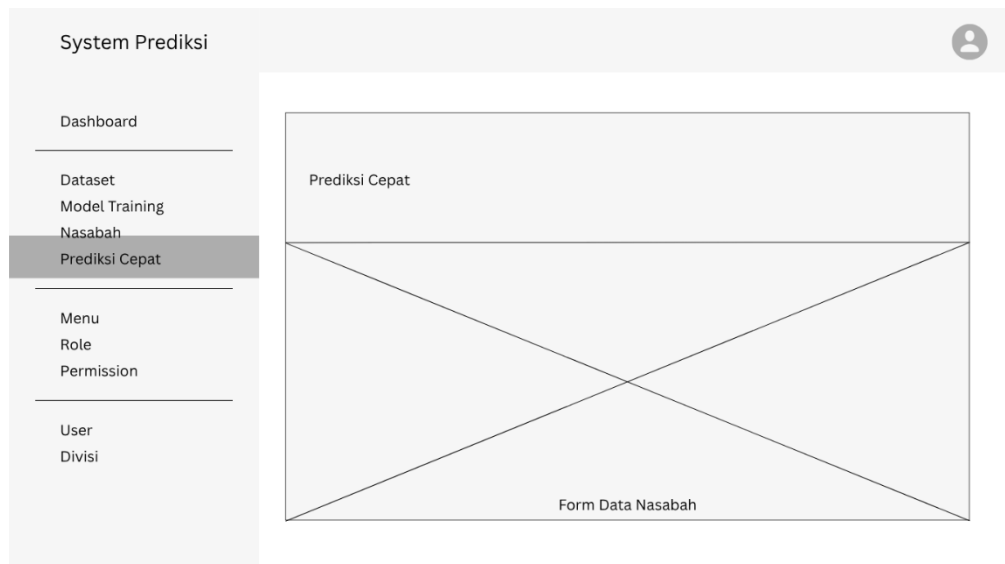
Gambar 3. 13 Halaman Dataset Admin



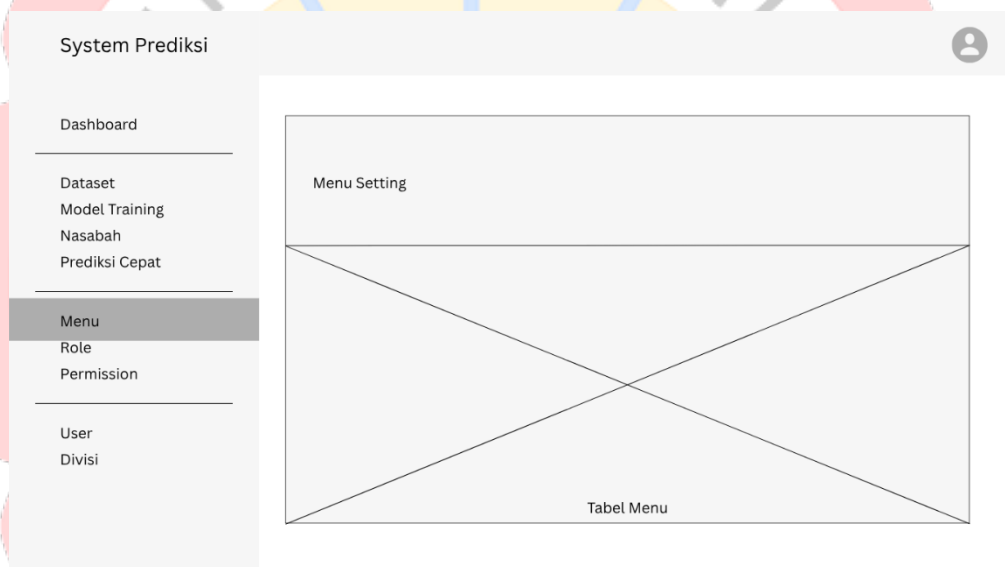
Gambar 3. 14 Halaman Model Training Admin



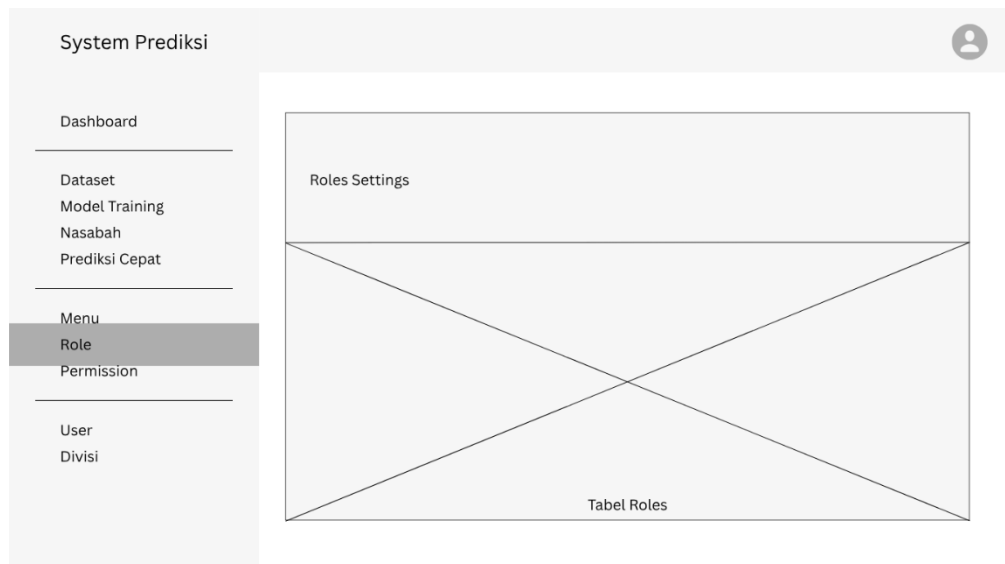
Gambar 3. 15 Halaman Nasabah Admin



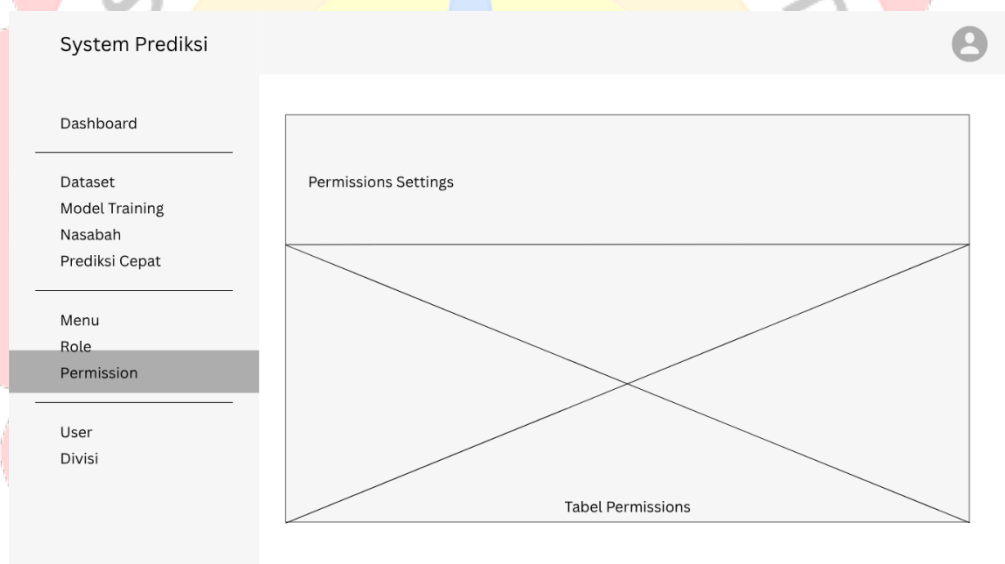
Gambar 3. 16 Halaman Prediksi Cepat Admin



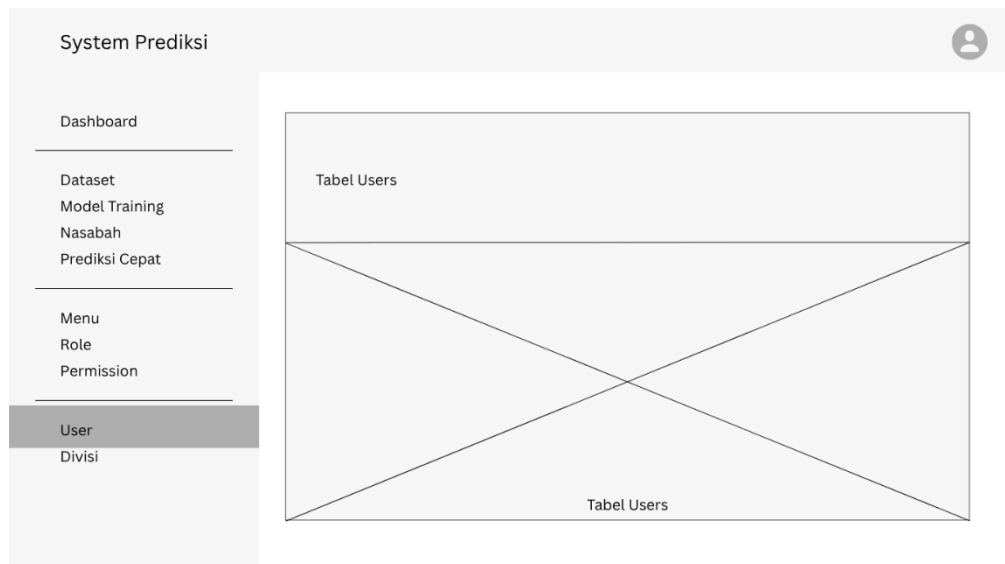
Gambar 3. 17 Halaman Menu Admin



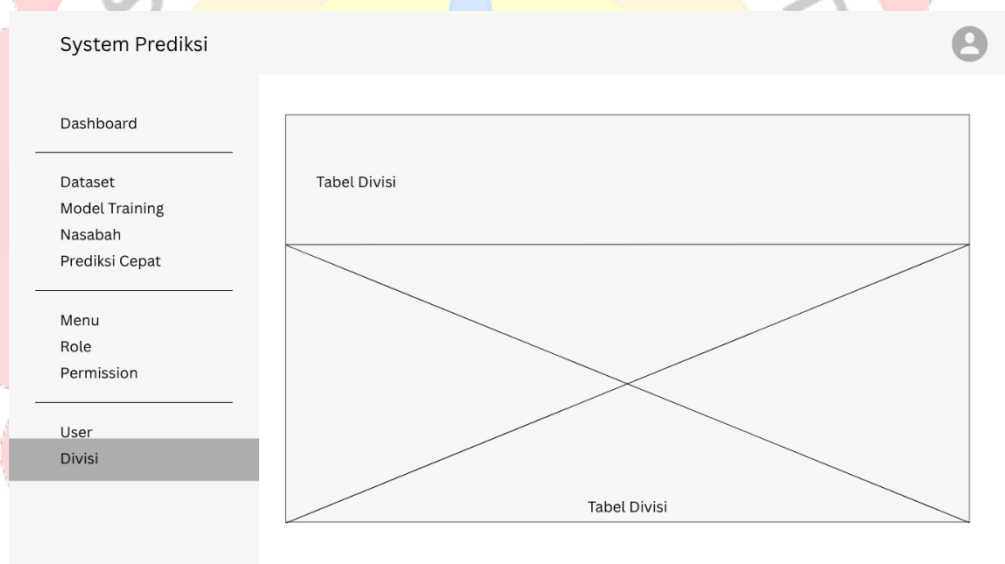
Gambar 3. 18 Halaman Role Admin



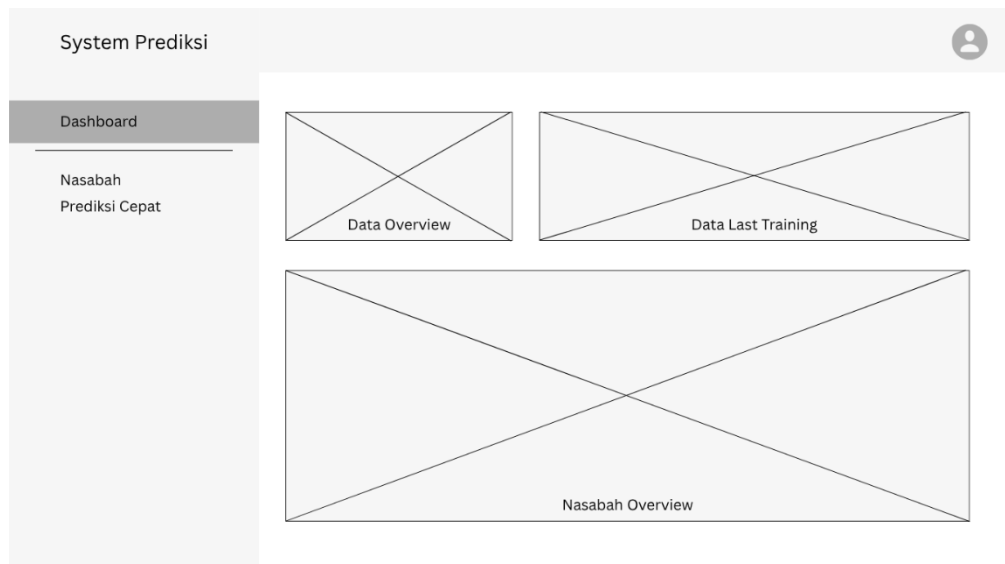
Gambar 3. 19 Halaman Permission Admin



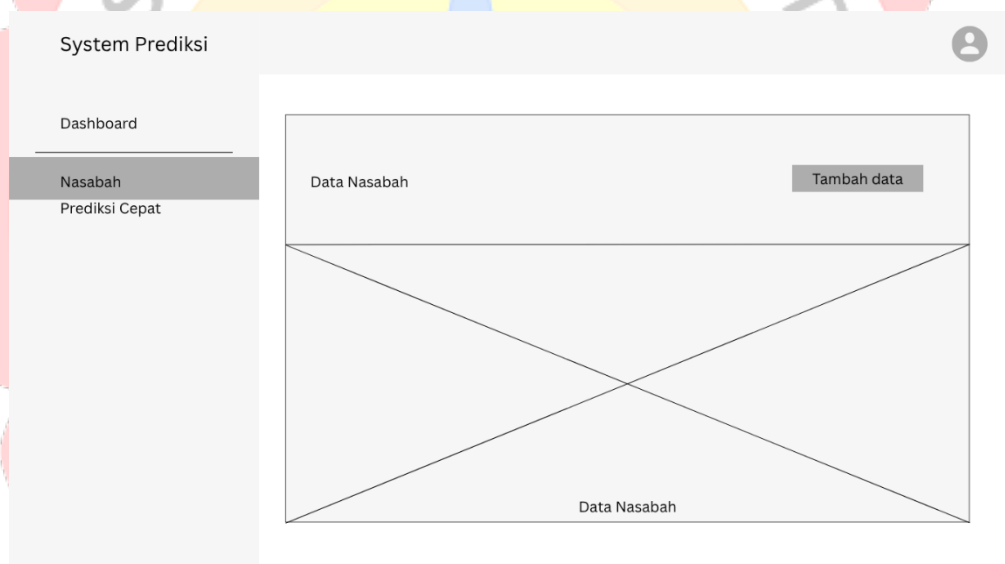
Gambar 3. 20 Halaman User Admin



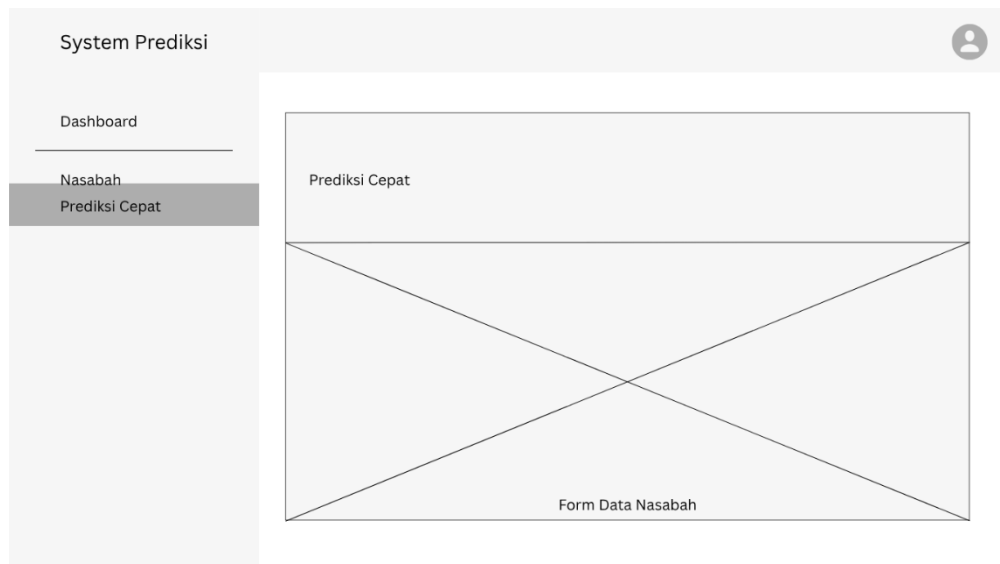
Gambar 3. 21 Halaman Divisi Admin



Gambar 3. 22 Halaman Dashboard *User*



Gambar 3. 23 Halaman Data Nasabah *User*



Gambar 3. 24 Halaman Prediksi

Gambar 3.11 sampai dengan 3.24 adalah desain yang telah dirancang untuk sebuah website sebelum dilakukan pengimplementasian. Untuk pengimplementasiannya terdapat dua (2) role, yaitu admin dan *user*. Untuk halaman login sebagai halaman awal sama baik untuk admin maupun untuk *user*. Akses menu pada role admin digambarkan pada Gambar 3.11 sampai dengan 3.21, sedangkan role *user* digambarkan pada Gambar 3.22 sampai dengan 3.24. Menu secara keseluruhan terdapat menu dashboard, dataset, model training, nasabah, prediksi cepat, menu, role, permission, user, divisi. Namun untuk role *user* hanya memiliki menu dashboard, nasabah, dan prediksi cepat.

3.8 Jadwal Penelitian

No.	Kegiatan	Keterangan	Maret				April				Mei				Juni				Juli			
			1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
1	Topik Penelitian	Membuat beberapa opsi terkait topik																				
2	Penyusunan Bab 1	Terkait masalah yang teridentifikasi, tujuan, manfaat, dan susunan lainnya																				
3	Penyusunan Bab 2	Mencari jurnal atau beberapa sumber sebagai pendukung penelitian																				
4	Penyusunan Bab 3	Wawancara Pengguna Sistem sebagai dasar pembentukan sistem																				
		Melakukan pemilihan data beserta atribut-atribut yang akan digunakan untuk penelitian																				
		Merancang tampilan yang akan dibangun																				
		Merancang database yang akan digunakan																				
5	Menyusun Bab 4	Mengimplementasikan metode menggunakan RapidMiner																				
		Mengimplementasikan metode menggunakan Website																				
		Melakukan uji coba blackbox																				
6	Menyusun Bab 5	Menarik Kesimpulan dari hasil penelitian																				

Gambar 3. 25 Jadwal Penelitian

Gambar 3.25 merupakan jadwal dalam melakukan penelitian, tahapan-tahapan yang dilakukan dari awal hingga akhir penelitian yang disusun dalam bentuk tabel atau diagram (Gantt Chart). Dengan adanya jadwal penelitian, kegiatan yang telah direncanakan lebih terstruktur dan menjadi lebih efektif.