

BAB II

LANDASAN TEORI DAN KERANGKA KONSEP

2.1 Tinjauan Pustaka

2.1.1 Pengertian Kredit

Istilah kredit berasal dari bahasa Latin *credo*, yang berarti ‘saya percaya’, yang merupakan kombinasi dari bahasa Sanskerta *cred* yang artinya ‘kepercayaan’ dan bahasa Latin *do* yang artinya ‘saya tempatkan’. Kredit yang diberikan oleh finance atas dasar kepercayaan sehingga pemberian kredit merupakan pemberian kepercayaan finance kepada nasabahnya[5]. Menurut [6] undang-undang no.10 tahun 1998 tentang perubahan undang-undang no.7 tahun 1992 undang-undang tentang pembiayaan kendaraan, dalam bab 1 ketentuan umum pasal 1, yang dimaksud dengan kredit adalah penyediaan uang atau tagihan yang dapat dipersamakan dengan itu, berdasarkan persetujuan atau kesepakatan pinjam-meminjam antara finance dan pihak lain yang mewajibkan pihak peminjam untuk melunasi hutangnya setelah jangka waktu tertentu dengan pemberian bunga.

Peranan finance sebagai lembaga keuangan berperan besar dalam pemberian kredit kepada masyarakat tepatnya kepada nasabahnya, dimana pemberian kredit merupakan kegiatan utamanya. Sebagai sebuah lembaga keuangan, aset terbesar yang dimiliki oleh finance adalah aset financial dan aset utama dari finance adalah kredit yang disalurkan kepada debitur dalam hal ini adalah nasabahnya.

Ada beberapa kriteria dalam penilaian kredit kepada nasabah yang harus dilakukan atau prinsip utama yang berkaitan dengan kondisi calon nasabah. Prinsip itu dikenal dengan istilah 5C [3] yaitu :

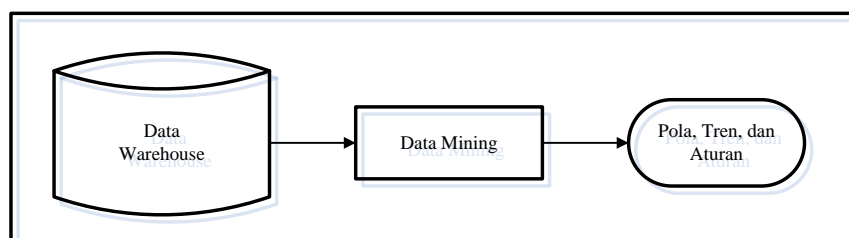
1. *Character*

Penilaian terhadap karakter atau kepribadian calon peminjam untuk memperkirakan kemungkinan bahwa peminjam dapat memenuhi kewajibannya.

2. *Capacity*
Kemampuan peminjam (*debitur*) untuk melakukan pembayaran. Kemampuan ini di ukur dari catatan prestasi peminjam di masa lalu.
3. *Capital*
Penilaian terhadap kemampuan modal yang dimiliki oleh calon peminjam, di ukur dengan posisi usaha atau jumlah penghasilan.
4. *Collateral*
Jaminan yang dimiliki calon peminjam. Penilaian ini untuk lebih meyakinkan bahwa jika suatu resiko kegagalan pembayaran terjadi, maka jaminan dapat dipakai sebagai pengganti kewajiban.
5. *Conditions*
Pihak *finance* akan melihat jenis usaha yang dilakukan oleh calon peminjam, hal ini dilakukan karena kondisi eksternal memiliki pengaruh yang kuat atau cukup besar dalam proses berjalannya usaha dari calon peminjam dalam jangka panjang.

1.1.2 *Data Mining*

Data mining atau penambangan data adalah perangkat lunak yang digunakan untuk menemukan pola tersembunyi, tren, maupun aturan-aturan yang terdapat dalam basis berukuran besar dan menghasilkan aturan-aturan yang digunakan untuk memperkirakan perilaku di masa mendatang[4].



Gambar 2.1 Prinsip *Data Mining*[7]

Data mining terdiri dari *extract*, *transform*, dan memuat data transaksi ke sistem *data warehouse*, disimpan dan mengelola data dalam sistem database multidimensi, menyediakan akses data untuk analisis bisnis dan profesional

teknologi informasi, menganalisis data oleh aplikasi perangkat lunak, menyajikan data dengan format yang berguna, seperti grafik atau tabel *Data mining* sering dikatakan berurusan dengan "penemuan pengetahuan" dalam basis data. Hal yang menarik, *data mining* menjadi perangkat yang membantu para pemakai untuk menemukan jawaban atas pertanyaan-pertanyaan yang tidak pernah mereka pikirkan sebelumnya.

Istilah[4] *data mining* memiliki hakikat sebagai disiplin ilmu yang tujuan utamanya adalah untuk menemukan, menggali, atau menambang pengetahuan dari data atau informasi yang kita miliki. Ada beberapa peran utama dalam *data mining* antara lain:

1. *Description*

Cara yang digunakan untuk menggambarkan sekumpulan data secara ringkas. Banyak cara yang digunakan dalam memberikan gambaran secara ringkas bagi sekumpulan data yang besar jumlahnya dan banyak macamnya, yaitu Deskripsi Grafis, Deskripsi Lokasi dan Deskripsi keragaman.

2. *Estimation*

Algoritma estimasi yang biasa digunakan adalah: *Linear Regression, Neural Network, Support Vector Machine*. Algoritma estimasi mirip dengan algoritma klasifikasi, tapi variabel target adalah berupa bilangan numerik dan bukan kategorikal (nominal). Model dibangun dari data dengan *record* yang lengkap, yang menyediakan nilai dari variabel sebagai prediktor, kemudian estimasi nilai dari variabel target ditentukan berdasarkan nilai dari variabel prediktor. Penentuan kebijakan atau suatu nilai pada proses yang akan dilakukan. Estimasi dapat dilakukan dari data-data lama yang akan diolah.

3. *Prediction*

Algoritma prediksi sama dengan algoritma estimasi dimana label/target/class bertipe numerik, bedanya adalah data yang digunakan merupakan data rentetan waktu (*data time series*). Sifat prediksi bisa menghasilkan *class* berdasarkan berbagai atribut yang kita sediakan. Penentuan hasil dari proses yang sedang berlangsung. Data-data yang

digunakan untuk prediksi berasal dari data yang ada saat proses sedang berlangsung. Istilah prediksi kadang digunakan juga untuk klasifikasi, tidak hanya untuk prediksi *time series*, karena sifatnya yang bisa menghasilkan *class* berdasarkan berbagai atribut yang kita sediakan.

4. *Classification*

Algoritma yang menggunakan data dengan *target/class/label* berupa nilai kategorikal (nominal). Pengelompokan data-data yang ada menjadi dalam kelompok yang sudah ditentukan nama kelompoknya. Metode yang cocok untuk klasifikasi, yakni: *Naïve Bayes*, *K-Nearest Neighbor*, *C4.5*, *ID3*, *CART*, *Linear Discriminant Analysis*, dan yang lainnya. Contoh, apabila *target/class/label* adalah pendapatan, maka bisa digunakan nilai nominal (kategorikal) pendapatan besar, menengah, kecil.

5. *Clustering*

Klastering adalah pengelompokan data, hasil observasi dan kasus ke dalam *class* yang mirip. Suatu klaster (*cluster*) adalah koleksi data yang mirip antara satu dengan yang lain, dan memiliki perbedaan bila dibandingkan dengan data dari klaster lain. Metode yang cocok untuk klastering, yakni: *K-Means*, *K-Medoids*, *Self-Organizing Map (SOM)*, *Fuzzy C-Means*, dan yang lainnya. Perbedaan utama algoritma klastering dengan klasifikasi adalah klastering tidak memiliki *target/class/label*, jadi termasuk *unsupervised learning*. Klastering sering digunakan sebagai tahap awal dalam proses *data mining*, dengan hasil klaster yang terbentuk akan menjadi input dari algoritma berikutnya yang digunakan.

6. *Association*

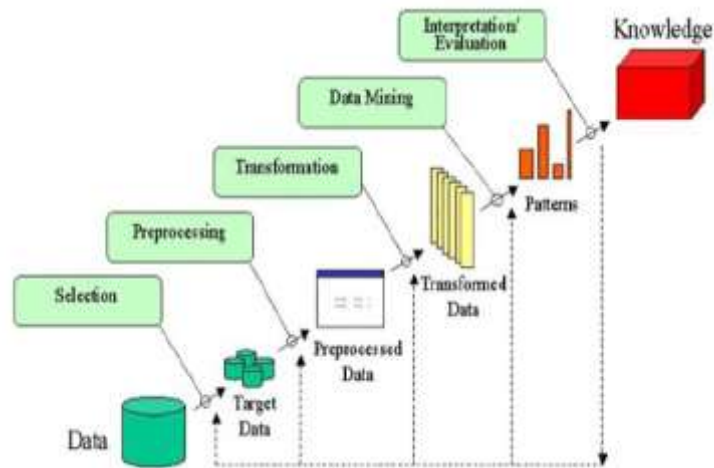
Algoritma *association rule* (aturan asosiasi) adalah algoritma yang menemukan atribut yang “jalan bersamaan”. Dalam dunia bisnis, sering disebut dengan *affinity analysis* atau *market basket analysis*. Algoritma *association rules* berangkat dari pola “*If antecedent, then consequent*,” bersamaan dengan pengukuran *support (coverage)* dan *confidence (accuracy)* yang terasosiasi dalam aturan. Algoritma *association rule*

diantaranya adalah: *Apriori algorithm*, *FP-Growth algorithm*, *GRI algorithm*.

Berikut beberapa alasan *data mining* menjadi penting saat ini:

1. Data yang tersedia sangat besar:
Selama *decade* terakhir harga *hardware* terutama harga *hardisk* telah menurun drastis. Bersama dengan ini, perusahaan telah menyimpan data yang sangat besar melalui banyak aplikasi. Dengan semua data ini untuk mengeksplorasi, perusahaan menginginkan dapat menemukan pola dan informasi yang tersembunyi sehingga dapat membantu mengarahkan strategi bisnis perusahaan menjadi lebih baik.
2. Meningkatkan kompetisi:
Kompetisi yang tinggi sebagai dampak dari modernisasi pasar dan distribusi seperti *internet* dan telekomunikasi. Perusahaan-perusahaan di seluruh dunia menghadapi persaingan dan kunci untuk keberhasilan bisnis adalah kemampuan untuk mempertahankan pelanggan dan memperoleh pelanggan baru. *Data mining* adalah teknologi yang memungkinkan perusahaan untuk menganalisis faktor-faktor yang mempengaruhi masalah ini.
3. Teknologi yang telah siap:
Teknologi *data mining* yang sebelumnya hanya ada dalam wilayah akademik, tapi sekarang banyak dari teknologi tersebut telah matang dan siap diterapkan dalam perusahaan. Algoritma yang lebih akurat, efektif dan dapat menangani data yang semakin banyak dan rumit. Selain itu pemrograman aplikasi antar muka *data mining* telah distandarisasi yang akan memungkinkan para pengembang untuk membangun aplikasi *data mining* akan lebih baik.

Sebagai suatu rangkaian proses, *data mining* dapat dibagi menjadi beberapa tahap proses yang diilustrasikan pada gambar II-2.



Gambar 2.2 Proses *Data Mining*[8]

Pada gambar 2.2 digambarkan proses *data mining* adalah data dipilih, dibersihkan, dan dilakukan *preprocessing* mengikuti pedoman dan *knowledge* dari ahli domain yang menangkap dan mengintegrasikan *data internal* dan *eksternal* ke dalam tinjauan organisasi secara menyeluruh. Penggunaan algoritma *data mining* dilakukan pada langkah ini untuk menggali data yang terintegrasi untuk memudahkan identifikasi informasi bernilai. Keluaran dari *data mining* dievaluasi untuk melihat apakah *knowledge domain* ditemukan dalam bentuk rule yang telah diekstraksi dari jaringan.

Adapun proses dari KDD(*Knowledge Discovery in Database*) sebagai berikut [8]:

1. *Data Selelection*

Pemilihan data dari sekumpulan data operasional perlu dilakukan tahap penggalan informasi sebelum KDD di mulai.

2. *Preporcessing*

Sebelum proses *data mining* dapat dilaksanakan, perlu dilakukan proses *cleaning* dengan tujuan untuk membuang duplikasi data, memeriksa data yang inkonsisten, dan memperbaiki kesalahan pada data, seperti kesalahan cetak (tipografi). Juga dilakukan proses *enrichment*,

yaitu proses “memperkaya” data yang sudah ada dengan data atau informasi lain yang relevan dan diperlukan untuk KDD, seperti data atau informasi eksternal.

3. *Transformation*

Proses *coding* pada data yang telah dipilih, sehingga data tersebut sesuai untuk proses *data mining*. Proses *coding* dalam KDD merupakan proses kreatif dan sangat tergantung pada jenis atau pola informasi yang akan dicari dalam *database*

4. *Data Mining*

Proses mencari pola atau informasi menarik dalam data terpilih dengan menggunakan teknik atau metode tertentu

5. *Interpretation / Evaluation*

Pola informasi yang dihasilkan dari proses *data mining* perlu ditampilkan dalam bentuk yang mudah dimengerti oleh pihak yang berkepentingan. Tahap ini merupakan bagian dari proses KDD yang disebut dengan *interpretation*. Tahap ini mencakup pemeriksaan apakah pola atau informasi yang ditemukan bertentangan dengan fakta atau hipotesa yang ada sebelumnya atau tidak.

Evaluasi adalah kunci ketika membuat aplikasi berbasis *data mining*. Ada berbagai macam cara melakukan evaluasi. Evaluasi ternyata tidak sesederhana yang kita bayangkan. Jika kita memiliki data yang kita gunakan dalam proses pelatihan, maka tidak serta merta kita menjadikan data tersebut sebagai indikator keberhasilan aplikasi yang kita buat. Oleh karena itu, kita membutuhkan metode tertentu guna memprediksi performa berdasarkan eksperimen untuk berbagai macam data selain *data training* tersebut.

Pada umumnya data yang cukup banyak dapat dimanfaatkan untuk pengujian. Hanya saja masalah yang kerap dijumpai adalah datanya. Oleh karena itu, kita harus memastikan data yang akan kita gunakan baik untuk pelatihan maupun untuk pengujian merupakan data yang berkualitas.

2.1.3 Klasifikasi

Klasifikasi dalam *data mining* merupakan metode pembelajaran data untuk memprediksi nilai dari sekelompok atribut. Algoritma klasifikasi akan menghasilkan sekumpulan aturan yang disebut *rule* yang akan digunakan sebagai *indicator* untuk dapat memprediksi kelas dari data yang ingin diprediksi[1]. Klasifikasi digunakan dalam banyak sekali bidang, dan secara teori algoritma klasifikasi sama seperti otak manusia.

Otak manusia mampu mengolah data yang sudah ada sebagai pengalaman dalam bertindak. Dalam *data mining* ada beberapa algoritma klasifikasi yang banyak digunakan dalam masyarakat atau dalam penelitian secara luas diantaranya adalah *Decision / classification trees*, *Bayesian classifier / Naïve Bayes classifiers*, *Neural networks*, Analisa Statistik, Algoritma Genetika, *Rough sets*, *K-Nearest Neighbor*, metode *Rule Based*, *Memory Based reasoning*, dan *Support Vector Machine (SVM)*.

Tujuan dari algoritma klasifikasi adalah untuk menemukan relasi antara beberapa variable yang tergolong dalam kelas yang sama. Relasi tersebut akan digambarkan dengan aturan-aturan agar dapat memprediksi kelas dari data yang *attribute* nya sudah diketahui.

Penilaian algoritma klasifikasi biasanya dilihat dari akurasi model. Akurasi model merupakan ketepatan model dalam memprediksi kelas data. Selain akurasi kecepatan pembentukan model, kemampuan algoritma dalam mengatasi data yang tidak relevan atau bahkan data yang tidak lengkap, serta kemampuan algoritma ketika diterapkan pada data jumlah besar maupun kecil.

2.1.4 K-Fold Cross Validation

Dalam penelitian ini,[4] metode yang digunakan untuk menguji pola klasifikasi adalah dengan metode *k-fold cross validation*. Dalam *k-fold cross validation* data dibagi menjadi *k* bagian, D_1, D_2, \dots, D_k , dan masing-masing D memiliki jumlah data yang sama. Kemudian akan dilakukan proses perulangan sebanyak *k*, dimana dalam setiap perulangan ke-*i*, D_1 akan dijadikan *data testing*, dan sisanya akan digunakan sebagai *data training*.

Pengujian dengan k=5 atau k=10 dapat digunakan untuk memperkirakan tingkat kesalahan yang terjadi, sebab data training pada setiap *fold* cukup berbeda dengan *data training* yang asli. Secara keseluruhan, 5 atau 10 *fold cross validation* sama-sama direkomendasikan dan disepakati bersama. Menghitung nilai akurasi dapat dilakukan dengan menggunakan persamaan:

$$\text{Akurasi} = \frac{\text{Jumlah klasifikasi benar}}{\text{Jumlah data uji}} \times 100 \% \quad (2.1)$$

2.1.5 Algoritma *Naïve Bayes*

Algoritma *Naïve Bayes* merupakan metode yang dapat digunakan untuk mengklasifikasikan sekumpulan data. Algoritma ini memanfaatkan metode probabilitas dan Statistik yang dikemukakan oleh ilmuwan Inggris Thomas Bayes, yaitu memprediksi probabilitas di masa depan berdasarkan pengalaman di masa sebelumnya [5][4].

Naïve Bayes merupakan *machine learning* yang menggunakan perhitungan probabilitas yang menggunakan konsep pendekatan *Bayesian*. Kata *Naïve*, yang terkesan merendahkan, berasal dari asumsi *independensi* pengaruh nilai suatu atribut dari probabilitas pada kelas yang diberikan terhadap nilai atribut lainnya. Penggunaan teorema Bayes pada algoritma *Naïve Bayes* yaitu dengan mengkombinasikan prior probability dan probabilitas bersyarat dalam sebuah rumus yang bisa digunakan untuk menghitung probabilitas tiap klasifikasi yang mungkin.

Rumus *Naïve Bayes* nya adalah :

$$P(H|X) = \frac{P(H)P(X|H)}{P(X)} \quad (2.2)$$

Keterangan :

- X = data dengan kelas yang belum diketahui
- H = hipotesis data X, merupakan suatu kelas yang spesifik
- P(H|X) = probabilitas hipotesis H berdasar kondisi X (posteriori probability)
- P(H) = probabilitas hipotesis H (posteriori probability)

$P(X|H)$ = probabilitas X berdasar kondisi H

$P(X)$ = probabilitas dari X

atau

$$Posterior Probability = \frac{Prior Probability \times likelihood}{evidance} \quad (2.3)$$

2.1.6 Algoritma *Decision Tree*

Decision Tree adalah sebuah struktur pohon, dimana setiap node pohon merepresentasikan atribut yang telah di uji, setiap cabang merupakan suatu pembagian hasil uji, dan node daun (*leaf*) merepresentasikan kelompok kelas tertentu. *Level node* teratas dari sebuah *Decision Tree* adalah node akar (*root*) yang biasanya berupa atribut yang paling memiliki pengaruh terbesar pada kelas tertentu[4].

Pohon keputusan sendiri merupakan pendekatan “*divide and conquer*” dalam mempelajari masalah dari sekumpulan data independen yang digambarkan dalam bagan pohon. Pohon keputusan juga merupakan sekumpulan pertanyaan yang tersusun secara sistematis, dimana setiap pertanyaan yang ada menentukan percabangan berdasarkan nilai atribut dan berhenti pada daun dari pohon yang merupakan prediksi dari kelas variabel.

Semua metode yang menghasilkan pohon keputusan bekerja mulai dari akar yang paling atas, dengan menguji data yang ada dan dipecah kedalam setiap cabang dari pohon yang lebih kecil. Keputusan dipilih berdasarkan kesesuaian antara data yang diuji dengan aturan dari cabang pohon keputusan yang ada hingga mencapai daun sehingga nilai kelas yang terdapat pada daun diberikan pada tupel data yang diperiksa tersebut.

Ada beberapa proses yang harus diperhatikan dalam pembentukan struktur pohon, yaitu :

- a) Pilih *root* berdasarkan *gain ratio* terbesar
- b) Pilih internal *root* / cabang *root* berdasar *gain ratio* terbesar setelah menghapus atribut yang telah terpilih sebagai *root*.
- c) Ulangi sampai semua atribut terhitung nilai *gain rationya*.

Dalam pembuatan pohon keputusan, setiap algoritma menerapkan ukuran pemilihan atribut yang berbeda-beda. Ukuran pemilihan atribut merupakan ukuran yang digunakan dalam menentukan kriteria yang terbaik untuk mengelompokkan *tuple*. Ukuran pemilihan atribut ini juga disebut sebagai *splitting rules* karena menentukan bagaimana data akan dipisahkan kesetiap cabang. C4.5 yang merupakan pengembangan dari ID3 menggunakan *Information gain* untuk ukuran pemilihan atribut.

Information gain diciptakan oleh Claude Shannon dengan mempelajari nilai informasi dari data, dan menggunakan nilai tersebut sebagai acuan dalam menentukan atribut yang akan digunakan dalam menyusun pohon keputusan. *Atribut* yang dipilih akan menghasilkan partisi dengan data yang lebih seragam, dan dapat menghasilkan pohon keputusan yang sesederhana mungkin dengan perulangan yang sedikit. Berikut persamaan data dalam *tuple* D.

$$Info(D) = \sum_{i=1}^n -p_i \log_2(p_i) \quad (2.4)$$

Dimana p_i merupakan probabilitas tuple dalam D yang menjadi kelas C_i dengan asumsi $|C_{i,D}|/|D|$. $Info(D)$ atau disebut juga *entropy* dari D merupakan rata rata informasi yang diperlukan untuk identifikasi tuple dalam D.

Jika Nilai A adalah nilai diskrit maka data D akan dipisahkan sejumlah nilai data A sehingga nilai setiap cabang akan murni dan sejenis. Setelah percabangan pertama, jumlah percabangan yang mungkin terjadi diukur dengan persamaan.

$$Info_A(D) = \sum_{j=1}^v \frac{|D_j|}{|D|} \times Info(D_j) \quad (2.5)$$

Dimana $\frac{|D_j|}{|D|}$ merupakan bobot dari partisi j. $Info_A(D)$ merupakan informasi yang diperlukan untuk mengklasifikasikan tuple dari D pada partisi A. Semakin kecil hasil persamaan ini, semakin baik pula partisi yang dihasilkan. Nilai dari sebuah atribut menentukan penting tidaknya atribut tersebut dalam penyusunan pohon keputusan. Semakin besar nilai atribut berarti semakin baik pula klasifikasi yang akan terbentuk jika cabang dipecah menurut atribut tersebut, untuk melihat nilai dari atribut A digunakan persamaan.

$$Gain(A) = Info(D) - Info_A(D) \quad (2.6)$$

2.1.7 Pemilihan Variabel

Pemilihan variabel yang juga disebut sebagai pemilihan atribut, digunakan pada *dataset* untuk menemukan pola yang penting dalam *data mining*. Pemilihan variabel digunakan untuk pengurangan dimensi pada *dataset*. Pemilihan variabel digunakan untuk melakukan eliminasi variabel yang tidak *relevan* dan *redundan*, yang dapat menyebabkan kebingungan dalam penggunaan variabel.

Pemilihan variabel dapat mengurangi dimensi data, hal ini memungkinkan lebih efektif dalam operasi agar lebih cepat dari beberapa algoritma data mining. Dengan adanya pemilihan variabel membuat algoritma data mining lebih cepat dan lebih efektif.

Penggunaan pemilihan variabel pada *dataset* yang menggunakan variabel bebas dapat meningkatkan performa model. Pemilihan variabel juga merupakan proses yang cukup memakan biaya, dan juga berten.tangan dalam asumsi awal, bahwa semua informasi diperlukan untuk mencapai akurasi yang maksimal.

Metode yang dapat digunakan untuk pemilihan variabel antara lain *Backward Elimination*, *Forward Selection*, *Genetic Algorithm*, dan yang lainnya. Metode-metode tersebut digunakan dalam penelitian *data mining* agar dapat menghasilkan variabel yang relevan dalam penelitian.

Pemilihan variabel dengan filter model ini lebih murah dalam komputasi karena tidak melibatkan induksi algoritma dalam prosesnya.

2.1.8 Pengujian Akurasi dan Validasi Metode Klasifikasi Data Mining

Untuk menguji model, pada penelitian ini, digunakan metode *Confusion Matrix*, dan kurva *ROC (Receiver Operating Characteristic)*.

1. *Confusion matrix*

Metode ini menggunakan tabel matriks seperti pada Tabel 2.1, jika data set hanya terdiri dari dua kelas, kelas yang satu dianggap sebagai positif dan yang lainnya *negative*

Tabel 2.1 Model *Confusion Matrix*

<i>Correct classification</i>	<i>Clasified as</i>	
	+	-
+	<i>True Positives</i>	<i>False Negatives</i>
-	<i>False Positives</i>	<i>True Negatives</i>

True positives adalah jumlah *record* positif yang diklasifikasikan sebagai positif, *false positives* adalah jumlah *record* negatif yang diklasifikasikan sebagai positif, *false negatives* adalah jumlah *record* positif yang diklasifikasikan sebagai negatif, *true negatives* adalah jumlah *record* negatif yang diklasifikasikan sebagai negative, kemudian masukkan data uji. Setelah data uji dimasukkan ke dalam *confusion matrix*, hitung nilai-nilai yang telah dimasukkan tersebut untuk dihitung jumlah *sensitivity (recall)*, *specificity*, *precision* dan *accuracy*. *Sensitivity* digunakan untuk membandingkan jumlah TP terhadap jumlah *record* yang positif sedangkan *specificity* adalah perbandingan jumlah TN terhadap jumlah *record* yang negatif. Untuk menghitung digunakan persamaan di bawah ini:

$$\text{Sensitivity} = \frac{TP}{P} \quad (2.7)$$

$$\text{Specificity} = \frac{TN}{N} \quad (2.8)$$

$$\text{Precision} = \frac{TP}{TP+FP} \quad (2.9)$$

$$\text{Accuracy} = \text{sensitivity} \frac{P}{P+N} + \text{specificity} \frac{N}{P+N}$$

Keterangan:

TP = jumlah *true positives*

TN = jumlah *true negatives*

P = jumlah *record* positif

N = jumlah *tupel* negatif

FP = jumlah *false positives*

2. Kurva ROC (*Receiver Operating Characteristic*)

Kurva ROC atau *Receiver Operating Characteristic* menunjukkan akurasi dan membandingkan klasifikasi secara visual. ROC mengekspresikan *confusion matrix*. ROC adalah grafik dua dimensi dengan *false positives* sebagai garis horisontal dan *true positives* sebagai garis vertikal[9][10].

2.1.9 Pengertian Nasabah

Nasabah adalah pihak yang menggunakan jasa bank atau lembaga keuangan bukan bank. Penghimpunan dana dan pemberian kredit merupakan pelayanan jasa perbankan yang utama dari semua kegiatan lembaga keuangan. Berdasarkan Pasal 1 angka (16) UU Perbankan diintroduksikan rumusan nasabah yaitu nasabah adalah pihak yang menggunakan jasa bank.

Rumusan tersebut kemudian diperinci pada butir berikutnya, yaitu sebagai berikut:

- a. Nasabah Penyimpan adalah nasabah yang menempatkan dananya di bank dalam bentuk simpanan berdasarkan perjanjian bank dengan nasabah yang bersangkutan.
- b. Nasabah Debitur adalah nasabah yang memperoleh fasilitas kredit atau pembiayaan berdasarkan prinsip syariah atau yang dipersamakan dengan itu berdasarkan perjanjian bank dengan nasabah yang bersangkutan.

2.2 Tinjauan Studi

Berikut adalah ringkasan dari beberapa penelitian sebelumnya yang terkait dengan klasifikasi *data mining*.

Tabel 2.2 Ringkasan Tinjauan Studi

No	Judul	Metode	Tujuan	Variabel/atribut	Hasil
1.	“Algoritma Naïve Bayes, Decision Tree, dan SVM untuk Klasifikasi Persetujuan Pembiayaan Nasabah Koperasi Syariah”	Pengujian menggunakan metode Algoritma Decision tree, Naïve Bayes dan SVM	menerapkan data mining untuk dapat membantu melakukan klasifikasi persetujuan pembiayaan nasabah	Status, tanggungan, usia, pendidikan, pekerjaan, pendapatan, status rumah, jaminan, jumlah pinjaman, lama pinjaman	Algoritma yang telah dilakukan diuji menggunakan tools Rapid Miner. Diperoleh hasil akurasi sebesar tertinggi 94% pada algoritma SVM.
2.	“Perbandingan Metode K-Nearest Neighbor, Naïve Bayes dan Decision Tree untuk Prediksi Kelayakan Pemberian Kredit”[1]	Algoritma K-NN, Naïve Bayes dan Decision Tree	Data ini mengacu dan dapat digunakan dalam memprediksi kelayakan pemberian kredit untuk calon debitur	Jenis kelamin, umur, Pendidikan terakhir, status pernikahan, pekerjaan, penghasilan, jumlah pinjaman, jenis pinjaman, jangka waktu	Decision Tree (J-48) dengan akurasi sebesar 92,21%, algoritma K-Nearest Neighbor memiliki tingkat akurasi sebesar 81,82% dan Naïve Bayes memiliki tingkat akurasi sebesar 81,83%.

Tabel 2.2 (lanjutan)

No	Judul	Metode	Tujuan	Variabel/atribut	Hasil
3.	“Aplikasi metode K-Nearest Neighbor dan analisis Diskriminan untuk analisis resiko kredit pada koperasi simpan pinjam di kopinkra sumber rejeki”[12]	Algoritma K-NN dan Analisis Diskriminan Metode pengujian : <i>K-Folds Cross Validation</i>	Dapat memprediksi potensi kredit macet dan menentukan kelancaran simpan pinjam di kopinkra sumber ejeke	penghasilan, pinjaman, jangka waktu dan jaminan	Ketepatan prediksi resiko dengan metode KNN 84,33% pada nilai K adalah 7 dan Metode Diskriminan sebesar 76,5%.
4.	“analisis klasifikasi pada nasabah kredit koperasi menggunakan decision tree c4.5 dan naïve bayes”[13]	Metode pengujian menggunakan decision tree c4.5 dan naïve bayes	untuk membentuk model decision tree C4.5 dan naïve bayes untuk klasifikasi nasabah kredit berdasarkan nilai kolektibilitasnya	Status nasabah, Umur, Desa, Kecamatan, Status Marital, Nilai Pinjaman, Jumlah angsuran	Decision tree C4.5 memiliki kecenderungan tingkat akurasi yang lebih tinggi yaitu sebesar 71,91%, 68,03%, dan 66,84%, sedangkan naïve bayes memiliki akurasi sebesar 67,01%, 64,66%, dan 65,82%.
5.	“Komparasi Algoritma Klasifikasi Data Mining untuk Evaluasi Pemberian Kredit”[14]	Algoritma C4.5 dan Naïve Bayes	Analisa kredit nasabah sebelum diabil keputusan pemberian kredit	Nama, Jenis Kelamin, Umur, Pinjaman, Tenor, Angsuran, Plafon Teoritis, Tunggalan Pokok, Tunggalan bunga	Tingkat akurasi yang lebih baik adalah menggunakan algoritma C4.5 yaitu nilai akurasinya 88,90% sedangkan dengan Naïve Bayes yaitu 80,00%

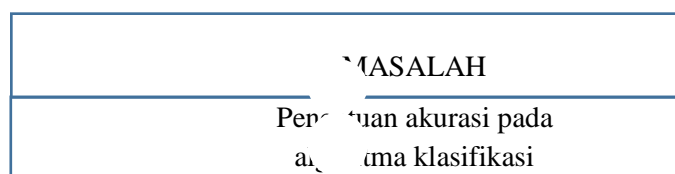
Pada tabel 2.2. Ringkasan tinjauan studi, hampir semua penelitian menggunakan metode klasifikasi *data mining*. Dalam hal ini penulis mencoba membandingkan dua metode klasifikasi *data mining* yaitu *Naïve Bayes* dan *Decision Tree* untuk melihat akurasi yang tepat dari penghitungan ke dua metode tersebut penulis menggunakan pengujian *k-folds cross validation* dengan $k=10$ agar dapat mengetahui tingkat akurasinya terhadap prediksi pemberian kredit pada finance. Nantinya hasil dari pengujian tersebut akan dipilih salah satu metode dan kemudian akan dimodifikasi modelnya sehingga diharapkan dapat menaikkan nilai pengujian baik akurasi, precision, recall dan lain-lain. Hasil pemilihan metode dan modifikasi model akan diterapkan dalam pengembangan prototype prediksi pemberian kredit untuk calon nasabah.

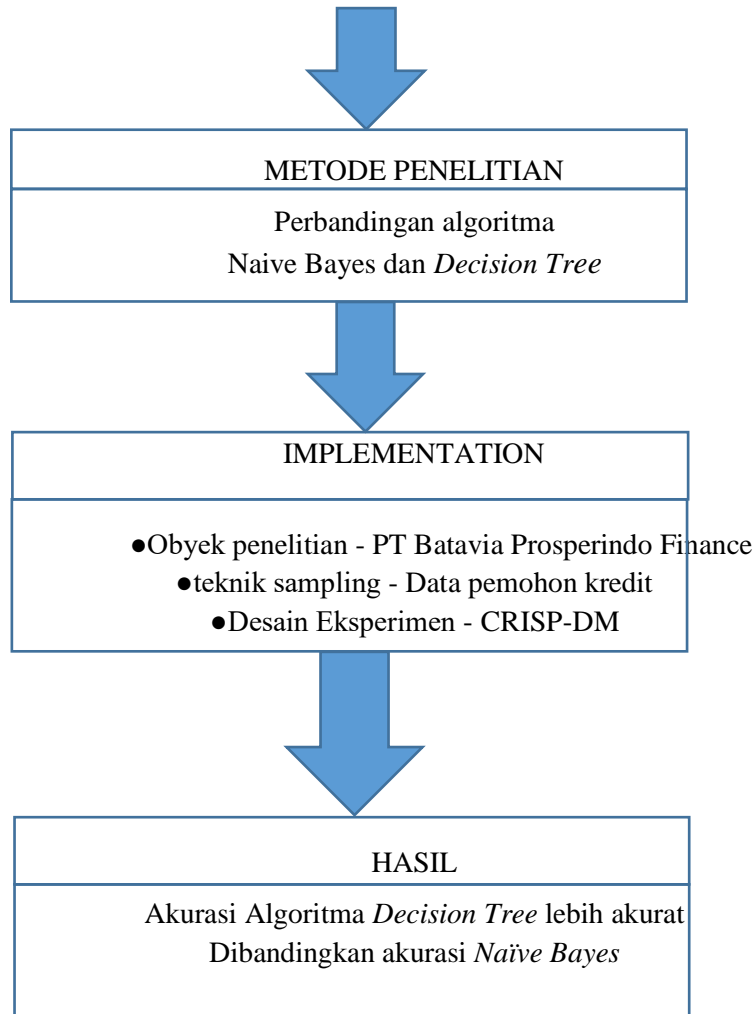
2.3 Tinjauan Obyek Penelitian

PT Batavia Prosperindo Finance Tbk adalah suatu perusahaan *finance* yang bergerak di bidang pembiayaan untuk kendaraan roda empat, terutama kendaraan bekas jenis Penumpang/Pribadi (Passenger) dan Niaga (Commercial) yang memiliki banyak penghargaan di dunia usaha dan bisnis perbankan, yang memiliki banyak anak cabang yang tersebar di seluruh Indonesia dan dikenal sebagai salah satu pembiayaan yang memiliki pelayanan terbaik kepada para debitur. Dengan penawaran dari suku bunga rendah dan proses pengajuan kredit tidak sulit menjadikan perusahaan ini menjadi tolak ukur bagi seluruh finance di Indonesia. Adapun visi dan misi dari perusahaan ini adalah pelopor dalam memberikan solusi finansial yang inovatif dan menjadikan hidup lebih bernilai untuk kebutuhan kendaraan dan permodalan dalam dunia usaha serta bisnis.

2.4 Kerangka Konsep / Pola Pikir Pemecahan

Penelitian ini dilakukan dengan judul “**Perbandingan Algoritma Naive Bayes Dan Decision Tree Dalam Memprediksi Approval Pemberian Kredit Studi Kasus : PT Batavia Prosperindo Finance Tbk. Cabang Pringsewu**”. Maka dapat dibangun kerangka konsep penelitian tahapan sebagai berikut:





Gambar 2.3 Kerangka Pemikiran

Penjelasan untuk kerangka Pemikiran pada gambar 2.3 yaitu:

a. Masalah / *Problems*

Problems atau permasalahan yang ada pada penelitian ini adalah penentuan akurasi pada algoritma Klasifikasi.

b. Metode Penelitian / *Approach*

Approach atau metode penelitian yang dipakai pada kasus ini adalah dengan perbandingan *Naive Bayes* dan *Decision Tree*.

c. *Implementation*

Penelitian ini menggunakan data Nasabah dalam pengajuan kredit pada PT Batavia Prosperindo Finance , dan urutan desain eksperimen menggunakan CRISP-DM (*Cross Standard Industry Process for Data Mining*)

d. Hasil / *Result*

Result atau hasil penelitian yang didapat adalah akurasi dari Algoritma *Decision Tree* lebih tepat akurasinya dibanding dengan metode *Naïve Bayes* untuk penentuan Prediksi kelayakan pemberian Kredit

2.5 Hipotesis

Dari ke dua klasifikasi data mining yang diuji yaitu *Naïve Bayes* dan *Decision Tree* merupakan metode yang dapat digunakan dalam menentukan prediksi pemberian kredit kepada nasabah.

1. Dengan menggunakan metode klasifikasi *Naïve Bayes* dan *Decision tree* diduga dapat digunakan untuk prediksi penentuan nasabah dalam melakukan pengajuan kembali pinjaman dengan evaluasi data nasabah sebelumnya.
2. Penggunaan kedua algoritma yaitu *Naïve Bayes* dan *Decision tree* salah satunya akan menghasilkan nilai yang lebih tinggi, nilai yang lebih tinggi yang nantinya akan dijadikan sebagai acuan dalam menentukan evaluasi nasabah melakukan pinjaman kembali.

Nilai akurasi dari *Decision Tree* diduga memiliki nilai yang lebih tinggi dibandingkan *Naïve Bayes*