

BAB II

LANDASAN TEORI DAN KERANGKA KONSEP

1. Tinjauan Pustaka

1. Pengertian Kredit

Istilah kredit berasal dari bahasa Latin *credo* yang berarti „saya percaya“, yang merupakan kombinasi dari bahasa Sansekerta *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. Menurut Undang- Undang No.10 tahun 1998 tentang perubahan Undang-Undang no.7 tahun 1992 tentang pembiayaan kendaraan yang dipaparkan dalam bab 1 ketentuan umum pasal 1 dijelaskan bahwa 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 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, diukur dengan posisi usaha atau jumlah penghasilan.

4. *Collateral*

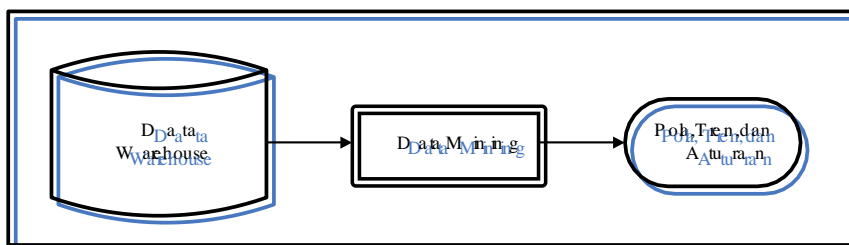
Jaminan yang dimiliki calon peminjam. Penilaian ini untuk lebih meyakinkan bahwa jika suatu resiko kegagalan pembayaran terjadi, maka jaminan dapat dipakai sebagai pengganti kewajiban.

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. 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.



Gambar 2. 1 Prinsip Data Mining

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 *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 pengelompokkan 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.

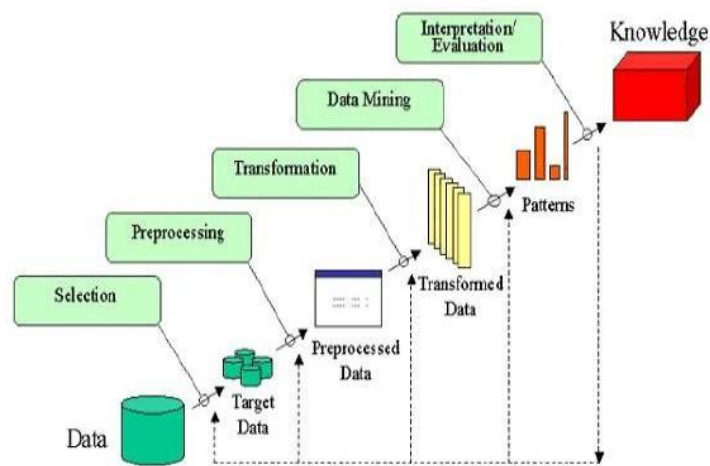
2. Meningkatkan kompetisi:

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 2.2



Gambar 2. 2 Proses Data Mining

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

(Mandala & Putri, 2018):

1. *Data Selection*

Pemilihan data dari sekumpulan data operasional perlu dilakukan tahap penggalian 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.

1. **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 (Wahyuningsih & Utari, 2018). 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. K-Fold Cross Validation

Dalam penelitian Menarianti, (2015) 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_i 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:

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

3. 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 (Menarianti, 2015).

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

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

4. 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 bertenangan 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.

5. 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*

Table 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:

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

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

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

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

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 vertical (Fibrianda & Bhawiyuga, 2018)(Ashari, 2013).

1. 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:

1. Nasabah Penyimpan adalah nasabah yang menempatkan dananya di bank dalam bentuk simpanan berdasarkan perjanjian bank dengan nasabah yang bersangkutan.
2. 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. Probability sampling

Pemilihan sampel populasi berkaitan dengan pengajuan kredit dari nasabah. Penarikan sampel merupakan proses pilihan sejumlah elemen dari populasi. *Probability sampling* adalah teknik pengambilan sampel yang memberikan peluang yang sama bagi setiap unsur atau anggota populasi untuk dipilih menjadi sampel (Wahyuningsih & Utari, 2018).

$$n = \frac{N}{1 + Ne^2}$$

Keterangan:

n: ukuran sampel N:

ukuran populasi

e: kelonggaran ketidak telitian atau derajat toleransi

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	Masalah	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	Algoritma yang telah dilakukan diuji menggunakan tools Rapid Miner. Diperoleh hasil akurasi sebesar tertinggi 94% pada algoritma SVM.
2.	“Implementasi algoritma Naïve Bayes dalam penentuan pemberian kredit	Algoritma Naïve Bayes	Masalah ini sebenarnya dapat diatasi dengan cara mengidentifikasi dan memprediksi nasabah dengan baik sebelum memberikan pinjaman dengan cara memperahitkan data historis pinjaman	Untuk memprediksi dan mengklasifikasi nasabah mana saja yang bermasalah dan tidak bermasalah, dan diharapkan mampu meningkatkan akurasi dalam menganalisa kelayakan kredit.
3.	“Aplikasi metode K-Nearest Neighbor dan analisis Diskriminan untuk analisis resiko kredit pada koperasi simpan pinjam di kopinkra sumber rejeki”(Muhammad Sholeh, 2014)	Algoritma K-NN dan Analisis Diskriminan Metode pengujian : <i>K-Folds Cross Validation</i>	memprediksi potensi kredit macet dan menentukan kelancaran simpan pinjam di kopinkra sumber rejeki	Ketepatan prediksi resiko dengan metode KNN 84,33% pada nilai K adalah 7 dan Metode Diskriminan sebesar 76,5%.

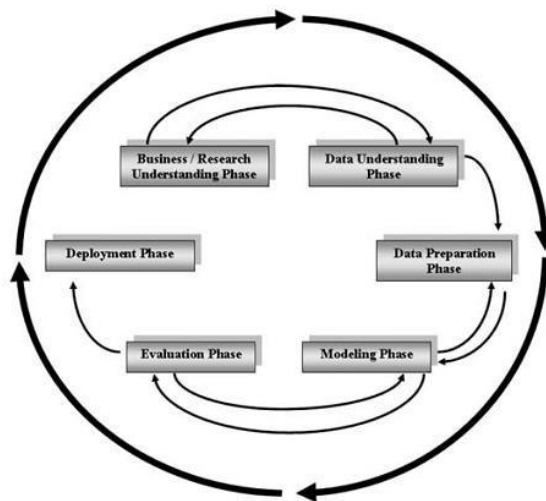
4.	“Analisis klasifikasi pada nasabah kredit koperasi menggunakan decision tree c4.5 dan naïve bayes”(Kholifah & Insani, 2016)	Metode pengujian menggunakan decision tree c4.5 dan naïve bayes	membentuk model decision tree C4.5 dan naïve bayes untuk klasifikasi nasabah kredit berdasarkan nilai kolektibilitasnya	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”(Masripah, 2016)	Algoritma C4.5 dan Naïve Bayes	Analisa kredit nasabah sebelum diabil keputusan pemberian kredit	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* 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 prediksi pemberian kredit untuk calon nasabah.

3. Tinjauan Obyek Penelitian

PT Batavia Prosperindo Finance Tbk adalah suatu perusahaan finance yang bergerak di bidang pembiayaan untuk kendaraan roda empat, perusahaan ini sudah berdiri sejak 15 februari 2018 yang beralamat di jalan A.Yani No.5c, Sidoharjo, Pringsewu Timur, Kecamatan Prinsewu,Kabupaten Pringsewu, Lampung. PT ini lebih mengolah pembiayaan 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.

4. Proses CRISP- DM



Gambar 2.4 Proses CRISP-DM

Dalam gambar 3.1 dijelaskan bahwa metode penelitian eksperimen, digunakan model proses CRISP-DM (*Cross Standard Industry Process for Data Mining*) yang terdiri dari 6 tahapan(Larose & Larose, 2018):

1. *Business Understanding*

Tahap pertama dalam CRISP-DM dan termasuk bagian yang cukup vital. Pada tahap ini membutuhkan pengetahuan dari objek bisnis, bagaimana membangun atau mendapatkan data, dan bagaimana untuk mencocokkan tujuan pemodelan untuk tujuan bisnis sehingga model terbaik dapat dibangun. Kegiatan yang dilakukan antara lain: menentukan tujuan dan persyaratan dengan jelas secara keseluruhan, menerjemahkan tujuan tersebut serta menentukan pembatasan dalam perumusan masalah data mining, dan selanjutnya mempersiapkan strategi awal untuk mencapai tujuan tersebut

2. *Data Understanding*

Secara garis besar untuk memeriksa data, sehingga dapat mengidentifikasi masalah dalam data. Tahap ini memberikan fondasi analitik untuk sebuah penelitian dengan membuat ringkasan (*summary*) dan mengidentifikasi potensi masalah dalam data. Tahap ini juga harus dilakukan secara cermat dan tidak terburu-buru, seperti pada visualisasi data, yang terkadang *insight*-nya sangat sulit didapat jika dihubungkan dengan *summary data* nya. Jika ada masalah pada tahap ini yang belum terjawab, maka akan mengganggu pada tahap *modeling*.

3. *Data Preparation*

Secara garis besar untuk memperbaiki masalah dalam data, kemudian membuat *variabel derived*. Tahap ini jelas membutuhkan pemikiran yang cukup matang dan usaha yang cukup tinggi untuk memastikan data tepat untuk algoritma yang digunakan.

Bukan berarti saat *Data Preparation* pertama kali dimana masalah-masalah pada data sudah diselesaikan, data sudah dapat digunakan hingga tahap terakhir. Tahap ini merupakan tahap yang sering ditinjau kembali saat menemukan masalah pada saat pembangunan model. Sehingga dilakukan iterasi sampai menemukan hal yang cocok dengan data. Tahap *sampling* dapat dilakukan disini dan data secara umum dibagi menjadi dua, *data training* dan *data testing*.

Kegiatan yang dilakukan antara lain: memilih kasus dan parameter yang akan dianalisis (*Select Data*), melakukan transformasi terhadap parameter tertentu (*Transformation*), dan melakukan pembersihan data agar data siap untuk tahap *modeling (Cleaning)*.

4. *Modelling*

Pada tahap ini dilakukan metode statistika dan *Machine Learning* untuk penentuan terhadap teknik *data mining*, alat bantu *data mining*, dan algoritma *data mining* yang akan diterapkan. Lalu selanjutnya adalah melakukan penerapan teknik dan algoritma data mining tersebut kepada data dengan bantuan alat bantu. Jika diperlukan penyesuaian data terhadap teknik data mining tertentu, dapat kembali ke tahap *data preparation*.

Beberapa modeling yang biasa dilakukan adalah *classification, scoring, ranking, clustering, finding relation*, dan *characterization*.

5. *Evaluasi*

Melakukan interpretasi terhadap hasil dari data mining yang dihasilkan dalam proses pemodelan pada tahap sebelumnya. Evaluasi dilakukan terhadap model yang diterapkan pada tahap sebelumnya dengan tujuan agar model yang ditentukan dapat sesuai dengan tujuan yang ingin dicapai dalam tahap pertama.

6. *Deployment*

Tahap *deployment* atau rencana penggunaan model adalah tahap yang paling dihargai dari proses CRISP-DM. Perencanaan untuk *Deployment* dimulai selama *Business Understanding* dan harus menggabungkan tidak hanya bagaimana untuk menghasilkan nilai model, tetapi juga bagaimana mengkonversi skor keputusan, dan bagaimana untuk menggabungkan keputusan dalam sistem operasional.

Pada akhirnya, rencana sistem *Deployment* mengakui bahwa tidak ada model yang statis. Model tersebut dibangun dari data yang diwakili data pada waktu tertentu, sehingga perubahan waktu dapat

menyebabkan berubahnya karakteristik data. Modelpun harus dipantau dan mungkin diganti dengan model yang sudah diperbaiki

Penelitian ini menggunakan metode studi pustaka sebagai tahap awal dengan mempelajari landasan teori mengenai *data mining* menggunakan algoritma klasifikasi yaitu *Naïve Bayes* pada beberapa literatur dan referensi lainnya. Referensi mencakup data-data dari jurnal, *internet*, *e-book*, dan dokumen lainnya yang berkaitan dengan penelitian ini.