

**SISTEM INFORMASI PENGAMBILAN KEPUTUSAN PENGAJUAN
KREDIT DENGAN ALGORITMA *K-NEAREST NEIGHBOUR*
(STUDI KASUS: KOPERASI SIMPAN PINJAM)**

**Tesis
untuk memenuhi sebagian persyaratan
mencapai derajat Sarjana S-2 Program Studi
Magister Sistem Informasi**



**Emerensye S. Y. Pandie
(J4F009017)**

**PROGRAM PASCASRAJANA
UNIVERSITAS DIPONEGORO
SEMARANG
2012**

TESIS

**SISTEM INFORMASI PENGAMBILAN KEPUTUSAN PENGAJUAN
KREDIT DENGAN ALGORITMA *K-NEAREST NEIGHBOUR*
(STUDI KASUS: KOPERASI SIMPAN PINJAM)**

**Oleh :
Emerensye S. Y. Pandie
J4F009017**

Telah diujikan dan dinyatakan lulus ujian tesis pada tanggal 25 Juli 2012 oleh tim penguji Program Pascasarjana Magister Sistem Informasi Universitas Diponegoro

Semarang, 25 Juli 2012
Mengetahui

Pembimbing I

Penguji I

Prof. Drs. Mustafid, M.Eng. Ph.D
NIP. 195505281980031002

Drs. Bayu Surarso, M.ASc, Ph.D
NIP. 196311051988031001

Pembimbing II

Penguji II

Beta Noranita, S.Si, M.Kom
NIP. 197308291998022 001

Dr. Vincencius Gunawan S.K., M.Si
NIP. 197105221997021001

Penguji III

Drs. Eko Adi Sarwoko, M.Kom
NIP. 196511071992031003

Mengetahui
Ketua Program Studi
Magister Sistem Informasi

Drs. Bayu Surarso, M.ASc, Ph.D
NIP. 196311051988031001

PERNYATAAN

Dengan ini saya menyatakan bahwa dalam tesis ini tidak terdapat karya yang pernah diajukan untuk memperoleh gelar akademik di suatu perguruan tinggi, dan sepanjang pengetahuan saya juga tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang lain, kecuali secara tertulis diacu dalam naskah ini dan disebutkan dalam daftar pustaka.

Semarang, 25 Juli 2012

Emerensye S. Y. Pandie

KATA PENGANTAR

Puji Tuhan atas berkat dan anugerah-Nya kepada penulis, sehingga penyusunan tesis dapat terselesaikan dalam rangka memenuhi persyaratan memperoleh gelar Magister Sistem Informasi pada Program Pasca Sarjana Universitas Diponegoro Semarang. Tesis dengan judul : “SISTEM INFORMASI PENGAMBILAN KEPUTUSAN PENGAJUAN KREDIT DENGAN ALGORITMA *K-NEAREST NEIGHBOUR* (STUDI KASUS: KOPERASI SIMPAN PINJAM)” ini berhasil diselesaikan tidak lepas dari adanya bantuan dan dukungan dari berbagai pihak. Pada kesempatan ini perkenankan penulis menyampaikan ucapan terima kasih yang setulus-tulusnya kepada :

1. Drs. Bayu Surarso, M.ASc, Ph.D, selaku Ketua Program Studi Magister Sistem Informasi dan Penguji I.
2. Prof. Drs. Mustafid, M.Eng. Ph.D, selaku Pembimbing I.
3. Beta Noranita, S.Si, M.Kom, selaku Pembimbing II.
4. Dr. Vincencius Gunawan S.K., M.Si, selaku Penguji II.
5. Drs. Eko Adi Sarwoko, M.Kom, selaku Penguji III.

Semoga tesis ini dapat bermanfaat untuk mengembangkan ilmu pengetahuan terutama dalam bidang sistem informasi.

Semarang, 25 Juli 2012

DAFTAR ISI

	Halaman
HALAMAN JUDUL	i
HALAMAN PENGESAHAN	ii
HALAMAN PERNYATAAN	iii
KATA PENGANTAR	iv
DAFTAR ISI	v
DAFTAR GAMBAR	vii
DAFTAR TABEL	ix
ABSTRAK	x
ABSTRACT	xi
BAB I. PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Perumusan Masalah	2
1.3 Batasan Masalah	2
1.4 Keaslian Penelitian	3
1.5 Tujuan Penelitian	4
1.6 Manfaat Penelitian	4
BAB II. TINJAUAN PUSTAKA	5
2.1 Tinjauan Pustaka	5
2.2 Landasan Teori	5
2.2.1 Konsep Sistem Pengambilan Keputusan.....	5
2.2.2 <i>Data mining</i>	6
2.2.3 Klasifikasi	9
2.2.4 Standarisasi Data	11
2.2.5 Algoritma <i>k-Nearest Neighbour</i> (k-NN)	13
2.2.6 <i>n-fold Cross Validation</i>	16
2.2.7 Kategori Kredit	19
2.2.8 Analisa Kredit " <i>The 5 C of Credit Analysis</i> "	19
2.2.8 Metode SDLC	20
BAB III. CARA PENELITIAN	22
5.1 Bahan Penelitian	22
5.2 Alat Penelitian	22
5.3 Jalan Penelitian	22
3.3.1 Tahapan <i>Data mining</i>	24
3.3.2 Kerangka Sistem	35
3.3.3 Pemodelan Proses	36
3.3.4 Pemodelan Data	40
3.3.5 Rancangan Tabel	41
3.3.6 Pemodelan Perangkat Lunak	42
3.3.7 Desain Antar Muka	47

BAB IV. HASIL PENELITIAN DAN PEMBAHASAN	52
4.1 Hasil Penelitian	52
4.2 Pembahasan	58
BAB V. KESIMPULAN DAN SARAN	66
5.1. Kesimpulan	66
5.2. Saran	66

DAFTAR PUSTAKA

DAFTAR GAMBAR

	Halaman
Gambar 2.1 Tahapan <i>data mining</i> CRISP-DM	9
Gambar 2.2 Pemodelan Prediktif	10
Gambar 2.3 Tahapan proses klasifikasi k-NN	11
Gambar 2.4 Prediksi data baru terhadap nilai k dalam k-NN	13
Gambar 2.5 Hubungan <i>data mining</i> dengan k-NN	15
Gambar 2.6 Contoh iterasi data dengan <i>10-fold cross validation</i>	17
Gambar 2.7 Tahapan metode penelitian SDLC	21
Gambar 3.1 Skema kategori kredit BI dan penerapannya pada koperasi	29
Gambar 3.2 Tabel data pola	31
Gambar 3.3 Grafik tingkat eror terhadap nilai k	34
Gambar 3.4 Kerangka sistem informasi penunjang keputusan	35
Gambar 3.5 Diagram konteks sistem informasi persetujuan kredit	36
Gambar 3.6 DFD Level 1 sistem informasi persetujuan kredit	37
Gambar 3.7 DFD Level 2 proses algoritma k-NN	39
Gambar 3.8 ERD Model	40
Gambar 3.9 <i>Flowchart</i> standarisasi data	43
Gambar 3.10 <i>Flowchart</i> proses kNN	45
Gambar 3.11 Menu utama	47
Gambar 3.12 Form anggota	48
Gambar 3.13 Form pola angsuran	48
Gambar 3.14 Form pinjaman	49
Gambar 3.15 Form proses penilaian	50
Gambar 3.16 Form verifikasi persetujuan	50
Gambar 3.17 Form input laporan persetujuan	51
Gambar 3.18 Form laporan persetujuan akhir	51
Gambar 4.1 Data pola kredit	52
Gambar 4.2 Data pengajuan kredit	53
Gambar 4.3 Atribut plafon	53
Gambar 4.4 Standarisasi atribut plafon	54
Gambar 4.5 Standarisasi data pola kredit	54
Gambar 4.6 Standarisasi data pengajuan kredit	54
Gambar 4.7 <i>Record</i> awal data pola kredit	55
Gambar 4.8 <i>Record</i> data pengajuan kredit	55
Gambar 4.9 Hasil perhitungan kNN terhadap data pengajuan kredit	56
Gambar 4.10 Hasil perhitungan kNN diurutkan <i>ascending</i>	56
Gambar 4.11 Hasil prediksi $k=7$	57
Gambar 4.12 Data pola kredit	58
Gambar 4.13 Input data anggota	59
Gambar 4.14 Input data peminjaman	59
Gambar 4.15 Daftar anggota koperasi	60
Gambar 4.16 Pilihan menu tampilan per- <i>record</i>	60
Gambar 4.17 Proses penilaian pinjaman	61

Gambar 4.18 Tampilan menu peringatan	62
Gambar 4.19 Tampilan menu saran/prediksi kategori kredit	62
Gambar 4.20 Hasil penghitungan jarak algoritma kNN dengan $k=7$	63
Gambar 4.21 Verifikasi persetujuan kredit.....	64
Gambar 4.22 Tampilan inputan laporan per-tanggal.....	65
Gambar 4.23 <i>Print Preview</i> laporan verifikasi.....	65

DAFTAR TABEL

	Halaman
Tabel 3.1 Anggota	24
Tabel 3.2 Aplikasi	25
Tabel 3.3 Angsuran	25
Tabel 3.4 Pinjam	26
Tabel 3.5 Jaminan	26
Tabel 3.6 Plafon.	27
Tabel 3.7 Kategori kredit BI	28
Tabel 3.8 Kategori kredit	30
Tabel 3.9 Konversi nilai nominal ke diskrit	32
Tabel 3.10 Hasil 10- <i>fold cross validation</i>	33
Tabel 3.11 Tabel rumusan	41
Tabel 3.12 Tabel anggota	41
Tabel 3.13 Tabel pinjam	42

ABSTRAK

Berdasarkan data dari dinas koperasi dan UMKM propinsi jawa tengah, koperasi menjadi penyangga kekuatan ekonomi sebesar 59% dengan total omset mencapai 25 trilyun pertahun. 20% diantara jumlah koperasi tersebut mengalami kegagalan usaha, kegagalan usaha paling besar dialami oleh koperasi yang bergerak dalam usaha koperasi jasa keuangan. Salah satu faktor kegagalan usaha ini dipengaruhi oleh penilaian kemampuan debitur yang kurang akurat, sehingga mengakibatkan kesalahan dalam memberikan kredit yang berujung pada kemacetan. penelitian ini menggunakan teknik data mining untuk menilai kemampuan nasabah berdasarkan data-data masa lalu. Dengan menggunakan tahapan data mining dihasilkan sejumlah 3.018 data nasabah yang dikategorikan dalam 3 kategori kredit, macet, tersendat dan lancar. Hasil filtrasi data tersebut diuji tingkat errornya menggunakan teknik cross validation dengan 20 fold dan 10 nilai k . Hasil pengujian menunjukkan persentase tingkat error data pada angka kurang dari 3.7% dan mencapai kestabilan data pada nilai $k=3$ sampai $k=11$. Berdasarkan kesimpulan tersebut dibuat aplikasi pendukung keputusan untuk menentukan tingkat kemampuan pengembalian nasabah dengan menggunakan algoritma k-nearest neighbor dengan nilai $k=7$.

Kata kunci : kredit, koperasi, data mining, k-nearest neighbour, cross validation, fold, k

ABSTRACT

Based on data from the department of cooperatives and SMEs Central Java province, the economic strength of cooperatives into a buffer of 59% with total turnover reaching 25 trillion per year. 20% between the number of cooperative efforts fail, the biggest business failure experienced by cooperatives engaged in financial services cooperative effort. One of the factors of business failure is influenced by the judgment debtor's ability to be less accurate, resulting in errors in providing the credit that leads to congestion. This study uses data mining techniques to assess the ability of customers based on past data. By using phase generated a number of 3018 data mining of customer data that fall within three categories of credit, stuck, stalled and smoothly. The results of filtration data was tested using a technique error rate by 20 fold cross validation and 10 values of k. Test results show the percentage error rate data on the number of less than 3.7% and achieve stability data on the value of $k = 3$ to $k = 11$. Based on these conclusions support the application made the decision to determine the level of customer returns by using k-nearest neighbor algorithm with $k = 7$.

Kata kunci : credit, cooperatives, data mining, k-nearest neighbour, cross validation, fold, k

DAFTAR PUSTAKA

- Chen, M.C., Huang, S.H., 2003. Credit scoring and rejected instances reassigning through evolutionary computation techniques. *Jurnal Expert System with Application* 24 (4), 433-441
- Han, J.K., 2001. *Data mining: Concept and Technique*. San Fransisco: Morgan Kaufmann Publisher
- Kohavi, R., 1995. A Study of Cross Validation and Bootstrap for Accuracy Estimation and Model Selection, *Lecture Note in Computer Science* 6440, 114-124
- Kusrini, Luthfi, E.T., 2009. *Algoritma Data mining*. Yogyakarta: Andi Offset.
- Masngudi, H., 1990. *Penelitian Tentang Sejarah Perkembangan Koperasi Indonesia*, Badan Penelitian Perkembangan Koperasi, Jakarta.
- Moertini, S.V., 2003. Towards the use of C4.5 Algorithm for classifying Banking Dataset. *Jurnal Integral* 8 (2), 105-116
- Mohd, S.S., Rayner, A., 2010. *Advanced Data mining and Applications 6th International Conference, ADMA 2010, Chongqing, China, November 19-21, 2010, Proceedings, Part I*
- Oktrivianto, R., 2008. *Aplikasi Data mining untuk Memprediksi Kelas Resiko Pemberian Kredit menggunakan Support Vector Machine (SVM)*. Tesis. Surabaya: Institut Teknologi Surabaya.
- Pramudiono, I., 2003. *Pengantar Data mining: Menambang Permata Pengetahuan di Gunung Data*. Website:<http://www.ilmukomputer.com>. Diakses tanggal 10 Januari 2011.
- Sucahyo, Y.G., 2003. *Data mining : Menggali Informasi yang Terpendam*. Website: <http://www.ilmukomputer.com>. Diakses tanggal 12 Januari 2011.
- Sujarwanto., 2011, 5809 Koperasi di Jawa Tengah Tidak Aktif. *Harian Suara Merdeka* tanggal 30 Juni 2011
- Sunaryanto, D., 2009. *Penerapan K-Nearest Neighbour untuk Prediksi Resiko Kredit di PT. Telkom Kandatel Surabaya Timur*. Tesis. Surabaya: Sekolah Tinggi Ilmu Komputer (STIKOM).

BAB I

PENDAHULUAN

1.1. Latar Belakang

Sebagai salah satu usaha bank yakni memberikan kredit, bank akan selalu memperhatikan resiko kredit yang terjadi dalam hal pemberian kredit, seperti nasabah yang tidak mampu membayar pinjaman yang telah diterima. Untuk mengurangi resiko yang terjadi pihak bank menganggap perlu mengambil suatu kebijakan yaitu dengan memperkuat pengawasan bank terhadap nasabah untuk mengatasi resiko tersebut (Pardede, 2004).

Didalam kegiatan perkreditan sering terjadi masalah kredit macet yang disebabkan oleh gagalnya pengembalian sebagian pinjaman yang diberikan kepada para peminjam. Masalah ini dapat diatasi, salah satunya dengan mengidentifikasi dan memprediksi nasabah dengan baik sebelum memberikan pinjaman dengan cara memperhatikan data historis pinjaman. Analisa kredit dapat menggunakan salah satu jenis algoritma *data mining* yaitu *Support Vector Machine* (Oktrivianto, 2008).

Teknologi *data mining* memiliki peranan penting dalam dunia bisnis yang semakin kompetitif. *Data mining* mampu memanfaatkan aset penting perusahaan yaitu data bisnis yang jumlahnya sangat besar sehingga dapat menghasilkan informasi yang belum pernah terpelajari sebelumnya. Bentuk penerapan *data mining* antara lain dengan membuat aplikasi yang memanfaatkan teknologi *data mining* menggunakan algoritma C4.5 untuk klasifikasi profil kredit nasabah perbankan (Kusuma, 2004).

Dalam mengantisipasi kredit macet karena meningkatnya jumlah pengajuan kredit pada PT. Telkom Kandatel Surabaya Timur maka dilakukan analisa dan pola kredit dari angsuran pinjaman menggunakan metode *k-Nearest Neighbour* (k-NN). Output dari aplikasi digunakan untuk melakukan *reminding call*, *rescheduling* dan *reconditioning* kepada nasabah (Sunaryanto, 2009).

BAB II

TINJAUAN PUSTAKA

2.1. Tinjauan Pustaka

Mengingat pentingnya analisis resiko dalam dunia perkreditan, para ahli telah melakukan penelitian dengan berbagai metode yang berbeda untuk menganalisa tingkat keberhasilan pemberian kredit. Sunaryanto dalam penelitiannya pada tahun 2009, menyimpulkan untuk memperkecil resiko kemacetan kredit, perlu dilakukan tindakan-tindakan lanjutan kepada debitur (nasabah) sebagai sarana untuk menurunkan tingkat kemacetan kredit dengan cara reminding call, rescheduling maupun reconditioning.

Veronika moertini dalam jurnal yang ditulis pada tahun 2003, melakukan perbandingan antara algoritma data mining C4.5 dengan algoritma ID3 untuk melakukan klasifikasi nasabah. kesimpulan yang didapat adalah semakin banyak atribut yang digunakan sebagai dasar proses klasifikasi tingkat kemacetan nasabah, maka hasilnya akan semakin baik. Demikian juga dengan jumlah data yang digunakan sebagai data training semakin besar tingkat keberhasilan klasifikasi kredit semakin tinggi.

Mu-chen menggunakan algoritma *Neural Networks* (NNs) yang berbasis *scoring model* untuk mengklasifikasi kredit bermasalah, dan *Genetic Algorithms* (GAs) untuk mempertimbangkan kembali pengajuan kredit yang ditolak. Kesimpulan yang didapat dari penelitian ini adalah NNs tepat digunakan sebagai alat komputasi untuk masalah klasifikasi kredit. Sedangkan GAs dapat digunakan pihak kreditor untuk penerimaan bersyarat dan selanjutnya menjelaskan kondisi penyebab pengajuan kredit di tolak.

2.2. Landasan Teori

2.2.1. Konsep Sistem Pengambilan Keputusan

Pengambilan keputusan dalam manajemen memegang peranan yang sangat penting, dimana keputusan yang merupakan hasil pemikiran akhir yang ditetapkan oleh manajer akan dilaksanakan oleh bawahan atau organisasi yang dipimpinnya.

Kesalahan dalam pengambilan keputusan bisa berakibat pada citra organisasi maupun kerugian materi. Sehingga bisa dikatakan pengambilan keputusan adalah suatu proses pemikiran dalam pemecahan masalah untuk memperoleh hasil yang akan dilaksanakan.

Sebuah sistem keputusan merupakan model dari sistem yang terdiri atas keputusan yang bersifat tertutup atau terbuka. Sistem keputusan tertutup menganggap bahwa keputusan dipisah dari masukan yang tidak diketahui lingkungannya. sistem keputusan terbuka memandang keputusan sebagian berada dalam suatu lingkungan yang rumit dan sebagian tak diketahui. Keputusan dipengaruhi oleh lingkungan dan pada gilirannya proses keputusan kemudian mempengaruhi lingkungan.

Dalam keputusan model tertutup, komputer bertindak sebagai sebuah alat penghitung untuk bisa menghitung hasil optimum. Dalam keputusan model terbuka, komputer bertindak sebagai pembantu bagi pengambilan keputusan dalam menghitung, menyimpan, mencari kembali, menganalisis data dan sebagainya. Perancangan tersebut memungkinkan manusia pengambil keputusan mengalokasikan tugas bagi dirinya atau pada komputer (Margianti dkk, 1994).

2.2.1. Data mining

Data mining sering disebut juga *Knowledge Discovery in Databases* (KDD) adalah suatu istilah yang digunakan untuk menguraikan penemuan pengetahuan di dalam database. Penemuan pengetahuan tersebut bisa berupa penjelasan tentang masa lalu ataupun prediksi masa depan. *Data mining* menggunakan teknik statistika, matematika, kecerdasan buatan, dan *machine learning* untuk mendapatkan informasi dan pengetahuan yang bermanfaat dalam database. Keluaran dari *data mining* bisa dipakai untuk memperbaiki pengambilan keputusan di masa depan.

Kemajuan luar biasa yang terus berlanjut dalam bidang *data mining* didorong oleh beberapa faktor (Kusrini,2009).

- a. Pertumbuhan yang cepat dalam kumpulan data.

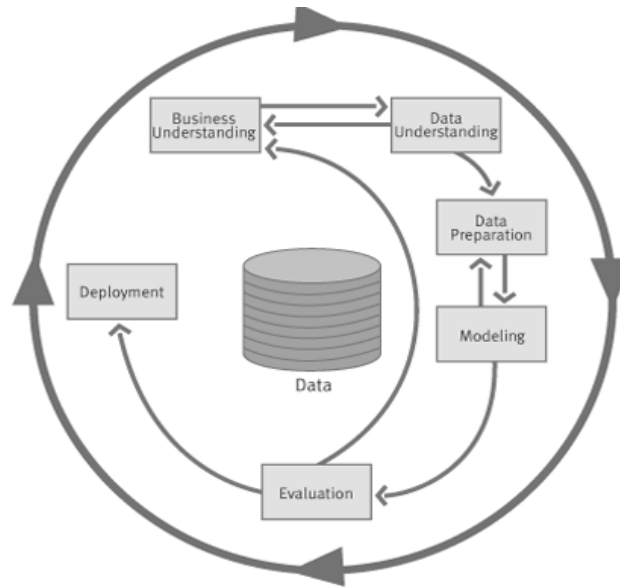
- b. Penyimpanan data dalam *warehouse* sehingga seluruh perusahaan memiliki akses ke dalam database.
- c. Adanya peningkatan akses data melalui navigasi web dan intranet.
- d. Tekanan kompetisi bisnis untuk meningkatkan penguasaan pasar dalam globalisasi ekonomi.
- e. Perkembangan teknologi perangkat lunak untuk *data mining*
- f. Perkembangan dalam kemampuan komputasi dan pengembangan kapasitas media penyimpanan.

Cross Industry Standard Proses for Data mining (CRISP-DM) yang dikembangkan tahun 1996 oleh analis dari beberapa industri seperti *Daimler Chrysler*, SPSS dan NCR menyediakan standar proses *data mining* dalam enam fase (Kusrini,2009).

1. Fase Pemahaman Bisnis (*Business Understanding Phase*)
 - a. Penentuan tujuan proyek dan kebutuhan secara detail dalam lingkup bisnis atau unit penelitian secara keseluruhan.
 - b. Menerjemahkan tujuan dan batasan menjadi formula dari permasalahan *data mining*.
 - c. Menyiapkan strategi awal untuk mencapai tujuan.
2. Fase Pemahaman Data (*Data Understanding Phase*)
 - a. Mengumpulkan data.
 - b. Menggunakan analisis penyelidikan data untuk mengenali lebih lanjut data dan pencarian pengetahuan awal.
 - c. Mengevaluasi kualitas data.
 - d. Jika diinginkan memilih sebagian kecil group data yang mungkin mengandung pola dari permasalahan.
3. Fase Pengolahan Data (*Data Preparation Phase*)
 - a. Menyiapkan kumpulan data yang akan digunakan untuk keseluruhan fase berikutnya.
 - b. Pilih kasus dan atribut yang ingin di analisa dan yang sesuai dengan analisis yang akan dilakukan.
 - c. Melakukan perubahan pada atribut jika dibutuhkan.

- d. Menyiapkan data awal sehingga siap untuk perangkat pemodelan.
4. Fase Pemodelan (*Modelling Phase*)
- a. Memilih dan mengaplikasikan teknik pemodelan yang sesuai.
 - b. Mengkalibrasi aturan model untuk mengoptimalkan hasil.
 - c. Memperhatikan kemungkinan bahwa beberapa teknik mungkin untuk digunakan pada permasalahan *data mining* yang sama.
 - d. Jika diperlukan, proses dapat kembali ke fase pengolahan data untuk menjadikan data ke dalam bentuk yang sesuai dengan spesifikasi kebutuhan teknik *data mining* tertentu.
5. Fase Evaluasi (*Evaluation Phase*)
- a. Mengevaluasi satu atau lebih model yang digunakan dalam fase pemodelan untuk mendapatkan kualitas dan efektivitas sebelum disebarkan untuk digunakan.
 - b. Menetapkan apakah terdapat model yang memenuhi tujuan pada fase awal.
 - c. Menentukan apakah terdapat permasalahan penting dari bisnis atau penelitian yang tidak tertangani dengan baik.
 - d. Mengambil keputusan berkaitan dengan penggunaan hasil dari *data mining*.
6. Fase Penyebaran (*Deployment Phase*)
- a. Menggunakan model yang dihasilkan.
 - b. Contoh sederhana penyebaran : pembuatan laporan.
 - c. Contoh kompleks penyebaran : penerapan proses *data mining* secara paralel pada departemen lain.

Hubungan antara setiap tahapan dalam tahapan data mining *CRISP-DM* seperti terlihat pada gambar 2.1

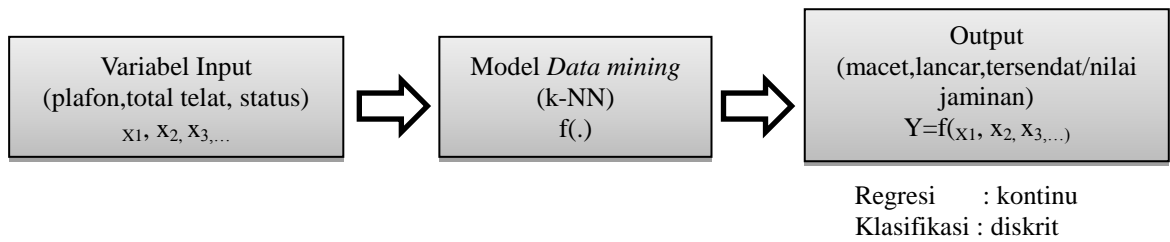


Gambar 2.1 Tahapan *Data mining* CRISP-DM
(Sumber: CRISP-DM.org)

2.2.2. Klasifikasi

Klasifikasi merupakan proses untuk menempatkan suatu objek ke dalam suatu kategori/kelas yang sudah didefinisikan sebelumnya berdasarkan model tertentu. *Data mining* merupakan penjelasan tentang masa lalu dan prediksi masa depan berdasarkan analisa pada sekelompok data. Dalam proses memprediksi masa depan ini *data mining* menggunakan beberapa model yaitu pemodelan prediktif dan deskriptif.

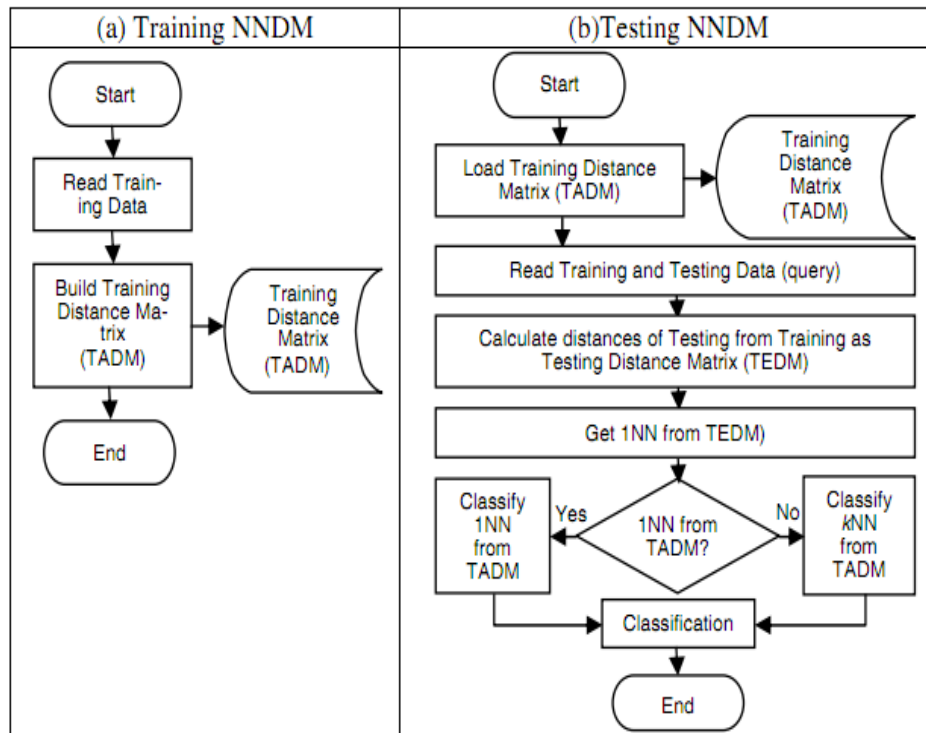
- a. Pemodelan prediktif diawali dengan pembentukan model untuk memprediksi hasil. Jika hasil pemodelan dalam bentuk kategori (diskrit) maka disebut sebagai klasifikasi sedangkan jika hasil pemodelan bernilai kontinu maka disebut regresi. Gambar 2.2 memperlihatkan contoh pemodelan prediktif beserta *output* dari pemodelan.



Gambar 2.2 Pemodelan Prediktif

- b. Pemodelan deskriptif atau lebih dikenal dengan istilah *clustering*, merupakan proses pengamatan terhadap kelompok data kemudian diikuti dengan pengelompokan data/*cluster* terhadap data yang mempunyai kesamaan ciri.

Dalam proses klasifikasi terdapat dua tahap yang harus dilewati yaitu tahap *learning* dan *testing*. Pada tahap *learning* sebagian data yang telah diketahui kelas datanya (data *training*) digunakan untuk membentuk model perkiraan. Pada tahap *testing*, model perkiraan yang sudah terbentuk diuji dengan sebagian data lainnya (data *testing*) untuk mengetahui akurasi dari model tersebut. Bila akurasinya dapat diterima maka model ini dapat dipakai untuk prediksi kelas data yang belum diketahui. Tahapan proses klasifikasi dengan k-NN seperti terlihat pada gambar 2.3 (Rayner, 2010)



Gambar 2.3 Tahapan Proses Klasifikasi k-NN (Sumber: Rayner, 2010)

2.2.3. Standarisasi Data

Proses algoritma k-NN menghitung jarak antar *record* untuk menentukan tingkat kemiripan. Satu kelemahan utama dalam menghitung jarak secara langsung dari data adalah dalam kasus di mana data memiliki rentang nilai pengukuran yang jauh berbeda. Misalnya, jika atribut pendapatan adalah pendapatan tahunan dalam rupiah, dan atribut umur adalah umur seseorang dalam tahun, maka pendapatan akan memiliki pengaruh yang jauh lebih tinggi dalam ukuran jarak dibanding umur, sehingga pengukuran jarak menjadi tidak relevan lagi. Salah satu solusinya adalah melakukan standarisasi data. Dalam proses standarisasi, data ditransformasi ke dalam rentang nilai [-1..1] atau [0..1]. Dalam penelitian ini digunakan standarisasi nilai dengan rentang [0..1] menggunakan rumusan standarisasi *min-max* seperti terlihat pada persamaan (1).

$$X_s = \frac{X - Min}{Max - Min} \quad (1)$$

Nilai X_s didefinisikan sebagai nilai setelah di standarisasi dimana X merupakan nilai *instance* atribut dan Min merupakan nilai minimum dalam atribut sebaliknya Max merupakan nilai maksimum dalam atribut. Setelah proses standarisasi maka bisa dilanjutkan dengan proses perhitungan jarak dengan algoritma k-NN. Proses standarisasi dapat digambarkan dalam tahapan algoritma *min-max*

Tahapan algoritma *min-max* berdasarkan persamaan (1)

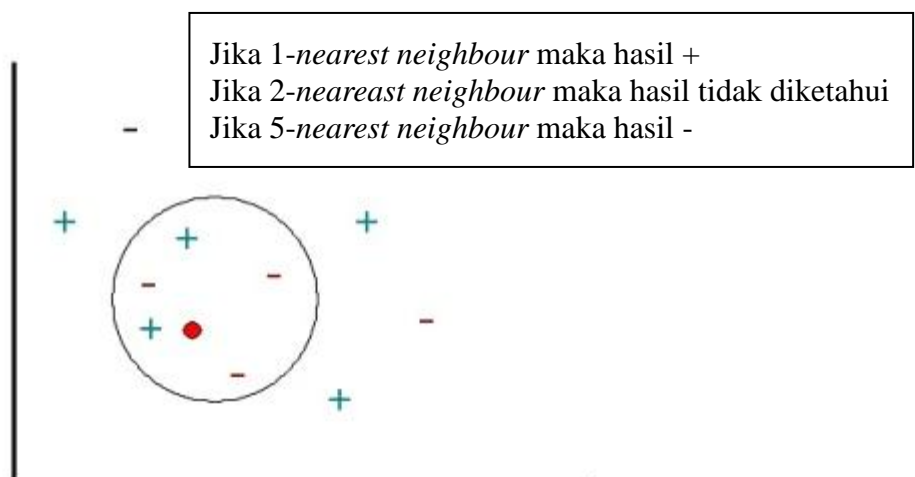
1. Baca data yang akan distandarisasi
2. Tentukan $I = 1$
3. Tentukan $L = 1$
4. Tentukan $J =$ jumlah atribut
5. Tentukan $N =$ jumlah *record*
6. While $I < J + 1$, Do
7. $Max_{(I)} = \text{Max}(\text{Instance Kolom ke-I})$
8. $Min_{(I)} = \text{Min}(\text{Instance Kolom ke-I})$
9. While $L < N+1$, Do
10. $A_{(L,I)} =$ *instance* baris ke-L kolom ke – I
11. $S_{(L,I)} = \frac{A_{(L,I)} - Min_{(I)}}{Max_{(I)} - Min_{(I)}}$
12. $A_{(L,I)} = S_{(L,I)}$
13. $L=L+1$
14. Loop (Step 9)
15. $I = I + 1$
16. $L=1$
17. Loop (Step 6)
18. Selesai

2.2.4. Algoritma *k-Nearest neighbour (k-NN)*

Algoritma *k-Nearest Neighbour (k-NN)* adalah algoritma pengklasifikasian data sederhana dimana penghitungan jarak terpendek dijadikan ukuran untuk mengklasifikasikan suatu kasus baru berdasarkan ukuran kemiripan. Algoritma ini telah digunakan dalam estimasi statistik dan pengenalan pola sejak awal tahun 1970-an. Algoritma *k-NN* tergolong dalam algoritma *supervised* yaitu proses pembentukan algoritma diperoleh melalui proses pembelajaran (*learning*) pada *record-record* lama yang sudah terklasifikasi dan hasil pembelajaran tersebut dipakai untuk mengklasifikasikan *record* baru dengan output yang belum diketahui.

Dalam algoritma *k-NN* sebuah data baru diklasifikasikan berdasarkan jarak data baru tersebut dengan tingkat kemiripan data baru terdekat terhadap data pola. Jumlah data tetangga terdekat ditentukan dan dinyatakan dengan *k*. Misalkan ditentukan $k=1$, maka kasus ini hanya diklasifikasikan untuk satu data dari tetangga terdekat. Jika nilai *k* didefinisikan berbeda oleh user, misal $k=5$ maka kasus dengan 5 jarak terpendek dipilih, kemudian diklasifikasi berdasarkan *instance* kelas target dimana kasus dengan jumlah mayoritas *instance* kelas target ditentukan sebagai klasifikasi untuk kasus baru.

Gambar 2.4 memperlihatkan bentuk representasi *k-NN* dengan nilai $k=1$, $k=2$ dan $k=5$



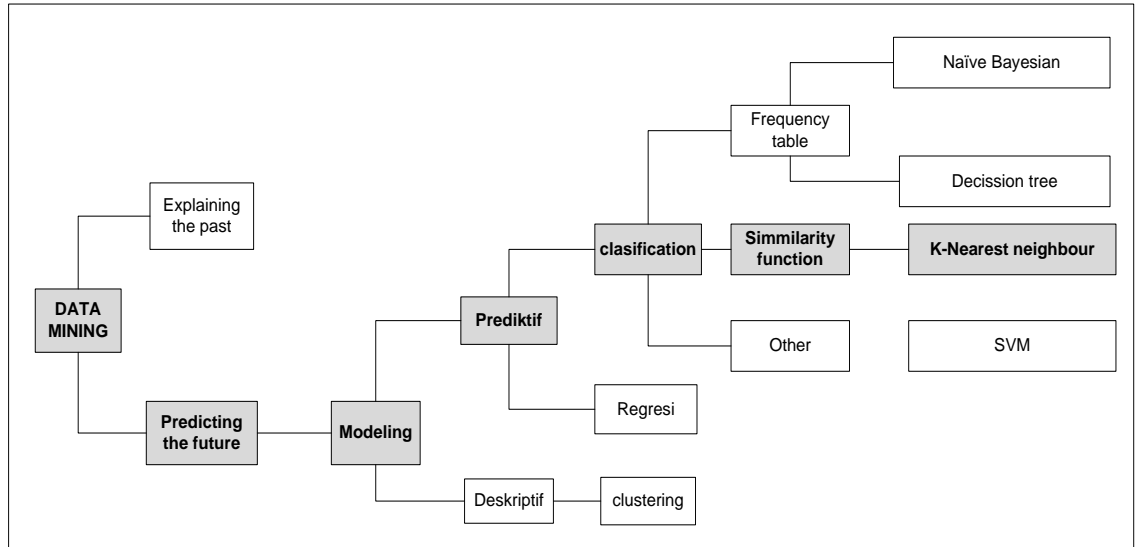
Gambar 2.4 prediksi data baru terhadap nilai *k* dalam *k-NN*

algoritma k-NN. Untuk lebih jelas melihat hubungan antara data mining Penentuan nilai k terbaik tergantung pada data. Nilai k yang tinggi bisa mengurangi efek *noise* pada klasifikasi, tetapi membuat batasan antara setiap kelas menjadi kabur. Sedangkan penentuan nilai $k=1$ belum tentu bisa menjawab permasalahan *data mining* dalam hal ini tingkat validitas. Nilai k terbaik dapat dipilih dengan optimasi parameter, misalnya dengan menggunakan *k-fold cross validation*. Untuk membedakan nilai k pada *cross validation* dengan nilai k pada k-NN, maka digunakan *n-fold cross validation* untuk mengacu kepada istilah yang sama yaitu *k-fold cross validation*.

Pada *n-fold cross validation*, data dibagi sejumlah n dan data ke- n digunakan sebagai data testing sedangkan data selain data ke- n sebagai data pembentuk pola atau data *training*. Pada proses fold ke n , Parameter nilai k tertentu digunakan untuk menguji validitas data terhadap data testing menggunakan algoritma k-NN. Pada akhir proses, rata-rata eror klasifikasi data dari masing-masing nilai k untuk seluruh *fold* dihitung. Dari proses tersebut diperoleh nilai k terbaik yang dapat dipakai sebagai k default dari algoritma k-NN.

k-NN efektif bila melibatkan data *training* dalam jumlah yang besar dan bisa mengurangi efek *noise*. Tetapi k-NN juga memiliki beberapa kelemahan dalam hal kebutuhan memori yang besar dalam menghitung jarak antar *record*, penentuan nilai k secara manual dan kecepatan dalam mengklasifikasi data tergantung dari banyaknya data.

Hubungan antara *data mining* dan k-NN adalah data mining merupakan pencarian pengetahuan dalam database yang digunakan untuk mendapatkan informasi yang dapat menjelaskan masa lalu maupun dipakai untuk memprediksi masa depan. Dalam memprediksi masa depan dibutuhkan model, jika model yang digunakan mempunyai tujuan pengelompokan data berupa atribut target maka termaksud dalam jenis pemodelan prediktif. Jika hasil pemodelan prediktif menghasilkan pengelompokan data dengan nilai diskrit maka disebut klasifikasi. Dalam mengklasifikasi data jika menggunakan fungsi berdasarkan tingkat kemiripan maka digunakan dan k-NN seperti terlihat pada gambar 2.5



Gambar 2.5 Hubungan *data mining* dengan k-NN
(Sumber: *Real Time Data mining*)

Secara umum untuk mendefinisikan jarak antara dua objek x dan y , digunakan rumus jarak Euclidean, seperti terlihat pada persamaan (2)

$$D_{xy} = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (2)$$

Jarak antara objek x dan y didefinisikan sebagai D_{xy} , dimana x_i merupakan *record* yang akan diprediksi dan y_i merupakan *record* data pola sedangkan nilai n didefinisikan sebagai jumlah atribut. Nilai i merujuk kepada record ke- i . Persamaan (2) merupakan rumus jarak untuk tipe data kontinu sehingga data nominal harus dikonversi ke bentuk numerik. Proses penghitungan jarak dapat digambarkan dalam tahapan algoritma k-NN

Tahapan algoritma k-NN berdasarkan persamaan (2)

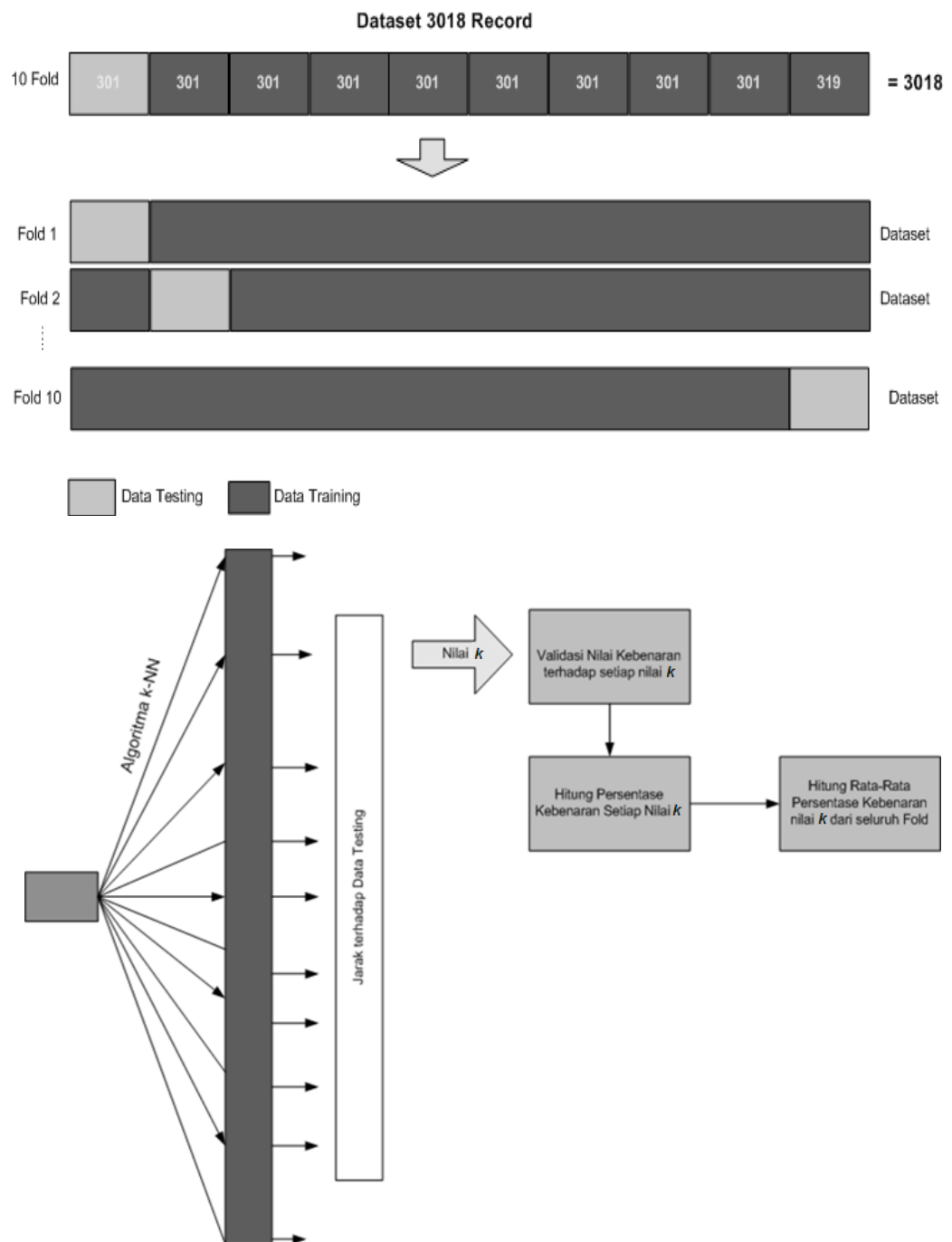
1. Baca data pola terstandarisasi
2. Baca data yang akan diprediksi terstandarisasi

3. Masukkan nilai k
4. Tentukan $J = \text{Jumlah Atribut}$
5. Tentukan $N = \text{jumlah record data pola}$
6. Tentukan $L = 1$
7. While not EOF
8. $I = 1$
9. $Y = 0$
10. While $I < J + 1$, Do
11. $A_{(L,I)} = \text{instance data baris ke-L kolom ke-I}$
12. $B_{(I)} = \text{instance data prediksi kolom ke - I}$
13. $X = (B_{(I)} - A_{(L,I)})^2$
14. $Y = Y + X$
15. $I = I + 1$
16. Loop (Step 10)
17. $Z_{(L)} = \text{SQRT}(Y)$
18. Replace jarak with $Z_{(L)}$
19. $L = L + 1$
20. Next 1
21. Loop (Step 7)
22. Sort data *ascending* by jarak
23. Filter data 1 to k
24. Hasil prediksi = nilai mayoritas

2.2.5. *n-fold Cross Validation*

n-fold cross validation atau *k-fold cross validation* merupakan salah satu metode yang digunakan untuk mengetahui rata-rata keberhasilan dari suatu sistem dengan cara melakukan perulangan dengan mengacak atribut masukan sehingga sistem tersebut teruji untuk beberapa atribut input yang acak. *n-fold cross validation* diawali dengan membagi data sejumlah *n-fold* yang diinginkan. Dalam proses *cross validation* data akan dibagi dalam n buah partisi dengan ukuran yang sama $D_1, D_2, D_3, \dots, D_n$ selanjutnya proses testing dan *training* dilakukan sebanyak n

kali. Dalam iterasi ke- i partisi D_i akan menjadi data testing dan sisanya akan menjadi data *training*. Untuk penggunaan jumlah fold terbaik untuk uji validitas, dianjurkan menggunakan *10-fold cross validation* dalam model. (Ron Kohavi, 1995). Contoh pembagian dataset dalam proses *10-fold cross validation* seperti terlihat pada gambar 2.6



Gambar 2.6 Contoh iterasi data dengan 10-fold cross validation

Gambar 2.6 memperlihatkan tahapan proses 10-fold cross validation dengan menggunakan model algoritma k-NN. Hitung jarak dari setiap data testing terhadap data *training*. Input parameter nilai *k* tertentu. Verifikasi hasil klasifikasi setiap nilai *k* dengan nilai klasifikasi sebenarnya dari data testing. Pada proses akhir dilakukan perhitungan rata-rata tingkat kebenaran atau tingkat eror dari tiap fold ke-*n* terhadap setiap nilai *k*.

Tahapan proses algoritma *n-fold cross validation* dari proses pengolahan *fold* tersebut dapat dijabarkan dalam tahapan algoritma *n-fold cross validation*.

- 1) Baca dataset terstandarisasi
- 2) Masukkan nilai *fold* (F)
- 3) Masukkan nilai *k*
- 4) T = jumlah *record* dataset
- 5) S = Jumlah *record* data testing (T/F)
- 6) Tetntukan L = 1
- 7) Tentukan M = 0
- 8) Partisi dataset sebanyak F, tiap partisi sebanyak S *record*
- 9) For I = 1 to F
- 10) F(I) = data testing
- 11) Not F(I) = data *training*
- 12) For N = 1 to S
- 13) For J = 1 to *k*
- 14) Jalankan fungsi algortima k-NN untuk setiap *record* (N) dalam tabel F(I) untuk nilai $k = J$
- 15) P = Hasil prediksi k-NN
- 16) H = Instance atribut target data testing ke-N
- 17) If H = P than Nilai = True; else Nilai = False
- 18) Replace hasil untuk K = J dengan nilai
- 19) J = J+1
- 20) Loop (step 13)

- 21) $N = N + 1$
- 22) Loop (step 12)
- 23) While $L < k + 1$
- 24) $M_{(I,L)} = \frac{\sum \text{Nilai False untuk } K=L}{T} \times 100\%$
- 25) Loop (step 24)
- 26) $I = I + 1$
- 27) Loop (step 9)
- 28) Selesai

2.2.6. Kategori Kredit

Berdasarkan SK Direksi BI NO.31/147/KEP/DIR tanggal 12 November 1998 tentang penggolongan kualitas kredit berdasarkan kemampuan membayar

1. Pembayaran tepat waktu, perkembangan rekening baik dan tidak ada tunggakan serta sesuai dengan persyaratan kredit. (Lancar)
2. Terdapat tunggakan pembayaran pokok dan/atau bunga sampai dengan 90 hari. (Dalam Perhatian Khusus)
3. Terdapat tunggakan pembayaran pokok dan/atau bunga yang telah melampaui 90 hari sampai dengan 180 hari. (Kurang Lancar)
4. Terdapat tunggakan pembayaran pokok dan/atau bunga yang telah melampaui 180 hari sampai dengan 270 hari. (Diragukan)
5. Terdapat tunggakan pembayaran pokok dan/atau bunga yang telah melampaui 270 hari. (Macet).

2.2.7. Analisa Kredit “*The 5 C of Credit Analysis*”

Dalam upaya memperkecil risiko dalam memberikan kredit, bank mempertimbangkan beberapa hal yang terkait dengan itikad baik (*willingness to pay*) dan kemampuan membayar (*ability to pay*) nasabah untuk melunasi kembali pinjaman beserta bunganya. Hal-hal tersebut tertuang dalam istilah “*The Five C of Credit Analysis*”.

1. Aspek Karakter (*character*), memuat data tentang penilaian terhadap karakter calon debitur dilihat dari berbagai sifat misalnya watak, kemauan,

kejujuran dan pengalaman hutang masa lalu.

2. Aspek Kemampuan (*capacity*), memuat data tentang penilaian terhadap kemampuan berusaha, kemampuan pemasaran, kemampuan membayar kembali hutangnya masa lalu dan hubungan dengan rekan usahanya.
3. Aspek Permodalan (*capital*), memuat data tentang penilaian terhadap data keuangan calon debitur yang meliputi harta lancar, harta tetap, hutang dan sebagainya.
4. Aspek Kondisi Ekonomi (*conditions of economi*), memuat data tentang kondisi usaha, kondisi rumah tangga, kondisi usaha yang berkaitan dengan kondisi ekonomi pada umumnya, dan sebagainya.
5. Aspek Jaminan (*collateral*), memuat data tentang agunan yang akan disediakan, kebendaannya, keberadaannya, kondisi jaminannya, nilai jualnya, penilaian terhadap kelayakannya, dan sebagainya.

2.2.8. Metode SDLC (*System Development Life Cycle*)

SDLC merupakan metode umum dalam pengembangan sistem informasi, dimana tahapan sistem dikerjakan secara terstruktur (Fatta, 2009).

1. Perencanaan

Merupakan langkah awal berupa pengumpulan informasi yang dibutuhkan sistem. Aktivitas tahap ini meliputi wawancara manajemen pengguna, merangkum pengetahuan yang didapat, mengestimasi cakupan proyek dan mendokumentasikan hasilnya. Tahapan ini akan menghasilkan laporan definisi masalah dan rangkuman tujuan yang ingin dicapai.

2. Analisa

Tujuan utama dari fase analisa adalah untuk memahami serta mendokumentasikan kebutuhan bisnis dan persyaratan proses dari sistem baru.

3. Desain

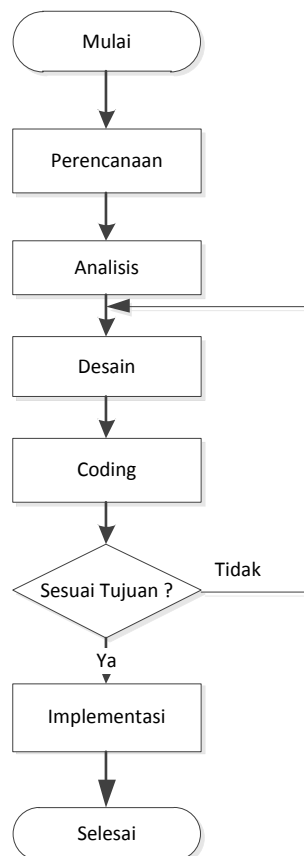
Tahapan mengubah kebutuhan yang masih berupa konsep menjadi spesifikasi sistem yang riil. Tahapan desain sistem dapat dibagi menjadi desain logis yang menghasilkan dokumen model data, rancangan tabel dan desain antar

muka dan desain fisik berupa aktifitas pemrograman/*coding*.

4. Implementasi

Terdiri atas tahap testing yaitu menguji hasil coding program aplikasi yang dihasilkan dari tahapan desain fisik dan tahap instalasi yaitu tindakan lanjutan setelah program lulus testing berupa penginstalan program pada organisasi.

Tahapan metode penelitian SDLC dapat digambarkan dalam tahapan proses seperti yang terlihat pada gambar 2.7



Gambar 2.7 Tahapan Metode Penelitian SDLC