

**DATA MINING PADA PENENTUAN KELAYAKAN KREDIT
MENGUNAKAN ALGORITMA K-NN BERBASIS FORWARD
SELECTION**

**DATA MINING ON CREDIT FEASIBILITY DETERMINATION
USING K-NN ALGORITHM BASED ON FORWARD SELECTION**

Sitti Harlina

*Jurusan Teknik Informatika, Sekolah Tinggi Manajemen Informatika Dan Komputer
Diponegara, Makassar*

Email : Sitiharlina76@gmail.com

Diterima: 5 Juli 2018/ Disetujui : 23 Juli 2018

ABSTRACT

Along with the development of economic and banking services, the term credit is one of the banking products, in determining credit worthiness required data, information about creditworthiness is an interesting issue to be in carefully. In line with business growth, the leasing or crediting sector (credit), has complex and complex problems and problems in which the data are grouped into two classes, good credit and bad credit. In banking activities the feasibility of credit determination is often the problem of bad debts caused by the failure of partial return of loans provided to the borrowers for it required accurate data prior to lending or credit. So it is appropriate to use data mining classification techniques. This study discusses the Data Mining on the determination of creditworthiness using the K-nn-based Forward Selection algorithm. K-nn algorithm applied to consumer data using credit financial services with the help of Forward Selection. The experimental results show that K-nn in classifying credit risk has accuracy value for K-nn at $K = 11$ accuracy = 68,30%, and for Forward Selection at $K = 9$ which accuracy level = 73,60%, using feature Forward Selection can produce a good degree of accuracy, so it can be applied to determine the risk of credit risk eligibility.

Keywords: *Data Mining, creditworthiness, K-nn Algorithm, Forward Selection.*

ABSTRAK

Seiring perkembangan ekonomi dan pelayanan jasa perbankan maka istilah kredit bukan lagi hal yang aneh karena kredit merupakan salah satu produk perbankan, dalam penentuan kelayakan kredit diperlukan data, informasi tentang kelayakan kredit merupakan masalah yang menarik untuk di teliti. Sejalan dengan pertumbuhan bisnis, sektor pembiayaan (leasing) atau lembaga pengkreditan (kredit), mempunyai persoalan dan masalah yang rumit dan kompleks dimana data tersebut dikelompokkan menjadi dua kelas, yaitu kredit baik dan kredit buruk. Didalam kegiatan perbankan kelayakan penentuan kredit sering terjadi masalah kredit macet yang disebabkan oleh gagalnya pengembalian sebagian pinjaman yang di berikan kepada para peminjam untuk itu diperlukan data yang akurat sebelum pemberian pinjaman atau kredit. Sehingga tepat menggunakan teknik klasifikasi data mining. Penelitian ini membahas mengenai Data Mining pada penentuan kelayakan kredit menggunakan algoritma K-nn berbasis Forward Selection. Algoritma K-nn yang diterapkan pada data konsumen yang menggunakan jasa keuangan kredit dengan bantuan Forward Selection. Hasil eksperimen menunjukkan bahwa K-nn dalam melakukan klasifikasi risiko kredit memiliki nilai akurasi untuk K-nn di $K=11$ accuracy = 68,30 %, dan untuk Forward Selection di $K=9$ yang tingkat

accuracy = 73,60%, dengan menggunakan fitur *Forward Selection* dapat menghasilkan tingkat akurasi yang baik, sehingga dapat diterapkan untuk menentukan risiko kelayakan risiko kredit.

Kata kunci : Data Mining, kelayakan kredit, Algoritma K-nn, Forward Selection.

I. PENDAHULUAN

Dalam dunia perbankan keberadaan kredit merupakan hal yang tak lepas dari salah satu bentuk usaha yang di jalankan oleh dunia perbankan. Dalam menjalankan kegiatan ini, bank akan selalu memperhatikan resiko kredit yang akan terjadi dalam hal pemberian kredit, untuk itu data tentang penentuan kelayakan kredit akan sangat diperlukan guna menunjang roda bisnis kehidupan perbankan.

Dunia perbankan saat ini tidak bisa lagi dipisahkan dengan keberadaan teknologi karena perkembangan teknologi telah memberi kemudahan hampir di setiap sektor usaha. Dalam usaha perbankan kemudahan dalam pencarian, penyajian dan penanganan, serta pengolahan data dan informasi amatlah penting .

Informasi tentang kelayakan kredit merupakan masalah yang menarik untuk di teliti. Sejalan dengan pertumbuhan bisnis, sektor pembiayaan (leasing) atau lembaga pengkreditan (kredit), mempunyai persoalan dan masalah yang rumit dan kompleks.

Salah satu kegiatan yang sangat tumbuh saat ini adalah pemberian kredit dengan implikasi risiko kerugian yang cukup besar. Bank sebagai lembaga keuangan memiliki kegiatan utama yaitu menghimpun dana masyarakat dalam bentuk tabungan dan lain-lain serta kemudian menyalurkannya kembali kepada masyarakat dalam bentuk kredit[1].

Didalam kegiatan perbankan kelayakan penentuan kredit sering terjadi masalah kredit macet yang disebabkan oleh gagalnya pengembalian sebagian pinjaman yang di berikan kepada para peminjam untuk itu diperlukan data yang akurat sebelum pemberian pinjaman atau kredit.

Masalah ini oleh pihak perbankan dan leasing telah disadari bahwa dalam menjalankan fungsi jasa-jasa keuangan, bank harus dapat mengidentifikasi dan memprediksi nasabah dengan baik salah satunya dengan penentuan kelayakan, mengidentifikasi dan memprediksi dengan cara mengelola berbagai jenis risiko keuangan secara efektif, agar dampak negatif tidak dapat terjadi dan menghindari atau menghilangkan kerugian yang besar akibat dari tidak dijalkannya manajemen risiko yang efektif dan disiplin.

Namun kegiatan berisiko tersebut harus diambil untuk mendapatkan peluang untuk mendapatkan keuntungan, dengan cara meminimalkan risiko yang akan timbul dengan manajemen risiko. Kegagalan akan berdampak kepada sistem perbankan dan bahkan sistem perekonomian. Risiko kredit adalah eksposur yang timbul sebagai akibat kegagalan pihak lawan (*counterparty*) memenuhi kewajibannya. Risiko ini timbul sebagai akibat dari kinerja satu atau lebih debitur yang buruk. Kinerja yang buruk dapat berasal dari ketidak mampuan debitur untuk memenuhi sebagian atau seluruh isi perjanjian kredit yang telah disepakati bersama. Yang menjadi dasar dari perhatian bank dalam hal ini adalah kondisi keuangan dan nilai pasar dari jaminan serta yang paling penting adalah karakter dari debitur [2].

Evaluasi risiko kredit merupakan masalah yang menarik untuk diteliti. Beberapa riset bidang komputer untuk mengurangi risiko kredit telah banyak dilakukan dalam rangka menghindari kehancuran suatu perusahaan pembiayaan. Untuk menyelesaikan masalah analisa risiko kredit, data akan dikelompokkan menjadi dua kelas, yaitu kredit baik dan kredit buruk. Sehingga tepat menggunakan teknik klasifikasi *data mining*. *Data mining* adalah suatu proses mencari pola atau informasi menarik dalam data terpilih dengan menggunakan teknik atau metode tertentu, yang bertujuan untuk menemukan pola secara otomatis atau semi otomatis dari data yang sudah

ada di dalam basis data yang dimanfaatkan untuk menyelesaikan suatu masalah [3]. *Data mining* memiliki beberapa teknik, diantaranya *clustering* dan klasifikasi. *Clustering* merupakan pengelompokan record, pengamatan atau memperhatikan dan membentuk kelas objek-objek yang memiliki kemiripan, kluster adalah kumpulan record yang memiliki kemiripan satu dengan yang lainnya dan memiliki ketidakmiripan dengan record-record dalam kluster lain. Klasifikasi bertujuan untuk membagi objek yang ditugaskan hanya ke salah satu nomor kategori yang disebut kelas [4].

Penelitian mengenai analisis kelayakan pemberian kredit untuk konsumen khususnya kredit kepemilikan barang dengan metode klasifikasi data mining telah banyak dilakukan. Diantaranya, hasil

penelitian yang dilakukan Tian-Shyug Lee, Chih-Chou Chiu, *et all* [5] yang menggunakan *data mining* dengan metode *Classification and Regresssion Tree* (CART) dan *Multivariate Adaptive Regression Splines* (MARS) untuk mengklasifikasikan kredit pelanggan dengan nilai tingkat akurasi sebesar 77,75%. Namun penelitian tersebut, metode CART memiliki kelemahan dalam mengatasi ketidak stabilan pemilah pada teknik pohon keputusan klasifikasi dan jika data berukuran besar dengan banyak atribut maka model yang terbentuk menjadi rumit dan sulit dipahami, sehingga perlu dilakukan pemangkasan (*pruning*) yang dapat mengurangi akurasi.

2. METODE PENELITIAN\

Data Mining

Data Mining merupakan analisis dari peninjauan kumpulan data untuk menemukan hubungan yang tidak di duga dan meringkas data dengan cara yang berbeda dengan sebelumnya, yang dapat dipahami dan bermanfaat bagi pemilik data.[6]

Data Mining merupakan bidang dari beberapa bidang keilmuan yang menyatukan teknik dari pembelajaran mesin, pengenalan pola, statistik, database, dan visualisasi untuk penanganan permasalahan pengambilan informasi dari database yang besar.[6]

Data Mining dibagi menjadi beberapa kelompok berdasarkan tugas yang dapat dilakukan, yaitu:

1. Deskripsi;
Deskripsi dari pola dan kecendrungan sering memberikan kemungkinan penjelasan untuk suatu pola atau kecendrungan.
2. Estimasi;
Estimasi hampir sama dengan klasifikasi, kecuali variabel target lebih ke arah numerik daripada ke arah kategori. Model dibangun menggunakan record lengkap yang menyediakan nilai dari variabel target sebagai nilai prediksi. Selanjutnya, pada peninjauan berikutnya estimasi nilai dari variabel target dibuat berdasarkan nilai variabel prediksi.
3. Prediksi;
Prediksi hampir sama dengan klasifikasi dan estimasi, kecuali bahwa dalam prediksi nilai dari hasil akan ada di masa mendatang.
4. Klasifikasi;
Dalam klasifikasi, terdapat target variabel kategori,
5. Pengklusteran
Pengklusteran merupakan pengelompokan record, pengamatan atau memperhatikan dan membentuk kelas objek – objek yang memiliki kemiripan satu dengan yang lainnya dan memiliki ketidakmiripan dengan record-record dalam kluster lain.
6. Asosiasi
Tugas asosiasi dalam data mining adalah menemukan atribut yang muncul dalam satu waktu.

Kredit

Kemampuan untuk melaksanakan suatu pembelian atau mengadakan suatu pinjaman dengan suatu janji, pembayaran akan dilaksanakan pada jangka waktu yang telah disepakati.[7] Pengertian kredit yang lebih mapan untuk kegiatan perbankan di Indonesia telah dirumuskan dalam Undang – Undang Pokok Perbankan No. 7 Tahun 1992 yang menyatakan bahwa kriteria adalah penyediaan uang / tagihan yang dapat dipersamakan dengan itu berdasarkan persetujuan / kesepakatan pinjam meminjam antara pihak bank dengan pihak lain yang mewajibkan pihak peminjam untuk melaksanakan dengan jumlah bunga sebagai imbalan.

Dalam praktek sehari – hari pinjaman kredit dinyatakan dalam bentuk perjanjian tertulis baik dibawah tangan maupun secara materiil. Dan sebagai jaminan pengaman, pihak peminjam akan memenuhi kewajiban dan menyerahkan jaminan baik bersifat kebendaan maupun bukan kebendaan.

Algoritma K-nn

Algoritma K-nn selalu diawali dengan pencarian nilai K, kemudian di lakukan penghitungan pada data ujia dan data latihan [8].

Nearest Neighbor adalah pendekatan untuk mencari kasus dengan menghitung kedekatan antara kasus baru dengan kasus lama, yaitu berdasarkan pada pencocokan bobot dari sejumlah fitur yang ada[9].

K-nearest neighbor (K-nn) termasuk kelompok instance-based learning. Algoritma ini juga merupakan salah satu teknik lazy learning. K-nn dilakukan dengan mencari kelompok k objek dalam data training yang paling dekat (mirip) dengan objek pada data baru atau data testing.

Ada banyak cara untuk mengukur jarak kedekatan antara data baru dengan data lama (data training), diantaranya euclidean distance dan manhattan distance (city block distance), yang paling sering digunakan adalah euclidean distance dari Bramer, yaitu:

$$\sqrt{(a_1 - b_1)^2 + (a_2 - b_2)^2 + \dots + (a_n - b_n)^2}$$

Dimana $a = a_1, a_2, \dots, a_n$, dan $b = b_1, b_2, \dots, b_n$ mewakili n nilai atribut dari dua record. Untuk atribut dengan nilai kategori, pengukuran dengan euclidean distance tidak cocok. Sebagai penggantinya, digunakan fungsi sebagai berikut[6], $\text{different}(a_i, b_i) = \{ 0 \text{ jika } a_i = b_i \}$ selainnya. Dimana a_i dan b_i adalah nilai kategori. Jika nilai atribut antara dua record yang dibandingkan sama maka nilai jaraknya 0, artinya mirip, sebaliknya, jika berbeda maka nilai kedekatannya 1, artinya tidak mirip sama sekali. Misalkan atribut warna dengan nilai merah dan merah, maka nilai kedekatannya 0, jika merah dan biru maka nilai kedekatannya 1. Untuk mengukur jarak dari atribut yang mempunyai nilai besar, seperti atribut pendapatan, maka dilakukan normalisasi. Normalisasi bisa dilakukan dengan min-max normalization atau Z-score standardization [6].

Jika data training terdiri dari atribut campuran antara numerik dan kategori, lebih baik gunakan min-max normalization [6].

Untuk menghitung kemiripan kasus, digunakan rumus [10]

$$\text{Similarity}(p, q) = \frac{\sum_{i=1}^n f(p_i, q_i) \times w_i}{w_i}$$

Keterangan :

P = Kasus baru

q = Kasus yang ada dalam penyimpanan

n = Jumlah atribut dalam tiap kasus

i = Atribut individu antara 1 sampai dengan n

f = Fungsi similarity atribut i antara kasus p dan kasus q

w = Bobot yang diberikan pada atribut ke-i

Forward Selection

Forward Selection merupakan salah satu metode pemodelan (pembangunan model linier) untuk menemukan kombinasi peubah yang “terbaik” dari suatu gugus variabel. Dalam Prosedur Forward selection, sekiranya variabel masuk kedalam persamaan maka tidak bisa dihilangkan.

Selain itu, Forward selection dapat berarti memasukkan variabel bebas yang memiliki korelasi yang paling erat dengan variabel tak bebasnya (variabel yang paling potensial untuk memiliki hubungan linier dengan Y). kemudian secara bertahap memasukkan variabel bebas yang potensial berikutnya dan nanti akan berhenti sampai tidak ada lagi variabel bebas yang potensial.

Prosedur Forward Selection dimulai dengan sebuah persamaan yang terdiri dari suku konstanta, tidak terdiri dari predictor variable. Variable pertama yang masuk ke dalam persamaan adalah variable yang memiliki simple correlation tertinggi dan signifikan dengan variable Y. Jika koefisien regresi signifikan berbeda dari 0 maka tetap dipakai dalam persamaan, dan dilakukan pencarian

variable kedua. Variabel yang masuk ke dalam persamaan sebagai variabel kedua adalah variabel yang memiliki korelasi tertinggi kedua dan masih signifikan dengan Y . kemudian koefisien regresi dari variabel kedua diuji. Jika signifikan, maka dilakukan pencarian terhadap variabel ketiga dengan cara yang sama. Prosedur dihentikan saat pemasukan variabel terakhir tidak memiliki koefisien regresi dan tidak signifikan atau semua variabel masuk dalam persamaan. Koefisien regresi yang signifikan dari variabel terakhir dilihat dari uji-t dari persamaan terakhir.

Prosedur *forward selection* sebagai berikut.

- a. Metode *forward selection* dimulai dengan tidak ada variabel bebas (model hanya berisi konstanta).
- b. Variabel yang pertama kali masuk dalam model adalah variabel X yang berkorelasi paling tinggi dengan Y dan nilai p -value-nya kurang dari nilai kritis α , kemudian buat persamaan regresinya dengan memasukkan variabel tersebut dalam model.
- c. Pemilihan variabel berikutnya adalah variabel yang memiliki korelasi parsial terbesar dengan Y dan nilai p -value-nya kurang dari nilai kritis α (variabel dengan nilai p -value lebih dari nilai kritik α diabaikan).
- d. Buat model regresinya dengan memasukkan variabel tersebut.
- e. Uji parameter yang telah ada di dalam model
- f. Begitu seterusnya, ulangi langkah c dan seterusnya sampai diperoleh model terbaik.

Klasifikasi

Klasifikasi adalah proses penemuan model atau fungsi yang menggambarkan dan membedakan kelas data atau konsep yang bertujuan agar bisa digunakan untuk memprediksi kelas dari objek yang label kelasnya tidak diketahui [11].

Algoritma klasifikasi yang banyak digunakan secara luas, yaitu *Decision / classification trees*, *Forward selection*, *Bayesian classifiers / Naïve Bayes*, *Neural Networks*, *Analisa Statistik*, *Algoritma Genetika*, *Rough sets*, *K-Nearest Neighbor*, *Metode Rule Based*, *Memory based reasoning*, dan *Support Vector Machines (SVM)*.

Klasifikasi data terdiri dari 2 (dua) langkah proses. Pertama adalah *learning* (fase *training*), dimana algoritma klasifikasi dibuat untuk menganalisa *data training* lalu di representasikan dalam bentuk *rule* klasifikasi. Proses kedua adalah klasifikasi, dimana data tes digunakan untuk memperkirakan akurasi dari *rule* klasifikasi [11].

Proses klasifikasi didasarkan pada empat komponen :

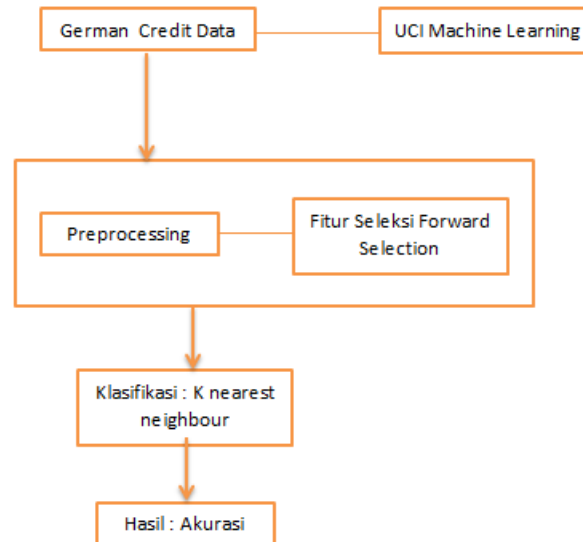
- a. **Kelas**
Variabel dependen yang berupa kategorikal yang merepresentasikan ‘label’ yang terdapat pada objek. Contohnya: resiko penyakit jantung, resiko kredit, *customer loyalty*, jenis gempa dll.
- b. **Predictor**
Variabel independen yang direpresentasikan oleh karakteristik (atribut) data. Contohnya: merokok, minum alkohol, tekanan darah, tabungan, aset, gaji
- c. **Training dataset**
Satu set data yang berisi nilai dari kedua komponen di atas yang digunakan untuk menentukan kelas yang cocok berdasarkan *predictor*.
- d. **Testing dataset**
Berisi data baru yang akan diklasifikasikan oleh model yang telah dibuat dan akurasi klasifikasi dievaluasi.

*Desain Eksperimen*Tabel 1: Atribut & Tipe *German Credit Data*

Atribut	Tipe
Status of existing checking account	Kategori
Duration in month	Numerik
Credit history	Kategori
Purpose	Kategori
Credit amount	Numerik
Savings account	Kategori
Present employment since	Kategori
Installment of disposable income	Numerik
Personal status n sex	Kategori
Other debtors/guarantors	Kategori
Present residence since	Numerik
Property	Kategori
Age	Numerik
Other installment plans	Kategori
Housing	Kategori
Existing credits at this bank	Numerik
Job	Kategori
Number of people being liable to provide maintenance for	Numerik
Telephone	Kategori
Foreign work	Kategori

Data yang digunakan pada penelitian ini berasal dari *University of California, Irvine (UCI) Machine Learning* dengan judul *German Credit data*. Dataset ini berjumlah 1000 *record* dan terdiri dari 20 atribut, dengan 1 class, 17 atribut bertipe numerik dan 13 atribut lainnya bertipe kategorikal[12].Eksperimen dilakukan dengan melakukan pengujian terhadap dataset serta klasifikasi risiko kredit menggunakan K - nn berbasis *Forward Selection* dan menerapkannya dalam RapidMiner. Eksperimen akan menghasilkan tingkat akurasi.

Adapun langkah-langkah eksperimen ini dapat dilihat di gambar 1.



Gambar 1: Alur Penelitian

3. HASIL DAN PEMBAHASAN

Setelah dilakukan klasifikasi terhadap dataset konsumen kredit menggunakan algoritma K -nn, maka didapatkan hasil perbandingan untuk tingkat akurasi dengan menggunakan fitur seleksi *Forward selection*.

Tabel 2 : Evaluasi Jerman Kredit dengan menggunakan Algoritma K -nn

<i>Algoritma</i>	<i>k-</i>	<i>Accuracy</i>
<i>K-nn</i>	<i>1</i>	<i>61,90%</i>
	<i>3</i>	<i>62,30%</i>
	<i>5</i>	<i>65,30%</i>
	<i>7</i>	<i>67,50%</i>
	<i>9</i>	<i>68,00%</i>
	<i>11</i>	<i>68,30%</i>

Tabel 3 : Evaluasi Jerman Kredit dengan menggunakan Algoritma K-nn berbasis *forward Selection*

<i>Algoritma</i>	<i>k-</i>	<i>Accuracy</i>
<i>K-nn +Forward Selection</i>	<i>1</i>	<i>69,60%</i>
	<i>3</i>	<i>70,00%</i>
	<i>5</i>	<i>70,00%</i>
	<i>7</i>	<i>73,00%</i>
	<i>9</i>	<i>73,60%</i>
	<i>11</i>	<i>73,20%</i>

Tabel 4 : Perbandingan Akurasi

<i>Algoritma</i>	<i>K-NN</i>	<i>K-NN berbasis Forward Selection</i>
<i>K-nn</i>	<i>K =11</i>	<i>68,30%</i>
<i>K-nn berbasis Forward selection</i>	<i>K = 9</i>	<i>73,60%</i>

Nilai akurasi tertinggi sebesar 73.60% dicapai ketika klasifikasi teks dilakukan dengan algoritma *K-nn* dengan berbasis *Forward Selection* sedangkan nilai akurasi terendah sebesar 61,90 % dicapai dengan menggunakan algoritma, *K-nn* tanpa penggunaan *Forward Selection*.

Penerapan metode *K-nn* berbasis *Forward Selection* dalam klasifikasi ini berpengaruh terhadap hasil akurasi yang didapatkan, karena dengan menerapkan *Forward Selection* maka akan menghilangkan atribut-atribut yang tidak relevan. Sehingga untuk klasifikasi risiko kelayakan kredit menggunakan *K-nn* ini didapatkan untuk akurasi tertinggi dengan menerapkan *Forward Selection* sebesar 73,60 %.

4. KESIMPULAN

Klasifikasi risiko penentuan kelayakan kredit menggunakan algoritma *K-NN* berbasis *Forward Selection* telah dilakukan dengan hasil akurasi 73,60%. Hasil ini diperoleh dengan menggunakan Algoritma *K-nn* berbasis *Forward Selection*. Penggunaan fitur *Forward Selection* dalam pemrosesan data akan mempengaruhi hasil pencapaian akurasi yang didapatkan. Beberapa hal masih perlu dilakukan untuk meningkatkan kinerja dari penelitian ini. Salah satunya adalah dengan melakukan pengujian *K-nn* berbasis *Forward Selection* terhadap data set yang berbeda.

5. SARAN

Saat ini penentuan kelayakan kredit telah banyak mempergunakan metode lain, dan dalam penelitian ini penulis menggunakan K-nn dengan berbasis *Forward Selection* dimana tingkat akurasi menunjukkan hasil yang tinggi, tetapi masih perlu penelitian lebih lanjut agar peningkatan akurasi dalam penentuan kelayakan kredit lebih meningkat.

UCAPAN TERIMA KASIH

Selama pembuatan penelitian ini, penulis ingin mengucapkan terima kasih kepada pihak-pihak berikut ini:

1. Bapak Dr. Purwanto sebagai pengajar matakuliah *Data Mining* yang sudah mengajarkan metode-metode penyelesaian masalah.
2. Ketua P4M STMIK Dipanegara dan Rekan-rekan Pascasarjana Magister Teknik Informatika Kelas Afiriasi 27 yang telah mendukung kami dalam berkarya.

Daftar Pustaka

- [1] Kasmir, 2012, *Bank dan Lembaga Keuangan lainnya*, Rajawali Pers.
- [2] D. Setiawan, 2007, *Analisis Terhadap Penerapan Manajemen Risiko Kredit pada PT. Bank Ekspor Indonesia*, Pascasarjana Universitas Gunadarma, Jakarta.
- [3] I. H. Witten dan E. Frank, 2011, pp.76,77 *Data Mining: Practical Machine Learning Tools and Techniques*, vol. 31, New Jersey: Morgan Kauffman Publishers.,
- [4] M. Bramer, 2013, *Principles of Data Mining*, London: Springer.
- [5] T. S. Lee, C. C. Chiu, Y. C. Chou dan C. J. Lu, 2004, *Mining the Customer Credit Using Classification and Regression Splines*, Fu-Jen Catholic University, Hsin Chuang.
- [6] Daniel T Larose, 2006, *Data Mining Methods And Models*, Published by John Wiley, WWW.Companion Web Site Available, & sons Inc. New Jersey Published Simultaneously-in Canada.
- [7] Astiko, 1996, *Manajemen Perkreditan*, Yogyakarta, Andi Offset
- [8] Tutus Praningsih, Indra Budi, 2017, *Sistem Prediksi Penyakit Kanker Serviks Menggunakan CART, Naïve Bayes, dan K-NN*, *Citec Journal, Creative Information Technology*. ISSN 2354-5771 volume 4, Februari 2017- April 2017: Universitas AMIKOM Yogyakarta.
- [9] Sri Wahyuningsih, Dyah Retno Utari, 2018, Perbandingan Metode K-Nearest Neighbour, Naïve Bayes dan Decision Tree untuk Prediksi Kelayakan Pemberian Kredit, Prosiding KNSI, STMIK Atma Luhur Pangkalpinang, Bangkabelitung.
- [10] Kusriani dan E. T. Luthfi, 2009, *Algoritma Data Mining*, Yogyakarta: Andi Publisher.
- [11] F. Gorunescu, 2011, *Data Mining: Concept, Models and Tsechniques*, Springer.
- [12] University of California , "Machine Learning Repository.[Online].Available repository, <https://archive.ics.uci.edu/ml/datasets/statlog+%28German+%29> +Credit+Data%29[diakses 19 October 2014].