

Algoritma Naïve Bayes, Decision Tree, dan SVM untuk Klasifikasi Persetujuan Pembiayaan Nasabah Koperasi Syariah

Naïve Bayes, Decision Tree, and SVM Algorithm for Classification of Sharia Cooperative Customer Financing Approval

Nurajijah, Dwiza Riana*)

Program Studi Magister Ilmu Komputer, STMIK Nusa Mandiri Jakarta
Jl. Kramat Raya No.18, Senen, Kota Jakarta Pusat, DKI Jakarta, Indonesia 10420

Cara sitasi: N. Nurajijah and D. Riana, "Algoritma Naïve Bayes, Decision Tree dan SVM untuk Prediksi Persetujuan Pembiayaan Nasabah Koperasi Syariah," *Jurnal Teknologi dan Sistem Komputer*, vol. 7, no. 2, 2019. doi: 10.14710/jtsiskom.7.2.2019.77-82, [Online].

Abstract - *The decision on financing approval in sharia cooperatives has a high risk of the inability of customers to pay their credit obligations at maturity or referred to as bad credit. To maintain and minimize risk, an accurate method is needed to determine the financing agreement. The purpose of this study is to classify sharia cooperative loan history data using the Naïve Bayes algorithm, Decision Tree and SVM to predict the credibility of future customers. The results showed the accuracy of Naïve Bayes algorithm 77.29%, Decision Tree 89.02% and the highest Support Vector Machine (SVM) 89.86%.*

Keywords - data mining; Naive Bayes; Decision Tree; SVM; credit financing

Abstrak - *Keputusan persetujuan pembiayaan pada koperasi syariah memiliki risiko tinggi atas ketidakmampuan nasabah dalam membayar kewajiban kreditnya pada saat jatuh tempo atau disebut dengan kredit macet. Untuk menjaga dan meminimalisir risiko tersebut dibutuhkan metode yang akurat untuk menentukan persetujuan pembiayaan. Tujuan penelitian ini adalah untuk melakukan klasifikasi data histori pinjaman nasabah koperasi syariah menggunakan algoritme Naïve Bayes, Decision Tree dan SVM untuk memprediksi kredibilitas calon nasabah selanjutnya. Hasil penelitian menunjukkan akurasi algoritme Naive Bayes 77,29%, Decision Tree 89,02% dan yang tertinggi Support Vector Machine (SVM) 89,86%.*

Kata Kunci - data mining; Naive Bayes; Decision Tree; SVM; pembiayaan kredit

I. PENDAHULUAN

Keputusan persetujuan pembiayaan atau kredit memiliki risiko tinggi atas ketidakmampuan nasabah dalam membayar kewajiban kreditnya pada saat jatuh

tempo [1]. Pengelolaan risiko kredit telah dianggap sebagai tugas yang paling penting bagi perusahaan dan lembaga keuangan lainnya [2]. Pada kasus permohonan kredit oleh nasabah, pengambil keputusan harus mampu mengambil keputusan yang tepat untuk menerima atau menolak permohonan kredit tersebut [3].

Teknik data mining sangat digunakan dalam industri perbankan yang membantu mereka bersaing di pasar dan menyediakan produk yang tepat untuk pelanggan yang tepat dengan risiko yang lebih kecil [4]. Data mining mengambil pengetahuan menggunakan data dengan jumlah besar yang disimpan dalam basis data, data warehouse atau informasi dalam repository [5].

Teknik data mining untuk klasifikasi resiko kredit telah banyak dilakukan dengan berbagai metode. Algoritme klasifikasi C4.5 dan Naïve Bayes dapat digunakan untuk membantu pengambilan keputusan pemberian kredit dan menganalisa kemampuan nasabah dalam membayar kredit [6]. Algoritme genetika diusulkan dalam [7] untuk diterapkan sebagai algoritme pencarian nilai parameter optimal sehingga meningkatkan akurasi klasifikasi terbaik pada *Support Vector Machine* (SVM). Algoritma *information gain* digunakan dalam [8] untuk menghitung bobot dari masing masing atribut dataset persetujuan kredit dengan klasifikasi menggunakan algoritme *K-Nearest Neighbour* (KNN). Prediksi tingkat kelancaran pembayaran kredit dengan menggunakan algoritme Naïve Bayes berbasis *Forward Selection* mampu memprediksi kelancaran pembayaran kredit ke depannya dalam [9]. Dalam [10], ditunjukkan bahwa SVM memiliki ketahanan dan kemampuan generalisasi yang lebih tinggi serta akurasi klasifikasi yang lebih stabil dibandingkan dengan algoritme yang lain.

Kajian beberapa algoritme yang sama untuk klasifikasi pemberian kredit belum dilakukan. Penelitian ini melakukan kajian penerapan algoritme Naïve Bayes, *Decision Tree* dan *SVM* pada klasifikasi data histori pembiayaan untuk membantu memprediksi calon nasabah koperasi syariah yang kredibel. Akurasi ketiga algoritme dalam klasifikasi pembiayaan nasabah dibandingkan. Hasil klasifikasi dapat digunakan oleh

*) Penulis korespondensi (Dwiza Riana)
Email: dwiza@nusamandiri.ac.id



Gambar 1. Kerangka penelitian

manajemen koperasi dalam membuat keputusan yang tepat untuk menyetujui atau menolak permintaan pinjaman atau pembiayaan dari nasabah.

II. METODE PENELITIAN

Metode penelitian yang digunakan mengacu pada *Cross-Standard Industry for Data Mining (CRISP-DM)* sebagai standarisasi pendekatan untuk data mining [11]. Tahapan dalam CRISP-DM antara lain *business understanding, data understanding, data preparation, modeling, evaluation* dan *deployment*. Alur pada penelitian ini diilustrasikan pada Gambar 1.

Penelitian ini menggunakan dataset pembiayaan [12] yang didapat dari salah satu koperasi syariah sebanyak 869 *record* selama bulan Januari 2015 sampai dengan September 2018 berlabel macet 291 *record* dan lancar 578 *record*. Dataset pembiayaan dibagi sebanyak 819 *record* (543 berlabel lancar dan 276 berlabel macet) sebagai masukan klasifikasi menggunakan *Rapidminer* dan 50 *record* (35 berlabel lancar dan berlabel 15 macet) sebagai *testing* pada sistem yang dihasilkan di tahap *deployment*.

Dataset pembiayaan yang telah dikumpulkan kemudian dilakukan penyeleksian menjadi 14 kriteria (Tabel 1). Dataset dilakukan klasifikasi menggunakan model Naïve Bayes (Algoritme 1), *decision tree* (Algoritme 2) dan SVM (Algoritme 3).

Pemodelan klasifikasi dataset pembiayaan menjadi kelas lancar dan macet menggunakan *Rapidminer* ditunjukkan pada Gambar 2. Evaluasi pengujian terhadap model untuk mendapatkan informasi model yang akurat. Evaluasi dan validasi menggunakan *confusion matrix, 10 k-fold cross validation* dan pengujian *T-Test*.

III. HASIL DAN PEMBAHASAN

Dataset pembiayaan sebanyak 819 *Record* digunakan sebagai masukan untuk klasifikasi algoritme Naïve Bayes, *Decision Tree* dan SVM menggunakan *Rapidminer* (Gambar 2). Hasil dari pengujian model

Algoritme 1. Naïve Bayes

$$1: \text{Persamaan Naïve Bayes: } P(y|X) = \frac{P(X|y)P(y)}{P(X)}$$

dengan P menunjukkan peluang, y variabel kelas (lancar/macet), X mewakili parameter / fitur.

2: X diberikan sebagai $X = (x_1, x_2, x_3, \dots, x_n)$ yang mewakili fitur yang ada di Tabel 1.

3: Dengan substitusi X, *rule* yang didapat selanjutnya untuk dimasukkan nilai dataset

$$P(y|x_1, \dots, x_n) = \frac{P(x_1 \vee y)P(x_2 \vee y) \dots P(x_n \vee y)P(y)}{P(x_1)P(x_2) \dots P(x_n)}$$

Algoritme 2. Decision Tree

1: *Data training*, berjumlah 819 *record* (543 berlabel lancar dan 276 berlabel macet)

2: Nilai *entropy* diperoleh dengan persamaan

$$\text{Entropy}(S) = \sum_{i=0}^n -p_i \cdot \log_2 p_i = (-543/819 \cdot \log_2 (543/819)) + (-276/819 \cdot \log_2 (276/819))$$

3: menghitung nilai *entropy* per-atribut

4: menghitung nilai *gain* untuk tiap atribut dan menentukan nilai *gain* tertinggi. Atribut dengan nilai *gain* tertinggi menjadi akar dari *tree* (pohon), yaitu

$$\text{Gain}(S, A) = \text{Entropy}(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} \times \text{Entropy}(S_i)$$

dengan S (himpunan kasus), n (jumlah partisi atribut A), pi (jumlah kasus pada partisi ke-i), |Si| (jumlah kasus pada partisi ke-i), dan |S| (jumlah kasus dalam S).

Algoritme 3. SVM

1: Titik data : $\{x_1, x_2, \dots, x_n\}$, merupakan atribut pada Tabel 1.

2: Kelas data : $y_i \{-1, +1\}$, -1 sebagai kelas macet dan +1 sebagai lancar.

3: Pasangan data dan kelas : $\{(x_i, y_i)\}_{i=1}^N$, N adalah banyaknya data.

4: Memaksimalkan fungsi

$$Ld = \sum_{i=1}^N a_i - \sum_{i=1}^N \sum_{j=1}^N a_i a_j y_i y_j K(x_i, x_j)$$

$$\text{syarat } 0 \leq a_i \leq C \text{ dan } \sum_{i=1}^N a_i y_i = 0$$

dengan Ld (dualitas langrange multiplier), a_i (nilai bobot setiap titik data), dan C (nilai konstanta)

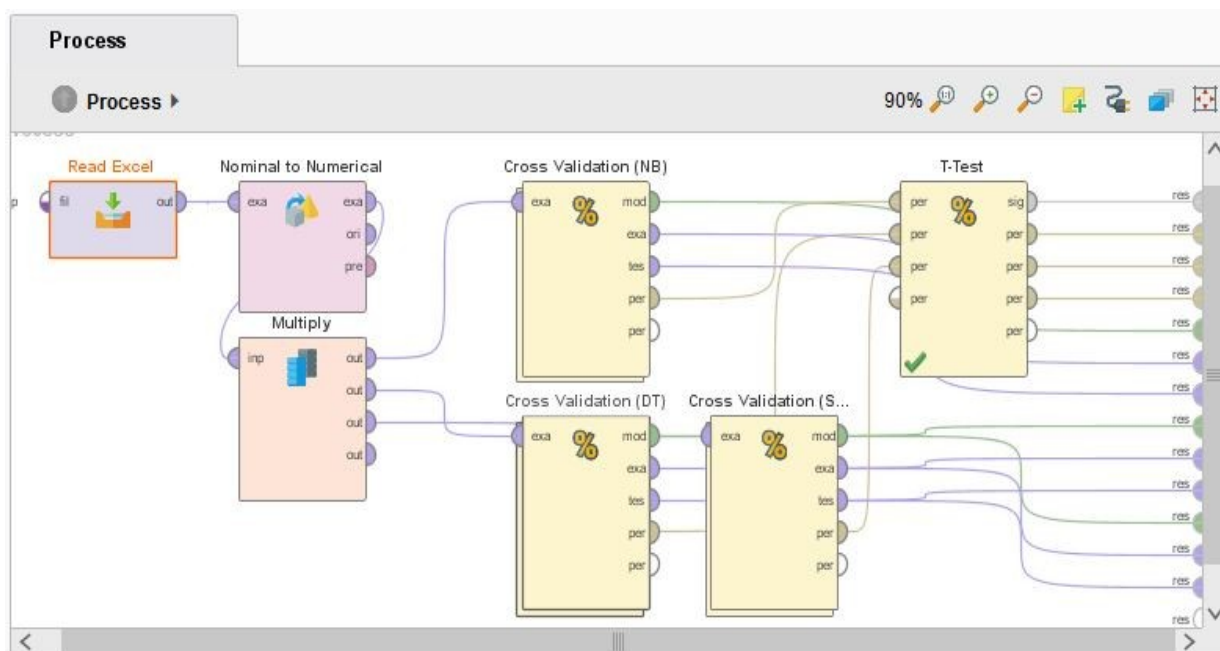
5: Menghitung nilai w dan b

$$w = \sum_{i=1}^N a_i y_i x_i \text{ dan } b = \frac{-1}{2}(w \cdot x^1 + w \cdot x)$$

6: Fungsi keputusan klasifikasi *sign* (f(x)):

$$f(x) = w \cdot x + b \text{ atau } f(x) = \sum_{i=1}^m a_i y_i K(x, x_i) + b$$

dengan m (jumlah titik data yang memiliki $a_i > 0$), $K(x, x_i)$ (fungsi kernel)



Gambar 2. Proses *modelling* klasifikasi dataset

Tabel 1. Data atribut

No	Atribut	Tipe	Keterangan
1	Jenis Kelamin	Binominal	P,W
2	Umur	Polinomial	<25, 25-50, >50
3	Status	Polinomial	Menikah, Lajang, Janda/Duda
4	Status Tempat Tinggal	Binominal	Sewa, Hak Milik
5	Jumlah Tanggungan	Polinomial	0, 1, 2-3, >3
6	Pendidikan	Polinomial	SD,SMP, SMA, Diploma, Sarjana
7	Pekerjaan	Polinomial	Pedagang, Karyawan
8	Penghasilan	Polinomial	<3jt, 3jt-5jt, >5jt
9	Flafond	Polinomial	<3jt, 3jt-5jt, >5jt-10jt, >10jt-50jt, >50jt
10	Jenis Pinjaman	Polinomial	Musyarakah, Murabahah
11	Jangka Waktu	Polinomial	Pendek, Menengah, Panjang
12	Cara Pembayaran	Polinomial	Harian, Mingguan, Bulanan
13	Jaminan	Polinomial	Elektronik, BPKB Motor, BPKB Mobil, Tabungan, Sertifikat
14	Keterangan	Binominal	Lancar, Macet

klasifikasi algoritme Naïve Bayes, *Decision Tree* dan SVM untuk menentukan nilai *accuracy*, *precision*, *recall* dan AUC. Validasi model menggunakan 10 *fold cross validation*.

Algoritme Naïve Bayes dalam klasifikasi pembiayaan nasabah koperasi menghasilkan presisi *class* lancar sebesar 90,29%, presisi *class* macet sebesar 61,97%, *recall class* lancar sebesar 73,66%, *recall class* macet sebesar 84,42% (Tabel 2), AUC sebesar 0,882 (Gambar 3) dan akurasi sebesar 77,29% (Tabel 5).

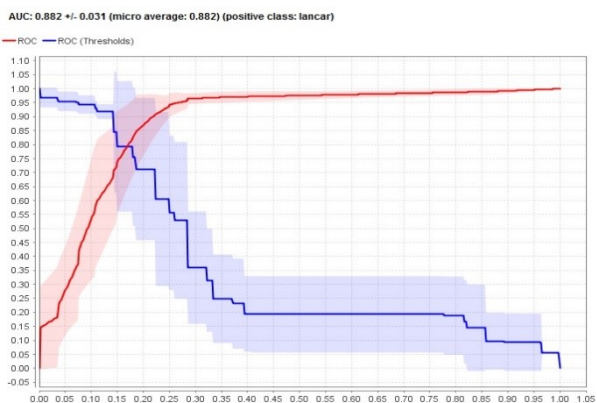
Penerapan algoritme *Decision Tree* dalam klasifikasi menghasilkan presisi *class* lancar sebesar 89,95%, presisi *class* macet sebesar 86,90%, *recall class* lancar sebesar 93,92%, *recall class* macet sebesar 79,35% (Tabel 3), AUC sebesar 0,882 (Gambar 4) dan akurasi sebesar 89,02% (Tabel 5).

Algoritma *Support Vector Machine* (SVM) sangat baik diterapkan dalam penelitian ini dengan hasil presisi *class* lancar sebesar 88,46%, presisi *class* macet sebesar 93,67%, *recall class* lancar sebesar 97,42%, *recall class* macet sebesar 75,00% (Tabel 4), AUC sebesar 0,939 (Gambar 5) dan akurasi sebesar 89,86% (Tabel 5).

Penelitian ini menyertakan atribut individu nasabah seperti status (menikah, lajang, janda/duda), status tempat tinggal, jumlah tanggungan, pendidikan yang tidak terdapat pada [6], [7], dan [9], sehingga menghasilkan akurasi yang lebih tinggi. Penerapan algoritme Naïve Bayes menghasilkan akurasi lebih tinggi dari [9] yang menerapkan *forward selection* pada algoritme Naïve Bayes dengan hasil akurasi sebesar 71,97%. Penerapan algoritme *decision tree* menghasilkan akurasi yang lebih tinggi dari hasil klasifikasi C4.5 [6] yang menghasilkan akurasi sebesar 88,90%. Penerapan algoritme SVM pada penelitian ini menghasilkan akurasi yang lebih tinggi dibanding [7] yang menerapkan algoritme genetika sebagai algoritme pencarian nilai parameter optimal pada SVM dengan hasil akurasi kernel *linear* 85,12%, *polynomial* 81,76%, RBF 77,22%, dan *Sigmoid* 78,70%.



Gambar 3. AUC algoritme Naive Bayes



Gambar 4. AUC algoritme Decision Tree



Gambar 5. AUC algoritme SVM

Dari ketiga algoritme yang digunakan, SVM menjadi algoritme yang paling tinggi tingkat akurasi. Sejalan dengan [10], SVM menghasilkan akurasi terbaik karena memiliki ketahanan dan generalisasi lebih tinggi dibanding algoritme lainnya termasuk algoritme C.45.

Pengujian *T-Test* (Tabel 6) menunjukkan perbedaan yang signifikan antara klasifikasi algoritme B sebesar Naive Bayes (77,29%) dengan C sebesar *Decision Tree* (89,02%) dan D sebesar SVM (89,86%). Algoritme *Decision Tree* dan SVM tidak terjadi perbedaan yang signifikan. Berdasarkan pengujian *T-test* tersebut, maka algoritme *Decision Tree* dan SVM dapat

Tabel 2. Confusion matrix algoritma Naive Bayes

	True Lancar	True Macet	Class precision
Pred. Lancar	400	43	90,29%
Pred. Macet	143	233	61,97%
Class recall	73,66%	84,42%	

Tabel 3. Confusion matrix Algoritma Decision Tree

	True Lancar	True Macet	Class precision
Pred. Lancar	510	57	89,95%
Pred. Macet	33	219	86,90%
Class recall	93,92%	79,35%	

Tabel 4. Confusion matrix algoritma SVM

	True Lancar	True Macet	Class precision
Pred. Lancar	529	69	88,46%
Pred. Macet	14	207	93,67%
Class recall	97,42%	75,00%	

Tabel 5. Pengujian algoritme Naive Bayes, Decision Tree dan SVM

	Naive Bayes	Decision Tree	SVM
Accuracy	77,29%	89,02%	89,86%
AUC	0,882	0,882	0,939

Tabel 6. Pengujian T-Test

	A	B	C	D
		0,773	0,890	0,899
0,773			0,000	0,000
0,890				0,481
0,899				

Tabel 7. Confusion matrix hasil testing Sistem Pendukung Persetujuan Pembiayaan (SPPP)

	True Lancar	True Macet	Class precision
Pred. Lancar	33	2	94,29%
Pred. Macet	1	14	93,33%
Class recall	97,06%	87,50%	

dipertimbangkan untuk diterapkan pada kasus klasifikasi persetujuan kredit dibandingkan algoritme Naive Bayes.

Berdasarkan evaluasi dari proses pengujian akurasi dan *T-test* terhadap model algoritme Naive Bayes, *Decision Tree* dan SVM, dapat dinyatakan bahwa kinerja terbaik dihasilkan oleh algoritme SVM. Dari hal tersebut, bobot dari *output* pengujian algoritme SVM pada Gambar 6 digunakan dalam pemodelan sistem pada tahap *deployment*. Tahap *deployment* menghasilkan sebuah sistem, yaitu Sistem Pendukung

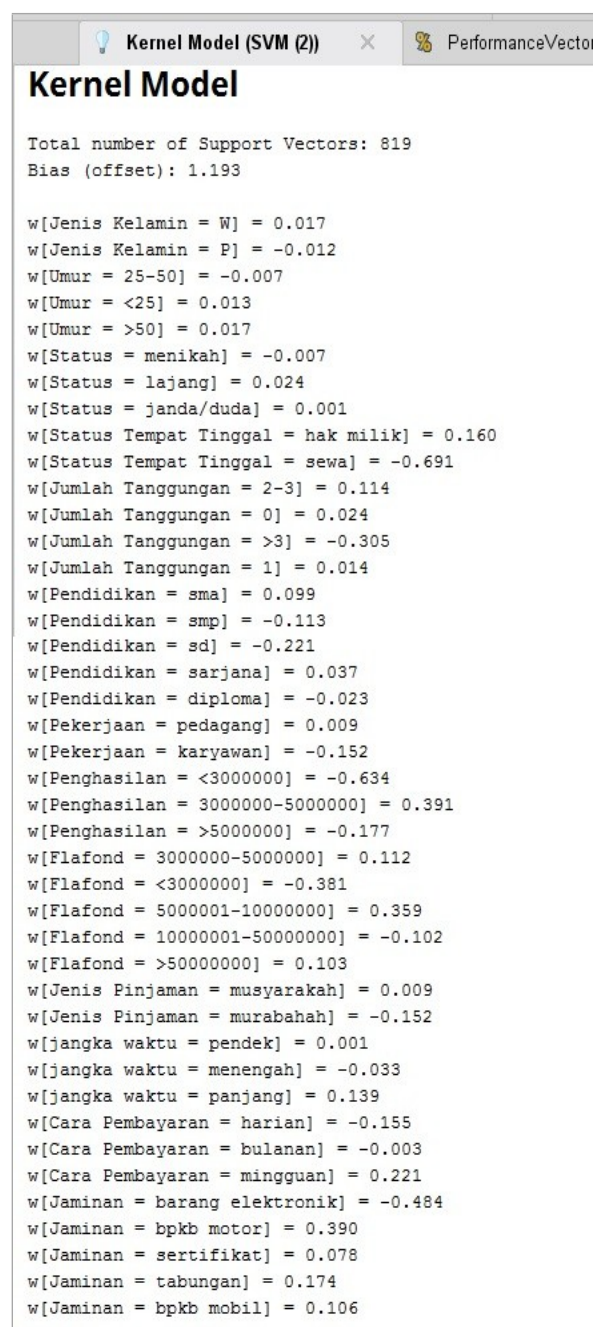
Persetujuan Pembiayaan (SPPP). Data *testing* yang digunakan dalam pengujian SPPP sebanyak 50 *record* dari dataset yang disisakan. Hasil prediksi sesuai sebanyak 47 *record* dan yang tidak sesuai sebanyak 3 *record* (Tabel 7) atau mempunyai akurasi 94%. SVM dapat diterapkan untuk prediksi tingkat kelancaran kredit dengan akurasi yang baik, selain [6] dengan C4.5 dan Naïve Bayes, [7] dengan algoritme genetika dan SVM, dan [8] dengan KNN.

IV. KESIMPULAN

SVM untuk klasifikasi data pembiayaan nasabah koperasi syariah ke dalam kelas macet dan lancar menghasilkan kinerja terbaik dengan akurasi 89,86% dan AUC 0,939 dibandingkan *Naïve Bayes* dan *Decision Tree*. Sistem Pendukung Persetujuan Pembiayaan (SPPP) menggunakan SVM menghasilkan akurasi klasifikasi sebesar 94% dari 50 *record* dataset.

DAFTAR PUSTAKA

- [1] R. A. Saraswati, "Peranan Analisis Laporan Keuangan, Penilaian Prinsip 5c Calon Debitur dan Pengawasan Kredit Terhadap Kreatifitas Pemberian Kredit," *Jurnal Nominal*, vol. I, no. 1, pp. 1–13, 2012.
- [2] N. Liu, E. J. Xia, and L. Yang, "Research and Application Of Pso-Bp Neural Networks in Credit Risk Assessment," in *2010 International Symposium on Computational Intelligence and Design*, Hangzhou, China, 2010, pp. 103–106.
- [3] S. A. Subekti, L. Linawati, and A. Setiawan, "Penggunaan Metode Fuzzy Mamdani untuk Membuat Keputusan Dalam Analisis Kredit," in *Seminar Nasional Matematika dan Pendidikan Matematika*, Yogyakarta, Indonesia, 2013, pp. 232–238.
- [4] M. Sudhakar and C. V. K. Reddy, "Two Step Credit Risk Assessment Model for Retail Bank Loan Applications Using Decision Tree Data Mining Technique," *International Journal of Advanced Research in Computer Engineering & Technology*, vol. 5, no. 3, pp. 705–718, 2016.
- [5] H. Jiawei, M. Kamber, J. Han, M. Kamber, and J. Pei, *Data Mining: Concepts and Techniques*. 2012.
- [6] S. Masripah, "Komparasi Algoritma Klasifikasi Data Mining untuk Evaluasi Pemberian Kredit," *Bina Insani ICT Journal*, vol. 3, no. 1, pp. 187–193, 2016.
- [7] J. Manurung, H. Mawengkang, and E. Zamzami, "Optimizing Support Vector Machine Parameters with Genetic Algorithm for Credit Risk Assessment," *Journal of Physics: Conference Series*, vol. 930, no. 1, 2017.
- [8] T. T. Chasanah and M. A. Al Karomi, "Sistem Pendukung Keputusan Persetujuan Kredit dengan Pemanfaatan Information Gain untuk Pembobotan



Gambar 6. Kernel model SVM

- Atribut Klasifikasi K-Nearest Neighbour," *Jurnal Ilmiah ICTech*, vol. 12, no. 2, pp. 14–19, 2017.
- [9] M. Hasan, "Prediksi Tingkat Kelancaran Pembayaran Kredit Bank Menggunakan Algoritma Naive Bayes Berbasis Berbasis Forward Selection," *ILKOM Jurnal Ilmiah*, vol. 9, no. 3, pp. 317–324, 2017.
 - [10] H. Yu, X. Huang, X. Hu, and H. Cai, "A Comparative Study on Data Mining Algorithms for Individual Credit Risk Evaluation," in *2010 International Conference on Management of e-Commerce and e-Government*, Chengdu, China, Oct. 2010, pp. 35–38.

- [11] M. North, *Data Mining for the Masses*. Global Text, 2012.
- [12] Nurajjah, D. Riana, "Dataset Pembiayaan", *Repository Nusamandiri*, 30 April 2019,

[Online]. Available: <https://repository.nusamandiri.ac.id/index.php/repo/viewitem/18584>