

Klasifikasi Status Kredit Nasabah Bank Menggunakan Teknik *Machine Learning: Support Vector Machine*

Kusnaeni¹, Miftahulhairah², Hartina Husain³, Muhammada Rifki Nisardi⁴

^{1,2,3,4}Jurusan sains, Institut Teknologi Bachruddin Jusuf Habibie, Parepare, Indonesia

¹kusnaeni25@ith.ac.id, ²miftahulhairah@ith.ac.id, ³hartinahusain@ith.ac.id, ⁴muhammadrifkinisardi@ith.ac.id

Abstract — Determining the credit status of customers in the banking sector has a crucial role. Therefore, precision and accuracy are required in classifying customer credit status. Machine learning techniques have become a popular solution for dealing with classification and clustering problems. One machine learning technique that is often used is support vector machine (SVM), which can classify objects using linear kernels. In this research, SVM is applied to separate customer data into good and bad classes. The classification results using SVM achieved an accuracy of 72%.

Keywords— Machine learning, Support Vector Machine, Classification

Abstrak — Penentuan status kredit pada nasabah di sektor perbankan memiliki peran krusial. Oleh karena itu, diperlukan ketelitian dan keakuratan dalam mengklasifikasi status kredit nasabah. Teknik *machine learning* menjadi solusi yang populer dalam menangani permasalahan klasifikasi dan klustering. Salah satu teknik *machine learning* yang sering digunakan adalah *support vector machine* (SVM), yang dapat mengklasifikasi objek dengan menggunakan kernel linear. Pada penelitian ini, SVM diterapkan untuk memisahkan data nasabah menjadi kelas good dan bad. Hasil klasifikasi menggunakan svm mencapai akurasi sebesar 72%.

Kata Kunci — Machine Learning, Support Vector Machine, Klasifikasi.

I. PENDAHULUAN

Penelitian ini difokuskan pada masalah klasifikasi dalam industri perbankan, yang melibatkan pengelompokan nasabah berdasarkan karakteristik mereka untuk menentukan status kredit nasabah di bank. Machine learning, sebagai teknik populer dalam klasifikasi dan klustering, menawarkan solusi potensial untuk meningkatkan efisiensi dan ketepatan dalam manajemen risiko di sektor perbankan. Salah satu metode klasifikasi yang umum digunakan dalam machine learning adalah Support Vector Machine (SVM). SVM dapat digunakan untuk melakukan prediksi, baik dalam kasus klasifikasi maupun regresi. Prinsip dasar dari SVM adalah linier classifier yang memungkinkan pemisahan linier antar kelompok data. Fungsi kernel dalam SVM berperan dalam memetakan data dari dimensi awal ke dimensi baru, memungkinkan penanganan data yang tidak linier.

Penelitian ini bertujuan untuk mengaplikasikan teknik machine learning, khususnya SVM, dalam mengklasifikasi nasabah bank berdasarkan karakteristik mereka. Klasifikasi ini diharapkan dapat membedakan

nasabah antara kelompok status kredit "Good" dan status kredit "Bad", memberikan dasar untuk perumusan kebijakan-kebijakan selanjutnya. Dengan fokus pada sektor perbankan, penelitian ini berusaha memberikan kontribusi pada pengembangan model klasifikasi yang efektif untuk manajemen risiko. Melalui penggabungan konsep SVM dan karakteristik nasabah, diharapkan penelitian ini dapat memberikan wawasan baru dan solusi yang dapat meningkatkan efisiensi serta ketepatan dalam pengelolaan risiko di industri perbankan. Hasil dari penelitian ini diharapkan dapat memberikan dasar yang kuat untuk perumusan kebijakan risiko yang lebih baik dalam industri perbankan.

II. METODOLOGI PENELITIAN

A. Data dan Variabel

Penelitian ini menggunakan data sekunder yang diperoleh dari laman: <https://archive.ics.uci.edu/dataset/144/statlog+german+credit+data> yang terdiri dari 1000 observasi dan 21 variabel. Variabel respon yang diamati adalah status kredit nasabah di negara Jerman. Variabel dalam penelitian dapat dilihat pada Tabel 1.

TABEL I
VARIABEL PENELITIAN

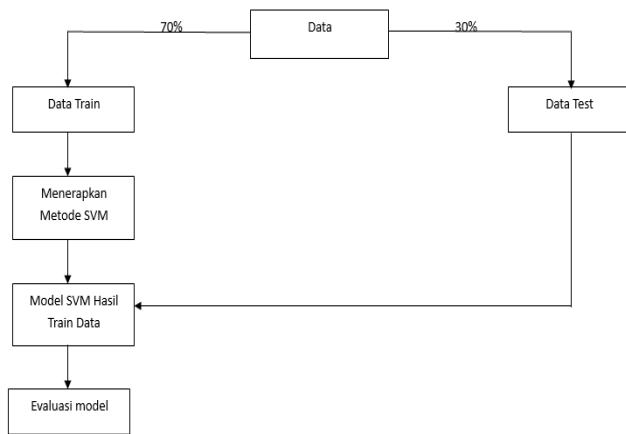
Variabel	Deskripsi
X_1	Status rekening giro yang ada
X_2	Durasi
X_3	Treking kredit
X_4	Tujuan Peminjaman
X_5	Jumlah kredit
X_6	Rekening tabungan/obligasi
X_7	Durasi Lama Kerja
X_8	Tingkat angsuran dalam persentase pendapatan yang dapat dibelanjakan
X_9	Status pernikahan dan jenis kelamin
X_{10}	Debitur/penjamin lainnya
X_{11}	Durasi Lama Menetap di Rumah yang bersangkutan
X_{12}	Properti
X_{13}	Usia
X_{14}	Paket cicilan lainnya
X_{15}	Status Rumah
X_{16}	Jumlah kredit yang ada di bank yang bersangkutan
X_{17}	Pekerjaan

X_{18}	Jumlah orang yang bertanggung jawab untuk menyediakan penjaminan
X_{19}	Telephone
X_{20}	Pekerja asing
Y	1 = Good, -1 = Bad

Pada Tabel 1 terlihat ada 20 variabel prediktor dan 1 variabel respon Y yang terdiri dari 2 kelas, Dimana kelas 1 menandakan status kredit nasabah Good dan kelas -1 menandakan status kredit nasabah Bad.

B. Metode dan Tahapan Analisis

Tahapan analisis yang dilakukan dalam penelitian ini digambarkan pada diagram alur sebagai berikut:



Gambar 1. Diagram alur tahapan analisis

1) Splitting Data

Penelitian ini menggunakan dua set data utama: data training dan data testing. Data training digunakan untuk memodelkan Support Vector Machine (SVM), sedangkan data testing digunakan untuk menguji akurasi model SVM yang telah dibuat. Pembagian data ini mengikuti komposisi sebesar 70% untuk data training dan 30% untuk data testing. Terdapat 700 data yang digunakan sebagai data training dan 300 data sebagai data testing. Proses ini memastikan bahwa model SVM dapat diuji dan dievaluasi dengan baik menggunakan dataset yang tidak digunakan selama proses pelatihan.

2) Menerapkan metode Support Vector Machine (SVM)

Metode machine learning untuk mengklasifikasikan status kredit nasabah di bank, dengan menggunakan algoritma Support Vector Machine (SVM). SVM dirancang untuk mencari hyperplane maksimal, yaitu fungsi matematis yang dapat efektif memisahkan dua kelas data. Fokus utama SVM adalah menciptakan hyperplane dengan jarak maksimal dari sampel terdekat dari masing-masing kelas, sehingga memungkinkan pemisahan yang optimal [1]. Dalam proses Support Vector Machine (SVM), tujuannya adalah memaksimalkan margin atau jarak antara pola pelatihan dan batas keputusan [2]. Algoritma ini memiliki beberapa keunggulan, termasuk kemampuan unggul dalam kinerja baik pada dataset kecil maupun besar, performa yang

optimal pada data dengan jumlah atribut yang banyak, serta kemudahan dalam implementasinya. Dalam menerapkan konsep SVM pada data, terdapat beberapa atribut penting yang perlu diperhatikan. Salah satunya adalah Kernel, yang merupakan fungsi yang digunakan untuk memetakan data ke dalam dimensi yang lebih tinggi. Proses ini memungkinkan data menjadi lebih mudah dipisahkan. hyperplane, sebagai batas keputusan, memisahkan kelas pada ruang fitur. Pada SVM linier, hyperplane yang terbentuk adalah garis, sedangkan pada SVM non-linier, hyperplane dapat berupa permukaan yang kompleks [3]. Atribut lainnya adalah Support Vector, yaitu data yang berada paling dekat dengan hyperplane dan memiliki pengaruh signifikan terhadap lokasi hyperplane tersebut. Margin, yang merupakan jarak antara hyperplane dan Support Vector yang terdekat, harus dioptimalkan agar kelas dapat dipisahkan dengan jarak yang optimal [4]. Dengan memahami dan memanfaatkan atribut-atribut ini, konsep SVM dapat diterapkan dengan lebih baik untuk membangun model klasifikasi yang efektif.

Misalkan m data training $\{(x_1, y_1), (x_2, y_2), \dots, (x_m, y_m)\}$ dimana $x_i \in R^n$ adalah sampel data dan $y_i \in \{1, -1\}$ adalah target atau kelas dari sampel data. Misalkan

juga bahwa data untuk kedua kelas terpisah secara linier (linearly separable) maka ingin dicari fungsi pemisah (hyperplane)

$$f(x) = xw + b = 0 \tag{1}$$

Dimana $w \in R^{n \times 1}$ adalah parameter bobot dan $b \in R$ adalah parameter bias, serta berlaku:

$$x_i w + b > 0 \text{ untuk } y_i = 1$$

$$x_i w + b < 0 \text{ untuk } y_i = -1$$

Misalnya terdapat hyperplane $H: xw + b = 0$ yang ingin dicari, dengan $H_1: xw + b = 1$ dan $H_2: xw + b = -1$ sebagai hyperplane dari kelas 1 dan kelas -1. Untuk mencari hyperplane H yang optimal, syaratnya adalah jarak antara H_1 dan H_2 harus sama, dan tidak ada sampel data yang terletak di antara H_1 dan H_2 , serta jarak antara H_1 dan H_2 harus maksimal. Dalam upaya memaksimalkan jarak antara H_1 dan H_2 , digunakan sampel data positif yang berada pada H_1 dan sampel data negatif yang berada pada H_2 . Sampel data ini disebut sebagai support vector karena perannya yang krusial dalam menentukan hyperplane optimal. Sampel data yang tidak termasuk dalam kategori support vector dapat diabaikan atau dipindahkan menuju H_1 dan H_2 , selama tidak melampaui masing-masing hyperplane [5]. Dengan memahami konsep ini, kita dapat mencari hyperplane optimal dalam konteks Support Vector Machine (SVM).

3) Model SVM hasil train data

Setelah menerapkan metode SVM pada data latih, kita akan mendapatkan sebuah model SVM yang dihasilkan dari proses pelatihan tersebut. Model ini kemudian digunakan untuk melakukan klasifikasi pada data uji yang sebelumnya telah dipisahkan dengan presentase 30% dari seluruh data. Hasil dari klasifikasi menggunakan data uji dengan metode SVM disebut sebagai hasil prediksi. Prediksi ini akan mencakup nilai variabel respon Y yang akan dibandingkan dengan nilai aktual dari variabel respon Y. Perbandingan hasil prediksi dengan nilai aktual ini akan menjadi dasar untuk mengukur akurasi dari model SVM. Evaluasi kinerja secara keseluruhan terhadap model SVM dapat diperoleh melalui analisis hasil perbandingan tersebut.

4) Evaluasi Model

Evaluasi kinerja dilakukan dengan memeriksa sejauh mana model yang dihasilkan oleh algoritma mampu berperforma. *Confusion Matrix* digunakan sebagai parameter utama untuk mengevaluasi model tersebut. Dalam fase ini, pengujian akan menghasilkan nilai akurasi, *sensitivity*, dan *specificity*. Akurasi mencerminkan proporsi prediksi yang benar, baik yang bersifat positif maupun negatif, dari seluruh data. Akurasi dapat dihitung menggunakan Persamaan (2). *Sensitivity* atau *Recall* mencerminkan proporsi prediksi positif yang benar dari seluruh data yang sebenarnya positif. *Sensitivity* dapat dihitung menggunakan Persamaan (3). *Specificity* mencerminkan proporsi prediksi negatif yang benar dari seluruh data yang sebenarnya negatif, dan dapat dihitung menggunakan Persamaan (4). Melalui evaluasi ini, dapat diperoleh pemahaman yang lebih mendalam tentang seberapa baik model mampu memprediksi dengan benar dan seberapa baik kinerja model dalam mengidentifikasi kelas tertentu, baik yang positif maupun negatif [6].

$$akurasi = \frac{Tp + Tn}{Tp + Fp + Fn + Tn} \times 100\% \quad (2)$$

$$sensitivity = \frac{Tp}{Tp + Fn} \times 100\% \quad (3)$$

$$specificity = \frac{Tn}{Tn + Fp} \times 100\% \quad (4)$$

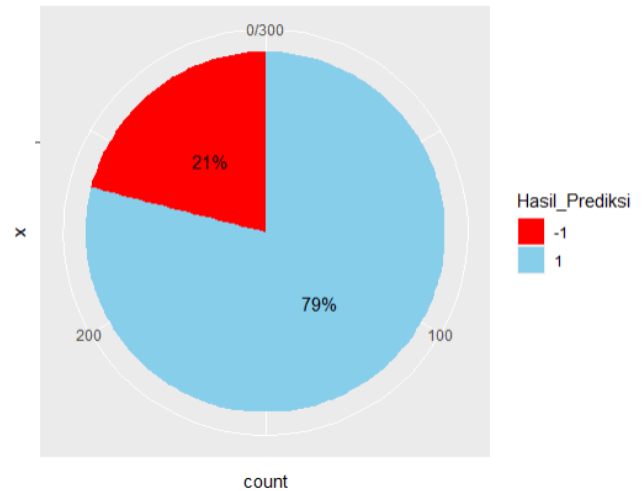
III. HASIL DAN PEMBAHASAN

A. Klasifikasi menggunakan Support Vector Machine

Analisis klasifikasi dilakukan setelah model diperoleh, dengan tujuan mengevaluasi kinerja model dalam mengklasifikasikan status kredit nasabah bank ke dalam dua kategori: "Good" dan "Bad". Tahap awal penelitian melibatkan pembagian data menjadi dua bagian, yakni data training yang digunakan untuk membentuk model Support Vector Machine, dan data testing yang digunakan untuk mengevaluasi kinerjanya. Pembagian data ini dilakukan dengan persentase 70% untuk data training dan

30% untuk data testing, dari total 1000 data yang tersedia. Setelah pembagian data, dilanjutkan dengan penerapan klasifikasi status kredit nasabah bank menggunakan metode Support Vector Machine dengan kernel linier.

PREDIKSI PERSENTASE STATUS NASABAH

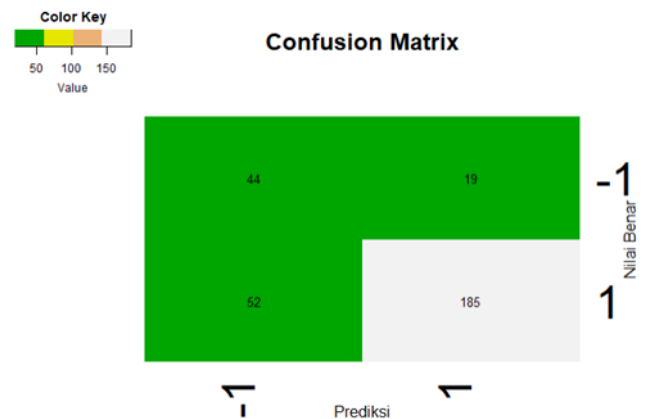


Hasil klasifikasi menggunakan SVM disajikan melalui pie chart, yang menunjukkan bahwa 79% dari total 100 nasabah tergolong dalam kategori "Good" dan 21% tergolong dalam kategori "Bad". Analisis ini memberikan gambaran visual mengenai distribusi status kredit nasabah berdasarkan model SVM yang telah dikembangkan.

Gambar 2. Diagram hasil prediksi status nasabah

B. Evaluasi Model

Dalam penelitian ini, tahap evaluasi model dilakukan menggunakan *confusion matrix*. *Confusion matrix* menyajikan informasi yang membandingkan hasil klasifikasi yang diberikan oleh sistem dengan hasil klasifikasi yang seharusnya. Dalam evaluasi kinerja model, terdapat tiga istilah utama dalam *confusion matrix*, yaitu *True Positive* (TP), *True Negative* (TN), *False Positive* (FP), dan *False Negative* (FN). Gambar 1 menunjukkan *confusion matrix* yang digunakan dalam penelitian ini. *Confusion matrix* membantu memberikan gambaran yang jelas tentang sejauh mana model mampu mengklasifikasikan data dengan benar atau memberikan prediksi yang tepat.



Gambar 3. Confusion Matrix

Dari *confusion matrix* yang telah diperoleh, langkah selanjutnya adalah menghitung nilai akurasi, *sensitivity*, dan *specificity* menggunakan rumus yang sesuai. Akurasi dihitung dengan menggunakan Persamaan 2, *sensitivity* dengan Persamaan 3, dan *specificity* dengan Persamaan 4. Proses perhitungan ini memungkinkan kita untuk mendapatkan informasi lebih rinci mengenai kinerja model. Akurasi memberikan gambaran umum tentang sejauh mana model mampu memberikan prediksi yang benar, sedangkan *sensitivity* menunjukkan kemampuan model dalam mengidentifikasi dengan benar data yang seharusnya masuk ke dalam kategori positif. Di sisi lain, *specificity* mengukur sejauh mana model dapat mengidentifikasi dengan benar data yang seharusnya masuk ke dalam kategori negatif. Dengan menggunakan rumus-rumus tersebut, evaluasi kinerja model menjadi lebih terperinci dan memungkinkan untuk mengidentifikasi kelebihan dan kekurangan dari model klasifikasi yang telah dikembangkan.

$$\text{akurasi} = \frac{229}{300} \times 100\% = 76.33\%$$

$$\text{sensitivity} = \frac{185}{204} \times 100\% = 90.68\%$$

$$\text{specificity} = \frac{44}{96} \times 100\% = 45.83\%$$

Nilai akurasi yang diperoleh sebesar 76.33% menunjukkan bahwa model klasifikasi mampu memprediksi kelas yang benar dari data *testing* sebesar 76.33%. *Sensitivity*, dengan nilai sebesar 90.68%, mengindikasikan bahwa model dapat memprediksi dengan benar dari semua data positif yang sebenarnya. Sementara itu, nilai *specificity* sebesar 45.83% menunjukkan perbandingan prediksi negatif yang benar dengan seluruh data negatif yang sebenarnya. Evaluasi ini memberikan gambaran lebih detail mengenai performa model, dengan akurasi yang mencerminkan tingkat keberhasilan umum, *sensitivity* yang menyoroti keberhasilan model dalam mengidentifikasi data positif, dan *specificity* yang menggambarkan kemampuan model dalam mengidentifikasi data negatif. Interpretasi nilai-nilai ini membantu dalam mengevaluasi sejauh mana model mampu memberikan prediksi yang akurat dan tepat.

IV. KESIMPULAN

Dengan menggunakan metode *machine learning Support Vector Machine* pada data *training* sebesar 70% dan data *testing* sebesar 30%, diperoleh nilai akurasi sebesar 76.33%, *sensitivity* sebesar 90.68%, dan *specificity* sebesar 45.83%. Hasil klasifikasi menggunakan metode *Support Vector Machine* pada data karakteristik nasabah dari

<https://archive.ics.uci.edu/dataset/144/statlog+german+cre>

[dit+data](#) menunjukkan bahwa 79% dari nasabah diprediksi masuk ke dalam kategori status kredit "Good" sementara 21% diprediksi masuk ke dalam kategori status kredit "Bad" Evaluasi ini memberikan gambaran komprehensif tentang kinerja model dalam memprediksi status kredit nasabah berdasarkan data karakteristik yang diberikan.

DAFTAR ACUAN

- [1] A. Wenda, "Support Vector Machine untuk Pengenalan Bentuk Manusia Menggunakan Kumpulan Fitur yang Dioptimalkan," vol. 11, no. 1, pp. 77–84, 2022.
- [2] P. A. Octaviani, Y. Wilandari, and D. Ispriyanti, "PENERAPAN METODE KLASIFIKASI SUPPORT VECTOR MACHINE (SVM) PADA DATA AKREDITASI SEKOLAH DASAR (SD) DI KABUPATEN MAGELANG," *J. GAUSSIAN*, vol. 3, pp. 811–820, 2014.
- [3] F. Abdusyukur, "ALGORITMA SUPPORT VECTOR MACHINE (SVM) UNTUK KLASIFIKASI PENCEMARAN NAMA BAIK KOMPUTA: Jurnal Ilmiah Komputer dan Informatika," *KOMPUTA J. Ilm. Komput. dan Inform. PENERAPAN*, vol. 12, no. 1, 2023.
- [4] R. K. Putri and M. Athoillah, "SUPPORT VECTOR MACHINE UNTUK IDENTIFIKASI BERITA HOAX TERKAIT VIRUS CORONA (COVID-19)," *J. Inform. J. Pengemb. IT*, vol. 6, no. 3, pp. 162–167, 2021.
- [5] R. Dasmasele, B. . Tomasouw, and Z. . Leleury, "PENERAPAN METODE SUPPORT VECTOR MACHINE (SVM) UNTUK MENDETEKSI PENYALAHGUNAAN NARKOBA," *J. Mat. Stat. DAN Ter.*, vol. 01, no. 02, pp. 111–122, 2022.
- [6] O. V. Putra, M. Z. Mustaqim, and D. Muriatmoko, "Transfer Learning untuk Klasifikasi Penyakit dan Hama Padi Menggunakan MobileNetV2," *Techno.COM*, vol. 22, no. 3, pp. 562–575, 2023.