

# Penerapan Fitur Seleksi dan Particle Swarm Optimization pada Algoritma Support Vector Machine untuk Analisis Credit Scoring

Abdul Razak Naufal, Akrim Teguh Suseno\*

Fakultas Sains dan Teknologi, Teknologi Komputer, ITSNU Pekalongan, Pekalongan, Indonesia

Email: <sup>1</sup>naufal@itsnupekalongan.ac.id, <sup>2,\*</sup>akrim@itsnupekalongan.ac.id

Email Penulis Korespondensi: akrim@itsnupekalongan.ac.id

Submitted: 12/10/2023; Accepted: 29/11/2023; Published: 30/11/2023

**Abstrak**—Pasca pandemi covid-19 lembaga-lembaga keuangan terutama lembaga keuangan mikro menghadapi kendala yang besar dalam berkontribusi terhadap peningkatan perekonomian baik ditingkat daerah maupun ditingkat nasional. Ratusan juta orang yang berpenghasilan rendah tidak memiliki kredit atau rekening bank karena mereka tidak memiliki riwayat kredit yang memadai untuk mendapatkan nilai kredit yang diberikan kepada mereka. Diperkirakan 1,7 miliar orang (31% dari populasi orang dewasa) tidak memiliki rekening di lembaga keuangan. Orang dewasa ini biasanya terkonsentrasi di negara-negara berkembang, khususnya di China 204 jt, India 357 juta dan Indonesia 102 juta orang. Oleh karena sangat sulit untuk melakukan prediksi yang akurat dalam menentukan kelayakan kredit kepada masyarakat berpenghasilan rendah. Koperasi adalah lembaga keuangan yang memiliki peran yang krusial dalam menyalurkan pembiayaan kepada anggota maupun masyarakat untuk mengembangkan usahanya. Proses penyaluran kredit yang tidak tepat dapat berpotensi negatif pada Lembaga keuangan tersebut, sehingga dapat mengalami resiko gagal bayar yang akibatnya koperasi jadi merugi. Risiko seperti ini biasa dikenal dengan *problem loans*, penyebabnya adalah lembaga keuangan gagal dalam melakukan analisis kredit kepada calon debiturnya. Maka dari itu diperlukan analisis perhitungan untuk meramalkan potensi gagal bayar atau *problem loans* tidak terjadi lagi. Kelayakan kredit atau *Credit scoring* adalah metode yang biasa dimanfaatkan untuk mengevaluasi risiko kredit yang berpotensi macet atau gagal bayar. Dalam penelitian ini kami akan memberi solusi menggunakan teknik klasifikasi dengan metode fitur seleksi untuk Algoritma *Particle Swarm Optimization* (PSO) pada Algoritma *Support Vector Machine* (SVM) untuk memprediksi risiko kredit calon debitur gagal dalam melakukan pembayaran pinjaman. Penerapan Algoritma SVM dalam penelitian *credit scoring* ini dikarenakan SVM bagus dalam klasifikasi data. Akan tetapi model SVM standar masih kurang menghasilkan hasil yang optimal dikarenakan susahnya menentukan parameter terbaik, oleh karena itu peneliti akan mengoptimasinya dengan algoritma *Feature Selection* dan PSO untuk menentukan parameter terbaik. Hasil penelitian dari kombinasi beberapa parameter menggunakan PSO-SVM ini diperoleh akurasi sebesar 87.23%, oleh karena itu penerapan metode ini terbukti dapat meningkatkan kinerja pada algoritma SVM untuk meningkatkan hasil akurasi dalam memprediksi kelayakan pemberian kredit.

**Kata Kunci:** Feature Selection; PSO; SVM; credit scoring; koperasi

**Abstract**—After the Covid-19 pandemic, the banking sector faced significant challenges in contributing to achieving national goals in terms of increasing living standards and equalizing the regional economy. Hundreds of millions of low-income people have no credit or bank accounts because they do not have sufficient credit history to warrant the credit scores assigned to them. An estimated 1.7 billion people (31% of the adult population) do not have an account with a financial institution. People today are usually concentrated in developing countries, especially in China 204 million, India 357 million and Indonesia 102 million people. Because it is very difficult to make accurate predictions in determining credit worthiness for low-income people. Cooperatives are financial institutions that have a crucial role in channeling financing to members and the community to develop their businesses. An inappropriate credit distribution process can have a negative effect on KSP, resulting in frequent losses. This risk is known as *problem loans*, the cause is the KSP's failure to analyze the credit of prospective debtors. Therefore, calculations are needed to detect opportunities for credit risk default by prospective debtors objectively and precisely so that loan problems do not occur. Credit scoring is a method used to evaluate credit risk in terms of loan applications from consumers [4]. In this research we will provide a solution using classification techniques with feature selection methods in the Particle Swarm Optimization (PSO) Algorithm and Support Vector Machine (SVM) to predict the credit risk of prospective debtors failing to make loan payments. The application of the SVM algorithm in credit scoring research is because SVM is good at data classification. However, the standard SVM model still does not produce optimal results due to the difficulty of determining the best parameters, therefore researchers will optimize it with the Feature Selection and PSO algorithms to determine the best parameters. The results from the combination of several parameters using PSO-SVM obtained an accuracy of 87.23%, therefore the application of this method was proven to improve the performance of the SVM algorithm to increase its accuracy results in predicting the feasibility of granting credit.

**Keywords:** Feature Selection; PSO; SVM; credit scoring; cooperative

## 1. PENDAHULUAN

Sejak Maret 2020 pasar keuangan mengalami guncangan hebat yang disebabkan oleh pandemi global Covid-19 [1]. Pasca pandemi covid-19 lembaga-lembaga keuangan terutama lembaga keuangan mikro menghadapi kendala yang besar dalam berkontribusi terhadap peningkatan perekonomian baik ditingkat daerah maupun ditingkat nasional [2]. Ratusan juta orang yang berpenghasilan rendah tidak memiliki kredit atau rekening bank karena mereka tidak memiliki riwayat kredit yang memadai untuk mendapatkan nilai kredit yang diberikan kepada mereka. Diperkirakan 1,7 miliar orang (31% dari populasi orang dewasa) tidak memiliki rekening di lembaga keuangan. Orang dewasa ini biasanya terkonsentrasi di negara-negara berkembang, khususnya di China

204 juta orang, India 357 juta orang dan Indonesia 102 juta orang [3]. Oleh karena itu sistem penilaian kredit yang akurat merupakan manajemen risiko yang penting bagi bank dan lembaga keuangan dalam beberapa dekade terakhir [4] guna mengurangi risiko gagal bayar.

Berdasarkan Undang-undang Nomor 10 Tahun 1998, kredit didasarkan pada penyediaan uang antara Lembaga keuangan dan pihak lain. Koperasi adalah lembaga keuangan yang mempunyai kontribusi penting dalam menumbuhkan ekonomi dan juga perdagangan, terutama dalam memberi bantuan pembiayaan kepada anggota atau Masyarakat untuk mengembangkan usahanya. Penyaluran kredit atau pemberian pinjaman mendominasi usaha lembaga keuangan terutama koperasi untuk membantu pemerintah dalam memakmurkan rakyatnya, kegiatan ini berdasarkan Permenkop dan UKM Nomor 15/Per/M.KUKM/2015 [5]. Pemberian pembiayaan atau pinjaman yang tidak tepat dapat berdampak pada keuangan lembaga keuangan, resiko ini sering disebut dengan kredit macet atau *problem loans*. Penyebab terjadinya kredit macet salah satunya adalah lembaga keuangan atau koperasi gagal dalam melakukan analisis kredit terhadap calon debiturnya, sehingga sangat diperlukan sekali sistem untuk mendeteksi kredit macet atau gagal bayar secara tepat supaya resiko kerugian dapat diminimalisir.

Prinsip kehati-hatian pada koperasi atau lembaga keuangan diantaranya kemampuan lembaga tersebut mampu dalam mengelola kelayakan kredit sehingga risiko terjadinya risiko kredit (*credit risk*) dapat monitoring dan dikontrol secara efektif. *Credit scoring* (skor kredit) adalah metode yang sering dipakai untuk mengevaluasi gagal bayar atau resiko kredit kepada *customer* [6]. Risiko kredit adalah risiko dampak paling signifikan bagi bank atau lembaga keuangan mana pun, penilaian risiko kredit yang akurat memengaruhi neraca dan laporan laba rugi organisasi, bahkan mungkin memengaruhi domain yang tampaknya tidak terkait, misalnya seperti pemasaran dan pengambilan keputusan [7], akibatnya pemberian kredit tanpa adanya metode prediksi yang akurat akan sangat berisiko terhadap calon debitur gagal dalam melakukan pembayaran kredit yang akhirnya lembaga keuangan akan merugi.

Tujuan dalam penelitian ini adalah untuk menghindari terbentuknya kredit macet, seseorang analisis kredit Lembaga keuangan terutama pada koperasi simpan pinjam wajib melakukan keputusan yang tepat dalam memberikan kredit maupun menolak kredit yang telah diajukan anggota. Untuk menemukenali kelayakan kredit ini di masa mendatang dibutuhkan sistem yang dapat memprediksi secara akurat. Salah satu metode yang dapat melakukan prediksi yang akurat ini adalah dengan teknik data mining. Teknik data mining merupakan teknik dalam mengekstraksi database yang besar untuk mencari pola-pola pelanggan, sehingga didapatkan prediksi yang akurat terhadap pemberian kredit sebagaimana yang telah dilakukan oleh beberapa peneliti terdahulu.

Pada tahun 2022, Riyadi *et al* melakukan penelitian dengan judul “*Analysis Of SVM And Naive Bayes Algorithm In Classification Of Nad Loans In Save And Loan Cooperatives*”, dalam penelitian tersebut membandingkan antara metode SVM dengan kernel linear, polynomial, RBF dan sigmoid dengan metode naïve bayes. Hasil yang diperoleh adalah naïve bayes memiliki nilai *accuracy* tertinggi dan SVM kernel RBF memiliki nilai F1, precision dan recall tertinggi yaitu 92,56%, sedangkan SVM dengan kernel sigmoid memiliki nilai tertinggi yaitu 95,44% untuk PoP kelas *true positive* dan PoA kelas *true negative* [8].

Penelitian yang dilakukan oleh Bumbungan pada tahun 2022 dengan Judul “Penerapan Particle Swarm Optimization (PSO) dalam Pemilihan Parameter Secara Otomatis pada Support Vector Machine (SVM) untuk Prediksi Kelulusan Mahasiswa Politeknik Amamapare Timika”. Hasil eksperimen dengan menggunakan algoritma Support Vector Machine (SVM) menghasilkan nilai *accuracy*, *recall*, *precision*, dan F1 *score* masing-masing dengan nilai 93,33%, 91,04%, 98,39%, dan 94,57%. Setelah algoritma SVM dilakukan optimasi PSO menghasilkan nilai *accuracy*, *recall*, *precision*, dan F1 *score* masing-masing dengan nilai 98,02%, 98,55%, 98,08%, dan 98,31% [9], hal ini terbukti bahwa penerapan PSO pada algoritma SVM dapat meningkatkan hasil akurasi.

Masih pada tahun 2022, penelitian yang dilakukan oleh Yao *et al* dengan judul “*A novel ensemble feature selection method by integrating multiple ranking information combined with an SVM ensemble model for enterprise credit risk prediction in the supply chain*” dalam penelitian Yao untuk memprediksi resiko kredit dalam perusahaan dengan menggunakan Fitur Seleksi dan SVM. Fitur seleksi digunakan untuk mengurangi kelemahan *class imbalance* pada metode SVM, yaitu lemah terhadap fitur yang berlebihan dan tidak relevan meningkatkan kesulitan pembelajaran model prediksi serta menyebabkan *overfitting* [10] sehingga dapat mempengaruhi prediksi. Hasil dalam eksperimen Yao mendapatkan kinerja prediksi yang baik yaitu AUC dan KS maksimum masing-masing sebesar 0,8772 dan 0,6363.

Dari beberapa hasil penelitian terdahulu, maka dalam penelitian ini *State of the art* nya adalah peneliti akan memberi solusi menggunakan teknik data mining dengan teknik klasifikasi dengan metode fitur seleksi dan optimasi parameter menggunakan algoritma *Particle Swarm Optimization* (PSO) pada algoritma *Support Vector Machine* (SVM) untuk prediksi kelayakan kredit. Penerapan Algoritma PSO-SVM dalam penelitian *credit scoring* ini dikarenakan SVM bagus dalam klasifikasi data. Model algoritma ini karena adanya kemampuan yang baik dalam menyamaratakan data juga mampu memecahkan masalah pada data dengan jumlah tertentu. Akan tetapi model SVM standar masih kurang menghasilkan hasil yang optimal, oleh karena itu peneliti akan mengoptimasinya dengan algoritma fitur seleksi dan PSO.

*Feature selection* adalah salah satu teknik ini mereduksi jumlah atribut atau fitur yang terlibat dalam menentukan nilai kelas atau label, serta mengurangi fitur yang tidak relevan [11] [12] dan data yang menyebabkan salah kelas target yang memberi efek negatif bagi aplikasi.

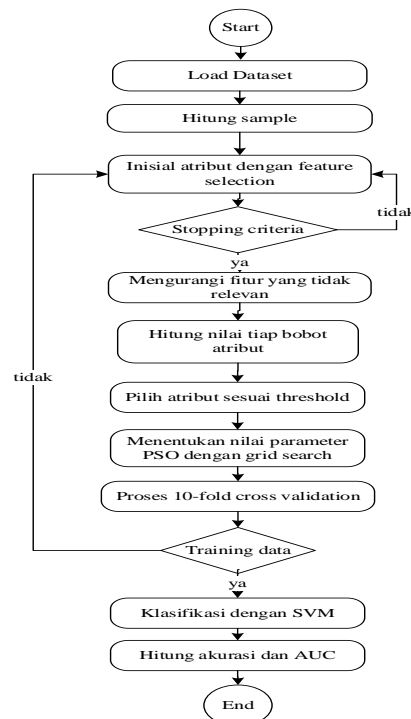
Algoritma PSO merupakan salah satu algoritma optimasi yang sering digunakan dalam Teknik data mining [13]. Algoritma PSO adalah algoritma optimisasi yang terinspirasi oleh perilaku kelompok dan gerakan koloni dalam alam. Algoritma ini digunakan untuk menemukan solusi terbaik dalam ruang pencarian parameter yang mungkin untuk masalah optimisasi. Cara kerja PSO mirip dengan perilaku kawanan burung atau sekumpulan partikel yang bergerak melalui ruang pencarian untuk mencari makanan atau tujuan tertentu. Setiap partikel dalam algoritma PSO mewakili solusi potensial untuk masalah optimisasi dan memiliki posisi serta kecepatan saat bergerak melalui ruang pencarian [5].

PSO telah digunakan dalam berbagai bidang, termasuk optimisasi parameter dalam machine learning, optimisasi perutean, optimisasi perancangan, dan masalah-masalah optimisasi kompleks lainnya. Keuntungan utama PSO adalah kemampuannya untuk mengeksplorasi ruang pencarian secara efisien dan menemukan solusi yang berkinerja baik dalam beragam masalah optimisasi. Algoritma PSO ini nanti yang akan peneliti gunakan dalam melakukan estimasi terhadap nilai parameter pada model SVM. Penentuan nilai parameter pada SVM terutama pada parameter  $\beta_i, i = 0, 1, \dots, k$ , yang bisa memberi nilai maksimal, sehingga dapat diperoleh nilai estimasi parameter model SVM berdasarkan algoritma PSO yang akan membantu meningkatkan akurasi pada algoritma SVM.

## 2. METODOLOGI PENELITIAN

### 2.1 Tahapan Penelitian

Eksperimen dalam penelitian ini adalah suatu kegiatan yang bertujuan untuk membuat kontribusi pengembangan terhadap ilmu pengetahuan [14]. Teknik Analisis data dalam penelitian ini adalah klasifikasi dengan pendekatan *machine learning*. Klasifikasi adalah pengelompokan data ke dalam beberapa kategori agar lebih mudah untuk diolah dan dianalisis. Model yang diusulkan dalam penelitian ini pada tahap *pre-processing* dengan menggunakan metode *feature selection* yang bertujuan untuk memberi bobot pada atribut atau fitur dalam dataset sehingga dapat menentukan atribut yang relevan dan meranking atribut yang paling berpengaruh terhadap kelasnya atau labelnya. Setelah dilakukan *training* data, untuk menentukan parameter digunakan metode *grid search* yang bisa menemukan nilai terbaik pada PSO-SVM. Semua tahapan yang sudah berjalan sesuai dengan metode yang diusulkan, maka tahap selanjutnya adalah mengukur *accuracy* model yang diusulkan dengan metode *confusion matrix*, setelah mendapatkan *accuracy* yang diharapkan kemudian model ini akan ditanam dalam sebuah aplikasi berbasis web. Pada Gambar 1 ditunjukkan alur metode yang diusulkan dalam penelitian ini menggunakan model *Feature Selection* pada Algoritma PSO – SVM.



Gambar 1. Diagram Model *Feature Selection* pada Algoritma PSO - SVM

Metode yang diusulkan dalam riset ini sebagaimana ditunjukkan pada Gambar 1, yaitu:

1. Pengumpulan Data

Tahap pertama dalam penelitian ini ialah mengumpulkan data. Pengumpulan data dilakukan di Lembaga Keuangan Mikro Syariah (LKMS) Kasuwari yang berlokasi di Kota Pekalongan. LKMS Kasuwari ini merupakan Lembaga keuangan yang berada dibawah naungan Otoritas Jasa Keuangan (OJK).

2. *Pre-processing Data*

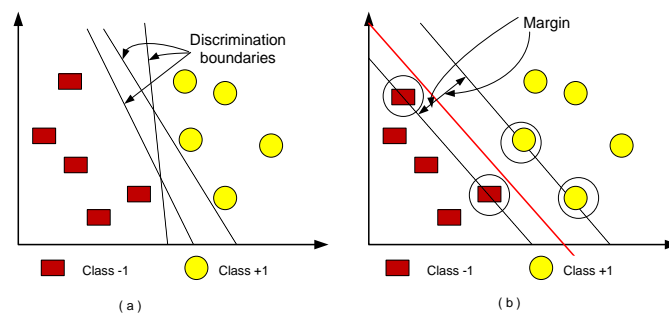
Tahap kedua adalah dengan melakukan *pre-processing* dengan data *sampling*. *Pre-processing* data adalah serangkaian langkah yang dilakukan untuk membersihkan, mentransformasi, dan menyiapkan data mentah sebelum data tersebut dapat digunakan untuk analisis atau pemodelan. *Pre-processing* data adalah tahap penting dalam analisis data, dan tujuannya adalah untuk memastikan bahwa data siap digunakan untuk mengambil wawasan atau membangun model yang akurat. Data hasil *sampling* difilter kembali menggunakan seleksi fitur dan kemudian atribut atau fitur diberi pembobotan. Setelah itu atribut-atribut itu dihitung nilainya, setelah semua atribut diketahui nilainya tahap selanjutnya ialah menentukan nilai *threshold* atau ambang batas antara atribut terpilih dengan atribut yang tidak terpilih.

3. Penentuan Parameter

Algoritma PSO akan digunakan untuk melakukan estimasi parameter yang optimal pada algoritma SVM, dengan cara memberikan nilai parameter yang tepat  $\beta_i, i = 0, 1, \dots, k$ , sehingga dapat dihasilkan nilai yang maksimum, serta diperoleh nilai estimasi parameter algoritma SVM pada PSO yang akan membantu meningkatkan akurasi pada algoritma SVM .

4. Evaluasi atau *Testing*

Algoritma SVM secara sederhana dapat dijelaskan sebagai usaha untuk mencari *hyperplane* terbaik yang bermanfaat sebagai pemisah antara dua buah kelas [15]. Gambar 2 menunjukkan *pattern* atau pola yang merupakan anggota dari dua *class*: +1 dan -1. Pola yang tergabung di kelas -1 dilambangkan dengan kotak berwarna merah, sedangkan pola pada *class* +1, dilambangkan dengan lingkaran warna kuning. Masalah *classification* yaitu mencari garis atau *hyperplane* yang telah memisahkan kelas itu. Garis pemisah (*discrimination boundaries*) sebagaimana Gambar 2 yaitu salah satu garis alternatif untuk pemisah yang memisahkan antara kedua kelas [16].



**Gambar 2** Kedua *Class* -1 dan +1 dipisahkan oleh *Hyperplane*

*Hyperplane* merupakan pemisah kedua *class* yang baik, *hyperplane* merupakan sebuah bidang atau permukaan linear yang memisahkan dua kelas data [17]. Dalam kasus klasifikasi biner (dua kelas), *hyperplane* ini berfungsi sebagai pembatas antara dua kelas, dan tujuannya adalah untuk memaksimalkan margin, yaitu jarak antara *hyperplane* dan titik terdekat dari setiap kelas (yang disebut sebagai *support vector*) [1]. *Hyperplane* ini adalah inti dari proses pembelajaran pada SVM, karena dengan menemukan *hyperplane* yang tepat [10] bisa mengoptimalkan *accuracy* dalam memprediksi risiko gagal bayar atau kredit.

5. Pengukuran *Performance*

Pada tahap selanjutnya setelah data dibagi dengan teknik *10-fold cross validation* yaitu data *training* dan data *testing*, kemudian diklasifikasi dengan algoritma SVM sebagai usaha untuk mencari *support vector* terbaik sebagai batas *margin* yang memisahkan kedua *class*. Pada tahapan ini dilakukan pengukuran *accuracy* prediksi dengan mengamati performa *accuracy*. Semakin besar *accuracy* otomatis akan mendekati model terbaik yang akurat dalam memprediksi *credit scoring*

## 2.2 Algoritma Particle Swarm Optimization (PSO)

Metode *hybrid* telah banyak digunakan diberbagai bidang. *Particle Swarm Optimization* (PSO) salah satu algoritma stokastik yang dikenalkan pertama dikenalkan oleh Eberhart dan Kennedy pada tahun 1995 [18]. Algoritma PSO adalah algoritma optimisasi yang terinspirasi oleh perilaku kelompok dan gerakan koloni burung dalam alam [13]. Algoritma ini digunakan untuk menemukan solusi terbaik dalam ruang pencarian parameter yang mungkin untuk masalah optimisasi. Cara kerja PSO mirip dengan perilaku kawanan burung atau sekumpulan partikel yang bergerak melalui ruang pencarian untuk mencari makanan atau tujuan tertentu. Setiap

partikel dalam algoritma PSO mewakili solusi potensial untuk masalah optimisasi dan memiliki posisi serta kecepatan saat bergerak melalui ruang pencarian [5] [19]. Algoritma ini beroperasi dengan cara berikut:

1. Inisialisasi Partikel: Pada awal iterasi, posisi dan kecepatan setiap partikel diinisialisasi secara acak dalam ruang pencarian.
2. Evaluasi Fitness: Setiap partikel dievaluasi berdasarkan seberapa baik solusi yang mereka wakili. Ini dilakukan dengan menggunakan fungsi fitness yang sesuai dengan masalah yang akan dipecahkan.
3. Perbarui PBest dan GBest: Setiap partikel melacak solusi terbaik yang pernah mereka temukan (Personal Best, PBest) dan solusi terbaik yang pernah ditemukan oleh seluruh kawanan (Global Best, GBest). Jika sebuah partikel menemukan solusi yang lebih baik, mereka memperbarui PBest dan GBest.
4. Perbarui Kecepatan dan Posisi: Setiap partikel memperbarui kecepatan dan posisi mereka berdasarkan PBest dan GBest. Gerakan partikel ini disinkronkan dengan partikel lain dalam kelompok.
5. Iterasi: Langkah-langkah di atas diulang sejumlah iterasi yang ditentukan sebelumnya atau hingga kondisi berhenti terpenuhi.
6. Hasil Akhir: Setelah sejumlah iterasi, PSO akan menghasilkan posisi terbaik yang mungkin sebagai solusi optimal atau mendekati solusi optimal untuk masalah yang diberikan.

Merujuk pada Chen *et al* [20] terdapat dua formula untuk membentuk algoritma PSO yang menggunakan  $k$  referensi iterasi saat ini, dan  $k + 1$  yang merepresentasikan pada iterasi pada tahap selanjutnya, yaitu:

Posisi partikel

$$X_{k+1}^i = X_k^i + V_{k+1}^i \quad (1)$$

Kecepatan partikel:

$$V_{k+1}^i = W_k V_k^i + c_1 r_1 (P_k^i - X_k^i) + c_2 r_2 (p_k^g - X_k^i) \quad (2)$$

di mana:

$X_k^i$  = posisi partikel

$V_k^i$  = kecepatan partikel

$p_k^i$  = posisi partikel individu terbaik

$p_k^g$  = posisi gerombolan atau kelompok terbaik

$w_k$  = berat inersia konstan

$c_1 r_1$  = parameter kognitif dan *social*

$c_2 r_2$  = bilangan acak yang bernilai antara 0 sampai dengan 1

Metode PSO yang dipakai dalam melakukan estimasi nilai parameter pada algoritma SVM, yaitu dengan memberikan nilai parameter  $\beta_i$ ,  $i = 0, 1, \dots, k$ , sehingga dapat dihasilkan nilai maksimal *Likelihood* pada persamaan (2). Sehingga diperoleh perkiraan nilai parameter algoritma SVM berdasarkan PSO.

### 2.3 Probabilitas Risiko Gagal Bayar

Resiko gagal bayar atau kredit macet merupakan salah satu dampak negatif dari pemberian pembiayaan atau pinjaman kepada anggota. Hubungan antara probability kredit macet dengan  $\Lambda(score_i)$ , adalah sebagai berikut:

a.  $Prob(Default_i) = \Lambda(score_i)$  jika terjadi kredit macet

b.  $Prob(No Default_i) = 1 - \Lambda(score_i)$  jika tidak terjadi kredit macet

Probabilitas gagal bayar (kredit macet) yakni  $0 \leq \Lambda(score_i) \leq 1$ . Berdasarkan interval ini, bagi debitur dengan  $\Lambda(score_i)$  pemberian pembiayaan yang rendah memiliki kemungkinan kredit macet yang rendah, sedangkan bagi debitur  $\Lambda(score_i)$  kredit yang tinggi, memiliki kemungkinan kredit macet yang tinggi pula. Oleh sebab itu, sebuah  $\Lambda(score_i)$  pembiayaan yang tinggi dapat berakibat pada penolakan pemberian pinjaman kepada nasabah

### 2.4 K-Fold Cross Validation

*K-Fold Cross Validation* merupakan salah satu teknik yang dapat dimanfaatkan untuk mengevaluasi model prediksi. Metode ini membagi data ke dalam  $k$  subset dan melakukan iterasi sebanyak  $k$  kali untuk *training* dan *testing* [16]. Metode ini menggunakan satu subset untuk dijadikan sebagai data uji dan subset interaktif lain yang digunakan sebagai data *training*, pada setiap iterasi yang dilakukan.

Dalam mengembangkan model melalui pelatihan data, model kemudian dievaluasi stabilitas dan kinerjanya menggunakan kinerja akurasi. Model dengan akurasi tertinggi akan diterapkan untuk pengujian menggunakan dataset pengujian.

### 2.5 Confusion Matrix

*Confusion matrix* adalah tabel yang digunakan untuk mengukur kinerja model klasifikasi dengan membandingkan prediksi model dengan nilai sebenarnya dari data. Matriks ini terdiri dari empat sel utama [21], yaitu:

- a. *True Positive* (TP): Jumlah pengamatan yang benar-benar positif dan diprediksi positif dengan benar oleh model.
- b. *True Negative* (TN): Jumlah pengamatan yang benar-benar negatif dan diprediksi negatif dengan benar oleh model.
- c. *False Positive* (FP): Jumlah pengamatan yang benar-benar negatif tetapi salah diprediksi sebagai positif oleh model (kesalahan tipe I).
- d. *False Negative* (FN): Jumlah pengamatan yang benar-benar positif tetapi salah diprediksi sebagai negatif oleh model (kesalahan tipe II).

*Confusion matrix* digunakan untuk menghitung berbagai metrik evaluasi klasifikasi seperti akurasi, presisi, *recall*, *F1-score*, dan lainnya, persamaan 4 adalah formula untuk mengukur akurasi.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (4)$$

Evaluasi performa dapat diukur dengan *accuracy*. *Recall* adalah metrik evaluasi yang mengukur kemampuan model untuk mengidentifikasi semua instansi yang seharusnya positif. Dalam istilah matriks konfusi, *recall* dihitung dengan rumus:

$$Recall = \frac{TP}{TP+FN} \quad (5)$$

*Recall* adalah penting dalam kasus di mana mendeteksi semua kasus positif adalah prioritas, bahkan jika itu mengakibatkan beberapa *false positives* (FP) [21]. Misalnya, dalam *credit scoring* untuk kelayakan kredit, lebih baik memiliki *recall* yang tinggi bahkan jika ini berarti lebih banyak orang yang dites berpotensi macet, untuk memastikan bahwa semua kasus kredit macet terdeteksi. *Recall* ini juga merupakan metrik evaluasi yang berguna dalam memahami performa model klasifikasi, terutama dalam konteks klasifikasi yang tidak seimbang (*imbalance class*) di mana kelas positif dan negatif tidak seimbang dalam jumlahnya, ini dapat memberi gambaran tentang sejauh mana model dapat mengenali kasus positif yang sebenarnya.

## 2.6 RapidMiner

*RapidMiner* adalah tools yang interaktif *user* untuk melakukan *machine learning* dan proses *mining* data. *RapidMiner* juga merupakan platform analitik data yang digunakan untuk menggali wawasan dari data bisnis dan ilmiah. Ini adalah perangkat lunak yang memungkinkan pengguna untuk mengimpor, mengelola, dan menganalisis data dari berbagai sumber. *RapidMiner* menyediakan alat visual untuk pemodelan data, pemrosesan bahasa alami, pemrosesan gambar, dan sebagainya [18].

*RapidMiner* memiliki beragam fitur, termasuk pemrosesan data ETL (*Extract, Transform, Load*), pemodelan prediksi, dan alat *machine learning* yang kuat. Ini juga mendukung bahasa pemrograman seperti Python dan R, sehingga pengguna yang terbiasa dengan bahasa pemrograman tersebut dapat mengintegrasikan kode mereka ke dalam alur kerja analitik. *RapidMiner* sering digunakan dalam berbagai industri, termasuk bisnis, ilmu pengetahuan, keuangan, dan lainnya untuk mengambil keputusan yang didasarkan pada data dan menganalisis tren. Ini adalah salah satu alat populer di bidang analitik data dan *machine learning*.

## 3. HASIL DAN PEMBAHASAN

Pada bab ini akan dijelaskan terkait standarisasi data, estimasi nilai parameter, pengujian parameter, serta diskusi ilustrasi bagaimana keputusan pemberian kelayakan kredit kepada calon peminjam.

### 3.1 Menstandarisasi Data

Standarisasi data adalah proses normalisasi data agar memiliki skala yang seragam. Hal ini membantu dalam analisis data, terutama dalam konteks *machine learning* dan statistik, di mana perbedaan skala antara variabel dapat memengaruhi hasil analisis. Terdapat beberapa metode yang umum digunakan untuk menstandarisasi data, termasuk:

- a. *Standarisasi Z-Score*: Metode ini mengubah setiap titik data menjadi nilai yang berada pada distribusi normal dengan rata-rata nol dan simpangan baku satu
- b. *Min-Max Scaling*: Dalam metode ini, data dikonversi ke rentang tertentu, biasanya antara 0 dan 1.
- c. *Robust Scaling*: Metode ini mirip dengan Min-Max Scaling, tetapi menggunakan kuartil sebagai referensi daripada nilai minimum dan maksimum. Ini lebih tahan terhadap *outlier*.
- d. *Unit Vector Scaling*: Dalam metode ini, setiap vektor data dikonversi menjadi vektor dengan panjang 1 (vektor satuan). Ini berguna dalam analisis data multivariabel

Pada eksperimen ini proses standarisasi data dengan memakai metode *Z-score* supaya lebih efektif dan efisien seperti yang ditampilkan pada Tabel 1 berikut ini.

Tabel 1. Contoh Sampel Data *Testing* pada Koperasi Kasuwari

Status pernikahan	Kelamin	Usia	Tanggungan	Tabungan	Pendapatan	Jaminan	Jml pinjaman	Tenor	Status
Menikah	Laki-laki	39	3	2.500.000	4.500.000	20.000.000	10.000.000	24	Lancar
Menikah	Perempuan	52	4	800.000	2.500.000	8.000.000	10.000.000	36	Macet
Menikah	Laki-laki	55	3	900.000	2.740.000	9.500.000	10.000.000	36	Macet
Belum	Laki-laki	21	0	3.500.000	5.500.000	25.000.000	20.000.000	24	Lancar
Menikah	Laki-laki	28	1	5.500.000	6.000.000	21.000.000	17.500.000	48	Lancar
Menikah	Laki-laki	42	3	2.340.000	4.500.000	19.500.000	15.000.000	24	Lancar
Menikah	Laki-laki	43	2	4.320.000	7.600.000	18.500.000	13.000.000	24	Lancar
Menikah	Perempuan	44	3	4.850.000	6.500.000	21.000.000	20.000.000	36	Lancar
Menikah	Laki-laki	29	1	2.320.000	4.650.000	25.000.000	25.000.000	24	Lancar
Menikah	Laki-laki	40	2	2.340.000	3.750.000	18.500.000	15.000.000	24	Lancar
...	...	...	...	...	...	...	...	...	...

Proses pengumpulan data merupakan isu penting dalam *machine learning*. Hal ini karena identifikasi berapa banyak contoh yang cukup untuk mendapatkan pengetahuan, membuat keputusan yang baik dan memvalidasi hasil merupakan tugas yang sangat sulit. Untuk mendapatkan inisial kumpulan data dengan kemampuan diskriminatif, kami menggunakan validasi silang k-fold dengan k=10. Untuk memisahkan kumpulan data masukan, masing-masing subsetnya adalah dipisahkan dengan mempertahankan proporsi yang hampir sama dalam distribusi kelas atas datanya. Misalnya menggunakan validasi silang k-fold dengan k=10, jika ada 2 nilai kelas (X- dan X+) dalam masalah klasifikasi P dengan 1000 contoh secara total, dan jumlah contoh mayoritas dan kelas minoritas XX ( , )+ berturut-turut adalah 800 dan 200. Maka, masing-masing subset akan berisi 200 *instance* dengan 160 data negatif (layak diberi kredit) dan 40 contoh positif (tidak layak diberi kredit)

### 3.2 Pre Processing

Dalam penelitian ini, algoritma baru berbasis PSO diusulkan secara berurutan untuk memperkenalkan contoh buatan dengan ciri-ciri diskriminatif yang tinggi. Gambar 4 dan Algoritma 1 menggambarkan proses umum SVM metode PSO. Algoritma SVM-PSO mulai menormalkan input Himpunan data. Setiap contoh  $x_i, x_i = (x_1, x_2, \dots, x_n) \in \mathbb{R}^n$  (dengan XT termasuk total kumpulan data masukan) dalam data masukan dinormalisasi.

Pada bagian ini, SVM dilatih menggunakan  $X_{tr} = \{x_1, \dots, x_n\}$  dan  $y_{tr} = \{+1, -1\}$  dimana  $x_{sv}$  dan  $y_{sv}$  adalah SV ke-i dari  $X_{tr}^+$ , dan  $x_{sv}$  tetangga  $x_{sv}$  Saya - mewakili sv ke-j- terdekat +, dan r mendefinisikan dimensi *instance*. Vektor awal  $v = 0$ ,  $k = 1, 2, \dots, d$ , dan algoritma memilih satu atau lebih entri acak dari array. Dalam percobaan hanya satu yang dipilih. Contoh buatan diperoleh dengan  $x_{vsg} = + \epsilon \cdot \text{Saya} + X_{tr}^+$  dan  $X_{tr}^-$ , idenya dalam hal ini langkahnya adalah mengidentifikasi *support vector* (SV). Dalam SVM solusinya adalah diberikan oleh sebagian kecil contoh yang disebut *support vector* (SV). Keberhasilan setiap SV ditentukan oleh posisinya dalam ruang fitur yang mendefinisikan *hyperplane* dan memberikan solusinya. *hyperplane* diperoleh pada langkah ini akan diperoleh Persamaan, dan fungsi keputusannya diberikan oleh Persamaan, yang miring karena ketidakseimbangan dalam memasukan data. Namun, SV yang diperoleh akan membantu kita menghasilkan SV baru contoh dan populasi awal untuk PSO [13]. Poin data ini adalah yang

paling representatif, namun efektivitasnya dalam beberapa kasus sulit diukur. Dalam kasus kami, algoritma PSO yang diterapkan menjamin efektivitas titik data yang dibuat.

Pada tahapan ini dilakukan penyiapan data awal, pemilihan atribut atau variabel yang akan dianalisis, validasi variabel yang ada serta persiapan transformasi data. Dilakukan juga persiapan data *training* dan *testing* sebanyak 200 data. Pada Tabel 2 menunjukkan atribut yang digunakan terdiri dari 10 atribut yaitu: Status perkawinan, Jenis kelamin, usia, tanggungan, tabungan, pendapatan, jaminan, nominal pembiayaan, tenor dan keterangan sebagai atribut.

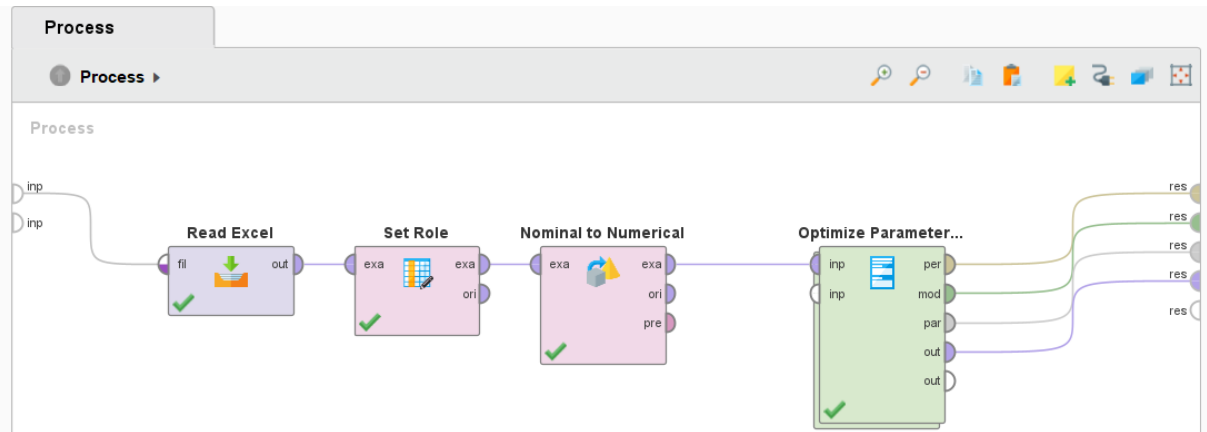
**Tabel 2.** Atribut dan Nilai Dataset

No	Nama Atribut	Nilai
1	Status Pernikahan	- Belum menikah
		- Sudah menikah
2	Jenis kelamin	- Laki-laki
		- Perempuan
3	Usia	- 20-55
		- 55-60
4	Jumlah tanggungan	- 0-2
		- 2-5
		- 5-6
5	Tabungan	- 100.000 – 1.000.000
		- 1.000.000 – 2.000.000
		- 2.000.000 – 3.000.000
		- > 4.000.000
6	Pendapatan	- 1.000.000 – 2.000.000
		- 2.000.000 – 3.000.000
		- > 4.000.000
7	Jaminan	- BPKB Kendaraan Bermotor
		- Sertifikat Tanah
		- Sertifikat Rumah
8	Jml_pinjaman	- 1.000.000 – 5.000.000
		- 5.000.000 – 10.000.000
		- 10.000.000 – 20.000.000
		- > 20.000.000
9	Tenor	- 0 – 12 Bulan
		- 13 – 24 Bulan
		- 25 – 36 Bulan
		- > 36 Bulan
10	Status	- Lancar
		- Macet

Pada Tabel 2 ditunjukkan *sample* 10 data yang pakai untuk data latih atau data *training*. Pada penelitian ini data *training* dan juga data *testing* yang digunakan sebanyak 200 data anggota

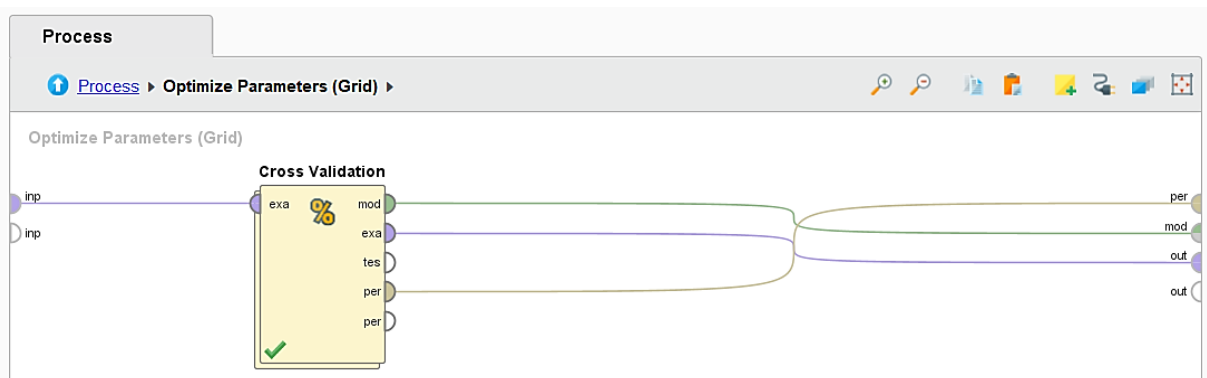
### 3.3 Implementasi

Eksperimen ini dilakukan untuk mengetahui seberapa besar akurasi dari performa training Model *Particle Swarm Optimizations* (PSO) pada *Support Vector Machine* (SVM). Untuk Teknik validasi menggunakan cross validation dengan 10 *fold cross validation* dari 70% dataset training. Bagian input parameter pada algoritma SVM di input nilai parameter Gamma, C dan Epsilon, karena setting parameter yang tepat ini sangat mempengaruhi akurasi dari PSO-SVM ini. Evaluasi model dalam eksperimen ini menggunakan Tools RapidMiner. Desain model yang akan digunakan dapat kita lihat pada Gambar 3, Gambar 4, dan Gambar 5 berikut ini.



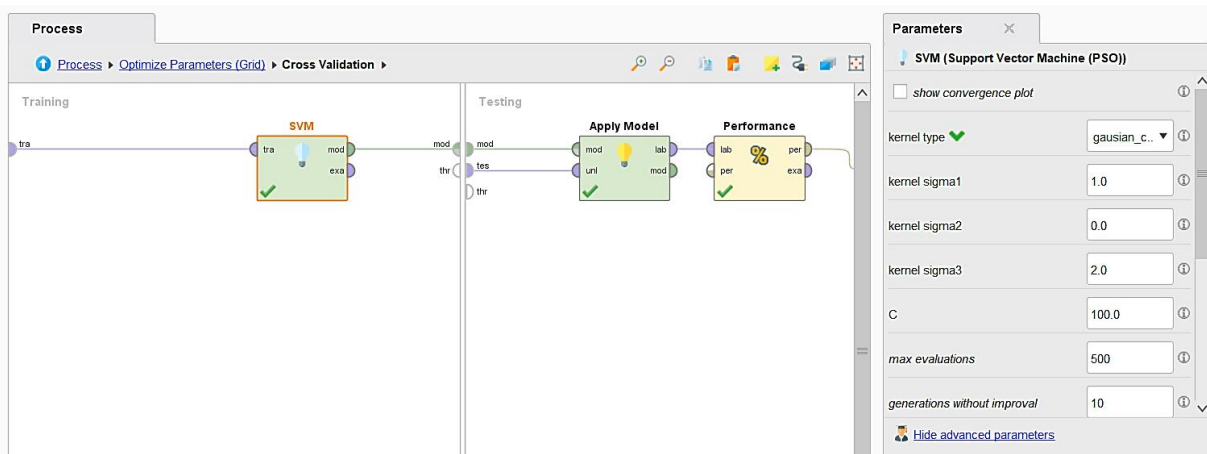
Gambar 3 Proses Import Excel dan Desain menggunakan Tools RapidMiner

1. *Read Excel*: Operator *Read Excel* digunakan untuk mengimpor data dari file Excel ke dalam alur kerja *RapidMiner*. Ini memungkinkan Anda untuk mengambil data dari spreadsheet Excel dan menggunakannya dalam analisis data.
2. *Set Role*: Operator *Set Role* digunakan untuk mendefinisikan peran atau jenis data yang dimiliki oleh kolom dalam dataset Anda. Ini adalah langkah penting dalam *pre-processing* data karena membantu *RapidMiner* memahami bagaimana kolom harus diolah selanjutnya
3. *Nominal to Numerical*: Operator ini mengubah tipe atribut non-numerik yang dipilih menjadi tipe numerik, juga memetakan semua nilai atribut ke nilai numerik.
4. *Optimize parameter*: Operator ini menemukan nilai optimal dari parameter yang dipilih untuk operator dalam subprocessesnya [21].



Gambar 4. Operator Cross Validation

5. *Cross Validation*: Operator ini melakukan validasi silang untuk memperkirakan kinerja statistik model pembelajaran dalam hal ini memakai model *Particle Swarm Optimization (PSO)* dan *Support Vector Machine (SVM)* [17]



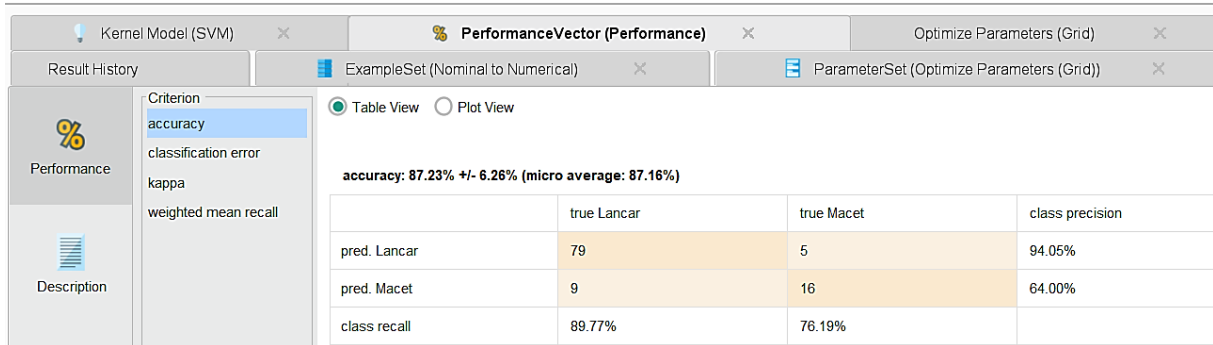
Gambar 5. Pengujian Model PSO-SVM Menggunakan Data Training dan Data Testing

6. *Validasi Model*: Metode klasifikasi yang digunakan dalam penelitian ini, yaitu SVM seperti yang ditunjukkan pada Gambar 5.
7. *Apply Model*: Operator ini digunakan untuk menerapkan model yang telah dikembangkan pada dataset baru atau data pengujian. Operator ini berfungsi untuk membuat prediksi pada data yang tidak digunakan selama pelatihan. Hasilnya adalah dataset baru yang berisi prediksi dari model tersebut.
8. *Performance*: Operator ini digunakan untuk mengevaluasi kinerja model pada dataset pengujian atau validasi. Ini memungkinkan Anda untuk mengukur sejauh mana model berkinerja dalam melakukan tugas tertentu, seperti klasifikasi atau regresi [22].

Hasil Pengujian Model PSO-SVM Menggunakan *Cross Validation* dan *Confusion Matrix*

1. Hasil pengujian K-10 *Cross Validation* menggunakan algoritma PSO-SVM.

Penelitian ini melakukan evaluasi model dengan menggunakan K-10 *cross validation*. menggunakan kernel dot dan radial, setting parameter Gamma dengan nilai 1, C dengan nilai 1,3 dan Epsilon dengan nilai 0,5 sebagaimana yang ditunjukkan pada Gambar 6 berikut ini:



Gambar 6. Hasil Pengujian K-10 *Cross Validation* Algoritma PSO-SVM

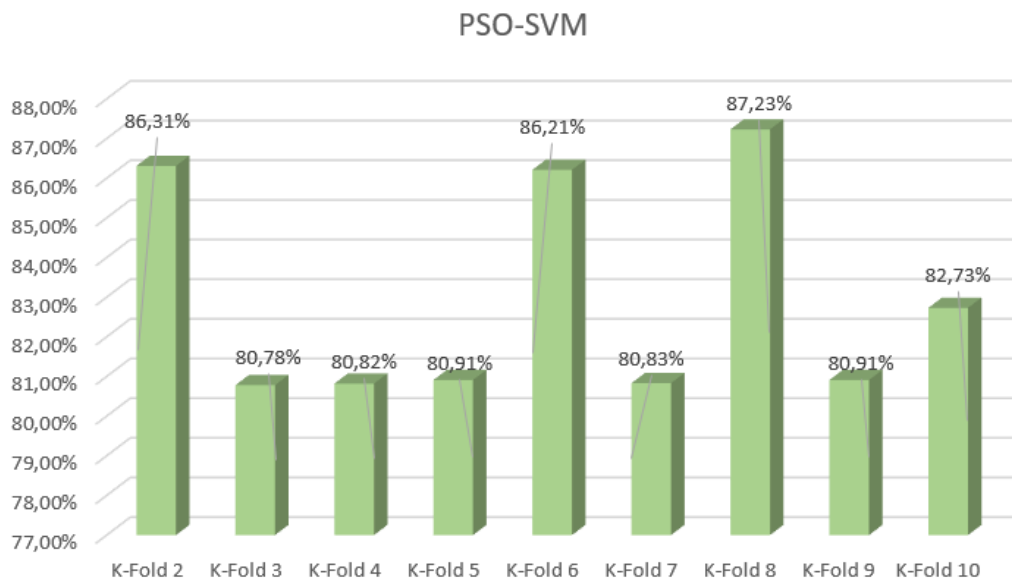
Dari proses evaluasi model SVM-PSO sesuai dengan persamaan (4) ini menghasilkan matrix *Accuracy* sebesar 87.23%. Berikut ini merupakan perhitungan Akurasi menggunakan *Confusion Matrix* dari Gambar 6 menggunakan persamaan 4.

$$\frac{79+16}{79+16+9+5} = \frac{95}{109} = 0.8723 = 87.23\%$$

Pada Tabel 3 menunjukkan hasil pengujian dengan menggunakan 10-fold *cross validation*, dimana hasil akurasi tertinggi tertinggi terdapat pada pengujian ke-8.

Tabel 3. Hasil Pengujian K-10 *Cross Validation*

Metode	K-Fold 2	K-Fold 3	K-Fold 4	K-Fold 5	K-Fold 6	K-Fold 7	K-Fold 8	K-Fold 9	K-Fold 10
PSO-SVM	86,31%	80,78%	80,82%	80,91%	86,21%	80,83%	87,23%	80,91%	82,73%



Gambar 7. Hasil Pengujian Algoritma PSO-SVM

Pada Gambar 7 dan pada Tabel 3 menunjukkan hasil pengujian Algoritma PSO-SVM menggunakan K-10 *Cross Validation*, pada pengujian ini didapatkan hasil terbaik adalah pada pengujian ke-8 yaitu mendapatkan *accuracy* sebesar 87.23%

#### 4. KESIMPULAN

Berdasarkan hasil penelitian yang dilakukan, dapat disimpulkan dalam penelitian ini fitur atau atribut yang digunakan dalam menentukan kelayakan kredit terdiri dari 10 fitur yaitu status pernikahan, jenis kelamin, usia, tanggungan, tabungan, pendapatan, nilai jaminan, nominal pinjaman, tenor dan keterangan fitur hasil. Dari hasil proses pengujian menggunakan algoritma yang telah dilakukan menggunakan tools RapidMiner. Penerapan Fitur seleksi dan algoritma *Particle Swarm Optimization* (PSO) pada *Support Vector Machine* (SVM) untuk memprediksi kelayakan kredit pada LKMS Kasuwari dengan menggunakan kernel dot dan radial, setting parameter Gamma dengan nilai 1, C dengan nilai 1.3 dan Epsilon dengan nilai 0,5 serta setting parameter Population Size dengan nilai 500 dan pengurangan atribut yang tidak relevan menggunakan seleksi fitur. Dari pengujian ini diperoleh akurasi tertinggi sebesar 87.23% dipengujian ke 8, oleh karena itu penerapan algoritma PSO dinyatakan terbukti dapat meningkatkan kinerja pada algoritma SVM untuk meningkatkan hasil akurasi dalam memprediksi kelayakan pemberian kredit.

#### REFERENCES

- [1] D. B. Vukovic, K. Romanyuk, S. Ivashchenko, and E. M. Grigorieva, "Are CDS spreads predictable during the Covid-19 pandemic? Forecasting based on SVM, GMDH, LSTM and Markov switching autoregression," *Expert Syst. Appl.*, vol. 194, no. January, p. 116553, 2022, doi: 10.1016/j.eswa.2022.116553.
- [2] R. Oktapiani, D. Prayudi, A. Fajria, N. S. Z. Nufus, and R. N. Lestari, "Sistem Pendukung Keputusan Untuk Menentukan Manajemen Kelayakan Pemberian Kredit Di Bank Mandiri Taspen Sukabumi Menggunakan Metode Analytic Hierarchy Process," *Indones. J. Softw. Eng.*, vol. 8, no. 1, pp. 36–45, 2022, doi: 10.31294/ijse.v8i1.12054.
- [3] V. B. Djeundje, J. Crook, R. Calabrese, and M. Hamid, "Enhancing credit scoring with alternative data," *Expert Syst. Appl.*, vol. 163, p. 113766, 2021, doi: 10.1016/j.eswa.2020.113766.
- [4] M. Abdoli, M. Akbari, and J. Shahrabi, "Bagging Supervised Autoencoder Classifier for credit scoring," *Expert Syst. Appl.*, vol. 213, no. PB, p. 118991, 2023, doi: 10.1016/j.eswa.2022.118991.
- [5] S. Purnama and A. P. Kusumawardhani, "Deteksi Peluang Gagal Bayar Calon Debitur Menggunakan Algoritma Particle Swarm Optimization (PSO) untuk Meningkatkan Kinerja Manajemen Risiko pada Koperasi Simpan Pinjam ABC," *KUBIK J. Publ. Ilm. Mat.*, vol. 6, no. 2, pp. 71–84, 2022, doi: 10.15575/kubik.v6i2.13835.
- [6] H. Yasin, A. R. Hakim, and A. Hoyyi, "Sistem Informasi Potensi Kredit Macet Berbasis Aplikasi Credit Scoring-Support Vector Machine (CS-SVM)," *Pros. Semin. Nas. VARIANSI*, vol. 1, no. 1, pp. 1–9, 2020, [Online]. Available: <https://ojs.unm.ac.id/variansistatistika/article/view/19493>
- [7] A. Markov, Z. Seleznyova, and V. Lapshin, "Credit scoring methods: Latest trends and points to consider," *J. Financ. Data Sci.*, vol. 8, pp. 180–201, 2022, doi: 10.1016/j.jfds.2022.07.002.
- [8] S. Riyadi, M. M. Siregar, K. fadhli F. Margolang, and K. Andriani, "Analysis Of SVM And Naive Bayes Algorithm In Classification Of Nad Loans In Save And Loan Cooperatives," *JURTEKSI (Jurnal Teknol. dan Sist. Informasi)*, vol. 8, no. 3, pp. 261–270, Aug. 2022, doi: 10.33330/jurteksi.v8i3.1483.
- [9] S. Bumbungan, Kusrini, and Kusnawi, "Penerapan Particle Swarm Optimization (PSO) dalam Pemilihan Parameter Secara Otomatis pada Support Vector Machine (SVM) untuk Prediksi Kelulusan Mahasiswa Politeknik Amamapare Timika," *J. Tek. AMATA*, vol. 4, no. 1, pp. 81–93, 2022, doi: 10.55334/jtam.v4i1.77.
- [10] G. Yao, X. Hu, and G. Wang, "A novel ensemble feature selection method by integrating multiple ranking information combined with an SVM ensemble model for enterprise credit risk prediction in the supply chain," *Expert Syst. Appl.*, vol. 200, no. January, p. 117002, 2022, doi: 10.1016/j.eswa.2022.117002.
- [11] D. T. Larose, *Data Mining Methods and Models*. Canada: John Wiley & Sons, Inc, 2007.
- [12] S. Maldonado, C. Bravo, J. López, and J. Pérez, "Integrated framework for profit-based feature selection and SVM classification in credit scoring," *Decis. Support Syst.*, vol. 104, pp. 113–121, 2017, doi: 10.1016/j.dss.2017.10.007.
- [13] J. Cervantes, F. Garcia-Lamont, L. Rodriguez, A. López, J. R. Castilla, and A. Trueba, "PSO-based method for SVM classification on skewed data sets," *Neurocomputing*, vol. 228, no. December 2015, pp. 187–197, 2017, doi: 10.1016/j.neucom.2016.10.041.
- [14] C. W. Dawson, *Projects in Computing and Information Systems*. Addison Wesley, 2009.
- [15] P. Danenas and G. Garsva, "Selection of Support Vector Machines based classifiers for credit risk domain," *Expert Syst. Appl.*, vol. 42, no. 6, pp. 3194–3204, 2015, doi: 10.1016/j.eswa.2014.12.001.
- [16] A. R. Naufal, R. Satria, and A. Syukur, "Penerapan Bootstrapping untuk Ketidakseimbangan Kelas dan

- Weighted Information Gain untuk Feature Selection pada Algoritma Support Vector Machine untuk Prediksi Loyalitas Pelanggan,” *J. Intell. Syst.*, vol. 1, no. 2, pp. 98–108, 2015.
- [17] A.- Amrin and O.- Pahlevi, “Implementation of Logistic Regression Classification Algorithm and Support Vector Machine for Credit Eligibility Prediction,” *J. Informatics Telecommun. Eng.*, vol. 5, no. 2, pp. 433–441, 2022, doi: 10.31289/jite.v5i2.6220.
- [18] M. Han, J., & Kamber, *Data Mining : Concepts and Techniques*, 3rd Editio. Morgan Kaufmann Publishers, 2012.
- [19] G. N. Kouziokas, “SVM kernel based on particle swarm optimized vector and Bayesian optimized SVM in atmospheric particulate matter forecasting,” *Appl. Soft Comput. J.*, vol. 93, p. 106410, 2020, doi: 10.1016/j.asoc.2020.106410.
- [20] S. Chen, J. qiang Wang, and H. yu Zhang, “A hybrid PSO-SVM model based on clustering algorithm for short-term atmospheric pollutant concentration forecasting,” *Technol. Forecast. Soc. Change*, vol. 146, no. May, pp. 41–54, 2019, doi: 10.1016/j.techfore.2019.05.015.
- [21] M. A. Witten, I. H., Frank, E., & Hall, *Data Mining Practical Machine Learning Tools and Techniques*, 3rd ed. USA: Morgan Kaufmann Publishers, 2011.
- [22] T. T. Muryono, A. Taufik, and I. Irwansyah, “Perbandingan Algoritma K-Nearest Neighbor, Decision Tree, dan Naive Bayes untuk Menentukan Kelayakan Pemberian Kredit,” *Infotech J. Technol. Inf.*, vol. 7, no. 1, pp. 35–40, Jun. 2021, doi: 10.37365/jti.v7i1.104.