

BAB I PENDAHULUAN UMUM

1.1. Latar Belakang

Pendekatan *machine learning* dalam analisis survival semakin berkembang seiring meningkatnya kompleksitas data, khususnya ketika jumlah variabel yang dianalisis jauh melebihi jumlah observasi. Kondisi seperti ini, menyebabkan menurunnya kinerja metode statistik klasik sehingga memerlukan pendekatan regularisasi yang lebih baik (Tibshirani et al., 2002). Metode survival seperti *Cox Proportional Hazard Regression* (CPHR), sering kali tidak berkinerja baik karena sulitnya memenuhi asumsi model dan permasalahan estimasi parameter (Belle et al., 2011; Sanz et al., 2020b). Ketidakseimbangan antara jumlah prediktor dan jumlah kejadian juga dapat menimbulkan bias estimasi, peningkatan varians, serta penurunan validitas uji signifikansi (Peduzzi et al., 1996). Oleh karena itu, pendekatan berbasis *machine learning* semakin banyak digunakan karena kemampuannya menangani kompleksitas data tanpa ketergantungan kuat pada asumsi parametrik (Fryan & Alazzam, 2023b; Tizi & Berrado, 2023b).

Pendekatan analisis *survival* diaplikasikan dalam banyak bidang seperti medis, industri, dan ekonomi (Sorayaie Azar et al., 2022b; Fryan & Alazzam, 2023b; Deng et al., 2024; Tonetto et al., 2024; Saha et al., 2025; Xu, 2025). Metode *machine learning* telah terbukti meningkatkan kinerja analisis *survival* seperti *Random Survival Forest* (RSF) (Papathanasiou et al., 2023; Germer et al., 2024; Jia et al., 2025), *Artificial Neural Network* (ANN) (Bae et al., 2023; Y. Deng et al., 2023; Ho et al., 2024; Meixide et al., 2024) ataupun mesin vektor pendukung yang lebih dikenal dengan *Support Vector Machine* (SVM) (Sanz et al., 2020b; Van Belle et al., 2011; Shen et al., 2021; Kiessling et al., 2023).

Keunggulan utama SVM terletak pada kemampuannya menangani data berdimensi tinggi dan pola non-linear melalui formulasi optimasi berbasis margin maksimum (Belle et al., 2011; Sanz et al., 2020b). Namun demikian, sebagian besar pengembangan SVM dalam analisis *survival* belum secara eksplisit mempertimbangkan struktur risiko yang heterogen antarindividu. Pada praktiknya, data *survival* sering kali bersifat tidak seimbang karena proporsi data tersensor lebih besar dibandingkan data *event*. Ketidakseimbangan ini banyak ditemukan pada berbagai konteks seperti kanker (Nguyen & Huynh, 2021a; Sorayaie Azar et al., 2022b; Lee, 2023; El Badisy et al., 2024; Germer et al., 2024), penyakit kronis (Sujatha et al., 2022a; Akazue et al., 2023), serta kegagalan usaha (Leurcharusmee et al., 2022; Tonetto et al., 2024), dan dapat menyebabkan model kurang sensitif terhadap kelompok berisiko tinggi.

Berbagai pengembangan *Survival SVM* telah dilakukan untuk mengatasi keterbatasan tersebut, salah satunya melalui Analisis *Survival Least Squares Support Vector Machine* (Surv-LSSVM) (Belle et al., 2011; Khotimah et al., 2017; Khoiri et al., 2019). Pendekatan LSSVM terbukti unggul melalui formulasi kuadratik pada variabel *slack*, namun sebagian besar pengembangannya masih memperlakukan seluruh pengamatan secara homogen. Kebaruan penelitian dalam studi ini terletak pada

pendekatan pembobotan pada metode SVM khususnya *Least Square Support Vector Machine* (LSSVM) untuk analisis *survival*. Metode yang dikembangkan diaplikasikan pada data di mana metode konvensional tidak menghasilkan kinerja yang baik. Kinerja model juga dievaluasi pada data dengan tingkat *censoring* yang bervariasi. Selain itu, jenis pembobotan yang digunakan dalam penelitian terdahulu masih terbatas dan umumnya tidak diturunkan langsung dari karakteristik data *survival*. Hingga saat ini, belum ada literatur yang mengevaluasi secara komprehensif mengenai integrasi estimator Kaplan–Meier sebagai fungsi pembobot dalam kerangka Surv-LSSVM, sehingga topik tersebut masih menjadi celah penelitian yang relevan untuk dikaji lebih lanjut.

Berdasarkan celah penelitian tersebut, penelitian ini mengusulkan pengembangan *Survival Weighted Least Squares Support Vector Machine* (Surv-WLSSVM). Model tersebut mengintegrasikan Surv-LSSVM dengan pembobotan berbasis estimator Kaplan–Meier. Kebaruan pendekatan ini terletak pada penentuan bobot yang diperoleh langsung dari data *survival*.

Selanjutnya, untuk meningkatkan kinerja model yang dikembangkan, digunakan metode pencarian parameter SVM yang optimal. Penelitian ini menerapkan algoritma *Particle Swarm Optimization* (PSO) untuk menentukan parameter regularisasi dan parameter *kernel* dalam model. Dibandingkan dengan algoritma genetika, penggunaan PSO menunjukkan keunggulan dalam hal akurasi, efisiensi iterasi, serta kesederhanaan mekanisme optimisasi yang digunakan (Algorithms et al., 2018; Çarkıt & Alçı, 2022). Selain itu, PSO juga telah terbukti memiliki kinerja yang lebih baik pada berbagai terapan lain, termasuk pemodelan invers stokastik dalam bidang geofisika (Pace et al., 2022).

Metode yang dikembangkan dalam penelitian, dievaluasi pada data simulasi dan data riil. Hal ini bertujuan mengukur kebaikan model pada berbagai macam skenario data. Beragam jumlah variabel, ukuran sampel, distribusi data dan proporsi data tersensor diaplikasikan dalam data simulasi. Kinerja model dievaluasi menggunakan *Concordance Index* (c-index), di mana model dengan nilai c-index yang lebih tinggi menunjukkan kemampuan kinerja yang lebih baik. Selanjutnya model diimplementasikan pada data pemberian ASI oleh Ibu di Indonesia untuk mengetahui variabel yang berpengaruh sesuai c-index yang dihasilkan dari model.

1.2. Rumusan Masalah

Rumusan masalah dalam penelitian ini adalah sebagai berikut:

1. Bagaimana penggunaan model *Survival Least Squares Support Vector Machine* (Surv-LSSVM) dengan fungsi *kernel Radial Basis Function* (RBF) dan *tuning parameter* berbasis algoritma *Particle Swarm Optimization* (PSO)?
2. Bagaimana mengembangkan metode *Survival Weighted Least Squares Support Vector Machine* (Surv-WLSSVM) menggunakan pembobotan berbasis estimator Kaplan–Meier?

3. Bagaimana kemampuan model Surv-WLSSVM berbasis pembobotan dalam menghasilkan indeks prognostik dikaitkan dengan waktu *survival* dibandingkan dengan model Surv-LSSVM yang tanpa pembobotan?
4. Bagaimana aplikasi model Surv-WLSSVM pada data pemberian ASI eksklusif oleh ibu di Indonesia, serta bagaimana interpretasi indeks prognostik yang dihasilkan dalam menjelaskan keberlangsungan pemberian ASI eksklusif?

1.3. Tujuan Penelitian

Berdasarkan uraian rumusan masalah sebelumnya, tujuan dalam penelitian ini adalah sebagai berikut:

1. Menganalisis penggunaan model *Survival Least Squares Support Vector Machine* (Surv-LSSVM) dengan fungsi *kernel Radial Basis Function* (RBF) dan *tuning parameter* berbasis algoritma *Particle Swarm Optimization* (PSO).
2. Mengembangkan metode *Survival Weighted Least Squares Support Vector Machine* (Surv-WLSSVM) menggunakan pembobotan berbasis estimator Kaplan–Meier.
3. Menganalisis kemampuan model Surv-WLSSVM dalam menghasilkan indeks prognostik dikaitkan dengan waktu *survival* dibandingkan model Surv-LSSVM yang tanpa pembobotan.
4. Mengaplikasikan model Surv-WLSSVM pada data pemberian ASI eksklusif oleh ibu di Indonesia, serta menginterpretasikan indeks prognostik yang dihasilkan dalam menjelaskan keberlangsungan pemberian ASI eksklusif.

1.4. Manfaat Penelitian

Manfaat yang ingin dicapai dari hasil penelitian ini antara lain:

1. Memberikan kontribusi metodologis dalam penanganan permasalahan analisis *survival* melalui penerapan metode *support vector machine* yang terintegrasi dengan skema pembobotan.
2. Menyediakan alternatif pendekatan *support vector machine* dalam analisis *survival*, khususnya melalui pengembangan berbasis pembobotan estimator Kaplan–Meier.
3. Menunjukkan potensi penggunaan pembobotan dalam kerangka *support vector machine* sebagai strategi untuk meningkatkan kinerja prediktif dan ketepatan estimasi pada analisis *survival*, sehingga membuka peluang pengembangan metodologi lanjutan.
4. Memberikan implikasi praktis melalui penerapan model Surv-WLSSVM pada studi kasus pemberian ASI eksklusif, yang dapat digunakan sebagai masukan berbasis bukti bagi pemangku kepentingan dalam upaya peningkatan kualitas dan keberlanjutan program ASI eksklusif di Indonesia.

1.5. Ruang Lingkup Penelitian

Pada penelitian ini, terdapat beberapa batasan masalah yaitu:

1. Penelitian merupakan kajian teoretis tentang penggunaan mesin vektor pendukung terboboti untuk analisis *survival*. Jenis pembobotan yang digunakan didasarkan pada estimator Kaplan–Meier atau rentang nilai bobot 0 sampai dengan 1.

2. Data yang digunakan adalah data simulasi dan data riil yang secara umum terdiri dari 3 kategori banyaknya variabel dengan distribusi tertentu, selain itu terdapat 4 kategori banyaknya sampel dengan beragam proporsi tersensor dalam data.
3. Proses *feature map* dalam model menggunakan fungsi *kernel Radial Basis Function* (RBF) dan evaluasi model menggunakan ukuran c-index.

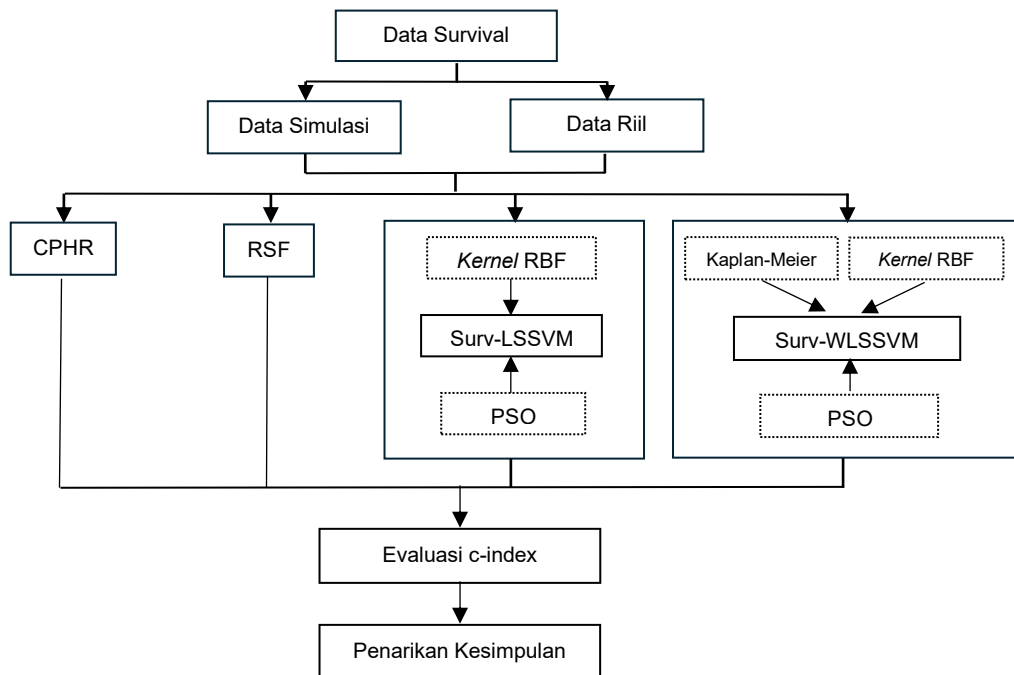
1.6. Alur Penelitian

Alur penelitian diuraikan sebagai berikut:

1. Implementasi model tanpa pembobotan, yaitu *Survival Least Square Support Vector Machine* (Surv-LSSVM), diikuti dengan proses *tuning parameter* untuk menentukan parameter optimal metode SVM serta evaluasi kinerja model menggunakan *Concordance Index* (c-index).
2. Mengembangkan model berbasis pembobotan yakni *Survival Weighted Least Square Support Vector Machine* (Surv-WLSSVM) dengan menggunakan *Kernel Radial Basis Function* (RBF) yang berperan sebagai *feature map* dan pembobotan berbasis estimator Kaplan-Meier.
3. Membangkitkan data simulasi dan mengumpulkan data riil serta melakukan pra-pemrosesan data. Data simulasi dibangkitkan menggunakan distribusi tertentu dengan beberapa skenario yang mempertimbangkan jumlah variabel, banyaknya sampel, dan proporsi tersensor.
4. Mengevaluasi nilai c-index yang diperoleh dari Model Surv-WLSSVM, Surv-LSSVM, *Random Survival Forest* (RSF) dan *Cox Proportional Hazard Regression* (CPHR). Langkah ini bertujuan mengukur kinerja model Surv-WLSSVM berbasis pembobotan dibandingkan dengan Surv-LSSVM yang tanpa pembobotan, serta metode yang umum digunakan dalam analisis *survival* yakni RSF dan CPHR.
5. Implementasi model Surv-WLSSVM pada data pemberian ASI eksklusif oleh Ibu di Indonesia, interpretasi indeks prognostik yang diperoleh, serta menentukan variabel yang berpengaruh.
6. Penarikan kesimpulan dan interpretasi performa kinerja Surv-WLSSVM berbasis estimator Kaplan-Meier.

Cara kerja algoritma PSO ditunjukkan dengan mengaplikasikan algoritma tersebut pada model terdahulu yakni Surv-LSSVM yang dijelaskan pada Bab II. Pengembangan model dalam penelitian ini serta perbandingan kinerjanya dengan metode lain dibahas pada Bab III. Selanjutnya, Model yang telah dikembangkan diaplikasikan pada data *survival* pemberian ASI eksklusif oleh ibu di Indonesia yang dipaparkan pada Bab VI. Hal ini bertujuan mengetahui implementasi dari model tersebut dalam memprediksi variabel berpengaruh serta uji Log-Rank pada indeks prognostik yang dihasilkan dalam model yang telah dikembangkan.

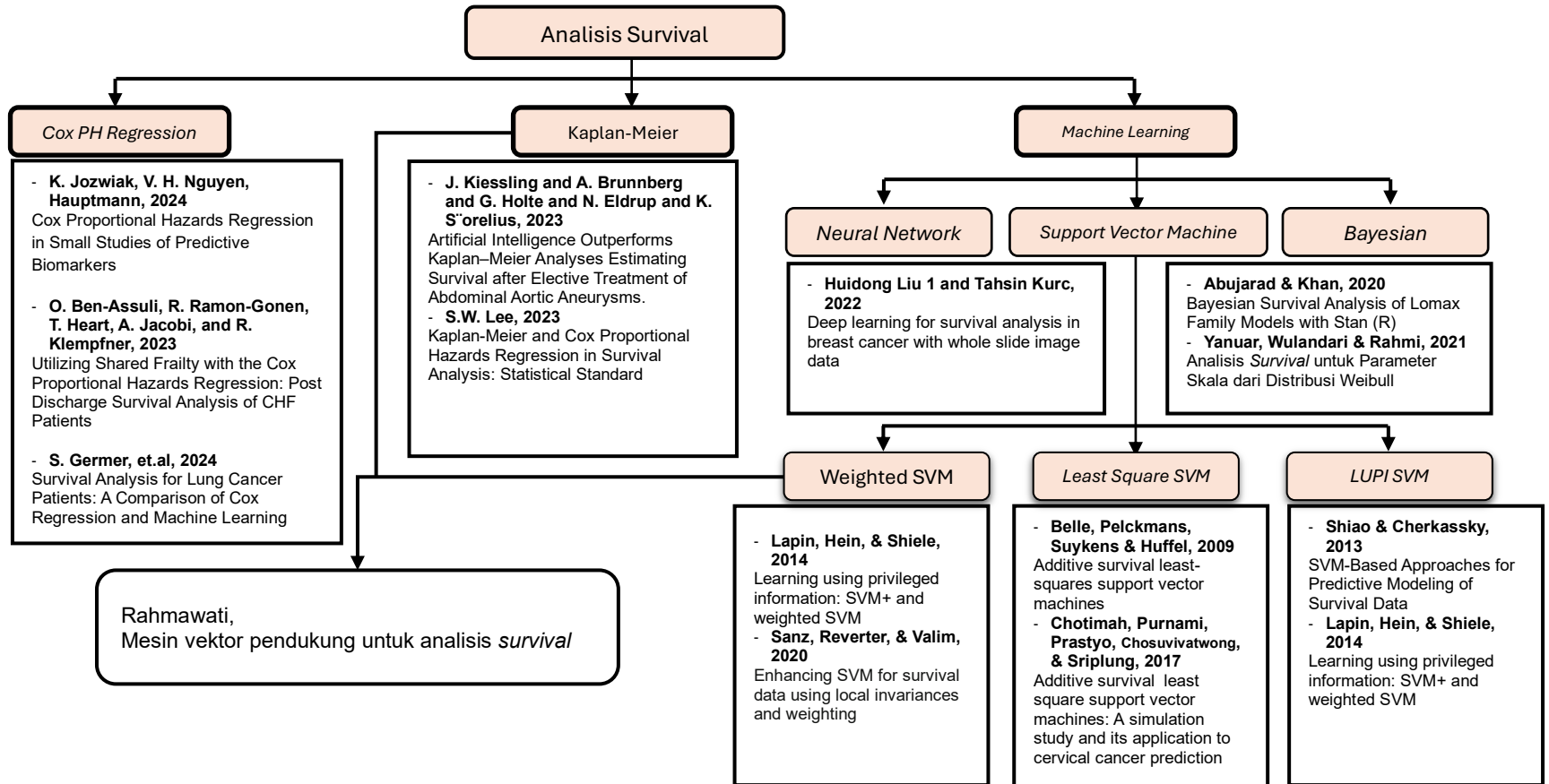
Bagan alur penelitian ditunjukkan dalam Gambar 1.1.



Gambar 1.1. Bagan Alur Penelitian

1.7. Orisinalitas Penelitian

Model *Cox Proportional Hazard Regression* (CPHR) dan Kaplan–Meier merupakan metode klasik analisis *survival* yang berkinerja baik pada data berdimensi kecil, namun keduanya memiliki keterbatasan pada data berdimensi tinggi dan kompleks. Model Cox bergantung pada asumsi *proportional hazard* dan linearitas, sedangkan Kaplan–Meier tidak mampu memodelkan pengaruh multivariat secara simultan. Keterbatasan ini mendorong penggunaan pendekatan *machine learning* untuk analisis *survival*, seperti *Bayesian*, *Neural Network*, *Random Survival Forest*, dan SVM. Metode LS-SVM sebagai pengembangan SVM mampu menangani hubungan non-linear tanpa bergantung pada asumsi statistik model CPHR (Van Belle et al., 2011). Selanjutnya, SVM terboboti menunjukkan potensi peningkatan kinerja melalui pembobotan pengamatan (Sanz et al., 2020b). Disertasi ini merupakan yang pertama secara eksplisit menurunkan bobot berbasis Kaplan–Meier dan mengintegrasikannya ke dalam kerangka optimisasi primal LS-SVM untuk data *survival*. *Research gap* inilah yang melatarbelakangi pengembangan *Weighted Least Square Support Vector Machine* (WLSSVM) dalam penelitian ini. Bagan orisinalitas penelitian ditunjukkan pada Gambar 1.2.



Gambar 1.2. Bagan Orisinalitas Penelitian

BAB II

ALGORITMA *PARTICLE SWARM OPTIMIZATION* UNTUK MENENTUKAN PARAMETER OPTIMAL MODEL *SURVIVAL LEAST SQUARE SUPPORT VECTOR MACHINE*

2.1. Abstrak

Analisis *Survival Least Square Support Vector Machine* (Surv-LSSVM) adalah suatu model yang bekerja dengan cara membentuk suatu fungsi prognostik sebagai fungsi *hazard* pada model *Cox Proportional Hazard Regression* (CPHR). Keoptimalan cara kerja metode ini bergantung pada parameter regularisasi dan parameter *kernel* dalam model yang menghasilkan nilai *Concordance Index* (c-index) yang optimal, semakin besar nilai c-index maka semakin baik model yang digunakan dalam analisis *survival*. *Particle Swarm Optimization* (PSO) adalah algoritma yang melakukan pencarian solusi optimum dengan cara menelusuri perilaku kawanan partikel dan mengeksplorasi ruang pencarian. Kemampuan PSO dalam melakukan pencarian solusi global dengan mengidentifikasi solusi lokal menjadikan algoritma ini mampu menentukan solusi optimum dengan lebih baik. Data yang digunakan dalam penelitian adalah Dataset Pemberian ASI oleh Ibu di Indonesia yang diperoleh dari lembaga survei *The Indonesia Family Life Survei wave 5* (IFLS-5). Data tersebut terdiri dari 12 variabel prediktor dan 718 individu. Hasil penelitian menunjukkan bahwa pada 150 iterasi, parameter regularisasi dan parameter *kernel* optimum masing-masing diperoleh sebesar 0,005 dan 0,246535 dengan nilai c-index sebesar 73,54%, hal ini menunjukkan bahwa parameter yang diperoleh dengan menggunakan algoritma PSO memberikan nilai c-index yang baik digunakan dalam model untuk menghasilkan prediksi yang tepat.

Kata kunci: *Survival Least Square Support Vector Machine*, *Concordance Index*, *Particle Swarm Optimization*, parameter regularisasi, parameter *kernel*

2.2. Pendahuluan

Bab ini menjadi fondasi optimasi parameter bagi model Surv-WLSSVM. Sebagai bahan pembelajaran pencarian parameter, digunakan model Surv-LSSVM. Data yang digunakan dalam proses pembelajaran tersebut adalah data pemberian ASI Eksklusif oleh Ibu di Indonesia. Pada bagian berikut disajikan uraian pendahuluan mengenai studi optimisasi parameter pada model Surv-LSSVM.

Analisis *survival* merupakan metode statistik untuk menganalisis data *time-to-event* yang dapat berupa kematian, kekambuhan penyakit, atau titik akhir penting lainnya. Metode yang paling umum digunakan adalah analisis *Cox Proportional Hazard Regression* (CPHR). Seiring berkembangnya ilmu statistik dan *machine learning*, beberapa pendekatan untuk analisis *survival* telah dilakukan diantaranya penalti *Ridge*, *Support Vector Machine*, *Random Survival Forests*, *Gradient Boosting*, *1-Nearest Neighbor*, dan *Multilayer Perceptron* (Sorayaie Azar et al., 2022a; Tizi & Berrado, 2023a; Fryan & Alazzam, 2023a). SVM merupakan metode *machine learning* yang bekerja sangat baik pada data berdimensi tinggi, khususnya ketika jumlah variabel lebih besar dibandingkan dengan ukuran sampel (Belle et al., 2011; Khotimah et al, 2017; Khoiri et al., 2019). SVM untuk analisis *survival* telah digunakan untuk memodelkan persoalan *time-to-event* berbasis klasifikasi biner (Sanz

et al., 2020a). Selain itu, modifikasi SVM untuk data *survival* berbasis klasifikasi telah diaplikasikan pada data penyakit kegagalan hati (Sujatha et al., 2022b). Pada penelitian dasar penggunaan SVM pada analisis *survival* adalah untuk menentukan fungsi prognostik sebagai fungsi *hazard* pada model CPHR.

Penggunaan fungsi *kernel* memungkinkan metode SVM bekerja lebih baik pada data yang tidak dapat dipisahkan secara linear. Kinerja SVM sangat bergantung pada pemilihan *hyperparameter*. Mengoptimalkan parameter ini sangat penting untuk meningkatkan akurasi model. Teknik optimasi parameter diantaranya *Weighted Superposition Attraction* (WSA), *Particle Swarm Optimization* (PSO), Algoritma *Firefly*, dan Algoritma Genetik. Pendekatan optimisasi ini telah banyak diaplikasikan di berbagai bidang, diantaranya bidang kesehatan seperti prediksi penyakit getah bening (Almayyan, 2016), bidang transportasi tentang masalah penjadwalan bis (Algorithms et al., 2018), dan sistem operasi tenaga listrik (Gaing, 2014). Algoritma PSO terbukti meningkatkan akurasi prediksi kekuatan angin pada metode *hybrid* berbasis model ARMA dan metode *clustering* (Wang et al., 2020). PSO untuk SVM juga digunakan untuk persoalan *feature selection*, pembangkitan data dan menghapus data *noise*, serta prediksi data curah hujan (Tu et al., 2007; Cervantes & Lopez-chau, 2015; Algorithms et al., 2017). Studi komparatif menunjukkan bahwa PSO unggul dibandingkan algoritma genetika dalam hal akurasi, efisiensi iterasi, dan kesederhanaan mekanisme optimisasi (Algorithms et al., 2018; Çarkit & Alçi, 2022). Selain itu, PSO juga unggul untuk pemodelan invers stokastik dalam geofisika (Pace et al., 2022). Penelitian ini mengusulkan pendekatan algoritma PSO untuk menentukan parameter optimum pada model *Least Square Support Vector Machine* (LSSVM) untuk analisis *survival*. Pencarian parameter didasarkan pada nilai *Concordance Index* (c-index) maksimum. Nilai c-index dalam analisis *survival* adalah mengukur keterurutan antara fungsi prognostik yang diperoleh dalam model dan waktu *survival* baik data tersensor maupun data yang tidak tersensor.

2.3. Analisis Survival

Analisis *survival* merupakan analisis data yang memperhatikan waktu dari awal pengamatan sampai suatu *event* terjadi dengan melihat variabel yang memengaruhi *event* tersebut. Terjadinya *event* dapat berupa kematian, munculnya suatu penyakit, atau peristiwa berupa kegagalan lainnya. Selain itu terdapat pula kegagalan yang bersifat positif seperti sembuhnya seseorang dari suatu penyakit dan seseorang yang mendapatkan pekerjaan.

Dataset dari suatu analisis *survival* terdiri dari variabel waktu, status tersensor dan variabel prediktor yang memengaruhi kejadian. Data dengan status tersensor merupakan data yang tidak dapat diamati secara utuh karena subyek pengamatan hilang atau sampai akhir penelitian subyek tersebut tidak mengalami *event* (Kleinbaum, 2012; Elisa & Wang, 2003). Terdapat tiga alasan suatu data mengalami penyensoran, yaitu subyek pengamatan yang diamati tidak mengalami suatu *event* sampai penelitian berakhir (*loss to follow-up*), subyek pengamatan hilang selama penelitian, dan subyek pengamatan ditarik dari penelitian karena meninggal di mana meninggal merupakan suatu peristiwa yang tidak diperhatikan oleh peneliti atau alasan yang lain, misalnya reaksi obat yang buruk atau risiko yang lain (Suykens & Vandewalle, 1999).

Setidaknya terdapat 3 tujuan dasar dari analisis *survival* yaitu (1) menghitung fungsi *survival* dan fungsi *hazard*, (2) membandingkan antara fungsi *survival* dan fungsi *hazard*, dan (3) melihat hubungan antara variabel prediktor terhadap waktu *survival*. Fungsi *survival* dan fungsi *hazard* dijelaskan sebagai berikut:

1. Fungsi *survival*, sering dilambangkan dengan $S(t)$ menggambarkan probabilitas bahwa suatu individu akan "bertahan hidup" atau tidak mengalami peristiwa tersebut hingga waktu t . Secara matematis dituliskan sebagai berikut:

$$S(t) = P(T > t)$$

di mana T adalah suatu variabel acak yang menunjukkan waktu hingga terjadinya *event*.

2. Sementara itu fungsi *hazard* $h(t)$ menunjukkan laju kegagalan individu pada saat t yang dirumuskan sebagai berikut:

$$h(t) = \frac{f(t)}{S(t)}$$

di mana $f(t)$ adalah fungsi kepadatan peluang dari distribusi waktu t .

Metode yang paling umum digunakan dalam menentukan fungsi *survival* adalah Metode Kaplan-Meier. Metode ini menggunakan pendekatan non-parametrik untuk memperkirakan fungsi *survival* dari data yang mengandung status tersensor. Hingga saat ini, metode ini umumnya diterapkan dalam bidang kesehatan (Mikhaeel et al., 2022; Lee, 2023; Jia et al., 2025).

Fungsi *survival* dapat dihitung dengan menggunakan rumus berikut:

$$S(t) = \prod_{i: t_i \leq t} 1 - \frac{d_i}{n_i} \quad (2.1)$$

di mana

t_i : waktu kejadian $ke - i$.

d_i : jumlah peristiwa (misalnya, kematian) pada waktu t_i .

n_i : jumlah individu yang "berisiko" tepat sebelum waktu t_i (jumlah individu yang masih belum mengalami peristiwa dan belum tersensor).

Sementara itu, metode yang paling umum digunakan untuk menentukan fungsi *hazard* adalah *Cox Proportional Hazard Regression* (CPHR) yang menggunakan pendekatan semi-parametrik dan mempertimbangkan pengaruh variabel prediktor.

$$h(t) = h_o(t) \exp \left(\sum_{i=1}^p \beta_i x_i \right) \quad (2.2)$$

di mana

$h_o(t)$: fungsi *baseline hazard* yang menggambarkan laju kegagalan ketika semua variabel prediktor bernilai nol

β_i : parameter regresi yang menunjukkan efek variabel prediktor terhadap laju kegagalan dengan $i = 1, 2, \dots, p$ dan p adalah banyaknya variabel

prediktor

$\exp\left(\sum_{i=1}^p \beta_i x_i\right)$: faktor risiko relatif yang bergantung pada variabel x

Fungsi *hazard* menghasilkan suatu ukuran yang dikenal sebagai *Hazard Ratio* (HR). Dalam analisis *survival*, ukuran ini digunakan untuk membandingkan laju kegagalan antarkelompok. Nilai HR menunjukkan tingkat kecenderungan terjadinya suatu *event* pada satu kelompok dibandingkan dengan kelompok lainnya.

2.4. Support Vector Machine

SVM adalah metode pembelajaran mesin yang merancang cara efisien dalam mempelajari *hyperplane* pemisah terbaik di ruang fitur berdimensi tinggi (Cristianini & Shawe-Taylor, 2000). Prinsip dasar SVM adalah menemukan *hyperplane* yang memisahkan data dengan margin maksimal. *Hyperplane* Optimal adalah *hyperplane* yang memaksimalkan margin, yang menghasilkan pemisahan kelas dan generalisasi yang lebih baik pada data baru (Gunn, 1998).

Fungsi objektif dari model SVM ditunjukkan pada persamaan (2.3)

$$\min_{\mathbf{w}, b, \xi} \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^n \xi_i \quad (2.3)$$

dengan kendala :

$$\mathbf{y}_i(\mathbf{x}_i \mathbf{w} + b) \geq 1 - \xi_i \quad \text{dan} \quad \xi_i \geq 0$$

di mana ξ_i adalah faktor kesalahan dan jika $\xi_i > 0$ maka terjadi kesalahan klasifikasi, C adalah parameter regularisasi, $\mathbf{w}$ adalah vektor bobot dan b adalah deviasi dari titik asal.

Parameter regularisasi C dalam SVM adalah alat penting yang membantu mengontrol kompleksitas model. Memahami dan memilih nilai C yang tepat adalah hal penting dalam membangun model SVM yang lebih baik.

Selanjutnya dibentuk fungsi *Lagrange* untuk *primal problem* yaitu.

$$L_p(\mathbf{w}, b, \xi) = \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^n \xi_i - \sum_{i=1}^n \alpha_i \{ \mathbf{y}_i(\mathbf{x}_i^T \mathbf{w} + b) - 1 + \xi_i \} - \sum_{i=1}^n \mu_i \xi_i \quad (2.4)$$

Persamaan *primal problem* diturunkan terhadap $\mathbf{w}, b, \xi_i$ diperoleh

$$\begin{cases} \frac{\partial L_p}{\partial \mathbf{w}} = 0 ; \quad \mathbf{w} - \sum_{i=1}^n \alpha_i \mathbf{y}_i \mathbf{x}_i = 0 \\ \frac{\partial L_p}{\partial b} = 0 ; \quad \sum_{i=1}^n \alpha_i \mathbf{y}_i = 0 \\ \frac{\partial L_p}{\partial \xi_i} = 0 ; \quad C - \alpha_i - \mu_i = 0 \end{cases} \quad (2.5)$$

Primal problem tidak dapat langsung diselesaikan karena parameter yang tidak diketahui nilainya yaitu α_i , sehingga muncul dua problem. Fungsi *Lagrange dual problem* adalah sebagai berikut :

$$L_D(\alpha) = \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j \mathbf{x}_i^T \mathbf{x}_j \quad (2.6)$$

Dengan demikian dua problem SVM secara lengkap adalah sebagai berikut

$$\max_{\alpha} L_D(\alpha) \quad (2.7)$$

dengan kendala :

$$0 \leq \alpha_i \leq C$$

$$\sum_{i=1}^n \alpha_i y_i = 0$$

Nilai $\alpha > 0$ terjadi pada titik yang berbeda tepat pada garis margin dan disebut sebagai *support vector*, sedangkan titik-titik dengan $\alpha = 0$ berada di luar garis margin.

Persamaan (2.6) menunjukkan terdapat perkalian skalar $\mathbf{x}_i^T \mathbf{x}_j$ yang dalam kasus non-linear SVM dapat ditransformasikan menjadi $\boldsymbol{\varphi}(\mathbf{x}_i)^T \boldsymbol{\varphi}(\mathbf{x}_j)$ dengan menggunakan pendekatan fungsi *kernel*. Transformasi tersebut mendefinisikan fungsi *kernel* sebagai $K(\mathbf{x}_i, \mathbf{x}_j) = \boldsymbol{\varphi}(\mathbf{x}_i)^T \boldsymbol{\varphi}(\mathbf{x}_j)$. Fungsi *kernel* yang sering digunakan adalah fungsi *kernel Radial Basis Function* (RBF) yang secara matematis dituliskan $K(\mathbf{x}_i, \mathbf{x}_j) = \exp\left(-\|\mathbf{x}_i - \mathbf{x}_j\|_2^2 / 2\sigma^2\right)$ (Belle et al., 2011; Khotimah et al., 2017; Khoiri et al., 2019). Fungsi *kernel* dalam analisis *survival* menggambarkan jarak waktu *survival* antar pasien. Semakin kecil nilai *kernel*, maka semakin berbeda karakteristiknya.

Model SVM mengalami perkembangan dengan beberapa modifikasi, salah satunya adalah model LS-SVM yang diperkenalkan oleh Suyken dan Vandewalle (1999). Fungsi tujuan dan fungsi kendala LS-SVM adalah sebagai berikut

$$\min_{\mathbf{w}, \xi} \frac{1}{2} \|\mathbf{w}\|^2 + \frac{1}{2} C \sum_{i=1}^n \xi_i^2 \quad (2.8)$$

dengan kendala:

$$y_i [\boldsymbol{\varphi}(\mathbf{x}_i)^T \mathbf{w} + b] = 1 - \xi_i \quad ; \quad i = 1, 2, \dots, n.$$

Peningkatan kinerja SVM dipengaruhi oleh pemilihan parameter regularisasi dan parameter *kernel* terbaik. Dalam analisis *survival* Kedua parameter dikatakan mencapai optimum jika mencapai c-index terbesar. Metode optimasi parameter yang dapat digunakan diantaranya Algoritma Genetik, *Particle Swarm Optimization* (PSO), dan Algoritma *Firefly*. Algoritma PSO merupakan metode iterasi yang telah banyak digunakan karena kesederhanaan algoritmanya dan mempertimbangkan perilaku historis dari nilai iterasi sebelumnya.

2.5. Analisis *Survival* dengan Pendekatan *Support Vector Machine*

Metode *Support Vector Machine* (SVM) dalam analisis *survival* telah diaplikasikan pada data *survival* berdimensi tinggi (Belle et al., 2011). Konsep dasar *Survival-SVM* adalah menentukan fungsi prognostik sebagai representasi fungsi hazard. Fungsi prognostik tersebut dinyatakan sebagai berikut:

$$\mathbf{u}(\mathbf{x}) = \mathbf{w}^T \boldsymbol{\phi}(\mathbf{x}) \quad (2.9)$$

di mana $u: \mathbb{R}^d \rightarrow \mathbb{R}$, $\mathbf{w}$ adalah suatu vektor parameter, dan $\boldsymbol{\phi}(\mathbf{x})$ merupakan transformasi (*feature map*) dari kovariat $\mathbf{x}$. Berdasarkan *Mercer's condition*, *feature map* tidak perlu terdefinisi secara eksplisit saat menggunakan *kernel* yang bersifat definit positif (Belle et al., 2011; Khotimah et al., 2017; Khoiri et al., 2019).

Fungsi tujuan dan fungsi kendala dari *Survival-SVM* adalah sebagai berikut :

$$\min_{\mathbf{w}, \xi} \quad \frac{1}{2} \mathbf{w}^T \mathbf{w} + \frac{\gamma}{2} \sum_i \sum_{j, i < j} r_{ij} \xi_{ij}; \gamma \geq 0 \quad (2.10)$$

dengan kendala :

$$\begin{aligned} \mathbf{w}^T \boldsymbol{\phi}(\mathbf{x}_j) - \mathbf{w}^T \boldsymbol{\phi}(\mathbf{x}_i) &\geq 1 - \xi_{ij}; \forall i < j \\ \xi_{ij} &\geq 0; \forall i < j \end{aligned}$$

di mana r_{ij} memenuhi:

$$r_{ij} = \begin{cases} 1; (t_i < t_j, \delta_i = 1) \\ 0; (t_i < t_j, \delta_i = 0) \end{cases}$$

yang merupakan indikator pembandingan antara individu $ke - i(\mathbf{x}_i, t_i, \delta_i)$ dan individu $ke - j(\mathbf{x}_j, t_j, \delta_j)$.

begitupun dengan model *Surv-LSSVM*, fungsi tujuan dan fungsi kendala adalah sebagai berikut:

$$\min_{\mathbf{w}, \xi} \quad \frac{1}{2} \mathbf{w}^T \mathbf{w} + \frac{1}{2} \gamma \sum_{i=1}^n \sum_{j, i < j}^n r_{ij} \xi_{ij}^2 \quad (2.11)$$

dengan kendala :

$$\mathbf{w}^T \boldsymbol{\phi}(\mathbf{x}_j) - \mathbf{w}^T \boldsymbol{\phi}(\mathbf{x}_i) = 1 + \xi_{ij}; \quad \forall i, j = 1, 2, \dots, n.$$

Penyelesaian dilakukan dengan membentuk persamaan *Lagrange* dan diturunkan terhadap parameter $\mathbf{w}$, ξ_{ij} dan α . Fungsi prognostik diperoleh sebagai berikut:

$$\hat{\mathbf{u}}^* = \hat{\mathbf{w}}^T \boldsymbol{\phi}(\mathbf{x}^*) = \hat{\alpha}^T \mathbf{D} \mathbf{K}_n(\mathbf{x}^*) \quad (2.12)$$

di mana α adalah pengali Lagrange, $\mathbf{D}$ adalah matriks yang berisi $\{-1,0,1\}$, dan $\mathbf{K}$ adalah matriks *kernel*.

2.6. Ukuran Kebaikan Model

Kriteria kebaikan model dalam analisis *survival* dapat menggunakan ukuran *Concordance Index* (c-index). Nilai c-index merepresentasikan proporsi pasangan waktu *survival* yang dapat diurutkan secara benar, sehingga mencerminkan tingkat ketepatan prediksi model (Jr, 2017; Khotimah et al., 2017; Khoiri et al., 2019). c-index mengukur kesesuaian urutan antara nilai fungsi prognostik dan waktu *survival*, baik pada data tersensor maupun tidak tersensor, serta mencerminkan tingkat kesalahan perankingan. Semakin tinggi nilai c-index, semakin baik kinerja model (Mahjub, 2016). Untuk suatu dataset $D = \{(x_i, t_i, \delta_i)\}_{i=1}^n$, nilai c-index dihitung menggunakan Persamaan (2.13):

$$c_{ij}(u) = \frac{\sum_{i=1}^n \sum_{i < j} r_{ij} I(u(\mathbf{x}_j) - u(\mathbf{x}_i))(t_j - t_i) > 0}{\sum_{i=1}^n \sum_{i < j} r_{ij}} \quad (2.13)$$

di mana r_{ij} telah didefinisikan pada Sub Bab 2.5, I adalah fungsi indikator, $u(x)$ adalah indeks prognostik, dan t adalah waktu *survival*. Fungsi prognostik model Surv-LSSVM yang dinyatakan dalam bentuk matriks dapat dilihat pada persamaan (2.12)

2.7. Particle Swarm Optimization

Particle Swarm Optimization (PSO) merupakan algoritma optimasi yang terinspirasi dari perilaku sosial hewan, seperti sekumpulan burung. Algoritma ini menggunakan populasi partikel yang bergerak dalam ruang pencarian dengan kecepatan yang diperbarui berdasarkan pengalaman terbaik individu (*local best*) dan seluruh populasi (*global best*). Pembaruan ini mengarahkan partikel menuju solusi dengan nilai biaya terendah yang pernah diperoleh (Zhang et al., 2023).

Posisi partikel ditentukan oleh kecepatan dan posisi terbaik di sekitarnya. Melalui proses iterasi, partikel secara bertahap memusatkan pencarian pada area solusi terbaik (Abido, 2017). Elemen dasar dalam algoritma PSO adalah partikel, waktu, populasi, kecepatan partikel, bobot inersia, *fitness*, *personal best*, dan *global best*. Bobot inersia digunakan untuk mengatur keseimbangan antara eksplorasi global dan eksploitasi lokal sehingga proses pencarian menjadi lebih stabil dan efektif (Z. Gaing, 2014, Papazoglou, 2023).

Langkah-langkah dalam algoritma PSO adalah sebagai berikut :

1. Inisialisasi populasi dari partikel dengan posisi dan kecepatan secara acak dalam suatu ruang penelusuran.
2. Evaluasi fungsi *fitness* optimasi yang diinginkan pada setiap partikel.

3. Membandingkan evaluasi *fitness* partikel dengan *personal best* (P_b). Jika nilai yang ada lebih baik dibandingkan dengan nilai P_b , maka P_b diset sama dengan nilai tersebut.
4. *Update* kecepatan dan posisi partikel. Kecepatan dihitung menggunakan persamaan berikut:

$$v_i^{t+1} = \theta v_i^t + \lambda \tau_1 [g^* - z_i^t] + \beta \tau_2 [z_i^{*(k)} - z_i^t] \quad (2.14)$$

di mana

v_i^{t+1} : kecepatan baru

v_i^t : kecepatan awal yang didefinisikan sama dengan 0 pada saat $t = 0$

τ_1 dan τ_2 : bilangan acak pada interval $[0,1]$

λ dan β : parameter pembelajaran atau koefisien percepatan yang umumnya menggunakan $\lambda \approx \beta \approx 2$

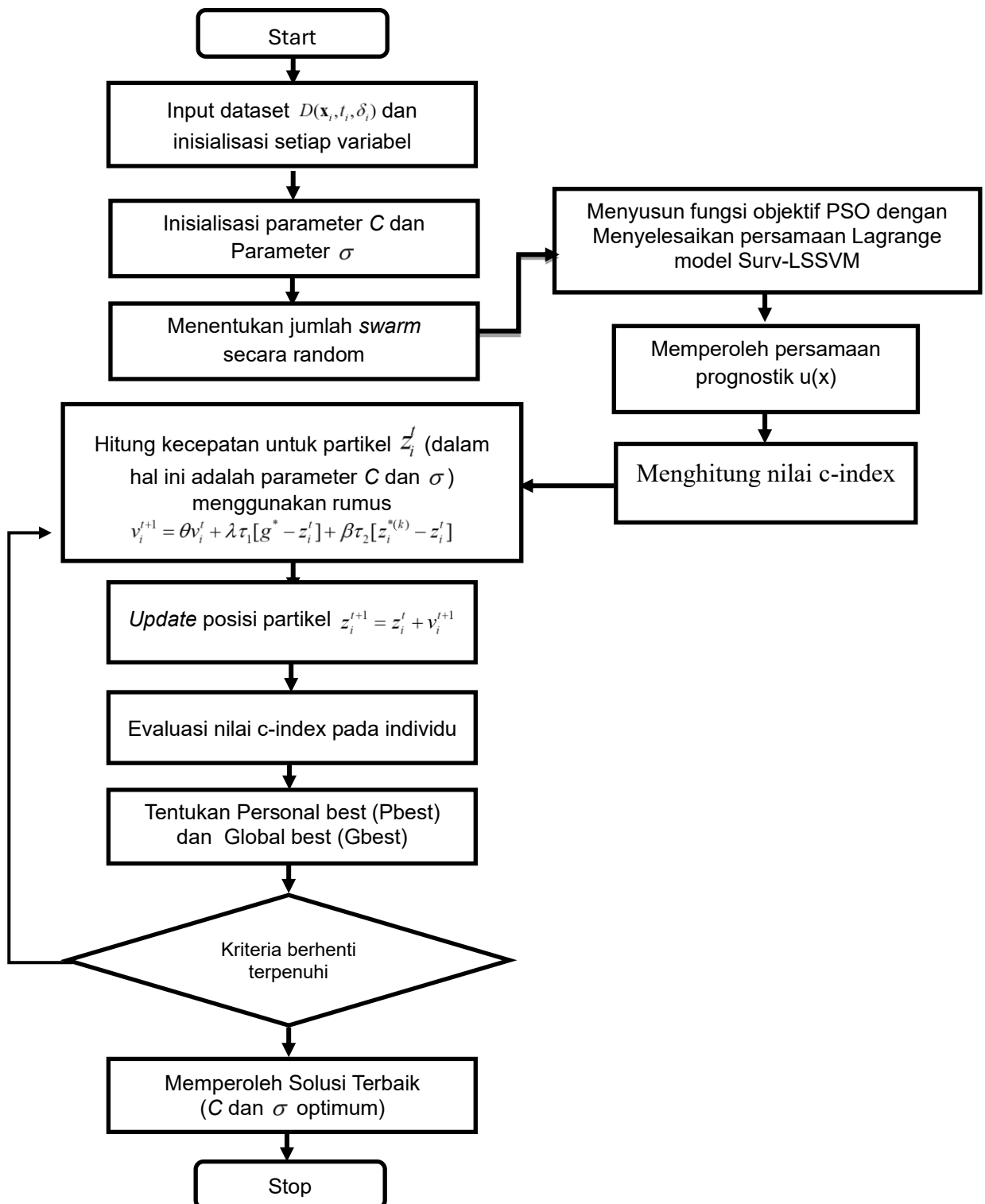
θ : fungsi inersia, konstan dan biasanya menggunakan $\theta \approx 0,5 \sim 0,9$

Posisi partikel baru ditentukan menggunakan persamaan $z_i^{t+1} = z_i^t + v_i^{t+1}$

5. Kembali ke langkah 2 sampai kriteria terpenuhi, umumnya iterasi berhenti pada nilai *fitness* yang cukup baik atau sampai pada jumlah maksimum iterasi.

Setiap partikel bergerak dalam ruang solusi dan menyimpan posisi terbaik yang pernah dicapai selama proses pencarian. Seiring berjalannya iterasi, kawanan partikel secara bertahap memusatkan pencarian pada area ruang solusi yang mengandung solusi dengan hasil terbaik (Ajibade, 2022).

Peran PSO dalam studi ini adalah menentukan parameter regularisasi dan parameter *kernel* pada model Surv-LSSVM untuk memperoleh c-index maksimum. Langkah-langkah algoritma PSO untuk Model Surv-LSSVM disajikan pada Gambar 2.1.



Gambar 2.1. Algoritma PSO model Surv-LSSVM

2.8. Dataset

Penelitian ini menggunakan data pemberian ASI oleh ibu kepada anak yang diperoleh dari lembaga survei *The Indonesia Family Life Survei wave 5* (IFLS-5) dan dapat diakses melalui link <https://www.rand.org>. Setelah dilakukan pra-pemrosesan data diperoleh dataset berukuran 1059 amatan. Dataset tersebut terdiri dari variabel respon dan variabel prediktor. Variabel respon adalah lama pemberian ASI eksklusif oleh ibu pada anak sejak usia 0 tahun. Status tersensor dalam penelitian ini adalah pemberian ASI eksklusif terpenuhi sampai 6 bulan dan status *event* pemberian ASI eksklusif gagal terpenuhi sebelum 6 bulan (Dwi & Astutik, 2019). Variabel prediktor ditunjukkan pada Tabel 2.1.

Table 2.1. Deskripsi variabel prediktor data pemberian ASI

Simbol	Variabel	Skala
x_1	Usia ibu	Interval
x_2	Pendidikan ibu	Ordinal
x_3	Status pernikahan	Nominal
x_4	Status pekerjaan	Nominal
x_5	Pengetahuan ibu	Ordinal
x_6	Tempat melahirkan	Ordinal
x_7	Pembantu persalinan	Nominal
x_8	Pelayanan antenatal	Nominal
x_9	Pelayanan postnatal	Nominal
x_{10}	Kondisi bayi lahir	Ordinal
x_{11}	Keadaan kesehatan	Nominal
x_{12}	Status ekonomi	Ordinal

2.9. Pra-pemrosesan Data

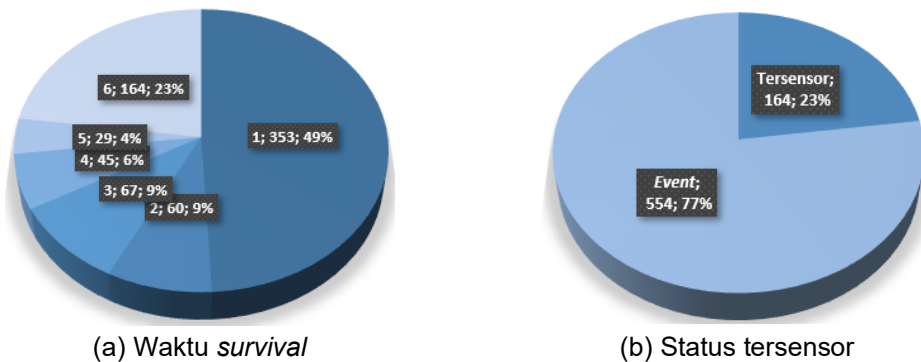
Tahapan pra-pemrosesan data terdiri dari *data collecting*, *data cleaning* dan *label encoding*.

1. *Data collecting*: data pemberian ASI oleh ibu dikumpulkan dari dokumen Household IFLS Wave 5 tahun 2014 berdasarkan 12 variabel penelitian sebagaimana tercantum dalam Tabel 2.1. Setiap variabel ditelusuri sesuai dengan aspek survei yang relevan, dan dilakukan proses *tracking* untuk melengkapi data pada setiap individu.
2. *Data cleaning*: dilakukan dengan menghapus data individu yang memiliki informasi tidak lengkap pada satu atau lebih variabel (*missing data*). Mengingat ukuran sampel yang besar, penghapusan data tersebut tidak memengaruhi representativitas data secara signifikan.
3. *Label encoding*: dilakukan untuk mengonversi variabel kategorik ke dalam bentuk numerik agar dapat digunakan dalam analisis. Setiap kategori diberikan label angka sesuai tingkatannya, misalnya pekerjaan ibu (*bekerja* = 1, *tidak bekerja* = 0) dan pendidikan ibu (*dasar* = 0, *menengah* = 1, *tinggi* = 2), serta diterapkan secara konsisten pada variabel kategorik lainnya.

2.10. Hasil dan Pembahasan

2.10.1. Analisis Deskriptif Dataset ASI

Deskripsi tentang waktu *survival* dan status tersensor dari data pemberian ASI eksklusif oleh ibu diperlihatkan pada Gambar 2.2.



Gambar 2.2. (a) Waktu *survival* dan (b) Status tersensor dataset pemberian ASI

Pemberian ASI oleh ibu dinyatakan dalam satuan bulan yakni berkisar 1 sampai dengan 6 bulan. Sebanyak 353 individu atau sekitar 49% dari jumlah ibu menyusui melakukan pemberian ASI selama 1 bulan, angka ini adalah paling besar dibandingkan dengan waktu *survival* lainnya. Selanjutnya secara berturut-turut persentase jumlah ibu menyusui adalah waktu *survival* 6 bulan sebesar 23%, 3 bulan sebanyak 9%, 2 bulan sebesar 9%, 4 bulan sebesar 6% dan 5 bulan sebesar 4%. Pemberian air putih atau makanan selain ASI sebelum berusia 6 bulan didefinisikan mengalami *event* dan terdapat sebanyak 554 individu atau sekitar 77%. Sedangkan Ibu yang hanya memberikan ASI selama 6 bulan didefinisikan data status tersensor dan terdapat sebanyak 164 atau 23%. Selain variabel waktu *survival* dan status tersensor terdapat beberapa variabel prediktor yang dideskripsikan pada Tabel 2.2.

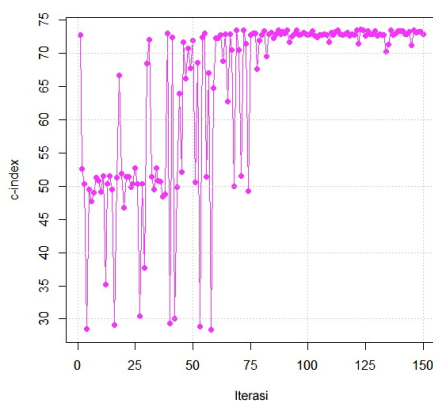
Tabel 2.2. Variabel prediktor data pemberian ASI

No.	Variabel	Kategori	Jumlah
1	Usia Ibu	< 20	102
		21 -35	560
		> 35	56
2	Pendidikan Ibu	Sekolah dasar	106
		Sekolah menengah	477
		Perguruan tinggi	135
3	Status Pernikahan	Menikah	709
		Berpisah	9
4	Status Pekerjaan	Tidak Bekerja	365

		Bekerja	353
5	Pengetahuan ibu	Baik	341
		Kurang Baik	377
6	Tempat Melahirkan	Fasilitas Kesehatan	615
		Lainnya	103
7	Pembantu persalinan	Tenaga profesional	685
		Lainnya	33
8	Pelayanan antenatal	< 4 kali kunjungan	84
		4 kali kunjungan	634
9	Pelayanan postnatal	Tidak	399
		Ya	319
10	Kondisi bayi lahir	Lebih besar	11
		Sama dengan lainnya	705
		Lebih kecil	2
11	Keadaan kesehatan	Tidak ada komplikasi	402
		Terdapat penyakit lainnya	316
12	Status ekonomi	Menggunakan asuransi/ keringanan pembiayaan	276
		Tidak menggunakan asuransi/ keringanan pembiayaan	442

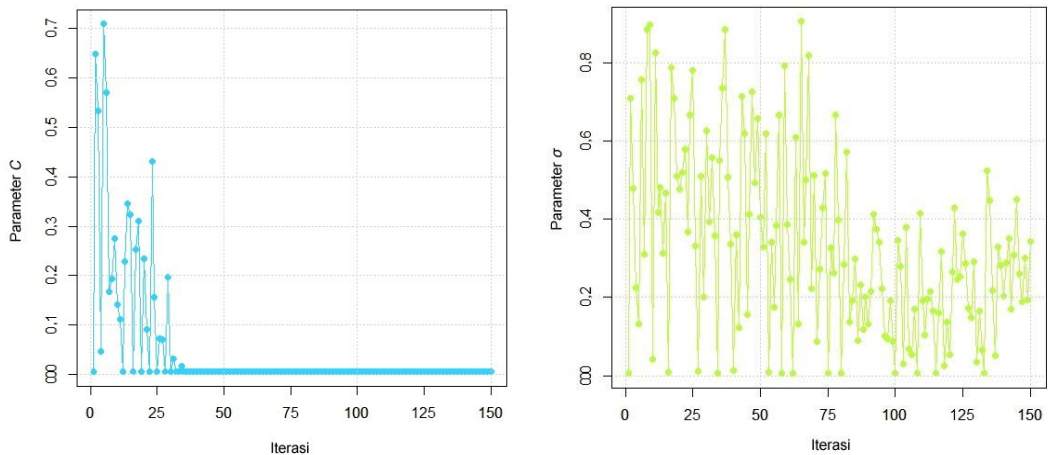
2.10.2. Pencarian Parameter Model Surv-LSSVM

Algoritma PSO dilakukan dengan terlebih dahulu menentukan *swarm* atau jumlah partikel secara acak, partikel dalam penelitian ini adalah parameter C dan σ . Kedua parameter tersebut dipilih secara acak dalam interval $[0,005 ; 1]$ dengan nilai awal masing-masing adalah 0,005. Perubahan posisi kedua partikel adalah untuk menemukan nilai *c-index* optimum. *Tuning parameter* model pada 150 iterasi diperlihatkan pada Gambar 2.3.



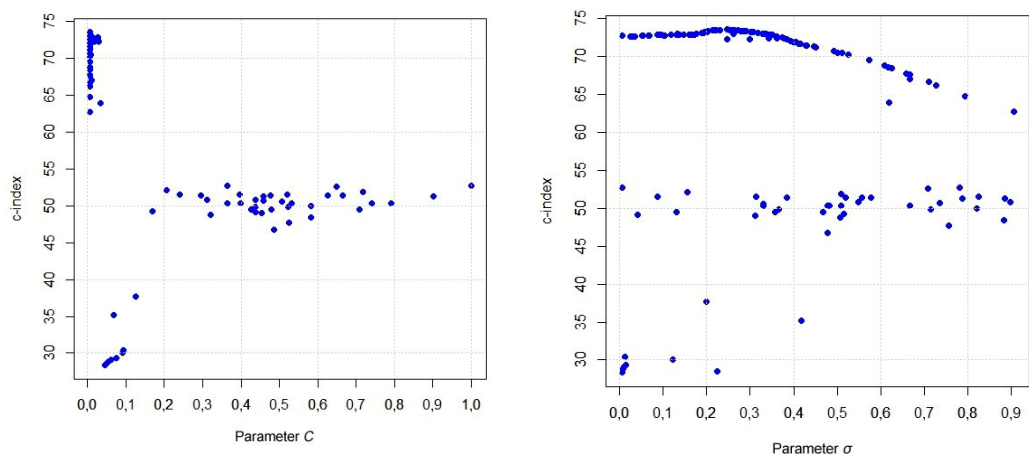
Gambar 2.3. c-index model Surv-LSSVM

Nilai c-index berfluktuasi cukup besar untuk menemukan nilai c-index optimum. Kombinasi nilai parameter C dan σ menentukan nilai c-index yang relatif konvergen pada iterasi ke-75. Nilai c-index maksimum diperoleh sebesar 73,542757 atau sekitar 73,54%. Kecenderungan nilai parameter C dan σ diperlihatkan pada Gambar 2.4.



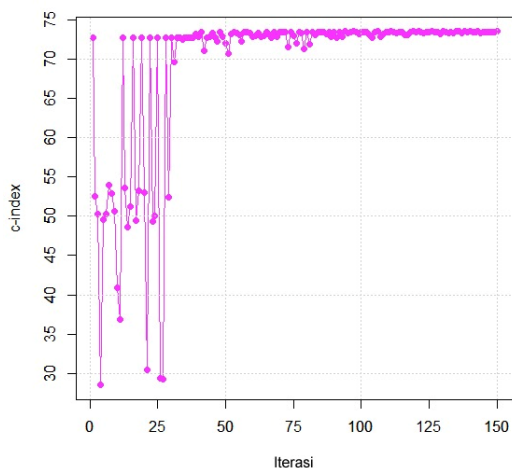
Gambar 2.4. Parameter C dan σ model Surv-LSSVM

Nilai parameter C berfluktuasi sangat besar di iterasi awal, hal ini disebabkan oleh pencarian partikel dalam algoritma PSO untuk menemukan solusi global yang memaksimalkan c-index. Kemampuan PSO untuk tidak terjebak pada solusi lokal dan terus menerus menemukan pencarian solusi global menjadikan metode unggul dalam menentukan nilai optimum. Kekonvergenan nilai parameter C terlihat pada saat mencapai iterasi ke-150 dengan nilai C optimum sebesar 0,005. Hal yang sama untuk parameter σ , nilai σ yang optimum diperoleh sebesar 0,246535. kecenderungan nilai kedua parameter tersebut dalam menemukan c-index optimum diperlihatkan pada Gambar 2.5.

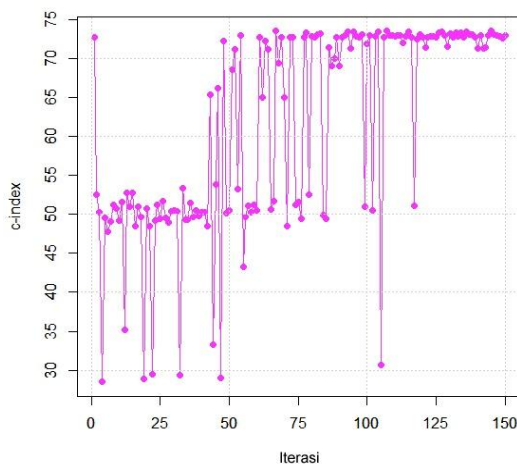


Gambar 2.5. Kecenderungan parameter C dan parameter σ untuk c-index optimum

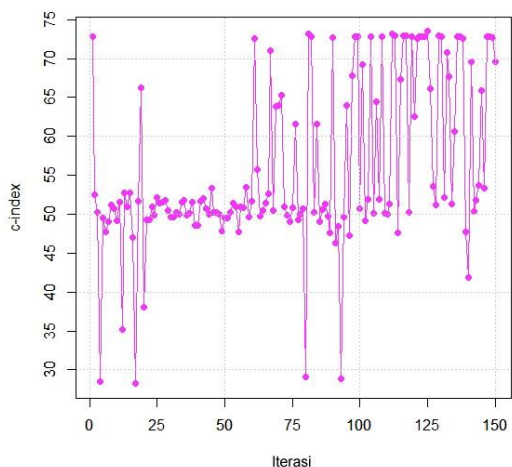
Berdasarkan Gambar 2.5, nilai c-index cenderung mencapai nilai optimal saat C berada pada rentang 0 sampai 0,1. Berbeda dengan parameter C , nilai c-index cenderung berfluktuasi sangat besar di sepanjang nilai σ yaitu 0 sampai dengan 1. Fluktuasi nilai kedua parameter dipengaruhi oleh pergerakan partikel dalam PSO berdasarkan posisi dan kecepatan partikel di setiap iterasi. Selanjutnya, Gambar 2.6, Gambar 2.7, dan Gambar 2.8 secara berturut-turut menyajikan pergerakan nilai c-index, parameter C , dan parameter σ pada swarm yang berbeda.



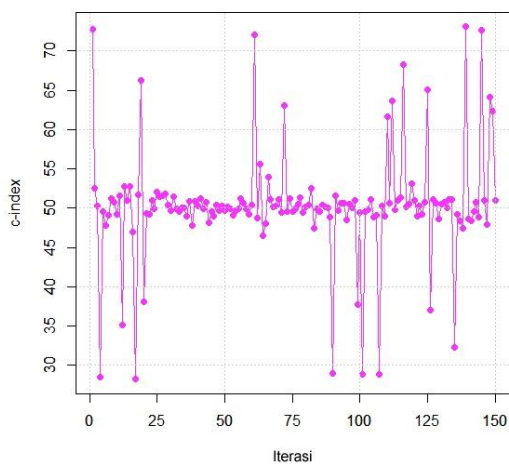
Swarm = 5



Swarm = 10

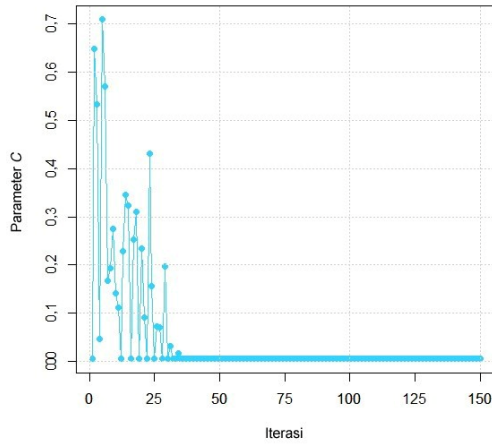


Swarm = 30

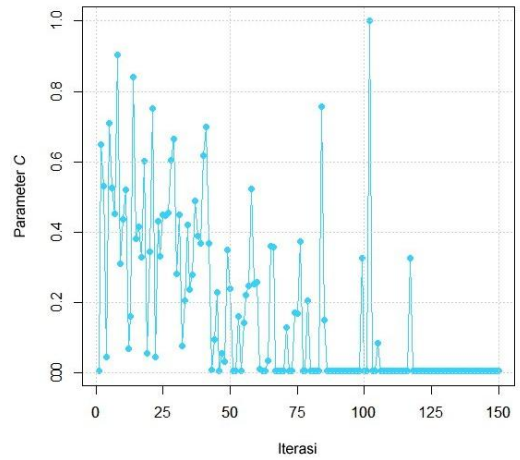


Swarm = 50

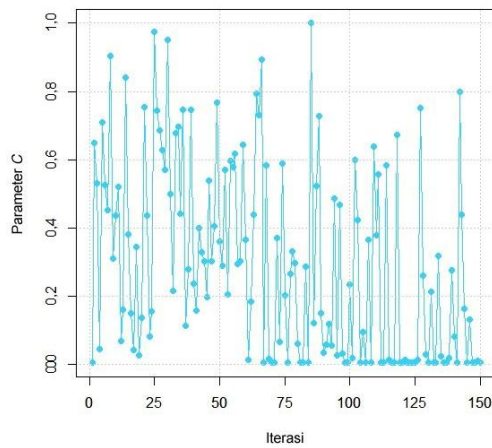
Gambar 2.6. c-index dengan *swarm* berbeda



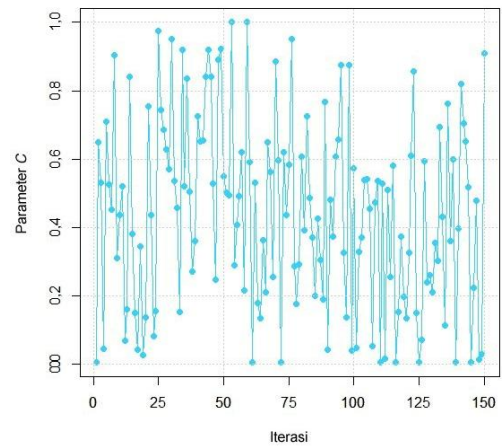
Swarm = 5



Swarm = 10

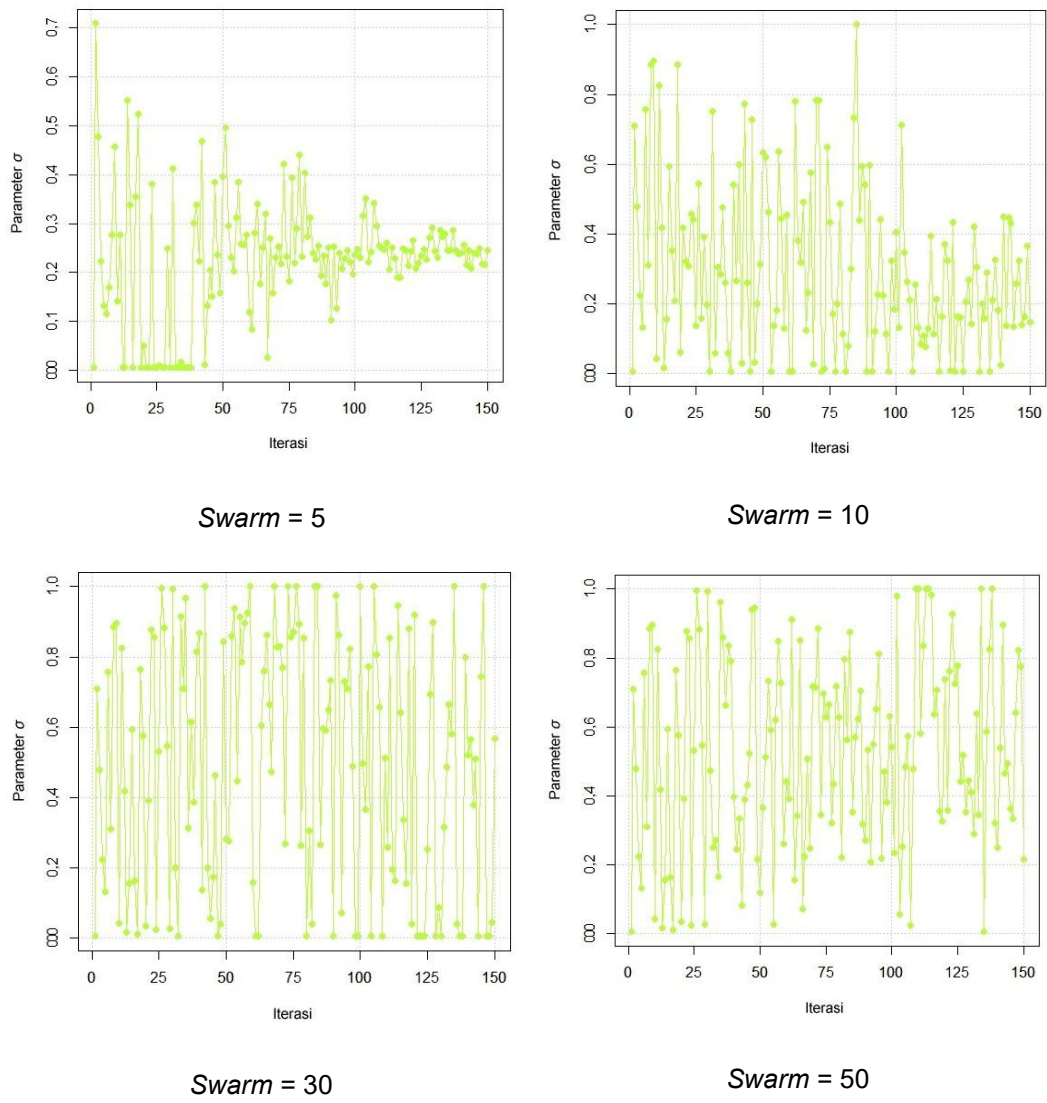


Swarm = 30



Swarm = 50

Gambar 2.7. Parameter C dengan *swarm* berbeda



Gambar 2.8. Parameter σ dengan *swarm* berbeda

Tabel 2.3 menunjukkan c-index maksimum untuk setiap *swarm* yang diujikan dalam model Surv-LSSVM.

Tabel 2.3. c-index maksimum untuk setiap *swarm*

<i>Swarm</i>	Iterasi optimum	c-index
5	143	73,542757
15	67	73,498006
30	125	73,522364
50	139	73,072587

Berdasarkan Tabel 2.3, nilai c-index yang diperoleh tidak menunjukkan perbedaan yang signifikan pada setiap ukuran *swarm*. *Swarm* dengan ukuran lebih kecil memungkinkan proses penelusuran ruang solusi berlangsung lebih cepat dibandingkan *swarm* yang lebih besar. Namun, seluruh konfigurasi *swarm* yang digunakan tetap menghasilkan nilai c-index terbaik, karena mekanisme pembaruan posisi dan kecepatan partikel dalam PSO mampu menjaga keseimbangan antara eksplorasi dan eksploitasi ruang pencarian sehingga konvergen pada solusi optimal yang serupa.

2.10.3. Pemilihan Variabel Model Surv-LSSVM

Analisis dilakukan untuk mengidentifikasi variabel yang memengaruhi pemberian ASI oleh Ibu. Setiap variabel dihapus dalam model Surv-LSSVM dan selanjutnya dilakukan perhitungan nilai c-index. Selisih nilai c-index sebelum dan setelah variabel dihapus merupakan ukuran seberapa besar pengaruh variabel tersebut dalam model. Perhitungan c-index untuk setiap variabel yang dihapus dilakukan dengan menggunakan nilai parameter yang diperoleh pada algoritma PSO. Hasil perhitungan tersebut disajikan pada Tabel 2.4.

Tabel 2.4. c-index setelah variabel dihapus dari model

Variabel Yang Dihapus	c-index Model melibatkan semua variabel	c-index setelah variabel dihapus	Selisih nilai c-index	HR	Log-Rank
x_1	73,542757	60,47346	13,06930	1,490811	20,9
x_2	73,542757	72,00536	1,537397	5,207303	304
x_3	73,542757	73,38471	0,158047	7,320529	364
x_4	73,542757	72,38461	1,158147	6,166396	337
x_5	73,542757	71,46438	2,078377	5,860961	323
x_6	73,542757	71,45220	2,090557	5,783786	320
x_7	73,542757	73,08505	0,457707	7,255549	361
x_8	73,542757	72,90520	0,637557	5,288719	305
x_9	73,542757	71,28566	2,257097	6,165908	335
x_{10}	73,542757	73,36998	0,172777	7,010644	355
x_{11}	73,542757	70,67614	2,866617	5,042102	294
x_{12}	73,542757	71,18115	2,361607	4,890741	285

Berdasarkan Tabel 2.4, penurunan nilai c-index paling besar jika variabel tersebut dihapus adalah variabel usia ibu (x_1), hal ini berarti variabel usia merupakan faktor terbesar yang memengaruhi pemberian ASI oleh ibu. Secara berurut variabel yang memiliki pengaruh penurunan c-index adalah kondisi kesehatan (x_{11}), status ekonomi (x_{12}), tempat melahirkan (x_6), pelayanan postnatal (x_9) Pengetahuan Ibu (x_5), Pendidikan ibu (x_2), pekerjaan ibu (x_4), pelayanan antenatal (x_6), tenaga pembantu persalinan (x_7), kondisi bayi lahir (x_{10}), dan status pernikahan (x_3).

2.11. Kesimpulan

Studi ini menggambarkan pencarian parameter regularisasi dan parameter *kernel* dalam model Surv-LSSVM. Nilai parameter optimal adalah yang menghasilkan nilai *Concordance Index* (c-index) maksimum. Algoritma PSO bekerja sangat baik untuk menemukan nilai parameter optimum. Hal tersebut dapat dilihat dari nilai c-index yang diperoleh, di mana nilai tersebut mengindikasikan bahwa model memiliki kemampuan yang baik dalam memprediksi fungsi prognostik. Kemampuan algoritma PSO dalam menemukan parameter optimum dipengaruhi oleh parameter yang digunakan untuk menghitung kecepatan partikel. Pemilihan rentang nilai dan jumlah partikel dalam *swarm* juga dapat meningkatkan kinerja PSO, namun pemilihan yang kurang tepat dapat meningkatkan beban komputasi dan memperlambat konvergensi. Model Surv-LSSVM yang diintegrasikan dengan algoritma PSO menghasilkan c-index yang lebih besar dari 70%. Model diimplementasikan pada pemberian ASI eksklusif oleh ibu di Indonesia dan menghasilkan dua variabel yang paling berpengaruh, yakni variabel usia ibu dan kondisi kesehatan ibu.

2.12. Daftar Pustaka

- Abido, M. A. (2002). Optimal power flow using particle swarm optimization. *International Journal of Electrical Power & Energy Systems*, 24(7), 563–571. [https://doi.org/10.1016/S0142-0615\(01\)00067-9](https://doi.org/10.1016/S0142-0615(01)00067-9)
- Algorithms, E., Kakkar, M. K., & Singla, J. (2018). Performance comparison of genetic algorithms and particle swarm optimization for model integer programming bus timetabling problem performance comparison of genetic algorithms and particle swarm optimization for model integer programming bus timetabling p. <https://doi.org/10.1088/1757-899X/332/1/012020>.
- Algorithms, P., Du, J., Liu, Y., Yu, Y., & Yan, W. (2017). A Prediction of precipitation data based on support vector machine and particle swarm optimization. <https://doi.org/10.3390/a10020057>.
- Almayyan, W. (2016). Lymph diseases prediction using random forest and particle swarm optimization. *Journal of Intelligent Learning Systems and Applications*, 8(3), 51–62. <https://doi.org/10.4236/jilsa.2016.83005>
- Belle, V. Van, Pelckmans, K., Huffel, S. Van, & Suykens, J. A. K. (2011). Improved performance on high-dimensional survival data by application of Survival-SVM. 27(1), 87–94. <https://doi.org/10.1093/bioinformatics/btq617>.
- Çarkıt, T., & Alçı, M. (2022). Comparison of the performances of heuristic optimization algorithms PSO, ABC and GA for parameter estimation in the discharge processes of Li-NMC battery. *Journal of Energy Systems*, 6(3), 387–400. <https://doi.org/10.30521/jes.1094106>
- Cervantes, J., & Lopez-chau, A. (2015). PSO-Based Method for SVM classification on skewed Data-Sets PSO-Based. March 2016. <https://doi.org/10.1007/978-3-319-22053-6>
- Cristianini, N., & Shawe-Taylor, J. (2000). An introduction to support vector machines and other *kernel*-based learning methods. Cambridge University Press.

- Dwi Tama, T., & Astutik, E. (2019). Exclusive breastfeeding survival and factors related to early breastfeeding cessation in Indonesia. In *Proceedings of the 2nd International Conference on Sports Sciences and Health 2018 (Advances in Health Sciences Research, Vol. 7, pp. 183–186)*. Atlantis Press. <https://doi.org/10.2991/icssh-18.2019.42>
- Elisa, T.L. & Wang, J.W. (2003). *Statistical methods for survival data analysis*. (2003). John Wiley & Sons. New Jersey. Canada.
- Fryan, L. H. Al, & Alazzam, M. B. (2023). Survival Analysis of Oncological Patients Using Machine Learning Method. *Healthcare (Switzerland)*, 11(1). <https://doi.org/10.3390/healthcare11010080>.
- Gaing, Z. (2014). Particle Swarm Optimization to Solving the Economic Dispatch Considering the Generator Constraints. 18(February), 1187–1195. <https://doi.org/10.1109/TPWRS.2003.814889>.
- Gunn, S. R. (1998). Support vector machines for classification and regression (Technical Report, Faculty of Engineering, Science and Mathematics, School of Electronics and Computer Science, University of Southampton). University of Southampton. <http://eprints.soton.ac.uk/id/eprint/256459>
- Jia, Y., Li, C., Feng, C., Sun, S., Cai, Y., Yao, P., Wei, X., Feng, Z., Liu, Y., Lv, W., Wu, H., Wu, F., Zhang, L., Zhang, S., & Ma, X. (2025). Prognostic prediction for inflammatory breast cancer patients using random survival forest modeling. *Translational Oncology*, 52. <https://doi.org/10.1016/j.tranon.2024.102246>
- Khoiri, H. A., Prastyo, D. D., & Purnami, S. W. (2019). A simulation study and application of feature selection on survival least square support vector machines. *AIP Conference Proceedings*, 2138. <https://doi.org/10.1063/1.5121121>
- Khotimah, C., Purnami, S. W., Prastyo, D. D., Chosuvivatwong, V., & Sriplung, H. (2017). Additive survival least square support vector machines: A simulation study and its application to cervical cancer prediction. *AIP Conference Proceedings*, 1905. <https://doi.org/10.1063/1.5012243>
- Kleinbaum, D. G. (2012). *Survival Analysis Self Learning (Third Edit)*. Springer. London
- Lee, S. W. (2023). Kaplan–Meier and Cox proportional hazards regression in survival analysis: Statistical standard and guideline of Life Cycle Committee. *Life Cycle*, 3. <https://doi.org/10.54724/lc.2023.e8>
- Mikhaeel, N. G., Heymans, M. W., Eertink, J. J., De Vet, H. C. W., Boellaard, R., Uhrsen, U. D., Ceriani, L., Schmitz, C., Wieggers, S., Uttmann, H., Lugtenburg, P. J., Zucca, E., Zwezerijnen, J. C., Hoekstra, O. S., Joshi, J., & Barrington, S. F. (2022). Proposed new dynamic prognostic index for diffuse large B-cell lymphoma: International Metabolic Prognostic Index. *Journal of Clinical Oncology*, 40, 2352–2360. <https://doi.org/10.1200/JCO.21>
- Sanz, H., Reverter, F., & Valim, C. (2020b). Enhancing SVM for survival data using local invariances and weighting. *BMC Bioinformatics*, 21(1). <https://doi.org/10.1186/s12859-020-3481-2>
- Sorayaie Azar, A., Babaei Rikan, S., Naemi, A., Bagherzadeh Mohasefi, J., Pirnejad, H., Bagherzadeh Mohasefi, M., & Wiil, U. K. (2022a). Application of machine learning

- techniques for predicting survival in ovarian cancer. *BMC Medical Informatics and Decision Making*, 22(1), 1–24. <https://doi.org/10.1186/s12911-022-02087-y>
- Sujatha, R., Chatterjee, J. M., Jhanjhi, N. Z., Tabbakh, T. A., & Almusaylim, Z. A. (2022b). Heart failure patient survival analysis with multi *kernel* support vector machine. *Intelligent Automation and Soft Computing*, 32(1), 115–129. <https://doi.org/10.32604/iasc.2022.019133>.
- Suykens, J.A.K. dan Vandewalle, J. (1999), “Least Squares Support Vector Machine Classifiers”, *Neural Processing Letters*, 9, 293-300.
- Tizi, W., & Berrado, A. (2023a). Machine learning for survival analysis in cancer research: A comparative study. *Scientific African*, 21, e01880. <https://doi.org/10.1016/j.sciaf.2023.e01880>
- Tu, C., Chuang, L., Chang, J., & Yang, C. (2007). Feature selection using PSO-SVM. *IAENG International Journal of Computer Science*, 33(1).
- Wang, Y., Wang, D., Tang, Y. (2020). Clustered hybrid wind power prediction model based on ARMA, PSO-SVM, and clustering methods. *IEEE Access*, 8, 17071–17079. <https://doi.org/10.1109/ACCESS.2020.2968390>
- Zhang, K., Xu, Y., Gao, W., Yu, K., & Liu, Y. (2023). Optimizing MRT data processing via comparative analysis of SA and PSO algorithms: A simulation and numerical study. *Optics Express*, 31(13), 20905. <https://doi.org/10.1364/oe.493697>

