

Evaluasi Performa Algoritma Logistic Regression Dan Support Vector Machine Classifier Dalam Kasus Perceraian

Felice Wang¹, Dewi Nasien²

¹Institut Bisnis dan Teknologi Pelita Indonesia, felice.wang@student.pelitaindonesia.ac.id

²Institut Bisnis dan Teknologi Pelita Indonesia, devi.nasien@lecturer.pelitaindonesia.ac.id

INFORMASI ARTIKEL

Sejarah Artikel:

Diterima Redaksi: 01 Desember 2025

Revisi Akhir: 30 Maret 2026

Diterbitkan Online: 30 April 2026

KATA KUNCI

Perceraian, Logistic Regression, Support Vector Machine, Machine Learning, Evaluasi Performa

KORESPONDENSI

E-mail:

felice.wang@student.pelitaindonesia.ac.id

ABSTRACT

Kasus perceraian yang terus meningkat setiap tahunnya menjadi perhatian penting bagi para peneliti, praktisi hukum, konselor keluarga, hingga pemerintah karena membawa dampak sosial dan psikologis yang signifikan. Penelitian ini bertujuan untuk menganalisis serta memprediksi kemungkinan terjadinya perceraian dengan memanfaatkan pendekatan *machine learning*. Dua algoritma klasifikasi yang dibandingkan dalam studi ini adalah *Logistic Regression* dan *Support Vector Classifier*, yang keduanya dikenal efektif dalam tugas prediksi berbasis data. Dataset penelitian mencakup berbagai indikator kualitas hubungan, antara lain tingkat kepercayaan, pola komunikasi, kepuasan emosional, intensitas konflik, serta variabel lain yang mencerminkan dinamika pasangan. Proses penelitian melibatkan tahapan pra-pemrosesan data, termasuk pembersihan, normalisasi, dan pembagian data latih serta uji. Kinerja kedua algoritma kemudian dievaluasi menggunakan metrik akurasi, presisi, *recall*, *F1-score*, dan analisis *confusion matrix*. Hasil menunjukkan bahwa *Logistic Regression* berhasil mencapai akurasi 100%, sedangkan *Support Vector Classifier* memberikan akurasi 97%. *Logistic Regression* dinilai unggul dari segi interpretabilitas dan efisiensi komputasi, sehingga sangat sesuai untuk model yang membutuhkan penjelasan yang mudah dipahami. Sementara itu, *Support Vector Classifier* menunjukkan performa yang kuat untuk pola data yang lebih kompleks dan berdimensi tinggi. Dengan demikian, pemilihan algoritma terbaik sangat bergantung pada karakteristik dataset serta tujuan analisis yang ingin dicapai.

1. PENDAHULUAN

Perceraian secara konseptual didefinisikan sebagai putusnya ikatan perkawinan yang sah antara suami dan istri melalui keputusan pengadilan yang berkekuatan hukum tetap, sehingga mengakhiri hak dan kewajiban mereka sebagai pasangan [1]. Secara yuridis di Indonesia, perceraian diatur oleh Undang-Undang Nomor 1 Tahun 1974 tentang Perkawinan, yang menegaskan bahwa perkawinan putus karena kematian, perceraian, dan atas putusan pengadilan. Beberapa penyebab utama peningkatan jumlah perceraian adalah komunikasi yang buruk, stres ekonomi, kurangnya dukungan sosial, dan perbedaan nilai dalam hubungan [2].

Pada awal periode, yaitu tahun 2008, jumlah perceraian masih tergolong rendah, sekitar 200 ribu kasus. Namun, sejak 2010 hingga 2015, angka tersebut terus meningkat dan stabil di kisaran 300 hingga 350 ribu kasus per tahun. Kecenderungan ini menunjukkan adanya perubahan dinamika sosial dan ekonomi yang turut memengaruhi keharmonisan rumah tangga. Peningkatan yang lebih tajam tampak pada 2018–2019, ketika jumlah perceraian menembus lebih dari 400 ribu kasus. Penurunan sempat terjadi pada tahun 2020, kemungkinan akibat

pembatasan kegiatan masyarakat selama pandemi *COVID-19* yang berdampak pada proses pengajuan perkara di pengadilan.

Setelah pandemi mereda, jumlah perceraian kembali melonjak dan mencapai puncaknya pada 2022 dengan lebih dari 500 ribu kasus, kemudian sedikit menurun pada 2023 menjadi sekitar 460 ribu kasus.

Secara keseluruhan, data ini mengindikasikan bahwa kasus perceraian di Indonesia cenderung meningkat dalam jangka panjang. Fenomena tersebut dapat dikaitkan dengan berbagai faktor, seperti perselisihan rumah tangga yang berkepanjangan, tekanan ekonomi, perubahan nilai dan peran sosial, serta meningkatnya kesadaran hukum terutama di kalangan perempuan.

Machine learning (ML) telah digunakan untuk memahami dan memprediksi fenomena sosial, seperti perceraian. Penelitian menunjukkan kemajuan besar dalam aplikasi ML pada masalah sosial dan keluarga dengan mengidentifikasi tanda-tanda awal perceraian melalui analisis data kuesioner pasangan dan faktor sosial dan psikologis [3]. Sebab jenis data perceraian, seperti Set Data Prediktor Perceraian, memiliki fitur psikososial numerik dan kategorikal, kedua algoritma yang sering digunakan dalam prediksi perceraian adalah *Logistic Regression* (LR) dan *Support Vector Classifier* (SVC).

LR berhasil ketika hubungan antara variabel independen dan dependen bersifat linier dan mudah diinterpretasikan, sehingga

peneliti dapat mengetahui kontribusi masing-masing faktor risiko terhadap perceraian [4]. Sebaliknya, *Support Vector Classifier* (SVC) berhasil ketika data berdimensi menghasilkan jarak terhadap *hyperplane* sehingga interpretasinya berfokus pada margin dan pemisahan data. [4]

Penelitian ini akan membandingkan secara langsung algoritma LR dan SVC untuk menentukan metode yang paling efisien untuk menemukan tanda-tanda awal perceraian. Penelitian ini diharapkan dapat membantu bidang psikologi keluarga, layanan sosial, dan kebijakan publik mengurangi perceraian.

2. TINJAUAN PUSTAKA

2.1. Pengertian Perceraian

Menurut hasil penelitian, perceraian diartikan sebagai berakhirnya ikatan perkawinan antara suami dan istri baik secara hukum maupun secara agama. Penelitian tersebut mengungkapkan bahwa perceraian merupakan bentuk kegagalan pasangan dalam menjaga keharmonisan rumah tangga yang umumnya disebabkan oleh konflik berkepanjangan, tekanan ekonomi, dan komunikasi yang tidak efektif. Berdasarkan data dari Badan Pusat Statistik, pada tahun 2022 tercatat lebih dari 500 ribu kasus perceraian di Indonesia, menunjukkan adanya peningkatan yang signifikan dari tahun sebelumnya serta menjadikan perceraian sebagai persoalan sosial yang perlu mendapatkan perhatian serius [5].

Dari paragraf tersebut, dapat disimpulkan bahwa perceraian tidak hanya sekadar proses hukum untuk mengakhiri hubungan suami istri, tetapi juga merupakan fenomena sosial dan psikologis yang timbul akibat adanya ketidakharmonisan, kesenjangan ekonomi, serta perbedaan nilai-nilai dalam kehidupan rumah tangga.

2.2. Pengertian Klasifikasi

Klasifikasi merupakan salah satu metode dalam ML yang berfungsi untuk mengelompokkan data ke dalam kategori atau kelas tertentu berdasarkan atribut atau karakteristik yang dimiliki. Proses ini dilakukan dengan menggunakan data latih yang telah diberi label, sehingga model dapat mempelajari pola-pola yang ada dan memprediksi kelas dari data baru yang belum dikenal. Dengan kata lain, klasifikasi termasuk dalam metode *supervised learning*, karena keberhasilan model sangat bergantung pada kualitas dan kuantitas data berlabel yang digunakan selama pelatihan [6].

Metode klasifikasi memiliki peran penting di berbagai bidang, misalnya dalam diagnosa medis untuk menentukan penyakit pasien berdasarkan gejala, dalam analisis teks untuk mengidentifikasi kategori dokumen, maupun dalam sistem deteksi penipuan finansial. tujuan utama klasifikasi adalah membangun model prediktif yang mampu mengenali pola dari data historis, sehingga dapat digunakan untuk pengambilan keputusan yang lebih cepat dan tepat [7].

2.2.1. Logistic Regression

LR adalah teknik statistik digunakan untuk memprediksi kemungkinan kejadian biner seperti perceraian. Metode ini menggunakan fungsi logistik untuk menghitung hubungan antara variabel independen dan dependen. Dalam analisis data sosial, LR dapat menjelaskan pengaruh variabel terhadap hasil klasifikasi. LR mudah untuk dipahami dan diinterpretasi, dan memiliki proses pelatihan cepat dan efisien. LR menghasilkan

output probabilitas dan cocok dalam hubungan linier antara fitur dan kelas.

Fungsi logistik sering digunakan dalam LR untuk memprediksi probabilitas terjadinya suatu peristiwa [8]. Fungsi ini berbentuk sigmoid, sehingga respons terhadap perubahan variabel prediktor bersifat non-linier, dengan *output* selalu berada di antara 0 dan 1. Secara matematis, fungsi logistik dinyatakan sebagai salah satu fungsi utama LR yang digunakan untuk memodelkan probabilitas terjadinya peristiwa biner. Fungsi logistik diwakili sebagai berikut:

$$P(Y = 1|X) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n)}} \quad (2.1)$$

Dimana:

$P(Y=1|X)$ adalah probabilitas terjadinya peristiwa.

β_0 adalah *intercept*.

$\beta_1, \beta_2, \dots, \beta_n$ adalah koefisien regresi untuk masing-masing variabel independen.

2.2.2. Support Vector Classifier

SVC adalah algoritma pembelajaran mesin yang digunakan untuk menyelesaikan masalah regresi dan klasifikasi. Mencari *hyperplane* optimal, yang dapat memisahkan kelas data dengan margin maksimum, adalah prinsip utama SVC. Ini memungkinkan model untuk melakukan generalisasi yang baik terhadap data baru [9].

Linier, polinomial, dan *Radial Basis Function* (RBF) kernel adalah beberapa jenis fungsi kernel yang biasa digunakan oleh SVC untuk memetakan data ke ruang berdimensi lebih besar. Ini memungkinkan data yang tidak dapat dipisahkan secara linier di ruang awal untuk dipisahkan pada ruang fitur yang lebih besar [10].

Kelebihan utama dari SVC adalah kemampuannya dalam menangani data berdimensi tinggi, ketahanan terhadap *overfitting* terutama dengan jumlah sampel yang terbatas, serta efektivitasnya dalam memproses data yang memiliki *outlier* melalui konsep *soft margin* [11].

Margin adalah jarak antara *hyperplane* dengan titik data terdekat dari masing-masing kelas dalam algoritma SVC. Titik terdekat itu disebut *support vectors* yang memastikan bahwa dua kelas data dapat dipisahkan dengan jelas dan akurat. SVC berusaha memaksimalkan margin karena margin yang lebih besar meningkatkan kemampuan model dalam melakukan generalisasi terhadap data baru [12].

Hyperplane pemisah pada SCV dirumuskan sebagai berikut:

$$w \cdot x + b = 0 \quad (2.2)$$

Dimana:

w adalah vektor bobot.

x adalah vektor fitur.

b adalah bias.

Prediksi kelas:

$$\hat{y} = \begin{cases} 1, & \text{jika } f(x) \geq 0 \\ 0, & \text{jika } f(x) \leq 0 \end{cases} \quad (2.3)$$

Mekanisme ini menunjukkan bahwa proses klasifikasi bergantung pada posisi relatif suatu titik data terhadap *hyperplane* pemisah. Dalam pembentukannya, nilai $f(x)$ sangat dipengaruhi oleh sejumlah sampel kunci yang berada paling dekat dengan margin pemisahan, yaitu *support vectors*. Titik-titik inilah yang menentukan bentuk dan orientasi *hyperplane*.

2.3. Teori Evaluasi Model

Evaluasi model adalah proses sistematis untuk menilai seberapa baik sebuah model pembelajaran mesin dapat memprediksi atau mengklasifikasikan data dengan benar. Tujuan utama dari evaluasi model adalah untuk mengukur tiga hal: akurasi (seberapa tepat model dalam melakukan prediksi), reliabilitas atau stabilitas (seberapa konsisten performa ketika diuji ulang), dan kemampuan generalisasi (seberapa baik model bekerja pada data baru yang belum digunakan dalam pelatihan) [13].

Dalam klasifikasi, beberapa metrik umum digunakan: akurasi (*accuracy*) untuk mengukur proporsi prediksi yang benar, presisi (*precision*) untuk menilai ketepatan model dalam memprediksi kelas positif, sensitivitas (*recall*) untuk menilai sejauh mana model menangkap seluruh data positif, dan *F1-score*, yang merupakan keseimbangan antara presisi dan sensitivitas. Untuk memastikan model baik pada data pelatihan dan stabil saat diuji pada data baru, evaluasi model juga dapat melibatkan teknik seperti *cross-validation* [13].

2.4. Metriks Evaluasi

Akurasi adalah salah satu metrik evaluasi yang sering digunakan untuk menilai kinerja model pembelajaran mesin, yang diukur berdasarkan persentase jumlah prediksi yang benar dibandingkan dengan total seluruh prediksi yang dibuat oleh model. Karena mudah dihitung dan ditafsirkan serta memberikan gambaran umum tentang bagaimana model bekerja, metrik ini sangat populer. Akurasi sebagai satu-satunya metrik evaluasi biasanya dianggap tepat ketika data digunakan seimbang, yaitu jumlah data untuk setiap kelas relatif sama. Dalam situasi seperti ini, akurasi dapat secara adil menunjukkan kemampuan model tanpa bias terhadap salah satu kelas (Miller, Portlock, Nyaga, & O'Sullivan, 2024).

Tabel 1 : Tabel Konfusi Matriks

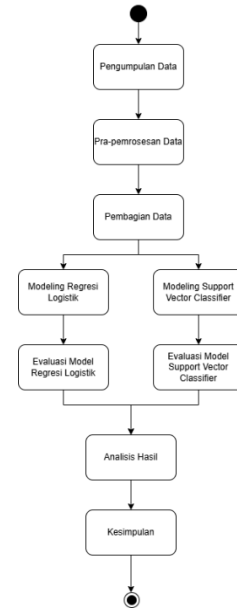
Nilai Prediksi	Nilai Aktual	
	<i>Positive</i>	<i>Negative</i>
<i>Positive</i>	TP (<i>True Positive</i>)	FP (<i>False Positive</i>)
<i>Negative</i>	FN (<i>False Negative</i>)	TN (<i>True Negative</i>)

Pada Tabel 1 *True Positive* (TP) adalah ketika model memprediksi positif, dan hasil aktual memang positif. Model memprediksi “bercerai” dan pasangan benar-benar bercerai. *False Positive* (FP) adalah ketika model memprediksi positif, tetapi hasil aktual negatif (*Type I Error*). Model memprediksi “bercerai” padahal pasangan masih menikah. *False Negative* (FN) adalah ketika model memprediksi negatif, tetapi hasil aktual positif (*Type II Error*). Model memprediksi “menikah” padahal pasangan bercerai. *True Negative* (TN) adalah ketika model memprediksi negatif, dan hasil aktual memang negatif. Model memprediksi “menikah” dan pasangan masih menikah. *Confusion matrix* tidak hanya memberikan gambaran umum tentang jumlah prediksi yang benar dan salah, tetapi juga menjadi dasar untuk menghitung metrik evaluasi lain seperti akurasi, presisi, sensitivitas, dan *F1-score* [13]. Misalnya, akurasi dapat dihitung dengan rumus:

$$Akurasi = \frac{TP+TN}{TP+TN+FP+FN} \quad (2.4)$$

3. METODOLOGI

Kerangka penelitian disusun untuk memberikan gambaran sistematis mengenai alur pelaksanaan penelitian, mulai dari identifikasi masalah hingga penarikan kesimpulan.



Gambar 1. Kerangka Penelitian

Alur yang digunakan untuk penelitian adalah seperti pada Gambar 1. Langkah-langkah yang dilakukan pada penelitian mencakup: (1) pengumpulan data yang didapatkan dari sebuah dataset pada Tahun 2024 yang disebut “Split or Stay: Divorce Predictor Dataset” yang terdiri dari 170 entri data. Data didapatkan dari Platform Kaggle. (2) Pra-pemrosesan Data dilakukan untuk mengubah data menjadi bentuk yang optimal untuk pemodelan dan analisis. Pra-pemrosesan meliputi penghapusan data tidak lengkap, inkonsisten, atau mengandung distorsi, normalisasi data, dan encoding data. (3) Pembagian data, Data dibagi menjadi 2 bagian, *dataset training* dan *dataset testing* yang dibagi dengan rasio 80:20. *Dataset training* digunakan untuk melatih model dan *dataset testing* digunakan untuk evaluasi model. (4) pelatihan model dengan algoritma SVC dan LR. (5) Evaluasi Hasil pelatihan model. Evaluasi hasil pelatihan model dilakukan dengan menggunakan metrik akurasi, presisi, sensitivitas, serta *F1-Score*. (6) Perbandingan hasil untuk kedua algoritma untuk mendapatkan kemampuan masing-masing algoritma. (7) Kesimpulan dari hasil penelitian.

4. HASIL DAN PEMBAHASAN

4.1. Hasil Penelitian

“*Dataset Prediction Divorce*” (DPS), yang diambil dari situs Kaggle, berisi 8250 data yang memiliki 54 atribut (fitur) yang mencakup pernyataan perilaku dan komunikasi pasangan suami istri, serta satu kolom kelas (*Class*) yang menunjukkan status hubungan. Kelas “0” menunjukkan bahwa pasangan telah bercerai, sedangkan kelas “1” menunjukkan bahwa pasangan masih menikah. Pengolahan data menggunakan bahasa *Python* melalui platform *Google Colab* menggunakan pustaka *NumPy*, *Pandas*, dan *scikit-learn* untuk komputasi, pemrosesan data, dan

implementasi model pembelajaran mesin dan data disimpan dalam format CSV. Dataset memiliki 55 kolom, yang masing-masing memiliki 54 atribut dan 1 kelas.

Tabel 2 : Deskripsi Data

	Atr1	...	Atr54	Class
count	170.000.000	...	170.000.000	170.000.000
mean	1.776.471	...	2.011.765	0.494118
std	1.627.257	...	1.667.611	0.501442
min	0.000000	...	0.000000	0.000000
25%	0.000000	...	0.000000	0.000000
50%	2.000.000	...	2.000.000	0.000000
75%	3.000.000	...	4.000.000	1.000.000
max	4.000.000	...	4.000.000	1.000.000

Ringkasan statistik untuk setiap fitur dihasilkan oleh fungsi Tabel 1 yang mencakup nilai rata-rata (*mean*), simpangan baku (*standard deviation*), nilai minimum pada 54 atribut, maksimum, dan kuartil (25%, 50%, dan 75%). Tujuan langkah ini adalah untuk mendapatkan pemahaman tentang distribusi data dan menemukan nilai ekstrem atau outlier.

Tabel 3 : Menunjukkan Data Tidak Ada yang Kosong.

	0
Atr1	0
Atr2	0
Atr3	0
Atr4	0

Hasil pemeriksaan menunjukkan bahwa tidak ada data dengan nilai duplikat. seluruh kolom terisi sepenuhnya, artinya tidak ada nilai yang hilang. Pada Tabel 2 diperlihatkan atribut 1-54 dan kelas dengan angka 0 yaitu tidak ada data yang kosong. Oleh karena itu, dataset dapat digunakan tanpa proses imputasi. Meskipun data duplikat tidak ditemukan, prosedur penghapusan duplikat tetap dilakukan sebagai langkah pencegahan untuk memastikan bahwa tidak ada baris data yang tercatat lebih dari satu kali. Keberadaan data duplikat dapat menyebabkan bias dalam proses pelatihan model dan menurunkan kualitas prediksi. Oleh karena itu, jumlah baris yang terduplikasi dihapus dengan menggunakan fungsi `df.drop_duplicates()`.

Karena model dalam ML memerlukan data numerik, kolom ini sudah sesuai atau tidak perlu diubah. Jumlah kemunculan setiap kategori pada kolom kelas dalam dataset ditampilkan dengan perintah `df['Class'].value_counts()`. Sebelum proses pelatihan model dimulai, pembaca dapat melihat proporsi masing-masing kelas secara lebih jelas dan memahami keseimbangan data melalui hasil perintah ini, yang dapat diketahui melalui distribusi data antara kelas "Menikah" dan "Bercera". Hasil perintah ini terkait dengan Tabel 3. *Output* menunjukkan distribusi kelas cukup seimbang, sehingga tidak diperlukan *resampling*.

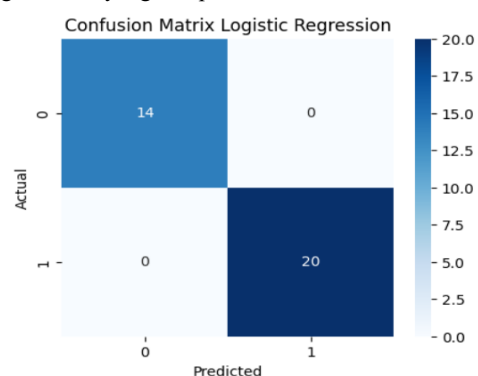
Tabel 4 : Jumlah Kelas

Class	count
0	86
1	84

Data dibagi menjadi dua bagian: fitur (X), yang mengandung variabel independen sebagai *input model*, dan label (Y), yang merupakan variabel dependen atau target yang ingin diprediksi, yaitu kolom kelas. Ini dilakukan untuk memudahkan proses pelatihan dan evaluasi, serta memungkinkan model untuk mempelajari secara terarah hubungan antara fitur dan label. Variabel X memiliki dimensi (170, 54) yang menunjukkan bahwa ada 170 data observasi dengan 54 variabel independen yang digunakan sebagai prediktor, dan variabel Y memiliki dimensi (170) yang menunjukkan bahwa variabel target memuat 170 nilai label, yang sama dengan jumlah data pada X. Variabel ini hanya memiliki satu kolom, yaitu kolom kelas, yang merupakan *output* dari model klasifikasi. Dataset telah konsisten dan siap untuk tahap pemodelan, seperti yang ditunjukkan oleh perubahan jumlah baris antara X dan Y. Data dibagi menjadi dua bagian menggunakan fungsi `Train_test_split` dari Pustaka `scikit-learn` dengan rasio 80:20, dimana 80 untuk *data training* dan 20 untuk *data testing*. Rasio 80:20 dipilih karena merupakan praktik umum yang memberikan keseimbangan optimal antara jumlah data untuk proses pembelajaran dan jumlah data untuk evaluasi model. Dengan 80% *data training*, model memperoleh informasi yang cukup untuk mempelajari pola dalam dataset, sementara 20% *data testing* dinilai sudah representatif untuk menilai performa model secara objektif.

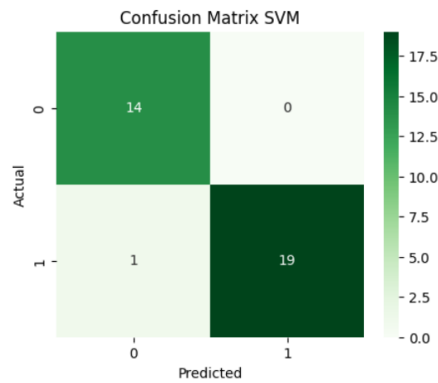
4.2. Hasil Evaluasi Model

Proses evaluasi dilakukan terhadap dua algoritma LR dan SVC, dengan proporsi *data testing* 20%. Hasil perbandingan metrik evaluasi kedua algoritma ditunjukkan bahwa LR mendapatkan akurasi, sensitivitas, presisi dan *F1-score* mendapatkan 100% sementara SVC mendapatkan 97% menghasilkan sedikit lebih rendah dari LR. Pendekatan ini menghitung metrik untuk setiap kelas secara terpisah, kemudian memberikan bobot sesuai proporsi jumlah sampel pada setiap kelas. Dengan demikian, nilai metrik menjadi lebih representatif terhadap distribusi data yang tidak seimbang. Penggunaan rata-rata berbobot ini penting agar performa model tidak bias pada kelas tertentu yang memiliki jumlah sampel lebih besar atau lebih kecil. Hasilnya menunjukkan bahwa model LR pada Gambar 2 ditemukan dengan akurat memprediksi seluruh data tanpa FP dan FN sehingga akurasi yang didapatkan 100%.



Gambar 2. Matriks untuk Logistic Regression

Berdasarkan Gambar 3, SVC menghasilkan sedikit kesalahan prediksi dengan 1 data FN, dimana model mengira pasangan tidak bercerai, tetapi sebenarnya bercerai. Namun, secara keseluruhan, model menunjukkan akurasi sebesar 97% dan hasil yang sangat baik.



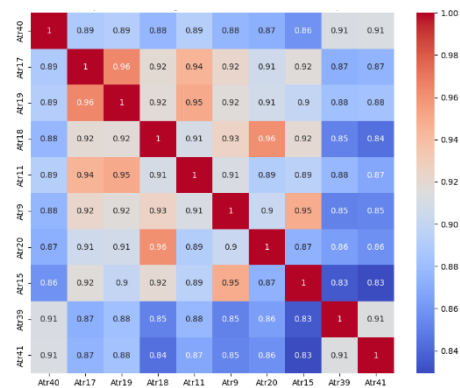
Gambar 3. Matriks untuk *Support Vector Classifier*

Pada model LR, beberapa *hyperparameter* ditetapkan untuk menjamin proses pelatihan berlangsung stabil dan menghasilkan model yang konsisten. Algoritma optimasi yang digunakan adalah *lbfgs*, yakni metode *quasi-Newton* yang dikenal efisien untuk kasus klasifikasi biner dan cocok digunakan pada dataset dengan jumlah fitur cukup banyak seperti DPS. Regularisasi yang dipilih adalah *L2 regularization*, yang berfungsi mengurangi risiko *overfitting* dengan menekan besarnya koefisien agar model tidak menjadi terlalu kompleks. Parameter *C* ditetapkan pada nilai 1.0, yang mencerminkan tingkat regularisasi moderat serta memberikan keseimbangan antara fleksibilitas model dan batasan terhadap kompleksitasnya. Batas iterasi ditentukan melalui *max_iter=100*, sehingga proses optimasi memiliki cukup kesempatan untuk mencapai titik konvergensi. Adapun *class_weight=None* dipertahankan karena distribusi kelas pada dataset relatif seimbang, sehingga tidak diperlukan penyesuaian bobot antar kelas. Dengan konfigurasi tersebut, model LR diharapkan mampu memberikan hasil prediksi yang stabil serta tidak rentan terhadap *overfitting*.

Pada model SVC, pengaturan *hyperparameter* dirancang agar algoritma mampu menangkap pola non-linier yang mungkin terdapat dalam data. Kernel yang dipilih adalah RBF, karena kernel ini memiliki fleksibilitas tinggi dan mampu memetakan data ke ruang berdimensi lebih tinggi sehingga pola pemisahan antar kelas dapat dipelajari lebih baik. Nilai regulasi *C=1.0* dipertahankan untuk memberikan keseimbangan antara margin yang lebih lebar dan toleransi terhadap kesalahan klasifikasi. Parameter *gamma=scale* digunakan karena pendekatan ini menghitung nilai *gamma* secara otomatis berdasarkan varians fitur, sehingga kompleksitas batas keputusan dapat menyesuaikan diri dengan karakteristik dataset. *Hyperparameter degree=3* tetap dicantumkan sebagai bagian dari konfigurasi standar meskipun tidak berdampak ketika kernel yang digunakan adalah RBF. Sementara itu, pengaturan *probability=False* digunakan demi efisiensi proses komputasi, mengingat perhitungan probabilitas memerlukan prosedur tambahan seperti *Platt Scaling* yang tidak diperlukan pada penelitian ini. Secara keseluruhan, konfigurasi ini dirancang agar SVC dapat beroperasi secara efektif sekaligus mempertahankan kemampuan generalisasi yang baik.

Gambar 4 menampilkan *heatmap* korelasi untuk 10 fitur yang memiliki hubungan terkuat terhadap variabel kelas. Berdasarkan visualisasi tersebut, seluruh fitur menunjukkan nilai korelasi yang sangat tinggi yaitu berada pada rentang 0.83 hingga 0.96. Hal ini mengindikasikan bahwa fitur-fitur tersebut saling memiliki hubungan linier yang kuat dan kemungkinan mengukur aspek

perilaku atau kondisi yang saling berkaitan. Beberapa pasangan fitur dengan korelasi tertinggi antara lain *Atr17* dan *Atr19* (0.96), *Atr18* dan *Atr20* (0.96), serta *Atr11* dan *Atr19* (0.95). Korelasi yang tinggi menunjukkan bahwa fitur-fitur tersebut memiliki pola yang mirip dan mungkin menggambarkan dimensi psikologis atau hubungan interpersonal yang sama. Secara keseluruhan, *heatmap* ini mendukung pemilihan fitur sebagai variabel penting dalam proses pemodelan. Tingginya korelasi antara fitur dan target memperkuat alasan penggunaan fitur-fitur tersebut dalam algoritma klasifikasi, karena berpotensi meningkatkan akurasi dan konsistensi prediksi model.



Gambar 4. *Heatmap* dengan 10 fitur korelasi terkuat terhadap kelas

Pengujian model juga dilakukan menggunakan metode *K-Fold Cross-Validation*, yang bertujuan untuk menilai stabilitas dan kemampuan generalisasi model secara lebih komprehensif. Pada penelitian ini digunakan skema *Cross-Validation* dengan jumlah fold sebanyak lima, sehingga dataset dibagi menjadi lima bagian yang masing-masing secara bergantian berfungsi sebagai *data testing*, sementara empat bagian lainnya digunakan sebagai *data training*. Dengan demikian, proses pelatihan dan pengujian dilakukan sebanyak lima kali, dan setiap iterasi memberikan nilai akurasi yang berbeda tergantung pada karakteristik subset data yang digunakan. Hasil pengujian menunjukkan bahwa model menghasilkan skor akurasi berturut-turut sebesar 0.8823, 1.0, 1.0, 1.0, dan 1.0. Dengan demikian, empat dari lima *fold* memberikan performa klasifikasi yang sempurna, sementara satu *fold* memperoleh akurasi sebesar 88.23%, yang masih tergolong sangat tinggi. Variasi kecil ini menunjukkan adanya distribusi data yang sedikit berbeda pada salah satu *fold*, tetapi tetap dalam rentang yang wajar. Secara keseluruhan, nilai akurasi yang tinggi pada hampir semua *fold* mengindikasikan bahwa model memiliki *robustness* yang sangat baik.

Rata-rata akurasi yang diperoleh dari proses *Cross-Validation* adalah sebesar 0.9764 atau 97.64%. Nilai rata-rata yang sangat tinggi ini mencerminkan bahwa model tidak hanya berkinerja baik pada *data training*, tetapi juga mampu mempertahankan performanya ketika diuji pada subset data yang benar-benar berbeda. Kondisi ini menunjukkan bahwa model memiliki kemampuan generalisasi yang sangat kuat serta potensi risiko *overfitting* yang rendah. Dengan demikian, hasil *Cross-Validation* memberikan bukti kuat bahwa model yang digunakan dalam penelitian ini stabil, andal, dan layak digunakan untuk tugas prediksi pada dataset perceraian yang digunakan.

Pada hasil kesimpulan yang didapatkan bahwa LR menunjukkan kinerja sempurna pada dataset perceraian linier dengan akurasi

dan *F1-score* sebesar 100%. Performa SVC dengan kernel linier juga sangat baik sebanyak 97%, tetapi sedikit lebih rendah karena sensitivitas terhadap margin antar kelas. Pola hubungan yang cukup kuat antara fitur dan label perceraian ditunjukkan oleh semua metrik evaluasi yang tinggi. Nilai akurasi yang sempurna pada LR kemungkinan disebabkan karena dataset tidak memiliki *noise* yang signifikan dan bersifat linier *separable*, sehingga model dapat mengklasifikasikan data dengan sangat baik.

5. KESIMPULAN DAN SARAN

5.1. Kesimpulan

Penelitian ini mengevaluasi efektivitas LR dan SVC dalam memprediksi potensi perceraian menggunakan DPS. Hasilnya menunjukkan bahwa LR adalah algoritma yang paling optimal, mencapai akurasi sempurna 100% pada *data testing* dan akurasi *Cross-Validation* yang stabil sebesar 97,6%. Meskipun SVC juga menunjukkan performa sangat baik dengan akurasi 97%, LR unggul dalam interpretabilitas yang tinggi dan efisiensi komputasi, menjadikannya pilihan yang lebih sesuai untuk penelitian sosial. Analisis koefisien LR mengungkapkan bahwa kualitas interaksi pasangan adalah faktor paling berpengaruh, pengetahuan mendalam tentang pasangan (Atr27) berfungsi sebagai pelindung, sedangkan perilaku kritik negatif (Atr54 dan Atr52) merupakan pemicu utama perceraian. Temuan ini menegaskan kemampuan pendekatan berbasis data dan model linier dalam menganalisis fenomena sosial kompleks seperti risiko perceraian.

5.2. Saran

Penelitian selanjutnya dianjurkan untuk melaksanakan analisis koefisien bobot yang lebih mendalam dan komprehensif terhadap seluruh 54 fitur yang diteliti. Langkah ini krusial untuk menyusun tatanan hierarki yang presisi mengenai faktor-faktor psiko-sosial yang paling signifikan dalam memengaruhi potensi perceraian. Meskipun LR telah memberikan kinerja superior, evaluasi lanjutan melalui implementasi model *ensemble* atau optimasi *hyperparameter* yang lebih intensif pada model SVC tetap wajib dilakukan. Upaya ini ditujukan untuk menilai batas kemampuan prediksi maksimal dari dataset ini, sekaligus memastikan seluruh potensi eksplorasi pendekatan ML telah dieksploitasi secara optimal.

DAFTAR PUSTAKA

- [1] A. Giddens and P. W. . Sutton, *Sociology*. Polity Press, 2021.
- [2] Y. Ansori and K. F. H. Holle, "Perbandingan Metode Machine Learning dalam Analisis Sentimen Twitter," *J. Sist. dan Teknol. Inf.*, vol. 10, no. 4, p. 429, Dec. 2022, doi: 10.26418/justin.v10i4.51784.
- [3] H. Muhiburrahman, "ANALISIS PERBANDINGAN METODE KLASIFIKASIOLOGISTIC REGRESSION DAN SUPPORT VECTOR MACHINE," 2023.
- [4] H. Nalatissifa, W. Gata, S. Diantika, and K. Nisa, "Perbandingan Kinerja Algoritma Klasifikasi Naive Bayes, Support Vector Machine (SVM), dan Random Forest untuk Prediksi Ketidakhadiran di Tempat Kerja," *J. Inform. Univ. Pamulang*, vol. 5, no. 4, p. 578, Dec. 2021, doi: 10.32493/informatika.v5i4.7575.
- [5] Badan Pusat Statistik, "Nikah dan Cerai Menurut Provinsi (kejadian), 2024," bps. Accessed: Dec. 01, 2025. [Online]. Available: <https://www.bps.go.id/id/statistics-table/3/VkhwVUuszTXJPVmq2ZFRKamNIZG9RMVo2VEdsbVVUMDkjMyMwMDAw/nikah-dan-cerai-menurut-provinsi--kejadian-.html?year=2024>
- [6] E. Alpaydin, *Introduction to machine learning*. The MIT Press, 2020.
- [7] J. Han, J. Pei, and H. Tong, *Data Mining*. Elsevier, 2023, doi: 10.1016/C2013-0-18660-6.
- [8] A. Zaidi and A. S. M. Al Luhayb, "Two Statistical Approaches to Justify the Use of the Logistic Function in Binary Logistic Regression," *Math. Probl. Eng.*, vol. 2023, no. 1, Jan. 2023, doi: 10.1155/2023/5525675.
- [9] J. Cao, M. Wang, Y. Li, and Q. Zhang, "Improved support vector machine classification algorithm based on adaptive feature weight updating in the Hadoop cluster environment," *PLoS One*, vol. 14, no. 4, p. e0215136, Apr. 2019, doi: 10.1371/journal.pone.0215136.
- [10] H. Kim, D. Zhang, L. Kim, and C.-H. Im, "Classification of Individual's discrete emotions reflected in facial microexpressions using electroencephalogram and facial electromyogram," *Expert Syst. Appl.*, vol. 188, p. 116101, Feb. 2022, doi: 10.1016/j.eswa.2021.116101.
- [11] R. Guido, S. Ferrisi, D. Lofaro, and D. Conforti, "An Overview on the Advancements of Support Vector Machine Models in Healthcare Applications: A Review," *Information*, vol. 15, no. 4, p. 235, Apr. 2024, doi: 10.3390/info15040235.
- [12] M. Kurniawan, M. Hakimah, and S. Agustini, "Perbandingan SVM dan Perceptron dengan Optimasi Heuristik," *J. Telemat.*, vol. 15, no. 2, pp. 85–92, Feb. 2021, doi: 10.61769/telematika.v15i2.356.
- [13] C. Miller, T. Portlock, D. M. Nyaga, and J. M. O'Sullivan, "A review of model evaluation metrics for machine learning in genetics and genomics," *Front. Bioinform.*, vol. 4, Sep. 2024, doi: 10.3389/fbinf.2024.1457619.