

PENGKLASIFIKASIAN INDEKS KEYAKINAN KONSUMEN DI INDONESIA TAHUN 2022 MENGGUNAKAN METODE *SUPPORT VECTOR MACHINE*

Aritana Udju Dima^{1*}, Yudi Setyawan², Rokhana Dwi Bekti³, Febriani Astuti⁴
^{1,2,3,4}Program Studi Statistika, Fakultas Sains dan Teknologi Informasi, Universitas AKPRIND Indonesia
 Email : arrytana@gmail.com
 *corresponding author

Abstract. *The Consumer Confidence Index (CCI) is one of the main indicators in assessing economic performance. The CCI provides an indication of future developments in household consumption and savings, based on answers regarding their expected financial situation, their sentiments about the general economic situation, unemployment and ability to save. The February 2023 Consumer Survey indicated consumer optimism regarding economic conditions, this was indicated by the February 2023 Consumer Confidence Index of 122.4, which was maintained at an optimistic level (CCI > 100). The data used is data from 18 cities/provinces in Indonesia sourced from Bank Indonesia (BI) accessed in February 2023. In this study, two kernels from the Support vector machine method were used, namely the RBF kernel and the Polynomial kernel to carry out the classification. on the consumer confidence index in 2022, by dividing the data into three proportions, namely 60:40, 75:25, and 80:20 and randomizing three times for training data, namely set.seed (159, 57 and 143) and using cost values of 0.1, 1, 10 and 100. With Gamma 1 and a parameter degree value of 2 and randomization of the accuracy again for the testing data as much as thirty times, the accuracy interval value for the RBF is 88.01% to 88.92%, so it can be concluded that the accuracy the RBF kernel is slightly better than the polynomial kernel which has a value of around 87.36%.*

Keywords: *Consumer Confidence Index, RBF Kernels, Polinomial Kernel, Cost Values, Gamma*

Abstrak. Indeks keyakinan konsumen (IKK) adalah salah satu indikator utama dalam menilai kinerja perekonomian. IKK memberikan indikasi perkembangan konsumsi dan tabungan rumah tangga di masa mendatang, berdasarkan jawaban mengenai situasi keuangan yang mereka harapkan, sentimen mereka tentang situasi ekonomi secara umum, pengangguran dan kemampuan menabung. Survei Konsumen Februari 2023 mengindikasikan optimisme konsumen terhadap kondisi ekonomi, hal tersebut terindikasi dari Indeks Keyakinan Konsumen Februari 2023 sebesar 122.4, tetap terjaga pada level optimis (IKK > 100). Data yang digunakan adalah data 18 kota/provinsi yang berada di Indonesia yang bersumber dari Bank Indonesia (BI) yang diakses pada bulan Februari tahun 2023. Pada penelitian kali ini menggunakan dua kernel dari metode *Support vector machine* yaitu kernel RBF dan kernel Polinomial untuk melakukan pengklasifikasian terhadap indeks keyakinan konsumen pada tahun 2022, dengan membagi data ke dalam tiga proporsi yaitu 60:40, 75:25, dan 80:20 dan dilakukan tiga kali pengacakan untuk data training yaitu set.seed (159, 57 dan 143) serta dengan menggunakan nilai cost sebesar 0.1, 1, 10 dan 100. Dengan Gamma 1 serta nilai derajat parameter sebesar 2 dan dilakukan pengacakan akurasi kembali untuk data testing sebanyak tiga puluh kali diperoleh nilai interval akurasi untuk RBF yaitu 88.01% sampai 88.92%, Jadi dapat disimpulkan bahwa akurasi kernel RBF sedikit lebih bagus dibandingkan kernel polinomial yang memiliki nilai sekitar 87.36%.

Kata kunci: Indeks Keyakinan Konsumen, Kernel RBF, Kernel Polinomial, Cost, Gamma

1. Pendahuluan

Indeks keyakinan konsumen (IKK) adalah salah satu indikator utama dalam menilai kinerja perekonomian. IKK memberikan indikasi perkembangan konsumsi dan tabungan rumah tangga di masa mendatang, berdasarkan jawaban mengenai situasi keuangan yang mereka harapkan, sentimen mereka tentang situasi ekonomi secara umum, pengangguran dan kemampuan menabung. Jika nilai indikator tersebut di atas 100 berarti terdapat peningkatan kepercayaan konsumen terhadap situasi ekonomi masa depan, sebagai akibatnya mereka cenderung tidak menabung dan lebih cenderung membelanjakan uang untuk pembelian besar

dalam 12 bulan ke depan. Nilai di bawah 100 menunjukkan sikap pesimis terhadap perkembangan ekonomi di masa depan, yang kemungkinan mengakibatkan kecenderungan untuk lebih banyak menabung dan mengkonsumsi lebih sedikit.

Di Indonesia terdapat dua lembaga yang secara rutin merilis IKK, yaitu Bank Indonesia (BI) dan Danareksa Research Institute (DRI). Dengan adanya IKK yang diperoleh dari hasil survei kepada masyarakat maka dapat diketahui persepsi masyarakat terhadap ekonomi saat ini dan di masa yang akan datang. IKK yang dirilis setiap bulan oleh BI adalah rata-rata sederhana dari Indeks Kondisi Ekonomi (IKE) saat ini dan Indeks Ekspektasi Konsumen (IKK) terhadap kondisi perekonomian 6 bulan mendatang. BI melakukan survei konsumen di 18 kota sebagai sampel sehingga dapat diperoleh informasi perbedaan dan perubahan IKK secara spasial.

Pada penelitian ingin dilakukan untuk mengklasifikasikan Indeks Keyakinan Kosumen di Indonesia dengan metode *Support Vector Machine* dikarenakan peneliti ingin melihat apakah masyarakat optimis terhadap pertumbuhan ekonomi kedepannya. Pada penelitian ini juga akan dibandingkan kernel yang terdapat dalam SVM yaitu kernel *Radial Basis Function* dan kernel Polinomial dengan membandingkan nilai akurasi tertinggi. Dari penelitian ini juga diharapkan dapat memberikan gambaran bagi pemerintah di masing-masing kota untuk mengambil kebijakan terhadap kepercayaan masyarakat.

2. Metode Penelitian

a. Sumber data

Data yang digunakan dalam penelitian adalah data sekunder yang diperoleh dari Bank Indonesia dan bisa di akses dengan website https://www.bi.go.id/id/publikasi/ruang-media/news-release/Pages/sp_2512023.aspx

b. Variabel Penelitian

Adapun variabel yang digunakan dalam penelitian terdiri dari variabel dependen dan variabel independen.

1. Variabel dependen (Y) dalam penelitian ini yaitu Data Indeks Keyakinan Konsumen yang dibagi dua kategori yaitu nilai diatas 100 merupakan optimis sedangkan nilai dibawa 100 tidak optimis
2. Variabel independen (X) yang digunakan sebanyak tiga variabel yaitu Indeks Harga Konsumen, Tabungan Masyarakat dan Jumlah Bank

c. Tahapan Analisis Data

1) Analisis Deskriptif

Statistik deskriptif dilakukan pada masing-masing variabel.

2) *Support Vector Machine*

Adapun langkah-langkah pada analisis menggunakan SVM adalah sebagai berikut:

a. Pengumpulan data

b. Membentuk data latih dan data uji dengan proporsi 60:40; 75:25 dan 80:20

c. Pada proses SVM, akan dicari fungsi hyperplane dengan menentukan fungsi kernel yang digunakan, setelah itu dihitung dengan matriks $m \times m$. Selanjutnya ditentukan nilai parameter C untuk mendapatkan nilai alpha atau support vector pada data training. Nilai parameter C yang digunakan adalah 0.1, 1, 10, dan 100. Setelah didapatkan nilai alpha, selanjutnya akan dicari nilai bias (b) menggunakan dan untuk membentuk persamaan *Hyperlane* pada kedua kernel digunakan

d. Menghitung akurasi klasifikasi kernel RBF berdasarkan *confusion matrix*

e. Menghitung akurasi klasifikasi kernel Polinomial berdasarkan *confusion matrix*

f. Menentukan tingkat akurasi klasifikasi terbaik antara kernel RBF dan kernel polinomial dengan membandingkan hasil akurasi klasifikasi menggunakan data testing

d. Metode Analisis Data

Support Vector Machine merupakan salah satu metode klasifikasi data mining. SVM adalah algoritma yang bekerja menggunakan pemetaan nonlinear untuk mengubah data pelatihan asli ke dimensi yang lebih tinggi. Dalam hal ini dimensi baru, akan mencari hyperplane untuk memisahkan secara linier dan dengan pemetaan nonlinier yang tepat ke dimensi lebih tinggi, data dari dua kelas selalu dapat dipisahkan dengan hyperplane tersebut (Ritonga, dkk 2018).

Hyperplane terbaik adalah hyperplane yang terletak diantara dua set objek dari dua kelas yang diperoleh dengan mengukur margin hyperplane dan mencari titik maksimalnya. Margin merupakan jarak antara hyperplane dengan titik terdekat dari masing-masing kelas (Athoillah, 2017).

Pada umumnya masalah dalam domain dunia nyata (*real world problem*) data yang diperoleh jarang yang bersifat linear separable. Kebanyakan bersifat *non-linear*. Dalam menyelesaikan *problem non-linear*, SVM dimodifikasi dengan memasukkan fungsi kernel (Nugroho. dkk, 2003).

Persamaan *Hyperplane* yang membatasi dua kelas pada SVM yaitu

$$H_1: x_i w + b \geq 1 \text{ untuk } y_1 = +1 \quad (1)$$

$$H_2: x_i w + b \leq 1 \text{ untuk } y_2 = -1 \quad (2)$$

Penggabungan dari persamaan (1) dan (2) menghasilkan pertidaksamaan :

$$y_i(x_i w + b) \geq 1, \text{ untuk } i = 1, 2, \dots, n \quad (3)$$

n merupakan jumlah data *Training*

Margin antara dua kelas dapat dihitung dengan mencari jarak antara kedua *hyperplane* H_1 atau H_2 . Setiap tupel pelatihan yang jatuh pada *hyperplane* H_1 atau H_2 yang memenuhi persamaan (1) disebut *support vector*. Jarak terdekat suatu titik di bidang H_1 terhadap pusat dapat dihitung dengan meminimalkan $x^T x$ dengan memperhatikan kendala $x_i w + b \geq 1$. Dengan menggunakan bentuk fungsi *Lagrange* dan turunan pertamanya diperoleh:

$$\left| \frac{(1-b)}{\|w\|} - \frac{(-b-1)}{\|w\|} \right| = \frac{2}{\|w\|} \quad (4)$$

Oleh karena memaksimalkan $\frac{1}{\|w\|}$ sama dengan meminimumkan $\|w\|^2$ dan untuk menyederhanakan penyelesaian ditambahkan faktor $\frac{1}{2}$ dengan demikian, model persamaannya menjadi:

$$\text{Min } \frac{1}{2} \|w\|^2 \quad (5)$$

$y_i(x_i w + b) \geq 1$, untuk $i = 1, 2, n$ (n merupakan jumlah data training).

Dalam klasifikasi kadang-kadang didapatkan bidang pemisah yang diambil tidak berbentuk linier sehingga diperlukan penyelesaian khusus untuk permasalahan ini. Untuk data-data yang tidak dapat dipisahkan secara linier tersebut ditambahkan variabel *slack* $\xi_i \geq 0$ ke pertidaksamaan (5) sehingga kendala dan fungsi tujuan menjadi

$$y_i(x_i w + b) - 1 + \xi_i \geq 0, \text{ untuk } i = 1, 2, n \quad (6)$$

Dengan n merupakan data training

$$\text{Min } \frac{1}{2} \|w\|^2 + C \sum_{i=1}^n \xi_i \tag{7}$$

Dengan $y_i(x_i w + b) - 1 + \xi_i \geq 0, \xi_i \geq 0$, untuk i

Hal ini dilakukan untuk mengurangi jumlah kesalahan margin antara dua kelas. Hal ini dilakukan untuk mengurangi jumlah kesalahan klasifikasi yang dinyatakan dengan variabel *slack* ξ_i . Penggunaan variabel ini adalah untuk mengatasi kasus ketidaklayakan (*infeasibility*) dan pembatas (*constraints*) dengan cara memberi pinalti untuk data yang tidak memenuhi pembatas tersebut. Pinalti ini dinotasikan dengan C dan perlu waktu lebih panjang. Untuk itu diperkenalkan pengali *Lagrange* α_i dengan $i = 1, 2, \dots, n$. Sehingga diperoleh persamaan sebagai berikut: (Rachman, 2012)

$$L_D = \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j K(x_i, x_j^T) \tag{8}$$

Untuk masalah pengklasifikasian kasus *non-linier* dapat menggunakan fungsi kernel atau disebut sebagai kernel trick dimana kita hanya cukup mengetahui fungsi kernel yang dipakai, dan tidak perlu mengetahui wujud dari fungsi *non-linear* ϕ (Nugroho, dkk. 2003). Selanjutnya hasil klasifikasi dari data x diperoleh dari persamaan berikut:

$$f(x) = \text{sign}(\sum_{i=1}^n \alpha_i y_i K(x_i^T, x_j) + b) \tag{9}$$

Pada penelitian ini menggunakan dua kernel yaitu kernel RBF dan kernel Polinomial. Berikut merupakan fungsi dari kedua kernel yang digunakan :

1. Kernel RBF

$$K(\vec{x}_i, \vec{x}_j) = \exp\left(-\gamma \frac{\|\vec{x}_i - \vec{x}_j\|}{2\sigma^2}\right) \tag{10}$$

2. Kernel Polinomial

$$K(\vec{x}_i, \vec{x}_j) = (\vec{x}_i \cdot \vec{x}_j + 1)^d \tag{11}$$

3. Hasil dan Pembahasan

- a. Analisis Deskriptif

Berikut merupakan tabel karakteristik setiap variabel yang digunakan

Tabel 1 Analisis Deskriptif

Keterangan	Rata-rata	Nilai Min	Nilai Maks	Standar Deviasi
IKK	120.8	78.5	152.5	15.69173636
IHK	104.63	102.19	107.42	1.113955495
TM	266.284	13.633	2.481.875	509424.6036
JB	159	30	471	138.1253481

- b. Pengklasifikasian Metode SVM

- 1) Kernel *Radial Basis Function*

Pada perhitungan menggunakan kernel *Radial Basis Function* (RBF) peneliti melakukan beberapa percobaan dalam penentuan nilai cost dan Gamma sehingga diperoleh beberapa nilai *cost* dan Gamma, untuk nilai *cost* yaitu 0.1, 1, 10, 100 dan

Gamma yaitu 1. Kemudian untuk proporsi digunakan data *training* dan data *testing* yang digunakan sebesar 60:40, 75:25 dan 80:20.

setelah dilakukan beberapa pengacakan dan percobaan nilai parameter *cost* pada data training diperoleh nilai parameter *cost* sebesar 100 dan Gamma sebesar 1 dengan data training 60% karena memiliki jumlah *support vector* yang paling rendah yaitu sebesar 31 pada pengacakan pertama (set. seed 159).

Selanjutnya dilakukan pengoptimalan fungsi *hyperplane* dengan menggunakan *lagrange multiplier* untuk setiap titik data. Penentuan alpha membutuhkan bantuan *software* karena jumlah model yang sangat banyak. Nilai $\alpha > 0$ merupakan titik data sebagai *support vector*, sedangkan nilai $\alpha < 0$ merupakan titik data yang bukan *support vector*. Sehingga dapat dikatakan bahwa tidak semua titik data mempunyai nilai $\alpha > 0$.

Berikut merupakan nilai alpha yang diperoleh dari proporsi terbaik

Tabel 2 Nilai Alpha RBF C =100

No	Nilai Alpha (α)	Data Ke..
1	67.618	14
2	3.374	17
3	100	19
...
31	2.711	121

Dari nilai alpha di atas dapat membantu untuk menemukan nilai bias, sehingga diperoleh nilai bias yaitu -1.660. Setelah diperoleh nilai alpha (α) dan bias, maka dapat dilakukan prediksi terhadap data testing sebagai berikut:

Tabel 3 Akurasi data Testing dari proporsi terbaik

	Nilai Cost	Akurasi
Data	0.1	87.36%
Testing	1	88.51%
40%	10	88.51%
	100	87.36%

Dengan perhitungan akurasi data testing dapat dilihat bahwa pada tabel 3 masih terdapat nilai akurasi lain yang memiliki nilai lebih besar sehingga perlu dilakukan estimasi interval lagi sehingga mendapat nilai akurasi data testing yang terbaik untuk model di atas.

Setelah dilakukan estimasi interval menggunakan model terbaik dengan menggunakan proporsi yang sama pada model terbaik yaitu 60:40, Cost = 100, Gamma = 1 dan dilakukan tiga puluh kali pengacakan maka diperoleh hasil seperti pada tabel berikut ini

Tabel 4 Pengacakan data Testing RBF

No	Pengacakan	RBF C= 100
		Akurasi data testing
1	Set.seed (24)	88.51%
2	Set.seed (100)	88.51%
...	...	''
30	Set.seed (27)	87.36%
	rata-rata	88.47%
	Standar Deviasi (s)	1.66284%

Setelah dilakukan pengacakan sebanyak tiga puluh kali untuk mencari estimasi interval diperoleh nilai rata-rata pada data testing sebesar 88.47% yang artinya memiliki kategori baik dan didapat nilai standar deviasi sebesar 1.66284% sehingga interval confidence 95% untuk akurasi adalah

$$\bar{X} - Z_{\alpha/2} \frac{s}{\sqrt{n}} \leq \mu \leq \bar{X} + Z_{\alpha/2} \frac{s}{\sqrt{n}}$$

Diperoleh

$$0.8787 \leq \mu \leq 0.8906 \text{ atau } 87.87\% \leq \mu \leq 89.06\%$$

Dari hasil perhitungan di atas diperoleh didapat interval konfiden yaitu antara 87.87% sampai 89.06%

2) Kernel Polinomial

Pada klasifikasi SVM dengan fungsi kernel polinomial terdapat parameter γ (Gamma), d (derajat parameter) dan C (cost). Penentuan parameter γ (Gamma) untuk fungsi pemisah dengan fungsi kernel polinomial dicobakan pada parameter 1, sedangkan penentuan derajat parameter yang digunakan adalah $d = 2$ dan untuk nilai C yang dicobakan yaitu 0.1, 1, 10, 100. Nilai C , γ dan d tersebut kemudian diterapkan pada data training dengan melakukan tiga kali pengacakan dan menggunakan tiga proporsi data *training testing* yang dicobakan.

Setelah dilakukan perhitungan menggunakan bantuan *software R-studio* untuk mencari jumlah *support vector* dengan tiga kali pengacakan menggunakan tiga proporsi data *training testing* yang berbeda didapat hasil nilai parameter *cost* sebesar 0.1, Gamma sebesar 1 dan menggunakan derajat parameter $d = 2$ pada data *training* 60% memiliki nilai yang bagus karena memiliki jumlah *support vector* yang paling rendah yaitu sebesar 36 pada pengacakan pertama (set.seed 159) dengan nilai alpha atau *lagrange multiplier* dapat dilihat pada tabel berikut

Tabel 5 Nilai Alpha Kernel Polinomial

No	Nilai Alpha	Data ke
1	0.1	2
2	0.1	3
3	0.1	13
...
36	0.0271	129

Dari nilai alpha pada tabel 5 dapat membantu untuk menemukan nilai bias, sehingga diperoleh nilai bias yaitu -1.000362. Setelah diperoleh nilai alpha (α) dan bias, maka dapat dilakukan prediksi terhadap data testing sebagai berikut:

Tabel 6 Proporsi terbaik kernel Polinomial

	Nilai Cost	Akurasi
	0.1	87.36%
Data Testing 40%	1	87.36%
	10	87.60%
	100	87.36%

Pada tabel 6 terdapat nilai akurasi lain yang berbeda sehingga perlu dilakukan estimasi interval lagi sehingga mendapat nilai akurasi data *testing* yang terbaik untuk model di atas.

Setelah dilakukan estimasi interval menggunakan model terbaik dengan menggunakan proporsi yang sama pada model terbaik yaitu 60:40, $Cost = 100$, $\Gamma = 1$, derajat parameter = 2 dan dilakukan tiga puluh kali pengacakan maka diperoleh hasil seperti pada Tabel berikut ini

Poly C = 0.1		
No	Pengacakan	Akurasi data testing
1	Set.seed (24)	87.36%
2	Set.seed (100)	87.36%
...

30	Set.seed (27)	87.36%
	rata-rata	87.36%

Setelah dilakukan pengacakan sebanyak tiga puluh kali diperoleh nilai akurasi yang sama pada data *testing* yaitu sebesar 87.36% yang artinya memiliki kategori baik

4. Kesimpulan

Berdasarkan penelitian yang telah dilakukan, maka kesimpulan yang didapatkan adalah sebagai berikut:

1. Hasil statistik deskriptif Indeks Keyakinan Konsumen yang didapat bahwa Indeks Keyakinan Konsumen pada tahun 2022 cenderung konstan pada angka 100 keatas yang artinya pada tahun 2022 kepercayaan masyarakat terhadap kondisi ekonomi rumah tangga memiliki respon yang optimis, dengan nilai rata-rata sebesar 120.8, nilai minimal sebesar 78.5, nilai maksimal sebesar 152.2 dan standar deviasinya sebesar 15.69173636.
2. Hasil klasifikasi Indeks Keyakinan Konsumen menggunakan metode SVM dengan fungsi Kernel RBF menghasilkan tingkat akurasi pada data training sebesar 93.02% dan data *testing* sebesar 88.47% dengan jumlah *Support vector* sebanyak 31, nilai *error* sebanyak 0.06971 dan standar deviasi sebesar 0.069761 yang diperoleh dari pemodelan dengan nilai *Cost* = 100 dan *Gamma* = 1 pada pengacakan pertama (set. Seed 159) dengan interval konfidens untuk akurasi data testing adalah 87.87% sampai 89.06%
3. Hasil klasifikasi Indeks Keyakinan Konsumen menggunakan metode SVM dengan fungsi Kernel Polynominal menghasilkan tingkat akurasi pada data training sebesar 87.60% dan data testing sebesar 87.36% dengan jumlah *Sopport vector* sebanyak 36, nilai *error* sebanyak 0.124031 dan standar deviasi sebesar 0 yang diperoleh dari pemodelan dengan nilai *cost* = 0.1, *d* = 2 dan *Gamma* = 1 pada pengacakan pertama (set. seed 159).
4. Berdasarkan hasil perbandingan akurasi dari kernel RBF dan Polinomial didapat hasil bahwa kernel RBF sedikit lebih baik dibandingkan kernel Polinomial karena memiliki nilai akurasi yang lebih tinggi dibandingkan model kernel polinomial dengan nilai akurasi data *training* sebesar 93.02% dan interval konfidens untuk akurasi adalah 87.87% sampai 89.06%.

Ucapan Terima Kasih

Penulis mengucapkan terima kasih kepada dosen pembimbing Bapak Drs. Yudi Setyawan, M.S.,M.Sc dan dosen pembimbing II Ibu Maria Titah Jatipaningrum, S.Si., M.Sc yang telah memberikan banyak waktu, tenaga, dan pikiran mulai dari awal sampai akhir laporan penelitian ini. Penulis juga mengucapkan terima kasih kepada kedua orang tua, saudara, serta teman-teman penulis yang selalu memberi dukungan, dorongan, serta doa dari awal sampai akhir penyusunan penelitian ini.

Daftar Pustaka

- [1]. Athoillah, (2017). *Pengenalan Wajah Menggunakan SVM Multi kernal dengan Pembelajaran yang Bertambah*. Surabaya: jurnal online Informatika.
- [2]. Noeryanti. dkk (2018). *Metode Statistika II*. Yogyakarta: AKPRIND PRESS.
- [3]. Nugroho, (2018, September). Analisis Sentimen Pada Media Sosial Twitter Menggunakan Naive Bayes Classifier Dengan Ekstrasi Fitur N-Gram. *Jurnal Sains Komputer & Informatika (J-SAKTI)*, 2, 200-209.
- [4]. Ritoga,dkk. (2018). Penerapan Metode Support Vector Machine (SVM) dalam Klasifikasi Kualits Pengelasan SMAW (Shield Metal ARC Welding). *Jurnal Ilmiah Edutic*, 5.