

Perbandingan Kernel Support Vector Machine Dalam Melakukan Klasifikasi Penundaan Biaya Kuliah Mahasiswa

Wiwi Widayani¹⁾, Harliana^{*2)}

¹⁾ Program Studi Sistem Informasi, Fakultas Ilmu Komputer, Universitas AMIKOM Yogyakarta
Jl Ring Road Utara, Condongcatur Sleman, Yogyakarta
Email wiwi.w@amikom.ac.id

²⁾ Program Studi Teknik Informatika, Fakultas Teknik, Universitas Muhadi Setiabudi
Jl Pangeran Diponegoro KM.2 Pesantunan Wanasari, Brebes
Email harliana.hifzhiya@gmail.com

Abstrak

Rekomendasi penundaan pembayaran kuliah merupakan salah satu bentuk kebijakan yang diambil oleh suatu Perguruan Tinggi Swasta terhadap mahasiswanya. Ketika seorang mahasiswa mengajukan permohonan penundaan pembayaran maka secara tidak langsung bagian keuangan harus dapat mengklasifikasi mahasiswa yang akan membayar tepat waktu dan yang gagal bayar. Berdasarkan hal tersebut, maka penelitian ini bertujuan untuk mendapatkan nilai akurasi tertinggi melalui algoritma SVM dalam memberikan rekomendasi penundaan pembayaran kuliah secara tepat bagi mahasiswa. Untuk mendapatkan akurasi tertinggi, pengujian dilakukan melalui 2 cara dengan 6 scenario pengujian, berdasarkan hasil pengujian pertama (membandingkan antara jumlah data training dan testing) diketahui bahwa nilai akurasi akan berbanding lurus dengan banyaknya jumlah data training yang digunakan. Selain itu pemilihan jenis kernel yang digunakan juga akan mempengaruhi besarnya nilai akurasi yang dihasilkan, dan hal ini sesuai dengan hasil pengujian kedua yang dilakukan.

Kata kunci: penundaan pembayaran kuliah, SVM, akurasi.

Abstract

The recommendation to postpone tuition payments is a University policy towards its students. When a student applies for a payment delay, the finance department indirectly must be able to classify the student who will pay on time and failed. Based on this, this study aims to obtain the highest accuracy value through the SVM algorithm in providing recommendations for proper postponement of tuition payments for students. To get the highest accuracy, testing is carried out in 2 ways with 6 testing scenarios, based on the results of the first test it is known that the accuracy value will be directly proportional to the large amount of training data used. In addition, the choice of kernel type used will also affect the amount of accuracy value produced, and this is in accordance with the results of the second test conducted.

Keywords: *postponement of tuition fees, SVM, accuracy*

1. PENDAHULUAN

Rekomendasi penundaan pembayaran kuliah merupakan salah satu bentuk kebijakan yang diambil oleh suatu Perguruan Tinggi Swasta terhadap mahasiswanya. Di Universitas Muhadi Setiabudi rekomendasi penundaan pembayaran kuliah akan diberikan kepada mahasiswa yang telah membayar minimal 25% dari total tagihan semester yang berjalan. Beberapa syarat prosedural yang digunakan untuk mendapatkan dispensasi pembayaran kuliah diantaranya pekerjaan dan penghasilan kedua orangtua, kepemilikan tempat tinggal, kepemilikan kendaraan, jumlah tanggungan keluarga, jumlah cicilan yang dimiliki keluarga, tagihan air dan listrik, dan

pajak kepemilikan kendaraan bermotor[1][2]. Ketika seorang mahasiswa mengajukan permohonan penundaan pembayaran kepada pihak kampus, maka semua syarat tersebut harus dilengkapi dengan surat pernyataan yang telah ditandatangani oleh orang tua / wali mengenai batas waktu kesanggupan sisa pembayaran yang akan dilakukan.

Saat bagian keuangan menyetujui permohonan rekomendasi penundaan pembayaran yang diajukan, maka secara tidak langsung bagian keuanganpun harus dapat memprediksi prosentase jumlah mahasiswa yang akan membayar tepat waktu dan yang gagal bayar. Algoritma fuzzy Mamdani pernah digunakan untuk menentukan besarnya rekomendasi penundaan pembayaran yang akan dilakukan oleh mahasiswa, berdasarkan *fuzzy rule* yang terbentuk didapatkan bahwa nilai dispensasi penundaan pembayaran berbanding lurus dengan besarnya potongan yang akan didapatkan mahasiswa[3]. Karena *fuzzy Mamdani* berprinsip pada nilai min-max maka akurasi yang dihasilkanpun hanya mampu berkisar pada 60-70%[4]. Berbeda dengan dua penelitian tersebut, pada penelitian ini peneliti akan mencoba menaikkan akurasi yang dihasilkan dalam memberikan rekomendasi penundaan pembayaran kuliah melalui metode klasifikasi SVM (*Support Vector Machine*). Hasil akhir klasifikasi ini selanjutnya dapat digunakan untuk memprediksi jumlah mahasiswa yang akan mendapatkan rekomendasi penundaan pembayaran serta memprediksi mahasiswa yang akan menepati pembayarannya ataupun yang akan gagal bayar. Pada penelitian ini algoritma SVM dipilih karena kemampuannya yang spesifik dalam mengontrol keputusannya melalui kernel yang digunakan[5]. Selain itu SVM juga memiliki nilai akurasi paling baik dalam melakukan klasifikasi namun dengan jumlah data yang tidak terlalu besar[6]. Dalam melakukan klasifikasi, ternyata kinerja SVM pernah dibandingkan dengan RF (*Random Forest*) dan SGD (*Stochastics Gradient Descent*) dalam melakukan klasifikasi kinerja *programmer*, hasilnya menunjukkan bahwa rata-rata nilai akurasi yang didapatkan oleh SVM (melalui perhitungan *cross validation*) menghasilkan nilai yang jauh lebih tinggi bila dibandingkan dengan RF dan SGD yaitu 81,3%; 74,4%; dan 80,1% [7]. Apabila dilihat dari waktu eksekusi, algoritma SVM memiliki waktu eksekusi diantara 0,1 detik sampai dengan kurang dari 0,3 detik dalam melakukan klasifikasi kualitas pengajuan kredit[8]. Berdasarkan hal tersebut maka penelitian ini bertujuan untuk mengukur nilai akurasi tertinggi melalui pendekatan 3 kernel (linier, polynominal degree, dan gaussian RBF) SVM. Selanjutnya kernel yang memiliki akurasi tertinggilah yang akan menghasilkan klasifikasi pemberian rekomendasi penundaan pembayaran kuliah secara tepat bagi mahasiswa.

2. TINJAUAN PUSTAKA

2.1 Klasifikasi

Klasifikasi merupakan suatu model/fungsi yang akan membedakan *object* menjadi kelas tertentu, dengan tujuan menentukan kelas dari *object* tersebut berdasarkan proses *learning* dan *testing* yang dilakukan [9]. Proses *learning / training* akan digunakan untuk mendapatkan model berdasarkan data training[9][10]. Sedangkan pada tahapan *testing*, sistem akan melakukan pengetesan data terhadap model yang telah didapatkan sebelumnya [9].

2.2 Support Vector Machine (SVM)

SVM merupakan salah satu algoritma klasifikasi yang dikembangkan oleh Boser, Guyon dan Vapnik yang digunakan untuk memisahkan data berdasarkan 2 kelas berbeda melalui garis optimal *hyperplane*[11]. *Hyperplane* inilah yang selanjutnya akan digunakan sebagai pemisah antara 2 kelas berdasarkan jarak margin [11]. Dimana margin sendiri merupakan jarak terdekat antara pola terdekat dengan *hyperplane* untuk setiap kelasnya, dan pola terdekat inilah yang selanjutnya akan disebut dengan *support vector* [12]. Persamaan yang digunakan oleh SVM ditunjukkan oleh persamaan 1 dan 2[8]

$$[(w^T \cdot x_i) + b] \geq 1 \text{ untuk } y_i = +1 \quad (1)$$

$$[(w^T \cdot x_i) + b] \leq -1 \text{ untuk } y_i = -1 \quad (2)$$

Ketika data tersebut bersifat linier, maka persamaan (1) dan (2) dapat dioptimalkan dengan fungsi *Lagrange Multiplier*, sedangkan jika data tersebut bersifat non-linier maka dapat menggunakan kernel [8]. Adapun tahapan dalam algoritma SVM tersebut adalah [13]:

1. Tentukan titik data: $\{x_1, x_2, \dots, x_n\}$ yang merupakan atribut yang digunakan
2. Tentukan kelas data: $y_i \{-1, +1\}$, dimana -1 sebagai mahasiswa yang tidak diizinkan mendapat rekomendasi penundaan pembayaran, dan 1 sebagai mahasiswa yang mendapat izin untuk mendapat rekomendasi penundaan pembayaran
3. Tentukan pasangan data dan kelas berdasarkan persamaan (3)

$$\{(x_i, y_i)\}_{i=1}^N \quad (3)$$

Dimana : N adalah banyaknya data

4. Maksimalkan fungsi berdasarkan persamaan (4)

$$Ld = \sum_{i=1}^N a_i - \sum_{i=1}^N \sum_{j=1}^N a_i a_j y_i y_j K(x_i x_j) \quad (4)$$

Dengan syarat:

$$0 \leq a_i \leq C \text{ dan } \sum_{i=1}^N a_i y_i = 0$$

Dimana:

Ld : dualitas langrange multipler

a_i : nilai bobot setiap titik data

C : nilai konstanta

5. Hitung nilai w dan b melalui persamaan (5) dan (6)

$$w = \sum_{i=1}^n a_i y_i x_i \quad (5)$$

$$b = \frac{-1}{2} (w \cdot x^i + w \cdot x) \quad (6)$$

6. Hitung fungsi keputusan klasifikasi $\text{sign}(f(x))$ berdasarkan persamaan (7)

$$f(x) = w \cdot x + b \quad (7)$$

Atau bisa juga menggunakan persamaan (8)

$$F(x) = \sum_{i=1}^m a_i y_i K(x, x_i) + b \quad (8)$$

Dimana:

m adalah jumlah titik data yang memiliki $a_i > 0$

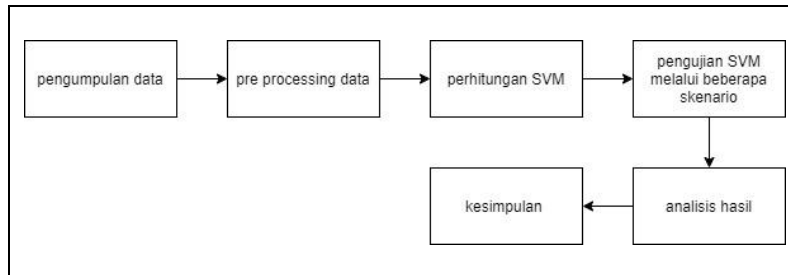
$k(x, x_i)$ adalah fungsi kernel yang digunakan

3. METODE PENELITIAN

Penelitian ini akan menggunakan 1.089 dataset yang berasal dari data mahasiswa Universitas Muhadi Setiabudi pada semester genap 2019. Adapun parameter yang digunakan dalam proses klasifikasi ini adalah keaktifan kuliah mahasiswa, nilai IPK mahasiswa, Pendidikan terakhir mahasiswa, pendapatan / penghasilan rata-rata orangtua / wali mahasiswa, jumlah anggota keluarga yang masih sekolah, total tunggakan mahasiswa, rata-rata hutang yang telah diselesaikan oleh mahasiswa, dan prosentase ketepatan pembayaran cicilan. Sebelum dataset digunakan akan dilakukan *pre-processing* terhadap data yang meliputi data *selection*, *cleaning*, dan *transformation*. Data *selection* digunakan untuk memilih parameter-parameter yang berpengaruh langsung terhadap pemberian rekomendasi penundaan pembayaran. Data *cleaning* digunakan untuk memilih dataset yang tidak bersifat null dan *missing value*, sedangkan *transformation* digunakan untuk merubah atribut menjadi bentuk yang akan memudahkan klasifikasi. Dari tahapan *pre-processing* ini peneliti mendapatkan 945 dataset yang siap digunakan untuk proses klasifikasi dengan SVM. Untuk mendapatkan hasil perbandingan terbaik antara data testing dan data training, maka dilakukan enam *scenario* perbandingan untuk data testing dan data training. Enam *scenario* ini selanjutnya akan dihitung akurasi dengan *confusion matrix* untuk

mendapatkan nilai akurasi. *Scenario* dengan nilai akurasi terbesar tersebut selanjutnya akan digunakan untuk mendapatkan kernel terbaik dalam proses klasifikasi.

Adapun rangkuman mengenai tahapan yang akan dilakukan pada penelitian ini terangkum pada Gambar 1.



Gambar 1. Alur penelitian

4. PEMBAHASAN

Untuk mendapatkan hasil yang maksimal, maka dalam penelitian ini penulis melakukan pengujian dengan 2 cara, yaitu:

- Pengujian melalui perbandingan jumlah data training dan testing, bertujuan untuk mencari nilai akurasi terbesar
- Pengujian jenis kernel, bertujuan untuk mengetahui rata-rata akurasi terbaik dari kernel yang akan digunakan

4.1 Pengujian akurasi secara *confussion matriks*

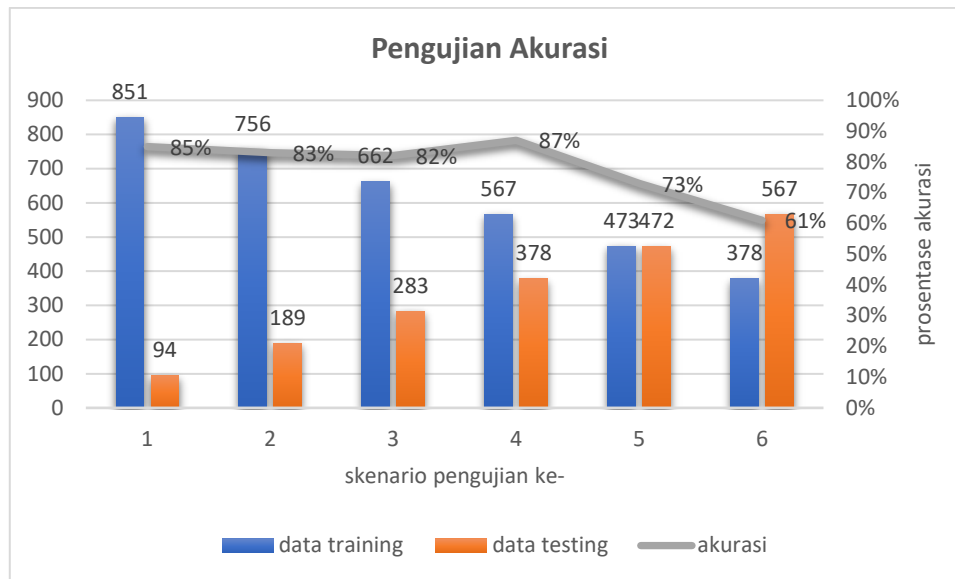
Pada pengujian pertama, peneliti menggunakan 6 rencana pengujian yang terangkum pada Tabel 1. Pengujian dihitung berdasarkan *confussion matriks* dimana parameter yang digunakan untuk masing-masing *sequential training* memiliki nilai yang tetap yaitu:

- λ : 0,5
- γ : 0,01
- c : 1
- iterasiMax : 200
- ε : 0,0001

Tabel 1. Rencana pengujian

No	Perbandingan	
	Jumlah data training	Jumlah data testing
Scenario ke-1	851	94
Scenario ke-2	756	189
Scenario ke-3	662	283
Scenario ke-4	567	378
Scenario ke-5	473	472
Scenario ke-6	378	567

Berdasarkan Gambar 2 terlihat bahwa nilai akurasi tertinggi terdapat pada *scenario* pengujian ke-4 yaitu 87% dengan perbandingan data training dan data testing adalah 60:40.



Gambar 2. Hasil pengujian akurasi secara *confussion matrix*

4.2 Pengujian jenis kernel

Pada pengujian kedua ini, penulis akan membandingkan rata-rata akurasi yang dihasilkan oleh kernel *Gaussian RBF*, *polynomial degree*, dan *linier*. Scenario yang digunakan untuk mendapatkan kernel terbaik adalah *scenario* ke-4 yaitu dengan perbandingan 60:40 untuk data testing dan data trainingnya.

a. Linier

Linier kernel SVM merupakan fungsi kernel yang baik digunakan ketika data sudah terpisah secara linier. Tabel 2 adalah tabel *confusion matrix* untuk data testing, sedangkan table 3 adalah table *confusion matrix* untuk data training.

Tabel 2. *Confusion matrix* data testing

Prediksi	Actual	
	Sesuai	Tidak sesuai
Sesuai	498	69
Tidak sesuai	0	0

Tabel 3. *Confusion matrix* data training

Prediksi	Actual	
	Sesuai	Tidak sesuai
Sesuai	370	8
Tidak sesuai	0	0

Berdasarkan tabel 2 dan 3, Langkah selanjutnya yaitu mencari nilai akurasi dari data testing dan data training melalui perhitungan *confusion matrix*. Hasil perhitungan ini selanjutnya terangkum pada Tabel 4.

Tabel 4. *Confusion matrix* data training

Dataset	Akurasi
Training dataset	0,4476
Testing dataset	0,4531

Ketika dilakukan analisis SVM dengan menggunakan kernel linier, didapatkan akurasi yang dihasilkan sebesar 0,4531 atau 45%.

b. Polynomial degree

Polinomial merupakan fungsi kernel non linier yang sangat cocok digunakan untuk permasalahan yang semua training dataset-nya dinormalisasi. Tabel 5 adalah tabel *confusion matrix* untuk data testing, sedangkan tabel 6 adalah tabel *confusion matrix* untuk data training.

Tabel 5. Confusion matrix data testing kernel polynominal

Prediksi	Actual	
	Sesuai	Tidak sesuai
Sesuai	403	164
Tidak sesuai	0	0

Tabel 6. Confusion matrix data testing kernel polynominal

Prediksi	Actual	
	Sesuai	Tidak sesuai
Sesuai	119	259
Tidak sesuai	0	0

Berdasarkan Tabel 5 dan 6, Langkah selanjutnya yaitu mencari nilai akurasi dari data testing dan data training melalui perhitungan *confusion matrix*. Hasil perhitungan ini selanjutnya terangkum pada Tabel 7 berikut.

Tabel 7. Confusion matrix data training kernel polynominal

Dataset	Akurasi
Training dataset	0,66102
Testing dataset	0,65891

Ketika dilakukan analisis SVM dengan menggunakan kernel polynominal, didapatkan akurasi yang dihasilkan sebesar 0,65891 atau 65%.

c. Gaussian RBF

Kernel RBF merupakan fungsi kernel yang digunakan ketika data tidak dapat terpisah secara linier, dimana dalam melakukan analisis dengan RBF akan dilakukan optimasi parameter cost dan gamma. Tabel 8 adalah tabel *confusion matrix* untuk data testing, sedangkan table 9 adalah tabel *confusion matrix* untuk data training.

Tabel 8. Confusion matrix data testing kernel RBF

Prediksi	Actual	
	Sesuai	Tidak sesuai
Sesuai	499	68
Tidak sesuai	21	546

Tabel 9. Confusion matrix data training kernel RBF

Prediksi	Actual	
	Sesuai	Tidak sesuai
Sesuai	335	43
Tidak sesuai	366	12

Berdasarkan Tabel 8 dan 9, Langkah selanjutnya yaitu mencari nilai akurasi dari data testing dan data training melalui perhitungan *confusion matrix*. Hasil perhitungan ini selanjutnya terangkum pada Tabel 10 berikut.

Tabel 10. Confusion matrix data training kernel polynominal

Dataset	Akurasi
Training dataset	0,96666
Testing dataset	0,97012

Ketika dilakukan analisis SVM dengan menggunakan kernel RBF, didapatkan akurasi yang dihasilkan sebesar 0,96666 atau 97%.

Berdasarkan pengujian yang telah dilakukan diatas, dapat diketahui bahwa kernel Gaussian RBF memiliki rata-rata akurasi tertinggi bila dibandingkan jenis kernel Polynomial dan linier, yaitu 97%. Hal ini menjelaskan bahwa kernel RBF ternyata mampu melakukan persebaran data yang lebih baik pada saat proses pemetaan datanya.

5. KESIMPULAN

Berdasarkan dua pengujian yang telah dilakukan melalui enam skenario pengujian, didapatkan bahwa ternyata akurasi akan dipengaruhi oleh perbandingan antara jumlah data training dan data testing serta jenis kernel yang digunakan. Ketika jumlah data training yang digunakan lebih banyak daripada data testing maka akurasi yang dihasilkanpun akan lebih dari 80%, hal ini terbukti dari hasil pengujian pada scenario 1 sampai dengan scenario 4 pada Gambar 2. Namun pada penelitian ini akurasi terbaik terlihat pada scenario keempat yaitu 87% dengan jumlah perbandingan 60:40. Sedangkan untuk jenis kernel, maka kernel Gaussian RBF lah yang memiliki tingkat akurasi tertinggi, yaitu 97% bila dibandingkan dengan kernel linier dan polynominal

DAFTAR PUSTAKA

- [1] F. D. Noviandha, I. F. Astuti, and A. H. Kridalaksana, "Sistem Pendukung Keputusan Untuk Penentuan Kategori Uang Kuliah Tunggal Dengan Metode Multifactor Evaluation Process (Studi Kasus : Universitas Mulawarman)," *Inform. Mulawarman J. Ilm. Ilmu Komput.*, vol. 13, no. 2, pp. 88–96, 2019, doi: 10.30872/jim.v13i2.1552.
- [2] H. Novianti, M. Mulya, and I. Jambak, "Sistem Pendukung Keputusan Penentuan Uang Kuliah Tunggal Dengan Metode Analitical Hierarchy Process (Ahp)," *JSI J. Sist. Inf.*, vol. 11, no. 1, pp. 1711–1721, 2019.
- [3] P. Poningsih and S. R. Andani, "Sistem Pendukung Keputusan Untuk Dispensasi Pembayaran Uang Kuliah," *Jurasik (Jurnal Ris. Sist. Inf. dan Tek. Inform.)*, vol. 3, no. 3, pp. 56–65, 2018, doi: 10.30645/jurasik.v3i0.66.
- [4] A. C. Mutia, A. F. Sundoro, A. Yajiddin, M. Khoirullah, and Q. Aini, "Review Penerapan Fuzzy Logic Sugeno Dan Mamdani Pada Sistem Pendukung Keputusan Prakiraan Cuaca Di Indonesia," *Semin. Nas. Sist. Inf. Indones.*, pp. 115–120, 2017.
- [5] W. Huang, Y. Nakamori, and S. Y. Wang, "Forecasting stock market movement direction with support vector machine," *Comput. Oper. Res.*, vol. 32, no. 10, pp. 2513–2522, 2005, doi: 10.1016/j.cor.2004.03.016.
- [6] H. Yu, X. Huang, X. Hu, and H. Cai, "A comparative study on data mining algorithms for individual credit risk evaluation," *Proc. - 2010 Int. Conf. Manag. e-Commerce e-Government, ICMecG 2010*, pp. 35–38, 2010, doi: 10.1109/ICMeCG.2010.16.
- [7] R. Umar, I. Riadi, and Purwono., "Perbandingan Metode SVM, RF dan SGD untuk Penentuan Model Klasifikasi Kinerja Programmer pada Aktivitas Media Sosial," *J. Resti*, vol. 4, no. 2, pp. 329–335, 2020.
- [8] I. T. A. Nur, N. Y. Setiawan, and F. A. Bachtiar, "Perbandingan Performa Metode Klasifikasi SVM, Neural Network, dan Naive Bayes untuk Mendeteksi Kualitas Pengajuan Kredit di Koperasi Simpan Pinjam," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 6, no. 4, pp. 444–449, 2019, doi: 10.25126/jtiik.2019641352.
- [9] A. I. Kurnia, M. T. Furqon, and B. Rahayudi, "Klasifikasi Kualitas Susu Sapi

-
- Menggunakan Algoritme Support Vector Machine (SVM) (Studi Kasus : Perbandingan Fungsi Kernel Linier dan RBF Gaussian),” *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 2, no. 11, pp. 4453–4461, 2018.
- [10] A. Purnamasari and S. Assegaff, “Penentuan Klasifikasi Tingkat Kesejahteraan Keluarga Menggunakan Metode Naive Bayes Pada Kecamatan Pasar Jambi,” *J. Manaj. Sist. Inf.*, vol. 4, no. 4, pp. 480–491, 2019.
- [11] V. A. Gunawan, I. I. Fitriani, and L. S. A. Putra, “Sistem Diagnosis Otomatis Identifikasi Penyakit Jantung Coroner Menggunakan Ekstraksi Ciri GLCM dan Klasifikasi SVM,” *Inform. Mulawarman J. Ilm. Ilmu Komput.*, vol. 15, no. 1, pp. 13–20, 2020, doi: 10.30872/jim.v15i1.2495.
- [12] A. A. Kasim and M. Sudarsono, “Algoritma Support Vector Machine (SVM) untuk Klasifikasi Ekonomi Penduduk Penerima Bantuan Pemerintah di Kecamatan Simpang Raya Sulawesi Tengah,” *Semin. Nas. APTIKOM 2019*, pp. 568–573, 2019.
- [13] N. Nurajijah and D. Riana, “Algoritma Naive Bayes, Decision Tree, dan SVM untuk Klasifikasi Persetujuan Pembiayaan Nasabah Koperasi Syariah,” *J. Teknol. dan Sist. Komput.*, vol. 7, no. 2, pp. 77–82, 2019, doi: 10.14710/jtsiskom.7.2.2019.77-82.

Biodata Penulis

Wiwi Widayani, M.Kom, lahir di Majalengka pada tahun 1983. Penulis pertama memperoleh gelar S.Kom di Jurusan Sistem Informasi STMIK Amikom Yogyakarta pada Tahun 2006. Kemudian melanjutkan pendidikan S2 pada STMIK Amikom Yogyakarta (sekarang telah menjadi Universitas) dan lulus pada Tahun 2015. Konsentrasi penelitian yang penulis pertama dalam yaitu bidang Artificial Intellience dan Rekayasa Perangkat Lunak. Saat ini penulis adalah salah satu dosen di Jurusan Sistem Informasi pada Universitas Amikom Yogyakarta.

Harliana, M.Cs, lahir di Cirebon pada tahun 1986. Penulis kedua memperoleh gelar S.T di Jurusan Teknik Informatika Universitas Muhammadiyah Cirebon pada Tahun 2008. Kemudian melanjutkan pendidikan S2 pada Universitas Gadjah Mada Yogyakarta dan lulus pada Tahun 2012. Konsentrasi penelitian yang penulis kedua dalam yaitu bidang *Artificial Intellience*, data mining, dan *image processing*. Saat ini penulis adalah salah satu dosen di Jurusan Teknik Informatika pada Universitas Muhadi Setiabudi Brebes