

Optimasi Metode Klasifikasi dengan Menggunakan Particle Swarm Optimization untuk Identifikasi Penyakit Diabetes Retinopathy

Toni Arifin*, Asti Herliana

Fakultas Teknik
Universitas BSI Bandung

Bandung

*toni.tfn@bsi.ac.id

Abstrak-Penyakit *diabetic retinopathy*, merupakan salah satu jenis penyakit yang bermula dari *diabetes mellitus*. Hingga dewasa ini, penyakit *diabetic retinopathy* masih sulit untuk dikenali pada tahap awal. Hal ini dimungkinkan karena sifat dari penyakit ini yang menyerang bagian dalam sistem saraf mata terlebih dulu. Hingga saat ini, para ahli masih mengandalkan pemeriksaan dengan menggunakan serangkaian tes pada mata serta memeriksa rekam medis dari pasien. Metode yang digunakan hingga saat ini, selain tentunya memerlukan proses yang panjang juga menghabiskan biaya yang tidak murah. Hal ini dapat menyulitkan penderita dengan kondisi ekonomi lemah. Penelitian yang dilakukan kali ini, bermaksud untuk mendapatkan metode klasifikasi terbaik yang dapat menghasilkan tingkat akurasi yang tinggi bila dikombinasikan dengan metode *Particle Swarm Optimization* (PSO). Setelah didapatkan metode klasifikasi terbaik, maka dikemudian hari penelitian ini dapat dikembangkan menjadi sebuah perangkat lunak pendeteksi *diabetic retinopathy*. Dari penelitian yang dilakukan dapat disimpulkan bahwa metode *Neural Network* (NN) merupakan metode terbaik dalam menghasilkan akurasi yang tinggi bila dikombinasikan dengan metode pemilihan fitur PSO.

Kata kunci: *Diabetic retinopathy*, Klasifikasi, Optimasi, *Particle Swarm Optimization*

1. Pendahuluan

Diabetes Mellitus (DM) merupakan salah satu penyakit dengan tingkat penderita yang cukup tinggi. Pada rentang antara Tahun 2007 sampai Tahun 2013, jumlah penderita diabetes di Indonesia mengalami peningkatan, dari 1.1% menjadi 2.1% [1]. Dengan tingginya penderita penyakit DM, maka akan lebih besar kemungkinan seseorang juga menderita penyakit komplikasi dari penyakit DM. Salah satu jenis penyakit yang merupakan penyakit komplikasi yang masih sulit disembuhkan dari DM adalah *Diabetic retinopathy* (DR). DR hingga saat ini masih sulit disembuhkan karena mayoritas penderita melakukan pemeriksaan di saat kondisi penyakit telah memasuki tahap berbahaya. Hal ini dikarenakan sifat dari penyakit DR ini yang tidak menunjukkan gejala yang terlihat bila masih pada tahap awal [2].

Dengan adanya fakta bahwa penyakit DR ini akan lebih mudah disembuhkan bila masih berada pada tahap awal, maka dibutuhkan suatu deteksi dini bagi para penderita DR agar tidak mencapai tahap lanjut. Berdasarkan penelitian yang telah dilakukan oleh Antal dan Hajdu pada tahun 2014 dihasilkan sebuah basis data yang dapat dijadikan acuan bagi para peneliti dalam mengklasifikasikan kondisi penderita DR yang dikenal dengan nama mesidor dataset [3]. Dengan adanya dataset ini, maka beberapa penelitian terdahulu telah dilakukan diantaranya penelitian mengenai

komparasi metode untuk memaksimalkan hasil klasifikasi telah dilakukan oleh para peneliti. Seperti yang dilakukan oleh Valverde, Garcia, Hornero, dan Galvez pada tahun 2016 dengan memanfaatkan data berupa data citra dari mesidor dataset. Pada penelitian ini, metode yang dikomparasikan terdiri dari *K-NN*, *Fuzzy C-Means*, *Neural Network*, *Support vector machine* dan *Multylayer Perceptron*. Pada penelitian ini, metode yang dikomparasikan terdiri dari *Fuzzy C-Means*, *K-NN*, *Multylayer Perceptron*, *Support vector machine* dan *Neural Network* [4].

Berdasarkan pada penelitian yang dilakukan oleh Valverde, Garcia, Hornero dan Galvez, maka pada penelitian kali ini juga akan dilakukan komparasi metode untuk mendapatkan metode terbaik yang di kemudian akan dapat dijadikan acuan dalam pembuatan perangkat lunak deteksi penyakit DR. Untuk dapat memaksimalkan hasil klasifikasi, maka dilakukan juga optimasi pemilihan fitur dengan menggunakan metode *Particle Swarm Optimization* (PSO). Metode optimasi pemilihan fitur dilakukan dengan mengacu kepada penelitian yang dilakukan oleh Herliana pada tahun 2016 [5] di mana pada penelitian ini dibuktikan bahwa dengan menggunakan metode optimasi pada pemilihan fitur, dapat memaksimalkan hasil klasifikasi pada dataset yang berjumlah besar. Selain itu, penggunaan metode optimasi untuk memaksimalkan hasil klasifikasi pada penelitian ini juga didasarkan pada penelitian yang dilakukan oleh Arifin pada tahun 2017 [6] di mana pada

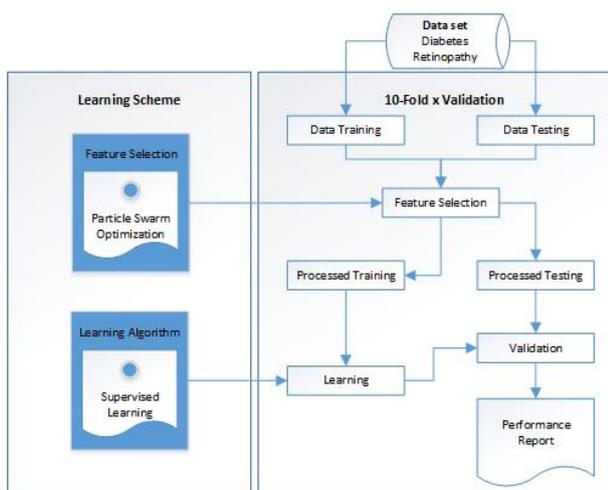
penelitian ini dibuktikan bahwa metode *Particle Swarm Optimization* (PSO) dapat digunakan untuk meningkatkan hasil klasifikasi pada dataset yang berjumlah besar dan berbentuk citra.

Tabel 1. Atribut pada dataset diabetes retinopathy

Name	Information
Attribute 0	The binary result of quality assessment
Attribute 1	The binary result of pre-screening
Attribute 2-7	The results of MA detection
Attribute 8-15	Exudates are represented by a set of points rather than the number of pixels constructing the lesions
Attribute 16	The euclidean distance of the center of the macula and the center of the optic disc to provide important information regarding the patients condition
Attribute 17	The diameter of the optic disc
Attribute 18	The binary result of the AM/FM-based classification
Attribute 19	Class label. 1 = contains signs of DR (Accumulative label for the Messidor classes 1, 2, 3), 0 = no signs of DR
Class	1 = contains signs of DR (Accumulative label for the Messidor classes 1, 2, 3), 0 = no signs of DR.

Tabel 2. Jumlah dataset berdasarkan class

No	Jenis Klasifikasi	Jumlah Record Dataset
1	Contains signs	611
2	No signs	540
Jumlah		1151



Gambar 1. Kerangka Penelitian

Berdasarkan beberapa penelitian terdahulu, maka pada penelitian kali ini menggunakan metode optimasi *Particle Swarm Optimization* yang dikombinasikan dengan beberapa

metode klasifikasi untuk mengetahui seberapa baikkah model yang diusulkan dalam proses klasifikasi, dan dapat memudahkan kinerja ahli dalam indentifikasi penyakit diabetes retinopathy.

2. METODE

Untuk memudahkan para ahli dalam mendeteksi dini penyakit Diabetes Rethinopathy, para peneliti telah berhasil membuat sebuah dataset hasil ekstraksi data citra retina mata yang terdiri dari 2 kelas yaitu kelas normal dan kelas yang terindikasi Diabetes Rethinopathy [7]. Dataset yang digunakan pada penelitian ini sebanyak 1151 data. Di bawah ini adalah tabel 1 dan 2 yang menjelaskan dataset yang digunakan.

a. Kerangka Penelitian

Berdasarkan gambaran dari kerangka penelitian di atas pada gambar 1, dapat dijelaskan melalui langkah-langkah berikut:

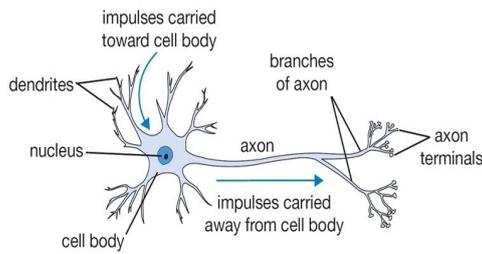
- Langkah pertama adalah memilih dataset yang akan digunakan dalam proses penelitian, dalam penelitian ini menggunakan dataset *Diabetic Rethinopathy*.
- Langkah selanjutnya pembagian dataset *Diabetic Rethinopathy* menjadi 10 bagian dengan menggunakan *10 fold cross validation*, semua bagian dataset akan dibagi menjadi data *training* dan data *testing*.
- Setelah pembagian data menjadi 2 bagian *training* dan *testing* dilakukan langkah seleksi fitur atau atribut dengan menggunakan *Particle Swarm Optimization*
- Langkah selanjutnya setelah didapatkan atribut terbaik adalah tahapan klasifikasi data dengan metode klasifikasi data mining.
- Sementara itu data *testing* yang telah terseleksi kemudian dilakukan validasi dengan data *training*.
- Setelah data *training* dan *testing* divalidasi dengan menggunakan metode klasifikasi. Langkah terakhir yaitu membandingkan hasil klasifikasi untuk mengetahui metode terbaik untuk klasifikasi dataset *Diabetic Rethinopathy*.

b. Deep Learning

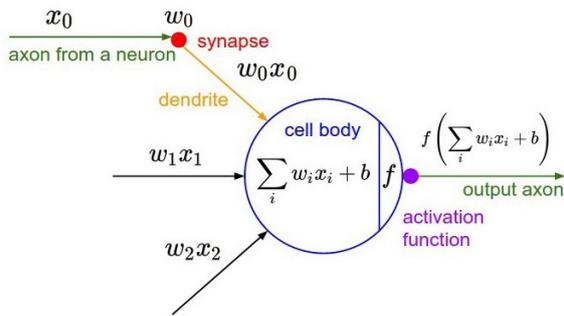
Deep learning mengacu pada algoritma *machine learning* dan andal dalam mengenali pola nonlinear yang kompleks pada data yang besar/*big data* [8], [9]. *Deep Learning* adalah sub-bidang dari *machine learning* yang dibuat oleh Hinton dan terinspirasi dari arsitektur otak manusia [9]. Dalam penelitian ini model metode klasifikasi dengan menggunakan *deep learning* akan diterapkan

c. Neural Network

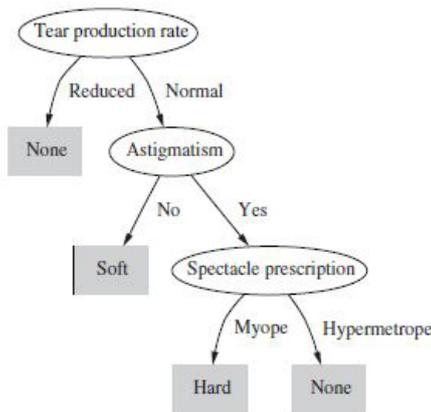
Neural Network adalah termasuk salah satu cabang dari bidang *soft computing*. Kemampuan dasar otak manusia menjadi acuan untuk diadopsi dalam *Neural Network*. Kemampuan tersebut adalah kemampuan untuk menerima stimulus, kemudian mengolah stimulus tersebut, dan memberikan tanggapan atau *output* sebagai hasil dari pengolahan stimulus. Variasi tanggapan atau *Output* merupakan fungsi dari dari variasi stimulasi dan kemampuan pengolahan yang dilakukan oleh otak[10]. Gambar 2 menyajikan ilustrasi dan formula dari algoritma *Neural Network* [11].



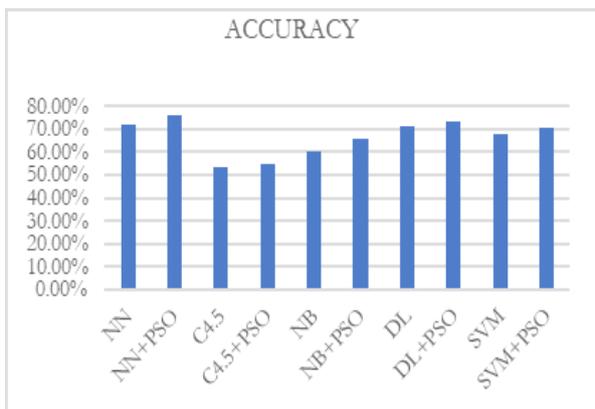
Gambar 2. Ilustrasi neural network



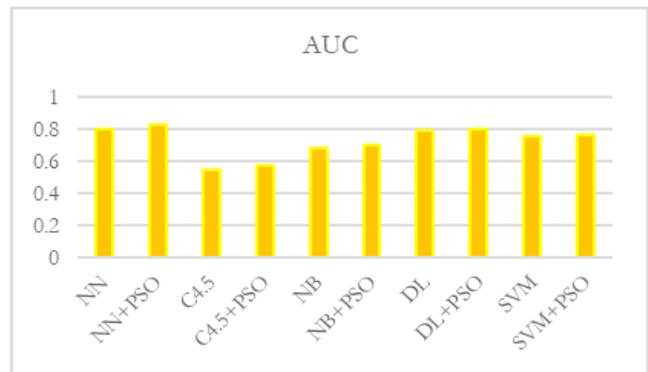
Gambar 3. Formula neural network



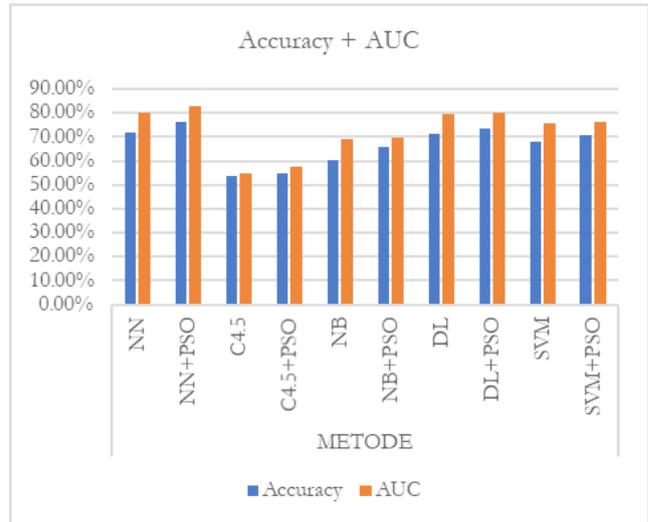
Gambar 4. Decision tree data lensa kontak



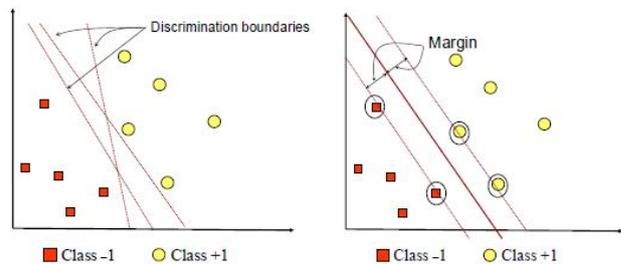
Gambar 5. Akurasi dari setiap metode



Gambar 6. AUC dari setiap metode



Gambar 7. Akurasi dan AUC dari setiap metode



Gambar 8. Proses SVM yang berusaha menemukan hyperplane terbaik untuk memisahkan keuda class -1 dan +1

d. C4.5 (Decision tree)

Decision tree merupakan model data mining yang bertujuan untuk mengekstraksi pengetahuan tersembunyi dari database yang besar [10]. Model decision tree ini kebanyakan diterapkan dalam analisis keputusan yang berperan dalam mengidentifikasi atau menentukan strategi yang paling sesuai untuk tujuan tertentu [12]. Gambar 4 merupakan contoh dari decision tree yang menunjukkan deskripsi struktural untuk data lensa kontak [10].

Tabel 3. Detail hasil eksperimen metode klasifikasi + *Particle Swarm Optimization* (PSO)

Keterangan	METODE									
	NN	NN+PSO	C4.5	C4.5+PSO	NB	NB+PSO	DL	DL+PSO	SVM	SVM+PSO
Accuracy	71.76%	76.11%	53.69%	55.08%	60.30%	66.03%	71.50%	73.50%	68.20%	70.90%
AUC	0.803	0.827	0.549	0.574	0.689	0.699	0.795	0.803	0.757	0.765

e. *Naïve Bayes*

Naïve bayes merupakan metode klasifikasi yang dikenal sederhana namun sangat efisien, model probabilistik dari proses klasifikasi naïve bayes didasarkan pada teorema bayes [13]. Metode ini merupakan salah satu metode yang memberikan manfaat perhitungan probabilitas dan statistik berdasarkan pengalaman di masa sebelumnya [14]. Kelebihan dari metode naïve bayes ini dalam proses klasifikasi adalah dalam jumlah data *training* yang diperlukan untuk menentukan estimasi parameter. Jumlah tersebut cukup kecil dibanding metode yang lain. Naïve bayes selalu bekerja jauh lebih baik dibanding metode klasifikasi lain [11][5].

$$P(H|X) = \frac{P(X|H) \cdot P(H)}{P(X)} \quad (1)$$

Keterangan:

- X : Data dengan *class* yang belum diketahui
- H : Hipotesis data merupakan suatu *class* spesifik
- $P(H|X)$: Probabilitas hipotesis H berdasar kondisi X (posteriori probabilitas)
- $P(H)$: Probabilitas hipotesis H (prior probabilitas)
- $P(X|H)$: Probabilitas X berdasarkan kondisi pada hipotesis H
- $P(X)$: Probabilitas X

f. *Support vector machine*

Support vector machine (SVM) merupakan algoritma *supervised learning* yang sangat populer [15]. *Support vector machine* lebih efektif dalam hal klasifikasi dan menghasilkan akurasi yang baik [16][6]. *Support vector machine* menunjukkan kinerja klasifikasi yang baik, dan secara komputasi sangat efisien untuk proses klasifikasi, selain itu SVM mewakili sekumpulan metode pembelajaran (*supervised learning*) yang banyak digunakan dalam pengenalan pola, analisis regresi dan klasifikasi [15][17].

g. *Particle Swarm Optimization* (PSO)

Particle Swarm Optimization (PSO) adalah teknik optimasi yang terinspirasi oleh perilaku sosial kawanan burung atau ikan. Karakteristik PSO secara umum adalah: sederhana konsepnya, komputasinya efisien, dan implementasinya mudah [18]. Sama halnya dengan GA (*Genetik Algorithm*), metode PSO adalah berbasis populasi, tetapi PSO tidak menggunakan konsep persaingan melainkan kerja sama. Partikel ke- i disajikan sebagai: $x_i = (x_{i,1}, x_{i,2}, \dots, x_{i,d})$ dalam ruang d -dimensi. Posisi partikel ke- i terbaik sebelumnya disimpan dan disajikan sebagai $pbest_i = (pbest_{i,1}, pbest_{i,2}, \dots, pbest_{i,d})$. Indeks partikel terbaik di antara semua partikel dalam kawanan *group* disajikan sebagai $gbest$. Kecepatan partikel ke- i disajikan sebagai: $v_i = (v_{i,1}, v_{i,2}, \dots, v_{i,d})$. Modifikasi kecepatan dan posisi tiap partikel dihitung menggunakan persamaan 2-3, memperhitungkan kecepatan saat ini dan jarak $pbest_i$ ke $gbest$. [12], [18].

$$v_{i,m} = w \cdot v_{i,m} + c_1 \cdot R \cdot (pbest_{i,m} - x_{i,m}) + c_2 \cdot R \cdot (gbest_m - x_{i,m}) \quad (2)$$

$$x_{id} = x_{i,m} + v_{i,m} \quad (3)$$

3. Hasil dan Diskusi

Data hasil eksperimen metode klasifikasi disajikan dalam Tabel 3. Hasil uji akurasi dari setiap metode yang diuji dapat dilihat pada Gambar 5. Sedangkan akurasi dan AUC dari setiap metode diilustrasikan pada Gambar 6-7.

Ilustrasi grafik dan tabel hasil eksperimen menggunakan berbagai metode klasifikasi berbasis *Particle Swarm Optimization* menunjukkan bahwa hasil eksperimen metode *Neural Network + Particle Swarm Optimization* adalah yang terbaik dari metode klasifikasi lain yang diuji, dengan akurasi 76.11% dan AUC 0.827%.

4. Kesimpulan

Berdasarkan hasil pembahasan pada penelitian kali ini, maka didapatkan beberapa kesimpulan sebagai berikut:

Penyakit *Diabetic retinopathy* (DR) merupakan salah satu jenis penyakit yang membutuhkan bantuan teknologi untuk melakukan deteksi dini, hal ini dikarenakan sifat dari penyakit ini yang tidak menunjukkan gejala yang nampak sampai menginjak tahap lanjut. Agar dapat membuat sebuah perangkat lunak yang dapat membantu para ahli dalam melakukan deteksi dini pada penyakit DR, maka perlu dilakukan sejumlah percobaan untuk mendapatkan metode terbaik yang nantinya akan digunakan dalam menentukan rule pada perangkat lunak deteksi dini penyakit DR ini.

Setelah melalui serangkaian uji coba, pada penelitian ini didapatkan bahwa dengan menggunakan metode *Neural Network* yang dikombinasikan dengan metode seleksi fitur *Particle Swarm Optimization* dapat memberikan hasil akurasi terbaik bila dibandingkan dengan metode lainnya yang juga diujicobakan pada penelitian kali ini. Penelitian ini dapat dijadikan acuan dalam menentukan rule untuk pengembangan perangkat lunak deteksi awal penyakit DR dengan menerapkan metode *Neural Network* yang dioptimasi menggunakan *Particle Swarm Optimization*.

5. Daftar Pustaka

- [1] T.E. Pardede, D. Rosdiana, E. Christianto, "Gambaran Pengendalian Diabetes Melitus Berdasarkan Parameter Indeks Massa Tubuh dan Tekanan Darah di Poli Rawat Jalan Penyakit Dalam RSUD Arifin Achmad Pekanbaru", JOM FK Vol.4 No.1 Februari 2017
- [2] I. M. Dewi, "Mengenal Lebih Jauh Retinopathy Diabetik," www.columbiaasia.com, 2018. [Online].

- Available: <https://www.columbiaasia.com/indonesia/health-articles/mengenal-lebih-jauh-retinopathy-diabetik>.
- [3] A. Antal, B., & Hajdu, "An ensemble-based system for automatic screening of *diabetic retinopathy*," *Knowl. - Based Syst.*, pp. 20–27, 2014.
- [4] M. L. Valverde, C., Garcia, M., Hornero, R., & Galvez, "Automated detection of *diabetic retinopathy* in retinal images," *Indian J. Ophthalmology*, vol. 64, no. 1, pp. 26–32, 2016.
- [5] A. Herliana, "Optimasi Klasifikasi Sel Tunggal Pap Smear Menggunakan Correlation Based Features Selection (Cfs) Berbasis C4 . 5 Dan Naive Bayes," *J. Inform.*, vol. 3, no. 2, pp. 148–155, 2016.
- [6] T. Arifin, "Implementasi Algoritma PSO Dan Teknik Bagging Untuk Klasifikasi Sel Pap Smear," *J. Inform.*, vol. 4, no. 2, pp. 155–162, 2017.
- [7] "Machine Learning Repository," 2018. [Online]. Available: <https://archive.ics.uci.edu/ml/datasets.html>.
- [8] C. Fan, F. Xiao, and Y. Zhao, "A short-term building cooling load prediction method using deep learning algorithms," *Appl. Energy*, vol. 195, pp. 222–233, 2017.
- [9] W. Sun, B. Zheng, and W. Qian, "Computer aided lung cancer diagnosis with deep learning algorithms," vol. 9785, p. 97850Z, 2016.
- [10] I. H. Witten, *Data Mining Practical Machine learning Tools and Techniques*, vol. 18 Suppl, no. 1. 2017.
- [11] S. A. Pattekari and A. Parveen, "Prediction system for heart disease using Naive Bayes," *Int. J. Adv. Comput. Math. Sci. ISSN*, vol. 3, no. 3, pp. 2230–9624, 2012.
- [12] Y. Zhang, S. Wang, P. Phillips, and G. Ji, "Binary PSO with mutation operator for feature selection using *decision tree* applied to spam detection," *Knowledge-Based Syst.*, vol. 64, pp. 22–31, 2014.
- [13] S. Raschka, "Naive Bayes and Text Classification I - Introduction and Theory," pp. 1–20, 2014.
- [14] A. Pamungkas, "Algoritma k-means clustering dan Naive Bayes classifier untuk Pengenalan Pola Tesktur," *pemrogramanmatlab.com*, 2016. [Online]. Available: <https://pemrogramanmatlab.com/2016/06/24/algoritma-k-means-clustering-dan-naive-bayes-classifier-untuk-pengenalan-pola-tesktur/>.
- [15] L. Khedher, J. Ramírez, J. M. Górriz, A. Brahim, and F. Segovia, "Early diagnosis of Alzheimer's disease based on partial least squares, principal component analysis and *Support vector machine* using segmented MRI images," *Neurocomputing*, vol. 151, no. P1, pp. 139–150, 2015.
- [16] G. Cavallaro, M. Riedel, M. Richerzhagen, J. A. Benediktsson, and A. Plaza, "On Understanding Big Data Impacts in Remotely Sensed Image Classification Using *Support vector machine* Methods," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, pp. 1–13, 2015.
- [17] C. Cortes and V. Vapnik, "*Support vector machine*," *Mach. Learn.*, pp. 1303–1308, 1995.
- [18] F. Marini and B. Walczak, "*Particle Swarm Optimization* (PSO). A tutorial," *Cbemor. Intell. Lab. Syst.*, vol. 149, pp. 153–165, 2015.