

PEMBOBOTAN ATRIBUT PSO UNTUK OPTIMASI SVM DALAM KASUS KELAYAKAN KREDIT BANK

Wiwik Setyaningsih¹, Sri Mujiyono²

^{1,2} Jurusan SI Teknik Informatika, Universitas Ngudi Waluyo

Email : wiwik2823@gmail.com¹,

Abstract— *Credit success in a bank plays a role in maintaining the survival of a bank. Therefore it is very necessary to measure creditworthiness accurately to classify customers with good credit and bad credit. Based on these conditions the right data mining technique to use is classification. One of the data mining classification techniques is Naïve Bayes Classifier (NBC), but the accuracy is still less than the C4.5 and SVM algorithms. This final report describes the steps of research using the Particle Swarm Optimizatin (PSO) algorithm to weight attributes to increase the accuracy of SVM. This study uses data set public German Credit Data. The validation process uses tenfold-cross validation, while testing the model using confusion matrix and ROC curve. The results show SVM accuracy increased from 74.6% to 76.50% after combined with PSO.*

Keywords— **Credit, German Credit Data, SVM-PSO.**

I. PENDAHULUAN

Kredit menurut UU Perbankan No. 10 tahun 1998 adalah penyediaan uang berdasarkan kesepakatan antara peminjam dan pemberi pinjaman yang mengharuskan peminjam melunasinya beserta bunganya dalam jangka waktu tertentu.

Kredit dalam perbankan sangat mempengaruhi pembiayaan perekonomian nasional dan berfungsi sebagai penggerak pertumbuhan ekonomi (Utari D, G.A., dkk, 2012). Proses penilaian kredit bagi bank tidak dapat diremehkan, karena bagi bank sangat penting untuk membedakan antara UKM baik dan buruk dalam hal kredit melalui proses penilaian kredit tersebut (Guidici, P., and Figini, S ,2009). Dalam kondisi keuangan yang sangat lemah dengan jumlah kredit bermasalah yang semakin besar, serta likuiditas yang semakin rendah dapat menyebabkan bank kesulitan untuk membiayai kegiatan usahanya (Triandaru, S., dan Budisantoso, T , 2009).

Seperti yang tercantum pada harian suara merdeka yang terbit pada tanggal 18 Maret 2013 bahwa Lembaga Penjamin Simpanan (LPS) telah melakukan pencabutan izin usaha perbankan kepada 47 Bank Perkreditan Rakyat (BPR) dan 1 bank umum sejak tahun 2006 hingga 31 Desember 2012. Oleh karena itu analisa kelayakan kredit sangat penting dilakukan untuk mencegah jumlah kredit bermasalah yang semakin besar guna menjaga kelangsungan hidup sebuah bank.

Beberapa peneliti mengembangkan berbagai teknik untuk permasalahan kredit, diantaranya dengan analisa statistik (konvensional), soft computing dan data mining. Dewasa ini pendekatan data mining lebih berkembang untuk mengatasi berbagai permasalahan termasuk permasalahan kredit bank.

Data mining adalah suatu proses yang bertujuan untuk menemukan pola secara otomatis atau semi otomatis dari data yang sudah ada di dalam basis data yang dimanfaatkan untuk menyelesaikan suatu masalah

(Witten, I. H., dkk, 2011). Data mining memiliki beberapa teknik, diantaranya klasifikasi dan clustering. Teknik klasifikasi adalah teknik pembelajaran yang digunakan untuk memprediksi nilai dari atribut kategori target (Vercellis, Carlo, 2009). Klasifikasi bertujuan untuk membagi objek yang ditugaskan hanya ke salah satu nomor kategori yang disebut kelas [15]. Clustering mengelompokkan objek atau data berdasarkan kemiripan antar data, sehingga anggota dalam satu kelompok memiliki banyak kemiripan dibandingkan dengan kelompok lain (Gorunescu, F., 2011). Untuk menyelesaikan masalah analisa resiko kredit data akan diklasifikasikan menjadi dua kelas, yaitu kredit baik dan kredit buruk. Sehingga tepat menggunakan teknik klasifikasi data mining. Metode yang paling populer digunakan untuk teknik klasifikasi adalah Decision Trees, Naïve Bayes Classifiers (NBC), Statistical analysis, dan lain lain (Gorunescu, F., 2011).

Beberapa peneliti telah menganalisa kelayakan kredit dengan metode klasifikasi data mining, diantaranya adalah yang dilakukan oleh Scott A. Z, Kevin B. K dan Ann E. N menggunakan Bayesian Network dan NBC (Zonneveldt, S., dkk, 2010). Henny Leidiyana pada tahun 2011 menggunakan algoritma C4.5, naïve bayes, dan neural network (Leidiyana, H., 2012). Dan Siti Marsipah pada tahun 2011 menggunakan algoritma C4.5 dan C4.5 berbasis Particle Swarm Optimization (PSO). SVM terbukti memiliki akurasi yang lebih tinggi dari C4.5 diberbagai kasus klasifikasi.

Kredit menurut UU Perbankan No. 10 tahun 1998 adalah penyediaan uang berdasarkan kesepakatan antara peminjam dan pemberi pinjaman yang mengharuskan peminjam melunasinya beserta bunganya dalam jangka waktu tertentu.

Kredit dalam perbankan sangat mempengaruhi pembiayaan perekonomian nasional dan berfungsi sebagai penggerak pertumbuhan ekonomi (Utari D, G.A.,

dkk, 2012). Proses penilaian kredit bagi bank tidak dapat diremehkan, karena bagi bank sangat penting untuk membedakan antara UKM baik dan buruk dalam hal kredit melalui proses penilaian kredit tersebut (Guidici, P., and Figini, S, 2009). Dalam kondisi keuangan yang sangat lemah dengan jumlah kredit bermasalah yang semakin besar, serta likuiditas yang semakin rendah dapat menyebabkan bank kesulitan untuk membiayai kegiatan usahanya (Triandaru, S., dan Budisantoso, T, 2009). Seperti yang tercantum pada harian suara merdeka yang terbit pada tanggal 18 Maret 2013 bahwa Lembaga Penjamin Simpanan (LPS) telah melakukan pencabutan izin usaha perbankan kepada 47 Bank Perkreditan Rakyat (BPR) dan 1 bank umum sejak tahun 2006 hingga 31 Desember 2012. Oleh karena itu analisa kelayakan kredit sangat penting dilakukan untuk mencegah jumlah kredit bermasalah yang semakin besar guna menjaga kelangsungan hidup sebuah bank.

Beberapa peneliti mengembangkan berbagai teknik untuk permasalahan kredit, diantaranya dengan analisa statistik (konvensional), soft computing dan data mining. Dewasa ini pendekatan data mining lebih berkembang untuk mengatasi berbagai permasalahan termasuk permasalahan kredit bank.

Data mining adalah suatu proses yang bertujuan untuk menemukan pola secara otomatis atau semi otomatis dari data yang sudah ada di dalam basis data yang dimanfaatkan untuk menyelesaikan suatu masalah (Witten, I. H., dkk, 2011). Data mining memiliki beberapa teknik, diantaranya klasifikasi dan clustering. Teknik klasifikasi adalah teknik pembelajaran yang digunakan untuk memprediksi nilai dari atribut kategori target (Vercellis, Carlo, 2009). Klasifikasi bertujuan untuk membagi objek yang ditugaskan hanya ke salah satu nomor kategori yang disebut kelas [15]. Clustering mengelompokkan objek atau data berdasarkan kemiripan antar data, sehingga anggota dalam satu kelompok memiliki banyak kemiripan dibandingkan dengan

B. KERANGKA KERJA

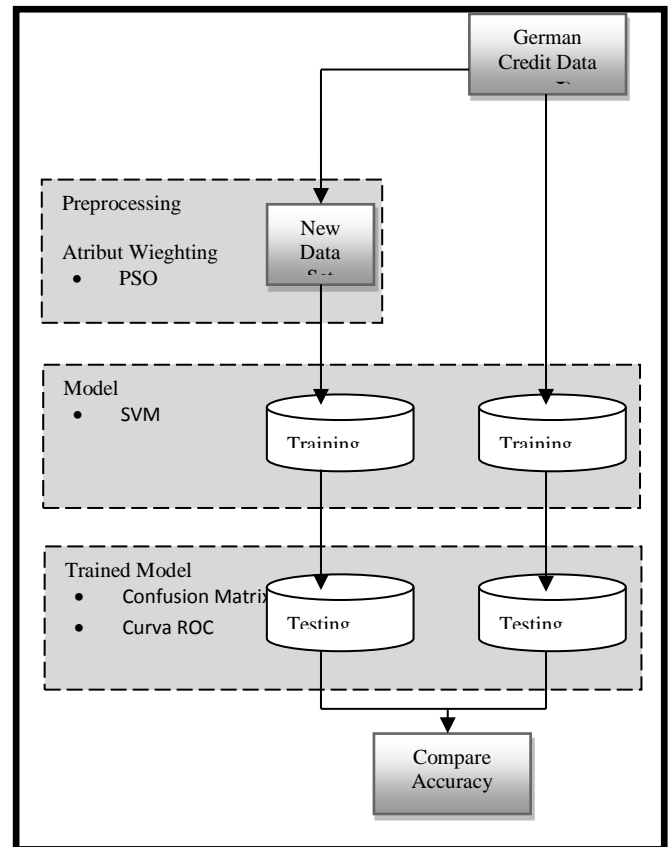
kelompok lain (Gorunescu, F. , 2011). Untuk menyelesaikan masalah analisa resiko kredit data akan diklasifikasikan menjadi dua kelas, yaitu kredit baik dan kredit buruk. Sehingga tepat menggunakan teknik klasifikasi data mining. Metode yang paling populer digunakan untuk teknik klasifikasi adalah Decision Trees, Naïve Bayes Classifiers (NBC), Statistical analysis, dan lain lain (Gorunescu, F. , 2011).

Beberapa peneliti telah menganalisa kelayakan kredit dengan metode klasifikasi data mining, diantaranya adalah yang dilakukan oleh Scott A. Z, Kevin B. K dan Ann E. N menggunakan Bayesian Network dan NBC (Zonneveldt, S., dkk, 2010). Henny Leidiyana pada tahun 2011 menggunakan algoritma C4.5, naïve bayes, dan neural network (Leidiyana, H , 2012). Dan Siti Marsipah pada tahun 2011 menggunakan algoritma C4.5 dan C4.5 berbasis Particle Swarm Optimization (PSO). SVM terbukti memiliki akurasi yang lebih tinggi dari C4.5 diberbagai kasus klasifikasi.

II. METODE PENELITIAN

A. Pengumpulan Data

Data yang digunakan pada penelitian ini berasal dari University of California, Irvine (UCI) Machine Learning dengan judul German Credit data. Data yang disediakan oleh Professor Dr. Hans Hofmann ini berjumlah 1000 record dan terdiri dari 20 atribut, dengan 7 atribut bertipe numerik dan 13 bertipe kategorikal.



1. Pre-processing

Dalam tahap ini, data yang diperoleh dari UCI dalam bentuk text bertipe txt. untuk dapat digunakan pada RapidMiner, data tersebut harus diubah dalam bentuk sheet bertipe csv atau xls.

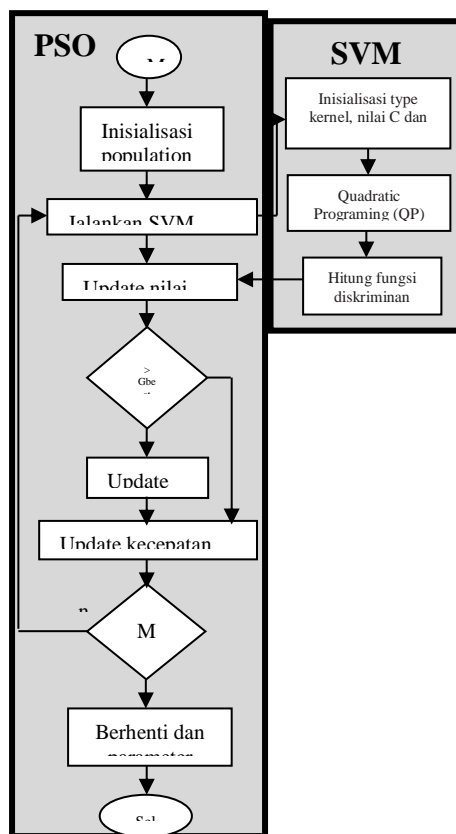
2. Pemodelan menggunakan SVM-PSO

Pada tahap ini PSO digunakan untuk meningkatkan performa dari SVM. Metode SVM-PSO yang diusulkan seperti gambar 3.8 (Melgani, F & Bazi, Y., 2008).

Implementasi PSO dilakukan dengan mengubah parameter population size dengan nilai antara 5 sampai 100. Setelah mendapatkan nilai population size dengan

akurasi tertinggi kemudian merubah nilai parameter maximum number of generation antara 30 sampai 250.

Implementasi SVM pada tahap ini menggunakan 4 kernel yang sering digunakan linier, polinomial, sigmoid dan radial. Percobaan pertama menerapkan kernel-kernel tersebut. Setelah ditemukan kernel dengan akurasi tertinggi, kemudian percobaan kedua mengubah parameter C dengan nilai antara 0,1 sampai dengan 1 dan antara 5 sampai dengan 35. Hasil percobaan dengan merubah beberapa parameter tersebut dibandingkan dan diambil akurasi yang paling tinggi.



Gambar 3.8 Flowchart usulan metode PSO-SVM.

3. Validasi dan evaluasi

Dalam tahap ini dilakukan validasi dan pengukuran keakuratan hasil yang dicapai

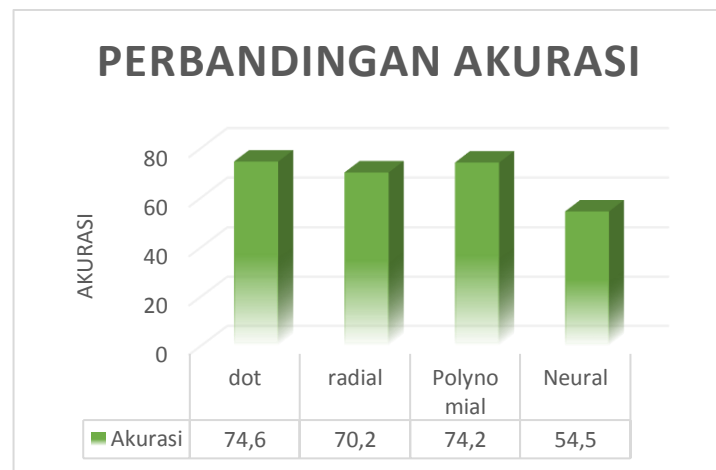
oleh model menggunakan teknik *confusion matrix* dan kurva ROC untuk pengukuran akurasi model.

III. HASIL PERCOBAAN

Percobaan pertama dilakukan dengan menggunakan SVM saja kemudian dengan SVM-PSO. Hasil akurasi percobaan pertama dilakukan dengan mengubah metode *dot*, *radial*, *polynomial* dan *neural*. Kemudian percobaan kedua dengan pembobotan atribut PSO.

Tabel 4.1 Hasil perbandingan akurasi percobaan pertama

Kernel	Akurasi	AUC
Dot	74,6	0,786
Radial	70,2	0,666
Polynomial	74,2	0,768
Neural	54,5	0,577



Gambar 4.1 Perbandingan akurasi antar kernel.

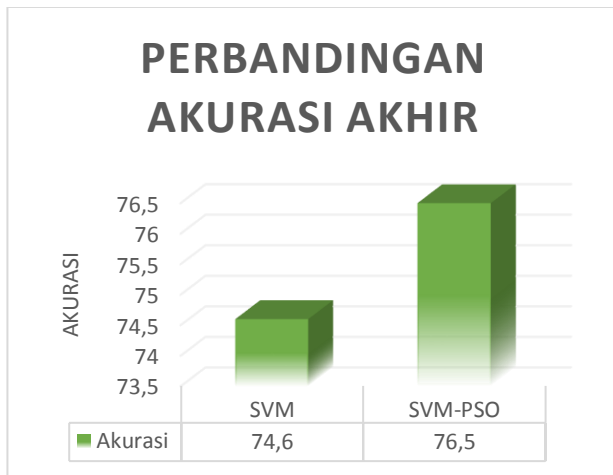
Percobaan kedua menghasilkan akurasi sebesar 76,5% dan AUC 0,791. Sehingga terbukti PSO dapat meningkatkan akurasi SVM pada data kelayakan kredit bank.

IV. PEMBAHASAN

Hasil dari seluruh percobaan menghasilkan akurasi seperti tercantum pada tabel dibawah ini :

Tabel 4.2 Hasil Percobaan

Metode	Akurasi	AUC
SVM	74,6	0,786
SVM-PSO	76,5	0,791



Gambar 4.2 Perbandingan akurasi akhir.

Terlihat bahwa metode yang diusulkan pada penelitian ini memiliki akurasi tertinggi dibandingkan dengan metode lainnya yaitu dengan akurasi 76,5% dan AUC 0,791. Metode SVM sebelumnya memiliki akurasi terendah dengan 74,6% dan AUC 0,786.

V. PENUTUP

Hasil penelitian menunjukkan metode yang diusulkan pada penelitian ini memiliki akurasi tertinggi dibandingkan dengan metode lainnya yaitu dengan akurasi 76,5% dan AUC 0,791. Metode SVM sebelumnya memiliki akurasi terendah dengan 74,6% dan AUC 0,786.

Dengan demikian untuk klasifikasi kelayakan kredit bank, metode SVM-PSO yang diusulkan terbukti memiliki akurasi yang lebih tinggi dibandingkan dengan metode SVM.

Untuk penelitian selanjutnya sebaiknya

menggunakan data privat atau data publik yang lainnya. Dan mencoba membandingkan dengan metode klasifikasi dan metode.

REFERENSI

- [1] Abraham, A., Grosan, C., & Ramos, V. (2006). *Swarm Intelligence In Data Mining*. Verlag Berlin Heidelberg: Springer.
- [2] Agmalaro, M. A., Kustiyo, A., & Akbar, A. R. (2013). Identifikasi Tanaman Buah Tropika Berdasarkan Tekstur Permukaan Daun Menggunakan Jaringan Syaraf Tiruan. *Jurnal Ilmu Komputer dan Agri-Informatika*, 2(2).
- [3] Anami, B. S., Suvarna, S. N., & Govardhan, A. (2010). A combined color, texture and edge features based approach for identification and classification of indian medicinal plants. *International Journal of Computer Applications*, 6(12), 45-51.
- [4] Ehsanirad, A., & Sharath Kumar, Y. H. (2010). Leaf recognition for plant classification using GLCM and PCA methods. *Oriental Journal of Computer Science and Technology*, 3(1), 31-36.
- [5] Elhariri, E., El-Bendary, N., & Hassanien, A. E. (2014, December). Plant classification system based on leaf features. In *Computer Engineering & Systems (ICCES), 2014 9th International Conference on* (pp. 271-276). IEEE.
- [6] Gorunescu, F. (2011). *Data Mining Concepts, Models And Techniques*. Verlag Berlin Heidelberg: Springer.
- [7] Hadiwijaya, T. 1986. *Cengkeh: Data dan Petunjuk ke Arah Swa Sembada*. PT. Gunung Agung, Jakarta.
- [8] Handayanna, F. "Penerapan Particle Swarm Optimization Untuk Seleksi Atribut Pada Metode Support Vector Machine Untuk Prediksi Penyakit Diabetes," Tesis Magister Ilmu Komputer. Sekolah Tinggi Manajemen Informatika dan Komputer Nusa Mandiri, 2012.
- [9] Haralick, R. M., Shanmugam, K., & Dinstein, I. H. (1973). Textural features for image classification. *Systems, Man and Cybernetics, IEEE Transactions on*, (6), 610-621.
- [10] Hastutiningrum, N.O, "Efek Minyak Atsiri Daun Cengkeh (*Syzygium aromaticum* L.) terhadap Mortalitas Larva *Anopheles aconitus*," Skripsi, Universitas Sebelas Maret Surakarta, 2010.
- [11] Ika Novichasari, S. "Klasifikasi Daun Cengkeh Berdasarkan Tekstur Permukaan Daun Menggunakan GLCM Dan PSO-SVM," Tesis Magister Ilmu Komputer. Universitas Dian Nuswantoro, 2015
- [12] Karunamoorthy, B., & Somasundereswari, D. (2013). A Defect Tea Leaf Identification Using Image Processing. *Przeglad Elektrotechniczny*, 89, 318-320.
- [13] Kennedy, J. (2010). Particle swarm optimization. In *Encyclopedia of Machine Learning* (pp. 760-766). Springer US.
- [14] Kulkarni, A. H., Rai, H. M., Jahagirdar, K. A., & Upparamani, P. S. (2013). A Leaf Recognition Technique for Plant Classification Using RBPNN and Zernike Moments. *International Journal of Advanced Research in Computer and Communication Engineering*, 2(1), 1-5.
- [15] Kusri, & Luthfi, E. T. (2009). *Algoritma Data Mining*. Yogyakarta: Andi Publishing.
- [16] Larese, M. G., Namias, R., Craviotto, R. M., Arango, M. R., Gallo, C., & Granitto, P. M. (2014). Automatic classification of legumes using leaf vein image features. *Pattern Recognition*, 47(1), 158-168.
- [17] Larose, D. T. (2005). *Discovering Knowledge in Data*. New Jersey: John Wiley & Sons, Inc.
- [18] Listia, R., & Harjoko, A. (2014). Klasifikasi Massa pada Citra Mammogram Berdasarkan Gray Level Cooccurrence Matrix (GLCM). *IJCCS-Indonesian Journal of Computing and Cybernetics Systems*, 8(1), 59-68.
- [19] Melgani, F and Bazi, Y., (2008). Classification of Electrocardiogram Signals With Support Vector Machines and Particle Swarm Optimization. *IEEE Transactions On Information Technology In Biomedicine*, Vol. 12, No. 5, September 2008.

- [20] Mohanaiah, P., Sathyanarayana, P., & GuruKumar, L. (2013). Image texture feature extraction using GLCM approach. *International Journal of Scientific and Research Publications*, 3(5), 1-3.
- [21] Moningka, F. F., Runtuuwu, S. D., & Paulus, J. M. (2012). RESPON PERTUMBUHAN TINGGI! DAN PRODUKSI TANAMAN CENGKEH (*Syzigium arom aticum L.*) TERHADAP PEMBERIAN PACLOBUTRAZOL. *Eugenia*, 18(2), 18-2.
- [22] Muiz, R. 2007. Pedoman Teknis Budidaya Cengkeh. Direktorat Jenderal Perkebunan. Jakarta
- [23] Nurjanah, N. (2004). Diversifikasi Penggunaan Cengkeh,
- [24] Pouteau, R., Meyer, J. Y., Taputuarai, R., & Stoll, B. (2012). Support vector machines to map rare and endangered native plants in Pacific islands forests. *Ecological Informatics*, 9, 37-46.
- [25] Raja, S. D. M., & Shanmugam, A. (2012). ANN and SVM Based War Scene Classification Using Invariant Moments and GLCM Features: A Comparative Study. *Mach. Learn*, 2(6), 869-873.
- [26] Rampun, A., Strange, H., & Zwiiggelaar, R. (2013, June). Texture segmentation using different orientations of GLCM features. In *Proceedings of the 6th International Conference on Computer Vision/Computer Graphics Collaboration Techniques and Applications* (p. 17). ACM.
- [27] Singh, K., Gupta, I., & Gupta, S. (2010). Svm-bdt pnn and fourier moment technique for classification of leaf shape. *International Journal of Signal Processing, Image Processing and Pattern Recognition*, 3(4), 67-78.
- [28] Situmeang, T.H, "Analisis Produksi, Konsumsi, dan Harga Cengkeh Indonesia," Skripsi, Institut Pertanian Bogor, 2008.
- [29] Soh, L. K., & Tsatsoulis, C. (1999). Texture analysis of SAR sea ice imagery using gray level co-occurrence matrices. *Geoscience and Remote Sensing, IEEE Transactions on*, 37(2), 780-795.
- [30] Stefan Fiel and Robert Sablatnig. Automated identification of tree species from images of the bark, leaves and needles.
- [31] Tu, C. J., Chuang, L. Y., Chang, J. Y., & Yang, C. H. (2007). Feature selection using PSO-SVM. *IAENG International journal of computer science*, 33(1), 111-116.
- [32] Vercellis, Carlo (2009). *Business Intelligent: Data Mining and Optimization for Decision Making*. Southern Gate, Chichester, West Sussex: John Willey & Sons, Ltd.
- [33] Wu, Xindong and Kumar, Vipin. (2009). *The Top Ten Algorithms in Data Mining*. Boca Raton: CRC Press.
- [34] Zeniarja, J. "Opinion Mining of Movie Review On Twitter Using Support Vector Machine With Particle Swarm Optimization," Tesis Master of Computer Science. Universiti Teknikal Malaysia Melaka. 2012.
- [35]
- [36] Rampun, A., Strange, H., & Zwiiggelaar, R. (2013, June). Texture segmentation using different orientations of GLCM features. In *Proceedings of the 6th International Conference on Computer Vision/Computer Graphics Collaboration Techniques and Applications* (p. 17). ACM.
- [37] Singh, K., Gupta, I., & Gupta, S. (2010). Svm-bdt pnn and fourier moment technique for classification of leaf shape. *International Journal of Signal Processing, Image Processing and Pattern Recognition*, 3(4), 67-78.
- [38] Situmeang, T.H, "Analisis Produksi, Konsumsi, dan Harga Cengkeh Indonesia," Skripsi, Institut Pertanian Bogor, 2008.
- [39] Soh, L. K., & Tsatsoulis, C. (1999). Texture analysis of SAR sea ice imagery using gray level co-occurrence matrices. *Geoscience and Remote Sensing, IEEE Transactions on*, 37(2), 780-795.
- [40] Stefan Fiel and Robert Sablatnig. Automated identification of tree species from images of the bark, leaves and needles.
- [41] Tu, C. J., Chuang, L. Y., Chang, J. Y., & Yang, C. H. (2007). Feature selection using PSO-SVM. *IAENG International journal of computer science*, 33(1), 111-116.
- [42] Vercellis, Carlo (2009). *Business Intelligent: Data Mining and Optimization for Decision Making*. Southern Gate, Chichester, West Sussex: John Willey & Sons, Ltd.
- [43] Wu, Xindong and Kumar, Vipin. (2009). *The Top Ten Algorithms in Data Mining*. Boca Raton: CRC Press.
- [44] Zeniarja, J. "Opinion Mining of Movie Review On Twitter Using Support Vector Machine With Particle Swarm Optimization," Tesis Master of Computer Science. Universiti Teknikal Malaysia Melaka. 2012.