

KLASIFIKASI MAHASISWA HER BERBASIS ALGORITMA SVM DAN DECISION TREE

Jajang Jaya Purnama^{*1}, Hendri Mahmud Nawawi², Susy Rosyida³, Ridwansyah⁴, Risnandar⁵

^{1,2}Program Studi Teknologi Komputer, Universitas Bina Sarana Informatika

^{3,4}Program Studi Teknik Informatika, Sekolah Tinggi Manajemen Informatika dan Komputer Nusa Mandiri

⁵Pusat Penelitian Informatika-LIPI

Email: ¹jajangja2412@bsi.ac.id, ²hendri.hiw@bsi.ac.id, ³susyrosyida@gmail.com, ⁴rdwansyah@gmail.com,

⁵risnandar01@gmail.com

*Penulis Korespondensi

(Naskah masuk: 16 Januari 2020, diterima untuk diterbitkan: 26 November 2020)

Abstrak

Mahasiswa di setiap perguruan tinggi dituntut untuk memperoleh pengetahuan dan keterampilan yang memenuhi syarat dengan prestasi akademik. Hasil dari pembelajaran mahasiswa didapat dari ujian teori dan praktek, setiap mahasiswa wajib menuntaskan nilai sesuai kriteria kelulusan minimum dari masing-masing dosen pengajar, jika dibawah batas minimum maka mahasiswa mengikuti her. Her adalah salah satu cara untuk menuntaskan kriteria kelulusan minimum. Mahasiswa yang mengikuti her setiap semesternya hampir mencapai angka yang relatif tinggi dari jumlah seluruh mahasiswa. Untuk mengurangi jumlah mahasiswa yang mengikuti her maka dibutuhkan sebuah metode yang dapat mengurangi hal tersebut, dengan metode *Support Vector Machine* (SVM) dan *Decision Tree* (DT). SVM dan DT adalah salah satu metode klasifikasi *supervised learning*. Oleh karena itu, dalam penelitian ini menggunakan SVM dan DT. SVM dapat menghilangkan hambatan pada data, memprediksi, mengklasifikasikan dengan sampling kecil dan dapat meningkatkan akurasi dan mengurangi kesalahan. Klasifikasi data siswa yang melakukan her/peningkatan dengan mengimprovisasi model kernel untuk visualisasi termasuk bar, histogram, dan sebaran begitu juga *Decision Tree* mempunyai kelebihan tersendiri. Dari hasil penelitian ini telah didapatkan akurasi dan presisi model DT lebih besar dibandingkan dengan SVM, akan tetapi untuk *recall* DT lebih kecil dibandingkan SVM.

Kata kunci: Data Mahasiswa, Her, SVM, Decision Tree

CLASSIFICATION OF HER STUDENTS BASED ON SVM ALGORITHM AND DECREE TREE

Abstract

Students in each tertiary institution are required to obtain knowledge and skills that meet the requirements with academic achievement. The results of student learning are obtained from the theory and practice exams, each student is required to complete grades according to the minimum graduation criteria of each teaching lecturer, if below the minimum limit then students take remedial. Remedial is one way to complete the minimum passing criteria. Students who take remedial every semester almost reach a relatively high number of the total number of students. To reduce the number of students who take remedial, a method that can reduce this is needed, with the Support Vector Machine (SVM) and Decision Tree (DT) methods. SVM and DT are one of the supervised learning classification methods. Therefore, in this study using SVM and DT. SVM can eliminate barriers to data, predict, classify with small sampling and can improve accuracy and reduce errors. Data classification of students who do remedial/improvements by improving the kernel model for visualization including bars, histograms, and distributions as well as the Decision Tree has its own advantages. From the results of this study it has been obtained that the accuracy and precision of DT models is greater than that of SVM, but for recall DT is smaller than SVM.

Keywords: Student Data, Remedial, SVM, Decision Tree

1. PENDAHULUAN

Ujian merupakan agenda rutinitas setiap semester yang harus dilakukan mahasiswa untuk

menunjang dan juga menuntaskan program studinya di sebuah perguruan tinggi (Mukhlis & Koentjoro, 2015) Salah satu faktor yang dapat mempengaruhi mahasiswa mendapatkan nilai kecil pada mata kuliah

yang diambil karena mahasiswa tersebut menjalani kuliah dan bekerja, sehingga mengakibatkan harus mengikuti remedial. Remedial bertujuan untuk mengevaluasi setiap mahasiswa yang nilai UAS dan UTS nya kurang dari ketentuan kriteria minimum dalam kurikulum yang berlaku pada perguruan tinggi. Setiap semesternya tidak sedikit mahasiswa yang mengikuti her atau remedial, sistem remedial yang diikuti dengan mengulang kembali mata kuliah tersebut atau memperbaiki nilai dari mata kuliah tersebut.

Perguruan tinggi harus mempunyai target dalam mengurangi mahasiswa yang mengikuti her. Untuk mencapai target tersebut perguruan tinggi dapat mencari sebuah keputusan agar mahasiswa tidak mengalami kegagalan dalam pelajaran dengan kumpulan data besar yang dapat menghabiskan waktu dan akurasi yang dihasilkan kurang akurat (Rao, Mangathayaru, & Rao, 2017). Dengan data yang besar dan pemilihan variabel yang sedikit itu dapat membuang banyak informasi yang tersedia, maka dibutuhkan *Support Vector Machine* (SVM) yang dapat sepenuhnya mengekstraksi sebuah informasi yang berguna terhadap indikator yang diberikan, meminimalkan kesalahan model dan memaksimalkan akurasi model.

SVM memiliki keunggulan komputasi dalam menangani klasifikasi multiclass (Sahu, Pujari, Kagita, Kumar, & Padmanabhan, 2015), dengan berdasarkan teori pembelajaran statistik dengan sampling kecil, dan dimensi tinggi (Li & Zhang, 2009). SVM dapat memodelkan masalah dunia nyata yang kompleks seperti klasifikasi teks dan gambar, pengenalan tulisan tangan, dan bioinformatika serta analisis biosequence. Pendekatan ini membantu menghilangkan hambatan pada data (Krishnan & K, 2018). Algoritma yang sangat kuat terutama digunakan untuk analisis regresi, prediksi dan klasifikasi margin berbasis kernel (Lee & Yoon, 2017). Kinerja SVM dengan mengandalkan parameter regularisasi dan kernel yang digunakan dalam mesin vector dukungan untuk memetakan data pembelajaran (non linier) ke dalam ruang fitur dimensi yang lebih tinggi di mana daya komputasi mesin pembelajaran linier meningkat. SVM juga dapat digunakan dalam meningkatkan akurasi sehingga mencapai yang tertinggi (Riyanto, Hamid, & Ridwansyah, 2019). SVM juga dapat meningkatkan hasil diagnosis dan mengurangi kesalahan (Zou, Zheng, Xu, & Chen, 2010).

Decision Tree adalah pendekatan yang paling kuat dalam penemuan pengetahuan dan *data mining*. *Decision Tree* termasuk teknologi penelitian besar dan kompleks data untuk menemukan pola yang bermanfaat. Gagasan ini sangat penting karena memungkinkan pemodelan dan ekstraksi pengetahuan dari sebagian besar data yang tersedia. Semua ahli teori dan spesialis terus mencari teknik untuk membuat proses lebih efisien, hemat biaya dan akurat. *Decision Tree* adalah alat yang sangat efektif

di banyak bidang seperti *data mining* dan teks, ekstraksi informasi, *machine learning* dan pengenalan pola (Sharma, Bhargava, & Mathuria, 2013). Algoritma *decision tree* termasuk klasifikasi yang sederhana namun efektif. *Decision tree* mempunyai kelebihan yaitu memberikan rules yang dapat dipahami oleh manusia. *Decision tree* memiliki beberapa kekurangan salah satunya membutuhkan waktu untuk menyortir semua atribut numerik untuk menentukan keputusan yang hasilnya harus simpul, ini menjadi salah satu yang memperlambat dan memakan ukuran memori terutama dataset yang diuji berjumlah jutaan (Ben-Haim & Tom-Tov, 2010).

Karakteristik dari dataset yang akan diuji dalam penelitian ini, mempunyai 2 *class* numerik kategori “her dan tidak her” sangat relevan dengan model SVM dan *decision tree*. Dengan penelitian ini SVM dan *decision tree* dapat membantu sebuah perguruan tinggi untuk mengurangi jumlah her guna meningkatkan mutu mahasiswa.

2. METODE PENELITIAN

2.1 Support Vector Machine

Support Vector Machine (SVM) adalah metode pembelajaran mesin baru berdasarkan teori pembelajaran statistik dan itu diklasifikasikan sebagai salah satu pendekatan komputasi yang dikembangkan oleh Vapnik (Mat Deris, Mohd Zain, & Sallehuddin, 2011). Awalnya, SVM dikembangkan untuk menyelesaikan masalah klasifikasi dua kelas tetapi kemudian dirumuskan dan diperluas untuk menyelesaikan klasifikasi multi-kelas (Tomar & Agarwal, 2015).

SVM memperkenalkan dua parameter biaya $C +$ dan $C -$ untuk menyatakan setiap biaya false negative dan false positive, dan mempertimbangkan formulasi primal berikut: (Gu, Sheng, Tay, Romano, & Li, 2017).

$$\min(w, b, s) = \frac{1}{2} (w, w) + C + \sum_{i \in S+} \epsilon_i + C - \sum_{i \in S-} \epsilon_i \quad (1)$$

SVM didasarkan pada teori dimensi VC dan prinsip minimum risiko struktur teori statistik, yang memetakan vector input ke dalam ruang fitur dimensi tinggi dengan transformasi non linier dan menggunakan fungsi kernel teori yang relevan untuk menggantikan operasi produk titik dalam ruang dimensi tinggi ketika fungsi keputusan optimal diselesaikan dalam ruang dimensi tinggi. Memecahkan masalah perhitungan regresi non linier di ruang sampel. Ide dasar regresi SVM adalah untuk melatih sampel yang diberikan (Y. Feng et al., 2019). SVM juga bekerja secara efektif untuk kategorisasi

sampel data yang tidak diklasifikasikan linier dengan memanfaatkan teori fungsi kernel (Tomar & Agarwal, 2015).

$$\{(X_i, Y_i) | x \in R^n, y \in R, i = 1, 2, \dots, m\} \quad (2)$$

di mana X_j adalah vector input; y_i adalah vector keluaran.

Model prediksi adalah:

$$f(x) = w^T \phi(x) + b \quad (3)$$

2.2 Kernel

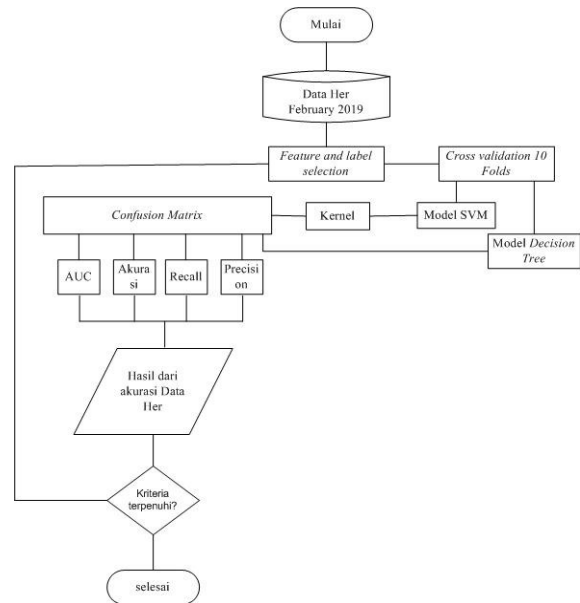
Fungsi kernel pada dasarnya digunakan untuk mengganti produk dalam operasi antara dua titik sampel dalam dimensi tinggi ruang (X. Feng, Ling, Zheng, Chen, & Xu, 2019). SVM bekerja dengan menyematkan data ke fitur dimensi tinggi ruang dan kernel dalam SVM memainkan peran utama dalam formulasi model. Inti trik dalam SVM memastikan bahwa kita tidak perlu menghitung fungsi embedding secara eksplisit selama kita dapat membangun matriks kernel yang tepat (Jiang, Ching, Yiu, & Qiu, 2018). Cara memilih fungsi kernel tergantung pada distribusi sampel data dan hubungan antara data sampel dan variabel yang diprediksi (X. Feng et al., 2019). Karena ruang fitur berbeda distribusi data yang berbeda, kinerja SVM tergantung sebagian besar pada pilihan fungsi kernel. Fungsi kernel bisa dibagi menjadi tipe lokal dan global, fungsi kernel lokal memiliki kemampuan belajar yang baik tetapi kemampuan generalisasi yang lemah, sementara fungsi kernel global memiliki kemampuan generalisasi yang baik tetapi kemampuan belajarnya lemah (X. Feng et al., 2019).

2.3 Decision Tree

Decision Tree dapat dibangun relatif cepat dibandingkan dengan metode klasifikasi lainnya. *Trees* dapat dengan mudah dikonversi menjadi pernyataan SQL yang dapat digunakan untuk mengakses database secara efisien. Pengklasifikasi *Decision Tree* memperoleh akurasi yang serupa dan terkadang lebih baik jika dibandingkan dengan metode klasifikasi lainnya (Priyam, Gupta, Rathee, & Srivastava, 2013).

2.4 Model Yang Diusulkan

Sebelum melakukan klasifikasi data kami melakukan proses pembersihan data seperti data kosong, data tidak lengkap yang dapat menurunkan hasil penelitian. Dengan algoritma yang kami usulkan akan dijelaskan secara rinci



Gambar 1. Model Penelitian yang diusulkan

Metode yang diusulkan oleh penulis metode supervised learning algoritma SVM dengan improvisasi kernel yang terlihat dalam gambar diatas dengan catatan data sudah tidak ada yang kosong atau bias, data yang sudah lengkap siap diuji menggunakan metode SVM dengan menyeleksi atribut dan pemilihan label terlebih dahulu. Setelah seleksi data akan diuji validasi dengan 10-cross-validation dengan menggunakan data training maupun data testing yang akan dilakukan dengan model SVM. Jika model diuji sudah baik maka penelitian selesai, jika masih belum memenuhi kriteria akan dilakukan peseleksian kembali terhadap data.

3. PENELITIAN TERKAIT

Penelitian tentang her pernah dilakukan oleh (Vasani & Gawali, 2013) yang dikumpulkan dari siswa lembaga politeknik, penelitian ini menggunakan model J48 dan NB dengan hasil j48 ROC 0.957, NB dengan ROC 0.975 menghapus beberapa atribut untuk meningkatkan akurasi. Alat tools yang dipakai oleh penulis adalah WEKA.

Peneliti lain telah melakukan prediksi mahasiswa yang mengikuti her dikarenakan mahasiswa yang tidak dapat memahami materi perkuliahan dan kurangnya dasar pengetahuan mereka adalah faktor yang paling signifikan dalam melemahkan efektifitas belajar mereka, dengan permasalahan bagaimana menemukan beberapa faktor yang terkait sehingga menemukan hasil untuk mencari tahu solusi dari permasalahan tersebut (Zhang, 2008)

Banyak penelitian telah dilakukan untuk mengambil struktur dari data yang diperhitungkan dalam pendekatan SVM. Sebagai contoh, kernel grafik dan urutan telah diusulkan untuk mengklasifikasikan data terstruktur yang sesuai

seperti bahan kimia senyawa atau sekuens DNA (Cuingnet, Glaunés, Chupin, Benali, & Colliot, 2013).

Penelitian lain tentang SVM dengan menyajikan kerangka kerja untuk memperkenalkan prior spasial dan anatomi dalam SVM untuk analisis citra otak berdasarkan operator regularisasi. Gagasan kedekatan berdasarkan pengetahuan anatomi sebelumnya antara titik-titik gambar didefinisikan oleh grafik (Cuingnet et al., 2013).

Penelitian lain dengan menggunakan algoritma SVM membuat indikator dasar untuk peramalan jangka pendek yang sangat penting untuk microgrid grup ladang minyak lepas pantai. Hasilnya menunjukkan bahwa algoritma DA-SVM memiliki langkah yang lebih mudah, kemampuan pencarian global yang lebih baik, akurasi prediksi yang lebih tinggi, dan kecepatan komputasi yang lebih baik (Y. Feng, Zhang, Yang, Li, & Zhang, 2019).

4. HASIL DAN PEMBAHASAN

Pengujian data her ini menggunakan dataset sebanyak 2264 yang didapatkan dari salah satu perguruan tinggi terkemuka yang ada di Indonesia. Data didapat dari hasil preprocessing, atribut JK (Jenis Kelamin) dengan value laki-laki menjadi 0, value perempuan menjadi 1, atribut KLH_SMBL_KRJ (Kuliah Sambil Kerja) dengan value tidak kerja menjadi 0, value kerja menjadi 1, value wirausaha menjadi 2, atribut jarak dengan value jauh menjadi 0, value dekat menjadi 1 dan atribut her dengan value mahasiswa pernah her menjadi 1, value mahasiswa tidak pernah her menjadi 0. Data set yang akan diuji sebagai berikut.

Tabel 1. Sample Dataset Her

NIM	USIA	J K	KLH_S MBL_K RJ	JARA K	HER
11170040	21	0	1	0	1
11170052	27	1	1	0	1
11170053	21	1	1	0	1
11170058	21	1	1	0	1
11170059	28	1	1	0	1
11170074	21	1	1	0	1
11170092	24	1	1	0	1
64191582	18	1	2	0	0
64191597	17	1	2	1	0
64191604	21	1	2	0	0
64191606	19	1	2	1	0
64191635	21	1	2	0	0
64191653	19	0	2	0	0
64191659	24	0	2	0	0

Dalam penelitian pengklasifikasi data her menggunakan algoritma *Support Vector Machine* (SVM), dalam penelitian ini menggunakan improvisasi model kernel untuk visualisasi diantaranya bars, histogram dan scatter.

A. Hasil pengujian algoritma

Hasil pengujian algoritma *Support Vector Machine* (SVM) terhadap dataset her dengan menggunakan model kernel dihasilkan sebagai berikut:

Tabel 2. Pengujian SVM

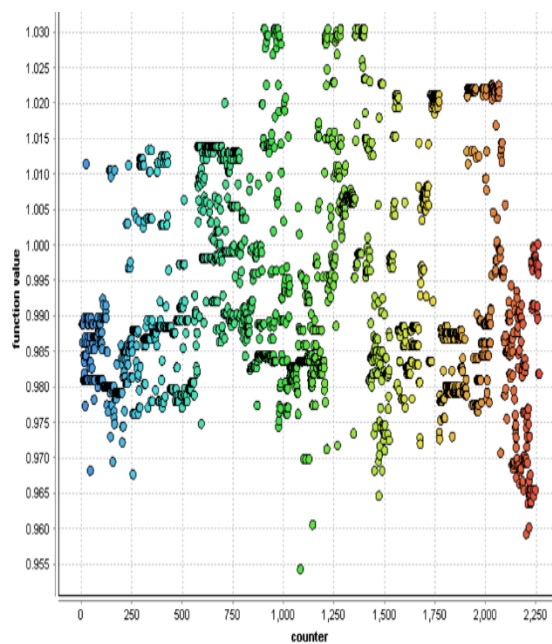
Akurasi	76.80 %
Precision	76.83 %
Recall	99.94 %
AUC	0.584

Dari tabel diatas pengujian SVM diatas dapat dijelaskan bahwa pengujian data her dengan model SVM memiliki akurasi 76.80 %, precision 76,83 % dan AUC 0.584 %

Tabel 3. Confusion matrix

	true 1	true 0	Calss precision
Pred. 1	0	1	0.00%
Pred. 2	524	1738	76.83% ^s
Class recall	0.00%	99.94%	

Dari tabel diatas *confusion matrix* SVM data her yang diprediksi positif dan ternyata positif ada 0, yang diprediksi positif dan ternyata bukan positif ada 1, yang diprediksi negatif ternyata bukan negatif ada 524, yang diprediksi negatif dan ternyata negatif ada 1738.



Gambar 2. Scatter SVM Data Her

Dari gambar 2 terlihat korelasi antara atribut yang satu dengan lainnya dan menunjukkan keeratan antara atribut sehingga dapat menunjukkan korelasi koefisiannya.

Dari semua atribut yang diuji terbukti dapat menyebabkan mahasiswa mendapatkan nilai yang kurang dan mengakibatkan mahasiswa mengikuti her terhadap beberapa mata kuliah. Oleh karena itu perguruan tinggi dapat mengantisipasi terhadap beberapa atribut yang menyebabkan permasalahan tersebut dengan metode yang sudah di uji coba SVM dan *Decision Tree*.

Berdasarkan hasil tersebut bisa diambil kesimpulan hasil yang akurasi terbesar yaitu Decision Tree sebesar 76.84% dan AUC sebesar 0.748. Tolak ukur dari feature-feature yang ada pada dataset her Februari 2019 bahwa mahasiswa yang her berdasarkan feature tersebut, kedepannya untuk bisa dijadikan evaluasi oleh perguruan tinggi tersebut dan dapat mengurangi mahasiswa yang mengikuti her.

Disarankan kedepannya dilakukan uji coba dengan menambahkan *feature* dataset her atau penambahan optimasi untuk memaksimalkan confusion matrix pada model

DAFTAR PUSTAKA

- BEN-HAIM, Y., & TOM-TOV, E. 2010. A streaming parallel decision tree algorithm. *Journal of Machine Learning Research*, 11, 849–872.
- CUINGNET, R., GLAUNÈS, J. A., CHUPIN, M., BENALI, H., & COLLIOT, O. 2013. Spatial and anatomical regularization of SVM: A general framework for neuroimaging data. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(3), 682–696. <https://doi.org/10.1109/TPAMI.2012.142>.
- FENG, X., LING, X., ZHENG, H., CHEN, Z., & XU, Y. 2019. Adaptive multi-kernel SVM with spatial-temporal correlation for short-term traffic flow prediction. *IEEE Transactions on Intelligent Transportation Systems*, 20(6), 2001–2013. <https://doi.org/10.1109/TITS.2018.2854913>
- FENG, Y., ZHANG, P., YANG, M., LI, Q., & ZHANG, A. 2019. Short term load forecasting of offshore oil field microgrids based on DA-SVM. *Energy Procedia*, 158, 2448–2455. <https://doi.org/10.1016/j.egypro.2019.01.318>.
- GU, B., SHENG, V. S., TAY, K. Y., ROMANO, W., & LI, S. 2017. Cross Validation Through Two-Dimensional Solution Surface for Cost-Sensitive SVM. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(6), 1103–1121. <https://doi.org/10.1109/TPAMI.2016.2578326>.
- JIANG, H., CHING, W. K., YIU, K. F. C., & QIU, Y. 2018. Stationary Mahalanobis kernel SVM for credit risk evaluation. *Applied Soft Computing Journal*, 71, 407–417. <https://doi.org/10.1016/j.asoc.2018.07.005>.
- KRISHNAN, R. R., & K, T. G. 2018. Important Fund Client Classification Based On Proportion SVM. *2018 International Conference on Circuits and Systems in Digital Enterprise Technology (ICCSDET)*. <https://doi.org/10.1109/ICCSDET.2018.8821169>.
- LEE, J., & YOON, T. 2017. Analysis of Relation between Aging and Telomere using Datamining – Apriori, Decision Tree, and Support Vector Machine(SVM). *ICACT2017*, 685–689.
- LI, H., & ZHANG, Y. 2009. An algorithm of soft fault diagnosis for analog circuit based on the optimized SVM by GA. *2009 9th International Conference on Electronic Measurement & Instruments*. <https://doi.org/10.1109/ICEMI.2009.5274151>.
- MAT DERIS, A., MOHD ZAIN, A., & SALLEHUDDIN, R. 2011. Overview of support vector machine in modeling machining performances. *Procedia Engineering*, 24, 308–312. <https://doi.org/10.1016/j.proeng.2011.11.2647>.
- MUKHLIS, H., & KOENTJORO. 2015. Pelatihan Kebersyukuran untuk Menurunkan Kecemasan Menghadapi Ujian Nasional pada Siswa SMA. *Gadjah Mada Journal Of Professional Psychology*, 1(3), 203–215.
- PRIYAM, A., GUPTA, R., RATHEE, A., & SRIVASTAVA, S. 2013. *Comparative Analysis of Decision Tree Classification Algorithms*. 334–337.
- RAO, N. V. K., MANGATHAYARU, N., & RAO, M. S. 2017. Evolution and prediction of radical multi-dimensional e-learning system with cluster based data mining techniques. *2017 International Conference on Trends in Electronics and Informatics (ICEI)*. <https://doi.org/10.1109/ICOEI.2017.8300793>.
- RIYANTO, V., HAMID, A., & RIDWANSYAH. 2019. Prediction of Student Graduation Time Using the Best Algorithm. *Indonesian Journal of Artificial Intelligence and Data Mining*, 2(2), 1–9. <https://doi.org/http://dx.doi.org/10.24014/ijaidm.v2i1.6424>.
- SAHU, S. K., PUJARI, A. K., KAGITA, V. R., KUMAR, V., & PADMANABHAN, V. 2015. GP-SVM: Tree Structured Multiclass SVM with Greedy Partitioning. *International Conference on Information Technology (ICIT)*. <https://doi.org/10.1109/ICIT.2015.24>.
- SHARMA, G., BHARGAVA, R., & MATHURIA, M. 2013. Decision Tree Analysis on J48 Algorithm. *International Journal of Advanced Research InComputer Science*

- and Software Engineering*, 3(6), 1114–1119. Retrieved from https://www.academia.edu/4375403/Decision_Tree_Analysis_on_J48_Algorithm_for_Data_Mining.
- TOMAR, D., & AGARWAL, S. 2015. Twin Support Vector Machine: A review from 2007 to 2014. *Egyptian Informatics Journal*, 16(1), 55–69. <https://doi.org/10.1016/j.eij.2014.12.003>
- VASANI, V. P., & GAWALI, R. D. 2013. Classification Performance Evaluation. *Encyclopedia of Systems Biology*, 3(3), 411–411. https://doi.org/10.1007/978-1-4419-9863-7_100213.
- ZHANG, R. 2008. A preliminary study on prediction models for English web-based remedial education: Application of data mining theory. *WSEAS Transactions on Advances in ...*, (December). Retrieved from <http://www.wseas.us/e-library/transactions/education/2008/27-409.pdf>.
- ZOU, C., ZHENG, E., XU, H., & CHEN, L. 2010. SVM-Based Multiclass Cost-sensitive Classification with Reject Option for Fault Diagnosis of Steam Turbine Generator. *2010 Second International Conference on Machine Learning and Computing*. <https://doi.org/10.1109/ICMLC.2010.26>.

Halaman ini sengaja dikosongkan