

Klasifikasi Stres berdasarkan Unggahan pada Media Sosial Twitter menggunakan Metode *Support Vector Machine* dan Seleksi Fitur *Information Gain*

Jeowandha Ria Wiyani¹, Indriati², Sutrisno³

Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Brawijaya
Email: ¹jeonwiyani@student.ub.ac.id, ²indriati.tif@ub.ac.id, ³trisno@ub.ac.id

Abstrak

Stres dapat menimpa siapapun dan stress yang berkepanjangan dapat menyebabkan gangguan kesehatan mental. Namun, masih banyak orang yang enggan untuk pergi menemui tenaga ahli untuk membicarakan kondisi kesehatan mentalnya dan memilih untuk berkeluh kesah pada media sosial yang mana salah satunya adalah Twitter. Para pengguna menuliskan keluh kesahnya pada Twitter, sehingga tweet Twitter dapat digunakan untuk mengidentifikasi kondisi stres seseorang dengan klasifikasi teks. Klasifikasi teks tingkat stres menggunakan metode *Support Vector Machine* dan seleksi fitur *Information Gain*. Data yang digunakan pada penelitian ini sebanyak 87 dokumen dengan rincian 29 dokumen kelas 'Berat', 29 dokumen kelas 'Sedang', dan 29 dokumen kelas 'Ringan'. Metode *K-Fold Cross Validation* dengan k senilai 5 digunakan pada proses pengujian, sehingga 80% data digunakan sebagai data latih dan 20% sisanya sebagai data uji. Perbandingan hasil antara metode *Support Vector Machine* saja dengan kombinasi metode *Support Vector Machine* dan *Information Gain* menghasilkan akurasi paling baik pada metode *Support Vector Machine* saja dengan akurasi sebesar 59,11%, presisi sebesar 29,99%, recall sebesar 38,67%, dan *f-measure* sebesar 33,53%

Kata kunci: *stres, klasifikasi, twitter, support vector machine, seleksi fitur, information gain*

Abstract

*Stress can happen to anyone and prolonged stress can cause mental health problems. However, many people continue to be unwilling to consult with mental health professionals about their concerns instead opting to complain on social media, such as Twitter. Many people use Twitter to vent their frustrations, making it possible to utilize text classification to determine someone's stress level from their tweets. In this work, the Support Vector Machine technique with Information Gain feature selection is used for text categorization. The data used in this study were 87 documents with details of 29 'Heavy' class documents, 29 'Medium' class documents, and 29 'Light' class documents. With a k value of 5, the test was run using the K-Fold Cross Validation method, and the distribution of training and test data was 80:20. The comparison of the results between the Support Vector Machine method alone with the combination of the Support Vector Machine and Information Gain methods produces the best accuracy on the Support Vector Machine method alone with an accuracy of 59.11%, precision of 29.99%, recall of 38.67%, and *f-measure* of 33.53%.*

Keywords: *stress, classification, twitter, support vector machine, feature selection, information gain*

1. PENDAHULUAN

Stres dapat menimpa siapapun, mulai dari anak-anak hingga lanjut usia. Sumber stress sangat beragam, bisa dari sekolah, tempat kerja, lingkungan keluarga, dan lain sebagainya. Stres buruk yang berkepanjangan dapat menyebabkan timbulnya penyakit (Lumban Gaol, 2016). Dalam kondisi yang parah, seseorang yang

mengalami gangguan kesehatan mental dapat menyakiti dirinya sendiri atau menyakiti orang lain, sehingga, penting untuk memberikan perhatian dan penanganan yang tepat. Namun, kebanyakan dari orang masih malu atau enggan untuk membicarakan kondisi kesehatan mentalnya dengan cara berbicara langsung kepada tenaga ahli, sehingga cara ini dinilai kurang efektif untuk mengurangi stres (Xue, et

al., 2014). Sehingga, perlu mencari cara baru untuk mengetahui kondisi stres seseorang. Banyak orang menggunakan *micro-blog* untuk mendapatkan informasi, mengekspresikan diri, membagikan kegiatan sehari-hari dan emosi, serta berinteraksi secara personal. Kondisi seseorang dapat dilihat melalui apa yang ia unggah beserta interaksinya di *micro-blog* (Xue, et al., 2014). Twitter merupakan media sosial yang menggunakan sistem *microblogging*, yang digunakan untuk mengunggah berbagai jenis unggahan. Media sosial, termasuk Twitter, mengakomodasi ekspresi linguistik dan informasi pribadi seseorang yang informatif sehingga dapat digunakan untuk menganalisis karakteristik dan ciri kepribadian seseorang (Wang, et al., 2020). Tweet berupa teks yang sudah didapatkan perlu diklasifikasikan secara otomatis untuk mempermudah proses analisis kondisi stres seseorang.

Klasifikasi teks yang dilakukan oleh Kusumahadi et al. (2019) yaitu mengklasifikasikan tiket help desk menggunakan pendekatan Support Vector Machine dengan empat kernel dan metode TF-IDF (Term Frequency-Inverse Document Frequency) dengan hasil keakuratan metode SVM dengan kernel linier adalah 78%, kernel polinomial 78%, kernel Gaussian 81%, dan kernel sigmoid 51%. Maulana dkk. (2019) mengelaskan tingkat stres berdasarkan unggahan tweet menggunakan teknik Improved k-Nearest Neighbor dengan pemilihan fitur Chi-Square yang memberikan hasil presisi rata-rata 70%, *recall* rata-rata 67,2 persen, akurasi rata-rata 83,3 persen, dan rata-rata *f-measure* 66,3 persen. Dengan presisi 83 persen, *recall* 80 persen, dan akurasi 80 persen, pendekatan Support Vector Machine juga diterapkan dalam penelitian oleh (Mutawalli, et al., 2019).

Support Vector Machine dapat digunakan untuk menyelesaikan klasifikasi linear maupun non-linear. Masalah-masalah klasifikasi teks Sebagian besar merupakan non-linear separable problem, sehingga metode klasifikasi SVM cocok digunakan (Truşcă, 2019). Terdapat masalah lain pada klasifikasi teks yaitu besarnya dimensi dari data yang akan digunakan, yang mana nantinya akan menurunkan performa model, sehingga seleksi fitur penting dilakukan untuk mengurangi masalah ini. Rizaldy & Santoso (2017), menggunakan metode Information Gain untuk pemilihan fitur kategorisasi dokumen berita dengan bahasa Indonesia. Metode ini digabungkan dengan

metode klasifikasi Support Vector Machine dan didapatkan akurasi sebesar 98,057%. Menggunakan Information Gain sebagai metode pemilihan fitur dan metode Support Vector Machine, (Purnamasari, et al., 2018) mampu mengidentifikasi tweet cyberbullying di Twitter dengan hasil akurasi 76,66 persen, presisi 72,22 persen, *recall* 86,66 persen, dan *f-measure* 78,78 persen.

2. LANDASAN PUSTAKA

2.2 Stres

Menurut (Selye, 1950) stres merupakan respon tidak spesifik dari tubuh terhadap suatu tuntutan. Reaksi stres fisiologis adalah respon tubuh terhadap segala perubahan, ancaman, atau tekanan yang terjadi baik dari luar maupun dalam diri. Selye membedakan stres menjadi dua macam, yaitu stres baik (*eustress*) dan stres buruk (*distress*). Beberapa penyebab stres yang umum di antaranya yaitu ancaman fisik, ancaman terhadap citra diri, kejadian penting dalam hidup, pertengkaran atau konflik dengan teman atau saudara atau rekan kerja, tenggat waktu yang sangat ketat, dan kehilangan sesuatu atau seseorang yang disayang

2.3 Twitter

Twitter merupakan layanan *microblogging* yang pada tahun 2009, tahun ketiga peluncuran, telah digunakan oleh lebih dari 41 juta orang dan terus bertambah dengan cepat hingga sekarang (Kwak, et al., 2010). Praktik umum untuk merespon suatu tweet yaitu dengan RT atau *retweet*, menggunakan simbol '#' diikuti dengan kata-kata yang dinamakan hash tag, dan menyebut akun pengguna Twitter lain dengan simbol '@' yang disebut dengan *mention*. Mekanisme dari proses *retweet* ini dapat digunakan untuk menyebarkan informasi pilihan penggunaannya sehingga Twitter dapat digunakan untuk menyebarkan informasi secara luas.

2.4 Klasifikasi Teks

Menurut Sebastiani (dalam Maulana, et al., 2019) menyatakan bahwa klasifikasi teks adalah proses otomatis pengkategorian dokumen berdasarkan kategori yang telah ditentukan yang didasarkan pada isi dari dokumen. Terdapat tiga jenis klasifikasi teks yaitu klasifikasi teks berbasis statistik, klasifikasi teks berbasis koneksi, dan klasifikasi teks berbasis aturan.

2.5 Preprocessing

Preprocessing adalah tahapan yang dilakukan pada dataset teks untuk persiapan sebelum diproses lebih lanjut (Maulana, et al., 2019). Proses yang dilakukan pada proses ini yaitu case folding, pembersihan data, tokenisasi, penghapusan stopword, dan stemming (Luqyana, et al., 2018).

2.5.1 Case Folding

Case folding merupakan tahapan awal yang dilakukan pada preprocessing. Pada tahapan ini dilakukan pengubahan huruf-huruf pada setiap kata menjadi lowercase. Hal ini dilakukan untuk menghindari terjadinya redundansi data (Hidayat, et al., 2020).

2.5.2 Pembersihan Data

Pembersihan data adalah proses menghilangkan karakter non-abjad, seperti tanda baca dan angka, dari dataset (Luqyana, et al., 2018). Hal ini dilakukan karena angka dan tanda baca tidak dapat digunakan sebagai fitur.

2.5.3 Stopword Removal

Pada tahapan stopword removal dilakukan proses penghilangan stopword atau kata umum tanpa arti penting. Hal ini dilakukan agar program tidak menyimpan terlalu banyak kata (Luqyana, et al., 2018).

2.5.4 Stemming

Stemming adalah proses mengubah kata berimbuhan menjadi kata dasar. Stemming dilakukan berdasarkan dua aturan, yaitu pendekatan kamus dan pendekatan aturan (Luqyana, et al., 2018). Pada penerapannya, hanya satu aturan yang dipilih.

2.5.5 Tokenisasi

Pada tahapan ini dilakukan proses pemisahan dokumen menjadi kata atau token dengan menghilangkan whitespace (Luqyana, et al., 2018). Nantinya, kata-kata atau token ini akan digunakan sebagai fitur.

2.6 Information Gain

Menurut Deng & Runger (dalam Purnamasari, et al., 2018), Information Gain (IG) adalah teknik pemilihan fitur untuk memperkirakan signifikansi maksimum suatu fitur. Nilai Information Gain suatu fitur dapat merepresentasikan seberapa penting fitur

tersebut untuk mewakili suatu kelas. Perhitungan Information Gain dilakukan pada semua fitur data latih pada tiap kelas dan fitur yang nilainya kurang dari threshold atau batas akan dihapus (Pradhana, et al., 2020). Rumus untuk seleksi fitur Information Gain dinyatakan pada persamaan (2.1).

$$IG(t) = -\sum_{i=1}^m P(C_i) \log P(C_i) + P(t) \sum_{i=1}^m P(C_i|t) \log P(C_i|t) + P(\bar{t}) \sum_{i=1}^{|C|} P(C_i|\bar{t}) \log P(C_i|\bar{t}) \tag{1}$$

Keterangan:

$IG(t)$: nilai *Information Gain* dari *term t*.

m : banyaknya kelas.

$P(C_i)$: peluang dari dokumen dengan kelas C_i .

$P(t)$: peluang dari kehadiran *term t*.

$P(C_i|t)$: peluang dari kelas C_i yang memiliki *term t*.

$P(\bar{t})$: peluang dari ketidakhadiran *term t*.

$P(C_i|\bar{t})$: peluang dari kelas C_i yang tidak terdapat *term t* di dalamnya.

2.7 Support Vector Machine

Support Vector Machine merupakan cara untuk mencari *hyperplane* paling optimal yang digunakan untuk memisahkan dua kelas pada bidang masukan (Nugroho, et al., 2003). *Hyperplane* pembeda dua kelas terbaik didapatkan dengan menaksir batas *hyperplane* dan menemukan titik maksimalnya. Metode ini dapat juga digunakan untuk menyelesaikan masalah berdimensi tinggi dengan menggunakan trik *kernel*. *Kernel Polynomial*, *kernel Gaussian*, dan *kernel Sigmoid* adalah tiga *kernel* umum yang digunakan untuk menangani masalah non-linear. Persamaan untuk *kernel* tersebut dinyatakan pada tabel 2.1.

Tabel 2.1 Tabel Persamaan Kernel pada SVM

Kernel	Persamaan
Linear	$K(\vec{x}_i, \vec{x}_j) = x_i^T x_j$
Polynomial	$K(\vec{x}_i, \vec{x}_j) = (\vec{x}_i \vec{x}_j + 1)^p$
RBF	$K(\vec{x}_i, \vec{x}_j) = \exp(-\gamma x - x' ^2)$
Sigmoid	$K(\vec{x}_i, \vec{x}_j) = \tanh(\alpha \vec{x}_i, \vec{x}_j + \beta)$

Proses pelatihan menggunakan algoritme *sequential training* dengan tahapan:

1. Inisialisasi parameter-parameter yang diperlukan dalam perhitungan SVM seperti α_i , ϵ , γ , λ , dan C .

α_i = alfa, sebagai *support vector*

ϵ = epsilon, untuk mencari nilai *error*

γ = *gamma*, untuk mengatur seberapa jauh lengkungan sebagai *decision boundary* dan hanya dibutuhkan ketika menggunakan kernel RBF atau *Gaussian*

λ = *lambda*, variabel nilai skalar.

C = *hyperparameter* pada SVM yang digunakan untuk mengontrol *error*, semakin rendah C maka akan semakin rendah juga *error*

2. Menghitung nilai kernel data dengan kernel linear, *polynomial*, *Gaussian*, atau *Sigmoid*.
3. Menghitung matriks *Hessian* dengan persamaan 2.

$$D_{ij} = y_i y_j (K(x, x') + \lambda^2) \tag{2}$$

4. Mendapatkan nilai dari E_i menggunakan persamaan 3, $\delta\alpha_i$ menggunakan persamaan 4, dan juga α_i menggunakan persamaan 5 hingga iterasi maksimal.

$$E_i = \sum_{i=1}^N \alpha_i D_{ij} \tag{3}$$

$$\delta\alpha_i = \min\{\max[\gamma(1 - E_i), -\alpha_i], C - \alpha_i\} \tag{4}$$

$$\alpha_i = \alpha_i + \delta\alpha_i \tag{5}$$

5. Menghitung bobot dan bias dengan persamaan 6 dan 7.

$$\begin{aligned} w \cdot x^+ &= \alpha_i * y_i * \\ K(x_i, x^+) ; w \cdot x^- &= \alpha_i * y_i * \\ K(x_i, x^-) \end{aligned} \tag{6}$$

$$b = -\frac{1}{2}(w \cdot x^+ + w \cdot x^-) \tag{7}$$

3 PERANCANGAN

Tahapan yang dilakukan pada sistem dimulai dengan *preprocessing*, yang kemudian dilanjutkan dengan seleksi fitur dengan metode IG, pembobotan kata atau fitur hasil seleksi, dan yang terakhir yaitu proses klasifikasi

menggunakan *Support Vector Machine* dengan diagram alir proses yang ditunjukkan pada Gambar 1.



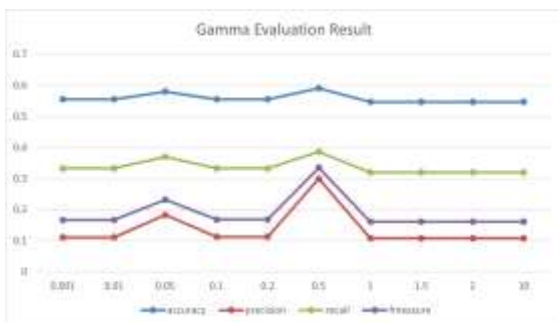
Gambar 1 Diagram Alir Proses

Langkah yang dilakukan pada tahap *preprocessing* yaitu *case folding*, *stopword removal*, *stemming*, dan tokenisasi. Pada penelitian ini tidak dilakukan proses *cleaning* pada *preprocessing* dikarenakan dataset yang tersedia sudah tidak mengandung karakter selain alfabet. Kemudian dilakukan seleksi fitur dengan metode IG, yang mana hanya kata atau fitur tertentu saja yang diambil berdasarkan nilai IG yang dimiliki fitur apakah memenuhi batas atau tidak. Setelah dilakukan seleksi fitur, kemudian dilakukan proses pembobotan kata menggunakan metode *tf-idf*. Setelah didapatkan bobot masing-masing kata hasil seleksi fitur, berikutnya dilakukan proses klasifikasi menggunakan *Support Vector Machine*.

4 HASIL DAN PEMBAHASAN

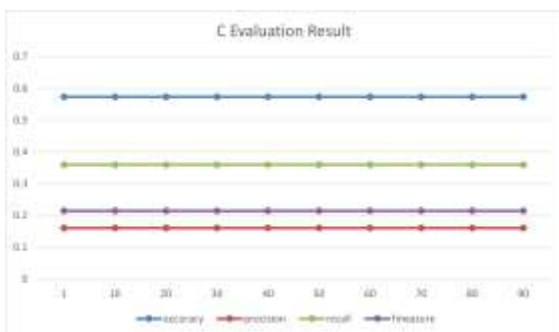
4.1 Pengujian Parameter *Support Vector Machine*

Pada pengujian ini dipilih tiga parameter yang akan diujikan, yaitu *gamma*, C dan *iterMax* dengan masing-masing parameter sebanyak 10 nilai. Nilai *default* dari parameter yaitu epsilon = 0.0001, gamma = 0.001, lambda = 0.5, C = 0.001, dan imax = 100 berdasarkan penelitian yang dilakukan oleh (Purnamasari, et al., 2018).



Gambar 2 Hasil Pengujian Nilai *Gamma*

Hasil dari pengujian parameter *gamma* ditunjukkan pada Gambar 2. Pada grafik tersebut, dapat dilihat bahwa akurasi, presisi, *recall*, dan *f-measure* tertinggi didapatkan pada nilai *gamma* 0.5 dengan akurasi sebesar 59,11%, presisi sebesar 29,99%, *recall* sebesar 38,67%, dan *f-measure* sebesar 33,53%. Parameter *gamma* menentukan seberapa jauh pengaruh antar titik data terhadap satu sama lain, dengan nilai *gamma* yang kecil berarti ‘jauh’ dan nilai *gamma* yang besar berarti ‘dekat’. Jika nilai *gamma* terlalu besar, radius area yang dipengaruhi oleh *support vector* menjadi terlalu sempit dan hanya *support vector* saja yang dianggap ‘dekat’. Sebaliknya, jika nilai *gamma* terlalu kecil, maka area yang dipengaruhi oleh *support vector* akan terlalu luas. Pada *delta alpha*, nilai *gamma* akan mempengaruhi nilai *alpha* yang akan menjadi *support vector*.



Gambar 3 Hasil Pengujian Nilai *C*

Gambar 3 menunjukkan hasil dari pengujian parameter *C*, yang mana menunjukkan bahwa didapatkan nilai akurasi, presisi, *recall* dan *f-measure* yang sama yaitu akurasi sebesar 59,11%, presisi sebesar 29,99%, *recall* sebesar 38,67%, dan *f-measure* sebesar 33,53%. Dengan demikian, ditunjukkan bahwa nilai parameter *C* tidak berpengaruh pada pengujian ini, sehingga dapat disimpulkan bahwa telah didapatkan nilai *C* yang optimal pada nilai 1.

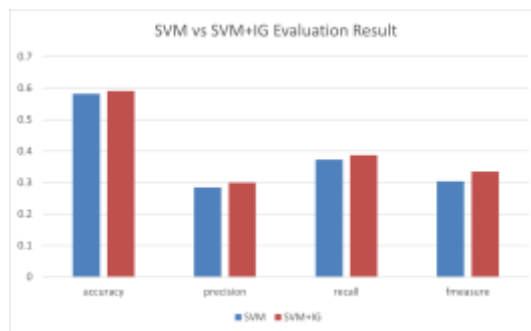


Gambar 4 Hasil Pengujian Nilai *iterMax*

Hasil pada pengujian nilai *iterMax*, yang ditunjukkan pada Gambar 4, nilai evaluasi terbaik didapatkan pada nilai *iterMax* 10 akurasi sebesar 59,11%, presisi sebesar 29,99%, *recall* sebesar 38,67%, dan *f-measure* sebesar 33,53%, yang kemudian mengalami penurunan pada nilai *iterMax* 15 dan kembali menaik pada nilai *iterMax* 20. Dan ini berlangsung pada nilai-nilai *iterMax* berikutnya. Hal ini dikarenakan nilai *iterMax* akan mempengaruhi nilai *alpha* yang didapatkan pada proses *sequential training*.

4.2 Pengujian Pengaruh Seleksi Fitur *Information Gain*

Pada pengujian ini dilakukan perbandingan kinerja sistem dari metode *Support Vector Machine* saja dengan gabungan metode *Support Vector Machine* dengan seleksi fitur *Information Gain*. Parameter yang dipergunakan pada pengujian ini diperoleh dari pengujian sebelumnya, yaitu nilai *gamma* = 0.5, nilai *C* = 1, nilai *lambda* = 0.5, nilai *epsilon* = 0.0001, dan nilai *iterMax* = 10. Untuk batas yang digunakan pada seleksi fitur *Information Gain* yaitu 90% dari *term* yang memiliki nilai *Information Gain* tertinggi yang didasarkan pada penelitian yang dilakukan oleh (Putri, et al., 2020). Hasil pengujian ditunjukkan pada Gambar 4.



Gambar 5 Hasil Pengujian SVM dengan SVM+IG

Gambar 4 menunjukkan bahwa didapatkan hasil bahwa nilai akurasi pada gabungan

metode *Support Vector Machine* dan *Information Gain* mendapat nilai lebih tinggi dibandingkan dengan hanya menggunakan metode *Support Vector Machine* saja dengan akurasi sebesar 59,11%, presisi sebesar 29,99%, *recall* sebesar 38,67%, dan *f-measure* sebesar 33,53%. Namun, hasil evaluasi antara metode *Support Vector Machine* saja dengan kombinasi metode *Support Vector Machine* dan *Information Gain* tidak begitu berbeda jauh. Hal ini dikarenakan pada saat proses seleksi fitur, fitur-fitur yang kurang mewakili dokumen-dokumen pada setiap kelas tidak lolos seleksi, sehingga hasil klasifikasi menjadi lebih baik. Namun, apabila *threshold* dari *Information Gain* diperkecil, dapat memperburuk keluaran klasifikasi dikarenakan fitur-fitur yang mewakili dokumen pada setiap kelas tidak akan lolos seleksi.

5 KESIMPULAN

Kesimpulan yang didapatkan berdasarkan hasil penelitian, pengujian, dan analisis yang telah dilakukan adalah nilai parameter terbaik yang didapatkan yaitu nilai *gamma* 0.5 dengan akurasi sebesar 59,11%, presisi sebesar 29,99%, *recall* sebesar 38,67%, dan *f-measure* sebesar 33,53%, nilai *C* 1 dengan akurasi sebesar 59,11%, presisi sebesar 29,99%, *recall* sebesar 38,67%, dan *f-measure* sebesar 33,53%, dan nilai *iterMax* 10 dengan akurasi sebesar 59,11%, presisi sebesar 29,99%, *recall* sebesar 38,67%, dan *f-measure* sebesar 33,53%. Untuk hasil pengujian pengaruh seleksi fitur *Information Gain* pada hasil klasifikasi menggunakan metode *Support Vector Machine* didapatkan bahwa nilai akurasi pada gabungan metode *Support Vector Machine* dan *Information Gain* mendapat nilai lebih tinggi dibandingkan dengan hanya menggunakan metode *Support Vector Machine* saja dengan akurasi sebesar 59,11%, presisi sebesar 29,99%, *recall* sebesar 38,67%, dan *f-measure* sebesar 33,53%. Namun, hasil evaluasi antara metode *Support Vector Machine* saja dengan kombinasi metode *Support Vector Machine* dan *Information Gain* tidak begitu berbeda jauh

Saran yang dapat diberikan yaitu diharapkan untuk menggunakan *kernel SVM* lain seperti polinomial, sigmoid atau linear. Selain itu, dapat dipertimbangkan untuk menggunakan metode lain untuk mencari kombinasi nilai parameter SVM yang optimal, serta menambah banyaknya data yang digunakan

untuk penelitian agar hasil klasifikasi menjadi lebih optimal.

6 DAFTAR PUSTAKA

- Bangsheng, S., 2013. Information Gain Feature Selection Based on Feature Interactions. s.l.:s.n.
- Deng, X., Liu, Q., Deng, Y. & Mahadevan, S., 2016. An improved method to construct basic probability assignment based on the confusion matrix for classification problem. *Information Science*, Volume 340-341, pp. 250-261.
- Grootendorst, M., 2019. Towards Data Science. [Online] Available at: <https://towardsdatascience.com/validating-your-machine-learning-model-25b4c8643fb7> [Accessed October 2021].
- Hidayat, A. R., Adikara, P. P. & Adinugroho, S., 2020. Klasifikasi Hoaks Kesehatan di Media Sosial menggunakan. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, Volume 4, pp. 1702-1708.
- Kusumahadi, S. H., Junaedi, H. & Santoso, J., 2019. Klasifikasi Helpdesk Menggunakan Metode Support Vector Machine. *Jurnal Informatika: Jurnal Pengembangan IT (JPIT)*, Volume 04, pp. 54-60.
- Kwak, H., Lee, C., Park, H. & Moon, S., 2010. What is Twitter, a Social Network or a News Media?. Raleigh, North Carolina, USA., International World Wide Web Conference Com-mittee (IW3C2)..
- Lumban Gaol, N. T., 2016. Teori Stres: Stimulus, Respons, dan Transaksional. *Buletin Psikologi*, Volume 24, pp. 1-11.
- Luqyana, W. A., Cholissodin, I. & Setya, P. R., 2018. Analisis Sentimen Cyberbullying pada Komentar Instagram dengan Metode Klasifikasi Support Vector Machine. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, Volume 2, pp. 4704-4713.
- Maulana, M. I., Indriati & Soebroto, A. A., 2019. Klasifikasi Tingkat Stres Berdasarkan Tweet pada Akun Twitter menggunakan Metode Improved k-Nearest Neighbor dan Seleksi Fitur Chi-square. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, Volume 3, pp. 6662-

- 6669.
- Mutawalli, L., Zaen, M. T. A. & Bagye, W., 2019. Klasifikasi Teks Sosial Media Twitter Menggunakan Support Vector Machine (Studi Kasus Penusukan Wiranto). *Jurnal Informatika & Rekayasa Elektronika*, 2(2), pp. 43-51.
- Nugroho, A. S., Witarto, A. B. & Handoko, D., 2003. *Support Vector Machine - Teori dan Aplikasinya dalam Bioinformatika*. s.l.:s.n.
- Pradhana, M. O., Indriati & Adinugroho, S., 2020. Analisis Sentimen Masyarakat terhadap Uji Coba LRT Jakarta Menggunakan Improved K-Nearest Neighbor dan Information Gain. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, Volume 4, pp. 1888-1896.
- Purnamasari, N. M. G. D., Fauzi, M. A., Indriati & Dewi, L. S., 2018. Identifikasi Tweet Cyberbullying pada Aplikasi Twitter Menggunakan Metode Support Vector Machine (SVM) dan Information Gain (IG) sebagai Seleksi Fitur. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 2(11), pp. 5326-5332.
- Putri, F. O., Indriati & Wihandika, R. C., 2020. Analisis Sentimen pada Ulasan Pengguna MRT Jakarta Menggunakan Metode Neighbor-Weighted K-Nearest Neighbordengan Seleksi Fitur Information Gain. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, Volume 4, pp. 2195-2203.
- Rizaldy, A. & Santoso, H. A., 2017. *Performance Improvement Of Support Vector Machine (SVM) With Information Gain On Categorization Of Indonesian News Documents*. Semarang, Indonesia, s.n.
- Selye, H., 1950. *Stress*. Montreal : Acta, Volume 1955.
- Thapliyal, H., Khalus, V. & Labrado, C., 2017. *Stress Detection and Management : A survey of wearable smart health devices.. IEEE Consumer Electronics Magazine* , pp. 64-69.
- Truşcă, M. M., 2019. Efficiency of SVM classifier with Word2Vec and Doc2Vec models. Bucharest, s.n.
- Wang, X., Zhang, H., Cao, L. & Feng, L., 2020. *Leverage Social Media for Personalized Stress Detection*. Seattle, WA, USA., ACM, pp. 2710-2718.
- Xue, Y. et al., 2014. *Detecting Adolescent Psychological Pressures from Micro-Blog*. Shenzhen, Springer, Cham, pp. 83-94.