

PERBANDINGAN PERFORMANSI DENGAN METODE CORRELATION BASED FEATURE SELECTION PADA LVQ 2

SURYA ADITYA GD¹, IIS AFRIANTY², SUWANTO SANJAYA³, RAHMAD
ABDILLAH⁴, LESTARI HANDAYANI⁵, FITRI INSANI⁶

1,2,3,4,5,6 Program Studi Teknik Informatika, Fakultas Sains dan Teknologi
Universitas Islam Negeri Sultan Syarif Kasim Riau

Jl.H.R Soebrantas No.155 KM.15 Simpang Baru Panam Pekanbaru, 28293

Email: ¹11950115225@students.uin-suska.ac.id, ²iis.afrianty@uin-suska.ac.id,

³suwantosanjaya@uin-suska.ac.id, ⁴rahmad.abdillah@uin-suska.ac.id,

⁵lestari.handayani@uin-suska.ac.id, ⁶fitri.insani@uin-suska.ac.id

ABSTRAK

Melakukan sebuah penelitian diperlukannya mengidentifikasi sebuah data yang sesuai dengan melakukan sebuah klasifikasi. Pengaruh dalam mendapatkan hasil akurasi yang maksimal dengan menentukan teknik penelitian secara tepat melalui proses klasifikasi. Pada penelitian ini melakukan perbandingan peningkatan performansi akurasi akurasi LVQ 2 dengan mengimplementasikan *Correlation Based Feature Selection* (CFS) pada *dataset* bertujuan keakuratan pengambilan data sampel dengan metode klasifikasi. Data parameter tulang tengkorak yang digunakan yaitu data pria dan wanita dengan jumlah data 2524 dan fitur 82. Penelitian LVQ 2 tanpa CFS dengan nilai *learning rate* (α) = 0.9 dan *window* 0.2 yang akurasi tertingginya memperoleh sebesar 77.05%, dan menggunakan CFS pada nilai α = 0.9 dan *window* = 0.3 hasil akurasi tertinggi yaitu 82,51%. Hal ini menunjukkan bahwa LVQ 2 menggunakan CFS sangat direkomendasikan baik dari segi performansi terhadap pada *dataset* Tengkorak dibandingkan LVQ 2 tanpa menggunakan CFS.

Kata Kunci : Akurasi, Perbandingan, LVQ 2, CFS, Tengkorak

I. PENDAHULUAN

Melakukan sebuah penelitian diperlukannya mengidentifikasi sebuah data yang sesuai dengan melakukan sebuah klasifikasi. teknik tersebut dilakukan dalam pengelompokkan suatu obyek yang terdapat dalam kategori telah ditentukan(Bishop, 2006). Teknik klasifikasi umumnya dilakukan dari sekumpulan data training dengan kelas yang sudah ditetapkan(Wibawa et al., 2018). Penelitian klasifikasi ini kerap sekali dilakukan. Sepeti pada halnya mengenai pada bidang forensik seperti antropologi yang bermanfaat dalam

pengulangan kembali dan ilustrasi biologis terhadap individu yang tidak diketahui asal usulnya. Kabar didapatkan dari hasil identifikasi digunakan pada kejadian peristiwa kematian atau kasus kejahatan. Menentukan jenis kelamin adalah salah satu cara mengidentifikasi yang sering dilakukan.

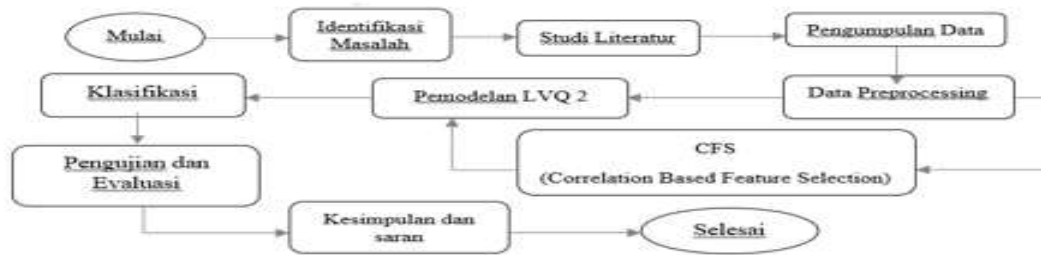
Cara menentukan hal tersebut memerlukan suatu data yang akan dianalisis metriknya beserta pengukuran morfologi pada tulang (Afrianty et al., 2022). Dalam menentukan jenis kelamin dan kerangka tulang yang sangat direkomendasikan setelah tulang pelvis yaitu tulang tengkorak (Toneva et al., 2021). Dataset tengkorak yang digunakan sebanyak 82 fitur sebagai perhitungan metriknya untuk penentuan jenis kelamin dari penelitian ini. Pada Penelitian LVQ 2 yang pernah diterapkan klasifikasi gangguan kehamilan trimester pertama mendapatkan akurasi sebesar 100% (Budianita et al., 2018).

Adanya penelitian yang sudah di klaim untuk meningkatkan hasil akurasi klasifikasi dengan *Correlation-Based Feature Selection* yang prinsip kerjanya dari algoritma ini yaitu menemukannya subset fitur tanpa mempengaruhi nilai dimensi sebuah dataset (Made et al., 2019). Penerapan teknik *Feature Selection* pada penelitian dengan metode *Correlation-Based Feature selection (CFS)* dalam pengambilan data yang digunakan berdasarkan nilai bobot yang akan dirangking dari yang terendah ke tertinggi (Pratama et al., 2022). Mengenai CFS ada juga yang digunakan pada *Support Vector Machine (SVM)* untuk implementasinya dengan Menggunakan Metode CFS untuk mengklasifikasikan hotel review analisis sentimen karena memiliki kelebihan dalam mengolah dataset yang besar (Ririanti & Purwinarko, 2021).

Penelitian ini melakukan pada *dataset* William White Howells Craniometric mengenai perbandingan performansi peningkatan akurasi menggunakan algoritma LVQ2. *Dataset* tempurung kepala manusia (tengkorak) merupakan kumpulan data pengukuran yang telah diperkirakan oleh Dr. William W Howells yang dikumpulkan pada tahun 1965 sampai 1980 (Auerbach, n.d.). Jumlah *Dataset* dengan 2524 data dan 82 fitur yang hanya mendapatkan akurasi tertinggi sebesar 77.05% pada *learning rate* (α) 0.9 dan *window* 0.2 (Darmila et al.,

2022). Peran penelitian ini yaitu evaluasi performansi algoritma LVQ 2 dengan metode *Correlation-Based Feature Selection (CFS)* maupun tidak menggunakannya pada *dataset* dengan menggunakan sampel dalam pengumpulan data dan membandingkan hasil akurasi yang masih bisa mendapatkan hasil lebih baik berdasarkan kriteria pengujian yang dipakai.

II.METODE PENELITIAN



Gambar 1. Tahap Alur Penelitian

Pada Gambar 1, menunjukkan alur penelitian yang diuraikan sebagai berikut:

1. Melakukan Pengumpulan Data

Data tengkorak yang dipakai pada penelitian ini berjumlah 1368 data pria dan 1156 data wanita. Ada 82 fitur pengukuran tengkorak manusia. Pada dibawah ini menjelaskan 10 fitur dari 82 fitur dengan parameter yang diberikan kode tiap fitur pada tulang tengkorak di Tabel 1 dan Tabel 2.

Tabel 1. 10 Fitur Parameter Tengkorak

Kode	Fitur Parameter Tulang Tengkorak
AUB	Biauricular breadth
OCC	Occipital Chord
PAS	Parietal subtense
PAC	Parietal Chord
FRS	Frontal subtense
NPH	Nasion - prosthion height
PAF	Parietal fraction
FRF	Frontal fraction
MAB	External palate breadth

Tabel 2. Parameter Tulang Tengkorak

CLASS	GOL	BNL	BPL	ZYB	FOL	NPH	NLH	NLB	ZMB	MAB	...	AUB
P	193	104	98	131	35	70	54	25	86	59	...	122
P	192	102	92	137	32	71	54	24	89	59	...	130
P	189	106	96	139	31	72	55	23	85	58	...	134
W	182	97	103	122	35	68	48	26	92	60	...	117
W	170	91	90	124	33	65	48	21	79	58	...	113
W	153	88	92	115	32	53	41	21	89	56	...	103

2. Data Preprocessing

Pada tahapan ini terlebih dahulu melakukan *cleaning* data, perubahan data atau transformasi data, normalisasi data, dan distribusi data.

A. Melakukan *Cleaning* Data

Pada *dataset* ini membersihkan beberapa fitur seperti *popnum* dan *population*. Fitur ini dibersihkan dikarenakan memberikan keterangan nama populasi dan populasi saja.

B. Melakukan Transformasi Data

Tahap ini melakukan perubahan pada kelas jenis kelamin yaitu Pria menjadi 1 dan Wanita menjadi 2.

C. Melakukan Normalisasi Data

Tahap ini proses pengukuran nilai fitur dari suatu data tertentu dalam nilai rentang tertentu, metode yang sederhana dilakukan terhadap data asli metode min-max hanya mengukur jarak fitur 0, dan 1 (Darmila et al., 2022).

Normalisasi dilakukan dengan menggunakan Persamaan 1.

$$X_{new} = \left(\frac{X_{old} - X_{min}}{X_{max} - X_{min}} (new_{max} - new_{min}) + new_{min} \right) \quad (1)$$

D. Distribusi Data

Tahap ini akan dikelompokkan menjadi 9 sampel dari tiap kelas. Pada Tabel dibawah ini yaitu melakukan distribusi data yang digunakan saat pengujian, melakukan sampel data setelah menggunakan (*CFS*) dan tidak menggunakan (*CFS*) dengan metode pencarian fitur *Best First Search* menggunakan

Direction Forward dan *Direction Backward Best First Search* yaitu suatu pencarian *heuristic* yang mengarahkan eksplorasi pada *node* dengan nilai pertama terbaiknya (Liana & Nudin, 2020) pada dataset untuk diseleksi beberapa fitur. Pada dataset ini dengan kedua kelas yaitu yaitu tulang tengkorak pria dan tulang tengkorak wanita pada tiap sampel (SI – S IX).

Tabel 3. *Dataset* sampel sebelum dan setelah CFS direction Forward dan Backward

Forward dan Backward									
Kelas	S I	S II	S III	S IV	S V	S VI	S VII	S VIII	S IX
Pria	1231	137	1094	274	958	410	821	547	684
Wanita	116	1040	231	925	347	809	462	694	578

3. CFS (*Correlation Based Feature Selection*)

Tahap ini data yang diperlukan seleksi fitur dengan metode *Correlation Based Feature Selection* (CFS) yang sudah ada pada machine learning dengan sisa fitur dipakai 78 Fitur.

4. Pemodelan Menggunakan LVQ 2

Data yang diperlukan model implementasi yang sesuai dengan teknik LVQ 2.

5. Klasifikasi dan *Testing*

Klasifikasi ini menggunakan *dataset* yang telah diketahui jenis kelasnya dengan melakukan proses *training* dan *testing*. Selanjutnya dilakukan pengujian performa dari LVQ 2.

6. Evaluasi Performansi Hasil Pengujian

Tahap ini yaitu bertujuan evaluasi terhadap performanya menggunakan LVQ2 dengan melakukan CFS dan tanpa melakukan CFS, Metode CFS ini digunakan matriks korelasi yang mempunyai kemampuan mendapatkan nilai berhubungan fitur tertinggi dari kelas tersebut, akan tetapi fitur tersebut juga tidak berhubungan dengan beberapa fitur(Nurul Yusufiyah & Gya Nur Rochman, 2021).

2.1 CFS

Pada Metode ini supaya mendapatkan nilai korelasinya yang tertinggi pada kelas tersebut, akan tetapi beberapa fitur ada juga tidak berkorelasi dengan fitur-fitur lainnya (Ririanti & Purwinarko, 2021). Algoritma ini dapat terdapat pada Persamaan (2).

$$Ms = \frac{nrcf}{\sqrt{n + n(n - 1) rff}} \quad (2)$$

Dimana ,

- Ms = "Merits" yang merupakan banyaknya fitur
- n = total fitur.
- rcf = Korelasi kelas fitur
- rff = interkorelasi fitur

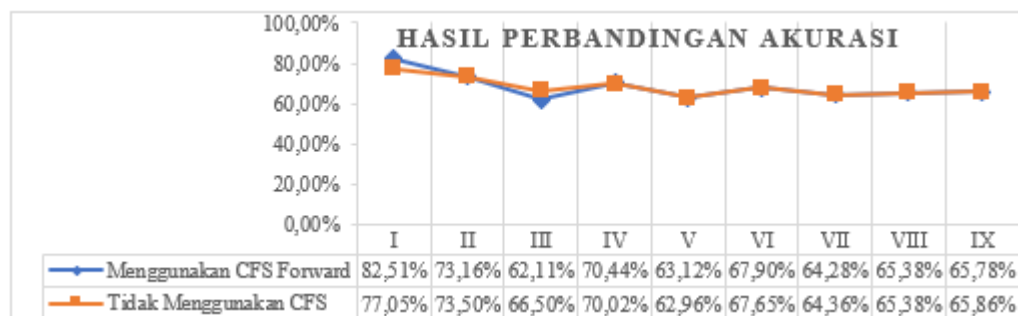
2.2 LVQ 2

Tahap ini dengan cara apabila kedua vector pemenang maka secara bersamaan diperbaharui asalkan kedua vektor tersebut memiliki range yang mirip. Lakukan pengurangan α (Budianita et al., 2018).

III.HASIL DAN PEMBAHASAN

Dataset tengkorak akan dibuat dalam 9 sampel yang dibagi menjadi data *training* dan *testing*. Data ini dilakukan pengujiannya dengan kriteria nilai pada *learning rate* yaitu 0.00001, 0.0001, 0.001, 0.01, 0.1, 0.4, 0.7, dan 0.9, dengan minimum α 0.01, dan pengurang α sebesar 0.01, nilai *window* 0, 0.3, 0.4, 0.1, 0.5, 0.2 dengan LVQ 2 menggunakan *k-fold cross validation*. Adapun nilai *epoch* yang digunakan yaitu 1000.

3.1 Hasil Akurasi Tertinggi Tidak Menggunakan CFS dan CFS BFS Forward LVQ 2

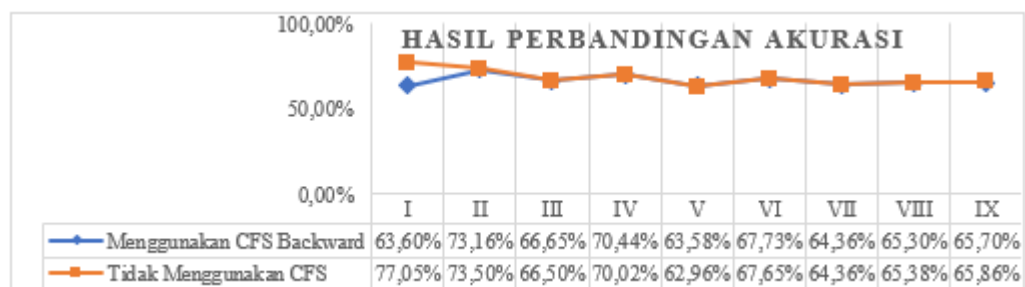


Gambar 2. Hasil perbandingan kedua akurasi yang tertinggi pada Sampel I – IX tidak menggunakan CFS (Darmila et al., 2022) dan menggunakan CFS metode pencarian BFS Forward

Gambar 2 membuktikan dari ke-9 sampel hasil akurasi yang tertinggi

dilakukan pengujian dengan LVQ 2. Hasil akurasi yang terbaik penelitian (Darmila et al., 2022) diperoleh pada sampel I nilai $\alpha = 0.9$ dan $window = 0.1$ sebesar 77,05%, Sedangkan menggunakan CFS mendapatkan hasil akurasi terbaik yaitu 82,51% pada nilai $\alpha = 0.9$ dan $window 0.3$. Pada sampel II pada penelitian (Darmila et al., 2022) atau yang tidak menggunakan CFS hasil akurasi tertinggi diperoleh dengan nilai $\alpha = 0.00001$ dan $window = 0.1$. Menggunakan CFS nilai $\alpha = 0.0001$ dengan $window 0.1$. Sampel III pada penelitian (Darmila et al., 2022) memperoleh akurasi tertinggi dengan nilai $\alpha = 0.4$ dan $window = 0.1$, sedangkan yang menggunakan CFS akurasi yang didapatkan dengan nilai α dan $window$ yang sama dengan sampel I. Sampel IV,V,VI,VII,VIII penelitian (Darmila et al., 2022) akurasi terbaiknya diperoleh seluruh dengan nilai α dan nilai $window$ yaitu 0.0. Pada sampel IX, akurasi tertingginya berada nilai $\alpha = 0.00001$ dan nilai $window$ yaitu 0.1 , 0.2 , 0.3 , 0.4 , hingga 0.5, sedangkan menggunakan CFS akurasi tertinggi diperoleh semua nilai α dan $window 0.0$ hanya pada sampel IV,VI,VII, dan VIII. Pada sampel V dan IX saja yang memperoleh akurasi terbaiknya di semua nilai α dan nilai $window 0.0$

3.2 Hasil Akurasi Tertinggi Tidak Menggunakan CFS dan CFS BFS Backward LVQ 2



Gambar 3. Hasil perbandingan kedua akurasi yang tertinggi pada sampel I – IX tidak menggunakan CFS dan menggunakan CFS metode pencarian BFS Backward

Gambar 3 membuktikan hasil perbandingan akurasi tertinggi dari 9 sampel dilakukan pengujian dengan LVQ 2 tanpa CFS dan menggunakan CFS. Hasil

akurasi penelitian (Darmila et al., 2022) yang terbaik diperoleh pada sampel I dengan $\alpha = 0.9$ dan $window = 0.1$ sebesar 77,05%, Sedangkan yang menggunakan CFS mendapatkan hasil akurasi terbaik dibawah dari hasil pada sampel I penelitian yang (Darmila et al., 2022) dengan nilai $\alpha = 0.0001$ dan $window = 0.1$ yaitu 63,60%. Pada sampel II penelitian (Darmila et al., 2022) akurasi tertinggi didapat pada nilai $\alpha = 0.00001$ dan $window = 0.1$. Menggunakan CFS memperoleh akurasi tertinggi dari sampel II dengan nilai $\alpha = 0.0001$ dengan $window$ yang 0.2, 0.3, 0.4, dan 0.5. Sampel III pada penelitian (Darmila et al., 2022) akurasi tertingginya diperoleh pada nilai α dan $window$ yang sama pada sampel I dan II, dan yang menggunakan CFS akurasi tertinggi yang didapatkan dengan nilai $\alpha = 0.9$ dan $window = 0.2$. Akurasi terbaik pada penggunaan seluruh nilai $\alpha = 0.0$ $window = 0.0$ terdapat di sampel IV, V, VI, VII, VIII penelitian (Darmila et al., 2022). Sementara sampel IX, akurasi tertingginya diperoleh dengan nilai $\alpha = 0.00001$ dan nilai $window = 0.1, 0.2, 0.3, 0.4,$ dan 0.5 (Darmila et al., 2022), Menggunakan CFS yang memperoleh akurasi cukup baik seluruh nilai α dan nilai $window 0.0$ hanya pada sampel IV, VI, VII, dan VIII. Pada sampel V dan IX saja yang memperoleh akurasi tertinggi seluruh nilai α dan nilai $window$ yaitu 0.2, 0.3, 0.4, 0.5.

3.3 Hasil Selisih Akurasi Tertinggi Perbandingan LVQ 2 dengan Data Sebelum dan Sesudah CFS BFS Forward dan Backward LVQ 2

Tabel 4. Hasil Selisih Akurasi Perbandingan Pada Penelitian sebelumnya dengan Menggunakan CFS pada pengujian Learning Rate $\alpha 0.1$ dan $window 0.1$

Pengujian	Sampel								
	I	II	III	IV	V	VI	VII	VIII	IX
LVQ2 - LVQ2 CFS Forward	+5,13%	+10,25%	+12,24%	-5,79%	-14,31%	+4,10%	-5,32%	-0,40%	+2,59%
LVQ2 - LVQ2 CFS Backward	-2,90%	-7,57%	+1,83%	-2,36%	+1,06%	+5,74%	-0,00%	+0,01%	+0,05%

Pada Tabel 4, Menjelaskan pada pengujian LVQ 2 (Darmila et al., 2022) dengan menggunakan CFS BFS forward pada setiap sampel dari I-IX Menunjukkan pada keterangan warna kuning menunjukkan bahwa yang menggunakan CFS BFS

forward lebih tinggi dari LVQ 2 (Darmila et al., 2022) , Sedangkan Tabel yang berwarna putih menunjukkan CFS BFS *forward* hasil selisih akurasi yang rendah daripada LVQ 2 (Darmila et al., 2022) dari ke-9 sampel ini selisih akurasi terbaik yang lebih banyak pada tabel warna kuning yaitu menggunakan CFS BFS *forward* yaitu jumlah selisih tertinggi yaitu pada sampel III sebesar 12,24%, sama halnya dengan CFS BFS *backward* lebih banyak mendapatkan selisih akurasi tertinggi dari pada LVQ 2 (Darmila et al., 2022) dengan jumlah selisih akurasi tertinggi pada sampel VI sebesar 5,74%.

IV.KESIMPULAN

Pada kesimpulan ini menjelaskan bahwa penerapan pelatihan dan pengujian CFS pada LVQ 2 sangat berpengaruh untuk tingkat akurasi pada klasifikasi. Mekanisme LVQ 2 dilakukan metode CFS pengambilan data terbaik menggunakan sampel I yang jumlah data *training* dan data *testing* 1231 dan 116. Metode CFS adalah metode yang paling tepat digunakan ketika masih keadaan dataset yang utuh setelah tahap preprocessing agar menemukan perbandingan ketika melakukan pengujian dengan LVQ 2 agar dataset tersebut dapat menemukan masing masing korelasi antar fiturnya sehingga mendapatkan pengujian lebih akurat ketika dilakukan pengujian pengambilan data dengan LVQ 2. Dengan demikian, implementasi evaluasi LVQ 2 terhadap CFS sangat direkomendasikan baik dari segi performansi pada *dataset* Tengkorak dibandingkan LVQ 2 tanpa menggunakan CFS. Untuk penelitian berikutnya dapat menggunakan LVQ2 dikombinasikan dengan metode *Feature Selection Gain Ratio*.

DAFTAR PUSTAKA

- Afrianty, I., Nasien, D., & Haron, H. (2022). Performance Analysis of Support Vector Machine in Sex Classification of The Sacrum Bone in Forensic Anthropology. *JURNAL TEKNIK INFORMATIKA*, 15(1), 63–72. <https://doi.org/10.15408/jti.v15i1.25254>
- Budianita, E., Sanjaya, S., Syafria, F., Teknik Informatika, J., Sains dan Teknologi, F., Sultan Syarif Kasim Riau Jl Soebrantas No, U. H., & Baru, S. (2018).

Penerapan Metode Learning Vector Quantization² (LVQ 2) Untuk Menentukan Gangguan Kehamilan Trimester I. *Jurnal Sains, Teknologi Dan Industri*, 15(2), 144–151.

- Darmila, Afrianty, I., Sanjaya, S., Abdillah, R., Iskandar, I., & Syafria, F. (2022). Evaluasi Perbandingan Performansi LVQ 1, LVQ 2, Dan LVQ 3 Dalam Klasifikasi Jenis Kelamin Menggunakan Tulang Tengkorak. *Jurnal INSTEK (Informatika Sains Dan Teknologi)*, 7(2), 344–353.
- Liana, L. I., & Nudin, S. R. (2020). Implementasi Algoritma Best-First Search untuk Aplikasi Mesin Pencari Handphone pada E-commerce (Apenphone). *Journal of Informatics and Computer Science (JINACS)*, 2(1).
- Made, I., Adnyana, B., Jln, S. B., & Puputan, R. (2019). Penerapan Feature Selection untuk Prediksi Lama Studi Mahasiswa. *JURNAL SISTEM DAN INFORMATIKA*, 13(2).
- Nurul Yusufiyah, H. K., & Gya Nur Rochman, J. P. (2021). Efektivitas Penggunaan Seleksi Ciri CFS pada Klasifikasi Ciri Bentuk Nodul Kanker Payudara dengan Citra Ultrasonografi. *Physics Education Research Journal*, 3(1), 11–18. <https://doi.org/10.21580/perj.2021.3.1.6667>
- Pratama, A. B., Budianita, E., Yanti, N., Mai Candra, R., & Informatika UIN Sultan Syarif Kasim Riau Jl Subrantas Km, T. (2022). Implementasi Metode Learning Vector Quantization (LVQ) Untuk Sentimen Analisis Terhadap Aplikasi Go-Jek Pada Playstore. *Jurnal Nasional Komputasi Dan Teknologi Informasi*, 5(3).
- Ririanti, N. P., & Purwinarko, A. (2021). Implementation of Support Vector Machine Algorithm with Correlation-Based Feature Selection and Term Frequency Inverse Document Frequency for Sentiment Analysis Review Hotel. *Scientific Journal of Informatics*, 8(2), 297–303. <https://doi.org/10.15294/sji.v8i2.29992>
- Toneva, D., Nikolova, S., Agre, G., Zlatareva, D., Hadjidekov, V., & Lazarov, N. (2021). Machine learning approaches for sex estimation using cranial measurements. *International Journal of Legal Medicine*, 135(3), 951–966. <https://doi.org/10.1007/s00414-020-02460-4>
- Wibawa, A. P., Guntur, M., Purnama, A., Akbar, M. F., & Dwiyanto, F. A. (2018). Metode-metode Klasifikasi. *Prosiding Seminar Ilmu Komputer Dan Teknologi Informasi*, 3(1).