

KLASIFIKASI PENYAKIT KANKER SERVIKS MENGGUNAKAN ALGORITMA SUPPORT VECTOR MACHINE (SVM)

Siti Silvia Arifin^{1*}, Amril Mutoi Siregar¹⁾, Ayu Ratna juwita¹⁾, Tohirin Al Mudzakir¹⁾

¹⁾ Teknik Informatika, Universitas Buana Perjuangan Karawang

*Email Korespondensi: lf17.sitiarifin@mhs.ubpkarawang.ac.id

ABSTRAK

Kanker serviks atau yang sering di sebut kanker leher rahim merupakan penyakit yang mematikan dan banyak sekali menyerang kaum wanita di seluruh dunia. di indonesia dari menteri kesehatan bahkan mencatat kanker ini menempati peringkat ke dua dari jenis kanker serviks yang sering di jumpai setelah kanker payudara. *support vactor machine* (SVM) adalah algoritma yang di gunakan untuk pengklasifikasi dengan membagi data menjadi dua kelas lalu menjadikan garis hayperplane untuk pemisah dua kelas dan margin untuk pemisah antar garis hayperplan dan support vactor. Seperti pada Penelitian kali ini yang bertujuan untuk mempermudah mengklasifikasikan data kanker serviks, karena dari permasalahan data kanker serviks memerlukan klasifikasi untuk menentukan mana saja data yang lebih dominan negatif atau positif dengan menerapkan algoritma support vector machine (SVM). data yang di ambil dari *website Archive.com* sebanyak 72 data dan 19 atribut dengan menggunakan data *training* sebanyak 59 data dan 4 atribut dimana di antaranya dukungan sosial instrumental, pengetahuan pemberdayaan, kemampuan pemberdayaan, keinginan pemberdayaan. Data di bagi menjadi dua yaitu menjadi data *testing* dan *training* dengan membagi 80:20 pada pengujian kali ini menggunakan *python* dengan data *training* dan menggunakan *tools orange* lalu menggunakan data *testing*. Dari hasil pengujian dengan menggunakan *orange* membagi data 80:20 memiliki nilai nilai *accuracy* sebesar 92,9%, dan *python* sebesar 87%.

Kata Kunci: hyperplane, kanker serviks, *support vactor machine* (SVM)

ABSTRACT

Cervical cancer or often called cervical cancer is a deadly disease and affects many women around the world. In Indonesia, the Ministry of Health even noted that this cancer is ranked second from the type of cervical cancer that is often encountered after breast cancer. support vector machine (SVM) is an algorithm used for classifiers by dividing the data into two classes and then making a hayperplane line to separate the two classes and a margin to separate the hayperplan and support vector lines. As in this study which aims to make it easier to classify cervical cancer data, because of cervical cancer data problems, it requires classification to determine which data is more dominantly negative or positive by applying the support vector machine (SVM) algorithm. data taken from the Archive.com website as many as 72 data and 19 attributes using training data as much as 59 data and 4 attributes which include instrumental social support, empowerment knowledge, empowerment ability, empowerment desire. The data is divided into two, namely into testing and training data by dividing 80:20 in this test using python with training data and using orange tools and then using testing data. From the test results using orange, dividing the data 80:20 has an accuracy value of 92.9%, and python by 87%.

Keywords: hyperplane, cervical cancer, *support vector machine* (SVM)

PENDAHULUAN

Penyakit kanker leher Rahim atau bisa d sebut kanker serviks sudah menjadi penyakit yang mematikan bagi kaum perempuan setelah kanker payudara dan kanker serviks sudah menempati urutan ke 1 kategori penyakit yang banyak menelan korban jiwa di dunia dan

di Indonesia sendiri menempati urutan nomer 2 untuk pembunuh terbanyak bagi kaum wanita.

Pada penelitian sebelumnya jumlah angka terjadinya kanker di Indonesia sebanyak (136,2 - 100.000 jiwa) dengan menempati urutan ke delapan di Asia Tenggara, sedangkan jumlah kasus kanker serviks total 23,4 kasus - 100.000 jiwa di Asia dengan rata-rata 13,9 kematian - 100.000 jiwa. Kemudian penelitian yang dilakukan Andani (Andanni, 2016) pada umumnya bahwa perempuan saat ini telah menjadi sasaran empuk untuk terpapar beberapa jenis kanker ganas. Karena pada dasarnya para kaum hawa masih saja banyak yang tidak menjaga pola makan sehat dan olah raga dengan cara merawat tubuh dengan sebaik mungkin agar terhindar dari penyakit yang mematikan.

Penelitian yang dilakukan Praningki and Budi (Praningki and Budi, 2018) saat menggunakan teknik data mining agar diimplementasikan ke sistem untuk melakukan prediksi kanker serviks yang berfokus pada dataset untuk diagnosa medis pasien yang akan dilakukan pengujian pada pap smear. dan penelitian ini menggunakan algoritma Classification And Regression Trees (CART), Naive Bayes, dan k-Nearest Neighbor (k-NN). Kemudian Tsani and Harliana, (Tsani and Harliana, 2019) mengenai kebijakan komputerisasi pembuatan suatu aplikasi pengelolaan citra agar mendeteksi kanker serviks menggunakan metode Canny dan transformasi powerlaw dari hasil parameter keberhasilan sistem yang di gunakan untuk membandingkan hasil diagnosa sistem dengan hasil diagnosa dokter 10 dari gambar diagnosa akurasi yang di dapat 80% rata - rata running time Canny terendah 0,058 ms, dan nilai MSE 11296.3. penelitian selesai Pratama, Wihandika and Ratnawati (Pratama, Wihandika and Ratnawati, 2018) pada saat implementasi sistem cerdas agar mendapatkan hasil klasifikasi dengan data prediksi kelulusan mahasiswa memakai parameter algoritma support vector machine (SVM) mengklasifikasi dan membagi 2 data kelas dengan menggunakan kernel Gaussian RBF dengan kombinasi nilai parameter $\lambda = 0,5$, konstanta $\gamma = 0,01$, dan ϵ (epsilon) = 0,001 itermax = 100, c = 1 menggunakan data latih sebanyak 170 data. Saat penelitian sudah di lakukan maka di dapat hasil rata-rata akurasi sebesar 80,55 %.

Berdasarkan dari permasalahan penelitian di atas dapat menjadi solusi pada penelitian kali ini yang bertujuan untuk mengklasifikasi apakah dataset penyakit kanker serviks dengan menggunakan algoritma SVM (Support Vektor Machine) tentunya dengan mencari nilai accuracy menggunakan python dan orange lalu mencari garis hyperplane untuk pembatas antar kelas dan mendapatkan garis margin yaitu pembatas support vector dan hyperplane dengan demikian penelitian kali ini mendapatkan prediksi terbaik dengan akurasi menggunakan orange (93%) dan python (93%) dapat di simpulkan bahwa metode algoritma SVM adalah metode klasifikasi terbaik yang di gunakan untuk penelitian data kanker serviks.

METODE PENELITIAN

Kanker Serviks

Menurut Nurashiah and Marlina (Nurashiah and Marlina, 2018) kanker serviks adalah kanker dengan menyebabkan kematian terbesar di dunia setelah kanker serviks menempati urutan nomer dua setelah kanker payudara dan di Indonesia kanker serviks menempati urutan pertama dengan total jumlah 15.000 kasus baru kanker serviks atau kanker leher rahim dengan jumlah total kematian sekitar 7.500 kasus per tahun 2015.

Klasifikasi

Menurut Indri Monika Parapat *et.al.*, (Parapat, Furqon and Sutrisno, 2018) klasifikasi merupakan suatu proses dari suatu model untuk mengklasifikasi suatu objek menggunakan atribut -atributnya. Klasifikasi data maupun dokumennya dapat

membangun aturan dengan aturan klasifikasi tertentu menggunakan data training yang sering di gunakan pada tahap pembelajaran dan pengujian data sebagai data testing.

Algoritma Support Vector Machine

Algoritma *Support Vector Machine* (SVM) merupakan metode efek pembelajaran yang baik tentunya operator pendukung dari beberapa jenis *kernel* salah satunya yaitu radial, polinomial, saraf, analisis varians, *epachnenikov*, *Gaussian* dan beberapa kombinasi kuadrat. Penerapan jenis kernel trik ini di berikan di bagian parameter Siregar *et al.*, (Siregar *et al.*, 2019) kemudian Menurut Pratama, Wiandika and Ratnawati (Pratama, Wiandika and Ratnawati, 2018) algoritma SVM adalah metode terbaik untuk sebuah permasalahan klasifikasi. Rancangan dari masalah klasifikasi dua kelas dan dibutuhkan training positif dan negative. Metode SVM cara pengerjaan nya yaitu dengan menggunakan prinsip *Structural Risk Minimization* yang tentunya agar mencari tahu hyperplane terbaik dengan memisahkan 2 *class* dan mencari garis *margin* untuk pembatas *support vector* dan *hyperplane*. Untuk dataset diberikan variabel X_i , sedangkan untuk kelas pada dataset diberikan variabel Y_i . Metode SVM membagi dataset menjadi 2 kelas. Kelas pertama yang dipisah oleh hyperplane bernilai 1, sedangkan kelas lainnya bernilai -1.

$$X_i.W + b \geq 1 \text{ untuk } Y_i = 1 \quad (1)$$

$$X_i.W + b \leq -1 \text{ untuk } Y_i = -1 \quad (2)$$

Keterangan :

X_i = data ke -i

W = nilai bobot support vector yang tegak lurus dengan hyperplane

b = nilai bias

Y_i = kelas data ke -i

Perhitungan menggunakan algoritma *Support Vector Machine* (SVM) pada data penyakit kanker serviks. Maka memiliki *feature* ($w_1 w_2 w_3 w_4 \dots n$) Formulasi yang akan digunakan adalah sebagai berikut :

$$\frac{1}{2} ||w||^2 = \frac{1}{2} (w^2 + w_2^2 + w_3^2 \dots w_4^2) \text{ dengan } y (w^1, x^1 + b)$$

$$i = 1, 2, 3, 4, \dots, N$$

$$y_i (w_1, x_1 + w_2, x_2 \dots \dots \dots n)$$

+1 untuk tidak terdeteksi

$$y_i (w_1, x_1 + w_2, x_2 \dots \dots \dots n)$$

-1 untuk yang terdeteksi

Confusion Matrix

Menurut Fibrianda and Bhawiyuga, (Fibrianda and Bhawiyuga, 2018) *confusion matrix* merupakan suatu alat yang memiliki fungsi untuk menganalisis apakah *classifier* baik dalam mengidentifikasi tupel termasuk golongan yang berbeda nilai pada *true positive* dan *true negative* memberitahu untuk *classifier* mengklasifikasikan data sedangkan *false positive* dan *false negative* memberitahu informasi ketika *cluster* mengklasifikasikan data secara tidak benar.

Tabel 1 *Confusion Matrix*

	Classified positif	Classified Negatif
Actual Positif	Tp	Fn
Actual Negatif	Fp	Tn

TP (*True Positive*) → total data dengan nilai *true positive* dan nilai prediksi positif FP (*false positive*) → total data dengan nilai *true negative* dan nilai prediksi positif FN (*false negative*) → jumlah data dengan nilai *true positive* dan nilai prediksi negatif TN (*True Negative*) → Total nilai data dengan nilai *true negative* dan nilai prediksi negatif.

Python

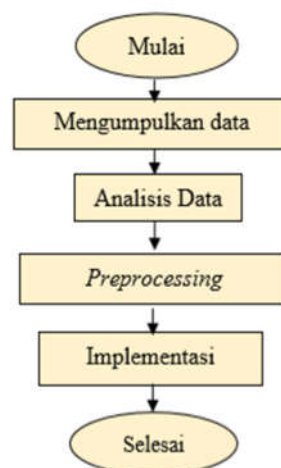
Python merupakan cara yang sering digunakan untuk menciptakan lingkungan yang terisolasi dengan dunia luar maka dari itu modul *python* yang dipasang tidak dapat mempengaruhi dunia global dan yang utama tidak mempengaruhi proyek lain. Ada beberapa alat untuk menciptakan lingkungan *online* dalam *python* termasuk *miniconda*. *Interpreter* dapat di andalkan dalam suatu yang sangat sulit dengan ketelitian juga perhitungan tingkat tinggi. (Teknologi, 2017)

Orange

Pada pemrograman *orange* menyediakan banyak *widget* untuk kita letakan di kanvas/papan gambar, lalu hubungkan dengan *widget* lain. Aplikasi *orange* hanya fokus pada analisis *tex cluster* dengan menggunakan data yang diperoleh dari *web scrapper*. (Sentiya et al., 2018)

Gambaran umum penelitian

Adapun gambaran umum penelitian di gunakan.



Gambar 1 Gambaran Penelitian

Pengumpulan data merupakan tahap yang bertujuan untuk memperoleh informasi-informasi atau data-data yang berhubungan dengan penelitian. Penelitian ini dimulai dengan melakukan pengumpulan data tentang penyakit kanker serviks dari (Sobar, Machmud and Wijaya, 2016) yang terdiri dari 72 data dan 19 atribut.

Preprocessing

Preprocessing/Cleaning Pada tahapan ini dilakukan pembersihan (*cleaning*) data yang menjadi fokus dalam *data mining*. Pembersihan data dilakukan untuk mengganti data yang memiliki *missing value* atau data yang tidak lengkap memiliki nilai default dengan menggunakan nilai median.

HASIL DAN PEMBAHASAN

Hasil penelitian

Pada permasalahan dataset kanker serviks sebelum di lakukan *preprocessing* di dapat dari situs archive.com pada tanggal 17 november 2020 yang terdiri dari 72 data dan 19 atribut. Setelah data terkumpul selanjutnya masuk pada tahap *preprocessing* untuk mengganti data yang memiliki *missing value* atau data yang tidak lengkap memiliki nilai default dengan menggunakan nilai median.

atribut-atribut yang bernilai relevan diantaranya dukungan sosial instrumental, pengetahuan pemberdayaan, kemampuan pemberdayaan dan keinginan pemberdayaan dan satu atribut *keterangan*. Dataset dibagi menjadi dua yakni data *training* dan data *testing* dengan pembagian data 80 : 20, dimana 80% data *training* dan 20% data *testing* Pada algoritma *support vector machine (SVM)* pengujian dilakukan dengan tiga tahapan menggunakan perhitungan manual, orange dan python.

Tabel 2 data *training* dan *testing*

kanker serviks	apresiasi dukungan sosial	dukungan sosial instrumental	pengetahuan pemberdayaan	kemampuan pemberdayaan	keinginan pemberdayaan
Positif	7	12	12	11	8
Positif	6	5	5	4	4
Positif	6	11	3	3	15
Positif	4	4	4	4	4
Positif	6	12	5	4	7
Positif	2	7	13	9	6
Positif	3	15	3	3	5
Positif	2	9	8	7	12
Positif	10	15	13	15	15
...	...	...	...	...	...

1. Hasil Pengujian Algoritma SVM menggunakan perhitungan manual

$$\frac{1}{2} ||w|^2 = \frac{1}{2} (w^2 + w_2^2 + w_3^2 \dots w_4^2) \text{ dengan } y (w^1, x^1 + b)$$

$$i = 1, 2, 3, 4, \dots, N$$

$$y_i (w_1, x_1 + w_2, x_2 \dots \dots \dots n)$$

+1 untuk tidak terdeteksi

$$y_i (w_1, x_1 + w_2, x_2 \dots \dots \dots n)$$

-1 untuk yang terdeteksi

Tabel 3 data yang akan di kelola

Dukungan sosial instrumental	pengetahuan pemberdayaan	kemampuan pemberdayaan	keinginan pemberdayaan	kanker serviks	Keterangan
12	10	10	9	1	Positif
7	8	5	3	1	Positif
12	12	12	12	0	Negatif
15	15	15	15	0	Negatif
14	15	14	15	0	Negatif

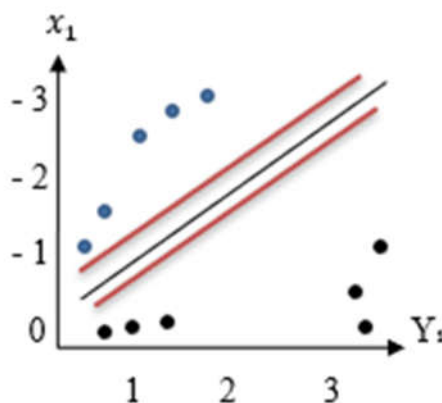
Masukan data yang di kelolah ke dalam $(w^2 + w_2^2 + w_3^2 \dots w_4^2)$ dengan $y (w^1, x^1 + b)$ agar menjadi seperti di bawah ini:

$$\begin{aligned}(w_1, 12) (w_2, 10) (w_3, 10) (w_4, 9) + b &> +1 \\(w_1, 7) (w_2, 8) (w_3, 5) (w_4, 3) + b &> +1 \\(w_1, 12) (w_2, 12) (w_3, 12) (w_4, 12) + b &> -1 \\(w_1, 15) (w_2, 15) (w_3, 15) (w_4, 15) - b &> -1 \\(w_1, 14) (w_2, 15) (w_3, 14) (w_4, 15) - b &> -1\end{aligned}$$

Selanjutnya melakukan persamaan linier dan eliminasi setiap W Setelah melalui proses perhitungan maka didapatlah hasil untuk visualisasi garis *hyperplane* dan *margin* dengan nilai yang didapat $w_1(1,672), w_2(-0,641), w_3(-1,672), w_4(-2,730), w_5(0,045), w_6(1,744), w_7(-15,077), w_8(3,366), w_9(-4,995), b(50,493)$

$$\begin{aligned}y &= +1 \text{ jika } w.z + b > 0 \\&= -1 \text{ jika } w.z + b < 0\end{aligned}$$

hasil perhitungan manual Gambar 2 dapat di lihat bahwa ada pattern yang membuat tahap anggota dari dua *category*: +1 dan -1 pada petran yang berwarna biru di simbolkan kategori -1 sedangkan warna hitam itu menandakan symbol +1 lalu pada garis 2 merah itu adalah garis *margin* yaitu untuk pembatas atar support vecktor dan garis *hyperplane* lalu untuk garis miring 1 di tangan garis *margin* itu adalah garis *hyperplane* sebagai pembatas 2 *class*.



Gambar 2 garis *margin* dan *hyperplane*

Tabel 4 hasil perhitungan manual

Dukungan sosial instrumental	pengetahuan pember-dayaan	kemampuan pember - dayaan	keinginan pembed- rdayaan	kanker serviks	Keterangan
12	10	10	9	1	Positif
7	8	5	3	0	Negatif
12	12	12	12	0	Negatif
15	15	15	15	0	Negatif
14	15	14	15	0	Positif
12	14	10	14	1	Positif
12	15	11	15	0	Negatif
10	12	13	12	0	Positif
15	15	13	15	0	Negatif
13	13	12	13	0	Positif

Hasil berdasarkan accuracy menggunakan *confusion matrix* pada *Orange*

		Predicted		
		Negative	Positif	Σ
Actual	Negative	10	0	10
	Positif	1	3	4
Σ		11	3	14

Gambar 3 *confusion matrix* pada *orange*

- a. Hasil *confusion matrix* negatif

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} = \frac{10 + 3}{10 + 3 + 0 + 1} = 93\%$$

$$Precision = \frac{TP}{TP + FP} = \frac{10}{10 + 0} = 90,9\%$$

$$Recall = \frac{TP}{TP + FN} = \frac{10}{10 + 1} = 1\%$$

$$F1 = 2 * \frac{(recall * precision)}{(recall + precision)} = 2 * \frac{(1 * 90,9)}{(1 + 90,9)} = 95,2\%$$

Maka dari hasil pengujian data dengan menggunakan tools *orange* accuracy di dapat sebesar 93% precision 90,9% recall 1% dan F1 95,2%

Hasil pengujian pada *Python* menggunakan algoritma support vector machine (SVM)

Pada accuracy yang di dapat dari pengujian menggunakan *python* mendapatkan nilai 93% dengan nilai precision 91%, recall 1% dan F1 95%

	precision	recall	f1-score	support
Negative	0.91	1.00	0.95	10
Positif	1.00	0.80	0.89	5
accuracy			0.93	15
macro avg	0.95	0.90	0.92	15
weighted avg	0.94	0.93	0.93	15

Gambar 4 Hasil accuracy pada *python*

Dari gambar 4 perhitungan hasil *accuracy Confusion Matrix* menggunakan metode algoritma *Support Vector Machine (SVM)* lalu menggunakan tools *orange* dan *python* maka dapat disimpulkan akurasi memiliki nilai yang sama yaitu sebesar 93% sehingga akurasi menggunakan algoritma SVM adalah akurasi yang baik dan tepat dalam menentukan klasifikasi penyakit kanker serviks lalu nilai *recall* pada *orange* negatif hasilnya 1 dan positif 0,75 lalu pada *precision* nilai negatif 0,91 sedangkan positif 1 dan F1 negatif hasilnya 0,952 lalu positif 0,857 untuk hasil *Confusion Matrix* pada *python* pada *recall* negatif bernilai 1 lalu pada positif 0,80 sedangkan *precision* untuk negatif hasilnya 0,91 pada positif yaitu 1 dan untuk *F1* negatif hasilnya 0,95 dan positif 0,89.

KESIMPULAN

Berdasarkan dari hasil pengujian yang telah di lakukan. Maka penulis dapat mengambil kesimpulan sebagai berikut.

1. Algoritma Support Vector Machine (SVM) telah berhasil di terapkan untuk klasifikasi penyakit kanker serviks karena dalam pengujian menggunakan *orange* dan *python* hasilnya sama dan sesuai.

2. Akurasi menggunakan algoritma *Support Vector Machine* (SVM) menggunakan 2 cara yaitu dengan accuracy pada orange dan python dimana hasil dari keduanya mendapatkan nilai yang sama yaitu sebesar 93% tentunya penelitian kaliini sangat akurat dengan menggunakan algoritma SVM.

REFERENSI

- [1] Andanni, R. M. (2016) 'Perancangan Strategi Kampanye Bahaya Kanker Serviks Bagi Remaja Putri Usia 13-18 Tahun'. Available at: <http://repository.its.ac.id/75601/>.
- [2] Fibrianda, M. F. and Bhawiyuga, A. (2018) 'Analisis Perbandingan Akurasi Deteksi Serangan Pada Jaringan Komputer Dengan Metode Naïve Bayes Dan Support Vector Machine (SVM)', *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 2(9), pp. 3112–3123.
- [3] Nurashiah, A. and Marlina, M. T. (2018) 'Efektivitas Pelatihan Konseling Kesehatan Reproduksi Terhadap Peningkatan Kompetensi Kader Posyandu Dalam Pelayanan Konseling Pencegahan Kanker Serviks di Kabupaten Kuningan Tahun 2018', *Jurnal Ilmu Kesehatan Bhakti Husada: Health Sciences Journal*, 9(2), pp. 34–39. doi: 10.34305/jikbh.v9i2.65.
- [4] Parapat, I. M., Furqon, M. T. and Sutrisno (2018) 'Penerapan Metode Support Vector Machine (SVM) Pada Klasifikasi Penyimpangan Tumbuh Kembang Anak', *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 2(10), pp. 3163–3169. Available at: <https://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/2577>.
- [5] Praningki, T. and Budi, I. (2018) 'Sistem Prediksi Penyakit Kanker Serviks Menggunakan CART, Naive Bayes, dan k-NN', *Creative Information Technology Journal*, 4(2), p. 83. doi: 10.24076/citec.2017v4i2.100.
- [6] Pratama, A., Wihandika, R. C. and Ratnawati, D. E. (2018) 'Implementasi Algoritme Support Vector Machine (SVM) untuk Prediksi Ketepatan Waktu Kelulusan Mahasiswa', *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 2(March), pp. 1704–1708.
- [7] Sentiya, A. et al. (2018) 'Analisis Text Clustering Akun Fanpage Shopee Indonesia Dengan Komentar Followers Menggunakan Tools Orange Data Mining', *Bina Darma Conference on Computer Science*, 2, pp. 1055–1067.
- [8] Siregar, A. et al. (2019) 'Comparison Study Of Term Weighting Optimally With SVM In Sentiment Analysis', (January). doi: 10.4108/eai.18-7-2019.2288508.
- [9] Sobar, Machmud, R. and Wijaya, A. (2016) *Cervical Cancer Behavior Risk Data Set, Center for Machine Learning and Intelligent Systems*. Available at: <https://archive.ics.uci.edu/ml/datasets/Cervical+Cancer+Behavior+Risk#>.
- [10] Teknologi, S. N. (2017) 'Python Merupakan Alternatif Yang Sederhana Dan', *Abacus A Journal Of Accounting Finance And Business Studies*, 1(November), pp. 1–22.
- [11] Tsani, N. B. and Harliana, H. (2019) 'Implementasi Deteksi Tepi Canny Dengan Transformasi Powerlaw Dalam Mendeteksi Stadium Kanker Serviks', *Jurnal Ilmiah Intech: Information Technology Journal of UMUS*, 1(01), pp. 22–33. doi: 10.46772/intech.v1i01.35.