

Optimasi Metode Algoritma C.45 Menggunakan Algoritma Genetika Untuk Penentuan Jenis Benang Pada Konveksi

¹Wisti Dwi Septiani, ²Indah Puspitorini, ³Murniyati, ⁴Diana Novita

^{1,2,3} Universitas Bina Sarana Informatika, Indonesia, ⁴ Universitas Esa Unggul, Indonesia

¹wisti.wst@bsi.ac.id, ²indah.itn@bsi.ac.id, ³murni.mni@bsi.ac.id, ⁴diana.novita@esaunggul.ac.id

ABSTRAK

PT. Nilosa Rama Buana adalah sebuah industri di bidang konveksi yang memproduksi produk konveksi baik dari instansi perkantoran, sekolah, perusahaan maupun pribadi. Produk yang dihasilkan seperti kaos sablon, kaos distro, sprei, selimut, pakaian bayi dan anak. Tujuan penelitian ini adalah memprediksi jenis rajutan, gramasi, dan jenis ukuran benang yang sesuai untuk membantu konveksi dalam pembuatan produk sehingga persiapan produksi menjadi lebih efisien menggunakan algoritma klasifikasi data mining. Metode yang digunakan dalam penelitian ini adalah implementasi Algoritma Genetika optimasi terhadap metode Algoritma C4.5, dilakukan optimasi untuk meningkatkan akurasi metode. Hasil penelitian ini adalah kenaikan nilai akurasi 45,18% dengan pengujian optimasi menggunakan metode Algoritma Genetika, juga dihasilkan pohon keputusan dari metode Decision Tree. Kesimpulan dari penelitian ini adalah metode optimasi Algoritma Genetika dapat membantu meningkatkan kehandalan dan rule yang dihasilkan dari pohon keputusan Algoritma C4.5 dapat dijadikan acuan untuk membantu konveksi dalam penentuan jenis benang pada persiapan produksi.

Kata Kunci: Algoritma C4.5, Algoritma Genetika, Data Mining, Metode Klasifikasi

PENDAHULUAN

Pada penelitian ini dilakukan optimasi terhadap Algoritma C4.5 menggunakan algoritma Genetika. Optimasi adalah proses menyelesaikan suatu masalah tertentu supaya berada pada kondisi yang paling menguntungkan dari suatu sudut pandang, yaitu berhubungan dengan pencarian nilai minimum atau nilai maksimum (Buani, 2016). Pada penerapan optimasi menggunakan algoritma genetika terdapat peningkatan akurasi dari 97,66% menjadi 99,33% pada algoritma Naïve Bayes untuk prediksi *fertility* (Buani, 2016) dan peningkatan akurasi dari 83,81% menjadi 86,47% pada algoritma C4.5 untuk prediksi *phising website* (Sunge, 2018). Decision Tree merupakan salah satu algoritma klasifikasi data mining. Menurut Gorunescu dalam (Sunge, 2018) Algoritma dalam klasifikasi yang banyak digunakan ialah Decision Tree. Dikarenakan sangat mudah dimengerti dan dijabarkan oleh banyak pengguna juga mudah dipahami dimana cabang pohon disimpulkan dalam bentuk klasifikasi. Tujuan dari penelitian ini adalah melakukan optimasi algoritma C4.5 menggunakan algoritma genetika untuk meningkatkan nilai akurasinya sehingga prediksi yang dihasilkan lebih baik dan akurat.

Seperti kita ketahui saat ini banyak konveksi yang menghasilkan produk seperti pakaian, kaos, sprei, dan lain-lain. Dalam pembuatan produk-produk tersebut dibutuhkan bahan-bahannya yang terbuat dari benang. Benang yang biasa digunakan adalah jenis katun yang mempunyai 4 jenis ukuran benangnya seperti benang 20, 24, 30, dan 40. Setiap ukurannya mempengaruhi ketebalan dari produk yang akan dibuat oleh konveksi. Setiap ukuran benang juga biasa digunakan untuk

membuat bahan kain dengan gramasi yang berbeda beda. Seperti benang 20 biasa digunakan untuk bahan kain dengan gramasi antara 180 sampai dengan 220 gram/m². Benang 24 biasa digunakan untuk bahan kain dengan gramasi antara 170 sampai dengan 210 gram/m². Benang 30 biasa digunakan untuk bahan kain dengan gramasi antara 140 sampai dengan 160 gram/m². Benang 30 biasa digunakan untuk bahan kain dengan gramasi antara 210 sampai dengan 230 gram/m². Penelitian ini dilakukan dengan pengambilan data di konveksi PT. Nilosa Rama Buana yang terletak di Kota Tangerang Selatan. Berdasarkan data yang didapatkan dilakukan penelitian sehingga dapat dihasilkan pengetahuan mengenai produk yang dihasilkan menggunakan ukuran benang apa, gramasinya, serta jenis rajutan benang tunggal atau ganda sehingga bisa digunakan untuk pengetahuan bagi pegawai baru yang akan membuat produksi maupun pebisnis konveksi lainnya.

TINJAUAN PUSTAKA

Data Mining

Data mining telah menarik banyak perhatian dalam dunia sistem informasi dan dalam masyarakat secara keseluruhan dalam beberapa tahun terakhir, karena ketersediaan luas dalam jumlah besar data dan kebutuhan segera untuk mengubah data tersebut menjadi informasi yang berguna dan pengetahuan. Informasi dan pengetahuan yang diperoleh dapat digunakan untuk aplikasi mulai dari pasar analisis, deteksi penipuan, dan retensi pelanggan, untuk pengendalian produksi dan ilmu pengetahuan eksplorasi (Han & Kamber, 2007). Adanya ketersediaan data yang melimpah, kebutuhan akan informasi atau pengetahuan sebagai sarana pendukung dalam pengambilan keputusan baik bagi individu, organisasi, perusahaan dan pemerintahan.

Banyaknya data, ditambah dengan kebutuhan untuk alat analisis data yang kuat, telah digambarkan sebagai kaya data tapi miskin informasi. Jumlah data yang tumbuh secara cepat, dikumpulkan dan disimpan dalam repositori data yang besar dan banyak, telah jauh melampaui kemampuan manusia untuk memahami data-data tersebut tanpa mampu mengelolah data tersebut. Akibatnya, data yang dikumpulkan dalam repositori data yang besar menjadi "kuburan data" (Han & Kamber, 2007).

Hal ini melatarbelakangi lahirnya suatu cabang ilmu pengetahuan baru yaitu *data mining*. *Data mining* adalah untuk mengekstrasikan atau "menambang" pengetahuan dari kumpulan banyak data (Han dan Kamber, 2007). *Data mining* adalah teknik yang merupakan gabungan metode-metode analisis data secara berkesinambungan dengan algoritma-algoritma untuk memproses data berukuran besar. *Data mining* merupakan proses menemukan informasi atau pola yang penting dalam basis data berukuran besar dan merupakan kegiatan untuk menemukan informasi atau pengetahuan yang berguna secara otomatis dari data yang jumlahnya besar.

Data mining, sering juga disebut *knowledge discovery in database (KDD)*, adalah kegiatan yang meliputi pengumpulan, pemakaian data historis untuk menemukan pola keteraturan, pola hubungan dalam set data berukuran besar (Santosa, 2007). Keluaran dari *data mining* ini dapat dijadikan untuk memperbaiki pengambilan keputusan di masa depan. Dalam *data mining* data disimpan secara elektronik dan diolah secara otomatis, atau setidaknya disimpan dalam komputer. *Data mining* adalah tentang menyelesaikan masalah dengan menganalisa data yang telah ada dalam database (Witten & Frank, 2011).

Algoritma C4.5

Berdasarkan tugasnya, *data mining* dikelompokkan menjadi 6 yaitu deskripsi, estimasi, prediksi, klasifikasi, clustering, dan asosiasi (Larose, 2005). Klasifikasi (taksonomi) adalah proses menempatkan objek tertentu (konsep) dalam satu set kategori, berdasarkan masing-masing objek (konsep) *property* (Gorunescu, 2011). Proses klasifikasi didasarkan pada empat komponen mendasar yaitu kelas, prediktor, *training set*, dan pengujian *dataset*.

Diantara model klasifikasi yang paling populer adalah *Decision/Classification Trees, Bayesian Classifiers/Naïve Bayes Classifiers, Neural Networks, Statistical Analysis, Genetic Algorithms,*

Rough Sets, K-Nearest Neighbor Classifier, Rule-based Methods, Memory Based Reasoning, Support Vector Machines (Gorunescu, 2011).

Algoritma C4.5 menggunakan konsep *information gain* atau *entropy reduction* untuk memilih pembagian yang optimal (Larose, 2005). Tahapan dalam membuat pohon keputusan dengan algoritma C4.5 (Gorunescu, 2011) yaitu:

1. Mempersiapkan data *training*, dapat diambil dari data histori yang pernah terjadi sebelumnya dan sudah dikelompokkan dalam kelas-kelas tertentu.
2. Menentukan akar dari pohon dengan menghitung nilai *gain* yang tertinggi dari masing-masing atribut atau berdasarkan nilai *index entropy* terendah. Sebelumnya dihitung terlebih dahulu nilai *index entropy*, dengan rumus:

$$Entropy(i) = - \sum_{j=1}^m f(i,j) \cdot \log_2 f(i,j)$$

3. Hitung nilai *gain* dengan rumus:

$$Entropy\ split = \sum_{i=1}^p \binom{n1}{n} \cdot IE(i)$$

4. Ulangi langkah ke-2 hingga semua record terpartisi. Proses partisi pohon keputusan akan berhenti disaat:
 - a. Semua tupel dalam *record* dalam simpul N mendapat kelas yang sama.
 - b. Tidak ada atribut dalam *record* yang dipartisi lagi.
 - c. Tidak ada *record* di dalam cabang yang kosong

Algoritma Genetika

Pada tahun 1970 Algoritma Genetika (GA) diperkenalkan oleh John Holland di Universitas Michigan, bahwa dari bagian masalah merupakan bentuk dari adaptasi dari alam maupun buatan yang dapat diformulasikan menjadi bagian genetika (Sunge, 2018). Menurut Desiani dan Muhammad dalam (Buani, 2018) Algoritma genetika merupakan suatu algoritma pencarian berdasarkan pada mekanisme seleksi alam dan genetika alam. Algoritma genetika dimulai dengan sekumpulan solusi awal (individu) yang disebut populasi. Satu hal yang sangat penting adalah bahwa satu individu menyatakan satu solusi. Populasi awal akan berevolusi menjadi populasi baru melalui serangkaian iterasi (generasi). Pada akhir iterasi, algoritma genetika mengembalikan satu anggota populasi yang terbaik sebagai solusi untuk masalah yang dihadapi. Pada setiap iterasi, proses evolusi yang terjadi adalah sebagai berikut:

- a) Dua individu dipilih sebagai orang tua (*parent*) berdasarkan mekanisme tertentu. Kedua *parent* ini kemudian dikawinkan melalui operator *crossover* (kawin silang) untuk menghasilkan dua individu anak atau *offspring*.
- b) Dengan probabilitas tertentu, dua individu anak ini mungkin mengalami perubahan gen melalui operator *mutation*.
- c) Suatu skema penggantian (*replacement scheme*) tertentu diterapkan sehingga menghasilkan populasi baru.

METODE PENELITIAN

Penelitian dilakukan dengan cara melakukan eksperimen dalam bentuk sistem penunjang keputusan untuk penentuan jenis benang. Metode penelitian memiliki 5 tahapan sebagai berikut:

1. Pengumpulan data

Data yang digunakan dalam penelitian ini adalah data produksi konveksi di PT. Nilosa Rama Buana. Dataset berisi atribut: *benang*, *gramasi*, *jenis_rajutan*, dan *jenis_produk* (atribut hasil

prediksi). Prediksi yang dilakukan adalah menentukan jenis ukuran benang, gramasi kain, dan jenis rajutan yang sesuai untuk membuat kaos sablon, kaos distro, pakaian bayi dan anak, sprei, selimut dan kaos kaki, dengan syarat di bawah ini (Puspitorini, Novita, Murniyati, & Septiani, 2023):

- a. Jenis ukuran benang : apakah 20, 24, 30 atau 40
 - b. Gramasi kain : apakah 180 sampai dengan 220 gram/m², 170 sampai dengan 210 gram/m², 140 sampai dengan 160 gram/m² atau 210 sampai dengan 230 gram/m².
 - c. Jenis Rajutan : Jarum tunggal atau Jarum ganda
2. Pengolahan Data Awal
Dilakukan *data validation*, *data integration and transformation*, dan *data size and dicritization*. Pada dataset diabetes ini semua atribut digunakan dan tidak ada data kosong.
 3. Model/Metode yang Diusulkan
Model yang diusulkan adalah Algoritma Decision Tree dengan seleksi fitur Algoritma Genetika.
 4. Eksperimen dan Pengujian Model/Metode
Eksperimen dilakukan dengan mengolah dataset konveksi menggunakan Algoritma Decision Tree dengan seleksi Algoritma Genetika. Tools yang digunakan adalah Rapidminer.
 5. Evaluasi dan Validasi Hasil
Evaluasi dan validasi diuji tingkat akurasinya dengan Confusion Matrix.

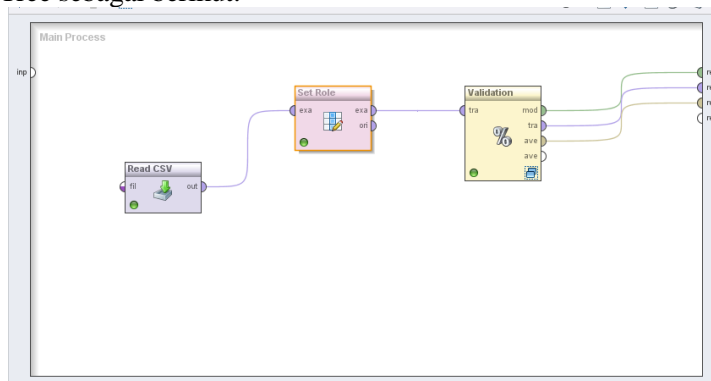
HASIL DAN PEMBAHASAN

Evaluasi dan Validasi

Eksperimen dan pengujian model dilakukan untuk menghitung dan mendapatkan rule-rule dari model yang diusulkan.

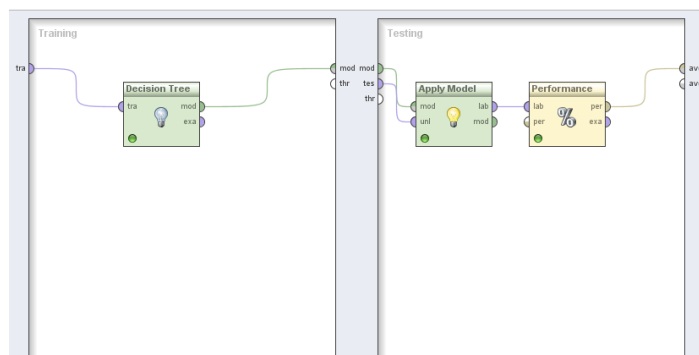
1. Decision Tree

Eksperimen pertama dilakukan dengan pengujian model K-Fold Cross Validation untuk Algoritma Decision Tree sebagai berikut:



Sumber: (Septiani, Puspitorini, Murniyati, & Novita, 2023)

Gambar 1. Pengujian Cross Validation

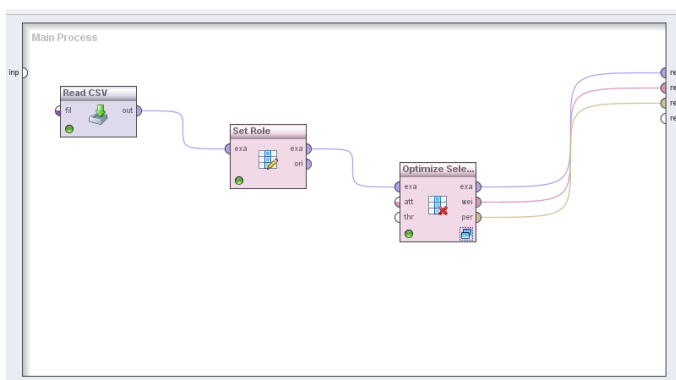


Sumber: (Septiani, Puspitorini, Murniyati, & Novita, 2023)

Gambar 2. Pengujian Metode Decision Tree

2. Algoritma Genetika

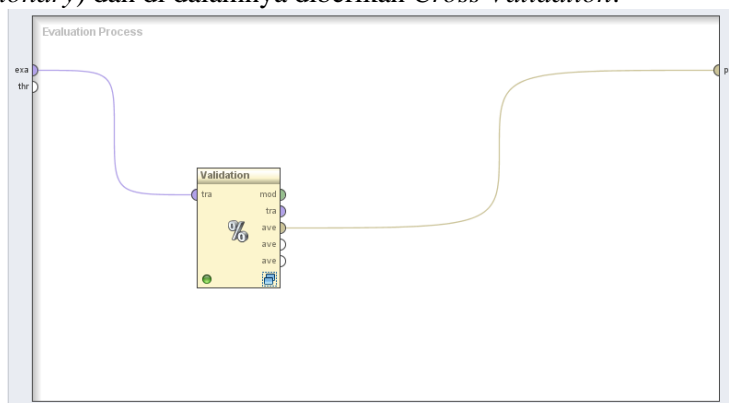
Pengujian menggunakan Algoritma C4.5 dengan fitur seleksi Algoritma Genetika sebagai berikut:



Sumber: (Septiani, Puspitorini, Murniyati, & Novita, 2023)

Gambar 3. *Optimize Selection (Evolutionary)*

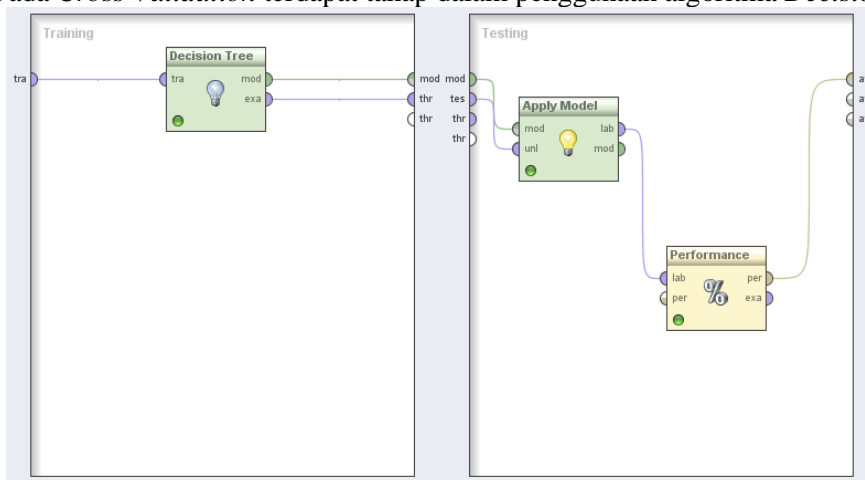
Berdasarkan gambar 3 dataset konveksi dihubungkan dengan *Feature Optimize Selection (Evolutionary)* dan di dalamnya diberikan *Cross Validation*.



Sumber: (Septiani, Puspitorini, Murniyati, & Novita, 2023)

Gambar 4. *Cross Validation*

Berdasarkan gambar 5 penggunaan *Cross Validation* dalam dataset konveksi terdiri dari *10-fold validation* Pada *Cross Validation* terdapat tahap dalam penggunaan algoritma *Decision Tree*.



Sumber: (Septiani, Puspitorini, Murniyati, & Novita, 2023)

Gambar 5. *Model Decision Tree*

Hasil Penelitian

Validasi terhadap model yang diterapkan yaitu Decision Tree untuk menguji bagaimana performa dari algoritma tersebut dan menghasilkan nilai akurasi 54,82% seperti pada Tabel 1 berikut:

Tabel 1. Akurasi Algoritma *Decision Tree*

| accuracy: 54.82% +/- 6.19% (mikro: 54.93%) | | | | | | | |
|--|------------------|------------------|---------------------|------------|--------------|----------------|-----------------|
| | true KAOS SABLON | true KAOS DISTRO | true Pakaian Bayi d | true SPREI | true SELIMUT | true KAOS KAKI | class precision |
| pred. KAOS SABLON | 0 | 0 | 0 | 0 | 0 | 0 | 0.00% |
| pred. KAOS DISTRO | 0 | 6 | 0 | 0 | 0 | 0 | 100.00% |
| pred. Pakaian Bayi d | 0 | 0 | 0 | 0 | 0 | 0 | 0.00% |
| pred. SPREI | 6 | 0 | 0 | 18 | 0 | 0 | 75.00% |
| pred. SELIMUT | 0 | 0 | 12 | 0 | 15 | 14 | 36.59% |
| pred. KAOS KAKI | 0 | 0 | 0 | 0 | 0 | 0 | 0.00% |
| class recall | 0.00% | 100.00% | 0.00% | 100.00% | 100.00% | 0.00% | |

Sumber: (Septiani, Puspitorini, Murniyati, & Novita, 2023)

Optimasi terhadap metode Decision Tree menggunakan Algoritma Genetika menghasilkan peningkatan akurasi yaitu 45,18% seperti pada Tabel 2 berikut:

Tabel 2. Akurasi Optimasi Algoritma Genetika

| accuracy: 100.00% +/- 0.00% (mikro: 100.00%) | | | | | |
|--|----------------|----------------|-----------------|------------------|-----------------|
| | true Dua Puluh | true Dua Empat | true Tiga Puluh | true Empat Puluh | class precision |
| pred. Dua Puluh | 9 | 0 | 0 | 0 | 100.00% |
| pred. Dua Empat | 0 | 9 | 0 | 0 | 100.00% |
| pred. Tiga Puluh | 0 | 0 | 12 | 0 | 100.00% |
| pred. Empat Puluh | 0 | 0 | 0 | 41 | 100.00% |
| class recall | 100.00% | 100.00% | 100.00% | 100.00% | |

Sumber: (Septiani, Puspitorini, Murniyati, & Novita, 2023)

Hasil dari penggunaan metode Decision Tree adalah pohon keputusan seperti pada gambar 6 sebagai berikut:



Sumber: (Septiani, Puspitorini, Murniyati, & Novita, 2023)

Gambar 6. Pohon Keputusan

Rule yang dihasilkan adalah sebagai berikut:

1. Jika jenis Rajutan Jarum Tunggal, Gramasi 180- 220, benang 20 maka produksi Kaos Sablon,
2. Jika jenis Rajutan Jarum Tunggal, Gramasi 170 – 210, benang 24 maka produksi Kaos Sablon
3. Jika jenis Rajutan Jarum Tunggal, Gramasi 140 – 160, benang 20 maka produksi Kaos Distro
4. Jika jenis Rajutan Jarum Tunggal, Gramasi 210 – 230, benang 24 maka produksi Kaos Distro
5. Jika jenis Rajutan Jarum Tunggal, Gramasi 210 – 230, benang 20 maka produksi Sprei
6. Jika jenis Rajutan Jarum Tunggal, Gramasi 210 – 230, benang 24 maka produksi Sprei
7. Jika jenis Rajutan Ganda, Gramasi 180- 200, benang 40 maka produksi Selimut

KESIMPULAN

Rules yang dihasilkan dari pengujian penggunaan fitur seleksi Algoritma Genetika yang diterapkan pada Algoritma C4.5 dapat dijadikan kontribusi dalam pengambilan keputusan terhadap penentuan jenis benang dalam produksi di konveksi. Evaluasi dalam pengujian menggunakan

Algoritma C4.5 dengan seleksi fitur Algoritma Genetika ini adanya peningkatan nilai akurasi sebesar 45,18% dari penelitian tanpa fitur seleksi yang memiliki akurasi hanya 54,82%. Sehingga dapat disimpulkan bahwa penggunaan fitur seleksi mampu meningkatkan nilai akurasi. Penelitian ini dapat dijadikan masukan untuk dilanjutkan kembali dengan metode optimasi lain seperti Adabbost dan PSO. Berdasarkan penelitian ini dapat diberikan saran untuk diadakannya penelitian lebih lanjut dengan melakukan pengujian dengan metode lain seperti SVM, Nural Network, ataupun komparasi dari beberapa metode klasifikasi data mining.

REFERENSI

- Buani, D. C. P. (2018). Prediksi Penyakit Hepatitis Menggunakan Algoritma Naive Bayes Dengan Seleksi Fitur Algoritma Genetika. *Jurnal Evolusi*, 6(2), 1–5. Retrieved from ejournal.bsi.ac.id
- Gorunescu, F. (2011). *Data Mining: Concepts and Techniques*. Verlag Berlin Heidelberg: Springer.
- Larose, D. T. (2005). *Discovering Knowledge in Database*. New Jersey: John Willey & Sons Inc.
- Puspitorini, I., Novita, D., Murniyati, & Septiani, W. D. (2023). Penentuan Jenis Benang Untuk Pembuatan Produk Pada Konveksi Menggunakan Metode Algoritma C4 . 5 Dengan Rapid Miner. 12, 385–392.
- Septiani, W. D., & Marlina. (2021). *Laporan Akhir Penelitian*. Jakarta.
- Septiani, W. D., Puspitorini, I., Murniyati, & Novita, D. (2023). Laporan Akhir Penelitian. *Universitas Bina Sarana Informatika*.
- Sunge, A. S. (2018). Optimasi Algoritma C4.5 Menggunakan Genetic Algoritma Dalam Memprediksi Website Phishing. *Seminar Nasional Inovasi Dan Tren (SNIT)*, 92.