

## Implementasi Algoritma C4.5 Pada Jumlah Penduduk Kota Prabumulih menggunakan Metode *Klasifikasi*

Devi Udariansyah<sup>1</sup>, Robbi Lesmana<sup>2\*</sup>

<sup>1,2</sup>Program Studi Teknik Informatika, Fakultas Ilmu Komputer,  
Universitas Bina Darma Palembang

Email : [devi.udariansyah@binadarma.ac.id](mailto:devi.udariansyah@binadarma.ac.id)<sup>1</sup>, [robbilesmana2017@gmail.com](mailto:robbilesmana2017@gmail.com)<sup>2\*</sup>

### Abstrak

Penelitian ini bertujuan untuk mengklasifikasikan jumlah penduduk berdasarkan pekerjaan dari tahun 2018 sampai dengan tahun 2020 pada badan pusat statistik dan dinas kependudukan pencatatan sipil kota prabumulih dengan menerapkan algoritma c4.5. penulis membatasi masalah hanya memberikan informasi data jumlah penduduk berdasarkan pekerjaan. data dalam penelitian ini adalah data set kependudukan yang di ambil langsung pada dinas kependudukan dan pencatatan sipil dan badan pusat statistik. Jenis pengambilan data yang digunakan menggunakan pengambilan data sekunder. metode yang digunakan penulis dalam penelitian ini menggunakan metode deskriptif kualitatif . metode analisis kebutuhan data yang digunakan menggunakan tahapan knowledge discovery-in data base. Berdasarkan hasil penelitian dan analisis yang telah dilakukan penulis dapat di simpulkan bahwa bertambah nya jumlah penduduk berdasarkan pekerjaan mengalami sedikit peningkatan. Hal ini dapat di lihat dari data hasil yang di keluarkan dengan klasifikasi masing-masing per kecamatan yang berbentuk pohon keputusan yang di bantu dengan software weka.

**Kata Kunci:** *Penduduk, Klasifikasi, Algoritma C4.5.*

### Abstract

This study aims to classify the population based on occupation from 2018 to 2020 at the Central Statistics Agency and the Population Service for Civil Registration in Prabumulih City by applying the c4.5 algorithm. The author limits the problem to only providing information on population data based on occupation. The data in this study are population data sets taken directly from the Population and Civil Registration Service and the Central Statistics Agency. The type of data collection used is secondary data collection. The method used by the author in this study uses a qualitative descriptive method. The data requirements analysis method used is the knowledge discovery-in database stage. Based on the results of research and analysis that has been carried out by the author, it can be concluded that the increase in the number of residents based on work has slightly increased. This can be seen from the result data that is issued with the classification of each per sub-district in the form of a decision tree which is assisted by the Weka software.

**Keywords:** *Population, Classification, Algorithm C4.5.*

### PENDAHULUAN

Perkembangan teknologi informasi saat ini sudah semakin berkembang kebutuhan akan sebuah informasi yang berkualitas sangatlah diperlukan dalam perkembangan teknologi, banyak mempengaruhi tatanan hidup atau sebuah aturan dan system tertentu dan dapat di manfaatkan di berbagai bidang salah satunya dalam bidang pencatatan data penduduk.

Teknologi informasi adalah suatu perangkat yang di gunakan seseorang untuk mengolah data, mendapat informasi dan sebagainya ini sangat berpengaruh di lingkungan masyarakat serta memberikan dampak yang negatif dan positif pula tergantung tergantung bagaimana seseorang menggunakannya begitupula dengan masyarakat yang berada dilingkungannya. (Novi Yona Sidratul Munti & Dwi Asril Syaifuddin, 2020)

Pada sebuah layanan informasi untuk pengolahan data instansi yang terkait memiliki subsistem dari informasi manajemen yang di gunakan untuk menyelesaikan masalah dari penyediaan informasi dan pelayanan informasi

mengenai jumlah penduduk di kota prabumulih berdasarkan pekerjaan.

Menurut dewi, dkk (2020) dampak pandemi covid ini tidak hanya melemahkan pertumbuhan ekonomi tetapi juga menambah jumlah pengangguran terbuka yang disebabkan adanya pemutusan hubungan kerja (PHK) yang semakin meluas di sejumlah sektor.

Untuk mengetahui jumlah penduduk yang statusnya bekerja yang disebabkan oleh pandemi covid-19 penulis tertarik untuk mencari informasi yang dapat berdasarkan data pertahun, data yang dapat pada instansi yang berkaitan penulis bisa mengambil sebuah keputusan yang tepat dalam menentukan berapa jumlah penduduk berdasarkan pekerjaan yang dapat dari instansi yang berkaitan.

Ketersediaan data yang melimpah, kebutuhan akan sebuah informasi yang akurat sebagai pendukung pengambilan keputusan untuk mengurangi kesalahan dalam perhitungan, dan dukungan di bidang infrastruktur teknologi informasi merupakan awal terbentuknya sebuah teknologi data mining. data mining yang dimaksud adalah untuk memberikan informasi data jumlah penduduk yang telah di data pada dinas kependudukan dan pencatatan sipil kota prabumulih menggunakan algoritma C4.5.

Menurut Hermawati (2013:3), data mining adalah suatu percobaan yang memanfaatkan lebih dari satu prosedur studi (*knowledge engineering*) untuk menyelidiki dan menarik pengetahuan (*know-how*) secara tidak sadar. (Sari et al., 2022)

Data mining di sisi lain adalah kegiatan meliputi pengumpulan, pemakaian pengumpulan dan histories untuk menentukan keteraturan, pola atau hubungan dalam set data berukuran besar. (Swastina, 2018)

Sebagai alat analisis, berikut ini ada kelebihan dan kekurangan data mining, yaitu: (Harmizi, 1978)

#### 1. Kelebihan Data Mining

- a. Data mining mampu menangani data dalam jumlah besar dan kompleks.
- b. Data mining dapat menangani data dengan berbagai macam tipe atribut.
- c. Data mining mampu mencari dan mengolah data secara semi-otomatis. Disebut semiotomatis karena dalam beberapa teknik data mining, diperlukan parameter yang harus di-input oleh user secara manual.
- d. Data mining dapat menggunakan pengalaman ataupun kesalahan terdahulu untuk meningkatkan kualitas dan hasil analisa sehingga mendapat hasil yang terbaik

#### 2. Kekurangan Data Mining

- a. Terkadang ada data yang lebih mudah diselesaikan dengan statistik manual dibandingkan dengan menggunakan teknik data mining.
- b. Data mining tidak dapat menemukan pengetahuan secara instant.
- c. Algoritma data mining cukup kompleks.
- d. Hasil dari data mining tidak dapat langsung digunakan, harus dianalisa dan diinterpretasi terlebih dahulu.

Algoritma C4.5 merupakan salah satu algoritma yang telah secara luas di gunakan khususnya di area machine learning yang memiliki beberapa perbaikan dari algoritma sebelumnya yaitu ID3, algoritma C4.5 dan ID3 model yang tak terpisahkan, karena membangun sebuah pohon keputusan. (Ardiansyah & Walim, 2018). Algoritma C4.5 merupakan kelompok algoritma decision tree. Algoritma ini mempunyai input berupa training samples dan samples. (Salsabilla, 2022)

## METODE

Jenis penelitian yang di gunakan dalam penelitian ini adalah metode deskriptif kualitatif, metode deskriptif kualitatif dapat di artikan prosedur-prosedur pemecahan masalah yang di selidiki dengan menggambarkan keadaan subjek atau objek dalam penelitian dapat berupa orang, lembaga, masyarakat dan yang lainnya yang pada saat sekarang berdasarkan fakta-fakta yang tampak atau apadanya. (Rizmayanti et al., 2021). Penelitian ini menggunakan teknik pengumpulan data sekunder, yaitu data-data yang akan menjadi sebuah bahan penelitian yang di peroleh langsung dari tempat penelitian pada dinas kependudukan dan pencatatan sipil kota prabumulih, yang dimana data-data tersebutlah yang akan menjadi bahan sebuah penelitian untuk di teliti. Data sekunder adalah data yang dikumpulkan melalui pihak kedua, berupa bukti, catatan atau laporan yang diperoleh. (Bukanaung et al., 2022)

Berikut adalah langkah-langkah metode pengambilan data yang di lakukan sebagai berikut :

1. Wawancara : pengumpulan data yang di lakukan dengan proses Tanya jawab
2. observasi : mengumpulkan data dengan cara mengadakan tinjauan secara langsung ke objek yang di teliti
3. studi pustaka : untuk mendapatkan data-data yang bersifat teoritis maka penulis melakukan pengumpulan

data dengan cara membaca dan mempelajari buku-buku, jurnal penelitian, makalah ataupun referensi lain yang berhubungan dengan masalah yang akan di bahas. (Alhapizi et al., 2020)

Teknik analisis data mining atau pengolahan data dalam penelitian ini menggunakan *knowledge discovery-in database* (KDD) yang di bantu dengan menggunakan software WEKA.

Untuk mempermudah penulis sehingga mendapatkan hasil yang akurat, adapun untuk menganalisis data mining ini ada beberapa tahapan yaitu sebagai berikut :

1. data selection (menciptakan himpunan data target)
2. cleaning (membuang duplikasi data atau pembersihan data)
3. transformation (pola yang akan di cari dalam basis data)
4. data mining (mencari pola atau informasi menarik dalam data terpilih dengan menggunakan teknik atau metode tertentu)

5. Evaluation (peroses pemeriksaan) **HASIL DAN PEMBAHASAN**

Setelah pengumpulan data yang telah di lakukan, maka langkah selanjutnya yang dilakukan adalah ke tahap pengolahan data yang telah terkumpul. Penelitian ini di lakukan pada badan pusat statistic dan dinas kependudukan dan pencatatan sipil kota prabumulih yang bertempat di jalan raya lintas timur Palembang – prabumulih No. KM 12, prabumulih, prabumulih barat kota prabumulih, sumatera selatan. Penelitian ini di lakukan pada bu;an November 2021 hingga bulan april 2022.

Dalam penelitian ini memiliki tahapan-tahapan analisis data mining yaitu data seleksi, pembersihan data, transformasi data, data mining, evaluasi. Seleksi data menciptakan suatu himpunan data dan data target, setelah dilakukan lanjut ke tahap permbersihan data untuk membuang duplikat data lalu bagian transformasi untuk mencari pola yang akan di cari dalam basis data dari excel ke format CSV, tahap data mining mencari informasi menarik dalam data terpilih dengan menggunakan teknik atau metode tertentu disini penuis menggunakan algoritma C4.5 dengan metode klasifikasi, tahap evcaluasi akhir dari peroses data mining.

1. *Data selection*

Dibawah ini merupakan seleksi data yang telah dilakukan karena tidak semua attribute pada sumber data dapat di gunakan seluruhnya sumber data yang di gunakan berasal dari data penduduk berdasarkan pekerjaan.

Table 1.Struktur tabel Variabel

No	Nama field	Type data	Field size	Keterangan
1	No_data	Varchar	16	Nomor Data
2	Nama_akte	Varchar	20	Nama Akte
3	Anak_KE	Int	2	Anak ke berapa
4	Tgl_AKTE	Date	-	Tanggal akte
5	Nama_p	Year	4	Nama pemilik
6	Alamat	Varchar	20	Alamat
7	Kec_p	Varchar	20	Kecamatan pemilik
8	Desa_p	Varchar	20	Desa pemilik
9	Rt	Varchar	2	RT
10	Rw	Varchar	3	RW
11	Tempat_lahir	Varchar	20	Tempat lahir
12	Tgl_lahir	Date	-	Tanggal lahir

13	Jenis_kelamin	Varchar	10	Jenis kelamin
14	Pekerjaan	Varchar	10	Pekerjaan
15	Pendidikan	Varchar	10	Pendidikan
16	Status_p	Varchar	10	Pendidikan
17	Kewarganegaraa n	Varchar	20	Kewarga negeraan

Dari 17 atribut tersebut kemudian di seleksi menjadi 8 atribut berdasarkan pekerjaan masing-masing.

Table 2 Struktur variable yang telah di seleksi

NO	Atribut	Type	Field	Keterangan
1	No_data	Varchar	20	Nomor data
2	Nm_p	Varchar	20	Nama pemilik
3	Alamat	Varchar	20	Alamat
4	Kec_p	Varchar	20	Kecamatan pemilik
5	pekerjaan	varchar	20	Pekerjaan

Tabel di atas menghasilkan himpunan data target pemilihan himpunan data, atau memfokuskan pada sub set variable atau sample data.

## 2. Data cleaning

Pembersihan data terhadap noise yang tidak di temukan berupa *missing value*, dan *redundan* data.

Table 3 pekerja tahun 2020

NO	KECAMATAN	Pekerjaan			
		PNS	TNI / POLRI	PARA MEDIS	TOTAL
1	2	3	4	5	6
1.	PRABUMULIH BARAT	550	173	117	840
2.	PRABUMULIH TIMUR	2.672	987	435	4.094
3	CAMBAI	268	62	53	383
4	RAMBANG KAPAK TENGAH	143	33	39	215
5	PRABUMULIH UTARA	548	101	104	753
6	PRABUMULIH SELATAN	462	100	92	654
<b>JUMLAH</b>		<b>4.643</b>	<b>1.456</b>	<b>840</b>	<b>6.939</b>

Table 4 pekerja tahun 2020

NO	KECAMATAN	Pekerjaan			TOTAL
		Petani	Buruh	Pegawai swasta	
1	2	3	4	5	6
1.	PRABUMULIH BARAT	3.274	2.508	1.325	7.107
2.	PRABUMULIH TIMUR	2.194	4.807	4.357	11.358
3	CAMBAI	4.477	603	341	5.421
4	RAMBANG KAPAK TENGAH	2.693	148	196	3.037
5	PRABUMULIH UTARA	273	3.619	1.371	5.263
6	PRABUMULIH SELATAN	1.220	1.718	762	3.700
<b>JUMLAH</b>		<b>14.131</b>	<b>13.403</b>	<b>8.3522</b>	<b>35.886</b>

Table 5 pekerja tahun 2020

NO	KECAMATAN	Pekerjaan		TOTAL
		Pedagang	Nelayan	
1	2	3	4	6
1.	PRABUMULIH BARAT	91	1	92
2.	PRABUMULIH TIMUR	829	1	830
3	CAMBAI	103	0	103
4	RAMBANG KAPAK TENGAH	19	-	19
5	PRABUMULIH UTARA	676	-	676
6	PRABUMULIH SELATAN	77	0	77
<b>JUMLAH</b>		<b>1.795</b>	<b>2</b>	<b>1.797</b>

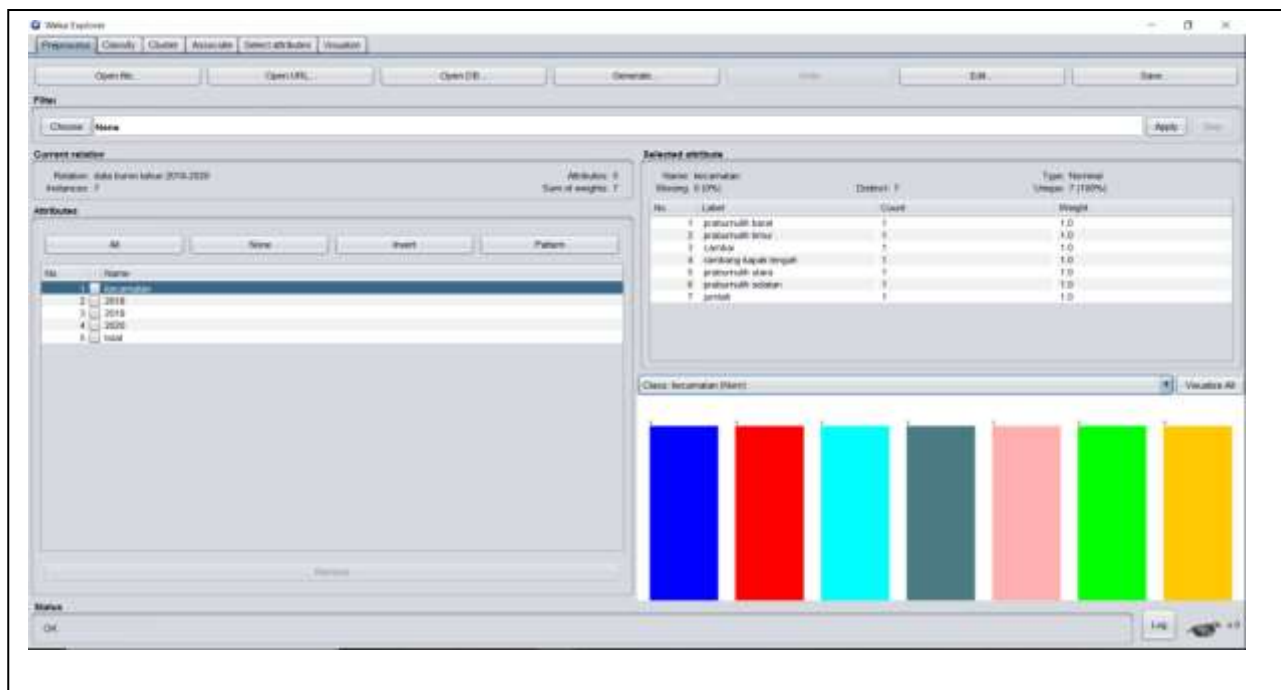
### 3. Transformasi data

Tahapan merubah data ke dalam bentuk yang sesuai untuk di mining tahapan ini juga sesuai untuk membentuk format data yang di terima di perangkat lunak data mining yang akan memproses nya.



Gambar 1 data set yang di ubah dari excel ke CSV

#### 4. Data Mining

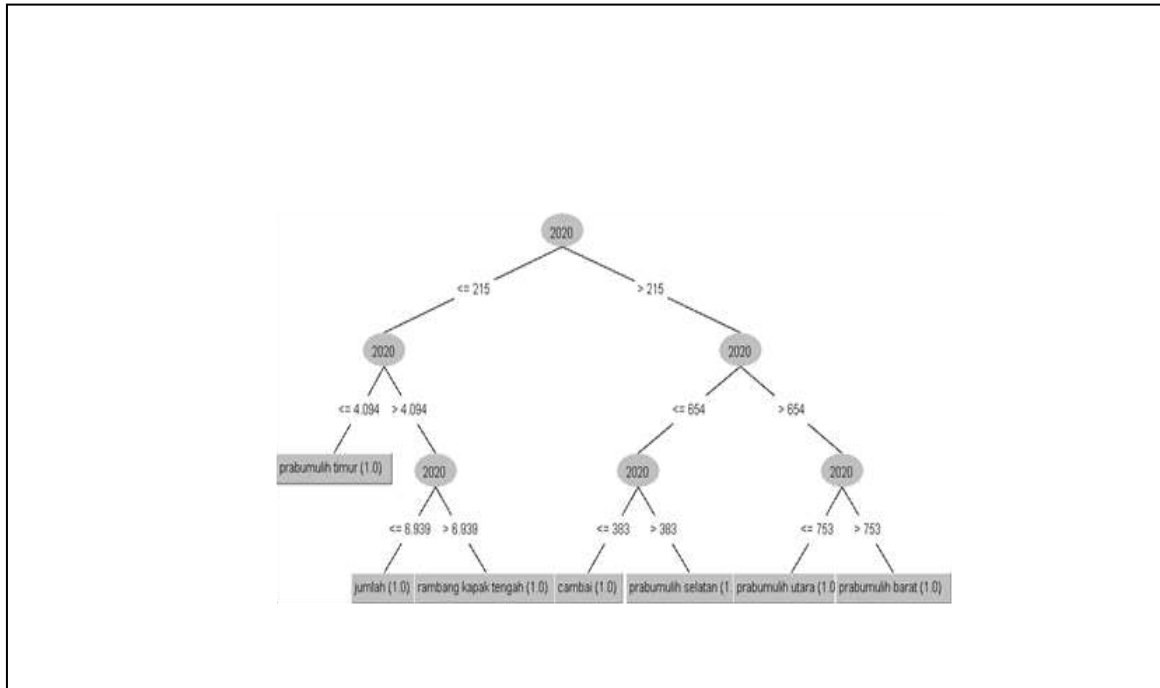


Gambar 2 tampilan awal kalsifikasi pekerja

Setelah melakukan transformasi data berikutnya adalah proses mining data, proses yang dilakukan untuk mencari pola atau mencari nilai yang unik dari suatu data. Berdasarkan gambar di atas dapat dilihat bahwa atribut yang mampu dibaca berjumlah 5 atribut dengan total data sebanyak 5. Data di atas siap di proses menggunakan weka karena attribute yang lain dianggap tidak penting dan jika tetap di pertahan kan attribute-attribute tersebut tidak akan memberikan informasi yang berguna.

#### 5. Evaluation

Tahap akhir dari analisis data mining. jika ada eror data atau lain sebagai nya di tahap ini akan mengevaluasi kesalahan-kesalahan data dan bisa diperbaiki atau memberika informasi



Gambar 3 pohon keputusan penduduk bekerja

1. Prabumulih timur = 4.049 orang yang bekerja sebagai pns, polisi dan tni
2. Prabumulih selatan = 654 orang yang bekerja sebagai pns, polisi dan tni
3. Prabumulih utara = 753 orang yang bekerja sebagai pns, polisi dan tni
4. Rambang kapak tengah = 215 orang yang bekerja sebagai pns, polisi dan tni
5. Cambia = 383 orang yang bekerja sebagai pns, polisi dan tni

Jumlah pekerja tahun 2020 adalah sebanyak 6.939 orang yang bekerja dan sisanya 840 merupakan pekerja dari prabumulih barat. Dari data tersebut dapat disimpulkan bahwa pekerja terbanyak di kecamatan prabumulih timur sebanyak 4.094 orang sebagai pns, polisi dan TNI

## SIMPULAN

Berdasarkan analisis data yang telah dilakukan maka dapat disimpulkan bahwa jumlah pekerja terbanyak ada pada prabumulih timur sebagai PNS, Polisi dan TNI dengan total penduduk yang bekerja sebanyak 4.094 sedangkan penduduk yang bekerja sebagai PNS, Polisi dan TNI dengan pekerja terendah pada kecamatan rambang kapak tengah berjumlah 215 orang sedangkan untuk mengetahui pekerja dari prabumulih barat dengan cara menghitung jumlah pekerja di kurangi dengan perkecamatan masing-masing sehingga total pekerja prabumulih barat dapat diketahui.

## DAFTAR PUSTAKA

- Alhapizi, M. R., Nasir, M., & Effendy, I. (2020). Penerapan Data Mining Menggunakan Algoritma K-Means Clustering Untuk Menentukan Strategi Promosi Mahasiswa Baru Universitas Bina Darma Palembang. *Journal of Software Engineering Ampera*, 1(1), 1–14. <https://doi.org/10.51519/journalsea.v1i1.10>
- Ardiansyah, D., & Walim, W. (2018). Algoritma c4.5 untuk klasifikasi calon peserta lomba cerdas cermat siswa smp dengan menggunakan aplikasi rapid miner. *Jurnal Inkofar*, 1(2), 5–12.
- Bukanaung, N. M., Rantung, S. V., Tambani, G. O., Sondakh, S. J., & Kotambunan, O. V. (2022). PELABUHAN PERIKANAN PANTAI TUMUMPA KOTA MANADO Latar Belakang yang harus dapat menjamin suksesnya aktivitas usaha perikanan . *Pelabuhan perikanan Tumumpa II Kecamatan Tuminting Kota Manado , Provinsi Sulawesi Utara . Letak Sulawesi Utara . proses untuk menerjemahkan stimulus yang masuk ke dalam alat indera manusia Rumusan Masalah “ Bagaimana persepsi pedagang perantara ikan terhadap layanan yang disediakan oleh Pelabuhan Perikanan Pantai Tumumpa Manad o .”* 10(2), 478–489.
- Harmizi, M. (1978). *Estimasi Data Mining Tingkat Kecelakaan Lalu Lintas Kabupaten Serdang Bedagai Menggunakan Algoritma Regresi Linier Berganda*. x.
- Novi Yona Sidratul Munti, & Dwi Asril Syaifuddin. (2020). Analisa Dampak Perkembangan Teknologi Informasi Dan

Komunikasi Dalam Bidang Pendidikan. *Jurnal Pendidikan Tambusai*, 4(2), 1799–1805.

- Rizmayanti, A. I., Hidayati, N., Nugraha, F. S., & Gata, W. (2021). Penerapan Data Mining Untuk Memprediksi Kompetensi Siswa Menggunakan Metode Decission Tree ( Studi Kasus Smk Multicomp Depok ). *Swabumi*, 9(1), 9–18. <https://doi.org/10.31294/swabumi.v9i1.8363>
- Salsabilla, T. (2022). *Implementasi algoritma c4.5 untuk klasifikasi produk laris sepeda motor honda pada cv cendana motor cepiring 1) 1,2). 7(2)*, 164–171.
- Sari, D. N., Judul, H., Manajemen, P. S., Ekonomi, F., Bisnis, D. A. N., & Surakarta, U. M. (2022). *SAAT PANDEMI COVID-19 DENGAN METODE CAMEL ( Kasus Pada Bank BUMN Yang Terdaftar Di Bursa Efek Indonesia )*.
- Swastina, L. (2018). Penerapan Algoritma C4 . 5 Untuk Penentuan Jurusan Mahasiswa. *Gema Aktualita*, 2(1), 93–98.