

Penerapan Algoritma C4.5 Dalam Memprediksi Pembelian Rokok Elektrik Pada Toko Silvape

Muhammad Gumilang Adha Thosca¹

¹ Fakultas Ilmu Komputer, Program Studi Teknik Informatika, Universitas Pamulang, Tangerang Selatan, Indonesia

E-mail: muhammadgumilangat@gmail.com

Abstrak- Toko Silvape adalah salah satu Toko rokok elektrik yang berada di Tangerang Selatan. Jika prediksi minat pelanggan tidak diperhatikan dengan baik maka bisa sangat merugikan bisnis penjualan. Prediksi minat pelanggan saat ini menjadi faktor yang sangat berpengaruh pada perkembangan suatu usaha, jika suatu pelaku bisnis bisa memprediksi pada pasar maka akan mudah baginya untuk memenuhi kebutuhan pelanggan. Berdasarkan latar belakang tersebut maka dengan menerapkan teknik data mining diharapkan dapat membantu Toko Silvape untuk memprediksi minat pelanggan.

Kata Kunci: Data Mining, Prediksi, Penjualan, Algoritma C4.5, Klasifikasi, RapidMiner

Abstract- Silvape shop is one of the e-cigarette shops located in South Tangerang. If the prediction of customer interest is not considered properly, it can be very detrimental to the sales business. Prediction of customer interest is currently a very influential factor in the development of a business, if a business person can predict the market it will be easy for him to meet customer needs. Based on this background, applying data mining techniques is expected to help Silvape Stores predict customer interest.

Keywords: Data Mining, Prediction, Sales, Algorithm C4.5, Classification, RapidMiner

1. PENDAHULUAN

Saat ini tingkat persaingan perdagangan di Indonesia semakin tinggi, termasuk persaingan pada industri rokok di Indonesia. Semakin pesatnya perkembangan industri rokok elektrik (vapor), membuat usahawan harus menentukan strategi pemasaran yang tepat agar mampu bertahan dan memenangkan persaingan. Perilaku merokok dilihat dari berbagai sudut pandang sangat merugikan, baik untuk diri sendiri maupun orang di sekelilingnya. Dilihat dari sisi individu yang bersangkutan, perilaku merokok sebagai bagian dari gaya hidup dan kebutuhan, serta melihat adanya bahaya dari merokok yang cukup tinggi, membuat munculnya inovasi teknologi dalam hal merokok dengan produk rokok elektrik atau vaporizer elektrik yang biasa disebut vape.

Adanya produk vape saat ini menyebabkan munculnya tren di berbagai kota di Indonesia yang disebut “vaping”. Nama Vaping sendiri berasal dari kata Vapor yang berarti uap. Cara kerja rokok elektrik sendiri adalah mengubah cairan menjadi uap yang menyerupai asap seperti sedang merokok. Meskipun menuai pro dan kontra, namun vaping semakin menyebar dan banyak disukai oleh masyarakat karena vaping sendiri disebut-sebut sebagai alternatif merokok yang lebih sehat. Vaping dapat digunakan sebagai media bagi para perokok aktif untuk perlahan-lahan berhenti merokok. Sensasi yang diberikan rokok elektrik hampir sama dengan rokok tembakau sehingga menjadi alternatif pengganti rokok tembakau. Karena mengandung bahan-bahan yang lebih aman dibandingkan dengan rokok tembakau, vaping dipromosikan lebih aman untuk kesehatan. Rokok elektronik juga menghasilkan uap air bukan asap sehingga aman untuk lingkungan sekitar. Dipercaya bahwa vape dapat menggantikan kebiasaan merokok yang tidak sehat dan berbahaya.

Toko Silvape merupakan suatu tempat penjualan macam-macam jenis rokok elektrik, berbagai macam bentuk dan berbagai macam model yang dijualnya, mulai dari yang Rebuildable Dripping Atomizer (RDA), Rebuildable Tank Atomizer (RTA), batre, kapas, coil sampai alat-alat bongkar pasangannya, toko silvape pun menjual barang dagangannya di marketplace seperti Tokopedia dan

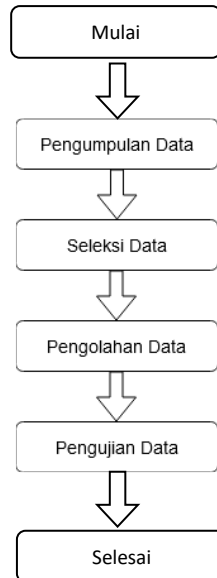
Shopee. Dengan harga jualnya yang cukup terjangkau, toko silvape pun memperoleh penjualan yang sangat banyak sehingga kehabisan stok barang. vaporizer toko silvape ini pun ingin mengetahui seberapa banyak penjualan di bulan berikutnya sehingga penjual dapat menyiapkan barang yang cukup untuk di jual di bulan berikutnya.

Salah satu wujud nyata dari upaya ini yaitu menghitung potensi pembelian seakurat mungkin sehingga akan diketahui seberapa besar pembelian pada toko silvape. Algoritma yang akan digunakan untuk memprediksi pembelian dalam kasus ini adalah algoritma C4.5, Metode ini akan berusaha mendapatkan prediksi pembelian dan dapat diketahui gambaran fluktuasi pembelian pada toko tersebut.

Tujuan penelitian ini adalah untuk menerapkan tingkat minat produk rokok elektrik mana yang banyak diminati pelanggan Toko Silvape dan untuk mengetahui tingkat akurasi dalam memprediksi pembelian rokok pada Toko Silvape dengan menggunakan algoritma C4.5.

2. METODE

Penerapan metodologi penelitian dilakukan dengan melakukan pengumpulan data, seleksi data, pengolahan data, serta pengujian data.



Gambar 1. Alur metodologi penelitian

1. **Pengumpulan Data**
 Pada tahap ini peneliti melakukan pengumpulan data. Pada bagian ini juga dijelaskan tentang bagaimana dan dari mana data dalam penelitian ini didapatkan, meliputi data kualitatif dan kuantitatif. Tahapan ini merupakan tahapan yang penting karena dapat berpengaruh terhadap hasil penelitian, sehingga dalam mengumpulkan data harus dilakukan dengan benar.
2. **Seleksi Data**
 Tahap ini dilakukan seleksi terhadap data database. Karena data yang diperoleh tidak semuanya digunakan, dipilih sesuai dengan atribut atau variabel yang dibutuhkan dalam penelitian dengan melakukan seleksi data sehingga menjadi dataset. ada tahap ini akan menghilangkan data yang null, data tidak valid, dan data yang ganda. karena data yang kosong ataupun tidak valid akan berpengaruh terhadap hasil yang diperoleh. Kemudian data akan diolah menggunakan Microsoft Office Excel 2016.

3. Metode yang Diusulkan
 Setelah semua data yang diperlukan telah dipilih, maka tahap penelitian selanjutnya adalah pengolahan data. Pada tahap ini akan dilakukan transformasi atau mengubah nilai atribut data ke dalam bentuk data yang sesuai agar data dapat diproses menggunakan algoritma C4.5. Sehingga akan diperoleh dataset utuh yang digunakan untuk proses ke tahapan selanjutnya.
4. Pengujian Data
 Pada tahap pengujian hasil akan dilakukan pengujian data baik secara manual dengan algoritma C4.5 dan menggunakan software RapidMiner.

3. ANALISA DAN PEMBAHASAN

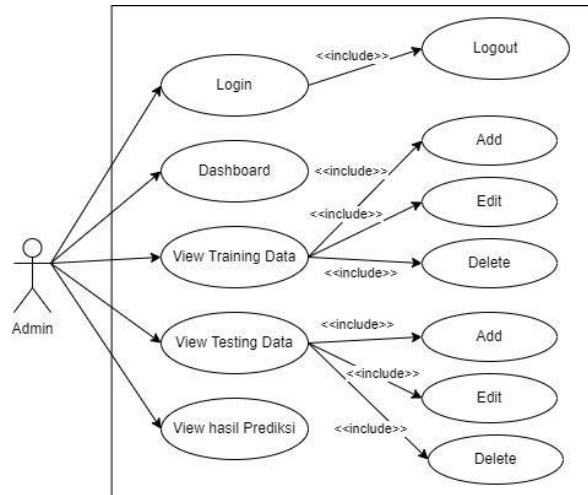
Hasil Penelitian Metode penelitian yang penulis lakukan menghasilkan informasi dan data yang digunakan sebagai acuan penulis sebagai dasar penelitian. Penelitian ini menggunakan sebagian data dari data penjualan Toko Silvape yaitu sebanyak 120 data untuk proses klasifikasi. Untuk melakukan pengolahan data di RapidMiner data yang digunakan adalah Data Training adalah 80% dari data keseluruhan populasi, sehingga total populasi data training berjumlah 96 data. 96 data tersebut untuk dijadikan data training, yang akan dipakai untuk membuat model atau rule. Selain itu juga Data testing adalah 20% dari data keseluruhan, total populasi data testing berjumlah 24 data yang akan dipakai untuk sample testing. Data training dan testing dipilih secara acak dan menghapus atribut nama barang, kategori, dan satuan karena atribut tersebut tidak berpengaruh saat pengolahan data di RapidMiner. Atribut yang dipakai hanya harga, ukuran, bentuk dan jumlah minat (banyak dan sedikit).

NO	ID	Nama Barang	Kategori	Satuan	Harga	Berat / gr	Ukuran	Bentuk	Penjualan	Jumlah Peminat
1	LQ-101	Krusty Crepes	Liquid	1	175.000	250	Berat	Besar	72	Banyak
2	LQ-102	Mines Two Zero	Liquid	1	200.000	250	Berat	Besar	50	Sedikit
3	LQ-103	Dalgona Coffe	Liquid	1	35.000	150	Ringan	Besar	322	Banyak
4	LQ-104	IceBerg	Liquid	1	120.000	150	Ringan	Besar	77	Banyak
5	LQ-105	Poppies By Tigac	Liquid	1	180.000	150	Ringan	Besar	87	Banyak
6	LQ-106	Coffe Delight	Liquid	1	130.000	150	Ringan	Besar	88	Banyak
7	LQ-107	Sunset Trip	Liquid	1	100.000	100	Ringan	Kecil	90	Banyak
8	LQ-108	King-Kong	Liquid	1	130.000	150	Ringan	Besar	75	Banyak
9	LQ-109	Paradewa Apple	Liquid	1	112.000	150	Ringan	Besar	83	Banyak
10	LQ-110	Banana Cake	Liquid	1	35.000	150	Ringan	Besar	200	Banyak
11	LQ-111	Bubblegum	Liquid	1	35.000	100	Ringan	Kecil	210	Banyak
12	LQ-112	Paradewa Mangga	Liquid	1	112.000	150	Ringan	Besar	85	Banyak
13	LQ-113	AlaCarte	Liquid	1	130.000	250	Berat	Besar	82	Banyak
14	LQ-114	Chocolate Cheese	Liquid	1	110.000	250	Berat	Besar	66	Banyak
15	LQ-115	Steezy	Liquid	1	120.000	150	Ringan	Besar	82	Banyak
16	LQ-116	Tokyo Banana	Liquid	1	180.000	150	Ringan	Besar	63	Sedikit
17	LQ-117	Extra Jon	Liquid	1	130.000	150	Ringan	Besar	78	Banyak
18	LQ-118	KW	Liquid	1	130.000	150	Ringan	Besar	67	Banyak
19	LQ-119	Munchies	Liquid	1	130.000	100	Ringan	Kecil	88	Banyak
20	LQ-120	Cushman	Liquid	1	125.000	150	Ringan	Besar	99	Banyak
21	LQ-121	Donat Paten	Liquid	1	140.000	100	Ringan	Kecil	89	Banyak
22	LQ-122	Strawberry Vanilla Crs	Liquid	1	250.000	150	Ringan	Besar	50	Sedikit
23	LQ-123	NYC Cheese	Liquid	1	145.000	100	Ringan	Kecil	57	Sedikit
24	LQ-124	Dessert Blubbery	Liquid	1	185.000	250	Berat	Besar	61	Sedikit

Gambar 2. Data Tarining

3.1. Use case diagram

Diagram *use case* menyajikan interaksi antara *use case* dan *actor*. Dimana aktor dapat berupa orang, peralatan atau sistem lain yang berinteraksi dengan sistem yang sedang dibangun. *Use case* pada penelitian ini dapat digambarkan sebagai berikut.



Gambar 3. Use case diagram

Tabel 1. Deskripsi Aktor

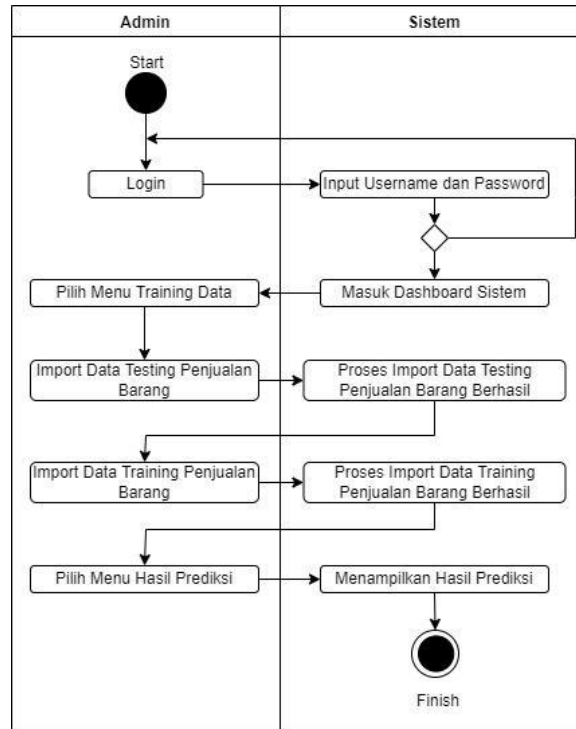
No	Nama Aktor	Deskripsi
1	Admin	Orang yang mengontrol <i>system</i> , yaitu mengedit, menambahkan dan menghapus data.

Tabel 2. Deskripsi Use Case

No	Nama Use Case	Aktivitas
1	<i>Login</i>	Melakukan <i>login</i> ke <i>system</i> yang menampilkan halaman untuk sistem
2	Melihat Data <i>Training</i>	Melihat Data <i>Training</i> , menambah Data <i>Training</i> , mengubah Data <i>training</i> dan menghapus data <i>training</i>
3	Melihat Data <i>Testing</i>	Melihat Data <i>Testing</i> , menambah Data <i>Testing</i> , mengubah Data <i>testing</i> dan menghapus data <i>testing</i>
4	Melihat Hasil Prediksi	Melihat hasil prediksi dari data <i>testing</i> yang diinput
5	<i>Log Out</i>	Keluar sistem

3.2. Activity diagram

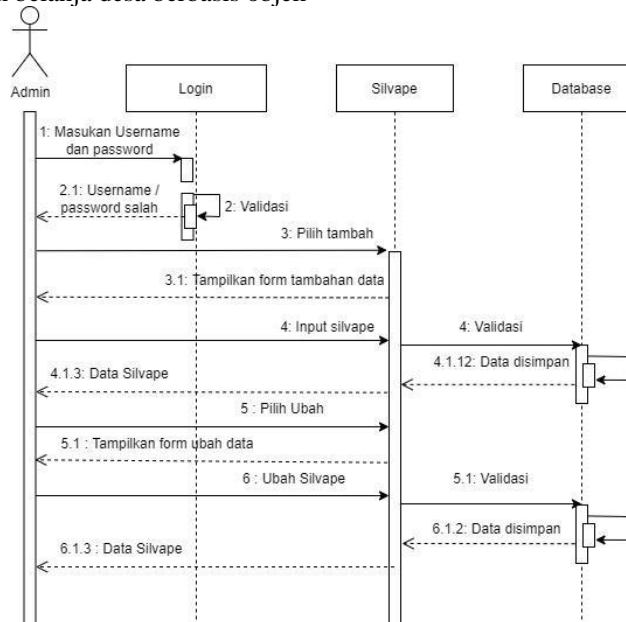
Activity diagram merupakan diagram yang menggambarkan berbagai alir aktivitas dalam suatu sistem, bagaimana masing-masing alir berawal, decision yang mungkin terjadi, dan bagaimana mereka berakhir. *Activity* diagram juga dapat menggambarkan proses paralel yang mungkin terjadi pada beberapa eksekusi. *Activity* diagram pada penelitian ini dapat digambarkan sebagai berikut.



Gambar 4. Activity diagram

3.3. Sequence diagram

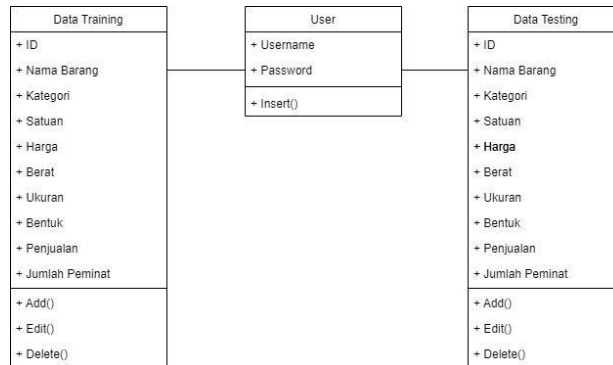
Sequence Diagram menggambarkan bagaimana sistem merespons kegiatan *user*. Sequence Diagram yang dibuat yaitu yang berhubungan langsung dengan kegiatan utama dari sistem informasi anggaran pendapatan dan belanja desa berbasis objek



Gambar 5. Sequence diagram

3.4. Class diagram

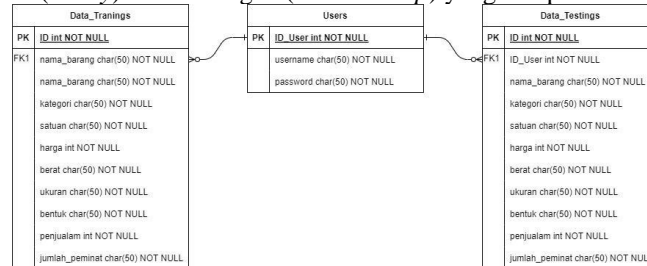
Class adalah deskripsi sekelompok objek dari property (atribut), sifat (operasi), relasi antar objek dan semantik yang umum. *Class* merupakan template untuk membentuk objek. Diagram *class* merupakan diagram yang terdiri dari sekumpulan objek yang memiliki atribut-atribut dan method. *Class* Diagram menggambarkan struktur dan deskripsi *class*, *package* dan objek serta hubungan satu sama lain.



Gambar 6. Class diagram

3.5. ERD

Entity Relationship Diagram adalah alat pemodelan data utama dan akan membantu mengorganisasi data dalam suatu proyek ke dalam suatu proyek ke dalam entitas-entitas dan menentukan hubungan antarentitas. Didalam hubungan ini tersebut dinyatakan yang utama dari ERD adalah menunjukkan objek data (*Entity*) dan hubungan (*Relationship*) yang ada pada *Entity* berikutnya.



Gambar 7. Entity Relationship Diagram

Sedangkan hasil pengukuran akurasi data yang diperoleh dari data training nilainya mencapai 97.22%. Dari tabel tersebut diketahui prediksi naik dengan *true* naik mencapai 19 barang dan *true* turun sebanyak 0 barang, dengan hasil pencapaian nilai presisi sebesar 100.00%. Sedangkan untuk prediksi penjualan barang turun untuk *true* naik mencapai 1 barang dan untuk *true* turun terdapat 16 barang dengan hasil pencapaian nilai presisi sebesar 94.12%. Untuk *recall* data barang yang sangat tinggi terdapat pada *class recall tue* naik mencapai 95.00% sedangkan untuk *class recall true* turun mencapai 100.00%.

1. Accuracy (Data Training)

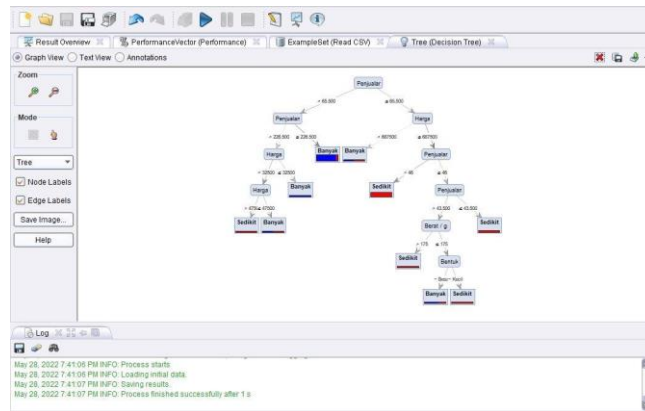
Accuracy didefinisikan sebagai tingkat kedekatan antara nilai prediksi dengan nilai aktual. Perhitungan akurasi dilakukan dengan cara membagi jumlah data yang diklasifikasi secara benar dengan total data sampel yang diuji.

2. Recall Data (Data Training)

Recall dihitung dengan cara membagi data benar yang bernilai positif (*true positif*) dengan hasil penjumlahan dari data benar yang bernilai positif (*true positif*) dan data salah yang bernilai negatif (*false negative*).

3. *Precision (Data Testing)*

Nilai *Precision* dihitung dengan cara membagi data benar yang bernilai positif (*true positif*) dibagi dengan jumlah data yang bernilai positif (*true positif*) dan data salah yang bernilai (*false positif*).



Gambar 8. *Precision (Data Testing)*

4. KESIMPULAN

Berdasarkan hasil evaluasi dan validasi hasil serta uji coba pada penelitian penulis dapat menyimpulkan bahwa:

- Dengan menerapkan metode algoritma C4.5, yaitu melakukan prediksi penjualan rokok elektrik menggunakan dapat memudahkan pemilik dan karyawan toko untuk melakukan proses prediksi produk secara sistematis dan efisien.
- Dengan menggunakan metode algoritma C4.5 menghasilkan akurasi yang cukup tinggi. Keakuratannya diuji menggunakan 120 data penjualan rokok elektrik. Pengujian dilakukan dengan menghitung nilai Confusion Matrix, maka nilai accuracy didapat sebesar 97.22%, sehingga algoritma C4.5 sangat efektif dalam memprediksi produk terlaris Toko Silvape dan dapat memaksimalkan peluang peminatan pelanggan untuk kemajuan tokonya.
- Dalam upaya implementasi pengembangannya system dibuat berbasis website, dengan adanya sistem berbasis web memberikan kemudahan bagi penjual toko silvape dalam menentukan penjualan produk terlaris pada toko silvape.

REFERENSI

- Anastasia C. R., dan Kemal A. S., (2020). Penerapan Data Mining Dengan Menggunakan Algoritma C4.5 Pada Klasifikasi Fasilitas Kesehatan Provinsi Di Indonesia, Jurnal Ilmiah KOMPUTASI 19(1). <https://ejournal.jak-stik.ac.id/index.php/komputasi/article/view/153/46>
- Erfan K, Agus S, dan Nadia H., (2021). PENERAPAN ALGORITMA DECISION TREE C4.5 BERBASIS PARTICLE SWARM OPTIMIZATION (PSO) UNTUK PREDIKSI TINGKAT KEPUASAN PELAYANAN OBAT, Jurnal Ilmiah Technologia 12(2). https://web.archive.org/web/20170520084805id_/http://ejournal.stikombali.ac.id:80/index.php/knsi/article/download/527/179
- Fitriana H., (2018). Penerapan data Mining dalam Pemilihan Mobil menggunakan Algoritma C4.5. Jurnal Konferensi Nasional Sistem & Informatika, 01(1). <https://vo.stmik-tasikmalaya.ac.id/index.php/vo/article/download/99/42>
- Heru S, & Qadafi K., (2017). Perancangan Sistem Informasi Manajemen Proyek Sistem Informasi Kontraktor. Jurnal Khatulistiwa 05(2).

[https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=&cad=rja&uact=8&ved=2ahUKEwiOxd-](https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=&cad=rja&uact=8&ved=2ahUKEwiOxd-SnOv5AhXpSGwGHUJeAg0QFnoEAcQAQ&url=https%3A%2F%2Fjournal.bsi.ac.id%2Fejournal%2Findex.php%2Fkhatulistiwa%2Farticle%2Fdownload%2F2886%2F1869&usg=AOvVaw2q-dGPJePqeb25F4oZ-CVN)

[SnOv5AhXpSGwGHUJeAg0QFnoEAcQAQ&url=https%3A%2F%2Fjournal.bsi.ac.id%2Fejournal%2Findex.php%2Fkhatulistiwa%2Farticle%2Fdownload%2F2886%2F1869&usg=AOvVaw2q-dGPJePqeb25F4oZ-CVN](https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=&cad=rja&uact=8&ved=2ahUKEwiOxd-SnOv5AhXpSGwGHUJeAg0QFnoEAcQAQ&url=https%3A%2F%2Fjournal.bsi.ac.id%2Fejournal%2Findex.php%2Fkhatulistiwa%2Farticle%2Fdownload%2F2886%2F1869&usg=AOvVaw2q-dGPJePqeb25F4oZ-CVN)

- Syaiful B, Akhyar L., (2020). Metode Klasifikasi Decision Tree Untuk Memprediksi Juara English Premier League. Jurnal Sintaksis: Pendidikan Guru Sekolah Dasar, IPA, IPS dan Bahasa Inggris,02(1). <https://jurnal.stkipalmaksum.ac.id/index.php/Sintaksis/article/download/47/57>
- Purwadi., (2018). Implementasi Data Mining Untuk Memprediksi Pola Pembelian Sepeda Motor Pada Showroom CV. Viva Mas Motors Dengan Metode Algoritma C4.5. Jurnal Sistem Informasi Kaputama, 02(2). <http://jurnal.kaputama.ac.id/index.php/JSIK/article/download/106/pdf>
- Rika, H., & Amrizal. (2021). Analisis Algoritma C4.5 Untuk Menentukan Faktor Pembelian Sepeda Bekas Pada Toko Sepeda Batam. Jurnal CBIS, 04(2). <https://mail.pbtv.co.id/index.php/cbis/article/view/4473/2087>
- Tri, B.T. (2018) Penerapan Data Mining Dengan Algoritma C4.5 Dalam Prediksi Penjualan, Skripsi, Universitas Pelita Bangsa, Indonesia. <https://ecampus.pelitabangsa.ac.id/pb/AmbilLampiran?ref=22987&jurusan=&jenis=Item&usingId=false&download=false&clazz=ais.database.model.file.LampiranLain>
- Widiya, I. (2020) Penerapan Data Mining Dengan Algoritma C4.5 Dalam Prediksi Penjualan pada toko Deshop Cikarang, Skripsi, Universitas Pelita Bangsa, Indonesia.