

PENERAPAN ALGORITMA NAÏVE BAYES UNTUK REKOMENDASI GENSET

Chandra Wijaya¹, Alyauma Hajjah²

^{1,2}Institut Bisnis dan Teknologi Pelita Indonesia, Jl. Ahmad Yani No. 78-88, Pekanbaru

E-mail: chandrawijaya@student.pelitaindonesia.ac.id¹,

alyauma.hajjah@lecturer.pelitaindonesia.ac.id²

ABSTRACT

At this time, technology is an important driver for a country. Technological developments in the economic field are marked by the emergence of online shops. PT. Yanmarindo Perkasa is a company engaged in the sale and purchase of machines, tools, spare parts and others. The goods sold are very diverse, causing confusion for consumers in choosing the goods they want to buy because goods of the same type have different types. To overcome this, a recommendation system is needed to provide recommendations for goods according to the criteria and needs of consumers. This recommendation system was built using the Naïve Bayes algorithm as a part of data mining. The Naïve Bayes algorithm is a classification method based on probability and statistics. This recommendation system is limited to generator engines where the training data used is a combination of specification data and generator sales data. Naïve Bayes algorithm is used to find the greatest probability of all instances on target attributes such as brand, fuel, capacity, voltage and ignition. From one of the sample test data entered, there is one generator that appears as a recommendation from the calculation of the Naïve Bayes algorithm. Based on this, it is proven that the Naïve Bayes algorithm can be used to provide generator recommendations according to the needs and criteria of consumers. With the application of this algorithm on the generator recommendation system, it is expected to provide accurate results and reduce consumer anxiety in finding generators.

Keywords: Recommendations, Criteria, Naïve Bayes, Generators

ABSTRAK

Pada saat ini, teknologi merupakan penggerak penting bagi suatu negara. Perkembangan teknologi di bidang ekonomi ditandai dengan munculnya *online shop*. PT. Yanmarindo Perkasa merupakan perusahaan yang bergerak di bidang jual beli mesin, perkakas, *sparepart* dan lain-lain. Barang yang dijual sangat beragam menyebabkan bingungnya konsumen dalam memilih barang yang ingin dibeli karena barang dengan jenis yang sama memiliki tipe berbeda. Untuk mengatasi hal tersebut, dibutuhkan suatu sistem rekomendasi untuk memberikan rekomendasi barang yang sesuai dengan kriteria dan kebutuhan konsumen. Sistem rekomendasi ini dibangun menggunakan algoritma *Naïve Bayes* sebagai salah satu bagian dari *data mining*. Algoritma *Naïve Bayes* adalah metode klasifikasi yang didasarkan pada probabilitas dan statistik. Sistem rekomendasi ini terbatas pada mesin genset dimana data *training* yang digunakan yaitu kombinasi dari data spesifikasi dan data penjualan genset. Algoritma *Naïve Bayes* digunakan untuk mencari probabilitas terbesar dari seluruh *instance* pada atribut target seperti merek, bahan bakar, kapasitas, tegangan dan penyalan. Dari salah satu contoh data uji yang dimasukkan, terdapat satu genset yang tampil sebagai hasil rekomendasi dari perhitungan algoritma *Naïve Bayes*. Berdasarkan hal tersebut, terbukti bahwa algoritma *Naïve Bayes* dapat digunakan untuk memberikan rekomendasi genset sesuai dengan kebutuhan dan kriteria konsumen. Dengan penerapan algoritma ini pada sistem rekomendasi genset, diharapkan dapat memberikan hasil yang akurat dan mengurangi kebimbangan konsumen dalam mencari genset.

Kata kunci: Rekomendasi, Kriteria, Naïve Bayes, Genset

1. PENDAHULUAN

Zaman sekarang ini, teknologi merupakan salah satu penggerak yang sangat penting bagi suatu negara. Pada bidang komunikasi, terlihat dari bentuk komunikasi yang sederhana sampai komunikasi elektronik [1]. Di bidang ekonomi, muncul berbagai perusahaan baru khususnya di bidang teknologi, *startup* menjadi salah satu jenis perusahaan yang paling banyak dibicarakan orang [2]. Karena hal tersebut banyak toko- toko kecil dan perusahaan besar yang melakukan penjualan secara online melalui *online shop*. *Online Shop* adalah fasilitas yang disajikan oleh internet untuk mempermudah masyarakat dalam berbelanja tanpa harus bertatap muka dengan pelanggan, tanpa harus antri dan tawar- menawar [3]. PT. Yanmarindo Perkasa adalah sebuah perusahaan yang bergerak di bidang penjualan mesin, perkakas dan *sparepart*. Mesin- mesin yang dijual salah satunya yaitu mesin genset. Genset yang dijual di perusahaan ini sangat banyak dengan kriteria, merek, tipe dan kualitas yang berbeda – beda. Akibatnya, konsumen menjadi kesulitan dalam menentukan genset yang sesuai dengan kebutuhan dan kriteria yang mereka inginkan karena banyaknya variasi genset tersebut. Sehingga mereka harus menelpon bahkan sampai meluangkan waktunya ke toko untuk melihat serta bertanya kepada sales mengenai genset yang cocok untuk dibeli. Sistem rekomendasi sudah pernah diterapkan oleh [4] untuk pencarian gedung serbaguna dan [5] sebagai penentu destinasi wisata di Bali. Untuk memberikan rekomendasi produk kepada pelanggan juga pernah dilakukan oleh [6] dengan menggunakan algoritma Apriori. Penelitian mengenai sistem rekomendasi menggunakan algoritma *Naïve Bayes* sendiri sudah pernah diangkat oleh [7], dimana objek penelitiannya tersebut merupakan pakaian wanita. Hasilnya, algoritma *Naïve Bayes* ini terbukti mampu memberikan rekomendasi pakaian wanita

kepada penggunanya. Berdasarkan permasalahan tersebut, diperlukan sebuah sistem rekomendasi yang dapat mengklasifikasikan genset dengan kriteria berbeda namun masih dalam jenis yang sama serta memberikan rekomendasi genset, sesuai dengan kebutuhan dan kriteria yang diinginkan oleh konsumen. Kemudian konsumen dapat langsung bertransaksi secara *online* barang yang direkomendasikan tersebut

2. METODOLOGI

2.1. Data Mining

Menurut Gartner Group, data mining adalah proses menemukan hubungan baru yang mempunyai arti, pola dan kebiasaan dengan memilah-milah sebagian besar data yang disimpan dalam media penyimpanan dengan menggunakan teknologi pengenalan pola seperti teknik statistik dan matematika.

Data mining menurut David Hand, Heikki Mannila, dan Padhraic Smyth dari MIT adalah analisa terhadap data (biasanya data yang berukuran besar) untuk menemukan hubungan yang jelas serta menyimpulkannya yang belum diketahui sebelumnya dengan cara terkini dipahami dan berguna bagi pemilik data tersebut.

Data mining merupakan kombinasi dari ilmu- ilmu yang menyatukan teknik dari pembelajaran mesin, pengenalan pola, statistik, database, dan visualisasi untuk penanganan permasalahan pengambilan informasi dari database yang besar [8]. Sederhananya, data mining adalah suatu proses untuk menggali dan menghasilkan informasi yang lebih berguna untuk pengambilan keputusan dari suatu kumpulan data yang selama ini tidak diketahui secara manual.

2.2. Naïve Bayes

Algoritma *Naïve Bayes* merupakan suatu metode klasifikasi yang didasarkan pada probabilitas dan statistik. Algoritma ini dikemukakan oleh seorang ilmuwan inggris yang bernama Thomas Bayes [7],

sehingga algoritma ini kemudian dikenal dengan istilah “*Naïve Bayes*”. Algoritma ini melakukan prediksi peluang di masa depan berdasarkan data yang ada pada masa sebelumnya sehingga dikenal dengan “Teorema Bayes”. Algoritma *Naïve Bayes* merupakan algoritma yang cukup sederhana karena hanya mengasumsikan klasifikasi atribut.

Pada algoritma *Naïve Bayes*, bila terdapat dua kejadian yang terpisah (seperti A dan B). Algoritma *Naïve Bayes* dirumuskan sebagai berikut :

$$P(H|X) = \frac{P(X|H)}{P(X)} P(H) \tag{1}$$

Naïve Bayes menggunakan asumsi yang sangat kuat (naif) akan independensi dari masing-masing kondisi atau kejadian, dimana masing-masing petunjuk saling bebas (independen) satu sama lain. Dengan asumsi tersebut, maka berlaku persamaan sebagai berikut :

$$P(H|X) = P(H) \prod_{i=1}^n P(X_i|H) \tag{2}$$

Dimana X = Data dengan *class* yang belum diketahui

H = Hipotesis data merupakan suatu *class* spesifik

P(H|X) = Probabilitas Hipotesis H berdasarkan kondisi X (*posterior* probabilitas)

P(H) = Probabilitas hipotesis H (*prior* probabilitas)

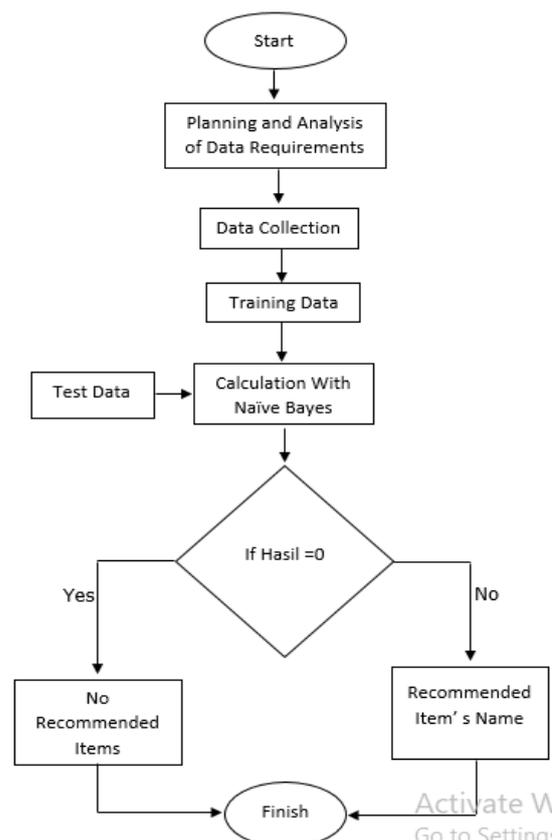
P(X|H) = Probabilitas X berdasarkan kondisi pada hipotesis H

P(X) = Probabilitas X

2.3. Kerangka Penelitian

Dalam mengimplementasikan algoritma *Naïve Bayes* ke dalam sistem rekomendasi yang dibangun, diperlukan kerangka penelitian, agar ketika sistem tersebut dalam proses pembuatan, hasil yang muncul tidak berbeda dari tujuan yang ingin dicapai. Seperti yang tertera pada Gambar 2, ketika penelitian dimulai, dilakukan perencanaan dan analisa kebutuhan data, kemudian dilakukan pengumpulan data yang digunakan dalam perhitungan *Naïve Bayes*. Beberapa data yang diperlukan seperti data gense, data

spesifikasi atau kriteria gense dan data penjualan gense (karena penelitian ini membahas gense saja). Data penjualan gense dan data spesifikasi gense akan saling dikaitkan dan dijadikan data *training*. Saat konsumen ingin menggunakan sistem, konsumen akan diberi pilihan mengenai kriteria gense apa yang dibutuhkan dan menjawab pertanyaan tersebut. Kriteria yang diinputkan konsumen ke dalam sistem, akan menjadi data uji. Data uji diproses menggunakan algoritma *Naïve Bayes* oleh sistem sesuai data *training* yang ada. Apabila hasil perhitungan adalah nol (0), maka sistem akan memberikan *output* berupa tulisan yang menandakan bahwa gense yang sesuai dengan kriteria yang dibutuhkan tidak ada atau tidak ada hasil rekomendasi yang sesuai. Selain dari hal tersebut, sistem akan menampilkan gense yang sesuai atau paling mendekati dengan kriteria yang diinginkan dan mengurutkan gense tersebut hasil perhitungan tertinggi ke terendah.



Gambar 2.1. Kerangka Penelitian

3. HASIL DAN PEMBAHASAN

3.1. Data Training

Metode yang digunakan dalam penelitian ini yaitu algoritma *Naïve Bayes*, sehingga dibutuhkan data *training* dalam perhitungan. Data *training* akan menjadi pedoman dari perhitungan data. Data *training* juga sering disebut sebagai *knowledge* karena data uji yang ada sangat dipengaruhi oleh data *training* yang terdapat di dalam *database*. Untuk data *training* yang ada sebanyak 209 data dan ini akan menjadi data awal untuk digunakan dalam perhitungan algoritma *Naïve Bayes*. Data *training* yang digunakan merupakan kombinasi antara data penjualan dan data genset beserta spesifikasinya. Terdapat 103 data genset dengan tipe berbeda yang

digunakan dalam penelitian ini. Untuk data penjualan genset yang digunakan yaitu data pada tahun 2020 sampai 2021. Tabel 1 adalah tabel inisial dari nama genset dan Tabel 2 merupakan *data training*.

Tabel 3.1. Inisial Nama Genset

Nama Genset	Inisial
FDG 7800 ECA	B31
SDP 1400	B32
SDP 2400	B33
SDP 7000	B34
SG 1500 X	B35
SG 3000 X	B36
SG 7500 X	B37
SG 10000 X	B38
GFH 2800 L	B39
GFH 5800 LX	B40

Tabel 3.2. Data Training

Bahan Bakar	Kapasitas (Watt)	Tegangan	Penyalan	Merek	Penjualan (Unit)	Inisial
Solar	4500	1 Phase	Engkol dan Starter	FRM	9	B31
Bensin Murni	1000	1 Phase	Engkol	SDP	12	B32
Bensin Murni	2000	1 Phase	Engkol dan Starter	SDP	11	B33
Bensin Murni	5000	1 Phase	Starter	SDP	4	B34
Bensin Murni	1000	1 Phase	Engkol	Shark	14	B35
Bensin Murni	2000	1 Phase	Starter	Shark	8	B36
Bensin Murni	5000	1 Phase	Starter	Shark	5	B37
Bensin Murni	6500	1 Phase	Starter	Shark	3	B38
Bensin Murni	1000	1 Phase	Engkol	Gambino	6	B39
Bensin Murni	2500	1 Phase	Engkol	Gambino	4	B40

3.2 Data Uji

Data uji merupakan data yang menjadi *input* atau masukan dalam sistem rekomendasi genset yang ada, dimana data tersebut akan dihitung sesuai dengan

metode yang dipakai yaitu algoritma *naïve bayes*. Tabel 3 adalah salah satu contoh dari data uji yang menjadi *input* dalam sistem rekomendasi genset.

Tabel 3.3. Data Uji

Bahan Bakar	Kapasitas (Watt)	Tegangan	Penyalaaan	Merek	Nama Genset
Bensin Campur	1- 2000	1 Phase	Engkol	Nishikawa	?????

3.3.Perhitungan

Dalam perhitungan ini, dilakukan penghitungan peluang setiap kelas kejadian ($P(H_i)$). Caranya yaitu membagikan jumlah setiap kelas kejadian ($n(H_i)$) dibagi dengan jumlah seluruh data ($n(S)$). Contohnya genset tipe FDG 7800 ECA dengan merek FRM (inisial B31), memiliki total jual 21 unit ($n(B31)$). Total penjualan genset pada data *training* yaitu 1267 unit ($n(S)$). Dengan begitu, 21 dibagi dengan 1267 menjadi 0,016574586 ($P(H_{B1})$). Lakukan hal yang sama pada data yang lainnya.

$$P(H_{B31}) = \frac{n(B31)}{n(S)} = \frac{21}{1267} = 0,016574586$$

Tabel 3.4. Perhitungan Setiap Kelas Kejadian Dibagi Total

Inisial	Merek	Total Penjuala n	$P(H_i)$
B31	FRM	21	0,016574586
B32	SDP	19	0,014996054
B33	SDP	20	0,01578532
B34	SDP	4	0,003157064
B35	Shark	24	0,018942384
B36	Shark	14	0,011049724
B37	Shark	5	0,00394633
B38	Shark	3	0,002367798
B39	Gambin o	6	0,004735596
B40	Gambin o	4	0,003157064

$$P(x_{Bensin Campur} | H_{B31}) = \frac{n(X_{Bensin Campur} | H_{B31})}{n(H_{B31})} = \frac{0}{21} = 0$$

Selanjutnya dilakukan pencarian probabilitas atribut dalam kelas ($P(X_i/H)$). Dimana jumlah *instance* dari data uji setiap kelas kejadian ($n(X_i/H)$) dibagi dengan jumlah *instance* setiap data uji ($n(H)$). Pada data uji, bahan bakar yang dipilih oleh konsumen merupakan bensin campur. Total penjualan genset bahan bakar bensin campur dengan nama genset FDG 7800 ECA merek FRM (inisial B31) yaitu 0 unit ($n(X_{Bensin Murni} | H_{B31})$). Total penjualan genset dengan tipe ini yaitu 21 unit ($n(H_{B31})$). Sehingga 0 dibagi dengan 21, maka didapat hasil 0 (nol) ($P(X_{Bensin Murni} | H_{B31})$).

$$P(x_{Bensin Campur} | H_{B34}) = \frac{n(X_{Bensin Campur} | H_{B34})}{n(H_{B34})} = \frac{0}{4} = 0$$

Contoh lainnya yaitu genset SDP 7000 merek SDP (inisial B34) dengan bahan bakar bensin murni, karena kriteria yang diinginkan merupakan bahan bakar bensin campur serta tidak ada genset tipe SDP 7000 merek SDP dengan bahan bakar bensin campur, maka $n(X_{Bensin Campur} | H_{B34})$ nilainya 0 (nol) dibagi dengan jumlah penjualan genset dengan tipe ini sebanyak 4 unit ($n(H_{B34})$), didapatkan nilai 0 (nol) ($P(X_{Bensin Campur} | H_{B34})$). Lakukan hal yang sama pada atribut kelas yang lain seperti kapasitas, merek, penyalaaan dan tegangan sesuai dengan data uji seperti Tabel 5.

Tabel 3.5. Perhitungan Atribut Dalam Kelas

Inisial	Total Penjualan	$P(X_1/H)$	$P(X_2/H)$	$P(X_3/H)$	$P(X_4/H)$	$P(X_5/H)$
B31	21	0	0	0	0	1
B32	19	0	1	0	1	1
B33	20	0	1	0	0	1
B34	4	0	0	0	0	1
B35	24	0	1	0	1	1
B36	14	0	1	0	0	1
B37	5	0	0	0	0	1
B38	3	0	0	0	0	1
B39	6	0	1	0	1	1
B40	4	0	0	0	1	1

Setelah dilakukan perhitungan *instance* setiap kelas, selanjutnya hasil dari seluruh perhitungan tadi dikalikan untuk

menemukan hasil akhir. Hasil perkalian dipaparkan pada Tabel 6.

Tabel 3.6 . Hasil Perhitungan

Inisial	$P(H)$	$P(X_1/H)$	$P(X_2/H)$	$P(X_3/H)$	$P(X_4/H)$	$P(X_5/H)$	$P(H/X)$
B31	0,016574586	0	0	0	0	1	0
B32	0,014996054	0	1	0	1	1	0
B33	0,01578532	0	1	0	0	1	0
B34	0,003157064	0	0	0	0	1	0
B35	0,018942384	0	1	0	1	1	0
B36	0,011049724	0	1	0	0	1	0
B37	0,00394633	0	0	0	0	1	0
B38	0,002367798	0	0	0	0	1	0
B39	0,004735596	0	1	0	1	1	0
B40	0,003157064	0	0	0	1	1	0

Berdasarkan hasil yang terdapat pada Tabel 6, data hasil akhir dengan nilai 0 (nol) akan dihapus atau tidak ditampilkan ke konsumen sebagai rekomendasi, sedangkan untuk data dengan hasil akhir bukan 0 (nol) akan diurutkan dari nilai tertinggi ke nilai terendah, kemudian akan tampil sebagai

genset yang direkomendasikan ke konsumen. Di sini, konsumen bisa melihat-lihat dan bebas memilih genset mana yang sesuai dengan konsumen tersebut. Hasil rekomendasi dari perhitungan *naïve bayes* dipaparkan pada Tabel 7.

Tabel 3.7. Hasil Rekomendasi Genset

Rangking	Nama Genset	$P(H/X)$
1	Genset NGG 1380 DCV NISHIKAWA	0.0165745856

Berdasarkan hasil rekomendasi yang terdapat pada Tabel 7, ditemukan satu jenis genset yang sesuai dengan kriteria yang konsumen inginkan dan memiliki nilai

$P(H/X)$ di atas 0 (nol). Genset inilah yang nantinya akan tampil sebagai hasil rekomendasi genset. Apabila terdapat lebih dari satu genset hasil hasil rekomendasi,

maka gensek yang ada akan diurutkan dari nilai $P(H/X)$ tertinggi ke terendah. Dari Tabel 7, terlihat bahwa gensek dengan tipe NGG 1380 DCV merek NISHIKAWA merupakan gensek yang paling direkomendasikan ke konsumen, karena memiliki nilai $P(H/X)$ tertinggi dengan nilai 0.0165745856.

4. KESIMPULAN

Berdasarkan apa yang telah dipaparkan pada pembahasan dan penelitian ini, maka dapat diambil kesimpulan sebagai berikut :

1. Sistem rekomendasi gensek yang dibangun dapat bekerja dan memberikan hasil rekomendasi gensek sebagaimana mestinya.
2. Penerapan algoritma *Naive Bayes* terbukti berhasil dalam memberikan rekomendasi gensek yang sesuai dengan kriteria konsumen berdasarkan data *training* yang ada. Hal ini dapat dibuktikan dari hasil akhir perhitungan, dimana gensek yang sesuai dengan kriteria konsumen bernilai tidak 0 (nol), yang kemudian akan diurutkan berdasarkan nilai tertinggi hingga terendah. Dari hasil rekomendasi, diketahui bahwa terdapat satu gensek yang sesuai dengan kriteria yang diinginkan konsumen yaitu gensek dengan tipe NGG 1380 DCV merek NISHIKAWA, dimana gensek ini memiliki nilai $P(H/X)$ tertinggi dengan nilai 0.0165745856.

5. SARAN

Berdasarkan sistem rekomendasi gensek yang telah dirancang dan dibangun menggunakan algoritma *Naive Bayes*, adapun saran untuk peneliti berikutnya yaitu :

1. Apabila ingin membangun sistem rekomendasi untuk suatu hal menggunakan algoritma ini, sebaiknya lebih diperhatikan terkait keanekaragamannya. Jika hal yang ingin diuji sangat beranekaragam, lebih baik menggunakan lebih banyak

kriteria, hal ini untuk meningkatkan keakuratan hasil perhitungan dengan algoritma ini.

2. Sistem rekomendasi barang yang dibangun hanya terbatas pada rekomendasi gensek saja. Lebih baik lagi jika sistem rekomendasi yang dibangun dapat digunakan untuk seluruh barang yang dijual pada PT. Yanmarindo Perkasa.

DAFTAR PUSTAKA

- [1] M. Zamroni, "Perkembangan Teknologi Komunikasi Dan Dampaknya Terhadap Kehidupan," *J. Pendidik.*, vol. X, no. 2, pp. 195–211, 2017.
- [2] M. A. Jaya, R. Ferdiana, and S. Fauziyati, "Analisis Faktor Keberhasilan Startup Digital di Yogyakarta Mardi Arya Jaya 1 , Ridi Ferdiana 2 , Silmi Fauziati 3," *Pros. SNATIF*, vol. 4, no. 1, pp. 167–173, 2017, [Online]. Available: <https://jurnal.umk.ac.id/index.php/SNATIF/article/view/1261>.
- [3] M. El Fikri, R. Ahmad, and R. Harahap, "Strategi Mengembangkan Kepuasan Pelanggan Online Shop Dalam Meningkatkan Penjualan (Studi Kasus Sabun Pyari)," *J. Manaj. Tools*, vol. 12, no. 1, pp. 87–105, 2020, doi: 10.1017/CBO9781107415324.004.
- [4] T. Badriyah *et al.*, "Rancang bangun aplikasi sistem rekomendasi pencarian gedung serbaguna," *J. Sist. Inf.*, vol. 9, no. 1, pp. 1–6, 2018, [Online]. Available: <http://ejournal.nusamandiri.ac.id/index.php/techno/article/view/105>.
- [5] Dwi Putra Githa and D. P. Singgih Putri, "Rancang Bangun Sistem Rekomendasi Destinasi Wisata Di Bali," *SINTECH (Science Inf. Technol. J.)*, vol. 2, no. 2, pp. 81–90, 2019, doi: 10.31598/sintechjournal.v2i2.388.
- [6] A. R. Riszky and M. Sadikin, "Data Mining Menggunakan Algoritma Apriori untuk Rekomendasi Produk

- bagi Pelanggan,” *J. Teknol. dan Sist. Komput.*, vol. 7, no. 3, pp. 103–108, 2019, doi: 10.14710/jtsiskom.7.3.2019.103-108.
- [7] R. Y. Hayuningtyas, “Penerapan Algoritma Naïve Bayes untuk Rekomendasi Pakaian Wanita,” *J. Inform.*, vol. 6, no. 1, pp. 18–22, 2019, doi: 10.31311/ji.v6i1.4685.
- [8] Y. Mardi, “Data Mining : Klasifikasi Menggunakan Algoritma C4.5,” *Edik Inform.*, vol. 2, no. 2, pp. 213–219, 2017, doi: 10.22202/ei.2016.v2i2.1465.
- [9] Y. S. Dwanoko, “Implementasi Software Development Life Cycle (SDLC) Dalam Penerapan Pembangunan Aplikasi Perangkat,” *J. Teknol. Inf.*, vol. 7, no. 2, pp. 83–94, 2016.