

Implementasi *Naïve Bayes* Dalam Memprediksi Permintaan Bahan Baku Di PT. Miyasaka Indonesia

Dyah Ayu Nurmumpuni^{*1}, Fery Salman Farisy², Novianti Maharani Putri³

^{*1,2}Sistem Komputer Universitas Banten, Serang

³Sistem Informasi STMIK Pranata Indonesia, Bekasi

e-mail: ^{*1}dnurmumpuni@univbanten.ac.id, ²ffarisy@univbanten.ac.id, ³noviantimaharaniputri@gmail.com

Abstrak

Penerapan algoritma *Naïve Bayes* dalam memprediksi kebutuhan bahan baku bertujuan untuk mengetahui persentase kontribusi masing-masing variabel terhadap faktor dominan dalam sistem permintaan bahan baku yang dikembangkan oleh PT. Miyasaka Indonesia. Hal ini digunakan sebagai acuan dalam mengklasifikasikan permintaan bahan baku agar sesuai dengan harapan dan kebutuhan perusahaan, serta untuk memberikan rekomendasi dan estimasi yang lebih tepat. Penelitian ini menggunakan pendekatan kuantitatif dengan menerapkan algoritma *Naïve Bayes* berdasarkan kriteria-kriteria yang ada dalam permintaan bahan baku, dan dianalisis menggunakan perangkat lunak RapidMiner. Berdasarkan hasil dari RapidMiner, diperoleh bahwa akurasi prediksi permintaan bahan baku dengan metode *Naïve Bayes* mencapai 95,12%, dengan nilai tertinggi pada kategori bahan baku dalam klasifikasi *performance vector*.

Kata Kunci: permintaan bahan baku, data mining, *Naïve Bayes*

Abstract

The implementation of the *Naïve Bayes* algorithm in predicting raw material demand aims to determine the percentage contribution of each variable to the dominant factor in the raw material request system developed by PT. Miyasaka Indonesia. This serves as a reference for classifying raw material demands to align with the company's expectations and needs, as well as to provide more accurate recommendations and estimations. This study employs a quantitative approach by applying the *Naïve Bayes* algorithm based on specific criteria for raw material requests, analyzed using RapidMiner software. Based on the results from RapidMiner, it was found that the accuracy of predicting raw material demand using the *Naïve Bayes* method reached 95.12%, with the highest value in the raw material category under the *performance vector* classification.

Keywords: raw material demand, data mining, *Naïve Bayes*

I. PENDAHULUAN

PT. Miyasaka Indonesia merupakan perusahaan yang bergerak di industri makanan yaitu tauco khas Jepang yang terletak di daerah Bekasi. Saat ini, PT. Miyasaka Indonesia dalam memprediksi permintaan bahan baku masih mengalami kendala, dikarenakan analisa data yang digunakan masih manual.

Dalam memprediksi jumlah permintaan bahan baku yang dibutuhkan setiap kegiatan produksi PT. Miyasaka Indonesia terkadang mengalami kelebihan stok bahan baku yang ada dalam gudang yang mengakibatkan kurang efisien dalam segi biaya. Persediaan barang didalam gudang perlu dijaga berkala. Stok barang adalah salah satu faktor penting yang harus diperhatikan oleh perusahaan (Nababan &

Vanessa Tanlim, 2019). Dibutuhkannya analisa data untuk mengolah dan menggali potensi data yang ada pada penyimpanan yang dapat membantu perusahaan untuk memprediksi permintaan bahan baku apa saja yang harus ditingkatkan maupun sebaliknya agar dapat meminimalisir bahan baku yang berlebih atau jarang dipakai dalam memenuhi kegiatan produksi perusahaan.

Saat ini banyak ilmu pengetahuan akan pemanfaatan data yaitu dengan menggunakan teknik *data mining*. *Data mining* dapat digunakan untuk mengekstrak (mengambil intisari) pengetahuan dari sekumpulan data sehingga mendapatkan struktur yang dapat mengerti manusia serta meliputi basis data dan manajemen data, prapemrosesan data, pertimbangan model dan inferensi, ukuran

ketertarikan, pertimbangan kompleksitas, pasca pemrosesan terhadap struktur yang ditemukan, visualisasi, maupun *online updating* (Laia, Buulolo, & Julyus, 2018).

Beberapa algoritma yang dapat digunakan untuk membantu analisis salah satunya adalah algoritma apriori. Algoritma apriori digunakan pengambilan data dengan aturan asosiatif (*association rule*) untuk menentukan hubungan asosiatif suatu kombinasi item (Simbolon, 2019). Algoritma apriori paling banyak digunakan dalam sistem dan strategi penjualan (Silalahi, 2020). Kekurangan algoritma ini yaitu membutuhkan waktu yang lama dalam proses pencarian frequent itemsets sehingga harus melakukan scanning database berulang kali setiap kombinasi item.

Algoritma lain yang dapat digunakan yaitu *frequent pattern growth (FP-Growth)*. Algoritma FP-Growth merupakan suatu alternatif algoritma yang digunakan untuk menentukan himpunan data yang paling sering muncul (*frequent itemset*) dalam sebuah himpunan data (Erwansyah, 2019).

Algoritma selanjutnya yang dapat digunakan adalah *Naïve Bayes*. *Naïve Bayes* yaitu metode yang mempunyai perhitungan matematik dasar yang sangat kuat serta dalam efisien klasifikasinya juga stabil (Wijayatun & Sulisty, 2016). Model *Naïve Bayes* memiliki tingkat kesalahan yang sangat minimum jika dibandingkan dengan algoritma klasifikasi lainnya (Wijayatun & Sulisty, 2016).

II. METODE PENELITIAN

Teknik Pengumpulan Data

Teknik pengumpulan data dilakukan untuk mendapatkan segala informasi dan data yang berhubungan dengan keperluan penyelesaian skripsi. Adapun hal yang dilakukan dengan 3 cara yaitu observasi, wawancara, dan studi pustaka.

Observasi merupakan metode yang dilakukan untuk mengidentifikasi data yang dilakukan secara sistematis baik dengan cara memperhatikan secara langsung atau tidak langsung objek yang diteliti serta mengambil data visual sesuai kebutuhan penelitian sehingga tidak ada data yang terlewat untuk mengimplementasikan *Naïve Bayes* dalam

memprediksi permintaan bahan baku di PT. Miyasaka Indonesia.

Wawancara sendiri merupakan teknik pengumpulan data, berita, fakta dan informasi dilapangan yang prosesnya bisa dilakukan dengan cara menanyakan langsung kepada pihak-pihak yang bisa memberikan informasi mengenai masalah yang sedang diteliti atau secara tidak langsung seperti melakukan telepon, email dan surat (wawancara tertulis).

Sedangkan studi pustaka adalah usaha yang dilakukan untuk menghimpun informasi yang relevan dengan topik dan masalah yang akan atau sedang diteliti. Informasi itu dapat diperoleh dari buku-buku ilmiah, laporan penelitian, karangan-karangan ilmiah, tesis dan disertasi, peraturan-peraturan, ketetapan-ketetapan, buku-buku tahunan, ensiklopedia dan sumber-sumber tertulis baik tercetak maupun elektronik serta internet.

Jenis dan Sumber Data

Data yang digunakan dalam penelitian ini yaitu data pemakaian bahan baku produksi bulan Oktober - Desember 2017 yang terdiri dari 204 record.

Data yang akan dipakai untuk penelitian merupakan data kuantitatif, data ini tersaji dalam bentuk angka yang sifatnya numerik.

Data penelitian diperoleh dari data internal perusahaan yaitu data pemakaian bahan baku di PT. Miyasaka Indonesia.

Variabel Penelitian

Variabel yang akan digunakan dalam proses *data mining* ditentukan berdasarkan tujuan penelitian. Variabel yang digunakan dalam penelitian ini yaitu bahan baku yang digunakan dalam proses produksi di PT. Miyasaka Indonesia yang diperoleh dari hasil studi lapangan.

Metode Penentuan Subjek

Algoritma *Naïve Bayes* menggunakan konsep standar deviasi atau nilai mean untuk memilih pembagian yang optimal. Setelah data yang didapat sebanyak 204 *record*, kemudian data tersebut dibagi menjadi dua secara acak untuk data *training* (80%) yaitu 163 record dan data *testing* (20%) yaitu 41 *record*.

Business Understanding

Penelitian ini akan dilakukan di PT. Miyasaka Indonesia berdasarkan masalah yang ditemukan, yang akan difokuskan untuk menganalisa permintaan bahan baku yang dibutuhkan, sesuai dengan kriteria yang dibutuhkan oleh perusahaan ini. Dan pembuatan analisis menggunakan metode *Naïve Bayes Classification*, untuk memenuhi permintaan bahan baku terhadap kebutuhan produksi dan memaksimalkan persediaan bahan baku didalam gudang, serta dapat memberikan tingkat produksi yang baik untuk perusahaan.

Tahapan Pengumpulan Data (*Aggregation*)

Proses pengumpulan data (*aggregation*) merupakan proses mengkombinasikan dua atau lebih atribut atau objek-objek kedalam satu atribut tunggal atau objek. Dalam pengumpulan data dilakukan dengan mengambil data langsung dari PT. Miyasaka Indonesia. Data yang digunakan dalam penelitian yaitu data transaksi pembelian bahan baku dengan melalui file data transaksi pembelian bahan baku pada PT. Miyasaka Indonesia.

Data yang didapat dari PT. Miyasaka Indonesia adalah data pembelian bahan baku dan pemakaian bahan baku produksi dari Oktober - Desember 2017 yang akan mempengaruhi dalam jumlah permintaan bahan baku yang ada didalam persediaan gudang, jumlah data sebanyak 204 *record* yang akan dijadikan nilai deviasi pada *data mining*. Tabel 1 merupakan data pemakaian bahan baku di PT.

Miyasaka Indonesia pada bulan Januari – Desember 2017.

Tahapan Pembersihan Data (*Cleaning Data*)

Data tahap pembersihan data semua atribut data yang digunakan, hanya menggunakan atribut-atribut data yang dianggap penting untuk digunakan. Oleh karena itu, perlu dilakukan proses pembersihan data, memperbaiki kesalahan data yang tidak diperlukan sehingga menyisakan data yang penting yang akan digunakan dalam tahap selanjutnya. Agar mudah diolah data rekapan yang masih merupakan data mentah harus diolah kembali berdasarkan data pemakaian bahan baku yang telah ditentukan pada tahap sebelumnya kedalam Microsoft Excel.

Pada data pemakaian bahan baku di PT. Miyasaka Indonesia yang telah dikumpulkan dilakukan *cleaning data* yaitu dengan menghilangkan *record-record* yang *noise* atau tidak lengkap dan *record* yang berulang. Atribut yang tidak diperlukan juga bisa dihilangkan, seperti kode barang, satuan dan tanggal pemakaian. Hal ini dilakukan karena atribut tersebut tidak berpengaruh kepada tahap proses konsep standar deviasi dan mean.

Pada data penelitian tahap pembersihan data dilakukan dengan menghapus beberapa data yaitu nomor, nomor faktur, tanggal faktur, kode barang, satuan, jumlah harga, dan tanggal pemakaian bahan baku.

Tabel 1. Data Awal

No	Tanggal	KodeBarang	NamaBarang	QTY	Satuan
1	10/4/2017	021204201	SOYBEAN GMO USA		
2	10/4/2017	021204102	RICE	372	KG
3	10/4/2017	021204101	SALT SP-4	205	KG
4	10/4/2017	021204105	RICE STARCH	1.302	KG
5	10/4/2017	021204301	MOLD SPORE	130.2	GR
6	10/4/2017	021204107	WATER	90	LTR
7	10/4/2017	021204108	POTASSIUM SORBATE	0.855	KG
8	10/5/2017	021204201	SOYBEAN GMO USA	330	KG
9	10/5/2017	021204102	021204102	RICE	339
...
204	12/28/2017	021204103	ETHANOL 96%	264.1	LTR

Tabel 2. Data Transformasi

No.	Supplier	Nama Barang	Qty	Harga	Skala Pemakaian
1	PT. PURNAMA	SOYBEAN GMO USA	>315	Rp 1000 - Rp 30000	TINGGI
2	PT. LOTTE SHOPPING INDONESIA	RICE STARCH	1.187-1.471	Rp 1000 - Rp 30000	NORMAL
3	PT. MASUYA GRAHA TRIKENCANA	MOLD SPORE	1.187-1.471	Rp 1000 - Rp 30000	NORMAL
4	PT. SINAR SOSRO	WATER	>87	Rp >30000	TINGGI
5	PT. LOTTE SHOPPING INDONESIA	POTASSIUM SORBATE	>513	Rp >30000	TINGGI
...
204	PD. LUMBUNG PANGAN MANDIRI	RICE	339-419	Rp 1000 - Rp 30000	NORMAL

Transformasi Data

Pada tahap ini merupakan tahap data yang disederhanakan atau dikelompokan. Data yang digunakan dalam penelitian ini adalah pembelian

bahan baku dan pemakaian bahan baku pada bulan Oktober - Desember 2017. Sumber data berupa aplikasi file dalam Microsoft Excel dengan jumlah 204 record data pemakaian bahan baku. Data akan ditransformasikan agar dapat mempermudah dalam pembacaan *data mining*. Penilaian skala pemakaian dapat dilihat pada tabel 3 dan data yang telah ditransformasikan dapat dilihat pada tabel 2.

Data Preparation

data yang akan digunakan adalah sejumlah data sebanyak 204 *record* yang kemudian akan dibagi dua yaitu data training dan data testing. Pembagian data yaitu 80:20 yang menghasilkan 163 data *training* dan 41 data *testing*.

Tabel 3. Penilaian Skala Pemakaian

QTY	Rendah	Normal	Tinggi
ETHANOL 96%	< 210	210 - 230	> 250
MOLD SPORE	< 118.7	118.7 - 147.1	> 147.1
POTASSIUM SORBATE	< 0.171	0.171 - 0.513	> 0.513
RICE	< 339	339 - 419	> 419
RICE STARCH	< 1.187	1.187 - 1.471	> 1.471
SALT SP-4	< 134	134 - 205	> 205
SOYBEAN GMO USA	< 300	300 - 315	> 315
WATER	< 67.5	67.5 - 87	> 87

Sebelum dilakukan tahap pengolahan data, perlu dilakukan tahap persiapan (*preprocessing*), dengan tujuan untuk mencari field data yang kosong (*missing value*) dan data dengan format yang tidak sesuai (*noise*) agar akurasi dari hasil pengolahan data dapat meningkat. Tahapan ini dilakukan dengan bantuan software rapidminer dengan data yang telah dikumpulkan akan di import kedalam rapidminer dengan bantuan pembacaan file excel.

Naïve Bayes Modelling

Pada tahap ini, data yang akan digunakan yaitu berupa data training yang terdiri dari 163 data. Diketahui, skala pemakaian “Rendah” sebanyak 52 data, skala pemakaian “Normal” sebanyak 48 data dan skala pemakaian “Tinggi” sebanyak 63 data.

Data yang diperoleh terdiri dari 4 atribut dan 1 atribut skala pemakaian yang anggota dan probabilitasnya adalah supplier bahan baku, nama barang, quantity, harga, dan skala pemakaian.

Tabel 4. Tabel Probabilitas

Atribut	Anggota	Prob.
Supplier	PD. Lumbang Pangan Mandiri	17
Bahan Baku	PT. Gemilang Karunia Abadi	2
	PT. Lotte Shopping Indonesia	35
	PT. Masuya Graha Trikencana	19
	PT. Purnama	33
	PT. Saltindo	22
Nama Barang	PT. Sinar Sosro	35
	Ethanol 96 %	2
	Mold Spore	19
	Potassium Sorbate	17
	Rice	17

	Rice Starch	18
	Salt SP-4	22
	Soybean GMO USA	33
	Water	35
Quantity	<210	1
	210 – 230	0
	>230	1
	<118.7	8
	118.7 - 147.1	8
	>147.1	3
	<0.171	5
	0.171 – 0.513	3
	>0.513	9
	<339	7
	339 -419	8
	>419	2
	<1.187	6
	1.187 – 1.471	9
	>1.471	3
	<134	2
	134 – 205	15
	>205	5
	<300	1
	300 - 315	0
>315	32	
67.5 – 87	4	
>87	8	
Harga	<Rp 1000,-	35
	Rp 1000,- sampai 30.000,-	109
	>Rp 30.000,-	19
Skala	Rendah	41
Pemakaian	Normal	59
	Tinggi	63

Pendefinisian variable

$C = \{C1=PT. Purnama, C2=Soybean GMO USA, C3=330, C4=Rp 6,700\}$.

Pendefinisian prior

Probabilitas prior yang terdiri dari skala pemakaian “Rendah” ($C1$), skala pemakaian “Normal” ($C2$) dan skala pemakaian “Tinggi” ($C3$)

$$P(C_1) = \frac{\text{skala pemakaian=rendah}}{\text{jumlah daya uji}} = \frac{41}{163} = 0,25 \quad (1)$$

$$P(C_2) = \frac{\text{skala pemakaian=normal}}{\text{jumlah daya uji}} = \frac{59}{163} = 0,36 \quad (2)$$

$$(C_3) = \frac{\text{skala pemakaian=tinggi}}{\text{jumlah daya uji}} = \frac{63}{163} = 0,39 \quad (3)$$

Menghitung probabilitas pada setiap kriteria, yaitu

$$P(H|X) = \frac{P(H|X).P(H)}{P(X)} \quad (4)$$

Beberapa perhitungan untuk kriteria supplier P(supplier), didapatkan:

PT. Purnama | rendah = 0,02

PT. Purnama | normal = 0

PT. Purnama | tinggi = 0,51

Untuk kriteria nama barang P(nama barang) didapatkan:

Soybean GMO USA | rendah = 0,02

Soybean GMO USA | normal = 0

Soybean GMO USA | tinggi = 0,51

Untuk P(Quantity) didapatkan:

>315 | rendah = 0

>315 | normal = 0

>315 | tinggi = 0,51

Untuk P(harga) didapatkan:

Rp 1,000 – Rp 30,000 | rendah = 0,02

Rp 1,000 – Rp 30,000 | normal = 0

Rp 1,000 – Rp 30,000 | tinggi = 0,51

Perhitungan dilakukan untuk semua kriteria terhitung sehingga didapatkan:

$P(\text{Rendah}) = 0,25 \times 0,02 \times 0,02 \times 0,02 = 0$

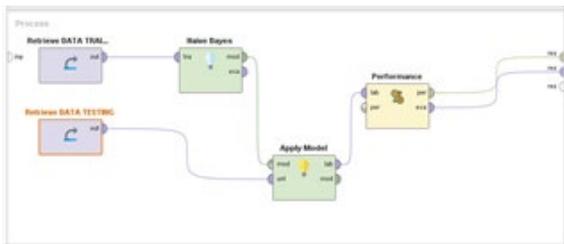
$P(\text{Normal}) = 0,36 \times 0 \times 0 \times 0 = 0$

$P(\text{Tinggi}) = 0,39 \times 0,51 \times 0,51 \times 0,51 \times 0,1 = 0,005173389$

Sehingga disimpulkan skala permintaan bahan baku kedelai adalah TINGGI.

Rapidminer Modelling

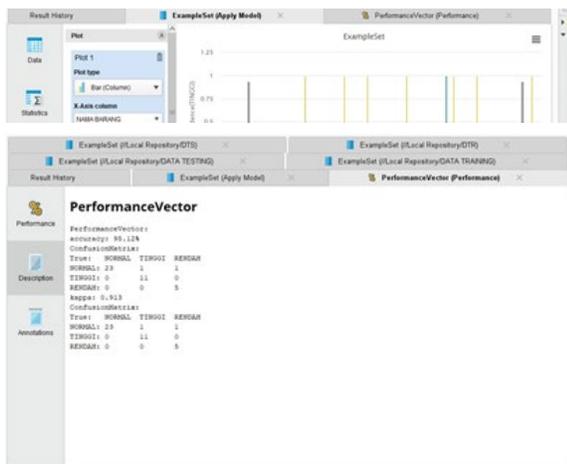
Data Training dan data testing diujikan pada software rapidminer untuk mendapatkan hasil yang lebih akurat. Proses dilakukan dengan menggunakan model *Naive Bayes* yang operatornya telah disediakan oleh rapidminer. Desain *rapidminer* yang disusun dapat dilihat pada gambar 1.



Gambar 1. Desain Penelitian dengan rapidminer

III. HASIL DAN PEMBAHASAN

Berdasarkan percobaan yang telah dilakukan menggunakan *rapidminer* maka, berikut merupakan hasil *performance* dari data yang diolah dengan memperoleh *accuracy* data yaitu 95.12 % , class recall 100 % untuk prediksi normal, 91.67 % untuk prediksi tinggi dan 83.33 % untuk prediksi rendah.



Gambar 4. Performance Vector

Class	Accuracy
pred rendah	83.33%
class recall	100.00%
class recall	91.67%
class recall	83.33%
class recall	100.00%

Gambar 3. Akurasi Data

Gambar 2 menunjukkan hasil data yang diolah dalam bentuk charts bar column. Gambar 3 menunjukkan nilai *accuracy* yang didapatkan dari pengolahan data, sedangkan gambar 4 menunjukkan *performance vector* dari model yang dihasilkan.

IV. KESIMPULAN

Berdasarkan hasil akurasi algoritma *Naive Bayes* dan metode *rapidminer* diperoleh hasil nilai *confusion matrix* yaitu *accuracy* data 95.12 % dan *class recall* 100 % untuk prediksi normal, 91.67 % untuk prediksi tinggi serta 83.33 % untuk prediksi rendah.

Dengan adanya perbaikan pada permintaan bahan baku yang dibutuhkan dapat membantu kinerja perusahaan dalam proses memenuhi kebutuhan bahan baku produksi.

Sebagai kesimpulan akhir, dengan adanya penerapan *data mining* algoritma *Naive Bayes* dengan metode *rapidminer* diharapkan mampu memprediksi permintaan bahan baku di PT. Miyasaka Indonesia.

V. REFERENSI

Astuti Hermawati, F. (2013). *Data mining*. (P. Christian, Ed.). Yogyakarta: CV Andi Offset.

Erwansyah, K. (2019). Implementasi Data mining Untuk Menganalisa Hubungan Data Penjualan Produk Bahan Kimia Terhadap Persediaan Stok Barang Menggunakan Algoritma FP (Frequent Pattern) Growth Pada PT . Grand Multi Chemicals. *Jurnal Teknologi Sistem Informasi Dan Sistem Komputer TGD*, 2(2), 30–40.

Fathoroni, A., Fathonah, N. S., Ardaryah, R., & Riza, N. (2020). *Buku Tutorial Sistem Pendukung Keputusan Penilaian Kinerja Dosen*. (R. M. Awangga, Ed.). Bandung: Kreatif Industri Nusantara.

Fatmawati. (2016). Perbandingan Algoritma Klasifikasi Data mining Model C 4.5 Dan Naive Bayes Untuk Prediksi Penyakit Diabetes. *Jurnal Techno Nusa Mandiri*, XIII(1), 50–59.

- Ferdianti, A. (2018). *Implementasi Manajemen Berbasis Sekolah Dalam Meningkatkan Prestasi Belajar Siswa*. (E. Wiji Astuti, Ed.) (1st ed.). Yogyakarta: CV. GRE Publishing.
- Habibi, R., & Suryansah, A. (2020). *Aplikasi Prediksi Jumlah Kebutuhan Perusahaan*. (R. Habibi & A. Suryansah, Eds.) (1st ed.). Bandung: Kreatif Industri Nusantara.
- Istijanto. (2005). *Riset Sumber Daya Manusia*. Jakarta: PT. Gramedia Pustaka Utama.
- Laia, D.-, Buulolo, E., & Julyus, M. (2018). Implementasi Data mining Untuk Memprediksi Pemesanan Driver Go-Jek Online Dengan Menggunakan Metode Naive Bayes. *KOMIK (Konferensi Nasional Teknologi Informasi Dan Komputer)*, 2, 434–439.
- Manalu, E., Sianturi, F. A., & Manalu, M. R. (2017). Penerapan Algoritma Naive Bayes Untuk Memprediksi Jumlah Produksi Barang Berdasarkan Data Persediaan Dan Jumlah Pemesanan Pada CV. Papa Dan Mama Pastries. *Jurnal Manik Penusa*, 1(2), 18.
- Muhammad, Z., Rahmadhani, R., Rizqifaluthi, H., & Yaqin, M. A. (2018). Process Mining Akademik Sekolah Menggunakan RapidMiner. *Jurnal Ilmu Komputer Dan Teknologi Informasi*, 10(2),
- Nababan, D., & Vanessa Tanlim, A. (2019). Analisis Persediaan Stok Barang Menggunakan Algoritma C 4.5 (Studi Kasus CV Harapan Raya). *Jurnal Ilmiah Informatika*, 07 No 01, 6–10.
- Pransiska, N., & Mirza, A. H. (2016). Penerapan Data mining Prediksi Penjualan Barang Elektronik Terlaris Menggunakan Algoritma Naive Bayes. *Bina Darma Conference on Computer Science*, 2157–2169.
- Silalahi, N. (2020). Penerapan Data mining Dalam Prediksi Penjualan Prabot Rumah Tangga Menggunakan Metode Apriori Pada Toko
- Hasanah Mart. *Buildings Of Informatics, Technology and Science (BITS)*, 2(1), 33–38.
- Simbolon, P. H. (2019). Implementasi Data mining Pada Sistem Persediaan Barang Menggunakan Algoritma Apriori (Studi Kasus : Srikandi Cash Credit Elektronik dan Furniture). *Jurnal Riset Komputer (JURIKOM)*, 6(4), 401– 406.
- Suyanto. (2017). *Data mining Untuk Klasifikasi Dan Klusterisasi Data*. Bandung: Informatika Bandung.
- Wijayatun, R., & Sulistyoy, Y. (2016). Prediksi Rating Film Menggunakan Metode Naïve Bayes. *Jurnal Teknik Elektro*, 8(2), 60–63.
- Wulandari, R. T. (2017). *Data mining Teori Dan Aplikasi Rapidminer (I)*. Yogyakarta: Gava Media.