

Klasifikasi Data Menggunakan Metode Naïve Bayes untuk Pemilihan Vendor

Tasya Hertianti Fitria*, Tacbir Hendro Pudjiantoro, Puspita Nurul Sabrina

Jurusan Informatika, Fakultas Sains dan Informatika

Universitas Jenderal Achmad Yani

Jl. Terusan Sudirman, Cimahi

*tasyahertiantif17@if.unjani.ac.id

Abstrak— *Data mining* adalah kombinasi dari teknologi *database*, kecerdasan buatan, mesin pembelajaran dan statistik. Klasifikasi adalah proses pengkategorian yang dilakukan pada sekumpulan dokumen. Terdapat banyak metode klasifikasi, salah satunya Metode *Naïve Bayes* yang digunakan pada penelitian ini. Metode *Naïve Bayes Classification* adalah pengklasifikasian statistik yang dapat digunakan untuk memprediksi probabilitas keanggotaan suatu class. Tujuan dari penelitian ini yaitu membuat perangkat lunak yang menerapkan metode *Naïve Bayes* untuk mencari informasi mengenai pengetahuan baru pengklasifikasian data dengan tingkat akurasi yang lebih baik serta mampu memberikan informasi mengenai *vendor* yang sesuai dengan jarak, harga, kuantitas dan tahun berdiri sehingga menghasilkan informasi bagi perusahaan mengenai *vendor* yang direkomendasikan dan tidak direkomendasikan. Penelitian ini menggunakan menggunakan data pembelian pada bulan Juni – Agustus 2019, *dataset* sebanyak 1066 *record*. Hasil akhir dari penelitian ini adalah nilai akurasi algoritma klasifikasi *Naïve Bayes* adalah 96.96%, untuk evaluasi menggunakan *ROC Curve* dan untuk *AUC* adalah 0.955 yaitu tingkat akurasi *Excellent Classification*.

Kata kunci— *Data Mining; Klasifikasi; Naïve Bayes; Vendor*

I. PENDAHULUAN

Data mining adalah kombinasi dari teknologi *database*, kecerdasan buatan, mesin pembelajaran dan statistik. Sederhananya interpretasi *data mining* adalah untuk 'menambang' pengetahuan yang dibutuhkan berbagai bidang dari sejumlah besar data [1]. Pada prosesnya *data mining* akan mengekstrak informasi dengan cara menganalisis adanya pola ataupun hubungan keterkaitan dari data yang berukuran besar. *Data mining* berkaitan dengan ilmu-ilmu lain seperti, *Database System*, *Statistic*, *Data Warehousing*, *Machine Learning*, *Information Retrieval*, dan *Komputasi tingkat tinggi*. *Data mining* adalah sebuah proses, sehingga dalam melakukan prosesnya harus sesuai dengan prosedur *CRISP-DM (Cross-Industry Standard Process for Data Mining)* [2]. Namun sejumlah data yang besar membutuhkan waktu untuk proses penambangan datanya, maka dari itu pada penelitian ini bertujuan untuk melakukan klasifikasi data menggunakan Metode *Naïve Bayes*. Klasifikasi adalah suatu proses pengkategorian yang dilakukan terhadap sekumpulan dokumen. Klasifikasi sangat penting untuk kemudahan pengguna dalam melakukan pencarian dokumen [3]. Mitchel, (1997), Youn ,

Jeong., (2009), menyatakan secara teoritis sebuah optimal *classifier Bayes theorem* adalah meminimalkan probabilitas *error rate exists* klasifikasi *Bayesian* adalah klasifikasi statistik yang bisa memprediksi probabilitas sebuah kelas yang dapat dihitung dengan *Teorema Bayesian* [4]. Metode *Naïve bayes* adalah salah satu metode yang digunakan untuk klasifikasi teks. *Naïve Bayes* menggunakan teori probabilitas sebagai dasar teori [5]. *Vendor* dalam arti harfiahnya, *vendor* adalah penjual. Namun, *vendor* memiliki artian yang lebih spesifik yakni pihak ketiga dalam *supply chain* istilah dalam industri yang menghubungkan produk dari produsen untuk sampai ke *customer* yang menjual barang kepada perusahaan yang dijual kembali atau dipergunakan oleh *user* dari perusahaan tersebut [6]. Metode *Naïve Bayes* digunakan pada penelitian ini karena dapat menghasilkan akurasi yang maksimal dengan data latih yang sedikit [7]. Data yang digunakan pada penelitian ini adalah data pembelian pada bulan Juni – Agustus 2019, *dataset* sebanyak 1066 *record*.

Menurut penelitian Nia Nuraeni (2017) yang berjudul “Penentuan Kelayakan Kredit Dengan Algoritma *Naïve Bayes Classifier*. Studi Kasus Bank Mayapada Mitra Usaha Cabang PGC” dalam penelitian tersebut klasifikasi algoritma data mining secara luas digunakan untuk menentukan kelayakan kredit dari salah satu *Naïve Bayes Classifier*, *Bayes Classifier* unggul dalam meningkatkan nilai akurasi yang tinggi tetapi lemah dalam pemilihan atribut. Hasil penelitian untuk nilai akurasi algoritma klasifikasi *Naïve Bayes Classifier* adalah 89.33% [2].

Selanjutnya menurut penelitian Tika Mahardhika dan Ifan Rizqa M.Kom (2019) yang berjudul “Klasifikasi Aspirasi Mahasiswa dengan *Naïve Bayes Classifier*” dalam penelitian tersebut Klasifikasi teks dokumen menjadi cara yang paling baik untuk menentukan katagori berdasarkan isi aspirasi mahasiswa. Metode *Naïve Bayes* digunakan karena metode ini mampu menghasilkan akurasi yang tinggi. Dengan 1000 dokumen data latih yang masingmasing 250 dokumen setiap katagori yaitu “Sarana dan Prasarana”, “Dosen”, “Kepegawaian dan Sistem Akademik”, “Saran dan Masukan” menghasilkan akurasi 90.20% [8].

Berdasarkan beberapa penelitian tersebut menyatakan bahwa Metode *Naïve Bayes* memiliki akurasi yang tinggi dalam mengatasi masalah pada proses pencarian data, sehingga pada PT. XYZ sebuah perusahaan yang bergerak di *garment*,

khususnya topi. Perusahaan bekerja sama dengan *vendor-vendor* yang menjual bahan baku topi. Dalam memilih *vendor*, terdapat kriteria yang digunakan untuk memilih *vendor* yang akan diajak bekerja sama yaitu jarak, harga, kuantitas dan tahun berdiri.

Seiring banyaknya *vendor* yang bekerja sama dengan perusahaan PT.XYZ maka semakin bertambah banyak pula data *vendor* untuk informasi perusahaan mengenai *vendor* yang sesuai dengan jarak, harga, kuantitas dan tahun berdiri sehingga menghasilkan informasi mengenai *vendor* yang direkomendasikan dan tidak direkomendasikan bagi perusahaan untuk diajak bekerja sama. *Data* ini akan bermanfaat untuk melakukan proses *data mining* yang akan mendapatkan informasi mengenai pengetahuan pada suatu algoritma yaitu pada algoritma *Naïve Bayes*.

Dengan peneliti melakukan klasifikasi *vendor* menggunakan *data* pembelian pada bulan Juni hingga Agustus 2019 ini memberitahu mengenai ke efektifan dan ke akuratan dari algoritma *Naïve Bayes*, sehingga semakin besar nilai akurasi maka akan semakin baik dalam pengklasifikasiannya. Pengklasifikasian ini berdasarkan atribut jarak, harga, kuantitas dan tahun berdiri dari *vendor* dengan menggunakan data sebanyak 1066 *record*.

Selain memberikan pengetahuan pada suatu algoritma mengenai ke efektifan dan ke akuratnya, juga mampu memberikan informasi bagi perusahaan mengenai rekomendasi *vendor* yang akan diajak bekerja sama.

II. METODE

A. Perolehan Data

Penelitian ini memperoleh *data* dengan melakukan *survey* ke perusahaan untuk melakukan wawancara dengan pihak perusahaan guna untuk mendapatkan informasi dan *data* mengenai perusahaan. *Data* yang digunakan pada penelitian yaitu data pembelian dari Bulan Juni hingga Bulan Agustus 2019.

B. Proses Knowledge Discovery In Databases (KDD)

KDD atau *Knowledge Discovery In Databases* adalah proses terorganisir untuk mengidentifikasi pola yang *valid*, baru, berguna, dan dapat dipahami dari kumpulan data yang besar dan kompleks [9].

1) Data Selection

Tahap *data selection* merupakan tahap dalam pemilihan *data* dari banyaknya sekumpulan *data*, tahap ini perlu dilakukan sebelum masuk dalam tahap penggalian informasi dimulai. *Data vendor* akan melalui proses seleksi terlebih dahulu lalu *data* hasil seleksi ini digunakan untuk proses *data mining*, selanjutnya disimpan dalam suatu berkas yang terpisah dari basis *data* operasional [10].

Pada tahap *data selection* ini data sebanyak 1066 record dan atribut sebanyak 9 atribut dilakukan penyeleksian untuk menghasilkan *data* yang dibutuhkan karena jumlah atribut yang digunakan terlalu banyak dan tidak semua atribut menjadi syarat atribut penentu. Sehingga jumlah atribut yang digunakan sebanyak 4 atribut. Berikut atribut yang digunakan sebagai acuan untuk klasifikasi vendor pada Tabel 1.

TABEL 1. ATRIBUT DATA PENELITIAN

No	Atribut	Keterangan
1	Harga	Atribut ini menjelaskan berapa besarnya biaya yang dibutuhkan dalam barang yang dibeli. Pihak perusahaan lebih tertarik dengan harga yang terjangkau.
2	Kuantitas	Atribut ini menjelaskan kuantitas barang yang tersedia pada <i>vendor</i> , jika kuantitas barang selalu tersedia ketika ada pemesanan maka ketersediaan barang pada <i>vendor</i> tersebut dikatakan baik, begitupun sebaliknya.
3	Jarak	Atribut ini menjelaskan jarak lokasi <i>vendor</i> dengan perusahaan.
4	Tahun Berdiri	Atribut ini menjelaskan seberapa lama perusahaan <i>vendor</i> tersebut berdiri.

2) Praprocessing Data

Tahap *praprocessing* merupakan langkah yang dilakukan sebelum langkah inti, dimana sebagian besar data ialah data mentah yang kotor seperti data tidak lengkap, *data noise* atau tidak konsisten. *Data* yang tidak berkualitas akan menghasilkan kualitas *mining* yang tidak baik sehingga untuk dapat kualitas *mining* yang baik.

Pada proses *praprocessing data* ini, *data* yang digunakan pada penelitian adalah *data* pembelian perusahaan dari bulan Juni 2019 - Agustus 2019 sebanyak 1199 *record*. Setelah dilakukannya *data cleaning* jumlah *data* yang digunakan menjadi 1066 *record*. *Data* tersebut diubah menjadi *dataset* yang dibagi menjadi 2 bagian yaitu *data* latih dan *data* uji. Pada *data* latih terdapat 836 *record* dan *data* uji 230 *record*.

3) Transformation Data

Pada teknik ini dilakukannya proses perubahan data menjadi bentuk data yang sesuai. Dalam melakukan transformasi data pada penelitian ini yaitu mengubah data yang memiliki tipe data character menjadi numeric. Pada Transformasi data ini menggunakan Normalisasi Data Min-Max.

a) Normalisasi Min-Max

Min-Max normalization merupakan metode normalisasi dengan melakukan transformasi *linier* terhadap *data* asli sehingga menghasilkan keseimbangan nilai perbandingan antar *data* saat sebelum dan sesudah proses [11].

Tahap ini dilakukan agar rentang nilai dapat seimbang pada setiap atribut yang digunakan pada penelitian. Untuk mendapatkan nilai dengan rentang 0-1 maka dilakukan normalisasi data min-max yang terdapat pada persamaan (1).

$$N_x = \frac{MinRange + (X - MinValue)(MaxRange - MinRange)}{MaxValue - MinValue} \quad (1)$$

Diketahui :

$$MinRange = 0$$

$$MaxRange = 1$$

Berikut *data* mentah yang diambil secara acak yang akan ditranformasi menggunakan normalisasi *min-max* pada Tabel 2.

TABEL 2. DATA MENTAH

Nama Vendor	Tahun Berdiri	Jarak	Kuantitas	Harga	Ket.
Wirasandi	2010	109 km	1533	550	Tidak Rekomendasi
Kian Jaya	2017	217 km	2	17500	Tidak Rekomendasi
Dunia Garment Accessories , Cv	2006	39 km	18	9833.34	Tidak Rekomendasi
Rajawali Globalindo	2002	108 km	18	4000	Rekomendasi
Cahaya Gesper	2013	7 km	112	680	Rekomendasi
Biar Jaya Sentosa, Pt	1985	145 km	40000	43.5	Rekomendasi
Cellindo Putra Foam	2003	156 km	601.7	11242.20	Tidak Rekomendasi
Mustika Jaya Abadi, Cv	2004	49 km	240	24545.45	Tidak Rekomendasi
Sinar Mas, Toko	2013	172 km	237	29500	Tidak Rekomendasi
Rajawali Globalindo	2002	108 km	398	300	Tidak Rekomendasi

Berikut *data* mentah yang telah ditranformasi menggunakan normalisasi *min-max* pada Tabel 3.

TABEL 3. DATA MENTAH SETELAH NORMALISASI

Nama Vendor	Tahun Berdiri	Jarak	Kuantitas	Harga	Ket.
Wirasandi	0.7812	0.4857	0.3879	0.0172	Tidak Rekomendasi
Kian Jaya	1	1	0	0.5926	Tidak Rekomendasi
Dunia Garment Accessories , Cv	0.6562	0.1524	0.004	0.334	Tidak Rekomendasi
Rajawali Globalindo	0.5312	0.4809	0.004	0.1343	Rekomendasi
Cahaya Gesper	0.875	0	0.0275	0.0216	Rekomendasi
Biar Jaya Sentosa, Pt	0	0.6571	1	0	Rekomendasi
Cellindo Putra Foam	0.5625	0.7095	0.15	0.3802	Tidak Rekomendasi

Nama Vendor	Tahun Berdiri	Jarak	Kuantitas	Harga	Ket.
Mustika Jaya Abadi, Cv	0.5937	0.2	0.0595	0.8318	Tidak Rekomendasi
Sinar Mas, Toko	0.875	0.7857	0.0588	1	Tidak Rekomendasi
Rajawali Globalindo	0.5312	0.4809	0.099	0.0087	Tidak Rekomendasi

4) Data Mining

Data Mining adalah inti dari proses KDD, melibatkan kesimpulan dari algoritma yang mengeksplorasi data, mengembangkan model dan menemukan pola yang sebelumnya tidak diketahui. *Model* digunakan untuk memahami fenomena dari data, analisis dan prediksi [9].

a) Klasifikasi

Klasifikasi adalah suatu proses pengkategorian yang dilakukan terhadap sekumpulan dokumen. Klasifikasi sangat penting untuk kemudahan pengguna dalam melakukan pencarian dokumen [3]. Pada penelitian ini menggunakan teknik klasifikasi untuk melakukan pengkategorian *vendor* yaitu direkomendasikan dan tidak direkomendasikan. Metode yang digunakan untuk pengklasifikasian pada penelitian ini yaitu metode *Naïve Bayes*.

b) Algoritma Naïve Bayes

Algoritma *Naïve Bayes* merupakan salah satu algoritma klasifikasi data dengan menggunakan metode probabilitas dan statistik. Metode ini pertama kali dikenalkan oleh ilmuwan Inggris Thomas Bayes yang dikenal sebagai teorema *Bayes*, digunakan untuk memprediksi peluang yang terjadi di masa depan berdasarkan pengalaman di masa sebelumnya [2].

Naïve Bayes bekerja sangat baik dibanding dengan metode lainnya karena memiliki tingkat akurasi yang lebih baik dibanding metode lainnya.

Berikut langkah-langkah perhitungan dalam *Naïve Bayes* [12]:

- Hitung probabilitas *bersyarat/likelihood*
- Hitung probabilitas *prior* untuk tiap kelas
- Hitung probabilitas *posterior*

Persamaan (2) merupakan bentuk umum dari teorema *bayes*:

$$P(H|X) = \frac{P(X|H) \cdot P(H)}{P(X)} \quad (2)$$

Keterangan:

X = *Data* dengan *class* yang belum diketahui.

H = Hipotesis data X merupakan suatu *class* spesifik.

P(H|X) = Probabilitas hipotesis H berdasarkan kondisi X (*posteriori probability*).

$P(H)$ = Probabilitas Hipotesis H (*prior probability*).

$P(X|H)$ = Probabilitas X berdasar kondisi pada Hipotesis H

$P(X)$ = Probabilitas dari X.

5) Interpretation / Evaluation

Pada tahap ini dilakukannya mengevaluasi dan menafsirkan pola yang ditambah sehubungan dengan tujuan yang ditentukan sebelumnya. Langkah ini berfokus pada pemahaman dan kegunaan *model* yang diinduksi. Pada langkah ini pengetahuan yang ditemukan juga didokumentasikan untuk penggunaan lebih lanjut.

III. HASIL DAN DISKUSI

A. Algoritma Naïve Bayes

Metode *Bayes rule* digunakan dan diterapkan untuk melakukan penghitungan terhadap *posterior* dan probabilitas dari data yang digunakan pada penelitian ini. Klasifikasi akhir dihasilkan dengan menggabungkan kedua informasi yaitu *prior* dan *posterior* untuk menghasilkan probabilitas menggunakan aturan *bayes*. Berikut langkah-langkah perhitungan *Naïve Bayes*:

1) Menghitung Probabilitas Prior

Langkah pertama dalam perhitungan *Naïve Bayes* adalah menghitung Probabilitas *Prior* dan Probabilitas *Posterior* dengan menggunakan *data* latih sebanyak 10 *record data* yang telah dinormalisasi *min-max*.

Berikut persamaan menghitung *Probabilitas Prior* dan *Probabilitas Posterior* dibawah ini:

$$\text{Total data} = 10$$

$$\text{Data rekomendasi} = 3$$

$$\text{Data tidak rekomendasi} = 7$$

$$P(\text{Rekomendasi}) = 3 : 10 = \mathbf{0.3}$$

$$P(\text{Tidak Rekomendasi}) = 7 : 10 = \mathbf{0.7}$$

Setelah nilai probabilitas didapatkan untuk tiap hipotesis dari *class*, langkah selanjutnya adalah melakukan penghitungan terhadap kondisi probabilitas tertentu (Probabilitas X) dengan menggunakan *data* berdasarkan probabilitas tiap hipotesis (Probabilitas H) atau yang disebut dengan probabilitas *Prior*. Untuk mengetahui hasil perhitungan dari probabilitas *Prior*, dilakukan penghitungan dengan cara menjumlah kasus secara rinci dari tiap-tiap atribut *variabel data*, berikut hasil perhitungan probabilitas *prior* dengan menggunakan Algoritma *Naïve Bayes* pada Tabel 4.

Setelah dilakukan pengolahan *data* latih diatas, maka terdapat dua *class* yang dibentuk pada *probabilitas Prior*, yaitu:

Class Rekomendasi

Class Tidak Rekomendasi

TABEL 4. PROBABILITAS PRIOR

Atribut		Jumlah Data	Rekomendasi	Tidak Rekomendasi	P(X Ci)	
					^a Rekomendasi	^b Tidak Rekomendasi
Tahun Berdiri	< 0.5	1	0	1	0	1
	> 0.5	9	3	6	0.333	0.667
Jarak	< 0.5	6	0	6	0	1
	> 0.5	4	1	3	0.25	0.75
Kuantitas	< 0.5	9	3	6	0.333	0.667
	> 0.5	1	0	1	0	1
Harga	< 0.5	7	2	5	0.286	0.714
	> 0.5	3	1	2	0.333	0.667
Total		10	3	7	0.3	0.7

2) Menghitung Probabilitas Posterior

Tahapan selanjutnya adalah menghitung probabilitas *Prior* untuk menentukan *class* terhadap kasus baru dengan cara menghitung *probabilitas Posterior*nya, hal tersebut dilakukan jika ditemukan kasus baru dalam pengolahan data. Berikut tabel *probabilitas posterior* untuk menghitung kasus baru yang ditemukan:

TABEL 5. PROBABILITAS POSTERIOR

Data X		P(X Ci)	
^c Atribut	^d Nilai (Value)	^e Rekomendasi	^f Tidak Rekomendasi
Tahun Berdiri	> 0.5	0.333	0.667
Jarak	< 0.5	0	1
Kuantitas	< 0.5	0.333	0.667
Harga	> 0.5	0.333	0.667

Setelah mengetahui nilai probabilitas dari setiap atribut pada probabilitas tiap *class* atau yang dirumuskan dengan $P(X|Ci)$, langkah selanjutnya adalah melakukan penghitungan terhadap total keseluruhan probabilitas tiap *class*. Berikut persamaan untuk menghitung probabilitas tiap *class*:

$$\begin{aligned}
 P(X | \text{Rekomendasi}) &= P(\text{Tahun Berdiri} = > 0.5 | \text{Rekomendasi}) * \\
 &P(\text{Jarak} = < 0.5 | \text{Rekomendasi}) * P(\text{Kuantitas} = < 0.5 | \\
 &\text{Rekomendasi}) * P(\text{Harga} = > 0.5 | \text{Rekomendasi}) \\
 &= 0.333 * 0 * 0.333 * 0.333 \\
 &= \mathbf{0}
 \end{aligned}$$

$P(X | \text{Tidak Rekomendasi}) = P(\text{Tahun Berdiri} = > 0.5 | \text{Tidak Rekomendasi}) * P(\text{Jarak} = < 0.5 | \text{Tidak Rekomendasi}) * P(\text{Kuantitas} = < 0.5 | \text{Tidak Rekomendasi}) * P(\text{Harga} = > 0.5 | \text{Tidak Rekomendasi})$

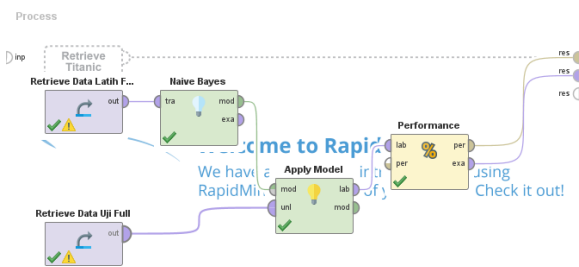
$$= 0.667 * 1 * 0.667 * 0.667$$

$$= \mathbf{0.2967}$$

$P(X|\text{Rekomendasi})P(\text{Rekomendasi}) = 0 * 0.3 = 0$

$P(X|\text{Tidak Rekomendasi})P(\text{Tidak Rekomendasi}) = 0.2967 * 0.7 = 0.2077$

Jadi hasil perhitungan pada probabilitas tiap *class* bahwa nilai $P(X|\text{Tidak Rekomendasi})$ lebih besar daripada nilai $P(X|\text{Rekomendasi})$, sehingga dapat disimpulkan bahwa dalam kasus pemilihan *vendor* tersebut akan masuk ke dalam klasifikasi **Tidak Rekomendasi**. Berikut gambar pengujian modelnya menggunakan *Tools Rapidminer* pada Gambar 1.



Gambar 1. Pengujian algoritma Naive Bayes

B. Interpretation / Evaluation

Pada *interpretation* atau *evaluation* ini dilakukannya proses untuk mengetahui tingkat keberhasilan klasifikasi algoritma *Naive Bayes* karena untuk mendapatkan nilai akurasi yang benar, maka diperlukan proses klasifikasi dan alat ukur yang tepat. Proses pengujian nilai akurasi ini dilakukan dengan menggunakan *tools software rapid miner* juga menggunakan model *Confussion Matrix* dan kurva *ROC (Receiver Operating Characteristic)*.

Pengujian nilai akurasi ini dilakukan menggunakan *data* latih sebanyak 836 *record* dan *data* uji sebanyak 230 *record*, 126 *record data* rekomendasi dan 104 *record data* Tidak direkomendasi.

C. Hasil Pengujian Algoritma Naive Bayes

Hasil dari pengujian algoritma *Naive Bayes* yang telah dilakukan adalah untuk mengukur tingkat akurasi dan *AUC (Area Under Cover)*.

1) Confussion Matrix

Hasil dari pengujian *tools rapidminer* dengan jumlah *data 1066 record*. Berikut tabel *confussion matrix* pada Gambar 2.

accuracy: 96.96%			
	true Tidak Rekomendasi	true Rekomendasi	class precision
pred. Tidak Rekomendasi	103	6	94.50%
pred. Rekomendasi	1	120	99.17%
class recall	99.04%	95.24%	

Gambar 2. Confussion Matrix algoritma Naive Bayes

Jumlah *True Positive (TP)* adalah 103 *record* diklasifikasikan sebagai **TIDAK REKOMENDASI** dan *False Negative (FN)* sebanyak 6 *record* diklasifikasikan sebagai **REKOMENDASI** tetapi **TIDAK REKOMENDASI**. Selanjutnya 120 *record* untuk *True Negative (TN)* diklasifikasikan sebagai **REKOMENDASI**, dan 1 *record False Positive (FP)* diklasifikasikan sebagai **REKOMENDASI** tetapi **TIDAK REKOMENDASI**.

Berdasarkan Gambar 2 tersebut menunjukkan bahwa tingkat akurasi dengan menggunakan algoritma *Naive Bayes Classifier* adalah sebesar 96.96% dan dapat dihitung untuk mencari nilai *accuracy*, *sensitivity*, *specificity*, *ppv*, dan *npv* pada persamaan dibawah ini:

$$acc = (tp + tn) / (tp + tn + fp + fn) \rightarrow acc = (103 + 120) / (103 + 120 + 1 + 6)$$

$$sensitivity = (tp) / (tp + fn) \rightarrow 103 / (103 + 6)$$

$$specitivity = (tn) / (tn + fp) \rightarrow 120 / (120 + 1)$$

$$ppv = (tp) / (tp + fp) \rightarrow 103 / (103 + 1)$$

$$npv = (tn) / (tn + fn) \rightarrow 120 / (120 + 6)$$

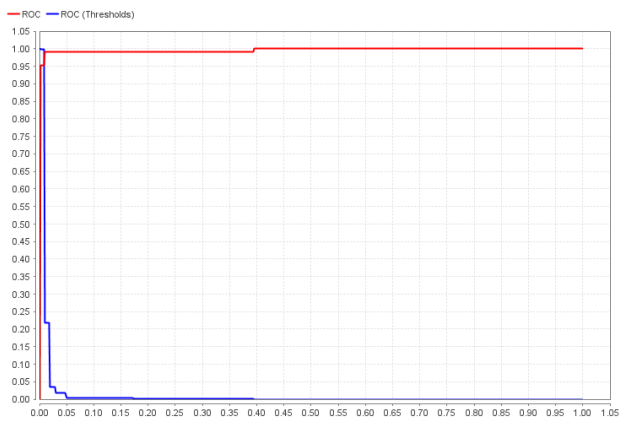
Hasil dari perhitungan persamaan untuk mencari nilai *accuracy*, *sensitivity*, *specificity*, *ppv*, dan *npv* pada Tabel ? dibawah ini:

TABEL 6. NILAI ACCURACY, SENSITIVITY, SPECIFICITY, PPV, DAN NPV

Parameter	Nilai (%)
Accuracy	96.96%
Sensitivity	94.49%
Specitivity	99.04%
PPV	95.24 %
NPV	95.24%

2) Evaluasi ROC Curve

Hasil dari Tabel 3 menampilkan grafik *ROC* dengan nilai *AUC (Area Under Cover)* sebesar 0.955 dengan nilai akurasi *Excellent Classification*.



Gambar 3. Nilai AUC dalam grafik ROC algoritma Naive Bayes

IV. KESIMPULAN

Hasil dari penelitian ini yaitu nilai akurasi algoritma klasifikasi *Naive Bayes* adalah 96.96%, untuk evaluasi menggunakan *ROC Curve* dan untuk AUC adalah 0.955 yaitu tingkat akurasi *Excellent Classification*.

Maka berdasarkan klasifikasi *vendor* menggunakan metode *Naive Bayes* berdasarkan atribut jarak, quantity, harga dan tahun berdiri yang dapat memberikan informasi serta keputusan yang dapat diambil dari hasil klasifikasi *vendor*.

Berdasarkan pengujian dan kesimpulan yang telah dilakukan, berikut saran dalam penelitian ini yaitu:

1. Bisa menggunakan metode lain atau menambah metode lain
2. Penambahan jumlah data dan atribut yang dapat memungkinkan untuk meningkatkan nilai akurasi
3. Pada PT.XYZ dapat ditingkatkan sistem untuk pemilihan *vendor*

DAFTAR PUSTAKA

- [1] Q. Zheng, Y. Li, and J. Cao, "Application of data mining technology in alarm analysis of communication network," *Comput. Commun.*, vol. 163, no. April, pp. 84–90, 2020, doi: 10.1016/j.comcom.2020.08.012.
- [2] N. Nuraeni, "Penentuan Kelayakan Kredit Dengan Algoritma Naive Bayes Classifier : Studi Kasus Bank Mayapada Mitra Usaha Cabang PGC," vol. III, no. 1, pp. 9–15, 2017.
- [3] A. Indriani, "Klasifikasi Data Forum dengan menggunakan Metode Naive Bayes Classifier," *Semin. Nas. Apl. Teknol. Inf. Yogyakarta*, vol. 1, no. 1, pp. 21–2014, 2014, [Online]. Available: www.bluefame.com,.
- [4] E. Junianto and B. Santosa, "Prediksi Yield Produk Asam Phosphate , Konsumsi Phosphate Rock dan Asam Sulfat untuk Pemilihan Supplier dengan Metoda Data Mining (Studi Kasus : PT . Petrokimia Gresik)," 2011.
- [5] A. Indranandita, B. Susanto, and A. Rahmat, "Sistem Klasifikasi Dan Pencarian Jurnal Dengan Menggunakan Metode Naive Bayes Dan Vector Space Model," *J. Inform.*, vol. 4, no. 2, 2011, doi: 10.21460/inf.2008.42.48.
- [6] K. B. Luhur and Sutandi, "Pengaruh Kualitas Produk terhadap Pemilihan Vendor di PT . Hitachi Power Systems Indonesia," vol. 4, no. 2, pp. 154–162, 2020.
- [7] R. N. Devita, H. W. Herwanto, and A. P. Wibawa, "Perbandingan Kinerja Metode Naive Bayes dan K-Nearest Neighbor untuk Klasifikasi Artikel Berbahasa Indonesia," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 5, no. 4, p. 427, 2018, doi: 10.25126/jtiik.201854773.
- [8] I. Rizqa, "Klasifikasi Aspirasi Mahasiswa Dengan Naive Bayes Classifier," *CSRID (Computer Sci. Res. Its Dev. Journal)*, vol. 11, no. 1, p. 01, 2021, doi: 10.22303/csrid.11.1.2019.01-09.
- [9] M. T. Elmore, "Introduction to knowledge discovery minitrack," *Proc. Annu. Hawaii Int. Conf. Syst. Sci.*, 2008, doi: 10.1109/HICSS.2008.246.
- [10] H. Junaedi, H. Budianto, I. Maryati, and Y. Melani, "Data Transformation pada Data Mining," *Pros. Konf. Nas. Inov. dalam Desain dan Teknol.*, vol. 7, pp. 93–99, 2011.
- [11] "Perbandingan Normalisasi Data untuk Klasifikasi Wine Menggunakan Algoritma K-NN," *CESS (Journal Comput. Eng. Syst. Sci.)*, vol. 4, no. 1, pp. 78–82, 2019.
- [12] M. A. Maricar and D. Pramana, "Perbandingan Akurasi Naive Bayes dan K-Nearest Neighbor pada Klasifikasi untuk Meramalkan Status Pekerjaan Alumni ITB STIKOM Bali," pp. 16–22, 2019, doi: 10.30864/jsi.v14i1.233.