



Comparative study of application of C4.5 algorithm and naïve bayes in selection partners of cooperation at pt. Citra pesona gemilang

Tri Hartati¹, Noer Hikmah², Heru Purwanto³

Information System Study Program¹
Department of Information System^{2,3}
Faculty Engineering and Informatics^{1,2,3}
Universitas Bina Sarana Informatika^{1,2,3}

tri.tri@bsi.ac.id¹, noer.nhh@bsi.ac.id², heru.hrp@bsi.ac.id³

Received: August 30, 2022 **Revised:** October 12, 2022 **Accepted:** October 27, 2022. **Issue Period:** Vol.6 No.2 (2022), Pp. 390-401

Abstrak: Saat ini banyak perusahaan dimana bidang usahanya berorientasi pada dunia pendidikan dimana mengedepankan teknologi sebagai sumber informasi, dan sebelum perusahaan melakukan kontrak kerjasama dengan vendor penyedia perangkat teknologi informasi maka ada banyak hal yang memerlukan pertimbangan dalam menentukan pemilihan mitra kerjasama dimana pertimbangan ini sangat penting dilakukan untuk mengetahui prospek keuntungan yang akan dihasilkan perusahaan dari kerjasama tersebut. Penelitian ini mengambil studi kasus pada PT. Citra Pesona Gemilang yang perusahaannya bergerak dibidang penyediaan berbagai macam kebutuhan sekolah. Mitra kerjasama dari PT Citra Pesona Gemilang adalah vendor dari perangkat teknologi. Dalam penelitian yang dilakukan, penulis menggunakan metode pengambilan keputusan dengan algoritma C4.5 dan Naïve Bayes untuk menentukan pemilihan mitra kerjasama terbaik. Pemilihan tersebut dilihat berdasarkan kualitas dan kuantitas produk/jasa, harga produk/jasa, legalitas dari vendor, keunggulan/kehandalan bermitra. Hasil dari penelitian ini adalah membandingkan tingkat akurasi yang dihasilkan oleh kedua algoritma yaitu algoritma C4.5 dan Naïve Bayes dalam menentukan pemilihan mitra kerjasama penyedia perangkat teknologi informasi.

Kata kunci: *algoritma C4.5, Naïve Bayes, Pemilihan Mitra Kerjasama*

Abstract: Currently, there are many companies whose line of business is education-oriented, which currently prioritizes technology as a source of information, and before a company enters into a cooperation contract with a vendor providing information technology equipment, many things need to be considered in determining the selection of partners where this consideration is very important to find out the prospect of profit that will be generated by the company from the cooperation. This research takes a case study at PT. Citra Pesona Gemilang whose company is engaged in providing various kinds of school needs. The cooperation partner of PT Citra Pesona Gemilang is a vendor of technology devices. In the research conducted, the authors use a decision-making method with the C4.5 algorithm and Naïve Bayes to determine the selection of the best cooperation partners. The selection is seen



DOI: 10.52362/jisicom.v6i2.927

Ciptaan disebarluaskan di bawah [Lisensi Creative Commons Atribusi 4.0 Internasional](https://creativecommons.org/licenses/by/4.0/).

based on the quality and quantity of the product/service, the price of the product/service, the legality of the vendor, the superiority/reliability of the partner. The result of this study is to compare the level of accuracy produced by the two algorithms, namely the C4.5 algorithm and the Naïve Bayes algorithm in determining the selection of partners in the provision of information technology tools.

Keywords: *algorithm C4.5, Naïve Bayes, Selection of Cooperation Partners*

I. PENDAHULUAN

Perkembangan teknologi saat ini memiliki peran yang sangat mendominasi dalam segala aspek kehidupan terutama bisnis. Peluang kesuksesan dan keunggulan perusahaan sangat bergantung pada upaya memanfaatkan secara optimal akan teknologi informasi. Dalam dunia bisnis sudah menjadi hal yang lumrah untuk melakukan persaingan untuk mencapai visi dan misi serta memperluas area bisnis. Tidak terkecuali PT Citra Pesona Gemilang yang memiliki target pencapaian bisnis dengan memanfaatkan kemajuan teknologi.

PT Citra Pesona Gemilang adalah sebuah perusahaan yang bergerak di bidang pelayanan penyediaan berbagai macam kebutuhan sekolah seperti perlengkapan sekolah, perangkat teknologi penunjang pembelajaran, layanan digital dan lain sebagainya. Dalam perjalanan bisnisnya PT Citra Pesona Gemilang melakukan kontrak kerjasama dengan beberapa vendor penyedia perangkat teknologi dan vendor pengembangan software. Permasalahan kemudian muncul pada saat proses pemilihan vendor sebagai mitra kerjasama dimana proses ini masih dilakukan secara konvensional sehingga menyebabkan kurang layaknya produk yang dikirimkan vendor, kendala pada software/aplikasi yang dikembangkan, waktu pengembangan software yang tidak sesuai/melebihi deadline. Hal ini tentu saja sangat berdampak pada kontrak kerjasama PT Citra Pesona Gemilang kepada sekolah-sekolah yang menjadi rekaan utama dari proses bisnis yang dijalankan. Untuk mengurangi kerugian yang disebabkan oleh kesalahan pemilihan vendor maka peneliti mencoba membuat perbandingan algoritma dari metode algoritma C4.5 dan metode Naïve Bayes dimana hasilnya diharapkan akan berguna untuk pengambilan keputusan dalam menentukan pemilihan mitra kerjasama (vendor) sehingga perusahaan dapat mengurangi kerugian yang terjadi selama ini dan dapat meningkatkan pelayanan kepada sekolah-sekolah yang menjadi rekaan bisnisnya.

II. METODE DAN MATERI

A. Metode Pengumpulan Data

Pengumpulan data diperlukan dalam penelitian untuk mendapatkan kejelasan dari permasalahan dan memberikan input bagi peneliti sehingga hasil yang didapat mendukung pengambilan keputusan yang bersifat manajerial.

1. Observasi

Pemahaman tentang rangkaian cara kerja sistem dan hasil atau output yang dikeluarkan oleh sistem sehingga dapat diketahui kekurangan atau kelemahan pada sistem yang digunakan.

2. Studi Pustaka

Peneliti juga melakukan kajian ilmu pengetahuan yang didapatkan melalui media informasi seperti jurnal, e-book, dan lainnya untuk mendapatkan gambaran solusi pemecahan masalah

B. Studi Literatur

1. Data Mining

Menurut Mardi, Data Mining adalah proses yang menggunakan teknik statistika, matematika, kecerdasan buatan dan machine learning untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai database besar. Data mining merupakan serangkaian proses yang digunakan untuk menggali nilai tambah dari suatu kumpulan data berupa pengetahuan yang selama ini tidak diketahui secara manual.[7]



DOI: 10.52362/jisicom.v6i2.927

Ciptaan disebarluaskan di bawah [Lisensi Creative Commons Atribusi 4.0 Internasional](https://creativecommons.org/licenses/by/4.0/).

Menurut Pramadhani dan Setiadi, secara sederhana data mining atau penambangan data dapat didefinisikan sebagai proses seleksi, eksplorasi dan pemodelan dari sejumlah besar data untuk menemukan pola atau kecenderungan yang biasanya tidak disadari keberadaannya.[10]

2. Algoritma Klasifikasi Data Mining

Klasifikasi adalah proses penemuan model (fungsi) yang menggambarkan dan membedakan kelas data atau konsep yang bertujuan agar bisa digunakan untuk memprediksi kelas pada objek yang label kelasnya tidak diketahui [4]. Berikut algoritma klasifikasi yang paling umum digunakan secara luas:

a. Decision/Clasification Tree

Digunakan untuk memprediksi keanggotaan suatu objek kedalam kategori kelas yang berbeda, berdasarkan kategori prediktor. Algoritma klasifikasi yang biasanya digunakan adalah: Hunt, Cart(C&RT), C4.5 dan C5.0, ID3, SLIQ, SPRINT, QUEST, DTREG, THAID, CHAID. Pohon keputusan (Decision Tree) merupakan salah satu metode penambangan data yang paling banyak digunakan karena mudah dalam pengaplikasiannya untuk menemukan pengetahuan yaitu dengan menganalisa secara sistematis serta dapat mengekstrak aturan sehingga tercapai tujuan klasifikasi/prediksi.

b. Bayesian Classifiers/Naïve Bayes Classifier

Klasifikasi Bayes adalah pengklasifikasian statistik yang dapat digunakan untuk memprediksi probabilitas keanggotaan suatu kelas. Klasifikasi bayes juga dikenal dengan istilah *Naïve Bayes*, *Idiot's Bayes*, *Simple Bayes*, dan *Independence Bayes*. Klasifikasi bayes didasarkan pada teorema bayes, diambil dari nama seorang ahli matematika yang juga menteri Prebyterian Inggris, Thomas Bayes (1702 – 1761).

3. Algoritma C4.5

Tahapan dalam membuat sebuah pohon keputusan dengan algoritma C4.5 adalah:

a. Data Training

Diambil dari data history atau data yang pernah terjadi sebelumnya dan kemudian dikelompokkan kedalam kelas-kelas yang berkesesuaian.

b. Menentukan akar pohon

Penentuan akar pohon dapat dilakukan dengan memilih atribut yang dianggap sebagai nilai gain tertinggi dari nilai gain atribut lainnya. Perhitungan nilai gain harus melalui perhitungan nilai entropy setiap atribut dengan rumus sebagai berikut:

$$Entropy(S) = \sum_{i=1}^n -p_i \cdot \log_2 p_i$$

Keterangan:

S = himpunan kasus

n = jumlah partisi S

p_i = proporsi S_i terhadap S

c. Menghitung Nilai Gain

Untuk menghitung nilai gain digunakan rumus:

$$Gain(S, A) = Entropy(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} * Entropy(S_i)$$

Keterangan:

S = himpunan kasus

A = fitur

n = jumlah partisi atribut A

$|S_i|$ = proporsi S_i terhadap S

$|S|$ = jumlah kasus dalam S

d. Loop

Lakukan pengulangan perhitungan dari proses 2 sehingga semua atribut yang berisi record terpartisi.



e. Stop

Proses akan berhenti apabila terjadi kondisi: semua record telah dikategorikan dalam kelas yang sama, tidak ada atribut yang dapat di partisi lagi, cabang pohon yang kosong tidak terdapat atribut/record.

4. Algoritma Naïve Bayes

Teorema bayes merupakan teorema yang biasa digunakan dalam bidang statistik dimana teorema ini dapat menghitung peluang untuk suatu hipotesis. Bayes optimal classifier digunakan untuk menghitung peluang masing-masing atribut yang berada pada suatu class dan menentukan class yang paling optimal diantara class lainnya. Pada dasarnya teorema bayes merupakan suatu formula yang dapat menjelaskan cara untuk memperbaharui probabilitas dari suatu hipotesis berdasarkan bukti/fakta baru yang dijumpai. Teorema ini juga merupakan persamaan matematis yang digunakan dalam statistik dan probabilitas untuk menghitung probabilitas bersyarat, dengan kata lain menghitung probabilitas dari suatu peristiwa berdasarkan hubungannya dengan peristiwa lain, dikenal dengan istilah 'Hukum Bayes' atau 'Aturan Bayes'. [8]

Peristiwa A => Kejadian B P (A|B)

Peristiwa B => Kejadian A P (B|A)

Logika => P (A|B) = P (B|A)

Formula dari teorema bayes adalah:

$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)}$$

dimana A dan B adalah dua kejadian dan $P(B) \neq 0$

$P(A|B)$ adalah **probabilitas bersyarat** dari peristiwa A yang terjadi mengingat B benar.

$P(B|A)$ adalah **probabilitas bersyarat** dari peristiwa B yang terjadi mengingat A benar.

$P(A)$ dan $P(B)$ adalah probabilitas peristiwa A dan peristiwa B yang terjadi secara independen satu sama lain (probabilitas marginal)

C. Desain Penelitian

Penelitian ini menggunakan desain model CRISP-DM (Cross Industries Standart Process For Data Mining) yang merupakan salah satu model datamining yang dibangun oleh lima perusahaan (Integral Solution Ltd, Teradata, Daimler AG, NCR Corporation, OHRA) [12]. Adapun tahapannya adalah:

1. Bussiness Understanding Phase

Pada tahapan ini dilakukan pemahaman terhadap bisnis yang dilakukan oleh perusahaan. Kemudian menentukan faktor faktor apa saja yang dibutuhkan dalam menjalankan bisnis sehingga perusahaan mendapatkan keuntungan dari proses bisnis. Implementasi data mining pada penelitian kali ini berhubungan dengan pemilihan mitra kerjasama yang akan mempengaruhi produktifitas perusahaan secara keseluruhan. Tujuannya adalah untuk mendapatkan pengetahuan tambahan mengenai faktor dan atribut pemilihan yang menjadi tolak ukur perhitungan kelayakan kerjasama.

2. Data Understanding Phase

Pengumpulan data dilakukan pada tahapan ini, dimana semua data yang dihasilkan akan mengalami penyeleksian dan pengolahan secara keseluruhan. Data yang dikumpulkan akan diintegrasikan secara langsung tanpa harus adanya proses warehouse dan tanpa adanya perpindahan data.

3. Data Preparation Phase

Fase ini merupakan tahapan dimana terjadi proses pemilihan dan pengolahan data secara random, hal ini dilakukan untuk dapat menghasilkan data yang akan dijadikan sebagai bahan modelling sehingga diharapkan hasil dari fase ini adalah data yang sesuai dengan kondisi perusahaan serta berkaitan erat dengan permasalahan pemilihan mitra kerjasama.

4. Modeling Phase

Dalam tahapan pemilihan terhadap data yang akan diolah akan menyesuaikan dengan proses mining. Klasifikasi data dilakukan dengan tujuan untuk melihat kelayakan vendor sebagai mitra kerjasama.



DOI: 10.52362/jisicom.v6i2.927

Ciptaan disebarluaskan di bawah [Lisensi Creative Commons Atribusi 4.0 Internasional](https://creativecommons.org/licenses/by/4.0/).

5. Evaluation Phase

Fase evaluasi merupakan tahapan dimana data akan dilakukan proses pengujian dan perbandingan. Pada fase ini sample data diambil secara random kemudian dilakukan perhitungan nilai secara manual selanjutnya dapat dilakukan perbandingan dengan menggunakan model yang digunakan. Pengujian terhadap model data bertujuan untuk menghasilkan model yang tepat dan akurat. Hal ini dapat dilakukan dengan teknik metode *Confusion Matrix* dan kurva ROC (*Receiver Operating Characteristic*).

6. Deployment Phase

Tahapan terakhir adalah penyebaran data, proses implementasi dari keseluruhan model yang telah dirancang pada tahapan modelling dan tahapan evaluasi berdasarkan dari hasil perhitungan nilai data yang dilakukan secara detail.

III. PEMBAHASAN DAN HASIL

A. Bussiness Understanding Phase

Berdasarkan data observasi yang penulis lakukan pada PT Citra Pesona Gemilang terkait vendor yang menjadi mitra kerjasama, diketahui bahwa terdapat pelayanan pengadaan perangkat teknologi informasi dan pengembangan software yang masih kurang optimal yang dilakukan oleh pihak mitra dikarenakan pada saat pemilihan masih tidak sesuai dengan kriteria standart. Sampai saat ini masih belum ditemukan algoritma yang memiliki tingkat akurasi tinggi untuk menentukan kelayakan pemilihan mitra kerjasama. Oleh karena itu peneliti akan mengkaji algoritma yang terbaik dalam menentukan pilihan vendor dengan mengbandingkan algoritma C4.5 dan algoritma Naïve Bayes.

B. Data Understanding Phase

Atribut atau variabel yang digunakan pada penelitian kali ini sebanyak delapan atribut yang dapat dilihat pada tabel 3.1 berikut.

Tabel 1 Atribut dan Nilai Kategori

No	Atribut/Variabel	Nilai Kategori
1.	Pengalaman Kerja Perusahaan	<5 tahun
		5 – 10 tahun
		>10 tahun
2.	Masa Kerjasama	<1 tahun
		1 – 5 tahun
		>5 tahun
3.	Bidang	Penyedia Hardware
		Pengembangan Software
		Tenaga IT
		Penyedia Perlengkapan
4.	Pelayanan	Mandiri
		Non Mandiri
5.	Skala Pelayanan	Antar Kota
		Antar Propinsi
		Antar Wilayah Kepulauan
6.	Harga Pelayanan	Murah
		Sedang
		Mahal
7.	Status Perusahaan	

8.	Remark	Swasta Kecil
		Swasta Menengah
		Swasta Besar
		Excellent
		Very Good
		Good
		Bad

C. Data Preparation Phase

Pada penelitian ini terdapat 50 record perusahaan mitra yang pernah bekerjasama dengan PT Citra Pesona Gemilang, dimana dari data data tersebut merupakan rekanaan bisnis yang memiliki kriteria layak ataupun tidak layak menjadi mitra bisnis, sehingga diperlukannya proses pengolahan kelayakan data agar didapatkan kualitas data terbaik. Teknik processing yang dilakukan meliputi tahapan:

1. *Validation*, proses ini akan mengecek dan mengeliminasi data yang bersifat *outlier/noise*, data tidak konsisten, ataupun data tidak lengkap/*missing value*.
2. *Integration and Transformation*, proses meningkatkan keakuratan dan keefisienan algoritma yang akan digunakan.
3. *Size reduction dicretization*, pengelompokan data kedalam data set yang memiliki atribut atau variabel dengan record yang lebih sedikit akan tetapi bersifat informatif.

Setelah dilakukan tahapan proses seperti yang dijelaskan diatas maka didapatkan data set yang akan digunakan untuk sample data training sebagai berikut:

Tabel 2 Sample Data Training

No	A	B	C	D	E	F	G	H
1	12	7	Hardware	Mandiri	Antar Kota	Sedang	Swasta Besar	Very Good
2	3	2	Hardware	Non Mandiri	Antar Kota	Mahal	Swasta Menengah	Good
3	10	4	Tenaga IT	Mandiri	Antar Kota	Mahal	Swasta Menengah	Very Good
4	27	8	Software	Mandiri	Antar Kota	Sedang	Swasta Menengah	Very Good
5	9	6	Hardware	Mandiri	Antar Kota	Sedang	Swasta Menengah	Good
6	12	2	Hardware	Non Mandiri	Antar Kota	Sedang	Swasta Menengah	Good
7	4	1	Perlengkapan	Mandiri	Antar propinsi	Mahal	Swasta Besar	Good
8	6	5	Software	Mandiri	Antar propinsi	Sedang	Swasta Besar	Very Good
9	19	3	Software	Mandiri	Antar Kota	Mahal	Swasta Menengah	Excellent
10	3	2	Software	Mandiri	Antar Kota	Mahal	Swasta Menengah	Good
11	11	6	Software	Mandiri	Antar Kota	Sedang	Swasta Menengah	Good
12	2	1	Tenaga IT	Mandiri	Antar Wilayah Kepulauan	Mahal	Swasta Besar	Bad



13	7	2	Hardware	Mandiri	Antar Kota	Sedang	Swasta Menengah	Very Good
14	8	4	Tenaga IT	Non Mandiri	Antar propinsi	Sedang	Swasta Besar	Very Good
15	17	6	Tenaga IT	Non Mandiri	Antar Kota	Mahal	Swasta Besar	Very Good
16	3	1	Perlengkapan	Mandiri	Antar Kota	Mahal	Swasta Besar	Bad
17	22	10	Hardware	Non Mandiri	Antar propinsi	Sedang	Swasta Besar	Very Good
18	19	2	Hardware	Mandiri	Antar Kota	Mahal	Swasta Menengah	Very Good
19	7	1	Hardware	Mandiri	Antar Kota	Mahal	Swasta Menengah	Good
20	11	1	Hardware	Mandiri	Antar Wilayah Kepulauan	Mahal	Swasta Besar	Good
21	19	4	Perlengkapan	Mandiri	Antar Kota	Mahal	Swasta Menengah	Good
22	3	2	Perlengkapan	Mandiri	Antar Kota	Sedang	Swasta Menengah	Bad
23	2	1	Tenaga IT	Mandiri	Antar propinsi	Mahal	Swasta Besar	Bad
24	21	8	Perlengkapan	Mandiri	Antar Kota	Sedang	Swasta Menengah	Very Good
25	6	1	Hardware	Mandiri	Antar Kota	Sedang	Swasta Menengah	Very Good
26	10	4	Software	Non Mandiri	Antar Kota	Sedang	Swasta Menengah	Very Good
27	8	2	Hardware	Non Mandiri	Antar Kota	Mahal	Swasta Menengah	Very Good
28	11	7	Tenaga IT	Mandiri	Antar Kota	Sedang	Swasta Menengah	Good
29	15	3	Software	Mandiri	Antar Kota	Mahal	Swasta Menengah	Good
30	7	5	Hardware	Mandiri	Antar propinsi	Mahal	Swasta Besar	Excellent

Keterangan:

A : Pengalaman Kerja Perusahaan

B : Masa Kerjasama

C : Bidang

D : Pelayanan

E : Skala Pelayanan

F : Harga Pelayanan

G : Status Perusahaan

H : Remark

D. Modeling Phase

Dalam melakukan klasifikasi data menjadi data set maka diperlukan proses modelling dengan menggunakan metode algoritma C4.5 dan algoritma Naïve Bayes.

1. Algoritma C4.5

Sample data training yang digunakan dalam proses modelling berjumlah 30 record dengan data testing sebanyak 15 record. Tahapan modelling menggunakan algoritma C4.5 adalah sebagai berikut:

- Menyiapkan data training dimana pada penelitian kali ini berjumlah 30 record.



DOI: 10.52362/jisicom.v6i2.927

Ciptaan disebarluaskan di bawah [Lisensi Creative Commons Atribusi 4.0 Internasional](https://creativecommons.org/licenses/by/4.0/).

b. Perhitungan Nilai Entropy

Pemasukan data training pada formula Entropy

$$\begin{aligned} Entropy(S) &= \sum_{i=1}^n -p_i \cdot \log_2 p_i \\ &= (-2/30 \cdot \log_2(2/30)) + (-13/30 \cdot \log_2(13/30)) + (-11/30 \cdot \log_2(11/30)) + (-4/30 \cdot \log_2(4/30)) \\ &= 1,70157512424704 \end{aligned}$$

c. Perhitungan Nilai Gain

Pemasukan data training pada formula Gain (Pengalaman Kerja Perusahaan)

$$\begin{aligned} Gain(S, A) &= Entropy(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} * Entropy(S_i) \\ &= 1,70157512424704 - ((7/30 * 0,985228136) + (10/30 * 1,156779649) \\ &\quad + (13/30 * 1,314320887)) \\ &= 0,516556465117353 \end{aligned}$$

Tabel 3 Hasil Perhitungan Nilai Entropy dan Gain

No	Atribut/Variabel	Nilai Kategori	Jumlah Data	Excellent	Very Good	Good	Bad	Entropy	Gain
1.	Pengalaman Kerja Perusahaan		30	2	13	11	4	1,701575124	0,516556465
		<5 tahun	7	0	0	3	4	0,985228136	
		5 – 10 tahun	10	1	7	2	0	1,156779649	
		>10 tahun	13	1	6	6	0	1,314320487	
2.	Masa Kerjasama		30	2	13	11	4	1,701575124	0,11014287
		<1 tahun	0	0	0	0	0	0	
		1 – 5 tahun	22	2	8	8	4	1,823067982	
		>5 tahun	8	0	5	3	0	0,954434003	
3.	Bidang		30	2	13	11	4	1,701575124	0,288029392
		Penyedia Hardware	12	1	6	5	0	1,325011211	
		Pengembangan Software	7	1	3	3	0	1,448815636	
		Tenaga IT	6	0	3	1	2	1,459147917	
		Penyedia Perlengkapan	5	0	1	2	2	1,521928095	
4.	Pelayanan		30	2	13	11	4	1,701575124	0,116426992
		Mandiri	23	2	8	9	4	1,804895651	
		Non Mandiri	7	0	5	2	0	0,863120569	
5.	Skala Pelayanan		30	2	13	11	4	1,701575124	0,131115886
		Antar Kota	22	1	10	9	2	1,561767711	
		Antar Propinsi	6	1	3	1	1	1,79248125	
		Antar Wilayah Kepulauan	2	0	0	1	1	1	
6.	Harga Pelayanan		30	2	13	11	4	1,701575124	0,155999393
		Murah	0	0	0	0	0	0	
		Sedang	14	0	9	4	1	1,198117421	
		Mahal	16	2	4	7	3	1,849601753	
7.	Status Perusahaan		30	2	13	11	4	1,701575124	0,105891205
		Swasta Kecil	0	0	0	0	0	0	

	Swasta Menengah	19	1	8	9	1	1,483226197
	Swasta Besar	11	1	5	2	3	1,789929075

Berdasarkan nilai gain pada setiap atribut pada tabel 3.3 maka dapat ditentukan simpul akar dari pohon keputusan yaitu atribut yang memiliki nilai gain tertinggi (atribut pengalaman kerja perusahaan).

2. Algoritma Naïve Bayes

Perhitungan algoritma NaïveBayes akan menggunakan data pada tabel 3.2 Sample Data Training. Dalam data training dapat diketahui nilai prior probabilitas yaitu dari probabilitas nilai excellent, very good, good dan bad masing masing atribut terhadap total keseluruhan data. Jumlah record dalam data training adalah 30 record (n) dengan 2 kasus excellent, 13 kasus very good, 11 kasus good dan 4 kasus bad.

Nilai prior probabilitas pada pemilihan mitra kerjasama adalah sebagai berikut:

$$P(\text{Excellent},n) = 2/30 = 0,07$$

$$P(\text{Very Good},n) = 13/30 = 0,43$$

$$P(\text{Good},n) = 11/30 = 0,37$$

$$P(\text{Bad},n) = 4/30 = 0,13$$

Tabel 4 Hasil Perhitungan Prior Probabilitas

No	Atribut/Variabel	Nilai Kategori	Jumlah Data	Excellent	Very Good	Good	Bad	P(B A)			
								Excellent	Very Good	Good	Bad
1.	Pengalaman Kerja Perusahaan		30	2	13	11	4	0,07	0,43	0,37	0,13
		<5 tahun	7	0	0	3	4	0,00	0,00	0,10	0,13
		5 – 10 tahun	10	1	7	2	0	0,03	0,23	0,07	0,00
		>10 tahun	13	1	6	6	0	0,03	0,20	0,20	0,00
2.	Masa Kerjasama		30	2	13	11	4	0,07	0,43	0,37	0,13
		<1 tahun	0	0	0	0	0	0,00	0,00	0,00	0,00
		1 – 5 tahun	22	2	8	8	4	0,07	0,27	0,27	0,13
		>5 tahun	8	0	5	3	0	0,00	0,17	0,10	0,00
3.	Bidang		30	2	13	11	4	0,07	0,43	0,37	0,13
		Penyedia Hardware	12	1	6	5	0	0,03	0,20	0,17	0,00
		Pengembangan Software	7	1	3	3	0	0,03	0,10	0,10	0,00
		Tenaga IT	6	0	3	1	2	0,00	0,10	0,03	0,07
		Penyedia Perlengkapan	5	0	1	2	2	0,00	0,03	0,07	0,07

4.	Pelayanan	30	2	13	11	4	0,07	0,43	0,37	0,13
	Mandiri	23	2	8	9	4	0,07	0,27	0,30	0,13
	Non Mandiri	7	0	5	2	0	0,00	0,17	0,07	0,00
5.	Skala Pelayanan	30	2	13	11	4	0,07	0,43	0,37	0,13
	Antar Kota	22	1	10	9	2	0,03	0,33	0,30	0,07
	Antar Propinsi	6	1	3	1	1	0,03	0,10	0,03	0,03
	Antar Wilayah Kepulauan	2	0	0	1	1	0,00	0,00	0,03	0,03
6.	Harga Pelayanan	30	2	13	11	4	0,07	0,43	0,37	0,13
	Murah	0	0	0	0	0	0,00	0,00	0,00	0,00
	Sedang	14	0	9	4	1	0,00	0,30	0,13	0,03
	Mahal	16	2	4	7	3	0,07	0,13	0,23	0,10
7.	Status Perusahaan	30	2	13	11	4	0,07	0,43	0,37	0,13
	Swasta Kecil	0	0	0	0	0	0,00	0,00	0,00	0,00
	Swasta Menengah	19	1	8	9	1	0,03	0,27	0,30	0,03
	Swasta Besar	11	1	5	2	3	0,03	0,17	0,07	0,10

Apabila ditemukan kasus baru yang mengharuskan perhitungan terhadap pemilihan mitra kerjasama, maka akan ditentukan berdasarkan pada tabel 3.4. Berikut adalah contoh apabila ditemukan kasus baru (*Posterior Data*):

Tabel 5 Probabilitas Posterior Data

Kasus Baru		P(A)			
Atribut	Nilai	Excellent	Very Good	Good	Bad
Pengalaman Kerja Perusahaan	16	0,03	0,20	0,20	0,00
Masa Kerjasama	3	0,07	0,27	0,27	0,13
Bidang	Tenaga IT	0,00	0,10	0,03	0,07
Pelayanan	Mandiri	0,07	0,27	0,30	0,13
Skala Pelayanan	Antar Kota	0,03	0,33	0,30	0,07
Harga Pelayanan	Mahal	0,07	0,13	0,23	0,10

Status Perusahaan	Swasta Menengah	0,03	0,27	0,30	0,03
Remark		0,30	1,57	1,63	0,53
P(B A)P(A)		0,02	0,68	0,60	0,07

Dari hasil perhitungan pada tabel 3.4 dan tabel 3.5 dengan nilai Remark Excellent=0,30 nilai Very Good=1,57 nilai Good=1,63 dan nilai Bad=0,53, maka didapatkan hasil perhitungan $P(B|A)P(A)$ tertinggi adalah “Very Good dengan nilai 0,68, sehingga dapat disimpulkan bahwa data baru termasuk kedalam klasifikasi Very Good dan dapat menjadi mitra kerjasama PT. Citra Pesona Gemilang.

E. Evaluation Phase

Tahapan berikutnya adalah tahap evaluasi dari hasil pengujian menggunakan algoritma C4.5 dan Naïve Bayes, dimana atribut remark dengan nilai terendah (excellent) mengalami eliminasi, sehingga yang diuji adalah nilai dari very good, good dan bad yang ditampilkan dalam bentuk confusion matrix sebagai berikut:

```

Correctly Classified Instances      34          97.1429 %
Incorrectly Classified Instances    1          2.8571 %
Kappa statistic                    0.9515
Mean absolute error                 0.0286
Root mean squared error             0.1195
Relative absolute error             7.1788 %
Root relative squared error         26.9177 %
Total Number of Instances          35

=== Detailed Accuracy By Class ===

      TP Rate  FP Rate  Precision  Recall  F-Measure  ROC Area  Class
      1       0.056    0.944      1     0.971    0.995  Verry Good
      0.929    0       1       0.929  0.963    0.995   Good
      1       0       1       1       1       1       Bad
Weighted Avg.   0.971  0.027    0.973  0.971  0.971    0.996

=== Confusion Matrix ===
  a  b  c  <-- classified as
17  0  0 | a = Verry Good
 1 13  0 | b = Good
 0  0  4 | c = Bad

```

Gambar 1 Confusion Matrix Algoritma C4.5

```

Correctly Classified Instances      33          94.2857 %
Incorrectly Classified Instances    2          5.7143 %
Kappa statistic                    0.9041
Mean absolute error                 0.0795
Root mean squared error             0.2059
Relative absolute error             19.9701 %
Root relative squared error         46.3737 %
Total Number of Instances          35

=== Detailed Accuracy By Class ===

      TP Rate  FP Rate  Precision  Recall  F-Measure  ROC Area  Class
      0.882    0       1       0.882  0.938    0.944  Verry Good
      1       0.095    0.875    1     0.933    0.939   Good
      1       0       1       1       1       1       Bad
Weighted Avg.   0.943  0.038    0.95   0.943  0.943    0.949

=== Confusion Matrix ===
  a  b  c  <-- classified as
15  2  0 | a = Verry Good
 0 14  0 | b = Good
 0  0  4 | c = Bad

```

Gambar 2 Confusion Matrix Algoritma Naïve Bayes

Perbandingan pengujian menggunakan algoritma C4.5 dengan algoritma Naïve Bayes dapat dilihat pada tabel komparasi Confusion Matrix sebagai berikut:

Tabel 3.6 Komparasi Nilai Accuracy, Precision, Recall

Komparasi	C4.5	Naïve Bayes
Accuracy	97,14	94,28
Precision	0,973	0,95
Recall	0,971	0,943

IV. KESIMPULAN

Kesimpulan :

1. Pemilihan vendor untuk mitra kerjasama PT Citra Pesona Gemilang dilakukan dengan menggunakan algoritma C4.5 dan *Naïve Bayes*.
2. Menentukan tingkat akurasi dari kedua algoritma untuk mengetahui kelayakan mitra kerjasama PT Citra Pesona Gemilang dan mengukur kinerja dari kedua algoritma tersebut menggunakan *confusion matrix* yang didapatkan hasil bahwa algoritma C4.5 memiliki nilai *Accuracy* lebih tinggi dibandingkan dengan algoritma *Naïve Bayes*.
3. Hasil yang didapat dari penelitian perbandingan algoritma C4.5 dan *Naïve Bayes* diharapkan mampu memberikan solusi terhadap permasalahan pemilihan vendor mitra kerjasama pada PT Citra Pesona Gemilang sehingga mampu meningkatkan keuntungan perusahaan dalam menjalankan bisnisnya.
4. Kedepannya diharapkan dilakukannya penelitian lanjutan untuk memperbaiki kelemahan pada penelitian kali ini, karena pada dasarnya dunia bisnis akan selalu berkembang mengikuti kemajuan teknologi informasi, tidak terkecuali pada PT Citra Pesona Gemilang.

REFERENSI

- [1]. Anggriawan. (2017). Gaya Komunikasi Pimpinan Terhadap Motivasi Kerja Karyawan Pada PT. Perusahaan Listrik Negara (PLN) Persero Area Pelayanan Di Samarinda. *eJournal Ilmu Komunikasi*.
- [2]. Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., & Wirth, R. (2000). *CRISP-DM 1.0: Step-By-Step Data Mining Guide*. USA: SPSS Inc.
- [3]. Eska. (2016). Penerapan Data Mining Untuk Prediksi Penjualan Wallpaper Menggunakan Algoritma C4.5. *JURTEKSI (Jurnal Teknologi dan Sistem Informasi)*.
- [4]. Han, Jiawei et al. (2012). *Data Mining Concepts and Techniques*. Waltham: Elsevier Inc.
- [5]. Hayuningtyas, R. Y. (2019). Penerapan Algoritma Naïve Bayes untuk Rekomendasi Pakaian Wanita. *Jurnal Informatika UBSI*.
- [6]. Kennardi, Akhmad Budi. (2016). Penerapan Data Mining Untuk Prediksi Nilai Tukar Petani Tanaman Pangan Di Indonesia Dengan Metode Linear Regression Dan Support Vector Machine. *Jurnal Informatika dan Bisnis*.
- [7]. Mardi, Y. (2016). Data Mining Klasifikasi Menggunakan Algoritma C4.5. *Jurnal Edik Informatika*.
- [8]. Marsani Asfi, Nopi Fitrianiingsih. (2020). Implementasi Algoritma Naïve Bayes Classifier sebagai Sistem Rekomendasi Pembimbing Skripsi. *InfoTekJar (Jurnal Nasional Informatika dan Teknologi Jaringan)*.
- [9]. O'Brien, James A. (2010). *Management Information System*. New York: Mc Graw Hill Irwin.
- [10]. Pramadani, Setiadi. (2014). Penerapan Data Mining Untuk Klasifikasi Prediksi Penyakit ISPA (Infeksi Saluran Pernapasan Akut) Dengan Algoritma Decision Tree. *Jurnal Sarjana Teknik Informatika*.
- [11]. Stanton, Jeffrey, Robert W. De Graaf. (2013). *An Introduction to Data Science*. New York: Creative Commons.
- [12]. Yogasetya Suhandi, Ike Kurniati, Siti Norma. (2020). Penerapan Metode Crisp-DM Dengan Algoritma K-Means Clustering Untuk Segmentasi Mahasiswa Berdasarkan Kualitas Akademik. *Jurnal Teknologi Informatika dan Komputer*.

