
KAJIAN PENERAPAN ALGORITMA C4.5, NAÏVE BAYE DAN NEURAL NETWORK UNTUK MEMENUHI PENILAIAN DATA KARYAWAN SERVICE LEVEL AGREEMENT DI BANK

ACHMAD FAISAL

echalleo@yahoo.com

Program Studi Informatika

Fakultas Teknik, Matematika dan Ilmu Pengetahuan Alam

Universitas Indraprasta PGRI

Jl. Nangka No. 58 C, Tanjung Barat, Jagakarsa, Jakarta Selatan 12530

Abstrak. Dunia Ekonomi dan Bisnis yang semakin kompetitif mendorong setiap *Organization Level* berpacu dengan waktu dalam memenuhi kebutuhan pelanggan (*Customer Needs*). Dalam bahasa *Management* kita kenal dengan istilah *Service Level Standard* dan dalam beberapa Aspek atau Kriteria juga disebut *Service Level Agreement*. Bank merupakan salah satu urat jantung perekonomian sebuah negara, tanpa Bank, bisa kita bayangkan bagaimana kita sulitnya menyimpan dan mengirim uang, memperoleh tambahan modal usaha atau melakukan transaksi perdagangan Internasional secara efektif dan aman. Suatu Bank akan dapat memberikan pelayanan yang baik/prima, apabila Bank tersebut memiliki standar layanan yang baik. Maka standar layanan harus dapat dilaksanakan dengan baik dan konsisten, maka dari itu diperlukan suatu alat yang dapat mendefinisikan dan memonitor layanan unit kerja dalam masalah pelayanan permasalahan. Salah satu media yang dapat digunakan untuk mendefinisikan dan memonitor layanan adalah dengan adanya *Service Level Agreement (SLA)*. SLA merupakan sistem yang dirancang untuk mengukur penilaian karyawan berdasarkan kinerja. Salah satu input yang digunakan dalam SLA adalah Penerapan KPI (*Key Performance Indicators*) yang diperoleh dari penerapan SLA berupa standar acuan, yang merupakan kesepakatan antara pihak yang membutuhkan pelayanan yang spesifik dan penyedia layanan yang bertanggung jawab untuk memberikan layanan tersebut. Dengan adanya SLA, penyedia layanan akan berkomitmen untuk memberikan layanan terbaik untuk mendukung kebutuhan bisnis nasabah. Dalam penelitian ini, dilakukan analisis komparasi tiga algoritma klasifikasi data mining yaitu algoritma C4.5, *Neural Network* dan *Naïve Bayes* dengan menggunakan 130 data set karyawan sehingga dapat diketahui algoritma yang paling akurat untuk memenuhi penilaian *Service Level Agreement*.

Kata Kunci : *classification, C4.5algorithm, Neural Network andNaiveBayes.*

Abstract. *The world economy and an increasingly competitive business encourages each Organization Level race against time to meet the needs of customers (Customer Needs). In the language we are familiar with the term Management Service Level Standard and in some aspects or criteria also called Service Level Agreement. A bank will be able to provide good service / prime, if the bank has a good standard of service. Then the service standards should be implemented properly and consistently, and therefore needs a tool that can define and monitor service unit in service problems. Currently there are many who still do not use a bank system that monitors the SLA with respect to support SLA success, it needs to be made or arranged a policy of SLAs that can be used as standards and guidelines, in the preparation of SLA and monitoring and evaluation of the implementation of the SLA. Purpose this study is the classification of the employee data is good or bad in the assessment, in this research, comparative algorithm C4.5, naïve Bayes and neural network is applied to employee data that is either in the assessment. This study aimed to measure the accuracy of a comparative study of three pieces in a classification algorithm in the assessment of bad employees. From the test results to measure the performance of the three algorithms using Cross Validation testing methods, Confusion Matrix and ROC curves, it is known that the C4.5 algorithm has the highest accuracy value, that is 88.00%, followed by the methods of Naïve Bayes with the accuracy of 86.00% and the lowest is the Neural Network method with 86.15% accuracy values.*

Keywords: classification, C4.5algorithm, Neural Network andNaiveBayes.

PENDAHULUAN

Dunia Ekonomi dan Bisnis yang semakin kompetitif mendorong setiap *Organization Level* berpacu dengan waktu dalam memenuhi kebutuhan pelanggan (*Customer Needs*). Dalam bahasa *Management* kita kenal dengan istilah *Service Level Standard* dan dalam beberapa Aspek atau Kriteria juga disebut *Service Level Agreement*. Bank merupakan salah satu urat jantung perekonomian sebuah negara, tanpa Bank, bisa kita bayangkan bagaimana kita sulitnya menyimpan dan mengirimkan uang, memperoleh tambahan modal usaha atau melakukan transaksi perdagangan Internasional secara efektif dan aman. Suatu Bank akan dapat memberikan pelayanan yang baik/prima, apabila Bank tersebut memiliki standar layanan yang baik. Maka standar layanan harus dapat dilaksanakan dengan baik dan konsisten, maka dari itu diperlukan suatu alat yang dapat mendefinisikan dan memonitor layanan unit kerja dalam masalah pelayanan permasalahan. Dengan diimplementasikannya modul ini, setiap karyawan dapat melakukan kegiatannya akan selalu berorientasi kepada peningkatan layanan sehingga dapat memenuhi SLA/KPI-nya, karena dengan demikian akan mempengaruhi kinerja di unit kerja. Penilaian Kinerja dilakukan dengan menggunakan Form yang telah ditetapkan dan diimplementasi melalui system komputersisasi secara terpusat.dalam *database*.

METODE

Menurut Clifford Woody dalam (Khothari, 2004), penelitian terdiri dari mendefinisikan dan mendefinisikan ulang masalah, merumuskan hipotesis atau menyarankan solusi, mengumpulkan, mengorganisir dan mengevaluasi data, membuat deduksi dan mencapai kesimpulan, dan berakhir pada menguji kesimpulan untuk menentukan apakah hasil dari penelitian tersebut sesuai dengan rumusan hipotesis.

Penelitian eksperimen menggunakan data dalam penelitiannya dan menghasilkan kesimpulan yang mampu dibuktikan oleh pengamatan atau percobaan. Dalam penelitian eksperimen, peneliti harus memberikan hipotesis kerja atau membuat dugaan pada kemungkinan hasil. Setelah itu peneliti bekerja mendapatkan data untuk membuktikan atau menyangkal hipotesis. Kemudian peneliti membangun desain eksperimen yang menurutnya akan dapat memanipulasi orang atau data yang bersangkutan sehingga memperoleh informasi yang dikehendaki (Kothari, 2004). Metode eksperimen menguji kebenaran sebuah hipotesis dengan statistik dan menghubungkannya dengan masalah penelitian. Desain eksperimen dibagi menjadi dua yaitu eksperimen absolut dan eksperimen komparatif. Eksperimen absolut mengarah terhadap dampak atau akibat yang dihasilkan dari eksperimen, sedangkan eksperimen komparatif membandingkan objek yang berbeda, misalnya membandingkan dua algoritma yang berbeda dengan melihat hasil statistik masing-masing algoritma tersebut manakah yang lebih baik (Kothari, 2004).

Studi pendahuluan kegiatan yang dilakukan pada saat studi pendahuluan yaitu mengumpulkan materi-materi kepustakaan yang berhubungan dengan pengambilan judul. Kemudian langkah selanjutnya yaitu survey langsung ke tempat penelitian Tata Usaha Fakultas Teknik Matematika dan Ilmu Pengetahuan Alam serta BAAK Universitas Indraprasta PGRI.

Data setelah dilakukannya survey maka didapatkan data yang akan digunakan dalam penelitian ini, yang terdiri dari beberapa atribut yaitu data mahasiswa dan lain-lain.

Pengolah Data setelah mendapatkan data, maka data diolah menggunakan metode data mining, yaitu pertama kali menggunakan aplikasi *microsoft office excel*.

Laporan setelah ketiga tahap diatas dilakukan maka disusunlah laporan penelitian ini. Kedalam bentuk tesis, sebagai pemenuhan persyaratan Magister Ilmu Komputer di STMIK Nusa Mandiri.

HASIL DAN PEMBAHASAN

Setelah tinjauan system selesai dilakukan, maka mendapatkan gambaran yang jelas apa yang harus dikerjakan. Untuk dapat mencapai keinginan yang dimaksud maka perlu dilakukan suatu

perancangan sistem. Tahap rancangan system ini merupakan prosedur untuk mengkonversi spesifikasi logis kedalam sebuah program penilaian karyawan

Memahami Situasi Masalah Yang Sedang Berjalan

Berdasarkan identifikasi masalah dari penelitian ini yang penulis lakukan, di dapat beberapa permasalahan diantaranya :

1. Pada setiap bulan, triwulan dan semester jumlah nilai karyawan mengalami penurunan. Untuk dapat meningkatkan nilai karyawan pada setiap periode, perlu diketahui permasalahan-permasalahan yang terjadi pada diri karyawan terutama karyawan yang bermasalah dalam bekerja.
2. Dengan kajian komparasi algoritma ini diharapkan dapat membantu pihak Bank dalam memprediksi nilai karyawan yang mengalami penurunan.

Rancangan Sistem

Adapun tujuan dari Rancangan penelitian ini adalah mengklasifikasi penilaian SLA yang mengalami penurunan nilai dengan algoritma *C4.5*, *Neural Network* dan *Naïve Bayes* serta mengetahui algoritma manakah yang memiliki tingkat akurasi paling tinggi dalam klasifikasi penurunan nilai SLA karyawan.

Algoritma C4.5

Berikut ini adalah langkah-langkah klasifikasi mahasiswa dengan algoritma C4.5.

1. Siapkan data *training* yaitu tabel 3.3 yang berjumlah 747 data.
2. Hitung jumlah kelas registrasi dan bermasalah berdasarkan nilai tiap atribut.
3. Hitung nilai *entropy* total dimana diketahui jumlah kelas yang registrasi berjumlah 656 dan kelas bermasalah berjumlah 91.

$$\begin{aligned} Entropy(S) &= \sum_{i=1}^n - p_i \cdot \log_2 p_i \\ &= (-100/130) * \log_2 (100/130) + (-91/130) * \log_2 (91/130) \\ &= 0,469 \end{aligned}$$

4. Hitung nilai *gain* untuk masing-masing atribut. Kemudian tentukan nilai *gain* tertinggi. Atribut dengan nilai *gain* tertinggi maka atribut tersebut dijadikan sebagai akar. Sebagai contoh hitung nilai *gain* untuk atribut jenis kelamin yaitu:

$$\begin{aligned} Gain(S, A) &= Entropy(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} * Entropy(S_i) \\ &= 0,469 - ((53/100 * 0,510) + (47/100 * 0,420)) = 0,00158935 \end{aligned}$$

Perhitungan nilai *entropy* dan *gain* untuk semua atribut dilakukan untuk mendapatkan nilai *gain* tertinggi yang akan dijadikan sebagai akar. Hasil perhitungannya terlihat di tabel 4.

Tabel 4 Hasil perhitungan nilai *entropy* dan *gain* untuk menentukan simpul akar.

Node		Jumlah Kasus (S)	Baik (S1)	Buruk (S2)	Entropy	Gain
1	Total	100	90	10	0,469	
	Jenis Kelamin					0,00158935
	Laki-Laki	53	47	6	0,510	
	Perempuan	47	43	4	0,420	
	Status pekerja					0,00125739
	Tetap	66	60	6	0,439	
	Kontrak	34	30	4	0,523	
	Pemahaman jasa Produk dan Jasa					0,00361910
	Baik	45	41	4	0,433	
	Buruk	21	18	3	0,592	
	Sangat Baik	34	31	3	0,431	
	Komunikasi					0,00014885
	Aktif	68	61	7	0,478	
	pasif	32	29	3	0,449	
	Kemampuan Kerja					0,14013976
	cukup	8	7	1	0,544	
	kurang	47	40	7	0,607	
	rendah	3	3			
	sangat mampu	39	37	2	0,292	
	tinggi	3	3			
	Kecermatan					0,12473316
	<51	3	1	2	0,918	
	>=90	62	60	2	0,206	
	61-70	30	27	3	0,469	

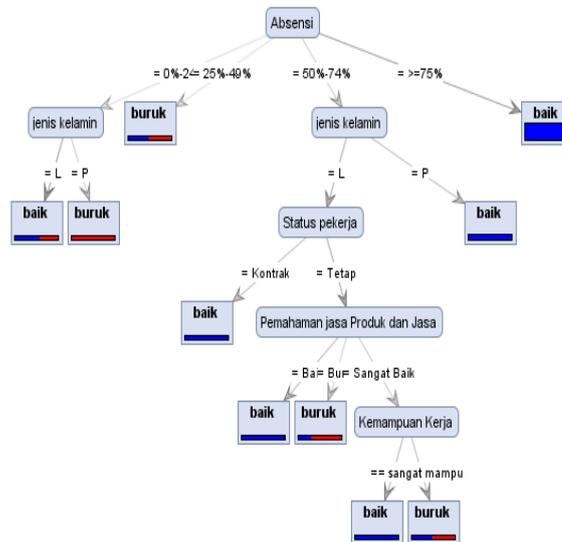
Berdasarkan hasil perhitungan *gain* pada tabel 4 terlihat atribut nilai smtr 1 mempunyai nilai *gain* tertinggi yaitu **0.14013976** sehingga atribut Nilai Karyawan dijadikan sebagai simpul akar dari pohon keputusan.

Tentukan simpul selanjutnya setelah simpul akar yaitu *node* 1.1 dengan menghitung nilai *entropy* dan *gain* semua atribut berdasarkan atribut Kemampuan Kerja dengan nilai (8) jumlah data, (7) jumlah Baik, (1) Buruk yang mempunyai nilai *entropy* dan *gain* 0,5435644.

Tabel 5 Perhitungan *entropy* dan *gain* untuk menentukan *node* 1.1

		Jumlah baik (S1)	buruk (S2)	Entropy	Gain
Kemampuan Kerja		8	7	1	0,5435644
Cukup					0,056047627
Jenis Kelamin	l	6	5	1	0,6500224
	p	2	2	0	0
Status pekerja					0,253323431
	Tetap	5	5	1	0,4643856
	Kontrak	3	3	0	0
Pemahaman jasa Produk dan Jasa					0,199203505
	Baik	3	3	0	0
	Buruk	3	2	1	0,9182958
	Sangat Baik	2	2	0	0
Komunikasi					0,293564443
	Aktif	4	4	0	0
	pasif	4	4	1	0,5
Kecermatan					0,092359384
	<51	1	1	0	0
	>=90	2	2	0	0
	61-70	5	4	1	0,7219281
	71-80	0	0	0	0
Absensi					0,199203505
	>=75%	3	2	1	0,9182958
	0%-24%	1	1	0	0
	25%-49%	1	1	0	0
	50%-74%	3	3	0	0
Penyelesaian pekerjaan					0,543564443
	Kurang	0	0	1	0
	Mampu	8	7	0	0

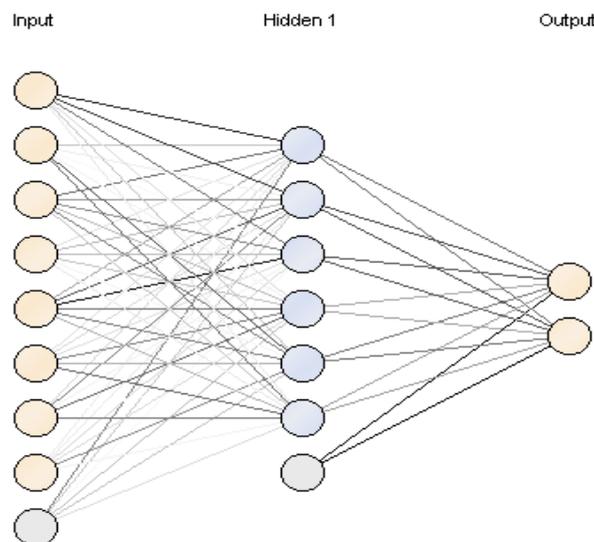
Pembentukan simpul-simpul dengan perhitungan *gain* diperoleh *decision tree* untuk klasifikasi mahasiswa bermasalah dalam registrasi terlihat seperti pada gambar 6.



Gambar 6 *Decision Tree* klasifikasi mahasiswa dengan C4.5

Algoritma Neural Network

Klasifikasi karyawan dengan *Neural Network* menggunakan data *training* yang sama dengan data *training* yang digunakan untuk klasifikasi data karyawan dengan algoritma C4.5 yaitu tabel 3.3 sejumlah 130 *record* data. Hasil pengolahan data *training* tersebut diperoleh algoritma C4.5 seperti pada gambar 4.2 dengan menggunakan metode algoritma C4.5 menghasilkan tiga *layer*, yaitu *input layer* yang terdiri dari delapan simpul yang terdiri dari delapan simpul yang sama dengan jumlah atribut *predictor* dan satu buah simpul bias. *Hidden layer* terdiri dari tujuh simpul yang terdiri dari delapan simpul ditambah satu simpul bias. *Output layer* yang merupakan hasil klasifikasi terdiri dari dua simpul yaitu baik dan Buruk.



Gambar. 7 *Neural Net* yang dihasilkan dengan algoritma *Neural Network*

Setiap data pada data *training*, dihitung input untuk simpul berdasarkan nilai input dan jaringan saat itu. Bobot awal untuk *input layer*, *hidden layer*, dan bias diinisialisasi secara acak (biasanya antara -0.1 sampai dengan 1.0). Simpul bias terdiri dari dua, yaitu pada *input layer* yang terhubung dengan simpul-simpul pada *hidden layer*, dan pada *hidden layer* yang terhubung pada *output layer*. Setelah semua nilai awal diinisialisasi, kemudian dihitung masukan, keluaran, dan *error*. Selanjutnya membangkitkan *output* untuk simpul menggunakan fungsi aktivasi sigmoid. Setelah didapat nilai dari fungsi aktivasi, hitung nilai *error* antara nilai yang diprediksi dengan nilai yang sesungguhnya. Setelah nilai *error* dihitung, selanjutnya dibalik ke *layer* sebelumnya (*backpropagated*). Nilai *Error* yang dihasilkan dari langkah sebelumnya digunakan untuk memperbarui bobot relasi. Hasil perhitungan akhir *backpropagation* fungsi aktivasi untuk simpul pada *hidden layer* terdapat pada Tabel 6. Kolom pertama pada Tabel 6 merupakan atribut yang dinyatakan berupa simpul pada *input layer* seperti pada Gambar 7. Sedangkan Kolom satu sampai delapan mewakili jumlah simpul pada *hidden layer*.

Tabel 6 Nilai bobot akhir Hidden Layer

No	Simpul	Hidden Layer (sigmoid)					
		1	2	3	4	5	6
1	JK	2.463	-0.637	-1.139	2.701	-1.265	-1.731
2	Status Pekerjaan	0.749	-2.928	-2.615	1.954	1.414	1.931
3	Pemahaman jasa produk dan jasa	-1.869	2.551	3.691	-2.014	0.786	0.407
4	Komunikasi	-1.602	1.768	0.574	-2.997	-0.370	0.062
5	Kemampuan Kerja	-0.306	1.077	0.775	-2.881	-0.095	0.622
6	Kecamatan	-2.512	-1.292	-2.091	-5.953	-3.918	-4.965
7	Absensi	-0.87	1.692	1.876	-1.873	-0.346	-0.439
8	Penyelesaian Pekerjaan	0.879	0.923	0.714	1.492	1.169	0.693
9	threshold	-0.547	-0.554	-0.266	-1.117	-0.372	0.068

Pada tabel 6 kolom simpul menerangkan atribut yang dinyatakan kolom simpul *input layer*. Sedangkan kolom satu sampai delapan menerangkan jumlah simpul pada *hidden layer*. Untuk nilai akhir fungsi aktivasi *output layer* dapat dilihat pada tabel 7 dibawah ini:

Tabel 7 Nilai bobot akhir Output Layer

Class	Output (sigmoid)						
	1	2	3	4	5	6	threshold
Baik	0.068	3.646	3.543	4.497	1.557	2.447	-4.369
Buruk	-1.627	-3.645	-3.545	-4.476	-1.545	-2.464	4.372

Pada tabel 7 Class terdiri dari Baik dan Buruk yang merupakan nilai hasil klasifikasi. Nilai yang terdapat pada kolom satu sampai delapan merupakan nilai bobot akhir pada *output layer*.

Algoritma Naïve Bayes

Penggunaan algoritma *Naïve Bayes* menggunakan data *training* pada Tabel 3.3 dimulai dengan melakukan perhitungan *probabilitas prior* untuk mengetahui nilai yang diterima dan tidak diterima untuk semua jumlah data. Pada data training jumlah data sebanyak 100 data, dimana kelas baik sebanyak 90 record dan yang Buruk sebanyak 10 record. Berikut hasil perhitungan *prior probability* terlihat pada tabel 8.

Tabel 8. Perhitungan Probabilitas *Prior*

Atribut		Kasus	Baik	Buruk	p(x c1)	
					Baik	Buruk
Total		100	90	10	0,9	0,1
Jenis Kelamin						
	Laki-Laki	53	47	6	0,886792453	0,113207547
	Perempuan	47	43	4	0,914893617	0,085106383
Status pekerja						
	Tetap	34	60	6	1,764705882	0,176470588
	Kontrak	66	30	4	0,454545456	0,060606061
Pemahaman Jasa Produk dan Jasa						
	Baik	45	41	4	0,911111111	0,088888889
	Buruk	21	18	3	0,857142857	0,142857143
	Sangat Baik	34	31	3	0,911764706	0,088235294
Komunikasi						
	Aktif	68	61	7	0,897058824	0,102941176
	pasif	32	29	3	0,90625	0,09375
Kecepatan Kerja						
	cukup	8	7	1	0,875	0,125
	kurang	47	40	0	0,85106383	0
	rendah	3	3	7	1	2,333333333
	sangat mampu	39	37	2	0,948717949	0,051282051
	tinggi	3	3	0	1	0
Kecermatan						
	<51	3	1	2	0,333333333	0,666666667
	>=90	62	60	2	0,967741936	0,032258066
	61-70	30	27	3	0,9	0,1
	71-80	5	2	3	0,4	0,6
Absensi						
	>=75%	65	63	2	0,969230769	0,030769231
	0%-24%	7	3	4	0,428571429	0,571428571
	25%-49%	4	2	2	0,5	0,5
	50%-74%	24	22	2	0,916666667	0,083333333
Penyelesaian pekerjaan						
	Kurang	7	83	10	11,85714286	1,428571429
	Mampu	93	7	0	0,075268817	0

Untuk menentukan nilai baru Baik atau Buruk, dilakukan perhitungan probabilitas *posterior* berdasarkan probabilitas *prior* yang telah dihitung sebelumnya pada Tabel 8. Perhitungan probabilitas *posterior* untuk menentukan data testing termasuk klasifikasi mana terdapat pada Tabel 9. Misalkan diambil sebuah data testing X dengan nilai seperti pada Tabel 9 kolom dua, untuk menentukan kelas mana, dilakukan perhitungan probabilitas *posterior* yang hasilnya terdapat pada Tabel 9 kolom tiga dan empat.

Tabel 9 Perhitungan Data Buruk Untuk Penerapan Algoritma Terpilih

	A	B	C	D	E	F	G	H	I
1	jenis kelamin	Status pekerja	Pemahaman jasa Produk dan Jasa	Komunikasi	Kemampuan Kerja	Kecermatan	Absensi	Penyelesaian pekerjaan	Remark
2	P	Kontrak	Buruk	Aktif	Rendah	<51	>=75%	Kurang	baik
3	L	Kontrak	Baik	Pasif	kurang	>=90	>=75%	Kurang	baik
4	L	Kontrak	buruk	Pasif	kurang	<51	>=75%	Kurang	buruk
5	L	Kontrak	Sangat Baik	Aktif	sangat mampu	61-70	0%-24%	mampu	baik
6	L	Kontrak	Buruk	Aktif	Cukup	<51	50%-74%	mampu	baik
7	P	Tetap	Baik	Aktif	Tinggi	71-80	25%-49%	mampu	baik
8	P	Tetap	Sangat Baik	Aktif	sangat mampu	>=90	>=75%	mampu	baik
9	P	Tetap	Buruk	Aktif	Cukup	61-70	0%-24	mampu	baik
10	L	Tetap	Buruk	Pasif	kurang	71-80	0%-24%	Kurang	Buruk
11	P	Tetap	Baik	Aktif	sangat mampu	>=90	>=75%	Kurang	baik

Evaluasi dan Validasi Model

Hasil dari pengujian model yang telah dilakukan yaitu dengan algoritma C4.5, *Neural Network*, dan *Naïve Bayes*, dilakukan pengujian tingkat akurasi dengan menggunakan *confussion matrix* dan kurva ROC/AUC (*Area Under Cover*).

Confusion Matrix

Tabel 10 merupakan perhitungan akurasi data *training* menggunakan algoritma C4.5. Diketahui data *training* terdiri dari 100 *record* data, 85 data diklasifikasikan baik dan 7 data diprediksi baik tetapi ternyata Buruk, 5 data dinyatakan buruk tetapi baik, dan, 3 data secara benar diklasifikasikan Buruk.

Tabel 10 Model *Confusion Matrix* data *training* untuk Metode C4.5

accuracy: 88.00% +/- 6.00% (mikro: 88.00%)			
	true baik	true buruk	class precision
pred. baik	85	7	92.39%
pred. buruk	5	3	37.50%
class recall	94.44%	30.00%	

Tabel 11. merupakan perhitungan akurasi data *testing* menggunakan algoritma C4.5. Diketahui data *testing* terdiri dari 30 *record* data, 26 data diklasifikasikan baik dan 5 data diprediksi baik tetapi ternyata Buruk, 2 data diprediksi buruk tapi baik dan 1 diprediksi buruk.

Tabel 11 Model *Confusion Matrix* data *testing* untuk Metode C4.5

accuracy: 80.00% +/- 13.54% (mikro: 79.41%)			
	true baik	true buruk	class precision
pred. baik	26	5	83.87%
pred. buruk	2	1	33.33%
class recall	92.86%	16.67%	

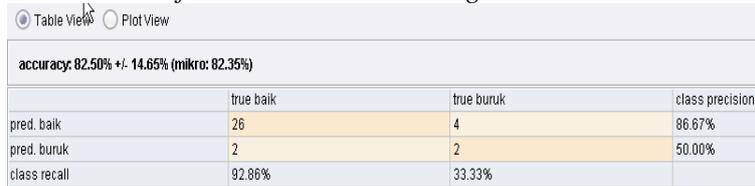
Tabel 11 merupakan perhitungan akurasi data *training* menggunakan algoritma *Neural Network*. Diketahui data *training* terdiri dari 100 *record* data, 84 data diklasifikasikan baik dan 8 data diprediksi baik tetapi ternyata Buruk, 6 data diklasifikasikan buruk tetapi ternyata Baik dan 2 data secara diklasifikasikan buruk.

Tabel 12 Model *Confusion Matrix* data *training* untuk Metode *Neural Network*

accuracy: 86.00% +/- 9.17% (mikro: 86.00%)			
	true baik	true buruk	class precision
pred. baik	84	8	91.30%
pred. buruk	6	2	25.00%
class recall	93.33%	20.00%	

Tabel 12 adalah *confusion matrix* untuk data *testing* dengan metode *neural network*. Diketahui dari 30 data *testing*, 26 diklasifikasikan baik, 4 data diprediksi baik tetapi ternyata Buruk, 2 data diprediksi dengan buruk klasifikasi Baik, dan 2 data diprediksi Buruk.

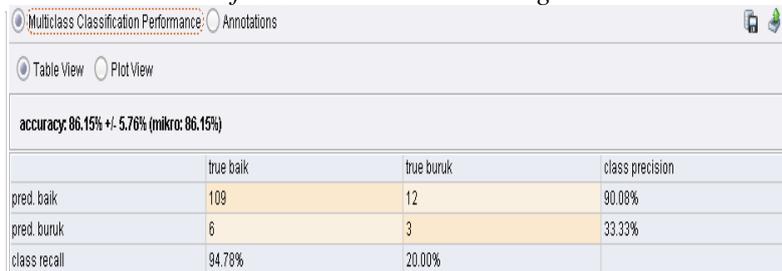
Tabel 13 Model *Confusion Matrix* data *testing* untuk Metode *Neural Network*



	true baik	true buruk	class precision
pred. baik	26	4	86.67%
pred. buruk	2	2	50.00%
class recall	92.86%	33.33%	

Tabel 14 merupakan perhitungan akurasi data *training* menggunakan algoritma *Naïve Bayes*. Diketahui data *training* terdiri dari 100 *record* data, 119 data diklasifikasikan Baik dan 12 data diprediksi benar tetapi ternyata Buruk, 6 data secara buruk diklasifikasikan Baik dan 3 data diprediksi buruk.

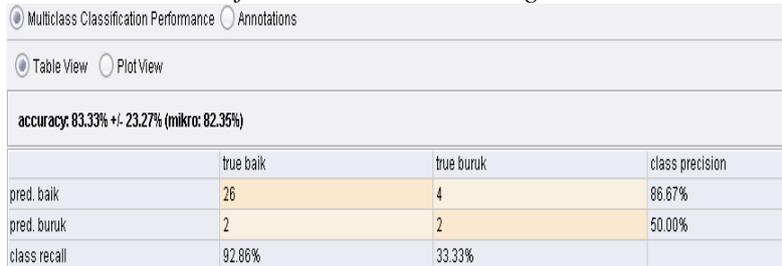
Tabel 14 Model *Confusion Matrix* data *training* untuk Metode *Naïve Bayes*



	true baik	true buruk	class precision
pred. baik	109	12	90.08%
pred. buruk	6	3	33.33%
class recall	94.78%	20.00%	

Tabel 15 merupakan perhitungan akurasi data *testing* menggunakan algoritma *Naïve Bayes*. Diketahui data *testing* terdiri dari 30 *record* data, 26 data diklasifikasikan baik dan 4 data diprediksi benar tetapi ternyata Buruk, 2 data secara benar diklasifikasikan Baik dan 2 data diprediksi Buruk..

Tabel 15 Model *Confusion Matrix* data *testing* untuk Metode *Naïve Bayes*



	true baik	true buruk	class precision
pred. baik	26	4	86.67%
pred. buruk	2	2	50.00%
class recall	92.86%	33.33%	

Dari hasil *confusion matrix* diatas, selanjutnya dilakukan perhitungan nilai *accuracy*, *precision*, dan *recall*. Perbandingan nilai *accuracy*, *precision*, dan *recall* yang telah dihitung untuk metode C4.5, *naïve bayes*, dan *neural network* dapat dilihat pada 16.

Tabel 16 Perbandingan Nilai *Accuracy*, *Precision*, dan *Recall*

	C4.5		<i>Neural Network</i>		<i>Naïve Bayes</i>	
	<i>training</i>	<i>testing</i>	<i>training</i>	<i>testing</i>	<i>training</i>	<i>testing</i>
<i>Accuracy</i>	88.00	80.00	86,00	82.50	86,15	83.33
<i>Precision</i>	37.50	33.33	25.00	50.00	33.33	50,00
<i>Recall</i>	30.00	16.67	20.00	33.33	20.00	33.33

Kurva ROC

Hasil perhitungan divisualisasikan dengan kurva ROC (*Receiver Operating Characteristic*) atau AUC (*Area Under Curve*). ROC memiliki tingkat nilai diagnosa yaitu (Gorunescu, 2011):

- a. Akurasi bernilai 0.90 – 1.00 = *excellent classification*
- b. Akurasi bernilai 0.80 – 0.90 = *good classification*
- c. Akurasi bernilai 0.70 – 0.80 = *fair classification*
- d. Akurasi bernilai 0.60 – 0.70 = *poor classification*
- e. Akurasi bernilai 0.50 – 0.60 = *failure*

Perbandingan hasil perhitungan nilai AUC untuk metode C4.5, *Neural Network*, dan *Naïve Bayes* dapat dilihat pada Tabel 17.

Tabel 17 Perbandingan Nilai AUC

	C4.5		<i>Neural Network</i>		<i>Naïve Bayes</i>	
	<i>training</i>	<i>testing</i>	<i>training</i>	<i>testing</i>	<i>training</i>	<i>testing</i>
AUC	0.772	0.859	0.496	0.444	0.587	0.679

Analisis Evaluasi Komparasi Model

Berdasarkan pengujian dan evaluasi hasil klasifikasi dengan algoritma C4.5, *Neural Network*, dan *Naïve Bayes* pada tabel 16, dapat kita lihat hasilnya yaitu tingkat akurasi menggunakan data *training* tertinggi adalah dengan algoritma C4.5 memiliki tingkat ROC paling tinggi, yaitu 0.772 dan pada *testing* algoritma C4.5 memiliki tingkat ROC yang paling tinggi, yaitu 0.859 termasuk dalam katagori *good classification*.

Berdasarkan kolom ROC pada tabel 16, pada data *training* algoritma C4.5 memiliki tingkat ROC paling tinggi, yaitu 0.772 dan pada *testing* algoritma C4.5 memiliki tingkat ROC yang paling tinggi, yaitu 0.859 termasuk dalam katagori *good classification*.

Dengan menggunakan perbandingan data *training* dengan data *testing*, yaitu 100 berbanding 30, maka perbandingan komparasinya untuk akurasi dapat dilihat dalam tabel 17:

Tabel 17 Perbandingan Akurasi

Metode	<i>Confussion Matrix</i>		Perbandingan Komparasi
	<i>Training</i>	<i>Testing</i>	
<i>Algoritma C4.5</i>	88.00%	80.00%	88.00%
<i>Neural Network</i>	86.00%	82.50%	86.00%
<i>Naïve Bayes</i>	86.15%	83.33%	86.15%

Berdasarkan hasil perbandingan akurasi pada tabel 4.15, Algoritma C4.5 memiliki tingkat akurasi, sehingga bisa digunakan untuk klasifikasi data karyawan yang bermasalah dalam penilaian dengan persentase 88.00%.

**Hasil dari Perancangan system dengan menampilkan :
Tampilan**

The screenshot shows a web application window titled "Data Klasifikasi Karyawan". The main content area is titled "Data Klasifikasi Karyawan" and contains a form with the following fields, each with a dropdown arrow:

- Jenis Kelamin
- Status Pekerjaan
- Pemahaman Jasa Produk Dan Jasa
- Komunikasi
- Kemampuan Kerja
- Kecermatan
- Absensi
- Penyelesaian Pekerjaan

Below the form are three buttons: "Tambah", "Proses", and "Keluar". To the right of the form is an illustration of a worker with gears and the text "XXXXXX" and "Klasifikasi Nilai karyawan".

PENUTUP

Simpulan

Dalam penelitian ini dilakukan pembuatan model menggunakan algoritma C4.5, *naïve bayes* dan *neural network* menggunakan data karyawan. Model yang dihasilkan, dikomparasi untuk mengetahui algoritma yang paling baik dalam mengatasi masalah pada karyawan yang mengalami nilai buruk. Untuk mengukur kinerja ketiga algoritma tersebut digunakan metode pengujian *Confusion Matrix* dan Kurva ROC, diketahui bahwa algoritma C4.5 memiliki nilai *accuracy* dan AUC paling tinggi.

Dengan demikian algoritma C4.5 merupakan algoritma terbaik dan dapat memberikan pemecahan dalam permasalahan karyawan yang mengalami nilai yang buruk.

Saran

1. Hasil penelitian ini diharapkan bisa digunakan pada pihak perusahaan, untuk lebih meningkatkan akurasi analisa data karyawan yang bermasalah dalam penilaian.
2. Model klasifikasi karyawan yang bermasalah dalam penilaian diharapkan dapat diterapkan pada sistem sehingga dapat dijadikan pendukung pengambilan kebijakan pihak manajemen untuk membantu mengurangi kesalahan pada pekerjaan
3. Untuk mendukung pengambilan keputusan dan pengembangan system informasi manajemen strategik, model ini dapat diterapkan pada perusahaan dengan menerapkan system yang menggunakan perangkat keras dan perangkat lunak, disertai dengan Penerapan KPI (*Key Performance Indicators*) yang diperoleh dari SLA (*Service Level Agreement*).

DAFTAR PUSTAKA

- Alpayadin, Ethem. 2010 **Introduction to Machine Learning**, The MIT Press, London, 2010
- Bramer, Max. 2007. **Principles of Data Mining**. London: Springer. ISBN-10: 1-84628-765-0, ISBN-13: 978-1-84628-765-7.
- Dua, S. & Xian Du. 2011. **Data Mining and Machine Learning in Cybersecurity**. USA: Taylor & Francis Group. ISBN-13: 978-1-4398-3943-0
- Gorunescu, F. 2011. **Data Mining Concept Model and Techniques**. Berlin: Springer. ISBN 978-3-642-19720-8
- Guidici, P. & Figini, S. 2009. **Applied Data Mining for Business and Industry (2nd ed)**. Italy. John Wiley & Sons, Ltd. ISBN: 978-0-470-05886-2

- Han, J., & Kamber, M. 2006. **Data Mining Concept and Tehniques**. San Fransisco: Morgan Kauffman. ISBN 13: 978-1-55860-901-3
- Larose, D. T. 2005. **Discovering Knowledge in Data**. New Jersey: John Willey & Sons, Inc. ISBN 0-471-66657-2.
- Meutia Agustina, Dwi. 2013, **Informasi Penilaian kinerja Pegawai Pada Badan Kepegawaian dan Diklat** Surabaya.
- Liao, T. W., & Triantaphyllou, E. 2007. **Recent Advances in Data Mining of Enterprise Data: Algorithms and Applications**. *Series on Computers and Operations Research, Vol 6*. USA: World Scientific. ISBN-13 978-981-277-985-4, ISBN-10 981-277-985-X
- Maimon, Oded., & Rokach, Lior. 2010. **Data Mining and Knowledge Discovery Handbook, 2nd Edition**. New York: Springer. ISBN 978-0-387-09822-7 Quinlan, (1993). *Algorithms C4.5*.
- Vercellis, Carlo. 2009. **Business Intelligence: Data Mining and Optimization for Decision Making**. United Kingdom: John Willey & Son
- Witten, I. H., Frank, E., Hall, M. A. 2011. **Data Mining Practical Machine Learning Tools and Techniques (3rd ed)**. USA: Elsevier
- Xindong Wu., & Vipin Kumar. 2009. **The Top Ten Algorithms in Data Mining**. New York: CRC Press.
- Sulistiono, Heru. 2013, **Kajian Penerapan Algoritma C4.5, Neural Network dan Neural Network Untuk Klasifikasi Mahasiswa Yang Bermasalah Dalam Registrasi : Studi Kasus Informatika Universitas Indraprasta PGRI**
- Xindong Wu., & Vipin Kumar. 2009. **The Top Ten Algorithms in Data Mining**. New York: CRC Press.