

Implementasi Algoritma C5.0 pada Penilaian Kinerja Pegawai Negeri Sipil

Putu Wiryastawan¹, Dewa Made Wiharta², I Made Sudarma³

Abstract— Employees become the spearhead of every Company, whether it is run the business on manufacture or service. The position and role of public government employees as an element of civil servant obligate them to provide fair public services. Refers to those facts, the performance of public government employees needs to be well managed. The algorithm C5.0 is one of the decision tree algorithm which can process the employees performance data become an input for decision-making. Based on evaluation result of 184 employees performance data, there was a high accuracy data in level 96.08%. Due to that result, the algorithm system can be developed become e-performance system which can predict or giving an advice in order to decision-making processes whether for assigning promotion, ranking or giving performance allowances.

Intisari— Dalam sebuah perusahaan, baik itu perusahaan produksi maupun jasa, pegawai menjadi ujung tombak dari perusahaan. Kedudukan dan peranan pegawai negeri sipil sebagai unsur aparatur negara yang bertugas sebagai abdi masyarakat, harus menyelenggarakan pelayanan secara adil kepada masyarakat. Dengan peranan yang sangat penting inilah maka kinerja pegawai perlu dikelola secara baik. Algoritma C5.0 merupakan salah satu algoritma pohon keputusan yang dapat memproses data kinerja pegawai menjadi sebuah aturan yang bisa dijadikan masukan dalam pengambilan keputusan. Berdasarkan hasil evaluasi terhadap 184 data kinerja pegawai didapat tingkat akurasi sebesar 96.08%. Dengan tingkat akurasi yang tinggi bisa dikembangkan menjadi sebuah sistem ekinerja yang dapat memberikan prediksi atau masukan dalam membuat keputusan baik untuk rekomendasi jabatan, kepangkatan maupun pemberian tunjangan kinerja.

Kata Kunci— data mining, algoritma C5.0, decision tree, kinerja pegawai.

I. PENDAHULUAN

Pegawai negeri sipil sebagai abdi masyarakat, harus menyelenggarakan pelayanan secara adil kepada masyarakat. Dengan peranan yang sangat penting inilah maka kinerja pegawai perlu dikelola secara baik. Kinerja pegawai perlu diukur untuk menentukan apakah profesionalisme pegawai telah tercapai secara maksimal. Penilaian kinerja pegawai telah dilakukan oleh berbagai organisasi sejak lampau.

¹Staff BKD Provinsi Bali, mahasiswa Pasca Sarjana, Jurusan Manajemen Sistem Informasi dan Komputer Universitas Udayana, Jalan PB Sudirman Denpasar Bali, 80232, INDONESIA (telp: 0361-239599; fax: -; e-mail: wiryastawan@gmail.com)

^{2, 3}Dosen, Magister Teknik Elektro Jalan PB Sudirman Denpasar Bali INDONESIA (telp: 0361-239599; fax: -; e-mail: wiharta@unud.ac.id, msudarma@unud.ac.id)

Penilaian kinerja diperlukan untuk proses kenaikan gaji, promosi jabatan, kenaikan dan penurunan pangkat. Pada saat ini pengolahan penilaian dan evaluasi kinerja aparatur masih dilakukan dengan manual, proses ini tentunya tidak akan menjadi masalah apabila data yang diolah dalam jumlah yang kecil, sedangkan bila data yang diolah dalam jumlah yang cukup besar maka akan memerlukan tenaga dan waktu yang cukup lama untuk mengelompokkannya. Untuk mengatasi masalah pengelompokan data yang besar dapat memanfaatkan teknologi *data mining*. Penelitian ini dilaksanakan pada Badan Kepegawaian Daerah Provinsi Bali, dan sampai saat ini pengolahan data pada sistem kepegawaian belum menggunakan *data mining*. Pemerintah Provinsi Bali telah mengembangkan berbagai sistem informasi untuk mendukung pelaksanaan tugas-tugas sesuai dengan tugas kedinasan masing-masing Organisasi Perangkat Daerah salah satunya adalah pengembangan sistem informasi manajemen kepegawaian (SIMPEG) yang merupakan sistem informasi yang mengelola data pegawai [1]. Pada penelitian ini mencoba menerapkan algoritma C5.0 yang sudah pernah dilakukan sebelumnya tetapi dengan variabel yang berbeda dan diaplikasikan pada data penilaian kinerja pegawai pada Badan Kepegawaian Daerah Provinsi Bali. Beberapa penelitian tentang algoritma C5.0 telah dilakukan, dimana hasil yang didapat bahwa algoritma C5.0 lebih baik dalam melakukan klasifikasi, seperti penelitian tentang perbandingan performansi algoritma *Decision Tree* C5.0, CART, dan CHAID, dimana didapat bahwa algoritma C5.0 lebih akurat dalam melakukan *credit scoring* [2]. Penelitian lain tentang perbandingan Algoritma C5.0 dengan CART Clasification. Hasil yang didapat bahwa tingkat akurasi algoritma C5.0 lebih baik daripada algoritma CART [3]. Sasaran penelitian ini adalah melakukan analisa klasifikasi terhadap data penilaian kinerja pegawai negeri sipil dengan menggunakan algoritma C5.0.

II. LANDASAN TEORI

Berikut ini akan dijelaskan teori pendukung yang menjadi landasan penelitian ini diantaranya:

A. Data Mining

Data mining bisa diartikan proses pencarian pola data yang tidak diketahui atau tidak diperkirakan sebelumnya [4]. *Data mining* merupakan salah satu tahapan dalam keseluruhan proses *knowledge discovery in database*. Secara umum ada beberapa teknik *data mining*, salah satu teknik *data mining* adalah klasifikasi. Adapun salah satu teknik klasifikasi adalah *decision tree*.



B. Decision tree

Decision tree atau pohon keputusan merupakan salah satu teknik klasifikasi. *Decision tree* adalah *top-down* pohon rekursif dari algoritma induksi, yang menggunakan ukuran seleksi atribut untuk memilih atribut yang diuji [5]. Dengan pohon keputusan, manusia dapat dengan mudah melihat hubungan antara faktor-faktor yang mempengaruhi suatu masalah. Pohon keputusan ini juga dapat menganalisa nilai suatu informasi yang terdapat dalam suatu alternatif pemecahan masalah. Pohon keputusan adalah salah satu metode klasifikasi yang mudah untuk diaplikasikan oleh manusia. Pohon keputusan bisa melakukan prediksi dengan menggunakan struktur pohon. Pohon keputusan mengubah data menjadi pohon keputusan serta menjadi beberapa aturan keputusan.

C. Algoritma C5.0

Algoritma C5.0 merupakan penyempurnaan dari algoritma ID3 dan C4.5. Dalam proses pembentukan pohon keputusan nilai informasi *gain* tertinggi akan terpilih sebagai *root* bagi *node* selanjutnya. Algoritma ini dimulai dengan semua data yang dijadikan akar dari pohon keputusan sedangkan atribut yang dipilih akan menjadi pembagi bagi sampel tersebut. Formula ukuran atribut adalah [5]:

$$\text{Info}(D) = -\sum_{i=1}^m p_i \log_2(p_i) \quad (1)$$

Dengan $\text{Info}(D)$ merupakan informasi yang dibutuhkan untuk mengklasifikasikan label kelas sebuah *tuple* di D . p_i adalah peluang bukan nol dengan sebuah *tuple* acak di D . Fungsi log menggunakan basis 2, karena informasi yang dikodekan dalam bit. $\text{Info}(D)$ juga dikenal sebagai *entropy*. Nilai *entropy* yang dihasilkan untuk mengklasifikasi *tuple* dari D berdasarkan partisi oleh A adalah [5]:

$$\text{Info}_A(D) = \sum_{j=1}^Y \frac{|D_j|}{D} \times \text{Info}(D_j) \quad (2)$$

Untuk mendapatkan nilai *information gain* pada atribut A selanjutnya digunakan rumus [5]:

$$\text{GAIN}(A) = \text{Info}(D) - \text{Info}_A(D) \quad (3)$$

$\text{Gain}(A)$ menyatakan berapa banyak cabang yang akan diperoleh pada A . Atribut A dengan *information gain* tertinggi. Informasi, $\text{Gain}(A)$, dipilih sebagai atribut pada node N .

D. Pruning

Pruning adalah proses yang dilakukan untuk memotong atau menghilangkan beberapa cabang (*branches*) yang tidak diperlukan [6]. Untuk menghitung nilai *error pruning* digunakan rumus [7]:

$$e = \frac{r + \frac{z^2}{2n} + z\sqrt{\frac{r}{n} - \frac{r^2}{n} + \frac{z^2}{4n^2}}}{1 + \frac{z^2}{n}} \quad (4)$$

Dimana:

r = nilai perbandingan *error rate*

n = total sample

$z = \Phi^{-1}(c)$

c = confidence level

E. Confusion Matrix

Confusion matrix merupakan sebuah metode untuk evaluasi yang menggunakan tabel *matrix*. Hasil Evaluasi dengan menggunakan *confusion matrix* menghasilkan nilai akurasi, serta laju *error*. Akurasi menyatakan jumlah data yang diklasifikasikan benar setelah dilakukan proses pengujian, sedangkan laju *error* digunakan untuk menghitung kesalahan identifikasi. Untuk menghitung akurasi adalah [8]

$$\text{Akurasi} = \frac{TP + TN}{TP + FN + FP + FN} \times 100 \quad (5)$$

Dimana TP adalah *True Positive* yaitu jumlah data positif yang terklasifikasi dengan benar oleh sistem, TN adalah *True Negative* yaitu jumlah data negatif yang terklasifikasi dengan benar oleh sistem, FN *False Negative* yaitu jumlah data negatif namun terklasifikasi salah oleh system dan FP adalah *False Positive*, yaitu jumlah data positif namun terklasifikasi salah oleh system. Laju *error* dapat dihitung sebagai berikut [9]:

$$\text{Laju error} = \frac{\text{Jumlah data diidentifikasi salah}}{\text{Jumlah seluruh data}} \times 100\% \quad (6)$$

III. METODE PENELITIAN

Secara garis besar penelitian ini dibagi menjadi 4 tahapan yaitu identifikasi masalah, penentuan variabel penelitian, klasifikasi dengan algoritma C5.0 dan melakukan perhitungan akurasi. Identifikasi masalah dilakukan karena adanya tingkat subjektivitas (suka atau tidak suka) yang tinggi pada pengukuran kinerja PNS, sehingga diperlukan suatu metode yang membantu sebagai acuan dalam melakukan penilaian kinerja PNS. Setelah dilakukan identifikasi masalah dilakukan penentuan variabel pada objek penelitian meliputi jabatan, sasaran kerja pegawai, orientasi pelayanan, integritas, komitmen, disiplin, kerjasama, dan kepemimpinan, dan status penilaian. Kemudian dibangunlah model klasifikasi dengan algoritma C5.0. Model C5.0 dapat membagi sampel berdasarkan nilai *information gain* terbesar. Variabel yang memiliki *information gain* terbesar akan dipilih.

Tahapan penelitian ini bisa dijabarkan sebagai berikut :

1. Lakukan pra proses terhadap data yang telah diambil, meliputi seleksi data, *data cleaning* dan transformasi data.
2. Untuk melakukan klasifikasi data dibagi menjadi data latih dan data uji. Data latih untuk membangun pohon keputusan, sedangkan data uji untuk menghitung akurasi pohon keputusan.

3. Selanjutnya dibangun model klasifikasi untuk membentuk pohon keputusan.
4. Penghitungan akurasi dihitung dari model klasifikasi, dengan menggunakan *confusion matrix*.

IV. HASIL DAN PEMBAHASAN

A. Deskripsi Data

Data yang diperoleh merupakan data penilaian kinerja yang perlu diolah. Pengolahan data diperlukan untuk proses penelitian, untuk itu perlu dilakukan klasifikasi terhadap data yang ada. Data dipilah berdasarkan jabatan karena ini terkait dengan atribut kepemimpinan, apabila jabatan staf maka atribut kepemimpinan diabaikan, dan apabila jabatan struktural maka nilai kepemimpinan terisi. Berikut proses klasifikasi data penilaian kinerja terhadap 275 data PNS

TABEL I
KLASIFIKASI DATA PEGAWAI

Jabatan	Jumlah Data	Memenuhi Syarat	
		Ya	Tidak
Staf	235	197	38
Struktural	40	38	2

Data dibagi menjadi data staf dan data pejabat struktural, pembagian data ini bertujuan untuk menghitung status penilaian pada variabel kepemimpinan. Variabel kepemimpinan terkait dengan dengan jabatan struktural. Apabila data merupakan pejabat struktural, variabel kepemimpinan harus terisi sedangkan apabila data jabatan merupakan staf maka variabel kepemimpinan dikosongkan Status penilaian bernilai ya dan tidak menyatakan jumlah atribut staff yang memenuhi syarat baik untuk naik pangkat atau mendapatkan promosi. Dari 235 data staf, terdapat 197 orang yang memenuhi syarat dan 38 orang yang tidak memenuhi syarat. Untuk atribut pejabat struktural terdapat 38 pejabat yang memenuhi syarat dan 2 orang pejabat yang tidak memenuhi syarat. Selanjutnya data tersebut kemudian diklasifikasi lagi berdasarkan kriteria penilaian yang telah ditetapkan yaitu sasaran kerja pegawai, orientasi pelayanan, integritas, komitmen, disiplin, kerjasama dan kepemimpinan. Klasifikasi berdasarkan kriteria penilaian dapat dilihat pada tabel II berikut:

TABEL II
KLASIFIKASI BERDASARKAN KRITERIA PENILAIAN

Sasaran Kerja Pegawai	Jumlah Data	Memenuhi Syarat	
		Ya	Tidak
Sangat Baik	16	16	0
Baik	205	181	24
Cukup	14	0	14
Buruk	0	0	0
Orientasi Pelayanan			
Sangat Baik	0	0	0
Baik	184	162	22
Cukup	51	35	16
Kurang	0	0	0

Buruk	0	0	0
Integritas			
Sangat Baik	0	0	0
Baik	191	169	22
Cukup	44	28	16
kurang	0	0	0
Buruk	0	0	0
Komitmen			
Sangat Baik	0	0	0
Baik	191	173	18
Cukup	44	24	20
kurang	0	0	0
Buruk	0	0	0
Disiplin			
Sangat Baik	0	0	0
Baik	185	168	17
Cukup	50	29	21
kurang	0	0	0
Buruk	0	0	0
Kerjasama			
Sangat Baik	0	0	0
Baik	190	172	18
Cukup	45	25	20
Kurang	0	0	0
Buruk	0	0	0
Kepemimpinan			
Sangat Baik	0	0	0
Baik	40	38	2
Cukup	0	0	0
Kurang	0	0	0
Buruk	0	0	0

Dari tabel II dapat dijelaskan bahwa kriteria penilaian dibagi menjadi sasaran kerja pegawai, orientasi pelayanan, integritas, komitmen, disiplin, kerjasama, kepemimpinan. Nilai masing-masing kriteria penilaian meliputi sangat baik, baik, cukup, kurang, buruk. Dari hasil tabel diatas nilai ya dan tidak menyatakan bahwa apabila salah satu atau beberapa kriteria penilaian memiliki nilai cukup, kurang dan buruk berpengaruh terhadap kriteria penilaian akhir, yaitu memenuhi syarat atau tidak. Nilai ya adalah nilai akhir yang bernilai sangat baik dan baik, sedangkan nilai tidak merupakan nilai akhir bernilai cukup, kurang, buruk.

B. Pembahasan

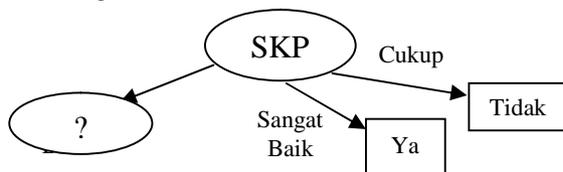
Proses pohon keputusan adalah mengubah bentuk data menjadi model pohon, pohon ini nanti akan menjadi aturan-aturan. Atribut yang dipilih sebagai *root* ditetapkan dengan menghitung nilai *gain* tertinggi. Sebelum menghitung nilai *gain* kita harus menghitung nilai *entropy* terlebih dahulu. Untuk perhitungan awal menggunakan data staf berjumlah 235 data, dilakukan dengan mengabaikan atribut kepemimpinan. Data dibagi menjadi data *training* dan data uji. Jumlah data *training* untuk awal pengujian sebanyak 184 data dan jumlah data uji sebanyak 51 data.



TABEL III
PERHITUNGAN NODE 1

No	Atribut	Nilai Atribut	Entropy	Information Gain
1	Total	Total	0.6415	
2	SKP	Baik	0.5777	0.1392
3	SKP	Cukup	0	0.1392
4	SKP	Sangat Baik	0	0.1392
5	orientasi_ pelayanan	Baik	0.5869	0.0126
6	orientasi_ pelayanan	Cukup	0.8631	0.0126
7	integritas	Baik	0.5144	0.0431
8	integritas	Cukup	0.9436	0.0431
9	komitmen	Baik	0.4435	0.1012
10	komitmen	Cukup	1	0.1012
11	disiplin	Baik	0.4138	0.1018
12	disiplin	Cukup	0.9928	0.1018
13	kerjasama	Baik	0.4312	0.0969
14	kerjasama	Cukup	0.9953	0.0969

Pada Perhitungan *node 1* didapat variabel SKP memiliki nilai *information gain* terbesar sehingga dijadikan *root*, selanjutnya dibuatlah pohon keputusan dengan SKP sebagai *root*. Pada perhitungan diatas atribut sangat baik dan cukup memiliki nilai *entropy* 0 sehingga tidak bercabang, tapi atribut baik masih harus dihitung kembali. Pohon keputusan yang terbentuk sebagai berikut



Gambar 1: Pohon Keputusan Node 1

Selanjutnya dilakukan proses perhitungan kembali. Proses perhitungan terus dilakukan hingga tidak terdapat cabang pada pohon keputusan. Hasil perhitungan dapat dilihat pada tabel IV. Setelah perhitungan selesai dilakukan, selanjutnya dilakukan evaluasi untuk mengetahui hasil perbandingan keputusan antara data uji dengan perhitungan algoritma C5.0.

TABEL IV
HASIL PERHITUNGAN ENTROPY DAN INFORMATION GAIN

No	Atribut	Nilai Atribut	Entropy	Information Gain
1	Total	Total	0.6415	
2	SKP	Baik	0.5777	0.1392

3	SKP	Cukup	0	0.1392
4	SKP	Sangat Baik	0	0.1392
5	orientasi_ pelayanan	Baik	0.5869	0.0126
6	orientasi_ pelayanan	Cukup	0.8631	0.0126
7	integritas	Baik	0.5144	0.0431
8	integritas	Cukup	0.9436	0.0431
9	komitmen	Baik	0.4435	0.1012
10	komitmen	Cukup	1	0.1012
11	disiplin	Baik	0.4138	0.1018
12	disiplin	Cukup	0.9928	0.1018
13	kerjasama	Baik	0.4312	0.0969
14	kerjasama	Cukup	0.9953	0.0969
15	disiplin	Total	0.5777	
16	disiplin	Baik	0.5561	0.0024
17	disiplin	Cukup	0.7025	0.0024
....
120	kerjasama	Cukup	0	0

C. Evaluasi

Setelah dilakukan proses perhitungan *entropy*, *information gain* dan pembuatan pohon keputusan maka dilakukan proses evaluasi dengan melakukan perbandingan keputusan antara hasil keputusan C5.0 dengan keputusan pada data uji. Data uji yang digunakan adalah sebanyak 51 data. Setelah dilakukan proses perbandingan keputusan didapat bahwa dari 43 keputusan data uji yang bernilai ya, algoritma C5.0 mampu mengidentifikasi 43 keputusan bernilai ya, sedangkan dari 8 keputusan data uji bernilai tidak, algoritma C5.0 hanya mampu mengidentifikasi 6 keputusan yang bernilai tidak dan 2 keputusan bernilai ya. Berdasarkan perbandingan tersebut algoritma ini memiliki akurasi sebesar 96,08%. Identifikasi keputusan C5.0 dapat dilihat pada tabel V.

TABEL V
IDENTIFIKASI KEPUTUSAN C5.0

	Diidentifikasi Ya oleh C5	Diidentifikasi Tidak oleh C5
Keputusan data uji: Ya = 43	43	0
Keputusan data uji: Tidak = 8	2	6

Dari tabel diatas didapat bahwa jumlah keputusan yang sesuai sebanyak 49 data dan total data uji sebanyak 51 data. Untuk menghitung akurasi sistem dalam mengklasifikasikan data testing digunakan teori *confusion matrix* yaitu dengan menentukan berapa nilai TP, TN, FP dan FN [10] kemudian dihitung dengan menggunakan persamaan (5) sebagai berikut:

$$Akurasi = \frac{43 + 6}{43 + 8} \times 100 = 96,07\%$$

Dari nilai akurasi yang dapat dijelaskan bahwa dengan melakukan pengujian sebanyak 51 data, algoritma C5.0 mampu mengidentifikasi sebanyak 49 data yang sesuai dengan data uji. Dari hasil data uji, 43 data bernilai ya dan 8 data bernilai tidak, didapat algoritma C5.0 mampu mengidentifikasi ya sebanyak 43 data dan tidak sebanyak 6 data, kesalahan identifikasi sebanyak 2 data. Untuk menghitung kesalahan identifikasi digunakan laju *error* dengan persamaan (6).

$$\text{Nilai laju } error \text{ didapat } \frac{0 + 2}{43 + 8} \times 100 = 3,93\%$$

Nilai laju *error* sebesar 3,93% menggambarkan tingkat kesalahan identifikasi yang dilakukan oleh algoritma C5.0, dari tabel V dapat dilihat terdapat 2 data yang diidentifikasi salah yaitu bernilai ya dari total 8 data uji yang seharusnya bernilai tidak.

Untuk mendapat akurasi yang baik, dilakukan pengujian dengan mengambil beberapa *sample* data secara bertahap dari total *sample* yang ada, lalu ditingkatkan jumlah *sample* yang diuji sampai total *sample* tercapai. Atribut kepemimpinan diperlakukan perhitungan khusus, karena ini terkait dengan data pejabat struktural. Dari total 235 *sample* jabatan staf, dan 40 *sample* jabatan struktural dilakukan model pengujian sebagai berikut:

1. 184 *sample* sebagai data *training* dan 51 *sample* sebagai data uji.
2. 118 *sample* sebagai data *training* dan 117 *sample* sebagai data uji.
3. 70 *sample* sebagai data *training* dan 165 *sample* sebagai data uji.
4. 32 *sample* sebagai data *training* dan 8 *sample* sebagai data uji.

Hasil akurasi pengujian beberapa *sample* data dapat dilihat pada tabel VI berikut

TABEL VI
PERBANDINGAN AKURASI DENGAN VARIASI DATA

Jumlah data <i>training</i>	Jumlah data uji	Total Data	Akurasi
184	51	235	96,08%
118	117	235	89,74%
70	165	235	66,06%
32	8	40	100%

IV. KESIMPULAN

Berdasarkan hasil evaluasi yang telah dilakukan dapat diambil beberapa kesimpulan sebagai berikut :

1. Algoritma C5.0 dapat memproses data kinerja pegawai menjadi sebuah pohon keputusan serta aturan-aturan yang berguna sebagai sebuah masukan. Hasil yang diperoleh bisa dikembangkan menjadi sebuah sistem penentu keputusan sehingga dapat dijadikan untuk membantu dalam menentukan keputusan kinerja pegawai.
2. Secara umum berdasarkan hasil evaluasi untuk data staf dengan data *training* sebanyak 184 data didapat tingkat

akurasi sebesar 96,08%, dimana salah satu faktor yang mempengaruhi akurasi adalah data kinerja tidak memenuhi syarat yang masih kecil, hal ini bisa ditingkatkan dengan menambah jumlah data *training* yang terkait dengan data tersebut.

REFERENSI

- [1] Laksmidewi, N.L.A., Linawati, Widyantara I.M.O., "Evaluasi Sistem Informasi Manajemen Kepegawaian dengan DS5 dan DS9 COBIT 4.1 Studi Kasus : Pemprov Bali". Denpasar: Teknologi Elektro, Vol.17.no. 1. Januari-April 2018.
- [2] W. Yusuf Y., "Perbandingan Performansi Algoritma Decision Tree C5.0, CART, Dan CHAID: Kasus Prediksi Status Resiko Kredit Di Bank X". Seminar, vol. 2007, no. Snati, pp. 0-3, 2007.
- [3] P. N. Patil, R. Lathi, and V. Chitre, "Comparison of C5.0 & CART Classification algorithms using pruning technique," Int. J. Eng. Res. Technol., vol. 1, no. 4, pp. 1-5, 2012.
- [4] Moss Larissa T. and Said Adelman, "Data Warehouse Project Mangement". Canada: Addison-Wesley, 2000.
- [5] Han, Jiawei dan Kamber, Micheline. "Data Mining: Concepts and Techniques. 3rd Edition", Morgan Kaufmann, Burlington. 2011
- [6] G. I. Sammut, Claude, Webb, "Encyclopedia of Machine Learning and Data Mining". Springer, 2011.
- [7] Larose, Daniel T., "Discovering Knowledge in Data: An Introduction to Data Mining". New Jersey : John Willey & Sons, Inc. 2005
- [8] M. Sokolova dan G. Lapalme, "A systematic analysis of performance measures for classification tasks," Inf. Process. Manag., vol. 45, no. 4, hal. 427-437, 2009.
- [9] Prasetyo, E. "Data Mining Mengolah Data Menjadi Informasi Menggunakan MATLAB". Yogyakarta: ANDI. 2012
- [10] N. L. Ratniasih, M. Sudarma, N. Gunantara. "Penerapan Text Mining dalam Spam Filtering untuk Aplikasi Chat," Denpasar: Teknologi Elektro, vol. 16, no. 3, September-Desember 2017.



{halaman ini sengaja di Kosongkan}