

## Penerapan Algoritma C4.5 Dalam Meningkatkan Kemampuan Bahasa Inggris Pada Mahasiswa

Septri Wanti Siahaan<sup>1</sup>; Kristin D R Sianipar<sup>2</sup>; P.P.P.A.N.W Fikrul Ilmi R.H Zer<sup>3</sup>;  
Dedy Hartama<sup>4</sup>

<sup>1, 2, 3, 4</sup>STIKOM Tunas Bangsa Pematangsiantar  
<sup>1</sup>septriwanti26@gmail.com

### ABSTRACT

*In this world, many languages from other countries can be used as a communication tool. One of them is English. Students who have education must know that learning English is very much needed because nobody knows what will happen in the years to come. It could be one factor to get a job the next few years is our expertise in speaking English. English is a global language used by people to communicate with other people. On this occasion, researchers conducted research to determine what factors can improve students' ability to speak English. To complete this research, researchers resolve by applying the existing algorithm in data mining, namely C4.5 Algorithm. The result of this research can be concluded that the factors that influence to improve students' ability in English are Listening from the environment. This research also has a purpose to motivate students to improve their ability to speak English.*

**Keywords:** *c4.5 algorithm, English, data mining, ability, improving*

### ABSTRAK

*Di dunia ini banyak bahasa dari negara-negara lain yang dapat dijadikan sebagai alat berkomunikasi. Salah satunya adalah bahasa inggris. Mahasiswa yang memiliki pendidikan harus tahu bahwa mempelajari bahasa inggris sangat dibutuhkan, sebab tidak ada yang tahu apa yang akan terjadi pada tahun-tahun yang akan datang. Bisa saja salah satu faktor untuk memperoleh pekerjaan beberapa tahun berikutnya adalah dengan kepiawaian kita dalam berbahasa inggris. Bahasa inggris adalah bahasa global yang digunakan oleh manusia untuk berkomunikasi satu sama lain. Pada kesempatan ini, peneliti melakukan penelitian untuk menentukan apa saja faktor yang dapat meningkatkan kemampuan mahasiswa dalam berbahasa inggris. Untuk menyelesaikan penelitian ini, peneliti menyelesaikan dengan menerapkan algoritma yang ada pada data mining, yaitu Algoritma C4.5. Hasil dari penelitian ini dapat disimpulkan bahwa faktor yang mempengaruhi untuk meningkatkan kemampuan mahasiswa dalam bahasa inggris adalah Mendengar dari Lingkungan. Penelitian ini pun memiliki tujuan yaitu untuk memotivasi mahasiswa dalam meningkatkan kemampuan mahasiswa dalam berbahasa inggris.*

**Kata kunci:** *algoritma c4.5, bahasa inggris, data mining, kemampuan, meningkatkan*

## **1. PENDAHULUAN**

Negara-negara di dunia memiliki bahasa yang mereka gunakan dalam berkomunikasi dengan orang lain. Di Indonesia menggunakan bahasa persatuan yaitu bahasa Indonesia. Namun, jika ingin berkomunikasi dengan orang asing dari negara lain, pasti akan kesulitan untuk mengerti apa yang mereka katakan. Hal itu dikarenakan manusia memakai bahasa yang mereka punya dan begitupun dengan orang asing, sehingga tidak ada titik temu dalam percakapan. Maka dari itu, bahasa Inggris digunakan sebagai bahasa global dalam berkomunikasi. Sehingga, nantinya dapat berkomunikasi dan mengerti apa yang ucapkan. Salah satu teknologi yang dapat digunakan untuk meningkatkan sistem pembelajaran adalah *computer* [1].

Kita mempelajari bahasa Inggris sudah sejak dari Sekolah Dasar (SD) ataupun ada juga yang mulai mempelajarinya dari taman kanak-kanak. Pada Perguruan Tinggi pun diberikan mata kuliah bahasa Inggris pada mahasiswa meskipun *output* dari studi tidak berhubungan dengan bahasa Inggris. Dari hal tersebut disimpulkan bahwa betapa pentingnya kemampuan dalam berbahasa asing, khususnya bahasa Inggris sebagai salah satu faktor kesuksesan pada akademik seseorang ataupun untuk mendapatkan karir bagus dalam pekerjaan [2]. Namun, walaupun telah mempelajari bahasa Inggris sudah sejak dari lama, masih banyak mahasiswa yang merasa kesulitan dalam mempelajari bahasa Inggris.

Mahasiswa kesulitan dalam mempelajari bahasa Inggris. Kesulitan yang dapat dialami seperti, kesulitan dalam membaca, menulis, mendengarkan dan lain-lain. Terkadang ada dari kita yang merasa tidak percaya diri dalam mengaplikasikan bahasa tersebut dalam kehidupan sehari-hari kita. Alasannya adalah karena takut diejek dan dianggap *sok jago* oleh orang lain.

Adapun juga masalah yang sering dialami mahasiswa yaitu kesulitan dalam membaca kosakata dan menerapkan *grammar* dalam bahasa Inggris. Peran pengajar sangat diperlukan untuk dapat meningkatkan kemampuan bahasa Inggris pada mahasiswa.

Dengan adanya permasalahan tersebut, peneliti berinisiatif untuk menentukan faktor apa yang dapat meningkatkan kemampuan mahasiswa dalam bahasa Inggris. Ada banyak cabang dari ilmu komputer yang dapat digunakan untuk menyelesaikan permasalahan yang sifatnya kompleks. Cabang ilmu komputer tersebut adalah *Artificial Intelligence* seperti *datamining* [3]–[9]. Berdasarkan permasalahan diatas, peneliti memanfaatkan *datamining* yaitu klasifikasi dalam menyelesaikan permasalahan untuk meningkatkan kemampuan bahasa Inggris mahasiswa. Algoritma pada *datamining* yang dipakai adalah Algoritma C4.5[10].

Tujuan dilakukannya penelitian ini akan membuat mahasiswa semakin termotivasi untuk meningkatkan kemampuan dalam berbahasa Inggris. Hasil penelitian ini menggunakan hitungan manual dan melakukan validasi menggunakan *software rapidminer* dan didapatkan bahwa faktor yang mempengaruhi dalam meningkatkan kemampuan bahasa Inggris mahasiswa adalah Mendengar dari Lingkungan.

## **2. METODE/PERANCANGAN PENELITIAN**

### **2.1. Membuat Dataset**

*Dataset* diperoleh dengan memberikan kuisioner yang dibuat dari *Google Form*. Kuisioner diberikan kepada mahasiswa untuk melengkapi penyelesaian penelitian ini.

**Tabel 1.** Data Responden

No	K1	K2	K3	K4	Kemampuan Bahasa Inggris
1.	3	4	5	1	Mampu
2.	4	3	1	2	Mampu
3.	1	5	5	3	Mampu
4.	3	5	2	2	Tidak Mampu
5.	5	4	4	3	Mampu
6.	4	1	2	2	Tidak Mampu
7.	5	4	5	5	Mampu
8.	4	5	5	3	Mampu
9.	3	4	3	1	Tidak Mampu
10.	1	3	2	3	Mampu
11.	4	2	3	3	Mampu
12.	5	4	1	5	Tidak Mampu
13.	2	2	3	3	Tidak Mampu
14.	4	4	5	5	Mampu
15.	5	5	3	5	Mampu
16.	1	4	3	2	Tidak Mampu
17.	4	1	4	4	Tidak Mampu
18.	5	2	2	3	Tidak Mampu
19.	2	5	5	1	Tidak Mampu
20.	4	4	5	4	Mampu
21.	4	3	1	2	Mampu
22.	1	5	5	3	Mampu
23.	3	5	2	2	Tidak Mampu
24.	4	4	5	5	Mampu
25.	5	5	3	5	Mampu
26.	1	4	3	2	Tidak Mampu
27.	4	1	4	4	Tidak Mampu
28.	5	2	2	3	Tidak Mampu
29.	2	5	5	1	Tidak Mampu
30.	1	5	5	3	Mampu
31.	3	5	2	2	Tidak Mampu
32.	5	4	4	3	Mampu
33.	3	4	3	1	Tidak Mampu
34.	1	3	2	3	Mampu
35.	4	2	3	3	Mampu
36.	5	4	1	5	Tidak Mampu
37.	1	4	3	2	Tidak Mampu
38.	4	1	4	4	Tidak Mampu
39.	4	1	2	2	Tidak Mampu
40.	5	4	5	5	Mampu
41.	4	5	5	3	Mampu
42.	3	4	3	1	Tidak Mampu
43.	1	3	2	3	Mampu
44.	3	4	5	1	Mampu
45.	4	3	1	2	Mampu
46.	2	5	5	1	Tidak Mampu
47.	5	4	1	5	Tidak Mampu
48.	2	2	3	3	Tidak Mampu
49.	4	4	5	5	Mampu
50.	5	5	3	5	Mampu
51.	2	5	5	1	Tidak Mampu
52.	4	4	5	4	Mampu
53.	4	3	1	2	Mampu

54.	5	4	1	5	Tidak Mampu
55.	2	2	3	3	Tidak Mampu
56.	5	4	4	3	Mampu
57.	4	1	2	2	Tidak Mampu
58.	5	4	5	5	Mampu
59.	4	1	2	2	Tidak Mampu
60.	5	4	5	5	Mampu
61.	4	3	1	2	Mampu
62.	1	5	5	3	Mampu
63.	3	5	2	2	Tidak Mampu
64.	5	4	4	3	Mampu
65.	1	4	3	2	Tidak Mampu
66.	4	1	4	4	Tidak Mampu
67.	4	1	2	2	Tidak Mampu
68.	2	5	5	1	Tidak Mampu
69.	4	4	5	4	Mampu
70.	4	3	1	2	Mampu

Penelitian ini mengumpulkan data dari 70 responden dengan membuat 4 kriteria, yaitu: Membaca Referensi (K1), Mendengar dari Lingkungan (K2), Mempraktekkan dalam Lingkungan (K3), dan Memanfaatkan Teknologi (K4). Dengan sub kriteria sebagai berikut:

- 5 = Sangat Setuju
- 4 = Setuju
- 3 = Cukup Setuju
- 2 = Kurang Setuju
- 1 = Sangat Kurang Setuju

## 2.2. Menerapkan Algoritma yang Digunakan

Pada penelitian ini menggunakan Algoritma C4.5 dalam menentukan faktor yang mempengaruhi mahasiswa dalam meningkatkan kemampuan bahasa Inggris mahasiswa. Algoritma C4.5 adalah algoritma yang terdapat pada teknik pengklasifikasian untuk menyelesaikan kasus atau pun masalah. Pohon keputusan (*decision tree*) merupakan dasar dari Algoritma C4.5. Algoritma C.45 adalah salah satu algoritma induksi pohon keputusan yaitu ID3 (*Iterative Dichotomiser 3*) [11].

Adapun rumus yang dipakai dalam Algoritma C4.5 yaitu:

- Menghitung *entropy*

$$Entropi (S) = \sum_{j=1}^k - p_j \log_2 p_j \quad (1)$$

Penjelasan:

- S = himpunan (*dataset*) kasus
- k = banyaknya partisi *dataset*

- Menghitung *gain*

$$Gain (A) = Entropi (S) - \sum_{i=1}^k \frac{|S_i|}{|S|} \times Entropi(S_i) \quad (2)$$

Penjelasan:

- S = himpunan (*dataset*) kasus
- A = atribut
- N = banyaknya partisi atribut
- $|S_i|$  = banyaknya jumlah kasus partisi ke-i
- $|S|$  = banyaknya jumlah kasus S

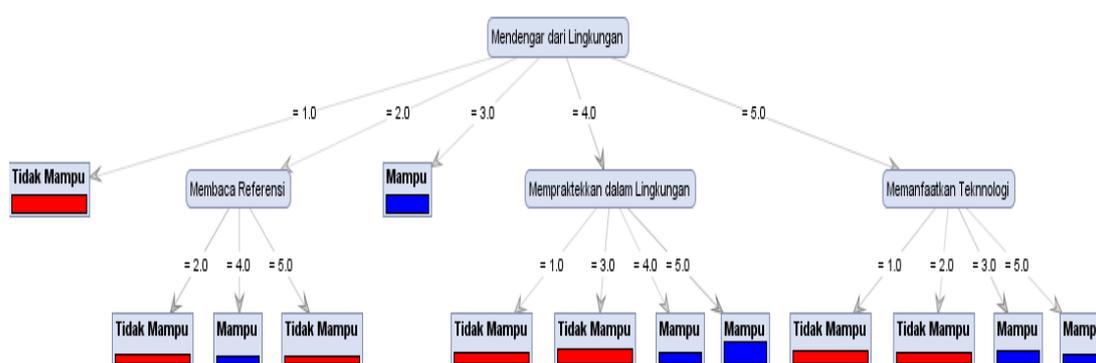
Kelebihan dari menggunakan Algoritma C4.5 adalah dapat membuat pohon keputusan (*decision tree*) sehingga menjadi efisien, dimana *decision tree* menangani atribut tipe diskrit dan tipe diskrit- numerik, mudah untuk diinterpretasikan dan memiliki tingkat akurasi yang dapat diterima [12]. Kelemahan Algoritma C4.5 salah satunya adalah terdapat pada skalabilitas yaitu *data training* hanya dapat digunakan dan disimpan secara keseluruhan pada waktu yang bersamaan di memori [13].

### 2.3. Melakukan Hitungan Manual

Pada penelitian ini dilakukan perhitungan manual dengan menggunakan rumus yang telah ada.

### 2.4. Validasi

Proses *validasi* dilakukan dengan menggunakan *software Rapidminer*. *Validasi* dilakukan dengan mengambil sampel dari *dataset* yang dimiliki sebanyak 70 responden dari data kuisisioner yang diberikan kepada mahasiswa dan kemudian data diuji dengan menggunakan *decision tree* (C4.5).



Gambar 1. Pohon Keputusan dari *dataset* 70 responden

*Decision Tree* memakai kinerja dalam bentuk pohon (*tree*) yang dimana pada masing-masing *node* yang dimiliki menjelaskan atribut yang ada, cabangnya menafsirkan nilai pada atribut yang ada dan *leaf* mempresentasikan kelas [14]. Pada pohon keputusan ada yang dikenal sebagai *root*. *Root* adalah *node* yang berada paling atas dari *decision tree*. *Decision Tree* adalah teknik pengklasifikasian data yang paling dikenal untuk digunakan pada *data mining*. Pohon keputusan menggunakan pembangunan yang relatif cepat. *Output* dari jenis model yang dibuat untuk dimengerti secara mudah.

### 2.5. Data Mining

Data mining adalah suatu istilah yang digunakan untuk menguraikan penemuan pengetahuan di dalam *database* [15]. Data mining juga dikenal dengan istilah *pattern recognition* merupakan suatu algoritma yang digunakan untuk pengolahan data guna menemukan pola yang tersembunyi

dari data yang diolah. Data yang diolah kemudian menghasilkan suatu pengetahuan baru yang bersumber dari data.

## 2.6. Decision Tree (Pohon Keputusan)

*Decision Tree* memakai kinerja dalam bentuk pohon (*tree*). Dalam pohon keputusan, masing-masing *node* yang dimiliki menjelaskan atribut yang ada, cabangnya menafsirkan nilai pada atribut yang ada dan *leaf* mempresentasikan kelas [14]. Pada pohon keputusan ada yang dikenal sebagai *root*. *Root* adalah *node* yang berada paling atas dari pohon keputusan. *Decision Tree* adalah metode pengklasifikasian data yang paling dikenal untuk digunakan pada *data mining*. Pohon keputusan menggunakan pembangunan yang relatif cepat. *Output* dari jenis model yang dibuat mudah untuk dimengerti. Pohon keputusan terbagi 3 jenis *node*, yaitu sebagai berikut [16]:

(a) *Root Node*, merupakan *node* yang berada paling atas. Pada *node* ini tidak memiliki *input* dan dapat juga tidak memiliki *output* atau memiliki *output* yang lebih dari satu.

(b) *Internal Node*, merupakan *node* yang bercabang. Pada *node* ini hanya memiliki 1 buah *input* dan memiliki *output* yaitu minimal 2.

(c) *Leaf node* atau *terminal node*, merupakan *node* yang diakhir. Pada *node* ini hanya memiliki 1 *input* dan tidak memiliki *output*.

## 2.7. Algoritma C4.5

Secara garis besar, langkah-langkah yang dilakukan oleh Algoritma C.45 dalam membentuk pohon keputusan adalah sebagai berikut [17]:

Input: sampel training, label training, atribut

1. Membuat simpul akar untuk pohon yang dibuat
2. Jika semua sampel positif, berhenti dengan suatu pohon dengan satu simpul akar, beri tanda (+)
3. Jika semua sampel negatif, berhenti dengan suatu pohon dengan satu simpul akar, beri tanda (-)
4. Jika atribut kosong, berhenti dengan suatu pohon dengan satu simpul akar, dengan label sesuai nilai yang terbanyak yang ada pada label training
5. Untuk yang lain, Mulai
  - a. A --- atribut yang mengklasifikasikan sampel dengan hasil terbaik (berdasarkan *gain rasio*)
  - b. Atribut keputusan untuk simpul akar ---- A
  - c. Untuk setiap nilai,  $v_i$ , yang mungkin untuk A
    - 1) Tambahkan cabang di bawah akar yang berhubungan dengan  $A = v_i$
    - 2) Tentukan sampel  $S_{v_i}$  sebagai subset dari sampel yang mempunyai nilai  $v_i$  untuk atribut A
    - 3) Jika sampel  $S_{v_i}$  kosong
      - i. Di bawah cabang tambahkan simpul daun dengan label = nilai yang terbanyak yang ada pada label training
      - ii. Yang lain tambah cabang baru di bawah cabang yang sekarang C 4.5 (sampel training, label training, atribut – [A]).
  - d. Berhenti Mengubah tree yang dihasilkan dalam beberapa rule. Jumlah rule sama dengan jumlah path yang mungkin dapat dibangun dari root sampai leaf node.

### 3. HASIL DAN PEMBAHASAN

Penelitian ini dilakukan dengan mengumpulkan data dari 70 responden dengan membuat 4 kriteria, yaitu: Membaca Referensi (K1), Mendengar dari Lingkungan (K2), Mempraktekkan dalam Lingkungan (K3), dan Memanfaatkan Teknologi (K4).

Dilakukan pengklasifikasian pada data responden. Perhitungan dilakukan dengan menggunakan *software Microsoft Excel*. Maka, selanjutnya melakukan perhitungan data menggunakan rumus yang ada pada Algoritma C4.5. Hasil dari perhitungan pada iterasi ke-1 dapat dilihat sebagai berikut:

**Tabel 2.** Hasil Perhitungan Iterasi ke-1

		Jlh Kasus	Mampu	Tidak Mampu	Entropy	Gain
		(S)	(S1)	(S2)		
Node 1		70	36	34	0,999411	
Membaca Referensi	Sangat Setuju	17	11	6	0,936667	
	Setuju	25	16	9	0,942683	
	Cukup Setuju	9	2	7	0,764205	
	Kurang Setuju	8	0	8	0	
	Sangat Kurang Setuju	11	7	4	0,94566	
						0,188403
Mendengar dari Lingkungan	Sangat Setuju	18	9	9	1	
	Setuju	27	16	11	0,975119	
	Cukup Setuju	9	9	0	0	
	Kurang Setuju	7	2	5	0,863121	
	Sangat Kurang Setuju	9	0	9	0	
						0,279838
Mempraktekkan dari Lingkungan	Sangat Setuju	23	18	5	0,755375	
	Setuju	8	4	4	1	
	Cukup Setuju	15	5	10	0,918296	
	Kurang Setuju	14	3	11	0,749595	
	Sangat Kurang Setuju	10	6	4	0,970951	
						0,151526
Memanfaatkan Teknologi	Sangat Setuju	14	10	4	0,863121	
	Setuju	7	3	4	0,985228	
	Cukup Setuju	20	15	5	0,811278	
	Kurang Setuju	19	6	13	0,899744	
	Sangat Kurang Setuju	10	2	8	0,721928	
						0,149121

Pada tabel 2 merupakan hasil perhitungan pada iterasi ke-1. Dapat dilihat bahwa kriteria Mendengar dari Lingkungan mendapatkan *gain* tertinggi dengan nilai adalah 0,27984 dan nilai *entropy* yang tertinggi adalah “Sangat Setuju” dengan nilai adalah 1.

Dari tabel 2, dikarenakan terdapat *root gain* tertinggi, maka dilanjutkan ke iterasi ke-2. Hasil perhitungan pada iterasi ke-2 dapat dilihat dibawah ini:

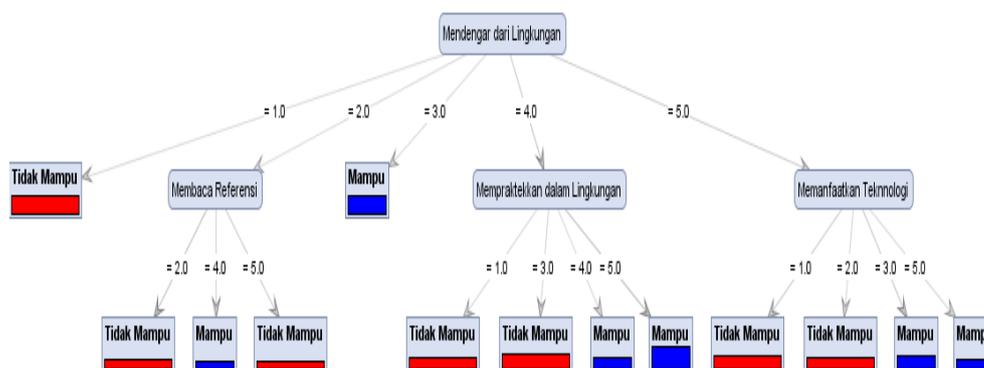
**Tabel 3.** Hasil Perhitungan Iterasi ke-2

		Jlh Kasus	Mampu	Tidak Mampu	Entropy	Gain
		(S)	(S1)	(S2)		
Node 2		18	9	9	1	
Membaca Referensi	Sangat Setuju	3	3	0	0	
	Setuju	2	2	0	0	
	Cukup Setuju	4	0	4	0	
	Kurang Setuju	5	0	5	0	
	Sangat Kurang Setuju	4	4	0	0	
						1
Mempraktekkan dari Lingkungan	Sangat Setuju	11	6	5	0,9940302	
	Setuju	0	0	0	0	
	Cukup Setuju	3	3	0	0	
	Kurang Setuju	4	0	4	0	
	Sangat Kurang Setuju	0	0	0	0	
						0,3925371
Memanfaatkan Teknologi	Sangat Setuju	3	3	0	0	
	Setuju	0	0	0	0	
	Cukup Setuju	6	6	0	0	
	Kurang Setuju	4	0	4	0	
	Sangat Kurang Setuju	5	0	5	0	
						1

Pada tabel 3 merupakan perhitungan pada iterasi ke-2 dan dapat dilihat bahwa telah didapatkan *gain* yang sama yaitu pada kriteria Membaca Referensi dan Memanfaatkan Teknologi dengan nilai *gain* adalah 1. Dan pada kriteria Mempraktekkan dalam Lingkungan mendapatkan *gain* terendah dengan nilai adalah 0,3925371.

Jika *entropy* telah bernilai 0 pada salah satu Semesta, hal itu menandakan bahwa sudah memiliki *leaf* atau daun.

Pada penelitian ini, kami melakukan validasi pada *software Rapidminer*. Kami menggunakan *decision tree* dan menghasilkan pohon keputusan sebagai berikut:



**Gambar 2.** Pohon Keputusan pada *rapidminer*

Berdasarkan pohon keputusan Gambar 2, dihasilkan pohon keputusan peningkatan kemampuan bahasa Inggris mahasiswa dalam bentuk teks yaitu, sebagai berikut:

Mendengar dari Lingkungan = 1.0: Tidak Mampu {Mampu=0, Tidak Mampu=9}

Mendengar dari Lingkungan = 2.0

| Membaca Referensi = 2.0: Tidak Mampu {Mampu=0, Tidak Mampu=3}

| Membaca Referensi = 4.0: Mampu {Mampu=2, Tidak Mampu=0}

| Membaca Referensi = 5.0: Tidak Mampu {Mampu=0, Tidak Mampu=2}

Mendengar dari Lingkungan = 3.0: Mampu {Mampu=9, Tidak Mampu=0}

Mendengar dari Lingkungan = 4.0

| Mempraktekkan dalam Lingkungan = 1.0: Tidak Mampu {Mampu=0, Tidak Mampu=4}

| Mempraktekkan dalam Lingkungan = 3.0: Tidak Mampu {Mampu=0, Tidak Mampu=7}

| Mempraktekkan dalam Lingkungan = 4.0: Mampu {Mampu=4, Tidak Mampu=0}

| Mempraktekkan dalam Lingkungan = 5.0: Mampu {Mampu=12, Tidak Mampu=0}

Mendengar dari Lingkungan = 5.0

| Memanfaatkan Teknologi = 1.0: Tidak Mampu {Mampu=0, Tidak Mampu=5}

| Memanfaatkan Teknologi = 2.0: Tidak Mampu {Mampu=0, Tidak Mampu=4}

| Memanfaatkan Teknologi = 3.0: Mampu {Mampu=6, Tidak Mampu=0}

| Memanfaatkan Teknologi = 5.0: Mampu {Mampu=3, Tidak Mampu=0}

#### 4. KESIMPULAN DAN SARAN

Dari hasil penelitian ini, dengan menggunakan Algoritma C4.5 dapat mempermudah dalam menentukan faktor apa yang dapat meningkatkan kemampuan bahasa Inggris pada mahasiswa. Dapat disimpulkan bahwa faktor yang mempengaruhi untuk meningkatkan kemampuan mahasiswa terhadap bahasa Inggris adalah kriteria “Mendengar dari Lingkungan” dengan proses perhitungan iterasi berhenti pada iterasi ke-2. Dengan adanya penelitian ini, dapat memberikan motivasi kepada orang lain terkhususnya mahasiswa, yaitu untuk meningkatkan kemampuan bahasa Inggris agar dapat digunakan pada masa yang akan datang dan menjadi salah satu sumber untuk mendapatkan pekerjaan.

Hasil dari perhitungan penelitian ini menggunakan pengaplikasian *data mining* yaitu menggunakan algoritma C4.5 yang dimana dalam melakukan perhitungan untuk mengetahui tingkat kemampuan mahasiswa dalam berbahasa Inggris harus memperhatikan nilai *Gain* yang paling tinggi dari atribut yang dipakai dalam penelitian ini.

Kelebihan dari dilakukannya penelitian ini adalah dapat dilakukannya proses validasi dengan menggunakan *software Rapidminer* sehingga menghasilkan pohon keputusan. Kekurangan dalam melakukan penelitian ini adalah pada saat proses perhitungan harus lebih hati-hati dalam menghitungnya sehingga hasil tidak berbeda dari yang dipakai di proses validasi menggunakan *software Rapidminer*. Pengembangan dari penelitian ini dapat diimplementasikan pada bidang lainnya, sehingga penelitian ini dapat menjadi acuan bagi peneliti lain dalam melakukan penelitian.

Saran dari kami peneliti dalam membuat penelitian ini adalah setelah melakukan perhitungan manual atau pun menggunakan *software* bantuan, ada baiknya melakukan validasi juga dengan bantuan *software Rapidminer*. Dengan begitu, lebih tepat jika dilakukan validasi.

#### UCAPAN TERIMA KASIH

Terima kasih kami ucapkan kepada semua penulis yang berpartisipasi dalam pembuatan penelitian ini. Kami ucapkan juga terima kasih kepada dosen yang telah membimbing kami dalam

penelitian ini. Semoga dengan adanya penelitian ini dapat memberikan wawasan baru kepada kami sebagai penulis dan juga sebagai pembaca.

#### **DAFTAR PUSTAKA**

- [1] D. Pembelajaran, K. Bahasa, and I. Di, “No Title,” no. 4, pp. 1 –23.
- [2] B. Inggris and S. Efektif, “Kesulitan Mahasiswa dalam Mencapai Pembelajaran Bahasa Inggris Secara Efektif,” no. 2, pp. 147–156, 2016.
- [3] W. Katrina, H. J. Damanik, F. Parhusip, D. Hartama, A. P. Windarto, and A. Wanto, “C.45 Classification Rules Model for Determining Students Level of Understanding of the Subject,” *J. Phys. Conf. Ser.*, vol. 1255, no. 12005, pp. 1 –7, 2019, doi: 10.1088/1742-6596/1255/1/012005.
- [4] M. Widyastuti, A. G. Fepdiani Simanjuntak, D. Hartama, A. P. Windarto, and A. Wanto, “Classification Model C.45 on Determining the Quality of Customer Service in Bank BTN Pematangsiantar Branch,” *J. Phys. Conf. Ser.*, vol. 1255, no. 12002, pp. 1 –6, 2019, doi: 10.1088/1742-6596/1255/1/012002.
- [5] Sudirman, A. P. Windarto, and A. Wanto, “Data mining tools | rapidminer: K-means method on clustering of rice crops by province as efforts to stabilize food crops in Indonesia,” *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 420, no. 1, 2018, doi: 10.1088/1757-899X/420/1/012089.
- [6] R. W. Sari, A. Wanto, and A. P. Windarto, “Implementasi Rapidminer Dengan Metode K-Means (Study Kasus: Imunisasi Campak Pada Balita Berdasarkan Provinsi),” *KOMIK (Konferensi Nas. Teknol. Inf. dan Komputer)*, vol. 2, no. 1, pp. 224–230, 2018, doi: 10.30865/komik.v2i1.930.
- [7] N. Rofiqo, A. P. Windarto, and D. Hartama, “Penerapan Clustering Pada Penduduk Yang Mempunyai Keluhan Kesehatan Dengan Datamining K-Means,” *KOMIK (Konferensi Nas. Teknol. Inf. dan Komputer)*, vol. 2, no. 1, pp. 216–223, 2018, doi:10.30865/komik.v2i1.929.
- [8] M. G. Sadewo, A. P. Windarto, and A. Wanto, “Penerapan Algoritma Clustering Dalam Mengelompokkan Banyaknya Desa/Kelurahan Menurut Upaya Antisipasi/ Mitigasi Bencana Alam Menurut Provinsi Dengan K-Means,” *KOMIK (Konferensi Nas. Teknol. Inf. dan Komputer)*, vol. 2, no. 1, pp. 311 –319, 2018, doi: 10.30865/komik.v2i1.943.
- [9] D. Hartama, A. Perdana Windarto, and A. Wanto, “The Application of Data Mining in Determining Patterns of Interest of High School Graduates,” *J. Phys. Conf. Ser.*, vol. 1339, no. 1, 2019, doi: 10.1088/1742-6596/1339/1/012042.
- [10] S. Kasus, U. Dehasen, S. Haryati, A. Sudarsono, and E. Suryana, “Implementasi Data Mining Untuk Memprediksi Masa Studi Mahasiswa Menggunakan Algoritma C4 . 5,” vol. 11, no. 2, pp. 130–138, 2015.
- [11] A. S. Febriarini and E. Z. Astuti, “Penerapan Algoritma C4 . 5 untuk Prediksi Kepuasan Penumpang Bus Rapid Transit (BRT) Trans Semarang,” pp. 95–103.
- [12] Han, J., Kamber, M., 2001, “Data Mining Concepts and Techniques”, Morgan Kaufman Pub., USA.
- [13] Veronica Sri Moertini, 2007, “Pengembangan Skalabilitas Algoritma Klasifikasi C4.5 Dengan Pendekatan Konsep Operator Relasi, studi kasus: pra-pengolahan dan klasifikasi citra batik”, Bandung.
- [14] O. Somantri, P. H. Bersama, S. Wiyono, and P. H. Bersama, “Model Data Mining Untuk Klasifikasi Tingkat Penguasaan Marteri Model Data Mining Untuk Klasifikasi Tingkat,” no. November, 2017.

- [15] D. A. Silitonga, A. P. Windarto, and D. Hartama, “Penerapan Metode K-Medoid pada Pengelompokan Rumah Tangga Dalam Perlakuan Memilah Sampah Menurut Provinsi,” *Semin. Nas. Sains Teknol. Inf. SENSASI 2019* ISBN, pp. 313–318, 2019.
- [16] No Title,” vol. 12, no. April, pp. 1–16, 2017.
- [17] Written, Frank. 2007. *Data Mining Practical Machine Learning tools and Techniques*. USA.