

Automated Essay Scoring untuk Bahasa Bali menggunakan Metode N-Gram String Similarity Matching

Made Agus Putra Subali¹, I Ketut Putu Suniantara²

^{1,2} Sistem Informasi,

Institut Teknologi dan Bisnis STIKOM Bali

Denpasar, Indonesia

email: ¹madeagusputrasubali@gmail.com, ²suniantara@stikom-bali.ac.id

Abstrak

Essay question merupakan model pertanyaan yang meminta jawaban dalam bentuk frasa atau kalimat, umumnya digunakan untuk mengukur kemampuan berpikir yang kompleks. Proses menilai jawaban pada soal uraian merupakan tugas yang rumit serta memerlukan waktu lebih lama. Adapun solusi dalam memudahkan proses penilaian pada soal uraian dapat dilakukan dengan menggunakan automated essay scoring. Salah satu metode yang dapat digunakan pada automated essay scoring adalah n-gram string similarity matching yang memiliki kelebihan dalam memperhatikan urutan kata maupun karakter pada suatu teks. Data yang digunakan dalam penelitian ini adalah lima belas soal beserta jawaban short sentence dalam Bahasa Bali. Berdasarkan hasil yang diperoleh antara metode yang diusulkan dan penilaian expert diperoleh akurasi sebesar 87%. Pada penelitian selanjutnya akan memperhatikan kata yang dianggap penting dengan memberikan bobot pada kata.

Kata kunci: *automated essay scoring, short answer, n-gram string similarity matching, bahasa bali.*

Abstract

Essay question is a question model that asks for answers in phrases or sentences, generally used to measure complex thinking skills. The assessing the answers to the description questions is a complex task and takes longer. The solution to facilitate the assessment process on essay questions can be done using automated essay scoring. One method that can be used in automated essay scoring is n-gram string similarity matching which has the advantage of paying attention to the order of words and characters in a text. The data used in this study were fifteen questions along with the answers to short sentences in Balinese. Based on the results obtained between the proposed method and the expert assessment, an accuracy of 87% was obtained. In future research, we will pay attention to words that are considered important by giving weight to words.

Keywords: *automated essay scoring, short answer, n-gram string similarity matching, bahasa bali.*

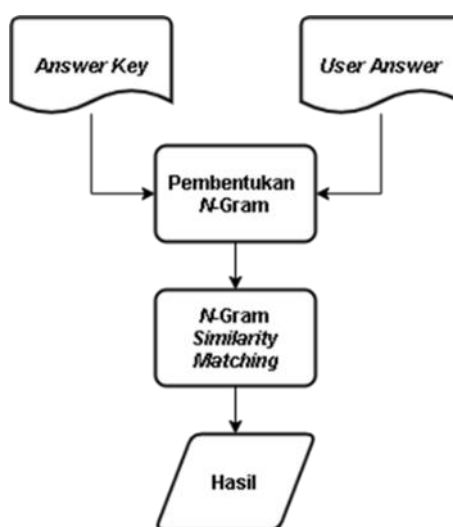
1. Pendahuluan

Essay question atau soal uraian merupakan model pertanyaan yang meminta jawaban dalam bentuk frasa ataupun kalimat, umumnya digunakan untuk mengukur kemampuan berpikir yang kompleks [1]. Proses menilai jawaban pada soal uraian merupakan tugas yang rumit serta memerlukan waktu lebih lama [2]. Adapun salah satu solusi dalam memudahkan proses penilaian pada soal uraian dapat dilakukan dengan menggunakan *automated essay scoring* (AES) [3]. Penggunaan AES pada kondisi tertentu sukses memberikan tingkat akurasi yang baik [4]. Metode AES seperti *cosine similarity* [5], [6] diusulkan oleh banyak peneliti. Penelitian terdahulu tentang *automated essay scoring* dilakukan oleh Fauzi, dkk. pada sistem *elearning* menggunakan metode *cosine similarity* untuk mengukur kemiripan antara *answer key* dan *user answer*, serta pada proses ekstraksi fitur menggunakan metode *n-gram*, hasil akurasi yang diperoleh sebesar 67% [6]. Penelitian *automated essay scoring* lainnya pernah dilakukan oleh Citawan, dkk. pada sistem *elearning* menggunakan metode *latent semantic analysis*, *cosine similarity*, dan *n-gram*, hasil akurasi yang diperoleh sebesar 78,65% [5]. Penelitian lain terkait AES Bahasa Bali pernah dilakukan oleh Subali, dkk. dimana dalam penelitian tersebut menerapkan sistem *question answering* menggunakan metode *cosine similarity* dan *rule-based*, sedangkan untuk data yang digunakan adalah 50 dokumen Bahasa Bali dan pada tahap pengujian menggunakan 20 pertanyaan, hasil dari penelitian tersebut memberikan akurasi sebesar 40% [7].

Metode AES menggunakan *cosine similarity* memiliki kelemahan pada saat proses pembentukan *bag of words* yang tidak memperhatikan struktur urutan kata pada kalimat [8], sehingga dua kalimat dengan komposisi kata yang sama dengan urutan kata yang berbeda akan memperoleh tingkat kemiripan yang tinggi. Pada penelitian terdahulu [5], [6] metode *n*-gram digunakan untuk mengatasi permasalahan tersebut dengan mengambil *n*-word pada kalimat, namun *n*-word memiliki kelemahan apabila diterapkan pada *short sentence* yang terdiri dari satu kata. Metode *n*-gram *string similarity matching* digunakan untuk mengatasi kelemahan tersebut, dimana *n*-gram dapat digunakan untuk memprediksi urutan karakter pada kata [9]. Data yang digunakan dalam penelitian ini adalah lima belas soal beserta jawaban *short sentence* dalam Bahasa Bali. Penelitian NLP dalam Bahasa Bali saat ini belum begitu jamak dilakukan, baik dari penggunaan data yang digunakan, teknik praproses, seperti *stop word removal* dan *stemmer*, serta proses utama yang melibatkan metode *data mining* masih jarang ditemui, padahal Bahasa Bali sendiri memiliki kaidah atau aturan baku dalam penggunaannya, selain itu dengan fokus pada topik Bahasa Bali kami turut melestarikan tradisi dengan menyediakan referensi terkait penelitian yang telah kami publikasikan.

2. Metode Penelitian

Pada Gambar 1 merupakan metode yang digunakan.



Gambar 10. Metode Sistem *Question Answering*

2.1. Answer Key dan User Answer

Pada tahap ini akan diberikan dua *input* yaitu *answer key* dan *user answer*. Pada *answer key* maupun *user answer* dilakukan proses mengubah setiap karakter ke bentuk *lowercase*.

2.2. Pembentukan N-Gram

Metode *n*-gram merupakan metode yang membentuk urutan *n* karakter dari setiap kata atau *string* [10]. Nilai *n* merupakan panjang karakter yang akan digunakan untuk memisahkan antara karakter pada kata. Panjang *n* satu sering disebut *uni*-gram, *n* dua disebut *bi*-gram, dan *n* tiga disebut *tri*-gram. Metode *n*-gram kerap kali digunakan untuk proses NLP [11].

Pada tahap ini *answer key* dan *user answer* dikonversi ke bentuk *bi*-gram, sebagai contoh diberikan kata pada *answer key* “*selem*” dan *user answer key* “*siap selem*”:

- *answer key*: kata “*selem*” menjadi *s, se, el, le, em, m*.
- *user answer*: kata “*siap selem*” menjadi *s, si, ia, ap, p*, *s, se, el, le, em, m*.

Berdasarkan hasil konversi ke bentuk *bi*-gram kata “*selem*” menghasilkan enam karakter sedangkan untuk kata “*siap selem*” menghasilkan sebelas karakter.

2.3. N-Gram String Similarity Matching

Hasil dari tahap pembentukan *n*-gram antara *answer key* dan *user answer* kemudian dilakukan perhitungan tingkat similaritasnya menggunakan metode *n*-gram *string similarity matching* [9] yang terlihat pada persamaan (1). Pada penelitian terdahulu yang dilakukan oleh Subali, dkk. menerapkan metode *n*-

gram *string similarity matching* sebagai solusi pada saat algoritma *rule-based stemmer* yang diusulkan tidak mampu mengenali kata berafiks atau kata imbuhan yang diberikan pada saat proses *stemming* [12].

$$dc = (2 \cdot c) \div (a + b) \quad (1)$$

Pada persamaan (1) menerima tiga parameter a, b, c yang diperoleh dari hasil tahapan pembentukan n -gram.

Keterangan:

dc : hasil perhitungan tingkat kemiripan.

a : jumlah n -gram pada kata pertama.

b : jumlah n -gram pada kata kedua.

c : jumlah n -gram yang sama antara a dan b .

Sebagai contoh tingkat similaritas antara kata *answer key* “*selem*” dan *user answer* “*siap selem*” memperoleh hasil 0,7.

$$dc = (2 \cdot 6) \div (6 + 11) = 0,7$$

Hasil tingkat similaritas yang dihasilkan berkisar antara 0-1. Semakin tinggi tingkat kemiripan berarti hasil similaritas mendekati 1 sebaliknya semakin rendah tingkat kemiripan, hasil similaritas mendekati 0. Pada penelitian ini nilai ambang batas tingkat similaritas untuk keterangan *true* yang ditetapkan adalah $\geq 0,6$.

3. Hasil dan Pembahasan

3.1. Data

Data penelitian yang digunakan adalah 15 pertanyaan Bahasa Bali yang diperoleh dari buku LKS SD pada mata pelajaran Bahasa Bali. Pada Tabel 1 merupakan kelima belas pertanyaan. Apabila diperhatikan panjang kalimat pada setiap jawaban terdiri dari satu hingga enam kata. Selain itu, terdapat pula kunci jawaban berupa karakter angka pada soal nomor sepuluh, hal ini tidak akan berakibat fatal apabila jawaban pengguna adalah karakter angka yang sama.

Tabel 1. Kelima Belas Pertanyaan

No.	Pertanyaan	Kunci Jawaban
1	<i>dadong dauh ngelah siap mabulu mawarna kenken?</i>	<i>putih</i>
2	<i>kuda siap putihe mataluh?</i>	<i>limolas</i>
3	<i>nyen ane ngelah siap putih?</i>	<i>dadong dauh</i>
4	<i>apa krana i kekua berang aking?</i>	<i>sawireh tuna amah</i>
5	<i>dija i lutung nepukin tongos melah?</i>	<i>di dauh tukad ada pondok melah</i>
6	<i>i kancil dadi juru kemit apa?</i>	<i>gong batara siwa</i>
7	<i>dugas i malu nyen ane matanduk?</i>	<i>i cicing</i>
8	<i>kenken warnan bok i kaki?</i>	<i>putih</i>
9	<i>ring dija i lutung masayuban?</i>	<i>di beten punyan kayune</i>
10	<i>i gusti ngurah rai padem pinanggal kuda?</i>	<i>20 november 1946</i>
11	<i>nyen adan nyamane i cupak?</i>	<i>i grantang</i>
12	<i>dija umah i benaru?</i>	<i>di tengah goane</i>
13	<i>buin akuda kone merine pan meri masisa?</i>	<i>aukud</i>
14	<i>nyen adan anak gede selem sane maakin pan meri?</i>	<i>i gede selem</i>
15	<i>dija i ketimun mas maumah?</i>	<i>desa dauh yeh</i>

3.2. Inisialisasi Answer Key dan User Answer

Sebagai contoh diberikan nilai “*selem*” untuk kunci jawaban dan “*siap selem*” untuk jawaban pengguna. Pada Gambar 2 merupakan inisialisasi untuk kunci jawaban dan jawaban pengguna.

```
key = 'selem' # Answer Key
user = 'siap selem' # User Answer
```

Gambar 2. Inisialisasi *Answer Key* dan *User Answer*

3.3. Pembentukan N-Gram

Pembentukan *n*-gram dilakukan dengan bantuan *library* nltk, dimana nilai kunci jawaban “selem” menjadi “*s, se, el, le, em, m*” serta jawaban pengguna “siap selem” menjadi “*s, si, ia, ap, p<space>, <space>s, se, el, le, em, m*”. Pada Gambar 3 di baris 6-18 merupakan perintah yang digunakan untuk pembentukan *n*-gram.

```

1  from nltk.util import ngrams
2
3  def ngram(key, user, n = 2, minimum = 0.55, maksimum = 1.0):
4      hasil = 0
5
6      a = list(ngrams(key,
7                n,
8                pad_left = True,
9                pad_right = True,
10               left_pad_symbol = '*',
11               right_pad_symbol = '*'))
12
13     b = list(ngrams(user,
14               n,
15               pad_left = True,
16               pad_right = True,
17               left_pad_symbol = '*',
18               right_pad_symbol = '*'))
19
20     dice = (2.0 * len(set(a) & set(b))) / (len(a) + len(b))
21
22     if(dice >= minimum):
23         if(dice <= maksimum):
24             hasil = dice
25
26     return dice

```

Gambar 3. Fungsi N-Gram String Similarity Matching

3.4. N-Gram String Similarity Matching

Proses menghitung tingkat kemiripan dilakukan dengan membandingkan antara *n*-gram pada kunci jawaban dan jawaban pengguna menggunakan persamaan (1). Pada Gambar 3 di baris 20 merupakan implementasi dari persamaan (1), apabila tingkat kemiripan yang diperoleh ≥ 0.6 dan ≤ 1 maka hasil akan ditampilkan. Pada Gambar 4 adalah hasil yang diperoleh sebesar 0.7.

0.7058823529411765

Gambar 4. Hasil Tingkat Kemiripan

3.5. Hasil dan Evaluasi

Pada Tabel 2 merupakan tingkat kemiripan antara kunci jawaban dan jawaban pengguna menggunakan metode yang diusulkan.

Tabel 2. Hasil Pengujian Akurasi

No.	Soal	User Answer	Kemiripan	Expert
1	Soal 1	putih	1	Benar
2	Soal 2	molas	0.71	Benar
3	Soal 3	i dadong	0.57	Benar
4	Soal 4	tuna amah	0.64	Benar
5	Soal 5	dauh tukad	0.47	Salah
6	Soal 6	sekaa gong	0.28	Salah
7	Soal 7	cicing	0.62	Benar
8	Soal 8	putih	1	Benar
9	Soal 9	beten punyan kayu	0.73	Benar
10	Soal 10	20 november 1946	1	Benar
11	Soal 11	gerantang	0.57	Benar
12	Soal 12	di tengah goa	0.86	Benar
13	Soal 13	ukud	0.72	Benar
14	Soal 14	i selem	0.76	Benar
15	Soal 15	dauh yeh	0.78	Benar

Berdasarkan pengujian yang telah dilakukan terdapat pertanyaan yang tidak memiliki kandidat jawaban lain (nomor satu, delapan, dan sepuluh), sedangkan pada pertanyaan nomor tiga, lima, enam, dan sebelas memperoleh skor di bawah nilai ambang batas, dan untuk pertanyaan nomor dua, empat, tujuh, sembilan, dua belas hingga lima belas memperoleh akurasi di atas nilai ambang batas.

Sedangkan penilaian yang diberikan oleh *expert* terdapat dua kandidat jawaban yang dinilai salah (nomor lima dan enam) dan terdapat ketidakseuaian penilaian antara tingkat kemiripan yang diperoleh dan penilaian *expert* pada nomor tiga dan sebelas. Berdasarkan perbandingan hasil tersebut akurasi yang diperoleh dari metode yang diusulkan adalah 87% berdasarkan perhitungan menggunakan persamaan (2).

$$s = \frac{t}{n} \times 100 \quad (2)$$

Keterangan:

s: hasil perhitungan akurasi.

t: total respons sesuai dari keseluruhan pertanyaan.

n: total pertanyaan.

4. Kesimpulan

Berdasarkan perbandingan hasil penilaian oleh *expert*, metode *n*-gram *string similarity matching* yang diterapkan pada kelima belas soal *short answer* Bahasa Bali dapat memperoleh akurasi sebesar 87%. Pada hasil pengujian, di nomor soal tiga, lima, enam, dan sebelas yang memperoleh nilai dibawah ambang batas dikarenakan selain secara struktur kata tidak sesuai dengan kunci jawaban, hal lain yang mempengaruhi adalah panjang kata antara kunci jawaban dan jawaban pengguna. Pada penelitian selanjutnya akan dilakukan pengembangan pada metode *n*-gram *string similarity matching* dengan memperhatikan panjang serta kata kunci pada kata yang dibandingkan.

Daftar Pustaka

- [1] J. O. Contreras, S. Hilles, and Z. B. Abubakar, "Automated Essay Scoring with Ontology based on Text Mining and NLTK Tools," 2018.
- [2] H. Chen, J. Xu, and B. He, "Automated Essay Scoring by Capturing Relative Writing Quality," *The Computer Journal*, vol. 57, no. 9, pp. 1318–1330, 2014.
- [3] Y. Salim, V. Stevanus, E. Barlian, A. C. Sari, and D. Suhartono, "Automated English Digital Essay Grader Using Machine Learning," 2019.
- [4] D. S. McNamara, S. A. Crossley, and R. Roscoe, "Natural language processing in an intelligent writing strategy tutoring system," *Behavior Research Methods*, vol. 45, pp. 499–515, 2013.
- [5] R. S. Citawan, V. C. Mawardi, and B. Mulyawan, "Automatic Essay Scoring in E-learning System Using LSA Method with N-Gram Feature for Bahasa Indonesia," 2017.
- [6] M. A. Fauzi, D. C. Utomo, B. D. Setiawan, and E. S. Pramukantoro, "Automatic Essay Scoring System Using N-Gram and Cosine Similarity for Gamification Based E-Learning," 2017.
- [7] M. A. P. Subali and P. Wijaya, "Sistem Question Answering untuk Bahasa Bali menggunakan Metode Rule-Based dan String Similarity," *Techno.COM*, vol. 20, no. 2, pp. 300–308, 2021.
- [8] A. R. Lahitani, A. E. Permanasari, and N. A. Setiawan, "Cosine similarity to determine similarity measure: Study case in online essay assessment," 2016.
- [9] T. M. Sembok and Z. A. Bakar, "Effectiveness of Stemming and N-Grams String Similarity Matching on Malay Documents," *International Journal of Applied Mathematics and Informatics*, vol. 5, no. 3, pp. 208–215, 2011.
- [10] I. F. Rozi, A. T. Firdausi, and K. Islamiyah, "Analisis Sentimen pada Twitter mengenai pasca Bencana menggunakan Metode Naïve Bayes dengan Fitur N-Gram," *Jurnal Informatika Polinema*, vol. 6, no. 2, pp. 33–39, 2020.
- [11] A. I. Fahma, I. Cholissodin, and R. S. Perdana, "Identifikasi Kesalahan Penulisan Kata (Typographical Error) pada Dokumen Berbahasa Indonesia menggunakan Metode N-Gram dan Levenshtein Distance," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 2, no. 1, pp. 53–62, 2018.
- [12] M. A. P. Subali and C. Fatichah, "Kombinasi Metode Rule-Based dan N-Gram Stemming untuk Mengenali Stemmer Bahasa Bali," *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIIK)*, vol. 6, no. 2, 2019.