

Implementasi Manhattan Distance dan Dice Similarity pada Sistem Ujian Esai Daring Berbahasa Indonesia

Faisal Rahutomo
Jurusan Teknologi Informasi
Politeknik Negeri Malang
faisal.polinema@gmail.com

Yoga Pramana Putra
Jurusan Teknologi Informasi
Politeknik Negeri Malang
pramanay@gmail.ac.id

Muhammad Haidar Ali
Jurusan Teknologi Infomasi
Politeknik Negeri Malang
haidariyyah96@gmail.com

Abstrak— Setiap proses pembelajaran memerlukan alat evaluasi untuk mengukur tingkat pemahaman siswa. Jenis evaluasi dapat berupa soal pilihan ganda, isian singkat dan esai. Beberapa penelitian mengungkapkan ujian esai lebih baik dari jenis evaluasi lainnya. Penilaian esai secara otomatis dibutuhkan untuk menghemat waktu pengajar dalam mengoreksi jawaban. Namun, pengembangan untuk penilaian esai masih terus dilakukan sampai sekarang. Tujuannya adalah untuk memperoleh nilai keakurasian yang lebih baik dari metode yang digunakan dalam penilaian. Berdasarkan permasalahan tersebut, penelitian ini mengusulkan analisis perbandingan metode kemiripan untuk penilaian ujian esai daring. Metode kemiripan yang dibandingkan adalah Dice Similarity, dan Manhattan Distance. Kedua metode menghasilkan nilai koefisien yang kemudian dibandingkan penilaian sistem dengan penilaian manual dengan skala sama. Data yang digunakan sebanyak 2162 data. Data ini diperoleh dari 50 siswa yang menjawab 40 soal (politik, olahraga, lifestyle dan teknologi). Data yang diperoleh dalam penelitian ini dapat digunakan untuk mendukung penelitian lain yang dapat diakses di alamat www.indonesian-ir.org. Penelitian ini menunjukkan bahwa skema Jaccard dengan stemming lebih kecil dibandingkan 2 skema kemiripan lainnya.

Kata kunci—*penilaian esai daring; skema kemiripan vektor; dice similarity; Manhattan Distance;*

I. PENDAHULUAN

Perkembangan teknologi informasi yang sedemikian maju telah banyak membantu manusia dalam segala bidang. Salah satunya adalah bidang pendidikan. Teknologi ini mengatasi keterbatasan waktu dan ruang dalam sebuah pembelajaran konvensional. Metode-metode juga banyak dikembangkan dalam sisi pembelajaran dan teknologi pendukung.

Setiap proses pembelajaran memerlukan suatu alat evaluasi untuk mengukur tingkat pemahaman siswa. Banyak macam dari jenis evaluasi, mulai dari soal pilihan ganda, isian singkat hingga esai. Beberapa penelitian mengungkapkan bahwa soal pilihan ganda dan isian singkat kurang memadai dalam proses belajar mengajar. Sebaliknya, ujian esai dapat melatih penyampaian suatu informasi secara verbal, ujian ini juga menuntut pemahaman yang lebih baik. Sehingga penilaian dalam soal esai dapat mengukur tingkat pemahaman lebih mendalam.

Banyak keuntungan yang dapat diperoleh dari penilaian esai otomatis dibandingkan penilaian secara tradisional. Dalam catatan di Inggris, guru menghabiskan 30% waktunya untuk mengoreksi jawaban siswa dan menghilangkan sekitar 30 milyar pound dalam setahun karena hal tersebut. Sehingga dapat dibayangkan keuntungannya apabila sebuah institusi pendidikan memiliki sebuah sistem untuk penilaian otomatis terutama untuk esai.

Saat ini, banyak pengembangan e-learning untuk penilaian ujian pilihan ganda, isian singkat dan esai. Namun, pengembangan untuk penilaian esai masih terus dilakukan sampai sekarang. Tujuannya adalah untuk memperoleh nilai keakurasian yang lebih baik dalam penilaian. Hal ini dikarenakan banyaknya metode dalam menyatakan kesesuaian jawaban siswa dengan kunci jawaban yang telah disediakan oleh guru. Sayangnya, belum ada suatu analisis mengenai perbandingan dari metode- metode (skema) yang banyak digunakan saat ini.

Beberapa penelitian mengenai ujian esai berbahasa Indonesia adalah Penilaian Esai Jawaban Bahasa Indonesia Menggunakan Metode SVM-LSA dengan Fitur Generik yang menghasilkan keakurasian 73%, An Attempt to Create an Automatic Scoring Tool of Short Text Answer in Bahasa Indonesia pada aplikasi moodle memiliki standar deviasi sebesar 3-30 dari berbagai macam jenis data yang diujikan [1]. Sehingga mendasari belum adanya perbandingan antara vektor-vektor kemiripan yang ada, padahal hal ini sangat penting dalam menentukan metode yang digunakan dalam pembuatan sistem penilaian ujian esai daring. Selain itu, data yang digunakan pada penelitian sebelumnya masih menggunakan data ukuran kecil (kelas dengan jumlah responden sedikit dan jenis soal tidak banyak). Berdasarkan hal tersebut, penelitian ini ditujukan untuk mengetahui perbandingan skema-skema kemiripan pada Cosine Similarity, Euclidean Distance dan Jaccard menggunakan 4 bidang soal (setiap bidang soal terdapat 10 soal) dengan 50 siswa. Dimana data yang diperoleh dari penelitian ini akan disediakan oleh peneliti untuk kepentingan penelitian lain. Di sisi lain, penggunaan stemming dalam dunia pengolahan kata, khususnya penilaian ujian esai daring belum pernah dikemukakan hasilnya, maka dari itu, penelitian ini akan mengungkapkan

perbedaan nilai error (perbedaan nilai manual dengan nilai sistem) menggunakan stemming.

II. TEORI KEMIRIPAN VEKTOR

Dokumen dapat direpresentasikan sebagai sebuah vektor dimana setiap komponen mengacu pada sebuah term. Kemudian nilai dari tiap-tiap komponen adalah angka kemunculan term dalam sebuah dokumen. Begitu dokumen direpresentasikan sebagai sebuah vektor, dapat dilakukan bermacam-macam operasi vektor.

Skema kemiripan merupakan sebuah metode untuk mencari koefisien yang menunjukkan seberapa mirip antara sebuah dokumen dengan dokumen lain. Dalam penelitian ini, skema kemiripan akan digunakan untuk mencari kemiripan antara jawaban siswa dan kunci soal. Beberapa proses yang perlu dilakukan dalam menghitung kemiripan sebuah dokumen adalah sebagai berikut:

A. Text Pre-Processing

Pada data teks perlu adanya text pre-processing, yaitu mengubah data teks menjadi data numerik yang dapat diolah. Tahap ini adalah tahap yang sangat penting sebelum memulai proses perhitungan penilaian otomatis karena pada proses ini bisa mempengaruhi akurasi dari penilaian. Dalam pre-processing ada beberapa tahap yang harus dilakukan. Dalam penelitian ini tahap pre-processing dibedakan menjadi 2, yaitu dengan menggunakan tahap stemming dan tanpa tahap stemming. Tahapan pre-processing dalam text terdiri dari case folding (mengubah teks menjadi huruf kecil), tokenizing (memecah teks menjadi kata), stemming (mengubah kata menjadi kata dasar), dan stopword (membuang kata yang tidak diperlukan) [2].

B. Vektor Kunci dan Jawaban

Vektor kunci dan jawaban hasil dari text pre-processing untuk jawaban siswa dan kunci jawaban masing-masing dapat direpresentasikan dengan term vectors dalam bentuk berikut:

$$d = (d_0, d_1, \dots, d_n) \quad (1)$$

Dimana setiap d_k mengidentifikasi term yang terdapat dalam dokumen jawaban d . Demikian juga pada kunci jawaban q direpresentasikan dalam term vectors, sehingga dirumuskan:

$$q = (q_0, q_1, \dots, q_n) \quad (2)$$

Dimana setiap q_k mengidentifikasi term yang terdapat pada jawaban q . Sehingga apabila ditentukan bobot (weight) pada setiap term untuk membedakan diantara term yang terdapat dalam jawaban siswa maupun kunci jawaban dapat dituliskan:

$$w_d = (w_{d0}, w_{d1}, \dots, w_{dn}) \quad (3)$$

Dan untuk setiap dokumen jawaban siswa

$$w_q = (w_{q0}, w_{q1}, \dots, w_{qn}) \quad (4)$$

Dimana w_{dk} merupakan bobot dari term t_k dalam jawaban d , sedangkan w_{qk} merupakan bobot term t_k dalam kunci jawaban.

1) Algoritma Stemming Sastrawi

Algoritma nazief andriani merupakan algoritma stemming khusus untuk bahasa Indonesia. Algoritma ini menggunakan beberapa aturan morfologi untuk menghilangkan affiks (awalan, imbuhan, dll) dari sebuah kata dan kemudian mencocokkannya dalam kamus akar kata (kata dasar). Jadi dasar utama algoritma ini adalah daftar kata dasar [3]. Langkah pertama yang dilakukan adalah mengumpulkan daftar kata dasar dalam bahasa Indonesia. Semakin lengkap daftarnya, semakin tinggi akurasi algoritma ini. Algoritma ini memiliki tahap-tahap sebagai berikut:

- Cari kata yang akan di-stem di dalam kamus, jika kata tersebut ditemukan maka kata tersebut adalah kata dasar dan algoritma berhenti. Jika tidak ada maka lanjutkan ke langkah-2.
- Hilangkan inflectional suffix (imbuhan infleksional) yaitu (“-lah”, “-kah”, “-tah”, “-ku”, “-mu”, “-nya”).
- Hapus derivation suffix (imbuhan turunan) yaitu (“-i”, “-an”, atau “-kan”). Jika kata ditemukan di kamus, maka algoritma berhenti.
- Hapus derivation prefix (awalan turunan) yaitu (“be-“, “di-“, “ke-“, “me-“, “pe-“, “se-“, “te-“).
- Lakukan recording.
- Jika semua langkah telah selesai tetapi tidak juga berhasil maka kata awal diasumsikan sebagai kata dasar. Proses selesai.

C. Bobot Kata

Term vektor selanjutnya dihitung term frequency (TF)-nya. TF merupakan kemunculan setiap term bila dibandingkan dengan term yang muncul pada dokumen tersebut. Bobot lokal suatu term i di dalam dokumen j (w_{ij}) dapat didefinisikan sebagai Persamaan 6.

$$w_{ij} = \frac{f_{ij}}{\sum f_{ij}} \quad (5)$$

Tabel 1 merupakan contoh dari term frequency matrix 4x5, dimana terdiri dari 4 term, yaitu air, kapur, bakar dan reaksi. Sedangkan setiap baris mendeskripsikan dokumen, yaitu kunci, jawaban 1, jawaban 2, jawaban 3 dan jawaban 4.

TABEL 1. TABEL SVM

Term	air	kapur	bakar	reaksi
Kunci	0,5	0,2	0	0,3
Jawaban 1	0,3	0,4	0,2	0,1
Jawaban 2	0,75	0,25	0	0
Jawaban 3	1	0	0	0
Jawaban 4	0,5	0,5	0	0

Term frequency matrix dihitung kemiripannya dengan menggunakan skema kemiripan Dice Similarity, dan Manhattan Distance.

1) Dice Similarity

Dice Similarity didefinisikan sebagai metode pengukuran kemiripan terhadap dua jumlah string istilah yang umum dalam string yang dibandingkan dan dibagi dengan jumlah total istilah yang ada di kedua string [4].

$$DiceSim(A, B) = \frac{2(X \cap Y)}{|X| + |Y|} \quad (6)$$

Dimana A dan B mewakili terms common dari dokumen a dan b; $W_{a,t}$ dan $W_{b,t}$ adalah bobot dari term t dalam dokumen a dan b secara respektif atau masing-masing.

2) Manhattan Distance

Manhattan distance adalah metode pengukuran jarak dua vektor. Manhattan distance menghitung jumlah dari perbedaan fitur antara dua vektor dengan n dimensi. Nilai yang dijumlahkan adalah nilai absolut dari masing-masing fitur yang dihitung perbedaannya. Berikut adalah persamaan Manhattan distance.

$$Manhattan(A, B) = \sum_{k=1}^n |w_{AK} - w_{BK}| \quad (7)$$

3) Error Rate

Setiap skema kemiripan akan dihitung nilai percentage error dan absolute error. Nilai percentage error menunjukkan seberapa besar perbedaan antara pengukuran dengan nilai fakta, lihat Persamaan 8. Nilai error yang kecil mengindikasikan bahwa tingkat kesalahan penilaian dari sistem semakin baik.

$$Nilai \% Error = \frac{Nilai Manual - Nilai Sistem}{Nilai Manual} \times 100\% \quad (8)$$

III. METODOLOGI PENELITIAN

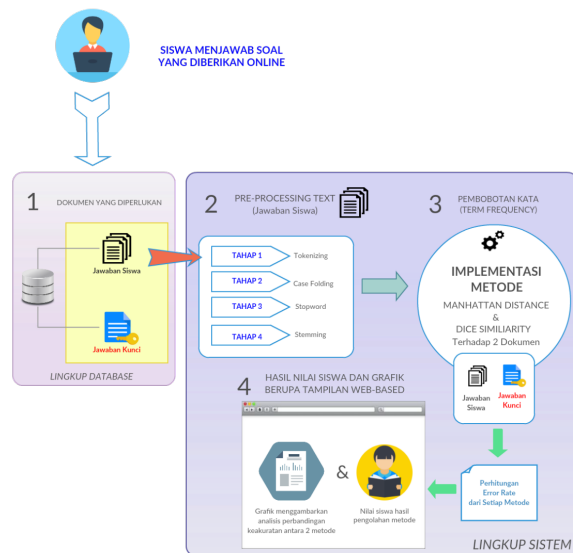
Metode penelitian dibagi menjadi beberapa phase. Phase tersebut secara detail akan dijelaskan pada yang akan terlampir di bawah. Gambar tersebut menunjukkan skema analisis dalam penelitian.

Masalah yang dihadapi adalah waktu yang terkuras banyak bagi para pengajar untuk mengkoreksi nilai esai dari setiap mahasiswa. Sehingga diperlukan solusi untuk mempermudah dan memangkas waktu para pengajar dalam memberikan nilai [5].

Fase pertama adalah membuat soal dan kunci jawaban, dimana masing-masing memiliki kategori. Fase kedua adalah siswa menjawab soal dengan sistem penilaian esai daring.

Fase ketiga adalah jawaban siswa dinilai oleh sistem. Penilaian sistem dilakukan dengan melakukan text-preprocessing dan perhitungan dengan 3 skema kemiripan. Fase keempat adalah memberi nilai manual jawaban siswa oleh 3 orang penilai yang kemudian diambil nilai rata-ratanya.

Fase kelima adalah menghitung nilai percentage error antara rata-rata penilaian manual dengan penilaian dari sistem. Berdasarkan fase ini, diperoleh nilai error masing-masing skema kemiripan [6]. Berikut adalah alur sistem skoring ujian online sebagaimana yang tertera pada Gambar 1.



Gambar 1. Alur Sistem Skoring Otomatis

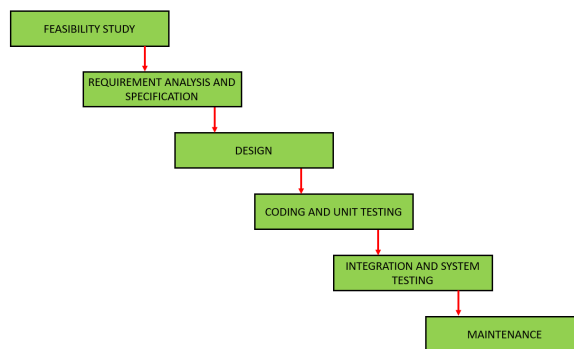
A. Pengembangan Sistem

Berdasarkan hal tersebut, penelitian ini ditujukan untuk mengetahui perbandingan skema-skema kemiripan pada Cosine Similarity, Euclidean Distance dan Jaccard menggunakan 4 bidang soal (setiap bidang soal terdapat 10 soal) dengan 50 siswa. Dimana data yang diperoleh dari penelitian ini akan disediakan oleh peneliti untuk kepentingan penelitian lain.

Di sisi lain penggunaan stemming dalam dunia pengolahan kata, khususnya penilaian ujian esai daring belum pernah dikemukakan hasilnya, maka dari itu, penelitian ini akan mengungkapkan perbedaan nilai error (perbedaan nilai manual dengan nilai sistem) menggunakan stemming dan tanpa stemming [7].

Metode dalam pengembangan sistem informasi ini menggunakan metode waterfall. Model waterfall merupakan sebuah metode pengembangan software yang bersifat sekuensial.

Metode ini dikenalkan oleh Royce pada tahun 1970 pada saat itu disebut sebagai siklus klasik dan sekarang ini lebih dikenal dengan sekuensial linier sebagaimana yang ditunjukkan di Gambar 2.



Gambar 2. Metode Waterfall

B. Pengumpulan Data

Pengambilan data dilakukan untuk mendukung analisis skema kemiripan dari Sistem Penilaian Ujian Esai Daring. Data yang dibutuhkan untuk membangun sistem ini terdiri dari dataset komprehensif berupa 40 soal yang dimana setiap 10 soalnya memiliki kategori yang berbeda-beda (Lifestyle, Politics, Sport, Technology).

Jawaban yang terkumpul yaitu sebanyak 2162 (dua ribu seratus enam puluh dua) teks jika dihitung total keseluruhannya. Selanjutnya, jawaban kunci dan jawaban siswa tersebut akan diolah oleh proses text pre-processing dan kedua metode Manhattan Distance dan Dice Similarity sebagai perbandingan keakuratannya [8].

Penilaian manual dilakukan oleh 3 orang. Pemilihan jumlah 3 dilakukan untuk menjaga objektivitas penilaian. Tiga penilai memiliki latarbelakang usia, gender, dan pekerjaan yang berbeda-beda. Nilai yang diberikan pada jawaban siswa memiliki rentang antara 0-100 [9]. Dalam pemberian nilai manual ini, penilai tetap memperhatikan kunci jawaban sebagai patokan penilaian.

C. Use Case Diagram

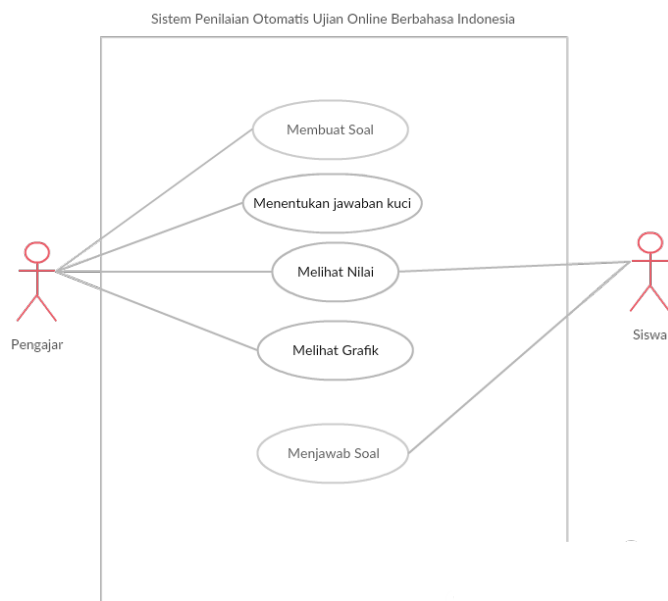
Peran use case disini digunakan untuk pemodelan serta menggambarkan kelakuan (behavior) sistem yang akan dibuat. Diagram use case mendeskripsikan sebuah interaksi antara satu atau lebih aktor dengan sistem yang akan dibuat.

Use Case juga berfungsi untuk mendeskripsikan sebuah interaksi antara satu atau lebih aktor dengan sistem yang akan dibuat [10].

Alur sistem secara keseluruhan untuk sistem penilaian otomatis ujian esai online berbahasa Indonesia dengan metode Manhattan Distance dan Dice Similarity [11]. Implementasi dalam bentuk usecase adalah sebagai berikut sebagaimana yang ditunjukkan pada Gambar 3.

REFERENSI

- [1] F. Rahutomo, Z. Hanif, R. Adi, and I. F. Rozi, "Implementasi Text Mining Pada Laman Blog di," pp. 101–109, 2018.
- [2] F. Rahutomo and A. Hafidh Ayatullah, "Indonesian Dataset Expansion of Microsoft Research Video Description Corpus and Its Similarity Analysis," *Kinet. Game Technol. Inf. Syst. Comput. Network, Comput. Electron. Control*, vol. 3, no. 4, p. 319, 2018.
- [3] F. Rahutomo, P. Y. Saputra, C. Febriawan, and P. Putra, "Implementasi Explicit Semantic Analysis Berbahasa Indonesia Menggunakan Corpus Wikipedia Indonesia," *J. Inform. Polinema*, vol. 4, no. 4, pp. 252–257, 2018.
- [4] M. Astiningrum *et al.*, "Implementasi nlp dengan konversi kata pada sistem chatbot konsultasi laktasi," vol. 5, no. November, pp. 46–52, 2018.
- [5] Aggrawal, C. C., & Zai, C. (n.d.). Mining Text Data.
- [6] Dalglish, T., Williams, J. M. G., Golden, A.-M. J., Perkins, N., Barrett, L. F., Barnard, P. J., ... Watkins, E. (2007). Text Mining Application and Theory. *Journal of Experimental Psychology: General* (Vol. 136).
- [7] Feldman, R., & Sanger, J. (n.d.). The Text Mining handbook. <https://doi.org/10.1017/CBO9780511546914>
- [8] H.Gomaa, W., & A. Fahmy, A. (2013). A Survey of Text Similarity Approaches. *International Journal of Computer Applications*, 68(13), 13–18. <https://doi.org/10.5120/11638-7118>
- [9] M.K, V., & K, K. (2016). A Survey on Similarity Measures in Text Mining. *Machine Learning and Applications: An International Journal*, 3(1), 19–28. <https://doi.org/10.5121/mlaj.2016.3103>
- [10] Muzzammil, T. R., Ginardi, R. V. H., & Purwitasari, D. (2016). Modul Klasifikasi Aduan dengan Pendekatan Kemiripan Teks pada Aplikasi Perangkat Bergerak Suara Warga (SURGA) Kota Kediri, 5(1), 52–57. <https://doi.org/10.2307/4126447>
- [11] Roshinta, T. A., & Faisal, R. (2016). Analisis Aspek-Aspek Ujian Esai Daring Berbahasa Indonesia, 01, 1–26. <https://doi.org/10.1021/jp0629036>
- [12] Shoaib, M., Daud, A., Sikandar, M., & Khiyal, H. (2014). An Improved Similarity Measure for Text Documents. *J. Basic. Appl. Sci. Res*, 4(6), 215–223. <https://doi.org/10.13140/2.1.4814.4006>



Gambar 3. Use Case sistem