

PERFORMANCE OF TEXT SIMILARITY ALGORITHMS FOR ESSAY ANSWER SCORING IN ONLINE EXAMINATIONS

Muhammad Riza Radyaka Susanto^{*1}, Husni Thamrin^{*2}, Naufal Azmi Verdikha³

^{1,2}Faculty of Informatics Engineering, Universitas Muhammadiyah Surakarta, Indonesia

³Faculty of Informatics Engineering, Universitas Muhammadiyah Kalimantan Timur, Indonesia

Email: ¹1200180136@student.ums.ac.id, ²husni.thamrin@ums.ac.id, ³nav651@umkt.ac.id

(Article received: May 3, 2023; Revision: May 23, 2023; published: December 23, 2023)

Abstract

The purpose of assessment is to determine learning success. Exams with question descriptions have several advantages, including ease of preparation and the ability to reveal student comprehension and originality. The problem with space is that it takes time to fix. Therefore, it is important to develop algorithms and software that automatically evaluate space. With the help of this algorithm and this software, you can solve some exam and assessment problems. This study aims to investigate similarity algorithms that approximate human patterns in evaluating ambiguous answers. This study examines his five similarity algorithms, including TF-IDF and LSA. The data was a collection of correct answers with a total of 371 texts. The similarity algorithm's performance was compared with human correction results. Evaluation was performed using Root Mean Square Error (RMSE). This study shows that his TF-IDF algorithm like Jaccard has the lowest his RMSE compared to human judgement. However, the LSA algorithm tended better to follow human rating patterns for descriptive tests..

Keywords: *algorithm, answer scoring, essay grading, assessment, rmse*

KINERJA ALGORITMA SIMILARITAS TEKS UNTUK PENILAIAN JAWABAN URAIAN PADA UJIAN DARING

Abstrak

Tujuan asesmen adalah untuk menentukan keberhasilan pembelajaran. Ujian dengan menggunakan soal uraian memiliki beberapa manfaat, antara lain mudah disiapkan dan mampu mengungkap pemahaman dan orisinalitas siswa. Masalah soal uraian adalah waktu koreksi yang lama, oleh karena itu sangat penting untuk membangun algoritma dan perangkat lunak untuk mengevaluasi jawaban uraian secara otomatis. Sehingga dengan bantuan algoritma dan perangkat lunak ini bisa menyelesaikan permasalahan dalam ujian atau asesmen. Penelitian ini bertujuan untuk mengeksplorasi algoritma similaritas yang mendekati pola manusia dalam menilai jawaban uraian. Penelitian ini mengamati lima algoritma similaritas termasuk TF-IDF dan LSA. Data berjumlah 371 berupa teks dari jawaban ujian uraian yang telah dikoreksi. Kinerja algoritma similaritas dibandingkan dengan hasil koreksi manusia. Evaluasi dilakukan menggunakan *Root Mean Square Error* (RMSE). Penelitian ini menunjukkan bahwa algoritma TF-IDF dengan similaritas Jaccard memiliki RMSE terendah dibandingkan hasil penilaian manusia. Namun algoritma LSA memperlihatkan kecenderungan yang lebih baik dalam mengikuti pola penilaian ujian uraian oleh manusia.

Kata kunci: algoritma, asesmen, rmse, soal uraian, penilaian jawaban uraian

1. PENDAHULUAN

Kegiatan pembelajaran selain dilaksanakan secara luring atau tatap muka di kelas, dapat dilaksanakan secara daring memanfaatkan teknologi informasi dan komunikasi. Kegiatan pembelajaran secara daring penuh atau blended dapat memastikan proses pembelajaran terus berjalan meskipun ada situasi yang mengekang terjadinya pertemuan tatap muka [1].

Setiap teknik pembelajaran baik luring ataupun daring memiliki masalah masing-masing [2]. Permasalahan pada pembelajaran daring dapat muncul dari sisi infrastruktur server, infrastruktur konektivitas, organisasi, kemampuan dalam menggunakan teknologi, maupun dalam kesiapan konten pembelajaran [3]. Selain itu, pembelajaran daring memiliki masalah pada proses asesmen. Asesmen merupakan kegiatan pendidik untuk mendapat gambaran keberhasilan pembelajaran dari setiap peserta didik [4]. Salah satu bentuk asesmen

adalah kegiatan ujian yang bersifat formatif maupun ujian akhir [5].

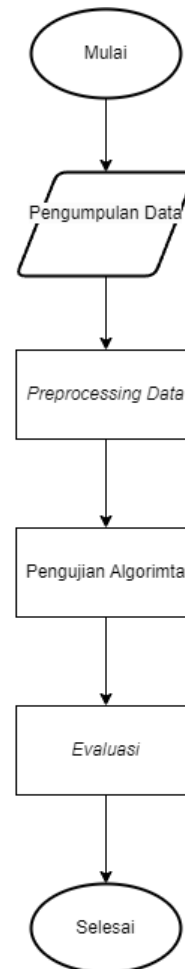
Kegiatan ujian daring biasanya menggunakan soal pilihan ganda karena sifatnya yang masif dan mudah dikoreksi, dan jarang digunakan soal uraian kecuali bertujuan untuk menguji kemampuan menulis. Soal uraian sebenarnya memiliki kelebihan yaitu mudah dibuat dan mampu menunjukkan tingkat pemahaman peserta didik pada berbagai level kognitif [6],[7]. Namun penggunaan soal uraian memiliki kelemahan yaitu proses koreksi yang membutuhkan waktu dan energi sebab setiap peserta didik bisa memiliki jawaban yang berbeda-beda dan mungkin berbentuk tulisan yang panjang [8], [9].

Tulisan ini menguraikan upaya mengatasi permasalahan koreksi soal uraian dengan memanfaatkan perangkat lunak. Upaya semacam ini telah dilakukan beberapa peneliti seperti oleh [10] yang menguji sistem penilaian otomatis dengan memanfaatkan teknik *Natural Language Processing* (NLP). Selain itu [11] telah mencoba algoritma similaritas TF-IDF untuk sebuah sistem pengkategorian jawaban menggunakan *Support Vector Machine* (SVM) yang diterapkan pada sistem penilaian otomatis jawaban esai berbahasa Jepang menggunakan metode *Latent Semantic Analysis* (LSA). Sedangkan [12] mencoba algoritma klasifikasi dan algoritma similaritas untuk menentukan nilai jawaban uraian. Sebelumnya [13] telah mencoba menerapkan algoritma similaritas untuk penilaian jawaban uraian singkat.

Penelitian kami menguji beberapa algoritma similaritas untuk menentukan skor jawaban uraian. Algoritma similaritas biasanya digunakan dalam *text mining*, deteksi plagiasi, dan klasifikasi [14]–[17]. Namun pada penelitian ini, algoritma similaritas digunakan untuk menentukan nilai mahasiswa dengan melihat kemiripan jawaban soal dengan jawaban kunci. Algoritma yang terbaik adalah algoritma yang menghasilkan nilai jawaban yang terdekat dengan jawaban yang dilakukan oleh manusia [18].

2. METODE PENELITIAN

Penelitian ini melalui beberapa tahapan yaitu seperti Gambar 1. Alur Diagram Penelitian berikut



Gambar 1. Alur Diagram Penelitian

2.1 Pengumpulan Data

Data dikumpulkan dari sebuah ujian semester yang melibatkan lebih dari 1000 mahasiswa. Jawaban mahasiswa disaring untuk diambil 450 data jawaban. Penyaringan dilakukan dengan mengambil jawaban dengan tingkat similaritas rendah di antara satu dengan yang lain, di mana similaritas dihitung menggunakan algoritma Cosine. Selain itu, dipilih jawaban yang memiliki jumlah kata di atas minimum dan di bawah maksimum. Data jawaban terpilih kemudian dikoreksi oleh tiga penilai (*rater*) dan setiap jawaban dikoreksi oleh dua orang.

Setiap jawaban diberi nilai antara 0 hingga 3 oleh penilai manusia (*human rater*). Selanjutnya data disaring lagi dengan memilih jawaban yang mendapat nilai koreksi tidak jauh berbeda antara satu penilai dengan penilai yang lain. Jika perbedaan nilai antara dua orang lebih dari 1 poin, maka jawaban tersebut disingkirkan. Akhirnya didapat sebanyak 371 data jawaban yang telah dinilai oleh dua orang dan nilai berada pada rentang 0 – 2.5.

2.2 Data Preprocessing

Pada tahapan ini, teks dibersihkan dari berbagai kode karakter yang tidak terbaca, angka, dan lain sebagainya yang memperlambat proses

komputasi. Pada tahapan ini juga dilakukan pembersihan spasi, dan pengulangan kata.

Setelah dilakukan *data preprocessing*, teks siap diolah dengan berbagai algoritma similaritas. Penelitian ini menguji 5 algoritma yaitu Jaccard, Cosine, TF-IDF dengan Cosine, TF-IDF dengan Jaccard, dan LSA.

2.3 Algoritma Similaritas Jaccard

Algoritma Jaccard similaritas merupakan algoritma untuk mengukur kesamaan dari dua buah objek [19]. Kesamaan dihitung dari nilai perpotongan (*intersection*) dua buah dokumen dibagi dengan nilai gabungan (*union*) dari kedua dokumen. Perhitungan similaritas Jaccard dilakukan dengan mengikuti persamaan (1).

$$J(A,B) = \frac{|A \cap B|}{|A \cup B|} \quad (1)$$

Pada persamaan (1), A dan B adalah vektor yang mewakili teks dari dokumen A dan B.

2.4 Algoritma Similaritas Cosine

Algoritma Cosine merupakan algoritma untuk mengukur similaritas atau kesamaan menggunakan kosinus sudut dari dua atau lebih vektor [20]. Algoritma Cosine dapat diekspresikan sebagai pembagian antara hasil kali skalar dua buah vektor dan nilai dari *Euclidean norm* setiap vektor seperti ditunjukkan pada persamaan (2).

$$\text{sim} = \frac{A \cdot B}{\|A\| \|B\|} = \frac{\sum_{i=1}^n A_i B_i}{\sqrt{\sum_{i=1}^n A_i^2} \sqrt{\sum_{i=1}^n B_i^2}} \quad (2)$$

Pada persamaan (2), A dan B adalah vektor yang mewakili teks dari dokumen A dan B.

2.5 TF-IDF (*Term Frequency-Inverse Document Frequency*)

TF-IDF merupakan suatu pengukuran mengabungkan dua buah konsep, yaitu *term frequency* (pengukuran kemunculan kata di sebuah dokumen) dan *inverse document frequency* (pengukuran kemunculan kata di seluruh dokumen). Nilai *tf* menunjukkan seberapa penting kata di dalam dokumen [21], [22]. Nilai frekuensi dokumen menunjukkan seberapa umum sebuah kata ada di berbagai dokumen. Nilai *tf-idf* dapat dihitung dengan menggunakan persamaan (3).

$$\text{idf} = \log\left(\frac{D}{d_f}\right), W = \text{tf} \times \text{idf} \quad (3)$$

Pada persamaan (3), nilai *tf* merupakan nilai *term frequency* atau jumlah suatu kata atau term pada dokumen. Nilai *idf* ditentukan dengan mendapatkan log dari D yaitu jumlah semua dokumen dibagi dengan *df* yaitu jumlah dari dokumen yang memiliki kata atau term. Sedangkan W merupakan bobot dari suatu kata yang ditentukan dengan perkalian antara nilai *tf* dan *idf*. Hasil dari proses TF-IDF adalah vektor yang menggambarkan

bobot kata dalam keseluruhan dokumen. Vektor tersebut selanjutnya menjadi dasar perhitungan skor jawaban, dan perhitungannya menggunakan algoritma similaritas seperti Jaccard atau Cosine.

2.6 LSA (*Latent Semantic Analysis*)

Latent semantic analysis adalah sebuah metode pengukuran yang menggunakan pendekatan statistik untuk menganalisa semantik sebuah kata di dalam dokumen [23], [24]. Pada algoritma LSA dilakukan perhitungan *term-document matrix* untuk membuat sebuah matrix dari vektor kata. Setelah itu, dilakukan *Singular Value Decomposition* (SVD) untuk mengurangi dimensi sehingga mempermudah dalam pemrosesan dokumen matriks. Pada SVD, matriks persegi panjang dipecah menjadi perkalian tiga matriks lainnya. Matriks pertama merepresentasikan entitas baris matriks asli sebagai vektor dari nilai faktor ortogonal. Matriks kedua merepresentasikan entitas kolom matriks asli dengan cara yang sama. Matriks ketiga adalah matriks diagonal yang berisi nilai skala sehingga jika ketiga matriks komponen dikalikan, matriks semula akan terbentuk (lihat persamaan 4).

$$A_{m \times n} = U_{m \times m} \times S_{m \times n} \times V_{n \times n}^T \quad (4)$$

Pada persamaan (4), A adalah matriks asli dengan m baris dan n kolom. Matriks U dan V merupakan matriks singular kanan dan kiri. Sedangkan S merupakan matriks diagonal.

Proses LSA menghasilkan vektor kata. Selanjutnya setiap teks jawaban dan teks kunci dapat direpresentasikan menggunakan vektor tersebut. Similaritas teks jawaban dengan teks kunci selanjutnya dihitung dengan algoritma Cosine.

2.7 Evaluasi

Kinerja algoritma diukur menggunakan *Root Mean Squared Error* (RMSE). *Error* yang dimaksud adalah perbedaan nilai yang dihitung oleh mesin menggunakan algoritma similaritas dengan nilai yang diberikan manusia (*rater*). Karena terdapat 371 data jawaban, maka terdapat 371 nilai *error* dari setiap data jawaban yang kemudian dikuadratkan, dijumlah dan diambil nilai akarnya menjadi nilai RMSE. Metrik ini digunakan karena nilai jawaban uraian bersifat angka kontinyu. Proses penilaian jawaban bukan proses klasifikasi, meskipun seringkali didekati sebagai proses klasifikasi. Karena merupakan nilai *error*, algoritma yang baik memiliki RMSE yang kecil.

3. HASIL DAN PEMBAHASAN

Pada penelitian ini dilakukan perhitungan nilai (skor) jawaban soal uraian menggunakan 5 algoritma. Nilai atau skor hasil perhitungan algoritma untuk setiap teks jawaban dibandingkan dengan nilai yang diberikan manusia (*rater*). Perbedaan nilai perhitungan algoritma dan nilai manusia dihitung sebagai nilai RMSE yang diperlihatkan pada Tabel

1. Tabel menunjukkan bahwa algoritma TF-IDF dengan perhitungan similaritas Jaccard memiliki nilai RMSE terkecil yaitu 0.3921. Nilai RMSE tersebut setara dengan persentase kesalahan rata-rata dalam memberi nilai sebesar 15.68%. Nilai dalam persen didapat karena nilai tertinggi untuk penilaian

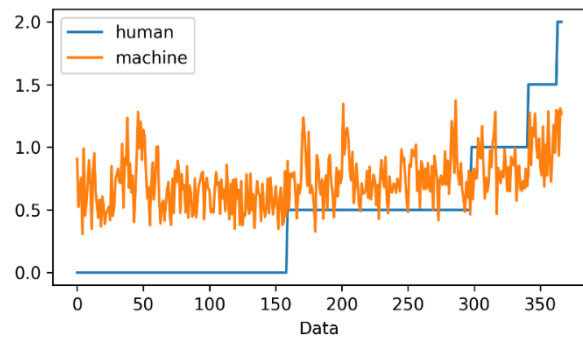
manusia adalah sebesar 2.5 dan rentang nilainya adalah 0 – 2.5. Nilai 2.5 dimiliki oleh 4 jawaban mahasiswa yang kemudian dijadikan jawaban kunci. Berikut tabel 1 yang menampilkan nilai RMSE dari setiap algoritma :

Tabel 1. Nilai RMSE Untuk Setiap Algoritma Yang Diuji

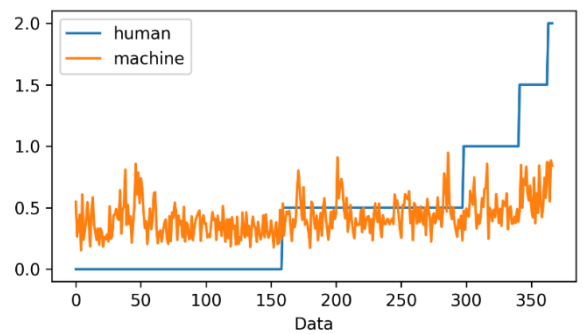
Algoritma	RMSE	Percent RMSE
Cosine	0.5158	20.63
Jaccard	0.4138	16.55
TF-IDF	0.4274	17.10
TF-IDF Jaccard	0.3921	15.68
LSA	0.5368	21.47

Tabel 1 menunjukkan bahwa algoritma LSA memiliki dengan RMSE terbesar yaitu di angka 0.5368 atau 21.47%. Algoritma lain memiliki RMSE di antara 0.4 hingga 0.5. Data pada tabel mengindikasikan bahwa algoritma TF-IDF dengan similaritas Jaccard merupakan algoritma terbaik untuk menentukan nilai jawaban uraian (*essay*) pada sebuah ujian berbasis komputer. Nilai yang diberikan memiliki kesalahan rata-rata terkecil terhadap nilai yang diberikan oleh penilai manusia (*human rater*).

Hasil penilaian yang lebih detil untuk setiap teks jawaban dapat diamati secara visual pada gambar 1-5. Gambar 2 memperlihatkan grafik penilaian dengan algoritma Cosine. Garis berwarna biru pada gambar merupakan nilai yang diberikan oleh manusia untuk jawaban mahasiswa. Garis berwarna oranye merupakan nilai yang diberikan oleh mesin. Gambar 1 menunjukkan bahwa algoritma Cosine memberikan nilai secara dominan pada rentang 0.5 sampai 1 meskipun penilaian oleh manusia bergerak dari angka nol hingga 2.5. Hal ini menunjukkan bahwa algoritma Cosine tidak cukup mampu memberi nilai yang mengikuti kecenderungan penilaian manusia. Gambar 3 memperlihatkan hasil penilaian menggunakan algoritma Jaccard. Rentang penilaian dominan berada di seputar angka 0.5 dan kurang mampu mengikuti kecenderungan penilaian manusia. Ketika nilai oleh *rater* bergerak ke angka 1, 1.5 bahkan 2, penilaian oleh algoritma Jaccard tidak berubah masih di seputar angka 0.5.

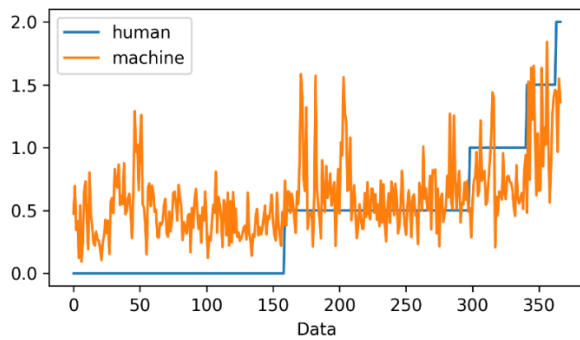


Gambar 2. Penilaian Algoritma Cosine

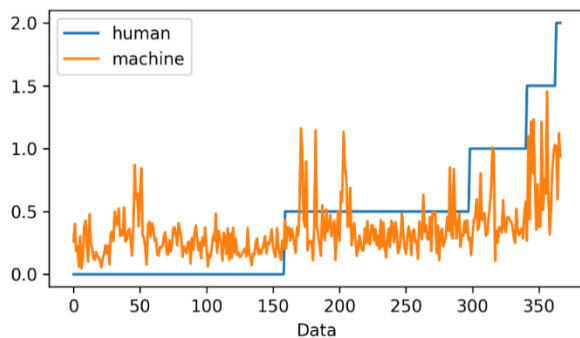


Gambar 3. Penilaian Algoritma Jaccard

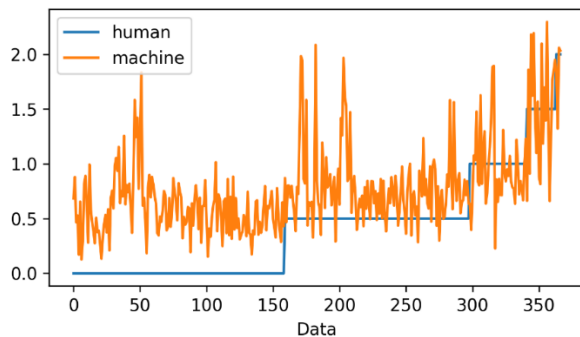
Hasil penilaian dengan algoritma TF-IDF diperlihatkan pada Gambar 4 dan Gambar 5, di mana masing-masing merupakan hasil penilaian TF-IDF yang dilanjutkan dengan perhitungan similaritas Cosine dan Jaccard. Hasil penilaian dengan algoritma TF-IDF berada di seputar 0.5 ketika penilaian manusia berada di angka 0 dan 0.5. Ketika penilaian manusia bertambah ke angka 1, 1.5 dan 2, hasil penilaian mesin memperlihatkan sedikit kenaikan. Kenaikan lebih tinggi lebih terlihat ketika digunakan algoritma Cosine dibanding penggunaan algoritma Jaccard untuk menentukan similaritas antara teks jawaban mahasiswa dengan teks jawaban kunci.



Gambar 4. Penilaian Algorithm TF-IDF Dengan Similaritas Cosine



Gambar 5. Penilaian Algorithm TF-IDF Dengan Similaritas Jaccard



Gambar 6. Penilaian Dengan Algorithm LSA

Gambar 6 memperlihatkan hasil penilaian menggunakan algoritma LSA. Hasil penilaian dominan di seputar angka 0.5 ketika penilaian manusia berada di angka 0 dan 0.5. Hasil penilaian mesin terlihat meningkat ketika penilaian manusia meningkat ke angka 1, 1.5 dan 2. Kesan visula menunjukkan algoritma LSA cenderung mengikuti hasil penilaian manusia, namun hasil penilaian tersebut memiliki deviasi yang besar dari satu titik ke titik yang lain.

Penelitian ini membandingkan penilaian yang dilakukan mesin dan penilaian yang dilakukan manusia. Penilaian mesin dihasilkan melalui perhitungan 5 algoritma, yaitu Cosine, Jaccard, TF-IDF dengan Cosine, TF-IDF dengan Jaccard, dan LSA. Dua algoritma melakukan perhitungan berdasarkan ada atau tidaknya kata pada teks jawaban, yaitu algoritma Cosine dan Jaccard. Dua

algoritma menghitung nilai berdasarkan pada frekuensi kata pada teks jawaban, yaitu algoritma TF-IDF yang dilanjutkan dengan perhitungan similaritas dengan Cosine dan Jaccard. Satu algoritma yaitu LSA mengekstraksi dan merepresentasikan penggunaan kontekstual kata dalam bentuk vektor dengan perhitungan statistik yang diterapkan pada kumpulan teks yang ada.

4. DISKUSI

Hasil penelitian menunjukkan bahwa algoritma TF-IDF untuk membentuk vektor kata dilanjutkan dengan algoritma Jaccard untuk perhitungan similaritas menghasilkan penilaian (skoring) dengan RMSE terendah yaitu 0.3921 atau 15.68%. RMSE dihitung dengan membandingkan nilai mesin dengan nilai aktual, yaitu nilai yang diberikan oleh manusia. Sebaliknya algoritma LSA memiliki RMSE terbesar yaitu 0.5368 atau 21.47%. Hasil ini sepertinya kontras dengan hasil penelitian yang diperoleh peneliti lain seperti [25]–[27] yang menunjukkan bahwa LSA memberikan hasil yang baik dan komparatif dibanding metode lain dalam proses penilaian otomatis.

Selain nilai RMSE, pengamatan visual terhadap hasil penilaian mesin menunjukkan bahwa algoritma LSA cenderung dapat mengikuti pola penilaian manusia. Namun algoritma ini menghasilkan penilaian yang memiliki deviasi sangat tinggi dan barangkali deviasi ini yang menyumbang nilai RMSE yang besar untuk algoritma LSA. Jika data penilaian cukup merata untuk setiap angka penilaian, dapat diduga RMSE algoritma LSA akan memiliki nilai terendah. Pernyataan ini dapat dibuktikan dengan menghitung RMSE untuk setiap segmen data dengan segmentasi didasarkan pada penilaian manusia di angka 0, 0.5, 1, 1.5 dan 2. Nilai RMSE untuk setiap segmen data tersebut adalah 0.651, 0.443, 0.354, 0.496, dan 0.354. Jika diambil rata-ratanya, maka RMSE untuk penilaian dengan algoritma LSA adalah 0.459. Nilai RMSE untuk setiap segmen data jika penilaian dilakukan dengan TF-IDF dilanjutkan perhitungan similaritas Jaccard adalah 0.292, 0.235, 0.598, 0.824, dan 1.100. Jika diambil rata-ratanya, didapat nilai RMSE untuk algoritma TF-IDF+Jaccard adalah sebesar 0.6099 yang lebih besar dibanding RMSE untuk LSA. Dapat ditarik pernyataan bahwa jika data penilaian jawaban merata untuk setiap angka nilai yang mungkin, RMSE untuk LSA berpeluang memiliki nilai terkecil.

5. KESIMPULAN

Hasil penelitian menunjukkan bahwa algoritma TF-IDF untuk mendapatkan vektor kata yang dilanjutkan dengan perhitungan similaritas Jaccard memiliki RMSE terkecil yaitu di angka 0.3921 atau 15.68% dibanding hasil penilaian manusia. Algoritma LSA memperlihatkan kecenderungan

yang paling tinggi untuk mengikuti pola penilaian manusia dan jika data penelitian merata untuk setiap segmen nilai manusia, algoritma LSA cenderung memberikan nilai yang paling mendekati penilaian manusia.

DAFTAR PUSTAKA

- [1] D. Yulianto and A. S. Nugraheni, "Efektivitas Pembelajaran Daring Dalam Pembelajaran Bahasa Indonesia," *Decod. J. Pendidik. Teknol. Inf.*, vol. 1, no. 1, pp. 33–42, Mar. 2021, doi: 10.51454/decode.v1i1.5.
- [2] P. Paudel, "Online education: Benefits, challenges and strategies during and after COVID-19 in higher education," *Int. J. Stud. Educ.*, vol. 3, no. 2, pp. 70–85, 2021.
- [3] N. Sabani, "Pembelajaran Daring Menghadapi Fenomena Pandemi Covid-19 (Systematic Literature Review)," *J. Psychol. Treat.*, vol. 1, no. 1, pp. 11–21, 2021.
- [4] I. Faza Ahmad, "ALTERNATIVE ASSESSMENT IN DISTANCE LEARNING IN EMERGENCIES SPREAD OF CORONAVIRUS DISEASE (COVID-19) IN INDONESIA," *J. Pedagog.*, vol. 07, no. 01, 2020, [Online]. Available: <https://ejournal.unuja.ac.id/index.php/pedagogik>
- [5] R. D. Mahande, F. A. Darmawan, and ..., "Customization and Usability Testing Auto Essay File Grading Lms Based Metacognitive Assessment in Engineering Faculty," ... *Tek. Inform.*, vol. 3, no. 2, 2022, [Online]. Available: <http://jutif.if.unsoed.ac.id/index.php/jurnal/article/view/118%0Ahttp://jutif.if.unsoed.ac.id/index.php/jurnal/article/download/118/65>
- [6] S. Vajjala, "Automated Assessment of Non-Native Learner Essays: Investigating the Role of Linguistic Features," *Int. J. Artif. Intell. Educ.*, vol. 28, no. 1, pp. 79–105, Mar. 2018, doi: 10.1007/s40593-017-0142-3.
- [7] M. S. Fahlevi and L. S. Zanthi, "Analisis Kesulitan Siswa dalam Menyelesaikan Soal Uraian pada Materi Bangun Ruang Sisi Datar," *J. Pembelajaran Mat. Inov.*, vol. 3, no. 4, 2020, doi: 10.22460/jpmi.v3i4.313-322.
- [8] H. Suwandi, H. Harlinda, and S. H. Mansyur, "Implementation of a School Information System Using Rapid Application Development Method," *J. Tek. Inform.*, vol. 3, no. 6, pp. 1501–1512, 2022, doi: 10.20884/1.jutif.2022.3.6.332.
- [9] J. Zeniarta, A. Salam, and I. Achsamu, "Sistem Koreksi Jawaban Esai Otomatis (E-Valuation) dengan Vector Space Model pada Computer Based Test (CBT)," *Semin. Nas. Din. Inform.*, pp. 91–96, 2020.
- [10] A. Rokade, B. Patil, S. Rajani, S. Revandkar, and R. Shedge, "Automated Grading System Using Natural Language Processing," in *Proceedings of the International Conference on Inventive Communication and Computational Technologies, ICICCT 2018*, Sep. 2018, pp. 1123–1127. doi: 10.1109/ICICCT.2018.8473170.
- [11] A. A. Putri Ratna, A. Kaltsum, L. Santiar, H. Khairunissa, I. Ibrahim, and P. D. Purnamasari, "Term Frequency-Inverse Document Frequency Answer Categorization with Support Vector Machine on Automatic Short Essay Grading System with Latent Semantic Analysis for Japanese Language," in *ICECOS 2019 - 3rd International Conference on Electrical Engineering and Computer Science, Proceeding*, 2019, pp. 293–298. doi: 10.1109/ICECOS47637.2019.8984530.
- [12] H. Thamrin, N. A. Verdikha, and A. Triyono, "Text Classification and Similarity Algorithms in Essay Grading," in *2021 4th International Seminar on Research of Information Technology and Intelligent Systems, ISRITI 2021*, 2021, pp. 201–205. doi: 10.1109/ISRITI54043.2021.9702808.
- [13] M. Jamaluddin, N. Yuniarti, A. Rahmani, and J. Hutahaean, "Aplikasi Penilaian Otomatis Ujian Esai Berbahasa Indonesia Menggunakan Algoritma K-Nearest Neighbor (Studi kasus MAN Cimahi)," *Pros. Ind. Res. Work. Natl. Semin.*, vol. 10, no. August 2019, pp. 314–324, 2020, [Online]. Available: <https://jurnal.polban.ac.id/ojs-3.1.2/proceeding/article/view/1404>
- [14] V. Nandini and P. Uma Maheswari, "Automatic assessment of descriptive answers in online examination system using semantic relational features," *J. Supercomput.*, vol. 76, no. 6, pp. 4430–4448, Jun. 2020, doi: 10.1007/s11227-018-2381-y.
- [15] S. M. C. Loureiro, J. Guerreiro, S. Eloy, D. Langaro, and P. Panchapakesan, "Understanding the use of Virtual Reality in Marketing: A text mining-based review," *J. Bus. Res.*, vol. 100, pp. 514–530, Jul. 2019, doi: 10.1016/j.jbusres.2018.10.055.
- [16] F. Sutomo *et al.*, "OPTIMIZATION OF THE K-NEAREST NEIGHBORS ALGORITHM USING THE OPTIMASI ALGORITMA K-NEAREST NEIGHBORS MENGGUNAKAN METODE," vol. 4, no. 1, pp. 125–130, 2023.

- [17] M. Milkhatun, A. A. F. Rizal, N. W. W. Asthiningsih, and A. J. Latipah, "Performance Assessment of University Lecturers: A Data Mining Approach," *Khazanah Inform. J. Ilmu Komput. dan Inform.*, vol. 6, no. 2, 2020.
- [18] R. Rosnelly, D. Hartama, M. Sadikin, C. P. Lubis, M. S. Simanjuntak, and S. Kosasi, "The Similarity of Essay Examination Results using Preprocessing Text Mining with Cosine Similarity and Nazief-Adriani Algorithms," *Turkish J. Comput. Math. Educ.*, vol. 12, no. 3, pp. 1415–1422, Apr. 2021, doi: 10.17762/TURCOMAT.V12I3.938.
- [19] R. Riyanto, "Implementation of the Jaccard Similarity Algorithm on Answer Type Description," *IJIS Int. J. Informatics Inf. Syst.*, vol. 5, no. 2, pp. 76–83, 2022, doi: 10.47738/ijis.v5i2.130.
- [20] H. Zhou, "Research of Text Classification Based on TF-IDF and CNN-LSTM," *J. Phys. Conf. Ser.*, vol. 2171, no. 1, 2022, doi: 10.1088/1742-6596/2171/1/012021.
- [21] N. Hidayat and L. Afuan, "Penilaian Ujian Otomatis untuk Soal Bertipe Essay pada PJJ APTIKOM menggunakan Cosine Similarity," *Semin. Nas. APTIKOM*, pp. 259–271, 2019.
- [22] H. Fan and Y. Qin, "Research on Text Classification Based on Improved TF-IDF Algorithm," vol. 147, no. Ncce, pp. 501–506, 2018, doi: 10.2991/ncce-18.2018.79.
- [23] V. V. Ramalingam, A. Pandian, P. Chetry, and H. Nigam, "Automated Essay Grading using Machine Learning Algorithm," *J. Phys. Conf. Ser.*, vol. 1000, no. 1, 2018, doi: 10.1088/1742-6596/1000/1/012030.
- [24] G. I. Marthasari, N. Hayatin, and M. Yuniarti, "Content Classification based-on Latent Semantic Analysis and Support Vector Machine (LSA-SVM)," *J. Transform.*, vol. 19, no. 2, p. 144, 2022, doi: 10.26623/transformatika.v19i2.2745.
- [25] J. Hoblos, "Experimenting with latent semantic analysis and latent dirichlet allocation on automated essay grading," in *2020 Seventh International Conference on Social Networks Analysis, Management and Security (SNAMS)*, 2020, pp. 1–7.
- [26] A. A. Putri Ratna, H. Khairunissa, A. Kaltsum, I. Ibrahim, and P. D. Purnamasari, "Automatic Essay Grading for Bahasa Indonesia with Support Vector Machine and Latent Semantic Analysis," in *ICECOS 2019 - 3rd International Conference on Electrical Engineering and Computer Science, Proceeding*, 2019, pp. 363–367. doi: 10.1109/ICECOS47637.2019.8984528.
- [27] A. A. P. Ratna, N. A. Wulandari, A. Kaltsum, I. Ibrahim, and P. D. Purnamasari, "Answer categorization method using K-means for Indonesian language automatic short answer grading system based on latent semantic analysis," in *2019 16th International Conference on Quality in Research (QIR): International Symposium on Electrical and Computer Engineering*, 2019, pp. 1–5.