

Rancang Bangun Sistem TOEFL Untuk Analisis Kelemahan Peserta Dengan Penerapan Algoritma K-Means Clustering

Arika Risma Nabella^{1*}, Hani Zulfia Zahro², Yosep Agus Pranoto³

^{1,2,3}Program Studi Teknik Informatika, Institut Teknologi Nasional Malang

*rismarika259@gmail.com

Abstrak

Kemampuan berbahasa Inggris menjadi salah satu keterampilan penting dalam berbagai sektor, terutama dalam konteks pendidikan dan profesional. Tes TOEFL digunakan untuk mengukur kemampuan bahasa Inggris, mencakup aspek Reading, Listening, Speaking, dan Writing. Namun, hasil tes TOEFL sering kali hanya berupa skor total tanpa memberikan informasi rinci terkait kelemahan peserta. Untuk itu, penelitian ini menggunakan algoritma K-Means Clustering karena kesederhanaannya, efisiensinya dalam mengelompokkan data multidimensi, dan kemampuannya menghasilkan pengelompokan yang bermakna secara statistik. Penelitian ini bertujuan untuk mengidentifikasi kelemahan peserta TOEFL di Laboratorium Bahasa ITN Malang. Data yang digunakan adalah skor dari 520 mahasiswa yang mencakup Reading, Structure and Written Expression, dan Listening. Metode ini mengelompokkan peserta ke dalam tiga kluster utama: Reading Boosters (C1), Grammar Builders (C2), dan Listening Improvers (C3). Proses pengelompokan melibatkan pemilihan centroid awal secara acak, perhitungan jarak menggunakan Euclidean Distance, dan iterasi hingga kluster stabil. Hasil analisis menunjukkan 36.54 % peserta masuk ke C1, 41.35% ke C2, dan 22.12% ke C3. Implementasi algoritma ini memberikan manfaat dalam mendukung pengembangan program pembelajaran yang lebih efektif.

Kata kunci : Analisis Kelemahan, K-Means Clustering, TOEFL

Abstract

English proficiency is a critical skill across various sectors, particularly in educational and professional contexts. The TOEFL test assesses English language skills, covering aspects such as Reading, Listening, Speaking, and Writing. However, TOEFL results often only provide a total score without detailed insights into participants' weaknesses. To address this, this study utilizes the K-Means Clustering algorithm for its simplicity, efficiency in processing multidimensional data, and ability to produce statistically meaningful groupings. The research aims to identify the weaknesses of TOEFL participants at the Language Laboratory of ITN Malang. The dataset comprises scores from 520 students across Reading, Structure and Written Expression, and Listening sections. This method classifies participants into three primary clusters: Reading Boosters (C1), Grammar Builders (C2), and Listening Improvers (C3). The clustering process involves randomly selecting initial centroids, calculating distances using the Euclidean Distance formula, and iterating until cluster stability is achieved. The analysis results show that 36.54% of participants fall into C1, 41.35% into C2, and 22.21% into C3. The implementation of this algorithm offers valuable insights for developing more effective learning programs.

Keywords : Weakness Analysis, K-Means Clustering, TOEFL.

1. Pendahuluan

Bahasa Inggris merupakan bahasa paling banyak digunakan dalam berkomunikasi secara internasional. Seiring berkembangnya teknologi dan dunia yang semakin terhubung, kemampuan

berbahasa Inggris menjadi sangat penting untuk berkomunikasi lintas negara dan ini menjadi salah satu life skill yang sangat penting bagi anak muda.^[1] Kemampuan berbahasa Inggris merupakan salah satu keterampilan yang sangat

penting di era globalisasi, khususnya dalam dunia bisnis, teknologi, pendidikan, dan komunikasi internasional.^[2] Dalam konteks profesional, kemampuan berbahasa Inggris sering kali menjadi kunci keberhasilan untuk berinteraksi dengan rekan kerja global, memahami dokumen teknis, serta mengikuti perkembangan teknologi dan informasi. Di bidang akademik, kemampuan ini juga sangat diperlukan, terutama bagi mahasiswa yang ingin melanjutkan studi ke jenjang internasional atau membaca referensi dalam bahasa Inggris.

Laboratorium Bahasa di Institut Teknologi Nasional (ITN) menghadapi tantangan signifikan dalam menganalisis skor TOEFL peserta ujian. Kesulitan ini muncul karena kurangnya metode yang efektif untuk mengevaluasi dan memahami hasil ujian, yang berdampak pada upaya meningkatkan mutu pendidikan. Tanpa analisis yang mendalam, sulit untuk mengidentifikasi kelemahan spesifik peserta, sehingga strategi perbaikan tidak dapat diterapkan secara tepat. Hal ini mengakibatkan ketidakpuasan terhadap kualitas pendidikan yang diberikan dan menghambat perkembangan kemampuan bahasa Inggris mahasiswa.

Sebagai solusi, penerapan algoritma K-Means Clustering dapat menjadi metode yang efektif untuk menganalisis hasil TOEFL. Algoritma K-Means Clustering adalah metode sederhana untuk analisis clustering yang bertujuan untuk

menentukan cara terbaik untuk membagi entitas ke- n menjadi kelompok yang disebut cluster.^[3] Dengan menggunakan K-Means Clustering, data skor dapat dikelompokkan berdasarkan kesamaan pola dan karakteristik peserta. Metode ini memungkinkan pengelompokan peserta berdasarkan area kelemahan mereka, seperti *listening, structure and written expression dan Reading*. Dengan demikian, pihak lab dapat lebih mudah memahami tantangan yang dihadapi oleh masing-masing kelompok dan merancang intervensi yang sesuai untuk meningkatkan kemampuan bahasa Inggris peserta.

Tujuan dari penelitian ini adalah untuk merancang sistem toefl yang dilengkapi dengan analisis kelemahan peserta di Lab Bahasa ITN Malang dengan menggunakan algoritma K-Means Clustering. Penelitian ini juga bertujuan untuk memberikan wawasan yang lebih baik mengenai kelemahan peserta dan merumuskan strategi pembelajaran yang lebih efektif. Dengan hasil analisis ini, diharapkan mutu pendidikan di Lab Bahasa ITN Malang dapat meningkat, sehingga peserta dapat mencapai skor TOEFL yang lebih baik dan mempersiapkan diri dengan lebih baik untuk tantangan akademik dan profesional di masa depan

2. Tinjauan Pustaka

2.1. Penelitian Terkait

Dalam melaksanakan penelitian ini, penulis mengacu pada penelitian sebelumnya yang relevan dengan topik yang sedang diteliti. Berikut adalah beberapa studi yang dijadikan referensi dalam penelitian ini:

- Menurut Adam & Laksana dalam penelitiannya yang berjudul Penerapan Algoritma K-Means Clustering Untuk Mengetahui Kemampuan Akademik Siswa Berbasis WEB dapat disimpulkan bahwa Algoritma K-Means Clustering membantu mengelompokkan siswa berdasarkan nilai raport dengan perhitungan komputasi sederhana. Bobot parameter ditentukan berdasarkan data siswa, sementara skala prioritas ditetapkan melalui wawancara dengan admin sekolah. Algoritma K-Means efektif dalam mengelompokkan data berdasarkan kesamaan fitur dan dapat digunakan untuk menganalisis kualitas nilai dalam kurikulum baru.[4]
- Berdasarkan penelitian yang dilakukan oleh Putra yang berjudul Analisis Klasifikasi Metode K-Means Pada Minat dan Bakat Anak Dimasa Pandemi, dapat diketahui bahwa dengan menganalisis kumpulan nilai ujian yang diperoleh siswa, dapat diketahui tingkat minat siswa terhadap mata pelajaran matematika. Hasil dari proses clustering

menggunakan algoritma K-means menunjukkan bahwa 45 siswa memiliki minat tinggi terhadap matematika, 48 siswa memiliki minat sedang, dan 29 siswa menunjukkan minat rendah pada mata pelajaran tersebut.[5]

- Menurut Nurfathullah dan Purnamasari dalam penelitiannya yang berjudul Implementasi K-Means untuk mengelompokkan provinsi di Indonesia Berdasarkan index Jumlah Pengangguran Terbuka dapat diambil kesimpulan bahwa pengelompokan provinsi menggunakan algoritma K-Means terbagi menjadi 3 cluster berdasarkan kemiripan karakteristik indeks pengangguran terbuka di Indonesia. Tiga klaster yang terbentuk terdiri dari klaster 0 (C0) yang meliputi 11 provinsi, klaster 1 (C1) yang mencakup 12 provinsi, dan klaster 2 (C2) yang melibatkan 11 provinsi. Berdasarkan hasil evaluasi menggunakan koefisien silhouette, diperoleh nilai sebesar 0,54, yang menunjukkan bahwa pengelompokan yang dilakukan sudah memenuhi kriteria untuk struktur klaster yang standar.[6]
- Menurut Amri pada penelitiannya yang berjudul Penerapan Algoritma K-Means Clustering Dalam Mengelompokkan Smartphone Yang Rekomendasi Berdasarkan Spesifikasi dapat ditarik kesimpulan bahwa algoritma K-Means

Clustering dapat memudahkan dalam mengelompokkan data penjualan smartphone di Konter Al-Afgani Cellular dengan baik. Berdasarkan hasil pengelompokan smartphone tersebut, pihak konter dapat menjadi fokus dalam menekan angka peningkatan stok barang yang paling rekomendasi. Untuk smartphone yang masuk ke dalam pengelompokan produk Paling Rekomendasi, pihak konter harus memperhatikan ketersedianya untuk menghindari kekurangan stok. Sedangkan untuk barang yang Kurang Rekomendasi, pihak konter tidak perlu melakukan stok barang yang berlebih, sehingga jumlah stok barang tidak menumpuk.[7]

- Berdasarkan Penelitian Maulana yang digunakan dalam Pengelompokan Data Kerusakan Rumah Akibat Bencana Alam Di Kabupaten Cirebon. Dari penelitiannya dapat disimpulkan data kerusakan rumah di Kabupaten Cirebon dapat dikelompokkan menjadi 2 kelompok menggunakan metode K-Means Clustering, Kerusakan terbanyak berada pada cluster 2 dengan total kerusakan 468 dan kerusakan rumah paling sedikit ada pada cluster 1 dengan total kerusakan 12. [8]

2.2. Landasan Teori

1. Konsep Dasar TOEFL

TOEFL (*Test of English as a Foreign Language*) adalah sebuah layanan tes yang diregistrasi oleh Lembaga Pengujian Pendidikan (*Educational Testing Service, ETS*) sejak tahun 1947, berlokasi di New Jersey, Amerika. Hingga saat ini, TOEFL masih digunakan sebagai syarat akademik; persyaratan masuk perguruan tinggi level atas, serta non-academic; persyaratan untuk beberapa posisi pekerjaan tertentu. Ada tiga jenis tes TOEFL yang diterbitkan oleh ETS, yaitu PBT (*Paper-Based Test*) TOEFL, CBT (*Computer-Based Test*) TOEFL, dan iBT (*Internet-Based Test*) TOEFL. Dengan kemajuan teknologi yang terus meningkat, banyak perguruan tinggi telah beralih menggunakan Computer-Based Test (CBT) dalam pelaksanaannya.^[9]

2. Data Mining

Data mining adalah proses untuk menggali pengetahuan dari dalam basis data. Dalam proses ini, digunakan berbagai teknik seperti statistik, matematika, kecerdasan buatan, dan pembelajaran mesin untuk mengekstrak dan mengenali informasi berharga serta pengetahuan yang berkaitan dari berbagai basis data yang besar. Data mining memiliki beberapa kategori, antara lain deskripsi, estimasi, prediksi, klasifikasi, pengelompokan, dan asosiasi.^[10]

3. Clustering

Clustering adalah teknik analisis data yang mengelompokkan data berdasarkan kesamaan karakteristik, tanpa label yang telah ditentukan sebelumnya. [11] Clustering adalah teknik analisis data yang digunakan untuk mengelompokkan atau membagi sekumpulan data menjadi beberapa kelompok yang memiliki kesamaan atau karakteristik tertentu. Tujuan dari teknik ini adalah agar data yang saling mirip dapat dikelompokkan dalam satu kelompok yang sama, sementara data yang berbeda akan dikelompokkan ke dalam kelompok yang berbeda pula.[12]

4. Algoritma K-Means

K-Means adalah algoritma clustering non-hierarkis yang menggunakan pendekatan pembagian data ke dalam cluster dengan jarak terdekat ke centroid. Metode ini sangat berguna untuk mengenali pola dan kecenderungan dalam data yang luas.[12]

Berikut tahapan Algoritma K-Means :

- 1) Menentukan jumlah cluster dan centroid awal secara acak.
- 2) Menghitung jarak Euclidean dari data ke centroid.
- 3) Mengelompokkan data berdasarkan jarak terdekat dan memperbarui centroid hingga stabil.

Algoritma ini sederhana dan efektif dalam mengelompokkan data multidimensi untuk analisis pola atau kelemahan tertentu.[13]

5. PHP (Hypertext Preprocessor)

PHP adalah bahasa pemrograman server-side yang digunakan untuk membuat aplikasi web dinamis. PHP sering digunakan untuk mengelola basis data, menghasilkan konten web, dan menangani formulir serta sesi pengguna. PHP bersifat server-side, artinya kode dijalankan di server dan hasilnya dikirim ke browser dalam bentuk HTML. Dengan PHP, situs web dapat dibuat dinamis, sehingga pemeliharaan menjadi lebih mudah dan efisien.[14]

6. Database

Database berfungsi sebagai penyimpanan terintegrasi untuk data operasional. Basis data relasional seperti MySQL atau PostgreSQL umum digunakan dalam penelitian ini untuk menyimpan dan mengelola data tes TOEFL.[15]

3. Metode Penelitian

3.1 Metode Pengumpulan Data

Data yang digunakan dalam penelitian ini meliputi skor TOEFL dari mahasiswa ITN. Malang yang diambil dari tiga aspek utama yaitu: Reading, Structure and Written Expression, serta Listening. Dengan jumlah data sebanyak 520 data. Beberapa tahapan untuk mengumpulkan data yang digunakan antara lain:

1. Dokumentasi

Data yang digunakan adalah hasil tes TOEFL mahasiswa yang telah dilaksanakan di Laboratorium Bahasa ITN Malang pada tahun

akademik ganjil 2023/2024. Data ini mencakup skor dari tiga bagian utama tes TOEFL: Reading, Structure and Written Expression, dan Listening.

2. Wawancara

Wawancara dilakukan dengan Kepala Laboratorium Bahasa ITN Malang untuk mendapatkan informasi tambahan mengenai proses pelaksanaan tes TOEFL di Laboratorium, kendala yang sering dihadapi peserta dalam mengerjakan tes TOEFL, perspektif terkait analisis hasil tes dan pentingnya identifikasi kelemahan peserta dan upaya laboratorium dalam meningkatkan kemampuan bahasa Inggris mahasiswa.

3. Analisis Sistem

Data yang telah dikumpulkan dari hasil tes peserta dianalisis menggunakan metode K-Means Clustering, yang merupakan salah satu algoritma unsupervised learning. Algoritma ini digunakan untuk mengelompokkan kelemahan peserta TOEFL berdasarkan hasil skor di setiap bagian tes, yaitu Listening, Structure, dan Reading. Proses analisis ini bertujuan untuk mengidentifikasi pola kelemahan yang serupa di antara peserta, sehingga setiap cluster yang dihasilkan dapat merepresentasikan kelompok peserta dengan karakteristik kelemahan tertentu

3.2 Lokasi Penelitian

Penelitian ini berlokasi di UPT Lab Bahasa Institut Teknologi Nasional Malang

4. Hasil dan Pembahasan

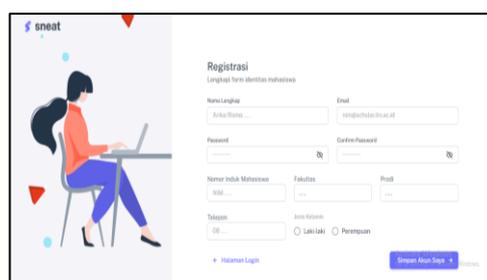
4.1. Hasil Penelitian

Pada tahap ini, dilakukan implementasi metode K-Means Clustering terhadap data hasil tes TOEFL mahasiswa. Hasil utama yang diperoleh meliputi:

1) Fitur Implementasi Sistem

a. Tampilan Registrasi Peserta

Tampilan ini digunakan untuk memasukkan data peserta, seperti nama, email, dan informasi lainnya sebelum memulai tes.



Gambar 1 Tampilan Registrasi

b. Tampilan Peserta Mulai Tes

Tampilan ini menampilkan opsi bagi peserta untuk memulai tes setelah berhasil melakukan registrasi, dengan petunjuk mengenai prosedur tes yang akan dijalani.

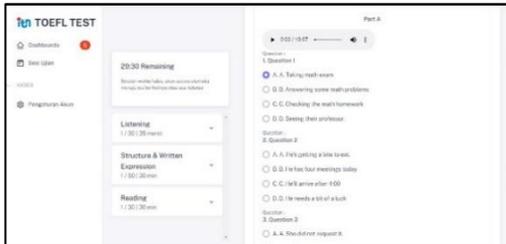


Gambar 2 Tampilan Mulai Tes

c. Tampilan Tes Peserta

Tampilan ini menunjukkan soal-soal tes yang harus dijawab oleh peserta, serta menyediakan

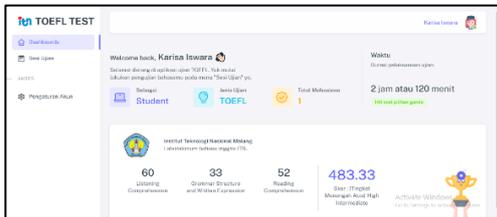
fungsi untuk menjawab soal dan melanjutkan ke soal berikutnya.



Gambar 3 Tampilan Tes

d. Tampilan Hasil Tes

Tampilan ini menyajikan untuk menyajikan hasil tes TOEFL peserta berdasarkan data yang telah diinputkan selama tes berlangsung. Hasil yang ditampilkan mencakup nilai dari tiga komponen utama dalam tes TOEFL, yaitu Listening, Structure and Writing Expressions, dan Reading.



Gambar 4 Hasil Tes

2) Hasil Clustering Data

Hasil Clustering data ini menunjukkan hasil perhitungan yang dilakukan pada sistem dan perhitungan secara manual oleh Excel. Berikut adalah hasil centroid pada iterasi terakhir yaitu iterasi ke- 23.

Tabel 1. Hasil Centroid pada Sistem

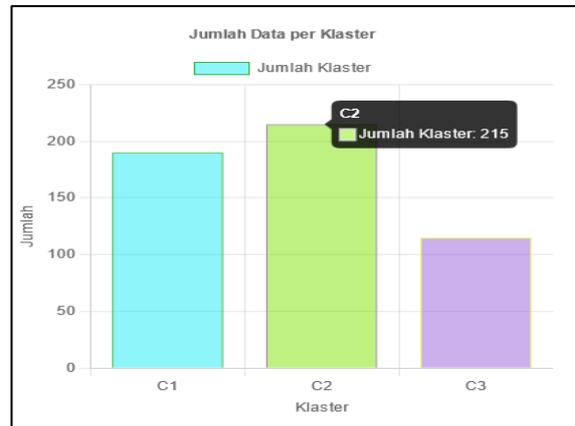
dari 520 data, C1 adalah 36.54 % (190 peserta), C2 adalah 41.35% (215 peserta), dan C3 adalah 22.12% (115 peserta).

Cluster	Sec 1	Sec 2	Sec 3
C1	39.3631	39.77894	39.0631
C2	41.7116	30.75814	32.1720
C3	52.2347	47.12173	42.7391

Tabel 2 Hasil Centroid pada sistem

Cluster	Sec 1	Sec 2	Sec 3
C1	39.3632	39.7789	39.0632
C2	41.7116	30.7581	32.1721
C3	52.2348	47.1217	42.7391

Pada table diatas menunjukkan bahwa perhitungan yang dilakukan sistem valid dengan menunjukkan hasil yang sama pada sistem dan perhitungan manual.



Gambar 5 Hasil Clustering

Gambar diatas merupakan hasil proses clustering, yang dihasilkan sistem menggunakan K-Means dalam 3 Cluster. Hasil Menunjukkan

4.1 Pembahasan

Pada bagian ini, penulis akan memaparkan tahapan-tahapan yang dilakukan dalam proses pengolahan data menggunakan algoritma K-

Means. Perhitungan data dilakukan dengan pendekatan jarak menggunakan metode Euclidean Distance. Adapun hasil yang diperoleh setelah pengujian dan langkah-langkah pengelompokan data menggunakan algoritma K-Means adalah sebagai berikut:

- 1) Menentukan Jumlah Cluster (K). Dalam kasus ini, jumlah cluster yang ditentukan adalah 3.

Keterangan :

K1 : Reading Boosters

K2 : Grammar Builders

K3 : Listening Improvers

- 2) Mentukan Centroid Awal Secara Random
Dalam kasus ini Centroid awal diambil dari data ke-421, ke-419, dan ke-451.

Centroid K1 : (28,36,31)

Centroid K2 : (61,20,22)

Centroid K3 : (68,25,22)

- 3) Hitung jarak setiap titik data ke setiap centroid menggunakan rumus Euclidean Distance.
- 4) Dengan jarak Euclidean antara dua objek, dihitung jarak setiap data b_i ke setiap centroid

$$d(A, K1) = \sqrt{(47 - 28)^2 + (42 - 36)^2 + (38 - 31)^2} \\ = 21.1187$$

$$d(A, K2) = \sqrt{(47 - 61)^2 + (32 - 20)^2 + (38 - 28)^2} \\ = 30.5941$$

$$d(A, K3) = \sqrt{(47 - 68)^2 + (42 - 25)^2 + (38 - 22)^2} \\ = 12.3288$$

- 5) Kelompokkan Data, Setiap titik data dikelompokkan ke cluster berdasarkan jarak terdekat ke centroid. Berdasarkan

perhitungan di atas, misalnya mendapatkan hasil pengelompokan sebagai berikut:

a. Cluster C1:Data {5,8,10,11,12,,.....}

b. Cluster C2:Data {419, 422, 442,.....}

c. Cluster C3:Data {1,2,3,4, 6, 7,.....}

- 6) Update Centroid, Setelah pengelompokan, hitung centroid baru untuk setiap cluster.
- 7) Ulangi langkah 3 hingga 5 sampai tidak ada perubahan pada anggota cluster atau centroid tidak berubah.
- 8) Perhitungan Selesai.

Hasil dari hasil perhitungan diperoleh 23 iterasi yang menunjukkan bahwa algoritma K-Means Clustering efektif dalam mengelompokkan peserta TOEFL ke dalam tiga kluster utama, yaitu Reading Boosters (C1), Grammar Builders (C2), dan Listening Improvers (C3). Distribusi peserta tes ke dalam kluster adalah :

C1 (Reading Boosters): 36.54 % (190 peserta)

C2 (Grammar Builders): 41.35% (215 peserta)

C3 (Listening Improvers): 22.12% (115 peserta)

Dampak pengembangan sistem ini sangat signifikan. Sistem ini meningkatkan efisiensi pengelompokan peserta TOEFL dan membantu mengidentifikasi kelemahan mereka secara akurat. Selain itu, sistem ini juga meningkatkan kualitas pengambilan keputusan untuk meningkatkan pembelajaran dan membantu institusi pendidikan dalam mengembangkan program pembelajaran yang lebih efektif. Hasil perancangan sistem ini menunjukkan bahwa

algoritma K-Means Clustering sangat efektif dalam mengelompokkan data dengan tingkat akurasi yang tinggi. Penggunaan metode Euclidean Distance juga efektif dalam menghitung jarak antara data dan centroid. Proses iteratif yang digunakan dalam algoritma ini membantu meningkatkan akurasi pengelompokan, sehingga sistem ini dapat menjadi alat bantu yang sangat berguna dalam pengembangan program pembelajaran.

5. Kesimpulan

Berdasarkan hasil analisis yang telah dilakukan, dapat disimpulkan bahwa algoritma K-Means Clustering mampu mempermudah proses pengelompokan peserta TOEFL untuk mengidentifikasi kelemahan mereka. Implementasi metode K-Means Clustering pada data hasil tes TOEFL mahasiswa Institut Teknologi Nasional Malang berhasil dilakukan dengan mengelompokkan peserta tes ke dalam tiga kluster utama, yaitu Reading Boosters (C1), Grammar Builders (C2), dan Listening Improvers (C3). Hasil analisis menunjukkan bahwa dari 520 data peserta, 36.54 % masuk ke kluster C1, 41.35% ke kluster C2, dan 22.12% ke kluster C3. Proses pengelompokan dimulai dengan pemilihan centroid awal secara acak, diikuti dengan perhitungan jarak setiap data ke centroid menggunakan metode Euclidean Distance. Proses iteratif ini menghasilkan pembagian

kluster yang bermanfaat untuk memahami kekuatan dan kelemahan peserta TOEFL, sehingga dapat menjadi dasar dalam pengambilan keputusan untuk meningkatkan pembelajaran.

6. Daftar Pustaka

- [1] J. P. Lubis, N. Z. N. Fitri, and S. C. Ridwan, "Pentingnya Menguasai Bahasa Inggris dan Faktor Yang Mempengaruhi Kemampuan Berbahasa Inggris," *Karimah Tauhid*, vol. 3, no. 3, pp. 3599–3605, 2024, doi: 10.30997/karimahtauhid.v3i3.12553.
- [2] R. Roinah, "Penggunaan Bahasa Inggris Pada Masyarakat Ekonomi Asean Di Masa Pandemi Covid-19," *Jurnal Cakrawala Ilmiah*, vol. 1, no. 12, pp. 3625–3634, 2022, doi: 10.53625/jcjournalcakrawalailmiah.v1i12.2445.
- [3] E. A. Saputra and Y. Nataliani, "Analisis Pengelompokan Data Nilai Siswa untuk Menentukan Siswa Berprestasi Menggunakan Metode Clustering K-Means," *Journal of Information Systems and Informatics*, vol. 3, no. 3, pp. 424–439, 2021, doi: 10.51519/journalisi.v3i3.164.
- [4] Adam Rifais and T. G. Laksana, "Penerapan Algoritma K-Means Clustering Untuk Mengetahui Kemampuan Akademik Siswa Berbasis WEB," *Jurnal Ilmiah Multidisiplin*, vol. 1, no. 2, pp. 157–183, Jul. 2024, doi: 10.62282/juilmu.v1i2.157-183.
- [5] P. H. Putra, A. Hasibuan, and E. A. Marpaung, "Analisis Klasifikasi Metode X-Means Pada Minat dan Bakat Anak Dimasa Pandemi," *Jurnal Sains, Teknologi dan Industri*, vol. 19, no. 2, pp. 424–429, 2022.
- [6] M. Nurfathullah and I. Purnamasari, "Implementasi K-Means Untuk Mengelompokkan Provinsi Di Indonesia Berdasarkan Indeks Jumlah Pengangguran Terbuka," 2024.

- [7] M. T. H. Amri Muliawan Nur, Muhammad Saiful, Hariman Bahtiar, "Penerapan Algoritma K-Means Clustering Dalam Mengelompokkan Smartphone Yang Rekomendasi Berdasarkan Spesifikasi," *Infotek : Jurnal Informatika dan Teknologi*, vol. 7, no. 2, pp. 478–488, 2024.
- [8] A. Maulana, R. Danar Dana, and N. D. Nuris, "Implementasi Algoritma K-Means Clustering Dalam Pengelompokan Data Kerusakan Rumah Akibat Bencana Alam Di Kabupaten Cirebon," 2024.
- [9] H. A. Crisdian *et al.*, "Pemanfaatan aplikasi english score dalam peningkatkan nilai TOEFL mahasiswa Tadris Bahasa Indonesia IAIN lhokseumawe," *Abdimas Siliwangi*, vol. 6, no. 1, pp. 83–89, 2023, doi: 10.22460/as.v7i2.22963.
- [10] S. Suhartini and R. Yuliani, "Penerapan Data Mining untuk Mengcluster Data Penduduk Miskin Menggunakan Algoritma K-Means di Dusun Bagik Endep Sukamulia Timur," *Infotek : Jurnal Informatika dan Teknologi*, vol. 4, no. 1, pp. 39–50, 2021, doi: 10.29408/jit.v4i1.2986.
- [11] Ramadhana, Islamiyah, and A. P. A. Masa, "Penerapan Data Mining Menggunakan Metode K-Means Clustering Pada Data Ekspor Batubara," *Adopsi Teknologi dan Sistem Informasi ATASI*, vol. 2, no. 1, pp. 35–42, Jun. 2023, doi: 10.30872/atasi.v2i1.595.
- [12] A. Zaki, I. Irwan, and I. A. Sembe, "Penerapan K-Means Clustering dalam Pengelompokan Data (Studi Kasus Profil Mahasiswa Matematika FMIPA UNM)," *Journal of Mathematics Computations and Statistics*, vol. 5, no. 2, p. 163, 2022, doi: 10.35580/jmathcos.v5i2.38820.
- [13] R. Kesuma Dinata, N. Hasdyna, and N. Azizah, "Analisis K-Means Clustering pada Data Sepeda Motor," 2020.
- [14] A. Yoraeni, A. Adetian, and A. Arfian, "Penerapan Model Water Fall Dalam Membangun Sistem Penjualan Berbasis Web Pada Nefertari Florist Bekasi," *Jurnal Interkom: Jurnal Publikasi Ilmiah Bidang Teknologi Informasi dan Komunikasi*, vol. 14, no. 4, pp. 4–12, 2020, doi: 10.35969/interkom.v14i4.57.
- [15] A. D. Hardiansyah, D. C. Nugrahaeni, P. Dewi, and M. Kom, *Perancangan Basis Data Sistem Informasi Perwira Tugas Belajar (Sipatubel) Pada Kementerian Pertahanan*. 2020