

BAB I

PENDAHULUAN

A. Latar Belakang

Pendidikan di era digital saat ini menuntut pengembangan kompetensi yang relevan dengan kebutuhan abad ke-21, salah satunya adalah kemampuan berpikir komputasional (*Computational Thinking/CT*) dan literasi informatika. Berpikir komputasional merupakan proses pemecahan masalah yang mencakup dekomposisi, pengenalan pola, abstraksi, dan algoritma, yang sangat penting dalam menghadapi tantangan teknologi modern [1]. Di Indonesia, kesadaran akan pentingnya kedua kompetensi ini semakin meningkat, terefleksikan melalui perubahan kurikulum pendidikan nasional. Sebelumnya, informatika seringkali menjadi mata pelajaran pilihan dan konsep CT belum terintegrasi secara luas dalam proses pembelajaran. Namun, dengan diluncurkannya "Kurikulum Merdeka" oleh Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi (Kemendikbudristek), informatika kini ditetapkan sebagai mata pelajaran wajib di jenjang sekolah menengah pertama (SMP) dan menengah atas (SMA) [2]. Sementara itu, berpikir komputasional diintegrasikan ke dalam mata pelajaran inti seperti Matematika, Bahasa Indonesia, dan Ilmu Pengetahuan Alam dan Sosial di jenjang sekolah dasar (SD). Perubahan kurikulum ini menandai sebuah langkah strategis untuk memperkuat fondasi informatika dan CT sejak dini.

Meskipun demikian, implementasi kebijakan ini menghadapi tantangan besar, terutama terkait dengan ketidaksetaraan kualitas pendidikan dan ketersediaan fasilitas teknologi informasi (TI) antara wilayah perkotaan dan daerah terpencil di Indonesia. Kesenjangan ini turut berkontribusi pada hasil PISA (*Programme for International Student Assessment*) yang menunjukkan bahwa siswa Indonesia belum mencapai performa optimal dibandingkan dengan negara-negara lain [2]. Di sisi lain, potensi siswa Indonesia dalam bidang informatika tetap menjanjikan, terbukti dari partisipasi dan pencapaian reguler mereka dalam Olimpiade Informatika Internasional (IOI) [2]. Hal ini mengindikasikan bahwa ada kelompok siswa dengan potensi berpikir komputasional yang tinggi, namun distribusinya mungkin tidak merata.

Dalam upaya menjembatani kesenjangan kompetensi dan mempromosikan berpikir komputasional serta informatika secara lebih luas, inisiatif internasional seperti Tantangan Bebras menjadi relevan. Tantangan Bebras adalah kontes tahunan yang bertujuan memotivasi dan mendorong siswa dari berbagai usia untuk mempelajari konsep-konsep dasar informatika dan mengembangkan pemikiran algoritmik [2]. Soal-soal Bebras dirancang agar dapat dijawab tanpa pengetahuan informatika sebelumnya dan tanpa memerlukan komputer, menjadikannya sangat cocok untuk meningkatkan keterampilan informatika di daerah dengan keterbatasan fasilitas TI. Indonesia sendiri telah menjadi anggota penuh Komunitas Bebras Internasional dan secara rutin menyelenggarakan Tantangan Bebras sejak tahun 2016 [3], dengan jumlah peserta yang terus menunjukkan peningkatan signifikan dari tahun ke tahun [4].

Peningkatan jumlah peserta Tantangan Bebras di Indonesia menunjukkan bahwa inisiatif ini semakin dikenal dan memiliki potensi besar sebagai sarana untuk meningkatkan keterampilan informatika dan kecerdasan komputasional siswa Indonesia. Mengingat pentingnya data dan wawasan yang diperoleh dari evaluasi kegiatan semacam ini, analisis data eksplorasi terhadap hasil Tantangan Bebras Indonesia menjadi krusial. Data hasil Tantangan Bebras, yang mencakup skor dan durasi waktu penyelesaian, dapat memberikan gambaran berharga mengenai tingkat kecerdasan komputasional siswa. Namun, dengan jumlah peserta yang terus meningkat, identifikasi pola dan pengelompokan siswa berdasarkan tingkat kecerdasan komputasional menjadi kompleks.

Untuk mengatasi hal tersebut, penelitian ini akan memanfaatkan data hasil Tantangan Bebras Indonesia dari tahun 2021 hingga 2024 dan menerapkan algoritma K-Means [5] untuk melakukan klusterisasi tingkat kecerdasan komputasional siswa, serta analisis eksploratif menggunakan EDA. Klusterisasi ini bertujuan untuk mengidentifikasi kelompok siswa dengan karakteristik kecerdasan komputasional yang serupa, berdasarkan skor nilai mereka. Hasil dari proses ini diharapkan tidak hanya memberikan gambaran menyeluruh mengenai kompetensi siswa, tetapi juga mampu mengungkap area yang memerlukan perhatian lebih, serta memberikan rekomendasi perbaikan terhadap pelaksanaan Tantangan Bebras Indonesia di masa mendatang. Dengan demikian, penerapan algoritma K-Means dalam penelitian ini diharapkan dapat mendukung optimalisasi strategi pembelajaran dan pengembangan kurikulum informatika yang lebih adaptif dan sesuai dengan kebutuhan beragam siswa di Indonesia.

B. Rumusan Masalah

Berdasarkan latar belakang yang telah diuraikan dan rencana penelitian Anda, berikut adalah rumusan masalah yang telah disesuaikan:

1. Bagaimana hasil *Exploratory Data Analysis* (EDA) dapat memberikan wawasan mengenai distribusi nilai siswa berdasarkan tahun (2021-2024) serta berdasarkan provinsi pada Tantangan Bebras Indonesia?
2. Bagaimana algoritma K-Means dapat digunakan untuk mengklusterisasi siswa peserta Tantangan Bebras Indonesia 2024 berdasarkan skor jawaban soal yang mencerminkan penguasaan keterampilan *Computational Thinking* (CT)?
3. Bagaimana menganalisis dan mendeskripsikan karakteristik setiap kluster siswa yang terbentuk, dengan mengidentifikasi CT *Skill* apa saja yang paling menonjol atau terbaik di dalam masing-masing kluster tersebut, serta meninjau sebaran skor Tantangan Bebras Indonesia 2024?

C. Batasan Masalah

Untuk memastikan penelitian ini tetap fokus dan dapat diselesaikan secara efektif, berikut adalah batasan-batasan masalah yang ditetapkan:

1. Fokus klasterisasi data tahun 2024 adalah pada tingkat kecerdasan komputasional siswa yang dianalisis menggunakan data hasil Tantangan Bebras Indonesia tahun 2024.
2. *Exploratory Data Analysis* (EDA) Tahun 2021-2024: *Exploratory Data Analysis*(EDA) akan mencakup data hasil Tantangan Bebras Indonesia dari tahun 2021 hingga 2024.
3. Klasterisasi menggunakan algoritma K-Means akan didasarkan pada variabel skor nilai pada setiap anak di Tantangan Bebras.
4. Data yang digunakan adalah data resmi hasil Tantangan Bebras Indonesia yang diperoleh dari pihak penyelenggara.
5. Penelitian ini tidak akan mengkaji faktor-faktor non-kognitif (misalnya, motivasi, latar belakang sosioekonomi) yang mungkin mempengaruhi tingkat kecerdasan komputasional siswa.

D. Tujuan Penelitian

Berdasarkan rumusan masalah dan batasan masalah yang telah ditetapkan, tujuan dari penelitian ini adalah sebagai berikut:

1. Melakukan *Exploratory Data Analysis* (EDA) terhadap hasil Tantangan Bebras Indonesia dari tahun 2021 hingga 2024 untuk mendapatkan wawasan mengenai tren dan distribusi nilai siswa secara keseluruhan, serta menganalisis distribusi nilai siswa berdasarkan provinsi.
2. Mengklasterisasi siswa peserta Tantangan Bebras Indonesia tahun 2024 ke dalam beberapa kelompok hasil klaster berdasarkan skor penguasaan keterampilan CT pada setiap soal menggunakan algoritma K-Means.
3. Menganalisis dan mendeskripsikan karakteristik dari setiap klaster yang terbentuk, sehingga dapat mengidentifikasi CT *skill* apa saja yang paling kuat atau menonjol di setiap klaster, serta memberikan gambaran mengenai profil skor di setiap kelompok siswa.

E. Metode Penelitian

Untuk mencapai tujuan penelitian yang telah ditetapkan, penelitian ini akan menggunakan pendekatan kuantitatif dengan metode klasterisasi dan analisis deskriptif. Berikut adalah tahapan-tahapan metodologi penelitian yang akan dilakukan:

1. Perencanaan Analisis Data

Tahap ini dilakukan untuk merumuskan tujuan analisis data dalam penelitian mengenai klasterisasi tingkat kecerdasan komputasional siswa menggunakan algoritma K-Means. Peneliti mengidentifikasi variabel atau fitur yang relevan untuk menggambarkan kemampuan *Computational Thinking* (CT) siswa, serta menentukan metode analisis yang tepat, yaitu kombinasi *Exploratory Data Analysis* (EDA) dan *Clustering* K-Means.

Perencanaan ini memastikan proses analisis dilakukan secara sistematis, mulai dari eksplorasi awal, pengolahan data, pembentukan klaster, hingga interpretasi hasil. Dengan perencanaan yang matang, analisis diharapkan dapat mengungkap pola-pola penguasaan CT siswa dan memberikan informasi yang bermanfaat bagi

guru, sekolah, maupun pembuat kebijakan pendidikan dalam merancang strategi pembelajaran yang lebih efektif.

2. Pengumpulan Data

Tahap pengumpulan data merupakan langkah krusial dalam penelitian ini. Data yang akan digunakan adalah data sekunder hasil Tantangan Bebras Indonesia dari tahun 2021 hingga 2024. Data ini diperoleh secara langsung dari dosen pembimbing 1 penulis yang memiliki akses resmi ke hasil Tantangan Bebras Indonesia. Kelengkapan data sangat diperhatikan, mencakup identitas siswa skor rinci untuk setiap soal yang dikerjakan, dan kategori CT *Skill* spesifik di tahun 2024 yang diukur oleh masing-masing soal (misalnya, *Decomposition*, *Pattern Recognition*, *Abstraction*, dan *Algorithmic Thinking*). Proses ini memastikan bahwa data yang digunakan valid dan relevan, menjadi fondasi kuat untuk analisis selanjutnya dalam penelitian ini.

3. Pembuatan Dataset Menggunakan Microsoft SQL Server

Tahap pembuatan dataset merupakan langkah fundamental yang memastikan data tersedia dalam format yang terstruktur dan siap untuk analisis lebih lanjut. Dalam penelitian ini, data hasil Tantangan Bebras Indonesia dari tahun 2021 hingga 2024 yang telah dikumpulkan dari sumber resmi akan diintegrasikan dan dikelola menggunakan Microsoft SQL Server. Penggunaan sistem manajemen basis data ini sangat strategis karena memungkinkan penyimpanan data yang efisien, terorganisir, dan dapat diakses dengan mudah. Proses ini akan melibatkan import data mentah ke dalam tabel-tabel yang relevan di SQL Server, memastikan setiap kolom data seperti identitas siswa (ID Peserta, jenjang pendidikan, kota, provinsi), skor pengerjaan setiap soal, dan kategori CT *Skill* yang diukur oleh soal tersebut, tersimpan secara akurat.

Dengan data yang tersimpan dalam database relasional, peneliti dapat melakukan berbagai operasi seperti penggabungan data dari tahun-tahun yang berbeda, pemfilteran, dan agregasi data awal menggunakan kueri SQL. Hal ini tidak hanya memfasilitasi proses pra-pemrosesan data selanjutnya, tetapi juga menjamin integritas dan konsistensi data sepanjang siklus penelitian, menjadi pondasi kokoh bagi analisis klusterisasi dan deskriptif yang akan dilakukan.

Selanjutnya, untuk memberikan pemahaman yang lebih jelas mengenai struktur data, penting untuk menjelaskan kategori Tantangan Bebras yang menjadi dasar pengelompokan data. Setiap jenjang peserta memiliki kategori tersendiri yang menentukan dimensi keterampilan berpikir komputasional (CT *Skill*) yang diukur. Tabel berikut merangkum kategori Tantangan Bebras berdasarkan jenjang pendidikan beserta dimensi CT *Skill* yang digunakan:

Kategori	Jenjang Pendidikan	Dimensi CT <i>Skill</i>
SiKecil	SD kelas 1-3	<i>Abstraction, Pattern Recognition, Evaluation, Logic</i>
Siaga	SD kelas 4-6	<i>Evaluation, Abstraction, Logic, Pattern Recognition, Algorithmic Thinking, Data Analysis</i>

Penggalang	SMP kelas 7-9	<i>Evaluation, Abstraction, Decomposition, Pattern Recognition, Algorithmic Thinking, Generalization</i>
Penegak	SMA kelas 10-12	<i>Evaluation, Abstraction, Logic, Pattern Recognition, Algorithmic Thinking, Data Analysis</i>

Tabel 1.1. Kategori Tantangan Bebras

4. *Pre-Processing Data*

Tahap ini penting untuk memastikan kualitas data sebelum analisis. Langkah-langkah yang akan dilakukan meliputi:

- Pembersihan Data: Menangani data yang hilang (*missing values*), duplikasi data, dan anomali (*outlier*) yang tidak wajar.
- Transformasi Data: Mengubah format data jika diperlukan agar sesuai untuk proses klusterisasi. Ini mungkin melibatkan normalisasi atau standarisasi skor.
- Ekstraksi Fitur CT Skill: Menghitung atau mengagregasi skor siswa untuk setiap kategori CT skill dari skor per soal yang tersedia. Misalnya, rata-rata skor pada semua soal yang mengukur 'dekomposisi' akan menjadi representasi penguasaan CT skill dekomposisi siswa.

5. *Clustering Menggunakan Algoritma K-Means*

Proses klusterisasi akan berfokus pada data Tantangan Bebras Indonesia tahun 2024 menggunakan algoritma k-means. Algoritma k-means adalah metode klustering dengan mengelompokkan data ke dalam k klaster yang telah ditentukan, memanfaatkan jarak Euclidean sebagai dasar pengukurannya [6]. k-means clustering memiliki keunggulan dalam skala yang besar dan efisiensi komputasional.

6. *Exploratory Data Analysis (EDA)*

Pada tahap ini, seluruh data Tantangan Bebras Indonesia dari tahun 2021 hingga 2024 akan dieksplorasi secara menyeluruh untuk mengungkap mengenai karakteristik dan pola yang terkandung di dalamnya. Analisis dimulai dengan statistik deskriptif, mencakup rata-rata dan rentang skor untuk memahami tendensi sentral serta sebaran data. Selanjutnya memvisualisasikan sebaran dan distribusi, seperti histogram untuk mengetahui pola distribusi skor dan box plot untuk membandingkan sebaran antarkelompok. Analisis tren dilakukan dengan membandingkan rata-rata skor tahunan dari 2021–2024 untuk mengidentifikasi pola peningkatan, penurunan, atau kestabilan dalam penguasaan keterampilan CT.

Terakhir, analisis geografis akan memvisualisasikan distribusi nilai rata-rata siswa berdasarkan provinsi, yang bertujuan untuk mengidentifikasi area-area dengan performa menonjol atau wilayah yang mungkin memerlukan perhatian lebih lanjut dalam pengembangan kecerdasan komputasional. Metodologi ini sejalan dengan rangka kerja EDA dalam konteks *learning analytics*, yang menekankan pentingnya analisis pendahuluan melalui visualisasi iteratif dan

statistik deskriptif untuk mengungkap pola kompleks dalam data pendidikan [8]. Pendekatan ini juga telah diterapkan dalam studi yang memadukan EDA dengan visualisasi untuk memprediksi dan menampilkan performa kognitif siswa melalui LMS, sehingga mampu menghasilkan wawasan yang komprehensif dari data yang dikumpulkan selama beberapa tahun [9].

7. Analisis dan Interpretasi Klaster

Tahapan ini adalah bagian paling krusial dari penelitian karena berfungsi untuk secara langsung menjawab tujuan yang telah ditetapkan. Setelah klaster terbentuk melalui algoritma K-Means, langkah selanjutnya adalah menganalisis karakteristik setiap klaster secara mendalam. Analisis ini meliputi pemeriksaan rata-rata skor total yang diperoleh siswa di setiap klaster dan menghitung rata-rata skor untuk setiap kategori *CT skill* (*Abstraction, Pattern Recognition, Evaluation, Logic, Algorithmic Thinking, Decomposition, Data Analysis, dan Generalization*) dalam klaster tersebut. Dari profil klaster yang terperinci ini, peneliti akan mengidentifikasi *CT skill* apa saja yang paling menonjol atau menunjukkan penguasaan terbaik di setiap kelompok siswa. Identifikasi ini sangat penting karena akan menjadi dasar bagi penamaan klaster yang bermakna dan deskriptif. Terakhir, tahap ini akan menghasilkan implikasi dari temuan tersebut, yaitu menarik kesimpulan yang komprehensif mengenai karakteristik klaster dan apa artinya bagi pemahaman kita tentang tingkat kecerdasan komputasional siswa di Indonesia.

Analisis klaster seperti yang dijelaskan pada [10] memungkinkan segmentasi subjek berdasarkan kesamaan karakteristik tanpa adanya asumsi awal terhadap struktur data. Dengan pendekatan ini, profil dari masing-masing klaster dapat diidentifikasi untuk mengetahui keterampilan *CT* yang paling dominan pada kelompok tertentu. Penelitian lain [11] juga menunjukkan bahwa pendekatan ini dapat mengidentifikasi tingkat penguasaan siswa terhadap masing-masing aspek *CT* secara spesifik. Dari profil ini, klaster kemudian dapat diberi nama yang deskriptif dan bermakna, seperti “Klaster Algoritma Dominan” atau “Klaster Pola Tinggi”. Tahapan ini diakhiri dengan penarikan implikasi berdasarkan karakteristik klaster, yang akan berkontribusi terhadap pemahaman lebih mendalam mengenai distribusi keterampilan *CT* siswa di Indonesia, serta rekomendasi kebijakan atau intervensi yang relevan.

F. Sistematika Penulisan

Secara sistematis isi dari laporan ini disusun sebagai berikut:

BAB I PENDAHULUAN

Bab ini menguraikan latar belakang, permasalahan, batasan, dan tujuan penelitian yang berfokus pada klasterisasi kecerdasan komputasional siswa Tantangan Bebras.

BAB II TINJAUAN PUSTAKA

Bab ini menyajikan tinjauan terhadap penelitian-penelitian terdahulu yang relevan sebagai dasar dan perbandingan dalam pelaksanaan penelitian ini.

BAB III LANDASAN TEORI

Bab ini menjelaskan konsep-konsep dasar terkait kecerdasan komputasional, Tantangan Bebras, klasterisasi, algoritma K-Means, Machine Learning, Python, dan Microsoft SQL Server yang menjadi landasan teori penelitian.

BAB IV DATASET DAN PENGEMBANGAN MODEL

Bab ini merinci proses persiapan dataset, termasuk analisis data saat mengkueri database dan pra-pemrosesan data, serta menjelaskan tahapan pengembangan model klasterisasi.

BAB V HASIL DAN PEMBAHASAN

Bab ini akan menyajikan secara rinci hasil dari implementasi analisis data dan evaluasi model, serta membahas interpretasi dari klaster yang terbentuk dan wawasan yang diperoleh dari EDA.

BAB VI PENUTUP

Bab ini berisi kesimpulan serta saran mengenai penelitian yang telah dilakukan oleh penulis.

BAB II

TINJAUAN PUSTAKA

Untuk menambah referensi dalam penelitian ini, penulis mencari berbagai penelitian terdahulu yang memiliki keterkaitan dengan topik yang penulis ambil. Studi-studi sebelumnya akan dijadikan sebagai dasar acuan serta bahan perbandingan dalam pelaksanaan dan penyelesaian penelitian ini.

Zulrahmadi melakukan penelitian yang bertujuan untuk mengetahui informasi terkait wilayah yang potensial sebagai penjualan sepeda motor Honda di Kabupaten Indragiri Hilir sehingga dapat digunakan oleh PT [12]. Capella Dinamik Nusantara untuk pengambilan keputusan. Penelitian tersebut dilakukan dengan clustering data menggunakan metode K-Means. Data dikumpulkan dari penjualan motor Honda pada PT. Capella Dinamik Nusantara dari tahun 2017 sampai dengan 2019 dengan menjadi 3 cluster, yaitu cluster C0 untuk wilayah kurang berpotensi, cluster C1 untuk wilayah cukup berpotensi, dan cluster C2 untuk wilayah sangat berpotensi. Hasil penelitian menghasilkan tiga cluster yaitu Cluster 0 (C0) yang dideskripsikan sebagai kurang berpotensi sebanyak lima wilayah desa, Cluster 1 (C1) yang dideskripsikan cukup berpotensi sebanyak tiga wilayah desa, Cluster 2 (C2) yang dideskripsikan sangat berpotensi sebanyak dua wilayah desa untuk penjualan motor. Iterasi terjadi sebanyak dua kali. Hasil cluster tersebut diharapkan bisa digunakan untuk mengambil keputusan terbaik sehingga keuntungan dari PT. Capella Dinamik Nusantara dapat meningkat dan dapat melakukan promosi yang tepat sasaran.

Parlina dkk melakukan penelitian dengan memanfaatkan algoritma K-Means untuk menentukan pegawai yang layak mengikuti *assessment center* [13]. Penelitian ini bertujuan untuk mengelompokkan data pegawai yang terdaftar dalam Program SDP (Staff Development Program) di PT. Bank Syariah Mandiri tahun 2017 menggunakan algoritma K-Means. Tujuan akhir dari pengelompokan ini adalah untuk mengetahui pegawai yang layak untuk lolos dalam program SPD sehingga dapat melakukan registrasi Assessment Center berdasarkan tiga kriteria utama, yaitu Kedisiplinan Pegawai, Target Kerja Pegawai, dan Kepatuhan terhadap Program SDP. Hasil dari penelitian ini menunjukkan bahwa tiga cluster berhasil terbentuk, masing-masing menunjukkan kelompok pegawai berdasarkan tingkat kecocokan mereka terhadap kriteria Program SDP. Pegawai dalam Cluster 1 dianggap Lolos dan siap mengikuti Assessment Center. Pegawai dalam Cluster 2 dikategorikan sebagai Hampir Lolos, tetapi memerlukan perbaikan administratif seperti kedisiplinan dari Juni hingga Oktober agar dapat mengikuti Assessment Center lanjutan. Pegawai dalam Cluster 3 dikategorikan sebagai Tidak Lolos, dan perlu memperbaiki kedisiplinan selama satu tahun penuh agar dapat dipertimbangkan untuk mengikuti program Assessment Center ini. Iterasi terjadi sebanyak tiga kali.

Selanjutnya Elin dkk melakukan penelitian dengan tujuan untuk memberikan kemudahan bagi PT XYZ, sebuah perusahaan distribusi minuman ringan di Lampung, dalam mengelola data transaksi penjualannya untuk mengetahui tingkat penjualan produk pada setiap cabangnya [7]. Penelitian ini juga bertujuan untuk membantu perusahaan meningkatkan akurasi peramalan penjualan serta merencanakan strategi bisnis yang sesuai, yang pada akhirnya akan memudahkan analisis dan pengambilan keputusan. Dengan demikian, perusahaan dapat mengarahkan upaya pemasaran dan sumber daya dengan lebih efektif untuk

dapat meningkatkan keuntungan dan pertumbuhan bisnis. Metode yang digunakan dalam penelitian ini adalah teknik data mining dengan algoritma K-means clustering menggunakan bahasa pemrograman Python. Data penjualan akan dikelompokkan ke dalam 3 cluster yaitu penjualan tertinggi, penjualan sedang, dan penjualan terendah. Penelitian ini menganalisis data penjualan minuman dari PT XYZ selama periode tahun 2022 yang terdiri dari 14 atribut dengan total 712 record yang tersebar di 12 wilayah Sumatera bagian selatan. Proses klasterisasi dilakukan dengan tahapan penentuan jumlah klaster, inisialisasi nilai centroid, perhitungan jarak terdekat menggunakan Euclidean Distance, pengelompokkan objek berdasarkan jarak terdekat dengan centroid, dan iterasi hingga nilai optimal tercapai. Untuk memvalidasi hasil clustering, penelitian ini menggunakan metode Silhouette Score untuk mengetahui jumlah klaster yang optimal. Hasil penelitian menunjukkan bahwa berdasarkan klasterisasi data penjualan berdasarkan region, penjualan dengan nilai tertinggi berada pada region 3 yang ditandai dengan warna hitam (tinggi), penjualan dengan nilai sedang berada pada region 11 yang ditandai dengan warna merah (sedang), dan penjualan dengan nilai terendah berada pada region 4 yang ditandai dengan warna hijau (rendah). Sementara itu, hasil klasterisasi untuk data penjualan berdasarkan outlet menunjukkan penjualan dengan klaster tinggi ditandai dengan warna hijau, klaster sedang ditandai dengan warna hitam, dan klaster rendah ditandai dengan warna merah. Pengujian menggunakan metode Silhouette Score menghasilkan skor 0,78 untuk klaster data penjualan berdasarkan region dan 0,58 untuk klaster data penjualan berdasarkan outlet. Kedua skor tersebut menunjukkan bahwa jumlah klaster yang dihasilkan masuk ke dalam kategori baik karena tidak mendekati -1, sehingga hasil klasterisasi dapat dijadikan rekomendasi bagi PT XYZ dalam menentukan strategi penjualan untuk meningkatkan keuntungan perusahaan.

Selanjutnya Eriskiannisa dkk melakukan penelitian dengan tujuan utama dari penelitian ini adalah menerapkan segmentasi pelanggan pada data transaksi *online retail* menggunakan algoritma K-Means clustering untuk mengidentifikasi karakteristik pelanggan dan menentukan strategi pemasaran yang lebih tepat sasaran [32]. Penelitian ini juga berfokus pada penentuan jumlah klaster optimal dengan metode Elbow, sehingga perusahaan dapat lebih memahami perilaku pelanggan dan meningkatkan efektivitas strategi pemasarannya. Metode penelitian yang digunakan meliputi studi literatur terkait segmentasi pelanggan dan K-Means pada rentang tahun 2018–2022, pengumpulan data transaksi *online retail* sebanyak 54.910 baris data dengan 8 kolom atribut (Invoice No, Stock Code, Description, Quantity, Invoice Date, Unit Price, Customer ID, Country), *exploratory data analysis* dan *data pre-processing* (penanganan *missing value*, *outlier*, transformasi data, penambahan kolom penjualan, serta normalisasi dan standarisasi data), implementasi K-Means dengan bahasa pemrograman Python melalui Google Colab, serta penentuan jumlah klaster optimal menggunakan metode Elbow. Hasil analisis menunjukkan bahwa jumlah klaster optimal adalah empat, dengan karakteristik sebagai berikut: 1) Klaster pertama terdiri dari pelanggan dengan jumlah pembelian rendah dan harga satuan tinggi, direkomendasikan diberikan diskon atau penawaran eksklusif, 2) Klaster kedua memiliki pembelian menengah dan harga cukup rendah, direkomendasikan diberi promo pembelian dalam jumlah tertentu, 3) Klaster ketiga memiliki pembelian tinggi dan harga rendah, direkomendasikan mendapatkan program hadiah atau undian, 4) Klaster keempat memiliki pembelian rendah dan harga tinggi, direkomendasikan diberi diskon berkala atau penawaran khusus. Temuan ini menegaskan bahwa kuantitas pembelian dan harga satuan berperan penting dalam memengaruhi perilaku pelanggan, serta menunjukkan bahwa K-Means dengan metode Elbow dapat memberikan hasil segmentasi yang efektif untuk mendukung strategi pemasaran.

Kemudian Nisa dkk melakukan penelitian yang bertujuan untuk melakukan analisis karakteristik pelanggan sebagai dasar penetapan segmentasi pelanggan dan *customer profiling* pada produk *digital service* Indihome milik PT. Telekomunikasi Indonesia [33]. Tujuan utama penelitian adalah untuk mengetahui perilaku pelanggan dan menerapkan strategi pemasaran yang tepat sehingga dapat mendatangkan keuntungan optimal bagi perusahaan, mengingat sebelumnya belum ada informasi yang memadai tentang karakteristik pelanggan. Dalam aspek metodologi, penelitian ini menggunakan kerangka kerja CRISP-DM (*Cross Standard Industries for Data Mining*) sebagai panduan sistematis dalam proses *data mining*. Algoritma K-Means dipilih sebagai metode clustering utama karena kemampuannya dalam mempartisi data ke dalam kelompok-kelompok berdasarkan kemiripan karakteristik. Untuk menentukan jumlah cluster yang optimal, peneliti menggunakan metode Elbow yang menganalisis nilai Sum Square. Data yang digunakan dalam penelitian ini terdiri dari 1.392 record yang merupakan 30% dari total 4.640 record data pelanggan yang tersedia. Data tersebut terdiri dari tiga jenis yaitu data pelanggan Indihome, data pelanggan *add-on*, dan data *churn* pelanggan. Atribut yang digunakan untuk proses clustering meliputi lama berlangganan, jumlah layanan yang diambil, dan total tagihan pelanggan. Melalui metode Elbow, ditentukan bahwa jumlah cluster optimal adalah $K = 3$, sehingga seluruh data pelanggan dikelompokkan ke dalam tiga segmen yang berbeda berdasarkan karakteristik tersebut. Hasil clustering pada simulasi terbaik menunjukkan bahwa Cluster 0 terdiri dari 396 pelanggan dengan kategori pelanggan yang memberikan keuntungan terbesar bagi perusahaan, memiliki waktu berlangganan 7-11 bulan, menggunakan 1-4 layanan, dan total tagihan berkisar 209-286 ribu rupiah. Cluster 1 berisi 286 pelanggan yang tanpa disadari memiliki potensi besar dalam memberikan keuntungan bagi perusahaan, dengan waktu berlangganan 7-11 bulan, menggunakan 1 layanan, dan total tagihan 275 ribu hingga 1,144 ribu rupiah. Sementara itu, Cluster 2 merupakan kelompok terkecil dengan 14 pelanggan yang dikategorikan sebagai pelanggan yang memberikan keuntungan lebih sedikit daripada biaya untuk memberikan pelayanan, memiliki waktu berlangganan 1-6 bulan, menggunakan 1-3 layanan, dan total tagihan 286-907,5 ribu rupiah.

Kemudian Dacwanda dan Nataliani melakukan penelitian dengan tujuan untuk menganalisis prestasi akademik siswa dengan mengelompokkan tingkat pencapaian mereka menggunakan metode k-means clustering [37]. Tujuan utama dari pengelompokan ini adalah membantu pihak sekolah memahami distribusi kemampuan siswa sehingga strategi pembelajaran dapat disesuaikan untuk meningkatkan kualitas pendidikan. Sampel penelitian melibatkan 266 siswa kelas IXA hingga IXH di SMP Negeri 2 Salatiga dengan data nilai dari 12 mata pelajaran sesuai kurikulum 2013, yang masing-masing memiliki dua komponen penilaian, yaitu pengetahuan dan keterampilan. Total terdapat 24 atribut yang dianalisis untuk menghasilkan pengelompokan yang lebih akurat dan representatif terhadap kemampuan akademik siswa selama satu semester pembelajaran. Proses analisis menggunakan perangkat lunak RapidMiner untuk mengolah data dan menjalankan algoritma k-means sekaligus menghitung validitas cluster menggunakan indeks Davies-Bouldin, yang menunjukkan bahwa tiga cluster merupakan jumlah yang paling optimal. Hasil analisis mengelompokkan siswa ke dalam tiga kategori, yaitu cluster pintar sebanyak 71 siswa, cluster sedang sebanyak 92 siswa, dan cluster cukup sebanyak 103 siswa. Temuan ini memberikan informasi yang sangat berguna bagi guru dan pihak sekolah dalam merancang metode pembelajaran yang lebih tepat sasaran, termasuk memberikan perhatian khusus bagi siswa yang membutuhkan pendampingan tambahan. Selain itu, hasil penelitian ini juga dapat menjadi dasar bagi pengembangan analisis lanjutan dengan menambahkan aspek non-akademik atau membandingkan metode k-means dengan algoritma clustering lainnya untuk evaluasi yang lebih komprehensif.

Lalu Saputra dan Nataliani melakukan penelitian yang bertujuan untuk mengelompokkan prestasi siswa berdasarkan nilai mata pelajaran dan absensi menggunakan metode k-means clustering [38]. Tujuannya adalah untuk membantu sekolah mengidentifikasi

siswa dengan kategori prestasi rendah, cukup, dan tinggi, sekaligus menentukan siswa top rank yang berpotensi menjadi teladan akademik. Objek penelitian ini adalah siswa kelas X–XII jurusan Bahasa SMAN 1 Tenganan pada tahun ajaran 2014–2017, dengan data yang mencakup rata-rata nilai rapor dari semester satu hingga lima, serta data kehadiran siswa. Sampel yang digunakan berjumlah 25 siswa, yang mewakili data dari angkatan tersebut, sehingga analisis dapat memberikan gambaran umum mengenai distribusi prestasi akademik dan tingkat kehadiran siswa. Proses analisis dilakukan dengan menggunakan perangkat lunak RStudio, yang diintegrasikan dengan algoritma k-means untuk mengelompokkan data ke dalam tiga cluster optimal berdasarkan metode Elbow dan Silhouette. Hasil analisis menunjukkan bahwa terdapat cluster rendah dengan 12 siswa, cluster cukup dengan 5 siswa, dan cluster tinggi dengan 8 siswa. Untuk mengidentifikasi siswa berprestasi terbaik dari cluster tinggi, metode *Simple Additive Weighting* (SAW) digunakan sehingga diperoleh siswa dengan peringkat teratas yang layak mendapatkan penghargaan atau perhatian khusus. Hasil penelitian ini diharapkan dapat digunakan pihak sekolah untuk mengevaluasi kualitas pembelajaran, memberikan motivasi bagi siswa, serta mengembangkan strategi pengajaran yang lebih tepat sasaran guna meningkatkan prestasi akademik di masa mendatang.

Dari berbagai penelitian terdahulu, penelitian yang akan dilakukan oleh penulis berfokus pada analisis data hasil Tantangan Bebras Indonesia tahun 2021 hingga 2024 dengan dua pendekatan utama, yaitu Exploratory Data Analysis (EDA) dan clustering menggunakan algoritma K-Means. Data yang akan dianalisis mencakup 156.753 entri peserta, yang selanjutnya akan diolah untuk menemukan pola dan informasi penting. Melalui EDA, penulis berencana memetakan sebaran nilai peserta berdasarkan provinsi serta menganalisis dimensi berpikir komputasional yang diukur dalam tantangan tersebut, seperti *Abstraction*, *Algorithmic Thinking*, *Decomposition*, *Pattern Recognition*, *Evaluation*, *Generalization*, *Logic*, dan *Data Analysis*. Pada tahap selanjutnya, metode K-Means Clustering akan diterapkan untuk mengelompokkan provinsi berdasarkan pola nilai, dengan pemilihan jumlah kluster optimal menggunakan metode Elbow dan Silhouette Score. Tujuan dari penelitian ini adalah untuk mengidentifikasi pola tersembunyi dalam data, memahami variasi kemampuan berpikir komputasional siswa secara geografis, serta memberikan rekomendasi berbasis data bagi pihak penyelenggara, pendidik, maupun pembuat kebijakan. Dengan pendekatan ini, diharapkan dapat teridentifikasi daerah-daerah yang memerlukan intervensi lebih lanjut maupun daerah dengan potensi yang dapat dijadikan acuan keberhasilan.

Tabel 2.1. Perbandingan Penelitian

Pembanding	Tujuan	Metode Algoritma	Jumlah Sampel	Jumlah Cluster	Objek	Tools
[12]	Mengetahui informasi terkait wilayah yang potensial sebagai	Clustering K-Means	1.213	3	Desa	-

	penjualan sepeda motor Honda di Kabupaten Indragiri Hilir					
[13]	Mengelompokkan data pegawai yang layak untuk lolos dalam program SPD (Staff Development Program)	Clustering K-Means	27	3	Pegawai	-
[7]	Memberikan kemudahan pada PT. XYZ untuk mengelola data transaksi penjualannya serta mengetahui tingkat penjualan produk pada setiap cabangnya di Sumatera	Clustering K-Means	712	3	Data Penjualan	Library Python: pandas, numpy, scikit-learn, matplotlib. Visualisasi: Scatter plot menggunakan matplotlib
[32]	Meneliti dan mengimplementasikan segmentasi pelanggan pada data transaksi <i>online retail</i> untuk mengidentifikasi karakteristik pelanggan dan menentukan strategi pemasaran yang lebih tepat.	Clustering K-Means	54.910	4	Data Penjualan	Library Python: Pandas, Numpy, Seaborn, Scikit-learn (K-Means, MinMaxScaler, StandardScaler, Metrics). Visualisasi: Scatter plot menggunakan matplotlib
[33]	Melakukan customer profiling pelanggan produk digital service Indihome menggunakan Algoritma K-Means serta mengetahui perilaku pelanggan dan menerapkan strategi pemasaran yang tepat untuk mendatangkan keuntungan bagi perusahaan	Clustering K-Means	1.392	3	Data Pelanggan	Library Python: pandas, numpy, Scikit-learn (K-Means, MinMaxScaler) Visualisasi: Scatter plot menggunakan matplotlib
[37]	Menganalisis prestasi akademik siswa dengan mengelompokkan tingkat pencapaian mereka menggunakan metode k-means clustering	Clustering K-Means	266	3	Siswa	RapidMiner untuk analisis data, penerapan algoritma K-Means, dan perhitungan validitas cluster menggunakan

						indeks Davies-Bouldin.
[38]	Mengelompokkan prestasi siswa berdasarkan nilai mata pelajaran dan absensi menggunakan metode k-means clustering	Clustering K-Means	25	3	Siswa	RStudio untuk analisis clustering K-Means dan metode <i>Simple Additive Weighting</i> (SAW) untuk menentukan siswa dengan peringkat tertinggi pada cluster tinggi.
Klasterisasi Kemampuan Berpikir Komputasional Siswa Berdasarkan Data Tantangan Bebras Menggunakan Algoritma K-Means	Tujuan penelitian ini adalah untuk menganalisis dan mengelompokkan data peserta Tantangan Bebras 2024 menggunakan EDA dan metode clustering.	EDA dan Clustering K-Means	156.753	3-4	Siswa	Microsoft SQL Server, Library Python: pandas, scikit-learn, matplotlib

BAB III

LANDASAN TEORI

3.1. Berpikir Komputasional (*Computational Thinking/CT*)

Berpikir komputasional adalah pendekatan dalam menyelesaikan masalah berdasarkan data masukan dengan menerapkan algoritma tertentu. Pendekatan ini melibatkan teknik-teknik yang biasa digunakan dalam pemrograman perangkat lunak. Namun, berpikir komputasional bukan berarti meniru cara kerja komputer, melainkan berpikir secara logis dan sistematis untuk merumuskan masalah ke dalam bentuk yang dapat diselesaikan secara komputasional, serta merancang solusi yang efektif dan efisien [14]. Berpikir komputasi (*Computational Thinking/CT*) telah diakui sebagai kompetensi krusial di abad ke-21, menekankan pada proses berpikir untuk memecahkan masalah dengan cara yang dapat dipahami dan dilaksanakan oleh komputer atau agen pemroses informasi lainnya. Konsep ini dipopulerkan pada awal abad ke-21, yang menyatakan bahwa CT adalah keterampilan fundamental bagi setiap orang, bukan hanya ilmuwan komputer, dan merupakan bagian dari kemampuan analitis esensial.

Berpikir komputasional (*Computational Thinking/CT*) telah diakui sebagai kompetensi krusial di abad ke-21, menekankan pada proses berpikir untuk memecahkan masalah dengan cara yang dapat dipahami dan dilaksanakan oleh komputer atau agen pemroses informasi lainnya. Konsep ini dipopulerkan pada awal abad ke-21, yang menyatakan bahwa CT adalah keterampilan fundamental bagi setiap orang, bukan hanya ilmuwan komputer, dan merupakan bagian dari kemampuan analitis esensial. Dalam penelitian yang dilakukan oleh Arifin dan Utomo [22], CT juga dikaitkan dengan kemampuan literasi digital siswa di Indonesia, di mana keterampilan ini menjadi bagian integral dari pembelajaran di era digital. Penelitian tersebut menunjukkan bahwa pengintegrasian CT dalam kurikulum Indonesia mendukung penguatan kecakapan abad ke-21 dan adaptasi terhadap tantangan revolusi industri 4.0. Kemampuan berpikir komputasional ini kemudian diukur dan dievaluasi secara luas melalui pelaksanaan Tantangan Bebras, yang dibahas pada subbab berikutnya.

3.2. Tantangan Bebras (*Bebras Challenge*)

Tantangan Bebras adalah kontes internasional tahunan yang bertujuan untuk mempromosikan informatika dan berpikir komputasi di kalangan siswa dan guru dari berbagai jenjang usia. Kontes ini terdiri dari tugas-tugas pemecahan masalah singkat yang didasarkan pada konsep informatika dan CT. Tantangan Bebras dirancang agar dapat dijawab tanpa memerlukan pengetahuan informatika sebelumnya atau penggunaan komputer, menjadikannya alat yang efektif untuk memperkenalkan konsep-konsep ini secara luas. Soal-soal Bebras diklasifikasikan menggunakan sistem dua dimensi yang mencakup subbidang informatika dan kompetensi CT yang diukur. Klasifikasi ini membantu dalam mengorganisir tugas dan memandu pengembangan soal baru yang relevan dengan kurikulum pendidikan informatika. Analisis data hasil Tantangan Bebras memberikan

wawasan tentang tingkat kompetensi siswa dan relevansi tingkat kesulitan soal [15].

Di Indonesia, pelaksanaan Bebras berada di bawah koordinasi Bebras Indonesia, yang telah bekerja sama dengan Kemdikbud dan universitas-universitas mitra. Menurut penelitian oleh Mayasari dan Huda [23], Bebras Challenge terbukti efektif dalam meningkatkan minat siswa terhadap informatika dan kemampuan berpikir logis. Penelitian tersebut menekankan bahwa pendekatan soal Bebras yang kontekstual dan menyenangkan membuatnya cocok diterapkan di berbagai jenjang pendidikan. Selain itu, hasil dari Tantangan Bebras juga dapat dijadikan tolak ukur untuk mengevaluasi kesiapan siswa dalam menghadapi tantangan digital yang kompleks. Untuk mengeksplorasi lebih jauh profil kemampuan siswa dari hasil Tantangan Bebras, penelitian ini menggunakan pendekatan klasterisasi yang dibahas dalam subbab berikutnya.

3.3. Klasterisasi

Klasterisasi (*Clustering*) adalah salah satu teknik *data mining* yang paling umum dan fundamental. Klasterisasi (*Clustering*) adalah tugas untuk mengelompokkan sekumpulan objek sedemikian rupa sehingga objek dalam kelompok yang sama (disebut kluster) lebih mirip satu sama lain daripada objek di kelompok lain [16]. Klasterisasi adalah proses kognitif dasar yang digunakan untuk mengatur objek ke dalam kategori, baik oleh manusia maupun mesin.

Dalam praktiknya, klasterisasi digunakan untuk berbagai tujuan, mulai dari segmentasi pasar, pengenalan pola, hingga pemetaan perilaku. Dalam konteks pendidikan, klasterisasi telah digunakan untuk mengelompokkan siswa berdasarkan kemampuan, gaya belajar, atau tingkat penguasaan materi. Penelitian oleh Nugroho dan Fitriani [24] menunjukkan bahwa penerapan metode klasterisasi dapat membantu guru dalam merancang strategi pembelajaran yang lebih sesuai dengan kebutuhan kelompok siswa yang berbeda. Dalam penelitian ini, teknik klasterisasi digunakan untuk mengidentifikasi pola penguasaan berpikir komputasi siswa berdasarkan hasil Tantangan Bebras. Untuk menerapkan teknik klasterisasi secara efisien dan terstruktur pada data siswa, algoritma K-Means dipilih sebagai metode utama dalam penelitian ini, sebagaimana dijelaskan pada subbab berikut.

3.4. Algoritma K-Means

Algoritma K-Means adalah algoritma klasterisasi berbasis partisi yang paling populer dan banyak digunakan karena kesederhanaan dan efisiensinya. K-Means mengelompokkan n objek data ke dalam k kluster yang telah ditentukan sebelumnya. Prinsip kerja K-Means melibatkan langkah-langkah iteratif sebagai berikut yaitu Inisialisasi (menentukan jumlah kluster k dan memilih k titik data awal sebagai centroid), Penugasan (setiap objek data ditugaskan ke kluster yang centroid-nya memiliki jarak terdekat), Pembaruan (centroid setiap kluster dihitung ulang sebagai nilai rata-rata dari semua objek data yang ditugaskan ke kluster tersebut), dan Iterasi (langkah penugasan dan pembaruan diulang hingga posisi centroid tidak lagi berubah secara signifikan atau jumlah iterasi maksimum tercapai). Dalam penelitian ini, K-Means akan menjadi algoritma utama untuk mengelompokkan siswa berdasarkan profil penguasaan CT skill yang diekstraksi

dari data Tantangan Bebras, memungkinkan identifikasi kelompok siswa dengan karakteristik kecerdasan komputasional yang serupa [17].

Algoritma K-Means efektif dalam mengelompokkan data non-hierarkis dalam waktu yang relatif singkat [25]. K-Means berhasil digunakan untuk mengelompokkan profil siswa berdasarkan hasil asesmen kompetensi matematika [25]. Hal ini menunjukkan potensi algoritma ini dalam mengelompokkan data siswa yang kompleks secara efisien, termasuk dalam konteks analisis kemampuan berpikir komputasi. Sebagai bagian dari metode kecerdasan buatan, K-Means merupakan salah satu contoh dari pendekatan machine learning yang akan dijelaskan lebih lanjut pada subbab berikutnya. Berikut ini adalah langkah-langkah algoritma k-means:

- a. Langkah pertama dalam proses klusterisasi menggunakan algoritma k-means adalah menentukan jumlah kluster (k) yang akan digunakan.
- b. Penetapan awal titik pusat (*Centroid*) dilakukan secara acak. Selanjutnya, nilai centroid diperbarui secara iteratif menggunakan rumus sebagai berikut [7]:

$$V_{ij} = \frac{1}{N_i} \sum_{k=0}^{N_i} X_{kj}$$

Keterangan:

V_{ij} : Rata-rata *Centroid* cluster ke- i untuk variable ke- j

N_i : Jumlah anggota cluster ke- i ,

k : Indeks cluster

j : Indeks variabel

X_{kj} : Nilai data ke- k variabel ke- j untuk cluster tersebut

- c. Untuk setiap data, dihitung jaraknya terhadap setiap centroid menggunakan rumus *Euclidean Distance* berikut [7]:

$$De = \sqrt{(xi - si)^2 + (yi - ti)^2}$$

Keterangan:

De : Jarak *Euclidean* antara data dan *Centroid*

I : Indeks objek/data

(x, y) : Koordinat objek

(s, t) : Koordinat *Centroid*

- d. Pengelompokkan data berdasarkan jarak terdekat terhadap titik *Centroid*.
- e. Melakukan iterasi dari Langkah 3 hingga 4 dilakukan secara berulang hingga nilai optimal pada *Centroid* tercapai.

3.5. Metode *Silhouette*

Metode *Silhouette*, dikembangkan oleh Rousseeuw [34], merupakan teknik evaluasi clustering yang mengukur seberapa sesuai sebuah objek dengan klasternya