

BAB I

PENDAHULUAN

1.1 Latar Belakang

Perkembangan teknologi informasi dan digitalisasi bisnis dalam beberapa dekade terakhir telah menghasilkan pertumbuhan data yang sangat pesat di berbagai sektor industri. Fenomena ini dikenal sebagai era *Big data*, yaitu kondisi di mana organisasi menghasilkan dan mengumpulkan data dalam jumlah yang sangat besar, dengan tingkat variasi dan kecepatan yang tinggi [48]. Data tersebut berasal dari berbagai sumber seperti transaksi bisnis, aktivitas pengguna pada platform digital, sistem operasional perusahaan, serta perangkat *Internet of Things* (IoT). Pertumbuhan data global ini diperkirakan mencapai skala zettabyte setiap tahunnya, sehingga menuntut organisasi untuk memiliki kemampuan dalam mengelola serta menganalisis data secara efektif [49].

Big data tidak hanya berkaitan dengan volume data yang besar, tetapi juga mencakup berbagai karakteristik lain seperti *variety*, *velocity*, *veracity*, dan *value* yang dikenal sebagai konsep 5V dalam *Big data* [50]. Keberadaan data dalam jumlah besar memberikan peluang bagi organisasi untuk memperoleh wawasan baru yang dapat mendukung pengambilan keputusan strategis. Melalui pendekatan *Big data Analytics*, organisasi dapat mengekstraksi informasi yang bernilai dari data yang kompleks dan tidak terstruktur sehingga mampu meningkatkan efisiensi operasional, memahami perilaku pelanggan, serta mengidentifikasi peluang bisnis baru [2]. Oleh karena itu, data saat ini dianggap sebagai salah satu aset strategis yang dapat memberikan keunggulan kompetitif bagi perusahaan yang mampu memanfaatkannya secara optimal.

Dalam konteks dunia bisnis, khususnya perusahaan yang bergerak di bidang proyek, konstruksi, logistik, maupun layanan profesional, aktivitas operasional perusahaan menghasilkan berbagai jenis data transaksi [51]. Data transaksi tersebut

dapat berupa *invoice*, faktur pajak, *purchase order*, data pembayaran, serta informasi terkait proyek yang sedang dijalankan. Data transaksi ini mencerminkan aktivitas bisnis perusahaan secara keseluruhan, termasuk nilai proyek, hubungan dengan klien, serta pola pembayaran yang terjadi dalam suatu periode tertentu. Namun, dalam praktiknya, banyak organisasi masih memanfaatkan data transaksi hanya sebagai dokumentasi administratif tanpa melakukan analisis yang lebih mendalam terhadap data tersebut.

Perusahaan yang menjadi objek dalam penelitian ini merupakan perusahaan yang bergerak di bidang jasa konstruksi dan pengelolaan proyek, yang dalam penelitian ini disamarkan sebagai PT XYZ untuk menjaga kerahasiaan data. Dalam operasionalnya, perusahaan ini menghasilkan data transaksi yang cukup besar, khususnya data *invoice* yang berkaitan dengan berbagai proyek yang telah dan sedang berjalan. Data *invoice* tersebut mencakup berbagai informasi penting seperti nilai transaksi, pelanggan, jenis proyek, serta detail administratif lainnya. Seiring berjalannya waktu, jumlah data yang tersimpan dalam sistem perusahaan semakin meningkat secara signifikan.

Data *invoice* dipilih sebagai objek penelitian karena merupakan representasi aktivitas transaksi perusahaan yang mencatat hubungan bisnis antara perusahaan dengan pelanggan. Informasi seperti nilai transaksi, frekuensi transaksi, serta riwayat transaksi pelanggan yang tersimpan pada data *invoice* dapat menggambarkan karakteristik perilaku transaksi setiap customer. Dengan demikian, data *invoice* tidak hanya berfungsi sebagai dokumen administratif, tetapi juga memiliki potensi untuk dimanfaatkan sebagai sumber informasi dalam proses analisis customer segmentation dan pengambilan keputusan bisnis berbasis data.

Namun, berdasarkan hasil observasi awal, terdapat permasalahan dalam pengelolaan data *invoice* dan transaksi pelanggan tersebut. Data yang dimiliki perusahaan belum terorganisir dan terklasifikasi secara optimal, sehingga menyulitkan dalam proses analisis dan pengambilan keputusan bisnis. Banyaknya jumlah data yang tersedia tidak diimbangi dengan struktur pengelolaan data yang

baik, sehingga informasi penting yang terkandung di dalamnya belum dapat dimanfaatkan secara maksimal. Kondisi ini menyebabkan data hanya berfungsi sebagai arsip administratif tanpa memberikan nilai tambah dalam bentuk insight bagi perusahaan, khususnya dalam memahami karakteristik pelanggan dan pola transaksi yang dimiliki.

Permasalahan ini juga berdampak pada sumber daya manusia di dalam perusahaan. Tidak semua karyawan mampu memahami isi dan struktur data yang tersedia, terutama karyawan baru yang belum memiliki pengalaman yang cukup dalam memahami pola transaksi perusahaan. Pengetahuan mengenai pelanggan prioritas, pelanggan dengan kontribusi transaksi tinggi, atau pola transaksi tertentu cenderung hanya dimiliki oleh karyawan yang telah lama bekerja di perusahaan. Hal ini menyebabkan terjadinya ketergantungan terhadap individu tertentu dalam memahami data perusahaan, yang pada akhirnya dapat menghambat efisiensi kerja, proses transfer pengetahuan, serta pengambilan keputusan bisnis secara cepat dan objektif.

Selain itu, tanpa adanya pengelompokan data pelanggan yang jelas, perusahaan mengalami kesulitan dalam mengidentifikasi pelanggan yang memberikan kontribusi terbesar terhadap pendapatan perusahaan maupun pelanggan yang memiliki tingkat aktivitas transaksi tinggi secara konsisten. Padahal, informasi tersebut sangat penting dalam mendukung strategi bisnis perusahaan, khususnya dalam menentukan prioritas *customer*, meningkatkan *customer relationship management* (CRM), serta menentukan strategi pelayanan dan pendekatan bisnis yang lebih tepat sasaran. Ketidakmampuan dalam mengelompokkan data pelanggan secara efektif juga dapat menghambat perusahaan dalam melakukan analisis tren transaksi, evaluasi hubungan pelanggan, serta pengembangan strategi bisnis jangka Panjang [52].

Selama ini informasi mengenai nilai transaksi maupun frekuensi transaksi pelanggan sebenarnya dapat diperoleh melalui proses rekapitulasi menggunakan aplikasi spreadsheet seperti Microsoft Excel. Namun, pendekatan tersebut hanya

menghasilkan ringkasan data dan belum mampu mengidentifikasi kelompok pelanggan yang memiliki karakteristik transaksi serupa secara otomatis berdasarkan banyak atribut secara bersamaan [53]. Akibatnya, proses identifikasi customer dengan pola transaksi yang mirip masih bergantung pada analisis manual dan interpretasi pengguna, terutama ketika jumlah data transaksi terus bertambah setiap tahunnya.

Berdasarkan permasalahan tersebut, diperlukan suatu pendekatan analisis data yang mampu mengelompokkan data transaksi pelanggan secara otomatis berdasarkan kemiripan karakteristiknya. Salah satu metode yang dapat digunakan adalah teknik *clustering* dalam *data mining*, yang memungkinkan pengelompokan data tanpa memerlukan label sebelumnya [54]. Dengan menerapkan teknik *clustering*, diharapkan data transaksi pelanggan dapat dikelompokkan menjadi beberapa segmen *customer* yang lebih terstruktur, sehingga memudahkan perusahaan dalam memahami karakteristik pelanggan, menentukan prioritas *customer*, serta mengidentifikasi pola-pola penting dalam transaksi bisnis perusahaan.

Pada penelitian ini, teknik clustering dipilih karena dataset perusahaan tidak memiliki label yang menunjukkan kategori customer, seperti pelanggan prioritas, pelanggan aktif, maupun pelanggan dengan kontribusi transaksi rendah. Oleh karena itu, pendekatan supervised learning tidak dapat diterapkan secara langsung. Clustering sebagai metode unsupervised learning memungkinkan proses pengelompokan customer dilakukan berdasarkan kemiripan karakteristik transaksi tanpa memerlukan label sebelumnya, sehingga pola-pola yang sebelumnya tidak terlihat melalui analisis deskriptif dapat diidentifikasi secara otomatis.

Teknik *clustering* telah banyak diterapkan dalam berbagai bidang penelitian maupun industri. Dalam sektor bisnis, *clustering* sering digunakan untuk melakukan segmentasi pelanggan, analisis perilaku konsumen, pengelompokan produk, serta analisis pola transaksi [55]. Melalui proses pengelompokan data, organisasi dapat memahami karakteristik kelompok data tertentu sehingga dapat

merancang strategi bisnis yang lebih efektif. Sebagai contoh, perusahaan dapat menggunakan *clustering* untuk mengidentifikasi kelompok pelanggan dengan nilai transaksi tinggi atau mengelompokkan produk berdasarkan pola penjualan tertentu.

Berbeda dengan analisis deskriptif yang hanya menampilkan informasi seperti total transaksi atau frekuensi transaksi setiap pelanggan, teknik clustering mampu mengelompokkan pelanggan berdasarkan kemiripan beberapa karakteristik transaksi secara bersamaan. Dengan demikian, perusahaan tidak hanya mengetahui siapa pelanggan dengan nilai transaksi terbesar, tetapi juga memahami kelompok pelanggan yang memiliki pola transaksi serupa sehingga strategi pemasaran dan pengelolaan hubungan pelanggan dapat disusun secara lebih tepat sasaran.

Berbagai algoritma *clustering* telah dikembangkan dalam bidang *data mining*, di antaranya *K-means*, *DBSCAN*, *Agglomerative Clustering*, dan *Gaussian Mixture Model* (GMM). Setiap algoritma *clustering* memiliki pendekatan yang berbeda dalam melakukan proses pengelompokan data. Algoritma *K-means* menggunakan *centroid*, *DBScan* Berbasis *density*, *Agglomerative* berbasis *hierarchy*, dan GMM berbasis probabilitas [56].

Perbedaan pendekatan yang digunakan oleh masing-masing algoritma *clustering* menyebabkan setiap algoritma memiliki keunggulan dan keterbatasan tersendiri dalam proses pengelompokan data. Beberapa algoritma mungkin memberikan hasil *cluster* yang lebih baik pada dataset tertentu, tetapi kurang optimal pada dataset lain yang memiliki karakteristik berbeda. Oleh karena itu, pemilihan algoritma *clustering* yang tepat menjadi faktor penting dalam memperoleh hasil analisis yang optimal.

Sejumlah penelitian sebelumnya telah melakukan analisis perbandingan algoritma *clustering* untuk mengetahui algoritma yang paling efektif dalam melakukan pengelompokan data. Hasil penelitian menunjukkan bahwa tidak terdapat satu algoritma *clustering* yang secara konsisten memberikan hasil terbaik pada semua jenis dataset. Performa algoritma *clustering* sangat bergantung pada

karakteristik data yang dianalisis, seperti distribusi data, jumlah atribut, serta keberadaan noise dalam dataset. Oleh karena itu, analisis komparatif terhadap beberapa algoritma *clustering* diperlukan untuk menentukan algoritma yang paling sesuai dengan dataset yang digunakan dalam penelitian.

Meskipun berbagai penelitian mengenai *clustering* telah dilakukan, sebagian besar penelitian sebelumnya lebih banyak berfokus pada analisis *clustering* pada data pelanggan, data penjualan, maupun data perilaku konsumen. Penelitian yang secara khusus membahas penerapan *clustering* pada data transaksi proyek atau data *invoice* perusahaan masih relatif terbatas. Padahal, data transaksi perusahaan memiliki karakteristik yang kompleks dan berpotensi memberikan wawasan bisnis yang penting apabila dianalisis menggunakan teknik *data mining*. Selain itu, sebagian besar penelitian terdahulu berhenti pada tahap evaluasi kualitas *clustering* menggunakan metrik teknis tanpa mengimplementasikan hasil segmentasi ke dalam sistem yang dapat dimanfaatkan secara langsung oleh pengguna bisnis maupun melakukan validasi terhadap penerimaan pengguna.

Berdasarkan permasalahan tersebut, penelitian ini bertujuan untuk melakukan analisis komparasi beberapa algoritma *clustering* dalam mengelompokkan data transaksi *invoice* perusahaan. Dataset yang digunakan dalam penelitian ini merupakan data transaksi proyek yang mencakup informasi mengenai nilai transaksi, proyek, serta atribut terkait lainnya dalam suatu periode tertentu. Melalui penerapan beberapa algoritma *clustering*, penelitian ini berupaya untuk mengidentifikasi pola segmentasi pelanggan berdasarkan karakteristik transaksi yang dimiliki.

Hasil *clustering* dari masing-masing algoritma kemudian akan dievaluasi menggunakan beberapa metrik evaluasi *clustering*, seperti *Silhouette Score*, *Davies-Bouldin Index*, dan *Calinski-Harabasz Index*, untuk menentukan algoritma yang menghasilkan kualitas *cluster* terbaik. Dengan melakukan analisis komparasi beberapa algoritma *clustering* terhadap dataset transaksi perusahaan, penelitian ini diharapkan dapat memberikan kontribusi dalam pemahaman mengenai metode

clustering yang paling optimal untuk analisis data transaksi bisnis. Selain itu, hasil penelitian ini juga diharapkan dapat memberikan wawasan yang bermanfaat bagi perusahaan dalam memahami pola data transaksi yang dimiliki serta mendukung pengambilan keputusan berbasis data di lingkungan bisnis.

1.2 Rumusan Masalah

Berdasarkan latar belakang tersebut, berikut adalah rumusan masalah untuk penelitian ini:

1. Bagaimana penerapan metode *clustering* dalam mengelompokkan data transaksi proyek berdasarkan dataset *invoice* perusahaan?
2. Bagaimana hasil perbandingan performa beberapa algoritma *clustering* dalam melakukan pengelompokan data transaksi proyek?
3. Algoritma *clustering* manakah yang menghasilkan kualitas *cluster* terbaik berdasarkan metrik evaluasi *clustering* yang digunakan dalam penelitian ini?

1.3 Batasan Masalah

Agar penelitian ini lebih terarah dan tidak meluas dari tujuan yang telah ditetapkan, maka diperlukan batasan masalah yang menjadi ruang lingkup penelitian. Adapun batasan masalah dalam penelitian ini adalah sebagai berikut:

1. Sumber Data Penelitian:

Dataset yang digunakan dalam penelitian ini merupakan data transaksi *invoice* perusahaan yang mencakup informasi terkait proyek, nilai transaksi, serta atribut terkait lainnya pada periode tertentu. Data yang digunakan merupakan data historis yang diperoleh dari sumber data Perusahaan.

2. Jenis Analisis Data:

Penelitian ini berfokus pada analisis data menggunakan teknik *data mining* dengan metode *clustering* untuk melakukan pengelompokan data transaksi proyek berdasarkan karakteristik tertentu.

3. Ruang Lingkup Hasil Penelitian:

Penelitian ini hanya berfokus pada implementasi hasil analisis dan perbandingan performa algoritma *clustering* dalam mengelompokkan data *invoice*. Penelitian ini tidak mencakup pengambilan Keputusan didalam perusahaan secara langsung dalam operasional perusahaan.

1.4 Tujuan dan Manfaat Penelitian

1.4.1 Tujuan Penelitian

Berdasarkan rumusan masalah yang telah dijelaskan sebelumnya, maka tujuan dari penelitian ini adalah sebagai berikut:

1. Menerapkan metode *clustering* untuk mengelompokkan data transaksi proyek berdasarkan dataset *invoice* perusahaan.
2. Menganalisis dan membandingkan performa beberapa algoritma *clustering* dalam mengelompokkan data transaksi proyek.
3. Mengidentifikasi algoritma *clustering* yang menghasilkan kualitas *cluster* terbaik berdasarkan metrik evaluasi *clustering* yang digunakan dalam penelitian ini.

1.4.2 Manfaat Penelitian

Penelitian ini diharapkan dapat memberikan manfaat baik secara teoritis maupun praktis, sebagai berikut:

1. Manfaat Teoritis

Penelitian ini diharapkan dapat memberikan kontribusi dalam pengembangan ilmu pengetahuan di bidang *Big data Analytics* dan *data mining*, khususnya dalam penerapan metode *clustering* untuk analisis data transaksi bisnis. Selain itu, penelitian ini juga dapat menjadi referensi bagi penelitian selanjutnya yang berkaitan dengan analisis komparasi algoritma *clustering* pada dataset bisnis atau transaksi perusahaan.

2. Manfaat Praktis

Penelitian ini diharapkan dapat memberikan wawasan bagi perusahaan dalam memahami pola data transaksi yang dimiliki melalui proses segmentasi data menggunakan metode *clustering*. Dengan adanya hasil pengelompokan data transaksi proyek, perusahaan dapat memperoleh gambaran mengenai karakteristik

proyek atau transaksi tertentu sehingga dapat membantu dalam pengambilan keputusan bisnis yang lebih efektif dan berbasis data.

Selain itu, hasil penelitian ini juga dapat memberikan rekomendasi algoritma *clustering* yang paling optimal untuk menganalisis data transaksi perusahaan berdasarkan kualitas *cluster* yang dihasilkan.

1.5 Sistematika Penulisan

Sistematika penulisan dalam penelitian ini disusun untuk memberikan gambaran yang jelas mengenai alur pembahasan serta struktur penyusunan laporan skripsi secara keseluruhan. Dengan adanya sistematika penulisan, pembaca dapat memahami urutan pembahasan yang disajikan dalam setiap bab sehingga memudahkan dalam mengikuti proses penelitian yang dilakukan. Secara umum, laporan skripsi ini terdiri dari lima bab utama yang saling berkaitan satu sama lain, yaitu Bab I Pendahuluan, Bab II Landasan Teori, Bab III Metodologi Penelitian, Bab IV Hasil dan Pembahasan, serta Bab V Kesimpulan dan Saran.

BAB I PENDAHULUAN

Bab ini merupakan bagian awal dari laporan skripsi yang berfungsi untuk memberikan gambaran umum mengenai penelitian yang dilakukan. Pada bab ini dijelaskan latar belakang penelitian yang memaparkan pentingnya pemanfaatan *Big data* dan teknik *data mining* dalam menganalisis data transaksi perusahaan. Selain itu, bab ini juga menjelaskan permasalahan yang melatarbelakangi penelitian serta alasan mengapa penelitian ini perlu dilakukan. Selanjutnya, pada bab ini juga disusun rumusan masalah yang menjadi fokus utama penelitian. Rumusan masalah ini dirumuskan berdasarkan permasalahan yang telah dijelaskan pada latar belakang penelitian.

Bab ini juga memuat batasan masalah yang bertujuan untuk memperjelas ruang lingkup penelitian agar penelitian tetap terarah dan tidak meluas ke pembahasan yang tidak relevan. Selain itu, pada bab ini juga dijelaskan tujuan penelitian yang ingin dicapai melalui proses analisis data yang dilakukan. Tujuan penelitian ini berkaitan langsung dengan rumusan masalah yang telah ditetapkan

sebelumnya. Selanjutnya, bab ini juga menjelaskan manfaat penelitian baik secara teoritis maupun praktis. Manfaat teoritis berkaitan dengan kontribusi penelitian terhadap perkembangan ilmu pengetahuan, khususnya dalam bidang *Big data Analytics* dan *data mining*. Sementara itu, manfaat praktis berkaitan dengan kegunaan hasil penelitian bagi pihak perusahaan maupun pihak lain yang berkepentingan. Bab ini diakhiri dengan sistematika penulisan yang menjelaskan susunan bab dalam laporan skripsi secara keseluruhan.

BAB II LANDASAN TEORI

Bab ini berisi kajian teori yang menjadi dasar dalam pelaksanaan penelitian. Pada bab ini dijelaskan berbagai konsep, teori, serta metode yang berkaitan dengan topik penelitian. Pembahasan pada bab ini meliputi teori mengenai *Big data*, konsep *data mining*, serta teknik *clustering* yang digunakan dalam proses analisis data. Selain itu, bab ini juga menjelaskan berbagai algoritma *clustering* yang digunakan dalam penelitian, seperti *K-means*, *DBSCAN*, *Agglomerative Clustering*, dan *Gaussian Mixture Model*.

Selain membahas teori yang berkaitan dengan algoritma *clustering*, pada bab ini juga dijelaskan mengenai metrik evaluasi *clustering* yang digunakan untuk menilai kualitas *cluster* yang dihasilkan oleh masing-masing algoritma. Metrik evaluasi tersebut meliputi *Silhouette Score*, *Davies-Bouldin Index*, serta *Calinski-Harabasz Index*. Pada bab ini juga dibahas mengenai tools atau perangkat lunak yang digunakan dalam penelitian, seperti bahasa pemrograman Python serta *library* yang digunakan untuk analisis data dan implementasi algoritma *clustering*.

Selain teori-teori yang mendukung penelitian, bab ini juga memuat pembahasan mengenai penelitian terdahulu yang relevan dengan topik penelitian. Penelitian terdahulu ini digunakan sebagai referensi untuk memahami perkembangan penelitian terkait serta untuk mengidentifikasi kesenjangan penelitian. Dengan mempelajari penelitian terdahulu, peneliti dapat mengetahui metode yang telah digunakan sebelumnya serta menemukan peluang untuk

melakukan pengembangan atau perbandingan metode dalam penelitian yang dilakukan.

BAB III METODOLOGI PENELITIAN

Bab ini menjelaskan metode yang digunakan dalam pelaksanaan penelitian. Pada bagian awal bab ini dijelaskan gambaran umum objek penelitian yang menjelaskan sumber data yang digunakan dalam penelitian, yaitu dataset transaksi *invoice* perusahaan. Dataset tersebut mencakup berbagai informasi terkait transaksi proyek, nilai transaksi, serta atribut lainnya yang relevan untuk dianalisis menggunakan teknik *data mining*.

Selanjutnya, pada bab ini juga dijelaskan metode penelitian yang digunakan dalam proses analisis data. Penelitian ini menggunakan pendekatan *data mining* dengan metode *clustering* untuk mengelompokkan data transaksi proyek berdasarkan karakteristik tertentu. Bab ini juga menjelaskan tahapan penelitian yang dilakukan mulai dari proses pengumpulan data, *preprocessing data*, transformasi data, hingga proses analisis menggunakan beberapa algoritma *clustering*.

Selain itu, pada bab ini juga dijelaskan teknik pengumpulan data yang digunakan dalam penelitian, yaitu data sekunder yang diperoleh dari sumber data perusahaan. Selanjutnya, dijelaskan pula variabel penelitian yang digunakan dalam proses analisis data, termasuk atribut atau fitur yang digunakan dalam proses *clustering*. Bab ini juga menjelaskan teknik analisis data yang digunakan untuk menerapkan algoritma *clustering* serta metode evaluasi yang digunakan untuk menilai kualitas *cluster* yang dihasilkan oleh masing-masing algoritma.

BAB IV HASIL DAN PEMBAHASAN

Bab ini menyajikan hasil dari proses analisis data yang telah dilakukan pada penelitian. Pada bab ini dijelaskan proses implementasi algoritma *clustering* terhadap dataset transaksi perusahaan. Hasil *clustering* yang diperoleh dari masing-

masing algoritma kemudian dianalisis untuk melihat pola pengelompokan data yang terbentuk.

Selain itu, pada bab ini juga dilakukan evaluasi terhadap hasil *clustering* menggunakan metrik evaluasi *clustering* seperti *Silhouette Score*, *Davies-Bouldin Index*, dan *Calinski-Harabasz Index*. Hasil evaluasi tersebut digunakan untuk membandingkan performa masing-masing algoritma *clustering* yang digunakan dalam penelitian. Melalui proses komparasi tersebut, dapat diketahui algoritma *clustering* mana yang memberikan kualitas *cluster* terbaik pada dataset yang digunakan.

Bab ini juga memuat pembahasan mengenai hasil analisis yang diperoleh serta interpretasi terhadap pola *cluster* yang terbentuk. Pembahasan ini bertujuan untuk memberikan pemahaman mengenai karakteristik kelompok data yang terbentuk serta implikasi dari hasil analisis tersebut terhadap pemahaman pola transaksi proyek perusahaan.

BAB V KESIMPULAN DAN SARAN

Bab ini merupakan bagian akhir dari laporan skripsi yang berisi kesimpulan dari seluruh proses penelitian yang telah dilakukan. Kesimpulan yang disajikan pada bab ini disusun berdasarkan hasil analisis dan pembahasan yang telah dijelaskan pada bab sebelumnya. Kesimpulan tersebut bertujuan untuk menjawab rumusan masalah yang telah ditetapkan pada awal penelitian.

Selain kesimpulan, bab ini juga memuat saran yang dapat diberikan berdasarkan hasil penelitian yang telah dilakukan. Saran tersebut dapat berupa rekomendasi untuk penelitian selanjutnya yang berkaitan dengan pengembangan metode analisis data maupun penerapan teknik *clustering* pada dataset yang lebih besar atau dengan pendekatan algoritma yang berbeda. Selain itu, saran juga dapat diberikan kepada pihak perusahaan sebagai referensi dalam memanfaatkan analisis data untuk mendukung pengambilan keputusan berbasis data.