



**PENGEMBANGAN ALGORITMA KUANTUM DBSCAN
UNTUK KLASTERISASI SUPPLIER**

KUALIFIKASI

MIFTAKHUL ZAEN

99223129

**PROGRAM DOKTOR TEKNOLOGI INFORMASI
UNIVERSITAS GUNADARMA**

2024

DAFTAR ISI

HALAMAN JUDUL	i
DAFTAR ISI.....	ii
DAFTAR GAMBAR.....	iii
DAFTAR GAMBAR.....	iv
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Batasan Penelitian	5
1.3 Rumusan Masalah	6
1.4 Tujuan Penelitian.....	6
1.5 Kontribusi Dan Manfaat Hasil Penelitian	6
BAB II TELAAH PUSTAKA	8
2.1 Klasterisasi	8
2.2 Algoritma DBSCAN	9
2.3 Kuantum	9
2.4 Komputasi Kuantum.....	10
2.5 Supplier	11
2.6 Penelitian Terdahulu.....	12
BAB III METODOLOGI PENELITIAN	17
3.1. Tahapan Penelitian	17
3.2. Algoritma Kuantum DBSCAN	22
3.3. Jadwal Penelitian	23

DAFTAR GAMBAR

Gambar 1.1 Index Pembangunan TIK di Indonesia (BPS 2023).....	1
Gambar 2.1 Ilustrasi Bit dan Qubit	11

DAFTAR GAMBAR

Gambar 1.1 Index Pembangunan TIK di Indonesia (BPS 2023)	1
Gambar 2.1 Ilustrasi Bit dan Qubit	11
Gambar 3.1 Tahapan Penelitian	17
Gambar 3. 2 <i>Flowchart</i> Kuantum DBSCAN	22

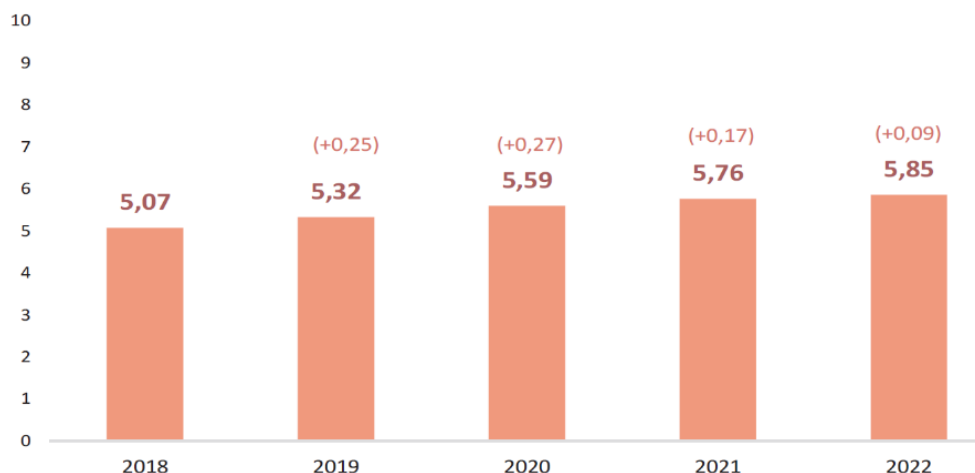
BAB I

PENDAHULUAN

1.1 Latar Belakang

Dalam era digital yang terus berkembang, pengolahan dan analisis data memiliki peran krusial dalam mengambil keputusan yang efektif di berbagai bidang, termasuk ilmu pengetahuan, bisnis, dan teknologi. Terlihat dalam beberapa dekade terakhir di Indonesia, Teknologi Informasi dan Teknologi (TIK) telah membantu pesatnya pertumbuhan industri, mendorong efisiensi dalam pemerintahan, dan operasional bisnis, serta mengembangkan landasan penting menuju perekonomian berbasis teknologi. Sehingga dapat mempermudah perusahaan dalam mencari informasi mengenai supplier dalam jumlah yang banyak.

Mengacu pada data Badan Pusat Statistik pembangunan teknologi informasi dan teknologi di Indonesia, menunjukkan perkembangan positif dalam lima tahun terakhir. Pada tahun 2018, nilai Indeks Pembangunan TIK tercatat sebesar 5,07 dan terus meningkat hingga tahun 2022 dengan nilai mencapai 5,85. Secara keseluruhan, peningkatan indeks yang terjadi dalam lima tahun sebesar 0,78 poin BPS (2023). Terlihat pada Gambar 1.1 menunjukkan perkembangan teknologi dan informasi dalam lima tahun terakhir.



Gambar 1.1 Index Pembangunan TIK di Indonesia (BPS 2023)

Berdasarkan Gambar 1.1 diatas dengan meningkatnya pembangunan teknologi dan komunikasi, hal ini dapat memudahkan dalam pencarian data. Data yang dikumpulkan dapat berjumlah banyak dan beragam, hal tersebut kemudian dikenal dengan istilah *big data*. *Big data* bukan hanya tentang banyak data, tetapi sebenarnya merupakan konsep baru yang memberi kesempatan untuk menemukan wawasan baru tentang data yang ada Kurasova, Marcinkevicius, Medvedev, Rapecka and Stefanovic (2014). Jenis *big data* yang tidak terstruktur menuntut analisis secara langsung yang lebih mendalam dibandingkan dengan dataset standar. Pemahaman yang lebih dalam mengenai nilai-nilai yang ada, seperti kebutuhan untuk mengatur dan memanipulasi kumpulan data yang begitu besar secara unik. Semakin banyaknya data yang tersedia dari berbagai sumber, hal ini juga membawa isu-isu rumit yang menuntut perbaikan yang cepat (Jawad & Al-Bakry, 2023).

Salah satu yang dapat digunakan dalam menganalisis *big data* adalah klusterisasi. Klusterisasi merupakan sebuah metode dari data *mining* yang membagi data ke dalam kluster berdasarkan nilai similaritas antara data. Klusterisasi bertujuan untuk mengelompokkan data yang tidak berlabel sehingga objek-objek data yang memiliki karakteristik dan atribut serupa akan berkumpul dalam satu kluster, sehingga kemiripan objek data dalam kluster yang sama akan lebih tinggi jika dibandingkan dengan objek data kluster yang lain Ikotun, Ezugwu, Abualigah, Abuhaija and Heming (2023). Klusterisasi terdapat metode yang dapat digunakan seperti, K-Means Clustering (Ikotun et al., 2023), KNN-Block DBSCAN (Chen, Zhou, Pei, Yu, Chen, Liu, Du, & Xiong, 2019), DBSCAN (Huang, Huang, Xiang, & Yan, 2021), Affinity Propagation (Duan, Liu, Li, Guo, & Yang, 2023).

Pengelompokan berbasis kepadatan data adalah metode untuk mengidentifikasi kelompok atau kluster dalam data berdasarkan kepadatan titik data. Metode ini mengasumsikan bahwa kluster terbentuk di area di mana titik-titik data terkumpul secara padat, sementara area dengan kepadatan data yang rendah dianggap sebagai *noise*. Tujuan dari pengelompokan data berbasis kepadatan adalah untuk mengidentifikasi daerah dengan bentuk yang berubah-

ubah, yang dapat diukur dengan kepadatan titik tertentu. Kluster yang teridentifikasi biasanya adalah wilayah dengan kepadatan tinggi, sedangkan *outlier* adalah wilayah dengan kepadatan rendah. Oleh karena itu, pengelompokan berbasis kepadatan adalah salah satu yang paling populer dalam mengelompokkan data berbasis kepadatan. Salah satu algoritma klustering yang digunakan untuk mengklasterisasi data berdasarkan kepadatan data adalah algoritma DBSCAN (Chen et al., 2019).

DBSCAN (*Density-based Spatial Clustering of Application with Noise*) adalah sebuah metode pembelajaran tanpa pengawasan, diusulkan pertama kali oleh Ester pada tahun 1996 yang menangani masalah pengelompokan secara efisien berdasarkan kepadatan de Moura Ventorim, Luchi, Rodrigues and Varejão (2021). DBSCAN memiliki kapasitas untuk mengidentifikasi titik-titik noise secara efisien dan tepat. Selain itu, DBSCAN juga dapat membedakan kluster dengan bentuk yang berubah-ubah. DBSCAN dapat diaplikasikan dalam berbagai bidang seperti pada bidang deteksi penipuan, khususnya dalam menganalisis transaksi keuangan untuk mengidentifikasi perilaku mencurigakan seperti yang dilakukan oleh Sivarajah, Kamal, Irani and Weerakkody (2017), dalam bidang transportasi dapat digunakan untuk menganalisis karakteristik perjalanan dan hotspot berdasarkan data trajektori taksi yang dilakukan oleh Du, Meng and Liu (2024), DBSCAN juga dapat diaplikasikan untuk mengelompokkan dan mengklasifikasikan cacat jalur kereta seperti yang dilakukan oleh Vishwakarma, DasGupta and Racherla (2024), dan pada bidang industri penerapan DBSCAN digunakan untuk menganalisis data pemeliharaan untuk prediksi kegagalan peralatan, meningkatkan keandalan dan efisiensi operasional (Wang, Li, Han, Sarkar, & Zhou, 2017).

Pada penelitian yang dilakukan oleh Zhang, Peng, Dou, Wu, Sun, Li, Zhang and Yu (2022) mengembangkan algoritma DBSCAN dengan memanfaatkan DRL untuk secara otomatis mencari parameter optimal DBSCAN tanpa bantuan manual, sehingga menghasilkan peningkatan signifikan dalam akurasi dan efisiensi pengelompokan DBSCAN, mencapai peningkatan akurasi pengelompokan hingga 26% dan 25% serta mampu menemukan parameter

dominan dengan efisiensi komputasi tinggi. Selain itu penelitian yang dilakukan oleh de Moura Ventorim et al. (2021) dengan menggabungkan algoritma BIRCH dan DBSCAN menunjukkan bahwa metode yang diusulkan mencapai hasil terbaik, dalam sebagian besar kasus menunjukkan variasi yang rendah. Berdasarkan hasil yang didapatkan terbukti bahwa dengan menggabungkan algoritma BIRCH dan DBSCAN sangat efisien dibandingkan dengan DBSCAN, terutama dalam skenario di mana DBSCAN terkendala waktu atau memori dalam memproses data besar. Meskipun DBSCAN memiliki solusi yang lebih baik dalam klusterisasi, tetapi DBSCAN masih memiliki kelemahan seperti DBSCAN memerlukan waktu yang relatif lama untuk data yang sangat besar karena perlu menghitung jarak antara semua pasangan titik dalam dataset (Luchi, Rodrigues, & Varejão, 2019).

Meningkatkan kinerja DBSCAN perlu mempelajari dan melakukan survei dari pengembangan DBSCAN yang telah dikembangkan oleh beberapa peneliti. Dengan mempelajari dan membandingkan beberapa algoritma, memungkinkan untuk mendapatkan informasi terkait kelebihan dan kekurangan mereka. Salah satu komputasi yang dapat digunakan untuk mengembangkan DBSCAN adalah komputasi kuantum. Algoritma kuantum menawarkan solusi potensial untuk mengatasi kelemahan DBSCAN, terutama dalam hal efisiensi komputasi dan kemampuan mengelola dataset yang sangat besar. Dalam era kuantum computing, kecepatan dan kapasitas pemrosesan yang signifikan dapat dimanfaatkan untuk meningkatkan algoritma pengelompokan tradisional seperti DBSCAN. Pada tahun 1982, Feynman menunjukkan bahwa komputer kuantum dapat mencapai peningkatan kecepatan yang signifikan dibandingkan komputer klasik pada masalah spesifik tertentu (Feynman, 2018).

Penelitian di bidang ini masih berkembang. Penelitian yang dilakukan oleh Lloyd, Mohseni and Rebentrost (2013) tentang algoritma kuantum untuk pemrosesan data menunjukkan potensi signifikan dari komputasi kuantum dalam pengelompokan dan analisis data. Meskipun belum spesifik mengenai DBSCAN, prinsip yang sama dapat diterapkan untuk mengembangkan versi kuantum dari algoritma pengelompokan ini.

Pada penelitian ini untuk mendukung transisi ke penggunaan komputasi kuantum dalam pengelompokan data, perlu dilakukan lebih banyak penelitian dalam pengembangan algoritma yang spesifik dirancang untuk arsitektur kuantum. Termasuk pengembangan *library* dan *framework* yang mendukung pemrograman kuantum. Meskipun tantangan masih ada, termasuk keterbatasan saat ini dalam teknologi komputer kuantum dan kebutuhan untuk lebih memahami bagaimana terbaik mengintegrasikan prinsip kuantum dengan kebutuhan pengelompokan data, arah penelitian ini bertujuan untuk mengatasi batasan algoritma DBSCAN dalam mengolah dataset besar.

Dalam upaya meningkatkan efisiensi dan akurasi pengelompokan *supplier* di era digital yang terus berkembang, penelitian ini bertujuan untuk mengembangkan algoritma DBSCAN dengan memanfaatkan kemampuan komputasi kuantum. Fokus utama dari penelitian ini adalah untuk mengatasi kelemahan utama DBSCAN, khususnya dalam hal penanganan dataset besar, dengan menerapkan prinsip-prinsip komputasi kuantum yang dapat meningkatkan signifikan dalam kecepatan dan efisiensi pemrosesan data. Melalui pengembangan DBSCAN kuantum, penelitian ini diharapkan dapat memberikan wawasan baru dalam klasterisasi *supplier*, memungkinkan analisis yang lebih mendalam dan pengambilan keputusan yang lebih tepat dalam pengelolaan rantai pasok, serta memberikan kontribusi pada kemajuan teknologi informasi dan komunikasi (TIK) dalam mendukung pertumbuhan industri dan perekonomian berbasis teknologi di Indonesia.

1.2 Batasan Penelitian

Masalah-masalah yang didiskusikan dalam penelitian ini dibatasi sebagai berikut. Batasan masalah pada penelitian ini adalah:

1. Data yang digunakan dalam penelitian ini adalah data supplier
2. Kriteria yang digunakan dalam mengelompokkan supplier hanya Harga, Kualitas, dan Waktu Pengiriman.
3. Bahasa pemrograman yang digunakan menggunakan Python.
4. Data yang digunakan hanya menggunakan data sintetik.

1.3 Rumusan Masalah

Berdasarkan latar belakang dan batasan masalah, diperoleh rumusan masalah untuk mengembangkan algoritma DBSCAN. Berikut merupakan rumusan masalah pada penelitian ini.

1. Bagaimana cara mengatasi keterbatasan algoritma DBSCAN klasik dalam mengelola dataset besar pada klasterisasi *supplier* dengan menggunakan komputasi kuantum?
2. Bagaimana peningkatan efisiensi dan akurasi yang dapat dicapai oleh DBSCAN kuantum dalam pengelompokan *supplier* dibandingkan dengan implementasi klasik nya?

1.4 Tujuan Penelitian

Berdasarkan rumusan masalah yang telah dipaparkan, tujuan penelitian dapat dijelaskan sebagai berikut.

1. Mengembangkan versi kuantum dari algoritma DBSCAN untuk mengatasi keterbatasan dalam penanganan dataset besar pada klasterisasi *supplier*.
2. Mengevaluasi peningkatan efisiensi dan akurasi yang dicapai oleh DBSCAN kuantum dalam pengelompokan *supplier* dibandingkan dengan pendekatan klasik.

1.5 Kontribusi Dan Manfaat Hasil Penelitian

Memberikan wawasan baru dalam penggunaan komputasi kuantum untuk meningkatkan algoritma pengelompokan data, khususnya DBSCAN, dalam konteks bisnis dan teknologi. Menyediakan kerangka kerja dan metodologi untuk implementasi dan evaluasi algoritma pengelompokan berbasis kuantum.

Manfaat dari hasil penelitian ini:

1. Bagi Akademisi: Menyediakan basis penelitian lebih lanjut tentang aplikasi komputasi kuantum dalam data science dan *machine learning*, khususnya dalam konteks pengelompokan data.

2. Bagi Industri: Memungkinkan pengambilan keputusan yang lebih efektif dan efisien dalam manajemen rantai pasok dengan menggunakan klasterisasi *supplier* yang ditingkatkan dengan memanfaatkan algoritma kuantum.
3. Bagi Pengembangan TIK: Memberikan kontribusi pada pengembangan teknologi informasi dan komunikasi dengan mengintegrasikan kemajuan komputasi kuantum dalam solusi bisnis dan teknologi.

Melalui penelitian ini, diharapkan dapat memperkuat pemahaman dan penerapan teknologi komputasi kuantum dalam mengatasi tantangan analitik data besar, khususnya dalam pengelompokan *supplier*, yang pada akhirnya mendukung pengambilan keputusan yang lebih cepat dan akurat dalam berbagai sektor industri.

BAB II

TELAAH PUSTAKA

2.1 Klasterisasi

Klaster merupakan pengelompokan objek berdasarkan kesamaan dalam suatu data ke dalam kelas yang sama yang disebut klaster Deng (2020). Metode klustering berpotensi untuk menetapkan titik-titik data tertentu ke dalam beberapa kelompok berdasarkan karakteristiknya Mavi, Zarbakhshnia, Mavi and Kazemi (2023). Menurut Han, Kamber and Pei (2006) klasterisasi merupakan proses pengelompokan sekumpulan data ke dalam beberapa klaster atau kelompok, sehingga yang berada dalam klaster atau kelompok memiliki tingkat kemiripan yang tinggi dan data antar klaster atau kelompok memiliki tingkat kemiripan yang rendah. Sehingga dapat disimpulkan klasterisasi merupakan pengelompokan objek atau data ke dalam klaster atau kelompok berdasarkan tingkat kesamaan objek atau data tersebut, dan objek atau data yang berada dalam klaster atau kelompok yang berbeda memiliki tingkat kemiripan yang rendah.

Klasterisasi data dikelompokkan dalam satu klaster berdasarkan sifat-sifat similaritas yang dimiliki oleh data. Sifat-sifat similaritas ini dapat diekstraksi berdasarkan objek data menggunakan berbagai metode pengukuran. Salah satu metode yang digunakan adalah pengukuran jarak (*distance measurement*). Pada bidang bisnis, analisis klaster membantu analis pasar untuk menganalisis dan menggambarkan karakteristik kelompok pelanggan, sehingga dapat memperoleh pemahaman tentang arah konsumsi, kemampuan konsumsi, kesediaan konsumsi, dan sebagainya dari kelompok pelanggan yang berbeda. Analisis klaster juga membantu untuk analisis data geografis seperti analisis curah hujan, geologi, dan data lainnya terkait erat dengan prediksi bencana alam yang tepat waktu (Deng, 2020).

2.2 Algoritma DBSCAN

DBSCAN (*Density-based Spatial Clustering of Application with Noise*) adalah sebuah metode pembelajaran tanpa pengawasan, diusulkan pertama kali oleh Ester pada tahun 1996 yang menangani masalah pengelompokan secara efisien berdasarkan kepadatan. DBSCAN memiliki keunggulan dibandingkan pengelompokan klasik lainnya. Salah satu keunggulannya adalah kemampuannya untuk mengidentifikasi kluster dalam bentuk yang berubah-ubah. Selain itu, DBSCAN memiliki kemampuan untuk membedakan pola *noise* pada data sehingga membuatnya memiliki sedikit berpengaruh pada pengelompokan (de Moura Ventorim et al., 2021).

Algoritma DBSCAN memiliki dua parameter, epsilon (EPS) dan MinPts, kedua parameter tersebut harus ditetapkan sebelumnya untuk menilai distribusi kepadatan titik Yang, Qian, Li, Gao, Wu, Liu and Zhao (2022). Algoritma DBSCAN mengidentifikasi semua kluster dengan secara sistematis menemukan semua titik inti dan memperluas setiap titik ke semua titik yang dapat dijangkau oleh kepadatan. Algoritma ini dimulai dengan sebuah titik sembarang p dan mengambil lingkungannya. Jika titik tersebut merupakan titik inti, maka ia akan memulai sebuah kluster baru yang diperluas dengan memasukkan semua titik di lingkungannya ke dalam kluster tersebut. Jika sebuah titik inti tambahan ditemukan di lingkungan tersebut, maka pencarian diperluas untuk memasukkan semua titik di lingkungannya. Jika tidak ada lagi titik inti yang ditemukan di lingkungan yang diperluas, maka kluster selesai, dan titik-titik yang tersisa dicari untuk melihat apakah titik inti lain dapat ditemukan untuk memulai kluster baru. Setelah memproses semua titik, titik-titik yang tidak dimasukkan ke dalam kluster harus menjadi titik noise (Hahsler, Piekenbrock, & Doran, 2019).

2.3 Kuantum

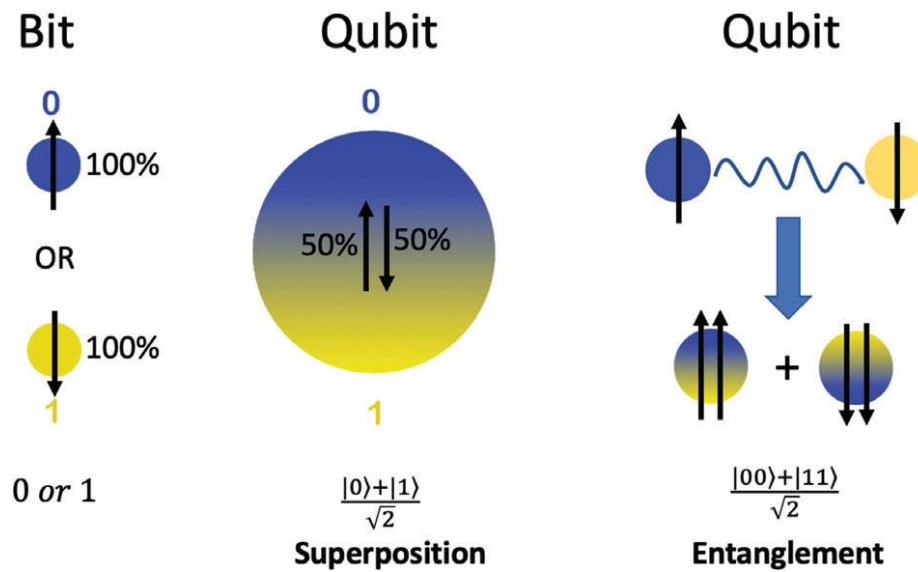
Kuantum adalah istilah yang berasal dari fisika dan merujuk pada konsep dasar dalam mekanika kuantum, yang merupakan cabang fisika yang mempelajari fenomena pada skala sangat kecil, seperti atom dan partikel subatomik. Mekanika kuantum mengubah sudut pandang dalam memahami dunia fisik, terutama pada

level mikroskopis, dengan prinsip-prinsip yang berbeda dari mekanika klasik (yang berlaku pada skala lebih besar, seperti benda sehari-hari). Teori kuantum (digunakan secara sinonim dengan istilah mekanika kuantum) adalah menyatakan "pertama dan terutama kalkulus untuk menghitung probabilitas hasil pengukuran yang dilakukan pada sistem fisik" yang disebut sistem kuantum (Leifer & Poulin, 2008).

Sistem kuantum biasanya merupakan kumpulan objek fisik mikroskopis yang tidak dapat dijelaskan oleh mekanika Newton klasik dalam pengamatan eksperimental. Contohnya adalah atom hidrogen, cahaya dengan intensitas yang sangat rendah, dan sejumlah kecil elektron dalam medan magnet.

2.4 Komputasi Kuantum

Menurut Nielsen and Chuang (2010) komputasi kuantum adalah cabang ilmu komputer dan fisika kuantum yang memanfaatkan fenomena mekanika kuantum seperti superposisi dan entanglement untuk melakukan operasi komputasi. Komputer kuantum melakukan komputasi kuantum, yang merupakan jenis komputasi yang mengumpulkan berbagai keadaan qubit, seperti superposisi, interferensi, dan keterikatan, untuk melakukan perhitungan (Harrow, Hassidim, & Lloyd, 2009). Berbeda dengan komputasi klasik yang menggunakan bit sebagai unit dasar informasi, yang hanya dapat berada dalam salah satu dari dua keadaan 0 atau 1, komputasi kuantum menggunakan qubit, yang dapat berada dalam superposisi dari kedua keadaan tersebut terlihat pada Gambar 2.1.



Gambar 2.1 Ilustrasi Bit dan Qubit
 Sumber: (Gill, Kumar, Singh, Singh, Kaur, Usman, & Buyya, 2022)

Berdasarkan Gambar 2.1 hal ini memungkinkan komputer kuantum untuk memproses dan menganalisis jumlah data yang jauh lebih besar dengan lebih cepat dibandingkan dengan komputer klasik tercepat yang ada saat ini. Superposisi qubit menyediakan akses ke ruang komputasi yang sangat besar yang dapat menyelesaikan banyak masalah dengan kompleksitas komputasi yang besar. Sebagai contoh, angka 3-bit pada waktu tertentu hanya dapat memiliki satu nilai dari himpunan delapan nilai yang mungkin (000, 001, 010, 011, 100, 101, 110, 111). Namun, keadaan 3 qubit dapat ditempatkan dalam superposisi dari seluruh delapan nilai tersebut (Gill et al., 2022)

Keuntungan utama dari komputasi kuantum terletak pada kemampuannya untuk menyelesaikan tugas-tugas tertentu, seperti faktorisasi bilangan besar dan pencarian *database*, dengan lebih efisien daripada metode komputasi klasik.

2.5 Supplier

Menurut Putri, Qadriah, Salat and Razi (2023) *Supplier* atau bisa disebut juga pemasok merupakan suatu perusahaan atau individu yang menyediakan sumber daya yang dibutuhkan oleh perusahaan dan para pesaing untuk memproduksi barang atau jasa tertentu. Pemasok secara intensif sangat mendukung proses

operasi perusahaan, biasanya dalam bentuk bahan baku yang belum jadi, sehingga kualitas dari pemasok dapat dilihat dari produk akhir yang nantinya akan dijual oleh perusahaan untuk pelanggan. Harga yang diberikan oleh pemasok memiliki dampak pada biaya produksi dan akan berdampak pada harga yang akan diberikan kepada pelanggan. Dalam konsep *supply chain management*, terdapat elemen yang penting dan salah satunya adalah bagian pemasok dimana pemasok memiliki peran penting dalam kelangsungan hidup suatu perusahaan. Dimana pemasok menjadi pihak yang menyediakan bahan baku mentah bagi perusahaan, apabila pemasok kurang bertanggung jawab dalam melakukan penyediaan bahan baku dari permintaan perusahaan. Masalah yang akan dihadapi perusahaan jika mengalami keterlambatan pasokan bahan baku bahkan akan kehabisan bahan baku dan mengakibatkan perusahaan mengalami kerugian.

Berdasarkan uraian diatas, dapat disimpulkan bahwa pemasok merupakan elemen yang sangat penting bagi perusahaan dan memiliki pengaruh yang signifikan terhadap proses operasional suatu perusahaan. Oleh karena itu, jika suatu perusahaan memiliki banyak pemasok maka suatu perusahaan diharapkan selektif dalam memilih pemasok. Karena jika salah dalam pemilihan pemasok, maka akan berdampak pada kegiatan operasional perusahaan terutama dalam hal penyediaan bahan baku perusahaan.

2.6 Penelitian Terdahulu

Penelitian terdahulu bertujuan untuk mengetahui perkembangan apa saja yang telah dilakukan pada algoritma DBSCAN. Penelitian tersebut dijadikan acuan dalam penelitian yang dilakukan. Berikut merupakan penelitian terdahulu pada algoritma DBSCAN.

Tabel 2. 1 Penelitian Terdahulu

Peneliti	Judul	Jurnal	Tujuan	Kelebihan	Kelemahan
(Chen et al., 2019)	KNN-BLOCK DBSCAN: Fast Clustering for Large-Scale Data	IEEE transactions on systems, man, and cybernetic: systems	Pengembangan algoritma KNN-BLOCK DBSCAN untuk mengatasi keterbatasan DBSCAN dalam menangani data besar	Algoritma ini dapat mengelompokkan data pada database yang besar	algoritma ini sangat dipengaruhi oleh parameter yang digunakan untuk mendapatkan kluster yang optimal
(Bechini, Marcelloni, & Renda, 2020)	TSF-DBSCAN: a Novel Fuzzy Density-based Approach for Clustering Unbounded Data Streams	IEEE Transactions on Fuzzy Systems	Pengembangan algoritma TSF-DBSCAN untuk pengelompokan data stream	Algoritma ini secara signifikan lebih baik dari algoritma DenStream dan D-Stream.	algoritma ini sangat dipengaruhi oleh parameter yang digunakan dalam data stream untuk mendapatkan kluster yang optimal
(de Moura Ventrone et al., 2021)	BIRCHSCAN: A sampling method for applying DBSCAN to large datasets	Expert Systems with Applications	Pengembangan algoritma BIRCHSCAN menggunakan metode sampling untuk meningkatkan efisiensi DBSCAN	Algoritma ini meningkatkan efisiensi DBSCAN sehingga dapat mengelompokkan dataset besar	Efektivitas dari algoritma ini sangat bergantung pada pemilihan parameter yang tepat
(Chen, Zhou, Bouguila, Wang, Chen, & Du, 2021)	BLOCK-DBSCAN: Fast clustering for large scale data	Pattern Recognition	Pengembangan algoritma BLOCK-DBSCAN untuk mengatasi keterbatasan DBSCAN dalam memproses data besar	Algoritma ini memiliki akurasi yang sama dengan DBSCAN dengan waktu proses yang lebih cepat	Kinerja pada algoritma ini bergantung pada pemilihan parameter
(Ma, Yang, Li, Chi, & Chen, 2023)	A Faster DBSCAN Algorithm Based on Self-Adaptive Determination of Parameters.	Procedia Computer Science	Menentukan parameter Eps dan Mints secara adaptif dan mengubah cara penjelajahan titik data untuk meningkatkan efisiensi waktu	Algoritma ini meningkatkan efisiensi dalam mengelompokkan data serta mempertahankan akurasi kluster yang terbentuk	Berdasarkan percobaan yang dilakukan terdapat akurasi dari kluster yang terbentuk mengalami penurunan

Peneliti	Judul	Jurnal	Tujuan	Kelebihan	Kelemahan
			proses pada DBSCAN		
Huang dkk 2023 (Huang, Ma, Liu, & Liu, 2023)	GriT-DBSCAN: A Spatial Clustering Algorithm for Very Large Databases	Pattern Recognition	Pengembangan algoritma klasterisasi spasial untuk mengatasi keterbatasan DBSCAN dalam menangani data besar	Algoritma ini meningkatkan kinerja DBSCAN pada dataset berdimensi tinggi dengan tetap mempertahankan akurasi klasterisasi	Algoritma ini tidak efektif jika dimensi dataset rendah
Zhang and Zhou 2023 (Zhang & Zhou, 2023)	WOA-DBSCAN: Application of Whale Optimization Algorithm in DBSCAN Parameter Adaption	IEEE Access	Algoritma ini bertujuan untuk mengatasi sensitivitas DBSCAN terhadap parameter inputnya, Eps dan MinPts. WOA-DBSCAN menentukan rentang parameter berdasarkan distribusi dataset menggunakan koefisien siluet sebagai fungsi objektif.	Hasil eksperimen menunjukkan bahwa WOA-DBSCAN secara signifikan meningkatkan akurasi dan efisiensi klasterisasi dibandingkan dengan DBSCAN tradisional dan algoritma optimasi terkait	Proses iteratif dalam WOA dan perhitungan koefisien siluet menambah beban komputasi, terutama untuk dataset besar. Sehingga algoritma ini menunjukkan penurunan stabilitas ketika diterapkan pada dataset berdimensi tinggi.
(Yang et al., 2022)	An efficient DBSCAN optimized by arithmetic optimization algorithm with opposition-based learning	<i>The journal of supercomputing</i>	Algoritma ini bertujuan untuk melakukan pencarian parameter Eps dan MinPts secara otomatis dan adaptif, dengan tujuan untuk meningkatkan kinerja DBSCAN	Algoritma ini menentukan parameter Eps dan MinPts secara otomatis. Klastering yang dihasilkan memiliki akurasi yang signifikan.	Diperlukan pengujian lebih lanjut pada dataset yang lebih besar, karena pada uji coba baru menggunakan 10 dataset.

Peneliti	Judul	Jurnal	Tujuan	Kelebihan	Kelemahan
(Wang, Ye, Du, Mao, Liu, Wu, & Wang, 2022)	AMD-DBSCAN: An Adaptive Multi-density DBSCAN for datasets of extremely variable density	<i>IEEE 9th International Conference on Data Science and Advanced Analytics (DSAA)</i>	Pengembangan AMD-DBSCAN bertujuan untuk menyesuaikan parameter secara adaptif untuk setiap cluster, sehingga menghasilkan hasil clustering yang lebih akurat dan efisien pada dataset dengan kepadatan yang sangat bervariasi.	Algoritma ini meningkatkan akurasi dan efisiensi dalam clustering dataset multi-kepadatan. Adaptasi parameter otomatis yang menghilangkan kebutuhan untuk penyesuaian parameter secara manual	Implementasi membutuhkan pemahaman yang lebih mendalam mengenai teknik adaptasi parameter

Berdasarkan Tabel 2.1 Penelitian Terdahulu dapat diketahui perkembangan DBSCAN dilakukan untuk meningkatkan kemampuan DBSCAN dalam pengelompokan yang besar, seperti yang dilakukan oleh (Chen et al., 2021; Chen et al., 2019; de Moura Ventrone et al., 2021; Huang et al., 2023) karena salah satu kelemahan dari algoritma DBSCAN yaitu kurang efektif dalam mengelompokkan dataset besar. Oleh karena itu dilakukan pengembangan untuk dapat mengelompokkan dataset besar. Selain itu, perkembangan DBSCAN juga dilakukan untuk dapat mengelompokkan data stream, yang dimana data stream selalu berubah ubah karena data stream bersifat real-time, seperti yang dilakukan oleh Bechini et al. (2020).

Hasil penelitian yang dilakukan oleh Chen et al. (2019) menunjukkan bahwa KNN-BLOCK DBSCAN memiliki akurasi yang tinggi dan performa yang lebih baik dibandingkan varian DBSCAN lainnya, termasuk ρ -approximate DBSCAN dan AnyDBC, terutama dalam kecepatan proses klastering tanpa mengorbankan akurasi. Selain itu, penelitian yang dilakukan oleh de Moura Ventrone et al. (2021) dengan mengembangkan algoritma BIRCHSCAN hasil penelitian menunjukkan bahwa BIRCHSCAN efektif dalam memproses dataset besar dengan hasil yang serupa dengan DBSCAN, tetapi dengan pemrosesan waktu yang lebih cepat. Selain itu, penelitian yang dilakukan oleh Chen et al. (2021)

dengan mengembangkan algoritma BLOCK-DBSCAN menunjukkan bahwa BLOCK-DBSCAN mengungguli varian DBSCAN lainnya seperti NQDBSCAN dan p -approximate DBSCAN dalam hal kecepatan dan akurasi, menjadikannya pendekatan yang menjanjikan untuk analisis dataset besar.

Penelitian yang dilakukan oleh Huang et al. (2021) dengan mengembangkan algoritma GriT-DBSCAN menunjukkan bahwa GriT-DBSCAN memiliki kompleksitas waktu linier terhadap ukuran dataset dan menawarkan kinerja yang lebih baik dibandingkan algoritma DBSCAN klasik. Penelitian yang dilakukan oleh Zhang and Zhou (2023) dengan mengembangkan algoritma WOA-DBSCAN menunjukkan bahwa hasil eksperimen secara signifikan meningkatkan akurasi dan efisiensi kluster dibandingkan dengan DBSCAN klasik. Selain penelitian yang dilakukan oleh Yang et al. (2022) dengan mengembangkan algoritma DBSCAN dengan *Arithmetic Optimization Algorithm with Opposition-Based Learning* bertujuan untuk pencarian Eps dan MinPts secara otomatis dan adaptif. Hasil dari algoritma ini menunjukkan hasil klustering yang lebih tinggi dibandingkan dengan K-means, Spectral Clustering, OPTICS, dan DPC.

Penelitian yang dilakukan oleh Wang et al. (2022) pengembangan algoritma *Adaptive Multi-Density* DBSCAN (AMD-DBSCAN) bertujuan untuk menyesuaikan parameter secara adaptif untuk setiap kluster, sehingga menghasilkan hasil klustering yang lebih akurat dan efisien pada dataset dengan kepadatan yang sangat bervariasi. Hasil dari algoritma ini peningkatan akurasi klustering hingga 24,7% dibandingkan dengan algoritma lain pada dataset multi-kepadatan dan 1,6% pada dataset tunggal. Selain itu, AMD-DBSCAN menunjukkan pengurangan waktu eksekusi rata-rata sebesar 75% dibandingkan dengan algoritma adaptif tradisional.

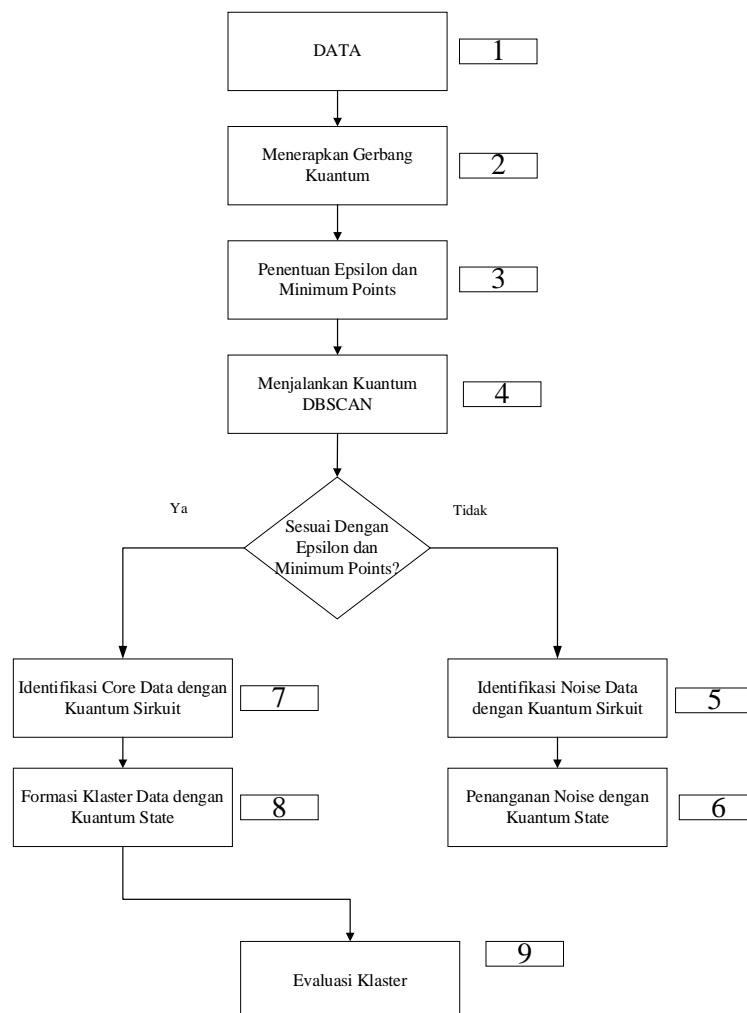
Berdasarkan pada tujuan pengembangan algoritma DBSCAN yang telah dilakukan pada penelitian terdahulu, yaitu untuk menjawab kelemahan dari algoritma DBSCAN yang kurang efektif dalam mengelompokkan dataset besar, peneliti tertarik untuk mengembangkan algoritma DBSCAN dengan memanfaatkan komputasi kuantum untuk mengelompokkan data *supplier*.

BAB III

METODOLOGI PENELITIAN

3.1. Tahapan Penelitian

Penelitian mengenai pengembangan algoritma DBSCAN dengan kuantum terdapat langkah-langkah yang dilakukan, seperti pada Gambar 3.1. Langkah-langkah yang dilakukan diantaranya yaitu, data, menerapkan gerbang kuantum, penentuan Epsilon dan Minimum points, menjalankan Kuantum DBSCAN, hingga evaluasi kluster.



Gambar 3.1 Tahapan Penelitian

Berdasarkan Gambar 3.1 Tahapan Penelitian dapat diketahui langkah-langkah yang dilakukan dalam pengembangan algoritma Kuantum DBSCAN. Langkah awal yang dilakukan yaitu pembuatan data supplier yang berisikan nama supplier, harga, kualitas, dan waktu pengiriman. Tahapan kedua yang dilakukan yaitu menerapkan gerbang kuantum berdasarkan data yang digunakan. Tahapan ketiga yaitu penentuan parameter yang dibutuhkan dalam menjalankan Kuantum DBSCAN. Parameter yang dibutuhkan yaitu terdiri dari *epsilon* dan *minimum poinst*.

Tahapan selanjutnya atau tahapan keempat yaitu menjalankan algoritma Kuantum DBSCAN. Tahapan selanjutnya yaitu terdapat kondisi jika kluster yang terbentuk dalam radius epsilon memiliki jumlah sebaran titik data kurang dari *minimum points*, maka kumpulan data tersebut selanjutnya akan di proses pada tahapan kelima. Pada tahapan kelima yaitu mengidentifikasi noise data dengan kuantum sirkuit. Selanjutnya kumpulan data tersebut yang telah diidentifikasi dikelompokkan ke dalam kluster noise.

Kluster Noise berisikan sekumpulan data yang tidak memenuhi syarat untuk termasuk ke dalam suatu kluster. Sedangkan jika kumpulan data dalam radius epsilon memiliki jumlah sebaran data sama atau lebih dari *minimum points*, maka kumpulan data tersebut akan dilanjutkan pada tahapan ketujuh, yaitu mengidentifikasi titik pusat data dengan kuantum sirkuit. Selanjutnya yaitu formasi kluster data dengan kuantum state. Pada tahapan ini setiap inti data yang memenuhi syarat Epsilon dan *Minimum points* akan membentuk satu kluster dengan setiap kluster yang terbentuk memiliki titik pusat data masing-masing. Selanjutnya yaitu tahap evaluasi kluster. Evaluasi kluster pada penelitian ini menggunakan *silhouette score* untuk mengetahui seberapa baik kluster yang terbentuk, dan menggunakan perhitungan waktu untuk seberapa lama waktu yang dibutuhkan untuk menjalankan algoritma kuantum DBSCAN. Berikut merupakan penjelasan untuk setiap tahapan yang dilakukan dalam penelitian ini.

3.1.1 Data

Tahap awal dalam penelitian diawali dengan pembuatan data, dimana data yang digunakan pada penelitian ini adalah data sintetis. Data sintetis digunakan untuk mendapatkan jumlah data yang besar, selain itu data sintetis juga bersifat fleksibel karena jumlah data yang digunakan dapat ditentukan sesuai dengan kebutuhan pengujian algoritma yang dikembangkan. Data sintetis yang dibuat berisikan nama supplier, harga, kualitas, dan waktu pengiriman. Pembuatan data sintetis dalam penelitian ini menggunakan metode perturbasi. Metode perturbasi dalam pembuatan data sintetis adalah teknik yang mengubah data asli sedemikian rupa sehingga data baru yang dihasilkan sulit dihubungkan kembali dengan sumber aslinya, namun tetap mempertahankan sifat-sifat statistik utama dari data tersebut. Pendekatan metode perturbasi yang digunakan dalam penelitian ini yaitu perturbasi aditif, dengan menambahkan *noise* secara acak ke dalam data asli.

3.1.2 Menerapkan Gerbang Kuantum

Pada tahap ini data yang digunakan akan diterapkan ke dalam gerbang kuantum. Salah satu metode yang dapat digunakan yaitu menerapkan gerbang rotasi R_y untuk melakukan rotasi terhadap qubit dalam rangkaian kuantum. Gerbang R_y mengubah keadaan qubit dengan merotasi amplitudo. Dengan menggunakan nilai input untuk menentukan sudut rotasi, fungsi ini memungkinkan pengaturan presisi terhadap keadaan kuantum, yang sangat penting dalam berbagai algoritma kuantum seperti pencarian kuantum atau simulasi kuantum.

3.1.3 Penentuan Epsilon dan Minimum Points

Pada tahap ini menentukan nilai *Epsilon* (Eps) dan *minimum points* (MinPts) untuk menentukan batas-batas kluster. Epsilon atau Eps digunakan untuk menentukan radius yang menentukan lingkungan di sekitar setiap titik data. Dua titik dianggap bertetangga jika jarak antara mereka kurang dari nilai Eps. *Minimum Points* atau MinPts untuk menentukan jumlah minimum titik yang diperlukan untuk membentuk sebuah kluster.

3.1.4 Menjalankan Kuantum DBSCAN

Pada tahapan ini yaitu menjalankan algoritma Kuantum DBSCAN. Algoritma Kuantum DBSCAN merupakan adaptasi dari algoritma DBSCAN klasikal yang dioptimalkan dalam lingkungan komputasi kuantum. Kuantum DBSCAN dirancang untuk mengatasi keterbatasan DBSCAN dalam menangani dataset besar.

3.1.5 Sesuai dengan Epsilon dan *Minimum Points*

Pada tahapan ini terdapat kondisi jika radius epsilon memiliki jumlah sebaran titik data kurang dari *minimum points*, maka kumpulan data tersebut selanjutnya akan di proses pada tahapan kelima, yaitu identifikasi noise data dengan kuantum sirkuit. Sedangkan jika kumpulan data dalam radius epsilon memiliki jumlah sebaran data sama atau lebih dari *minimum points*, maka kumpulan data tersebut akan dilanjutkan pada tahapan ketujuh yaitu identifikasi *core* data dengan kuantum sirkuit.

3.1.6 Identifikasi *Noise* Data dengan Kuantum Sirkuit

Pada tahapan ini menggunakan rangkaian kuantum untuk mengidentifikasi noise data. Noise data adalah data yang tidak memiliki cukup banyak tetangga yang sesuai dengan *minimum points* dalam radius epsilon yang telah ditentukan. Noise data juga dapat dijelaskan sebagai data yang memiliki jarak tidak cukup dekat atau memiliki jarak yang jauh dengan inti data untuk dianggap bagian dari kluster Suatu data dapat dikatakan *noise* ketika memiliki data kurang dari jumlah *minimum points* dalam radius epsilon.

3.1.7 Penanganan *Noise* dengan *Quantum State*

Pada tahapan ini Mengelola data noise yang telah diidentifikasi menggunakan teknik kuantum untuk memisahkan atau mengelompokkan *noise* secara terpisah. Dalam DBSCAN klasik, *noise* adalah titik data yang tidak

termasuk dalam kluster apa pun. Titik-titik ini tidak memiliki cukup tetangga dalam radius Epsilon (Eps) atau tidak terhubung ke core point.

3.1.8 Identifikasi Core Data dengan Kuantum Sirkuit

Pada tahapan ini menggunakan rangkaian kuantum untuk mengidentifikasi Data inti. Data inti adalah data yang memiliki cukup banyak tetangga yang sesuai dengan *minimum points* dalam radius epsilon yang telah ditentukan. Suatu data dapat di katan inti ketika memiliki data lebih banyak atau sama dengan jumlah minimum points dalam radius epsilon. Data tersebut nantinya akan dikumpulkan menjadi sebuah kluster. Pada setiap kluster yang terbentuk hanya memiliki satu inti data, dan jika terdapat data lain yang menjadi inti data tersebut akan dikelompokkan ke dalam kluster yang lain

3.1.9 Formasi Kluster data dengan Quantum State

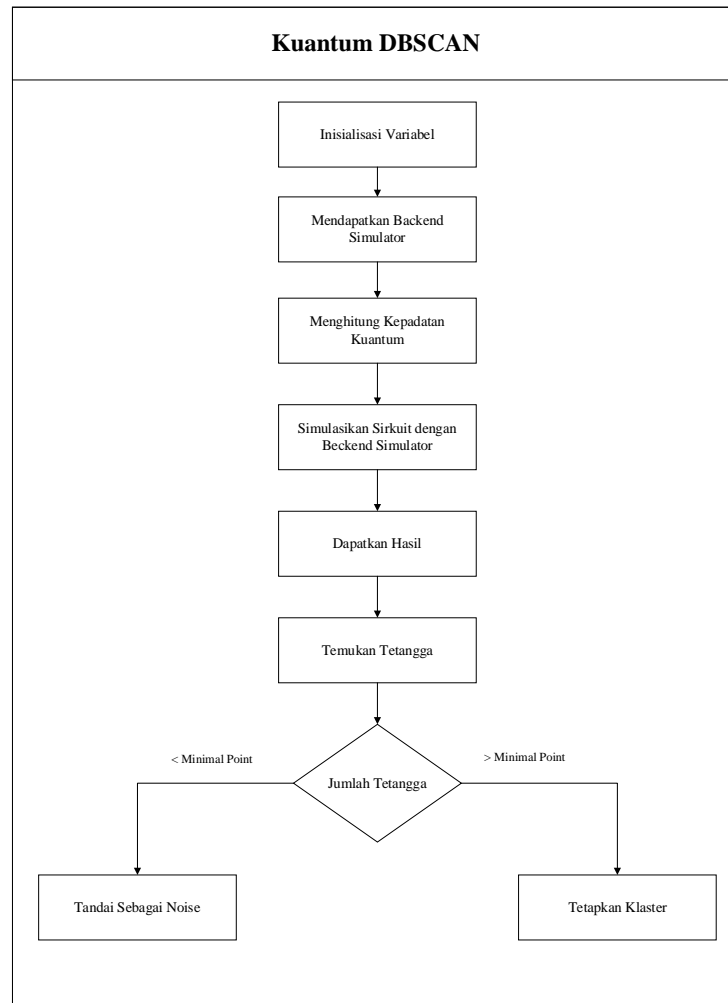
Pada tahapan ini menentukan seberapa banyak kluster yang dihasilkan berdasarkan kepadatan jarak dalam rentang epsilon. Data inti yang memiliki cukup banyak tetangga dalam rentang epsilon dan memiliki jumlah tetangga lebih dari *minimum points* akan dibentuk menjadi kluster. Jumlah kluster yang terbentuk bergantung pada jumlah data inti yang memiliki cukup banyak tetangga untuk dianggap sebagai kluster.

3.1.10 Evaluasi Kluster

Pada tahapan ini mengevaluasi kluster yang terbentuk menggunakan *silhouette score* untuk mengetahui seberapa baik kluster yang terbentuk, dan menggunakan perhitungan waktu untuk seberapa lama waktu yang dibutuhkan untuk menjalankan algoritma kuantum DBSCAN. Nilai pada *silhouette score* memiliki nilai rentang 0 hingga 1, kluster yang terbentuk akan dikatakan baik jika mendekati nilai 1.

3.2. Algoritma Kuantum DBSCAN

Kuantum DBSCAN merupakan adaptasi dari algoritma DBSCAN klasik yang digunakan untuk mengelompokkan data dengan memanfaatkan kemampuan komputasi kuantum. Berikut merupakan langkah-langkah dalam Kuantum DBSCAN.



Gambar 3. 2 Flowchart Kuantum DBSCAN

Berdasarkan Gambar 3.2 *Flowchart* Kuantum DBSCAN menjelaskan mengenai tahapan yang dilakukan dalam menjalankan algoritma Kuantum DBSCAN. Langkah pertama yang dilakukan dalam proses Kuantum DBSCAN adalah inisiasi variabel yang digunakan untuk menjalankan algoritma. Variabel yang dibutuhkan melibatkan data, penentuan parameter seperti epsilon dan *minimum points*, dan penandaan titik-titik data yang sudah dikunjungi. Inisiasi

variabel digunakan untuk menentukan bagaimana data akan di proses lebih lanjut dalam algoritma. Selanjutnya yaitu mendapatkan *backend* simulator. *Backend* simulator berfungsi sebagai platform komputasi kuantum yang mensimulasikan operasi komputer kuantum. Langkah berikutnya adalah menghitung kepadatan kuantum untuk setiap titik data. Proses ini dilakukan dengan membuat sirkuit kuantum yang merepresentasikan setiap titik data dan menjalankannya pada simulator. Hasil dari eksekusi sirkuit ini akan memberikan gambaran mengenai kepadatan di sekitar titik data tersebut, yang kemudian digunakan untuk menentukan apakah titik tersebut adalah inti dari suatu kluster atau tidak. Langkah selanjutnya yaitu mensimulasikan sirkuit dengan *backend* simulator. Pada tahap ini, sirkuit kuantum yang telah dibuat untuk setiap titik data disimulasikan menggunakan *backend* simulator. Proses simulasi ini melibatkan penghitungan jarak antar titik dalam ruang data berdasarkan hasil dari eksekusi kuantum. Proses ini bertujuan untuk mengidentifikasi tetangga dari setiap titik data, yang merupakan inti dari metode klustering Kuantum DBSCAN. Langkah selanjutnya yaitu mendapatkan hasil. Hasil yang didapatkan berupa hasil dari pengukuran sirkuit kuantum yang menunjukkan distribusi titik data dalam radius epsilon. Langkah selanjutnya yaitu menemukan tetangga berdasarkan hasil simulasi kuantum, algoritma menentukan tetangga untuk setiap titik yang sesuai dalam radius epsilon. Jika jumlah tetangga yang ditemukan melebihi jumlah *minimum points* yang telah ditetapkan, maka titik tersebut dapat dianggap sebagai bagian dari sebuah kluster. Titik-titik dengan cukup tetangga ditetapkan ke dalam kluster yang ada, atau digunakan untuk membuat kluster baru. Jika jumlah tetangga kurang dari nilai *minimum points* maka titik data dianggap sebagai *noise*. Dalam konteks DBSCAN, *noise* adalah titik-titik data yang tidak cukup dekat dengan titik lain untuk dianggap sebagai bagian dari kluster mana pun.

3.3. Jadwal Penelitian

Jadwal Penelitian digunakan untuk merencanakan, mengatur, dan mengelola waktu, sumber daya, dan tugas-tugas dalam rangka mencapai tujuan yang telah ditetapkan. Berikut merupakan rencana kegiatan pada penelitian ini.

DAFTAR PUSTAKA

- Bechini, A., Marcelloni, F., & Renda, A. (2020). TSF-DBSCAN: A novel fuzzy density-based approach for clustering unbounded data streams. *IEEE Transactions on Fuzzy Systems*, 30(3), 623-637.
- Chen, Y., Zhou, L., Bouguila, N., Wang, C., Chen, Y., & Du, J. (2021). BLOCK-DBSCAN: Fast clustering for large scale data. *Pattern Recognition*, 109, 107624.
- Chen, Y., Zhou, L., Pei, S., Yu, Z., Chen, Y., Liu, X., Du, J., & Xiong, N. (2019). KNN-BLOCK DBSCAN: Fast clustering for large-scale data. *IEEE transactions on systems, man, and cybernetics: systems*, 51(6), 3939-3953.
- de Moura Ventrorm, I., Luchi, D., Rodrigues, A. L., & Varejão, F. M. (2021). BIRCHSCAN: A sampling method for applying DBSCAN to large datasets. *Expert Systems with Applications*, 184, 115518.
- Deng, D. (2020). DBSCAN clustering algorithm based on density. 2020 7th international forum on electrical engineering and automation (IFEEA),
- Du, J., Meng, C., & Liu, X. (2024). Analysis of Urban Residents' Travelling Characteristics and Hotspots Based on Taxi Trajectory Data. *Applied Sciences*, 14(3), 1279.
- Duan, Y., Liu, C., Li, S., Guo, X., & Yang, C. (2023). An automatic affinity propagation clustering based on improved equilibrium optimizer and t-SNE for high-dimensional data. *Information Sciences*, 623, 434-454.
- Feynman, R. P. (2018). Simulating physics with computers. *Int. j. Theor. phys*, 21(6/7).
- Gill, S. S., Kumar, A., Singh, H., Singh, M., Kaur, K., Usman, M., & Buyya, R. (2022). Quantum computing: A taxonomy, systematic review and future directions. *Software: Practice and Experience*, 52(1), 66-114.
- Hahsler, M., Piekenbrock, M., & Doran, D. (2019). dbscan: Fast density-based clustering with R. *Journal of Statistical Software*, 91, 1-30.

- Han, J., Kamber, M., & Pei, J. (2006). Data mining: concepts and techniques morgan kaufmann [J]. *ISBN, 13*, 978-971.
- Harrow, A. W., Hassidim, A., & Lloyd, S. (2009). Quantum algorithm for linear systems of equations. *Physical review letters*, *103*(15), 150502.
- Huang, X., Ma, T., Liu, C., & Liu, S. (2023). GriT-DBSCAN: A spatial clustering algorithm for very large databases. *Pattern Recognition*, *142*, 109658.
- Huang, Y., Huang, W., Xiang, X., & Yan, J. (2021). An empirical study of personalized advertising recommendation based on DBSCAN clustering of sina weibo user-generated content. *Procedia Computer Science*, *183*, 303-310.
- Ikotun, A. M., Ezugwu, A. E., Abualigah, L., Abuhaija, B., & Heming, J. (2023). K-means clustering algorithms: A comprehensive review, variants analysis, and advances in the era of big data. *Information Sciences*, *622*, 178-210. <https://doi.org/https://doi.org/10.1016/j.ins.2022.11.139>
- Jawad, W. K., & Al-Bakry, A. M. (2023). Big Data Analytics: A Survey. *Iraqi Journal for Computers and Informatics*, *49*(1), 41-51.
- Kurasova, O., Marcinkevicius, V., Medvedev, V., Rapecka, A., & Stefanovic, P. (2014). Strategies for big data clustering. 2014 IEEE 26th international conference on tools with artificial intelligence,
- Leifer, M. S., & Poulin, D. (2008). Quantum graphical models and belief propagation. *Annals of Physics*, *323*(8), 1899-1946.
- Lloyd, S., Mohseni, M., & Rebentrost, P. (2013). Quantum algorithms for supervised and unsupervised machine learning. *arXiv preprint arXiv:1307.0411*.
- Luchi, D., Rodrigues, A. L., & Varejão, F. M. (2019). Sampling approaches for applying DBSCAN to large datasets. *Pattern Recognition Letters*, *117*, 90-96.
- Ma, B., Yang, C., Li, A., Chi, Y., & Chen, L. (2023). A Faster DBSCAN Algorithm Based on Self-Adaptive Determination of Parameters. *Procedia Computer Science*, *221*, 113-120. <https://doi.org/https://doi.org/10.1016/j.procs.2023.07.017>

- Mavi, R. K., Zarbakhshnia, N., Mavi, N. K., & Kazemi, S. (2023). Clustering sustainable suppliers in the plastics industry: A fuzzy equivalence relation approach. *Journal of Environmental Management*, 345, 118811.
- Nielsen, M. A., & Chuang, I. L. (2010). *Quantum computation and quantum information*. Cambridge university press.
- Putri, M. A., Qadriah, L., Salat, J., & Razi, Z. (2023). SISTEM PENDUKUNG KEPUTUSAN DALAM PEMILIHAN SUPPLIER BAHAN BANGUNAN DI TOKO JASA HAI UJONG RIMBA MENGGUNAKAN METODE MOORA. *Jurnal Real Riset*, 5(2), 365-371.
- Sivarajah, U., Kamal, M. M., Irani, Z., & Weerakkody, V. (2017). Critical analysis of Big Data challenges and analytical methods. *Journal of business research*, 70, 263-286.
- Vishwakarma, A., DasGupta, A., & Racherla, V. (2024). Detecting and Locating Track Defects Using Clustering Algorithm. *Available at SSRN 4711483*.
- Wang, J., Li, C., Han, S., Sarkar, S., & Zhou, X. (2017). Predictive maintenance based on event-log analysis: A case study. *IBM Journal of Research and Development*, 61(1), 11: 121-111: 132.
- Wang, Z., Ye, Z., Du, Y., Mao, Y., Liu, Y., Wu, Z., & Wang, J. (2022). AMD-DBSCAN: An Adaptive Multi-density DBSCAN for datasets of extremely variable density. 2022 IEEE 9th International Conference on Data Science and Advanced Analytics (DSAA),
- Yang, Y., Qian, C., Li, H., Gao, Y., Wu, J., Liu, C.-J., & Zhao, S. (2022). An efficient DBSCAN optimized by arithmetic optimization algorithm with opposition-based learning. *The Journal of Supercomputing*, 78(18), 19566-19604.
- Zhang, R., Peng, H., Dou, Y., Wu, J., Sun, Q., Li, Y., Zhang, J., & Yu, P. S. (2022). Automating DBSCAN via deep reinforcement learning. Proceedings of the 31st ACM International Conference on Information & Knowledge Management,
- Zhang, X., & Zhou, S. (2023). WOA-DBSCAN: Application of Whale Optimization Algorithm in DBSCAN Parameter Adaption. *IEEE Access*.