

PENGEMBANGAN ALGORITMA KUANTUM DBSCAN UNTUK
KLASTERISASI SUPPLIER

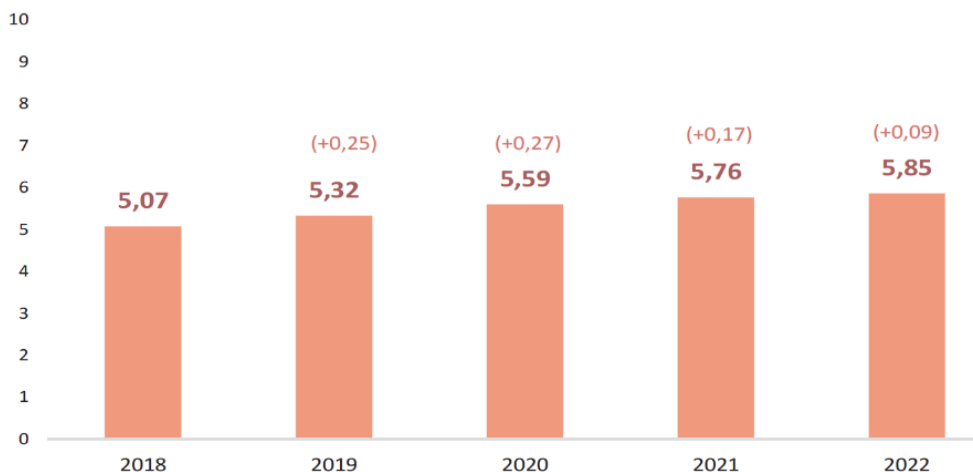
BAB I

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 et al. (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 klasterisasi. Klasterisasi merupakan sebuah metode dari data *mining* yang membagi data ke dalam klaster berdasarkan nilai similaritas antara data. Klasterisasi 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 et al. (2023). Klasterisasi terdapat metode yang dapat digunakan seperti, K-Means Clustering (Ikotun et al., 2023), KNN-Block DBSCAN (Chen et al., 2019), DBSCAN (Huang et al., 2021), Affinity Propagation (Duan et al., 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 pemisah antar kluster. 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 paradigma yang paling populer. Salah satu algoritma klustering yang digunakan untuk mengklusterisasi 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 et al. (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 et al. (2017), dalam bidang transportasi dapat digunakan untuk menganalisis karakteristik perjalanan dan hotspot berdasarkan data trajektori taksi yang dilakukan oleh Du et al. (2024), DBSCAN juga dapat diaplikasikan untuk mengelompokkan dan mengklasifikasikan cacat jalur kereta seperti yang dilakukan oleh Vishwakarma et al. (2024), dan pada bidang industri penerapan DBSCAN digunakan untuk menganalisis data pemeliharaan untuk prediksi kegagalan peralatan, meningkatkan keandalan dan efisiensi operasional (Wang et al., 2017).

Pada penelitian yang dilakukan oleh Zhang et al. (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 sambil menunjukkan variasi yang rendah. Ini juga menunjukkan kompromi yang memuaskan antara kualitas dan efisiensi komputasi, terbukti sangat efisien

dibandingkan dengan pesaing, terutama dalam skenario di mana DBSCAN tidak layak karena kendala waktu atau memori. 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 et al. (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 et al. (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 dan kompleks, 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.

1.3 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.

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 ρ -approximate DBSCAN dalam hal kecepatan dan akurasi, menjadikannya pendekatan yang menjanjikan untuk analisis dataset besar. Selain itu, 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.

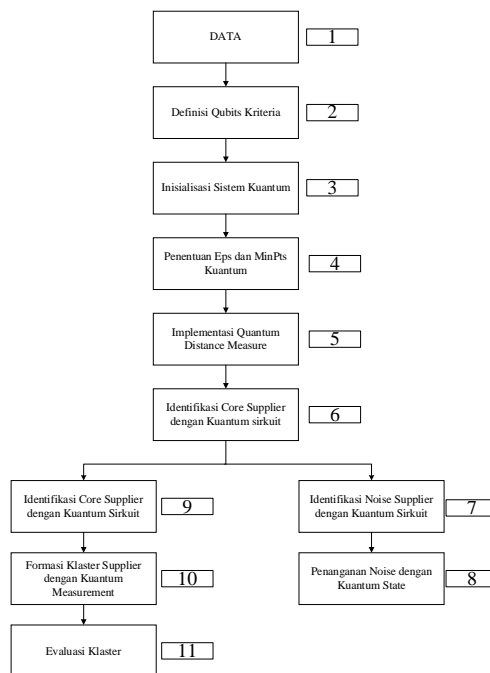
Berdasarkan pada tujuan pengembangan algoritma DBSCAN yang telah dilakukan pada penelitian terdahulu, yaitu untuk menjawab kelemahan dari algoritma DBSCAN yang kurang efektif dalam mengelompokkan dataset besar, peneliti tertarik untuk mengembangkan algoritma DBSCAN dengan memanfaatkan komputasi kuantum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

BAB III

METODOLOGI PENELITIAN

3.1 Tahapan Penelitian

Dalam penelitian mengenai pengembangan algoritma DBSCAN dengan kuantum terdapat langkah-langkah yang dilakukan, seperti pada gambar 3.1. Langkah-langkah yang dilakukan diantaranya yaitu, pengumpulan data, definisi qubits kriteria, inisialisasi sistem kuantum, hingga evaluasi kluster.



Gambar 3.1 Tahapan Penelitian

1. Data

Tahap awal dalam penelitian diawali dengan pembuatan data, dimana data yang digunakan pada penelitian ini adalah data sintetik. Data sintetik digunakan untuk mendapatkan jumlah data yang besar, selain itu data sintetik juga bersifat fleksibel karena jumlah data yang digunakan dapat ditentukan sesuai dengan kebutuhan pengujian algoritma yang dikembangkan. Data

sintetik yang dibuat berisikan nama *supplier*, harga, kualitas, dan waktu pengiriman.

2. Definisi Qubits Kriteria

Pada tahap ini kriteria yang digunakan untuk pengelompokan *supplier* diubah menjadi representasi kuantum menggunakan qubits. Setiap kriteria mungkin diwakili oleh satu atau lebih qubits tergantung pada kompleksitas yang diperlukan. Kriteria yang digunakan dalam pengelompokan *supplier* yaitu, Harga, Kualitas, dan Waktu Pengiriman.

3. Inisialisasi Sistem Kuantum

Pada tahapan ini melakukan persiapan awal dari komputer kuantum yaitu mengatur qubits ke state awal dan memastikan semua qubits berada dalam keadaan awal sebelum operasi kuantum dijalankan. Pada tahapan ini juga menentukan jumlah qubits yang digunakan.

4. Implementasi *Quantum Distance Measure*

Pada tahapan ini melakukan penerapan metode untuk mengukur jarak antar *supplier* dalam ruang kuantum dengan menggunakan prinsip-prinsip mekanika kuantum. Tahapan ini digunakan dalam proses pengelompokan data menggunakan Quantum DBSCAN karena jarak antar *supplier* akan digunakan untuk menentukan klaster

5. Penentuan Eps dan MinPts Kuantum

Pada tahap ini menentukan nilai nilai Epsilon atau Eps dan *minimum points* (MinPts) dalam konteks kuantum untuk menentukan batas-batas klaster. 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 klaster.

6. Identifikasi Core *Supplier* dengan Kuantum Sirkuit

Pada tahapan ini menggunakan rangkaian kuantum untuk mengidentifikasi *supplier* ini (*core supplier*). *Supplier* inti adalah *supplier* yang memiliki

cukup banyak tetangga yang sesuai dengan minpts dalam radius epsilon yang telah ditentukan.

7. Identifikasi Noise *Supplier* dengan Kuantum Sirkuit

Pada tahapan ini mengidentifikasi *supplier* noise atau *outlier* yang memiliki jarak tidak cukup dekat atau memiliki jarak yang jauh dengan *supplier* lain untuk dianggap bagian dari kluster.

8. Penanganan Noise dengan *Quantum State*

Pada tahapan ini Mengelola *supplier* 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.

9. Identifikasi Core *Supplier* dengan *Quantum Circuit*

Pada tahapan ini mengidentifikasi titik-titik data yang berada dalam jarak Epsilon atau Eps dari titik inti tetapi tidak memiliki cukup tetanga untuk masuk ke dalam kluster dengan menggunakan kuantum sirkuit.

10. Formasi Kluster *Supplier* dengan Quantum Measurement

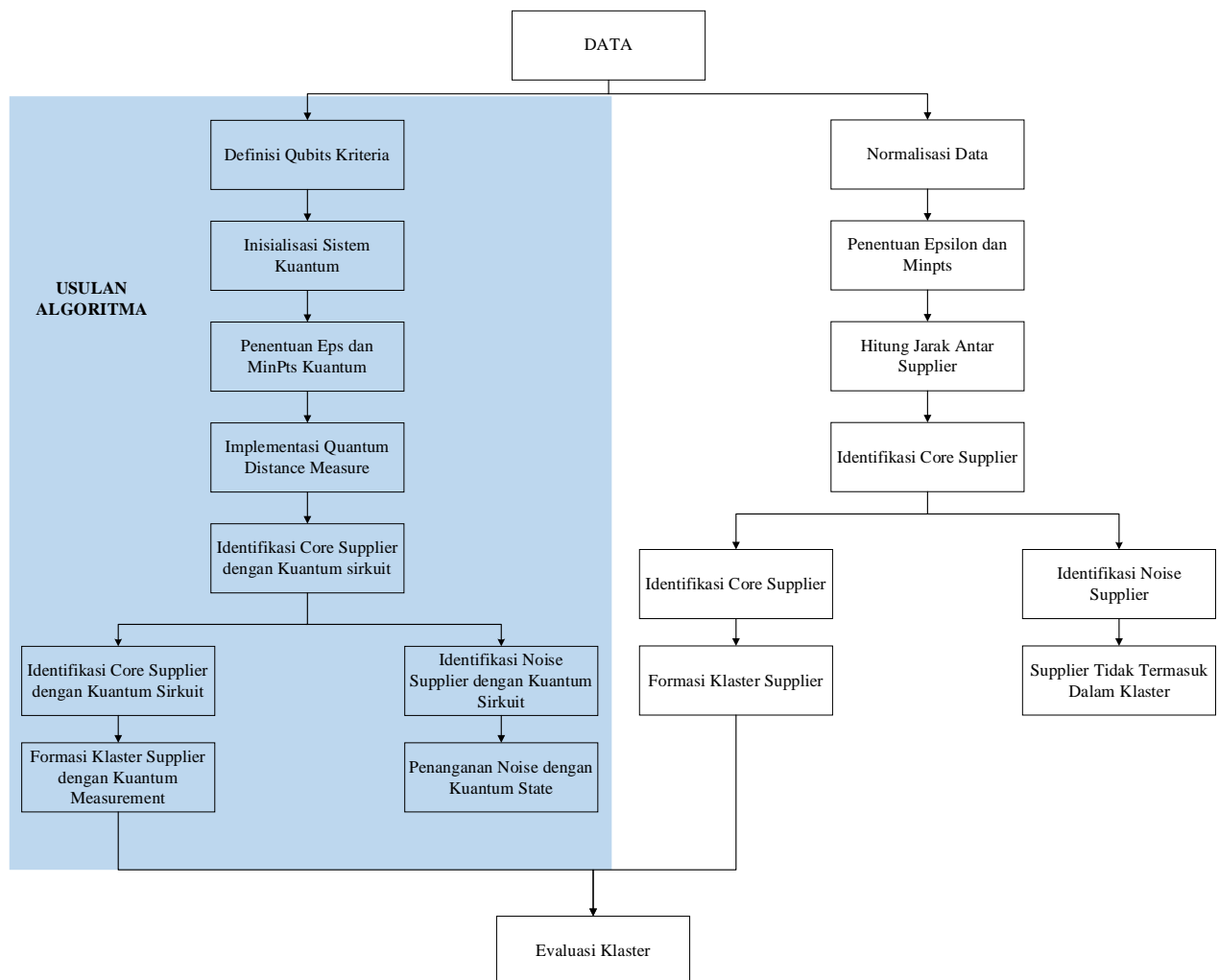
Pada tahapan ini membentuk kluster *supplier* dengan mengukur *state* kuantum yang telah diubah melalui interaksi antar qubits yang mewakili *supplier*.

11. Evaluasi Kluster

Tahap terakhir di mana kualitas dan keefektifan kluster yang terbentuk dievaluasi. Tahapan ini bertujuan untuk menilai seberapa baik kluster yang terbentuk menggunakan.

3.2 Rangkuman Langkah-Langkah Penelitian

Setelah mengembangkan algoritma kuantum DBSCAN selanjutnya membandingkannya dengan algoritma DBSCAN untuk mengetahui seberapa baik algoritma DBSCAN jika dibandingkan dengan algoritma klasiknya. Langkah-langkah tersebut dapat dilihat pada Gambar 3.2 Rangkuman Langkah-Langkah Prosedur Penelitian.



Gambar 3.2 Rangkuman Langkah-Langkah Prosedur Penelitian