

**IMPLEMENTASI ALGORITMA HDBSCAN UNTUK IDENTIFIKASI  
HALTE TRANSJAKARTA DENGAN TINGKAT KEPADATAN  
TERTINGGI BERDASARKAN DATA TRANSAKSI**

**SKRIPSI**

**Oleh:**

**DELA AMELIA SITORUS  
NIM. 221112198**



**PROGRAM STUDI S-1 TEKNIK INFORMATIKA  
FAKULTAS INFORMATIKA  
UNIVERSITAS MIKROSKIL  
MEDAN  
2026**

# **IMPLEMENTATION OF THE HDBSCAN ALGORITHM TO IDENTIFY TRANSJAKARTA BUS STOPS WITH THE HIGHEST DENSITY BASED ON TRANSACTION DATA**

## **UNDERGRADUATE THESIS**

**By:**

**DELA AMELIA SITORUS  
ID NUMBER. 221112198**



**UNDERGRADUATE PROGRAM OF COMPUTER SCIENCE  
FACULTY OF INFORMATICS  
UNIVERSITAS MIKROSKIL  
MEDAN  
2026**

LEMBARAN PENGESAHAN

**IMPLEMENTASI ALGORITMA HDBSCAN UNTUK IDENTIFIKASI  
HALTE TRANSJAKARTA DENGAN TINGKAT KEPADATAN  
TERTINGGI BERDASARKAN DATA TRANSAKSI**

**SKRIPSI**

Diajukan untuk Melengkapi Persyaratan Guna  
Mendapatkan Gelar Sarjana  
Program Studi S-1 Teknik Informatika

Oleh:

**DELA AMELIA SITORUS**  
NIM. 221112198

Disetujui Oleh:

|                                                                                     |                                                                                       |
|-------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------|
| Dosen Pembimbing I,                                                                 | Dosen Pembimbing II,                                                                  |
|  |  |
| <u>Sunarvo Winardi, S.Kom., M.T.</u>                                                | <u>Apriyanto Halim, S.Kom., M.Kom.</u>                                                |

Medan, 22 Juli 2026  
Diketahui dan Disahkan Oleh:

Ketua Program Studi  
S-1 Teknik Informatika,

  
Carles Juliandy, S.Kom., M.Kom.

## HALAMAN PERNYATAAN

Saya yang membuat pernyataan ini adalah mahasiswa Program Studi S-1 Teknik Informatika Universitas Mikroskil Medan dengan identitas mahasiswa sebagai berikut:

Nama : Dela Amelia Sitorus

NIM : 221112198

Saya telah melaksanakan penelitian dan penulisan Skripsi dengan judul dan tempat penelitian sebagai berikut:

Judul Skripsi : Implementasi Algoritma HDBSCAN untuk Identifikasi Halte TransJakarta dengan Tingkat Kepadatan Tertinggi Berdasarkan Data Transaksi

Tempat Penelitian : -

Alamat Tempat Penelitian : -

No. Telp. Tempat Penelitian : -

Sehubungan dengan Skripsi tersebut, dengan ini saya menyatakan dengan sebenar-benarnya bahwa penelitian dan penulisan Skripsi tersebut merupakan hasil karya saya sendiri (tidak menyuruh orang lain yang mengerjakannya) dan semua sumber, baik yang dikutip maupun dirujuk, telah saya nyatakan dengan benar. Bila di kemudian hari ternyata terbukti bahwa bukan saya yang mengerjakannya (membuatnya), maka saya bersedia dikenakan sanksi yang telah ditetapkan oleh Universitas Mikroskil Medan, yakni pencabutan ijazah yang telah saya terima dan ijazah tersebut dinyatakan tidak sah.

Selain itu, demi pengembangan ilmu pengetahuan, saya menyetujui untuk memberikan kepada Universitas Mikroskil Medan Hak Bebas Royalti Non-eksklusif (Non-exclusive Royalty Free Right) atas Skripsi saya beserta perangkat yang ada (jika diperlukan). Dengan hak ini, Universitas Mikroskil Medan berhak menyimpan, mengalihmedia/formatkan, mengelola dalam bentuk pangkalan data (database), merawat, dan mempublikasikan Skripsi saya, secara keseluruhan atau hanya sebagian atau hanya ringkasannya saja dalam bentuk format tercetak dan/atau elektronik, selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik hak cipta. Menyatakan juga bahwa saya akan mempertahankan hak eksklusif saya untuk menggunakan seluruh atau sebagian isi Skripsi saya guna pengembangan karya di masa depan, misalnya dalam bentuk artikel, buku, ataupun perangkat lunak/sistem informasi.

Demikian pernyataan ini saya perbuat dengan sungguh-sungguh, dalam keadaan sadar dan tanpa ada tekanan dari pihak manapun.

Medan, 1 Juli 2026

Saya yang membuat pernyataan,



Dela Amelia Sitorus

# IMPLEMENTASI ALGORITMA HDBSCAN UNTUK IDENTIFIKASI HALTE TRANSJAKARTA DENGAN TINGKAT KEPADATAN TERTINGGI BERDASARKAN DATA TRANSAKSI

## Abstrak

TransJakarta merupakan sistem Bus Rapid Transit (BRT) di Jakarta yang menghasilkan data transaksi penumpang dalam jumlah besar sehingga dapat dimanfaatkan untuk mengidentifikasi pola kepadatan halte sebagai dasar pengambilan keputusan operasional. Penelitian ini bertujuan mengidentifikasi halte dengan tingkat kepadatan transaksi tertinggi menggunakan algoritma Hierarchical Density-Based Spatial Clustering of Applications with Noise (HDBSCAN). Metode penelitian mengacu pada Knowledge Discovery in Databases (KDD) yang meliputi seleksi data, praproses, transformasi, penerapan algoritma, dan evaluasi hasil. Dataset yang digunakan berupa data sintetik (dummy) transaksi TransJakarta dari Kaggle sebanyak 4.000.200 baris, yang setelah praproses menjadi 3.995.020 baris data bersih dengan tujuh fitur. Parameter optimal diperoleh melalui grid search dengan nilai *min\_cluster\_size* 150 dan *min\_samples* 200, menghasilkan nilai *Silhouette Coefficient* sebesar 0,5309. Hasil clustering membagi halte ke dalam empat klaster, yaitu rendah, sedang, tinggi, dan sangat tinggi. Halte GBK Pintu 7, BNN LRT, dan Jln. Swadaya Raya termasuk dalam klaster dengan tingkat kepadatan transaksi tertinggi. Temuan ini menunjukkan bahwa HDBSCAN berpotensi mendukung identifikasi kepadatan transaksi sebagai dasar perencanaan operasional dan pemerataan layanan TransJakarta berbasis data.

**Kata kunci:** *clustering, HDBSCAN, TransJakarta, kepadatan transaksi, KDD.*

## Abstract

TransJakarta is a Bus Rapid Transit (BRT) system in Jakarta that generates a large volume of passenger transaction data, which can be used to identify bus stop density patterns as a basis for operational decision-making. This study aims to identify bus stops with the highest transaction density using the Hierarchical Density-Based Spatial Clustering of Applications with Noise (HDBSCAN) algorithm. The research methodology is based on Knowledge Discovery in Databases (KDD), which includes data selection, preprocessing, transformation, algorithm application, and result evaluation. The dataset used consists of synthetic (dummy) TransJakarta transaction data from Kaggle, totaling 4,000,200 rows, which, after preprocessing, resulted in 3,995,020 rows of clean data with seven features. Optimal parameters were obtained through a grid search with *min\_cluster\_size* set to 150 and *min\_samples* set to 200, yielding a *Silhouette Coefficient* of 0.5309. The clustering results divided the bus stops into four clusters: low, medium, high, and very high. The GBK Gate 7, BNN LRT, and Jln. Swadaya Raya stops are included in the cluster with the highest transaction density. These findings indicate that HDBSCAN has the potential to support the identification of transaction density as a basis for data-driven operational planning and the equitable distribution of TransJakarta services.

**Keywords:** *clustering, HDBSCAN, TransJakarta, transaction density, KDD.*

## KATA PENGANTAR

Puji syukur penulis panjatkan ke hadirat Tuhan Yang Maha Esa atas segala berkat dan anugerah-Nya, sehingga penulis dapat menyelesaikan skripsi yang berjudul **"Implementasi Algoritma HDBSCAN untuk Identifikasi Halte TransJakarta dengan Tingkat Kepadatan Tertinggi Berdasarkan Data Transaksi"** ini dengan baik dan tepat waktu.

Skripsi ini disusun sebagai salah satu syarat untuk memperoleh gelar Sarjana Komputer pada Program Studi S-1 Teknik Informatika, Fakultas Informatika, Universitas Mikroskil Medan. Penelitian ini dilakukan karena tingginya volume transaksi penumpang TransJakarta menyebabkan tingkat kepadatan yang berbeda-beda di setiap halte, sehingga diperlukan sebuah pendekatan berbasis data untuk mengidentifikasi halte dengan kepadatan tertinggi secara lebih objektif dan terukur, yang diharapkan dapat mendukung pengambilan keputusan operasional TransJakarta ke depannya.

Dalam proses penyusunan skripsi ini, penulis menyadari bahwa skripsi ini tidak dapat diselesaikan dengan baik tanpa adanya bimbingan, bantuan, dukungan, serta doa dari berbagai pihak. Oleh karena itu, pada kesempatan ini penulis ingin menyampaikan ucapan terima kasih yang sebesar-besarnya kepada:

1. Bapak Sunaryo Winardi, S.Kom., M.T., yang telah meluangkan waktu untuk membimbing serta memberikan masukan dan saran selama proses penyusunan skripsi ini, selaku Dosen Pembimbing I sekaligus Dekan Fakultas Informatika Universitas Mikroskil Medan.
2. Bapak Apriyanto Halim, S.Kom., M.Kom., yang juga telah meluangkan waktu untuk membimbing serta memberikan masukan dan saran selama proses penyusunan skripsi ini, selaku Dosen Pembimbing II.
3. Bapak Hardy, S.Kom., M.Sc., Ph.D., selaku Rektor Universitas Mikroskil Medan.
4. Bapak Carles Juliandy, S.Kom., M.Kom., selaku Ketua Program Studi S-1 Teknik Informatika, Fakultas Informatika, Universitas Mikroskil Medan.
5. Seluruh Bapak/Ibu Dosen Universitas Mikroskil yang telah mendidik dan memberikan bimbingan kepada penulis selama kuliah di Universitas Mikroskil.
6. Kedua orang tua yang tiada hentinya melangitkan doa-doa terbaik serta senantiasa memberikan dukungan kepada anak perempuannya ini dalam memperjuangkan masa depan. Dengan penuh kasih sayang, pengorbanan, dan kerja keras, mereka selalu berusaha memberikan yang terbaik bagi borunya agar dapat meraih cita-cita dan masa

depan yang lebih baik. Terima kasih atas segala perjuangan, kesabaran, dan cinta yang tulus, sehingga penulis dapat melewati setiap proses dan akhirnya menyelesaikan pendidikan ini.

7. Seluruh teman-teman seperjuangan yang selalu memberikan semangat, bantuan, dan motivasi selama proses penyusunan skripsi ini.
8. Anak perempuan tunggal yang menjadi harapan terbesar bagi kedua orang tuanya, Dela Amelia Sitorus. Ya, diri saya sendiri. Apresiasi sebesar-besarnya untuk diri sendiri yang telah bertahan dan terus berjuang hingga mampu sampai di titik ini. Terima kasih telah memilih untuk tetap melangkah, bekerja keras, dan tidak menyerah meskipun perjalanan yang dilalui tidak selalu mudah. Terima kasih untuk setiap malam yang dilewati dalam kelelahan, setiap pagi yang disambut dengan kekhawatiran, namun tetap dijalani dengan penuh harapan hingga akhirnya berhasil dilalui. Terima kasih kepada hati yang tetap ikhlas meskipun banyak hal terjadi di luar prediksi. Teruslah belajar, tetap rendah hati, dan senantiasa mensyukuri setiap nikmat yang Tuhan berikan. Tetap semangat dalam meraih setiap impian, dan semoga kebahagiaan selalu menyertaimu di mana pun berada. Semoga setiap langkah kecilmu senantiasa diperkuat oleh Tuhan, dikelilingi orang-orang baik yang tulus, serta setiap doa dan mimpimu dijawab pada waktu yang terbaik. Amin.

Penulis menyadari bahwa skripsi ini masih memiliki berbagai keterbatasan dan belum sepenuhnya sempurna. Hal tersebut disebabkan oleh keterbatasan pengetahuan, pengalaman, serta kemampuan yang dimiliki penulis. Oleh karena itu, penulis sangat mengharapkan kritik dan saran yang membangun dari berbagai pihak sebagai bahan evaluasi untuk penyempurnaan karya ini. Semoga skripsi ini dapat memberikan manfaat, menambah wawasan, serta menjadi referensi yang berguna bagi pembaca maupun pihak-pihak yang membutuhkan.

Medan, 1 Juli 2026

Dela Amelia Sitorus

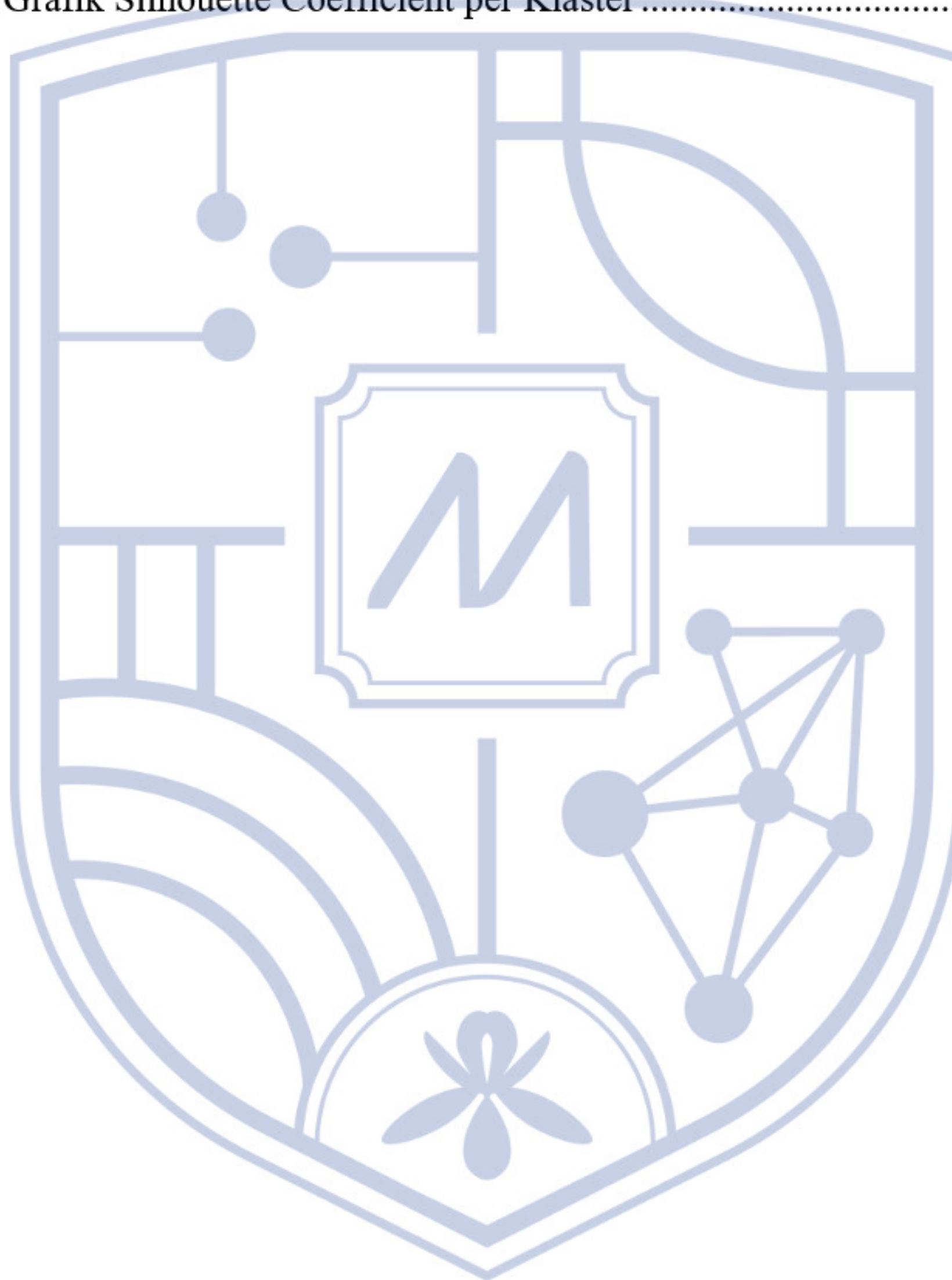
## DAFTAR ISI

|                                                                                              |             |
|----------------------------------------------------------------------------------------------|-------------|
| <b><u>Abstrak</u></b> .....                                                                  | <b>i</b>    |
| <b><u>KATA PENGANTAR</u></b> .....                                                           | <b>ii</b>   |
| <b>DAFTAR ISI</b> .....                                                                      | <b>iv</b>   |
| <b>DAFTAR GAMBAR</b> .....                                                                   | <b>vi</b>   |
| <b>DAFTAR TABEL</b> .....                                                                    | <b>vii</b>  |
| <b>DAFTAR LAMPIRAN</b> .....                                                                 | <b>viii</b> |
| <b>BAB I PENDAHULUAN</b> .....                                                               | <b>1</b>    |
| 1.1 Latar Belakang .....                                                                     | 1           |
| 1.2 Rumusan Masalah .....                                                                    | 2           |
| 1.3 Tujuan .....                                                                             | 3           |
| 1.4 Manfaat .....                                                                            | 3           |
| 1.5 Ruang Lingkup .....                                                                      | 3           |
| <b>BAB II KAJIAN LITERATUR</b> .....                                                         | <b>5</b>    |
| 2.1 Knowledge Discovery in Databases (KDD) .....                                             | 5           |
| 2.2 Data Mining .....                                                                        | 6           |
| 2.3 Clustering .....                                                                         | 7           |
| 2.4 Praproses Data .....                                                                     | 8           |
| 2.5 HDBSCAN (Hierarchical Density-Based Spatial Clustering of Applications with Noise) ..... | 8           |
| 2.6 Evaluasi Clustering .....                                                                | 11          |
| 2.7 Sistem Transportasi .....                                                                | 12          |
| 2.8 Konsep Kepadatan Penumpang .....                                                         | 13          |
| 2.9 Faktor-Faktor yang Mempengaruhi Kepadatan Penumpang .....                                | 14          |
| 2.10 TransJakarta .....                                                                      | 15          |
| 2.11 Kerangka Pemikiran .....                                                                | 17          |
| <b>BAB III TAHAPAN PELAKSANAAN</b> .....                                                     | <b>19</b>   |
| 3.1 Gambaran Umum Tahapan Penelitian .....                                                   | 19          |
| 3.2 Seleksi Data .....                                                                       | 21          |
| 3.3 Praproses Data .....                                                                     | 22          |
| 3.4 Transformasi Data .....                                                                  | 24          |
| 3.5 Cara Kerja Algoritma HDBSCAN .....                                                       | 31          |
| 3.6 Pengujian .....                                                                          | 43          |

|                                                               |           |
|---------------------------------------------------------------|-----------|
| <b>BAB IV HASIL DAN PEMBAHASAN.....</b>                       | <b>44</b> |
| 4.1 Implementasi Algoritma HDBSCAN .....                      | 44        |
| 4.1.1 Hasil Seleksi Data .....                                | 45        |
| 4.1.2 Hasil Praproses Data .....                              | 46        |
| 4.1.3 Hasil Transformasi .....                                | 47        |
| 4.1.4 Hasil Grid Search Parameter HDBSCAN .....               | 48        |
| 4.1.5 Hasil Clustering HDBSCAN dengan Parameter Optimal ..... | 49        |
| 4.1.6 Hasil Evaluasi Silhouette Coefficient .....             | 51        |
| 4.2 Pembahasan .....                                          | 52        |
| 4.2.1 Karakteristik Klaster Hasil HDBSCAN .....               | 52        |
| 4.2.2 Identifikasi Halte dengan Kepadatan Tertinggi .....     | 54        |
| <b>BAB V PENUTUP .....</b>                                    | <b>55</b> |
| 5.1 Kesimpulan .....                                          | 55        |
| 5.2 Saran .....                                               | 55        |
| <b>DAFTAR PUSTAKA .....</b>                                   | <b>57</b> |
| <b>DAFTAR RIWAYAT HIDUP.....</b>                              | <b>64</b> |
| <b>LAMPIRAN.....</b>                                          | <b>65</b> |
| Lampiran A. Dataset Penelitian.....                           | 65        |
| Lampiran B. Hasil Seleksi Data.....                           | 66        |
| Lampiran C. Dataset Penelitian.....                           | 66        |
| Lampiran D. Hasil Transformasi Data.....                      | 67        |
| Lampiran E. Hasil Grid Search.....                            | 68        |
| Lampiran F. Hasil Clustering.....                             | 69        |
| Lampiran G. Hasil Evaluasi SC.....                            | 69        |
| Lampiran H. Perhitungan Manual.....                           | 69        |
| H.1 Perhitungan Normalisasi Min-Max.....                      | 69        |
| H.2 Euclidean Distance .....                                  | 70        |
| H.3 Perhitungan Mutual Reachability Distance .....            | 70        |

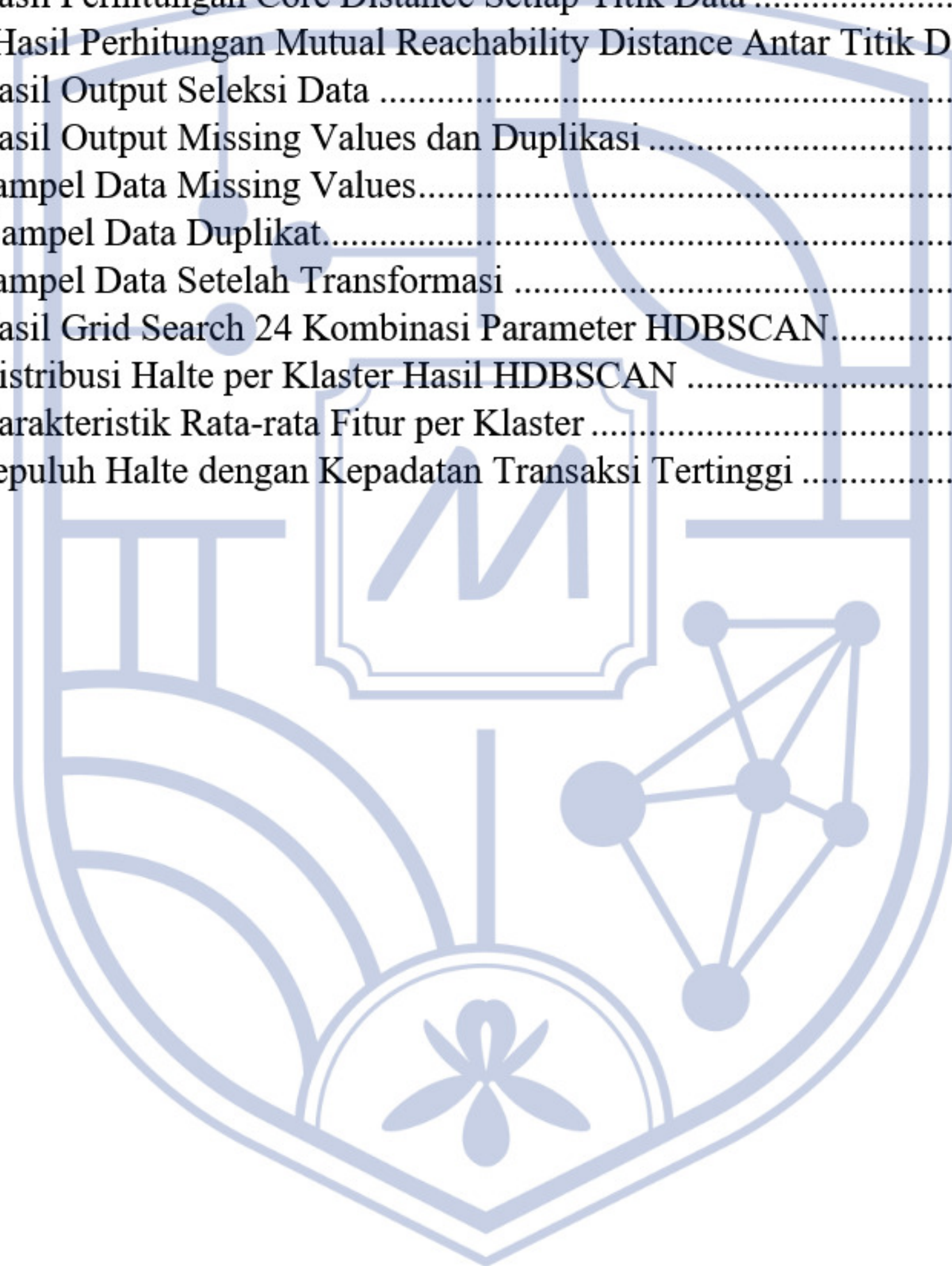
## DAFTAR GAMBAR

|                                                                                 |    |
|---------------------------------------------------------------------------------|----|
| Gambar 2. 1 Proses Knowledge Discovery in Databases (KDD)[20].....              | 5  |
| Gambar 2. 2 Ilustrasi Hasil Clustering Menggunakan Algoritma HDBSCAN [40] ..... | 11 |
| Gambar 2. 3 Bus Listrik TransJakarta Koridor 1 Blok M–Kota[66].....             | 16 |
| Gambar 2. 4 Mekanisme Tap-On-Bus (TOB) TransJakarta[67] .....                   | 17 |
| Gambar 3. 1 Metode Tahapan Penelitian .....                                     | 20 |
| Gambar 3. 2 Flowchart Cara Kerja Algoritma HDBSCAN .....                        | 31 |
| Gambar 3. 3 Flowchart Implementasi Algoritma HDBSCAN.....                       | 42 |
| Gambar 4. 1 Hasil Clustering Halte TransJakarta Menggunakan Algoritma HDBSCAN.. | 49 |
| Gambar 4. 2 Grafik Silhouette Coefficient per Kluster .....                     | 51 |



## DAFTAR TABEL

|                                                                                   |    |
|-----------------------------------------------------------------------------------|----|
| Tabel 3. 1 Fitur yang Digunakan dalam Clustering.....                             | 22 |
| Tabel 3. 2 Hasil Output Missing Values dan Duplikasi .....                        | 23 |
| Tabel 3. 3 Sampel Data Missing Values.....                                        | 23 |
| Tabel 3. 4 Sampel Data Duplikat.....                                              | 23 |
| Tabel 3. 5 Sampel Data Sebelum di Transformasi .....                              | 24 |
| Tabel 3. 6 Sampel Data Setelah Transformasi .....                                 | 31 |
| Tabel 3. 7 Fitur Perhitungan Jarak.....                                           | 33 |
| Tabel 3. 8 Ringkasan seluruh jarak: .....                                         | 36 |
| Tabel 3. 9 Hasil Perhitungan Core Distance Setiap Titik Data .....                | 37 |
| Tabel 3. 10 Hasil Perhitungan Mutual Reachability Distance Antar Titik Data ..... | 39 |
| Tabel 4. 1 Hasil Output Seleksi Data .....                                        | 45 |
| Tabel 4. 2 Hasil Output Missing Values dan Duplikasi .....                        | 46 |
| Tabel 4. 3 Sampel Data Missing Values.....                                        | 46 |
| Tabel 4. 4 Sampel Data Duplikat.....                                              | 47 |
| Tabel 4. 5 Sampel Data Setelah Transformasi .....                                 | 47 |
| Tabel 4. 6 Hasil Grid Search 24 Kombinasi Parameter HDBSCAN.....                  | 48 |
| Tabel 4. 7 Distribusi Halte per Kluster Hasil HDBSCAN .....                       | 50 |
| Tabel 4. 8 Karakteristik Rata-rata Fitur per Kluster .....                        | 52 |
| Tabel 4. 9 Sepuluh Halte dengan Kepadatan Transaksi Tertinggi .....               | 54 |



## DAFTAR LAMPIRAN

|                                                            |    |
|------------------------------------------------------------|----|
| Tabel A. 1 Sampel Dataset Transaksi TransJakarta .....     | 65 |
| Tabel A. 2 Sampel Data Hari Libur Nasional .....           | 65 |
| Tabel A. 3 Sampel Data Curah Hujan .....                   | 66 |
| Tabel B. 1 Hasil Seleksi Data .....                        | 66 |
| Tabel C. 1 Hasil Output Missing Values dan Duplikasi ..... | 66 |
| Tabel C. 2 Sampel Data Missing Values .....                | 66 |
| Tabel C. 3 Sampel Data Duplikat .....                      | 67 |
| Tabel D. 1 Sampel Data Sebelum Normalisasi .....           | 67 |
| Tabel D. 2 Setelah Transformasi .....                      | 67 |
| Tabel E. 1 Hasil Grid Search .....                         | 68 |
| Gambar F. 1 Hasil Clustering .....                         | 69 |
| Gambar G.1 Hasil Grafik .....                              | 69 |

