



SKRIPSI

KLASTERISASI INDUSTRI KECIL MENENGAH (IKM) DI KOTA SURABAYA MENGGUNAKAN METODE *BALANCED ITERATIVE REDUCING AND CLUSTERING USING HIERARCHIES* (BIRCH)

FITRI INDAH SARI
NPM 210830102025

DOSEN PEMBIMBING
Dr. Ir. Mohammad Idhom, S.P., S.Kom., M.T.
Trimono, S.Si, M.Si

**KEMENTERIAN PENDIDIKAN TINGGI, SAINS, DAN TEKNOLOGI
UNIVERSITAS PEMBANGUNAN NASIONAL VETERAN JAWA TIMUR
FAKULTAS ILMU KOMPUTER
PROGRAM STUDI SAINS DATA
SURABAYA
2025**



SKRIPSI

KLASTERISASI INDUSTRI KECIL MENENGAH (IKM) DI KOTA SURABAYA MENGGUNAKAN METODE *BALANCED ITERATIVE REDUCING AND CLUSTERING USING HIERARCHIES* (BIRCH)

FITRI INDAH SARI
NPM 21083010025

DOSEN PEMBIMBING

Dr. Ir. Mohammad Idhom, S.P., S.Kom., M.T.
Trimono, S.Si, M.Si

**KEMENTERIAN PENDIDIKAN, KEBUDAYAAN, RISET, DAN TEKNOLOGI
UNIVERSITAS PEMBANGUNAN NASIONAL VETERAN JAWA TIMUR
FAKULTAS ILMU KOMPUTER
PROGRAM STUDI SAINS DATA
SURABAYA
2025**



SKRIPSI

**KLASTERISASI INDUSTRI KECIL MENENGAH
(IKM) DI KOTA SURABAYA MENGGUNAKAN
METODE *BALANCED ITERATIVE REDUCING
AND CLUSTERING USING HIERARCHIES*
(BIRCH)**

FITRI INDAH SARI
NPM 21083010025

DOSEN PEMBIMBING
Dr. Ir. Mohammad Idhom, S.P., S.Kom., M.T.
Trimono, S.Si, M.Si

**KEMENTERIAN PENDIDIKAN TINGGI, SAINS, DAN TEKNOLOGI
UNIVERSITAS PEMBANGUNAN NASIONAL VETERAN JAWA TIMUR
FAKULTAS ILMU KOMPUTER
PROGRAM STUDI SAINS DATA
SURABAYA
2025**

LEMBAR PENGESAHAN

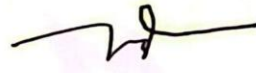
KLASTERISASI INDUSTRI KECIL MENENGAH (IKM) DI KOTA SURABAYA MENGGUNAKAN METODE *BALANCED ITERATIVE REDUCING AND CLUSTERING USING HIERARCHIES* (BIRCH)

Oleh :
FITRI INDAH SARI
NPM. 21083010025

Telah dipertahankan di hadapan dan diterima oleh Tim Penguji Sidang Skripsi Program Studi Sains Data Fakultas Ilmu Komputer Universitas Pembangunan Nasional Veteran Jawa Timur Pada tanggal 05 Juni 2025:

Menyetujui,

Dr. Ir. Mohammad Idhom, S.P., S.Kom., M.T.
NIP. 19830310 202121 1 006



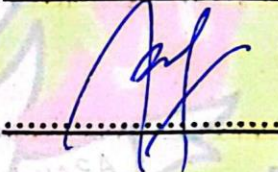
(Pembimbing I)

Trimono, S.Si., M.Si
NIP. 19950908 202203 1 003



(Pembimbing II)

Dr. Eng. Ir. Dwi Arman Prasetya, S.T., M.T.,
IPU., Asean, Eng.
NIP. 19801205 200501 1 002



(Ketua Penguji)

Alfan Rizaldy Pratama, S.Tr.T., M.Tr.Kom.
NIP. 19990606 202406 1 001



(Penguji I)

Mengetahui,
Dekan Fakultas Ilmu Komputer


Prof. Dr. Ir. Novirina Hendrasarie, MT.
NIP. 19681126 199403 2 001

LEMBAR PERSETUJUAN

**KLASTERISASI INDUSTRI KECIL MENENGAH (IKM) DI KOTA
SURABAYA MENGGUNAKAN METODE *BALANCED ITERATIVE
REDUCING AND CLUSTERING USING HIERARCHIES (BIRCH)***

Oleh:
FITRI INDAH SARI
NPM. 21083010025

Telah disetujui untuk mengikuti Ujian Skripsi



Menyetujui,

**Koordinator Program Studi Sains Data
Fakultas Ilmu Komputer**

Dr. Eng. Ir. Dwi Arman Prasetya, S.T., M.T., IPU., Asean, Eng.
NIP. 19801205 200501 1 002

SURAT PERNYATAAN BEBAS PLAGIASI

Saya yang bertanda tangan di bawah ini:

Nama Mahasiswa : Fitri Indah Sari
NPM : 21083010025
Program : Sarjana (S1)
Program Studi : Sains Data
Fakultas : Fakultas Ilmu Komputer

Menyatakan bahwa dalam dokumen ilmiah Skripsi ini tidak terdapat bagian dari karya ilmiah lain yang telah diajukan untuk memperoleh gelar akademik di suatu lembaga Pendidikan Tinggi, dan juga tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang/lembaga lain, kecuali yang secara tertulis disitasi dalam dokumen ini dan disebutkan secara lengkap dalam daftar pustaka.

Dan saya menyatakan bahwa dokumen ilmiah ini bebas dari unsur-unsur plagiasi. Apabila di kemudian hari ditemukan indikasi plagiat pada Skripsi ini, saya bersedia menerima sanksi sesuai dengan peraturan perundang-undangan yang berlaku.

Demikian surat pernyataan ini saya buat dengan sesungguhnya tanpa ada paksaan dari siapapun juga dan untuk dipergunakan sebagaimana mestinya.



Surabaya, 05 Juni 2025
Yang Membuat Pernyataan,



FITRI INDAH SARI
NPM. 21083010025

ABSTRAK

Nama Mahasiswa / NPM : Fitri Indah Sari / 21083010025
Judul Skripsi : Klasterisasi Industri Kecil Menengah (IKM) di Kota Surabaya Menggunakan Metode *Balanced Iterative Reducing and Clustering Using Hierarchies* (BIRCH)
Dosen Pembimbing : 1. Dr. Ir. Mohammad Idhom, S.P., S.Kom., M.T.
2. Trimono, S.Si., M.Si

Industri Kecil dan Menengah (IKM) memiliki peran strategis dalam mendorong pertumbuhan ekonomi lokal, termasuk di Kota Surabaya. Namun, permasalahan utama yang dihadapi pemerintah daerah adalah belum optimalnya strategi pembinaan dan pengembangan IKM akibat belum adanya segmentasi yang tepat berdasarkan karakteristik usaha. Ketidakterpaduan tersebut menyebabkan alokasi sumber daya dan kebijakan yang bersifat umum dan kurang responsif terhadap kebutuhan spesifik masing-masing kelompok IKM. Oleh karena itu, diperlukan pendekatan analitis yang objektif untuk mengelompokkan IKM berdasarkan atribut usaha guna mendukung pengambilan keputusan yang lebih efektif dan terarah. Penelitian ini bertujuan menerapkan metode klasterisasi *Balanced Iterative Reducing and Clustering using Hierarchies* (BIRCH) dalam mengelompokkan IKM di Surabaya berdasarkan kombinasi variabel numerik dan kategorik, yaitu luas lahan, modal awal, jumlah tenaga kerja, skala usaha, jenis usaha, dan risiko usaha. Pemilihan metode BIRCH didasarkan pada keunggulannya yang efisien dalam menangani dataset berukuran besar dengan struktur data hierarkis. Klasterisasi dilakukan terhadap 31.465 entitas IKM menggunakan enam variabel utama tersebut. Hasil klasterisasi membentuk tiga klaster utama dengan distribusi jumlah anggota sebanyak 24, 211, dan 31.230 entitas. Evaluasi menggunakan *silhouette score* sebesar 0,89 menunjukkan kualitas pemisahan antar klaster yang sangat baik, dengan tingkat kemiripan internal tinggi dan perbedaan signifikan antar kelompok. Selain itu, sistem berbasis web yang diusulkan menggunakan *framework Streamlit* memungkinkan pengguna berinteraksi dengan hasil analisis secara intuitif dan mudah dipahami, sehingga mendukung pengambilan keputusan yang lebih efektif.

Kata kunci : *Balanced Iterative Reducing and Clustering using Hierarchies* (BIRCH), Industri Kecil Menengah (IKM), Klasterisasi, *Silhouette Coefficient*, *Streamlit*

Halaman ini sengaja dikosongkan

ABSTRACT

Student Name / NPM : Fitri Indah Sari / 21083010025
Thesis Title : *Clustering of Small and Medium Enterprises (SMEs) in Surabaya City Using the Balanced Iterative Reducing and Clustering Using Hierarchies (BIRCH) Method*
Advisor : 1. Dr. Ir. Mohammad Idhom, S.P., S.Kom., M.T.
2. Trimono, S.Si., M.Si

Small and Medium Industries (SMIs) have a strategic role in driving local economic growth, including in Surabaya City. However, the main problem faced by the local government is the suboptimal strategy of coaching and development of SMEs due to the absence of proper segmentation based on business characteristics. This lack of integration has led to the allocation of resources and policies that are general in nature and less responsive to the specific needs of each group of SMEs. Therefore, an objective analytical approach is needed to cluster SMEs based on business attributes to support more effective and targeted decision-making. This study aims to apply the Balanced Iterative Reducing and Clustering using Hierarchies (BIRCH) clustering method in grouping SMEs in Surabaya based on a combination of numerical and categorical variables, namely land area, initial capital, number of workers, business scale, business type, and business risk. The selection of BIRCH method is based on its efficient advantage in handling large datasets with hierarchical data structure. Clustering was conducted on 31,465 SME entities using the six main variables. The clustering results formed three main clusters with a distribution of members of 24, 211, and 31,230 entities. Evaluation using a silhouette score of 0.89 indicates an excellent quality of separation between clusters, with a high degree of internal similarity and significant differences between groups. In addition, the proposed web-based system using the Streamlit framework allows users to interact with the analysis results in an intuitive and easy-to-understand manner, thus supporting more effective decision-making.

Keywords: *Balanced Iterative Reducing and Clustering using Hierarchies (BIRCH), Clustering, Silhouette Coefficient, Small and Medium Enterprises (SMEs), Streamlit*

Halaman ini sengaja dikosongkan

KATA PENGANTAR

Puji syukur kehadiran Allah SWT atas segala rahmat, hidayah dan karunia-Nya, sehingga penulis dapat menyelesaikan skripsi yang berjudul “**Klasterisasi Industri Kecil Menengah (IKM) di Kota Surabaya Menggunakan Metode *Balanced Iterative Reducing and Clustering Using Hierarchies (BIRCH)***” secara tepat waktu. Dalam proses penyusunan skripsi ini, penulis menerima banyak dukungan, bimbingan, dan bantuan dari berbagai pihak. Oleh karena itu, dengan penuh kerendahan hati, penulis mengucapkan terima kasih kepada:

1. Ibu Prof. Dr. Ir. Novirina Hendrasarie, M.T. selaku Dekan Fakultas Ilmu Komputer Universitas Pembangunan Nasional “Veteran” Jawa Timur
2. Bapak Dr. Eng. Ir. Dwi Arman Prasetya, ST., MT., IPU selaku Koordinator Program Studi Sains Data Universitas Pembangunan Nasional “Veteran” Jawa Timur.
3. Bapak Dr. Ir. Mohammad Idhom, S.P., S.Kom., M.T, selaku Dosen Pembimbing I, atas bimbingan, arahan, kesabaran, dan dedikasi yang luar biasa dalam membimbing penulis, sejak tahap awal hingga akhir penyusunan skripsi ini.
4. Bapak Trimono, S.Si., M.Si., selaku Dosen Pembimbing II, yang senantiasa memberikan masukan, koreksi, dan semangat yang sangat berarti dalam proses penyempurnaan skripsi ini.
5. Kedua orang tua serta keluarga, yang dengan tulus dan tanpa pamrih telah menjadi sumber kekuatan dan doa tiada henti dalam setiap langkah penulis. Segala perjuangan, kasih sayang, dan pengorbanan yang diberikan tidak akan pernah tergantikan oleh apapun.
6. Adik penulis, yang telah menjadi penyemangat dan pelipur lara dalam setiap masa sulit, serta selalu memberikan dukungan dengan caranya yang sederhana namun bermakna.
7. Barudal well, kos atas langit, dan teman-teman yang kebersamaan penulis selama proses penyusunan skripsi berlangsung serta selalu memberikan masukan, dukungan, dan motivasi untuk mampu menyelesaikan skripsi ini.

8. Dinas Perindustrian dan Tenaga Kerja Kota Surabaya, yang telah memberikan izin, data, dan informasi yang diperlukan dalam mendukung kelancaran penelitian ini.
9. Penulis menyadari bahwa pencapaian ini tidak terlepas dari usaha dan keyakinan yang ditanamkan kepada diri sendiri. Untuk itu, penulis mengucapkan terima kasih kepada diri sendiri atas kesediaan untuk terus belajar, tumbuh, dan tidak berhenti mencoba, meski sering kali dirundung rasa lelah dan ketidakpastian. Terima kasih telah memilih untuk terus melangkah, meskipun tidak selalu mudah.
10. Penulis menyadari bahwa tidak semua pihak dapat disebutkan satu per satu. Oleh karena itu, penulis menyampaikan terima kasih yang sebesar-besarnya kepada semua pihak yang telah memberikan dukungan, baik secara langsung maupun tidak langsung, dalam penyusunan skripsi ini.

Penulis menyadari bahwa di dalam penyusunan skripsi ini banyak terdapat kekurangan. Untuk itu kritik dan saran yang membangun dari semua pihak sangat diharapkan demi kesempurnaan penulisan skripsi ini. Akhirnya, dengan segala keterbatasan yang penulis miliki semoga laporan ini dapat bermanfaat bagi semua pihak umumnya dan penulis pada khususnya.

Surabaya, 27 Mei 2025

Penulis

DAFTAR ISI

LEMBAR JUDUL.....	i
LEMBAR PENGESAHAN	ii
LEMBAR PERSETUJUAN	iii
SURAT PERNYATAAN BEBAS PLAGIASI.....	iv
ABSTRAK	v
ABSTRACT	vii
KATA PENGANTAR.....	ix
DAFTAR ISI.....	xi
DAFTAR GAMBAR.....	xv
DAFTAR TABEL	xvii
DAFTAR LAMPIRAN	xix
DAFTAR NOTASI.....	xxi
BAB I PENDAHULUAN.....	1
1.1. Latar Belakang.....	1
1.2. Rumusan Masalah.....	3
1.3. Batasan Masalah	4
1.4. Tujuan Penelitian	4
1.5. Manfaat Penelitian	5
BAB II TINJAUAN PUSTAKA.....	7
2.1. Penelitian Terdahulu.....	7
2.2. Dasar Teori	13
2.1.1. Industri Kecil Menengah (IKM)	13
2.1.2. <i>Clustering</i>	14
2.1.3. Standarisasi Data	15
2.1.4. <i>Balanced Iterative Reducing and Clustering Using Hierarchies</i> (BIRCH).....	16
2.1.5. <i>Silhouette Coefficient</i>	24
2.1.6. <i>Streamlit</i>	26
BAB III DESAIN DAN IMPLEMENTASI SISTEM	29
3.1. Variabel Penelitian dan Sumber Data.....	29

3.2. Langkah Analisis.....	31
3.2.1. Identifikasi Masalah.....	31
3.2.2. Pengumpulan Data dan Studi Literatur.....	31
3.2.3. <i>Preprocessing</i> Data.....	32
3.2.4. <i>Clustering</i> dengan Algoritma BIRCH	33
3.2.5. Evaluasi Model	33
3.2.6. Analisis Hasil Klasterisasi	34
3.2.7. Pembuatan <i>Graphical User Interface</i> (GUI)	34
3.3. Desain Sistem.....	35
BAB IV HASIL DAN PEMBAHASAN	39
4.1. Analisis Deskriptif Data Awal	39
4.2. <i>Preprocessing</i> Data	43
4.2.1. Penanganan <i>Missing Values</i>	43
4.2.2. <i>Label Encoding</i>	45
4.2.3. Standarisasi Data	47
4.3. Analisis Awal Model BIRCH	49
4.3.1. Inisialisasi Parameter Model BIRCH	49
4.3.2. Struktur Model BIRCH	50
4.3.3. Simulasi Perhitungan Manual.....	56
4.4. Klasterisasi Menggunakan BIRCH	59
4.4.1. Penentuan Konfigurasi Parameter	59
4.4.2. Penerapan Metode BIRCH	62
4.5. Evaluasi Efisiensi Waktu dan Memori.....	65
4.6. Evaluasi Hasil Klasterisasi.....	67
4.7. Analisis Hasil Klasterisasi	68
4.7.1. Analisis Statistik Variabel Numerik pada Tiap Klaster.....	73
4.7.2. Distribusi Variabel Kategorikal per Klaster	75
4.7.3. Interpretasi Klaster.....	77
4.8. Implementasi GUI.....	80
4.8.1. Menu <i>About</i>	81
4.8.2. Menu <i>Upload</i> Data	81
4.8.3. Menu <i>Preprocessing</i>	82

4.8.4. Menu <i>Clustering</i>	85
4.8.5. Menu Evaluasi dan Analisis Kluster	88
4.8.6. Menu <i>Download</i>	89
BAB V PENUTUP	91
5.1. Kesimpulan.....	91
5.2. Saran Pengembangan.....	93
DAFTAR PUSTAKA	95
LAMPIRAN.....	99

Halaman ini sengaja dikosongkan

DAFTAR GAMBAR

Gambar 2.1 Struktur CF-Tree [24]	20
Gambar 2.2 Diagram alir algoritma BIRCH	22
Gambar 3.1 Diagram Alir	31
Gambar 3.2 Desain Sistem	35
Gambar 4.1 Ilustrasi <i>CFTree</i>	51
Gambar 4.2 Visualisasi pairplot variabel numerik	69
Gambar 4.3 Visualisasi pairplot kombinasi 1	70
Gambar 4.4 Visualisasi pairplot kombinasi 2	71
Gambar 4.5 Visualisasi pairplot variabel kategorik	72
Gambar 4.6 Visualisasi kluster	78
Gambar 4.7 Tampilan menu <i>about</i>	81
Gambar 4.8 Tampilan menu <i>upload</i> data	81
Gambar 4.9 Tampilan setelah unggah data	82
Gambar 4.10 Tampilan hasil statistika deskriptif	83
Gambar 4.11 Tampilan penanganan <i>missing value</i>	83
Gambar 4.12 Tampilan proses <i>encoding</i>	84
Gambar 4.13 Tampilan hasil standarisasi	84
Gambar 4.14 Tampilan klustering menggunakan model .pkl	85
Gambar 4.15 Tampilan pemilihan parameter	86
Gambar 4.16 Tampilan dataset hasil klustering	86
Gambar 4.17 Tampilan visualisasi PCA 3D	87
Gambar 4.18 Tampilan distribusi kluster	87
Gambar 4.19 Tampilan <i>silhouette score</i>	88
Gambar 4.20 Tampilan ringkasan statistik	88
Gambar 4.21 Tampilan interpretasi kluster	89
Gambar 4.22 Tampilan menu <i>download</i> hasil kluster	89

Halaman ini sengaja dikosongkan

DAFTAR TABEL

Tabel 2.1 Penelitian Terdahulu	7
Tabel 2.2 Indeks <i>Silhouette Coefficient</i>	26
Tabel 3.1 Variabel Penelitian	29
Tabel 3.2 Struktur Data	30
Tabel 4.1 Data Awal	39
Tabel 4.2 Statistik Deskriptif Variabel Numerik	40
Tabel 4.3 Distribusi Variabel Kategorik	41
Tabel 4.4 Jumlah <i>missing values</i>	43
Tabel 4.5 Jumlah <i>missing values</i> akhir	44
Tabel 4.6 Data hasil <i>encoding</i>	47
Tabel 4.7 Data hasil standarisasi	48
Tabel 4.8 Data perhitungan manual	57
Tabel 4.9 Hasil node awal	57
Tabel 4.10 Hasil node kedua	58
Tabel 4.11 Hasil node akhir	59
Tabel 4.12 Hasil evaluasi parameter	61
Tabel 4.13 Hasil klasterisasi	63
Tabel 4.14 Hasil evaluasi efisiensi waktu dan memori	67
Tabel 4.15 Hasil evaluasi jumlah klaster	68
Tabel 4.16 Analisis statistik variabel luas tanah	73
Tabel 4.17 Analisis statistik variabel modal awal	74
Tabel 4.18 Analisis statistik variabel tenaga kerja	74
Tabel 4.19 Distribusi variabel skala usaha	75
Tabel 4.20 Distribusi variabel skala usaha	76
Tabel 4.21 Distribusi variabel jenis perusahaan	76

Halaman ini sengaja dikosongkan

DAFTAR LAMPIRAN

Lampiran 1 <i>Dataset</i> Penelitian.....	99
Lampiran 2 <i>Source Code</i> Program.....	100
Lampiran 3 GUI Streamlit	101
Lampiran 4 LoA Jurnal	102

Halaman ini sengaja dikosongkan

DAFTAR NOTASI

x_i	:	Data ke- i
$\bar{x}$	:	Rata-rata data
s	:	Standar deviasi dari data
N	:	Jumlah total data dalam <i>cluster</i>
LS	:	Jumlah koordinat data linier dari semua titik dalam <i>cluster</i>
SS	:	Jumlah koordinat kuadrat dari semua titik dalam Intensitas
P_i	:	Titik data individual dalam cluster
$ P_i ^2$	:	Kuadrat dari titik data individual dalam cluster saturation
CF_{12}	:	Gabungan CF_1 dan CF_2
N_1	:	Jumlah total titik data dalam <i>cluster</i> pertama
N_2	:	Jumlah total titik data dalam <i>cluster</i> kedua
$\overrightarrow{LS_1}$	:	Jumlah koordinat dari titik-titik dalam <i>cluster</i> pertama
$\overrightarrow{LS_2}$	:	Jumlah koordinat dari titik-titik dalam <i>cluster</i> kedua
SS_1	:	Jumlah kuadrat dari koordinat titik dalam <i>cluster</i> pertama
SS_2	:	Jumlah kuadrat dari koordinat titik dalam <i>cluster</i> kedua
$D2$	:	Centroid dari <i>cluster</i>
R	:	Radius dari <i>cluster</i>
$a(i)$	:	Rata-rata jarak objek (i) ke semua objek lain dalam cluster A
$d(I, j)$	:	Jarak antara data objek i dan objek j
$ A $	:	Jumlah total objek dalam <i>cluster</i> A
$d(i, C)$	:	Rata-rata jarak objek (i) ke semua objek lain pada <i>cluster</i> C
$ C $	:	Jumlah total objek dalam <i>cluster</i> C

$b(i)$: *Cluster* tetangga dari objek (i) yang mencapai nilai minimum

$s(i)$: Nilai *Silhouette Coefficient*