

**SISTEM PENEMPATAN KARYAWAN BERDASARKAN
KEAHLIAN DAN KINERJA DENGAN METODE *BIRCH* DI
PT. SINAR MENARA DELI (AGUNG PODOMORO LAND)**

SKRIPSI

OLEH:

MHD.WAHYUDA KASEA PRATAMA MUNTHE

178160093



**PROGRAM STUDI INFORMATIKA
FAKULTAS TEKNIK
UNIVERSITAS MEDAN AREA
MEDAN
2023**

UNIVERSITAS MEDAN AREA

© Hak Cipta Di Lindungi Undang-Undang

1. Dilarang Mengutip sebagian atau seluruh dokumen ini tanpa mencantumkan sumber
2. Pengutipan hanya untuk keperluan pendidikan, penelitian dan penulisan karya ilmiah
3. Dilarang memperbanyak sebagian atau seluruh karya ini dalam bentuk apapun tanpa izin Universitas Medan Area

Document Accepted 14/8/24

Access From (repository.uma.ac.id)14/8/24

**SISTEM PENILAIAN KARYAWAN BERDASARKAN
KEAHLIAN DAN KINERJA DENGAN METODE *BIRCH* DI PT.
SINAR MENARA DELI (AGUNG PODOMORO LAND)**

SKRIPSI

Diajukan sebagai Salah Satu Syarat untuk Memperoleh
Gelar Sarjana di Fakultas Teknik
Universitas Medan Area

OLEH :

MHD WAHYUDA KASEA PRATAMA MUNTHE

178160093

**PROGRAM STUDI INFORMATIKA
FAKULTAS TEKNIK
UNIVERSITAS MEDAN AREA
MEDAN**

2023

UNIVERSITAS MEDAN AREA

© Hak Cipta Di Lindungi Undang-Undang

1. Dilarang Mengutip sebagian atau seluruh dokumen ini tanpa mencantumkan sumber
2. Pengutipan hanya untuk keperluan pendidikan, penelitian dan penulisan karya ilmiah
3. Dilarang memperbanyak sebagian atau seluruh karya ini dalam bentuk apapun tanpa izin Universitas Medan Area

Document Accepted 14/8/24

Access From (repository.uma.ac.id)14/8/24

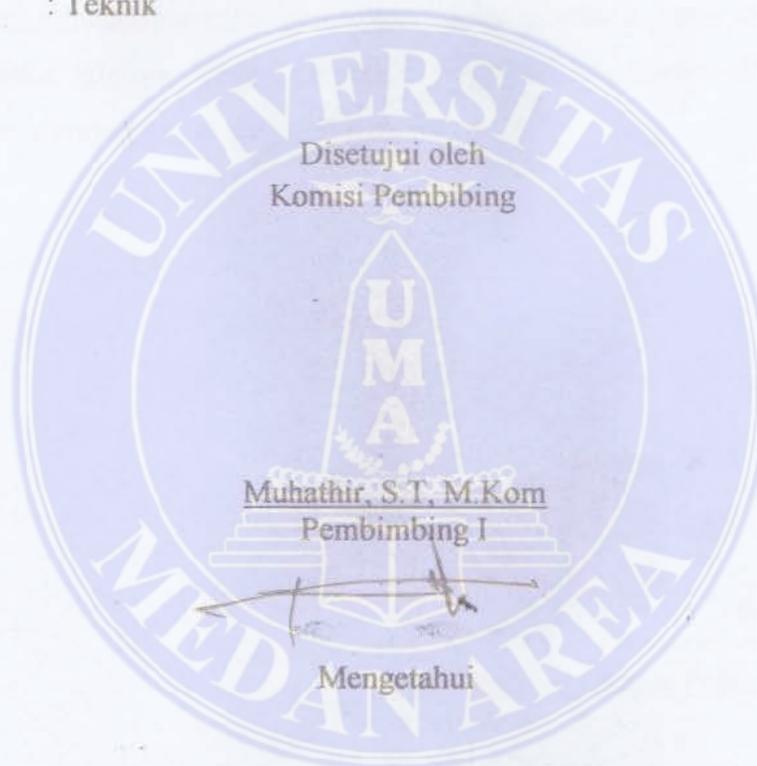
LEMBAR PENGESAHAN

Judul Skripsi : Sistem Penilaian Karyawan Berdasarkan Keahlian dan Kinerja
dengan Metode *Birch* di PT. Sinar Menara Deli (Agung Podomoro
Land)

Nama : Mhd Wahyuda Kasea Pratama Munthe

NPM : 178160093

Fakultas : Teknik



Tanggal Lulus : 25 November 2023

HALAMAN PERNYATAAN

Saya menyatakan bahwa skripsi yang saya susun, sebagai syarat memperoleh gelar sarjana di Universitas Medan Area merupakan hasil karya tulis saya sendiri. Adapun bagian-bagian tertentu dalam penulisan skripsi ini yang saya kutip dari hasil karya orang lain telah dituliskan sumbernya secara jelas sesuai dengan norma, kaidah, dan etika penulisan ilmiah.

Saya bersedia menerima saksi pencabutan gelar akademik yang saya peroleh dan sanksi-sanksi lainnya dengan peraturan yang berlaku, apabila di kemudian hari ditemukan adanya plagiat dalam skripsi ini.

Medan, 29 Agustus 2023



Mhd Wahyuda Kasea Pratama Munthe
178160093

HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR/SKRIPSI/TESIS KEPENTINGAN AKADEMIS

Sebagai sivitas akademik Universitas Medan Area, Saya yang bertanda tangan di bawah ini:

Nama : Mhd Wahyuda Kasea Pratama Munthe
NPM : 178160093
Program Studi : Teknik Informatika
Fakultas : Teknik
Jenis Karya : Skripsi

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Medan Area **Hak Bebas Royalti Noneksklusif (*Non-Exclusive Royalty-Free Right*)** Atas karya ilmiah saya yang berjudul :

Sistem Penempatan Karyawan Berdasarkan Keahlian dan Kinerja dengan Metode *Birch* di PT. Sinar Menara Deli (Agung Podomoro Land).

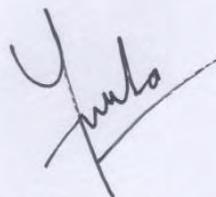
Beserta perangkat yang ada (jika diperlukan). Dengan hak bebas royalti noneksklusif ini Universitas Medan Area berhak menyimpan, berhak mengalihmedia/format-kan, mengelola dalam bentuk pangkalan data (database), merawat, dan memublikasikan skripsi saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik hak cipta.

Demikian pernyataan ini saya buat dengan sebenarnya.

Dibuat di : Medan

Pada tanggal : 25 November 2023

Yang menyatakan



(Mhd Wahyuda Kasea Pratama Munthe)

UNIVERSITAS MEDAN AREA

© Hak Cipta Di Lindungi Undang-Undang

Document Accepted 14/8/24

1. Dilarang Mengutip sebagian atau seluruh dokumen ini tanpa mencantumkan sumber
2. Pengutipan hanya untuk keperluan pendidikan, penelitian dan penulisan karya ilmiah
3. Dilarang memperbanyak sebagian atau seluruh karya ini dalam bentuk apapun tanpa izin Universitas Medan Area

Access From (repository.uma.ac.id)14/8/24

ABSTRAK

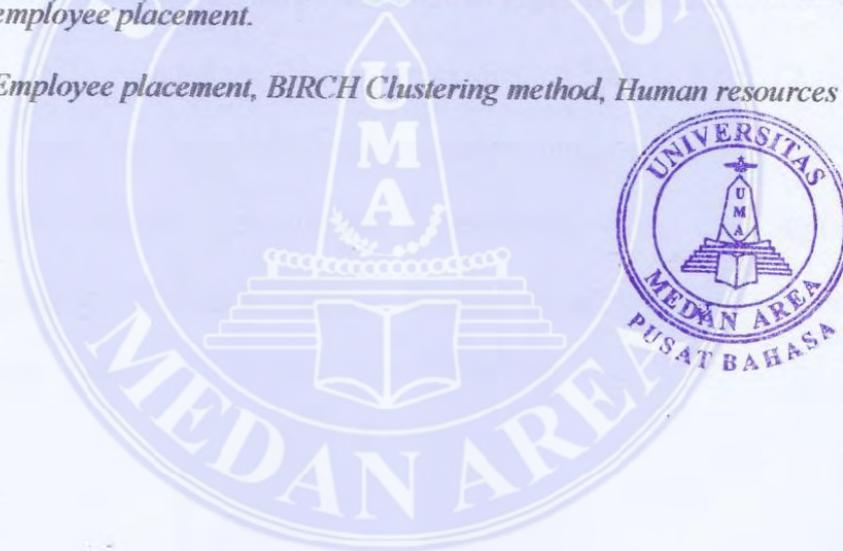
Menentukan cara untuk memperbaiki nilai sebuah pola penempatan yang berguna dalam kumpulan data yang besar telah menarik minat yang besar, dan salah satunya masalah yang paling banyak dipelajari di perusahaan PT. Sinar Menara Deli (Agung Podomoro Land). Penelitian ini mengimplementasikan metode algoritma Birch mengelompokkan sejumlah data numerik yang besar dengan mengintegrasikan pengelompokan secara hirarki dan metode pengelompokan lain seperti partisi berulang pada dataset yang mencakup beragam atribut terkait karyawan dalam suatu perusahaan identifikasi Cluster. Tujuan dari penelitian ini adalah untuk mengidentifikasi performance karyawan yang signifikan dalam data tersebut. Selain itu, penelitian ini mencakup mengenai pengalaman, kinerja, dan kompetensi serta berbagai aspek perusahaan. Hasil dari penelitian ini memiliki dampak positif dalam konteks manajemen sumber daya manusia dan pengambilan keputusan penempatan karyawan untuk menunjang karir. Dengan menggunakan nilai $threshold=2$ dan $branch\ factor=30$, dengan kelompok nilai 200 karyawan terbaik di masing-masing divisi, baik dan cukup mendapatkan hasil bahwa kelompok yang terbaik terdiri dari kumpulan posisi jabatan di setiap masing-masing divisi yaitu divisi Marketing, Finance, Operasional sampai IT. Nilai evaluasi menggunakan V-measure, AMI, dan ARI dan diperoleh sebesar 0.01457. Nilai yang mendekati 1 menunjukkan kluster yang baik dengan pemisahan yang jelas antara cluster dan anggota cluster yang serupa dan menunjukkan kualitas cluster yang cukup baik. Kesimpulan yang diperoleh dari penelitian ini memberikan wawasan yang baik untuk manajemen sumber daya manusia, membantu mereka dalam pengambilan keputusan yang baik untuk penempatan karyawan.

Kata Kunci: Metode BIRCH *Clustering*, Penempatan Karyawan, Sumber Daya Manusia

ABSTRACT

Determining methods to improve the value of a placement pattern useful in large datasets has garnered significant interest, particularly in the case of PT. Sinar Menara Deli (Agung Podomoro Land). This research implemented the BIRCH algorithm to cluster a large amount of numerical data by integrating hierarchical clustering with other clustering methods such as iterative partitioning on a dataset encompassing various attributes related to employees within a company's cluster identification. The aim of this study was to identify significant employee performance within the data. Additionally, the research encompassed aspects of experience, performance, and competence as well as various company-related aspects. The results of this study had a positive impact in the context of human resource management and employee placement decision-making to support career advancement. Using a threshold value of 2 and a branch factor of 30, with a group of the top 200 employees in each division, it was found that the best group consisted of job positions within each respective division, including Marketing, Finance, Operations, and IT. Evaluation metrics using V-measure, AMI, and ARI yielded a value of 0.01457. A value approaching 1 indicates good clustering with clear separation between clusters and similar cluster members, demonstrating a fairly good cluster quality. The conclusions drawn from this research provided valuable insights for human resource management, aiding in better decision-making for employee placement.

Keywords: *Employee placement, BIRCH Clustering method, Human resources*



RIWAYAT HIDUP

Penulis merupakan putra atau anak ke-1 (Pertama) dari ayah Mhd Risjon Munthe dan ibu Sri Widiastuti yang dilahirkan di Medan, pada tanggal 05 Desember 1999. Pada tahun 2017, Penulis lulus dari SMK Namira Tech Nusantara Jurusan Teknik Komputer dan Jaringan, lalu kemudian terdaftar sebagai mahasiswa Fakultas Teknik Prodi Teknik Informatika Universitas Medan Area.

Selama mengikuti perkuliahan, Penulis juga pernah bekerja di Universitas Medan Area sebagai outsourcing perpustakaan, dan juga berbagai kompetisi mulai dari teater antar SMK. Selanjutnya Penulis melaksanakan Praktek Kerja Lapangan (PKL) di PT. Telkom Indonesia Tbk yaitu perusahaan milik negara.

Selain itu, Penulis mendapatkan kesempatan untuk mengikuti dan melaksanakan program magang di beberapa perusahaan, salah satunya perusahaan di PT. Sinar Menara Deli.

KATA PENGANTAR

Puji syukur kehadiran Allah SWT Yang Maha Esa atas, rahmat, dan hidayah-Nya, sehingga saya bisa menyelesaikan penyusunan skripsi dengan judul “ Sistem Penilaian Karyawan Berdasarkan Keahlian dan Kinerja dengan Metode *Birch* di PT. Sinar Menara Deli (Agung Podomoro Land) ”.

Pada kesempatan ini, penulis mengucapkan banyak terima kasih kepada pihak-pihak yang telah memberikan banyak dukungan seperti dukungan motivasi, dukungan informasi materi, dan arahan, sehingga penulis dapat menyelesaikan skripsi ini dengan sebaik-baiknya:

1. Allah SWT Yang Maha Esa, berkat rahmat dan hidayah-Nya skripsi ini dapat terselesaikan,
2. Orang Tua Bapak dan Ibu Saya yang telah mendukung, memberikan semangat, motivasi, dan banyak perhatian serta memenuhi segala kebutuhan saya selama masa penyusunan tugas akhir/skripsi ini,
3. Bapak Prof. Dr. Dadan Ramdan, M.Eng., M.Sc., selaku Rektor Universitas Medan Area,
4. Bapak Dr. Eng., Supriatno, S.T., M.T., selaku Dekan Fakultas Teknik Universitas Medan Area yang telah memberikan banyak masukan, kritik, saran, dan motivasi kepada penulis serta membimbing penulis dalam menyelesaikan tugas akhir/skripsi ini,
5. Ibu Susilawati, S.Kom., M.Kom., selaku Wakil Dekan Fakultas Teknik Universitas Medan Area yang telah memberikan, kritik dan saran kepada penulis dalam menyelesaikan tugas akhir/skripsi ini,

6. Bapak Rizki Muliono, S.Kom., M.Kom., selaku Ketua Program Studi Teknik Informatika Universitas Medan Area,
7. Bapak Muhathir ST, M.Kom., selaku Dosen Pembimbing I yang telah memberikan arahan, bimbingan, semangat, dan motivasi kepada penulis hingga penyusunan tugas akhir/skripsi ini terselesaikan,
8. Serta semua pihak yang telah membantu penulis dalam menyelesaikan tugas akhir/skripsi ini, yang Namanya tidak bisa disebutkan satu persatu.

Terimakasih banyak.

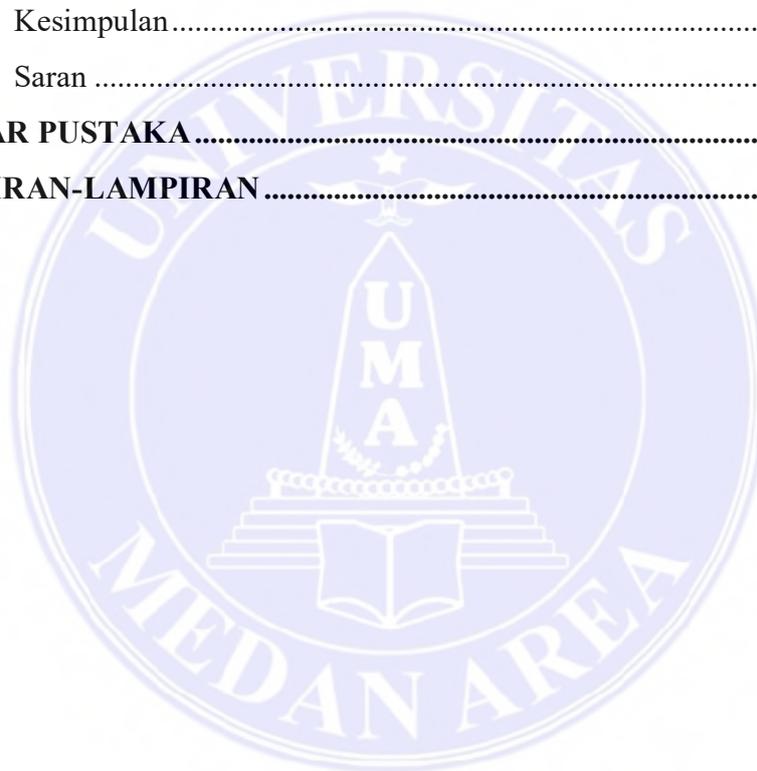
Penulis menyadari bahwa Tugas Akhir/Skripsi ini masih jauh dari kata sempurna, untuk itu penulis mengharapkan kritik dan saran yang membangun dari pembaca untuk pengembangan selanjutnya.

Medan, 29 September 2023

DAFTAR ISI

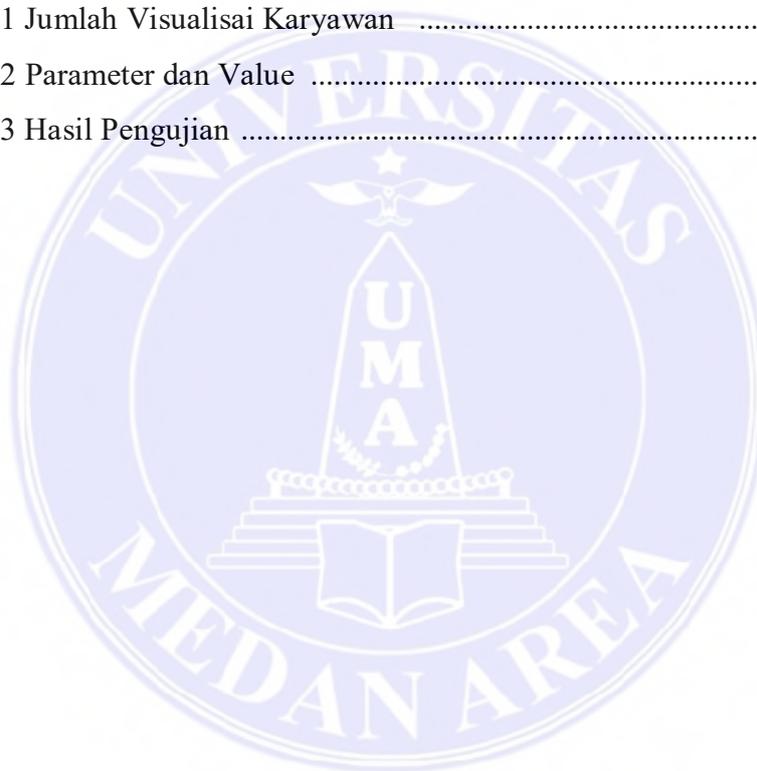
HALAMAN PERNYATAAN	Error! Bookmark not defined.
HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI	Error! Bookmark not defined.
ABSTRAK	vi
ABSTRACT	Error! Bookmark not defined.
RIWAYAT HIDUP	viii
KATA PENGANTAR	ix
DAFTAR TABEL	xiii
DAFTAR GAMBAR	xiv
BAB I	1
PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	4
1.3 Batasan Masalah.....	4
1.4 Tujuan Penelitian	5
1.5 Manfaat Penelitian	5
BAB II	6
LANDASAN TEORI	6
2.1 Penempatan.....	6
2.2 Metode Clustering.....	8
2.3 Metode BIRCH	10
2.4 Metrik Evaluasi untuk Algoritma Clustering	21
2.5 Penelitian Terdahulu	23
BAB III	27
METODE PENELITIAN	27
3.1 Tahapan Penelitian	27
3.2 Metode Pengumpulan Data	29
3.3 Proses Kerja Metode BIRCH	29
3.4 Hipotesis	38
3.5 Metode Penyelesaian	38

BAB IV	40
HASIL DAN PEMBAHASAN	40
4.1 Hasil	40
4.1.1 Penentuan Jumlah Clustering	40
4.1.2 Initial Data Visualisasi	40
4.1.3 Clustering Menggunakan BIRCH	41
4.2 Pembahasan	49
BAB V	51
KESIMPULAN DAN SARAN	51
5.1 Kesimpulan	51
5.2 Saran	51
DAFTAR PUSTAKA	53
LAMPIRAN-LAMPIRAN	56



DAFTAR TABEL

Tabel 2.1 Penelitian Terdahulu	23
Tabel 3.1 Data Karyawan	31
Tabel 3.2 Data Kriteria	31
Tabel 3.3 Data Karyawan Setelah Konversi Nilai	32
Tabel 3.4 Data Karyawan Setelah Normalisasi	32
Tabel 4.1 Jumlah Visualisasi Karyawan	40
Tabel 4.2 Parameter dan Value	42
Tabel 4.3 Hasil Pengujian	43



DAFTAR GAMBAR

Gambar 2.1 <i>Flowchart</i> Algoritma <i>Birch</i>	15
Gambar 2.2 Skema Fase pada <i>Birch</i>	16
Gambar 3.1 Proses Analisa Data	27
Gambar 3.2 <i>Flowchart</i> dari Metode <i>Birch</i>	30
Gambar 3.3 Struktur CF Tree Setelah Pengisian Data 1	34
Gambar 3.4 Struktur CF Tree Setelah Pengisian Data 2	35
Gambar 3.5 Struktur CF Tree Setelah Pengisian Data 3	36
Gambar 3.6 Struktur CF Tree Setelah Pengisian Data 4	38
Gambar 4.1 Grafik <i>Scatter</i> dari Data Karyawan	39
Gambar 4.2 Hasil Pengelompokkan Data Karyawan BF30 T0.5 <i>I</i>	41
Gambar 4.3 Grafik <i>Scatter</i> dari Hasil <i>Clustering</i> BF30 T0,5	43
Gambar 4.4 Hasil Pengelompokkan Data Karyawan BF30 T1	44
Gambar 4.5 Grafik <i>Scatter</i> dari Hasil <i>Clustering</i> BF30 T1	44
Gambar 4.6 Hasil Pengelompokkan Data Karyawan BF30 T1.5	45
Gambar 4.7 Grafik <i>Scatter</i> dari Hasil <i>Clustering</i> BIRCH BF30 T1.5	45
Gambar 4.8 Hasil Pengelompokkan Data Karyawan BF30 T2.....	46
Gambar 4.9 Grafik <i>Scatter</i> dari Hasil <i>Clustering</i> BIRCH BF30 T2	46

BAB I

PENDAHULUAN

1.1 Latar Belakang

Podomoro merupakan salah satu perusahaan swasta yang bergerak di bidang tempat perbelanjaan (*mall*). Penempatan kerja pegawai di Podomoro terdiri 2 (dua) bagian diantaranya adalah penempatan awal dan penempatan ulang. Penempatan awal ditujukan untuk pegawai baru yang telah lulus seleksi penerimaan pegawai. Sementara itu, pelaksanaan mutasi pada pegawai Podomoro Pada prinsipnya, tujuannya adalah untuk melakukan perpindahan dan pengisian jabatan bagi pegawai yang memiliki kapasitas, kompetensi, dan integritas yang memadai guna memenuhi posisi tertentu, sehingga pelaksanaan tugas dapat berjalan secara efisien dan efektif. Seleksi pegawai yang akan mengisi posisi kosong didasarkan pada evaluasi kompetensi, latar belakang pendidikan, dan pengalaman yang dimilikinya. Terdapat beberapa faktor yang memotivasi terjadinya mutasi, yakni aspirasi individu pegawai dan kebutuhan perusahaan atau organisasi itu sendiri.

Penempatan kerja pegawai pada Podomoro Masih terdapat hambatan yang menghambat pencapaian kinerja yang efektif pada pegawai, dikarenakan dalam penempatan kerja masih ada praktik nepotisme, yang berujung pada penempatan di jabatan yang tidak sejalan dengan kemampuan mereka. Kegagalan dalam penempatan kerja ini menghasilkan kinerja pegawai yang belum optimal, dan hal ini pada akhirnya memengaruhi pencapaian tujuan yang diharapkan oleh

Podomoro. Untuk menyelesaikan permasalahan tersebut, maka dapat diterapkan metode *data mining*.

Data mining adalah proses untuk mendapatkan informasi yang berguna dari gudang basis data yang besar. *Data mining* Dapat dimanfaatkan untuk memecahkan masalah dalam ranah intelektual, ekonomi, dan bisnis, yang terbagi menjadi enam tugas utama, yakni klasifikasi, estimasi, prediksi, afinitasi, *clustering*, deskripsi dan penentuan profil (Budiman, Safitr, & Ispriyanti, 2019). Penilaian *data mining* yang digunakan dalam penelitian ini adalah nilai kerjasama, nilai loyalitas, nilai kerajinan, nilai disiplin dan nilai ketelitian. *Clustering* adalah proses pengelompokan data, catatan, atau observasi berdasarkan kesamaan objek atau atribut tertentu. Ada berbagai algoritma klastering yang telah digunakan oleh peneliti sebelumnya, *K-Means* (Zakiyah, Merlina, & Mayangky, 2022), *Improved K-Means* (Putri, Cahya, & Fauzi, 2019), *Fuzzy C-Means* (Novianti, Yasmin, & Novitasari, 2022), *DBSCAN* (Tukiyat & Djohan, 2022), *K-Medoids (PAM)* (Marlina, Putri, Fernando, & Ramadhan, 2020), *CLARA* (Fitrayana & Saputro, 2022) dan *BIRCH Clustering* (Nur Fadlilah, 2022). Setiap algoritma memiliki kelebihan dan kekurangan masing-masing, namun prinsip algoritma sama, yaitu mengelompokkan data sesuai dengan karakteristik dan mengukur jarak kemiripan antar data dalam satu kelompok (Pramesti, Furqon, & Dewi, 2019).

Berdasarkan permasalahan yang ada pada Podomoro, maka perlu diselesaikan untuk mempermudah Podomoro dalam penempatan karyawan dengan memanfaatkan algoritma BIRCH dalam pengelompokan data karyawan. Algoritma BIRCH merupakan metode pengelompokan yang bersifat hirarki (*hierarchical clustering method*). Algoritma BIRCH sangat sesuai untuk set data

yang sangat besar, atau untuk *streaming* data, karena kemampuannya untuk menemukan solusi pengelompokan yang baik dengan hanya satu pemindaian data (Ramadhani, 2019). Algoritma BIRCH dirancang untuk mengelompokkan sejumlah data numerik yang besar dengan mengintegrasikan pengelompokan secara hirarki dan metode pengelompokan lain seperti partisi berulang (Shauma, Purwanto, & Niovianty, 2019). Adapun aturan yang digunakan dalam penempatan karyawan berdasarkan keahlian mencakup pencapaian target dan penilaian atasan.

Beberapa penelitian sejenis yang pernah dilakukan sebelumnya seperti Penentuan Penempatan Karyawan Baru di PDAM Kisaran dengan Metode SMART, dimana dengan menggunakan metode SMART (*Simple Multi Attribute Rating Technique*) pada sistem pendukung ini maka akan lebih efektif dalam penentuan penempatan karyawan baru serta membantu Departemen sumber daya manusia dalam mengambil keputusannya (Azhar, 2019). Penelitian lainnya yaitu Sistem Pendukung Keputusan Penempatan Karyawan Berdasarkan *Test Dominant-Influence-Steady-Compliance* (DISC) Menggunakan Metode *Profile Matching*, dimana hasil dari penelitian ini berupa rekomendasi karyawan yang dapat menempati suatu jabatan sesuai dengan kepribadiannya dengan keakurasian data yang di uji sebesar 85,8% (Setiawati, Abdillah, & Hadiana, 2019). Sementara itu, penelitian mengenai metode BIRCH yaitu Pengelompokan Menggunakan Data *Tweet* Kecelakaan Pendekatan *Text Mining* dan Algoritma BIRCH, dimana hasil yang diperoleh dari penelitian ini adalah data yang diperoleh dari *kaggle* menghasilkan 1545 cluster dengan *silhouette coefficient* bernilai 0.1159964638217295 sedangkan dari data hasil *crawling twitter* menghasilkan

487 cluster dengan *silhouette coefficient* bernilai 0.7262655918349612 (Nur Fadlilah, 2022).

Berdasarkan penjabaran diatas, penulis tertarik untuk mengangkat topik penempatan karyawan pada Podomoro ini. Oleh karena itu, penulis mengambil skripsi dengan judul “**Sistem Penempatan Karyawan Berdasarkan Keahlian dan Kinerja dengan Metode BIRCH**”.

1.2 Rumusan Masalah

Berdasarkan penjelasan latar belakang di atas, maka yang menjadi rumusan masalah adalah:

1. Bagaimana meningkatkan kinerja dari karyawan pada Podomoro.
2. Bagaimana melakukan proses penempatan karyawan berdasarkan keahlian dan kinerjanya dengan menggunakan metode BIRCH.

1.3 Batasan Masalah

Batasan masalah dalam tugas akhir ini mencakup :

1. Penelitian ditujukan khusus untuk karyawan yang bekerja di lapangan.
2. Kriteria yang digunakan mencakup:
 - a. Penilaian keahlian.
 - b. Penilaian kinerja.

1.4 Tujuan Penelitian

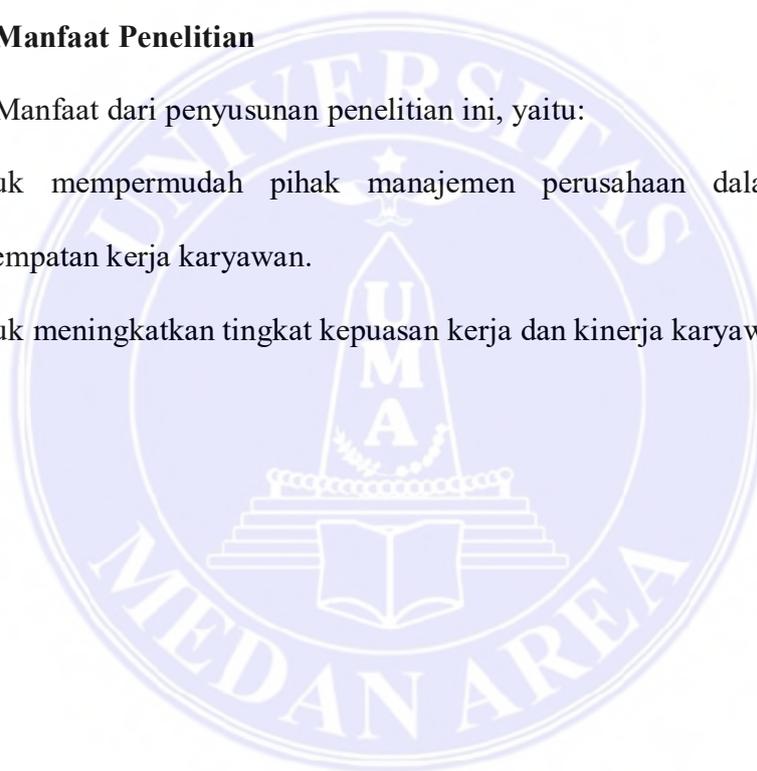
Tujuan penyusunan penelitian ini, yaitu:

1. Meningkatkan performance karyawan pada Podomoro berdasarkan kinerja dan keahlian.
2. Melakukan proses penempatan karyawan berdasarkan keahlian dan kinerjanya dengan menggunakan metode BIRCH.

1.5 Manfaat Penelitian

Manfaat dari penyusunan penelitian ini, yaitu:

1. Untuk mempermudah pihak manajemen perusahaan dalam melakukan penempatan kerja karyawan.
2. Untuk meningkatkan tingkat kepuasan kerja dan kinerja karyawan.



BAB II

LANDASAN TEORI

2.1 Penempatan

Penempatan Penempatan karyawan atau sumber daya manusia (SDM) merupakan langkah pengisian posisi jabatan guna melaksanakan tugas-tugas yang telah ditentukan oleh organisasi. Proses penempatan karyawan memegang peranan penting karena berkaitan dengan berbagai kepentingan organisasi, dan penempatan ini dilakukan baik terhadap karyawan baru maupun yang sudah berpengalaman. Memilih penempatan kerja yang cocok dengan kemampuan atau keterampilan yang dimiliki oleh karyawan akan secara signifikan memfasilitasi perkembangan karier mereka di masa mendatang. (Sahadewa & Rahmawati, 2021).

Menurut Hasibuan, penempatan merujuk pada tahap setelah seleksi, di mana calon karyawan yang telah diterima ditempatkan pada jabatan atau pekerjaan yang sesuai dengan kebutuhan, dan dalam proses ini, juga diberikan kewenangan kepada individu tersebut. Sementara itu, Rivai menyatakan bahwa penempatan karyawan mencakup penugasan ulang seorang karyawan ke pekerjaan baru, yang membawa tanggung jawab, wewenang, serta kesempatan untuk pengalaman dan perkembangan karier.

Berdasarkan pandangan para ahli di atas, dapat disimpulkan bahwa penempatan karyawan dan pengembangan karier memiliki hubungan yang erat terhadap kinerja karyawan. Dengan menempatkan karyawan sesuai dengan kompetensinya, mereka dapat menguasai tugas-tugas pekerjaan dengan baik. Hal

ini memiliki kaitan penting dengan pengembangan karier, yang dapat berjalan lebih optimal ketika karyawan berada dalam lingkungan pekerjaan yang sesuai. Sehingga, keselarasan antara penempatan karyawan dan pengembangan karier berkontribusi terhadap peningkatan kinerja karyawan secara keseluruhan (Sahadewa & Rahmawati, 2021).

Menurut Hasibuan, indikator penempatan kerja karyawan adalah sebagai berikut (Khairiyah, Febriani, & Yanti, 2020):

1. Orientasi adalah proses memberikan informasi kepada karyawan baru mengenai hak, kewajiban, tugas, tanggung jawab, peraturan perusahaan, sejarah, dan struktur organisasi. Karyawan baru juga diperkenalkan kepada rekan kerja yang sudah berpengalaman. Orientasi bertujuan untuk membuat karyawan baru merasa diterima dalam lingkungan kerjanya, sehingga mereka dapat dengan nyaman menjalankan tugas mereka tanpa rasa canggung. Tugas ini biasanya diemban oleh Unit Sumber Daya Manusia (USDM) atau departemen personalia perusahaan yang bersangkutan.
2. Induksi adalah proses memengaruhi perilaku karyawan baru yang telah dipekerjakan agar mereka patuh terhadap peraturan perusahaan dan norma sosial yang berlaku. Melalui induksi ini, diharapkan karyawan baru dapat beradaptasi dengan lingkungan kerja mereka, memungkinkan mereka untuk menjalankan tugas-tugas mereka dengan efektif dan efisien. Proses induksi ini biasanya dilakukan oleh atasan langsung dan karyawan senior dalam unit kerja yang bersangkutan.

3. Pengetahuan (knowledge) adalah kumpulan informasi yang terstruktur, yang umumnya terdiri dari fakta atau prosedur yang diterapkan secara langsung dalam kinerja suatu organisasi.
4. Keterampilan (skill) adalah kemampuan yang ditunjukkan dalam kinerja melalui tindakan yang dapat diamati atau melalui serangkaian tindakan yang mengarah pada hasil yang dapat diamati.

2.2 Metode Clustering

Clustering, juga dikenal sebagai klasifikasi, merujuk pada proses membagi objek-objek dalam dataset ke dalam kelompok-kelompok objek yang memiliki kesamaan. Objek-objek ini dapat berupa data numerik, kategorikal, atau kombinasi keduanya. Kemampuan algoritma clustering adalah untuk mengungkapkan struktur yang mendasari data, yang dapat digunakan dalam berbagai aplikasi seperti klasifikasi, taksonomi numerik, pengolahan gambar, pengenalan pola, kedokteran, ekonomi, ekologi, kecerdasan buatan, penambangan data, pemodelan, dan identifikasi. Berbeda dengan klasifikasi, clustering tidak melibatkan label kelas yang diketahui sebelumnya, dan tujuannya adalah untuk menemukan pola alami atau struktur tersembunyi dalam data. *clustering* tidak melibatkan variabel target, dan tujuannya adalah membagi keseluruhan set data menjadi sub-kelompok yang relatif homogen. (Chusyairi & Saputra, 2019).

Clustering Clustering pada data adalah proses pengelompokan himpunan data yang memiliki atribut kelas yang belum dijelaskan. Secara konseptual, *clustering* bertujuan untuk memaksimalkan kesamaan intra-kelas dan meminimalkan kesamaan antar-kelas. Sebagai contoh, dalam suatu himpunan

objek, tahap awal dapat mengelompokkan objek-objek tersebut menjadi beberapa kelompok, dan selanjutnya dapat diatur menjadi kelompok-kelompok yang lebih terstruktur berdasarkan karakteristik klasifikasinya. *Cluster* juga dapat diartikan sebagai kelompok. Dalam analisis clustering, esensinya adalah menghasilkan sejumlah *cluster* (kelompok). Sebelum melakukan analisis, penting untuk memahami bahwa himpunan data tertentu memiliki kemiripan di antara anggotanya. Oleh karena itu, anggota dengan karakteristik yang serupa dikelompokkan dalam satu atau lebih kelompok. Tujuan utama dari clustering data adalah meminimalkan fungsi tujuan yang telah ditentukan dalam proses *clustering*. Umumnya, tujuan ini melibatkan pengurangan variasi dalam suatu cluster dan peningkatan variasi antara *cluster*. Dengan cara ini, analisis *clustering* berusaha untuk membentuk kelompok-kelompok yang berbeda satu sama lain dengan karakteristik yang serupa di dalam setiap kelompok. (Muliono & Sembiring, 2019).

Analisis *cluster* merupakan salah satu alat yang digunakan dalam proses data mining dengan tujuan mengelompokkan objek-objek ke dalam cluster tertentu. Cluster adalah kumpulan objek data yang memiliki kesamaan satu sama lain dalam kelompok yang sama. Tujuan dari analisis cluster bukanlah mengkorelasikan objek satu dengan yang lain, tetapi untuk mengidentifikasi objek-objek yang memiliki kesamaan dan karakteristik yang dapat dibedakan dari kelompok lain.

Objek yang termasuk dalam satu cluster cenderung memiliki homogenitas yang lebih tinggi dibandingkan dengan objek dari cluster yang berbeda. Analisis cluster bertujuan untuk memisahkan objek-objek ke dalam kelompok berdasarkan

kesamaan karakteristiknya. Jumlah cluster yang dihasilkan dalam analisis tergantung pada jumlah data dan variasi karakteristik objek tersebut. Dengan menggunakan analisis cluster, kita dapat mengidentifikasi pola dan struktur dalam data yang tidak terlihat pada bagian pertama. (Dani, Wahyuningsih, & Rizki, 2019).

Hasil *cluster* dikatakan baik dengan ciri-ciri sebagai berikut (Dani, Wahyuningsih, & Rizki, 2019):

1. Memiliki homogenitas yang besar antar anggota pada *cluster* yang sama (*withincluster*).
2. Memiliki heterogenitas yang besar antar *cluster* yang satu dengan *cluster* lainnya (*between-cluster*).

Pada analisis *cluster*, terdapat dua metode yang dapat digunakan, yaitu metode hierarki dan non-hierarki. Pengelompokan hierarki (*hierarchical clustering*) adalah metode analisis cluster dengan membangun hierarki kelompok. Strategi untuk pengelompokan hierarki umumnya terbagi menjadi dua jenis, yaitu *agglomerative* (pemusatan) dan *divisive* (penyebaran). Metode *agglomerative* (pemusatan) merupakan pendekatan pengelompokan hierarki yang dimulai dari bawah ke atas (*bottom-up*). Metode *agglomerative* (pemusatan) biasanya digunakan dalam bidang ekonomi dan masyarakat. (Dani, Wahyuningsih, & Rizki, 2019).

2.3 Metode BIRCH

Algoritma BIRCH merupakan metode pengelompokan yang bersifat hirarki (*hierarchical clustering method*). Algoritma BIRCH dikembangkan oleh

Tian Zhang, Raghu Ramakrishnan, dan Miron Livny pada tahun 1996. Algoritma ini dapat menemukan pengelompokan dengan baik hanya dengan melakukan satu kali scan data dan dapat meningkatkan kualitas lebih lanjut dengan beberapa scan tambahan. BIRCH sangat sesuai untuk set data yang sangat besar, atau untuk *streaming* data, karena kemampuannya untuk menemukan solusi pengelompokan yang baik dengan hanya satu pemindaian data. Secara opsional, algoritma dapat melakukan pemindaian lebih lanjut melalui data untuk meningkatkan kualitas pengelompokan. BIRCH menangani set data besar dengan kompleksitas waktu dan efisiensi ruang yang lebih unggul daripada algoritma lainnya. Algoritma BIRCH dirancang untuk mengelompokkan sejumlah data numerik yang besar dengan mengintegrasikan pengelompokan sejumlah data numerik yang besar dengan mengintegrasikan pengelompokan secara hirarki dan metode pengelompokan lain seperti partisi berulang.

BIRCH adalah metode pengelompokan untuk kumpulan data yang sangat besar. BIRCH dapat membuat masalah pengelompokan besar bisa diselesaikan dengan berkonsentrasi pada porsi yang padat, dan menggunakan ringkasan yang ringkas. Rangkaian tersebut menggunakan pengukuran yang menangkap kedekatan alami data. Pengukuran ini dapat disimpan dan diperbarui secara bertahap di pohon yang tinggi yang seimbang. BIRCH dapat bekerja dengan jumlah memori yang diberikan dan kompleksitas I/O lebih dari satu pemindaian data (Shauma, Purwanto, & Niovianty, 2019). Kontribusi penting dari BIRCH adalah formulasi tentang pengelompokan, masalah dengan cara yang sesuai untuk itu dataset yang sangat besar, dengan membuat waktu dan memori kendala

eksplisit. Selain itu, BIRCH memiliki keunggulan dibandingkan pendekatan berbasis jarak sebelumnya.

Adapun keunggulan dari BIRCH adalah sebagai berikut:

1. BIRCH bersifat local (tidak seperti global) di masing-masing keputusan pengelompokan dibuat tanpa memindai semua data poin atau semua *cluster* yang ada saat ini. Ini menggunakan pengukuran yang mencerminkan kedekatan alami titik, dan pada saat yang sama dapat dipertahankan secara bertahap selama proses pengelompokan.
2. BIRCH mengeksploitasi pengamatan bahwa ruang data biasanya tidak ditempati secara seragam, dan karenanya tidak setiap titik data sama pentingnya untuk tujuan pengelompokan. Wilayah titik yang padat diperlakukan secara kolektif sebagai satu kelompok. Poin di daerah jarang diperlakukan sebagai outliers dan dihapus secara optional.
3. BIRCH memanfaatkan sepenuhnya memori yang tersedia tidak untuk diturunkan subkluster terbaik (untuk memastikan akurasi) sambil meminimalkan I/O (untuk memastikan efisiensi). Proses pengelompokan dan pengurangan terorganisir dan ditandai dengan penggunaan in-memory, *height balanced* dan struktur pohon yang sangat sibuk. Disebabkan oleh fitur-fitur ini, waktu menjalankannya secara linear *scalable*.
4. Jika menyiasati fase 4-5 opsional, BIRCH adalah metode tambahan yang tidak memerlukan keseluruhan dataset terlebih dahulu, dan hanya memindai dataset satu kali.

BIRCH adalah singkatan dari *Balance Iterative Reducing Clustering using Hierarchies*. Sebelum membahas lebih lanjut tentang algoritma, penting untuk

memahami konsep Clustering Feature (CF) *Tree*. *Clustering Feature* merupakan representasi yang ringkas dari titik data dalam suatu cluster dan mengandung informasi yang mencukupi untuk menghitung properti *intra-cluster*. *CF Tree* adalah pohon seimbang tinggi (*height-balance tree*) yang memiliki dua parameter: Faktor balancing B (*Balance B*) dan ambang batas T (*threshold T*). Setiap simpul *non-leaf* (bukan simpul daun) dalam *CF Tree* berisi maksimal B entri, dengan format $\{Cf_i, Child_i\}$, di mana i berkisar dari 1, 2, 3, ..., B. *Child_i* berfungsi sebagai petunjuk ke anak ke- i , sementara Cf_i berisi CF dari sub-kelompok yang diwakili oleh anak tersebut. Simpul daun memiliki entri L dengan format $\{CF_i\}$, di mana i berkisar dari 1 hingga L. Selain itu, simpul daun memiliki dua petunjuk terkait: *prev* dan *next*, yang menunjuk ke simpul daun sebelumnya dan berikutnya dalam pohon, memudahkan navigasi saat melewati simpul-simpul daun. Proses penyisipan titik data ke dalam *CF Tree* mirip dengan proses penyisipan dalam *B+ tree*. Pengembangan *CF Tree* dilakukan secara inkremental, artinya proses pemindaian data dan penempatannya ke dalam cluster yang sesuai dilakukan secara inkremental. Algoritma BIRCH diakui sebagai salah satu teknik pengelompokan pertama yang dapat menangani data yang mengandung *noise*. Algoritma ini terdiri dari total empat fase, di mana fase 2 dan fase 4 bersifat opsional (tidak selalu diperlukan) (Shauma, Purwanto, & Niovianty, 2019).

Algoritma BIRCH (*Balanced Iterative Reducing and Clustering Using Hierarchies*) memanfaatkan prinsip *Cluster Feature* (CF) untuk menggambarkan informasi tentang subkluster yang ada, sehingga mengurangi kompleksitas klusterisasi. Kelebihan Algoritma BIRCH terletak pada kemampuannya dalam menangani data yang besar dan memprosesnya secara akurat, dengan

menghilangkan banyak *outlier*. Pemilihan apakah suatu area dalam data dianggap sebagai outlier atau subkluster didasarkan pada kepadatan area tersebut. Area yang padat dengan data akan dianggap sebagai subkluster, sedangkan area yang jarang dianggap sebagai *outlier*. Dalam hierarki klasterisasi, data dipartisi secara berurutan ke dalam beberapa level/tingkatan. Ketika level meningkat, data dikelompokkan secara hierarkis. Pendekatan BIRCH memungkinkan klasterisasi data secara efisien dengan menggunakan Cluster Feature untuk mengumpulkan ringkasan informasi mengenai subkluster, yang kemudian digunakan untuk membangun struktur hierarkis (dendrogram) dari subkluster tersebut. (Shauma, Purwanto, & Niovianty, 2019).

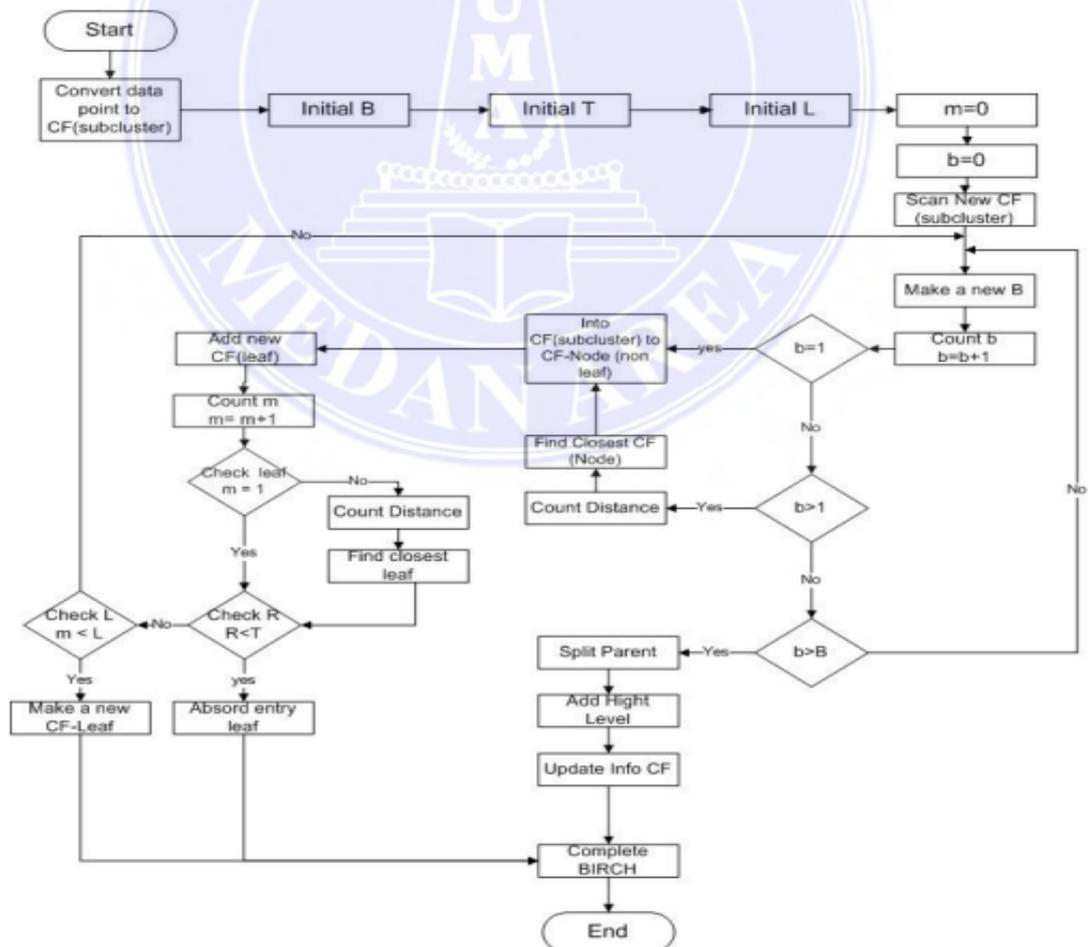
Pada fase pertama, algoritma BIRCH melakukan pemindaian *linier* terhadap seluruh titik data untuk membangun inisialisasi *CF Tree*. Setiap titik data diisikan ke dalam pohon dengan mencari simpul daun yang paling dekat. Jika simpul daun tersebut dapat menampung titik data baru tanpa melampaui ambang batas (*threshold*), nilai CF untuk inputan ini diperbarui. Jika tidak, sebuah simpul daun baru akan dibuat. Dalam kasus di mana sebuah simpul daun berisi lebih dari L entri setelah proses penyisipan, simpul daun tersebut kemungkinan akan dipecah.

Fase kedua bersifat opsional, dan tujuannya adalah mengurangi ukuran *CF Tree* hingga jumlah simpul daun yang diinginkan tercapai. Pada tahap ini, outlier yang lebih banyak dikeluarkan, dan subkluster yang lebih padat digabungkan ke dalam kluster yang lebih besar.

Fase ketiga melibatkan penggunaan algoritma *clustering* hierarkis untuk mengelompokkan nilai CF dari simpul-simpul daun. Data *point* dapat disesuaikan menjadi subkluster, di mana setiap subkluster diwakili oleh vektor CF. Biasanya, algoritma clustering dapat langsung diterapkan pada subkluster-subkluster ini karena informasi dalam vektor CF biasanya sudah cukup untuk menghitung jarak di antara mereka.

Fase keempat, yang juga bersifat opsional, melibatkan penggabungan kluster atau pengelompokkan tambahan, dan tergantung pada tujuan analisis yang lebih lanjut. (Shauma, Purwanto, & Niovianty, 2019).

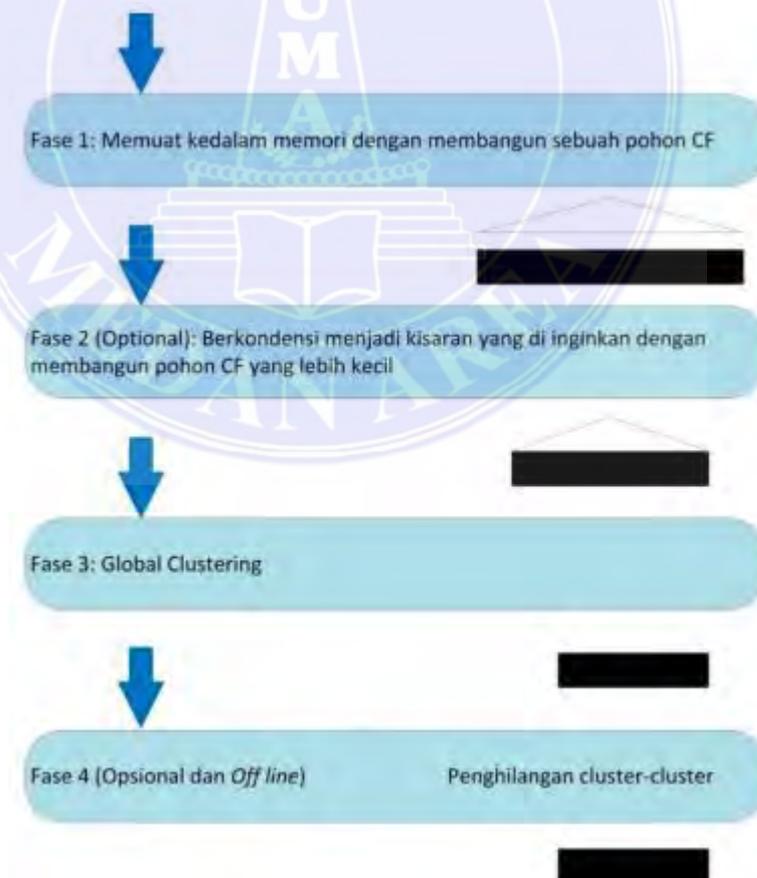
Adapun *flowchart* pada Algoritma BIRCH dapat dilihat pada gambar 2.1 (Venkatkumar & Shardaben, 2019).



Gambar 2.1 *Flowchart* Algoritma BIRCH

Algoritma BIRCH memiliki beberapa keuntungan yaitu dengan hanya BIRCH pemindaian tunggal dapat melakukan yang baik pengelompokkan dataset dan dengan pemindaian tambahan lebih lanjut dapat meningkatkan kualitas, tidak perlu menginput data set terlebih dahulu karena sudah mengikuti pendekatan bertahap, dan keputusan pengelompokkan tidak dibuat dengan mempertimbangkan semua titik data dan semua *cluster*. Semua bagian dari data point yang sangat banyak diikumpulkan sebagai satu *cluster* sedangkan outlier dihapus (Venkatkumar & Shardaben, 2019).

Namun, apabila algoritma BIRCH dikelompokkan per fase, maka dapat digambarkan seperti terlihat pada gambar 2.2.



Gambar 2.2 Skema Fase pada BIRCH

Sumber: (Venkatkumar & Shardaben, 2019)

1. Fase 1 Memuat Kedalam Memori dengan Membangun Sebuah Pohon CF

Data yang tersedia discan dan sebuah CF *tree* di dalam memori dibangun. Titik data yang sangat banyak di *subclusterkan* (dikelompokkan) dan kelompokkan pencilan (*Outliers*) akan dihapus. Saat memindai dan memasukkan data ke dalam *tree* jika kita kehabisan memori maka nilai ambang (*Threshold* (T)) awal adalah meningkat dan *tree* (*Tree*) CF baru yang lebih kecil dibangun lagi dan *Outliers* dihapus.

Dalam penerapannya pada fase 1, algoritma BIRCH memperkenalkan dua konsep yaitu *clustering feature* (CF) dan *clustering feature tree* (CF *tree*) yang digunakan untuk menggambarkan ringkasan *cluster*.

a. *Clustering Feature*

Clustering feature merupakan informasi mengenai *subcluster* dari obyek data. CF juga merupakan seperangkat dari tiga statistik ringkasan yang mewakili satu set titik data dalam satu *cluster*. Meringkas *cluster* menggunakan *clustering feature* dapat menghindari menyimpan informasi secara rinci mengenai objek data. Pada algoritma ini hanya perlu ukuran ruang yang konstan untuk menyimpan setiap *clustering fearture* yang dibangun, ini adalah kunci mengapa BIRCH dikatakan memiliki space yang efisien dan sangat efektif. CF memiliki rumus sebagai berikut:

$$CF = (N, LS, SS) \quad (1)$$

$$LS = \sum_{P_i \in N} P_i^\omega \quad (2)$$

$$SS = \sum_{P_i \in N} |P_i^\omega|^2 \quad (3)$$

Keterangan:

N = Jumlah data di *cluster*.

LS = Jumlah koordinat data linear pada N.

SS = Jumlah koordinat kuadrat dari data.

Salah satu mekanisme dari algoritma BIRCH melibatkan penggabungan cluster dalam kondisi tertentu. Konsep Aditivitas Teorema menyatakan bahwa CF (Cluster Feature) untuk dua cluster dapat digabungkan dengan menambahkan item dalam pohon CF masing-masing. Cluster feature dapat digabungkan dengan cluster yang memiliki jarak yang dekat selama masih memenuhi threshold yang ditentukan. Sebagai contoh, jika terdapat cluster CF1 = (N, LS, SS) dan CF2 = (N, LS, SS), cluster feature baru dapat terbentuk dengan menggabungkan CF1 dan CF2. dengan rumus:

$$CF_{12} = N_1 + N_2, LS_1 + LS_2, SS_1 + SS_2 \quad (4)$$

b. *Clustering Feature Tree (CF-Tree)*

BIRCH menggunakan konsep *Cluster feature* (CF) untuk meringkas informasi tentang setiap *subcluster* yang diatur dalam *height balance tree* yang disebut *CF-tree*. *Clustering feature tree* merupakan sebuah *height balanced tree* yang menyimpan *clustering feature* untuk pengelompokkan hirarki. Sebuah *clustering feature tree* memiliki 2 parameter, yaitu *factor branching* (B) dan nilai *threshold* (T). *Branching factor* adalah jumlah anggota maksimum *child*. Sedangkan *Threshold* (T) merupakan nilai ambang batas untuk diameter atau jarak dalam *cluster* pada daun (*leaf*). Setiap non *leaf note* mengandung

sebagian masukan B dalam bentuk $[CF_i, child_i]$ dimana $i = 1, 2, \dots, B$, “ $child_i$ ” adalah petunjuk ke- i untuk *child node* dan CF_i adalah *Clustering Feature* dari *subcluter* diwakili oleh *child*. Jadi *non-leaf node* mewakili *cluster* yang terdiri dari semua *subcluster* yang diwakili oleh masukannya (*entry*). Selain itu, setiap *leaf node* memiliki dua petunjuk (*pointer*) yaitu “previous” dan “next” yang dilakukan untuk menyatukan semua *leaf node* secara Bersama-sama untuk melakukan scan data secara efisien. Sebuah *leaf node* juga merupakan *cluster* yang terdiri dari semua *subcluster* yang ditunjukkan oleh setiap masukannya. Tapi semua *entry* dalam *leaf node* harus memenuhi persyaratan *Threshold* yaitu nilai jaraknya (diameter atau radius) harus kurang dari nilai T. Ukuran *tree* adalah fungsi *threshold*. Semakin besar *threshold* maka semakin kecil *tree* yang dihasilkan.

Langkah-langkah pembangunan CF-Tree adalah sebagai berikut:

- 1) Untuk setiap *subcluster* CF yang diberikan, BIRCH membandingkan lokasi catatan itu dengan lokasi dari masing-masing CF di *root node*, menggunakan Euclidean distance. BIRCH meneruskan *subcluster* CF masuk ke *root node* CF paling dekat dengan *subcluster* CF.
- 2) Catatan kemudian turun ke *node child non-leaf* dari simpul-simpul CF dipilih langkah a. BIRCH membandingkan lokasi catatan dengan lokasi setiap CF *non-leaf*. BIRCH meneruskan catatan yang masuk ke simpul *non-leaf* CF yang paling dekat dengan catatan yang masuk.
- 3) Catatan kemudian turun ke *node child leaf* dari simpul *non-leaf* CF yang dipilih pada langkah b. BIRCH membandingkan lokasi catatan dengan

lokasi setiap daun. BIRCH untuk sementara melewati catatan yang masuk ke daun yang paling dekat dengan *subcluster* CF yang masuk.

- 4) Lakukan salah satu (a) atau (b):
 - a) Sebuah. Jika jari-jari dari *leaf* yang dipilih termasuk catatan baru tidak melebihi *Threshold* T, maka *subcluster* CF yang masuk ditugaskan ke *leaf* itu. *Leaf* dan semua CF induknya diperbarui untuk memperhitungkan titik data baru.
 - b) Jika jari-jari *leaf* yang dipilih termasuk *subcluster* CF melebihi *Thershold* T, maka daun baru terbentuk, terdiri dari catatan yang masuk saja. CF induk diperbarui ke akun untuk titik data baru.

Sekarang jika langkah (b) dieksekusi dan sudah ada maksimum *leaf* L dalam *leaf node*, maka *leaf node* dibagi menjadi dua *leaf node*. *Node* daun paling jauh CF digunakan sebagai benih *leaf node*, dengan CF tersisa ditugaskan pada *leaf node* mana yang lebih dekat. Jika simpul induk penuh, pisahkan simpul induk, dan seterusnya. Setiap *leaf node* CF dapat dilihat sebagai *subcluster*. Pada langkah *cluster*, *subcluster* ini akan digabungkan menjadi *cluster*. Untuk *cluster* yang diberikan, biarkan centroid *cluster* menjadi:

$$D2 = \frac{\sqrt{(N_1SS_2)+(N_2SS_1)-2LS_1LS_2}}{N_1N_2} \quad (5)$$

Dengan radiusnya:

$$R = \frac{\sqrt{SS-LS^2}}{n} \quad (6)$$

2. Fase 2 berkondensasi menjadi kisaran yang diinginkan dengan membangun pohon CF yang lebih kecil (optional)

Fase 2 yang bersifat opsional bertindak sebagai celah penghubung antara fase 1 dan fase 3. Memindai entri simpul daun (*leaf node*) dengan demikian memadatkan titik data ke dalam *cluster* yang lebih besar dan menghapus *outlier* yang singkatnya menggabungkan data dan membangun bahkan *Tree CF* yang lebih kecil.

3. Fase 3 Global clustering

Fase 3 bertindak sebagai memperbaiki yang tidak diinginkan efek dari urutan *input* yang miring dan pemisahan yang disebabkan oleh *paging*. Dalam fase ini pengelompokan global atau semi-global algoritma diterapkan pada *CF leaf-node entries*. Algoritma yang digunakan disini adalah pengelompokan hierarkis aglomeratif.

4. Fase 4 Penghilangan cluster-cluster (optional dan off line)

Fase ini merupakan fase opsional dan juga bisa disebut sebagai fase pemurnian *cluster*. Setelah fase 3 mungkin ada beberapa ketidakakuratan yang bisa diatasi dengan melakukan tambahan melewati dataset dan menugaskan kembali titik ini ke *centroid* terdekat diperoleh dari fase 3. Dengan demikian menyempurnakan klaster lebih lanjut. Kompleksitas waktu dari algoritma ini adalah $O(N)$.

2.4 Metrik Evaluasi untuk Algoritma Clustering

Beberapa metrik evaluasi yang digunakan untuk mengevaluasi metode BIRCH adalah metode ARI, AMI dan *V-Measure*.

1. Metode *Adjusted Rank Index* (ARI)

Adjusted Rank Index (ARI) adalah ukuran kesamaan antara dua pengelompokan data. Metode ini adalah metode perbaikan dan metode *Rank Index*, yang merupakan ukuran dasar kesamaan antara dua pengelompokan, namun memiliki kelemahan karena sensitif terhadap peluang. *Adjusted Rank Index* memperhitungkan fakta bahwa beberapa kesepakatan antara dua pengelompokan dapat terjadi secara kebetulan, dan *Adjusted Rank Index* disesuaikan untuk memperhitungkan kemungkinan ini. Proses perhitungannya adalah sebagai berikut:

- a. Anggap N adalah jumlah sampel pada *dataset*.
- b. Anggap C1 dan C2 adalah dua buah cluster berbeda pada *dataset*.
- c. Anggap a adalah jumlah pasangan dari sampel yang berada pada cluster yang sama pada C1 dan C2.
- d. Anggap b adalah jumlah pasangan dari sampel yang berbeda cluster pada C1 dan C2.
- e. Hitung nilai Rand Index RI sebagai $RI = (a + b) / N^2$.
- f. Hitung nilai *Adjusted Rand Index* ARI sebagai:

$$ARI = \frac{\sum_{ij} \binom{n_{ij}}{2} - \left[\sum_i \binom{a_i}{2} \sum_j \binom{b_j}{2} \right] / \binom{n}{2}}{\frac{1}{2} \left[\sum_i \binom{a_i}{2} + \sum_j \binom{b_j}{2} \right] - \left[\sum_i \binom{a_i}{2} \sum_j \binom{b_j}{2} \right] / \binom{n}{2}}$$

where n_{ij}, a_i, b_j are values from the contingency table.

Semakin tinggi nilai ARI, maka semakin dekat dua buah *cluster* satu sama lain.

Nilai ARI berada dalam interval nilai dari -1 sampai 1, dimana 1 mengindikasikan bahwa kedua cluster sama persis, 0 mengindikasikan sebuah

persetujuan acak dan -1 mengindikasikan bahwa kedua *cluster* sama sekali berbeda.

2. Metode *Adjusted Mutual Information* (AMI)

Adjusted Mutual Information (AMI) adalah penyesuaian skor *Mutual Information* yang memperhitungkan peluang. Metode ini mengoreksi efek kesepakatan semata-mata karena kebetulan antar pengelompokan, mirip dengan cara *Adjusted Rank Index* mengoreksi *Rank Index*. Hal ini erat kaitannya dengan variasi informasi. Namun ukuran yang disesuaikan tidak lagi bersifat metrik. Untuk dua buah *cluster* U dan V, rumusan dari metode *Adjusted Mutual Information* (AMI) dapat dijabarkan sebagai berikut:

$$AMI(U, V) = \frac{MI(U, V) - E(MI(U, V))}{avg(H(U), H(V)) - E(MI(U, V))}$$

3. Metode *V-Measure*

V-Measure adalah rata-rata harmonik antara *homogeneity* dan *completeness*.

Rumusan dari *V-Measure* adalah sebagai berikut:

$$v = (1 + \beta) * \text{homogeneity} * \text{completeness} / (\beta * \text{homogeneity} + \text{completeness})$$

2.5 Penelitian Terdahulu

Beberapa penelitian sejenis yang pernah dilakukan sebelumnya dapat dirincikan sebagai berikut:

Tabel 2.1 Penelitian Terdahulu

Pengarang (Tahun)	Judul	Keterangan
Zulfi Azhar (2019)	Penentuan Penempatan Karyawan Baru di	Penggunaan metode SMART (Simple Multi

		<p>PDAM Kisaran dengan Metode SMART</p>	<p>Attribute Rating Technique) dalam sistem pendukung ini akan meningkatkan efektivitas dalam menentukan penempatan karyawan baru, sekaligus memberikan dukungan bagi Departemen Sumber Daya Manusia dalam proses pengambilan keputusan.</p>
<p>Irma Setiawati, Gunawan Abdillah, Asep Id Hadiana (2019)</p>		<p>Sistem Pendukung Keputusan Penempatan Karyawan Berdasarkan Test Dominant-Influence-Steady-Compliance (DISC) Menggunakan Metode Profile Matching</p>	<p>Studi ini bertujuan untuk mengembangkan sebuah sistem pendukung keputusan untuk penempatan karyawan berdasarkan hasil uji DISC. Hasil tes DISC diambil sebagai parameter pendukung dan diolah menggunakan metode Profile Matching untuk mendapatkan</p>

		perangkingan karyawan. Hasil akhir dari penelitian ini adalah rekomendasi karyawan yang sesuai dengan karakteristik jabatan, dengan tingkat akurasi pengujian mencapai 85,8%.
Iftah Nur Fadlilah (2022)	Pengelompokan Menggunakan Data Tweet Kecelakaan Pendekatan Text Mining dan Algoritma BIRCH	Penelitian ini menggunakan salah satu algoritma dalam data mining yang terintegrasi yakni algoritma BIRCH yang sebelumnya melalui proses pendekatan text mining. Algoritma BIRCH merupakan algoritma yang penemuan kelompok yang bagus dengan hanya menggunakan satu kali scan data. Hasil yang diperoleh dari

		<p>penelitian ini adalah data yang diperoleh dari kaggle menghasilkan 1545 cluster dengan silhouette coefficient bernilai 0.1159964638217295 sedangkan dari data hasil crawling twitter menghasilkan 487 cluster dengan silhouette coefficient bernilai 0.7262655918349612.</p>
--	--	---

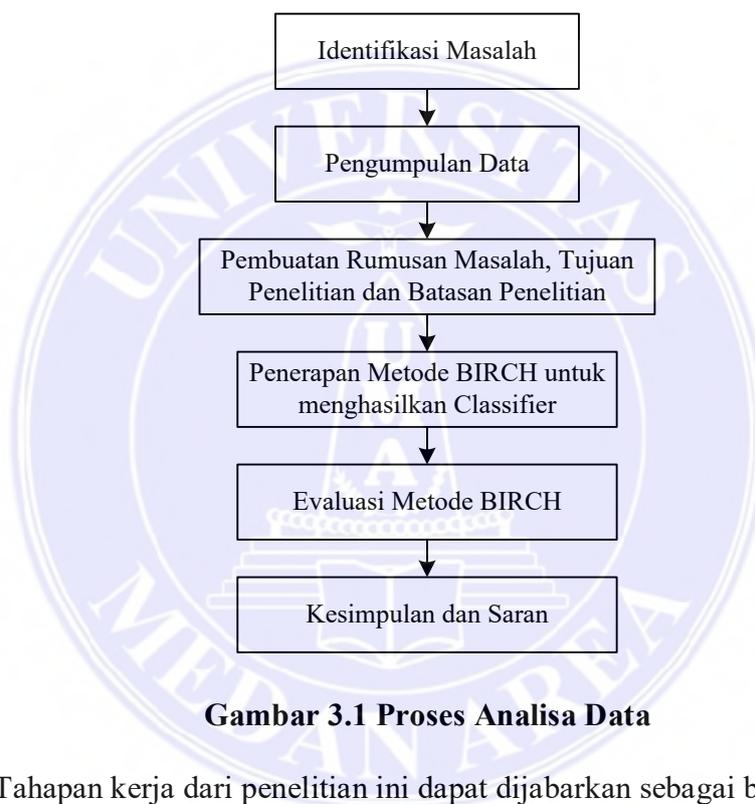


BAB III

METODE PENELITIAN

3.1 Tahapan Penelitian

Tahapan penelitian dalam penyusunan laporan penelitian ini secara garis besar digambarkan pada Gambar 3.1.



Gambar 3.1 Proses Analisa Data

Tahapan kerja dari penelitian ini dapat dijabarkan sebagai berikut:

1. Identifikasi Masalah

Langkah pertama dari penelitian ini adalah mengidentifikasi masalah yang akan diselesaikan.

2. Pengumpulan Data

Pengumpulan data pada penelitian ini diambil dari perusahaan Podomoro.

Bahan penelitian yang digunakan dalam proses penelitian ini berasal dari studi

jurnal dan buku pendukung. Dalam melakukan penelitian ini, alat yang digunakan untuk keberhasilan penelitian adalah sebagai berikut :

1. Kebutuhan Perangkat Keras (*Hardware*)

Dalam merancang dan menguji sistem, *hardware* yang digunakan adalah satu set lengkap perangkat komputer yang memiliki spesifikasi sebagai berikut:

- a. Intel Core i3.
- b. RAM 4 GB.
- c. Harddisk 2 TB.
- d. Monitor LCD dengan resolusi layar minimal 1024 x 768.
- e. Keyboard dan Mouse.

2. Kebutuhan Perangkat Lunak (*Software*)

Adapun perangkat lunak untuk merancang program ini adalah:

- a. Sistem operasi Windows 7/8/10.
- b. Microsoft Word 2016.

3. Pembuatan Rumusan Masalah, Tujuan Penelitian dan Batasan Penelitian.

Pada tahapan ini, akan disusun rumusan masalah, tujuan penelitian dan batasan dari penelitian ini.

4. Penerapan Metode BIRCH untuk menghasilkan *Classifier*.

Tahapan selanjutnya, adalah menerapkan metode BIRCH untuk menghasilkan *classifier* yang akan digunakan untuk melakukan proses klasifikasi karyawan.

5. Evaluasi Metode BIRCH.

Classifier yang diperoleh pada tahapan sebelumnya akan digunakan untuk melakukan proses segmentasi pelanggan. *Dataset* pengujian akan digunakan untuk mengevaluasi *classifier* yang diperoleh dari metode BIRCH.

6. Kesimpulan dan Saran.

Tahapan terakhir dari penelitian ini adalah menarik kesimpulan dari hasil pengujian yang dilakukan dan memberikan saran untuk penelitian selanjutnya.

3.2 Metode Pengumpulan Data

Untuk melakukan pengumpulan data yang diperlukan dalam penelitian ini, penulis menggunakan beberapa metode pengumpulan data, yaitu:

1. Metode Pengamatan (Observasi)

Penulis melakukan pengamatan dengan cara mengamati dan melakukan uji coba beberapa data perhitungan pengambil keputusan yang ada sebagai panduan dalam penelitian ini. Selain itu, penulis juga mencari *dataset* yang cocok untuk digunakan sebagai studi kasus dalam penelitian ini.

2. Metode Studi Pustaka

Penulis melakukan pengumpulan data melalui internet maupun buku yang sesuai dengan topik yang dibahas. Penulis akan mengumpulkan berbagai jurnal ilmiah dari *internet* terutama mengenai metode BIRCH.

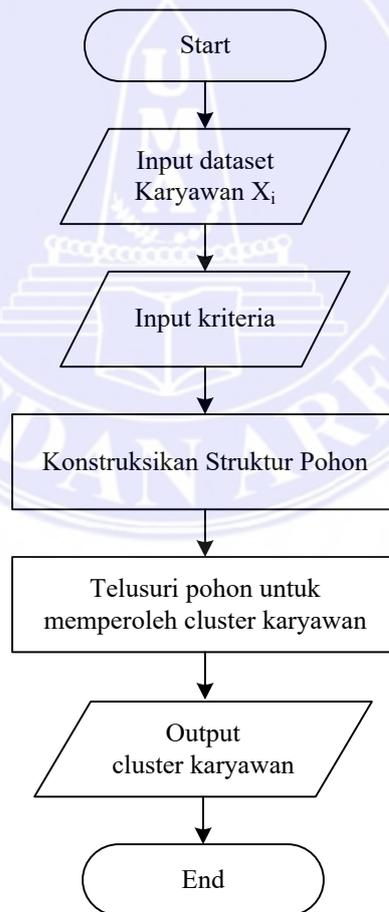
3.3 Proses Kerja Metode BIRCH

Metode BIRCH merupakan salah satu metode *clustering* yang menerapkan konsep *cluster feature* (CF). Metode BIRCH memerlukan data *input* berupa

dataset karyawan, ukuran ketidaksamaan (mencakup pengalaman kerja karyawan dan tingkat pendidikan) dan parameter *input* kriteria. Hasil *output* dari metode BIRCH adalah label *cluster* untuk setiap elemen dalam *dataset* karyawan.

Proses kerja dari metode BIRCH akan dimulai dari pengisian *dataset* karyawan dan kriteria. Setelah itu, sistem akan mengkonstruksikan struktur pohon untuk setiap karyawan. Setelah itu, menelusuri pohon untuk menghasilkan *cluster* dari karyawan. Terakhir, sistem akan menghasilkan *cluster* (kelompok) dari karyawan.

Prosedur kerja dari metode BIRCH dapat digambarkan dalam bentuk *flowchart* seperti terlihat pada gambar berikut:



Gambar 3.2 *Flowchart* dari Metode BIRCH

Agar dapat lebih memahami proses kerja dari metode BIRCH dalam melakukan proses penempatan karyawan, maka diberikan sebuah contoh sederhana berikut ini:

Misalkan diketahui terdapat 4 orang karyawan dengan perincian data sebagai berikut:

Tabel 3.1 Data Karyawan

Titik	Kode	Nama	Penilaian	Penilaian
	Karyawan	Karyawan	Keahlian	Kinerja
1	B0001-94	NURMIATI	80	60
2	B0002-96	GUSTIANA	70	70
3	B0003-97	MISNANI	60	80
4	B0004-97	SUPRIATIK	70	50

Kriteria yang digunakan dalam melakukan pengelompokan dapat dirincikan sebagai berikut:

Tabel 3.2 Data Kriteria

Kode Kriteria	Nama Kriteria	Bobot
K1	Keahlian	1
K2	Kinerja	1

Proses kerja dari metode BIRCH adalah sebagai berikut:

1. Lakukan proses normalisasi terhadap data terlebih dahulu.

Proses normalisasi akan dilakukan terhadap data karyawan pada tabel 3.1 berdasarkan kriteria penilaian pada tabel 3.2.

Tabel 3.3 Data Karyawan Setelah Konversi Nilai

Titik	Kode Karyawan	Nama Karyawan	K1	K2
1	B0001-94	NURMIATI	80	60
2	B0002-96	GUSTIANA	70	70
3	B0003-97	MISNANI	60	80
4	B0004-97	SUPRIATIK	70	50
Nilai Maksimum			80	80

Proses normalisasi:

Kode Karyawan: B0001-97 (NURMIATI)

K1: nilai K1/nilai maksimum = $80/80 = 1$

K2: nilai K2/nilai maksimum = $60/80 = 0.75$

Proses perhitungan akan dilakukan terhadap karyawan lainnya. Hasil yang diperoleh adalah sebagai berikut:

Tabel 3.4 Data Karyawan Setelah Normalisasi

Titik	Kode Karyawan	Nama Karyawan	K1	K2
1	B0001-94	NURMIATI	1	0.75
2	B0002-96	GUSTIANA	0.88	0.88
3	B0003-97	MISNANI	0.75	1
4	B0004-97	SUPRIATIK	0.88	0.63

2. Konversi data *point* menjadi *cluster feature* (CF).

CF = (N, LS, SS), dimana N = jumlah karyawan dalam *cluster*, LS = nilai kriteria, SS = LS^2

Contoh:

Kode Karyawan: B0001-97 (NURMIATI)

$$K1: 1^2 = 1$$

$$K2: 0.75^2 = 0.5625$$

Setiap karyawan akan dimasukkan ke setiap *cluster* yaitu karyawan 1 ke CF_1 , karyawan 2 ke CF_2 , karyawan 3 ke CF_3 dan karyawan 4 ke CF_4 .

$$CF = [N, LS, SS]$$

$$CF_1 = [1; (1, 0.75); (1, 0.5625)]$$

$$CF_2 = [1; (0.88, 0.88); (0.7744, 0.7744)]$$

$$CF_3 = [1; (0.75, 1); (0.5625, 1)]$$

$$CF_4 = [1; (0.88, 0.63); (0.7744, 0.3969)]$$

3. Tentukan nilai B, T dan L.

Misalkan nilai $B = 2$, $T = 1.5$ dan $L = 2$.

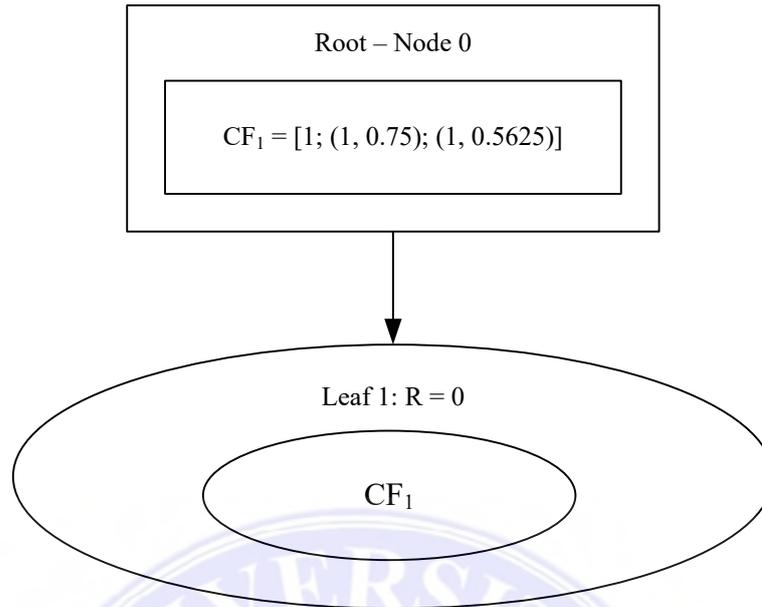
4. Karyawan pertama akan masuk kedalam *root* dan membentuk di *CF Node Non leaf* dan masuk ke *leaf*.

$CF_1 = [1; (1, 0.75); (1, 0.5625)]$ dimasukkan sebagai *root*.

5. Hitung radius dari *cluster*.

$$R = \sqrt{(1 - 1^2) + (0.5625 - \lfloor 0.75 \rfloor^2)} = 0$$

$R < T$, maka *node* dapat dibuat.



Gambar 3.3 Struktur CF Tree Setelah Pengisian Data 1

Saat ini, telah diperoleh sebuah *cluster* yang berisi karyawan 1.

6. Setelah karyawan CF_1 masuk, kemudian karyawan CF_2 akan mulai masuk juga kedalam struktur pohon CF dan akan ada pembaharuan informasi mengenai CF. Setelah bergabung dengan CF maka akan dilakukan pengecekan terhadap radius untuk bergabung dengan CF-Leaf. Karyawan 1 akan digabungkan dengan karyawan 2 ke dalam sebuah *cluster*. Setelah itu, akan dilakukan pengecekan nilai radius dari hasil gabungan tersebut. Apabila nilai radius $<$ nilai T, maka kedua karyawan tersebut boleh digabung menjadi satu cluster. Jika tidak, maka akan dibuat sebuah *cluster* baru yang berisi karyawan kedua.

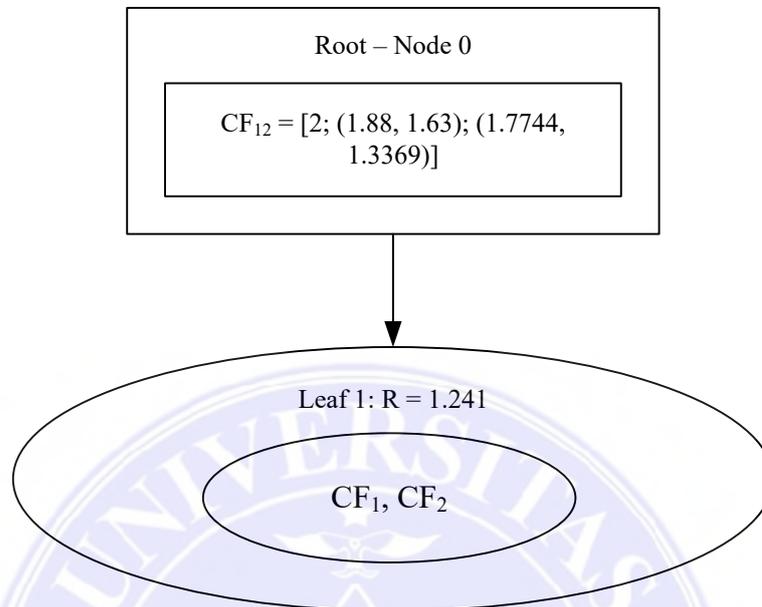
$$CF_{12} = CF_1 + CF_2$$

$$CF_{12} = [2; (1.88, 1.63); (1.7744, 1.3369)]$$

7. Hitung radius dari *cluster*.

$$R = \sqrt{((1.7744 - 1.88)^2) + ((1.3369 - 1.63)^2)} / 2 = 1.241$$

Karena $R < T$, maka CF_1 dan CF_2 dapat digabungkan. Hal ini berarti bahwa karyawan 1 dan karyawan 2 dapat digabung menjadi satu *cluster*.



Gambar 3.4 Struktur CF Tree Setelah Pengisian Data 2

8. Setelah karyawan CF_1 dan CF_2 , kemudian karyawan CF_3 akan mulai masuk juga ke dalam struktur pohon CF dan akan ada pembaharuan informasi mengenai CF. Tetapi setelah bergabung dengan CF maka akan di lakukan pengecekan terhadap radius untuk bergabung dengan CF-Leaf

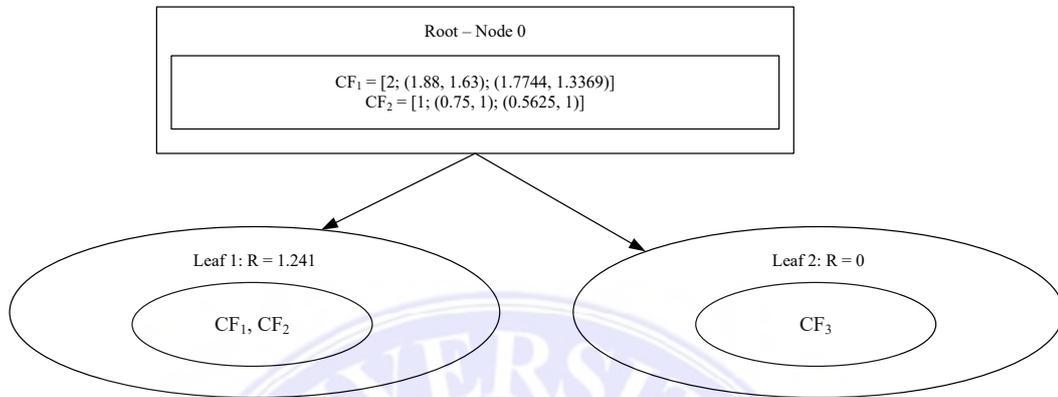
$$CF_{123} = [3; (2.63, 2.63); (2.3369, 2.3369)]$$

$$R = \sqrt{((2.3369 - 2.63)^2) + (2.3369 - [2.63])^2)}/3 = 1.747$$

Karena $R > T$, maka CF_{12} dan CF_3 tidak dapat digabungkan. Hal ini berarti bahwa karyawan ketiga tidak dapat dimasukkan ke dalam *cluster* yang sama dengan karyawan pertama dan kedua, sehingga akan dibuat sebuah *cluster* baru yang berisi karyawan ketiga.

Disini nilai R sudah melebihi nilai T sehingga perlu ada *split* pada CF Node parent. Informasi CF akan di perbaharui menjadi CF yang semula CF_{12} karena

sudah menjadi satu *cluster* akan berubah menjadi CF_1 dan menampung seluruh ringkasan dari subcluster. Kemudian akan terbentuk CF_2 hasil *split parent*.



Gambar 3.5 Struktur CF Tree Setelah Pengisian Data 3

9. Setelah karyawan CF_1 , CF_2 dan CF_3 , kemudian karyawan CF_4 akan mulai masuk juga ke dalam struktur pohon CF dan akan ada pembaharuan informasi mengenai CF. Tetapi karena sudah mempunyai 2 buah *cluster* (yaitu *cluster* pertama yang berisi karyawan pertama dan kedua, serta *cluster* kedua yang berisi karyawan ketiga), maka harus ditentukan terlebih dahulu, karyawan keempat akan dimasukkan ke dalam *cluster* yang mana. Caranya adalah dengan menghitung nilai jarak terdekat dengan menggunakan rumus *distance* D_2 . *Cluster* yang dipilih adalah *cluster* yang memiliki nilai D_2 terkecil. Setelah ini dapat bergabung dengan CF-*leaf* yang memenuhi nilai *threshold*.

Hitung jarak subcluster CF_4 dengan CF-Node(CF_1)

$$D_2 = \frac{\sqrt{N_1SS_2 + N_2SS_1 - 2LS_1LS_2}}{N_1N_2}$$

$$2SS_2 + SS_1 = (2.8994, 3.3369)$$

$$2LS_1LS_2 = (2.82, 3.26)$$

$$D_2 = \frac{\sqrt{(2.8994 - 2.82) + (3.3369 - 3.26)}}{2 * 1}$$

$$D_2 = 0.19767$$

Hitung jarak subcluster CF₄ dengan CF-Node(CF₂)

$$SS_2 + SS_1 = (1.3369, 1.3969)$$

$$2LS_1LS_2 = (0.945, 2)$$

$$D_2 = \frac{\sqrt{(1.3369 - 0.945) + (1.3969 - 2)}}{1 * 1}$$

$$D_2 = 0.4595$$

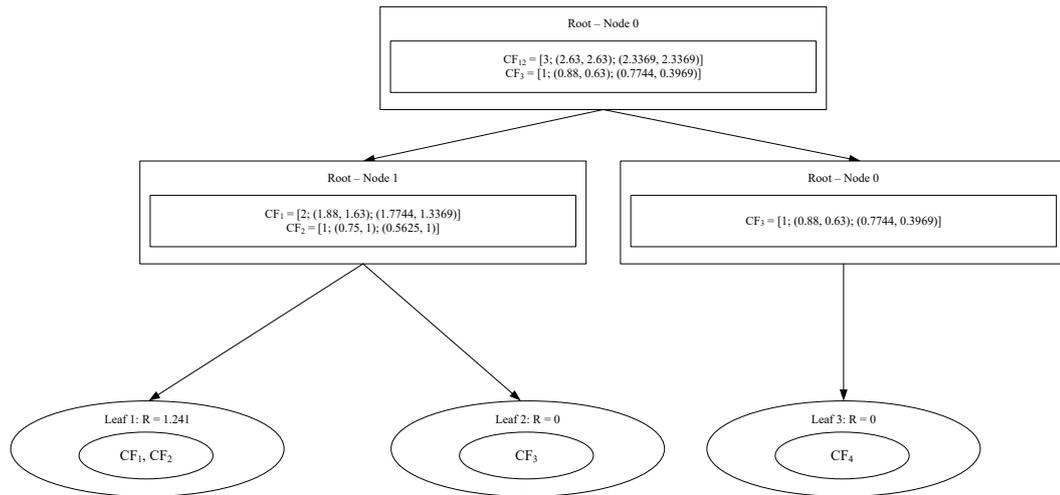
10. Dari perhitungan di atas, didapatkan bahwa karyawan CF₄ lebih dekat dengan *cluster* pertama CFNode (CF₁), karena nilai D₂ ke *cluster* pertama yang paling kecil, sehingga karyawan CF₄ akan masuk ke *cluster* pertama CF₁. Kemudian lakukan perhitungan radius.

$$CF_{124} = [3; (2.76, 2.26); (2.5488, 1.7338)]$$

$$R = \sqrt{((2.5488 - 2.76)^2) + (1.7338 - [2.26])^2)} / 3 = 1.6775$$

Karena nilai R sudah melebihi T, yang berarti bahwa karyawan keempat tidak dapat digabungkan ke dalam *cluster* CF₁, sehingga perlu dibuat sebuah *cluster* CF baru lagi.

Tetapi *leaf* sudah di batasi nilainya 2 (L=2) yang berarti tidak boleh ada penambahan *leaf* ketiga. Sehingga perlu *split parent root node*.



Gambar 3.6 Struktur CF Tree Setelah Pengisian Data 4

11. Dari hasil struktur *CF Tree* diatas, maka dapat diperoleh informasi berikut:

Cluster-1: B0001-94 (NURMIATI) dan B0002-96 (GUSTIANA)

Cluster-2: B0003-97 (MISNANI)

Cluster-3: B0004-97 (SUPRIATIK)

3.4 Hipotesis

Pada penelitian ini, akan dilakukan proses klasifikasi karyawan menjadi beberapa segmen yaitu divisi *marketing*, *accounting* dan HRD. Jenis segmentasi yang digunakan adalah proses segmentasi berdasarkan nilai. Adapun beberapa variabel *input* yang digunakan mencakup pengalaman kerja, tingkat pendidikan, kesehatan fisik dan mental, nilai kerjasama, nilai kerajinan dan nilai kebersihan.

3.5 Metode Penyelesaian

Langkah penyelesaian dari proses klasifikasi karyawan dengan metode BIRCH dapat dijabarkan sebagai berikut:

1. Proses pembacaan *dataset*.

Langkah pertama adalah membaca *dataset* yang akan digunakan dalam proses *training* dan *testing*. Proses pembacaan *dataset* akan dilakukan dengan membaca *file* dari *local drive*. Adapun *file dataset* yang digunakan adalah *file .csv*. Untuk memilih *file* dari komputer, maka dapat menggunakan perintah berikut:

```
from google.colab import files
uploaded = files.upload()
```

Perintah diatas akan membuka kotak dialog untuk pemilihan *file csv* yang akan digunakan. *User* dapat mengklik tombol *Choose Files* dan pilih *file csv* yang diinginkan. Setelah itu tunggu sampai *file* selesai di-*upload*. Terakhir, ketikkan *code* berikut ini untuk membaca dan menyimpan *dataset* ke variabel *X*.

```
import io
X = pd.read_csv(io.BytesIO(uploaded['data.csv']))
```

2. Proses pembuatan *classifier*.

Tahapan selanjutnya adalah membuat *classifier* dengan menggunakan metode BIRCH. Caranya adalah dengan menjalankan perintah berikut:

```
brc = Birch(branching_factor=50, n_clusters=None, threshold=1.5)
brc.fit(X)
```

3. Proses penentuan *cluster* (label) untuk setiap data.

Tahapan terakhir adalah menentukan *cluster* yang dihasilkan untuk setiap karyawan. Caranya adalah dengan menjalankan perintah berikut:

```
labels = brc.predict(X)
```

BAB V

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Dari pembahasan pada bab-bab sebelumnya, maka akhirnya penelitian pada skripsi ini dapat diambil beberapa kesimpulan, antara lain.

1. Metode BIRCH ini tidak memerlukan proses penentuan nilai awal untuk setiap *cluster* dan juga jumlah *cluster* yang akan digunakan dalam perhitungan, sehingga metode ini mampu memperbaiki kelemahan dari metode *Clustering* dan *K-Means Clustering*.
2. Proses penempatan karyawan dapat dilakukan dengan menentukan nilai dari setiap kriteria yang digunakan untuk setiap karyawan.
3. Hasil *clustering* terbaik akan diperoleh untuk nilai parameter *branching factor* = 30 dan nilai *threshold* = 2.

5.2 Saran

Dalam pembuatan sistem ini, tentu masih terdapat kekurangan. Oleh karena itu, ada beberapa saran yang dapat dijadikan bahan pertimbangan apabila ada yang berminat untuk mengembangkan skripsi ini.

1. Bagi peneliti lain yang ingin melakukan penelitian mengenai penempatan karyawan, maka disarankan untuk menggabungkan metode BIRCH dengan metode lainnya yang sejenis, seperti metode klasifikasi, agar dapat ditingkatkan akurasi dari hasil klasifikasi yang diperoleh.

2. Disarankan untuk membandingkan metode BIRCH dengan metode klasifikasi lainnya sehingga dapat diketahui kelebihan dan kelemahan setiap metode.



DAFTAR PUSTAKA

- Azhar, Z. (2019). Penentuan Penempatan Karyawan Baru di PDAM Kisaran dengan Metode SMART. *JURTEKSI (Jurnal Teknologi dan Sistem Informasi)*, IV(2), 179-184.
- Budiman, S. A., Safitr, D., & Ispriyanti, D. (2019). Perbandingan Metode K-Mean dan Metode DBSCAN Pada Pengelompokan Rumah Kost Mahasiswa di Kelurahan Tembalang Semarang. *Jurnal Gaussian*, 5(4), 757-762.
- Chusyairi, A., & Saputra, P. R. (2019). Pengelompokan Data Puskesmas Banyuwangi Dalam Pemberian Imunisasi Menggunakan Metode K-Means Clustering. *Telematika*, 12(2), 139-148.
- Dani, A. T., Wahyuningsih, S., & Rizki, N. A. (2019). Penerapan Hierarchical Clustering Metode Agglomerative pada Data Runtun Waktu. *Jambura Journal of Mathematics*, 1(2), 64-78.
- Fitrayana, P. R., & Saputro, S. D. (2022). Algoritme Clustering Large Application (CLARA) untuk Menangani Data Outlier. *PRISMA, Prosiding Seminar Nasional Matematika* 5, 5, 721-725.
- Khairiyah, Y., Febriani, & Yanti, B. (2020). Pengaruh Penempatan Kerja dan Mutasi Terhadap Kinerja Pegawai Rumah Sakit Universitas Andalas Padang. *Menara Ekonomi*, VI(2), 89-98.
- Marlina, D., Putri, N. F., Fernando, A., & Ramadhan, A. (2020). Implementasi Algoritma K-Medoids dan K-Means untuk Pengelompokan Wilayah Sebaran Cacat pada Anak. *Jurnal CoreIT*, 4(2), 64-71.
- Muliono, R., & Sembiring, Z. (2019). Data Mining Clustering Menggunakan Algoritma K-Means Untuk Klasterisasi Tingkat Tridarma Pengajaran Dosen. *CESS (Journal of Computer Engineering System and Science)*, 4(2), 272-279.
- Novianti, F., Yasmin, Y., & Novitasari, D. (2022). Penerapan Algoritma Fuzzy C-Means (FCM) dalam Pengelompokan Provinsi di Indonesia berdasarkan Indikator Penyakit Menular Manusia. *JUMANJI*, 6(1), 23-33.

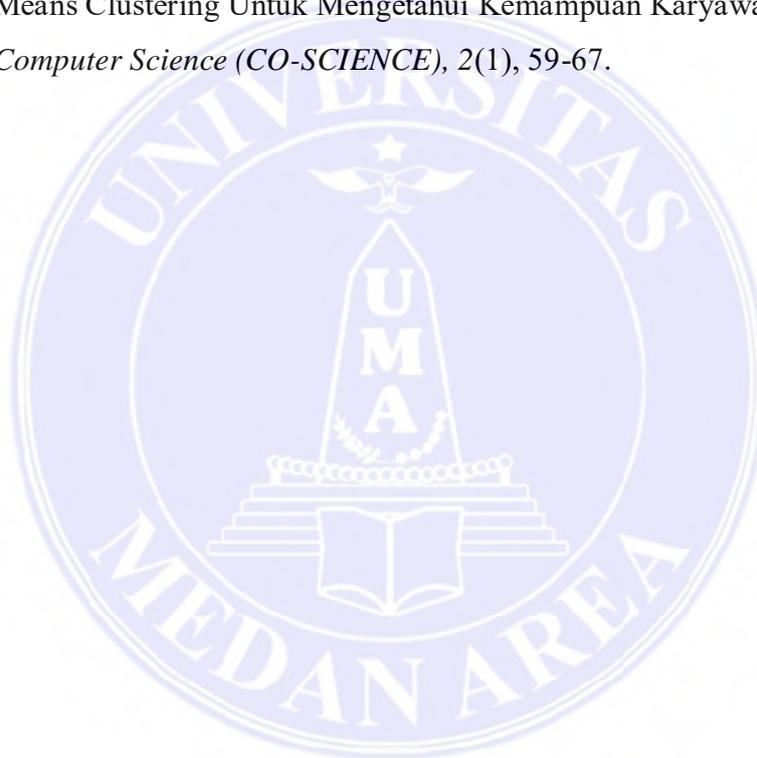
- Nur Fadlilah, I. (2022). Pengelompokan Data Tweet Kecelakaan Menggunakan Pendekatan Text Mining dan Algoritma Birch. *Universitas Islam Negeri Maulana Malik Ibrahim Malang*, 1-96.
- Onibala, A., Saerang, I., & Dotulong, L. (2019). Analisis Perbandingan Prestasi Kerja Karyawan Tetap Dan Karyawan Tidak Tetap di Kantor Sinode GMIM. *Jurnal EMBA*, 5(2), 380-387.
- Pramesti, D. F., Furqon, M. T., & Dewi, C. (2019). Implementasi Metode K-Medoids Clustering Untuk Pengelompokan Data Potensi Kebakaran Hutan/Lahan Berdasarkan Persebaran Titik Panas (Hotspot). *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 1(9), 723-732.
- Putri, A. H., Cahya, R., & Fauzi, M. (2019). Implementasi Algoritma Improved K-Means Pada Portal Jurnal Internasional. *Research Gate*, 1-9.
- Ramadhani, F. (2019). Cluster Big Data Dengan Balance Iterative Reducing Clustering Using Hierarchies (BIRCH). *Universitas Sumatera Utara*.
- Sahadewa, I., & Rahmawati, P. I. (2021). Pengaruh Penempatan Karyawan Dan Pengembangan Karier Terhadap Kinerja Karyawan Pada Handara Golf and Resort Bali. *Jurnal Manajemen Perhotelan dan Pariwisata (JMPP)*, 4(1), 10-20.
- Setiawati, I., Abdillah, G., & Hadiana, A. (2019). Sistem Pendukung Keputusan Penempatan Karyawan Berdasarkan Test Dominant-Influence-Steady-Compliance (DISC) Menggunakan Metode Profile Matching. *Prosiding SNST ke-7*, F(5), 25-30.
- Shauma, A., Purwanto, Y., & Niovianty, A. (2019). Deteksi Anomali Trafik Menggunakan Algoritma BIRCH dan DBSCAN menggunakan Streaming Traffic Anomaly Traffic Detection with BIRCH dan DBSCAN Algorithm for Streaming Traffic. *e-Proceeding of Engineering*, 5004-5012.
- Sitinjak, D. D., Maman, & Suwita, J. (2020). Analisa dan Perancangan Sistem Informasi Administrasi Kursus Bahasa Inggris Pada Intensive English Course di Ciledug Tangerang. *Jurnal IPSIKOM*, 1-19.
- Sutabri, T. (2019). *Analisis Sistem Informasi*. Yogyakarta: Andi.
- Tukiyat, & Djohan, Y. (2022). Analisis Penyebaran Pandemi Covid-19 di Kota Jakarta Menggunakan Metode Clustering K-Means dan Density Based

Spatial Clustering of Application With Noise. *Jurnal Informatika*, 9(1), 43-54.

Venkatkumar, I. A., & Shardaben, S. J. (2019). Comparative study of Data Mining Clustering algorithms. *IEEE International Conference on Data Science and Engineering (ICDSE)*, 1-7.

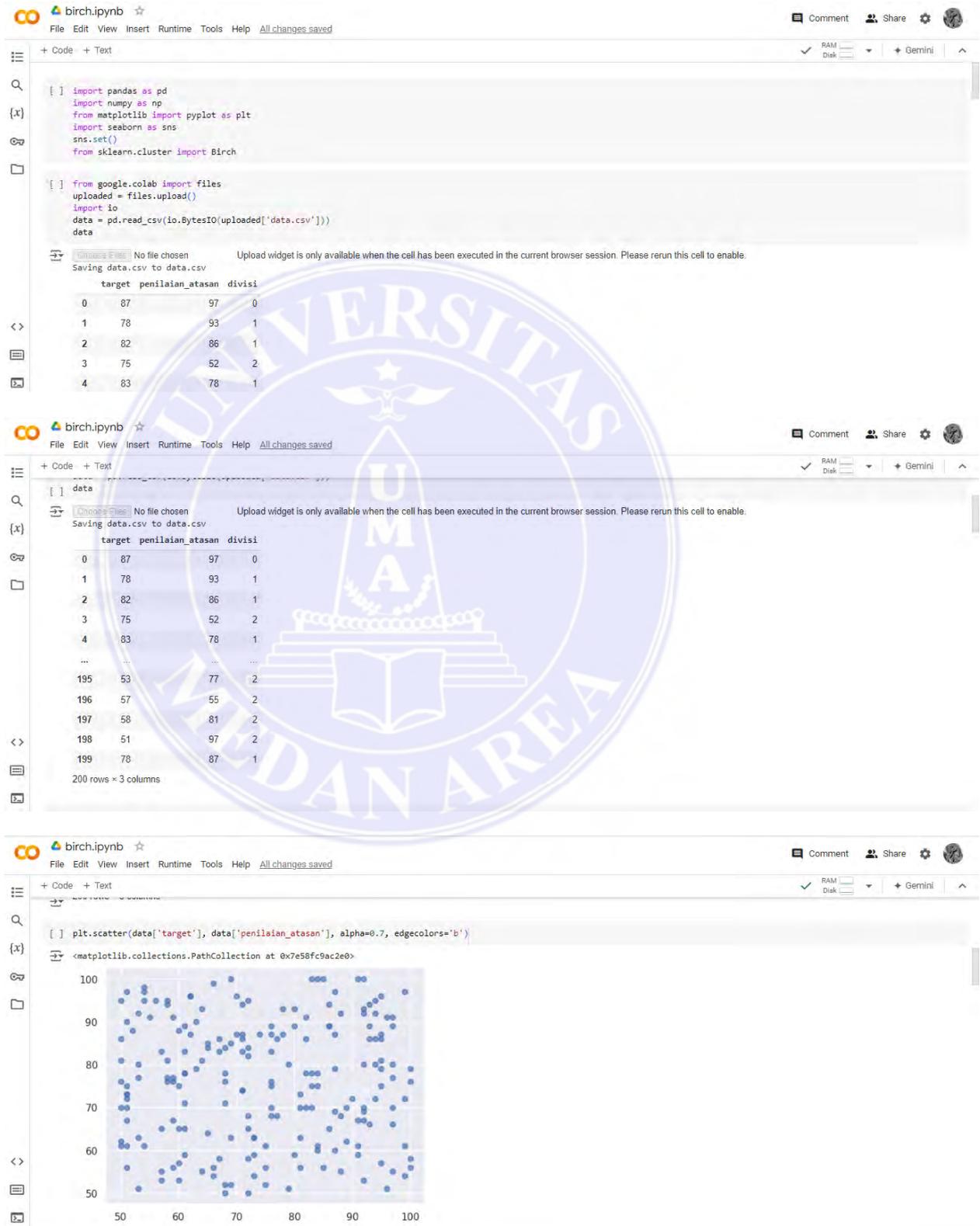
Yona, M. (2018). Pengaruh Penempatan Kerja, Keahlian (SKILL) Kepuasan Kerja Terhadap Peningkatan Kinerja Karyawan Pada PT. OSI ELECTRONICS BATAM. *Jurnal BENING*, 5(1), 27-38.

Zakiyah, D., Merlina, N., & Mayangky, N. A. (2022). Penerapan Algoritma K-Means Clustering Untuk Mengetahui Kemampuan Karyawan IT. *Computer Science (CO-SCIENCE)*, 2(1), 59-67.



LAMPIRAN-LAMPIRAN

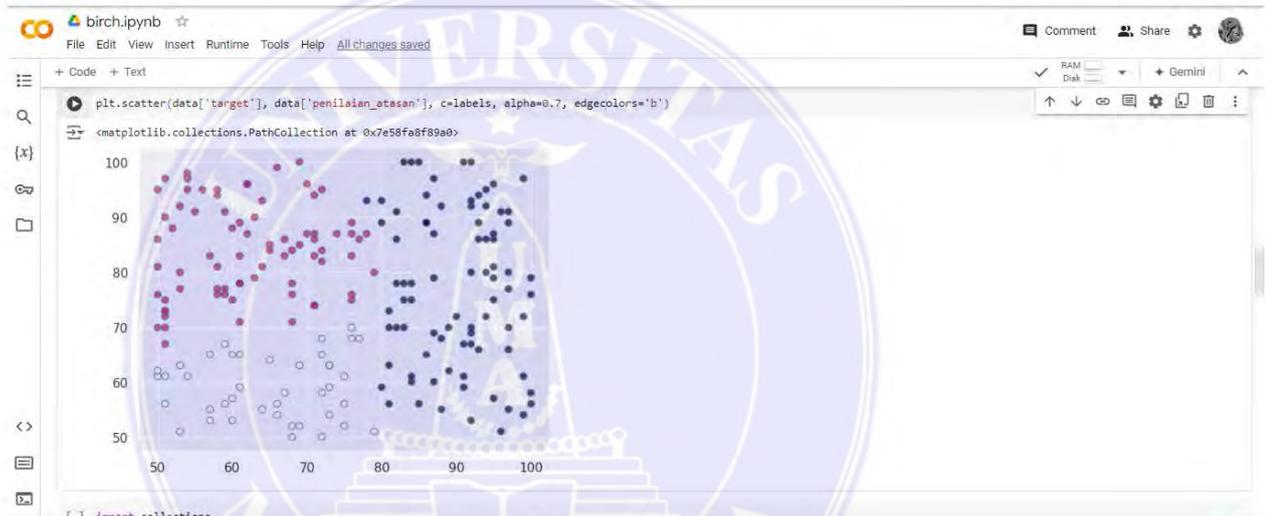
Lampiran 1 Coding Program



```

birch.ipynb
File Edit View Insert Runtime Tools Help All changes saved
+ Code + Text
[ ] brc = Birch(branching_factor=150, n_clusters=3, threshold=2)
brc.fit(data)
+ Birch
Birch(branching_factor=150, threshold=2)
[ ] Start coding or generate with AI.
[ ] labels = brc.predict(data)
labels
array([[0, 0, 0, 2, 0, 0, 0, 0, 1, 1, 0, 2, 0, 1, 0, 0, 0, 0, 2, 0, 1, 0,
0, 1, 1, 1, 0, 0, 1, 1, 1, 0, 2, 0, 0, 1, 0, 0, 0, 0, 0, 0, 1, 0,
1, 0, 1, 1, 0, 1, 1, 2, 1, 2, 0, 1, 0, 1, 1, 2, 0, 0, 0, 1, 0, 0,
0, 1, 0, 1, 0, 1, 0, 0, 1, 0, 0, 1, 1, 1, 0, 1, 2, 1, 1, 1, 0, 2,
1, 0, 1, 1, 0, 1, 2, 2, 2, 1, 1, 0, 0, 0, 2, 0, 2, 1, 0, 0, 1, 1,
0, 2, 1, 0, 0, 0, 0, 2, 0, 0, 0, 1, 0, 1, 0, 1, 2, 0, 0, 0,
0, 0, 2, 0, 1, 0, 1, 1, 2, 0, 2, 2, 1, 0, 2, 0, 1, 0, 2, 2, 2,
0, 0, 2, 0, 1, 0, 1, 1, 1, 0, 2, 0, 0, 1, 0, 1, 1, 0, 1, 0, 1, 1,
2, 0, 0, 1, 2, 1, 0, 2, 0, 1, 1, 1, 1, 2, 1, 0, 0, 0, 2, 1, 2, 1,
1, 1])
plt.scatter(data['target'], data['penilaian_otasan'], c=labels, alpha=0.7, edgecolors='b')
<matplotlib.collections.PathCollection at 0x7e58fa8f89a0>

```



```

birch.ipynb
File Edit View Insert Runtime Tools Help All changes saved
+ Code + Text
[ ] import collections
counter = collections.Counter(labels)
counter
Counter({2: 37, 1: 84, 0: 79})
[ ] from sklearn.metrics import adjusted_rand_score
ari = adjusted_rand_score(data['divisi'], labels)
ami = metrics.adjusted_mutual_info_score(data['divisi'], labels)
v_measures = metrics.v_measure_score(data['divisi'], labels)
print(ari)
print(ami)
print(v_measures)
0.814579224265027592
1.0
1.0
[ ] from sklearn import metrics
score_funcs = [
    ("V-measure", metrics.v_measure_score),
    ("ARI", metrics.adjusted_rand_score),
    ("AMI", metrics.adjusted_mutual_info_score),
]

```

```
birch.ipynb
File Edit View Insert Runtime Tools Help All changes saved
+ Code + Text
[ ]
n_samples = data.count()
n_classes = 3
plots = []
names = []

sns.color_palette("colorblind")
plt.figure(1)

def fixed_classes_uniform_labelings_scores(
    score_func, n_samples, n_classes, n_runs=3, n_threshold=4
):
    scores = np.zeros((n_threshold, n_runs))
    labels_a = data['divisi']
    n_threshold_range = np.array([0.5, 1, 1.5, 2])

    for i in range(n_threshold):
        for j in range(n_runs):
            brc = Birch(branching_factor=50, n_clusters=3, threshold=n_threshold_range[i])
            brc.fit(data)
            labels_b = labels
            scores[i, j] = score_func(labels_a, labels_b)

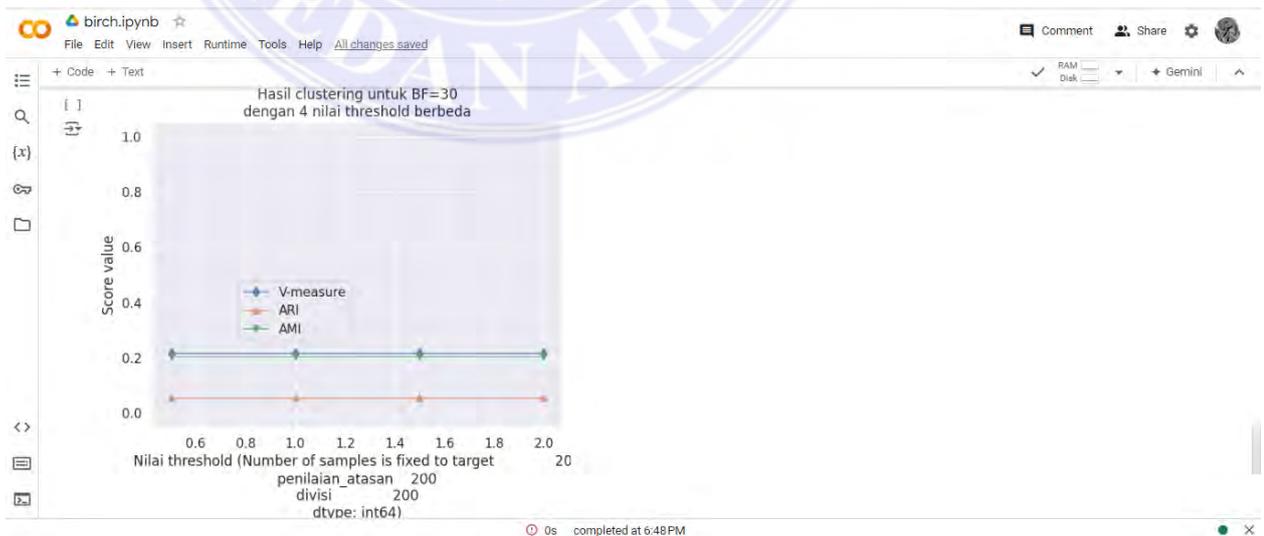
    return scores

for marker, (score_name, score_func) in zip("d^vx.", score_funcs):
    scores = fixed_classes_uniform_labelings_scores(
```

```
birch.ipynb
File Edit View Insert Runtime Tools Help All changes saved
+ Code + Text
[ ]
for marker, (score_name, score_func) in zip("d^vx.", score_funcs):
    scores = fixed_classes_uniform_labelings_scores(
        score_func, n_samples, n_classes=3
    )
    n_threshold_range = np.array([0.5, 1, 1.5, 2])
    plots.append(
        plt.errorbar(
            n_threshold_range,
            scores.mean(axis=1),
            scores.std(axis=1),
            alpha=0.8,
            linewidth=1,
            marker=marker,
        )[0]
    )
    names.append(score_name)

plt.title(
    "Hasil clustering untuk BF=30\n"
    "dengan 4 nilai threshold berbeda"
)

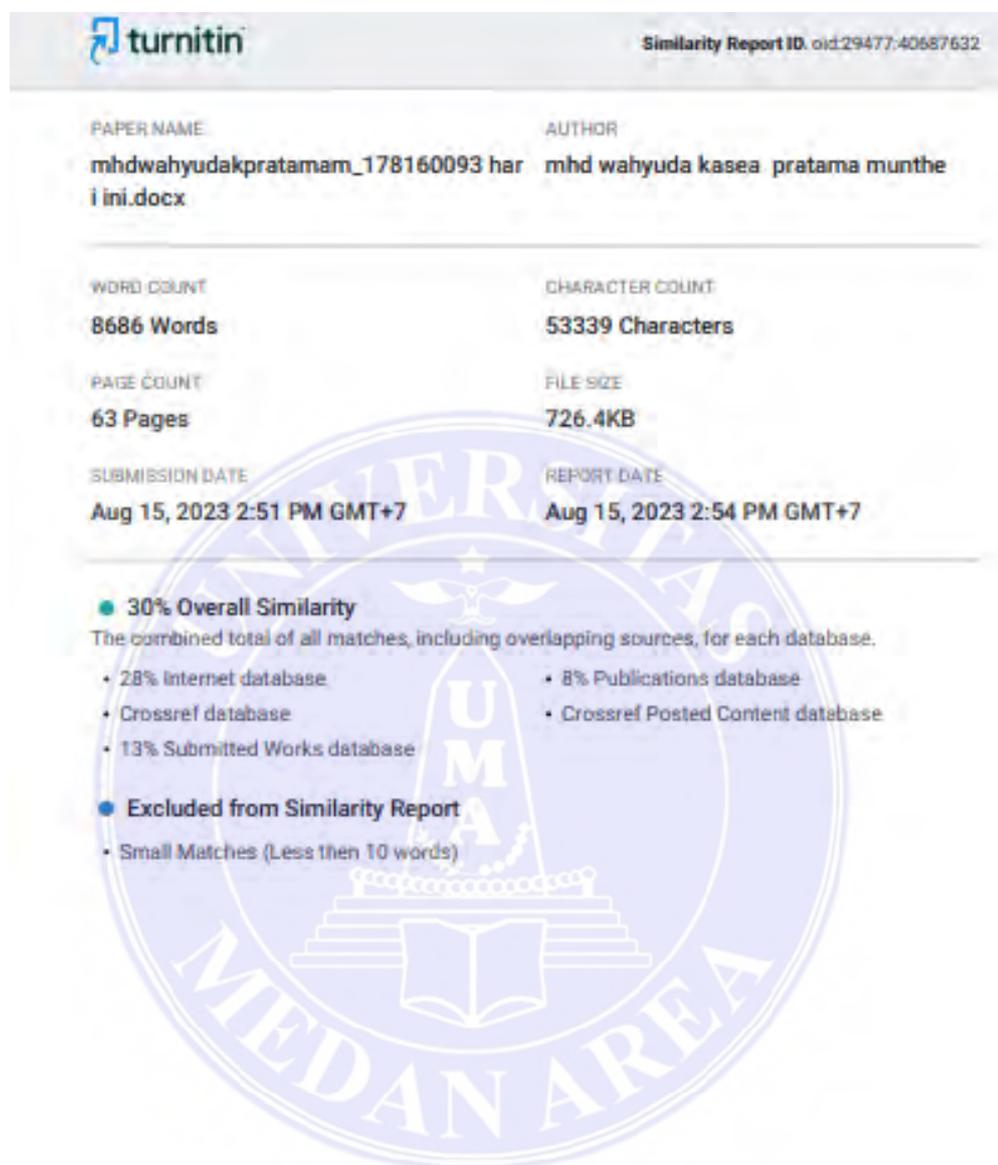
plt.xlabel(f"Nilai threshold (Number of samples is fixed to {n_samples})")
plt.ylabel("Score value")
plt.ylim(bottom=0.05, top=1.05)
plt.legend(plots, names, bbox_to_anchor=(0.5, 0.5))
plt.show()
```



Lampiran 2 SK Pembimbing Skripsi

	UNIVERSITAS MEDAN AREA FAKULTAS TEKNIK	
<small>Kampus I : Jalan Kolam Nomor 1 Medan Estate/Jalan PBSI Nomor 1 ☎ (061) 7366878, 7360168, 7364348, 7366781, Fax, (061) 7366998 Medan 20223 Kampus II : Jalan Setiabudi Nomor 79 / Jalan Sei Serayu Nomor 70 A, ☎ (061) 8225602, Fax, (061) 8226331 Medan 20122 Website: www.teknik.uma.ac.id E-mail: univ_medanarea@uma.ac.id</small>		
Nomor	: 382 /FT.6/01.10/V/2023	23 Mei 2023
Lamp	: -	
H a l	: Pergantian Dosen Pembimbing	
<p>Yth. Pembimbing Tugas Akhir Muhathir, ST, M.Kom di Tempat</p>		
<p>Dengan hormat, sehubungan dengan adanya perubahan dosen pembimbing pada SK pembimbing nomor 246/FT.6/01.10/VIII/2022 pada tanggal 15 Agustus 2022 maka perlu diterbitkan kembali SK Pembimbing Skripsi baru atas nama mahasiswa berikut :</p>		
N a m a	: Mhd. Wahyuda Kasea Pratama Munthe	
N P M	: 178160093	
Jurusan	: Teknik Informatika	
<p>Oleh karena itu kami mengharapkan kesediaan saudara :</p>		
1. Muhathir, ST, M.Kom	(Sebagai Pembimbing)	
<p>Adapun Tugas Akhir Skripsi berjudul :</p>		
<p>“Sistem Penilaian Karyawan Berdasarkan Keahlian dan Kinerja dengan Metode <i>BIRCH</i> di PT. Sinar Menara Deli (Podomoro Grup)”</p>		
<p>SK Pembimbing ini berlaku selama enam bulan terhitung sejak SK ini diterbitkan. Jika proses pembimbing melebihi batas waktu yang telah ditetapkan, SK ini dapat ditinjau ulang.</p>		
<p>Demikian kami sampaikan, atas kesediaan saudara diucapkan terima kasih.</p>		
<p style="text-align: right;">  Dr. Rahmad Syah, S. Kom, M. Kom</p>		

Lampiran 3 Cek turnitin



The image shows a Turnitin Similarity Report for a document titled 'mhdwahyudakpratamam_178160093 har i ini.docx'. The report includes the following details:

PAPER NAME	AUTHOR
mhdwahyudakpratamam_178160093 har i ini.docx	mhd wahyuda kasea pratama munthe

WORD COUNT	CHARACTER COUNT
8686 Words	53339 Characters

PAGE COUNT	FILE SIZE
63 Pages	726.4KB

SUBMISSION DATE	REPORT DATE
Aug 15, 2023 2:51 PM GMT+7	Aug 15, 2023 2:54 PM GMT+7

30% Overall Similarity
The combined total of all matches, including overlapping sources, for each database.

- 28% Internet database
- 8% Publications database
- Crossref database
- Crossref Posted Content database
- 13% Submitted Works database

Excluded from Similarity Report

- Small Matches (Less than 10 words)

The background of the report features a large, faint watermark of the Universitas Medan Area logo, which includes a book and the text 'UNIVERSITAS MEDAN AREA'.