**JURUSAN TEKNIK INFORMATIKA**

**FAKULTAS TEKNOLOGI INFORMASI**

**INSTITUT TEKNLOGI SEPULUH NOPEMBER**

**USULAN TUGAS AKHIR**

# IDENTITAS PENGUSUL

**NAMA : Ramadhan Rosihadi Perdana**

**NRP : 5112100032**

**DOSEN WALI : Diana Purwitasari, S.Kom, M.Sc.**

**DOSEN PEMBIMBING : 1. Dr. Chastine Fatichah, S.Kom, M.Kom  
 2. Rully Soelaiman, S.Kom., M.Kom.**

# JUDUL TUGAS AKHIR

“Implementasi Algoritma *Incremental Overlapping* *Clustering*  Berbasis *Tree*”

# LATAR BELAKANG

Dengan bertambah pesatnya informasi yang terus diperbaharui dalam waktu singkat, sehingga objek dalam informasi tersebut dapat dikelompokkan ke dalam beberapa kelompok berbeda sekaligus.Proses pengelompokkan ini disebut dengan istilah *clustering* atau klasterisasi.Dan metode yang digunakan untuk dapat mengelompokkan objek ke dalam beberapa kelompok berbeda sekaligus dinamakan *overlaping clustering.*

Pada proses klasterisasi biasanya digunakan *dataset* yang bersifat statik, dimana tidak ada proses modifikasi berupa penambahan,penghapusan,ataupun pengubahan data. Sehingga proses analisis dapat diterapkan langsung terhadap keseluruhan *dataset.*Ketika dalam proses klasterisasi terdapat banyak perubahan *dataset,* maka dilakukan proses *reclustering* terhadap keseluruhan data tersebut. Namun jika *dataset* berukuran besar dan laju pertambahan ukurannya sangat cepat, proses klasterisasi ulang ini tidaklah efisien.

Untuk mengatasi permasalahan klasterisasi ulang yang tidak efisien ketika dilakukan proses klaterisasi ulang pada *dataset* berukuran besar, maka proses *reclustering* dilakukan secara *incremental* atau secara bertahap sehingga proses *clustering* dapat dilakukan secara terus menerus tanpa harus memproses ulang seluruh *dataset.*

Permasalahan *incremental overlapping clustering* diatas dapat diselesaikan dengan menggunakan metode TIOC-TWD karena pada metode ini tidak dilakukan proses *recluster* semua data secara keseluruhan ketika terdapat proses modifikasi pada datadata,namun hanya sebagian data saja yang akan di *recluster* dan sekaligus mampu menggolongkan objek dalam suatu *dataset* kedalam lebih dari satu kategori.

Pada tugas akhir ini, penulis melakukan penelitian mengenai penggunaan algoritma TIOC-TWD ( *A tree-based incremental overlapping clustering method using the three-way decision theory* ) untuk proses klasterisasi *dataset*  terkait.

# RUMUSAN MASALAH

Rumusan masalah yang diangkat dalam tugas akhir ini adalah sebagai berikut :

1. Bagaimana mengimplementasikan algoritma TIOC-TWD ke dalam program?
2. Bagaimana kualitas *cluster* yang dihasilkan oleh algoritma TIOC-TWD saat sebelum dan sesudah proses *update*?
3. Bagaimana hasil kinerja algoritma TIOC-TWD dalam hal akurasi dan kecepatan proses?

# BATASAN MASALAH

1. Implementasi algoritma menggunakan bahasa pemrograman C++.
2. Proses *incremental* dilakukan ketika terdapat modifikasi pada *dataset.*

# TUJUAN PEMBUATAN TUGAS AKHIR

1. Menganalisis efisiensi algoritma TIOC-TWD pada proses dalam hal waktu eksekusi.
2. Mengukur dan menganalisis kualitas klaster yang dihasilkan algoritma TIOC-TWD saat sebelum dan sesudah terjadi proses modifikasi data.

# MANFAAT TUGAS AKHIR

Pengerjaan tugas akhir ini diharapkan dapat memberikan solusi dari permasalahan *incremental overlapping clustering* untuk tujuan pengelompokkan data.

# TINJAUAN PUSTAKA

Dalam mengerjakan Tugas Akhir ini, terdapat beberapa tinjauan pustaka yang digunakan, yaitu :

1. ***Incremental overlapping clustering***

*Incremental overlapping clustering* adalah proses menentukan *cluster* pada suatu data yang terus dimodifikasi dan objek yang di *cluster* dapat dikelompokkan pada lebih dari satu kelompok Terdapat beberapa penelitian yang merumuskan metode untuk *incremental clustering* atau *overlapping clustering*, namun sedikit penelitian yang mencoba merumuskan gabungan dari kedua model klasterisasi tersebut.

Pada tahun 1998, untuk pertama kalinya dirumuskan metode yang menangani permasalahan *incremental clustering* bernama *Incremental* DBSCAN dengan berpacu pada algoritma klasterisasi DBSCAN [1] . Kemudian pada tahun 2009 terdapat penelitian yang mengklaim memiliki kinerja lebih baik dalam hal kecepatan proses jika dibandingkan dengan algoritma *Incremental* DBSCAN meskipun *cluster* yang dihasilkan sama [2]. Setelah itu di tahun 2012 dipublikasikan metode *incremental clustering* yang berdasarkan jarak pada bentuk *cluster* yang tidak beraturan, namun algoritma pada tahapannya perlu untuk menscan keseluruhan *dataset* [3].Satu tahun kemudian,pada tahun 2013 dirumuskan algoritma *incremental clustering* yang dapat memaksimalkan hubungan jarak antar pola pada *cluster* yang sama [4]. Dan pada tahun 2014 dipublikasikan algoritma *incremental co-clustering* yang menghitung pembagian objek dan pembagian fitur secara simultan, namun tidak dapat mengklaster objek kedalam lebih dari satu kelompok [5].

Permasalahan *incremental overlapping clustering* juga telah dikaji oleh para akademisi dan menghasilkan metode-metode yang dapat digunakan untuk menyelesaikan permasalahan *incremental overlapping clustering*.

1. ***Three-way decision clustering***  
    *Three-way decision clustering* adalah metode klasterisasi yang mengambil keputusan untuk pembentukan klaster dengan tiga opsi, yaitu :
   1. Menerima sebuah objek sebagai bagian dalam sebuah kriteria jika tingkat kecocokan kriteria objek tersebut lebih besar atau sama dengan level yang ditetapkan.
   2. Menolak sebuah objek sebagai bagian dalam sebuah kriteria jika tingkat kecocokan kriteria objek tersebut berada di bawah level yang ditetapkan
   3. Tidak menerima atau menolak sebuah objek atau bisa dikatakan tidak memberikan komitmen objek tersebut termasuk kriteria atau tidak.

Model *three-way decision clustering* digunakan untuk *overlapping clustering* sedangkan untuk *crisp clustering* metode yang dipakai dikategorikan sebagai *two-way decision clustering*. Berbeda dengan *crisp clustering* yang memodelkan klasternya sebagai himpunan dari objek-objek, pada model *three-way decision clustering* ini *cluster* di representasikan dalam bentuk interval.

Contoh : Ck = [ Cklower , Ckupper]  
 Yang mana Cklower merupakan batasan bawah dari *cluster* Ck dan Ckupper merupakan batasan atas dari *cluster* Ck.

# RINGKASAN ISI TUGAS AKHIR

## Deskripsi dan Gambaran Sistem

Sistem yang akan dibangun dalam Tugas Akhir ini berupa aplikasi desktop sederhana yang berfungsi untuk melakukan proses *clustering* sekumpulan data dengan menggunakan metode *Tree based incremental overlapping clustering method using the three-way decision theory* (TIOC-TWD). Data masukan berupa *dataset,* kemudian dilakukan *preprocessing* dan dilanjutkan dengan proses *clustering.*

Berikut urutan proses-proses yang dilakukan sistem digambarkan dalam bentuk *flow diagram* dari awal (*input*) sampai akhir (*output*) dari sistem.

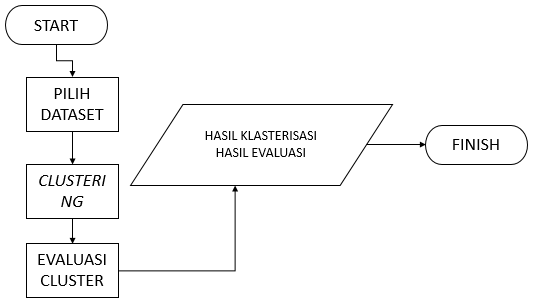


Figure 1 *Flow Diagram* Sistem

**9.1.1 Proses Clustering**

Setelah menyiapkan *dataset* yang akan diklaster, dilanjutkan dengan tahapan inti yaitu *clustering*, dengan metode TIOC-TWD.

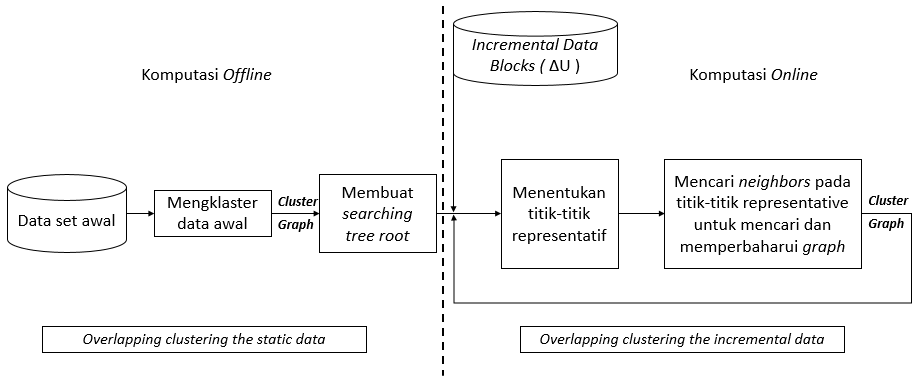


Figure 2 Proses metode TIOC-TWD

* + - 1. ***Clustering* *dataset* awal**

Pada tahapan pertama, *dataset* awal sebelum terjadi modifikasi apapun di*cluster* menggunakan *three-way decisions* untuk proses *static overlapping clustering.*Adapun ­sub proses pada proses *clustering* data awal adalah sebagai berikut :

* + - 1. Menghitung jarak antar objek menggunakan *Euclidean Distance.*
      2. Menentukan titik-titik representatif. Titik representatif merupakan titik fiksi, tidak mewakili objek dalam sistem. Suatu titik dikatakan representatif apabila kepadatan data pada wilayah titik tersebut lebih besar atau sama dengan nilai *density threshold* yang ditetapkan.
      3. Membangun *undirected graph* berdasarkan titik-titik representatif yang telah didapatkan dengan menggunakan konsep dari metode *three-way decision.*
      4. **Membangun *searching tree***

Pemikiran dasar dari metode TIOC-TWD ini adalah untuk merepresentasikan *incremental data* sebagai titik-titik representatif terlebih dahulu, yang mana hal ini memberikan keuntungan untuk menghemat waktu komputasi dibandingkan dengan metode yang berdasarkan objek.Berangkat dari hal ini maka dibangunlah *searching tree* untuk menyimpan titik-titik representatif,sehingga memungkinkan untuk mencari dan memperbaharui operasi-operasi pada stuktur *tree.*

* + - 1. ***Clustering incremental data***

Setelah dilakukan modifikasi terhadap data bersangkutan, maka dilanjutkan dengan proses *clustering incremental data.*Metode TIOC-TWD tidak perlu untuk meng*cluster* keseluruhan data lagi,namun hanya data yang mengalami perubahan saja ( ∆U ).Pada proses ini ditentukan titik-titik representatif didalam ∆U, kemudian mencari dan memperbaharui hubungan antar *graph* yang telah dibuat pada proses sebelumnya.

Pokok pemikiran dari proses ini adalah untuk mencari data-data yang berdekatan dengan titik-titik representatif dan memperbaharui wilayah yang bersangkutan pada *tree* dan *graph.*

* + 1. **Evaluasi *Clustering***

Pada tahapan ini dilakukan pengukuran terhadap kualitas *cluster* yang dihasilkan oleh algoritma TIOC-TWD.Adapun metode pengukuran yang digunakan adalah *Silhoutte Coefficient.* Dengan metode ini, akan diukur nilai *silhoutte coefficient* tiap *cluster* dan nilai rata-rata dari semua *cluster*, dari nilai *silhoute coefficient* ini akan diketahui bagaimana kualitas dari *cluster.*

# METODOLOGI

## Penyusunan Proposal Tugas Akhir

Tahap pertama dalam proses pengerjaan Tugas Akhir ini adalah menyusun proposal Tugas Akhir. Pada proposal tugas akhir ini diajukan sebuah permasalahan *cluster* model tumpang tindih pada data set yang terus diperbaharui dalam waktu singkat yang akan diselesaikan menggunakan algoritma yang berkaitan dengan permasalahan tersebut.

Proposal Tugas Akhir ini berisi tentang deskripsi pendahuluan dari tugas akhir yang akan dibuat. Pendahuluan terdiri atas hal yang menjadi latar belakang diajukannya usulan tugas akhir, rumusan masalah yang diangkat, batasan masalah untuk tugas akhir, tujuan dari pembuatan tugas akhir, dan manfaat dari hasil pembuatan tugas akhir. Selain itu dijabarkan pula tinjauan pustaka yang digunakan sebagai referensi pendukung pembuatan tugas akhir. Sub bab metodologi berisi penjelasan mengenai tahapan penyusunan tugas akhir mulai dari penyusunan proposal hingga penyusunan buku tugas akhir. Terdapat pula sub bab jadwal kegiatan yang menjelaskan jadwal pengerjaan tugas akhir.

## Studi literatur

Pada studi literatur ini, akan dilakukan studi mendalam mengenai algoritma *search tree*, *three way decision,* dan algoritma-algoritma lain untuk menyelesaikan permasalahan *cluster* model tumpang tindih pada data set yang terus diperbaharui dalam waktu singkat. Adapun sumber studi yang dilakukan didapat dari buku-buku, papers, internet, serta materi perkuliahan yang berkaitan.

## Implementasi Algoritma

Implementasi algoritma adalah tahapan membangun program untuk meng*cluster* data set dengan *cluster* model tumpang tindih pada data set yang terus diperbaharui dalam waktu singkat. Implementasi algoritma yang dibangun menggunakan bahasa pemrograman C++ dengan menggunakan IDE Dev-C++ 5.11.

## Pengujian dan evaluasi

Tahap pengujian dan evaluasi akan dilakukan dengan menguji program yang sudah dibuat untuk meng*cluster* data set terkait. Evaluasi dilakukan dengan menghitung nilai akurasi dari *cluster* yang dihasilkan oleh program.

## Penyusunan Buku Tugas Akhir

Pada tahap ini dilakukan penyusunan laporan yang menjelaskan dasar teori dan metode yang digunakan dalam tugas akhir ini serta hasil dari implementasi aplikasi perangkat lunak yang telah dibuat. Sistematika penulisan buku tugas akhir secara garis besar antara lain:

1. Pendahuluan
   1. Latar Belakang
   2. Rumusan Masalah
   3. Batasan Tugas Akhir
   4. Tujuan
   5. Metodologi
   6. Sistematika Penulisan
2. Tinjauan Pustaka
3. Desain dan Implementasi
4. Pengujian dan Evaluasi
5. Kesimpulan dan Saran
6. Daftar Pustaka

# JADWAL KEGIATAN

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Tahapan** | **2015** | | | **2016** | | | |
| **Oktober** | **Nopember** | **Desember** | **Januari** | **Februari** | **Maret** | **April** |
| **Penyusunan Proposal** |  |  |  |  |  |  |  |
| **Studi Literatur** |  |  |  |  |  |  |  |
| **Perancangan Sistem** |  |  |  |  |  |  |  |
| **Implementasi** |  |  |  |  |  |  |  |
| **Pengujian dan Evaluasi** |  |  |  |  |  |  |  |
| **Penyusunan Buku** |  |  |  |  |  |  |  |

# DAFTAR PUSTAKA

|  |  |
| --- | --- |
|  |  |

|  |  |
| --- | --- |
| [1] | H.-P. K. J. S. M. W. X. X. M. Ester, "Incremental clustering for mining in a data warehousing environment," *VLDB, vol. 98,* p. 323–333, 1998. |
| [2] | P. G. K. V. P. D. P. S. N. Goyal, "An efficient density based incremental clustering algorithm in data warehousing environment," *2009 International Conference on Computer Engineering and Applications, IPCSIT,* vol. 2, p. 482–486, 2009. |
| [3] | O. V. R. L. S. N. K. B. B.K. Patra, "Distance based incremental clustering for mining clusters of arbitrary shapes, in: Pattern Recognition and Machine Intelligence," *Springer,* p. 229–236, 2013. |
| [4] | N. A. N. Y. M. I. R. Ibrahim, "Incremental mitosis: discovering clusters of arbitrary shapes and densities in dynamic data," *11th International Conference on Machine Learning and Applications,* vol. 1, no. IEEE Computer Society, pp. 102-107, 2012. |
| [5] | D. I. R. M. R.G. Pensa, "Hierarchical co-clustering: off-line and incremental approaches," *Data Mining Knowl,* vol. 1, pp. 31-64, 2014. |
| [6] | W. Wang, "A Two-Layer Night-Time Vehicle Detector," 2015. |
| [7] | B. Y. Pratama, "Saintek UIN Malang," [Online]. Available: http://saintek.uin-malang.ac.id/Mirror/ilmukomputer/Batra-Operasi-Morfologi-Pada-Citra-Biner.pdf. [Accessed 2015]. |
| [8] | M. Lie, "Gofat Wordpress," [Online]. Available: https://gofat.wordpress.com/2012/04/12/face-detection-dengan-metoda-haar-cascade/. [Accessed 2015]. |
| [9] | R. Cucchiar, "Ist Psu Education," [Online]. Available: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.454.2188&rep=rep1&type=pdf. [Accessed 2015]. |
| [10] | "Thesis Binus," 2015. [Online]. Available: http://thesis.binus.ac.id/Asli/Bab2/2007-3-00373-MTIF%20Bab%202.pdf. |
| [11] | C. A. B. Larsen, "Fast and effective text mining using linear-time document clustering," *Proceedings of the Fifth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM,* p. 16–22, 1999. |
| [12] | J. G. A. Strehl, "Cluster ensembles—a knowledge reuse framework for combining multiple partitions," *J. Machine Learn. Res. 3 ,* p. 583– 617, 2003. |